Parametric inference for Hawkes processes with a general kernel

Yoann Potiron[1],

[1] *Faculty of Business and Commerce, Keio University, e-mail:* potiron@fbc.keio.ac.jp

Abstract: We develop inference for point processes when its intensity has a parametric form. The inference procedure is based on maximum likelihood estimation. Under ergodicity of the point process intensity and its derivative, we show the central limit theorem of the inference procedure. As an application, we consider Hawkes mutually exciting processes. We assume that the kernel has a general form and is parametric. We show the ergodicity of the Hawkes process intensity and its derivative. Moreover, we obtain the central limit theorem of the inference procedure for Hawkes processes. In particular, we allow for kernels with power distribution, under some smoothness assumptions on the kernel shape. The proofs are based on the application of Burkholder—Davis—Gundy inequalites.

Keywords and phrases: point processes, Hawkes mutually exciting processes, parametric inference, general kernel, central limit theorem, ergodicity.

1. Introduction

This paper concerns parametric inference for point processes. The main stylized fact in this strand of literature, the presence of event clustering in time, motivates the so-called Hawkes mutually exciting processes (see Hawkes (1971a) and Hawkes (1971b)). We define the point process N_t of dimension d as the number of events from the starting time 0 to the time t and λ_t its intensity. A standard definition of Hawkes mutually exciting processes is given by

$$\lambda_t = \nu^* + \int_0^t h(t-s)\, dN_s. \tag{1}$$

Here, ν^* is a d dimensional Poisson baseline and h is a $d \times d$ dimensional kernel matrix. The point processes are mutually exciting in the sense that the diagonal components $h^{(i,i)}$ are self-exciting terms for the i-th process and non-diagonal components $h^{(i,j)}$ are cross-exciting terms for the i-th process made by events from the j-th process. The particular case $h = 0$ corresponds to a classical Poisson process, so that we can view Hawkes processes as a natural extension of Poisson processes.

The main application of Hawkes processes lies in seismology (see Rubin (1972), Ozaki (1979), Vere-Jones and Ozaki (1982) and Ogata (1978), Ogata (1988)). There are also applications in quantitative finance (see Chavez-Demoulin, Davison and McNeil (2005), Embrechts, Liniger and Lin (2011), Bacry et al. (2013), Jaisson and Rosenbaum (2015), Jaisson and Rosenbaum (2016), Clinet and Yoshida (2017)). Some applications are also in financial econometrics (see Chen and Hall (2013), Clinet and Potiron (2018), Kwan, Chen and Dunsmuir (2023), Potiron and Volkov (2025)). We can also find some applications in biology (see Reynaud-Bouret and Schbath (2010) and Donnet, Rivoirard and Rousseau (2020)). See also Liniger (2009) and Hawkes (2018) with the references therein.

There are many theoretical results for Hawkes processes in statistics. Hawkes and Oakes (1974) provide a Poisson cluster process representation for the Hawkes process. Brémaud and Massoulié (1996) study stability of nonlinear Hawkes processes. Zhu (2015) considers large deviations for Markovian nonlinear Hawkes processes. Roueff, von Sachs and Sansonnet (2016), Roueff and Von Sachs (2019), Cheysson and Lang (2022), Mammen and Müller (2023), Erdemlioglu et al. (2025a) study locally stationary Hawkes processes. Potiron et al. (2025a) and Potiron et al. (2025b) introduce Hawkes processes with Itô semimartingale baseline. Potiron (2025) consider a more general baseline. The microstructure of stochastic volatility models with self-excitation is investigated in Horst and Xu (2022). Horst and Xu (2021) and Horst and Xu (2024+) give functional limit theorems for Hawkes processes. Xu (2024) studies diffusion approximations for self-excited systems. Karim, Laeven and Mandjes (2025+) introduce compound multivariate Hawkes processes.

In this paper, we consider Hawkes processes, where the kernel has a general form and is parametric. The inference procedure is based on maximum likelihood estimation (MLE). Ogata (1978) shows the central limit theorem (CLT) of the inference procedure for an ergodic stationary point process. However, the definition of ergodicity is vague in that paper. Most of the papers on parametric inference for Hawkes processes make this ergodicity assumption (see Bowsher (2007), Large (2007) and Cavaliere et al. (2023), Assumption 1(b) and Remark 2.1).

In fact, Clinet and Yoshida (2017) exhibit the conditions required, i.e. ergodicity of the Hawkes intensity process and its derivative jointly. They consider general point processes and derive the CLT of the inference procedure in Theorem 3.11 (p. 1809) under these ergodicity assumptions. They also show these ergodicity assumptions in the case of a Hawkes process with exponential kernel in Theorem 4.6 (p. 1821). The proofs rely heavily on the Markov property of the exponential distribution.

Kwan (2023) considers the non-exponential kernel case but the author mentions that such case is challenging since the Hawkes intensity process is non-Markovian, thus rendering standard Markov tools inapplicable. Consequently, the author can only show the ergodicity for the Hawkes intensity process and for its derivative (see Theorem 4.3.2, p. 91), but not jointly. Thus, he can only show the consistency of the inference procedure in Theorem 3.4.3 (p. 73). When the kernel follows a generalized gamma distribution, Potiron and Volkov (2025) can show that the ergodicity assumptions are satisfied and also obtain the CLT of the inference procedure. This is due to the exponentially decreasing nature of the kernel. With a general kernel, Costa et al. (2020) (Theorem 1.2, p. 884) and Graham (2021) (Theorem 5.4, p. 2856) shows the ergodicity of the Hawkes processes, but not its intensity. See also Section 3.2 (p. 893) in Reynaud-Bouret and Roy (2007).

All these results are useful, but the obtained CLT for the inference procedure are restricted to exponentially decreasing kernels, which are restrictive for applications. In finance, there is empirical evidence that the kernel decays as the power distribution (see Bacry, Dayri and Muzy (2012) and Hardiman, Bercot and Bouchaud (2013)). Consequently, we extend the literature in two directions. First, we weaken the assumptions from the point process theory in Clinet and Yoshida (2017), since they do not allow for kernels with power distribution. More specifically, we consider a different approach in the proofs that circumvents the use of the Sobolev embedding theorem (see Theorem 4.12 (p. 85) in Adams and Fournier (2003)). This new approach is based on the application of Burkholder—Davis—Gundy inequalites (see Expression (2.1.32) in Jacod and Protter (2012) (p. 39)). Under ergodicity of the point process intensity and its derivative, we show the CLT of the inference procedure (see Theorem 1). Second, we show the ergodicity of the Hawkes intensity process and its derivative, in case of a general kernel. Moreover, we show the CLT of the inference procedure (see Theorem 2). This is the main result of this paper. This extends Clinet and Yoshida (2017) and Potiron and Volkov (2025), who obtain ergodicity by proving first that the Hawkes intensity process and its derivative is mixing by Markov arguments. In particular, we allow for kernels with power distribution, under some smoothness assumptions on the kernel shape.

The remainder of this paper is organized as follows. We consider inference for point processes in Section 2. We study the Hawkes processes with a general kernel case in Section 3. The proofs of the CLT for point processes are gathered in Section 4. The proofs of the CLT for Hawkes processes with a general kernel are given in Section 5. Finally, we provide concluding remarks in Section 6.

2. Parametric inference for point processes

In this section, we develop inference for point processes when its intensity has a parametric form. The inference procedure is based on MLE. Under ergodicity of the point process intensity and its derivative, we show the CLT of the inference procedure. In particular, we weaken the assumptions from the point process theory in Clinet and Yoshida (2017), since they do not allow for kernels with power distribution. More specifically, we consider a different approach in the proofs that circumvents the use of the Sobolev embedding theorem. This new approach is based on the application of Burkholder—Davis—Gundy inequalites.

We start with an introduction to the point process N_t of dimension d. For any index $i = 1, \dots, d$, each component of the point process $N_t^{(i)}$ counts the number of events between 0 and t for the i-th process. Here, we denote the i-th component of a vector V by $V^{(i)}$. We define $N_t^{(i)}$ as a simple point process on the space $\mathbb{R}^+$, i.e. a family $\{N^{(i)}(C)\}_{C \in \mathcal{B}(\mathbb{R}^+)}$ of random variables with values in the space $\overline{\mathbb{N}} = \mathbb{N} \cup \{+\infty\}$. Here, $\mathcal{B}(S)$ denotes the Borel σ-algebra on the space S for any space S. Moreover, $N^{(i)}(C) = \sum_{k \in \mathbb{N}} \mathbf{1}_C(T_k^{(i)})$ and $\{T_k^{(i)}\}_{k \in \mathbb{N}}$ is a sequence of event times, which are $\mathbb{R}^+$-valued and random. We assume that the time of the first event $T_0^{(i)}$ is equal to 0 a.s. and the following times are increasing for each process a.s. Namely, we assume that

$$\mathbb{P}(T_0^{(i)} = 0 \text{ and } T_k^{(i)} < T_{k+1}^{(i)} \text{ for } k \in \mathbb{N}_* \text{ and } i = 1, \dots, d) = 1. \tag{2}$$

Here, we define for any space S such that $0 \in S$ the space without zero as S_*. We also assume that no events happen at the same time for different processes a.s., i.e. $\mathbb{P}(T_k^{(i)} \neq T_l^{(j)}$ for $k, l \in \mathbb{N}_*$ and $i, j = 1, \dots, d$ s.t. $i \neq j) = 1$.

Let $\mathcal{B} = (\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in \mathbb{R}^+}, \mathbb{P})$ be a filtered probability space which satisfies the usual conditions. For any $t \in \mathbb{R}$, we denote the filtration generated by some stochastic process X as $\mathcal{F}_t^X = \sigma\{X_s : 0 \leq s \leq t\}$. We assume that, for any $t \in \mathbb{R}^+$, the filtration generated by the point process N_t is included in the main filtration, i.e. $\mathcal{F}_t^N \subset \mathcal{F}_t$. Any stochastic process $\{\lambda_t\}_{t \in \mathbb{R}^+}$ which satisfies the following properties is called an $\mathcal{F}_t$-intensity of N_t. First, we have that

$$\mathbb{E}[N((a, b]) \mid \mathcal{F}_a] = \mathbb{E}\Big[\int_a^b \lambda_s ds \mid \mathcal{F}_a\Big] \text{ a.s.}$$

for all intervals $(a, b] \subset \mathbb{R}^+$. Second, the process λ_t is $\mathcal{F}_t$-progressively measurable, of dimension d where each component $\lambda_t^{(i)}$ takes its values in the space of nonnegative real numbers $\mathbb{R}^+$. Intuitively, the intensity corresponds to the expected number of events given the past information, i.e.

$$\lambda_t = \lim_{u \to 0} \mathbb{E}\Big[\frac{N_{t+u} - N_t}{u} \mid \mathcal{F}_t\Big] \text{ a.s.}$$

We note that $N((a, b])$ is a.s. finite if and only if $\int_a^b \lambda_s ds$ is a.s. finite. For background on point processes, the reader can consult Jacod (1975), Jacod and Shiryaev (2003), Daley and Vere-Jones (2003), and Daley and Vere-Jones (2008).

The present work is concerned with multidimensional point processes N_t admitting an $\mathcal{F}_t$-intensity which has a parametric form. More specifically, we introduce the parameter space Θ, consisting of n parameters. We also introduce the family of intensities $\lambda_t(\theta)$ for any parameter $\theta \in \Theta$. We assume that the intensity process $\lambda_t(\theta)$ is of dimension d where each component $\lambda_t^{(i)}(\theta)$ takes its value in the space of nonnegative real numbers $\mathbb{R}^+$ for any parameter $\theta \in \Theta$. Finally, we assume the existence of the true parameter $\theta^* \in \Theta$ such that

$$\lambda_t = \lambda_t(\theta^*). \tag{3}$$

For any parameter $\theta \in \Theta$, we rely on the log likelihood process (see Ogata (1978) and Daley and Vere-Jones (2003))

$$l_T(\theta) = \sum_{i=1}^d \int_0^T \log(\lambda_t^{(i)}(\theta)) dN_t^{(i)} - \sum_{i=1}^d \int_0^T \lambda_t^{(i)}(\theta) dt. \tag{4}$$

Here, 0 is the starting time and T is the final time. Then, the MLE is defined as a maximizer of the log likelihood process between 0 and T, i.e.

$$\widehat{\theta}_T \in \operatorname{argmax}_{\theta \in \Theta} l_T(\theta).$$

In this paper, we focus on the stochastic processes $X_t = (\lambda_t(\theta^*), \lambda_t(\theta), \partial_\theta \lambda_t(\theta))$ taking values in the space E^d where $E = \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^n$. We denote by $X_t^{(i)} \in E$ the i-th component of X_t. Here, $\partial_\theta G(\theta)$ denotes the vector of partial derivatives for any function $G(\theta)$, i.e. $\partial_\theta G(\theta) = \frac{\partial G}{\partial \theta}(\theta)$. We denote by $C_b(E, F)$ the space of bounded and continuous functions from E to F. In what follows, we provide the definition of ergodicity. This corresponds to Definition 3.1 (p. 1805) in Clinet and Yoshida (2017). See also Definition C1 in Supplement C of Potiron and Volkov (2025).

Definition 1. We say that X is ergodic if for any $i = 1, \dots, d$ there exists a function $\pi^{(i)} : C_b(E, \mathbb{R}) \to \mathbb{R}$ such that for any $\psi \in C_b(E, \mathbb{R})$ we have

$$\frac{1}{T} \int_0^T \psi(X_s^{(i)}) ds \xrightarrow{\mathbb{P}} \pi^{(i)}(\psi).$$

Since the space of bounded functions is not large enough to establish CLT, we introduce a bigger space in the following definition. We denote this bigger space by $C_\uparrow(E, \mathbb{R})$. This corresponds to Definition 3.7 (p. 1806) in Clinet and Yoshida (2017).

Definition 2. We denote by $C_\uparrow(E, \mathbb{R})$ the set of continuous functions $\psi : (u, v, w) \to \psi(u, v, w)$ from E to $\mathbb{R}$ that satisfy

(a) ψ is continuous on $\mathbb{R}_*^+ \times \mathbb{R}_*^+ \times \mathbb{R}^n$.
(b) ψ is of polynomial growth in u, v, w, $\frac{\mathbf{1}_{\{u>0\}}}{u}$ and $\frac{\mathbf{1}_{\{v>0\}}}{v}$.
(c) For any $(u, v, w) \in E$, we have $\psi(0, v, w) = \psi(u, 0, w) = 0$.

Lemma 1 extends the starting space of the limit function π from $C_b(E, \mathbb{R})$ to $C_\uparrow(E, \mathbb{R})$ and gives a more explicit form. More specifically, it shows that, for any index $i = 1, \dots, d$ and any parameter $\theta \in \Theta$, there exists a probability measure $\pi_\theta^{(i)}$ on the space E such that, for any $\psi \in C_\uparrow(E, \mathbb{R})$, we have

$$\pi^{(i)}(\psi) = \int_E \psi(u, v, w) \pi_\theta^{(i)}(du, dv, dw).$$

Moreover, Lemma 2 ensures that the family of intensities $\lambda_t(\theta)$ does not explode on any compact space based on these ergodicity assumptions.

