*Proceedings of the 2025 Winter Simulation Conference*
*E. Azar, A. Djanatliev, A. Harper, C. Kogler, V. Ramamohan, A. Anagnostou, and S. J. E. Taylor, eds.*

# EXACT IMPORTANCE SAMPLING FOR A LINEAR HAWKES PROCESS

Liang Feng,[1] and Alex Shkolnik[1]

[1]Dept. of Statistics and Applied Probability, University of California, Santa Barbara, US

## ABSTRACT

We develop exact importance sampling schemes for (linear) Hawkes processes to efficiently compute their tail probabilities. We show that the classical approach of exponential twisting, while conceptually simple to apply, leads to simulation estimators that are difficult to implement. These difficulties manifest in either a simulation bias or unreasonable computational costs. We mimic exponential twisting with an exponential martingale approach to achieve identical variance reduction guarantees but without the aforementioned challenges. Numerical tests compare the two, and present benchmarks against plain Monte Carlo.

## 1 INTRODUCTION

Hawkes (1971) introduced a highly influential point process model in which the inter-arrival intensity is deterministic. By now, Hawkes processes have found a wide array of application in the sciences, engineering, statistics, operations research and mathematical finance (Laub et al. 2021). There is also a large and growing literature on the Monte Carlo simulation of these processes. However, remarkably few techniques for rare-event simulation for Hawkes processes exist. We make some progress in filling this gap.

For constants $\mu, \kappa, v > 0$ we consider a process $X = (X_t)_{t \geq 0}$ given by,

$$X_t = \mu + (X_0 - \mu)e^{-\kappa t} + v \int_0^t e^{-\kappa(t-s)} \mathrm{d}N_s \tag{1}$$

where $N$ is a univariate point process of intensity $X$ (e.g., Brémaud (1981)).

The process $N$ takes values in $\mathbb{N} = \{0, 1, 2, \dots\}$ and is called a "linear" Hawkes process as its intensity is a linear function of the state $X$. Fixing $T > 0$, we consider estimating the probability $\mathrm{P}(\mathscr{E}) = \mathrm{P}(N_T \geq n)$. It is well-known that the plain Monte Carlo estimator $1_{\mathscr{E}}$ of $\mathrm{P}(\mathscr{E})$ has a squared relative error that is given by $1/\mathrm{P}(\mathscr{E}) - 1$, which diverges as $\mathrm{P}(\mathscr{E})$ vanishes (e.g., Asmussen and Glynn (2007), Chapter VI). Consequently, plain Monte Carlo simulation of tail probabilities for a Hawkes process is highly inefficient.

A classic approach to this problem is exponential twisting, an importance sampling method that dates back to the seminal work of Siegmund (1976). This estimator is based on an exponential change of measure (ECM), which uses

$$Z_T(\theta) = \frac{\mathrm{dP}_\theta}{\mathrm{dP}} = e^{\theta N_T} / \mathrm{E}(e^{\theta N_T}) \tag{2}$$

as its Radon-Nikodym derivative, to induce an importance measure $\mathrm{P}_\theta$. The paramater $\theta \in \mathbb{R}$ is called the exponential twist and the associated simulation estimator $Q = 1_{\mathscr{E}} / Z_T(\theta)$ satisfies $\mathrm{E}_{\mathrm{P}_\theta}(Q) = \mathrm{P}(\mathscr{E})$ allowing for importance sampling under $\mathrm{P}_\theta$. The theoretical guarantees for the variance reduction obtained by using $Q$ are well understood (Asmussen and Glynn 2007, Chapter VI.2).

In this paper we point out that exponential twisting of a Hawkes process encounters significant practical challenges. First, the transform $\mathrm{E}(e^{\theta N_T})$ is challenging to evaluate numerically, so that sampling via transform inversion is not robust. Second, no black-box $\mathrm{P}_\theta$-simulation of $N_T$ is possible (Asmussen and Glynn 2007, Proposition 7.1), i.e., naive acceptance/rejection is not feasible because (2) is unbounded when

$\theta > 0$, which corresponds to a reduction in variance. Lastly, the point process $N$ is not in the parametric family of Hawkes processes under the ECM. We prove $N$ admits the $\mathrm{P}_\theta$-intensity $e^{\theta+\nu q}X$ with $X$ per (1) where $q$ is a $\mathbb{R}$-valued function of time that admits no closed-form. This makes intensity-based simulation schemes difficult to implement. In particular, most methods for simulating point processes are based either on time-scaling or thinning; (see Shkolnik et al. (2024) for a treatment of time-scaling for processes with jumps). The former leads to simulation bias (Giesecke and Shkolnik 2022). Thinning (e.g., Ogata (1981)) can and does, in the above setting, lead to unreasonably long simulation times.

There is by now a large and growing literature on applications of Hawkes processes (Hawkes 1973; Chavez-Demoulin et al. 2005; Hewlett 2006; Large 2007; Bowsher 2007; Bacry et al. 2015; Mei and Eisner 2017). In parallel, there is also significant work on their simulation Magris (2019), Kirchner (2017), Chen (2021), Dassios and Zhao (2013), Møller and Rasmussen (2005), Møller and Rasmussen (2006). However, these algorithms do not address the tail estimation problem $\mathrm{P}(N_T \geq n)$.

Rare-event simulation for Hawkes processes was considered in El Maazouz and Bennouna (2018), but without theoretical guarantees on variance reduction. Such guarantees for "generalized" Hawkes processes are presented in Zhang et al. (2009) and Zhang et al. (2015) but for large $T$ asymptotics. This leads to qualitatively different applications, and requires stationarity. We consider a fixed time $T$ and "large intensity" and large $k = \gamma\mu$ asymptotics where $\mu$ tends to infinity. See Giesecke and Shkolnik (2010) and Giesecke and Shkolnik (2025) and the references therein for this rare-event regime along with importance sampling estimators that may be applied to a Hawkes process. However, the conditions required in these papers for optimality are not met by model (1).

We develop an importance sampling scheme that has all the variance reduction guarantees of the ECM, but also allows for exact sampling (i.e., exact samples of the rare event and the Radon-Nikodym derivative may be obtained). The approach is based on exponential martingales that mimic the variance reduction properties of the ECM. See Chen et al. (2019), Chen et al. (2019) and Chen et al. (2025) for related approaches that use changes of measure to develop unbiased estimators. The resulting algorithm can further take advantage of the paths of of $N$ (rather than just the terminal value $N_T$) to incorporate early stopping criteria leading to significant run-time advantages. We prove logarithmic efficiency of our importance sampling estimator to guarantee a fixed precision with a number of trials that grows subexponentially (in $|\log \mathrm{P}(\mathscr{E})|$). We test the estimator numerically, comparing its performance to plain Monte Carlo, as well as a biased implementation of exponential twisting.

