Conditional Importance Sampling for Event Counting Processes

Baheo Kim[*] and Alex Shkolnik[†]

September 9, 2021

Abstract

This article develops *conditional* importance sampling for a wide range of point process models for event counting. We design, implement and test a simple and practically efficient importance sampling scheme for tail probabilities. Our approach incorporates adaptive probability measure changes conditional on the tail event of interest. Numerical results demonstrate its application to estimating default clustering likelihoods in a stochastic network, pricing interest rate derivatives, and calculating overflow probabilities of dynamic queuing systems with efficient performance.

[*]Korea University Business School, Seoul, 02841, South Korea, Email: `baehokim@korea.ac.kr`.

[†]Department of Statistics and Applied Probability, University of California, Santa Barbara, CA and Consortium for Data Analytics in Risk, University of California, Berkeley, CA. Email: `shkolnik@ucsb.edu`.

1. Introduction

The problem of estimating the likelihood that a given number of events is observed by some horizon arises naturally in a wide range of disciplines (e.g. finance, queuing, insurance, healthcare, etc.). For instance, investors and fixed-income managers need to consider the odd chance a portfolio of defaultable assets (e.g. loans, bonds, etc.) experiences a large number of credit events. As such, tail probability is essential to estimate several popular (financial) risk measures, including value at risk and expected shortfall. Insurance risk management asks for gauging the probability of ruin due to excessive number of claims. In reliability studies, the likelihood of systemic breakdown caused by failure of too many system components in too short a time is often emphasized. The conditional tail distribution of health events caused by virus pandemics must be accurately estimated so that health authorities can minimize economic costs and social disruption. Unbiased and efficient algorithms for these problems are important not only to provide numerical values, but also to investigate how tail phenomena occur in a particular model via simulation.

This paper develops a novel, facile and fast Monte Carlo (MC) estimator of rare event probabilities described in the examples above. We focus on event timing models whose dynamics may be prescribed by a (multivariate) point process. Such models and related simulation algorithms are widely used in the literature. The events arrive at some stochastic rate, or through an *intensity* process that is typically modeled as a system of stochastic differential equations. The proposed algorithms accommodate any intensity specification, provided it can be simulated. However, plain MC is inadequate when intensity is too small to produce a sufficient number of events to achieve the rare event threshold, and hence can be highly inaccurate for calculating tail probabilities.

It is well-known that plain Monte Carlo is extremely inefficient to estimate rare event probabilities. More precisely, the number of simulation trials required to estimate small probability events broadly scales in proportion to one over its square-root [AG07, Chapter VI]. Importance sampling (IS) is a technique that facilitates efficient rare event simulation. For our specific focus, IS entails sampling under an *importance* measure (distinct from the reference distribution) where the simulated event times occur more frequently. A classical measure change that yields such outcomes is known as exponential twisting, which was originally designed by [Sie76] for computing gambler's ruin probabilities. However, its application typically hinges on the knowledge of a problem related moment generating function (MGF). Computing the

MGF is usually impossible except for simple event timing models; e.g., when the total intensity follows a (doubly-stochastic) Poisson process. As such, MGF is typically intractable for event time intensities governed by correlated stochastic processes with contagion or clustering interactions under the model specifications typically referred to as *self-exciting* point processes in the literature. This prohibits the application of IS schemes based on exponential twisting to many empirically motivated models of event timing.

Motivated by recent work such as [GS20], [BR16], [GS14] and [VEW12], we utilize the Girsanov-Meyer, exponential martingale approach to construct the measure change. This is further coupled with a sequence of zero-variance distributions that converge (weakly) to the proposed importance measure. The latter step provides further efficiency gains by ensuring each simulated path hits the rare event with probability one. This mathematical novelty extends the scope of the algorithm to other parts of the distribution besides the (right) tail. Perhaps more importantly, the limiting measure possesses attractive properties for simulation. While the event counting process has the interpretation of a *limit* Poisson process with *zero* intensity conditional on the tail event of interest under the auxiliary conditional importance measure, its arrival times are uniformly distributed up to the chosen horizon. The remaining stochastic variables (pertaining to the original intensity process) have the same dynamics as under the reference measure. Thus, remarkably, the simulation is simpler and faster than the process generating plain MC estimators. Moreover, all simulation bias of the estimator is captured by the Radon-Nikodym derivative between the importance and reference measures. The proposed approach can also reduce sampling bias, which often plagues event time simulation estimators; see [GS20] for an in-depth discussion. To our knowledge, the limit of measure change and its application to estimating probability of rare events are new in this paper.

Related studies have usually considered asymptotic optimality of IS estimators, which demands that the second moment of the estimator under the importance measure decay at twice the rate of decay of the rare event probability. However, a tradeoff usually arises between the restrictive conditions required for optimality and algorithm specifications that are too difficult to implement in general. The proposed algorithms pursue weaker notions of optimality while preserving both speed and simplicity; i.e., significant gains in speed can often offset increased variance by incorporating more simulation trials (albeit at the slower, CLT square-root rate). In this regard, we prove that the proposed *conditional* IS (cIS) scheme outperforms plain MC under mild conditions of rare events, providing asymptotic variance reduction. We further show that the cIS estimator is asymptotically optimal, but only under restrictive conditions, when the intensity is sufficiently small for the rare event threshold. This ensures the proposed estimator is preferred compared with the standard approach to under rare event regimes.

Previous research has investigated tail estimation in the context of event timing simulation; refer to [GS20] and [GS14] for a thorough review of this literature. We do not discuss exponential twisting studies here, as its scope is restricted in our setting; see [RT09] for a broad range of applications in rare event simulation techniques. Several studies have developed interacting particle schemes to address such difficulties (e.g., [DG05], [CFV09], [CC10]). The current paper is most closely related to [GS10] which develops an asymptotically optimal IS scheme for Markov chain models under a strict set of assumptions.

The proposed algorithms are numerically tested on several applications. Our first numerical example is related to credit risk management using a popular reduced-form model of correlated name-by-name default timing in a network. The model incorporates frailty and self-exciting features central to corporate debt modeling; e.g., see [DEHS09] and [AGS18] among many others. The second example studies the case of zero-coupon bond pricing under a non-trivial short-rate model with jump diffusion over various time-to-maturities. In the third example, the proposed algorithm is applied for estimating overflow probabilities of a series of time-inhomogeneous queues, which can be further extended to a variety of consequences such as catastrophic failures to meet the demand for intensive healthcare units, urgent transportation services, emergency call centers, and so forth. Our experiments illustrate superior performance of the proposed cIS scheme compared with plain MC.

The remainder of this paper is structured as follows. Section 2 formulates the problem of interest. Section 3 develops the proposed conditional IS scheme. Section 4 analyzes asymptotic efficiency of the conditional IS estimator, and Section 5 provides numerical examples. Finally, Section 6 concludes the paper. Mathematical proofs and computational details are provided as online Appendices.

2. Problem Formulation

This section outlines the formulation of point processes, followed by the overall objective and the main scope of this article.

2.1. Preliminaries. Fix a measure space $(\Omega, \mathcal{F})$ equipped with a right-continuous and complete information filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t\geq 0}$ satisfying the usual conditions. Let

P be the probability measure of interest on $(\Omega, \mathcal{F})$, where $\mathrm{E}^{\mathrm{P}}(\cdot)$ denotes the expectation under P. In a statistical application, such as real-world risk management, one typically takes P as the physical probability measure. On the other hand, P can be taken as a risk-neutral measure for derivatives pricing purposes. In Section 3, we further construct auxiliary conditional importance sampling measures on the measure space $(\Omega, \mathcal{F})$.

On the complete probability space $(\Omega, \mathcal{F}, \mathrm{P})$, we consider for some $n \in \mathbb{N}$ a set of distinct stopping times $\{T_i\}_{i=1}^n$, which are totally inaccessible in the sense that they can never be predicted by an increasing sequence of stopping times. We further assume $T_i < \infty$ for all $i = 1, \ldots, n$ almost surely, and each event time T_i is associated with an event indicator process $N_t^i = 1_{\{T_i \leq t\}}$ for $t \geq 0$, where $1_{\mathscr{A}}$ is the indicator of $\mathscr{A} \in \mathcal{F}$.

We define $\mathrm{N} = (N^1, \ldots, N^n)$ and let $\{\tau_i\}_{i=1}^n$ be the *order* statistics of $\{T_i\}_{i=1}^n$. Thus, τ_j is the j^{th} arrival time of the aggregate counting process $\overline{N}_t = \sum_{i=1}^n N_t^i$, where we let $\tau_0 = 0$ for notational consistency. Suppose that each N^i admits an intensity, $p^i = X^i(1 - N^i)$, for an integrable and càdlàg process $X^i > 0$. In other words, each p^i is a conditional mean rate of event occurrence of N^i in the sense that

$$p_t^i = \lim_{\Delta \downarrow 0} \frac{1}{\Delta} \mathrm{P}\left(N_{t+\Delta}^i - N_t^i > 0 \middle| \mathcal{F}_t\right),$$

and $N_t^i - \int_0^t p_s^i ds$ forms a martingale for $i = 1, 2, \ldots, n$. It follows that $\overline{N}$ admits $\overline{p} = \sum_{i=1}^n p^i$ as its intensity, where $\overline{p}$ is strictly positive almost everywhere on $[0, \tau_n)$.

2.2. Objective. Considering a sequence of *rare events* $\{\mathcal{E}_k\}_{k=1}^\infty$ with a rarity parameter $k \in \mathbb{N}$, we wish to compute the sequence of rare event probabilities $Y_k = \mathrm{P}(\mathcal{E}_k) \to 0$ as k increases. A typical example of such events is $\mathcal{E}_k = \left\{\overline{N}_T \geq k\right\}$ or $\mathcal{E}_k = \left\{\overline{N}_T = k\right\}$ for a fixed horizon $T > 0$. In addition, it is also of our interest to analyze the asymptotic properties of $\mathcal{E}_k = \left\{\overline{N}_T \geq \mu k\right\}$ for some fixed $\mu \in (0, 1)$ by taking the rarity parameter $k = n \uparrow \infty$.

In principle, a Monte Carlo (MC) sampling scheme can be characterized by an absolutely continuous change of measure from P to *some* sampling measure Q based on the identity $Y_k = \mathrm{E}^{\mathrm{Q}}(\mathcal{L}_k 1_{\mathcal{E}_k}) = \mathrm{P}(\mathcal{E}_k)$, where $\mathrm{E}^{\mathrm{Q}}(\cdot)$ denotes the expectation under Q, and $\mathcal{L}_k$ is the Radon-Nikodym derivative of P with respect to Q. For a given k, the true value of Y_k could be estimated by generating a set of independent and identically distributed (i.i.d.) Q-samples $\{\hat{y}_m\}_{m=1}^M$ of $\mathcal{L}_k 1_{\mathcal{E}_k}$ for some $M \in \mathbb{N}$ representing the number of simulation trials. Then, we take the empirical mean $\hat{Y}_M = \frac{1}{M} \sum_{m=1}^M \hat{y}_m$ as the Q-estimator of Y_k. The root mean square error, defined as $\mathrm{RMSE} = \sqrt{\mathrm{bias}^2 + \mathrm{SE}^2}$

is commonly used to evaluate performance across different MC schemes, where the bias is given by $\mathrm{E}^{\mathrm{Q}}(\hat{Y}_M) - Y_k$ and the standard error (SE) is the Q-sample standard deviation of $\{\hat{y}_m\}_{m=1}^{M}$ divided by $\sqrt{M}$. Our goal is to achieve relatively fast RMSE convergence by adopting an efficient simulation measure specific to the *rare* events of interest.

In particular, the *plain* Monte Carlo (pMC) scheme takes $\mathrm{Q} = \mathrm{P}$, which implies that $\mathcal{L}_k = 1$ and $Y_k = \mathrm{E}^{\mathrm{P}}(1_{\mathcal{E}_k}) = \mathrm{P}(\mathcal{E}_k)$. The pMC method is computationally inefficient for estimating rare event probabilities, as the number of pMC simulation trials required to achieve a given relative precision tends to infinity as the rarity parameter increases. The central limit theorem shows that we need a factor of Ψ more replications to reduce the confidence interval of pMC estimators by $\sqrt{\Psi}$. As such, heavy computational burden is unavoidable for simulating rare events with a small probability in that the pMC methods generally fail to induce reasonable precision of confidence intervals unless the number of simulation trials $M \gg 1/Y_k$, where $Y_k \to 0$ as the rarity parameter $k \uparrow \infty$; e.g., refer to [RT09] and the references therein for related discussions.

In turn, we propose the *conditional importance sampling* (cIS) measure, denoted by Q_k, under which we have $1_{\mathcal{E}_k} = 1$ almost surely and $Y_k = \mathrm{E}^{\mathrm{Q}_k}(\mathcal{L}_k)$. Our main objective is to improve efficiency in the estimation of rare event probabilities, and thereby reducing the *standard error* (SE), when the pMC method is not practically appropriate. In this sense, this article is complementary to [GS20] whose main objective is reducing the *bias*; see [GS20] for a rigorous discussion on the issues of bias reduction. Thus, we investigate the robustness properties of the simulation estimators with respect to rarity by focusing on the reduction of SE under the cIS scheme.

3. Conditional Importance Sampling Measure

We propose a method of *conditional* IS (cIS) by incorporating adaptive measure changes conditional on the event of interest. For this purpose, define $\mathcal{E}_k = \left\{\overline{N}_T \geq k\right\}$ and

$$\mathcal{L}_k = \exp\left(-\int_0^{\tau_k} \overline{p}_s ds\right) \prod_{j=1}^{k} \overline{p}_{\tau_j-} \frac{T^k}{k!}$$

for a given simulation horizon $T > 0$ and some $k \in \mathbb{N}$. We then construct the cIS simulation measure specific to the event $\mathcal{E}_k$ as $\mathrm{Q}_k(\mathscr{A}) = \mathrm{E}^{\mathrm{P}}\left(\frac{1_{\mathcal{E}_k \cap \mathscr{A}}}{\mathcal{L}_k}\right)$ for all $\mathscr{A} \in \mathcal{F}_{\tau_k}$, where Q_k is designed to be absolutely continuous with respect to P.

Theorem 3.1. *The following statements are true:*

(i) The cIS simulation measure Q_k *is a well-defined probability measure on* $(\Omega, \mathcal{F}_{\tau_k})$.

(ii) $\mathrm{P}(\mathcal{E}_k) = \mathrm{E}^{\mathrm{Q}_k}(\mathcal{L}_k)$*; i.e., the estimator of* $\mathrm{P}(\mathcal{E}_k)$ *under* Q_k *is* $\mathcal{L}_k$.

(iii) If $\{u_i\}_{i=1}^k$ *denotes the order statistics of* k *i.i.d. uniform random variables on* $[0, T]$*, we have* $\tau_i \stackrel{\mathcal{D}}{=} u_i$ *under* Q_k *for* $i = 1, \ldots, k$.

(iv) $\mathrm{Q}_k\left(I_j = i \,\middle|\, \mathcal{H}_{\tau_j -}\right) = \mathrm{P}\left(I_j = i \,\middle|\, \mathcal{F}_{\tau_j -}\right) = p^i_{\tau_j -}\Big/\overline{p}_{\tau_j -}$ *holds for all* $j = 1, \ldots, k$*, where* I_j *is the component of* N *at which the* j^{th} *event of* $\overline{N}$ *occurs and* $\mathcal{H}_t = \mathcal{F}_t \vee \sigma\left\{\tau_j\right\}_{j=1}^k$.