We have now all the ingredients to derive the form of the asymptotic covariance matrix. If we consider a vector $z \in \mathbb{R}^n$, we define the tensor product as $z^{\otimes 2} = zz^T \in \mathbb{R}^{n \times n}$. Then, we define the asymptotic Fisher information matrix Γ of dimension $n \times n$ as

$$\Gamma = \sum_{i=1}^d \int_E w^{\otimes 2} \frac{1}{u} \pi_{\theta^*}^{(i)}(du, dv, dw). \tag{5}$$

The Fisher information matrix measures the amount of information that the intensity λ_t carries about the parameter θ^*. Formally, it is the expected value of the observed information. The Fisher information matrix is used to calculate the covariance matrices associated with MLE. In other words, Γ^{-1} is the asymptotic covariance matrix. In Expression (37) from the proof of Theorem 1, we show that we can reexpress the asymptotic Fisher information matrix as $\Gamma = -\lim_{T\to\infty}\frac{1}{T}\mathbb{E}\big[\partial_\theta^2 l_T(\theta^*)\big]$. Here, $\partial_\theta^2 G(\theta)$ denotes the Hessian matrix for any function $G(\theta)$. i.e. $\partial_\theta^2 G(\theta) = \frac{\partial^2 G}{\partial^2\theta}(\theta)$.

Before introducing the assumptions, we first need to introduce some notation. We define $\overline{\Theta}$ as the closure space of Θ. If x is a real number, a vector, a matrix or a tensor, we define the sum of the absolute values of its components as $|x| = \sum_i |x_i|$. If Y is a random variable, we define its L^p norm as $||Y||_p = \mathbb{E}[|Y|^p]^{1/p}$. Then, we define the limit of the normalized deviation between the log likelihood at the parameter value $\theta\in\Theta$ and the log likelihood at the true parameter value θ^* as

$$Y(\theta) = \sum_{i=1}^{d}\int_E \Big(\log\Big(\frac{v}{u}\Big)u - (v-u)\Big)\mathbf{1}_{\{u>0,v>0\}}\pi_\theta^{(i)}(du,dv,dw). \tag{6}$$

As the true parameter value θ^* is a maximum argument of the log likelihood limit, we have by definition that $Y(\theta)\leq 0$ for any parameter $\theta\in\Theta$. Finally, we define F for any $i=1,\dots,d$ and any parameter $\theta\in\Theta$ as

$$F_t^{(i)}(\theta) = \frac{(\partial_\theta\lambda_t^{(i)})^{\otimes 2}(\theta)}{\lambda_t^{(i)}(\theta)^2}.$$

We now introduce a set of assumptions required for the CLT of the parametric inference procedure based on MLE. In particular, we weaken the assumptions from the point process theory in Clinet and Yoshida (2017), since they do not allow for kernels with power distribution.

Assumption 1. (a) The family of intensities $\lambda : \Omega\times\mathbb{R}^+\times\Theta\to\mathbb{R}_+^d$ is $\mathcal{F}\otimes\mathcal{B}(\mathbb{R}^+)\otimes\mathcal{B}(\Theta)$-measurable.

(b) We assume that $\Theta\subset\mathbb{R}^n$ is such that its closure $\overline{\Theta}$ is a compact space.
(c) For any $\theta\in\Theta$, the stochastic processes X are ergodic in the sense of Definition 1.
(d) We have $\sup_{t\in\mathbb{R}^+}||\sup_{\theta\in\Theta}|\lambda_t(\theta)|\,||_2 < +\infty$.
(e) For any $\theta\in\overline{\Theta}-\{\theta^*\}$, we have $Y(\theta)\neq 0$.
(f) For any time $s\in\mathbb{R}^+$ a.e., we have a.s. that the intensity process $\theta\to\lambda_s(\theta)$ is twice continuously differentiable from the parameter space Θ to the space $\mathbb{R}_+^d$ and there exists a continuous extension to $\overline{\Theta}$.
(g) For any $\theta\in\Theta$ and $T>0$, we have $\mathbb{P}(\int_0^T|\partial_\theta\lambda_t(\theta)|dt<\infty) = 1$ and $\mathbb{P}(\int_0^T|\partial_\theta^2\lambda_t(\theta)|dt<\infty)=1$.
(h) For any $i=1,\dots,d$, we have

$$\begin{aligned}\sup_{t\in\mathbb{R}^+}\Big|\Big|\,\Big|\partial_\theta\Big(\frac{\partial_\theta\lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)}\Big)\Big|\lambda_t^{(i)}(\theta^*)\mathbf{1}_{\{\lambda_t^{(i)}>0\}}\Big|\Big|_2 &< +\infty,\\ \sup_{t\in\mathbb{R}^+}\Big|\Big|\,\Big|\partial_\theta F_t^{(i)}(\theta^*)\Big|\lambda_t^{(i)}(\theta^*)\mathbf{1}_{\{\lambda_t^{(i)}>0\}}\Big|\Big|_1 &< +\infty,\\ \sup_{t\in\mathbb{R}^+}\Big|\Big|\,\Big|(\partial_\theta^2\lambda_t^{(i)})(\theta^*)\Big|\,\Big|\frac{\partial_\theta\lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)}\Big|\mathbf{1}_{\{\lambda_t^{(i)}>0\}}\,\Big|\Big|_1 &< +\infty.\end{aligned}$$

Assumption 1 (a) is natural and is equal to the first statement in Assumption [A1] from Clinet and Yoshida (2017). Assumption 1 (b) is the assumption about the parameter space and is weaker than the framework from Clinet and Yoshida (2017) and Potiron and Volkov (2025) where the parameter space satisfies the assumptions from the Sobolev embedding theorem. Assumption 1 (c) corresponds exactly to Assumption [A3] from Clinet and Yoshida (2017). Assumptions 1 (d), (g) and (h) are weaker than Assumptions [A2] (i) and (ii) from Clinet and Yoshida (2017), who requires the finiteness of the L^p norm of intensity and some derivatives for any integer $p \in \mathbb{N}_*$. Assumption 1 (e) is required for the non-degeneracy of the inference procedure and is Assumption [A4] from Clinet and Yoshida (2017). Namely, we have that the true parameter value θ^* is the only maximum argument of the log likelihood limit. Finally, Assumption 1 (f) only requires that the intensity is twice continuously differentiable whereas Assumption [A1] (ii) in Clinet and Yoshida (2017) needs that the intensity is continuously differentiable three times.

In the theorem that follows, we state the CLT of the inference procedure based on MLE. We consider asymptotics when the final time diverges to infinity, i.e. $T \to +\infty$. In particular, we weaken the assumptions from the point process theory in Clinet and Yoshida (2017), since they do not allow for kernels with power distribution. More specifically, we consider a different approach in the proofs that circumvents the use of the Sobolev embedding theorem. This new approach is based on the application of Burkholder—Davis—Gundy inequalites. In the theorem and what follows, ξ is defined as an n dimensional standard normal vector.

Theorem 1. *We assume that Assumption 1 holds. As $T \to +\infty$, we have the CLT of the inference procedure based on MLE, i.e.*

$$\sqrt{T}(\widehat{\theta}_T - \theta^*) \xrightarrow{\mathcal{D}} \Gamma^{-1/2}\xi. \tag{7}$$

3. Hawkes processes with a general kernel case

In this section, we consider Hawkes mutually exciting processes. We assume that the kernel has a general form and is parametric. We show the ergodicity of the Hawkes process intensity and its derivative. Moreover, we obtain the CLT of the inference procedure for the Hawkes processes. This is the main result of this paper. In particular, we allow for kernels with power distribution, under some smoothness assumptions on the kernel shape.

We consider Hawkes processes, where the kernel has a general form and is parametric. More specifically, we introduce for any $\theta \in \Theta$ the family of intensities

$$\lambda_t(\theta) = \nu + \int_0^t h(t-s,\kappa)\, dN_s. \tag{8}$$

Here, h is a $d \times d$ dimensional kernel matrix. The point processes are mutually exciting in the sense that the diagonal components $h^{(i,i)}$ are self-exciting terms for the i-th process and non-diagonal components $h^{(i,j)}$ are cross-exciting terms for the i-th process made by events from the j-th process. Moreover, ν consists of d baseline parameters, while κ consists of $n-d$ kernel parameters. We assume that the parameter θ has the form $\theta = (\nu, \kappa)$, and that they belong to the parameter space $\Theta = (\Theta_\nu, \Theta_\kappa)$. We also assume that $n \geq 2d$. Finally, we assume the existence of the true parameter $\theta^* \in \Theta$ such that

$$\lambda_t = \lambda_t(\theta^*). \tag{9}$$

Here, we assume that the parameter θ^* has the form $\theta^* = (\nu^*, \kappa^*)$, where $\nu^* \in \Theta_\nu$ and $\kappa^* \in \Theta_\kappa$.

For a matrix ϕ, we denote its spectral radius as $\rho(\phi)$. For any $t \in \mathbb{R}^+$, we denote by κ_t^+ the maximum argument parameter of $\rho(h(t,\kappa))$. It is defined through

$$\rho(h(t,\kappa_t^+)) = \sup_{\kappa\in\Theta_\kappa} \rho(h(t,\kappa)). \tag{10}$$

Then, we define the $d \times d$ dimensional matrix ϕ as

$$\phi = \int_0^\infty h(s,\kappa_s^+)ds.$$

For any $t \in \mathbb{R}^+$, we denote by $\kappa_{t,2}^+$ the maximum argument parameter of $\rho(h^2(t,\kappa))$. It is defined through

$$\rho(h^2(t,\kappa_{t,2}^+)) = \sup_{\kappa\in\Theta_\kappa} \rho(h^2(t,\kappa)). \tag{11}$$

Then, we define the $d \times d$ dimensional matrix ϕ_2 as

$$\phi_2 = \int_0^\infty h^2(s,\kappa_{s,2}^+)ds.$$

For any $i = 1,\ldots,d$, $j = 1,\ldots,d$ and any $t \in \mathbb{R}^+$, we denote by $k_{t,3}^{(i,j)}$ the maximum argument of $\left|\partial_\kappa h^{(i,j)}(t,\kappa)^{(k)}\right|$. It is defined through

$$\left|\partial_\kappa h^{(i,j)}(t,\kappa)^{(k_{t,3}^{(i,j)})}\right| = \sup_{k=1,\ldots,n-d} \left|\partial_\kappa h^{(i,j)}(t,\kappa)^{(k)}\right|. \tag{12}$$

Then, we define the $d \times d$ dimensional matrix $\phi_3(\kappa)$ as

$$\phi_3^{(i,j)}(\kappa) = \int_0^\infty \left|\partial_\kappa h^{(i,j)}(s,\kappa)^{(k_{t,3}^{(i,j)})}\right|ds,$$

for any index $i = 1,\ldots,d$ and any index $j = 1,\ldots,d$. For any index $i = 1,\ldots,d$, any index $j = 1,\ldots,d$ and any time $t \in \mathbb{R}^+$, we denote by $(k_{t,4}^{(i,j)}, l_{t,4}^{(i,j)})$ the maximum argument of $\left|\partial_\kappa^2 h^{(i,j)}(t,\kappa)^{(k,l)}\right|$. It is defined through

$$\left|\partial_\kappa^2 h^{(i,j)}(t,\kappa)^{(k_{t,4}^{(i,j)},l_{t,4}^{(i,j)})}\right| = \sup_{k,l=1,\ldots,n-d} \left|\partial_\kappa^2 h^{(i,j)}(t,\kappa)^{(k,l)}\right|. \tag{13}$$

Then, we define the $d \times d$ dimensional matrix $\phi_4(\kappa)$ as

$$\phi_4^{(i,j)}(\kappa) = \int_0^{\infty} \left|\partial_\kappa^2 h^{(i,j)}(s,\kappa)^{(k_{t,4}^{(i,j)}, l_{t,4}^{(i,j)})}\right| ds,$$

for any index $i = 1, \dots, d$ and any index $j = 1, \dots, d$. Moreover, we define the $d \times d$ dimensional matrix ϕ_5 as

$$\phi_5^{(i,j)} = \int_0^{\infty} \left|\partial_\kappa h^{(i,j)}(s,\kappa^*)^{(k_{t,3}^{(i,j)})}\right|^2 ds,$$

for any index $i = 1, \dots, d$ and any index $j = 1, \dots, d$. Finally, we define the $d \times d$ dimensional matrix ϕ_6 as

$$\phi_6^{(i,j)} = \int_0^{\infty} \left|\partial_\kappa^2 h^{(i,j)}(s,\kappa^*)^{(k_{t,4}^{(i,j)}, l_{t,4}^{(i,j)})}\right|^2 ds,$$

for any $i = 1, \dots, d$ and $j = 1, \dots, d$.

Finally, Lemma 15 from Section 5 states that X_t is stable, i.e. for any $\theta \in \Theta$ and any $i = 1, \dots, d$ there exists an $\mathbb{R}_+^*$-valued random variable $\lambda_l^{(i)}(\theta)$ such that we have

$$X_T^{(i)} \xrightarrow{\mathcal{D}} \left(\lambda_l^{(i)}(\theta^*), \lambda_l^{(i)}(\theta), \partial_\theta \lambda_l^{(i)}(\theta)\right).$$

From Lemma 16, X_t is also ergodic in the sense of Definition 1 for any $\theta \in \Theta$. Moreover, for any $i = 1, \dots, d$ we have the more explicit expression of the limit function as

$$\pi^{(i)}(\psi) = \mathbb{E}\left[\psi(\lambda_l^{(i)}(\theta^*), \lambda_l^{(i)}(\theta), \partial_\theta \lambda_l^{(i)}(\theta))\right].$$

We now introduce a set of assumptions required for the CLT of the parametric inference procedure for Hawkes processes. In particular, we allow for kernels with power distribution, under some smoothness assumptions on the kernel shape.

Assumption 2. (a) There exists $\nu_- \in \mathbb{R}_*^+$ such that for any $\nu \in \Theta_\nu$ and any $i = 1, \dots, d$ we have $\nu^{(i)} > \nu_-$.

(b) For any kernel parameter $\kappa \in \Theta_\kappa$ and any time $t \in \mathbb{R}^+$, we have $h(t,\kappa) > 0$.

(c) We assume that $\Theta \subset \mathbb{R}^n$ is such that its closure $\overline{\Theta}$ is a compact space.

(d) We have $\rho(\phi) < 1$ and $\rho(\phi_2) < +\infty$.

(e) For any time $s \in \mathbb{R}^+$ a.e., we have the kernel function $\kappa \to h(s,\kappa)$ is twice continuously differentiable from the kernel parameter space Θ_κ to the space $\mathbb{R}_+^{d\times d}$ and there exists a continuous extension to $\overline{\Theta}_\kappa$.

(f) There exists $\phi_+ \in [0,1)$ such that for any $\kappa \in \Theta_\kappa$ we have $\rho(\phi_3(\kappa)) < \phi_+$ and $\rho(\phi_4(\kappa)) < \phi_+$.

(g) We have $\rho(\phi_5) < +\infty$ and $\rho(\phi_6) < +\infty$.

(h) We have $\mathbb{P}\big(\lambda_l(\theta^*) = \lambda_l(\theta)\big) = 1$ implies that $\theta^* = \theta$.

Assumption 2 (a) imply that the point processes are well-defined and is also required in the simpler case of heterogeneous Poisson processes without a kernel (see Daley and Vere-Jones (2003)). Assumption 2 (b) are restrictive for kernels with inhibitory effects. Assumption 2 (c) corresponds exactly to Assumption 1 (b) and is weaker than the framework from Clinet and Yoshida (2017) and

Potiron and Volkov (2025) where the parameter space satisfies the assumptions from the Sobolev embedding theorem.

