



HAL
open science

Strong-mixing rates for Hawkes processes and application to Whittle estimation from count data

Felix Cheyssou, Gabriel Lang

► **To cite this version:**

Felix Cheyssou, Gabriel Lang. Strong-mixing rates for Hawkes processes and application to Whittle estimation from count data. 2021. hal-03117924v1

HAL Id: hal-03117924

<https://hal.science/hal-03117924v1>

Preprint submitted on 21 Jan 2021 (v1), last revised 22 Sep 2023 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

STRONG-MIXING RATES FOR HAWKES PROCESSES AND APPLICATION TO WHITTLE ESTIMATION FROM COUNT DATA

Felix Cheyssou^{1,2,3} & Gabriel Lang¹

¹ *MIA-Paris, AgroParisTech, Paris-Saclay, France*

² *Epidemiology and Modeling of bacterial Evasion to Antibacterials Unit (EMEA), Institut Pasteur, Paris, France*

³ *Anti-infective Evasion and Pharmacoepidemiology Team, Centre for Epidemiology and Public health (CESP), Inserm / UVSQ, France*

Abstract. In this paper, we study the time series generated by the event counts of the stationary Hawkes process. Using the cluster properties of the stationary Hawkes process, we prove an upper bound for its strong-mixing coefficient, and for its count series', provided that the reproduction kernel has a finite $(1 + \beta)$ -th order moment (for a $\beta > 0$). When the exact locations of points are not observed, but only counts over fixed time intervals, we propose a spectral approach to the estimation of Hawkes processes, based on Whittle's likelihood. This approach provides consistent and asymptotically normal estimates provided common regularity conditions on the reproduction kernel. Simulated datasets illustrate the performances of the estimation, notably, of the Hawkes reproduction mean and kernel, even with relatively large time intervals.

Keywords. Hawkes process, Strong mixing, Bartlett spectrum, Whittle estimation, Time series

1 Introduction

In this paper, we study the time series generated by the event counts of a Hawkes process. We propose a spectral approach to the parametric estimation of the process from its count data sequence and prove strong-mixing properties necessary to get good asymptotic properties for the estimator.

Hawkes eponymous self-exciting point processes [1, 2] form a family of models for point processes for which the occurrence of any event increases temporarily the probability of further events occurring. Importantly, the Hawkes process exhibits clustering properties: it is a special case of the Poisson cluster process, where each cluster is a subcritical continuous-time Galton-Watson tree with Poisson offspring distribution [3]. These self-exciting and clustering properties are appealing in point process modeling, and while first applications concerned almost exclusively seismology [4, 5], the use of Hawkes

processes quickly spread to many other disciplines, including neurophysiology [6], finance [7], genomics [8] and epidemiology [9].

Parameter estimation of Hawkes processes has been studied thoroughly when events are fully observed, first relying on spectral analysis techniques [4], then on maximum likelihood methods [5, 10, 11]. However, when count data are only observed in discrete time (*i.e.* the timeline is cut into bins and the numbers of events in each bin is counted) these methods are not directly applicable. Kirchner [12] proposed a non-parametric estimation, approximating the distribution of the bin-count sequence by an INAR(∞) sequence, and showed that the INAR conditional least-square estimation yields consistent and asymptotically normal estimates for the underlying Hawkes process when the binsize tends to zero [13]. However, while adapted for most event data which live on relatively discrete time grids, these estimates are biased for those with large binsize.

Another approach often taken for the estimation of Hawkes processes, for example when the processes are multivariate [14] or when the immigration intensity is a renewal process [15], is using an Expectation Maximization algorithm which considers the branching structure of the process as a latent variable. For a process sampled in discrete time, an analogous approach which would consider the arrival times as latent variables is unfortunately not adapted, since there is no closed form for their conditional distribution given the event counts. Stochastic expectation maximization algorithms [16], which approximate this conditional distribution, do not alleviate this issue since usual convergence results are based on likelihoods of the exponential families [17], which excludes Hawkes processes.

In this paper we revisit Adamopoulos's spectral approach to the parametric estimation of Hawkes processes [4]. Using the Bartlett spectrum of the Hawkes process (*i.e.* the spectral density of the covariance measure of the process), Adamopoulos defined as an estimator the minimiser of the log-spectral likelihood, first introduced by Whittle [18]. We extend these results to processes whose event counts are only observed in discrete time.

Section 3 contains our first important result: we establish an upper bound on the strong-mixing coefficient of the Hawkes process, and therefore of its count series, using its cluster properties. Rosenblatt [19] introduced the strong-mixing coefficient to formalise a measure of dependence between random variables, which has been useful in proving moment inequalities and central limit theorems [20, 21]. While Hawkes processes have already been shown to be strongly-mixing [3, 22], this has not yet led to any development in their estimation. Only recent results [23], that hold specifically because the σ -algebras generated by countable sets are poorer than those generated by continuous sets, enable the mixing properties to be used in the estimation of Hawkes processes.

In Section 4, we focus on the estimation of Hawkes processes from their count data. We adapt Adamopoulos's work by taking into account the aliasing of the spectral density caused by sampling the process in discrete time. we introduce a correction to the spectral density function by taking into account the spectral aliasing that occurs when the process

is sampled in discrete time. Then, using the strong-mixing condition and the work of Dzhaparidze [24] on Whittle's method, we propose a consistent and asymptotically normal estimator to the parameters of the Hawkes process.

Finally, Section 5 provides some numerical experiments to illustrate the results of the two preceding sections and in Section 6, we discuss some of the appealing features of this approach.

2 The Hawkes process and its count series

2.1 Notation

In this paper, we consider *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, *i.e.* the atoms of N , is denoted $\{T_i\}$. For a function f on \mathbb{R} , we write

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

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

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

2.2 The stationary 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

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

The constant $\eta > 0$ is called *immigration intensity* and the measurable function $h : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ *reproduction function*.

Moreover, the Hawkes process is a specific case of the Poisson *cluster process* [3]. Briefly, the process consists of a stream of *immigrants*, the cluster centres, 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 superimposition of all branching processes:

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

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 , *i.e.* $Z_0 = 1$ 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 $z\mu$, where $\mu := \int_{\mathbb{R}} h(t)dt$. 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 series

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 count series, according to whether the interval endpoints live on the real line or on a regular grid respectively (see Figure 1):

Definition 1. *The count data time series (or count series in short) with binsize Δ associated to a point process N is the process $(X_t)_{t \in \mathbb{R}} = (N(t\Delta, (t+1)\Delta])_{t \in \mathbb{R}}$ or $(X_k)_{k \in \mathbb{Z}} = (N(k\Delta, (k+1)\Delta])_{k \in \mathbb{Z}}$, generated by the count measure on intervals of size Δ .*

