Importance Sampling for Event Timing Models

Kay Giesecke* and Alexander Shkolnik†

September 8, 2011

This version: October 31, 2011‡

Abstract

This paper provides an efficient Monte Carlo method for estimating rare-event probabilities in point process models of correlated event timing, which have applications in finance, insurance, engineering, and many other areas. It develops an importance sampling scheme for the tail of the distribution of the total event count at a fixed horizon, and provides conditions guaranteeing the asymptotic optimality of the resulting estimator. The change of measure differs from the widely used exponential twisting. The algorithm applies to point process models with arbitrary stochastic intensity dynamics. Numerical tests illustrate its performance.

*Department of Management Science & Engineering, Stanford University, Stanford, CA 94305, USA, Phone (650) 723 9265, Fax (650) 723 1614, email: giesecke@stanford.edu, web: www.stanford.edu/∼giesecke .

†Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA 94305, USA, email: ads2@stanford.edu.

‡We thank conference participants at the 2010 Fields Institute Workshop on Numerical Methods in Finance, the 2010 INFORMS Annual Meeting in Austin, the 4th Western Conference in Mathematical Finance, and the 7th International Congress on Industrial and Applied Mathematics for comments. We are also grateful to Chris Rogers and Stefan Weber for insightful discussions.

1 Introduction

This paper develops an efficient Monte Carlo method for estimating rare-event probabilities in stochastic models of event timing, which are common in many disciplines. In finance, event timing models are used to measure the credit risk in a portfolio of defaultable assets such as loans and corporate bonds, value securities exposed to default risk, and analyze security order books on exchanges. In insurance, event timing models are used to price policies and estimate reserves. In engineering, event timing models are used to evaluate performance measures of component systems. At the center of many applications are rare events, i.e., the atypical behavior of the underlying system. For example, the measurement of portfolio credit risk focuses on the likelihood of a large number of defaults. Insurance risk management centers around the probability of ruin due to a large number of claims. Reliability studies emphasize the likelihood of systemic breakdown caused by the failure of a sufficiently large number of components.

Monte Carlo methods are widely used to analyze event timing models. Standard Monte Carlo simulation is, however, highly inefficient for estimating rare-event probabilities. This is because the number of simulation trials required to estimate the probability of interest to a given relative precision scales in rough proportion to one over the square-root of the probability. A large number of trials may be required to obtain sufficiently accurate estimates of rare-event probabilities.

This paper develops and evaluates a provably efficient importance sampling (IS) scheme for estimating rare-event probabilities in event timing models. An event timing model is represented by a point process N with values in $\{0,1\}^n$. The components of N are indicator processes of n different events, which are typically correlated. A component jumps from 0 to 1 at the arrival of a given event, such as the default of a firm, the placement of a security order, the filing of an insurance claim, or the failure of a system component. It is governed by a stochastic intensity process, which represents the conditional event arrival rate. We do not impose a particular structure on the vector intensity process. Therefore, the IS scheme we propose applies to virtually any stochastic intensity model formulated in the application areas discussed above. This is a feature that distinguishes it from the rare-event schemes developed in prior work.

Motivated by the applications outlined above, we focus on estimating the tail of the total event count C_T at a horizon T. More precisely, we consider the probability that C_T exceeds some fraction of n in the asymptotic regime $n \to \infty$. Our IS scheme for this probability differs from the exponential twisting schemes widely used in the rare-event literature. When applied in our setting, exponential twisting amounts to directly transforming the law of C_T. The asymptotically optimal twist is suggested by the Gärtner-Ellis theorem. Its practical implementation requires knowledge of the transform of C_T. This transform is, however, difficult to compute for many stochastic intensity models of interest. This limits the scope of exponential twisting for event timing models.

We propose IS schemes that entail a change of measure induced by "twisting" the intensity of N. A particularly convenient family of intensity twists transforms the process C counting the events of N into a stopped Poisson process. The variance-minimizing Poisson rate is asymptotically optimal, but only under restrictive conditions. An alternative family of intensity twists results in a shift to the intensity of C. Asymptotic optimality is shown to hold under relatively mild conditions. The optimal twist satisfies a saddle point condition for a generalized birth process with random birth rates given by the reference measure intensity of C evaluated just prior to an event. This measure change reduces to exponential twisting if C is a stopped birth process under the reference measure, in which case exponential twisting is indeed implementable. Unlike exponential twisting, however, our measure change is implementable even in the case of a general intensity model, because it does not require knowledge of the transform of C. Sampling under our importance measure is not harder than sampling under the reference measure.

Numerical experiments illustrate the performance of our IS schemes for an empirically motivated point process model of correlated default timing. In this model, the intensities of the components of N satisfy a system of coupled jump-diffusion SDEs whose diffusion and jump terms are correlated across the components. The numerical results indicate that the second IS scheme outlined above performs well even for portfolios subject to significant contagion (i.e., self-exciting) effects.

The rest of this introduction discusses prior work. Section 2 formulates the IS problem. Section 3 discusses an exponential twisting approach. Section 4 develops and analyzes intensity twisting. Section 5 provides conditions guaranteeing asymptotic optimality and Section 6 analyzes the optimal twist. Section 7 explains the IS algorithm and Section 8 provides numerical results. Section 9 concludes. Proofs are in the Appendix.

1.1 Related literature

Previous research has studied rare-event algorithms for point process models of event timing. Bassamboo & Jain (2006) develop an asymptotically optimal IS scheme for doubly-stochastic models. In these models, a component intensity is a function of independent risk factor processes, some of which are common to all components. Conditional on the common factors, the event times are independent, facilitating the application of exponential twisting to the factors and the conditional count. Our IS estimators are constructed differently; they also apply when the doubly-stochastic assumption fails. This renders them useful for the analysis of self-exciting models, in which an event may have a direct impact on the intensities of other events. Such formulations are especially important in the analysis of portfolio credit risk, where they address contagion effects, and reliability, where they address the interaction of system components.

Several articles analyze rare-event schemes for Markov chain models, in which a component intensity is a function of time and state of the point process. Giesecke & Shkolnik (2010) develop an asymptotically optimal IS scheme. Their measure change transforms the event counting process into a stopped Poisson process. Carmona & Crépey (2010) examine an interacting particle scheme (Del Moral & Garnier (2005)). Carmona, Fouque & Vestal (2009) study an interacting particle scheme for a model in which the event times are the hitting times of correlated geometric Brownian motions.

Giesecke, Kakavand, Mousavi & Takada (2010) develop an interacting particle scheme, and Deng, Giesecke & Lai (2011) an asymptotically optimal sequential resampling scheme for event timing models with arbitrary component intensities. The algorithms proposed in these papers are formulated in terms of a mimicking Markov chain M whose value at t has the same distribution as N_t, for any fixed t. The construction of M boils down to the computation of a conditional expectation of a component intensity of N. For some model families, the computation is straightforward. For others, it is burdensome, which degrades performance. There are also formulations for which the computation seems difficult if not intractable. The IS scheme developed in this paper eliminates the need to construct a mimicking chain; it is formulated in terms of N.

Zhang, Blanchet, Giesecke & Glynn (2011) develop an asymptotically optimal IS scheme for affine point processes, i.e., point processes with intensities driven by an affine jump-diffusion. They consider a "large horizon" asymptotic regime. Our IS estimators are designed for indicator point processes rather than non-terminating point processes. Moreover, we consider a "large system" asymptotic regime.

There is a substantive literature on efficient IS for highly reliable Markovian systems, see Goyal, Shahabuddin, Heidelberger, Nicola & Glynn (1992), Juneja & Shahabuddin (2001), Shahabuddin (1994), Nakayama (1996), and many others. Here, a component system is modeled by a time-homogenous continuous-time Markov chain, and the analysis is based on the embedded discrete-time Markov chain. We take a point process approach to event timing (without a role for repair), permitting an arrival to depend also on exogenous sources of randomness governed by SDEs, and time. In our formulation, the inter-event times need not be exponentially distributed, but are allowed to follow a much more general distribution as implied by the vector SDE governing the component intensities. While we analyze the system as the number of components increases, the aforementioned articles parametrize the transition rates by a rarity parameter, and analyze the behavior of the system when these rates tend to 0.

There is also a large literature on efficient IS for copula models of event timing, see Bassamboo, Juneja & Zeevi (2008), Chen & Glasserman (2008), Joshi & Kainth (2004), Glasserman & Li (2005), Glasserman, Kang & Shahabuddin (2008), and others. This literature focuses on portfolio credit risk; the joint distribution of event times is modeled by a copula function and marginal distributions. The specification results in conditionally independent event times, a feature the IS schemes exploit.

2 Event timing and importance sampling

Fix a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a right-continuous and complete information filtration $(\mathcal{F}_t)_{t\geq 0}$. Consider n almost surely distinct stopping times $\tau^i > 0$. These represent the arrival times of events such as corporate defaults, insurance claims, or component failures. Associated to the τ^i are indicator processes N^i given by $N^i_t \triangleq \mathbb{1}(\tau^i \leq t)$, where $\mathbb{1}(A)$ is the indicator function of an event $A \in \mathcal{F}$. The indicator point process $N \triangleq (N^1, \dots, N^n) \in \{0,1\}^n$ is a stochastic model for the timing of n events.

We suppose that each N^i admits an intensity. More precisely, for each i there is an integrable and progressively measurable process p^i which is strictly positive on $[0, \tau^i)$ and 0 elsewhere, such that a martingale is defined by

$$N^i_t - \int_0^t p^i(s)\, ds. \tag{1}$$

The process p^i is called the intensity of N^i. It represents the conditional event arrival rate. For example, if $p^i(s) = \alpha\,\mathbb{1}(\tau^i > s)$ for some constant $\alpha > 0$, then N^i indicates the first jump of a Poisson process, and the τ^i are independent exponentially distributed random variables. If $p^i(s) = f(s, N_s)$ for a function $f > 0$, then N^i is a time-inhomogenous Markov chain. More generally, the intensity process $p = (p^1, \dots, p^n)$ is governed by a system of coupled SDEs that incorporates also other sources of randomness. For example, p may follow a jump-diffusion process whose diffusion and jump terms are correlated across the components. The correlation between the p^i induces a dependence structure for the vector N. In this paper we do not impose a specific structure on p.

Motivated by the applications outlined in the Introduction, we focus on estimating the tail of the total event count in the system at some fixed horizon. More precisely, we consider the asymptotic regime $n \to \infty$. For fixed $T > 0$ and $\mu \in (0,1)$, we analyze the sequence of events $\{\xi_n\}_1^\infty$, where $\xi_n \triangleq \{C_T \geq \mu n\}$ and $C \triangleq 1_n \cdot N$ is the process counting the events of N.[1] We assume that $\{\xi_n\}$ is a *rare event* sequence:

$$y_n \triangleq \mathbb{P}(\xi_n) = \mathbb{P}(C_T \geq \mu n) \to 0 \quad \text{as} \quad n \to \infty\,. \tag{2}$$

The relative error of the plain Monte Carlo estimator $\mathbb{1}(\xi_n)$ of the rare-event probability y_n tends to infinity asymptotically. Thus, the number of simulation trials required to achieve a given relative precision tends to infinity with n, implying that the estimator $\mathbb{1}(\xi_n)$ is highly inefficient. We develop an importance sampling (IS) estimator of the rare-event probabilities $\{y_n\}$ that addresses this issue.

[1] Depending on the application, other events may be of interest. In portfolio credit risk, for example, one may be interested in the event that the loss from default exceeds a large level. This is the event $\{\ell_n \cdot N \geq x\}$, where ℓ_n is an n-vector of positive random variables. A component of ℓ_n describes the loss at a default. Under mild conditions on ℓ_n, the approach described below extends to this event. For clarity, below we focus on the event ξ_n. Details of the extension are available upon request.

