This article was downloaded by: [2001:638:807:212:a8af:506c:25a1:b254] On: 09 November 2026, At: 05:59
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA



Mathematics of Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Reducing Bias in Event Time Simulations via Measure Changes

Kay Giesecke, Alexander Shkolnik

To cite this article:
Kay Giesecke, Alexander Shkolnik (2022) Reducing Bias in Event Time Simulations via Measure Changes. Mathematics of Operations Research 47(2):969-988. https://doi.org/10.1287/moor.2021.1156

Full terms and conditions of use: https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2022, INFORMS

Please scroll down for article—it is on subsequent pages

With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

http://pubsonline.informs.org/journal/moor

MATHEMATICS OF OPERATIONS RESEARCH
Vol. 47, No. 2, May 2022, pp. 969–988
ISSN 0364-765X (print), ISSN 1526-5471 (online)

Downloaded from informs.org by [2001:638:807:212:a8af:506c:25a1:b254] on 09 November 2026, at 05:59 . For personal use only, all rights reserved.

Reducing Bias in Event Time Simulations via Measure Changes

Kay Giesecke,[a] **Alexander Shkolnik**[b]

[a] Department of Management Science and Engineering, Stanford University, Stanford, California 94305; [b] Department of Statistics and Applied Probability, University of California, Santa Barbara, California 93106

Contact: giesecke@stanford.edu, https://orcid.org/0000-0002-5380-2918 (KG); shkolnik@ucsb.edu, https://orcid.org/0000-0003-4523-8592 (AS)

Received: October 27, 2016
Revised: August 1, 2017; February 2, 2018
Accepted: June 22, 2018
Published Online in Articles in Advance: February 2, 2022

MSC2020 Subject Classification: Primary: 65C05, 60G55, 60G44; Secondary: 90-10, 91-10, 60J76

https://doi.org/10.1287/moor.2021.1156

Copyright: © 2022 INFORMS

Abstract. Stochastic point process models of event timing are common in many areas, including finance, insurance, and reliability. Monte Carlo simulation is often used to perform computations for these models. The standard sampling algorithm, which is based on a time-change argument, is widely applicable but generates biased simulation estimators. This article develops and analyzes a change of probability measure that can reduce or even eliminate the bias without restricting the scope of the algorithm. A result of independent interest offers new conditions that guarantee the existence of a broad class of point process martingales inducing changes of measure. Numerical results illustrate our approach.

Funding: This work was supported by the Center for Risk Management Research at UC Berkeley.

Keywords: simulation bias • point processes • event timing • changes of measure • time scaling • Monte Carlo methods

1. Introduction

In the context of Monte Carlo (MC) simulation, changes of probability measure have by and large been employed to reduce the variance of a simulation estimator.[1] Samples are generated under an auxiliary probability measure Q instead of the reference probability P. The resulting distortion is corrected by reweighting the estimator by the likelihood ratio of the two measures. A carefully chosen Q can reduce the simulation variance. In this article, by contrast, we construct and analyze a change of measure to reduce or even eliminate the simulation bias.

We consider the simulation of point processes, which are widely used in finance, economics, queuing, actuarial science, and other disciplines as models for the timing of events. Some examples of event timing include corporate bankruptcies, market crashes, equipment failures, and insurance claim arrivals. The point process dynamics are specified by an intensity, or conditional event arrival rate. A deterministic intensity generates a Poisson process, which is easily sampled exactly. Many applications call for a stochastic intensity process, which specifies a richer point process model. A standard sampling approach for such point processes harnesses a result of Meyer [40], which states that under mild conditions, a point process can be constructed by *time-changing* a Poisson process by the compensator (cumulative intensity); that is, the point process arrivals are the hitting times of the compensator for a sequence of levels determined by the Poisson arrival times. This sampling approach has a very broad scope.

A major disadvantage of the time-scaling approach is that it usually generates biased simulation estimators. For most point process models of interest, only a skeleton of the compensator process can be generated.[2] A skeleton, however, is insufficient to determine the exact hitting times. The required approximation of the hitting times introduces bias into the simulation estimator. The analysis of Shkolnik et al. [47] suggests that the convergence rate of the estimator's error is of weak order one for a suitable class of functions.

We develop a change of probability measure that can reduce or even eliminate the bias of the time-scaling algorithm without restricting its scope of applicability. We construct a simulation measure, Q, under which the point process intensity is a pure jump process with jumps occurring at the event times. The Q-intensity approximates the reference measure (P) intensity, and the two coincide at the event times. The corresponding compensator is a piecewise-linear approximation of the P-compensator, and therefore can be easily inverted to exactly determine the hitting (i.e., event) times under Q. The Radon–Nikodym derivative between P and Q transforms the Q-samples into the desired P-samples.

Our approach eliminates the need to approximate the event times on a discretized time grid, given a skeleton of the compensator. It avoids the dominant (order one) source of bias of the time-scaling algorithm. Analogous to time scaling, the remaining potential sources of bias in our algorithm are the sampling of intensity values and their numerical integration, which is required to evaluate the Radon–Nikodym derivative. Exact methods can

Downloaded from informs.org by [2001:638:807:212:a8af:506c:25a1:b254] on 09 November 2026, at 05:59 . For personal use only, all rights reserved.

potentially be used to eliminate the intensity sampling bias.[3] Assuming exact intensity samples, we develop a bound for any residual bias of the simulation estimator due to integration. We also show that if a certain bridge transform can be evaluated for the model at hand, then the integration bias can be entirely eliminated and the algorithm is exact. This is possible for a broad class of jump-diffusion intensity models.[4]

Remarkably few adjustments to the standard time-scaling algorithm are required to incorporate our change of measure. The new algorithm has several advantages over the standard scheme, in addition to the reduction in simulation bias. The implementation is simpler. The simulation of the estimators can be vectorized and parallelized to achieve significant computational gains. Finally, the improved accuracy of higher-order schemes and unbiased methods can be harnessed to mitigate the potential remaining sources of bias described above.

Our measure change is based on a result that is of independent interest. We provide mild conditions under which a class of important, point process local martingales (Brémaud [8, theorem VI.T2]) is a class of (true) martingales. This facilitates the construction of measure changes for point processes and extends the results of Blanchet and Ruf [5], Brémaud [8], Collin-Dufresne et al. [15], Giesecke and Zhu [25], Sokol and Hansen [48], and others.

The remainder of the introduction discusses the related literature. Section 2 introduces our point process model for event timing, states the computational problem, and reviews the standard time-scaling algorithm. Section 3 introduces the theoretical change of measure results underlying our algorithms. Sections 4 and 5 develop and analyze the plain and exact versions of our scheme. A numerical study illustrating the performance of the algorithms is found in Section 6.

1.1. Related Literature

Prior work has examined the simulation of point processes, with recent focus placed on exact algorithms. Daley and Vere-Jones [17] review some approaches, including time scaling. Because of its broad scope, the time-scaling algorithm appears to be the most widely used scheme in applications requiring the simulation of point processes. See, for instance, Duffie and Singleton [20] for a discussion of the time-scaling algorithm in the context of default timing applications, where one is interested in estimating financial loss due to default in portfolios of bonds, or the computation of prices of securities backed by these pools. Giesecke and Kim [23] show that the time-scaling algorithm can generate a downward-biased estimator of portfolio risk measures, which in turn might lead to insufficient risk capital reserves. Our results supply adjustments to the time-scaling algorithm that are guaranteed to significantly improve the performance of the simulation estimator. These adjustments do not restrict the broad scope of the algorithm.

Lewis and Shedler [38], Ogata [42], Glasserman and Merener [31], Li [39], and others develop thinning algorithms. These schemes are exact if the intensity is bounded (possibly by a process), and the intensity, the bound, and the arrival times of the point process governed by the bound can be sampled exactly. However, many point process models of interest in applications have unbounded intensities that are driven by (jump) diffusions. Giesecke et al. [27] develop thinning-type schemes to treat these cases. The schemes allow one to control (and in some cases even eliminate) the bias. However, they require knowledge of transition and hitting time distributions for the intensity process. The sampling algorithms developed in this article cover unbounded intensities. Our standard algorithm does not require the computation of any nontrivial quantities, and thus has a very broad scope. The exact version of the algorithm requires only the computation of a bridge transform of the compensator,[5] which is tractable for a large class of jump-diffusion intensity models.

Giesecke et al. [26, 28] use projection arguments to develop exact sampling schemes. Their schemes require the computation of certain point process filters, which is straightforward for some models but difficult for others. This is not unlike the exact version of our algorithm, for which certain precomputations are required. Duffie and Singleton [20] and Giesecke and Kim [23] discuss exact inverse transform schemes. These algorithms require the evaluation of a bridge transform, and the sampling of values from this transform, which can be computationally expensive and difficult to implement. The exact version of our algorithm requires only the evaluation of that transform, not the sampling of values from it.

Finally, there is growing interest in the exact sampling of jump-diffusion processes with state-dependent coefficient and jump intensity functions (e.g., Casella and Roberts [10], Giesecke and Smelov [24], Pollock et al. [44]). Here, the jump times of the process generate the point process of interest. Unlike our approach, these methods are currently limited to one-dimensional processes. However, the methods can potentially be used in the implementation of our algorithms (see Chen et al. [11]).

Downloaded from informs.org by [2001:638:807:212:a8af:506c:25a1:b254] on 09 November 2026, at 05:59 . For personal use only, all rights reserved.

2. Problem Formulation

2.1. Event Timing Models

Fix a complete probability space $(\Omega, \mathcal{F}, \mathrm{P})$ with a right-continuous and complete information filtration $\mathbb{F} = \{\mathcal{F}_s\}_{s\geq 0}$. Consider $n \in \mathbb{N}$ distinct and totally inaccessible stopping times $\{\tau_i\}_{i=1}^n$. The variable τ_i represents the arrival time of some event i, for example, a loan default, an insurance claim, a failure of a system component. An event indicator process N^i is defined as

$$N_t^i = \mathbb{1}_{\{\tau_i \leq t\}}$$

for $t \geq 0$, where $\mathbb{1}_{\mathcal{A}}$ is the indicator of $\mathcal{A} \in \mathcal{F}$. The process $\boldsymbol{N} = (N^1, \ldots, N^n)$ takes values in $\mathbb{S} = \{0,1\}^n$ and is a model for the timing of n events.

A standard approach to constructing the process $\boldsymbol{N}$ proceeds as follows.[6] Let $\mathbb{H}_0$ be a right-continuous, complete filtration, and let $\mathbb{H}_i = \{\mathcal{H}_s^i\}_{s\geq 0}$ be the progressive enlargement of $\mathbb{H}_0$ (see Protter [45, chapter VI.3]) by the random times $\{\tau_j\}_{j\neq i}$, defined below. Let $\{\xi_i\}_{i=1}^n$ be a collection of independent and identically distributed (i.i.d.) standard exponential random variables with each $\sigma(\xi_i)$ and $\mathcal{H}_\infty^i$ independent. Then, for a continuous, increasing, $\mathbb{H}_i$-adapted process H^i that starts at zero, we set[7]

$$\tau_i = \inf\{t > 0 : H_t^i \geq \xi_i\}. \tag{1}$$

The filtration $\mathbb{F}$ is then taken as $\vee_{i=1}^n \mathbb{H}_i$. The independence of $\sigma(\xi_i)$ and $\mathcal{H}_\infty^i$ ensures that the knowledge of the event time is not involved in its own construction. However, information about the events other than itself may be used.[8]

The pairs (H^i, ξ_i) are closely related to the compensator of $\boldsymbol{N}$, which governs the joint law of the event times. By the Doob–Meyer decomposition, this compensator is the unique, increasing, predictable process $\boldsymbol{A} = (A^1, \ldots, A^n)$ with $A^i = 0$ such that every $M^i = N^i - A^i$ is a local $\mathbb{F}$-martingale. In fact, each M^i is a martingale.[9] As the event times are totally inaccessible, A^i is continuous (Protter [45, theorem III.24]) and $A_t^i = H_{t\wedge\tau_i}^i$. Moreover, $\boldsymbol{A}$ defines a change of time under which $\boldsymbol{N}$ is a vector of independent Poisson processes stopped at their first arrival times (Meyer [40]). If every $\tau_i < \infty$ almost surely, the $\{A_\infty^i\}$ are i.i.d. standard exponentials, corresponding to the $\{\xi_i\}$ in (1).[10]

It is convenient and standard to assume that each H^i is absolutely continuous with respect to the Lebesgue measure, that is,

$$H_t^i = \int_0^t h_s^i \mathrm{d}s$$

for some nonnegative and $\mathbb{H}_i$-progressively measurable process h^i. We consider $\boldsymbol{h} = (h^1, \ldots, h^n)$ to be the modeling primitive but do not impose any particular structure. In applications, $\boldsymbol{h}$ might be governed by a system of stochastic differential equations (SDEs) that represent diffusive as well as jump sources of randomness.[11] The process $\boldsymbol{h}$ uniquely specifies the law of $\boldsymbol{N}$, by (1). It can be interpreted as a conditional arrival rate, because

$$\begin{aligned}\frac{1}{\delta}\mathrm{P}(N_{s+\delta}^i - N_s^i = 1 \mid \mathcal{F}_s) &= \mathbb{1}_{\{\tau_i > s\}} \frac{1}{\delta}\mathrm{E}(A_{s+\delta}^i - A_s^i \mid \mathcal{F}_s) \\ &\approx h_s^i \mathbb{1}_{\{\tau_i > s\}}\end{aligned}$$

for small $\delta > 0$ with equality almost surely as $\delta \downarrow 0$ by dominated convergence (see footnote 9 for the requisite integrability condition). We let

$$p_s^i = h_s^i \mathbb{1}_{\{\tau_i > s\}}$$

denote the Lebesgue density of the compensator process A^i. The vector process $\boldsymbol{p} = (p^1, \ldots, p^n)$ is referred to as the intensity of $\boldsymbol{N}$ (Brémaud [8]).

2.2. Computational Problem

The combinatorial (in $\mathbb{S}$) aspect of many event timing models limits the scope of possible computational methods. In particular, given a Borel $f : \mathbb{S}^{[0,\infty)} \to \mathbb{R}$, we are interested in computing the expectation[12]

$$y = \mathrm{E}(f(\boldsymbol{N})). \tag{2}$$

Analytical or semianalytical (i.e., deterministic) methods to compute the expectations in (2) usually have exponential complexity. For example, Jarrow and Yu [32] provide an elementary model that leads to a formula for the joint and marginal distributions of the $\{\tau_i\}_{i=1}^n$ with $O(2^n)$ terms (see Collin-Dufresne et al. [15, section 4.3]).[13] It is not known whether this may be improved, but the problem is NP-hard (reducible to the independent set

Downloaded from informs.org by [2001:638:807:212:a8af:506c:25a1:b254] on 09 November 2026, at 05:59 . For personal use only, all rights reserved.

problem), and one may craft instances that take exponential time to approximate. A common approach to estimating (2) that overcomes these difficulties is Monte Carlo simulation.

