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 theory 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 theory 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.

MSC2020 subject classifications: Primary 60G55, 62M09; secondary 60F05.
Keywords and phrases: Hawkes mutually exciting processes, general kernel, point processes, inference, central limit theory, ergodicity.

1. Introduction

This paper concerns 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)). If we define N_t as the aggregated number of events up to time t and λ_t its corresponding intensity, a standard definition of Hawkes doubly exciting processes is given by

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

Here, d is the number of point processes, 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 applied probability. 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) and Mammen and Müller (2023) study locally stationary Hawkes processes. See also Potiron et al. (2025) and Erdemlioglu, Potiron and Volkov (2025). 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 theory (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 inference for Hawkes processes with parametric kernel make this ergodicity assumption (see, e.g., 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. 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 itself but not for its derivative. 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. 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.

For any space S such that $0 \in S$, we define the space without zero as S_*. For any space S, $\mathcal{B}(S)$ is the Borel σ-algebra on the space S. For a vector V, we denote its i-th component as $V^{(i)}$. In what follows, we introduce the multidimensional point process N_t. We denote its dimension as d. For $i = 1, \cdots, d$, each component of the point process $N_t^{(i)}$ counts the number of events between 0 and t for the i-th process. We define $N_t^{(i)}$ as a simple point process on the space of positive real numbers $\mathbb{R}^+$, i.e. a family $\{N^{(i)}(C)\}_{C \in \mathcal{B}(\mathbb{R}^+)}$ of random variables with values in the space of natural integers $\mathbb{N}$. 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 first time is equal to 0 and the following times are increasing for each process a.s., i.e. $\mathbb{P}(T_0^{(i)} = 0 \text{ and } T_k^{(i)} < T_{k+1}^{(i)} \text{ for } k \in \mathbb{N}_* \text{ and } i = 1, \cdots, d) =$

1. 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 $i, j = 1, \cdots, d$ 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 nonnegative $\mathcal{F}_t$-progressively measurable process $\{\lambda_t\}_{t \in \mathbb{R}^+}$, which is d dimensional and satisfies $\mathbb{E}[N((a,b]) \mid \mathcal{F}_a] = \mathbb{E}\big[\int_a^b \lambda_s ds \,\big|\, \mathcal{F}_a\big]$ a.s. for all intervals $(a, b]$, is called an $\mathcal{F}_t$-intensity of N_t. Intuitively, the intensity corresponds to the expected number of events given the past information, i.e.

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

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 $\theta \in \Theta$. Finally, we assume the existence of the true parameter $\theta^* \in \Theta$ such that

$$\lambda_t = \lambda_t(\theta^*). \tag{2}$$

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

$$l_T(\theta) = \sum_{i=1}^{d} \int_0^T \log(\lambda_t^{(i)}(\theta)) dN_t^{(i)} - \sum_{i=1}^{d} \int_0^T \lambda_t^{(i)}(\theta) dt. \tag{3}$$

Here, 0 is the starting time and T is the final time. Then, the MLE is defined as the 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 $\mathcal{C}_b(E, F)$ 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 the supplementary materials of Potiron and Volkov (2025).

Definition 1. We say that X is ergodic if for any $i = 1, \cdots, d$ 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 shows that for any $i = 1,\cdots,d$ there exists a probability measure $\pi_{\theta^*}^{(i)}$ on E such that, for any $\psi \in C_\uparrow(E,\mathbb{R})$, we have

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

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 as

$$Y(\theta) = \sum_{i=1}^{d} \int_E \Big(\log\Big(\frac{v}{u}\Big)u - (v-u)\Big)\pi_{\theta^*}^{(i)}(du,dv,dw). \tag{4}$$

For $k \in \mathbb{N}$, we denote by $C^k(E,F)$ the space of functions which are k times continuously differentiable from E to F. We define $\overline{\Theta}$ as the closure space of Θ. If x is a real number, a vector or a matrix, we define the sum of the absolute values of its components as $|x| = \sum_i |x|_i$. If X is a random variable, we define its L^p norm as $||X||_p = \mathbb{E}[X^p]^{1/p}$. For any $i = 1,\cdots,d$ and any parameter $\theta \in \Theta$, we define F 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 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 $\lambda : \Omega \times \mathbb{R}^+ \times \Omega \to \mathbb{R}^+$ 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 $s \in \mathbb{R}^+$ a.e., we have a.s. that $\theta \to \lambda_s(\theta)$ is in $C^2(\Theta,\mathbb{R}_+^d)$ 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 $i = 1, \cdots, d$, 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 spacee and is weaker than the framework from Clinet and Yoshida (2017) 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 $p \in \mathbb{N}_*$. Assumption 1 (e) is required for the non-degeneracy of the inference procedure and is Assumption [A4] from Clinet and Yoshida (2017). Finally, Assumption 1 (f) only requires that the intensity is continuously differentiable twice whereas Assumption [A1] (ii) in Clinet and Yoshida (2017) needs that the intensity is continuously differentiable three times.

