Sorbonne Université

LPSM

Doctoral School **École Doctorale Sciences Mathématiques de Paris Centre**

University Department **Laboratoire de Probabilités, Statistique et Modélisation**

Thesis defended by **Miguel Martínez Herrera**

In order to become Doctor from Sorbonne Université

Academic Field **Applied Mathematics**

Speciality **Statistics**

Inference of non-linear or imperfectly observed Hawkes processes

Thesis supervised by

Arnaud Guyader	Supervisor
Anna Bonnet	Co-Supervisor
Maxime Sangnier	Co-Supervisor

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

LPSM
LABORATOIRE DE PROBABILITÉS STATISTIQUE & MODÉLISATION

Colophon

Doctoral dissertation entitled “Inference of non-linear or imperfectly observed Hawkes processes”, written by Miguel Martínez Herrera, completed on August 21, 2024, typeset with the document preparation system LaTeX and the yathesis class dedicated to theses prepared in France.

This thesis has been prepared at

Laboratoire de Probabilités, Statistique et Modélisation

Sorbonne Université
Campus Pierre et Marie Curie
4 place Jussieu
75005 Paris
France

☎ +33 1 57 27 93 16
Web Site `https://www.lpsm.paris/`

Inference of non-linear or imperfectly observed Hawkes processes

Abstract

The Hawkes point process is a popular statistical tool to analyse temporal patterns. Modern applications propose extensions of this model to account for specificities in each field of study, which in turn complexifies the task of inference. In this thesis, we advance different approaches for the parametric estimation of two submodels of the Hawkes process in univariate and multivariate settings. Motivated by the modelling of complex neuronal interactions observed from spike train data, our first study focuses on accounting for both inhibition and excitation effects between neurons, modelled by the non-linear Hawkes process. We derive a closed-form expression of the log-likelihood in order to implement a maximum likelihood procedure. As a consequence of our approach, we gain access to a goodness-of-fit scheme allowing us to establish ad hoc model selection methods to estimate the interaction network in the multivariate setting. The second part of this thesis focuses on studying Hawkes process data noised by two different alterations: adding or removing points. The absence of knowledge on the noise dynamics makes classical inference procedures intractable or computationally expensive. Our solution is to leverage the spectral analysis of point processes to establish an estimator obtained by maximising the spectral log-likelihood. By deriving the spectral densities of the noised processes and by establishing identifiability conditions on our model, we show that the spectral inference method does not necessitate any information on the structure of the noise, effectively circumventing this issue. An additional result of the study of Hawkes processes with missing points is that it gives access to a subsampling paradigm to enhance the estimation methods by introducing a penalisation parameter. We illustrate the efficiency of all of our methods through reproducible numerical implementations.

Keywords: hawkes processes, parametric inference, identifiability, inhibition, spectral theory, neuronal data

Résumé

Le processus ponctuel de Hawkes est un outil statistique très répandu pour analyser des dynamiques temporelles. Les applications modernes des processus de Hawkes proposent des extensions du modèle initial pour prendre en compte certaines caractéristiques spécifiques à chaque domaine d'étude, ce qui complexifie les tâches d'inférence. Dans cette thèse, nous proposons différentes contributions à l'estimation paramétrique de deux variantes du processus de Hawkes dans les cadres univarié et multivarié. Motivée par la modélisation d'interactions complexes au sein d'une population de neurones, notre première étude porte sur la prise en compte conjointe d'effets excitateurs et inhibiteurs entre les signaux émis par les neurones au cours du temps, modélisés par un processus de Hawkes non-linéaire. Dans ce modèle, nous obtenons une expression explicite de la log-vraisemblance qui nous permet d'implémenter une procédure de maximum de vraisemblance. Nous établissons également une méthode de sélection de modèle qui fournit notamment une estimation du réseau d'interactions dans le cadre multivarié. La deuxième partie de cette thèse est consacrée à l'étude des processus de Hawkes bruités par deux types d'altérations : l'ajout ou la suppression de certains points. Le manque d'information lié à ces mécanismes de bruit rend les méthodes classiques d'inférence non-applicables ou numériquement coûteuses. Notre solution consiste à s'appuyer sur l'analyse spectrale des processus ponctuels afin d'établir un estimateur obtenu en maximisant la log-vraisemblance spectrale. Nous obtenons l'expression des densités spectrales des processus bruités et, après avoir établi des conditions d'identifiabilité pour nos différents modèles, nous montrons que cette méthode d'inférence ne nécessite pas de connaître la structure du bruit, contournant ainsi le problème d'estimation. Notre étude sur les processus bruités donne accès à une méthode de sous-échantillonnage qui nous permet d'améliorer les approches d'estimation en introduisant un paramètre de pénalisation. Nous illustrons la performance des différentes méthodes proposées à travers des implémentations numériques reproductibles.

Mots clés : processsus de hawkes, inférence paramétrique, identifiabilité, inhibition, théorie spectrale, données neuronales

Laboratoire de Probabilités, Statistique et Modélisation
Sorbonne Université – Campus Pierre et Marie Curie – 4 place Jussieu – 75005 Paris – France

Contents

1 Introduction

This chapter is a general introduction to Hawkes processes and the challenges explored in this manuscript. After a succinct presentation of Hawkes processes with excitation, we contextualise the state-of-the-art literature concerning estimation methods in Section 1.1 and present the main questions that guide our research. This allows us to exhibit the two paradigms that are studied in this work: inhibition and imperfect data. Section 1.2 pertains to the parametric estimation of both univariate and multivariate Hawkes processes with potential inhibiting interactions (detailed in Chapters 3 and 4). Section 1.3 presents our contributions to the study of imperfectly observed exciting Hawkes processes (detailed in Chapters 5 and 6). A general outline of this manuscript is described in Section 1.4.

Outline of the current chapter

1.1 Statistics for Hawkes processes

1.1.1 The self-exciting point process

In probability and statistics, modelling random collections of points in a certain space is commonly done through a point process. When studying point processes on the real line $\mathbb{R}$ (or the half line $\mathbb{R}_{\geq 0}$), the natural order of this space often incurs an ordering of any countable sequence of points $(T_k)_{k\in\mathbb{Z}}$: we talk then of *temporal* point processes.

In statistics, it is a common question to analyse the dynamics describing the occurrences of a certain phenomenon: the time of arrival of buses at a bus stop, the apparition of symptoms in a population or earthquake incidents in a region of the world. The simplest model for point processes is the homogeneous Poisson process, where the waiting times between any two event times are i.i.d distributed as an exponential random variables with parameter $\lambda > 0$, known as the intensity of the process. A natural extension of this model is obtained by allowing the intensity to be a deterministic non-negative function $\lambda : \mathbb{R} \to \mathbb{R}_{\geq 0}$, adding a temporal dependence on point arrivals.

The Hawkes process

In 1971, Alan G. Hawkes introduces a past-dependent model, initially called self-exciting point processes (Hawkes 1971), which will later be known as the Hawkes point process. Let $\mathcal{H}_t = \sigma(\{T_k \mid k \in \mathbb{Z}, T_k \leq t\})$ denote the past history of a process N for any $t \in \mathbb{R}$, we define the conditional intensity function $\lambda : \mathbb{R} \to \mathbb{R}_{>0}$ of this process as:

$$\lambda(t \mid \mathcal{H}_t) = \lim_{h\to 0} \frac{\mathbb{E}[N([t, t+h]) \mid \mathcal{H}_t]}{h} .$$

The Hawkes process is then defined by the following expression of the conditional intensity function:

$$\lambda(t \mid \mathcal{H}_t) = \mu + \sum_{T_k \leq t} h(t - T_k) . \tag{1.1}$$

Remark. *In the literature, it is common to omit writing the history $\mathcal{H}_t$ as it is directly implied that λ has access to the entire past of the process. We follow this convention throughout this work, unless marked otherwise.*

The term μ is often referred to as the baseline intensity which dictates a constant rate of occurrences, similar to a homogeneous Poisson process. The dependence on the past is represented by the second term in Equation (1.1), where each event time that precedes t contributes to the intensity function through the interaction function $h : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, also calles the kernel.

Positivity of h represents an excitation effect between points: each point T_k in the past increases the value of the intensity function, which in turn increases the rate at which points in the future appear. For stability reasons, h is assumed to converge to 0 as $t \to +\infty$, representing the rate at which the effects from the past are "forgotten".

Different shapes of h allow to account for different effects. On the one hand, strictly decreasing functions such as exponential (Ozaki 1979; Ogata 1988) or power-law kernels (Zhang 2016) represent instantaneous effects, which appear as spikes in the conditional intensity function (Equation (1.1)). On the other hand, Rayleigh or gamma kernels (Lesage et al. 2022), where the maximum of the function is not at the origin $t = 0$, can be chosen to represent delayed effects. Figure 1.1 illustrates two Hawkes processes with two different kernels. The properties of the conditional intensity function are intrinsically connected to those of h, and so when working

in parametric settings, the choice of the kernel is essential and comes with its advantages and inconveniences.

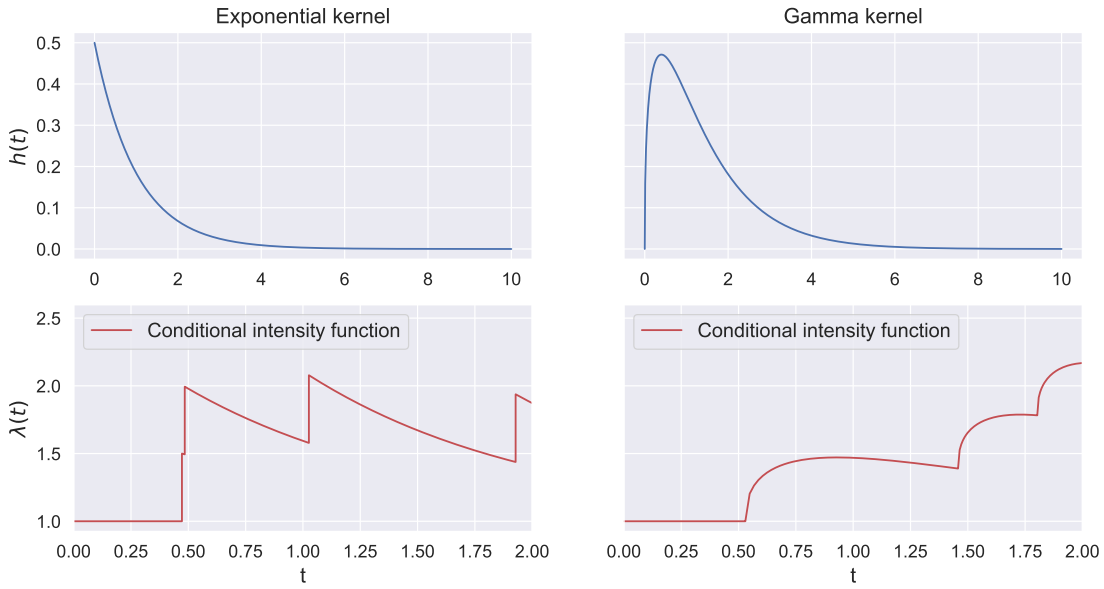

Figure 1.1: Kernel functions (top) and respective conditional intensity functions (bottom) of a Hawkes process started at $t = 0$ with baseline intensity $\mu = 1$.

Clustering and branching structures

A direct consequence of the expression of the intensity function (Equation (1.1)) is that the Hawkes process model can be interpreted as a Poisson cluster process (Hawkes et al. 1974). The most practical way of defining a cluster process (Bartlett 1963) is from a generative point of view. We begin by simulating a homogenenous Poisson process on the real line $\mathbb{R}$ with parameter μ: these points T_k^c are often called ancestors, immigrants or cluster centers. Each ancestor generates an inhomogeneous Poisson process with intensity function $h(\cdot - T_k^c)$, forming a family of children points, also called descendants. The iteration is repeated with each new point generating its own subprocess until no descendants are generated. In the end, the cluster process is formed by the union of both ancestors and descendants. The particularity of a Hawkes process is that the support of the function h is a subset of $\mathbb{R}_{\geq 0}$, meaning that each occurrence influences solely the future of the process.

This kind of process is also known as the Poisson branching process (Lewis 1964) as it describes a generation dynamic similar to the Galton-Watson branching process (see Watson et al. (1875) for the discrete time version and Harris (2002, Chapter III) for the generalised version). More precisely, the Hawkes process can be seen as a time-continuous branching process with immigration: let us assume we observe the arrival of an immigrant at a time T_k^c. This immigrant will produce a first generation of children, forming the first set of points, or branches. Then, each child becomes a parent and generates a new set of branches. A tree is then formed by the union of each immigrant and all of its branches, and then the Hawkes process is once again formed by the union of all trees.

A visualisation of both structures is illustrated in Figure 1.2 for a Hawkes process with

exponential kernel. The clustering (equivalently branching) structure is one of the properties that made Hawkes processes so attractive in the literature. From a theoretical point of view, the many existing developments of branching theory allowed to quickly obtain results concerning existence, stability and stationarity. A main example concerns the existence of a self-exciting Hawkes process: in order for the point process to have a finite number of points inside any bounded set, a necessary and sufficient condition (Hawkes et al. 1974, Lemma 1) is:

$$\int_0^{+\infty} |h(t)| \, \mathrm{d}t < 1 \,,$$

which is derived from the subcriticality condition of Galton-Watson processes.

From a numerical point of view, the definition of a Poisson cluster process, as given above, provides a simulation algorithm. An inference method inspired from branching process theory was adapted to the study of Hawkes processes (Veen et al. 2008).

The aforementioned properties, and the quick developments that followed in the literature, created a growing interest in other scientific fields. From this, different versions of the original self-exciting Hawkes process appeared in the literature, each one trying to accommodate to more complex phenomena.

The multivariate Hawkes process

Probably the most natural and useful extension is the definition of the *multivariate* Hawkes process. Instead of studying a single phenomenon (a univariate process), we define a d-variate Hawkes process as d individual point processes $(N_i)_{i=1:d}$ interacting together, where each process is defined by a conditional intensity function λ^i:

$$\lambda^i(t) = \mu_i + \sum_{j=1}^{d} \sum_{T_k^j \leq t} h_{ij}(t - T_k^j) \,, \qquad \text{for any } t \in \mathbb{R}.$$

In this formulation, each process N_j has its own event times $(T_k^j)_{k\in\mathbb{Z}}$ and baseline intensity $\mu_j > 0$. This model introduces interactions between processes, which appear in the second term of the intensity. The kernel functions h_{ij} represent the effect of points from N_j on process N_i. This model allows to study a group of individuals linked by a network of interactions, adding a new dimension of study for such processes.

From modelling infection spread all the way to analysing clusters of earthquakes, the study of Hawkes processes from a statistical point of view quickly became a central topic of interest.

1.1.2 Inference and applications

Statistical estimation

Statistical challenges for Hawkes processes with excitation focus around estimating both the baseline intensity μ and the interaction function h.

To our knowledge, the very first paper implementing an estimation procedure in a parametric framework is Adamopoulos (1976). In his work, the author proposes a study of exponential kernels for both univariate and bivariate Hawkes processes through the spectral log-likelihood, closely related to time series theory. The exponential kernel is often chosen as a parametric interaction function because in this case the univariate Hawkes process becomes a Markov process which presents the advantage of simplifying the expression of the intensity function. This is

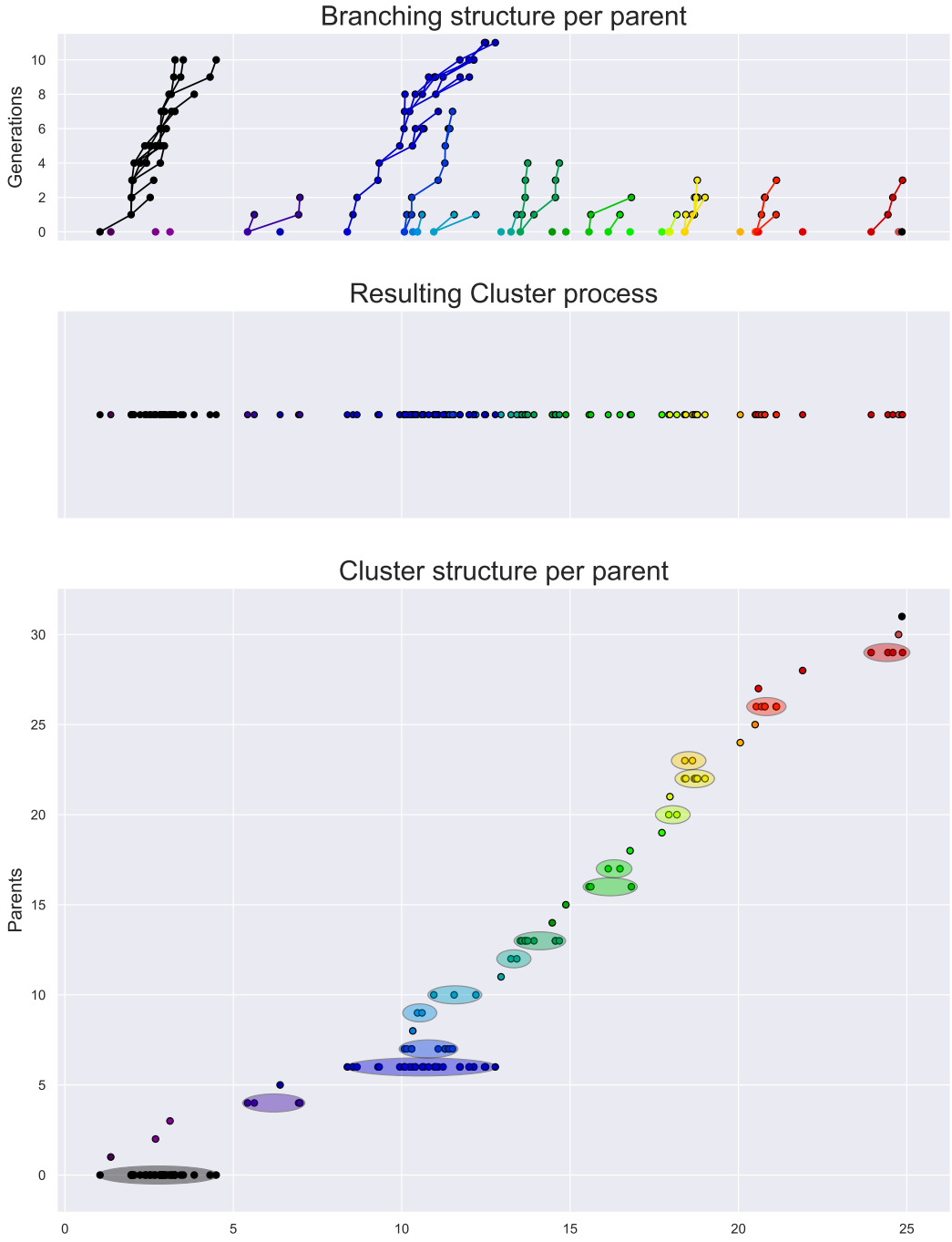

Figure 1.2: Illustration of both branching (top) and clustering (bottom) structures of an exponential Hawkes process (middle). Each color represents a single cluster/tree. Simulation is done thanks to the cluster process algorithm.

exemplified in the implementation of the maximum likelihood estimator (MLE) in Ozaki (1979). The expression of the log-likelihood of a general point process, for an observation in $[0, T]$, is:

$$\ell_T(\theta) = -\int_0^T \lambda(u)\,\mathrm{d}u + \sum_{k=1}^{N(T)} \log(\lambda(T_k^-))\,, \tag{1.2}$$

where $N(T)$ is the number of points in the observation window. Other works address estimation in the parametric setting, such as Ogata (1988) for other kernel functions using the MLE, Veen et al. (2008) leveraging the branching structure, Bacry et al. (2020) minimising a least-squares contrast and Da Fonseca et al. (2013) via the method of moments.

The literature regarding non-parametric inference is also vast with implementations of the MLE in Guo et al. (2018) and a penalised variant optimised by an Expectation-Maximisation algorithm. The least-squares minimisation is often used as shown in Reynaud-Bouret et al. (2009), Eichler et al. (2016), and Kirchner (2017). Other methods include solving Wiener-Hopf equations (Bacry et al. 2016) or approximating the interaction functions with autoregressive models (Kirchner 2017).

All of these references concern frequentist approaches of statistics, and so it is essential to mention that an equal effort has been made from a Bayesian point of view. Rasmussen (2013) proposes two procedures through log-likelihood estimation: one through the classical intensity function and another similar to Veen et al. (2008) with the branching structure. In Lemonnier et al. (2014), the authors propose an approximation through exponential kernels and taking advantage of the Markovian properties. The multivariate case is deeply studied in Donnet et al. (2020) with illustrations on estimating the underlying interaction graph.

This is jus a small sample of the plethora of approaches that have been developed in order to study these kind of processes.

Applications and generalisations

The temporal-dependency structure of Hawkes process has motivated its use accross a variety of application fields with many generalisations to obtain more explicative models.

The historical example is the study of seismic activity where the occurrence of an earthquake is usually followed by a number of smaller tremors known as aftershocks. This behaviour is similar to the self-exciting effect modelled by the Hawkes process and associated with the clustering structure, as shown in Adamopoulos (1976). As in seismology the spatial placement and magnitude of each quake are important factors to take into account, applications in this field tend to include this information via the concept of marked point processes. Each event time has an associated mark representing the detected magnitude (Ogata 1988) and the location of each epicenter (Ogata 1998; Kwon et al. 2023).

A similar approach is proposed in the study of social media interactions for the analysis of subject trends. The excitation effect appears in the form of reposting where users have the option to share a news post among their followers which in turn may continue to spread the information in their social circles. In this context, the impact of each repost is dependent on the influence of the account and the impact of a topic on a certain community. This is represented for example by including information in the form of number of followers (Mishra et al. 2016) again through marks, or by adding an additional dimension to the process (Pinto et al. 2015) to represent the impact of a topic over another.

In criminology, it is commonly assumed that delinquent acts may incite other crimes in a population, which can be modelled by a self-exciting Hawkes process. As a branch of social sciences, it is important to account for different factors that affect human behaviour such as time

of the day, demographics and social interactions. A way of accounting for these effects is to adapt the Hawkes process to time and spatial dependent baseline intensities. A time-varying baseline intensity can represent periodic criminal behaviour (Lewis et al. 2011) with higher values for nightime. A space-dependent baseline is used to represent the link between residential density and burglary (Mohler et al. 2011) or the spatial distributions of gangs in a city (Linderman et al. 2014).

Many other fields include finance (Embrechts et al. 2011; Bacry et al. 2013; Roueff et al. 2019; Lotz 2024), genomics (Reynaud-Bouret et al. 2009; Carstensen et al. 2010), ecology (Denis et al. 2024; Nicvert et al. 2024), epidemiology (Rizoiu et al. 2018; Chiang et al. 2022), TV browsing behaviour (Xu et al. 2016), event data streams in football (Baouan et al. 2023; Narayanan et al. 2023), each one often accompanied with new formulations of the Hawkes process.

1.1.3 Challenges and contributions

Our work is motivated by the applications of Hawkes processes in neurobiology for the study of neuronal activity data (Reynaud-Bouret et al. 2013; Lambert et al. 2018; Duval et al. 2022). Neurons in the brain are connected through a deep network of synapses allowing them to communicate through electrical impulses usually studied by measuring the action potential or spikes emitted as the membrane potential of the cells is depolarised. These measurements generate a spike train of instants when a neuron activates and influences the membrane potential of connecting neighbours.

These exchanges can appear either as excitatory or inhibitory to respectively incite or stop other neurons from activating. The first effect can be clearly modelled by a multivariate Hawkes process but the original modelling does not account for inhibiting effects. In practice, this requires to allow for the interaction functions h_{ij} to take negative values, introducing the concept of Hawkes processes with inhibition.

In order to guarantee the non-negativity of the intensity function, a common practice in the literature is to consider the concept of *non-linear Hawkes processes*. The main difficulty is that the branching structure of Hawkes processes is not valid any longer and the intensity function presents more complex behaviour, so pre-established inference methods are not accessible anymore. The first part of this manuscript pertains to the study of inhibiting Hawkes processes in order to propose estimation methods in the parametric setting. We present an overview of the studied model and our contributions in Section 1.2.

The second part of this manuscript is focused around studying imperfect data. Collecting spike trains data is a procedure that can present measurement errors, which can appear as missing neuronal activation instants or by attributing spikes to the wrong neuron.

This setting represents a common statistical framework of accounting for missing data in an observed process. In our context, not having a full knowledge on the past history of the process makes the conditional intensity function intractable. Our contributions, as summarised in Section 1.3, are focused on exhibiting an inference paradigm through the spectral analysis of point processes for observations noised either by adding points from an external process or by deleting event times.

Although the models inbetween chapters may differ, our contributions follow a common thread by addressing the following four questions.

- What approaches can we take to establish parametric estimation procedures for more complex Hawkes processes dynamics?
- Under which conditions are our statistical models identifiable?
- How to evaluate and select the best estimators through data-driven methods?
- Which modern statistical tools can we leverage to improve our inference procedures?

1.2 Factoring in inhibition for Hawkes processes

1.2.1 Context and related works

The original Hawkes process was proposed as a way of modelling the effect of excitation between points, which is characterised by the positivity of the interaction function h. Our first goal in this work is to study the opposite effect which is commonly known in the literature as *inhibition*.

An inhibiting Hawkes process consists in modelling repulsion between points: each event will reduce the chances of others occurring during a certain period of time. Mathematically, a way of translating this effect is by allowing h to take negative values. However, the non-negativity condition on λ prevents us from adopting this approach without adding some constraints.

One of the most common solutions is the *non-linear* Hawkes process model. In the univariate setting, let $h : \mathbb{R}_{\geq 0} \to \mathbb{R}$ and $\Phi : \mathbb{R} \to \mathbb{R}_{\geq 0}$ be two measurable functions. We define a univariate non-linear Hawkes process N on the real half-line $\mathbb{R}_{\geq 0}$ with event times $(T_k)_{k\in\mathbb{N}}$ by the conditional intensity function, for any $t \in \mathbb{R}_{\geq 0}$:

$$\lambda(t) = \Phi\left(\mu + \sum_{T_k \leq t} h(t - T_k)\right) . \tag{1.3}$$

By allowing h to be negative, the function Φ has to be a non-linear function. Existence of such processes is ensured as long as Φ is an L-Lipschitz function (Brémaud et al. 1996, Theorem 1) such that:

$$L \int_0^{+\infty} |h(t)| \, \mathrm{d}t < 1 .$$

Multiple choices exist in the literature like a clipped exponential function (Chornoboy et al. 1988; Carstensen et al. 2010; Gerhard et al. 2017), a softplus function (Mei et al. 2017), a sigmoid function (Menon et al. 2018), among others.

In our work, we choose the positive part (ReLU) $\Phi(\cdot) = (\cdot)^+ = \max(0, \cdot)$ like in Lemonnier et al. (2014), Hansen et al. (2015), Lu et al. (2018), and Costa et al. (2020). The intensity function (Equation (1.3)) becomes:

$$\lambda(t) = \left(\mu + \sum_{T_k \leq t} h(t - T_k)\right)^+ , \tag{1.4}$$

and the extension to a d-variate Hawkes process is:

$$\lambda^i(t) = \left(\mu_i + \sum_{j=1}^{d} \sum_{T_k^j \leq t} h_{ij}(t - T_k^j) \right)^+ , \tag{1.5}$$

for each process N_i.

Our general contribution in the context of Hawkes processes with inhibition is to provide parametric estimation procedures in a frequentist framework through maximum likelihood estimation. To our knowledge, and by the time of publication of both corresponding papers, other frequentist approaches (Reynaud-Bouret et al. 2014; Bacry et al. 2016) do not model inhibition but can provide negative estimations. In Bayesian contexts, works modelling non-linear Hawkes processes focus around the log-likelihood. In Deutsch et al. (2023), the compensator is approximated through a Simpson's rule to obtain posterior samples in a parametric setting. Works in non-parametric settings include Sulem et al. (2024) that derives posterior concentration rates and Sulem et al. (2023) proposing a variational Bayes method and a sparsity-inducing version based on a model selection paradigm.

1.2.2 Maximum Likelihood Estimation for Hawkes Processes with self-excitation or inhibition

In Chapter 3, we focus on the study of the univariate Hawkes process N with intensity function described by Equation (1.4). In order to establish the maximum likelihood estimator for a parametrised model of the intensity with parameter $\theta \in \Theta$, it is necessary to obtain a closed-form expression of the log-likelihood:

$$\ell_T(\theta) = -\int_0^T \lambda_\theta(u)\,\mathrm{d}u + \sum_{k=1}^{N(T)} \log(\lambda_\theta(T_k^-))\,.$$

For Hawkes processes with excitation, the linearity of the intensity function results in an explicit expression of ℓ_T without much trouble, as shown in Ozaki (1979). The main difficulty when including inhibition is that the cumulated effects from the past events may saturate process N, meaning that its intensity is null for a certain period of time. In the work of Lemonnier et al. (2014), the authors propose to approximate the computation in this context by ignoring the non-linear function. This allows them to circumvent this problem at the cost of assuming that the inhibition effects are small enough to be negligible.

In order for us to obtain an exact computation of the log-likelihood that accounts for inhibition, we introduce two novel concepts: the *underlying* intensity function $\lambda^\star$ and the *restart times* $(T_k^\star)_{k\geq 1}$. We define $\lambda^\star \colon \mathbb{R}_{\geq 0} \to \mathbb{R}$, for any $t \geq 0$, as:

$$\lambda^\star(t) = \mu + \sum_{T_k \leq t} h(t - T_k)\,,$$

and the restart times $T_k^\star$ for any integer $k > 0$ as:

$$T_k^\star = \inf\{t \geq T_k \mid \lambda(t) > 0\}\,.$$

The advantage of working with $\lambda^\star$ instead of λ is that it inherits properties such as smoothness and strict monotonicity from the interaction function h between any two consecutive event times

T_k and T_{k+1}. If assumed so, it follows that $T_k^\star$ is the moment from which both functions coincide. Our main contribution is to provide a closed-form expression of the compensator $\int_0^T \lambda_\theta(u)\,\mathrm{d}u$, which we establish in Proposition 1.2.1.

Proposition 1.2.1. *For any $t > 0$:*

$$\int_0^t \lambda(u)\,\mathrm{d}u = \begin{cases} \mu t & \text{if } t < T_1 \\ \mu T_1 + \displaystyle\sum_{k=1}^{N(t)-1} \int_{T_k^\star}^{T_{k+1}} \lambda^\star(u)\,\mathrm{d}u + \int_{T_{N(t)}^\star}^{t} \lambda^\star(u)\,\mathrm{d}u & \text{if } t \geq T_1\,, \end{cases} \tag{1.6}$$

with the conventions that the sum is equal to 0 if $N(t) = 1$ and the last integral is equal to 0 if $t < T^*_{N(t)}$.

If we consider the exponential kernel $h\colon t \mapsto \alpha \mathrm{e}^{-\beta t}$ for parameters $\alpha > 0$, $\beta > 0$, then the restart times read:

$$T_k^\star = T_k + \beta^{-1} \log\left(\frac{\lambda_0 - \lambda^\star(T_k)}{\lambda_0}\right) \mathbb{1}_{\lambda^\star(T_k) < 0}\,,$$

and, for any integer $k \geq 1$ and any $\tau \in [T_k^*, T_{k+1}]$, each integral simplifies to :

$$\int_{T_k^\star}^{\tau} \lambda^\star(u)\,\mathrm{d}u = \lambda_0(\tau - T_k^\star) + \beta^{-1}(\lambda^\star(T_k) - \lambda_0)(\mathrm{e}^{-\beta(T_k^\star - T_k)} - \mathrm{e}^{-\beta(\tau - T_k)})\,.$$

This result allows to establish the maximum likelihood estimation procedure that accounts for both self-exciting and self-inhibiting effects, while keeping the same computational complexity as the method implemented in previous works. In particular, for the exponential kernel function, the computational complexity of ℓ_T is $O(N(T))$ thanks to the Markovian property and, as shown in our numerical study, the derived estimations are better than by using an approximation framework (as presented in Lemonnier et al. (2014)) especially for high levels of inhibition effects.

A second consequence of Proposition 1.2.1 is that it gives access to a goodness-of-fit measure for our estimations by leveraging the Time Change theorem of point processes (Daley et al. 2003, Theorem 7.4.IV). In particular, this allows us to introduce a hypothesis testing procedure that evaluates the quality of our estimations on an independent set of observations.

1.2.3 Inference of multivariate exponential Hawkes processes with inhibition and application to neuronal activity

Chapter 4 builds upon the bases established in Chapter 3 to derive an estimation procedure for multivariate Hawkes processes $N = (N_1, \ldots, N_d)$. In this setting the log-likelihood of N reads:

$$\ell_T(\theta) = \sum_{i=1}^{d} \left(-\int_0^T \lambda_\theta^i(u)\,\mathrm{d}u + \sum_{k=1}^{N_i(T)} \log(\lambda_\theta^i(T_k^{i-})) \right),$$

where λ_θ^i and $(T_k^i)_k$ are respectively the candidate intensity and event times of N_i. Process N can be seen as a univariate point process with its event times $(T_{(k)})_k$ being the ordered union of $(T_k^i)_k$ for all integers $i = \{1, \ldots, d\}$. To alleviate the notations, we will omit the subscript θ unless said otherwise. Similarly to the univariate case, we propose to study the underlying

intensity functions $\lambda^{i\star}$:

$$\lambda^{i\star}(t) = \mu_i + \sum_{j=1}^{d} \sum_{T_k^j \leq t} h_{ij}(t - T_k^j)\,,$$

but this time the monotonicity condition of h_{ij} is not enough to retrieve useful properties for this functions.

This is a consequence of the increased complexity of the dynamics of each subprocess that allows both exciting and inhibiting effects to exist simultaneously. In the multivariate setting, it is possible for a subprocess N_i to be saturated ($\lambda^i(t) = 0$) at an instant t and to receive an external excitation of another subprocess. So, in order to simplify our problem, we restrict our study to the exponential kernel case $h_{ij}(t) = \alpha_{ij}\mathrm{e}^{-\beta_{ij}t}$ with the following mild assumption.

Assumption 1. *For each $i \in \{1, \dots, d\}$, there exists $\beta_i \in \mathbb{R}_{>0}$ such that $\beta_{ij} = \beta_i$ for all $j \in \{1, \dots, d\}$.*

This allows us to recover the strict monotony of $\lambda^{i\star}$ between any two event times $T_{(k)}$ and $T_{(k+1)}$ of the general process N. With this, we adapt the expression of the restart times $(T_{(k)}^{i\star})_{k\geq 1}$ and the compensator $\int_0^t \lambda^i(u)\,\mathrm{d}u$ for each process N_i and every event time $T_{(k)}$, allowing us to establish our first contribution in the form of Proposition 1.2.2.

Proposition 1.2.2. *Let us suppose that Assumption 1 is granted. Then, for each $i \in \{1, \dots, d\}$ and any $k \geq 1$, we have:*

$$T_{(k)}^{i\star} = \min\left(t_k^\star, T_{(k+1)}\right). \tag{1.7}$$

where

$$t_k^\star = \left(T_{(k)} + \beta_i^{-1}\log\left(\frac{\mu_i - \lambda^{i\star}(T_{(k)})}{\mu_i}\right)\mathbb{1}_{\{\lambda^{i\star}(T_{(k)})<0\}}\right).$$

And, for each $i \in \{1, \dots, d\}$, the compensator of the process N_i reads, for any $t \geq 0$:

$$\int_0^t \lambda^i(u)\,\mathrm{d}u = \begin{cases} \mu_i t & \text{if } t < T_{(1)} \\ \mu_i T_{(1)} + \sum_{k=1}^{N(t)} J_k & \text{if } t \geq T_{(1)}\,, \end{cases} \tag{1.8}$$

where for all integer $k \in \{1, \dots, N(t)\}$:

$$J_k = \mu_i\left[\min(t, T_{(k+1)}) - T_{(k)}^{i\star}\right] + \beta_i^{-1}\left(\lambda^{i\star}(T_{(k)}) - \mu_i\right)\left[\mathrm{e}^{-\beta_i(T_{(k)}^{i\star} - T_{(k)})} - \mathrm{e}^{-\beta_i(\min(t, T_{(k+1)}) - T_{(k)})}\right].$$

With this result, we establish an exact procedure to compute the log-likelihood ℓ_T^i of each process N_i, giving us access to the complete log-likelihood ℓ_T and in turn, access to the maximum likelihood estimation method. Let us note that as in the univariate case, we also get access to a goodness-of-fit procedure via the Time Change theorem.

Another complication in the multivariate setting is that the identifiability of our statistical model for the exponential Hawkes process is not as straightforward as in the univariate case. To partially answer this question, we provide a sufficient condition in Theorem 1.2.1 to ensure the identifiability of the intensity functions.

Theorem 1.2.1 (Identifiability)**.** *For all $i \in \{1, \dots, d\}$, let $\theta_i \in \Theta$ be a parameter for λ^i.*

Let us assume that a.s. for every $(i, j) \in \{1, \dots, d\}^2$, $i \neq j$, there exist an event time τ from process N_j, and an event time $\tau_+ > \tau$ from process N_i, such that:

1. *$\lambda_{\theta_i}^i(\tau^-) > 0$;*

2. there are only events of process N^j in the interval $[\tau, \tau_+)$.

Then,

$$\forall i \in \{1, \dots, d\},\ \lambda^i_{\theta_i}(t) = \lambda^i_{\theta'_i}(t)\ \text{a.e.} \qquad \Longleftrightarrow \qquad \forall i \in \{1, \dots, d\},\ \theta_i = \theta'_i\,.$$

Overall, this condition allows to avoid pathological situations either when some process N_i is not observed (strong external inhibition effects), or when there is a cyclical observation of points (strong internal inhibition effects).

Finally, we present a solution for estimating the interaction network of a multivariate Hawkes process. As it is common in real-data contexts, it is often unrealistic to assume that all functions h_{ij} are non-null, meaning that all processes interact with each other. For this, we propose three different data-driven techniques based on a thresholding approach inspired by principal component analysis, and based on the construction of confidence intervals (Student and empirical). Our results on synthetic data show a significant improvement with respect to other state-of-the-art methods (Lemonnier et al. 2014; Bacry et al. 2020). In particular, when the interaction matrix $(\alpha_{ij})_{i,j=1:d}$ contains null entries, we implement our post-hoc procedures to obtain an estimation of the connectivity network before re-estimating all non-null entries on a reduced model. This provides the best estimations, in the sense of the ℓ_2 relative error, among the considered models, as confirmed by our results on real-data concerning the analysis of neuronal activity.

1.3 Something is amiss: spectral methods for imperfect data

1.3.1 Context and related works

In the second part of the thesis, we turn the spotlight to a common issue in statistics: establishing efficient estimation procedures when the observed data is noised in some sense.

In the context of point processes, a natural approach to define noised models is through the operations of thinning, superposition and jittering.

1. The **superposition** of two point processes X and Y, usually noted $X + Y$ is the point process defined by the ordered union of event times $(T^X_k)_k$ and $(T^Y_k)_k$.

2. The **thinning** of a point process X corresponds to erasing some event times T^X_k through a random rule. When point are erased independently from each other with a common probability $1 - p$, we talk about a p-thinning of process X.

3. The **jittering** of a point process consists in the random displacement of points.

The main interest of studying such alterations is to account for various measurement errors in real-world data. The most commonly studied in the literature is jittering which tends to appear when spatial imprecisions add a layer of uncertainty. Literature tends to focus on Poisson processes (Antoniadis et al. 2006; Hohage et al. 2016; Bonnet et al. 2022b), with Hawkes processes being studied in Trouleau et al. (2019) and Deutsch et al. (2020).

Thinning is often used to represent missing data but has been scarcely studied in the literature. Mei et al. (2019) propose a method to complete a sequence of times that have missing points via a Long Short-Term Memory neural network model based on the Hawkes process.

The study of superposition allows to account for observations that are contaminated by points originated from an external process, but are indistinguishable from those of the original process. To our knowledge, the only work that studies this kind of noise is Lund et al. (2000), which in fact tackles a general point process that is potentially noised by the three aforementioned

mechanisms. Their work establishes an approach by maximising the conditional log-likelihood dependent on the original process with an application to Hawkes processes. In practice and for numerical reasons, they propose a local maximiser of this quantity in a small set of parameters when having access to an estimation of the unnoised process.

Other imperfect data models have been presented in Linderman et al. (2014) where the authors favour a Bayesian approach through latent variables to study processes with missing marks, unobserved time intervals or unobserved subprocesses. The last work we mention here is Cheysson et al. (2022), where the exact locations of the points of a Hawkes process are not known. Instead, the only information available is the number of points that fall inside equally-sized intervals. They leverage then the spectral theory of point processes and time series in order to provide an estimation procedure that does not necessitate the precise location of event times.

Our contribution is to make use of the spectral theory of point processes, centered around the Bartlett spectrum (Bartlett 1964) and the Whittle estimator (Whittle 1952), in order to propose estimators for Hawkes processes altered by superposition (Chapter 5) and thinning (Chapter 6). Another contribution is to leverage the obtained inference method to incorporate a subsampling paradigm through thinning for small observation windows. The existing spectral theory for Hawkes processes is limited to exciting interactions, so we restrict our models to this framework.

1.3.2 Spectral analysis for the inference of noisy Hawkes processes

Chapter 5 is dedicated to the study of the superposition of a Hawkes process H with exciting kernel functions and an independent homogeneous Poisson process P with intensity $\lambda_0 > 0$. We consider their superposition, that we note $N = H + P$, and we assume that we have no knowledge on the origin of each event time T_k^N (either Hawkes or Poisson). The intensity function of this resulting process is intractable and any method based on it is therefore inaccessible.

As mentioned previously, we take inspiration from Cheysson et al. (2022) in order to leverage the spectral approach of point processes to circumvent this issue. The Bartlett spectrum Γ^N of N is a measure defined by the means of the Fourier transform of its reduced covariance measure. When this measure is absolutely continuous, its density is noted f^N and often called the spectral density of the process. We can establish an estimation procedure of a parametric model for f^N (with parameter $\theta \in \Theta$) through the periodogram function I^T. For an observation $(T_k^N)_{k\geq 1}$ in $[0, T]$, the periodogram is defined, for any $\omega \in \mathbb{R}$, as:

$$I^T(\omega) = \frac{1}{T} \sum_{k=1}^{N(T)} \sum_{l=1}^{N(T)} \mathrm{e}^{-2\pi \mathrm{i}\omega(T_k^N - T_l^N)} \,. \tag{1.9}$$

Asymptotically, this quantity follows an exponential distribution with mean $f^N(\omega)$ and so it is possible to define the maximum spectral log-likelihood estimator $\hat{\theta}$ as:

$$\hat{\theta} \in \arg\max_{\theta\in\Theta} - \frac{1}{T} \sum_{k=1}^{M} \left(\log\left(f_\theta^N(\omega_k)\right) + \frac{I^T(\omega_k)}{f_\theta^N(\omega_k)} \right), \tag{1.10}$$

where $\omega_k = k/T$ for $k \in \{1, \dots, M\}$. What makes this quantity so useful in our context is that it can be exactly computed without any information on the source of event times $(T_k^N)_k$.

Our first result establishes the expression of the spectral density function for the superposition of two independent point processes as exhibited in Proposition 1.3.1.

Proposition 1.3.1. *Let X and Y be two independent and stationary point processes, admitting respective spectral densities f^X and f^Y.*

Then $N = X + Y$ admits a spectral density function f^N and

$$f^N = f^X + f^Y \,. \tag{1.11}$$

As the spectral density of a Hawkes process is explicit (Daley et al. 2003, Example 8.2(e)), we obtain the general form of the spectral density of our noisy Hawkes process, which in turn makes it possible to establish an estimation procedure by maximising the spectral log-likelihood.

Our main contribution in the univariate setting is to establish conditions for the identifiability of f^N for different parametrisations of the kernel function h. The first result is related to the classical exponential kernel. We show that the model is identifiable if and only if one of the four parameters (three for the Hawkes process and one for the Poisson process) is fixed, even though the distribution of N is uniquely determined without this constraint. Furthermore, we prove that the noisy Hawkes process with a uniform kernel function defines an identifiable spectral model.

In the multivariate case, all tools from spectral theory exist and results are easily extended concerning the estimation procedure. Our contributions specifically concern the bivariate exponential Hawkes process noised by a bivariate Poisson process with common parameter λ_0. We present sufficient conditions on the interaction matrix $(\alpha_{ij})_{1\leq i,j\leq 2}$ that characterise either the identifiability or non-identifiability of the corresponding statistical model.

Our work is concluded by a numerical study in both the univariate and bivariate cases when identifiability conditions are verified. We show in both cases that our estimators perform particularly well, especially for processes with strong excitation effects and as the observation horizon T increases. Our final contribution is to propose an ad-hoc procedure in the bivariate case that allows to determine the support of the interaction matrix by considering the empirical 5% quantiles of a sample of estimations, which in turn enhances the performance of our methods.

1.3.3 A numerical exploration of thinned Hawkes processes through spectral theory

Finally, Chapter 6 turns to the analysis of thinning for point processes. Let H_p be a process issued from thinning a univariate Hawkes process with a probability $1-p$ of erasing each point T_k, for $k \in \mathbb{Z}$. As previously mentioned, any method based on the conditional intensity function is unavailable, so we turn to spectral theory.

All the tools presented in the previous subsection can still be implemented here, in particular the expression of the periodogram (Equation (1.9)) remains unchanged. As a consequence, it can still be exactly computed even though we do not observe all points of the original process. In order to implement an inference procedure as defined by Equation (1.10), we need to obtain an expression of the p-thinned process H_p. This is presented in Proposition 1.3.2.

Proposition 1.3.2. *Let H be a stationary point process admitting a spectral density function f^H and let $m_1 = \mathbb{E}[\lambda(0)]$ be its average intensity. For any $p \in (0,1]$, let H_p be a p-thinning of H.*

Then, H_p admits a spectral density function, denoted f^{H_p}, such that for any $\omega \in \mathbb{R}$:

$$f^{H_p}(\omega) = p^2 f^H(\omega) + p(1-p)m_1 \,. \tag{1.12}$$

This result is obtained by leveraging the spectral theory of marked point processes as presented in Brémaud et al. (2005).

We focus our study on the exponential kernel function and we prove that the derived statistical model is identifiable if and only if one parameter is fixed, similar to the aforementioned superposition scenario. We present numerical results under this condition by fixing the value

of the thinning parameter p. In particular, we show that our estimation method is robust with respect to the level of thinning, with remarkable results when enough event times are available for inference.

Our last contribution in this work is to make use of the thinning operation as a subsampling procedure when a single observation of H is available in a small window of time. The use of subsampling for point processes has been done previously in the literature (Møller et al. 2003; Cronie et al. 2024) in order to improve estimation procedures. In our case, we focus on enhancing the quality of our spectral estimations by combining an ℓ_2 penalisation with an averaged estimator obtained by thinning the unique observation of H. We show that our proposed procedure provides in general the best estimations, in the sense of the relative ℓ_2 error, when compared to another subsampling procedure and displays a substantial improvement by reducing the high bias observed for the non-penalised estimator.

1.4 Outline of the manuscript

Chapter 2 proposes a succinct presentation of point processes theory, from a random measure approach, and the formalism of Hawkes processes.

The first part of this thesis concerns the study of the inhibition.

- Chapter 3 presents a parametric estimation method for univariate Hawkes processes that accounts for both self-exciting or self-inhibiting interactions. This is a joint work with Anna Bonnet and Maxime Sangnier, published in *Statistics and Probability Letters* (Bonnet et al. 2021).

  Code is freely available at `https://github.com/migmtz/hawkes-inhibition-expon`

- Chapter 4 presents a parametric estimation method for multivariate exponential Hawkes processes with exciting and inhibiting interactions along with a model selection procedure. This is a joint work with Anna Bonnet and Maxime Sangnier, published in *Statistics and Computing* (Bonnet et al. 2023).

  Code is freely available at `https://github.com/migmtz/multivariate-hawkes-inhibition`

The second part consists in the study of imperfect observations of a Hawkes process realisation, similar to missing data formulations.

- Chapter 5 presents a parametric estimation method for exciting Hawkes processes whose event times are noised by those of a homogenenous point process, leveraging point process spectral theory. This is a joint work with Anna Bonnet, Felix Cheysson and Maxime Sangnier and has been submitted for publication (Bonnet et al. 2024).

  Code is available at `https://github.com/migmtz/noisy-hawkes-process`

- Chapter 6 presents the analysis of an inference method for thinned univariate Hawkes processes through spectral theory. Additionally, it presents an improvement of ℓ_2 penalisation through thinning subsampling for the spectral estimator.

  This chapter is an ongoing joint work with Felix Cheysson.

2 An introduction to Hawkes processes study

Outline of the current chapter

2.1 Introduction

Point processes are a fundamental concept in stochastic modelling, providing a framework for analysing events that occur randomly over time. Traditionally, a temporal point process is presented through the concept of an ordered collection of random variables in the real line, each one representing the occurrence of an event. A standard approach to study point processes is by defining random measures, which are random variables with realisations on the space of boundedly finite measures. In particular, this framework gives access to many results from set theory and measure theory as largely shown in the literature (Cox et al. 1980; Baddeley 2006; Daley et al. 2008; Baccelli et al. 2020).

In this chapter, we present a succinct introduction to these concepts in order to present the submodel of past-dependent point processes: the Hawkes model (Hawkes 1971). Our goal here is twofold: the first one is to introduce numerous notations and definitions that are used throughout this manuscript, serving as a general reference for further chapters. The second is to present a short review of the original self-exciting Hawkes processes and some numerical procedures from the literature.

We introduce the concept of temporal point processes through random measure theory in Section 2.2 with a particular interest on the concept of conditional intensity functions. With this background, we formally introduce the univariate self-exciting Hawkes processes in Section 2.3 with the two classical extensions to multivariate and non-linear settings, the latter being often used to model inhibition effects. We follow up with a general presentation of some classic inference methods in Section 2.4 and we conclude with Section 2.5 by presenting two methods of simulation that are commonly used in the context of Hawkes processes.

2.2 Random measures and point processes

2.2.1 Random measures

Let $\mathcal{X} \subseteq \mathbb{R}$ denote either the real half-line $\mathbb{R}_{\geq 0}$ or the real line $\mathbb{R}$ and $\mathcal{B}_{\mathcal{X}}$ be the Borel σ-algebra of $\mathcal{X}$.

Definition 2.2.1. A measure ν on $\mathcal{X}$ is said to be *boundedly finite* if for any bounded $B \in \mathcal{B}_{\mathcal{X}}$, $\nu(B) < +\infty$ and we define $\mathcal{M}_{\mathcal{X}}$ as the set of boundedly finite measures on $\mathcal{X}$.

For $\mathcal{M}_{\mathcal{X}}$, we may obtain Borel σ-algebras generated by the mappings $\nu \mapsto \nu(B)$, for all $B \in \mathcal{B}_{\mathcal{X}}$, that we note $\mathcal{B}(\mathcal{M}_{\mathcal{X}})$.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

Definition 2.2.2 (Random measure)**.** A *random measure* ξ on $\mathcal{X}$ is a measurable mapping from Ω into $\mathcal{M}_{\mathcal{X}}$. The distribution of a random measure ξ is the probability measure induced by the probability measure $\mathbb{P}$.

An equivalent definition of a random measure is obtained through the following result.

Proposition 2.2.1. *(Baccelli et al. 2020, Proposition 1.1.7)*

For a mapping $\xi : (\Omega, \mathcal{F}) \to (\mathcal{M}_{\mathcal{X}}, \mathcal{B}(\mathcal{M}_{\mathcal{X}}))$, we define for any $B \in \mathcal{B}_{\mathcal{X}}$ the mapping $\xi(B) : (\Omega, \mathcal{F}) \to (\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ as:

$$(\xi(B))(\omega) = (\xi(\omega))(B)\,.$$

Then, ξ is a random measure if and only if, for all $B \in \mathcal{B}_{\mathcal{X}}$, $\xi(B)$ is measurable.

With this result, it means that we can work with a random measure ξ in two ways:

- For any $\omega \in \Omega$, a realisation $\xi(\omega)$ is a measure on $\mathcal{X}$.
- For any $B \in \mathcal{B}_{\mathcal{X}}$, $\xi(B)$ is a random variable on $\mathcal{X}$. Furthermore, $\{\xi(B)\}_{B \in \mathcal{B}_{\mathcal{X}}}$ is a stochastic process.

In practice, as with other random variables we tend to omit the term ω when working with random measures and using ξ either as a measure or random variable is to be understood by the context. Another direct result from Proposition 2.2.1 is that, for any finite sequence $(B_k)_{k=1:K}$ of Borel sets, we can define the random vector:

$$(\xi(B_1), \ldots, \xi(B_K))\,.$$

It turns out that a characterisation of any random measure is given by the distribution of all of these vectors.

Proposition 2.2.2. *(Daley et al. 2008, Corollary 9.2.IV)*
Let ξ be a random measure on $\mathcal{X}$. The finite-dimensional distributions of ξ are the probability distribution of the vectors $(\xi(B_1), \dots, \xi(B_K))$ for all integers $K \geq 1$ and all sequences of Borel sets $(B_k)_{k=1:K}$.
The distribution of ξ is fully given by its finite-dimensional distributions.

With this result, a way of defining any random measure is by defining its finite-dimensional distributions. However, not any set of finite-dimensional distributions corresponds to those of a random measure, but a set of necessary and sufficient conditions is proposed in Daley et al. (2008, Conditions 9.2.V-VI), namely Kolmogorov extension theorem conditions and additivity and continuity conditions. Finally, let us introduce the notion of stationarity. For any $B \in \mathcal{B}_{\mathcal{X}}$, for any $x \in \mathcal{X}$, we define $B + x$ as the set:

$$B + x = \{b + x \mid b \in B\}.$$

Definition 2.2.3. A random measure ξ is said to be *stationary* if for any $B \in \mathcal{B}_{\mathcal{X}}$ and for any $x \in \mathcal{X}$, $N(B)$ and $N(B + x)$ follow the same distribution.

2.2.2 Point processes

Let us now turn to the study of point processes. For this, we need the definition of a counting measure.

Definition 2.2.4. A measure ν on $\mathcal{X}$ is a counting measure if for any $B \in \mathcal{B}_{\mathcal{X}}$, $\nu(B) \in \mathbb{N} \cup \{+\infty\}$. We note $\mathcal{N}_{\mathcal{X}}$ the set of all boundedly finite counting measures on $\mathcal{X}$.

As done previously, we can equip the space $\mathcal{N}_{\mathcal{X}}$ with a Borel σ-algebra $\mathcal{B}(\mathcal{N}_{\mathcal{X}})$ generated by the mappings $\nu \mapsto \nu(B)$, for all $B \in \mathcal{B}_{\mathcal{X}}$.

Definition 2.2.5 (Point processes and simple point processes)**.** A *point process* N on $\mathcal{X}$ is a measurable mapping from a $(\Omega, \mathcal{F}, \mathbb{P})$ into $(\mathcal{N}_{\mathcal{X}}, \mathcal{B}(\mathcal{N}_{\mathcal{X}}))$.

Intuitively, a point process N is a random measure (Baccelli et al. 2020, Corollary 1.6.4.) such that, for any $\omega \in \Omega$, the realisation $N(\omega)$ is a counting measure and so in particular all previous results from random measures hold for a point process. For instance, for any $B \in \mathcal{B}_{\mathcal{X}}$, $N(B)$ will denote the measure of B by N, which allows us to introduce the definition of a simple point process.

Definition 2.2.6. A point process N is *simple* if for any $x \in \mathcal{X}$, $N(\{x\}) \in \{0, 1\}$ a.s.

It is very common in the literature to work uniquely with simple point processes, but before further discussing this, let us exhibit the following result from the theory of point processes.

Theorem 2.2.1. *(Last et al. 2017, Corollary 6.5)*
Let N be a point process on $\mathcal{X}$ and let δ_x be the Dirac measure on x, for any $x \in \mathcal{X}$. There exists a random variable $\kappa \in \mathbb{N} \cup \{+\infty\}$ and a sequence of random variables $(T_1, T_2, \dots)$ such that:

$$N = \sum_{k=1}^{\kappa} \delta_{T_k}, \quad a.s.$$

We refer to the variables T_k as the event times *or* points *of process N.*

This theorem reunites the definition of a point process as a random measure with other formalisms of point processes as random points observed in a space $\mathcal{X}$. A simple point process is then a point process without coinciding points, in other words, for any $i \neq j$, $T_i \neq T_j$ almost surely. From now on, we will assume that all point processes are simple.

Let us now turn to some particularities of working in a temporal setting. As $\mathcal{X}$ is a well-ordered set, it is practical to order the event times of N. The indexation of the sequence of event times slightly differs whether we work on $\mathbb{R}_{\geq 0}$ or $\mathbb{R}$. Unlike Theorem 2.2.1 may suggest, it is a common adopted convention, when working on the entire real line, that the event times are indexed by $\mathbb{Z}$ such that

$$\ldots < T_{-2} < T_{-1} < T_0 \leq 0 < T_1 < T_2 < \ldots .$$

This way, positive indices will always denote points *after* the origin of time $t = 0$. Another usual convention is to denote $N(t) = N([0, t])$ the number of points in the interval $[0, t]$ especially used in inference contexts. This allows also to see $(N(t))_{t\in\mathcal{X}}$ as a right-continuous stochastic process.

