



**HAL**  
open science

## The non-linear discrete-time Hawkes process

Lorick Huang, Mahmoud Khabou

► **To cite this version:**

Lorick Huang, Mahmoud Khabou. The non-linear discrete-time Hawkes process. 2022. hal-03220800v2

**HAL Id: hal-03220800**

**<https://hal.science/hal-03220800v2>**

Preprint submitted on 4 Aug 2022

**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.

# The non-linear discrete-time Hawkes process

Lorick Huang\*

Mahmoud Khabou†

June 2022

## Abstract

The non-linear Hawkes process is a point process for which the occurrence of future events depends on the history, either by excitation or inhibition. This property made it popular in many fields, such as neuro-sciences and social-dynamics. In this paper we propose a tractable hidden Markov chain time series as a discrete-time Hawkes process. Our model allows for cross-excitation and inhibition between components, as well as for exogenous random noise on the intensity. We then prove a convergence theorem when the time scale goes to zero. Finally, we suggest a parametric calibration method for the continuous-time Hawkes process based on the regression on the discrete-time approximation.

## 1 Introduction

In this paper we introduce a multi-variate time series  $(\mathbf{P}_n)_{n \in \mathbb{N}}$  in  $\mathbb{N}^d$  that represents the bin-count sequence of some events on regular time intervals of size  $h > 0$ . Examples of integer valued time series have been studied in the past, such as the  $g$ -functions for  $\{0, 1\}$  valued sequences by Berbee [5] and the integer valued auto-regressive of order  $p$  (INAR( $p$ )) process by Alzaid and Al-Osh[1].

The specificity of our time series resides in its hidden Markov chain structure, in which the intensity depends on the history of the process. In fact, for any  $n \in \mathbb{N}$  and any  $i = 1, \dots, d$ , the bin-count variable  $P_n^i$  is assumed to be the realisation of a Poisson distribution of parameter  $h\psi_i(\mathbf{U}_n)$ , where  $(\mathbf{U}_n)_{n \in \mathbb{N}}$  is a predictable tensor Markov chain, *i.e.*  $\mathbf{U}_n \in \sigma(\mathbf{P}_0, \dots, \mathbf{P}_{n-1}, \mathbf{W}_0, \dots, \mathbf{W}_{n-1})$ , where  $(\mathbf{W}_n)_{n \in \mathbb{N}}$  are some exogenous random perturbations and  $(\psi_i)_{i=1, \dots, d}$  are non-negative functions. The choice of the Poisson distribution for the integer valued random variable is natural: we assume that the time bins are reasonably small, so that one would observe at most one event most of the time, while still allowing for the rare occurrence of two or more events. This, for instance, is not possible with the simpler choice of a Bernoulli distribution (as in [29]), even though that for very small bin size, the two choices are virtually the same. The aforementioned model, called the non-linear discrete-time Hawkes process (by an analogy with the continuous-time Hawkes process made in section 4) allows for self and cross excitation as well as inhibition, which makes it a good candidate for modelling systems in which the occurrence of an event triggers (excitation) or discourages (inhibition) other events.

The main result of this paper is proving that the continuous-time process  $(\mathbf{U}_{\lfloor t/h \rfloor}, \sum_{n=0}^{\lfloor t/h \rfloor} \mathbf{P}_n)_{t \in \mathbb{R}_+}$  converges in distribution to a standard continuous-time non-linear Hawkes process, as the time step  $h$  goes to zero. The linear uni-variate Hawkes process was introduced in [18] as an example of a self-exciting point process. It was then extended to the more general non-linear multivariate process in [8]. Initially used in the context of seismic activity in [25], Hawkes process found applications in many fields such as genome analysis [28], portfolio credit risk [14], micro-structure noise [3] and social networks [4]. Due to its clear branching structure [19] and the possibility of obtaining closed formulae for its asymptotic moments, the linear Hawkes process has been studied extensively in the literature. Linear Hawkes processes with exponentially decaying intensities have been particularly dealt with in many articles. In fact, the memorylessness of the exponential kernel ensures that the intensity is a Markov process that solves a stochastic differential equation (*SDE*). For instance, Giesecke and Kim [17] devised a *self-exciting point process with CIR intensity*

---

\*INSA de Toulouse, IMT UMR CNRS 5219, Université de Toulouse, 135 avenue de Rangueil 31077 Toulouse Cedex 4 France.  
Email: lorick.huang@insa-toulouse.fr

†INSA de Toulouse, IMT UMR CNRS 5219, Université de Toulouse, 135 avenue de Rangueil 31077 Toulouse Cedex 4 France.  
Email: mahmoud.khabou@insa-toulouse.fr

$(N_t)_{t \in \mathbb{R}_+}$  whose intensity  $(\lambda_t)_{t \in \mathbb{R}_+}$  solves the SDE

$$d\lambda_t = \beta(\mu - \lambda_t)dt + \sigma\sqrt{\lambda_t}dW_t + \alpha dN_t,$$

where  $(W_t)_{t \in \mathbb{R}_+}$  is a Brownian motion that can be seen as a persistent market noise. Dassios and Zhao [10] suggested an efficient simulation procedure for this process, based on the fact that its evolution follows the well-studied *CIR* dynamics between jumps.

Despite the explicit formulae obtained thanks to its Markov properties, the simple exponential kernel is a limiting choice. In fact, it does not account for longer memories nor for delays in the excitation.

Similarly, the linearity's tractability comes at an expense: unlike the non-linear Hawkes process, the linear process does not allow for inhibition, nor for bounded excitation. Moreover, the linear process cannot be perturbed by a simple Gaussian diffusion as it can drive the intensity below zero.

The link between time series and linear Hawkes processes have been studied in the seminal works of Kirchner [21], where the linear Hawkes process was proven to be a continuous version of the INAR( $\infty$ ) time series and vice-versa. The proof of convergence relies on the tightness of the sequence in the space of point processes then showing that all the sub-sequential weak limits follow the same Hawkes dynamics.

To the best of our knowledge, the non-linear Hawkes process has no intuitive discrete-time equivalent in the literature. The time series introduced in our paper is a recursively constructed approximation for the Hawkes processes with Erlang kernels (*i.e.* the product of a polynomial and an exponential) that have a Markov structure. We also allow for the presence of an exogenous Gaussian perturbation that happens continuously in time. We then use specific Markov process techniques to show the weak convergence in the space of *càdlàg* functions.

The goal of this discrete-time approximation is twofold; on one hand, most data are recorded on regular time intervals (*e.g.* high-frequency financial data) which makes a discrete-time process more realistic, on the other hand its simulation is more intuitive and tractable than the classic Ogata's thinning algorithm ([24]) albeit at the expense of introducing discretisation bias. Moreover, from a practical point of view, the convergence theorem suggests that the parametric calibration of the time-series yields a good estimation of the Hawkes process' parameters, which is verified numerically.

The paper is structured as follows: Section 2 gives the recursive construction of the discrete-time model and explores its Markov structure. In section 3 we give the definition of the non-linear Hawkes process, to which we add a Gaussian perturbation and then we show that it is a Markov process if the kernels are Erlang functions. Section 4 establishes the link between the discrete-time model and the continuous-time process via the convergence of the infinitesimal generators. Finally, in section 5 we run a parametric calibration of the Hawkes process based on a regression on its discrete-time approximation. The experimental results are given with no theoretical guarantees.

## 2 The discrete-time model

### 2.1 Preliminaries

Throughout this paper,  $d \in \mathbb{N}$  denotes the number of components and  $h > 0$  denotes the width of a time bin. Let  $\phi_1, \dots, \phi_d$  be a family of non-negative functions, called the jump-rate functions. We set the following assumption

**Assumption 2.1.** *The functions  $\phi_1, \dots, \phi_d$  are bounded, i.e.*

$$\|\phi\|_\infty = \max_{i=1, \dots, d} \sup_{s \in \mathbb{R}} \phi_i(s) < \infty.$$

For any vector  $\mathbf{u} = (u_1, \dots, u_d)$  of non-negative real numbers, we say that the vector  $\boldsymbol{\xi}$  follows the multivariate Poisson distribution of parameter  $\mathbf{u}$  if the variables  $\xi_1, \dots, \xi_d$  are independent and

$$\xi_i \sim \text{Pois}(u_i), \quad \forall i = 1, \dots, d.$$

for a more concise notation, we simply write

$$\boldsymbol{\xi} \sim \text{Pois}(\mathbf{u}).$$

We denote by  $N_q$  an independent realization of the multivariate standard normal distribution of dimension  $q$ .

## 2.2 Construction

For the sake of better understanding, we start this section by building our discrete-time model in the uni-variate setting, with a memory parameter  $p$  equal to zero. Knowing the history of the process until  $n-1$ , the bin-count sequence  $P_n$  is supposed to be drawn from a Poisson distribution of parameter  $h\phi(a_n)$ , where  $(a_n)_{n \in \mathbb{N}}$  is the predictable Markov chain that is defined recursively:

$$a_n = \mu(1 - e^{-\beta h}) + e^{-\beta h} a_{n-1} + \alpha P_{n-1} + \sigma \sqrt{h} \mathcal{N}(0, 1),$$

where  $\alpha \in \mathbb{R}$ ,  $\sigma \geq 0$ ,  $\beta > 0$  and  $\phi$  is a positive function, for some fixed initial state. The hidden Markov chain  $a$  can be seen as a *pre-intensity* that takes into account the recent past of the process as well as the exogenous noise (the Gaussian variable), which is turned into the intensity by the action of  $\phi$ .

We simulate the uni-variate bin count sequence  $P$ , its cumulative sum  $H$  as well as the intensity  $\phi((a_n)_{n \in \mathbb{N}})$  (scaled by  $h$ ) and the pre-intensity  $(a_n)_{n \in \mathbb{N}}$  for the jump-rate  $\phi(x) = \ln(1 + e^x) \wedge 40$ , a decay parameter  $\beta = 5$ , a diffusion parameter  $\sigma = 2$  and a baseline pre-intensity  $\mu = 2$ .

Figure 1 gives an illustration in the *self-exciting* case of  $\alpha = 2$  whereas the *self-inhibiting* case  $\alpha = -2$  is illustrated on figure 2.

We now generalize this process to the multivariate setting for a general memory parameter  $p$ . This construction can be seen as an Euler scheme of the continuous-time Hawkes process' intensity, which will be defined in Section 3.

**Definition 2.2.** For an integer  $p \in \mathbb{N}$ , let  $\alpha = (\alpha_q^{ij})_{i,j=1,\dots,d} \in \mathbb{R}^{(p+1) \times d \times d}$ . Let  $\beta = (\beta_{ij})_{i,j=1,\dots,d} \in \mathcal{M}_d(\mathbb{R}_+^*)$ ,

$\omega \in (\mathbb{R}_+^*)^d$ ,  $\sigma \in \mathbb{R}_+^d$  and  $\mu \in \mathbb{R}^d$ .

The discrete time Hawkes process  $\mathbf{H} = (H_n^1, \dots, H_n^d)_{n \in \mathbb{N}}$  is defined by

$$\mathbf{H}_n = \sum_{k=1}^n \mathbf{P}_k,$$

where  $\mathbf{P}_k \sim \text{Pois}(h\phi(\mathbf{a}_k))$ ,  $\forall k \geq 1$  and the "pre-intensities"  $(\mathbf{a}_k)_{k \in \mathbb{N}}$  are defined as

$$a_{k+1}^i = \mu_i + \sum_{j=1}^d \sum_{q=0}^p \alpha_q^{ij} u_{k+1}^{ij,q} + \sigma_i y_{k+1}^i,$$

where  $(u^{ij,q})_{i,j=1,\dots,d}$  and  $(y^i)_{i=1,\dots,d}$  are respectively the discrete cascade of memory terms and the discrete cascade of perturbation terms, defined recursively by:

$$\begin{cases} y_{k+1}^i &= e^{-\omega_i h} y_k^i + \sqrt{h} \mathcal{N}(0, 1) \\ u_{k+1}^{ij,p} &= e^{-\beta_{ij} h} (u_k^{ij,p} + h u_k^{ij,p-1}), \\ &\vdots \\ u_{k+1}^{ij,q} &= e^{-\beta_{ij} h} (u_k^{ij,q} + h u_k^{ij,q-1}), \quad \forall i, j = 1, \dots, d, \\ &\vdots \\ u_{k+1}^{ij,0} &= e^{-\beta_{ij} h} u_k^{ij,0} + P_k^j. \end{cases} \quad (1)$$

With the given initial states  $u_0^{ij,q} = y_0^i = 0$ , for every  $i, j = 1, \dots, d$  and  $q = 0, \dots, p-1$  and  $\mathbf{P}_0 = \mathbf{0}_d$ .

**Remark 2.3.** For each  $i = 1, \dots, d$  the sequence  $(\mathbf{P}_n)$  can be seen as the number of events of a continuous-time counting process  $(\mathbf{N}_t)_{t \geq 0}$  observed in a time bin of size  $h$ , i.e.

$$\mathbf{P}_n = \mathbf{N}_{nh} - \mathbf{N}_{(n-1)h}.$$

That is why a bin-size parameter  $h$  appears in the definition.

The system 1 in Definition 2.2 can be put under the matrix form.

Let

$$\mathbf{U}_k^{ij} = \begin{pmatrix} u_k^{ij,p} \\ \vdots \\ u_k^{ij,0} \end{pmatrix} \in \mathbb{R}^{p+1}, \quad \mathbf{Y}_k = \begin{pmatrix} y_k^1 \\ \vdots \\ y_k^d \end{pmatrix} \in \mathbb{R}^d. \quad (2)$$

The memory and perturbation cascades are perfectly described by the family  $((\mathbf{U}^{ij})_{i,j=1,\dots,d}, \mathbf{Y})$ .

Let  $\boldsymbol{\alpha}^{ij} = (\alpha_p^{ij}, \dots, \alpha_0^{ij})$  and  $K = \begin{pmatrix} 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix} \in \mathcal{M}_{p+1}(\mathbb{R})$ , with 1 on the super-diagonal and 0 elsewhere.

From the definition of  $K$  we set  $M_{ij}^h = e^{-\beta_{ij}h}(I + hK)$ , where  $I$  is the identity matrix.  $\mathbf{e}_{p+1}$  designates the vector of  $\mathbb{R}^{p+1}$  with 0 everywhere and 1 in the last component and  $\Omega^h = \text{diag}(e^{-\omega_1h}, \dots, e^{-\omega_dh})$ .