The case $\rho(\phi) < 1$ in Assumption 2 (d) states that the spectral radius of the kernel integral when evaluated at the maximum argument parameter of $\rho(h(t, \kappa))$ is strictly smaller than unity. This is slightly stronger than the assumption which is necessary to obtain a stationary intensity with finite first moment (see Lemma 1 (p. 495) in Hawkes and Oakes (1974) and Theorem 1 (p. 1567) in Brémaud and Massoulié (1996)). The case $\rho(\phi_2) < +\infty$ in Assumption 2 (d) ensures that the spectral radius of the kernel integral, when squared and evaluated at the maximum argument parameter of $\rho(h^2(t, \kappa))$, is finite. Assumption 2 (d) is used to prove Assumption 1 (d).

Assumption 2 (e) requires some smoothness assumptions on the kernel shape and is used to show Assumption 1 (f). Moreover, the case $\rho(\phi_3(\kappa)) < \phi_+$ in Assumption 2 (f) states that the spectral radius of the kernel derivative integral when evaluated at the maximum argument of $\left|\partial_\kappa h^{(i,j)}(t, \kappa)^{(k)}\right|$ is strictly smaller than ϕ_+ uniformly in the space parameter value. The case $\rho(\phi_4(\kappa)) < \phi_+$ in Assumption 2 (f) ensures that the spectral radius of the kernel second derivative integral when evaluated at the maximum argument of $\left|\partial_\kappa^2 h^{(i,j)}(t, \kappa)^{(k,l)}\right|$ is strictly smaller than ϕ_+ uniformly in the space parameter value. Assumption 2 (f) yields Assumption 1 (g). It is required since Assumption 1 (g) considers the intensity derivatives $\partial_\theta \lambda_t(\theta)$ and Hessian matrix of the intensity, i.e. $\partial_\theta^2 \lambda_t(\theta)$.

In addition, the case $\rho(\phi_5) < +\infty$ in Assumption 2 (g) ensures that the spectral radius of the kernel derivative integral, when squared and evaluated at the maximum argument parameter of $\left|\partial_\kappa h^{(i,j)}(t, \kappa)^{(k)}\right|$, is finite. When $\rho(\phi_6) < +\infty$ in Assumption 2 (g), we have that the spectral radius of the kernel second derivative integral, when squared and evaluated at the maximum argument parameter of $\left|\partial_\kappa^2 h^{(i,j)}(t, \kappa)^{(k,l)}\right|$, is finite. Assumption 2 (g) implies Assumption 1 (h). It is necessary since Assumption 1 (h) considers the product of the intensity derivatives $\partial_\theta \lambda_t(\theta^*)$ and the intensity Hessian matrix $\partial_\theta^2 \lambda_t(\theta^*)$. Finally, Assumption 2 (h) is required for the non-degeneracy of the parametric inference procedure and gives Assumption 1 (e).

In the theorem that follows, we state the CLT of the parametric inference procedure for Hawkes processes. The kernel has a general form and is parametric. The inference procedure is based on MLE. We consider asymptotics when the final time diverges to infinity, i.e. $T \to +\infty$. This is the main result of this paper. This extends Clinet and Yoshida (2017) and Potiron and Volkov (2025), who obtain ergodicity by proving first that the Hawkes intensity process and its derivative is mixing by Markov arguments. In particular, we allow for kernels with power distribution, under some smoothness assumptions on the kernel shape.

Theorem 2. *We assume that Assumption 2 holds. As $T \to +\infty$, we have the CLT of the inference procedure for Hawkes processes where the kernel has a general form and is parametric, i.e.*

$$\sqrt{T}(\widehat{\theta}_T - \theta^*) \xrightarrow{\mathcal{D}} \Gamma^{-1/2}\xi. \tag{14}$$

4. Proofs of the CLT for point processes

In this section, we give the proofs of the CLT of the parametric inference procedure for point processes, i.e. Theorem 1. All the theoretical results refer to the convergence $T \to +\infty$.

In what follows, the constant C refers to a generic constant, which can differ from line to line. For a measure μ, we denote by $L^1(\mu)$ the space of functions that are integrable with respect to μ.

Since the functions that we will be using in our proofs will not necessarily be bounded, we extend from $C_b(E, \mathbb{R})$ to $C_\uparrow(E, \mathbb{R})$ the space of functions in which the ergodicity assumption holds. We also give a more explicit form to the functions $\pi(\psi)$. The following lemma is Proposition 3.8 (pp. 1806-1807) in Clinet and Yoshida (2017). The proof follows the arguments from the proof of Proposition 3.8 (pp. 1822-1824) in Clinet and Yoshida (2017).

Lemma 1. *We assume that Assumptions 1 (a), (b) and (c) hold. For any parameter $\theta \in \Theta$, we have*

(a) The ergodicity assumption 1 (c) still holds for any $\psi \in C_\uparrow(E, \mathbb{R})$. In particular, the function $\pi^{(i)}(\psi)$ can be extended to $C_\uparrow(E, \mathbb{R})$ for any $\theta \in \Theta$ and any $i = 1, \ldots, d$. Moreover, the convergence is uniform in $\theta \in \Theta$ for any $\psi \in C_\uparrow(E, \mathbb{R})$.

(b) For any $i = 1, \ldots, d$ and any $\theta \in \Theta$, there exists a probability measure $\pi_\theta^{(i)}$ on E such that, for any $\psi \in C_\uparrow(E, \mathbb{R})$, we have $\pi^{(i)}(\psi) = \int_E \psi(u, v, w) \pi_\theta^{(i)}(du, dv, dw)$. In particular, $C_\uparrow(E, \mathbb{R}) \subset L^1(\pi_\theta^{(i)})$.

Proof of Lemma 1. We can use the arguments from the proof of Proposition 3.8 (pp. 1822-1824) in Clinet and Yoshida (2017). □

We introduce now the following lemma, which ensures that the family of intensities does not explode on any compact space. Its proof is a direct consequence to Lemma 1.

Lemma 2. *We assume that Assumptions 1 (a), (b) and (c) hold. For any parameter $\theta \in \Theta$, the intensity process $\lambda_t(\theta)$ is a.s. locally integrable on the space $\mathbb{R}^+$.*

Proof of Lemma 2. This is a direct consequence to Lemma 1 with the function $\psi(u, v, w) = v$ for any $(u, v, w) \in E$. □

We define the normalized deviation between the log likelihood at the parameter value $\theta \in \Theta$ and the log likelihood at the true parameter value as

$$Y_T(\theta) = \frac{1}{T}(l_T(\theta) - l_T(\theta^*)). \tag{15}$$

We define the compensated point process as

$$M_t = N_t - \int_0^t \lambda_s(\theta^*) ds. \tag{16}$$

By definition of a compensator, we have that M_t is an $\mathcal{F}_t$-local martingale.

In the following lemma, we will prove the consistency of Y_T to Y uniformly in the parameter $\theta \in \Theta$. This weakens the assumptions used in Lemma 3.10 (p. 1807) from Clinet and Yoshida (2017). We consider a different approach in the proofs that circumvents the use of the Sobolev embedding theorem. This new approach is based on the application of Burkholder—Davis—Gundy inequalites.

Lemma 3. *We assume that Assumptions 1 (a), (b), (c) and (d) hold. We have the uniform consistency*

$$\sup_{\theta\in\Theta} \mid Y_T(\theta) - Y(\theta) \mid \xrightarrow{\mathbb{P}} 0.$$

Proof of Lemma 3. We can rewrite $Y_T(\theta)$ as

$$\begin{aligned}
Y_T(\theta) &= \frac{1}{T}(l_T(\theta) - l_T(\theta^*)) \\
&= \frac{1}{T}\Big(\sum_{i=1}^{d}\int_0^T \log(\lambda_t^{(i)}(\theta))dN_t^{(i)} - \sum_{i=1}^{d}\int_0^T \lambda_t^{(i)}(\theta)dt \\
&\quad - \sum_{i=1}^{d}\int_0^T \log(\lambda_t^{(i)}(\theta^*))dN_t^{(i)} + \sum_{i=1}^{d}\int_0^T \lambda_t^{(i)}(\theta^*)dt\Big) \\
&= \frac{1}{T}\sum_{i=1}^{d}\int_0^T \log\Big(\frac{\lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta^*)}\Big)dN_t^{(i)} - \frac{1}{T}\sum_{i=1}^{d}\int_0^T (\lambda_t^{(i)}(\theta) - \lambda_t^{(i)}(\theta^*))dt \\
&= \frac{1}{T}\sum_{i=1}^{d}\int_0^T \log\Big(\frac{\lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta^*)}\Big)dM_t^{(i)} \\
&\quad - \frac{1}{T}\sum_{i=1}^{d}\int_0^T \Big(\lambda_t^{(i)}(\theta) - \lambda_t^{(i)}(\theta^*) - \log\Big(\frac{\lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta^*)}\Big)\lambda_t^{(i)}(\theta^*)\Big)dt.
\end{aligned}$$

Here, we use Definition (15) in the first equality, Definition (4) in the second equality, algebraic manipulation in the third equality, Definition (16) and algebraic manipulation in the fourth equality. We define $I_T^{(i)}(\theta)$ as

$$I_T^{(i)}(\theta) = \frac{1}{T}\int_0^T \log\Big(\frac{\lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta^*)}\Big)dM_t^{(i)}.$$

We also define $II_T^{(i)}(\theta)$ as

$$II_T^{(i)}(\theta) = \frac{1}{T}\int_0^T \Big(\lambda_t^{(i)}(\theta) - \lambda_t^{(i)}(\theta^*) - \log\Big(\frac{\lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta^*)}\Big)\lambda_t^{(i)}(\theta^*)\Big)dt.$$

We first show that the martingale term disappears uniformly asymptotically in probability, i.e. that

$$\sup_{\theta\in\Theta}\Big|\sum_{i=1}^{d} I_T^{(i)}(\theta)\Big| \xrightarrow{\mathbb{P}} 0. \tag{17}$$

Since L^2 convergence implies convergence in probability, it is sufficient to show Expression (17) that

$$\mathbb{E}\Big[\Big(\sup_{\theta\in\Theta}\Big|\sum_{i=1}^{d} I_T^{(i)}(\theta)\Big|\Big)^2\Big] \to 0. \tag{18}$$

By the triangular inequality, we can deduce that

$$\mathbb{E}\Big[\Big(\sup_{\theta\in\Theta}\Big|\sum_{i=1}^{d} I_T^{(i)}(\theta)\Big|\Big)^2\Big] \le \mathbb{E}\Big[\Big(\sup_{\theta\in\Theta}\sum_{i=1}^{d}\Big|I_T^{(i)}(\theta)\Big|\Big)^2\Big]. \tag{19}$$

Then, the definition of $I_T^{(i)}(\theta)$ yields

$$\mathbb{E}\Big[\Big(\sup_{\theta\in\Theta}\sum_{i=1}^{d}\Big|I_T^{(i)}(\theta)\Big|\Big)^2\Big] = \mathbb{E}\Big[\Big(\sup_{\theta\in\Theta}\sum_{i=1}^{d}\Big|\frac{1}{T}\int_0^T \log\Big(\frac{\lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta^*)}\Big)dM_t^{(i)}\Big|\Big)^2\Big]. \tag{20}$$

By supremum properties, we get

$$\begin{aligned}\mathbb{E}\Big[\Big(\sup_{\theta\in\Theta}\sum_{i=1}^{d}\Big|\frac{1}{T}\int_0^T \log\Big(\frac{\lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta^*)}\Big)dM_t^{(i)}\Big|\Big)^2\Big] \\ \le \mathbb{E}\Big[\Big(\sum_{i=1}^{d}\sup_{\theta\in\Theta}\Big|\frac{1}{T}\int_0^T \log\Big(\frac{\lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta^*)}\Big)dM_t^{(i)}\Big|\Big)^2\Big].\end{aligned} \tag{21}$$

By the triangular inequality and the definition of martingale (16), we can deduce that

$$\begin{aligned}\mathbb{E}\Big[\Big(\sum_{i=1}^{d}\sup_{\theta\in\Theta}\Big|\frac{1}{T}\int_0^T \log\Big(\frac{\lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta^*)}\Big)dM_t^{(i)}\Big|\Big)^2\Big] \\ \le C\mathbb{E}\Big[\Big(\sum_{i=1}^{d}\sup_{\theta\in\Theta}\frac{1}{T}\int_0^T \Big|\log\Big(\frac{\lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta^*)}\Big)\Big|dM_t^{(i)}\Big)^2\Big].\end{aligned} \tag{22}$$

Again by supremum properties, we can deduce that

$$\begin{aligned}C\mathbb{E}\Big[\Big(\sum_{i=1}^{d}\sup_{\theta\in\Theta}\frac{1}{T}\int_0^T \Big|\log\Big(\frac{\lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta^*)}\Big)\Big|dM_t^{(i)}\Big)^2\Big] \\ \le C\mathbb{E}\Big[\Big(\sum_{i=1}^{d}\frac{1}{T}\int_0^T \sup_{\theta\in\Theta}\Big|\log\Big(\frac{\lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta^*)}\Big)\Big|dM_t^{(i)}\Big)^2\Big].\end{aligned} \tag{23}$$

By an algebraic manipulation, we can deduce that

$$\begin{aligned}C\mathbb{E}\Big[\Big(\sum_{i=1}^{d}\frac{1}{T}\int_0^T \sup_{\theta\in\Theta}\Big|\log\Big(\frac{\lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta^*)}\Big)\Big|dM_t^{(i)}\Big)^2\Big] \\ \le \sum_{i=1}^{d}\frac{C}{T^2}\mathbb{E}\Big[\Big(\int_0^T \sup_{\theta\in\Theta}\Big|\log\Big(\frac{\lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta^*)}\Big)\Big|dM_t^{(i)}\Big)^2\Big].\end{aligned} \tag{24}$$

We are going to apply Burkholder—Davis—Gundy inequalites (see Expression (2.1.32) in Jacod and Protter (2012) (p. 39)). In the notation of the book, we set $p = 2$ and introduce the stochastic process $\widetilde{M}_t$ of dimension d where for $i = 1, \ldots, d$ its i-th component is defined as

$$\widetilde{M}_t^{(i)} = \int_0^t \sup_{\theta\in\Theta} \Big| \log\Big(\frac{\lambda_u^{(i)}(\theta)}{\lambda_u^{(i)}(\theta^*)}\Big)\Big| dM_u^{(i)}.$$

First, we can show that $\widetilde{M}_t$ is an $\mathcal{F}_t$-local martingale by an application of Lemma 1 with the fact that M_t is an $\mathcal{F}_t$-local martingale. Then, we can apply Burkholder—Davis—Gundy inequalites and we obtain

$$\sum_{i=1}^d \frac{C}{T^2}\mathbb{E}\Big[\Big(\int_0^T \sup_{\theta\in\Theta}\Big|\log\Big(\frac{\lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta^*)}\Big)\Big| dM_t^{(i)}\Big)^2\Big] \tag{25}$$
$$\leq \sum_{i=1}^d \frac{C}{T^2}\mathbb{E}\Big[\int_0^T \sup_{\theta\in\Theta}\log\Big(\frac{\lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta^*)}\Big)^2 \lambda_t^{(i)}(\theta^*)^2 dt\Big].$$