2.2.3 Integrals and moment measures

We can define the stochastic integral of a measurable function $f : \mathcal{X} \to \mathbb{R}_{\geq 0}$ against a point process N with event times $(T_k)_{k\in\mathbb{Z}}$ as:

$$\int_{\mathcal{X}} f(t)\, N(\mathrm{d}t) = \sum_{k\in\mathbb{Z}} f(T_k)\,,$$

which is similar to the Lebesgue-Stieljes integral. We can see that the integral defined this way is a random variable and by taking $f = \mathbb{1}_B$ for any $B \in \mathcal{B}_{\mathcal{X}}$, we can retrieve the measure of B:

$$\int_{\mathcal{X}} \mathbb{1}_B(t)\, N(\mathrm{d}t) = \sum_{k\in\mathbb{Z}} \mathbb{1}_{T_k\in B} = N(B)\,.$$

Let us now define the moment measures of N.

Definition 2.2.7. Let N be a point process. Whenever it exists, we note M_k the *k-th moment measure* of N defined, for any $(B_1, \ldots, B_k) \in \mathcal{B}_{\mathcal{X}}^k$, as:

$$M_k(B_1, \ldots, B_k) = \mathbb{E}[N(B_1) \ldots N(B_k)]\,.$$

In particular, the first moment measure M_1 is also known as the *intensity measure* of N, also denoted Λ.

With the intensity measure, we are able to introduce the most classical point process, the Poisson process.

Definition 2.2.8. Let Λ be a measure on $\mathcal{X}$ absolutely continuous with respect to the Lebesgue measure ℓ and let $\lambda : \mathcal{X} \to \mathbb{R}_{\geq 0}$ be its Radon-Nikodym derivative.

A point process N is a *Poisson process* with intensity function λ if it verifies the following conditions:

1. For any $B \in \mathcal{B}_{\mathcal{X}}$, $N(B)$ follows a Poisson distribution with parameter $\Lambda(B) = \int_B \lambda(t)\,\mathrm{d}t$.
2. For any integer K and any set of disjoint Borel sets $(B_k)_{k=1:K}$, the variables $(N(B_k))_{k=1:K}$ are independent.

If λ is a constant function, then N is called a *homogeneous* Poisson process.

The Poisson process is sometimes referred to as a process with independent increments because of the second property. In particular, when the intensity function λ is constant, we can further characterise the process by its inter-arrival times.

Theorem 2.2.2. *(Last et al. 2017, Theorem 7.2)*
N is a homogeneous Poisson process with intensity $\lambda > 0$ if and only if the inter-arrival times $(T_{k+1} - T_k)_{k\in\mathbb{Z}}$ are i.i.d, all following an exponential distribution with parameter λ.

2.2.4 The conditional intensity function

In the previous subsection, we defined the Poisson process through its intensity function, which allows to account for different time-dependent dynamics. The concept that will allow us to properly introduce the Hawkes process is the conditional intensity function. For this, let us note $\mathcal{H}_t$ the *history* of a point process N up to the instant $t \in \mathcal{X}$. $\mathcal{H}_t$ is the σ-algebra generated by the event times T_k:

$$\mathcal{H}_t = \sigma(\{T_k \mid k \in \mathbb{Z}, T_k \leq t\}) .$$

Definition 2.2.9. Let N be a simple point process on $\mathcal{X}$, and $(\mathcal{H}_t)_t$ the histories of N. The *conditional intensity function* λ of N is defined, for any $t \in \mathcal{X}$, as:

$$\lambda(t \mid \mathcal{H}_t) = \lim_{h\to 0} \frac{\mathbb{E}[N([t, t+h]) \mid \mathcal{H}_t]}{h} . \tag{2.1}$$

Let us remark that by Proposition 7.2.IV in Daley et al. (2003), a point process is entirely determined by its conditional intensity function. In general, the conditional intensity function λ is a random variable for each $t \in \mathcal{X}$ as it is expressed through a conditional expectation. However, if λ is a deterministic function, then it coincides with the intensity function of a Poisson process as given in Definition 2.2.8, which justifies keeping the same notation λ.

It is a common practice to omit the term $\mathcal{H}_t$ in Equation (2.1) as a conditional intensity is usually understood as having access to the entire past history (unless it is to avoid ambiguity). Intuitively λ quantifies the instantaneous probability of observing a point on an interval of infinitesimal size h. As N is a simple point process, the expectation in Equation (2.2.8) corresponds to $\mathbb{P}(N([t, t+h])) = 1$ for h small enough. Another common interpretation (see Hawkes (1971), for an example) of the conditional intensity function for simple point processes is through the following properties:

$$\begin{cases} \mathbb{P}(N([t, t+h]) = 1) = \lambda(t)h + o(h) \\ \mathbb{P}(N([t, t+h]) = 0) = 1 - \lambda(t)h + o(h) \\ \mathbb{P}(N([t, t+h]) \geq 2) = o(h) . \end{cases}$$

A very formal presentation of the conditional intensity function can be found in Daley et al. (2003, Chapter 7) through the concept of Janossy densities. In a nutshell, given a number K of observed points in a set B, the Janossy densities describe the probability distributions of points $(T_k)_{k=1:K}$ inside B.

Definition 2.2.10. Let N be a simple point process and λ its conditional intensity function. The *compensator* Λ of N is the random measure defined, for any $B \in \mathcal{B}_\mathcal{X}$ as:

$$\Lambda(B) = \int_B \lambda(t)\, \mathrm{d}t .$$

In particular, the expectation of Λ is the first-order moment measure of N. If N is stationary, then this measure is a constant multiple of the Lebesgue measure ℓ.

Again, the concept of compensator coincides with the intensity measure in Definition 2.2.8 whenever λ is deterministic. Similarly to the notation of point processes, we note $\Lambda(t) = \Lambda([0,t])$ for any $t \in \mathcal{X}$. The compensator is a highly studied quantity for point processes defined through conditional intensity functions. An important result concerning Λ is the time change theorem:

Theorem 2.2.3. *(Daley et al. 2003, Theorem 7.4.IV)*
Let $(X_k)_{k\in\mathbb{Z}}$ be an a.s. increasing sequence of random variables in $\mathcal{X}$ and $\mathcal{F}_t = \sigma(\{X_k \mid k \in \mathbb{Z}, X_k \leq t\})$, for all $t \in \mathbb{R}$. Let $\bar{\lambda}(\cdot \mid \mathcal{F}_\cdot)$ be a non-negative function such that, for all $t \in \mathbb{R}$, $\bar{\lambda}(t \mid \mathcal{F}_t)$ is $\mathcal{F}_t$-measurable and let $\bar{\Lambda}$ be defined, for any $t \in \mathbb{R}$, as:

$$\bar{\Lambda}(t) = \int_{-\infty}^{t} \bar{\lambda}(u \mid \mathcal{F}_u)\,\mathrm{d}u\,.$$

Then, the process $\sum_{k\in\mathbb{Z}} \delta_{X_k}$ is a point process with conditional intensity function $\bar{\lambda}$ if and only if $\sum_{k\in\mathbb{Z}} \delta_{\bar{\Lambda}(X_k)}$ is a homogeneous Poisson process with unit intensity.

It is then possible to rescale the event times of any observed point process N in order to obtain a realisation of a homogeneous Poisson process. This is a very useful result in order to establish goodness-of-fit procedures for point processes when an estimation $\hat{\Lambda}$ of Λ is available. For example, by transforming the observed event times $(T_k)_k$ with $\hat{\Lambda}$, we can then test whether the inter-arrival times $(\hat{\Lambda}(T_{k+1}) - \hat{\Lambda}(T_k))_k$ are exponentially distributed (see Theorem 2.2.2).

Lastly, we conclude this section by introducing the expression of the likelihood.

Proposition 2.2.3. *(Daley et al. 2003, Theorem 7.2.III)*
Let N be a point process with conditional intensity function λ. Let $(T_k)_{k=1:N(T)}$ be the realisation of N in the interval $[0,T]$.
Then the likelihood L_T of N reads:

$$L_T = \left(\prod_{k=1}^{N(T)} \lambda(T_k^-) \right) \mathrm{e}^{-\Lambda(T)}\,, \tag{2.2}$$

where $\lambda(T_k^-) = \lim_{t\to T_k^-}(\lambda(t))$.
The log-likelihood of N is:

$$\ell_T = \sum_{k=1}^{N(T)} \log(\lambda(T_k^-)) - \Lambda(T)\,. \tag{2.3}$$

The likelihood can be introduced by means of the Janossy densities. Let us note that, in the case of point processes on the entire real line $\mathbb{R}$, it is necessary to know all the points $T_k \leq T$ in order to compute the likelihood, which is usually unavailable. In practice, it is often assumed that no points occured in $(-\infty, 0)$ and so an adapted version of λ is used in Equation 2.2 (see Ogata (1978)).

2.3 Hawkes process

2.3.1 The self-exciting Hawkes process

The Hawkes process, as introduced in Hawkes (1971), is defined as a point process in $\mathbb{R}$ as follows:

Definition 2.3.1. Let $\mu > 0$ and $h : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be a measurable function such that

$$\|h\|_1 = \int_0^{+\infty} |h(t)| \, \mathrm{d}t < 1 \, .$$

A *Hawkes process* N is a point process on $\mathbb{R}$ defined by the conditional intensity function:

$$\forall t \in \mathbb{R}, \qquad \lambda(t) = \mu + \int_{-\infty}^{t} h(t-s) \, N(\mathrm{d}s) \, . \tag{2.4}$$

μ is known as the *baseline intensity* and h as the *interaction* or *kernel function*.

This process was initially referred to as a self-exciting point process due to the integral in Equation (2.4) that represents the positive contribution of all past points up to time t to the intensity function λ. Let us recall that:

$$\int_{-\infty}^{t} h(t-s) \, N(\mathrm{d}s) = \sum_{T_k \leq t} h(t - T_k) \, ,$$

and so each event time T_k increases the baseline intensity by an additive factor of $h(t - T_k)$, hence the self-excitation effect.

The condition

$$\int_0^{+\infty} h(t) \, \mathrm{d}t < 1 \, ,$$

for the interaction function ensures that N is boundedly finite a.s. and so that N is a point process. Another result of this condition is that $\lim_{t \to +\infty} h(t) = 0$, meaning that the effect of any point T_k dissipates the further they are in the past.

Another implication of this condition is that a Hawkes process defined in $\mathbb{R}$ is a stationary point process (Hawkes et al. 1974, Lemma 1) which allows to establish the following result.

Proposition 2.3.1. *(Hawkes 1971, Equation (9)) Let N be a stationary Hawkes process. Then the average intensity of N reads:*

$$\forall t \in \mathbb{R}, \qquad \mathbb{E}[\lambda(t)] = \frac{\mu}{1 - \|h\|_1} \, .$$

The stationarity of process N is verified for processes only in $\mathbb{R}$ and so this result does not hold for a process in $\mathbb{R}_{\geq 0}$. On the one hand, it is often more practical for theoretical reasons to work in $\mathbb{R}$, as stationarity tends to facilitate establishing certain results (see Hawkes (1971) for example in the context of spectral theory). On the other hand, working with a process in $\mathbb{R}_{\geq 0}$ is often preferred for application purposes, as it is unrealistic to suppose that the practitioner has access to an infinite number of points in the past before $t = 0$. To define such a process, it suffices to replace the lower bound in the integral of Equation (2.4) by 0.

2.3.2 The multivariate Hawkes process

The multivariate version of a self-exciting Hawkes process was introduced in Hawkes (1971) as follows.

Definition 2.3.2. A multivariate Hawkes process $N = (N_1, \dots, N_d)$ of dimension d is defined by d point processes $(N_i)_{i=1:d}$. For each integer i, we note $(T_k^i)_k$ the event times of process N_i and its conditional intensity function λ^i reads:

$$\lambda^i(t) = \mu_i + \sum_{j=1}^{d} \int_{-\infty}^{t} h_{ij}(t-s)\, N_j(\mathrm{d}s) = \mu_i + \sum_{j=1}^{d} \sum_{T_k^j \le t} h_{ij}(t - T_k^j)\,.$$

For any integers i, j, $\mu_i > 0$ is the baseline intensity of N_i and $h_{ij} \colon \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ is a measurable function.

For a multivariate Hawkes process, each intensity function λ^i is influenced by every point T_k^j of process N_j, for all integers j. Function h_{ij} represents the influence of a point from N_j to process N_i. We define the matrix $S = (\|h_{ij}\|_1)_{1 \ge i,j \ge d}$ and a condition for all N^i to be proper point processes is to control the spectral radius $\rho(S) < 1$ (Bacry et al. 2015). This condition ensures again the stationarity of all processes when defined in $\mathbb{R}$.

When $d = 1$, we retrieve the original expression of a Hawkes process from Definition 2.3.1 and so we talk about a univariate Hawkes process. Defining a multivariate Hawkes process as a tuple is the more common formulation for any kind of point process, but it is sometimes practical to consider process N as the superposition of all subprocesses N_i. N is then characterised by the ordered union of all event times $(T_k^i)_{k \in \mathbb{Z}}$ for all integers i, forming a single point process in $\mathbb{R}$ with event times $(T_{(k)})$ and conditional intensity function:

$$\lambda(t) = \sum_{i=1}^{d} \lambda^i(t)\,.$$

This perspective allows to extend all results from general point process theory to multivariate processes, including inference methods.

2.3.3 Non-linear Hawkes processes

A more general formulation of a past-dependent point process inspired by the Hawkes process was proposed in Brémaud et al. 1996 and known as the non-linear Hawkes process.

Definition 2.3.3. The univariate *non-linear Hawkes process* N on $\mathbb{R}$ is defined by the conditional intensity function:

$$\lambda(t) = \Phi\left(\mu + \int_{-\infty}^{t} h(t-s)\, N(ds)\right), \tag{2.5}$$

where $\mu > 0$ and $\Phi : \mathbb{R} \to \mathbb{R}_{\ge 0}$ and $h \colon \mathbb{R}_{\ge 0} \to \mathbb{R}$ are two measurable functions. Φ is known as the activation function and h as the *interaction* or *kernel function*.

The usual formulation of the self-exciting Hawkes process in Definition 2.3.1 can be retrieved by taking the identity function $\Phi \colon x \mapsto x$ and by adding a positivity constraint on h. Because of this, it is often referred to as the *linear* Hawkes process.

By allowing h to take negative values in Equation (2.5), the non-linear Hawkes process can model an *inhibiting* effect between points where each point decreases the chances of other points occuring, in opposition to the excitation effect.

To guarantee the existence of such a process, either Φ needs to be upper-bounded or Φ to be L-lipschitz, for $L > 0$, such that:

$$L\|h^+\|_1 < 1\,,$$

where $(\cdot)^+ \colon x \mapsto \max(0, x)$ denotes the positive part function.

A multivariate version of this process can be defined as for the classical Hawkes process in Definition 2.3.2. In this setting, existence can be retrieved either by upper-boundness of Φ or by imposing that the spectral radius of matrix $S^+ = (L\|h^+_{ij}\|_1)_{1\geq i,j\geq d}$ is stricly smaller than 1 (Sulem et al. 2024, Lemma 2.1). For instance, if Φ is the positive part function, this condition can be retrieved by upper-bounding the intensity function by the intensity of the linear process N^+ with interaction functions $(h^+_{ij})_{i,j}$, for all integers i, j (Deutsch et al. 2023, Theorem 2).

This function, often called ReLU activation, is the most common choice in the literature for the non-linear function (Lemonnier et al. 2014; Lu et al. 2018; Costa et al. 2020), with other examples include clipped exponential (Chornoboy et al. 1988; Carstensen et al. 2010), sigmoid (Menon et al. 2018) and softplus functions (Mei et al. 2017). This model of inhibiting interactions for the Hawkes process is often referred to as *additive* inhibition and other works in the literature have proposed alternative versions such as the self-limiting Hawkes process with a multiplicative exponential term dependent on the number of points in a moving interval (Olinde et al. 2020), and the mean-field Hawkes process with multiplicative inhibition (Duval et al. 2022).

2.4 Inference for Hawkes processes

In this section we will review some classical methods for the estimation of the baseline intensity μ and interaction function h of the linear Hawkes process N. The literature on inference procedures in the non-linear setting is scarcer and some will be mentioned in Section 2.4.3.

2.4.1 MLE for Hawkes process with exponential kernel

A common method is maximisation of the log-likelihood:

$$\ell_T = \sum_{k=1}^{N(T)} \log(\lambda(T_k^-)) - \Lambda(T)\,,$$

as shown in Ozaki (1979). A multivariate version of it is presented in Embrechts et al. (2011) and Guo et al. (2018) with expected mathematical properties including consistency and asymptotic normality (Simon et al. 2017), mainly developed in parametric settings.

We present here the implementation of the maximum likelihood estimation method for univariate self-exciting Hawkes processes. Let N be a Hawkes process in $\mathbb{R}_{\geq 0}$ and we assume that the function h is parametrised by an exponential distribution such that:

$$h(t) = \alpha \mathrm{e}^{-\beta t}\,, \quad \text{for } t \geq 0,$$

with $\alpha > 0$ and $\beta > 0$. The existence condition in Definition 2.3.1 reads:

$$\|h\|_1 = \frac{\alpha}{\beta} < 1\,,$$

and so we assume that $\alpha < \beta$. By defining the following parametric model:

$$\mathcal{P} = \left\{\lambda_\theta \mid \theta = (\mu, \alpha, \beta) \in \mathbb{R}^3_{\geq 0}, \alpha < \beta\right\}\,,$$

the log-likelihood (2.3) becomes a function of parameter θ:

$$\ell_T(\theta) = \sum_{k=1}^{N(T)} \log(\lambda_\theta(T_k^-)) - \Lambda_\theta(T)\,,$$

with Λ_θ the compensator of λ_θ.

The choice of the exponential kernel function for N implies that the intensity λ is Markovian in the sense that in each interval $[T_k, T_{k+1})$ the expression is solely dependent on the value of $\lambda(T_k)$. For any integer $k \geq 1$ and any $t \in [T_k, T_{k+1})$, the intensity reads:

$$\begin{aligned}\lambda(t) = \mu + \int_0^{+\infty} \alpha \mathrm{e}^{-\beta(t-s)}\, N(\mathrm{d}s) &= \mu + \sum_{j=1}^{k} \alpha \mathrm{e}^{-\beta(t-T_j)} \\ &= \mu + \mathrm{e}^{-\beta(t-T_k)} \sum_{j=1}^{k} \alpha \mathrm{e}^{-\beta(T_k-T_j)} \\ &= \mu + \mathrm{e}^{-\beta(t-T_k)} \left(\lambda(T_k) - \mu\right)\,. \end{aligned} \tag{2.6}$$

We can then easily integrate λ in $[T_k, T_{k+1})$:

$$\begin{aligned}\int_{T_k}^{T_{k+1}} \lambda(t)\,\mathrm{d}t &= \mu(T_{k+1} - T_k) + (\lambda(T_k) - \mu) \int_{T_k}^{T_{k+1}} \mathrm{e}^{-\beta(t-T_k)}\,\mathrm{d}t \\ &= \mu(T_{k+1} - T_k) + (\lambda(T_k) - \mu)\beta^{-1}(1 - \mathrm{e}^{-\beta(T_{k+1}-T_k)})\,.\end{aligned}$$

Then, for any interval $[0, T]$, the compensator Λ can be computed piecewise:

$$\begin{aligned}\Lambda(T) &= \int_0^T \lambda(t)\,\mathrm{d}t \\ &= \int_0^{T_1} \mu\,\mathrm{d}t + \sum_{k=1}^{N(T)-1} \int_{T_k}^{T_{k+1}} \lambda(t)\,\mathrm{d}t + \int_{T_{N(T)}}^{T} \lambda(t)\,\mathrm{d}t \\ &= \mu T + \beta^{-1} \left(\sum_{k=1}^{N(T)-1} (\lambda(T_k) - \mu)(1 - \mathrm{e}^{-\beta(T_{k+1}-T_k)}) + (\lambda(T_{N(T)}) - \mu)(1 - \mathrm{e}^{-\beta(T-T_k)}) \right)\,.\end{aligned}$$

The consequence of this expression is that the terms $\lambda(T_k)$ (and $\lambda(T_k^-)$) can be computed recursively by Equation 2.6 and so Λ has a computational complexity of $O(N(T))$. The log-likelihood also has a complexity of $O(N(T))$ and can be efficiently computed as described in Algorithm 1:

In practice, we obtain an estimation $\hat{\theta}$ of a parameter $\theta = (\mu, \alpha, \beta)$ by maximising the log-likelihood.

Other methods implemented in parametric settings are the method of moments (Da Fonseca et al. 2013), by leveraging the spectral theory of point processes to maximise a spectral version of the log-likelihood (Adamopoulos 1976) or through an Expectation-Maximisation procedure to estimate the branching structure of the Hawkes process (Veen et al. 2008).

Algorithm 1: Computation of ℓ_T

Input Parameters μ, α, β, time horizon T and sequence of event times $(T_k)_{k=1:N(T)}$;
Initialization Initialize $\ell_T = \log(\mu) - \mu T$, $\lambda(T_1) = \mu + \alpha$;
for $k \geq 1$ **do**
  Compute $\lambda(T_{k+1}^-) = \mu + (\lambda(T_k) - \mu)\mathrm{e}^{-\beta(T_{k+1} - T_k)}$;
  Update $\ell_T = \ell_T + \log(\lambda(T_{k+1}^-)) - \beta^{-1}(\lambda(T_k) - \mu)(1 - \mathrm{e}^{-\beta(T_{k+1} - T_k)})$;
  Compute $\lambda(T_{k+1}) = \lambda(T_{k+1}^-) + \alpha$;
end
Update $\ell_T = \ell_T - \beta^{-1}(\lambda(T_k) - \mu)(1 - \mathrm{e}^{-\beta(T - T_k)})$;
return Log-likelihood ℓ_T

2.4.2 Least-squares minimisation

Inference procedures in non-parametric settings tend to be based on the minimisation of a least-squares contrast. Let us assume that we want to estimate an intensity function λ in an interval $[l, L]$ through a function $\hat{\lambda}$. We define the squared error of $\hat{\lambda}$ as:

$$\|\hat{\lambda} - \lambda\|_2^2 = \int_l^L (\hat{\lambda}(t) - \lambda(t))^2 \,\mathrm{d}t \tag{2.7}$$

$$= \int_l^L \hat{\lambda}^2(t) \,\mathrm{d}t - 2\int_l^L \hat{\lambda}(t)\lambda(t) \,\mathrm{d}t + \int_l^L \lambda^2(t) \,\mathrm{d}t\,. \tag{2.8}$$

In order to minimise this quantity with respect to $\hat{\lambda}$, we need to remove the dependency on λ as this quantity is unavailable in estimation contexts. We may remark that the last term can be completely ignored as it does not depend on $\hat{\lambda}$ but the middle term has to be treated differently.

As done in Reynaud-Bouret et al. (2014), we can leverage the fact that a valid approximation of $\lambda(t)\,\mathrm{d}t$ is $N(t)$. This is justified by the fact that locally $\mathbb{E}[N(t)] = \mathbb{E}[\Lambda(t)]$ and that λ is the Stieltjes-Lebesgue derivative of Λ. We can then minimise the following expression:

$$\int_l^L \hat{\lambda}^2(t) \,\mathrm{d}t - 2\int_l^L \hat{\lambda}(t) N(\mathrm{d}t)\,,$$

for an observation of process N.

Non-parametric approaches to minimise this quantity tend to approximate the interaction function h by histograms (Lemonnier et al. 2014; Reynaud-Bouret et al. 2014) or by using autoregressive models (Kirchner 2017).

Other non-parametric methods include solving Wiener-Hopf equations (Bacry et al. 2016), optimising a penalised log-likelihood through an EM algorithm (Lewis et al. 2011) and fitting second and third-order cumulants (Achab et al. 2016).

2.4.3 A Bayesian estimation approach

The Bayesian approach for the inference of Hawkes processes consists in estimating the posterior distribution for the parameters of the intensity function. The method we present here follows the procedure by Rasmussen (2013) which leverages the expression of the likelihood. We present an application for the exponential kernel function $h(\cdot) = \alpha\mathrm{e}^{-\beta\cdot}$.

Let $\theta = (\mu, \alpha, \beta) \sim \Pi$ be a random vector of parameters with density π and, conditionally on

θ, let N be a linear Hawkes process with intensity λ_θ parametrised by θ. We define the likelihood L_T (Equation (2.2)) of N, in a time window $[0,T]$ and for a history $\mathcal{H}_T$, as:

$$L_T(\theta) = \left(\prod_{k=1}^{N(T)} \lambda_\theta(T_k^-) \right) \mathrm{e}^{-\Lambda_\theta(T)} .$$

By the Bayes rule, the posterior distribution of θ conditional on the observation of N in $[0,T]$ admits a density $\pi(\cdot \mid \mathcal{H}_T)$ proportional to the prior π and the likelihood L_T:

$$\pi(\theta \mid \mathcal{H}_T) \propto \pi(\theta) L_T(\theta) .$$

An estimator $\hat{\theta} = \mathbb{E}[\theta \mid \mathcal{H}_T]$ can be approximated by sampling the posterior distribution of $\theta \mid \mathcal{H}_T$, which has the advantage of providing credibility intervals.

For this, we implement a Metropolis-Hastings algorithm with multivariate isotropic normal distribution as the proposition kernel. Let us illustrate the principle with the baseline intensity μ: for a current parameter μ_l, a candidate $\tilde{\mu}_l \mid \mu_l \sim \mathcal{N}(\mu_l, \sigma^2)$ is sampled and the acceptance ratio reads:

$$H_\mu = \frac{\pi(\tilde{\mu}_l, \alpha, \beta)}{\pi(\mu_l, \alpha, \beta)} \left(\prod_{k=1}^{N(T)} \frac{\lambda_{(\tilde{\mu}_l,\alpha,\beta)}(T_k^-)}{\lambda_{(\mu_l,\alpha,\beta)}(T_k^-)} \right) \mathrm{e}^{-(\tilde{\mu}_l - \mu_l)T} .$$

This allows us to obtain approximatively a sample according to the posterior distribution and provides an estimation $\hat{\theta}$ by averaging.

Further works in Bayesian contexts in the multivariate setting include Blundell et al. (2012) for Hawkes processes for the study of a social interaction graph, Donnet et al. (2020) with theoretical results concerning the concentration rates of posterior distributions. These methods have been recently adapted to include inhibition through non-linear Hawkes processes. In Deutsch et al. (2023), the authors propose to compute the log-likelihood by approximating the compensator by a Simpson's rule in a parametric setting. A non-parametric estimation method is proposed in Sulem et al. (2024) which derives posterior concentration rates and in Sulem et al. (2023) through a variational Bayes procedure with a sparsity-inducing paradigm.

2.5 Simulation

2.5.1 Ogata's thinning simulation

The most classic method to simulate point processes with an intensity function λ is obtained through a rejection algorithm. An important condition to implement this method is for the intensity function to be upper-bounded by a constant $\lambda^\star$. We may then simulate N by first simulating a homogeneous Poisson process with intensity $\lambda^\star$ and then keeping each point with probability $\lambda(t)/\lambda^\star$.

In general, it is not possible to find an upper bound for the intensity of a Hawkes process. Additionally, the intensity function is dependent on each simulated point so each time a point appears, the intensity has to be updated. In order to account for these particularities of a Hawkes process, one of the most used methods is Ogata's thinning algorithm (Ogata 1981). To simulate the points of a process N in a window of time $[0,T]$ (with the convention that $T_0 = 0$), we implement the following paradigm.

1. Assume that event time $T_k < T$ has been simulated.
2. Compute an upper bound $\lambda^\star$ of $t \mapsto \lambda(t \mid \mathcal{H}_{T_k})$ on the interval $[T_k, T]$.
3. Simulate a candidate point t_{cand} according to an exponential distribution with mean $1/\lambda^\star$.
4. Compute the intensity $\lambda(t_{cand})$.
5. Accept the candidate with probability $\lambda(t_{cand})/\lambda^\star$.

This algorithm generalises the case when a global upper bound is known by allowing for upper bounds to be local instead.

For the simulation of a Hawkes process, we recall that the interaction function h is such that:

$$\|h\|_1 < 1\,,$$

so in particular it is upper-bounded and attains its upper bound noted $\|h\|_\infty$. We can then sequentially determine a local upper bound. Without any other constraints on h, for a set of simulated event times $(T_1, \dots, T_k)$, a local upper bound for $t \mapsto \lambda(t \mid \mathcal{H}_{T_k})$ on the interval $[T_k, +\infty]$ (assuming no other points) is given by:

$$\lambda^\star = \mu + k\|h\|_\infty\,.$$

This is clearly a very rough bound and can be easily improved under certain conditions, for example if h is a decreasing function. In this case, then λ is piecewise decreasing and so the upper bound becomes:

$$\lambda^\star = \lambda(T_k) + h(0)\,.$$

Choosing a smaller upper-bound improves the simulation time by reducing the overall number of candidates and rejections.

Algorithm 2 displays the adaptation of Ogata's thinning algorithm to Hawkes processes for a non-increasing interaction function.

Algorithm 2: Thinning algorithm for monotone self-exciting Hawkes process.

Input Parameters μ, h a non-increasing function, and a stopping criteria (end-time T or maximal number of jumps N_{max});
Initialization Initialize $\lambda_k = \mu$, $t_k = 0$ and list of times $\mathcal{T} = \emptyset$;
while Stopping criteria not fulfilled **do**
  Set $\lambda^\star = \lambda_k$;
  Generate candidate time $t_{cand} = t_k - \frac{\log(U_1)}{\lambda^\star}$, $U_1 \sim U([0,1])$;
  Compute intensity $\lambda_k = \lambda(t_{cand})$ using sequence of times $\mathcal{T}$;
  Sample $U_1 \sim U([0,1])$;
  if $U_1 \leq \frac{\lambda_k}{\lambda^\star}$ **then**
    Add t_{cand} to sequence of times $\mathcal{T}$;
    Update $\lambda_k = \lambda_k + h(0)$;
  end
  Set $t_k = t_{cand}$;
end
return the sequence of jumps $\mathcal{T}$.

2.5.2 Simulation through branching theory

Another simulation algorithm is based on what is called the branching structure of a Hawkes process. By Definition 2.3.1, the Hawkes process can be seen as a branching Poisson process as described below.

1. Let N_c be a homogeneous Poisson process with intensity μ. All event times $(T_k^c)_{k\in\mathbb{Z}}$ are known as parents.
2. Each parent T_k^c generates a subsidiary process of descendants C_k of event times as follows.
   - A first generation $C_{k,1}$ is generated as an inhomogenous Poisson process with intensity $h(\cdot - T_k^c)$.
   - Each child point t in $C_{k,1}$ generates a new generation $C_{k,2}$ inhomogenous Poisson process with intensity $h(\cdot - t)$.
   - This is repeated for each generated point until no new points are born. C_k is formed by the union of all points in each subprocess $C_{k,j}$.
3. The process N composed of the ordered union of all parents $(T_k^c)_k$ and all points in every subprocess $(C_k)_k$ is a Hawkes process.

Branching Poisson processes were introduced in Bartlett (1963) and Lewis (1964) and allow to leverage the theory of branching processes to study Hawkes processes. In particular, this procedure can be used to simulate a Hawkes process N as presented in Algorithm 3.

Algorithm 3: Branching simulation algorithm for self-exciting Hawkes process

Input Parameters μ, h a positive function, and an end-time T;
Initialization Initialize list of times $\mathcal{T}$, $t_k = 0$ and auxiliary empty list L_{aux};
Generate the number of parents N_0 according to a Poisson distribution with parameter μT;
Generate parent event times $(T_k^c)_k$ as N_0 independent and uniformly distributed points in $[0, T]$, that are then sorted;
Add all points $(T_k^c)_k$ to L_{aux} and $\mathcal{T}$;
while L_{aux} is not empty **do**
  for *t in* L_{aux} **do**
    Generate a number of children with Poisson distribution of parameter $\|h\|_1$;
    Generate the children times distributed according to the probability density function $h(\cdot - t)/\|h\|_1$.;
    Add all points inside the simulation window $[0, T]$ to L_{aux} and $\mathcal{T}$;
  end
end
return Ordered sequence of jumps $\mathcal{T}$.

The fact that such an algorithm will end, in other words that each subsidiary process will end up by not generating any point, is a direct consequence of the condition $\|h\|_1 < 1$. This is a result of Galton-Watson process theory, as each point tends to generate on average less than a children, the branches will "die out" eventually.

This algorithm presents some advantages when compared to Ogata's procedure. First, it does not require to compute an upper-bound for the intensity function, which as seen previously, is an important step in Algorithm 2. Second, this procedure does not have a rejection step and there

is no need to compute the value λ each time that a point is simulated, which can greatly reduce the overall computation time. Overall, the only constraint on h imposed by Algorithm 3 is to be able to simulate according to the probability density function $h(\cdot - t)/\|h\|_1$. Nevertheless, this method is not longer valid to simulate non-linear Hawkes processes that lack the branching structure of the linear Hawkes process.

3 Maximum Likelihood Estimation for Hawkes Processes with self-excitation or inhibition

In this chapter, we present a maximum likelihood method for estimating the parameters of a univariate Hawkes process with self-excitation or inhibition. Our work generalizes techniques and results that were restricted to the self-exciting scenario. The proposed estimator is implemented for the classical exponential kernel and we show that, in the inhibition context, our procedure provides more accurate estimations than current alternative approaches.

Outline of the current chapter

3.1 Introduction

The Hawkes model is a point process observed on the real line, which generally corresponds to the time, where any previously encountered event has a direct influence on the chances of future events occurring. This past-dependent mathematical model was introduced in Hawkes (1971) and its first application was to model earthquakes occurrences (Ogata 1988; Ogata 1998). Since then, Hawkes processes have been widely used in various fields, for instance finance (Bacry et al. 2013), social media (Rizoiu et al. 2017; Mishra et al. 2016), epidemiology (Rizoiu et al. 2018), sociology (Linderman et al. 2014) and neuroscience (Reynaud-Bouret et al. 2014).

The main advantage of Hawkes processes is their ability to model different kinds of relationships between phenomena through an unknown kernel or transfer function. The Hawkes model

was originally introduced as a self-exciting point process where the appearance of an event increases the chances of another one triggering. Several estimation procedures have been proposed for the kernel function, both in parametric (Ogata 1988; Da Fonseca et al. 2013; Ozaki 1979) and nonparametric (Reynaud-Bouret et al. 2014; Bacry et al. 2016) frameworks.

However, the inhibition setting, where the presence of an event decreases the chance of another occurring, has drawn less attention in the literature, although it can be of great interest in several fields, in particular in neuroscience (Lambert et al. 2018). In this inhibition context, the cluster representation (Hawkes et al. 1974) on which is based the construction of a self-exciting Hawkes process, is no longer valid. While the existence and the construction of such nonlinear processes can be found in recent works for the univariate (Costa et al. 2020) and multivariate (Chen et al. 2017) cases, statistical estimation of the kernel function has been hardly addressed. A first approach consists in computing an approximation of the likelihood as if the intensity function could take negative values, and optimizing it to get a maximum likelihood estimator (Lemonnier et al. 2014). Alternatively, the type of interaction (excitation or inhibition) can be considered as a hidden variable, giving rise to a very practical estimation method (Mei et al. 2017).

In this chapter, we propose a maximum likelihood procedure that can handle both excitation and inhibition scenarios for a univariate Hawkes process. Our approach is based on an explicit computation of the likelihood for any type of monotone kernel functions, which is facilitated by the introduction of the natural concept of restart points. The latter are the times when the intensity function, that can be null on some intervals, become strictly positive again. We show that these restart points have a closed-form expression when the kernel is exponential, which allows us to rewrite and maximize the likelihood without approximations that are proposed for instance in Lemonnier et al. (2014). Our estimator is implemented in Python (the code is freely available online[1]). We also propose a numerical study which shows the good performance of our exact estimation procedure compared to approximated approaches, especially when the intensity function is frequently equal to zero.

To outline the chapter, besides a quick introduction to self-regulating Hawkes processes (also referred to as self-correcting Hawkes processes or Hawkes processes with inhibition), Section 3.2 introduces the concepts of underlying intensity function and restart points. General results concerning the compensator and the exact maximum likelihood estimation procedure are described in Section 3.3. At last, after a brief discussion about goodness-of-fit in Section 3.4, Section 3.5 concludes with a numerical study of the estimation error.

3.2 The Hawkes process

Let N be a point process on $\mathbb{R}_+^*$, here $\mathbb{R}_+^* = \{x > 0 : x \in \mathbb{R}\}$, and $(T_k)_{k\geq 1}$ its associated event times (with convention $T_0 = 0$). For any $t \geq 0$, let us note $N(t) = \sum_{k\geq 1} \mathbb{1}_{T_k \leq t}$ the number of events in $[0, t]$ (where $\mathbb{1}$. stands for the indicator function), and λ its conditional intensity function (Daley et al. 2003):

$$\lambda(t) = \lim_{h\to 0} \frac{\mathbb{P}(N(t+h) - N(t) > 0)}{h}\,.$$

[1] `https://github.com/migmtz/hawkes-inhibition-expon`

A univariate Hawkes process is a point process defined by the conditional intensity function:

$$\lambda(t) = \left(\lambda_0 + \int_0^t h(t-s)\,\mathrm{d}N(s)\right)^+ = \left(\lambda_0 + \sum_{T_k \le t} h(t-T_k)\right)^+, \tag{3.1}$$

where $x^+ = \max(0, x)$ denotes the positive part of any real value x, $\lambda_0 \in \mathbb{R}_+^*$ is the baseline intensity and $h : \mathbb{R}_+ \to \mathbb{R}$ is the kernel, which is assumed to be a monotone measurable function with $\lim_{t\to\infty} h(t) = 0$. The kernel function h is the key component of a Hawkes process: it translates the influence (generally assumed to fade away over time) of a past event over the process. Here, h is allowed to take negative values, meaning that it can model both self-exciting and self-regulating Hawkes processes.

Working with such Hawkes processes may prove to be difficult as the positive part function is non-linear. In particular, while computing the compensator function (Daley et al. 2003)

$$\Lambda(t) = \int_0^t \lambda(t)\,\mathrm{d}t\,, \quad \forall t \ge 0, \tag{3.2}$$

is very easy in the self-exciting case (by linearity of the intensity), it becomes more challenging for the self-regulating Hawkes process. As it is the keystone to derive the likelihood function (and then to obtain a parametric estimation method), our first contribution is to provide an exact expression of the compensator.

For this purpose, let us first introduce the *underlying intensity function* and the *restart time*, two quantities which will allow us to derive the computation of the likelihood of a monotone Hawkes process, in a framework unifying self-correcting and self-exciting Hawkes processes.

Definition 3.2.1. Let the *underlying intensity function* of N be:

$$\lambda^\star(t) = \lambda_0 + \int_0^t h(t-s)\,\mathrm{d}N(s)\,.$$

In addition, let the *restart time* $T_k^\star$ be, for any positive integer k:

$$T_k^\star = \inf\left\{t \ge T_k \mid \lambda(t) > 0\right\},$$

along with its corresponding *cooldown interval* $\tau_k^* = T_k^* - T_k$.

As illustrated in Figure 3.1, $\lambda^\star$ corresponds to the intensity λ as if it were allowed to take negative values. Moreover, as the kernel is assumed to be monotone, the restart time associated to one occurrence can be interpreted as the first moment after this occurrence from which λ and $\lambda^\star$ become equal (in particular, the restart time and the occurrence time coincide if the intensity function is nonnegative at this time, see Figure 3.1):

$$T_k^\star = \inf\left\{t \ge T_i \mid \forall t \in (T_k^*, T_{k+1}),\, \lambda(t) = \lambda^\star(t)\right\}.$$

3.3 Maximum likelihood estimation and the exponential model

Assume a parametric model $\mathcal{P} = \{\lambda_\theta, \theta \in \Theta\}$ for the conditional intensity function λ, where θ contains unknown quantities such as the baseline λ_0 and the kernel h. Then, with convention $log(t) = -\infty$ for $t \le 0$, the log-likelihood ℓ_t of any $\theta \in \Theta$ with respect to the observations

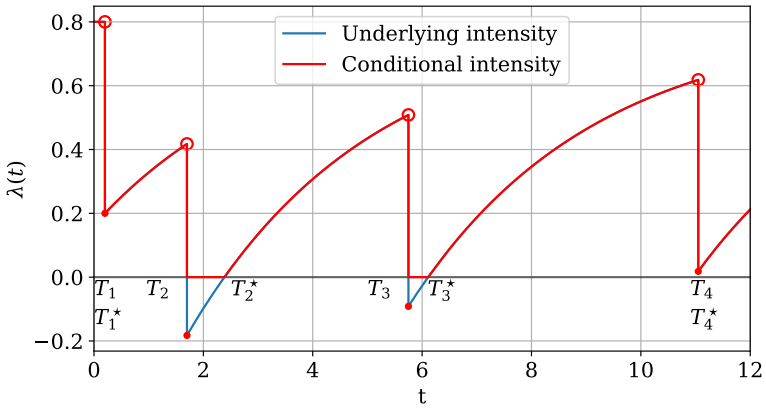

Figure 3.1: Example of the intensity (red curve) and underlying intensity (blue curve) for a self-regulating Hawkes process, with the associated restart times. We only see the negative values of the blue curve since they precisely correspond to the values for which the two intensity functions are not equal.

$T_1, \dots, T_{N(t)}$ in the time interval $[0, t]$ is (Daley et al. 2003, Proposition 7.2.III.), (Ozaki 1979):

$$\ell_t(\theta) = \sum_{k=1}^{N(t)} \log\left(\lambda_\theta(T_k^-)\right) - \Lambda_\theta(t)\,, \tag{3.3}$$

where the compensator Λ_θ is defined as in Equation (3.2) and $\lambda_\theta(T_k^-) = \lim_{t \to T_k^-} \lambda_\theta(t)$.

Equation (3.3) reveals the importance of being able to compute the compensator Λ (equivalently Λ_θ) in order to provide a practical implementation of the maximum likelihood estimator of λ. Thus, a first contribution of this work lies in Proposition 3.3.1, which establishes a decomposition of the compensator Λ using the underlying intensity function $\lambda^\star$ and the restart times $T_1^\star, \dots, T_{N(t)}^\star$.

Proposition 3.3.1. *For any $t > 0$, the compensator Λ can be expressed as:*

$$\Lambda(t) = \begin{cases} \lambda_0 t & \text{if } t < T_1 \\ \lambda_0 T_1 + \displaystyle\sum_{k=1}^{N(t)-1} \int_{T_k^\star}^{T_{k+1}} \lambda^\star(u)\,\mathrm{d}u + \int_{T_{N(t)}^\star}^{t} \lambda^\star(u)\,\mathrm{d}u & \text{if } t \geq T_1\,, \end{cases} \tag{3.4}$$

with the conventions that the sum is equal to 0 if $N(t) = 1$ and the last integral is equal to 0 if $t < T_{N(t)}^*$.

Proof. This comes directly from splitting the integral of $\Lambda(t) = \int_0^t \lambda(t)\,\mathrm{d}t$ on the intervals $[T_k, T_{k+1})$ $(k \in \{0, \dots, N(t) - 1\})$ and $[T_{N(t)}, t]$, and by remarking that, since h is monotone, $\forall t \in [T_k, T_{k+1})$, $\lambda(t) = \lambda^\star(t) \mathbb{1}_{[T_k^*, T_{k+1})}(t)$.

□

In order to give an explicit computation of the quantity $\int_{T_k^\star}^{T_{k+1}} \lambda^\star(u)\,\mathrm{d}u$ (equivalently $\int_{T_{N(t)}^\star}^{t} \lambda^\star(u)\,\mathrm{d}u$) which appears in Proposition 3.3.1, we focus on the classical scenario where we consider an ex-

ponential kernel $h(t) = \alpha e^{-\beta t}$, for some $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}_+^*$. Let us notice that α can be either positive or negative, meaning that the process may be either self-exciting or self-regulating.

Then, the underlying intensity function can be written as:

$$\lambda^\star(t) = \lambda_0 + \int_0^t \alpha e^{-\beta(t-s)} \, \mathrm{d}N(s) \, . \tag{3.5}$$

The forthcoming proposition steps forward in computing the compensator for an exponential kernel.

Proposition 3.3.2 (Compensator for exponential kernel)**.** *Let $t > 0$ and $k \in \{1, \dots, N(t)\}$. The restart times read:*

$$T_k^\star = T_k + \beta^{-1} \log\left(\frac{\lambda_0 - \lambda^\star(T_k)}{\lambda_0}\right) \mathbb{1}_{\lambda^\star(T_k) < 0} \, ,$$

and the compensator is expressed as in Equation (3.4)*, with, for any $\tau \in [T_k^*, T_{k+1}]$:*

$$\int_{T_k^\star}^{\tau} \lambda^\star(u) \, \mathrm{d}u = \lambda_0(\tau - T_k^\star) + \beta^{-1}(\lambda^\star(T_k) - \lambda_0)(e^{-\beta(T_k^\star - T_k)} - e^{-\beta(\tau - T_k)}) \, .$$

Proof. The proof is in 3.A. □

Corollary 3.3.1 (Log-likelihood for exponential kernel)**.** *Let*

$$\mathcal{P} = \left\{ \lambda_\theta = \bar{\lambda_0} + \int_0^t \bar{\alpha} e^{-\bar{\beta}(t-s)} \, \mathrm{d}N(s) : \theta = (\bar{\lambda_0}, \bar{\alpha}, \bar{\beta}) \in \Theta \right\} \, , \tag{3.6}$$

be a parametric exponential model for the conditional intensity function λ with $\Theta = \mathbb{R}_+^* \times \mathbb{R} \times \mathbb{R}_+^*$, along with the candidate compensator Λ_θ, the underlying intensity function $\lambda_\theta^\star$ and the restart times $T_{\theta,1}^\star, \dots, T_{\theta,N(t)}^\star$ associated to λ_θ (see Equation (3.2) *and Definition 3.2.1).*

For any $\theta = (\bar{\lambda_0}, \bar{\alpha}, \bar{\beta}) \in \Theta$, by denoting

$$\Lambda_{\theta,k} = \bar{\lambda_0}(T_k - T_{\theta,k-1}^\star) + \bar{\beta}^{-1}(\lambda_\theta^\star(T_{k-1}) - \bar{\lambda_0})(e^{-\bar{\beta}(T_{\theta,k-1}^\star - T_{k-1})} - e^{-\bar{\beta}(T_k - T_{k-1})}) \, ,$$

the log-likelihood reads (with convention $\log(x) = -\infty$ for $x \leq 0$):

$$\begin{aligned} \ell_t(\theta) = &\log \bar{\lambda_0} - \bar{\lambda_0} T_1 + \sum_{k=2}^{N(t)} \left[\log\left(\bar{\lambda_0} + (\lambda_\theta^\star(T_{k-1}) - \bar{\lambda_0}) e^{-\bar{\beta}(T_k - T_{k-1})} \right) - \Lambda_{\theta,k} \right] \\ &- \left[\bar{\lambda_0}(t - T_{\theta,N(t)}^\star) + \bar{\beta}^{-1}(\lambda_\theta^\star(T_{N(t)}) - \bar{\lambda_0}) \left(e^{-\bar{\beta}(T_{\theta,N(t)}^\star - T_{N(t)})} - e^{-\bar{\beta}(t - T_{N(t)})} \right) \right] \mathbb{1}_{t > T_{\theta,N(t)}^\star} \, . \end{aligned} \tag{3.7}$$

Proof. By Equation (3.8) in the proof of Proposition 3.3.2,

$$\lambda_\theta^\star(T_k^-) = \begin{cases} \bar{\lambda_0} & \text{if } k = 1, \\ \bar{\lambda_0} + (\lambda_\theta^\star(T_{k-1}) - \bar{\lambda_0}) e^{-\bar{\beta}(T_k - T_{k-1})} & \text{if } k \geq 2. \end{cases}$$

Combining this expression with Propositions 3.3.1 and 3.3.2 leads to the result. □

Corollary 3.3.1 exhibits that the log-likelihood for self-regulating Hawkes processes with an exponential kernel can be evaluated in $O(N(t))$ operations (by computing iteratively the quantities $T^{\star}_{\theta,k}$ and $\Lambda_{\theta,k}$ appearing in the summation of Equation (3.7)), as already known for self-exciting exponential Hawkes processes Laub (2014, Chapter 4.2). For other monotone kernels without the Markov property, evaluating the log-likelihood with the method proposed here requires $O(N(t)^2)$ operations, similarly to existing approaches for self-exciting Hawkes processes.

3.4 Goodness-of-fit

Even though computing the compensator Λ (equivalently Λ_θ) was clearly motivated by maximum likelihood estimation, it turns out that it is of great benefit to assess goodness-of-fit, and in particular to check the validity of a maximum likelihood estimation. This is possible thanks to the Time Change Theorem, a result originally stated for inhomogeneous Poisson processes.

Theorem 3.4.1 ((Daley et al. 2003, Theorem 7.4.IV)). *Assume that Λ is continuous, monotone and $\Lambda(t) \xrightarrow[t\to\infty]{} \infty$ a.s. Then a.s., a sequence of event times $(U_k)_{k\geq 1}$ is a realization of N if and only if $(\Lambda(U_k))_{k\geq 1}$ is a realization of a homogeneous Poisson process with unit intensity.*

Let us note that we can find applications of Theorem 3.4.1 to self-exciting Hawkes processes in the literature (Laub 2014, Chapter 5). Since for self-regulating Hawkes processes Λ is still monotone, this result can also be applied in our case.

To be more precise, let us consider $\theta \in \Theta$ and the null hypothesis: "$(U_k)_{k\geq 1}$ is a realization of an exponential Hawkes process with parameter θ". This hypothesis can be tested by applying a Kolmogorov-Smirnov test between the empirical distribution of $(\Lambda_\theta(U_{k+1}) - \Lambda_\theta(U_k))_{k\geq 1}$ and an exponential distribution with parameter 1. This procedure is illustrated in Table 3.1, Section 3.5.

3.5 Numerical Results

This section is aimed at assessing the maximum likelihood estimation method for self-regulating Hawkes processes, based on the exact computation of the compensator Λ_θ in the exponential model (3.6) (Corollary 3.3.1). This procedure is compared to the approximated maximum likelihood estimation proposed in Lemonnier et al. (2014), which consists in approximating Λ_θ by:

$$\Lambda_\theta^{LM}(t) = \int_0^t \lambda_\theta^{\star}(u)\,\mathrm{d}u\,.$$

This optimization procedure is performed with the L-BFGS-B algorithm from the Scipy package (with $(1,0,1)$ as a starting guess and a bounds argument such that $\lambda_0 \geq 0, \alpha \in \mathbb{R}, \beta \geq 0$). In other words, estimators are:

$$\hat{\theta} \in \arg\max_{\theta\in\Theta} \left\{ \ell_{T_{N_{max}}}(\theta) = \sum_{k=1}^{N_{max}} \log\left(\lambda_\theta(T_k^-)\right) - \Lambda_\theta(T_{N_{max}}) \right\},$$

where $N_{max} = 200$ is the total number of jumps and Λ_θ can be replaced by Λ_θ^{LM} to obtain the approximated likelihood proposed in Lemonnier et al. (2014).

The comparison between the exact and the approximated estimation procedure is based on simulated data sets coming from self-correcting Hawkes processes of the form (3.6) with 6 different values of $\theta = (\bar{\lambda}_0, \bar{\alpha}, \bar{\beta}) \in \Theta$ (see Table 3.1) which have been chosen in order to explore different

scenarios, in particular depending on whether the intensity function is frequently null or not. Observations are sets of time jumps generated with a sampling algorithm (see the algorithm in 3.B and Python implementation online), which is a particular case of Ogata's thinning simulation method (Ogata 1981) that can handle Hawkes processes with either self-excitation or inhibition.

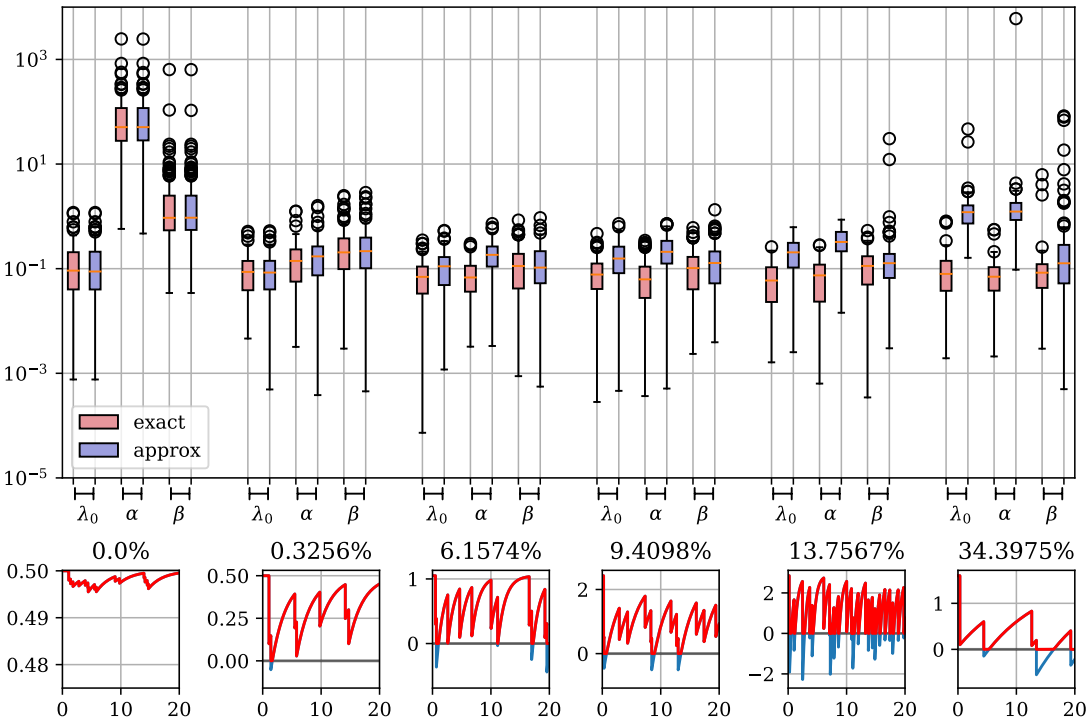

Figure 3.2: Top panel: relative absolute errors of estimations $\hat{\theta} = (\hat{\lambda}_0, \hat{\alpha}, \hat{\beta})$. Bottom panel: example of simulated intensities for each set of values $\theta = (\bar{\lambda}_0, \bar{\alpha}, \bar{\beta})$ with the corresponding average percentage of time when the intensities are equal to zero.

Figure 3.2 represents the relative absolute errors of estimations $\hat{\theta} = (\hat{\lambda}_0, \hat{\alpha}, \hat{\beta})$ for each of the 6 simulated models. We observe that the exact approach provides more accurate estimations than the approximated procedure (as illustrated in the boxplots of Figure 3.2 and by the p-values of the goodness-of-fit tests in Table 3.1). As expected, the more time the conditional intensity equals 0 (from left to right in Figure 3.2), the greater the differences between the two procedures. Furthermore, the leftmost boxplot confirms that when the underlying intensity is nonnegative both methods are mostly identical. Let us note that in this case the estimation of $\bar{\alpha}$ is rather wrong (the estimation of $\bar{\beta}$ is impacted consequently) probably because its value is close to 0 compared to the magnitude of $\bar{\lambda}_0$.

3.6 Discussion

In this chapter we proposed a maximum likelihood approach for Hawkes processes that can handle both self-exciting and self-regulating scenarios, the first case being already covered in the literature and the latter being our main contribution. For this purpose, we define the concepts

	Parameters			Estimations			
	$\bar{\lambda}_0$	$\bar{\alpha}$	$\bar{\beta}$	$\hat{\lambda}_0$	$\hat{\alpha}$	$\hat{\beta}$	p-value
Exact	0.5	-0.001	0.4	0.52	0.03	2.13	0.38
Approx				0.52	0.03	2.11	0.38
Exact	0.5	-0.2	0.4	0.51	-0.21	0.45	0.42
Approx				0.51	-0.22	0.47	0.42
Exact	1.05	-0.75	0.8	1.06	-0.76	0.83	0.43
Approx				1.12	-0.88	0.84	0.35
Exact	2.43	-0.98	0.4	2.59	-1.00	0.38	0.53
Approx				2.90	-1.21	0.40	0.45
Exact	2.85	-2.5	1.8	2.81	-2.56	1.87	0.36
Approx				1.55×10^4	-1.63×10^7	3.04	0.07
Exact	1.6	-0.75	0.1	1.62	-0.76	0.11	0.42
Approx				2.72×10^7	-2.31×10^{10}	0.27	2.14×10^{-05}

Table 3.1: Quantitative assessment of the numerical study: sets of true parameters (left), average estimations over 50 repetitions (middle) and average p-values over 50 independent realisations for the test of Section 3.4.

of underlying intensity function and restart times when working with monotone kernel functions. In particular we obtain exact expressions of the compensator for the exponential Hawkes process which is the key step of the estimation procedure. We present numerical results on synthetic data that show the efficiency of our procedure, with a substantial improvement compared to approximated approaches when the intensity function is frequently null.

From a theoretical point of view, future work will consist in adapting analytical results to study the convergence of our estimator in the self-regulating case. Regarding modeling, it would be of great interest to consider kernel functions outside the classical exponential scenario. Another important step is the extension of our concepts and algorithms to the multivariate version of the process, which is not straightforward since in the multivariate setting the expression of the restart times are no longer explicit. This last point is essential in order to target real-world datasets since in many applications, being limited to the univariate case will lead to detect self-excitation. However, a model that accounts for potential inhibition effects is of great interest when considering interactions between events of different natures, which will typically be modeled by a multivariate process. This multidimensional extension is the object of a future work, with a further perspective to use our procedure in neuroscience applications in order to detect attraction and repulsion effects between neurons.