IS entails an absolutely continuous change of measure from $\mathbb{P}$ to an importance measure $\mathbb{Q}$, and uses the identity

$$y_n = \mathbb{E}_{\mathbb{Q}}\left(Z_n \mathbb{1}(\xi_n)\right) \tag{3}$$

where Z_n is the Radon-Nikodym derivative of $\mathbb{P}$ with respect to $\mathbb{Q}$ and $\mathbb{E}_{\mathbb{Q}}$ denotes expectation with respect to $\mathbb{Q}$. The corresponding IS estimator of y_n is defined by

$$Y_n \triangleq Z_n \mathbb{1}(\xi_n). \tag{4}$$

It satisfies $\mathbb{E}_{\mathbb{Q}}(Y_n) = \mathbb{P}(\xi_n) = y_n$, as dictated by (3). Our goal is to design an importance measure $\mathbb{Q}$ such that Y_n is "more efficient" than the plain MC estimator $\mathbb{1}(\xi_n)$. To measure the efficiency of Y_n, we consider its relative error, defined by

$$\varsigma_n \triangleq \sqrt{\mathbb{E}_{\mathbb{Q}}(Y_n^2) \,/\, y_n^2 - 1}\,. \tag{5}$$

Note that if $\mathbb{Q} = \mathbb{P}$, we have $Y_n = \mathbb{1}(\xi_n)$. In this case, the relative error $\sqrt{1/y_n - 1} \to \infty$ in the rare event regime (2). Ideally, one would seek an importance measure $\mathbb{Q}$ for which ς_n is bounded. However, this property is very hard to achieve, and typically $\varsigma_n \to \infty$ as $n \to \infty$. We consider a weaker form of efficiency for which the exponential rate of growth of $\varsigma_n \uparrow \infty$ is asymptotically negligible relative to the exponential rate of decay of the estimator y_n. More precisely, $\liminf_{n\to\infty} \log(\varsigma_n)/\log(y_n) = 0$ or equivalently

$$\liminf_{n\to\infty} \frac{\log \mathbb{E}_{\mathbb{Q}}(Y_n^2)}{\log(y_n)} = 2\,. \tag{6}$$

An IS estimator Y_n satisfying condition (6) is called *asymptotically optimal*.

3 Conventional exponential twisting

Letting $\{S_k\}$ denote the arrival times of C, we may write $\xi_n = \{C_T \geq \mu n\} = \{S_{\lceil \mu n \rceil} \leq T\}$. The arrival time $S_{\lceil \mu n \rceil}$ is a sum of $\lceil \mu n \rceil$ random variables. This property suggests a classical approach to the design of $\mathbb{Q}$, which goes back to Siegmund (1976). Consider the cumulant generating function (CGF) of the scaled $\lceil \mu n \rceil$th arrival time $nS_{\lceil \mu n \rceil}$, given by

$$\Lambda_n(\theta) \triangleq \log \mathbb{E}\left(\exp\left(-\theta n S_{\lceil \mu n \rceil}\right)\right). \tag{7}$$

For finite $\Lambda_n(\theta)$ we take the Radon-Nikodym derivative Z_n as

$$Z_n(\theta) = \exp\left(\theta n S_{\lceil \mu n \rceil} + \Lambda_n(\theta)\right). \tag{8}$$

This choice is known as exponential twisting of $nS_{\lceil \mu n \rceil}$ and the parameter θ is called the exponential twist; see Asmussen & Glynn (2007) for a discussion. We write $\mathbb{Q} = \mathbb{Q}_\theta$ to highlight the dependence of the importance measure on θ and denote by $\mathbb{E}_\theta$ its expectation.

Since $S_{\lceil \mu n \rceil} \le T$ on ξ_n, it follows that for $\theta \ge 0$ the second moment under $\mathbb{Q}_\theta$ of the estimator $Y_n = Z_n(\theta)\,\mathbb{1}(\xi_n)$ satisfies the upper bound

$$\frac{1}{n} \log \mathbb{E}_\theta \left(Y_n^2\right) \le 2\theta T + 2\frac{1}{n}\Lambda_n(\theta). \tag{9}$$

The decay of the probability $\mathbb{P}(\xi_n)$ can be analyzed using the Gärtner-Ellis theorem from large deviations theory. We define the scaled limiting CGF

$$\Lambda(\theta) \triangleq \lim_{n\to\infty} \frac{1}{n}\Lambda_n(\theta)$$

and make the following standard assumption.

Assumption 3.1. *The limit $\Lambda(\theta)$ exists as an extended real number for all $\theta \in \mathbb{R}$ and is finite in some neighborhood of the origin. Moreover, there exists some $\theta^* > 0$ such that $\Lambda(\cdot)$ is continuously differentiable in some neighborhood of θ^* with $\nabla\Lambda(\theta^*) = -T$.*

Under these conditions, the Gärtner-Ellis theorem (Theorem 2.3.6 in Dembo & Zeitouni (1998)) and the argument used to obtain (9) imply that

$$\liminf_{n\to\infty} \frac{1}{n} \log\left(\mathbb{P}(\xi_n)\right) = \theta^* T + \Lambda(\theta^*). \tag{10}$$

From this, condition (6) readily follows because (9) and Jensen's inequality imply that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}_{\theta^*}\left(Y_n^2\right) = 2\theta^* T + 2\Lambda(\theta^*), \tag{11}$$

and consequently, the IS estimator Y_n with $\theta = \theta^*$ is asymptotically optimal. The optimal twist θ^* has an intuitive interpretation. First, dominated convergence implies

$$-\frac{1}{n}\nabla\Lambda_n(\theta n) = \mathbb{E}_\theta\left(S_{\lceil \mu n \rceil}\right)$$

and, whenever we can interchange the limit and the differential operator, $\nabla\Lambda(\theta^*) = -T$. It follows that $\mathbb{E}_{\theta^*}\left(S_{\lceil \mu n \rceil}\right) = T$ in the limit as $n \to \infty$, which is commonly referred to as the saddle point condition. In fact, one can show that $S_{\lceil \mu n \rceil} \to T$ in $\mathbb{Q}_{\theta^*}$-probability, implying that

$$\frac{1}{n} \log Z_n(\theta^*) \to \theta^* T + \Lambda(\theta^*) \tag{12}$$

as $n \to \infty$ in $\mathbb{Q}_{\theta^*}$-probability. Since the Radon-Nikodym derivative of $\mathbb{P}$ with respect to the zero variance measure $\mathbb{P}(\cdot \mid \xi_n)$ is constant on the rare event, (12) suggests, in the same sense, that $\mathbb{Q}_{\theta^*}$ approximates the zero variance measure. See Sadowsky & Bucklew (1990) for further discussion of the use of large deviations theory for optimal IS.

While generating an asymptotically optimal IS estimator, exponential twisting has a narrow scope in the point process setting. Its implementation requires the computation of the CGF of $nS_{\lceil \mu n \rceil}$, which only few models p of the intensity permit. For example, suppose the component intensities $p^i(s) = \alpha^i \mathbb{1}(s < \tau^i)$ for constants $\alpha^i > 0$, in which case C is a birth process stopped at S_n. Then $S_{\lceil \mu n \rceil}$ is a sum of $\lceil \mu n \rceil$ independent exponential random variables and has a gamma distribution. Also tractable is the Markov chain case, for which $p^i(s)$ is some deterministic function of N_s that vanishes for $s \geq \tau^i$; see Bucklew, Ney & Sadowsky (1990). Finally, a doubly-stochastic formulation, in which $p^i(s)$ for $s < \tau^i$ is a deterministic function of a factor process common to all components, may permit the computation of the CGF by a conditioning argument. In this formulation, (8) is an alternative to the asymptotically optimal measure change for doubly-stochastic models proposed by Bassamboo & Jain (2006). Their measure change is based on exponential twisting of the law of the integrated factor process and of the conditional law of the event count given the integrated factors.

4 Intensity twisting

Instead of directly applying exponential twisting to the scaled time $nS_{\lceil \mu n \rceil}$, we propose to induce a change of measure by "twisting" the $\mathbb{P}$-intensity process $p = (p^1, \ldots, p^n)$. Twisting the intensity changes the dynamics of the point process N, and this results in an adjustment to the law of $S_{\lceil \mu n \rceil}$. Intuitively, the intensity needs to be increased in order to transform ξ_n into a non-rare event under the importance measure $\mathbb{Q}$.

4.1 Radon-Nikodym derivative

Let $q = (q^1, \ldots, q^n)$ where $q^i(s) \triangleq \lambda_n(s) p^i(s)$ for $s \geq 0$ and λ_n is some nonnegative, progressively measurable twisting process. Take the Radon-Nikodym derivative as

$$Z_n = \exp(K_n) \exp\left(- \int_0^{S_{\lceil \mu n \rceil}} (1 - \lambda_n(s)) p_n(s) \, ds \right) \tag{13}$$

where $p_n \triangleq 1_n \cdot p$ is the $\mathbb{P}$-intensity of the counting process C and

$$K_n = - \sum_{k=1}^{\lceil \mu n \rceil} \log(\lambda_n(S_k -)) \, . \tag{14}$$

Proposition 4.1. *Suppose the twisting process $\lambda_n \geq 1$ and $\mathbb{E}(\exp(-K_n)) < \infty$. Then $\mathbb{Q}$ defined via (13) is a probability measure and N admits $\mathbb{Q}$-intensity $q(s) = \lambda_n(s) p(s)$ on the interval $[0, S_{\lceil \mu n \rceil})$.*

Remark 4.2. *The twisting process λ_n applies to all elements of the vector p. Element specific twisting processes could be introduced to distinguish between different components of N. However, our formulation suffices because the event of interest $\xi_n = \{C_T \geq \mu n\}$ depends only on the number of events of N at T; the component configuration is irrelevant.*

Remark 4.3. *A sufficient condition for* $\mathbb{E}(\exp(-K_n))$ *to be finite is*

$$\mathbb{E}\left(\int_0^{S_n} \lambda_n(s)^n p_n(s)\, ds\right) < \infty. \tag{15}$$

This follows from the fact that N *admits* $\mathbb{P}$*-intensity* p*, thus*

$$\mathbb{E}(\exp(-K_n)) \leq \sup_i \mathbb{E}\left(\int_0^\infty \lambda_n(s-)^n dN_s^i\right) = \sup_i \mathbb{E}\left(\int_0^\infty \lambda_n(s)^n p^i(s)\, ds\right)$$

where we sum over i *to obtain the desired upper bound. Observe that if* λ_n *is almost surely bounded above by some constant* c*, the left side of* (15) *is bounded above by* $c^n n$ *by Meyer's time change theorem.*

We propose and analyze two different specifications for the twisting process λ_n.

4.2 Poisson measure change

Adopting the convention $c/0 = \infty$ for any $c > 0$, take

$$\lambda_n(s) \triangleq \gamma / p_n(s) \tag{16}$$

for all $s \geq 0$ where $\gamma > 0$ is a constant to be determined and $p_n = 1_n \cdot p$ is the $\mathbb{P}$-intensity of the counting process C. We write $\mathbb{Q} = \mathbb{Q}_\gamma$ to indicate the dependence of the importance measure on γ and let $\mathbb{E}_\gamma$ denote its expectation. Under the hypotheses of Proposition 4.1, on the interval $[0, S_{\lceil \mu n \rceil})$ the $\mathbb{Q}_\gamma$-intensity of N^i takes the form

$$q^i(s) = p^i(s) \frac{\gamma}{p_n(s)}. \tag{17}$$

The $\mathbb{Q}_\gamma$-intensity of C is $q_n \triangleq 1_n \cdot q = \gamma$ on $[0, S_{\lceil \mu n \rceil})$ so C is a $\mathbb{Q}_\gamma$-Poisson process with rate γ on this interval. Moreover, at event time S_k, the component $I_k \in \{1, \ldots, n\}$ of N at which the kth event occurs has conditional distribution

$$\mathbb{Q}_\gamma(I_k = i \mid \mathcal{F}_{S_k-}) = \frac{p^i(S_k-)}{p_n(S_k-)} = \mathbb{P}(I_k = i \mid \mathcal{F}_{S_k-}) \qquad i = 1, 2, \ldots, n. \tag{18}$$

The second equality follows from an argument similar to the one used to prove Theorem II.T15 of Brémaud (1980). The relative ordering property (18) and the Poisson property of C render the generation of the IS estimator $Y_n = Z_n \mathbb{1}(\xi_n)$ straightforward. The relative ordering property also ensures that the Radon-Nikodym derivative Z_n behaves well with respect to the component configuration of N, in the sense that the ratio of the right side of (18) to the left side is bounded.