If we consider a vector $z \in \mathbb{R}^n$, we define the tensor product as $z^{\otimes 2} = z \times z^T \in \mathbb{R}^{n \times n}$. We define the $n \times n$ dimensional Fisher information matrix Γ as

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

The inverse of the Fisher information matrix, i.e. Γ^{-1}, is the asymptotic covariance matrix. Also ξ is defined as a n dimensional standard normal vector.

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.

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.*

$$\sqrt{T}(\widehat{\theta}_T - \theta^*) \xrightarrow{\mathcal{D}} \Gamma^{-1/2}\xi. \tag{6}$$

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

$$\lambda_t(\theta) = \nu + \int_0^t h(t-s,\kappa)\,dN_s. \tag{7}$$

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

$$\lambda_t = \lambda_t(\theta^*). \tag{8}$$

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

$$h(t,\kappa_t^+) = \sup_{\kappa\in\Theta_\kappa} \rho(h(t,\kappa)). \tag{9}$$

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

$$h^2(t,\kappa_{t,2}^+) = \sup_{\kappa\in\Theta_\kappa} \rho(h^2(t,\kappa)). \tag{10}$$

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 $i = 1, \cdots, d$, $j = 1, \cdots, d$ 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

$$\left|\partial_\kappa h^{(i,j)}(t,\kappa)^{(k_{t,3}^{(i,j)})}\right| = \sup_{k=1,\cdots,d} \left|\partial_\kappa h^{(i,j)}(t,\kappa)^{(k)}\right|. \tag{11}$$

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 $i = 1, \cdots, d$ and $j = 1, \cdots, d$. For any $i = 1, \cdots, d$, $j = 1, \cdots, d$ and any $t \in \mathbb{R}^+$, 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

$$\left|\partial_\kappa^2 h^{(i,j)}(t,\kappa)^{(k_{t,4}^{(i,j)}, l_{t,4}^{(i,j)})}\right| = \sup_{k,l=1,\cdots,d} \left|\partial_\kappa^2 h^{(i,j)}(t,\kappa)^{(k,l)}\right|. \tag{12}$$

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 $i = 1, \cdots, d$ and $j = 1, \cdots, d$. Moreover, we define the $(n-d) \times (n-d)$ 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 $i = 1, \cdots, d$ and $j = 1, \cdots, d$. 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 $i = 1, \cdots, d$ and $j = 1, \cdots, d$.

We now introduce a set of assumptions required for the CLT of the 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) We assume that $\Theta \subset \mathbb{R}^n$ is such that its closure $\overline{\Theta}$ is a compact space.

(b) There exists $\nu_- \in \mathbb{R}_*^+$ such that for any $\nu \in \Theta_\nu$ and any $i = 1, \ldots, d$ we have $\nu^{(i)} > \nu_-$.
(c) For any $\kappa \in \Theta_\kappa$ and any $t \in \mathbb{R}^+$, we have $h(t,\kappa) > 0$.
(d) We have $\rho(\phi) < 1$ and $\rho(\phi_2) < 1$.
(e) For any $s \in \mathbb{R}^+$ a.e., we have $\kappa \to h(s,\kappa)$ is in $\mathcal{C}^2(\Theta_\kappa, \mathbb{R}_+^d)$ and there exists a continuous extension to $\overline{\Theta}_\kappa$.
(f) For any $\kappa \in \Theta_\kappa$, we have $\rho(\phi_3(\kappa)) < 1$ and $\rho(\phi_4(\kappa)) < 1$.

(g) We have $\rho(\phi_5) < 1$ and $\rho(\phi_6) < 1$.

Assumptions 2 (a) and (b) imply that the point processes are well-defined and are also required in the simpler case of heterogeneous Poisson processes without a kernel (see Daley and Vere-Jones (2003)). Assumption 2 (c) are restrictive for kernels with inhibitory effects. Assumption 2 (d) is slightly stronger than the condition 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)). Assumption 2 (d) is used to prove Assumption 1 (d). Assumption 2 (e) is required to show Assumption 1 (f). Moreover, Assumption 2 (f) yields Assumption 1 (g). Finally, Assumption 2 (g) implies Assumption 1 (h).

In the theorem that follows, we state the CLT of the 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.*

$$\sqrt{T}(\widehat{\theta}_T - \theta^*) \xrightarrow{\mathcal{D}} \Gamma^{-1/2}\xi. \tag{13}$$

4. Proofs of the CLT for point processes

In this section, we give the proofs of the CLT of the 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 need to 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 $\pi_{\theta^*}(\psi, \theta)$. 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 $\theta \in \Theta$, we have*

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

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

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 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

$$Y_T(\theta) = \frac{1}{T}(l_T(\theta) - l_T(\theta^*)). \tag{14}$$

We define the compensated point process as

$$M_t = N_t - \int_0^t \lambda_s(\theta^*)ds. \tag{15}$$

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 $\theta \in \Theta$. 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 2. *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 2. 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 (14) in the first equality, Definition (3) in the second equality, algebraic manipulation in the third equality, Definition (15) 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

$$\sup_{\theta\in\Theta}\Big|\sum_{i=1}^d I_T^{(i)}(\theta)\Big| \xrightarrow{\mathbb{P}} 0. \tag{16}$$