3.A Proof of Proposition 3.3.2

Let us begin by expressing the underlying intensity function between two event times. First, $\lambda^\star(t) = \lambda_0$ for $t \in [0, T_1)$. Then, for any $k \in \mathbb{N}$, for all $t \in [T_k, T_{k+1})$, the underlying intensity is differentiable in t and

$$(\lambda^\star)'(t) = -\beta(\lambda^\star(t) - \lambda_0)\,,$$

with the left condition: $\lambda_k^\star := \lambda^\star(T_k)$. Solving this differential equation leads to

$$\lambda^\star(t) = \lambda_0 + (\lambda_k^\star - \lambda_0)\mathrm{e}^{-\beta(t-T_k)}\,. \tag{3.8}$$

Now, by definition of the restart time $T_k^\star = \inf\{t \geq T_k \mid \lambda(t) > 0\}$, we have that if $\lambda_k^\star \geq 0$, then $T_k^\star = T_k$. Otherwise, as $\lambda^\star$ is continuous on the interval $[T_k, T_{k+1})$, we obtain $T_k^\star$ by solving for t: $\lambda^\star(t) = 0$. Thus, by Equation (3.8):

$$\lambda^\star(T_k^\star) = 0 \iff T_k^\star = T_k + \beta^{-1} \log\left(\frac{\lambda_0 - \lambda_k^\star}{\lambda_0}\right).$$

Gathering both situations, we obtain the first part of Proposition 3.3.2:

$$T_k^\star = T_k + \beta^{-1} \log\left(\frac{\lambda_0 - \lambda_k^\star}{\lambda_0}\right) \mathbb{1}_{\lambda_k^\star < 0}$$

.

Let now $k \in \{1, \ldots, N(t)\}$ and $\tau \in [T_k^\star, T_{k+1}]$. By Equation (3.8),

$$\begin{aligned}\int_{T_k^\star}^{\tau} \lambda^\star(u)\,\mathrm{d}u &= \int_{T_k^\star}^{\tau} \left(\lambda_0 + (\lambda_k^\star - \lambda_0)\mathrm{e}^{-\beta(u - T_k)}\right) \mathrm{d}u \\ &= \lambda_0(\tau - T_k^\star) + \beta^{-1}(\lambda_k^\star - \lambda_0)(\mathrm{e}^{-\beta(T_k^\star - T_k)} - \mathrm{e}^{-\beta(\tau - T_k)}),\end{aligned}$$

which is the second part of Proposition 3.3.2.

3.B Simulation algorithm

Algorithm 4 builds upon Ogata's thinning simulation method (Ogata 1981, Proposition 1) in order to handle Hawkes processes with either self-excitation or inhibition.

Algorithm 4: Thinning algorithm for monotone Hawkes process.

Input Parameters λ_0, h a monotone function, and a stopping criteria (end-time T or maximal number of jumps N_{max});
Initialization Initialize $\lambda_k = \lambda_0$, $t_k = 0$ and list of times $\mathcal{T} = \emptyset$;
while Stopping criteria not fulfilled **do**
  Set $\lambda_{max} = \max(\lambda_0, \lambda_k)$;
  Generate candidate time $t_{cand} = t_k - \frac{\log(U_1)}{\lambda_{max}}$, $U_1 \sim U([0,1])$;
  Estimate intensity $\lambda_k = \lambda(t_{cand})$ using sequence of times $\mathcal{T}$;
  Sample $U_2 \sim U([0,1])$;
  if $U_2 \leq \frac{\lambda_k}{\lambda_{max}}$ **then**
    Add t_{cand} to sequence of times $\mathcal{T}$;
    Update $\lambda_k = \lambda_k + \alpha$;
  end
  Set $t_k = t_{cand}$;
end
return the sequence of jumps $\mathcal{T}$.

4

Inference of multivariate exponential Hawkes processes with inhibition and application to neuronal activity

The multivariate Hawkes process is a past-dependent point process used to model the relationship of event occurrences between different phenomena. Although the Hawkes process was originally introduced to describe excitation interactions, which means that one event increases the chances of another occurring, there has been a growing interest in modelling the opposite effect, known as inhibition. In this chapter, we focus on how to infer the parameters of a multidimensional exponential Hawkes process with both excitation and inhibition effects. Our first result is to prove the identifiability of this model under a few sufficient assumptions. Then we propose a maximum likelihood approach to estimate the interaction functions, which is, to the best of our knowledge, the first exact inference procedure in the frequentist framework. Our method includes a variable selection step in order to recover the support of interactions and therefore to infer the connectivity graph. A benefit of our method is to provide an explicit computation of the log-likelihood, which enables in addition to perform a goodness-of-fit test for assessing the quality of estimations. We compare our method to standard approaches, which were developed in the linear framework and are not specifically designed for handling inhibiting effects. We show that the proposed estimator performs better on synthetic data than alternative approaches. We also illustrate the application of our procedure to a neuronal activity dataset, which highlights the presence of both exciting and inhibiting effects between neurons.

Outline of the current chapter

4.1 Introduction

A Hawkes process is a point process in which each point is commonly associated with event occurrences in time. In this past-dependent model, every event time impacts the probability that other events take place subsequently. These processes are characterised by the conditional intensity function, seen as an instantaneous measure of the probability of event occurrences. Since their introduction in Hawkes (1971), Hawkes processes have been applied in a wide variety of fields, for instance in seismology (Ogata 1988), social media (Rizoiu et al. 2017), criminology (Olinde et al. 2020) and neuroscience (Lambert et al. 2018).

The multidimensional version of this model, referred to as the multivariate Hawkes process, describes the appearance of different types of events, the occurrences of which are influenced by all past events of all types. Each interaction between two types of events is encoded in kernel functions, also called interaction functions. Originally this model takes only into account mutually exciting interactions - an event increases the chances of others occurring - by assuming that all kernel functions are non-negative. A specificity of self-exciting Hawkes processes is their branching structure, also known as cluster structure. Introduced in Hawkes et al. (1974), this parallel between Hawkes processes and branching theory has provided the first theoretical background for the self-exciting Hawkes model, in particular existence and expected number of points on a finite interval. Estimation methods in the literature are vast including maximum likelihood estimators (Ozaki 1979; Guo et al. 2018) and method of moments (Da Fonseca et al. 2013). Nonparametric approaches include an EM procedure introduced in Lewis et al. (2011), estimations obtained via the solution of Wiener-Hopf equations (Bacry et al. 2016) or by approximating the process through autoregressive models (Kirchner 2017) or through functions in reproducing kernel Hilbert spaces (Yang et al. 2017).

Although the self-exciting Hawkes process remains widely studied, there has been a growing interest in modeling the opposite effect, known as inhibition, in which the probability of observing an event is lowered by the occurrence of certain events. In practice, this amounts

to considering negative kernel functions. In order to maintain the positivity of the intensity function, a non-linear operator is added to the expression which in turns entails the loss of the cluster representation. This model known as the non-linear Hawkes process was first presented in Brémaud et al. (1996), where existence of such processes was proved via construction using bi-dimensional marked Poisson processes. Such approach of analysis has been used in the literature as in Chen et al. (2017), where a coupling process is established and leveraged to obtain theoretical guarantees on cross-analysis covariance. Another approach is presented in Costa et al. (2020), where renewal theory allows to obtain limit theorems for processes with bounded support kernel functions. Estimation methods focus mainly on nonparametric methods for general interactions and non-linear functions, as found in Bacry et al. (2016) and Sulem et al. (2024).

In the last years, alternative models have been designed in order to take into account inhibiting effects in Hawkes processes. An example is the neural Hawkes process, presented in Mei et al. (2017) and Zuo et al. (2020), which combines a multivariate Hawkes process and a recurrent neural network architecture. In Duval et al. (2022), a multiplicative model considers two sets of neuronal populations, one exciting and another inhibiting, and each intensity function is the product of two non-linear functions (one for each group). Another model is presented in Olinde et al. (2020) and called self-limiting Hawkes process. It includes the inhibition as a multiplicative term in front of a the traditional self-exciting intensity function.

In this chapter, we present a maximum likelihood estimation method for multivariate Hawkes processes with exponential kernel functions, that works for both exciting and inhibiting interactions, as modelled by Brémaud et al. (1996) and Chen et al. (2017). This work builds upon the methodology for the univariate case, presented in Bonnet et al. (2021), by focusing in the intervals where the intensity function is positive. We show that, under a weak assumption on the kernel functions, these intervals can be determined exactly. We can then write for each dimension the integral of the intensity function, known in the literature as the compensator, which in turn provides an explicit expression of the log-likelihood. This enables to build the corresponding maximum likelihood estimator and we complete our procedure with a variable selection step to recover the significant interactions within the whole process. This is of particular interest since it provides a graphical interpretation of the model and it can also be used a reduction dimension tool. Our numerical procedure is implemented in Python and freely available on GitHub.[1] As a by-product of our method, the closed-form expression of the compensator also allows to assess goodness-of-fit via the Time Change Theorem and multiple testing. We carry out a numerical study on simulated data and on a neuronal activity dataset (Petersen et al. 2016; Radosevic et al. 2019). The performance of our approach is compared to estimations obtained via approximations from Bacry et al. (2020) and Lemonnier et al. (2014), and we show that our method not only achieves better estimations but is capable of identifying correctly the interaction network of the process.

To outline this chapter, Section 4.2 presents the multivariate Hawkes process framework and reviews the literature regarding inference of non-linear Hawkes processes. In Section 4.3, we detail our procedure, including the maximum likelihood estimation, variable selection and goodness-of-fit test to assess the quality of the estimations. We also address the question of identifiability of the model, that we prove under a few sufficient conditions. The whole procedure is illustrated on simulated data in Section 4.4 and applied to a neuronal activity dataset in Section 4.5. In Section 4.6 we discuss our contributions and its current limitations along with interesting perspectives for future work.

[1] `https://github.com/migmtz/multivariate-hawkes-inhibition`

4.2 The multivariate Hawkes process

4.2.1 Definition

A multivariate Hawkes process $N = (N^1, N^2, \dots, N^d)$ of dimension d is defined by d point processes on $\mathbb{R}_+^*$, denoted $N^i \colon \mathcal{B}(\mathbb{R}_+^*) \to \mathbb{N}$, where $\mathcal{B}(\mathbb{R}_+^*)$ is the Borel algebra on the set of positive numbers.

Each process N^i can be characterised by its associated event times $\left(T_k^i\right)_k$ and its conditional intensity function, defined for all $t \geq 0$ by

$$\begin{aligned} \lambda^i(t) &= \left(\mu_i + \sum_{j=1}^{d} \int_0^t h_{ij}(t-s)\, \mathrm{d}N^j(s) \right)^+ \\ &= \left(\mu_i + \sum_{j=1}^{d} \sum_{T_k^j \leq t} h_{ij}(t - T_k^j) \right)^+ , \end{aligned} \tag{4.1}$$

where $x^+ = \max(0, x)$. Here, the quantity $\mu_i \in \mathbb{R}_+^*$ is called the baseline intensity and each interaction or kernel function $h_{ij} \colon \mathbb{R}_+^* \to \mathbb{R}$ represents the influence of the process N^j on the process N^i and T_k^j corresponds to the k-th event time of process N^j.

Remark. *The positive-part function in Equation* (4.1) *is needed to ensure the non-negativity of λ^i in the presence of strong inhibiting effects, that is when some interaction functions h_{ij} are sufficiently negative compared to positive contributions. Concretely, the positive part does not affect the intensity function if inhibiting effects are in minority compared to the positive contributions (exciting effects or baseline intensities).*

Remark. *Equation* (4.1) *may question the reader for two reasons. First, it is the cadlag definition of a the conditional intensity of a Hawkes process. It's our choice to prefer it to the caglad version but all the results presented here can be written in this setting. Second, it is considered that the history is empty for $t < 0$. It is a common choice for statistical inference (a finite amount a times is observed) while the infinite history is preferred for a probabilistic analysis based on a stationary assumption.*

For each process N^i and for all $t \geq 0$, let us note $N^i(t) = \sum_{k \geq 1} \mathbb{1}_{T_k^i \leq t}$ the measure of $(0, t]$ and the compensator

$$\Lambda^i(t) = \int_0^t \lambda^i(u)\, \mathrm{d}u \,.$$

The process N can be seen as a point process on $\mathbb{R}_+^*$, where for any $B \in \mathcal{B}(\mathbb{R}_+^*)$, $N(B) = \sum_{i=1}^d N^i(B)$. Similarly to a univariate process, N can be characterised by its conditional intensity λ (also called total intensity):

$$\lambda(t) = \sum_{i=1}^{d} \lambda^i(t) \,, \tag{4.2}$$

and by its compensator

$$\Lambda(t) = \int_0^t \lambda(u)\, \mathrm{d}u = \sum_{i=1}^{d} \Lambda^i(t) \,.$$

From this point of view, the process N is associated to event times $\left(T_{(k)}\right)_k = \left(T_{u_k}^{m_k}\right)_k$,

corresponding to the ordered sequence composed of $\bigcup_{i=1}^{d}\{T_k^i \mid k>0\}$, and we may define, for every $t \geq 0$, $N(t) = \sum_{k\geq 1} \mathbb{1}_{T_{(k)}\leq t} = \sum_{i=1}^{d} N^i(t)$. Here, $(u_k)_k$ is the random ordering sequence and $(m_k)_k$ the sequence of marks that make it possible to identify to which dimension each time corresponds. These marks can be written as

$$m_k = \sum_{j=1}^{d} j \mathbb{1}_{N^j(\{T_{(k)}\})=1} \,.$$

Remark. *A more detailed introduction of multivariate point processes via the concept of marked point processes can be found in Daley et al. (2003, Chapter 6.4)*

As the aim of this chapter is to describe a practical methodology for estimating the conditional intensities $\lambda_1, \ldots, \lambda_d$ via maximising the log-likelihood, the latter quantity has to be made explicit. Let $t \geq 0$ and assume that event times $\left\{T_k^i : 1 \leq k \leq N_i(t), 1 \leq i \leq d\right\}$ are observed in the interval $(0,t]$. Then, given a parametric model $\mathcal{P} = \{(\lambda^1_{\theta_1}, \ldots, \lambda^d_{\theta_d})\colon \theta = (\theta_1, \ldots, \theta_d) \in \Theta_1 \times \cdots \times \Theta_d\}$ (and associated compensators $\Lambda^1_{\theta_1}, \ldots, \Lambda^d_{\theta_d}$) for conditional intensity functions $\lambda^1, \ldots, \lambda^d$, for every $\theta \in \Theta$, the log-likelihood $\ell_t(\theta)$ reads Daley et al. (2003, Proposition 7.3.III.)

$$\ell_t(\theta) = \sum_{i=1}^{d} \ell_t^i(\theta_i) \,,$$

with

$$\ell_t^i(\theta_i) = \sum_{k=1}^{N^i(t)} \log \lambda^i_{\theta_i}(T_k^{i-}) - \Lambda^i_{\theta_i}(t) \,, \tag{4.3}$$

where $\lambda^i_{\theta_i}(T_k^{i-}) = \lim_{t\to T_k^{i-}} \lambda^i_{\theta_i}(t)$ and with convention $\log(x) = -\infty$ for $x \leq 0$.

The heart of the problem in deriving a maximum likelihood estimator for the conditional intensities λ^i is being able to evaluate exactly the compensator values $\Lambda^i_\theta(t)$ for every possible $\theta \in \Theta$, which requires to determine when the conditional intensities λ^i are non-zero. The forthcoming sections clear this point up.

4.2.2 Related work

Estimation methods for Hawkes processes have focused mainly on self-exciting interactions (by assuming $h_{ij} \geq 0$). In Ozaki (1979), the author presents the maximum likelihood estimation method for univariate processes with exponential kernel $h(t) = \alpha \mathrm{e}^{-\beta t}$ ($\alpha > 0$, $\beta > 0$), the same method being established in Mishra et al. (2016) for the power law kernel function $h(t) = \frac{\alpha\beta}{(1+\beta t)^{1+\gamma}}$ ($\alpha > 0$, $\beta > 0$, $\gamma > 0$). In Chen et al. (2018) the maximum likelihood method is presented for the multivariate version with exponential kernel, while Bacry et al. (2020) proposed an inference method based on optimising a least-squares criterion. Other methods in the parametric setting include spectral analysis (Adamopoulos 1976), EM algorithm (Veen et al. 2008) and method of moments (Da Fonseca et al. 2013).

Estimators of the interaction functions h_{ij} are also presented in a nonparametric setting. For instance, Yang et al. (2017) proposed to estimate h_{ij} in a reproducing kernel Hilbert space. Reynaud-Bouret et al. (2014) proposed a decomposition of the interaction functions h_{ij} on a histogram basis with bounded support, the estimation of which are obtained by minimising a least-squares contrast. Hawkes processes with excitation have also been studied in a Bayesian context, with likelihood-based approaches, as in Rasmussen (2013) for the univariate case and

in Donnet et al. (2020) for multivariate processes.

Although inhibiting effects in Hawkes processes were first mentioned in Brémaud et al. (1996), they have only met a growing interest in the last decade. Concerning inference, most of the known methods are not designed for handling the inhibiting case: nevertheless some are able in practice to estimate negative interactions by minimising a least-squares criterion, but without guaranteeing that the estimated intensity functions remain non-negative (Reynaud-Bouret et al. 2014; Bacry et al. 2020). A similar approach is proposed in Lemonnier et al. (2014) for maximum likelihood estimation, where the compensator Λ^i is approximated by integrating the conditional intensity λ^i without the positive part function (see Equation (4.1)). Obviously, these methods should perform well when the intensities remain mostly positive, but it is unclear how they will adapt to scenarios when the intensities are frequently equal to zero due to inhibiting terms. A similar remark is mentioned in Bacry et al. (2016) concerning their estimation method, that can provide negative estimations of the interactions only if there is a negligible chance of the intensities to be null.

Inference procedures that are specifically dedicated to Hawkes processes with inhibition are scarcer in the literature. Sulem et al. (2024) presents various results for non-linear Hawkes processes including inhibition effects: existence, stability and Bayesian estimation for kernel functions with bounded support. Deutsch et al. (2023) presents choices of priors for Bayesian estimation based on a new reparametrisation of the process.

Lastly, Bonnet et al. (2021) presents a maximum likelihood estimation adapted to the univariate Hawkes process with inhibition and monotone kernel functions. The decisive contribution of this work is to give, for an exponential kernel $h(t) = \alpha \mathrm{e}^{-\beta t}$ ($\alpha \in \mathbb{R}$, $\beta > 0$), a closed-form expression of restart times, which are defined as the instants at which the single conditional intensity becomes non-zero. This makes possible to compute explicitly the compensator and then the log-likelihood. Yet, this study is limited to the univariate case. It has to be noted that a formalism similar to Bonnet et al. (2021) for multivariate Hawkes processes is mentioned in Deutsch et al. (2023), but used neither for maximum likelihood estimation, nor for goodness-of-fit tests.

This work goes a step forward in estimation of multivariate Hawkes processes with inhibition, by providing the first exact maximum likelihood method for exponential interactions $h_{ij}(t) = \alpha_{ij}\mathrm{e}^{-\beta_{ij}t}$, combined with a variable selection procedure. As it will be explained in the next section, the proposed approach also enables to perform standard goodness-of-fit tests.

4.3 Estimation and goodness-of-fit

4.3.1 Introductive example

Before motivating and explaining the estimation procedure proposed in this chapter, we present an example of multivariate Hawkes process. Figure 4.1 depicts in red conditional intensities λ^1 and λ^2 for a realisation of a 2-dimensional Hawkes process (see the forthcoming section for the definition of underlying intensities). The existence of such a process (along with its stationarity) is ensured by controlling the spectral radius $\rho(S^+) < 1$ of the matrix $S^+ = (\|h^+_{ij}\|_1)_{ij}$ (Deutsch et al. 2023). Similar results with slightly different conditions can be found in Brémaud et al. (1996) and Sulem et al. (2024). The simulation has been carried out with baselines $\mu_1 = 0.5$ and $\mu_2 = 1.0$, and with exponential kernels $h_{ij}(t) = \alpha_{ij}\mathrm{e}^{-\beta_{ij}t}$ parameterised by:

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} = \begin{pmatrix} -1.9 & 3.0 \\ 0.9 & -0.7 \end{pmatrix}, \qquad \text{and} \qquad \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} = \begin{pmatrix} 2.0 & 20.0 \\ 3.0 & 2.0 \end{pmatrix}.$$

These kernels have been chosen such that both processes are self-inhibiting ($\alpha_{11}, \alpha_{22} < 0$) but inter-exciting ($\alpha_{12}, \alpha_{21} > 0$).

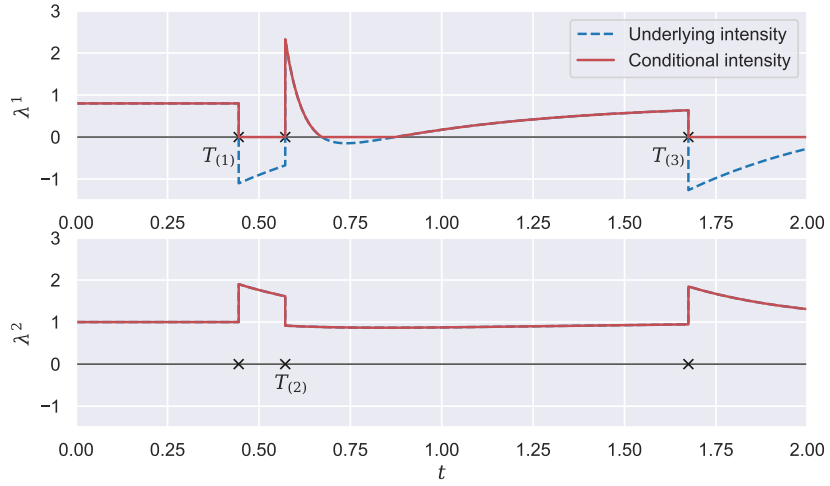

Figure 4.1: Simulation of a 2-dimensional Hawkes process. Each cross corresponds to an event time, and each $T_{(k)}$ is shown in its corresponding process.

The goal of this work is to establish a parametric estimation method, via maximum likelihood estimation, that is able to handle both excitation and inhibition frameworks in the multivariate case. For this purpose, it is necessary to compute explicitly the log-likelihood $\ell_t(\theta)$ (see Equation (4.3)) and in particular to evaluate the compensator Λ_θ^i, expressed as an integral of λ_θ^i. For the latter, the main challenge is to determine when conditional intensities λ^i are non-zero, that is on which intervals they are tailored by the exponential interaction functions and not by the positive-part operator.

In Bonnet et al. (2021), the authors solved this challenge for univariate processes by remarking that the conditional intensity is monotone between two event times. Figure 4.1 illustrates that this is not necessarily true for multivariate processes (here, between $T_{(2)}$ and $T_{(3)}$). This constitutes the major difficulty we have to cope with.

4.3.2 Underlying intensity and restart times in the multivariate setting

From now on, let us focus on the exponential model (Hawkes 1971), where each interaction function h_{ij} is then defined as

$$h_{ij}(t) = \alpha_{ij} \mathrm{e}^{-\beta_{ij} t} \,,$$

with $\alpha_{ij} \in \mathbb{R}$ and $\beta_{ij} \in \mathbb{R}_+^*$ for $i, j \in \{1, \dots, d\}$. For each $i \in \{1, \dots, d\}$, the underlying intensity function $\lambda^{i\star}$ is defined as in Bonnet et al. (2021) for the univariate case:

$$\lambda^{i\star}(t) = \mu_i + \sum_{j=1}^{d} \int_0^t h_{ij}(t-s) \, \mathrm{d}N^j(s) \,.$$

This quantity coincides with the conditional intensity λ^i when it is non-zero, and is non-positive otherwise. In particular, we can observe that $\lambda^i(t) = \big(\lambda^{i\star}(t)\big)^+$ (see Figure 4.1).

As explained in the previous section, the main difficulty of the multivariate exponential setting is the non-monotony of conditional intensities λ^i between two event times. Determining intervals where λ^i is non-zero (that is when $\lambda^{i\star}$ is positive) would require to numerically find the roots of a high-degree polynomial, which is expensive and inexact. To alleviate this problem, we introduce Assumption 2.

Assumption 2. *For each $i \in \{1, \dots, d\}$, there exists $\beta_i \in \mathbb{R}_+^*$ such that $\beta_{ij} = \beta_i$ for all $j \in \{1, \dots, d\}$.*

Remark. *This model with constant recovery rates β_i has been studied before in the works of Ogata (1981) in the self-exciting version of a 2-dimensional Hawkes process. Intuitively, this assumption considers the situation where the rate of "dissipation" of any internal or external effect is dependent only on the receiving phenomenon. For instance, for neuronal interactions, each activation from neuron j will have an impact on a connected neuron i dependent on both neurons $(\alpha_{ij})_{ij}$ but the "recovery" time can be assumed to depend only on the connected neuron i $(\beta_i)_i$.*

As we will see in Lemma 4.3.1, this assumption enables to recover the monotony of the conditional intensities between two times. It remains now to determine when the underlying intensity $\lambda^{i\star}$ is negative. To do so, we define the restart times in the multivariate framework, to be, for any k and i:

$$T_{(k)}^{i\star} = \min\left(\inf\left\{t \geq T_{(k)} \colon \lambda^{i\star}(t) \geq 0\right\}, T_{(k+1)}\right).$$

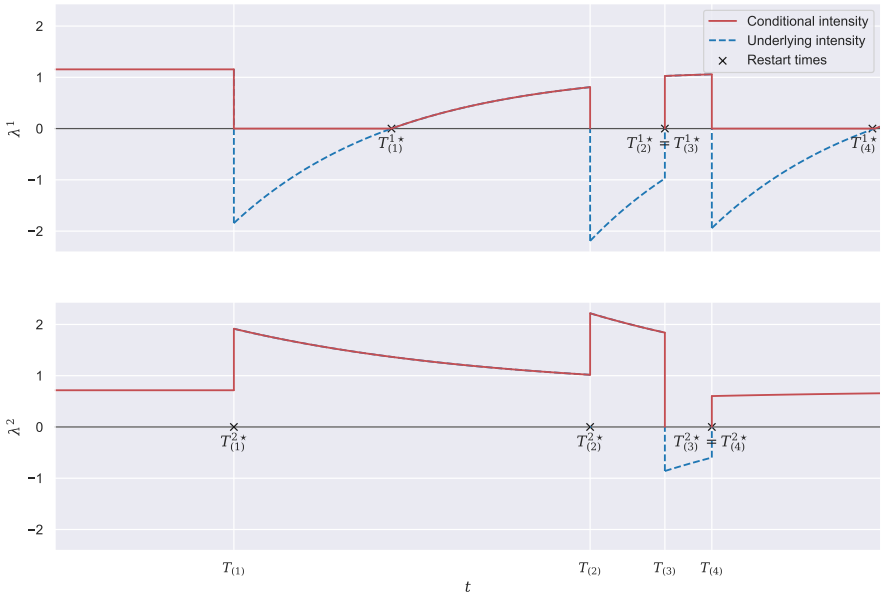

Figure 4.2: Illustration of restart times $(T_{(k)}^\star)_{1\leq k\leq 4}$ for each subprocess of a 2-dimensional process associated with event times $(T_{(k)})_{1\leq k\leq 4}$.

As exemplified in Figure 4.2, the restart time $T_{(k)}^{i\star}$ (associated to the sub-process i) corresponds to the first instant between $T_{(k)}$ and $T_{(k+1)}$ from which $\lambda^{i\star}(t)$ becomes non-negative (or $T_{(k+1)}$ if this instant does not exist). Intuitively, it means that $\lambda^i(t) = \lambda^{i\star}(t)$ on $(T_{(k)}^{i\star}, T_{(k+1)})$ and

$\lambda^i(t) = 0$ elsewhere on $(T_{(k)}, T_{(k+1)})$. This is formalised in Lemma 4.3.1. In particular, it appears that the restart time $T^{i\star}_{(k)}$ can be expressed as a function of

$$T_{(k)} + \beta_i^{-1} \log\left(\frac{\mu_i - \lambda^{i\star}(T_{(k)})}{\mu_i}\right),$$

which is the single root to the equation $\mu_i + (\lambda^{i\star}(T_{(k)}) - \mu_i)\mathrm{e}^{-\beta_i(t-T_{(k)})} = 0$ on the interval $[T_{(k)}, +\infty)$ when $\lambda^{i\star}(T_{(k)}) < 0$ (see top panel of Figure 4.2). Then, Proposition 4.3.1 gives an explicit formulation of the compensator of each subprocessi N^i, which is needed to compute the log-likelihood (see Equation (4.3)). The proofs of these two results are presented respectively in Appendix 4.A and Appendix 4.B.

Lemma 4.3.1. *If Assumption 2 is granted, then for each $i \in \{1, \dots, d\}$ and any $k \geq 1$:*

$$T^{i\star}_{(k)} = \min\left(t^\star_k, T_{(k+1)}\right). \tag{4.4}$$

where

$$t^\star_k = \left(T_{(k)} + \beta_i^{-1} \log\left(\frac{\mu_i - \lambda^{i\star}(T_{(k)})}{\mu_i}\right)\mathbb{1}_{\{\lambda^{i\star}(T_{(k)})<0\}}\right)$$

Furthermore, for all $t \in (T_{(k)}, T_{(k+1)})$,

$$\lambda^i(t) = \begin{cases} \lambda^{i\star}(t) > 0 & \text{if } t \in (T^{i\star}_{(k)}, T_{(k+1)}) \\ 0 & \text{otherwise.} \end{cases}$$

Proposition 4.3.1. *[Compensator for multivariate exponential kernels] Let us suppose that Assumption 2 is granted. For each $i \in \{1, \dots, d\}$ the compensator Λ^i of the process N^i reads, $\forall t \geq 0$:*

$$\Lambda^i(t) = \begin{cases} \mu_i t & \text{if } t < T_{(1)} \\ \mu_i T_{(1)} + \sum_{k=1}^{N(t)} J_k & \text{if } t \geq T_{(1)}\,, \end{cases} \tag{4.5}$$

where for all integer $k \in \{1, \dots, N(t)\}$:

$$J_k = \mu_i\left[\min(t, T_{(k+1)}) - T^{i\star}_{(k)}\right] + \beta_i^{-1}\left(\lambda^{i\star}(T_{(k)}) - \mu_i\right)\left[\mathrm{e}^{-\beta_i(T^{i\star}_{(k)} - T_{(k)})} - \mathrm{e}^{-\beta_i(\min(t, T_{(k+1)}) - T_{(k)})}\right].$$

4.3.3 Identifiability and likelihood computation

As already mentioned in Section 4.2, let $t \geq 0$ and assume that event times $\left\{T^i_k : 1 \leq k \leq N_i(t), 1 \leq i \leq d\right\}$ are observed in the interval $(0, t]$. We consider the parametric exponential model for a multivariate Hawkes process of dimension d, defined by

$$\mathcal{P} = \left\{(\lambda^1, \dots, \lambda^d) \colon \lambda^1 \in \mathcal{P}^1, \dots, \lambda^d \in \mathcal{P}^d\right\},$$

where for each $i \in \{1, \dots, d\}$, $\mathcal{P}^i$ is the exponential parametric model for the process N^i:

$$\mathcal{P}^i = \left\{ \lambda^i_{\theta_i} = \left(\mu_i + \sum_{j=1}^{d} \int_{-\infty}^{t} \alpha_{ij} \mathrm{e}^{-\beta_i(t-s)}\, \mathrm{d}N^j(s)\right)^+ : \theta_i = (\mu_i, \alpha_{i1}, \dots, \alpha_{id}, \beta_i) \in \Theta \right\},$$

where $\Theta = \mathbb{R}^\star_+ \times \mathbb{R}^d \times \mathbb{R}^\star_+$. For a candidate set of intensities $(\lambda^1_{\theta_1}, \dots, \lambda^d_{\theta_d})$, the underlying intensity functions are denoted $\lambda^{i\star}_{\theta_i}$ ($i \in \{1, \dots, d\}$), the compensators $\Lambda^i_{\theta_i}$ and the restart times $(T^{i\star}_{\theta_i,(k)})_k$.

Now, given a realisation $\left(T_{(k)}\right)_{k>0}$ of a multivariate exponential Hawkes process, Theorem 4.3.1 establishes the correspondence between the conditional intensities and the parameters.

Theorem 4.3.1 (Identifiability)**.** *Let $N = (N^1, \dots, N^d)$ be a multivariate Hawkes process defined by a set of intensity functions $(\lambda^1_{\theta_1}, \dots, \lambda^d_{\theta_d}) \in \mathcal{P}$, for some $\theta = (\theta_1, \dots, \theta_d) \in \Theta^d$. Let also $\left(T_{(k)}\right)_{k>0}$ be a realisation of N and $\mathcal{F}_t$ be the corresponding filtration.*

Let us assume that a.s. for every $(i, j) \in \{1, \dots, d\}^2$, $i \neq j$, there exist an event time τ from process N^j, and an event time $\tau_+ > \tau$ from process N^i, such that:

1. *$\lambda^i_{\theta_i}(\tau^-) > 0$;*
2. *there are only events of process N^j in the interval $[\tau, \tau_+)$.*

Then, for any $\theta' \in \Theta^d$,

$$\forall i \in \{1, \dots, d\},\ \lambda^i_{\theta_i}(t \mid \mathcal{F}_t) = \lambda^i_{\theta'_i}(t \mid \mathcal{F}_t)\ \text{a.e.} \qquad \Longleftrightarrow \qquad \theta = \theta'\,.$$

The proof is presented in Appendix 4.C. To the best of our knowledge, the only identifiability result for non-linear multivariate Hawkes processes is given by Sulem et al. (2024) but only if interaction functions h_{ij} have a bounded support. Their proof strongly relies on this assumption since they extend to the multivariate case the renewal properties proved by Costa et al. (2020) for non-linear univariate Hawkes processes with a bounded kernel. Their proof also requires some assumptions to ensure that one process is not totally inhibited, which is also a consequence of the assumptions that we propose, as discussed in Section 4.3.4.

As expected, Proposition 4.3.1 makes it possible to compute explicitly the log-likelihood expressed in Equation (4.3) for multivariate exponential Hawkes processes. This is formalised in Corollary 4.3.1 (and proved in Appendix 4.D).

Corollary 4.3.1. *Let $i \in \{1, \dots, d\}$ and $k \geq 2$ an integer. Let us denote*

$$S^i_k := T_{(N(T^i_k)-1)} = T_{(\max\{\ell \in \mathbb{N}^* : T_{(\ell)} < T^i_k\})}\,,$$

the time preceding directly T^i_k, the k^{th} observation of process N^i.

Then, for all $\theta \in \Theta$, the log-likelihood of process N^i reads:

$$\ell^i_t(\theta_i) = \log \mu_i + \sum_{k=2}^{N^i(t)} \log\left(\mu_i + (\lambda^{i\star}_{\theta_i}(S^i_k) - \mu_i)\mathrm{e}^{-\beta_i(T^i_k - S^i_k)}\right) - \Lambda^i_{\theta_i}(t)\,, \tag{4.6}$$

where $\Lambda^i_{\theta_i}$ is given by Equation (4.5) and with convention $\log(x) = -\infty$ for $x \leq 0$.

Algorithm 5 in Appendix 4.E presents the iterative computation of the likelihood using Equation (4.6). In particular, the complexity of the computation is $O(N(t) \times d)$. It is then possible to establish the Maximum Likelihood Estimator, which we will refer to as (MLE). These estimators will be denoted by a tilde: $(\tilde{\mu}_i)_i$, $(\tilde{\alpha}_{ij})_{ij}$, $(\tilde{\beta}_i)_i$ and $(\tilde{h}_{ij})_{ij}$.

4.3.4 On identifiability conditions

In the previous section we presented a result on the identifiability of multivariate Hawkes process with inhibition via Theorem 4.3.1. Let us mention that identifiability of parameters μ_i and β_i do not require any assumption. The challenge of the proof lies in identifying parameters α_{ij}, which is achieved thanks to Conditions 1 and 2 of Theorem 4.3.1. Condition 1 allows to control strong

inhibition scenarios by ensuring that each subprocess' intensity is positive sufficiently often, and not only at its own event times, while Condition 2 enables to disentangle the contributions of each subprocess. We believe that these assumptions are only sufficient and could be weakened at the cost of a more intricate analysis.

In this section we will discuss this set of conditions by providing two examples of parameters α_{ij} that allow to apply this result along with one counter-example.

Examples for which the conditions are fulfilled

Let us begin with an example of a two-dimensional process. In this situation, as soon as both processes have an infinite amount of events, Conditions 1 and 2 boil down to finding indexes $k \geq 1$ and $k' \geq 1$ such that $\lambda^1(T_k^{2^-}) > 0$ and $\lambda^2(T_{k'}^{1^-}) > 0$. Indeed, it is then enough to consider $(\tau, \tau_+) = (T_k^2, T^1_{N(T_k^2)+1})$ for $(i,j) = (1,2)$ and $(\tau, \tau_+) = (T_{k'}^1, T^2_{N(T_{k'}^1)+1})$ for $(i,j) = (2,1)$.

Example 1. *Let us assume that N is a two-dimensional Hawkes process with the following matrix for parameters α_{ij}:*

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} = \begin{pmatrix} \alpha_{11} & 0.0 \\ \alpha_{21} & \alpha_{22} \end{pmatrix},$$

such as $\alpha_{11} \leq 0$, $\alpha_{21} \geq 0$ and $\alpha_{22} > 0$.

Both processes contain an infinite number of events as process N^1 can be seen as a one-dimensional Hawkes process and the second one has a lower-bounded intensity $\lambda^2 > \mu^2$. Now, since $\lambda^2(t) > 0$ for all $t \geq 0$, we have $\lambda^2(T_1^{1^-}) > 0$. Then, let us first remark that the event times of process N^1 occur independently of N^2, as $\alpha_{12} = 0$. So, for any event time T_ℓ^1 of N^1, the restart times can be written as if N^1 was a univariate Hawkes process (see Bonnet et al. (2021)):

$$T_{(\ell)}^{1\star} = T_\ell^1 + \beta_1^{-1} \log\left(\frac{\mu_1 - \lambda^{1\star}(T_\ell^1)}{\mu_1}\right) \mathbb{1}_{\{\lambda^{1\star}(T_\ell^1) < 0\}}.$$

For $t > T_{(\ell)}^{1\star}$ small enough, both $\lambda^1(t)$ and $\lambda^2(t)$ are positive, meaning that the next event time can come either from N^1 or from N^2. If we consider an infinite sequence of event times, we will eventually observe an event T_k^2 of N^2 such that $\lambda^1(T_k^{2^-}) > 0$.

This gives a set of Hawkes processes with inhibition that verify the assumptions of Theorem 4.3.1. For higher dimensions, the multiplicity of all possible connections between processes complicates the study of general cases from a theoretical point of view. Example 2 illustrates a case where the conditions are fulfilled by considering identically distributed processes.

Example 2. *Let us consider a d-dimensional Hawkes process, as well as $\mu, \alpha_+, \alpha_-, \beta$ such that for any i and for any $j \neq i$:*

$$\begin{aligned} \mu_i = \mu > 0\,, \quad \alpha_{ii} = \alpha_- \leq 0\,, \\ \beta_i = \beta > 0\,, \quad \alpha_{ij} = \alpha_+ \geq 0\,. \end{aligned}$$

As each process has the same parameters for μ_i and β_i along with the exact same interactions, all processes are identically distributed and so in order to verify Conditions 1 and 2 it is enough to verify them for $i = 1$ and $j = 2$.

Fulfilling both conditions amounts to verifying that, with non-zero probability, we can find $k \geq 1$ such that:

1. $\lambda^1(T_k^{2^-}) > 0$;

2. $T_{(N(T_k^2)+1)}$ *is an event from process* N^1*.*

Furthermore, let us notice that as all processes are cross-exciting, if $\lambda^1(T_k^{2^-}) > 0$ *for a certain* k*, then* $\lambda^1(t) > 0$ *for* $t > T_k^2$ *small enough, so* $T_{(N(T_k^2)+1)}$ *can come from process* N^1*. It remains now to verify that with non zero probability* $\lambda^1(T_k^{2^-}) > 0$ *for a certain* $k \geq 1$*.*

But the opposite would entail that almost surely, for all indexes $k \geq 1$*,* $\lambda^1(T_k^{2^-}) = 0$*. Yet, as all processes are identically distributed, this means that, for all* $j \neq 2$*,* $\lambda^j(T_k^{2^-}) = 0$ *almost surely. It would then follow that, for every* i *and for every* $j \neq i$*, for all* $k > 1$*,* $\lambda^j(T_k^{i^-}) = 0$ *almost surely. Let us consider* $T_k^{i_0}$ *for a fixed* k *and* i_0*. As each process* N^j *for* $j \neq i$ *is then excited with the same parameter* α_+ *and they are identically distributed, then* $\lambda^j(T_k^{i_0})$ *are identically distributed and independent conditionally on history* $\mathcal{F}_{T_k^{i_0}}$*. It follows then that the restart times associated to* $T_k^{i_0}$ *of each process are independent and identically distributed. Consequently, there is a non-zero probability that at least two processes regenerate at roughly the same time before another time of process* N^{i_0} *(as it self-inhibits). Once that for* j_0, j_1*,* λ^{j_0} *and* λ^{j_1}*, are positive, the next event time may come from either process, and so either* $\lambda^{j_1}(T^-_{(N(T_k^{i_0})+1)}) > 0$ *with* $T_{(N(T_k^{i_0})+1)}$ *from process* N^{j_0}*, or the inverse, which contradicts the fact that for all* i, j *and for all* $k > 1$*,* $\lambda^j(T_k^i) = 0$*. So for at least one pair* (i, j) *and one* k*, we must have* $\lambda^j(T_k^{i^-}) > 0$ *and so it has to occur for all pairs, in particular* $(1, 2)$*.*

In the next section we present Example 3 of a specific set of parameters for which both conditions are not necessarily met.

Example where the conditions are not fulfilled

Example 3. *Let us consider the Hawkes process defined by the following parameters:*

$$\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix} = \begin{pmatrix}10^5\\ 0.1\end{pmatrix}, \quad \begin{pmatrix}\alpha_{11} & \alpha_{12}\\ \alpha_{21} & \alpha_{22}\end{pmatrix} = \begin{pmatrix}1.0 & 1.0\\ -10^6 & -10^6\end{pmatrix}, \quad \begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix} = \begin{pmatrix}1.0\\ 10^{-5}\end{pmatrix}.$$

The probability that the first event time $T_{(1)}$ *is from process* N^1 *is* $10^5/(10^5 + 0.1) \approx 1$*. If the first event is indeed from process* N^1*, then process* N^2 *is strongly inhibited and in that case the restart time* $T^{2\star}_{(1)}$ *is equal to*

$$T^{2\star}_{(1)} = T_1 + 10^5 \log(10^7)\,,$$

which is very far from T_1*. As* λ^1 *is lower-bounded by* 10^5*, the next candidate of process* N^1 *is roughly distributed as an exponential random variable with parameter* 10^5 *so the next event time is with high probability of process* N^1*. If this is the case, process* N^2 *is further inhibited, and with probability going exponentially quickly to* 1*, all next event times will come from process* N^1*, preventing us from granting Conditions 1 and 2.*

4.3.5 Recovering the graph of interactions

The aim of this section is to describe methodologies able to estimate non-null interactions between subprocesses, which boils down to detecting parameters such that $\alpha_{ij} \neq 0$. Recovering interactions has an interest, first, in providing a graphical interpretation of the Hawkes model, as it describes which subprocesses are actually connected within the whole process. Moreover, the graph of interactions can also be used as a reduction dimension tool, for instance if we focus on the dynamic of one single subprocess, the activity of which can be impacted by a sub-network

of surrounding processes that we want to identify, as investigated in Bonnet et al. (2022a) for neuronal activity.

Estimating non-null interactions is a challenging topic, which has been particularly studied for linear regression (see for instance Tibshirani (1996)). Regarding Hawkes processes with inhibition, this is even more demanding because of the non-differentiability of the log-likelihood. Therefore, we propose, in the subsequent sections, two post hoc techniques (i.e. after computing the MLE estimator) not related to numerical optimization, and additionally, having the benefit to scale easily to high-dimensional processes.

Thresholding

The first method that will be referred to as (MLE-ε) is obtained by adding a thresholding step to the classic Maximum Likelihood Estimation (MLE). This is similar to the cumulative percentage of total variation approach presented in Principal Component Analysis Joliffe (2002, Section 6.1.1). All absolute estimated values $|\tilde{\alpha}_{ij}|$ are arranged in increasing order $(|\tilde{\alpha}_{(k)}|)_{k\in\{1,\dots,d^2\}}$. We compute then the cumulative sums $s_k = \sum_{l=1}^{k} |\tilde{\alpha}_{(l)}|$ and write $S := s_{d^2}$ the sum of all absolute estimated values. Then all estimations $\tilde{\alpha}_{(k)}$ such that

$$s_k < \varepsilon S\,,$$

are set to zero, for a threshold $\varepsilon \in (0,1)$. Subsequently, all non-null estimations $\tilde{\alpha}_{ij}$ are then re-estimated by maximising the log-likelihood.

The choice of an optimal threshold level ε requires a way of comparing estimations, which is achieved thanks to the goodness-of-fit procedure described in Section 4.3.6, and which is illustrated in Section 4.4.2.

Confidence interval

The second method is applicable when a sample $(N_1,\dots,N_n)$ of n realisations of a multivariate Hawkes process is available. For every i,j, we average all estimations $\tilde{\alpha}_{ij}$ over the realisations $N_1,\dots,N_n$ to obtain $\bar{\alpha}_{ij}$ and then we determine a confidence interval around each estimation at a given confidence level $1-\eta$. Then each estimation for which the confidence interval contains 0 is set to zero. Subsequently, all non-null estimations $\tilde{\alpha}_{ij}$ are re-estimated.

In this work we consider two different confidence intervals.

- (CfE) corresponds to the empirical interval

$$\left[\alpha_{(\lfloor \frac{\eta}{2} n \rfloor)}, \alpha_{(\lceil (1-\frac{\eta}{2}) n \rceil)}\right]\,,$$

  where, $(\alpha_{(k)})_{k\in\{1,\dots,n\}}$ is the sequence of the sorted estimations of α_{ij}, and $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ are respectively the floor and the ceiling functions.

- (CfSt) corresponds to

$$\left[\bar{\alpha}_{ij} - t_{1-\frac{\eta}{2}} s_n, \bar{\alpha}_{ij} + t_{1-\frac{\eta}{2}} s_n\right]\,,$$

  where s_n is the empirical standard deviation of the sample and $t_{1-\frac{\eta}{2}}$ is the quantile of level $1-\frac{\eta}{2}$ of the Student distribution with $n-1$ degrees of freedom. This corresponds to a confidence interval obtained for normally distributed estimators. For Hawkes processes with exclusively exciting interactions, estimations obtained through the MLE procedure are asymptotically normal as proven in Ogata (1978, Theorem 5) and as discussed in Laub

(2014). For processes with inhibiting interactions, asymptotic normality is still an open question but is in all likelihood true. However, one has to be careful when using this estimator, in particular a small number of observations could imply that the asymptotic normality is not achieved. In practice, normality can be tested thanks to a Kolmogorov-Smirnov test.

This method of selection through confidence intervals can be seen as testing the following hypothesis at a confidence level $1-\eta$ for every i, j:

$$\begin{cases} \mathcal{H}_0 : \alpha_{ij} = 0\,, \\ \mathcal{H}_1 : \alpha_{ij} \neq 0\,. \end{cases}$$

We can then compute the corresponding p-value for each test and set to zero all parameters for which the null hypothesis is not rejected. As we test d^2 different hypotheses, it is essential to incorporate multiple testing procedures. For this purpose, we choose the Benjamini-Hochberg method, consisting in adapting the rejection threshold of each p-value. This method enables to control the false discovery rate (FDR). If we denote V the number of rejected true null hypothesis and R the number of rejected true alternative hypotheses, the FDR is defined as

$$FDR = \mathbb{E}\left[\frac{V}{R+V}\right].$$

In other words, we control the expected number of true null hypotheses (i.e. parameter α_{ij} is equal to zero) rejected by our testing method. The B-H procedure considers the ordered p-values $(p_{(k)})_{k\in\{1,\dots,d^2\}}$ and compares each one to the adapted rejection threshold $(1-\eta)\frac{k}{d^2}$. Then, we determine the largest $K \in \{1,\dots,d^2\}$ such that $p_{(K)} < (1-\eta)\frac{K}{d^2}$ and we reject all hypothesis such that $p_{(k)} \leq p_{(K)}$

4.3.6 Goodness-of-fit

As a benefit of our approach, it is possible to perform a goodness-of-fit test for assessing the quality of estimations. This is particularly useful when choosing between several estimations (such as those introduced before), in particular to choose an optimal level of thresholding for the (MLE-ε) method. The closed-form expression of the compensator given in Proposition 4.3.1 enables the use of the Time Change Theorem for inhomogeneous Poisson processes Daley et al. (2003, Proposition 7.4.IV). For any i, the sequence of transformed times $(\Lambda^i(T_k^i))_k$ is a realisation of a homogeneous Poisson process with unit-intensity if and only if $(T_k^i)_k$ is a realisation of a point process with intensity λ^i.

We can then define for any $\theta \in \Theta$ the null hypothesis

$$\mathcal{H}_i : \text{“}(T_k^i)_k \text{ is a realisation of a point process with intensity } \lambda^i_{\theta_i}\text{”}.$$

The hypothesis is then tested via a Kolmogorov-Smirnov test between the empirical distribution $(\Lambda^i_\theta(T^i_{k+1}) - \Lambda^i_\theta(T^i_k))_{k\geq 1}$ and an exponential distribution with parameter 1. We obtain then d different tests with p-values $(p_i)_{i\geq 1}$. Using multiple testing approaches can help in determining correctly estimated processes.

Lastly, we can obtain an additional test by considering the entire sequence of times $(T_{(k)})_{k\geq 1}$ and the total intensity λ. We obtain then the null hypothesis

$$\mathcal{H}_{tot} : \text{“}(T_{(k)})_k \text{ is a realisation of a point process with intensity } \lambda_\theta\text{”},$$

with corresponding p-value p_{tot}. This value is obtained by way of a Kolmogorov-Smirnov test between the empirical distribution of $(\Lambda_\theta(T_{(k+1)}) - \Lambda_\theta(T_{(k)}))_{k\geq 1}$ and an exponential distribution with parameter 1, this time using the total compensator of the process.

In the forthcoming sections, this testing procedure is applied to several realisations of event times, that are independent of the considered estimator. This enables to assess properly the accuracy of estimations, without knowing the true conditional intensities. This is particularly interesting for real-world data.

Let us mention that the goodness-of-fit procedure is not only an assessment of the overall fit between the model and the observations but it provides also a tool to calibrate the threshold for the (MLE-ε) method. Indeed, for each threshold level ε chosen over a grid, we compute all p-values (one for each subprocess, one for p_{tot}) and we choose the value of ε that maximises the mean p-value. For the other two methods, (CfE) and (CfSt), the model selection procedure is described in Section 4.3.5 and the goodness-of-fit is only performed afterwards in order to establish the quality of the estimations.

4.4 Illustration on synthetic datasets

4.4.1 Simulation procedure

In order to assess the performance of the maximum likelihood estimation method, we simulate different data by using Ogata's thinning method (Ogata 1981). This method consists in defining a piecewise constant function λ^+ such that for any $k \geq 1$ and any $t \in [T_{(k)}, T_{(k+1)})$, $\lambda^+(t) \geq \lambda(t)$. For this, we define λ^+ for any $t \in [T_{(k)}, T_{(k+1)})$ as

$$\lambda^+(t) = \sum_{i=1}^{d} \left(\mu_i + \sum_{j=1}^{d} \int_0^t \alpha_{ij}^+ \mathrm{e}^{-\beta_i (T_{(k)} - s)} \, \mathrm{d}N^j(s) \right),$$

which corresponds to considering only the positive interactions.

Four different parameter sets are considered: three sets for 2-dimensional Hawkes processes and a last one for a 10-dimensional process. Table 4.1 presents the parameters used in Dimension 2. All scenarios contain at least one negative interaction ($\alpha_{ij} < 0$). Scenario (1) is a Hawkes process where all parameters are non-null whereas Scenarios (2) and (3) are chosen to study the performance of our methods when estimating null interactions (α_{12} for Scenario (2) and α_{21} for Scenario (3)). All simulations have exactly 5000 event times in total.

Scenario	(1)	(2)	(3)
$\begin{pmatrix}\mu^1\\ \mu^2\end{pmatrix}$	$\begin{pmatrix}0.5\\ 1.0\end{pmatrix}$	$\begin{pmatrix}0.7\\ 1.0\end{pmatrix}$	$\begin{pmatrix}1.2\\ 1.0\end{pmatrix}$
$\begin{pmatrix}\alpha_{11} & \alpha_{12}\\ \alpha_{21} & \alpha_{22}\end{pmatrix}$	$\begin{pmatrix}-1.9 & 3.0\\ 1.2 & 1.5\end{pmatrix}$	$\begin{pmatrix}0.2 & 0.0\\ -0.6 & 1.2\end{pmatrix}$	$\begin{pmatrix}-1.0 & 0.1\\ 0.0 & -0.8\end{pmatrix}$
$\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}$	$\begin{pmatrix}5.0\\ 8.0\end{pmatrix}$	$\begin{pmatrix}3.0\\ 2.0\end{pmatrix}$	$\begin{pmatrix}0.3\\ 0.5\end{pmatrix}$

Table 4.1: Parameters for simulations of two-dimensional Hawkes processes.

In order to carry out the hypothesis testing procedure, we simulate a sample of Hawkes

processes independent from the one used for the estimation. Each testing sample contains as many realisations as the estimation sample. All p-values presented in the chapter correspond to the average obtained over all realisations.

4.4.2 Proposed methods and comparison to existing procedures

The main focus of this chapter is to assess the performance of the maximum likelihood estimator to correctly detect the interacting functions of our processes without ambiguity, estimators are denoted with a tilde: $(\tilde{\mu}_i)_i$, $(\tilde{\alpha}_{ij})_{ij}$, $(\tilde{\beta}_i)_i$ and $(\tilde{h}_{ij})_{ij}$. In this chapter we propose four methods previously introduced in Section 4.3.5:

- (MLE) The estimator obtained by minimising the opposite of the log-likelihood $-\sum_{i=1}^{d} \ell_t^i(\theta)$ (see Equation (4.6)). The log-likelihood is computed via Algorithm 5 and the minimisation is done with the L-BFGS-B method (Byrd et al. 1995).
- (MLE-ε) The estimator obtained by adding a thresholding step to the previous method to determine the non-null estimations. The value of ε is chosen such that it maximises the mean over all p-values obtained.
- (CfE) The estimator whose support is obtained through the empirical confidence intervals.
- (CfSt) The estimator using Student distributed intervals after verification of normality of the estimations.

The latter three methods are specially interesting for Scenarios (2) and (3) of the 2-dimensional processes, and also for the 10-dimensional setting.

Remark. *Another option considered for (MLE-ε) is to use instead the values $|\tilde{\alpha}_{ij}/\tilde{\beta}_i|$ for the thresholding. Numerical results slightly differ between the two methods, with the retained method showing better overall p-values.*

For an informative assessment of the proposed approaches, we compare their performance to estimation methods from the literature. However, up to our knowledge, there is no other parametric estimation methods designed for inhibiting processes, that is for handling negative values of α_{ij}. As a consequence, we chose to include estimation methods developed for exciting processes, that are nonetheless able to produce negative estimations of α_{ij}. This is the case for three popular approaches described below.

1. (Approx) The first one (Lemonnier et al. 2014) is obtained by approaching the compensator $\Lambda^i(t)$ (in each log-likelihood $\ell_t^i(\theta)$) by

$$\int_0^t \lambda^{i\star}(u)\,\mathrm{d}u\,.$$

   In the case where all interactions are positive, this integral is equal to the compensator. The difference is when interactions are negative as this integral takes into account the negative values of the underlying intensity function. The minimisation is done in the same way as for (MLE) using the L-BFGS-B method.

2. The other two methods minimise the least-squares loss approximation defined in Reynaud-Bouret et al. (2014) and Bacry et al. (2020) as:

$$R_t(\theta) = \int_0^t (\lambda_\theta(u))^2\,\mathrm{d}u - \frac{2}{t}\sum_{k=1}^{N(t)} \lambda_\theta^{m_k}(T_{(k)}^-)\,,$$

which is an observable approximation of $\|\lambda_\theta - \lambda\|_t^2 = \int_0^t (\lambda_\theta(u) - \lambda(u))^2 \, \mathrm{d}u$ up to a constant term. In Bacry et al. (2020), all interactions are assumed to be positive, however the implemented version of this method in the package `tick` Bacry et al. (2018) allows to retrieve negative values. For this, we consider two different kernel functions from this implementation:

- (Lst-sq) $h_{ij}(t) = \alpha_{ij}\beta_{ij}e^{-\beta_{ij}t}$, where β_{ij} is fixed beforehand by the practitioner. In practice, we fix $\beta_{ij} = \beta_i$ to be consistent with our model (see Assumption 2). The only solver in the implementation that provides negative values is BFGS, which is limited to work with an ℓ_2-penalty. The grid of values $\{1, 10, \dots, 10^6\}$ is considered for the regularisation constant. To obtain the best estimation for this method, we choose the constant that minimises the relative squared error over all estimated parameters.
- (Grid-lst-sq) $h_{ij}(t) = \sum_{u=1}^{U} \alpha_{ij}^u \beta^u e^{-\beta^u t}$, with $(\beta^u)_u$ a fixed grid of parameters. In our case, we choose $U = d$ and the grid contains each parameter β_i. Intuitively, by applying an ℓ_1-penalty, this method would be able to retrieve the corresponding parameter β_i for each process. However, in practice, the implementation uses BFGS as optimiser and is limited to work with an ℓ_2-penalty. As for (Lst-sq), the regularisation parameter is chosen over the grid of values by minimising the relative squared error.