By the inequality $\log(x) \leq 1 + x$, we can deduce that

$$\sum_{i=1}^d \frac{C}{T^2}\mathbb{E}\Big[\int_0^T \sup_{\theta\in\Theta}\log\Big(\frac{\lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta^*)}\Big)^2 \lambda_t^{(i)}(\theta^*)^2 dt\Big] \tag{26}$$
$$\leq \sum_{i=1}^d \frac{C}{T^2}\mathbb{E}\Big[\int_0^T \sup_{\theta\in\Theta}\Big(1+\frac{\lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta^*)}\Big)^2 \lambda_t^{(i)}(\theta^*)^2 dt\Big].$$

This can be reexpressed as

$$\sum_{i=1}^d \frac{C}{T^2}\mathbb{E}\Big[\int_0^T \sup_{\theta\in\Theta}\Big(1+\frac{\lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta^*)}\Big)^2 \lambda_t^{(i)}(\theta^*)^2 dt\Big] \tag{27}$$
$$= \sum_{i=1}^d \frac{C}{T^2}\mathbb{E}\Big[\int_0^T \sup_{\theta\in\Theta}\Big(\lambda_t^{(i)}(\theta^*)^2 + 2\lambda_t^{(i)}(\theta)\lambda_t^{(i)}(\theta^*) + \lambda_t^{(i)}(\theta)^2\Big) dt\Big].$$

By Tonelli theorem, this can be reexpressed as

$$\sum_{i=1}^d \frac{C}{T^2}\mathbb{E}\Big[\int_0^T \sup_{\theta\in\Theta}\Big(\lambda_t^{(i)}(\theta^*)^2 + 2\lambda_t^{(i)}(\theta)\lambda_t^{(i)}(\theta^*) + \lambda_t^{(i)}(\theta)^2\Big) dt\Big] \tag{28}$$
$$= \sum_{i=1}^d \frac{C}{T^2}\int_0^T \mathbb{E}\Big[\sup_{\theta\in\Theta}\Big(\lambda_t^{(i)}(\theta^*)^2 + 2\lambda_t^{(i)}(\theta)\lambda_t^{(i)}(\theta^*) + \lambda_t^{(i)}(\theta)^2\Big) dt\Big].$$

By Expressions (18) to (28) with Assumption 1 (d), we can prove Expression (17). To prove that $\mid \sum_{i=1}^d II_T^{(i)}(\theta) + Y(\theta) \mid \xrightarrow{\mathbb{P}} 0$, we can use Lemma 1. □

In the following lemma, we will prove the consistency of the parametric inference procedure based on MLE. This weakens the assumptions used in Theorem 3.9 (p. 1807) from Clinet and Yoshida (2017).

Lemma 4. *We assume that Assumptions 1 (a), (b), (c), (d) and (e) hold. We have the consistency of the parametric inference procedure based on MLE, i.e.*

$$\widehat{\theta}_T \xrightarrow{\mathbb{P}} \theta^*.$$

Proof of Lemma 4. By its definition (6), we can deduce that $Y(\theta) \leq 0$ for any parameter $\theta \in \Theta$ and $Y(\theta^*) = 0$. By Assumption 1 (e), we have that θ^* is a global maximum of Y. Finally, we can conclude by an application of Lemma 3 with Theorem 5.7 (p. 45) in Van der Vaart (2000) □

In the following lemma, we give a more explicit form to the partial derivatives and the Hessian matrix of the log likelihood. This extends Lemma A.1 (p. 1824) in Clinet and Yoshida (2017).

Lemma 5. *We assume that Assumptions 1 (a), (b), (c), (f) and (g) hold. For any parameter $\theta \in \Theta$ and any time $T > 0$, we have that $l_T(\theta)$ is a.s. finite and is differentiable a.s. with partial derivatives equal to*

$$\partial_\theta l_T(\theta) = \sum_{i=1}^{d} \int_0^T \frac{\partial_\theta \lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta)} \mathbf{1}_{\{\lambda_t^{(i)}(\theta)>0\}} dN_t^{(i)} - \sum_{i=1}^{d} \int_0^T \partial_\theta \lambda_t^{(i)}(\theta) dt. \tag{29}$$

Moreover, we have that $l_T(\theta)$ is differentiable twice a.s. and that its Hessian matrix satisfies

$$\begin{aligned} \partial_\theta^2 l_T(\theta) &= \sum_{i=1}^{d} \int_0^T \partial_\theta \Big(\frac{\partial_\theta \lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta)} \Big) \mathbf{1}_{\{\lambda_t^{(i)}(\theta)>0\}} dN_t^{(i)} \\ &\quad - \sum_{i=1}^{d} \int_0^T \partial_\theta^2 \lambda_t^{(i)}(\theta) dt. \end{aligned} \tag{30}$$

Proof. We define $I_T^{(i)}(\theta)$ for any parameter $\theta \in \Theta$ as

$$I_T^{(i)}(\theta) = \int_0^T \log(\lambda_t^{(i)}(\theta)) dN_t^{(i)}.$$

We also define $II_T^{(i)}(\theta)$ as

$$II_T^{(i)}(\theta) = \int_0^T \lambda_t^{(i)}(\theta) dt.$$

From the definition (4), we have the decomposition

$$l_T(\theta) = \sum_{i=1}^{d} \Big(I_T^{(i)}(\theta) - II_T^{(i)}(\theta) \Big).$$

First, we show that, for any parameter $\theta \in \Theta$, any time $T > 0$ and any index $i = 1, \dots, d$, we have that $I_T^{(i)}(\theta)$ is a.s. finite and a.s. differentiable with partial derivatives equal to

$$\partial_\theta I_T^{(i)}(\theta) \quad = \quad \int_0^T \partial_\theta \log(\lambda_t^{(i)}(\theta)) dN_t^{(i)}. \tag{31}$$

By the assumption on the times of the point process (2), $I_T^{(i)}(\theta)$ can be reexpressed as

$$I_T^{(i)}(\theta) = \sum_{k \in \mathbb{N}_* \text{ s.t. } 0 < T_k^{(i)} < T} \log\left(\lambda_{T_k^{(i)}}^{(i)}(\theta)\right). \tag{32}$$

By Lemma 2 and compacity of the segment $[0, T]$, we have that the number of terms in the sum is a.s. finite. We also have that each term in the sum is a.s. finite as the intensity process $\lambda_t^{(i)}(\theta)$ takes its values in $\mathbb{R}^+$ by definition. Then, we can deduce that $I_T^{(i)}(\theta)$ is a.s. finite. As the intensity process is differentiable a.e. a.s. by Assumption 1 (f) and by linearity of the derivative, we can deduce that $I_T^{(i)}(\theta)$ is a.s. differentiable and that a.s.

$$\partial_\theta I_T^{(i)}(\theta) = \sum_{k \in \mathbb{N}_* \text{ s.t. } 0 < T_k^{(i)} < T} \partial_\theta \log\left(\lambda_{T_k^{(i)}}^{(i)}(\theta)\right).$$

By Equation (32), this equality can be reexpressed as Equation (31). Finally, we get by rewriting the sum as an integral that

$$\partial_\theta I_T^{(i)}(\theta) \quad = \quad \int_0^T \frac{\partial_\theta \lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta)} \mathbf{1}_{\{\lambda_t^{(i)}(\theta) > 0\}} dN_t^{(i)}. \tag{33}$$

For any parameter $\theta \in \Theta$, any time $T > 0$ and any index $i = 1, \dots, d$, we have that $II_T^{(i)}(\theta)$ is a.s. finite and differentiable with partial derivatives equal to

$$\partial_\theta II_T^{(i)}(\theta) \quad = \quad \int_0^T \partial_\theta \lambda_t^{(i)}(\theta) dt. \tag{34}$$

By Lemma 2 and compacity of the segment $[0, T]$, $II_T^{(i)}(\theta)$ is a.s. finite. To show Equation (34), we show that the conditions for dominated convergence theorem are satisfied. We get by Assumption 1 (g) that the conditions are satisfied. Thus, we can deduce Equation (34). Finally, Equations (33) and (34) yield Equation (29). We can show Equation (30) with similar arguments. □

In the following lemma, we give a different form to the Hessian matrix of the log likelihood. This is a direct application of Lemma 5.

Lemma 6. *We assume that Assumptions 1 (a), (b), (c), (f) and (g) hold. For any parameter $\theta \in \Theta$ and any time $T > 0$, we have that the Hessian matrix of $l_T(\theta)$ can be rewritten as*

$$\begin{aligned}\partial_\theta^2 l_T(\theta) \quad = \quad & \sum_{i=1}^{d} \int_0^T \partial_\theta \Big(\frac{\partial_\theta \lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta)} \Big) \mathbf{1}_{\{\lambda_t^{(i)}(\theta)>0\}} dM_t^{(i)} \\ & - \sum_{i=1}^{d} \int_0^T (\partial_\theta \lambda_t^{(i)})^{\otimes 2}(\theta) \frac{\lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta)^2} \mathbf{1}_{\{\lambda_t^{(i)}(\theta)>0\}} dt \\ & + \sum_{i=1}^{d} \int_0^T (\partial_\theta^2 \lambda_t^{(i)})(\theta) \frac{\lambda_t^{(i)}(\theta) - \lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta)} \mathbf{1}_{\{\lambda_t^{(i)}(\theta)>0\}} dt. \qquad (35)\end{aligned}$$

Proof. The proof of this lemma is a direct application of Lemma 5. □

We provide the proof of Theorem 1 in what follows. This weakens the assumptions used in Theorem 3.11 (p. 1809) from Clinet and Yoshida (2017). We consider a different approach in the proofs that circumvents the use of the Sobolev embedding theorem. This new approach is based on the application of Burkholder—Davis—Gundy inequalites.

Proof of Theorem 1. By Assumption (f), we have that l_t is twice continuously differentiable in $\theta \in \Theta$ a.s. for any time $t \in \mathbb{R}^+$. Thus, we can apply the mean value theorem. We obtain that

$$\partial_\theta l_T(\widehat{\theta}_T) = \partial_\theta l_T(\theta^*) + \partial_\theta^2 l_T(\zeta)(\widehat{\theta}_T - \theta^*),$$

where ζ is between $\widehat{\theta}_T$ and θ^*. Since $\widehat{\theta}_T$ the maximizer of l_T by definition, we can deduce that $\partial_\theta l_T(\widehat{\theta}_T) = 0$. This yields that

$$0 = \partial_\theta l_T(\theta^*) + \partial_\theta^2 l_T(\zeta)(\widehat{\theta}_T - \theta^*).$$

If we multiply by $\frac{-\Gamma^{-1}}{\sqrt{T}}$ on both sides of the equation, we obtain that

$$0 = \frac{-\Gamma^{-1}}{\sqrt{T}} \partial_\theta l_T(\theta^*) + \frac{-\Gamma^{-1}}{\sqrt{T}} \partial_\theta^2 l_T(\zeta)(\widehat{\theta}_T - \theta^*).$$

This equation can be reexpressed as

$$0 = \frac{-\Gamma^{-1}}{\sqrt{T}} \partial_\theta l_T(\theta^*) + \frac{-\Gamma^{-1}}{T} \partial_\theta^2 l_T(\zeta) \sqrt{T}(\widehat{\theta}_T - \theta^*).$$

To prove the theorem, it remains to show that

$$\frac{-\Gamma^{-1}}{\sqrt{T}} \partial_\theta l_T(\theta^*) \quad \overset{\mathcal{D}}{\to} \quad \Gamma^{-1/2}\xi, \qquad (36)$$

$$\frac{-\Gamma^{-1}}{T} \partial_\theta^2 l_T(\zeta) \quad \overset{\mathbb{P}}{\to} \quad I_n. \qquad (37)$$

Here, I_n is the identity matrix of dimension $n \times n$. Then, the theorem follows using Slutsky's theorem.

We prove now Equation (36). Equation (29) from Lemma 5 can be reexpressed as

$$\partial_\theta l_T(\theta^*) = \sum_{i=1}^d \int_0^T \frac{\partial_\theta \lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)} \mathbf{1}_{\{\lambda_t^{(i)}>0\}} dM_t^{(i)}. \tag{38}$$

By Equation (38), we have that

$$\frac{-\Gamma^{-1}}{\sqrt{T}} \partial_\theta l_T(\theta^*) = \frac{-\Gamma^{-1}}{\sqrt{T}} \sum_{i=1}^d \int_0^T \frac{\partial_\theta \lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)} \mathbf{1}_{\{\lambda_t^{(i)}>0\}} dM_t^{(i)}.$$

For $u \in [0,1]$, we define S_u as

$$S_u = \frac{-\Gamma^{-1}}{\sqrt{T}} \sum_{i=1}^d \int_0^{uT} \frac{\partial_\theta \lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)} \mathbf{1}_{\{\lambda_t^{(i)}>0\}} dM_t^{(i)}. \tag{39}$$

We use Corollary VIII.3.24 (p. 476) in Jacod and Shiryaev (2003). We can calculate that

$$\begin{aligned}\langle S, S\rangle_u &= \frac{\Gamma^{-2}}{T} \sum_{i=1}^d \int_0^{uT} \partial_\theta \lambda_t^{(i)}(\theta^*)^2 \mathbf{1}_{\{\lambda_t^{(i)}>0\}} dt \\ &\xrightarrow{\mathbb{P}} u\Gamma^{-1}.\end{aligned}$$

Here, we define the quadratic variation of a stochastic process X at time t as $\langle X, X\rangle_t$. We also define ΔX_t as the jump of the process X at time t, i.e. $\Delta X_t = X_t - X_{t^-}$. We prove now that Lindeberg's condition is satisfied. We introduce

$$I = \mathbb{E}\Big[\sum_{s\leq u} |\Delta S_s|^2 \mathbf{1}_{|\Delta S_s|>a}\Big].$$

For any $a > 0$, we have

$$\begin{aligned} I &\leq \mathbb{E}\Big[\frac{1}{a}\sum_{s\leq u} |\Delta S_s|^3\Big] \\ &= \mathbb{E}\Big[\frac{1}{a}\sum_{s\leq u} \Big|\Delta\Big(\frac{-\Gamma^{-1}}{\sqrt{T}} \sum_{i=1}^d \int_0^{sT} \frac{\partial_\theta \lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)} \mathbf{1}_{\{\lambda_t^{(i)}>0\}} dM_t^{(i)}\Big)\Big|^3\Big] \\ &\leq C\mathbb{E}\Big[\frac{1}{a}\sum_{s\leq u} \Big|\Delta\Big(\frac{-\Gamma^{-1}}{\sqrt{T}} \sum_{i=1}^d \int_0^{sT} \frac{\partial_\theta \lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)} \mathbf{1}_{\{\lambda_t^{(i)}>0\}} dN_t^{(i)}\Big)\Big|^3\Big]. \end{aligned}$$