Since L^2 convergence implies convergence in probability, it is sufficient to show Expression (16) that

$$\mathbb{E}\Big[\Big(\sup_{\theta\in\Theta}\Big|\sum_{i=1}^d I_T^{(i)}(\theta)\Big|\Big)^2\Big] \to 0. \tag{17}$$

By the triangular inequality, we can deduce that

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

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

$$\mathbb{E}\Big[\Big(\sup_{\theta\in\Theta}\sum_{i=1}^d \Big|I_T^{(i)}(\theta)\Big|\Big)^2\Big] = \mathbb{E}\Big[\Big(\sup_{\theta\in\Theta}\sum_{i=1}^d \Big|\frac{1}{T}\int_0^T \log\Big(\frac{\lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta^*)}\Big) dM_t^{(i)}\Big|\Big)^2\Big]. \tag{19}$$

By supremum properties, we get

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

By the triangular inequality, we can deduce that

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

Again by supremum properties, we can deduce that

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

By an algebraic manipulation, we can deduce that

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

By an application of Burkholder-Davis-Gundy inequalites (see Expression (2.1.32) in Jacod and Protter (2012) (p. 39)) with the fact that M_t is an $\mathcal{F}_t$-local martingale, we obtain

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

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

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

This can be reexpressed as

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

By Tonelli theorem, this can be reexpressed as

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

By Expressions (17) to (27) with Assumption 1 (d), we can prove Expression (16). To prove that $|\sum_{i=1}^{d} II_T^{(i)}(\theta) - Y(\theta)| \xrightarrow{\mathbb{P}} 0$, we can use Lemma 1. $\square$

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

Lemma 3. *We assume that Assumptions 1 (a), (b), (c), (d) and (e) hold. We have the consistency of the inference procedure based on MLE, i.e.*

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

Proof of Lemma 3. By the definition (14), we can deduce that $Y_T(\theta) \leq 0$ for any $\theta \in \Theta$ and $\overline{Y}(\theta^*) = 0$. By Assumption 1 (e), we have that θ^* is a global maximum of Y. We can then conclude by Lemma 2. $\square$

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

$$\partial_\theta l_T(\theta) = \sum_{i=1}^{d} \int_0^T \frac{\partial_\theta \lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta)} \mathbf{1}_{\{\lambda_t^{(i)}>0\}} dN_t^{(i)} - \sum_{i=1}^{d} \int_0^T \partial_\theta \lambda_t^{(i)}(\theta) dt. \tag{28}$$

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

$$\partial_\theta^2 l_T(\theta) = \sum_{i=1}^{d} \int_0^T \partial_\theta \Big(\frac{\partial_\theta \lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta)} \Big) \mathbf{1}_{\{\lambda_t^{(i)}>0\}} dN_t^{(i)} - \sum_{i=1}^{d} \int_0^T \partial_\theta^2 \lambda_t^{(i)}(\theta) dt \tag{29}$$

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

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

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

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

From the definition (3), 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 $\theta \in \Theta$, any $T > 0$ and any $i = 1, \cdots, d$, we have that $I_T^{(i)}(\theta)$ is a.s. finite and a.s. differentiable with partial derivatives equal to

$$\partial_\theta I_T^{(i)} \quad = \quad \int_0^T \partial_\theta \log(\lambda_t^{(i)}(\theta)) dN_t^{(i)}. \tag{30}$$

Since $N_t^{(i)}$ is a simple point process, $I_T^{(i)}(\theta)$ can be reexpressed as

$$I_T^{(i)}(\theta) = \sum_{k \in \mathbb{N}_* \text{ s.t. } 0 < T_k^{(i)} < T} \log\left(\lambda_{T_k^{(i)}}^{(i)}(\theta)\right). \tag{31}$$

Since $N_t^{(i)}$ is a simple point process, we have that the number of terms in the sum and each term are a.s. finite. 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 (31), this equality can be reexpressed as Equation (30). Finally, we get by differentiating the term inside the integral that

$$\partial_\theta I_T^{(i)} \quad = \quad \int_0^T \frac{\partial_\theta \lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta)} \mathbf{1}_{\{\lambda_t^{(i)} > 0\}} dN_t^{(i)}. \tag{32}$$

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

$$\partial_\theta II_T^{(i)}(\theta) \quad = \quad \int_0^T \partial_\theta \lambda_t^{(i)}(\theta) dt. \tag{33}$$

Since $N_t^{(i)}$ is a simple point process, $II_T^{(i)}(\theta)$ is a.s. finite. To show Equation (33), 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 (33). Finally, Equations (32) and (33) yield Equation (28). We can show Equation (29) 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 4.

