



**HAL**  
open science

# Sur l'approximation des processus de Hawkes : des séries temporelles aux théorèmes centraux limites quantitatifs

Mahmoud Khabou

► **To cite this version:**

Mahmoud Khabou. Sur l'approximation des processus de Hawkes : des séries temporelles aux théorèmes centraux limites quantitatifs. Probabilités [math.PR]. Université Paul Sabatier - Toulouse III, 2022. Français. NNT : 2022TOU30196 . tel-03979216

**HAL Id: tel-03979216**

**<https://theses.hal.science/tel-03979216>**

Submitted on 8 Feb 2023

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

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



# THÈSE

En vue de l'obtention du

## DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

---

---

Présentée et soutenue le *05/12/2022* par :

**Mahmoud KHABOU**

Sur l'approximation des processus de Hawkes:  
des séries temporelles aux théorèmes centraux limites quantitatifs

---

---

### JURY

CHARLES BORDENAVE  
MARIANNE CLAUSEL  
LAURE COUTIN  
NOUFEL FRIKHA  
EVA LÖCHERBACH  
GIOVANNI PECCATI  
ANTHONY RÉVEILLAC

Université d'Aix-Marseille  
Université de Lorraine  
Université Toulouse III  
Université Paris I  
Université Paris I  
Université du Luxembourg  
Université Toulouse III-INSA

Rapporteur  
Examinatrice  
Examinatrice  
Examinateur  
Présidente du jury  
Rapporteur  
Directeur de thèse

---

### École doctorale et spécialité :

*MITT : Domaine Mathématiques : Mathématiques appliquées*

### Unité de Recherche :

*Institut de Mathématiques de Toulouse (UMR 5219)*

### Directeur de Thèse :

*Anthony RÉVEILLAC*

### Rapporteurs :

*Charles BORDENAVE et Giovanni PECCATI*



*Lente hora, celeriter anni.*

An hour passes slowly, the years go  
by quickly.  
L'heure est lente, les années passent  
vite.



---

## Remerciements

Tout d'abord, mes remerciements vont à mon directeur de thèse Anthony Réveillac. Anthony m'a offert un excellent encadrement scientifique mais également de précieux conseils sur le monde académique. Grâce à son optimisme et une certaine attitude décontractée, j'ai passé trois belles années (et trois mois).

Je tiens également à remercier Charles Bordenave et Giovanni Peccati d'avoir accepté de rapporter cette thèse. Je suis très honoré par l'intérêt qu'ils ont montré pour mes résultats. Leurs travaux ont été cruciaux pour maints résultats prouvés dans ce manuscrit.

Mes remerciements s'adressent aussi aux membres du jury : Marianne Clausel et Laure Coutin, deux excellentes mathématiciennes avec lesquelles j'attends de travailler avec impatience, Noufel Frikha, spécialiste des schémas d'approximation stochastiques que j'ai pu recruter à mon jury avec la phrase "*Saviez-vous que vous avez le même nom de famille que ma grand-mère ?*" et Eva Löcherbach, experte des processus de Hawkes avec qui j'ai eu la chance d'échanger à plusieurs occasions.

Plusieurs chapitres de cette thèse sont le fruit de collaboration avec d'autres chercheurs, je souhaiterais donc bien les remercier. Merci à Caroline Hillairet pour sa gentillesse et sa réactivité. Merci également à Lorick Huang que j'ai pu côtoyer ces trois dernières années pour ses rappels à l'ordre quand j'avais une flemme prolongée. Travailler avec vous fût un honneur et un privilège. Je tiens aussi à remercier mon "grand-père de thèse" Nicolas Privault que j'ai pu croiser brièvement en Pologne.

Je suis également très reconnaissant aux secrétariats de l'EDMITT, de l'IMT et de l'INSA de Toulouse, et plus spécifiquement à Sandrine Bosc pour résoudre tous les problèmes administratifs en un temps record.

Pendant ces trois années, j'ai pu faire la connaissance de plusieurs doctorants formidables (dont une bonne partie a acquis le titre de docteur) sans lesquels mon doctorat n'aurait pas été le même. Je souhaiterais bien les remercier pour leur présence. J'aimerais bien exprimer ma gratitude à (docteurs) Laetitia Colombani, Michèle Romanos et Clément Steiner pour leur patience, même quand je posais les questions les plus naïves. Merci à Alban Gossard qui soutiendra quelques jours après moi et qui partage le même calvaire administratif. Je tiens à remercier mes collègues de l'INSA Toulouse, qu'ils soient doctorants : Mustapha Allabou, Julie Binard, Hugo Boulenc, Iain Henderson, Franck Kouassi, Hippolyte Labarrière, Nathanaël Munier et Amandine Pépiot, ou non-doctorants (aucun autre qualificatif ne me vient à l'esprit) : Romain Duboscq, Aldéric Joulin, Mouhamad Jradeh, Loïc Lacouture, Olivier Mazet, Laure Pédèches, Violaine Roussier-Michon, Olivier Roustant, Sandrine Scott et Pierre Weiss.

Merci et bonne chance à mes collègues de l'IMT : Lucas De Lara, Alberto González Sanz, Viviana Grasselli, Fu-Hsuan Ho, Samawel Jaballi, Anthony Muraro et

Denis Rochette.

J'aimerais bien remercier ma famille pour son soutien.

Enfin, merci à vous cher lecteur pour l'intérêt que vous portez à mon manuscrit de thèse et veuillez m'excuser pour les coquilles. Maintenant, passons aux maths !

# Contents

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introduction générale</b>  | <b>1</b>  |
| 1.1      | Processus ponctuels   | 2         |
| 1.1.1    | Processus de Poisson  | 2         |
| 1.1.2    | Processus de Hawkes   | 4         |
| 1.2      | Approximation par des séries temporelles  | 11        |
| 1.2.1    | État de l'art   | 12        |
| 1.2.2    | Contributions de la thèse   | 14        |
| 1.3      | Approximation gaussienne du processus de Hawkes normalisé                               | 17        |
| 1.3.1    | État de l'art   | 17        |
| 1.3.2    | Contributions de la thèse   | 18        |
| 1.4      | Le processus de Hawkes marqué   | 26        |
| <b>2</b> | <b>The nonlinear discrete time Hawkes process</b>                                       | <b>27</b> |
| 2.1      | Introduction  | 28        |
| 2.2      | The discrete-time model   | 30        |
| 2.2.1    | Preliminaries   | 30        |
| 2.2.2    | Construction  | 30        |
| 2.2.3    | Stability   | 34        |
| 2.3      | The continuous-time Hawkes process  | 36        |
| 2.3.1    | Preliminaries   | 36        |
| 2.3.2    | Definition and construction   | 36        |
| 2.3.3    | Markov properties of the Hawkes process with memory kernels given by an Erlang function | 38        |
| 2.4      | Convergence of the discrete-time Hawkes process to the continuous-time Hawkes process   | 41        |
| 2.4.1    | Preliminaries   | 41        |
| 2.4.2    | Main result   | 43        |
| 2.5      | Parameters calibration using regression   | 46        |
| 2.5.1    | Calibration experiment for the perturbed Hawkes process                                 | 48        |
| 2.5.2    | Calibration with a lower degree   | 49        |
| 2.6      | Lemmata   | 49        |
| <b>3</b> | <b>The Malliavin-Stein method for Hawkes functionals</b>                                | <b>61</b> |
| 3.1      | Introduction  | 61        |
| 3.2      | Notations and preliminaries   | 65        |
| 3.2.1    | Elements of stochastic analysis on the Poisson space                                    | 65        |
| 3.2.2    | Representation of the compound Hawkes process   | 67        |
| 3.2.3    | Malliavin analysis of the compound Hawkes process                                       | 69        |
| 3.2.4    | Elements on Stein's method  | 74        |
| 3.3      | Main results  | 75        |



|          |  |            |
|----------|--|------------|
| 3.3.1    | The general result . . . . .   | 75         |
| 3.3.2    | Application to the Hawkes process . . . . .  | 78         |
| 3.4      | Appendix . . . . .   | 87         |
| 3.4.1    | General estimates . . . . .  | 88         |
| 3.4.2    | Specific estimates for the exponential and the Erlang's kernels . . . . .          | 89         |
| <b>4</b> | <b>Gaussian approximation of multivariate compound Hawkes processes</b>            | <b>99</b>  |
| 4.1      | Introduction . . . . .   | 99         |
| 4.2      | Notations and preliminaries . . . . .  | 103        |
| 4.2.1    | General notations . . . . .  | 103        |
| 4.2.2    | Elements of stochastic analysis on the multivariate Poisson space . . . . .        | 104        |
| 4.2.3    | Definition of the multivariate compound Hawkes process . . . . .                   | 106        |
| 4.2.4    | Malliavin's analysis of the multivariate compound Hawkes process . . . . .         | 108        |
| 4.2.5    | Multivariate Stein's method . . . . .  | 110        |
| 4.3      | Main results . . . . .   | 111        |
| 4.3.1    | General bound . . . . .  | 111        |
| 4.3.2    | Bounds for the CLT . . . . .   | 115        |
| 4.3.3    | Proof of Theorem 4.1.5 . . . . .   | 124        |
| 4.4      | Lemmata . . . . .  | 125        |
| <b>5</b> | <b>The marked Hawkes process as a health insurance model</b>                       | <b>135</b> |
| 5.1      | Introduction . . . . .   | 135        |
| 5.2      | Quantitative CLT for the marked Hawkes process . . . . .                           | 137        |
| 5.3      | Large deviation principle . . . . .  | 140        |
| 5.3.1    | The Laplace transform of the Poisson cluster . . . . .                             | 141        |
| 5.3.2    | The Gärtner-Ellis theorem . . . . .  | 143        |
| 5.4      | The asymptotic ruin probability of the insurance policy . . . . .                  | 145        |
| 5.4.1    | Theoretical results . . . . .  | 145        |
| 5.4.2    | A numerical example . . . . .  | 147        |
| 5.5      | Lemmata . . . . .  | 150        |
| <b>6</b> | <b>Conclusions and perspectives</b>  | <b>157</b> |
| 6.1      | The approximation of the nonlinear Hawkes process by time series . . . . .         | 157        |
| 6.2      | The Malliavin-Setin method and the quantitative CLT for Hawkes processes . . . . . | 158        |
| 6.3      | The risk process driven by a marked Hawkes process . . . . .                       | 159        |

# Introduction générale

---

Cette thèse a pour but de démontrer des résultats théoriques sur les processus de Hawkes, à savoir l'approximation des processus de Hawkes dans deux directions :

1. L'approximation par des séries temporelles sur toute la droite réelle.
2. La quantification en temps long de l'approximation gaussienne du processus de Hawkes renormalisé dans le cadre d'un théorème central limite.

On commence ce chapitre introductif par la définition et la construction des processus de Hawkes, vus comme des processus ponctuels. Ensuite, une série temporelle d'approximation est proposée. Cette série temporelle est prouvée de converger vers le processus en temps continu (voir chapitre 2).

On développe par la suite des éléments de calcul de Malliavin pour les processus en temps discret. Ceci nous permet d'obtenir (de manière informelle) des formules de dualité, utilisées par la suite pour quantifier la convergence vers la limite gaussienne. La preuve détaillée sera donnée au chapitre 3 pour les processus en une dimension et au chapitre 4 pour les processus composés multivariés. Malgré le lien fait entre la représentation en série temporelle et le calcul de Malliavin dans ce chapitre introductif, le chapitre 2 est entièrement indépendant des chapitres 3 et 4.

Finalement, on introduit au chapitre 5 un nouveau modèle d'assurances basé sur le processus de Hawkes marqué dont on étudie la probabilité de ruine asymptotique.

## Contents

---

|  |           |
|--|-----------|
| <b>1.1 Processus ponctuels . . . . .</b>                             | <b>2</b>  |
| 1.1.1 Processus de Poisson . . . . .                                 | 2         |
| 1.1.2 Processus de Hawkes . . . . .                                  | 4         |
| <b>1.2 Approximation par des séries temporelles . . . . .</b>        | <b>11</b> |
| 1.2.1 État de l'art . . . . .  | 12        |
| 1.2.2 Contributions de la thèse . . . . .                            | 14        |
| <b>1.3 Approximation gaussienne du processus de Hawkes normalisé</b> | <b>17</b> |
| 1.3.1 État de l'art . . . . .  | 17        |
| 1.3.2 Contributions de la thèse . . . . .                            | 18        |
| <b>1.4 Le processus de Hawkes marqué . . . . .</b>                   | <b>26</b> |

---

## 1.1 Processus ponctuels

Les processus ponctuels sont la manière la plus naturelle de représenter l'arrivée d'événements aléatoires (points) selon un axe réel (temps). On peut citer par exemple les processus de vie et de mort, la désintégration d'une particule physique, ou bien le nombre de thons qui passent par un détroit. Afin de formaliser la définition de ces processus, commençons par définir l'espace des configurations

$$\Omega^N = \left\{ \sum_{k=1}^n \delta_{\tau_k}, 0 = \tau_0 < \tau_1 < \dots < \tau_n, n \in \mathbb{N} \cup \{\infty\} \right\}.$$

Naturellement, on confond chaque réalisation  $\omega \in \Omega^N$  avec la mesure ponctuelle

$$N(A)(\omega) = \omega(A) = \sum_{k=1}^{\infty} \delta_{\tau_k}(A),$$

où  $A$  est un borélien de  $\mathbb{R}_+$  et  $\delta$  est la masse de Dirac.  $N(A)$  est donc une variable aléatoire à valeurs dans  $\mathbb{N}$  et on note  $N_t = N((0, t])$ . L'intégrale d'une fonction contre la mesure ponctuelle  $N(dt)$  est

$$\int_{[0,t]} f(s)N(ds) = \sum_{0 \leq \tau_k \leq t} f(\tau_k).$$

*Remarque 1.1.1.* Il est tout à fait possible d'étendre notre définition à l'intégralité de la droite réelle  $\mathbb{R}$ . Ceci est le cas si l'on suppose que le processus a commencé il y a très longtemps (en  $t = -\infty$ ) pour des raisons de stationnarité.

Au processus  $N$  on attribue l'historique ou la filtration  $\mathcal{F}^N$  qui contient toute l'information jusqu'à l'instant  $t$

$$\mathcal{F}_t^N = \sigma(N(A), A \in \mathcal{B}([0, t])),$$

ainsi que l'intensité conditionnelle

$$\lambda_t = \lim_{dt \rightarrow 0^+} \frac{\mathbb{E}[N((t, t+dt)) | \mathcal{F}_{t-}^N]}{dt},$$

quand cette limite existe. L'intensité conditionnelle mesure le taux instantané d'arrivée d'événements, par unité de temps. Elle peut être vue comme la projection prévisible du processus  $N$ , en effet on montre que si  $\lambda$  existe alors le processus  $M_t^C = \int_{[0,t]} C_s(N(ds) - \lambda_s ds)$  où  $C$  est un processus prévisible est une  $\mathcal{F}^N$ -martingale (cf. [17]).

### 1.1.1 Processus de Poisson

Il est possible de construire un processus ponctuel  $N$  pour lequel l'intensité conditionnelle  $\lambda_t = \lambda$  une constante positive et déterministe. En effet, si l'on prend  $(W_k)_{k \in \mathbb{N}}$  des variables aléatoires indépendantes et identiquement distribuées (*i.i.d*)

de loi exponentielle  $\mathcal{E}(\lambda)$  et que l'on définit de manière récursive  $\tau_k = \tau_{k-1} + W_k$  pour  $k \geq 1$ , il est possible de démontrer que le processus défini par  $N_t = \sum_{k \in \mathbb{N}} \mathbb{1}_{\tau_k \leq t}$  vérifie

$$\lim_{dt \rightarrow 0^+} \frac{\mathbb{E}[N((t, t + dt)) | \mathcal{F}_{t-}^N]}{dt} = \lambda,$$

et que, plus généralement (*cf.* [25] par exemple)

$$\begin{aligned} \mathbb{P}[N((a_1, b_1]) = k_1, \dots, N((a_m, b_m]) = k_m] \\ = e^{\lambda(b_1 - a_1)} \frac{(\lambda(b_1 - a_1))^{k_1}}{k_1!} \times \dots \times e^{\lambda(b_m - a_m)} \frac{(\lambda(b_m - a_m))^{k_m}}{k_m!} \end{aligned} \quad (1.1.1)$$

dès que  $(a_1, b_1], \dots, (a_m, b_m]$  sont disjoints. Le processus  $N$  est dit un processus de Poisson *homogène*, et d'après (1.1.1) il vérifie les propriétés suivantes :

1. Stationnarité : les lois dépendent uniquement de la taille de l'intervalle et pas de leurs positions.
2. Indépendance des incréments : le nombre d'évènements dans un intervalle n'a aucune influence sur le nombre d'évènements aux intervalles suivants, du moment qu'ils sont disjoints.

Ce processus garantit la distribution la plus équirépartie des évènements dans un intervalle. En effet, il est possible de prouver que, sachant que le nombre de points dans un intervalle  $A$  est  $n$ , ces points sont distribués selon une loi uniforme.

Il se peut cependant que la distribution des points ne soit pas équirépartie, par exemple dû à un phénomène de saisonnalité qui impliquerait une forte mobilité des thons en été (par opposition à une sédentarité hivernale) [88]. Dans ce cas il est judicieux d'utiliser un processus de comptage  $N$  avec une intensité variable  $(\lambda_t)_{t \in \mathbb{R}_+}$ . Ce choix est suffisant pour modéliser les saisonnalités régulières mais il devient obsolète du moment où les variables qui influent l'arrivée sont elles mêmes aléatoires, on peut par exemple penser aux fluctuations de la température entre les années. On pourrait donc, en toute généralité passer à un processus doublement stochastique, ou bien un processus de Cox [23]. L'intensité  $\lambda_t(\omega')$  est un processus aléatoire construit sur un espace probabilisé  $\Omega^\lambda$ . Si l'espace  $\Omega^\lambda$  est différent de  $\Omega^N$ , le processus  $N$  devient un processus de Poisson (éventuellement inhomogène) d'intensité  $(\lambda_t(\omega'))_{t \in \mathbb{R}_+}$  sachant  $\omega'$ .

Ce choix reste cependant trop général : les variables pouvant influencer l'intensité du processus ponctuel sont trop nombreuses et leur mesure peut être trop coûteuse. On peut donc naturellement penser à la dépendance de l'intensité du passé du processus.

Cette dépendance est assez logique, on peut par exemple supposer l'existence d'une certaine hiérarchie au sein d'un banc de thons, et que certains poissons sont des poissons "alpha" qui dirigent leurs groupes. Le passage d'un thon est corrélé avec l'observation d'autres poissons juste après.

### 1.1.2 Processus de Hawkes

Les processus de Hawkes sont des processus ponctuels à intensité aléatoire exhibant une auto-régression. On commence cette section par donner la définition générale du processus de Hawkes univarié

**Définition 1.1.2.** Soit  $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}$ ,  $\psi : \mathbb{R} \rightarrow \mathbb{R}_+$  et  $\mu \in \mathbb{R}$ .

On dit qu'un processus ponctuel  $N$  est un *processus de Hawkes* si son intensité conditionnelle  $\lambda$  est de la forme

$$\begin{aligned}\lambda_t &= \psi \left( \mu + \int_{[0,t)} \phi(t-s) dN_s \right) \\ &= \psi \left( \mu + \sum_{\tau_i < t} \phi(t - \tau_i) \right),\end{aligned}$$

où les  $(\tau_i)_{i \in \mathbb{N}}$  sont les instants de saut de  $N$ .

#### 1.1.2.1 Cas linéaire

Le processus introduit par Alan G. Hawkes [46] est un processus ponctuel dont l'intensité aléatoire dépend du passé, proche ou loin, des événements. Il s'agit d'un processus auto-régressif où la réalisation d'un événement augmente la probabilité d'observer d'autres points selon la dynamique

$$\lambda_t = \mu + \int_{[0,t)} \phi(t-s) dN_s, \quad (1.1.2)$$

où  $\mu \in \mathbb{R}_+$  joue le rôle d'intensité de base, c'est à dire le taux d'arrivée des points indépendamment du passé et  $\phi$  est un *noyau d'auto-interaction* qui dans le cas linéaire est responsable de l'auto-excitation.

*Remarque 1.1.3.* Si le processus est défini sur toute la droite réelle  $\mathbb{R}$ , l'équation (1.1.2) devient

$$\lambda_t = \mu + \int_{(-\infty,t)} \phi(t-s) dN_s.$$

Ce choix est fait surtout quand on s'intéresse à la stationnarité de  $N$ .

Afin que la relation (1.1.2) ait un sens, il faut évidemment que  $\phi$  soit positive. Ceci n'est pourtant pas suffisant pour avoir un processus ponctuel non-explosif (c'est à dire  $\lim_{n \rightarrow +\infty} \tau_n = +\infty$  presque sûrement). Bacry *et al.* [7] ont montré qu'une condition suffisante de non-explosion est

$$\int_0^t \phi(s) ds < +\infty$$

pour tout  $t \geq 0$ .

Dans la plupart des applications, supposer la non-explosion ne suffit pas. En effet, on aimerait bien avoir une certaine stabilité dans la dynamique. Ceci revient à exiger

la stationnarité (éventuellement asymptotique) du processus et de son intensité. Une condition nécessaire pour avoir une intensité stationnaire est la stagnation de son premier moment. L'espérance de l'équation (1.1.2) donne

$$\begin{aligned} m &:= \lim_{t \rightarrow +\infty} \mathbb{E}[\lambda_t] = \mu + \lim_{t \rightarrow +\infty} \int_0^t \phi(t-s) \mathbb{E}[dN_s], \\ &= \mu + \lim_{t \rightarrow +\infty} \int_0^t \phi(s) \mathbb{E}[dN_{t-s}], \\ &= \mu + \lim_{t \rightarrow +\infty} \int_0^t \phi(s) \mathbb{E}[\lambda_{t-s}] ds, \\ &= \mu + \|\phi\|_1 m. \end{aligned}$$

Cette équation en  $m$  a une solution positive unique  $m = \mu/(1 - \|\phi\|_1)$  si et seulement si

$$\|\phi\|_1 = \int_0^{+\infty} \phi(s) ds \in [0, 1). \quad (1.1.3)$$

En fait, la condition (1.1.3) est une condition nécessaire et suffisante pour l'existence d'un processus de Hawkes  $N^*$  construit sur  $\mathbb{R}$  tel que  $N^*(C+t)$  et  $N^*(C)$  ont la même loi pour tout  $t \in \mathbb{R}$  cf. [26] par exemple.

On suppose dorénavant que la condition (1.1.3) est vérifiée.

### Le processus linéaire vu comme un processus de branchement

Le processus défini via l'équation (1.1.2) est donné de façon implicite, en effet, on ne sait pas *a priori* s'il existe un processus qui résout l'équation. Dans le cas linéaire, Hawkes et Oakes [47] ont donné une preuve constructive de l'existence d'un processus ponctuel  $N$  dont l'intensité satisfait (1.1.2). Il s'agit d'un processus de branchement défini de la manière récursive suivante :

1. On commence par les points ancestraux  $N^0$  qui arrivent selon un processus de Poisson homogène d'intensité  $\mu$ .
2. Indépendamment de tout, chaque ancêtre  $a \in N^0$  génère des descendants  $N_a^1$  selon un processus de Poisson inhomogène d'intensité évanescence  $\phi(\cdot - a)$ . L'ensemble des descendants forme la première génération

$$N^1 = \bigcup_{a \in N^0} N_a^1.$$

3. Supposons les générations construites jusqu'à la génération  $k$  où  $k$  est un entier donné. Alors, indépendamment de tout, chaque point  $x$  de la génération  $N^k$  engendre des descendants  $N_x^{k+1}$  selon un processus de Poisson inhomogène d'intensité  $\phi(\cdot - x)$ . La  $k+1$ -ème génération est donc

$$N^{k+1} = \bigcup_{x \in N^k} N_x^{k+1}$$

4. Le processus est la réunion de tous les points, toutes générations confondues

$$N = \bigcup_{k \in \mathbb{N}} N^k.$$

On remarque que la construction se termine en temps fini sur tout intervalle fini  $[0, A]$ . En effet, comme il est indiqué dans la thèse de Chevallier [21], chaque ancêtre génère des descendants dont le nombre moyen est  $\|\phi\|_1 < 1$ . Il s'agit donc d'un processus de Galton-Watson sous-critique dont l'extinction aura lieu avec probabilité 1.

Le processus de Hawkes linéaire a été largement étudié dans la littérature. Sa structure permet d'avoir une formule explicite de sa moyenne asymptotique ainsi que de dériver une loi des grands nombres et un théorème de la limite centrale avec une variance explicite [7]. De plus, grâce à sa structure de processus de branchement, Stabile et Torrisi [86] ont explicité un principe de grandes déviations pour le processus de Hawkes linéaire.

Cette simplicité a cependant un coût ; le processus de Hawkes linéaire n'autorise que l'excitation comme auto-regression. La positivité de  $\phi$  implique que l'arrivée de chaque point engendre l'augmentation de l'intensité  $\lambda$  et donc de la probabilité de voir plus d'évènements dans le futur. Cependant, dans certaines applications on observe un effet d'auto-inhibition : la réalisation d'un évènement repousse les autres évènements. Ceci est par exemple le cas dans certains réseaux neuronaux [19]. Pour palier à ce manque, on introduit le processus de Hawkes non linéaire.

### 1.1.2.2 Cas non linéaire

Soit  $\psi$  une fonction de  $\mathbb{R}$  à valeurs positives, et  $\phi$  une fonction à valeurs réelles. Un processus ponctuel  $N$  est un processus de Hawkes non linéaire si son intensité  $\lambda$  suit la dynamique suivante

$$\lambda_t = \psi \left( \mu + \int_{[0,t)} \phi(t-s) dN_s \right). \quad (1.1.4)$$

La fonction  $\psi$  est appelée *fonction d'intensité* et elle est généralement supposée croissante. Si l'on suppose  $\phi$  négative, l'arrivée d'un point, disons à l'instant  $\tau_n$ , fait diminuer la quantité

$$X_t = \int_{[0,t)} \phi(t-s) dN_s = \sum_{\tau_k < t} \phi(t - \tau_k)$$

de  $|\phi(t - \tau_n)|$  et comme  $\psi$  est croissante, ceci aura l'effet de diminuer  $\lambda$  et donc la probabilité d'avoir d'autres sauts dans le futur.

Contrairement au processus linéaire, la représentation comme un processus de branchement n'est plus valable pour construire le processus de Hawkes non linéaire. On a donc recours à une autre technique plus générale.

### La procédure d'amincissement (ou thinning)

Cette procédure, introduite par Lewis et Shedler [60] pour les processus de Poisson inhomogènes a été adaptée par Brémaud et Massoulié [18] pour la construction du processus de Hawkes non-linéaires (*cf.* [69] pour une approche de simulation pour les processus linéaires). Elle consiste à extraire un processus ponctuel défini sur  $\mathbb{R}_+$  (ou  $\mathbb{R}$  plus généralement) à partir d'une mesure de Poisson bi-dimensionnelle, qui joue le rôle d'aléa sous-jacent.

Posons  $\Pi(dt, d\theta)$  une mesure de Poisson sur  $\mathbb{R}_+^2$  d'intensité  $dt d\theta$ , c'est-à-dire

$$\mathbb{E} \left[ \int_B f(t, \theta) \Pi(dt, d\theta) \right] = \int_B f(t, \theta) dt d\theta,$$

sur tout borélien  $B$  de  $\mathbb{R}_+^2$  et pour toute fonction  $f$ . Tout comme le processus de Poisson homogène, la mesure de Poisson est une façon de simuler des points "équi-répartis" dans le plan. Étant donnée une trajectoire positive prévisible  $(\lambda_t)_{t \in \mathbb{R}_+}$ , c'est-à-dire mesurable par rapport à la tribu

$$\mathcal{F}_{t-}^{\Pi} = \sigma(\Pi(A, B), A \in \mathcal{B}([0, t]), B \in \mathcal{B}(\mathbb{R}_+)),$$

le processus ponctuel  $N$  défini par les points de  $\Pi$  dans la bande  $([0, \lambda_t])_{t \in \mathbb{R}_+}$  est un processus ponctuel d'intensité  $\lambda$ . En effet, le processus de Hawkes non-linéaire est défini trajectoriellement de façon unique comme la solution à l'équation différentielle stochastique (EDS)

$$\begin{cases} N_t = \int_{[0, t] \times \mathbb{R}_+} \mathbf{1}_{\theta \leq \lambda_s} \Pi(ds, d\theta), \\ \lambda_t = \psi \left( \mu + \int_{[0, t] \times \mathbb{R}_+} \phi(t-s) \mathbf{1}_{\theta \leq \lambda_s} \Pi(ds, d\theta) \right) \end{cases} \quad (1.1.5)$$

*Remarque 1.1.4.* Le système (1.1.5) est triangulaire. En effet, on peut écrire la deuxième équation en fonction de la dynamique de l'inconnue  $\lambda$  et de l'aléa sous-jacent  $\Pi$ , sans le processus  $N$ .

Ceci justifie que l'on s'intéresse surtout à  $\lambda$ , les sauts de  $N$  y étant inclus.

*Remarque 1.1.5.* La construction par amincissement reste valable pour les processus de Hawkes linéaires. En effet, pour avoir un processus linéaire il suffit de prendre  $\psi(x) = x$  et  $\phi(x) \geq 0$  pour tout  $x \in \mathbb{R}_+$ .

L'amincissement d'une mesure de Poisson pour en extraire un processus de Hawkes non-linéaire d'intensité  $\lambda_t = (\mu + X_t)_+$  où

$$X_t = \int_{[0, t]} \frac{-1}{2(1 + (t-s)^2)} dN_s,$$

est illustré sur la figure 1.1

Concernant l'existence d'une version stationnaire d'un processus de Hawkes  $N$  non linéaire, Brémaud et Massoulié [18] donnent deux conditions suffisantes :

1. La fonction  $\psi$  est bornée.
2. La fonction  $\psi$  est  $L$ -lipschitzienne et  $L\|\phi\|_1 < 1$ .



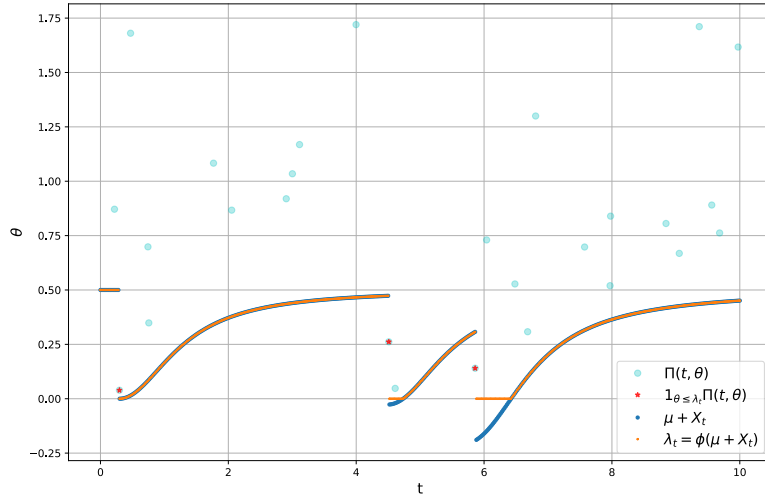


FIGURE 1.1 – Amincissement avec une intensité de base  $\mu = 0.5$ . Les points acceptés (dans la bande  $[0, \lambda_t]$ ) sont marqués par des étoiles rouges. La "pré-intensité"  $\mu + X_t$  peut aller en dessous de zéro, mais pas l'intensité qui en est la partie positive.

En effet, si on suppose la première condition satisfaite, le processus  $N$  est stochastiquement dominé par un processus de Poisson d'intensité  $\|\psi\|_\infty$  issu d'un thinning de la même mesure de Poisson sous-jacente. La logique derrière la deuxième condition est la suivante : comme  $\psi$  est  $L$ -lipschitzienne, on a la majoration suivante

$$\begin{aligned} \lambda_t = \psi \left( \mu + \int_{[0,t)} \phi(t-s) dN_s \right) &\leq \psi(\mu) + L \left| \int_{[0,t)} \phi(t-s) dN_s \right|, \\ &\leq \psi(\mu) + \int_{[0,t)} L|\phi(t-s)| dN_s. \end{aligned}$$

Il est donc possible de majorer stochastiquement par un processus de Hawkes linéaire de noyau  $L|\phi|$ .

On tient à préciser que cette deuxième condition peut être relaxée d'avantage, en remplaçant  $L\|\phi\|_1$  par  $L\|(\phi)_+\|_1$ .

### 1.1.2.3 Influence du choix du noyau

Jusqu' alors, on a vu que le choix du noyau  $\phi$  avait une influence sur la nature du processus  $N$ . Mais cette influence ne se résume pas au poids total du noyau  $\|\phi\|_1$ , mais la forme de  $\phi$  y joue un rôle. En effet, si  $\phi$  était concentrée autour de zéro, alors les évènements les plus récents sont ceux qui ont la plus grande influence sur le futur. Par contre, si le noyau avait une queue lourde, les évènements du passé, même très lointain, ont toujours une influence sur l'intensité. Il en découle alors, que ni le processus  $N$  ni son intensité  $\lambda$  ne sont des processus de Markov pour un

noyau général, car le passé lointain peut toujours influencer la dynamique. Cependant, pour le choix particulier

$$\phi(t) = \alpha e^{-\beta t}, \quad \beta > 0,$$

Errais *et al.* [35] ont démontré que le processus  $X$  (respectivement le vecteur  $(X, N)$ ) où  $X_t = \int_{[0,t)} \alpha e^{-\beta(t-s)} dN_s$  était un processus de Markov (bi-dimensionnel) dans le cas linéaire.

En effet, grâce à la mémoire courte de la fonction exponentielle, on a

$$e^{\beta t} X_t = \alpha \int_{[0,t)} e^{\beta s} dN_s,$$

ce qui donne par différenciation

$$(dX_t + \beta X_t dt) e^{\beta t} = \alpha e^{\beta t} dN_t,$$

et après réarranger les termes

$$dX_t = -\beta X_t dt + \alpha dN_t. \tag{1.1.6}$$

Grâce à cette EDS on obtient une formule de Dynkin pour  $\lambda$  (dans le cas linéaire)

$$\partial_t \mathbb{E}[f(\lambda_t)] = \mathbb{E} [\beta(\mu - \lambda_t) f'(\lambda_t) + \lambda_t (f(\lambda_t + \alpha) - f(\lambda_t))]$$

dont découlent les expressions explicites pour les moments de  $\lambda$  et  $N$ , si le processus de Hawkes est linéaire. Par exemple, l'espérance de l'intensité d'un processus de Hawkes de noyau  $\phi(t) = \alpha e^{-\beta t}$ ,  $\alpha > 0$  est

$$\mathbb{E}[\lambda_t] = \frac{\mu}{1 - \|\phi\|_1} \left( 1 - \|\phi\|_1 e^{(\alpha - \beta)t} \right),$$

où  $\|\phi\|_1 = \alpha/\beta \neq 1$ . Il est possible de voir que cette espérance converge vers la valeur asymptotique  $\mu/(1 - \|\phi\|_1)$  sus-mentionnée si et seulement si  $\|\phi\|_1 < 1$ , ou de manière équivalente  $\alpha < \beta$ .

*Remarque 1.1.6.* L'EDS (1.1.6) de la pré-intensité est vraie pour les processus de Hawkes non-linéaires aussi. Cependant, on n'a pas de formule explicite pour les moments de l'intensité dans ce cas à cause de la non-linéarité.

Le choix d'un noyau exponentiel, garantissant la structure processus de Markov déterministe par morceaux (*PDMP*) à l'intensité, permet aussi d'obtenir un algorithme de simulation exacte des temps de saut d'un processus de Hawkes linéaire. Ceci a été fait par Dassios et Zhao [28], en remarquant qu'entre deux instants de saut  $\tau_k$  et  $\tau_{k+1}$ , l'intensité évolue selon la dynamique déterministe

$$\lambda_t = (\lambda_{\tau_k+} - \mu) e^{-\beta(t-\tau_k)} + \mu.$$

L'avantage de cet algorithme par rapport à la procédure de thinning ordinaire c'est qu'il ne requiert pas le parcours de tout l'historique pour simuler le point suivant, grâce à la structure Markovienne des intensités à noyau exponentiel.

De manière similaire, Duarte *et al.* [30] ont présenté un algorithme pour simuler les processus de Hawkes non-linéaires satisfaisant les conditions de stabilité. Dans le même article, on montre que la pré-intensité  $X$ , en tant que processus de Markov converge vers une distribution stationnaire avec une vitesse exponentielle (toujours en supposant que les conditions de stabilité sont vérifiées). Ceci est prouvé avec des techniques classiques des processus de Markov comme les potentiels de Foster-Lyapunov et la récurrence de Harris.

On tient à préciser que les résultats de cet article sont établis non seulement pour des noyaux exponentiels, mais pour une classe plus générale qui est les noyaux de type Erlang. Les noyaux de type Erlang sont des produits entre une exponentielle et un polynôme. Le caractère Markov est préservé pour cette classe de noyaux, car ils résolvent une équation différentielle linéaire à coefficients constants, quoique en dimension supérieure. pour s'en convaincre, supposons que

$$X_t = \int_{[0,t)} \alpha \frac{(t-s)^p}{p!} e^{-\beta(t-s)} dN_s$$

c'est-à-dire le produit entre un monome et ne exponentielle, et posons les processus auxiliaires  $U_t^i = \int_{[0,t)} \alpha (t-s)^i (i!)^{-1} e^{-\beta(t-s)} dN_s$  pour  $i = 0 \dots, p-1$ . Alors, le vecteur  $\mathbf{X}_t = (X_t, U_t^{p-1}, \dots, U_t^0)$ , appelé *cascade* markovienne, résout l'équation différentielle stochastique suivante

$$d\mathbf{X}_t = (K - \beta I)\mathbf{X}_t dt + \mathbf{e}_{p+1} dN_t,$$

où  $K$  est la matrice avec 1 à la première sur-diagonale et 0 ailleurs, et  $\mathbf{e}_{p+1}$  est le vecteur de  $\mathbb{R}^{p+1}$  contenant 1 à la dernière composante et zéro ailleurs.

#### 1.1.2.4 Généralisations

Dans certains domaines d'application, les évènements d'un processus ponctuel n'ont pas seulement un impact sur le processus donné, mais aussi sur d'autres processus ponctuels. Ceci est le cas par exemple pour les réseaux de neurones, où le déclenchement d'un signal par un neurone peut exciter ou bien inhiber un neurone voisin (*cf.* [80] par exemple). On définit donc le processus de Hawkes multivarié  $\mathbf{N} = (N_1, \dots, N_d)$  pour un système de  $d$  particules, avec un intensité  $\lambda^i$  de chaque particule qui suit la dynamique

$$\lambda_t^i = \psi_i \left( \mu_i + \sum_{j=1}^d \int_{[0,t)} \phi_{ij}(t-s) dN_s^j \right), \quad i = 1, \dots, d.$$

Le noyau  $\phi_{ij}$  encode l'influence de la composante  $j$  du système sur la composante  $i$ . Et donc si  $\phi_{ij} > 0$  la composante  $j$  a un effet d'excitation sur la  $i$ -ème composante, un effet d'inhibition si  $\phi_{ij} < 0$  et aucune influence si  $\phi_{ij} = 0$ .

Les résultats établis pour les processus uni-variés sont transposables aux processus

multi-variés. En fait, dans le cas linéaire ( $\psi_i(x) = x, \forall i = 1, \dots, d$ ) la condition pour avoir une espérance d'intensité asymptotique devient

$$sp(K) < 1,$$

où  $K = (\|\phi_{ij}\|_1)_{i,j=1,\dots,d}$  et le résultat, donnée explicitement dans [7]

$$\begin{cases} T^{-1}\mathbf{N}_T & \longrightarrow (I - K)^{-1}\boldsymbol{\mu} \text{ p.s.} \\ \mathbb{E}[\boldsymbol{\lambda}_T] & \longrightarrow (I - K)^{-1}\boldsymbol{\mu} \end{cases} \quad (1.1.7)$$

quand  $T$  croît à l'infini. Pour plus de détails sur les processus de Hawkes multivariés on réfère à la thèse de Liniger [62].

Une autre façon d'étendre les processus de Hawkes consiste à considérer les processus composés, où à chaque saut correspond une perte ponctuelle  $Y_i$  qui peut modéliser la force d'un tremblement de terre ou le montant d'un sinistre signalé à un assureur. On définit alors le processus de perte totale

$$L_t = \sum_{k=1}^{N_t} Y_k,$$

où les  $Y_k$  sont des variables *i.i.d* indépendantes de  $N$ . On suppose également que l'intensité  $\lambda$  est influencée par la taille de la perte : une grosse perte déclenche plus de répliques dans le marché, un gros séisme est plus susceptible d'être suivi par des tremblements de terre etc. L'intensité de  $N$  devient

$$\begin{aligned} \lambda_t &= \mu + \int_{[0,t)} \phi(t-s) dL_s \\ &= \mu + \sum_{\tau_k < t} \phi(t - \tau_k) Y_k. \end{aligned}$$

De même, les résultats de stabilité et les expressions des moments dans le cas markovien linéaire restent transposable au processus composé [35].

## 1.2 Approximation par des séries temporelles

Avant d'entamer a thèse, j'avais suivi des cours de physique. Par conséquent, le seul moyen qui m'est paru judicieux pour construire et simuler intuitivement l'intégrale

$$\lambda_t = \mu + \int_{[0,t)} \phi(t-s) dN_s$$

était d'appliquer le conseil de Mark Kac "*Be wise, discretize!*"<sup>1</sup>. J'étais prêt à perdre un peu de précision afin d'aviter la construction en tant que processus de branchement qui m'est parue très exigeante en terme de mémoire et assez déconnectée de la dynamique en temps. Comme on va le voir dans cette section, la construction d'un processus de Hawkes sur une grille de temps discrets est directe, mais comme c'est le cas parfois en mathématiques, c'est le passage à la limite qui n'est pas évident.

1. Soyez sage, discrétisez!

### 1.2.1 État de l'art

Quoique le processus de Hawkes soit défini sur un continuum de temps ( $\mathbb{R}_+$  ou  $\mathbb{R}$ ), les données dans la vraie vie sont souvent acquises de manière discrète dans le temps, disons par exemple chaque  $h = 1$  ms. Ici  $h$  joue le rôle d'un pas de temps. Les séries temporelles définies sur  $\mathbb{N}$  (ou  $\mathbb{Z}$ ) deviennent un outil privilégié pour décrire l'évolution sur une grille de temps  $(t_i)_{i \in \mathbb{N}} = (ih)_{i \in \mathbb{N}}$ .

La manière la plus naturelle pour définir un processus ponctuel à temps discret sur une grille de temps est de discrétiser sa dynamique en remplaçant

$$N_t = \int_{[0,t]} dN_s$$

par

$$\tilde{N}_{t_k}^h = \sum_{i=1}^k \Delta \tilde{N}_{t_i}^h,$$

où  $\Delta \tilde{N}_{t_i}^h$  mesure le nombre de sauts dans l'intervalle  $((i-1)h, ih]$ . La notion d'intensité conditionnelle est directement transposable dans le cadre discret ; en effet,  $\tilde{\lambda}_{t_k}^h$  s'agit de la meilleure approximation de  $\Delta \tilde{N}_{t_k}^h$  sachant  $\tilde{\mathcal{F}}_{t_{k-1}}^h = \sigma(\Delta \tilde{N}_{t_0}^h, \dots, \phi \tilde{N}_{t_{k-1}}^h)$ , c'est-à-dire

$$\mathbb{E} \left[ \Delta \tilde{N}_{t_k}^h \mid \tilde{\mathcal{F}}_{t_{k-1}}^h \right] = \tilde{\lambda}_{t_k}^h h. \quad (1.2.1)$$

Il est important de conserver le pas de temps  $h$  dans la définition, en effet, à intensité égale, l'observation d'un saut est d'autant plus probable que l'intervalle est long. Plus spécifiquement en se rappelant que l'on s'intéresse exclusivement aux processus ponctuels simples (c'est à dire qui n'autorisent qu'un seul saut à un instant donné), on peut définir  $\Delta \tilde{N}_{t_k}^h$  conditionnellement à  $\tilde{\mathcal{F}}_{t_{k-1}}^h$  comme une loi de Bernoulli de paramètre  $\tilde{\lambda}_{t_k}^h h$ , ou plus précisément

$$\Delta \tilde{N}_{t_k}^h = \mathbb{1}_{U_k \leq \tilde{\lambda}_{t_k}^h h}, \text{ sachant } \tilde{\mathcal{F}}_{t_{k-1}}^h, \quad (1.2.2)$$

où  $(U_k)_{k \in \mathbb{N}}$  est une suite *i.i.d* de variables uniformes sur  $[0, 1]$ .

*Remarque 1.2.1.* La suite de variables uniformes  $(U_k)_{k \in \mathbb{N}}$  est l'équivalent discret de l'aléa sous-jacent  $\Pi$  à partir duquel on extrait le processus ponctuel par amincissement.

Pour conférer une structure auto-excitatrice à notre processus discret, l'intensité à temps discret se définit de la manière auto-regressive suivante :

$$\begin{aligned} \tilde{\lambda}_{t_k}^h &= \mu + \sum_{i=1}^{k-1} \phi(t_k - t_i) \Delta \tilde{N}_{t_i}^h, \\ &= \mu + \sum_{i=1}^{k-1} \phi((k-i)h) \mathbb{1}_{U_i \leq \tilde{\lambda}_{t_i}^h h}. \end{aligned}$$

qui n'est rien d'autre que la discrétisation de (1.1.2). On remarque que la somme s'arrête à  $k-1$  et non pas  $k$  afin de garantir la prévisibilité de  $\tilde{\lambda}^h$ .

Cette définition intuitive des processus de Hawkes discrets figure dans le papier de Seol [83] dans lequel il prove une loi des grands nombres ainsi qu'un théorème centrale limite pour le processus  $\tilde{N}^h$ .

Malgré sa simplicité et sa proximité d'un processus ponctuel quand la taille du pas de temps  $h$  tend vers zéro, le choix d'une variable de Bernoulli pour définir l'accroissement  $\Delta\tilde{N}$  a ses inconvénients :

1. Le fait que le paramètre d'une loi de Bernoulli (ici  $\tilde{\lambda}^h h$ ) puisse dépasser 1 rend les calculs moins directs : il faudra toujours séparer les deux cas  $\tilde{\lambda}^h h \leq 1$  et  $\tilde{\lambda}^h h > 1$  qui n'est pas à exclure.
2. Lors de l'échantillonnage pour des valeurs moyennement petits de  $h$ , il se peut que l'on observe la réalisation de deux événements ou plus dans le même intervalle. Ceci n'est pas autorisé par les lois de Bernoulli.

C'est pour ces raisons que Kirchner [56] construit son processus discret avec des variables de Poisson

$$\Delta\tilde{N}_{t_k}^h = \text{Pois}\left(\tilde{\lambda}_{t_k}^h h\right), \text{ sachant } \tilde{\mathcal{F}}_{t_{k-1}},$$

qui satisfont la relation (1.2.1). Plus précisément, il propose le modèle INAR( $\infty$ ) (auto-regressif entier d'ordre infini) suivant

**Définition 1.2.2.** Soient  $\epsilon_n$  des variables *i.i.d* de loi  $\text{Pois}(\mu h)$  pour tout  $n \in \mathbb{Z}$ , et  $\xi_l^{(n,k)}$  des variables aléatoires de loi  $\text{Pois}(h\phi(kh))$  indépendamment sur  $n \in \mathbb{Z}$ ,  $k \in \mathbb{N}$ ,  $l \in \mathbb{N}$  et des  $(\epsilon_n)$ . Une série temporelle INAR( $\infty$ ) est une suite de variables aléatoires  $(P_n)_{n \in \mathbb{Z}}$  qui résout le système des equations aux différences stochastiques

$$P_n = \sum_{k=1}^{\infty} \sum_{l=1}^{P_{n-k}} \xi_l^{(n,k)} + \epsilon_n, \quad n \in \mathbb{Z}.$$

On appelle  $K_h = \sum_{k=1}^{\infty} \phi(kh)h$  la moyenne de reproduction.

La suite  $P_n$  dans cette définition modélise le nombre d'évènement dans l'intervalle  $((n-1)h, nh]$ , c'est à dire

$$\Delta\tilde{N}_{t_n}^h = P_n.$$

*Remarque 1.2.3.* Bien que les séries temporelles INAR( $\infty$ ) aient été introduites dans les travaux de Kirchner, les séries INAR( $p$ ) ou  $p$  est un entier fini ont été introduites dans les travaux de Alzaid et Al-Osh [2].

L'originalité du travail de Kirchner est de formaliser le lien entre son modèle de série temporelle (INAR( $\infty$ )) et le processus de Hawkes linéaire en temps continu défini par la relation (1.1.2). En effet, si l'on définit la mesure aléatoire

$$\tilde{N}^h(A) = \sum_{k:kh \in A} P_k, \quad A \in \mathcal{B}(\mathbb{R}),$$

alors

$$\tilde{N}^h \longrightarrow N$$

faiblement dans la topologie de Skorokhod quand la taille de l'intervalle  $h$  tend vers 0, si  $\|\phi\|_1 < 1$  (dont découle  $K_h < 1$  pour tout  $h$  assez petit). La convergence faible dans la topologie de Skorokhod peut être vue comme la convergence

$$\lim_{h \rightarrow 0^+} \int_{\mathbb{R}} f(t) \tilde{N}^h(dt) = \int_{\mathbb{R}} f(t) N(dt)$$

faiblement pour toute fonction  $f$  positive, continue et bornée.

*Remarque 1.2.4.* La dernière limite est une conséquence de la convergence faible dans la topologie de Skorokhod et non pas une définition. Pour la définition de la convergence faible dans la topologie de Skorokhod on réfère à [14].

## 1.2.2 Contributions de la thèse

On remarque que la série temporelle définie par le modèle INAR( $\infty$ ) donne une construction de l'équivalent discret de  $N$  en contournant  $\lambda$ . Ceci nous rappelle la construction du processus de Hawkes linéaire par les processus de branchement. On ne peut donc pas, *a priori*, étendre ce type de séries temporelles aux processus de Hawkes non linéaires.

### 1.2.2.1 Approximation de l'EDS par la méthode d'Euler

Afin d'étendre l'approximation par des séries temporelles aux processus non-linéaires, on propose au chapitre 2 une autre construction qui repose sur la dynamique markovienne de la pré-intensité d'un processus dont le noyau est une fonction exponentielle. Rappelons que dans ce cas l'intensité est  $\lambda_t = \psi(\mu + X_t)$  où  $X$  résout l'EDS (1.1.6)

$$dX_t = -\beta X_t dt + \alpha dN_t.$$

Si l'on suppose maintenant que notre processus temporel est défini sur une grille de temps discrets, l'approximation la plus intuitive de (1.1.6) est un schéma d'Euler construit récursivement

$$\tilde{X}_{t_{k+1}}^h = e^{-\beta h} \tilde{X}_{t_k}^h + \alpha P_k, \text{ sachant } \tilde{X}_{t_k}^h \quad (1.2.3)$$

où  $P_k$  est une suite de variables aléatoires générées selon la loi

$$P_k \sim \text{Pois} \left( h\psi(\mu + \tilde{X}_{t_k}^h) \right).$$

La suite  $(\tilde{X}_{t_k}^h)_{k \in \mathbb{N}}$  est une chaîne de Markov, ainsi que le vecteur  $\left( (P_k, \tilde{X}_{t_{k+1}}^h) \right)_{k \in \mathbb{N}}$ . On remarque qu'entre deux instants de sauts consécutifs, la pré-intensité en temps

discret évolue de manière déterministe. Il est toutefois possible d'envisager une dynamique d'évolution selon laquelle la pré-intensité est impactée à chaque instant par un aléa externe. Ceci peut être vu comme un bruit de fond dans les réseaux neuronaux, ou bien l'influence totale des nouvelles du marché, etc. On suggère donc de modifier (1.2.3) de la façon suivante

$$\tilde{X}_{t_{k+1}}^h = e^{-\beta h} \tilde{X}_{t_k}^h + \alpha P_k + \sigma \sqrt{h} G_k, \text{ sachant } \tilde{X}_{t_k}^h \quad (1.2.4)$$

où les  $(G_k)_{k \in \mathbb{N}}$  sont des *i.i.d* gaussiennes centrées réduites, telles que pour tous  $k \in \mathbb{N}$   $G_k$  est indépendante de la variable de Poisson  $P_k$  et  $\sigma > 0$ . Puisque la fonction  $\psi$  prend des valeurs positives, l'éventualité d'une valeur fortement négative de  $G_k$  ne pose pas de problème pour la définition de  $\tilde{X}^h$ . Même avec la perturbation gaussienne, la suite  $(\tilde{X}_{t_k}^h)$  reste une chaîne de Markov et on peut donc écrire son opérateur associé évalué à une fonction quelconque  $f$  définie sur  $\mathbb{R}$  :

$$\begin{aligned} \mathcal{T}^h f(x) &:= \mathbb{E} \left[ f(\tilde{X}_{t_{k+1}}^h) | \tilde{X}_{t_k}^h = x \right], \\ &= \mathbb{E}^{\mathcal{N}} \left[ f \left( e^{-\beta h} x + \sigma \sqrt{h} G \right) + f \left( e^{-\beta h} x + \alpha + \sigma \sqrt{h} G \right) h \psi(\mu + x) \right] e^{-h \psi(\mu + x)} + R_h(x), \end{aligned}$$

où  $\mathbb{E}^{\mathcal{N}}$  est l'intégration contre la mesure gaussienne centrée réduite,  $G \sim \mathcal{N}(0, 1)$  et  $R_h$  est un reste.