To estimate y in (2) via MC simulation, we generate a collection $\{\hat{Y}^{(k)}\}_{k=1}^{m}$ of m i.i.d. samples $\hat{Y}$ of $f(\boldsymbol{N})$. The empirical average $\hat{y}_m = \sum_{k=1}^{m} \hat{Y}^{(k)}/m$ is the MC estimate of y. Unbiased estimates have $y = \mathrm{E}(\hat{y}_m)$, but often the sample $\hat{Y}$ is biased. In this case, $y = \mathrm{E}(\hat{y}_m) + \text{bias}$, where the bias term is due to $\hat{Y}$ being sampled using an approximation to the law of $\boldsymbol{N}$, not the exact law. The bias typically persists even when the number of samples m is increased. To see this, consider the mean-squared error (MSE) of the estimate

$$\mathrm{MSE}(\hat{y}_m) = \mathrm{E}((\hat{y}_m - y)^2) = \sigma_{\hat{Y}}^2/m + \text{bias}^2, \tag{3}$$

where $\sigma_{\hat{Y}}^2 = \mathrm{Var}(\hat{Y})$ is the variance of the sample. Increasing the number of trials m reduces the error $\sigma_{\hat{Y}}^2/m$, but the bias remains unaffected. The bias may be reduced only by increasing the computational effort associated with obtaining the samples. When $\hat{Y}$ can be sampled from the exact law of $f(\boldsymbol{N})$, (3) implies that $\mathrm{MSE}(\hat{y}_m) \to 0$ as $m \uparrow \infty$. The choice between a biased or an exact estimator is typically based on a trade-off between computational effort and desired precision (see Asmussen and Glynn [1], Glasserman [30]).

As discussed in Section 1.1, there are several schemes, biased and exact, for sampling $f(\boldsymbol{N})$. The scheme with the widest scope is based on Meyer's time change (see Section 2.3; cf. Shaked and Shanthikumar [46], Yu [50]).

2.3. Time-Scaling Scheme

The time-scaling construction (1) of $\boldsymbol{N}$ suggests a basic scheme for simulating the event times $\{\tau_i\}$ for a given h using independent, exponential random variables. Algorithm 1 summarizes the steps. Step 2 contains the only nontrivial part of the procedure.[14] We see that the time-scaling estimator is unbiased only if exact samples from the law of $\boldsymbol{H}^{-1}$ can be drawn. This is rarely possible for models common in applications. For most models of h, the time integral in (4) will be approximated. A standard approach is to discretize time as $0 < t_1 < \cdots < t_j < \cdots$, and simulate values of $\boldsymbol{h}$ at the $\{t_j\}$ to approximate the integral. Although this approach has a broad scope (only samples of the skeleton of h are required), it introduces bias into the MC estimator.

Algorithm 1 (Time Scaling; Generates a Sample $\hat{Y}$ of $Y = f(\boldsymbol{N})$)

1. Draw i.i.d. standard exponential random variables $\xi_1, \ldots, \xi_n$.
2. For every ξ_i generate a path of $\hat{h}^i$ (and $\hat{H}^i$) up to $\hat{\tau}_i$ such that

$$\int_0^{\hat{\tau}_i} \hat{h}_s^i \mathrm{d}s = \xi_i. \tag{4}$$

3. Set $\hat{N}_s^i = \mathbb{1}_{\{\hat{\tau}_i \le s\}}$ for $s \ge 0$ and $i = 1, \ldots, n$ and **return** $f(\hat{\boldsymbol{N}})$.

Notes. $\hat{X}$ refers to a sample of a random variable X.

We distinguish the following three potential sources of bias (throughout, $\Delta_g = t_j - t_{j-1}$):

1. *Intensity discretization.* Often the samples of $\boldsymbol{h}$ at the grid points $\{t_j\}$ are generated using a discretization scheme such as Euler's method. The use of such schemes introduces error into the samples. The analysis of this source of bias is relatively standard (see Asmussen and Glynn [1, chapter X]).

2. *Numerical integration.* The approximation of the integral in (4) given the values $\{\hat{h}_{t_j}^i\}$ also introduces bias. A variety of quadrature rules may be used. It is often convenient to consider the discretization of $\mathrm{d}H_s^i = h_s^i \mathrm{d}s$ to analyze the error in conjunction with that of the intensity process discretization.

3. *Event time location.* The $\hat{\tau}_i$ must be placed within a grid cell $(t_{j-1}, t_j]$ such that $\hat{H}_{t_{j-1}}^i < \xi_i$ and $\hat{H}_{t_j}^i \ge \xi_i$ per (4). Various heuristics are available. The right and midpoint rules set $\hat{\tau}_i = t_j$ and $\hat{\tau}_i = (t_{j-1} + t_j)/2$, respectively. Taking $\hat{h}^i$ linear on $(t_{j-1}, t_j]$, the equation $\hat{H}_{\hat{\tau}_i}^i = \xi_i$ is approximated by taking the larger root of $a\hat{\tau}_i^2 + b\hat{\tau}_i + c = 0$ for $a = \frac{1}{2}(\hat{H}_{t_j}^i - \hat{H}_{t_{j-1}}^i)/(t_j - t_{j-1})$, $b = \hat{h}_{t_{j-1}}^i$ and $c = \hat{H}_{t_{j-1}}^i - \xi_i$. One could also interpolate H^i directly to solve for the crossing time. However, irrespective of the heuristic, the error in general is $O(\Delta_g)$.

The bias due to (1) can be reduced or eliminated by higher-order discretization, unbiased, or exact methods, respectively.[15] Higher-order numerical integration methods can also help mitigate the bias due to (2). If $\boldsymbol{H}$ is simulated exactly at the grid points,[16] the bias due to (1) and (2) vanishes. In any case, the event time location $O(\Delta_g)$ bias, (3), remains.[17] It dominates the error of the time-scaling estimator. For further discussion, see Shkolnik et al. [47].

Downloaded from informs.org by [2001:638:807:212:a8af:506c:25a1:b254] on 09 November 2026, at 05:59 . For personal use only, all rights reserved.

3. Simulation Measure

We develop a change of measure from P to a simulation measure Q that can reduce or even eliminate the bias of the time-scaling scheme (see Sections 4 and 5). This section establishes the theoretical foundations of our approach. To avoid cumbersome technical difficulties, we assume where it is relevant that the information filtration $\mathbb{F}$ is augmented by the null sets of Q (see Remark 3.5).

As $\boldsymbol{N}$ is specified in terms of an intensity process, a natural way to induce a change of measure is to adjust the intensity process $\boldsymbol{p}$ of $\boldsymbol{N}$; that is, we require every N^i admit a $(\mathrm{Q},\mathbb{F})$-intensity q^i so that $N^i_\cdot - \int_0^\cdot q^i_s \mathrm{d}s$ forms a $(\mathrm{Q},\mathbb{F})$-martingale. To this end, we let $\boldsymbol{u} = (u^1,\dots,u^n)$ be a process such that every u^i is nonnegative, $\mathbb{H}_i$-predictable and locally bounded. For each i we define

$$q^i = u^i p^i \text{ and } \mathrm{Q}^i_t = \int_0^t u^i_s h^i_s \mathrm{d}s. \tag{5}$$

Without loss of generality, and throughout this section, we take the predictable version of the intensity p (also of $\boldsymbol{h}$) see Brémaud [8, theorem II.T13]).[18] Under the simulation measure Q, our MC estimator of (2) takes the form

$$Y = f(\boldsymbol{N})L_\infty \tag{6}$$

where L_∞ is a positive random variable given by[19]

$$L_\infty = \exp\left(\sum_{i=1}^n (\mathrm{Q}^i - H^i)_{\tau_i} - N^i_\infty \log(u^i_{\tau_i})\right). \tag{7}$$

Under the mild conditions to be stated in Lemma 3.2 below, L_∞ is the Radon–Nikodym derivative of P with respect to Q. Consequently,

$$y = \mathrm{E}_\mathrm{Q}(Y). \tag{8}$$

We now provide the technical results to establish a simulation measure Q. We let $Z_\infty = 1/L_\infty$ throughout.[20]

Theorem 3.1. *Suppose that for every i we have either $\mathrm{Q}^i_\infty = \infty$ or $H^i_\infty < \infty$ P-almost surely. Then, $\mathrm{Q} = Z_\infty \mathrm{P}$ defines an absolutely continuous probability measure. Moreover, $\boldsymbol{N}$ admits the Q-intensity $\boldsymbol{q} = (q^1,\dots,q^n)$ per (5).*

The proof of Theorem 3.1 is deferred to Appendix A and concludes that the random variable $Z_\infty \geq 0$ has $\mathrm{E}(Z_\infty) = 1$ and thus defines the absolutely continuous change of measure. From the predictable version of the Girsanov–Meyer theorem (Protter [45, chapter III.41]) one can deduce the adjustment of any local $(\mathrm{P},\mathbb{F})$-martingale under passage from P to Q. In particular,

$$X_\cdot - \sum_{i=1}^n \int_0^\cdot (u^i_s - 1)\mathrm{d}\langle X, N^i\rangle_s \tag{10}$$

is a local Q-martingale whenever X is a local P-martingale. The Q-intensity claim of Theorem 3.1 follows directly from this fact by taking $X = N^i$. Furthermore, it is easy to see that only the local martingales that have jumps in common with at least one of the $\{N^i\}$ are adjusted under the change of measure.

To our knowledge, Theorem 3.1 is new. Taking $Z_t = \mathrm{E}(Z_\infty \mid \mathcal{F}_t)$ as in (9), the Doob martingale Z (rather, its càdlàg version) is a special case of the Doléan–Dade stochastic exponential that has long been known to be a local P-martingale (Brémaud [8]). However, a true martingale is required for the existence of Q. Conditions for this in the literature,[21] one way or another, place integrability assumptions on Z. In contrast, we prove Theorem 3.1 by showing that $\{Z_{\theta_k}\}_{k\in\mathbb{N}}$ is uniformly integrable for suitable localization times $\{\theta_k\}_{k\in\mathbb{N}}$. The conclusion then follows by Vitali's convergence theorem (Bogachev [7, theorem 4.5.4]). Integrability conditions are avoided as Theorem 3.1 essentially requires that if $\tau_i < \infty$ P-almost surely, then also $\tau_i < \infty$ Q-almost surely.

Recent, relevant work on construction of measure changes from local martingales may be found in Blanchet and Ruf [5]. Their approach recommends showing that the measures induced by the $\{Z_{\theta_k}\}_{k\in\mathbb{N}}$ are tight. Unfortunately, it does not yield the desired conclusions for a large enough class of Q-intensities. In particular, Blanchet and Ruf [5, theorem 4] supplies the conclusions of Theorem 3.1 in the case when $\boldsymbol{u}$ is bounded and deterministic. The approach does not seem to generalize and we pursue the more direct method through exhibiting uniform integrability. The extra effort turns out to be necessary as the change of measure we will require has a stochastic and unbounded $\boldsymbol{u}$.

Theorem 3.1 alone is not sufficient to establish the identity in (8). The measure Q must also be equivalent to P. There is a simple sufficient condition.

Downloaded from informs.org by [2001:638:807:212:a8af:506c:25a1:b254] on 09 November 2026, at 05:59 . For personal use only, all rights reserved.

Lemma 3.2. *Suppose the conditions of Theorem 3.1 hold. Then, if for every i, we have either* (i) $H^i_\infty = \infty$ *and* $u^i_s = 0$ *only when* $h^i_s = 0$ *or* (ii) $H^i_\infty < \infty$ *and* $(H^i - Q^i)_\infty$ *is* P*-almost surely finite, the measures* P *and* Q *are equivalent.*

Proof. Because $h^i_{\tau_i} > 0$ P-almost surely (Brémaud [8, theorem II.T12]), it suffices to observe that $Z_\infty > 0$ under the stated conditions. □

For computational purposes, a change of measure that is only absolutely continuous is still useful. Such a measure was introduced in Collin-Dufresne et al. [15]. Under this probability, no events occur up to some finite time.

Corollary 3.3. *Let $\mathcal{T}_i \subset [0,\infty)$ and $u^i_s = \mathbb{1}_{\{s \notin \mathcal{T}_i\}}$ for $s \geq 0$ such that $\mathcal{T}_i$ has finite Lebesgue measure and every u^i is predictable. Let* P_0 *denote the corresponding* Q *of Theorem 3.1. Then,* P_0*-almost surely, event i will not occur in $\mathcal{T}_i$.*

The measure P_0 will be used in Section 5 for estimator analysis and in Section 6 for certain event time probability calculations.

Remark 3.4. In contrast to Collin-Dufresne et al. [15, theorem 1 and equation 2.3], by Theorem 3.1, no integrability conditions are required for the existence of P_0 and the sets $\{\mathcal{T}_i\}$ are further allowed to be random (P_0-almost surely finite).

Remark 3.5. When h^i is positive on some $\mathcal{S} \subset \mathcal{T}_i$ and $\mathcal{S}$ has positive Lebesgue measure, the measures P and P_0 are not equivalent because $P_0(\tau_i \in \mathcal{S}) = 0$ but $P(\tau_i \in \mathcal{S}) > 0$. In this case, $\mathbb{F}$ must be augmented by all subsets of every $\{\tau_i \in \mathcal{T}_i\}$ when a complete filtration is required (see Collin-Dufresne et al. [15]).

In applications, changes of measure are often applied over a sequence of random times. To address this, let $\{T_\ell\}_{\ell \geq 0}$ be a sequence of stopping times with respect to some subfiltration $\{\mathcal{G}_s\}_{s \geq 0}$ of $\mathbb{F}$. For example, T_ℓ could denote the ℓ th arrival of the event counting process $\sum_{i=1}^n N^i$. Then, $Z_{T_\ell} = E(Z_\infty \mid \mathcal{G}_{T_\ell})$ induces an absolutely continuous measure change from P to Q restricted to $\mathcal{G}_{T_\ell}$. By the conditional change of measure formula, for $\mathcal{F}$-measurable X and $\ell \leq k$,

$$Z_{T_\ell} E_Q(X \mid \mathcal{G}_{T_\ell}) = E(XZ_{T_k} \mid \mathcal{G}_{T_\ell}) \tag{11}$$

(Q-almost surely) whenever $X \geq 0$ or XZ_∞ is P-integrable. If $Z_{T_k} > 0$ then, given $\mathcal{G}_{T_\ell}$, we may simulate a Y under an equivalent probability Q by setting $X = Y/Z_{T_k}$. The Radon–Nikodym derivative is Z_{T_ℓ}/Z_{T_k} and, effectively, we set $u^i_s = 1$ for every i and $s \geq T_k$ in (7) so that P and Q coincide after T_k.