(v) Any P*-local martingale that has no jumps in common with* $\overline{N}$ *is a* Q_k*-local martingale. Thus, a* P*-Brownian motion is also a* Q_k*-Brownian motion.*

The proof of Theorem 3.1 is provided in Appendix A.1. To our best knowledge, the cIS simulation measure Q_k is unique albeit its simplicity and versatility; the Girsanov-Meyer exponential martingale approach to construct the measure change is coupled with a sequence of zero-variance distributions that weakly converge to the proposed importance measure. While the mathematical setting is similar to [GS20], the limit of measure change and its application to estimating probability of rare events are new to our knowledge. We illustrate the intuition of Theorem 3.1 in the following examples.

Example 1. (Homogeneous Poisson process) Suppose that $\overline{N}_t$ follows a homogeneous Poisson process with its intensity $\overline{p}_t = \lambda > 0$ for all $t \in [0, T]$. Theorem 3.1 implies that

$$\mathrm{P}\left(\overline{N}_T \geq k\right) = \mathrm{E}^{\mathrm{Q}_k}(\mathcal{L}_k) = \frac{(\lambda T)^k}{k!}\mathrm{E}^{\mathrm{Q}_k}\left(e^{-\lambda \tau_k}\right),$$

where the Q_k-distribution of $\{\tau_1, \ldots, \tau_k\}$ is that of the uniform order statistics on $[0, T]$. This means that $\beta_k = \tau_k / T$ follows the $\mathrm{Beta}(k, 1)$ distribution under Q_k and the Laplace transform of τ_k is related to the moment generating function of β_k taking the form of

$$\mathrm{E}^{\mathrm{Q}_k}\left(e^{-\lambda \tau_k}\right) = \mathrm{E}^{\mathrm{Q}_k}\left(e^{-\lambda T \beta_k}\right) = {}_1F_1(k, k+1, -\lambda T),$$

which can be expressed as the Kummer confluent hypergeometric function with parameters $k, k+1$ and $-\lambda T$.

Example 2. (Inhomogeneous Poisson process) Suppose that $\overline{N}_t = \sum_{i=1}^n N_t^i$ is a time-inhomogeneous Poisson process for some $n \in \mathbb{N}$ with its intensity $\overline{p}_t = n \cdot t$ for all $t \in [0, T]$. For a fixed $k \leq n$, we apply Theorem 3.1 to compute $\mathrm{P}\left(\overline{N}_T \geq k\right) = \mathrm{E}^{\mathrm{Q}_k}\left(\mathcal{L}_k\right)$, where

$$\mathcal{L}_k = \exp\left(-\int_0^{\tau_k} \overline{p}_s ds\right) \prod_{j=1}^k \overline{p}_{\tau_j -} \frac{T^k}{k!} = \frac{(nT)^k}{k!} \exp\left(-\frac{n}{2}\tau_k^2\right) \prod_{j=1}^k \tau_j .$$

Recall that the Q_k-distribution of $\{\tau_1, \ldots, \tau_k\}$ is that of the uniform order statistics on $[0, T]$. This implies that

$$V_1 \triangleq \frac{\tau_1}{\tau_2},\ V_2 \triangleq \left(\frac{\tau_2}{\tau_3}\right)^2, \ldots, V_{k-1} \triangleq \left(\frac{\tau_{k-1}}{\tau_k}\right)^{k-1}, V_k \triangleq \left(\frac{\tau_k}{T}\right)^k$$

are independent standard uniform random variables under Q_k. Hence, we can derive

$$\begin{aligned}
\mathrm{P}\left(\overline{N}_T \geq k\right) = \mathrm{E}^{\mathrm{Q}_k}\left(\mathcal{L}_k\right) &= \frac{(nT)^k}{k!}\mathrm{E}^{\mathrm{Q}_k}\left(\exp\left(-\frac{n}{2}\tau_k^2\right)\prod_{j=1}^k \tau_j\right) \\
&= \frac{(nT^2)^k}{k!}\mathrm{E}^{\mathrm{Q}_k}\left(\exp\left(-\frac{nT^2}{2}V_k^{\frac{2}{k}}\right)V_k\right)\prod_{j=1}^{k-1}\mathrm{E}^{\mathrm{Q}_k}\left(V_j\right) \\
&= \frac{(nT^2/2)^k}{k!} \cdot 2\mathrm{E}^{\mathrm{Q}_k}\left(\exp\left(-\frac{nT^2}{2}U^{\frac{2}{k}}\right)U\right) \\
&= \frac{(nT^2/2)^k}{k!}\,{}_1F_1\left(k, k+1, -\frac{nT^2}{2}\right),
\end{aligned}$$

where U is a Q_k-standard uniform random variable. Note that this result is a special case of **Example 1** with $\lambda = \frac{nT}{2}$.

Algorithm 1 Conditional Monte Carlo to estimate $\mathrm{P}\left(\overline{N}_T \geq k\right)$

1: **procedure** CONDITIONALIS(k, T, M) $\triangleright$ M is the number of cIS trials
2: Initialize $\hat{Y} \leftarrow 0$
3: **for** $m \in \{1, \ldots, M\}$ **do**
4: Draw an ordered sample of $\{\tau_1, \ldots, \tau_k\}$ from $U(0, T)$
5: Compute $\mathcal{L}_k$ from a conditional sample of $\overline{p}$ under Q_k
6: Update $\hat{Y} \leftarrow \hat{Y} + \mathcal{L}_k$
7: **end for**
8: **return** $\hat{Y}/M$
9: **end procedure**

Estimating tail probabilities under Q_k provides nontrivial efficiency gains by ensuring each simulated path hits the rare event with probability one. More importantly, the limiting measure possesses attractive properties for simulation. While $\overline{p}$ is no longer the intensity of $\overline{N}$ under Q_k, the arrival times of $\overline{N}$ are uniformly distributed up to the chosen horizon. The remaining stochastic variables (pertaining to the original intensity process) have the same dynamics as under the reference measure. Thus, remarkably, the simulation is simpler and faster than the process generating plain MC estimators. Algorithm 1 summarizes the proposed cIS algorithm, which provides computational efficiency in that sampling the skeleton of $\overline{p}$ on $\{\tau_1, \ldots, \tau_k\}$ is relatively cheap, since events times are uniformly distributed on $[0, T]$ under Q_k. In particular, $\mathcal{E}_k = \left\{\overline{N}_T \geq k\right\}$ always occurs under Q_k, as $\mathcal{L}_k$ derives the zero-variance conditional measure change with respect to $\mathcal{E}_k$.

The cIS algorithm generates an unbiased estimator of $Y = \mathrm{P}\left(\overline{N}_T \geq k\right)$, if one can evaluate $\mathcal{L}_k$ exactly under Q_k. Thus, the cIS estimator is unbiased, if one can exactly evaluate the time-integrated transform $\exp\left(-\int_0^{\tau_k} \overline{p}_s ds\right)$ without biases. Furthermore, the unbiased cIS estimator $\mathcal{L}_k$ under Q_k can be sampled efficiently in the sense that

$$
\begin{aligned}
\mathrm{P}\left(\overline{N}_T \geq k\right) &= \mathrm{E}^{\mathrm{Q}_k}\left(\mathcal{L}_k\right) \\
&= \frac{T^k}{k!}\mathrm{E}^{\mathrm{Q}_k}\left(\prod_{j=1}^{k} \overline{p}_{\tau_j-} \exp\left(-\int_{\tau_{j-1}}^{\tau_j} \overline{p}_s ds\right)\right) \\
&= \frac{T^k}{k!}\mathrm{E}^{\mathrm{Q}_k}\left(\prod_{j=1}^{k} \overline{p}_{\tau_j-}\mathrm{E}^{\mathrm{Q}_k}\left(\exp\left(-\int_{\tau_{j-1}}^{\tau_j} \overline{p}_s ds\right)\middle| \tau_{j-1}, \tau_j, \overline{p}_{\tau_{j-1}}, \overline{p}_{\tau_j-}\right)\right),
\end{aligned}
$$

where the last equality holds when $\overline{p}$ satisfies the Markov property between the consecutive event times; i.e., the *inter-arrival* intensity is a Markov process. This implies that an unbiased estimator of $\mathrm{P}\left(\overline{N}_T \geq k\right)$ is available by sampling $\tau_1, \tau_2, \ldots, \tau_j$ under Q_k, if exact samples of $\overline{p}_{\tau_j-}$ can be simulated conditional on $\overline{p}_{\tau_{j-1}}$ for $j = 1, 2, \ldots, k$ and the bridge transform $\mathrm{E}^{\mathrm{Q}_k}\left(\exp\left(-\int_{\tau_{j-1}}^{\tau_j} \overline{p}_s ds\right)\middle| \tau_{j-1}, \tau_j, \overline{p}_{\tau_{j-1}}, \overline{p}_{\tau_j-}\right)$ can be evaluated analytically. We demonstrate such cases in Sections 5.1 and 5.2 with numerical illustrations.

Furthermore, the cIS algorithm does not require *ad-hoc* level selection or tuning procedures to search for any *optimal* parameter specific to the event $\mathcal{E}_k = \left\{\overline{N}_T \geq k\right\}$ for estimating $\mathrm{P}\left(\overline{N}_T \geq k\right)$, as the proposed cIS scheme ensures each simulated path hits the rare event with probability one. This feature extends the scope of the cIS algorithm to other parts of the distribution besides the (right) tail. Corollary 3.2 states that we can subsequently apply a version of the cIS algorithm to estimate $Y_k = \mathrm{P}(\overline{N}_T = k)$ for $k = 1, \ldots, n$.

Corollary 3.2. *The following statements are true:*

(i) There exists a version of the cIS simulation measure $\widehat{\mathrm{Q}}_k$ in the sense that $\mathrm{P}\left(\widehat{\mathcal{E}}_k\right) = \mathrm{E}^{\widehat{\mathrm{Q}}_k}\left(\widehat{\mathcal{L}}_k\right)$, where the limit conditional measure $\widehat{\mathrm{Q}}_k$ is redefined for the modified event set denoted by $\widehat{\mathcal{E}}_k = \{\overline{N}_T = k\}$ and $\widehat{\mathcal{L}}_k$ is given by

$$\widehat{\mathcal{L}}_k = \exp\left(-\int_0^T \overline{p}_s ds\right) \prod_{j=1}^{k} \overline{p}_{\tau_j -} \frac{T^k}{k!}$$

for a fixed simulation horizon $T > 0$.

(ii) Statements (iii)–(v) of Theorem 3.1 still hold with $\widehat{\mathrm{Q}}_k$.

Example 3. (Homogeneous Poisson process) We revisit **Example 1** with the homogeneous Poisson process with its intensity $\overline{p}_t = \lambda$ for all $t \in [0, T]$. Corollary 3.2 implies that

$$\mathrm{P}\left(\overline{N}_T = k\right) = \mathrm{E}^{\widehat{\mathrm{Q}}_k}\left(\widehat{\mathcal{L}}_k\right) = \frac{(\lambda T)^k}{k!} e^{-\lambda T},$$

which is a well-known probability of observing the event $\left\{\overline{N}_T = k\right\}$ under P.

Example 4. (Inhomogeneous Poisson process) We next revisit **Example 2** with the time-inhomogeneous Poisson process with its intensity $\overline{p}_t = n \cdot t$ for all $t \in [0, T]$. For a fixed $k \leq n$, Corollary 3.2 implies that $\mathrm{P}\left(\overline{N}_T = k\right) = \mathrm{E}^{\widehat{\mathrm{Q}}_k}\left(\widehat{\mathcal{L}}_k\right)$, where

$$\widehat{\mathcal{L}}_k = \exp\left(-\int_0^T \overline{p}_s ds\right) \prod_{j=1}^{k} \overline{p}_{\tau_j -} \frac{T^k}{k!} = \frac{(nT)^k}{k!} \exp\left(-\frac{n}{2}T^2\right) \prod_{j=1}^{k} \tau_j .$$

With the same definition of $\{V_1, V_2, \ldots, V_k\}$ as

$$V_1 \triangleq \frac{\tau_1}{\tau_2},\ V_2 \triangleq \left(\frac{\tau_2}{\tau_3}\right)^2, \ \ldots, V_{k-1} \triangleq \left(\frac{\tau_{k-1}}{\tau_k}\right)^{k-1}, V_k \triangleq \left(\frac{\tau_k}{T}\right)^k,$$

which are independent standard uniform random variables under $\widehat{\mathrm{Q}}_k$, we can derive

$$\mathrm{P}\left(\overline{N}_T = k\right) = \mathrm{E}^{\widehat{\mathrm{Q}}_k}\left(\widehat{\mathcal{L}}_k\right) = \frac{(nT)^k}{k!} \mathrm{E}^{\widehat{\mathrm{Q}}_k}\left(\exp\left(-\frac{n}{2}T^2\right) \prod_{j=1}^{k} \tau_j\right)$$

$$
\begin{aligned}
&= \frac{(nT^2)^k}{k!} \exp\left(-\frac{nT^2}{2}\right) \prod_{j=1}^{k} \mathrm{E}^{\widehat{\mathrm{Q}}_k}\left(V_j\right) \\
&= \frac{(nT^2/2)^k}{k!} \exp\left(-\frac{nT^2}{2}\right),
\end{aligned}
$$

which is a special case of **Example 3** with $\lambda = \frac{nT}{2}$.

4. Asymptotic Properties

4.1. Short-time/small-intensity regime. We consider small horizons T for which $y_T = \mathrm{P}(\overline{N}_T \geq \lceil \mu n \rceil)$ is a small probability as $y_T \downarrow 0$ as $T \downarrow 0$. Note that this is equivalent to assuming that the intensity of each even is small, e.g., for $\epsilon > 0$

Theorem 4.1. *The cIS estimator of y_T has bounded relative error. Moreover, if $\overline{p}_0$ is determistic cIS has vanishing relative error.*

Proof. For $\mathscr{E}_T = \{\overline{N}_T \geq k\}$ we have that $\mathcal{L}_T$ and $1_{\mathscr{E}_T}$ are Q- and P-estmators of y_T. The squared relative error of $\mathcal{L}_T$ is (let u_j be the j uniform statistic on $[0, T]$)

$$
\begin{aligned}
\frac{\mathrm{Var}_{\mathrm{Q}}(\mathcal{L}_L)}{y_T^2} &= \frac{\mathrm{E}_Q(\mathcal{L}_T^2) - \mathrm{E}(\mathcal{L}_T)^2}{\mathrm{E}(\mathcal{L}_T)^2} \\
&= \frac{\mathrm{E}\big(\exp\big(-2\int_0^{u_k} \overline{p}_s \,\mathrm{d}s\big) \prod_{j=1}^{k} \overline{p}_{u_j-}^2\big)}{\mathrm{E}\big(\exp\big(-\int_0^{u_k} \overline{p}_s \,\mathrm{d}s\big) \prod_{j=1}^{k} \overline{p}_{u_j-}\big)^2} - 1
\end{aligned}
$$

As $T \downarrow 0$ the relative error has the limit $\sqrt{\mathrm{E}(\overline{p}_0^{2k})/\mathrm{E}(\overline{p}_0^k)^2 - 1}$ which is clearly bounded (in fact we have vanishing relative error provided that $\overline{p}_0$ is not random). ∎

4.2. Small-intensity regime. Consider $N^{\epsilon,i}$ with intensity ϵp^i, i.e.,

$$
N_t^{\epsilon,i} - \int_0^t \epsilon p_s^i \,\mathrm{d}s
$$

forms a martingale.

