Article

Coarse-Grained Hawkes Processes

Shinsuke Koyama [*]

Article

Coarse-Grained Hawkes Processes

Shinsuke Koyama

Department of Interdisciplinary Statistical Mathematics, Institute of Statistical Mathematics, Tokyo 190-8562, Japan; skoyama@ism.ac.jp

Abstract: When analyzing real-world event data, it is often the case that bin-count processes are observed instead of precise event time-stamps along a continuous timeline, owing to practical limitations in measurement accuracy. In this work, we propose a modeling framework for aggregated event data generated by multivariate Hawkes processes. The introduced model, termed the coarse-grained Hawkes process, effectively captures the second-order statistical characteristics of the bin-count representation of the Hawkes process, particularly when the bin size is large relative to the typical support of the excitation kernel. Building upon this model, we develop a method for inferring the underlying Hawkes process from bin-count observations, and demonstrate through simulation studies that the proposed approach performs comparably to, or even surpasses, existing techniques, while maintaining computational efficiency in parameter estimation.

Keywords: Hawkes process; aggregated data; count time series; coarse-grained modeling

---

1. Introduction

Generally, past events in natural and social systems facilitate the occurrence of future events through a self-exciting mechanism. The Hawkes process [1,2], a class of self-exciting point processes, has become a popular tool for modeling such processes continuously. Its defining intensity function, or instantaneous event rate, is composed of a baseline rate augmented by the cumulative influence of prior events, thereby capturing the self-exciting mechanism wherein each event elevates the probability of subsequent occurrences. Owing to the prevalence of self-excitation across diverse domains, the Hawkes process has been extensively employed in a wide range of disciplines, including seismology [3,4], neurophysiology [5,6], genomics [7], finance [8,9], social media analytics [10–12], criminology [13,14], terrorism studies [15], and traffic incident analysis [16].

The Hawkes process is applicable to scenarios wherein individual events are distinguishable, as it characterizes a series of discrete occurrences along a continuous timeline. Nevertheless, it proves insufficient for modeling count-based sequences in which the precise timing of events is unobserved and only their aggregated counts within successive intervals are available. A representative case arises in epidemiological studies, where individual infection events are not monitored in real time, but rather, the daily incidence rates are recorded. Consequently, there is a need to develop a count-based time series model that retains the salient features of the Hawkes process for such contexts.

Multiple methodologies have been proposed for fitting the Hawkes process to bin-count data. Kirchner demonstrated that the distribution of bin-count sequences generated by Hawkes processes can be effectively approximated using an integer-valued autoregressive (INAR) model, wherein conditional least squares estimation is employed to infer the underlying Hawkes process [17,18]. Shlomovich et al. introduced a Monte Carlo expectation-maximization (MC-EM) framework, which integrates an efficient sampling algorithm for the latent event times with an EM procedure to maximize the likelihood function [19,20]. Similarly, Chen et al. developed a Pseudo-Marginal Metropolis-Hastings (PMMH) algorithm for maximum likelihood estimation, wherein the likelihood is approximated via a sequential Monte Carlo approach [21]. Alternatively, Cheysson and Lang advocated for a spectral estimation technique grounded in the Whittle likelihood within the univariate context [22].

Although Kirchner's approach provides consistent and asymptotically normal estimators for the underlying Hawkes process as the bin width approaches zero, it introduces bias when the bin size exceeds the typical support of the excitation function, due to the INAR model's neglect of intra-bin excitation dynamics. In contrast, the MC-EM, PMMH and spectral estimation methods yield less biased results by accounting for excitation effects occurring within each bin. Nonetheless, the MC-EM and PMMH techniques are computationally intensive, whereas the spectral method is validated only in the univariate setting.

Accordingly, the objective of this study is to develop a computationally efficient methodology that extends naturally to the multivariate setting. Our proposed approach involves constructing a count-based time series model that approximates the bin-count sequence generated by Hawkes processes, upon which an estimation procedure is built. Unlike Kirchner's method, which relies on a straightforward discretization, namely, evaluating the intensity function at discrete time points, to relate the Hawkes and INAR processes [17], our model is derived through a coarse graining procedure. The resulting framework, termed the coarse-grained Hawkes process, is defined in a conceptually simple and analytically tractable form, while effectively preserving the second-order statistical characteristics of the bin-count Hawkes process, even in regimes where the bin size is large relative to the excitation function's effective range.

The structure of this paper is as follows. Section 2 provides a concise overview of Hawkes processes. Section 3 introduces the coarse-grained Hawkes process along with the proposed estimation methodology. In Section 4, we assess the approximation accuracy of the coarse-grained Hawkes process in representing the bin-count Hawkes process. Additionally, a simulation study is conducted to benchmark the performance of the proposed estimation technique against existing approaches. Section 5 concludes with a discussion of the findings.

2. Review of Hawkes Processes

In this section, we present a brief review of Hawkes processes. Let $\{t_i\}_{i \in \mathbb{N}}$ denote a sequence of non-negative random variables representing the occurrence times of events on $\mathbb{R}_+$, satisfying $t_i < t_{i+1}$ for $i \in \mathbb{N}$. Define the associated counting process by $N(t) = \sum_{i \in \mathbb{N}} \mathbf{1}_{t_i \leq t}$, and let $\mathcal{H}_t = \{N(u) : u < t\}$ denote the history of events up to, but not including, time t. We consider a point process such that

$$P(N(t + \delta t) - N(t) = 1 \mid \mathcal{H}_t) = \lambda(t)\delta t + o(\delta t),$$
$$P(N(t + \delta t) - N(t) > 1 \mid \mathcal{H}_t) = o(\delta t),$$

as $\delta t \to 0$, where $\lambda(t)$ is the conditional intensity function that uniquely characterizes the point process. A univariate Hawkes process is then defined by the conditional intensity function

$$\lambda(t) = \mu + \int_0^t \phi(t - u)dN(u),$$

where $\mu \geq 0$ is the baseline intensity and $\phi(\cdot)$ is a non-negative excitation kernel satisfying $\phi(t) = 0$ for $t < 0$ [1]. This formulation captures the self-exciting nature of the process, wherein the intensity at any given time depends on the historical sequence of events. The branching ratio is defined as

$$\alpha := \int_0^\infty \phi(t)dt,$$

which quantifies the expected number of subsequent events triggered by a single occurrence. The process admits a stationary distribution provided that $\alpha < 1$.

The univariate Hawkes process naturally generalizes to the multivariate case [1,2]. For $d \in \mathbb{N}$, a d-dimensional Hawkes process $\boldsymbol{N}(t) = (N_1(t), \ldots, N_d(t))^\mathsf{T}$ comprises d jointly defined point processes on $\mathbb{R}_+$. The corresponding vector-valued conditional intensity function is given by

$$\boldsymbol{\lambda}(t) = \boldsymbol{\mu} + \int_0^t \Phi(t-s)d\boldsymbol{N}(s),$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)^\mathsf{T} \in \mathbb{R}^d_{\geq 0}$ is the baseline intensity vector, and the excitation kernel $\Phi = (\phi_{ij})_{1 \leq i,j \leq d}$ is a matrix-valued function with non-negative entries satisfying $\phi_{ij}(t) = 0$ for $t < 0$. Each entry $\phi_{ij}(\cdot)$ describes the influence of process j on the intensity of process i, thus capturing both self-excitation and mutual excitation among the components. The multivariate Hawkes process is asymptotically stationary if the spectral radius of the branching matrix

$$A := \int_0^\infty \Phi(t)dt$$

is strictly less than one.

Finally, we briefly recall the martingale properties of point processes that are instrumental for our methodological developments. For a comprehensive treatment, see [23]. Define the process $\boldsymbol{M}(t) = \boldsymbol{N}(t) - \boldsymbol{\Lambda}(t)$, where $\boldsymbol{\Lambda}(t) = \int_0^t \boldsymbol{\lambda}(u)du$. Then $\boldsymbol{M}(t)$ is a martingale with respect to the filtration $\mathcal{H}_t$, satisfying the property that the conditional expectation of its increment is zero:

$$E[\boldsymbol{M}(t) - \boldsymbol{M}(s) \mid \mathcal{H}_s] = \boldsymbol{0}, \quad t > s.$$

Moreover, the conditional covariance matrix of the martingale increment is given by

$$E\left[(\boldsymbol{M}(t) - \boldsymbol{M}(s))(\boldsymbol{M}(t) - \boldsymbol{M}(s))^\mathsf{T} \mid \mathcal{H}_s\right] = \mathrm{diag}(E[\boldsymbol{\Lambda}(t) - \boldsymbol{\Lambda}(s) \mid \mathcal{H}_s]), \quad t > s,$$

which follows from the quadratic variation of the martingale.

3. Coarse-Grained Hawkes Process

3.1. Motivation

We consider a scenario wherein the exact timing of individual events is unobserved; instead, we observe an aggregated count of these latent continuous-time events within discrete time intervals. We define

$$\boldsymbol{X}_n = \boldsymbol{N}(n\Delta t) - \boldsymbol{N}((n-1)\Delta t), \quad n \in \mathbb{N},$$

as the bin-wise event counts over intervals of size Δt. When $\{\boldsymbol{N}(t) : t \in \mathbb{R}_+\}$ constitutes a Hawkes process, the corresponding discrete-time vector process $\{\boldsymbol{X}_n : n \in \mathbb{N}\}$ is referred to as the binned Hawkes process [19,20].

Although a closed-form representation of the probability distribution of binned Hawkes processes is not available, an approximate formulation can be derived by discretizing the conditional intensity function as follows:

$$\boldsymbol{\lambda}_n = (\lambda_{1n}, \ldots, \lambda_{dn})^\mathsf{T} = \boldsymbol{\mu}\Delta t + \sum_{k=1}^{n-1} \Phi(k\Delta t)\Delta t \boldsymbol{X}_k, \tag{1}$$

and assuming that the bin counts follow Poisson distributions:

$$\boldsymbol{X}_n \mid \boldsymbol{X}_1, \ldots, \boldsymbol{X}_{n-1} \sim \prod_{j=1}^d \mathrm{Poisson}(\lambda_{jn}). \tag{2}$$

This method, known as the binned Poisson approximation, converges in distribution to the true Hawkes process as $\Delta \to 0$. However, its accuracy degrades with larger Δt due to discretization errors in (1) and the conditional independence assumption in (2), which neglects the intra-bin excitation

effect. Nonetheless, it facilitates a tractable estimation procedure: the log-likelihood of the sequence $X_1, \ldots, X_n$ is given by

$$\log P(X_1, \ldots, X_n) = \sum_{k=1}^{n} \sum_{j=1}^{d} \left(X_{jk} \log \lambda_{jk} - \lambda_{jk} \right).$$

Parameter estimation is then performed via maximization of this log-likelihood function [19,20,24].

In light of the limitations of this approximation for larger Δt, we propose an alternative count time series model that more accurately approximates the binned Hawkes process while retaining computational simplicity for inference. We elucidate the underlying heuristic in the univariate setting.