Section 2 introduces exponential twisting and the rare-event regime. Section 3 discusses the challenges of implementing estimators that use exponential twisting. Section 4 develops an exact estimator for the tail probability of a Hawkes process and its theoretical guarantees. The latter results are proved in Section A. Section 5 illustrates the properties of the estimator on numerical examples. Appendices B–C contain auxiliary proofs and calculations.

## 2 RARE-EVENT SIMULATION VIA EXPONENTIAL TWISTING

A common approach to analyzing the tail behavior of $N_T$ leverages an asymptotic analysis. To this end, let $(N^\mu)_{\mu\geq 0}$ be a family of point processes, each $N^\mu$ with intensity $X^\mu$ solving the stochastic differential equation,

$$\mathrm{d}X_t^\mu = \kappa(\mu - X_t^\mu)\,\mathrm{d}t + \nu\,\mathrm{d}N_t^\mu \qquad X_0^\mu = \mu x_0 \in \mathbb{R}_+. \tag{3}$$

Itô's formula may be applied to show that $(X^\mu, N^\mu)$ follows equation (1) so that $(N^\mu)$ forms a family of Hawkes processes indexed by parameter $\mu$.

For a fixed constant $\gamma > 0$, we consider the tail events

$$\mathscr{E}_\mu = \{N_T^\mu \geq \gamma\mu\} \tag{4}$$

so that $(\mathscr{E}_\mu)_{\mu\in\mathbb{N}}$ forms a rare-event sequence if $\mathrm{P}(\mathscr{E}_\mu) \to 0$ as $\mu \to \infty$. These asymptotics concern a large intensity regime and the tail behavior of $N^\mu/\mu$.

Exponential twisting entails an estimator of $\mathrm{P}(\mathscr{E}_\mu)$ given by,

$$Q_{\theta,\mu} = \frac{1_{\mathscr{E}_\mu}}{Z_T^\mu(\theta)}; \qquad Z_T^\mu(\theta) = \exp(\theta N_T^\mu - \Lambda_\mu(\theta)) \tag{5}$$

for $\theta \in \mathbb{R}$ and where $\Lambda_\mu(\cdot)$ is the cumulant generating function, i.e.,

$$\Lambda_\mu(\theta) = \log \mathrm{E}(e^{\theta N_T^\mu}) \tag{6}$$

which takes values in $(-\infty,\infty]$. When $\Lambda_\mu(\theta)$ is finite, $\mathrm{E}(Z_T^\mu(\theta) = 1$ which induces a probability $\mathrm{P}_\theta = Z_T^\mu(\theta)\mathrm{P}$ (i.e., $\mathrm{P}_\theta(A) = \mathrm{E}(1_A Z_T^\mu(\theta))$ for all measurable events $A$). This importance measure thus satisfies,

$$\mathrm{E}_\theta(Q_{\theta,\mu}) = \mathrm{P}(\mathscr{E}_\mu) \tag{7}$$

where $\mathrm{E}_\theta$ denotes the expectation with respect to $\mathrm{P}_\theta$. This summarizes the application of the ECM in (2) to derive the exponential twisting estimator $Q_{\theta,\mu}$.

The finiteness and some additional properties of $\Lambda_\mu$ in (6) govern the rate of decay of $\mathrm{P}(\mathscr{E}_\mu)$ via large deviations theory.

**Lemma 1** For all $\mu > 0$ we have $\Lambda_\mu(\theta) = \mu\Lambda_1(\theta)$. The set,

$$\mathbb{D}_N = \{\theta \in \mathbb{R} : \Lambda_1(\theta) < \infty\} \tag{8}$$

has a nonempty interior $\mathbb{D}_N^o$ with $0 \in \mathbb{D}_N^o$ and $\Lambda_1(\cdot)$ is differentiable on $\mathbb{D}_N^o$. Moreover, $\nabla\Lambda_1(\theta) = \gamma$ has a (unique) solution $\theta > 0$ whenever

$$\gamma > \left(\frac{\kappa T}{\kappa - \nu}\right) + \left(\frac{\sqrt{\nu}}{\kappa - \nu}\right)^2 (e^{-(\kappa-\nu)T} - 1) = \gamma_0\,. \tag{9}$$

*Proof.* The first claim follows from a direct calculation of $\Lambda_\mu(\theta)$ based on equations (14)–(15) below. The existence of $\theta_+ > 0$ such that $\Lambda_1(\theta_+) < \infty$ follows from Zhu (2013), Lemma 2 which along with $N_T^1 \geq 0$ justifies the claim regarding $\mathbb{D}_N^o$. The remaining claims follow from Dembo and Zeitouni (2010), Lemma 2.2.5 recognizing that $\nabla\Lambda_\mu(0) = \mathrm{E}(N_T^\mu) = \mu\gamma_0$ (see Dembo and Zeitouni (2010), Exercise 2.3.25). The latter is obtained by computing $\mathrm{E}(X_T)$ via (1) and using that $(N_t - \int_0^t X_s\,\mathrm{d}s)_{t\geq 0}$ forms a martingale to calculate $\mu\gamma_0 = \int_0^T \mathrm{E}(X_s)\,\mathrm{d}s$. □

Equipped with Lemma 1 we have that the Legendre-Fenchel transform

$$\Lambda_1^*(y) = \sup_{\theta\in\mathbb{R}}\{\theta y - \Lambda_1(\theta)\} \tag{10}$$

has a maximizer that is attained at $\theta \in \mathbb{D}_N$ that solves $\nabla\Lambda_1(\theta) = y$. As a consequence of Cramér's theorem in $\mathbb{R}$ (Dembo and Zeitouni 2010, Chapter 2),

$$\lim_{\mu\to\infty}\frac{1}{\mu}\log\mathrm{P}(\mathscr{E}_\mu) = -\inf_{y>\gamma}\Lambda_1^*(y) = -\Lambda_1^*(\gamma) < 0 \tag{11}$$

under condition (9) (note that $\Lambda_1^*(\gamma) = \theta\gamma - \Lambda_1(\theta) \geq \theta(\gamma - \gamma_0) > 0$ for $\theta > 0$ as in Lemma 1 and where we applied Jensen's inequality for the bound).

In particular, we have $\mathrm{P}(\mathscr{E}_\mu) \to 0$ exponentially in $\mu \to \infty$. Moreover, to analyze the performance of the estimator $Q_{\theta,\mu}$ in (5), consider

$$\liminf_{\mu\to\infty}\frac{\log\mathrm{E}_\theta(Q_{\theta,\mu}^2)}{\log\mathrm{P}(\mathscr{E}_\mu)} = f\,. \tag{12}$$

Jensen's inequality reveals that $f \leq 2$ and an estimator that achieves $f = 2$ is called logarithmically efficient (Asmussen and Glynn 2007). This guarantees a fixed degree of precision with a number of importance sampling trials that is subexponential in $\mu$. Plain Monte Carlo (i.e., $Q_{0,m} = 1_{\mathscr{E}_\mu}$) has $f = 1$.