Theorem 4.2. *The cIS estimator of $y_\epsilon = \mathrm{P}(\overline{N}_T^\epsilon \geq k)$ has bounded relative error. Moreover, if $\overline{p}_0$ is determistic cIS has vanishing relative error.*

4.3. Large system regime. In this section, we develop conditions guaranteeing asymptotic efficiency of the cIS estimator in terms of variance reduction. We adopt the rare event regime $y_n = \mathrm{P}(\overline{N}_T \geq \lceil \mu n \rceil) \to 0$ as $n \uparrow \infty$ for some fixed $\mu \in (0, 1)$.

Observe that

$$\frac{1}{n}\mathrm{E}^{\mathrm{P}}(\overline{N}_T) = \int_0^T \frac{1}{n}\mathrm{E}^{\mathrm{P}}(\overline{p}_s)\, ds\,.$$

Thus, a sufficient condition for the rare event regime is given by

$$\mathrm{E}^{\mathrm{P}}(\overline{p}_t) < \frac{\mu n}{T} \tag{1}$$

for all $0 \le t \le T$ and sufficiently large n. By letting $\mathcal{L}_{\lceil \mu n \rceil} \triangleq \mathcal{L}_n$ and $\mathrm{Q}_{\lceil \mu n \rceil} \triangleq \mathrm{Q}_n$ for notational simplicity in this section, we compare the variance of the plain MC estimator $\mathrm{Var}^{\mathrm{P}}(1_{\{\overline{N}_T \ge \lceil \mu n \rceil\}}) = y_n(1-y_n)$ with the Q_n-variance of the estimator $\mathcal{L}_n$ to obtain conditions for asymptotic variance reduction of the cIS estimator. Subsequent theorem shows that the proposed cIS estimator asymptotically outperforms pMC in terms of variance reduction under mild rare event conditions.

Theorem 4.3. *(Sufficient condition for asymptotic variance reduction) Suppose that*

$$\frac{1}{\lceil \mu n \rceil}\sum_{k=1}^{\lceil \mu n \rceil} \overline{p}_{\tau_k -} < \frac{2\mu n}{eT} \tag{2}$$

holds Q_n-almost surely for sufficiently large n. Then, the cIS estimator achieves asymptotic variance reduction; i.e.,

$$\frac{\mathrm{Var}^{\mathrm{Q}_n}(\mathcal{L}_n)}{\mathrm{Var}^{\mathrm{P}}(1_{\{\overline{N}_T \ge \lceil \mu n \rceil\}})} \to 0 \qquad as \quad n \uparrow \infty\,. \tag{3}$$

Remark 4.4. *The condition in (2), which is analogous to the condition in (1), intuitively requires the average of aggregate event count intensity at each event time not to be excessively large under the conditional importance sampling measure. In other words, the performance of the cIS algorithm becomes more pronounced when the target probability gets even smaller.*

Remark 4.5. *If we relax the condition in (2) in the sense that*

$$\mathrm{Q}_n\left(\frac{1}{\lceil \mu n \rceil}\sum_{k=1}^{\lceil \mu n \rceil} \overline{p}_{\tau_k -} < \frac{2\mu n}{eT}\right) > 1 - \varepsilon$$

for some small $\varepsilon > 0$ and sufficiently large n, the cIS estimator achieves asymptotic sample *variance reduction with success rate at least $1 - \varepsilon$.*

Corollary 4.6. *(Sufficient condition for asymptotic optimality) Assume that $\overline{N}_T$ has a light-tailed distribution in the sense that there exists some constant $\gamma > 0$ satisfying*

$$\lim_{n\uparrow\infty} \frac{1}{n} \log \mathrm{P}\left(\overline{N}_T \geq \lceil \mu n \rceil\right) = -\gamma \tag{4}$$

and suppose that

$$\frac{1}{\lceil \mu n \rceil} \sum_{k=1}^{\lceil \mu n \rceil} \overline{p}_{\tau_k -} < \frac{2\mu n}{eT} \cdot e^{-\gamma/\mu} \tag{5}$$

holds Q_n-almost surely for sufficiently large n. Then, the cIS estimator achieves asymptotic optimality for variance reduction; i.e.,

$$\lim_{n\uparrow\infty} \frac{\log \mathrm{E}^{\mathrm{Q}_n}(\mathcal{L}_n^2)}{\log \mathrm{E}^{\mathrm{Q}_n}(\mathcal{L}_n)} = 2. \tag{6}$$

Remark 4.7. *It follows from Jensen's inequality that the lefthand side of equation (6) is bounded from above by 2. Corollary 4.6 implies that the cIS estimator is asymptotically optimal under relatively restrictive conditions, when $\overline{N}_T$ has a light-tailed distribution and the intensity is sufficiently small for the rare event threshold.*

5. Numerical Examples

This section illustrates the performance of the proposed cIS scheme under different settings to show its versatility and practical effectiveness in terms of variance reduction.

5.1. Networked default clustering. Multivariate counting process models are widely used to measure and manage correlated default risk in the system. For given $n \in \mathbb{N}$ defaultable names, the stopping times $\{T_i\}_{i=1}^n$ whose order statistics form $\{\tau_i\}_{i=1}^n$ model the default times of the individual constituents of the system, and $\overline{N}_T = \sum_{i=1}^n 1_{\{T_i \leq T\}}$ is the total default count by a fixed time $T > 0$ as the central (systemic) quantity of interest.

Our objective is to estimate the tail probability $\mathrm{P}\left(\overline{N}_T \geq k\right)$ for large k, where P refers to the physical probability measure. In this regard, we presume that there exists a systematic risk factor $\eta^0 \geq 0$ and a set of idiosyncratic factor processes $\{\eta^i\}_{i=1}^n$ such that the state process $\mathrm{X} = (X^1, \ldots, X^n)$ satisfies

$$X^i = \omega_i \eta^0 + \eta^i \tag{7}$$

for $i \in \{1, 2, \ldots, n\}$. Each default indicator process $N_t^i = 1_{\{T_i \leq t\}}$ admits $p_t^i = X_t^i(1 - N_t^i)$ as its P-intensity for $t \geq 0$. The systematic factor loading for the i^{th} name is denoted by $\omega_i > 0$ in equation (7). We assume that η^0 satisfies the SDE with some $\kappa_0 > 0$, $\theta_0 > 0$ and $\sigma_0 > 0$ given by

$$d\eta_t^0 = \kappa_0(\theta_0 - \eta_t^0)dt + \sigma_0\sqrt{\eta_t^0}dW_t^0,$$

where W^0 is a standard Brownian motion under P. Similarly, we also assume that the P-dynamics of η^i is given by

$$d\eta_t^i = \kappa_i(\theta_i - \eta_t^i)dt + \sigma_i\sqrt{\eta_t^i}dW_t^i + \delta_i \cdot dN_t,$$

where $(W^1, \ldots, W^n)$ is a vector of mutually independent standard Brownian motions under P, $\kappa_i > 0$ is the mean-reversion rate, $\theta_i > 0$ is the long-run mean level, $\sigma_i > 0$ is the diffusive volatility, and the vector $\delta_i = (\delta_{i1}, \ldots, \delta_{in}) \geq 0$ represents name i's sensitivity to defaults in the system for $i = 1, \ldots, n$; see, for example, [GKZ11] for a similar model specification.

We next introduce the *interarrival* intensity process h^{ij} satisfying

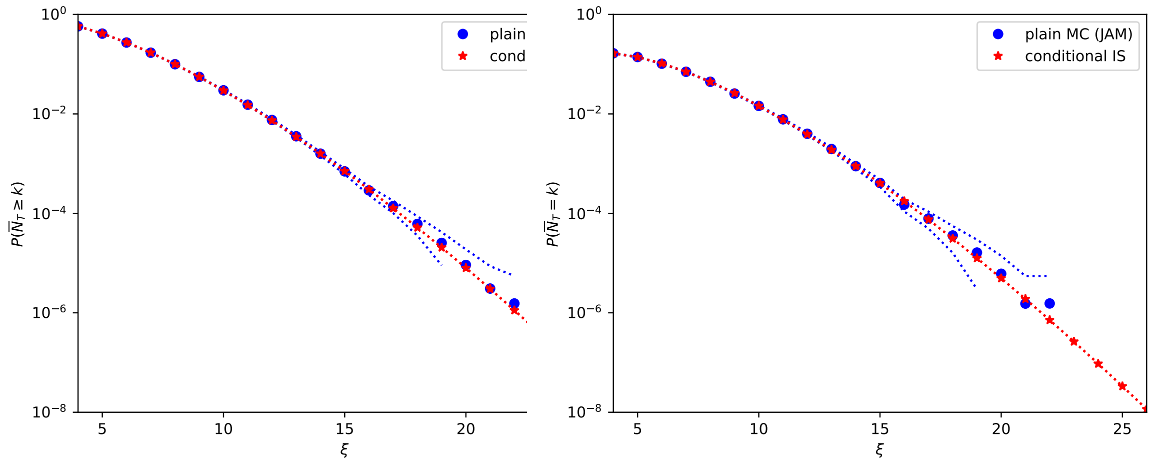
$$\eta_t^i = \sum_{j=1}^{n} h_{t-\tau_{j-1}}^{ij} 1_{\{\overline{N}_t = j-1\}}, \tag{8}$$

where $\tau_0 = 0$ and $h_0^{ij} = \eta_{\tau_{j-1}}^i$ for $j = 1, \ldots, n$. As the process h^{ij} follows the Feller diffusion expressed as $dh_t^{ij} = \kappa_i(\theta_i - h_t^{ij})dt + \sigma_i\sqrt{h_t^{ij}}dW_t^i$, the distribution of h_t^{ij} for $t > 0$ given h_0^{ij} is non-central chi-squared up to a scale factor. Specifically, [CIR85] showed that h_s^{ij} can be drawn from the non-central chi-squared transition density from h_0^{ij} as

$$\mathrm{P}\left(h_s^{ij} \leq z \,\middle|\, h_0^{ij}\right) = F_{\chi_d'^2(\nu)}\left(\frac{4\kappa_i z}{\sigma_i^2(1 - e^{-\kappa_i s})}\right), \tag{9}$$

where $d = \frac{4\kappa_i\theta_i}{\sigma^2}$ is the degree of freedom, $\nu = \frac{4\kappa_i e^{-\kappa_i s} h_0^{ij}}{\sigma_i^2(1-e^{-\kappa_i s})}$ is the non-centrality parameter. In this setup, it is shown by [BK06] that the bridge transform of the interarrival intensity process can be computed analytically for $a > 0$; see Appendix B.1 for the computational details. The closed-form evaluation of the bridge transform makes it possible to eliminate the source of bias from numerically integrating the sample paths of the intensity processes.

We estimate the tail probability $\mathrm{P}(\overline{N}_T \geq k)$ for portfolio size $n = 100$ with time horizon $T = 1$ year. For each name of $i = 1, \ldots, n$, we uniformly draw ω_i from $[0, 1]$, κ_i from $[0.5, 1.5]$, θ_i from $[0.001, 0.051]$, and set $\sigma_i = \min(\sqrt{2\kappa_i\theta_i}, \bar{\sigma}_i)$, where $\bar{\sigma}_i$ is



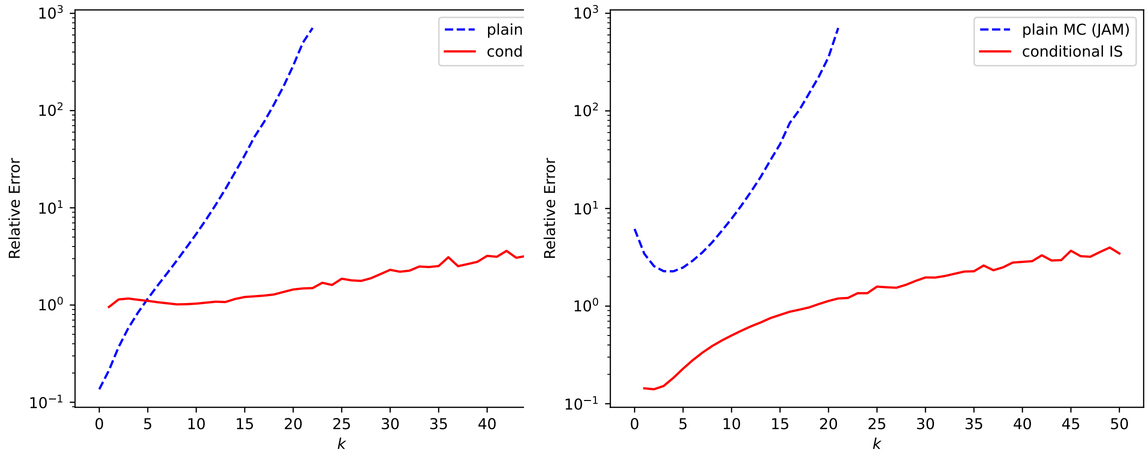
(a) *Estimates of* $\mathrm{P}\left(\overline{N}_T \geq k\right)$ (b) *Estimates of* $\mathrm{P}\left(\overline{N}_T = k\right)$

Figure 1. *Estimated probabilities for* $T = 1$ *year implied by the default intensity model*

This figure shows the cIS and pMC estimates (markers) and 99% confidence intervals (dotted lines) for $\mathrm{P}(\overline{N}_T \geq k)$ in Panel (a), and for $\mathrm{P}(\overline{N}_T = k)$ in Panel (b). We set $T = 1$ year and draw $M = 5 \times 10^5$ samples, where confidence intervals extending below zero are omitted.

drawn from $[0, 0.2]$ uniformly. We require that the parameters of η^0 and η^i satisfy the Feller condition to ensure $x^i > 0$ holds under P almost surely for $i = 1, \ldots, n$. We set the initial value $\eta^i_0 = \theta_i$. For the systematic factors, we set $\theta_0 = 0.02$, $\kappa_0 = 1.0$, $\sigma_0 = 0.1$, and $\eta^0_0 = \theta_0$. The jump sensitivities are constructed by drawing each δ_{ij} from $[0, 1/n]$ uniformly. The selected parameters model a system with $\mathrm{P}(\overline{N}_T = 0) = 0.0281186615220496$.

We compare the cIS estimator with a *bias-free* version of the pMC estimator, known as the *jump approximation method* (JAM) estimator originally proposed by [GS20] so that both schemes generate unbiased estimators of $Y = \mathrm{P}\left(\overline{N}_T \geq k\right)$. As shown in Table 1, the exact pMC method fails to generate rare events and return $[0, 0]$ as the *empirical* confidence interval of $\mathrm{P}(\overline{N}_T \geq k)$ for $k \geq 23$ without causing any bias. However, the event $\{\overline{N}_T \geq k\}$ always occurs under the cIS simulation measure Q_k for any choice of k. As a result, a substantial variance reduction was achieved by the cIS scheme for rare event probability estimation, and the efficiency gain becomes more pronounced as the 'rarity parameter' k increases.



