The Annals of Statistics
2022, Vol. 50, No. 3, 1722–1746
https://doi.org/10.1214/22-AOS2173

SPECTRAL ESTIMATION OF HAWKES PROCESSES FROM COUNT DATA

BY FELIX CHEYSSON[a] AND GABRIEL LANG[b]

UMR MIA-Paris, Université Paris-Saclay, AgroParisTech, INRAE, [a]*felix@cheysson.fr,* [b]*gabriel.lang@agroparistech.fr*

This paper presents a parametric estimation method for ill-observed linear stationary Hawkes processes. When the exact locations of points are not observed, but only counts over time intervals of fixed size, methods based on the likelihood are not feasible. We show that spectral estimation based on Whittle's method is adapted to this case and provides consistent and asymptotically normal estimators, provided a mild moment condition on the reproduction function. Simulated data sets and a case-study illustrate the performances of the estimation, notably of the reproduction function even when time intervals are relatively large.

1. Introduction. Hawkes processes, introduced in (Hawkes (1971a, 1971b)), form a family of models for point processes, which exhibit both self-exciting (i.e., the occurrence of any event increases temporarily the probability of further events occurring) and clustering properties: they are special cases of the Poisson cluster process, where each cluster is a continuous-time Galton–Watson tree with a Poisson offspring process, the intensity of which is called the reproduction function (Hawkes and Oakes (1974)). As they exhibit self-exciting and clustering properties, Hawkes processes are appealing in point process modeling, and while first applications concerned almost exclusively seismology (Adamopoulos (1976), Ogata (1988)), their use quickly spread to many other disciplines, including neurophysiology (Chornoboy, Schramm and Karr (1988)), finance (Bacry, Mastromatteo and Muzy (2015), Bowsher (2003), Chavez-Demoulin, Davison and McNeil (2005)), genomics (Reynaud-Bouret and Schbath (2010)) and epidemiology (Meyer, Elias and Höhle (2012)); see also Reinhart (2018) for a review of Hawkes processes and their applications.

Parameter estimation of Hawkes processes has been studied thoroughly when events are fully observed, relying mainly on maximum likelihood methods (Ogata (1978, 1988), Ozaki (1979)). Here, we consider that the arrival times are not observed, but interval censored: the timeline is cut into regular bins corresponding to, for example, days or weeks and the numbers of events in each bin is counted. Exact maximum likelihood methods are no more applicable to such bin-count data: since the resulting process is no longer a point process but a time series, either the point process must be reconstructed from the count data, or the estimation method of the process must be adapted to time series.

For the first strategy, one could assign arbitrarily to each point a location within its interval, for example, by uniformly drawing them in the interval (Meyer, Elias and Höhle (2012)), or attempt a more sophisticated approach such as an expectation maximization algorithm. Historically for Hawkes processes, this algorithm has been used for multivariate processes (Olson and Carley (2013)) or when the immigration intensity is a renewal process (Wheatley, Filimonov and Sornette (2016)), treating the genealogy as a latent variable. For an interval censored process, an analogous approach which would consider the arrival times as latent variables is unfortunately not adapted, since there is no closed form for the conditional distribution of the arrival times given the event counts. Stochastic expectation maximization

Received March 2020; revised October 2021.
MSC2020 subject classifications. Primary 60G55; secondary 37A25, 62M15.
Key words and phrases. Count data, Hawkes process, spectral estimation, strong mixing, Whittle method.

algorithms (Celeux, Chauveau and Diebolt (1995)), which approximate this conditional distribution, do not alleviate this issue since usual convergence results and simulation methods are based on likelihoods of the exponential families (Delyon, Lavielle and Moulines (1999)), which excludes Hawkes processes.

For the second strategy, Kirchner (2016) showed that the distribution of the bin-count sequence of the Hawkes process can be approximated by an INAR(∞) sequence and proposed a nonparametric estimation method for the Hawkes process. In particular, the conditional least-square estimates of the INAR process yields consistent and asymptotically normal estimates for the underlying Hawkes process when the bin size tends to zero (Kirchner (2017)). Unfortunately, while this method is adapted when the bin size can be chosen arbitrarily small, for example, when the data are collected continuously (seismology, finance, etc.), it is biased when the data has been collected with large bin size or when the events cannot precisely be located in time (Kirchner (2017)), as is often the case for biological, ecological and health data sets. In particular, when the bin size is larger than the typical range of the reproduction function, this strategy is not satisfactory, since the INAR model ignores the interaction within bins.

In this article, as in Kirchner (2017), we adapt an existing time series estimation method to the case of bin-count sequences from Hawkes processes. Following Adamopoulos (1976), we use the Bartlett spectrum of the Hawkes process (i.e., the spectral density of the covariance measure of the process) to define as an estimator the minimizer of the log-spectral likelihood, first introduced by Whittle (1952). To establish the asymptotic properties of the Whittle estimator, we look at strong mixing properties for the Hawkes processes.

Rosenblatt (1956) introduced the strong mixing coefficient to measure the dependence between σ-algebras, which sparked decades of interest in the theory of weak dependence for time series and random fields (see Bradley (2005) for a review of mixing conditions). The mixing conditions provide very strong inequalities and coupling methods (Doukhan (1994), Rio (2017)) to achieve proofs of asymptotic properties for parameter estimates, provided that the mixing coefficients decrease fast enough. However, these coefficients are formulated with respect to rich σ-algebras and, therefore, difficult to bound even for very simple models.

Westcott (1972) extended the definition of mixing to point processes proving, for example, that cluster Poisson point processes are mixing in the ergodic sense (Westcott (1971)). Yet, without precise information on strong mixing coefficients, the weak dependence framework did not lead to much statistical development in the modeling of point processes. Recent works addressed the computation of strong mixing coefficients for some classes of point processes (Heinrich and Pawlas (2013), Poinas, Delyon and Lavancier (2019)), building on the results for time series and random fields and using the fact that the σ-algebras generated by countable sets are poorer than those generated by continuous sets.

For practical reason, the absolute regularity mixing coefficients are often preferred since they can be easily computed for Markov processes and functions thereof (Davydov (1974)). Notably, Hawkes processes with exponential reproduction function are piecewise deterministic Markovian processes (Oakes (1975)), and one would hope to compute absolute regularity mixing coefficients. However, since this would not extend to other reproduction functions, we instead establish a strong mixing condition with polynomial decay rate, which holds for any reproduction function, provided it has a finite moment of order $1 + \delta$, $\delta > 0$. In turn, this proves that our proposed estimation method leads to consistent and asymptotically normal estimators for the parameters of Hawkes processes from bin-count data.

Section 2 recalls definitions and sets notation used in the paper. Section 3 contains our first important result: we establish strong mixing properties for the Hawkes process and its bin-count sequences. Using the cluster and positive association properties, we relate the strong mixing coefficients to those of a single time-continuous Galton–Watson tree, then control the covariance between arrival times using results from elementary Galton–Watson theory. In Section 4, we focus on the estimation of Hawkes processes from bin-count data. We derive the spectral density function of the bin-count sequence, taking into account the aliasing caused by sampling the process in discrete time. Then, using the strong mixing condition and the work of Dzhaparidze (1986) on Whittle's method, we propose a consistent and asymptotically normal estimator to the parameters of the Hawkes process. Sections 5 and 6 provide respectively some numerical experiments and a real-life case study to illustrate the results of the two preceding sections. Finally, in Section 7, we discuss some of the appealing features and extensions of this approach. The code used in the paper, both for the simulation and the case study, is publicly available (https://github.com/fcheysson/code-spectral-hawkes).

2. The Hawkes process and its count process.

2.1. *Notation.* In this paper, we consider *simple locally finite point processes* on the measure space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \ell)$, where $\mathcal{B}(A)$ denotes the Borel σ-algebra of A and ℓ the Lebesgue measure. A point process N on $\mathbb{R}$ may be defined as a measurable map from a probability space $(\mathcal{X}, \mathcal{F}, \mathbb{P})$ to the measurable space $(\mathfrak{N}, \mathcal{N})$ of locally finite counting measures on $\mathbb{R}$. The corresponding random set of points, that is, the set of the atoms of N, is denoted $\{T_i\}$. For a function f on $\mathbb{R}$, we write

$$N(f) := \int_{\mathbb{R}} f(t)N(\mathrm{d}t) = \sum_i f(T_i)$$

the integral of f with respect to N. Finally, for a Borel set A, the cylindrical σ-algebra $\mathcal{E}(A)$ generated by N on A is defined by

$$\mathcal{E}(A) := \sigma\big(\{N \in \mathfrak{N} : N(B) = m\}, B \in \mathcal{B}(A), m \in \mathbb{N}\big).$$

2.2. *The stationary linear Hawkes process.* A stationary self-exciting point process, or *Hawkes process*, on the real line $\mathbb{R}$ is a point process N with conditional intensity function

$$\begin{aligned}\lambda(t) &= \eta + \int_0^t h(t-u)N(\mathrm{d}u)\\ &= \eta + \sum_{T_i<t} h(t-T_i)\end{aligned}$$

for $t \in \mathbb{R}$. The constant $\eta > 0$ is called the *immigration intensity* and the measurable function $h : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ the *reproduction function*. The reproduction function can be further decomposed as $h = \mu h^*$, where $\mu = \int_{\mathbb{R}} h(t)\,\mathrm{d}t < 1$ is called the *reproduction mean* and h^* is a true density function, $\int_{\mathbb{R}} h^*(t)\,\mathrm{d}t = 1$, called the *reproduction kernel*.

Moreover, the linear Hawkes process is a specific case of the Poisson *cluster process* (Hawkes and Oakes (1974)). Briefly, the process consists of a stream of *immigrants*, the cluster centers, which arrive according to a Poisson process N_c with intensity measure η. Then an immigrant at time T_i generates *offsprings* according to an inhomogenous Poisson process $N_1(\cdot|T_i)$ with intensity measure $h(\cdot - T_i)$. These in turn independently generate further offsprings according to the same law, and so on ad infinitum. The *branching processes* $N(\cdot|T_i)$, consisting of an immigrant at time T_i and all their *descendants*, are therefore independent. Finally, the Hawkes process N is defined as the superposition of all branching processes:

$$\forall A \in \mathcal{B}(\mathbb{R}), \quad N(A) = N_c\big(N(A|\cdot)\big).$$

This cluster representation links to the usual Galton–Watson theory. Without loss of generality, consider one branching process whose immigrant has time 0. Define Z_k as the number of points of generation k, that is, $Z_0 = 1$ for the immigrant, then Z_1 denotes the number of offsprings that the immigrant generates, Z_2 the number of offsprings that the offsprings of the immigrants generate, etc. Then $(Z_k)_{k\in\mathbb{N}}$ is a Galton–Watson process.

In particular, $(Z_{k+1}|Z_k = z)(k, z \in \mathbb{N})$ follows a Poisson distribution with parameter $z\mu$. Then, by the usual Galton–Watson theory, a sufficient condition for the existence of the Hawkes process is $\mu < 1$, which ensures that the total number of descendants of any immigrant is finite with probability 1 and has finite mean. This condition also ensures that the process is strictly stationary.

2.3. *Count processes.* We are interested in the time series generated by the event counts of the Hawkes process, that is, the series obtained by counting the events of the process on intervals of fixed length. We give a definition for both time-continuous and discrete time bin-count processes, according to whether the interval endpoints live on the real line or on a regular grid, respectively.

DEFINITION 1. The bin-count process with bin size Δ associated to a point process N is the process $(X_t)_{t\in\mathbb{R}} = \{N((t\Delta, (t+1)\Delta])\}_{t\in\mathbb{R}}$ generated by the count measure on intervals of size Δ. The restriction of the bin-count process on $\mathbb{Z}$, $(X_k)_{k\in\mathbb{Z}}$, is called the bin-count sequence associated to N.

3. Strong mixing properties.

Here, we control the strong mixing coefficients of Hawkes processes and their associated bin-count processes. We recall that, for a probability space $(\mathcal{X}, \mathcal{F}, \mathbb{P})$ and $\mathcal{A}, \mathcal{B}$ two sub σ-algebras of $\mathcal{F}$, Rosenblatt's strong mixing coefficient is defined as the measure of dependence between $\mathcal{A}$ and $\mathcal{B}$ (Rosenblatt (1956)):

$$\alpha(\mathcal{A}, \mathcal{B}) := \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{A}, B \in \mathcal{B}\}.$$

This definition can be adapted to a point process N on $\mathbb{R}$, by defining (see Poinas, Delyon and Lavancier (2019))

$$\alpha_N(r) := \sup_{t\in\mathbb{R}} \alpha(\mathcal{E}_{-\infty}^{t}, \mathcal{E}_{t+r}^{\infty}),$$

where $\mathcal{E}_a^b$ stands for $\mathcal{E}((a, b])$, that is, the σ-algebra generated by the cylinder sets on the interval $(a, b]$, and $\mathcal{E}_a^\infty = \sigma(\bigcup_{b>a} \mathcal{E}_a^b)$. For the corresponding sequence $(X_k)_{k\in\mathbb{Z}}$, the strong mixing coefficient takes the form

$$\alpha_X(r) := \sup_{n\in\mathbb{Z}} \alpha(\mathcal{F}_{-\infty}^{n}, \mathcal{F}_{n+r}^{\infty}),$$

where $\mathcal{F}_a^b$ stands for the σ-algebra generated by $\{X_k : k \in \mathbb{Z}, a \le k \le b\}$.