**Theorem 1** (Logarithmic efficiency of ECM) $Q_{\theta,\mu}$ achieves $f=2$ for $\theta\in\mathbb{D}_N^o$ the solution of $\nabla\Lambda_1(\theta)=\gamma$ and under condition (9).

*Proof.* By definition (5), for any $\theta>0$, the numerator in (12) is bounded as,

$$\log \mathrm{E}_\theta(Q_{\theta,\mu}^2)\leq \log \mathrm{E}_\theta(\exp(-2\theta\gamma+2\Lambda_\mu(\theta)))=-2(\theta\gamma\mu-\Lambda_\mu(\theta)). \tag{13}$$

Consequently, using (10)–(11) and the relation $\Lambda_\mu(\theta)=\mu\Lambda_1(\theta)$ from Lemma 1, the theorem holds for $\theta>0$ solving $\nabla\Lambda_1(\theta)=\gamma$. □

## 3 SIMULATION BIAS OF EXPONENTIAL TWISTING

Below, we omit the superscript $\mu$ in (3) and refer to the solution $(X,N)$ per (1) for any $\mu$. To implement the exponential twisting of $N_T$ we rely on the computation of the conditional transform of a Hawkes process. For $t\leq T$ and $(\mathscr{F})_{t\geq 0}$ the filtration generated by $N$,

$$\mathrm{E}\left(e^{\theta(N_T-N_t)}\,|\,\mathscr{F}_t\right)=\exp\left(p(t)+q(t)X_t\right) \tag{14}$$

where $p$ and $q$ solve the ODE system (Keller-Ressel and Mayerhofer 2015),

$$\begin{aligned} \dot{p}&=-\kappa\mu q & p(T)&=0,\\ \dot{q}&=\kappa q-(e^{\theta+vq}-1) & q(T)&=0. \end{aligned} \tag{15}$$

These (Ricatti) equations do not admit a closed-form solution. As a consequence, we show that the estimator $Q_{\theta,\mu}$ in (5) is difficult to implement.

The following result, which should be standard, does not appear in the simulation literature on Hawkes processes to our knowledge. Its proof is deferred to Appendix B and leverages Girsanov-Meyer theory.

**Proposition 1** Under probability $\mathrm{P}_\theta$ with $\theta\in\mathbb{D}_N^o$, the process $X$ solves (1) with a point process $N$ that admits a $\mathrm{P}_\theta$-intensity $e^{\theta+vq}X$.

Notably, $N$ is not a Hawkes process under $\mathrm{P}_\theta$. Denoting by $T_k$ the $k$th jump time of $N$ and letting $h(x,t)=\mu+(x-\mu)e^{-\kappa t}$, we have that

$$H_{\theta,k}(x,t)=\int_0^t e^{\theta+vq(T_k+s)}h(x,s)\,\mathrm{d}s \tag{16}$$

defines the $\mathrm{P}_\theta$-cumulative hazard function for $T_{k+1}$ given $X_{T_k}=x$.

In particular, $\mathrm{P}_\theta\left(T_{k+1}-T_k>t\,|\,X_{T_k},T_k\right)=e^{-H_{\theta,k}(X_{T_k},t)}$. This leads to a sequential simulation algorithm for the $\mathrm{P}_\theta$-jump times $(T_k)_{k\geq 1}$ via a sequence of i.i.d. standard exponentials $(\xi_k)_{k\geq 1}$. The times $(T_k)$ generate the path of the point process $N$ under $\mathrm{P}_\theta$, but require using (16) to solve $H_{\theta,k}(X_{T_k},\cdot)=\xi_{k+1}$ in terms of $q$ solving the ODE in (15). The latter admits no closed-form solution. We can only compute $q(t_j)$ at a discretization point $t_j$ by using a numerical ODE solver. We describe this approach, which generates simulation bias, and then explain why exact samples via thinning are not practically useful.

Consider times $0=t_0<t_1<t_2<\cdots<t_J=T$ for some integer $J\geq 1$ and spacings $\delta_j=t_{j+1}-t_j$. Let $q_\dagger$ be a right-continuous step-function approximation of $q$ which is constant on each $[t_j,t_{j+1})$. Analogously to (16) define,

$$H^\dagger_{\theta,k}(x,t)=\int_0^t e^{\theta+vq_\dagger(T_k+s)}h(x,s)\,\mathrm{d}s \tag{18}$$

so that $H^\dagger_{\theta,k}(x,t)$ approximates $H_{\theta,k}(x,t)$ but is easily evaluated. In particular,

$$H^\dagger_{\theta,k}(x,t_{j+1})-H^\dagger_{\theta,k}(x,t_j)=e^{\theta+vq_\dagger(t_j)}H(x,\delta_j) \tag{19}$$

**Algorithm 1 (Discretization + Thinning).** Generates a (biased) sample $T_{k+1}$ of the $(k+1)$st jump time of $N$ under $\mathrm{P}_\theta$ given $X_{T_k}=x$, $T_k=s$ and exponential $\xi_{k+1}$ with discretization times $(t_j)$. Step 3 ensures $T_{k+1}$ is exact.

1. Find index $\ell$ such that $H^\dagger_{\theta,k}(x,t_\ell)\le\xi_{k+1}$ and $H^\dagger_{\theta,k}(x,t_{\ell+1})>\xi_{k+1}$.
2. Solve $H^\dagger_{\theta,k}(x,t_\ell)+e^{\theta+vq(t_\ell)}H(x_\ell,t)=\xi_{k+1}$ where $x_\ell=h(x,t_\ell-s)$ via,

$$t=T_{k+1}=s+\frac{a+W_0(be^{-a})}{\kappa}$$
$$b=\frac{x-\mu}{\mu}$$
$$a=\frac{\xi_{k+1}-H^\dagger_{\theta,k}(x,t_\ell)}{e^{\theta+vq_\dagger(t_\ell)}\mu/\kappa}-b$$

   where $W_0$ is the principal branch of the Lambert $W$ function.
3. (Thinning*) Accept $T_{k+1}$ as the exact jump time with probability,

$$e^{\theta+v\,q(T_{k+1}-)}/e^{\theta+v\,q_\dagger(T_{k+1}-)}. \tag{17}$$

   and reject otherwise, restarting from scratch at step 1.

*Valid if $q\le q_\dagger$ and note, $q(T_{k+1})=q(T_{k+1}-)$ and $q_\dagger(T_{k+1})=q_\dagger(T_{k+1}-)$ almost surely.

where $H(x,t)=\int_0^t h(x,s)\,\mathrm{d}s=\mu t-\left(\frac{\mu-x}{\kappa}\right)(1-e^{-\kappa t})$ and a telescoping sum can be used to compute (18). $H$ is the P-cumulative hazard function of $T_{k+1}$.

Algorithm 1 samples the jump times $(T_k)$ sequentially by solving the approximation equation $H^\dagger_{\theta,k}(X_{T_k},t)=\xi_{k+1}$. These samples $(T_k)$ yield a biased path of $N$ on $[0,T]$ under $\mathrm{P}_\theta$. This allows us to assemble a biased estimator (5) and the cumulant generating function $\Lambda_\mu(\theta)$ may be precomputed by solving (15). Appendix C details Step 2 that involves the Lambert $W$ function.