Comme la taille  $h$  du pas de temps est petite, on peut écrire un développement de Taylor de  $\mathcal{T}^h$  en  $h$ , ce qui donne

$$\mathcal{T}^h f(x) = f(x) + h \mathcal{A}f(x) + O(h^2),$$

avec

$$\mathcal{A}f(x) := \frac{\sigma^2}{2} f''(x) - \beta x f'(x) + \psi(\mu + x) (f(x + \alpha) - f(x)),$$

la première partie du générateur  $\mathcal{A}_{OU}f(x) = \frac{\sigma^2}{2} f''(x) - \beta x f'(x)$  correspond à un processus d'Ornstein-Uhlenbeck et la deuxième partie  $\mathcal{A}_J f(x) = \psi(\mu + x) (f(x + \alpha) - f(x))$  à un processus de saut. En effet, ce générateur infinitésimal correspond à la dynamique markovienne

$$dX_t = -\beta X_t dt + \alpha dN_t + \sigma dW_t \quad (1.2.5)$$

avec  $N$  un processus ponctuel d'intensité  $\psi(\mu + X)$  et  $W$  est un mouvement brownien standard indépendant de  $N$ .

On peut donc dire que le processus  $\tilde{X}^h$  construit selon la dynamique (1.2.4) approche, en tant que chaîne de Markov, le processus  $X$  solution de l'EDS (1.2.5). Ceci découle de la convergence de  $h^{-1}(\mathcal{T}^h - I)$  vers  $\mathcal{A}$  selon les résultats de [36]. Cette convergence est formalisée dans le théorème suivant (le théorème 2.4.5 dans le chapitre 2) :



**Théorème 1.2.5.** Soit  $N$  le processus de Hawkes d'intensité  $\psi(\mu + X)$  où

$$X_t = \mu + \int_{[0,t)} e^{-\beta(t-s)} (\alpha dN_s + \sigma dW_s),$$

avec  $W$  un mouvement brownien indépendant de  $N$ . Soient  $\tilde{X}^h$  et  $P$  les suites définies récursivement selon (1.2.4). On pose  $X^h$  et  $N^h$  les processus càdlàg constants par morceaux définis sur  $\mathbb{R}_+$  de la façon suivante :

- $X^h$  coïncide avec  $\tilde{X}^h$  sur les points  $t_k$ , pour  $k \in \mathbb{N}$ .
- $N_t^h = \sum_{i=1}^{\lfloor t/h \rfloor} P_k$ .

Alors, si la fonction  $\psi$  est bornée on a

$$(X^h, N^h) \xrightarrow{h \rightarrow 0} (X, N)$$

faiblement dans la topologie de Skorokhod.

*Remarque 1.2.6.* L'hypothèse " $\psi$  est bornée" est purement technique dans la preuve. Il est très probable que l'on puisse étendre la convergence au cas où  $\|\psi\|_L \|\phi\|_1 < 1$ .

Ce résultat formalise l'approximation d'un processus de Hawkes par un schéma d'Euler dans le cas d'un noyau exponentiel. Le choix d'une simple exponentielle étant un peu restrictif, on a généralisé le résultat à la classe plus riche des noyaux d'Erlang formés par le produit d'une exponentielle et d'un polynôme c'est à dire

$$\phi(t) = e^{-\beta t} \sum_{q=0}^p \alpha_q \frac{t^q}{q!}.$$

Ceci est possible car les fonctions  $t \mapsto t^q e^{-\beta t}$  résolvent une équation différentielle en dimension  $q + 1$ . Donc modulo l'ajout de quelques processus auxiliaires, formant ainsi une cascade markovienne, l'intensité reste un processus markovien.

Plus généralement, le résultat de convergence reste vrai pour les noyaux de la forme

$$\phi(t) = \sum_{q=0}^p \alpha_q \frac{t^q}{q!} e^{-\beta_q t}, \quad (1.2.6)$$

quitte à augmenter la dimension du vecteur. D'un point de vue pratique, cette classe est assez riche pour approcher tout noyaux dans l'espace  $L_1$  (cf. Kammler [53]), ce qui signifie qu'elle est capable d'approcher tout processus de Hawkes. En effet, étant donné un noyau  $\phi$  dans  $L_1(\mathbb{R}_+)$ , il existe une suite de noyaux  $(\phi_p)_{p \in \mathbb{N}}$  de la forme (1.2.6) telle que  $\lim_{p \rightarrow +\infty} \|\phi_p - \phi\|_1 = 0$  et une suite de processus de Hawkes  $(N^p)_{p \in \mathbb{N}}$  de noyaux  $(\phi_p)_{p \in \mathbb{N}}$  avec

$$\mathbb{E} [d_{TV}(N^p, N)_t] \leq C_T \int_0^t |\phi_p(s) - \phi(s)| ds$$

pour tout  $t \in [0, T]$ , et où la distance  $d_{TV}(\cdot)_t$  correspond à la variation totale entre les deux mesures jusqu'à  $t$ , cf. [78].

Malgré la nouveauté de l'approximation des processus de Hawkes par un schéma d'Euler, cette technique est classique pour les équations différentielles stochastiques, éventuellement avec des sauts. On renvoie à l'article de Frikha et Li [39] pour un résultat d'approximation des EDS dont l'aléa est un processus de Lévy.

1.2.2.2 Calibration de la série temporelle approximante

Comme application, on propose l'approximation des paramètres d'un processus de Hawkes en temps continu via la calibration de sa série temporelle approximante : On observe un processus  $N$  sur l'intervalle  $[0, T]$ , que l'on divise en  $n$  sous-intervalles de taille  $h = T/n$  et on définit

$$P_k = N_{kh} - N_{(k-1)h}, \quad k = 1, \dots, n$$

comme suite de comptage. On suppose que le processus de Hawkes sous-jacent a l'intensité  $\psi(\mu + X)$  où  $X_t = \mu + \int_{[0,t)} \alpha e^{-\beta(t-s)} dN_s$ . Les paramètres  $\alpha$ ,  $\beta$  et  $\mu$  sont inconnus et la fonction  $\psi$  est connue. La méthode utilisée dans le chapitre 2 est la minimisation d'un estimateur de moindres carrés conditionnels

$$Q_n = \sum_{k=1}^n (P_k - \mathbb{E}[P_k | \mathcal{F}_{(k-1)h}])^2,$$

inspirée du travail de Klimko et Nelson [57]. L'espérance conditionnelle du processus de comptage n'est rien d'autre que l'intensité  $\psi(\mu + \tilde{X}^h)$  où  $\tilde{X}^h$  est construite récursivement

$$\tilde{X}_{t_{k+1}}^h = e^{-\beta h} \tilde{X}_{t_k}^h + \alpha P_k,$$

avec la condition initiale  $\tilde{X}_{t_0}^h = 0$ . Si la fonction  $\psi$  est dérivable, on peut construire les dérivées partielles  $\partial_\alpha Q_n$ ,  $\partial_\beta Q_n$  et  $\partial_\mu Q_n$  de récursivement et l'utiliser afin de minimiser  $Q_n$ .

Les résultats de la calibration pour le processus de Hawkes tri-dimensionnel avec des noyaux d'Erlang sont illustrés sur les figures du chapitre 2.

1.3 Approximation gaussienne du processus de Hawkes normalisé

1.3.1 État de l'art

On se place dorénavant dans le cadre d'un processus de Hawkes linéaire ( $\psi(x) = x$  et  $\phi \geq 0$ ) en temps continu. Si la condition de stabilité (c'est à dire  $\|\phi\|_1 < 1$ ) est satisfaite, on sait que, moralement

$$”\lambda_t \simeq \frac{\mu}{1 - \|\phi\|_1}”$$

et avec cette approximation en tête, Bacry *et al.* [7] ont montré la loi des grands nombres

$$\frac{N_T}{T} \xrightarrow{T \rightarrow +\infty} \frac{\mu}{1 - \|\phi\|_1}$$

presque sûrement. Quant au théorème central limite (TCL), il est prouvé grâce aux résultats prouvés sur les martingales (*cf.* [51]). En effet la variation quadratique de

la martingale  $M_T = N_T - \int_0^T \lambda_s ds$  n'est rien d'autre que la cumulation des sautes, c'est-à-dire  $N$

$$[M]_T = N_T$$

et donc si l'on normalise la martingale  $F_T = M_T/\sqrt{T}$  on a

$$[F]_T = \frac{N_T}{T} \xrightarrow{T \rightarrow +\infty} \sigma^2$$

où

$$\sigma^2 = \frac{\mu}{1 - \|\phi\|_1}.$$

*Remarque 1.3.1.* Concernant les processus de Hawkes non-linéaires, on n'a toujours pas de formules explicites pour les théorèmes limites. Zhu [92] est parvenu à démontrer une loi des grands nombres et un théorème central limite qui font intervenir la mesure stationnaire (et donc pas explicites) pour les processus non-linéaires.

Une fois le TCL est démontré, l'étape suivante est d'estimer la distance entre la martingale normalisée  $F$  et sa limite (en loi) gaussienne  $\mathcal{N}(0, \sigma^2)$ .

Pour cela, on introduit la distance de Wasserstein mesure l'écart entre deux lois de probabilité représentées par deux variables aléatoires  $X$  et  $Y$

$$d_W(X, Y) := \sup_{f \in Lip_1} |\mathbb{E}[f(X)] - \mathbb{E}[f(Y)]|, \quad (1.3.1)$$

où  $Lip_1$  est l'espace des fonctions dérivables lipschitziennes de constante Lipschitz inférieure ou égale à 1.

*Remarque 1.3.2.* La définition (1.3.1) est en réalité la représentation duale (due à Kantorovich) de la définition originale

$$d_W(\nu, \nu') = \inf_{\gamma \in \Gamma(\nu, \nu')} \int_{x, y} |x - y| d\gamma(\nu, \nu'),$$

où  $\Gamma(\nu, \nu')$  est l'ensemble des couplages des probabilités  $\nu$  et  $\nu'$ .

Dans son article [91], Torrisi donne une borne sur la distance de Wasserstein entre la martingale issue du Hawkes (éventuellement non-linéaire) normalisée  $F_T$  et une variable gaussienne quelconque

$$d_W(F_T, \mathcal{N}(0, \gamma^2)) \leq C.$$

Torrisi utilise des méthodes similaires aux nôtres (la méthode de Malliavin-Stein qui sera explicitée dans la suite) mais sa borne n'est pas assez précise et ne permet pas de retrouver le théorème central limite.

## 1.3.2 Contributions de la thèse

### 1.3.2.1 Lemme de Stein et opérateur divergence

Notre objectif est de donner une quantification de la distance entre la martingale du Hawkes et sa limite gaussienne qui permet de retrouver la convergence du TCL et en donne la vitesse.

On commence par énoncer le résultat suivant : si l'une des variables (disons  $Y$ ) est gaussienne centrée de variance  $\sigma^2$ , la distance de Wasserstein est bornée par une quantité ne dépendant que de  $X$  :

$$d_W(X, Y) \leq \sup_{f \in Lip_2} |\mathbb{E} [Xf(X) - \sigma^2 f'(X)]|, \quad (1.3.2)$$

où

$$Lip_2 = \{f \in \mathcal{C}^2(\mathbb{R}), \|f'\|_\infty \leq 1 \text{ et } \|f''\|_\infty \leq 2\}.$$

Cette borne se base sur ce lemme prouvé par Stein [87]

**Lemme 1.3.3.** *Une variable aléatoire  $Y$  suit une loi gaussienne centrée de variance  $\sigma^2$  si et seulement si, pour toute fonction  $f$  différentiable par morceaux avec  $\mathbb{E} [|f'(\mathcal{N}(0, \sigma^2))|] < +\infty$  on a*

$$\mathbb{E} [Yf(Y) - \sigma^2 f'(Y)] = 0.$$

*Remarque 1.3.4.* Un sens de cette équivalence est trivial, il s'agit d'une application du résultat  $p'(x) = p(x)/\sigma^2$  où  $p(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right) / \sqrt{2\pi\sigma^2}$  est la densité gaussienne de variance  $\sigma^2$ .

En effet, ce lemme signifie heuristiquement qu'une variable  $X$  est proche de la distribution gaussienne si l'on a dans un certain sens

$$\mathbb{E} [Xf(X) - \sigma^2 f'(X)] \simeq 0,$$

pour une classe assez large de fonctions  $f$ . L'application de ces résultats à la martingale normalisée  $F_T = M_T/\sqrt{T}$  associée au processus de Hawkes s'écrit

$$d_W(F_T, \mathcal{N}(0, \sigma^2)) \leq \sup_{f \in Lip_2} |\mathbb{E} [F_T f(F_T) - \sigma^2 f'(F_T)]|.$$

Bien qu'elle soit plus simple que la borne originale (1.3.1), cette borne reste difficile à quantifier car elle fait intervenir le supremum sur une famille de fonctions. Il faudra donc faire de sorte que  $f$  n'intervienne que par ses dérivées première et seconde, qui sont bornées par des constantes. Pour ce faire, on a besoin d'une formule d'intégration par parties afin de décomposer l'espérance  $\mathbb{E}[F_T f(F_T)]$  et de faire intervenir les dérivées de  $f$ .

Tout d'abord, on remarque que la martingale normalisée  $F$  est l'intégrale d'une entité déterministe contre le processus ponctuel compensé (la définition la plus générale est l'intégrale contre l'aléa sous-jacent  $\Pi$  compensé cf. (1.1.5))

$$\begin{aligned} F_T &= \frac{M_T}{\sqrt{T}} = \frac{1}{\sqrt{T}} \left( N_T - \int_0^T \lambda_s ds \right) \\ &= \int_{\mathbb{R}_+} \frac{\mathbb{1}_{\{s \leq T\}}}{\sqrt{T}} (dN_s - \lambda_s ds) \\ &:= \delta^N \left( \left( \frac{\mathbb{1}_{\{s \leq T\}}}{\sqrt{T}} \right)_{s \in \mathbb{R}_+} \right). \end{aligned}$$

L'opérateur  $\delta^N(\cdot) = \int_{\mathbb{R}_+} (\cdot)(dN_s - \lambda_s ds)$  est appelé *divergence* par rapport au processus de Hawkes.

### 1.3.2.2 Dérivation heuristique de l'adjoint de la divergence

Dans les paragraphes qui suivent, on sacrifiera un peu de rigueur mathématique afin d'établir -heuristiquement- une relation d'intégration par parties pour les processus de Hawkes, dans le but de majorer la distance de Wasserstein entre la martingale normalisée et sa limite gaussienne. On se place dans le cadre du processus ponctuel en temps discret avec des sauts de Bernoulli. L'aléa sous-jacent est une famille *i.i.d* de variables uniformes  $(U_i)_{i \in \mathbb{N}}$  sur  $[0, 1]$ . On appelle la filtration engendrée par ces variables

$$\mathcal{F}_k^U = \sigma(U_1, \dots, U_k).$$

Le processus ponctuel  $\tilde{N}^h$  est défini sur une grille de temps  $(t_i)_{i \in \mathbb{N}}$  de taille  $h$

$$\begin{aligned} \tilde{N}_{t_k}^h &= \sum_{i=1}^k \Delta \tilde{N}_{t_i}^h \\ &= \sum_{i=1}^k \mathbb{1}_{U_i \leq \tilde{\lambda}_{t_i}^h h} \end{aligned}$$

et il définit une filtration  $\mathcal{F}_k^{\tilde{N}^h} = \sigma(\Delta \tilde{N}_{t_1}^h, \dots, \Delta \tilde{N}_{t_k}^h)$ . L'intensité  $\tilde{\lambda}^h$  est un processus prévisible par rapport à  $\tilde{N}^h$ , c'est-à-dire

$$\tilde{\lambda}_{t_{k+1}}^h \in \mathcal{F}_k^{\tilde{N}^h}.$$

On remarque que l'information sur les  $U$  contient l'information sur  $\tilde{N}^h$ , autrement dit

$$\mathcal{F}^{\tilde{N}^h} \subset \mathcal{F}^U.$$

L'analogie entre le processus en temps discret et le processus en temps continu se trouve dans le tableau suivant

| Temps continu  | Temps discret  |
|--|--|
| $dN_s = \int_{\mathbb{R}_+} \mathbb{1}_{\{\theta \leq \lambda_s\}} \Pi(ds, d\theta)$ | $\Delta \tilde{N}_{t_i}^h = \mathbb{1}_{U_i \leq \tilde{\lambda}_{t_i}^h h}$                                   |
| $\lambda_s$  | $\tilde{\lambda}_{t_i}^h$  |
| $\int_0^{+\infty} f(s) dN_s$   | $\sum_{i=0}^{+\infty} f(t_i) \Delta \tilde{N}_{t_i}^h$   |
| $\int_0^{+\infty} f(s) ds$   | $\sum_{i=0}^{+\infty} f(t_i) h$  |
| $\delta^N(f) = \int_{\mathbb{R}_+} f(s) (dN_s - \lambda_s ds)$                       | $\delta^{\tilde{N}^h}(f) = \sum_{i=0}^{+\infty} f(t_i) (\Delta \tilde{N}_{t_i}^h - \tilde{\lambda}_{t_i}^h h)$ |

Soit  $(Z_{t_i})_{i \in \mathbb{N}}$  un processus déterministe, nul à partir d'un certain rang (afin de supposer l'existence des intégrales infinies) et  $V = g\left((\mathbb{1}_{U_i \leq \tilde{\lambda}_{t_i}^h h})_{i \in \mathbb{N}}\right)$  une fonctionnelle générale de  $\tilde{N}^h$  (de manière équivalente une variable mesurable par rapport à

$\mathcal{F}_\infty^{\tilde{N}^h} = \sigma(\mathbb{1}_{U_k \leq \tilde{\lambda}_{t_k}^h}, k \in \mathbb{N})$ ). On cherche à évaluer l'espérance du produit entre  $V$  et l'intégrale de  $Z$  contre  $\Delta \tilde{N}^h$

$$\begin{aligned} \mathbb{E} \left[ V \sum_{k=0}^{+\infty} Z_{t_k} \mathbb{1}_{U_k \leq \tilde{\lambda}_{t_k}^h} \right] &= \sum_{k=0}^{+\infty} Z_{t_k} \mathbb{E} \left[ V \mathbb{1}_{U_k \leq \tilde{\lambda}_{t_k}^h} \right] \quad \text{car } Z \text{ est déterministe} \\ &= \sum_{k=0}^{+\infty} Z_{t_k} \mathbb{E} \left[ g \left( (\mathbb{1}_{U_i \leq \tilde{\lambda}_{t_i}^h})_{i \in \mathbb{N}} \right) \mathbb{1}_{U_k \leq \tilde{\lambda}_{t_k}^h} \right] \\ &= \sum_{k=0}^{+\infty} Z_{t_k} \mathbb{E} \left[ \mathbb{E} \left[ g \left( (\mathbb{1}_{U_i \leq \tilde{\lambda}_{t_i}^h})_{i \in \mathbb{N}} \right) \mathbb{1}_{U_k \leq \tilde{\lambda}_{t_k}^h} \middle| U_i, i \neq k \right] \right]. \end{aligned}$$

Sachant toutes les indicatrices  $\mathbb{1}_{U_i \leq \tilde{\lambda}_{t_i}^h}$  à l'exception de  $k$ ,  $\tilde{\lambda}_{t_k}^h$  est connue, étant mesurable par rapport à  $\mathcal{F}_{k-1}^U$ . Ceci signifie que l'on peut traiter  $\mathbb{1}_{U_k \leq \tilde{\lambda}_{t_k}^h}$  comme une variable de Bernoulli de paramètre  $\tilde{\lambda}_{t_k}^h$  connu, et donc

$$\begin{aligned} &\mathbb{E} \left[ g \left( (\mathbb{1}_{U_i \leq \tilde{\lambda}_{t_i}^h})_{i \in \mathbb{N}} \right) \mathbb{1}_{U_k \leq \tilde{\lambda}_{t_k}^h} \middle| U_i, i \neq k \right] \\ &= \mathbb{E} \left[ g \left( \mathbb{1}_{U_1 \leq \tilde{\lambda}_{t_1}^h}, \dots, \mathbb{1}_{U_k \leq \tilde{\lambda}_{t_k}^h}, \mathbb{1}_{U_{k+1} \leq \tilde{\lambda}_{t_{k+1}}^h}, \dots \right) \mathbb{1}_{U_k \leq \tilde{\lambda}_{t_k}^h} \middle| U_i, i \neq k \right] \\ &= \int_0^1 g \left( \mathbb{1}_{U_1 \leq \tilde{\lambda}_{t_1}^h}, \dots, \mathbb{1}_{u \leq \tilde{\lambda}_{t_k}^h}, \mathbb{1}_{U_{k+1} \leq \tilde{\lambda}_{t_{k+1}}^h} (\mathbb{1}_{u \leq \tilde{\lambda}_{t_k}^h})^h, \dots \right) \mathbb{1}_{u \leq \tilde{\lambda}_{t_k}^h} du. \end{aligned}$$

On insiste sur la dépendance des  $\tilde{\lambda}_{t_i}^h, i \geq k+1$  de la valeur de  $U_k$ . En effet, cette dépendance est binaire et ne dépend que de la position relative  $U_k \leq \tilde{\lambda}_{t_k}^h$  ou  $U_k > \tilde{\lambda}_{t_k}^h$  (il s'agit de l'équivalent discret du lemme 3.2.14 au chapitre 3) et par conséquent

$$\begin{aligned} &\mathbb{E} \left[ g \left( (\mathbb{1}_{U_i \leq \tilde{\lambda}_{t_i}^h})_{i \in \mathbb{N}} \right) \mathbb{1}_{U_k \leq \tilde{\lambda}_{t_k}^h} \middle| U_i, i \neq k \right] \\ &= g \left( \mathbb{1}_{U_1 \leq \tilde{\lambda}_{t_1}^h}, \dots, 1, \mathbb{1}_{U_{k+1} \leq \tilde{\lambda}_{t_{k+1}}^h \circ \varepsilon_{t_k}^+}, \dots \right) \times 1 \times \tilde{\lambda}_{t_k}^h \\ &\quad + g \left( \mathbb{1}_{U_1 \leq \tilde{\lambda}_{t_1}^h}, \dots, 0, \mathbb{1}_{U_{k+1} \leq \tilde{\lambda}_{t_{k+1}}^h \circ \varepsilon_{t_k}^-}, \dots \right) \times 0 \times (1 - \tilde{\lambda}_{t_k}^h) \\ &= g \left( \mathbb{1}_{U_1 \leq \tilde{\lambda}_{t_1}^h}, \dots, 1, \mathbb{1}_{U_{k+1} \leq \tilde{\lambda}_{t_{k+1}}^h \circ \varepsilon_{t_k}^+}, \dots \right) \tilde{\lambda}_{t_k}^h \\ &= (V \circ \varepsilon_{t_k}^+) \tilde{\lambda}_{t_k}^h. \end{aligned}$$

Où  $\varepsilon_{t_k}^+$  est l'opérateur d'ajout de point à  $t_k$  : on impose que  $\mathbb{1}_{U_k \leq \tilde{\lambda}_{t_k}^h}$  soit égale à 1, autrement dit, on impose que  $\tilde{N}^h$  saute à l'instant  $k$  tandis que  $\varepsilon_{t_k}^-$  impose qu'il ne saute pas. Il est important de noter que ces opérateurs n'ont aucune action sur l'aléa sous-jacent en dehors du saut ajouté (ou retranché) à  $\tilde{N}^h$ . Donc

$$\mathbb{E} \left[ V \sum_{k=0}^{+\infty} Z_{t_k} \Delta \tilde{N}_{t_k}^h \right] = \mathbb{E} \left[ \sum_{k=0}^{+\infty} Z_{t_k} (V \circ \varepsilon_{t_k}^+) \tilde{\lambda}_{t_k}^h \right]. \quad (1.3.3)$$

*Remarque 1.3.5.* Jusqu'alors, on a négligé l'événement  $\tilde{\lambda}_{t_k}^h h > 1$ , dont la probabilité n'est pas nulle pour un processus de Hawkes linéaire. Cependant, comme l'on prend  $h$  petit, cette probabilité est négligeable. On rappelle que l'objectif de cette section est d'établir un "calcul de Malliavin" intuitif pour les processus ponctuels construits par amincissement et en particulier les processus de Hawkes.

L'écart entre une fonctionnelle  $V$  de la trajectoire du processus ponctuel  $(\tilde{N}_{t_i}^h)_{i \in \mathbb{N}}$  forcée à sauter à  $t_k$  (c'est à dire composée avec  $\varepsilon_{t_k}^+$ ) et la même fonctionnelle avec une trajectoire forcée à ne pas sauter à  $t_k$  (composée avec  $\varepsilon_{t_k}^-$ ) définit la *dérivée de Malliavin* de  $V$  en  $t_k$

$$D_{t_k} V := V \circ \varepsilon_{t_k}^+ - V \circ \varepsilon_{t_k}^-.$$

Comme le pas de temps  $h$  est petit, et sachant que  $\tilde{\lambda}_{t_i}^h$  reste d'ordre  $O(1)$  (avec probabilité proche de l'unité) et donc, moralement, la probabilité que le processus  $\tilde{N}_{t_k}^h$  saute à un instant  $t_k$  fixé est d'ordre  $O(h)$  et que avec probabilité d'ordre  $1 - O(h)$  on a  $V \circ \varepsilon_{t_k}^- = V$ . Ceci signifie que, grossièrement

$$D_{t_k} V = "V \circ \varepsilon_{t_k}^+ - V + O(h)."$$

On montre maintenant que la dérivée de Malliavin correspond à l'opérateur adjoint de la divergence et on donne la formule de dualité ou d'intégration par parties. L'espérance du produit d'une divergence et une fonctionnelle  $V$  de la trajectoire du processus ponctuel s'écrit

$$\begin{aligned} \mathbb{E} \left[ V \delta^{\tilde{N}^h} (Z) \right] &= \mathbb{E} \left[ V \sum_{k=0}^{+\infty} Z_{t_k} (\Delta \tilde{N}_{t_k}^h - \tilde{\lambda}_{t_k}^h h) \right] \\ &= \mathbb{E} \left[ V \sum_{k=0}^{+\infty} Z_{t_k} \Delta \tilde{N}_{t_k}^h \right] - \mathbb{E} \left[ V \sum_{k=0}^{+\infty} Z_{t_k} \tilde{\lambda}_{t_k}^h h \right], \end{aligned}$$

et en utilisant (1.3.3) on a

$$\begin{aligned} \mathbb{E} \left[ V \delta^{\tilde{N}^h} (Z) \right] &= \mathbb{E} \left[ \sum_{k=0}^{+\infty} Z_{t_k} (V \circ \varepsilon_{t_k}^+) \tilde{\lambda}_{t_k}^h h \right] - \mathbb{E} \left[ V \sum_{k=0}^{+\infty} Z_{t_k} \tilde{\lambda}_{t_k}^h h \right] \\ &= \mathbb{E} \left[ \sum_{k=0}^{+\infty} Z_{t_k} (V \circ \varepsilon_{t_k}^+ - V) \tilde{\lambda}_{t_k}^h h \right] \\ &= \mathbb{E} \left[ \sum_{k=0}^{+\infty} Z_{t_k} D_{t_k} V \tilde{\lambda}_{t_k}^h h \right] + O(h), \end{aligned}$$

car on a supposé que  $Z$  était à support compact et que  $\tilde{\lambda}_{t_k}^h = O(1)$  dans un certain sens. Cette égalité nous donne alors l'adjoint de la divergence ; il s'agit de l'intégrale de la dérivée de Malliavin. Pour des résultats montrés plus rigoureusement sur le calcul de Malliavin discret, on réfère à la thèse d'Halconruy [45] et au livre de Privault [77].

*Remarque 1.3.6.* Il est possible d'étendre la formule d'intégration par parties aux processus  $(Z_{t_k})_{k \in \mathbb{N}}$  prévisibles et pas nécessairement déterministes. On se contente des processus déterministes car la martingale qui nous intéresse est la divergence d'un processus déterministe.

Comme on l'a démontré, le processus de Hawkes discret  $\tilde{N}^h$  converge vers le processus de Hawkes  $N$  quand le pas de temps  $h$  tend vers zéro. On peut donc, toujours de manière heuristique, étendre l'intégration par parties au processus en temps continu. Pour une fonctionnelle  $V$  du processus en temps continu  $V \in \mathcal{F}_\infty^N$  et un processus déterministe  $(Z_t)_{t \in \mathbb{R}_+}$  on a

$$\mathbb{E} [V \delta^N(Z)] = \mathbb{E} \left[ \int_{\mathbb{R}_+} Z_t D_t V \lambda_t dt \right]. \quad (1.3.4)$$

L'opérateur  $D_t$  ici correspond à la différence

$$\begin{aligned} D_t V &= V \circ \varepsilon_t^+ - V \circ \varepsilon_t^- \\ &= V \circ \varepsilon_t^+ - V \quad \text{presque sûrement} \end{aligned}$$

où l'opération  $\varepsilon_t^+$  correspond à l'ajout d'un saut au processus de Hawkes à l'instant  $t$ , et  $\varepsilon_t^-$  correspond à l'enlèvement du saut éventuel à l'instant  $t$ , qui n'a pas d'effet réel sur la fonctionnelle, car à un instant  $t$  fixé au préalable, avec probabilité 1 le processus  $N$  ne saute pas. On note que le saut ajouté au processus de Hawkes, est en réalité ajouté à l'aléa sous-jacent  $\Pi(t, \theta)$ , à une altitude  $\theta$  qui n'intervient que par sa position relative par rapport à  $\lambda_t$ . L'opération est donc équivalente à ajouter un point  $(t, 0)$  ou bien  $(t, \lambda_t)$  à l'aléa  $\Pi$  comme on l'illustre sur la figure 1.2.

### 1.3.2.3 La méthode de Malliavin-Stein appliquée au processus de Hawkes

La combinaison de la formule d'intégration par parties (3.2.6) et la borne (1.3.2) sur la distance de Wasserstein entre la divergence d'un certain aléa et une variable aléatoire gaussienne est appelée la méthode de *Malliavin-Stein*, dont les pionniers sont Nourdin et Peccati [65].

On rappelle que notre objectif est d'éliminer toute présence de  $f$  dans la borne de la distance de Wasserstein, tout en autorisant la présence éventuelle de  $f'$  et  $f''$ , étant majorées par des constantes. Dans le cas de la martingale issue du processus de Hawkes c'est le terme  $\mathbb{E} [F_T f(F_T)]$  qui nous gêne, on le déconstruit à l'aide de la relation (1.3.4) (*cf.* relation (3.2.6) au chapitre 3)

$$\begin{aligned} \mathbb{E} [F_T f(F_T)] &= \mathbb{E} \left[ \delta^N \left( \left( \frac{\mathbb{1}_{\{s \leq T\}}}{\sqrt{T}} \right)_{s \in \mathbb{R}_+} \right) f(F_T) \right] \\ &= \mathbb{E} \left[ \int_{\mathbb{R}_+} \frac{\mathbb{1}_{\{t \leq T\}}}{\sqrt{T}} D_t f(F_T) \lambda_t dt \right] \\ &= \frac{1}{\sqrt{T}} \mathbb{E} \left[ \int_0^T D_t f(F_T) \lambda_t dt \right] \end{aligned}$$



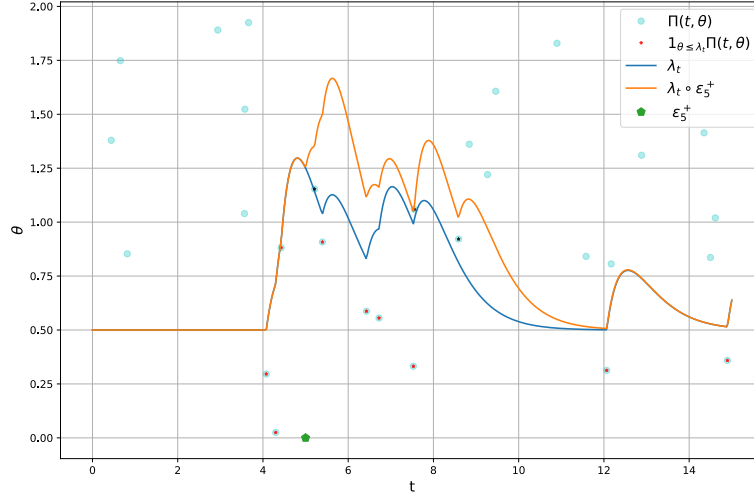


FIGURE 1.2 – Équivalence entre l'ajout d'un point au processus de Hawkes  $N$  à l'instant  $t = 5$  et l'ajout du point  $(5, 0)$  à l'aléa sous-jacent  $\Pi$ . Les points conservés par  $\lambda_t \circ \varepsilon_5^+$  et pas par  $\lambda_t$  sont marqués en noir. Les points marqués en rouge sont les points conservés par les deux processus ponctuels. Le noyau utilisé est  $\phi(t) = 1.5te^{-2t}$ .

On tient à insister sur le fait que l'aléa qui conduit  $\lambda$  et  $\lambda \circ \varepsilon_5^+$  reste strictement le même.

Puisque la dérivée de Malliavin correspond à la différence entre le processus et sa version avec un saut ajouté, elle peut être vue comme une dérivée discrète et on peut appliquer la formule de Taylor en  $F_T$

$$\begin{aligned} D_t f(F_T) &= f(F_T) \circ \varepsilon_t^+ - f(F_T) \\ &= f(F_T \circ \varepsilon_t^+) - f(F_T) \quad \text{car le saut est ajouté au processus} \\ &= f'(F_T) D_t F_T + \frac{1}{2} f''(\bar{F}) |D_t F_T|^2 \end{aligned}$$

pour un certain  $\bar{F}$  aléatoire. On a donc tous les ingrédients pour prouver le résultat suivant (théorème 3.3.10 dans le chapitre 3)

**Théorème 1.3.7.** *On suppose que le noyau  $\phi$  satisfait la condition de stabilité  $\|\phi\|_1 < 1$  et que, en outre,  $\int_0^{+\infty} u\phi(u)du < +\infty$ . On rappelle que  $\sigma^2 = \mu/(1 - \|\phi\|_1)$  et que  $F_T = (N_T - \int_0^T \lambda_t dt)/\sqrt{T}$ . Il existe alors, une constante  $C$  indépendante de  $T$  telle que*

$$d_W(F_T, \mathcal{N}(0, \sigma^2)) \leq \frac{C}{\sqrt{T}} + \frac{1}{T} \mathbb{E} \left[ \left| \int_0^T \lambda_t (D_t M_T - 1) dt \right| \right], \quad (1.3.5)$$

où  $M_T = N_T - \int_0^T \lambda_t dt$  est la martingale issue du processus de Hawkes.

Le deuxième terme dans la borne n'a généralement pas de majorant direct. Apparemment on est allé un peu trop vite dans les majorations précédentes. On peut, cependant, expliciter sa dynamique si le noyau  $\phi$  est exponentiel ou bien Erlang<sup>2</sup>. Ceci est une conséquence du caractère markovien de l'intensité pour ces types de noyaux. Dans ce cas là, on a la majoration plus explicite

$$d_W(F_T, \mathcal{N}(0, \sigma^2)) \leq \frac{C}{\sqrt{T}}.$$

*Remarque 1.3.8.* Quoique la borne en  $1/\sqrt{T}$  ne soit démontrée que pour les processus de Hawkes de noyaux exponentiels ou Erlang, on a un espoir raisonnable de pouvoir l'étendre à une classe beaucoup plus riche de noyaux. C'est le cas car tout noyau intégrable  $\phi$  peut être arbitrairement approché par un noyau  $\phi_p$  qui satisfait une dynamique markovienne.

*Remarque 1.3.9.* Bien qu'il soit utilisé en combinaison avec la méthode de Stein dans ce manuscrit, le calcul de Malliavin a des applications différentes de l'approximation gaussienne. On renvoie à [37] pour des applications dans la finance et à [38] pour le calcul de sensibilités d'un processus aléatoire arrêté.

D'un point de vue pratique, il est potentiellement plus intéressant d'étudier le processus de Hawkes "compensé" par une entité déterministe, idéalement simple à calculer, que de majorer la martingale, où le compensateur est une intégrale stochastique. On pose alors

$$Y_T = \frac{N_T - \sigma^2 T}{\sqrt{T}}$$

avec  $\sigma^2 = \frac{\mu}{1 - \|\phi\|_1}$  qui est à la fois l'espérance asymptotique de  $\lambda$ , la limite presque sûre de  $N_T/T$  et la variance de la gaussienne dans le théorème central limite. On arrive à prouver dans le chapitre 3 que si le noyau est exponentiel ou Erlang

$$d_W(Y_T, \mathcal{N}(0, \tilde{\sigma}^2)) \leq \frac{C}{\sqrt{T}}.$$

La nouvelle variance  $\tilde{\sigma}^2 = \frac{\mu}{(1 - \|\phi\|_1)^3}$  est plus grande que  $\sigma^2$  ce qui est attendu. En fait, on s'attend à ce que  $Y_T$  fluctue plus que la martingale compensée  $F_T$ . Cette majoration montre, grossièrement, que si l'on remplace  $N_T$  par  $\sigma^2 T + \mathcal{N}(0, \tilde{\sigma}^2 \sqrt{T})$ , alors l'erreur commise est d'ordre  $O(1)$ .

Au chapitre 4 on généralise ce résultat aux processus de Hawkes composés linéaires multi-dimensionnels et sur un nombre fini d'instant (appelées multi-marginales) sur un intervalle  $[0, T]$ . On obtient que la convergence vers une limite gaussienne (avec une matrice de covariance explicite) se fait toujours avec la vitesse  $O(1/\sqrt{T})$ .

Par contre, on n'établit pas une borne entre la distance entre  $M$  (ou  $Y$ ) vue comme un processus sur le long de l'intervalle  $[0, T]$  et la limite brownienne. On renvoie le lecteur à [12] pour les dernières avancées dans cette direction.

---

2. Ce qu'on appelle noyau d'Erlang dans cette section n'est qu'un cas particulier des noyaux d'Erlang définis dans la section précédente. Il s'agit ici des fonctions de la forme  $\phi(t) = \alpha t e^{-\beta t}$ .

## 1.4 Le processus de Hawkes marqué

Au chapitre 5, on propose un cadre plus applicatif en considérant le processus de Hawkes *marqué* comme un modèle d'une police d'assurance santé. On suppose qu'un assureur reçoit des sinistres  $(X_i)_{i \in \mathbb{N}^*}$  qui sont associées à des charges médicales de la part d'un client à des instants  $(\tau_i)_{i \in \mathbb{N}^*}$ . La charge totale de l'assureur jusqu'à l'instant  $T$  est

$$L_T = \sum_{i=1}^{N_T} g(X_i),$$

où  $N$  est un processus ponctuel. La fonction  $g$  prend en compte un plafond éventuel sur les remboursements, ou bien un contrat de réassurance qui prend en charge les remboursements à partir d'un certain montant.

Le processus d'arrivée des sinistres  $N$  suit une dynamique de Hawkes linéaire influencée par la taille des sinistres

$$\lambda_t = \mu + \sum_{\tau_i < t} \phi(t - \tau_i) b(X_i).$$

La fonction  $b$  a le rôle naturel suivant : Les grands sinistres correspondent à des actes médicaux lourds tels qu'une intervention chirurgicale par exemple. Ces interventions sont très souvent suivies par d'autres prestations médicales comme les séances de kinésithérapie dans le cas d'une intervention chirurgicale. On peut par exemple penser à une fonction de la forme  $b(x) = \mathbb{1}_{x \geq Q}$  où  $Q$  est un seuil de "gravité" en dessous duquel le sinistre ne sera pas la source d'autres sinistres qui seraient dans la continuité du premier.

Nous suivons dans le chapitre 5 la méthodologie introduite en [86] et reprise en [54] pour donner la probabilité de ruine asymptotique, dont la quantification est un problème crucial dans la théorie du risque. L'originalité de notre travail réside principalement dans le modèle présenté (contrairement à [54] où les marques sont indépendantes des sinistres) et dans les démonstrations plus détaillées de certains résultats énoncés.

Les résultats théoriques seront illustrés par un exemple numérique, qui sera comparé au modèle classique d'un processus de Poisson composé (approximation de Cramér-Lundberg).

# The nonlinear discrete time Hawkes process

---

## Abstract

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

## Contents

---

|            |  |           |
|------------|--|-----------|
| <b>2.1</b> | <b>Introduction</b>  | <b>28</b> |
| <b>2.2</b> | <b>The discrete-time model</b>   | <b>30</b> |
| 2.2.1      | Preliminaries  | 30        |
| 2.2.2      | Construction   | 30        |
| 2.2.3      | Stability  | 34        |
| <b>2.3</b> | <b>The continuous-time Hawkes process</b>  | <b>36</b> |
| 2.3.1      | Preliminaries  | 36        |
| 2.3.2      | Definition and construction  | 36        |
| 2.3.3      | Markov properties of the Hawkes process with memory kernels given by an Erlang function      | 38        |
| <b>2.4</b> | <b>Convergence of the discrete-time Hawkes process to the continuous-time Hawkes process</b> | <b>41</b> |
| 2.4.1      | Preliminaries  | 41        |
| 2.4.2      | Main result  | 43        |
| <b>2.5</b> | <b>Parameters calibration using regression</b>   | <b>46</b> |
| 2.5.1      | Calibration experiment for the perturbed Hawkes process                                      | 48        |
| 2.5.2      | Calibration with a lower degree  | 49        |
| <b>2.6</b> | <b>Lemmata</b>   | <b>49</b> |

---

## 2.1 Introduction

In this chapter we introduce a multivariate time series  $(\mathbf{P}_n)_{n \in \mathbb{N}}$  in  $\mathbb{N}^d$  that represents the bin-count sequence of some events on regular time intervals of size  $h > 0$ . Examples of integer valued time series have been studied in the past, such as the  $g$ -functions for  $\{0, 1\}$  valued sequences by Berbee [10] and the integer valued autoregressive of order  $p$  (INAR( $p$ )) process by Alzaid and Al-Osh [2].

The specificity of our time series resides in its hidden Markov chain structure, in which the intensity depends on the history of the process. In fact, for any  $n \in \mathbb{N}$  and any  $i = 1, \dots, d$ , the bin-count variable  $P_n^i$  is assumed to be the realisation of a Poisson distribution of parameter  $h\psi_i(\mathbf{U}_n)$ , where  $(\mathbf{U}_n)_{n \in \mathbb{N}}$  is a predictable tensor Markov chain, *i.e.*  $\mathbf{U}_n \in \sigma(\mathbf{P}_0, \dots, \mathbf{P}_{n-1}, \mathbf{W}_0, \dots, \mathbf{W}_{n-1})$ , where  $(\mathbf{W}_n)_{n \in \mathbb{N}}$  are some exogenous random perturbations and  $(\psi_i)_{i=1, \dots, d}$  are non-negative functions. The choice of the Poisson distribution for the integer valued random variable is natural: we assume that the time bins are reasonably small, so that one would observe at most one event most of the time, while still allowing for the rare occurrence of two or more events. This, for instance, is not possible with the simpler choice of a Bernoulli distribution (as in [83]), even though that for very small bin size, the two choices are virtually the same. The aforementioned model, called the nonlinear discrete-time Hawkes process (by an analogy with the continuous-time Hawkes process made in Section 2.4) allows for self and cross excitation as well as inhibition, which makes it a good candidate for modelling systems in which the occurrence of an event triggers (excitation) or discourages (inhibition) other events.

The main result of this chapter is proving that the continuous-time process  $(\mathbf{U}_{\lfloor t/h \rfloor}, \sum_{n=0}^{\lfloor t/h \rfloor} \mathbf{P}_n)_{t \in \mathbb{R}_+}$  converges in distribution to a standard continuous-time nonlinear Hawkes process, as the time step  $h$  goes to zero.

The linear univariate Hawkes process was introduced by Hawkes in [46] as an example of a self-exciting point process. It was then extended to the more general nonlinear multivariate process in [18]. Initially used in the context of seismic activity in [70], Hawkes process found applications in many fields such as genome analysis [82], portfolio credit risk [35], micro-structure noise [6] and social networks [5].

Due to its clear branching structure [47] and the possibility of obtaining closed formulae for its asymptotic moments, the linear Hawkes process has been studied extensively in the literature. Linear Hawkes processes with exponentially decaying intensities have been particularly dealt with in many articles. In fact, the memorylessness of the exponential kernel ensures that the intensity is a Markov process that solves a stochastic differential equation (SDE). For instance, Giesecke and Kim [42] devised a *self-exciting point process with CIR intensity*  $(N_t)_{t \in \mathbb{R}_+}$  whose intensity  $(\lambda_t)_{t \in \mathbb{R}_+}$  solves the SDE

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

where  $(W_t)_{t \in \mathbb{R}_+}$  is a Brownian motion that can be seen as a persistent market noise.

Dassios and Zhao [27] suggested an efficient simulation procedure for this process, based on the fact that its evolution follows the well-studied *CIR* dynamics between jumps.

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

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

The link between time series and linear Hawkes processes have been studied in the seminal works of Kirchner [56], where the linear Hawkes process was proven to be a continuous version of the  $\text{INAR}(\infty)$  time series and vice-versa. The proof of convergence relies on the tightness of the sequence in the space of point processes then showing that all the sub-sequential weak limits follow the same Hawkes dynamics.

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

The goal of this discrete-time approximation is twofold; on one hand, most data are recorded on regular time intervals (*e.g.* high-frequency financial data) which makes a discrete-time process more realistic, on the other hand its simulation is more intuitive and tractable than the classic Ogata's thinning algorithm [69] albeit at the expense of introducing discretization bias. Moreover, from a practical point of view, the convergence theorem suggests that the parametric calibration of the time-series yields a good estimation of the Hawkes process' parameters, which is verified numerically.

The chapter is structured as follows: Section 2.2 gives the recursive construction of the discrete-time model and explores its Markov structure. In Section 2.3 we give the definition of the nonlinear Hawkes process, to which we add a Gaussian perturbation and then we show that it is a Markov process if the kernels are Erlang functions. Section 2.4 establishes the link between the discrete-time model and the continuous-time process via the convergence of the infinitesimal generators. Finally, in section 2.5 we run a parametric calibration of the Hawkes process based on a regression on its discrete-time approximation. The experimental results are given with no theoretical guarantees.

## 2.2 The discrete-time model

### 2.2.1 Preliminaries

Throughout this chapter,  $d \in \mathbb{N}$  denotes the number of components and  $h > 0$  denotes the width of a time bin.

Let  $\psi_1, \dots, \psi_d$  be a family of non-negative functions, called the jump rate functions. We set the following assumption

**Assumption 2.2.1.** The functions  $\phi_1, \dots, \psi_d$  are bounded, *i.e.*

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

For any vector  $\mathbf{a} = (a_1, \dots, a_d)$  in  $\mathbb{R}^d$  we denote

$$\boldsymbol{\psi}(\mathbf{a}) = (\psi_1(a_1), \dots, \psi_d(a_d)).$$

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

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

For a more concise notation, we simply write

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

We denote by  $\mathcal{N}(0, I_q)$  the standard multivariate normal distribution of dimension  $q$ .

### 2.2.2 Construction

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

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

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

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

Figure 2.1 gives an illustration in the *self-exciting* case of  $\alpha = 2$  whereas the *self-inhibiting* case  $\alpha = -2$  is illustrated on Figure 2.2.

We now generalize this process to the multivariate setting for a general memory parameter  $p$ . This construction can be seen as an Euler scheme of the continuous-time Hawkes process' intensity, which will be defined in Section 2.3.

*Remark 2.2.2.* For the sake of simplicity, we drop the dependence of the processes  $(P, A, U, Y, \dots)$  on the time step  $h$ .

**Definition 2.2.3.** For an integer  $p \in \mathbb{N}$ , let  $\alpha = (\alpha_q^{ij})_{\substack{i,j=1,\dots,d \\ q=0,\dots,p}} \in \mathbb{R}^{(p+1) \times d \times d}$ . Let  $\beta = (\beta_{ij})_{i,j=1,\dots,d} \in \mathcal{M}_d(\mathbb{R}_+^*)$ ,  $\omega \in (\mathbb{R}_+^*)^d$ ,  $\sigma \in \mathbb{R}_+^d$  and  $\mu \in \mathbb{R}^d$ .

Let  $\mathbf{G} = (\mathbf{G}_k)_{k \in \mathbb{N}}$  be a sequence of *i.i.d* random variables of distribution  $\mathcal{N}(0, I_d)$ . On a different probability space, let  $\mathbf{P} = (\mathbf{P}_k)_{k \in \mathbb{N}}$  be a  $d$ -variate count series such that  $\mathbf{P}_0 = 0$  and set  $\mathcal{F}_k^h = \sigma(\mathbf{P}_0, \dots, \mathbf{P}_k, \mathbf{G}_0, \dots, \mathbf{G}_k)$  the  $\sigma$ -field generated by  $\mathbf{P}$  and  $\mathbf{G}$  until the  $k$ -th time bin.

We say that  $\mathbf{P}$  is a  $d$ -variate nonlinear Poisson autoregression with the memory parameter  $p$ , sensitivity parameters  $\alpha$ , decay parameter  $\beta$ , baseline pre-intensity  $\mu$ , jump rates  $\psi$  and noise parameters  $(\sigma, \omega)$  if

$$\mathbf{P}_{k+1} | \mathcal{F}_k^h \sim \text{Pois}(h\phi(\mathbf{A}_{k+1}))$$

where the pre-intensity  $\mathbf{A} = (A^1, \dots, A^d)$  is

$$A_{k+1}^i = \mu_i + \sum_{j=1}^d \sum_{q=0}^p \alpha_q^{ij} U_{k+1}^{ij,q} + \sigma_i Y_{k+1}^i,$$

where  $(U^{ij,q})_{\substack{i,j=1,\dots,d \\ q=0,\dots,p}}$  and  $(Y^i)_{i=1,\dots,d}$  are respectively the discrete cascade of memory terms and the discrete cascade of perturbation terms, defined recursively by:

$$\begin{cases} Y_{k+1}^i &= e^{-\omega_i h} Y_k^i + \sqrt{h} G_k^i \\ U_{k+1}^{ij,p} &= e^{-\beta_{ij} h} (U_k^{ij,p} + h U_k^{ij,p-1}), \\ &\vdots \\ U_{k+1}^{ij,q} &= e^{-\beta_{ij} h} (U_k^{ij,q} + h U_k^{ij,q-1}), \quad i, j = 1, \dots, d, \\ &\vdots \\ U_{k+1}^{ij,0} &= e^{-\beta_{ij} h} U_k^{ij,0} + P_k^j. \end{cases} \quad (2.2.1)$$

With the given initial states  $U_0^{ij,q} = Y_0^i = 0$ , for every  $i, j = 1, \dots, d$  and  $q = 0, \dots, p$ .