**Proposition 2.4.** *The cascade  $(\mathbf{U}_n, \mathbf{Y}_n)_{n \in \mathbb{N}}$  where  $\mathbf{U}_n = (\mathbf{U}_n^{ij})_{i,j=1,\dots,d}$  is a Markov chain that follows the matrix dynamics*

$$\begin{cases} \mathbf{U}_{n+1}^{ij} &= M_{ij}^h \mathbf{U}_n^{ij} + P_n^j \mathbf{e}_{p+1} \\ \mathbf{Y}_{n+1} &= \Omega^h \mathbf{Y}_n + \sqrt{h} \mathcal{N}_d \end{cases}$$

Its one step generator  $\mathcal{T}^h$  evaluated at  $(\mathbf{u}, \mathbf{y}) \in ((\mathbb{R}^{p+1})^{d \times d} \times \mathbb{R}^d)$  for a function  $f : ((\mathbb{R}^{p+1})^{d \times d} \times \mathbb{R}^d) \rightarrow \mathbb{R}$  has the expression

$$\begin{aligned} \mathcal{T}^h f(\mathbf{u}, \mathbf{y}) &= \mathbb{E} [f(\mathbf{U}_{n+1}, \mathbf{Y}_{n+1}) | (\mathbf{U}_n, \mathbf{Y}_n) = (\mathbf{u}, \mathbf{y})], \\ &= e^{-h \sum_{i=1}^d \phi_i(\tilde{a}^i(\mathbf{u}, \mathbf{y}))} \mathbb{E}^{\mathcal{N}} \left[ f \left( (M_{ij}^h \mathbf{u}^{ij})_{i,j=1,\dots,d}, \Omega^h \mathbf{y} + \sqrt{h} \mathcal{N}_d \right) \right] \\ &\quad + e^{-h \sum_{i=1}^d \phi_i(\tilde{a}^i(\mathbf{u}, \mathbf{y}))} \sum_{m=1}^d \mathbb{E}^{\mathcal{N}} \left[ f \left( (M_{ij}^h \mathbf{u}^{ij} + \mathbf{1}_{m=j} \mathbf{e}_{p+1})_{i,j=1,\dots,d}, \Omega^h \mathbf{y} + \sqrt{h} \mathcal{N}_d \right) \right] h \phi_m(\tilde{a}^m(\mathbf{u}, \mathbf{y})) \\ &\quad + R_2(\mathbf{u}, \mathbf{y}). \end{aligned}$$

Where the remainder  $R_2$  is

$$R_2(\mathbf{u}, \mathbf{y}) = \mathbb{E}^{\mathcal{N}} \left[ \sum_{i_1 + \dots + i_d \geq 2} f \left( (M_{ij}^h \mathbf{u}^{ij} + i_m \mathbf{e}_{p+1})_{i,j=1,\dots,d}, \Omega^h \mathbf{y} + \sqrt{h} \mathcal{N}_d \right) \prod_{m=1}^d \frac{(h \phi_m(\tilde{a}^m(\mathbf{u}, \mathbf{y})))^{i_m}}{i_m!} e^{-h \phi_m(\tilde{a}^m(\mathbf{u}, \mathbf{y}))} \right]$$

and for every  $i = 1, \dots, d$

$$\tilde{a}^i(\mathbf{u}, \mathbf{y}) = \mu_i + \sigma_i g^i + \sum_{j=1}^d \langle \boldsymbol{\alpha}^{ij}, \mathbf{u}^{ij} \rangle.$$

$\mathbb{E}^{\mathcal{N}}$  stands for the integration with respect to the standard normal distribution in  $d$ -dimensions.

*Proof.* The matrix equation is simply a compact formulation of the recursive definition of  $(g^i, b^{ij,q})_{i,j=1,\dots,d, q=1,\dots,p}$ .

For the one-step generator that knowing the state  $\mathbf{U}_n = \mathbf{u}$  and  $\mathbf{Y}_n = \mathbf{y}$ ,  $P_n^1, \dots, P_n^d$  are independent Poisson variables where

$$\begin{aligned} P_n^i &\sim \mathcal{P} \left( h \phi_i \left( \mu_i + \sigma_i g_n^i + \sum_{j=1}^d \sum_{q=0}^p \alpha_q^{ij} b_n^{ij,q} \right) \right), \\ &\sim \mathcal{P} \left( h \phi_i \left( \mu_i + \sigma_i g^i + \sum_{j=1}^d \langle \boldsymbol{\alpha}^{ij}, \mathbf{u}^{ij} \rangle \right) \right), \\ &\sim \mathcal{P} (h \phi_i (\tilde{a}^i(\mathbf{u}, \mathbf{y}))). \end{aligned}$$

The expectation operator  $\mathbb{E}$  stands for the integration with respect to both the Poisson variable and the Gaussian variable. We denote by  $\mathbb{E}^P$  the sum against the Poisson distributions and  $\mathbb{E}^{\mathcal{N}}$  the expectation with respect to the multivariate standard normal distribution. Since the Poisson variables and the Gaussian variables are built

in two separate spaces we have that:

$$\begin{aligned} \mathcal{T}_h f(\mathbf{u}, \mathbf{y}) &= \mathbb{E} [f(\mathbf{U}_{n+1}, \mathbf{Y}_{n+1}) | (\mathbf{U}_n, \mathbf{Y}_n) = (\mathbf{u}, \mathbf{y})], \\ &= \mathbb{E}^{\mathcal{N}} \left[ \mathbb{E}^P \left[ f \left( (M_{ij}^h \mathbf{U}_n^{ij} + P_n^j \mathbf{e}_{p+1})_{i,j=1,\dots,d}, \Omega^h \mathbf{Y}_n + \sqrt{h} \mathcal{N}_d \right) | (\mathbf{U}_n, \mathbf{Y}_n) = (\mathbf{u}, \mathbf{y}) \right] \right], \\ &= \mathbb{E}^{\mathcal{N}} \left[ \sum_{i_1, \dots, i_d=0}^{+\infty} f \left( (M_{ij}^h \mathbf{u}^{ij} + i_j \mathbf{e}_{p+1})_{i,j=1,\dots,d}, \Omega^h \mathbf{y} + \sqrt{h} \mathcal{N}_d \right) \mathbb{P} [P_n^1 = i_1, \dots, P_n^d = i_d | (\mathbf{U}_n, \mathbf{Y}_n) = (\mathbf{u}, \mathbf{y})] \right]. \end{aligned}$$

Keeping in mind that  $P_n^i \sim \mathcal{P}(h\phi_i(\tilde{a}^i(\mathbf{u}, \mathbf{y})))$  are independent we have that

$$\begin{aligned} \mathcal{T}_h f(\mathbf{u}, \mathbf{y}) &= \sum_{i_1, \dots, i_d=0}^{+\infty} \mathbb{E}^{\mathcal{N}} \left[ f \left( (M_{ij}^h \mathbf{u}^{ij} + i_j \mathbf{e}_{p+1})_{i,j=1,\dots,d}, \Omega^h \mathbf{y} + \sqrt{h} \mathcal{N}_d \right) \prod_{j=1}^d \frac{(h\phi_j(\tilde{a}^j(\mathbf{u}, \mathbf{y})))^{i_j}}{i_j!} e^{-h\phi_j(\tilde{a}^j(\mathbf{u}, \mathbf{y}))} \right], \\ &= e^{-h \sum_{i=1}^d \phi_i(\tilde{a}^i(\mathbf{u}, \mathbf{y}))} \mathbb{E}^{\mathcal{N}} \left[ f \left( (M_{ij}^h \mathbf{u}^{ij})_{i,j=1,\dots,d}, \Omega^h \mathbf{y} + \sqrt{h} \mathcal{N}_d \right) \right] \\ &\quad + e^{-h \sum_{i=1}^d \phi_i(\tilde{a}^i(\mathbf{u}, \mathbf{y}))} \sum_{m=1}^d \mathbb{E}^{\mathcal{N}} \left[ f \left( (M_{ij}^h \mathbf{u}^{ij} + \mathbf{1}_{m=j} \mathbf{e}_{p+1})_{i,j=1,\dots,d}, \Omega^h \mathbf{y} + \sqrt{h} \mathcal{N}_d \right) \right] h\phi_m(\tilde{a}^m(\mathbf{u}, \mathbf{y})) \\ &\quad + R_2(\mathbf{u}, \mathbf{y}). \end{aligned}$$

the last equality is obtained by separating according to the three scenarii

- Exactly zero jumps.
- One jump at exactly one component  $j \in \llbracket 1, d \rrbracket$ .
- Two or more jumps.

□

In order to illustrate the discrete time Hawkes process we simulate it for  $h = 0.2$ ,  $d = 2$  and  $p = 1$ . The jump-rate functions are chosen to be  $\phi_1(x) = \max(x, 0) \wedge 40$  and  $\phi_2(x) = \log(1 + e^x) \wedge 40$ . The decay rates are chosen to be  $\beta_{11} = \beta_{12} = 5$ ,  $\beta_{21} = \beta_{22} = 6$  and the baseline parameters are  $(\mu_1, \mu_2) = (2, 3)$ . The results are shown on figures 3 and 4 in the appendix 7.

Notice how  $H_1$  (or  $H_2$ ) has jumps whose sizes are larger than one. This is possible because the jumps are chosen to be Poisson variables (along the lines of [20]), unlike the alternative choice of a Bernoulli variable (cf. [29]). Even though the two choices are equivalent when  $h \rightarrow 0$ , the Poisson approach allows multiple jumps within a single time bin and allows the possibility  $h\phi_i(a_k^i) > 1$ .

**Remark 2.5.** *Even though in Proposition 2.4 the cascade  $(\mathbf{U}_n, \mathbf{Y}_n)_{n \in \mathbb{N}}$  were proven to be a Markov chain, it is straightforward to show that  $(\mathbf{U}_n, \mathbf{Y}_n, \mathbf{H}_n)_{n \in \mathbb{N}}$  is also a Markov chain. From now on, we mainly focus on the pre-intensity, because it already contains the jumps of  $\mathbf{H}$ .*

## 2.3 Stability

We show that if the jump rates are bounded, the pre-intensities  $\left( \left( \mathbf{U}_k^{ij} \right)_{i,j=1,\dots,d} \right)_{k \in \mathbb{N}}$  have bounded averages and are thus almost surely finite.

**Proposition 2.6.** *Let  $(\mathbf{U}_n, \mathbf{Y}_n)_{n \in \mathbb{N}}$  be the Markov chain defined in 2 following the recursive formula of Proposition 2.4 and assume assumption 2.1 is in force. Then there is a positive constant  $C$  independent of  $n$  such that*

$$\mathbb{E} \left[ \sum_{i,j=1}^d \sum_{q=0}^p u_n^{ij,q} + |y_n^i| \right] \leq C$$

for any  $n \in \mathbb{N}$ .

*Proof.* We start with  $\mathbf{Y}_n$ , using the triangular inequality and the mean of the folded normal distribution

$$\mathbb{E}[|y_{k+1}^i|] \leq e^{-\omega_i h} \mathbb{E}[|y_k^i|] + \sqrt{\frac{2h}{\pi}},$$

hence, using the discrete version of Gronwall's lemma:

$$\mathbb{E}[|y_{n+1}^i|] \leq \sqrt{\frac{2h}{\pi}} \frac{1}{1 - e^{-\omega_i h}}.$$

For  $\mathbf{U}_n$ , we set the decreasing positive sequence  $(\pi_0^{ij}, \dots, \pi_q^{ij})$  and we define  $D_{k+1}^{ij} = \sum_{q=0}^p \mathbb{E}[u_{k+1}^{ij,q}] \pi_q^{ij} \beta_{ij}^{q+1}$ . The dynamics of  $\mathbf{U}_{k+1}^{ij}$  yield

$$\begin{aligned} D_{k+1}^{ij} &= e^{-\beta_{ij} h} \left( \sum_{q=0}^p \mathbb{E}[u_k^{ij,q}] \pi_q^{ij} \beta_{ij}^{q+1} + h \sum_{q=1}^p \mathbb{E}[u_k^{ij,q-1}] \pi_q^{ij} \beta_{ij}^{q+1} \right) + \mathbb{E}[P_k^j] \pi_0^{ij} \beta_{ij}, \\ &= e^{-\beta_{ij} h} \left( \mathbb{E}[u_k^{ij,p}] \pi_p^{ij} \beta_{ij}^{p+1} + \sum_{q=0}^{p-1} \mathbb{E}[u_k^{ij,q}] \left( \pi_q^{ij} + h \beta_{ij} \pi_{q+1}^{ij} \right) \beta_{ij}^{q+1} \right) + \mathbb{E}[P_k^j] \pi_0^{ij} \beta_{ij}, \\ &\leq e^{-\beta_{ij} h} \left( \mathbb{E}[u_k^{ij,p}] \pi_p^{ij} \beta_{ij}^{p+1} + \sum_{q=0}^{p-1} \mathbb{E}[u_k^{ij,q}] \left( \pi_q^{ij} + h \beta_{ij} \pi_{q+1}^{ij} \right) \beta_{ij}^{q+1} \right) + h \|\phi\|_{\infty} \pi_0^{ij} \beta_{ij}. \end{aligned}$$

Using the fact that  $1 + x \leq e^x$  we get

$$\begin{aligned} e^{-\beta_{ij} h} \left( \pi_q^{ij} + h \beta_{ij} \pi_{q+1}^{ij} \right) &= \pi_q^{ij} e^{-\beta_{ij} h} \left( 1 + h \beta_{ij} \frac{\pi_{q+1}^{ij}}{\pi_q^{ij}} \right), \\ &\leq \pi_q^{ij} \exp \left( h \beta_{ij} \left( \frac{\pi_{q+1}^{ij}}{\pi_q^{ij}} - 1 \right) \right), \end{aligned}$$

where  $\frac{\pi_{q+1}^{ij}}{\pi_q^{ij}} - 1 < 0$ . Finally, we obtain that

$$D_{k+1}^{ij} \leq \kappa_{ij} D_k^{ij} + h \|\phi\|_{\infty} \pi_0^{ij} \beta_{ij}$$

with

$$\kappa_{ij} = \max_{q=0, \dots, p-1} \left\{ \exp \left( h \beta_{ij} \left( \frac{\pi_{q+1}^{ij}}{\pi_q^{ij}} - 1 \right) \right), e^{-\beta_{ij} h} \right\} < 1,$$

hence the result.  $\square$

The result above does not prove stability in variation (*cf.* [8]), but it ensures that the self-excitation (in case  $\alpha$  is positive) is not explosive. In fact, the result of Proposition 2.6 entails that (using Markov's inequality)

$$\mathbb{P}[u_n^{ij,q} \geq M] \leq \frac{C}{M}, \quad \forall n \in \mathbb{N}.$$

The result also ensures that the sequence  $(\mathbf{U}_n, \mathbf{Y}_n)_{n \in \mathbb{N}}$  is tight, thus having weakly convergent sub-sequence, which in turn can be used to prove the convergence towards a stationary distribution. We chose not to prove it in this paper, as the long-time behaviour has already been studied in [11] for the continuous-time Hawkes process, using classical Markov process techniques.

### 3 The continuous-time Hawkes process

#### 3.1 Preliminaries

Let  $\Omega^d$  be the space of configurations, where

$$\Omega := \left\{ \omega = \sum_{i=1}^n \delta_{(t_i, \theta_i)}, 0 = t_0 < t_1 < \dots < t_n, \theta_i \in \mathbb{R}_+, n \in \mathbb{N} \cup \{+\infty\} \right\}.$$

Let  $\mathcal{F}$  be the  $\sigma$ -field associated to the vague topology on  $\Omega^d$ , and  $\mathbb{P}^\Gamma$  the Poisson measure under which the family

$$\mathbf{\Gamma} = (\Gamma^j)_{j=1, \dots, d}$$

where

$$\Gamma^j([0, t] \times [0, b]) (\omega^j) := \omega^j([0, t] \times [0, b]), \quad (t, b) \in \mathbb{R}_+^2, \quad j = 1, \dots, d$$

is a family of independent homogeneous Poisson processes with intensity measures  $dt \otimes d\theta$ . We set  $\mathbb{F}^\Gamma = (\mathcal{F}_t^\Gamma)_{t \geq 0}$  to be the natural filtration of  $(\Gamma^j)_{j=1, \dots, d}$ . The expected value with respect to the Poisson measure is denoted by  $\mathbb{E}^\Gamma$ .

On a separate space  $\Omega'$ , we construct a standard  $d$ -variate Brownian motion  $\mathbf{W} = (W^1, \dots, W^d)$  and we denote its filtration by  $\mathbb{F}^\mathbf{W} = (\mathcal{F}_t^\mathbf{W})_{t \geq 0}$ , and by  $\mathbb{P}^\mathbf{W}$  the associated probability. The expected value with respect to the Brownian measure is denoted by  $\mathbb{E}^\mathbf{W}$ .

We define  $\mathbb{P} = \mathbb{P}^\Gamma \otimes \mathbb{P}^\mathbf{W}$  and  $\mathbb{F} = \mathbb{F}^\Gamma \vee \mathbb{F}^\mathbf{W}$ . The expected value with respect to  $\mathbb{P}$  is denoted by  $\mathbb{E}$ .

### 3.2 Definition and construction

We start by defining the non-linear multivariate Hawkes process in the general framework.