The Radon-Nikodym derivative (13) takes the form

$$Z_n(\gamma) = \exp\left(\gamma S_{\lceil \mu n \rceil} - \lceil \mu n \rceil \log(\gamma) + \sum_{k=1}^{\lceil \mu n \rceil} \log(p_n(S_k-)) - \int_0^{S_{\lceil \mu n \rceil}} p_n(s)\, ds\right), \tag{19}$$

allowing us to analyze the dependence of the $\mathbb{Q}_\gamma$-variance of the IS estimator $Y_n = Z_n(\gamma)\,\mathbb{1}(\xi_n)$ on the twisting parameter γ. Since $S_{\lceil \mu n \rceil} \leq T$ on ξ_n, for $\gamma \geq 1$ we obtain the upper bound

$$\log \mathbb{E}_\gamma\left(Y_n^2\right) \leq -\lceil \mu n \rceil \log \gamma + \gamma T + \log \mathbb{E}\left(Z_n(1) \exp\left(-S_{\lceil \mu n \rceil}\right) \mathbb{1}(\xi_n)\right) \tag{20}$$

which is minimized by $\gamma^* = \lceil \mu n \rceil / T$. This Poisson rate is variance-optimal in the sense that (20) is tight in probability since $S_{\lceil \mu n \rceil} \to T$ in $\mathbb{Q}_{\gamma^*}$-probability, as is the case for $\mathbb{Q}_{\theta^*}$ To verify this, we can apply the Lindeberg CLT to $S_{\lceil \mu n \rceil}$, the sum of $\lceil \mu n \rceil$ i.i.d. exponential random variables, each with parameter γ^*. Furthermore, γ^* is the unique rate satisfying $\mathbb{E}_{\gamma^*}\left(S_{\lceil \mu n \rceil}\right) = T$, the saddle point condition suggested by exponential twisting.

Observe that if C is a $\mathbb{P}$-Poisson process stopped at S_n with intensity $p_n(s) = \alpha\,\mathbb{1}(S_n > s)$ for some constant $\alpha > 0$, then $Z_n(\gamma)$ simplifies to

$$\begin{aligned} Z_n(\gamma) &= \exp\left((\gamma - \alpha)\, S_{\lceil \mu n \rceil} - \lceil \mu n \rceil \log(\alpha/\gamma)\right) \\ &= \exp\left((\gamma - \alpha)\, S_{\lceil \mu n \rceil} + \Lambda_n\left((\gamma - \alpha)/n\right)\right) \end{aligned}$$

where Λ_n is the CGF of $nS_{\lceil \mu n \rceil}$, given by (7). Thus, in this setting the measure change is equivalent to exponential twisting of $nS_{\lceil \mu n \rceil}$ with twist $\theta = (\gamma - \alpha)/n$. It is asymptotically optimal to take $\gamma = \gamma^*$ in this case. This derivation suggests that the IS estimator under $\mathbb{Q}_{\gamma^*}$ is close to optimal whenever the $\mathbb{P}$-intensity of C is roughly constant asymptotically. True conditions for optimality are indeed restrictive. Note that Proposition 4.1 already requires that $p_n \leq \gamma$ almost surely under $\mathbb{P}$. We thus expect this measure change to work well only for bounded intensity specifications with small variance.

4.3 An alternative measure change

Consider the choice

$$\lambda_n(s) \triangleq \beta n / p_n(s) + 1 \tag{21}$$

for $s \geq 0$ and a constant $\beta \geq 0$ to be determined. We scale the intensity twist by n since the variance-optimal Poisson rate γ^* is $O(n)$. We write $\mathbb{Q} = \mathbb{Q}_\beta$ and let $\mathbb{E}_\beta$ denote the associated expectation. Under the hypotheses of Proposition 4.1, on the interval $[0, S_{\lceil \mu n \rceil})$ the $\mathbb{Q}_\beta$-intensity of N^i takes the form

$$q^i(s) = p^i(s)\left(\frac{\beta n}{p_n(s)} + 1\right). \tag{22}$$

The $\mathbb{Q}_\beta$-intensity of C is $q_n(s) = p_n(s) + \beta n$ on $[0, S_{\lceil \mu n \rceil})$ so our measure change results in a shift of the intensity of C by βn. The relative ordering property (18) continues to hold under the specification (21).

The Radon-Nikodym derivative of $\mathbb{P}$ with respect to $\mathbb{Q}_\beta$ is

$$Z_n(\beta) = \exp\left(\beta n S_{\lceil \mu n \rceil} + K_n(\beta)\right) \tag{23}$$

where we indicate the dependence of K_n in (14) on β:

$$K_n = K_n(\beta) = -\sum_{k=1}^{\lceil \mu n \rceil} \log\left(1 + \beta n / p_n(S_k-)\right). \tag{24}$$

Observe that if C is a birth process under $\mathbb{P}$ with intensity $p_n(s) = n\alpha_n^{C_s}$ specified by a sequence $\{\alpha_n^k\}$ of positive constants with $\alpha_n^n = 0$, then this measure change corresponds to exponential twisting. To see this, compute the corresponding CGF of $nS_{\lceil \mu n \rceil}$ as

$$\Lambda_n(\beta) = \log \mathbb{E} \exp\left(-\beta n S_{\lceil \mu n \rceil}\right) = -\sum_{k=1}^{\lceil \mu n \rceil} \log\left(1 + \beta / \alpha_n^{k-1}\right) = K_n(\beta) \tag{25}$$

which verifies that (23) is equivalent to exponential twisting with twist $\theta = \beta$. If $\inf_k \alpha_n^k$ and $\sup_k \alpha_n^k$ are strictly positive and finite, differentiation and limit can be interchanged, and the optimal exponential twist satisfies the saddle point condition $\mathbb{E}_\beta\left(S_{\lceil \mu n \rceil}\right) = T$ in the limit as $n \to \infty$, where

$$\mathbb{E}_\beta\left(S_{\lceil \mu n \rceil}\right) = \frac{1}{n} \sum_{k=1}^{\lceil \mu n \rceil} \frac{1}{\alpha_n^{k-1} + \beta}.$$

In the general case, where C is not necessarily a birth process, we may compute $\mathbb{E}_\beta\left(S_{\lceil \mu n \rceil}\right)$ by noting that $0 = \nabla \mathbb{E}\left(Z_n(\beta)^{-1}\right)$. This, by dominated convergence under the hypotheses of Proposition 4.1 and Assumption 3.1, leads to

$$0 = \mathbb{E}\left(\nabla Z_n(\beta)^{-1}\right) = -n\mathbb{E}_\beta\left(S_{\lceil \mu n \rceil}\right) - \mathbb{E}_\beta\left(\nabla K_n(\beta n)\right)$$

which implies the identity

$$\mathbb{E}_\beta\left(S_{\lceil \mu n \rceil}\right) = \sum_{k=1}^{\lceil \mu n \rceil} \mathbb{E}_\beta\left(1/q_n(S_k-)\right). \tag{26}$$

In Sections 6 and 7, (26) will be used to compute β at the saddle point.

5 Asymptotic Optimality

The relation between the importance measure $\mathbb{Q}_\beta$ and the exponential twisting measure $\mathbb{Q}_\theta$ is the key to proving asymptotic optimality. Recall from (12) that

$$\frac{1}{n} \log Z_n(\theta^*) = \theta^* S_{\lceil \mu n \rceil} + \frac{1}{n} \Lambda_n(\theta^*)$$

converges in $\mathbb{Q}_{\theta^*}$-probability to the exponential rate of decay of $\mathbb{P}(\xi_n)$. For $\beta \geq 0$ the Radon-Nikodym derivative of $\mathbb{P}$ with respect to $\mathbb{Q}_\beta$ satisfies

$$\frac{1}{n}\log Z_n(\beta) = \beta S_{\lceil \mu n \rceil} + \frac{1}{n}K_n(\beta). \tag{27}$$

The random variable K_n is the stochastic analog of the CGF Λ_n. If (27) converges in $\mathbb{Q}_\beta$-probability we would like the limit to be one in (12) which suggests setting $\beta = \theta^*$. We will show that, under some assumptions on the rate of convergence in (27), this setting yields an asymptotically optimal IS estimator $Y_n = Z_n(\beta)\,\mathbb{1}(\xi_n)$.

Define the $\mathbb{Q}_\beta$-CGF of $n^{-1}K_n(\beta)$ by

$$\Psi_n(\rho) \triangleq \log \mathbb{E}_\beta\big(\exp\big(-\rho n^{-1}K_n(\beta)\big)\big). \tag{28}$$

The following result, variants of which can be found in Sadowsky & Bucklew (1990), Dembo & Zeitouni (1998, Exercise 2.3.25) and others, is useful.

Lemma 5.1. *Let $\{c_n\}$ be a sequence with $c_n \to \infty$ such that the limit*

$$\Psi(\rho) = \limsup_{n\to\infty} \frac{1}{c_n}\Psi_n(\rho c_n)$$

is finite in some neighborhood of the origin with $\nabla\Psi(0) = -K(\beta)$. Then,

$$\frac{1}{n}K_n(\beta) \to K(\beta) \tag{29}$$

in $\mathbb{Q}_\beta$-probability, exponentially in c_n.

By (23), we have that

$$\Psi_n(-n) = \log \mathbb{E}_\beta(\exp(K_n(\beta))) = \log \mathbb{E}\left(\exp\left(-\beta n S_{\lceil \mu n \rceil}\right)\right) = \Lambda_n(\beta), \tag{30}$$

and it follows by Assumption 3.1 that for $\beta = \theta^*$ the function $\Psi_n(\cdot)$ is finite on $[-n, \infty)$ for all n sufficiently large. Markov's inequality implies that

$$\limsup_{n\to\infty} \frac{1}{n}\log \mathbb{Q}_\beta\left(\frac{1}{n}K_n(\beta) \geq \Lambda(\beta) + \delta\right) \leq -\delta$$

so $n^{-1}K_n(\beta) \to K(\beta)$ in $\mathbb{Q}_\beta$-probability leads to

$$K(\beta) \leq \Lambda(\beta). \tag{31}$$

Since $S_{\lceil \mu n \rceil} \leq T$ on ξ_n we see that Y_n assumes large values (larger than order $\mathbb{P}(\xi_n)$) with exceedingly small probabilities under $\mathbb{Q}_\beta$ with $\beta = \theta^*$ as $n \to \infty$.

We can also apply Lemma 5.1 to $S_{\lceil \mu n \rceil}$. Proposition 4.1 requires that $\mathbb{E}(\exp(-K_n(\beta)))$ is finite. This implies that the $\mathbb{Q}_\beta$-CFG of $nS_{\lceil \mu n \rceil}$ is finite. Under conditions of Lemma

5.1 in terms of the CGF of $nS_{\lceil \mu n \rceil}$ having derivative $-T$ at the origin, $S_{\lceil \mu n \rceil} \to T$ in $\mathbb{Q}_\beta$-probability. When $\beta = \theta^*$ we have that (27) tends to the exponential rate of decay of $\mathbb{P}(\xi_n)$. However, as explained in Glasserman & Wang (1997), this convergence does not guarantee asymptotic optimality.

To prove asymptotic optimality, we observe that the second moment of Y_n satisfies

$$\frac{1}{n} \log \mathbb{E}_\beta \left(Y_n^2 \right) = \frac{1}{n} \log \mathbb{E}_\beta \left(\exp \left(2\beta n S_{\lceil \mu n \rceil} + 2K_n(\beta) \right) \mathbb{1}(\xi_n) \right).$$

We will show that if $c_n/n \to \infty$, then

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{E}_\beta \left(Y_n^2 \right) \le 2\beta T + 2K(\beta) \le 2\beta T + 2\Lambda(\beta)$$

where the second inequality follows from (31). The upper bound is minimized by $\beta = \theta^*$, yielding asymptotic optimality of Y_n. The above discussion is made precise in the following optimality theorem.

Theorem 5.2. *Suppose the conditions of Lemma 5.1 hold with $\beta = \theta^*$ of Assumption 3.1 and some sequence $\{c_n\}$ such that $c_n/n \to \infty$. Then, the IS estimator $Y_n = Z_n(\beta^*) \mathbb{1}(\xi_n)$ is asymptotically optimal for $\beta^* = \theta^*$.*

Remark 5.3. *The main condition for asymptotic optimality is the weak dependence of the random variables $\{n^{-1} p_n(S_k-)\}$ under $\mathbb{Q}_{\beta^*}$. Recall that*

$$K_n(\beta^*) = -\sum_{k=1}^{\lceil \mu n \rceil} \log \left(1 + \beta^* n / p_n(S_k-) \right).$$

Thus, optimality requires the weak law of large numbers $n^{-1} K_n(\beta^*) \to K(\beta^*)$ under $\mathbb{Q}_{\beta^*}$ to apply with convergence exponentially in c_n such that $c_n/n \to \infty$.

6 The optimal intensity twist

The optimal intensity twist β^* of the previous section is prescribed by the equation $\nabla K(\beta^*) = -T$. We have shown under suitable conditions that $K(\beta) = \Lambda(\beta)$ so $\beta^* = \theta^*$, the optimal exponential twist of Section 3. Here, under SLLN assumptions, we relate the optimal twist to the values of the process p prior to the event times under $\mathbb{Q}_{\beta^*}$. Based on these results we will be able to design an optimal IS scheme.