The point process N (resp., the sequence (X_k)) is said to be strongly mixing if $\alpha_N(r)$ (resp. $\alpha_X(r)$) $\to 0$ as $r \to \infty$. Intuitively, the strong mixing condition conveys that the dependence between past and future events decreases uniformly to zero as the time gap between them increases. Note that, since $\mathcal{F}_a^b \subset \mathcal{E}((a\Delta, (b+1)\Delta])$, we have that $\alpha_X(r) \le \alpha_N((r-1)\Delta)$ for all $r > 1$.

We here state the first important result of this article.

THEOREM 1. *Let N be a stationary Hawkes process on $\mathbb{R}$ with reproduction function $h = \mu h^*$, where $\mu = \int_{\mathbb{R}} h(t)\,dt < 1$ and $\int_{\mathbb{R}} h^*(t)\,dt = 1$. Suppose that there exists $\delta > 0$ such that the reproduction kernel h^* has a finite moment of order $1 + \delta$:*

$$\nu_{1+\delta} := \int_{\mathbb{R}} t^{1+\delta} h^*(t)\,dt < \infty.$$

Then N is strongly mixing and

$$\alpha_N(r) = \mathcal{O}(r^{-\delta}).$$

Furthermore, if h^* admits finite exponential moments, that is, there exists $a_0 > 0$ such that

$$\int_{\mathbb{R}} e^{a_0|t|} h^*(t)\,\mathrm{d}t < \infty,$$

then there exists $a \in (0, a_0]$ such that

$$\alpha_N(r) = \mathcal{O}(e^{-ar}).$$

Notably, this theorem can be extended in the case where the immigration intensity η is allowed to vary with respect to time.

COROLLARY 1. *Let N be a Hawkes process on $\mathbb{R}$ with reproduction function as in Theorem 1, and with nonconstant immigration intensity $\eta : t \mapsto \eta(t)$. Suppose that there exists $\delta > 0$ such that the reproduction kernel h^* has a finite moment of order $1 + \delta$, and that $\eta(\cdot)$ is nonnegative and bounded. Then N is strongly mixing and*

$$\alpha_N(r) = \mathcal{O}(r^{-\delta}).$$

Furthermore, if h^* admits finite exponential moments, then there exists $a > 0$ such that

$$\alpha_N(r) = \mathcal{O}(e^{-ar}).$$

In brief, the proof has two parts: first, we rescale the problem to a single continuous-time Galton–Watson tree using the cluster representation of the Hawkes process; second, we derive an upper bound for the strong mixing coefficients of the tree. The idea for the latter is that since the Galton–Watson process goes extinct almost surely and the reproduction kernel h^* has a finite moment, then the probability that there exists an offspring of generation k at a far distance from the immigrant goes quickly to 0 when k increases. We refer to Appendix A for the detailed proof of the theorem.

Finally, as an immediate consequence of Theorem 1, we get the following corollary for Hawkes bin-count process.

COROLLARY 2. *Let N be a Hawkes process as in Corollary 1, and $(X_k)_{k\in\mathbb{Z}} = \{N((k\Delta, (k+1)\Delta])\}_{k\in\mathbb{Z}}$ its associated bin-count sequence. Then (X_k) is strongly mixing and*

$$\alpha_X(r) = \mathcal{O}(r^{-\delta}). \tag{1}$$

Furthermore, if h^* admits finite exponential moments, then there exists $a > 0$ such that

$$\alpha_X(r) = \mathcal{O}(e^{-a\Delta r}).$$

4. Parametric estimation of bin-count sequences. In this section, we apply the strong mixing properties of the Hawkes bin-count sequence to parametric estimation using a spectral approach. First, we derive the spectral density function for both the time-continuous and discrete time Hawkes bin-count processes. Then using Whittle's method, we define a parametric estimator of a Hawkes process from its bin-count data.

4.1. *Spectral analysis*. We recall that the *Bartlett spectrum* of a second-order stationary point process N on $\mathbb{R}$ is defined as the unique, nonnegative, symmetric measure Γ on the Borel sets such that, for any rapidly decaying function φ on $\mathbb{R}$ (see Daley and Vere-Jones (2003), Proposition 8.2.I, equation (8.2.2))

$$\mathrm{Var}\big(N(\varphi)\big) = \int_{\mathbb{R}} \big|\widetilde{\varphi}(\omega)\big|^2 \Gamma(\mathrm{d}\omega), \tag{2}$$

where $\widetilde{\cdot}$ denotes the Fourier transform

$$\widetilde{\varphi}(\omega) = \int_{\mathbb{R}} \varphi(s) e^{-i\omega s}\,\mathrm{d}s.$$

By polarising relation (2), we get, for any rapidly decaying functions φ and ψ on $\mathbb{R}$:

$$\mathrm{Cov}\big(N(\varphi), N(\psi)\big) = \int_{\mathbb{R}} \widetilde{\varphi}(\omega)\widetilde{\psi^*}(\omega)\Gamma(\mathrm{d}\omega), \tag{3}$$

where $\psi^*(u) = \psi(-u)$, so that $\widetilde{\psi^*}$ is the complex conjugate of $\widetilde{\psi}$.

For the stationary Hawkes process, the Bartlett spectrum admits a density given by (see Daley and Vere-Jones (2003), Example 8.2(e))

$$\gamma(\omega) = \frac{m}{2\pi}\big|1 - \widetilde{h}(\omega)\big|^{-2}, \tag{4}$$

where $m := \mathbb{E}[N(0,1]] = \eta(1 - \int_{\mathbb{R}} h(t)\,\mathrm{d}t)^{-1}$.

For a time-continuous process, the spectral density f_c forms a Fourier pair with the autocovariance function r_c:

$$f_c(\omega) = \int_{\mathbb{R}} r_c(u) e^{-i\omega u}\,\mathrm{d}u, \qquad r_c(u) = \frac{1}{2\pi}\int_{\mathbb{R}} f_c(\omega) e^{i\omega u}\,\mathrm{d}\omega.$$

Using the second relation with the Bartlett spectrum of a stationary Hawkes process, we derive the spectral density of the time-continuous bin-count process with bin size Δ.

PROPOSITION 1. *Let N be a stationary Hawkes process on $\mathbb{R}$, and $\{X_t\}_{t\in\mathbb{R}} = \{N(t\Delta, (t+1)\Delta]\}_{t\in\mathbb{R}}$ the associated bin-count process. Then X_t has a spectral density function given by*

$$f_c(\omega) = m\Delta\,\mathrm{sinc}^2\left(\frac{\omega}{2}\right)\left|1 - \widetilde{h}\left(\frac{\omega}{\Delta}\right)\right|^{-2}. \tag{5}$$

PROOF. Let $\varphi = \mathbb{1}_{(0,\Delta]}$ and $\psi = \mathbb{1}_{(\Delta u, \Delta(u+1)]}$. We have

$$\widetilde{\varphi}(\omega) = \int_0^{\Delta} e^{-i\omega s}\,\mathrm{d}s = \frac{i}{\omega}\big[e^{-i\omega\Delta} - 1\big],$$

$$\widetilde{\psi^*}(\omega) = \int_{-\Delta(u+1)}^{-\Delta u} e^{-i\omega s}\,\mathrm{d}s = \frac{i}{\omega} e^{i\omega\Delta u}\big[1 - e^{i\omega\Delta}\big].$$

Then, using (3) and (4), the autocovariance function of X_t is

$$\begin{aligned} r_c(u) &= \mathrm{Cov}(X_0, X_u) \\ &= \mathrm{Cov}\big(N(\varphi), N(\psi)\big) \\ &= \int_{\mathbb{R}} \frac{1}{\omega^2} e^{i\omega\Delta u}\big|e^{i\omega\Delta} - 1\big|^2 \Gamma(\mathrm{d}\omega) \\ &= \frac{1}{2\pi}\int_{\mathbb{R}} m\Delta\,\mathrm{sinc}^2\left(\frac{\omega}{2}\right)\left|1 - \widetilde{h}\left(\frac{\omega}{\Delta}\right)\right|^{-2} e^{i\omega u}\,\mathrm{d}\omega. \end{aligned}$$

□

For a discrete-time process, the spectral density f_d once again forms a Fourier pair with the autocovariance function r_d:

$$f_d(\omega) = \sum_{k\in\mathbb{Z}} r_d(u)e^{-i\omega u}, \qquad r_d(u) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f_d(\omega)e^{i\omega u}\,\mathrm{d}\omega.$$

It turns out that the spectral density f_d of a time-continuous process sampled in discrete time can be related to the density f_c of the process, by taking into account spectral aliasing, which folds high frequencies back onto the apparent spectrum:

$$\begin{aligned} r_c(u) &= \frac{1}{2\pi}\int_{\mathbb{R}} f_c(\omega)e^{i\omega u}\,\mathrm{d}\omega \\ &= \frac{1}{2\pi}\sum_{k\in\mathbb{Z}}\int_{(2k-1)\pi}^{(2k+1)\pi} f_c(\omega)e^{i\omega u}\,\mathrm{d}\omega \\ &= \frac{1}{2\pi}\sum_{k\in\mathbb{Z}}\int_{-\pi}^{\pi} e^{i2k\pi u} f_c(\omega+2k\pi)e^{i\omega u}\,\mathrm{d}\omega \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{k\in\mathbb{Z}} f_c(\omega+2k\pi)e^{i\omega u}\,\mathrm{d}\omega, \end{aligned}$$

where the last equality follows when $u \in \mathbb{Z}$ and from an application of Fubini's theorem.

For the bin-count sequence associated with a stationary Hawkes process, this leads to the following corollary.

COROLLARY 3. *Let N be a stationary Hawkes process on $\mathbb{R}$, and $(X_k)_{k\in\mathbb{Z}} = \{N((k\Delta, (k+1)\Delta])\}_{k\in\mathbb{Z}}$ the associated bin-count sequence. Then X_k has a spectral density function given by*

$$f_d(\omega) = \sum_{k\in\mathbb{Z}} f_c(\omega+2k\pi),$$

where $f_c(\cdot)$ is the function defined in (5).

4.2. *Whittle estimation.* For a stationary linear process $(X_k)_{k\in\mathbb{Z}}$ with spectral density $f_\theta(\cdot)$, θ an unknown parameter vector, both Hosoya (1974) and Dzhaparidze (1974), building on the cornerstone laid by Whittle (1952), proposed as an estimator of θ the minimizer

$$\widehat{\theta}_n = \arg\min_{\theta\in\Theta}\mathcal{L}_n(\theta), \tag{6}$$

where

$$\mathcal{L}_n(\theta) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\log f_\theta(\omega) + \frac{I_n(\omega)}{f_\theta(\omega)}\right)\mathrm{d}\omega \tag{7}$$

is the log-spectral likelihood of the process, and $I_n(\omega) = (2\pi n)^{-1}|\sum_{k=1}^{n} X_k e^{-ik\omega}|^2$ is the periodogram of the partial realization $(X_k)_{1\le k\le n}$. They also gave the asymptotic properties of the estimator under appropriate regularity conditions.

Dzhaparidze (1986) extended these results to more general cases, and in particular to stationary processes verifying Rosenblatt's mixing conditions. The following conditions and theorems are thus adaptations of those found in Dzhaparidze (1986), Theorem II.7.1 and II.7.2, for stationary Hawkes bin-count sequences.

THEOREM 2. *Let N be a Hawkes process on $\mathbb{R}$ with reproduction function $h = \mu h^*$, where $\mu = \int_{\mathbb{R}} h(t)\,\mathrm{d}t < 1$ and $\int_{\mathbb{R}} h^*(t)\,\mathrm{d}t = 1$, and $(X_k)_{k\in\mathbb{Z}} = (N(k\Delta, (k+1)\Delta])_{k\in\mathbb{Z}}$ its associated bin-count sequences with spectral density function f_θ. Assume the following regularity conditions on f_θ:*

(A1) *The true parameter* θ_0 *belongs to a compact set* Θ *of* $\mathbb{R}^p$.

(A2) *For all* $\theta_1 \neq \theta_2$ *in* Θ, *then* $f_{\theta_1} \neq f_{\theta_2}$ *for almost all* ω.

(A3) *The function* f_θ^{-1} *is differentiable with respect to* θ *and its derivatives* $(\partial/\partial\theta_k) f_\theta^{-1}$ *are continuous in* $\theta \in \Theta$ *and* $-\pi \leq \omega \leq \pi$.