The clear problem with Steps 1–2 of Algorithm 1 is that the $(T_k)$ are biased samples. These are jump times of a point process $N$ with intensity $e^{\theta+vq_\dagger}X$ (not $e^{\theta+vq}X$) which may be "*thinned*" to obtain an exact samples, but at computational cost. Thinning is a form of acceptance rejection in which we accept the sample $T_{k+1}$ with probability (17). This is accomplished by Step 3 (Thinning), but requires a call to the ODE solver of (15), and since the $T_k$ are random, the values $q(T_k)$ cannot be precomputed, leading to impractically long algorithm run times.

## 4 EXACT AND OPTIMAL IMPORTANCE SAMPLING

We develop an importance sampling scheme that approximates exponential twisting, is computationally less costly than a thinning approach for the latter, and produces exact samples under the importance measure. The basic idea is to sample a point process with intensity $e^{\theta+vq_\dagger}X$ for a step-function $q_\dagger$ approximating $q$ in (15) (per Algorithm 1), and then to correct for the resulting bias by a set of weights that are obtained from a change of measure.

To this end, the process that performs the bias correction is given by

$$W_t(\theta)=\exp\left(\int_0^t\left(e^{\theta+vq(s)}-e^{\theta+vq_\dagger(s)}\right)X_s\,\mathrm{d}s\right)\prod_{k=1}^{N_t}\frac{e^{v\,q_\dagger(T_k)}}{e^{v\,q(T_k)}} \tag{20}$$

where we recall that the $(T_k)$ denote the jump times of $N$, and we define

$$Z_T^\dagger(\theta) = Z_T(\theta) W_T(\theta) \qquad (\theta \in \mathbb{D}_N^o) \tag{21}$$

where $Z_T(\theta) = e^{\theta N_T - \Lambda_\mu(\theta)}$ is the Radon-Nikodym for exponential twisting.

Provided $Z_T^\dagger(\theta)$ is a Radon-Nikodym derivative, we remark that it has the decomposition $\frac{d\mathrm{P}_\theta}{d\mathrm{P}} \times \frac{d\mathrm{P}_\theta^\dagger}{d\mathrm{P}_\theta}$ where $\mathrm{P}_\theta^\dagger = Z_T^\dagger(\theta)\mathrm{P}$ (c.f. above (7)). This decomposition highlights that the weight $W_T = \frac{d\mathrm{P}_\theta^\dagger}{d\mathrm{P}_\theta}$ is a Radon-Nikodym derivative.

The following theorem is an analog of Proposition 1 that performs the extra work to ensure $\mathrm{P}_\theta^\dagger$ is a probability (this was readily granted for $\mathrm{P}_\theta$).

**Theorem 2** Let $\theta \in \mathbb{D}_N^o$. Then, for $\sup_{t\le T} |q_\dagger(t) - q(t)|$ sufficiently small, $\mathrm{P}_\theta^\dagger = Z_T^\dagger(\theta)\mathrm{P}$ is an equivalent probability under which $X$ solves $(1)$ with a point process $N$ that admits the $\mathrm{P}_\theta^\dagger$-intensity $e^{\theta + v q_\dagger} X$. Moreover,

$$Z_T^\dagger(\theta) = \exp\Big(\int_0^T \big(1 - e^{\theta + v q_\dagger(s)}\big) X_s \, ds\Big) \prod_{k=1}^{N_T} e^{\theta + v\, q_\dagger(T_k)}. \tag{22}$$

**Remark 1** $\delta = \sup_{t\le T} |q_\dagger(t) - q(t)|$ sufficiently small is judged relative to the distance $\theta_+ - \theta$ where $\theta_+$ is the largest value for which $\Lambda_1(\theta_+)$ is finite. Our proofs guarantee that taking $v\delta(1 + 2e^{\theta + \sup_{t\le T} q(t)}) < \theta_+ - \theta$ suffices.

The proof is based on the theory of stochastic exponentials and is deferred to Appendix A. Using (22), we define the importance sampling estimator,

$$Q_{\theta,\mu}^\dagger = 1_{\mathscr{E}_\mu} \exp\Big(\int_0^T \big(e^{\theta + v q_\dagger^\mu(s)} - 1\big) X_s^\mu \, ds\Big) \prod_{k=1}^{N_T^\mu} e^{-(\theta + v\, q_\dagger^\mu(T_k^\mu))} \tag{23}$$

where $W_T^\mu$ denotes $W_T$ in (20) with $X$ replaced by the $X^\mu$ (solving (3)) with jump times $(T_k^\mu)$ and $q_\dagger$ replaced by $q_\dagger^\mu$, a step-function approximation over a discretization $0 = t_0^\mu < t_1^\mu < \cdots < t_{J_\mu}^\mu = T$ for an integer sequence $J_\mu \to \infty$ and with $\max_j (t_{j+1}^\mu - t_j^\mu) \to 0$ as $\mu \to \infty$. A standard choice is $t_j^\mu = jT/\lceil \mu \rceil$.

We remark that $Z_T^\dagger(\theta)$ in (22) (but not (21)) is essential for exact sampling. In particular, (22) may be exactly evaluated (see below), and the values of the function $q_\dagger$ may be precomputed with by solving (15) at finitely many points.

We note that $Q_{\theta,\mu}^\dagger = Q_{\theta,\mu}/W_T^\mu(\theta)$ for $Q_{\theta,\mu}$ in (5) which highlights the connection to exponential twisting. By the equivalence of the measure change,

$$\mathrm{E}_\theta^\dagger(Q_{\theta,\mu}^\dagger) = \mathrm{P}(\mathscr{E}_\mu)$$

where $\mathrm{E}_\theta^\dagger$ denotes an expectation with respect to $\mathrm{P}_\theta^\dagger$. This estimator further matches the variance reduction properties of ECM (c.f., Theorem 1).

**Theorem 3** (Logarithmic efficiency) $Q_{\theta,\mu}^\dagger$ achieves $f = 2$ in (12), i.e.,

$$\liminf_{\mu\to\infty} \frac{\log \mathrm{E}_\theta^\dagger\big((Q_{\theta,\mu}^\dagger)^2\big)}{\log \mathrm{P}(\mathscr{E}_\mu)} = 2$$

provided $\sup_{t\le T} |q_\dagger^\mu(t) - q(t)| \to 0$ and $\theta \in \mathbb{D}_N^o$ solves $\nabla\Lambda_1(\theta) = \gamma$ in (9).