We define the variables

$$\rho_n^k \triangleq \frac{1}{n} p_n(S_k-) = \frac{1}{n} \sum_{i=1}^{n} p^i(S_k-) \tag{32}$$

where the k superscript indicates that the sum above contains precisely $n-k$ strictly positive terms. We also define the means

$$\alpha_n^k(\beta) \triangleq \mathbb{E}_\beta\big(\rho_n^k\big) \tag{33}$$

and make the following assumption.

Assumption 6.1. *Each $\alpha_n^k(\beta)$ is continuously differentiable on $[0,\mu/T]$. For each $k \in \mathbb{N}$ the limit $\alpha^k(\beta) \triangleq \lim_{n\to\infty} \alpha_n^k(\beta)$ exists on $[0,\mu/T]$ with $\inf_k \alpha^k(\beta)$ and $\sup_k \alpha^k(\beta)$ finite and bounded away from zero. Moreover,*

$$\sup_k \big|\, \rho_n^k - \alpha_n^k(\beta) \,\big| \to 0$$

$\mathbb{Q}_\beta$-almost surely as $n \to \infty$ for all $\beta \in [0,\mu/T]$.

Recall from the discussion in Section 3 that for large values of n we approximate θ^* by the solution of $\mathbb{E}_\theta\big(S_{\lceil \mu n\rceil}\big) = T$, the saddle point condition. By analogy we may approximate β^* by solving $\mathbb{E}_\beta\big(S_{\lceil \mu n\rceil}\big) = T$. Recalling the identity (26), we have for $\beta > 0$

$$\left| \mathbb{E}_\beta\big(S_{\lceil \mu n\rceil}\big) - \frac{1}{n}\sum_{k=1}^{\lceil \mu n\rceil} \frac{1}{\alpha_n^k(\beta)+\beta} \right| \le \frac{1}{\beta^2}\,\mathbb{E}_\beta\left(\beta \wedge \sup_k \big|\, \rho_n^k - \alpha_n^k(\beta) \,\big|\right)$$

and applying the dominated convergence theorem we obtain

$$\lim_{n\to\infty} \mathbb{E}_\beta\big(S_{\lceil \mu n\rceil}\big) = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{\lceil \mu n\rceil} \frac{1}{\alpha_n^k(\beta)+\beta}\,, \tag{34}$$

assuming the limit exists (this also holds with $\beta = 0$). Setting the above expression equal to T we obtain the saddle point estimate for β^*. Note that $\beta = \mu/T$ yields that the right side exceeds T and for $\beta = 0$, it is strictly less than T, otherwise ξ_n is not a rare event. Then, under a continuity (in β) assumption on the limit in (34), we have by the intermediate value theorem that the desired $\beta^* \in [0,\mu/T]$ exists.

Unfortunately this approach does not yield the optimal solution β^* to $\nabla K(\beta^*) = -T$. Under Assumption 6.1 we have that the limit $K(\beta)$ takes the form

$$K(\beta) = -\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{\lceil \mu n\rceil} \log\big(1+\beta/\alpha_n^k(\beta)\big)\,. \tag{35}$$

The right side of (35) may be recognized from (25) as the limiting scaled CGF of a birth process with constant rates $\{n\alpha^k\}$ independent of β (setting $\alpha_n^k(\beta) = \alpha^{k-1}$). In this setting $\nabla K(\beta)$ is given by (34). In general, however, defining

$$\varphi_n(\beta) \triangleq \frac{1}{n}\sum_{k=1}^{\lceil \mu n\rceil} \frac{1}{\alpha_n^k(\beta)+\beta}\big(1-\beta\,\nabla\log\big(\alpha_n^k(\beta)\big)\big) \tag{36}$$

we make the following assumption.

Assumption 6.2. *The functions* $\{\varphi_n\}$ *converge uniformly to some finite* φ *on* $[0, \mu/T]$ *such that* $T < \varphi(\mu/T)$.

Then $\nabla K(\beta) = -\lim_{n\to\infty} \varphi(\beta)$ and the asymptotically optimal β^* is the solution to

$$\lim_{n\to\infty} \varphi(\beta^*) = T\,. \tag{37}$$

Note that if $\nabla\alpha^k(\beta) = 0$, we obtain the saddle point condition. That is, if the limits α^k are not too sensitive to changes in β, the optimal β^* is roughly at the saddle point. The following proposition makes precise the above discussion and justifies the interchange of the limit and derivative.

Proposition 6.3. *Under Assumptions 6.1 and 6.2 the asymptotically optimal* β^* *exists in* $[0, \mu/T]$ *and satisfies* $\nabla K(\beta^*) = -T$ *if and only if* (37) *holds.*

Remark 6.4. *Assumption 6.1 alone is not sufficient for asymptotic optimality. By Egoroff's theorem, for any* $\varepsilon > 0$ *there exist for each* k *a set* E_k *over which this convergence of* ρ_n^k *is uniform and such that* $\mathbb{Q}_\beta(E_k^c) < \varepsilon$. *To achieve asymptotic optimality we must also require that* $\mathbb{Q}_\beta(\cap E_k^c) = c_n$ *for any sequence* $\{c_n\}$ *with* $\limsup_{n\to\infty} c_n/n < \infty$. *This condition, independently of Section 5, yields optimality of* $Y_n = Z_n(\beta^*)\,\mathbb{1}(\xi_n)$ *as well.*

We provide sufficient conditions under which the ρ_n^k satisfy the strong law of large numbers required by Proposition 6.3. The following is a corollary of the strong law of large numbers of Etemadi (1983).

Proposition 6.5. *Suppose that* $p_n(s) \le p_{n+1}(s)$ $\mathbb{Q}_\beta$*-almost surely for all* $s \ge 0$ *and for all* $k \le n$ *there exist constants* $\nu, \eta > 0$ *such that for all* $n \in \mathbb{N}$

$$Cov_\beta\left(p^i(S_k-),\, p^j(S_k-)\right) \le \nu\,(|i-j|+1)^{-\eta}. \tag{38}$$

Then $\sup_k \left|\rho_n^k - \alpha_n^k(\beta)\right| \to 0$ *as* $n \to \infty$ $\mathbb{Q}_\beta$*-almost surely.*

To understand the condition of Proposition 6.5, suppose the components of N are modeled as a network on a complete weighted graph. The weight of each link (i, j) is assigned as $|i-j|$, the distance in the positions of components i and j in N. Then (38) says that the correlation between any two components decays in inverse proportion to the distance between them. Condition (38) implies that $\mathrm{Var}_\beta\left(\rho_n^k\right) \to 0$ as $n \to \infty$ which is consistent with Assumption 6.1.

7 IS Algorithms

This section outlines the IS algorithms for the importance measures $\mathbb{Q}_{\gamma^*}$ and $\mathbb{Q}_{\beta^*}$ analyzed above. The algorithm for $\mathbb{Q}_{\gamma^*}$ is particularly simple and illustrates the basic structure of both schemes. The asymptotically optimal measure $\mathbb{Q}_{\beta^*}$ requires a more complex algorithm because one needs to iteratively compute the parameter β^*.

7.1 Algorithm for $\mathbb{Q}_{\gamma^*}$

For fixed $\mu \in (0,1)$, we estimate the probability of $\xi_n = \{C_T \geq \mu n\}$ using IS with importance measure $\mathbb{Q}_{\gamma^*}$. This entails the sampling of the IS estimator $Y_n = Z_n(\gamma^*)\,\mathbb{1}(\xi_n)$ under $\mathbb{Q}_{\gamma^*}$. We first evaluate $\mathbb{1}(\xi_n)$ by drawing a Poisson variable with parameter $\lceil \mu n \rceil$. If $\mathbb{1}(\xi_n) = 0$ (the case for roughly half of the samples), there is no need to generate the Radon-Nikodym derivative $Z_n(\gamma^*)$. If $\mathbb{1}(\xi_n) = 1$, we generate $Z_n(\gamma^*)$.

With $\gamma^* = \lceil \mu n \rceil / T$, we initialize $S_0 = 0$ and $N_0 = 0_n = (0, \ldots, 0)$ and perform the following steps to obtain a sample $\hat{Y}_n$ of the IS estimator:

(1) Generate a Poisson($\lceil \mu n \rceil$) variable C_T and if $\mathbb{1}(\xi_n) = 0$, return $\hat{Y}_n = 0$.

(2) Generate C_T uniform random variables $\{S_k\}$ sorted on the interval $[0, T]$.

(3) Generate p_n under $\mathbb{Q}_{\gamma^*}$ on the interval $[0, S_{\lceil \mu n \rceil})$.

(4) Evaluate $Z_n(\gamma^*)$ according to (19) and return $\hat{Y}_n = Z_n(\gamma^*)$.

A collection of independent samples $\hat{Y}_n$ are averaged to obtain an IS estimate of $y_n = \mathbb{E}_{\gamma^*}(Y_n) = \mathbb{P}(\xi_n)$. The first two steps are computationally simple. The dynamics of p_n under $\mathbb{Q}_{\gamma^*}$ required at Step (3) are implied by Girsanov's theorem and Proposition 4.1.

7.2 Algorithm for $\mathbb{Q}_{\beta^*}$

The algorithm for the asymptotically optimal importance measure $\mathbb{Q}_{\beta^*}$ is more complex than the algorithm for $\mathbb{Q}_{\gamma^*}$. This is because the event count C does not necessarily follow a Poisson process under the measure $\mathbb{Q}_{\beta^*}$. Moreover, the optimal parameter $\beta^* \in [0, \mu/T]$ must generally be determined during the simulation.

Suppose some value of β is given. Then, we initialize $S_0 = 0$ and $N_0 = 0_n$ and perform the following steps to obtain a sample $\hat{Y}_n$ of the IS estimator $Y_n = Z_n(\beta)\,\mathbb{1}(\xi_n)$:

(1) Generate (p_n, C) under $\mathbb{Q}_\beta$ on the interval $[0, S_{\lceil \mu n \rceil})$.

(2) If $\mathbb{1}(\xi_n) = 0$ return $\hat{Y}_n = 0$;
Else evaluate $Z_n(\beta)$ according to (23) and return $\hat{Y}_n = Z_n(\beta)$.

In most cases of interest the optimal value β^* for the twisting parameter cannot be computed prior to the simulation. To address this issue, we embed the IS trial above into a loop, allowing us to approximate β^* during the simulation. The derivative of α_n^k in (33) is computed similarly to (26) as

$$\nabla \alpha_n^k(\beta) = \sum_{j=1}^{\lceil \mu n \rceil} \mathbb{E}_\beta\left(\frac{\rho_n^k}{\beta + \rho_n^j}\right) - \mathbb{E}_\beta\left(n S_{\lceil \mu n \rceil} \rho_n^k\right). \tag{39}$$

The fixed n optimal parameter $\beta_n^* \in [0, \mu/T]$ is the solution of $\varphi_n(\beta) = T$ where φ_n is defined in (36). Since it is typically difficult to compute the expectations in (33) and (39) analytically, we approximate them by their sample means $\hat{\alpha}_n^k(\beta)$ and $\nabla\hat{\alpha}_n^k(\beta)$, obtained from samples of the underlying random variables generated during the simulation. We converge to the fixed n optimal β_n^* by iteratively applying the mapping

$$\psi_n(\beta) = \frac{\beta}{T}\,\varphi_n(\beta) \tag{40}$$

to find its fixed point in $[0, \mu/T]$. Initialize $\hat{\alpha}_n^k = \nabla\hat{\alpha}_n^k = 0_{\lceil \mu n \rceil}$ and $\beta = \mu/T$, and perform the following steps:

(I) Until $\hat{\alpha}_n^k(\beta)$ and $\nabla\hat{\alpha}_n^k(\beta)$ converge within some tolerance:

  (a) Run Step (1) of the IS trial with β.

  (b) Update $\hat{\alpha}_n^k(\beta)$ and $\nabla\hat{\alpha}_n^k(\beta)$ based on realizations in Step (a).

(II) Update $\beta = \psi_n(\beta)$ and continue to Step (I).