(a) *Relative errors for* $\mathrm{P}\left(\overline{N}_T \geq k\right)$ (b) *Relative errors for* $\mathrm{P}\left(\overline{N}_T = k\right)$

Figure 2. *Relative errors across different simulation schemes*
This figure shows the relative errors of pMC and cIS schemes for $\mathrm{P}\left(\overline{N}_T \geq k\right)$ in Panel (a) and for $\mathrm{P}\left(\overline{N}_T = k\right)$ in Panel (b). We set $T = 1$ year and draw $M = 5 \times 10^5$ samples.

Figure 1 shows cIS and pMC estimates of Y_k and their 99% confidence intervals for $M = 5 \times 10^5$ trials of cIS and exact pMC, respectively. The estimated probabilities of $Y_k = \mathrm{P}\left(\overline{N}_T \geq k\right)$ are shown in Panel (a), while Panel (b) illustrates the estimates of $Y_k = \mathrm{P}\left(\overline{N}_T = k\right)$ for $k \in \{5, 6, \ldots, 25\}$. Note that the full distribution of $\overline{N}_T$ can be estimated from the cIS scheme in the context of Corollary 3.2. In Figure 2, the performance of cIS is compared with that of pMC (JAM), where the simulation efficiency of each scheme is measured by its *relative error* defined as standard error divided by the true probability to be estimated. Our findings confirm that the cIS scheme can substantially reduce the variances of both $\mathrm{P}\left(\overline{N}_T \geq k\right)$ and $\mathrm{P}\left(\overline{N}_T = k\right)$ estimates for large k, compared with exact pMC for defaultable portfolios subject to non-trivial self-exciting contagion effects.

Risk managers and/or policymakers should consider possible failure of an abnormally large fraction of the total population in the system. The system default rate is denoted by $D_T \triangle \overline{N}_T/n$, where the distribution of $D_T \in [0, 1]$ represents the likelihood of failure by time $T > 0$ for any fraction of the population in the system; see [GK11] for similar failure-based measures of systemic risk. Quantile based tail-risk measures such as value-at-risk and expected shortfall are often used to measure and quantify the downside credit risk in the system. It is noteworthy that these tail-risk measures are crucial for the systemic risk analysis, but can be significantly underestimated by pMC. Applying the cIS scheme is beneficial, as it can precisely estimate the tail distribution of $\overline{N}_T$.

5.2. Zero-coupon bond pricing. An ample literature has emphasized the important role of jumps for interest rate dynamics. For instance, [Joh04] highlights the indispensability of jumps in continuous-time interest rate models both statistically and economically. However, most short-rate models with jump diffusion do not allow closed-form expressions for their integral transforms, which makes it difficult to apply the models to obtain unbiased pricing estimators of interest rate derivates analytically.

Let the short-rate process r_t follow the SDE under the pricing measure P given by

$$dr_t = \kappa\left(\theta - r_t\right)dt + \sigma\sqrt{r_t}dW_t + \delta_t d\pi_t, \tag{10}$$

where W_t is a standard Brownian motion, π_t is a pure jump process with intensity $\Lambda_t = \Lambda(r_t)$ and jump size $\delta_t = \delta(r_t)$. We assume that jump intensity π takes the form $\Lambda(r_t) = \lambda_0 + \lambda_1 r_t$ for $\lambda_0, \lambda_1 > 0$, and state-dependent jump sensitivities for π are specified as $\delta(r_t) = r_t\left(\exp\left(\eta_t\right) - 1\right)$, where $\eta_t \sim \text{Normal}\left(\left(\theta - r_t\right)u, v^2\right)$ for $u, v > 0$. Therefore, the log-normal jump components allow both positive and negative jumps, while keeping r_t nonnegative. A generalized version of the log-normal jump size model was introduced by [Joh04], and [GS13] investigate the case of positive jump sizes uniformly distributed between two strictly positive values. Sign-indefinite jumps provide additional sources for mean reversion; i.e. there is greater chance of positive jumps at lower interest rates than $\theta > 0$, and higher chance of negative jumps at high interest rates. As such, both positive and negative jumps occur, while r_t keeps its non-negativity and mean-reversion.

Estimating the zero-coupon bond price under the cIS scheme is a special case with $n = 1$. For consistent notation throughout the paper, we suppose an event indicator process N^1, which admits $p^1 = r\left(1 - N^1\right)$ as its P-intensity, where P is the risk-neutral pricing measure. Jump times for N^1 and $\overline{N}$ are identical P-almost surely for $n = 1$ by design, whereas jumps of π can occur prior to the first (and the last) jump time of $\overline{N}$. Thus, the price of a zero-coupon bond with unit face value and time-to-maturity $T > 0$ takes the form of the time-integrated transform of r given by

$$B_0(T) = \mathrm{E}^{\mathrm{P}}\left(e^{-\int_0^T r_t dt}\right) = \mathrm{P}\left(\tau_1 > T\right) = 1 - \mathrm{E}^{\mathrm{Q}_{k=1}}\left(\mathcal{L}_{k=1}\right),$$

where τ_1 is the first jump arrival time of $\overline{N}$ and $\mathrm{Q}_{k=1}$ is the cIS simulation measure conditional on the event $\{\overline{N}_T \geq 1\} = \{\overline{N}_T = 1\}$ with $n = 1$, taking $\mathcal{L}_{k=1}$ as the cIS estimator of $B_0(T)$. This implies that the measure $\mathrm{Q}_{k=1}$ in Theorem 3.1 is identical to the measure $\widehat{\mathrm{Q}}_{k=1}$ in Corollary 3.2. Hence, the zero-coupon bond price can be interpreted as the P-probability that no events of $\overline{N}$ occur during $[0, T]$ in a doubly stochastic counting process with its P-intensity $r\,(1 - \overline{N}) \geq 0$.

The joint process $(r, \overline{N})$ is self-affecting owing to the conflated dependence structure between the state of r, the jump intensity $\Lambda(r)$, and the jump size $\delta(r)$. This implies that jump times of r cannot be generated independently while drawing a sample path of r. However, significant computational advantage can be attained by taking a sequence of measure changes under the state-dependent specification of P-intensities for the jump term π. That is, the jump-counting process π and the event-counting process $\overline{N}$ satisfy a doubly stochastic property in that they are conditionally independent given the state of r. Jump arrivals of π can be sampled from the *exact* JAM method as originally proposed by [GS20]; refer to Appendix B.2 for details.

[BN12] propose a multinomial tree model to obtain approximate bond prices. Their proposed methodology superimposes mixed jump-diffusion trees by recombining multinomial jump trees on the diffusion tree for the stochastic short-rate model extended with various jump types. Although the multinomial tree does not generate simulation errors, its discretized nature and linear approximation should produce non-negligible biases; see [BN12] for details.

We adopt the *trapezoidal discretization* scheme as the benchmark pMC method by dividing time interval $[0, T]$ into $\lceil \sqrt{M} \rceil$ equal time steps. Specifically, we set the number of discretization time steps equal to the square-root of the number of simulation trials, as suggested by [DG95]. Assuming that at most one jump can occur at each discretized time point, we sequentially generated $r_j \triangleq r_{t_j}$ conditional on r_{j-1} as $r_j \approx r_{j-} + \delta_j \Delta\pi_j$, where r_{j-} can be drawn from the noncentral chi-squared transition density from r_{j-1} from equation (9). Furthermore, we approximated the probability of jump arrival at time t_j as $\mathrm{P}\left(\Delta\pi_j = 1\right) \approx (\lambda_0 + \lambda_1 r_{j-})h$ for small $h > 0$. We finally obtained pMC estimators of the zero-coupon bond price using the trapezoidal rule $\frac{1}{M}\sum_{i=1}^{M} \exp\left(-h\sum_{j=1}^{\lceil\sqrt{M}\rceil} \frac{r_{j-}+r_{j-1}}{2}\right)$. The pMC estimator bias for a given number of time discretization steps was estimated using 10^9 trials to establish estimator expectations, and then taking the difference from the true value, which was estimated using the proposed cIS algorithm with 10^{10} iterations.

Table 2 and Figure 3 show numerical results on zero-coupon bond pricing for cIS, multinomial tree, and trapezoidal discretization schemes. As shown, RMSE conver-

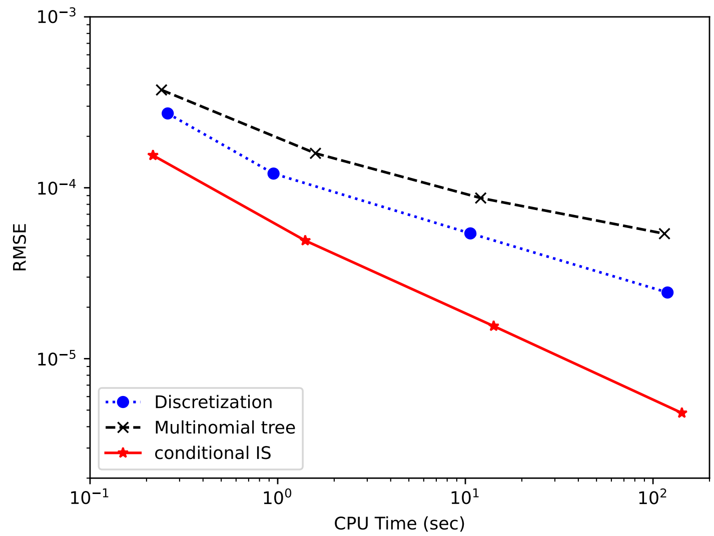


Figure 3. *Convergence of RMSE for estimated bond prices*
Root mean squared error (RMSE) convergence for bond prices under the short-rate model in (10). The selected parameters are $(\kappa, \theta, \sigma, \lambda_0, \lambda_1, u, v, r_0, T) = (0.1, 0.08, 0.05, 1.0, 5.0, 1.0, 1.0, 0.05, 3.0)$.

gence rate for the proposed cIS scheme is substantially faster the other schemes.

In addition, the *ultra short tenor* yield of zero-coupon bonds is investigated as the yield is given by $Y_0(T) = -\frac{\log B_0(T)}{T}$ by taking $T \downarrow 0$ to emphasize the usefulness of the proposed cIS scheme. Figure 4 shows that the cIS scheme is superior to the bias-free pMC method for estimating ultra-short tenor spot-rate curves under high frequency trading settlements. This finding suggests that the proposed cIS algorithm can facilitate immediate settlement of transactions on blockchain by employing an interbank money market model for ultra-short tenor interest rates that should be estimated at intraday level.

5.3. Overflow probabilities of time-varying queues. Consider a sequence of $M_t/M_t/1$ queues with infinite waiting spaces, where the sequence $\left\{C_t^{(n)} : n = 1, 2, \ldots\right\}$ denotes the number of customers present in the n-th queuing system at time $t \geq 0$. In the n-th single-server queue, we take a *top-down* formulation by presuming an ordered sequence of arrival times, denoted by $\left\{U_k^{(n)} : k = 1, 2, \ldots\right\}$. Similarly, we assume a strictly increasing sequence of *potential* service times denoted by $\left\{V_k^{(n)} : k = 1, 2, \ldots\right\}$, where $U_k^{(n)}$ and $V_k^{(n)}$ are stopping times on $(\Omega, \mathcal{F}, \mathrm{P})$. Note that some of the *potential* service times cannot be accepted in the set of the *actual* service times, denoted by

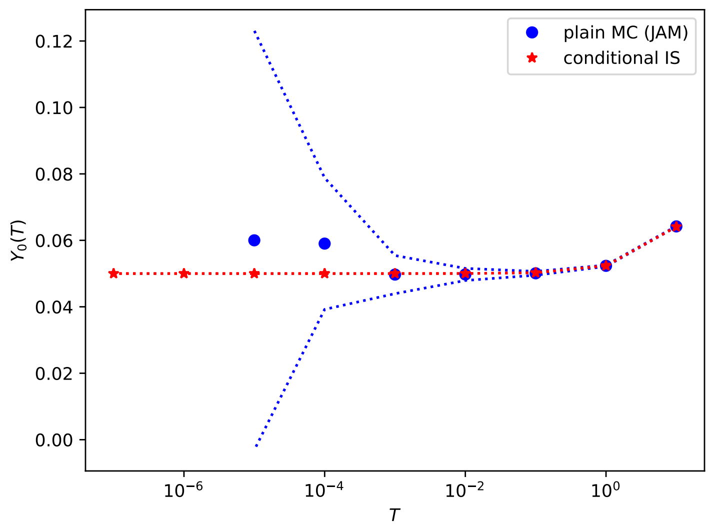


Figure 4. *Estimated ultra short tenor yield of zero-coupon bonds*
The cIS and pMC estimates (markers) and 99% confidence intervals (dashed lines) for annualized yields with respect to T with $M = 10^7$ under the short-rate model in (10). The selected parameters are $(\kappa, \theta, \sigma, \lambda_0, \lambda_1, u, v, r_0) = (0.1, 0.08, 0.05, 1.0, 5.0, 1.0, 1.0, 0.05)$.

$\left\{R_k^{(n)} : k = 1, 2, \ldots\right\}$, if the queue is empty and idle at that time.

We define the counting processes for arrivals, *potential* services and *actual* services as

$$A_t^{(n)} = \sum_{k \geq 1} 1_{\left\{U_k^{(n)} \leq t\right\}}, \quad D_t^{(n)} = \sum_{k \geq 1} 1_{\left\{V_k^{(n)} \leq t\right\}}, \quad S_t^{(n)} = \sum_{k \geq 1} 1_{\left\{R_k^{(n)} \leq t\right\}}.$$

This implies that we have

$$C_t^{(n)} = C_0^{(n)} + A_t^{(n)} - D_t^{(n)} + G_t^{(n)},$$

where $G_t^{(n)} \triangleq D_t^{(n)} - S_t^{(n)} \geq 0$. We posit that $A_t^{(n)}$ and $D_t^{(n)}$ are independent time-inhomogeneous Poisson processes with non-trivial arrival rate $\lambda_t^{(n)} > 0$ and *potential* service rate $\mu_t^{(n)} > 0$, respectively. This means that $A_t^{(n)} - \int_0^t \lambda_t^{(n)}$ and $D_t^{(n)} - \int_0^t \mu_t^{(n)}$ are (local) martingales for every $t \geq 0$. We assume that $\lambda_t^{(n)}$ and $\mu_t^{(n)}$ are constructed from the *base rate functions* with periodic patterns given by

$$\lambda_t = \overline{\lambda} + \alpha \sin(2\pi t/\tau) \quad \text{and} \quad \mu_t = \overline{\mu} + \beta \sin((2\pi(t - \nu)/\tau).$$

Specifically, the arrival rate in the n-th queuing system is specified as a periodic process $\lambda_t^{(n)} \triangleq n\lambda_t$, where customers are served first-in-first-out upon arrival. In turn, the *potential* service rate is also time-varying as $\mu_t^{(n)} \triangleq n\mu_t$, where the *actual* service times of the n-th queue are exponentially distributed with the rate of $\nu_t^{(n)} = \mu_t^{(n)} 1_{\left\{X_t^{(n)} > 0\right\}}$. It follows that we have $\nu_t^{(n)} = \mu_t^{(n)}$ unless the server is idle; see [XMS15] and [Rid20] for similar settings. Note that the stability condition of the queue is given by $\overline{\lambda} < \overline{\mu}$.