We explain why estimator in (23) may be evaluated exactly (in the form stated). Taking $q_\dagger$ to be step functions with $q_\dagger(t) = q(t_j)$ for $t_j \le t < t_{j+1}$, their values may be precomputed by solving the ODEs (15) prior to the simulation. In this way, the random $q_\dagger(T_k)$ do not inflict the computational costs that the $q(T_k)$ in (17) have. Letting $(\tau_j)$ be the sorted times $(t_j) \cup (T_k)$, we have

$$\int_{\tau_j}^{\tau_{j+1}} \left(e^{\theta + v q_\dagger(s)} - 1\right) X_s \, \mathrm{d}s = \left(e^{\theta + v q_\dagger(\tau_j)} - 1\right) H(X_{\tau_j}, \tau_{j+1} - \tau_j)$$

with $H(x,t)$ given below (19). Therefore, the integral in (23) may also be evaluated exactly. Indeed Steps 1–2 of Algorithm 1 may still be used to generate the jump times $(T_k)$. These are exact under $\mathrm{P}_\theta^\dagger$ (but not under $\mathrm{P}_\theta^\dagger$ without thinning Step 3). The correction stems from the Radon-Nikodym that $\mathrm{P}_\theta^\dagger$ uses.

Lastly, conditional Monte Carlo may be applied for faster algorithm run times. Since $\mathscr{E}_\mu$ is $\mathscr{F}_{T \wedge T_{\lceil \gamma\mu \rceil}}$-measurable (for $(\mathscr{F}_t)_{t \ge 0}$ the natural filtration of $N^\mu$),

$$\mathrm{E}_\theta^\dagger \left(Z_T^\dagger(\theta) \,|\, \mathscr{F}_{T \wedge T_{\lceil \gamma\mu \rceil}}\right) = Z_{T \wedge T_{\lceil \gamma\mu \rceil}}^\dagger(\theta) \tag{24}$$

which means that the algorithm may be terminated when we reach $\gamma\mu$ events (i.e., (4)). This only requires we replace $T$ in (23) by $T \wedge T^\mu_{\lceil \gamma\mu \rceil}$, i.e, not all of the jumps simulated by exponential twisting are necessary. And since conditional Monte Carlo always reduces variance, simulating them is counterproductive.

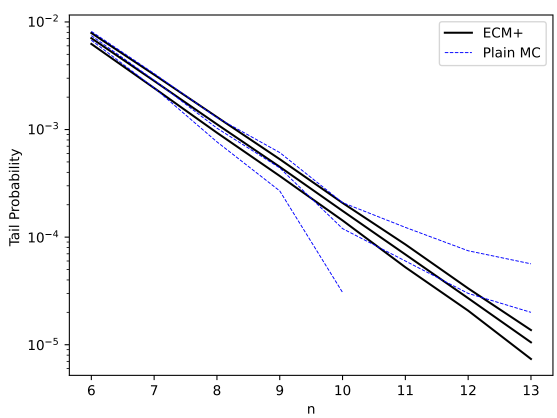


Figure 1: Log-log plot of the probability $\mathrm{P}(N_T \ge n)$ versus $n$ for ECM+ and plain Monte Carlo with 99% confidence intervals. Lower confidence limits below zero are omitted. Model parameters are in Table 2.

## 5 NUMERICAL EXAMPLES

We present two examples. The first compares the exponential twisting estimator $Q_{\theta,\mu}$ in (5) (referred to as ECM) and our exact estimator $Q_{\theta,n}^\dagger$ in (23) (referred to as ECM+). The second illustrates the variance reduction of $Q_{\theta,\mu}^\dagger$ relative to the plain Monte Carlo estimator $1_{\{N_T \ge n\}}$ of $z = \mathrm{P}(N_T \ge n)$.

Our implementation of ECM follows Algorithm 1 with a uniform discretization (i.e., $t_j - t_{j-1} = \delta$) without thinning (Step 3). Both ECM and ECM+ compute the twist $\theta$ via $\Lambda_\mu(\theta) = n$. For each probability estimate $\hat{z}_m$, we compute (upper and lower) %99 confidence intervals via $\pm 2.58\,\hat{\sigma}_m/\sqrt{m}$ where $m$ is the number of trials and $\hat{\sigma}_m$ is the sample variance of the estimator. The latter is used to compute the variance reduction ratio (VRR) for any two methods. The relative error (RE) is estimated via $\hat{\sigma}_m/\hat{z}_m$. The relative bias is computed as $|\hat{z}_m - z|/z$ where $z$ is computed to a high precision with $M \gg m$ trials.

Table 1 reports on the simulation bias of the ECM. The estimated probabilities of ECM are not contained by the confidence intervals produced by ECM+ until $\delta$ becomes small. The run times of ECM in the last row of the table is almost three times slower than for ECM+. For the earlier rows, we see that the ECM estimator generates significant simulation bias. The corresponding variance reduction ratios are deceptive. While ECM appears to have lower variance, this is due to the large downward bias of the its probability estimate. The apparent differences in variance reduction vanish for smaller $\delta$ values.

Table 2 reports on the variance reduction obtained by ECM+ relative to plain Monte Carlo. ECM+ results in a dramatic variance reduction. Figure 1 augments Table 2 with probability estimates and confidence intervals.

Table 1: Bias of the ECM probability $\mathrm{P}(N_T \geq 8)$ estimate for model parameters $\mu = 0.3, \kappa = 0.3, v = 1.0$ and $X_0 = 0.15$. The number of trials for ECM and ECM+ is set to $10^6$ per row. The lower (LCI) and upper (UCI) limits of the 99% confidence intervals for ECM+ are in columns 3–4.

| $\delta$ | ECM Prob. | ECM+ LCI | ECM+ UCI | VR | REL. BIAS (%) |
|---|---|---|---|---|---|
| $2^{-2}$ | 5.678e–04 | 8.787e–04 | 1.198e–03 | 0.003 | 45.33 |
| $2^{-3}$ | 9.029e–04 | 1.097e–03 | 1.150e–03 | 0.192 | 19.63 |
| $2^{-4}$ | 1.064e–03 | 1.101e–03 | 1.130e–03 | 0.793 | 4.58 |
| $2^{-5}$ | 1.105e–03 | 1.098e–03 | 1.125e–03 | 0.959 | 0.57 |
| $2^{-6}$ | 1.117e–03 | 1.102e–03 | 1.128e–03 | 0.993 | 0.15 |

Table 2: Variance reduction ratio (VRR) and relative error (RE) estimates obtained by ECM+ relative to plain Monte Carlo for model parameters $\mu = 0.5, \kappa = 0.25, v = 0.5$ and $X_0 = 0.75$. The reported probability and confidence intervals (C.I.) is computed by ECM+ with $\delta = 1/n$. Plain Monte Carlo is run for 100,000 trials to obtain results for all rows simultaneously and ECM+ takes 1/8th the time budget to complete each row.

| $n$ | $\mathrm{P}(N_T \geq n)$ | 99% C.I. | RE | VRR | # trials |
|---|---|---|---|---|---|
| 6 | 6.77e–03 | 8.24e–04 | 1.15e+01 | 4.28e+01 | 1,700 |
| 7 | 2.84e–03 | 4.03e–04 | 1.86e+01 | 8.42e+01 | 1,400 |
| 8 | 1.07e–03 | 1.86e–04 | 3.11e+01 | 1.65e+02 | 1,200 |
| 9 | 4.40e–04 | 8.00e–05 | 4.77e+01 | 4.16e+02 | 1,100 |
| 10 | 1.45e–04 | 3.25e–05 | 9.13e+01 | 7.55e+02 | 1,000 |
| 11 | 7.48e–05 | 1.65e–05 | 1.29e+02 | 1.62e+03 | 900 |
| 12 | 2.47e–05 | 6.33e–06 | 1.83e+02 | 6.22e+03 | 800 |
| 13 | 1.12e–05 | 3.16e–06 | 2.24e+02 | 2.23e+04 | 600 |

**ACKNOWLEDGMENTS**

We thank Baeho Kim for suggestions regarding the implementation of our algorithms.