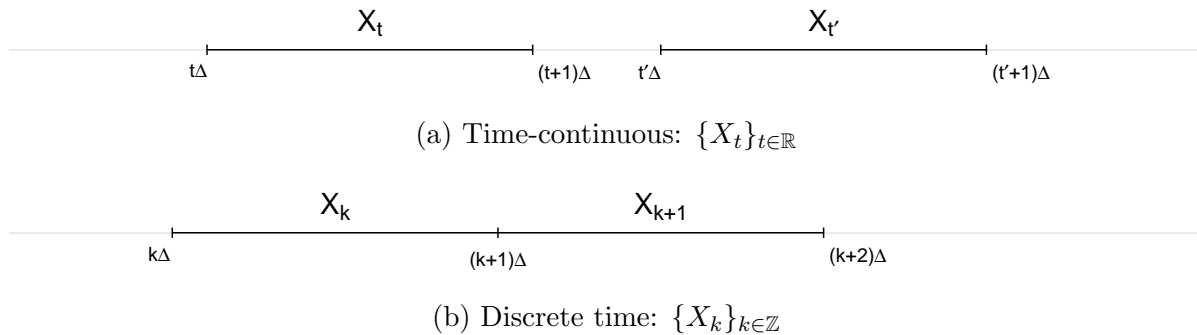


Figure 1: Count series with binsize Δ

3 Strong-mixing properties

Here, we control the strong-mixing coefficients of Hawkes processes and their associated count series. 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} [19]:

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

This definition can be adapted to a point process N on \mathbb{R} , by defining (see [23])

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

where \mathcal{E}_a^b stands for $\mathcal{E}((a, b])$, *i.e.* the σ -algebra generated by the cylinder sets on the interval $(a, b]$. For a given 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 $(X_k)_{a \leq k \leq b}$.

The point process N (resp. the sequence (X_k)) is said to be strongly-mixing if $\alpha_N(r)$ (resp. $\alpha_X(r)$) $\rightarrow 0$ as $r \rightarrow \infty$. Intuitively, the strong-mixing condition conveys that the past and the future of the process are asymptotically independent the further they are separated. Note that, since $\mathcal{F}_a^b \subset \mathcal{E}((a, b])$, we have that $\alpha_{(X_k)}(r) \leq \alpha_N(r)$ for all r .

We here state the first important result of this article:

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

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

Then N is strongly-mixing and

$$\alpha_N(r) = \mathcal{O}\left(\frac{1}{r^\beta}\right).$$

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 distribution 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 to Theorem 1, we get the following corollary for Hawkes count series:

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

$$\alpha_X(r) = \mathcal{O}\left(\frac{1}{r^\beta}\right).$$

4 Parametric estimation of count series

In this section, we apply the strong-mixing properties of the Hawkes count series to parametric estimation using a spectral approach. First, we derive the spectral density function for both the time-continuous and discrete time Hawkes count series. Then using Whittle's method, we define a parametric estimator of a Hawkes process from its 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, non-negative, symmetric measure Γ on the Borel sets such that, for any rapidly decaying functions φ and ψ on \mathbb{R} , (see [25, Proposition 8.2.I])

$$\text{Cov}(N(\varphi), N(\psi)) = \int_{\mathbb{R}} \tilde{\varphi}(\omega) \tilde{\psi}^*(\omega) \Gamma(d\omega), \quad (1)$$

where $\psi^*(u) = \psi(-u)$, and $\tilde{\cdot}$ denotes the Fourier transform:

$$\tilde{\varphi}(\omega) = \int_{\mathbb{R}} e^{-i\omega s} \varphi(s) ds.$$

For the stationary Hawkes process, the Bartlett spectrum admits a density given by (see [25, Example 8.2(e)])

$$\gamma(\omega) = \frac{m}{2\pi} \left| 1 - \tilde{h}(\omega) \right|^{-2} \quad (2)$$

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

We then derive the spectral density of the time-continuous count series with binsize 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 count series. Then X_t has a spectral density function given by*

$$f_{X_t}(\omega) = m \Delta \text{sinc}^2\left(\frac{\omega}{2}\right) \left| 1 - \tilde{h}\left(\frac{\omega}{\Delta}\right) \right|^{-2}. \quad (3)$$

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

$$\begin{aligned}\tilde{\varphi}(\omega) &= \int_0^\Delta e^{-i\omega s} ds = \frac{i}{\omega} [e^{-i\omega\Delta} - 1], \\ \tilde{\psi}^*(\omega) &= \int_{-\Delta(u+1)}^{-\Delta u} e^{-i\omega s} ds = \frac{i}{\omega} e^{i\omega\Delta u} [1 - e^{i\omega\Delta}].\end{aligned}$$

Then, using (1) and (2), the autocovariance function of X_t is

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

□

For the Hawkes process sampled in discrete time on a regular unit time grid, we must take into account spectral aliasing, which folds high frequencies back onto the apparent spectrum:

Corollary 2. *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 count series. Then X_k has a spectral density function given by*

$$f_{X_k}(\omega) = \sum_{k \in \mathbb{Z}} f_{X_t}(\omega + 2k\pi)$$

where $f_{X_t}(\cdot)$ is the function defined in (3).

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 [26] and Dzhaparidze [27], building on the cornerstone laid by Whittle [18], proposed as an estimator of θ the minimizer

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta} \mathcal{L}_n(\theta) \tag{4}$$

where

$$\mathcal{L}_n(\theta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(\log f_\theta(\omega) + \frac{I_n(\omega)}{f_\theta(\omega)} \right) d\omega \tag{5}$$

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

Dzhaparidze [24] 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 [24, Theorem II.7.1 and II.7.2] for stationary Hawkes count series.

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

- (A1) *The true value θ_0 of the parameter θ 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 $\beta > 0$ such that the reproduction kernel h^ has a finite moment of order $2 + \beta$. Then the estimator $\hat{\theta}_n$ defined as in (4) (with $\mathcal{L}_n(\theta)$ given by (5)), is consistent, i.e. $\hat{\theta}_n \rightarrow \theta_0$ in probability.*

Proof. The only condition from Dzhaparidze [24, 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:

$$\sum_{r=1}^{\infty} (\alpha_X(r))^{1-2/\gamma} < \infty. \quad (6)$$

Since the stationary Hawkes process admit finite exponential moments if h^* has a moment of order $\delta \in (0, 1]$ [28, Theorem 4], $\mathbb{E}[|X_k|^{2\gamma}]$ is finite for any γ . Then using Corollary 1 there always exists a $\gamma > 2$ that satisfies (6). \square

Define the matrix Γ_θ , which would actually be the limit as $n \rightarrow \infty$ of the Fisher's information matrix if the process (X_k) were Gaussian [24, Section II.2.2], by the relation:

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

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

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

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 count series 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 $\hat{\theta}_n$ is asymptotically normal and

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

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 [29] 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 count series. 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 Pareto kernel whose higher moments are not finite.

All simulations and analysis have been done using R [30] and the package *hawkesbow*, freely available online¹. The code and estimate datasets are also openly accessible².