The central idea is to replace the crude discretization with the expected value of the conditional intensity integrated over each time bin. The expected count within the nth bin is $\int_{(n-1)\Delta t}^{n\Delta t} \lambda(t)dt$, which depends on the trajectory $\{N(u) : 0 < u \leq n\Delta t\}$. Consequently, we consider the conditional expectation, given the bin counts $\{X_1, \ldots, X_n\}$:

$$\begin{aligned}
\lambda_n &= E\left[\int_{(n-1)\Delta t}^{n\Delta t} \lambda(t)dt \mid X_1, \ldots, X_n \right] \\
&= \sum_{k=1}^{n-1} E\left[\int_{(n-1)\Delta t}^{n\Delta t} \left(\int_{(k-1)\Delta t}^{k\Delta t} \phi(t - u)dN(u) \right) dt \mid X_k \right] \\
&\quad + E\left[\int_{(n-1)\Delta t}^{n\Delta t} \left(\int_{(n-1)\Delta t}^{t} \phi(t - u)dN(u) \right) dt \mid X_n \right].
\end{aligned} \tag{3}$$

As these conditional expectations do not admit closed-form expressions, we approximate them under the assumption that event times are uniformly distributed within each bin. Accordingly, the first term on the right-hand side of (3) is approximated as

$$\begin{aligned}
&E\left[\int_{(n-1)\Delta t}^{n\Delta t} \left(\int_{(k-1)\Delta t}^{k\Delta t} \phi(t - u)dN(u) \right) dt \mid X_k \right] \\
&\approx \int_{(k-1)\Delta t}^{k\Delta t} \cdots \int_{(k-1)\Delta t}^{k\Delta t} \frac{1}{\Delta t^{X_k}} \left(\sum_{i=1}^{X_k} \int_{(n-1)\Delta t}^{n\Delta t} \phi(t - t_i)dt \right) dt_1 \cdots dt_{X_k} \\
&= \phi_{n-k} X_k,
\end{aligned}$$

where ϕ_{n-k} denotes the coarse-grained kernel, defined as

$$\phi_{n-k} = \frac{1}{\Delta t} \int_{(k-1)\Delta t}^{k\Delta t} \int_{(n-1)\Delta t}^{n\Delta t} \phi(t - u)dudt. \tag{4}$$

Similarly, the second term on the right-hand side of (3) is approximated by

$$\begin{aligned}
&E\left[\int_{(n-1)\Delta t}^{n\Delta t} \left(\int_{(n-1)\Delta t}^{t} \phi(t - u)dN(u) \right) dt \mid X_n \right] \\
&= E\left[\int_{(n-1)\Delta t}^{n\Delta t} \left(\int_{(n-1)\Delta t}^{n\Delta t} \phi(t - u)dN(u) \right) dt \mid X_n \right] \approx \phi_0 X_n,
\end{aligned}$$

where we have utilized the causal property $\phi(t) = 0$ for $t < 0$. Substituting these approximations back into (3) yields

$$\lambda_n \approx \mu\Delta t + \sum_{k=1}^{n} \phi_{n-k} X_k. \tag{5}$$

In contrast to the formulation in (1), equation (5) incorporates the coarse-grained kernel, capturing both inter-bin excitation and the intra-bin self-excitation effect through the term $\phi_0 X_n$. Building upon this approximation, we will formally define the coarse-grained Hawkes process in the subsequent section.

3.2. Definition

Firstly, we formally define the coarse-grained kernel, previously derived heuristically in (4).

Definition 1. *Let $\phi(t)$ be a nonnegative excitation kernel defined on the real line, satisfying $\phi(t) = 0$ for $t < 0$. The coarse-grained kernel $\{\phi_k\}_{k \in \mathbb{N}_0}$ with bin size $\Delta t > 0$ is defined by*

$$\phi_k = \begin{cases} \psi_0, & k = 0, \\ \psi_k - \psi_{k-1}, & k = 1, 2, \ldots, \end{cases} \tag{6}$$

where

$$\psi_k = \frac{1}{\Delta t} \int_{k\Delta t}^{(k+1)\Delta t} \left(\int_0^t \phi(u) du \right) dt. \tag{7}$$

It can be readily verified that the formulation (6)–(7) coincides with the expression in (4).

Lemma 1. *The coarse-grained kernel satisfies*

$$\sum_{k=0}^{\infty} \phi_k = \int_0^{\infty} \phi(u) du =: \alpha.$$

Moreover, it holds that $\phi_0 < \alpha$.

Proof. See Appendix A.1. □

Lemma 1 guarantees that the total mass of the coarse-grained kernel equals the integral of the excitation kernel, i.e., the branching ratio. Furthermore, it ensures that the kernel does not collapse to the degenerate case $\phi_0 = \alpha$ as long as $\Delta t < \infty$.

Based on the coarse-grained kernel, we introduce the coarse-grained Hawkes process. Consider a probability space equipped with a sequence of d-dimensional, integer-valued random vectors $\{X_k\}_{k \in \mathbb{Z}}$. Define the history up to bin k as $H_k = \{X_j : j \leq k\}$. We consider the d-variate extension of the coarse-grained process given in (5):

$$\lambda_n = \mu \Delta t + \sum_{k \leq n} \Phi_{n-k} X_k, \quad n \in \mathbb{Z}, \tag{8}$$

where $\mu \in \mathbb{R}_{\geq 0}^d$ is the baseline intensity vector, and each $\Phi_{n-k} \in \mathbb{R}^{d \times d}$ is a matrix whose elements are given by the coarse-grained kernels for bin size Δt. Define the residual process as

$$\Delta M_n = X_n - \lambda_n, \quad n \in \mathbb{Z}, \tag{9}$$

and impose the following conditional moment properties, analogous to the martingale conditions in point process theory:

$$E[\Delta M_n \mid H_{n-1}] = 0, \tag{10}$$

$$E[\Delta M_n \Delta M_n^\mathsf{T} \mid H_{n-1}] = \mathrm{diag}(E[\lambda_n \mid H_{n-1}]). \tag{11}$$

The coarse-grained Hawkes process is then defined as a sequence $\{X_k\}$ satisfying (8)–(11). Importantly, this definition does not fully specify the probability law of the process, but constrains its behavior through second-order statistical structure.

The term $\Phi_0 X_n$ on the right-hand side of (8) encapsulates the effect of intra-bin excitation, which induces both cross-correlations and overdispersion in the count statistics. Assuming that the spectral

radius of Φ_0 is strictly less than one, the conditional expectation and covariance matrix of X_n given H_{n-1} are respectively

$$E[X_n \mid H_{n-1}] = (I - \Phi_0)^{-1}\left(\mu\Delta t + \sum_{k \leq n-1} \Phi_{n-k}X_k\right) =: \lambda_n^*, \tag{12}$$

$$\mathrm{Var}(X_n \mid H_{n-1}) = (I - \Phi_0)^{-1}\mathrm{diag}(\lambda_n^*)(I - \Phi_0^\mathsf{T})^{-1}, \tag{13}$$

where I denotes the $d \times d$ identity matrix. Derivations of these expressions are provided in Appendix A.2. Due to the presence of the matrix factor $(I - \Phi_0)^{-1}$, the conditional covariance matrix is generally non-diagonal, capturing cross-correlations, and its diagonal elements exceed those of λ_n^*, reflecting overdispersion. When Φ_0 is the zero matrix, implying the absence of intra-bin excitation, the conditional variance reduces to $\mathrm{Var}(X_n \mid H_{n-1}) = \mathrm{diag}(\lambda_n^*)$, corresponding to the conditional independence of Poisson-distributed counts.

3.3. Stationary Process

We investigate the coarse-grained Hawkes process under the assumption of stationarity and derive its second-order statistical properties. Assuming stationarity, equations (8) and (10) yield

$$\lambda = E[X_n] = (I - A)^{-1}\mu\Delta t, \quad n \in \mathbb{Z}, \tag{14}$$

where

$$A := \sum_{n=0}^{\infty} \Phi_n$$

denotes the branching ratio matrix. Consequently, the spectral radius of A must be strictly less than one.

To derive the second-order moment of the stationary coarse-grained Hawkes process, we first establish a white noise property of the residual process.

Lemma 2. *Let $\{X_n\}$ denote a coarse-grained Hawkes process whose branching ratio matrix has a spectral radius less than one. Then the residual process (9) forms a stationary sequence satisfying $E[\Delta M_n] = 0$, $n \in \mathbb{Z}$, and*

$$E[\Delta M_n \Delta M_{n'}^\mathsf{T}] = \delta_{nn'}\mathrm{diag}(\lambda), \quad n, n' \in \mathbb{Z}, \tag{15}$$

where $\delta_{nn'}$ denotes the Kronecker delta.

Proof. See Appendix A.3. □

As a consequence of the white noise structure of the residual process, the stationary coarse-grained Hawkes process admits a moving average representation of infinite order,

$$X_n - \lambda = \sum_{k=0}^{\infty} \Psi_k \Delta M_{n-k}, \tag{16}$$

where the kernel matrices Ψ_k are defined by

$$\Psi_k = \sum_{j=0}^{\infty} (\Phi^{(*j)})_k,$$

with $\Phi^{(*j)}$ denoting the j-fold convolution of Φ, and $(\Phi^{(*0)})_k = \delta_{k0}I$. The derivation is provided in Appendix A.4. Using this representation, the autocovariance structure is obtained analogously to linear time series models.

Theorem 1. *Let $\{X_n\}$ be a coarse-grained Hawkes process with a branching ratio matrix whose spectral radius is less than one. Then the autocovariance matrix of $\{X_n\}$ is given by*

$$R_j := \mathrm{Cov}(X_n, X_{n+j}) = \sum_{l=0}^{\infty} \Psi_{l-j}\mathrm{diag}(\lambda)\Psi_l^{\mathsf{T}}, \quad j \in \mathbb{Z}.$$

Proof. See Appendix A.5. □

The spectral density matrix of the stationary coarse-grained Hawkes process is obtained by taking the Fourier transform of the autocovariance sequence:

$$F_{\Delta t}^{(cg)}(\omega) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty} R_n e^{-i\omega n} = \frac{1}{2\pi}\widehat{\Psi}(-\omega)\mathrm{diag}(\lambda)\widehat{\Psi}^{\mathsf{T}}(\omega),$$

where $\widehat{\Psi}(\omega) = \sum_{n=0}^{\infty} \Psi_n e^{-i\omega n}$ is the Fourier transform of the effective kernel. Alternatively, it can be expressed via the Fourier transform of the coarse-grained kernel matrix, $\widehat{\Phi}(\omega) = \sum_{n=0}^{\infty} \Phi_n e^{-i\omega n}$, as

$$F_{\Delta t}^{(cg)}(\omega) = \frac{1}{2\pi}(I - \widehat{\Phi}(-\omega))^{-1}\mathrm{diag}(\lambda)(I - \widehat{\Phi}^{\mathsf{T}}(\omega))^{-1}, \tag{17}$$

utilizing the identity $\widehat{\Psi}(\omega) = (I - \widehat{\Phi}(\omega))^{-1}$.

In summary, under the condition that the spectral radius of the branching ratio matrix is less than one, the expected value of the process remains constant over time, and its autocovariance depends solely on the lag between observations, not their absolute positions in time. Accordingly, the coarse-grained Hawkes process is weakly stationary.

3.4. Approximation to Hawkes Process

We now establish a rigorous connection between the original Hawkes process and its coarse-grained counterpart. Specifically, we examine the second-order statistical properties of both processes under the assumption of stationarity. A summary of the statistical properties of the stationary Hawkes process is provided in Appendix B.

Since the sum of the coarse-grained kernel coincides with the integral of the excitation kernel (see Lemma 1), the branching ratio matrices, and consequently, the conditions for stationarity, are identical for both processes. Under stationarity, the expected event counts of the coarse-grained Hawkes process (14) coincide with those of the original Hawkes process (A6).

We proceed to compare the spectral density matrices of the two processes, focusing in particular on the convergence behavior of the spectral density of the coarse-grained process toward that of the original Hawkes process.

Theorem 2. *The spectral density matrix of the coarse-grained Hawkes process satisfies*

$$F_{\Delta t}^{(cg)}(\nu\Delta t) = F^{(hw)}(\nu)\Delta t + O(\Delta t^3), \tag{18}$$

as $\Delta t \to 0$, where $F^{(hw)}(\nu)$ denotes the spectral density matrix (A5) of the original Hawkes process.