Lemma 5. *We assume that Assumptions 1 (a), (f) and (g) hold. For any* $\theta \in \Theta$

and any $T > 0$, we have that the Hessian matrix of $l_T(\theta)$ can be rewritten as

$$\begin{aligned}
\partial_\theta^2 l_T(\theta) &= \sum_{i=1}^{d} \int_0^T \partial_\theta \Big(\frac{\partial_\theta \lambda_t^{(i)}(\theta)}{\lambda_t^{(i)}(\theta)} \Big) \mathbf{1}_{\{\lambda_t^{(i)} > 0\}} dM_t^{(i)} \\
&\quad - \sum_{i=1}^{d} \int_0^T (\partial_\theta \lambda_t^{(i)})^{\otimes 2}(\theta) \frac{\lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta)^2} \mathbf{1}_{\{\lambda_t^{(i)} > 0\}} dt \\
&\quad + \sum_{i=1}^{d} \int_0^T (\partial_\theta^2 \lambda_t^{(i)})(\theta) \frac{\lambda_t^{(i)}(\theta) - \lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta)} \mathbf{1}_{\{\lambda_t^{(i)} > 0\}} dt. \qquad (34)
\end{aligned}$$

Proof. The proof of this lemma is a direct application of Lemma 4. □

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 continuously differentiable in $\theta \in \Theta$ a.s. for any $t \in \mathbb{R}^+$. Thus, we can apply a Taylor expansion. 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

$$\begin{aligned}
\frac{-\Gamma^{-1}}{\sqrt{T}} \partial_\theta l_T(\theta^*) &\xrightarrow{\mathcal{D}} \Gamma^{-1/2} \xi, \qquad (35) \\
\frac{-\Gamma^{-1}}{T} \partial_\theta^2 l_T(\zeta) &\xrightarrow{\mathbb{P}} 1. \qquad (36)
\end{aligned}$$

Here, 1 is the $n \times n$ dimensional unity matrix. Then, the theorem follows using Slutsky's theorem.

We prove now Equation (35). Equation (28) from Lemma 4 can be reexpressed as

$$\partial_\theta l_T(\theta^*) = \sum_{i=1}^{d} \int_0^T \frac{\partial_\theta \lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)} \mathbf{1}_{\{\lambda_t^{(i)}>0\}} dM_t^{(i)}. \tag{37}$$

By Equation (37), we have that

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

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

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

We use Theorem $VIII.3.24$ in Jacod and Shiryaev (2003). We can calculate that

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

We define ΔS_s as the jump of the process S at time s. We prove now that Lindeberg's condition is satisfied. For any $a > 0$, we have

$$\begin{aligned} \mathbb{E}\Big[\sum_{s\le u} |\Delta S_s|^2 \mathbf{1}_{|\Delta S_s|>a}\Big] &\le \mathbb{E}\Big[\frac{1}{a}\sum_{s\le u} |\Delta S_s|^3\Big] \\ &= \mathbb{E}\Big[\frac{1}{a}\sum_{s\le u} \Big|\frac{-\Gamma^{-1}}{\sqrt{T}} \sum_{i=1}^{d} \int_0^{uT} \frac{\partial_\theta \lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)} \mathbf{1}_{\{\lambda_t^{(i)}>0\}} dM_t^{(i)}\Big|^3\Big] \\ &= \mathbb{E}\Big[\frac{1}{a}\sum_{s\le u} \Big|\frac{-\Gamma^{-1}}{\sqrt{T}} \sum_{i=1}^{d} \int_0^{uT} \frac{\partial_\theta \lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)} \mathbf{1}_{\{\lambda_t^{(i)}>0\}} dN_t^{(i)}\Big|^3\Big]. \end{aligned}$$

Here, we use the fact that $\mathbf{1}_{|\Delta S_s|>a} \le \frac{1}{a}|\Delta S_s|$ in the inequality, the first equality is due to Definition (38), and the second equality is explained by the fact that the compensator term does not jump. Then, we have

$$\begin{aligned} \mathbb{E}\Big[\sum_{s\le u} |\Delta S_s|^2 \mathbf{1}_{|\Delta S_s|>a}\Big] &\le \mathbb{E}\Big[\frac{1}{a}\sum_{s\le u} \Big|\frac{-\Gamma^{-1}}{\sqrt{T}} \sum_{i=1}^{d} \int_0^{uT} \frac{\partial_\theta \lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)} \mathbf{1}_{\{\lambda_t^{(i)}>0\}} dN_t^{(i)}\Big|^3\Big] \\ &= \mathbb{E}\Big[\frac{1}{a}\sum_{i=1}^{d} \int_0^{uT} \Big|\frac{-\Gamma^{-1}}{\sqrt{T}} \frac{\partial_\theta \lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)}\Big|^3 \mathbf{1}_{\{\lambda_t^{(i)}>0\}} dN_t^{(i)}\Big] \\ &= \mathbb{E}\Big[\frac{1}{a}\sum_{i=1}^{d} \int_0^{uT} \Big|\frac{-\Gamma^{-1}}{\sqrt{T}} \frac{\partial_\theta \lambda_t^{(i)}(\theta^*)}{\lambda_t^{(i)}(\theta^*)}\Big|^3 \lambda_t^{(i)}(\theta^*) dt\Big]. \end{aligned}$$

Here, the first and second equalities are a consequence of the form of $dN_t^{(i)}$. We can continue to bound the Linderberg's term by

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

Finally, we can bound the Linderberg's term by

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

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 (35) is satisfied.