Our objective is to obtain an efficient estimation of the sequence of overflow probabilities

$$Y_n \triangleq \mathrm{P}\left(C_T^{(n)} \geq \lceil \xi n \rceil \,\middle|\, C_0^{(n)} = \lfloor cn \rfloor\right), \quad n = 1, 2, \ldots$$

for a finite overflow horizon $T > 0$; i.e., our interest lies in a rare-event sequence such that $Y_n \downarrow 0$ as $n \to \infty$ for some $\xi \gg c \geq 0$. In our numerical study, we estimate Y_n for $n = 1, \ldots, 25$ based on the three different schemes: (i) Plain A/R scheme, (ii) A/R with rate exchange, and (iii) cIS with tilting scheme.

Due to the periodic nature of arrival and service rates, the acceptance/rejection (A/R) scheme proposed by [LS79] is typically applicable, where its extended version is applied in the default clustering simulation by [GKZ11]. [PW89] introduced a simple method of importance sampling by exchanging the arrival rate with the service rate of the bottleneck queue for estimating the overflow probability in tandem queues. This approach was extended by [FLA91] to overflows of the population in any Jackson network. See Appendix B.3 for details of the implementation of the 'plain A/R' and 'A/R with rate exchange' schemes.

To apply the 'cIS with tilting' scheme, we draw samples of $A_T^{(n)}$ and $D_T^{(n)}$ under the *tilting* measures for Poisson distributions and then adjust the bias for efficiency; see [Sie76], [SB90], [Buc04] and references therein. Specifically, we introduce a tilting parameter θ associated with the selected tilting measure P_θ for $A_T^{(n)} \sim \text{Poi}\left(n\overline{\lambda}e^{\theta}T\right)$ and $D_T^{(n)} \sim \text{Poi}\left(n\overline{\mu}e^{-\theta}T\right)$, where $A_T^{(n)}$ and $D_T^{(n)}$ are still independent under P_θ; see Appendix B.3 for the optimal choice of θ along with the details of algorithmic description. It follows that we have $Y_n = \mathrm{P}(\mathcal{H}_n) = \mathrm{E}^{\mathrm{P}}\left(1_{\mathcal{H}_n}\right) = \mathrm{E}^{\mathrm{P}_\theta}\left(1_{\mathcal{H}_n} Z_\theta\right)$, where Z_θ is the Radon-Nikodym derivative of P with respect to P_θ.

We then apply the proposed cIS methodology along with the optimal tilting scheme. Adopting Corollary 1 in the context of the mutual independence between $A_T^{(n)}$ and $D_T^{(n)}$, we define the Radon-Nikodym derivative

$$\mathcal{L}_{(\alpha,\beta)} = \exp\left(-\int_0^T \left(\lambda_t^{(n)} + \mu_t^{(n)}\right) dt\right) \prod_{i=1}^{\alpha} \lambda^{(n)}_{U_i^{(n)}} \frac{T^\alpha}{\alpha!} \prod_{j=1}^{\beta} \mu^{(n)}_{V_j^{(n)}} \frac{T^\beta}{\beta!}$$

and introduce the cIS simulation measure $\mathrm{Q}_{(\alpha,\beta)}$ specific to $\mathcal{E}_n(A;\alpha) \cap \mathcal{E}_n(D;\beta)$ for $\mathcal{E}_n(A;\alpha) \triangleq \left\{A_T^{(n)} = \alpha\right\}$ and $\mathcal{E}_n(D;\beta) \triangleq \left\{D_T^{(n)} = \beta\right\}$. By doing so, we rewrite the overflow probability as $Y_n = \mathrm{E}^{\mathrm{P}_\theta}\left(\mathrm{E}^{\mathrm{Q}^\theta_{(A,D)}}\left(1_{\mathcal{H}_n}\mathcal{L}^\theta_{(A,D)}\right)\right)$ with the notational abuse of $A = A_T^{(n)}$ and $D = D_T^{(n)}$, where $\mathcal{L}^\theta_{(A,D)} \triangleq \mathcal{L}_{(A,D)} Z_\theta$ is given by

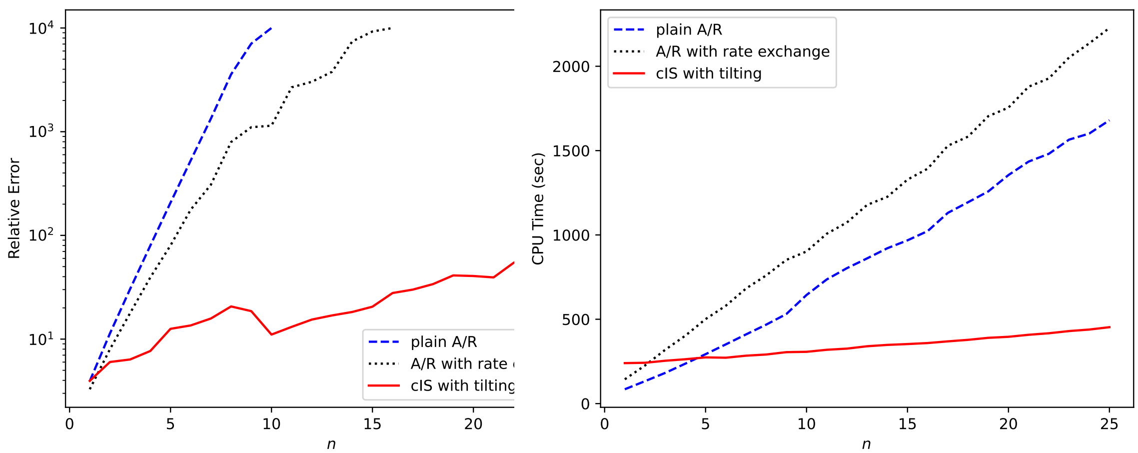
$$\mathcal{L}^\theta_{(A,D)} = \exp\left(-n\int_0^T \left(\lambda_t - \overline{\lambda}e^\theta + \mu_t - \overline{\mu}e^{-\theta}\right) dt\right) \prod_{i=1}^{A} \frac{\lambda_{U_i^{(n)}}}{\overline{\lambda}e^\theta} \prod_{j=1}^{D} \frac{\mu_{V_j^{(n)}}}{\overline{\mu}e^{-\theta}} . \tag{11}$$

Note from Corollary 1 that, under $\mathrm{Q}^\theta_{(A,D)}$, both arrival times and *potential* service times are independent and uniformly distributed on $[0, T]$. This facilitates the evaluation of $\mathcal{L}^\theta_{(A,D)} 1_{\mathcal{H}_n}$ under the cIS simulation measure $\mathrm{Q}^\theta_{(A,D)}$.

For our numerical analysis, the parameter set is chosen as

$$(\overline{\lambda}, \overline{\mu}, \alpha, \beta, \tau, \nu, T, \xi, c) = (0.5, 0.7, 0.25, 0.2, 0.25, 0.05, 1.0, 2.0, 0.1)$$

to compute the estimated overflow probabilities of a sequence of time-varying queues for $n = 1, 2, \ldots, 25$. We compare the performance and efficiency of the three different schemes by drawing $M = 10^8$ samples as shown in Figure 5. Panel (a) shows that



(a) *Relative errors*

(b) *Time elapsed (in seconds)*

Figure 5. *Simulation efficiency measures across different simulation schemes*
This figure shows the estimated relative errors in Panel (a) and the the CPU time (in seconds) to generate the samples in Panel (b). The parameter set is chosen as $(\bar{\lambda}, \bar{\mu}, \alpha, \beta, \tau, \nu, T, \xi, c) = (0.5, 0.7, 0.25, 0.2, 0.25, 0.05, 1.0, 2.0, 0.1)$.

the proposed cIS scheme is computationally more efficient than the other schemes in the estimation of the overflow probabilities as the 'rarity parameter' n increases. Specifically, the 'plain A/R' and 'A/R with rate exchange' schemes fail to achieve an accurate estimation of Y_n for large n, as no overflow events were observed from $M = 10^8$ samples based on the plain A/R scheme for $n \geq 11$ and the A/R with rate exchange scheme for $n \geq 17$. Panel (b) illustrates the required CPU running times in seconds in order to generate the samples. It confirms the substantial outperformance of the proposed cIS method by showing that the 'cIS with tilting' scheme requires the smallest computational budget to obtain those estimates of the overflow probabilities for $n \geq 5$.

The overflow example of dynamic queuing systems can be extended to numerous applications. Such examples include the consequences of being unable to meet the demand for intensive care service of critical patients in the context of healthcare applications, catastrophic failures in heavy loaded transportation systems, dropping calls in an emergency call center at a given time, and losses of information in a busy telecommunication system.

6. Conclusion

This paper proposes a simple and efficient importance sampling method for a wide range of counting processes. The proposed algorithm facilitates conditional Monte Carlo simulation based on the limit of conditional probability measures specific to the event of interest. Numerical results demonstrate superior performance of the proposed method to generate efficient simulation estimators of small probabilities of clustered defaults in a network, exact fixed-income security prices under jump-diffusion interest rate models, and overflow probabilities of a sequence of dynamic queuing systems.

A. Proofs

A.1. Proof of Theorem 3.1.

(i) Following the convention $1/0 = \infty$, we define for $j \in \mathbb{N}$

$$\mathcal{Z}_\infty^j = \exp\left(\int_0^{\tau_n} (\overline{p}_s - j^{-1})ds\right)\prod_{i=1}^{n}(j\,\overline{p}_{\tau_i-})^{-1} \tag{12}$$

and $\mathcal{Z}_t^j = \mathrm{E}^{\mathrm{P}}\left(\mathcal{Z}_\infty^j \,\middle|\, \mathcal{F}_t\right)$ $\forall t \geq 0$. This implies that we have

$$\mathcal{Z}_t^j = \prod_{i=1}^{n} \exp\left(\int_0^{t\wedge T_i} p_s^i(1-u_s^i)ds + N_{t\wedge T_i}^i \log(u_{T_i-}^i)\right),$$

where we take $u_t^i = (j\,\overline{p}_t)^{-1}$ for all $t \geq 0$ and $i \in \{1, 2, \ldots, n\}$ as a nonnegative and càdlàg process. (Recall that $\{\tau_i\}_{i=1}^\infty$ is the order statistics of $\{T_i\}_{i=1}^\infty$.) Accordingly, our assumption $\tau_n < \infty$ P-almost surely implies that $\mathcal{Z}_t^j$ is a (uniformly integrable) Doob martingale with $\mathrm{E}^{\mathrm{P}}(\mathcal{Z}_\infty^j) = 1$; see Theorem 3.1 in [GS20]. Then, we define $\widetilde{\mathrm{Q}}^j \triangleq \mathcal{Z}_\infty^j \mathrm{P}$ as an absolutely continuous probability measure in the sense that $\widetilde{\mathrm{Q}}^j(\mathscr{A}) = \mathrm{E}^{\mathrm{P}}(\mathcal{Z}_\infty^j 1_{\mathscr{A}})$ holds for all $\mathscr{A} \in \mathcal{F}$.

Since $\mathcal{E}_k = \left\{\overline{N}_T \geq k\right\} \in \mathcal{G}_k \triangleq \mathcal{F}_{\tau_k}$, we have

$$\begin{aligned}
\mathrm{E}^{\mathrm{P}}\left(\mathcal{Z}_{\tau_k}^j 1_{\mathcal{E}_k \bigcap \mathscr{A}}\right) &= \mathrm{E}^{\mathrm{P}}\left(\mathrm{E}^{\mathrm{P}}\left(\mathcal{Z}_{\tau_k}^j 1_{\mathcal{E}_k \bigcap \mathscr{A}} \,\middle|\, \mathcal{G}_k\right)\right) \\
&= \mathrm{E}^{\mathrm{P}}\left(1_{\mathcal{E}_k \bigcap \mathscr{A}} \mathrm{E}^{\mathrm{P}}\left(\mathcal{Z}_\infty^j \,\middle|\, \mathcal{G}_k\right)\right) = \widetilde{\mathrm{Q}}^j(\mathcal{E}_k \bigcap \mathscr{A})
\end{aligned} \tag{13}$$

for all $\mathscr{A} \in \mathcal{G}_k$ from the martingale property of Z_t^j. Note that $\overline{N}$ is a $\widetilde{\mathrm{Q}}^j$-Poisson process with rate $1/j$ on $[0, T \wedge \tau_n)$, which is a direct consequence of the definition of $\mathcal{Z}_\infty^j$ in (12) and Theorem 3.1 in [GS20]. Thus, we have $\widetilde{\mathrm{Q}}^j(\mathcal{E}_k) = \frac{e^{-T/j}}{j^k}\left(\frac{T^k}{k!} + o(j)\right)$, where $o(j) \downarrow 0$ as $j \uparrow \infty$. Furthermore, it follows that

$$\mathcal{Z}_{\tau_k}^j 1_{\mathcal{E}_k \bigcap \mathscr{A}} = \frac{e^{-\tau_k/j}}{\mathcal{L}_k\, j^k} 1_{\mathcal{E}_k \bigcap \mathscr{A}} \frac{T^k}{k!}.$$

Hence, $j^k \widetilde{\mathrm{Q}}^j(\mathcal{E}_k)$ can be expressed as

$$\mathrm{E}^{\mathrm{P}}\left(\frac{1_{\mathcal{E}_k}}{\mathcal{L}_k} e^{-\tau_k/j} \frac{T^k}{k!}\right) = e^{-T/j}\left(\frac{T^k}{k!} + o(j)\right)$$

by taking $\mathscr{A} = \Omega$ in equation (13). Since $\frac{1_{\mathcal{E}_k}}{\mathcal{L}_k} e^{-\tau_k/j} \frac{T^k}{k!}$ is increasing as $j \uparrow \infty$ and nonnegative, by monotone convergence we have

$$\mathrm{Q}_k(\Omega) = \mathrm{E}^{\mathrm{P}}\left(\frac{1_{\mathcal{E}_k}}{\mathcal{L}_k}\right) = 1,$$

which proves that $\mathrm{Q}_k \ll \mathrm{P}$ is a well-defined probability measure on $(\Omega, \mathcal{G}_k)$.

(ii) Notice that P is absolutely continuous with respect to Q_k on $\mathcal{E}_k$ by construction. We subsequently have

$$\mathrm{P}(\mathcal{E}_k) = \mathrm{E}^{\mathrm{P}}(1_{\mathcal{E}_k}) = \mathrm{E}^{\mathrm{P}}\left(\mathcal{L}_k \frac{1_{\mathcal{E}_k}}{\mathcal{L}_k}\right) = \mathrm{E}^{\mathrm{Q}_k}(\mathcal{L}_k),$$

as $\mathcal{L}_k > 0$ holds under P almost surely by our assumption.