5.1 Procedure

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

$$\lambda(t) = \eta + \mu \int \beta e^{-\beta(t-u)} dN(u),$$

i.e. 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 1,000 realisations of the Hawkes process on the interval $[0, T]$ with parameter values $\eta = 1$, $\mu = 0.5$ and $\beta = 1$. 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 2). Since the latter use the full information

¹<https://github.com/fcheysson/hawkesbow>

²<https://github.com/fcheysson/code-spectral-hawkes>

on the location of events, they are arguably better than any estimate based on the count series, and provide a best case scenario for the Whittle estimates when the bin size tends to 0. With an exponential kernel, a set of 1,000 simulations and their Whittle estimation with $T = 1000$ and bin size $\Delta = 1$ takes approximately 4 minutes on a laptop computer with an i5 Intel CPU.

Pareto kernel. We now consider a stationary Hawkes process with a Pareto reproduction kernel: $h_\gamma^*(t) = \gamma a^\gamma t^{-\gamma-1}$ for $t \geq a$. We recall that the moments of a Pareto 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) $\gamma = 1$, the mean is infinite and the process does not satisfy the condition of Theorem 1; (ii) $\gamma = 2$, the process is strongly-mixing, but the variance is infinite and the process does not satisfy the assumptions of Theorem 2; (iii) $\gamma = 3$, the process is strongly-mixing and satisfies the assumptions of Theorems 2 and 3, but the moments of order 3 and higher do not exist.

Similarly to the exponential kernel, we simulated 1,000 simulations of the Hawkes process for each $\gamma \in \{1, 2, 3\}$, with parameter values $\eta = 1$, $\mu = 0.5$, and $a_3 = 2/3$ for $\gamma = 3$, $a_2 = 1/2$ for $\gamma = 2$, such that the Pareto kernels h_3^* and h_2^* and the exponential kernel have the same first-order moment. For the Pareto kernel h_1^* , we chose $a_1 = 1/3$ arbitrarily. We could not compare the Whittle estimates to those of maximum likelihood, since the latter were computationally too expensive as the likelihood criterion contained a large number of discontinuity points with respect to the kernel position parameter a : $p(p-1)/2$ discontinuity points, with p the number of events of the process, though $\mathcal{O}(p)$ discontinuity points for acceptable ranges of a . Estimation figures can be found in Appendix B. With a Pareto kernel, a set of 1,000 simulations and their Whittle estimation with $T = 1000$ and bin size $\Delta = 1$ takes approximately 14 minutes on a laptop computer with an i5 Intel CPU.

5.2 Interpretation

Exponential kernel. For $T = 100$ and small binsizes, the Whittle estimates fare almost as well as the maximum likelihood estimates. The estimation deteriorates massively for higher binsizes, notably for the exponential kernel rate β . This is intuitive, since large binsizes with respect to the kernel scale make it difficult to detect interactions between points. This can be related to the probability that a point in a bin has an offspring in the same bin: assuming the stationarity of the process, this probability is equal to $\Delta^{-1} \int_0^\Delta \int_u^\Delta \beta e^{-\beta(t-u)} dt du = 1 - (\beta\Delta)^{-1}(1 - e^{-\beta\Delta})$. For example, with $\beta = 1$ and $\Delta = 2$, we get a probability of 0.57, *i.e.* 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 binsizes.

To further illustrate the asymptotic properties of the estimation, notably its rate of

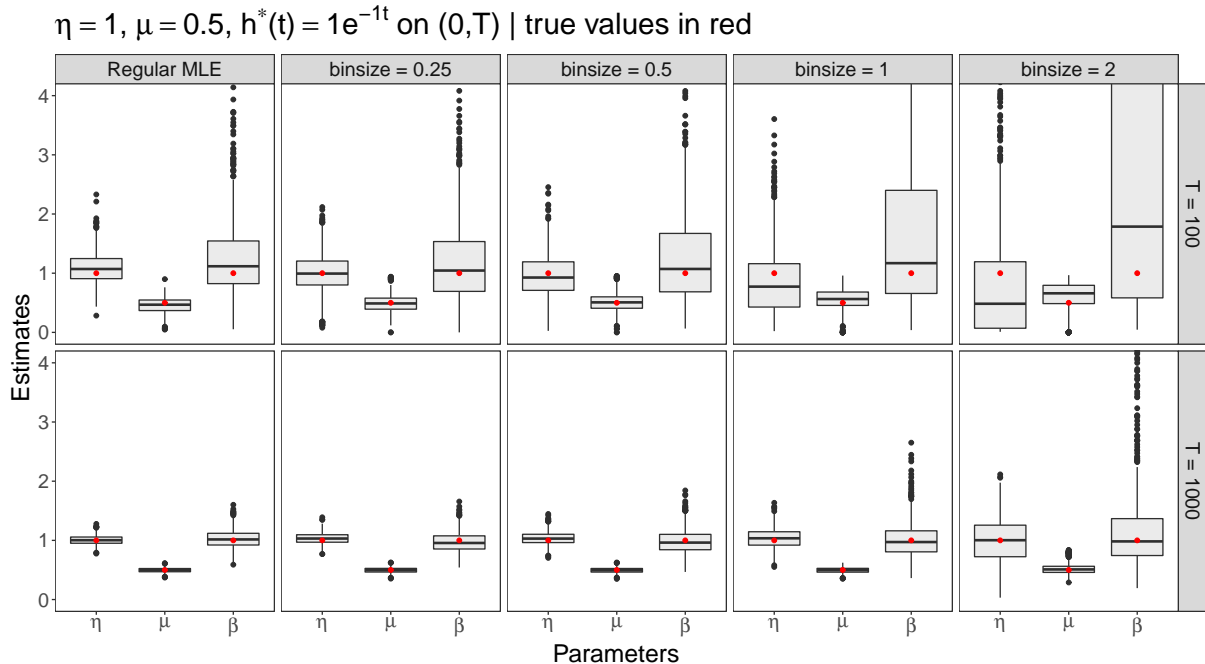


Figure 2: Estimates of parameters η , μ and β for 1,000 simulations of the stationary Hawkes process with kernel $h^*(t) = \beta e^{-\beta t}$ on the interval $[0, T]$. True values (in red) are: $\eta = 1$, $\mu = 0.5$, $\beta = 1$. The left column refers to the maximum likelihood estimates. The other columns refer to the Whittle estimates according to different binsizes.

convergence, we compute the mean square error, defined by $\text{MSE} = S^{-1} \sum (\hat{\theta}_n - \theta_0)^2$, for the estimates of each set of $S = 1,000$ simulations at given T s and binsizes (Figure 3). For large T s, the slope of the mean square error with respect to T reaches -1 (in log-log scale) for all parameters and almost all binsizes, illustrating the $\mathcal{O}(n^{-1})$ rate of convergence stated in Theorem 3. For small T s 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 T s. Finally note that, for reasonable binsizes ($\Delta \leq 1$), the Whittle estimates of the reproduction mean μ have a MSE comparable to those of the maximum likelihood.