Here, we use Markov inequality in the first inequality, the equality is due to Definition (39), and the second inequality is explained by the fact that the term

$$\sum_{i=1}^d \int_0^{sT} \frac{\partial_\theta \lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)} \mathbf{1}_{\{\lambda_t^{(i)}>0\}} \lambda_t^{(i)} dt$$

does not jump and the triangular inequality. Then, we have

$$\begin{aligned} I &\leq C\mathbb{E}\Big[\frac{1}{a}\sum_{s\leq u}\Big|\Delta\Big(\frac{-\Gamma^{-1}}{\sqrt{T}}\sum_{i=1}^{d}\int_0^{sT}\frac{\partial_\theta\lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)}\mathbf{1}_{\{\lambda_t^{(i)}>0\}}dN_t^{(i)}\Big)\Big|^3\Big] \\ &\leq C\mathbb{E}\Big[\frac{1}{a}\Big(\frac{-\Gamma^{-1}}{\sqrt{T}}\sum_{i=1}^{d}\sum_{k\in\mathbb{N}_*\text{ s.t. }0<T_k^{(i)}<uT}\Big|\frac{\partial_\theta\lambda_{T_k^{(i)}}^{(i)}(\theta^*)}{\lambda_{T_k^{(i)}}^{(i)}(\theta^*)}\mathbf{1}_{\{\lambda_{T_k^{(i)}}^{(i)}>0\}}\Big|^3\Big)\Big] \\ &= C\mathbb{E}\Big[\frac{1}{a}\sum_{i=1}^{d}\int_0^{uT}\Big|\frac{-\Gamma^{-1}}{\sqrt{T}}\frac{\partial_\theta\lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)}\Big|^3\lambda_t^{(i)}(\theta^*)\mathbf{1}_{\{\lambda_t^{(i)}>0\}}dt\Big]+o_{\mathbb{P}}(1). \end{aligned}$$

Here, the second inequality is obtained by the assumption on the times of the point process (2) and the triangular inequality, and the equality is due to an extension of Theorem 2 from Bacry et al. (2013). Then, we can continue to bound the Linderberg's term by

$$\begin{aligned} I &\leq C\mathbb{E}\Big[\frac{1}{a}\sum_{i=1}^{d}\int_0^{uT}\Big|\frac{-\Gamma^{-1}}{\sqrt{T}}\frac{\partial_\theta\lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)}\Big|^3\lambda_t^{(i)}(\theta^*)\mathbf{1}_{\{\lambda_t^{(i)}>0\}}dt\Big]+o_{\mathbb{P}}(1) \\ &= C\mathbb{E}\Big[\frac{1}{a}\Big|\frac{-\Gamma^{-1}}{\sqrt{T}}\Big|^3\sum_{i=1}^{d}\int_0^{uT}\Big|\frac{\partial_\theta\lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)}\Big|^3\lambda_t^{(i)}(\theta^*)\mathbf{1}_{\{\lambda_t^{(i)}>0\}}dt\Big]+o_{\mathbb{P}}(1) \\ &= C\mathbb{E}\Big[\frac{1}{a}\Big|\frac{-\Gamma^{-1}}{\sqrt{T}}\Big|^3\sum_{i=1}^{d}\int_0^{uT}\frac{|\partial_\theta\lambda_t^{(i)}(\theta^*)|^3}{\lambda_t^{(i)}(\theta^*)^2}\mathbf{1}_{\{\lambda_t^{(i)}>0\}}dt\Big]+o_{\mathbb{P}}(1). \end{aligned}$$

Finally, we can bound the Linderberg's term by

$$\begin{aligned} I &\leq C\mathbb{E}\Big[\frac{1}{a}\Big|\frac{-\Gamma^{-1}}{\sqrt{T}}\Big|^3\sum_{i=1}^{d}\int_0^{uT}\frac{|\partial_\theta\lambda_t^{(i)}(\theta^*)|^3}{\lambda_t^{(i)}(\theta^*)^2}\mathbf{1}_{\{\lambda_t^{(i)}>0\}}dt\Big]+o_{\mathbb{P}}(1) \\ &= \frac{C}{a}\Big|\frac{-\Gamma^{-1}}{\sqrt{T}}\Big|^3\sum_{i=1}^{d}\int_0^{uT}\mathbb{E}\Big[\frac{|\partial_\theta\lambda_t^{(i)}(\theta^*)|^3}{\lambda_t^{(i)}(\theta^*)^2}\mathbf{1}_{\{\lambda_t^{(i)}>0\}}\Big]dt+o_{\mathbb{P}}(1) \\ &\leq \frac{CuT}{a}\Big|\frac{-\Gamma^{-1}}{\sqrt{T}}\Big|^3+o_{\mathbb{P}}(1) \\ &\to 0. \end{aligned}$$

Here, we use Tonelli theorem in the equality, and Assumption 1 (h) in the second inequality. We have thus shown that Lindeberg's condition holds, i.e. Equation (36) is satisfied.

We prove now Equation (37), i.e. that $\frac{-\Gamma^{-1}}{T}\partial_\theta^2 l_T(\zeta)\xrightarrow{\mathbb{P}} I_n$. Then, it is sufficient to prove that

$$|\Gamma+T^{-1}\partial_\theta^2 l_T(\zeta)|\xrightarrow{\mathbb{P}}0.$$

If we define V as a shrinking ball centered on θ^*, it is then sufficient to show that

$$\sup_{\theta\in V} |\Gamma + T^{-1}\partial_\theta^2 l_T(\theta)| \xrightarrow{\mathbb{P}} 0. \tag{40}$$

We define $I_T^{(i)}(\theta)$ as

$$I_T^{(i)}(\theta) = \int_0^T \partial_\theta\Big(\frac{\partial_\theta \lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta)}\Big)\mathbf{1}_{\{\lambda_t^{(i)}(\theta)>0\}} dM_t^{(i)}.$$

We also define $II_T^{(i)}(\theta)$ as

$$II_T^{(i)}(\theta) = \int_0^T (\partial_\theta \lambda_t^{(i)})^{\otimes 2}(\theta)\frac{\lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta)^2}\mathbf{1}_{\{\lambda_t^{(i)}(\theta)>0\}} dt.$$

Finally, we define $III_T^{(i)}(\theta)$ as

$$III_T^{(i)}(\theta) = \int_0^T (\partial_\theta^2 \lambda_t^{(i)})(\theta)\frac{\lambda_t^{(i)}(\theta) - \lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta)}\mathbf{1}_{\{\lambda_t^{(i)}(\theta)>0\}} dt.$$

By Equation (30) from Lemma 5, Expression (40) can be reexpressed as

$$\sup_{\theta\in V}\Big|\Gamma + T^{-1}\sum_{i=1}^d \Big(I_T^{(i)}(\theta) - II_T^{(i)}(\theta) - III_T^{(i)}(\theta)\Big)\Big| \xrightarrow{\mathbb{P}} 0. \tag{41}$$

By Assumption 1 (h), we can prove with the same arguments from the proof of Expression (17) that

$$\sup_{\theta\in V}\Big|T^{-1}\sum_{i=1}^d I_T^{(i)}(\theta)\Big| \xrightarrow{\mathbb{P}} 0. \tag{42}$$

By Assumption 1 (c), we obtain that

$$\sup_{\theta\in V}\Big|\Gamma - T^{-1}\sum_{i=1}^d II_T^{(i)}(\theta^*)\Big| \xrightarrow{\mathbb{P}} 0. \tag{43}$$

We can deduce by the triangular inequality and supremum properties that

$$\begin{aligned}\sup_{\theta\in V}\Big|T^{-1}\sum_{i=1}^d II_T^{(i)}(\theta) - T^{-1}\sum_{i=1}^d II_T^{(i)}(\theta^*)\Big| \qquad (44)\\ \leq \sum_{i=1}^d \sup_{\theta\in V}\Big|T^{-1}\big(II_T^{(i)}(\theta) - II_T^{(i)}(\theta^*)\big)\Big|.\end{aligned}$$

By the definition of $II_T^{(i)}(\theta)$, we have that

$$\begin{aligned}
& \sup_{\theta\in V}\Big|T^{-1}\big(II_T^{(i)}(\theta)-II_T^{(i)}(\theta^*)\big)\Big| \qquad (45)\\
= \quad & \sup_{\theta\in V}\Big|T^{-1}\int_0^T\Big(\frac{(\partial_\theta\lambda_t^{(i)})^{\otimes 2}(\theta)}{\lambda_t^{(i)}(\theta)^2}-\frac{(\partial_\theta\lambda_t^{(i)})^{\otimes 2}(\theta^*)}{\lambda_t^{(i)}(\theta^*)^2}\Big)\lambda_t^{(i)}(\theta^*)\mathbf{1}_{\{\lambda_t^{(i)}>0\}}dt\Big|.
\end{aligned}$$

By Assumption (f), we have that $F_t^{(i)}$ is continuously differentiable in the parameter $\theta\in\Theta$ a.s. for any $t\in\mathbb{R}^+$. Thus, we can apply the mean value theorem. We obtain that

$$F_t^{(i)}(\theta)-F_t^{(i)}(\theta^*)=\partial_\theta F_t^{(i)}(\widetilde{\theta})(\theta-\theta^*), \qquad (46)$$

where $\widetilde{\theta}$ is between θ and θ^*. As $\theta\in V$ and $\theta^*\in V$, we also have $\widetilde{\theta}\in V$. By Equations (45) and (46), we can deduce that

$$\begin{aligned}
& \sup_{\theta\in V}\Big|T^{-1}\big(II_T^{(i)}(\theta)-II_T^{(i)}(\theta^*)\big)\Big| \qquad (47)\\
\leq \quad & \sup_{\theta\in V}\Big|T^{-1}\int_0^T\partial_\theta F_t^{(i)}(\widetilde{\theta})(\theta-\theta^*)\lambda_t^{(i)}(\theta^*)\mathbf{1}_{\{\lambda_t^{(i)}(\theta)>0\}}dt\Big|.
\end{aligned}$$

Then, we obtain by the triangular inequality, supremum and norm properties that

$$\begin{aligned}
& \sup_{\theta\in V}\Big|T^{-1}\int_0^T\partial_\theta F_t^{(i)}(\widetilde{\theta})(\theta-\theta^*)\lambda_t^{(i)}(\theta^*)\mathbf{1}_{\{\lambda_t^{(i)}>0\}}dt\Big| \qquad (48)\\
\leq \quad & T^{-1}\int_0^T\sup_{\theta\in V}\Big|\partial_\theta F_t^{(i)}(\widetilde{\theta})\Big|\Big|\theta-\theta^*\Big|\lambda_t^{(i)}(\theta^*)\mathbf{1}_{\{\lambda_t^{(i)}>0\}}dt.
\end{aligned}$$

By Assumption (f), we have that $\partial_\theta F_t^{(i)}$ is continuous in θ^* a.s. for any $t\in\mathbb{R}^+$. As V is a shrinking ball centered at the parameter θ^*, we get that

$$\begin{aligned}
& T^{-1}\int_0^T\sup_{\theta\in V}\Big|\partial_\theta F_t^{(i)}(\widetilde{\theta})\Big|\Big|\theta-\theta^*\Big|\lambda_t^{(i)}(\theta^*)\mathbf{1}_{\{\lambda_t^{(i)}>0\}}dt \qquad (49)\\
= \quad & T^{-1}\int_0^T\sup_{\theta\in V}\Big|\partial_\theta F_t^{(i)}(\theta^*)\Big|\Big|\theta-\theta^*\Big|\lambda_t^{(i)}(\theta^*)\mathbf{1}_{\{\lambda_t^{(i)}>0\}}dt+o_{\mathbb{P}}(1).
\end{aligned}$$

We define $s(V)$ as the size of the shrinking ball V. Then, we can deduce that

$$\begin{aligned}
& T^{-1}\int_0^T\sup_{\theta\in V}\Big|\partial_\theta F_t^{(i)}(\theta^*)\Big|\Big|\theta-\theta^*\Big|\lambda_t^{(i)}(\theta^*)\mathbf{1}_{\{\lambda_t^{(i)}>0\}}dt \qquad (50)\\
\leq \quad & T^{-1}s(V)\int_0^T\Big|\partial_\theta F_t^{(i)}(\theta^*)\Big|\lambda_t^{(i)}(\theta^*)\mathbf{1}_{\{\lambda_t^{(i)}>0\}}dt.
\end{aligned}$$

By Assumption 1 (h), we get that

$$T^{-1}s(V)\int_0^T\Big|\partial_\theta F_t^{(i)}(\theta^*)\Big|\lambda_t^{(i)}(\theta^*)\mathbf{1}_{\{\lambda_t^{(i)}>0\}}dt\xrightarrow{\mathbb{P}}0. \qquad (51)$$

By Expressions (45) to (51), we can deduce that

$$\sup_{\theta \in V} \left| T^{-1} \big(II_T^{(i)}(\theta) - II_T^{(i)}(\theta^*) \big) \right| \xrightarrow{\mathbb{P}} 0. \tag{52}$$

By Assumption 1 (h), we can prove with the same arguments from the proof of Expression (52) that

$$\sup_{\theta \in V} \left| T^{-1} \sum_{i=1}^{d} III_T^{(i)}(\theta) \right| \xrightarrow{\mathbb{P}} 0. \tag{53}$$

Finally, we can deduce Equation (37) by the use of Expressions (40), (41), (42), (43), (52) and (53). □

5. Proofs of the CLT for Hawkes processes with a general kernel

In this section, we give the proofs of the CLT for Hawkes processes where the kernel is parametric and general, i.e. Theorem 2.

We first show the following lemma, which corresponds to Assumption 1 (a).

Lemma 7. *We assume that Assumptions 2 (a) and (b) hold. Then, the family of intensities* $\lambda : \Omega \times \mathbb{R}^+ \times \Theta \to \mathbb{R}_+^d$ *defined in Equation (8) is* $\mathcal{F} \otimes \mathcal{B}(\mathbb{R}^+) \otimes \mathcal{B}(\Theta)$ *measurable.*

Proof of Lemma 7. First, we get by Definition (8), Assumptions 2 (a) and (b) that the intensity process is nonnegative, namely $\lambda_t \geq 0$ for any time $t \in \mathbb{R}^+$ and any $\omega \in \Omega$. Then, we can deduce that the intensity process λ_t is $\mathcal{F} \otimes \mathcal{B}(\mathbb{R}^+) \otimes \mathcal{B}(\Theta)$ measurable by Definition (8). □

We now prove the following lemma, which corresponds to Assumption 1 (b).

Lemma 8. *We assume that Assumption 2 (c) holds. Then, the parameter space* $\Theta \subset \mathbb{R}^n$ *is such that its closure* $\overline{\Theta}$ *is a compact space.*

Proof of Lemma 8. The statement of the lemma corresponds exactly to Assumption 2 (c) □

The next lemma is Lemma A.2 (p. 1825) from Clinet and Yoshida (2017).

Lemma 9. *We assume that Assumptions 2 (a), (b) and (c) hold. Let an integer* $p \in \mathbb{N}_*$ *and* f_t *be a stochastic process such that* $f_t^{2^p}$ *is a.s. locally integrable on the space* $\mathbb{R}^+$*. Then, we have for any index* $i = 1, \ldots, d$ *that*

$$\mathbb{E}\Big[\Big| \int_0^T f_t dN_t^{(i)} \Big|^{2^p} \Big] \leq \mathbb{E}\Big[\int_0^T f_t^{2^p} \lambda_t^{(i)}(\theta^*) dN_t^{(i)} \Big] + \mathbb{E}\Big[\Big| \int_0^T f_t^2 \lambda_t^{(i)}(\theta^*) dN_t^{(i)} \Big|^{2^{p-1}} \Big].$$

Proof of Lemma 9. This is a direct application of Lemma A.2 (p. 1825) from Clinet and Yoshida (2017). □

We now show the following lemma, which corresponds to Assumption 1 (d). This complements Lemma C4 in Supplement C of Potiron and Volkov (2025).