(iii) We construct a conditional measure

$$\widetilde{\mathrm{Q}}_k^j \triangleq \frac{1_{\mathcal{E}_k}}{\widetilde{\mathrm{Q}}^j(\mathcal{E}_k)} Z_\infty^j \mathrm{P}$$

on $(\Omega, \mathcal{G}_k)$ so that $\widetilde{\mathrm{Q}}_k^j(\mathscr{A}) = \widetilde{\mathrm{Q}}^j(\mathscr{A} \,|\mathcal{E}_k)$ for all $\mathscr{A} \in \mathcal{G}_k$. By Fatou's lemma with $\mathcal{Z}_\infty^j \geq 0$, we have

$$\begin{aligned}
\liminf_{j\uparrow\infty} \widetilde{\mathrm{Q}}_k^j(\mathscr{A}) &= \liminf_{j\uparrow\infty} \frac{\mathrm{E}^{\mathrm{P}}\left(1_{\mathscr{A}\bigcap\mathcal{E}_k} \mathrm{E}^{\mathrm{P}}\left(\mathcal{Z}_\infty^j \middle| \mathcal{G}_k\right)\right)}{\widetilde{\mathrm{Q}}^j(\mathcal{E}_k)} \\
&= \liminf_{j\uparrow\infty} \frac{\mathrm{E}^{\mathrm{P}}\left(1_{\mathscr{A}\bigcap\mathcal{E}_k} \mathcal{Z}_{\tau_k}^j\right)}{\widetilde{\mathrm{Q}}^j(\mathcal{E}_k)} \\
&\geq \mathrm{E}^{\mathrm{P}}\left(\frac{1_{\mathscr{A}\bigcap\mathcal{E}_k}}{\mathcal{L}_k} \liminf_{j\uparrow\infty} \frac{e^{-\tau_k/j}\frac{T^k}{k!}}{e^{-T/j}\left(\frac{T^k}{k!} + o(j)\right)}\right) \\
&= \mathrm{E}^{\mathrm{P}}\left(\frac{1_{\mathscr{A}\bigcap\mathcal{E}_k}}{\mathcal{L}_k}\right) = \mathrm{Q}_k(\mathscr{A}).
\end{aligned}$$

Therefore, by the Portmanteau theorem, we have that $\widetilde{\mathrm{Q}}_k^j \Rightarrow \mathrm{Q}_k$ as $j \uparrow \infty$. We also have

$$\widetilde{\mathrm{Q}}_k^j = \frac{1_{\mathcal{E}_k}\widetilde{\mathrm{Q}}^j}{\widetilde{\mathrm{Q}}^j(\mathcal{E}_k)} = \frac{1_{\mathcal{E}_k}\left(\mathcal{Z}_{\tau_k}^j \big/ \mathcal{Z}_{\tau_k}^1\right)\widetilde{\mathrm{Q}}^1}{\widetilde{\mathrm{Q}}^j(\mathcal{E}_k)} = \frac{1_{\mathcal{E}_k} e^{(1-j^{-1})\tau_k}\widetilde{\mathrm{Q}}^1}{e^{-T/j}\left(\frac{T^k}{k!} + o(j)\right)}.$$

Now consider a Q_k-continuity set

$$A_{(t_1,\ldots,t_k)} = \{\tau_1 \leq t_1, \tau_2 \leq t_2, \ldots, \tau_k \leq t_k\}$$

for any $0 \leq t_1 \leq t_2 \leq \cdots \leq t_k \leq T$. As $\tau_k \leq T$ and $e^{\tau_k - T/j} \leq e^{(1-j^{-1}}\tau_k) \leq e^{\tau_k}$ must hold on $\mathcal{E}_k \subseteq A_{(t_1,\ldots,t_k)}$), we have

$$\frac{\mathrm{E}^{\widetilde{\mathrm{Q}}^1}\left(1_{A_{(t_1,\ldots,t_k)}} e^{\tau_k}\right)}{\frac{T^k}{k!} + o(j)} \leq \widetilde{\mathrm{Q}}_k^j\left(A_{(t_1,\ldots,t_k)}\right) \leq \frac{\mathrm{E}^{\widetilde{\mathrm{Q}}^1}\left(1_{A_{(t_1,\ldots,t_k)}} e^{\tau_k}\right)}{e^{-T/j}\left(\frac{T^k}{k!} + o(j)\right)},$$

where [Mey71]'s time-change theorem implies that

$$\mathrm{E}^{\widetilde{\mathrm{Q}}^1}\left(1_{A_{(t_1,\ldots,t_k)}}e^{\tau_k}\right)=\int_0^{t_1}\int_{x_1}^{t_2}\cdots\int_{x_{k-1}}^{t_k}dx_k\cdots dx_2dx_1=\frac{T^k}{k!}F(t_1,\ldots,t_k),$$

where $F(t_1,\ldots,t_k)$ is the distribution of the $\{u_i\}_{i=1}^k$. It follows that

$$\frac{\frac{T^k}{k!}F(t_1,\ldots,t_k)}{\frac{T^k}{k!}+o(j)}\leq\widetilde{\mathrm{Q}}_k^j\left(A_{(t_1,\ldots,t_k)}\right)\leq\frac{\frac{T^k}{k!}F(t_1,\ldots,t_k)}{e^{-T/j}\left(\frac{T^k}{k!}+o(j)\right)}.$$

Taking $j\uparrow\infty$, we conclude that $\widetilde{\mathrm{Q}}_k^j\left(A_{(t_1,\ldots,t_k)}\right)\to F(t_1,\ldots,t_k)$, which proves the uniform arrivals of the first k event times on $[0,T]$ under the limit measure Q_k.

(iv) Let $m\in\{1,\ldots,k\}$ and consider $I_m\in\{1,\ldots,n\}$. By the conditional change of measure formula, we have

$$\frac{\mathrm{E}^{\widetilde{\mathrm{Q}}^j}\left(1_{\mathcal{E}_k}\middle|\mathcal{H}_{\tau_k-}\right)}{\widetilde{\mathrm{Q}}^j(\mathcal{E}_k)}\widetilde{\mathrm{Q}}_k^j\left(I_m=i\middle|\mathcal{H}_{\tau_m-}\right)=\frac{\mathrm{E}^{\widetilde{\mathrm{Q}}^j}\left(1_{\mathcal{E}_k\bigcap\{I_m=i\}}\middle|\mathcal{H}_{\tau_m-}\right)}{\widetilde{\mathrm{Q}}^j(\mathcal{E}_k)}.$$

Since $1_{\mathcal{E}_k}=0$ if $\tau_m>T$ and τ_m is $\mathcal{H}_{\tau_m-}$-measurable, we obtain

$$\begin{aligned}\widetilde{\mathrm{Q}}_k^j\left(I_m=i\middle|\mathcal{H}_{\tau_m-}\right)&=1_{\{\tau_m\leq T\}}\frac{\widetilde{\mathrm{Q}}^j\left(\mathcal{E}_k\bigcap\{I_m=i\}\middle|\mathcal{H}_{\tau_m-}\right)}{\widetilde{\mathrm{Q}}^j\left(\mathcal{E}_k\middle|\mathcal{H}_{\tau_m-}\right)}\\&=1_{\{\tau_m\leq T\}}\widetilde{\mathrm{Q}}^j\left(I_m=i\middle|\mathcal{H}_{\tau_m-}\right)\\&=1_{\{\tau_m\leq T\}}\frac{p^i_{\tau_m-}u_{\tau_m-}}{\overline{p}_{\tau_m-}u_{\tau_m-}}=1_{\{\tau_m\leq T\}}\frac{p^i_{\tau_m-}}{\overline{p}_{\tau_m-}}.\end{aligned}$$

Since $\widetilde{\mathrm{Q}}_k^j(\tau_m\leq T)=1$ holds for all $j\in\{1,\ldots,k\}$, we conclude that

$$\mathrm{E}^{\widetilde{\mathrm{Q}}_k^j}\left(1_{\mathscr{A}}1_{\{I_m=i\}}\right)=\mathrm{E}^{\widetilde{\mathrm{Q}}_k^j}\left(1_{\mathscr{A}}\frac{p^i_{\tau_m-}}{\overline{p}_{\tau_m-}}\right)$$

for all $\mathscr{A}\in\mathcal{H}_{\tau_m-}$. Taking $j\uparrow\infty$ by the Portmanteau theorem we obtain

$$\mathrm{Q}_k\left(I_m=i\middle|\mathcal{H}_{\tau_m-}\right)=\frac{p^i_{\tau_m-}}{\overline{p}_{\tau_m-}}=\mathrm{P}\left(I_m=i\middle|\mathcal{F}_{\tau_m-}\right).$$

(v) This is a direct consequence of the Girsanov-Meyer theorem [Pro05, Theorem III.41].

A.2. Proof of Corollary 3.2. This follows directly from the arguments in the proof of Theorem 3.1 with some slight adjustments that lead to significant simplification. We redefine the sequence of conditional measures for the modified set $\widehat{\mathcal{E}}_k \triangleq \{\overline{N}_T = k\}$. Observe that on $\widehat{\mathcal{E}}_k$, we have $\mathcal{Z}_T^j / \widehat{\mathrm{Q}}_T^j(\widehat{\mathcal{E}}_T^k) = \frac{e^{T/j}}{\widehat{\mathcal{L}}_T^k}$. We follow the same arguments that take $j \uparrow \infty$ as in the proof of Theorem 3.1. Statement (iii) is a direct consequence of the new definition of $\widehat{\mathcal{E}}_k$. Statements (iv) and (v) are proved based on the same methodology as Theorem 3.1.

A.3. Proof of Theorem 4.3. Under our hypothesis we have $y_n \to 0$ under $n \uparrow \infty$. Expanding (3), we obtain

$$\lim_{n\uparrow\infty} \frac{\mathrm{E}^{\mathrm{P}}(\mathcal{L}_n 1_{\{\overline{N}_T \geq \lceil \mu n \rceil\}}) - y_n^2}{y_n(1-y_n)} = \lim_{n\uparrow\infty} \frac{\mathrm{E}^{\mathrm{P}}(\mathcal{L}_n 1_{\{\overline{N}_T \geq \lceil \mu n \rceil\}})}{y_n}. \tag{14}$$

Note that we have

$$\mathrm{E}^{\mathrm{P}}(\mathcal{L}_n 1_{\{\overline{N}_T \geq \lceil \mu n \rceil\}}) = \mathrm{E}^{\mathrm{Q}_T^{\lceil \mu n \rceil}}((\mathcal{L}_n)^2) \geq y_n^2$$

by Jensen's inequality. Then, the definition of $\mathcal{L}_n$ yields

$$\mathrm{E}^{\mathrm{P}}(\mathcal{L}_n 1_{\{\overline{N}_T \geq \lceil \mu n \rceil\}}) = \frac{T^{\lceil \mu n \rceil}}{\lceil \mu n \rceil !} \mathrm{E}^{\mathrm{P}}\Big(\underbrace{\exp\Big(-\int_0^{\tau_{\lceil \mu n \rceil}} \overline{p}_s \, ds\Big)}_{:=A} \underbrace{1_{\{\overline{N}_T \geq \lceil \mu n \rceil\}} \prod_{k=1}^{\lceil \mu n \rceil} \overline{p}_{\tau_k -}}_{:=B} \Big).$$

Since the two random variables A and B under the expectation are negatively associated and that $\int_0^{\tau_{\lceil \mu n \rceil}} \overline{p}_s \, ds$ is equal in P-law to $\lceil \mu n \rceil$ standard i.i.d. exponential [Mey71,], we obtain

$$\mathrm{E}^{\mathrm{P}}(\mathcal{L}_n 1_{\{\overline{N}_T \geq \lceil \mu n \rceil\}}) < \frac{T^{\lceil \mu n \rceil}}{2^{\lceil \mu n \rceil} \lceil \mu n \rceil !} \mathrm{E}^{\mathrm{P}}\Big(1_{\{\overline{N}_T \geq \lceil \mu n \rceil\}} \prod_{k=1}^{\lceil \mu n \rceil} \overline{p}_{\tau_k -} \Big). \tag{15}$$

Notice that we have

$$\mathrm{E}^{\mathrm{P}}\left(\prod_{k=1}^{\lceil \mu n \rceil} \overline{p}_{\tau_k -} \,\middle|\, \overline{N}_T \geq \lceil \mu n \rceil \right) = \frac{\mathrm{E}^{\mathrm{P}}\Big(1_{\{\overline{N}_T \geq \lceil \mu n \rceil\}} \prod_{k=1}^{\lceil \mu n \rceil} \overline{p}_{\tau_k -} \Big)}{\mathrm{E}^{\mathrm{P}}\Big(1_{\{\overline{N}_T \geq \lceil \mu n \rceil\}} \Big)} = \frac{\mathrm{E}^{\mathrm{Q}_n}\Big(\mathcal{L}_n \prod_{k=1}^{\lceil \mu n \rceil} \overline{p}_{\tau_k -} \Big)}{\mathrm{E}^{\mathrm{Q}_n}(\mathcal{L}_n)},$$

which implies that

$$\mathrm{E}^{\mathrm{P}}\left(\prod_{k=1}^{\lceil \mu n \rceil} \overline{p}_{\tau_k -} \,\middle|\, \overline{N}_T \geq \lceil \mu n \rceil \right) = \frac{\mathrm{E}^{\mathrm{Q}_n}\Big(\prod_{k=1}^{\lceil \mu n \rceil} \overline{p}_{\tau_k -}^2 \exp\Big(-\int_{\tau_{k-1}}^{\tau_k} \overline{p}_s ds\Big)\Big)}{\mathrm{E}^{\mathrm{Q}_n}\Big(\prod_{k=1}^{\lceil \mu n \rceil} \overline{p}_{\tau_k -} \exp\Big(-\int_{\tau_{k-1}}^{\tau_k} \overline{p}_s ds\Big)\Big)}.$$

Applying our assumption in (2) with the inequality of arithmetic and geometric means yields

$$\mathrm{E}^{\mathrm{P}}\left(\prod_{k=1}^{\lceil\mu n\rceil}\overline{p}_{\tau_k-}\,\middle|\,\overline{N}_T \geq \lceil\mu n\rceil\right) < \left(\frac{2\mu n}{eT}\right)^{\lceil\mu n\rceil}. \tag{16}$$

Then, we combine (15) and (16) to obtain by iterated expectations

$$\mathrm{E}^{\mathrm{P}}(\mathcal{L}_n 1_{\{\overline{N}_T\geq\lceil\mu n\rceil\}}) < \frac{e^{-\lceil\mu n\rceil}\lceil\mu n\rceil^{\lceil\mu n\rceil}}{\lceil\mu n\rceil!}y_n, \tag{17}$$

which provides that the right side of (14) tends to zero for large n as required.