For  $i = 1, \dots, d$ , consider the events times  $\tau_1^i, \tau_2^i, \dots$  associated with the  $i$ -th component and define the counting process

$$N_t^i = \sum_{i \geq 1} \mathbf{1}_{\tau_i^i \leq t}.$$

To  $\mathbf{N} = (N^i)_{i=1, \dots, d}$  we associate a predictable intensity vector  $\boldsymbol{\lambda} = (\lambda^i)_{i=1, \dots, d}$  such that

$$\mathbb{P}[N_{t+dt}^i - N_t^i = 1 | \mathcal{F}_{t-}] = \lambda_t^i dt$$

which measures how likely it is for  $N^j$  to jump between  $t$  and  $t + dt$ , right before  $t$ .

Let  $(\delta_{ij})_{i, j=1, \dots, d}$  be a set of real integrable (not necessarily non-negative) functions on  $\mathbb{R}_+$  and  $\phi$  as defined in subsection 2.

The process  $\mathbf{N}$  is called a non-linear Hawkes process of memory kernel  $\delta$  and jump rate  $\phi$  if its intensity follows the dynamics

$$\begin{aligned} \lambda_t^i &= \phi_i \left( \mu_i + \int_{[0, t)} \left( \sum_{j=1}^d \delta_{ij}(t-s) dN_s^j \right) + \sigma_i e^{-\omega_i(t-s)} dW_s^i \right), \\ &= \phi_i \left( \mu_i + \sum_{j=1}^d \sum_{\tau_n^j < t} \delta_{ij}(t-\tau_n^j) + \int_{[0, t)} \sigma_i e^{-\omega_i(t-s)} dW_s^i \right), \end{aligned}$$

for some  $\omega_i > 0$ , for every  $i = 1, \dots, d$ . [8] constructed the non-linear Hawkes process as the result of the embedding from a Poisson measure. We give the result in the following theorem.

**Theorem 3.1.** *Let  $\mathbf{\Gamma}$  a  $d$ -dimension Poisson measure on  $\mathbb{R}_+^2$  and  $\mathbf{W}$  a standard  $d$ -variate Brownian motion. Let  $\phi$  and  $\delta$  be functions as defined in the beginning of this subsection. Let  $\mu \in \mathbb{R}^d$ ,  $\boldsymbol{\omega} \in (\mathbb{R}_+^*)^d$  and  $\boldsymbol{\sigma} \in \mathbb{R}_+^d$ . The following SDE has a unique solution  $(\mathbf{N}, \boldsymbol{\lambda})$  with  $\mathbf{N}$   $\mathcal{F}$ -measurable and  $\boldsymbol{\lambda}$   $\mathcal{F}$ -predictable*

$$\begin{cases} N_t^i = \int_{(0, t] \times \mathbb{R}_+} \mathbf{1}_{\{\theta \leq \lambda_s^i\}} \Gamma^i(ds, d\theta), & t \geq 0, \quad i = 1, \dots, d, \\ \lambda_t^i = \phi_i \left( \mu_i + \int_{[0, t)} \left( \sum_{j=1}^d \delta_{ij}(t-s) dN_s^j \right) + \sigma_i e^{-\omega_i(t-s)} dW_s^i \right) & t \geq 0 \quad i = 1, \dots, d. \end{cases} \quad (3)$$

**Remark 3.2.** In [8], the assumption that the functions  $\phi_1, \dots, \phi_d$  are bounded is relaxed and replaced by the condition that  $\phi_i$  is  $l_i$ -Lipschitz and that the matrix

$$\left( l_i \int_0^{+\infty} |\delta_{ij}(s)| ds \right)_{i,j=1,\dots,d}$$

has a spectral radius strictly less than one. This guarantees that  $\mathbf{N}$  exists on  $\mathbb{R}_+$  and has a stationary distribution.

In our definition a Gaussian perturbation was added to the original definition of the intensity. This can be seen as the effect of exogenous factors on each component of the system, that do not occur on the jump times of a point process, but continuously in time. On the other hand, the unperturbed Hawkes intensity follows a piece-wise deterministic dynamics.

In the following section, we choose the memory kernels from a specific family of functions and we show how this choice guarantees that the Hawkes process is Markov, without being too restrictive.

### 3.3 Markov properties of the Hawkes process with memory kernels given by an Erlang function

From now on, we restrict the choice for the memory kernels to be in the Erlang functions family

$$\delta_{ij}(s) = \left( \sum_{q=0}^p \alpha_q^{ij} \frac{s^q}{q!} \right) e^{-\beta_{ij}s},$$

for all  $s \geq 0$  and  $\forall i, j = 1, \dots, d$ . Thus, the intensity takes the form

$$\lambda_t^i = \phi_i \left( \mu_i + \sum_{j=1}^d \int_{[0,t)} \left( e^{-\beta_{ij}(t-s)} \sum_{q=0}^p \alpha_q^{ij} \frac{(t-s)^q}{q!} dN_s^j \right) + \sigma_i \int_{[0,t)} e^{-\omega_i(t-s)} dW_s^i \right).$$

When the memory kernels are Erlang functions, the stability condition in remark 3.2 becomes

**Assumption 3.3.** The functions  $\phi_i$  are  $l_i$ -Lipschitz for  $i = 1, \dots, d$  and the matrix

$$\left( l_i \sum_{q=0}^p \frac{|\alpha_0^{ij}|}{\beta_{ij}^{q+1}} \right)_{i,j=1,\dots,d}$$

has a spectral radius strictly less than one.

**Remark 3.4.** This is a special class of Erlang memory kernels. The more general approach would be to make the  $(\beta_i)_{i=1,\dots,d}$  dependent on  $q$ , along the lines of [11]. This choice would allow to approximate every  $L^1(\mathbb{R}_+, \mathbb{R})$  kernel, hence approximating any Hawkes process with an integrable kernel. In fact, it has been proven in [27] that for any  $T$  there is a constant  $C_T$  such that for two Hawkes processes  $\mathbf{N}$  and  $\tilde{\mathbf{N}}$  of memory kernels  $\delta$  and  $\tilde{\delta}$  respectively, we have that

$$\sum_{i=1}^d \mathbb{E} \left[ \|N^i - \tilde{N}^i\|_{TV} \right] \leq C_T \sum_{i,j=1}^d \int_0^T |\delta_{ij}(s) - \tilde{\delta}_{ij}(s)| ds,$$

where  $\|\cdot\|_{TV}$  is the total variation distance on  $[0, T]$ .

The results in this paper are entirely adaptable to those kernels, we chose however to write them down solely for  $\beta$  independent from  $q$  because the computations are less tedious.

We now show that, the vector formed by the continuous-time memory and perturbation cascades

$$v_t^{ij,q} := \int_{[0,t)} \frac{(t-s)^q}{q!} e^{-\beta_{ij}(t-s)} dN_s^j, \quad (4)$$

and

$$z_t^i := \int_{[0,t)} e^{-\omega_i(t-s)} dW_s^i, \quad (5)$$

is a  $(p+1)d^2 + d$ -variate Markov process that solves a certain SDE.

**Theorem 3.5.** Let  $\mathbf{N}$  be a  $d$ -variate Hawkes process with Erlang memory kernels. For every  $i = 1, \dots, d$ , set  $\boldsymbol{\alpha}^{ij} = (\alpha_p^{ij}, \dots, \alpha_0^{ij}) \in \mathbb{R}^{p+1}$ .

Let  $\mathbf{V}_t^{ij} = (v_t^{ij,p}, \dots, v_t^{ij,0}) \in \mathbb{R}^{p+1}$  and  $\mathbf{Z}_t = (z_t^1, \dots, z_t^d) \in \mathbb{R}^d$ , such that

$$\begin{aligned}\lambda_t^i &= \phi_i \left( \mu_i + \sigma_i z_t^i + \sum_{j=1}^d \langle \boldsymbol{\alpha}^{ij}, \mathbf{V}_t^{ij} \rangle \right), \\ &= \phi_i (\tilde{a}^i(\mathbf{V}_t, \mathbf{Z}_t)).\end{aligned}$$

Then,  $(\mathbf{V}, \mathbf{Z}) = ((\mathbf{V}^{ij})_{i,j=1,\dots,d}, (z^i)_{i=1,\dots,d})$  is a Markov process that follows the SDE:

$$\begin{cases} d\mathbf{V}_t^{ij} &= (K - \beta_{ij}I)\mathbf{V}_t^{ij} dt + dN_t^j \mathbf{e}_{p+1}, \\ d\mathbf{Z}_t &= -\text{diag}(\omega_1, \dots, \omega_d)\mathbf{Z}_t dt + d\mathbf{W}_t \end{cases}$$

where  $\mathbf{e}_{p+1} = (0, \dots, 0, 1)$  and  $\mathbf{W} = (W^1, \dots, W^d)$ .

*Proof.* For the auxiliary processes  $v_t^{ij,k}$  for fixed  $i, j \in \llbracket 1, \dots, d \rrbracket$  and  $q \in \llbracket 1, \dots, p \rrbracket$  we have

$$\begin{aligned}v_t^{ij,q} e^{\beta_{ij}t} &= \frac{1}{q!} \int_0^t (t-s)^q e^{\beta_{ij}s} dN_s^j, \\ &= \frac{1}{q!} \sum_{l=0}^q C_q^l t^l \int_0^t (-s)^{q-l} e^{\beta_{ij}s} dN_s^j,\end{aligned}$$

which yields after differentiation

$$\left( dv_t^{ij,q} + \beta_{ij} v_t^{ij,q} dt \right) e^{\beta_{ij}t} = \frac{1}{q!} \left( dt \sum_{l=0}^q C_q^l t^{l-1} \int_0^t (-s)^{q-l} e^{\beta_{ij}s} dN_s^j + \sum_{l=0}^q C_q^l t^l \int_0^t (-s)^{q-l} e^{\beta_{ij}s} dN_s^j \right),$$

and since

$$C_q^l = \frac{q!}{(q-l)!(l-1)!} = q \frac{(q-1)!}{(q-1-(l-1))!(l-1)!} = q C_{q-1}^{l-1},$$

we have

$$\begin{aligned}\left( dv_t^{ij,q} + \beta_{ij} v_t^{ij,q} dt \right) e^{\beta_{ij}t} &= \frac{1}{q!} dt \sum_{l=1}^q q C_{q-1}^{l-1} t^{l-1} \int_0^t (-s)^{q-l} e^{\beta_{ij}s} dN_s^j, \\ &= \frac{1}{(q-1)!} dt \sum_{l=0}^{q-1} C_{q-1}^l t^l \int_0^t (-s)^{q-1-l} e^{\beta_{ij}s} dN_s^j, \\ &= \frac{1}{(q-1)!} dt \int_0^t (t-s)^{q-1} e^{\beta_{ij}s} dN_s^j,\end{aligned}$$

and after re-arranging the terms

$$\begin{aligned}dv_t^{ij,q} &= \left( -\beta_{ij} v_t^{ij,q} + \int_0^t (t-s)^{q-1} e^{-\beta_{ij}(t-s)} dN_s^j \right) dt, \\ &= \left( -\beta_{ij} v_t^{ij,q} + v_t^{ij,q-1} \right) dt.\end{aligned}$$

For the last term

$$v_t^{ij,0} = \int_0^t e^{-\beta_{ij}(t-s)} dN_s^j,$$

a straightforward differentiation yields

$$dv_t^{ij,0} = -\beta_{ij} v_t^{ij,0} dt + dN_t^j.$$

The SDE for  $D^i$  is obtained identically, in fact, it is a multi-variate Ornstein-Uhlenbeck process.  $\square$

**Remark 3.6.** Note that the vector process formed by the memory and perturbation cascades as well as the point process  $(\mathbf{V}, \mathbf{Z}, \mathbf{N})$  is also a Markov process. But since it makes the notations and the computations more tedious, we simply consider  $(\mathbf{V}, \mathbf{Z})$  that already contains all the jump information. In fact, the jumps of  $N^i$  are exactly the points of the underlying Poisson measure under the curve  $\phi_i(\tilde{a}^i(\mathbf{V}_t, \mathbf{Z}_t))$ :

$$dN_t^i = \int_{\mathbb{R}_+} \mathbb{1}_{\{\theta \leq \phi_i(\tilde{a}^i(\mathbf{V}_t, \mathbf{Z}_t))\}} \Gamma^i(dt, d\theta).$$

This remark is the continuous-time analogue of Remark 2.5.

In the next proposition we give the infinitesimal generator for the process  $(\mathbf{V}, \mathbf{Z})$  defined in the last theorem.

**Proposition 3.7.** We recall that  $\mathbf{e}_{p+1}$  is the vector that has 1 at the  $p+1$ -th entry and 0 elsewhere. Let

$$f: (\mathbb{R}^{p+1})^{d \times d} \times \mathbb{R}^d \rightarrow \mathbb{R} \\ (\mathbf{v}, \mathbf{z}) \mapsto f(\mathbf{v}, \mathbf{z}).$$

be a  $\mathcal{C}^1$  function in the first variable and  $\mathcal{C}^2$  in the second variable. Then the infinitesimal generator  $\mathcal{A}$  of the process  $(\mathbf{V}, \mathbf{Z})$  at  $f$  is

$$\begin{aligned} \mathcal{A}f(\mathbf{v}, \mathbf{z}) &= \sum_{i=1}^d \frac{\partial^2 f}{\partial z_i^2}(\mathbf{v}, \mathbf{z}) + \sum_{i,j=1}^d \langle (K - \beta_{ij}I)\mathbf{v}_{ij}, \nabla_{\mathbf{v}_{ij}} f(\mathbf{v}, \mathbf{z}) \rangle - \langle \text{diag}(\omega_1, \dots, \omega_d)\mathbf{z}, \nabla_{\mathbf{z}} f(\mathbf{v}, \mathbf{z}) \rangle \\ &\quad + \sum_{m=1}^d \phi_m(\tilde{a}^m(\mathbf{v}, \mathbf{z})) (f((\mathbf{v}_{ij} + \mathbb{1}_{j=m}\mathbf{e}_{p+1})_{i,j=1,\dots,d}, \mathbf{z}) - f(\mathbf{v}, \mathbf{z})), \end{aligned}$$

where  $\mathbf{v} = (\mathbf{v}_{ij})_{i,j=1,\dots,d}$ ,  $\mathbf{z} = (z_1, \dots, z_d)$  and  $\tilde{a}^m(\mathbf{v}, \mathbf{z}) = \mu_m + \sigma_m z_m + \sum_{j=1}^d \langle \boldsymbol{\alpha}^{mj}, \mathbf{v}_{mj} \rangle$ .

*Proof.* For the unperturbed uni-variate case, the expression of the generator can be found in [11]. The second derivatives are a result of the Gaussian perturbation. This is more precisely the generator of a multi-variate Ornstein Uhlenbeck process.  $\square$

For the linear Hawkes process, ( $\phi_i(x) = x, \forall i = 1, \dots, d$  and  $\delta_{ij} \geq 0, \forall i, j = 1, \dots, d$ ) the Gaussian perturbation with constant volatilities poses a problem, since the intensity is no longer guaranteed to remain non-negative. That is why the volatility must be a variable function in the intensity. For instance, the intensity of an uni-variate linear Hawkes process with an exponential kernel  $\delta(t) = \alpha e^{-\beta t}$  is assumed to solve the SDE

$$d\lambda_t = \beta(\mu - \lambda_t)dt + \alpha dN_t + \sigma \sqrt{\lambda_t} dW_t$$

to ensure that the intensity remains positive (cf. [17] and [10]), with the upper bound  $\sigma^2 \leq 2\mu\beta$ . The existence of these "jump-diffusion" processes as well as other results on the affine point processes can be found in [12].

## 4 Convergence of the discrete-time Hawkes process to the continuous-time Hawkes process

### 4.1 Preliminaries

Let  $\mathcal{D}(\mathbb{R}_+, E)$  be the space of right continuous processes with left limits (*càdlàg*) on a metric space  $E$ . We denote by " $\Rightarrow$ " the weak convergence of processes in the Skorokhod topology (cf. [6] for example).