Results on bivariate Hawkes processes

We generate 25 realisations for each parameter set given in Table 4.1 and we estimate the parameters for each individual simulation. We begin this comparison by competing the proposed methods with both (Approx) and (Lst-sq) which are the two methods with the same kernel functions considered in this chapter. Figure 4.3 displays the relative squared errors for each group of parameters (baselines $(\mu_i)_i$, interaction terms $(\alpha_{ij})_{i,j}$ and delay factors $(\beta_i)_i$) by considering vector norms.

First, we observe that delay factors $(\beta_i)_i$ (last column of Figure 4.3) are similarly estimated by all approaches. Let us recall that (Lst-sq) is not included in the comparison of delay factors: since it requires to provide a value for these parameters (they are not estimated), it was given the true values of $(\beta_i)_i$ as input. An alternative offered by `tick` is to provide a grid of values, but this approach, denoted (Grid-lst-sq), is included in the comparison at the end of the section because of its difference with the exponential model considered here.

Then, regarding the baseline intensities $(\mu_i)_i$ and the interaction factors $(\alpha_{ij})_{ij}$, the proposed methods outperform the two other approaches. In all Scenarios, (MLE-ε), (CfE) and (CfSt) appear to perform almost identically as they retrieve the same supports and from then, the re-estimations are the same. In Scenario (2), all estimation methods perform reasonably well. This can be explained by the weak inhibiting effect of the interaction $1 \to 2$, leaving the intensity almost always positive. The slight difference between (MLE-ε) and the confidence intervals comes from the fact that (MLE-ε) is applied individually to each estimation so for some estimations it does not set any values to zero.

In Scenario (1), the performance of (Approx) and (Lst-sq) is altered, in particular for the $(\tilde{\alpha}_{ij})_{ij}$ estimations, because the inhibiting effect is stronger than in Scenario (2). The major changes appear in Scenario (3), where both (Approx) and (Lst-sq) obtain very high relative errors. More precisely, they fail to explain the interactions between the two processes (see the estimations $(\tilde{\alpha}_{ij})_{ij}$ in the middle column of Figure 4.3), which is compensated by a wrong estimation $(\tilde{\mu}_i)_i$ of baseline intensities. This is not surprising since Scenario (1), and even more Scenario (3), were designed so that the intensity functions are frequently equal to zero, which induces major differences between true and underlying intensities. Since (Approx) and (Lst-sq) are both based on assuming that these two functions are almost equal, the violation of

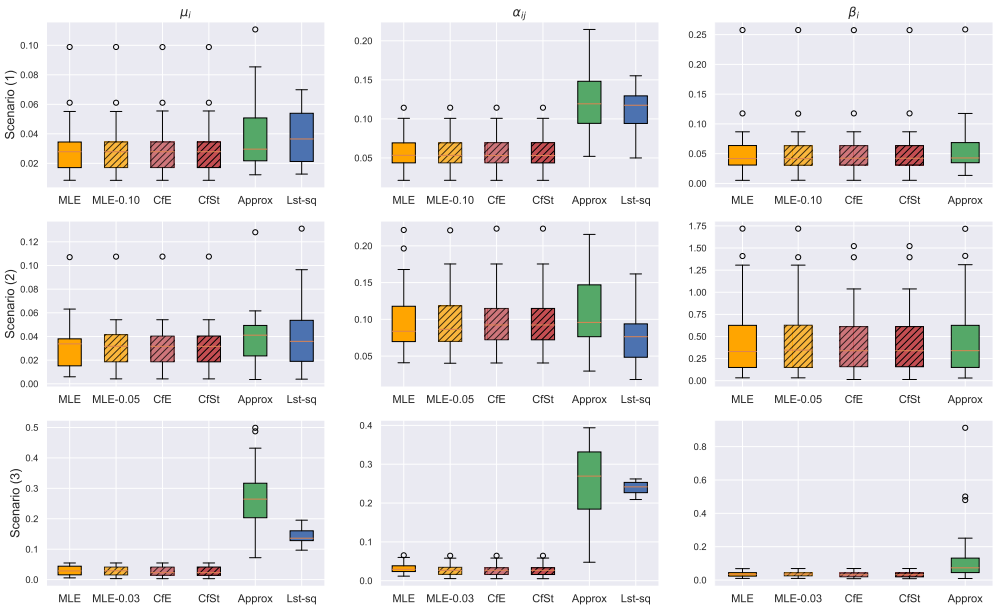

Figure 4.3: Boxplots of the relative squared error for each group of parameters $((\mu_i)_i$, $(\alpha_{i,j})_{i,j}$ and $(\beta_i)_i)$ for 25 realisations of a two-dimensional Hawkes processes. (Lst-sq) does not appear in the last column because it is provided with the true values of $(\beta_i)_i$. The proposed methods are (MLE), (MLE-ε), (CfE) and (CfSt).

this assumption causes large estimation errors. As expected, the proposed methods, which are developed to handle such inhibiting scenarios, provide accurate estimations.

These results are confirmed by the outcomes of the goodness-of-fit test displayed in Table 4.2. It shows indeed the averaged p-values for each scenario using both the true parameters and all four estimations from Figure 4.3 with 25 simulations different from the ones used for estimation. In particular, we can see that our methods obtain high p-values, being very close to those obtained using the true parameters. Table 4.2 also highlights when parameters are incorrectly estimated. For instance, in Scenario (1), (Approx) correctly estimate Process 2 but provides less accurate estimations for Process 1 (the p-value is almost half the one obtained with the true parameters), which is the one characterised by a self-inhibiting behaviour. In addition, at least one of the proposed methods obtains the highest value for p_{tot} in each scenario, which illustrates the ability of these procedures to reconstruct the complete process N. Let us note that the very low p-values obtained by (Approx) and (Lst-sq) for Scenario (3) confirm the ability of the goodness-of-fit procedure to detect when the parameter estimations strongly differ from the true parameters.

Lastly, let us investigate the estimations obtained via (Grid-lst-sq), which can be used in practice as a way to estimate the parameters β_i by providing a grid of possible parameters. Let us mention that both of the previous comparisons (boxplots and p-values) cannot be done here due to the difference in the number of parameters, but we can compare the methods in terms of reconstructions $\tilde{h}_{ij}$ of the interaction functions h_{ij}. For this purpose, we analyse Figure 4.4, which represents the estimated interaction functions $\tilde{h}_{ij}$ for all methods in Scenario (3). Interestingly, we see that (Grid-lst-sq) performs similarly to (Lst-sq), while the latter is fed with all true values $(\beta_i)_i$ for each interaction. However, we see that (Grid-lst-sq) suffers from the same difficulties

p-value	Scenario (1)			Scenario (2)			Scenario (3)		
	p_1	p_2	p_{tot}	p_1	p_2	p_{tot}	p_1	p_2	p_{tot}
True	0.492	0.438	0.430	0.535	0.468	0.479	0.510	0.623	0.338
MLE	0.440	0.442	0.398	0.483	0.461	0.485	0.549	0.638	0.357
MLE-ε CfE CfSt	0.440	0.442	0.398	0.488	0.461	0.491	0.549	0.574	0.327
Approx	0.257	0.442	0.358	0.483	0.452	0.459	**0.0**	**0.007**	**0.0**
Lst-sq	0.154	0.438	0.392	0.534	0.463	0.478	**0.0**	**0.0**	**0.0**

Table 4.2: Average p-values for estimations of two-dimensional Hawkes processes for all scenarios. The values are averaged over 25 simulations. In bold the p-values correspond to a rejected hypothesis at a confidence level of 0.95.

than (Approx) and (Lst-sq), which was expected since it relies on the same unvalid assumption. Let us note that we chose to display the results for Scenario (3) since it highlights the main differences between the compared approaches but the reconstructions for Scenarios (1) and (2) can be found in Appendix 4.F (Figures 4.13 and 4.14).

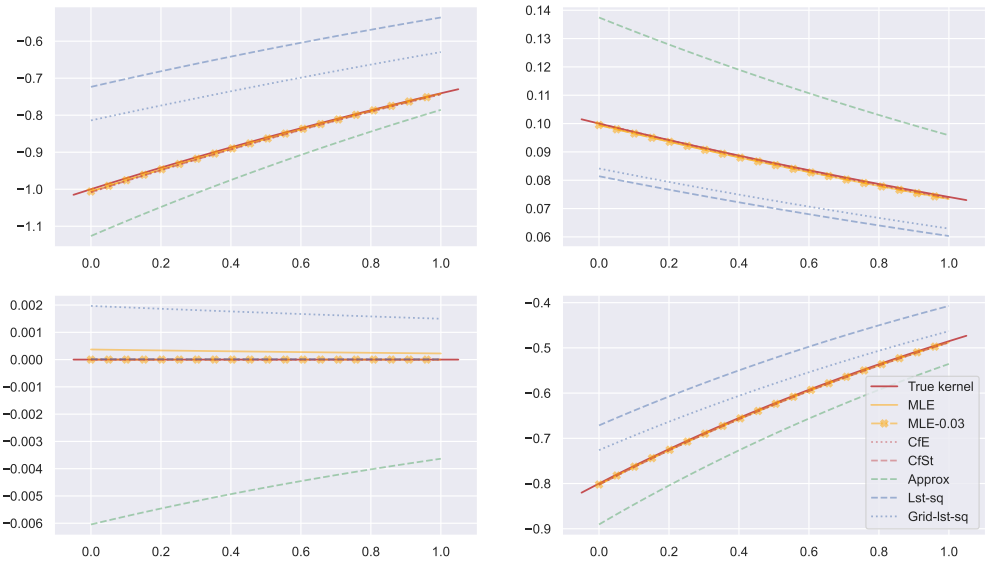

Figure 4.4: Reconstruction of interaction functions h_{ij} for Scenario (3) of two-dimensional Hawkes processes along with all estimated functions $\tilde{h}_{ij}$. The real function is plotted in red and 25 estimations are averaged for each method.

A 10-dimensional Hawkes process

A 10-dimensional Hawkes process is simulated based on a set of parameters corresponding to the quantities $(\text{sign}(\alpha_{ij})\|h_{ij}\|_1)_{ij} = (\alpha_{ij}/\beta_i)_{ij}$ displayed in Figure 4.5. The chosen parameters fulfil

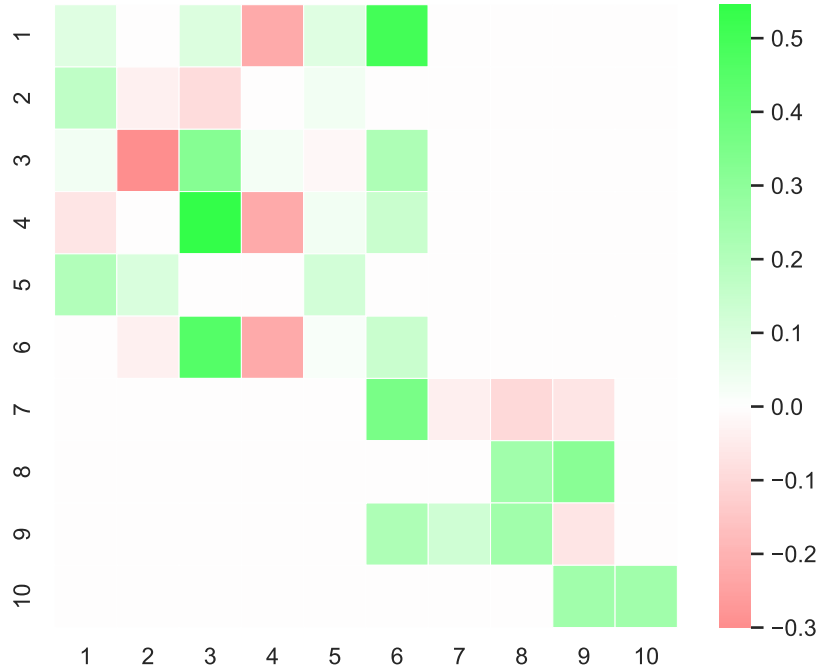

Figure 4.5: Heatmap of real parameters $(\text{sign}(\alpha_{ij})\|h_{ij}\|_1)_{ij} = (\alpha_{ij}/\beta_i)_{ij}$ for the 10-dimensional simulation.

the existence condition $\|\rho(S^+)\| < 1$.

The corresponding estimations $\tilde{\alpha}_{ij}$ and $\tilde{\beta}_i$ are averaged over 25 realisations and displayed in Figure 4.6. The heatmap representation is convenient for high-dimensional processes as it allows us to see whether the signs of each interaction are well-estimated and whether the null-interactions are correctly detected.

In this example we decided to keep only (Approx) and (Lst-sq) as comparison methods as these are the ones with the same parameterisation for the kernel functions. Among the four methods considered, (Approx) is the only one that wrongly estimates the sign of some interactions, represented by the black boxes in the second row matrix. (MLE) and (Lst-sq) correctly retrieve the sign of each interactions but are unable to detect the null interactions: this is not surprising since (MLE) does not contain a regularisation step and the (Lst-sq) estimator is implemented with a ℓ_2-penalty which does not provide a sparse solution. On the one hand, (MLE-ε) is in this case quite conservative by setting a single value equal to 0 compared to (MLE). On the other hand, both confidence intervals methods improve the number of interactions whose sign is correctly estimated. Interestingly, we see that (CfE) sets more values equal to zero than it should (purple boxes) whereas (CfSt) denotes the opposite effect by not detecting null interactions (orange boxes). Overall, (CfSt) obtains the best results in terms of support recovery and sign estimations by committing only two errors. Table 4.3 summarises the p-values for each hypothesis as described in Section 4.3.6. All of the proposed methods obtain overall better p-values with no particularly low values, which is not the case for (Approx) (see p_4 and p_{tot}) and for (Lst-sq) (see p_8 and p_{10}). Although the p-values all exceed 5%, they remain substantially smaller than those obtained with the alternative methods.

Finally, we compare the relative squared errors for each group of parameters (see Figure 4.8).

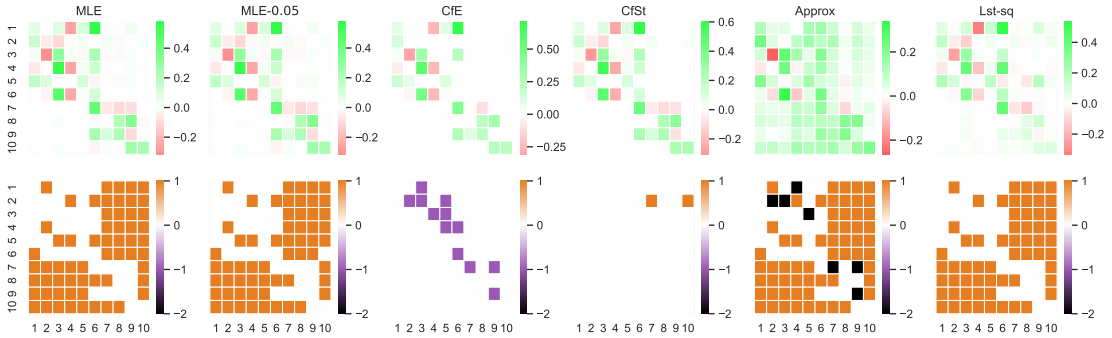

Figure 4.6: Top row corresponds to the heatmap for each estimation method. Bottom row corresponds to errors made with respect to real parameters from Figure 4.5. A value of 1 (orange) shows an undetected 0 (non-null estimation for $\alpha_{ij} = 0$), a value of -1 (purple) shows a non-null value set to 0 and a value of -2 (black) shows a non-null value whose sign is wrongly estimated. Each of the compared approaches is described in Section 4.4.2.

p-value	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}	p_{tot}
True	0.437	0.608	0.435	0.517	0.534	0.45	0.43	0.47	0.533	0.509	0.464
MLE	0.451	0.619	0.399	0.466	0.506	0.464	0.424	0.386	0.45	0.483	0.434
MLE-0.05	0.454	0.622	0.392	0.466	0.505	0.427	0.418	0.392	0.495	0.482	0.432
CfE	0.424	0.532	0.393	0.528	0.532	0.301	0.444	0.427	0.516	0.505	0.439
CfSt	0.452	0.633	0.375	0.474	0.527	0.462	0.431	0.422	0.488	0.493	0.465
Approx	0.411	0.376	0.475	0.077	0.485	0.411	0.3	0.384	0.285	0.436	0.085
Lst-sq	0.422	0.63	0.344	0.456	0.416	0.439	0.411	0.096	0.579	0.157	0.423

Table 4.3: p-values for estimations of a ten-dimensional Hawkes process. The values are averaged over 25 simulations. p_{tot} corresponds to testing whether the estimated intensity function corresponds to a multivariate Hawkes process N as defined in Section 4.3.6.

Similarly to the two-dimensional case, all proposed methods perform significantly better than alternative approaches regarding the estimation of all parameters. In addition, it can be noticed that inaccurate estimations of the parameters α_{ij} tend to deteriorate the estimations of μ_i, which suggests an effect of compensation between these parameters. Regarding the estimator (CfSt), which shows the best averaged performance, it can be remarked that it also exhibits a large variance, in particular when estimating β_i.

Figure 4.7 illustrates for (CfST) the ordered p-values for hypothesis $\mathcal{H}_0 : \alpha_{ij} = 0$.

This can be explained by this estimator providing a very sparse solution (as seen in Figure 4.6) and therefore taking into account less observations for estimating the coefficients β_i.

An important question for any inference method, especially in a high-dimensional setting, is its computational cost. Table 4.4 shows the average estimation time (over 25 realisations), all times being total estimation time. More precisely, for our 3 model selection methods (MLE-ε), (CfE) and (CfSt), it takes into account the total times, including the first (MLE) estimation in addition to the re-estimation over the support.

Although the difference between (Lst-sq) and all other methods is substantial, let us recall

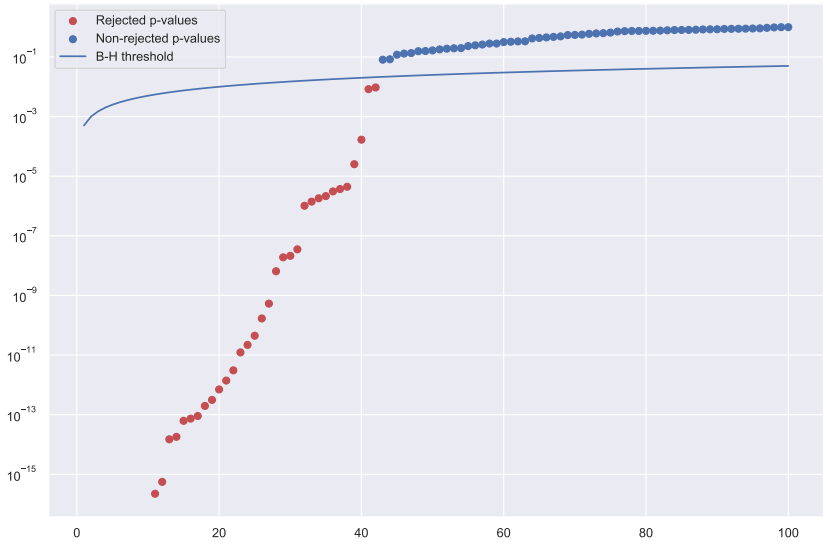

Figure 4.7: Ordered p-values corresponding to support estimation of method (CfSt). Red points correspond to rejected null hypothesis $\mathcal{H}_0 : \alpha_{ij} = 0$ and blue points to non-rejected ones.

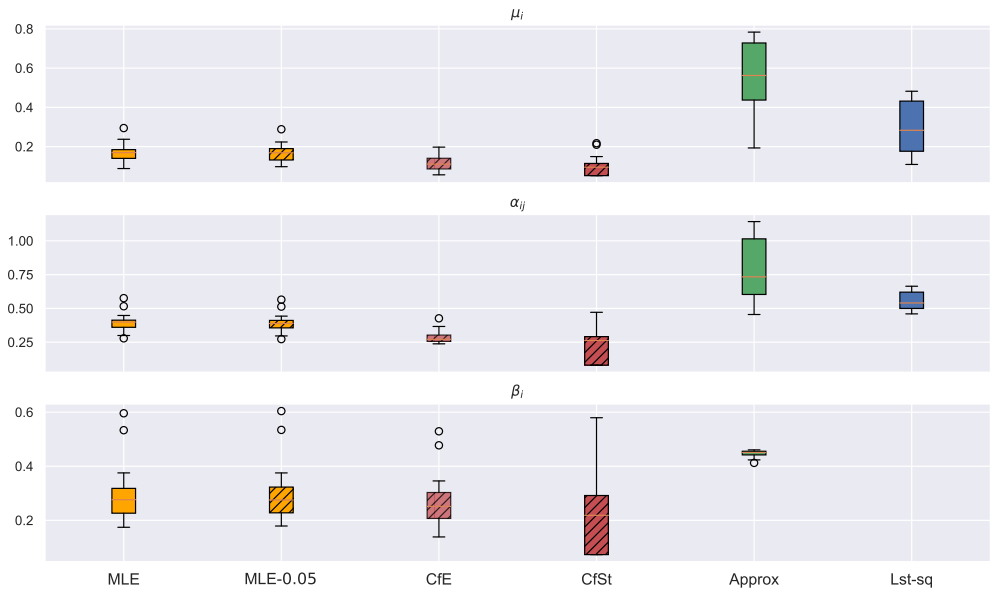

Figure 4.8: Boxplots of the relative squared error for each group of parameters $((\mu_i)_i$, $(\alpha_{ij})_{i,j}$ and $(\beta_i)_i)$ for 25 realisations of a ten-dimensional Hawkes processes. (Lst-sq) does not appear in the last row because it is provided with the true values of $(\beta_i)_i$. The proposed methods are (MLE), (MLE-ε), (CfE) and (CfSt).

that (Lst-sq) requires to be provided with parameters β_i, which offers two numerical advantages: it does not need to optimise for parameters β_i which are the more difficult parameters to estimate and it includes a pre-computation step (of exponential terms) that accelerates all internal

	MLE	MLE-ε	CfE	CfSt	Approx	Lst-sq	Grid-lst-sq
Computing time	87.6	178.2	147.2	152.2	51.4	1.32	47.7

Table 4.4: Average computing time in seconds for all estimation methods, averaged over 25 realisations of a ten-dimensional Hawkes process. The time shown for (MLE-ε) is the total time of estimation for 7 different values of ε. Similarly for `tick` methods, the time is for 7 different levels of penalisation (as done in the estimations of Section 4.4.2). For the (Grid-lst-sq), we provided a search grid for β_i that contains 12 values, including the true one.

computations. In order to provide a fairer comparison, we include the computing time of the (Grid-lst-sq) method which can be considered as an alternative for estimating these parameters when given a search grid for β_i (here with 12 values, including the true parameters). As expected, the computational cost of the MLE method is higher than the alternative approaches (Approx) and (Grid-lst-sq), designed for the linear model. It remains nevertheless that (Lst-sq) and (Grid-lst-sq) are both implemented in a compiled language (C++), which is always faster than an interpreted language such as that used in the proposed package (Python). However, all computation times remain quite reasonable, even for the methods that include a selection model step.

4.4.3 Robustness on misspecified models

In this section, we address the question of the robustness of our estimator regarding the misspecification of the kernel function. More precisely, we generate a two-dimensional Hawkes process with power-law kernels, which are commonly used in the literature (Mishra et al. 2016; Ogata 1988) as an alternative to the exponential kernel for modelling a slower convergence to zero. For all i, j we define the power-law kernel as:

$$h_{ij}(t) = \frac{\alpha_{ij}\beta_{ij}}{(1+\beta_{ij}t)^{1+\gamma}}\,,$$

with $\beta_{ij} > 0$, $\gamma > 0$ and $\alpha_{ij} \in \mathbb{R}$ in order to allow inhibition effects. Let us remark that in the general case each kernel function could be given a different parameter γ_{ij} but here we fix the same parameter for all interactions, which is a similar condition as Assumption 2.

We propose to investigate different scenarios in order to model different behaviours. In all cases, we set the values of α_{ij} in order to have both excitation and inhibition effects and the values of μ_i as follows:

$$\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix} = \begin{pmatrix}1.0\\ 1.0\end{pmatrix}\,, \quad \begin{pmatrix}\alpha_{11} & \alpha_{12}\\ \alpha_{21} & \alpha_{22}\end{pmatrix} = \begin{pmatrix}0.1 & 1.5\\ 1.0 & -0.5\end{pmatrix}\,.$$

- Scenarios γ : we set

$$\begin{pmatrix}\beta_{11} & \beta_{12}\\ \beta_{21} & \beta_{22}\end{pmatrix} = \begin{pmatrix}1.0 & 1.1\\ 1.2 & 1.0\end{pmatrix}\,,$$

  all values being similar in order to be close to Assumption 2. Then we study the effect of parameter γ which controls how the kernel functions decrease to zero. We set

$$\gamma \in \{2.0, 4.0, 6.0, 8.0\}\,.$$

- Scenario β : we keep the same values of μ_i, α_{ij} as in Scenarios γ, we fix $\gamma = 4.0$ and we choose very different values of $\beta_{i,j}$:

$$\begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} = \begin{pmatrix} 1.0 & 2.0 \\ 0.1 & 1.0 \end{pmatrix},$$

  so that Assumption 2 is violated and therefore the intensity function of process N^2 is frequently non-monotonous between two event times.

We expect that our estimator should adapt better to Scenarios γ (in particular large values of γ would correspond to a fast convergence to zero) than to Scenario β.

Figure 4.9 represents the heatmap for $\text{sign}(\alpha_{ij})\|h_{ij}\|_1 = \alpha_{ij}/\gamma$ and $\text{sign}(\tilde{\alpha}_{ij})\|\tilde{h}_{ij}\|_1 = \tilde{\alpha}_{ij}/\tilde{\beta}_i$ as well as whether the type of each interaction is correctly estimated (excitation or inhibition). We first notice that in most cases, our method is robust enough to differentiate between exciting and inhibiting interactions, the only errors concerning parameter α_{11} that is close to zero. For Scenarios γ, we can observe as expected that the bigger differences are obtained for smaller values of γ. Although the signs of the interactions are correctly estimated in Scenario β, we can observe substantial errors regarding the estimations of both α_{21} and α_{22}.

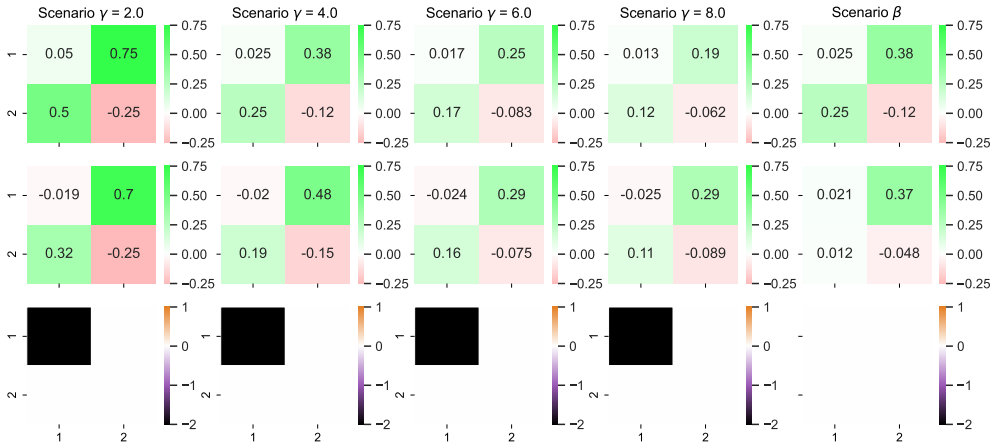

Figure 4.9: Top row corresponds to the heatmap $\text{sign}(\alpha_{ij})\|h_{ij}\|_1 = \alpha_{ij}/\gamma$. Middle row corresponds to the estimated heatmap $\text{sign}(\tilde{\alpha}_{ij})\|\tilde{h}_{ij}\|_1 = \tilde{\alpha}_{ij}/\tilde{\beta}_i$. Bottom row corresponds to errors when estimating the nature of the interaction. A black box represents a value α_{ij} whose sign is wrongly estimated.

Table 4.5 shows the average p-values associated with the goodness-of-fit measure presented in Section 4.3.6. It confirms that for all Scenarios γ, the p-values are smaller for smaller values of γ, all p-values remaining greater than 5%. However, the p-values associated with Scenario β are always equal to zero, which means that the goodness-of-fit is able to detect that the interactions are not correctly estimated.

To conclude, if the true model is not too far from an exponential model and Assumption 2 holds, our procedure can adapt and provide reasonable estimations. If not, our estimator cannot adjust but we are able to detect incorrect estimations thanks to the goodness-of-fit procedure.

		p_1	p_2	p_{tot}
Scenario γ	$\gamma = 2.0$	0.140	0.275	0.068
	$\gamma = 4.0$	0.484	0.536	0.482
	$\gamma = 6.0$	0.381	0.503	0.408
	$\gamma = 8.0$	0.408	0.506	0.330
Scenario β		**0.0**	**0.0**	**0.0**

Table 4.5: Average p-values for estimations of two-dimensional Hawkes processes for all scenarios in the misspecified power-law model. The values are averaged over 25 simulations. In bold the p-values correspond to a rejected hypothesis at a confidence level of 0.95.

4.5 Application on neuronal data

4.5.1 Preprocessing and data description

In this section we present the results obtained by our estimation method applied to a collection of 10 trials consisting in the measurement of spike trains of 223 neurons from the lumbar spinal of a red-eared turtle. This data are first presented in Petersen et al. (2016) and then also analysed in Bonnet et al. (2022a) to study how the activity of a group of neurons impacts the membrane potential's dynamic of another neuron. In particular, recovering the connectivity graph allows to isolate a subnetwork which activity impacts the dynamics of one given neuron. Events were registered for 40 seconds and in order to take into account eventual stationarity we only consider the events that took place on the interval $[11, 24]$ (see Bonnet et al. (2022a) for further details). Among all trials, each neuron recording contains between 54 and 4621 event times. Furthermore, we divide our samples in a training set consisting on all events in half the interval $[11, 17.5]$ and a test set consisting on the remaining window $[17.5, 24]$, in particular each neuron has at least 15 event times in each set. The training sets are used for obtaining the estimations and the test sets for performing the goodness-of-fit tests.

4.5.2 Resampling

As only ten realisations are available, this can obviously limit the performance of both confidence intervals methods. In order to counter this problem, we perform a resampling method obtained as follows:

1. We sample 3 realisations at random (N_1, N_2, N_3), without replacement and by taking the order into account. From now on, we consider that each realisation takes place in the time interval $[0, 6.5]$ (instead of $[11, 17.5]$).

2. We cumulate all 3 realisations by considering that process N_1 takes place in $[0, 6.5]$ then process N_2 in $[6.5, 13]$ and finally N_3 in $[13, 19.5]$. This creates a single realisation N in the interval $[0, 19.5]$.

This approach is proposed in Reynaud-Bouret et al. (2014, Section 3.4). In our case, we repeat this process 20 times to obtain another sample of realisations. These new sample will be used for both the (MLE) method (presented as (resampled-MLE)) and for both (CfE) and (CfSt).

4.5.3 Goodness-of-fit and multiple testing procedure

Since we are dealing with a high-dimensional setting, it is crucial to account for a multiple testing correction which is performed through the Benjamini-Hochberg procedure as described in Section 4.3.5. In this case, the adapted rejection threshold corresponds to $\frac{0.05k}{d+1}$ and represented in Figure 4.10 by a blue line. This is particularly useful in order to determine the best value of ε for (MLE-ε) as we do not have prior knowledge regarding the sparsity of the neuronal connections.

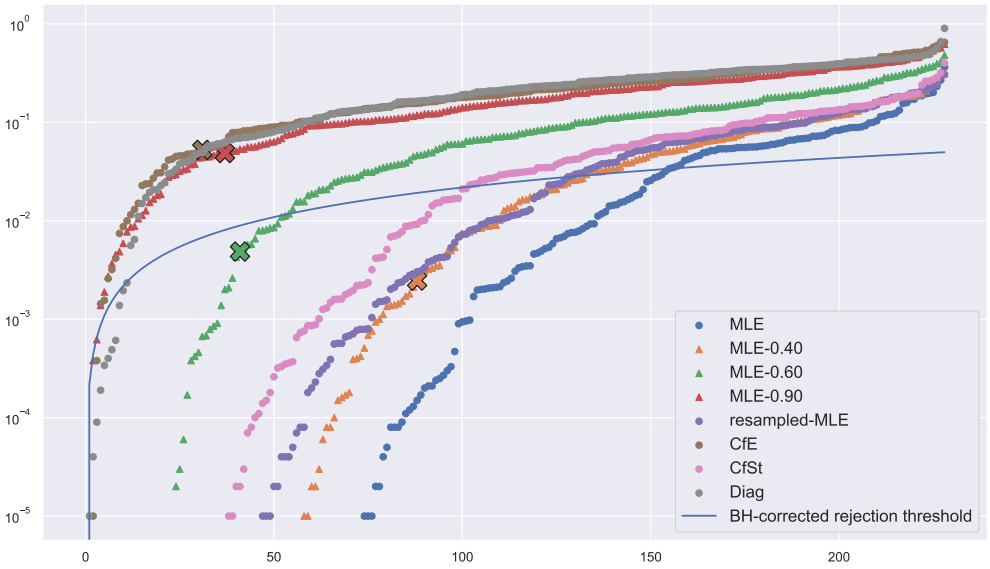

Figure 4.10: Ordered p-values for all hypothesis tests $\mathcal{H}_i$ and $\mathcal{H}_{tot}$. p_{tot} appears as a cross for each model (if there appears no cross for one given method, it means that the corresponding p_{tot} is equal to zero). The blue curve corresponds to the adapted rejection threshold from the B-H procedure, so all tests whose p-value are under the line are rejected.

Figure 4.10 shows the ordered p-values for each hypothesis $\mathcal{H}_i$ along with hypothesis $\mathcal{H}_{tot}$ displayed with a bold cross, all total p-values also being summarised in Table 4.6.

	MLE	MLE-0.40	MLE-0.60	MLE-0.90	resampled-MLE	CfE	CfSt	Diag
p_{tot}	0.0	0.002	0.004	**0.048**	0.0	**0.053**	0.0	0.0

Table 4.6: Values of p_{tot} for each estimation method for the neuronal dataset. In bold appear the p-values above the rejection threshold after Benjamini-Hochberg procedure.

We first note that for most methods, the p-values p_{tot} associated with the $\mathcal{H}_{tot}$ hypothesis are either equal to 0 or under the rejection threshold. In particular, it is the case for the MLE-ε approach for small values of ε (i.e. weak sparsity scenarios) but as we increase the threshold ε, the p-values appear to increase, with the best estimation being achieved for $\varepsilon = 0.90$.

This suggests that the simpler the model the better p-values we obtain so we decided to include another approach, named "Diag", consisting in setting all $\alpha_{ij} = 0$ for $i \neq j$. This corresponds to a model where there exists no interaction between neurons and we keep only self-interactions: in other words, each neuron is seen as a univariate Hawkes process with three

parameters $(\mu_i, \alpha_{ii}, \beta_i)$. Although most hypotheses $\mathcal{H}_i$ are not rejected by the method, the total p-value p_{tot} is zero which suggests that although such a model could explain each dimension individually, it is unable to explain the neurons' interactions as a whole interconnected process. Although the total p-values are generally quite low, they remain above the rejection threshold for (MLE-0.90) and (CfE). Let us notice that the high value of the best threshold (0.90) is consistent with the (CfE) estimator that provides an even sparser solution with 4.26% non-null interactions.

Let us also recall that the (CfSt) method relies on the assumption that the MLE estimator is asymptotically normal. Here we have a low number of repetitions (10 trials) in a high-dimensional setting where some neurons are rarely observed which could explain why this estimator does not perform well. Moreover, in this context, a Kolmogorov-Smirnov test of normality is likely to provide high p-values for such small-sized samples even for non-normal distributions.

Finally, the model that best describes the complete process N is (CfE) with the highest value for p_{tot} and with almost all hypotheses, including $\mathcal{H}_{tot}$, not rejected. This suggests indeed that the estimations provided by (CfE) are the best fit for explaining the entire process as well as each individual subprocess.

4.5.4 Estimation results

Figure 4.11 illustrates the obtained estimation for all parameters for the (CfE) method. Let us recall that the estimation for (CfE) is obtained by using the resampled trials: the support is determined by using the empirical quantiles confidence intervals after an estimation through (MLE) and then all parameters are re-estimated over each trial. A single estimation is obtained by averaging over all trials.

Although the heatmap matrix corresponding to $(\text{sign}(\tilde{\alpha}_{ij}))_{ij}$ contains only 4.26% of non-null entries, there remain many significant interactions. Interestingly, among them we detect all types of interactions: mutual excitation, mutual inhibition, self-excitation, self-inhibition. This supports the relevance of carefully accounting for inhibition when developing inference procedures.

We also notice that the diagonal contains mostly non-null entries (all but 6), which highlights the major effect of self-interactions, among which some are negative and some are positive. Although it is possible that different neurons actually show different patterns, some being self-exciting and other self-inhibiting, there exists another hypothesis. We might indeed observe a combination of effects from which we cannot distinguish: on the one hand, a self-exciting behaviour and on the other hand, a refractory period following a spike during which a neuron cannot spike again. This could also explain why the order of magnitude of the β_i estimations, which describe the duration until an effect vanishes, is different from a neuron to another. It would be of great interest to propose another modelling that could account for both effects and thereby helping us to provide additional information to support or refute this hypothesis.

Another striking phenomenon is the behaviour of neuron 13, which seems to interact with many other neurons: it contains indeed 69% non-null receiving interactions (row) and 57% non-null giving interactions (column). Further analysis shows that this neuron spikes only in one out of ten trials so that it could indicate an inaccurate estimation. However, the p-value associated with this neuron's subprocess is not rejected by our goodness-of-fit procedure, which suggests that the corresponding estimation is actually accurate. Therefore, this neuron could either play central role among the whole network or be connected to an unobserved neuron with a central role. On the opposite side, some neurons exhibit only a few connections, in particular there is one neuron that only receives interactions without giving, while another one gives without receiving.

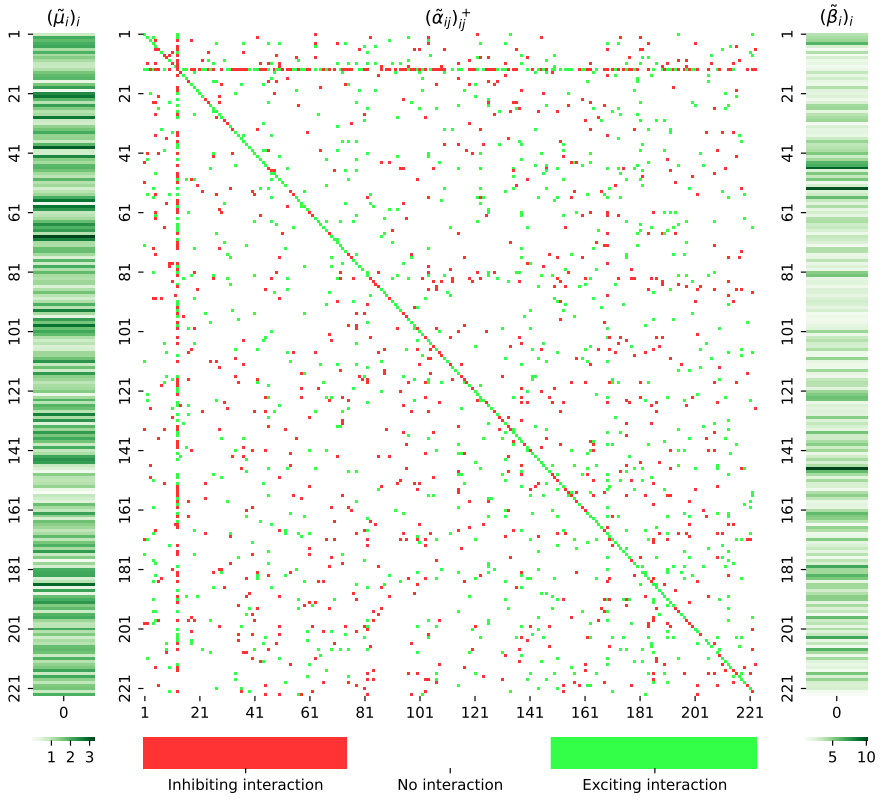

Figure 4.11: Heatmap $(\text{sign}(\alpha_{ij}))_{ij}$ of (CfE) estimation on 223 neurons.

4.6 Discussion

In this chapter, we proposed a methodology for estimating the parameters of multivariate exponential Hawkes processes with both exciting and inhibiting effects. Our first contribution was to provide a few sufficient conditions to ensure the identifiability of a commonly used model. Then we developed and implemented a maximum likelihood estimator combined with a variable selection procedure that enables to detect the significant interactions inside the whole process. While our framework is more general than the usual linear Hawkes model, there remain two main limitations, the first one being the exponential distribution of the kernels, the second one the assumption that the delay factors β_{ij} only depend on the receiving process N^i. If it is essential to assume a parametric form for the kernel functions in order to maintain the concepts of our approach, it would be of great interest to consider some extensions to account for potential combined effects, as already mentioned in Section 4.5.4. It would be notably relevant to include multi-scale effects or to consider a potential lag between an event time and its actual impact. Regarding the assumption on the delay factors, while it is quite standard, it could be a limitation of our approach when considering heterogeneous phenomena.

Going over this assumption would lead us to explore numerical integration methods and would considerably increase the computational time of the estimation procedure. This is obviously detrimental since, in practice, time sequences are increasingly abundant and large. On the other hand, improving the computational effectiveness of estimation procedures for Hawkes processes

is a current direction of research (Bompaire et al. 2018).

Our work focuses on the computational aspects of both maximum likelihood estimation and variable selection. It is of natural interest to provide further theoretical study of the asymptotic behaviour of our estimator, as done for exciting Hawkes processes (Guo et al. 2018). This work is currently under investigation.

Let us also highlight that, because of the physical constraints of the experiment, only a fraction of the neuronal network is observed, which raises the question of interpretability of the estimated interactions. Indeed, the latter do not take into account the interactions with neurons that are outside the observed network. Very recent results tackle the consistency of estimated interactions in a partially observed network (Reynaud-Bouret et al. 2021). A necessary condition to recover interactions in the subnetwork requires in particular to have a large number of interactions within the full network. Regarding the neuronal application, it could be of great interest to further investigate the interpretability of the inferred interactions and connectivity graph in light of the aforementioned work.

4.A Proof of Lemma 4.3.1

In order to prove Lemma 4.3.1, let us first state a preliminary result.

Lemma 4.A.1. *If Assumption 2 is granted, then for each $i \in \{1, \dots, d\}$ and any $k \geq 1$:*

$$\begin{aligned} &\forall t \in [T_{(k)}, T_{(k+1)}), \\ &\lambda^{i\star}(t) = \mu_i + \left(\lambda^{i\star}(T_{(k)}) - \mu_i\right) \mathrm{e}^{-\beta_i (t - T_{(k)})} . \end{aligned}$$

Proof. Let $i \in \{1, \dots, d\}$. For any $k \geq 1$, the underlying intensity function $\lambda^{i\star}$ in the interval $[T_{(k)}, T_{(k+1)})$ can be written:

$$\lambda^{i\star}(t) = \mu_i + \sum_{j=1}^{d} \sum_{\ell=1}^{N^j(t)} \alpha_{ij} \mathrm{e}^{-\beta_{ij}(t - T_\ell^j)} .$$

This function is differentiable in the open interval $(T_{(k)}, T_{(k+1)})$ and we obtain:

$$(\lambda^{i\star})'(t) = -\sum_{j=1}^{d} \beta_{ij} \sum_{\ell=1}^{N^j(t)} \alpha_{ij} \mathrm{e}^{-\beta_{ij}(t - T_\ell^j)} .$$

By using Assumption 2 that for all $j \in \{1, \dots, d\}$, $\beta_{ij} = \beta_i \in \mathbb{R}_+^*$, we obtain the following differential equation:

$$(\lambda^{i\star})'(t) = -\beta_i \left(\lambda^{i\star}(t) - \mu_i\right) ,$$

which by solving on the interval gives:

$$\lambda^{i\star}(t) = \mu_i + \left(\lambda^{i\star}(T_{(k)}) - \mu_i\right) \mathrm{e}^{-\beta_i (t - T_{(k)})} .$$

□

Proof of Lemma 4.3.1. Let $i \in \{1, \dots, d\}$ and $k \geq 1$. By Lemma 4.A.1, $\forall t \in [T_{(k)}, T_{(k+1)})$:

$$\lambda^{i\star}(t) = \mu_i + \left(\lambda^{i\star}(T_{(k)}) - \mu_i\right) \mathrm{e}^{-\beta_i (t - T_{(k)})} .$$

In particular, the derivative of the underlying intensity function is of opposite sign as $(\lambda^{i\star}(T_{(k)})-\mu_i)$. Let us distinguish two cases, referring to Figure 4.12 for a better understanding:

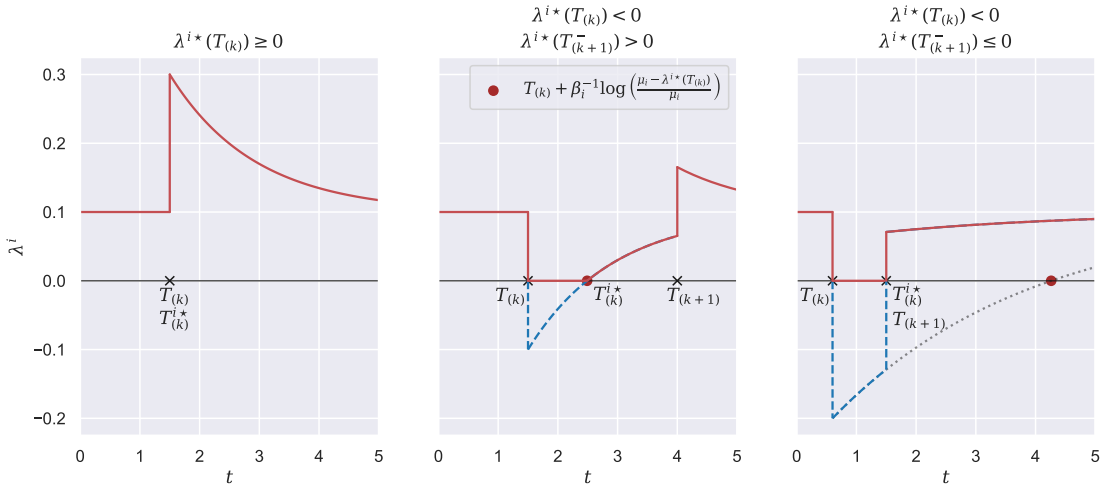

Figure 4.12: Illustration of three possible scenarios for restart times $T_{(k)}^{i\star}$ depending on the sign of $\lambda^{i\star}(T_{(k)})$ and $\lambda^{i\star}(T_{(k+1)}^-) = \lim_{t\to T_{(k+1)}^-} \lambda^{i\star}(t)$. The dotted line in the last scenario shows the equation $\mu_i + (\lambda^{i\star}(T_{(k)}) - \mu_i)\mathrm{e}^{-\beta_i(t-T_{(k)})} = 0$ and the term $T_{(k)} + \beta_i^{-1}\log\left(\frac{\mu_i - \lambda^{i\star}(T_{(k)})}{\mu_i}\right)$ as its only root.

- If $\lambda^{i\star}(T_{(k)}) \geq 0$, then,

$$\begin{aligned} T_{(k)}^{i\star} = T_{(k)} &= \min\left(T_{(k)}, T_{(k+1)}\right) \\ &= \min\left(t_k^\star, T_{(k+1)}\right). \end{aligned}$$

  If $(\lambda^{i\star}(T_{(k)})-\mu_i) \geq 0$, then $\lambda^{i\star}$ is decreasing and lower-bounded by μ_i. If $(\lambda^{i\star}(T_{(k)})-\mu_i) < 0$ then $\lambda^{i\star}$ is increasing and lower-bounded by zero. In both cases, for any $t \in (T_{(k)}^{i\star}, T_{(k+1)})$, $\lambda^{i\star}(t) > 0$ and then $\lambda^i(t) = \lambda^{i\star}(t)$.

- If $\lambda^{i\star}(T_{(k)}) < 0$, then $\left(\lambda^{i\star}\left(T_{(k)}\right) - \mu_i\right) < 0$ so $\lambda^{i\star}$ is strictly increasing and by continuity and by Lemma 4.A.1, there exists a unique $t^\star > T_{(k)}$ such that $\mu_i + \left(\lambda^{i\star}(T_{(k)})\right) - \mu_i)\mathrm{e}^{-\beta_i(t^\star - T_{(k)})} = 0$.

  We obtain:

$$t^\star = T_{(k)} + \beta_i^{-1}\log\left(\frac{\mu_i - \lambda^{i\star}(T_{(k)})}{\mu_i}\right).$$

  By denoting $\lambda^{i\star}(T_{(k+1)}^-) := \lim_{t\to T_{(k+1)}^-} \lambda^{i\star}(t)$:

  - If $\lambda^{i\star}(T_{(k+1)}^-) > 0$, then $t^\star < T_{(k+1)}$ by strict increasingness and so by definition $T_{(k)}^{i\star} = t^\star = t_k^\star$. Lastly, for any $t \in (T_{(k)}, T_{(k+1)})$, if $t \in (T_{(k)}, T_{(k)}^{i\star}]$, $\lambda^{i\star}(t) \leq 0$ and then $\lambda^i(t) = 0$, while if $t \in (T_{(k)}^{i\star}, T_{(k+1)})$, $\lambda^{i\star}(t) > 0$ and then $\lambda^i(t) = \lambda^{i\star}(t)$.

– If $\lambda^{i\star}(T_{(k+1)}^-) \leq 0$, then by strict increasingness $t^\star > T_{(k+1)}$ and so $T_{(k)}^{i\star} = T_{(k+1)}$. In this case, for all $t \in (T_{(k)}, T_{(k+1)})$, $\lambda^{i\star}(t) < 0$ so $\lambda^i(t) = 0$. Moreover, $(T_{(k)}^{i\star}, T_{(k+1)}) = \emptyset$ so we never have $\lambda^i(t) = \lambda^{i\star}(t)$.

Combining all scenarios achieves the proof. □

4.B Proof of Proposition 4.3.1

Proof. For each $i \in \{1, \dots, d\}$ and $\forall t \geq 0$, with convention $T_{(0)} = 0$ and $T_{(0)}^{i\star} = 0$:

$$\begin{aligned}\Lambda^i(t) = \int_0^t \lambda^i(u)\,\mathrm{d}u &= \sum_{k=0}^{N(t)} \int_{T_{(k)}}^{T_{(k+1)}} \lambda^i(u)\mathbb{1}_{u\leq t}\,\mathrm{d}u \\ &= \sum_{k=0}^{N(t)} \int_{T_{(k)}^{i\star}}^{T_{(k+1)}} \lambda^{i\star}(u)\mathbb{1}_{u\leq t}\,\mathrm{d}u\,,\end{aligned}$$

where the last equation comes from Lemma 4.3.1. Then, for $k = 0$:

$$\begin{aligned}\int_{T_{(0)}^{i\star}}^{T_{(1)}} \lambda^{i\star}(u)\mathbb{1}_{u\leq t}\,\mathrm{d}u &= \int_{T_{(0)}^{i\star}}^{T_{(1)}} \mu^i\mathbb{1}_{u\leq t}\,\mathrm{d}u \\ &= \mu_i \min(t, T_{(1)})\,,\end{aligned}$$

and for every $k \in \{1, \dots, N(t)\}$, by Lemma 4.A.1:

$$\begin{aligned}&\int_{T_{(k)}^{i\star}}^{T_{(k+1)}} \lambda^{i\star}(u)\mathbb{1}_{u\leq t}\,\mathrm{d}u \\ =&\int_{T_{(k)}^{i\star}}^{T_{(k+1)}} \left[\mu_i + \left(\lambda^{i\star}(T_{(k)}) - \mu_i\right)\mathrm{e}^{-\beta_i(t-T_{(k)})}\right]\mathbb{1}_{u\leq t}\,\mathrm{d}u \\ =&\mu_i\left(\min(t, T_{(k+1)}) - T_{(k)}^{i\star}\right) \\ &+ \beta_i^{-1}\left(\lambda^{i\star}(T_{(k)}) - \mu_i\right) \\ &\left(\mathrm{e}^{-\beta_i(T_{(k)}^{i\star} - T_{(k)})} - \mathrm{e}^{-\beta_i(\min(t, T_{(k+1)}) - T_{(k)})}\right)\,.\end{aligned}$$

□

4.C Proof of Theorem 4.3.1

Proof. We only need to prove that if $\lambda_{\theta_i}^i(t \mid \mathcal{F}_t) = \lambda_{\theta_i'}^i(t \mid \mathcal{F}_t)$ a.e. for every $i \in \{1, \dots, d\}$ then $\theta = \theta'$. In this proof, both intensities are considered with respect to the same filtration $\mathcal{F}_t$ so we will omit it from the rest of the proof.

Let $\theta, \theta' \in \Theta$. Let us assume that $\lambda_{\theta_i}^i(t) = \lambda_{\theta_i'}^i(t)$ for all $t < T$. In order to prove equality between the two parameters we will first prove that $\mu_i = \mu_i{}'$, then $\beta_i = \beta_i'$ and lastly that $\alpha_{ij} = \alpha_{ij}'$ for every i, j.

- For any i, as $\lambda^i_{\theta_i}(t) = \lambda^i_{\theta'_i}(t)$ a.e., then

$$\begin{aligned} & \Lambda^i_{\theta_i}(T_{(1)}) = \Lambda_{\theta'_i}(T_{(1)}) \\ \implies \quad & \mu_i T_{(1)} = {\mu_i}' T_{(1)} \\ \implies \quad & \mu_i = {\mu_i}' . \end{aligned}$$

- For any i, let us choose $T_{(k+1)}$ such that it is an event of process N^i. As

$$\mathbb{P}(T_{(k+1)} \text{ is an event of } N^i \mid \mathcal{F}_{T^-_{(k+1)}}) = \frac{\lambda^i_{\theta_i}(T^-_{(k+1)})}{\lambda_\theta(T^-_{(k+1)})} ,$$

then $\lambda^i_{\theta_i}(T^-_{(k+1)}) > 0$. By definition of $T^{i\star}_{(k),\theta}$,

$$T^{i\star}_{(k),\theta} < T_{(k+1)} \text{ a.s.} .$$

Furthermore, $\lambda^i_{\theta_i}(t) = 0$ for $t \in (T_{(k)}, T^{i\star}_{(k),\theta_i})$ and $\lambda^i_{\theta_i}(t) > 0$ for $t \in (T^{i\star}_{(k),\theta_i}, T_{(k+1)})$. Then, as we assumed that $\lambda^i_{\theta_i}(t) = \lambda^i_{\theta'_i}(t)$ a.e., we can conclude that $T^{i\star}_{(k),\theta_i} = T^{i\star}_{(k),\theta'_i}$. By differentiating the intensity functions on the interval $(T^{i\star}_{(k),\theta_i}, T_{(k+1)})$ as in the proof of Lemma 4.3.1 we obtain:

$$\begin{aligned} & (\lambda^i_{\theta_i})'(t) = (\lambda^i_{\theta'_i})'(t) \text{ a.e.} \\ \implies & -\beta_i(\lambda^i_{\theta_i}(t) - \mu_i) = -\beta'_i(\lambda^i_{\theta'_i}(t) - \mu_i) \text{ a.e.} \\ \implies & (\beta_i - \beta'_i)(\lambda^i_{\theta_i}(t) - \mu_i) = 0 \text{ a.s.} \\ \implies & \int_{T^{i\star}_{(k),\theta}}^{T_{(k+1)}} (\beta_i - \beta'_i)(\lambda^i_{\theta_i}(t) - \mu_i)\,\mathrm{d}t = 0 \\ \implies & (\beta_i - \beta'_i) \int_{T^{i\star}_{(k),\theta}}^{T_{(k+1)}} (\lambda^i_{\theta_i}(t) - \mu_i)\,\mathrm{d}t = 0 . \end{aligned}$$

Additionnally, $|\lambda^i_{\theta_i}(t) - \mu_i| > 0$ a.s. for all $t \in (T^{i\star}_{(k),\theta_i}, T_{(k+1)})$ as $\lambda^i_{\theta_i}$ is monotone and converges to μ_i. Then it follows that the integral is non-zero and so $\beta_i = \beta'_i$.