## A PROOF OF THEOREMS 2 & 3

The next two lemmas are needed for the proof of both theorems. Their proofs are omitted for brevity, but these bounds follow from first principles.

**Lemma 2** $\sup_{t \leq T} X_t \leq \max(\mu, X_0) + vN_T$.

**Lemma 3** Suppose $\sup_{t \leq T} |q_{\dagger}(t) - q(t)| \leq \delta$ for $\delta > 0$. Then for any $\alpha \in \mathbb{R}$, there exist constants $c_\alpha, C_\alpha > 0$ with $\sup_{t \leq T} W_t^\alpha \leq \exp\left(c_\alpha \mu \delta + C_\alpha \delta N_T\right)$.

**Remark 2** The constants $c_\alpha, C_\alpha$ of Lemma 3 depend on $\alpha, T$ and the parameters $\kappa$ and $v$ in (1) (but not $\mu$ or $\delta$).

*Proof of Theorem 2.* By Brémaud (1981), Chapter VI, T2 the process $W(\theta)$ defined in (20) is a local $\mathrm{P}_\theta$-martingale. Here, we remark that $q$ is continuous and $q_\dagger(T_k-) = q_\dagger(T_k)$ almost surely. Below, we show $W(\theta)$ is further a $\mathrm{P}_\theta$-martingale for $\theta \in \mathbb{D}_N$, so that $\mathrm{E}_\theta(W_T(\theta)) = 1$ and it then follows that the point process $N$ has $\mathrm{P}_\theta^\dagger$-intensity $e^{\theta+vq_\dagger}X$ (Brémaud 1981, Chapter VI, T3).

By Protter (2005), Chapter 1, Theorem 51 the local $\mathrm{P}_\theta$-martingale $W(\theta)$ is a $\mathrm{P}_\theta$-martingale provide that $\mathrm{E}_\theta\left(\sup_{t\le T} W_t(\theta)\right) < \infty$.

Applying Lemma 3 with $\alpha = 1$ and $\delta > 0$, we obtain

$$\begin{aligned}\mathrm{E}_\theta\big(\sup_{t\le T} W_t(\theta)\big) &= \mathrm{E}\big(Z_T \sup_{t\le T} W_t(\theta)\big) = \mathrm{E}\big(\sup_{t\le T} W_t(\theta)\exp(\theta N_T - \Lambda_\mu(\theta))\big)\\ &\le \mathrm{E}\big(\exp(c_1\delta\max(\mu,X_0) - \Lambda_\mu(\theta) + (\theta + C_1\delta)N_T)\big)\\ &= \exp\big(c_1\delta\max(\mu,X_0) - \Lambda_\mu(\theta) + \Lambda_\mu(\theta + C_1\delta)\big)\end{aligned}$$

Choosing $\delta > 0$ sufficiently small so that $\theta + C_1\delta \in \mathbb{D}_N$ yields the desired martingale property for $W(\theta)$. Now, $\mathrm{P}_\theta^\dagger = W_T(\theta)\mathrm{P}_\theta$ defines a change of measure which may be written as $\mathrm{P}_\theta^\dagger = Z_T(\theta)W_T(\theta)\mathrm{P}$ by definition of $\mathrm{P}_\theta^\dagger$.

We turn to establishing that $Z_T^\dagger(\theta) = Z_T(\theta)W_T(\theta)$ equals the right side of (22). Since $X$ and $q_\dagger$ that determine (22) are almost surely finite on $[0,T]$, we have $\mathrm{P}(Z_T^\dagger(\theta) > 0) = 1$ so that $\mathrm{P}_\theta^\dagger$ and $\mathrm{P}$ are equivalent measures as claimed.

We consider the Doob martingale $Z = Z(\theta)$ in (27), the density of the Radon-Nikodym derivative $Z_T(\theta) = \frac{\mathrm{dP}_\theta}{\mathrm{dP}}$ of the ECM. We have,

$$Z_t(\theta) = \exp\left(\theta N_t + \int_0^t \dot{q}(s)X_s\,\mathrm{d}s + \int_0^t q(s)\,\mathrm{d}X_s + p(t) - p(0)\right)$$

using that $\Lambda_\mu(\theta) = p(0) + q(0)X_0$ and the integration by parts formula,

$$q(t)X_t = q(0)X_0 + \int_0^t q(s)\,\mathrm{d}X_s + \int_0^t X_s\dot{q}(s)\,\mathrm{d}s \tag{25}$$

Continuing by substituting the expression for $p$ and $\dot{q}$ from (15),

$$\begin{aligned}Z_t(\theta) &= \exp\left(\theta N_t + \int_0^t (\kappa q(s) - e^{\theta+vq(s)} - 1)X_s\mathrm{d}s + \int_0^t q(s)\,\mathrm{d}X_s - \int_0^t \kappa\mu q(s)\mathrm{d}s\right)\\ &= \exp\left(\theta N_T + \int_0^t (1 - e^{\theta+vq(s)})X_s\,\mathrm{d}s + \int_0^t q(s)\,\mathrm{d}X_s - \int_0^t q(s)\kappa(\mu - X_s)\,\mathrm{d}s\right)\end{aligned}$$

and using that $\mathrm{d}X_t = \kappa(\mu - X_t)\,\mathrm{d}t + v\,\mathrm{d}N_t$ (see (3)) yields,

$$\begin{aligned}Z_t(\theta) &= \exp\left(\theta N_T - \int_0^t (e^{\theta+vq(s)} - 1)X_s\mathrm{d}s + v\int_0^t q(s)\,\mathrm{d}N_s\right)\\ &= \exp\left(\int_0^t (X_s - e^{\theta+vq(s)}X_s)\mathrm{d}s\right)\prod_{k=1}^{N_t} e^{\theta+vq(T_k)}.\end{aligned}$$

Considering $W_T(\theta)$ per (20), it follows by direct calculation that $Z_T(\theta)W_T(\theta)$ equals the right side of (22) as required.