Once a sufficiently accurate estimate of β_n^* is obtained we may perform only the IS trials to obtain samples of Y_n. At Step (I) the tolerance for the convergence of $\hat{\alpha}_n^k$ and its derivative should be set large at first to prevent unnecessary trials with a suboptimal β. The tolerance can then be reduced as β converges. The scheme above is guaranteed to converge if ψ_n is a contractive mapping on $[0, \mu/T]$. If this property fails one can update β at Step (2) based on some other root finding scheme such as Newton's or bisection method applied to $\varphi_n(\beta) - T$.

There are alternatives to estimating β_n^*. When the gradients of α_n^k are small, they are difficult to estimate, and instead should be approximated by zero resulting in ψ_n taking a simpler from. From the discussion in Section 6, it follows that in this case β_n^* approximately satisfies the saddle point condition $\mathbb{E}_\beta\big(S_{\lceil \mu n \rceil}\big) = T$. Then we can take ψ_n of the form

$$\psi_n(\beta) = \frac{\beta}{T}\,\mathbb{E}_\beta\big(S_{\lceil \mu n \rceil}\big)\,,$$

where we estimate $\mathbb{E}_\beta\big(S_{\lceil \mu n \rceil}\big)$ by its sample mean. The effectiveness of this choice will depend on the $\mathbb{Q}_\beta$-variance of the arrival time $S_{\lceil \mu n \rceil}$.

8 Numerical Results

We illustrate the performance of our IS schemes for a model of correlated default timing in a portfolio of n firms. In this setting, a component of N indicates the default of a portfolio constituent firm. We consider the $\mathbb{P}$-intensity model

$$p^i = (w^i \nu^0 + \nu^i)(1 - N^i) \tag{41}$$

where $w^i \geq 0$ is a parameter and ν^i is a risk factor governed by the SDE

$$d\nu_t^i = \kappa^i \left(\bar{\nu}^i - \nu_t^i\right) dt + \sigma^i \sqrt{\nu_t^i}\, dW_t^i + \delta^i dJ_t, \quad \nu_0^i > 0, \tag{42}$$

for $i = 0, 1, \ldots, n$. Here, $(W^0, W^1, \ldots, W^n)$ is a standard Brownian motion and J is a jump process defined by $J = \Delta \cdot N$, where Δ is an n-vector of nonnegative parameters. The $\kappa^i, \bar{\nu}^i, \sigma^i$ and δ^i represent nonnegative reversion speed, reversion level, diffusive volatility and jump sensitivity parameters. We take $\delta^0 = 0$.

The model (41)–(42) was proposed by Giesecke, Kim & Zhu (2011). It addresses several of the sources of default clustering identified in empirical research, see Azizpour, Giesecke & Schwenkler (2010) and Das, Duffie, Kapadia & Saita (2007). The common factor ν^0 represents a source of systematic risk all firms are exposed to. It generates diffusive correlation between the intensities; the parameters w^i govern the magnitude of that correlation. The feedback term $\delta^i dJ_t$ generates self-excitation; a default begets more defaults. This models the contagion effects observed in credit markets.

The model (41)–(42) is difficult to analyze using existing rare-event schemes. If $\delta^i > 0$ for at least one $i \geq 1$, then N is no longer doubly-stochastic, and the scheme of Bassamboo & Jain (2006) no longer applies. The model (41)–(42) falls also outside the scope of the rare-event schemes for Markov chain models developed by Carmona & Crépey (2010) and Giesecke & Shkolnik (2010). The resampling schemes of Giesecke et al. (2010) and Deng et al. (2011) apply to the model (41)–(42). However, the practical implementation of these schemes for the model (41)–(42) requires a nontrivial computation that has not yet been worked out, to our knowledge.[2] In contrast, the IS algorithms developed in this paper can be readily applied to the model (41)–(42).

The model parameters are selected as follows. Let $U(a, b)$ be the uniform distribution on $[a, b]$. For each name $i = 1, \ldots, n$, we draw κ^i from $U(0.5, 1.5)$, $\bar{\nu}^i$ from $U(0.001, 0.051)$ and set σ^i to $\min\{\sqrt{2\kappa^i \bar{\nu}^i}, \bar{\sigma}^i\}$, where each $\bar{\sigma}^i$ is drawn from $U(0, 0.2)$. The restriction on σ^i ensures that the Feller condition is satisfied, guaranteeing that ν^i is strictly positive almost surely. We initialize $\nu_0^i = \bar{\nu}^i$. The common factor is initialized as $\nu_0^0 = \bar{\nu}^0 = 0.025$, and we set $\kappa^0 = 1$ and $\sigma^0 = 0.1$ so that the Feller condition is satisfied. Each common factor weight w^i is drawn from $U(0, 1)$. The components of Δ, which along with the δ^i specify the impact of defaults, are drawn from $U(0, 2/n)$. (The scaling $1/n$ reflects the fact that J depends on the number of names.)

We analyze two portfolios, one exhibiting low and the other high levels of contagion. They are designed by selecting different values for the sensitivities δ^i. For Portfolio I we draw each δ^i from $U(0, 2)$ and for Portfolio II we draw it from $U(0, 5)$.

We estimate the probability of $\xi_n = \{C_1 \geq \mu n\}$ for various values of $\mu \in (0, 1)$ for portfolios of $n = 100$ firms. We compare the IS estimators based on $\mathbb{Q}_{\gamma^*}$ and $\mathbb{Q}_{\beta^*}$ with a plain

[2] The resampling schemes require the construction of a mimicking Markov chain M for N. The conditional expectation $\mathbb{E}(p_t^i \mid N_t = \cdot)$ representing the transition function of a component of M seems hard to compute for the model (41)–(42) unless the δ^i are all 0.

Monte Carlo (MC) estimator. The plain MC estimator $\mathbb{1}(\xi_n)$ is obtained by simulating the counting process C using the $\mathbb{P}$-intensity $p_n = 1_n \cdot p$ up to time $T = 1$. This is based on a standard time-scaling scheme, which involves the discretization of the intensity (see Section 2.2 in Giesecke et al. (2010)). We use a time step of 10^{-3} and apply the trapezoidal rule to compute the integral (the compensator to C) at the grid points.[3] Linear interpolation is used to approximate the arrival times between the grid points. The intensity values at the grid points are sampled from scaled non-central chi-squared densities describing the transition laws of the ν^i between event times. At an event time S_k, the index of the defaulter is drawn from (18), and the (pre-jump) intensity samples are revised by the appropriate contribution from the jump term.

We use the algorithms described in Section 7 to generate the IS estimators $Y_n = Z_n \mathbb{1}(\xi_n)$. The IS scheme based on $\mathbb{Q}_{\gamma^*}$ requires the generation of the $\mathbb{P}$-intensity p_n of C under $\mathbb{Q}_{\gamma^*}$ (Step (3)). Since the jump times S_k of p_n under $\mathbb{Q}_{\gamma^*}$ are already determined at Step (2) of the algorithm, it remains to generate p_n between the S_k, which also involves the sampling of the index of a defaulter from (18). The paths of p_n are required because $Z_n(\gamma^*)$ is a function of $\int_0^{S_{\lceil \mu n \rceil}} p_n(s)ds$, see (19). We approximate them on a discrete time grid with time step 10^{-3} and use the trapezoidal rule to compute the integral.[4] The values of p_n at the grid points are sums of variables drawn independently from scaled non-central chi-squared densities, because p_n is a sum of independent Feller diffusions between defaults. Here we use the fact that the $\mathbb{P}$-Brownian motions W^i driving the ν^i are also $\mathbb{Q}_{\gamma^*}$-Brownian motions, by Girsanov's and Lévy's theorems (the measure change affects only local $\mathbb{P}$-martingales that have non-zero covariation with $Z_n(\gamma^*)$). Thus, between events ν^i has the same dynamics under both measures.

The IS scheme based on $\mathbb{Q}_{\beta^*}$ requires the generation of (p_n, C) under $\mathbb{Q}_{\beta^*}$. To generate C, we use a time-scaling scheme similar to the one used to generate the plain MC estimator. This is based on the discretization of the $\mathbb{Q}_{\beta^*}$-intensity $q_n = 1_n \cdot q = p_n + n\beta^*$ of C, and also leads to the required values of $p_n(S_k-)$. The processes q_n and p_n have the same dynamics under $\mathbb{Q}_{\beta^*}$, and the $\mathbb{P}$-Brownian motions W^i driving the ν^i are also $\mathbb{Q}_{\beta^*}$-Brownian motions. Thus, we can again sample the values of q_n at the discretization grid points between the S_k from scaled non-central chi-squared densities. As above, the index of a defaulter is drawn from (18) at an event time S_k. The tolerances for the convergence of the sample means $\hat{\alpha}_n^k$ and $\nabla\hat{\alpha}_n^k$ are based on their 90% confidence intervals. We initialize both tolerances to the value $\mu/T = \mu$ and reduce them by a factor of 0.9 each time β is updated. We regard the estimates of β^* as converged whenever the relative error between the current and previous estimate is bounded by 10^{-3}. The initial guess for β^* was set to $\mu/T = \mu$ and for each

[3]The fine grid ensures a negligible discretization bias at a relatively significant computational cost. An alternative to discretization is the scheme of Giesecke et al. (2011), which generates asymptotically unbiased estimators at a lower cost than time-scaling. We do not use this scheme in our experiments in order to be consistent with the intensity discretization required in the IS scheme based on $\mathbb{Q}_{\gamma^*}$ (the Radon-Nikodym derivative $Z_n(\gamma^*)$ depends on the path of p_n).

[4]At time t with $C_t = m - 1$ the time step is computed as $\min\{10^{-3}, S_m - t\}$.

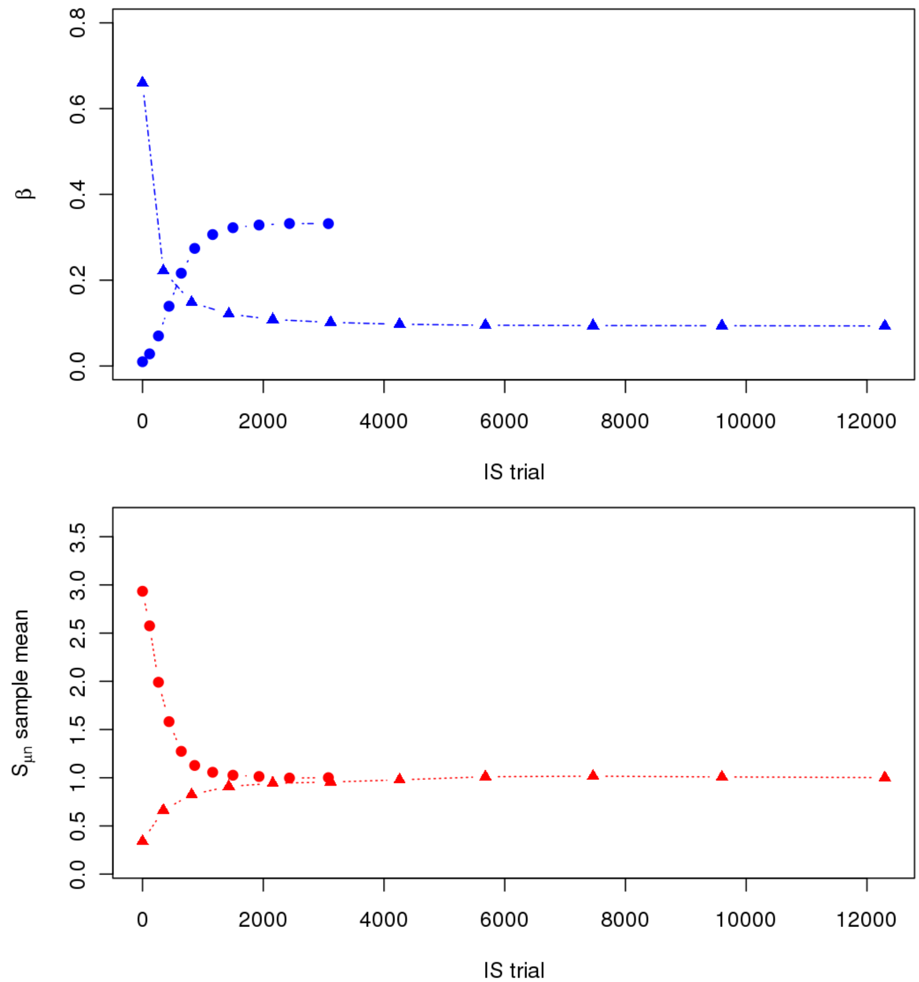

Figure 1: *Upper Panel*: Convergence of β to the optimal parameter β^* during two simulations of Portfolio II. The decreasing curve shows convergence starting at $\beta = 0.66$ to the optimal value $\beta^* \approx 0.09341$ for the case $\mu = 0.32$. The increasing curve shows convergence from initial $\beta = 0.01$ to the optimal value $\beta^* \approx 0.3320$ for the case $\mu = 0.66$. *Lower Panel*: Convergence of the mean of $S_{\lceil \mu n \rceil}$ to $T = 1$.

successive value of μ we have used the previous optimal value as the new starting value.