Pareto kernel. Performances for the point estimates are remarkably similar amongst all values of γ . Both the immigration intensity η and the reproduction mean μ exhibit the optimal rate of convergence $\mathcal{O}(n^{-1})$ throughout all T s considered for almost all binsizes. On the other hand, the estimates for the Pareto kernel position a show a curious behaviour. While for binsizes 0.5, 1 and 2, the mean square error with respect to T asymptotically reaches the ideal slope of -1 (though with an order of magnitude between binsize 0.5 and

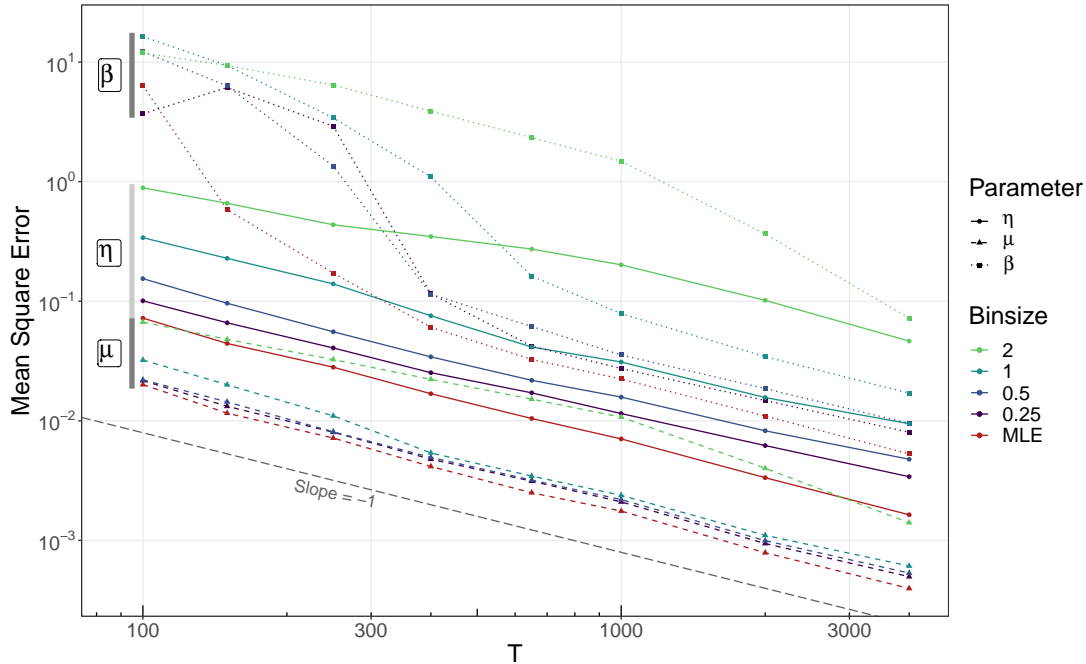


Figure 3: Mean square error of the estimates of parameters η , μ and β for 1,000 simulations of the stationary Hawkes process with kernel $h^*(t) = \beta e^{-\beta t}$ on the interval $[0, T]$, in log-log scale. The dashed grey line represents the ideal slope of -1 , *i.e.* a rate of convergence of $\mathcal{O}(n^{-1})$.

binsize 1, and another between binsize 1 and binsize 2), it does not seem to have reached a similar asymptotic regime for binsize 0.25, which exhibits MSE almost comparable to binsize 2. We are not able to explain this behaviour.

Interestingly, that the point estimates exhibit good asymptotic behaviours for all values of γ , binsize 0.25 excepted, even though the Pareto kernels h_2^* and h_1^* do not satisfy the assumptions of Theorems 2 and 3 would suggest that the upper bound found on the strong-mixing coefficient is not optimal. This can be expected since Theorem 1 only proves a strong-mixing condition for kernel with finite moments $1 + \beta$, $\beta > 0$, while it has been shown that all stationary Hawkes processes are strongly-mixing, irrespective of their reproduction kernels [3, 22]. Nevertheless, the assumption on the moments of the kernels are mild enough that the spectral approach developed in this article can be useful for applications in many disciplines.

6 Conclusion

In this article, we propose a Whittle estimation procedure for the Hawkes process from their count series. 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 \tilde{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 [11], making it more efficient than our approach); *(iv)* it is particularly well-adapted to applications where the binsize cannot be chosen arbitrarily, *i.e.* the events are only counted in bins of fixed size.

A direct extension of the results proved hereby concerns non causal Hawkes processes, for which the reproduction kernel h^* may take non-negative values on \mathbb{R}_- . Indeed, all but Lemmas 7 and 8 from Appendix A are directly applicable to non causal Hawkes. For Lemma 8, split the integral into two: one from $-\infty$ to $t + r/2$, the other from $t + r/2$ to ∞ . The first integral is treated as written. For the second integral, Lemma 7 can be extended using a symmetry argument regarding the location of the immigrant and the interval considered. Then the spectral estimation procedure proposed here is applicable to non causal Hawkes processes, with a central limit theorem for its estimator.

We expect that the results of our paper also hold in a multivariate setup, with minimal modifications. Strong-mixing properties were derived using some properties of the Galton-Watson tree that extend to the multitype process. Moreover, the spectral analysis of the Hawkes count series can be straightforwardly extended to the multivariate case, using the results of Daley and Vere-Jones on the multivariate Bartlett spectrum of mutually exciting point processes [25, Example 8.3(c)]. Nevertheless, we decided to concentrate on the univariate case for brevity and added clarity.

7 Acknowledgements

The authors would like to thank François Roueff who suggested the use of Whittle's method for the estimation of Hawkes processes from their count series and Anna Bonnet for her insightful remarks on multivariate Hawkes processes.

A Proof of Theorem 1

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

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

where $\mathbb{1}_{\mathcal{A}}(N)$ is the indicator function of the cylinder set \mathcal{A} , *i.e.* 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 $(A_i)_{1 \leq i \leq k}$ and $(B_j)_{1 \leq j \leq l}$, and for all coordinate-wise increasing functions $F : \mathbb{N}^k \rightarrow \mathbb{R}$ and $G : \mathbb{N}^l \rightarrow \mathbb{R}$, it satisfies

$$\text{Cov}\left(F(N(A_1), \dots, N(A_k)), G(N(B_1), \dots, N(B_l))\right) \geq 0.$$

We start by stating a useful property (see [31, Section 2.1, key property (e)]), which follows from Hawkes processes being infinitely divisible processes:

Proposition 2. *The stationary Hawkes process is positively associated.*

Using this proposition and Poinas et al.'s work on associated point processes [23], 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.

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,*

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

where $M_c(\cdot)$ refers to the first-order moment of the centre process N_c .