- Let us prove the equality $\alpha_{ij} = \alpha'_{ij}$. By the assumption made on the event times, for any i, j with $i \neq j$, there exists two event times $\tau < \tau_+$ with τ an event time from process N_j and τ_+ an event time from process N_i such that:

  1. $\lambda^i_{\theta_i}(\tau^-) > 0$;
  2. there are only events of process N^j in the interval $[\tau, \tau_+)$.

Let τ and τ_+ be two such event times. As a reminder, $T_{(N(\tau)-1)}$ corresponds to the event time before τ and similarly for τ_+. For $t \in [T_{(N(\tau)-1)}, \tau)$, as $\lambda^i_{\theta_i}(t) = \lambda^i_{\theta'_i}(t)$ a.e. and by Condition 1:

$$\lambda^{i\star}_{\theta_i}(\tau^-) = \lambda^{i\star}_{\theta'_i}(\tau^-) .$$

By using the following equalities (proven in the previous points of the proof)

$$\mu_i = \mu_i{}' \,, \qquad \beta_i = \beta_i' \,,$$
$$T^{i\star}_{(N(\tau)-1),\theta_i} = T^{i\star}_{(N(\tau)-1),\theta_i'} \,,$$

it follows that

$$\lambda^{i\star}_{\theta_i}(\tau^-) = \lambda^{i\star}_{\theta_i'}(\tau^-)$$
$$\implies \sum_{l=1}^{d} \alpha_{il} \sum_{T_k^l<\tau} \mathrm{e}^{-\beta_i(\tau - T_k^l)} = \sum_{l=1}^{d} \alpha'_{il} \sum_{T_k^l<\tau} \mathrm{e}^{-\beta_i(\tau - T_k^l)}$$
$$\implies \sum_{l=1}^{d} (\alpha_{il} - \alpha'_{il}) A^l = 0 \,, \tag{4.7}$$

where

$$A^l = \sum_{T_k^l<\tau} \mathrm{e}^{-\beta_i(\tau - T_k^l)} \,.$$

We can then write Equation (4.7) by replacing τ by τ_+ as $\lambda^i_{\theta_i}(\tau_+^-) > 0$ because τ_+ is an event time of process N^i. We obtain then

$$\sum_{l=1}^{d} (\alpha_{il} - \alpha'_{il}) B^l = 0 \,, \tag{4.8}$$

where

$$B^l = \sum_{T_k^l<\tau_+} \mathrm{e}^{-\beta_i(\tau_+ - T_k^l)} \,.$$

By definition of event times τ and τ_+, all events on the interval $[\tau, \tau_+)$ are from process N^j so we obtain for all $l \neq j$,

$$B^l = A^l \mathrm{e}^{-\beta_i(\tau_+ - \tau)} \,.$$

For $l = j$,

$$B^j = A^j \mathrm{e}^{-\beta_i(\tau_+ - \tau)} + \sum_{\tau \le T_k^j<\tau_+} \mathrm{e}^{-\beta_i(\tau_+ - T_k^j)} \,,$$

where the second term of the right hand side is positive as interval $[\tau, \tau_+)$ contains at least one event, τ, from process N^j. We can rewrite then Equation (4.8):

$$\sum_{l=1}^{d} (\alpha_{il} - \alpha'_{il}) B^l = 0$$

$$\implies \mathrm{e}^{-\beta_i(\tau_+ - \tau)} \sum_{l=1}^{d} (\alpha_{il} - \alpha'_{il}) A^l + (\alpha_{ij} - \alpha'_{ij}) \sum_{\tau \le T_k^j<\tau_+} \mathrm{e}^{-\beta_i(\tau_+ - T_k^j)} = 0 \,.$$

By Equation (4.7), the first term is null and the second sum is non-zero as interval $[\tau, \tau_+)$ contains at least an event from process N^j. It follows that:

$$\begin{aligned}(\alpha_{ij} - \alpha'_{ij}) \sum_{\tau \le T_k^j < \tau_+} \mathrm{e}^{-\beta_i(\tau_+ - T_k^j)} = 0 \\ \implies \quad (\alpha_{ij} - \alpha'_{ij}) = 0\,.\end{aligned}$$

It follows that for every $j \neq i$, $\alpha_{ij} = \alpha'_{ij}$. It remains to prove that $\alpha_{ii} = \alpha'_{ii}$. For this, let T_2^i be the second event time of process N^i. We can the write the equality

$$\begin{aligned}\lambda_{\theta_i}^{i\star}(T_2^i) - \lambda_{\theta'_i}^{i\star}(T_2^i) = 0\,, \\ \implies \sum_{l=1}^{d} (\alpha_{il} - \alpha'_{il}) \sum_{T_k^l < T_2^i} \mathrm{e}^{-\beta_i(\tau - T_k^l)} = 0\,, \\ \implies (\alpha_{ii} - \alpha'_{ii}) \sum_{T_k^i < T_2^i} \mathrm{e}^{-\beta_i(\tau - T_k^l)} = 0\,,\end{aligned}$$

as for $j \neq i$, $\alpha_{ij} = \alpha'_{ij}$. The sum is non-zero as it contains the event T_1^i and so we obtain $\alpha_{ii} = \alpha'_{ii}$.

This achieves the proof. □

4.D Proof of Corollary 4.3.1

Proof. For all $i \in \{1, \dots, d\}$, $\theta \in \Theta$ and $k \in \mathbb{N}^*$,

$$\begin{aligned}\log \lambda_{\theta_i}^i(T_k^{i-}) &= -\infty \mathbb{1}_{\lambda_{\theta_i}^i(T_k^{i-})=0} \\ &\quad + \log \lambda_{\theta_i}^i(T_k^{i-}) \mathbb{1}_{\lambda_{\theta_i}^i(T_k^{i-})>0} \\ &= -\infty \mathbb{1}_{\lambda_{\theta_i}^{i\star}(T_k^{i-})\le 0} \\ &\quad + \log \lambda_{\theta_i}^{i^\star}(T_k^{i-}) \mathbb{1}_{\lambda_{\theta_i}^{i\star}(T_k^{i-})>0} \\ &= \log \lambda_{\theta_i}^{i^\star}(T_k^{i-}) \\ &= \log \lim_{t \to T_k^{i-}} \lambda_{\theta_i}^{i^\star}(t).\end{aligned}$$

Now, for $k = 1$, $\lambda_{\theta_i}^{i^\star}(T_k^{i-}) = \mu_i$, and for $k \ge 2$, let us note $q = N(T_k^i) - 1$. Then,

$$[T_{(q)}, T_{(q+1)}) = [T_{(N(T_k^i)-1)}, T_{(N(T_k^i))}) = [S_k^i, T_k^i)\,,$$

and by Lemma 4.A.1, $\forall t \in [T_{(q)}, T_{(q+1)})$:

$$\begin{aligned}\lambda_\theta^{i\star}(t) &= \mu_i + \left(\lambda_{\theta_i}^{i\star}(T_{(q)}) - \mu_i\right) \mathrm{e}^{-\beta_i(t - T_{(q)})} \\ &= \mu_i + \left(\lambda_{\theta_i}^{i\star}(S_k^i) - \mu_i\right) \mathrm{e}^{-\beta_i(t - S_k^i)}\,.\end{aligned}$$

Thus,

$$\begin{aligned}\lim_{t\to T_k^{i-}} \lambda_{\theta_i}^{i\star}(t) &= \lim_{t\to T_{(q)}^{-}} \lambda_{\theta_i}^{i\star}(t) \\ &= \mu_i + \left(\lambda_{\theta_i}^{i\star}(S_k^i) - \mu_i\right) \mathrm{e}^{-\beta_i (T_k^i - S_k^i)} .\end{aligned}$$

To conclude, by Equation (4.3),

$$\begin{aligned}\ell_t^i(\theta_i) &= \sum_{k=1}^{N^i(t)} \log \lambda_{\theta_i}^i(T_k^{i-}) - \Lambda_{\theta_i}^i(t) \\ &= \log \lambda_{\theta_i}^i(T_1^{i-}) + \sum_{k=2}^{N^i(t)} \log \lambda_{\theta_i}^i(T_k^{i-}) - \Lambda_{\theta_i}^i(t) \\ &= \log \mu_i + \sum_{k=2}^{N^i(t)} \log \left(\mu_i + \left(\lambda_{\theta_i}^{i\star}(S_k^i) - \mu_i\right) \mathrm{e}^{-\beta_i (T_k^i - S_k^i)}\right) \\ &\quad - \Lambda_{\theta_i}^i(t) .\end{aligned}$$

□

4.E Algorithm for computing the log-likelihood

This section presents Algorithm 5 for computing the log-likelihood $\ell_t(\theta)$ by leveraging the results from Corollary 4.3.1.

4.F Reconstructed interaction functions for synthetic data

This section presents the reconstruction of interaction functions h_{ij} along with the estimated functions $\tilde{h}_{ij}$ from the two-dimensional Hawkes processes simulations as described in Section 4.4.2. Figure 4.13 and Figure 4.14 correspond respectively to the estimations for Scenarios (1) and (2) from Table 4.1.

Algorithm 5: Computation of the log-likelihood $\ell_t(\theta)$ of a multivariate exponential Hawkes process.

Input Parameters μ_i, α_{ij}, β_i for $i, j \in \{1, \dots, d\}$, list of event times and marks $(T_{(k)}, m_k)_{k=1:N(t)}$;

Initialisation Initialise for all i, $\Lambda_k^i = \mu_i T_{(1)}$, $\lambda^{i\star}(T_{(k)}^-) = \mu_i$, $\lambda_k^{i\star} = \mu_i + \alpha_{im_1}$ and $\ell_t(\theta) = \log(\lambda^{m_1\star}(T_{(k)}^-)) - \sum_{i=1}^d \Lambda_k^i$;

for $k = 2$ *to* $N(t)$ **do**

Compute for all i, $T_{(k-1)}^{i\star} = \min\left(T_{(k-1)} + \beta_i^{-1} \log\left(\frac{\mu_i - \lambda_k^{i\star}}{\mu_i}\right) \mathbb{1}_{\{\lambda_k^{i\star} < 0\}}, T_{(k)}\right)$;

Compute for all i,
$\Lambda_k^i = \mu_i(T_{(k)} - T_{(k-1)}^{i\star}) + \beta_i^{-1}(\lambda_k^{i\star} - \mu_i)(\mathrm{e}^{-\beta_i(T_{(k-1)}^{i\star} - T_{(k-1)})} - \mathrm{e}^{-\beta_i(T_{(k)} - T_{(k-1)})})$;

Compute for all i, $\lambda^{i\star}(T_{(k)}^-) = \mu_i + (\lambda_k^{i\star} - \mu_i)\mathrm{e}^{-\beta_i(T_{(k)} - T_{(k-1)})}$;

Update $\ell_t(\theta) = \ell_t(\theta) + \log(\lambda^{m_k\star}(T_{(k)}^-)) - \sum_{i=1}^d \Lambda_k^i$;

Compute for all i, $\lambda_k^{i\star} = \lambda^{i\star}(T_{(k)}^-) + \alpha_{im_k}$;

end

Compute for all i, $T_{(N(t))}^{i\star} = \min\left(T_{(N(t))} + \beta_i^{-1} \log\left(\frac{\mu_i - \lambda_k^{i\star}}{\mu_i}\right) \mathbb{1}_{\{\lambda_k^{i\star} < 0\}}, t\right)$;

Compute for all i,
$\Lambda_k^i = \left[\mu_i(t - T_{(N(t))}^{i\star}) + \beta_i^{-1}(\lambda_k^{i\star} - \mu_i)(\mathrm{e}^{-\beta_i(T_{N(t)}^{i\star} - T_{(N(t))})} - \mathrm{e}^{-\beta_i(t - T_{(N(t))})})\right] \mathbb{1}_{\{t > T_{(N(t))}^{i\star}\}}$;

Update $\ell_t(\theta) = \ell_t(\theta) - \sum_{i=1}^d \Lambda_k^i$;

return Log-likelihood $\ell_t(\theta)$.

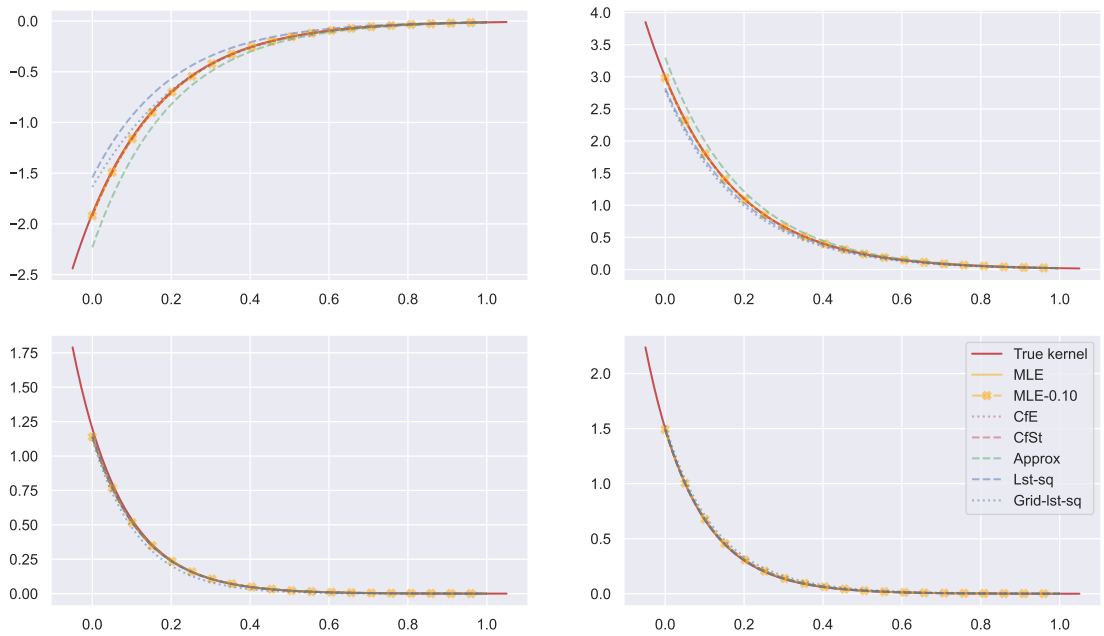

Figure 4.13: Reconstruction of interaction functions h_{ij} for Scenario (1) of two-dimensional Hawkes processes along with all estimated functions $\tilde{h}_{ij}$. The real function is plotted in red and 25 estimations are averaged for each method.

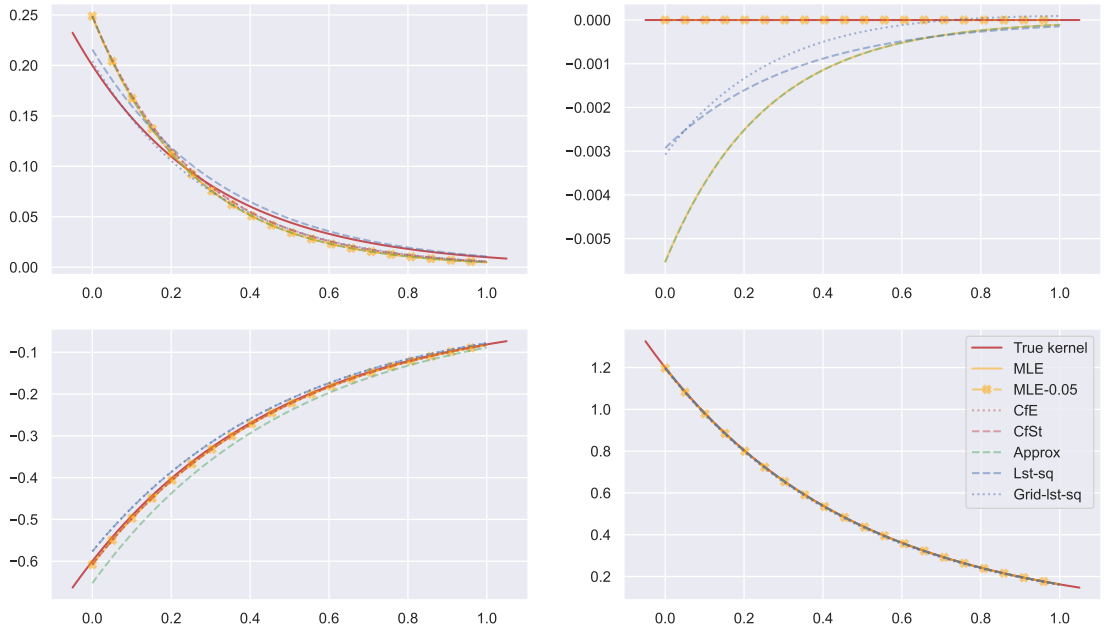

Figure 4.14: Reconstruction of interaction functions h_{ij} for Scenario (2) of two-dimensional Hawkes processes along with all estimated functions $\tilde{h}_{ij}$. The real function is plotted in red and 25 estimations are averaged for each method.

5 Spectral analysis for the inference of noisy Hawkes processes

Classic estimation methods for Hawkes processes rely on the assumption that observed event times are indeed a realisation of a Hawkes process, without considering any potential perturbation of the model. However, in practice, observations are often altered by some noise, the form of which depends on the context. It is then required to model the alteration mechanism in order to infer accurately such a noisy Hawkes process. While several models exist, we consider, in this work, the observations to be the indistinguishable union of event times coming from a Hawkes process and from an independent Poisson process. Since standard inference methods (such as maximum likelihood or Expectation-Maximisation) are either unworkable or numerically prohibitive in this context, we propose an estimation procedure based on the spectral analysis of second order properties of the noisy Hawkes process. Novel results include sufficient conditions for identifiability of the ensuing statistical model with exponential interaction functions for both univariate and bivariate processes. Although we mainly focus on the exponential scenario, other types of kernels are investigated and discussed. A new estimator based on maximising the spectral log-likelihood is then described, and its behaviour is numerically illustrated on synthetic data. Besides being free from knowing the source of each observed time (Hawkes or Poisson process), the proposed estimator is shown to perform accurately in estimating both processes.

Outline of the current chapter

5.1 Introduction

Hawkes processes, introduced in Hawkes (1971), are a class of point processes that have been originally used to model self-exciting phenomena and more recently other types of past-dependent behaviours. Their fields of applications are wide and include for instance seismology (Ogata 1988; Ogata 1998), neuroscience (Chornoboy et al. 1988; Lambert et al. 2018), criminology (Olinde et al. 2020), finance (Embrechts et al. 2011; Bacry et al. 2015) and biology (Gupta et al. 2018), to mention a few. Consequently, there has been a deep focus on estimation techniques for Hawkes processes. Among them, let us mention maximum likelihood approaches (Ogata 1978; Ozaki 1979; Guo et al. 2018), methods of moments (Da Fonseca et al. 2013), least-squares contrast minimisation (Reynaud-Bouret et al. 2014; Bacry et al. 2020), Expectation-Maximisation (EM) procedures (Lewis et al. 2011), and methods using approximations through autoregressive models (Kirchner 2017).

All of these methods are based on the assumption that the history of the process has been accurately observed, although this information is partial or noised in many contexts, in particular due to measurement or detection errors. Different models, described notably in Lund et al. (2000), have been proposed to handle such errors for spatial point processes, with inference methods based on approximating the likelihood depending on each noise scenario. The first scenario, called *displacement*, is when the event times are observed with a shift. If deconvolution methods to recover the unnoised process are standard approaches for simpler point processes such as Poisson processes (see for instance Antoniadis et al. (2006) and Bonnet et al. (2022b) or the review by Hohage et al. (2016)), the literature for Hawkes processes remains scarce and consists of the work of Trouleau et al. (2019) where the event times are observed with a delay and the work of Deutsch et al. (2020) where the shift follows a Gaussian distribution. The latter also explores the framework where some event times are undetected, which is similar to that studied in Mei et al. (2019). This setting can either be referred to as *thinning* when the observations are randomly missing, or *censoring* when complete regions are unobserved. The last scenario, called *Superposition of ghost points* by Lund et al. (2000), is the focus of this chapter and describes situations where additional points are coming from an external point process, in our case a Poisson process. A real-world application that motivated this work comes from spike trains analysis in neuroscience: the membrane potential, which is a continuous signal, is recorded and when it exceeds a certain threshold, one considers that an event time, called a spike, occurred. However, since the original signal is noised, it is possible to detect spikes that do not correspond to real events. Regarding inference of such a noisy process, let us highlight that exact likelihood approaches are intractable due to the unknown origin of each occurrence

(Hawkes process or noise process) while methods based on inferring this missing information, for instance Expectation-Maximisation algorithms, are computationally too demanding.

Inspired by the work of Cheysson et al. (2022) when the event times of a Hawkes process are aggregated, we turn to spectral analysis (Section 5.2.1) to propose a novel estimation procedure that allows to infer the noisy Hawkes process. On the way, we present a general characterisation of the Bartlett spectrum of the superposition of two independent processes (Section 5.2.2) and identifiability results of the statistical model in the univariate (Section 5.3) and the bivariate (Section 5.4) settings. Inference in these two settings is numerically illustrated in Section 5.5. Before starting, we present the mathematical setting (Section 5.1.1) and review the references related to spectral approaches (Section 5.1.2).

5.1.1 Mathematical setting

Let $H = (H_1, \dots, H_d)$ be a stationary multivariate Hawkes process on $\mathbb{R}$ defined by its conditional intensity functions λ_i^H ($i \in \{1, \dots, d\}$): for all $t \in \mathbb{R}$,

$$\lambda_i^H(t) = \mu_i + \sum_{j=1}^{d} \int_{-\infty}^{t} h_{ij}(t-s)\, H_j(\mathrm{d}s) = \mu_i + \sum_{j=1}^{d} \sum_{T_k^{H_j} \le t} h_{ij}(t - T_k^{H_j})\,, \tag{5.1}$$

where $\mu_i > 0$ is the baseline intensity of process H_i, $h_{ij} : \mathbb{R} \to \mathbb{R}_{\ge 0}$ is the interaction or kernel function describing the effect of process H_j on process H_i, and $(T_k^{H_j})_{k\ge 1}$ denotes the event times of H_j.

By defining the matrix $S = (\|h_{ij}\|_1)_{1\le i,j\le d}$, where

$$\|h_{ij}\|_1 = \int_{-\infty}^{+\infty} h_{ij}(t)\,\mathrm{d}t\,,$$

the stationarity condition of H reduces to controlling the spectral radius of S: $\rho(S) < 1$ (Brémaud et al. 1996).

The goal of this work is to study a noisy version of the Hawkes process where the sequences of event times $(T_k^{H_1})_{k\ge1}$, …, $(T_k^{H_d})_{k\ge1}$ of H are contaminated by the event times from another process P. Since the latter process is aimed at modeling an external noise mechanism due to errors in the detection of event times, it is naturally assumed that each subprocess of H is perturbed uniformly with the same level of noise, such that P is chosen to be a multivariate Poisson process with same intensity across subprocesses. Formally, let $P = (P_1, \dots, P_d)$ be a multivariate homogeneous Poisson process on $\mathbb{R}$, supposed to be independent from H, where each univariate process P_i has the same intensity $\lambda_0 > 0$. We note the event times $(T_k^{P_i})_{k\ge1}$ ($i \in \{1, \dots, d\}$). We consider then the point process $N = (N_1, \dots, N_d)$ defined as the superposition of H and P (Definition 5.2.1): the sequence of event times $(T_k^{N_i})_{k\ge1}$ of N_i ($i \in \{1, \dots, d\}$) is the ordered union of $(T_k^{H_i})_{k\ge1}$ and $(T_k^{P_i})_{k\ge1}$.

Throughout this chapter we will refer to N as the noisy Hawkes process, and it will be assumed that event times of N are observed without knowing their origin (Hawkes or Poisson process). Our goal is to estimate both processes (i.e. the baselines μ_i, the kernels h_{ij} and the shared Poisson intensity λ_0) from the sole observation of $(T_k^{N_i})_{k\ge1}$, $i \in \{1, \dots, d\}$.

Inference procedures for point processes often leverage the intensity functions in order to devise maximum likelihood and method of moment estimators (Ogata 1988; Ozaki 1979; Da Fonseca et al. 2013). Here, the process of interest N being a superposition of two independent point processes, the intensity of each subprocess N_i reads (for any integer $i \in \{1, \dots, d\}$): for all

$t \in \mathbb{R}$,

$$\lambda_i^N(t) = \lambda_0 + \mu_i + \sum_{j=1}^{d} \int_{-\infty}^{t} h_{ij}(t-s)\, H_j(\mathrm{d}s)\,. \tag{5.2}$$

However, it appears from Equation (5.2) that usual estimators cannot be designed from intensity functions $\lambda_1^N, \dots, \lambda_d^N$ since they are based on H, which is indistinguishable from P in our setting.

In order to estimate the Hawkes and the Poisson components of N, we propose to leverage the spectral analysis of point processes, recently advocated by Cheysson et al. (2022) for inference of aggregated Hawkes processes. It consists in considering, for a multivariate point process, its matrix-valued spectral density function, denoted $\mathbf{f} : \mathbb{R} \to \mathbb{C}^{d\times d}$, which is related to second-order measures (Bartlett 1963). Given some observed times $(T_k^{N_1})_{k\geq 1}, \dots, (T_k^{N_d})_{k\geq 1}$, (in a prescribed time window $[0, T]$), the spectral density is linked to cross-periodograms, defined for all pairs $(i, j) \in \{1, \dots, d\}^2$ and all $\omega \in \mathbb{R}$ by:

$$I_{ij}^T(\omega) = \frac{1}{T} \sum_{k=1}^{N_i(T)} \sum_{l=1}^{N_j(T)} \mathrm{e}^{-2\pi \mathrm{i}\omega(T_k^{N_i} - T_l^{N_j})}\,, \tag{5.3}$$

where $N_i(t) = N_i([0, t))$. Indeed, considering the matrix-valued function $\mathbf{I}^T : \omega \in \mathbb{R} \mapsto (I_{ij}^T(\omega))_{1\leq i,j\leq d}$, the aforementioned link is to be understood as $\mathbf{I}^T(\omega)$ (for all $\omega \in \mathbb{R}$) being asymptotically distributed according to a complex Wishart distribution with one degree of freedom and scale matrix $\mathbf{f}(\omega)$ (Tuan 1981; Villani et al. 2022). In particular, this implies that $\mathbb{E}[\mathbf{I}^T(\omega)] = \mathbf{f}(\omega)$. Moreover, it is noteworthy that the periodogram $\mathbf{I}^T(\omega)$ can be computed regardless of knowing the source of the event times. This paves the way for estimation.

As it happens, in the scope of statistical inference, a parametric model for the matrix-valued spectral density function is considered:

$$\mathcal{P} = \left\{ \mathbf{f}_\theta^N : \mathbb{R} \to \mathbb{C}^{d\times d}, \theta = (\mu, \gamma, \lambda_0) \in \Theta \right\}\,,$$

where γ is a parameter that characterises the interaction functions. Then, for $\omega_k = k/T$ ($k \in \{1, 2, \dots\}$), it can be shown that $(\mathbf{I}^T(\omega_k))_{k\geq 1}$ are asymptotically independent, leading to the approximate spectral log-likelihood (Brillinger 2012; Düker et al. 2019; Villani et al. 2022):

$$\ell_T(\theta) = -\frac{1}{T} \sum_{k=1}^{M} \left\{ \log\left(\det\left(\mathbf{f}_\theta^N(\omega_k)\right)\right) + \operatorname{tr}\left(\mathbf{f}_\theta^N(\omega_k)^{-1} \mathbf{I}^T(\omega_k)\right) \right\}, \tag{5.4}$$

where det and tr are respectively the determinant and trace of matrices. Then, the so-called Whittle (or spectral) estimator $\mathbf{f}_{\hat\theta}^N$ of $\mathbf{f}$ (Whittle 1952) is obtained for $\hat\theta \in \Theta$ such that

$$\hat\theta \in \arg\max_{\theta\in\Theta}\ \ell_T(\theta)\,.$$

5.1.2 Related works

The spectral analysis of point processes was introduced in Bartlett (1963) and extended to 2-dimensional point processes in Bartlett (1964). Subsequent research works focusing on the theoretical properties of the Bartlett spectrum include Daley (1971), Daley et al. (2003), and Tuan (1981) for temporal settings and Mugglestone et al. (1996), Mugglestone et al. (2001), and Rajala et al. (2023) for spatial contexts.

Despite this, practical applications remain scarcer in the literature. Adamopoulos (1976)

studies earthquake arrivals through the analysis of Hawkes processes and Karavasilis et al. (2007) analyses the cross-correlation of bivariate point processes in the context of muscular stimulation.

In a recent contribution by Cheysson et al. (2022), the authors employ the spectral analysis of point processes to infer an aggregated Hawkes process. More precisely, the observations are assumed to come from a standard Hawkes process but only the counts of events on fixed intervals are available. By leveraging the properties of the Bartlett spectrum, they propose an estimator obtained by means of maximisation of the spectral log-likelihood.

The main advantage of the spectral approach is its obvious ability to handle different kinds of partial observations. That is why we propose an inference procedure for noisy Hawkes processes based on Bartlett's spectral density.

Spectral results concerning the Hawkes model are built on the linearity of the intensity function along with its branching properties. The first spectral analysis of linear Hawkes processes appears in the original paper Hawkes (1971) and was then developed in Daley et al. (2003) in both univariate and multivariate contexts. Explicit expressions of the spectral density of a Hawkes process are available as long as the Fourier transform of the kernel functions are known, which allows to work with a wide array of parametrisations.

In this chapter, we mainly focus on studying identifiability of the statistical model with the classic exponential kernel (and also give insights about identifiability with other kernels). We subsequently derive a parametric inference procedure for estimating the parameters of a process composed of the superposition of a linear Hawkes process and a homogeneous Poisson process.

5.2 Spectral analysis

5.2.1 The Bartlett spectrum

In this section, we formally introduce the concept of matrix-valued spectral measure $\mathbf{\Gamma} : \mathbb{R} \to \mathbb{C}^{d\times d}$ for a multivariate stationary point process $N = (N_1, \dots, N_d)$. This is an extension of the Bartlett spectrum introduced by Bartlett (1963) for the analysis of univariate point processes. Let $\mathcal{S}$ be the space of real functions on $\mathbb{R}$ with rapid decay (Daley et al. 2003, Chapter 8.6.1):

$$\mathcal{S} = \left\{ f \in C^\infty, \forall k \in \{1, 2, \dots\}, \forall r \in \{1, 2, \dots\}, \sup_{x\in\mathbb{R}} \left| x^r f^{(k)}(x) \right| < \infty \right\},$$

where C^∞ is the set of smooth functions from $\mathbb{R}$ to $\mathbb{R}$ and $f^{(k)}$ is the k^{th} derivative of $f \in C^\infty$.

Then, the Bartlett spectrum of N is the matrix-valued function $\mathbf{\Gamma}^N : \omega \in \mathbb{R} \mapsto (\Gamma^N_{ij}(\omega))_{1\le i,j\le d} \in \mathbb{C}^{d\times d}$, such that for all $1 \le i, j \le d$, Γ^N_{ij} is a measure on $\mathbb{R}$ verifying (Daley et al. 2003, Equation 8.4.13):

$$\forall (\varphi, \psi) \in \mathcal{S} \times \mathcal{S} : \quad \operatorname{cov}\left(\int_{\mathbb{R}} \varphi(x)\, N_i(\mathrm{d}x), \int_{\mathbb{R}} \psi(x)\, N_j(\mathrm{d}x) \right) = \int_{\mathbb{R}} \tilde{\varphi}(\omega)\tilde{\psi}(-\omega)\, \Gamma^N_{ij}(\mathrm{d}\omega)\,,$$

where for all $f \in \mathcal{S}$, $\tilde{f} : \mathbb{R} \to \mathbb{C}$ denotes the Fourier transform of f:

$$\forall \omega \in \mathbb{R} : \quad \tilde{f}(\omega) = \int_{\mathbb{R}} f(x) \mathrm{e}^{-2\pi \mathrm{i} x\omega}\, \mathrm{d}x\,.$$

If, for all $1 \le i, j \le d$, the measure Γ^N_{ij} is absolutely continuous, we can define the matrix-valued spectral density function of N, denoted $\mathbf{f}^N : \mathbb{R} \to \mathbb{C}^{d\times d}$, such that for all $\omega \in \mathbb{R}$, $\mathbf{f}^N(\omega) = (f^N_{ij}(\omega))_{1\le i,j\le d}$ with $\Gamma^N_{ij}(\mathrm{d}\omega) = f^N_{ij}(\omega)\, \mathrm{d}\omega$.

From a practical point of view, the spectral density $\mathbf{f}^N$ can be derived from the reduced covariance densities. Let $\mathcal{B}_{\mathbb{R}}^c$ be the collection of all bounded Borel sets on $\mathbb{R}$ and $\ell_{\mathbb{R}} : \mathcal{B}_{\mathbb{R}}^c \to \mathbb{R}_{\geq 0}$ the Lebesgue measure on $\mathbb{R}$. For all $i \in \{1, \dots, d\}$, the first moment measure of process N_i is defined as $A \in \mathcal{B}_{\mathbb{R}}^c \mapsto \mathbb{E}[N_i(A)]$ and, by stationarity of the process, it comes:

$$\forall A \in \mathcal{B}_{\mathbb{R}}^c : \quad \mathbb{E}[N_i(A)] = m_i^N \ell_{\mathbb{R}}(A)\,,$$

where $m_i^N = \mathbb{E}[N_i([0,1))]$ is the mean intensity of process N_i. Then, for all $(i,j) \in \{1, \dots, d\}^2$ the second order moment measure $M_{ij}^N : \mathcal{B}_{\mathbb{R}}^c \times \mathcal{B}_{\mathbb{R}}^c \to \mathbb{R}_{\geq 0}$ is defined by (Daley et al. 2003, Section 5.4):

$$\forall (A,B) \in \mathcal{B}_{\mathbb{R}}^c \times \mathcal{B}_{\mathbb{R}}^c : \quad M_{ij}^N(A,B) = \mathbb{E}[N_i(A) N_j(B)] = \int_{A \times B} M_{ij}^N(\mathrm{d}x, \mathrm{d}y)\,.$$

Now, as the process N is stationary, M_{ij}^N can be decomposed in a product of $\ell_{\mathbb{R}}$ and a so-called reduced measure $\breve{M}_{ij}^N : \mathcal{B}_{\mathbb{R}}^c \to \mathbb{R}_{\geq 0}$, such that for any bounded measurable function $g : \mathbb{R}^2 \to \mathbb{R}$ with bounded support (Daley et al. 2003, Equation 8.1.1a):

$$\int_{\mathbb{R}^2} g(x,y)\, M_{ij}^N(\mathrm{d}x, \mathrm{d}y) = \int_{\mathbb{R}} \int_{\mathbb{R}} g(x, x+u)\, \ell_{\mathbb{R}}(\mathrm{d}x)\, \breve{M}_{ij}^N(\mathrm{d}u)\,, \tag{5.5}$$

which leads to the definition of the reduced covariance measure $\breve{C}_{ij}^N : \mathcal{B}_{\mathbb{R}}^c \to \mathbb{R}_{\geq 0}$:

$$\forall B \in \mathcal{B}_{\mathbb{R}}^c : \quad \breve{C}_{ij}^N(B) = \breve{M}_{ij}^N(B) - m_i^N m_j^N \ell_{\mathbb{R}}(B)\,. \tag{5.6}$$

Since the right-hand side is the difference of two positive, positive-definite measures (Daley et al. 2003, Section 8.6), we can define the Fourier transform of $\breve{C}_{ij}$ as the difference of their Fourier transforms (see for example (Pinsky 2008, Equation 5.2.1) for the Fourier transform of a measure). The resulting quantity comes out to correspond exactly to the spectral density function f_{ij}^N:

$$\forall \omega \in \mathbb{R} : \quad f_{ij}^N(\omega) = \int_{\mathbb{R}} \mathrm{e}^{-2\pi \mathrm{i} x \omega}\, \breve{M}_{ij}^N(\mathrm{d}x) - m_i^N m_j^N \delta(\omega)\,, \tag{5.7}$$

where δ is the Dirac delta function.

5.2.2 Superposition of processes and noisy Hawkes process

The model we study considers the superposition of two point processes that we define as follows.

Definition 5.2.1 (Superposition of processes)**.** Let X and Y be two independent and stationary multivariate point processes with same dimension d. The superposition of X and Y, denoted $N = X + Y$, is the stationary multivariate point process defined for any integer $1 \leq i \leq d$ as:

$$\forall A \in \mathcal{B}_{\mathbb{R}}^c : \quad N_i(A) = X_i(A) + Y_i(A).$$

It comes from the definition that if X and Y have respectively event times $(T_k^{X_1})_{k\geq 1}$, …, $(T_k^{X_d})_{k\geq 1}$ and $(T_k^{Y_1})_{k\geq 1}$, …, $(T_k^{Y_d})_{k\geq 1}$, the sequences of event times of $N = X+Y$ are the ordered unions of $(T_k^{X_1})_{k\geq 1}$ and $(T_k^{Y_1})_{k\geq 1}$ up to $(T_k^{X_d})_{k\geq 1}$ and $(T_k^{Y_d})_{k\geq 1}$. In addition, Proposition 5.2.1 below states that the spectral density of N can be obtained easily from those of X and Y.

Proposition 5.2.1. *Let X and Y be two independent and stationary multivariate point processes with same dimension d, admitting respective spectral densities $\mathbf{f}^X$ and $\mathbf{f}^Y$.*

Then $N = X + Y$ admits a matrix-valued spectral density function $\mathbf{f}^N$ and

$$\mathbf{f}^N = \mathbf{f}^X + \mathbf{f}^Y \,. \tag{5.8}$$

Proof. The proof is given in Appendix 5.A. □

We are now ready to define the noisy Hawkes model, which is the superposition of a Hawkes process and a homogeneous Poisson process.

Definition 5.2.2 (Noisy Hawkes process)**.** Let H be a multivariate Hawkes process and $P = (P_1, \dots P_d)$ be a multivariate homogenous Poisson process independent from H and with common intensity $\lambda_0 > 0$ (i.e. for any integer $1 \leq i \leq d$, P_i is a univariate Poisson process with constant intensity λ_0). The superposition $N = H + P$ is called a noisy Hawkes process.

Let us remark that, since the process P is aimed at modeling some kind of background noise, it is naturally assumed that subprocesses of P share the same intensity λ_0. However, if identifiability results from the forthcoming sections are specific to this assumption, the presentation done in this section can be trivially extended to a multivariate Poisson process with different intensities.

Now, let us recall that we aim at analysing the process $N = H + P$ through an observation of event times $(T_k^{N_1})_{k\geq 1}, \dots, (T_k^{N_d})_{k\geq 1}$ with the inability of distinguishing between the times of H and P. To do so from a statistical estimation perspective, we leverage the spectral log-likelihood (Equation (5.4)), which makes use of the spectral density of N. The latter can be computed thanks to Proposition 5.2.1 and Example 8.3(c) from Daley et al. (2003). Indeed, let $\tilde{h} : \mathbb{R} \to \mathbb{C}^{d\times d}$ be the matrix-valued Fourier transform of the interaction functions:

$$\forall w \in \mathbb{R} : \quad \tilde{h}(\omega) = \left(\tilde{h}_{ij}(\omega)\right)_{1\leq i,j\leq d} \,,$$

under stationarity conditions, the spectral density $\mathbf{f}^H$ of the Hawkes process H (defined as in Equation (5.1)) is (Daley et al. 2003, Equation (8.3.11)):

$$\forall w \in \mathbb{R} : \quad \mathbf{f}^H(\omega) = \left(I_d - \tilde{h}(-\omega)\right)^{-1} \operatorname{diag}\left(m^H\right) \left(I_d - \tilde{h}(\omega)^T\right)^{-1} , \tag{5.9}$$

where I_d is the identity matrix of dimension d and $\operatorname{diag}\left(m^H\right)$ is the diagonal matrix formed by the vector of the mean intensities:

$$\begin{pmatrix} m_1^H \\ \vdots \\ m_d^H \end{pmatrix} = \left(I_d - \tilde{h}(0)\right)^{-1} \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_d \end{pmatrix} .$$

Moreover, since the homogeneous Poisson process P (with common intensity λ_0) is a Hawkes process with null interactions, its spectral density $\mathbf{f}^P$ results easily from Equation (5.9):

$$\forall w \in \mathbb{R} : \quad \mathbf{f}^P(\omega) = \lambda_0 I_d \,,$$

which leads to the spectral density of N:

$$\forall w \in \mathbb{R} : \quad \mathbf{f}^N(\omega) = \mathbf{f}^H(\omega) + \lambda_0 I_d \,, \tag{5.10}$$

with $\mathbf{f}^H$ expressed in Equation (5.9). Given this result, the inference procedure consists in maximising the spectral log-likelihood described in Equation (5.4).

In the forthcoming sections, we focus on this statistical estimation problem in low-dimensional settings ($d = 1$ and $d = 2$) and provide identifiability results when interactions are exponential.

5.3 The univariate noisy Hawkes process

5.3.1 General setting

Let us start with univariate processes ($d = 1$). In this case, Equation (5.10) simplifies, as stated in Corollary 5.3.1.

Corollary 5.3.1. *Let N be a noisy Hawkes process defined by the superposition of a stationary Hawkes process H (with baseline intensity $\mu > 0$ and kernel function $h : \mathbb{R} \to \mathbb{R}_{\geq 0}$) and an independent homogeneous Poisson process P (with constant intensity $\lambda_0 > 0$). Then the spectral density f^N of N reads:*

$$\forall \omega \in \mathbb{R}, \quad f^N(\omega) = \frac{\mu}{(1 - \|h\|_1)\left|1 - \tilde{h}(\omega)\right|^2} + \lambda_0 \,.$$

Proof. This is straightforward from Equations (5.9) and (5.10). See also Daley et al. (2003, Example 8.2(e)) for the spectral density of a univariate exciting Hawkes process. □

The estimation procedure also simplifies since the periodogram of N and the spectral log-likelihood (see Equations (5.3) and (5.4)) respectively read:

$$\forall \omega \in \mathbb{R}, \quad I^T(\omega) = \frac{1}{T}\left|\sum_{k=1}^{N(T)} \mathrm{e}^{-2\pi \mathrm{i} \omega T_k^N}\right|^2 ,$$

where $(T_k^N)_{k \geq 1}$ is the sequence of event times of the noisy Hawkes process N, and:

$$\forall \theta \in \Theta, \quad \ell_T(\theta) = -\frac{1}{T}\sum_{k=1}^{M}\left(\log\left(f_\theta^N(\omega_k)\right) + \frac{I^T(\omega_k)}{f_\theta^N(\omega_k)}\right). \tag{5.11}$$

As explained in Section 5.1.1, the Whittle estimator $\hat{\theta}$ is then obtained by maximising the function ℓ_T. In the next section, the exponential parametric model $\mathcal{Q}$ for f^N is described and analysed.

5.3.2 Exponential model

Let us consider the classic exponential kernel for the Hawkes process H:

$$\forall t \in \mathbb{R} : \quad h(t) = \alpha\beta \mathrm{e}^{-\beta t} \mathbb{1}_{t \geq 0} \,, \tag{5.12}$$

with $0 < \alpha < 1$ and $\beta > 0$. The kernel parameter is thus $\gamma = (\alpha, \beta)$ and the statistical model for a univariate noisy Hawkes process becomes:

$$\mathcal{Q} = \left\{ f_\theta^N : \mathbb{R} \to \mathbb{C}, \theta = (\mu, \alpha, \beta, \lambda_0) \in \mathbb{R}_{>0} \times (0,1) \times \mathbb{R}_{>0} \times \mathbb{R}_{>0} \right\} .$$

The exponential kernel parametrisation has been widely studied from an inference point of view (see for instance Ozaki (1979) and Bacry et al. (2016)). In particular, its Fourier transform

is:

$$\forall \omega \in \mathbb{R} : \quad \tilde{h}(\omega) = \frac{\alpha\beta}{\beta + 2\pi \mathrm{i}\omega}\,, \tag{5.13}$$

and the spectral density f^H of a univariate Hawkes process with baseline intensity $\mu > 0$ and exponential kernel is (Hawkes 1971):

$$\forall \omega \in \mathbb{R} : \quad f^H(\omega) = m^H \left(1 + \frac{\beta^2\alpha(2-\alpha)}{\beta^2(1-\alpha)^2 + 4\pi^2\omega^2}\right), \quad \text{where } m^H = \frac{\mu}{1-\alpha}\,.$$

It results that the spectral density $f_\theta^N \in \mathcal{Q}$ of a univariate noisy Hawkes process is:

$$\forall \omega \in \mathbb{R} : \quad f_\theta^N(\omega) = \left(\frac{\mu}{1-\alpha}\beta^2\alpha(2-\alpha)\right)\frac{1}{\beta^2(1-\alpha)^2 + 4\pi^2\omega^2} + \left(\frac{\mu}{1-\alpha} + \lambda_0\right). \tag{5.14}$$

Now, we should be ready for implementing the inference procedure based on maximising the spectral log-likelihood as expressed in Equation (5.11). However, it appears that the model $\mathcal{Q}$, as currently defined, is not identifiable (see Proposition 5.3.1). Hopefully, this model becomes identifiable when restricted to only three parameters, thus allowing the practicability of the proposed estimation method, as numerically illustrated in Section 5.5.1.

Proposition 5.3.1 (Identifiability in the univariate setting)**.** *The model $\mathcal{Q}$ is not identifiable. In particular, for any admissible parameter $\theta = (\mu, \alpha, \beta, \lambda_0)$ there exists an infinite number of admissible parameters θ' such that $f_\theta^N = f_{\theta'}^N$.*

However, the four collections of models defined below are identifiable:

1. *for all $\mu^\circ > 0$, $\mathcal{Q}_\mu = \left\{f^N_{(\mu,\alpha,\beta,\lambda_0)} \in \mathcal{Q}, \mu = \mu^\circ\right\}$;*
2. *for all $\alpha^\circ \in (0,1)$, $\mathcal{Q}_\alpha = \left\{f^N_{(\mu,\alpha,\beta,\lambda_0)} \in \mathcal{Q}, \alpha = \alpha^\circ\right\}$;*
3. *for all $\beta^\circ > 0$, $\mathcal{Q}_\beta = \left\{f^N_{(\mu,\alpha,\beta,\lambda_0)} \in \mathcal{Q}, \beta = \beta^\circ\right\}$;*
4. *for all $\lambda_0^\circ > 0$, $\mathcal{Q}_{\lambda_0} = \left\{f^N_{(\mu,\alpha,\beta,\lambda_0)} \in \mathcal{Q}, \lambda_0 = \lambda_0^\circ\right\}$.*

Proof. The proof is given in Appendix 5.B. □

The previous discussion raises the question of whether this non-identifiability also extends to the distribution of the noisy Hawkes process. It turns out that, in the exponential case, Markov properties of the intensity function λ^H of the underlying Hawkes process help ensuring identifiability. Indeed, from the definition of the Hawkes process H and the stationarity of the Poisson process P, $\left(\lambda^N(t)\right)_{t\geq 0}$ is a Markov process: it decreases with rate $\beta(\lambda^H(t) - \mu) = \beta(\lambda^N(t) - \mu - \lambda_0)$, and the jumps occurring from the Hawkes process, with rate $\lambda^H(t) = \lambda^N(t) - \lambda_0$, are of size $\alpha\beta$, while the jumps occurring from the Poisson component, with rate λ_0, have no impact on the intensity of the process. Then $\left(\lambda^N(t), N(t)\right)_{t\geq 0}$ is also a Markov process, more specifically a piecewise deterministic Markov process (Davis 1984). This allows us to use results from Dassios et al. (2011) on the distribution of exponential Hawkes processes to show that the distribution of the exponential noisy Hawkes process is identifiable.

Proposition 5.3.2. *Let $(\mu, \alpha, \beta, \lambda_0)$ and $(\mu', \alpha', \beta', \lambda_0')$ be two admissible 4-tuples for the exponential noisy Hawkes model $\mathcal{Q}$, and N and N' respectively defined by these two tuples (Equation (5.2) with kernel (5.12)). Then, if N and N' have same distribution and $\lambda^N(0)$ (respec-*

tively $\lambda^{N'}(0)$) is distributed according to the stationary distribution of $\left(\lambda^{N}(t)\right)_{t\geq 0}$ (respectively $\left(\lambda^{N'}(t)\right)_{t\geq 0}$), then $(\mu, \alpha, \beta, \lambda_0) = (\mu', \alpha', \beta', \lambda_0')$.

Proof. See Appendix 5.C. □

A consequence of this result is that non-identifiability of our proposed method in the exponential case is a shortcoming of the spectral approach itself rather than an underlying property of the noisy Hawkes process, presumably stemming from the fact that the spectral density only encodes the second order moments of the process. In the forthcoming section, we briefly investigate whether this issue arises when considering other reproduction kernels.

5.3.3 Beyond the exponential model

Let N be a noisy Hawkes process defined by the superposition of a stationary Hawkes process with baseline intensity μ and kernel function αh, with $\alpha \in (0,1)$, $h : \mathbb{R} \to \mathbb{R}_{\geq 0}$ and $\|h\|_1 = 1$, and a homogeneous Poisson process with intensity λ_0. Per Corollary 5.3.1, its spectral density is given by

$$\forall \omega \in \mathbb{R} : \quad f^N(\omega) = \frac{\mu}{(1-\alpha)\left|1 - \alpha\tilde{h}(\omega)\right|^2} + \lambda_0 \,.$$

While it may be difficult to show that a model is identifiable from the spectral density expression, it may prove fruitful to look at its Taylor expansion around 0, and analyse the Taylor coefficients. For example, considering the uniform kernel and the corresponding Taylor expansion of the spectral density up to order 2, we get the following proposition.

Proposition 5.3.3. *Let us consider a rectangle interaction function*

$$h : t \in \mathbb{R} \mapsto \phi^{-1} \mathbb{1}_{0 \leq t \leq \phi} \,,$$

for some kernel parameter $\phi > 0$, and the corresponding statistical model for a univariate noisy Hawkes process:

$$\mathcal{R} = \left\{ f_\theta^N : \mathbb{R} \to \mathbb{C}, \theta = (\mu, \alpha, \phi, \lambda_0) \in \mathbb{R}_{>0} \times (0,1) \times \mathbb{R}_{>0} \times \mathbb{R}_{>0} \right\} .$$

Then $\mathcal{R}$ is identifiable.

Proof. See Appendix 5.D. □

This last proposition shows that non-identifiability of the spectral approach for the noisy exponential Hawkes process is more a consequence of the choice of the reproduction function h rather than a general shortcoming of the spectral approach. It is unexpected that the exponential reproduction function, usually chosen because the Markov properties for the resulting Hawkes intensity simplify calculations (Ozaki 1979; Da Fonseca et al. 2013; Duarte et al. 2019), seems to be here the main culprit of non-identifiability for our proposed spectral approach. Still, we will show how, by imposing some constraints on the modelling of multivariate noisy Hawkes processes, we are able to ensure identifiability of the model even in this case.

5.4 The bivariate noisy Hawkes process

This section addresses bivariate noisy Hawkes processes ($d = 2$). More precisely, for such a process $N = H + P$, where H is a bivariate Hawkes process (see Equation (5.1)) and P a Poisson

process with shared intensity λ_0, Corollary 5.4.1 gives the closed-form expression of the spectral density $\mathbf{f}^N$.

Corollary 5.4.1. *Let $N = (N_1, N_2)$ be a bivariate noisy Hawkes process defined by the superposition of a stationary Hawkes process $H = (H_1, H_2)$ (with baseline intensities $\mu_1 > 0$ and $\mu_2 > 0$, and kernel functions $h_{ij} : \mathbb{R} \to \mathbb{R}_{\geq 0}$, $1 \leq i, j \leq 2$) and an independent homogeneous Poisson process $P = (P_1, P_2)$ (with same constant intensity $\lambda_0 > 0$). Then the spectral density $\mathbf{f}^N$ of N reads:*

$$\forall \omega \in \mathbb{R}, \quad \mathbf{f}^N(\omega) = \begin{pmatrix} f_{11}^H(\omega) + \lambda_0 & f_{12}^H(\omega) \\ f_{21}^H(\omega) & f_{22}^H(\omega) + \lambda_0 \end{pmatrix},$$

where

$$\begin{cases} f_{11}^H(\omega) = \frac{m_1^H \left|1-\tilde{h}_{22}(\omega)\right|^2 + m_2^H \left|\tilde{h}_{12}(\omega)\right|^2}{\left|\left(1-\tilde{h}_{11}(\omega)\right)\left(1-\tilde{h}_{22}(\omega)\right) - \tilde{h}_{12}(\omega)\tilde{h}_{21}(\omega)\right|^2} \\ f_{12}^H(\omega) = \frac{m_1^H \left(1-\tilde{h}_{22}(-\omega)\right)\tilde{h}_{21}(\omega) + m_2^H \left(1-\tilde{h}_{11}(\omega)\right)\tilde{h}_{12}(-\omega)}{\left|\left(1-\tilde{h}_{11}(\omega)\right)\left(1-\tilde{h}_{22}(\omega)\right) - \tilde{h}_{12}(\omega)\tilde{h}_{21}(\omega)\right|^2} \end{cases},$$

and

$$m_1^H = \frac{\mu_1 \left(1 - \|h_{22}\|_1\right) + \mu_2 \|h_{12}\|_1}{\left(1 - \|h_{11}\|_1\right)\left(1 - \|h_{22}\|_1\right) - \|h_{12}\|_1 \|h_{21}\|_1},$$

and f_{22}^H, f_{21}^H and m_2^H are obtained by symmetry of all indices.

Proof. This is straightforward from Equations (5.9) and (5.10). □

Then, the estimation procedure is exactly that described in Section 5.1.1, which is based on computing the cross periodogram $\mathbf{I}^T$ and on maximising the spectral log-likelihood ℓ_T (see Equations (5.3) and (5.4)). Now, similarly to the univariate case detailed in Section 5.3.2, we consider exponential interaction functions, i.e. for $1 \leq i, j \leq 2$:

$$\forall t \in \mathbb{R} : \quad h_{ij}(t) = \alpha_{ij} \beta_i \mathrm{e}^{-\beta_i t} \mathbb{1}_{t \geq 0},$$

with $\alpha_{ij} \geq 0$ and $\beta_i > 0$. The kernel parameter is thus $\gamma = (\alpha, \beta)$, where $\alpha \in \mathbb{R}_{\geq 0}^{2\times 2}$ and $\beta \in \mathbb{R}_{>0}^2$ and the statistical model for a bivariate noisy Hawkes process becomes:

$$\mathcal{Q}_\Lambda = \left\{ \mathbf{f}_\theta^N : \mathbb{R} \to \mathbb{C}^{2\times 2}, \theta = (\mu, \alpha, \beta, \lambda_0) \in \mathbb{R}_{>0}^2 \times \Lambda \times \mathbb{R}_{>0}^2 \times \mathbb{R}_{>0}, \beta \in \Omega_\alpha \right\},$$

where $\Lambda \subset \{\alpha \in \mathbb{R}_{\geq 0}^{2\times 2} : \rho(\alpha) < 1\}$ is subset of matrices α that will be specified later, and for all $\alpha \in \mathbb{R}_{\geq 0}^{2\times 2}$,

$$\Omega_\alpha = \left\{ \beta \in \mathbb{R}_{>0}^2, \beta_1 = 1 \text{ if } \alpha_{11} = \alpha_{12} = 0, \beta_2 = 1 \text{ if } \alpha_{21} = \alpha_{22} = 0 \right\},$$

is a subset of admissible values for β. The definition of Ω_α takes into account that when a row, say the first one, of the interaction matrix α is null, then the corresponding kernels verify $h_{11} = 0$ and $h_{12} = 0$ independently of the value of β_1. Thus, identifiability for the parameter β_1 is hopeless, which justifies that we get rid of it (by fixing it to an arbitrary value) from the model.

Remark. *Different versions of the multivariate exponential model exist. A first convention assumes that there is a unique $\beta \in \mathbb{R}_{>0}$ shared by all kernel functions (Chevallier et al. 2019; Bacry et al. 2020). A second and less restrictive option, which is that we opt for in this work, assumes that the recovery rate $\beta_i \in \mathbb{R}_{>0}$ for each subprocess N_i ($1 \leq i \leq d$) is shared among received interactions (Bonnet et al. 2023). These choices allow for simplified derivations of estimators in the time domain and in the frequency domain, as shown below.*

The aim of this section is to study identifiability of model $\mathcal{Q}_\Lambda$. A broad analysis (i.e. for $\Lambda = \{\alpha \in \mathbb{R}_{\geq 0}^{2\times 2} : \rho(\alpha) < 1\}$) being out of reach for complexity reasons, we exhibit some situations (i.e. subsets Λ) for which non-identifiability (Proposition 5.4.1) or identifiability (Proposition 5.4.2) can be proved.