We prove now Equation (36), i.e. that $\frac{-\Gamma^{-1}}{T}\partial_\theta^2 l_T(\zeta)\xrightarrow{\mathbb{P}}1$. 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

$$\sup_{\theta\in V}|\Gamma+T^{-1}\partial_\theta^2 l_T(\theta)|\xrightarrow{\mathbb{P}}0. \tag{39}$$

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

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

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

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

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

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

By Equation (29) from Lemma 4, Expression (39) can be reexpressed as

$$\sup_{\theta \in V} \Big| \Gamma + T^{-1} \sum_{i=1}^d \Big(I_T^{(i)}(\theta) - II_T^{(i)}(\theta) + III_T^{(i)}(\theta) \Big) \Big| \xrightarrow{\mathbb{P}} 0. \tag{40}$$

By Assumption 1 (h), we can prove with the same arguments from the proof of Expression (16) that

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

By Assumption 1 (c), we obtain that

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

We can deduce by the triangular inequality and supremum properties that

$$\sup_{\theta \in V} \Big| T^{-1} \sum_{i=1}^d II_T^{(i)}(\theta) - T^{-1} \sum_{i=1}^d II_T^{(i)}(\theta^*) \Big| \leq \sum_{i=1}^d \sup_{\theta \in V} \Big| T^{-1} \big(II_T^{(i)}(\theta) - II_T^{(i)}(\theta^*) \big) \Big|. \tag{43}$$

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

$$\begin{aligned} & \sup_{\theta \in V} \Big| T^{-1} \big(II_T^{(i)}(\theta) - II_T^{(i)}(\theta^*) \big) \Big| \qquad (44) \\ = \; & \sup_{\theta \in V} \Big| T^{-1} \int_0^T \Big(\frac{(\partial_\theta \lambda_t^{(i)})^{\otimes 2}(\theta)}{\lambda_t^{(i)}(\theta)^2} - \frac{(\partial_\theta \lambda_t^{(i)})^{\otimes 2}(\theta^*)}{\lambda_t^{(i)}(\theta^*)^2} \Big) \lambda_t^{(i)}(\theta^*) \mathbf{1}_{\{\lambda_t^{(i)}>0\}} dt \Big|. \end{aligned}$$

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

$$F_t^{(i)}(\theta) - F_t^{(i)}(\theta^*) = \partial_\theta F_t^{(i)}(\widetilde{\theta})(\theta - \theta^*), \tag{45}$$

where $\widetilde{\theta}$ is between θ and θ^*. By Equations (44) and (45), we can deduce that

$$\begin{aligned} & \sup_{\theta \in V} \Big| T^{-1} \big(II_T^{(i)}(\theta) - II_T^{(i)}(\theta^*) \big) \Big| \qquad (46) \\ \leq \; & \sup_{\theta \in V} \Big| T^{-1} \int_0^T \partial_\theta F_t^{(i)}(\widetilde{\theta})(\theta - \theta^*) \lambda_t^{(i)}(\theta^*) \mathbf{1}_{\{\lambda_t^{(i)}>0\}} dt \Big|. \end{aligned}$$

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

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

By Assumption (f), we have that $\partial_\theta F_t^{(i)}$ is continuous in θ^* a.s. for any $t\in\mathbb{R}^+$. Thus, we get that

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

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

$$\begin{aligned}
&T^{-1}\int_0^T \sup_{\theta\in V}\Big|\partial_\theta F_t^{(i)}(\theta^*)\Big|\Big|\theta-\theta^*\Big|\lambda_t^{(i)}(\theta^*)\mathbf{1}_{\{\lambda_t^{(i)}>0\}}dt \qquad (49)\\
\leq\;& T^{-1}s(V)\int_0^T \Big|\partial_\theta F_t^{(i)}(\theta^*)\Big|\lambda_t^{(i)}(\theta^*)\mathbf{1}_{\{\lambda_t^{(i)}>0\}}dt.
\end{aligned}$$

By Assumption 1 (h), we get that

$$T^{-1}s(V)\int_0^T \Big|\partial_\theta F_t^{(i)}(\theta^*)\Big|\lambda_t^{(i)}(\theta^*)\mathbf{1}_{\{\lambda_t^{(i)}>0\}}dt \xrightarrow{\mathbb{P}} 0. \qquad (50)$$

By Expressions (44) to (50), we can deduce that

$$\sup_{\theta\in V}\Big|T^{-1}\big(II_T^{(i)}(\theta)-II_T^{(i)}(\theta^*)\big)\Big| \xrightarrow{\mathbb{P}} 0. \qquad (51)$$

By Assumption 1 (h), we can prove with the same arguments from the proof of Expression (51) that

$$\sup_{\theta\in V}\Big|T^{-1}\sum_{i=1}^d III_T^{(i)}(\theta)\Big| \xrightarrow{\mathbb{P}} 0. \qquad (52)$$

Finally, we can deduce Equation (36) by the use of Expressions (39), (40), (41), (42), (51) and (52). □

5. Proofs 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 6. *We assume that Assumptions 2 (a), (b) and (c) hold. Then, the family of intensities $\lambda : \Omega \times \mathbb{R}^+ \times \Omega \to \mathbb{R}^+$ defined in Equation (7) is $\mathcal{F} \otimes \mathcal{B}(\mathbb{R}^+) \otimes \mathcal{B}(\Theta)$-measurable.*