Proof. See Appendix A.6. □

Considering the binned Hawkes process, its spectral density matrix $F_{\Delta t}^{(hw)}(\omega)$ behaves as (see Appendix B)

$$F_{\Delta t}^{(hw)}(\nu\Delta t) = F^{(hw)}(\nu)\Delta t + O(\Delta t^3),$$

as $\Delta t \to 0$, which matches the expansion in (18) up to second-order terms in Δt. This leads to the following corollary.

Corollary 1. *The spectral density matrix of the coarse-grained Hawkes process approximates that of the binned Hawkes process to third-order accuracy as*

$$F_{\Delta t}^{(cg)}(\omega) = F_{\Delta t}^{(hw)}(\omega) + O(\Delta t^3),$$

as $\Delta t \to 0$.

For comparative purposes, consider the binned Poisson approximation defined by equations (1)–(2). Its spectral density matrix $F_{\Delta t}^{(po)}(\omega)$ behaves as (see Appendix A.6)

$$F_{\Delta t}^{(po)}(\omega) = F_{\Delta t}^{(hw)}(\omega) + O(\Delta t^2).$$

Therefore, the coarse-grained Hawkes process yields a spectral approximation to the binned Hawkes process that is accurate to a higher order than the binned Poisson approximation.

3.5. Parameter Estimation Method

We address the problem of estimating the parameters of a Hawkes process from binned count data. Assume that we observe a sequence of binned event counts $\{X_1, \ldots, X_n\}$ of length n generated by a d-dimensional Hawkes process, whose excitation kernel matrix $\Phi(\theta)$ is specified by a parametric form with an unknown parameter vector θ. We propose a parameter estimation method for θ based on the coarse-grained Hawkes process framework.

To this end, we utilize the AR(∞) representation of the coarse-grained Hawkes process (see Appendix A.4),

$$X_n - \lambda = \sum_{k=1}^{\infty} \Phi_k^*(\theta)(X_{n-k} - \lambda) + \Delta M_n^*(\theta),$$

where $\Phi_k^*(\theta) = (I - \Phi_0(\theta))^{-1}\Phi_k(\theta)$, and $\Delta M_n^*(\theta) = (I - \Phi_0(\theta))^{-1}\Delta M_n(\theta)$ is a zero-mean, cross-correlated white noise sequence satisfying

$$E[\Delta M_n^*(\theta)\Delta M_{n'}^*(\theta)^{\mathsf{T}}] = \delta_{nn'}(I - \Phi_0(\theta))^{-1}\mathrm{diag}(\lambda)(I - \Phi_0(\theta)^{\mathsf{T}})^{-1}$$
$$=: \delta_{nn'}\Lambda^*(\theta), \quad n, n' \in \mathbb{Z}.$$

Based on this representation, we define a loss function in quadratic form, whose minimizer yields an estimator of the parameter vector:

$$\mathcal{L}_n(\theta) = \sum_{k=1}^{n}\left(Y_k - \sum_{j=1}^{k-1}\Phi_j^*(\theta)Y_{k-j}\right)^{\mathsf{T}}\Lambda^{*-1}(\theta)\left(Y_k - \sum_{j=1}^{k-1}\Phi_j^*(\theta)Y_{k-j}\right) + n\log|\Lambda^*(\theta)|, \qquad (19)$$

where $Y_k = X_k - \lambda$. In practice, the unknown mean vector λ is replaced by the empirical mean $\hat{\lambda} = n^{-1}\sum_{k=1}^{n}X_k$. The loss function then simplifies to

$$\mathcal{L}_n(\theta) = \sum_{k=1}^{n}\left(Y_k - \sum_{j=0}^{k-1}\Phi_j(\theta)Y_{k-j}\right)^{\mathsf{T}}\mathrm{diag}(\hat{\lambda})^{-1}\left(Y_k - \sum_{j=0}^{k-1}\Phi_j(\theta)Y_{k-j}\right) - 2n\log|I - \Phi_0(\theta)|, \qquad (20)$$

where the constant term $\log\mathrm{diag}(\hat{\lambda})$ has been omitted. The first term on the right-hand side of (20) corresponds to a weighted quadratic loss, while the second term serves as a regularization component that discourages trivial solutions of the form $\Phi_k(\theta) = \delta_{k0}I$. The optimal parameter estimate is obtained by minimizing the loss function with respect to θ.

The estimation procedure is summarized as follows:

Algorithm 1 Estimation procedure for Hawkes processes

1: Given the observed bin-count sequence $\{X_1, \ldots, X_n\}$, compute the empirical mean $\hat{\lambda} = n^{-1} \sum_{k=1}^{n} X_k$ and center the data as $Y_k = X_k - \hat{\lambda}$ for $k = 1, \ldots, n$.
2: Determine the parameter estimate $\hat{\theta}$ by minimizing the loss function (20).
3: Using the estimated parameter vector $\hat{\theta}$, compute the estimate of the baseline intensity as

$$\hat{\mu} = (I - \hat{A})\hat{\lambda}/\Delta,$$

where $\hat{A} = \sum_{n=0}^{\infty} \Phi_n(\hat{\theta})$ denotes the estimated branching ratio matrix.

---

4. Numerical Experiments

4.1. Assessment of Second-Order Characteristics

As established in Corollary 1, the coarse-grained Hawkes process asymptotically approximates the spectral density matrix of the binned Hawkes process as the bin size $\Delta t \to 0$. In this section, we conduct numerical investigations to evaluate the validity and robustness of this approximation for increasing values of Δt. To this end, we consider a bivariate Hawkes process ($d = 2$), where each component of the excitation kernel matrix is defined as

$$\phi_{ij}(t) = \alpha_{ij} g_{ij}(t), \quad t \geq 0,$$

with $g_{ij}(t)$ denoting a normalized kernel satisfying $\int_0^{\infty} g_{ij}(t)dt = 1$. The coefficient α_{ij} specifies the branching ratio from component j to component i, and $g_{ij}(t)$ characterizes the distribution of waiting times for event excitation. We specifically focus on a symmetric bivariate Hawkes process where the parameters satisfy $\mu_1 = \mu_2 = \mu$, $g_{ij}(t) = g(t)$ ($1 \leq i, j \leq 2$), and the excitation kernel matrix takes the form

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} = \begin{pmatrix} \alpha^{(s)} & \alpha^{(c)} \\ \alpha^{(c)} & \alpha^{(s)} \end{pmatrix}.$$

The stationarity condition for the process holds if $\alpha^{(s)} + \alpha^{(c)} < 1$. Under this constraint, the second-order statistical structure of the symmetric Hawkes process is described by the power spectral density (PSD) of each component (the diagonal elements of the spectral density matrix) and the cross-spectral density (CSD) between them (the off-diagonal elements).

We illustrate our findings using an exponential kernel defined as

$$g(t) = \beta e^{-\beta t}, \quad t \geq 0,$$

where β^{-1} denotes the expected waiting time before excitation. Figure 1 presents the PSD and CSD, along with the auto- and cross-covariance functions, of the binned Hawkes process with parameters $\alpha^{(s)} = 0.4$, $\alpha^{(c)} = 0.3$, and $\beta = 1$, for bin sizes $\Delta t = 0.1$ (a), 1 (b), and 2 (c). These are depicted using blue dotted lines.
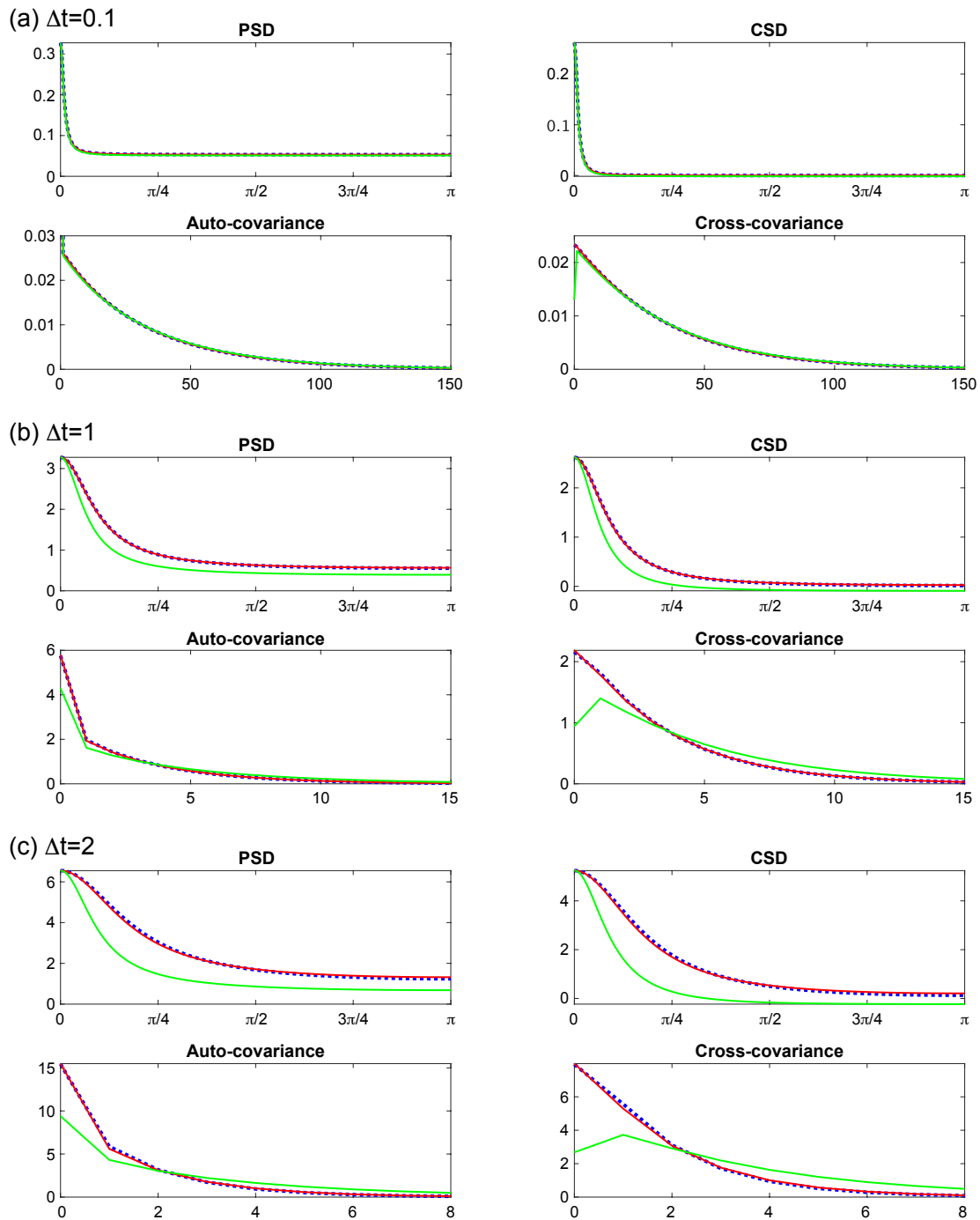
The associated coarse-grained Hawkes process is constructed using the coarse-grained kernel derived from the exponential function:

$$g_k = \begin{cases} 1 - (1 - e^{-\beta \Delta t})/\beta \Delta t, & k = 0, \\ (e^{\beta \Delta t} + e^{-\beta \Delta t} - 2)e^{-\beta k \Delta t}/\beta \Delta t, & k = 1, 2, \ldots \end{cases}$$

In the same figures, the four second-order statistics of the coarse-grained process are represented by red lines, while those corresponding to the binned Poisson approximation are shown in green.