Convergence plots for β^* are shown in Figure 1. The results suggest that starting at higher values for the initial β is preferable. Caution has to be taken when initializing β to values near zero since $\psi_n(0) = 0$, a fixed point. As illustrated, the rate of convergence starting near zero may be slow. Figure 1 also shows the convergence of the sample mean of $S_{\lceil \mu n \rceil}$ to T suggesting that the optimal value lies near the saddle point $\mathbb{E}_\beta\big(S_{\lceil \mu n \rceil}\big) = 1$. This is reasonable since each $\nabla \hat{\alpha}_n^k$ is indeed small relative to $\hat{\alpha}_n^k$ for both portfolios.

A single plain MC experiment leads to estimators $\mathbb{1}(\xi_n)$ for any value $\mu \in (0, 1)$. The IS estimators $Y_n = \mathbb{1}(\xi_n) Z_n$ depend on μ through the underlying importance measure. Thus,

μn	$\mathbb{Q}_{\beta^*}$-IS Estimate $\pm$ 95% CI	$\mathbb{Q}_{\beta^*}$-IS Var	Var Ratio ($\mathbb{Q}_{\beta^*}$)	Var Ratio ($\mathbb{Q}_{\gamma^*}$)
2	9.2150E−01 ± 3.7276E−03	7.2341E−02	9.7855E−01	6.0879E−02
4	7.1970E−01 ± 6.2250E−03	2.0174E−01	1.0040E+00	3.4276E−01
6	4.7594E−01 ± 5.9429E−03	1.8387E−01	1.3563E+00	8.3306E−01
8	2.8020E−01 ± 3.8658E−03	7.7804E−02	2.5716E+00	1.4675E+00
10	1.4447E−01 ± 2.3082E−03	2.7737E−02	4.4793E+00	2.4682E+00
12	6.9894E−02 ± 1.2889E−03	8.6483E−03	7.5298E+00	3.5518E+00
14	3.1562E−02 ± 6.6722E−04	2.3177E−03	1.2914E+01	6.1388E+00
16	1.3174E−02 ± 3.1217E−04	5.0736E−04	2.5338E+01	1.0769E+01
18	5.1605E−03 ± 1.3566E−04	9.5816E−05	5.2955E+01	2.3032E+01
20	1.9614E−03 ± 6.0333E−05	1.8951E−05	9.9412E+01	3.1242E+01
22	6.5714E−04 ± 2.1882E−05	2.4929E−06	2.5557E+02	9.7345E+01
24	2.2860E−04 ± 8.7531E−06	3.9888E−07	5.5142E+02	2.8836E+02
26	6.9839E−05 ± 2.9529E−06	4.5397E−08	1.5969E+03	9.9169E+02
28	2.1805E−05 ± 9.8427E−07	5.0437E−09	6.4435E+03	3.0083E+03
30	6.0130E−06 ± 3.1161E−07	5.0551E−10	9.8909E+03	3.7827E+03
32	1.6072E−06 ± 9.2551E−08	4.4594E−11	5.6061E+04	3.3981E+04
34	3.8886E−07 ± 2.3816E−08	2.9530E−12	8.4660E+05	4.2395E+05
36	9.7045E−08 ± 7.4591E−09	2.8966E−13	8.6309E+06	5.3467E+06

Table 1: Simulation results for Portfolio I. The second column reports the $\mathbb{Q}_{\beta^*}$-IS estimate of $\mathbb{P}(C_1 \geq \mu n)$ along with a 95% confidence interval, and the third column the sample variance of the estimator. The fourth column reports the variance ratio, given by the sample variance of the plain MC estimator of $\mathbb{P}(C_1 \geq \mu n)$ over the sample variance of the $\mathbb{Q}_{\beta^*}$-IS estimator. The last column reports the variance ratio for the $\mathbb{Q}_{\gamma^*}$-IS estimator.

separate IS experiments have to be performed for the desired values of μ. To facilitate a meaningful comparison of plain MC and IS estimators, we run 400K plain MC trials and 18 separate IS experiments of approximately 20K (for $\mathbb{Q}_{\beta^*}$) and 40K (for $\mathbb{Q}_{\gamma^*}$) trials each, leading to IS estimates of $\mathbb{P}(\xi_n)$ for 18 values of μ for each importance measure. The plain MC experiment takes the same amount of time as the 18 separate IS experiments, for each of the two importance measures. The experiments were implemented in the R package, running on a Linux 64bit Intel® Core™ 2 Quad CPU at 2.40GHz.

Table 1 reports the results for Portfolio I and Figure 2 plots the plain MC and IS estimates along with confidence intervals. The variance ratio, computed as the variance of the plain MC estimator over the variance of an IS estimator, indicates significant variance reduction for both importance measures. The measure $\mathbb{Q}_{\beta^*}$ achieves higher variance reduction than the measure $\mathbb{Q}_{\gamma^*}$, for all values of μ considered. For small values of μ the IS scheme with $\mathbb{Q}_{\beta^*}$ has the optimal β^* near zero and the algorithm reduces to plain Monte Carlo as highlighted by a variance ratio of approximately 1. The IS scheme with $\mathbb{Q}_{\gamma^*}$ does not perform as well

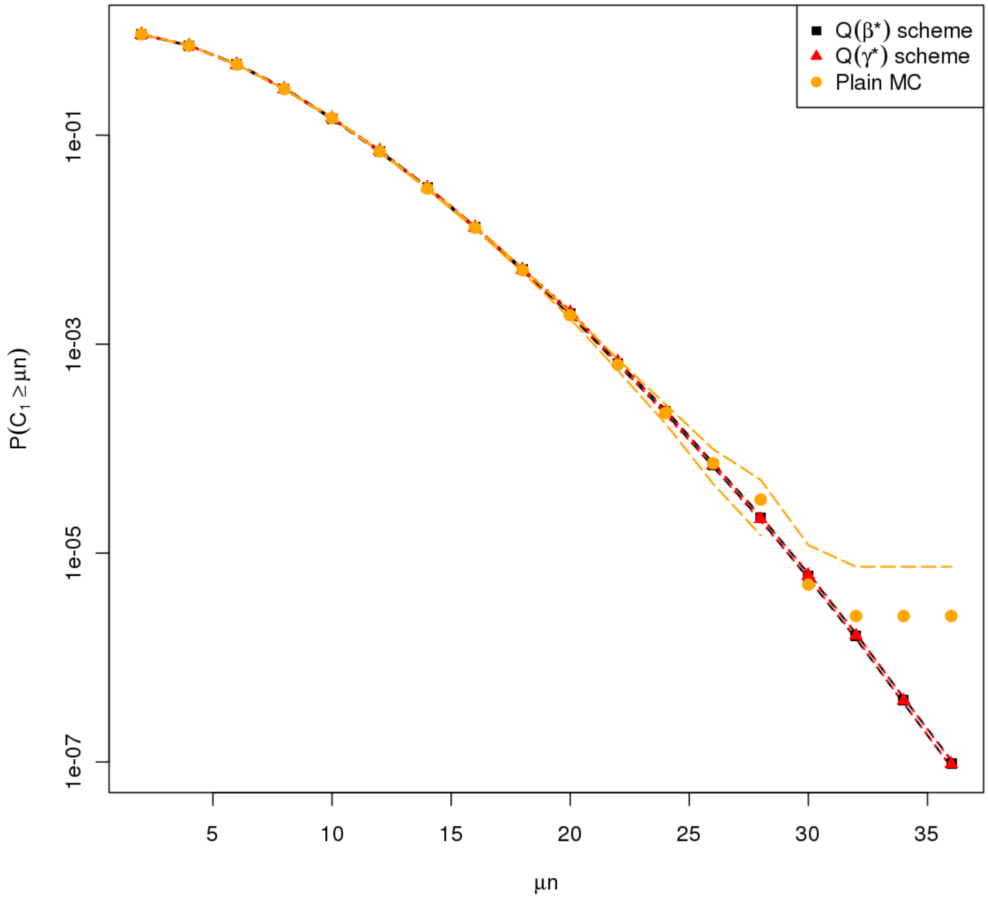

Figure 2: IS and plain MC estimates (markers) and 95% confidence intervals (dashed lines). Lower confidence levels extending below zero are not shown.

because the Radon-Nikodym derivative is skewed for large probabilities. Remarkably, as we move into the center of the distribution we already begin to see variance reduction. This is due to the fact that for this portfolio the relative error of each IS estimator remains roughly constant for all values of μ, suggesting better performance than the optimality criterion (6). In contrast, the plain Monte Carlo relative error increases with μ as discussed in Section 2.

Table 2 reports the results for the high contagion Portfolio II. Figure 3 plots the estimates and confidence intervals. We consider larger values of μ since defaults occur more frequently. While both schemes obtain significant variance reduction, clearly the $\mathbb{Q}_{\beta^*}$-IS scheme outperforms the $\mathbb{Q}_{\gamma^*}$-IS scheme. While also true this is much less evident in the low contagion scenario of Portfolio I. In the high contagion setting the values $p_n(S_k)$ (and hence $p_n(S_k-)$) increase rapidly with k, the number of events. This results in a highly skewed Radon-Nikodym derivative $Z_n(\gamma^*)$. This is not the case for $Z_n(\beta^*)$ since $K_n(\beta^*)$

μn	$\mathbb{Q}_{\beta^*}$-IS Estimate ± 95% CI	$\mathbb{Q}_{\beta^*}$-IS Var	Var Ratio ($\mathbb{Q}_{\beta^*}$)	Var Ratio ($\mathbb{Q}_{\gamma^*}$)
32	2.4529E−02 ± 6.2536E−04	2.0360E−03	1.1475E+01	6.7408E−01
34	1.6084E−02 ± 4.3809E−04	9.9916E−04	1.5900E+01	3.1734E−02
36	1.0520E−02 ± 3.1462E−04	5.1535E−04	2.0189E+01	5.1549E−01
38	6.7985E−03 ± 2.1650E−04	2.4402E−04	2.7617E+01	5.9328E−01
40	4.2230E−03 ± 1.4913E−04	1.1578E−04	3.5523E+01	1.2561E+00
42	2.5003E−03 ± 9.8816E−05	5.0836E−05	4.8321E+01	1.5028E+00
44	1.4346E−03 ± 5.8494E−05	1.7813E−05	7.8624E+01	1.6784E+00
46	8.0243E−04 ± 3.6886E−05	7.0835E−06	1.1355E+02	7.2956E+00
48	4.0861E−04 ± 1.8218E−05	1.7279E−06	2.7188E+02	1.8795E+00
50	2.3000E−04 ± 1.3306E−05	9.2173E−07	2.4676E+02	3.1866E+01
52	1.0780E−04 ± 6.2009E−06	2.0018E−07	4.4955E+02	6.8605E+01
54	5.1820E−05 ± 4.0897E−06	8.7075E−08	3.7323E+02	3.8096E+01
56	2.1106E−05 ± 1.4801E−06	1.1406E−08	6.5757E+02	1.8556E+01
58	8.6270E−06 ± 8.2096E−07	3.5088E−09	7.1250E+02	1.9154E+02
60	3.3727E−06 ± 2.9739E−07	4.6043E−10	5.4297E+03	4.7664E+02
62	1.1006E−06 ± 9.9266E−08	5.1301E−11	N/A	N/A
64	3.3889E−07 ± 2.8783E−08	4.3130E−12	N/A	N/A
66	1.0368E−07 ± 1.2327E−08	7.9110E−13	N/A	N/A

Table 2: Simulation results for Portfolio II. The second column reports the $\mathbb{Q}_{\beta^*}$-IS estimate of $\mathbb{P}(C_1 \geq \mu n)$ along with a 95% confidence interval, and the third column the sample variance of the estimator. The fourth column reports the variance ratio, given by the sample variance of the plain MC estimator of $\mathbb{P}(C_1 \geq \mu n)$ over the sample variance of the $\mathbb{Q}_{\beta^*}$-IS estimator. The last column reports the variance ratio for the $\mathbb{Q}_{\gamma^*}$-IS estimator.

is close to zero for large $p_n(S_k-)$, ensuring that the Radon-Nikodym derivative is well bounded. In fact, the performance of the $\mathbb{Q}_{\beta^*}$-IS estimator remains largely unaffected by the choice of the jump sensitivities δ^i. In contrast, the $\mathbb{Q}_{\gamma^*}$-IS scheme becomes highly unreliable as the δ^i's are increased beyond some threshold.