4. Bias-Reducing Algorithm

We design an algorithm that avoids the event time location bias of the time-scaling scheme (Section 2.3). The idea is to sample $\boldsymbol{N}$ under a more convenient measure Q, and then use (6) as an estimator of (2). For practical reasons, we assume throughout that $\boldsymbol{h}$ is a càdlàg process and thus $\boldsymbol{h}_-$ and $\boldsymbol{p}_-$ form the predictable versions of $\boldsymbol{h}$ and $\boldsymbol{p}$, respectively. We assume every h^i is strictly positive. One possible relaxation of this condition is described in footnote 22.

4.1. Design of the Algorithm

Let S_ℓ denote the ℓ th arrival time of the event counting process $C = \sum_{i=1}^n N^i$ (with $S_0 = 0$). We design a simulation measure Q under which the paths of C are step functions. This ensures that the arrivals times $\{S_\ell\}_{\ell=0}^n$ are simple to simulate and is key to obtaining a bias reduction.

We first define the step function values,

$$\lambda_\ell = \sum_{i=1}^n p^i_{S_\ell} \quad \ell = 0, 1, \ldots, n. \tag{12}$$

To construct the measure, we apply Theorem 3.1 with every $u^i = \bar{u}$ in (5) for

$$\bar{u} = \frac{\lambda_C}{\sum_{i=1}^n p^i}. \tag{13}$$

Observe that $\bar{u}_-$ is a locally bounded, predictable version of $\bar{u}$ as required, but $u^i = \bar{u}$ is not $\mathbb{H}_i$-adapted as written because N^i (and thus p^i) is not $\mathbb{H}_i$-adapted. However, this shortcoming is easily addressed in the same way as in the construction of $\boldsymbol{N}$ of Section 2.1 (see footnote 8). Also, conditions of Lemma 3.2 are satisfied under our assumption of positivity of the $\{h^i\}$; thus, (8) holds. This assumption prevents the division by zero in (13) and can easily be relaxed.[22]

Downloaded from informs.org by [2001:638:807:212:a8af:506c:25a1:b254] on 09 November 2026, at 05:59 . For personal use only, all rights reserved.

Although each component Q-intensity q^i is a complicated process, the sum over the components, by construction, has the promised simpler form:

$$\sum_{i=1}^{n} q^i = \sum_{i=1}^{n} u^i h^i (1 - N^i) = \bar{u} \sum_{i=1}^{n} h^i (1 - N^i) = \lambda_C. \tag{14}$$

Paths of the process λ_C in (14) are step function approximations to paths of $\sum_{i=1}^{n} p^i$ on the (stochastic) grid $0 = S_0 < S_1 < \cdots < S_n$. Thus, the Q-intensity of the counting process C is a pure jump process and simulating its paths is elementary. Indeed, (14) stays constant on $[S_{\ell-1}, S_\ell)$ with value $\lambda_{\ell-1}$. Given $\mathcal{F}_{S_{\ell-1}}$, the Q-law of $S_\ell - S_{\ell-1}$ is exponential with parameter $\lambda_{\ell-1}$. Figure 1 illustrates a path of (14) with the process $\sum_{i=1}^{n} p^i$ superimposed. Substituting all $u^i = \bar{u}_-$ into (7), the Radon–Nikodym derivative L_∞ simplifies to

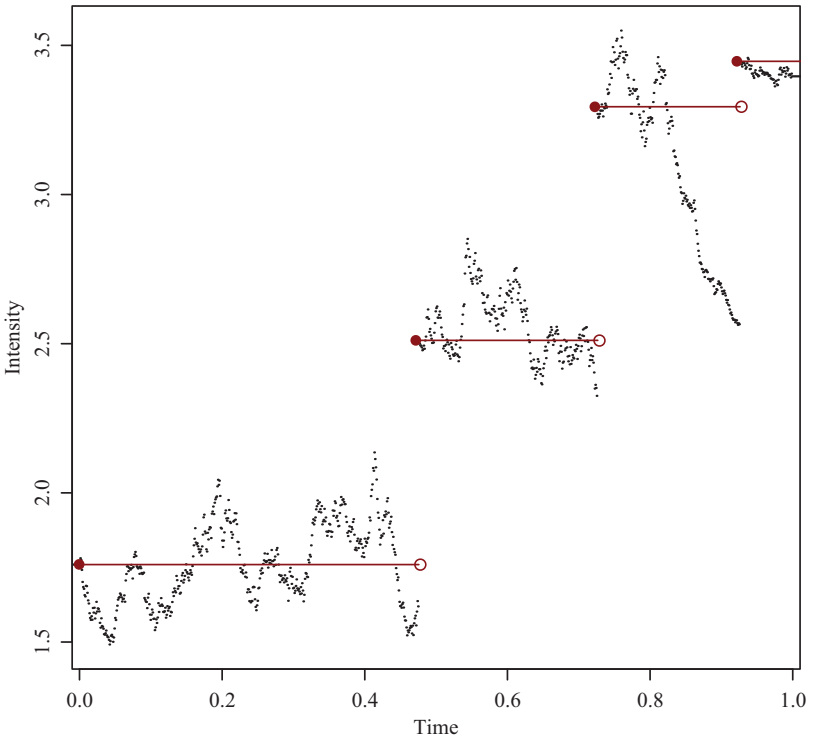
$$\exp\left(\sum_{\ell=1}^{n} \lambda_{\ell-1} (S_\ell - S_{\ell-1}) - \int_0^\infty p_s^\ell \, \mathrm{d}s - \log\left(\frac{\lambda_{\ell-1}}{\sum_{i=1}^{n} p^i_{S_\ell -}} \right) \right). \tag{15}$$

To sample $\boldsymbol{N}$ under Q, it is sufficient to sample $\lambda_{\ell-1}$ followed by an exponential η_ℓ with that rate to form S_ℓ. At time S_ℓ, the component $I_\ell \in \{1, 2, \ldots, n\}$ of $\boldsymbol{N}$ that experiences an event has the $\mathcal{F}_{S_\ell -}$-conditional law $q^i_{S_\ell -} / \sum_{i=1}^{n} q^i_{S_\ell -}$ (generalization of Brémaud [8, theorem II.T15]). This simplifies to

$$\mathrm{Q}(I_\ell = i \mid \mathcal{F}_{S_\ell -}) = \frac{p^i_{S_\ell -}}{\sum_{i=1}^{n} p^i_{S_\ell -}} \quad i = 1, \ldots, n. \tag{16}$$

Finally, (15) is evaluated and a sample of $f(\boldsymbol{N}) L_\infty$ is returned. We refer to this scheme as the plain jump approximation method (JAM).

Figure 1. (Color online) Illustration of a path of the Q-intensity of the event-counting process C for the model in Section 6. The dotted path illustrates a realization of the process $\sum_{i=1}^{n} p^i$ under Q. The step function (solid) illustrates the corresponding realization of the Q-intensity $\lambda_C = \sum_{i=1}^{n} q^i$ of C in (14). The times of increment of C coincide with the jump times of λ_C as each such jump time also marks the occurrence of a jump in $\sum_{i=1}^{n} p^i$. The (time) length of the ℓ th step $\eta_\ell = S_\ell - S_{\ell-1}$ of λ_C, conditionally on $\mathcal{F}_{S_{\ell-1}}$, is an exponentially distributed random variable of rate $\lambda_{\ell-1}$.



Downloaded from informs.org by [2001:638:807:212:a8af:506c:25a1:b254] on 09 November 2026, at 05:59 . For personal use only, all rights reserved.

Algorithm 2 (Plain JAM; Generates a Sample $\hat{Y}$ of $Y = f(\boldsymbol{N})L_\infty$ with L_∞ Given by (15) Under Q)

Initialize $\chi = \{1, \dots, n\}$ and $\ell = 0$.

1. Generate samples of the jump sizes $\{\Delta h^i_{S_\ell}\}_{i\in\chi}$.
2. Draw an exponential random variable $\eta_{\ell+1}$ with rate
$$\hat{\lambda}_\ell = \sum_{i\in\chi} \hat{h}^i_{S_\ell -} + \Delta \hat{h}^i_{S_\ell}.$$
3. Increment $\ell \leftarrow \ell + 1$ and set $\hat{S}_\ell = \hat{S}_{\ell-1} + \eta_\ell$.
4. Sample $(\sum_{i\in\chi} H^i_{S_\ell}, \boldsymbol{h}_{S_\ell -})$ and set
$$(R_\ell, r_\ell) = \sum_{i\in\chi} \left(\hat{H}^i_{S_\ell} - \hat{H}^i_{S_{\ell-1}}, \hat{h}^i_{S_\ell -}\right). \tag{17}$$
5. If $\ell = n$ (or some other stopping condition $\star$) then **return**
$$\hat{Y} = f(\hat{\boldsymbol{N}}) \exp\left(\sum_{k=1}^{n} \hat{\lambda}_{k-1}\eta_k - R_k - \log(\hat{\lambda}_{k-1}/r_k)\right). \tag{18}$$
6. Draw index $\hat{I}_\ell$ from the (conditional) distribution
$$\{\hat{h}^i_{S_\ell -}/r_\ell\}_{i\in\chi}.$$
7. Set $\hat{N}^{\hat{I}_\ell}_s = \mathbb{1}_{\{S_\ell \le s\}}$ for $s \ge 0$ and $\chi \leftarrow \chi \backslash \{\hat{I}_\ell\}$. Go to Step 1.

Notes. When terminating at some stopping time T, the L_∞ in (15) may be replaced by $\mathrm{E}(L_\infty \mid \mathcal{F}_T)$, which is equivalent to setting $\bar{u}_s$ in (13) to one for all $s \ge T$.

Notes. $\hat{X}$ refers to a sample of a random variable X.

4.2. Implementation of the Plain JAM

Algorithm 2 summarizes the steps. The initial rate of the Q-intensity λ_C of C is of course $\lambda_0 = \sum_{i=1}^n h^i_0$. In Step 1 and Step 2, with $\ell = 0$, we adopt the usual convention $h^i_{0-} = 0$ so that $\Delta h^i_{S_\ell} = h^i_0$. Otherwise, the jump sizes are drawn according to a given intensity specification. Step 2 and Step 3 generate the interarrival times $\{\eta_\ell\}$ and arrival times $\{S_\ell\}$ of the event counting process C. In Step 5, if the algorithm terminates, $\hat{\boldsymbol{N}}$ is a Q-sample of $\boldsymbol{N}$. The Radon–Nikodym derivative (15) may be computed at the end (which is more efficient when $f(\hat{\boldsymbol{N}}) = 0$ often) per (18) or during the simulation to reduce storage. The distribution in Step 6 matches that of (16). Step 7 records the component of $\hat{\boldsymbol{N}}$ (drawn in Step 6) for which an event has occurred.

Although Step 4 is relatively straightforward to implement, there is a significant degree of subtlety involved. In this step, we sample from the $(\mathrm{Q}, \mathcal{F}_{S_{\ell-1}})$-conditional joint density of $(\sum_{i\in\chi} H^i, \boldsymbol{h}_-)$ at the random time S_ℓ, where $i \in \chi$ if and only if $N^i_{S_\ell} = 0$. This sampling procedure is typically very challenging under P because S_ℓ and $(\sum_{i\in\chi} H^i, \boldsymbol{h}_-)$ are dependent (conditionally on $\mathcal{F}_{S_{\ell-1}}$). Indeed, H^i and h^i are the P-compensator and P-intensity of N^i for $i \in \chi$ prior to arrival time S_ℓ. Under Q, this (conditional) dependence vanishes.

The sampling of $(\sum_{i\in\chi} H^i_{S_\ell}, \boldsymbol{h}_{S_\ell -})$ under Q may be performed in several ways. The most straightforward approach involves the sampling of the skeleton of $\boldsymbol{h}$ on the discretized interval $[S_{\ell-1}, S_\ell)$ and assembling R_ℓ in (17) by numerical integration. For instance, constructing a time grid $S_{\ell-1} = s_0 < s_1 < \cdots < s_d = S_\ell$, we obtain the samples $\{\hat{h}^i_{s_j}\}_{j=1}^d$ for each $i \in \chi$ by some discretization or exact method. Then, $\hat{h}^i_{s_d}$ is a sample of $h^i_{S_\ell -}$, and by the trapezoid rule,
$$R_\ell = \sum_{i\in\chi} \sum_{j=1}^{d} \frac{(s_j - s_{j-1})}{2} (\hat{h}^i_{s_j} + \hat{h}^i_{s_{j-1}}). \tag{19}$$

This is analogous to the numerical integration of the compensator in the time-scaling scheme (Algorithm 1), but with the notable exception that the termination point s_d is given and does not have to be simultaneously estimated. To highlight the similarities further, we note that Algorithm 1 may be implemented using the P-compensator of C together with distribution (16) (cf. footnote 8).

4.3. Bounding the Bias

The plain JAM involves two potential sources of bias. The first pertains to the samples of $f(\boldsymbol{N})$. When the skeleton of $\boldsymbol{h}$ is not sampled exactly (under Q), the samples of $f(\boldsymbol{N})$ will be biased. A second source of bias pertains to the

Downloaded from informs.org by [2001:638:807:212:a8af:506c:25a1:b254] on 09 November 2026, at 05:59 . For personal use only, all rights reserved.

Radon–Nikodym derivative in (15). This requires the numerical integration of the Q-sample paths of $\boldsymbol{h}$. Even if the skeleton of $\boldsymbol{h}$ is sampled exactly, its numerical integration introduces bias into the JAM estimator.

The sources of bias for the plain JAM algorithm mirror the first two sources of bias for the time-scaling scheme discussed in Section 2.3. However, the plain JAM algorithm does not suffer from the dominating event time location bias of time scaling. This is because after the change of measure, the inversion of the compensator of the event count C is exact. Moreover, higher-order discretization, unbiased methods (Chen et al. [12]), or exact methods (Blanchet and Zhang [6]) can further mitigate the bias of the JAM estimator.

Without assuming a specific intensity model, it seems difficult to quantify the bias of the plain JAM estimator. However, we can get some insight into the bias due to numerical integration. Suppose that the samples of the skeleton of $\boldsymbol{p}$ are exact. Then the Q-samples of $f(\boldsymbol{N})$ are exact. The integration bias given by

$$\mathscr{E} = \sum_{i=1}^{n} \left(\int_0^\infty p_s^i \mathrm{d}s - \mathscr{D}(p^i) \right),$$

where $\mathscr{D}(p^i)$ is the quadrature approximation that numerically integrates the skeleton of a path p^i over $[0, \tau_i)$. We have the following result.