Proposition 5.4.1 (Non-identifiability in the bivariate setting)**.** *The model $\mathcal{Q}_\Lambda$ is not identifiable in the three situations:*

1. $\Lambda = \left\{ \begin{pmatrix} \alpha_{11} & 0 \\ 0 & \alpha_{22} \end{pmatrix}, 0 \leq \alpha_{11}, \alpha_{22} < 1 \right\}$, *that is for diagonal matrices α (with possibly null entries).*
2. $\Lambda = \left\{ \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ 0 & 0 \end{pmatrix}, 0 < \alpha_{11} < 1, \alpha_{12} > 0 \right\}$, *that is for matrices with positive entries in the first row and null entries in the second.*
3. $\Lambda = \left\{ \begin{pmatrix} 0 & 0 \\ \alpha_{21} & \alpha_{22} \end{pmatrix}, \alpha_{21} > 0, 0 < \alpha_{22} < 1 \right\}$, *that is for matrices with null entries in the first row and positive entries in the second.*

Proof. The proof is given in Appendix 5.E. □

Remark. *The proof of Proposition 5.4.1, Situation 1 reveals that non-identifiability stands actually for each subprocess (considered as a univariate process), such that all the submodels with null cross-interactions built by fixing α_{11} or α_{22} to zero, or by keeping them away from zero are also not identifiable.*

Proposition 5.4.2 (Identifiability in the bivariate setting)**.** *The model $\mathcal{Q}_\Lambda$ is identifiable in the six situations:*

1. $\Lambda = \left\{ \begin{pmatrix} \alpha_{11} & 0 \\ \alpha_{21} & 0 \end{pmatrix}, 0 \leq \alpha_{11} < 1, \alpha_{21} > 0 \right\}$, *that is for matrices α with null entries in the second column and a positive entry on the antidiagonal.*
2. $\Lambda = \left\{ \begin{pmatrix} 0 & \alpha_{12} \\ 0 & \alpha_{22} \end{pmatrix}, \alpha_{12} > 0, 0 \leq \alpha_{22} < 1 \right\}$, *that is for matrices α with null entries in the first column and a positive entry on the antidiagonal.*
3. $\Lambda = \left\{ \begin{pmatrix} \alpha_{11} & 0 \\ \alpha_{21} & \alpha_{22} \end{pmatrix}, 0 < \alpha_{11} < 1, \alpha_{21} > 0, 0 \leq \alpha_{22} < 1 \right\}$, *that is for matrices α with positive entries in the first column and null upper right entry.*
4. $\Lambda = \left\{ \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ 0 & \alpha_{22} \end{pmatrix}, 0 \leq \alpha_{11} < 1, \alpha_{12} > 0, 0 < \alpha_{22} < 1 \right\}$, *that is for matrices α with a null lower left entry and positive entries in the second column.*

Proof. The proofs of Situations 1 and 3 are respectively in Appendices 5.F and 5.G. The other situations are obtained by symmetry of all indices. □

Several lessons can be learnt from Propositions 5.4.1 and 5.4.2 and Remark 5.4. First, the statistical model $\mathcal{Q}_\Lambda$ is not identifiable if H reduces to a bivariate homogenous Poisson process (Proposition 5.4.1, Situation 1 with $\alpha_{11} = \alpha_{22} = 0$) or to two independent univariate Hawkes processes (Proposition 5.4.1, Situation 1 with $\alpha_{11} > 0$ and $\alpha_{22} > 0$) even if the noise P shares the same intensity λ_0 for both subprocesses. This result actually still holds true for a dimension $d > 2$.

Second, Proposition 5.4.2 tells in a nutshell that there must exist cross-interactions in the Hawkes process H (i.e. $\alpha_{12} > 0$ or $\alpha_{21} > 0$) for the model $\mathcal{Q}_\Lambda$ to be identifiable. However, interactions must not come from a Poisson subprocess and reach a self-exciting Hawkes subprocess (Proposition 5.4.1, Situations 2 and 3), but rather they must reach a self-neutral (i.e. with null self-excitation) Hawkes subprocess (Proposition 5.4.2, Situations 1 and 2) or come from a self-exciting Hawkes subprocess (Proposition 5.4.2, Situations 3 and 4).

5.5 Numerical results

This section numerically illustrates the behaviour of the proposed estimator $\hat{\theta}$ (described in Section 5.1.1) for noisy exponential Hawkes processes. It investigates the effect of horizon T and hyperparameter M in the univariate setting (Section 5.5.1), and the impact of model sparsity and interaction strength in the bivariate setting (Section 5.5.2).

In the whole study, point processes are simulated thanks to Ogata's thinning method (Ogata 1981) and numerical optimisation of the spectral log-likelihood is performed via the L-BFGS-B method (Byrd et al. 1995), implemented in the `scipy.optimize.minimize` Python function. Both simulation and estimation algorithms are freely available as a Python package on GitHub[1]

5.5.1 Univariate setting

Simulation and estimation scenarios

Data simulation We consider observations $(T_k^N)_{1\leq k\leq N(T)}$ coming from a univariate exponential noisy Hawkes process $N = H + P$, where P is a Poisson process with intensity $\lambda_0 > 0$ and H is a Hawkes process with baseline intensity $\mu = 1$ and kernel given by Equation (5.12) with parameters $\alpha = 0.5$ and $\beta = 1$.

In order to get close to stationarity while coping with inability to generate a process on the whole line $\mathbb{R}$, N is simulated on the window $[-100, T]$ with no points in $(-\infty, -100)$ but only observations falling in $[0, T]$ are considered.

The forthcoming section will illustrate the convergence of $\hat{\theta}$, thanks to its behaviour with respect to varying horizon $T \in \{250, 500, 1000, \dots, 8000\}$, and the impact of the intensity λ_0 of the noise process P on estimation accuracy, via varying noise-to-signal ratio $\lambda_0/m^H \in \{0.2, 0.4, \dots, 2.0\}$ (given the average intensity $m^H = \mu/(1-\alpha) = 2$ of H).

Statistical models According to Proposition 5.3.1, which states four collections of identifiable models for univariate exponential Hawkes processes, estimation is successively performed in models $\mathcal{Q}_\mu$, $\mathcal{Q}_\alpha$, $\mathcal{Q}_\beta$, $\mathcal{Q}_{\lambda_0}$, where the known parameter is fixed to the value of the generated process (see above).

In addition, the behaviour of $\hat{\theta}$ will be assessed thanks to its relative error $\|\hat{\theta} - \theta^\star\|_2/\|\theta^\star\|_2$ (where $\theta^\star = (\mu, \alpha, \beta, \lambda_0)$ is the vector of the parameters of the generated process) averaged over 50 different trials.

Convergence, computation time and influence of parameter M

Up to now, the hyperparameter M, appearing in the spectral log-likelihood (Equations (5.4) and (5.11)) and determining the number of frequencies tested with the spectral density, has been let

[1] `https://github.com/migmtz/noisy-hawkes-process`

free. However, the theoretical literature suggests that its choice has a lot of influence on the convergence of the estimation procedure, and has to be guided by $M/T \xrightarrow[T\to\infty]{} \infty$ (Pham 1996).

Since the rate of convergence is not specified but has an effect on the computational efficiency of the spectral estimator, we propose to study the compatible case $M = N(T)\log N(T)$ and the economy case $M = N(T)$.

Figure 5.1 displays the relative errors $\|\hat{\theta} - \theta^\star\|_2/\|\theta^\star\|_2$ with respect to both the simulation horizon T (top panels) and the estimation time (bottom panels). As expected, the quality of the estimations improves as T increases, and this independently of which parameter is fixed. Estimations are slightly better when considering $M = N(T)\log N(T)$ (orange line) especially for smaller values of T but the trade-off is a ten times higher computation time. Therefore, the performance benefit of taking $M = N(T)\log N(T)$ rather than $M = N(T)$ seems minor when compared to the computational cost. For this reason, the forthcoming numerical experiments will be performed with $M = N(T)$.

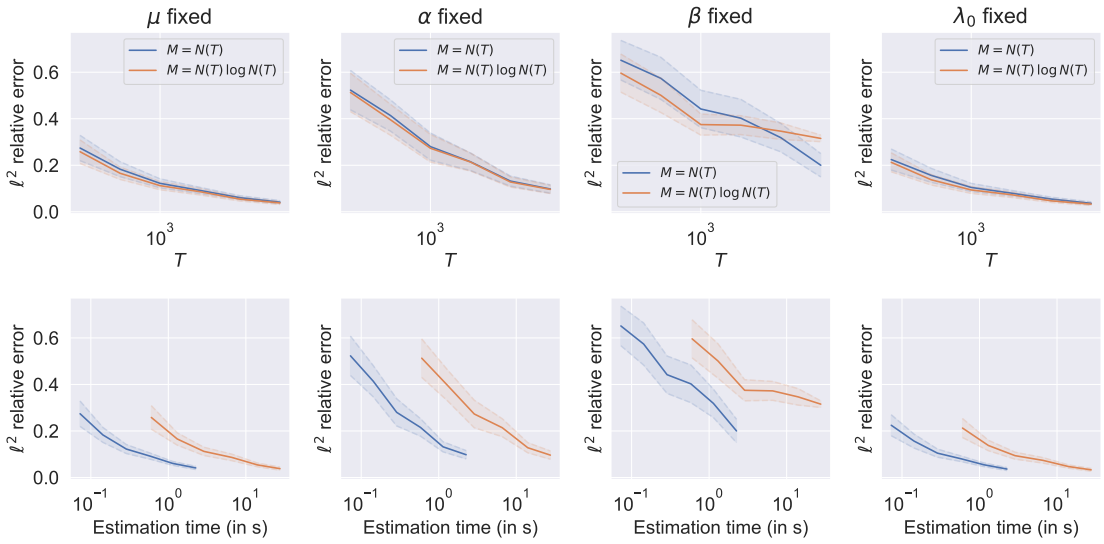

Figure 5.1: Relative estimation error with confidence bands (± 1.96 empirical standard deviation) respectively for μ, α, β and λ_0 fixed (columns from left to right) with respect to the time horizon T (top) and the computation time (bottom) in logarithmic scale. Level of noise $\lambda_0 = 1.6$.

Influence of the noise level

Figure 5.2 shows the relative error with respect to the ratio λ_0/m_H, obtained when increasing the value of λ_0 while keeping the other parameters fixed, for a given horizon $T = 8000$.

First, we can see that the value of λ_0 has a low impact on the quality of estimations. However, let us notice that when β is fixed, the average error is substantially larger than when any of the other parameters is fixed. This could be explained by a compensation phenomenon inside the triplet (μ, α, λ_0) which occurs as our method implicitly adjusts the estimation to the mean intensity of the noisy Hawkes process:

$$m^N = \lambda_0 + \frac{\mu}{1-\alpha},$$

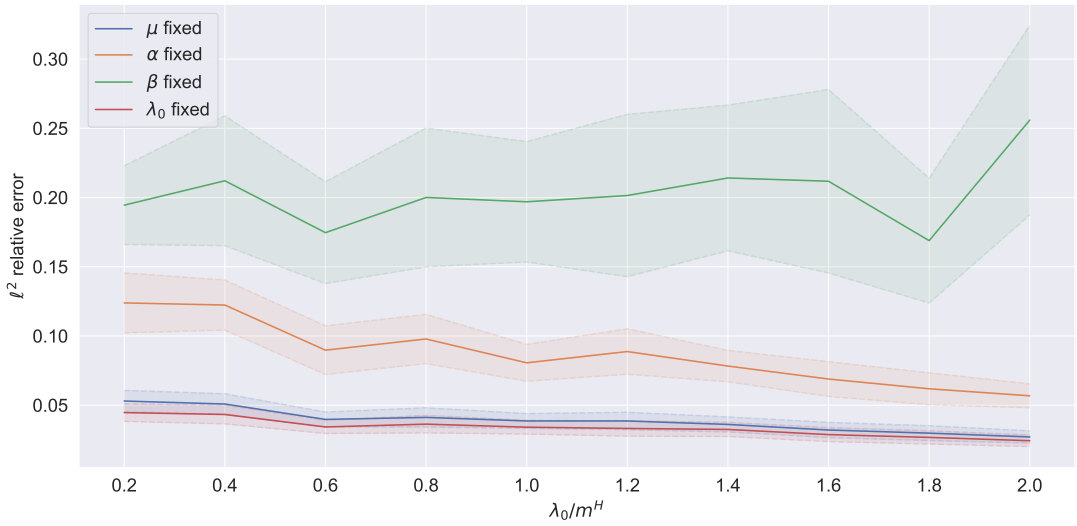

Figure 5.2: Relative estimation error with confidence bands (± 1.96 empirical standard deviation) respectively for μ, α, β and λ_0 fixed with respect to the noise-to-signal ratio for the maximal horizon $T = 8000$.

which is indeed a quantity independent of β.

This numerical compensation is illustrated in Figure 5.3, where we can see that overestimating μ is systematically balanced by underestimating α and λ_0 and vice versa, whereas the estimated mean intensities remain accurate. In this experiment, the level of noise has been arbitrarily fixed to $\lambda_0 = 1.2$ but the observed behaviour appears similarly for all possible values of λ_0 used in the previous section.

When performing estimation with α fixed (the case for which the average error is the second larger, as illustrated in Figure 5.2), the compensation appears only between $\hat{\mu}$ and $\hat{\lambda}_0$ whereas $\hat{\beta}$ does not seem impacted, as shown in Figure 5.4.

5.5.2 Bivariate setting

This section illustrates numerical estimation of bivariate exponential noisy Hawkes processes (see Section 5.4) when conditions of identifiability are met (Proposition 5.4.2). We carry out two different studies, exploring different scenarios: Section 5.5.2 studies the influence of the strength of the cross-interaction between the two subprocesses and Section 5.5.2 investigates the performance of the estimator with and without knowledge of the null components. Indeed, since identifiability conditions depend on knowing which components are non-null, an information that is unlikely to be available in practical applications, we compare the performance of the estimator for both the reduced model $\mathcal{Q}_\Lambda$, where the null components are known, and the complete model,

$$\mathcal{Q} = \left\{ \mathbf{f}_\theta^N : \mathbb{R} \to \mathbb{C}^{2\times 2}, \theta = (\mu, \alpha, \beta, \lambda_0) \in \mathbb{R}^2_{>0} \times \mathbb{R}^{2\times 2}_{\geq 0} \times \mathbb{R}^2_{>0} \times \mathbb{R}_{>0}, \rho(\alpha) < 1 \right\},$$

with no prior information.

Throughout this section, we consider a Hawkes process with $\mu = \begin{pmatrix} 1.0 \\ 1.0 \end{pmatrix}$ and $\beta = \begin{pmatrix} 1.0 \\ 1.3 \end{pmatrix}$. In

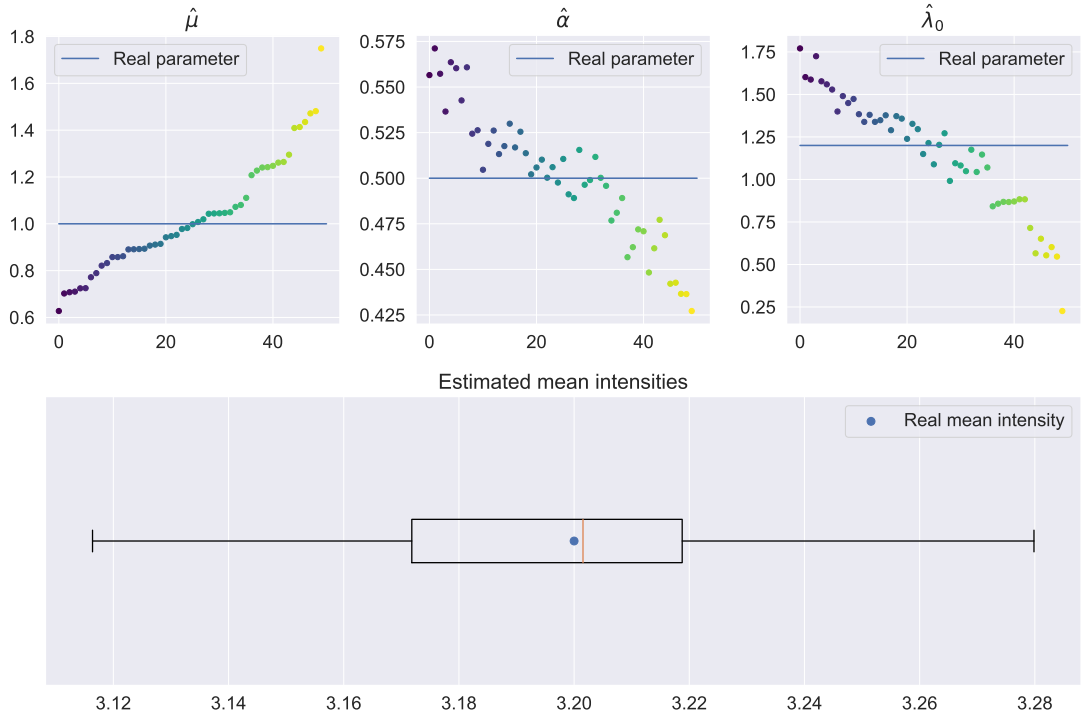

Figure 5.3: Estimations $\hat{\mu}, \hat{\alpha}$ and $\hat{\lambda}_0$ (of μ, α, and λ_0) for $\lambda_0 = 1.2$ when β is fixed, sorted by the values of $\hat{\mu}$ (top). In all plots, each color corresponds to one of the 50 repetitions. Boxplot of estimated mean intensities (bottom).

addition, fortified by the analysis of the univariate setting, the Poisson intensity is chosen to be $\lambda_0 = 0.5$ (the level of noise does not appear to have a significant impact on the quality of estimations, Figure 5.2) and it is considered $M = N(T)$ (which provides accurate estimations in a reasonable amount of time, Figure 5.1).

Influence of the cross-interaction

Let us consider one of the identifiable scenarios where the only non-null interaction in the Hawkes process is one of the two cross-interactions (see Proposition 5.4.2, Situation 1). More precisely, we consider the reduced model $\mathcal{Q}_\Lambda$, where

$$\Lambda = \left\{ \begin{pmatrix} 0 & 0 \\ \alpha_{21} & 0 \end{pmatrix} : \alpha_{21} > 0 \right\}.$$

The Hawkes process is then simulated with different levels of cross-interaction: $\alpha_{21} \in \{0.2, 0.4, 0.6, 0.8\}$, and estimations are obtained by optimising the spectral log-likelihood on the non-null parameters μ_1, μ_2, α_{21}, β_2, and λ_0.

Figure 5.5 illustrates the influence of the true parameter α_{21} on the quality of the estimations, through the relative error of the estimations for the different values of α_{21} and an increasing range of horizons T. As a complement to what has been observed in Figure 5.1, our estimator appears to behave particularly well for higher values of T, but also for higher values of α_{21}.

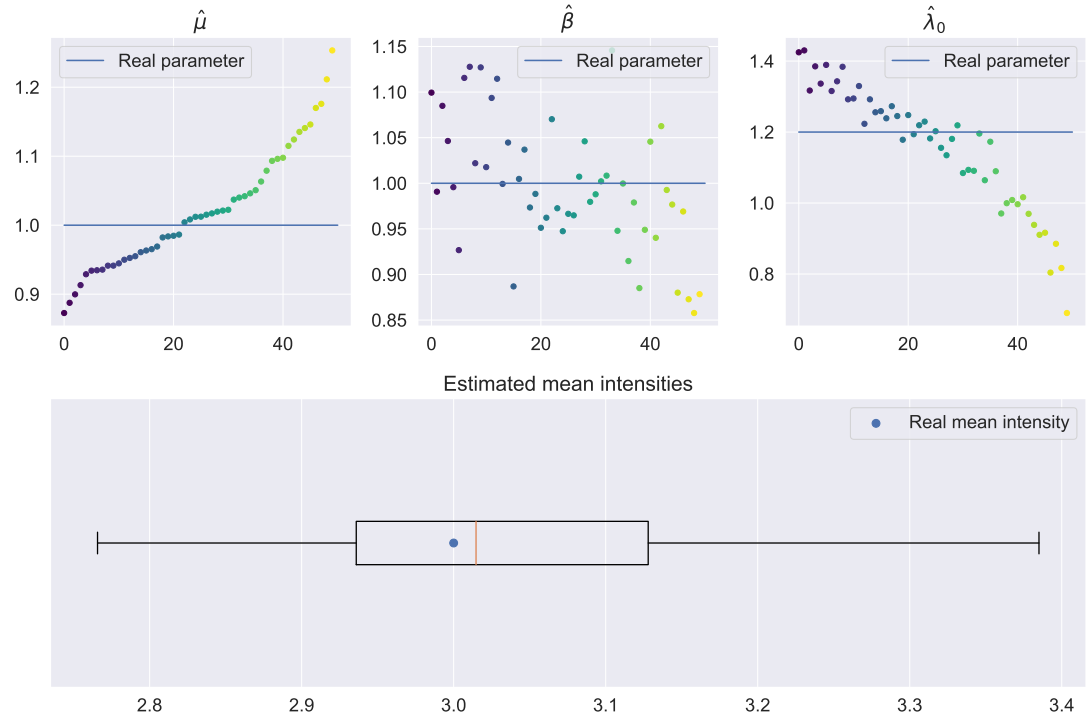

Figure 5.4: Estimations $\hat{\mu}, \hat{\beta}$ and $\hat{\lambda}_0$ (of μ, β, and λ_0) for $\lambda_0 = 1.2$ when α is fixed, sorted by the values of $\hat{\mu}$ (top). In all plots, each color corresponds to one of the 50 repetitions. Boxplot of estimated mean intensities (bottom).

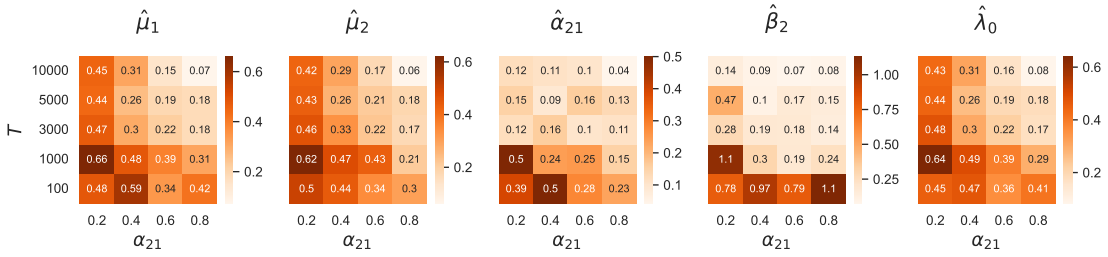

Figure 5.5: Heatmap of relative errors for each estimation $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\alpha}_{21}$, $\hat{\beta}_2$, and $\hat{\lambda}_0$ for different levels of interaction α_{21} (x-axis) and horizons T (y-axis).

This is not surprising since, for smaller values of α_{21}, the Hawkes process behaves closely to a homogeneous Poisson process, and as proven in Proposition 5.4.1, the superposition of two Poisson processes leads to a non-identifiable model. Lower interactions necessitate then higher values of T to obtain satisfactory results. Inversely, for average and high interaction magnitudes, we start to obtain small errors for horizon values around $T = 3000$.

Influence of null interactions

In this section we simulate 50 repetitions with a fixed horizon $T = 3000$ for two identifiable scenarios regarding the Hawkes process $H = (H_1, H_2)$.

Scenario 1 The matrix of interactions is:

$$\alpha = \begin{pmatrix} 0.5 & 0 \\ 0.4 & 0 \end{pmatrix},$$

corresponding to Proposition 5.4.2, Situation 1. In other terms, H_1 excites both subprocesses whereas H_2 has no influence on the dynamics (See Figure 5.6, left).

Scenario 2 The matrix of interactions is:

$$\alpha = \begin{pmatrix} 0.5 & 0 \\ 0.4 & 0.4 \end{pmatrix},$$

corresponding to Proposition 5.4.2, Situation 3. In other terms, H_1 excites both subprocesses and H_2 excites itself (See Figure 5.6, right).

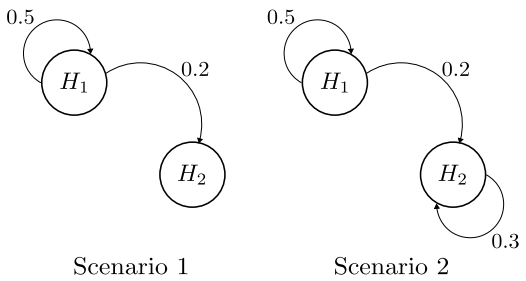

Figure 5.6: Interactions in the two numerical scenarios considered.

Graphics in the left column of Figure 5.7 present the boxplots of each parameter estimation when considering their respective reduced models $\mathcal{Q}_\Lambda$. These results show that our method provides unbiased estimates of all parameters and is particularly efficient at inferring the interaction matrix α (estimations have very low variance). The larger variances are observed for parameters μ_1, μ_2 and λ_0, which is probably due to compensation effects already mentioned in Section 5.5.1.

Estimating in the reduced model $\mathcal{Q}_\Lambda$ requires prior knowledge on the null parameters α_{ij} $(1 \leq i, j \leq 2)$, which is unlikely in practical applications. Therefore, we compare the results obtained in the reduced model to those in the complete model $\mathcal{Q}$ (see the right column graphics in Figure 5.7). If the estimates of α and β seem still empirically unbiased, we observed several deteriorations compared to the previous results. First, we notice a bias in the estimates of μ_1, μ_2 and λ_0: more precisely, μ_1 and μ_2 are overestimated while λ_0 is underestimated in both scenarios. Moreover, we observe in Scenario 2 some outlier estimations for the α_{ij} $(1 \leq i, j \leq 2)$ coefficients, which did not appear when considering the reduced model.

Fortunately, Figure 5.7 also suggests that our estimator is able to detect the null interactions in the full model, which allows to re-estimate the parameters in the reduced model. To do so, we propose to look at the 5%-empirical quantile of each term of the estimated interaction matrix $\hat{\alpha}$, which are summarised in Table 5.1.

We can notice that the empirical 5%-quantiles are set to zero for each real null parameters α_{ij} $(1 \leq i, j \leq 2)$ in both scenarios. This suggests that when enough repetitions are available,

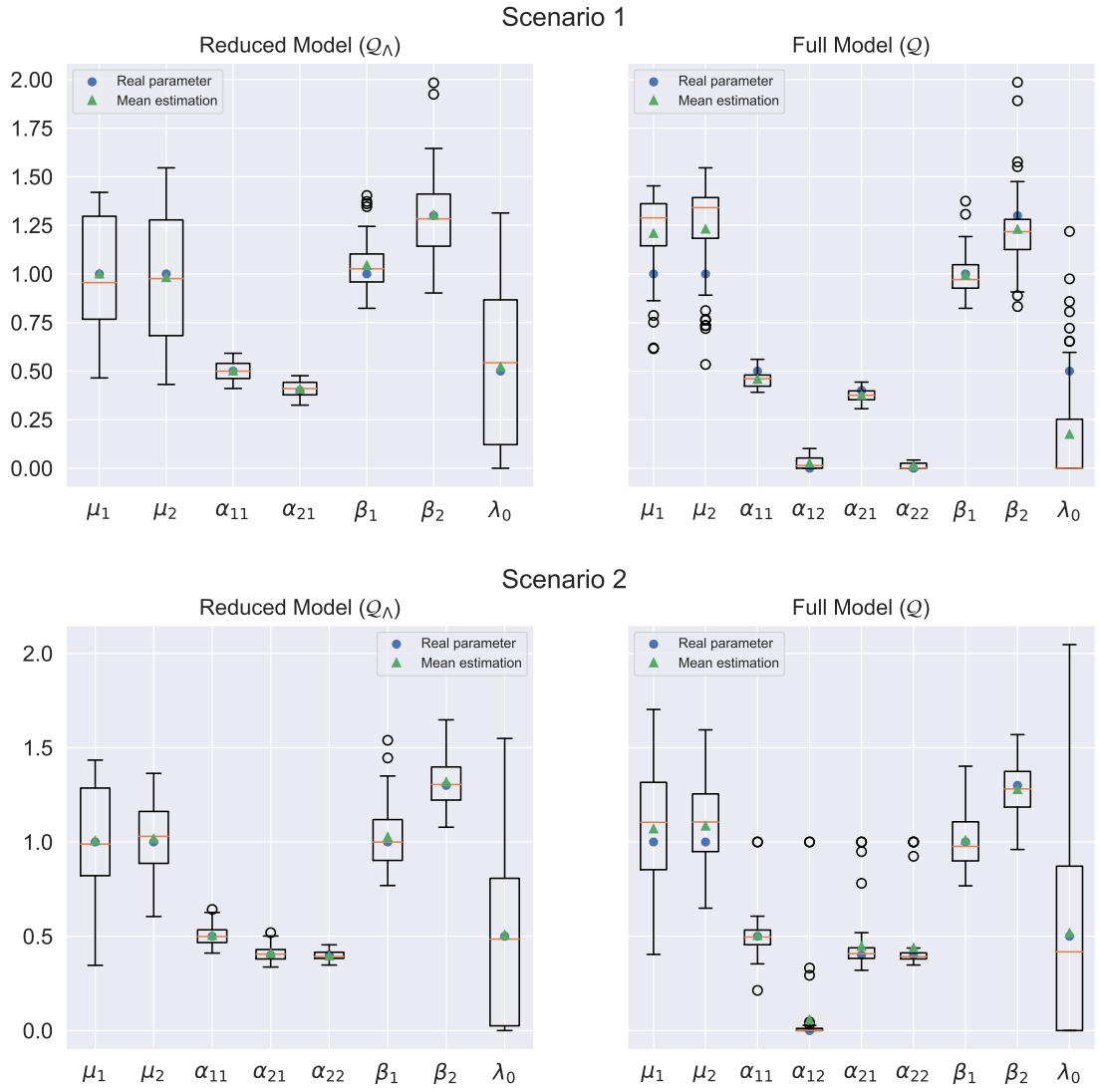

Figure 5.7: Boxplots of parameter estimations in the reduced model $\mathcal{Q}_\Lambda$ (left) and full model $\mathcal{Q}$ (right) in Scenario 1 (top) and Scenario 2 (bottom) (50 trials). Average estimation (green triangle) is to be compared to true parameter (blue point).

	α_{11}	α_{12}	α_{21}	α_{22}
Scenario 1	0.40	0.00	0.32	0.00
Scenario 2	0.41	0.00	0.34	0.35

Table 5.1: 5%-quantile of each parameter of the estimated interaction matrix $\hat{\alpha}$. The true parameter α_{12} is null in both scenarios and α_{22} is null only in Scenario 1.

it is possible to use these empirical quantiles to estimate the null interactions. An estimation

procedure when no prior information is known about the support of the interaction graph would then consist in a three-step approach: first, estimating all parameters in the complete model $\mathcal{Q}$; second, computing the 5%-quantiles for all estimated interactions parameters in matrix $\hat{\alpha}$, and defining the support of α to be entries corresponding to a positive empirical quantile; finally, re-estimating all parameters in the reduced model defined by the support of α. Let us remark that the proposed support estimation step boils down to correspond to a multiple test that, when the noisy Hawkes process has significantly more that 2 dimensions, can be corrected thanks to usual procedures such as Bonferroni and Benjamini-Hochberg methods.

5.6 Discussion

In this chapter, we propose a spectral approach to estimate the parameters of a noisy Hawkes process, the performance of which is illustrated on an extended numerical study. Although we highlight the great benefit of considering a spectral analysis when standard inference methods are not available, we also bring out identifiability issues that may arise, either from the model itself or from the spectral approach. While we exhibit several identifiable and non-identifiable scenarios in both univariate and bivariate contexts, a general result on identifiability is still to be established, in particular in higher dimensions. For this purpose, the number of non-null cross-interactions and the choice of the kernel functions appear to be key elements in order to obtain identifiability guarantees.

More generally, we believe that the spectral analysis can provide efficient estimators in many frameworks of inaccurately or partially observed data. A natural extension of this work is to consider alternative forms for the noise process, for instance an heterogeneous Poisson process. Another topic of interest is to investigate another mechanism for the noise when some points are randomly missing. This would be complementary to our work since it would allow to model both false positive and false negative occurrences. In practice, this could be of great interest for applications, especially for the tracking of epidemics. Finally, let us mention that this chapter focuses on the linear Hawkes model, which excludes notably inhibition phenomenons. Though the nonlinear framework is particularly interesting for many applications, for instance in neuroscience, all the spectral theory for point processes only exists in a linear context so that we believe that developing a spectral inference procedure for nonlinear processes would be very challenging and remains a widely open topic.

5.A Proof of Proposition 5.2.1

Lemma 5.A.1. *Let X and Y be two independent and stationary multivariate point processes with same dimension d. If they admit second order moment measures, denoted respectively $(M^X_{ij})_{1\leq i,j\leq d}$ and $(M^Y_{ij})_{1\leq i,j\leq d}$, then the process $N = X + Y$ also admits a second order moment measure, noted $(M^N_{ij})_{1\leq i,j\leq d}$, and for any pair $(A, B) \in (\mathcal{B}^c_{\mathbb{R}})^2$ and all $1 \leq i, j \leq d$:*

$$M^N_{ij}(A, B) = M^X_{ij}(A, B) + M^Y_{ij}(A, B) + (m^X_i m^Y_j + m^X_j m^Y_i)\ell_{\mathbb{R}}(A)\ell_{\mathbb{R}}(B)\,, \tag{5.15}$$

where for all $i \in \{1, \dots, d\}$ and $m^X_i = \mathbb{E}[X_i([0, 1))]$ and $m^Y_i = \mathbb{E}[Y_i([0, 1))]$. Furthermore, the reduced measure of N reads:

$$\breve{M}^N_{ij}(B) = \breve{M}^X_{ij}(B) + \breve{M}^Y_{ij}(B) + (m^X_i m^Y_j + m^X_j m^Y_i)\ell_{\mathbb{R}}(B)\,. \tag{5.16}$$

Proof. Let $(A, B) \in (\mathcal{B}_{\mathbb{R}}^c)^2$. Then for any $1 \leq i, j \leq d$:

$$\begin{aligned}
M_{ij}^N(A,B) &= \mathbb{E}[N_i(A)N_j(B)] \\
&= \mathbb{E}[(X_i(A)+Y_i(A))(X_j(B)+Y_j(B))] \\
&= \mathbb{E}[X_i(A)X_j(B)] + \mathbb{E}[X_i(A)Y_j(B)] + \mathbb{E}[X_j(A)Y_i(B)] + \mathbb{E}[Y_j(A)Y_j(B)] \\
&= M_{ij}^X(A,B) + \mathbb{E}[X_i(A)]\mathbb{E}[Y_j(B)] + \mathbb{E}[X_j(A)]\mathbb{E}[Y_i(B)] + M_{ij}^Y(A,B) \\
&= M_{ij}^X(A,B) + m_i^X m_j^Y \ell(A)\ell(B) + m_j^X m_i^Y \ell(A)\ell(B) + M_{ij}^Y(A,B)\,,
\end{aligned}$$

where the last line comes from the stationarity, which implies that $\mathbb{E}[X_i(A)] = m_i^X \ell(A)$.

By applying Equation (5.5) to the function $g(x,y) = \mathbb{1}_{x\in[0,1]}\mathbb{1}_{y-x\in B}$ for any $B \in \mathcal{B}_{\mathbb{R}}^c$, we can remark that:

$$\int_{\mathbb{R}^2} g(x,y)\, M_{ij}^N(\mathrm{d}x,\mathrm{d}y) = \int_{[0,1]} \ell_{\mathbb{R}}(\mathrm{d}x) \int_B \breve{M}_{ij}^N(\mathrm{d}u) = \breve{M}_{ij}^N(B)\,.$$

In particular this equality is also true if we replace N by X and Y. By leveraging Equation (5.15) on the left-side integral, we obtain:

$$\begin{aligned}
\int_{\mathbb{R}^2} g(x,y)\, M_{ij}^N(\mathrm{d}x,\mathrm{d}y) &= \breve{M}_{ij}^X(B) + \breve{M}_{ij}^Y(B) + (m_i^X m_j^Y + m_j^X m_i^Y) \int_{x\in[0,1]} \int_{y\in B+x} \ell_{\mathbb{R}}(\mathrm{d}y)\, \ell_{\mathbb{R}}(\mathrm{d}x) \\
&= \breve{M}_{ij}^X(B) + \breve{M}_{ij}^Y(B) + (m_i^X m_j^Y + m_j^X m_i^Y)\ell_{\mathbb{R}}([0,1])\ell_{\mathbb{R}}(B) \\
&= \breve{M}_{ij}^X(B) + \breve{M}_{ij}^Y(B) + (m_i^X m_j^Y + m_j^X m_i^Y)\ell_{\mathbb{R}}(B)\,,
\end{aligned}$$

which achieves the proof.

□

Proof of Proposition 5.2.1

By definition of the spectral density, for all $\omega \in \mathbb{R}$:

$$\begin{aligned}
f_{ij}^N(\omega) &= \int_{\mathbb{R}} \mathrm{e}^{-2\pi \mathrm{i} x\omega}\, \breve{M}_{ij}^N(\mathrm{d}x) - m_i^N m_j^N \delta(\omega) \\
&= \int_{\mathbb{R}} \mathrm{e}^{-2\pi \mathrm{i} x\omega}\, \breve{M}_{ij}^X(\mathrm{d}x) + \int_{\mathbb{R}} \mathrm{e}^{-2\pi \mathrm{i} x\omega}\, \breve{M}_{ij}^Y(\mathrm{d}x) \\
&\quad + (m_i^X m_j^Y + m_j^X m_i^Y) \int_{\mathbb{R}} \mathrm{e}^{-2\pi \mathrm{i} x\omega}\, \ell_{\mathbb{R}}(\mathrm{d}x) - (m_i^X + m_i^Y)(m_j^X + m_j^Y)\delta(\omega) \\
&= \int_{\mathbb{R}} \mathrm{e}^{-2\pi \mathrm{i} x\omega}\, \breve{M}_{ij}^X(\mathrm{d}x) - m_i^X m_j^X \delta(\omega) + \int_{\mathbb{R}} \mathrm{e}^{-2\pi \mathrm{i} x\omega}\, \breve{M}_{ij}^Y(\mathrm{d}x) - m_i^Y m_j^Y \delta(\omega) \\
&\quad + (m_i^X m_j^Y + m_j^X m_i^Y) \int_{\mathbb{R}} \mathrm{e}^{-2\pi \mathrm{i} x\omega}\, \ell_{\mathbb{R}}(\mathrm{d}x) - (m_i^X m_j^Y + m_i^Y m_j^X)\delta(\omega) \\
&= f_{ij}^X(\omega) + f_{ij}^Y(\omega) + (m_i^X m_j^Y + m_j^X m_i^Y) \int_{\mathbb{R}} \mathrm{e}^{-2\pi \mathrm{i} x\omega}\, \mathrm{d}x - (m_i^X m_j^Y + m_i^Y m_j^X)\delta(\omega)\,. \qquad (5.17)
\end{aligned}$$

By properties of the Dirac measure, for all $\omega \in \mathbb{R}$:

$$\int_{\mathbb{R}} \mathrm{e}^{-2\pi \mathrm{i} x\omega} \delta(x)\, \mathrm{d}x = 1\,,$$

and so by duality of the Fourier transform Pinsky (2008, Proposition 5.2.4.) it follows that:

$$\int_{\mathbb{R}} \mathrm{e}^{-2\pi \mathrm{i} x\omega} \, \mathrm{d}x = \delta(\omega) \, .$$

The last two terms of Equation (5.17) being equal, we obtain $f_{ij}^N = f_{ij}^X + f_{ij}^Y$.

5.B Proof of Proposition 5.3.1

We first show that equality of two spectral densities with different parameters is equivalent to a system of equations (Equations (5.18)). This result will be used then to prove both parts of Proposition 5.3.1.

Let $(\mu, \alpha, \beta, \lambda_0)$ and $(\mu', \alpha', \beta', \lambda_0')$ be two admissible 4-tuples for the exponential noisy Hawkes model. Let us assume that for all $\omega \in \mathbb{R}$:

$$f^N_{(\mu,\alpha,\beta,\lambda_0)}(\omega) = f^N_{(\mu',\alpha',\beta',\lambda_0')}(\omega).$$

Thanks to Equation (5.14), this equality implies the following system of equations:

$$\begin{cases} \dfrac{\mu}{1-\alpha} + \lambda_0 = \dfrac{\mu'}{1-\alpha'} + \lambda_0' & (\omega \to +\infty) \\ \dfrac{\mu}{1-\alpha}\dfrac{\beta^2\alpha(2-\alpha)}{\beta^2(1-\alpha)^2} + \dfrac{\mu}{1-\alpha} + \lambda_0 = \dfrac{\mu'}{1-\alpha'}\dfrac{\beta'^2\alpha'(2-\alpha')}{\beta'^2(1-\alpha')^2} + \dfrac{\mu'}{1-\alpha'} + \lambda_0' & (\omega = 0) \\ \dfrac{\mu}{1-\alpha}\dfrac{\beta^2\alpha(2-\alpha)}{\beta^2(1-\alpha)^2+4\pi^2} + \dfrac{\mu}{1-\alpha} + \lambda_0 = \dfrac{\mu'}{1-\alpha'}\dfrac{\beta'^2\alpha'(2-\alpha')}{\beta'^2(1-\alpha')^2+4\pi^2} + \dfrac{\mu'}{1-\alpha'} + \lambda_0' & (\omega = 1). \end{cases}$$

The first equality can be used to simplify the second and third equalities, leading to:

$$\begin{cases} \dfrac{\mu}{1-\alpha} + \lambda_0 = \dfrac{\mu'}{1-\alpha'} + \lambda_0' \\ \dfrac{\mu}{1-\alpha}\dfrac{\alpha(2-\alpha)}{(1-\alpha)^2} = \dfrac{\mu'}{1-\alpha'}\dfrac{\alpha'(2-\alpha')}{(1-\alpha')^2} \\ \dfrac{\mu}{1-\alpha}\dfrac{\beta^2\alpha(2-\alpha)}{\beta^2(1-\alpha)^2+4\pi^2} = \dfrac{\mu'}{1-\alpha'}\dfrac{\beta'^2\alpha'(2-\alpha')}{\beta'^2(1-\alpha')^2+4\pi^2} \, . \end{cases}$$

Now, given that $\alpha \in (0,1)$ (same for α'), the last two equalities lead to:

$$\frac{\beta^2(1-\alpha)^2}{\beta^2(1-\alpha)^2+4\pi^2} = \frac{\beta'^2(1-\alpha')^2}{\beta'^2(1-\alpha')^2+4\pi^2} \, ,$$

which in turn implies $\beta(1-\alpha) = \beta'(1-\alpha')$.

Thus, it results the following system of equations:

$$\begin{cases} \dfrac{\mu}{1-\alpha} + \lambda_0 = \dfrac{\mu'}{1-\alpha'} + \lambda_0' & \text{(5.18a)} \\ \dfrac{\mu\alpha(2-\alpha)}{(1-\alpha)^3} = \dfrac{\mu'\alpha'(2-\alpha')}{(1-\alpha')^3} & \text{(5.18b)} \\ \beta(1-\alpha) = \beta'(1-\alpha')\,. & \text{(5.18c)} \end{cases}$$

Now, Equations (5.18) lead to

$$\begin{cases} \dfrac{\mu}{1-\alpha} + \lambda_0 = \dfrac{\mu'}{1-\alpha'} + \lambda_0' \\ \dfrac{\mu}{1-\alpha}\beta^2\alpha(2-\alpha) = \dfrac{\mu'}{1-\alpha'}\beta'^2\alpha'(2-\alpha') \\ \beta(1-\alpha) = \beta'(1-\alpha')\,, \end{cases}$$

which, by Equation (5.14), implies straightforwardly $f^N_{(\mu,\alpha,\beta,\lambda_0)} = f^N_{(\mu',\alpha',\beta',\lambda_0')}$. Consequently,

$$f^N_{(\mu,\alpha,\beta,\lambda_0)} = f^N_{(\mu',\alpha',\beta',\lambda_0')} \iff \text{Equations (5.18)}\,.$$

The model $\mathcal{Q}$ is not identifiable.

Let $\tau > -\lambda_0$ and $\lambda_0' = \lambda_0 + \tau > 0$.

Then, by denoting $\kappa = \frac{\mu}{1-\alpha}\beta^2\alpha(2-\alpha)$, Equations (5.18) are equivalent to:

$$\begin{cases} \mu' = \dfrac{\beta(1-\alpha)(\frac{\mu}{1-\alpha}-\tau)}{\sqrt{\beta^2(1-\alpha)^2 + \frac{\kappa}{\frac{\mu}{1-\alpha}-\tau}}} \\ \alpha' = 1 - \dfrac{\beta(1-\alpha)}{\sqrt{\beta^2(1-\alpha)^2 + \frac{\kappa}{\frac{\mu}{1-\alpha}-\tau}}} \\ \beta' = \sqrt{\beta^2(1-\alpha)^2 + \frac{\kappa}{\frac{\mu}{1-\alpha}-\tau}}\,. \end{cases} \tag{5.19}$$

From (5.19), it is clear that for all $\tau \in \left(-\lambda_0, \frac{\mu}{1-\alpha}\right) \setminus \{0\}$, $(\mu', \alpha', \beta', \lambda_0 + \tau) \neq (\mu, \alpha, \beta, \lambda_0)$ is an admissible parameter for $\mathcal{Q}$ and $f^N_{(\mu',\alpha',\beta',\lambda_0+\tau)} = f^N_{(\mu,\alpha,\beta,\lambda_0)}$. Consequently, $\mathcal{Q}$ is not identifiable.

The reduced model defined by a triplet of admissible parameters is identifiable.

It will be shown that, for admissible parameters, $f^N_{(\mu,\alpha,\beta,\lambda_0)} = f^N_{(\mu',\alpha',\beta',\lambda_0')}$ implies

$$\mu = \mu' \iff \alpha = \alpha' \iff \beta = \beta' \iff \lambda_0 = \lambda_0'\,,$$

from which we can deduce identifiability of the four collections of models mentioned in Proposition 5.3.1. Indeed, let, for instance, $\alpha^\circ \in (0,1)$ and consider

$$\mathcal{Q}_\alpha = \left\{ f^N_{(\mu,\alpha,\beta,\lambda_0)} \in \mathcal{Q}, \alpha = \alpha^\circ \right\}\,.$$

Then for all $f^N_{(\mu,\alpha,\beta,\lambda_0)} \in \mathcal{Q}_\alpha$ and $f^N_{(\mu',\alpha',\beta',\lambda'_0)} \in \mathcal{Q}_\alpha$,

$$f^N_{(\mu,\alpha,\beta,\lambda_0)} = f^N_{(\mu',\alpha',\beta',\lambda'_0)} \implies \begin{cases} \mu = \mu' \iff \alpha = \alpha' \iff \beta = \beta' \iff \lambda_0 = \lambda'_0 \\ \alpha = \alpha^\circ = \alpha' \end{cases} \implies \begin{cases} \mu = \mu' \\ \beta = \beta' \\ \lambda_0 = \lambda'_0 \\ \alpha = \alpha' \,. \end{cases}$$

So, let us assume that $f_{(\mu,\alpha,\beta,\lambda_0)} = f_{(\mu',\alpha',\beta',\lambda'_0)}$, for some admissible parameters. As shown beforehand this implies Equations (5.18). From Equation (5.18c), we establish that

$$\alpha = \alpha' \iff \beta = \beta' \,,$$

since, $\beta > 0$ and $\alpha < 1$.

From Equation (5.18b), since $\alpha > 0$, it is clear that $\alpha = \alpha' \implies \mu = \mu'$. Conversely, if $\mu = \mu' > 0$, Equation (5.18b) becomes:

$$g(\alpha) = \frac{\alpha(2-\alpha)}{(1-\alpha)^3} = \frac{\alpha'(2-\alpha')}{(1-\alpha')^3} = g(\alpha') \,.$$

In particular, g is a strictly increasing function on $(0,1)$ and so it can be deduced that $\alpha = \alpha'$. So

$$\mu = \mu' \iff \alpha = \alpha' \,.$$

Finally, as $\mu = \mu' \iff \alpha = \alpha'$, from Equation (5.18a) we conclude that $\alpha = \alpha' \implies \lambda_0 = \lambda'_0$. Let us now assume that $\lambda_0 = \lambda'_0$, Equation (5.18a) then reads

$$\frac{\mu}{1-\alpha} = \frac{\mu'}{1-\alpha'} \,.$$

As $\mu > 0$ and $\mu' > 0$, Equation (5.18b) can be then reduced to

$$s(\alpha) := \frac{\alpha(2-\alpha)}{(1-\alpha)^2} = \frac{\alpha(2-\alpha)}{(1-\alpha)^2} = s(\alpha') \,,$$

where s is a strictly increasing function on $(0,1)$ and so $\alpha = \alpha'$. With this we have proved that $\alpha = \alpha' \iff \lambda_0 = \lambda'_0$ which achieves the proof.

5.C Proof of Proposition 5.3.2

Let $(\mathcal{H}^N_t)_{t\geq 0}$ (respectively $(\mathcal{H}^H_t)_{t\geq 0}$) be the natural filtration associated with $(\lambda^N(t))_{t\geq 0}$ (respectively $(\lambda^H(t))_{t\geq 0}$). Let us now consider the conditional survival function of the first non-negative jump τ_1 of N given the past (before 0): for all $t \geq 0$,

$$\begin{aligned} \mathbb{P}\left(\tau_1 > t \mid \mathcal{H}^N_0\right) &= \mathbb{P}\left(N(t) = 0 \mid \mathcal{H}^N_0\right) \\ &= \mathbb{P}\left(P(t) = 0, H(t) = 0 \mid \mathcal{H}^N_0\right) \\ &= \mathbb{P}\left(P(t) = 0 \mid \mathcal{H}^N_0\right) \mathbb{P}\left(H(t) = 0 \mid \mathcal{H}^N_0\right) && \text{(by independence)} \\ &= \mathbb{P}\left(P(t) = 0\right) \mathbb{P}\left(H(t) = 0 \mid \mathcal{H}^H_0\right) && \text{(by definition)} \\ &= \mathrm{e}^{-\lambda_0 t} e^{-\mu(t-u_t) - u_t \lambda^H(0)} \,, \end{aligned}$$

with $u_t = (1 - \mathrm{e}^{-\beta t})\beta^{-1}$ and where we have used that, given that $H(t) = 0$, for all $u \in [0, t]$,

$$\lambda^H(u) = \mu + \left(\lambda^H(0) - \mu\right) \mathrm{e}^{-\beta u}.$$

Let us remark that the last equality can also be deduced from Dassios et al. (2011, Corollary 3.3) with the same notation u_t:

$$\mathbb{P}\left(H(t) = 0 \mid \mathcal{H}_0^H\right) = \exp\left(-\int_0^{u_t} \frac{\mu\beta v}{1-\beta v}\,\mathrm{d}v\right) \mathrm{e}^{-u_t \lambda^H(0)}.$$

It appears that $\mathbb{P}\left(\tau_1 > t \mid \mathcal{H}_0^N\right)$ depends on the past only through $\lambda^H(0)$. But $\lambda^H(0) = \lambda^N(0) - \lambda_0$ and $\lambda^N(0)$ is distributed according to the stationary distribution of $\left(\lambda^N(t)\right)_{t\geq 0}$. It results that $\lambda^H(0)$ is distributed according to the stationary distribution of $\left(\lambda^H(t)\right)_{t\geq 0}$, the Laplace transform of which is given by Dassios et al. (2011, Corollary 3.1). Plugging in this result, we have:

$$\begin{aligned}\mathbb{P}\left(\tau_1 > t\right) &= \mathrm{e}^{-\lambda_0 t - \mu(t-u_t)} \mathbb{E}\left[\mathrm{e}^{-u_t \lambda^H(0)}\right] \\ &= \mathrm{e}^{-\lambda_0 t - \mu(t-u_t)} \exp\left(-\int_0^{u_t} \frac{\mu\beta v}{\beta v + \mathrm{e}^{-\alpha\beta v} - 1}\mathrm{d}v\right).\end{aligned}$$

For ease of derivation, let $G : t \in \mathbb{R}_{\geq 0} \mapsto -\log \mathbb{P}\left(\tau_1 > t\right)$. The function G is differentiable and has derivative D given by:

$$D(t) = \lambda_0 + \mu\left(1 - \mathrm{e}^{-\beta t}\right) + \frac{\mu\left(1 - \mathrm{e}^{-\beta t}\right)\mathrm{e}^{-\beta t}}{\exp\left(-\alpha\left(1 - \mathrm{e}^{-\beta t}\right)\right) - \mathrm{e}^{-\beta t}}.$$

Now define in the same manner τ_1', G', and D' from the process N' with parameter $(\mu', \alpha', \beta', \lambda_0')$, with $\lambda^{N'}(0)$ being distributed according to the stationary distribution of $\left(\lambda^{N'}(t)\right)_{t\geq 0}$, and assume that N and N' have the same distribution. Then it follows that both τ_1 and τ_1' have the same distribution, so that $G(t) = G'(t)$ for all $t \geq 0$. Since G and G' are everywhere differentiable and have the same initial value $G(0) = G'(0) = 0$, it results that $D(t) = D'(t)$ for all $t \geq 0$.

We want to establish a system of four equations satisfied by the parameters that leads to the equality of the 4-tuples. First, noting that $\lim_{t\to\infty} D(t) = \lambda_0 + \mu$, we get

$$\lambda_0 + \mu = \lambda_0' + \mu'. \tag{5.20}$$

Then, since $\lim_{t\to 0} D(t) = \lambda_0 + \mu/(1-\alpha)$, we get, using Equation (5.20),

$$\frac{\mu\alpha}{1-\alpha} = \frac{\mu'\alpha'}{1-\alpha'}. \tag{5.21}$$

To highlight two other equations on the parameters, we establish the Taylor expansion of $D(t)$ around $t = 0$ up to order 2. After some calculation, we find that

$$D(t) = \lambda_0 + \frac{\mu}{1-\alpha} + \frac{\mu\alpha}{1-\alpha}\left(\frac{1}{2}\beta t\frac{\alpha-2}{1-\alpha} + \frac{1}{12}\beta^2 t^2 \frac{\alpha^3 - \alpha^2 - 3\alpha + 6}{(1-\alpha)^2} + o\big(t^2\big)\right).$$

From the first-order term of the expansion, Equation (5.21) and the equality of D and D', we find that

$$\beta\frac{\alpha-2}{1-\alpha} = \beta'\frac{\alpha'-2}{1-\alpha'}. \tag{5.22}$$

Then the second-order term of the expansion can be rewritten

$$\frac{\mu\alpha}{1-\alpha}\frac{1}{12}\left(\frac{\beta(\alpha-2)}{1-\alpha}\right)^2\frac{\alpha^3-\alpha^2-3\alpha+6}{(\alpha-2)^2},$$

so that from Equations (5.21) and (5.22), we find that

$$g(\alpha)=\frac{\alpha^3-\alpha^2-3\alpha+6}{(\alpha-2)^2}=\frac{\alpha'^3-\alpha'^2-3\alpha'+6}{(\alpha'-2)^2}=g(\alpha')\,. \tag{5.23}$$

$\alpha\mapsto g(\alpha)$ can be shown to be strictly increasing for $\alpha\in(0,1)$, so that Equation (5.23) yields that $\alpha=\alpha'$. With this remark, one can easily show from the system composed by Equations (5.20)–(5.23) that the tuples $(\mu,\alpha,\beta,\lambda_0)$ and $(\mu',\alpha',\beta',\lambda_0')$ are equal.