Further assume that there exists a $\delta > 0$ *such that the reproduction kernel* h^* *has a finite moment of order* $2 + \delta$. *Then the estimator* $\widehat{\theta}_n$ *defined as in* (6) (*with* $\mathcal{L}_n(\theta)$ *given by* (7)), *is consistent*, *that is*, $\widehat{\theta}_n \to \theta_0$ *in probability.*

PROOF. The only condition from Dzhaparidze (1986), Theorem II.7.1, that we need to verify is that there exists a $\gamma > 2$ such that $\mathbb{E}[|X_k|^{2\gamma}]$ is finite and the following inequality holds:

$$\sum_{r=1}^{\infty} (\alpha_X(r))^{1-2/\gamma} < \infty. \tag{8}$$

Since the stationary Hawkes process admit finite exponential moments if h^* has a moment of order $\delta \in (0, 1]$ (Roueff, von Sachs and Sansonnet (2016), Theorem 4), $\mathbb{E}[|X_k|^{2\gamma}]$ is finite for any γ. Then using equation (1) from Corollary 2 there always exists a $\gamma > 2$ that satisfies (8). □

Define the matrix Γ_θ, which would actually be the limit as $n \to \infty$ of the Fisher's information matrix if the process (X_k) were Gaussian (Dzhaparidze (1986), Section II.2.2), by the relation:

$$\Gamma_\theta = \left(\frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial \theta_k} \log f_\theta(\omega) \frac{\partial}{\partial \theta_l} \log f_\theta(\omega) \, \mathrm{d}\omega \right)_{1 \leq k,l \leq p}.$$

Since (X_k) is not Gaussian, the asymptotic properties of the Whittle estimator depends on the fourth-order statistics of the process and we define the following matrix:

$$C_{4,\theta} = \left(\frac{1}{8\pi} \int \int_{-\pi}^{\pi} f_{4,\theta}(\omega_1, -\omega_1, -\omega_2) \frac{\partial}{\partial \theta_k} \frac{1}{f_\theta(\omega_1)} \frac{\partial}{\partial \theta_l} \frac{1}{f_\theta(\omega_2)} \, \mathrm{d}\omega_1 \, \mathrm{d}\omega_2 \right)_{1 \leq k,l \leq p},$$

where $f_{4,\theta}(\cdot, \cdot, \cdot)$ is the fourth-order cumulant spectral density of the process. We have the following result.

THEOREM 3. *Let* N *be a Hawkes process as in Theorem* 2, *and* $(X_k)_{k \in \mathbb{Z}} = (N(k\Delta, (k + 1)\Delta])_{k \in \mathbb{Z}}$ *its associated bin-count sequences with spectral density function* f_θ. *Assume conditions* (A1), (A2), (A3) *and*:

(A4) *The function* f_θ *is twice differentiable with respect to* θ *and its second derivatives* $(\partial^2/\partial\theta_k\partial\theta_l) f_\theta$ *are continuous in* $\theta \in \Theta$ *and* $-\pi \leq \omega \leq \pi$.

Then the estimator $\widehat{\theta}_n$ *is asymptotically normal and*

$$n^{1/2}(\widehat{\theta}_n - \theta_0) \underset{n \to \infty}{\sim} \mathcal{N}(0, \Gamma_{\theta_0}^{-1} + \Gamma_{\theta_0}^{-1} C_{4,\theta_0} \Gamma_{\theta_0}^{-1}).$$

REMARK. The computation of the integral of the fourth-order cumulant spectra in C_{4,θ_0} is not straightforward. We refer to the work of Shao (2010) for an elegant way to compute an estimate of this integral.

5. Simulation study. We illustrate the estimation procedure and asymptotic properties of the spectral approach for Hawkes bin-count sequences. To highlight the different theorems of the previous sections, we consider two kernels h^* for the reproduction function: the exponential kernel for which all moments exist and the power law kernel whose higher moments are not finite.

The following simulations and estimations have been implemented with our package *hawkesbow*, freely available online (https://cran.r-project.org/web/packages/hawkesbow/index.html), written in both R (R Core Team (2019)) and C++ using Rcpp (Eddelbuettel and François (2011)).

5.1. *Simulation procedure.*

5.1.1. *Exponential kernel.* We first consider a stationary Hawkes process with exponentially decaying reproduction function:

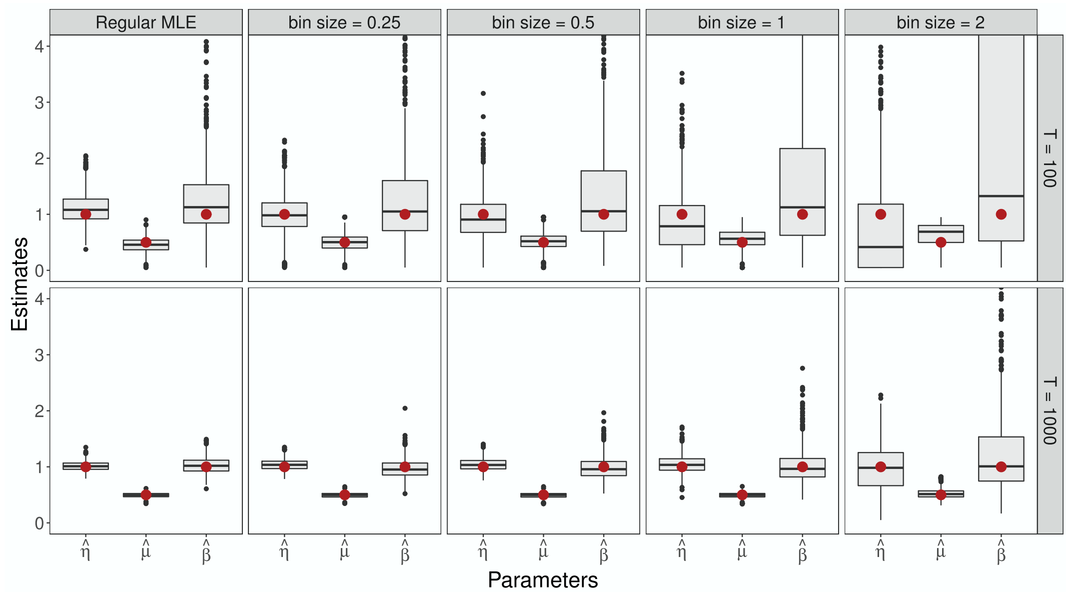
$$\lambda(t) = \eta + \mu \int_{-\infty}^{t} \beta e^{-\beta(t-u)} N(\mathrm{d}u), \quad t \in \mathbb{R},$$

that is, with reproduction kernel $h^*(t) = \beta e^{-\beta t}$ for $t \geq 0$. Note that the process verifies the conditions of both Theorems 2 and 3.

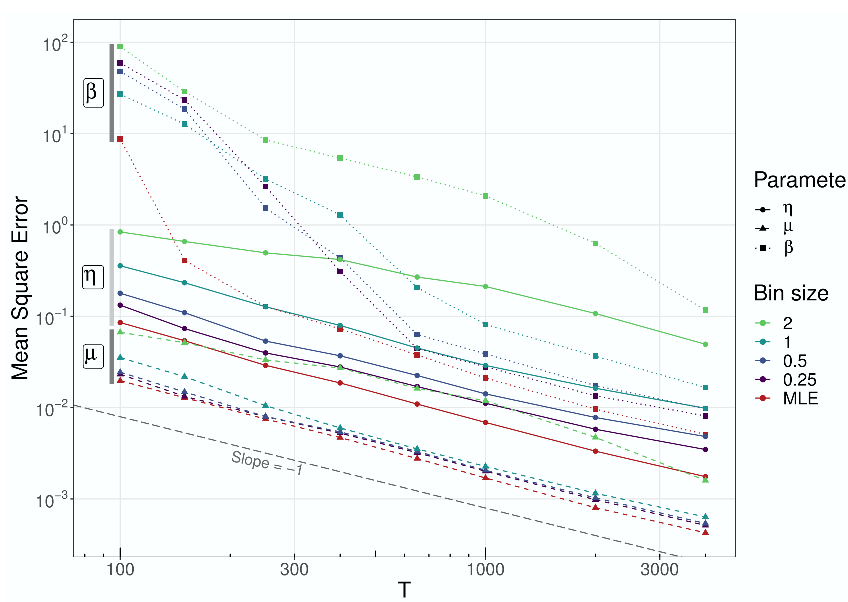
Using the cluster representation of the Hawkes process, we simulated 1000 realizations of the Hawkes process on the interval $[0, T]$ with parameter values $\eta = 1$, $\mu = 0.5$ and $\beta = 1$ and a burn-in interval $[-100, 0]$. For each of the simulations, we created four time series by counting the events in bins of size $\Delta = 0.25$, 0.5, 1 or 2, respectively. We then estimated the parameters η, μ and β as in Section 4.2 for each of the four time series. We compared these estimates to the usual maximum likelihood estimates (Figure 1(a)), for which it is assumed no event lies outside of $[0, T]$. Since the latter use the full information on the location of events, they are arguably better that any estimate based on the bin-count sequences, and provide a best case scenario for the Whittle estimates when the bin size tends to 0. Minimization of the log-spectral likelihood, respectively, maximization of the likelihood, was carried out using a limited-memory BFGS optimisation algorithm with bound constraints (Liu and Nocedal (1989), Nocedal (1980))—$0 < \mu < 1$ and $\eta, \beta > 0$—available in function optim from R, respectively, in function nloptr from package NLOPTR (Johnson). With an exponential kernel, a set of 1000 simulations and their Whittle estimation with $T = 1000$ and bin size $\Delta = 1$ takes approximately 2 minutes on a laptop computer with an i5-6300HQ Intel CPU.

5.1.2. *Power law kernel.* We now consider a stationary Hawkes process with a power law reproduction kernel: $h_\gamma^*(t) = \gamma a^\gamma (a + t)^{-\gamma-1}$ for $t \geq 0$. We recall that the moments of a power law distribution are all finite up to, but not including, the order γ. We illustrate the theorems of the previous sections by considering three cases for the shape, with each increasingly satisfying the necessary assumptions: (i) $\gamma = 0.5$, the process does not satisfy the condition of Theorem 1; (ii) $\gamma = 1.5$, the process is strongly mixing and satisfies the condition of Theorem 1, but not the assumptions of Theorem 2; (iii) $\gamma = 2.5$, the process is strongly mixing and satisfies the assumptions of Theorems 1, 2 and 3.

As for the exponential kernel, we simulated 1000 realizations of the Hawkes process for each $\gamma \in \{0.5, 1.5, 2.5\}$, with parameter values $\eta = 1$, $\mu = 0.5$, and scale parameter $a = 1.5$ chosen such that the power law kernel $h_{2.5}^*$ and the exponential kernel have the same first-order moment. For the power law kernels $h_{1.5}^*$ and $h_{0.5}^*$, we kept the scale parameter the same so that the simulations can be compared with the kernel $h_{2.5}^*$. As for the exponential kernel, the Whittle estimates of η, μ and γ were compared to the usual maximum likelihood estimates, with the scale parameter kept fixed to its true value $a = 1.5$. Estimation figures can



(a) Estimates of parameters η, μ and β for 1,000 simulations on the interval $[0,T]$. True values (larger dots) are: $\eta = 1$, $\mu = 0.5$, $\beta = 1$. The left column refers to the maximum likelihood estimates. The other columns refer to the Whittle estimates according to different bin sizes.



(b) Mean square error of the estimates of parameters η, μ and β for 1,000 simulations on the interval $[0,T]$, in log-log scale. The dashed grey line represents the ideal slope of -1, that is, a rate of convergence of $\mathcal{O}(n^{-1})$.

FIG. 1. *Performance of the Whittle estimates for the stationary Hawkes process with immigration intensity* $\eta = 1$, *reproduction mean* $\mu = 0.5$ *and reproduction kernel* $h^*(t) = \beta e^{-\beta t}$, *where* $\beta = 1$.

be found in Appendix B. The constraint bounds for the optimisation routines were $0 < \mu < 1$ and $\eta, \gamma > 0$. With a power law kernel, a set of 1000 simulations and their Whittle estimation with $T = 1000$ and bin size $\Delta = 1$ takes approximately 14 minutes on a laptop computer with an i5-6300HQ Intel CPU.

5.2. *Results and interpretation.*

5.2.1. *Exponential kernel.* For $T = 100$ and small bin sizes, the Whittle estimates fare almost as well as the maximum likelihood estimates (see Figure 1(a)). The estimation deteriorates massively for higher bin sizes, notably for the exponential kernel rate β. This is intuitive, since large bin sizes with respect to the kernel scale make it difficult to detect interactions between points. This can be related to the probability that an offspring and its parent belong in the same bin: assuming the stationarity of the process, this probability is equal to $\Delta^{-1} \int_0^\Delta \int_u^\Delta \beta e^{-\beta(t-u)}\,\mathrm{d}t\,\mathrm{d}u = 1 - (\beta\Delta)^{-1}(1 - e^{-\beta\Delta})$. For example, with $\beta = 1$ and $\Delta = 2$, we get a probability of 0.57, that is, 57% of the information about the interaction of the Hawkes process is located within bins, with only 43% remaining between bins. Thankfully, by increasing T, the asymptotic properties ensure that the Whittle estimates improve, even for large bin sizes.