From these comparisons, we observe that for small bin sizes ($\Delta t = 0.1 < \beta^{-1}$), both the coarse-grained Hawkes process and the binned Poisson approximation closely reproduce the second-order behavior of the binned Hawkes process (Figure 1a). However, for $\Delta t = 1$, which is comparable to the mean waiting time, the Poisson approximation significantly deteriorates (Figure 1b), whereas the

coarse-grained Hawkes process continues to provide a high-fidelity approximation. Even for a larger bin size $\Delta t = 2$, exceeding the characteristic time scale β^{-1}, the coarse-grained model remains accurate (Figure 1c).

Figure 1. Power spectral density (PSD), cross-spectral density (CSD), auto-covariance, and cross-covariance functions for the three processes at bin sizes $\Delta t = 0.1$ (a), 1 (b), and 2 (c). The blue dotted line corresponds to the binned Hawkes process, while the red and green lines depict the coarse-grained Hawkes process and the binned Poisson approximation, respectively. The parameters of the Hawkes process are set as $\mu = 1$, $\alpha^{(s)} = 0.4$, $\alpha^{(c)} = 0.3$, and $\beta = 1$. The coarse-grained Hawkes process provides a close approximation to the binned Hawkes process, whereas the binned Poisson approximation exhibits noticeable degradation for $\Delta t = 1$ and 2.

To quantitatively assess the fidelity of the approximations, we introduce a divergence measure based on the log-determinant of the spectral density matrices. Specifically, let $F_{\Delta t}^{(\bullet \bullet)}(\omega)$ denote the
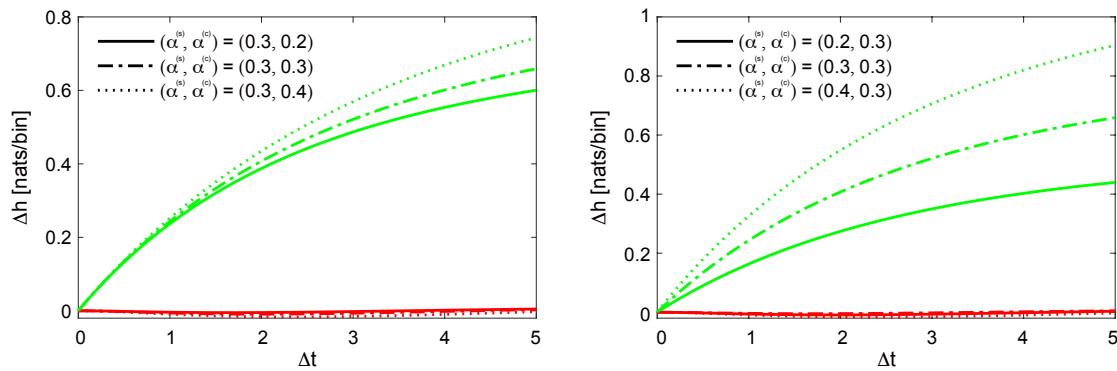
spectral density matrix of the approximating process, where '••' indicates either the coarse-grained (cg) or Poisson (po) approximation. Then, the information loss relative to the binned Hawkes process $F_{\Delta t}^{(hw)}(\omega)$ is defined as

$$\Delta h^{(\bullet\bullet)} = \frac{1}{4\pi} \int_{-\pi}^{\pi} \log \frac{|F_{\Delta t}^{(hw)}(\omega)|}{|F_{\Delta t}^{(\bullet\bullet)}(\omega)|} d\omega,$$

which corresponds to the gap in maximum entropy rates under spectral constraints.

Figure 2 plots the information loss for both approximations as a function of Δt, across varying values of $\alpha^{(s)}$ and $\alpha^{(c)}$. The results clearly indicate that the coarse-grained Hawkes process incurs minimal information loss across a wide range of parameter settings. In contrast, the accuracy of the binned Poisson approximation degrades with increasing Δt, with the loss exacerbated further as $\alpha^{(s)}$ and $\alpha^{(c)}$ increase.

We additionally examined the case where the excitation kernel follows a power-law distribution instead of an exponential decay. The qualitative behavior remained consistent (see Appendix C). These findings collectively demonstrate that the coarse-grained Hawkes process offers a substantially improved approximation of the second-order dynamics of the binned Hawkes process, particularly for larger bin widths.

Figure 2. Information loss as a function of Δt for the coarse-grained Hawkes process (red line) and the binned Poisson approximation (green line). The parameters are set to $\mu = 1$ and $\beta = 1$. The information loss associated with the binned Poisson approximation increases with larger values of Δt, $\alpha^{(s)}$, and $\alpha^{(c)}$, whereas the information loss incurred by the coarse-grained Hawkes process remains negligible across all configurations.

4.2. Parameter Estimation

We now investigate the efficacy of the proposed estimation method in inferring the parameters of a Hawkes process from bin-count data. Specifically, we consider an asymmetric bivariate Hawkes process with exponential excitation kernels given by $g_{ij}(t) = \beta_{ij} e^{-\beta_{ij} t}$ ($1 \le i, j \le 2$). The parameters of the Hawkes process are set as follows:
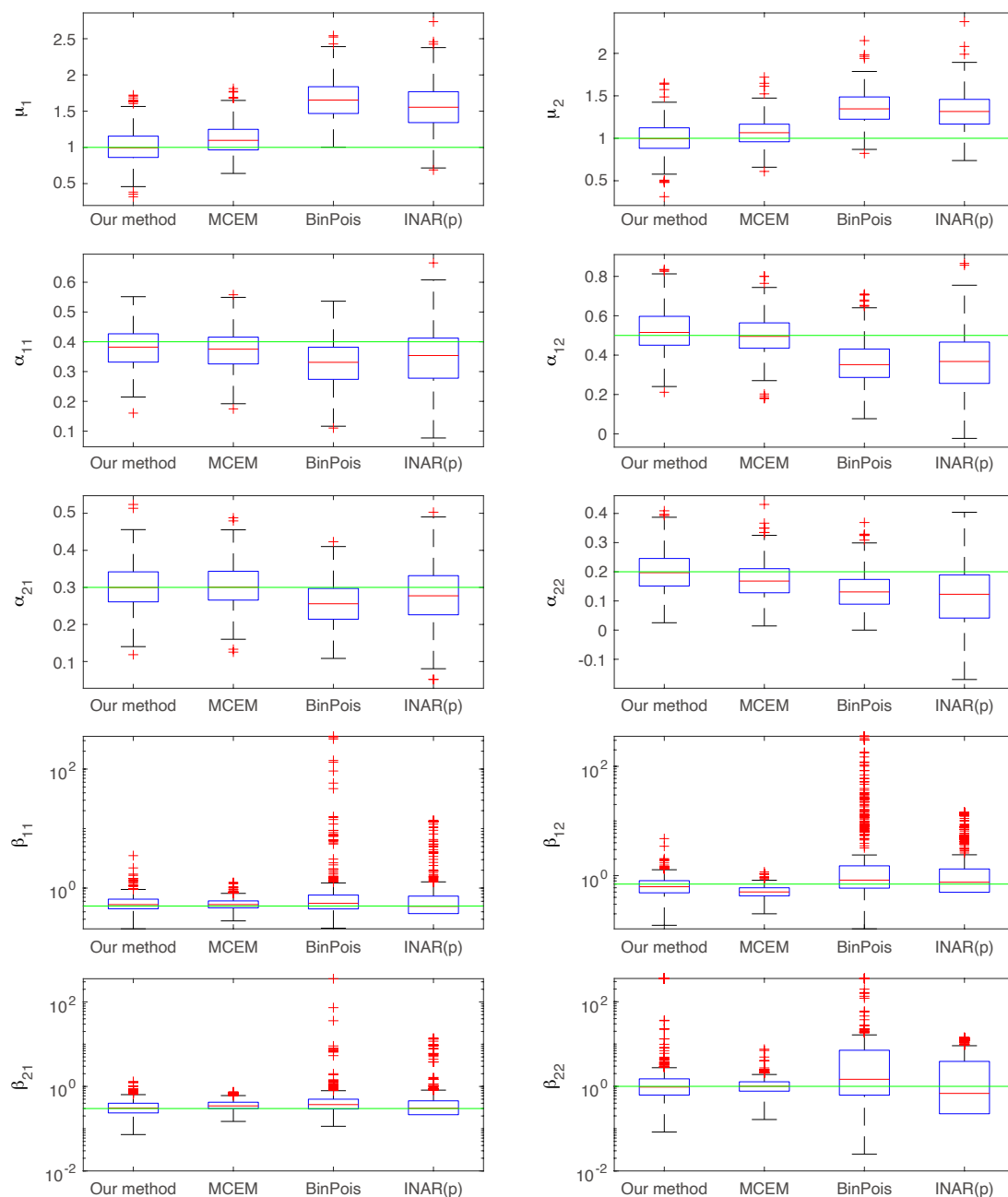
$$\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} = \begin{pmatrix} 0.4 & 0.5 \\ 0.3 & 0.2 \end{pmatrix}, \quad \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} = \begin{pmatrix} 0.5 & 0.7 \\ 0.3 & 1.0 \end{pmatrix}.$$

The numerical experiments were conducted as follows. First, realizations of the Hawkes process were generated over the interval $[0, T]$ and subsequently discretized into bin-count sequences using bin size Δt. The ten model parameters $\{\mu_i, \alpha_{ij}, \beta_{ij}\}_{1 \le i,j \le 2}$ were then estimated from these sequences.

To assess the performance of our proposed method, we compare it against three established approaches. The first is the MC-EM algorithm proposed in [20], for which we employed the publicly available implementation [25]. The second method is maximum likelihood estimation (MLE) applied to the binned Poisson approximation. The third method involves conditional least squares estimation for the INAR(p) process, as introduced in [18]. Since the INAR(p) framework yields nonparametric
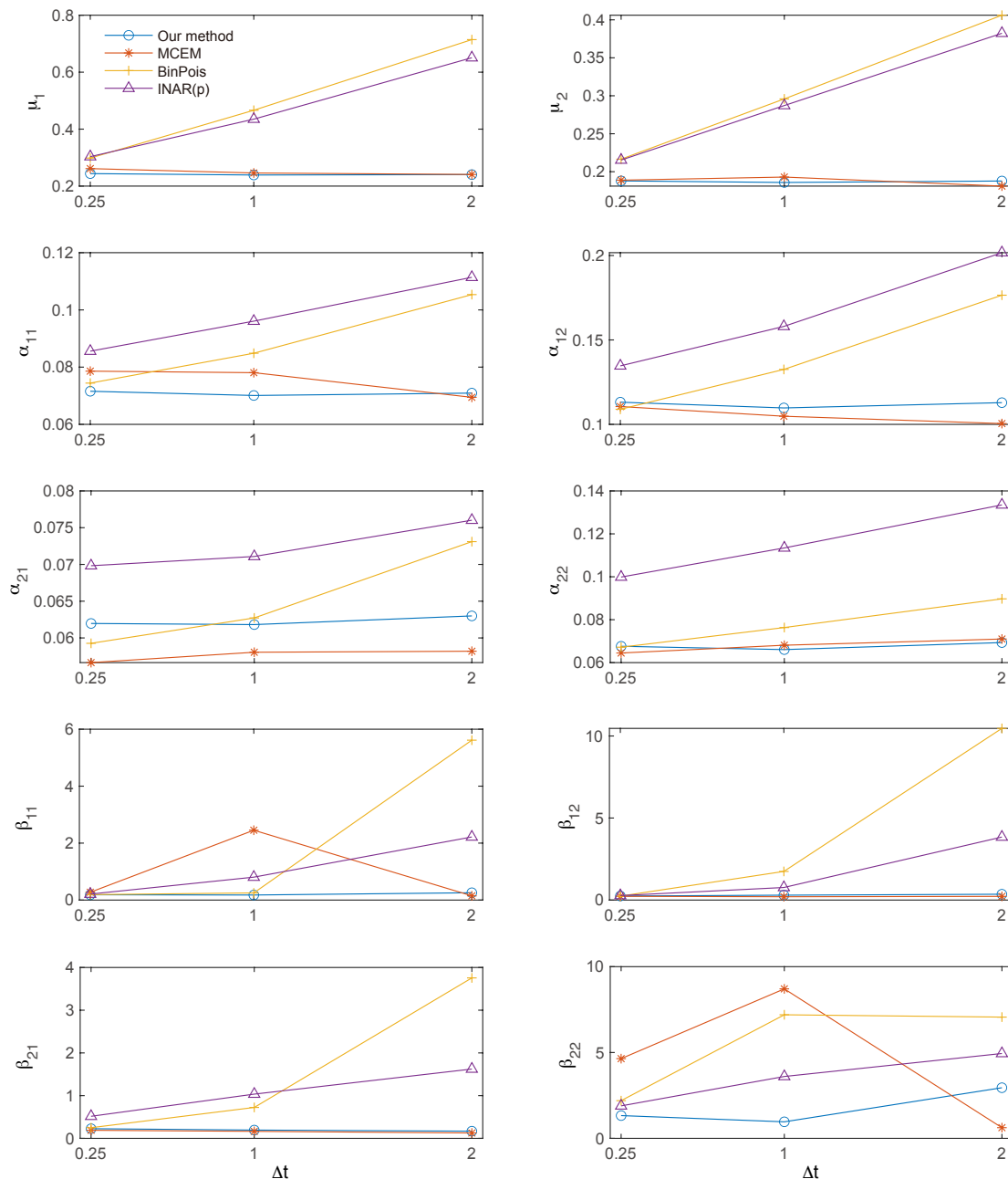
estimates of the excitation kernels, we obtained the corresponding parametric kernel parameters by fitting an exponential function to the nonparametric estimates [19,20].