A.4. Proof of Corollary 4.6. The proof of Corollary 4.6 follows immediately from the light-tailed assumption given by (4) and the proof of Theorem 4.3. Specifically, the inequality of arithmetic and geometric means along with the condition (??) implies that

$$\mathrm{E}^{\mathrm{P}}\left(\prod_{k=1}^{\lceil\mu n\rceil}\overline{p}_{\tau_k-}\,\middle|\,\overline{N}_T \geq \lceil\mu n\rceil\right) < \left(\frac{2\mu n}{eT}e^{-\gamma/\mu}\right)^{\lceil\mu n\rceil}$$

holds. Similarly to (17), we obtain by iterated expectations

$$\mathrm{E}^{\mathrm{Q}_n}(\mathcal{L}_n^2) = \mathrm{E}^{\mathrm{P}}(\mathcal{L}_n 1_{\{\overline{N}_T\geq\lceil\mu n\rceil\}}) < \frac{e^{-\lceil\mu n\rceil}\lceil\mu n\rceil^{\lceil\mu n\rceil}}{\lceil\mu n\rceil!}e^{-\gamma n}y_n.$$

Note that the inequality of Stirling's formula guided by [Rob55] confirms that

$$\sqrt{2\pi\lceil\mu n\rceil}\,\lceil\mu n\rceil^{\lceil\mu n\rceil}e^{-\lceil\mu n\rceil}e^{\frac{1}{12\lceil\mu n\rceil+1}} < \lceil\mu n\rceil!$$

is valid for all $n \geq 1$. It follows that we have

$$\limsup_{n\uparrow\infty}\frac{1}{n}\log\mathrm{E}^{\mathrm{Q}_n}(\mathcal{L}_n^2) \leq -2\gamma. \tag{18}$$

By Jensen's inequality, the inequality

$$\lim_{n\uparrow\infty}\frac{\log\mathrm{E}^{\mathrm{Q}_n}(\mathcal{L}_n^2)}{\log\mathrm{E}^{\mathrm{Q}_n}(\mathcal{L}_n)} \leq 2 \tag{19}$$

always holds. Combining (18) and (19) completes the proof, as we have $y_n = \mathrm{P}\left(\overline{N}_T \geq \lceil\mu n\rceil\right) = \mathrm{E}^{\mathrm{Q}_n}(\mathcal{L}_n)$ with our assumption (4).

B. Computational Details

This appendix describes computational details and the concrete algorithms of the numerical examples in Section 5. The programming code is written in Julia and available upon request.

B.1. Networked default clustering. For the notational simplicity, we abuse the notation for the interarrival intensity process h^{ij} in equation (8) as h by assuming that h satisfies the following SDE given by

$$dh_t = \kappa(\theta - h_t)\,dt + \sigma\sqrt{h_t}dB_t$$

for some $\kappa > 0, \theta > 0, \sigma > 0$ and $h_0 > 0$, where B is a Brownian motion. We consider a transform taking the form of

$$\begin{aligned}\varphi_t(v;u) &\triangleq \mathrm{E}\Big(\exp(vh_t - u\int_0^t h_s ds)\big|h_0\Big)\\ &= \exp\big(a_t(v;u) + b_t(v;u)h_0\big).\end{aligned}$$

Then, the functions $a_t(v;u)$ and $b_t(v;u)$ can be expressed as

$$\begin{aligned}a_t(v;u) &= \frac{\kappa\theta}{\sigma^2}\left(\kappa t - 2\log\left(\cosh(0.5\gamma_u t) + \frac{\kappa - \sigma^2 v}{\gamma_u}\sinh(0.5\gamma_u t)\right)\right)\\ b_t(v;u) &= \frac{1}{\sigma^2}\left(\kappa - \gamma_u \tanh\left(0.5\gamma_u t + \operatorname{atanh}\left(\frac{\kappa - \sigma^2 v}{\gamma_u}\right)\right)\right),\end{aligned}$$

where $\gamma_u = \sqrt{\kappa^2 + 2u\sigma^2}$. Taking the derivative with respect to v and applying $v = 0$ will yield

$$\begin{aligned}a_t'(0;u) &= \frac{2\kappa\theta\sinh(0.5\gamma_u t)}{\gamma_u\cosh(0.5\gamma_u t) + \kappa\sinh(0.5\gamma_u t)}\\ a_t''(0;u) &= \frac{\sigma^2}{2\kappa\theta}\big(a_t'(0;u)\big)^2\end{aligned}$$

and

$$\begin{aligned}b_t'(0;u) &= \frac{\gamma_u^2\Big(\operatorname{sech}(0.5\gamma_u t + \operatorname{atanh}(\frac{\kappa}{\gamma_u}))\Big)^2}{2u\sigma^2}\\ b_t''(0;u) &= \frac{b_t'(0;u)}{u}\left(\gamma_u \tanh\big(0.5\gamma_u t + \operatorname{atanh}(\frac{\kappa}{\gamma_u})\big) - \kappa\right).\end{aligned}$$

Taking the derivative with respect to v with $(v,u) = (0,1)$ yields

$$\begin{aligned}\varphi_t'(0;1) &= \mathrm{E}\left(h_t\exp\left(-\int_0^t h_s ds\right)\Big|h_0\right)\\ &= \varphi_t(0;1)(a_t'(0;1) + b_t'(0;1)h_0).\end{aligned}$$

Finally, taking the derivative with respect to v twice with $(v,u) = (0,1)$ yields

$$\varphi_t''(0;1) = \mathrm{E}\left(h_t^2\exp\left(-\int_0^t h_s ds\right)\Big|h_0\right)$$

$$= \varphi_t'(0;1)(a_t'(0;1) + b_t'(0;1)h_0) + \varphi_t(0;1)(a_t''(0;1) + b_t''(0;1)h_0)$$
$$= \varphi_t(0;1)\Big((a_t'(0;1) + b_t'(0;1)h_0)^2 + (a_t''(0;1) + b_t''(0;1)h_0)\Big).$$

We also require the bridge transform in the form of

$$\Psi_t \triangleq \mathrm{E}\left(\exp\left(-\int_0^t h_s ds\right)\middle| h_0, h_t\right)$$

for some $t > 0$. The closed-form solution for Ψ_t conditional on h_0 and h_t can be obtained by applying equation (13) in [BK06], which is given by

$$\Psi_t = \frac{\gamma_1 e^{-0.5(\gamma_1-\kappa)t}(1-e^{-\kappa t})}{\kappa(1-e^{-\gamma_1 t})} \cdot \exp\left(\frac{h_0+h_t}{\sigma^2}\left(\frac{\kappa(1+e^{-\kappa t})}{1-e^{-\kappa t}} - \frac{\gamma_1(1+e^{-\gamma_1 t})}{1-e^{-\gamma_1 t}}\right)\right) \cdot A,$$

where A can be evaluated with $d = \frac{2\kappa\theta}{\sigma^2} - 1$ as

$$A = \frac{I_d\left(\sqrt{h_0 h_t}\frac{4\gamma_1 e^{-0.5\gamma_1 t}}{\sigma^2(1-e^{-\gamma_1 t})}\right)}{I_d\left(\sqrt{h_0 h_t}\frac{4\kappa e^{-0.5\kappa t}}{\sigma^2(1-e^{-\kappa t})}\right)}.$$

Here, I_ν denotes the modified Bessel function of the first kind of order ν, and a trapezoidal approximation is applied when t is too small to evaluate the modified Bessel function numerically. For a fixed k, Algorithm 2 summarizes the proposed cIS algorithm to estimate $\mathrm{P}(\overline{N}_T \geq k)$ with $T > 0$.

B.2. Zero-coupon bond pricing. This appendix illustrates the technical details of simulating jump diffusion in Section 5.2. The basic concept is to construct a simulation measure such that jump arrival intensity is a pure jump process with jumps occurring at jump arrival times of π. Specifically, let

$$0 = \mathcal{S}_0 < \mathcal{S}_1 < \mathcal{S}_2 < \dots$$

be the ordered jump arrival times of π. We constructed a jump-simulation measure, Q^J, such that jump arrival times are sampled based on the twisted jump arrival intensity whose paths are piecewise constant between two consecutive jumps; i.e., Q^J is constructed with twisted jump intensities $\psi_k = \Lambda(r_{\mathcal{S}_k})$ for $k \geq 0$. The Radon-Nikodym derivative of Q^J with respect to P is given by $\mathcal{Z}_\infty^J$, which takes the form of

$$\mathcal{Z}_\infty^J = \exp\left(\int_0^\infty \Big(\Lambda(r_s) - \psi_{\pi_s}\Big)ds\right)\prod_{i=1}^\infty \frac{\psi_{i-1}}{\Lambda(r_{\mathcal{S}_{i-}})}. \tag{20}$$

Algorithm 2 The cIS algorithm to estimate $\mathrm{P}(\overline{N}_T \geq k)$ with $T > 0$ under the default intensity model

procedure CIS_TAILPROBABILITY(k, T, M) ▷ M is the number of cIS trials
  Initialize $\hat{Y} \leftarrow 0$
  for $m \in \{1, \ldots, M\}$ **do**
    Initialize $\mathcal{K} \leftarrow \{1, \ldots, n\}$ and $X_0^i \leftarrow \omega^i \eta_0^0 + \eta_0^i \;\; \forall i \in \mathcal{K}$
    Set $t \leftarrow 0$ and $\mathcal{L}_k \leftarrow T^k / k!$
    Draw an ordered sample of $\{\tau_1, \ldots, \tau_k\}$ from $U(0, T)$
    for $j \in \{1, \ldots, k\}$ **do**
      Sample $\eta_{\tau_j}^0$ given η_t^0 from the transition density of h^{0j}
      Update $\mathcal{L}_k \leftarrow \mathcal{L}_k \times \Psi^{0j}\left(\sum_{i \in \mathcal{K}} \omega^i,\ \tau_j - t;\ h_0^{0j},\ h_{\tau - t}^{0j} \right)$
      for $i \in \mathcal{K}$ **do**
        Sample $\eta_{\tau_j}^i$ given η_t^i from the transition density of h^{ij}
        Update $\mathcal{L}_k \leftarrow \mathcal{L}_k \times \Psi^{ij}\left(1,\ \tau_j - t;\ h_0^{ij},\ h_{\tau_j - t}^{ij}\right)$
        Set $X_{\tau_j}^i \leftarrow \omega^i \eta_{\tau_j}^0 + \eta_{\tau_j}^i$
      end for
      Update $\mathcal{L}_k \leftarrow \mathcal{L}_k \times \sum_{i \in \mathcal{K}} X_{\tau_j}^i$
      Draw index $I_j = m$ from the distribution $\left\{ X_{\tau_j}^m / \sum_{i \in \mathcal{K}} X_{\tau_j}^i \right\}_{m \in \mathcal{K}}$ and set $t \leftarrow \tau_j$
      Update $\mathcal{K} \leftarrow \mathcal{K} - \{m\}$ and $\eta_t^i \leftarrow \eta_t^i + \delta_{im} \;\forall i \in \mathcal{K}$
    end for
    Update $\hat{Y} \leftarrow \hat{Y} + \mathcal{L}_k$
  end for
  return $\hat{Y} / M$
end procedure

The Doob martingale of $\mathcal{Z}^J_\infty$ is given by $\mathcal{Z}^J_{\tau_1} = \mathrm{E}^{\mathrm{P}}\left(\mathcal{Z}^J_\infty \middle| \mathcal{F}_{\tau_1}\right) = \left(\mathcal{L}^J_{\tau_1}\right)^{-1}$ for $\tau_1 > 0$. In this sense, the expression of $\mathcal{L}^J_{\tau_1}$ is given by

$$\mathcal{L}^J_{\tau_1} = \exp\left(\int_0^{\tau_1}\left(-\Lambda(r_s) + \psi_{\pi_s}\right) ds\right) \prod_{i=1}^{\pi_{\tau_1}} \frac{\Lambda(r_{\mathcal{S}_{i-}})}{\psi_{i-1}}. \tag{21}$$

Therefore, we obtain $B_0(T) = 1 - \mathrm{E}^{\mathrm{Q}_{k=1}}\left(\mathrm{E}^{\mathrm{Q}^J}(\mathcal{L}_\star)\right)$, where $\mathcal{L}_\star = \mathcal{L}_{k=1}\,\mathcal{L}^J_{\tau_1}$.

We then introduce the *interjump* short-rate process ρ^j satisfying

$$r_t = \sum_{j\ge 1} \rho^j_{t-\mathcal{S}_{j-1}} 1_{\{\pi_t = j-1\}},$$

where $\rho^j_0 = r_{\mathcal{S}_{j-1}}$ for $j \ge 1$. As the process ρ^j follows the Feller diffusion without jump arrivals, the exact pMC scheme is applicable. Furthermore, we can avoid sampling the time-integrated transform of the jump intensity $\Lambda(\rho^j_t) = \lambda_0 + \lambda_1 \rho^j_t$ from its Markov property by adopting a closed-form expression for the bridge transform of the interjump short-rate process ρ^j_t. This ensures unbiased estimators with smaller simulation errors. Algorithm 3 summarizes the cIS algorithm to estimate zero-coupon bond prices with maturity $T > 0$.

B.3. Overflow probabilities of time-varying queues. This appendix describes the technical details of the algorithms, which are applied in Section 5.3

(i) **(Plain A/R scheme)** By letting $\Delta_n = \lceil \xi n\rceil - \lfloor cn \rfloor$ and $\mathcal{H}_n = \left\{A^{(n)}_T - D^{(n)}_T + G^{(n)}_T \ge \Delta_n\right\}$, the evaluation of $1_{\mathcal{H}_n}$ is feasible conditional on the simulated $\left(U^{(n)}_1, \ldots, U^{(n)}_A\right)$ and $\left(V^{(n)}_1, \ldots, V^{(n)}_D\right)$ by eliminating *false* service times while the server is idle. In turn, the pMC estimator of $Y_n = \mathrm{E}^{\mathrm{P}}\left(1_{\mathcal{H}_n}\right)$ is given by

$$\hat{Y}_n \approx \frac{1}{M}\sum_{m=1}^{M} 1_{\mathcal{H}_n},$$

where M is the number of generated samples.