Proposition 4.1. *Suppose the skeleton of the process $\boldsymbol{h}$ is sampled exactly under* Q. *Fix $\rho \in [1, \infty)$, and for $X = |\mathscr{E}| / E(|\mathscr{E}|)$ define $\Lambda(\theta) = \log \mathrm{E}(e^{\theta X})$ for $\theta \in \mathbb{R}$. Then, if $\mathrm{E}(|\mathscr{E}|) \le \delta/4$, for $x = \Lambda(\delta\rho)/(\delta\rho)$, the bias of Algorithm 2 satisfies*

$$|\,\mathrm{bias}\,| \le \|f\|_\mu \left(\frac{e^{\delta x} - 1}{\delta x} + 4/\delta \right) \mathrm{E}(|\mathscr{E}|^\rho)^{1/\rho}$$

where $\|f\|_\mu = \mathrm{E}(|f(\boldsymbol{N})|^\mu)^{1/\mu}$ must have $1/\rho + 1/\mu = 1$.

The proof is in Appendix B. The error $\mathrm{E}(|\mathscr{E}|^\rho)$ of the intensity integral may be improved through higher-order methods. In contrast, the time-scaling scheme with a discretization spacing Δ_g retains the $O(\Delta_g)$ bias (see Section 2.3).

5. Exact Algorithm

We show that under some conditions, the bias of the JAM estimator can be completely eliminated. We keep the technical assumptions on $\boldsymbol{h}$ of Section 4.

5.1. Exact Sampling with a Bridge Transform

We develop an exact algorithm that avoids sampling the integral $\boldsymbol{H}$. Instead, assume a Markovian setting in which exact samples of a skeleton of $\boldsymbol{h}$ may be obtained and the bridge transform

$$\phi^\chi_{(t,x)}(s,z) = \mathrm{E}_{\mathrm{P}_0}\left(e^{-\sum_{i\in\chi}(H_t^i - H_s^i)} \mid \boldsymbol{h}_t = x, \boldsymbol{h}_s = z \right) \tag{20}$$

may be evaluated for every $\chi \subseteq \{1, \ldots, n\}$ and admissible (t, x) and (s, z). Here, probability P_0 of Corollary 3.3 with all $\mathcal{T}_i = [s, t]$ facilitates the computation by ensuring no jumps in the $\{N^i\}$ on $[s, t]$. Such transforms of the bridge of $\boldsymbol{h}$ are available for a wide class of models including Lévy processes, Bessel diffusions (Pitman and Yor [43]), and affine processes (Kang and Kang [33]).

The algorithm is, essentially, a conditional Monte Carlo scheme. Constructing Q from $u^i = \bar{u}$ in (13), we consider the Q-estimator

$$Y^{\mathcal{G}} = f(\boldsymbol{N}) \mathrm{E}_{\mathrm{Q}}(L_\infty \mid \mathcal{G}) \tag{21}$$

for a σ-algebra $\mathcal{G}$ such that $f(\boldsymbol{N})$ is $\mathcal{G}$-measurable. Then, we have the relation $y = \mathrm{E}_{\mathrm{Q}}(Y^{\mathcal{G}})$ as required by (2) and the Radon–Nikodym derivative

$$L_\infty^{\mathcal{G}} = \mathrm{E}_{\mathrm{Q}}(L_\infty \mid \mathcal{G})$$

may be viewed as a filtered likelihood given the observations in $\mathcal{G}$. We construct $\mathcal{G}$ as follows.[23] Take $\{\eta_\ell\}_{\ell=1}^n$ (see Algorithm 2) such that each η_ℓ has the $\mathcal{F}_{S_{\ell-1}}$-conditional distribution of an exponential random variable of rate $\lambda_{\ell-1}$ in (12).

Let $T_\ell = \sum_{k=1}^{\ell} \eta_k$ and define

$$\mathcal{G} = \sigma(\{T_\ell, \boldsymbol{h}_{T_\ell -}, I_\ell, \boldsymbol{h}_{T_\ell}\}_{\ell=0}^n). \tag{22}$$

The variables in $\mathcal{G}$ are sufficient in order to assemble $f(\boldsymbol{N})$ as in Algorithm 2. Given T_ℓ, a sample of $\boldsymbol{h}_{T_\ell}$ and the $\{I_\ell\}$

Downloaded from informs.org by [2001:638:807:212:a8af:506c:25a1:b254] on 09 November 2026, at 05:59 . For personal use only, all rights reserved.

yield a sample of λ_ℓ in (12). The next event occurs in time $\eta_{\ell+1}$, exponentially distributed with rate λ_ℓ. The conditional distribution (16) identifies which event has occurred, as in Algorithm 2.

Note that $T_\ell = S_\ell$ and $\eta_\ell = S_\ell - S_{\ell-1}$ under Q but not under P by construction. Here, one can take either the nominal $(\mathrm{Q},\mathbb{F})$-intensity $\boldsymbol{q} = \bar{u}\boldsymbol{p}$ or consider a change of filtration to events restricted to $\mathcal{G}$. The two approaches are equivalent from the simulation perspective as highlighted by the fact that each λ_ℓ is $\mathcal{G}$-measurable. The Radon–Nikodym derivative L_∞, however, is not $\mathcal{G}$-measurable in general, and $L_\infty^{\mathcal{G}}$ must be computed for the scheme to apply.

For $1 \le \ell \le n$, we let

$$G_w^{\ell-1} = \left(\frac{\sum_{k=\ell}^n h^{I_k}_{T_\ell-}}{\lambda_{\ell-1}}\right) e^{\lambda_{\ell-1} w}, \tag{23}$$

which is $\mathcal{G}$-measurable so that $L_\infty^{\mathcal{G}}$ (see (15)) may be written in the form

$$L_\infty^{\mathcal{G}} = \mathrm{E}_{\mathrm{Q}}\left(e^{-\sum_{i=1}^n \int_0^\infty p^i_s \mathrm{d}s} \mid \mathcal{G}\right) \prod_{\ell=1}^n G^{\ell-1}_{\eta_\ell}. \tag{24}$$

In general, we do not have an explicit formula for the conditional expectation in (24). However, in a Markovian setting a formula for $L_\infty^{\mathcal{G}}$ can be derived. The corresponding scheme, which we refer to as the exact JAM, is provided in Algorithm 3.

Theorem 5.1. *For every $0 \le \ell \le n$ and $s \ge 0$, suppose $\boldsymbol{h}$ satisfies the strong Markov property at S_ℓ over $\mathcal{T} = [S_\ell, S_\ell + s)$ under P_0 of Corollary 3.3 with every $\mathcal{T}_i = \mathcal{T}$. Assume the transform ϕ in (20) may be evaluated for every subset $\chi \subseteq \{1,\dots,n\}$ and admissible (t,x) and (s,z). Then, $L_\infty^{\mathcal{G}}$ in (24) is given by*

$$L_\infty^{\mathcal{G}} = \prod_{\ell=1}^n G^{\ell-1}_{\eta_\ell} \phi^{\{I_k\}_{k=\ell}^n}_{(S_{\ell-1},\lambda_{\ell-1})}(S_\ell, \boldsymbol{h}_{S_\ell-}). \tag{25}$$

Moreover, whenever the Q*-skeleton of $\boldsymbol{h}$ may be sampled exactly, Algorithm 3 generates exact samples of the* Q*-estimator $Y^{\mathcal{G}}$ in (21).*

The proof of Theorem 5.1 is deferred to Appendix C. Equation (25) is derived from (24) by iteratively conditioning over intervals of the form $[S_\ell, S_{\ell+1})$. The Markov property is required for the conditional independence of past and future events in $\mathcal{G}$ given the present. Note that the Q-skeleton of $\boldsymbol{h}$ is significantly easier to sample than its P analog. We require exact samples of $\boldsymbol{h}$ at the successive times $\{S_\ell\}$ between which no jumps in the $\{N^i\}$ occur. However, under Q, $S_\ell = T_\ell$, and so $\{\boldsymbol{h}_{S_{\ell-1}+\cdot}\}$ is $\mathcal{F}_{S_{\ell-1}}$-conditionally independent of S_ℓ.

Remark 5.2. Recall that $\{I_k\}_{k=\ell}^n$ in (25) is the $\mathcal{F}_{S_{\ell-1}}$-measurable collection of components of $\mathbf{N}$ at which no event has occurred by time $S_{\ell-1}$. The strong Markov property is checked under P_0 over a time intervals on which no jumps in the components of $\boldsymbol{N}$ occur. Here, $\boldsymbol{h}$ may be taken as regular Markov (Dynkin [22]), a class that includes both homogeneous and inhomogeneous Markov processes. The strong Markov property for this setting may be found in Dynkin [22] and Kuznetsov [35]. The extension to the (time-)inhomogeneous case is treated by considering the homogeneous (space–time) process $\tilde{\boldsymbol{h}}$ for which $\tilde{\boldsymbol{h}}_s = (s, \boldsymbol{h}_s)$.

Algorithm 3 (Exact JAM; Generates a Sample $\hat{Y}$ of $Y^{\mathcal{G}} = f(\boldsymbol{N})L_\infty^{\mathcal{G}}$ with $L_\infty^{\mathcal{G}}$ Given by (25) Under Q)

Initialize $\chi = \{1,\dots,n\}$ and $\ell = 0$.

1. Generate samples of the jump sizes $\{\Delta h^i_{S_\ell}\}_{i\in\chi}$.
2. Draw an exponential random variable $\eta_{\ell+1}$ with rate

$$\hat{\lambda}_\ell = \sum_{i\in\chi} \hat{h}^i_{S_\ell-} + \Delta \hat{h}^i_{S_\ell}. \tag{26}$$

3. Increment $\ell \leftarrow \ell + 1$ and set $\hat{S}_{\ell_i} = \hat{S}_{\ell-1} + \eta_\ell$.
4. Sample $\boldsymbol{h}_{S_\ell-}$ and set $r_\ell = \sum_{i\in\chi} \hat{h}_{S_\ell-}$.
5. If $\ell = n$ (or some other stopping condition $\star$) then **return**

$$\hat{Y} \quad = f(\hat{\boldsymbol{N}}) \prod_{k=0}^{\ell} \left(\frac{r_k}{\hat{\lambda}_k}\right) e^{-\hat{\lambda}_k \eta_k} \phi^{\{\hat{I}_j\}_{j=k}^n}_{(\hat{S}_{k-1},\hat{\lambda}_{k-1})}(\hat{S}_k, \hat{\boldsymbol{h}}_{S_k-}). \tag{27}$$

Downloaded from informs.org by [2001:638:807:212:a8af:506c:25a1:b254] on 09 November 2026, at 05:59 . For personal use only, all rights reserved.

6. Draw index $\hat{I}_\ell$ from the (conditional) distribution

$$\left\{\hat{h}^i_{S_{\ell-}}/r_\ell\right\}_{i\in\chi}.$$

7. Set $\hat{N}^{\hat{I}_\ell}_s = \mathbb{1}_{\{S_\ell \le s\}}$ for $s \ge 0$ and $\chi \leftarrow \chi\backslash\{\hat{I}_\ell\}$. Go to Step 1.

Notes. Terminating at some stopping time T implies we take the restriction of $\mathcal{G}$ to $[0,T]$. Then $\hat{S}_n$ is replaced by T and corresponding adjustments to (27) are made (i.e., each $u^i_s = 1$ for $s > T$).

Notes. $\hat{X}$ refers to a sample of a random variable X.

5.2. Exact Sampling Without a Bridge Transform

An alternative algorithm exists when the bridge transform (20) is not available by drawing samples from the law

$$\mathscr{L}_{t,s,\chi}(\cdot) = \mathrm{P}_0\left(\left(\textstyle\sum_{i\in\chi}(H^i_t - H^i_s), \boldsymbol{h}_t\right) \in \cdot \mid \mathscr{F}_s\right), \tag{28}$$

where $0 \le s < t$ and measure P_0 of Corollary 3.3 with every $\mathcal{T}_i = [s,t]$. Given such samples, we follow the plain JAM Algorithm 2 but avoid the discretization of the interval $[S_{\ell-1}, S_\ell)$ in the implementation of Step 4. When the samples obtained in this step are exact, then clearly so is the simulation estimator.

No events occur under P_0 during $[s,t]$, simplifying the computations by bypassing the event times $\{\tau_i\}_{i\in\chi}$ in the evaluation of (28). In the context of Step 4 of Algorithm 2, the events in χ^c have already occurred. The joint law (28) is computable (semi)analytically for many models of interest (e.g., Cheng and Scaillet [14], Duffie et al. [21]).

To prove correctness of the scheme, observe that for Borel $\mathscr{B} \subseteq \mathbb{R}_+ \times \mathbb{R}^n_+$,

$$\begin{aligned}\mathscr{L}_{t,s,\chi}(\mathscr{B}) &= \mathrm{E}_\mathrm{P}\left((Z_t/Z_s)\mathbb{1}_{\left\{\left(\sum_{i\in\chi}(H^i_t - H^i_s), \boldsymbol{h}_t\right)\in\mathscr{B}, C_t - C_s = 0\right\}} \mid \mathscr{F}_s\right)\\ &= \mathrm{E}_\mathrm{Q}\left(\mathbb{1}_{\left\{\left(\sum_{i\in\chi}(H^i_t - H^i_s), \boldsymbol{h}_t\right)\in\mathscr{B}, C_t - C_s = 0\right\}} \mid \mathscr{F}_s\right)\end{aligned}$$

by identity (11). Some attractive properties of measure Q facilitate the rest of the analysis. By construction of Algorithm 2, we have $S_\ell = S_{\ell-1} + \eta_\ell$, and because $\left(\sum_{i\in\chi}(H^i - H^i_{S_{\ell-1}}), \boldsymbol{h}\right)$ and η_ℓ are $\mathcal{F}_{S_{\ell-1}}$-conditionally Q-independent,

$$\begin{aligned}\mathscr{L}_{t,s,\chi}(\mathscr{B}) = \mathrm{Q}\Big(&\left(\textstyle\sum_{i\in\chi}(H^i_{S_\ell} - H^i_{S_{\ell-1}}), \boldsymbol{h}_{S_{\ell-}}\right) \in \mathscr{B} \mid \mathscr{F}_{S_{\ell-1}} \vee \sigma(\eta_\ell)\Big),\\ &S_{\ell-1} = s,\\ &S_{\ell-1} + \eta_\ell = t = S_\ell\end{aligned}$$

Q-almost surely for all Borel $\mathscr{B}$. This justifies the use of P_0 in (28) for obtaining exact samples of (R_ℓ, r_ℓ) at the interarrival times $\{S_\ell\}_{\ell=0}^n$ in Step 4 of Algorithm 2 (a procedure far more complicated under measure P). The law $\mathscr{L}_{t,s,\chi}$ may be sampled by acceptance/rejection or by using the method of inversion.