The cumulative sum of  $\mathbf{P}$ , that is the time series  $\mathbf{H} = (\mathbf{H}_n)_{n \in \mathbb{N}}$  where

$$\mathbf{H}_n = \sum_{k=0}^n \mathbf{P}_k,$$

is referred to as the discrete-time Hawkes process (*DTHP* for short).



*Remark 2.2.4.* For each  $i = 1, \dots, d$  the sequence  $(\mathbf{P}_n)$  can be seen as the number of events of a continuous-time counting process  $(\mathbf{N}_t)_{t \geq 0}$  observed in a time bin of size  $h$ , i.e.

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

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

The System (2.2.1) in Definition 2.2.3 can be put under the matrix form.

Let

$$\mathbf{U}_k^{ij} = \begin{pmatrix} U_k^{ij,p} \\ \vdots \\ U_k^{ij,0} \end{pmatrix} \in \mathbb{R}^{p+1}, \quad \mathbf{Y}_k = \begin{pmatrix} Y_k^1 \\ \vdots \\ Y_k^d \end{pmatrix} \in \mathbb{R}^d. \quad (2.2.2)$$

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

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

super-diagonal and 0 elsewhere.

From the definition of  $K$  we set  $M_{ij}^h = e^{-\beta_{ij}h}(I + hK)$ , where  $I$  is the identity matrix.

$\mathbf{e}_{p+1}$  designates the vector of  $\mathbb{R}^{p+1}$  with 0 everywhere and 1 in the last component and  $\Omega^h = \text{diag}(e^{-\omega_1 h}, \dots, e^{-\omega_d h})$ .

*Proposition 2.2.5.* The cascade  $(\mathbf{U}, \mathbf{Y}) = ((\mathbf{U}_n, \mathbf{Y}_n))_{n \in \mathbb{N}}$  where  $\mathbf{U}_n = (\mathbf{U}_n^{ij})_{i,j=1,\dots,d}$  is a Markov chain that follows the matrix dynamics

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

Its transition operator  $\mathcal{T}_h$  evaluated at  $(\mathbf{u}, \mathbf{y}) \in ((\mathbb{R}^{p+1})^{d \times d} \times \mathbb{R}^d)$  for a bounded measurable function  $f : ((\mathbb{R}^{p+1})^{d \times d} \times \mathbb{R}^d) \rightarrow \mathbb{R}$  has the expression

$$\begin{aligned} &\mathcal{T}_h f(\mathbf{u}, \mathbf{y}) \\ &= \mathbb{E} [f(\mathbf{U}_{n+1}, \mathbf{Y}_{n+1}) | (\mathbf{U}_n, \mathbf{Y}_n) = (\mathbf{u}, \mathbf{y})], \\ &= e^{-h \sum_{i=1}^d \psi_i(\tilde{a}^i(\mathbf{u}, \mathbf{y}))} \mathbb{E}^{\mathcal{N}} \left[ f \left( \left( M_{ij}^h \mathbf{u}^{ij} \right)_{i,j=1,\dots,d}, \Omega^h \mathbf{y} + \sqrt{h} \mathbf{G}' \right) \right] \\ &\quad + e^{-h \sum_{i=1}^d \psi_i(\tilde{a}^i(\mathbf{u}, \mathbf{y}))} \sum_{m=1}^d \mathbb{E}^{\mathcal{N}} \left[ f \left( \left( M_{ij}^h \mathbf{u}^{ij} + \mathbf{1}_{m=j} \mathbf{e}_{p+1} \right)_{i,j=1,\dots,d}, \Omega^h \mathbf{y} + \sqrt{h} \mathbf{G}' \right) \right] h \psi_m(\tilde{a}^m(\mathbf{u}, \mathbf{y})) \\ &\quad + R_2(\mathbf{u}, \mathbf{y}). \end{aligned}$$

Where the remainder  $R_2$  is

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

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

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

$\mathbb{E}^{\mathcal{N}}$  stands for the integration with respect to the standard normal distribution in  $d$ -dimensions, and  $\mathbf{G}'$  is a variable of distribution  $\mathcal{N}(0, I_d)$ .

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

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

$$\begin{aligned} P_n^i &\sim \text{Pois} \left( h\psi_i \left( \mu_i + \sigma_i Y_n^i + \sum_{j=1}^d \sum_{q=0}^p \alpha_q^{ij} U_n^{ij,q} \right) \right) \\ &\sim \text{Pois} \left( h\psi_i \left( \mu_i + \sigma_i y^i + \sum_{j=1}^d \langle \boldsymbol{\alpha}^{ij}, \mathbf{u}^{ij} \rangle \right) \right) \\ &\sim \text{Pois} (h\psi_i (\tilde{a}^i(\mathbf{u}, \mathbf{y}))). \end{aligned}$$

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

$$\begin{aligned} \mathcal{T}_h f(\mathbf{u}, \mathbf{y}) &= \mathbb{E} [f(\mathbf{U}_{n+1}, \mathbf{Y}_{n+1}) | (\mathbf{U}_n, \mathbf{Y}_n) = (\mathbf{u}, \mathbf{y})] \\ &= \mathbb{E}^{\mathcal{N}} \left[ \mathbb{E}^P \left[ f \left( (M_{ij}^h \mathbf{U}_n^{ij} + P_n^j \mathbf{e}_{p+1})_{i,j=1,\dots,d}, \Omega^h \mathbf{Y}_n + \sqrt{h} \mathbf{G}_n \right) \middle| (\mathbf{U}_n, \mathbf{Y}_n) = (\mathbf{u}, \mathbf{y}) \right] \right] \\ &= \mathbb{E}^{\mathcal{N}} \left[ \sum_{i_1, \dots, i_d=0}^{+\infty} f \left( (M_{ij}^h \mathbf{u}^{ij} + i_j \mathbf{e}_{p+1})_{i,j=1,\dots,d}, \Omega^h \mathbf{y} + \sqrt{h} \mathbf{G}' \right) \right. \\ &\quad \left. \mathbb{P} \left[ P_n^1 = i_1, \dots, P_n^d = i_d \middle| (\mathbf{U}_n, \mathbf{Y}_n) = (\mathbf{u}, \mathbf{y}) \right] \right]. \end{aligned}$$

The exchange of the infinite sum and the expectation is possible because  $f$  is bounded measurable (Fubini-Lebesgue Theorem). Keeping in mind that, knowing

$(\mathbf{U}_n, \mathbf{Y}_n) = (\mathbf{u}, \mathbf{y})$ , the variables  $P_n^1, \dots, P_n^d$  are independent and  $P_n^i \sim \text{Pois}(h\psi_i(\tilde{a}^i(\mathbf{u}, \mathbf{y})))$ , we have that

$$\begin{aligned} & \mathcal{T}_h f(\mathbf{u}, \mathbf{y}) \\ &= \sum_{i_1, \dots, i_d=0}^{+\infty} \mathbb{E}^{\mathcal{N}} \left[ f \left( (M_{ij}^h \mathbf{u}^{ij} + i_j \mathbf{e}_{p+1})_{i,j=1, \dots, d}, \Omega^h \mathbf{y} + \sqrt{h} \mathbf{G}' \right) \prod_{j=1}^d \frac{(h\psi_j(\tilde{a}^j(\mathbf{u}, \mathbf{y})))^{i_j}}{i_j!} e^{-h\psi_j(\tilde{a}^j(\mathbf{u}, \mathbf{y}))} \right] \\ &= e^{-h \sum_{i=1}^d \psi_i(\tilde{a}^i(\mathbf{u}, \mathbf{y}))} \mathbb{E}^{\mathcal{N}} \left[ f \left( (M_{ij}^h \mathbf{u}^{ij})_{i,j=1, \dots, d}, \Omega^h \mathbf{y} + \sqrt{h} \mathbf{G}' \right) \right] \\ &+ e^{-h \sum_{i=1}^d \psi_i(\tilde{a}^i(\mathbf{u}, \mathbf{y}))} \sum_{m=1}^d \mathbb{E}^{\mathcal{N}} \left[ f \left( (M_{ij}^h \mathbf{u}^{ij} + \mathbb{1}_{m=j} \mathbf{e}_{p+1})_{i,j=1, \dots, d}, \Omega^h \mathbf{y} + \sqrt{h} \mathbf{G}' \right) \right] h\psi_m(\tilde{a}^m(\mathbf{u}, \mathbf{y})) \\ &+ R_2(\mathbf{u}, \mathbf{y}). \end{aligned}$$

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

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

□

In order to illustrate the discrete time Hawkes process we simulate it for  $h = 0.2$ ,  $d = 2$  and  $p = 1$ . The jump-rate functions are chosen to be  $\psi_1(x) = \max(x, 0) \wedge 40$  and  $\psi_2(x) = \log(1 + e^x) \wedge 40$ . The decay rates are chosen to be  $\beta_{11} = \beta_{12} = 5$ ,  $\beta_{21} = \beta_{22} = 6$  and the baseline parameters are  $(\mu_1, \mu_2) = (2, 3)$ . The results are shown on Figures 2.3 and 2.4 in the Appendix 2.6.

Notice how  $H_1$  (or  $H_2$ ) has jumps whose sizes are larger than one. This is possible because the jumps are chosen to be Poisson variables (along the lines of [55]), unlike the alternative choice of a Bernoulli variable (*cf.* [83]). Even though the two choices are equivalent when  $h \rightarrow 0$ , the Poisson approach allows multiple jumps within a single time bin and allows the possibility  $h\psi_i(A_k^i) > 1$ .

*Remark 2.2.6.* Even though in Proposition 2.2.5 the cascade  $(\mathbf{U}_n, \mathbf{Y}_n)_{n \in \mathbb{N}}$  were proven to be a Markov chain, it is straightforward to show that  $(\mathbf{U}_n, \mathbf{Y}_n, \mathbf{H}_n)_{n \in \mathbb{N}}$  is also a Markov chain. From now on, we mainly focus on the pre-intensity, because it already contains the jumps of  $\mathbf{H}$ .

### 2.2.3 Stability

We show that if the jump rates are bounded, the cascade  $(\mathbf{U}, \mathbf{Y})$  has a bounded average and is thus almost surely finite.

*Proposition 2.2.7.* Let  $(\mathbf{U}_n, \mathbf{Y}_n)_{n \in \mathbb{N}}$  be the Markov chain defined by (2.2.2) following the recursive formula of Proposition 2.2.5 and assume Assumption 2.2.1 is in force. Then there is a positive constant  $C$  independent of  $n$  such that

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

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

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

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

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

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

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

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

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

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

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

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

with

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

hence the result.  $\square$

The result above does not prove stability in variation (*cf.* [18]), but it ensures that the self-excitation (in case  $\boldsymbol{\alpha}$  is positive) is not explosive. In fact, the result of Proposition 2.2.7 entails that (using Markov's inequality)

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

The result also ensures that the sequence  $(\mathbf{U}_n, \mathbf{Y}_n)_{n \in \mathbb{N}}$  is tight, thus having weakly convergent sub-sequence, which in turn can be used to prove the convergence towards a stationary distribution. We chose not to prove it in this paper, as the long-time behaviour has already been studied in [30] for the continuous-time Hawkes process, using classical Markov process techniques.

## 2.3 The continuous-time Hawkes process

### 2.3.1 Preliminaries

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

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

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

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

where

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

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

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

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

### 2.3.2 Definition and construction

We start by defining the nonlinear multivariate Hawkes process in the general framework.

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

$$N_t^i = \sum_{j \geq 1} \mathbb{1}_{\tau_j^i \leq t}.$$

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

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

which measures how likely it is for  $N^i$  to jump between  $t$  and  $t + dt$ , right before  $t$ . Let  $(\phi_{ij})_{i,j=1,\dots,d}$  be a set of real integrable (not necessarily non-negative) functions on  $\mathbb{R}_+$  and  $\boldsymbol{\psi}$  as defined in Subsection 2.2.

The process  $\mathbf{N}$  is called a nonlinear Hawkes process of memory kernel  $\phi$  and jump rate  $\boldsymbol{\psi}$  if its intensity follows the dynamics

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

for some  $\omega_i > 0$ , for every  $i = 1, \dots, d$ . Brémaud and Massoulié [18] constructed the nonlinear Hawkes process as the result of the embedding from a Poisson measure. We give the result in the following theorem.

*Theorem 2.3.1.* Let  $\boldsymbol{\Pi}$  a  $d$ -dimension Poisson measure on  $\mathbb{R}_+^2$  and  $\mathbf{W}$  a standard  $d$ -variate Brownian motion. Let  $\boldsymbol{\psi}$  and  $\phi$  be functions as defined in the beginning of this subsection. Let  $\boldsymbol{\mu} \in \mathbb{R}^d$ ,  $\boldsymbol{\omega} \in (\mathbb{R}_+^*)^d$  and  $\boldsymbol{\sigma} \in \mathbb{R}_+^d$ . The following SDE has a unique solution  $(\mathbf{N}, \boldsymbol{\lambda})$  with  $\mathbf{N}$   $\mathcal{F}$ -measurable and  $\boldsymbol{\lambda}$   $\mathcal{F}$ -predictable

$$\begin{cases} N_t^i = \int_{(0,t] \times \mathbb{R}_+} \mathbf{1}_{\{\theta \leq \lambda_s^i\}} \boldsymbol{\Pi}^i(ds, d\theta), & t \geq 0, \quad i = 1, \dots, d, \\ \lambda_t^i = \psi_i \left( \mu_i + \int_{[0,t)} \left( \sum_{j=1}^d \phi_{ij}(t-s) dN_s^j \right) + \sigma_i e^{-\omega_i(t-s)} dW_s^i \right) & t \geq 0 \quad i = 1, \dots, d. \end{cases} \quad (2.3.1)$$

*Remark 2.3.2.* In [18], the assumption that the functions  $\psi_1, \dots, \psi_d$  are bounded can be relaxed and replaced by the condition that  $\psi_i$  is  $l_i$ -Lipschitz and that the matrix

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

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

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

a piece-wise deterministic dynamics.

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

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

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

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

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

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

When the memory kernels are Erlang functions, the stability condition in Remark 2.3.2 becomes

**Assumption 2.3.3.** The functions  $\psi_i$  are  $l_i$ -Lipschitz for  $i = 1, \dots, d$  and the matrix

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

has a spectral radius strictly less than one.

*Remark 2.3.4.* This is a special class of Erlang memory kernels. The more general approach would be to make the  $(\beta_i)_{i=1,\dots,d}$  dependent on  $q$ , along the lines of [30]. This choice would allow to approximate every  $L^1(\mathbb{R}_+, \mathbb{R})$  kernel, hence approximating any Hawkes process with an integrable kernel. In fact, it has been proven in [78] that for any  $T$  there is a constant  $C_T$  such that for two Hawkes processes  $\mathbf{N}$  and  $\tilde{\mathbf{N}}$  of memory kernels  $\phi$  and  $\tilde{\phi}$  respectively, we have that

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

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

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

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

$$V_t^{ij,q} := \int_{[0,t)} \frac{(t-s)^q}{q!} e^{-\beta_{ij}(t-s)} dN_s^j, \quad (2.3.2)$$

and

$$Z_t^i := \int_{[0,t)} e^{-\omega_i(t-s)} dW_s^i, \quad (2.3.3)$$

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

*Theorem 2.3.5.* Let  $\mathbf{N}$  be a  $d$ -variate Hawkes process with Erlang memory kernels. For every  $i = 1, \dots, d$ , set  $\boldsymbol{\alpha}^{ij} = (\alpha_p^{ij}, \dots, \alpha_0^{ij}) \in \mathbb{R}^{p+1}$ .

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

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

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

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

where  $\mathbf{e}_{p+1} = (0, \dots, 0, 1)$  and  $\mathbf{W} = (W^1, \dots, W^d)$  is a standard  $d$ -variate Brownian motion.

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

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

which yields after differentiation

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

and since

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



we have

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

and after re-arranging the terms

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

For the last term

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

a straightforward differentiation yields

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

The SDE for  $(Z^1, \dots, Z^d)$  is obtained identically, in fact, it is a multivariate Ornstein-Uhlenbeck process.  $\square$

*Remark 2.3.6.* Note that the vector process formed by the memory and perturbation cascades as well as the point process  $(\mathbf{V}, \mathbf{Z}, \mathbf{N})$  is also a Markov process. But since it makes the notations and the computations more tedious, we simply consider  $(\mathbf{V}, \mathbf{Z})$  that already contains all the jump information.

In fact, the jumps of  $N^i$  are exactly the points of the underlying Poisson measure under the curve  $\psi_i(\tilde{a}^i(\mathbf{V}_t, \mathbf{Z}_t))$ :

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

This remark is the continuous-time analogue of Remark 2.2.6.

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

*Proposition 2.3.7.* We recall that  $\mathbf{e}_{p+1}$  is the vector that has 1 at the  $p+1$ -th entry and 0 elsewhere. Let

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

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

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

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

*Proof.* For the unperturbed univariate case, the expression of the generator can be found in [30]. The second derivatives are a result of the Gaussian perturbation. This is more precisely the generator of a multivariate Ornstein-Uhlenbeck process.  $\square$

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

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

to ensure that the intensity remains positive (*cf.* [42] and [27]), with the upper bound  $\sigma^2 \leq 2\mu\beta$ . The existence of these "jump-diffusion" processes as well as other results on the affine point processes can be found in [31].

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

### 2.4.1 Preliminaries

Let  $\mathcal{D}(\mathbb{R}_+, E)$  be the space of right continuous processes with left limits (*càdlàg*) on a metric space  $E$ . We denote by " $\implies$ " the weak convergence of processes in the Skorokhod topology (*cf.* [14] for example).

Throughout this chapter,  $\hat{\mathcal{C}}(E)$  denotes the set of continuous functions from  $E$  to  $\mathbb{R}$  that vanish at infinity and the functional convergence is in the uniform norm denoted by  $\|\cdot\|$ . In the Markov framework, the weak convergence of a family of processes in the Skorokhod topology can be deduced directly from the behavior of its transition operators. This is formulated in the following theorem in [36]:

*Theorem 2.4.1.* Let  $E$  be locally compact and separable. For  $h > 0$ , let  $\nu_h(\mathbf{x}, \cdot)$  be a transition function on  $E \times \mathcal{B}(E)$  such that the operator  $\mathcal{T}_h$  defined by

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

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

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

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

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

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

$$X^h \xrightarrow{h \rightarrow 0} X.$$

The convergence for the semi-groups is usually proved by showing the convergence of the infinitesimal generators on a subset of functions called the *core* of the limiting generator.

The definition of the core of a generator can be found in Section 3 in [36]. In our proofs we use the following sufficient condition for  $D$  to be a core which is stated in Proposition 1.3.3 in [36]: Let  $\mathcal{A}$  be the generator of a contraction semi-group  $\mathcal{T}(t)$  defined on a Banach functional space  $L$ . Let  $D$  a dense subset of  $L$ . If there exists a dense subset  $D_0$  such that  $D_0 \subset D \subset L$  and if for any  $t \geq 0$ ,  $\mathcal{T}(t) : D_0 \rightarrow D$ , then  $D$  is a core for  $\mathcal{A}$ .

*Proposition 2.4.2.* Let  $L$  be a functional Banach space. For  $h > 0$ , let  $\mathcal{T}_h$  be a linear contraction on  $L$  and put

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

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

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

*Proof.* This is a modification of Theorem 1.6.5 in [36] page 31. □

*Remark 2.4.3.* Even though the auxiliary processes defined by (2.3.2) are predictable (hence left continuous with right limits or *càglàd*), the theorems used in this section prove the convergence in  $\mathcal{D}(\mathbb{R}_+, E)$ . That is why we replace our processes with their *càdlàg* version that has the exact same jumps and dynamics.

### 2.4.2 Main result

In this section we prove that the discrete time Hawkes process defined in Section 2.2.2 is an approximation of the continuous-time multivariate nonlinear Hawkes process. To do so we provide a convergence theorem for the auxiliary processes that contain the jumps and the dynamics. Note that it is possible to prove the results for the augmented process  $(\mathbf{V}, \mathbf{Z}, \mathbf{N})$ , but this is not necessary as  $(\mathbf{V}, \mathbf{Z})$  is sufficient to reconstruct  $\mathbf{N}$  (cf. Remark 2.3.6).

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

*Proposition 2.4.4.* Let  $\mathcal{T}_h$  be the discrete cascade's transition operator defined in 2.2.5 and let  $\mathcal{A}$  be the generator for the auxiliary processes of a Hawkes process defined in Proposition 2.3.7. Set

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

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

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

*Proof.* Let  $f \in \mathcal{S}(E)$  and  $h > 0$ . Using Lemma 2.6.2 we have that

$$\begin{aligned} & \mathcal{T}_h(\mathbf{u}, \mathbf{y}) \\ &= e^{-h \sum_{i=1}^d \psi_i(\tilde{a}^i(\mathbf{u}, \mathbf{y}))} \mathbb{E}^{\mathcal{N}} \left[ f \left( \left( M_{ij}^h \mathbf{u}^{ij} \right)_{i,j=1,\dots,d}, \Omega^h \mathbf{y} + \sqrt{h} \mathbf{G}' \right) \right] \\ &+ e^{-h \sum_{i=1}^d \psi_i(\tilde{a}^i(\mathbf{u}, \mathbf{y}))} \sum_{m=1}^d \mathbb{E}^{\mathcal{N}} \left[ f \left( \left( M_{ij}^h \mathbf{u}^{ij} + \mathbb{1}_{m=j} \mathbf{e}_{p+1} \right)_{i,j=1,\dots,d}, \Omega^h \mathbf{y} + \sqrt{h} \mathbf{G}' \right) \right] h \psi_m(\tilde{a}^m(\mathbf{u}, \mathbf{y})) \\ &+ O(h^2), \end{aligned}$$

with  $\mathbf{G}' \sim \mathcal{N}(0, I_d)$ . Note that the subscript  $i, j = 1, \dots, d$  has been omitted to make the notations less tedious.

For a  $d$ -variate Gaussian variable  $\mathbf{G}'$ , we write a Taylor expansion in the second component

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

where the remainder  $R_r$  satisfies

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

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

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

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

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

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

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

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

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

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

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

Applying the same arguments to the second derivative yields

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

Finally, using the fact that the jump rates  $\psi_1, \dots, \psi_d$  are bounded and the Taylor expansion of the exponential

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

we conclude that

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

which yields the result.  $\square$

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

*Theorem 2.4.5.* Let  $(\mathbf{U}_n, \mathbf{Y}_n, \mathbf{H}_n)_{n \in \mathbb{N}}$  be the Markov chain defined in Proposition 2.2.5.

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

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

(*cf.* Figure 2.5 for an illustration).

Then

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

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

*Proof.* It is sufficient to prove the convergence of  $(\tilde{\mathbf{V}}_t^h, \tilde{\mathbf{Z}}_t^h)_{t \in \mathbb{R}_+}$  to  $(\mathbf{V}_t, \mathbf{Z}_t)_{t \in \mathbb{R}_+}$  as  $h \rightarrow 0$  (*cf.* Remarks 2.2.6 and 2.3.6). Since the Schwartz space  $\mathcal{S}(E)$  is a core for the infinitesimal generator  $\mathcal{A}$  (*cf.* Lemma 2.6.5), Propositions 2.4.2 and 2.4.4 ensure that

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

for each  $f$  in the Banach space (*cf.* Lemma 2.6.1)  $\hat{\mathcal{C}}(E)$ .

Theorem 2.4.1 yields that the process

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

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

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

$$\psi_1(x) = (x)_+ \wedge 40, \quad \psi_2(x) = \log(1 + e^x) \wedge 40$$

and the memory kernels are

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

The convergence is illustrated in the transition from Figure 2.5 ( $h = 0.125$ ) to Figure 2.6 ( $h = 5 \cdot 10^{-3}$ ).

## 2.5 Parameters calibration using regression

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

For the exponential kernels, Ozaki [72] proved that the likelihood's computation is linear in the number of events, numerical results are given in the univariate case as well. Da Fonseca and Zaatour [24] derived an estimation of the parameters using the stationary moments, which can be explicitly derived for the linear Hawkes process with exponential kernels. This method is not applicable in the nonlinear framework since the expressions of the first moments are not known explicitly.

Since the kernel's shape is usually not known in advance and since the exponential family is not rich enough (*e.g.* power laws, non monotonous functions), methods for non-parametric estimation have been derived for the linear Hawkes process. The first of which is based on the numerical resolution of a system of Wiener-Hopf equations, introduced by Bacry and Muzy [9]. The second is introduced by Kirchner [55] and by Eichler *et al.* [33] based on time series. These methods are close to the ones we present in this paper. The reader can refer to [84] for a comparison between the time series method and an E-M method.

The nonlinear framework has also been studied, though to a lesser extent. For the parametric estimation, Lemonnier and Vayatis [59] derived an algorithm for the log-likelihood maximisation for kernels composed by an exponential mixture and a ReLU jumps rate. For a more general choice of the jump rate functions as well as kernels, Sulem *et al.* [89] obtained concentration rates of the posterior distribution on the parameters.

When it comes to the non parametric approach, Reynaud-Bouret *et al.* [80] devised a method for the non-parametric estimation for multivariate Hawkes processes with ReLU jump rates.

In this section, we assume that both Assumptions 2.2.1 (boundedness of the jump-rate) and 2.3.3 (jump-rate functions are  $l_i$ -Lipschitz and the matrix  $\left( l_i \sum_{q=0}^p \frac{|\alpha_0^{ij}|}{\beta_{ij}^{q+1}} \right)_{i,j=1,\dots,d}$  has a spectral radius less than one) are in force. In fact, if assumption 2.3.3 is not respected, the pre-intensity  $\tilde{a}^i$  can grow to infinity, before being capped by the upper bound  $\|\psi_i\|_\infty$ , transforming the Hawkes process into a Poisson process of intensity  $\|\psi_i\|_\infty$  from which no other parameters can be recovered. This is the problem of

identifiability which is discussed in more details in [89].

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

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

For the sake of brevity, we set

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

The underlying process is supposed to follow multivariate Hawkes dynamics

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

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

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

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

for a fixed  $h > 0$ .

General results about CLS estimators for time-series, including asymptotic bias and normality can be found in the seminal work of [57].

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

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

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

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



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

$$\begin{aligned}\partial_{\mu_i} Q_{M+1} &= \partial_{\mu_i} Q_M - 2hZ \\ \nabla_{\alpha^{ij}} Q_{M+1} &= \nabla_{\alpha^{ij}} Q_M - 2h\mathbf{U}_M^{ij} Z \\ \partial_{\beta_{ij}} Q_{M+1} &= \partial_{\beta_{ij}} Q_M - 2h \left\langle \alpha^{ij}, \partial_{\beta_{ij}} \mathbf{U}_{M+1}^{ij} \right\rangle Z \\ \partial_{\beta_{ij}} \mathbf{U}_{M+1}^{ij} &= M^{ij} \left( \partial_{\beta_{ij}} \mathbf{U}_M^{ij} - h\mathbf{U}_M^{ij} \right),\end{aligned}$$

where

$$Z = (\psi_i)' \left( \mu_i + \sum_{j=1}^d \langle \alpha^{ij}, \mathbf{U}_{M+1}^{ij} \rangle \right) \left( P_{M+1}^i - h\psi_i \left( \mu_i + \sum_{j=1}^d \langle \alpha^{ij}, \mathbf{U}_{M+1}^{ij} \rangle \right) \right).$$

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

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

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

with jump-rates

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

and baseline intensities

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

The minimization of  $Q_M$  is done using the SLSQP method of the function `scipy.optimize.minimize` with Assumption 2.3.3 as a constraint.

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

The minimization results are then averaged (for five calibrations) and are given in Figure 2.7.

### 2.5.1 Calibration experiment for the perturbed Hawkes process

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

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

and

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

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

The result of the CLS minimisation is given on Figures 2.8, 2.9 and 2.10.

### 2.5.2 Calibration with a lower degree

We now simulate 15 continuous-time Hawkes processes with kernels

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

with jump-rates

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

and baseline intensities

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

The Gaussian perturbation's parameters are

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

and

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

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

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

The exponential kernels composed with the average parameters are illustrated in Figure 2.11. When it comes to the baseline intensities refer to Figure 2.12

## 2.6 Lemmata

*Lemma 2.6.1.* Set  $E = (\mathbb{R}^{p+1})^{d \times d} \times \mathbb{R}^d$ . Then the space  $\hat{\mathcal{C}}(E)$  of continuous functions that vanish at infinity equipped with the uniform norm is a Banach space.

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

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

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

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

Let  $p \in \mathbb{N}$  and  $\mathbf{x} \in E$

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

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

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

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

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

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

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

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

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

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

*Lemma 2.6.2.* Assume that Assumption 2.2.1 is in force.

Let  $R_2$  be the remainder defined in Proposition 2.2.5. Then, uniformly in  $\mathbf{u}$  and  $\mathbf{y}$ ,

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

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

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

where  $\mathbf{G} \sim \mathcal{N}(0, I_d)$ . Since  $f$  is bounded we have that

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

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

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

□

*Lemma 2.6.3.* For  $i, j = 1, \dots, d$ , let  $M_{ij}^h = e^{-\beta_{ij}h}(I+hK)$  and  $\Omega^h = \text{diag}(e^{-\omega_1 h}, \dots, e^{-\omega_d h})$  be the matrices defined in Section 2.2.2. Then

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

and

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

*Proof.* This is a simple second order Taylor expansion. □

*Lemma 2.6.4.* Set  $E = (\mathbb{R}^{p+1})^{d \times d} \times \mathbb{R}^d$ .

Let  $\mathcal{T}$  be the semi-group associated with the Markov process defined by the auxiliary processes (2.3.2) and (2.3.3), *i.e.* defined for every  $t \geq 0$  and  $f \in \hat{\mathcal{C}}(E)$  by

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

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

*Proof.* It is trivial to see that for any  $t \geq 0$ ,  $\mathcal{T}(t)$  is a contraction (*i.e.*  $\|\mathcal{T}(t)f\| \leq \|f\|$  for every  $f \in \hat{\mathcal{C}}(E)$ ) and positive (*i.e.* for every non-negative function  $f$ ,  $\mathcal{T}(t)f$  is a non-negative function).

Now we show that it is strongly continuous. To do so, one must show that it has the Feller property

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

First, by solving the SDE in Theorem 2.3.5 with the initial conditions  $\mathbf{V}_0^{ij} = \mathbf{v}^{ij}$  and  $\mathbf{Z}_0 = \mathbf{z}$  we have

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

and

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

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

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

This, combined with the weak continuity (trivial) yields that  $\mathcal{T}$  is a Feller semi-group (cf. the first chapter in [16]).  $\square$

*Lemma 2.6.5.* Let  $E = (\mathbb{R}_+^{(p+1)})^{d \times d} \times \mathbb{R}^d$ .

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

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

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

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

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

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

where  $\xi_t^{1,ij} = \int_{[0,t)} e^{-(K-\beta_{ij}I)(t-s)} \mathbf{e}_{p+1} dN_s^j$ , and  $\xi_t^{2,ij} = \int_0^t \text{diag}(e^{-\omega_1(t-s)}, \dots, e^{-\omega_d(t-s)}) d\mathbf{W}_s$  are two finite random variables. Since  $\mathcal{S}(E)$  is stable by composition with an affine transform (cf. [40]) and since the differentiation and the expected value can be exchanged ( $f$  is bounded), then  $\mathcal{T}(t)f \in \mathcal{S}(E)$ .  $\square$

## Appendix: Figures

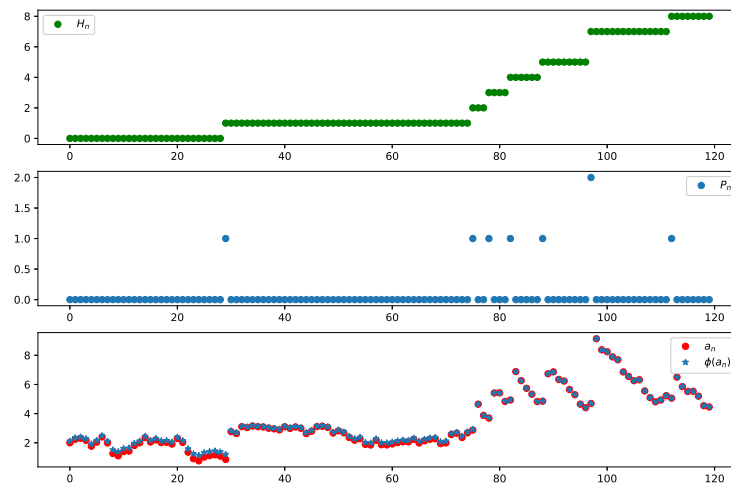


Figure 2.1 – The process is *self-exciting* for  $\alpha = 2$ . This is seen as events tend to be clustered in time. Notice how one bin contains two events.

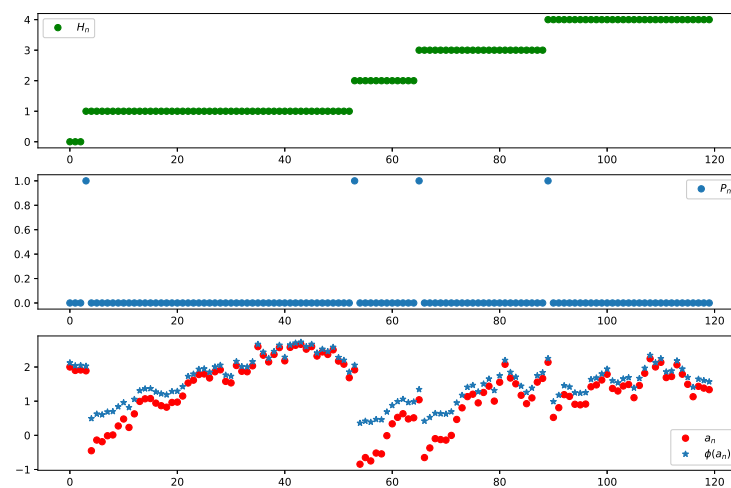


Figure 2.2 – The process is *self-inhibiting* for  $\alpha = -2$ . Events tend to be distanced from one another.

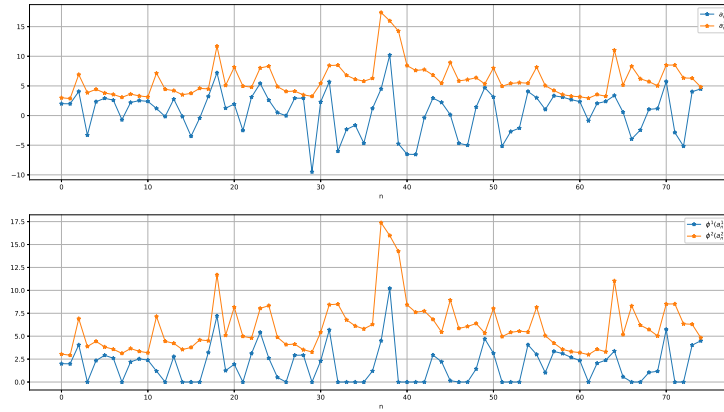


Figure 2.3 – A plot of a realisation of  $\mathbf{a}_k = (\tilde{a}^m(\mathbf{U}_k, \mathbf{Y}_k))_{m=1,2}$  (upper plot) as well as  $\phi(\mathbf{a}_k)$  (lower plot) for  $k \leq 75$ . The interaction parameters are  $(\alpha_1^{11}, \alpha_0^{11}) = (0, 2)$ ,  $(\alpha_1^{12}, \alpha_0^{12}) = (15, -3)$ ,  $(\alpha_1^{21}, \alpha_0^{21}) = (-6, 4)$  and  $(\alpha_1^{22}, \alpha_0^{22}) = (10, 0)$ . The perturbation parameters are  $(\sigma_1, \sigma_2) = (0.1, 0.2)$ .

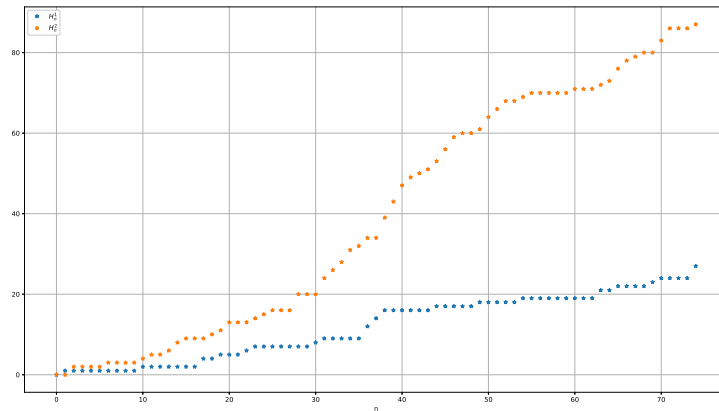


Figure 2.4 – A realisation of  $\mathbf{H}_k$  for  $k \leq 75$ .

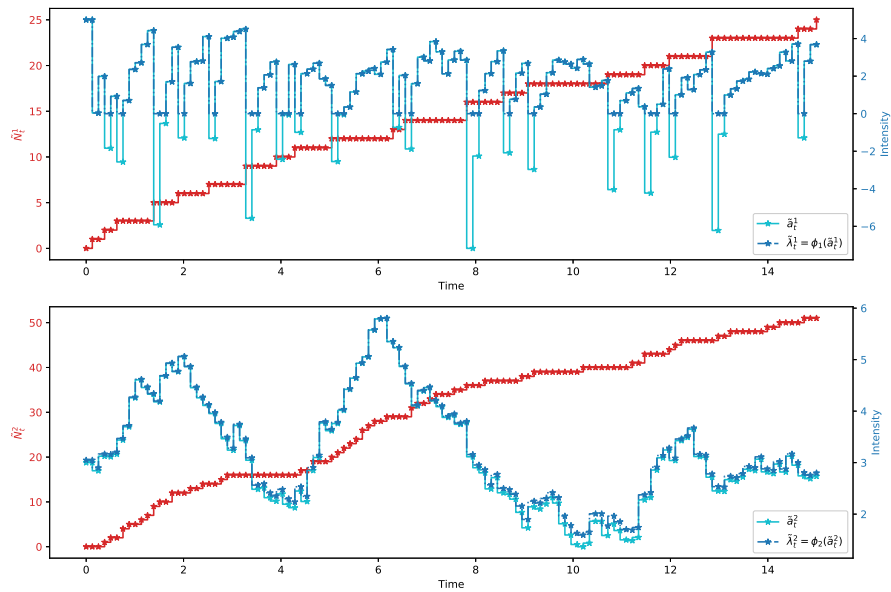


Figure 2.5 – Simulation for  $h = 0.125$ . The baseline intensities are  $(\mu_1, \mu_2) = (5, 3)$ .

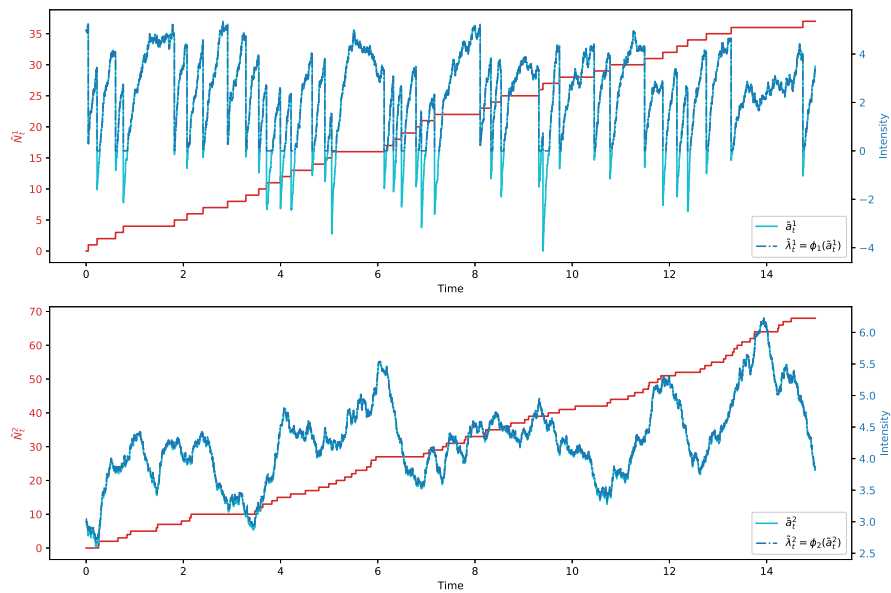


Figure 2.6 – Simulation for  $h = 5 \cdot 10^{-3}$ . The Gaussian perturbation parameters are  $\sigma_1 = 1.5$  and  $\sigma_2 = 0.5$ . The graph is drawn in simple continuous lines for visibility.



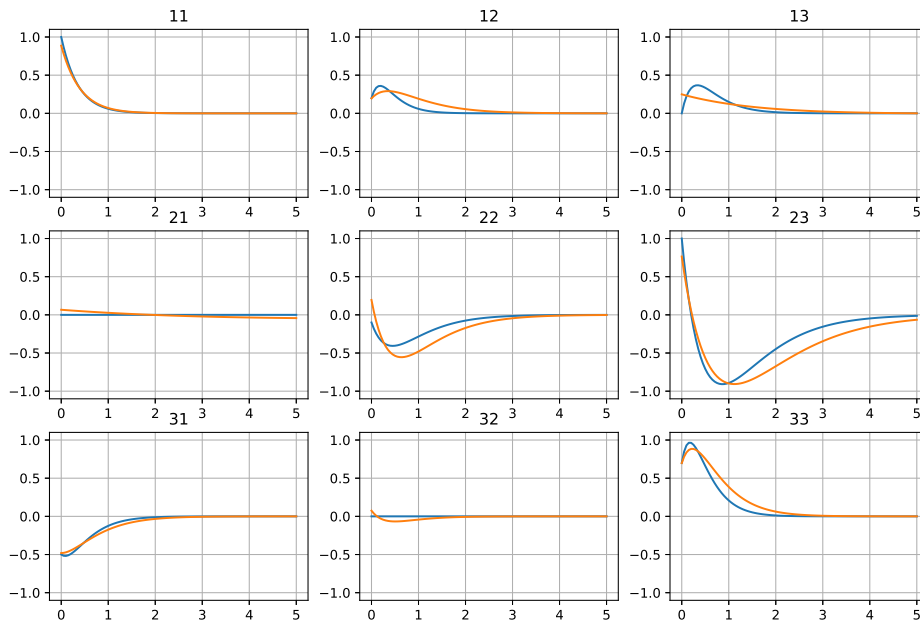


Figure 2.7 – The blue curves are the actual kernels and the orange curves are the calibration results. The estimated baseline intensities are  $\hat{\boldsymbol{\mu}} = (0.37, 1.3, 1.5)$ .

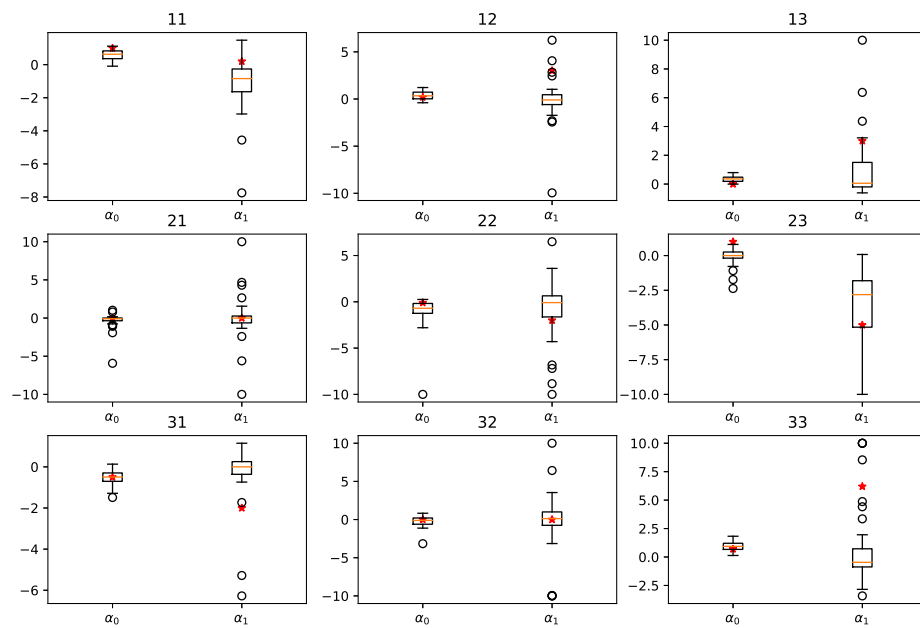


Figure 2.8 – Calibrations of the coefficient  $\alpha_0^{ij}$  and  $\alpha_1^{ij}$  for  $i, j = 1, \dots, d$ . The estimation is clearly less precise for  $\alpha_1$  with a larger variance. The red stars are the markers of the ground truth value.

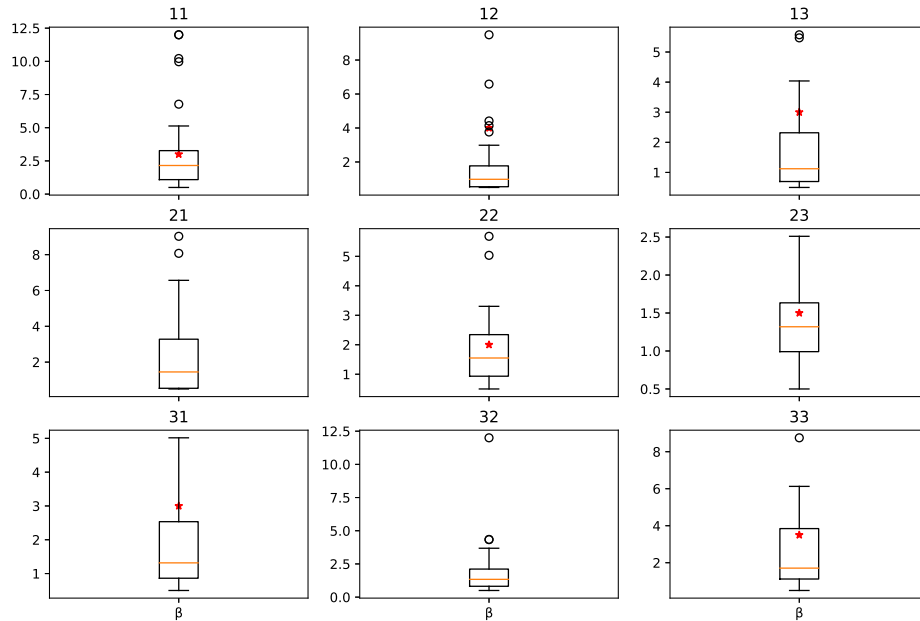


Figure 2.9 –  $\beta_{21}$  and  $\beta_{32}$  are not defined because the kernels are equal to zero.

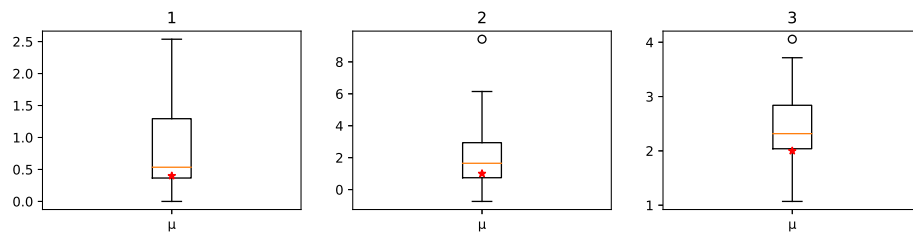


Figure 2.10 – The estimator tends to yield larger values for the baseline intensities. This is due to the aforementioned bias, where some self-excitations are ignored.

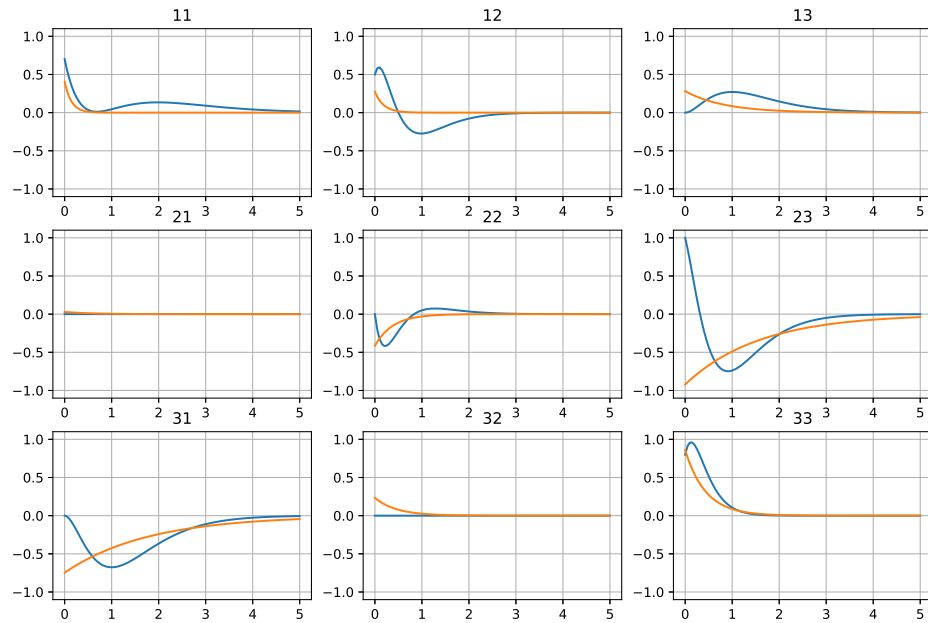


Figure 2.11 – The overall effect (excitation, inhibition, independence) are well captured by the exponential kernels. For composite kernels (*e.g.*  $\delta_{12}$  and  $\delta_{22}$ ) the exponential kernel captures the more influential behaviour.

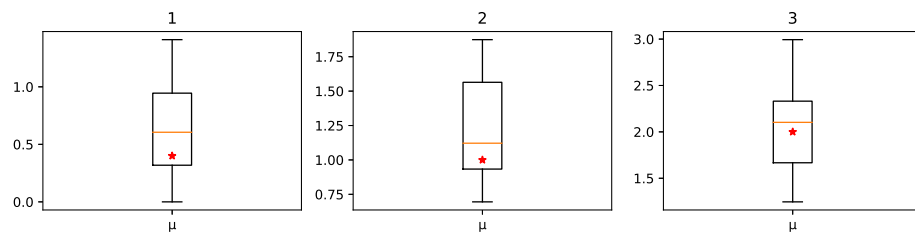


Figure 2.12 – Again, the baseline intensities are over estimated because of the bias.



# The Malliavin-Stein method for Hawkes functionals

---

## Abstract

In this chapter, following Nourdin-Peccati’s methodology, we combine the Malliavin calculus and Stein’s method to provide general bounds on the Wasserstein distance between the law of functionals of a compound Hawkes process and the one of a Gaussian random variable. To achieve this, we rely on the Poisson embedding representation of a Hawkes process to provide a Malliavin calculus for the Hawkes processes, and more generally for compound Hawkes processes. As an application, we close a gap in the literature by providing a quantitative Central Limit Theorem for the compound Hawkes process.

## Contents

---

|            |   |           |
|------------|---|-----------|
| <b>3.1</b> | <b>Introduction</b>   | <b>61</b> |
| <b>3.2</b> | <b>Notations and preliminaries</b>                              | <b>65</b> |
| 3.2.1      | Elements of stochastic analysis on the Poisson space            | 65        |
| 3.2.2      | Representation of the compound Hawkes process                   | 67        |
| 3.2.3      | Malliavin analysis of the compound Hawkes process               | 69        |
| 3.2.4      | Elements on Stein’s method                                      | 74        |
| <b>3.3</b> | <b>Main results</b>   | <b>75</b> |
| 3.3.1      | The general result  | 75        |
| 3.3.2      | Application to the Hawkes process                               | 78        |
| <b>3.4</b> | <b>Appendix</b>   | <b>87</b> |
| 3.4.1      | General estimates   | 88        |
| 3.4.2      | Specific estimates for the exponential and the Erlang’s kernels | 89        |

---

## 3.1 Introduction

Hawkes processes provide a generalization of the Poisson process to model a sequence of arrivals over time of some type of events, that present self-exciting feature, in the sense that each arrival increases the rate of future arrivals for some period of time. This class of counting processes allows one to capture self-exciting phenomena as opposed to inhomogeneous Poisson processes or Cox processes. Introduced by Hawkes in [46], there was historically a first boom in their application in seismology.