To further illustrate the asymptotic properties of the estimation, notably its rate of convergence, we compute the mean square error, defined by $\mathrm{MSE} = S^{-1} \sum(\widehat{\theta}_n - \theta_0)^2$, for the estimates of each set of $S = 1000$ simulations at given Ts and bin sizes (Figure 1(b)). For large Ts, the slope of the mean square error with respect to T reaches -1 (in log-log scale) for all parameters and almost all bin sizes, illustrating the $\mathcal{O}(n^{-1})$ rate of convergence stated in Theorem 3. For small Ts and both the Whittle and the maximum likelihood estimation methods, the estimates of the immigration intensity η and reproduction mean μ have already reached the optimal rate of convergence, while the MSE for the exponential kernel rate β is up to one and a half orders of magnitude higher than what would be expected by extrapolating the MSE for large Ts. Finally note that, for reasonable bin sizes ($\Delta \le 1$), the Whittle estimates of the reproduction mean μ have a MSE comparable to those of the maximum likelihood.

5.2.2. *Power law kernel.* Performances for the point estimates are remarkably similar for $\gamma = 2.5$ and $\gamma = 1.5$. As in the exponential case, both the immigration intensity η and the reproduction mean μ exhibit the optimal rate of convergence $\mathcal{O}(n^{-1})$ throughout all Ts considered for all bin sizes, while the shape parameter γ exhibits this asymptotic regime for sufficiently large Ts ($T \ge 400$). When $\gamma = 0.5$, the Whittle estimates for both the immigration intensity η and the reproduction mean μ do not considerably improve for the range of Ts considered, in contrast to the maximum likelihood estimates which approach the optimal asymptotic regime for large Ts. For all three kernels, the estimates show a curious behavior: for large Ts, almost all estimates for the bin size 0.25 have larger MSE than for bin sizes 0.5 and 1.

Interestingly, the point estimates exhibit good asymptotic behaviors for $\gamma = 1.5$ even though the power law kernel $h^*_{1.5}$ does not satisfy the assumptions of Theorems 2 and 3, but do not for $\gamma = 0.5$. This could suggest that the condition on the kernel moments in Theorem 1 is too restrictive, and probably lies between 0.5 and 1.5. Nevertheless, it is mild enough that the spectral approach developed in this article can be useful for applications in many disciplines.

6. Case study: Transmission of measles in Tokyo.

Measles is a highly contagious viral disease, primarily transmitted via droplets and manifesting as a febrile rash illness. Despite worldwide efforts to eradicate the disease, it has sprung back in developed countries mainly through imported cases and nonvaccinated individuals, generating minor outbreaks. As the infectious period of measles begins before symptoms are first apparent, in some cases the carriers may be diagnosed after their offsprings. For this reason, we propose to adapt the Hawkes process to reflect this apparent noncausal situation.

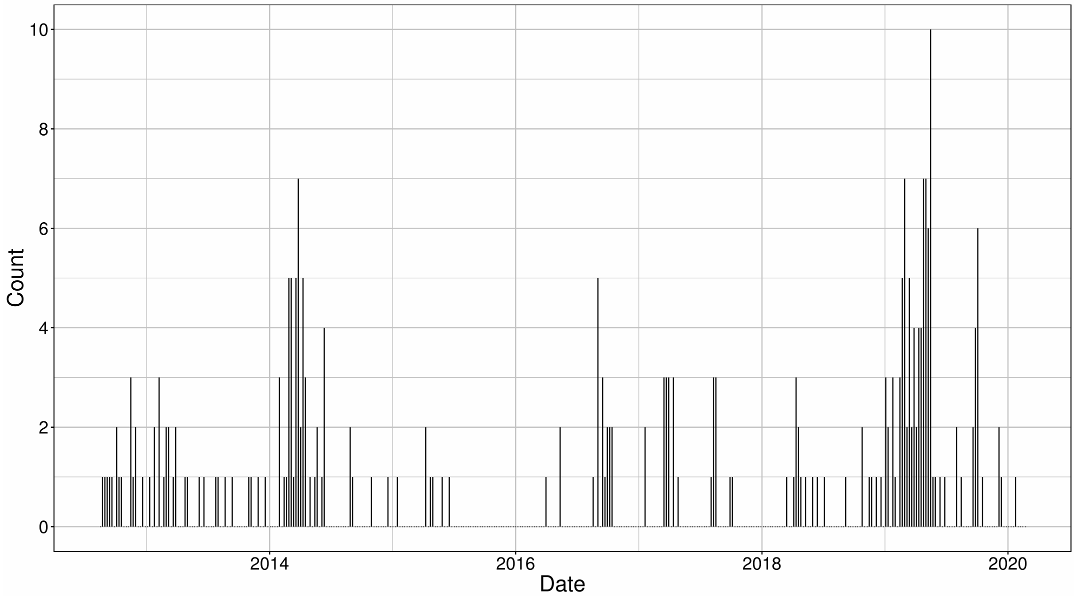


FIG. 2. *Weekly count of measles cases in Tokyo. Between the third week of August*, 2012 *and the third week of February*, 2020, 264 *cases of measles have been declared in the prefecture of Tokyo*.

6.1. *Extension to a noncausal framework.* We consider a natural extension of linear Hawkes processes by relaxing the condition that the reproduction kernel h^* has support on $\mathbb{R}_{\geq 0}$. Such a process can be defined through the cluster representation presented in Section 2.2 by allowing the offsprings to be generated in the past, that is, by allowing h^* to take positive values on $\mathbb{R}_{<0}$. We will call this process a noncausal Hawkes process, even if its conditional intensity function is intractable.

The results proved in Section 3 can be directly extended to noncausal Hawkes processes. Indeed, all proofs but those of Lemmas 7 and 8 from Appendix A remain identical. For Lemma 8, split the integral into two: one from $-\infty$ to $t + r/2$, the other from $t + r/2$ to $+\infty$. The first integral is treated as written. For the second integral, Lemma 7 can be adapted using a symmetry argument regarding the location of the immigrant and the interval considered. In consequence, the spectral estimation procedure proposed in Section 4 remains applicable, with consistent and asymptotically normal estimators.

6.2. *Estimation of the contagion function.* In Japan, measles is a notifiable disease: all diagnosed cases must be reported to the government, then investigated to contain potential outbreaks. The Japanese National Institute of Infectious Diseases publishes weekly reports as well as surveillance data tables for all notifiable diseases (https://www.niid.go.jp/niid/en/survaillance-data-table-english.html). We here consider the number of measles cases in the prefecture of Tokyo, from August 2012 to today (Figure 2). We model the weekly count data using a Hawkes process with Gaussian kernel:

$$h^*(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-\nu)^2}{2\sigma^2}\right), \quad t \in \mathbb{R},$$

where ν can be related to the incubation period and σ to the transmission period, then estimate the parameters η, μ, ν and σ as in Section 4.2. We treat the process as stationary because the impact of the seasonality was small compared to local variability.

For the Gaussian kernel, we find $\widehat{\nu} = 9.8$ days and $\widehat{\sigma} = 5.9$ days, corresponding to an interquartile range of 7.9 days. These estimates can be related to clinical features of the virus:

the incubation period of measles averages 10–12 days, while the transmission occurs usually from 4 days before to 4 days after rash onset (Centers for Disease Control and Prevention (2015)). For the immigration intensity and reproduction mean, we find $\widehat{\eta} = 0.040$ day^{-1} and $\widehat{\mu} = 0.72$. Interestingly, we find that cases with unknown source of transmission (i.e., immigrants of the model) represent $1 - \widehat{\mu} = 28\%$ of all measles cases, a figure close to the data found in (Nishiura, Mizumoto and Asai (2017), Figure 3), which reports 23 imported cases amongst 106 contagious events in Japan, 2016.

6.3. *Goodness-of-fit diagnostics.* Assessing the goodness-of-fit of a Hawkes model to the observed data is usually achieved via residual analysis (Ogata (1988)) where the residuals, which are obtained through an application of the random time change theorem (see Papangelou (1972) or Daley and Vere-Jones (2003), Theorem 7.4.IV), are expected to behave like a unitary Poisson point process. Here, since the arrival times of the process are not observed, we instead use the spectral approach to goodness-of-fit diagnostics for time series models proposed by Paparoditis (2000).

Using the notation of Section 4.2, the test is based on the distance between a kernel estimator of the normalized periodogram ordinates,

$$\widehat{q}(\omega, \widehat{\theta}) = \frac{1}{nh} \sum_{j=-m}^{m} K\left(\frac{\omega - \omega_j}{h}\right) \frac{I_n(\omega_j)}{f_{\widehat{\theta}}(\omega_j)}, \tag{9}$$

and its expected value under the null hypothesis, leading to the test statistic given by

$$S_{n,h}(\widehat{\theta}) = nh^{1/2} \int_{-\pi}^{\pi} \left(\frac{1}{nh} \sum_{j=-m}^{m} K\left(\frac{\omega - \omega_j}{h}\right)\left(\frac{I_n(\omega_j)}{f_{\widehat{\theta}}(\omega_j)} - 1\right)\right)^2 \mathrm{d}\omega,$$

where $\omega_j = 2\pi j/n$ are the studied frequencies, $m = \lfloor (n-1)/2 \rfloor$, K denotes the kernel and h the bandwidth used to smooth the rescaled periodogram ordinates. Then, under some regularity assumptions on the studied process (X_k) and the kernel K, as $n \to \infty$ and $h \sim n^{-\rho}$ for some $0 < \rho < 1$ (Paparoditis (2000), Theorem 2),

$$S_{n,h}(\widehat{\theta}) - \mu_h \to \mathcal{N}(0, \tau^2),$$

where

$$\mu_h = h^{-1/2} \int_{-\pi}^{\pi} K^2(x)\,\mathrm{d}x \quad \text{and} \quad \tau^2 = \frac{1}{\pi} \int_{-2\pi}^{2\pi} \left[\int_{-\pi}^{\pi} K(u) K(u+x)\,\mathrm{d}u \right]^2 \mathrm{d}x.$$

Then the null hypothesis, that is, that the true density function of the process lies in the postulated class of density functions,

$$\mathcal{H}_0 : f \in \mathcal{F}_\Theta = \{f_\theta, \theta \in \Theta\},$$

can be rejected at an asymptotical α-level if $S_{n,h}(\widehat{\theta}) > \mu_h + u_{1-\alpha}\tau$, where $u_{1-\alpha}$ is the $(1-\alpha)$-th quantile of the standard normal distribution.

For the measles data set, we calculated the test statistic using the Epanechnikov kernel given by $K(x) = 3(1 - (x/\pi)^2)/2$ for $|x| \leq \pi$. With this choice of kernel, $\mu_h = (12\pi/5)h^{-1/2}$ and $\tau^2 = 2672\pi^2/385$. This leads to the asymptotic p-values $p = 0.61$ and $p = 0.96$ for the bandwidths $h = 0.05$ and $h = 0.10$, respectively. Bootstrap approximation of the distribution of the test statistic under the null (Paparoditis (2000), Section 4) using 1000 replicates yields similar p-values: $p = 0.55$ and $p = 0.97$, respectively. Hence the chosen Hawkes model seems to correctly reproduce the spectral characteristics of the data (Figure 3(a)).

Additionally, it is possible to derive a goodness-of-fit diagnostic plot, which gives helpful information were the postulated model to be rejected, by looking at the asymptotic behavior of the statistic $\widehat{q}$ given in (9). Indeed, under the null (Paparoditis (2000), Section 5),

$$Q^2(\omega, \widehat{\theta}) = \frac{nh(\widehat{q}(\omega, \widehat{\theta}) - s_h(\omega))^2}{\frac{1}{2\pi}\int_{-\pi}^{\pi} K^2(u)\,\mathrm{d}u} \to \chi_1^2,$$

where $s_h(\omega) = (nh)^{-1}\sum_{j=-m}^{m} K((\omega - \omega_j)/h)$. Then a plot of the test statistic $Q^2(\cdot, \widehat{\theta})$ can be used to diagnose the frequencies at which the fit of the model must be reevaluated by comparing the values of $Q^2(\omega, \widehat{\theta})$ against the $(1-\alpha)$-th quantile of the χ_1^2 distribution (Figure 3(b)).