6. Numerical Study

Credit risk management often requires the computation of the distribution of defaults in a portfolio of positions such as loans, corporate bonds, and other securities exposed to default risk. In this setting, the stopping times $\{\tau_i\}$ are the default times of the individual positions in the portfolio. We test our algorithms in this application context and compare them to the time-scaling scheme.

6.1. Model Specification and Methods

A popular and computationally challenging intensity specification used in credit risk is a factor model with contagion. For risk factors $\{X^k\}_{k=0}^n$ and nonnegative parameters $\{w_i\}_{i=1}^n$, we take

$$p^i = h^i(1 - N^i), \tag{29}$$

$$h^i = w_i X^0 + X^i. \tag{30}$$

Each risk factor X^k is taken to be a solution to the SDE

$$\mathrm{d}X^k_s = \kappa_k(x_k - X^k_s)\mathrm{d}s + \sigma_k\sqrt{X^k_s}\mathrm{d}B^k_s + \sum_{i\ne k}\delta_{ki}\mathrm{d}N^i_s, \quad X^k_0 > 0. \tag{31}$$

Here, $(B^0, B^1, \ldots, B^n)$ is a standard Brownian motion. Parameter $x_k > 0$ represents factor k's long-run mean, $\kappa_k > 0$ is factor k's mean-reversion rate, and $\sigma_k > 0$ is factor k's diffusive volatility. Last, the parameter $\delta_{ki} \ge 0$ describes risk factor k's sensitivity to default in position i of the portfolio. We further let $\delta_{0i} = 0$ for all $1 \le i \le n$, which is a standard modeling assumption.

Downloaded from informs.org by [2001:638:807:212:a8af:506c:25a1:b254] on 09 November 2026, at 05:59 . For personal use only, all rights reserved.

Model (29)–(31) and its special cases are analyzed by Giesecke et al. [27] and many others. It addresses several of the sources of default clustering identified in empirical research (Das et al. [18] and Azizpour et al. [2]). Factor X^0 represents a source of systematic risk to which all the assets are exposed. All individual intensities are correlated through this common factor, and the parameters $\{w_i\}$ govern the magnitude of that correlation. The jump term $\delta_{ki}\mathrm{d}N^i$ generates feedback; a default in position i increases the intensity of position k by a sensitivity δ_{ki}. This feedback models the contagion effects that are widely observed in credit markets.

Model (29)–(31) is difficult to analyze and simulate using existing methods. When the $\{\delta_{ki}\}$ are positive, it is not known how to efficiently compute the joint distribution of the $\{\tau_i\}$ analytically (see Section 2.2). Because the intensities are not bounded, an exact simulation scheme based on thinning (Lewis and Shedler [38], Ogata [42]) cannot be applied. The exact Markov chain projection and simulation algorithm of Giesecke et al. [28] applies. However, the presence of the factor X^0 renders the computation of the intensity projection intractable. The exact and asymptotically exact schemes of Giesecke et al. [27] may be applied. However, the implementation is not trivial, for example, the exact scheme requires sampling from a complicated hitting-time distribution by the inverse method. These complications stem from the fact that the interarrival intensity must be sampled at the time of an upcoming default (these two random variables are dependent). Similar considerations also apply to the exact scheme of Giesecke and Kim [23], which may be extended to the model (29)–(31).

Time scaling (Algorithm 1) applies to the model because each risk factor X_k may be simulated. In between any two successive default times, each X_k is a CIR (Cox-Ingersoll-Ross) process Cox et al. [16] which has a well-known transition law in terms of a noncentral chi-square distribution (Glasserman [30, section 3.4]). At the default times, the intensities are updated with the corresponding jump sizes $\{\delta_{ki}\}$. Thus, exact samples of the skeleton of the intensity may be obtained, and the bias is due only to the approximation of the default times and the numerical integration that are required by the scheme (Section 2.3).

The plain JAM is implementable for model (29)–(31) because time scaling is feasible. The method (Algorithm 2) has virtually the same running time as a comparable implementation of time scaling.[24] Indeed, the computations required by these schemes are nearly identical. CIR risk factors may be sampled exactly in between the default times as for time scaling. The Q-samples of the $\{\tau_i\}$ are exact, whereas the final simulation estimator is biased only by the numerical integration present in the Radon–Nikodym derivative. Thus, we expect a reduction in the bias of the plain JAM estimator relative to that of time scaling.

The exact JAM (Algorithm 3) also applies because the bridge transform (20) is analytically tractable for a CIR risk factor. Note that no sampling from this transform is needed; only its evaluation is required. This leads to significant efficiency gains in contrast to the use of this transform by the other sampling methods in the aforementioned papers. For similar reasons, this method is also a better alternative to the exact sampling scheme supplied in Section 5.2.[25]

6.2. Computing the Probability of No Defaults

Few closed-form formulae (for even moderate n) are known for model (29)–(31). An exception is the probability that no defaults occur by some time $T > 0$, that is, $f(z) = \mathbb{1}_{\{\sum_{i=1}^n z_i = 0\}}$ in (2),

$$y = \mathrm{P}(C_T = 0) \tag{32}$$

The computation is facilitated by using probability P_0 of Corollary 3.3 (with $\mathscr{T}_i = [0, T]$), which yields $\mathrm{P}(\cap_{i\in\chi}\{N_T^i = 0\}) = \mathrm{E}_{\mathrm{P}_0}(e^{-\sum_{i\in\chi}H_T^i})$. Then,

$$y = \psi_0\Big(\sum_{i=1}^n w_i\Big)\prod_{i=1}^n \psi_i(1), \tag{33}$$

where ψ_k is the Laplace transform $\psi_k(v) = \mathrm{E}_{\mathrm{P}_0}(e^{-v\int_0^T X_s^k \mathrm{d}s})$. Evaluation of the transform is straightforward. Because a Brownian motion is a P-local martingale without jumps, it is also a Q-local martingale (see (10)). Thus, the P_0-dynamics of a risk factor X^k on $[0, T]$ are those of (31) but with every $\delta_{ki} = 0$. The risk factors $\{X^k\}_0^n$ are mutually P_0-independent on $\mathscr{F}_T$, and thus, the product form in (33) is justified. To perform computations under P, define $\tilde{X}^k$ as

$$\mathrm{d}\tilde{X}_s^k = \kappa_k\Big(x_k - \tilde{X}_s^k\Big)\mathrm{d}s + \sigma_k\sqrt{\tilde{X}_s^k}\,\mathrm{d}W_s^k, \quad \tilde{X}_0^k > 0 \tag{34}$$

Downloaded from informs.org by [2001:638:807:212:a8af:506c:25a1:b254] on 09 November 2026, at 05:59 . For personal use only, all rights reserved.

on $[0,T]$, where W^k is a standard P-Brownian motion. The P-law of $\tilde{X}^k$ coincides with the P_0-law of X^k on $\mathcal{F}_T$. The transform ψ_k may now be computed using a well-known closed formula in Cox et al. [16] for the transform

$$\psi_k(v) = \mathrm{E}(e^{-v\int_0^T \tilde{X}_s^k \mathrm{d}s}). \tag{35}$$

We denote by $\hat{y}_{\text{exact}}$ the value of y computed using (33) and (35).

6.3. Model Parameters and Implementation Details

The parameters of the model (29)–(31) are selected as follows. We consider a portfolio of size $n = 100$ with time horizon $T = 1$. Let $U(a, b)$ be the uniform distribution on $[a,b]$. For each position $k = 1,\ldots,n$, we draw κ_k from $U(0.5,1.5)$ and x_k from $U(0.005,0.025)$, and set σ_k to $\min\{\sqrt{2x_k\kappa_k}, \bar{\sigma}_k\}$, where each $\bar{\sigma}_k$ is drawn from $U(0.1,0.2)$. The restriction on σ_k ensures the Feller condition is satisfied which guarantees X^k is strictly positive almost surely. We initialize each X^k at x_k times a sample from $U(0.5,1.25)$. Jump sensitivities are constructed by drawing each δ_{ki} from $U(0,0.1)$, and each factor sensitivity, $w_{k'}$ is drawn from $U(0,0.02)$. For the systematic factor X^0, we take $x_0^0 = 0.0119582202766817646$, $\kappa_0 = 1$, $\sigma_0 = 0.2$ and initialize X^0 as x_0 times a sample from $U(0.5,1.25)$. Our choice of x_0 ensures $\hat{y}_{\text{exact}} = 0.16$ to machine precision. These parameters model a high credit quality portfolio but, in practice, they would be calibrated to data.

The number of MC trials is denoted by m and the estimate of y by $\hat{y}_m$ (see Section 2.2). The central limit theorem is used to construct all confidence intervals; that is, for $z \in (0,1)$ and m large, $y \in [\hat{y}_m - \zeta_m, \hat{y}_m + \zeta_m]$, where

$$\zeta_m = \mathcal{N}^{-1}\left(\frac{1+z}{2}\right)\hat{\sigma}_{\hat{Y}}/\sqrt{m} \tag{36}$$

with $100z\%$ confidence. Here, $\mathcal{N}$ denotes the standard normal distribution function, and $\hat{\sigma}^2_{\hat{Y}}$ denotes the sample variance of the estimator Y.

A uniform grid spacing Δ_g is used to discretize the intensity for time scaling and the plain JAM (the exact JAM does not discretize). Both time scaling and the plain JAM utilize the trapezoid rule (see (19)) to compute integrals of the intensity process. Time scaling employs a linear interpolation of $\boldsymbol{h}$ for the default times (see Section 2.3). The interevent CIR risk factors (34) are sampled exactly on the time grid (Glasserman [30, section 3.4]). The closed form for the bridge transform in (20), used for the exact JAM, is found in Broadie and Kaya [9].

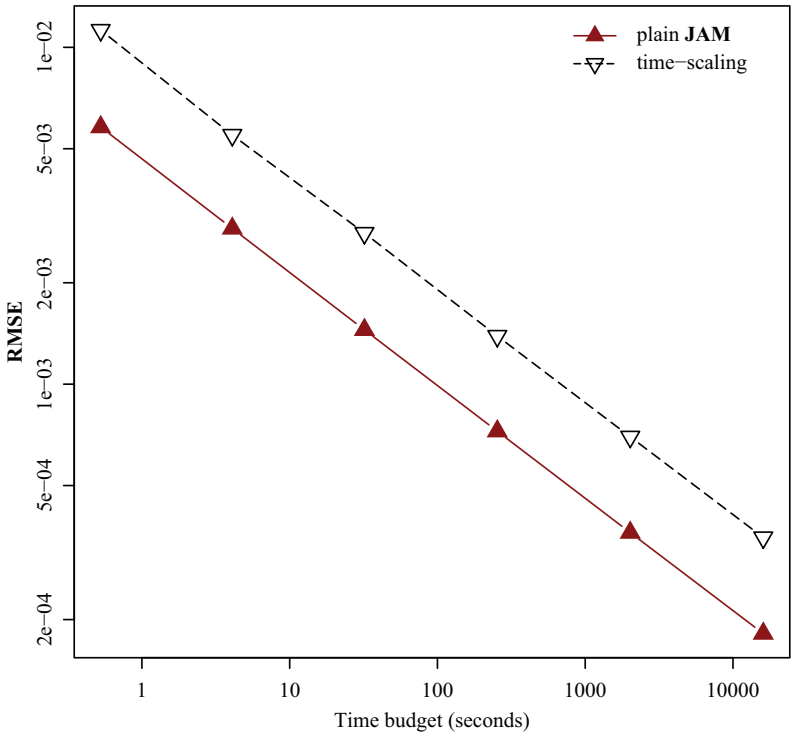
Table 1. Comparison of the performance of the (plain/exact) JAM and time scaling for estimating y in (32).

	$\lvert\hat{y}_{\text{exact}} - \hat{y}_m\rvert$			
	99%-conf.			# trials (m)
Δ_g	Time scaling	Plain JAM	Exact JAM	
$\frac{1}{2^5}$	0.009597	0.000041		40,448,520
	0.000152	**0.000151**		
$\frac{1}{2^6}$	0.004670	0.000038		20,825,646
	0.00010	**0.000210**		
$\frac{1}{2^7}$	0.002418	0.000102		10,625,031
	0.000292	**0.000294**		
$\frac{1}{2^8}$	0.001169	0.000025		5,361,749
	0.000410	**0.000413**		
$\frac{1}{2^9}$	0.000617	0.000392		2,702,421
	0.000576	**0.000583**		
$\frac{1}{2^{10}}$	0.000847	0.000013		1,359,483
	0.000813	**0.000821**		
Not applicable			0.000026	10^8
			0.000093	

Notes. The analytically computed value is $\hat{y}_{\text{exact}} = 0.16$. The computational budget time for all methods was 4,096 seconds. The column Δ_g reports the (uniform) grid spacing utilized. The second, third, and fourth columns report the estimate error $\lvert\hat{y}_{\text{exact}} - \hat{y}_m\rvert$ (the estimate of $\lvert\text{bias}\rvert$) for each discretization spacing Δ_g. The numbers in bold provide the margin of error ζ_m computed from (36). If the margin is smaller than $\lvert\hat{y}_{\text{exact}} - \hat{y}_m\rvert$, then $\hat{y}_{\text{exact}}$ is not inside the computed 99% confidence interval (99%-conf.). As the number of trials for both schemes are nearly identical, the fifth column reports them for the time-scaling scheme only.

Downloaded from informs.org by [2001:638:807:212:a8af:506c:25a1:b254] on 09 November 2026, at 05:59 . For personal use only, all rights reserved.

Figure 2. (Color online) A log-log plot of the RMSE versus the computational budget (time) allowed for the simulation. From left to right, the triangle markers correspond to discretization size Δ_g in Table 2. The number of trials m in the computation of the RMSE at each point may be recovered via the relation $\Delta_g = 2/\sqrt{m}$. However, the estimates of the bias and variance $\sigma^2_{\hat{Y}}$ are computed with 10^9 trials as in Table 2.



6.4. Numerical Results

Table 1 compares the performance of time scaling, the plain JAM, and the exact JAM. It reports the absolute differences between the computed estimates and $\hat{y}_{\text{exact}}$ with 99% confidence intervals. Time-scaling confidence intervals miss the analytically computed value $\hat{y}_{\text{exact}}$ for all six discretization sizes Δ_g. Plain JAM estimates always include $\hat{y}_{\text{exact}}$ within the confidence interval. The exact JAM estimate deviates from $\hat{y}_{\text{exact}}$ only in the fifth significant digit.

The performance of the estimator is well summarized by the root-mean-squared error (RMSE), the square root of (3). The statistic balances the variance of the estimator and its bias. To reduce the bias one could, at the expense of computational effort, refine the discretization Δ_g. Figure 2 illustrates the RMSE estimates for time scaling and the plain JAM when $\Delta_g = 2/\sqrt{m}$. The run times for the two methods are virtually identical, and averages are

Table 2. Time-scaling and plain JAM estimates of the simulation bias and variance $\sigma^2_{\hat{Y}}$ in (3).