5.D Proof of Proposition 5.3.3

Lemma 5.D.1. *Assume all moments of h are finite: $\forall n\geq 0, m_n=\int t^n h(t)\,\mathrm{d}t<\infty$. Then the spectral density f^N of the noisy Hawkes process N has a Taylor expansion around 0 given by:*

$$\forall t\in\mathbb{R}:\quad f^N(t)=\frac{\mu}{(1-\alpha)^3}+\frac{\mu}{(1-\alpha)^3}\sum_{q\geq 1}\left(\frac{\alpha}{1-\alpha}\sum_{n\geq 1}b_n t^{2n}\right)^q+\lambda_0\,, \tag{5.24}$$

with

$$b_n=\frac{(-1)^n(2\pi)^{2n}a_n}{(2n)!}\,,\qquad \text{and}\qquad a_n=2m_{2n}-\frac{\alpha}{1-\alpha}\sum_{k=1}^{2n-1}\binom{2n}{k}(-1)^k m_k m_{2n-k}\,. \tag{5.25}$$

Proof. Introduce the Taylor expansion of the Fourier transform of h:

$$\forall t\in\mathbb{R}:\quad \tilde{h}(t)=\sum_{n\geq 0}(\mathrm{i}\tau)^n\frac{m_n}{n!},$$

where $\tau=-2\pi t$. Then, given that $m_0=1$,

$$\begin{aligned}
\left|1-\alpha\tilde{h}(t)\right|^2&=\left(1-\alpha\sum_{n\geq 0}(\mathrm{i}\tau)^n\frac{m_n}{n!}\right)\left(1-\alpha\sum_{n\geq 0}(-\mathrm{i}\tau)^n\frac{m_n}{n!}\right)\\
&=1-2\alpha\sum_{n\geq 0}(-1)^n\tau^{2n}\frac{m_{2n}}{(2n)!}+\alpha^2\sum_{n\geq 0}\sum_{k=0}^{n}(-1)^k(\mathrm{i}\tau)^n\frac{m_{n-k}m_k}{(n-k)!k!}\\
&=1-\alpha\sum_{n\geq 0}2m_{2n}\frac{(-1)^n\tau^{2n}}{(2n)!}+\alpha^2\sum_{n\geq 0}\sum_{k=0}^{2n}(-1)^k m_k m_{2n-k}\binom{2n}{k}\frac{(-1)^n\tau^{2n}}{(2n)!}\\
&=(1-\alpha)^2\left(1-\frac{\alpha}{1-\alpha}\sum_{n\geq 1}a_n\frac{(-1)^n\tau^{2n}}{(2n)!}\right).
\end{aligned}$$

Inverting this expression and taking the Taylor expansion of $x\mapsto(1-x)^{-1}$ around 0 yields the desired result. □

Remark. *For $n = 1, 2$, a_n is given by*

$$\begin{cases} a_1 &= 2m_2 + \frac{\alpha}{1-\alpha} 2m_1^2 \\ a_2 &= 2m_4 + \frac{\alpha}{1-\alpha}\left(8m_1 m_3 - 6m_2^2\right), \end{cases}$$

so that the Taylor expansion of f^N up to order 5 is:

$$f^N(t) = a + c_1 t^2 + c_2 t^4 + o(t^5),$$

with

$$\begin{aligned} a &= \frac{\mu}{(1-\alpha)^3} + \lambda_0 \\ c_1 &= 4\frac{\mu\alpha\pi^2}{(1-\alpha)^4}\left[-m_2 - \frac{\alpha}{1-\alpha} m_1^2\right] \\ c_2 &= 16\frac{\mu\alpha\pi^4}{(1-\alpha)^4}\left[\frac{m_4}{12} + \frac{\alpha}{1-\alpha}\left(\frac{m_1 m_3}{3} + \frac{3m_2^2}{4}\right) + \left(\frac{\alpha}{1-\alpha}\right)^2 2m_2 m_1^2 + \left(\frac{\alpha}{1-\alpha}\right)^3 m_1^4\right]. \end{aligned}$$

Proof of Proposition 5.3.3

Let $(\mu, \alpha, \phi, \lambda_0)$ and $(\mu', \alpha', \phi', \lambda_0')$ be two admissible 4-tuples for the uniform noisy Hawkes model $\mathcal{R}$, and assume that, for all $\omega \in \mathbb{R}$,

$$f^N_{(\mu,\alpha,\phi,\lambda_0)}(\omega) = f^N_{(\mu',\alpha',\phi',\lambda_0')}(\omega).$$

First, noting that $\lim_{\omega\to\infty} f^N_{(\mu,\alpha,\phi,\lambda_0)}(\omega) = \lambda_0 + \mu/(1-\alpha)$, we get

$$\lambda_0 + \frac{\mu}{1-\alpha} = \lambda_0' + \frac{\mu'}{1-\alpha'}. \tag{5.26}$$

Then, since $f^N_{(\mu,\alpha,\phi,\lambda_0)}(0) = \lambda_0 + \mu/(1-\alpha)^3$ and using Equation (5.26), we get:

$$\frac{\mu\alpha(2-\alpha)}{(1-\alpha)^3} = \frac{\mu'\alpha'(2-\alpha')}{(1-\alpha')^3}. \tag{5.27}$$

Now, since the Taylor expansions, given by Equation (5.24), of $f^N_{(\mu,\alpha,\phi,\lambda_0)}$ and $f^N_{(\mu',\alpha',\phi',\lambda_0')}$ around 0 coincide, their respective Taylor coefficients $(c_1, c_2, \ldots)$ and $(c_1', c_2', \ldots)$ are equal. Plugging in the moments of the uniform distribution on $[0, \phi]$,

$$m_n = \frac{\phi^n}{n+1}, \qquad \text{for all } n \geq 0,$$

the first order coefficient c_1 can be written

$$c_1 = -\frac{\pi}{3}\frac{\mu\alpha(2-\alpha)}{(1-\alpha)^3}\phi^2\frac{4-\alpha}{(1-\alpha)^2(2-\alpha)},$$

so that, with the use of Equation (5.27), we get

$$\phi^2\frac{4-\alpha}{(1-\alpha)^2(2-\alpha)} = \phi'^2\frac{4-\alpha'}{(1-\alpha')^2(2-\alpha')}. \tag{5.28}$$

Similarly, by plugging the moments in the second order coefficient c_2, we get

$$c_2 = \frac{\pi^4}{15}\frac{\mu\alpha(2-\alpha)}{(1-\alpha)^3}\left(\phi^2\frac{4-\alpha}{(1-\alpha)^2(2-\alpha)}\right)^2\frac{(2-\alpha)(\alpha^3-8\alpha^2+18\alpha+4)}{(4-\alpha)^2}\,,$$

so that, by Equations (5.27) and (5.28),

$$g(\alpha) = \frac{(2-\alpha)(\alpha^3-8\alpha^2+18\alpha+4)}{(4-\alpha)^2} = \frac{(2-\alpha')(\alpha'^3-8\alpha'^2+18\alpha'+4)}{(4-\alpha')^2} = g(\alpha')\,. \tag{5.29}$$

But $\alpha \mapsto g(\alpha)$ can be shown to be strictly increasing for $\alpha \in (0,1)$, so that Equation (5.29) yields that $\alpha = \alpha'$. Finally, it is easily proven from the system composed by Equations (5.26)–(5.29) that the tuples $(\mu, \alpha, \beta, \lambda_0)$ and $(\mu', \alpha', \beta', \lambda_0')$ are equal.

5.E Proof of Proposition 5.4.1

Let N be a bivariate noisy Hawkes process parametrised by the exponential model $\mathcal{Q}_\Lambda$. We will prove that if either of the conditions in Proposition 5.4.1 is fulfilled then $\mathcal{Q}_\Lambda$ is not identifiable.

- **Condition 1:** Let
$$\Lambda = \left\{\begin{pmatrix}\alpha_{11} & 0\\ 0 & \alpha_{22}\end{pmatrix}, 0 \le \alpha_{11}, \alpha_{22} < 1\right\}.$$
For any admissible $\theta = (\mu, \alpha, \beta, \lambda_0)$, the spectral matrix $\mathbf{f}_\theta^N$ is diagonal such that, for any integer $i \in \{1,2\}$:
$$f_\theta^N{}_{ii}(\omega) = \frac{\mu_i}{1-\alpha_{ii}}\left(\frac{\beta_i^2\alpha_{ii}(2-\alpha_{ii})}{\beta_i^2(1-\alpha_{ii})^2+4\pi^2\omega^2}\right)+\left(\frac{\mu_i}{1-\alpha_{ii}}+\lambda_0\right).$$
We will show that there exists an admissible $\theta' = (\mu', \alpha', \beta', \lambda_0')$ such that $\theta' \neq \theta$ and $\mathbf{f}_\theta^N = \mathbf{f}_{\theta'}^N$. Let $\tau > -\lambda_0$ and $\lambda_0' = \lambda_0 + \tau$.
For every $i \in \{1,2\}$, on the one hand, if $\alpha_{ii} \neq 0$ then, as shown in the univariate case (Appendix 5.B), letting $\kappa_i = \frac{\mu_i}{1-\alpha_{ii}}\beta_i^2\alpha_{ii}(2-\alpha_{ii})$ and defining the parameters:
$$\begin{cases}\mu_i' = \dfrac{\beta_i(1-\alpha_{ii})(\frac{\mu_i}{1-\alpha_{ii}}-\tau)}{\sqrt{\beta_i^2(1-\alpha_{ii})^2+\frac{\kappa_i}{\frac{\mu_i}{1-\alpha_{ii}}-\tau}}}\\ \alpha_{ii}' = 1-\dfrac{\beta_i(1-\alpha_{ii})}{\sqrt{\beta_i^2(1-\alpha_{ii})^2+\frac{\kappa_i}{\frac{\mu_i}{1-\alpha_{ii}}-\tau}}}\\ \beta_i' = \sqrt{\beta_i^2(1-\alpha_{ii})^2+\frac{\kappa_i}{\frac{\mu_i}{1-\alpha_{ii}}-\tau}}\,,\end{cases} \tag{5.30}$$
leads to $f_\theta^N{}_{ii} = f_{\theta'}^N{}_{ii}$.
On the other hand, if $\alpha_{ii} = 0$ then, $f_\theta^N{}_{ii} = \mu_i + \lambda_0$ and it is enough to consider:
$$\begin{cases}\mu_i' = \mu_i - \tau\\ \alpha_{ii}' = 0\\ \beta_i' = \beta_i\,,\end{cases}$$
to get $f_\theta^N{}_{ii} = f_{\theta'}^N{}_{ii}$.

In both cases, for $\tau \in \left(-\lambda_0, \min_{1\leq i\leq 2}\left\{\frac{\mu_i}{1-\alpha_{ii}}\right\}\right) \setminus \{0\}$, we obtain an admissible parameter $\theta' \neq \theta$ and $\mathbf{f}_\theta^N = \mathbf{f}_{\theta'}^N$. Thus, the model is not identifiable.

- **Condition 2 and 3:** Without loss of generality, let

$$\Lambda = \left\{ \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ 0 & 0 \end{pmatrix}, 0 < \alpha_{11} < 1, \alpha_{12} > 0 \right\}.$$

For any admissible parameter $\theta = (\mu, \alpha, \beta, \lambda_0)$ and for any $\omega \in \mathbb{R}$, the spectral matrix reads:

$$\mathbf{f}_\theta^N(\omega) = \begin{bmatrix} f_\theta^N{}_{11}(\omega) & f_\theta^N{}_{12}(\omega) \\ f_\theta^N{}_{21}(\omega) & f_\theta^N{}_{22}(\omega) \end{bmatrix} = \begin{bmatrix} f_\theta^N{}_{11}(\omega) & \dfrac{\mu_2\beta_1\alpha_{12}}{\beta_1(1-\alpha_{11}) - 2\pi\mathrm{i}\omega} \\ \dfrac{\mu_2\beta_1\alpha_{12}}{\beta_1(1-\alpha_{11}) + 2\pi\mathrm{i}\omega} & \mu_2 + \lambda_0 \end{bmatrix}$$

with

$$f_\theta^N{}_{11}(\omega) = \left(\frac{\mu_1}{1-\alpha_{11}} + \frac{\mu_2\alpha_{12}}{1-\alpha_{11}}\right) \frac{\beta_1^2\alpha_{11}(2-\alpha_{11})}{\beta_1^2(1-\alpha_{11})^2 + 4\pi^2\omega^2} + \frac{\mu_2\beta_1^2\alpha_{12}^2}{\beta_1^2(1-\alpha_{11})^2 + 4\pi^2\omega^2} + \frac{\mu_1}{1-\alpha_{11}} + \frac{\mu_2\alpha_{12}}{1-\alpha_{11}} + \lambda_0 .$$

Let us introduce the following constants:

$$\begin{cases} A = \mu_2 + \lambda_0 \\ B = \mu_2\beta_1\alpha_{12} \\ C = \beta_1(1-\alpha_{11}) \\ D = \frac{\mu_1+\mu_2\alpha_{12}}{1-\alpha_{11}} + \lambda_0 \\ E = (\frac{\mu_1+\mu_2\alpha_{12}}{1-\alpha_{11}})\beta_1^2\alpha_{11}(2-\alpha_{11}) + \mu_2\beta_1^2\alpha_{12}^2 , \end{cases}$$

which allow us to rewrite the spectral matrix as:

$$\mathbf{f}_\theta^N(\omega) = \begin{bmatrix} \dfrac{E}{C^2 + 4\pi^2\omega^2} + D & \dfrac{B}{C - 2\pi\mathrm{i}\omega} \\ \dfrac{B}{C + 2\pi\mathrm{i}\omega} & A \end{bmatrix}. \tag{5.31}$$

Let $\tau \in \mathbb{R} \setminus \{\mu_2, (\mu_1 + \mu_2\alpha_{12})/(1-\alpha_{11})\}$ and:

$$\kappa_\tau = \frac{(\mu_1 + \mu_2\alpha_{12})\alpha_{11}(2-\alpha_{11})(\mu_2 - \tau) - \tau\mu_2^2\alpha_{12}^2(1-\alpha_{11})}{(\mu_2 - \tau)(\mu_1 + \mu_2\alpha_{12} - \tau(1-\alpha_{11}))} .$$

Now, consider the parameter $\theta' = (\mu', \alpha', \beta', \lambda_0')$, defined as follows:

$$\begin{cases} \mu_1' = \frac{\mu_1 - \tau(1-\alpha_{11})}{\sqrt{(1-\alpha_{11})^2+\kappa_\tau}} \\ \mu_2' = \mu_2 - \tau \\ \alpha_{11}' = 1 - \frac{1-\alpha_{11}}{\sqrt{(1-\alpha_{11})^2+\kappa_\tau}} \\ \alpha_{12}' = \frac{\mu_2\alpha_{12}}{(\mu_2-\tau)\sqrt{(1-\alpha_{11})^2+\kappa_\tau}} \\ \beta_1' = \beta_1\sqrt{(1-\alpha_{11})^2+\kappa_\tau} \\ \lambda_0' = \lambda_0 + \tau\,. \end{cases} \tag{5.32}$$

The goal will be to show that there exists τ such that θ' is well defined, admissible for $\mathcal{Q}_\Lambda$, such that $\theta \neq \theta'$ and $\mathbf{f}_\theta^N = \mathbf{f}_{\theta'}^N$.

First, in order for θ' to be an admissible parameter, let us remark that $\rho(S) = \alpha_{11}$ and so we obtain the following constrains:

$$\begin{cases} \mu_1' > 0 \\ \mu_2' > 0 \\ 0 < \alpha_{11}' < 1 \\ \alpha_{12}' > 0 \\ \beta_1' > 0 \\ \lambda_0' > 0 \end{cases} \iff \begin{cases} \mu_1 - \tau(1-\alpha_{11}) > 0 \\ \mu_2 - \tau > 0 \\ \frac{1-\alpha_{11}}{\sqrt{(1-\alpha_{11})^2+\kappa_\tau}} < 1 \\ (1-\alpha_{11})^2 + \kappa_\tau > 0 \\ \lambda_0 + \tau > 0 \end{cases} \iff \begin{cases} \tau < \frac{\mu_1}{1-\alpha_{11}} \\ \tau < \mu_2 \\ \kappa_\tau > 0 \\ \tau > -\lambda_0\,. \end{cases} \tag{5.33}$$

Since $\tau < \mu_1/(1-\alpha_{11}) \implies \mu_1 + \mu_2\alpha_{12} - \tau(1-\alpha_{11}) > 0$, the third inequality becomes

$$\tau < \frac{(\mu_1+\mu_2\alpha_{12})\alpha_{11}(2-\alpha_{11})\mu_2}{(\mu_1+\mu_2\alpha_{12})\alpha_{11}(2-\alpha_{11}) + \mu_2^2(1-\alpha_{11})\alpha_{12}^2}\,.$$

Then for

$$\tau \in \left(-\lambda_0, \min\left\{\frac{\mu_1}{1-\alpha_{11}}, \mu_2, \frac{(\mu_1+\mu_2\alpha_{12})\alpha_{11}(2-\alpha_{11})\mu_2}{(\mu_1+\mu_2\alpha_{12})\alpha_{11}(2-\alpha_{11}) + \mu_2^2(1-\alpha_{11})\alpha_{12}^2}\right\}\right) \setminus \{0\}\,,$$

the right-hand side of Equations (5.33) is verified and so θ' defined by Equations (5.32) is well defined, admissible for $\mathcal{Q}_\Lambda$ and $\theta' \neq \theta$.

Then, we can show that:

$$\begin{cases} \mu_2' + \lambda_0' = \mu_2 + \lambda_0 = A \\ \mu_2'\beta_1'\alpha_{12}' = \mu_2\beta_1\alpha_{12} = B \\ \beta_1'(1-\alpha_{11}') = \beta_1(1-\alpha_{11}) = C \\ \frac{\mu_1'+\mu_2'\alpha_{12}'}{1-\alpha_{11}'} + \lambda_0' = \frac{\mu_1+\mu_2\alpha_{12}}{1-\alpha_{11}} + \lambda_0 = D \\ \left(\frac{\mu_1'+\mu_2'\alpha_{12}'}{1-\alpha_{11}'}\right)\beta_1'^2\alpha_{11}'(2-\alpha_{11}') + \mu_2'\beta_1'^2\alpha_{12}'^2 = \left(\frac{\mu_1+\mu_2\alpha_{12}}{1-\alpha_{11}}\right)\beta_1^2\alpha_{11}(2-\alpha_{11}) + \mu_2\beta_1^2\alpha_{12}^2 = E\,. \end{cases}$$

This assures that, for all $\omega \in \mathbb{R}$:

$$\mathbf{f}^N_{\theta'}(\omega) = \begin{bmatrix} \dfrac{E}{C^2+4\pi^2\omega^2} + D & \dfrac{B}{C-2\pi\mathrm{i}\omega} \\ \dfrac{B}{C+2\pi\mathrm{i}\omega} & A \end{bmatrix} = \mathbf{f}^N_{\theta}(\omega)\,,$$

Thus, the model is not identifiable.

5.F Proof of Proposition 5.4.2, Situations 1 and 2

Without loss of generality, let

$$\Lambda = \left\{ \begin{pmatrix} \alpha_{11} & 0 \\ \alpha_{21} & 0 \end{pmatrix}, 0 \leq \alpha_{11} < 1, \alpha_{21} > 0 \right\}.$$

Then, for any admissible parameter $\theta = (\mu, \alpha, \beta, \lambda_0)$ and all $\omega \in \mathbb{R}$:

$$\begin{cases} {f^N_\theta}_{11}(\omega) = \dfrac{\mu_1}{1-\alpha_{11}} \dfrac{\beta_1^2 + 4\pi^2\omega^2}{\beta_1^2(1-\alpha_{11})^2 + 4\pi^2\omega^2} + \lambda_0 \\ {f^N_\theta}_{12}(\omega) = \dfrac{\mu_1}{1-\alpha_{11}} \dfrac{\beta_1^2 + 4\pi^2\omega^2}{\beta_1^2(1-\alpha_{11})^2 + 4\pi^2\omega^2} \dfrac{\alpha_{21}\beta_2}{\beta_2 + 2\pi\mathrm{i}\omega} \\ {f^N_\theta}_{22}(\omega) = \dfrac{\mu_1}{1-\alpha_{11}} \dfrac{\beta_1^2 + 4\pi^2\omega^2}{\beta_1^2(1-\alpha_{11})^2 + 4\pi^2\omega^2} \dfrac{\alpha_{21}^2\beta_2^2}{\beta_2^2 + 4\pi^2\omega^2} + \mu_2 + \dfrac{\mu_1\alpha_{21}}{1-\alpha_{11}} + \lambda_0\,, \end{cases}$$

and ${f^N_\theta}_{21}(\omega) = \overline{{f^N_\theta}_{12}(\omega)}$.

These expressions can be reformulated as follows:

$$\begin{cases} {f^N_\theta}_{11}(\omega) = m_1^H \dfrac{1}{|1 - \tilde{h}_{\theta\,11}(\omega)|^2} + \lambda_0 \\ {f^N_\theta}_{12}(\omega) = ({f^N_\theta}_{11}(\omega) - \lambda_0)\tilde{h}_{\theta\,21}(\omega) \\ {f^N_\theta}_{22}(\omega) = \dfrac{|{f^N_\theta}_{12}(\omega)|^2}{{f^N_\theta}_{11}(\omega) - \lambda_0} + \lim\limits_{\omega' \to +\infty} {f^N_\theta}_{22}(\omega')\,. \end{cases}$$

where $m_1^H = \mu_1/(1-\alpha_{11})$, $m_2^H = \mu_2 + \mu_1\alpha_{21}/(1-\alpha_{11})$ and $\lim_{\omega'\to+\infty} {f^N_\theta}_{22}(\omega') = m_2^H + \lambda_0$.

Let $\theta' = (\mu', \alpha', \beta', \lambda_0')$ be an admissible parameter such that $\mathbf{f}^N_\theta = \mathbf{f}^N_{\theta'}$. For $\omega = 0$, it comes ${f^N_\theta}_{22}(0) = {f^N_{\theta'}}_{22}(0)$, which implies that

$$\frac{|{f^N_\theta}_{12}(0)|^2}{{f^N_\theta}_{11}(0) - \lambda_0} = \frac{|{f^N_{\theta'}}_{12}(0)|^2}{{f^N_{\theta'}}_{11}(0) - \lambda_0'}\,.$$

Since $|{f^N_\theta}_{12}(0)|^2 = \mu_1\alpha_{21}(1-\alpha_{11})^{-3} \neq 0$ (as $\alpha_{11} < 1$ and $\alpha_{21} > 0$), ${f^N_\theta}_{12}(0) = {f^N_{\theta'}}_{12}(0)$ and ${f^N_\theta}_{11}(0) = {f^N_{\theta'}}_{11}(0)$, it results that:

$$\lambda_0 = \lambda_0'\,.$$

Now, for all $\omega \in \mathbb{R}$, since ${f^N_\theta}_{12}(\omega) = {f^N_{\theta'}}_{12}(\omega)$, it comes $({f^N_\theta}_{11}(\omega) - \lambda_0)\tilde{h}_{\theta\,21}(\omega) = ({f^N_{\theta'}}_{11}(\omega) - \lambda_0')\tilde{h}_{\theta',21}(\omega)$, then $\tilde{h}_{\theta\,21}(\omega) = \tilde{h}_{\theta',21}(\omega)$. By Equation (5.13), it results that:

$$\frac{\alpha_{21}\beta_2}{\beta_2 + 2\pi\mathrm{i}\omega} = \frac{\alpha_{21}'\beta_2'}{\beta_2' + 2\pi\mathrm{i}\omega}\,.$$

For $\omega = 0$, this simplifies to

$$\alpha_{21} = \alpha'_{21}\,.$$

For $\omega = 1$, as $\alpha_{21} \neq 0$ it comes $\beta_2/(\beta_2 + 2\pi\mathrm{i}) = \beta'_2/(\beta'_2 + 2\pi\mathrm{i})$, then

$$\beta_2 = \beta'_2\,.$$

Now, by considering the limits as $\omega \to \infty$, we obtain the two following equations:

$$\begin{cases} \dfrac{\mu_1}{1-\alpha_{11}} + \lambda_0 = \dfrac{\mu'_1}{1-\alpha'_{11}} + \lambda_0 \\ \mu_2 + \dfrac{\mu_1\alpha_{21}}{1-\alpha_{11}} + \lambda_0 = \mu'_2 + \dfrac{\mu'_1\alpha_{21}}{1-\alpha'_{11}} + \lambda_0\,, \end{cases}$$

which, using that $\mu_1 > 0$, can be simplified to:

$$\begin{cases} \dfrac{\mu_1}{1-\alpha_{11}} = \dfrac{\mu'_1}{1-\alpha'_{11}} \\ \mu_2 = \mu'_2\,. \end{cases}$$

Now, as ${f_\theta^N}_{11}(0) = \mu_1/(1-\alpha_{11})^3$, we have the two following equations:

$$\begin{cases} \dfrac{\mu_1}{1-\alpha_{11}} = \dfrac{\mu'_1}{1-\alpha'_{11}} \\ \dfrac{\mu_1}{(1-\alpha_{11})^3} = \dfrac{\mu'_1}{(1-\alpha'_{11})^3}\,, \end{cases}$$

which imply

$$\begin{cases} \mu_1 = \mu'_1 \\ \alpha_{11} = \alpha'_{11}\,. \end{cases}$$

At last, since ${f_\theta^N}_{11}(1) = {f_{\theta'}^N}_{11}(1)$ and $\mu_1 > 0$,

$$\frac{\beta_1^2 + 4\pi^2}{\beta_1^2(1-\alpha_{11})^2 + 4\pi^2} = \frac{{\beta'_1}^2 + 4\pi^2}{{\beta'_1}^2(1-\alpha_{11})^2 + 4\pi^2}\,,$$

which implies

$$\alpha_{11}(2-\alpha_{11})\beta_1^2 = \alpha_{11}(2-\alpha_{11}){\beta'_1}^2\,.$$

Either $\alpha_{11} > 0$, so the previous equation leads to

$$\beta_1 = \beta'_1\,,$$

or $\alpha_{11} = 0$, so $\alpha'_{11} = \alpha_{11} = 0$ and $\beta_1 = \beta'_1 = 1$ (since θ and θ' are admissible for the model) .

This proves that $\mathbf{f}_\theta^N = \mathbf{f}_{\theta'}^N \implies \theta = \theta'$, which achieves the proof.

5.G Proof of Proposition 5.4.2, Situations 3 and 4

Without loss of generality, let

$$\Lambda = \left\{ \begin{pmatrix} \alpha_{11} & 0 \\ \alpha_{21} & \alpha_{22} \end{pmatrix}, 0 < \alpha_{11} < 1, \alpha_{21} > 0, 0 \leq \alpha_{22} < 1 \right\}.$$

Then, for any admissible parameter $\theta = (\mu, \alpha, \beta, \lambda_0)$ and all $\omega \in \mathbb{R}$:

$$\begin{cases} f^N_{\theta\ 11}(\omega) = \dfrac{\mu_1}{1-\alpha_{11}} \dfrac{\beta_1^2 \alpha_{11}(2-\alpha_{11})}{\beta_1^2(1-\alpha_{11})^2 + 4\pi^2\omega^2} + \dfrac{\mu_1}{1-\alpha_{11}} + \lambda_0 \\ f^N_{\theta\ 12}(\omega) = \left(f^N_{\theta\ 11}(\omega) - \lambda_0\right) \dfrac{\alpha_{21}\beta_2}{\beta_2^2(1-\alpha_{22})^2 + 4\pi^2\omega^2} \left(\beta_2(1-\alpha_{22}) - 2\pi \mathrm{i}\omega\right) \\ f^N_{\theta\ 22}(\omega) = \left(\dfrac{\mu_1\alpha_{21}}{(1-\alpha_{22})(1-\alpha_{11})} + \dfrac{\mu_2}{1-\alpha_{22}} \right) \dfrac{\beta_2^2 + 4\pi^2\omega^2}{\beta_2^2(1-\alpha_{22})^2 + 4\pi^2\omega^2} + \\ \qquad \dfrac{\mu_1\alpha_{21}^2\beta_2^2}{1-\alpha_{11}} \dfrac{\beta_1^2 + 4\pi^2\omega^2}{(\beta_2^2(1-\alpha_{22})^2 + 4\pi^2\omega^2)(\beta_1^2(1-\alpha_{11})^2 + 4\pi^2\omega^2)} + \lambda_0 \,, \end{cases}$$

and $f^N_{\theta\ 21}(\omega) = \overline{f^N_{\theta\ 12}(\omega)}$.

Let $\theta' = (\mu', \alpha', \beta', \lambda_0')$ be an admissible parameter such that $\mathbf{f}^N_\theta = \mathbf{f}^N_{\theta'}$. We start by showing that $\lambda_0 = \lambda_0'$. From $\mathfrak{Re}\left(f^N_{\theta\ 12}(1)\right) = \mathfrak{Re}\left(f^N_{\theta'\ 12}(1)\right)$ and $\mathfrak{Im}\left(f^N_{\theta\ 12}(1)\right) = \mathfrak{Im}\left(f^N_{\theta'\ 12}(1)\right)$, the following system of equations is established:

$$\begin{cases} \dfrac{(f^N_{\theta\ 11}(1) - \lambda_0)\alpha_{21}\beta_2}{\beta_2^2(1-\alpha_{22})^2 + 4\pi^2} \beta_2(1-\alpha_{22}) = \dfrac{(f^N_{\theta'\ 11}(1) - \lambda_0')\alpha_{21}'\beta_2'}{{\beta_2'}^2(1-\alpha_{22}')^2 + 4\pi^2} \beta_2'(1-\alpha_{22}') \\ \dfrac{(f^N_{\theta\ 11}(1) - \lambda_0)\alpha_{21}\beta_2}{\beta_2^2(1-\alpha_{22})^2 + 4\pi^2} = \dfrac{(f^N_{\theta'\ 11}(1) - \lambda_0')\alpha_{21}'\beta_2'}{{\beta_2'}^2(1-\alpha_{22}')^2 + 4\pi^2} \,. \end{cases} \tag{5.34}$$

Since

$$f^N_{\theta\ 11}(1) - \lambda_0 = \frac{\mu_1}{1-\alpha_{11}} \frac{\beta_1^2 + 4\pi^2}{\beta_1^2(1-\alpha_{11})^2 + 4\pi^2} \neq 0 \,,$$

and $\alpha_{21}\beta_2 > 0$, Equations (5.34) imply that

$$\beta_2(1-\alpha_{22}) = \beta_2'(1-\alpha_{22}') \,. \tag{5.35}$$

Now, from $f^N_{\theta\ 12}(0) = \left(f^N_{\theta\ 11}(0) - \lambda_0\right)\alpha_{21}/(1-\alpha_{22})$ and $f^N_{\theta\ 11}(0) - \lambda_0 = \mu_1/(1-\alpha_{11})^3 \neq 0$, it comes $\alpha_{21} = f^N_{\theta\ 12}(0)(1-\alpha_{22})/\left(f^N_{\theta\ 11}(0) - \lambda_0\right)$ and the second equality of Equations (5.34) implies that

$$\frac{f^N_{\theta\ 11}(1) - \lambda_0}{f^N_{\theta\ 11}(0) - \lambda_0} \frac{f^N_{\theta\ 12}(0)\beta_2(1-\alpha_{22})}{\beta_2^2(1-\alpha_{22})^2 + 4\pi^2} = \frac{f^N_{\theta'\ 11}(1) - \lambda_0'}{f^N_{\theta'\ 11}(0) - \lambda_0'} \frac{f^N_{\theta'\ 12}(0)\beta_2'(1-\alpha_{22}')}{{\beta_2'}^2(1-\alpha_{22}')^2 + 4\pi^2} \,. \tag{5.36}$$

By Equation (5.35)

$$\frac{f^N_{\theta\ 12}(0)\beta_2(1-\alpha_{22})}{\beta_2^2(1-\alpha_{22})^2 + 4\pi^2} = \frac{f^N_{\theta'\ 12}(0)\beta_2'(1-\alpha_{22}')}{{\beta_2'}^2(1-\alpha_{22}')^2 + 4\pi^2} \,,$$

and since $\alpha_{21} > 0$, ${f_\theta^N}_{12}(0) \neq 0$ and Equation (5.36) leads to

$$\frac{{f_\theta^N}_{11}(1) - \lambda_0}{{f_\theta^N}_{11}(0) - \lambda_0} = \frac{{f_{\theta'}^N}_{11}(1) - \lambda_0'}{{f_{\theta'}^N}_{11}(0) - \lambda_0'}\,.$$

Since ${f_\theta^N}_{11}$ is strictly decreasing, ${f_\theta^N}_{11}(1) - {f_\theta^N}_{11}(0) = {f_{\theta'}^N}_{11}(1) - {f_{\theta'}^N}_{11}(0) \neq 0$, and it comes $\lambda_0 = \lambda_0'$.

Now, the expression of ${f_\theta^N}_{11}$ is similar to that of the univariate spectral density f_θ^N in Proposition 5.3.1. Thus, following the proof in Appendix 5.B, from ${f_\theta^N}_{11} = {f_{\theta'}^N}_{11}$ and $\lambda_0 = \lambda_0'$ (since $\alpha_{11} > 0$ and $\alpha_{11}' > 0$) it comes:

$$\begin{cases} \mu_1 = \mu_1' \\ \alpha_{11} = \alpha_{11}' \\ \beta_1 = \beta_1'\,. \end{cases}$$

Next, from ${f_\theta^N}_{12}(0) = {f_{\theta'}^N}_{12}(0)$, it comes:

$$\left({f_\theta^N}_{11}(0) - \lambda_0\right)\frac{\alpha_{21}}{1-\alpha_{22}} = \left({f_{\theta'}^N}_{11}(0) - \lambda_0'\right)\frac{\alpha_{21}'}{1-\alpha_{22}'}\,,$$

which implies, with ${f_\theta^N}_{11}(0) - \lambda_0 = {f_{\theta'}^N}_{11}(0) - \lambda_0' \neq 0$:

$$\frac{\alpha_{21}}{1-\alpha_{22}} = \frac{\alpha_{21}'}{1-\alpha_{22}'}\,. \tag{5.37}$$

Now, from $\lim_{\omega'\to\infty} {f_\theta^N}_{22}(\omega') = \lim_{\omega'\to\infty} {f_{\theta'}^N}_{22}(\omega')$,

$$\frac{\mu_1\alpha_{21}}{(1-\alpha_{11})(1-\alpha_{22})} + \frac{\mu_2}{1-\alpha_{22}} = \frac{\mu_1\alpha_{21}'}{(1-\alpha_{11})(1-\alpha_{22}')} + \frac{\mu_2'}{1-\alpha_{22}'}\,,$$

and leveraging Equation (5.37), it comes:

$$\frac{\mu_2}{1-\alpha_{22}} = \frac{\mu_2'}{1-\alpha_{22}'}\,. \tag{5.38}$$

Moreover,

$$\begin{aligned} {f_\theta^N}_{22}(0) &= \left(\frac{\mu_1\alpha_{21}}{(1-\alpha_{11})(1-\alpha_{22})} + \frac{\mu_2}{1-\alpha_{22}} + \frac{\mu_1\alpha_{21}^2}{(1-\alpha_{11})^3}\right)\frac{1}{(1-\alpha_{22})^2} + \lambda_0 \\ &= \left(\frac{\mu_1}{1-\alpha_{11}}\frac{\alpha_{21}}{1-\alpha_{22}} + \frac{\mu_2}{1-\alpha_{22}}\right)\frac{1}{(1-\alpha_{22})^2} + \frac{\mu_1}{(1-\alpha_{11})^3}\frac{\alpha_{21}^2}{(1-\alpha_{22})^2} + \lambda_0\,. \end{aligned}$$

Thus, by Equations (5.37) and (5.38), we obtain from ${f_\theta^N}_{22}(0) = {f_{\theta'}^N}_{22}(0)$:

$$\frac{1}{(1-\alpha_{22})^2} = \frac{1}{(1-\alpha_{22}')^2}\,,$$

which implies

$$\alpha_{22} = \alpha_{22}'\,.$$

To conclude, from Equations (5.35), (5.38) and (5.37), $\beta_2 = \beta_2'$, $\mu_2 = \mu_2'$ and $\alpha_{21} = \alpha_{21}'$. This proves that $\mathbf{f}_\theta^N = \mathbf{f}_{\theta'}^N \implies \theta = \theta'$, which achieves the proof.

6

A numerical exploration of thinned Hawkes processes through spectral theory

Standard estimation methods for Hawkes processes with missing information are either intractable or computationally expensive. Nonetheless this is a common scenario when studying real-world data where detection procedures may miss event times, which can be modelled through the thinning alteration of point processes. We assume that we observe a thinned version of a Hawkes process and we propose a parametric estimation procedure by maximising a version of the log-likelihood from spectral analysis theory. We propose identifiability conditions for the ensuing statistical model for the classical exponential kernel functions and the performance of our estimator is studied on simulated data. Our inference method gives access to a subsampling method for our point processes via the thinning operation. We propose to use this paradigm to enhance the performance of a penalised version of the spectral estimator. Numerically, we show that, in a small-sized sample context with a short observation windows, this estimator performs better than alternative approaches.

Outline of the current chapter

6.1 Introduction

The Hawkes process model, introduced for the first time in Hawkes (1971), is a past-dependent point process used to characterise self-exciting interactions among event times. Its applications are numerous among various fields with examples including sociology (Linderman et al. 2014), biology (Gupta et al. 2018; Lambert et al. 2018; Rizoiu et al. 2018), seismology (Ogata 1988; Ogata 1998) and finance (Bacry et al. 2013; Bacry et al. 2015; Hawkes 2018), to mention a few.

As a result, there are many efforts focused around proposing estimation methods. These include estimators obtained through likelihood maximisation (Ogata 1978; Ozaki 1979), least-squares error minimisation (Reynaud-Bouret et al. 2014; Bacry et al. 2020) and method of moments (Da Fonseca et al. 2013), in both parametric and non-parametric settings.

Most of these methods tend to study the model under the assumption that there is no missing data in the samples used for estimation. However, if this assumption does not hold, it is common for estimators to be biased. Some works in the literature consider different missing or imperfect data contexts such as binned observations in Cheysson et al. (2022), jittering (or random displacement) as exemplified in Antoniadis et al. (2006) and Bonnet et al. (2022b), superposition as in Bonnet et al. (2024).

Jittering and superposition are two of the three most common alterations on point processes, along with thinning. Thinning is more often used in simulation scenarios for point processes, with the main example being Ogata's thinning procedure (Ogata 1981) or to accelerate estimations as a method of downsampling (Li et al. 2019). Nonetheless, the use of thinning to model missing data is scarcer in the literature, which could be useful in order to account for sporadic detection errors.

A main work focusing on noised data for both point processes and proposing numerical applications to Hawkes processes can be found in Lund et al. (2000). They propose a conditional log-likelihood model for point process that have been noised through the three classic operations in the field: jittering, superposition and thinning. Their method consists in establishing an expression of such a log-likelihood when the base point process (or an estimation of it) is available. They obtain an approximation of a maximiser through some simplifications and an iterative optimisation on a small space of parameters. The main obstacle that arises in this context is that it is essential to have some information on the ground process in order to estimate the parameters associated with each noise.

In this chapter, we take inspiration from the spectral estimation methods as presented in Cheysson et al. (2022) and Bonnet et al. (2024). Spectral theory for point processes was properly introduced for the first time in Bartlett (1963), focusing around the Bartlett spectrum. Works concerning the spectral study and estimation of point processes subsequently appeared in the literature with numerous theoretical approaches in temporal and spatial contexts (Daley 1971; Tuan 1981; Mugglestone et al. 2001; Rajala et al. 2023). Practical applications are scarcer, with an application to multiple clustering models, including Hawkes processes, in Adamopoulos (1976), in Karavasilis et al. (2007) to stationary bivariate processes and in Roueff et al. (2019) to a locally stationary version of the Hawkes process.

In our work, we explicit the form of the Bartlett spectrum for a point process thinned with independent and identically distributed probabilities. In particular, we follow a similar scheme as the one presented for the superposition of point processes in Bonnet et al. (2024) to establish an estimation method by maximising the spectral log-likelihod, with a particular interest in the univariate Hawkes process with exponential kernel.

A second contribution of our work focuses on the thinning operation as a tool to provide a subsampling scheme for point processes. This approach has been explored mainly in spatial contexts (Møller et al. 2003; Moradi et al. 2019; Cronie et al. 2024) showing the advantages of

considering thinning for point processes, either for improving estimations, or for proposing cross-validation procedures in spatio-temporal contexts, as in Coeurjolly et al. (2024). Our effort in this chapter consists in illustrating its use for spectral estimators when combined with a penalised version of the log-likelihood.

After a general presentation of the mathematical setting in Section 6.2 for both thinning and spectral theory of point processes, we introduce the results concerning the spectral density function for thinned point processes in Section 6.3.1. We leverage this result in the case of a Hawkes process with exponential kernel in Section 6.3.2 along with an identifiability condition for the statistical model. Finally, we illustrate the performance of our estimation procedure in the context of missing data in Section 6.4.1, and for enhancing estimation through ℓ_2 penalisation with subsampling in Section 6.4.2

6.2 Mathematical setting

6.2.1 The p-thinned point processes

In this chapter, we will focus on the stationary univariate Hawkes process on the real line $\mathbb{R}$, noted H. The Hawkes process can be characterised by its conditional intensity function defined, for all $t \in \mathbb{R}$, by:

$$\lambda(t) = \mu + \int_{-\infty}^{t} h(t-s)\, H(\mathrm{d}s) = \mu + \sum_{T_k \leq t} h(t - T_k)\,, \tag{6.1}$$

where $\mu > 0$ is the baseline intensity, and $h\colon \mathbb{R} \to \mathbb{R}_{\geq 0}$ is the kernel function modelling the self-exciting behaviour of past events $(T_k)_{k\in\mathbb{Z}}$.

In order for H to be a point process (i.e. an a.s. finite measure in any bounded set B), a sufficient and necessary condition (Hawkes 1971) is that

$$\|h\|_1 = \int_{\mathbb{R}} h(t)\,\mathrm{d}t < 1\,.$$

Let $\mathcal{B}^c_{\mathbb{R}}$ denote the set of bounded Borel sets on $\mathbb{R}$, and let, for any $B \in \mathcal{B}^c_{\mathbb{R}}$:

$$H(B) = \sum_{k\in\mathbb{Z}} \mathbb{1}_{T_k \in B}$$

be the number of event times in B.

In this chapter, we will study a thinned version of process H defined as follows:

Definition 6.2.1. For any $p \in (0, 1]$, we denote H_p a p-thinning of a point process H, defined for any $B \in \mathcal{B}^c_{\mathbb{R}}$ as:

$$H_p(B) = \sum_{k\in\mathbb{Z}} \mathbb{1}_{T_k \in B} Z_k$$

where $(Z_k)_{k\in\mathbb{Z}}$ is an i.i.d. collection of Bernoulli random variables of parameter p. We refer to H_p as a p-thinned point process.

In practice, the event times of process H_p correspond to a subset of $(T_k)_{k\in\mathbb{Z}}$ where each point is erased with a random probability $1 - p$. Setting $p = 1$ keeps all original points from process H which is not a proper thinning per se, but we include this value in order to lighten notations in the incoming sections.

6.2.2 Spectral theory for point processes

For a point process H, we define the first and second-order measures M_1 and M_2, for any $A, B \in \mathcal{B}^c_{\mathbb{R}}$, by:

$$M_1(A) = \mathbb{E}[H(A)]\,, \qquad M_2(A,B) = \mathbb{E}[H(A)H(B)]\,.$$

Under stationarity conditions, it follows that (Daley et al. 2003, Proposition 8.1.I), for any $A, B \in \mathcal{B}^c_{\mathbb{R}}$:

$$M_1(A) = m_1 \ell_{\mathbb{R}}(A)\,,$$

where $m_1 = \mathbb{E}[H([0,1])]$ is usually known as the average intensity of process H.

The spectral analysis approach of point processes is based on the Bartlett spectrum Γ (Bartlett 1963), a complex-valued measure on $\mathbb{R}$ associated with the second-order moment measure M_2. To introduce it properly, let us consider the Schwartz space $\mathcal{S}$ defined as:

$$\mathcal{S} = \left\{ f \in C^\infty, \forall k \in \{1,2,\dots\}, \forall r \in \{1,2,\dots\}, \sup_{x\in\mathbb{R}} \left| x^r f^{(k)}(x) \right| < \infty \right\}\,,$$

where C^∞ denotes the set of infinitely differentiable functions f from $\mathbb{R}$ to $\mathbb{R}$, and $f^{(k)}$ denotes the k^{th}-order derivative of f. For any $f \in \mathcal{S}$, we define its Fourier transform $\tilde{f}$, for any $\omega \in \mathbb{R}$, by:

$$\tilde{f}(\omega) = \int_{\mathbb{R}} f(x) \mathrm{e}^{-2\pi \mathrm{i} x\omega}\, \mathrm{d}x\,.$$

The Bartlett spectrum of a point process H is the measure Γ such that, for any $\varphi \in \mathcal{S}$ (Brémaud et al. 2005, Definition 2):

$$\mathbb{V}\left[\int_{\mathbb{R}} \varphi(x)\, H(\mathrm{d}x) \right] = \int_{\mathbb{R}} |\tilde{\varphi}(\omega)|^2\, \Gamma(\mathrm{d}\omega)\,. \tag{6.2}$$

Existence of such a measure is established for any stationary point process (Daley et al. 2003, Proposition 8.2.I.(a)), and by polarisation of Equation (6.2), we obtain for any $\varphi, \psi \in \mathcal{S}$:

$$\mathrm{cov}\left(\int_{\mathbb{R}} \varphi(x)\, H(\mathrm{d}x), \int_{\mathbb{R}} \psi(x)\, H(\mathrm{d}x) \right) = \int_{\mathbb{R}} \tilde{\varphi}(\omega)\tilde{\psi}(-\omega)\, \Gamma(\mathrm{d}\omega)\,. \tag{6.3}$$

Whenever Γ is an absolutely continous measure, we denote $f : \mathbb{R} \to \mathbb{C}$ its Radon-Nikodym derivative, known as the spectral density of process H.

Another important quantity in the spectral theory of point processes is the periodogram $I^T \colon \mathbb{R} \to \mathbb{C}$. For a realisation $(T_k)_{k=1:N(T)}$ of a process H in a time window $[0,T]$, the periodogram is defined, for all $\omega \in \mathbb{R}$, as:

$$I^T(\omega) = \frac{1}{T} \sum_{k=1}^{N(T)} \sum_{l=1}^{N(t)} \mathrm{e}^{-2\pi \mathrm{i}\omega(T_k - T_l)}\,.$$

Furthermore, for any sequence $(\omega_k)_{k=1:M}$, such that $\omega_k \neq \omega_l$, for all integers $k \neq l$, the random variables $(I^T(\omega_k))_{k=1:M}$ are asymptotically independent and exponentially distributed (Tuan 1981) with respective parameter $(1/f(\omega_k))_{k=1:M}$.

In this chapter, we will work on a parametric setting, so let a statistical model $\mathcal{P}$ for the spectral density be defined as:

$$\mathcal{P} = \{ f_\theta \colon \mathbb{R} \to \mathbb{C}, \theta \in \Theta \}\,.$$

We can then define the *spectral* log-likelihood $\ell_T(\theta)$ for an observation $(T_k)_{k=1:N(T)}$ of H as:

$$\ell_T(\theta) = -\frac{1}{T}\sum_{k=1}^{M} \log(f_\theta(\omega_k)) + \frac{I^T(\omega_k)}{f_\theta(\omega_k)}\,. \tag{6.4}$$

We can then introduce the Whittle estimator $\hat{\theta}$ (Whittle 1952) as:

$$\hat{\theta} = \arg\max_{\theta\in\Theta}\ \ell_T(\theta)\,.$$

6.3 Parametric estimation of a thinned process

6.3.1 The spectrum of a thinned point process

The goal of this chapter is to establish a parametric estimation method for the observation of a thinned Hawkes process. We will leverage the study of spectral quantities on marked point processes, which can be found in Brémaud et al. (2002) and Brémaud et al. (2005).

For this purpose, let us introduce an alternative way of viewing the p-thinning of a point process through marked point process theory (as also done in Cronie et al. (2024)). In this section, H will denote a stationary point process on $\mathbb{R}$ with Bartlett spectrum Γ^H and spectral density function f^H.

We define the marked point process $\bar{H}$ associated with H, with marks Z_k on a metric space $\mathcal{K}$ as the collection of points $(T_k, Z_k)_{k\in\mathbb{Z}} \in (\mathbb{R}\times\mathcal{K})^{\mathbb{Z}}$ (see Daley et al. (2003, Chapter 6.4) for a more thorough presentation of marked point processes). The random marks Z_k are usually used to represent underlying information on the event times of a point process H, which is often referred to as the *ground* process. In our work, we will restrict ourselves to the case where the random variables $(Z_k)_{k\in\mathbb{Z}}$ are independent and identically distributed. In this setting, process $\bar{H}$ is well-defined (Daley et al. 2003, 6.4.IV(a)).

We can then view a p-thinning of H as a marked version $\bar{H}$ where $\mathcal{K} = \{0,1\}$ and the $(Z_k)_{k\in\mathbb{Z}}$ are a collection of Bernoulli random variables of parameter p. This way we may define the thinned process H_p, for any $B \in \mathcal{B}^c$, as:

$$H_p(B) = \bar{H}(B\times\{1\})\,.$$

Under this scope, we will apply the results of Brémaud et al. (2005), that we adapt to our notations.

Theorem 6.3.1 (Brémaud et al. (2005, Theorem 2))**.** *Let H be a stationary point process with Bartlett spectrum measure Γ and let $\bar{H}$ be a marked version of H with i.i.d. marks Z_k with shared distribution Z on a metric space $\mathcal{K}$. Let $\varphi^\star, \psi^\star$ be measurable functions from $\mathbb{R}\times\mathcal{K}\to\mathbb{R}$, such that:*

- $\displaystyle\int_{\mathbb{R}} \mathbb{E}\left[|\varphi^\star(x,Z)|\right]\mathrm{d}x < +\infty\,, \qquad \int_{\mathbb{R}} \mathbb{E}\left[|\psi^\star(x,Z)|\right]\mathrm{d}x < +\infty\,.$
- $\displaystyle\int_{\mathbb{R}} \mathbb{E}\left[\varphi^\star(x,Z)^2\right]\mathrm{d}x < +\infty\,, \qquad \int_{\mathbb{R}} \mathbb{E}\left[\psi^\star(x,Z)^2\right]\mathrm{d}x < +\infty\,.$
- *By denoting $\bar{\varphi} : x \to \mathbb{E}\left[\varphi^\star(x,Z)\right]$ and $\bar{\psi} : x \to \mathbb{E}\left[\psi^\star(x,Z)\right]$,*

$$\bar{\varphi}, \bar{\psi} \in \mathcal{S}\,.$$

Then, it follows that:

$$\operatorname{cov}\left(\sum_{k\in\mathbb{Z}}\varphi^\star(T_k,Z_k),\sum_{k\in\mathbb{Z}}\psi^\star(T_k,Z_k)\right)=\int_{\mathbb{R}}\tilde{\tilde{\varphi}}(\omega)\tilde{\tilde{\psi}}(-\omega)\,\Gamma^H(\mathrm{d}\omega) + \int_{\mathbb{R}}\operatorname{cov}\left(\tilde{\varphi}^\star(\omega,Z),\tilde{\psi}^\star(-\omega,Z)\right)M_1(\mathrm{d}\omega)\,, \tag{6.5}$$

where, for any $\omega\in\mathbb{R}$, $\tilde{\varphi}^\star(\omega,Z)$ (resp. $\tilde{\psi}^\star(\omega,Z)$) denotes the Fourier transform of the function $x\to\varphi^\star(x,Z)$ (resp. $x\to\psi^\star(x,Z)$).

The utility of this result lies on the link that it establishes between the covariance of a marked point process $\bar{H}$ and the Bartlett spectrum of its ground process H. This allows to establish an expression of the spectral density of the p-thinning of a process H, as presented in the following proposition:

Proposition 6.3.1. *Let H be a stationary point process admitting a Bartlett spectrum Γ^H defined as in Equation (6.2). We assume that Γ^H is absolutely continuous and we note f^H the spectral density function. Let m_1 be the average intensity of H.*

For any $p\in(0,1]$, let H_p be a p-thinning of H as explicited in Definition 6.2.1.

Then, H_p admits a spectral density function, denoted f^{H_p}, such that for any $\omega\in\mathbb{R}$:

$$f^{H_p}(\omega)=p^2f^H(\omega)+p(1-p)m_1\,. \tag{6.6}$$

Proof. The proof is given in Appendix 6.A □

Let us remark that Equation (6.6) can be found in Daley et al. (2003, Equation 8.3.5), where they obtain this result by term identification of a bivariate point process. The proof presented in this chapter is an alternative way of establishing this expression, illustrating the usefulness of leveraging the spectral theory of point processes.

6.3.2 The p-thinned Hawkes process estimation and the exponential kernel

For the rest of this chapter, H will denote a stationary Hawkes process defined by the intensity λ as in Equation (6.1).

As shown in Hawkes (1971), the spectral density of a Hawkes process reads, for all $\omega\in\mathbb{R}$:

$$f^H(\omega)=\frac{\mu}{(1-\|h\|_1)|1-\tilde{h}(\omega)|^2}\,.$$

Corollary 6.3.1, presents the spectral density of the p-thinned univariate Hawkes process:

Corollary 6.3.1. *Let H be a stationary univariate Hawkes process with baseline intensity $\mu>0$, and kernel function $h\colon\mathbb{R}\to\mathbb{R}_{>0}$, as defined by Equation (6.1). Let $p\in(0,1]$ and H_p a p-thinning of H. Then the spectral density f^{H_p} of H_p is:*

$$\forall\omega\in\mathbb{R}\,,\quad f^{H_p}(\omega)=p^2\frac{\mu}{(1-\|h\|_1)|1-\tilde{h}(\omega)|^2}+p(1-p)\frac{\mu}{1-\|h\|_1}\,. \tag{6.7}$$

Proof. The proof is direct by applying Proposition 6.3.1 and with the expression of the Hawkes process spectral density and m_1 the mean number of points of a Hawkes process given by $m_1=\mu/(1-\|h\|_1)$. □

We can define the statistical model:

$$\mathcal{P} = \{f_\theta \colon \mathbb{R} \to \mathbb{C}, \theta = (\mu, \gamma, p) \in \Theta\} \,,$$

where γ is a parameter of the interaction function h. From now on, we will drop the superscript on the spectral density function as specifying θ removes any ambiguity concerning the value of p and whether the point process is thinned or not. It is then possible to estimate θ by maximising the spectral log-likelihood (Equation (6.4)).

Let us now focus on the exponential interaction function, defined as:

$$\forall t \in \mathbb{R}\,, \quad h(t) = \alpha\beta \mathrm{e}^{-\beta t} \mathbb{1}_{t \geq 0}\,,$$

where $\alpha \in (0,1)$ and $\beta > 0$.

The Fourier transform of h is explicit and reads:

$$\tilde{h}(\omega) = \frac{\alpha\beta}{\beta + 2\pi \mathrm{i}\omega}\,,$$

and so Equation (6.7) reduces to:

$$\forall \omega \in \mathbb{R}\,, \quad f_\theta(\omega) = \frac{\mu p}{1-\alpha}\left(1 + p\frac{\beta^2\alpha(2-\alpha)}{\beta^2(1-\alpha)^2 + 4\pi^2\omega^2}\right)\,. \tag{6.8}$$

In this context, the statistical model of f_θ is:

$$\mathcal{Q} = \{f_\theta \colon \mathbb{R} \to \mathbb{C}, \theta = (\mu, \alpha, \beta, p) \in \mathbb{R}_{>0} \times (0,1) \times \mathbb{R}_{>0} \times (0,1]\}\,.$$

Similarly to the superposition case shown in Bonnet et al. (2024, Proposition 3.2), model $\mathcal{Q}$ is not identifiable in the general setting but identifiability can be retrieved, as shown in Proposition 6.3.2, as long as one parameter of the model is fixed.

Proposition 6.3.2. *The model $\mathcal{Q}$ is identifiable if and only if one of the parameters in the 4-uplet (μ, α, β, p) is fixed.*

In particular, for any admissible parameter $\theta = (\mu, \alpha, \beta, p)$, for any $\kappa \in (0, 1/p]$, let:

$$\begin{cases} \mu' = \dfrac{\mu}{\kappa(1-\alpha)\sqrt{1+\frac{1}{\kappa}\left(\frac{1}{(1-\alpha)^2}-1\right)}} \\ \alpha' = 1 - \dfrac{1}{\sqrt{1+\frac{1}{\kappa}\left(\frac{1}{(1-\alpha)^2}-1\right)}} \\ \beta' = \beta(1-\alpha)\sqrt{1+\frac{1}{\kappa}\left(\frac{1}{(1-\alpha)^2}-1\right)} \\ p' = \kappa p\,. \end{cases} \tag{6.9}$$

Then, $\theta' = (\mu', \alpha', \beta', p')$ is an admissible parameter such that $f_\theta = f_{\theta'}$.

Proof. The proof is given in Appendix 6.B. □

The non-identifiability of the full model (four unknown parameters) limits the implementation of an estimation method for $\theta \in \Theta$, but this problem is avoided as long as one of the parameters is known beforehand. In this work we will focus on the scenario where the value of p is known in advance, and so the previous result ensures that, for any $p^\star \in (0,1]$, the reduced model:

$$\mathcal{Q}_p = \{ f_\theta \colon \mathbb{R} \to \mathbb{C}, \theta = (\mu, \alpha, \beta, p) \in \Theta = \mathbb{R}_{>0} \times (0,1) \times \mathbb{R}_{>0} \times (0,1], p = p^\star \} \,,$$

is identifiable. We can then define an estimator $\hat{\theta} = (\hat{\mu}, \hat{\alpha}, \hat{\beta}, \hat{p})$ of θ with $\hat{p} = p^\star$, as presented at the end of Section 6.2.2, as:

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \ell_T(\theta) = \arg\max_{\theta \in \Theta} \left(-\frac{1}{T} \sum_{k=1}^{M} \log(f_\theta(\omega_k)) + \frac{I^T(\omega_k)}{f_\theta(\omega_k)} \right) .$$

6.4 Numerical illustrations

In this section, we will study the estimation of p-thinned Hawkes processes obtained by maximising the spectral lof-likelihood defined in Equation (6.4) under the mathematical model $\mathcal{Q}_p$, and unless said otherwise, p will be fixed to the true value $p^\star$ used for thinning.

Simulation is done by leveraging the cluster and branching representation of self-exciting Hawkes processes (Hawkes et al. 1974), which is explicited in Møller et al. (2005, Algorithm 2). In order to approximate the stationarity condition of the Hawkes process, for any simulation window $[0, T]$ considered in our study, we will initially simulate the process in $[-100, T]$ to avoid edge effects.

Our numerical procedure is implemented in Python and optimisation of all considered functions in this section will be done via the minimize method with the L-BFGS-B solver (Byrd et al. 1995).

To evaluate the performance of our estimator $\hat{\theta}$ against any set of parameters θ, we will consider its relative ℓ_2 error defined as:

$$\frac{\|\hat{\theta} - \theta\|_2}{\|\theta\|_2} \,.$$

6.4.1 Spectral estimator for missing data

In this subsection, we will study the performance of estimator $\hat{\theta}$ for different levels of p. We simulate 100 different Hawkes processes $(H^1, \dots, H^{100})$ with exponential kernel with parameters:

$$\mu = 1.25 \,, \quad \alpha = 0.5 \,, \quad \beta = 1.5 \,,$$

and the following grid of values for thinning:

$$p \in \{0.1, 0.2, \dots, 0.9\} \,.$$

We perform then a p-thinning of each simulation obtaining the thinned sample $(H_p^1, \dots, H_p^{100})$. For each level p, we will consider an observation window $[0, T_p]$ with T_p chosen so that the average number of points after thinning is constant equal to 5000. By remarking that:

$$\mathbb{E}[H_p([0, T_p])] = p\mathbb{E}[H([0, T_p])] = \frac{p\mu}{1 - \alpha} T_p \,,$$

it follows that $T_p = 5000(1 - \alpha)/(p\mu)$. This ensures that for each scenario, all estimations take into account the same amount of information on average.