□

*Proof of Theorem 3.* By definition, $Q^{\dagger}_{\theta,\mu} = W^{\mu}_T 1_{\mathscr{E}_\mu}/Z^{\mu}_T = \frac{\mathrm{dP}}{\mathrm{dP}^{\dagger}_{\theta}} 1_{\mathscr{E}_\mu}$ as guaranteed by Theorem 2. Let $\theta > 0$ be such that $\Lambda_{N_T}(\theta) = \gamma$ (per Lemma 1).

Passing from $\mathrm{P}^{\dagger}_{\theta}$ to P we obtain,

$$\begin{aligned}
\frac{1}{\mu}\log \mathrm{E}^{\dagger}_{\theta}((Q^{\dagger}_{\theta,\mu})^2) &= \frac{1}{\mu}\log \mathrm{E}\left(\frac{W^{\mu}_T}{Z^{\mu}_T} 1_{\mathscr{E}_\mu}\right) = \frac{1}{\mu}\log \mathrm{E}\left(e^{-\theta N^{\mu}_T + \Lambda_{N^{\mu}_T}(\theta)} W^{\mu}_T 1_{\mathscr{E}_\mu}\right) \\
&\le -\frac{1}{\mu}(\theta\gamma\mu - \Lambda_\mu(\theta)) + \frac{1}{\mu}\log \mathrm{E}(W^{\mu}_T 1_{\mathscr{E}_\mu}) \\
&\le -\Lambda^*_1(\gamma) + \frac{1}{\mu}\log \mathrm{E}(W^{\mu}_T 1_{\mathscr{E}_\mu}) \qquad (26)
\end{aligned}$$

where $\Lambda^*_1$ is defined in (10) and the last step uses Lemma 1.

Let $\rho > 1$. Applying Hölder's inequality and Lemma 3 with $\alpha = \frac{\rho}{\rho-1}$,

$$\begin{aligned}
\frac{1}{\mu}\log \mathrm{E}(W^{\mu}_T 1_{\mathscr{E}_\mu}) &\le \frac{1}{\mu}\log\left(\left(\mathrm{E}(1^{\rho}_{\mathscr{E}_\mu})\right)^{\frac{1}{\rho}} \left(\mathrm{E}((W^{\mu}_T)^{\alpha})\right)^{\frac{1}{\alpha}}\right) \\
&= \frac{1}{\rho\mu}\log \mathrm{P}(\mathscr{E}_\mu) + \frac{1}{\alpha\mu}\log \mathrm{E}((W^{\mu}_T)^{\alpha}) \\
&\le \frac{1}{\rho\mu}\log \mathrm{P}(\mathscr{E}_\mu) + c_\alpha\delta_\mu + \frac{1}{\mu\alpha}\log \mathrm{E}(e^{\alpha C_\alpha \delta_\mu N^{\mu}_T}) \\
&\le \frac{1}{\rho\mu}\log \mathrm{P}(\mathscr{E}_\mu) + c_\alpha\delta_\mu + \frac{1}{\mu\alpha}\Lambda_\mu(\alpha C_\alpha \delta_\mu) \\
&= \frac{1}{\rho\mu}\log \mathrm{P}(\mathscr{E}_\mu) + c_\alpha\delta_\mu + \frac{1}{\alpha}\Lambda_1(\alpha C \delta_\mu)
\end{aligned}$$

where $\delta_\mu = \max_j (t^{\mu}_{j+1} - t^{\mu}_j)$ and we applied Lemma 1 in the last step.

As $\mu \to \infty$ we have $\delta_\mu \to 0$, and by continuity of $\Lambda_1$, we further obtain that $\Lambda_1(\alpha C_\alpha \delta_\mu) \to 0$. With $\rho$ (and hence $\alpha$) fixed, taking $\mu \to \infty$,

$$\limsup_{\mu\to\infty} \frac{1}{\mu}\log \mathrm{E}(W^{\mu}_T 1_{\mathscr{E}_\mu}) \le \frac{1}{\rho}\limsup_{\mu\to\infty}\frac{1}{\mu}\log \mathrm{P}(\mathscr{E}_\mu) = -\frac{1}{\rho}\Lambda^*_1(\gamma)$$

As this inequality holds for all $\rho > 1$, taking $\rho \downarrow 1$ and using (26), we obtain

$$\limsup_{\mu\to\infty}\frac{1}{\mu}\log \mathrm{E}^{\dagger}_{\theta}((Q^{\dagger}_{\theta,\mu})^2) \le -2\Lambda^*_1(\gamma).$$

The same argument as in the proof of Theorem 1 completes the proof. □

## B AUXILIARY PROOFS

*Proof of Proposition 1.* We use Girsanov-Meyer theory (Protter 2005, Chapter III.8). Consider the Doob martingale $Z$ defined via $Z_t = \mathrm{E}(Z_T(\theta) \mid \mathscr{F}_t)$,

$$Z_t = \exp\left(\theta N_t + p(t) + q(t)X_t - \Lambda_\mu(\theta)\right) \qquad (27)$$

where we applied (14). Here, $(\mathscr{F})_{t\ge 0}$ is the filtration generated by $N$.

Let $A_t = \int_0^t X_s \,\mathrm{d}s$, the compensator of the point process $N$. We find the adjustment of the local P-martingale $M = N - A$ upon passing to $\mathrm{P}_\theta$.

Computing the quadratic variation $[M,Z]$ of $M$ and $Z$ yields,

$$[M,Z]_t = \sum_{s<t} \Delta N_s \Delta Z_s - \int_0^t Z_s \, dA_s = \int_0^t Z_{s-} e^{\theta + vq(s)} \, dN_s - \int_0^t Z_s X_s \, ds\,,$$

which has the predictable P-compensator, $\langle M,Z\rangle_t = \int_0^t Z_s(e^{\theta+vq(s)} - 1)X_s \, ds$. Applying the predictable version of the Girsanov-Meyer theorem (Protter 2005, Chapter III, Theorem 40) yields that $G$ is a local $\mathrm{P}_\theta$-martingale where

$$G_t = M_t - \int_0^t \frac{1}{Z_{s-}} d\langle M,Z\rangle_s = N_t - \int_0^t e^{\theta+vq(s)} X_s \, ds$$

This implies that the $\mathrm{P}_\theta$-intensity of $N$ is given by $e^{\theta+vq}X$ as claimed. □

## C INVERSE OF THE CUMULATIVE HAZARD FUNCTION

We compute the inverse of the cumulative hazard function $H$ appearing in (19) to justify Algorithm 1, i.e., for $\xi > 0$ we solve

$$H(x,t) = \mu t - \left(\frac{\mu - x}{\kappa}\right)(1 - e^{-\kappa t}) = \xi\,.$$

We have $t = \frac{\xi}{\mu} + \frac{\mu-x}{\kappa\mu} + \frac{x-\mu}{\kappa\mu} e^{-\kappa t}$. Letting $a = \frac{\kappa\xi}{\mu} - b$ where $b = \frac{x-\mu}{\mu}$, equivalently

$$t = \frac{a}{\kappa} + \frac{b}{\kappa} e^{-\kappa t}\,. \tag{28}$$

The solution for $t$ is well known in terms of the Lambert $W$ function, i.e.,

$$t = \frac{a + W(be^{-a})}{\kappa}$$

with $W$ computed in terms of two possible branches $W_0$ (called the principal branch) or $W_{-1}$. We show that (28) corresponds to the principal branch.