	$(\widehat{\text{bias}} \mid \hat{\sigma}^2_{\hat{Y}})$				Time/10^4
Δ_g	Time scaling		Plain JAM		Trials (sec)
$\frac{1}{2^5}$	−0.009605	**0.140894**	−0.000029	**0.137683**	1.285
$\frac{1}{2^6}$	−0.004665	**0.137546**	−0.000028	**0.137688**	2.496
$\frac{1}{2^7}$	−0.002411	**0.136134**	−0.000019	**0.1376832**	4.892
$\frac{1}{2^8}$	−0.001190	**0.1350705**	−0.000011	**0.1376565**	9.695
$\frac{1}{2^9}$	−0.000598	**0.135045**	−0.000007	**0.1376544**	19.23
$\frac{1}{2^{10}}$	−0.000301	**0.135035**	−0.000015	**0.137632**	38.24

Notes. All estimates are computed with $m = 10^9$ trials. Here, $\widehat{\text{bias}} = \hat{y}_{\text{exact}} - \hat{y}_m$, and $\sigma^2_{\hat{Y}}$ (bold) denotes the sample variance. The analytically computed probability is $\hat{y}_{\text{exact}} = 0.16$ to machine precision. For time scaling, $\sigma^2_{\hat{Y}}$ may be computed exactly via $\hat{y}_{\text{exact}}(1 - \hat{y}_{\text{exact}})$.

Downloaded from informs.org by [2001:638:807:212:a8af:506c:25a1:b254] on 09 November 2026, at 05:59 . For personal use only, all rights reserved.

reported in Table 2. Estimates of the bias and variance for 10^9 trials are reported for each discretization. Figure 2 demonstrates that the plain JAM systematically reduces the bias component of the RMSE relative to that of time scaling.

Although the run times of time scaling and the plain JAM are similar as expected (see further discussion in Section 6.1), the exact JAM is orders of magnitude faster. Table 1 reports that the number of trials completed by the exact JAM (within the given time budget) is larger by a factor of 70 than that of time scaling with the uniform spacing $\Delta_g \approx 0.001$. This gain in performance is proportional to Δ_g^{-1} because the bridge transform evaluation does not depend on Δ_g.

Acknowledgments

The authors thank Lisa Goldberg, Baeho Kim, two anonymous referees, and participants of the 2015 Risk Seminar at the Center for Risk Management Research in Berkeley, California; the 2016 Joint Mathematics Meeting in Seattle, Washington; and the 2017 Annual INFORMS Meeting in Houston, Texas, for helpful comments on an earlier version of this paper.

Appendix A. Proof of Theorem 3.1

For a stopping time θ, denote by X^θ an X stopped at θ, that is, $X_s^\theta = X_{s\wedge\theta}$.

We let $Z(\ell)$ denote the Z in (9) but with $u^i = 1$ for all $i > \ell$. We proceed by induction on $\ell \le n$. The case $\ell = 0$ is trivial (i.e., $Z_\infty(0) = 1$). Suppose $Z(\ell)$ defines a probability measure $\mathrm{Q}_\ell = Z_\infty(\ell)\mathrm{P}$ and N^i admits the Q_ℓ-intensity $q^i = u^i p^i$ for $i \le \ell$ and p^i for $i > \ell$. We augment the $\mathbb{F}$ by the null sets of Q_ℓ (see Remark 3.5). We then construct $\mathrm{Q}_{\ell+1}$ by taking any $u^{\ell+1}$ satisfying our hypotheses. To ease the notation, we drop the sub- and superscript $(\ell+1)$ from the $\tau_{\ell+1}$, $\xi_{\ell+1}$, $u^{\ell+1}$, $h^{\ell+1}$, $H^{\ell+1}$, $\mathrm{Q}^{\ell+1}$, $p^{\ell+1}$, $N^{\ell+1}$, $M^{\ell+1}$ and $\mathbb{H}_{\ell+1}$ of Section 2.1 (i.e., consider a single stopping time τ on $(\Omega, \mathcal{F}, \mathrm{Q}_\ell)$).

For $0 \le t \le \infty$, let Z be the càdlàg, $\mathbb{F}$-adapted process[26] defined by

$$Z_t = \exp\left(\int_0^t (1-u_s)p_s \mathrm{d}s + N_t \log(u_\tau)\right).$$

If $\mathrm{E}_{\mathrm{Q}_\ell}(Z_\infty) = 1$, we may define $\mathrm{Q}_{\ell+1} = Z_\infty \mathrm{Q}_\ell$. And furthermore, if N^i admits a $\mathrm{Q}_{\ell+1}$-intensity $q = up$, the claim extends by induction to all n stopping times.

We first show that $\mathrm{E}_{\mathrm{Q}_\ell}(Z_\infty) = 1$. For $k \in \mathbb{N}$ define the $\mathbb{H}$-stopping time

$$\theta_k = \inf\{t \ge 0 : (H+\mathrm{Q})_t \ge k\}. \tag{A.1}$$

It is not difficult to show that $\{\theta_k\}_{k\in\mathbb{N}}$ is a sequence of localization times[27] for the local $(\mathrm{Q}_\ell, \mathbb{F})$-martingale Z. Indeed, by integration by parts, we have

$$Z_t = 1 + \int_0^t Z_{s-}(u_s - 1)\mathrm{d}M_s, \tag{A.2}$$

and consequently Z is a local $(\mathrm{Q}_\ell, \mathbb{F})$-martingale (Protter [45, theorem IV.29]). It is easy to verify (because $M = N - A$ is a $(\mathrm{Q}_\ell, \mathbb{F})$-martingale) that

$$\mathrm{E}(\sup_{t\ge 0} Z_t^{\theta_k}) \le 1 + \mathrm{E}\left(\int_0^{\theta_k} Z_{s-}(1+u_s)p_s \, \mathrm{d}s\right).$$

Note that $p_s Z_{s-} \le h_s e^{(H+\mathrm{Q})_s}$ because $p_s = h_s \mathbb{1}_{\{s<\tau\}}$ and $N_{\tau-} = 0$. By the definition of θ_k, we have that $(H+\mathrm{Q})_s \le k$ for all $s \le \theta_k \le \infty$, and therefore

$$\mathrm{E}\left(\int_0^{\theta_k} Z_{s-}(1+u_s)p_s \, \mathrm{d}s\right) \le k e^k.$$

Therefore, Z^{θ_k} is a unit-mean uniformly integrable $(\mathrm{Q}_\ell, \mathbb{F})$-martingale (Protter [45, theorem I.51]), and we may define the measure $\mathrm{Q}_{\ell+1}^k = Z_\infty^{\theta_k} \mathrm{Q}_\ell$. To prove the existence of the limiting probability $\mathrm{Q}_{\ell+1} = \mathrm{Q}_{\ell+1}^\infty$, it suffices show that $\{Z_{\theta_k}\}_{k\in\mathbb{N}}$ (note, $Z_\infty^{\theta_k} = Z_{\theta_k}$) is uniformly integrable. Because $\mathrm{E}_{\mathrm{Q}_\ell}(Z_{\theta_k}\mathbb{1}_{\mathcal{A}}) = \mathrm{Q}_{\ell+1}^k(\mathcal{A}) \le 1$ and finite families are uniformly integrable, it suffices to prove

$$\limsup_{k\uparrow\infty} \mathrm{E}_{\mathrm{Q}_\ell}(Z_{\theta_k}\mathbb{1}_{\{Z_{\theta_k}>z\}}) \to 0 \quad (z \uparrow \infty). \tag{A.3}$$

To this end we write, using the fact that $Z_{\theta_k} = Z_\infty$ on $\{\theta_k \ge \tau\}$,

$$\mathrm{E}_{\mathrm{Q}_\ell}(Z_{\theta_k}\mathbb{1}_{\{Z_{\theta_k}>z\}}) \le \mathrm{E}_{\mathrm{Q}_\ell}(Z_{\theta_k}\mathbb{1}_{\{\theta_k<\tau, Z_{\theta_k}>z\}}) \tag{A.4}$$

$$+ \mathrm{E}_{\mathrm{Q}_\ell}(Z_\infty \mathbb{1}_{\{Z_\infty>z\}}). \tag{A.5}$$

Downloaded from informs.org by [2001:638:807:212:a8af:506c:25a1:b254] on 09 November 2026, at 05:59 . For personal use only, all rights reserved.

Because Z^{θ_k} is a unit-mean martingale and $Z \geq 0$, Fatou's lemma yields

$$\mathrm{E}_{\mathrm{Q}_\ell}(Z_\infty) \leq \liminf_{k\uparrow\infty} \mathrm{E}_{\mathrm{Q}_\ell}(Z_\infty^{\theta_k}) \leq \liminf_{k\uparrow\infty} \liminf_{t\uparrow\infty} \mathrm{E}_{\mathrm{Q}_\ell}(Z_t^{\theta_k}) = 1.$$

Hence, Z_∞ is Q_ℓ-integrable and the corresponding term in (A.5) now vanishes as $z \uparrow \infty$. To deal with the term on the right of (A.4), we invoke our induction hypothesis and that θ_k is $\mathcal{H}_\infty$-measurable, so that by construction of τ in (1),

$$\begin{aligned}\mathrm{Q}_\ell(\tau > \theta_k \mid \mathcal{H}_\infty) &= \mathrm{Q}_\ell(\xi > H_{\theta_k} \mid \mathcal{H}_\infty)\\ &= e^{-H_{\theta_k}}.\end{aligned}$$

The last step follows by the independence of $\sigma(\xi)$ and $\mathcal{H}_\infty$ (Section 2.1). Then, because H_{θ_k}, Q_{θ_k}, and θ_k are $\mathcal{H}_\infty$-measurable, the right side of (A.4) reduces to

$$\begin{aligned}\mathrm{E}_{\mathrm{Q}_\ell}(Z_{\theta_k}\mathbb{1}_{\{\theta_k<\tau, Z_{\theta_k}>z\}}) &= \mathrm{E}_{\mathrm{Q}_\ell}(e^{(H-\mathrm{Q})_{\theta_k}}\mathrm{Q}_\ell(\tau > \theta_k \mid \mathcal{H}_\infty)\mathbb{1}_{\{(H-\mathrm{Q})_{\theta_k}>\log z\}})\\ &= \mathrm{E}(Z_\infty(\ell)e^{-\mathrm{Q}_{\theta_k}}\mathbb{1}_{\{(H-\mathrm{Q})_{\theta_k}>\log z\}}).\end{aligned}$$

The case $\mathrm{Q}_\infty = \infty$ is addressed by the much simpler bound

$$\mathrm{E}_{\mathrm{Q}_\ell}(Z_{\theta_k}\mathbb{1}_{\{\theta_k<\tau, Z_{\theta_k}>z\}} \leq \mathrm{E}(Z_\infty(\ell)e^{-\mathrm{Q}_{\theta_k}})$$

because the right side vanishes by an application of Fatou's lemma with $k \uparrow \infty$ along with the facts that $\mathrm{Q} \geq 0$ and $\theta_k \uparrow \infty$ P-almost surely.

The case $H_\infty < \infty$ implies that $(H - \mathrm{Q})_\infty < \infty$ P-almost surely, and so

$$\mathrm{E}_{\mathrm{Q}_\ell}(Z_{\theta_k}\mathbb{1}_{\{\theta_k<\tau, Z_{\theta_k}>z\}}) \leq \mathrm{E}(Z_\infty(\ell)\mathbb{1}_{\{(H-\mathrm{Q})_{\theta_k}>\log z\}}).$$

which vanishes upon taking $z \uparrow \infty$ (Fatou's lemma). Thus, (A.3) holds.

We have now shown that the family $\{Z_{\theta_k}\}_{k\in\mathbb{N}}$ is uniformly integrable with $\lim_{k\uparrow\infty} Z_{\theta_k} = Z_\infty$ Q_ℓ-almost surely, and therefore, by Vitali's convergence theorem (Bogachev [7, theorem 4.5.4]), we have $\mathrm{E}_{\mathrm{Q}_\ell}(Z_\infty) = 1$ as required.

To prove N admits $(\mathrm{Q}_{\ell+1}, \mathbb{F})$-intensity $q = up$, note that $Z_t = \mathrm{E}_{\mathrm{Q}_\ell}(Z_\infty \mid \mathcal{F}_t)$. Applying Protter [45, theorem IV.12], because $[M, M] = N$, we have

$$[M, Z]_t = \int_0^t (u_s - 1)Z_{s-}\mathrm{d}N_s,$$

and hence, by Protter [45, theorem IV.29], $\langle M,Z\rangle_t = \int_0^t Z_s(u_s - 1)p_s\,\mathrm{d}s$. By the predictable version of the Girsanov–Meyer theorem (Protter [45, theorem III.41]), the process $N - \mathrm{Q}^\tau$ is a $(\mathrm{Q}_{\ell+1}, \mathbb{F})$-martingale (see (10)) because

$$\begin{aligned}M_t - \int_0^t (1/Z_{s-})\mathrm{d}\langle M,Z\rangle_s &= M_t - \int_0^t (u_s - 1)p_s\,\mathrm{d}s\\ &= N_t - \mathrm{Q}_t^\tau,\end{aligned}$$

and $N \leq 1$ (see footnote 9). The claim holds by definition of intensity. □

Appendix B. Proof of Proposition 4.1

Take $Y = f(\boldsymbol{N})L_\infty$ as an exact Q-sample of the estimator where L_∞ is a (exact) sample of (15) and $\mathscr{E}$ is the associated integration bias for the paths of the process p. Because the samples of the skeleton of $\boldsymbol{h}$ generated by Algorithm 2 are exact, the samples of $(\lambda_\ell, \eta_\ell)$ are exact, and hence, so are those of $f(\boldsymbol{N})$. Moreover, it is easy to see that the Radon–Nikodym derivative $\hat{L}_\infty$ in (18) generated by Algorithm 2 has the same Q-law as $L_\infty e^{\mathscr{E}}$. Let $f(\hat{\boldsymbol{N}})$ denote the (exact) sample corresponding to $\hat{L}_\infty$ and set $\hat{Y} = f(\hat{\boldsymbol{N}})\hat{L}_\infty$. It follows that

$$\begin{aligned}|\,\mathrm{bias}\,| &= |\,\mathrm{E}_{\mathrm{Q}}(\hat{Y}) - \mathrm{E}_{\mathrm{Q}}(Y)\,| = |\,\mathrm{E}_{\mathrm{Q}}(f(\boldsymbol{N})L_\infty e^{\mathscr{E}} - f(\boldsymbol{N})L_\infty)\,|\\ &= |\,\mathrm{E}(f(\boldsymbol{N})(e^{\mathscr{E}} - 1))\,|\\ &\leq \sum_{k=1}^{\infty} \mathrm{E}(|f(\boldsymbol{N})\,\|\,\mathscr{E}|^k)/k!.\end{aligned}$$

Applying Cauchy–Schwartz to $|f(\boldsymbol{N})\,\|\,\mathscr{E}|^k = |f(\boldsymbol{N})\,\|\,\mathscr{E}|^{k-\rho/2}\sqrt{|\,\mathscr{E}|^\rho}$ does not yield a sharp bound.