7. Conclusion. In this article, we establish a strong mixing condition with polynomial decay rate for stationary Hawkes processes, then propose a Whittle estimation procedure from their count data. To our knowledge, this is the first work investigating strong mixing conditions for the estimation of Hawkes processes. This approach has appealing features: (i) it has good asymptotic properties, similar to maximum likelihood estimation; (ii) it is easy to implement and flexible, since the only user-specified input is the Fourier transform $\widetilde{h}$ of the reproduction kernel h^*; (iii) it is computationally efficient, with a complexity of $\mathcal{O}(n \log n)$, n the number of bins, from calculating the periodogram with a fast Fourier transform, compared to $\mathcal{O}(p^2)$, p the number of events, for the maximum likelihood method (except when the kernel is exponential, in which case the complexity is reduced to $\mathcal{O}(p)$ with minimal efforts (Ozaki (1979)), making it more efficient than our approach); (iv) it is particularly well adapted to applications where the bin size cannot be chosen arbitrarily, that is, the events are only counted in bins of fixed size.

APPENDIX A: PROOF OF THEOREM 1 AND COROLLARY 1

By definition, for a given Hawkes process N, we have

$$\alpha_N(r) := \sup_{t\in\mathbb{R}} \alpha\big(\mathcal{E}_{-\infty}^{t}, \mathcal{E}_{t+r}^{\infty}\big) = \sup_{t\in\mathbb{R}} \sup_{\substack{\mathcal{A}\in\mathcal{E}_{-\infty}^{t} \\ \mathcal{B}\in\mathcal{E}_{t+r}^{\infty}}} \big|\mathrm{Cov}\big(\mathbb{1}_{\mathcal{A}}(N), \mathbb{1}_{\mathcal{B}}(N)\big)\big|,$$

where $\mathbb{1}_{\mathcal{A}}(N)$ is the indicator function of the cylinder set $\mathcal{A}$, that is, for an elementary cylinder set $\mathcal{A}_{B,m} = \{N \in \mathfrak{N} : N(B) = m\}$, $\mathbb{1}_{\mathcal{A}_{B,m}}(N) = 1$ if $N(B) = m$ and 0 otherwise.

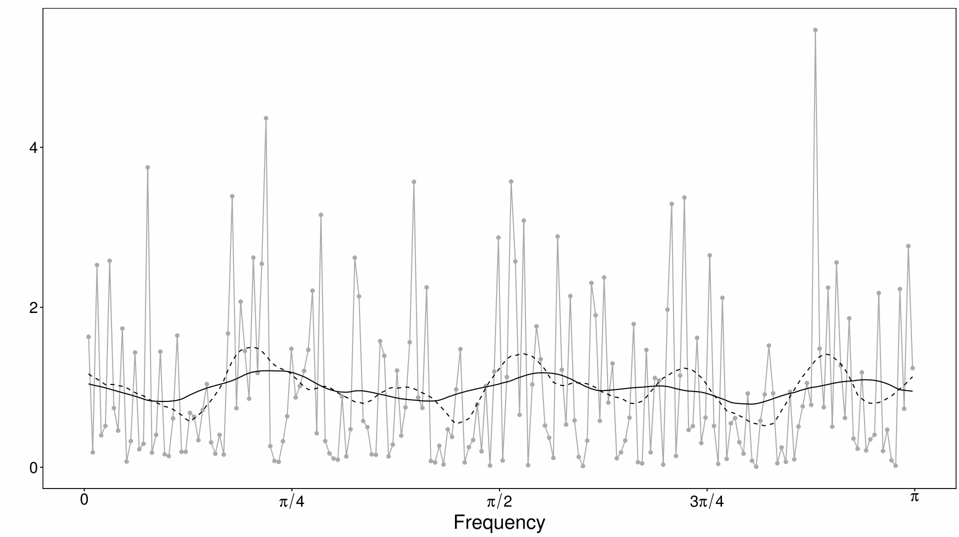
We recall that a point process N is said to be positively associated if, for all families of pairwise disjoint Borel sets $(A_i)_{1\le i\le k}$ and $(B_j)_{1\le j\le l}$, and for all coordinatewise increasing functions $F : \mathbb{N}^k \to \mathbb{R}$ and $G : \mathbb{N}^l \to \mathbb{R}$, it satisfies

$$\mathrm{Cov}\big(F\big(N(A_1), \dots, N(A_k)\big), G\big(N(B_1), \dots, N(B_l)\big)\big) \ge 0.$$

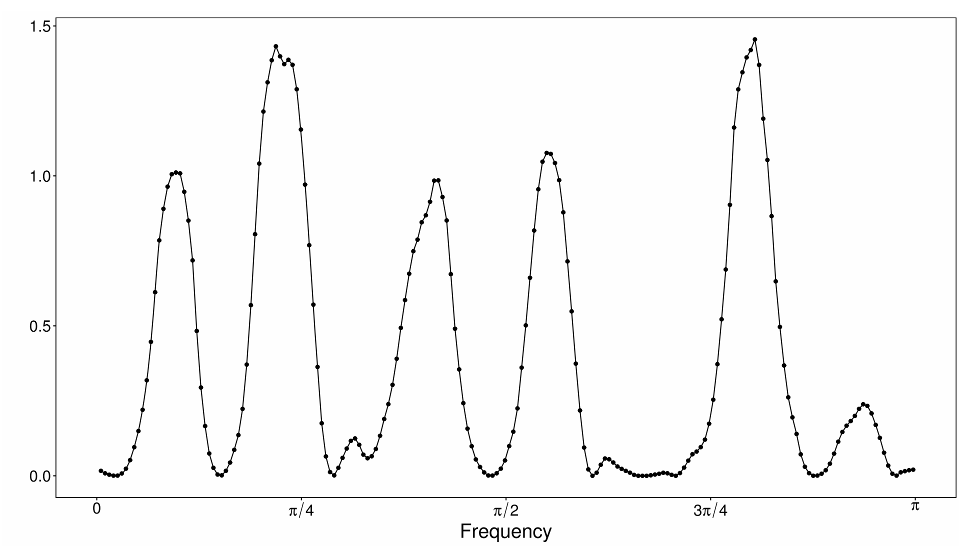
We start by stating a useful property (see Gao and Zhu (2018), Section 2.1, key property (e)), which follows from Hawkes processes being infinitely divisible processes:

PROPOSITION 2. *The Hawkes process is positively associated.*

Using this proposition and Poinas, Delyon and Lavancier's work on associated point processes (Poinas, Delyon and Lavancier (2019)), the following lemma controls the covariance of the indicator functions by the covariance of the count measure of the process, then rescale the problem to a single branching process, thanks to the independence between clusters of a Hawkes process.



(a) Plot of the normalized periodogram ordinates $I(\omega_j)/f_{\widehat{\theta}}(\omega_j)$ (grey lines) and of the kernel estimate $\widehat{q}(\omega_j, \widehat{\theta})$ with bandwidths $h = 0.05$ (black dashed curve) and $h = 0.10$ (black solid curve).



(b) Plot of the statistic $Q^2(\omega_j, \widehat{\theta})$ with bandwidth $h = 0.10$. Note that the value of the 95% quantile of the χ_1^2 distribution is 3.84, well above the drawing box of the plot.

FIG. 3. *Goodness-of-fit diagnostic plots for the measles data set.*

LEMMA 1. *Let $s, t, u \in \mathbb{R}$ and $r > 0$ such that $s < t < t + r < u$, and let $\mathcal{A} \in \mathcal{E}_s^t$, $\mathcal{B} \in \mathcal{E}_{t+r}^u$. Then,*

$$\left|\operatorname{Cov}\left(\mathbb{1}_{\mathcal{A}}(N), \mathbb{1}_{\mathcal{B}}(N)\right)\right| \leq \int \left|\operatorname{Cov}\left(N\left((s,t]|y\right), N\left((t+r,u]|y\right)\right)\right| M_c(\mathrm{d}y),$$

where $N(\cdot|y)$ denotes the branching process consisting of an immigrant at time y and all its descendants, and $M_c(\cdot)$ refers to the first-order moment of the centre process N_c.

PROOF. Using Proposition 2 and (Poinas, Delyon and Lavancier (2019), Theorem 2.5), we have

$$\left|\mathrm{Cov}\left(\mathbb{1}_{\mathcal{A}}(N), \mathbb{1}_{\mathcal{B}}(N)\right)\right| \leq \left|\mathrm{Cov}\left(N\left((s,t]\right), N\left((t+r,u]\right)\right)\right|.$$

Then, conditioning on the cluster centre process N_c (see, e.g., Daley and Vere-Jones (2003), Exercise 6.3.4):

$$\begin{aligned}
&\mathrm{Cov}\left(N\left((s,t]\right), N\left((t+r,u]\right)\right) \\
&\quad = \int \mathrm{Cov}\left(N\left((s,t]|y\right), N\left((t+r,u]|y\right)\right) M_c(\mathrm{d}y) \\
&\qquad + \int \mathbb{E}\left[N\left((s,t]|x\right)\right]\mathbb{E}\left[N\left((t+r,u]|y\right)\right] C_c(\mathrm{d}x \times \mathrm{d}y),
\end{aligned}$$

where $M_c(\cdot)$ and $C_c(\cdot)$ refer to the first-order moment measure and the covariance measure of the centre process N_c, respectively. Since the centre process is Poisson, $C_c \equiv 0$ and the second term is zero. □

We are now interested in deriving an upper bound for the covariance of counts of a typical single branching process $N(\cdot|y)$. Without loss of generality, we consider a cluster whose immigrant is located at time $y = 0$. Let Z_k denote the number of points of generation k, and by $Z_k^{(s,t]}$ those that are located in the interval $(s,t]$ $(s, t \in \mathbb{R})$. Note that, for generation 0, there is a single immigrant located at time 0. By definition, we have

$$N\left((s,t]|0\right) = \sum_{k=0}^{+\infty} Z_k^{(s,t]}.$$

Then the covariance between two intervals for a branching process is

$$\mathrm{Cov}\left(N\left((s,t]|0\right), N\left((t+r,u]|0\right)\right) = \sum_{k=0}^{+\infty}\sum_{l=0}^{+\infty} \mathrm{Cov}\left(Z_k^{(s,t]}, Z_l^{(t+r,u]}\right).$$

Before continuing further, we will need a few results on the Galton–Watson process $(Z_k)_{k\in\mathbb{N}}$:

LEMMA 2. *For $k \geq 0$, the expectation, variance and second-order moment of Z_k are*

$$\begin{aligned}
\mathbb{E}[Z_k] &= \mu^k, \\
\mathrm{Var}(Z_k) &= \mu^k \sum_{j=0}^{k-1} \mu^j = \mu^k \frac{1-\mu^k}{1-\mu}, \\
\mathbb{E}\left[Z_k^2\right] &= \mu^k \sum_{j=0}^{k} \mu^j = \mu^k \frac{1-\mu^{k+1}}{1-\mu}.
\end{aligned}$$

PROOF. Call ϕ_k the probability-generating function of Z_k:

$$\forall s \in [0,1], \quad \phi_k(s) = \mathbb{E}\left[s^{Z_k}\right].$$

It is well known for a Galton–Watson process that $(\phi_k)_{k\in\mathbb{N}}$ verifies

$$\forall k \in \mathbb{N}, \quad \phi_{k+1} = \phi_k \circ \phi_1,$$

where in our case ϕ_1 is the probability-generating function of a Poisson process with parameter μ. Differentiating the recurrence relation up to order 2 then evaluating it at $s = 1$ gives the following relations:

$$\phi'_{k+1}(1) = \phi'_1(1)\phi'_k(1),$$
$$\phi''_{k+1}(1) = \phi''_1(1)\phi'_k(1) + \big(\phi'_1(1)\big)^2\phi''_k(1),$$

where $\phi'_k(1)$ and $\phi''_k(1)$ are related to the moments of the process by

$$\mathbb{E}[Z_k] = \phi'_k(1), \qquad \operatorname{Var}(Z_k) = \phi''_k(1) + \phi'_k(1) - \big(\phi'_k(1)\big)^2.$$

Finally, plugging in the initial conditions for the Poisson variable Z_1, $\phi'_1(1) = \mu$ and $\phi''_1(1) = \mu^2$, yields the expected result. □

LEMMA 3. *For $0 \le k \le l$, the covariance and second-order product moment of (Z_k) are*

$$\operatorname{Cov}(Z_k, Z_l) = \mu^l \sum_{j=0}^{k-1} \mu^j = \mu^l \frac{1-\mu^k}{1-\mu},$$
$$\mathbb{E}[Z_k Z_l] = \mu^l \sum_{j=0}^{k} \mu^j = \mu^l \frac{1-\mu^{k+1}}{1-\mu}.$$

PROOF. This is a straightforward recurrence, noting that

$$\begin{aligned}\operatorname{Cov}(Z_k, Z_{k+h}) &= \operatorname{Cov}\left(Z_k, \sum_{i=1}^{+\infty} \mathbb{1}_{\{Z_{k+h-1}\ge i\}} Z_{1,i}\right)\\ &= \mathbb{E}[Z_{1,1}]\operatorname{Cov}\left(Z_k, \sum_{i=1}^{+\infty} \mathbb{1}_{\{Z_{k+h-1}\ge i\}}\right)\\ &= \mu \operatorname{Cov}(Z_k, Z_{k+h-1}),\end{aligned}$$

wherein $Z_{1,i}$ denotes the number of offsprings of the point i of generation $k+h-1$, is independent of $Z_{1,j}$ $(i \neq j)$, of Z_{k+h-1} and of Z_k, and has the same distribution as Z_1. □

Let T_i^k denote the time of arrival of the ith point of generation k. It has a parent T_j^{k-1} (when $k > 0$). Let Δ_i^k be the associated inter-arrival time, that is, $\Delta_i^k = T_i^k - T_j^{k-1}$. Then, for each point i of generation k, there exists a sequence $(\alpha_{i,k}^{(j)})_{1\le j\le k}$, with $\alpha_{i,k}^{(k)} = i$, denoting the indices of the ancestors of T_i^k, such that

$$T_i^k = \sum_{j=1}^{k} \Delta^j_{\alpha_{i,k}^{(j)}}.$$

As a consequence, we get the following lemma.