Throughout this paper,  $\hat{\mathcal{C}}(E)$  denotes the set of continuous functions from  $E$  to  $\mathbb{R}$  that vanish at infinity and the functional convergence is in the uniform norm denoted by  $\|\cdot\|$ . In the Markov framework, the weak convergence of a family of processes in the Skorokhod topology can be deduced directly from the behavior of its transition operators. This is formulated in the following theorem in [15]:

**Theorem 4.1.** Let  $E$  be locally compact and separable. For  $h > 0$ , let  $\nu_h(\mathbf{x}, \cdot)$  be a transition function on  $E \times \mathcal{B}(E)$  such that the operator  $\mathcal{T}_h$  defined by

$$\mathcal{T}_h f(\mathbf{x}) = \int_E f(\mathbf{y}) \nu_h(\mathbf{x}, d\mathbf{y})$$

satisfies  $\mathcal{T}_h : \hat{\mathcal{C}}(E) \rightarrow \hat{\mathcal{C}}(E)$ . Suppose that  $\mathcal{T}(t)$  is a Feller semi-group on  $\hat{\mathcal{C}}(E)$ . Assume that  $h \downarrow 0$  and suppose that for every  $f \in \hat{\mathcal{C}}(E)$ ,

$$\lim_{h \downarrow 0} \mathcal{T}_h^{\lfloor t/h \rfloor} f = \mathcal{T}(t)f, \quad t \geq 0.$$

for each  $h > 0$ , let  $(Y_k^h)_{k \in \mathbb{N}}$  be a Markov chain on  $E$  with a transition function  $\nu_h$ , and suppose that  $(Y_0^h)_{h>0}$  has a limiting distribution when  $h \downarrow 0$ . Define

$$X_t^h = Y_{\lfloor t/h \rfloor}^h.$$

Then there is a Markov process  $X$  corresponding to  $\mathcal{T}(t)$  with sample paths in  $\mathcal{D}(\mathbb{R}_+, E)$  and

$$X^h \Rightarrow X.$$

The convergence for the semi-groups is usually proved by proving the convergence of the infinitesimal generators on a subset of functions called "the core" of the limiting generator. For a generator  $\mathcal{A}$  of a contraction semi-group  $\mathcal{T}(t)$  defined on a Banach functional space  $L$ , we suppose that there are two dense subsets  $D$  and  $D_0$  such that  $D_0 \subset D \subset L$ . We say that  $D$  is a core for  $\mathcal{A}$  if for any  $t \geq 0$ ,  $\mathcal{T}(t) : D_0 \rightarrow D$ .

**Proposition 4.2.** *Let  $L$  be a functional Banach space. For  $h > 0$ , let  $\mathcal{T}_h$  be a linear contraction on  $L$  and put*

$$\mathcal{A}_h = \frac{\mathcal{T}_h - I}{h},$$

where  $I$  stands for the identity operator. Let  $\mathcal{T}(t)$  be a strongly continuous contraction semi-group on  $L$  with generator  $\mathcal{A}$ . Let  $D$  be a core of  $\mathcal{A}$ , then the following are equivalent

1. For each  $f \in L$ ,  $\lim_{h \downarrow 0} \mathcal{T}_h^{\lfloor t/h \rfloor} f = \mathcal{T}(t)f, \forall t \geq 0$ .
2. For each  $f \in D$ , there exists a sequence  $f_h \in L$  such that  $\lim_{h \downarrow 0} f_h = f$  and  $\lim_{h \downarrow 0} \mathcal{A}_h f_h = \mathcal{A}f$ .

*Proof.* This is a modification of theorem 6.5 in [15] page 31. □

**Remark 4.3.** *Even though the auxiliary processes defined in 4 are predictable (hence left continuous with right limits or càglàd), the theorems used in this section prove the convergence in  $\mathcal{D}(\mathbb{R}_+, E)$ . That is why we replace our processes with their càdlàg version that has the exact same jumps and dynamics.*

## 4.2 Main result

In this section we prove that the discrete time Hawkes process defined in section 2.2 is an approximation of the continuous-time multivariate non-linear Hawkes process. To do so we provide a convergence theorem for the auxiliary processes that contain the jumps and the dynamics. Note that it is possible to prove the results for the augmented process  $(\mathbf{V}, \mathbf{Z}, \mathbf{N})$ , but this is not necessary as  $(\mathbf{V}, \mathbf{Z})$  is sufficient to reconstruct  $\mathbf{N}$  (cf. Remark 3.6).

For the initial state, there is nothing to prove since we take it always to be equal to zero. To prove the convergence of the process on  $\mathcal{D}(\mathbb{R}_+, E)$  we state the following result

**Proposition 4.4.** *Let  $\mathcal{T}_h$  be the discrete cascade's transition operator defined in 2.4 and let  $\mathcal{A}$  be the generator for the auxiliary processes of a Hawkes process defined in 3.7. Set*

$$\mathcal{A}_h = \frac{\mathcal{T}_h - I}{h}.$$

Then, for any  $f \in \mathcal{S}(E)$  (the Schwartz space of functions) one has

$$\lim_{h \downarrow 0} \mathcal{A}_h f = \mathcal{A}f.$$

*Proof.* Let  $f \in \mathcal{S}(E)$  and  $h > 0$ . Using Lemma 7.2 we have that

$$\begin{aligned} \mathcal{T}_h(\mathbf{u}, \mathbf{y}) &= e^{-h \sum_{i=1}^d \phi_i(\tilde{a}^i(\mathbf{u}, \mathbf{y}))} \mathbb{E}^{\mathcal{N}} \left[ f \left( (M_{ij}^h \mathbf{u}^{ij})_{i,j=1,\dots,d}, \Omega^h \mathbf{y} + \sqrt{h} \mathcal{N}_d \right) \right] \\ &+ e^{-h \sum_{i=1}^d \phi_i(\tilde{a}^i(\mathbf{u}, \mathbf{y}))} \sum_{m=1}^d \mathbb{E}^{\mathcal{N}} \left[ f \left( (M_{ij}^h \mathbf{u}^{ij} + \mathbf{1}_{m=j} \mathbf{e}_{p+1})_{i,j=1,\dots,d}, \Omega^h \mathbf{y} + \sqrt{h} \mathcal{N}_d \right) \right] h \phi_m(\tilde{a}^m(\mathbf{u}, \mathbf{y})) \\ &+ O(h^2). \end{aligned}$$

Note that the subscript  $i, j = 1, \dots, d$  has been omitted to make the notations less tedious. For a  $d$ -variate Gaussian variable  $\mathcal{N}_d$ , we write a Taylor expansion in the second component

$$\begin{aligned} f\left((M_{ij}^h \mathbf{u}^{ij}), \Omega^h \mathbf{y} + \sqrt{h} \mathcal{N}_d\right) &= f\left((M_{ij}^h \mathbf{u}^{ij}), \Omega^h \mathbf{y}\right) + \sqrt{h} \langle \mathcal{N}_d, (\nabla_{\mathbf{y}} f)\left((M_{ij}^h \mathbf{u}^{ij}), \Omega^h \mathbf{y}\right) \rangle \\ &\quad + \frac{h}{2} \langle \mathcal{N}_d, (\text{Hess}_{\mathbf{y}} f)\left((M_{ij}^h \mathbf{u}^{ij}), \Omega^h \mathbf{y}\right) \mathcal{N}_d \rangle \\ &\quad + h^{3/2} \sum_{|\mathbf{r}|=3} R_r\left((M_{ij}^h \mathbf{u}^{ij}), \Omega^h \mathbf{y}\right) (\mathcal{N}_d)^{\mathbf{r}}, \end{aligned}$$

where the remainder  $R_r$  satisfies

$$\begin{aligned} |R_r\left((M_{ij}^h \mathbf{u}^{ij}), \Omega^h \mathbf{y}\right)| &\leq \max_{|\mathbf{s}|=3} \sup_{\mathbf{z} \in \mathbb{R}^d} |D^{\mathbf{s}} f(\mathbf{z})|, \\ &\leq C, \end{aligned}$$

since  $f$  is in the Schwartz space. Hence by taking the expected value with respect to the centered  $d$ -variate Gaussian we have that

$$\mathbb{E}^{\mathcal{N}} \left[ f\left((M_{ij}^h \mathbf{u}^{ij}), \Omega^h \mathbf{y} + \sqrt{h} \mathcal{N}_d\right) \right] = f\left((M_{ij}^h \mathbf{u}^{ij}), \Omega^h \mathbf{y}\right) + \frac{h}{2} \sum_{i=1}^d \frac{\partial^2 f}{\partial y_i^2}\left((M_{ij}^h \mathbf{u}^{ij}), \Omega^h \mathbf{y}\right) + O(h^{3/2}).$$

Now, a second order Taylor expansion with a Lagrange remainder in the first component yields

$$\begin{aligned} f\left((M_{ij}^h \mathbf{u}_{ij}), \Omega^h \mathbf{y}\right) &= f\left(\mathbf{u}_{ij}, \Omega^h \mathbf{y}\right) + \sum_{i,j=1}^d \langle (M_{ij}^h - I) \mathbf{u}_{ij}, \nabla_{\mathbf{u}_{ij}} f(\mathbf{u}_{ij}, \Omega^h \mathbf{y}) \rangle \\ &\quad + \sum_{i,j=1}^d \sum_{|\mathbf{s}|=2} \frac{1}{\mathbf{s}!} D^{\mathbf{s}} f\left(\left((1-c)I + cM_{ij}^h\right) \mathbf{u}_{ij}, \Omega^h \mathbf{y}\right) \left((M_{ij}^h - I) \mathbf{u}_{ij}\right)^{\mathbf{s}} \end{aligned}$$

for some  $c \in (0, 1)$ . Since for any  $i, j = 1, \dots, d$   $M_{ij}^h - I = h(K - \beta_{ij}I) + O(h^2)$ ,

$$\begin{aligned} f\left((M_{ij}^h \mathbf{u}_{ij}), \Omega^h \mathbf{y}\right) &= f\left(\mathbf{u}_{ij}, \Omega^h \mathbf{y}\right) + h \sum_{i,j=1}^d \langle (K - \beta_{ij}I) \mathbf{u}_{ij}, \nabla_{\mathbf{u}_{ij}} f(\mathbf{u}_{ij}, \Omega^h \mathbf{y}) \rangle \\ &\quad + \sum_{i,j=1}^d \langle O(h^2) \mathbf{u}_{ij}, \nabla_{\mathbf{u}_{ij}} f(\mathbf{u}_{ij}, \Omega^h \mathbf{y}) \rangle \\ &\quad + h^2 \sum_{i,j=1}^d \sum_{|\mathbf{s}|=2} \frac{1}{\mathbf{s}!} D^{\mathbf{s}} f\left(\left((1-c)I + cM_{ij}^h\right) \mathbf{u}_{ij}, \Omega^h \mathbf{y}\right) \left((K - \beta_{ij}I + O(h)) \mathbf{u}_{ij}\right)^{\mathbf{s}}. \end{aligned}$$

Since  $f$  is taken to be in the Schwartz space, it follows that its derivatives multiplied by any polynomial are uniformly bounded, hence

$$f\left((M_{ij}^h \mathbf{u}_{ij}), \Omega^h \mathbf{y}\right) = f\left(\mathbf{u}_{ij}, \Omega^h \mathbf{y}\right) + h \sum_{i,j=1}^d \langle (K - \beta_{ij}I) \mathbf{u}_{ij}, \nabla_{\mathbf{u}_{ij}} f(\mathbf{u}_{ij}, \Omega^h \mathbf{y}) \rangle + O(h^2)$$

A similar argument for the second component of the function and its gradient gives us that

$$\begin{aligned} f\left((M_{ij}^h \mathbf{u}_{ij}), \Omega^h \mathbf{y}\right) &= f\left(\mathbf{u}_{ij}, \mathbf{y}\right) + h \sum_{i,j=1}^d \langle (K - \beta_{ij}I) \mathbf{u}_{ij}, \nabla_{\mathbf{u}_{ij}} f(\mathbf{u}_{ij}, \mathbf{y}) \rangle \\ &\quad - h \langle \text{diag}(\omega_1, \dots, \omega_d) \mathbf{y}, \nabla_{\mathbf{y}} f(\mathbf{u}_{ij}, \mathbf{y}) \rangle + O(h^2). \end{aligned}$$

Applying the same arguments to the second derivative yields

$$\begin{aligned} \mathbb{E}^{\mathcal{N}} \left[ f\left((M_{ij}^h \mathbf{u}^{ij}), \Omega^h \mathbf{y} + \sqrt{h} \mathcal{N}_d\right) \right] &= f\left(\mathbf{u}_{ij}, \mathbf{y}\right) + h \sum_{i,j=1}^d \langle (K - \beta_{ij}I) \mathbf{u}_{ij}, \nabla_{\mathbf{u}_{ij}} f(\mathbf{u}_{ij}, \mathbf{y}) \rangle \\ &\quad - h \langle \text{diag}(\omega_1, \dots, \omega_d) \mathbf{y}, \nabla_{\mathbf{y}} f(\mathbf{u}_{ij}, \mathbf{y}) \rangle + \sum_{i=1}^d \frac{\partial^2 f}{\partial y_i^2}\left(\mathbf{u}^{ij}, \mathbf{y}\right) + O(h^{3/2}). \end{aligned}$$

Finally, using the fact that the jump rates  $\phi_1, \dots, \phi_d$  are bounded and the Taylor expansion of the exponential

$$e^{-h \sum_{i=1}^d \phi_i(\tilde{a}^i(\mathbf{u}, \mathbf{y}))} = 1 - h \sum_{i=1}^d \phi_i(\tilde{a}^i(\mathbf{u}, \mathbf{y})) + O(h^2)$$

we conclude that

$$\mathcal{T}_h(t)f((\mathbf{u}_{ij}), \mathbf{y}) = f((\mathbf{u}_{ij}), \mathbf{y}) + h\mathcal{A}f((\mathbf{u}_{ij}), \mathbf{y}) + O(h^{3/2}),$$

which yields the result.  $\square$

We now state the main result, establishing the convergence of the discrete-time Hawkes process to the continuous-time Hawkes process.

**Theorem 4.5.** *Let  $(\mathbf{U}_n, \mathbf{Y}_n, \mathbf{H}_n)_{n \in \mathbb{N}}$  be the Markov chain defined in 2.4.*

*For a bin size  $h > 0$ , set  $(\tilde{\mathbf{V}}_t^h, \tilde{\mathbf{Z}}_t^h, \tilde{\mathbf{N}}_t^h)_{t \in \mathbb{R}_+}$  the càdlàg process that coincides with  $(\mathbf{U}_n, \mathbf{Y}_n, \mathbf{H}_n)$  on the  $n$ -th time bin, i.e.*

$$(\tilde{\mathbf{V}}_t^h, \tilde{\mathbf{Z}}_t^h, \tilde{\mathbf{N}}_t^h) = (\mathbf{U}_{\lfloor t/h \rfloor}, \mathbf{Y}_{\lfloor t/h \rfloor}, \mathbf{H}_{\lfloor t/h \rfloor}),$$

(cf. Figure 5 for an illustration).

Then

$$(\tilde{\mathbf{V}}_t^h, \tilde{\mathbf{Z}}_t^h, \tilde{\mathbf{N}}_t^h)_{t \in \mathbb{R}_+} \Longrightarrow (\mathbf{V}_t, \mathbf{Z}_t, \mathbf{N}_t)_{t \in \mathbb{R}_+},$$

in distribution as the bin size  $h$  goes to zero.