Downloaded from informs.org by [2001:638:807:212:a8af:506c:25a1:b254] on 09 November 2026, at 05:59 . For personal use only, all rights reserved.

To ease the notation, set $\epsilon = \mathrm{E}(|\mathscr{E}|)$. By Holder's inequality, we have $\mathrm{E}(|f(\boldsymbol{N})||\mathscr{E}|^k) \le \|f\|_\mu \mathrm{E}(|\mathscr{E}|^{\rho k})^{1/\rho}$. Let $x \ge 1$. By Minkowski's inequality,

$$\begin{aligned}\mathrm{E}(|\mathscr{E}|^{\rho k})^{1/\rho} &= \mathrm{E}((|\mathscr{E}|^k \mathbb{1}_{\{|\mathscr{E}|\le 2x\epsilon\}} + |\mathscr{E}|^k \mathbb{1}_{\{|\mathscr{E}|>2x\epsilon\}})^\rho)^{1/\rho} \\ &= \mathrm{E}(|\mathscr{E}|^{\rho k} \mathbb{1}_{\{|\mathscr{E}|\le 2x\epsilon\}})^{1/\rho} + \mathrm{E}(|\mathscr{E}|^{\rho k} \mathbb{1}_{\{|\mathscr{E}|>2x\epsilon\}})^\rho)^{1/\rho} \\ &\le (2x\epsilon)^{k-1} \mathrm{E}(|\mathscr{E}|^\rho)^{1/\rho} + \mathrm{E}'(|\mathscr{E}|^{\rho k} e^{-\theta X + \Lambda(\theta)} \mathbb{1}_{\{|\mathscr{E}|>2x\epsilon\}})^{1/\rho}\end{aligned} \quad , \tag{B.1}$$

where in the second term we change the measure to $\mathrm{P}' = e^{\theta X - \Lambda(\theta)}\mathrm{P}$ with respect to which the expectation E′ above is taken. Temporarily, assume $\Lambda(\theta) < \infty$.

The series for the first term in (B.1) is simple. Because $2\epsilon \le \delta$,

$$\sum_{k=1}^{\infty} \frac{(2x\epsilon)^{k-1}}{k!} \mathrm{E}(|\mathscr{E}|^\rho)^{1/\rho} \le \left(\frac{e^{\delta x} - 1}{\delta x}\right) \mathrm{E}(|\mathscr{E}|^\rho)^{1/\rho}. \tag{B.2}$$

We deal with the second term next. Let $\theta = 2z\epsilon > 0$. Then,

$$\begin{aligned}\mathrm{E}'(|\mathscr{E}|^{\rho k} e^{-\theta X + \Lambda(\theta)} \mathbb{1}_{\{|\mathscr{E}|>2x\epsilon\}}) &= \mathrm{E}'(e^{-z|\mathscr{E}| + \Lambda(\theta) - z|\mathscr{E}| + \rho k \log|\mathscr{E}|} \mathbb{1}_{\{|\mathscr{E}|>2x\epsilon\}}) \\ &\le e^{-(\theta x - \Lambda(\theta))} \mathrm{E}'(e^{-z|\mathscr{E}| + \rho k \log|\mathscr{E}|}).\end{aligned}$$

Note that $f(a) = -za + \rho k \log(a)$ has $f'(a) = -z + \rho k/a = 0$ at $a = \rho k/z$, and thus $f \le -\rho k + \rho k \log(\rho k/z)$. It follows that[28]

$$\begin{aligned}\sum_{k=1}^{\infty} \mathrm{E}'(e^{-z|\mathscr{E}| + \rho k \log|\mathscr{E}|})^{1/\rho}/k! &\le \sum_{k=1}^{\infty} e^{-k + k\log(\rho k/z)}/k! \\ &\le \frac{z}{z-\rho} - 1 = \frac{2\rho\epsilon}{\theta - 2\rho\epsilon}\end{aligned}.$$

Setting $\theta = \delta\rho$ and $x = \Lambda(\delta\rho)/(\delta\rho)$, we have that $\theta x - \Lambda(\theta) = 0$ and $\theta - 2\rho\epsilon \ge \delta\rho/2$ because $\epsilon < \delta/4$. Consequently, we deduce that

$$\sum_{k=1}^{\infty} \mathrm{E}'(|\mathscr{E}|^{\rho k} e^{-\theta X + \Lambda(\theta)} \mathbb{1}_{\{|\mathscr{E}|>2x\epsilon\}})^{1/\rho} \le (4/\delta)\epsilon \le (4/\delta)\mathrm{E}(|\mathscr{E}|^\rho)^{1/\rho}. \tag{B.3}$$

Assembling bounds (B.2) and (B.3) completes the proof. Note that if $\Lambda(\delta\rho) = \infty$, then we have $x = \infty$ so that the bound holds trivially because $\lim_{y\uparrow\infty}(e^y - 1)/y = \infty$.

Appendix C. Proof of Theorem 5.1

Consider the recursive definition $V_{\ell-1} = \mathrm{E}_{\mathrm{Q}}(W_{\eta_\ell}^{\ell-1} V_\ell \mid \mathcal{G}_\ell)$ for $1 \le \ell \le n$ with $V_n = 1$, where $\mathcal{G}_\ell = \mathcal{F}_{S_{\ell-1}} \vee \mathcal{G}$ (see (22)) and the process $W^{\ell-1}$ is given by

$$W_w^{\ell-1} = \exp\left(-\sum_{k=\ell}^{n} H^{I_k}_{S_{\ell-1}+w} - H^{I_k}_{S_{\ell-1}}\right) \quad w \ge 0.$$

It is easy to check that $L_\infty^{\mathcal{G}} = V_0 \prod_{\ell=1}^n G_{\eta_\ell}^{\ell-1}$ with $G^{\ell-1}$ as in (23). We prove by induction that (under our assumptions) the V_ℓ is $\mathcal{G}_\ell$-measurable and

$$V_{\ell-1} = V_\ell \phi^{\{I_k\}_{k=\ell}^n}_{(S_{\ell-1}, \lambda_{\ell-1})}(S_\ell, \boldsymbol{h}_{S_\ell -}) \quad \mathrm{Q}\text{-almost surely.} \tag{C.1}$$

Then, the formula for $L_\infty^{\mathcal{G}}$ in (25) follows immediately. As the transform ϕ can be evaluated, and the variables in $\mathcal{G}$ can be sampled exactly by our hypotheses, the samples of $L_\infty^{\mathcal{G}}$ in Algorithm 3 (provided by (27)) are exact. Because $f(\boldsymbol{N})$ is assembled only from the variables in $\mathcal{G}$ the remaining conclusion follows.

The main calculations involve $\mathrm{E}_{\mathrm{Q}}(W_w^{\ell-1} \mid \mathcal{G}_\ell)$ and a change of measure from Q to P_0 so as to appeal to the strong Markov property. These steps are facilitated by Theorem 3.1 and Corollary 3.3. Recall that Q is constructed via $u^i = \bar{u}$ in (13) and $\mathrm{Q} = Z_\infty \mathrm{P}$ for $Z_\infty = L_\infty^{-1}$ as in (7). For any $w > 0$, we construct P_0 by taking each $u^i = \bar{u}$ only on $[0, S_{\ell-1})$ while letting $u_s^i = \mathbb{1}_{\{s > S_{\ell-1}+w\}}$ elsewhere. That is, no events occur on $[S_{\ell-1}, S_{\ell-1} + w)$ under P_0. We let $\tilde{Z}_\infty$ denote the corresponding Radon–Nikodym derivative so that $\mathrm{P}_0 = \tilde{Z}_\infty \mathrm{P}$. Observe that

$$G_w^{\ell-1} \tilde{Z}_\infty = \mathcal{N}_w^{\ell-1} Z_{S_{\ell-1}+w}, \tag{C.2}$$

where $\mathcal{N}_w^{\ell-1} = \prod_{k=\ell}^n \mathbb{1}_{\{N^{I_k}_{S_{\ell-1}+w} = 0\}}$ and $Z_{S_{\ell-1}+w} = \mathrm{E}(Z_\infty \mid \mathcal{F}_{S_{\ell-1}+w})$. By (11) and (C.2) we have, because $G_w^{\ell-1}$ and $\mathcal{N}_w^{\ell-1}$ are $\mathcal{G}_\ell$-measurable, that

Downloaded from informs.org by [2001:638:807:212:a8af:506c:25a1:b254] on 09 November 2026, at 05:59 . For personal use only, all rights reserved.

$$\begin{aligned}\mathrm{E}(G_w^{\ell-1}\tilde{Z}_\infty \mid \mathcal{G}_\ell)\mathrm{E}_{\mathrm{P}_0}(W_w^{\ell-1} \mid \mathcal{G}_\ell) &= \mathrm{E}(W_w^{\ell-1}G_w^{\ell-1}\tilde{Z}_\infty \mid \mathcal{G}_\ell)\\ &= \mathrm{E}(W_w^{\ell-1}\mathcal{N}_w^{\ell-1}Z_{S_{\ell-1}+w} \mid \mathcal{G}_\ell)\\ &= \mathrm{E}(Z_{S_{\ell-1}+w} \mid \mathcal{G}_\ell)\mathrm{E}_{\mathrm{Q}}(W_w^{\ell-1} \mid \mathcal{G}_\ell)\end{aligned}$$

Q-almost surely whenever $0 < w < \eta_\ell$ (i.e., $\mathcal{N}_w^{\ell-1} = 1$ $\mathrm{Q}(\cdot \mid \mathcal{G}_\ell)$-almost surely because no event may occur before η_ℓ time has passed). It follows that

$$\mathrm{E}_{\mathrm{Q}}(W_w^{\ell-1} \mid \mathcal{G}_\ell) = \mathrm{E}_{\mathrm{P}_0}(W_w^{\ell-1} \mid \mathcal{G}_\ell) 0 < w < \eta_\ell, \mathrm{Q}\text{-almostsurely}. \tag{C.3}$$

Because $W^\ell \le 1$, we have $\mathrm{E}_{\mathrm{Q}}(W_{\eta_\ell}^{\ell-1} \mid \mathcal{G}_\ell) = \lim_{w\uparrow\eta_\ell}\mathrm{E}_{\mathrm{Q}}(W_w^{\ell-1} \mid \mathcal{G}_\ell)$. Applying identity (C.3) and dominated convergence for conditional expectations yields $\mathrm{E}_{\mathrm{Q}}(W_{\eta_\ell}^{\ell-1} \mid \mathcal{G}_\ell) = \mathrm{E}_{\mathrm{P}_0}(W_{\eta_\ell}^{\ell-1} \mid \mathcal{G}_\ell)$. For $0 \le s < t$, define the transform

$$\varphi_{(s,x)}^{\chi}(t) = \mathrm{E}_{\mathrm{P}_0}\left(e^{-\sum_{i\in\chi}(H_t^i - H_s^i)} \mid \boldsymbol{h}_s = x\right). \tag{C.4}$$

For any $\chi \subseteq \{1,\dots,n\}$ (recall the $\mathcal{F}_{S_{\ell-1}}$-measurable collection of events $\{I_k\}_{k=\ell}^n$ that have not occurred by time $S_{\ell-1}$), on the set $\{\{I_k\}_{k=\ell}^n = \chi\}$,

$$\begin{aligned}\mathrm{E}_{\mathrm{P}_0}(W_{\eta_\ell}^{\ell-1} \mid \mathcal{F}_{S_{\ell-1}}\vee\sigma(T_\ell)) &= \mathrm{E}_{\mathrm{P}_0}(W_{\eta_\ell}^{\ell-1} \mid \sigma(S_{\ell-1}, \boldsymbol{h}_{S_{\ell-1}}, T_\ell, \{I_k\}_{k=\ell}^n))\\ &= \varphi_{(S_{\ell-1},\boldsymbol{h}_{S_{\ell-1}})}^{\chi}(T_\ell)\end{aligned} \tag{C.5}$$

by the strong P_0-Markov property of $\boldsymbol{h}$ at $S_{\ell-1}$ (Dynkin [22, theorem 3.2]). Note that T_ℓ becomes known as soon as $\boldsymbol{h}_{S_{\ell-1}}$ is known (i.e., η_ℓ is exponential with rate $\lambda_{\ell-1}$), which facilitates the application of the strong Markov property. We now have the Markov property for the sequence $\{W_{\eta_\ell}^{\ell-1}\}$, which is time reversible (cf. Léonard et al. [37]), that is,

$$\mathrm{E}_{\mathrm{P}_0}(W_{\eta_\ell}^{\ell-1} \mid \mathcal{G}_\ell) = \mathrm{E}_{\mathrm{P}_0}(W_{\eta_\ell}^{\ell-1} \mid \sigma(\{I_k\}_{k=\ell}^n) \vee \sigma(S_{\ell-1}, \boldsymbol{h}_{S_{\ell-1}}, T_\ell, \boldsymbol{h}_{T_\ell-})).$$

This and (C.5) imply $\mathrm{E}_{\mathrm{P}_0}(W_{\eta_\ell}^{\ell-1} \mid \mathcal{G}_\ell) = \phi_{(S_{\ell-1},\boldsymbol{h}_{S_{\ell-1}})}^{\chi}(T_\ell, \boldsymbol{h}_{T_\ell-})$, where we use regular conditional (on $\boldsymbol{h}_{T_\ell-}$) probability to evaluate the bridge transform ϕ^χ.

Finally, using (C.3),

$$\mathrm{E}_{\mathrm{Q}}(W_{\eta_\ell}^{\ell-1} \mid \mathcal{G}_\ell) = \phi_{(S_{\ell-1},\boldsymbol{h}_{S_{\ell-1}})}^{\chi}(T_\ell, \boldsymbol{h}_{T_\ell-}) \ \mathrm{Q}\text{-almost surely}$$

on $\{\{I_k\}_{k=\ell}^n = \chi\}$ and $T_\ell = S_\ell$ under Q by construction as required.

To conclude, we have that $V_{\ell-1} = V_\ell \mathrm{E}_{\mathrm{Q}}(W_{\eta_\ell}^{\ell-1} \mid \mathcal{G}_\ell) = V_\ell \phi_{(S_{\ell-1},\lambda_{\ell-1})}^{\{I_k\}_{k=\ell}^n}(S_\ell, \boldsymbol{h}_{S_\ell-})$ is $\mathcal{G}_\ell$-measurable. The base case ($\ell = n$) holds as $V_n = 1$ is trivially $\mathcal{G}_\ell$-measurable. Hence, (C.1) holds for all $\ell \le n$, which completes the proof. □

Endnotes

[1] Other applications of measure changes include reducing the mixing time in Markov chain Monte Carlo and performing a sensitivity analysis on discrete event systems.