Proof. Using Proposition 2 and [23, Theorem 2.5], we have

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

Then, conditioning by the cluster centre process N_c (see for example [25, Exercise 6.3.4]):

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

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. \square

We are now interested in deriving an upper bound for the covariance of counts of a single branching process. Without loss of generality, we consider a cluster whose immigrant is located at time 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 $(s, t]$. By definition, we have

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

Then, the covariance between two intervals for a branching process is

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

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

Lemma 2. *The expectation, variance and second-order moment of Z_k are*

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

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

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

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

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

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 in $s = 1$ gives the following relations:

$$\begin{aligned} \phi'_{k+1}(1) &= \phi'_1(1)\phi'_k(1), \\ \phi''_{k+1}(1) &= \phi''_1(1)\phi'_k(1) + (\phi'_1(1))^2\phi''_k(1), \end{aligned}$$

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

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

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. \square

Lemma 3. *The covariance and second-order product moment of (Z_k) are*

$$\begin{aligned} \text{Cov}(Z_k, Z_l) &= \mu^{k \vee l} \sum_{j=0}^{k \wedge l - 1} \mu^j = \mu^{k \vee l} \frac{1 - \mu^{k \wedge l}}{1 - \mu}, \\ \mathbb{E}[Z_k Z_l] &= \mu^{k \vee l} \sum_{j=0}^{k \wedge l} \mu^j = \mu^{k \vee l} \frac{1 - \mu^{k \wedge l + 1}}{1 - \mu}, \end{aligned}$$

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

Proof. This is a straightforward recurrence, noting that

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

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 . \square

Let T_i^k denote the time of arrival of the i -th 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, *i.e.* $\Delta_i^k = T_i^k - T_j^{k-1}$. Then, for each point i of generation k , there exists a sequence $(\alpha_{i,k}^{(j)})_{1 \leq j \leq k}$, with $\alpha_{i,k}^{(k)} = i$, denoting the indices of the ancestors of T_i^k , such that

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

For the stationary Hawkes process, the Δ_i^k are independent of all other Δ_j^l , and identically distributed according to the measure h^* . As a consequence, we get the following lemma:

Lemma 4. *For $k \in \mathbb{N}$ and $1 \leq i, j \leq Z_k$,*

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

(ii) For $m > 1$, there is an upper bound on the m -th moment of T_1^k :

$$\mathbb{E}[(T_1^k)^m] \leq k^m \mathbb{E}[(\Delta_1^1)^m] = k^m \nu_m$$

where $\nu_m := \int_{\mathbb{R}} t^m h^*(t) dt$.

Proof. This uses the upper bound on the m -th moment of a sum of random variables from [32]:

$$\mathbb{E}[(T_1^k)^m] \leq k^{m-1} \sum_{j=1}^k \mathbb{E}\left[\left(\Delta_{\alpha_{1,k}^{(j)}}^j\right)^m\right].$$

\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 stationary 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 from 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. This is analogous to the backwards simulation of a Wright-Fisher process without the constant population size restriction.

We state a useful lemma 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*

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

Proof. Writing the expression of the covariance then adding and subtracting 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 $\text{Cov}\left(Z_k^{(s,t]}, Z_l^{(t+r,u]}\right)$:

Lemma 7. *Suppose that there exists $m > 1$ such that $\nu_m < \infty$, and $l \geq 0$. Then*

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

Proof. We have

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

Then, by Lemmas 5 and 6,

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

For the first term,

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

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

$$\begin{aligned}
\text{Cov} \left(\mathbb{1}_{\{T_i^k \in (s,t)\}}, \mathbb{1}_{\{T_j^l \in (t+r,u)\}} \right) &= \mathbb{E} \left[\mathbb{1}_{\{T_i^k \in (s,t)\}} \mathbb{1}_{\{T_j^l \in (t+r,u)\}} \right] - \mathbb{E} \left[\mathbb{1}_{\{T_i^k \in (s,t)\}} \right] \mathbb{E} \left[\mathbb{1}_{\{T_j^l \in (t+r,u)\}} \right] \\
&\geq -\mathbb{E} \left[\mathbb{1}_{\{T_j^l \in (t+r,u)\}} \right] \\
&\geq -\frac{l^m \nu_m}{(t+r)^m},
\end{aligned}$$

The second term is straightforward,

$$\begin{aligned}
\left| \mathbb{E} \left[\mathbb{1}_{\{T_i^k \in (s,t)\}} \right] \mathbb{E} \left[\mathbb{1}_{\{T_j^l \in (t+r,u)\}} \right] \right| &\leq \mathbb{E} \left[\mathbb{1}_{\{T_j^l \in (t+r,u)\}} \right] \\
&\leq \frac{l^m \nu_m}{(t+r)^m}.
\end{aligned}$$

Then:

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

using Lemma 3 for the last inequality. \square

Straightforwardly, since $\sum \mu^k$ and $\sum l^m \mu^l$ are summable for $m \geq 0$, we get the following lemma:

Lemma 8. *Suppose that there exists $m > 1$ such that $\nu_m < \infty$. Then,*

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

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

$$\begin{aligned}
|\text{Cov}(\mathbb{1}_{\mathcal{A}}(N), \mathbb{1}_{\mathcal{B}}(N))| & \leq \int_{\mathbb{R}} \left| \text{Cov} \left(N((s, t] | y), N((t+r, u] | y) \right) \right| M_c(dy) \\
& = \int_{-\infty}^t \left| \text{Cov} \left(N((s, t] | y), N((t+r, u] | y) \right) \right| M_c(dy) \\
& = \mathcal{O} \left(\int_{-\infty}^t \frac{1}{(t+r-y)^m} dy \right) \\
& = \mathcal{O} \left(\frac{1}{r^{m-1}} \right).
\end{aligned}$$

This upper bound is valid for any $s, u \in \mathbb{R}$, therefore holds for $\mathcal{A} \in \mathcal{E}_{-\infty}^t, \mathcal{B} \in \mathcal{E}_{t+r}^{\infty}$. \blacksquare