*Proof.* It is sufficient to prove the convergence of  $(\tilde{\mathbf{V}}_t^h, \tilde{\mathbf{Z}}_t^h)_{t \in \mathbb{R}_+}$  to  $(\mathbf{V}_t, \mathbf{Z}_t)_{t \in \mathbb{R}_+}$  as  $h \rightarrow 0$  (cf. Remark 2.5 and 3.6). Since the Schwartz space  $\mathcal{S}(E)$  is a core for the infinitesimal generator  $\mathcal{A}$  (cf. Lemma 7.5), Proposition 4.2 ensures that

$$\lim_{h \downarrow 0} \mathcal{T}_h^{\lfloor t/h \rfloor} f = \mathcal{T}(t)f,$$

for each  $f$  in the Banach space (cf. Lemma 7.1)  $\hat{\mathcal{C}}(E)$ .

Theorem 4.1 yields that the process

$$(\tilde{\mathbf{V}}_t^h, \tilde{\mathbf{Z}}_t^h) = (\mathbf{U}_{\lfloor t/h \rfloor}, \mathbf{Y}_{\lfloor t/h \rfloor}),$$

converges to the auxiliary process  $(\mathbf{V}, \mathbf{Z})$  of the non linear multivariate Hawkes process, weakly in the Skorokhod metric.  $\square$

To illustrate the convergence, we simulate  $\tilde{a}^m(\tilde{\mathbf{V}}_t^h, \tilde{\mathbf{Z}}_t^h)$  in the bi-variate case for  $t \in [0, 15]$  for two different bin sizes. We recall that  $\tilde{a}^m(\mathbf{v}, \mathbf{z}) = \mu_m + \sigma_m z_m + \sum_{j=1}^d \langle \boldsymbol{\alpha}^{mj}, \mathbf{v}_{mj} \rangle$ . The jump rates are

$$\phi_1(x) = (x)_+ \wedge 40, \quad \phi_2(x) = \log(1 + e^x) \wedge 40$$

and the memory kernels are

$$\begin{pmatrix} \delta_{11}(s) & \delta_{12}(s) \\ \delta_{21}(s) & \delta_{22}(s) \end{pmatrix} = \begin{pmatrix} 5(s-1)e^{-5s} & s(\frac{s}{2} - 6)e^{-6s} \\ 0 & s(\frac{3s}{2} + 4)e^{-4s} \end{pmatrix}.$$

The convergence is illustrated in the transition from figure 5 ( $h = 0.125$ ) to figure 6 ( $h = 5.10^{-3}$ ).

## 5 Parameters calibration using regression

The estimation of linear Hawkes processes have been thoroughly studied in the linear case, both in a parametric and a non-parametric way.

For the exponential kernels, [26] proved that the computation likelihood function is linear in the number of events, numerical results are given in the 1-d case as well. [9] derived an estimation of the parameters using the stationary moments that can be explicitly derived for the linear Hawkes process with exponential kernels. This method is not applicable in the non-linear framework since the expressions of the first moments are not known explicitly.

Since the kernel's shape is usually not known in advance and since the exponential family is not rich enough (e.g. power laws, non monotonous functions), methods for non-parametric estimation have been derived for

the linear Hawkes process. The first of which is based on the numerical resolution of a system of Wiener-Hopf equations, introduced by [2]. The second is introduced by [20] and by [13] based on time series. These methods are close to the ones we present in this paper. The reader can refer to [30] for a comparison between the time series method and an E-M method.

The non-linear framework has also been studied, though to a lesser extent. For the parametric estimation, [23] derived an algorithm for the log-likelihood maximisation for kernels composed by an exponential mixture and a ReLU jumps rate. For a more general choice of the jump rate functions as well as kernels, [31] obtained concentration rates of the posterior distribution on the parameters. When it comes to the non parametric approach, Reynaud-Bouret *et al.* devised a method for the non-parametric estimation for multivariate Hawkes processes with ReLU jump-rates.

In this section, we assume that both assumptions 2.1 (boundedness of the jump-rate) and 3.3 (jump-rate functions are  $l_i$ -Lipschitz and the matrix  $\left( l_i \sum_{q=0}^p \frac{|\alpha_q^{ij}|}{\beta_j^{q+1}} \right)_{i,j=1,\dots,d}$  has a spectral radius less than one) are in force. In fact, if assumption 3.3 is not respected, the pre-intensity  $\tilde{a}^i$  can grow to infinity, before being capped by the upper bound  $\|\phi_i\|_\infty$ , transforming the Hawkes process into a Poisson process of intensity  $\|\phi_i\|_\infty$  from which no other parameters can be recovered. This is the problem of identifiability which is discussed in more details in [31].

We then implement a conditional least-squares (CLS) regression on the observed discrete time version  $(\mathbf{P}_k)_{k=1,\dots,M} = (\mathbf{H}_k - \mathbf{H}_{k-1})_{k=1,\dots,M}$  of an underlying Hawkes process  $(\mathbf{N}_t)_{t \in [0,T]}$ , with time bins  $h = T/M$ . Let  $\mathcal{F}_k = \sigma(\mathbf{P}_1, \dots, \mathbf{P}_k)$  be the filtration that contains the information up to  $k$ . We seek to find

$$\min_{(\alpha, \beta, \boldsymbol{\mu}) \in (\mathbb{R}^{(p+1) \times d \times d}) \times \mathcal{M}_d(\mathbb{R}_+) \times \mathbb{R}_+^d} \sum_{k=1}^M \|\mathbf{P}_k - \mathbb{E}[\mathbf{P}_k | \mathcal{F}_{k-1}]\|^2.$$

For the sake of brevity, we set

$$\begin{aligned} Q_M &= \sum_{k=1}^M \|\mathbf{P}_k - \mathbb{E}[\mathbf{P}_k | \mathcal{F}_{k-1}]\|^2, \\ &= \sum_{k=1}^M \sum_{i=1}^d (P_k^i - \mathbb{E}[P_k^i | \mathcal{F}_{k-1}])^2. \end{aligned}$$

The underlying process is supposed to follow multivariate Hawkes dynamics

$$\lambda_t^i = \phi_i \left( \mu_i + \sum_{j=1}^d \int_{[0,t)} \delta_{ij}(t-s) dN_s^j + \sigma_i \int_{[0,t)} e^{-\omega_i(t-s)} dW_s^i \right),$$

with known jump-rates  $\phi$ , unknown baseline intensities  $\boldsymbol{\mu}$  and unknown kernels  $\delta_{ij}(s) = e^{-\beta_{ij}s} \sum_{q=0}^p \alpha_q^{ij} \frac{s^q}{q!}$ , where  $p$  is unknown. We do not estimate the Gaussian perturbation's parameters.

Since we have proved that the discrete time Hawkes process converges to the continuous-time Hawkes process, we set

$$P_k^i = N_{hk}^i - N_{h(k-1)}^i$$

for a fixed  $h > 0$ .

General results about CLS estimators for time-series, including asymptotic bias and normality can be found in the seminal work of [22].

Using the definition of the discrete time Hawkes process,  $Q_M$  can be computed linearly in the number of points as

$$\begin{aligned} Q_{M+1} &= Q_M + \sum_{i=1}^d (P_{M+1}^i - \mathbb{E}[P_{M+1}^i | \mathcal{F}_M])^2, \\ &= Q_M + \sum_{i=1}^d \left( P_{M+1}^i - h\phi_i \left( \mu_i + \sum_{j=1}^d \langle \boldsymbol{\alpha}^{ij}, \mathbf{U}_{M+1}^{ij} \rangle \right) \right)^2, \end{aligned}$$

where the  $\mathbf{U}^{ij}$  are also computed recursively

$$\mathbf{U}_{M+1}^{ij} = M^{ij}\mathbf{U}_M^{ij} + P_M^j \mathbf{e}_{p+1}.$$

If the jump-rates are assumed to be differentiable almost everywhere, the same holds for the gradients which are computed recursively

$$\begin{aligned} \partial_{\mu_i} Q_{M+1} &= \partial_{\mu_i} Q_M - 2h(\phi_i)' \left( \mu_i + \sum_{j=1}^d \langle \boldsymbol{\alpha}^{ij}, \mathbf{U}_{M+1}^{ij} \rangle \right) \left( P_{M+1}^i - h\phi_i \left( \mu_i + \sum_{j=1}^d \langle \boldsymbol{\alpha}^{ij}, \mathbf{U}_{M+1}^{ij} \rangle \right) \right), \\ \nabla_{\boldsymbol{\alpha}^{ij}} Q_{M+1} &= \nabla_{\boldsymbol{\alpha}^{ij}} Q_M - 2h\mathbf{U}_M^{ij}(\phi_i)' \left( \mu_i + \sum_{j=1}^d \langle \boldsymbol{\alpha}^{ij}, \mathbf{U}_{M+1}^{ij} \rangle \right) \left( P_{M+1}^i - h\phi_i \left( \mu_i + \sum_{j=1}^d \langle \boldsymbol{\alpha}^{ij}, \mathbf{U}_{M+1}^{ij} \rangle \right) \right), \\ \partial_{\beta_{ij}} Q_{M+1} &= \partial_{\beta_{ij}} Q_M - 2h \langle \boldsymbol{\alpha}^{ij}, \partial_{\beta_{ij}} \mathbf{U}_{M+1}^{ij} \rangle (\phi_i)' \left( \mu_i + \sum_{j=1}^d \langle \boldsymbol{\alpha}^{ij}, \mathbf{U}_{M+1}^{ij} \rangle \right) \left( P_{M+1}^i - h\phi_i \left( \mu_i + \sum_{j=1}^d \langle \boldsymbol{\alpha}^{ij}, \mathbf{U}_{M+1}^{ij} \rangle \right) \right), \\ \partial_{\beta_{ij}} \mathbf{U}_{M+1}^{ij} &= M^{ij} \left( \partial_{\beta_{ij}} \mathbf{U}_M^{ij} - h\mathbf{U}_M^{ij} \right). \end{aligned}$$

Note that unlike the linear Hawkes process with Erlang kernels, this cannot be simplified to a computation in the number of events.

As a first trial we simulate an unperturbed 3–dimensional DTHP with  $h = 0.1$  for  $M = 5000$ . The kernels are

$$\delta(s) = \begin{pmatrix} (1 + 0.2s)e^{-3s} & (0.2 + 3s)e^{-4s} & 3se^{-3s} \\ 0 & (-0.1 - 2s)e^{-2s} & (1 - 5s)e^{-1.5s} \\ (-0.5 - 2s)e^{-3s} & 0 & (0.7 + 6.2s)e^{-3.5s} \end{pmatrix},$$

with jump-rates

$$\phi(x) = ((x)_+ \wedge 40 \quad \ln(1 + e^x) \wedge 40 \quad \frac{4}{1 + e^{-(x-3)}}),$$

and baseline intensities

$$\boldsymbol{\mu} = (0.4 \quad 1 \quad 2).$$

The minimization of  $Q_M$  is done using the SLSQP method of the function `scipy.optimize.minimize` with assumption 3.3 as a constraint.

The initial guesses are  $\hat{\alpha}_q^{ij} \sim \mathcal{U}([0, 1])$ ,  $\hat{\beta}_{ij} = 1$  and  $\hat{\mu}^i = (\phi^i)^{-1} \left( \frac{\sum_{k=1}^M P_k^i}{T} \right)$ .

The minimization results are then averaged (for five calibrations) and are given in figure 7.

## 5.1 Calibration experiment for the perturbed Hawkes process

The underlying dynamics are assumed to follow a continuous-time 3–variate Hawkes process with a Gaussian perturbation of parameters

$$\boldsymbol{\omega} = (3 \quad 4 \quad 5)$$

and

$$\boldsymbol{\sigma} = (1 \quad 2 \quad 2).$$

We simulate 35 Monte Carlo samples for  $t \in [0, 100]$ , which we regroup in 35 DTHP arrays with a time step  $h = 0.25$ . The discretisation introduces bias, by discarding the possibility that a given event in a given bin be the origin of a later event in the same bin.

The result of the CLS minimisation is given in figures 8, 9 and 10.

## 5.2 Calibration with a lower degree

We now simulate 15 continuous-time Hawkes processes with kernels

$$\delta(s) = \begin{pmatrix} (0.7 - 2s + 1.5s^2)e^{-1.5s} & (0.5 + 4s - 10s^2)e^{-3s} & 2s^2e^{-2s} \\ 0 & (5s + 6s^2)e^{-3s} & (1 - 10s^2)e^{-2.5s} \\ -5s^2e^{-2s} & 0 & (0.8 + 7s + 8s^2)e^{-5s} \end{pmatrix},$$

with jump-rates

$$\phi(x) = ((x)_+ \wedge 40 \quad \ln(1 + e^x) \wedge 40 \quad \frac{4}{1+e^{-(x-3)}}),$$

and baseline intensities

$$\boldsymbol{\mu} = (0.4 \quad 1 \quad 2).$$

The Gaussian perturbation's parameters are

$$\boldsymbol{\omega} = (3 \quad 4 \quad 5)$$

and

$$\boldsymbol{\sigma} = (1 \quad 2 \quad 2).$$

The processes are then discretized on a grid with a time-step  $h = 0.25$ . We minimise the conditional least squares  $Q_M$  with assumed kernels of the form

$$\delta_{ij}(s) = \alpha_0^{ij} e^{-\beta_{ij}s}.$$

The exponential kernels composed with the average parameters are illustrated in figure 11. When it comes to the baseline intensities refer to figure 12

## 6 Conclusion

In this paper, we introduced a tractable and intuitive Markov Chain approximation of the multivariate non-linear Hawkes process with a Gaussian perturbation, and we proved that this approximation converges when the time-bin size goes to zero. This result can be extended to a wide array of similar processes with Markov dynamics, for instance for a perturbation parameter  $\sigma$  that depends on the system's state, a baseline intensity formed by a mixture of Erlang and trigonometric functions to account for seasonality or for jumps with random sizes.

A parametric calibration method has been studied numerically but without theoretical guarantees on the convergence of the estimated parameters to the ground truth parameters. The experiments show that the calibrated parameters are close to the actual kernels despite the Gaussian perturbations and the bias introduced by the discretization. This method can be enhanced for instance by adding a penalisation or by decreasing the variance.

## 7 Lemmata

**Lemma 7.1.** *Set  $E = (\mathbb{R}^{p+1})^{d \times d} \times \mathbb{R}^d$ . Then the space  $\hat{\mathcal{C}}(E)$  of continuous functions that vanish at infinity equipped with the uniform norm is a Banach space.*

*Moreover, the space of  $C^\infty$  functions of compact support  $\mathcal{C}_c^\infty(E)$  is dense in  $\hat{\mathcal{C}}(E)$ .*

*Proof.* Let  $(f_n)_{n \in \mathbb{N}}$  be a Cauchy sequence in  $\hat{\mathcal{C}}(E)$ . For every  $\mathbf{x} \in E$ , the sequence  $(f_n(\mathbf{x}))_{n \in \mathbb{N}}$  satisfies

$$|f_n(\mathbf{x}) - f_m(\mathbf{x})| \leq \|f_n - f_m\|, \quad \forall n, m \in \mathbb{N},$$

and thus, is a Cauchy sequence in  $E$ . Since  $E$  is complete, the sequence converges to the point-wise limit  $f(\mathbf{x})$ . Let  $p \in \mathbb{N}$  and  $\mathbf{x} \in E$

$$\begin{aligned} |f_p(\mathbf{x}) - f(\mathbf{x})| &= |f_p(\mathbf{x}) - \lim_{n \rightarrow +\infty} f_n(\mathbf{x})|, \\ &= \lim_{n \rightarrow +\infty} |f_p(\mathbf{x}) - f_n(\mathbf{x})|, \\ &\leq \lim_{n \rightarrow +\infty} \|f_p - f_n\|. \end{aligned}$$

Let  $\epsilon > 0$  and  $M$  such that for  $m, n \geq M$ ,  $\|f_m - f_n\| \leq \epsilon$ . If  $p \geq M$

$$|f_p(\mathbf{x}) - f(\mathbf{x})| \leq \epsilon.$$

Thus  $(f_p)_{p \in \mathbb{N}}$  converges to  $f$  uniformly, hence  $f$  is continuous.