Lemma 10. *We assume that Assumptions 2 (a), (b), (c) and (d) hold. Then, we have*

$$\sup_{t\in\mathbb{R}^+}\Big|\Big|\sup_{\theta\in\Theta}|\lambda_t(\theta)|\Big|\Big|_2<+\infty.$$

Proof of Lemma 10. We first prove that

$$\sup_{t\in\mathbb{R}^+}\Big|\Big|\sup_{\theta\in\Theta}|\lambda_t(\theta)|\Big|\Big|_1<+\infty. \tag{54}$$

We have

$$\begin{aligned}\Big|\Big|\sup_{\theta\in\Theta}|\lambda_t(\theta)|\Big|\Big|_1 &= \mathbb{E}\Big[\sup_{\theta\in\Theta}|\lambda_t(\theta)|\Big]\\ &= \mathbb{E}\Big[\sup_{\theta\in\Theta}\Big\{\sum_{i=1}^d\lambda_t^{(i)}(\theta)\Big\}\Big]\\ &= \mathbb{E}\Big[\sup_{\theta\in\Theta}\Big\{\sum_{i=1}^d\nu^{(i)}+\sum_{j=1}^d\int_0^t h^{(i,j)}(t-s,\kappa)\,dN_s^{(j)}\Big\}\Big].\end{aligned}$$

Here, we use the definition of the norm $||\ ||_1$ in the first equality, the definition of $|\ |$ in the second equality, and Definition (8) in the third equality. Then, we have

$$\begin{aligned}\Big|\Big|\sup_{\theta\in\Theta}|\lambda_t(\theta)|\Big|\Big|_1 &= \mathbb{E}\Big[\sup_{\theta\in\Theta}\Big\{\sum_{i=1}^d\nu^{(i)}+\sum_{j=1}^d\int_0^t h^{(i,j)}(t-s,\kappa)dN_s^{(j)}\Big\}\Big]\\ &\le \mathbb{E}\Big[C+\sup_{\theta\in\Theta}\Big\{\sum_{i=1}^d\sum_{j=1}^d\int_0^t h^{(i,j)}(t-s,\kappa)dN_s^{(j)}\Big\}\Big]\\ &= C+\mathbb{E}\Big[\sup_{\theta\in\Theta}\Big\{\sum_{i=1}^d\sum_{j=1}^d\int_0^t h^{(i,j)}(t-s,\kappa)dN_s^{(j)}\Big\}\Big].\end{aligned}$$

Here, we use Assumption 2 (c) in the inequality, and expectation properties in the second equality. Then, we obtain

$$\begin{aligned}\Big|\Big|\sup_{\theta\in\Theta}|\lambda_t(\theta)|\Big|\Big|_1 &\le C+\mathbb{E}\Big[\sup_{\theta\in\Theta}\Big\{\sum_{i=1}^d\sum_{j=1}^d\int_0^t h^{(i,j)}(t-s,\kappa)dN_s^{(j)}\Big\}\Big]\\ &= C+\mathbb{E}\Big[\sup_{\kappa\in\Theta_\kappa}\Big\{\sum_{i=1}^d\sum_{j=1}^d\int_0^t h^{(i,j)}(t-s,\kappa)dN_s^{(j)}\Big\}\Big]\\ &\le C+\mathbb{E}\Big[\sum_{i=1}^d\sum_{j=1}^d\int_0^t h^{(i,j)}(t-s,\kappa_s^+)dN_s^{(j)}\Big]\end{aligned}$$

Here, we use the fact that the kernel depends only on the parameter κ in the equality, and Definition (10) in the second inequality. Then, we obtain

$$\begin{aligned}\Big|\Big|\sup_{\theta\in\Theta}|\lambda_t(\theta)|\Big|\Big|_1 &\leq C+\mathbb{E}\Big[\sum_{i=1}^{d}\sum_{j=1}^{d}\int_0^t h^{(i,j)}(t-s,\kappa_s^+)dN_s^{(j)}\Big]\\ &= C+\sum_{i=1}^{d}\sum_{j=1}^{d}\int_0^t h^{(i,j)}(t-s,\kappa_s^+)\Big|\Big|\lambda_t^{(j)}\Big|\Big|_1 ds\\ &\leq C+\sum_{i=1}^{d}\sum_{j=1}^{d}\sup_{t\in\mathbb{R}^+}\Big|\Big|\lambda_t^{(j)}\Big|\Big|_1\int_0^t h^{(i,j)}(t-s,\kappa_s^+)ds.\end{aligned}$$

Here, we use martingale properties in the equality and supremum properties in the second inequality. By Assumption 2 (d), there exists a real positive number h^+ which is smaller than unity, i.e. $0 < h^+ < 1$. This real positive number h^+ satisfies for any $i = 1, \ldots, d$ and any $j = 1, \ldots, d$ that

$$\int_0^t h^{(i,j)}(t-s,\kappa_s^+)ds \leq h^+. \tag{55}$$

Then, we have

$$\begin{aligned}\Big|\Big|\sup_{\theta\in\Theta}|\lambda_t(\theta)|\Big|\Big|_1 &\leq C+\sum_{i=1}^{d}\sum_{j=1}^{d}\sup_{t\in\mathbb{R}^+}\Big|\Big|\lambda_t^{(j)}\Big|\Big|_1\int_0^t h^{(i,j)}(t-s,\kappa_s^+)ds\\ &\leq C+\sum_{i=1}^{d}\sum_{j=1}^{d}\sup_{t\in\mathbb{R}^+}\Big|\Big|\lambda_t^{(j)}\Big|\Big|_1 h^+\\ &\leq C+\sup_{t\in\mathbb{R}^+}\Big|\Big|\sup_{\theta\in\Theta}|\lambda_t(\theta)|\Big|\Big|_1 h^+.\end{aligned}$$

Here, we use Expression (55) in the second inequality, the definition of $|\ |$ and supremum properties in the third inequality. By taking the supremum over the time $t \in \mathbb{R}^+$ on the left side of the expression, we can deduce that

$$\sup_{t\in\mathbb{R}^+}\Big|\Big|\sup_{\theta\in\Theta}|\lambda_t(\theta)|\Big|\Big|_1 \leq C+h^+\sup_{t\in\mathbb{R}^+}\Big|\Big|\sup_{\theta\in\Theta}|\lambda_t(\theta)|\Big|\Big|_1. \tag{56}$$

Since $0 < h^+ < 1$, Expression (56) implies

$$\sup_{t\in\mathbb{R}^+}\Big|\Big|\sup_{\theta\in\Theta}|\lambda_t(\theta)|\Big|\Big|_1 \leq \frac{C}{1-h^+}. \tag{57}$$

Thus, we have shown Expression (54). Finally, the lemma can be shown by an application of Lemma 9 with Assumption 2 (d). □

We now show the following lemma, which corresponds to Assumption 1 (f).

Lemma 11. *We assume that Assumptions 2 (a), (b), (c), (d) and (e) hold. For any $s \in \mathbb{R}^+$ a.e., we have a.s. that the intensity $\theta \to \lambda_s(\theta)$ is twice continuously differentiable from the parameter space Θ to the space $\mathbb{R}^d_+$. Moreover, there exists a continuous extension to $\overline{\Theta}$.*

Proof of Lemma 11. By Definition (8), we have

$$\lambda_t(\theta) = \nu + \int_0^t h(t-s,\kappa)\,dN_s.$$

Since the baseline parameter ν is twice continuously differentiable from the parameter space Θ to the space $\mathbb{R}_+^d$ and there exists a continuous extension to $\overline{\Theta}$, it remains to show the lemma with

$$\lambda_{t,h}(\theta) = \int_0^t h(t-s,\kappa)\,dN_s.$$

For any index $i = 1,\ldots,d$, the intensity can be rewritten as

$$\lambda_{t,h}^{(i)}(\theta) = \sum_{j=1}^{d} \int_0^t h^{(i,j)}(t-s,\kappa)\,dN_s^{(j)}.$$

By the assumption on the times of the point process (2), $\lambda_{t,h}^{(i)}(\theta)$ can be reexpressed as

$$\lambda_{t,h}^{(i)}(\theta) = \sum_{j=1}^{d} \sum_{k\in\mathbb{N}_* \text{ s.t. } 0<T_k^{(j)}<t} h^{(i,j)}\big(t-T_k^{(j)},\kappa\big).$$

By Lemma 2 and compacity of the segment $[0,T]$, we have that the number of terms in the sum and each term are a.s. finite. Then, we can deduce that $\lambda_{t,h}^{(i)}(\theta)$ is a.s. finite. As the kernel is differentiable a.e. by Assumption 2 (e), we can deduce that $\lambda_{t,h}^{(i)}(\theta)$ is a.s. differentiable and

$$\partial_\theta \lambda_{t,h}^{(i)}(\theta) = \sum_{j=1}^{d} \sum_{k\in\mathbb{N}_* \text{ s.t. } 0<T_k^{(j)}<t} \partial_\theta h^{(i,j)}\big(t-T_k^{(j)},\kappa\big).$$

As the kernel is differentiable twice a.e. by Assumption 2 (e), we can deduce that $\lambda_{t,h}^{(i)}(\theta)$ is a.s. differentiable and

$$\partial_\theta^2 \lambda_{t,h}^{(i)}(\theta) = \sum_{j=1}^{d} \sum_{k\in\mathbb{N}_* \text{ s.t. } 0<T_k^{(j)}<t} \partial_\theta^2 h^{(i,j)}(t-T_k^{(j)},\kappa).$$

Thus, we have a.s. that the intensity process $\theta \to \lambda_s(\theta)$ for any time $s \in \mathbb{R}^+$ a.e. is twice continuously differentiable from the parameter space Θ to the space $\mathbb{R}_+^d$, and there exists a continuous extension to $\overline{\Theta}$. □

We now show the following lemma, which corresponds to Assumption 1 (g).

Lemma 12. *We assume that Assumptions 2 (a), (b), (c), (d), (e) and (f) hold. For any $\theta \in \Theta$ and $T > 0$, we have $\mathbb{P}(\int_0^T |\partial_\theta \lambda_t(\theta)| dt < \infty) = 1$ and $\mathbb{P}(\int_0^T |\partial_\theta^2 \lambda_t(\theta)| dt < \infty) = 1$.*

Proof of Lemma 12. By Definition (8), we have

$$\lambda_t(\theta) = \nu + \int_0^t h(t-s,\kappa)\,dN_s.$$

Since ν satisfies $\mathbb{P}(\int_0^T |\partial_\theta \nu| dt < \infty) = 1$ and $\mathbb{P}(\int_0^T |\partial_\theta^2 \nu| dt < \infty) = 1$, it remains to show the lemma with

$$\lambda_{t,h}(\theta) = \int_0^t h(t-s,\kappa)\,dN_s.$$

For $i = 1,\dots,d$, $\lambda_{t,h}^{(i)}(\theta)$ can be rewritten as

$$\lambda_{t,h}^{(i)}(\theta) = \sum_{j=1}^d \int_0^t h^{(i,j)}(t-s,\kappa)\,dN_s^{(j)}.$$

By the assumption on the times of the point process (2), $\lambda_{t,h}^{(i)}(\theta)$ can be reexpressed as

$$\lambda_{t,h}^{(i)}(\theta) = \sum_{j=1}^d \sum_{k\in\mathbb{N}_* \text{ s.t. } 0<T_k^{(j)}<t} h^{(i,j)}\big(t-T_k^{(j)},\kappa\big).$$

By Lemma 2 and compacity of the segment $[0,T]$, we have that the number of terms in the sum and each term are a.s. finite. Then, we can deduce that $\lambda_{t,h}^{(i)}(\theta)$ is a.s. finite. As the kernel is differentiable a.e. by Assumption 2 (e), we can deduce that $\lambda_{t,h}^{(i)}(\theta)$ is a.s. differentiable and

$$\partial_\theta \lambda_{t,h}^{(i)}(\theta) = \sum_{j=1}^d \sum_{k\in\mathbb{N}_* \text{ s.t. } 0<T_k^{(j)}<t} \partial_\theta h^{(i,j)}\big(t-T_k^{(j)},\kappa\big).$$

By the assumption on the times of the point process (2), $\partial_\theta \lambda_{t,h}^{(i)}(\theta)$ can be reexpressed as

$$\partial_\theta \lambda_{t,h}^{(i)}(\theta) = \sum_{j=1}^d \int_0^t \partial_\theta h^{(i,j)}(t-s,\kappa)\,dN_s^{(j)}.$$

Then, we obtain $\mathbb{P}(\int_0^T |\partial_\theta \lambda_t(\theta)| dt < \infty) = 1$ by Assumption 2 (f). As the kernel is differentiable twice a.e. by Assumption 2 (e), we can deduce that $\lambda_{t,h}^{(i)}(\theta)$ is a.s. differentiable twice and

$$\partial_\theta^2 \lambda_{t,h}^{(i)}(\theta) = \sum_{j=1}^d \sum_{k\in\mathbb{N}_* \text{ s.t. } 0<T_k^{(j)}<t} \partial_\theta^2 h^{(i,j)}(t-T_k^{(j)},\kappa).$$

By the assumption on the times of the point process (2), $\partial_\theta^2 \lambda_{t,h}^{(i)}(\theta)$ can be reexpressed as

$$\partial_\theta^2 \lambda_{t,h}^{(i)}(\theta) = \sum_{j=1}^{d} \int_0^t \partial_\theta^2 h^{(i,j)}(t-s,\kappa)\, dN_s^{(j)}.$$

Finally, we obtain $\mathbb{P}(\int_0^T |\partial_\theta^2 \lambda_t(\theta)| dt < \infty) = 1$ by Assumption 2 (f). □

We now show the following lemma, which corresponds to Assumption 1 (h).