Proof of Lemma 6. The proof can be deduced from its definition (7). □

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

Lemma 7. *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 7. We first prove that

$$\sup_{t \in \mathbb{R}^+} \Big|\Big| \sup_{\theta \in \Theta} |\lambda_t(\theta)| \ \Big|\Big|_1 < +\infty. \tag{53}$$

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 (7) in the third equality. Then, we have

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

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

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

Here, we use the fact that the kernel depends only on the parameter κ in the first equality, and Definition (12) in the second inequality. Then, we obtain

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

Here, we use point process properties in the first equality, and expectation properties in the second equality. Then, we have

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

Here, we use Assumption 2 (d) to get the existence of $h^+<1$ in the first equality, the definition of $|\ |$ and supremum properties in the second inequality. Thus, we can deduce that

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

Since $h^+<1$, Expression (54) implies Expression (53). Finally, the lemma can be shown by an application of Burkholder-Davis-Gundy inequalites with Assumption 2 (d). □

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

Lemma 8. *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 $\theta \to \lambda_s(\theta)$ is in $C^2(\Theta, \mathbb{R}_+^d)$ and there exists a continuous extension to $\overline{\Theta}$.*

Proof of Lemma 8. By Definition (7), we have

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

Since ν is in $C^2(\Theta, \mathbb{R}_+^d)$ 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 $i = 1, \cdots, d$, the intensity can be rewritten as

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

Since $N_t^{(i)}$ is a simple point process, $\lambda_{t,h}^{(i)}(\theta)$ can be reexpressed as

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

Since $N_t^{(i)}$ is a simple point process, 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 $\theta \to \lambda_s(\theta)$ for any $s \in \mathbb{R}^+$ a.e. is in $C^2(\Theta, \mathbb{R}_+^d)$, and there exists a continuous extension to $\overline{\Theta}$. $\square$

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

Lemma 9. *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 9. By Definition (7), 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 $i = 1, \cdots, d$, $\lambda_{t,h}^{(i)}(\theta)$ can be rewritten as

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

Since $N_t^{(i)}$ is a simple point process, $\lambda_{t,h}^{(i)}(\theta)$ can be reexpressed as

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

Since $N_t^{(i)}$ is a simple point process, 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 $N_t^{(i)}$ is a simple point process, $\partial_\theta \lambda_{t,h}^{(i)}(\theta)$ 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

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

Since $N_t^{(i)}$ is a simple point process, $\partial_\theta^2 \lambda_{t,h}^{(i)}(\theta)$ 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). $\square$

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

Lemma 10. *We assume that Assumption 2 holds. For any $i = 1, \cdots, d$, 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 10. 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 (b) and (c), 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 (b), 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 (7), we obtain

$$I < +\infty. \tag{55}$$

We define II as

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

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 (b) and (c), 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 Assumptions 2 (b), (g) and Equation (7), we obtain

$$II < +\infty. \tag{56}$$

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 (b) and (c), 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 (b), 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 Assumption 2 (g) and Equation (7), we obtain

$$III < +\infty. \tag{57}$$

We can prove the lemma with Expressions (55), (56) and (57). □

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 the supplementary materials of Potiron and Volkov (2025).

Definition 3. We say that X is C-mixing, for some set of functions C from E to $\mathbb{R}$, if for any ϕ, $\psi \in C$ and $i = 1, \cdots, d$, we have

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

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 the supplementary materials of Potiron and Volkov (2025).

Lemma 11. *We assume that Assumption 2 holds. For any $\theta \in \Theta$, X_t is $C_b(E,\mathbb{R})$-mixing in the sense of Definition 3.*

Proof of Lemma 11. We first define the truncation of $X_T^{(i)}$ at time $t \leq T$ as

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

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

$$\begin{aligned}\mu_T^{(i)} \leq\ & \sup_{s\in\mathbb{R}^+}\big|\operatorname{Cov}\big[\phi(X_s^{(i)}),\psi(X_{s+T}^{(i)})-\psi(\widetilde{X}_{s+\sqrt{T},s+T}^{(i)})\big]\big| \\ &+\sup_{s\in\mathbb{R}^+}\big|\operatorname{Cov}\big[\phi(X_s^{(i)}),\psi(\widetilde{X}_{s+\sqrt{T},s+T}^{(i)})\big]\big|.\end{aligned} \tag{58}$$

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 $I_T^{(i)}$, covariance and supremum properties, we can deduce that

$$I_T^{(i)} \leq \sup_{s\in\mathbb{R}^+}\operatorname{Var}\big[\phi(X_s^{(i)})\big]\sup_{s\in\mathbb{R}^+}\operatorname{Var}\big[\psi(X_{s+T}^{(i)})-\psi(\widetilde{X}_{s+\sqrt{T},s+T}^{(i)})\big]. \tag{59}$$

By Lemmas 7 and 10, we get

$$\sup_{s\in\mathbb{R}^+}\operatorname{Var}\big[\phi(X_s^{(i)})\big] \leq C. \tag{60}$$