Figure 3 presents boxplots of the estimated values for each of the ten parameters across 500 simulated realizations of the binned Hawkes process, using $T = 1000$ and $\Delta t = 2$. It is apparent that both the binned Poisson MLE and the INAR(p) method produce significantly biased estimates, which is expected given that these approaches disregard excitation effects within each bin. Furthermore, we observe a substantial number of outliers in the estimates of β_{ij} across all four methods, indicating that estimation of the kernel scales exhibits higher variance than estimation of the baseline intensities or branching ratios. For instance, our method yielded estimates of β_{22} that deviated by a factor of 100 from the true value in 16 out of the 500 trials, highlighting the challenges in accurately estimating β_{ij} when the bin size is large relative to the kernel time scale.

Figure 3. Boxplots of the estimated values for each of the ten model parameters. The green solid line indicates the ground truth values. Note that the INAR(p) method may yield negative values, which are omitted when log-scaled axes are used.

Figure 4 displays the root mean squared error (RMSE) of the parameter estimates across various bin sizes. For the RMSE of $\hat{\beta}_{ij}$, extreme outliers with values exceeding 100 were excluded to mitigate their undue influence. Overall, the proposed method demonstrates superior performance compared to both the binned Poisson MLE and the INAR(p) approach, and achieves accuracy comparable to the MC-EM algorithm. Additionally, both the MC-EM and the proposed methods consistently maintain low RMSE values across different bin sizes, whereas the RMSEs for the binned Poisson MLE and the INAR(p) approach increase with larger bin sizes. A similar trend is observed for $\hat{\beta}_{ij}$, albeit with more fluctuation due to outliers.
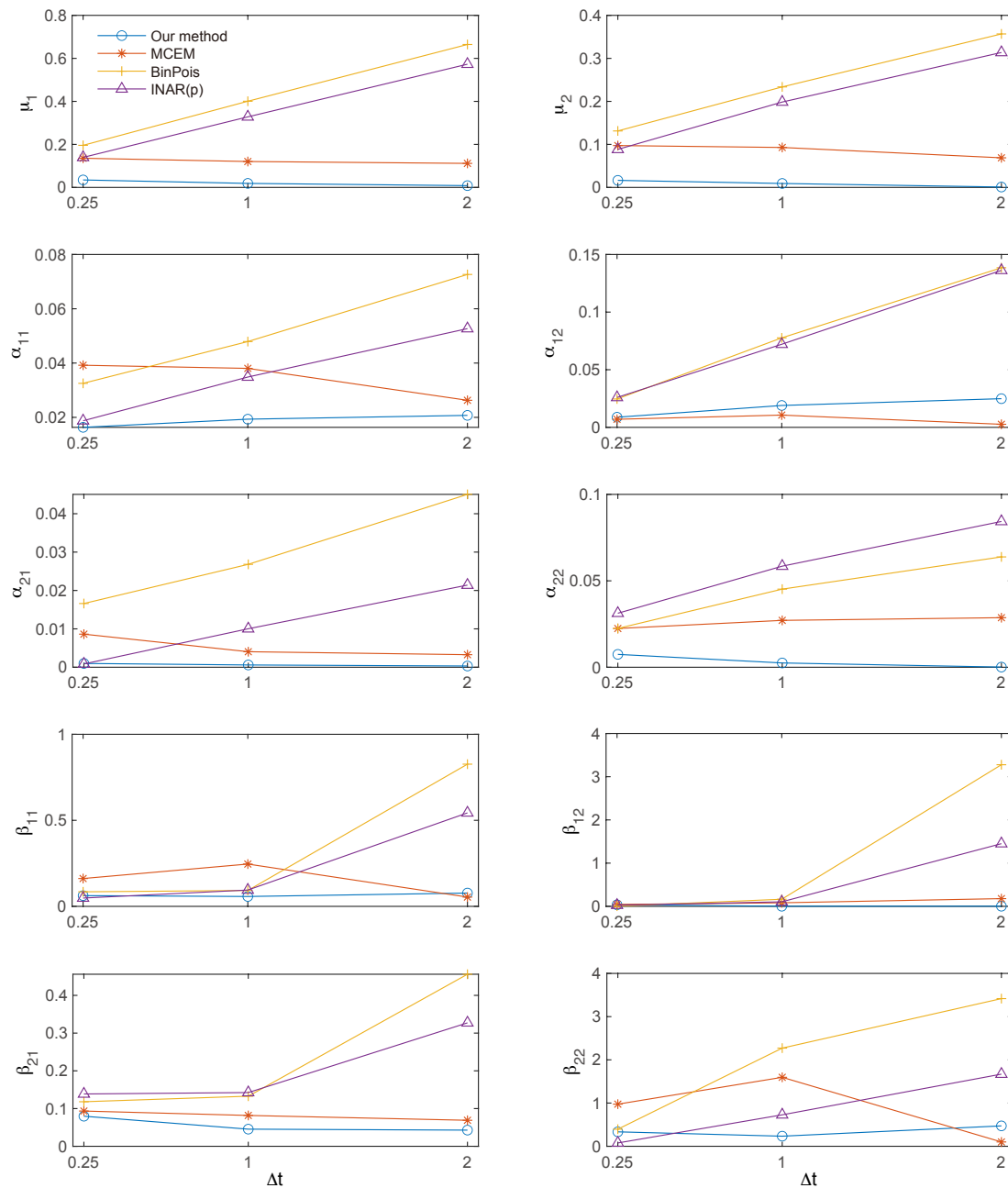

Figure 4. Root mean squared error (RMSE) of the parameter estimates. Overall, the RMSE of the proposed method is comparable to that of the MC-EM algorithm, and significantly lower than that of the binned Poisson approximation and the INAR(p) method.

Figures 5 and 6 illustrate, respectively, the bias and standard deviation components of the RMSE for each of the ten parameters. The proposed method yields the lowest bias among all methods,
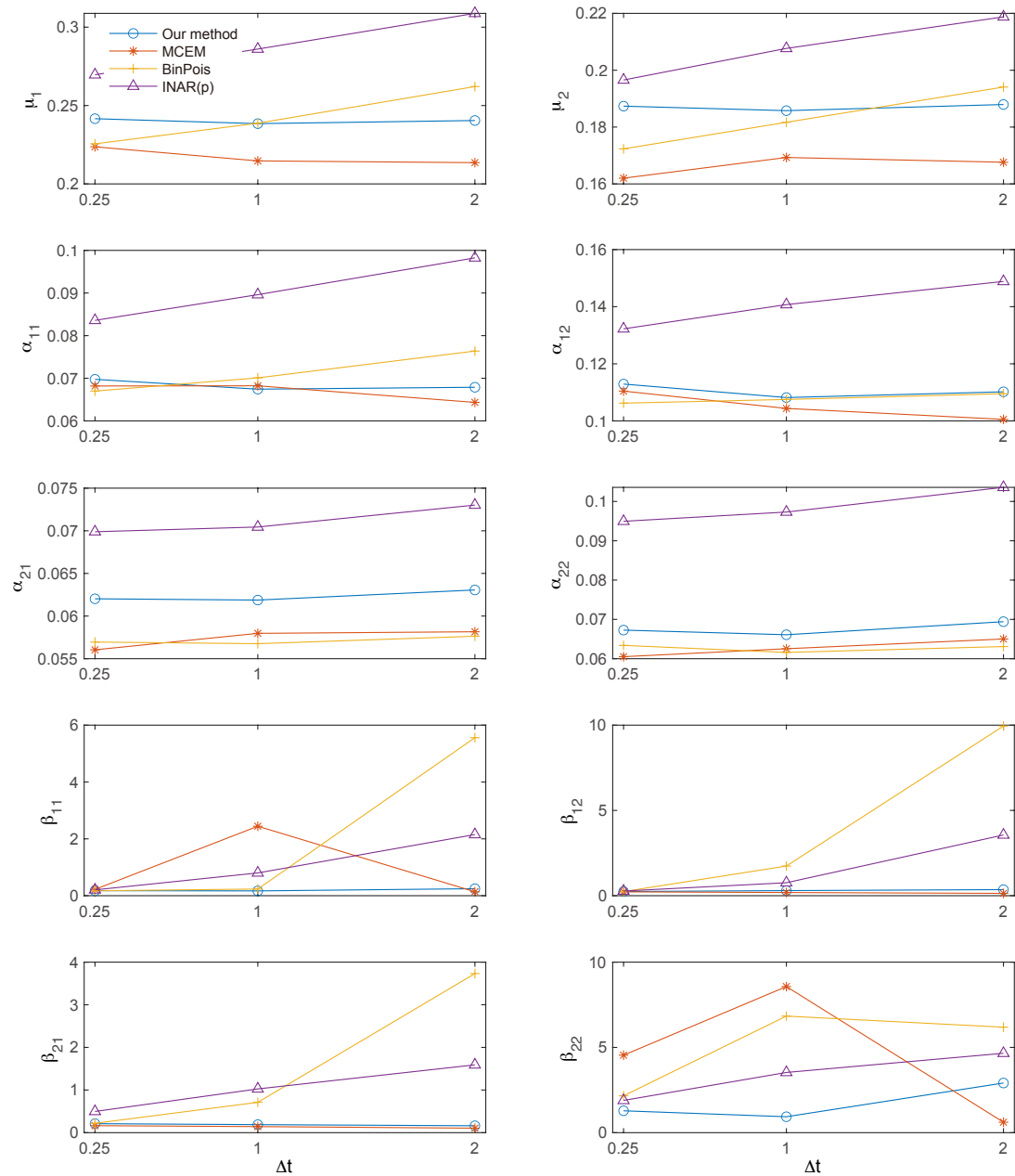
with only a few exceptions (Figure 5); meanwhile, the MC-EM algorithm attains the lowest standard deviation (Figure 6).

Figure 5. Bias in the estimated parameters. The proposed method consistently achieves the lowest bias among the four methods, except for a few isolated cases.