Since then, they have been widely used in many different fields, among which neurosciences [81], social network models, biology [82], population dynamics [22], finance or insurance. In finance, they are used to model for example credit risk contagion [35], order book or the microstructure noise's feature of financial markets [8]. In insurance also, some risks exhibit self-exciting features, as it is the case for cyber risk [13].

Naturally in all these applications, the long term behaviour of the Hawkes process  $N := (N_t)_{t \geq 0}$  (or of functionals of  $N$ ) is of interest. Although we will focus only on (central) limit theorems for Hawkes functionals, we mention that large deviation principles have also been intensively studied in the literature (see [15] for the first work in that direction for linear Hawkes processes with many extensions in the non-linear case, and more recently [93]).

The first Central limit theorem for Hawkes processes is stated in the seminal paper [47]. The analysis of limit theorems for such processes have been further derived, notably in [7]. For the sake of the presentation, we recall that a Hawkes process is a counting process  $N := (N_t)_{t \geq 0}$  with stochastic intensity  $\lambda := (\lambda_t)_{t \geq 0}$  given by

$$\lambda_t = \mu + \int_{(0,t)} \phi(t-s) dN_s,$$

where  $\mu \geq 0$  and  $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  are two given parameters. In the literature, processes starting at minus infinity have been considered, especially in applications to biology. We limit ourselves here to the case where the intensity starts at zero, firstly in order to fit the paradigm set by Bacry *et. al.* in [7], and secondly, because of the applications we have in mind, namely cyber risk and insurance (which will be part of an upcoming work). The methodology we consider could possibly be extended to the case where  $(\lambda_t)_{t \geq 0}$  starts at  $-\infty$ , at the price of a modification of the filtrations we consider. This would change some of the techniques we used, and we chose not to pursue this extension. Under an integrability assumption (which reads as Assumption 3.3.2 below), according to [7, Theorem 2],

$$\left( \frac{N_{Tv} - \int_0^{Tv} \mathbb{E}[\lambda_s] ds}{\sqrt{T}} \right)_{v \in [0,1]} \xrightarrow[T \rightarrow +\infty]{\mathcal{L}-\mathcal{S}} (\tilde{\sigma} W_v)_{v \in [0,1]}, \tag{3.1.1}$$

where  $\mathcal{L}-\mathcal{S}$  stands for the convergence in law as a process in the Skorokhod topology with  $\tilde{\sigma} > 0$  an explicit constant depending only on  $\mu$  and  $\phi$ , and  $W$  a Brownian motion. Note that we consider here a particular case of the results in [7] which hold for multidimensional Hawkes processes. This result is a consequence of the martingale limit theorem that can be found as [7, Lemma 7] (once again we present a particular case of this result)

$$\left( \frac{N_{Tv} - \int_0^{Tv} \lambda_s ds}{\sqrt{T}} \right)_{v \in [0,1]} \xrightarrow[T \rightarrow +\infty]{\mathcal{L}-\mathcal{S}} (\sigma W_v)_{v \in [0,1]}, \tag{3.1.2}$$

with  $\sigma > 0$  an explicit constant depending only on  $\mu$  and  $\phi$  (but still different from  $\tilde{\sigma}$  in (3.1.1)). Naturally, limit theorems (3.1.1) and (3.1.2) come as a second step after a law of large numbers for Hawkes processes has been derived; program which has been performed in [7].

Generalizations of limit theorems (3.1.1) and (3.1.2) have been obtained in [41], [94], [92], [52], [50], [49] for different functionals of the Hawkes process (to mention a few references) in different contexts.

In spite of this large variety of (functional) limit theorems, it appears that a quantification of this convergence is not available in the literature. To be more specific, we are not aware of bounds for the 1-marginal convergence in law (by taking  $v = 1$  in (3.1.2)):

$$\frac{N_T - \int_0^T \lambda_s ds}{\sqrt{T}} \xrightarrow[T \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2). \quad (3.1.3)$$

Several authors (see [76], [11] or [91]) have used the Nourdin-Peccati methodology to provide bounds for the normal approximation of counting processes but no result in the literature allows one to quantify convergence (3.1.3). The closest result is the one of [91]. Even though this reference is not dedicated to the CLT above, and consider more general counting processes than the linear Hawkes process, a general bound is presented (as [91, Theorem 3.1] and more precisely as [91, Relations (5.2) and (5.4)]) for the CLT (3.1.3). Denoting  $d_W$  the Wasserstein distance, these estimates in [91] lead to :

$$d_W \left( \frac{N_T - \int_0^T \lambda_s ds}{\sqrt{T}}; \mathcal{N}(0, \sigma^2) \right) \leq B_T, \quad \text{with } B_T \geq \frac{2\sqrt{2/\pi} \|\phi\|_1 (2 - \|\phi\|_1)}{1 - \|\phi\|_1};$$

which obviously does not converge to 0 as  $T$  tends to  $+\infty$  (we refer to Remark 3.3.8 for a discussion on this matter). In this chapter, we consider a compound Hawkes process  $X := (X_t)_{t \geq 0}$ , of the form

$$X_t = \sum_{i=1}^{N_t} Y_i, \quad t \geq 0,$$

with  $(Y_i)_{i \geq 1}$  a sequence of iid random variables with distribution  $\nu$  (and independent of  $N$ ) and we provide quantitative limit theorems for  $X$  in the spirit of (3.1.2)-(3.1.3). To this end, we adapt the (by now classical) Malliavin-Stein approach (initiated in [66] that we will refer to Nourdin-Peccati's approach) to derive general bounds in the Wasserstein distance between a class of functionals of the compound Hawkes process and a given Gaussian density (see Theorem 4.3.4 and especially Relation (3.3.3)). More precisely, we make use of the Poisson embedding representation of the Hawkes process (initially introduced in [18]) which allows one to write the Hawkes process  $N$  together with its intensity process  $\lambda$  and the compound process  $X$  as the unique solution of an SDE driven by a Poisson measure  $\Pi$  on  $\mathbb{R}_+^2 \times \mathbb{R}$  (see Theorem 4.2.6



below) to derive a Malliavin calculus with respect to the Hawkes process. This completes the analysis initiated for a general Hawkes process in [48], where already a Mecke formula has been obtained. With this material at hand, we can then consider a large class of Hawkes functionals  $F$  (taking the form of a divergence operator with respect to the baseline Poisson measure  $\Pi$ ) for which we adapt the Malliavin-Stein method of [66] (or more precisely in this Poisson context of [73]) to derive bounds between the Wasserstein distance of any functional  $F$  as above and any distribution  $\mathcal{N}(0, \sigma^2)$  with  $\sigma^2 > 0$ .

Our approach includes as a particular case of these functionals, the compound Hawkes process itself (centered) and counterpart of Convergence (3.1.3) for the Hawkes process  $N$  can be quantified in the Wasserstein distance  $d_W$  (see Theorem 3.3.10 and Theorem 3.3.12). For instance, in the famous cases of exponential and Erlang's kernels<sup>1</sup>

$$d_W \left( \frac{X_T - m \int_0^T \lambda_s ds}{\sqrt{T}}, G \right) \leq \frac{C_{\phi, \nu}}{\sqrt{T}}, \quad \forall T > 0, \quad G \sim \mathcal{N}(0, \sigma^2 \vartheta^2),$$

where  $C_{\phi, \nu}$  and  $\sigma^2, \vartheta^2$  are explicit constants and  $m := \int_{\mathbb{R}} x \nu(dx)$  (see Theorem 3.3.12 for a precise statement). As Theorem 3.3.10 suggests, we were not able to prove the general  $T^{-1/2}$  rate for any kernel  $\phi$ . Indeed, the term in the bound requires some specific estimates on some cross-correlations of the intensity process of the given Hawkes process that we were not able to perform in a general framework. These correlation estimates are for sure of interest outside the scope of this chapter. We refer to Section 3.3.2.1 for a review of our results on quantitative counterparts of limit theorems (3.1.2)-(3.1.3) regarding the (compound) Hawkes process.

The chapter is organised as follows. First we derive in Section 3.2 elements of stochastic analysis for the Hawkes process including the Poisson embedding SDE and the definition of a Malliavin calculus for the Hawkes process (following the paper [48]). The main element of the Malliavin-Stein method is also presented in this section. General bounds for Hawkes functionals are presented in Section 4.3 and the bounds associated to generalization for the Central Limit Theorem 3.1.2 are presented in Section 3.3.2, starting with a review of our quantitative limit theorems for the compound Hawkes process in Section 3.3.2.1. Technical lemmata including some results on exponential and Erlang kernels are contained in the Appendix (Section 3.4).

---

1. The exponential kernel refers to  $\phi(t) := \alpha e^{-\beta t}$ , whereas Erlang's kernel refers to  $\phi(t) := \alpha t e^{-\beta t}$  where in both cases  $\alpha, \beta$  are well-chosen positive constants. The Erlang kernel used in this Chapter is a special case of the kernels used in Chapter 2.

## 3.2 Notations and preliminaries

In this section we list all the mathematical framework we will use in our analysis. More precisely, we first recall some general elements of stochastic analysis in Section 3.2.1. Then with this material at hand, we make precise in Section 3.2.2 the representation of the compound Hawkes process as a solution to an SDE with respect to a Poisson random measure. The former representation will turn to be fundamental in our analysis and we provide in Section 3.2.3 the Malliavin derivative of the compound Hawkes process and give an integration by parts formula in Theorem 3.2.19 which completes the Mecke formula obtained in [48]. Finally, we recall the main elements of Stein's method in Section 4.2.5.

For  $E$  a topological space, we set  $\mathcal{B}(E)$  the  $\sigma$ -algebra of Borel sets.

### 3.2.1 Elements of stochastic analysis on the Poisson space

Let  $\nu$  be a Borel measure on  $\mathbb{R}$  with  $\nu(\mathbb{R}) = 1$  and  $\nu(\{0\}) = 0$ .

In this section we model a Hawkes process and a compound Hawkes process using the Poisson embedding representation (or thinning procedure) in the spirit of [48]. To this end, we need three variables for the Poisson measure :  $t$  for the jump times,  $x$  will stand for the size the jump (with distribution  $\nu$ ) and  $\theta$  will play the role of an auxiliary variable required for the representation itself (according to the thinning algorithm).

Let the space of configurations

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

Each path of a counting process is represented as an element  $\omega^\Pi$  in  $\Omega^\Pi$  which is a  $\mathbb{N}$ -valued measure on  $\mathbb{R}_+^2 \times \mathbb{R}$ . Let  $\mathcal{F}^\Pi$  be the  $\sigma$ -field associated to the vague topology on  $\Omega^\Pi$ , and  $\mathbb{P}^\Pi$  the Poisson measure under which the counting process  $\Pi$  defined as :

$$\Pi([0, t] \times [0, b] \times (-\infty, y])(\omega) := \omega([0, t] \times [0, b] \times (-\infty, y]), \quad t \geq 0, (b, y) \in \mathbb{R}_+ \times \mathbb{R},$$

is an homogeneous Poisson process with intensity measure  $dt \otimes d\theta \otimes \nu$ , that is, for any  $(t, b, y) \in [0, T] \times \mathbb{R}_+ \times \mathbb{R}$ ,  $\Pi([0, t] \times [0, b] \times (-\infty, y])$  is a Poisson random variable with intensity  $bt\nu((-\infty, y])$ .

We set  $\mathbb{F}^\Pi := (\mathcal{F}_t^\Pi)_{t \geq 0}$  the natural history of  $\Pi$ , that is  $\mathcal{F}_t^\Pi := \sigma(\Pi(\mathcal{T} \times B))$ ,  $\mathcal{T} \subset \mathcal{B}([0, t])$ ,  $B \in \mathcal{B}(\mathbb{R}_+ \times \mathbb{R})$ . Let also,  $\mathcal{F}_\infty^\Pi := \lim_{t \rightarrow +\infty} \mathcal{F}_t^\Pi = \vee_{t \geq 0} \mathcal{F}_t^\Pi$ . The expectation with respect to  $\mathbb{P}^\Pi$  is denoted by  $\mathbb{E}[\cdot]$ . For  $t \geq 0$ , we denote by  $\mathbb{E}_t[\cdot]$  the conditional expectation  $\mathbb{E}[\cdot | \mathcal{F}_t^\Pi]$ .

We describe some elements of stochastic analysis on the Poisson space, especially the adding point operator, the Malliavin derivative and its dual operator : the divergence.

**Definition 3.2.1** (Adding point operator). We define for  $(t, \theta, x)$  in  $\mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}$  the measurable maps

$$\begin{aligned} \varepsilon_{(t,\theta,x)}^+ : \Omega^\Pi &\rightarrow \Omega^\Pi \\ \omega &\mapsto \varepsilon_{(t,\theta,x)}^+(\omega), \end{aligned}$$

where for any  $A$  in  $\mathcal{B}(\mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R})$

$$(\varepsilon_{(t,\theta,x)}^+(\omega))(A) := \omega(A \setminus (t, \theta, x)) + \mathbf{1}_A(t, \theta, x),$$

with

$$\mathbf{1}_A(t, \theta, x) := \begin{cases} 1, & \text{if } (t, \theta, x) \in A, \\ 0, & \text{else.} \end{cases}$$

*Lemma 3.2.2.* Let  $t \geq 0$  and  $F$  be an  $\mathcal{F}_t^\Pi$ -measurable random variable. Let  $v > t$  and  $(\theta, x) \in \mathbb{R}_+ \times \mathbb{R}$ . It holds that

$$F \circ \varepsilon_{(v,\theta,x)}^+ = F, \quad \mathbb{P} - a.s..$$

**Definition 3.2.3** (Malliavin derivative). For  $F$  in  $L^2(\Omega, \mathcal{F}_\infty^\Pi, \mathbb{P})$ , we define  $DF$  the Malliavin derivative of  $F$  as

$$D_{(t,\theta,x)}F := F \circ \varepsilon_{(t,\theta,x)}^+ - F, \quad (t, \theta, x) \in \mathbb{R}_+^2 \times \mathbb{R}.$$

*Remark 3.2.4.* Our approach follows [75, 77]. In our setting, the measure  $\nu$  induces a Malliavin calculus for the compound Poisson process as a Lévy process. In that realm, the previous Malliavin calculus was extended to general Lévy processes in [85]. There the Malliavin derivative takes the form of :  $D_{(t,\theta,x)}F := \frac{F \circ \varepsilon_{(t,\theta,x)}^+ - F}{x}$  but the intensity measure is not exactly the one we consider here. Hence, here we decided to follow the classical approach (without the  $1/x$  normalisation) with compensator  $\nu$  (we refer to [64] or [77, Section 6.7] for more details) as our baseline process is the Poisson process  $\Pi$  and not the compound process  $\int x \Pi(dt, d\theta, dx)$ .

The following definition is a by-product of [75, Théorème 1] (see also [68]).

**Definition 3.2.5.** Let  $\mathcal{I}$  be the sub-sigma field of  $\mathcal{B}(\mathbb{R}_+^2 \times \mathbb{R}) \otimes \mathcal{F}_\infty^\Pi$  of stochastic processes  $Z := (Z_{(t,\theta,x)})_{(t,\theta,x) \in \mathbb{R}_+^2 \times \mathbb{R}}$  in  $L^1(\Omega \times \mathbb{R}_+^2 \times \mathbb{R}, \mathbb{P} \otimes dt \otimes d\theta \otimes \nu)$  such that

$$D_{(t,\theta,x)}Z_{(t,\theta,x)} = 0, \quad \text{for a.a. } (t, \theta, x) \in \mathbb{R}_+^2 \times \mathbb{R}.$$

*Remark 3.2.6.* Let  $(t_0, \theta_0, x_0) \in \mathbb{R}_+^2 \times \mathbb{R}$ ,  $(s, t) \in \mathbb{R}_+^2$  with  $t_0 < s < t$ . For  $\mathcal{T} \in \{(s, t), (s, t], [s, t), [s, t]\}$  and  $B \in \mathcal{B}(\mathbb{R}_+^2 \times \mathbb{R})$ , we have that

$$N \circ \varepsilon_{(t_0, \theta_0, x_0)}^+(\mathcal{T} \times B) = N(\mathcal{T} \times B).$$

**Definition 3.2.7.** We set  $\mathcal{S}$  the set of stochastic processes  $Z := (Z_{(t,\theta,x)})_{(t,\theta,x) \in \mathbb{R}_+^2 \times \mathbb{R}}$  in  $\mathcal{I}$  such that :

$$\mathbb{E} \left[ \int_{\mathbb{R}_+^2 \times \mathbb{R}} |Z_{(t,\theta,x)}|^2 dt d\theta \nu(dx) \right] + \mathbb{E} \left[ \left( \int_{\mathbb{R}_+^2 \times \mathbb{R}} Z_{(t,\theta,x)} \Pi(dt, d\theta, dx) \right)^2 \right] < +\infty,$$

where  $\int_{\mathbb{R}_+^2 \times \mathbb{R}} Z_{(t,\theta,x)} \Pi(dt, d\theta, dx)$  is understood in the sense of the Stieltjes integral.

For  $Z$  in  $\mathcal{S}$ , we set the divergence operator with respect to  $\Pi$  as

$$\delta^\Pi(Z) := \int_{\mathbb{R}_+^2 \times \mathbb{R}} Z_{(t,\theta,x)} \Pi(dt, d\theta, dx) - \int_{\mathbb{R}_+^2 \times \mathbb{R}} Z_{(t,\theta,x)} dt d\theta \nu(dx). \quad (3.2.1)$$

We conclude this section with the integration by parts formula on the Poisson space (see [75, Remarque 1]).

*Proposition 3.2.8* (See e.g. [75]). Let  $F$  be in  $L^2(\Omega, \mathcal{F}_\infty^\Pi, \mathbb{P})$  and  $Z$  be in  $\mathcal{S}$ . We have that

$$\mathbb{E} [F \delta^\Pi(Z)] = \mathbb{E} \left[ \int_{\mathbb{R}_+^2 \times \mathbb{R}} Z_{(t,\theta,x)} D_{(t,\theta,x)} F dt d\theta \nu(dx) \right]. \quad (3.2.2)$$

### 3.2.2 Representation of the compound Hawkes process

We first recall the definition of a Hawkes process.

**Definition 3.2.9** (Standard Hawkes process, [46]). Let  $\mu > 0$  and  $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  be a non-negative map with  $\|\phi\|_1 := \int_0^{+\infty} \phi(u) du < 1$ . A standard Hawkes process  $N := (N_t)_{t \geq 0}$  with parameters  $\mu$  and  $\phi$  is a counting process such that

- (i)  $N_0 = 0$ ,  $\mathbb{P}$ -a.s.,
- (ii) its  $(\mathbb{F}^\Pi)$ -predictable intensity process is given by

$$\lambda_t := \mu + \int_{(0,t)} \phi(t-s) dN_s, \quad t \geq 0,$$

that is for any  $0 \leq s \leq t$  and  $A \in \mathcal{F}_s^\Pi$ ,

$$\mathbb{E} [\mathbf{1}_A (N_t - N_s)] = \mathbb{E} \left[ \int_{(s,t]} \mathbf{1}_A \lambda_r dr \right].$$

The previous definition can be generalized as follows.

**Definition 3.2.10** (Generalized Hawkes process). Let  $v \geq 0$ ,  $h^v$  be a  $\mathcal{F}_v^\Pi$ -measurable random variable with valued in  $\mathbb{N}$ ,  $\mu^v := (\mu^v(t))_{t \geq v}$  a positive map such that  $\mu^v(t)$  is  $\mathcal{F}_v^\Pi$ -measurable for any  $t \geq v$ , and  $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  be a non-negative map with  $\|\phi\|_1 < 1$ . A Hawkes process on  $[v, +\infty)$  with parameters  $\mu^v$ ,  $h^v$  and  $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is a  $(\mathbb{F}^\Pi)$ -adapted counting process  $N := (N_t)_{t \geq v}$  such that

- (i)  $N_v = h^v$ ,  $\mathbb{P}$ -a.s.,

(ii) its ( $\mathbb{F}^\Pi$ -predictable) intensity process is given by

$$\lambda_t := \mu^v(t) + \int_{(v,t)} \phi(t-s) dN_s, \quad t \geq v,$$

that is for any  $v \leq s \leq t$  and  $A \in \mathcal{F}_s^\Pi$ ,

$$\mathbb{E} [\mathbf{1}_A(N_t - N_s) | \mathcal{F}_v^\Pi] = \mathbb{E} \left[ \int_{(s,t)} \mathbf{1}_A \lambda_r dN \middle| \mathcal{F}_v^\Pi \right].$$

**Definition 3.2.11** (Compound Hawkes process). Let  $\mu > 0$ ,  $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  be a non-negative map with  $\|\phi\|_1 < 1$  and  $(Y_i)_{i \geq 1}$  be iid random variables, with common distribution  $\nu$ , and independent of the Hawkes process  $N$  with parameters  $\mu$  and  $\phi$ . We name compound Hawkes process  $X := (X_t)_{t \geq 0}$  a stochastic process with representation :

$$X_t = \sum_{i=1}^{N_t} Y_i, \quad t \geq 0. \quad (3.2.3)$$

We now represent a Hawkes process (and a compound Hawkes process) as the unique solution to an SDE driven by  $\Pi$ . This representation relies on the "Poisson embedding" (or "Thinning Algorithm" (see *e.g.* [4], [18, 22, 25, 69] and references therein). The next result is an extension of [48, Theorem 3.3].

*Theorem 3.2.12* (See [48]). Let  $\mu > 0$  and  $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  such that  $\|\phi\|_1 < 1$ . The SDE below admits a unique solution  $(X, N, \lambda)$  with  $N$  (resp.  $\lambda$ )  $\mathbb{F}^\Pi$ -adapted (resp.  $\mathbb{F}^\Pi$ -predictable)

$$\begin{cases} X_t = \int_{(0,t] \times \mathbb{R}_+ \times \mathbb{R}} x \mathbf{1}_{\{\theta \leq \lambda_s\}} \Pi(ds, d\theta, dx), & t \geq 0, \\ N_t = \int_{(0,t] \times \mathbb{R}_+ \times \mathbb{R}} \mathbf{1}_{\{\theta \leq \lambda_s\}} \Pi(ds, d\theta, dx), & t \geq 0, \\ \lambda_t = \mu + \int_{(0,t)} \phi(t-u) dN_u, & t \geq 0. \end{cases} \quad (3.2.4)$$

According to [48, Theorem 3.3], uniqueness holds pathwise. In addition,  $(N, \lambda)$  is a Hawkes process in the sense of Definition 3.2.9. We set  $\mathbb{F}^N := (\mathcal{F}_t^N)_{t \geq 0}$  (respectively  $\mathbb{F}^X := (\mathcal{F}_t^X)_{t \geq 0}$ ) the natural filtration of  $N$  (respectively of  $X$ ) and  $\mathcal{F}_\infty^N := \lim_{t \rightarrow +\infty} \mathcal{F}_t^N$  (respectively  $\mathcal{F}_\infty^X := \lim_{t \rightarrow +\infty} \mathcal{F}_t^X$ ). Obviously  $\mathcal{F}_t^N \subset \mathcal{F}_t^X \subset \mathcal{F}_t^\Pi$  as  $N$  is completely determined by the jump times of  $N$  which are exactly those of  $X$ .

Finally,  $X$  is a compound Hawkes process in the sense of Definition 3.2.11.

*Remark 3.2.13.* We have decided to include in our analysis the case of compound Hawkes processes and not simply the one of Hawkes processes. By choosing  $\nu(dx) = \delta_1(dx)$  (the Dirac measure concentrated at  $x = 1$ ) one obviously recover the construction of the Hawkes process as in [48] and in that case one can just consider  $\Pi$  to be a Poisson measure on  $\mathbb{R}_+^2$ .

### 3.2.3 Malliavin analysis of the compound Hawkes process

We now wish to describe the impact of the Malliavin derivative on the Hawkes process. Once again this material relies on the one provided in [48]. The objective of this section is to derive an integration by parts formula for the Hawkes functionals as Theorem 3.2.19.

*Lemma 3.2.14.* Let  $t$  and  $v$  in  $\mathbb{R}_+$ ,  $(\theta, x)$  and  $(\theta_0, x_0)$  in  $\mathbb{R}_+ \times \mathbb{R}$ , it holds that :

$$\begin{aligned} & \mathbf{1}_{\{\theta \leq \lambda_t\}} \mathbf{1}_{\{\theta_0 \leq \lambda_t\}} (X_v \circ \varepsilon_{(t,\theta,x)}^+, N_v \circ \varepsilon_{(t,\theta,x)}^+, \lambda_v \circ \varepsilon_{(t,\theta,x)}^+)_{v \geq 0} \\ &= \mathbf{1}_{\{\theta \leq \lambda_t\}} \mathbf{1}_{\{\theta_0 \leq \lambda_t\}} (X_v \circ \varepsilon_{(t,\theta_0,x)}^+, N_v \circ \varepsilon_{(t,\theta_0,x)}^+, \lambda_v \circ \varepsilon_{(t,\theta_0,x_0)}^+)_{v \geq 0}. \end{aligned}$$

In other words, as long as the two marks  $\theta$  and  $\theta_0$  remain below the current intensity, the corresponding added-point Hawkes processes coincide independently of the values of the two marks. Moreover, the adding point operator has the exact same effect on  $N$  and  $\lambda$  regardless of the value of the  $x$ -component. However, this  $x$ -component matters for the compound process  $X$ .

*Proof.* The proof relies on the definition of the operator  $\varepsilon_{(t,\theta,x)}^+$  and on the structure of the Hawkes process. Indeed, the intensity is impacted (on  $(t, +\infty)$ ) by the addition of a jump to  $N$  at time  $t$ , but the value of this jump for  $N$  is equal to 1, and this for any  $\theta$  such that  $\theta \leq \lambda_t$ . In other words : for  $v \geq t$  on  $\{\theta_0 \leq \lambda_t\}$

$$\begin{aligned} N_v \circ \varepsilon_{(t,\theta_0,x_0)}^+ &= N_{t-} + \left( \int_{[t,v] \times \mathbb{R}_+ \times \mathbb{R}} \mathbf{1}_{\{\theta \leq \lambda_u\}} \Pi(d\theta, du, dy) \right) \circ \varepsilon_{(t,\theta_0,x_0)}^+ \\ &= N_{t-} + \mathbf{1}_{\{\theta_0 \leq \lambda_t\}} + \left( \int_{(t,v] \times \mathbb{R}_+ \times \mathbb{R}} \mathbf{1}_{\{\theta \leq \lambda_u\}} \Pi(d\theta, du, dy) \right) \circ \varepsilon_{(t,\theta_0,x_0)}^+ \\ &= N_{t-} + \mathbf{1}_{\{\theta_0 \leq \lambda_t\}} + \int_{(t,v] \times \mathbb{R}} \int_{\mathbb{R}_+} \mathbf{1}_{\{\theta \leq \lambda_u \circ \varepsilon_{(t,\theta_0,x_0)}^+\}} \Pi(d\theta, du, dy), \end{aligned}$$

and  $N_v \circ \varepsilon_{(t,\theta_0,x_0)}^+ = N_v$  for any  $v < t$ . Note that it is part of the definition of the adding point operator (see Definition 4.2.2) to remove the possible natural jump at time  $t$ , which explain why we move from the integral  $\int_{[t,v]}$  to  $\int_{(t,v]}$  in the computations above. Similarly, for  $v \geq t$

$$\begin{aligned} X_v \circ \varepsilon_{(t,\theta_0,x_0)}^+ &= X_{t-} + \left( \int_{[t,v] \times \mathbb{R}_+ \times \mathbb{R}} y \mathbf{1}_{\{\theta \leq \lambda_u\}} \Pi(d\theta, du, dy) \right) \circ \varepsilon_{(t,\theta_0,x_0)}^+ \\ &= X_{t-} + x_0 \mathbf{1}_{\{\theta_0 \leq \lambda_t\}} + \int_{(t,v] \times \mathbb{R}} \int_{\mathbb{R}_+} y \mathbf{1}_{\{\theta \leq \lambda_u \circ \varepsilon_{(t,\theta_0,x_0)}^+\}} \Pi(d\theta, du, dy), \end{aligned}$$

and  $X_v \circ \varepsilon_{(t,\theta_0,x_0)}^+ = X_v$  for any  $v < t$ . In a similar fashion,  $\lambda_v \circ \varepsilon_{(t,\theta_0,x_0)}^+ = \lambda_v$  for any  $v \leq t$  and for  $v > t$ ,

$$\begin{aligned} \lambda_v \circ \varepsilon_{(t,\theta_0,x_0)}^+ &= \left( \mu + \int_{(0,t)} \phi(v-u) dN_u + \int_{[t,v]} \phi(v-u) dN_u \right) \circ \varepsilon_{(t,\theta_0,x_0)}^+ \\ &= \mu + \int_{(0,t)} \phi(v-u) dN_u + \phi(v-t) + \int_{(t,v)} \phi(v-u) d(N_u \circ \varepsilon_{(t,\theta_0,x_0)}^+). \end{aligned}$$

In other words,  $(X \circ \varepsilon_{(v, \theta_0, x_0)}^+, N \circ \varepsilon_{(v, \theta_0, x_0)}^+, \lambda \circ \varepsilon_{(v, \theta_0, x_0)}^+)$  solves the same equation for any  $\theta_0$  such that  $\theta_0 \leq \lambda_t$  (pathwise, up to a modification, and in the SDE sense).  $\square$

*Proposition 3.2.15.* Let  $F$  be a  $\mathcal{F}_\infty^X$ -measurable random variable. Then for any  $t \geq 0$ , and  $\forall \theta, \theta_0 \geq 0, \forall x \in \mathbb{R}$ ,

$$\mathbf{1}_{\{\theta \leq \lambda_t\}} \mathbf{1}_{\{\theta_0 \leq \lambda_t\}} D_{(t, \theta, x)} F = \mathbf{1}_{\{\theta \leq \lambda_t\}} \mathbf{1}_{\{\theta_0 \leq \lambda_t\}} D_{(t, \theta_0, x)} F, \quad \mathbb{P} - a.s..$$

*Proof.* This follows from Lemma 3.2.14 and the fact that any  $\mathcal{F}_T^X$ -measurable random variable  $F$  is a limit of a random variable of the form  $\varphi((X_{t_1}, N_{t_1}, \lambda_{t_1}), \dots, (X_{t_n}, N_{t_n}, \lambda_{t_n}))$ , with  $0 \leq t_1 < \dots < t_n \leq T, n \geq 1$  and  $\varphi$  a Borelian map from  $\mathbb{R}^{3n}$  to  $\mathbb{R}$ . More precisely,  $N$  is itself a functional of  $X$  as it is a counting process whose jumps coincide with those of  $X$ . But the fact of distinguishing  $X$  and  $N$  allows one to note that the Malliavin derivative of a functional of  $N$  does not depend on the  $x$  variable for instance.  $\square$

This motivates the introduction of the following notation.

**Definition 3.2.16.** Let  $t \geq 0, x \in \mathbb{R}$ . For a  $\mathcal{F}_\infty^X$ -measurable random variable  $F$ , we set

$$D_{(t, \lambda_t, x)} F := \mathbf{1}_{\{\theta = \lambda_t\}} D_{(t, \theta, x)} F.$$

In particular, for any  $(\theta, x) \in \mathbb{R}_+ \times \mathbb{R}$ , on the set  $\{\theta \leq \lambda_t\}$ , Proposition 3.2.15 reads as

$$D_{(t, \lambda_t, x)} F = D_{(t, \theta, x)} F.$$

As a consequence, if  $(F_s)_{s \geq 0}$  is a  $\mathbb{F}^X$ -measurable (resp. predictable) process, then  $D_{(t, \lambda_t, x)} F_s = 0$  for  $s < t$  (resp. for  $s \leq t$ ) since  $D_{(t, \lambda_t, x)} F_s = F_s \circ \varepsilon_{(t, \lambda_t, x)}^+ - F_s$  using Lemma 3.2.2. With this notation at hand, we determine the Malliavin derivative of the compound Hawkes process.

*Proposition 3.2.17.* Let  $t \geq 0$  and  $x \in \mathbb{R}$ . We have

$$(D_{(t, \lambda_t, x)} X_s, D_{(t, \lambda_t, x)} N_s, D_{(t, \lambda_t, x)} \lambda_s) = \begin{cases} (x + \hat{X}_s^t, 1 + \hat{N}_s^t, \hat{\lambda}_s^t), & s \geq t, \\ (0, 0, 0), & s < t \end{cases}$$

where the equality is understood pathwise and in the SDE sense and where  $(\hat{X}_s^t, \hat{N}_s^t, \hat{\lambda}_s^t)_{s \geq t}$  is the unique solution to the SDE

$$\begin{cases} \hat{N}_s^t = \int_{(t, s] \times \mathbb{R}_+ \times \mathbb{R}} \mathbf{1}_{\{\lambda_u < \theta \leq \lambda_u + \hat{\lambda}_u^t\}} \Pi(du, d\theta, dy), & s \geq t, \\ \hat{X}_s^t = \int_{(t, s] \times \mathbb{R}_+ \times \mathbb{R}} y \mathbf{1}_{\{\lambda_u < \theta \leq \lambda_u + \hat{\lambda}_u^t\}} \Pi(du, d\theta, dy), & s \geq t, \\ \hat{\lambda}_s^t = \phi(s - t) + \int_{(t, s)} \phi(s - u) d\hat{N}_u^t, & s > t, \hat{\lambda}_t^t = 0. \end{cases} \quad (3.2.5)$$

In addition,  $(\hat{N}_s^t, \hat{\lambda}_s^t)_{s \in [t, +\infty)}$  is a generalized Hawkes process, with initial intensity that is not bounded away from 0.

*Proof.* We follow the proof of Lemma 3.2.14 to apply the adding point operator to SDEs (4.2.3). Fix  $t \geq 0$  and  $x \in \mathbb{R}$ . According to Definition 3.2.16,  $(D_{(t,\lambda_t,x)}X, D_{(t,\lambda_t,x)}N, D_{(t,\lambda_t,x)}\lambda)$  is given as  $D_{(t,\lambda_t,x)}X_s = X_s \circ \varepsilon_{(t,\lambda_t,x)}^+ - X_s$ ,  $D_{(t,\lambda_t,x)}N_s = N_s \circ \varepsilon_{(t,\lambda_t,x)}^+ - N_s$ ,  $D_{(t,\lambda_t,x)}\lambda_s = \lambda_s \circ \varepsilon_{(t,\lambda_t,x)}^+ - \lambda_s$ . Besides,  $D_{(t,\lambda_t,x)}X_s = D_{(t,\lambda_t,x)}N_s = 0$  for  $s < t$  and  $D_{(t,\lambda_t,x)}\lambda_s = 0$  for  $s \leq t$ .

We start by proving that  $D_{(t,\lambda_t,x)}\lambda_s \geq 0$ , for any  $s \geq t$ ,  $\mathbb{P}$ -a.s.. From Relation (4.2.3), the intensity  $\lambda$  is solution to the SDE :

$$\lambda_s = \mu + \int_{(0,s) \times \mathbb{R}_+ \times \mathbb{R}} \phi(s-u) \mathbf{1}_{\{\theta \leq \lambda_u\}} \Pi(du, d\theta, dx);$$

which we rewrite as

$$\lambda_s = \mu_{s,t} + \int_{(t,s) \times \mathbb{R}_+ \times \mathbb{R}} \phi(s-u) \mathbf{1}_{\{\theta \leq \lambda_u\}} \Pi(du, d\theta, dx),$$

with  $\mu_{s,t} := \mu + \int_{(0,t) \times \mathbb{R}_+ \times \mathbb{R}} \phi(s-u) \mathbf{1}_{\{\theta \leq \lambda_u\}} \Pi(du, d\theta, dx)$ . Thus

$$\lambda_s \circ \varepsilon_{(t,\lambda_t,x)}^+ = \mu_{s,t} + \phi(s-t) + \int_{(t,s) \times \mathbb{R}_+ \times \mathbb{R}} \phi(s-u) \mathbf{1}_{\{\theta \leq \lambda_u \circ \varepsilon_{(t,\lambda_t,x)}^+\}} \Pi(du, d\theta, dx),$$

which in turn implies that  $\lambda \circ \varepsilon_{(t,\lambda_t,x)}^+$  and  $\lambda$  coincide on  $(0, t)$  and  $\lambda_t \circ \varepsilon_{(t,\lambda_t,x)}^+ - \lambda_t = \phi(0) \geq 0$ . We let  $\tau_0 := t$  and for  $n \geq 1$

$$\tau_n := \inf\{u > \tau_{n-1}, N_u \circ \varepsilon_{(t,\lambda_t,x)}^+ - N_{u-} \circ \varepsilon_{(t,\lambda_t,x)}^+ \neq 0\} \wedge \inf\{u > \tau_{n-1}, N_u - N_{u-} \neq 0\}.$$

So  $\tau_1$  is the first time after  $t$  where  $N \circ \varepsilon_{(t,\lambda_t,x)}^+$  or  $N$  has a jump. Note that by definition on  $[t, \tau_1)$ ,  $\lambda \circ \varepsilon_{(t,\lambda_t,x)}^+ - \lambda \geq 0$ , thus if  $N$  has a jump at time  $\tau_1$ , so has  $N \circ \varepsilon_{(t,\lambda_t,x)}^+$ ; but the converse is not true. This means that  $\lambda \circ \varepsilon_{(t,\lambda_t,x)}^+ - \lambda \geq 0$  on  $[\tau_1, \tau_2)$ . By induction this proves that  $\lambda \circ \varepsilon_{(t,\lambda_t,x)}^+ - \lambda \geq 0$  on  $[0, \tau_n)$  for any  $n \geq 1$ . As by definition  $\lambda \geq \mu$  on  $\mathbb{R}_+$ , any jump of  $N$  is a jump of the homogeneous Poisson process with intensity  $\mu$  (built by thinning from  $\Pi$ ) and thus that  $\lim_{n \rightarrow +\infty} \tau_n = +\infty$ ,  $\mathbb{P}$ -a.s. which gives

$$\lambda_s \circ \varepsilon_{(t,\lambda_t,x)}^+ - \lambda_s \geq 0, \quad \forall s \geq 0, \mathbb{P} - a.s..$$

With this result at hand we are able to write that

$$\mathbf{1}_{\{\theta \leq \lambda_u \circ \varepsilon_{(t,\lambda_t,x)}^+\}} - \mathbf{1}_{\{\theta \leq \lambda_u\}} = \mathbf{1}_{\{\lambda_u < \theta \leq \lambda_u \circ \varepsilon_{(t,\lambda_t,x)}^+\}},$$

which is of interest for describing the Malliavin derivatives of  $N$  and  $X$ . Let  $s \geq t$ , according to Lemma 3.2.14 and the previous claim we have that

$$\begin{aligned} & D_{(t,\lambda_t,x)}N_s \\ &= N_{t-} + 1 + \int_{(t,s) \times \mathbb{R}_+ \times \mathbb{R}} \mathbf{1}_{\{\theta \leq \lambda_u \circ \varepsilon_{(t,\lambda_t,x)}^+\}} \Pi(d\theta, du, dy) - \left( N_t + \int_{(t,s) \times \mathbb{R}_+ \times \mathbb{R}} \mathbf{1}_{\{\theta \leq \lambda_u\}} \Pi(d\theta, du, dy) \right) \\ &= -(N_t - N_{t-}) + 1 + \int_{(t,s) \times \mathbb{R}_+ \times \mathbb{R}} \mathbf{1}_{\{\lambda_u \leq \theta \leq \lambda_u \circ \varepsilon_{(t,\lambda_t,x)}^+\}} \Pi(d\theta, du, dy) \\ &= 1 + \int_{(t,s) \times \mathbb{R}_+ \times \mathbb{R}} \mathbf{1}_{\{\lambda_u \leq \theta \leq \lambda_u \circ \varepsilon_{(t,\lambda_t,x)}^+\}} \Pi(d\theta, du, dy), \quad \mathbb{P} - a.s.. \end{aligned}$$



$$\begin{aligned}
& D_{(t,\lambda_t,x)} X_s \\
&= X_{t-} + x + \int_{(t,s] \times \mathbb{R}_+ \times \mathbb{R}} \mathbf{1}_{\{\theta \leq \lambda_u \circ \varepsilon_{(t,\lambda_t,x)}^+\}} y \Pi(d\theta, du, dy) - \left( X_t + \int_{(t,s] \times \mathbb{R}_+ \times \mathbb{R}} \mathbf{1}_{\{\theta \leq \lambda_u\}} y \Pi(d\theta, du, dy) \right) \\
&= -(X_t - X_{t-}) + x + \int_{(t,s] \times \mathbb{R}_+ \times \mathbb{R}} \mathbf{1}_{\{\lambda_u \leq \theta \leq \lambda_u \circ \varepsilon_{(t,\lambda_t,x)}^+\}} y \Pi(d\theta, du, dy) \\
&= x + \int_{(t,s] \times \mathbb{R}_+ \times \mathbb{R}} \mathbf{1}_{\{\lambda_u \leq \theta \leq \lambda_u \circ \varepsilon_{(t,\lambda_t,x)}^+\}} y \Pi(d\theta, du, dy), \quad \mathbb{P} - a.s..
\end{aligned}$$

Similarly, for  $s > t$ , we have that

$$\begin{aligned}
D_{(t,\lambda_t,x)} \lambda_s &= \phi(s-t) + \int_{(t,s)} \phi(s-u) d(N_u \circ \varepsilon_{(t,\lambda_t,x)}^+ - dN_u) \\
&= \phi(s-t) + \int_{(t,s)} \phi(s-u) d(D_{(t,\lambda_t,x)} N_u).
\end{aligned}$$

For  $s = t$ ,  $D_{(t,\lambda_t,x)} \lambda_s = 0$  as  $\lambda$  is  $\mathbb{F}^N$ -predictable. Writing  $\lambda_u \circ \varepsilon_{(t,\lambda_t,x)}^+ = \lambda_u + D_{(t,\lambda_t,x)} \lambda_u$  proves that  $(D_{(t,\lambda_t,x)} X, D_{(t,\lambda_t,x)} N, D_{(t,\lambda_t,x)} \lambda)$  is solution to SDE (4.2.4) which admits a unique solution following [48, Theorem 3.3].

The last claim follows by proving that  $(\hat{N}^t, \hat{\lambda}^t)$  is a Hawkes process on  $[t, +\infty)$ . Indeed, for  $t \leq s_1 \leq s_2$ , we have that :

$$\begin{aligned}
\mathbb{E} [\hat{N}_{s_2}^t - \hat{N}_{s_1}^t | \mathcal{F}_{s_1}^N] &= \mathbb{E} \left[ \mathbb{E} [\hat{N}_{s_2}^t - \hat{N}_{s_1}^t | \mathcal{F}_{s_1}^\Pi] | \mathcal{F}_{s_1}^N \right] \\
&= \mathbb{E} \left[ \int_{s_1}^{s_2} \mathbb{E} \left[ \int_{\mathbb{R}_+} \mathbf{1}_{\{\lambda_u \leq \theta \leq \lambda_u + \hat{\lambda}_u^t\}} d\theta | \mathcal{F}_{s_1}^\Pi \right] dt | \mathcal{F}_{s_1}^N \right] \\
&= \int_{s_1}^{s_2} \mathbb{E} [\hat{\lambda}_u^t | \mathcal{F}_{s_1}^N] dt.
\end{aligned}$$

Note that the intensity  $(\hat{\lambda}_s^t)_{s \in [t, +\infty)}$  can reach zero.  $\square$

*Remark 3.2.18.* In a similar fashion to Proposition 3.2.15, as a byproduct of the previous proposition, we get that the  $x$ -component of the mark does not affect the counting process. More precisely, for any functional  $F$  of  $(N, \lambda)$ , we have  $\forall t \geq 0, \forall \theta, \theta_0 \geq 0, \forall x, y \in \mathbb{R}$ ,

$$\mathbf{1}_{\{\theta \leq \lambda_t\}} \mathbf{1}_{\{\theta_0 \leq \lambda_t\}} D_{(t,\theta,x)} F = \mathbf{1}_{\{\theta \leq \lambda_t\}} \mathbf{1}_{\{\theta_0 \leq \lambda_t\}} D_{(t,\theta_0,y)} F, \quad \mathbb{P} - a.s..$$

In other words, in that case, the Malliavin derivative does not depend on the variable  $x$ .

We conclude this section by re-writing the integration by parts formula (3.2.2) for the Hawkes process.

*Theorem 3.2.19.* Set  $\mathcal{Z} := (\mathcal{Z}_{(t,\theta)})_{(t,\theta) \in \mathbb{R}_+^2}$  the stochastic process defined as

$$\mathcal{Z}_{(t,\theta)} := \mathbf{1}_{\{\theta \leq \lambda_t\}}, \quad (t, \theta) \in \mathbb{R}_+^2.$$

Let  $Z := (Z_{(t,x)})_{(t,x) \in \mathbb{R}_+ \times \mathbb{R}}$  be a  $\mathbb{F}^X$ -predictable process satisfying

$$\mathbb{E} \left[ \int_{\mathbb{R}_+ \times \mathbb{R}} |Z_{(t,x)}|^2 \lambda_t dt \nu(dx) + \left( \int_{\mathbb{R}_+ \times \mathbb{R}} Z_{(t,x)} \lambda_t dt \nu(dx) \right)^2 \right] < \infty.$$

It holds that

- (i)  $Z\mathcal{Z} = (Z_{(t,x)} \mathbf{1}_{\{\theta \leq \lambda_t\}})_{(t,\theta,x) \in \mathbb{R}_+^2 \times \mathbb{R}}$  belongs to  $\mathcal{S}$  (recall Definition 3.2.7).
- (ii) For any  $\mathcal{F}_\infty^X$ -measurable random variable  $F$  with  $\mathbb{E}[|F|^2] < +\infty$ ,

$$\mathbb{E} [F \delta^\Pi(Z \mathbf{1}_{\{\theta \leq \lambda_t\}})] = \mathbb{E} \left[ \int_{\mathbb{R}_+ \times \mathbb{R}} \lambda_t Z_{(t,x)} D_{(t,\lambda_t,x)} F dt \nu(dx) \right], \quad (3.2.6)$$

where  $D_{(t,\lambda_t,x)} F = D_{(t,\theta,x)} F \mathbf{1}_{\{\theta \leq \lambda_t\}}$  for any  $\theta \geq 0$ ,  $x \in \mathbb{R}$ .

*Proof.* By construction,  $\lambda$  is  $\mathbb{F}^N$ -predictable. Hence, for any  $(t, \theta, x)$  in  $\mathbb{R}_+^2 \times \mathbb{R}$ ,  $D_{(t,\theta,x)} \mathbf{1}_{\{\theta \leq \lambda_t\}} = 0$ . So,  $\mathcal{Z}$  belongs to  $\mathcal{I}$ . In addition, as  $Z$  and  $\lambda$  are predictable, our assumptions imply that

$$\begin{aligned} & \mathbb{E} \left[ \left( \int_{\mathbb{R}_+^2 \times \mathbb{R}} Z_{(t,x)} \mathcal{Z}_{(t,\theta)} \Pi(dt, d\theta, dx) \right)^2 \right] \\ & \leq 2 \left( \mathbb{E} \left[ \left( \int_{\mathbb{R}_+^2 \times \mathbb{R}} Z_{(t,x)} \mathcal{Z}_{(t,\theta)} (\Pi(dt, d\theta, dx) - dt d\theta \nu(dx)) \right)^2 \right] \right. \\ & \quad \left. + \mathbb{E} \left[ \left( \int_{\mathbb{R}_+^2 \times \mathbb{R}} Z_{(t,x)} \mathcal{Z}_{(t,\theta)} dt d\theta \nu(dx) \right)^2 \right] \right) \\ & = 2 \left( \mathbb{E} \left[ \int_{\mathbb{R}_+^2 \times \mathbb{R}} |Z_{(t,x)} \mathcal{Z}_{(t,\theta)}|^2 dt d\theta \nu(dx) \right] + \mathbb{E} \left[ \left( \int_{\mathbb{R}_+^2 \times \mathbb{R}} Z_{(t,x)} \lambda_t dt \nu(dx) \right)^2 \right] \right) \\ & = 2 \left( \mathbb{E} \left[ \int_{\mathbb{R}_+ \times \mathbb{R}} |Z_{(t,x)}|^2 \lambda_t dt \nu(dx) \right] + \mathbb{E} \left[ \left( \int_{\mathbb{R}_+ \times \mathbb{R}} Z_{(t,x)} \lambda_t dt \nu(dx) \right)^2 \right] \right) < +\infty \end{aligned}$$

and

$$\mathbb{E} \left[ \int_0^{+\infty} \int_{\mathbb{R}_+ \times \mathbb{R}} |Z_{(t,x)} \mathcal{Z}_{(t,\theta)}|^2 d\theta dt \nu(dx) \right] = \mathbb{E} \left[ \int_0^{+\infty} \int_{\mathbb{R}} |Z_{(t,x)}|^2 \lambda_t dt \nu(dx) \right] < +\infty.$$

This proves (i). In particular,  $\delta^\Pi(Z\mathcal{Z})$  is well-defined and the integration by parts

formula (3.2.2) is in force. It gives that

$$\begin{aligned} \mathbb{E} [F\delta^\Pi(Z\mathcal{Z})] &= \mathbb{E} \left[ \int_{\mathbb{R}_+^2 \times \mathbb{R}} Z_{(t,x)} \mathbf{1}_{\{\theta \leq \lambda_t\}} D_{(t,\theta,x)} F d\theta dt \nu(dx) \right] \\ &= \mathbb{E} \left[ \int_0^{+\infty} \int_{\mathbb{R}} Z_{(t,x)} \int_{\mathbb{R}_+} \mathbf{1}_{\{\theta \leq \lambda_t\}} D_{(t,\theta,x)} F d\theta \nu(dx) dt \right] \\ &= \mathbb{E} \left[ \int_0^{+\infty} \int_{\mathbb{R}} Z_{(t,x)} \int_0^{\lambda_t} D_{(t,\lambda_t,x)} F d\theta \nu(dx) dt \right] \\ &= \mathbb{E} \left[ \int_0^{+\infty} \lambda_t \int_{\mathbb{R}} Z_{(t,x)} D_{(t,\lambda_t,x)} F \nu(dx) dt \right]. \end{aligned}$$

□

### 3.2.4 Elements on Stein’s method

Though Stein’s method has been introduced by C. M. Stein in [87], the combination of the Malliavin calculus with Stein’s method (and known as the Nourdin-Peccati’s approach) has been initiated in [66] for the approximation of Gaussian functionals, and subsequently extended in [73] for Poisson functionals (which is closer to this chapter). We introduce in this section Stein’s original approach which allows one to derive inequality (3.2.7) below. Then, in the proof of Theorem 4.3.4, we will adapt [73] to transform the right-hand side of (3.2.7) into the right-hand side of (3.3.3) using the Malliavin calculus. We refer to [65] for a complete exposition of the original Stein method and of the Nourdin-Peccati approach.

**Definition 3.2.20.** Let  $F$  and  $G$  two random variables defined on some  $(\Omega, \mathcal{F}_\infty^\Pi, \mathbb{P})$ . We recall that the Wasserstein distance between the distributions  $\mathcal{L}_F$  and  $\mathcal{L}_G$  (or simply between  $F$  and  $G$ ) as :

$$d_W(F, G) := \sup_{h \in \text{Lip}} \left| \mathbb{E}[h(F)] - \mathbb{E}[h(G)] \right|,$$

with  $\text{Lip} := \{h : \mathbb{R} \rightarrow \mathbb{R} \text{ differentiable a.e. with } \|h'\|_\infty \leq 1\}$ .