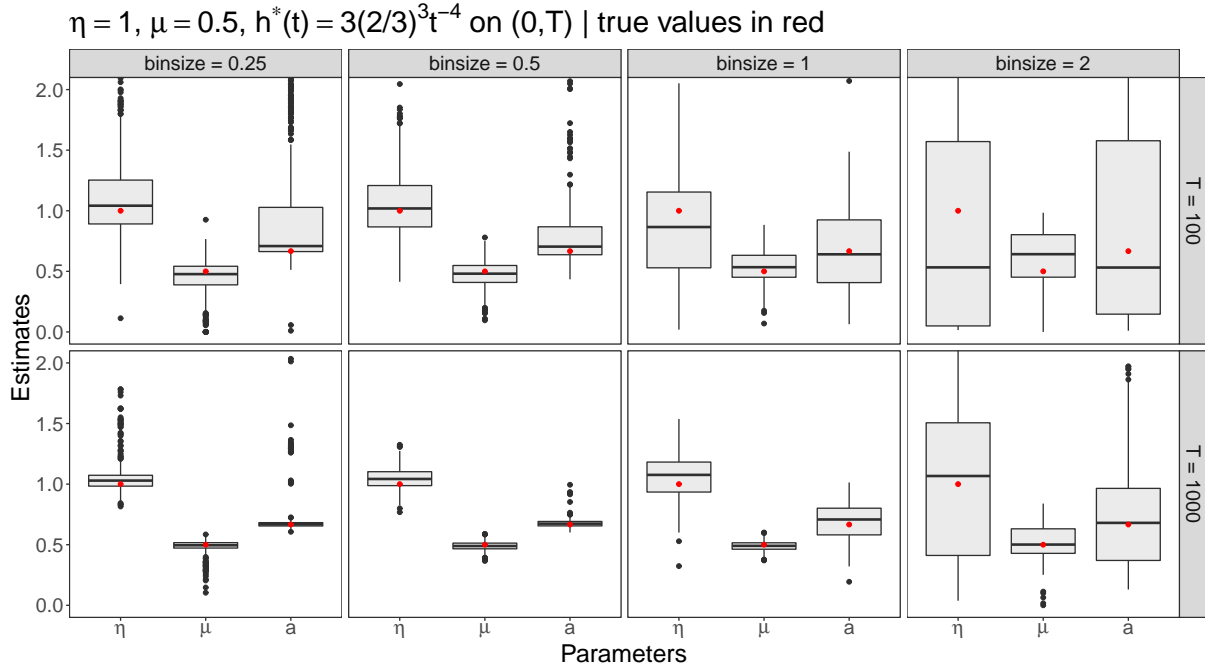


Figure 4: Estimates of parameters η , μ and a for 1,000 simulations of the stationary Hawkes process with kernel $h_3^*(t) = 3a^3 t^{-4}$ on the interval $[0, T]$. True values (in red) are: $\eta = 1$, $\mu = 0.5$, $a = 2/3$.

B Figures of Section 5

References

- [1] A. G. Hawkes. Spectra of Some Self-Exciting and Mutually Exciting Point Processes. *Biometrika*, 58(1):83–90, 1971.
- [2] A. G. Hawkes. Point Spectra of Some Mutually Exciting Point Processes. *J. R. Stat. Soc. Ser. B*, 33(3):438–443, 1971.
- [3] A. G. Hawkes and D. Oakes. A cluster process representation of a self-exciting process. *J. Appl. Probab.*, 11(03):493–503, sep 1974.
- [4] L. Adamopoulos. Cluster models for earthquakes: Regional comparisons. *J. Int. Assoc. Math. Geol.*, 8(4):463–475, 1976.
- [5] Y. Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *J. Am. Stat. Assoc.*, 83(401):9–27, 1988.
- [6] E. S. Chornoboy, L. P. Schramm, and A. F. Karr. Maximum likelihood identification of neural point process systems. *Biol. Cybern.*, 59(4-5):265–275, 1988.
- [7] E. Bacry, I. Mastromatteo, and J.-F. Muzy. Hawkes Processes in Finance. *Mark. Microstruct. Liq.*, 01(01):1550005, jun 2015.

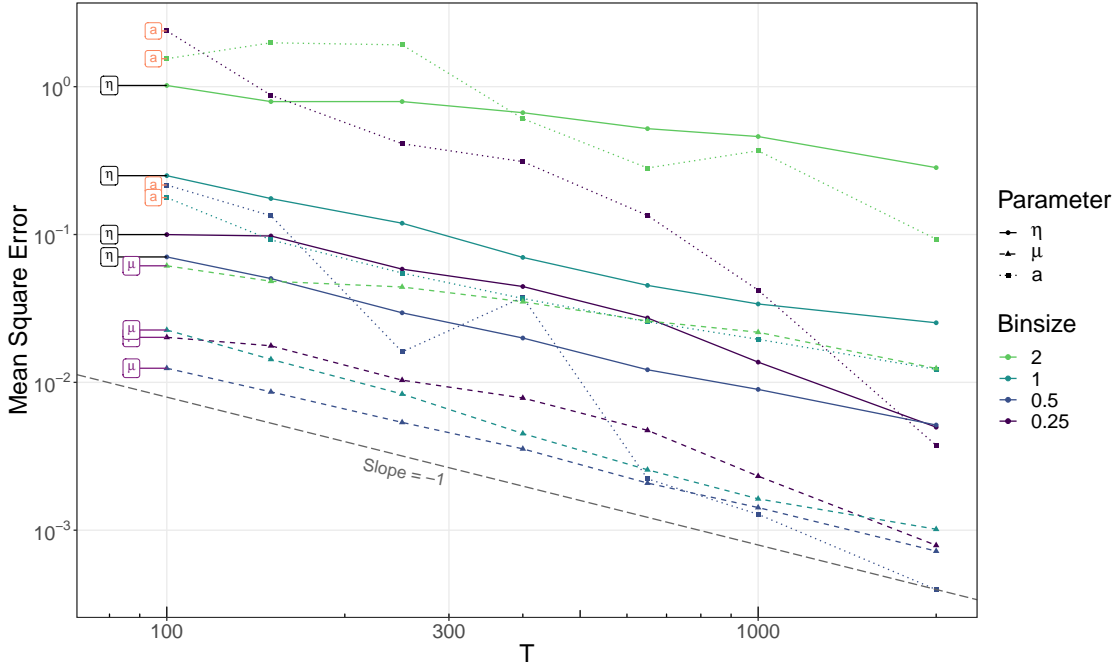


Figure 5: Mean square error of the estimates of parameters η , μ and a for 1,000 simulations of the stationary Hawkes process with kernel $h_3^*(t) = 3a^3t^{-4}$ on the interval $[0, T]$, in log-log scale. The dashed grey line represents the ideal slope of -1 , *i.e.* a rate of convergence of $\mathcal{O}(n^{-1})$.

- [8] P. Reynaud-Bouret and S. Schbath. Adaptive estimation for hawkes processes; Application to genome analysis. *Ann. Stat.*, 38(5):2781–2822, 2010.
- [9] S. Meyer, J. Elias, and M. Höhle. A Space-Time Conditional Intensity Model for Invasive Meningococcal Disease Occurrence. *Biometrics*, 68(2):607–616, 2012.
- [10] Y. Ogata. The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Ann. Inst. Stat. Math.*, 30(1):243–261, 1978.
- [11] T. Ozaki and Y. Ogata. Maximum likelihood estimation of Hawkes’ self-exciting point processes. *Ann. Inst. Stat. Math.*, 31(1):145–155, dec 1979.
- [12] M. Kirchner. Hawkes and INAR(∞) processes. *Stoch. Process. their Appl.*, 126(8): 2494–2525, aug 2016.
- [13] M. Kirchner. An estimation procedure for the Hawkes process. *Quant. Financ.*, 17 (4):571–595, apr 2017.
- [14] J. F. Olson and K. M. Carley. Exact and approximate EM estimation of mutually exciting hawkes processes. *Stat. Inference Stoch. Process.*, 16(1):63–80, 2013.
- [15] S. Wheatley, V. Filimonov, and D. Sornette. The Hawkes process with renewal immigration & its estimation with an EM algorithm. *Comput. Stat. Data Anal.*, 94: 120–135, feb 2016.