Lemma 13. *We assume that Assumptions 2 (a), (b), (c) and (g) hold. For any $i = 1, \dots, d$, we have*

$$\begin{aligned}
\sup_{t\in\mathbb{R}^+} \Big|\Big| \Big|\partial_\theta\Big(\frac{\partial_\theta \lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)}\Big)\Big| \lambda_t^{(i)}(\theta^*) \mathbf{1}_{\{\lambda_t^{(i)}>0\}} \Big|\Big|_2 &< +\infty,\\
\sup_{t\in\mathbb{R}^+} \Big|\Big| \Big|\partial_\theta F_t^{(i)}(\theta^*)\Big| \lambda_t^{(i)}(\theta^*) \mathbf{1}_{\{\lambda_t^{(i)}>0\}} \Big|\Big|_1 &< +\infty,\\
\sup_{t\in\mathbb{R}^+} \Big|\Big| \Big|(\partial_\theta^2 \lambda_t^{(i)})(\theta^*)\Big| \Big|\frac{\partial_\theta \lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)}\Big| \mathbf{1}_{\{\lambda_t^{(i)}>0\}} \Big|\Big|_1 &< +\infty.
\end{aligned}$$

Proof of Lemma 13. We define I as

$$I = \sup_{t\in\mathbb{R}^+} \Big|\Big| \Big|\partial_\theta\Big(\frac{\partial_\theta \lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)}\Big)\Big| \lambda_t^{(i)}(\theta^*) \mathbf{1}_{\{\lambda_t^{(i)}>0\}} \Big|\Big|_2.$$

By Assumptions 2 (a) and (b), we can deduce that

$$I = \sup_{t\in\mathbb{R}^+} \Big|\Big| \Big|\partial_\theta\Big(\frac{\partial_\theta \lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)}\Big)\Big| \lambda_t^{(i)}(\theta^*) \Big|\Big|_2.$$

By derivative formula, we obtain

$$I = \sup_{t\in\mathbb{R}^+} \Big|\Big| \Big|\frac{\partial_\theta^2 \lambda_t^{(i)}(\theta^*) - (\partial_\theta \lambda_t^{(i)})^{\otimes 2}(\theta^*)}{\lambda_t^{(i)}(\theta^*)^2}\Big| \lambda_t^{(i)}(\theta^*) \Big|\Big|_2.$$

This can be reexpressed as

$$I = \sup_{t\in\mathbb{R}^+} \Big|\Big| \Big|\frac{\partial_\theta^2 \lambda_t^{(i)}(\theta^*) - (\partial_\theta \lambda_t^{(i)})^{\otimes 2}(\theta^*)}{\lambda_t^{(i)}(\theta^*)}\Big| \Big|\Big|_2.$$

By Assumption 2 (a), we can deduce that

$$I < \frac{1}{\nu_-} \sup_{t\in\mathbb{R}^+} \Big|\Big| \Big|\partial_\theta^2 \lambda_t^{(i)}(\theta^*) - (\partial_\theta \lambda_t^{(i)})^{\otimes 2}(\theta^*)\Big| \Big|\Big|_2.$$

By Assumption 2 (g) and Equation (8), we obtain

$$I < +\infty. \tag{58}$$

We define II as

$$II = \sup_{t \in \mathbb{R}^+} \Big|\Big| \ \Big|\partial_\theta F_t^{(i)}(\theta^*)\Big| \Big|\lambda_t^{(i)}(\theta^*) \mathbf{1}_{\{\lambda_t^{(i)} > 0\}} \Big|\Big|_1.$$

By definition of $F_t^{(i)}(\theta^*)$, we can deduce that

$$II = \sup_{t \in \mathbb{R}^+} \Big|\Big| \ \Big|\partial_\theta \Big(\frac{(\partial_\theta \lambda_t^{(i)})^{\otimes 2}(\theta^*)}{\lambda_t^{(i)}(\theta^*)^2} \Big) \Big| \lambda_t^{(i)}(\theta^*) \mathbf{1}_{\{\lambda_t^{(i)} > 0\}} \Big|\Big|_1.$$

By Assumptions 2 (a) and (b), we can deduce that

$$II = \sup_{t \in \mathbb{R}^+} \Big|\Big| \ \Big|\partial_\theta \Big(\frac{(\partial_\theta \lambda_t^{(i)})^{\otimes 2}(\theta^*)}{\lambda_t^{(i)}(\theta^*)^2} \Big) \Big| \lambda_t^{(i)}(\theta^*) \Big|\Big|_1.$$

By extending the arguments from the proof of the case $I < +\infty$ with Assumptions 2 (a), (g) and Equation (8), we obtain

$$II < +\infty. \tag{59}$$

We define III as

$$III = \sup_{t \in \mathbb{R}^+} \Big|\Big| \ \Big|(\partial_\theta^2 \lambda_t^{(i)})(\theta^*)\Big| \ \Big| \frac{\partial_\theta \lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)} \Big| \mathbf{1}_{\{\lambda_t^{(i)} > 0\}} \ \Big|\Big|_1.$$

By Assumptions 2 (a) and (b), we can deduce that

$$III = \sup_{t \in \mathbb{R}^+} \Big|\Big| \ \Big|(\partial_\theta^2 \lambda_t^{(i)})(\theta^*)\Big| \ \Big| \frac{\partial_\theta \lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)} \Big| \ \Big|\Big|_1.$$

By Assumption 2 (a), we can deduce that

$$III < \frac{1}{\nu_-} \sup_{t \in \mathbb{R}^+} \Big|\Big| \ \Big|(\partial_\theta^2 \lambda_t^{(i)})(\theta^*)\Big| \ \Big|\partial_\theta \lambda_t^{(i)}(\theta^*)\Big| \ \Big|\Big|_1.$$

By extending the arguments from the proof of the case $I < +\infty$ with Assumption 2 (g) and Equation (8), we obtain

$$III < +\infty. \tag{60}$$

We can prove the lemma with Expressions (58), (59) and (60). □

The following definition introduces the notion of mixing. This corresponds to the definition from Section 3.4 in Clinet and Yoshida (2017). See also Definition C2 in Supplement C of Potiron and Volkov (2025).

Definition 3. We say that X is mixing if for any ϕ, $\psi \in C_b(E, \mathbb{R})$ and $i = 1, \dots, d$, we have as $T \to \infty$ that

$$\mu_T^{(i)} = \sup_{s \in \mathbb{R}^+} \Big| \operatorname{Cov}\big[\phi(X_s^{(i)}), \psi(X_{s+T}^{(i)})\big] \Big| \to 0.$$

The following lemma states that X_t is mixing in the sense of Definition 3. This extends Lemma A.6 (p. 1834) in Clinet and Yoshida (2017) and Proposition C1 (i) in Supplement C of Potiron and Volkov (2025).

Lemma 14. *We assume that Assumptions 2 (a), (b), (c), (d), (e), (f) and (g) hold. For any $\theta \in \Theta$, X_t is mixing in the sense of Definition 3.*

Proof of Lemma 14. We first define the truncation of $X_T^{(i)}$ at time $t \leq T$ as

$$\widetilde{X}_{t,T}^{(i)} = \Big(\lambda_t^{(i)}(\theta^*), \sum_{j=1}^{d} \int_t^T h^{(i,j)}(T-u,\theta) dN_u^{(i)}, \sum_{j=1}^{d} \int_t^T \partial_\theta \big(h^{(i,j)}(T-u,\theta)\big) dN_u^{(i)}\Big).$$

Then, we can reexpress $\mu_T^{(i)}$ as

$$\begin{aligned}\mu_T^{(i)} &= \sup_{s \in \mathbb{R}^+} \big| \operatorname{Cov}\big[\phi(X_s^{(i)}), \psi(X_{s+T}^{(i)})\big]\big| \\ &= \sup_{s \in \mathbb{R}^+} \big| \operatorname{Cov}\big[\phi(X_s^{(i)}), \psi(X_{s+T}^{(i)}) - \psi(\widetilde{X}_{s+\sqrt{T},s+T}^{(i)}) + \psi(\widetilde{X}_{s+\sqrt{T},s+T}^{(i)})\big]\big|.\end{aligned}$$

Here, we use Definition 3 in the first equality. Using the triangular inequality, covariance and supremum properties, we can bound $\mu_T^{(i)}$ as

$$\begin{aligned}\mu_T^{(i)} \leq\ & \sup_{s \in \mathbb{R}^+} \big| \operatorname{Cov}\big[\phi(X_s^{(i)}), \psi(X_{s+T}^{(i)}) - \psi(\widetilde{X}_{s+\sqrt{T},s+T}^{(i)})\big]\big| \\ & + \sup_{s \in \mathbb{R}^+} \big| \operatorname{Cov}\big[\phi(X_s^{(i)}), \psi(\widetilde{X}_{s+\sqrt{T},s+T}^{(i)})\big]\big|.\end{aligned} \tag{61}$$

We define $I_T^{(i)}$ as

$$I_T^{(i)} = \sup_{s \in \mathbb{R}^+} \big| \operatorname{Cov}\big[\phi(X_s^{(i)}), \psi(X_{s+T}^{(i)}) - \psi(\widetilde{X}_{s+\sqrt{T},s+T}^{(i)})\big]\big|.$$

We also define $II_T^{(i)}$ as

$$II_T^{(i)} = \sup_{s \in \mathbb{R}^+} \big| \operatorname{Cov}\big[\phi(X_s^{(i)}), \psi(\widetilde{X}_{s+\sqrt{T},s+T}^{(i)})\big]\big|.$$

By the definition of $I_T^{(i)}$, Cauchy–Schwarz inequality and supremum properties, we can deduce that

$$I_T^{(i)} \leq \sup_{s \in \mathbb{R}^+} \operatorname{Var}\big[\phi(X_s^{(i)})\big] \sup_{s \in \mathbb{R}^+} \operatorname{Var}\big[\psi(X_{s+T}^{(i)}) - \psi(\widetilde{X}_{s+\sqrt{T},s+T}^{(i)})\big]. \tag{62}$$

By Lemmas 10 and 13, we get

$$\sup_{s \in \mathbb{R}^+} \operatorname{Var}\big[\phi(X_s^{(i)})\big] \leq C. \tag{63}$$

Since $\sqrt{T} \to \infty$, by Assumption 2 (d), we obtain

$$\sup_{s \in \mathbb{R}^+} \operatorname{Var}\big[\psi(X_{s+T}^{(i)}) - \psi(\widetilde{X}_{s+\sqrt{T},s+T}^{(i)})\big] \to 0. \tag{64}$$

By Expressions (62), (63) and (64), we can deduce that

$$I_T^{(i)} \to 0. \tag{65}$$

As $T \to \infty$, by Assumption 2 (d), we obtain

$$II_T^{(i)} \to 0. \tag{66}$$

By Expressions (61), (65) and (66), we can deduce that

$$\mu_T^{(i)} \to 0.$$

□

The following lemma states that X_t is stable. This extends Lemma A.6 (p. 1834) in Clinet and Yoshida (2017) and Proposition C1 (ii) in Potiron and Volkov (2025).

Lemma 15. *We assume that Assumptions 2 (a), (b), (c), (d), (e), (f) and (g) hold. For any $\theta \in \Theta$, X_t is stable, i.e. for any $i = 1, \dots, d$ there exists an $\mathbb{R}_+^*$-valued random variable $\lambda_l^{(i)}(\theta)$ such that we have*

$$X_T^{(i)} \xrightarrow{\mathcal{D}} \big(\lambda_l^{(i)}(\theta^*), \lambda_l^{(i)}(\theta), \partial_\theta \lambda_l^{(i)}(\theta)\big).$$

Proof of Lemma 15. The proof is obtained by an application of Theorem 1 in Brémaud and Massoulié (1996) and Lemma 4 in Brémaud and Massoulié (1996) with Assumption 2 (d). □

The following lemma states that X_t is ergodic in the sense of Definition 1. Moreover, it delivers a more explicit expression of the limit function $\pi(\psi)$. This extends Lemma 3.16 (p. 1815) in Clinet and Yoshida (2017) and Proposition C1 (iii) in Potiron and Volkov (2025).

Lemma 16. *We assume that Assumptions 2 (a), (b), (c), (d), (e), (f) and (g) hold. For any $\theta \in \Theta$, X_t is ergodic in the sense of Definition 1. Moreover, for any $i = 1, \dots, d$ we have*

$$\pi^{(i)}(\psi) = \mathbb{E}\big[\psi(\lambda_l^{(i)}(\theta^*), \lambda_l^{(i)}(\theta), \partial_\theta \lambda_l^{(i)}(\theta))\big].$$

Proof of Lemma 16. For $\psi \in C_b(E, \mathbb{R})$, we define $V^{(i)}(\psi)$ as

$$V^{(i)}(\psi) = \frac{1}{T} \int_0^T \psi(X_s^{(i)}) ds. \tag{67}$$

To show that X_t is ergodic, it is sufficient to show that $V^{(i)}(\psi) \xrightarrow{\mathbb{P}} \pi^{(i)}(\psi)$ where

$$\pi^{(i)}(\psi) = \mathbb{E}\big[\psi(\lambda_l^{(i)}(\theta^*), \lambda_l^{(i)}(\theta), \partial_\theta \lambda_l^{(i)}(\theta))\big].$$

Since L^2 convergence implies convergence in probability, it is sufficient to show L^2 convergence. Since for any random variable X and any nonrandom $a \in \mathbb{R}$ we have $\mathbb{E}[(X-a)^2] = \mathrm{Var}[X] + (\mathbb{E}[X]-a)^2$, we can deduce that

$$\mathbb{E}\big[(V^{(i)}(\psi) - \pi^{(i)}(\psi))^2\big] = \mathrm{Var}[V^{(i)}(\psi)] + (\mathbb{E}[V^{(i)}(\psi)] - \pi^{(i)}(\psi))^2. \quad (68)$$

We define $I^{(i)}$ as

$$I^{(i)} = \mathrm{Var}[V^{(i)}(\psi)].$$

We also define $II^{(i)}$ as

$$II^{(i)} = (\mathbb{E}[V^{(i)}(\psi)] - \pi^{(i)}(\psi))^2.$$

We have

$$\begin{aligned} I^{(i)} &= \mathrm{Var}[V^{(i)}(\psi)] \\ &= \mathrm{Var}\Big[\frac{1}{T}\int_0^T \psi(X_s^{(i)})ds\Big] \\ &= \frac{1}{T^2}\mathrm{Var}\Big[\int_0^T \psi(X_s^{(i)})ds\Big]. \end{aligned}$$

Here, we use the definition of $I^{(i)}$ in the first equality, the definition of $V^{(i)}(\psi)$ in the second equality, and the fact that for any nonrandom $a \in \mathbb{R}$ and any random variable X we have $\mathrm{Var}[aX] = a^2\,\mathrm{Var}[X]$ in the third equality. Then, we have

$$\begin{aligned} I^{(i)} &= \frac{1}{T^2}\mathrm{Var}\Big[\int_0^T \psi(X_s^{(i)})ds\Big] \\ &= \frac{1}{T^2}\lim_{K\to\infty}\mathrm{Var}\Big[\frac{T}{K}\sum_{k=0}^{K-1}\psi(X_{kT/K}^{(i)})\Big] \\ &= \frac{1}{T^2}\lim_{K\to\infty}\frac{T^2}{K^2}\mathrm{Var}\Big[\sum_{k=0}^{K-1}\psi(X_{kT/K}^{(i)})\Big]. \end{aligned}$$

Here, we use the approximation of the Riemann sum in the second equality as the random function $\psi(X_s^{(i)})$ is Riemann integrable for any $\omega \in \Omega$, and an application of the dominated convergence theorem in the third equality. Then, we have

$$\begin{aligned} I^{(i)} &= \frac{1}{T^2}\lim_{K\to\infty}\frac{T^2}{K^2}\mathrm{Var}\Big[\sum_{k=0}^{K-1}\psi(X_{kT/K}^{(i)})\Big] \\ &= \frac{1}{T^2}\lim_{K\to\infty}\frac{T^2}{K^2}\sum_{k=0}^{K-1}\sum_{l=0}^{K-1}\mathrm{Cov}\Big[\psi(X_{kT/K}^{(i)}), \psi(X_{lT/K}^{(i)})\Big] \\ &= \frac{1}{T^2}\int_0^T\int_0^T \mathrm{Cov}\Big[\psi(X_s^{(i)}), \psi(X_u^{(i)})\Big]dsdu. \end{aligned}$$

Here, we use Bienayme's identity in the second equality. By Definition 3, we obtain

$$I^{(i)} \leq \frac{1}{T^2} \int_0^T \int_0^T \mu^{(i)}_{|s-u|} dsdu.$$

A split of the integral into two terms leads to

$$\begin{aligned} I^{(i)} \leq & \frac{1}{T^2} \int_0^T \int_0^T \mu^{(i)}_{|s-u|} \mathbf{1}_{\{|s-u| \leq \sqrt{T}\}} dsdu \\ & + \frac{1}{T^2} \int_0^T \int_0^T \mu^{(i)}_{|s-u|} \mathbf{1}_{\{|s-u| > \sqrt{T}\}} dsdu. \end{aligned} \tag{69}$$

By Lemma 14, there exists $\mu^{(i)}_+ > 0$ such that for any $t \geq 0$ we have $\mu^{(i)}_t \leq \mu^{(i)}_+$. Then, we obtain that

$$\begin{aligned} & \frac{1}{T^2} \int_0^T \int_0^T \mu^{(i)}_{|s-u|} \mathbf{1}_{\{|s-u| \leq \sqrt{T}\}} dsdu \\ \leq & \frac{\mu^{(i)}_+}{T^2} \int_0^{nT} \int_0^{nT} \mathbf{1}_{\{|s-u| \leq \sqrt{nT}\}} dsdu. \end{aligned} \tag{70}$$