[2] Usually, the intensity is sampled at a set of grid points (by an exact or discretization method) and then these values are integrated to obtain samples of the compensator skeleton, that is, the finite-dimensional representation of the compensator sample path.

[3] See, for example, Beskos and Roberts [3], Chen and Huang [13] and Giesecke and Smelov [24].

[4] See Pitman and Yor [43], Broadie and Kaya [9], and Kang and Kang [33].

[5] More precisely, the Laplace transform of the compensator with respect to the law of the bridge of the interarrival intensity process.

[6] See, for example, Norros [41], Shaked and Shanthikumar [46], and Bielecki and Rutkowski [4].

[7] We adopt the standard convention $\inf\emptyset = \infty$.

[8] This is facilitated by the progressive enlargement of $\mathbb{H}_0$ by the times $\{\tau_j\}_{j\neq i}$ to obtain $\mathbb{H}_i$. The construction guarantees the $\{\tau_i\}$ are totally inaccessible stopping times with respect to $\mathbb{F}$. A more explicit approach would progressively enlarge $\mathbb{H}_0$ by the order statistics $\{\tau_{(\ell)}\}$ of the $\{\tau_\ell\}$ up to, but not including, τ_i to yield $\mathbb{H}_i$. The ℓ th order statistic $\tau_{(\ell)}$ is the first hitting time of $\sum_{k=1}^\ell \xi_k$ by the compensator $\sum_{i=1}^n H_{\cdot\wedge\tau_i}^i$ of $\sum_{i=1}^n N^i$. The information regarding the identity of the events is embedded via a "marked process" (Brémaud [8, theorem II.T15]).

[9] To see this, observe that $\mathrm{E}(\sup_{s\le t} \mid M_s^i \mid) \le 1 + \mathrm{E}(A_t^i)$ and let $\{\theta_k\}_{k=1}^\infty$ be a sequence of localization times for M_i. Then, $\mathrm{E}(M_{t\wedge\theta_k}^i) = 0$, and by monotone convergence, $\mathrm{E}(A_t^i) \le 1$.

[10] Note that $\tau_i < \infty$ if and only if $H_\infty^i = \infty$. In general, if the $\{\theta_i\}$ are i.i.d. standard exponentials, each θ_i independent of $\mathcal{H}_\infty^i$, then the $\{A_\infty^i + \theta\mathbb{1}_{\{\tau_i=\infty\}}\}$ are i.i.d. standard exponentials.

[11] See, for example, Lando [36], Duffie and Garleanu [19], Jarrow and Yu [32], Collin-Dufresne et al. [15], and Giesecke et al. [29].

Downloaded from informs.org by [2001:638:807:212:a8af:506c:25a1:b254] on 09 November 2026, at 05:59 . For personal use only, all rights reserved.

[12] The algorithms of Sections 4 and 5 extend to expectations of the form $\mathrm{E}(f(\boldsymbol{N}, X))$, where X is a random variable. Note that (2) encompasses $\mathrm{E}(g(\boldsymbol{N}_T))$ for fixed $T > 0$ and $g : \mathbb{S} \to \mathbb{R}$.

[13] Throughout, O corresponds to the standard big O notation.

[14] The algorithm is easily adapted to sampling of $g(\boldsymbol{N}_T)$ for a stopping time T and function g. In this case, one may wish to terminate the path of $\hat{h}^i$ if $\hat{\tau}_i$ is already known to exceed T.

[15] One may also combine simulation runs with differing discretization spacings Δ_g (see (Talay [49, section 5.7, Romberg extrapolations]) to significantly improve the order of the approximation.

[16] Although feasible for some models of $\boldsymbol{h}$, this may be prohibitively expensive.

[17] When bridge sampling of the intensity is possible, one could reduce the uncertainty in the event time to an arbitrary precision with a bisection search (cf. Asmussen and Glynn [1, chapter X, 2b]). Another conceivable approach might reduce the time-grid spacing as $\hat{H}^i$ approaches the level ξ_i. Both heuristics are computationally expensive, difficult to analyze, and challenging to implement. More importantly, these approaches limit the scope of the scheme.

[18] In most practical settings, the predictable version is easy to determine. For instance if X is càdlàg (or more generally a process with left limits) then X_- is the predictable version of X.

[19] For complete rigor, let $L_t = 1/Z_t$ with Z in (9), a Doob martingale. Taking $t \uparrow \infty$ yields (7) with $N^i_\infty \log(u^i_{\tau_i}) = \lim_{t\uparrow\infty} N^i_t \log(u^i_{t\wedge\tau^i}) = \log(u^i_{\tau_i})$ if $\tau_i < \infty$ and zero otherwise.

[20] Consistently with the convention $1/0 = \infty$, Z_∞ is the expression in (7) with the exponent negated. It is the last element of the Doob martingale Z defined by $Z_t = \mathrm{E}(Z_\infty \mid \mathcal{F}_t)$ and

$$Z_t = \exp\left(\sum_{i=1}^n (H^i - \mathrm{Q}^i)_{t\wedge\tau_i} + N^i_t \log(u^i_{t\wedge\tau_i})\right).$$

[21] See, for example, Brémaud [8], Kusuoka [34], Giesecke and Zhu [25], and Sokol and Hansen [48].

[22] In the case of division by zero in (13), that is, if $\sum_{i=1}^n p^i_s = 0$ for all $s \in \mathscr{T}$, without loss of generality we may set $\bar{u}_s = 1$ for all $s \in \mathscr{T}$. Then $\sum_{i=1}^n q^i = 0$ on $\mathscr{T}$, and no events occur under Q during $\mathscr{T}$ (the conditions of Lemma 3.2 are satisfied). The Q-event times may then be generated via inversion of the Q-compensator of C, which takes a simple piecewise linear form.

[23] When estimating expectations of $f(\boldsymbol{N}, X)$, the new variable X would be included in $\mathcal{G}$.

[24] Note that unlike time scaling, the plain/exact JAM may be vectorized and parallelized.

[25] We do not implement this scheme, as it provides no further insight into our methods.

[26] Here, it is important that u is predictable in order to define the integral on the right of (A.2).

[27] These times are distinct from those utilized in the proof of the related (Brémaud [8, theorem VI.2]) and do not form a localizing sequence for the local martingales therein. Equation (A.1) localizes Z because we have $|N| \leq 1$. Our $\{\theta_k\}$ are $\mathcal{H}_\infty$-measurable; a fact crucial to our proof.

[28] Because $k! \geq k^k e^{-k}$, yields $e^{-k+k\log(\rho k/z)}/k! \leq e^{-k}(\rho k/z)^k/k! \leq (\rho/z)^k$.

References

[1] Asmussen S, Glynn P (2007) *Stochastic Simulation—Algorithms and Analysis* (Springer, New York).
[2] Azizpour S, Giesecke K, Schwenkler G (2018) Exploring the sources of default clustering. *J. Financial Econom.* 129(1):154–183.
[3] Beskos A, Roberts G (2005) Exact simulation of diffusions. *Ann. Appl. Probab.* 15(4):2422–2444.
[4] Bielecki TR, Rutkowski M (2004) *Credit Risk: Modeling, Valuation and Hedging* (Springer-Verlag, Berlin Heidelberg).
[5] Blanchet J, Ruf J (2016) A weak convergence criterion for constructing changes of measure. *Stochastic Models* 32(2):233–252.
[6] Blanchet J, Zhang F (2020) Exact simulation for multivariate Itô diffusions. *Adv. Appl. Probab.* 52(4):1003–1034.
[7] Bogachev VI (2007) *Measure Theory*, vol. 1 (Springer-Verlag, Berlin Heidelberg).
[8] Brémaud P (1981) *Point Processes and Queues—Martingale Dynamics* (Springer-Verlag, New York).
[9] Broadie M, Kaya O (2006) Exact simulation of stochastic volatility and other affine jump diffusion processes. *Oper. Res.* 54(2):217–231.
[10] Casella B, Roberts GO (2011) Exact simulation of jump-diffusion processes with Monte Carlo applications. *Methodol. Comput. Appl. Probab.* 13(3):449–473.
[11] Chen G, Shkolnik A, Giesecke K (2019) Unbiased simulation estimators for jump-diffusions. *Proc. 2019 Winter Simulation Conf.* (IEEE Press, Piscataway, NJ), 890–901.
[12] Chen G, Shkolnik A, Giesecke K (2020) Unbiased simulation estimators for path integrals of diffusions. *Proc.* 2020 *Winter Simulation Conf.* (IEEE Press, Piscataway, NJ), 227–288.
[13] Chen N, Huang Z (2013) Localization and exact simulation of Brownian motion-driven stochastic differential equations. *Math. Oper. Res.* 38(3):591–616.
[14] Cheng P, Scaillet O (2007) Linear-quadratic jump-diffusion modeling. *Math. Finance* 17(4):575–598.
[15] Collin-Dufresne P, Goldstein R, Hugonnier J (2004) A general formula for the valuation of defaultable securities. *Econometrica* 72:1377–1407.
[16] Cox J, Ingersoll J, Ross S (1985) A theory of the term structure of interest rates. *Econometrica* 53:385–408.
[17] Daley DJ, Vere-Jones D (2003) *An Introduction to the Theory of Point Processes—Volume I: Elementary Theory and Methods* (Springer-Verlag, New York).
[18] Das S, Duffie D, Kapadia N, Saita L (2007) Common failings: How corporate defaults are correlated. *J. Finance* 62:93–117.
[19] Duffie D, Garleanu N (2001) Risk and valuation of collateralized debt obligations. *Financial Anal. J.* 57(1):41–59.

[20] Duffie D, Singleton K (1999) Simulating correlated defaults. Working Paper, Stanford University, Stanford, CA.
[21] Duffie D, Pan J, Singleton K (2000) Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* 68(6):1343–1376.
[22] Dynkin EB (1973) Regular Markov processes. *Russian Math. Surveys* 28(2):33–64.
[23] Giesecke K, Kim B (2007) Estimating tranche spreads by loss process simulation. Henderson SG, Biller B, Hsieh MH, Shortle J, Tew JD, Barton RR, eds. *Proc. 2007 Winter Simulation Conf.* (Institute of Electrical and Electronics Engineers, Piscataway, NJ), 967–975.
[24] Giesecke K, Smelov D (2013) Exact sampling of jump diffusions. *Oper. Res.* 61(4):894–907.
[25] Giesecke K, Zhu S (2013) Transform analysis for point processes and applications in credit risk. *Math. Finance* 23(4):742–762.
[26] Giesecke K, Kakavand H, Mousavi M (2011) Exact simulation of point processes with stochastic intensities. *Oper. Res.* 59(5):1233–1245.
[27] Giesecke K, Kim B, Zhu S (2011) Monte Carlo algorithms for default timing problems. *Management Sci.* 57(12):2115–2129.
[28] Giesecke K, Kakavand H, Mousavi M, Takada H (2010) Exact and efficient simulation of correlated defaults. *SIAM J. Financial Math.* 1:868–896.
[29] Giesecke K, Spiliopoulos K, Sowers R, Sirignano J (2015) Large portfolio asymptotics for loss from default. *Math. Finance* 25(1):77–114.
[30] Glasserman P (2003) *Monte Carlo Methods in Financial Engineering* (Springer-Verlag, New York).
[31] Glasserman P, Merener N (2004) Convergence of a discretization scheme for jump-diffusion processes with state-dependent intensities. *Proc. Royal Soc. Lond. A.* 460(2041):111–127.
[32] Jarrow RA, Yu F (2001) Counterparty risk and the pricing of defaultable securities. *J. Finance* 56(5):555–576.
[33] Kang C, Kang W (2013) Transform formulae for linear functionals of affine processes and their bridges on positive semidefinite matrices. *Stochastic Process. Appl.* 123(6):2419–2445.
[34] Kusuoka S (1999) A remark on default risk models. *Adv. Math. Econom.* 1(1):69–82.
[35] Kuznetsov S (1984) Nonhomogeneous Markov processes. *J. Soviet Math.* 25(5):1380–1498.
[36] Lando D (1998) On Cox processes and credit risky securities. *Rev. Derivatives Res.* 2:99–120.
[37] Léonard C, Rœlly S, Zambrini JC, (2014) Reciprocal processes. A measure-theoretical point of view. *Probab. Surveys* 11:237–269.
[38] Lewis P, Shedler G (1979) Simulation of nonhomogeneous poisson processes by thinning. *Naval Logist. Quart.* 26(3):403–413.
[39] Li T (2007) Analysis of explicit tau-leaping schemes for simulating chemically reacting systems. *Multiscale Model. Simultion* 6(2):417–436.
[40] Meyer PA (1971) *Démonstration simplifée d'un théorème de Knight.* Séminaire de Probabilités V. Lecture Note in Mathematics vol. 191 (Springer-Verlag, Berlin), 191–195.
[41] Norros I (1986) A compensator representation of multivariate life length distributions, with applications. *Scandinavian J. Statist.* 13(2): 99–112.
[42] Ogata Y (1981) On Lewis' simulation method for point processes. *IEEE Trans. Inform. Theory* 27(1):23–31.
[43] Pitman J, Yor M (1982) A decomposition of Bessel bridges. Z. Wahrscheinlichkeitstheorie verw. *Gebiete* 59:425–457.
[44] Pollock M, Johansen AM, Roberts GO (2016) On the exact and ε-strong simulation of (jump) diffusions. *Bernoulli* 22(2):794–856.
[45] Protter P (2005) *Stochastic Integration and Differential Equations* (Springer-Verlag, New York).
[46] Shaked M, Shanthikumar JG (1987) The multivariate hazard construction. *Stochastic Process. Appl.* 24(2):241–258.
[47] Shkolnik A, Giesecke K, Teng G, Wei Y (2022) Numerical solution of jump-diffusion SDEs. *Oper. Res.* Forthcoming.
[48] Sokol A, Hansen NR (2015) Exponential martingales and changes of measure for counting processes. *Stochastic Anal. Appl.* 33(5):823–843.
[49] Talay D (1995) Simulation of stochastic differential systems. Krée P, Wedig W, eds. *Probabilistic Methods in Applied Physics.* Lecture Notes in Physics, vol. 451 (Springer, Berlin), 54–96.
[50] Yu F (2007) Correlated defaults in intensity-based models. *Math. Finance* 17(2):155–173.

Downloaded from informs.org by [2001:638:807:212:a8af:506c:25a1:b254] on 09 November 2026, at 05:59 . For personal use only, all rights reserved.