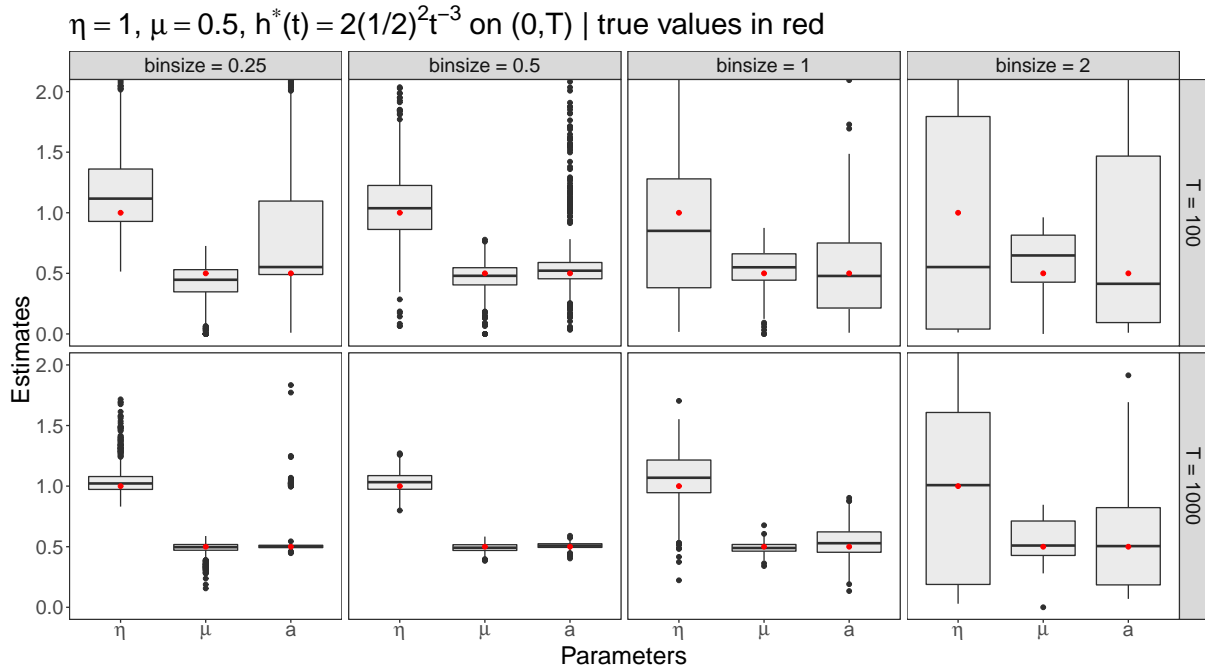


Figure 6: Estimates of parameters η , μ and a for 1,000 simulations of the stationary Hawkes process with kernel $h_2^*(t) = 2a^2 t^{-3}$ on the interval $[0, T]$. True values (in red) are: $\eta = 1, \mu = 0.5, a = 1/2$.

- [16] G. Celeux, D. Chauveau, and J. Diebolt. On Stochastic Versions of the EM Algorithm. Technical report, INRIA, 1995.
- [17] B. Delyon, M. Lavielle, and E. Moulines. Convergence of a stochastic approximation version of the EM algorithm. *Ann. Stat.*, 27(1):94–128, 1999.
- [18] P. Whittle. Some results in time series analysis. *Scand. Actuar. J.*, 1952(1-2):48–60, 1952.
- [19] M. Rosenblatt. A Central Limit Theorem and a Strong Mixing Condition. *Proc. Natl. Acad. Sci.*, 42(1):43–47, jan 1956.
- [20] P. Doukhan. *Mixing: Properties and Examples*. Springer-Verlag, New York, 1994.
- [21] E. Rio. *Asymptotic Theory of Weakly Dependent Random Processes*, volume 80 of *Probability Theory and Stochastic Modelling*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2017.
- [22] M. Westcott. On Existence and Mixing Results for Cluster Point Processes. *J. R. Stat. Soc. Ser. B*, 33(2):290–300, 1971.
- [23] A. Poinas, B. Delyon, and F. Lavancier. Mixing properties and central limit theorem for associated point processes. *Bernoulli*, 25(3):1724–1754, aug 2019.
- [24] K. Dzhaparidze. *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series*. Springer Series in Statistics. Springer New York, New

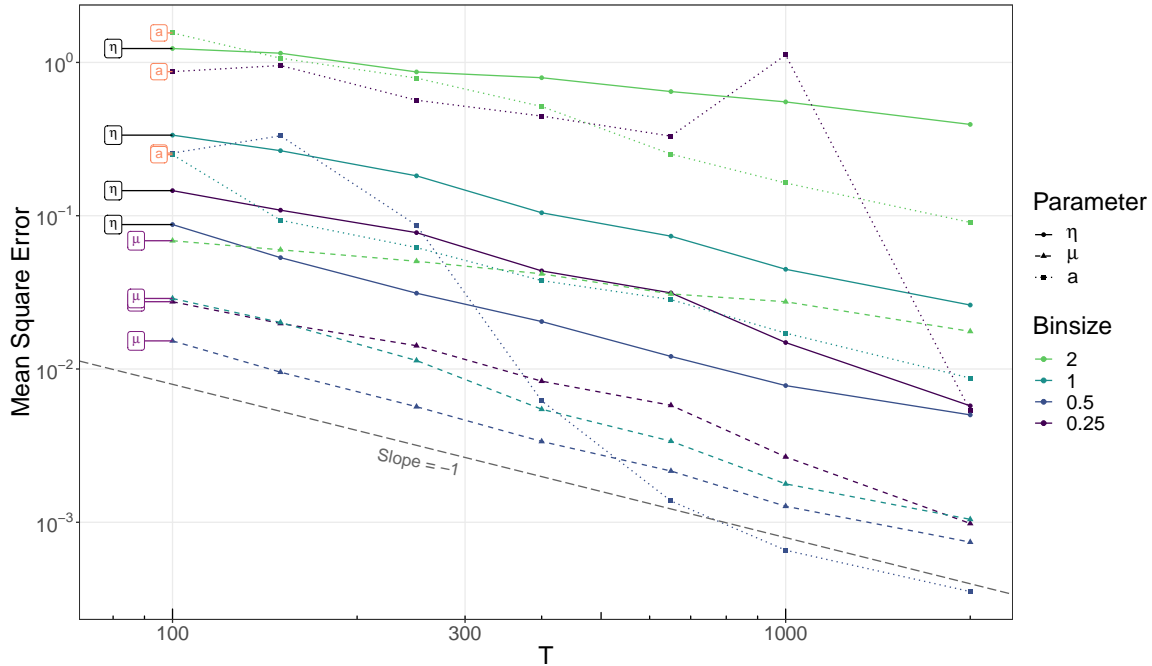


Figure 7: Mean square error of the estimates of parameters η , μ and a for 1,000 simulations of the stationary Hawkes process with kernel $h_2^*(t) = 2a^2t^{-3}$ on the interval $[0, T]$, in log-log scale. The dashed grey line represents the ideal slope of -1 , *i.e.* a rate of convergence of $\mathcal{O}(n^{-1})$.

York, NY, 1986.