Then, we can deduce that

$$\frac{\mu^{(i)}_+}{T^2} \int_0^{nT} \int_0^{nT} \mathbf{1}_{\{|s-u| \leq \sqrt{nT}\}} dsdu \to 0. \tag{71}$$

By Expressions (70) and (71), we can deduce that

$$\frac{1}{T^2} \int_0^T \int_0^T \mu^{(i)}_{|s-u|} \mathbf{1}_{\{|s-u| \leq \sqrt{T}\}} dsdu \to 0. \tag{72}$$

We also have

$$\begin{aligned} & \frac{1}{T^2} \int_0^T \int_0^T \mu^{(i)}_{|s-u|} \mathbf{1}_{\{|s-u| > \sqrt{T}\}} dsdu \\ \leq & \sup_{y > \sqrt{T}} \mu^{(i)}_y \frac{1}{T^2} \int_0^T \int_0^T \mathbf{1}_{\{|s-u| > \sqrt{T}\}} dsdu. \end{aligned} \tag{73}$$

Then, we obtain

$$\sup_{y > \sqrt{T}} \mu^{(i)}_y \frac{1}{T^2} \int_0^T \int_0^T \mathbf{1}_{\{|s-u| > \sqrt{T}\}} dsdu \quad \leq \quad \sup_{y > \sqrt{T}} \mu^{(i)}_y. \tag{74}$$

Since $\mu^{(i)}_T \to 0$ by an application of Lemma 14, we can also deduce that

$$\sup_{y > \sqrt{T}} \mu^{(i)}_y \to 0. \tag{75}$$

Expressions (73), (74) and (75) imply that

$$\frac{1}{T^2}\int_0^T\int_0^T \mu_{|s-u|}^{(i)} \mathbf{1}_{\{|s-u|>\sqrt{T}\}} dsdu \to 0. \tag{76}$$

Expressions (69), (72) and (76) yield that

$$I^{(i)} \to 0. \tag{77}$$

By the definitions of $II^{(i)}$ and $V^{(i)}$, we have

$$II^{(i)} = \Big(\mathbb{E}\Big[\frac{1}{T}\int_0^T \psi(X_s^{(i)})ds\Big] - \pi^{(i)}(\psi)\Big)^2.$$

By Fubini's theorem with Lemmas 10 and 13, we obtain

$$II^{(i)} = \Big(\frac{1}{T}\int_0^T \mathbb{E}\Big[\psi(X_s^{(i)})\Big]ds - \pi^{(i)}(\psi)\Big)^2. \tag{78}$$

By Lemma 15, we have that

$$X_T^{(i)} \xrightarrow{\mathcal{D}} \big(\lambda_l^{(i)}(\theta^*), \lambda_l^{(i)}(\theta), \partial_\theta \lambda_l^{(i)}(\theta)\big).$$

Since convergence in distribution implies convergence in expectation of any bounded function, we obtain that

$$\mathbb{E}[\psi(X_T^{(i)})] \to \mathbb{E}\big[\psi(\lambda_l^{(i)}(\theta^*), \lambda_l^{(i)}(\theta), \partial_\theta \lambda_l^{(i)}(\theta))\big].$$

By the definition of $\pi^{(i)}(\psi)$, we can deduce

$$\mathbb{E}[\psi(X_T^{(i)})] \to \pi^{(i)}(\psi). \tag{79}$$

Expressions (78) and (79) imply that

$$II^{(i)} \to 0. \tag{80}$$

By Expressions (68), (77) and (80), we can deduce

$$\mathbb{E}\big[(V^{(i)}(\psi) - \pi^{(i)}(\psi))^2\big].$$

□

We now show the following lemma, which corresponds to Assumption 1 (e). This complements Lemma A.7 (p. 1836) in Clinet and Yoshida (2017) and extends Lemma C6 in Supplement C of Potiron and Volkov (2025).

Lemma 17. *We assume that Assumption 2 holds. Then, for any $\theta \in \overline{\Theta} - \theta^*$ we have $Y(\theta) \neq 0$.*

Proof of Lemma 17. We assume that $\theta \in \overline{\Theta}$ and that $Y(\theta) = 0$. By Definition (6), we can deduce that

$$0 = \sum_{i=1}^{d} \int_E \Big(\log \Big(\frac{v}{u} \Big) u - (v - u) \Big) \pi_{\theta^*}^{(i)}(du, dv, dw).$$

By Lemma 15, this can be reexpressed as

$$0 = \sum_{i=1}^{d} \mathbb{E} \Big[\log \Big(\frac{\lambda_l^{(i)}(\theta)}{\lambda_l^{(i)}(\theta^*)} \Big) \lambda_l^{(i)}(\theta^*) - \big(\lambda_l^{(i)}(\theta) - \lambda_l^{(i)}(\theta^*) \big) \Big]. \tag{81}$$

we also have by definition for any index $i = 1, \dots, d$ and any $\omega \in \Omega$ that

$$0 \geq \log \Big(\frac{\lambda_l^{(i)}(\theta)}{\lambda_l^{(i)}(\theta^*)} \Big) \lambda_l^{(i)}(\theta^*) - \big(\lambda_l^{(i)}(\theta) - \lambda_l^{(i)}(\theta^*) \big). \tag{82}$$

Expressions (81) and (82) yield a.s.

$$0 = \log \Big(\frac{\lambda_l^{(i)}(\theta)}{\lambda_l^{(i)}(\theta^*)} \Big) \lambda_l^{(i)}(\theta^*) - \big(\lambda_l^{(i)}(\theta) - \lambda_l^{(i)}(\theta^*) \big).$$

We can then deduce that a.s.

$$\lambda_l(\theta^*) = \lambda_l(\theta).$$

Finally, we obtain $\theta^* = \theta$ by Assumption 2 (h). ☐

We now give the proof of Theorem 2. This is based on an application of Theorem 1 with the previous lemmas.

Proof of Theorem 2. The proof is an application of Theorem 1 with Lemmas 7, 10 , 11, 12, 13, 16 and 17. ☐

6. Conclusion

In this paper, we have developed inference for point processes when its intensity has a parametric form. The inference procedure was based on MLE. Under ergodicity of the point process intensity and its derivative, we have shown the CLT of the inference procedure. As an application, we have considered Hawkes mutually exciting processes, where the kernel has a general form and is parametric. We have shown the ergodicity of the Hawkes process intensity and its derivative. Moreover, we have obtained the CLT of the inference procedure for Hawkes processes. In particular, we have allowed for kernels with power distribution, under some smoothness assumptions on the kernel shape. The proofs were based on the application of Burkholder—Davis—Gundy inequalites.

As an application of the technology, Erdemlioglu et al. (2025b) show that Hawkes processes with a periodic log-logistic kernel satisfies Assumption 2. This requires to study deeply some smoothness properties of the log-logistic distribution, when seen as a function of its parameters.

Acknowledgments

The author would like to thank the chief editor Yoshiyuki Ninomiya, one anonymous co-editor, one anonymous associate editor and two anonymous referees.

Funding

The author was supported in part by Japanese Society for the Promotion of Science Grants-in-Aid for Scientific Research (B) 23H00807.

References

Adams, R. A. and Fournier, J. J. (2003). *Sobolev spaces* **140**. Academic press.

Bacry, E., Dayri, K. and Muzy, J.-F. (2012). Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data. *The European Physical Journal B* **85** 1–12.

Bacry, E., Delattre, S., Hoffmann, M. and Muzy, J.-F. (2013). Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Processes and their Applications* **123** 2475–2499.

Bowsher, C. G. (2007). Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics* **141** 876–912.

Brémaud, P. and Massoulié, L. (1996). Stability of nonlinear Hawkes processes. *Annals of Probability* 1563–1588.

Cavaliere, G., Lu, Y., Rahbek, A. and Stærk-Østergaard, J. (2023). Bootstrap inference for Hawkes and general point processes. *Journal of Econometrics* **235** 133–165.

Chavez-Demoulin, V., Davison, A. and McNeil, A. (2005). Estimating value-at-risk: a point process approach. *Quantitative Finance* **5** 227–234.

Chen, F. and Hall, P. (2013). Inference for a nonstationary self-exciting point process with an application in ultra-high frequency financial data modeling. *Journal of Applied Probability* **50** 1006–1024.

Cheysson, F. and Lang, G. (2022). Spectral estimation of Hawkes processes from count data. *Annals of Statistics* **50** 1722–1746.

Clinet, S. and Potiron, Y. (2018). Statistical inference for the doubly stochastic self-exciting process. *Bernoulli* **24** 3469–3493.

Clinet, S. and Yoshida, N. (2017). Statistical inference for ergodic point processes and application to limit order book. *Stochastic Processes and their Applications* **127** 1800–1839.

Costa, M., Graham, C., Marsalle, L. and Tran, V. C. (2020). Renewal in Hawkes processes with self-excitation and inhibition. *Advances in Applied Probability* **52** 879–915.

Daley, D. J. and Vere-Jones, D. (2003). *An introduction to the theory of point processes: Elementary theory and methods* **1**, 2nd ed. Springer New York, NY.

Daley, D. J. and Vere-Jones, D. (2008). *An introduction to the theory of point processes: General theory and structure* **2**, 2nd ed. Springer New York, NY.

Donnet, S., Rivoirard, V. and Rousseau, J. (2020). Nonparametric Bayesian estimation for multivariate Hawkes processes. *Annals of Statistics* **48** 2698–2727.

Embrechts, P., Liniger, T. and Lin, L. (2011). Multivariate Hawkes processes: an application to financial data. *Journal of Applied Probability* **48** 367–378.

Erdemlioglu, D., Potiron, Y., Volkov, V. and Xu, T. (2025a). Estimation of time-dependent latency with locally stationary Hawkes processes. *Working paper available at https://www.fbc.keio.ac.jp/~potiron/Erdemlioglu2025workingpaper.pdf*.

Erdemlioglu, D., Potiron, Y., Xu, T. and Volkov, V. (2025b). Estimation of latency for Hawkes processes with a polynomial periodic kernel. *Working paper available at https://www.fbc.keio.ac.jp/~potiron/Erdemlioglu2025workingpaperestimationlatency.pdf*.

Graham, C. (2021). Regenerative properties of the linear Hawkes process with unbounded memory. *Annals of Applied Probability* **31** 2844–2863.

Hardiman, S. J., Bercot, N. and Bouchaud, J.-P. (2013). Critical reflexivity in financial markets: a Hawkes process analysis. *The European Physical Journal B* **86** 1–9.

Hawkes, A. G. (1971a). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58** 83–90.

Hawkes, A. G. (1971b). Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological)* 438–443.

Hawkes, A. (2018). Hawkes processes and their applications to finance: a review. *Quantitative Finance* **18** 193–198.

Hawkes, A. G. and Oakes, D. (1974). A cluster process representation of a self-exciting process. *Journal of Applied Probability* **11** 493–503.

Horst, U. and Xu, W. (2021). Functional limit theorems for marked Hawkes point measures. *Stochastic Processes and their Applications* **134** 94–131.

Horst, U. and Xu, W. (2022). The microstructure of stochastic volatility models with self-exciting jump dynamics. *Annals of Applied Probability* **32** 4568–4610.

Horst, U. and Xu, W. (2024+). Functional limit theorems for Hawkes processes. *To appear in Probability Theory and Related Fields*.

Jacod, J. (1975). Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **31** 235–253.

Jacod, J. and Protter, P. E. (2012). *Discretization of processes*. Springer Berlin, Heidelberg.

Jacod, J. and Shiryaev, A. (2003). *Limit theorems for stochastic processes*, 2nd ed. Springer Berlin, Heidelberg.

Jaisson, T. and Rosenbaum, M. (2015). Limit theorems for nearly unstable

Hawkes processes. *Annals of Applied Probability* **25** 600–631.

Jaisson, T. and Rosenbaum, M. (2016). Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes. *Annals of Applied Probability* **26** 2860–2882.

Karim, R. S., Laeven, R. J. and Mandjes, M. R. (2025+). Compound Multivariate Hawkes Processes: Large Deviations and Rare Event Simulation. *To appear in Bernoulli*.

Kwan, T. (2023). Asymptotic analysis and ergodicity of the Hawkes process and its extensions, PhD thesis, UNSW Sydney.

Kwan, T., Chen, F. and Dunsmuir, W. (2023). Alternative asymptotic inference theory for a nonstationary Hawkes process. *Journal of Statistical Planning and Inference* **227** 75–90.

Large, J. (2007). Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets* **10** 1-25.

Liniger, T. (2009). Multivariate Hawkes processes, PhD thesis, ETH Zurich.

Mammen, E. and Müller, M. (2023). Nonparametric estimation of locally stationary Hawkes processes. *Bernoulli* **29** 2062–2083.

Ogata, Y. (1978). The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics* **30** 243–261.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical association* **83** 9–27.

Ozaki, T. (1979). Maximum likelihood estimation of Hawkes' self-exciting point processes. *Annals of the Institute of Statistical Mathematics* **31** 145–155.

Potiron, Y. (2025). Nonparametric inference for Hawkes processes with a stochastic time-dependent baseline. *Working paper available at https://www.fbc.keio.ac.jp/~potiron/Potiron2025nonparametricworkingpaper.pdf*.

Potiron, Y. and Volkov, V. (2025). Mutually exciting point processes with latency. *To appear in Journal of the American Statistical Association*.

Potiron, Y., Scaillet, O., Volkov, V. and Yu, S. (2025a). High-frequency estimation of Itô semimartingale baseline for Hawkes processes. *Working paper available at https://www.fbc.keio.ac.jp/~potiron/Potiron2025workingpaper.pdf*.

Potiron, Y., Scaillet, O., Volkov, V. and Yu, S. (2025b). Estimation of branching ratio for Hawkes processes with Itô semimartingale baseline. *Working paper available at https://www.fbc.keio.ac.jp/~potiron/Potiron2025estimationworkingpaper.pdf*.

Reynaud-Bouret, P. and Roy, E. (2007). Some non asymptotic tail estimates for Hawkes processes. *Bulletin Belgium Mathematical Society Simon Stevin* **13** 883–896.

Reynaud-Bouret, P. and Schbath, S. (2010). Adaptive estimation for Hawkes processes; application to genome analysis. *Annals of Statistics* **38** 2781–2822.

Roueff, F., von Sachs, R. and Sansonnet, L. (2016). Locally stationary Hawkes processes. *Stochastic Processes and their Applications* **126** 1710–1743.

Roueff, F. and Von Sachs, R. (2019). Time-frequency analysis of locally stationary Hawkes processes. *Bernoulli* **25** 1355–1385.

Rubin, I. (1972). Regular point processes and their detection. *IEEE Transactions on Information Theory* **18** 547–557.

Van der Vaart, A. W. (2000). *Asymptotic statistics*. Cambridge university press.

Vere-Jones, D. and Ozaki, T. (1982). Some examples of statistical estimation applied to earthquake data: I. Cyclic Poisson and self-exciting models. *Annals of the Institute of Statistical Mathematics* **34** 189–207.

Xu, W. (2024). Diffusion approximations for self-excited systems with applications to general branching processes. *Annals of Applied Probability* **34** 2650–2713.

Zhu, L. (2015). Large deviations for Markovian nonlinear Hawkes processes. *Annals of Applied Probability* **25** 548–581.