LEMMA 4. *For $k \in \mathbb{N}$ and $1 \le i, j \le Z_k$:*

(i) *T_i^k and T_j^k are identically distributed, with distribution function equal to the k-multiple convolution of h^* with itself,*

(ii) *For $\delta > 0$, there is an upper bound on the mth moment of T_1^k:*

$$\mathbb{E}\big[(T_1^k)^{1+\delta}\big] \le k^{1+\delta}\mathbb{E}\big[(\Delta_1^1)^{1+\delta}\big] = k^{1+\delta}\nu_{1+\delta},$$

where $\nu_{1+\delta} := \int_{\mathbb{R}} t^{1+\delta} h^*(t)\,\mathrm{d}t$.

PROOF. Statement (i) follows from the variables Δ_i^k being independent of Δ_j^l for $(i,k) \neq (j,l)$, and identically distributed with density function h^*. For statement (ii), the upper bound of the $(1+\delta)$-th order moment of T_1^k can be obtained using the following Hölder's inequality:

$$T_1^k = \sum_{j=1}^{k} \Delta^j_{\alpha^{(j)}_{1,k}} \leq \left(\sum_{j=1}^{k} 1\right)^{\frac{\delta}{1+\delta}} \left(\sum_{j=1}^{k} (\Delta^j_{\alpha^{(j)}_{1,k}})^{1+\delta}\right)^{\frac{1}{1+\delta}}. \qquad \square$$

Additionally, since for any point of the branching process offsprings are generated by a Poisson process, the arrival times, say Δ_i^k, are independent from the number of offsprings generated at the current or past generations. Conversely, since the reproduction mean μ does not depend on the time, the number of offsprings generated at any generation, say Z_l, are independent from the past arrival times. Consequently, we have the following lemma.

LEMMA 5. *For $k, l \in \mathbb{N}$ and $1 \leq i \leq Z_k$, T_i^k and Z_l are independent.*

REMARK. This lemma separates the genealogy of the Galton–Watson process (Z_k) from the arrival times (T_i^k) of the branching process, analogously to how the Poisson process is a binomial process with Poisson-distributed number of points. Then a cluster in a Hawkes process is equivalent to a Galton–Watson process (Z_k), upon which the ancestors $(\alpha_{i,k}^{(k-1)})$ are drawn equiprobably from the Z_{k-1} possible ancestors and the (Δ_i^k) independently with distribution function h^*. Intuitively, since each point j of generation $k-1$ generates offsprings according to the same intensity measure, then each point of generation k has ancestor j with equiprobability.

Finally, we will need the following identity for the covariance of the product of independent random variables.

LEMMA 6. *Let $(X_i^k)_{i,k\in\mathbb{N}}$ and $(Y_j^l)_{j,l\in\mathbb{N}}$ be two collections of random variables such that, for all $i, j, k, l \in \mathbb{N}$, the variables X_i^k and Y_j^l are independent. Then*

$$\operatorname{Cov}(X_i^k Y_i^k, X_j^l Y_j^l) = \mathbb{E}[X_i^k X_j^l] \operatorname{Cov}(Y_i^k, Y_j^l) + \mathbb{E}[Y_i^k]\mathbb{E}[Y_j^l] \operatorname{Cov}(X_i^k, X_j^l).$$

PROOF. Writing the expression of the covariance then adding and substracting the term $\mathbb{E}[X_i^k X_j^l]\mathbb{E}[Y_i^k]\mathbb{E}[Y_j^l]$ yields the relation. $\square$

We can now derive an upper bound for $\operatorname{Cov}(Z_k^{(s,t]}, Z_l^{(t+r,u]})$.

LEMMA 7. *Let $s, t, u \in \mathbb{R}$ and $r > 0$ such that $s < t < t+r < u$. Suppose that there exists $\delta > 0$ such that $\nu_{1+\delta} < \infty$. Then*

$$\left|\operatorname{Cov}(Z_k^{(s,t]}, Z_l^{(t+r,u]})\right| \leq 2 \frac{l^{1+\delta} \nu_{1+\delta}}{(t+r)^{1+\delta}} \mu^{k \vee l} \frac{1-\mu^{k\wedge l+1}}{1-\mu},$$

where $k \vee l = \max(k,l)$ and $k \wedge l = \min(k,l)$.

PROOF. We have

$$\begin{aligned}\operatorname{Cov}(Z_k^{(s,t]}, Z_l^{(t+r,u]}) &= \operatorname{Cov}\left(\sum_{i=1}^{Z_k} \mathbb{1}_{\{T_i^k \in (s,t]\}}, \sum_{j=1}^{Z_l} \mathbb{1}_{\{T_j^l \in (t+r,u]\}}\right) \\ &= \sum_{i=1}^{+\infty}\sum_{j=1}^{+\infty} \operatorname{Cov}(\mathbb{1}_{\{Z_k \geq i\}} \mathbb{1}_{\{T_i^k \in (s,t]\}}, \mathbb{1}_{\{Z_l \geq j\}} \mathbb{1}_{\{T_j^l \in (t+r,u]\}}).\end{aligned}$$

Then, by Lemmas 5 and 6,

$$\begin{aligned}
&\operatorname{Cov}(\mathbb{1}_{\{Z_k\geq i\}}\mathbb{1}_{\{T_i^k\in(s,t]\}}, \mathbb{1}_{\{Z_l\geq j\}}\mathbb{1}_{\{T_j^l\in(t+r,u]\}}) \\
&\quad = \mathbb{E}[\mathbb{1}_{\{Z_k\geq i\}}\mathbb{1}_{\{Z_l\geq j\}}]\operatorname{Cov}(\mathbb{1}_{\{T_i^k\in(s,t]\}}, \mathbb{1}_{\{T_j^l\in(t+r,u]\}}) \\
&\qquad + \mathbb{E}[\mathbb{1}_{\{T_i^k\in(s,t]\}}]\mathbb{E}[\mathbb{1}_{\{T_j^l\in(t+r,u]\}}]\operatorname{Cov}(\mathbb{1}_{\{Z_k\geq i\}}, \mathbb{1}_{\{Z_l\geq j\}}).
\end{aligned}$$

For the first term,

$$\begin{aligned}
&\operatorname{Cov}(\mathbb{1}_{\{T_i^k\in(s,t]\}}, \mathbb{1}_{\{T_j^l\in(t+r,u]\}}) \\
&\quad = \mathbb{E}[\mathbb{1}_{\{T_i^k\in(s,t]\}}\mathbb{1}_{\{T_j^l\in(t+r,u]\}}] - \mathbb{E}[\mathbb{1}_{\{T_i^k\in(s,t]\}}]\mathbb{E}[\mathbb{1}_{\{T_j^l\in(t+r,u]\}}] \\
&\quad \leq \mathbb{E}[\mathbb{1}_{\{T_j^l\in(t+r,u]\}}] \\
&\quad \leq \mathbb{P}(T_j^l \geq t+r) \\
&\quad \leq \frac{\mathbb{E}[(T_1^l)^{1+\delta}]}{(t+r)^{1+\delta}} \\
&\quad \leq \frac{l^{1+\delta}\nu_{1+\delta}}{(t+r)^{1+\delta}},
\end{aligned}$$

using Markov's inequality for the second to last inequality, and Lemma 4 for the last one. Similarly,

$$\begin{aligned}
&\operatorname{Cov}(\mathbb{1}_{\{T_i^k\in(s,t]\}}, \mathbb{1}_{\{T_j^l\in(t+r,u]\}}) \\
&\quad = \mathbb{E}[\mathbb{1}_{\{T_i^k\in(s,t]\}}\mathbb{1}_{\{T_j^l\in(t+r,u]\}}] - \mathbb{E}[\mathbb{1}_{\{T_i^k\in(s,t]\}}]\mathbb{E}[\mathbb{1}_{\{T_j^l\in(t+r,u]\}}] \\
&\quad \geq -\mathbb{E}[\mathbb{1}_{\{T_j^l\in(t+r,u]\}}] \\
&\quad \geq -\frac{l^{1+\delta}\nu_{1+\delta}}{(t+r)^{1+\delta}}.
\end{aligned}$$

The second term is straightforward,

$$\big|\mathbb{E}[\mathbb{1}_{\{T_i^k\in(s,t]\}}]\mathbb{E}[\mathbb{1}_{\{T_j^l\in(t+r,u]\}}]\big| \leq \mathbb{E}[\mathbb{1}_{\{T_j^l\in(t+r,u]\}}] \leq \frac{l^{1+\delta}\nu_{1+\delta}}{(t+r)^{1+\delta}}.$$

Then

$$\begin{aligned}
&\left|\sum_{i=1}^{+\infty}\sum_{j=1}^{+\infty}\operatorname{Cov}(\mathbb{1}_{\{Z_k\geq i\}}\mathbb{1}_{\{T_i^k\in(s,t]\}}, \mathbb{1}_{\{Z_l\geq j\}}\mathbb{1}_{\{T_j^l\in(t+r,u]\}})\right| \\
&\quad \leq \frac{l^{1+\delta}\nu_{1+\delta}}{(t+r)^{1+\delta}}\left|\sum_{i=1}^{+\infty}\sum_{j=1}^{+\infty}\mathbb{E}[\mathbb{1}_{\{Z_k\geq i\}}\mathbb{1}_{\{Z_l\geq j\}}] + \sum_{i=1}^{+\infty}\sum_{j=1}^{+\infty}\operatorname{Cov}(\mathbb{1}_{\{Z_k\geq i\}}, \mathbb{1}_{\{Z_l\geq j\}})\right| \\
&\quad = \frac{l^{1+\delta}\nu_{1+\delta}}{(t+r)^{1+\delta}}\left|\mathbb{E}[Z_k Z_l] + \operatorname{Cov}(Z_k, Z_l)\right| \\
&\quad \leq 2\frac{l^{1+\delta}\nu_{1+\delta}}{(t+r)^{1+\delta}}\mu^{k\vee l}\frac{1-\mu^{k\wedge l+1}}{1-\mu},
\end{aligned}$$

using Lemma 3 for the last inequality. □

Straightforwardly, since $\sum \mu^k$ and $\sum l^{1+\delta}\mu^l$ are summable for $\delta > 0$, we get the following lemma.

LEMMA 8. *Let $s, t, u \in \mathbb{R}$ and $r > 0$ such that $s < t < t + r < u$. Suppose that there exists $\delta > 0$ such that $\nu_{1+\delta} < \infty$. Then*

$$\left|\operatorname{Cov}(N((s,t]|0), N\big((t+r,u]|0)\big)\right| = \mathcal{O}\left(\frac{1}{(t+r)^{1+\delta}}\right).$$

All that is left to prove Theorem 1 and Corollary 1 is to integrate the upper bound with respect to the first-moment measure of the center process. Using the notation of Lemmas 1 and 8, and with $M_c(\cdot) = \eta(\cdot)\ell(\cdot)$ where $\ell(\cdot)$ is the Lebesgue measure,

$$\begin{aligned}
\left|\operatorname{Cov}\big(\mathbb{1}_{\mathcal{A}}(N), \mathbb{1}_{\mathcal{B}}(N)\big)\right| &\leq \int_{\mathbb{R}} \left|\operatorname{Cov}(N((s,t]|y), N\big((t+r,u]|y)\big)\right| M_c(\mathrm{d}y) \\
&= \int_{-\infty}^{t} \left|\operatorname{Cov}(N((s,t]|y), N\big((t+r,u]|y)\big)\right| M_c(\mathrm{d}y) \\
&= \mathcal{O}\left(\int_{-\infty}^{t} \frac{1}{(t+r-y)^{1+\delta}} \eta(y)\,\mathrm{d}y\right) \\
&= \mathcal{O}\big(r^{-\delta}\big),
\end{aligned}$$

where the last inequality follows from the boundedness of $\eta(\cdot)$. This upper bound is valid for any $s, u \in \mathbb{R}$ therefore holds for $\mathcal{A} \in \mathcal{E}_{-\infty}^{t}$, $\mathcal{B} \in \mathcal{E}_{t+r}^{\infty}$.