Since $\sqrt{T}\to\infty$, by Equation (8) and Assumption 2 (d), we obtain

$$\sup_{s\in\mathbb{R}^+}\operatorname{Var}\big[\psi(X_{s+T}^{(i)})-\psi(\widetilde{X}_{s+\sqrt{T},s+T}^{(i)})\big]\to 0. \tag{61}$$

By Expressions (59), (60) and (61), we can deduce that

$$I_T^{(i)} \to 0. \tag{62}$$

As $T \to \infty$, by Equation (8) and Assumption 2 (d), we obtain

$$II_T^{(i)} \to 0. \tag{63}$$

By Expressions (58), (62) and (63), 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 12. *We assume that Assumption 2 hold. For any $\theta \in \Theta$, X_t is stable, i.e. there exists an $\mathbb{R}_+^*$-valued random variable $\lambda_l^{(i)}$ such that for any $i = 1, \cdots, d$ 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 12. The proof is obtained by an application of Theorem 1 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. This extends Lemma 3.16 (p. 1815) in Clinet and Yoshida (2017) and Proposition C1 (iii) in Potiron and Volkov (2025).

Lemma 13. *We assume that Assumptions 2 hold. For any $\theta \in \Theta$, X_t is ergodic in the sense of Definition 1.*

Proof of Lemma 13. For $\psi \in C_b(E, \mathbb{R})$, we define $V^{(i)}(\psi)$ as

$$V^{(i)}(\psi) = \frac{1}{T}\int_0^T \psi(X_s^{(i)})ds. \tag{64}$$

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

$$\mathbb{E}\big[(V^{(i)}(\psi) - \pi^{(i)}(\psi))^2\big] \quad = \quad \mathrm{Var}[V^{(i)}(\psi)] + (\mathbb{E}[V^{(i)}(\psi)] - \pi^{(i)}(\psi))^2. \tag{65}$$

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

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

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 $\operatorname{Var}[aX] = a^2 \operatorname{Var}[X]$ in the third equality. Then, we have

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

Here, we use the approximation of the Riemann sum in the second equality, and an application of the dominated convergence theorem in the third equality. Then, we have

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

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

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

A split of the integral into two terms leads to

$$I^{(i)} \leq \frac{1}{T^2}\int_0^T\int_0^T \mu_{|s-u|}^{(i)}\mathbf{1}_{\{|s-u|\leq\sqrt{T}\}}dsdu + \frac{1}{T^2}\int_0^T\int_0^T \mu_{|s-u|}^{(i)}\mathbf{1}_{\{|s-u|>\sqrt{T}\}}dsdu. \quad (66)$$

By Lemma 11, there exists $\mu_+^{(i)} > 0$ such that for any $t \geq 0$ we have $\mu_t^{(i)} \leq \mu_+^{(i)}$. Then, we obtain that

$$\frac{1}{T^2}\int_0^T\int_0^T \mu_{|s-u|}^{(i)}\mathbf{1}_{\{|s-u|\leq\sqrt{T}\}}dsdu \quad\leq\quad \frac{\mu_+^{(i)}}{T^2}\int_0^{nT}\int_0^{nT}\mathbf{1}_{\{|s-u|\leq\sqrt{nT}\}}dsdu \tag{67}$$

Then, we can deduce that

$$\frac{\mu_+^{(i)}}{T^2}\int_0^{nT}\int_0^{nT}\mathbf{1}_{\{|s-u|\leq\sqrt{nT}\}}dsdu \to 0. \tag{68}$$

By Expressions (67) and (68), we can deduce that

$$\frac{1}{T^2}\int_0^T\int_0^T \mu_{|s-u|}^{(i)}\mathbf{1}_{\{|s-u|\leq\sqrt{T}\}}dsdu \to 0. \tag{69}$$

We also have

$$\frac{1}{T^2}\int_0^T\int_0^T \mu_{|s-u|}^{(i)}\mathbf{1}_{\{|s-u|>\sqrt{T}\}}dsdu \quad\leq\quad \sup_{y>\sqrt{T}}\mu_y^{(i)}\frac{1}{T^2}\int_0^T\int_0^T\mathbf{1}_{\{|s-u|>\sqrt{T}\}}dsdu \tag{70}$$

Then, we obtain

$$\sup_{y>\sqrt{T}}\mu_y^{(i)}\frac{1}{T^2}\int_0^T\int_0^T\mathbf{1}_{\{|s-u|>\sqrt{T}\}}dsdu \quad\leq\quad \sup_{y>\sqrt{T}}\mu_y^{(i)}. \tag{71}$$

Since $\mu_T^{(i)} \to 0$ by an application of Lemma 11, we can also deduce that

$$\sup_{y>\sqrt{T}}\mu_y^{(i)} \to 0. \tag{72}$$

Expressions (70), (71) and (72) imply that

$$\frac{1}{T^2}\int_0^T\int_0^T \mu_{|s-u|}^{(i)}\mathbf{1}_{\{|s-u|>\sqrt{T}\}}dsdu \to 0. \tag{73}$$

Expressions (66), (69) and (73) yield that

$$I^{(i)} \to 0. \tag{74}$$

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 7 and 10, we obtain

$$II^{(i)} = \Big(\frac{1}{T}\int_0^T\mathbb{E}\Big[\psi(X_s^{(i)})\Big]ds - \pi^{(i)}(\psi)\Big)^2. \tag{75}$$

By Lemma 12, 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

$$\mathbb{E}[\psi(X_T^{(i)})] \to \pi^{(i)}(\psi). \tag{76}$$

Expressions (75) and (76) imply that

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

By Expressions (65), (74) and (77), 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 extends Lemma A.7 (p. 1836) in Clinet and Yoshida (2017) and Lemma C6 in the supplementary materials of Potiron and Volkov (2025).