In conclusion, the proposed method matches the performance of the MC-EM algorithm while outperforming both the binned Poisson MLE and the INAR(p) method, particularly in terms of bias reduction. It provides robust and stable estimates for the baseline intensities μ_i and the branching ratios α_{ij}, with estimation accuracy largely unaffected by bin size. Accurate estimation of the kernel scales β_{ij} is feasible when the bin size is smaller than the characteristic kernel scale, but becomes unreliable as the bin size increases.

Figure 6. Standard deviation (STD) in the parameter estimates. The MC-EM algorithm generally attains the lowest standard deviation among the four methods.

5. Discussion

In this study, we introduced the coarse-grained Hawkes process as an analytical approximation to the binned Hawkes process. Unlike conventional discretization techniques, the proposed framework incorporates a coarse-grained excitation kernel that systematically accounts for intra-bin excitations. Consequently, the coarse-grained Hawkes process faithfully reproduces the second-order statistical properties of the binned Hawkes process, even when the bin size exceeds the characteristic timescale of the excitation kernel. Moreover, we demonstrated that the proposed approach enables stable estimation of Hawkes process parameters from bin-count data. In particular, both the branching ratios and baseline intensities can be reliably inferred, irrespective of the temporal resolution of the bin-count sequences.

A central distinction between our approach and the Monte Carlo Expectation-Maximization (MC-EM) algorithm lies in the treatment of latent event times within the bin-count data. Whereas the MC-EM method necessitates Monte Carlo sampling from the conditional distribution of unobserved events,

resulting in considerable computational burden, our method employs a parsimonious assumption that events are uniformly distributed within each bin. This assumption facilitates the analytical derivation of approximate conditional expectations. Despite its simplicity, the proposed method achieves estimation accuracy on par with that of the MC-EM algorithm while offering substantial computational advantages.

We further highlight a theoretical connection between our estimation framework and the spectral method, previously validated in the univariate setting [22]. Applying the Fourier transform, the loss function in equation (19) asymptotically approximates a spectral likelihood for large n:

$$\mathcal{L}_n(\boldsymbol{\theta}) \approx \sum_{m=1}^{n} \left(\widehat{\boldsymbol{X}}_m^* F_{\Delta t}^{(cg)}(\omega_m)^{-1} \widehat{\boldsymbol{X}}_m + \log |F_{\Delta t}^{(cg)}(\omega_m)| \right),$$

where $\omega_m = 2\pi m/n$,

$$\widehat{\boldsymbol{X}}_m = \frac{1}{\sqrt{2\pi n}} \sum_{k=1}^{n} (\boldsymbol{X}_k - \boldsymbol{\lambda}) e^{-2\pi ikm/n},$$

and $\widehat{\boldsymbol{X}}_m^*$ denotes the Hermitian transpose. Notably, the spectral likelihood for the multivariate binned Hawkes process can be obtained by replacing $F_{\Delta t}^{(cg)}(\omega_m)$ with the spectral density matrix of the binned Hawkes process.

Finally, we emphasize the potential extension of the proposed framework to nonstationary time series. Owing to its formulation in the time domain, the coarse-grained Hawkes process is amenable to integration within state-space modeling paradigms [26,27], offering a promising direction for modeling nonstationary dynamics. We propose this as a compelling avenue for future research.

Conflicts of Interest: The author declares no conflicts of interest.

Appendix A. Proofs

Appendix A.1. Proof of Lemma 1

It follows from equations (6) and (7) that

$$\sum_{k=0}^{\infty} \phi_k = \lim_{n\to\infty} \{\psi_0 + (\psi_1 - \psi_0) + \cdots + (\psi_n - \psi_{n-1})\} = \lim_{n\to\infty} \psi_n$$

$$= \lim_{n\to\infty} \frac{1}{\Delta t} \int_{n\Delta t}^{(n+1)\Delta t} \left(\int_0^t \phi(u)du \right) dt = \lim_{n\to\infty} \frac{1}{\Delta t} \int_0^{\Delta t} \left(\int_0^{n\Delta t+s} \phi(u)du \right) ds.$$

Since $\int_0^{n\Delta t+s} \phi(u)du$ is nonnegative and monotonically nondecreasing with respect to n, we may invoke the dominated convergence theorem to interchange the limit and the integral, yielding

$$\sum_{k=0}^{\infty} \phi_k = \frac{1}{\Delta t} \int_0^{\Delta t} \left(\lim_{n\to\infty} \int_0^{n\Delta t+s} \phi(u)du \right) ds = \frac{1}{\Delta t} \int_0^{\Delta t} \left(\int_0^{\infty} \phi(u)du \right) ds = \alpha.$$

Furthermore, since $\int_0^t \phi(u)du$ is continuous, nonnegative, and monotonically nondecreasing in t, we obtain

$$\phi_0 = \frac{1}{\Delta t} \int_0^{\Delta t} \left(\int_0^t \phi(u)du \right) dt < \frac{1}{\Delta t} \int_0^{\Delta t} \left(\int_0^{\infty} \phi(u)du \right) dt = \alpha.$$

Appendix A.2. Derivation of (12) and (13)

From equations (8) and (10), we have

$$E[\boldsymbol{X}_n \mid H_{n-1}] = E\left[\boldsymbol{\mu}\Delta t + \sum_{k \leq n} \Phi_{n-k}\boldsymbol{X}_k \mid H_{n-1}\right]$$

$$= \boldsymbol{\mu}\Delta t + \sum_{k \leq n-1} \Phi_{n-k}\boldsymbol{X}_k + \Phi_0 E[\boldsymbol{X}_n \mid H_{n-1}].$$

$$\therefore \ (I - \Phi_0)E[\boldsymbol{X}_n \mid H_{n-1}] = \boldsymbol{\mu}\Delta t + \sum_{k \leq n-1} \Phi_{n-k}\boldsymbol{X}_k.$$

Since the spectral radius of Φ_0 is strictly less than one, the matrix $(I - \Phi_0)$ is invertible, and all entries of $E[\boldsymbol{X}_n \mid H_{n-1}]$ are nonnegative. Hence, the conditional mean vector of $\boldsymbol{X}_n$ is given by expression (12).

To derive (13), observe that

$$\boldsymbol{X}_n - \boldsymbol{\lambda}_n^* = (I - \Phi_0)^{-1}\left((I - \Phi_0)\boldsymbol{X}_n - \boldsymbol{\mu}\Delta t - \sum_{k \leq n-1} \Phi_{n-k}\boldsymbol{X}_k\right) = (I - \Phi_0)^{-1}\Delta \boldsymbol{M}_n.$$

Therefore,

$$\text{Var}(\boldsymbol{X}_n \mid H_{n-1}) = E\left[(\boldsymbol{X}_n - \boldsymbol{\lambda}_n^*)(\boldsymbol{X}_n - \boldsymbol{\lambda}_n^*)^{\mathsf{T}} \mid H_{n-1}\right]$$

$$= (I - \Phi_0)^{-1}E\left[\Delta \boldsymbol{M}_n \Delta \boldsymbol{M}_n^{\mathsf{T}} \mid H_{n-1}\right](I - \Phi_0^{\mathsf{T}})^{-1}$$

$$= (I - \Phi_0)^{-1}\text{diag}(E[\boldsymbol{\lambda}_n \mid H_{n-1}])(I - \Phi_0^{\mathsf{T}})^{-1},$$

where the last equality follows from equation (11). Finally, since $E[\boldsymbol{\lambda}_n \mid H_{n-1}] = \boldsymbol{\lambda}_n^*$ by (10), we obtain equation (13).

Appendix A.3. Proof of Lemma 2

From equation (10), it follows that

$$E[\Delta \boldsymbol{M}_n] = E[E[\Delta \boldsymbol{M}_n \mid H_{n-1}]] = \boldsymbol{0}.$$

To derive equation (15), observe that for $n < n'$ (and, by symmetry, for $n \neq n'$),

$$E[\Delta \boldsymbol{M}_n \Delta \boldsymbol{M}_{n'}^{\mathsf{T}}] = E\left[\Delta \boldsymbol{M}_n E[\Delta \boldsymbol{M}_{n'}^{\mathsf{T}} \mid H_{n'-1}]\right] = \boldsymbol{0}.$$

Moreover, applying equation (11), we obtain

$$E[\Delta \boldsymbol{M}_n \Delta \boldsymbol{M}_n^{\mathsf{T}}] = E\left[E[\Delta \boldsymbol{M}_n \Delta \boldsymbol{M}_n^{\mathsf{T}} \mid H_{n-1}]\right] = E[\text{diag}(E[\boldsymbol{\lambda}_n \mid H_{n-1}])] = \text{diag}(\boldsymbol{\lambda}),$$

which completes the proof.

Appendix A.4. AR(∞) and MA(∞) representations

Using equations (9) and (14), the rate equation (8) can be reformulated as

$$\boldsymbol{X}_n - \boldsymbol{\lambda} = \sum_{k=0}^{\infty} \Phi_k(\boldsymbol{X}_{n-k} - \boldsymbol{\lambda}) + \Delta \boldsymbol{M}_n, \qquad (A1)$$

$$\therefore \ (I - \Phi_0)(\boldsymbol{X}_n - \boldsymbol{\lambda}) = \sum_{k=1}^{\infty} \Phi_k(\boldsymbol{X}_{n-k} - \boldsymbol{\lambda}) + \Delta \boldsymbol{M}_n.$$

Under the stationarity condition, the inverse $(I - \Phi_0)^{-1}$ exists, yielding the autoregressive representation

$$X_n - \lambda = \sum_{k=1}^{\infty} \Phi_k^*(X_{n-k} - \lambda) + \Delta M_n^*,$$

where $\Phi_k^* = (I - \Phi_0)^{-1}\Phi_k$ and $\Delta M_n^* = (I - \Phi_0)^{-1}\Delta M_n$.

To obtain a moving average representation, we apply the formal z-transform to both sides of equation (A1):

$$\widehat{Y}(z) = \widehat{\Phi}(z)\widehat{Y}(z) + \Delta\widehat{M}(z),$$

where $\widehat{Y}(z) = \sum_{n=-\infty}^{\infty}(X_n - \lambda)z^{-n}$, $\widehat{\Phi}(z) = \sum_{n=-\infty}^{\infty}\Phi_n z^{-n}$ and $\Delta\widehat{M}(z) = \sum_{n=-\infty}^{\infty}\Delta M_n z^{-n}$. Solving for $\widehat{Y}(z)$ yields

$$\widehat{Y}(z) = \widehat{\Psi}(z)\Delta\widehat{M}(z),$$

where the transfer function is defined by

$$\widehat{\Psi}(z) = (I - \widehat{\Phi}(z))^{-1} = \sum_{j=0}^{\infty} \widehat{\Phi}^j(z).$$

Applying the inverse z-transform then gives the MA(∞) process:

$$X_n - \lambda = \sum_{k=0}^{\infty} \Psi_k \Delta M_{n-k},$$

where $\Psi_k = \sum_{j=0}^{\infty}(\Phi^{(*j)})_k$.

Appendix A.5. Proof of Theorem 1

Using equations (15) and (16), the autocovariance matrix of $\{X_n\}$ is given by

$$
\begin{aligned}
\mathrm{Cov}(X_n, X_{n+j}) &= E[(X_n - \lambda)(X_{n+j} - \lambda)^{\mathsf{T}}] \\
&= E\left[\left(\sum_{k=0}^{\infty} \Psi_k \Delta M_{n-k}\right)\left(\sum_{l=0}^{\infty} \Psi_l \Delta M_{n+j-l}\right)^{\mathsf{T}}\right] \\
&= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \Psi_k E[\Delta M_{n-k}\Delta M_{n+j-l}^{\mathsf{T}}]\Psi_l^{\mathsf{T}} \\
&= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \Psi_k \mathrm{diag}(\lambda)\delta_{k,l-j}\Psi_l^{\mathsf{T}} \\
&= \sum_{l=0}^{\infty} \Psi_{l-j}\mathrm{diag}(\lambda)\Psi_l^{\mathsf{T}},
\end{aligned}
$$

which completes the proof.