For some  $p$  such that  $\|f_p - f\| \leq \epsilon/2$ , Let  $K_p$  be the compact such that  $\forall x \notin K_p$ :

$$|f_p(\mathbf{x})| \leq \epsilon/2.$$

We have for any  $\mathbf{x} \notin K_p$

$$\begin{aligned} |f(\mathbf{x})| &\leq |f(\mathbf{x}) - f_p(\mathbf{x})| + |f_p(\mathbf{x})|, \\ &\leq \|f - f_p\| + \epsilon/2, \\ &\leq \epsilon. \end{aligned}$$

Which proves that  $\hat{\mathcal{C}}(E)$  is a Banach space.

For the density of the  $C_c^\infty(E)$ , it is an application of Stone-Weierstrass' theorem, thanks to the existence of bump functions.  $\square$

**Lemma 7.2.** *Assume that assumption 2.1 is in force.*

*Let  $R_2$  be the remainder defined in Proposition 2.4. Then, uniformly in  $\mathbf{u}$  and  $\mathbf{y}$ ,*

$$R_2(\mathbf{u}, \mathbf{y}) = O(h^2).$$

*Proof.* We recall that for a function  $f \in D$ ,

$$R_2(\mathbf{u}, \mathbf{y}) = \mathbb{E}^{\mathcal{N}} \left[ \sum_{i_1 + \dots + i_d \geq 2} f \left( (M_{ij}^h \mathbf{u}^{ij} + i_m \mathbf{e}_{p+1})_{i,j=1,\dots,d}, \Omega^h \mathbf{y} + \sqrt{h} \mathcal{N}_d \right) \prod_{m=1}^d \frac{(h\phi_m(\tilde{a}^m(\mathbf{u}, \mathbf{y})))^{i_m}}{i_m!} e^{-h\phi_m(\tilde{a}^m(\mathbf{u}, \mathbf{y}))} \right].$$

Since  $f$  is bounded we have that

$$\begin{aligned} |R_2(\mathbf{u}, \mathbf{y})| &\leq \mathbb{E}^{\mathcal{N}} \left[ \sum_{i_1 + \dots + i_d \geq 2} \|f\|_\infty \prod_{m=1}^d \frac{(h\phi_m(\tilde{a}^m(\mathbf{u}, \mathbf{y})))^{i_m}}{i_m!} e^{-h\phi_m(\tilde{a}^m(\mathbf{u}, \mathbf{y}))} \right], \\ &= \|f\|_\infty \sum_{i_1 + \dots + i_d \geq 2} \prod_{m=1}^d \frac{(h\phi_m(\tilde{a}^m(\mathbf{u}, \mathbf{y})))^{i_m}}{i_m!} e^{-h\phi_m(\tilde{a}^m(\mathbf{u}, \mathbf{y}))}. \end{aligned}$$

Since  $i_1 + \dots + i_m \leq 2$  and  $h < 1$ ,  $h^{i_m} \leq h^2$  for any  $m = 1, \dots, d$ . Hence

$$\begin{aligned} |R_2(\mathbf{u}, \mathbf{y})| &\leq \|f\|_\infty h^2 \sum_{i_1 + \dots + i_d \geq 2} \prod_{m=1}^d \frac{\phi_m(\tilde{a}^m(\mathbf{u}, \mathbf{y}))^{i_m}}{i_m!} e^{-h\phi_m(\tilde{a}^m(\mathbf{u}, \mathbf{y}))}, \\ &\leq \|f\|_\infty h^2 \sum_{i_1 + \dots + i_d \geq 2} \prod_{m=1}^d \frac{\|\phi_m\|^{i_m}}{i_m!}, \\ &\leq \|f\|_\infty h^2 e^{\sum_{m=1}^d \|\phi_m\|}. \end{aligned}$$

$\square$

**Lemma 7.3.** *For  $i, j = 1, \dots, d$ , let  $M_{ij}^h = e^{-\beta_{ij}h}(I + hK)$  and  $\Omega^h = \text{diag}(e^{-\omega_1 h}, \dots, e^{-\omega_d h})$  be the matrices defined in section 2.2. Then*

$$M_{ij}^h = I + h(K - \beta_{ij}I) + O(h^2),$$

and

$$\Omega^h = I - \text{diag}(\omega_1, \dots, \omega_d)h + O(h^2).$$

*Proof.* This is a simple second order Taylor expansion.  $\square$

**Lemma 7.4.** *Set  $E = (\mathbb{R}^{p+1})^{d \times d} \times \mathbb{R}^d$ .*

*Let  $\mathcal{T}$  be the semi-group associated with the Markov process defined by the auxiliary processes 4 and 5, i.e. defined for every  $t \geq 0$  and  $f \in \hat{\mathcal{C}}(E)$  by*

$$\mathcal{T}(t)f(\mathbf{v}, \mathbf{z}) = \mathbb{E}[f(\mathbf{V}_t, \mathbf{Z}_t) | (\mathbf{V}_0, \mathbf{Z}_0) = (\mathbf{v}, \mathbf{z})].$$

*Then  $\mathcal{T}$  is a Feller semi-group.*

*Proof.* It is trivial to see that for any  $t \geq 0$ ,  $\mathcal{T}(t)$  is a contraction (*i.e.*  $\|\mathcal{T}(t)f\| \leq \|f\|$  for every  $f \in \hat{\mathcal{C}}(E)$ ) and positive (*i.e.* for every non-negative function  $f$ ,  $\mathcal{T}(t)f$  is a non-negative function). Now we show that it is strongly continuous. To do so, one must show that it has the Feller property

$$\forall f \in \hat{\mathcal{C}}(E), t \geq 0, \mathcal{T}(t)f \in \hat{\mathcal{C}}(E).$$

First, by solving the SDE in 3.5 with the initial conditions  $\mathbf{V}_0^{ij} = \mathbf{v}^{ij}$  and  $\mathbf{Z}_0 = \mathbf{z}$  we have

$$\begin{aligned} \mathbf{V}_t^{ij} &= e^{(K-\beta_{ij}I)t} \mathbf{v}^{ij} + \int_{[0,t)} e^{-(K-\beta_{ij}I)(t-s)} \mathbf{e}_{p+1} dN_s^j, \\ &= e^{(K-\beta_{ij}I)t} \mathbf{v}^{ij} + \boldsymbol{\xi}_t^{1,ij} \end{aligned}$$

and

$$\begin{aligned} \mathbf{Z}_t &= \text{diag}(e^{-\omega_1 t}, \dots, e^{-\omega_d t}) \mathbf{z} + \int_0^t \text{diag}(e^{-\omega_1(t-s)}, \dots, e^{-\omega_d(t-s)}) d\mathbf{W}_s, \\ &= \text{diag}(e^{-\omega_1 t}, \dots, e^{-\omega_d t}) \mathbf{z} + \boldsymbol{\xi}_t^{2,ij}, \end{aligned}$$

where  $\boldsymbol{\xi}_t^{1,ij}$  and  $\boldsymbol{\xi}_t^{2,ij}$  are two finite random variables. Since  $f$  is bounded it is possible to exchange limits and integrals and one has for any fixed  $t \geq 0$ :

$$\begin{aligned} \lim_{(\mathbf{v}, \mathbf{z}) \rightarrow \infty} \mathcal{T}(t)f(\mathbf{v}, \mathbf{z}) &= \lim_{(\mathbf{v}, \mathbf{z}) \rightarrow \infty} \mathbb{E} \left[ f \left( (e^{(K-\beta_{ij}I)t} \mathbf{v}^{ij} + \boldsymbol{\xi}_t^{1,ij})_{i,j=1, \dots, d}, \text{diag}(e^{-\omega_1 t}, \dots, e^{-\omega_d t}) \mathbf{z} + \boldsymbol{\xi}_t^{2,ij} \right) \right], \\ &= \mathbb{E} \left[ \lim_{(\mathbf{v}, \mathbf{z}) \rightarrow \infty} f \left( (e^{(K-\beta_{ij}I)t} \mathbf{v}^{ij} + \boldsymbol{\xi}_t^{1,ij})_{i,j=1, \dots, d}, \text{diag}(e^{-\omega_1 t}, \dots, e^{-\omega_d t}) \mathbf{z} + \boldsymbol{\xi}_t^{2,ij} \right) \right], \\ &= \mathbb{E} [0], \\ &= 0. \end{aligned}$$

This, combined with the weak continuity (trivial) yields that  $\mathcal{T}$  is a Feller semi-group (*cf.* the first chapter in [7]).  $\square$

**Lemma 7.5.** Let  $E = (\mathbb{R}_+^{(p+1)})^{d \times d} \times \mathbb{R}^d$ .

Let  $\mathcal{S}(E)$  be the space of Schwartz functions, *i.e.* the  $\mathcal{C}^\infty$  functions such that

$$\sup_{\mathbf{x} \in E} |\mathbf{x}^{\mathbf{m}} D^{\mathbf{n}} f(\mathbf{x})| < +\infty,$$

for any multi-indices  $\mathbf{m}$  and  $\mathbf{n}$ . Then  $\mathcal{S}(E)$  is a core for  $\mathcal{A}$ .

*Proof.* The density of  $\mathcal{S}(E)$  in  $\hat{\mathcal{C}}(E)$  is a consequence of the density of  $\mathcal{C}_c^\infty(E)$ .

Let  $f \in \mathcal{S}(E)$  and  $t \geq 0$ . Then

$$\begin{aligned} \mathcal{T}(t)f(\mathbf{v}, \mathbf{z}) &= \mathbb{E} [f(\mathbf{V}_t, \mathbf{Z}_t) | (\mathbf{V}_0, \mathbf{Z}_0) = (\mathbf{v}, \mathbf{z})], \\ &= \mathbb{E} \left[ f \left( (e^{(K-\beta_{ij}I)t} \mathbf{v}^{ij} + \boldsymbol{\xi}_t^{1,ij})_{i,j=1, \dots, d}, \text{diag}(e^{-\omega_1 t}, \dots, e^{-\omega_d t}) \mathbf{z} + \boldsymbol{\xi}_t^{2,ij} \right) \right], \end{aligned}$$

where  $\boldsymbol{\xi}_t^{1,ij} = \int_{[0,t)} e^{-(K-\beta_{ij}I)(t-s)} \mathbf{e}_{p+1} dN_s^j$ , and  $\boldsymbol{\xi}_t^{2,ij} = \int_0^t \text{diag}(e^{-\omega_1(t-s)}, \dots, e^{-\omega_d(t-s)}) d\mathbf{W}_s$  are two finite random variables. Since  $\mathcal{S}(E)$  is stable by composition with an affine transform (*cf.* [16]) and since the differentiation and the expected value can be exchanged ( $f$  is bounded), then  $\mathcal{T}(t)f \in \mathcal{S}(E)$ .  $\square$

## Data availability statement

Data sharing not applicable to this article as no datasets were analysed during the current study.

## Appendix: Figures

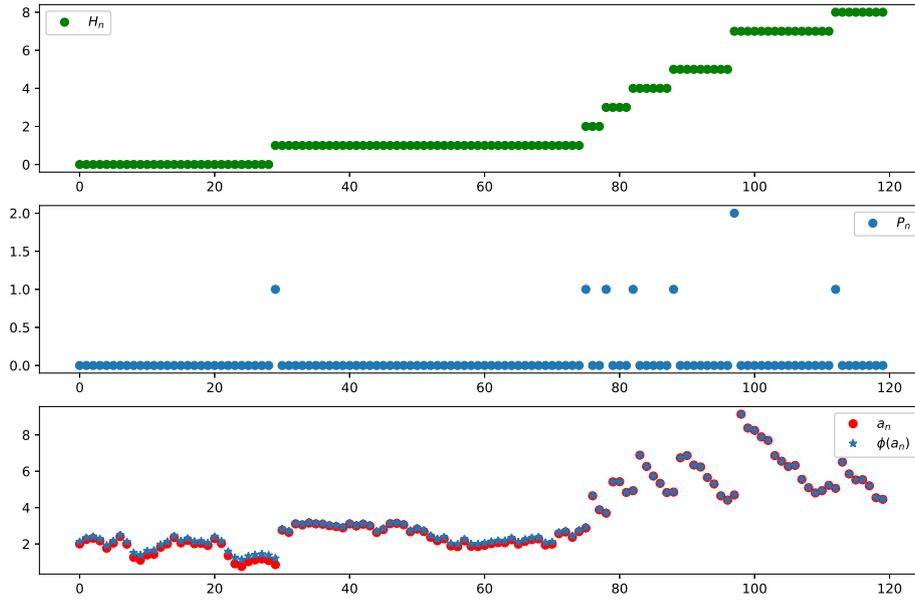


Figure 1: The process is *self-exciting* for  $\alpha = 2$ . This is seen as events tend to be clustered in time. Notice how one bin contains two events.

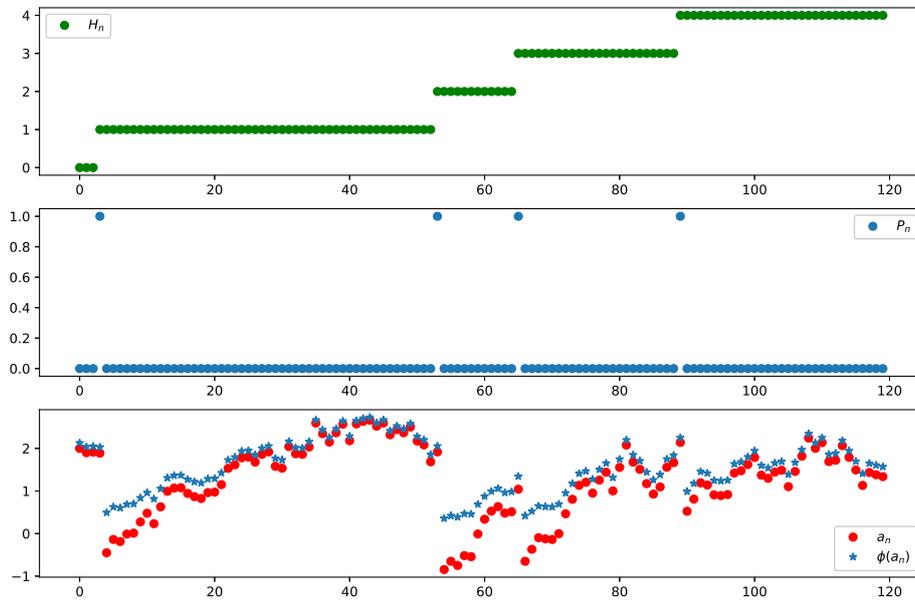


Figure 2: The process is *self-inhibiting* for  $\alpha = -2$ . Events tend to be distanced from one another.

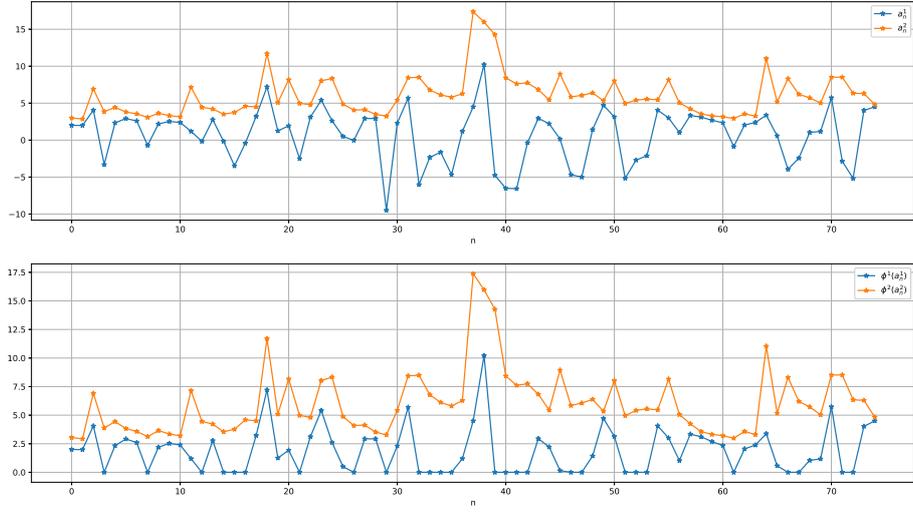


Figure 3: A plot of a realisation of  $\mathbf{a}_k = (\tilde{a}^m(\mathbf{U}_k, \mathbf{Y}_k))_{m=1,2}$  (upper plot) as well as  $\phi(\mathbf{a}_k)$  (lower plot) for  $k \leq 75$ . The interaction parameters are  $(\alpha_1^{11}, \alpha_0^{11}) = (0, 2)$ ,  $(\alpha_1^{12}, \alpha_0^{12}) = (15, -3)$ ,  $(\alpha_1^{21}, \alpha_0^{21}) = (-6, 4)$  and  $(\alpha_1^{22}, \alpha_0^{22}) = (10, 0)$ . The perturbation parameters are  $(\sigma_1, \sigma_2) = (0.1, 0.2)$ .