We assume that  $F$  is centered. Let  $G \sim \mathcal{N}(0, \sigma^2)$ . We set

$$\mathfrak{F}_W^0 := \{f : \mathbb{R} \rightarrow \mathbb{R}, \text{ twice differentiable with } \|f'\|_\infty \leq 1, \|f''\|_\infty \leq 2, f(0) = 0\}.$$

Consider  $h$  in  $\text{Lip}$ . C. M. Stein proved in [87], that there exists a function  $f_h$  in  $\mathfrak{F}_W^0$  solution to the functional equation (named Stein’s equation) :

$$h(x) - \mathbb{E}[h(G)] = \sigma^2 f'_h(x) - x f_h(x), \quad \forall x \in \mathbb{R}.$$

Plugging  $F$  in this equation and taking the expectation, we get that :

$$|\mathbb{E}[h(F)] - \mathbb{E}[h(G)]| = |\mathbb{E}[\sigma^2 f'_h(F) - F f_h(F)]|.$$

Hence,

$$d_W(F, G) \leq \sup_{f \in \mathfrak{F}_W^0} |\mathbb{E}[\sigma^2 f'(F) - F f(F)]|. \tag{3.2.7}$$

In addition, the right hand side is equal to 0 if and only if  $F \sim \mathcal{N}(0, \sigma^2)$ .

*Remark 3.2.21.* The original result proven by C. M. Stein in [87] was the existence of a function  $f_h$  in

$$\mathfrak{F}_W := \{f : \mathbb{R} \rightarrow \mathbb{R}, \text{ twice differentiable with } \|f'\|_\infty \leq 1, \|f''\|_\infty \leq 2\}.$$

The sup on  $\mathfrak{F}_W^0$  in (3.2.7) coincides with a sup on  $\mathfrak{F}_W$  if  $F$  is centered. To see that, it is enough to replace  $f \in \mathfrak{F}_W$  by  $f - f(0)$  and use the fact that  $\mathbb{E}[f(0)F] = 0$ .

### 3.3 Main results

We present in Section 4.3 a general bound on the Wasserstein distance between a given Hawkes functional and a centered Gaussian distribution. This bound is then applied in Section 3.3.2 to provide bounds for CLTs for compound Hawkes processes. We refer to Section 3.3.2.1 for a review of our results on quantitative counterparts of limit theorems (3.1.2)-(3.1.3) regarding the compound Hawkes process.

#### 3.3.1 The general result

Recall that we consider the compound Hawkes process given as the unique solution to SDE(4.2.3)

$$\begin{cases} X_t = \int_{(0,t] \times \mathbb{R}_+ \times \mathbb{R}} x \mathbf{1}_{\{\theta \leq \lambda_s\}} \Pi(ds, d\theta, dx), & t \geq 0 \\ N_t = \int_{(0,t] \times \mathbb{R}_+ \times \mathbb{R}} \mathbf{1}_{\{\theta \leq \lambda_s\}} \Pi(ds, d\theta, dx), & t \geq 0 \\ \lambda_t = \mu + \int_{(0,t)} \phi(t-u) dN_u, & t \geq 0. \end{cases}$$

**Assumption 3.3.1.** Throughout this chapter we recall that  $\nu$  is a probability measure on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  with  $\nu(\{0\}) = 0$ . In addition, we assume that :

$$m := \int_{\mathbb{R}} x \nu(dx) < +\infty, \quad \vartheta^2 := \int_{\mathbb{R}} x^2 \nu(dx) < +\infty. \quad (3.3.1)$$

We assume the following standard stability condition, that is crucial to obtain classical ergodic properties for the Hawkes process.

**Assumption 3.3.2.** The mapping  $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is such that

$$\|\phi\|_1 = \int_0^{+\infty} \phi(u) du < 1.$$

This assumption of subcritical regime is fundamental for our analysis. Under this assumption, one can define

$$\chi := \sum_{n \geq 1} \phi^{(*n)}, \quad (3.3.2)$$

with  $\phi^{(*n)}$  the  $n$ -th convolution of  $\phi$  with itself. We have that

$$\int_0^{+\infty} \chi(t) dt = \int_0^{+\infty} \sum_{n \geq 1} \phi^{(*n)}(t) dt = \sum_{n \geq 1} \int_0^{+\infty} \phi^{(*n)}(t) dt = \sum_{n \geq 1} \|\phi\|_1^n = \frac{\|\phi\|_1}{1 - \|\phi\|_1}.$$

**Assumption 3.3.3.** The mapping  $\phi$  is such that

$$\int_0^{+\infty} u\phi(u)du < +\infty.$$

*Theorem 3.3.4.* Consider

(i)  $Z := (Z_{(t,x)})_{(t,x) \in \mathbb{R}_+ \times \mathbb{R}}$  a  $\mathbb{F}^X$ -predictable stochastic process such that

$$\mathbb{E} \left[ \int_{\mathbb{R}_+ \times \mathbb{R}} |Z_{(t,x)}|^2 \lambda_t dt\nu(dx) + \left( \int_{\mathbb{R}_+ \times \mathbb{R}} Z_{(t,x)} \lambda_t dt\nu(dx) \right)^2 \right] < \infty.$$

(ii)  $\mathcal{Z} := (\mathcal{Z}_{(t,\theta)})_{(t,\theta) \in \mathbb{R}_+^2}$  the stochastic process defined as

$$\mathcal{Z}_{(t,\theta)} = \mathbf{1}_{\{\theta \leq \lambda_t\}}, \quad (t, \theta) \in \mathbb{R}_+^2.$$

(iii)  $F := \delta^\Pi(Z\mathcal{Z}) = \int_{\mathbb{R}_+^2 \times \mathbb{R}} Z_{(t,x)} \mathcal{Z}_{(t,\theta)} \Pi(dt, d\theta, dx) - \int_{\mathbb{R}_+ \times \mathbb{R}} Z_{(t,x)} \lambda_t dt\nu(dx)$ .

Let  $\gamma > 0$ . Then, letting  $G \sim \mathcal{N}(0, \gamma^2)$ ,

$$d_W(F, G) \leq \mathbb{E} \left[ \left| \gamma^2 - \int_{\mathbb{R}_+ \times \mathbb{R}} Z_{(t,x)} \lambda_t D_{(t,\lambda_t,x)} F dt\nu(dx) \right| \right] + \mathbb{E} \left[ \int_{\mathbb{R}_+ \times \mathbb{R}} |Z_{(t,x)}| \lambda_t |D_{(t,\lambda_t,x)} F|^2 dt\nu(dx) \right]. \quad (3.3.3)$$

*Proof.* Note first of all that by Theorem 3.2.19,  $F$  is well-defined. Then, let  $f$  in  $\mathfrak{F}_W^0$  (see Definition 3.2.20). As  $F = \delta^\Pi(Z\mathcal{Z})$ , the integration by parts formula (3.2.6) leads to

$$\begin{aligned} \mathbb{E} [f(F)F] &= \mathbb{E} [f(F)\delta^\Pi(Z\mathcal{Z})] \\ &= \mathbb{E} \left[ \int_{\mathbb{R}_+ \times \mathbb{R}} Z_{(t,x)} \lambda_t D_{(t,\lambda_t,x)} f(F) dt\nu(dx) \right] \\ &= \mathbb{E} \left[ \int_{\mathbb{R}_+ \times \mathbb{R}} Z_{(t,x)} \lambda_t \left( f(F \circ \varepsilon_{(t,\lambda_t,x)}^+) - f(F) \right) dt\nu(dx) \right]. \end{aligned}$$

Using a Taylor expansion, (with  $\bar{F}$  be a random element between  $F_T \circ \varepsilon_{(t,\lambda_t,x)}^+$  and  $F$ ) we have

$$f(F \circ \varepsilon_{(t,\lambda_t,x)}^+) - f(F) = f'(F) D_{(t,\lambda_t,x)} F + \frac{1}{2} f''(\bar{F}) |D_{(t,\lambda_t,x)} F|^2.$$

Hence,

$$\begin{aligned} &|\mathbb{E} [\gamma^2 f'(F) - f(F)F]| \\ &= \left| \mathbb{E} \left[ f'(F) \left( \gamma^2 - \int_{\mathbb{R}_+ \times \mathbb{R}} Z_{(t,x)} \lambda_t D_{(t,\lambda_t,x)} F dt\nu(dx) \right) \right] - \frac{1}{2} \mathbb{E} \left[ \int_{\mathbb{R}_+ \times \mathbb{R}} Z_{(t,x)} \lambda_t f''(\bar{F}) |D_{(t,\lambda_t,x)} F|^2 dt\nu(dx) \right] \right| \\ &\leq \|f'\|_\infty \mathbb{E} \left[ \left| \gamma^2 - \int_{\mathbb{R}_+ \times \mathbb{R}} Z_{(t,x)} \lambda_t D_{(t,\lambda_t,x)} F dt\nu(dx) \right| \right] + \frac{\|f''\|_\infty}{2} \mathbb{E} \left[ \int_{\mathbb{R}_+ \times \mathbb{R}} |Z_{(t,x)}| \lambda_t |D_{(t,\lambda_t,x)} F|^2 dt\nu(dx) \right] \\ &\leq \mathbb{E} \left[ \left| \gamma^2 - \int_{\mathbb{R}_+ \times \mathbb{R}} Z_{(t,x)} \lambda_t D_{(t,\lambda_t,x)} F dt\nu(dx) \right| \right] + \mathbb{E} \left[ \int_{\mathbb{R}_+ \times \mathbb{R}} |Z_{(t,x)}| \lambda_t |D_{(t,\lambda_t,x)} F|^2 dt\nu(dx) \right], \end{aligned}$$

as  $f$  belongs to  $\mathfrak{F}_W^0$ . □

In particular, we obtain the following corollary.

*Corollary 3.3.5.* For each  $T > 0$ , consider  $Z^T := (Z_{(t,x)}^T)_{(t,x) \in \mathbb{R}_+ \times \mathbb{R}}$  defined as :

$$Z_{(t,x)}^T := x\alpha_t \mathbf{1}_{t \in [0,T]}, \quad t \geq 0, x \in \mathbb{R},$$

where  $(\alpha_t)_{t \in [0,T]}$  is a  $(\mathcal{F}_t^X)_{t \in [0,T]}$ -predictable process.

Let also  $\gamma > 0$ ,  $G \sim \mathcal{N}(0, \gamma^2)$ ,  $m := \int_{\mathbb{R}} x\nu(dx)$  and

$$\begin{aligned} F_T &:= \delta^\Pi(Z^T \mathcal{Z}) = \int_{(0,T] \times \mathbb{R}_+ \times \mathbb{R}} \alpha_s x \mathbf{1}_{\{\theta \leq \lambda_s\}} (\Pi(ds, d\theta, dx) - dsd\theta\nu(dx)) \\ &= \int_{(0,T]} \alpha_t dx_t - m \int_0^T \alpha_t \lambda_t dt. \end{aligned}$$

Then

$$\begin{aligned} d_W(F_T, G) &\leq \mathbb{E} \left[ \left| \gamma^2 - \int_0^T \alpha_t \lambda_t \int_{\mathbb{R}} x D_{(t,\lambda_t,x)} F_T \nu(dx) dt \right| \right] \\ &\quad + \mathbb{E} \left[ \int_0^T |\alpha_t| \lambda_t \int_{\mathbb{R}} |x| |D_{(t,\lambda_t,x)} F_T|^2 \nu(dx) dt \right]. \end{aligned} \quad (3.3.4)$$

Before going further we make the following remark showing that the decomposition above is sharp in case of a Poisson process.

*Remark 3.3.6.* Assume  $\phi \equiv 0$  so that  $N$  is an homogeneous Poisson process with intensity  $\mu$ . Let  $m = \int_{\mathbb{R}} y\nu(dy)$  and  $X$  is a compound Poisson process which can thus be represented as :

$$X_t = \sum_{i=1}^{N_T} Y_i,$$

with  $(Y_i)_{i \geq 1}$  iid random variables independent of  $N$  with distribution  $\nu$ . Assume in addition that  $\int_{\mathbb{R}} |x|^3 \nu(dx) < +\infty$ . We have that

$$F_T = \frac{X_T - \mu m T}{\sqrt{T}},$$

$\lambda_t = \mu$  and

$$D_{(t,\lambda_t,x)} F_T = \frac{X_T + x - X_T}{\sqrt{T}} = \frac{x}{\sqrt{T}}.$$

In addition,

$$\gamma^2 = \mu \vartheta^2 = \mu \int_{\mathbb{R}} y^2 \nu(dy).$$

As a consequence, for  $Z_{(t,x)}^T = \frac{x}{\sqrt{T}}$  for all  $t \in [0, T]$  in the previous corollary, we get

$$\mathbb{E} \left[ \left| \gamma^2 - \frac{1}{\sqrt{T}} \int_0^T \int_{\mathbb{R}} \lambda_t x D_{(t,\lambda_t,x)} F_T dt \nu(dx) \right| \right] = |\gamma^2 - \mu \vartheta^2| = 0.$$

Hence, the speed of convergence is completely contained in that case by the term

$$\frac{1}{\sqrt{T}} \mathbb{E} \left[ \int_0^T \int_{\mathbb{R}} \lambda_t |x| |D_{(t,\lambda_t,x)} F_T|^2 dt \nu(dx) \right] = \frac{\mu \int_{\mathbb{R}} |x|^3 \nu(dx)}{\sqrt{T}}$$

and thus we recover a bound with central speed  $T^{-1/2}$ .

*Remark 3.3.7.* Before concluding this section we would like to make a comment regarding the Nourdin-Peccati methodology we applied with a slight modification. Indeed, one realises that the key ingredient is to consider a random variable  $F$  of the form  $F = \delta(u)$  where  $u$  is a given process (belonging to an appropriate class). In the Nourdin-Peccati's approach (in both Gaussian and Poisson frameworks), one consider a centered random variable  $F$  that then naturally belongs to the domain of the Ornstein-Uhlenbeck operator  $L$ , which can be defined as  $LF = -\delta(DF)$ , where once again in both Gaussian and Poisson frameworks  $D$  is the Malliavin derivative and  $\delta$  the divergence operator. Hence, writing  $F = LL^{-1}F = -\delta(DL^{-1}F)$  one gets back to the previous divergence form.

Coming back to the notations of Theorem 4.3.4 (and choosing  $\nu(dx) = \delta_1(dx)$  for simplicity of notations), for a Hawkes functional  $F$  of the form  $F = \delta^{\Pi}(ZZ)$ , by adapting the Nourdin-Peccati methodology to our framework (so by adapting [73]), one would obtain the inequality (once again using the notations of Theorem 4.3.4):

$$d_W(F, G) \leq \mathbb{E} \left[ \left| \gamma^2 - \int_0^{+\infty} \lambda_t (-D_{(t,\lambda_t)} L^{-1}F) D_{(t,\lambda_t)} F dt \right| \right] + \mathbb{E} \left[ \int_0^{+\infty} |D_{(t,\lambda_t)} L^{-1}F| \lambda_t |D_{(t,\lambda_t)} F|^2 dt \right]. \tag{3.3.5}$$

However, as  $F$  is of the form  $F = \delta^{\Pi}(ZZ)$ , bounds (3.3.3) and (3.3.5) coincide as :

$$-D_{(t,\lambda_t)} L^{-1}F = Z_t, \quad \mathbb{P} \otimes \lambda_t dt - \text{a.e.},$$

or equivalently

$$-D_{(t,\theta)} L^{-1}F = Z_t \mathcal{Z}_{(t,\theta)}, \quad \mathbb{P} \otimes dt \otimes d\theta - \text{a.e..}$$

Indeed, letting  $\rho_{(t,\theta)} := -D_{(t,\theta)} L^{-1}F$ , it holds that

$$\delta^{\Pi}(\rho) = -\delta^{\Pi}(-\rho) = -\delta^{\Pi}(DL^{-1}F) = LL^{-1}F = F = \delta^{\Pi}(ZZ).$$

Finally note that the fact that the bound rewrites in a simpler form for divergence form functionals has already been observed and exploited (see for instance [76], [11] or [91]).

### 3.3.2 Application to the Hawkes process

#### 3.3.2.1 Overview of our results

We aim to provide the speed of convergence for the convergence of the renormalized Hawkes process when  $T$  tends to  $+\infty$ .

We recall that in the case of a (non-compound) Hawkes process (that is  $X_t = N_t$ ,  $\nu(dx) = \delta_1(dx)$  or equivalently  $Y_i \equiv 1$  in (3.2.3)), it has been proved as [7, Lemma 7] that

$$\left( \frac{N_{Tv} - \int_0^{Tv} \lambda_s ds}{\sqrt{T}} \right)_{v \in [0,1]} \xrightarrow[T \rightarrow +\infty]{\mathcal{L}\text{-}\mathcal{S}} (\sigma W_v)_{v \in [0,1]},$$

(with  $W$  a Brownian motion and  $\sigma^2 := \frac{\mu}{1 - \|\phi\|_1}$ ) which in particular implies

$$\frac{N_T - \int_0^T \lambda_s ds}{\sqrt{T}} \xrightarrow[T \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2). \quad (3.3.6)$$

In what follows, we provide for a compound Hawkes process  $X := (X_t)_{t \in [0, T]}$ ,

$$X_t = \sum_{i=1}^{N_t} Y_i,$$

(( $Y_i$ ) $_{i \geq 1}$  iid with common distribution  $\nu$  and independent of  $N$ ) the counterpart of the normal convergence (3.3.6). Let

$$F_T := \frac{X_T - m \int_0^T \lambda_t dt}{\sqrt{T}}, \quad \sigma^2 := \frac{\mu}{1 - \|\phi\|_1}; \quad \vartheta^2 = \int_{\mathbb{R}} y^2 \nu(dy).$$

More precisely

1. We give, as Theorem 3.3.10, a general bound for the speed of convergence (with respect to  $T$ ) of the convergence (3.3.6) in the generalized case of a compound Hawkes process.
2. We prove, as Theorem 3.3.12 (see this result for a precise statement), for a compound Hawkes process, and in case of an exponential kernel  $\phi(u) = \alpha e^{-\beta u}$ ,  $u \geq 0$  (with  $0 < \alpha < \beta$ ) or an Erlang kernel  $\phi(u) = \alpha u e^{-\beta u}$ ,  $u \geq 0$  (with  $0 < \alpha < \beta^2$ ) the speed of convergence  $O\left(\frac{1}{\sqrt{T}}\right)$  for the counterpart of convergence (3.3.6) :

$$d_W(F_T, \mathcal{N}(0, \sigma^2 \vartheta^2)) \leq \frac{C_{\alpha, \beta, \nu}}{\sqrt{T}};$$

3. We provide, as Theorem 3.3.13, for the Hawkes process and in case of an exponential kernel or of an Erlang kernel, a speed of convergence (with respect to the Wasserstein distance) of the modified CLT in the spirit of [7, Theorem 2] as follows :

$$d_W(Y_T, \mathcal{N}(0, \tilde{\sigma}^2)) \leq \frac{\tilde{C}_{\alpha, \beta}}{\sqrt{T}},$$

where  $Y_T := \frac{N_T - \int_0^T \mathbb{E}[\lambda_t] dt}{\sqrt{T}}$  and  $\tilde{\sigma}^2 = \frac{\mu}{(1 - \|\phi\|_1)^3}$ .



*Remark 3.3.8.* We would like to discuss here the work [91] which is the closest to ours. In [91], the author makes use of the Poisson embedding representation to apply the Nourdin-Peccati methodology. So on that regard our work follows the same point of view. However, the remark (initially pointed out in [48] and fully exploited in this chapter) stated as Lemma 3.2.14, which is valid for Hawkes processes but not for general counting processes with a stochastic intensity, is exploited to provide new bounds in Theorem 3.3.10 below for the Gaussian approximation of Hawkes functionals. In [91], the general case of a stochastic intensity point process is considered (including non-linear Hawkes processes), but then calls for some estimates that turn out to fail to be sharp enough in the particular case of a linear Hawkes process. For instance, according to [91, Theorem 3.1] (see also [91, Relations (5.2) and (5.4)])

$$d_W \left( \frac{N_T - \mu T}{\sqrt{T}}; \mathcal{N}(0, \sigma^2) \right) \leq B_T, \quad \text{with } B_T \geq \frac{2\sqrt{2/\pi} \|\phi\|_1 (2 - \|\phi\|_1)}{1 - \|\phi\|_1};$$

that cannot converge to 0 as  $T$  tends to infinity.

### 3.3.2.2 Quantitative Central Limit Theorems for compound Hawkes processes

In order to make precise some of the statements below we recall that the Malliavin derivative of  $X$ ,  $N$  and  $\lambda$  involves the following parametrized system (see Proposition 3.2.17).

*Notation 3.3.9.* For fixed  $t \geq 0$ , we denote by  $(\hat{X}_s^t, \hat{N}_s^t, \hat{\lambda}_s^t)_{s \geq t}$  the unique solution to the SDE

$$\begin{cases} \hat{N}_s^t = \int_{(t,s] \times \mathbb{R}_+ \times \mathbb{R}} \mathbf{1}_{\{\lambda_u < \theta \leq \lambda_u + \hat{\lambda}_u^t\}} \Pi(du, d\theta, dy), & s \geq t, \\ \hat{X}_s^t = \int_{(t,s] \times \mathbb{R}_+ \times \mathbb{R}} y \mathbf{1}_{\{\lambda_u < \theta \leq \lambda_u + \hat{\lambda}_u^t\}} \Pi(du, d\theta, dy), & s \geq t, \\ \hat{\lambda}_s^t = \phi(s-t) + \int_{(t,s)} \phi(s-u) d\hat{N}_u^t, & s > t, \hat{\lambda}_t^t = 0. \end{cases}$$

In addition we introduce different compensated  $\mathbb{F}^\Pi$ -martingale processes for the (added-point) Hawkes process and the corresponding compound Hawkes process

$$M_s := X_s - m \int_0^s \lambda_u du, \quad s \in [0, T], \quad \hat{M}_s^t := \hat{X}_s^t - m \int_t^s \hat{\lambda}_u^t du, \quad s \in [t, T].$$

$$\mathcal{M}_s := N_s - \int_0^s \lambda_u du, \quad s \in [0, T], \quad \hat{\mathcal{M}}_s^t := \hat{N}_s^t - \int_t^s \hat{\lambda}_u^t du, \quad s \in [t, T].$$

*Theorem 3.3.10.* Assume Assumptions 3.3.1, 3.3.2 and 3.3.3 are in force.

For  $T > 0$ , let  $Z_{(t,x)}^T := \frac{x}{\sqrt{T}}$ ,  $t \in [0, T]$  and recall

$$F_T = \delta^\Pi(Z^T \mathcal{Z}^T) = \frac{X_T - m \int_0^T \lambda_t dt}{\sqrt{T}}.$$

Let also  $G \sim \mathcal{N}(0, \sigma^2 \vartheta^2)$  with  $\sigma^2 := \frac{\mu}{1 - \|\phi\|_1}$  (recall  $\vartheta^2 = \int_{\mathbb{R}} y^2 \nu(dy)$ ). There exists a constant  $C_{\phi, \nu} > 0$  depending on  $\phi$  and  $\nu$  only such that for any  $T > 0$

$$d_W(F_T, G) \leq \frac{C_{\phi, \nu}}{\sqrt{T}} + \frac{|m|}{T} \mathbb{E} \left[ \left| \int_0^T \lambda_t \hat{M}_T^t dt \right| \right],$$

where  $\hat{M}^t$  is defined in Notation 3.3.9.

*Proof.* To ease the presentation, we chose to provide here the main steps of the proof and to postpone in Section 3.4 the proof of general results concerning the convergence of moments of order 2. By Theorem 4.3.4 we have that

$$d_W(F_T, G) \leq A_1 + A_2$$

where

$$A_1 := \mathbb{E} \left[ \left| \sigma^2 \vartheta^2 - \frac{1}{\sqrt{T}} \int_0^T \int_{\mathbb{R}} x \lambda_t D_{(t, \lambda_t, x)} F_T dt \nu(dx) \right| \right],$$

$$A_2 := \frac{1}{\sqrt{T}} \mathbb{E} \left[ \int_0^T \int_{\mathbb{R}} \lambda_t |x| |D_{(t, \lambda_t, x)} F_T|^2 dt \nu(dx) \right].$$

Note that

$$\sigma^2 \vartheta^2 = \lim_{T \rightarrow +\infty} \mathbb{E}[|F_T|^2] = \lim_{T \rightarrow +\infty} \frac{1}{T} \mathbb{E}[|M_T|^2] = d \lim_{T \rightarrow +\infty} \frac{1}{T} \mathbb{E}[N_T] = d \lim_{T \rightarrow +\infty} \frac{1}{T} \int_0^T \mathbb{E}[\lambda_t] dt,$$

and we will quantify the speed of convergence. Before going further, recall that according to Proposition 3.2.17,

$$D_{(t, \lambda_t, x)} F_T = \frac{1}{\sqrt{T}} D_{(t, \lambda_t, x)} M_T = \frac{1}{\sqrt{T}} \left( x + \hat{M}_T^t \right) \quad (3.3.7)$$

where  $(\hat{M}_s^t := \hat{X}_s^t - m \int_t^s \hat{\lambda}_u du)_{s \in [t, T]}$ , and  $(\hat{N}^t, \hat{\lambda}^t)$  is defined by (4.2.4) and recalled in Notation 3.3.9. We treat both terms  $A_1$  and  $A_2$  separately.

#### Term $A_1$

We have

$$\begin{aligned} A_1 &= \mathbb{E} \left[ \left| \sigma^2 \vartheta^2 - \frac{1}{\sqrt{T}} \int_0^T \int_{\mathbb{R}} x \lambda_t D_{(t, \lambda_t, x)} F_T dt \nu(dx) \right| \right] \\ &= \mathbb{E} \left[ \left| \sigma^2 \vartheta^2 - \frac{\vartheta^2}{T} \int_0^T \lambda_t dt - \frac{1}{T} \int_0^T \int_{\mathbb{R}} x \lambda_t \hat{M}_T^t dt \nu(dx) \right| \right] \\ &\leq \left| \sigma^2 \vartheta^2 - \frac{\vartheta^2}{T} \int_0^T \mathbb{E}[\lambda_t] dt \right| + \frac{\vartheta^2}{T} \mathbb{E} \left[ \left| \int_0^T (\lambda_t - \mathbb{E}[\lambda_t]) dt \right| \right] + \frac{|m|}{T} \mathbb{E} \left[ \left| \int_0^T \lambda_t \hat{M}_T^t dt \right| \right] \\ &=: A_{1,1} + \vartheta^2 A_{1,2} + |m| A_{1,3}. \end{aligned}$$

By Lemma 3.4.1 (in Section 3.4),

$$A_{1,1} = O\left(\frac{1}{T}\right). \quad (3.3.8)$$

We turn to term  $A_{1,2}$ . According to the second line of [7, Lemma 4] one has

$$N_T - \mathbb{E}[N_T] = \mathcal{M}_T + \int_0^T \chi(T-s)\mathcal{M}_s ds,$$

where  $\mathcal{M}_t = N_t - \int_0^t \lambda_s ds$ . By subtracting  $N_T$  from the equation one gets

$$-\mathbb{E}[N_T] = -\int_0^T \lambda_s ds + \int_0^T \chi(T-s)\mathcal{M}_s ds,$$

thus since  $\mathbb{E}[N_T] = \int_0^T \mathbb{E}[\lambda_s] ds$

$$\int_0^T \lambda_s - \mathbb{E}[\lambda_s] ds = \int_0^T \chi(T-s)\mathcal{M}_s ds,$$

where we recall that we defined  $\chi$  in (3.3.2). This means that

$$A_{1,2} \leq \frac{1}{T} \int_0^T \chi(T-s) \mathbb{E}[|\mathcal{M}_s|] ds \leq \frac{1}{T} \int_0^T \chi(T-s) \mathbb{E}[|\mathcal{M}_s|^2]^{1/2} ds \leq \frac{1}{T} \int_0^T \chi(T-s) \mathbb{E}[N_s]^{1/2} ds.$$

According to the proof of [7, Lemma 5], with Assumption 3.3.3 that  $\int_0^{+\infty} s\phi(s) ds < +\infty$ , then

$C := \int_0^{+\infty} s\chi(s) ds < +\infty$  and a fortiori  $\forall p \in [0, 1]$ ,  $\int_0^{+\infty} s^p \chi(s) ds < +\infty$ . Following once again [7, Lemma 4],

$$\mathbb{E}[N_t] = \mu t + \int_0^t \chi(t-s) s ds = \mu t + \int_0^t \chi(s)(t-s) ds \leq (\mu + \|\chi\|_1)t + C.$$

Hence, using  $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$  we have the following inequality (for some  $B > 0$ )

$$\mathbb{E}[N_t]^{1/2} \leq B(\sqrt{t} + 1).$$

Now the term  $A_{1,2}$  becomes bounded by

$$\begin{aligned} A_{1,2} &\leq \frac{B}{T} \left( \int_0^T \chi(T-s) \sqrt{s} ds + \int_0^T \chi(T-s) ds \right) \\ &\leq \frac{B}{T} \left( \int_0^T \chi(s) \sqrt{T-s} ds + \|\chi\|_1 \right) \\ &\leq \frac{B}{\sqrt{T}} \left( \int_0^T \chi(s) \sqrt{\frac{T-s}{T}} ds + \frac{1}{\sqrt{T}} \|\chi\|_1 \right) \\ &\leq \frac{B}{\sqrt{T}} \left( \int_0^T \chi(s) ds + \frac{1}{\sqrt{T}} \|\chi\|_1 \right) \\ &\leq \frac{B\|\chi\|_1}{\sqrt{T}} \left( 1 + \frac{1}{\sqrt{T}} \right) \end{aligned}$$

leading to

$$A_{1,2} = O\left(\frac{1}{\sqrt{T}}\right). \quad (3.3.9)$$

Combining (3.3.8) and (3.3.9) we get that

$$A_1 = O\left(\frac{1}{\sqrt{T}}\right) + \frac{|m|}{T} \mathbb{E} \left[ \left| \int_0^T \lambda_t \hat{M}_T^t dt \right| \right].$$

**Term  $A_2$**

Recall that

$$\begin{aligned} A_2 &= \frac{1}{\sqrt{T}} \mathbb{E} \left[ \int_0^T \int_{\mathbb{R}} \lambda_t |x| |D_{(t, \lambda_t, x)} F_T|^2 \nu(dx) dt \right] \\ &\leq \frac{2 \int_{\mathbb{R}} |x|^3 \nu(dx)}{T^{3/2}} \mathbb{E} \left[ \int_0^T \lambda_t dt \right] + \frac{2}{T^{3/2}} \mathbb{E} \left[ \int_0^T \int_{\mathbb{R}} |x| \lambda_t |\hat{M}_T^t|^2 \nu(dx) dt \right] \\ &= \frac{2 \int_{\mathbb{R}} |x|^3 \nu(dx)}{T^{3/2}} \mathbb{E} \left[ \int_0^T \lambda_t dt \right] + \frac{2 \int_{\mathbb{R}} |x| \nu(dx)}{T^{3/2}} \mathbb{E} \left[ \int_0^T \lambda_t |\hat{M}_T^t|^2 dt \right] \\ &=: 2 \left( \int_{\mathbb{R}} |x|^3 \nu(dx) A_{2,1} + \int_{\mathbb{R}} |x| \nu(dx) A_{2,2} \right). \end{aligned}$$

By Lemma 3.4.1, we immediately get that

$$A_{2,1} = O\left(\frac{1}{\sqrt{T}}\right). \quad (3.3.10)$$

Finally, for Term  $A_{2,2}$ , using the predictable feature of  $\lambda$  and the fact that the Poisson measure  $\Pi$  is compensated by the measure  $dt \otimes d\theta \otimes \nu$ , it holds that

$$\begin{aligned} A_{2,2} &= \frac{1}{T^{3/2}} \mathbb{E} \left[ \int_0^T \lambda_t |\hat{M}_T^t|^2 dt \right] \\ &= \frac{1}{T^{3/2}} \int_0^T \mathbb{E} \left[ \lambda_t \mathbb{E} \left[ |\hat{M}_T^t|^2 | \mathcal{F}_t^\Pi \right] \right] dt \\ &= \frac{1}{T^{3/2}} \int_0^T \mathbb{E} \left[ \lambda_t \mathbb{E} \left[ \int_t^T \int_{\mathbb{R}} \int_{\mathbb{R}_+} x^2 \mathbf{1}_{\{\theta \leq \hat{\lambda}_u^t\}} \Pi(du, d\theta, dx) | \mathcal{F}_t^\Pi \right] \right] dt \\ &= \frac{\vartheta^2}{T^{3/2}} \int_0^T \mathbb{E} \left[ \lambda_t \int_t^T \mathbb{E} \left[ \hat{\lambda}_s^t | \mathcal{F}_t^\Pi \right] ds \right] dt \end{aligned}$$

where we used for the last equality the identity  $\hat{N}_t^t = 0$ . Using Lemma 3.4.2,

$$|A_{2,2}| \leq \vartheta^2 \frac{\|\phi\|_1 (1 + \|\chi\|_1)}{T^{3/2}} \int_0^T \mathbb{E}[\lambda_t] dt,$$

and since  $\int_0^T \mathbb{E}[\lambda_t] dt = \mathbb{E}[N_T] \leq B(T+1)$  where  $B$  is a positive constant (see computations for Term  $A_{1,1}$  above), we have

$$A_{2,2} = O\left(\frac{1}{\sqrt{T}}\right). \quad (3.3.11)$$

Combining (3.3.10) and (3.3.11), we get that

$$A_2 = O\left(\frac{1}{\sqrt{T}}\right).$$

□

*Remark 3.3.11.* For simplicity let  $R_T := \frac{1}{T} \mathbb{E} \left[ \left| \int_0^T \lambda_t \hat{M}_T^t dt \right| \right]$ . According to Theorem 3.3.10, determining the speed of convergence for a general Hawkes process requires the speed of convergence for  $R_T$ . Unfortunately, we have not been able to give such estimates for a general kernel  $\phi$ . This is the only part of our analysis that fails in complete generality, and the main difficulty can be attributed to the self-exciting feature of the intensity. Indeed, in the Poisson case, that is when  $\phi \equiv 0$ , we have that  $R_T = 0$  for any  $T$ , as discussed in Remark 3.3.6. Dealing with this term calls for a precise statement of the correlation between the original intensity  $\lambda$  and the transformed one (more precisely of the Malliavin derivative) Hawkes martingale  $\hat{M}$  (defined in Notation 3.3.9). It is worth noticing that getting estimates on the correlation of the Hawkes process itself is already quite challenging for a general kernel and constitutes an active area of research (see *e.g.* [49]). However, in the particular cases of the exponential and Erlang kernels, we can provide suitable estimates on  $R_T$  as we will see in Theorem 3.3.12.

*Theorem 3.3.12.* Assume an exponential kernel  $\phi(u) = \alpha e^{-\beta u}$ ,  $u \geq 0$  (with  $0 < \alpha < \beta$ ) or an Erlang kernel  $\phi(u) = \alpha u e^{-\beta u}$ ,  $u \geq 0$  (with  $0 < \alpha < \beta^2$ ).

Then, using the notations of Theorem 3.3.10 ( $F_T = \frac{N_T - \int_0^T \lambda_t dt}{\sqrt{T}}$ ;  $G \sim \mathcal{N}(0, \sigma^2 \vartheta^2)$ ), there exists in both cases a constant  $C_{\alpha, \beta, \nu} > 0$  depending only on  $\alpha, \beta$  and  $\nu$  such that for any  $T > 0$

$$d_W(F_T, G) \leq \frac{C_{\alpha, \beta, \nu}}{\sqrt{T}}.$$

*Proof.* Obviously, Assumptions (3.3.2)-(3.3.3) are in force. By Theorem 3.3.10, we need to estimate the quantity

$$R_T = \frac{1}{T} \mathbb{E} \left[ \left| \int_0^T \lambda_t \hat{M}_T^t dt \right| \right] = \frac{1}{T} \mathbb{E} \left[ \left| \int_0^T \lambda_t (D_{(t, \lambda_t)} M_T - 1) dt \right| \right],$$

where  $\hat{M}^t$  is defined in Notation 3.3.9. In Lemma 4.4.5 we prove that  $R_T = O(\frac{1}{\sqrt{T}})$ . □

### 3.3.2.3 Alternative quantitative Limit Theorem for the exponential and Erlang Hawkes processes

In this section we consider the Hawkes process  $N$  (alternatively one can set  $\nu(dx) = \delta_1(dx)$  or  $Y_i \equiv 1$  in Representation (3.2.3)). It has been proven in [7] that as  $T$  goes to infinity

$$Y_T \xrightarrow[T \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \tilde{\sigma}^2),$$

where  $Y_T = \frac{N_T - \int_0^T \mathbb{E}[\lambda_t] dt}{\sqrt{T}}$  is a centered and normalized Hawkes process and  $\tilde{\sigma}^2 = \frac{\mu}{(1 - \|\phi\|_1)^3}$ .

The goal of this section is to provide the speed of convergence of  $Y_T$  using Wasserstein metric between  $F_T$  and its Gaussian limit that we have established in the last paragraph.

*Theorem 3.3.13.* Set  $Y_T = \frac{N_T - \int_0^T \mathbb{E}[\lambda_t] dt}{\sqrt{T}}$  and  $\tilde{\sigma}^2 = \frac{\mu}{(1 - \|\phi\|_1)^3}$ .

Assume an exponential kernel  $\phi(u) = \alpha e^{-\beta u}$ ,  $u \geq 0$  (with  $0 < \alpha < \beta$ ) or an Erlang kernel  $\phi(u) = \alpha u e^{-\beta u}$ ,  $u \geq 0$  (with  $0 < \alpha < \beta^2$ ).

Then, if  $G \sim \mathcal{N}(0, \tilde{\sigma}^2)$  there exists in both cases a constant  $C_{\alpha, \beta} > 0$  depending only on  $\alpha, \beta$  such that for any  $T > 0$

$$d_W(Y_T, G) \leq \frac{C_{\alpha, \beta}}{\sqrt{T}}.$$

*Proof.* As we show in Lemma 3.4.7, it is possible to link  $Y_T$  to  $F_T$  via the relation

$$\frac{Y_T}{\gamma} = F_T + \mathfrak{R}_T$$

where  $\gamma = \frac{1}{1 - \|\phi\|_1}$  is a positive constant and  $\mathfrak{R}_T$  is a "small remainder" whose expressions depend on the kernel. According to (3.2.7)

$$d_W\left(\frac{Y_T}{\gamma}, \frac{G}{\gamma}\right) \leq \sup_{f \in \mathfrak{F}_W^0} \left| \mathbb{E} \left[ \frac{\tilde{\sigma}^2}{\gamma^2} f' \left( \frac{Y_T}{\gamma} \right) - \frac{Y_T}{\gamma} f \left( \frac{Y_T}{\gamma} \right) \right] \right|,$$

where  $G$  is a Gaussian of variance  $\tilde{\sigma}^2 = \frac{\mu}{(1 - \|\phi\|_1)^3}$ .

By applying a Taylor expansion on  $f$  and  $f'$ :

$$f\left(\frac{Y_T}{\gamma}\right) = f(F_T + \mathfrak{R}_T) = f(F_T) + \mathfrak{R}_T f'(X^*),$$

$$f'\left(\frac{Y_T}{\gamma}\right) = f'(F_T + \mathfrak{R}_T) = f'(F_T) + \mathfrak{R}_T f''(\bar{X}),$$

with  $X^*, \bar{X}$  in  $[F_T \wedge (F_T + \mathfrak{R}_T), F_T \vee (F_T + \mathfrak{R}_T)]$  two random variables. Thus:

$$\begin{aligned} & d_W\left(\frac{Y_T}{\gamma}, \frac{G}{\gamma}\right) \\ & \leq \sup_{f \in \mathfrak{F}_W^0} \left| \mathbb{E} \left[ \frac{\tilde{\sigma}^2}{\gamma^2} f'(F_T + \mathfrak{R}_T) - (F_T + \mathfrak{R}_T) f(F_T + \mathfrak{R}_T) \right] \right| \\ & = \sup_{f \in \mathfrak{F}_W^0} \left| \mathbb{E} \left[ \frac{\tilde{\sigma}^2}{\gamma^2} (f'(F_T) + \mathfrak{R}_T f''(\bar{X})) - (F_T + \mathfrak{R}_T) (f(F_T) + \mathfrak{R}_T f'(X^*)) \right] \right| \\ & = \sup_{f \in \mathfrak{F}_W^0} \left| \mathbb{E} \left[ \frac{\tilde{\sigma}^2}{\gamma^2} f'(F_T) - F_T f(F_T) + \frac{\tilde{\sigma}^2}{\gamma^2} \mathfrak{R}_T f''(\bar{X}) - \mathfrak{R}_T f(F_T) - F_T \mathfrak{R}_T f'(X^*) - \mathfrak{R}_T^2 f'(X^*) \right] \right| \\ & \leq \sup_{f \in \mathfrak{F}_W^0} \left| \mathbb{E} \left[ \frac{\tilde{\sigma}^2}{\gamma^2} f'(F_T) - F_T f(F_T) \right] \right| \\ & + \sup_{f \in \mathfrak{F}_W^0} \mathbb{E} \left[ \left| \frac{\tilde{\sigma}^2}{\gamma^2} \mathfrak{R}_T f''(\bar{X}) \right| + |\mathfrak{R}_T f(F_T)| + |F_T \mathfrak{R}_T f'(X^*)| + |\mathfrak{R}_T^2 f'(X^*)| \right]. \end{aligned}$$

The choice of the variance yields  $\frac{\tilde{\sigma}^2}{\gamma^2} \frac{\mu}{(1 - \|\phi\|_1)^3} \cdot (1 - \|\phi\|_1)^2 = \sigma^2$ . In Theorem 3.3.12 we proved that

$$\sup_{f \in \mathfrak{F}_W^0} \left| \mathbb{E} [\sigma^2 f'(F_T) - F_T f(F_T)] \right| \leq O\left(\frac{1}{\sqrt{T}}\right).$$

And since  $f$  is in  $\mathfrak{F}_W^0$ , the first and second derivatives are bounded and the Wasserstein distance is thus bounded by

$$d_W \left( \frac{Y_T}{\gamma}, \frac{G}{\gamma} \right) \leq C \left( \frac{1}{\sqrt{T}} + \mathbb{E}[|\mathfrak{R}_T|] + \sup_{f \in \mathfrak{F}_W^0} \mathbb{E}[|\mathfrak{R}_T f(F_T)|] + \mathbb{E}[|\mathfrak{R}_T F_T|] + \mathbb{E}[|\mathfrak{R}_T^2|] \right)$$

for a positive constant  $C$  that does not depend on  $T$ . In the computations below  $C$  will denote a constant independent of  $T$  which may change from line to line.

We now try to simplify the upper bound. An application of Cauchy-Schwarz inequality yields

$$\mathbb{E}[|\mathfrak{R}_T|] \leq \sqrt{\mathbb{E}[\mathfrak{R}_T^2]}, \quad \mathbb{E}[|\mathfrak{R}_T F_T|] \leq \sqrt{\mathbb{E}[\mathfrak{R}_T^2] \mathbb{E}[F_T^2]},$$

and

$$\begin{aligned} \mathbb{E}[\mathfrak{R}_T f(F_T)]^2 &\leq \mathbb{E}[\mathfrak{R}_T^2] \mathbb{E}[|f(F_T)|^2] \\ &\leq \mathbb{E}[\mathfrak{R}_T^2] \mathbb{E}[\|f'\|_\infty^2 |F_T|^2], \quad \text{using mean value equality and the fact that } f \in \mathfrak{F}_W^0 \\ &\leq \mathbb{E}[\mathfrak{R}_T^2] \mathbb{E}[|F_T|^2]. \end{aligned}$$

In order to have an upper bound on  $\mathbb{E}[|F_T|^2]$ , we recall that  $F_T = \frac{M_T}{\sqrt{T}} = \frac{N_T - \int_0^T \lambda_t dt}{\sqrt{T}}$  where  $M$  is a martingale. This means that  $\mathbb{E}[|M_T|^2] = \mathbb{E}[N_T] = O(T)$  (the last estimate can be found in [35]), which yields  $\mathbb{E}[|F_T|^2] \leq C$  and

$$\mathbb{E}[\mathfrak{R}_T^2] \mathbb{E}[|F_T|^2] \leq C \mathbb{E}[\mathfrak{R}_T^2]$$

and the Wasserstein distance is now bounded by

$$d_W \left( \frac{Y_T}{\gamma}, \frac{G}{\gamma} \right) \leq C \left( \frac{1}{\sqrt{T}} + \sqrt{\mathbb{E}[\mathfrak{R}_T^2]} \right).$$

To perform an estimate on  $\mathbb{E}[\mathfrak{R}_T^2]$  we need to distinguish the case of Exponential and Erlang kernel.

### Case 1 : the kernel is an exponential function

In this case the remainder term writes down as  $\mathfrak{R}_T = \frac{\mathbb{E}[\lambda_T] - \lambda_T}{\beta \sqrt{T}}$ . Hence

$$\begin{aligned} \mathbb{E}[\mathfrak{R}_T^2] &= \mathbb{E} \left[ \left( \frac{\mathbb{E}[\lambda_T] - \lambda_T}{\beta \sqrt{T}} \right)^2 \right] \\ &= \frac{\mathbb{E}[(\lambda_T - \mathbb{E}[\lambda_T])^2]}{\beta^2 T} \\ &= \frac{\text{Var}(\lambda_T)}{\beta^2 T}, \quad \text{since the second moment of } \lambda \text{ is bounded (according to Lemma 3.4.5),} \\ &= O \left( \frac{1}{T} \right). \end{aligned}$$

**Case 2 : the kernel is an Erlang function**

In this case the squared remainder becomes bounded by:

$$\begin{aligned} \mathbb{E}[\mathfrak{R}_T^2] &= \mathbb{E} \left[ \left( \frac{\mathbb{E}[\lambda_T] - \lambda_T}{\beta\sqrt{T}} + \frac{\mathbb{E}[\xi_T] - \xi_T}{\beta^2\sqrt{T}} \right)^2 \right] \\ &\leq 2\mathbb{E} \left[ \left( \frac{\mathbb{E}[\lambda_T] - \lambda_T}{\beta\sqrt{T}} \right)^2 + \left( \frac{\mathbb{E}[\xi_T] - \xi_T}{\beta^2\sqrt{T}} \right)^2 \right], \quad \text{since } 2ab \leq a^2 + b^2, \\ &\leq \frac{2\text{Var}(\lambda_T)}{\beta^2 T} + \frac{2\text{Var}(\xi_T)}{\beta^2 T} \\ &\leq O\left(\frac{1}{T}\right) \end{aligned}$$

since the second moment of  $\xi$  is bounded (according to the proof of Lemma 3.4.5). Hence, in both cases one has

$$d_W \left( \frac{Y_T}{\gamma}, \frac{G}{\gamma} \right) \leq \frac{C}{\sqrt{T}}.$$

And since  $d_W(\frac{Y_T}{\gamma}, \frac{G}{\gamma}) = \frac{1}{\gamma} d_W(Y_T, G)$  (replace  $f(x)$  with  $f_\gamma(x) = \frac{f(\gamma x)}{\gamma}$  in the Wasserstein distance definition) we conclude that there is a positive constant  $C_{\alpha,\beta}$  that does not depend on  $T$  such that

$$d_W(Y_T, G) \leq \frac{C_{\alpha,\beta}}{\sqrt{T}}.$$

□

**Conclusion :** In this chapter we have computed bounds associated to Central Limit Theorems for the compound Hawkes process, using a Mallavin-Stein approach, also known as Nourdin-Peccati's approach. Since a compound Hawkes process is a natural model for the cumulative loss process of an insurance portfolio exhibiting self-exciting features, such bounds are of particular interest for the ruin theory. This is a work in progress.

**3.4 Appendix**

Most of the estimates presented in this section are focused on the Hawkes process and its intensity. As the reader will figure out, the  $x$ -variable (representing the role of the random variables  $(Y_i)_{i \geq 1}$ ) will be factored out of the computations. Hence, the technology is focused on the Hawkes process together with its intensity as if  $\nu(dx) = \delta_1(dx)$ .

We recall the original system  $(X, N, \lambda)$  as (4.2.3) and  $(\hat{X}^t, \hat{N}^t, \hat{\lambda}^t)$  describing the Malliavin derivative recalled in Notation 3.3.9.



### 3.4.1 General estimates

Throughout this section, we assume that Assumptions 3.3.1, 3.3.2 and 3.3.3 are in force. We first recall the following lemma whose statement and proof are given in [7, Theorem 2].

*Lemma 3.4.1.* For any  $T > 0$ , and recalling that  $\sigma^2 = \frac{\mu}{1 - \|\phi\|_1}$ ,

$$\left| \sigma^2 - \frac{1}{T} \int_0^T \mathbb{E}[\lambda_t] dt \right| = O\left(\frac{1}{T}\right).$$

We provide an upper bound for the integral of the first moment of  $\hat{\lambda}^t$  the intensity of the (added point) Hawkes process.

*Lemma 3.4.2.* We make use of the notation of Proposition 3.2.17. Let  $T > 0$  and  $0 \leq s \leq t \leq T$ . The following estimate holds

$$\int_t^T \mathbb{E}_t[\hat{\lambda}_s^t] ds \leq \|\phi\|_1 (1 + \|\chi\|_1).$$

*Proof.* Recall that for  $t \geq 0$ ,  $\mathbb{E}_t[\cdot]$  stands for the conditional expectation  $\mathbb{E}[\cdot | \mathcal{F}_t^\Pi]$ . Taking the conditional expectation in Equation ((4.2.4)) for  $0 \leq t \leq s$  leads to

$$\mathbb{E}_t[\hat{\lambda}_s^t] = \phi(s - t) + \int_{(t,s)} \phi(s - u) \mathbb{E}_t[\hat{\lambda}_u^t] du.$$

For  $t \geq 0$  set  $f_t(s) := \mathbb{1}_{s \geq t} \mathbb{E}_t[\hat{\lambda}_s^t]$  and  $\phi_t(s) := \mathbb{1}_{s \geq t} \phi(s - t)$ , the last equation becomes

$$f_t(s) = \phi_t(s) + \int_0^s \phi(s - u) f_t(u) du.$$

A straightforward application of [7, Lemma 3] yields

$$f_t(s) = \phi_t(s) + \int_0^s \chi(s - u) \phi_t(u) du,$$

and by integrating between  $t$  and  $T$  one obtains

$$\begin{aligned} \int_t^T \mathbb{E}_t[\hat{\lambda}_s^t] ds &= \int_t^T f_t(s) ds \\ &= \int_t^T \phi_t(s) ds + \int_t^T \int_0^s \chi(s - u) \phi_t(u) du ds \\ &= \int_t^T \phi_t(s) ds + \int_t^T \int_t^s \chi(s - u) \phi_t(u) du ds \\ &= \int_t^T \phi_t(s) ds + \int_t^T \int_t^T \mathbb{1}_{s \geq u} \chi(s - u) \phi_t(u) du ds. \end{aligned}$$

Since all the involved functions are positive, we use Fubini's theorem and exchange the integrals to get

$$\begin{aligned} \int_t^T \mathbb{E}_t[\hat{\lambda}_s^t] ds &= \int_t^T \phi_t(s) ds + \int_t^T \int_u^T \chi(s-u) ds \phi_t(u) du \\ &\leq \|\phi\|_1 + \int_t^T \|\chi\|_1 \phi_t(u) du \\ &\leq \|\phi\|_1 (1 + \|\chi\|_1). \end{aligned}$$

□

### 3.4.2 Specific estimates for the exponential and the Erlang's kernels

#### 3.4.2.1 Definition and some properties of the exponential and the Erlang Hawkes processes

**Definition 3.4.3** (Exponential Hawkes process). A counting process  $N$  as in Definition 3.2.9 is referred to

- (i) an exponential Hawkes process if there exist  $(\alpha, \beta) \in \mathbb{R}_+^2$  such that

$$\alpha < \beta \quad \text{and} \quad \phi(u) := \alpha e^{-\beta u}, \quad u \in \mathbb{R}_+. \quad (3.4.1)$$

- (ii) an Erlang Hawkes process if there exist  $(\alpha, \beta) \in \mathbb{R}_+^2$  such that

$$\alpha < \beta^2 \quad \text{and} \quad \phi(u) := \alpha u e^{-\beta u}, \quad u \in \mathbb{R}_+. \quad (3.4.2)$$

*Remark 3.4.4.* Obviously, the two kernels above ((3.4.1)-(3.4.2)) satisfy Assumptions 3.3.2 and (3.3.3).