We now turn to an upper bound for the strong mixing coefficient when the reproduction kernel h^* admits a finite exponential moment, that is, there exists $a_0 > 0$ such that

$$M(a_0) := \int_{\mathbb{R}} e^{a_0|t|} h^*(t)\,\mathrm{d}t < \infty.$$

Choose $a \in (0, a_0]$ such that $1 < M(a) < \mu^{-1}$. Then, by substituting in Lemma 7 the Markov inequality with

$$\begin{aligned}
\mathbb{P}\big(T_j^l \geq t + r\big) &\leq \mathbb{E}\big[\exp\big(a\big|T_1^l\big|\big)\big] e^{-a(t+r)} \\
&\leq \mathbb{E}\big[\exp\big(a\big|\Delta_1^1\big|\big)\big]^l e^{-a(t+r)} \\
&= M(a)^l e^{-a(t+r)},
\end{aligned}$$

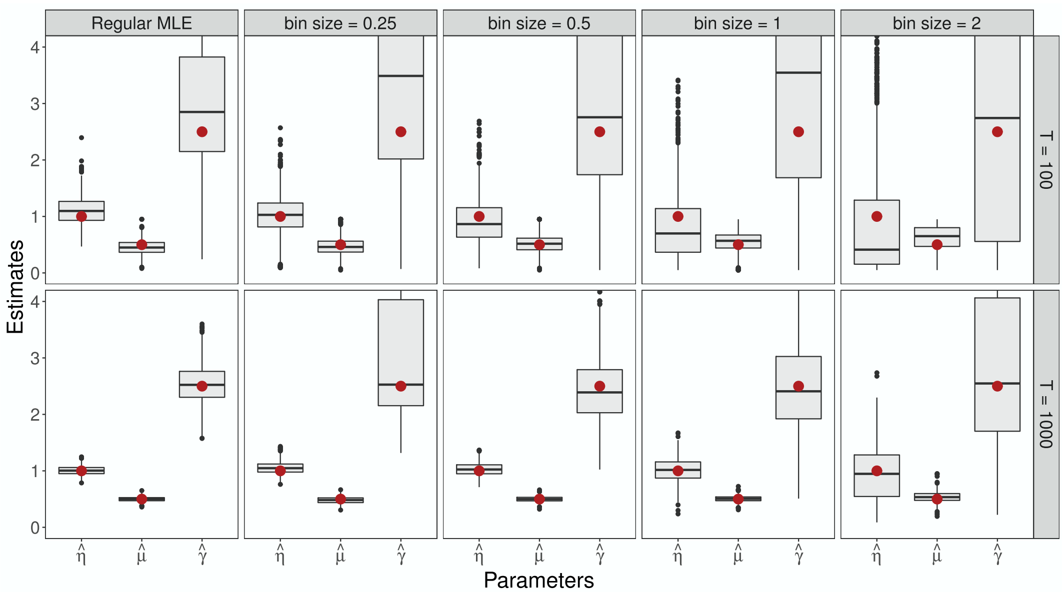
the term $\sum M(a)^l \mu^l$ is again summable, and Lemma 8 turns into

$$\left|\operatorname{Cov}(N((s,t]|0), N\big((t+r,u]|0)\big)\right| = \mathcal{O}\big(e^{-a(t+r)}\big).$$

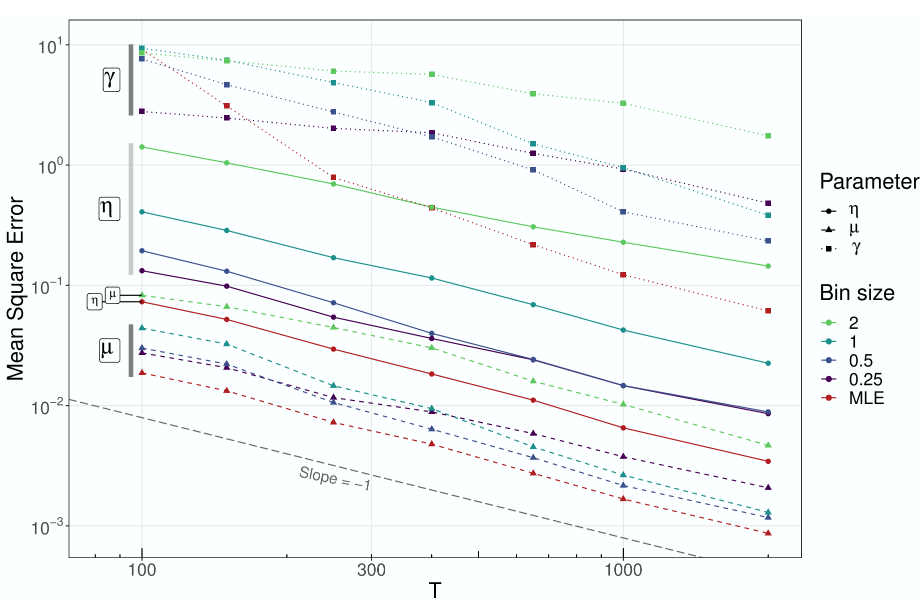
Finally, by integrating with respect to the first-moment measure of the centre process, we get the desired result:

$$\left|\operatorname{Cov}\big(\mathbb{1}_{\mathcal{A}}(N), \mathbb{1}_{\mathcal{B}}(N)\big)\right| = \mathcal{O}\big(e^{-ar}\big).$$

APPENDIX B: FIGURES OF SECTION 5

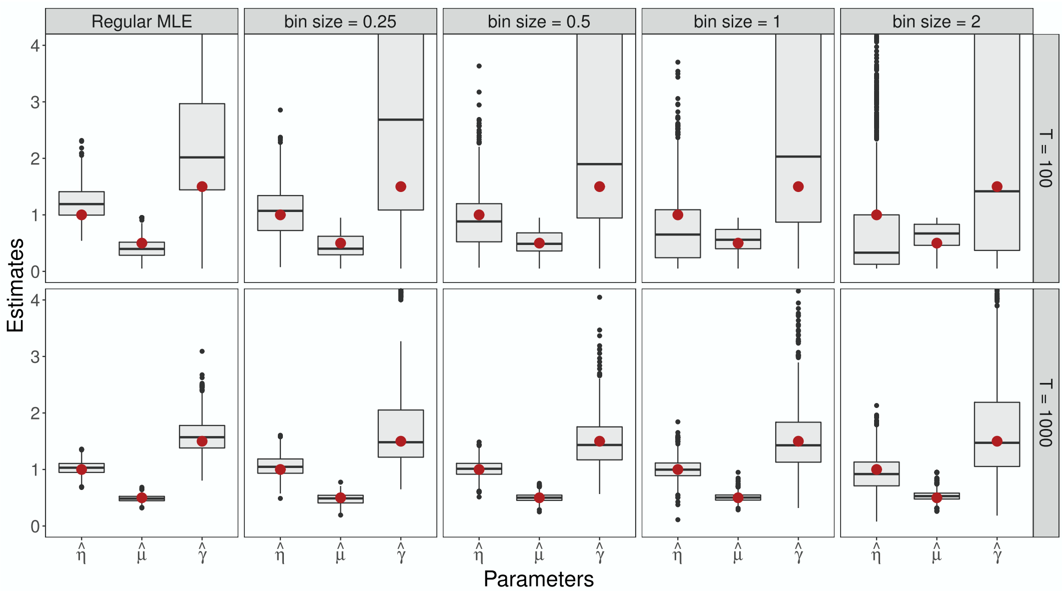


(a) Estimates of parameters η, μ and γ for 1,000 simulations on the interval $[0, T]$. True values (larger dots) are: $\eta = 1$, $\mu = 0.5$, $\gamma = 2.5$. The left column refers to the maximum likelihood estimates. The other columns refer to the Whittle estimates according to different bin sizes.

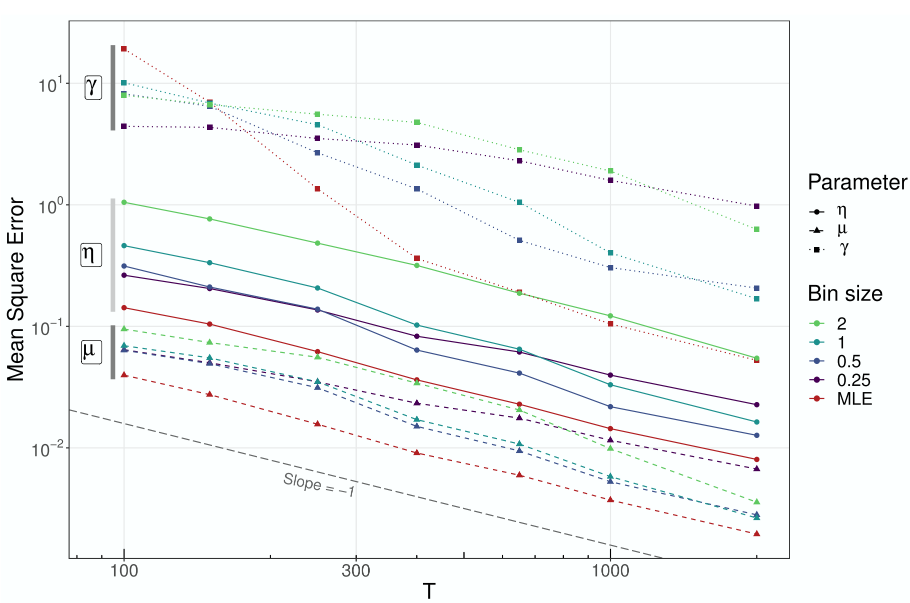


(b) Mean square error of the estimates of parameters η, μ and γ for 1,000 simulations on the interval $[0, T]$, in log-log scale. The dashed grey line represents the ideal slope of -1, that is, a rate of convergence of $\mathcal{O}(n^{-1})$.

FIG. 4. *Performance of the Whittle estimates for the stationary Hawkes process with immigration intensity* $\eta = 1$, *reproduction mean* $\mu = 0.5$, *and reproduction kernel* $h^*(t) = \gamma a^\gamma (a + t)^{-\gamma - 1}$, *where* $\gamma = 2.5$ *and* $a = 1.5$.

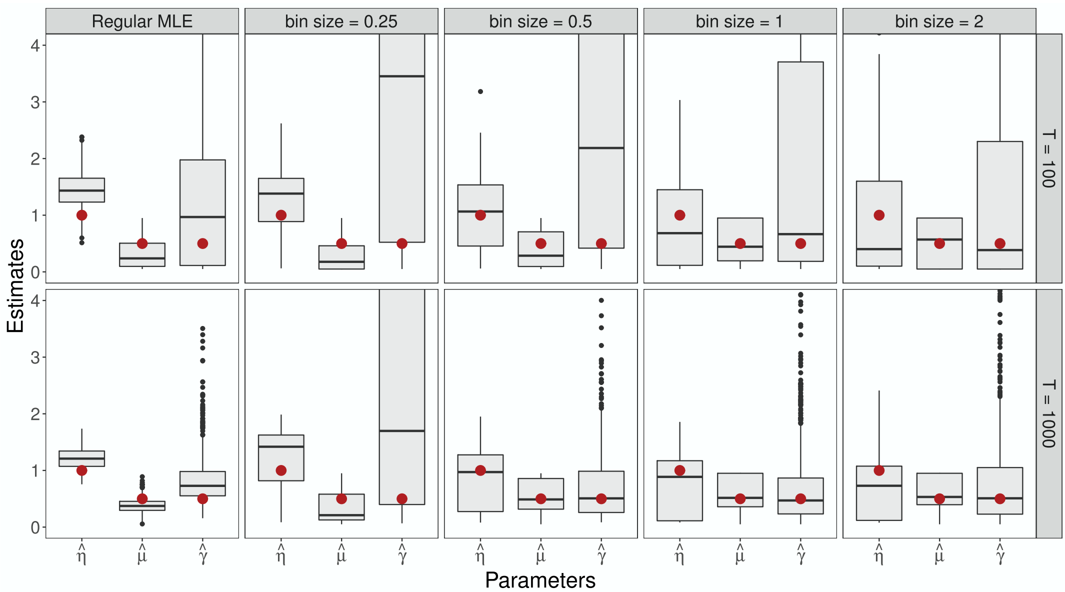


(a) Estimates of parameters η, μ and γ for 1,000 simulations on the interval $[0, T]$. True values (larger dots) are: $\eta = 1$, $\mu = 0.5$, $\gamma = 1.5$. The left column refers to the maximum likelihood estimates. The other columns refer to the Whittle estimates according to different bin sizes.