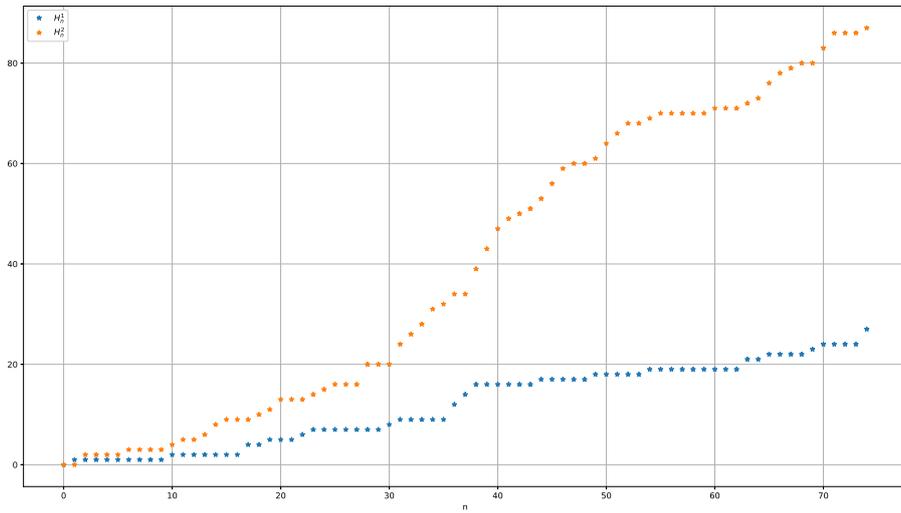


Figure 4: A realisation of  $\mathbf{H}_k$  for  $k \leq 75$ .

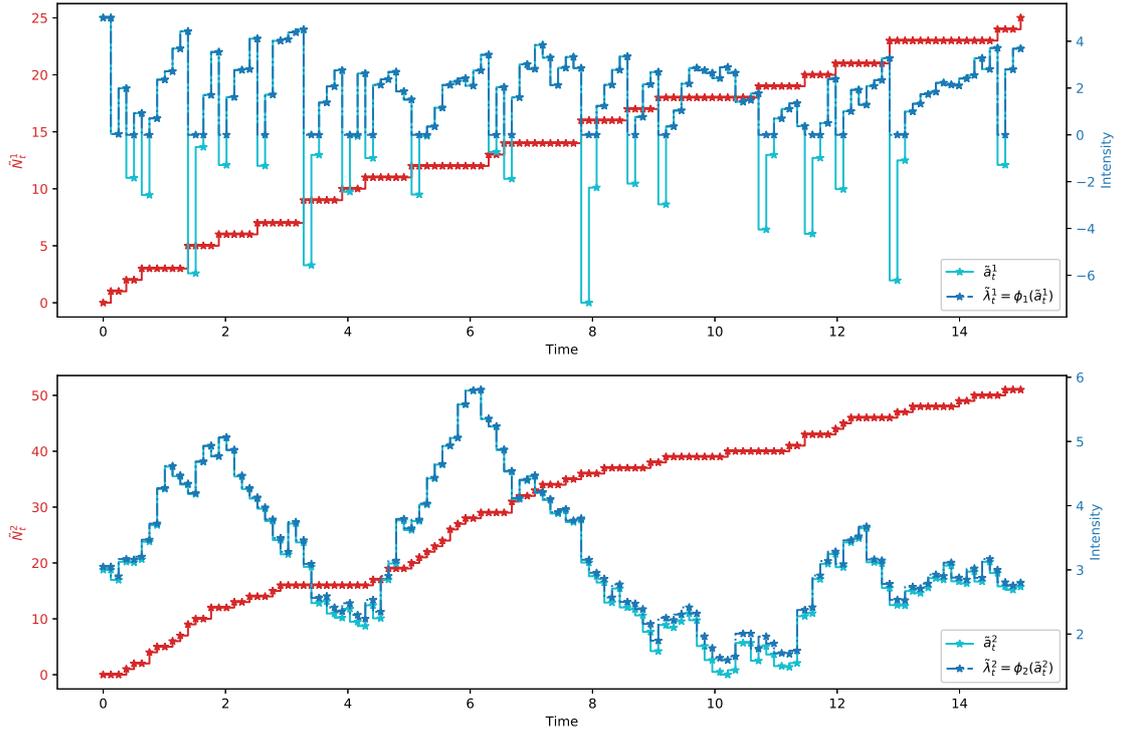


Figure 5: Simulation for  $h = 0.125$ . The baseline intensities are  $(\mu_1, \mu_2) = (5, 3)$ .

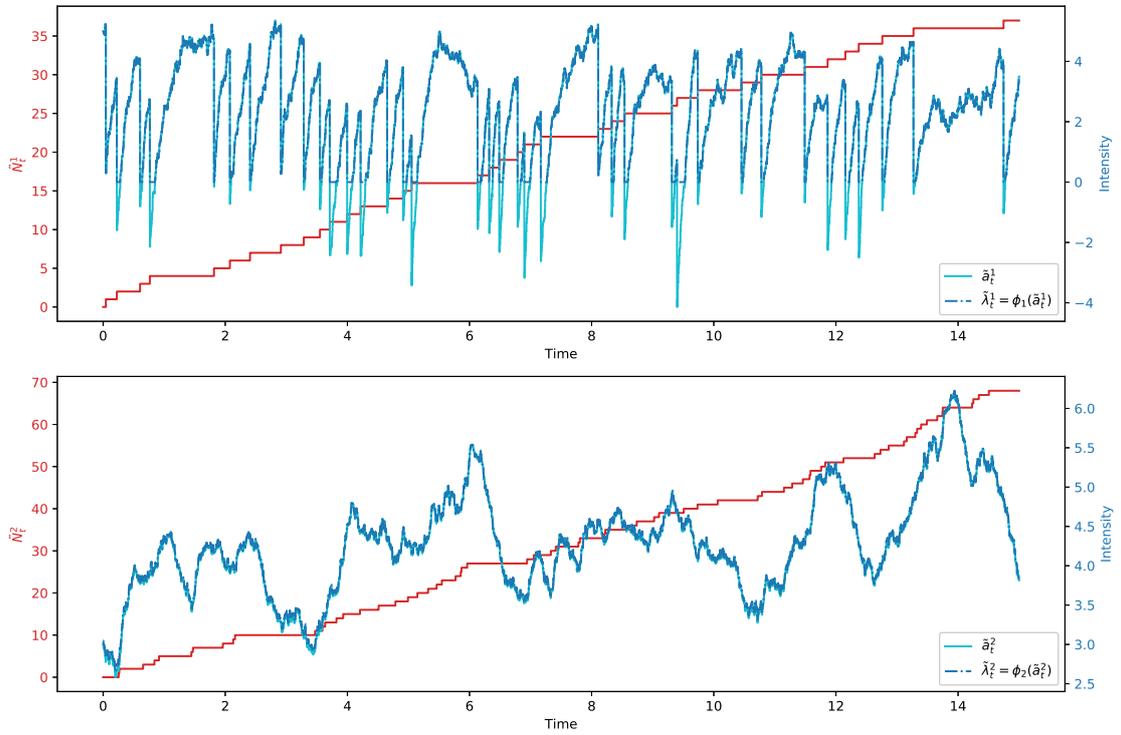


Figure 6: Simulation for  $h = 5 \cdot 10^{-3}$ . The Gaussian perturbation parameters are  $\sigma_1 = 1.5$  and  $\sigma_2 = 0.5$ . The graph is drawn in simple continuous lines for visibility.

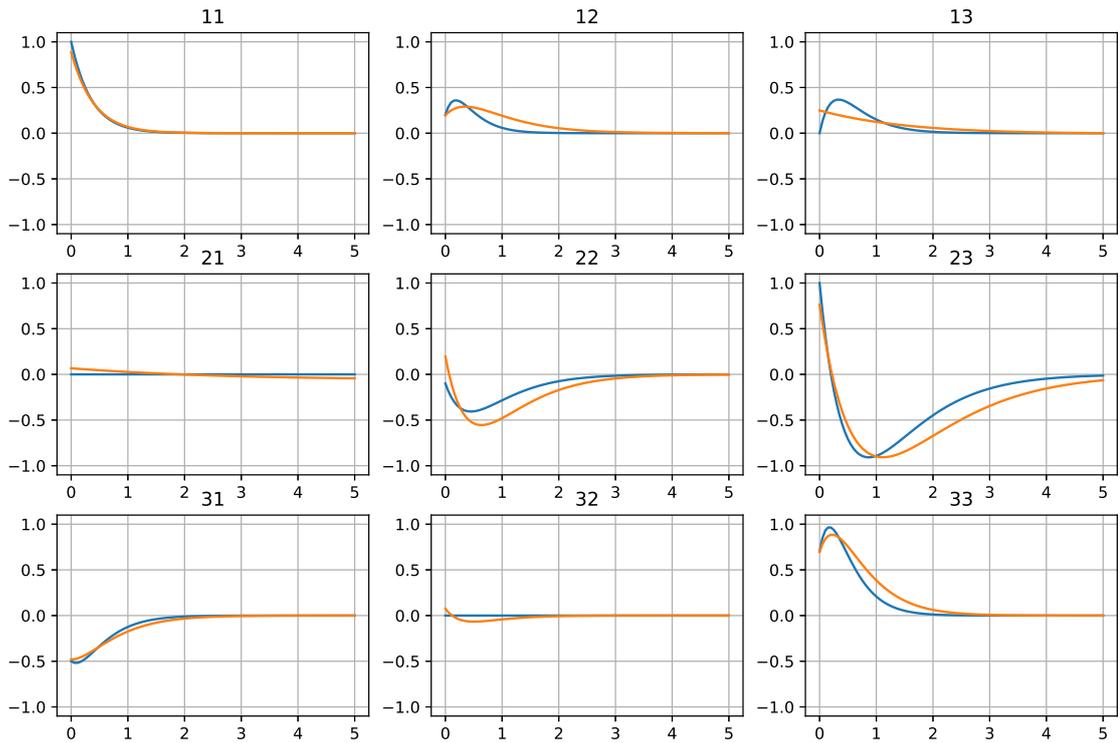


Figure 7: The blue curves are the actual kernels and the orange curves are the calibration results. The estimated baseline intensities are  $\hat{\boldsymbol{\mu}} = (0.37, 1.3, 1.5)$ .

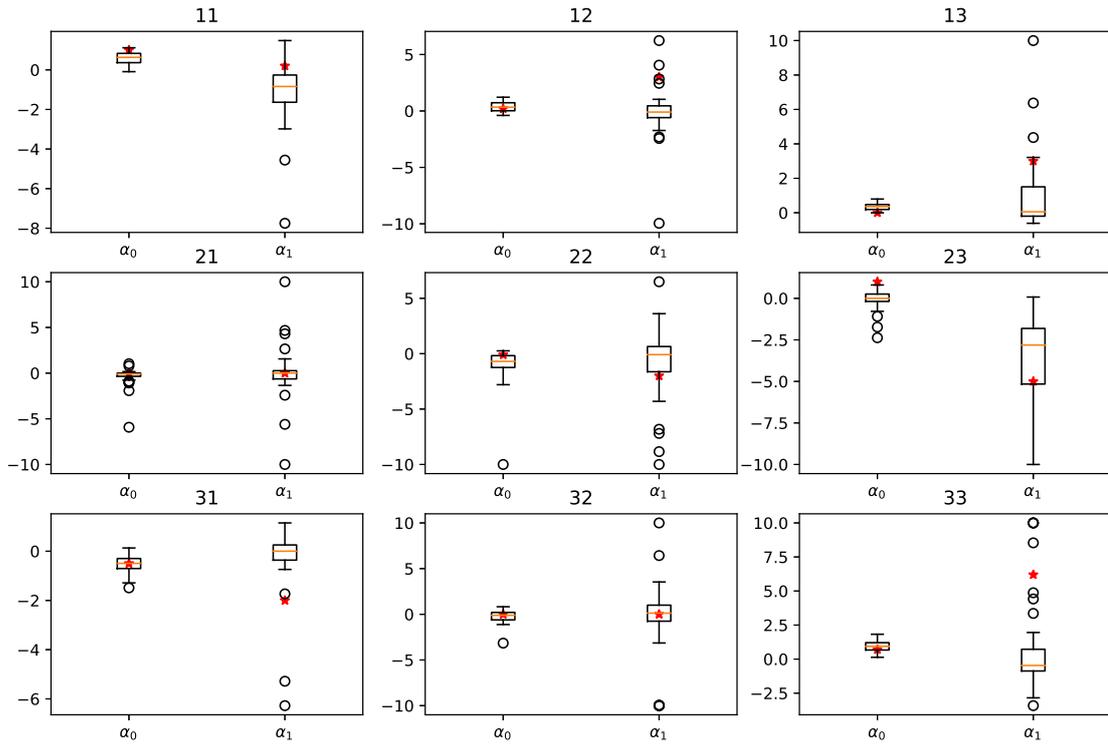


Figure 8: Calibrations of the coefficient  $\alpha_0^{ij}$  and  $\alpha_1^{ij}$  for  $i, j = 1, \dots, d$ . The estimation is clearly less precise for  $\alpha_1$  with a larger variance. The red stars are the markers of the ground truth value.

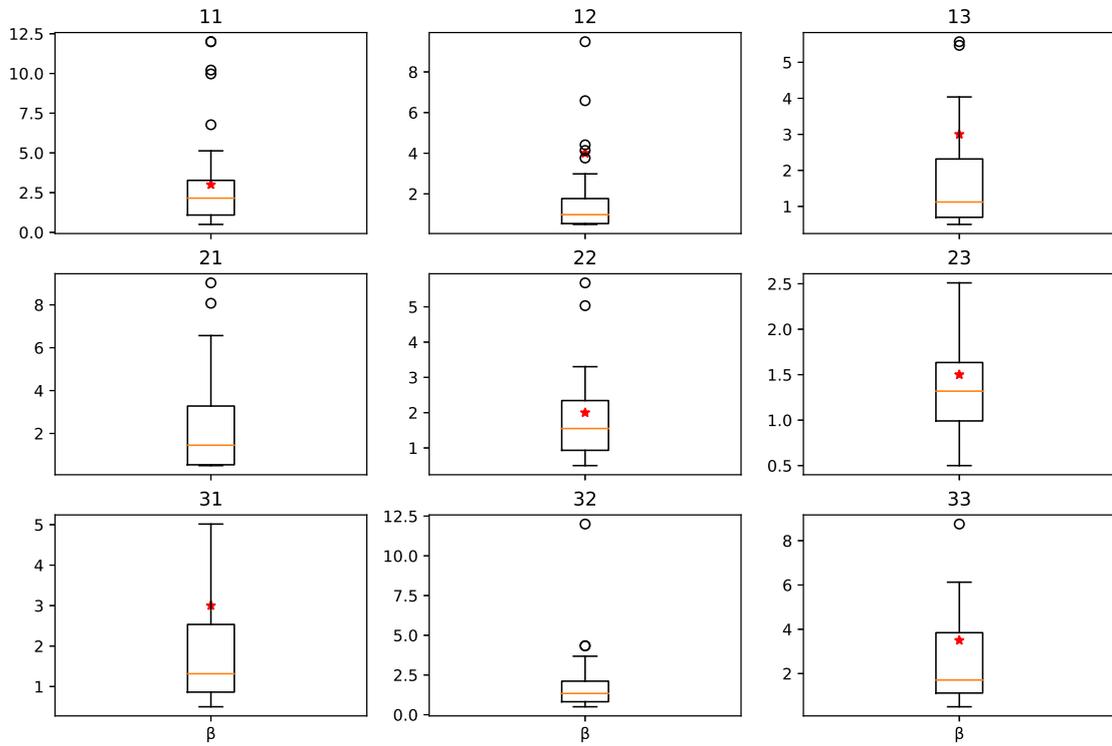


Figure 9:  $\beta_{21}$  and  $\beta_{32}$  are not defined because the kernels are equal to zero.