*Lemma 3.4.5.* Assume that  $N$  is an exponential or an Erlang Hawkes process (that is Condition (3.4.1) or (3.4.2) is in force for some parameters  $\alpha, \beta$ ). Then there is a positive constant  $C$  (depending only on the parameters  $\alpha$  and  $\beta$ ) such that for any  $t \geq 0$

$$\mathbb{E}[\lambda_t^2] \leq C.$$

*Proof.* We treat the two cases (3.4.1) or (3.4.2) separately.

- If  $\phi(u) = \alpha e^{-\beta u}$ ,  $\alpha < \beta$

According to [35], the intensity is a Markov process that satisfies the following Dynkin formula:

$$\frac{d}{dt} \mathbb{E}[g(\lambda_t)] = \mathbb{E}[\beta(\mu - \lambda_t)g'(\lambda_t) + \lambda_t(g(\lambda_t + \alpha) - g(\lambda_t))].$$

This formula can be used to have the variance of the intensity  $\lambda_t$  and an upper bound explicitly.

— If  $\phi(u) = \alpha u e^{-\beta u}$

Even though the intensity is no longer Markov it is possible to see it as a part of a 'Markov cascade' (cf [30]) with an auxiliary process  $(\xi_t)_{t \geq 0}$ . The Dynkin formula for the vector process  $(\lambda_t, \xi_t)_{t \geq 0}$  is

$$\frac{d}{dt} \mathbb{E}[g(\lambda_t, \xi_t)] = \mathbb{E}[(\xi_t + \beta(\mu - \lambda_t)) \partial_\lambda g(\lambda_t, \xi_t) - \beta \xi_t \partial_\xi g(\lambda_t, \xi_t) + \lambda_t (g(\lambda_t, \xi_t + \alpha) - g(\lambda_t, \xi_t))].$$

After taking  $g(\lambda, \xi) = \lambda^2$ ,  $g(\lambda, \xi) = \xi^2$  and  $g(\lambda, \xi) = \lambda \xi$ , we obtain the following system

$$\begin{cases} \partial_t \mathbb{E}[\lambda_t^2] &= -2\beta \mathbb{E}[\lambda_t^2] + 2\mathbb{E}[\lambda_t \xi_t] + 2\beta \mu \mathbb{E}[\lambda_t], \\ \partial_t \mathbb{E}[\xi_t^2] &= -2\beta \mathbb{E}[\xi_t^2] + 2\alpha \mathbb{E}[\lambda_t \xi_t] + \alpha^2 \mathbb{E}[\lambda_t], \\ \partial_t \mathbb{E}[\lambda_t \xi_t] &= \alpha \mathbb{E}[\lambda_t^2] + \mathbb{E}[\xi_t^2] - 2\beta \mathbb{E}[\lambda_t \xi_t] + \beta \mu \mathbb{E}[\xi_t]. \end{cases} \quad (3.4.3)$$

This is a linear system that can be put under the following matrix form:

$$\frac{d}{dt} \begin{pmatrix} \mathbb{E}[\lambda_t^2] \\ \mathbb{E}[\xi_t^2] \\ \mathbb{E}[\lambda_t \xi_t] \end{pmatrix} = \begin{pmatrix} -2\beta & 0 & 2 \\ 0 & -2\beta & 2\alpha \\ \alpha & 1 & -2\beta \end{pmatrix} \begin{pmatrix} \mathbb{E}[\lambda_t^2] \\ \mathbb{E}[\xi_t^2] \\ \mathbb{E}[\lambda_t \xi_t] \end{pmatrix} + \begin{pmatrix} 2\beta \mu \mathbb{E}[\lambda_t] \\ \alpha^2 \mathbb{E}[\lambda_t] \\ \beta \mu \mathbb{E}[\xi_t] \end{pmatrix}.$$

The matrix has three distinct negative eigenvalues:  $v = -2\beta$  and  $v_\pm = -2\beta \pm 2\sqrt{\alpha}$ .

Since  $\mathbb{E}[\lambda_t]$  and  $\mathbb{E}[\xi_t]$  are both bounded by a constant, we conclude that there is  $C > 0$  such that

$$\mathbb{E}[\lambda_t^2] \leq C.$$

□

*Lemma 3.4.6.* Assume that  $\phi(u) = \alpha e^{-\beta u}$  (with  $\alpha < \beta$ ) or  $\phi(u) = \alpha u e^{-\beta u}$  (with  $\alpha < \beta^2$ ).

We remind that  $A_{1,3} = \frac{1}{T} \mathbb{E} \left[ \left| \int_0^T \lambda_t \hat{M}_T^t dt \right|^2 \right]$  and that  $\hat{M}^t$  is defined in Notation 3.3.9. For any  $T > 0$ , we have

$$A_{1,3} = O\left(\frac{1}{\sqrt{T}}\right).$$

*Proof.* We proceed in three steps. First we provide a general estimate for term  $A_{1,3}$  for a general kernel  $\phi$ . Then we make use of this estimate in the two particular cases of exponential and Erlang kernels (3.4.1)-(3.4.2).

### Step 1 : a general estimate

In the following lines we provide an estimate for term  $A_{1,3}$  for a general kernel  $\phi$  satisfying Assumptions 3.3.2 and 3.3.3. The idea is to bound the term  $A_{1,3}$  as tightly as possible. To do so, we start with a Cauchy-Schwarz inequality to get

$$A_{1,3} \leq \frac{1}{T} \mathbb{E} \left[ \left| \int_0^T \lambda_t \hat{M}_T^t dt \right|^2 \right]^{\frac{1}{2}}.$$

From now on we are interested in the term  $\mathbb{E} \left[ \left| \int_0^T \lambda_t \hat{M}_T^t dt \right|^2 \right]$ . By expanding the square one obtains

$$\begin{aligned} \mathbb{E} \left[ \left| \int_0^T \lambda_t \hat{M}_T^t dt \right|^2 \right] &= 2\mathbb{E} \left[ \int_0^T \int_0^t \lambda_t \lambda_s \hat{M}_T^t \hat{M}_T^s ds dt \right] \\ &= 2 \int_0^T \int_0^t \mathbb{E}[\lambda_t \lambda_s \mathbb{E}_t[\hat{M}_T^t \hat{M}_T^s]] ds dt \\ &= 2 \int_0^T \int_0^t \mathbb{E}[\lambda_t \lambda_s \mathbb{E}_t[\hat{M}_T^t (\hat{M}_T^s - \hat{M}_t^s)]] ds dt \\ &= 2 \int_0^T \int_0^t \mathbb{E}[\lambda_t \lambda_s \mathbb{E}_t \left[ [\hat{M}^t, (\hat{M}^s - \hat{M}_t^s)]_T \right]] ds dt, \end{aligned}$$

where we have used the fact that for any  $t$ ,  $\hat{M}^t$  is a martingale with  $\hat{M}_t^t = 0$ . Hence

$$\begin{aligned} &\mathbb{E}_t[[\hat{M}^t, (\hat{M}^s - \hat{M}_t^s)]_T] \\ &= \mathbb{E}_t \left[ \sum_{t < u \leq T} \Delta_u \hat{M}^t \Delta_u \hat{M}^s \right] \\ &= \mathbb{E}_t \left[ \int_{(t,T) \times E} x \mathbf{1}_{\{\theta \leq \hat{\lambda}_u^t\}} \Pi(du, d\theta, dx) \int_{(t,T) \times E} x \mathbf{1}_{\{\theta \leq \hat{\lambda}_u^s\}} \Pi(du, d\theta, dx) \right] \\ &= \mathbb{E}_t \left[ \int_{(t,T) \times E} x^2 \mathbf{1}_{\{\theta \leq \hat{\lambda}_u^t\}} \mathbf{1}_{\{\theta \leq \hat{\lambda}_u^s\}} du d\theta \nu(dx) \right] \\ &= \vartheta^2 \mathbb{E}_t \left[ \int_t^T \min(\hat{\lambda}_u^t, \hat{\lambda}_u^s) du \right]. \end{aligned}$$

The last entity can be bounded as follows

$$\mathbb{E}_t[[\hat{M}^t, (\hat{M}^s - \hat{M}_t^s)]_T] \leq \vartheta^2 \int_t^T \min(\mathbb{E}_t[\hat{\lambda}_u^t], \mathbb{E}_t[\hat{\lambda}_u^s]) du.$$

Since the intensity  $\lambda$  of a Hawkes process is positive, multiplying by it preserves the inequality, leading to

$$\begin{aligned} A_{1,3} &\leq \frac{1}{T} \mathbb{E} \left[ \left| \int_0^T \lambda_t \hat{M}_T^t dt \right|^2 \right]^{\frac{1}{2}} \\ &\leq \vartheta^2 \frac{\sqrt{2}}{\sqrt{T}} \left( \frac{1}{T} \int_0^T \int_0^t \mathbb{E} \left[ \lambda_t \lambda_s \int_t^T \min(\mathbb{E}_t[\hat{\lambda}_u^t], \mathbb{E}_t[\hat{\lambda}_u^s]) du \right] ds dt \right)^{1/2} \end{aligned} \quad (3.4.4)$$

$$=: \vartheta^2 \frac{\sqrt{2}}{\sqrt{T}} \left( \frac{1}{T} I_T \right)^{1/2}, \quad (3.4.5)$$

with obvious notation. The rest of the proof consists in specifying estimates of Quantity  $I_T$  for the two particular Hawkes processes under interest.

**Step 2 : the exponential case**

We assume an exponential kernel (3.4.1)  $\phi(u) = \alpha e^{-\beta u}$  with  $0 < \alpha < \beta$ .

The main advantage of the Markov framework of the exponential kernel is the fact that many formulae for  $\lambda$  and  $\hat{\lambda}$  are known explicitly. In fact, given a starting time  $t$ ,  $\hat{\lambda}^t$  is defined for  $u \geq t$  and it satisfies the following SDE (using once again Notation 3.3.9)

$$d\hat{\lambda}_u^t = (\alpha - \beta)\hat{\lambda}_u^t du + \alpha d\mathcal{M}_u^t,$$

with  $\mathcal{M}_u^t = \hat{N}_u^t - \int_t^u \hat{\lambda}_s^t ds$ , which yields after taking the initial conditions into account

$$\mathbb{E}_t[\hat{\lambda}_u^t] = \alpha e^{(\alpha-\beta)(u-t)} \quad \text{and} \quad \mathbb{E}_t[\hat{\lambda}_u^s] = \hat{\lambda}_t^s e^{(\alpha-\beta)(u-t)}.$$

The bound on  $I_T$  becomes (from now on everything is written up to a positive multiplicative constant  $C$  depending on  $\alpha, \beta$  that may differ from line to line)

$$\begin{aligned} I_T &\leq C \int_0^T \int_0^t \mathbb{E}[\lambda_t \lambda_s \int_t^T \min(\alpha, \hat{\lambda}_t^s) e^{(\alpha-\beta)(u-t)} du] ds dt \\ &= C \int_0^T \int_0^t \mathbb{E}[\lambda_t \lambda_s \min(\alpha, \hat{\lambda}_t^s)] \int_t^T e^{(\alpha-\beta)(u-t)} du ds dt \\ &= C \int_0^T \int_0^t \mathbb{E}[\lambda_t \lambda_s \min(\alpha, \hat{\lambda}_t^s)] ds (1 - e^{(\alpha-\beta)(T-t)}) dt \\ &= C \int_0^T \int_0^t \mathbb{E}[\lambda_s \mathbb{E}_s[\lambda_t \min(\alpha, \hat{\lambda}_t^s)]] ds (1 - e^{(\alpha-\beta)(T-t)}) dt \\ &\leq C \int_0^T \int_0^t \mathbb{E}[\lambda_s \mathbb{E}_s[\lambda_t \hat{\lambda}_t^s]] ds (1 - e^{(\alpha-\beta)(T-t)}) dt. \end{aligned}$$

Since  $\hat{\lambda}_t^s$  starts at  $s$ , it is independent from  $\mathcal{F}_s^\Pi$  thus  $\mathbb{E}_s[\lambda_t \hat{\lambda}_t^s] = \mathbb{E}_s[\lambda_t] \mathbb{E}[\hat{\lambda}_t^s] = \mathbb{E}_s[\lambda_t] \alpha e^{(\alpha-\beta)(t-s)}$ .

Hence the Cauchy-Schwarz inequality together with Lemma 3.4.5 entail :

$$\begin{aligned} I_T &\leq C \int_0^T \int_0^t \mathbb{E}[\lambda_s \mathbb{E}_s[\lambda_t]] e^{(\alpha-\beta)(t-s)} ds (1 - e^{(\alpha-\beta)(T-t)}) dt \\ &= C \int_0^T \int_0^t \mathbb{E}[\lambda_s \lambda_t] e^{(\alpha-\beta)(t-s)} ds (1 - e^{(\alpha-\beta)(T-t)}) dt \\ &\leq C \int_0^T \int_0^t \mathbb{E}[\lambda_s^2]^{\frac{1}{2}} \mathbb{E}[\lambda_t^2]^{\frac{1}{2}} e^{(\alpha-\beta)(t-s)} ds (1 - e^{(\alpha-\beta)(T-t)}) dt, \quad \text{using the Cauchy-Schwarz inequality} \\ &\leq C \int_0^T \int_0^t e^{(\alpha-\beta)(t-s)} ds (1 - e^{(\alpha-\beta)(T-t)}) dt, \quad \text{using Lemma 3.4.5} \\ &= C \int_0^T (1 - e^{(\alpha-\beta)t})(1 - e^{(\alpha-\beta)(T-t)}) dt \\ &= C \int_0^T 1 - e^{(\alpha-\beta)(T-t)} - e^{(\alpha-\beta)t} + e^{(\alpha-\beta)T} dt \\ &\leq C(T + e^{(\alpha-\beta)T} - 1 + T e^{(\alpha-\beta)T}) \quad \text{since } \alpha - \beta < 0 \\ &\leq C T. \end{aligned}$$

Finally using (3.4.4) we conclude that  $A_{1,3} = O\left(\frac{1}{\sqrt{T}}\right)$ .

### Step 3 : the Erlang case

We assume an Erlang kernel (3.4.2)  $\phi(u) = \alpha u e^{-\beta u}$  with  $0 < \alpha < \beta^2$ .

Even though the intensity  $\lambda$  is no longer a Markov process, it is possible to "Markovize" it by taking an auxiliary process  $\xi$  into account.

In the case of the transformed vanishing process  $\hat{\lambda}^s, \hat{\xi}_u^s = \alpha e^{-\beta(u-s)} + \alpha \int_s^u e^{-\beta(u-v)} d\hat{N}_v^s$ , and the process  $(\hat{\lambda}^s, \hat{\xi}^s)$  follows the SDE:

$$\begin{cases} d\hat{\lambda}_u^s &= -\beta\hat{\lambda}_u^s du + \hat{\xi}_u^s du \\ d\hat{\xi}_u^s &= -\beta\hat{\xi}_u^s du + \alpha d\hat{N}_u^s. \end{cases}$$

This yields after solving the system

$$\begin{pmatrix} \hat{\lambda}_u^s \\ \hat{\xi}_u^s \end{pmatrix} = \frac{\sqrt{\alpha}}{2} \left( \left( \int_s^u e^{(\sqrt{\alpha}-\beta)(s-v)} d\hat{\mathcal{M}}_v^s + 1 \right) e^{(\sqrt{\alpha}-\beta)(u-s)} \begin{pmatrix} 1 \\ \sqrt{\alpha} \end{pmatrix} \right. \\ \left. - \left( \int_s^u e^{-(\sqrt{\alpha}+\beta)(s-v)} d\hat{\mathcal{M}}_v^s + 1 \right) e^{-(\sqrt{\alpha}+\beta)(u-s)} \begin{pmatrix} 1 \\ -\sqrt{\alpha} \end{pmatrix} \right).$$

And finally

$$\begin{pmatrix} \mathbb{E}_t[\hat{\lambda}_u^s] \\ \mathbb{E}_t[\hat{\xi}_u^s] \end{pmatrix} = \frac{1}{2} \left( \hat{\lambda}_t^s + \frac{\hat{\xi}_t^s}{\sqrt{\alpha}} \right) e^{(\sqrt{\alpha}-\beta)(u-t)} \begin{pmatrix} 1 \\ \sqrt{\alpha} \end{pmatrix} + \frac{1}{2} \left( \hat{\lambda}_t^s - \frac{\hat{\xi}_t^s}{\sqrt{\alpha}} \right) e^{-(\sqrt{\alpha}+\beta)(u-t)} \begin{pmatrix} 1 \\ -\sqrt{\alpha} \end{pmatrix}. \quad (3.4.6)$$

Once again we make use of the general estimate (3.4.4) and estimate the quantity

$$I_T = \int_0^T \int_0^t \mathbb{E}[\lambda_t \lambda_s \int_t^T \min(\mathbb{E}_t[\hat{\lambda}_u^t], \mathbb{E}_t[\hat{\lambda}_u^s]) du] ds dt,$$

We have

$$I_T \leq \int_0^T \int_0^t \mathbb{E}[\lambda_t \lambda_s \int_t^T \mathbb{E}_t[\hat{\lambda}_u^s] du] ds dt.$$

We chose  $\mathbb{E}_t[\hat{\lambda}_u^s]$  on purpose, because since the process starts earlier it has more chances of vanishing at time  $t$ .

$$\begin{aligned} I_T &\leq C \int_0^T \int_0^t \mathbb{E}[\lambda_t \lambda_s \int_t^T \frac{1}{2} \left( \hat{\lambda}_t^s + \frac{\hat{\xi}_t^s}{\sqrt{\alpha}} \right) e^{(\sqrt{\alpha}-\beta)(u-t)} + \frac{1}{2} \left( \hat{\lambda}_t^s - \frac{\hat{\xi}_t^s}{\sqrt{\alpha}} \right) e^{-(\sqrt{\alpha}+\beta)(u-t)} du] ds dt \\ &= C \int_0^T \int_0^t \mathbb{E}[\lambda_t \lambda_s \left( \hat{\lambda}_t^s + \frac{\hat{\xi}_t^s}{\sqrt{\alpha}} \right) \int_t^T e^{(\sqrt{\alpha}-\beta)(u-t)} du] ds dt \\ &\quad + C \int_0^T \int_0^t \mathbb{E}[\lambda_t \lambda_s \left( \hat{\lambda}_t^s - \frac{\hat{\xi}_t^s}{\sqrt{\alpha}} \right) \int_t^T e^{-(\sqrt{\alpha}+\beta)(u-t)} du] ds dt \\ &\leq C \int_0^T (1 - e^{(\sqrt{\alpha}-\beta)(T-t)}) \int_0^t \mathbb{E}[\lambda_t \lambda_s \left( \hat{\lambda}_t^s + \frac{\hat{\xi}_t^s}{\sqrt{\alpha}} \right)] ds dt \\ &\quad + C \int_0^T (1 - e^{-(\sqrt{\alpha}+\beta)(T-t)}) \int_0^t \mathbb{E}[\lambda_t \lambda_s \left( \hat{\lambda}_t^s - \frac{\hat{\xi}_t^s}{\sqrt{\alpha}} \right)] ds dt \\ &:= C(I_T^+ + I_T^-). \end{aligned}$$

Like in the Markov case,  $\hat{\lambda}^s$  and  $\hat{\xi}^s$  both start at time  $s$ , thus  $\mathbb{E}_s[\lambda_t \hat{\lambda}_t^s] = \mathbb{E}_s[\lambda_t] \mathbb{E}[\hat{\lambda}_t^s]$  and the same holds for  $\hat{\xi}^s$  as well. In addition, since  $\mathbb{E}_s[\hat{\lambda}_t^s] = \mathbb{E}[\hat{\lambda}_t^s]$  and since  $\hat{\lambda}_s^s = 0$ , Equation (3.4.6) yields

$$\mathbb{E}[\hat{\lambda}_t^s] = \frac{\sqrt{\alpha}}{2} (e^{(\sqrt{\alpha}-\beta)(t-s)} - e^{-(\sqrt{\alpha}+\beta)(t-s)})$$

and

$$\frac{\mathbb{E}[\hat{\xi}_t^s]}{\sqrt{\alpha}} = \frac{\sqrt{\alpha}}{2} (e^{(\sqrt{\alpha}-\beta)(t-s)} + e^{-(\sqrt{\alpha}+\beta)(t-s)}).$$

Hence

$$\begin{aligned} I_T^+ &= \int_0^T (1 - e^{(\sqrt{\alpha}-\beta)(T-t)}) \int_0^t \mathbb{E}[\lambda_s \mathbb{E}_s[\lambda_t (\hat{\lambda}_t^s + \frac{\hat{\xi}_t^s}{\sqrt{\alpha}})]] ds dt \\ &= \int_0^T (1 - e^{(\sqrt{\alpha}-\beta)(T-t)}) \int_0^t \mathbb{E}[\lambda_s \mathbb{E}_s[\lambda_t] \mathbb{E}[\hat{\lambda}_t^s + \frac{\hat{\xi}_t^s}{\sqrt{\alpha}}]] ds dt \\ &= \int_0^T (1 - e^{(\sqrt{\alpha}-\beta)(T-t)}) \int_0^t \mathbb{E}[\lambda_t \lambda_s] \sqrt{\alpha} e^{(\sqrt{\alpha}-\beta)(t-s)} ds dt. \end{aligned}$$

As for the exponential kernel, it is possible to have bounds on the intensity's second moment using Lemma 3.4.5, which yields by the Cauchy-Schwarz inequality

$$\mathbb{E}[\lambda_t \lambda_s] \leq \mathbb{E}[\lambda_s^2]^{\frac{1}{2}} \mathbb{E}[\lambda_t^2]^{\frac{1}{2}} \leq C.$$

Finally

$$\begin{aligned} I_T^+ &\leq \int_0^T (1 - e^{(\sqrt{\alpha}-\beta)(T-t)}) \int_0^t e^{(\sqrt{\alpha}-\beta)(t-s)} ds dt \\ &\leq \int_0^T (1 - e^{(\sqrt{\alpha}-\beta)(T-t)}) (1 - e^{(\sqrt{\alpha}-\beta)t}) dt \\ &\leq \int_0^T 1 - e^{(\sqrt{\alpha}-\beta)t} - e^{(\sqrt{\alpha}-\beta)(T-t)} + e^{(\sqrt{\alpha}-\beta)T} dt \\ &= O(T). \end{aligned}$$

Following the same lines we get for  $|I_T^-|$  :

$$\begin{aligned} I_T^- &= \int_0^T (1 - e^{-(\sqrt{\alpha}+\beta)(T-t)}) \int_0^t \mathbb{E}[\lambda_s \mathbb{E}_s[\lambda_t (\hat{\lambda}_t^s - \frac{\hat{\xi}_t^s}{\sqrt{\alpha}})]] ds dt \\ &= \int_0^T (1 - e^{-(\sqrt{\alpha}+\beta)(T-t)}) \int_0^t \mathbb{E}[\lambda_s \mathbb{E}_s[\lambda_t] \mathbb{E}[\hat{\lambda}_t^s - \frac{\hat{\xi}_t^s}{\sqrt{\alpha}}]] ds dt \\ &= - \int_0^T (1 - e^{-(\sqrt{\alpha}+\beta)(T-t)}) \int_0^t \mathbb{E}[\lambda_t \lambda_s] \sqrt{\alpha} e^{-(\sqrt{\alpha}+\beta)(t-s)} ds dt. \end{aligned}$$

So  $|I_T^-| = O(T)$ , and since  $I_T \leq I_T^+ + |I_T^-|$ , we conclude using once again (3.4.4) that

$$A_{1,3} = O\left(\frac{1}{\sqrt{T}}\right).$$

□

*Lemma 3.4.7.* Set  $Y_T = \frac{N_T - \int_0^T \mathbb{E}[\lambda_t] dt}{\sqrt{T}}$  and  $F_T = \frac{N_T - \int_0^T \lambda_t dt}{\sqrt{T}}$ .

Assume that  $\phi(u) = \alpha e^{-\beta u}$  (with  $\alpha < \beta$ ) or  $\phi(u) = \alpha u e^{-\beta u}$  (with  $\alpha < \beta^2$ ).

Then  $Y_T$  and  $F_T$  satisfy the relation:

$$\frac{Y_T}{\gamma} = F_T + \mathfrak{R}_T,$$

where

$$\gamma = \frac{1}{1 - \|\phi\|_1} = \begin{cases} \frac{1}{1 - \alpha/\beta}, & \text{if } \phi(u) = \alpha e^{-\beta u} \\ \frac{1}{1 - \alpha/\beta^2}, & \text{if } \phi(u) = \alpha u e^{-\beta u} \end{cases}$$

and

$$\mathfrak{R}_T = \begin{cases} \frac{\mathbb{E}[\lambda_T] - \lambda_T}{\beta\sqrt{T}}, & \text{if } \phi(u) = \alpha e^{-\beta u} \\ \frac{\mathbb{E}[\lambda_T] - \lambda_T}{\beta\sqrt{T}} + \frac{\mathbb{E}[\xi_T] - \xi_T}{\beta^2\sqrt{T}}, & \text{if } \phi(u) = \alpha u e^{-\beta u} \end{cases}$$

*Proof.* We remind that the advantage of the exponential and Erlang kernels is the SDE. This SDE will be used to eliminate the integral in  $Y_T$  and  $F_T$  in order to obtain alternative expressions.

#### Case 1 : the kernel is an exponential function

We recall that  $\|\phi\|_1 = \frac{\alpha}{\beta}$  and that the intensity is a solution to the following SDE:

$$d\lambda_t = \beta(\mu - \lambda_t)dt + \alpha dN_t.$$

We start by integrating the SDE between 0 and  $T$ :

$$\begin{aligned} \int_0^T d\lambda_t &= \int_0^T \beta(\mu - \lambda_t)dt + \alpha dN_t \\ \lambda_T - \mu &= \beta\mu T - \beta \int_0^T \lambda_t dt + \alpha N_T. \end{aligned}$$

After rearranging the terms,  $F_T$  can be put under the form:

$$F_T = \frac{\lambda_T - \mu + (\beta - \alpha)N_T - \beta\mu T}{\beta\sqrt{T}}.$$

When it comes to  $Y_T$ , we start by taking the expected value of the SDE:

$$\begin{aligned} d\mathbb{E}[\lambda_t] &= \beta(\mu - \mathbb{E}[\lambda_t])dt + \alpha d\mathbb{E}[N_t] \\ &= \beta(\mu - \mathbb{E}[\lambda_t])dt + \alpha\mathbb{E}[\lambda_t]dt \end{aligned}$$

and after integrating with respect to time:

$$\begin{aligned} \mathbb{E}[\lambda_T] - \mu &= \beta\mu T + (\alpha - \beta) \int_0^T \mathbb{E}[\lambda_t] dt \\ &= \beta\mu T + (\alpha - \beta) \left( \int_0^T \mathbb{E}[\lambda_t] dt - N_T \right) + (\alpha - \beta)N_T. \end{aligned}$$



And finally:

$$Y_T = \frac{\mathbb{E}[\lambda_T] - \mu + (\beta - \alpha)N_T - \beta\mu T}{(\beta - \alpha)\sqrt{T}}.$$

These alternative expressions allow us to deduce a simple relation between  $F_T$  and  $Y_T$ :

$$\beta\sqrt{T}F_T - (\beta - \alpha)\sqrt{T}Y_T = \lambda_T - \mathbb{E}[\lambda_T]$$

which is equivalent to:

$$\frac{Y_T}{\gamma} = F_T + \mathfrak{R}_T$$

where  $\gamma = \frac{1}{1 - \|\phi\|_1} = \frac{1}{1 - \alpha/\beta}$  and  $\mathfrak{R}_T = \frac{\mathbb{E}[\lambda_T] - \lambda_T}{\beta\sqrt{T}}$ .

**Case 2 : the kernel is an Erlang function**

In this case  $\|\phi\|_1 = \frac{\alpha}{\beta^2}$  and the vector  $(\lambda_t, \xi_t)$  where  $\xi_t = \int_{[0,t)} \alpha e^{-\beta(t-s)} dN_s$  follows satisfies the following SDE:

$$\begin{cases} d\lambda_t = \xi_t dt + \beta(\mu - \lambda_t) dt \\ d\xi_t = -\beta\xi_t dt + \alpha dN_T. \end{cases}$$

We integrate the system:

$$\begin{cases} \lambda_T - \mu = \int_0^T \xi_t dt + \beta\mu T - \beta \int_0^T \lambda_t dt \\ \xi_T = -\beta \int_0^T \xi_t dt + \alpha N_T. \end{cases}$$

We eliminate  $\int_0^T \xi_t dt$  in the system to obtain:

$$\begin{aligned} \lambda_T - \mu &= \frac{\alpha}{\beta} N_T - \frac{1}{\beta} \xi_T + \mu\beta T - \beta \int_0^T \lambda_t dt \\ &= \frac{\alpha}{\beta} N_T - \frac{1}{\beta} \xi_T + \mu\beta T + \beta\sqrt{T}F_T - \beta N_T \end{aligned}$$

and after re-arranging the terms:

$$\beta\sqrt{T}F_T = \lambda_T - \mu + \frac{\beta^2 - \alpha}{\beta} N_T + \frac{1}{\beta} \xi_T - \mu\beta T.$$

When it comes to  $Y_T$  we take the expected value of the SDE:

$$\begin{cases} d\mathbb{E}[\lambda_t] = (\mathbb{E}[\xi_t] + \beta(\mu - \mathbb{E}[\lambda_t])) dt \\ d\mathbb{E}[\xi_t] = -\beta\mathbb{E}[\xi_t] dt + \alpha\mathbb{E}[\lambda_t] dt. \end{cases}$$

After taking the integral and eliminating  $\int_0^T \mathbb{E}[\xi_t]dt$  in the system:

$$\begin{aligned}\mathbb{E}[\lambda_T] - \mu &= -\frac{1}{\beta}\mathbb{E}[\xi_T] + \frac{\alpha}{\beta} \int_0^T \mathbb{E}[\lambda_t]dt + \beta\mu T - \beta \int_0^T \mathbb{E}[\lambda_t]dt \\ &= -\frac{1}{\beta}\mathbb{E}[\xi_T] + \beta\mu T + \left(\frac{\alpha}{\beta} - \beta\right) \int_0^T \mathbb{E}[\lambda_t]dt \\ &= -\frac{1}{\beta}\mathbb{E}[\xi_T] + \beta\mu T + \frac{\beta^2 - \alpha}{\beta} \sqrt{T}Y_T + \frac{\alpha - \beta^2}{\beta} N_T\end{aligned}$$

which yields after re-arranging the terms:

$$\frac{\beta^2 - \alpha}{\beta} \sqrt{T}Y_T = \mathbb{E}[\lambda_T] - \mu + \frac{\beta^2 - \alpha}{\beta} N_T + \frac{1}{\beta}\mathbb{E}[\xi_T] - \mu\beta T.$$

Combining these expressions on  $F_T$  and  $Y_T$  yields the following relation:

$$\beta\sqrt{T}F_T - \frac{\beta^2 - \alpha}{\beta} \sqrt{T}Y_T = \lambda_T - \mathbb{E}[\lambda_T] + \frac{1}{\beta}(\xi_T - \mathbb{E}[\xi_T])$$

which is equivalent to:

$$\frac{Y_T}{\gamma} = F_T + \mathfrak{R}_T$$

with  $\gamma = \frac{1}{1-\|\phi\|_1} = \frac{1}{1-\alpha/\beta^2}$  and  $\mathfrak{R}_T = \frac{\mathbb{E}[\lambda_T] - \lambda_T}{\beta\sqrt{T}} + \frac{\mathbb{E}[\xi_T] - \xi_T}{\beta^2\sqrt{T}}$ . □



# Gaussian approximation of multivariate compound Hawkes processes

---

## Abstract

In this chapter, we provide upper bounds on the  $d_2$  distance between a large class of functionals of a multivariate compound Hawkes process and a given Gaussian vector. This is proven using Malliavin's calculus defined on an underlying Poisson embedding. The upper bound is then used to infer the speed of convergence of Central Limit Theorems for the multivariate compound Hawkes process with exponential kernels as the observation time  $T$  goes to infinity.

## Contents

---

|  |            |
|--|------------|
| <b>4.1 Introduction</b> . . . . .  | <b>99</b>  |
| <b>4.2 Notations and preliminaries</b> . . . . .                                 | <b>103</b> |
| 4.2.1 General notations . . . . .  | 103        |
| 4.2.2 Elements of stochastic analysis on the multivariate Poisson space          | 104        |
| 4.2.3 Definition of the multivariate compound Hawkes process . . . . .           | 106        |
| 4.2.4 Malliavin's analysis of the multivariate compound Hawkes process . . . . . | 108        |
| 4.2.5 Multivariate Stein's method . . . . .                                      | 110        |
| <b>4.3 Main results</b> . . . . .  | <b>111</b> |
| 4.3.1 General bound . . . . .  | 111        |
| 4.3.2 Bounds for the CLT . . . . .   | 115        |
| 4.3.3 Proof of Theorem 4.1.5 . . . . .   | 124        |
| <b>4.4 Lemmata</b> . . . . .   | <b>125</b> |

---

## 4.1 Introduction

Hawkes processes have been used to model events that exhibit self-exciting properties. Initially introduced by A.G. Hawkes in 1971 [46] to model seismic activities, Hawkes processes became popular in other fields like credit risk [35] or microstructure in finance [7]. In recent years, the need to model systems with interacting components has been growing, that is how the multivariate Hawkes process emerged.

In addition to its self-exciting properties, the multivariate Hawkes process can model mutually-exciting phenomena and it has been used in a wide array of fields such that neuro-science [63, 81], social networks [1], cyber-security [13] and financial econometrics [34].

In this chapter, we consider the multivariate compound process (or total loss process)  $(\mathbf{L}_t)_{t \geq 0}$  defined component-wise at a time  $t$  as

$$\left\{ \begin{array}{l} L_t^1 = \sum_{k=1}^{N_t^1} Y_k^1, \quad (Y_k^1)_{k=1, \dots} \text{ are } i.i.d., \\ \vdots \\ L_t^d = \sum_{k=1}^{N_t^d} Y_k^d, \quad (Y_k^d)_{k=1, \dots} \text{ are } i.i.d., \end{array} \right. \quad (4.1.1)$$

where  $(Y_k^j)_{k \in \mathbb{N}^*, j=1, \dots, d}$  are *i.i.d* random variables and  $\mathbf{N}_t = (N_t^1, \dots, N_t^d)$  is a family of  $d$  point processes each having an intensity  $(\lambda_t^1, \dots, \lambda_t^d)$  such that

$$\lambda_t^i dt = \mathbb{E} [N_{t+dt}^i - N_t^i | \mathcal{F}_{t-}], \quad i = 1, \dots, d.$$

The process  $\mathbf{L}$  is called a multivariate compound Hawkes process if its intensity  $\boldsymbol{\lambda}$  follows the dynamics

$$\left\{ \begin{array}{l} \lambda_t^1 = \mu^1 + \int_{[0,t)} \sum_{k=1}^d \phi_{1k}(t-s) dL_s^k, \\ \vdots \\ \lambda_t^d = \mu^d + \int_{[0,t)} \sum_{k=1}^d \phi_{dk}(t-s) dL_s^k, \end{array} \right. \quad (4.1.2)$$

where  $\boldsymbol{\mu} = (\mu^1, \dots, \mu^d) \in \mathbb{R}_+^d$  plays the role of a baseline intensity and  $\phi = (\phi_{ik})_{i,k=1, \dots, d}$  is a family of non-negative integrable kernels on  $\mathbb{R}_+$ .

This dynamics can also be expressed under the following matrix form

$$\boldsymbol{\lambda}_t = \boldsymbol{\mu} + \int_{[0,t)} \phi(t-s) d\mathbf{L}_s.$$

In some applications, the question about the Hawkes process' longtime behaviour naturally arises. In the special case  $Y_k^j \equiv 1$ , and under the condition that the spectral radius of the matrix  $S := \|\phi\|_1$  is strictly less than one ( $\rho(S) < 1$ ), Bacry *et al.* [7] proved the following martingale Central Limit Theorem (CLT) for  $(\mathbf{L}_t)_{t \geq 0} = (\mathbf{N}_t)_{t \geq 0}$

$$\left( \frac{\mathbf{N}_{Tv} - \int_0^{Tv} \boldsymbol{\lambda}_t dt}{\sqrt{T}} \right)_{v \in [0,1]} \xrightarrow[T \rightarrow +\infty]{} (\boldsymbol{\Sigma}^{1/2} W_v)_{v \in [0,1]}$$

where the convergence takes place in law for the Skorokhod topology and where  $\boldsymbol{\Sigma}$  is a diagonal matrix that depends only on  $\boldsymbol{\mu}$  and  $S$ . From this martingale result,

they derived the alternative CLT

$$\left( \frac{\mathbf{N}_T - \int_0^T \mathbb{E}[\boldsymbol{\lambda}_t] dt}{\sqrt{T}} \right)_{v \in [0,1]} \xrightarrow{T \rightarrow +\infty} \left( (I_d - S)^{-1} \Sigma^{1/2} W_v \right)_{v \in [0,1]}.$$

Until very recently, the quantification of the speed of this convergence has not been thoroughly studied. In the univariate 1-marginal case ( $v = 1$ ), in chapter 3 we found a type bound on the Wasserstein distance between the normalized one-dimensional Hawkes process and its Gaussian limit

$$d_W \left( \frac{N_T - \int_0^T \mathbb{E}[\lambda_t] dt}{\sqrt{T}}, \mathcal{N}(0, \tilde{\sigma}^2) \right) = O \left( \frac{1}{\sqrt{T}} \right), \quad \tilde{\sigma}^2 = \frac{\mu}{(1 - \|\phi\|_1)^3}$$

in case the kernel is an exponential  $\phi(u) = \alpha e^{-\beta u}$  (with  $0 < \alpha < \beta$ ) or an Erlang function  $\phi(u) = \alpha u e^{-\beta u}$  (with  $0 < \alpha < \beta^2$ ). This result has been derived thanks to an approach introduced by Nourdin and Peccati [65] which combines Malliavin's calculus with Stein's method.

*Remark 4.1.1.* The combined use of the Poisson embedding with the Nourdin-Peccati approach for Hawkes processes appeared for the first time in [91]. However the  $O(\frac{1}{\sqrt{T}})$  upper bound was proven in chapter 3 for the first time.

In this article we prove a generalization of the quantification result to the multivariate case using a version of Malliavin's calculus that is adapted to higher dimensions. Following the lines of [74], we prove an upper bound on the distance between a vector of divergences with respect to the multivariate Poisson process and a given centered multivariate Gaussian in the  $d_2$  metric. Unlike the Wasserstein metric which remains relevant in the multivariate normal space [65], the  $d_2$  metric is more suitable for the multi-dimensional Poisson space.

After defining compound multivariate Hawkes process as the result of the thinning of Poisson measures (as introduced in [18]) in section 4.2.3, we introduce the elements of multivariate Malliavin's calculus on the Hawkes process in sections 4.2.4 and 4.2.5. Finally, in section 4.3.1, we prove a general bound on a class of multivariate Hawkes functionals. As a first application in section 4.3.2 we give a bound on the Wasserstein distance between the multivariate and multimarginal normalized martingale and its Gaussian limits. Then we use those results to show that if the kernels take the exponential form

$$\phi_{ij}(u) = \alpha_{ij} e^{-\beta_{ij} u},$$

and under the following assumptions

**Assumption 4.1.2** (Stability 1). The spectral radius of the matrix  $B^{-1} A \text{diag}(m^1, \dots, m^d)$  satisfies

$$\rho \left( B^{-1} A \text{diag}(m^1, \dots, m^d) \right) < 1,$$

with  $A = (\alpha_{ij})_{ij}$ ,  $B^{-1} = \text{diag}(\frac{1}{\beta_1}, \dots, \frac{1}{\beta_d})$  and  $(m^1, \dots, m^d) = (\mathbb{E}[Y^1], \dots, \mathbb{E}[Y^d])$ .

**Assumption 4.1.3** (Stability 2). Set  $V = B - A \operatorname{diag}(m^1, \dots, m^d)$ . The eigenvalues of  $V$  are positive.

**Assumption 4.1.4** (Third moment). The measures  $\nu^1, \dots, \nu^d$  of  $Y_1^1, \dots, Y_1^d$  have finite third moments.

the following result holds:

*Theorem 4.1.5.* Let  $\mathbf{L}_T$  be a multivariate compound Hawkes process.

Set  $\mathbf{Y}'_T = \frac{\mathbf{L}_T - \operatorname{diag}(m^1, \dots, m^d)(B - A \operatorname{diag}(m^1, \dots, m^d))^{-1} B \boldsymbol{\mu} T}{\sqrt{T}}$ .

Set  $\tilde{C} = (J\sqrt{C})(J\sqrt{C})^\top = J C J^\top$ , where  $J = (I_d - \operatorname{diag}(m^1, \dots, m^d) B^{-1} A)^{-1}$  and  $C = \operatorname{diag}(\sigma_1^2, \dots, \sigma_d^2)$  where for any  $j = 1, \dots, d$

$$\sigma_j^2 = \int x^2 \nu^j(dx) \left[ (B - A \operatorname{diag}(m^1, \dots, m^d))^{-1} B \boldsymbol{\mu} \right]^j.$$

Let  $\tilde{\mathbf{G}} \sim \mathcal{N}(0, \tilde{C})$ . Then, under the Assumptions 4.1.2, 4.1.3 and 4.1.4, there exists a constant  $K$  that does not depend on  $T$  such that

$$d_2(\mathbf{Y}'_T, \tilde{\mathbf{G}}) \leq \frac{K}{\sqrt{T}},$$

for any  $T > 0$ .

*Proof.* Cf. section 4.3.3. □

*Remark 4.1.6.* Assumption 4.1.2 is a special case of the assumption

$$\rho \left( \|\phi\|_1 \operatorname{diag}(m^1, \dots, m^d) \right) < 1,$$

since in the case of an exponential kernel one has

$$\|\phi\|_1 = (\|\phi_{ij}\|_1)_{ij} = \left( \|\alpha_{ij} e^{\beta_i \cdot}\|_1 \right)_{ij} = \left( \frac{\alpha_{ij}}{\beta_i} \right)_{ij} = B^{-1} A.$$

In the case of a univariate compound Hawkes process ( $d = 1$ ), this is equivalent to assuming

$$\|\phi\|_1 m = \frac{\alpha}{\beta} m < 1.$$

*Remark 4.1.7.* If the memory parameters  $(\beta_i)_{i=1, \dots, d}$  are the same for every particle (i.e.  $\phi_{ij}(u) = \alpha_{ij} e^{-\beta u}$ ), Assumptions 4.1.2 and 4.1.3 become equivalent.

For instance, we take the bivariate compound Hawkes process  $\mathbf{L}_T = (L_T^1, L_T^2)$  where

$$\begin{cases} L_t^1 = \sum_{k=1}^{N_t^1} Y_k^1, \\ L_t^2 = \sum_{k=1}^{N_t^2} Y_k^2, \end{cases}$$

with *i.i.d* claims  $(Y_k^i)_{i=1,2,k \in \mathbb{N}}$  with common exponential distribution  $\mathcal{E}(1)$ . The intensities are assumed to follow the dynamics

$$\lambda_t = \boldsymbol{\mu} + \int_{[0,t)} A e^{-\beta(t-s)} d\mathbf{L}_s.$$

For stability, it is enough to choose the parameters such that  $\rho(A) < \beta$ . For instance, by setting

$$A = \begin{pmatrix} \frac{1}{2} & 2 \\ 2 & \frac{1}{2} \end{pmatrix},$$

any  $\beta > \frac{5}{2}$  satisfies Assumptions 4.1.2 and 4.1.3. Assumption 4.1.4 is satisfied because the exponential distribution has moments of every order.

Figure 4.1 illustrates the convergence of the compensated loss (*cf.* Theorem 4.1.5) to its centered Gaussian limit  $\mathcal{N}(0, \tilde{C})$  (also defined in Theorem 4.1.5)

On the other hand, showing the speed of convergence in the  $d_2$  metric is a bit more complicated, mainly for two reasons:

1. We do not know for which function the upper bound is reached for the sake of simulation.
2. Since the quantity  $\mathbb{E}[f(\mathbf{Y}'_T)]$  cannot be computed directly, it has to be approximated with a Monte Carlo estimation  $\frac{1}{n} \sum_{k=1}^n f(\mathbf{Y}'_T^k)$ . This means that for large times, the term in  $\frac{1}{\sqrt{T}}$  can be eclipsed by the slow decay of the Monte Carlo estimator which is in  $\frac{1}{\sqrt{n}}$ .

We can still illustrate the behavior of  $\frac{1}{n} \sum_{k=1}^n f(\mathbf{Y}'_T^k) - \mathbb{E}[f(\mathbf{G})]$  for  $\mathbf{G} \sim \mathcal{N}(0, \tilde{C})$ , where  $f$  is a "well behaved" function for which  $\mathbb{E}[f(\mathbf{G})]$  is known explicitly. For instance, Figure 4.2 shows the evolution in time for  $f(\mathbf{x}) = e^{-\frac{1}{4}\|\mathbf{x}\|^2}$ .

## 4.2 Notations and preliminaries

In this section we generalize the mathematical framework introduced in Section 2 of chapter 3. We then proceed to define the multivariate Hawkes loss as the result of a thinning procedure from a Poisson measure. We finally recall some elements of Stein's method.

### 4.2.1 General notations

1. Let  $d$  be an integer. For any two vectors  $\mathbf{u} = (u^i)_{i=1,\dots,d}$  and  $\mathbf{v} = (v^i)_{i=1,\dots,d}$  in  $\mathbb{R}^d$ , the product  $\mathbf{u}\mathbf{v}$  is defined as the vector of  $\mathbb{R}^d$  such that  $\mathbf{u}\mathbf{v} = (u^i v^i)_{i=1,\dots,d}$ .
2. Let  $\theta \in \mathbb{R}$  and  $\mathbb{1}_{\theta \leq u^i}$  be the indicator of the set  $\{\theta \leq u^i\}$ . We define  $\mathbf{1}_{\theta \leq \mathbf{u}}$  as the element of  $\mathbb{R}^d$  such that  $\mathbf{1}_{\theta \leq \mathbf{u}} = (\mathbb{1}_{\theta \leq u^i})_{i=1,\dots,d}$ .
3. Similarly, if  $\boldsymbol{\nu}$  is a vector of  $d$  measures and  $\mathbf{f}$  is a vector of  $d$  functions, we define  $\int \mathbf{f}(x) \boldsymbol{\nu}(dx) = (\int f^i(x) \nu^i(dx))_{i=1,\dots,d}$ .



4. The inner product  $\langle \cdot, \cdot \rangle$  corresponds to the Euclidean inner product and  $\|\cdot\|$  is its norm.
5. The operator norm of a matrix  $A \in \mathcal{M}_d(\mathbb{R})$  is

$$\|A\|_{op} := \sup_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|.$$

6. For every function  $g : \mathbb{R}^d \rightarrow \mathbb{R}$ , let

$$\|g\|_{Lip} := \sup_{\mathbf{x} \neq \mathbf{y}} \frac{|g(\mathbf{x}) - g(\mathbf{y})|}{\|\mathbf{x} - \mathbf{y}\|}.$$

7. If  $g \in \mathcal{C}^1(\mathbb{R}^d)$  (continuously differentiable) then we write

$$M_2(g) := \sup_{\mathbf{x} \neq \mathbf{y}} \frac{\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|}{\|\mathbf{x} - \mathbf{y}\|}.$$

8. Similarly, if  $g \in \mathcal{C}^2(\mathbb{R}^d)$  (twice continuously differentiable) then

$$M_3(g) := \sup_{\mathbf{x} \neq \mathbf{y}} \frac{\|\text{Hess } f(\mathbf{x}) - \text{Hess } f(\mathbf{y})\|_{op}}{\|\mathbf{x} - \mathbf{y}\|}.$$

#### 4.2.2 Elements of stochastic analysis on the multivariate Poisson space

Let  $d$  be a positive integer. Let  $\nu^1, \dots, \nu^d$  be a family of integrable probability measures on  $\mathbb{R}_+$  such that  $\nu^i(\{0\}) = 0$  for every  $i = 1, \dots, d$ , and define  $m^i = \int_{\mathbb{R}_+} x \nu^i(dx)$ .

In this section, every component of the multivariate compound Hawkes process is obtained through the thinning of 3-component Poisson measure.

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

$$\Omega := \left\{ \omega = \sum_{i=1}^n \delta_{(t_i, \theta_i, x_i)}, 0 = t_0 < t_1 < \dots < t_n, (\theta_i, x_i) \in \mathbb{R}_+^2, n \in \mathbb{N} \cup \{+\infty\} \right\}.$$

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

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

where

$$\Pi^j([0, t] \times [0, b] \times [0, y]) (\omega^j) := \omega^j([0, t] \times [0, b] \times [0, y]), \quad (t, b, y) \in \mathbb{R}_+^3, \quad j = 1, \dots, d$$

is a family of independent homogeneous Poisson processes with intensity measures  $dt \otimes d\theta \otimes d\nu^j$ , that is,

$$\begin{aligned} \mathbb{P}[\Pi^1([0, t_1] \times [0, b_1] \times [0, y_1]) = n_1, \dots, \Pi^d([0, t_d] \times [0, b_d] \times [0, y_d]) = n_d] \\ = \prod_{j=1}^d \frac{(t_j b_j \nu_j([0, y_j]))^{n_j} \exp(-t_j b_j \nu_j([0, y_j]))}{n_j!}. \end{aligned}$$

We set  $\mathbb{F}^\Pi = (\mathcal{F}_t^\Pi)_{t \geq 0}$  to be the natural filtration of  $(\Pi^j)_{j=1, \dots, d}$ . The expectation with respect to  $\mathbb{P}$  is denoted by  $\mathbb{E}[\cdot]$ . For  $t \geq 0$ , we denote by  $\mathbb{E}_t[\cdot]$  the conditional expectation  $\mathbb{E}[\cdot | \mathcal{F}_t^\Pi]$ .

*Remark 4.2.1.* It is possible to define all the Poisson measures on the same probability space  $\Omega^d$  by taking

$$\tilde{\Pi}^j([0, t] \times [0, b] \times [0, y]) (\omega) = \Pi^j([0, t] \times [0, b] \times [0, y]) (\omega^j), \quad (t, b, y) \in \mathbb{R}_+^3,$$

for any  $\omega = (\omega^1, \dots, \omega^d) \in \Omega^d$ .

From now on we confound  $(\tilde{\Pi}^j)_{j=1, \dots, d}$  with  $(\Pi^j)_{j=1, \dots, d}$ .

We now generalize the operators defined in chapter 3 for the 1-dimensional setting. We start with the component-wise shift operator.