Appendix A.6. Proof of Theorem 2

We provide a proof of Theorem 2 in the univariate setting. The extension to the multivariate case follows by applying the same reasoning component-wise to each element of the spectral density matrix. To begin, we establish the following approximation result for the Fourier transforms of the excitation kernel and its coarse-grained counterpart.

Lemma A1. *Let $\widetilde{\phi}(v) = \int_0^{\infty} \phi(t)e^{-ivt}dt$ denote the continuous-time Fourier transform of the excitation kernel, and let $\widehat{\phi}(\omega) = \sum_{k=0}^{\infty} \phi_k e^{-i\omega k}$ denote the discrete-time Fourier transform of the corresponding coarse-grained kernel. Then, in the limit as $\Delta t \to 0$, the following approximation holds:*

$$\widehat{\phi}(v\Delta t) = \widetilde{\phi}(v) + O(\Delta t^2). \tag{A2}$$

Proof. From (6) and (7), we obtain

$$
\begin{aligned}
\widehat{\phi}(v\Delta t) &= \sum_{k=0}^{\infty} \phi_k e^{-ivk\Delta t} \\
&= \psi_0 + (\psi_1 - \psi_0)e^{-iv\Delta t} + (\psi_2 - \psi_1)e^{-2iv\Delta t} + \cdots \\
&= \lim_{n\to\infty}\left[\sum_{k=0}^{n-1} \psi_k e^{-ivk\Delta t}(1 - e^{-iv\Delta t}) + \psi_n e^{-ivn\Delta t}\right] \\
&= \lim_{n\to\infty}\left[\sum_{k=0}^{n-1} \int_{k\Delta t}^{(k+1)\Delta t} \varphi(t)e^{-ivk\Delta t}dt \cdot \frac{1 - e^{-iv\Delta t}}{\Delta t} + \psi_n e^{-ivn\Delta t}\right],
\end{aligned}
\tag{A3}
$$

where $\varphi(t) = \int_0^t \phi(u)du$. To evaluate the integral, consider

$$
\begin{aligned}
\int_{k\Delta t}^{(k+1)\Delta t} \varphi(t)e^{-ivt}dt &= \int_{k\Delta t}^{(k+1)\Delta t} \varphi(t)\left[e^{-ivk\Delta t} - ive^{-ivk\Delta t}(t - k\Delta t) + O(\Delta t^2)\right]dt \\
&= \int_{k\Delta t}^{(k+1)\Delta t} \varphi(t)e^{-ivk\Delta t}dt - Q_k + O(\Delta t^3),
\end{aligned}
\tag{A4}
$$

where

$$
Q_k = ive^{-ivk\Delta t}\int_0^{\Delta t} \varphi(t + k\Delta t)t\,dt.
$$

By Taylor's theorem, for each $t \in [0, \Delta t]$, there exists $c \in (k\Delta t, (k+1)\Delta t)$ such that

$$
\varphi(t + k\Delta t) = \varphi(k\Delta t) + \dot{\varphi}(c)t = \varphi(k\Delta t) + \phi(c)t.
$$

Note that c generally depends on t. Then Q_k is approximated as

$$
Q_k = ive^{-ivk\Delta t}\int_0^{\Delta t}[\varphi(k\Delta t) + \phi(c)t]t\,dt = ive^{-ivk\Delta t} \cdot \frac{\varphi(k\Delta t)}{2}\Delta t^2 + O(\Delta t^3).
$$

Substituting into (A4), we obtain

$$
\int_{k\Delta t}^{(k+1)\Delta t} \varphi(t)e^{-ivk\Delta t}dt = \int_{k\Delta t}^{(k+1)\Delta t} \varphi(t)e^{-ivt}dt + ive^{-ivk\Delta t} \cdot \frac{\varphi(k\Delta t)}{2}\Delta t^2 + O(\Delta^3).
$$

Therefore, for $n \gg 1$ and $\Delta t \sim n^{-1}$, it follows that

$$
\begin{aligned}
\sum_{k=0}^{n-1} \int_{k\Delta t}^{(k+1)\Delta t} \varphi(t)e^{-ivk\Delta t}dt &= \int_0^{n\Delta t} \varphi(t)e^{-ivt}dt + \frac{iv\Delta t}{2}\sum_{k=0}^{n-1}\varphi(k\Delta t)e^{-ivk\Delta t}\Delta t + O(\Delta t^2) \\
&= \int_0^{n\Delta t} \varphi(t)e^{-ivt}dt + \frac{iv\Delta t}{2}\int_0^{n\Delta t} \varphi(t)e^{-ivt}dt + O(\Delta t^2).
\end{aligned}
$$

Substituting back into (A3) yields

$$
\begin{aligned}
\widehat{\phi}(v\Delta t) &= \lim_{n\to\infty}\left[\left(\int_0^{n\Delta t} \varphi(t)e^{-ivt}dt + \frac{iv\Delta t}{2}\int_0^{n\Delta t} \varphi(t)e^{-ivt}dt + O(\Delta t^2)\right)\right. \\
&\qquad \left. \times \left(iv + \frac{v^2\Delta t}{2} + O(\Delta t^2)\right) + \psi_n e^{-ivn\Delta t}\right] \\
&= \lim_{n\to\infty}\left(iv\int_0^{n\Delta t} \varphi(t)e^{-ivt}dt + \psi_n e^{-ivn\Delta t}\right) + O(\Delta t^2).
\end{aligned}
$$

The first term on the right-hand side can be further simplified as

$$iv \int_0^{n\Delta t} \varphi(t)e^{-ivt}dt = -\left[\varphi(t)e^{-ivt}\right]_0^{n\Delta t} + \int_0^{n\Delta t} \dot{\varphi}(t)e^{-ivt}dt$$

$$= -\varphi(n\Delta t)e^{-ivn\Delta t} + \int_0^{n\Delta t} \dot{\varphi}(t)e^{-ivt}dt.$$

Consequently,

$$\widehat{\phi}(v\Delta t) = \lim_{n\to\infty}\left(\int_0^{n\Delta t} \phi(t)e^{-ivt}dt - \varphi(n\Delta t)e^{-ivn\Delta t} + \psi_n e^{-ivn\Delta t}\right) + O(\Delta t^2)$$

$$= \int_0^{\infty} \phi(t)e^{-ivt}dt + O(\Delta t^2),$$

where we have used the identity $\lim_{n\to\infty}\psi_n = \varphi(\infty)$.   □

Now we complete a proof of Theorem 2. According to equation (17), the spectral density of the univariate coarse-grained Hawkes process is expressed as

$$f^{(cg)}_{\Delta t}(\omega) = \frac{\lambda\Delta t}{2\pi}\frac{1}{|1-\widehat{\phi}(\omega)|^2}.$$

Substituting equation (A2) into the expression above yields

$$f^{(cg)}_{\Delta t}(v\Delta t) = \frac{\lambda\Delta t}{2\pi}\frac{1}{|1-\widetilde{\phi}(v)+O(\Delta t^2)|^2}$$

$$= \frac{\lambda\Delta t}{2\pi}\left(\frac{1}{|1-\widetilde{\phi}(v)|^2}+O(\Delta t^2)\right)$$

$$= f^{(hw)}(v)\Delta t + O(\Delta t^3),$$

thereby completing the proof of Theorem 2 in the univariate setting.

For comparison, consider the binned Poisson approximation introduced in equations (1)–(2), wherein the discretized excitation kernel is defined by $\phi_k = \phi(k\Delta t)\Delta t$ for $k = 1, 2, \ldots$. The discrete-time Fourier transform of this kernel is given by

$$\widehat{\phi}(v\Delta t) = \sum_{k=1}^{\infty}\phi(k\Delta t)e^{-ivk\Delta t}\Delta t = \widetilde{\phi}(v) + O(\Delta t),$$

as $\Delta \to 0$. Observe that the approximation order is lower than that of equation (A2). Consequently, the spectral density of the binned Poisson approximation is asymptotically given by

$$f^{(po)}_{\Delta t}(v\Delta t) = f^{(hw)}(v)\Delta t + O(\Delta t^2).$$

Appendix B. Second-order properties of the stationary Hawkes process

This appendix provides a concise summary of the second-order statistical characteristics of the stationary Hawkes process. We present only those results pertinent to the analysis in this paper, and refer the reader to [22,28,29] for detailed derivations.

• **Spectral Density Matrix of the Stationary Hawkes Process**:

$$F^{(hw)}(v) = \frac{1}{2\pi}(I-\widetilde{\Phi}(v))^{-1}\text{diag}(\boldsymbol{\lambda})(I-\widetilde{\Phi}^{\mathsf{T}}(v))^{-1}, \quad v \in (-\infty, \infty), \qquad (A5)$$

where $\widetilde{\Phi}(\nu) = \int_0^\infty \Phi(t)e^{-i\nu t}dt$ denotes the Fourier transform of the excitation kernel matrix, and

$$\lambda = E[\lambda(t)] = (I - A)^{-1}\mu,$$

represents the stationary (mean) intensity.

- **Expected Value of the Binned Stationary Hawkes Process**:

$$E[X_k] = \lambda\Delta t = (I - A)^{-1}\mu\Delta t. \tag{A6}$$

- **Spectral Density Matrix of the Binned Stationary Hawkes Process**:

$$F_{\Delta t}^{(hw)}(\omega) = \Delta t \sum_{k=-\infty}^{\infty} \text{sinc}^2\left(\frac{\omega + 2\pi k}{2}\right) F^{(hw)}\left(\frac{\omega + 2\pi k}{\Delta t}\right), \quad \omega \in [-\pi, \pi). \tag{A7}$$

To evaluate (A7) in the limit as $\Delta t \to 0$, we employ the following asymptotic expansion:

$$\text{sinc}\left(\frac{\nu\Delta t + 2\pi k}{2}\right) = \begin{cases} 1 - \frac{\nu^2\Delta t^2}{12} + O(\Delta t^4), & k = 0, \\ \frac{\nu^2\Delta t^2}{4\pi^2 k^2} + O(\Delta t^3), & k \neq 0. \end{cases}$$

Substituting into (A7) yields the approximation:

$$\begin{aligned} F_{\Delta t}^{(hw)}(\nu\Delta t) &= \left(1 - \frac{\nu^2\Delta t^2}{12} + O(\Delta t^4)\right) F^{(hw)}(\nu)\Delta t \\ &+ \sum_{k\neq 0}\left(\frac{\nu^2\Delta t^2}{4\pi^2 k^2} + O(\Delta t^3)\right) F^{(hw)}\left(\nu + \frac{2\pi k}{\Delta t}\right)\Delta t \\ &= F^{(hw)}(\nu)\Delta t + O(\Delta t^3). \end{aligned}$$