- [25] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods*. Probability and its Applications. Springer, New York, 2003.
- [26] Y. Hosoya. *Estimation problems on stationary time series models*. Ph.d. dissertation, Yale University, 1974.
- [27] K. O. Dzhaparidze. A New Method for Estimating Spectral Parameters of a Stationary Regular Time Series. *Theory Probab. Its Appl.*, 19(1):122–132, dec 1974.
- [28] F. Roueff, R. von Sachs, and L. Sansonnet. Locally stationary Hawkes processes. *Stoch. Process. their Appl.*, 126(6):1710–1743, jun 2016.
- [29] X. Shao. A self-normalized approach to confidence interval construction in time series. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 72(3):343–366, jun 2010.
- [30] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019.
- [31] X. Gao and L. Zhu. Functional central limit theorems for stationary Hawkes processes and application to infinite-server queues. *Queueing Syst.*, 90(1-2):161–206, 2018.
- [32] B. von Bahr and C.-G. Esseen. Inequalities for the r -th Absolute Moment of a Sum

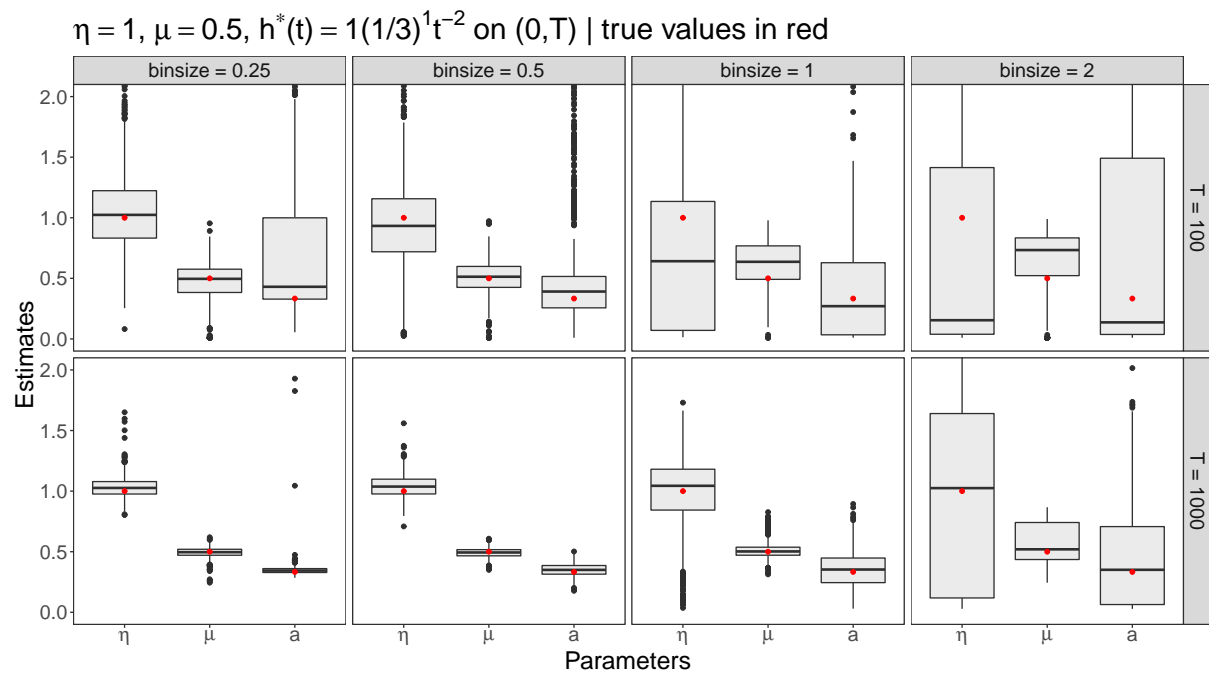


Figure 8: Estimates of parameters η , μ and a for 1,000 simulations of the stationary Hawkes process with kernel $h_1^*(t) = a^1 t^{-2}$ on the interval $[0, T]$. True values (in red) are: $\eta = 1, \mu = 0.5, a = 1/3$.

of Random Variables, $1 \leq r \leq 2$. *Statistics (Ber)*, 36(1):299–303, 1965.

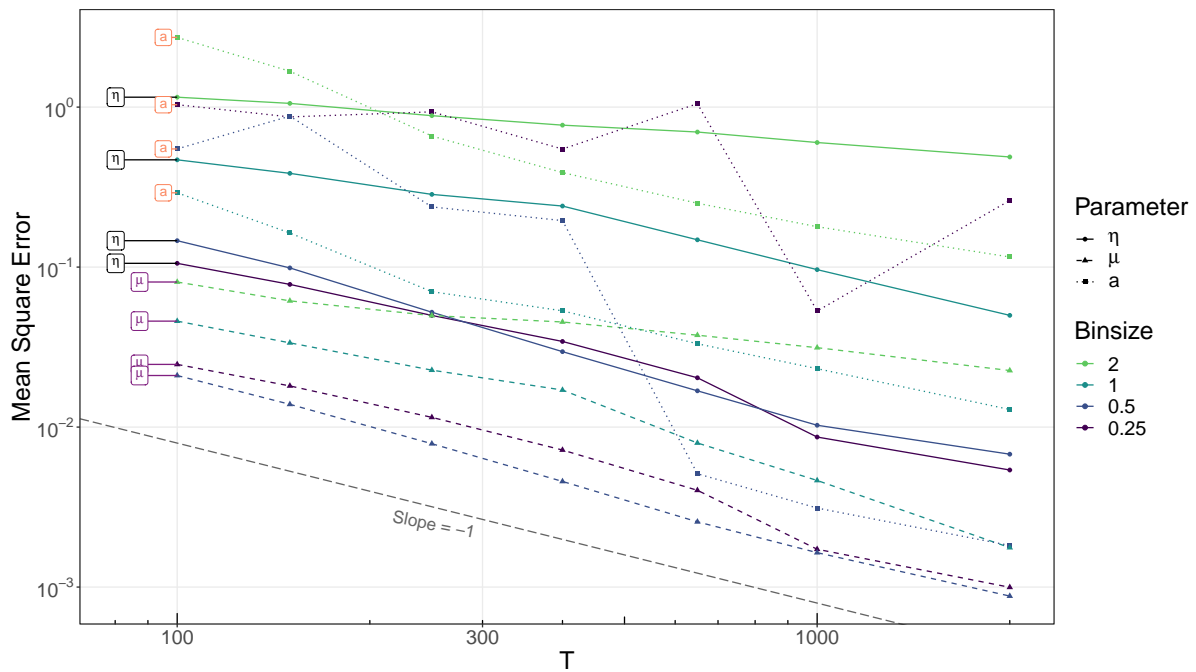


Figure 9: Mean square error of the estimates of parameters η , μ and a for 1,000 simulations of the stationary Hawkes process with kernel $h_1^*(t) = a^1 t^{-2}$ on the interval $[0, T]$, in log-log scale. The dashed grey line represents the ideal slope of -1 , *i.e.* a rate of convergence of $\mathcal{O}(n^{-1})$.