Lemma 14. *We assume that Assumption 2 holds. Then, for any $\theta \in \overline{\Theta} - \theta^*$ we have $Y(\theta) \neq 0$.*

Proof of Lemma 14. We assume that $\theta \in \overline{\Theta}$ and that $Y(\theta) = 0$. By Definition (4), 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 13, this can be reexpressed as

$$0 = \sum_{i=1}^{d} \mathbb{E}\Big[\log\Big(\frac{\lambda_l^{(i)}(\theta)}{\lambda_l^{(i)}(\theta^*)} \Big) \lambda_l^{(i)}(\theta^*) - \big(\lambda_l^{(i)}(\theta) - \lambda_l^{(i)}(\theta^*)\big) \Big]. \tag{78}$$

For any $i = 1, \cdots, d$ we also have by definition

$$0 \geq \log\Big(\frac{\lambda_l^{(i)}(\theta)}{\lambda_l^{(i)}(\theta^*)} \Big) \lambda_l^{(i)}(\theta^*) - \big(\lambda_l^{(i)}(\theta) - \lambda_l^{(i)}(\theta^*)\big). \tag{79}$$

Expressions (78) and (79) 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.

$$\lambda_l^{(i)}(\theta^*) = \lambda_l^{(i)}(\theta).$$

By injectivity of the function $\theta \mapsto (\lambda_l^{(1)}(\theta), \cdots, \lambda_l^{(d)}(\theta))$, we obtain $\theta^* = \theta$. □

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 6, 7 , 8, 9, 10, 13 and 14. □

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.

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. *Eur. Phys. J. 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. *Stoch. Process. Appl.* **123** 2475–2499.

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

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

Chavez-Demoulin, V., Davison, A. and McNeil, A. (2005). Estimating value-at-risk: a point process approach. *Quant. 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. *J. Appl. Probab.* **50** 1006–1024.

Cheysson, F. and Lang, G. (2022). Spectral estimation of Hawkes processes from count data. *Ann. Stat.* **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. *Stoch. Process. Appl.* **127** 1800–1839.

Costa, M., Graham, C., Marsalle, L. and Tran, V. C. (2020). Renewal in Hawkes processes with self-excitation and inhibition. *Adv. Appl. Probab.* **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. *Ann. Stat.* **48** 2698–2727.

Embrechts, P., Liniger, T. and Lin, L. (2011). Multivariate Hawkes processes: an application to financial data. *J. Appl. Probab.* **48** 367–378.

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

Graham, C. (2021). Regenerative properties of the linear Hawkes process with unbounded memory. *Ann. Appl. Probab.* **31** 2844–2863.

Hardiman, S. J., Bercot, N. and Bouchaud, J.-P. (2013). Critical reflexivity in financial markets: a Hawkes process analysis. *Eur. Phys. J. 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. *J. Roy. Stat. Soc. Ser. B* 438–443.

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

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

Horst, U. and Xu, W. (2021). Functional limit theorems for marked Hawkes point measures. *Stoch. Process. Appl.* **134** 94–131.

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

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

Jacod, J. (1975). Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales. *Z. Wahrscheinlichkeitstheorie und verw. 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. *Ann. Appl. Probab.* **25** 600–631.

Jaisson, T. and Rosenbaum, M. (2016). Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes. *Ann. Appl. Probab.* **26** 2860.

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. *J. Stat. Plan. Inference* **227** 75–90.

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. *Ann. I. Stat. Math.* **30** 243–261.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Am. Stat. Assoc.* **83** 9–27.

Ozaki, T. (1979). Maximum likelihood estimation of Hawkes' self-exciting point processes. *Ann. I. Stat. Math.* **31** 145–155.

Potiron, Y. and Volkov, V. (2025). Mutually exciting point processes with latency. *To appear in J. Am. Stat. Assoc.*

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

Reynaud-Bouret, P. and Roy, E. (2007). Some non asymptotic tail estimates for Hawkes processes. *Bull. Belg. Math. Soc. Simon Stevin* **13** 883–896.

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

Roueff, F., von Sachs, R. and Sansonnet, L. (2016). Locally stationary Hawkes processes. *Stoch. Process. Appl.* **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 Trans. Inf. Theory* **18** 547–557.

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

Xu, W. (2024). Diffusion approximations for self-excited systems with applications to general branching processes. *Ann. Appl. Probab.* **34** 2650–2713.

Zhu, L. (2015). Large deviations for Markovian nonlinear Hawkes processes. *Ann. Appl. Probab.* **25** 548–581.