Appendix C. Power-Law Distribution

A power-law distribution for waiting times is defined as

$$g(t) = \frac{\beta\gamma}{(1 + \beta t)^{1+\gamma}}, \quad t \geq 0,$$

for $\beta > 0$ and $\gamma > 0$. It is well-known that the moments of a power-law distribution exist and are finite for all orders strictly less than the exponent γ. In particular, the expected waiting time is finite if $\gamma > 1$, and is given by

$$\int_0^\infty t g(t)dt = \frac{1}{\beta(\gamma - 1)}.$$

The corresponding coarse-grained kernel is given by

$$g_k = \begin{cases} 1 - \left[(1+\beta\Delta)^{1-\gamma} - 1\right]/\beta(1-\gamma)\Delta, & k = 0, \\ -\left[\{1+\beta(k+1)\Delta\}^{1-\gamma} - 2\{1+\beta k\Delta\}^{1-\gamma} + \{1+\beta(k-1)\Delta\}^{1-\gamma}\right]/\beta(1-\gamma)\Delta, & k = 1, 2, \ldots, \end{cases}$$
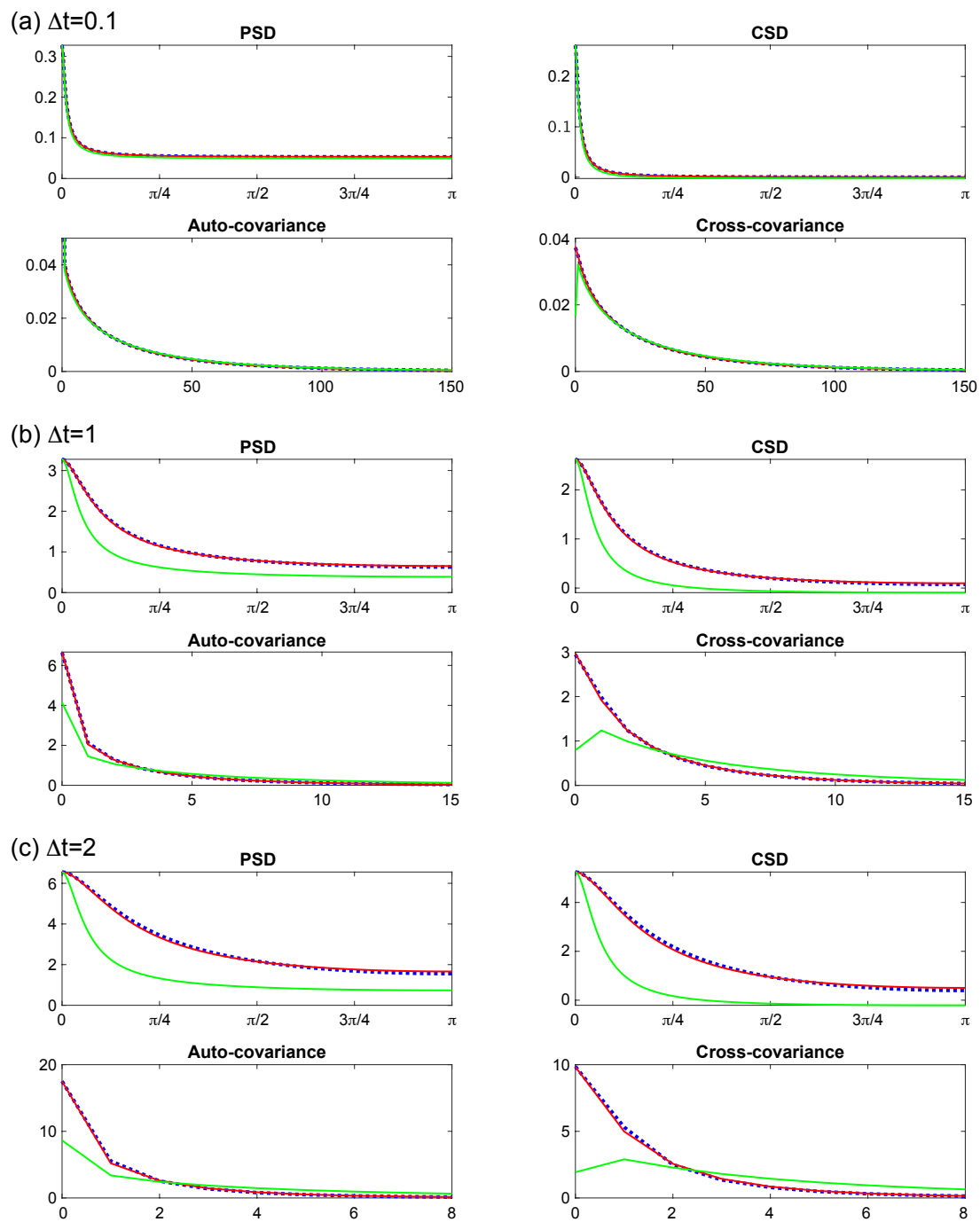
for $\gamma \neq 1$, and

$$g_k = \begin{cases} 1 - \log(1+\beta\Delta)/\beta\Delta, & k = 0, \\ -[\log\{1+\beta(k+1)\Delta\} - 2\log\{1+\beta k\Delta\} + \log\{1+\beta(k-1)\Delta\}]/\beta\Delta, & k = 1, 2, \ldots, \end{cases}$$

for $\gamma = 1$.

Figure A1 illustrates the power spectral density (PSD), cross-spectral density (CSD), auto-covariance, and cross-covariance of the binned Hawkes process (blue dotted lines), the coarse-grained Hawkes process (red lines), and the binned Poisson approximation (green lines), respectively. Fig-

ure A2 presents the information loss associated with the coarse-grained Hawkes process (red lines) and the binned Poisson approximation (green lines), respectively. These results are qualitatively consistent with those obtained using the exponential kernel, as shown in Figures 1–2.

Figure A1. Same as Figure 1, but for the power-law kernel. Parameters of the Hawkes process are set to $\mu = 1$, $\alpha^{(s)} = 0.4$, $\alpha^{(c)} = 0.3$, $\beta = 1$, and $\gamma = 2$.
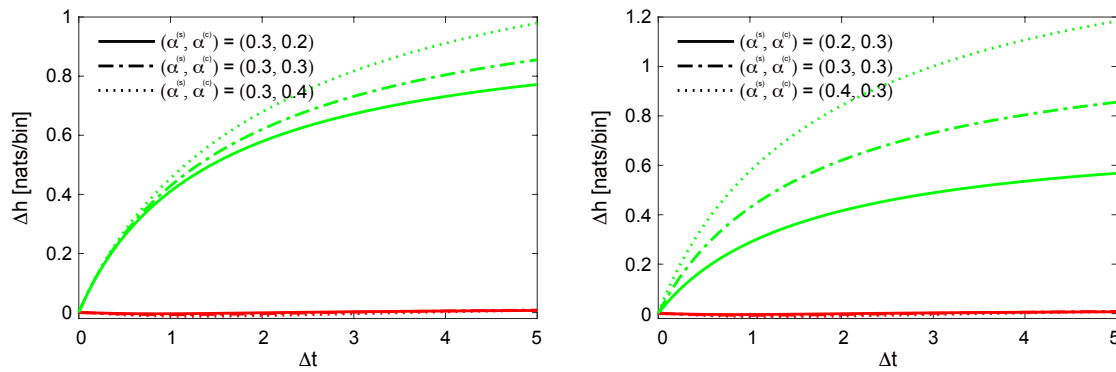
Figure A2. Same as Figure 2, but for the power-law kernel. Parameters are set to $\mu = 1$, $\beta = 1$, and $\gamma = 2$.

References

1. Hawkes, A.G. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **1971**, *58*, 83–90.
2. Hawkes, A.G. Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological)* **1971**, *33*, 438–443.
3. Adamopoulos, L. Cluster models for earthquakes: Regional comparisons. *Journal of the International Association for Mathematical Geology* **1976**, *8*, 463–475.
4. Ogata, Y. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association* **1988**, *83*, 9–27.
5. Chornoboy, E.S.; Schramm, L.P.; Karr, A.F. Maximum likelihood identification of neural point process systems. *Biological Cybernetics* **1988**, *59*, 265–275.
6. Pernice, V.; Staude, B.; Cardanobile, S.; Rotter, S. How structure determines correlations in neuronal networks. *PLoS Computational Biology* **2011**, *7*, e1002059.
7. Reynaud-Bouret, P.; Schbath, S. Adaptive estimation for Hawkes processes; application to genome analysis. *The Annals of Statistics* **2010**, *38*, 2781–2822.
8. Bacry, E.; Mastromatteo, I.; Muzy, J.F. Hawkes processes in finance. *Market Microstructure and Liquidity* **2015**, *1*, 1550005.
9. Hawkes, A.G. Hawkes processes and their applications to finance: a review. *Quantitative Finance* **2018**, *18*, 193–198.
10. Fox, E.W.; Short, M.B.; Schoenberg, F.P.; Coronges, K.D.; Bertozzi, A.L. Modeling E-mail Networks and Inferring Leadership Using Self-Exciting Point Processes. *Journal of the American Statistical Association* **2016**, *111*, 564–584.
11. Kobayashi, R.; Lambiotte, R. TiDeH: time-dependent Hawkes process for predicting retweet dynamics. *Proceedings of the International AAAI Conference on Web and Social Media* **2016**, *10*, 191–200.
12. Koyama, S.; Shinomoto, S. Statistical physics of discovering exogenous and endogenous factors in a chain of events. *Physical Review Research* **2020**, *2*, 043358.
13. Mohler, G.; Short, M.B.; Brantingham, P.J.; Schoenberg, F.P.; Tita, G.E. Self-Exciting Point Process Modeling of Crime. *Journal of the American Statistical Association* **2011**, *106*, 100–108.
14. Zhuang, J.; Mateu, J. A semiparametric spatiotemporal Hawkes-type point process model with periodic background for crime data. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* **2019**, *182*, 919–942.
15. Lewis, E.; Mohler, G.; Brantingham, P.J.; Bertozzi, A.L. Self-exciting point process models of civilian deaths in Iraq. *Security Journal* **2012**, *25*, 244–264.
16. Kalair, K.; Connaughton, C.; Loro, P.A.D. A non-parametric Hawkes process model of primary and secondary accidents on a UK smart motorway. *Journal of the Royal Statistical Society Series. C (Applied Statistics)* **2021**, *70*, 80–97.
17. Kirchner, M. Hawkes and INAR(∞) processes. *Stochastic Processes and their Applications* **2016**, *126*, 2494–2525.
18. Kirchner, M. An estimation procedure for the Hawkes process. *Quantitative Finance* **2017**, *17*, 571–595.
19. Shlomovich, L.; Cohen, E.A.K.; Adams, N.; Patel, L. Parameter estimation of binned Hawkes processes. *Journal of Computational and Graphical Statistics* **2022**, *31*, 990–1000.
20. Shlomovich, L.; Cohen, E.A.K.; Adams, N. A parameter estimation method for multivariate binned Hawkes processes. *Statistics and Computing* **2022**, *32*, 98.

21. Chen, F.; Kwan, T.K.J.; Stindl, T. Estimating the Hawkes Process From a Discretely Observed Sample Path. *Journal of Computational and Graphical Statistics* **2025**, *0*, 1–13.

22. Cheysson, F.; Lang, G. Spectral estimation of Hawkes processes from count data. *The Annals of Statistics* **2022**, *50*, 1722 – 1746.

23. Daley, D.; Vere-Jones, D. *An Introduction to the Theory of Point Processes Volume II: General Theory and Structure*, 2nd ed.; Springer: New York, 2008.

24. Mark, B.; Raskutti, G.; Willett, R. Network estimation from point process data. *IEEE Transactions on Information Theory* **2019**, *65*, 2953–2975.

25. Shlomovich, L. https://github.com/lshlomovich/MCEM_Multivariate_Hawkes.

26. Durbin, J.; Koopman, S. *Time series analysis by state space methods*; Oxford University Press, 2001.

27. Kitagawa, G. *Introduction to Time Series Modeling*; Chapman and Hall/CRC, 2010.

28. Bacry, E.; Dayri, K.; Muzy, J.F. Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data. *The European Physical Journal B* **2012**, *85*, 157.

29. Bacry, E.; Muzy, J.F. First- and second-order statistics characterization of Hawkes processes and non-parametric estimation. *IEEE Transactions on Information Theory* **2016**, *62*, 2184–2202.