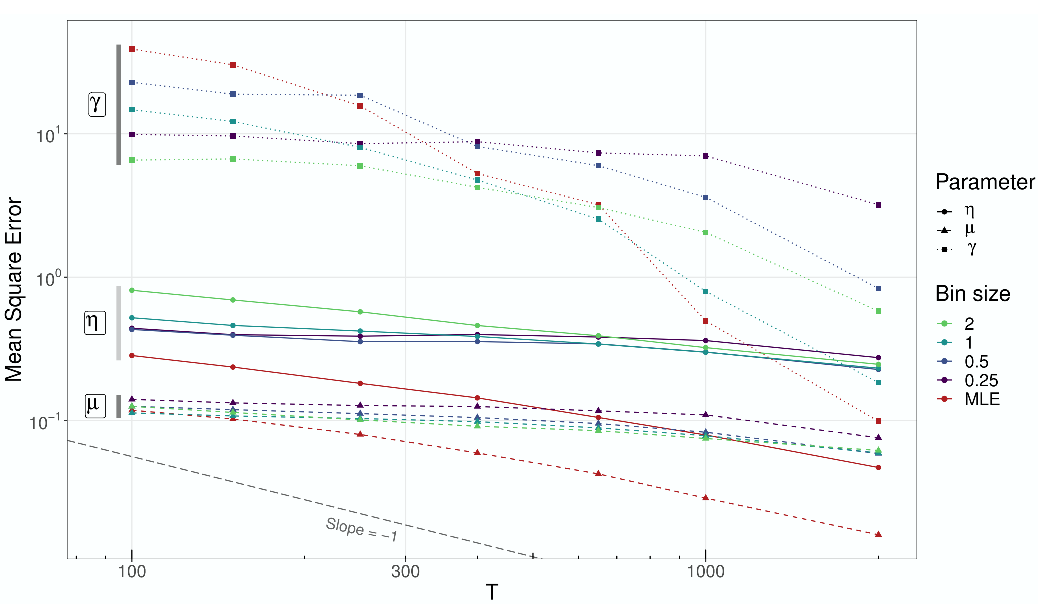


(b) Mean square error of the estimates of parameters η, μ and γ for 1,000 simulations on the interval $[0, T]$, in log-log scale. The dashed grey line represents the ideal slope of -1, that is, a rate of convergence of $\mathcal{O}(n^{-1})$.

FIG. 5. *Performance of the Whittle estimates for the stationary Hawkes process with immigration intensity* $\eta = 1$, *reproduction mean* $\mu = 0.5$, *and reproduction kernel* $h^*(t) = \gamma a^\gamma (a + t)^{-\gamma-1}$, *where* $\gamma = 1.5$ *and* $a = 1.5$.



(a) Estimates of parameters η, μ and γ for 1,000 simulations on the interval $[0, T]$. True values (larger dots) are: $\eta = 1$, $\mu = 0.5$, $\gamma = 0.5$. The left column refers to the maximum likelihood estimates. The other columns refer to the Whittle estimates according to different bin sizes.



(b) Mean square error of the estimates of parameters η, μ and γ for 1,000 simulations on the interval $[0, T]$, in log-log scale. The dashed grey line represents the ideal slope of -1, *i.e.* a rate of convergence of $\mathcal{O}(n^{-1})$.

FIG. 6. *Performance of the Whittle estimates for the stationary Hawkes process with immigration intensity* $\eta = 1$, *reproduction mean* $\mu = 0.5$, *and reproduction kernel* $h^*(t) = \gamma a^\gamma (a + t)^{-\gamma - 1}$, *where* $\gamma = 0.5$ *and* $a = 1.5$.

Acknowledgments. The authors would like to thank François Roueff who suggested the use of Whittle's method for the estimation of Hawkes processes from bin-count data and Theorem 1's extension to exponentially decaying reproduction kernels. The authors would also like to thank the three anonymous reviewers who helped, through their remarks and suggestions, to considerably improve this article. During this work, Felix Cheysson was a Ph.D. student of UMR MIA-Paris, Université Paris-Saclay, AgroParisTech, INRAE; Epidemiology and Modeling of bacterial Evasion to Antibacterials Unit (EMEA), Institut Pasteur and Centre de recherche en Epidémiologie et Santé des Populations (CESP), Université Paris-Saclay, UVSQ, Inserm.

REFERENCES

ADAMOPOULOS, L. (1976). Cluster models for earthquakes: Regional comparisons. *J. Int. Assoc. Math. Geol.* **8** 463–475.

BACRY, E., MASTROMATTEO, I. and MUZY, J.-F. (2015). Hawkes processes in finance. *Mark. Microstruct. Liq.* **1** 1550005.

BOWSHER, C. G. (2003). Modelling security market events in continuous time: Intensity based, multivariate point process models. *SSRN Electron. J.* 1–39.

BRADLEY, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2** 107–144. MR2178042 https://doi.org/10.1214/154957805100000104

CELEUX, G., CHAUVEAU, D. and DIEBOLT, J. (1995). On Stochastic Versions of the EM Algorithm. Technical Report No. RR-2514 INRIA.

CENTERS FOR DISEASE CONTROL AND PREVENTION (2015). *Epidemiology and Prevention of Vaccine-Preventable Diseases*, 13 ed. Public Health Foundation, Washington D.C.

CHAVEZ-DEMOULIN, V., DAVISON, A. C. and MCNEIL, A. J. (2005). Estimating value-at-risk: A point process approach. *Quant. Finance* **5** 227–234. MR2240248 https://doi.org/10.1080/14697680500039613

CHORNOBOY, E. S., SCHRAMM, L. P. and KARR, A. F. (1988). Maximum likelihood identification of neural point process systems. *Biol. Cybernet.* **59** 265–275. MR0961117 https://doi.org/10.1007/BF00332915

DALEY, D. J. and VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes. Vol. I*, 2nd ed. *Probability and Its Applications* (*New York*). Springer, New York. MR1950431

DAVYDOV, J. A. (1974). Mixing conditions for Markov chains. *Theory Probab. Appl.* **18** 312–328.

DELYON, B., LAVIELLE, M. and MOULINES, E. (1999). Convergence of a stochastic approximation version of the EM algorithm. *Ann. Statist.* **27** 94–128. MR1701103 https://doi.org/10.1214/aos/1018031103

DOUKHAN, P. (1994). *Mixing*: *Properties and Examples*. *Lecture Notes in Statistics* **85**. Springer, New York. MR1312160 https://doi.org/10.1007/978-1-4612-2642-0

DZHAPARIDZE, K. O. (1974). A new method for estimating spectral parameters of a stationary regular time series. *Theory Probab. Appl.* **19** 122–132.

DZHAPARIDZE, K. (1986). *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series*. *Springer Series in Statistics*. Springer, New York. MR0812272 https://doi.org/10.1007/978-1-4612-4842-2

EDDELBUETTEL, D. and FRANÇOIS, R. (2011). Rcpp: Seamless R and C++ integration. *J. Stat. Softw.* **40** 1–18.

GAO, X. and ZHU, L. (2018). Functional central limit theorems for stationary Hawkes processes and application to infinite-server queues. *Queueing Syst.* **90** 161–206. MR3850052 https://doi.org/10.1007/s11134-018-9570-5

HAWKES, A. G. (1971a). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58** 83–90. MR0278410 https://doi.org/10.1093/biomet/58.1.83

HAWKES, A. G. (1971b). Point spectra of some mutually exciting point processes. *J. Roy. Statist. Soc. Ser. B* **33** 438–443. MR0358976

HAWKES, A. G. and OAKES, D. (1974). A cluster process representation of a self-exciting process. *J. Appl. Probab.* **11** 493–503. MR0378093 https://doi.org/10.2307/3212693

HEINRICH, L. and PAWLAS, Z. (2013). Absolute regularity and Brillinger-mixing of stationary point processes. *Lith. Math. J.* **53** 293–310. MR3097306 https://doi.org/10.1007/s10986-013-9209-5

HOSOYA, Y. (1974). Estimation problems on stationary time series models. Ph.D. dissertation, Yale Univ.

JOHNSON, S. G. The NLopt nonlinear-optimization package.

KIRCHNER, M. (2016). Hawkes and INAR(∞) processes. *Stochastic Process. Appl.* **126** 2494–2525. MR3505235 https://doi.org/10.1016/j.spa.2016.02.008

KIRCHNER, M. (2017). An estimation procedure for the Hawkes process. *Quant. Finance* **17** 571–595. MR3620953 https://doi.org/10.1080/14697688.2016.1211312

LIU, D. C. and NOCEDAL, J. (1989). On the limited memory BFGS method for large scale optimization. *Math. Program.* **45** 503–528. MR1038245 https://doi.org/10.1007/BF01589116

MEYER, S., ELIAS, J. and HÖHLE, M. (2012). A space–time conditional intensity model for invasive meningococcal disease occurrence. *Biometrics* **68** 607–616. MR2959628 https://doi.org/10.1111/j.1541-0420.2011.01684.x

NISHIURA, H., MIZUMOTO, K. and ASAI, Y. (2017). Assessing the transmission dynamics of measles in Japan, 2016. *Epidemics* **20** 67–72. https://doi.org/10.1016/j.epidem.2017.03.005

NOCEDAL, J. (1980). Updating quasi-Newton matrices with limited storage. *Math. Comp.* **35** 773–782. MR0572855 https://doi.org/10.2307/2006193

OAKES, D. (1975). The Markovian self-exciting process. *J. Appl. Probab.* **12** 69–77. MR0362522 https://doi.org/10.1017/s0021900200033106

OGATA, Y. (1978). The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Ann. Inst. Statist. Math.* **30** 243–261. MR0514494 https://doi.org/10.1007/BF02480216

OGATA, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Amer. Statist. Assoc.* **83** 9–27.

OLSON, J. F. and CARLEY, K. M. (2013). Exact and approximate EM estimation of mutually exciting Hawkes processes. *Stat. Inference Stoch. Process.* **16** 63–80. MR3029333 https://doi.org/10.1007/s11203-013-9074-1

OZAKI, T. (1979). Maximum likelihood estimation of Hawkes' self-exciting point processes. *Ann. Inst. Statist. Math.* **31** 145–155. MR0541960 https://doi.org/10.1007/BF02480272

PAPANGELOU, F. (1972). Integrability of expected increments of point processes and a related random change of scale. *Trans. Amer. Math. Soc.* **165** 483–506. MR0314102 https://doi.org/10.2307/1995899

PAPARODITIS, E. (2000). Spectral density based goodness-of-fit tests for time series models. *Scand. J. Stat.* **27** 143–176. MR1774049 https://doi.org/10.1111/1467-9469.00184

POINAS, A., DELYON, B. and LAVANCIER, F. (2019). Mixing properties and central limit theorem for associated point processes. *Bernoulli* **25** 1724–1754. MR3961228 https://doi.org/10.3150/18-BEJ1033

R CORE TEAM (2019). R: A Language and Environment for Statistical Computing.

REINHART, A. (2018). A review of self-exciting spatio-temporal point processes and their applications. *Statist. Sci.* **33** 299–318. MR3843374 https://doi.org/10.1214/17-STS629

REYNAUD-BOURET, P. and SCHBATH, S. (2010). Adaptive estimation for Hawkes processes; application to genome analysis. *Ann. Statist.* **38** 2781–2822. MR2722456 https://doi.org/10.1214/10-AOS806

RIO, E. (2017). *Asymptotic Theory of Weakly Dependent Random Processes. Probability Theory and Stochastic Modelling* **80**. Springer Berlin Heidelberg, Berlin, Heidelberg.

ROSENBLATT, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci. USA* **42** 43–47. MR0074711 https://doi.org/10.1073/pnas.42.1.43

ROUEFF, F., VON SACHS, R. and SANSONNET, L. (2016). Locally stationary Hawkes processes. *Stochastic Process. Appl.* **126** 1710–1743. MR3483734 https://doi.org/10.1016/j.spa.2015.12.003

SHAO, X. (2010). A self-normalized approach to confidence interval construction in time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 343–366. MR2758116 https://doi.org/10.1111/j.1467-9868.2009.00737.x

WESTCOTT, M. (1971). On existence and mixing results for cluster point processes. *J. Roy. Statist. Soc. Ser. B* **33** 290–300. MR0317443

WESTCOTT, M. (1972). The probability generating functional. *J. Aust. Math. Soc.* **14** 448–466. MR0324772

WHEATLEY, S., FILIMONOV, V. and SORNETTE, D. (2016). The Hawkes process with renewal immigration & its estimation with an EM algorithm. *Comput. Statist. Data Anal.* **94** 120–135. MR3412815 https://doi.org/10.1016/j.csda.2015.08.007

WHITTLE, P. (1952). Some results in time series analysis. *Skand. Aktuarietidskr.* **35** 48–60. MR0049539 https://doi.org/10.1080/03461238.1952.10414182