Since $t \mapsto H(x,t)$ increases from zero to infinity continuously and $\xi > 0$, there is only one solution of (28) which is positive. The principle branch $W_0$ always takes larger values than $W_{-1}$ and hence must correspond to this positive solution. In summary, we must have $t = a/\kappa + W_0(be^{-a})/\kappa$.

## REFERENCES


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

Bacry, E., I. Mastromatteo, and J.-F. Muzy. 2015. "Hawkes Processes in Finance". *Market Microstructure and Liquidity* 1(01):1550005.

Bowsher, C. G. 2007. "Modelling Security Market Events in Continuous Time: Intensity Based, Multivariate Point Process Models". *Journal of Econometrics* 141(2):876–912.

Brémaud, P. 1981. *Point Processes and Queues: Martingale Dynamics*. New York: Springer-Verlag, Inc.

Chavez-Demoulin, V., A. C. Davison, and A. J. McNeil. 2005. "Estimating Value-at-Risk: A Point Process Approach". *Quantitative Finance* 5(2):227–234.

Chen, G., A. Shkolnik, and K. Giesecke. 2019. "Unbiased Simulation Estimators for Jump-Diffusions". In *2019 Winter Simulation Conference*, 890–901 https://doi.org/10.1109/WSC40007.2019.9004767.

Chen, G., A. Shkolnik, and K. Giesecke. 2025. "Unbiased Simulation Estimators for Multivariate Jump-Diffusions". Working Paper.

Chen, X. 2021. "Perfect Sampling of Hawkes Processes and Queues with Hawkes Arrivals". *Stochastic Systems* 11(3):264–283.

Dassios, A., and H. Zhao. 2013. "Exact Simulation of Hawkes Process with Exponentially Decaying Intensity". *Electronic Communications in Probability* 18:1–13.

Dembo, A., and O. Zeitouni. 2010. *Large Deviations Techniques and Applications*. 2nd ed. Berlin Heidelberg: Springer, Inc.

El Maazouz, Y., and M. A. Bennouna. 2018. *Simulating Rare Events: Hawkes Process Applied to Twitter*. PhD Thesis, Ecole Polytechnique, Palaiseau, France.

Giesecke, K., and A. Shkolnik. 2022. "Reducing Bias in Event Time Simulations via Measure Changes". *Mathematics of Operations Research* 47(2):969–988.

Giesecke, K., and A. Shkolnik. 2025. "Asymptotically Optimal Importance Sampling for Default Timing". Working Paper.

Giesecke, K., and A. D. Shkolnik. 2010. "Importance Sampling for Indicator Markov Chains". In *2010 Winter Simulation Conference*, 2742–2750 https://doi.org/10.1109/WSC.2010.5678969.

Hawkes, A. G. 1971. "Spectra of Some Self-Exciting and Mutually Exciting Point Processes". *Biometrika* 58(1):83–90.

Hawkes, A. G. 1973. "Cluster Models for Earthquakes – Regional Comparisons". *Bulletin of the International Statistical Institute* 45(3):454–461.

Hewlett, P. 2006. "Clustering of Order Arrivals, Price Impact and Trade Path Optimisation". In *Workshop on Financial Modeling with Jump Processes*, 6–8. Ecole Polytechnique / Citeseer.

Keller-Ressel, M., and E. Mayerhofer. 2015. "Exponential Moments of Affine Processes". *The Annals of Applied Probability* 25:714–752.

Kirchner, M. 2017. "An Estimation Procedure for the Hawkes Process". *Quantitative Finance* 17(4):571–595.

Large, J. 2007. "Measuring the Resiliency of an Electronic Limit Order Book". *Journal of Financial Markets* 10(1):1–25.

Laub, P. J., Y. Lee, and T. Taimre. 2021. *The Elements of Hawkes Processes*. Cham: Springer, Inc.

Magris, M. 2019. "On the Simulation of the Hawkes Process via Lambert-W Functions". Working Paper.

Mei, H., and J. M. Eisner. 2017. "The Neural Hawkes Process: A Neurally Self-Modulating Multivariate Point Process". In *Advances in Neural Information Processing Systems*, Volume 31, 6757–6767. Curran Associates, Inc.

Møller, J., and J. G. Rasmussen. 2005. "Perfect Simulation of Hawkes Processes". *Advances in Applied Probability* 37(3):629–646.

Møller, J., and J. G. Rasmussen. 2006. "Approximate Simulation of Hawkes Processes". *Methodology and Computing in Applied Probability* 8:53–64.

Ogata, Y. 1981. "On Lewis' Simulation Method for Point Processes". *IEEE Transactions on Information Theory* 27(1):23–31.

Protter, P. 2005. *Stochastic Integration and Differential Equations*. 2nd ed. Berlin Heidelberg: Springer, Inc.

Shkolnik, A., K. Giesecke, G. Teng, and Y. Wei. 2024. "Numerical Solution of Jump-Diffusion SDEs". Operations Research (forthcoming).

Siegmund, D. 1976. "Importance Sampling in the Monte Carlo Study of Sequential Tests". *The Annals of Statistics* 4(4):673–684.

Zhang, X., J. Blanchet, K. Giesecke, and P. W. Glynn. 2015. "Affine Point Processes: Approximation and Efficient Simulation". *Mathematics of Operations Research* 40(4):797–819.

Zhang, X.-W., P. W. Glynn, K. Giesecke, and J. Blanchet. 2009. "Rare Event Simulation for a Generalized Hawkes Process". In *2009 Winter Simulation Conference*, 1291–1298 https://doi.org/10.1109/WSC.2009.5429693.

Zhu, L. 2013. "Central Limit Theorem for Nonlinear Hawkes Processes". *Journal of Applied Probability* 50(3):760–771.

## AUTHOR BIOGRAPHIES

**LIANG FENG** Liang Feng is a PhD candidate at the Department of Statistics and Applied Probability at University of California, Santa Barbara. His research topic is Monte Carlo simulation of stochastic process. His email address is liangfeng@ucsb.edu.

**ALEX SHKOLNIK** Alex Shkolnik is an Assistant Professor at the Department of Statistics and Applied Probability at the University of California, Santa Barbara and a Research Fellow at the Consortium for Data Analytics in Risk at the University of California, Berkeley where he was a postdoctoral scholar. Alex received his PhD in computational mathematics and engineering from Stanford University in 2015. His research interests include Monte Carlo simulation of stochastic processes, high-dimensional statistics and quantitative financial risk management. His email address is shkolnik@ucsb.edu.