9 Conclusion

This paper develops and tests provably efficient importance sampling algorithms for the estimation of rare-event probabilities in point process models of event timing. We do not impose a particular structure on the stochastic intensity processes governing event times, which guarantees the broad applicability of these schemes. The change of measure underlying our algorithms is induced by twisting the intensities. A family of such twists is shown to lead to asymptotically optimal estimators of the tail of the distribution of the

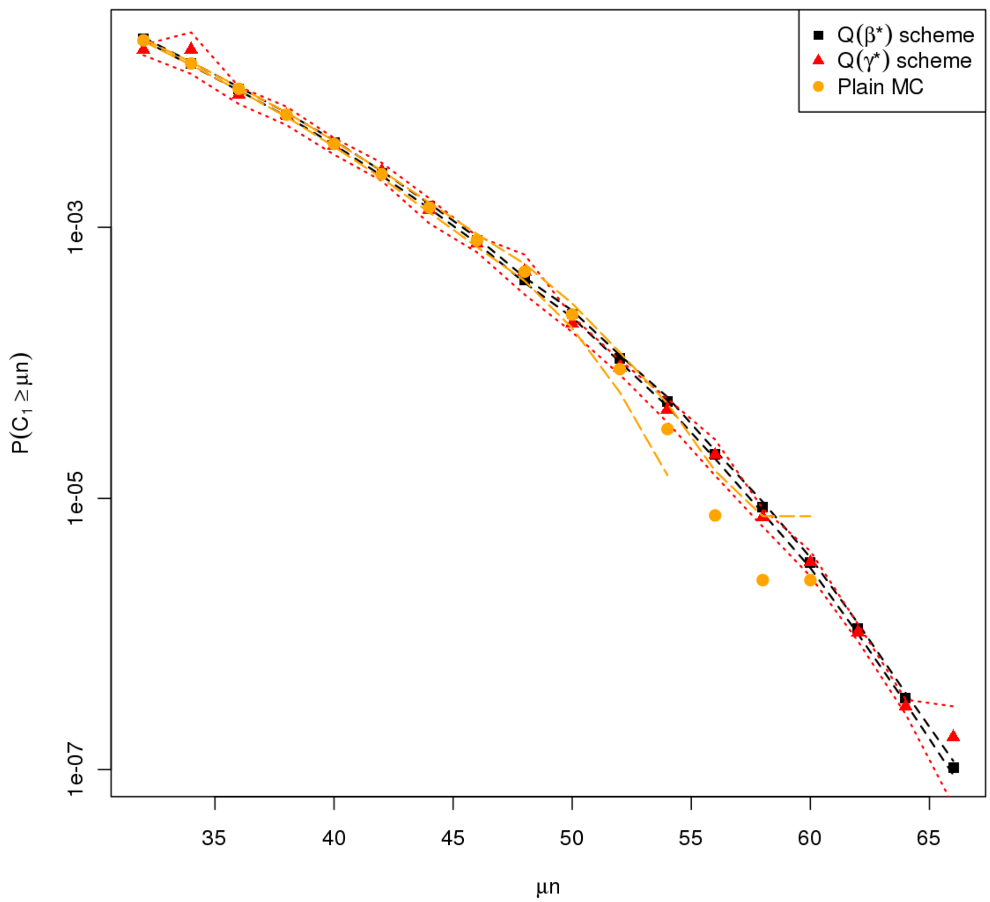

Figure 3: IS and plain MC estimates (markers) and 95% confidence intervals (dashed lines). Lower confidence levels extending below zero are not shown.

total event count at a fixed horizon. Numerical results illustrate the performance of the associated rare-event estimators.

A Proofs

Proof of Proposition 4.1. The hypotheses imply that each $\lambda_n(S_k-)$ is $\mathbb{P}$-a.s. finite for all $k = 1, \ldots, \lceil \mu n \rceil$. Stieltjes integration by parts implies that the process L defined by

$$L_t = \exp\left(\int_0^{t \wedge S_{\lceil \mu n \rceil}} \log\left(\lambda_n(s-)\right) dC_s + \int_0^{t \wedge S_{\lceil \mu n \rceil}} \left(1 - \lambda_n(s)\right) p_n(s)\, ds \right)$$

satisfies the equation

$$L_t = 1 + \int_0^t L_{s-} (\lambda_n (s-) - 1) \mathbb{1}(s \le S_{\lceil \mu n \rceil}) dM_s$$

where $M = C - \int_0^\cdot p_n(s)\, ds$ is a $\mathbb{P}$-martingale by (1). Theorem 29 of Chapter IV of Protter (2004) and the fact that all adapted and left continuous processes are locally bounded, imply that L is a local $\mathbb{P}$-martingale. Since $\lambda_n \ge 1$, we have $\sup_{t \ge 0} L_t \le \exp(-K_n)$ which is $\mathbb{P}$-integrable by assumption. Thus, by Theorem 51 in Chapter I of Protter (2004), the process L is a uniformly integrable $\mathbb{P}$-martingale.

By optional stopping, $\mathbb{E}(L_{S_{\lceil \mu n \rceil}}) = L_0 = 1$ and we can take $\mathbb{Q}(A) = \mathbb{E}(\mathbb{1}(A) L_{S_{\lceil \mu n \rceil}})$ for all $A \in \mathcal{F}_{S_{\lceil \mu n \rceil}}$. Since our hypotheses imply that $L_{S_{\lceil \mu n \rceil}} > 0$ almost surely under $\mathbb{P}$, the probability measures $\mathbb{Q}$ and $\mathbb{P}$ are equivalent and $Z_n = 1/L_{S_{\lceil \mu n \rceil}}$ is the Radon-Nikodym derivative of $\mathbb{P}$ with respect to $\mathbb{Q}$.

That N admits $\mathbb{Q}$-intensity $q(s) = \lambda_n(s) p(s)$ on the interval $[0, S_{\lceil \mu n \rceil})$ follows from Girsanov's theorem; see Jacod (1975) or VI.T3 in Brémaud (1980). □

Proof of Theorem 5.2. Fix $\delta > 0$ and let $\beta = \theta^*$ of Assumption 3.1. Let

$$G_n = \{n^{-1} K_n(\beta) \ge K(\beta) + \delta\}.$$

By Lemma 5.1 it follows that the $\mathbb{Q}_\beta(G_n) \to 0$ exponentially in c_n as $n \to \infty$. Fixing $\eta > 0$ for $\|X\|_n = \mathbb{E}_\beta(X^n)^{\frac{1}{n}}$ we obtain

$$\begin{aligned} \frac{1}{n} \Psi_n(-\eta n) &= \frac{1}{n} \log \mathbb{E}_\beta(\exp(\eta K_n(\beta))) \\ &= \log \left\| \exp\left(\eta n^{-1} K_n(\beta)\right) \right\|_n. \end{aligned}$$

Applying Minkowski's inequality and the fact that $K_n(\beta) \le 0$ for all $\beta \ge 0$

$$\begin{aligned} \frac{1}{n} \Psi_n(-\eta n) &\le \log(\exp(\eta(K(\beta) + \delta)) + \|\mathbb{1}(G_n)\|_n) \\ &\le \delta\eta + \eta K(\beta) + \log(1 + \exp(-\eta K(\beta) - \eta\delta) \|\mathbb{1}(G_n)\|_n) \\ &\le \delta\eta + \eta K(\beta) + \exp(-\eta K(\beta) - \eta\delta) \mathbb{Q}_\beta(G_n)^{1/n} \\ &\le \delta\eta + \eta K(\beta) + \exp\left(-\eta K(\beta) - \eta\delta + n^{-1} \log \mathbb{Q}_\beta(G_n)\right). \end{aligned}$$

Since $c_n/n \to \infty$ taking the limit as $n \to \infty$ and then $\delta \to 0$ we obtain

$$\lim_{n \to \infty} \frac{1}{n} \Psi_n(-\eta n) \le \eta K(\beta) \tag{43}$$

We apply this to the second moment of the estimator as follows.

$$\begin{aligned}\frac{1}{n}\log\mathbb{E}_\beta\left(Y_n^2\right) &= \frac{1}{n}\log\mathbb{E}_\beta\left(\exp\left(2\beta n S_{\lceil\mu n\rceil}+2K_n(\beta)\right)\mathbb{1}(\xi_n)\right)\\ &\le 2\beta T+\frac{1}{n}\log\mathbb{E}_\beta(\exp(2K_n(\beta)))\\ &\le 2\beta T+\frac{1}{n}\Psi_n(-2n)\end{aligned}$$

so that

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}_\beta\left(Y_n^2\right)\le 2\beta T+2K(\beta)$$

Setting $\eta=1$, taking $n\to\infty$ in (43) and combining with (30) yields $\Lambda(\beta)\le K(\beta)$ and

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}_\beta\left(Y_n^2\right)\le 2\beta T+2\Lambda(\beta) \tag{44}$$

The minimizer β^* of the bound in (44) satisfies $\nabla\Lambda(\beta)=-T$ which exists by Assumption 3.1 since $\beta=\theta^*$.

Furthermore, by the Gärtner-Ellis theorem, Theorem 2.3.6 (Dembo & Zeitouni 1998), applied to $\{S_{\lceil\mu n\rceil}\in(0,T)\}\subset\{S_{\lceil\mu n\rceil}\le T\}$ we have that

$$\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}(\xi_n)\ge-\inf_{(0,T)\cap E}\Lambda^*(s) \tag{45}$$

where E is the set of exposed points of the Fenchel-Legendre transform[5] of Λ given by $\Lambda^*(s)=\sup_\theta\{-\theta s-\Lambda(\theta)\}$. Since $\nabla\Lambda(\theta)$ is continuous (and monotone by convexity of Λ) it follows that $B_T(\delta)\subset E$ for some $\delta>0$, so $(0,T)\cap E$ is nonempty.

We show that $\lambda^*\triangleq\inf_{(T-\delta,T)}\Lambda^*(s)=\Lambda^*(T)$. Since T is the exposed point of Λ^* with exposing hyperplane $\theta^*>0$, it follows that Λ^* is decreasing as $s\uparrow T$ so $\lambda^*\ge\Lambda^*(T)$. Now, assume that $\lambda^*>\Lambda^*(T)$. Then there exists some λ such that $\Lambda^*(T)<\lambda<\lambda^*\le\Lambda^*(s)$ for all $s\in(T-\delta,T)$. Clearly this contradicts the fact that Λ^* is continuous (by convexity) on $B_T(\delta)$. Therefore, $\lambda^*\le\Lambda^*(T)$ and consequently $\lambda^*=\Lambda^*(T)$. Finally,

$$\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}(\xi_n)\ge-\inf_{(0,T)\cap E}\Lambda^*(s)\ge-\lambda^*=-\Lambda^*(T)=\theta^*T+\Lambda(\theta^*) \tag{46}$$

as desired. Asymptotic optimality of Y_n under importance measure $\mathbb{Q}_{\beta^*}$ with $\beta^*=\theta^*$ follows by (44) and Jensen's inequality. □

Proof of Proposition 6.3. Suppose that $n^{-1}K_n(\beta)\to K$ in $\mathbb{Q}_\beta$-probability for some constant K. First, we prove that $K=K(\beta)$ as defined in (35) under the assumptions. Define

$$\kappa_n(\beta)\triangleq-\sum_{k=1}^{\lceil\mu n\rceil}\log\left(1+\beta/\alpha_n^k(\beta)\right)$$

[5] We adopt the convention which uses $-\theta$ instead of θ due to our definition of the CGF Λ.

for which we have the upper bound

$$\begin{aligned} K_n(\beta) - \kappa_n(\beta) &= \sum_{k=1}^{\lceil \mu n \rceil} \log\left(\frac{1+\beta/\alpha_n^k(\beta)}{1+\beta/\rho_n^k}\right) \\ &= \sum_{k=1}^{\lceil \mu n \rceil} \log\left(1+\beta\frac{1/\alpha_n^k(\beta) - 1/\rho_n^k}{1+\beta/\rho_n^k}\right) \\ &\leq \beta \lceil \mu n \rceil \sup_k \left|1/\rho_n^k - 1/\alpha_n^k(\beta)\right| \end{aligned}$$

The lower bound $-\beta \lceil \mu n \rceil \sup_k \left|1/\rho_n^k - 1/\alpha_n^k(\beta)\right|$ follows similarly. We now have that

$$\left| n^{-1}K_n(\beta) - n^{-1}\kappa_n(\beta) \right| \leq \beta \sup_k \left|1/\rho_n^k - 1/\alpha_n^k(\beta)\right|$$

and by Assumption 6.1, $n^{-1}K_n(\beta) \to \lim_{n\to\infty} n^{-1}\kappa_n(\beta) = K(\beta)$ as $n \to \infty$ $\mathbb{Q}_\beta$-a.s.