**Definition 4.2.2** (Shift operator). Let  $j \in \llbracket 1, d \rrbracket$ . We define for  $(t, \theta, x)$  in  $\mathbb{R}_+^3$  the measurable maps

$$\begin{aligned} \varepsilon_{(t, \theta, x)}^{j+} : \Omega^d &\rightarrow \Omega^d \\ (\omega^1, \dots, \omega^j, \dots, \omega^d) &\mapsto (\omega^1, \dots, \varepsilon_{(t, \theta, x)}^+(\omega^j), \dots, \omega^d), \end{aligned}$$

where for any  $A$  in  $\mathcal{B}(\mathbb{R}_+^3)$

$$\left( \varepsilon_{(t, \theta, x)}^+(\omega^j) \right) (A) := \omega^j (A \setminus (t, \theta, x)) + \mathbb{1}_A(t, \theta, x),$$

with

$$\mathbb{1}_A(t, \theta, x) := \begin{cases} 1, & \text{if } (t, \theta, x) \in A, \\ 0, & \text{else.} \end{cases}$$

**Definition 4.2.3** (Malliavin's derivative). For  $F$  in  $\mathcal{L}^2(\Omega^d, \mathcal{F}_\infty, \mathbb{P})$ , we define  $D^j F$  the Malliavin's derivative of  $F$  as

$$D_{(t, \theta, x)}^j F := F \circ \varepsilon_{(t, \theta, x)}^{j+} - F, \quad (t, \theta, x) \in \mathbb{R}_+^3.$$

If  $\mathbf{F} = (F^1, \dots, F^n)$  for some  $n \geq 2$  where  $F^i \in \mathcal{L}^2(\Omega^d, \mathcal{F}_\infty, \mathbb{P}) \quad \forall i = 1, \dots, n$ ,

$$D_{(t, \theta, x)}^j \mathbf{F} := (F^i \circ \varepsilon_{(t, \theta, x)}^{j+} - F^i)_{i=1, \dots, n}, \quad (t, \theta, x) \in \mathbb{R}_+^3$$

**Definition 4.2.4.** Let  $\mathcal{I}$  be the sub-sigma field of  $\mathcal{B}(\mathbb{R}_+^3) \otimes \mathcal{F}^N$  of stochastic processes  $Z := (Z_{(t, \theta, x)})_{(t, \theta, x) \in \mathbb{R}_+^3}$  in  $\mathcal{L}^1(\Omega \times \mathbb{R}_+^3, \mathbb{P} \otimes dt \otimes d\theta \otimes \nu)$  such that

$$D_{(t, \theta, x)}^j Z_{(t, \theta, x)} = 0, \quad \text{for a.a. } (t, \theta, x) \in \mathbb{R}_+^3, \quad \text{for all } j = 1, \dots, d.$$

**Definition 4.2.5** (Divergence operator). We set  $\mathcal{S}$  the set of stochastic processes  $Z := (Z_{(t,\theta,x)})_{(t,\theta,x) \in \mathbb{R}_+^3}$  in  $\mathcal{I}$  such that for every  $j \in \llbracket 1, d \rrbracket$ :

$$\mathbb{E} \left[ \int_{\mathbb{R}_+^3} |Z_{(t,\theta,x)}|^2 dt d\theta \nu^j(dx) \right] + \mathbb{E} \left[ \left( \int_{\mathbb{R}_+^3} Z_{(t,\theta,x)} \Pi^j(dt, d\theta, dx) \right)^2 \right] < +\infty,$$

where  $\int_{\mathbb{R}_+^3} Z_{(t,\theta,x)} \Pi^j(dt, d\theta, dx)$  is understood in the sense of the Stieltjes integral.

For  $Z$  in  $\mathcal{S}$ , we set the divergence operator with respect to  $\Pi^j$  as

$$\delta^j(Z) := \int_{\mathbb{R}_+^3} Z_{(t,\theta,x)} \Pi^j(dt, d\theta, dx) - \int_{\mathbb{R}_+^3} Z_{(t,\theta,x)} dt d\theta \nu^j(dx). \quad (4.2.1)$$

For a vector  $\mathbf{Z} \in \mathcal{S}^d$ , the divergence operator with respect to  $\mathbf{\Pi}$  is defined as

$$\delta^{\mathbf{\Pi}}(\mathbf{Z}) = \sum_{i=1}^d \delta^i(Z^i),$$

and if  $Z = (\mathbf{Z}^1, \dots, \mathbf{Z}^n) \in \mathcal{S}^{d \times n}$ , is a matrix, its divergence is defined as

$$\begin{aligned} \delta^{\mathbf{\Pi}}(Z) &= (\delta^{\mathbf{\Pi}}(\mathbf{Z}^1), \dots, \delta^{\mathbf{\Pi}}(\mathbf{Z}^n)), \\ &= \sum_{i=1}^d (\delta^i(Z^{i1}), \dots, \delta^i(Z^{in})). \end{aligned}$$

### 4.2.3 Definition of the multivariate compound Hawkes process

In this subsection we give a definition of the multivariate compound Hawkes process in the Markov framework. Consider the events times  $\tau_1^j, \tau_2^j, \dots$  associated with the  $j$ -th component and define the counting process

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

Now assume that each event  $\tau_i^j$  corresponds to a random "loss"  $Y_i^j \sim \nu^j$  such that the variables  $(Y_i^j)_{i \geq 1}$  are independent and identically distributed (*i.i.d.*).

The compound process  $L_t^j$  of the total loss attributed to the  $j$ -th component is defined as

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

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

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

which measures how likely it is for  $N^j$  to jump between  $t$  and  $t + dt$ , right before  $t$ . The process  $\mathbf{L}_t = (L_t^1, \dots, L_t^d)_{t \geq 0}$  is called a multivariate compound Hawkes process if its intensity vector  $\boldsymbol{\lambda}_t$  follows the dynamics

$$\begin{aligned} \lambda_t^j &= \mu^j + \int_{[0,t)} \sum_{k=1}^d \phi_{jk}(t-s) dL_s^k, \\ &= \mu^j + \sum_{k=1}^d \sum_{\tau_i^k < t} \phi_{jk}(t - \tau_i^k) Y_i^k, \end{aligned}$$

where  $(\mu^1, \dots, \mu^d) \in \mathbb{R}_+^d$  and  $(\phi_{i,j})_{i,j=1,\dots,d}$  are non-negative integrable functions.

In this article, we restrict ourselves to the case where the kernels  $\phi$  are a family of exponential functions

$$\phi_{jk}(u) = \alpha_{jk} e^{-\beta_j u}$$

where  $(\beta_j)_{j=1,\dots,d} \in (\mathbb{R}_+)^d$  and  $A = (\alpha_{ij})_{i,j=1,\dots,d} \in \mathcal{M}_d(\mathbb{R}_+)$ . The intensity can be expressed under matrix form

$$\boldsymbol{\lambda}_t = \boldsymbol{\mu} + \int_{[0,t)} e^{-B(t-s)} A d\mathbf{L}_s, \quad (4.2.2)$$

with  $B = \text{diag}(\beta_1, \dots, \beta_d)$ . Furthermore, we recall Assumptions 4.1.2, 4.1.3 and 4.1.4.

We now introduce an equivalent definition of the multivariate compound Hawkes process, presented as the result of a stochastic differential equation (SDE) with respect to a family of random Poisson measures. This thinning result has been introduced in [18].

*Theorem 4.2.6.* Let  $\boldsymbol{\Pi} = (\Pi^1, \dots, \Pi^d)$  be a family of independent Poisson measures (as presented in Section 4.2.2). Let  $\boldsymbol{\mu} \in \mathbb{R}_+^d$  and  $A \in \mathcal{M}_d(\mathbb{R}_+)$  and  $(\beta_1, \dots, \beta_d) \in \mathbb{R}_+^d$  such that the Assumption 4.1.2 is verified. The SDE below admits a unique solution  $(\mathbf{L}, \mathbf{N}, \boldsymbol{\lambda})$  with  $\mathbf{H}$  and  $\mathbf{L}$  (resp.  $\boldsymbol{\lambda}$ )  $\mathbb{F}^{\boldsymbol{\Pi}}$ -adapted (resp.  $\mathbb{F}^{\boldsymbol{\Pi}}$ -predictable)

$$\begin{cases} L_t^j = \int_{(0,t] \times \mathbb{R}_+^2} x \mathbb{1}_{\{\theta \leq \lambda_s^j\}} \Pi^j(ds, d\theta, dx), & t \geq 0, \quad j = 1, \dots, d, \\ N_t^j = \int_{(0,t] \times \mathbb{R}_+^2} \mathbb{1}_{\{\theta \leq \lambda_s^j\}} \Pi^j(ds, d\theta, dx), & t \geq 0, \quad j = 1, \dots, d, \\ \lambda_t^j = \mu^j + \int_{(0,t)} \sum_{k=1}^d \alpha_{jk} e^{-\beta_j(t-u)} dL_u^k, & t \geq 0 \quad j = 1, \dots, d. \end{cases} \quad (4.2.3)$$

We set  $\mathbb{F}^{\mathbf{N}} := (\mathcal{F}_t^{\mathbf{N}})_{t \geq 0}$  (respectively  $\mathbb{F}^{\mathbf{L}} := (\mathcal{F}_t^{\mathbf{L}})_{t \geq 0}$ ) the natural filtration of  $\mathbf{N}$  (respectively of  $\mathbf{L}$ ) and  $\mathcal{F}_\infty^{\mathbf{N}} := \lim_{t \rightarrow +\infty} \mathcal{F}_t^{\mathbf{N}}$  (respectively  $\mathcal{F}_\infty^{\mathbf{L}} := \lim_{t \rightarrow +\infty} \mathcal{F}_t^{\mathbf{L}}$ ). Obviously  $\mathcal{F}_t^{\mathbf{N}} \subset \mathcal{F}_t^{\mathbf{L}} \subset \mathcal{F}_t^{\boldsymbol{\Pi}}$  as  $\mathbf{N}$  is completely determined by the jump times of  $\mathbf{N}$  which are exactly those of  $\mathbf{L}$ .

*Proof.* This is merely a  $d$ -dimensional version of Theorem 4.2.6 in chapter 3.  $\square$

**Definition 4.2.7.** Let  $\mathbf{L}$  and  $\boldsymbol{\lambda}$  be the processes defined in Theorem 4.2.6. The multivariate Hawkes martingale  $\mathbf{M}$  is the process defined as

$$\mathbf{M}_T = \mathbf{L}_T - \text{diag} \left( m^1, \dots, m^d \right) \int_0^T \boldsymbol{\lambda}_t dt$$

for any  $T \geq 0$ .

#### 4.2.4 Malliavin's analysis of the multivariate compound Hawkes process

This subsection is a generalization of Section 3.3 of chapter 3 where we examine the impact of the  $i$ -th Malliavin's derivative on the  $j$ -th component of  $\mathbf{L}$ ,  $\mathbf{N}$  and  $\boldsymbol{\lambda}$ .

*Proposition 4.2.8* (3.2.15 in chapter 3). Let  $t$  and  $v$  in  $\mathbb{R}_+$  and  $(\theta, \theta_0, x)$  in  $\mathbb{R}_+^3$ . For every  $j \in \llbracket 1, \dots, d \rrbracket$  it holds that :

$$\begin{aligned} & \mathbb{1}_{\{\theta \leq \lambda_t^j\}} \mathbb{1}_{\{\theta_0 \leq \lambda_t^j\}} (\mathbf{L}_v \circ \varepsilon_{(t, \theta, x)}^{j+}, \boldsymbol{\lambda}_v \circ \varepsilon_{(t, \theta, x)}^{j+})_{v \geq 0} \\ &= \mathbb{1}_{\{\theta \leq \lambda_t^j\}} \mathbb{1}_{\{\theta_0 \leq \lambda_t^j\}} (\mathbf{L}_v \circ \varepsilon_{(t, \theta_0, x)}^{j+}, \boldsymbol{\lambda}_v \circ \varepsilon_{(t, \theta_0, x)}^{j+})_{v \geq 0}. \end{aligned}$$

which entails that for every  $\mathcal{F}_\infty^{\mathbf{L}}$ -measurable random variable  $\mathbf{F}$

$$\mathbb{1}_{\{\theta \leq \lambda_t^j\}} \mathbb{1}_{\{\theta_0 \leq \lambda_t^j\}} D_{(t, \theta, x)}^j \mathbf{F} = \mathbb{1}_{\{\theta \leq \lambda_t^j\}} \mathbb{1}_{\{\theta_0 \leq \lambda_t^j\}} D_{(t, \theta_0, x)}^j \mathbf{F}, \quad \mathbb{P} - a.s.$$

where the derivative is applied to each component of  $\mathbf{F}$ . The last equation allows us to define

$$D_{(t, \lambda_t^j, x)}^j \mathbf{F} := \mathbb{1}_{\{\theta = \lambda_t^j\}} D_{(t, \theta, x)}^j \mathbf{F}, \quad \forall (\theta, x) \in \mathbb{R}_+^2$$

as well as its vector version

$$\begin{aligned} D_{(t, \boldsymbol{\lambda}_t, x)} \mathbf{F} &= (D_{(t, \lambda_t^j, x)}^j \mathbf{F})_{j=1, \dots, d} \\ &:= (\mathbb{1}_{\{\theta = \lambda_t^j\}} D_{(t, \theta, x)}^j \mathbf{F})_{j=1, \dots, d}, \quad \forall (\theta, x) \in \mathbb{R}_+^2. \end{aligned}$$

*Proof.* For each  $i \in \llbracket 1, d \rrbracket$  and  $v \geq t$ , we have conditionally on  $\theta_0 \leq \lambda_t^j$

$$L_v^i \circ \varepsilon_{(t, \theta_0, x)}^{+j} = L_{t-}^i + x \mathbb{1}_{i=j} + \int_{(t, v) \times \mathbb{R}_+^2} y \mathbb{1}_{\{\theta \leq \lambda_u^i \circ \varepsilon_{(t, \theta_0, x)}^{+j}\}} \Pi^i(d\theta, du, dy),$$

and  $L_v^i \circ \varepsilon_{(t, \theta_0, x)}^{+j} = L_v^i$  if  $t > v$ . Similarly for  $\boldsymbol{\lambda}$ , we have

$$\begin{aligned} \lambda_v^i \circ \varepsilon_{(t, \theta_0, x)}^{+j} &= \left( \mu_i + \int_{(0, t)} \sum_{k=1}^d e^{-\beta(v-u)} \alpha_{ik} dL_u^k + \int_{[t, v]} \sum_{k=1}^d e^{-\beta(v-u)} \alpha_{ik} dL_u^k \right) \circ \varepsilon_{(t, \theta_0, x)}^{+j} \\ &= \mu_i + \int_{(0, t)} \sum_{k=1}^d e^{-\beta(v-u)} \alpha_{ik} dL_u^k + x \alpha_{ij} e^{-\beta(v-t)} + \int_{(t, v)} \sum_{k=1}^d e^{-\beta(v-u)} \alpha_{ik} d(L_u^k \circ \varepsilon_{(t, \theta_0, x)}^{+j}). \end{aligned}$$

In other words,  $(\mathbf{L} \circ \varepsilon_{(t,\theta_0,x)}^{+j}, \boldsymbol{\lambda} \circ \varepsilon_{(t,\theta_0,x)}^{+j})$  solves the same (path-wise and in the SDE sense) equation for any  $\theta_0$  such that  $\theta_0 \leq \lambda_t$ .

For the second equality, *cf.* the proof of Proposition 3.2.15.  $\square$

*Remark 4.2.9.* It is also possible to include the process  $\mathbf{N}$  in these results. From now on,  $\mathbf{N}$  will be omitted and we will focus exclusively on  $\mathbf{L}$  and  $\boldsymbol{\lambda}$ .

Now we give the Malliavin's derivative of the multivariate compound Hawkes process as well as its intensity. To do so we start with introducing some notations. The vector  $\mathbf{e}_i$  is the element of  $\mathbb{R}^d$  that has 1 in the  $i$ -th component and zero elsewhere.

*Proposition 4.2.10.* Let  $t \geq 0$  and  $x \in \mathbb{R}$ . For every  $j \in \llbracket 1, \dots, d \rrbracket$ , we have

$$(D_{(t,\lambda_t^j,x)}^j \mathbf{L}_s, D_{(t,\lambda_t^j,x)}^j \boldsymbol{\lambda}_s) = \begin{cases} (x\mathbf{e}_j + \hat{\mathbf{L}}_s^{j,t,x}, \hat{\boldsymbol{\lambda}}_s^{j,t,x}), & s \geq t, \\ (0, 0), & s < t, \end{cases}$$

where the equality is understood path-wise and in the SDE sense and where  $(\hat{\mathbf{L}}_s^{j,t,x}, \hat{\boldsymbol{\lambda}}_s^{j,t,x})_{s \geq t} = ((\hat{L}_s^{i,j,t,x}, \hat{\lambda}_s^{i,j,t,x})_{s \geq t})_{i=1,\dots,d}$  is the unique solution to the SDE

$$\begin{cases} \hat{L}_s^{i,j,t,x} = \int_{(t,s] \times \mathbb{R}_+^2} y \mathbf{1}_{\{\lambda_u^i \leq \theta \leq \lambda_u^i + \hat{\lambda}_u^{i,j,t,x}\}} \Pi^i(du, d\theta, dy), & s \geq t, \\ \hat{\lambda}_s^{i,j,t,x} = x\alpha_{ij}e^{-\beta_i(s-t)} + \sum_{k=1}^d \int_{(t,s)} \alpha_{ik} e^{-\beta_i(s-u)} d\hat{L}_u^{k,j,t,x}, & s > t, \hat{\lambda}_t^{i,j,t,x} = x\alpha_{ij}, \end{cases} \quad (4.2.4)$$

or under the matrix form

$$\begin{cases} \hat{\mathbf{L}}_s^{j,t,x} = \int_{(t,s] \times \mathbb{R}_+^2} y \mathbf{1}_{\{\lambda_u \leq \theta \leq \lambda_u + \hat{\lambda}_u^{j,t,x}\}} \mathbf{\Pi}(du, d\theta, dy), & s \geq t, \\ \hat{\boldsymbol{\lambda}}_s^{j,t,x} = xe^{-B(s-t)} \mathbf{A}_{\cdot j} + \int_{(t,s)} e^{-B(s-u)} \mathbf{A} d\hat{\mathbf{L}}_u^{j,t,x}, & s > t, \hat{\boldsymbol{\lambda}}_t^{j,t,x} = x\mathbf{A}_{\cdot j}, \end{cases} \quad (4.2.5)$$

where the first integral is understood as a vector of integrals with respect to each Poisson measure and the indicator of a vector is understood as a vector of indicators. The process  $(\hat{\mathbf{L}}_s^{j,t,x}, \hat{\boldsymbol{\lambda}}_s^{j,t,x})_{s \in [t, +\infty)}$  is a generalized multivariate compound Hawkes process, with an intensity vector that is not bounded away from 0.

*Proof.* It is sufficient to apply the same procedure of the proof of Proposition 3.2.17 to each component of the vectors. Note that in this proposition, the subscript contains  $x$ , because in our case the size of the jump has an impact on  $\hat{\boldsymbol{\lambda}}$ 's behavior, since it is an integral with respect to  $d\hat{\mathbf{L}}$  and not merely  $d\tilde{\mathbf{N}}$ .  $\square$

*Remark 4.2.11.* Note that the process  $(\hat{\mathbf{L}}_s^{j,t,x}, \hat{\boldsymbol{\lambda}}_s^{j,t,x})_{s \geq t}$  defined above is equal in distribution to the (generalized) Hawkes process  $(\tilde{\mathbf{L}}_s^{j,t,x}, \tilde{\boldsymbol{\lambda}}_s^{j,t,x})_{s \geq t}$  defined as a solution to the SDE

$$\begin{cases} \tilde{\mathbf{L}}_s^{j,t,x} = \int_{(t,s] \times \mathbb{R}_+^2} y \mathbf{1}_{\{0 \leq \theta \leq \tilde{\boldsymbol{\lambda}}_u^{j,t,x}\}} \tilde{\boldsymbol{\Pi}}(du, d\theta, dy), & s \geq t, \\ \tilde{\boldsymbol{\lambda}}_s^{j,t,x} = x e^{-B(s-t)} \mathbf{A}_j + \int_{(t,s)} e^{-B(s-u)} \text{Ad} \tilde{\mathbf{L}}_u^{j,t,x}, & s > t, \quad \tilde{\boldsymbol{\lambda}}_t^{j,t,x} = x \mathbf{A}_j, \end{cases} \quad (4.2.6)$$

where  $\tilde{\boldsymbol{\Pi}}$  is a family of Poisson measures independent from  $\boldsymbol{\Pi}$  but of the same distributions.

#### 4.2.5 Multivariate Stein's method

Stein's method is based on an alternative characterization of the Gaussian distribution. The combination of this characterization with elements of Malliavin's calculus (known as the Nourdin-Peccati approach) provides us with an estimation of the distance between a random variable and a Gaussian.

*Lemma 4.2.12.* Let  $C$  be a  $d \times d$  real symmetric positive definite matrix and  $\mathbf{Y}$  an  $\mathbb{R}^d$  random variable. Then  $\mathbf{Y} \sim \mathcal{N}_d(0, C)$  if and only if for every twice differentiable function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  such that  $\mathbb{E}[|\langle C, \text{Hess } f(\mathbf{Y}) \rangle_{H.S}| + |\langle \mathbf{Y}, \nabla f(\mathbf{Y}) \rangle|] < +\infty$  we have

$$\mathbb{E}[\langle C, \text{Hess } f(\mathbf{Y}) \rangle_{H.S} - \langle \mathbf{Y}, \nabla f(\mathbf{Y}) \rangle] = 0$$

where  $\langle A, B \rangle_{H.S} = \text{tr}(AB^T) = \sum A_{ij} B_{ij}$  is the Hilbert-Schmidt matrix inner product and Hess is the Hessian operator.

*Proof.* Cf. Lemma 2 in [20] for example. □

This shows that if for a sufficiently large class of functions  $f$  one has

$$\mathbb{E}[\langle C, \text{Hess } f(\mathbf{Y}) \rangle_{H.S} - \langle \mathbf{Y}, \nabla f(\mathbf{Y}) \rangle] \simeq 0$$

in some sense, then the random variable  $\mathbf{Y}$ 's distribution is fairly "close" to  $\mathcal{N}_d(0, C)$ . We now introduce the function classes as well as the metric that we will use to quantify how close the given distance of a random variable to a Gaussian distribution.

**Definition 4.2.13.** 1. The distance  $d_2$  between two integrable random variables  $\mathbf{X}$  and  $\mathbf{Y}$  is given by

$$d_2(\mathbf{X}, \mathbf{Y}) := \sup_{g \in \mathcal{H}} |\mathbb{E}[g(\mathbf{X})] - \mathbb{E}[g(\mathbf{Y})]|,$$

where

$$\mathcal{H} := \{g \in \mathcal{C}^2(\mathbb{R}^d), \text{ such that } \|g\|_{Lip} \leq 1 \text{ and } M_2(g) \leq 1\}.$$

2. Set  $C \in \mathcal{S}_d^{++}(\mathbb{R})$  and  $g \in \mathcal{H}$  and let  $\mathbf{G} \sim \mathcal{N}_d(0, C)$ . We define  $U_0g$  as the function

$$U_0g(\mathbf{x}) = \int_0^1 \frac{1}{2t} \mathbb{E} \left[ g(\sqrt{t}\mathbf{x} + \sqrt{1-t}\mathbf{G}) - g(\mathbf{G}) \right] dt.$$

According to [20],  $U_0g$  is a solution to the differential equation (with unknown  $f$ )

$$g(\mathbf{x}) - \mathbb{E}[g(\mathbf{G})] = \langle C, \text{Hess } f(\mathbf{x}) \rangle_{H.S} - \langle \mathbf{x}, \nabla f(\mathbf{x}) \rangle,$$

such that

$$\sup_{\mathbf{x} \in \mathbb{R}^d} \|U_0g(\mathbf{x})\| \leq \|C^{-1}\|_{op} \|C\|_{op}^{1/2} \text{ and } M_3(U_0g) \leq \frac{\sqrt{2\pi}}{4} \|C^{-1}\|_{op}^{3/2} \|C\|_{op}$$

Combining the two points in the previous definition, it is possible to prove (following the lines of Lemma 2.17 in [74]) the following upper bound of the  $d_2$  distance between a centered variable  $\mathbf{F}$  and  $\mathbf{G} \sim \mathcal{N}_d(0, C)$ :

$$d_2(\mathbf{F}, \mathbf{G}) \leq \sup_{f \in \mathcal{F}_C} \left| \mathbb{E} \left[ \langle C, \text{Hess } f(\mathbf{F}) \rangle_{H.S} - \langle \mathbf{F}, \nabla f(\mathbf{F}) \rangle \right] \right|,$$

where  $\mathcal{F}_C$  is the functional space

$$\mathcal{F}_C := \{g \in \mathcal{C}^2(\mathbb{R}^d), \text{ such that } M_2(g) \leq \|C^{-1}\|_{op} \|C\|_{op}^{1/2}, M_3(g) \leq \frac{\sqrt{2\pi}}{4} \|C^{-1}\|_{op}^{3/2} \|C\|_{op}\}.$$

*Remark 4.2.14.* Unlike the univariate Poisson space case (*cf.* chapter 3) or the multivariate variables on a Gaussian space [67], the Wasserstein distance  $d_W$  is not well adapted to our computations. Nevertheless, if  $d_2(\mathbf{F}_T, \mathbf{G}) \rightarrow 0$  as  $T$  goes to infinity we have that  $\mathbf{F}_T$  converges to the Gaussian  $\mathbf{G}$  in distribution (*cf.* Remark 2.16 in [74]).

## 4.3 Main results

### 4.3.1 General bound

In this section, we give a bound for the  $d_2$  distance between any random variable that can be expressed as a divergence with respect to the Poisson measure and the Gaussian distribution.

*Lemma 4.3.1.* Let  $n$  be an integer and let  $\mathbf{F} = (F^1, \dots, F^n)$  be an  $\mathcal{F}_\infty^L$ -measurable variable such that  $\mathbb{E}[\|\mathbf{F}\|^2] \leq +\infty$ . For all  $\phi \in \mathcal{C}^3(\mathbb{R}^n)$  with bounded derivatives, for any  $k = 1, \dots, n$ , there exists a random  $\bar{\mathbf{F}}$  such that

$$D_{(t,\theta,x)}^k \partial_i \phi(\mathbf{F}) = \langle \nabla \partial_i \phi(\mathbf{F}), D_{(t,\theta,x)}^k \mathbf{F} \rangle + \frac{1}{2} \langle D_{(t,\theta,x)}^k \mathbf{F}, \text{Hess } \partial_i \phi(\bar{\mathbf{F}}) D_{(t,\theta,x)}^k \mathbf{F} \rangle, \quad (t, \theta, x) \in \mathbb{R}_+^3$$

where

$$|\langle \mathbf{y}, \text{Hess } \partial_k \phi(\bar{\mathbf{F}}) \mathbf{y} \rangle| \leq \|\mathbf{y}\|^2 M_3(\phi), \quad \mathbf{y} \in \mathbb{R}^d$$



*Proof.* The equality is merely an application of the multivariate Taylor-Young Theorem, combined with the fact that

$$\begin{aligned} D_{(t,\theta,x)}^k \partial_i \phi(\mathbf{F}) &= \partial_i \phi(\mathbf{F} \circ \varepsilon_{(t,\theta,x)}^{k+}) - \partial_i \phi(\mathbf{F}), \\ &= \partial_i \phi(\mathbf{F} + D_{(t,\theta,x)}^k \mathbf{F}) - \partial_i \phi(\mathbf{F}). \end{aligned}$$

When it comes to the upper bound on the rest, we have using Cauchy-Schwarz

$$\begin{aligned} |\langle \mathbf{y}, \text{Hess } \partial_i \phi(\bar{\mathbf{F}}) \mathbf{y} \rangle| &\leq \|\mathbf{y}\| \|\text{Hess } \partial_i \phi(\bar{\mathbf{F}}) \mathbf{y}\|, \\ &\leq \|\mathbf{y}\|^2 \|\text{Hess } \partial_i \phi(\bar{\mathbf{F}})\|_{op}. \end{aligned}$$

Using Schwarz's Theorem we have that

$$\begin{aligned} \text{Hess } \partial_i \phi(\bar{\mathbf{F}}) &= \partial_i \text{Hess } \phi(\bar{\mathbf{F}}), \\ &= \lim_{h \rightarrow 0} \frac{\text{Hess } \phi(\bar{\mathbf{F}} + h \mathbf{e}_i) - \text{Hess } \phi(\bar{\mathbf{F}})}{h}, \end{aligned}$$

and the result follows using the fact that the norm  $\|\cdot\|_{op}$  is continuous.  $\square$

This lemma will be useful in proving the following result.

*Theorem 4.3.2.* Let  $n \in \mathbb{N}^*$ . Let  $Z = \left( Z_{(t,\theta,x)}^{ki} \right)_{(t,\theta,x) \in \mathbb{R}_+^3}$  be a stochastic process in  $\mathcal{S}^{d \times n}$ . Set  $\mathbf{F} = \delta^{\mathbf{N}}(Z)$  the divergence of  $Z$ . Then, letting  $\mathbf{G} \sim \mathcal{N}_d(0, C)$  (for  $C \in \mathcal{S}_n^{++}(\mathbb{R})$ ) we have

$$\begin{aligned} d_2(\mathbf{F}, \mathbf{G}) &\leq \|C^{-1}\|_{op} \|C\|_{op}^{1/2} \sum_{i,j=1}^n \mathbb{E} \left[ \left\| C_{ij} - \sum_{k=1}^d \int_{\mathbb{R}_+^3} Z_{(t,\theta,x)}^{ki} D_{(t,\theta,x)}^k F^j dt d\theta \nu^k(dx) \right\|^2 \right] \\ &\quad + \frac{\sqrt{2\pi}}{8} \|C^{-1}\|_{op}^{3/2} \|C\|_{op} \sum_{i=1}^n \sum_{k=1}^d \mathbb{E} \left[ \int_{\mathbb{R}_+^3} \left| Z_{(t,\theta,x)}^{ki} \right| \left\| D_{(t,\theta,x)}^k \mathbf{F} \right\|^2 dt d\theta \nu^k(dx) \right]. \end{aligned}$$

If in particular  $Z_{(t,\theta,x)}^{ki} = \mathbb{1}_{\theta \leq \lambda_t^k} U_{(t,x)}^{ki}$ , the upper bound takes the form

$$\begin{aligned} d_2(\mathbf{F}, \mathbf{G}) &\leq \|C^{-1}\|_{op} \|C\|_{op}^{1/2} \sum_{i,j=1}^n \mathbb{E} \left[ \left\| C_{ij} - \sum_{k=1}^d \int_{\mathbb{R}_+^2} \lambda_t^k U_{(t,x)}^{ki} D_{(t,\lambda_t^k,x)}^k F^j dt \nu^k(dx) \right\|^2 \right] \\ &\quad + \frac{\sqrt{2\pi}}{8} \|C^{-1}\|_{op}^{3/2} \|C\|_{op} \sum_{i=1}^n \sum_{k=1}^d \mathbb{E} \left[ \int_{\mathbb{R}_+^2} \lambda_t^k \left| U_{(t,x)}^{ki} \right| \left\| D_{(t,\lambda_t^k,x)}^k \mathbf{F} \right\|^2 dt \nu^k(dx) \right]. \end{aligned}$$

*Proof.* We start by recalling the definition of the  $d_2$  distance between  $\mathbf{F}$  and  $\mathbf{G}$

$$d_2(\mathbf{F}, \mathbf{G}) = \sup_{g \in \mathcal{H}} |\mathbb{E}[g(\mathbf{F}) - g(\mathbf{G})]|.$$

Using an interpolation argument (as in [67] or [20]) it can be proved that there exists a family  $g_\eta$  such that

1.  $g_\eta \in C^\infty(\mathbb{R}^n)$ .
2.  $\|g - g_\eta\|_\infty \rightarrow 0$  when  $\eta$  goes to zero.
3.  $M_2(g_\eta) \leq M_2(g)$  and  $M_3(g_\eta) \leq M_3(g)$  using Young's convolution inequality.

For any  $\epsilon > 0$ , there exists  $\eta$  such that

$$\begin{aligned} |\mathbb{E}[g(\mathbf{F}) - g(\mathbf{G})]| &\leq |\mathbb{E}[g_\eta(\mathbf{F}) - g_\eta(\mathbf{G})]| + \epsilon, \\ &= |\mathbb{E}[\langle C, \text{Hess } f(\mathbf{F}) \rangle_{H.S} - \langle \mathbf{F}, \nabla f(\mathbf{F}) \rangle]| + \epsilon, \end{aligned}$$

where  $f = U_0 g_\eta$ , defined in Definition 4.2.13.

Since  $g_\eta$  belongs to  $C^\infty(\mathbb{R}^n)$  the same holds for  $f$ , which we assume from now on. The expansion of the scalar products yields:

$$\begin{aligned} \mathbb{E}[\langle C, \text{Hess } f(\mathbf{F}) \rangle_{H.S} - \langle \mathbf{F}, \nabla f(\mathbf{F}) \rangle] &= \mathbb{E}\left[\sum_{i,j=1}^n C_{ij} \partial_{ij}^2 f(\mathbf{F}) - \sum_{i=1}^n F^i \partial_i f(\mathbf{F})\right], \\ &= \sum_{i,j=1}^n C_{ij} \mathbb{E}[\partial_{ij}^2 f(\mathbf{F})] - \sum_{i=1}^n \mathbb{E}[F^i \partial_i f(\mathbf{F})]. \end{aligned}$$

By definition,  $\mathbf{F}$  is defined as the divergence of the matrix  $Z$ , thus for each  $i \in \llbracket 1, n \rrbracket$

$$F^i = \sum_{k=1}^d \delta^k(Z^{ki}),$$

which entails that

$$\mathbb{E}[F^i \partial_i f(\mathbf{F})] = \sum_{k=1}^d \mathbb{E}[\delta^k(Z^{ki}) \partial_i f(\mathbf{F})].$$

Set  $\mathbb{E}_{\neq k}[\cdot] = \mathbb{E}[\cdot | \Pi^1, \dots, \Pi^{k-1}, \Pi^{k+1}, \dots, \Pi^d]$ , which stands for the expected value knowing all the counting measures except for the  $k$ -th one. This notation is introduced in order use the integration by parts formula in [75], (cf. Proposition 3.2.8) which is available for univariate processes. Hence

$$\begin{aligned} \mathbb{E}[F^i \partial_i f(\mathbf{F})] &= \sum_{k=1}^d \mathbb{E}[\mathbb{E}_{\neq k}[\delta^k(Z^{ki}) \partial_i f(\mathbf{F})]] \\ &= \sum_{k=1}^d \mathbb{E}\left[\mathbb{E}_{\neq k}\left[\int_{\mathbb{R}_+^3} Z_{(t,\theta,x)}^{ki} D_{(t,\theta,x)}^k \partial_i f(\mathbf{F}) dt d\theta \nu^k(dx)\right]\right] \\ &= \sum_{k=1}^d \mathbb{E}\left[\int_{\mathbb{R}_+^3} Z_{(t,\theta,x)}^{ki} D_{(t,\theta,x)}^k \partial_i f(\mathbf{F}) dt d\theta \nu^k(dx)\right]. \end{aligned}$$

Using Lemma 4.3.1, we have that

$$\begin{aligned} D_{(t,\theta,x)}^k \partial_i f(\mathbf{F}) &= \langle \nabla \partial_i f(\mathbf{F}), D_{(t,\theta,x)}^k \mathbf{F} \rangle + \frac{1}{2} \langle D_{(t,\theta,x)}^k \mathbf{F}, \text{Hess } \partial_i f(\bar{\mathbf{F}}) D_{(t,\theta,x)}^k \mathbf{F} \rangle \\ &= \sum_{j=1}^n \partial_{ij}^2 f(\mathbf{F}) D_{(t,\theta,x)}^k F^j + \frac{1}{2} \langle D_{(t,\theta,x)}^k \mathbf{F}, \text{Hess } \partial_i f(\bar{\mathbf{F}}) D_{(t,\theta,x)}^k \mathbf{F} \rangle, \end{aligned}$$

for some random  $\bar{\mathbf{F}}$ . Hence

$$\begin{aligned} \mathbb{E}[F^i \partial_i f(\mathbf{F})] &= \sum_{k=1}^d \mathbb{E} \left[ \int_{\mathbb{R}_+^3} Z_{(t,\theta,x)}^{ki} \sum_{j=1}^n \partial_{ij}^2 f(\mathbf{F}) D_{(t,\theta,x)}^k F^j dt d\theta \nu^k(dx) \right] \\ &\quad + \sum_{k=1}^d \frac{1}{2} \mathbb{E} \left[ \int_{\mathbb{R}_+^3} Z_{(t,\theta,x)}^{ki} \langle D_{(t,\theta,x)}^k \mathbf{F}, \text{Hess } \partial_i f(\bar{\mathbf{F}}) D_{(t,\theta,x)}^k \mathbf{F} \rangle dt d\theta \nu^k(dx) \right] \\ &= \sum_{j=1}^n \mathbb{E} \left[ \partial_{ij}^2 f(\mathbf{F}) \sum_{k=1}^d \int_{\mathbb{R}_+^3} Z_{(t,\theta,x)}^{ki} D_{(t,\theta,x)}^k F^j dt d\theta \nu^k(dx) \right] \\ &\quad + \sum_{k=1}^d \frac{1}{2} \mathbb{E} \left[ \int_{\mathbb{R}_+^3} Z_{(t,\theta,x)}^{ki} \langle D_{(t,\theta,x)}^k \mathbf{F}, \text{Hess } \partial_i f(\bar{\mathbf{F}}) D_{(t,\theta,x)}^k \mathbf{F} \rangle dt d\theta \nu^k(dx) \right]. \end{aligned}$$

Then

$$\begin{aligned} &\mathbb{E}[\langle C, \text{Hess } f(\mathbf{F}) \rangle_{H.S} - \langle \mathbf{F}, \nabla f(\mathbf{F}) \rangle] \\ &= \sum_{i,j=1}^n \mathbb{E} \left[ \partial_{ij}^2 f(\mathbf{F}) \left( C_{ij} - \sum_{k=1}^d \int_{\mathbb{R}_+^3} Z_{(t,\theta,x)}^{ki} D_{(t,\theta,x)}^k F^j dt d\theta \nu^k(dx) \right) \right] \\ &\quad - \frac{1}{2} \sum_{i=1}^n \sum_{k=1}^d \mathbb{E} \left[ \int_{\mathbb{R}_+^3} Z_{(t,\theta,x)}^{ki} \langle D_{(t,\theta,x)}^k \mathbf{F}, \text{Hess } \partial_i f(\bar{\mathbf{F}}) D_{(t,\theta,x)}^k \mathbf{F} \rangle dt d\theta \nu^k(dx) \right]. \end{aligned}$$

By taking the absolute value, using the triangular inequality and the fact that for any  $f \in \mathcal{F}_C$  and any  $(i, j) \in [1, d]^2$

$$\|\partial_{ij}^2 f\|_\infty \leq M_2(f) \leq \|C^{-1}\|_{op} \|C\|_{op}^{1/2},$$

we obtain the bound

$$\begin{aligned}
& \left| \mathbb{E}[\langle C, \text{Hess } f(\mathbf{F}) \rangle_{H.S} - \langle \mathbf{F}, \nabla f(\mathbf{F}) \rangle] \right| \\
& \leq \|C^{-1}\|_{op} \|C\|_{op}^{1/2} \sum_{i,j=1}^n \mathbb{E} \left[ \left| C_{ij} - \sum_{k=1}^d \int_{\mathbb{R}_+^3} Z_{(t,\theta,x)}^{ki} D_{(t,\theta,x)}^k F^j dt d\theta \nu^k(dx) \right| \right] \\
& \quad + \frac{1}{2} \sum_{i,j=1}^n \sum_{k=1}^d \mathbb{E} \left[ \int_{\mathbb{R}_+^3} |Z_{(t,\theta,x)}^{ki}| \left| \langle D_{(t,\theta,x)}^k \mathbf{F}, \text{Hess } \partial_i f(\bar{\mathbf{F}}) D_{(t,\theta,x)}^k \mathbf{F} \rangle \right| dt d\theta \nu^k(dx) \right] \\
& \leq \|C^{-1}\|_{op} \|C\|_{op}^{1/2} \sum_{i=1}^n \mathbb{E} \left[ \left| C_{ij} - \sum_{k=1}^d \int_{\mathbb{R}_+^3} Z_{(t,\theta,x)}^{ki} D_{(t,\theta,x)}^k F^j dt d\theta \nu^k(dx) \right| \right] \\
& \quad + \frac{\sqrt{2\pi}}{8} \|C^{-1}\|_{op}^{3/2} \|C\|_{op} \sum_{i=1}^n \sum_{k=1}^d \mathbb{E} \left[ \int_{\mathbb{R}_+^3} |Z_{(t,\theta,x)}^{ki}| \left\| D_{(t,\theta,x)}^k \mathbf{F} \right\|^2 dt d\theta \nu^k(dx) \right],
\end{aligned}$$

and finally

$$\begin{aligned}
d_2(\mathbf{F}, \mathbf{G}) & \leq \|C^{-1}\|_{op} \|C\|_{op}^{1/2} \sum_{i,j=1}^n \mathbb{E} \left[ \left| C_{ij} - \sum_{k=1}^d \int_{\mathbb{R}_+^3} Z_{(t,\theta,x)}^{ki} D_{(t,\theta,x)}^k F^j dt d\theta \nu^k(dx) \right| \right] \\
& \quad + \frac{\sqrt{2\pi}}{8} \|C^{-1}\|_{op}^{3/2} \|C\|_{op} \sum_{i=1}^n \sum_{k=1}^d \mathbb{E} \left[ \int_{\mathbb{R}_+^3} |Z_{(t,\theta,x)}^{ki}| \left\| D_{(t,\theta,x)}^k \mathbf{F} \right\|^2 dt d\theta \nu^k(dx) \right].
\end{aligned}$$

For the second point, assume that  $\forall k \in \llbracket 1, d \rrbracket$  and  $\forall i \in \llbracket 1, n \rrbracket$

$$Z_{(t,\theta,x)}^{ki} = \mathbb{1}_{\theta \leq \lambda_t^k} U_{(t,x)}^{ki}.$$

Keeping in mind that according to Definition 4.2.8, we have

$$\mathbb{1}_{\theta \leq \lambda_t^k} D_{(t,\theta,x)}^k F^j = D_{(t,\lambda_t^k,x)}^k F^j$$

whenever  $\theta$  is less than  $\lambda_t^k$ , thus

$$\begin{aligned}
d_2(\mathbf{F}, \mathbf{G}) & \leq \|C^{-1}\|_{op} \|C\|_{op}^{1/2} \sum_{i,j=1}^n \mathbb{E} \left[ \left| C_{ij} - \sum_{k=1}^d \int_{\mathbb{R}_+^2} \lambda_t^k U_{(t,x)}^{ki} D_{(t,\lambda_t^k,x)}^k F^j dt \nu^k(dx) \right| \right] \\
& \quad + \frac{\sqrt{2\pi}}{8} \|C^{-1}\|_{op}^{3/2} \|C\|_{op} \sum_{i=1}^n \sum_{k=1}^d \mathbb{E} \left[ \int_{\mathbb{R}_+^2} \lambda_t^k |U_{(t,x)}^{ki}| \left\| D_{(t,\lambda_t^k,x)}^k \mathbf{F} \right\|^2 dt \nu^k(dx) \right].
\end{aligned}$$

□

### 4.3.2 Bounds for the CLT

The following theorem is our first main result, where we give the bound for the  $d_2$  distance between a vector formed by the normalized Hawkes martingale

$$\mathbf{F}_T = \frac{\mathbf{M}_T}{\sqrt{T}} = \frac{\mathbf{L}_T - \text{diag}(m^1, \dots, m^d) \int_0^T \boldsymbol{\lambda}_t dt}{\sqrt{T}}$$

and its Gaussian limit as  $T$  goes to infinity.

*Theorem 4.3.3.* Fix  $p$  and  $d$  in  $\mathbb{N}^*$ . Let  $(\mathbf{L}_t)_{t \geq 0}$  be a compound multivariate Hawkes process whose intensity  $\lambda$  follows the dynamics 4.2.2 and let  $0 < v_1 < \dots < v_p \leq 1$  be  $p$  distinct positive numbers. Set

$$\mathbf{\Gamma}_T = \left( F_{v_1 T}^1, \dots, F_{v_p T}^1, \dots, F_{v_1 T}^d, \dots, F_{v_p T}^d \right) \in \mathbb{R}^{p \cdot d},$$

and

$$\hat{C} = \text{diag}(C^1, \dots, C^d) \in \mathcal{S}_{p \cdot d}(\mathbb{R})$$

the block diagonal matrix such that  $\forall n \in \llbracket 1, d \rrbracket$

$$C_{ij}^n = C_{ji}^n = \int x^2 \nu^n(dx) \sqrt{\frac{v_i}{v_j}} \left[ \left( B - A \text{diag}(m^1, \dots, m^d) \right)^{-1} B \mu \right]^n, \quad \forall 1 \leq i \leq j \leq p.$$

Let  $\mathbf{G} \sim \mathcal{N}_d(0, \hat{C})$ . Then there is a constant  $K > 0$  independent of  $T$  such that

$$d_2(\mathbf{\Gamma}_T, \mathbf{G}) \leq \frac{K}{\sqrt{T}}.$$

*Proof.* Throughout this proof,  $K$  is a positive constant that does not depend on  $T$  and that is susceptible to change from one line to the other. We also set  $v_0 = v_p$ . For each  $i \in \llbracket 1, p \cdot d \rrbracket$  and  $k \in \llbracket 1, d \rrbracket$ , we define the matrix process

$$Z_{(t, \theta, x)}^{ki} = \mathbb{1}_{\theta \leq \lambda_t^k} \mathbb{1}_{k=i \div p + 1} \mathbb{1}_{t \leq v_i \% p T} \frac{x}{\sqrt{v_i \% p T}},$$

where  $i \% p$  (respectively  $i \div p$ ) is the remainder (respectively the quotient) of the Euclidean division of  $i$  by  $p$ . It is possible to write the set of integers from 1 to  $p \cdot d$  as a partition of  $d$  disjoint intervals  $S_1 \cup \dots \cup S_d$  each of cardinal  $p$ . Thus for any  $n = 1, \dots, d$ ,  $n = i \div p + 1$  if and only if  $i \in S_n$ .

We assume  $i \in S_n$  and compute the  $i$ -th component of the divergence of  $Z$

$$\begin{aligned} (\delta(Z))^i &= \sum_{k=1}^d \delta^k(Z^{ki}), \\ &= \delta^n(Z^{ni}), \\ &= \int_{\mathbb{R}_+^3} Z_{(t, \theta, x)}^{ni} (\Pi^n(dt, d\theta, dx) - dt d\theta \nu^n(dx)), \\ &= \frac{1}{\sqrt{v_i \% p T}} \int_0^{v_i \% p T} \int_{\mathbb{R}_+^2} x \mathbb{1}_{\theta \leq \lambda_t^n} (\Pi^n(dt, d\theta, dx) - dt d\theta \nu^n(dx)), \\ &= \Gamma_T^i. \end{aligned}$$

Using the second equality of Theorem 4.3.2, the  $d_2$  distance is bounded by

$$\begin{aligned}
& d_2(\Gamma_T, \mathbf{G}) \\
& \leq \|\hat{C}^{-1}\|_{op} \|\hat{C}\|_{op}^{1/2} \sum_{i,j=1}^{pd} \mathbb{E} \left[ \left\| \hat{C}_{ij} - \sum_{k=1}^d \int_{\mathbb{R}_+^2} \lambda_t^k \mathbf{1}_{t \leq v_i \%_p T} \mathbf{1}_{k=i \div p+1} \frac{x}{\sqrt{v_i \%_p T}} D_{(t, \lambda_t^k, x)}^k \Gamma_T^j dt \nu^k(dx) \right\|^2 \right] \\
& \quad + \frac{\sqrt{2\pi}}{8} \|\hat{C}^{-1}\|_{op}^{3/2} \|\hat{C}\|_{op} \sum_{i=1}^{pd} \sum_{k=1}^d \mathbb{E} \left[ \int_{\mathbb{R}_+^2} \lambda_t^k \mathbf{1}_{t \leq v_i \%_p T} \mathbf{1}_{k=i \div p+1} \frac{x}{\sqrt{v_i \%_p T}} \left\| D_{(t, \lambda_t^k, x)}^k \Gamma_T \right\|^2 dt \nu^k(dx) \right] \\
& \leq \|\hat{C}^{-1}\|_{op} \|\hat{C}\|_{op}^{1/2} \sum_{n_1, n_2=1}^d \sum_{i \in S_{n_1}} \sum_{j \in S_{n_2}} \mathbb{E} \left[ \left\| \hat{C}_{ij} - \int_{\mathbb{R}_+} \int_0^{v_i \%_p T} \lambda_t^{n_1} \frac{x}{\sqrt{v_i \%_p T}} D_{(t, \lambda_t^{n_1}, x)}^{n_1} \Gamma_T^j dt \nu^{n_1}(dx) \right\|^2 \right] \\
& \quad + \frac{\sqrt{2\pi}}{8} \|\hat{C}^{-1}\|_{op}^{3/2} \|\hat{C}\|_{op} \sum_{n_1=1}^d \sum_{i \in S_{n_1}} \mathbb{E} \left[ \int_{\mathbb{R}_+} \int_0^{v_i \%_p T} \lambda_t^{n_1} \frac{x}{\sqrt{v_i \%_p T}} \left\| D_{(t, \lambda_t^{n_1}, x)}^{n_1} \Gamma_T \right\|^2 dt \nu^{n_1}(dx) \right] \\
& \leq \|\hat{C}^{-1}\|_{op} \|\hat{C}\|_{op}^{1/2} \sum_{n_1, n_2=1}^d \sum_{i \in S_{n_1}} \sum_{j \in S_{n_2}} \mathbb{E} \left[ \left\| \hat{C}_{ij} - \int_{\mathbb{R}_+} \int_0^{v_i \%_p T} \lambda_t^{n_1} \frac{x}{\sqrt{v_i \%_p T}} D_{(t, \lambda_t^{n_1}, x)}^{n_1} \Gamma_T^j dt \nu^{n_1}(dx) \right\|^2 \right] \\
& \quad + \frac{\sqrt{2\pi}}{8} \|\hat{C}^{-1}\|_{op}^{3/2} \|\hat{C}\|_{op} \sum_{n_1, n_2=1}^d \sum_{i \in S_{n_1}} \sum_{j \in S_{n_2}} \mathbb{E} \left[ \int_{\mathbb{R}_+} \int_0^{v_i \%_p T} \lambda_t^{n_1} \frac{x}{\sqrt{v_i \%_p T}} \left| D_{(t, \lambda_t^{n_1}, x)}^{n_1} \Gamma_T^j \right|^2 dt \nu^{n_1}(dx) \right] \\
& \leq K \sum_{n_1, n_2=1}^d \sum_{i \in S_{n_1}} \sum_{j \in S_{n_2}} \left( A_1^{i,j} + A_2^{i,j} \right),
\end{aligned}$$

where

$$A_1^{i,j} = \mathbb{E} \left[ \left\| \hat{C}_{ij} - \int_{\mathbb{R}_+} \int_0^{v_i \%_p T} \lambda_t^{n_1} \frac{x}{\sqrt{v_i \%_p T}} D_{(t, \lambda_t^{n_1}, x)}^{n_1} \Gamma_T^j dt \nu^{n_1}(dx) \right\|^2 \right],$$

and

$$A_2^{i,j} = \mathbb{E} \left[ \int_{\mathbb{R}_+} \int_0^{v_i \%_p T} \lambda_t^{n_1} \frac{x}{\sqrt{v_i \%_p T}} \left| D_{(t, \lambda_t^{n_1}, x)}^{n_1} \Gamma_T^j \right|^2 dt \nu^{n_1}(dx) \right].$$

In both terms we have dependence on  $n_1$  and  $n_2$  which are respectively functions of  $i$  and  $j$ . For  $i, j = 1, \dots, d$ . we will treat each term separately.