(ii) **(A/R with rate exchange)** For notational simplicity, we abuse some of the notations for fixed n and $T > 0$ as

$$A \triangleq A^{(n)}_T,\ \ D \triangleq D^{(n)}_T,\ \ U_k \triangleq U^{(n)}_k,\ \ V_k \triangleq V^{(n)}_k$$

for $k = 1, 2, \ldots$. We then introduce the Radon-Nikodym derivative for a given pair of (A, D) defined as

$$\mathcal{L}^{\leftrightarrow}_{(A,D)} = \prod_{i=1}^{A} \frac{\lambda_{U_i}}{\mu_{U_i}} \prod_{j=1}^{D} \frac{\mu_{V_j}}{\lambda_{V_j}}$$

Algorithm 3 The cIS algorithm to estimate a zero-coupon bond price with $T > 0$

procedure CIS_BONDPRICING(T, M) $\triangleright$ M is the number of cIS trials
  Initialize $\hat{Y} \leftarrow 0$
  for $m \in \{1, \ldots, M\}$ **do**
    Set $\mathcal{L}_\star \leftarrow T,\ r_t \leftarrow r_0$ and draw τ from $U(0, T)$
    while $\tau > 0$ **do**
      Draw Δ from the exponential distribution with rate $\Lambda(r_t) = \lambda_0 + \lambda_1 r_t$
      if $\Delta \geq \tau$ **then**
        Sample $r_{t+\tau}$ given r_t from the transition density of r without jump arrivals
        Compute $A = e^{-\lambda_0 \tau}\Psi(1 + \lambda_1,\ \tau;\ r_t,\ r_{t+\tau})$
        Update $\mathcal{L}_\star \leftarrow \mathcal{L}_\star \times r_{t+\tau} \times A e^{\Lambda(r_t)\tau}$ and $\tau \leftarrow 0$
      else
        Sample $r_{t+\Delta}$ given r_t from the transition density of r without jump arrivals
        Compute $A = e^{-\lambda_0 \Delta}\Psi(1 + \lambda_1,\ \Delta;\ r_t,\ r_{t+\Delta})$
        Set $\mathcal{L}_\star \leftarrow \mathcal{L}_\star \times A e^{\Lambda(r_t)\Delta}\Lambda(r_{t+\Delta})/\Lambda(r_t)$
        Draw $\eta_{t+\Delta}$ from $\mathcal{N}((\theta - r_{t+\Delta})u, v^2)$ and set $\delta_{t+\Delta} \leftarrow r_{t+\Delta}(\exp(\eta_{t+\Delta}) - 1)$
        Update $r_t \leftarrow r_{t+\Delta} + \delta_{t+\Delta}$ and $\tau \leftarrow \tau - \Delta$
      end if
    end while
    Update $\hat{Y} \leftarrow \hat{Y} + \mathcal{L}_\star$
  end for
  return 1 - $\hat{Y}/M$
end procedure

to construct the rate-exchange measure $P^{\leftrightarrow}$ from the reference measure P in the sense that

$$Y_n = \mathrm{P}\left(\mathcal{H}_n\right) = \mathrm{E}^{\mathrm{P}^{\leftrightarrow}}\left(1_{\{\mathcal{H}_n\}}\mathcal{L}^{\leftrightarrow}_{(A,D)}\right)$$

holds. Under the rate-exchange measure $\mathrm{P}^{\leftrightarrow}_{(A,D)}$, we cam draw $A_t^{(n)}$ and $D_t^{(n)}$ by applying the A/R scheme as if they are time-inhomogeneous Poisson processes with rate $\mu_t^{(n)}$ and $\lambda_t^{(n)}$ for $t \geq 0$, respectively. Hence, we obtain the estimator of Y_n by

$$\hat{Y}_n \approx \frac{1}{M} 1_{\{\mathcal{H}_n\}}\mathcal{L}^{\leftrightarrow}_{(A,D)},$$

based on exchanging the rates between A and D under the importance sampling measure $\mathrm{P}^{\leftrightarrow}$.

(iii) **(cIS with tilting scheme)** We follow the same notations as the ones used in Section 5.3. A heuristic approach guiding the optimal path to overflow equals the most likely path suggests the optimal choice of θ as

$$\left(n\overline{\lambda}e^{\theta} - n\overline{\mu}e^{-\theta}\right)T + cn = \xi n \quad \Rightarrow \quad \theta = \log\left(\frac{\xi - c + \sqrt{(\xi - c)^2 + 4\overline{\lambda}\overline{\mu}T^2}}{2\overline{\lambda}T}\right),$$

which yields

$$Y_n = \mathrm{P}(\mathcal{H}_n) = \mathrm{E}^{\mathrm{P}}\left(1_{\mathcal{H}_n}\right) = \mathrm{E}^{\mathrm{P}_\theta}\left(1_{\mathcal{H}_n} Z_\theta\right),$$

where Z_θ is the Radon-Nikodym derivative of P with respect to P_θ. Combining the proposed cIS methodology with the optimal tilting scheme, we obtain the estimator of Y_n by

$$\hat{Y}_n \approx \frac{1}{M} 1_{\{\mathcal{H}_n\}}\mathcal{L}^{\theta}_{(A,D)},$$

where the event times of A and D are independent and uniformly distributed on $[0, T]$, and $\mathcal{L}^{\theta}_{(A,D)}$ is given by the equation (11). Algorithm 4 shows the detailed algorithm for implementing the 'cIS with tilting' scheme.

Algorithm 4 Overflow probability estimation under the 'cIS with tilting' scheme

procedure cIS_TILTING_OVERFLOWPROB(n, M) ▷ M is the number of simulated samples

Initialize $\hat{Y} \leftarrow 0$ and set $\theta \leftarrow \log\left(\frac{\xi - c + \sqrt{(\xi-c)^2 + 4\overline{\lambda}\overline{\mu}T^2}}{2\overline{\lambda}T}\right)$

for $m \in \{1, 2, \ldots, M\}$ **do**

Draw $\alpha \sim \text{Poi}\left(n\overline{\lambda}e^{\theta}T\right)$ ▷ Simulated number of arrivals

Generate an ordered sample of $\{U_1, \cdots, U_\alpha\}$ from $U(0, T)$ ▷ Simulated arrival times

Draw $\beta \sim \text{Poi}\left(n\overline{\mu}e^{-\theta}T\right)$ ▷ Simulated *potential* number of departures

Generate an ordered sample of $\left\{V_1, \cdots, V_\beta\right\}$ from $U(0, T)$ ▷ Simulated *potential* service times

if $\lfloor nc \rfloor + \alpha - \beta \geq \lceil nY \rceil$ **then**

$Z \leftarrow 1$

else

$\gamma \leftarrow 1$ ▷ γ counts the number of *false* service times while the server is idle

for $k \in \{1, \ldots, \beta\}$ **do**

$i \leftarrow \sum_{j \in \{U_1, \cdots, U_\alpha\}} 1_{\{V_k > U_j\}}$

if $\lfloor nc \rfloor + i - k + \gamma < 0$ **then**

$\gamma \leftarrow \gamma + 1$

end if

end for

$Z \leftarrow 1_{\{\lfloor nc \rfloor + \alpha - \beta + \gamma \geq \lceil nY \rceil\}}$

end if

$\mathcal{L}^{\theta}_{(\alpha,\beta)} \leftarrow \exp\left(-n\int_0^T \left(\lambda_t - \overline{\lambda}e^{\theta} + \mu_t - \overline{\mu}e^{-\theta}\right) dt\right) \prod_{i=1}^{\alpha} \frac{\lambda_{U_i}}{\overline{\lambda}e^{\theta}} \prod_{j=1}^{\beta} \frac{\mu_{V_j}}{\overline{\mu}e^{-\theta}}$

$\hat{Y} \leftarrow \hat{Y} + \mathcal{L}^{\theta}_{(\alpha,\beta)} \times Z$

end for

return $\hat{Y} \leftarrow \hat{Y}/M$ ▷ Overflow probability estimator

end procedure

References

[AG07] S. Asmussen and P Glynn. *Stochastic Simulation – Algorithms and Analysis*. Springer, New York, 2007.

[AGS18] Shahriar Azizpour, Kay Giesecke, and Gustavo Schwenkler. Exploring the sources of default clustering. *Journal of Financial Economics*, 129(1):154–183, 2018.

[BK06] M. Broadie and O. Kaya. Exact simulation of stochastic volatility and other affine jump diffusion processes. *Operations Research*, 54(2):217–231, 2006.

[BN12] Natalia Beliaeva and Sanjay Nawalkha. Pricing American interest rate options under the jump-extended constant-elasticity-of-variance short rate models. *Journal of Banking and Finance*, 36:151–163, 2012.

[BR16] J. Blanchet and J. Ruf. A weak convergence criterion for constructing changes of measure. *Stochastic Models*, 32(2):233–252, 2016.

[Buc04] J. A. Bucklew. *Introduction to Rare Event Simulation*. Springer-Verlag, New York, 2004.

[CC10] René Carmona and Stéphane Crépey. Particle methods for the estimation of Markovian credit portfolio loss distributions. *International Journal of Theoretical and Applied Finance*, 13(4):577–602, 2010.

[CFV09] René Carmona, Jean-Pierre Fouque, and Douglas Vestal. Interacting particle systems for the computation of rare credit portfolio losses. *Finance and Stochastics*, 13(4):613–633, 2009.

[CIR85] J. Cox, J. Ingersoll, and S. Ross. A theory of the term structure of interest rates. *Econometrica*, 53:385–408, 1985.

[DEHS09] Darrell Duffie, Andreas Eckner, Guillaume Horel, and Leandro Saita. Frailty correlated default. *Journal of Finance*, 64:2089–2123, 2009.

[DG95] Darrell Duffie and Peter Glynn. Efficient Monte Carlo estimation of security prices'. *Annals of Applied Probability*, 4(5):897–905, 1995.

[DG05] Pierre Del Moral and Josselin Garnier. Genealogical particle analysis of rare events. *Annals of Applied Probability*, 15:2496–2534, 2005.

[FLA91] M. R. Frater, T. M. Lemon, and B. D. O. Anderson. Optimally efficient estimation of the statistics of rare events in queueing networks. *IEEE Transactions on Automatic Control*, 36:1395–1405, 1991.

[GK11] Kay Giesecke and Baeho Kim. Systemic risk: What defaults are telling us. *Management Science*, 57(8):1387–1405, 2011.

[GKZ11] Kay Giesecke, Baeho Kim, and Shilin Zhu. Monte carlo algorithms for default timing problems. *Management Science*, 57(12):2115–2129, 2011.

[GS10] Kay Giesecke and Alexander Shkolnik. Importance sampling for indicator Markov chains. In B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, editors, *Proceedings of the 2010 Winter Simulation Conference*, pages 967–975. IEEE Press, 2010.

[GS13] Kay Giesecke and Dmitry Smelov. Exact sampling of jump diffusions. *Operations Research*, 61(4):894–907, 2013.

[GS14] Kay Giesecke and Alexander Shkolnik. Optimal importance sampling of default losses. Working Paper, Stanford University, 2014.

[GS20] Kay Giesecke and Alexander Shkolnik. Reducing bias in event time simulation via measure changes. Mathematics of Operations Research, Forthcoming, 2020.

[Joh04] Michael Johannes. The statistical and economic role of jumps in continuous-time interest rate models. *Journal of Finance*, 59(1):227–260, 2004.

[LS79] P. A. W Lewis and G. S. Shedler. Simulation of nonhomogeneous poisson processes by thinning. *Naval Research Logistics Quarterly*, 26(3):403—413, 1979.

[Mey71] P.-A. Meyer. Démonstration simplifée d'un théorème de knight. *Séminaire de Probabilités V. Lecture Notes in Mathematics 191*, pages 191–195, 1971.

[Pro05] Philip Protter. *Stochastic Integration and Differential Equations*. Springer-Verlag, New York, 2005.

[PW89] S. Parekh and J. Walrand. A quick simulation method for excessive backlogs in networks of queues. *IEEE Transactions on Automatic Control*, 34:54–66, 1989.

[Rid20] Ad Ridder. Rare-event analysis and simulation of queues with time-varying rates. August 2020. YouTube presentation in the 14th International Conference in Monte Carlo & Quasi-Monte Carlo Methods in Scientific Computing.

[Rob55] H. Robbins. A remark on Stirling's formula. *The American Mathematical Monthly*, 62(1):26–29, 1955.

[RT09] G. Rubino and B. Tuffin. *Rare Event Simulation using Monte Carlo Methods*. John Wiley, 2009.

[SB90] J. Sadowsky and J. A. Bucklew. On large deviations theory and asymptotically efficient monte carlo simulation. *IEEE Transition of Information Theory*, 36:579–588, 1990.

[Sie76] D Siegmund. Importance sampling in the Monte Carlo study of sequential tests. *The Annals of Statistics*, 4(4):673–684, 1976.

[VEW12] E. Vanden-Eijnden and J. Weare. Rare event simulation of small noise diffusions. *Communications on Pure and Applied Mathematics*, 65(12):1770–1803, 2012.

[XMS15] Yaofei Xiong, Duncan J. Murdoch, and David A. Stanford. Perfect sampling of a single-server queue with periodic poisson arrivals. *Queueing Systems*, 80:15—33, 2015.

Table 1. *Estimates of* $P(\overline{N}_T \geq k)$ *for* $T = 1$ *year under the default intensity model*

k	pMC Estimate	cIS Estimate	pMC Relative Error	cIS Relative Error	Variance Reduction Ratio
5	0.412192318	0.413040302	1.16067900	1.10336106	1.102056543
6	0.273098276	0.274012591	1.57685662	1.06868475	2.162629238
7	0.169399224	0.170499163	2.12938486	1.04419869	4.105068100
8	0.098904979	0.100056099	2.88625621	1.01629616	7.880946291
9	0.054844707	0.055701032	3.94623136	1.02030140	14.50277367
10	0.029220524	0.029499693	5.44344754	1.03533313	27.12244654
11	0.014958138	0.015061468	7.61512492	1.05983324	50.92129466
12	0.007331477	0.007389213	10.8574088	1.08842431	97.95858220
13	0.003574434	0.003470969	15.4774840	1.07479839	219.9174754
14	0.001652759	0.001590929	22.6404779	1.15468716	414.9156103
15	0.000710317	0.00070593	34.2718602	1.21971791	799.3503795
16	0.000326411	0.000303376	50.3776475	1.22958016	1943.260429
17	0.000145299	0.000126311	76.3227857	1.24494897	4973.359027
18	6.54620E-05	5.16438E-05	111.020962	1.28187186	12052.09551
19	3.39612E-05	2.05575E-05	151.108144	1.35760861	33810.69244
20	1.71527E-05	7.98745E-06	213.815107	1.45511983	99569.89798
21	6.49714E-06	3.02651E-06	354.100529	1.48279312	262816.2317
22	3.26046E-06	1.12316E-06	500.199438	1.49347978	945289.1945
23	N/A	4.08726E-07	N/A	1.71653526	N/A
24	N/A	1.45143E-07	N/A	1.59962853	N/A

Note. This table reports cIS and pMC (JAM) estimates of $Y = P\left(\overline{N}_T \geq k\right)$ for 20 values of k. The pMC and cIS estimates from $M = 5 \times 10^5$ samples are reported in the second and third columns, while the fourth and fifth columns report the relative errors of the pMC and cIS estimators, respectively. The last column reports the variance reduction ratio defined as the sample variance of the exact pMC estimator over that of the unbiased cIS estimator.

Table 2. *Bond price estimation results under the short-rate model in (10) with maturity $T = 3$ years*

Method	Trials (K)	Steps	Bias	SE	RMSE	Time (sec)
Trapezoidal Discretization	10	100	5.2978E-05	2.6705E-04	2.7225E-04	0.2594
	50	225	2.1852E-05	1.1908E-04	1.2107E-04	0.9502
	250	500	8.5263E-06	5.3402E-05	5.4078E-05	10.6361
	1,250	1,120	5.3556E-06	2.3793E-05	2.4389E-05	119.5019
Multinomial Tree	N/A	1,000	-3.7366E-04	0	3.7366E-04	0.2403
		2,500	-1.5910E-04		1.5910E-04	1.5892
		7,000	-8.7096E-05		8.7096E-05	12.0744
		20,000	-5.3870E-05		5.3870E-05	115.0961
Conditional Importance Sampling	50	N/A	0	1.5477E-04	1.5477E-04	0.2164
	500			4.9092E-05	4.9092E-05	1.4017
	5,000			1.5513E-05	1.5513E-05	14.1515
	50,000			4.8123E-06	4.8123E-06	142.4679

1.25**Note.** Bond prices are estimated under the short-rate model in (10). The selected parameters are $(\kappa, \theta, \sigma, \lambda_0, \lambda_1, u, v, r_0) = (0.1, 0.08, 0.05, 1.0, 5.0, 1.0, 1.0, 0.05)$. The true price of the bond as estimated by the cIS algorithm with $M = 10^{10}$ iterations is 0.8447251.