Next, it is easy to show $\nabla n^{-1}\kappa_n(\beta) = -\varphi_n(\beta)$. Since φ_n converges to some φ uniformly and $n^{-1}\kappa_n(\beta) \to K(\beta)$ it follows that $\nabla K(\beta) = -\varphi(\beta)$. Thus $\varphi(\beta^*) = T$ if and only if $\nabla K(\beta^*) = -T$. To show that such β^* exists in $[0, \mu/T]$ we first show that $\varphi(0) \geq T$. Similarly to the derivation of (34) we have setting $\beta = 0$ that

$$0 = \lim_{h\downarrow 0} \frac{1}{h}\mathbb{E}\left(1/Z_n(h) - 1\right)$$

and note that the integrand is bounded above by $\exp(-K_n(h))$ which is integrable under the hypetheses of Proposition 4.1. Applying the dominatated convergence theorem yields

$$\mathbb{E}\left(S_{\lceil \mu n \rceil}\right) = \frac{1}{n}\sum_{k=1}^{\lceil \mu n \rceil} \mathbb{E}\left(1/\rho_n^k\right)$$

and

$$\left| \mathbb{E}\left(S_{\lceil \mu n \rceil}\right) - \frac{1}{n}\sum_{k=1}^{\lceil \mu n \rceil} 1/\alpha_n^k(0) \right| \leq \sup_k \left|1/\rho_n^k - 1/\alpha_n^k(0)\right| \to 0$$

as $n \to \infty$ $\mathbb{P}$-a.s. under Assumption 6.1. It follows that $\lim_{n\to\infty} \mathbb{E}\left(S_{\lceil \mu n \rceil}\right) = \varphi(0)$ and by Markov's inequality

$$\mathbb{P}(\xi_n^c) = \mathbb{P}\left(S_{\lceil \mu n \rceil} > T\right) \leq \mathbb{E}\left(S_{\lceil \mu n \rceil}\right)/T.$$

Thus, if $\varphi(0) < T$ then $\limsup_{n\to\infty} \mathbb{P}(\xi_n^c) < 1$ and it follows that $\limsup_n \mathbb{P}(\xi_n) > 0$ i.e. ξ_n is not a rare event sequence. Therefore, $\varphi(0) \geq T$ and we have assumed that $\varphi(\mu/T) < T$. By Assumption 6.1 each φ_n is continuous and since $\varphi_n \to \varphi$ uniformly on $[0, \mu/T]$ it follows that φ is continuous on $[0, \mu/T]$. By the intermediate value theorem it follows that β^* exists in $[0, \mu/T]$. □

Proof of Proposition 6.5. Define $\rho_n^k(i) \triangleq n^{-1} p^i(S_k-)$ and suppose without loss of generality that $\mathrm{Cov}_\beta\left(\rho_n^k(i), \rho_n^k(j)\right) \leq \nu/2\left(|i-j|+1\right)^{-2\eta}$ for all $k \leq n$, all $n \in \mathbb{N}$ and some constants $\nu, \eta > 0$ independent of n. Again w.l.o.g. assume that $\eta \in (0, \frac{1}{2})$, then

$$\begin{aligned}
\mathrm{Var}_\beta\left(\rho_n^k\right) &= \sum_{i=1}^n \sum_{j=1}^n \mathrm{Cov}_\beta\left(\rho_n^k(i), \rho_n^k(j)\right) \\
&\leq \frac{\nu}{2n^2} \sum_{i=1}^n \sum_{j=1}^n \left(|i-j|+1\right)^{-2\eta} \\
&\leq \frac{\nu}{2n^2} \left(\sum_{|i-j|+1<\sqrt{n}} \left(|i-j|+1\right)^{-2\eta} + \sum_{|i-j|+1 \geq \sqrt{n}} \left(|i-j|+1\right)^{-2\eta} \right) \\
&\leq \frac{\nu}{2n^2}\left(n + n^2 n^{-\eta}\right) \\
&\leq \nu / n^\eta
\end{aligned}$$

Now let $n_\ell = \lceil a^\ell \rceil$ for any $a > 1$. Then,

$$\sum_{n=1}^\infty \mathbb{Q}_\beta\left(\rho_{n_\ell}^k - \alpha_{n_\ell}^k(\beta) \geq \varepsilon\right) \leq \varepsilon^{-2} \sum_{n=1}^\infty \mathrm{Var}_\beta\left(\rho_{n_\ell}^k\right) \leq \varepsilon^{-2} \nu \sum_{n=1}^\infty n_\ell^{-\eta} < \infty$$

where the last step follows since $a > 1$. By the Borel-Cantelli lemma $\rho_{n_\ell}^k - \alpha_{n_\ell}^k(\beta) \to 0$ as $n \to \infty$ $\mathbb{Q}_\beta$-almost surely. Now, let n be such that $n_\ell \leq n < n_{\ell+1}$ and $b_\ell = n_{\ell+1}/n_\ell$, by monotonicity of p_n in n we have

$$\rho_n^k - \alpha_n^k(\beta) \leq b_\ell\, \rho_{n_{\ell+1}}^k - \alpha_n^k(\beta) \leq b_\ell\left(\rho_{n_{\ell+1}}^k - \alpha_{n_{\ell+1}}^k(\beta)\right) + b_\ell\, \alpha_{n_{\ell+1}}^k(\beta) - \alpha_n^k(\beta)$$

and since we have $\lim_{\ell\to\infty} b_\ell = a$ and

$$\limsup_{n\to\infty} \sup_k \left(\rho_n^k - \alpha_n^k(\beta)\right) \leq (a-1) \sup_k \alpha^k(\beta).$$

Similarly

$$\rho_n^k - \alpha_n^k(\beta) \geq 1/b_\ell\, \rho_{n_\ell}^k - \alpha_n^k(\beta) \geq 1/b_\ell\left(\rho_{n_\ell}^k - \alpha_{n_\ell}^k(\beta)\right) + 1/b_\ell\, \alpha_{n_\ell}^k(\beta) - \alpha_n^k(\beta)$$

which yields

$$\liminf_{n\to\infty} \sup_k \left(\rho_n^k - \alpha_n^k(\beta)\right) \geq (1/a - 1) \sup_k \alpha^k(\beta)$$

Since $a > 1$ is arbitrary and $\sup_k \alpha^k(\beta)$ is finite we have

$$\lim_{n\to\infty} \sup_k \left|\rho_n^k - \alpha_n^k(\beta)\right| \to 0$$

as $n \to \infty$ $\mathbb{Q}_\beta$-a.s. as desired. □

References

Asmussen, Soren & Peter Glynn (2007), *Stochastic Simulation – Algorithms and Analysis*, Springer, New York.

Azizpour, Shahriar, Kay Giesecke & Gustavo Schwenkler (2010), Exploring the sources of default correlation. Working Paper, Stanford University.

Bassamboo, Achal & Sachin Jain (2006), Efficient importance sampling for reduced form models in credit risk, *in* L. F.Perrone, F. P.Wieland, J.Liu, B. G.Lawson, D. M.Nicol & R. M.Fujimoto, eds, 'Proceedings of the 2006 Winter Simulation Conference', IEEE Press, pp. 741–748.

Bassamboo, Achal, Sandeep Juneja & Assaf Zeevi (2008), 'Portfolio credit risk with extremal dependence: Asymptotic analysis and efficient simulation', *Operations Research* **56**(3), 593–606.

Brémaud, Pierre (1980), *Point Processes and Queues – Martingale Dynamics*, Springer-Verlag, New York.

Bucklew, James A., Peter Ney & John S. Sadowsky (1990), 'Monte Carlo simulation and large deviations theory for uniformly recurrent Markov chains', *Journal of Applied Probability* **27**(1), 44–59.

Carmona, René, Jean-Pierre Fouque & Douglas Vestal (2009), 'Interacting particle systems for the computation of rare credit portfolio losses', *Finance and Stochastics* **13**(4), 613–633.

Carmona, René & Stéphane Crépey (2010), 'Particle methods for the estimation of Markovian credit portfolio loss distributions', *International Journal of Theoretical and Applied Finance* **13**(4), 577–602.

Chen, Zhiyong & Paul Glasserman (2008), 'Fast pricing of basket default swaps', *Operations Research* **56**(2), 286–303.

Das, Sanjiv, Darrell Duffie, Nikunj Kapadia & Leandro Saita (2007), 'Common failings: How corporate defaults are correlated', *Journal of Finance* **62**, 93–117.

Del Moral, Pierre & Josselin Garnier (2005), 'Genealogical particle analysis of rare events', *Annals of Applied Probability* **15**, 2496–2534.

Dembo, Amir & Ofer Zeitouni (1998), *Large Deviations Techniques and Applications*, Springer-Verlag, New York.

Deng, Shaojie, Kay Giesecke & Tze Leung Lai (2011), Sequential importance sampling and resampling for dynamic portfolio credit risk. *Operations Research*, forthcoming.

Etemadi, Nasrollah (1983), 'On the law of large numbers for nonnegative random variables', *Journal of Multivariate Analysis* **13**, 187–193.

Giesecke, Kay & Alexander Shkolnik (2010), Importance sampling for indicator Markov chains, *in* B.Johansson, S.Jain, J.Montoya-Torres, J.Hugan & E.Yücesan, eds, 'Proceedings of the 2010 Winter Simulation Conference', IEEE Press, pp. 967–975.

Giesecke, Kay, Baeho Kim & Shilin Zhu (2011), Monte carlo algorithms for default timing problems. *Management Science*, forthcoming.

Giesecke, Kay, Hossein Kakavand, Mohammad Mousavi & Hideyuki Takada (2010), 'Exact and efficient simulation of correlated defaults', *SIAM Journal of Financial Mathematics* **1**, 868–696.

Glasserman, Paul & Jingyi Li (2005), 'Importance sampling for portfolio credit risk', *Management Science* **51**(11), 1643–1656.

Glasserman, Paul, Wanmo Kang & Perwez Shahabuddin (2008), 'Fast simulation of multifactor portfolio credit risk', *Operations Research* **56**(5), 1200–1217.

Glasserman, Paul & Yashan Wang (1997), 'Counterexamples in importance sampling for large deviations probabilities', *Annals of Applied Probability* **7**(3), 731–746.

Goyal, Ambuj, Perwez Shahabuddin, Philip Heidelberger, Victor F. Nicola & Peter W. Glynn (1992), 'A unified framework for simulating markovian models of highly dependable systems', *IEEE Transactions On Computers* **41**(1), 36–51.

Jacod, Jean (1975), 'Multivariate point processes: Predictable projection, radon-nikodym derivatives, representation of martingales', *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **31**, 235–253.

Joshi, Mark & Dherminder Kainth (2004), 'Rapid and accurate development of prices and greeks for nth to default credit swaps in the Li model', *Quantitative Finance* **4**, 266–275.

Juneja, Sandeep & Perwez Shahabuddin (2001), 'Efficient simulation of markov chains with small transition probabilities', *Management Science* **47**(4), 547–562.

Nakayama, Mervin (1996), 'General conditions for bounded relative error in simulations of highly reliable markovian systems', *Advances in Applied Probability* **28**, 687–727.

Protter, Philip (2004), *Stochastic Integration and Differential Equations*, Springer-Verlag, New York.

Sadowsky, John S. & James A. Bucklew (1990), 'On large deviations theory and asymptotically efficient Monte Carlo simulation', *IEEE Transactions of Information Theory* **36**(3), 579–588.

Shahabuddin, Perwez (1994), ‘Importance sampling for the simulation of highly reliable markovian systems’, *Management Science* **40**(3), 333–352.

Siegmund, David (1976), ‘Importance sampling in the monte carlo study of sequential tests’, *The Annals of Statistics* **4**(4), 673–684.

Zhang, Xiaowei, Jose Blanchet, Kay Giesecke & Peter W. Glynn (2011), Affine point processes: asymptotic analysis and rare event simulation. Working Paper, Stanford University.