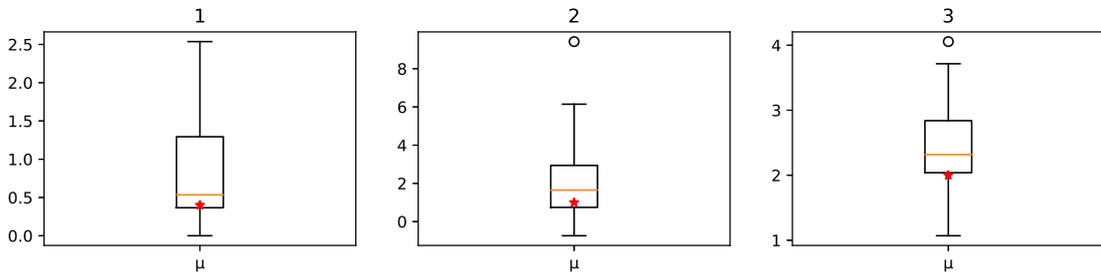


Figure 10: The estimator tends to yield larger values for the baseline intensities. This is due to the aforementioned bias, where some self-excitations are ignored.

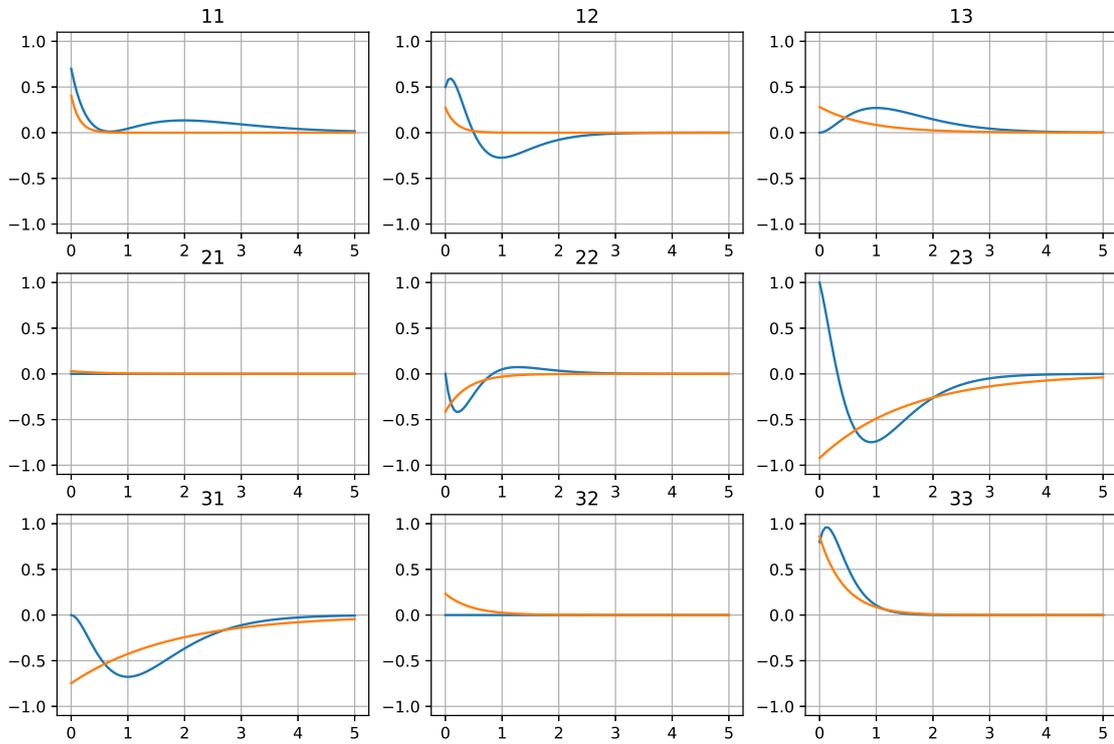


Figure 11: The overall effect (excitation, inhibition, independence) are well captured by the exponential kernels. For composite kernels (e.g.  $\delta_{12}$  and  $\delta_{22}$ ) the exponential kernel captures the more influential behaviour.

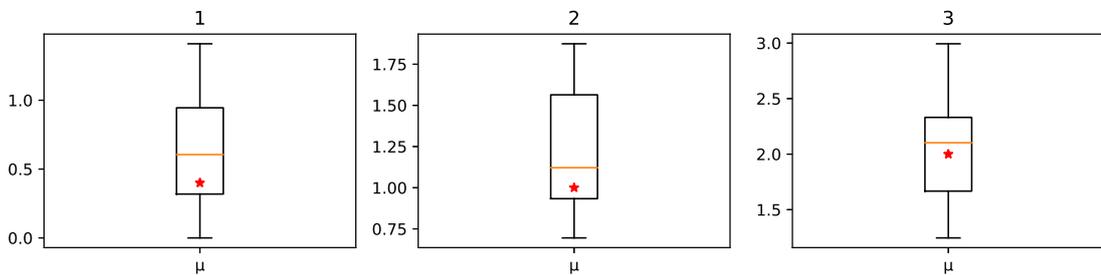


Figure 12: Again, the baseline intensities are over estimated because of the bias.

## References

- [1] Abdulhamid A. Alzaid and Mohamed A. Al-Osh. “An integer-valued  $p$ th-order autoregressive structure (INAR( $p$ )) process”. In: *J. Appl. Probab.* 27.2 (1990), pp. 314–324. ISSN: 0021-9002. DOI: 10.2307/3214650. URL: <https://doi.org/10.2307/3214650>.
- [2] Emmanuel Bacry and Jean-François Muzy. “First- and second-order statistics characterization of Hawkes processes and non-parametric estimation”. In: *IEEE Trans. Inform. Theory* 62.4 (2016), pp. 2184–2202. ISSN: 0018-9448. DOI: 10.1109/TIT.2016.2533397. URL: <https://doi.org/10.1109/TIT.2016.2533397>.
- [3] Emmanuel Bacry et al. “Modelling microstructure noise with mutually exciting point processes”. In: *Quant. Finance* 13.1 (2013), pp. 65–77. ISSN: 1469-7688. DOI: 10.1080/14697688.2011.647054. URL: <https://doi.org/10.1080/14697688.2011.647054>.
- [4] Emmanuel Bacry et al. “Sparse and low-rank multivariate Hawkes processes”. In: *J. Mach. Learn. Res.* 21 (2020), Paper No. 50, 32. ISSN: 1532-4435.
- [5] Henry Berbee. “Chains with infinite connections: uniqueness and Markov representation”. In: *Probab. Theory Related Fields* 76.2 (1987), pp. 243–253. ISSN: 0178-8051. DOI: 10.1007/BF00319986. URL: <https://doi.org/10.1007/BF00319986>.
- [6] Patrick Billingsley. *Convergence of probability measures*. Second. Wiley Series in Probability and Statistics: Probability and Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1999, pp. x+277. ISBN: 0-471-19745-9. DOI: 10.1002/9780470316962. URL: <https://doi.org/10.1002/9780470316962>.
- [7] Björn Böttcher, René Schilling, and Jian Wang. *Lévy matters. III*. Vol. 2099. Lecture Notes in Mathematics. Lévy-type processes: construction, approximation and sample path properties, With a short biography of Paul Lévy by Jean Jacod, Lévy Matters. Springer, Cham, 2013, pp. xviii+199. ISBN: 978-3-319-02683-1; 978-3-319-02684-8. DOI: 10.1007/978-3-319-02684-8. URL: <https://doi.org/10.1007/978-3-319-02684-8>.
- [8] Pierre Brémaud and Laurent Massoulié. “Stability of nonlinear Hawkes processes”. In: *Ann. Probab.* 24.3 (1996), pp. 1563–1588. ISSN: 0091-1798. DOI: 10.1214/aop/1065725193. URL: <https://doi.org/10.1214/aop/1065725193>.
- [9] José Da Fonseca and Riadh Zaatour. “Hawkes Process: Fast Calibration, Application to Trade Clustering, and Diffusive Limit”. In: *Journal of Futures Markets* 34.6 (2014), pp. 548–579. DOI: <https://doi.org/10.1002/fut.21644>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/fut.21644>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/fut.21644>.
- [10] Angelos Dassios and Hongbiao Zhao. “Efficient simulation of clustering jumps with CIR intensity”. In: *Oper. Res.* 65.6 (2017), pp. 1494–1515. ISSN: 0030-364X. DOI: 10.1287/opre.2017.1640. URL: <https://doi.org/10.1287/opre.2017.1640>.
- [11] Aline Duarte, Eva Löcherbach, and Guilherme Ost. “Stability, convergence to equilibrium and simulation of non-linear Hawkes processes with memory kernels given by the sum of Erlang kernels”. In: *ESAIM Probab. Stat.* 23 (2019), pp. 770–796. ISSN: 1292-8100. DOI: 10.1051/ps/2019005. URL: <https://doi.org/10.1051/ps/2019005>.
- [12] Darrell Duffie, Jun Pan, and Kenneth Singleton. “Transform analysis and asset pricing for affine jump-diffusions”. In: *Econometrica* 68.6 (2000), pp. 1343–1376. ISSN: 0012-9682. DOI: 10.1111/1468-0262.00164. URL: <https://doi.org/10.1111/1468-0262.00164>.
- [13] Michael Eichler, Rainer Dahlhaus, and Johannes Dueck. “Graphical modeling for multivariate Hawkes processes with nonparametric link functions”. In: *J. Time Series Anal.* 38.2 (2017), pp. 225–242. ISSN: 0143-9782. DOI: 10.1111/jtsa.12213. URL: <https://doi.org/10.1111/jtsa.12213>.
- [14] Eymen Errais, Kay Giesecke, and Lisa R. Goldberg. “Affine point processes and portfolio credit risk”. In: *SIAM J. Financial Math.* 1.1 (2010), pp. 642–665. DOI: 10.1137/090771272. URL: <https://doi.org/10.1137/090771272>.
- [15] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. Characterization and convergence. John Wiley & Sons, Inc., New York, 1986, pp. x+534. ISBN: 0-471-08186-8. DOI: 10.1002/9780470316658. URL: <https://doi.org/10.1002/9780470316658>.

- [16] Antonio Galbis and Enrique Jordá. “Composition operators on the Schwartz space”. In: *Rev. Mat. Iberoam.* 34.1 (2018), pp. 397–412. ISSN: 0213-2230. DOI: 10.4171/RMI/989. URL: <https://doi.org/10.4171/RMI/989>.
- [17] Kay Giesecke and Baeho Kim. “Estimating tranche spreads by loss process simulation”. In: *2007 Winter Simulation Conference*. 2007, pp. 967–975. DOI: 10.1109/WSC.2007.4419693.
- [18] Alan G. Hawkes. “Spectra of some self-exciting and mutually exciting point processes”. In: *Biometrika* 58 (1971), pp. 83–90. ISSN: 0006-3444. DOI: 10.1093/biomet/58.1.83. URL: <https://doi.org/10.1093/biomet/58.1.83>.
- [19] Alan G. Hawkes and David Oakes. “A cluster process representation of a self-exciting process”. In: *J. Appl. Probability* 11 (1974), pp. 493–503. ISSN: 0021-9002. DOI: 10.2307/3212693. URL: <https://doi.org/10.2307/3212693>.
- [20] Matthias Kirchner. “An estimation procedure for the Hawkes process”. In: *Quant. Finance* 17.4 (2017), pp. 571–595. ISSN: 1469-7688. DOI: 10.1080/14697688.2016.1211312. URL: <https://doi.org/10.1080/14697688.2016.1211312>.
- [21] Matthias Kirchner. “Hawkes and INAR( $\infty$ ) processes”. In: *Stochastic Process. Appl.* 126.8 (2016), pp. 2494–2525. ISSN: 0304-4149. DOI: 10.1016/j.spa.2016.02.008. URL: <https://doi.org/10.1016/j.spa.2016.02.008>.
- [22] Lawrence A. Klimko and Paul I. Nelson. “On conditional least squares estimation for stochastic processes”. In: *Ann. Statist.* 6.3 (1978), pp. 629–642. ISSN: 0090-5364. URL: [http://links.jstor.org/sici?sici=0090-5364\(197805\)6:3%3C629:OCLSEF%3E2.0.CO;2-5&origin=MSN](http://links.jstor.org/sici?sici=0090-5364(197805)6:3%3C629:OCLSEF%3E2.0.CO;2-5&origin=MSN).
- [23] Remi Lemonnier and Nicolas Vayatis. “Nonparametric Markovian Learning of Triggering Kernels for Mutually Exciting and Mutually Inhibiting Multivariate Hawkes Processes”. In: *Machine Learning and Knowledge Discovery in Databases*. Ed. by Toon Calders et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 161–176. ISBN: 978-3-662-44851-9.
- [24] Yosihiko Ogata. “On Lewis’ simulation method for point processes”. In: *IEEE Transactions on Information Theory* 27.1 (1981), pp. 23–31. DOI: 10.1109/TIT.1981.1056305.
- [25] Yosihiko Ogata. “Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes”. In: *Journal of the American Statistical Association* 83.401 (1988), pp. 9–27. ISSN: 01621459. URL: <http://www.jstor.org/stable/2288914> (visited on 06/03/2022).
- [26] T. Ozaki. “Maximum likelihood estimation of Hawkes’ self-exciting point processes”. In: *Ann. Inst. Statist. Math.* 31.1 (1979), pp. 145–155. ISSN: 0020-3157. DOI: 10.1007/BF02480272. URL: <https://doi.org/10.1007/BF02480272>.
- [27] Mads Bonde Raad, Susanne Ditlevsen, and Eva Löcherbach. “Stability and mean-field limits of age dependent Hawkes processes”. In: *Ann. Inst. Henri Poincaré Probab. Stat.* 56.3 (2020), pp. 1958–1990. ISSN: 0246-0203. DOI: 10.1214/19-AIHP1023. URL: <https://doi.org/10.1214/19-AIHP1023>.
- [28] Patricia Reynaud-Bouret and Sophie Schbath. “Adaptive estimation for Hawkes processes; application to genome analysis”. In: *The Annals of Statistics* 38.5 (2010), pp. 2781–2822. DOI: 10.1214/10-AOS806. URL: <https://doi.org/10.1214/10-AOS806>.
- [29] Youngsoo Seol. “Limit theorems for discrete Hawkes processes”. In: *Statist. Probab. Lett.* 99 (2015), pp. 223–229. ISSN: 0167-7152. DOI: 10.1016/j.spl.2015.01.023. URL: <https://doi.org/10.1016/j.spl.2015.01.023>.
- [30] Leigh Shlomovich, Edward AK Cohen, and Niall Adams. “A Parameter Estimation Method for Multivariate Aggregated Hawkes Processes”. In: *arXiv preprint arXiv:2108.12357* (2021).
- [31] Deborah Sulem, Vincent Rivoirard, and Judith Rousseau. “Bayesian estimation of nonlinear Hawkes process”. In: *arXiv preprint arXiv:2103.17164* (2021).