We will consider two different estimators in order to show the advantage of taking thinning into account. For each H_p^i, the first estimator $\hat{\theta}_p$ is obtained by estimating in the correct model

$\mathcal{Q}_p$ and we will compare it to the estimator $\hat{\theta}_1$ obtained by maximising Equation (6.4) for $p = 1$. Estimator $\hat{\theta}_1$ corresponds to making the assumption that observations $(H_p^1, \dots, H_p^{100})$ are univariate Hawkes processes that have not been thinned.

Figure 6.1 represents the boxplots of relative square errors for each parameter estimation in $\hat{\theta}_p$ and $\hat{\theta}_1$ in logarithmic scale. As expected, $\hat{\theta}_p$ performs significantly better than $\hat{\theta}_1$, in particular for higher levels of thinning (lower values of p). We can see that estimations are especially improved when estimating μ and α, which are the two parameters controlling the average number of points of H, that we recall here:

$$\mathbb{E}[H([0,T])] = m_1 T_p = \frac{\mu T_p}{1-\alpha}\,.$$

Let us remark that the performance of $\hat{\theta}$ is fairly consistent accross all considered values of p. This is particularly encouraging as the more points are erased, interaction between points are attenuated and so the optimisation task is more difficult.

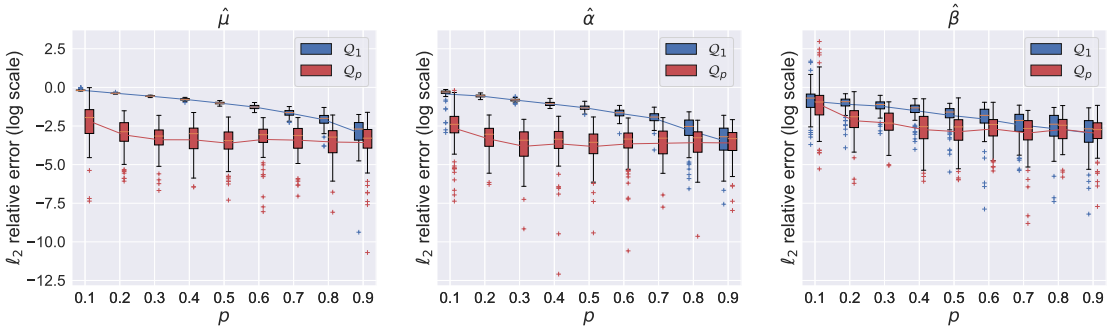

Figure 6.1: Boxplots of ℓ_2 relative error in log-scale with respect to level of thinning p. Estimations for each parameter on the correct thinned model (red boxes) and on the model without accounting for thinning (blue boxes).

6.4.2 p-thinning as a subsampling method

In this section, we assume that we are provided with a single and short observation H of a Hawkes process for the same parameters:

$$\mu = 1.25\,, \quad \alpha = 0.5\,, \quad \beta = 1.5\,,$$

in an observation window of $[0, 50]$. The lack of multiple repetitions and the small window size tend to negatively affect estimation procedures for point processes.

A common practice to improve estimations in such contexts is to consider a penalised version of the to-be-optimised function, in our case, the spectral log-likelihood (Equation (6.4)). We define then the ℓ_2-penalised spectral log-likelihood, for any penalisation parameter $L \geq 0$, as:

$$\ell_T(\theta) - L\|\theta\|_2\,. \tag{6.10}$$

A penalised estimator can then by obtained by maximising this quantity. In this penalised setting, we will illustrate how combining the ℓ_2 penalisation with a subsampling procedure by thinning provides better results than other approaches.

Benchmark estimators. As our work in this chapter concerns mainly the study of spectral approaches, we will solely take into consideration estimators obtained through such means. We will compare then four estimators:

- $\hat{\theta}$: the estimator obtained by maximising the non-penalised spectral log-likelihood (Equation (6.4)) on the single observation H.
- $\hat{\theta}^L$: the estimator obtained by maxising the penalised spectral log-likelihood (Equation (6.10)) on the single observation H.
- $\hat{\theta}^L_{partition}$: we partition the observation window $[0, T]$ in I equally sized intervals:

$$\left[\frac{iT}{I}, \frac{(i+1)T}{I}\right],$$

  for $i \in \{0, \dots, I-1\}$. For each window, we obtain a penalised estimator by maximising Equation (6.10), and finally we obtain $\hat{\theta}^L_{partition}$ by averaging all estimations. This is similar to subsampling by means of covering regions (Possolo 1991; Politis et al. 1999; Guan et al. 2007).
- $\hat{\theta}^L_{thinning}$: we perform S different thinning procedures for a fixed $p \in (0, 1)$ providing the p-thinned observations $H_1, \dots, H_S$ of H. For each H_s, we obtain a penalised estimation by maximising Equation (6.10) on the statistical model $\mathcal{Q}_p$ and we obtain $\hat{\theta}^L_{thinning}$ by averaging all estimations.

$\hat{\theta}^L_{partition}$ corresponds to a common practice in point processes theory used to split an observation into multiple samples, which is particularly useful for large values of T.

Our proposed estimator $\hat{\theta}^L_{thinning}$ can be seen as an estimator obtained through a subsampling procedure, similar to the Bernoulli subsampling scheme (see for example Särndal et al. (2003, Chapter 3.2)).

Choice of hyperparameters. We focus on studying the efficiency of all four methods and so we consider the following grids for the hyper-parameters I, p and L:

$$I \in \{2, 3, 4, 5\}, \quad p \in \{0.1, \dots, 0.9\}, \quad L \in \{10^{-6}, 10^{-5}, \dots, 10^{2}\}.$$

Concerning the subsampling parameter of $\hat{\theta}^L_{thinning}$, we fix it at $S = 3$ but we remark that all presented results are consistent for other values of S. Proposing a model selection procedure to choose the optimal hyper-parameters I, p, L is out of the scope of this work, so we solely present the best estimators obtained through each method in the sense of the relative ℓ_2 error with respect to the true parameters of the model.

Results. Figure 6.2 represents the ℓ_2 relative error for each estimator over 1000 different simulations. We ordered all curves so that the error of our proposed estimator $\hat{\theta}^L_{thinning}$ is shown in ascending order. We can initially observe that the non-penalised estimations $\hat{\theta}$ often attain high ℓ_2 relative errors with high variability, which are greatly improved by the penalised methods. Overall, our estimations $\hat{\theta}^L_{thinning}$ present lower errors than the other estimations (dashed lines).

This can be confirmed in Table 6.1 displaying the proportion of times that each estimator outperforms the others. The non-penalised method provides by far the weakest estimator with a notably high MSRE when estimating β, which is consistently known as the hardest parameter

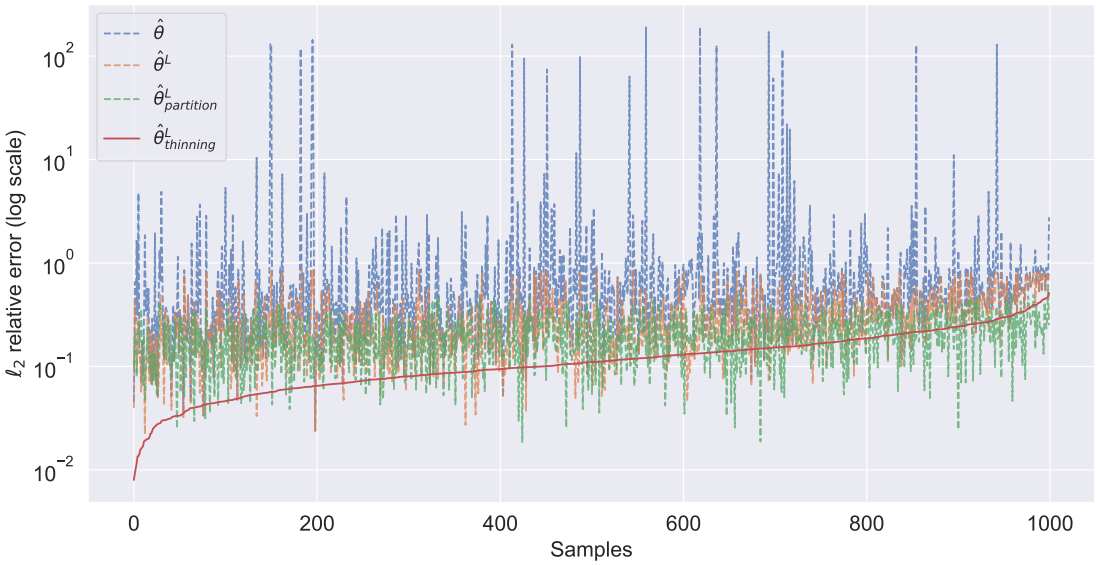

Figure 6.2: ℓ_2 relative error of all four estimators without penalisation (blue), with penalisation (orange), with penalisation and subsampling by partition (green) and our proposed method with penalisation and subsampling by thinning (red). 1000 independent samples are shown and samples are sorted so to display the error of $\hat{\theta}^L_{thinning}$ in ascending order.

to estimate in other parametric inference settings (see in Lemonnier et al. (2014) the discussion on the non-concavity of the log-likelihood). The three penalised methods greatly improve the estimations, in particular for β, significantly reducing MSRE all around. We notice that the lowest MSRE are obtained by our thinning method, with it being in most scenarios the best method all around. This is a testimony of the advantages of the subsampling and penalisation procedure that we propose in this work.

Estimator	MSRE $\hat{\mu}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\theta}$	% best
$\hat{\theta}$	0.18	0.13	5.14×10^2	2.85×10^2	1%
$\hat{\theta}^L$	0.12	0.09	0.13	0.12	6.7%
$\hat{\theta}^L_{partition}$	0.08	0.07	0.04	0.05	22.3%
$\hat{\theta}^L_{thinning}$	0.03	0.04	0.02	0.02	70%

Table 6.1: Mean square relative error (MSRE) of (μ, α, β) along with the MSRE of θ. Last column shows the proportion of times that each estimator achieves the lower relative ℓ_2 error accross all 1000 simulations.

6.5 Discussion

In this chapter, we leverage the spectral theory of point processes to present a parametric estimation procedure for thinned versions of univariate Hawkes processes. A first motivation for studying this model concerns the study of imperfect data and we showcase the efficiency of the spectral analysis to tackle such problems. This approach does not come without some issues, as we show that the underlying spectral density model is not identifiable without additional information on the parameter to be estimated. However, we highlight the efficiency of our estimation method once that identifiability is recovered, which is a step towards improving estimation in situations with potential missing event times. Our second contribution is to make use of thinning as a tool to perform subsampling in scenarios with short observations of processes. By introducing a penalised version of the spectral log-likelihood and combining this with subsampling, we demonstrate that we are able to significantly improve estimations. This is encouraging for real-world applications, such as in health or biology, where acquiring sizeable quantities of data can be difficult.

The main extension of our work would be to adapt our methods to higher dimensional data in order to consider multiple inter-connected phenomena, for instance in epidemiology, social interaction or neurobiology. In particular, our idea of introducing a subsampling approach for penalised optimisation methods is highly motivated by the high interest in Lasso approaches for multivate point processes (Reynaud-Bouret et al. 2013; Hansen et al. 2015; Bacry et al. 2020). This kind of regularisation is used in order to determine a sparse matrix of interaction for high-dimensional networks where only few connections between individuals exist. As these approaches tend to require a high amount of information or samples in order to perform well, adapting our subsampling scheme would make their use more accessible. A last avenue of exploration would be to propose a model selection procedure adapted to spectral approaches in order to choose the hyper-parameters of the penalised and subsampling methods in practical contexts where information on the real parameters is unavailable.

6.A Proof of Proposition 6.3.1

Let H_p be a p-thinning of a stationary point process H admitting a Bartlett spectrum Γ^H and a spectral density function f^H. The stationarity of H_p is given by the fact that the variables $(Z_k)_{k\in\mathbb{Z}}$ are i.i.d. and so for any integer r, for any $B_1 \dots, B_r \in \mathcal{B}^c$, the random vector $(N_p(B_k)_{k=1:r})$ has the same distribution as $(N_p(B_k+t)_{k=1:r})$, for any $t \in \mathbb{R}$. We will then denote Γ^{H_p} and f^{H_p} respectively the Bartlett spectrum and spectral density function of H_p

In order to establish Equation (6.6), we will leverage Theorem 6.3.1. We consider then the marked version $\bar{H}$ of H where the marks Z_k are i.i.d. Bernoulli distributions with common probability p.

Let $\varphi, \psi \in \mathcal{S}$, we define, for all $x, z \in \mathbb{R} \times \{0, 1\}$, the functions:

$$\varphi^\star(x, z) = \varphi(x)z\,, \qquad \psi^\star(x, z) = \psi(x)z\,.$$

Let us verify that these functions satisfy the conditions of Theorem 6.3.1. Without loss of generality, we will work uniquely with $\varphi^\star$, as the arguments are exactly the same for $\psi^\star$.

For any $x \in \mathbb{R}$, $\varphi^\star(x, Z) = \varphi(x)Z$ for Z a Bernoulli distribution of parameter p. It follows that $\varphi^\star(x, Z)$ admits a first and second order moment, and as $\varphi \in \mathcal{S}$, $\varphi(x)Z$ is integrable and

twice integrable, which shows that:

$$\int_{\mathbb{R}} \mathbb{E}\left[|\varphi^\star(x,Z)|\right] \mathrm{d}x < +\infty\,, \qquad \int_{\mathbb{R}} \mathbb{E}\left[\varphi^\star(x,Z)^2\right] \mathrm{d}x < +\infty\,.$$

Furthermore, for any $x \in \mathbb{R}$, $\bar{\varphi}(x) = \varphi(x)\mathbb{E}[Z] = p\varphi(x)$, and so, as the Schwartz space is closed under scalar mutliplication, it follows that $\bar{\varphi} \in \mathcal{S}$.

We can then apply Equation (6.5) to our marked process $\bar{H}$. For this, let us notice that:

$$\sum_{k\in\mathbb{Z}} \varphi^\star(T_k, Z_k) = \sum_{k\in\mathbb{Z}} \varphi(T_k) Z_k = \int_{\mathbb{R}} \varphi(t) H_p(\mathrm{d}t)\,,$$

with the same expression holding for $\psi^\star$ and ψ. So, the left-hand side of Equation (6.5) reads:

$$\begin{aligned}\operatorname{cov}\left(\sum_{k\in\mathbb{Z}} \varphi^\star(T_k,Z_k), \sum_{k\in\mathbb{Z}} \psi^\star(T_k,Z_k)\right) &= \operatorname{cov}\left(\int_{\mathbb{R}} \varphi(t)H_p(\mathrm{d}t), \int_{\mathbb{R}} \psi(t) H_p(\mathrm{d}t)\right) \\ &= \int_{\mathbb{R}} \tilde{\varphi}(\omega)\tilde{\psi}(-\omega)\,\Gamma^{H_p}(\mathrm{d}\omega)\,, \end{aligned} \tag{6.11}$$

where the last equality comes from Equation (6.3).

For the left-hand side, let us remark that $\bar{\varphi}(x) = p\varphi(x)$ for any $x \in \mathbb{R}$ and so, for any $\omega \in \mathbb{R}$,

$$\tilde{\bar{\varphi}}(\omega) = p\tilde{\varphi}(\omega)\,,$$

and,

$$\tilde{\phi}^\star(\omega, Z) = \tilde{\phi}(\omega) Z\,.$$

The right-hand side of Equation (6.5) becomes:

$$\begin{aligned}\int_{\mathbb{R}} \tilde{\bar{\varphi}}(\omega)\tilde{\bar{\psi}}(-\omega)\,\Gamma^H(\mathrm{d}\omega) &+ \int_{\mathbb{R}} \operatorname{cov}\left(\tilde{\varphi}^\star(\omega,Z), \tilde{\psi}^\star(-\omega,Z)\,M_1(\mathrm{d}\omega)\right) \\ &= \int_{\mathbb{R}} p^2\tilde{\varphi}(\omega)\tilde{\psi}(-\omega)\,\Gamma^H(\mathrm{d}\omega) + \int_{\mathbb{R}} \tilde{\varphi}(\omega)\tilde{\psi}(\omega)\operatorname{cov}(Z,Z)\,M_1(\mathrm{d}\omega) \\ &= \int_{\mathbb{R}} \tilde{\varphi}(\omega)\tilde{\psi}(-\omega)\left(p^2\Gamma^H(\mathrm{d}\omega) + p(1-p)M_1(\mathrm{d}\omega)\right). \end{aligned} \tag{6.12}$$

By combining both sides (Equations (6.11) and (6.12)) it follows that, for any $\varphi, \psi \in \mathcal{S}$:

$$\int_{\mathbb{R}} \tilde{\varphi}(\omega)\tilde{\psi}(-\omega)\,\Gamma^{H_p}(\mathrm{d}\omega) = \int_{\mathbb{R}} \tilde{\varphi}(\omega)\tilde{\psi}(-\omega)\left(p^2\Gamma^H(\mathrm{d}\omega) + p(1-p)M_1(\mathrm{d}\omega)\right).$$

As this equality holds for any functions in the Schwartz space, by duality of the Fourier transform (Pinsky 2008), then,

$$\Gamma_p = p^2\Gamma + p(1-p)M_1\,,$$

and as $M_1 = m_1\ell_{\mathbb{R}}$,

$$f_p(\omega) = p^2 f(\omega) + p(1-p)m_1\,,$$

which achieves the proof.

6.B Proof of Proposition 6.3.2

Let $\theta = (\mu, \alpha, \beta, p)$ and $\theta' = (\mu', \alpha', \beta', p')$ be two admissible parameters for model $\mathcal{Q}$.

In order to prove that model $\mathcal{Q}$ is identifiable if and only if one of four parameters is fixed, we will begin by retrieving the system of Equations (6.9) and verify that it defines an admissible parameter for $\mathcal{Q}$. Subsequently we will verify that by fixing one parameter, we retrieve the identifiability of the model.

We begin by assuming that $f_\theta = f_{\theta'}$. By Equation (6.8), this equality reads:

$$\forall \omega \in \mathbb{R}, \quad \frac{\mu p}{1-\alpha}\left(1 + p\frac{\beta^2\alpha(2-\alpha)}{\beta^2(1-\alpha)^2 + 4\pi^2\omega^2}\right) = \frac{\mu' p'}{1-\alpha'}\left(1 + p'\frac{\beta'^2\alpha'(2-\alpha')}{\beta'^2(1-\alpha')^2 + 4\pi^2\omega^2}\right).$$

Let us remark that a sufficient condition for both sides to be equal, for all $\omega \in \mathbb{R}$, is that the following system of equations is verified:

$$\begin{cases} \frac{\mu p}{1-\alpha} = \frac{\mu' p'}{1-\alpha'} \\ p\beta^2\alpha(2-\alpha) = p'\beta'^2\alpha'(2-\alpha') \\ \beta(1-\alpha) = \beta'(1-\alpha'), \end{cases} \tag{6.13}$$

as this would mean that the three constants (w.r.t. ω) in both sides are equal. In fact this is also a necessary condition as the three equalities can be retrieved for $f(0)$, $\lim_{\omega\to+\infty} f(\omega)$ and $f(1)$ and combining the equations.

Let $\kappa \in (0, 1/p)$ and $p' = \kappa p$, which reduces the system of Equations (6.13) to:

$$\begin{cases} \dfrac{\mu}{1-\alpha} = \dfrac{\mu'\kappa}{1-\alpha'} & \text{(6.14a)} \\ \beta^2\alpha(2-\alpha) = \kappa\beta'^2\alpha'(2-\alpha') & \text{(6.14b)} \\ \beta(1-\alpha) = \beta'(1-\alpha'), & \text{(6.14c)} \end{cases}$$

As $\alpha < 1$, $\alpha' < 1$, $\beta > 0$ and $\beta' > 0$, Equation (6.14c) can be expressed as:

$$\beta'^2 = \frac{\beta^2(1-\alpha)^2}{(1-\alpha')^2},$$

and by replacing β'^2 in Equation (6.14b), we obtain:

$$\begin{aligned} & \frac{\alpha(2-\alpha)}{(1-\alpha)^2} = \kappa\frac{\alpha'(2-\alpha')}{(1-\alpha')^2} \\ \iff \quad & \frac{1-(1-\alpha)^2}{(1-\alpha)^2} = \kappa\frac{1-(1-\alpha')^2}{(1-\alpha')^2} \\ \iff \quad & 1 + \frac{1}{\kappa}\left(\frac{1}{(1-\alpha)^2} - 1\right) = \frac{1}{(1-\alpha')^2}, \end{aligned} \tag{6.15}$$

and as $\alpha \in (0, 1)$, the left-side term is positive and so:

$$\alpha' = 1 - \frac{1}{\sqrt{1 + \frac{1}{\kappa}\left(\frac{1}{(1-\alpha)^2} - 1\right)}}.$$

We can then obtain the explicit expressions of μ' and β' with Equations (6.14a) and (6.14c) providing the following system:

$$\begin{cases} \mu' = \frac{\mu(1-\alpha')}{\kappa(1-\alpha)} \\ \alpha' = 1 - \frac{1}{\sqrt{1+\frac{1}{\kappa}\left(\frac{1}{(1-\alpha)^2}-1\right)}} \\ \beta' = \beta\frac{1-\alpha}{1-\alpha'} \\ p' = \kappa p \end{cases} \iff \begin{cases} \mu' = \frac{\mu}{\kappa(1-\alpha)\sqrt{1+\frac{1}{\kappa}\left(\frac{1}{(1-\alpha)^2}-1\right)}} \\ \alpha' = 1 - \frac{1}{\sqrt{1+\frac{1}{\kappa}\left(\frac{1}{(1-\alpha)^2}-1\right)}} \\ \beta' = \beta(1-\alpha)\sqrt{1+\frac{1}{\kappa}\left(\frac{1}{(1-\alpha)^2}-1\right)} \\ p' = \kappa p\,. \end{cases} \tag{6.16}$$

To verify that, for any $\kappa \in (0, 1/p)$, parameter $\theta' = (\mu', \alpha', \beta', p')$ is an admissible parameter, we have to make sure that $\theta' \in \Theta = \mathbb{R}_{>0} \times (0,1) \times \mathbb{R}_{>0} \times (0,1)$. Let $\theta = (\mu, \alpha, \beta, p) \in \Theta$, $\kappa \in (0, 1/p)$ and θ' defined by Equations (6.16). For such a κ, $p' \in (0,1)$ is immediately verified.

The rest of the conditions are verified if and only if:

$$\sqrt{1+\frac{1}{\kappa}\left(\frac{1}{(1-\alpha)^2}-1\right)} > 1 \iff \frac{1}{(1-\alpha)^2} - 1 > 0,.$$

This is immediate as $\alpha \in (0,1)$ and so $\theta' \in \Theta$. From Equations (6.16) we can see that $\theta' = \theta$ and we have proven that $f_\theta = f_{\theta'}$, so model $\mathcal{Q}$ is not identifiable.

Lastly, let us show that if any of the four parameters is known then the model defined by the remaining triplet is identifiable. By considering the previously established equations (see Equations (6.15) and (6.16)), we know that for $\theta = (\mu, \alpha, \beta, p) \in \Theta$ and $\theta' = (\mu', \alpha', \beta', p') \in \Theta$:

$$f_\theta = f_{\theta'} \iff \begin{cases} \mu' = \frac{\mu(1-\alpha')}{\kappa(1-\alpha)} \\ \frac{\alpha(2-\alpha)}{(1-\alpha)^2} = \kappa\frac{\alpha'(2-\alpha')}{(1-\alpha')^2} \\ \beta' = \beta\frac{1-\alpha}{1-\alpha'} \\ p' = \kappa p \end{cases}. \tag{6.17}$$

- If $\mu = \mu'$, the system of Equations (6.17) reduces to:

$$\begin{cases} \kappa = \frac{(1-\alpha')}{(1-\alpha)} \\ \kappa\frac{\alpha'(2-\alpha')}{(1-\alpha')^2} = \frac{\alpha(2-\alpha)}{(1-\alpha)^2} \\ \beta' = \frac{\beta}{\kappa} \\ p' = \kappa p \end{cases} \iff \begin{cases} \kappa = \frac{(1-\alpha')}{(1-\alpha)} \\ \frac{\alpha'(2-\alpha')}{(1-\alpha')} = \frac{\alpha(2-\alpha)}{(1-\alpha)} \\ \beta' = \frac{\beta}{\kappa} \\ p' = \kappa p \end{cases}.$$

  As $\alpha \in (0,1)$ and $\alpha' \in (0,1)$, the second equation implies that $\alpha = \alpha'$ and so $\kappa = 1$ and $\beta' = \beta$.

- If $\alpha = \alpha'$, the second equation in (6.17) implies directly that $\kappa = 1$ and so all other equalities follow.

- If $\beta = \beta'$, the third equation in (6.17) implies that $\alpha = \alpha'$ and by the previous point all other equalities hold.

- If $p = p'$, $\kappa = 1$ and the rest of equalities are verified.

  This shows that whenever one of the parameters is fixed, $f_\theta = f_{\theta'}$ implies $\theta = \theta'$. This achieves the proof.

7 Discussion

The Hawkes process model is used accross a vast multitude of application fields, and so establishing inference procedures to model more complex phenomena is undeniably important. Our work provided novel statistical tools for the estimation of Hawkes processes within the frequentist parametric framework. Our effort was centered around establishing methodological procedures for two different submodels, with code implementations of our estimators and numerical illustrations of their efficiency.

The first part of our work focused on the study of inhibiting interactions for Hawkes processes by establishing the maximum likelihood estimation procedure. In Chapter 3, we implemented this method under the assumption of monotony of the kernel function, which allowed us to derive a closed-form expression of the log-likelihood. In Chapter 4, we extended our previous work to the multivariate setting for exponentially-shaped interactions which revealed the challenges of establishing the identifiability of our model. This issue arised primarily from our use of the positive part function in the non-linear Hawkes process model as, by taking null values, it hinders the accurate identification of interaction parameters. On the other hand, the choice of this function helps modelling a closer behaviour to the classical linear Hawkes model and simplifies the computation of the intensity function. Prioritising positive non-linear functions is a way to circumvent this problem, as shown in the Bayesian context in Sulem et al. (2024), for a shifted version of the positive part function. By adapting our procedure to an approximation of the positive part function, for example through the softplus function, the resulting method could be a satisfying compromise between model fidelity and theoretical guarantees.

Numerically, we illustrated the improvements in our estimations by incorporating a support inference step for the interaction matrix, as evidenced by our goodness-of-fit hypothesis testing. Our methods relied on data-based paradigms which are highly dependant on sample size. In practical applications, such as medical data analysis with limited observations, it is important to propose more robust methods. Future research could explore Lasso penalisation procedures, in the same vein as the work of Bacry et al. (2020), where both an ℓ_1 and trace-norm penalisations are tuned by the means of data-driven weights and that can benefit from cross-validation procedures. We believe this is a promising research field with novel contributions as shown in the work of Lotz (2024) on sparsity tests for Hawkes processes.

In our application to the study of neuronal activity, the results exhibited a clear self-interaction behaviour for each studied neuron. This is consistent with the known behaviour of neurons as

they present a post-activation refractory period, reflecting an incapacity to reactivate after a synaptic impulse is produced. The fact that our estimations are not consistently self-inhibitory is most likely due to the choice of a monotone kernel in the form of the exponential function. In reality, when membrane potential is measured, the refractory effect is delayed which is better represented by a non-monotone function suggesting that extending our work for more complex interaction functions is a clear path to obtaining more faithful models.

The second part of our work examined noisy observations of linear Hawkes processes. In Chapter 5, we modelled the presence of exogeneous points by superposing a homogeneous Poisson process to the Hawkes process. Chapter 6 adresses the absence of points through a thinning representation of the original point process. Our spectral approach, utilising the Bartlett spectrum and the periodogram, enabled us to establish an estimation procedure by optimising a spectral version of the log-likelihood. The main difficulty encountered was about identifiability of the statistical models, which derived directly from the definition of the spectral density function. To partially solve this problem, we provided conditions on the kernel functions that allowed us to recover identifiability of our model. Our method will greatly benefit from more general results, particularly for higher dimensional processes with more complex interactions. Another possible solution could be found in the recent works by Roueff et al. (2016) and Roueff et al. (2019), that establish an equivalent expression of the Bartlett spectrum by relaxing the stationarity condition to a local stationarity. This could help to incorporate more general shapes of noise which in turn could help to better identify the noise dynamics.

Our numerical illustrations on synthetic data showed the efficiency of our approach to perform this difficult estimation task and the next step is to apply our results to real-world data. This can help to provide a better insight in contexts where measurement errors are more likely to appear. In particular, it can be interesting to combine both models studied in Chapters 5 and 6in order to study the spread of illnesses through screening. In this context, the presence of false positives and false negatives can be modelled by the superposition and thinning operations and so our model can clearly help to obtain better interpretations of disease transmission dynamics.

Finally, in the context of subsampling through thinning, a significant issue concerns the lack of evaluation methods for spectral estimators in unsupervised settings. As suggested in the recent works by Cronie et al. (2024) and Coeurjolly et al. (2024), subsampling paradigms can greatly improve the estimation procedures and allow for more complex methods as cross-validation paradigms to be available for the study of point processes. Establishing asymptotic results for our estimators could partially address this issue by providing a way of establishing asymptotic confidence intervals. Encouragingly, there has been recent advances concerning these guarantees for the periodogram in the works by Rajala et al. (2023) and Yang et al. (2024), which pave the way for the study of the periodogram.

Bibliography

Achab, M., E. Bacry, S. Gaïffas, I. Mastromatteo, and J-F. Muzy (2016). "Uncovering Causality from Multivariate Hawkes Integrated Cumulants". In: *Journal of Machine Learning Research* 18, 192:1–192:28.

Adamopoulos, L. (1976). "Cluster models for earthquakes: Regional comparisons". In: *Journal of the International Association for Mathematical Geology* 8.4, pp. 463–475.

Antoniadis, A. and J. Bigot (2006). "Poisson inverse problems". In: *Annals of Statistics* 34.5, pp. 2132–2158.

Baccelli, F., B. Błaszczyszyn, and M. Karray (Jan. 2020). *Random Measures, Point Processes, and Stochastic Geometry*.

Bacry, E, S. Delattre, M. Hoffmann, and J.F. Muzy (2013). "Scaling limits for Hawkes processes and application to financial statistics". In: *Stochastic Processes and Applications* 123, pp. 2475–2499.

Bacry, E., M. Bompaire, P. Deegan, S. Gaïffas, and S.V. Poulsen (2018). "tick: a Python Library for Statistical Learning, with an emphasis on Hawkes Processes and Time-Dependent Models". In: *Journal of Machine Learning Research* 18.1, pp. 7937–7941.

Bacry, E., M. Bompaire, S. Gaïffas, and J.F. Muzy (2020). "Sparse and low-rank multivariate Hawkes processes". In: *Journal of Machine Learning Research* 21.1, pp. 1–32.

Bacry, E., I. Mastromatteo, and J-F. Muzy (2015). "Hawkes Processes in Finance". In: *Market Microstructure and Liquidity* 01.01, p. 1550005.

Bacry, E. and J.F. Muzy (2016). "First- and Second-Order Statistics Characterization of Hawkes Processes and Non-Parametric Estimation". In: *IEEE Transactions on Information Theory* 62.4, pp. 2184–2202.

Baddeley, A. (Jan. 2006). "Spatial Point Processes and their Applications". In: *Stochastic Geometry: Lectures given at the C.I.M.E. 2004, Lecture Notes in Mathematics 1892*.

Baouan, A., S Coustou, M. Lacome, S. Pulido, and M. Rosenbaum (2023). "Crediting football players for creating dangerous actions in an unbiased way: the generation of threat (GoT) indices". In: Preprint at `https://arxiv.org/abs/2304.05242`.

Bartlett, M. S. (1963). "The Spectral Analysis of Point Processes". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 25.2, pp. 264–296.

— (1964). "The Spectral Analysis of Two-Dimensional Point Processes". In: *Biometrika* 51.3/4, pp. 299–311.

Blundell, C., J. Beck, and K.A. Heller (2012). "Modelling Reciprocating Relationships with Hawkes Processes". In: *Advances in Neural Information Processing Systems*. Ed. by F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger. Vol. 25. Curran Associates, Inc.

Bompaire, M., E. Bacry, and S. Gaïffas (2018). "Dual optimization for convex constrained objectives without the gradient-Lipschitz assumption". In: Preprint at https://arxiv.org/abs/1807.03545.

Bonnet, A, F. Cheysson, M. Martinez Herrera, and M. Sangnier (2024). "Spectral analysis for noisy Hawkes processes inference". In: Preprint at https://arxiv.org/abs/2405.12581.

Bonnet, A, M. Martinez Herrera, and M. Sangnier (2023). "Inference of multivariate exponential Hawkes processes with inhibition and application to neuronal activity". In: *Statistics and Computing* 33.91.

Bonnet, A., C. Dion-Blanc, F. Gindraud, and S. Lemler (2022a). "Neuronal network inference and membrane potential model using multivariate Hawkes processes". In: *Journal of Neuroscience Methods* 372, p. 109550.

Bonnet, A., C. Lacour, F. Picard, and V. Rivoirard (2022b). "Uniform deconvolution for Poisson point processes". In: *Journal of Machine Learning Research* 23.1.

Bonnet, A., M. Martinez Herrera, and M. Sangnier (2021). "Maximum likelihood estimation for Hawkes processes with self-excitation or inhibition". In: *Statistics & Probability Letters* 179, p. 109214.

Brémaud, P and L. Massoulié (2002). "Power Spectra of General Shot Noises and Hawkes Point Processes with a Random Excitation". In: *Advances in Applied Probability* 34.1, pp. 205–222.

— (1996). "Stability of nonlinear Hawkes processes". In: *The Annals of Probability* 24.3, pp. 1563–1588.

Brémaud, P., L. Massoulié, and A. Ridolfi (Dec. 2005). "Power spectra of random spike fields and related processes". In: *Advances in Applied Probability* 37.

Brillinger, D. R. (2012). *Statistical Inference for Stationary Point Processes*. New York, NY: Springer New York, pp. 499–543.

Byrd, R.H., P. Lu, J. Nocedal, and C. Zhu (1995). "A Limited Memory Algorithm for Bound Constrained Optimization". In: *SIAM Journal on Scientific and Statistical Computing* 16.5, pp. 1190–1208.

Carstensen, L., A. Sandelin, O. Winther, and N.R. Hansen (2010). "Multivariate Hawkes process models of the occurrence of regulatory elements". In: *BMC Bioinformatics* 11.1, p. 456.

Chen, J.M., A.G. Hawkes, E. Scalas, and M. Trinh (2018). "Performance of information criteria for selection of Hawkes process models of financial data". In: *Quantitative Finance* 18.2, pp. 225–235.

Chen, S., A. Shojaie, E. Shea-Brown, and D. Witten (2017). "The Multivariate Hawkes Process in High Dimensions: Beyond Mutual Excitation". In: Preprint at https://arxiv.org/abs/1707.04928.

Chevallier, J., A. Duarte, E. Löcherbach, and G. Ost (2019). "Mean field limits for nonlinear spatially extended Hawkes processes with exponential memory kernels". In: *Stochastic Processes and their Applications* 129.1, pp. 1–27.

Cheysson, F. and G. Lang (2022). "Spectral estimation of Hawkes processes from count data". In: *The Annals of Statistics* 50.3, pp. 1722–1746.

Chiang, W-H., X. Liu, and G. Mohler (2022). "Hawkes process modeling of COVID-19 with mobility leading indicators and spatial covariates." In: *International Journal of Forecasting* 38.2, pp. 505–520.

Chornoboy, E. S., L. P. Schramm, and A. F. Karr (1988). "Maximum likelihood identification of neural point process systems". In: *Biological Cybernetics* 59.4, pp. 265–275.

Coeurjolly, J-F., A-L. Fougères, T. Espinasse, and M. Ribatet (2024). "Spatio-temporal point process intensity estimation using zero-deflated subsampling applied to a lightning strikes dataset in France". In: Preprint at https://arxiv.org/abs/2403.11564.

Costa, M., C. Graham, L. Marsalle, and V. Tran (2020). "Renewal in Hawkes processes with self-excitation and inhibition". In: *Advances in Applied Probability* 52.3, pp. 879–915.

Cox, D.R. and V. Isham (1980). *Point Processes*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

Cronie, O., M. Moradi, and C.A.N. Biscio (June 2024). "A cross-validation-based statistical theory for point processes". In: *Biometrika* 111.2, pp. 625–641.

Da Fonseca, J. and R. Zaatour (2013). "Hawkes Process: Fast Calibration, Application to Trade Clustering, and Diffusive Limit". In: *Journal of Futures Markets* 34.6, pp. 548–579.

Daley, D. J. (1971). "Weakly Stationary Point Processes and Random Measures". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 33.3, pp. 406–428.

Daley, D.J. and D. Vere-Jones (2003). *An introduction to the theory of point processes. Vol. I*. Second. New York: Springer-Verlag.

— (2008). *An introduction to the theory of point processes. Vol. II*. Second. New York: Springer-Verlag.

Dassios, A. and H. Zhao (2011). "A dynamic contagion process". In: *Advances in Applied Probability* 43.3, pp. 814–846.

Davis, M. H. A. (1984). "Piecewise-Deterministic Markov Processes : A General Class of Non-Diffusion Stochastic Models". In: *Journal of the Royal Statistical Society Series B (Methodological)* 46.3, pp. 353–388.

Denis, C., C. Dion-Blanc, R.E. Lacoste, L. Sansonnet, and Y. Bas (June 2024). "Bats monitoring: a classification procedure of bats behaviours based on Hawkes processes". In: *Journal of the Royal Statistical Society Series C: Applied Statistics*, qlae024.

Deutsch, I. and G.J. Ross (2020). "ABC Learning of Hawkes Processes with Missing or Noisy Event Times". In: Preprint at `https://arxiv.org/abs/2006.09015`.

— (2023). "Estimating Product Cannibalisation in Wholesale using Multivariate Hawkes Processes with Inhibition". In: Preprint at `https://arxiv.org/abs/2201.05009`.

Donnet, S., V. Rivoirard, and J. Rousseau (2020). "Nonparametric Bayesian estimation for multivariate Hawkes processes". In: *Annals of Statistics* 48.5, pp. 2698–2727.

Duarte, A., E. Löcherbach, and G. Ost (2019). "Stability, convergence to equilibrium and simulation of non-linear Hawkes processes with memory kernels given by the sum of Erlang kernels". In: *ESAIM: Probability and Statistics* 23, pp. 770–796. eprint: `1610.03300`.

Düker, M-C. and V. Pipiras (2019). "Asymptotic results for multivariate local Whittle estimation with applications". In: *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pp. 584–588.

Duval, C., E. Luçon, and C. Pouzat (2022). "Interacting Hawkes processes with multiplicative inhibition". In: *Stochastic Processes and their Applications* 148, pp. 180–226.

Eichler, M., R. Dahlhaus, and J. Dueck (2016). "Graphical Modeling for Multivariate Hawkes Processes with Nonparametric Link Functions". In: *Journal of Time Series Analysis* 38, pp. 225–242.

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

Gerhard, F., M. Deger, and W. Truccolo (2017). "On the stability and dynamics of stochastic spiking neuron models: Nonlinear Hawkes process and point process GLMs." In: *PLoS computational biology* 13.2.

Guan, Y. and D. R. Afshartous (2007). "Test for independence between marks and points of marked point processes: a subsampling approach". In: *Environmental and Ecological Statistics* 14.2, pp. 101–111.

Guo, X., A. Hu, R. Xu, and J. Zhang (2018). "Consistency and Computation of Regularized MLEs for Multivariate Hawkes Processes". In: Preprint at `https://arxiv.org/abs/1810.02955`.

Gupta, A., M. Farajtabar, B. Dilkina, and H. Zha (July 2018). "Discrete Interventions in Hawkes Processes with Applications in Invasive Species Management". In: *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*. International Joint Conferences on Artificial Intelligence Organization, pp. 3385–3392.

Hansen, N.R., P. Reynaud-Bouret, and V. Rivoirard (2015). "Lasso and probabilistic inequalities for multivariate point processes". In: *Bernoulli* 21.1, pp. 83–143.

Harris, T.E. (2002). *The Theory of Branching Processes*. Dover phoenix editions. Dover Publications.

Hawkes, A.G. (1971). "Spectra of Some Self-Exciting and Mutually Exciting Point Processes". In: *Biometrika* 58.1, pp. 83–90.

— (2018). "Hawkes processes and their applications to finance: a review". In: *Quantitative Finance* 18.2, pp. 193–198.

Hawkes, A.G. and D. Oakes (1974). "A cluster process representation of a self-exciting process". In: *Journal of Applied Probability* 11.3, pp. 493–503.

Hohage, Thorsten and Frank Werner (2016). "Inverse problems with Poisson data: statistical regularization theory, applications and algorithms". In: *Inverse Problems* 32.9, p. 093001.

Joliffe, I.T. (2002). *Principal Component Analysis*. Second. New York: Springer-Verlag.

Karavasilis, G. J. and A. G. Rigas (2007). "Spectral analysis techniques of stationary point processes used for the estimation of cross-correlation: Application to the study of a neurophysiological system". In: *2007 15th European Signal Processing Conference*, pp. 2479–2483.

Kirchner, M. (2017). "An estimation procedure for the Hawkes process". In: *Quantitative Finance* 17.4, pp. 571–595.

Kwon, J., Y. Zheng, and M. Jun (2023). "Flexible spatio-temporal Hawkes process models for earthquake occurrences". In: *Spatial Statistics* 54, p. 100728.

Lambert, R.C., C. Tuleau-Malot, T. Bessaih, V. Rivoirard, Y. Bouret, N. Leresche, and P. Reynaud-Bouret (2018). "Reconstructing the functional connectivity of multiple spike trains using Hawkes models". In: *Journal of Neuroscience Methods* 297, pp. 9–21.

Last, G. and M. Penrose (2017). *Lectures on the Poisson Process*. Institute of Mathematical Statistics Textbooks. Cambridge University Press.

Laub, P. (2014). "Hawkes Processes: Simulation, Estimation, and Validation". In: *Bachelor's Thesis, University of Queensland*.

Lemonnier, R. and N. Vayatis (2014). "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 T. Calders, F. Esposito, E. Hüllermeier, and R. Meo. Vol. 8725. Berlin: Springer Berlin Heidelberg, pp. 161–176.

Lesage, L., M. Deaconu, A. Lejay, J.A. Meira, G. Nichil, and R. State (2022). "Hawkes Processes Framework With a Gamma Density As Excitation Function: Application to Natural Disasters for Insurance". In: *Methodology and Computing in Applied Probability* 24.4, pp. 2509–2537.

Lewis, E. and G. Mohler (2011). "A Nonparametric EM Algorithm for Multiscale Hawkes Processes". In: *Journal of Nonparametric Statistics* 1, pp. 1–20.

Lewis, Peter A. W. (1964). "A Branching Poisson Process Model for the Analysis of Computer Failure Patterns". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 26.3, pp. 398–441.

Li, T. and Y. Ke (2019). "Thinning for Accelerating the Learning of Point Processes". In: *Advances in Neural Information Processing Systems*. Ed. by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett. Vol. 32. Curran Associates, Inc.

Linderman, S. and R Adams (2014). "Discovering Latent Network Structure in Point Process Data". In: *Proceedings of the 31st International Conference on Machine Learning*. Ed. by

E. P. Xing and T. Jebara. Vol. 32. Beijing, China: Proceedings of Machine Learning Research, pp. 1413–1421.

Lotz, A. (2024). "A sparsity test for multivariate Hawkes processes". In: Preprint at `https://arxiv.org/abs/2405.08640`.

Lu, X and F. Abergel (2018). "High-dimensional Hawkes processes for limit order books: modelling, empirical analysis and numerical calibration". In: *Quantitative Finance* 18.2, pp. 249–264.

Lund, J. and M. Rudemo (2000). "Models for Point Processes Observed with Noise". In: *Biometrika* 87.2, pp. 235–249.

Mei, H. and J. Eisner (2017). "The Neural Hawkes Process: A Neurally Self-Modulating Multivariate Point Process". In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. California: Curran Associates Inc., pp. 6757–6767.

Mei, H., G. Qin, and J. Eisner (2019). "Imputing Missing Events in Continuous-Time Event Streams". In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri and Ruslan Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. PMLR, pp. 4475–4485.

Menon, A. and Y. Lee (Apr. 2018). "Proper Loss Functions for Nonlinear Hawkes Processes". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 32.

Mishra, S., M.A. Rizoiu, and L. Xie (2016). "Feature Driven and Point Process Approaches for Popularity Prediction". In: *Proceedings of the 25th ACM International Conference on Information and Knowledge Management, CIKM 2016, Indianapolis, IN, USA, October 24-28, 2016*. New York: ACM, pp. 1069–1078.

Mohler, G., M. Short, P. Brantingham, F. Schoenberg, and G. Tita (Mar. 2011). "Self-Exciting Point Process Modeling of Crime". In: *Journal of the American Statistical Association* 106, pp. 100–108.

Møller, J. and J.G. Rasmussen (2005). "Perfect Simulation of Hawkes Processes". In: *Advances in Applied Probability* 37.3, pp. 629–646.

Møller, J. and R.P. Waagepetersen (2003). *Statistical Inference and Simulation for Spatial Point Processes*. First. New York, NY: Chapman and Hall/CRC.

Moradi, M. M., O. Cronie, E. Rubak, R. Lachieze-Rey, J. Mateu, and A. Baddeley (2019). "Resample-smoothing of Voronoi intensity estimators". In: *Statistics and Computing* 29.5, pp. 995–1010.

Mugglestone, M.A. and E. Renshaw (1996). "A practical guide to the spectral analysis of spatial point processes". In: *Computational Statistics & Data Analysis* 21.1, pp. 43–65.

— (2001). "Spectral tests of randomness for spatial point patterns". In: *Environmental and Ecological Statistics* 8.3, pp. 237–251.

Narayanan, S., I. Kosmidis, and P. Dellaportas (2023). "Flexible marked spatio-temporal point processes with applications to event sequences from association football". In: *Journal of the Royal Statistical Society Series C: Applied Statistics* 72.5, pp. 1095–1126.

Nicvert, L., S. Donnet, M. Keith, M. Peel, M.J. Somers, K.H. Swanepoel, J. Venter, H. Fritz, and S. Dray (2024). "Using the multivariate Hawkes process to study interactions between multiple species from camera trap data". In: *Ecology* 105.4, e4237.

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

— (1981). "On Lewis' simulation method for point processes". In: *IEEE Transactions on Information Theory* 27, pp. 23–30.

— (1988). "Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes". In: *Journal of the American Statistical Association* 83.1, pp. 9–27.

Ogata, Y. (1998). "Space-Time Point-Process Models for Earthquake Occurrences". In: *Annals of the Institute of Statistical Mathematics* 50.2, pp. 379–402.

Olinde, J. and M.B. Short (2020). "A Self-limiting Hawkes Process: Interpretation, Estimation, and Use in Crime Modeling". In: *2020 IEEE International Conference on Big Data (Big Data)*, pp. 3212–3219.

Ozaki, T. (1979). "Maximum likelihood estimation of Hawkes' self-exciting point processes". In: *Annals of the Institute of Statistical Mathematics* 31.1, pp. 145–155.

Petersen, P.C. and R.W. Berg (2016). "Lognormal firing rate distribution reveals prominent fluctuation-driven regime in spinal motor networks". In: *eLife* 5, e18805.

Pham, D.T. (1996). "Blind separation of instantaneous mixture of sources via an independent component analysis". In: *IEEE Transactions on Signal Processing* 44.11, pp. 2768–2779.

Pinsky, M.A. (2008). *Introduction to Fourier Analysis and Wavelets*. Graduate studies in mathematics. American Mathematical Society.

Pinto, J.C.L., T. Chahed, and E. Altman (2015). "Trend detection in social networks using Hawkes processes". In: *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 1441–1448.

Politis, D. N., J. P. Romano, and M. Wolf (1999). *Subsampling Marked Point Processes*. New York, NY: Springer New York, pp. 138–158.

Possolo, A. (1991). "Subsampling a Random Field". In: *Lecture Notes-Monograph Series* 20, pp. 286–294.

Radosevic, M., A. Willumsen, P.C. Petersen, H. Lindén, M. Vestergaard, and R.W. Berg (2019). "Decoupling of timescales reveals sparse convergent CPG network in the adult spinal cord". In: *Nature Communications* 10.2937, pp. 1–14.

Rajala, T.A., S.C. Olhede, J.P. Grainger, and D.J. Murrell (2023). "What is the Fourier Transform of a Spatial Point Process?" In: *IEEE Transactions on Information Theory*.

Rasmussen, J.G. (2013). "Bayesian Inference for Hawkes Processes". In: *Methodology and Computing in Applied Probability* 15.3, pp. 623–642.

Reynaud-Bouret, P., A. Muzy, and I. Bethus (2021). "Towards a mathematical definition of functional connectivity". In: *Comptes rendus. Mathématique* 359, pp. 481–492.

Reynaud-Bouret, P., V. Rivoirard, F. Grammont, and C. Tuleau-Malot (2014). "Goodness-of-Fit Tests and Nonparametric Adaptive Estimation for Spike Train Analysis". In: *The Journal of Mathematical Neuroscience* 4.3, p. 3.

Reynaud-Bouret, P., V. Rivoirard, and C. Tuleau-Malot (2013). "Inference of functional connectivity in Neurosciences via Hawkes processes". In: pp. 317–320.

Reynaud-Bouret, P. and S. Schbath (2009). "Adaptive estimation for Hawkes processes; application to genome analysis". In: *The Annals of Statistics* 38.

Rizoiu, M., S. Mishra, Q. Kong, M. Carman, and L. Xie (2018). "SIR-Hawkes: linking epidemic models and Hawkes processes to model diffusions in finite populations". In: *Proceedings of the 2018 World Wide Web Conference*. International World Wide Web Conferences Steering Committee, pp. 419–428.

Rizoiu, M.A., Y. Lee, S. Mishra, and L. Xie (2017). "A Tutorial on Hawkes Processes for Events in Social Media". In: Preprint at `https://arxiv.org/abs/1708.06401`.

Roueff, F. and R. von Sachs (2019). "Time-frequency analysis of locally stationary Hawkes processes". In: *Bernoulli* 25.2, pp. 1355–1385.

Roueff, F., R. von Sachs, and L. Sansonnet (2016). "Locally stationary Hawkes processes". In: *Stochastic Processes and their Applications* 126.6, pp. 1710–1743.

Särndal, C.E., B. Swensson, and J. Wretman (2003). *Model Assisted Survey Sampling*. Springer Series in Statistics. New York, NY: Springer New York.

Simon, C. and Y. Nakahiro (2017). "Statistical inference for ergodic point processes and application to Limit Order Book". In: *Stochastic Processes and their Applications* 127.6, pp. 1800–1839.

Sulem, D., V. Rivoirard, and J. Rousseau (2023). "Scalable and adaptive variational Bayes methods for Hawkes processes". In: Preprint at `https://arxiv.org/abs/2212.00293`.

— (2024). "Bayesian estimation of nonlinear Hawkes processes". In: *Bernoulli* 30.2, pp. 1257–1286.

Tibshirani, R. (1996). "Regression Shrinkage and Selection via the Lasso". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 58.1, pp. 267–288.

Trouleau, W., J. Etesami, M. Grossglauser, N. Kiyavash, and P. Thiran (2019). "Learning Hawkes Processes Under Synchronization Noise". In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri and Ruslan Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. PMLR, pp. 6325–6334.

Tuan, P.D. (1981). "Estimation of the Spectral Parameters of a Stationary Point Process". In: *The Annals of Statistics* 9.3, pp. 615–627.

Veen, A. and F.P. Schoenberg (2008). "Estimation of Space–Time Branching Process Models in Seismology Using an EM–Type Algorithm". In: *Journal of the American Statistical Association* 103.482, pp. 614–624.

Villani, M., M. Quiroz, R. Kohn, and R. Salomone (2022). "Spectral Subsampling MCMC for Stationary Multivariate Time Series with Applications to Vector ARTFIMA Processes". In: *Econometrics and Statistics*.

Watson, H. W. and F. Galton (1875). "On the Probability of the Extinction of Families". In: *The Journal of the Anthropological Institute of Great Britain and Ireland* 4, pp. 138–144.

Whittle, P (1952). "Some results in time series analysis". In: *Scandinavian Actuarial Journal* 1952.1-2, pp. 48–60.

Xu, H., M. Farajtabar, and H. Zha (2016). "Learning Granger Causality for Hawkes Processes". In: *Proceedings of The 33rd International Conference on Machine Learning*. Ed. by Maria Florina Balcan and Kilian Q. Weinberger. Vol. 48. Proceedings of Machine Learning Research. New York, New York, USA: PMLR, pp. 1717–1726.

Yang, J. and Y. Guan (2024). "Fourier analysis of spatial point processes". In: Preprint at `https://arxiv.org/abs/2401.06403`.

Yang, Y., J. Etesami, N. He, and N. Kiyavash (2017). "Online Learning for Multivariate Hawkes Processes". In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett. Vol. 30. California: Curran Associates, Inc.

Zhang, C. (2016). "Modeling High Frequency Data Using Hawkes Processes with Power-law Kernels1". In: *Procedia Computer Science* 80. International Conference on Computational Science 2016, ICCS 2016, 6-8 June 2016, San Diego, California, USA, pp. 762–771.

Zuo, S., H. Jiang, Z. Li, T. Zhao, and H. Zha (2020). "Transformer Hawkes Process". In: *Proceedings of the 37th International Conference on Machine Learning*. California: PMLR, pp. 11692–11702.

Inference of non-linear or imperfectly observed Hawkes processes

Abstract

The Hawkes point process is a popular statistical tool to analyse temporal patterns. Modern applications propose extensions of this model to account for specificities in each field of study, which in turn complexifies the task of inference. In this thesis, we advance different approaches for the parametric estimation of two submodels of the Hawkes process in univariate and multivariate settings. Motivated by the modelling of complex neuronal interactions observed from spike train data, our first study focuses on accounting for both inhibition and excitation effects between neurons, modelled by the non-linear Hawkes process. We derive a closed-form expression of the log-likelihood in order to implement a maximum likelihood procedure. As a consequence of our approach, we gain access to a goodness-of-fit scheme allowing us to establish ad hoc model selection methods to estimate the interaction network in the multivariate setting. The second part of this thesis focuses on studying Hawkes process data noised by two different alterations: adding or removing points. The absence of knowledge on the noise dynamics makes classical inference procedures intractable or computationally expensive. Our solution is to leverage the spectral analysis of point processes to establish an estimator obtained by maximising the spectral log-likelihood. By deriving the spectral densities of the noised processes and by establishing identifiability conditions on our model, we show that the spectral inference method does not necessitate any information on the structure of the noise, effectively circumventing this issue. An additional result of the study of Hawkes processes with missing points is that it gives access to a subsampling paradigm to enhance the estimation methods by introducing a penalisation parameter. We illustrate the efficiency of all of our methods through reproducible numerical implementations.

Keywords: hawkes processes, parametric inference, identifiability, inhibition, spectral theory, neuronal data

---

Résumé

Le processus ponctuel de Hawkes est un outil statistique très répandu pour analyser des dynamiques temporelles. Les applications modernes des processus de Hawkes proposent des extensions du modèle initial pour prendre en compte certaines caractéristiques spécifiques à chaque domaine d'étude, ce qui complexifie les tâches d'inférence. Dans cette thèse, nous proposons différentes contributions à l'estimation paramétrique de deux variantes du processus de Hawkes dans les cadres univarié et multivarié. Motivée par la modélisation d'interactions complexes au sein d'une population de neurones, notre première étude porte sur la prise en compte conjointe d'effets excitateurs et inhibiteurs entre les signaux émis par les neurones au cours du temps, modélisés par un processus de Hawkes non-linéaire. Dans ce modèle, nous obtenons une expression explicite de la log-vraisemblance qui nous permet d'implémenter une procédure de maximum de vraisemblance. Nous établissons également une méthode de sélection de modèle qui fournit notamment une estimation du réseau d'interactions dans le cadre multivarié. La deuxième partie de cette thèse est consacrée à l'étude des processus de Hawkes bruités par deux types d'altérations : l'ajout ou la suppression de certains points. Le manque d'information lié à ces mécanismes de bruit rend les méthodes classiques d'inférence non-applicables ou numériquement coûteuses. Notre solution consiste à s'appuyer sur l'analyse spectrale des processus ponctuels afin d'établir un estimateur obtenu en maximisant la log-vraisemblance spectrale. Nous obtenons l'expression des densités spectrales des processus bruités et, après avoir établi des conditions d'identifiabilité pour nos différents modèles, nous montrons que cette méthode d'inférence ne nécessite pas de connaître la structure du bruit, contournant ainsi le problème d'estimation. Notre étude sur les processus bruités donne accès à une méthode de sous-échantillonnage qui nous permet d'améliorer les approches d'estimation en introduisant un paramètre de pénalisation. Nous illustrons la performance des différentes méthodes proposées à travers des implémentations numériques reproductibles.

Mots clés : processsus de hawkes, inférence paramétrique, identifiabilité, inhibition, théorie spectrale, données neuronales

---

LPSM
LABORATOIRE DE PROBABILITÉS STATISTIQUE & MODÉLISATION