**Term  $A_1^{i,j}$**

First we start by computing the Malliavin's derivative in the  $i$ -th direction of  $\Gamma_T^j$ . By linearity of the derivative operator

$$\begin{aligned}
D_{(t, \lambda_t^{n_1}, x)}^{n_1} \Gamma_T^j &= D_{(t, \lambda_t^{n_1}, x)}^{n_1} F_{v_j \%_p T}^{n_2}, \\
&= \frac{1}{\sqrt{v_j \%_p T}} D_{(t, \lambda_t^{n_1}, x)}^{n_1} \left( L_{v_j \%_p T}^{n_2} - m^{j \%_p} \int_0^{v_j \%_p T} \lambda_s^{n_2} ds \right),
\end{aligned}$$

where  $n_2 = j \div p + 1$ . This yields using Proposition 4.2.10

$$\begin{aligned} D_{(t, \lambda_t^{n_1}, x)}^{n_1} \Gamma_T^j &= \frac{\mathbb{1}_{t \leq v_j \% p T}}{\sqrt{v_j \% p T}} \left( x \mathbb{1}_{n_1 = n_2} + \hat{L}_{v_j \% p T}^{n_2, n_1, t, x} - m^{j \% p} \int_t^{v_j \% p T} \hat{\lambda}_s^{n_2, n_1, t, x} ds \right) \\ &= \frac{\mathbb{1}_{t \leq v_j \% p T}}{\sqrt{v_j \% p T}} \left( x \mathbb{1}_{n_1 = n_2} + \hat{M}_{v_j \% p T}^{n_2, n_1, t, x} \right). \end{aligned}$$

Thus

$$\begin{aligned} A_1^{i,j} &= \mathbb{E} \left[ \left| \hat{C}_{ij} - \int_{\mathbb{R}_+} \int_0^{v_i \% p T} \lambda_t^{n_1} \frac{x}{\sqrt{v_i \% p T}} \frac{\mathbb{1}_{t \leq v_j \% p T}}{\sqrt{v_j \% p T}} \left( x \mathbb{1}_{n_1 = n_2} + \hat{M}_{v_j \% p T}^{n_2, n_1, t, x} \right) dt \nu^{n_1}(dx) \right| \right] \\ &= \mathbb{E} \left[ \left| \hat{C}_{ij} - \frac{1}{\sqrt{v_i \% p v_j \% p T}} \int_{\mathbb{R}_+} \int_0^{(v_i \% p \wedge v_j \% p) T} x \lambda_t^{n_1} \left( x \mathbb{1}_{n_1 = n_2} + \hat{M}_{v_j \% p T}^{n_2, n_1, t, x} \right) dt \nu^{n_1}(dx) \right| \right], \\ &\leq \mathbb{E} \left[ \left| \hat{C}_{ij} - \frac{\mathbb{1}_{n_1 = n_2}}{\sqrt{v_i \% p v_j \% p T}} \int_{\mathbb{R}_+} \int_0^{(v_i \% p \wedge v_j \% p) T} x^2 \lambda_t^{n_1} dt \nu^{n_1}(dx) \right| \right] \\ &\quad + \frac{1}{\sqrt{v_i \% p v_j \% p T}} \mathbb{E} \left[ \left| \int_{\mathbb{R}_+} \int_0^{(v_i \% p \wedge v_j \% p) T} x \lambda_t^{n_1} \hat{M}_{v_j \% p T}^{n_2, n_1, t, x} dt \nu^{n_1}(dx) \right| \right] \\ &\leq \mathbb{E} \left[ \left| \hat{C}_{ij} - \frac{\mathbb{1}_{n_1 = n_2}}{\sqrt{v_i \% p v_j \% p T}} \int_{\mathbb{R}_+} x^2 \nu^{n_1}(dx) \int_0^{(v_i \% p \wedge v_j \% p) T} \mathbb{E}[\lambda_t^{n_1}] dt \right| \right] \\ &\quad + \frac{\mathbb{1}_{n_1 = n_2}}{\sqrt{v_i \% p v_j \% p T}} \int_{\mathbb{R}_+} x^2 \nu^{n_1}(dx) \mathbb{E} \left[ \left| \int_0^{(v_i \% p \wedge v_j \% p) T} \lambda_t^{n_1} - \mathbb{E}[\lambda_t^{n_1}] dt \right| \right] \\ &\quad + \frac{1}{\sqrt{v_i \% p v_j \% p T}} \mathbb{E} \left[ \left| \int_{\mathbb{R}_+} \int_0^{(v_i \% p \wedge v_j \% p) T} x \lambda_t^{n_1} \hat{M}_{v_j \% p T}^{n_2, n_1, t, x} dt \nu^{n_1}(dx) \right| \right] \\ &\leq A_{1,1}^{i,j} + \mathbb{1}_{n_1 = n_2} A_{1,2}^{i,j} + A_{1,3}^{i,j}. \end{aligned}$$

We start with the term  $A_{1,1}^{i,j}$ . If  $i$  and  $j$  are not in the same interval  $S_{n_1}$  (*i.e.*  $n_1 = n_2$ ),  $\hat{C}_{ij} = 0$ , thus  $A_{1,1}^{i,j} = 0$ . Note that due to symmetry arguments  $i$  and  $j$  are exchangeable, which allows us to assume that  $i$  and  $j$  are in the same interval  $S_{n_1}$  and that  $i \leq j$  which is equivalent to  $i \% p \leq j \% p$ . In this case, we have

$$\hat{C}_{ij} = C_{i \% p j \% p}^{n_1} = \int_{\mathbb{R}_+} x^2 \nu^{n_1}(dx) \sqrt{\frac{v_i \% p}{v_j \% p}} \left[ \left( B - A \operatorname{diag}(m^1, \dots, m^d) \right)^{-1} B \boldsymbol{\mu} \right]^{n_1}.$$

According to Lemma 4.4.2 it is possible to put the intensity's expectation under the form

$$\mathbb{E}[\boldsymbol{\lambda}_t] = \left( B - A \operatorname{diag}(m^1, \dots, m^d) \right)^{-1} B \boldsymbol{\mu} + Q e^{-tV} \boldsymbol{\mu},$$

where  $Q$  and  $V$  are matrices such that all the eigenvalues of  $V$  are positive. Thus

$$\begin{aligned} & \frac{1}{\sqrt{v_{i\%p}v_{j\%p}T}} \int_0^{v_{i\%p}T} \mathbb{E}[\boldsymbol{\lambda}_t] dt \\ &= \sqrt{\frac{v_{i\%p}}{v_{j\%p}}} \left( B - A \operatorname{diag}(m^1, \dots, m^d) \right)^{-1} B \boldsymbol{\mu} + \frac{1}{\sqrt{v_{i\%p}v_{j\%p}T}} Q \int_0^{v_{i\%p}T} e^{-Vt} dt \boldsymbol{\mu} \\ &= \sqrt{\frac{v_{i\%p}}{v_{j\%p}}} \left( B - A \operatorname{diag}(m^1, \dots, m^d) \right)^{-1} B \boldsymbol{\mu} + \frac{1}{\sqrt{v_{i\%p}v_{j\%p}T}} QV^{-1} (I_d - e^{-VTv_{i\%p}}) \boldsymbol{\mu}, \end{aligned}$$

which means that

$$\begin{aligned} A_{1,1}^{i,j} &= \mathbb{E} \left[ \left\| \hat{C}_{ij} - \frac{1}{\sqrt{v_{i\%p}v_{j\%p}T}} \int_{\mathbb{R}_+} x^2 \nu^{n_1}(dx) \int_0^{v_{i\%p}T} \mathbb{E}[\boldsymbol{\lambda}_t^{n_1}] dt \right\| \right] \\ &\leq \mathbb{E} \left[ \left\| \hat{C}_{ij} - \sqrt{\frac{v_{i\%p}}{v_{j\%p}}} \int_{\mathbb{R}_+} x^2 \nu^{n_1}(dx) \left[ (B - A \operatorname{diag}(m^1, \dots, m^d))^{-1} B \boldsymbol{\mu} \right]^{n_1} \right\| \right] \\ &\quad + \mathbb{E} \left[ \left\| \int_{\mathbb{R}_+} x^2 \nu^{n_1}(dx) \frac{1}{\sqrt{v_{i\%p}v_{j\%p}T}} [QV^{-1} (I_d - e^{-VTv_{i\%p}}) \boldsymbol{\mu}]^{n_1} \right\| \right] \\ &= \int_{\mathbb{R}_+} x^2 \nu^{n_1}(dx) \frac{1}{\sqrt{v_{i\%p}v_{j\%p}T}} \mathbb{E} \left[ \left\| [QV^{-1} (I_d - e^{-VTv_{i\%p}}) \boldsymbol{\mu}]^{n_1} \right\| \right] \\ &= O\left(\frac{1}{T}\right) \quad \text{because } e^{-VTv_{i\%p}} \rightarrow 0. \end{aligned}$$

For the term  $A_{1,2}^{i,j}$  we use the second equality of Lemma 4.4.3, hence

$$\begin{aligned} \mathbb{E} \left[ \left\| \int_0^{v_{i\%p}T} \boldsymbol{\lambda}_t - \mathbb{E}[\boldsymbol{\lambda}_t] dt \right\| \right] &= \mathbb{E} \left[ \left\| \int_0^{v_{i\%p}T} e^{-(B - A \operatorname{diag}(m^1, \dots, m^d))(v_{i\%p}T-s)} A \mathbf{M}_s ds \right\| \right] \\ &\leq \mathbb{E} \left[ \left\| \int_0^{v_{i\%p}T} e^{-(B - A \operatorname{diag}(m^1, \dots, m^d))(v_{i\%p}T-s)} A \mathbf{M}_s \right\| ds \right] \\ &\leq \int_0^{v_{i\%p}T} \left\| e^{-(B - A \operatorname{diag}(m^1, \dots, m^d))(v_{i\%p}T-s)} A \right\|_{op} \mathbb{E} [\|\mathbf{M}_s\|] ds \\ &\leq \int_0^{v_{i\%p}T} \left\| e^{-(B - A \operatorname{diag}(m^1, \dots, m^d))(v_{i\%p}T-s)} A \right\|_{op} \mathbb{E} [\|\mathbf{M}_s\|^2]^{1/2} ds \end{aligned}$$

where the last inequality comes from Cauchy-Schwarz.



Keeping in mind Ito's isometry we have

$$\begin{aligned}
\mathbb{E} [\|\mathbf{M}_s\|^2] &= \mathbb{E} \left[ \sum_{i=1}^d |M_s^i|^2 \right] \\
&= \sum_{i=1}^d \mathbb{E} \left[ |M_s^i|^2 \right] \\
&= \sum_{i=1}^d \mathbb{E} \left[ [M^i]_s \right], \quad \text{where } [M^i]_s \text{ is the quadratic variation} \\
&= \sum_{i=1}^d \int_{\mathbb{R}_+} x^2 \nu^i(dx) \int_0^s \mathbb{E} [\lambda_u^i] du \\
&\leq K(s+1),
\end{aligned}$$

which implies that  $\mathbb{E} [\|\mathbf{M}_s\|^2]^{1/2} \leq K(\sqrt{s}+1)$ .

Using the fact that the operator norm is sub-multiplicative and Lemma 4.4.1 we have that

$$\begin{aligned}
\left\| e^{-(B-A \operatorname{diag}(m^1, \dots, m^d))(v_i \%_p T - s)} A \right\|_{op} &\leq \left\| e^{-(B-A \operatorname{diag}(m^1, \dots, m^d))(v_i \%_p T - s)} \right\|_{op} \|A\|_{op} \\
&\leq K \left( 1 + (v_i \%_p T - s)^{d-1} \right) e^{-\rho_d(v_i \%_p T - s)}.
\end{aligned}$$

Combining these inequalities yields

$$\begin{aligned}
A_{1,2}^{i,j} &= \frac{1}{\sqrt{v_i \%_p v_j \%_p T}} \mathbb{E} \left[ \left\| \int_0^{v_i \%_p T} \lambda_t^i - \mathbb{E}[\lambda_t^i] dt \right\| \right], \\
&\leq \frac{1}{\sqrt{v_i \%_p v_j \%_p T}} \mathbb{E} \left[ \left\| \int_0^{v_i \%_p T} \boldsymbol{\lambda}_t - \mathbb{E}[\boldsymbol{\lambda}_t] dt \right\| \right], \\
&\leq \frac{K}{\sqrt{v_i \%_p v_j \%_p T}} \int_0^{v_i \%_p T} e^{-\rho_d(v_i \%_p T - s)} \left( 1 + (v_i \%_p T - s)^{d-1} \right) (\sqrt{s}+1) ds, \\
&\leq O \left( \frac{1}{\sqrt{T}} \right), \quad \text{cf. the bound on } A_{1,2} \text{ in chapter 3.}
\end{aligned}$$

For the term  $A_{1,3}^{i,j}$  we start by noticing that the integral with respect to  $\nu^i(dx)$  is equivalent to taking the expectation of a random variable  $X$  of law  $\nu^i$ . Hence

$$\begin{aligned}
A_{1,3}^{i,j} &= \mathbb{E} \left[ \left\| \frac{1}{\sqrt{v_i \%_p v_j \%_p T}} \int_0^{v_i \%_p T} \int_{\mathbb{R}_+} x \lambda_t^{n_1} \hat{M}_{v_j \%_p T}^{n_2, n_1, t, x} \nu^i(dx) dt \right\| \right] \\
&= \frac{1}{\sqrt{v_i \%_p v_j \%_p T}} \mathbb{E} \left[ \left\| \mathbb{E}_X \left[ \int_0^{v_i \%_p T} X \lambda_t^{n_1} \hat{M}_{v_j \%_p T}^{n_2, n_1, t, X} dt \right] \right\| \right] \\
&\leq \frac{1}{\sqrt{v_i \%_p v_j \%_p T}} \mathbb{E} \left[ \left[ \mathbb{E}_X \left[ \int_0^{v_i \%_p T} X \lambda_t^{n_1} \hat{M}_{v_j \%_p T}^{n_2, n_1, t, X} dt \right]^2 \right]^{1/2} \right],
\end{aligned}$$

and by Jensen's inequality

$$\begin{aligned} A_{1,3}^{i,j} &\leq \frac{1}{\sqrt{v_{i\%p} v_{j\%p} T}} \mathbb{E} \left[ \mathbb{E}_X \left[ \left( \int_0^{v_{i\%p} T} X \lambda_t^{n_1} \hat{M}_{v_{j\%p} T}^{n_2, n_1, t, X} dt \right)^2 \right] \right]^{1/2}, \\ &\leq \frac{1}{\sqrt{v_{i\%p} v_{j\%p} T}} \left( \frac{1}{T} I_T \right)^{1/2}, \end{aligned}$$

where

$$I_T := \mathbb{E} \left[ \mathbb{E}_X \left[ \left( \int_0^{v_{i\%p} T} X \lambda_t^{n_1} \hat{M}_{v_{j\%p} T}^{n_2, n_1, t, X} dt \right)^2 \right] \right]. \quad (4.3.1)$$

Using Lemma 4.4.5 we have that

$$I_T \leq KT,$$

and therefore

$$A_{1,3}^{i,j} \leq O\left(\frac{1}{\sqrt{T}}\right).$$

This shows that

$$A_1^{i,j} = O\left(\frac{1}{\sqrt{T}}\right). \quad (4.3.2)$$

**Term  $A_2^{i,j}$**

Keeping in mind that

$$A_2^{i,j} = \mathbb{E} \left[ \int_{\mathbb{R}_+} \int_0^{v_{i\%p} T} \lambda_t^{n_1} \frac{x}{\sqrt{v_{i\%p} T}} \left| D_{(t, \lambda_t^{n_1}, x)}^{n_1} \Gamma_T^j \right|^2 dt \nu^{n_1}(dx) \right],$$

and that

$$D_{(t, \lambda_t^{n_1}, x)}^{n_1} \Gamma_T^j = \frac{\mathbb{1}_{t \leq v_{j\%p} T}}{\sqrt{v_{j\%p} T}} \left( x \mathbb{1}_{n_1 = n_2} + \hat{M}_{v_{j\%p} T}^{n_2, n_1, t, x} \right),$$

we have

$$\begin{aligned} A_2^{i,j} &= \frac{1}{\sqrt{v_{i\%p} v_{j\%p} T^{3/2}}} \mathbb{E} \left[ \int_0^{v_{i\%p} T \wedge v_{j\%p} T} \int_{\mathbb{R}_+} x \lambda_t^{n_1} \left( x \mathbb{1}_{n_1 = n_2} + \hat{M}_{v_{j\%p} T}^{n_2, n_1, t, x} \right)^2 \nu^{n_1}(dx) dt \right] \\ &\leq \frac{2}{\sqrt{v_{i\%p} v_{j\%p} T^{3/2}}} \mathbb{E} \left[ \int_0^{v_{i\%p} T} \int_{\mathbb{R}_+} x^3 \lambda_t^{n_1} \mathbb{1}_{n_1 = n_2} + x \lambda_t^{n_1} \left| \hat{M}_{v_{j\%p} T}^{n_2, n_1, t, x} \right|^2 \nu^{n_1}(dx) dt \right] \\ &= \frac{2K \cdot \mathbb{1}_{n_1 = n_2}}{T^{3/2}} \int_{\mathbb{R}_+} x^3 \nu^{n_1}(dx) \int_0^{v_{i\%p} T} \mathbb{E}[\lambda_t^{n_1}] dt \\ &\quad + \frac{2K}{T^{3/2}} \int_0^{v_{i\%p} T} \int_{\mathbb{R}_+} x \mathbb{E} \left[ \lambda_t^{n_1} \left| \hat{M}_{v_{j\%p} T}^{n_2, n_1, t, x} \right|^2 \right] \nu^{n_1}(dx) dt \\ &:= A_{2,1}^i + A_{2,2}^{i,j}. \end{aligned}$$

According to Lemma 4.4.2 and to the fact that  $(B - A \text{diag}(m^1, \dots, m^d))$  has only positive eigenvalues we have that  $\int_0^{v_{i\%p}T} \mathbb{E}[\lambda_t^i] dt \leq KT$  and hence

$$A_{2,1}^i = O\left(\frac{1}{\sqrt{T}}\right).$$

For the final term it holds that

$$\begin{aligned} A_{2,2}^{i,j} &= \frac{2K}{T^{3/2}} \int_0^{v_{i\%p}T} \int_{\mathbb{R}_+} x \mathbb{E} \left[ \lambda_t^{n_1} \left| \hat{M}_{v_j\%pT}^{n_2, n_1, t, x} \right|^2 \right] \nu^{n_1}(dx) dt \\ &= \frac{K}{T^{3/2}} \int_0^{v_{i\%p}T} \int_{\mathbb{R}_+} x \mathbb{E} \left[ \lambda_t^{n_1} \mathbb{E}_t \left[ \left| \hat{M}_{v_j\%pT}^{n_2, n_1, t, x} \right|^2 \right] \right] \nu^{n_1}(dx) dt \\ &= \frac{K}{T^{3/2}} \int_0^{v_{i\%p}T} \int_{\mathbb{R}_+} x \mathbb{E} \left[ \lambda_t^{n_1} \mathbb{E}_t \left[ \int_t^{v_j\%pT} \hat{\lambda}_s^{n_2, n_1, t, x} ds \right] \right] \nu^{n_1}(dx) dt. \end{aligned}$$

Solving the SDE 4.4.3 with initial condition  $\mathbb{E}_t [\tilde{\lambda}_t^{n_1, t, x}] = x \mathbf{A}_{\cdot i}$  we have that

$$\mathbb{E}_t [\tilde{\lambda}_s^{n_2, n_1, t, x}] = x \left( e^{-(B - A \text{diag}(m^1, \dots, m^d))(s-t)} \mathbf{A}_{\cdot n_1} \right)^{n_2},$$

which is integrable. It follows that

$$\begin{aligned} A_{2,2}^{i,j} &\leq \frac{K}{T^{3/2}} \int_0^{v_{i\%p}T} \int_{\mathbb{R}_+} x^2 \mathbb{E} [\lambda_t^{n_1}] \nu^{n_1}(dx) dt, \\ &\leq \frac{K}{T^{3/2}} \int_{\mathbb{R}_+} x^2 \nu^{n_1}(dx) K v_{i\%p}T, \\ &\leq O\left(\frac{1}{\sqrt{T}}\right). \end{aligned}$$

And finally

$$A_2^{i,j} = O\left(\frac{1}{\sqrt{T}}\right). \quad (4.3.3)$$

Combining (4.3.2) and (4.3.3) we conclude that

$$d_2(\mathbf{\Gamma}_T, \mathbf{G}) = O\left(\frac{1}{\sqrt{T}}\right).$$

□

*Corollary 4.3.4.* Fix  $d \in \mathbb{N}^*$ . Let  $(\mathbf{L}_t)_{t \geq 0}$  be a compound multivariate Hawkes process whose intensity  $\lambda$  follows the dynamics 4.2.2. Assume Assumptions 4.1.2, 4.1.3 and 4.1.4 are in force.

Set  $C = \text{diag}(\sigma_1^2, \dots, \sigma_d^2)$  where for any  $j = 1, \dots, d$

$$\sigma_j^2 = \int x^2 \nu^j(dx) \left[ (B - A \text{diag}(m^1, \dots, m^d))^{-1} B \boldsymbol{\mu} \right]^j$$

and let  $\mathbf{G} \sim \mathcal{N}(0, C)$ .

Then there exists a constant  $K > 0$  independent from  $T$  such that

$$d_2(\mathbf{F}_T, \mathbf{G}) \leq \frac{K}{\sqrt{T}}.$$

*Proof.* This is merely an application of Theorem 4.3.3 for  $p = 1$  and  $v_1 = 1$ .  $\square$

*Remark 4.3.5.* The specific bounds  $\|g\|_{Lip} \leq 1$  and  $M_2(g) \leq 1$  could have been relaxed to  $\|g\|_{Lip} \leq K_1$  and  $M_2(g) \leq K_2$  where  $K_1$  and  $K_2$  are two arbitrary positive constants.

As a final result in this section, we consider the slightly modified process

$$\mathbf{Y}_T = \frac{\mathbf{L}_T - \text{diag}(m^1, \dots, m^d) \int_0^T \mathbb{E}[\boldsymbol{\lambda}_s] ds}{\sqrt{T}}$$

and we study its behaviour as  $T$  goes to infinity.

*Theorem 4.3.6.* Set  $\tilde{C} = \left(J\sqrt{C}\right)^t \left(J\sqrt{C}\right) = JC^t J$ , where  $J = (I_d - \text{diag}(m^1, \dots, m^d)B^{-1}A)^{-1}$  and  $C$  is defined in Corollary 4.3.4.

Let  $\tilde{\mathbf{G}} \sim \mathcal{N}(0, \tilde{C})$ . There exists a constant  $K$  that does not depend on  $T$  such that

$$d_2(\mathbf{Y}_T, \tilde{\mathbf{G}}) \leq \frac{K}{\sqrt{T}},$$

for any  $T > 0$ .

*Proof.* Thanks to Lemma 4.4.6 we have that

$$J^{-1}\mathbf{Y}_T = \mathbf{F}_T + J^{-1}\mathbf{R}_T.$$

Let  $f \in \mathcal{H}$  and set  $f_J : \mathbf{x} \mapsto \frac{f(J\mathbf{x})}{\|J\|_{op}}$ , this function is clearly in  $\mathcal{C}^2(\mathbb{R}^d)$  and  $\|f_J\|_{Lip} \leq 1$  and  $M_2(f_J) \leq \|J\|_{op}$ .

Let  $\tilde{\mathbf{G}} \sim \mathcal{N}(0, \tilde{C})$ . We have

$$\begin{aligned} \left| \mathbb{E}[f(\mathbf{Y}_T)] - \mathbb{E}[f(\tilde{\mathbf{G}})] \right| &= \|J\|_{op} \left| \mathbb{E}[f_J(J^{-1}\mathbf{Y}_T)] - \mathbb{E}[f_J(J^{-1}\tilde{\mathbf{G}})] \right| \\ &= \|J\|_{op} \left| \mathbb{E}[f_J(\mathbf{F}_T + J^{-1}\mathbf{R}_T)] - \mathbb{E}[f_J(\mathbf{G})] \right|, \end{aligned}$$

where  $\mathbf{G} = J^{-1}\tilde{\mathbf{G}} \sim \mathcal{N}(0, C)$ .

Using a Taylor expansion, there exists a random  $\mathbf{Y}^*$  such that

$$f_J(\mathbf{F}_T + J^{-1}\mathbf{R}_T) = f_J(\mathbf{F}_T) + \langle \nabla f_J(\mathbf{Y}^*), J^{-1}\mathbf{R}_T \rangle,$$

hence

$$\begin{aligned}
 \left| \mathbb{E}[f(\mathbf{Y}_T)] - \mathbb{E}[f(\tilde{\mathbf{G}})] \right| &= \|J\|_{op} \left| \mathbb{E}[f_J(\mathbf{F}_T)] - \mathbb{E}[f_J(\mathbf{G})] + \mathbb{E}[\langle \nabla f_J(\mathbf{Y}^*), J^{-1} \mathbf{R}_T \rangle] \right| \\
 &\leq \|J\|_{op} \left| \mathbb{E}[f_J(\mathbf{F}_T)] - \mathbb{E}[f_J(\mathbf{G})] \right| + \|J\|_{op} \left| \mathbb{E}[\langle \nabla f_J(\mathbf{Y}^*), J^{-1} \mathbf{R}_T \rangle] \right| \\
 &\leq \|J\|_{op} \sup_{g \in \mathcal{N}_J} \left| \mathbb{E}[g(\mathbf{F}_T)] - \mathbb{E}[g(\mathbf{G})] \right| + \|J\|_{op} \left| \mathbb{E}[\langle \nabla f_J(\mathbf{Y}^*), J^{-1} \mathbf{R}_T \rangle] \right|,
 \end{aligned}$$

where  $\mathcal{N}_J = \{f \in \mathcal{C}^2(\mathbb{R}^d) \text{ such that } \|f\|_{Lip} \leq 1 \text{ and } M_2(f) \leq \|J\|_{op}\}$ . Using Cauchy-Schwarz's inequality (twice) the second term is bounded as follows

$$\begin{aligned}
 \left| \mathbb{E}[\langle \nabla f_J(\mathbf{Y}^*), J^{-1} \mathbf{R}_T \rangle] \right| &\leq \mathbb{E}[\|\nabla f_J(\mathbf{Y}^*)\| \|J^{-1} \mathbf{R}_T\|] \\
 &\leq \left| \mathbb{E}[\|\nabla f_J(\mathbf{Y}^*)\|^2]^{1/2} \mathbb{E}[\|J^{-1} \mathbf{R}_T\|^2]^{1/2} \right| \\
 &\leq \left| \mathbb{E}[\|f_J\|_{Lip}^2]^{1/2} \mathbb{E}[\|J^{-1}\|_{op}^2 \|\mathbf{R}_T\|^2]^{1/2} \right| \\
 &= \|J^{-1}\|_{op} \mathbb{E}[\|\mathbf{R}_T\|^2]^{1/2}.
 \end{aligned}$$

Thanks to Corollary 4.3.4 and Remark 4.3.5 there exists a positive constant  $K$  (independent from  $f$  and  $T$  and that can change from one line to another) such that

$$\left| \mathbb{E}[f(\mathbf{Y}_T)] - \mathbb{E}[f(\tilde{\mathbf{G}})] \right| \leq K \left( \frac{1}{\sqrt{T}} + \mathbb{E}[\|\mathbf{R}_T\|^2]^{1/2} \right).$$

And using Lemma 4.4.4 we have that

$$\begin{aligned}
 \mathbb{E}[\|\mathbf{R}_T\|^2] &= \mathbb{E} \left[ \left\| \text{diag}(m^1, \dots, m^d) \left( B - A \text{diag}(m^1, \dots, m^d) \right)^{-1} \frac{\mathbb{E}[\boldsymbol{\lambda}_T] - \boldsymbol{\lambda}_T}{\sqrt{T}} \right\|^2 \right] \\
 &\leq \left\| \text{diag}(m^1, \dots, m^d) \left( B - A \text{diag}(m^1, \dots, m^d) \right)^{-1} \right\|_{op}^2 \frac{1}{T} \mathbb{E}[\|\mathbb{E}[\boldsymbol{\lambda}_T] - \boldsymbol{\lambda}_T\|^2] \\
 &\leq \frac{K}{T} \sum_{i=1}^d \mathbb{E}[(\lambda_T^i - \mathbb{E}[\lambda_T^i])^2], \quad \text{and by virtue of Lemma 4.4.4,} \\
 &\leq \frac{K}{T}.
 \end{aligned}$$

And finally

$$\left| \mathbb{E}[f(\mathbf{Y}_T)] - \mathbb{E}[f(\tilde{\mathbf{G}})] \right| = O\left(\frac{1}{\sqrt{T}}\right),$$

hence the result.  $\square$

### 4.3.3 Proof of Theorem 4.1.5

We start by recalling that

$$\mathbf{Y}_T = \frac{L_T - \text{diag}(m^1, \dots, m^d) \int_0^T \mathbb{E}[\boldsymbol{\lambda}_s] ds}{\sqrt{T}},$$

and that

$$\begin{aligned} \mathbf{Y}'_T &= \frac{\mathbf{L}_T - \text{diag}(m^1, \dots, m^d) (B - A \text{diag}(m^1, \dots, m^d))^{-1} B \boldsymbol{\mu} T}{\sqrt{T}} \\ &= \frac{\mathbf{L}_T - \text{diag}(m^1, \dots, m^d) V^{-1} B \boldsymbol{\mu} T}{\sqrt{T}}, \end{aligned}$$

where  $V = B - A \text{diag}(m^1, \dots, m^d)$ .

Using Lemma 4.4.2, we have that

$$\mathbb{E}[\boldsymbol{\lambda}_t] = V^{-1} B \boldsymbol{\mu} + e^{-Vt} (I_d - V^{-1} B) \boldsymbol{\mu},$$

hence

$$\begin{aligned} \int_0^T \mathbb{E}[\boldsymbol{\lambda}_t] dt &= V^{-1} B \boldsymbol{\mu} T + \int_0^T e^{-Vt} dt (I_d - V^{-1} B) \boldsymbol{\mu} \\ &= V^{-1} B \boldsymbol{\mu} T + (I_d - e^{-VT}) V^{-1} (I_d - V^{-1} B) \boldsymbol{\mu}. \end{aligned}$$

We deduce that

$$\mathbf{Y}'_T = \mathbf{Y}_T - \mathbf{R}'_T,$$

with  $\mathbf{R}'_T = \frac{\text{diag}(m^1, \dots, m^d) (I_d - e^{-VT}) (V^{-1} - V^{-2} B) \boldsymbol{\mu}}{\sqrt{T}}$ .

Let  $\tilde{\mathbf{G}} \sim \mathcal{N}(0, \tilde{C})$  as defined in Theorem 4.1.5 and let  $f \in \mathcal{H}$ . Using a Taylor expansion, we have for some  $\mathbf{X}$

$$f(\mathbf{Y}'_T) - f(\tilde{\mathbf{G}}) = f(\mathbf{Y}_T) - f(\tilde{\mathbf{G}}) - \langle \nabla f(\mathbf{X}), \mathbf{R}'_T \rangle.$$

Since  $\sup_{\mathbf{x} \in \mathbb{R}^d} \|\nabla f(\mathbf{x})\|_2 \leq 1$  and  $\mathbf{Y}'_T = O\left(\frac{1}{\sqrt{T}}\right)$ , we deduce that

$$d_2(\mathbf{Y}'_T, \tilde{\mathbf{G}}) \leq d_2(\mathbf{Y}_T, \tilde{\mathbf{G}}) + O\left(\frac{1}{\sqrt{T}}\right),$$

which yields using Theorem 4.3.6

$$d_2(\mathbf{Y}'_T, \tilde{\mathbf{G}}) = O\left(\frac{1}{\sqrt{T}}\right).$$

## 4.4 Lemmata

*Lemma 4.4.1.* Set  $V = B - A \text{diag}(m^1, \dots, m^d)$ . Assume that Assumption 4.1.3 is in force. Then there are positive constants  $K$  and  $\rho_d$  such that

$$\|e^{-tV}\|_{op} \leq K(1 + t^{d-1})e^{-\rho_d t}$$

for any  $t \geq 0$ .

*Proof.* First, we call  $V$ 's eigenvalues  $\rho_1 \geq \dots \geq \rho_d$  and we recall that they are positive. Using Jordan-Chevalley's decomposition we can write

$$V = P \operatorname{diag}(\rho_1, \dots, \rho_d) P^{-1} + V_{nil}$$

where  $V_{nil}$  is a nilpotent matrix that commutes with  $P \operatorname{diag}(\rho_1, \dots, \rho_d) P^{-1}$ . Let  $t \geq 0$ , taking the exponential yields

$$\begin{aligned} e^{-tV} &= e^{-tP \operatorname{diag}(\rho_1, \dots, \rho_d) P^{-1} - tV_{nil}} \\ &= e^{-tP \operatorname{diag}(\rho_1, \dots, \rho_d) P^{-1}} e^{-tV_{nil}} \\ &= P \operatorname{diag}(e^{-\rho_1 t}, \dots, e^{-\rho_d t}) P^{-1} \sum_{j=1}^{d-1} t^j \frac{(-V_{nil})^j}{j!}. \end{aligned}$$

Since the operator norm is sub-multiplicative and using the triangle inequality

$$\begin{aligned} \|e^{-tV}\|_{op} &\leq \|P\|_{op} \|P^{-1}\|_{op} \|\operatorname{diag}(e^{-\rho_1 t}, \dots, e^{-\rho_d t})\|_{op} \sum_{j=1}^{d-1} t^j \frac{\|V_{nil}\|_{op}^j}{j!} \\ &\leq K \|\operatorname{diag}(e^{-\rho_1 t}, \dots, e^{-\rho_d t})\|_{op} (1 + t^{d-1}) \\ &\leq K \|\operatorname{diag}(e^{-\rho_1 t}, \dots, e^{-\rho_d t})\|_{\infty} (1 + t^{d-1}), \end{aligned}$$

where the last inequality comes from the fact that all norms are equivalent in finite dimension.  $\square$

*Lemma 4.4.2.* If  $\boldsymbol{\lambda}$  follows the dynamics (4.2.2) then for each  $t$

$$\begin{aligned} \mathbb{E}[\boldsymbol{\lambda}_t] &= \left( B - A \operatorname{diag}(m^1, \dots, m^d) \right)^{-1} B \boldsymbol{\mu} \\ &\quad + e^{-(B - A \operatorname{diag}(m^1, \dots, m^d))t} \left( I_d - (B - A \operatorname{diag}(m^1, \dots, m^d))^{-1} B \right) \boldsymbol{\mu}. \end{aligned}$$

*Proof.* First we prove that if the kernels are exponential, then  $\boldsymbol{\lambda}$  is a Markov process that follows an SDE. By multiplying (4.2.2) by  $e^{Bt}$  we get

$$\begin{aligned} e^{Bt} \boldsymbol{\lambda}_t &= e^{Bt} \boldsymbol{\mu} + \int_{[0,t)} e^{Bs} \operatorname{Ad} \boldsymbol{L}_s, \quad \text{and by differentiating,} \\ e^{Bt} (d\boldsymbol{\lambda}_t + B\boldsymbol{\lambda}_t dt) &= e^{Bt} B\boldsymbol{\mu} dt + e^{Bt} \operatorname{Ad} \boldsymbol{L}_t. \end{aligned}$$

Hence the SDE

$$d\boldsymbol{\lambda}_t = B(\boldsymbol{\mu} - \boldsymbol{\lambda}_t) dt + \operatorname{Ad} \boldsymbol{L}_t. \quad (4.4.1)$$

By taking the expected value of (4.4.1) and keeping in mind that  $d\mathbb{E}[\boldsymbol{L}_s] = \operatorname{diag}(m^1, \dots, m^d) \mathbb{E}[\boldsymbol{\lambda}_s] ds$ , the intensity's expectation is the solution of the multivariate ODE

$$d\mathbb{E}[\boldsymbol{\lambda}_t] = B\boldsymbol{\mu} dt - \left( B - A \operatorname{diag}(m^1, \dots, m^d) \right) \mathbb{E}[\boldsymbol{\lambda}_t] dt, \quad (4.4.2)$$

with initial condition  $\mathbb{E}[\boldsymbol{\lambda}_0] = \boldsymbol{\mu}$ . Assumption 4.1.3 guarantees that all the eigenvalues of  $(B - A \operatorname{diag}(m^1, \dots, m^d))$  are positive, thus the matrix is invertible and we get the result.  $\square$

*Lemma 4.4.3.* The difference between the intensity and its expected value can be written as

$$\lambda_t - \mathbb{E}[\lambda_t] = \int_{[0,t)} e^{-(B-A \operatorname{diag}(m^1, \dots, m^d))(t-s)} \operatorname{Ad} \mathbf{M}_s,$$

where

$$\mathbf{M}_t = \mathbf{L}_t - \int_0^t \operatorname{diag}(m^1, \dots, m^d) \lambda_s \operatorname{d}s.$$

Its integral with respect to time is

$$\int_0^T \lambda_t - \mathbb{E}[\lambda_t] \operatorname{d}t = \int_0^T e^{-(B-A \operatorname{diag}(m^1, \dots, m^d))(T-s)} \mathbf{A} \mathbf{M}_s \operatorname{d}s.$$

*Proof.* By taking the difference between (4.4.1) and (4.4.2) we can verify that  $\lambda - \mathbb{E}[\lambda]$  is a solution of the SDE

$$\operatorname{d}(\lambda_t - \mathbb{E}[\lambda_t]) = -(B - A \operatorname{diag}(m^1, \dots, m^d)) (\lambda_t - \mathbb{E}[\lambda_t]) + \operatorname{Ad} \mathbf{M}_t$$

with the initial condition  $\lambda_0 - \mathbb{E}[\lambda_0] = \mathbf{0}$ . Solving the SDE (using variation of parameters) yields

$$\lambda_t - \mathbb{E}[\lambda_t] = \int_{[0,t)} e^{-(B-A \operatorname{diag}(m^1, \dots, m^d))(t-s)} \operatorname{Ad} \mathbf{M}_s.$$

For the second equality, we start by taking the integral with respect to time until the instant  $T$

$$\begin{aligned} \int_0^T \lambda_t - \mathbb{E}[\lambda_t] \operatorname{d}t &= \int_0^T \int_{[0,t)} e^{-(B-A \operatorname{diag}(m^1, \dots, m^d))(t-s)} \operatorname{Ad} \mathbf{M}_s \operatorname{d}t \\ &= \int_0^T \int_0^T \mathbb{1}_{s < t} e^{-(B-A \operatorname{diag}(m^1, \dots, m^d))(t-s)} \operatorname{Ad} \mathbf{M}_s \operatorname{d}t, \quad \text{and using Fubini's identity} \\ &= \int_0^T \int_s^T e^{-(B-A \operatorname{diag}(m^1, \dots, m^d))(t-s)} \operatorname{d}t \operatorname{Ad} \mathbf{M}_s \\ &= \int_0^T \int_0^{T-s} e^{-(B-A \operatorname{diag}(m^1, \dots, m^d))u} \operatorname{d}u \operatorname{Ad} \mathbf{M}_s, \quad \text{using a change of variables} \\ &= \int_0^T \phi(T-s) \operatorname{Ad} \mathbf{M}_s, \end{aligned}$$

where  $\phi$  is the anti-derivative of  $u \rightarrow e^{-(B-A \operatorname{diag}(m^1, \dots, m^d))u}$  that vanishes at zero. In absence of common jumps, the integration by parts formula is

$$0 = \phi(0) \mathbf{A} \mathbf{M}_T - \phi(T) \mathbf{A} \mathbf{M}_0 = \int_0^T \operatorname{d}(\phi(T-s)) \mathbf{A} \mathbf{M}_s + \phi(T-s) \operatorname{Ad} \mathbf{M}_s,$$

hence

$$\int_0^T \lambda_t - \mathbb{E}[\lambda_t] \operatorname{d}t = \int_0^T e^{-(B-A \operatorname{diag}(m^1, \dots, m^d))(T-s)} \mathbf{A} \mathbf{M}_s \operatorname{d}s.$$

□



*Lemma 4.4.4.* For any fixed  $n, m$  in  $\llbracket 1, d \rrbracket$  we have for any  $T \geq 0$

$$\mathbb{E}[(\lambda_T^n - \mathbb{E}[\lambda_T^n])(\lambda_T^m - \mathbb{E}[\lambda_T^m])] \leq K$$

where  $K$  is a positive constant independent from  $T$ .

*Proof.* Using the first equality of Lemma 4.4.3 we have

$$\begin{aligned} & (\boldsymbol{\lambda}_T - \mathbb{E}[\boldsymbol{\lambda}_T]) (\boldsymbol{\lambda}_T - \mathbb{E}[\boldsymbol{\lambda}_T])^\top \\ &= \int_{[0, T)} e^{-(B-A \operatorname{diag}(m^1, \dots, m^d))(T-s)} \operatorname{Ad} \mathbf{M}_s \int_{[0, T)} d\mathbf{M}_s^\top A^\top e^{-(B-A \operatorname{diag}(m^1, \dots, m^d))^\top (T-s)}, \end{aligned}$$

which yields using the multivariate Ito isometry

$$\begin{aligned} & \mathbb{E} \left[ (\boldsymbol{\lambda}_T - \mathbb{E}[\boldsymbol{\lambda}_T]) (\boldsymbol{\lambda}_T - \mathbb{E}[\boldsymbol{\lambda}_T])^\top \right] \\ &= \int_{[0, T)} e^{-(B-A \operatorname{diag}(m^1, \dots, m^d))(T-s)} \operatorname{Ad} \mathbb{E} \left[ [\mathbf{M}]_s \right] A^\top e^{-(B-A \operatorname{diag}(m^1, \dots, m^d))^\top (T-s)}, \end{aligned}$$

where  $d[\mathbf{M}]$  is the quadratic variation matrix's infinitesimal growth

$$\begin{aligned} d([\mathbf{M}]_s)_{ij} &= d[M^i, M^j]_s \\ &= \mathbb{1}_{i \neq j} d[M^i]_s \\ &= \mathbb{1}_{i \neq j} \mathbb{E}[\lambda_s^i] ds. \end{aligned}$$

Since the expected value of  $\lambda_s^i$  is bounded by a constant independent from  $T$  and since all of  $(B - A \operatorname{diag}(m^1, \dots, m^d))$ 's eigenvalues are positive we obtain the result.  $\square$

*Lemma 4.4.5.* Assume that Assumptions 4.1.2 and 4.1.4 are in force. For  $i, j \in \llbracket 1, d \rrbracket$  let  $I_T$  be the quantity defined in (4.3.1), then we have

$$\frac{1}{T} I_T \leq K.$$

*Proof.* Let  $m \leq n$  be two integers in  $\llbracket 1, p \rrbracket$  and  $i, j$  in  $\llbracket 1, d \rrbracket$ . We start by expanding (4.3.1)

$$\begin{aligned} I_T &= \mathbb{E} \left[ \mathbb{E}_X \left[ \left( \int_0^{v_m T} X \lambda_t^i \hat{M}_{v_n T}^{j, i, t, X} dt \right)^2 \right] \right] \\ &= 2 \mathbb{E} \left[ \mathbb{E}_X \left[ \int_0^{v_m T} \int_0^t X^2 \lambda_t^i \hat{M}_{v_n T}^{j, i, t, X} \lambda_s^i \hat{M}_{v_n T}^{j, i, s, X} ds dt \right] \right] \\ &= 2 \int_{\mathbb{R}_+} \int_0^{v_m T} \int_0^t x^2 \mathbb{E} \left[ \lambda_t^i \lambda_s^i \mathbb{E}_t \left[ \hat{M}_{v_n T}^{j, i, t, x} \hat{M}_{v_n T}^{j, i, s, x} \right] \right] ds dt \nu^i(dx), \end{aligned}$$

where we recall that  $\hat{M}_{v_n T}^{j,i,t,x} = \hat{L}_{v_n T}^{j,i,t,x} - m^j \int_t^{v_n T} \hat{\lambda}_s^{j,i,t,x} ds$  is a martingale. Thus the product's expectation is

$$\mathbb{E}_t \left[ \hat{M}_{v_n T}^{j,i,t,x} \hat{M}_{v_n T}^{j,i,s,x} \right] = \mathbb{E}_t \left[ \hat{M}_{v_n T}^{j,i,t,x} \left( \hat{M}_{v_n T}^{j,i,s,x} - \hat{M}_t^{j,i,s,x} \right) \right] + \mathbb{E}_t \left[ \hat{M}_{v_n T}^{j,i,t,x} \hat{M}_t^{j,i,s,x} \right],$$

the last term vanishes since for any  $t \leq v_n T \leq v_n T$ ,  $\mathbb{E}_t \left[ \hat{M}_{v_n T}^{j,i,t,x} \hat{M}_t^{j,i,s,x} \right] = \hat{M}_t^{j,i,s,x} \mathbb{E}_t \left[ \hat{M}_{v_n T}^{j,i,t,x} \right] = 0$ . Thus

$$\begin{aligned} \mathbb{E}_t \left[ \hat{M}_{v_n T}^{j,i,t,x} \hat{M}_{v_n T}^{j,i,s,x} \right] &= \mathbb{E}_t \left[ \int_t^{v_n T} \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} y^2 \mathbb{1}_{\theta \leq \hat{\lambda}_u^{j,i,t,x}} \mathbb{1}_{\theta \leq \hat{\lambda}_u^{j,i,s,x}} d\theta \nu^j(dy) du \right] \\ &= \mathbb{E}_t \left[ \int_t^{v_n T} \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} y^2 \mathbb{1}_{\theta \leq \min(\hat{\lambda}_u^{j,i,t,x}, \hat{\lambda}_u^{j,i,s,x})} d\theta \nu^j(dy) du \right] \\ &= \int_{\mathbb{R}_+} y^2 \nu^j(dy) \int_t^{v_n T} \mathbb{E}_t \left[ \min(\hat{\lambda}_u^{j,i,t,x}, \hat{\lambda}_u^{j,i,s,x}) \right] du \\ &\leq K \int_t^{v_n T} \mathbb{E}_t \left[ \hat{\lambda}_u^{j,i,s,x} \right] du. \end{aligned}$$

As shown in Remark 4.2.11,  $\hat{\lambda}_u^{i,s,x}$  has the same dynamics as a Hawkes process  $\tilde{\lambda}_u^{i,s,x}$  that satisfies -if we follow the lines of the proof of Lemma 4.4.2- the SDE

$$d\tilde{\lambda}_u^{i,s,x} = -B\tilde{\lambda}_u^{i,s,x} du + A d\tilde{L}_u^{i,s,x} \quad (4.4.3)$$

whose solution yields

$$\mathbb{E}_t \left[ \tilde{\lambda}_u^{i,s,x} \right] = e^{-V(u-t)} \tilde{\lambda}_t^{i,s,x},$$

where  $V = B - A \text{diag}(m^1, \dots, m^d)$  whose eigenvalues  $\rho_1, \dots, \rho_d$  are positive. Hence

$$\begin{aligned} \mathbb{E}_t \left[ \hat{M}_{v_n T}^{j,i,t,x} \hat{M}_{v_n T}^{j,i,s,x} \right] &\leq K \left[ \int_t^{v_n T} e^{-V(u-t)} \tilde{\lambda}_t^{i,s,x} du \right]^j \\ &= K \left[ V^{-1} \left( I_d - e^{-V(v_n T-t)} \right) \tilde{\lambda}_t^{i,s,x} \right]^j. \end{aligned}$$

By plugging this inequality in  $I_T$ 's expression we get

$$\begin{aligned}
 I_1 &\leq K \int_{\mathbb{R}_+} \int_0^{v_m T} \int_0^t x^2 \mathbb{E} \left[ \lambda_t^i \lambda_s^i \left[ V^{-1} \left( I_d - e^{-V(v_n T - t)} \right) \tilde{\lambda}_t^{i,s,x} \right]^j \right] ds dt \nu^i(dx) \\
 &= K \int_{\mathbb{R}_+} \int_0^{v_m T} \int_0^t x^2 \left[ V^{-1} \left( I_d - e^{-V(v_n T - t)} \right) \mathbb{E} \left[ \lambda_t^i \lambda_s^i \tilde{\lambda}_t^{i,s,x} \right]^j \right] ds dt \nu^i(dx) \\
 &= K \int_{\mathbb{R}_+} \int_0^{v_m T} \int_0^t x^2 \left[ V^{-1} \left( I_d - e^{-V(v_n T - t)} \right) \mathbb{E} \left[ \lambda_s^i \mathbb{E}_s \left[ \lambda_t^i \tilde{\lambda}_t^{i,s,x} \right] \right]^j \right] ds dt \nu^i(dx) \\
 &= K \int_{\mathbb{R}_+} \int_0^{v_m T} \int_0^t x^2 \left[ V^{-1} \left( I_d - e^{-V(v_n T - t)} \right) \mathbb{E} \left[ \lambda_s^i \mathbb{E}_s \left[ \lambda_t^i \tilde{\lambda}_t^{i,s,x} \right] \right]^j \right] ds dt \nu^i(dx),
 \end{aligned}$$

since  $\tilde{\lambda}_t^{i,s,x}$  starts at  $s$ ,  $\mathbb{E}_s \left[ \lambda_t^i \tilde{\lambda}_t^{i,s,x} \right] = \mathbb{E}_s \left[ \lambda_t^i \right] \mathbb{E} \left[ \tilde{\lambda}_t^{i,s,x} \right]$ . By solving the expectation value of the SDE (4.4.3) with the initial condition  $\mathbb{E} \left[ \tilde{\lambda}_s^{i,s,x} \right] = \mathbf{A}_i x$  we get

$$\mathbb{E} \left[ \tilde{\lambda}_t^{i,s,x} \right] = e^{-V(t-s)} \mathbf{A}_i x,$$

which yields after being plugged in the last inequality

$$\begin{aligned}
 I_T &\leq K \int_{\mathbb{R}_+} \int_0^{v_m T} \int_0^t x^2 \left[ V^{-1} \left( I_d - e^{-V(v_n T - t)} \right) e^{-V(t-s)} \mathbf{A}_i x \right]^j \mathbb{E} \left[ \lambda_s^i \mathbb{E}_s \left[ \lambda_t^i \right] \right] ds dt \nu^i(dx) \\
 &\leq K \int_{\mathbb{R}_+} \int_0^{v_m T} \int_0^t x^3 \left[ V^{-1} \left( I_d - e^{-V(v_n T - t)} \right) e^{-V(t-s)} \mathbf{A}_i \right]^j \mathbb{E} \left[ \lambda_s^i \lambda_t^i \right] ds dt \nu^i(dx).
 \end{aligned}$$

By combining Cauchy-Schwarz's inequality and Lemma 4.4.4 we have that

$$\mathbb{E} \left[ \lambda_s^i \lambda_t^i \right] \leq \mathbb{E} \left[ |\lambda_s^i|^2 \right]^{1/2} \mathbb{E} \left[ |\lambda_t^i|^2 \right]^{1/2} \leq K,$$

thus

$$\begin{aligned}
 I_T &\leq K \int_{\mathbb{R}_+} x^3 \nu^i(dx) \int_0^{v_m T} \left[ V^{-1} \left( I_d - e^{-V(v_n T - t)} \right) \int_0^t e^{-V(t-s)} ds \mathbf{A}_i \right]^j dt \\
 &\leq K \int_{\mathbb{R}_+} x^3 \nu^i(dx) \int_0^{v_m T} \left[ V^{-1} \left( I_d - e^{-V(v_n T - t)} \right) V^{-1} \left( I_d - e^{-Vt} \right) \mathbf{A}_i \right]^j dt \\
 &\leq K \int_0^{v_m T} \left[ V^{-2} \left( I_d - e^{-V(v_n T - t)} \right) \left( I_d - e^{-Vt} \right) \mathbf{A}_i \right]^j dt \\
 &= K \int_0^{v_m T} \left[ V^{-2} \left( I_d - e^{-V(v_n T - t)} - e^{-Vt} + e^{-Vv_n T} \right) \mathbf{A}_i \right]^j dt,
 \end{aligned}$$

and since  $v_m T \leq v_n T$ ,

$$\begin{aligned} I_T &\leq K \int_0^{v_n T} \|I_d\|_{op} + \left\| e^{-V(v_n T - t)} \right\|_{op} + \|e^{-Vt}\|_{op} + \|e^{-VT}\|_{op} dt \\ &\leq K \int_0^{v_n T} 1 + (1 + (v_n T - t)^{d-1}) e^{-(\beta - \rho_d)(v_n T - t)} + (1 + t^{d-1}) e^{-(\beta - \rho_d)t} \\ &\quad + (1 + v_n T^{d-1}) e^{-(\beta - \rho_d)v_n T} dt \\ &\leq KT. \end{aligned}$$

□

*Lemma 4.4.6.* Set  $\mathbf{F}_T = \frac{\mathbf{L}_T - \text{diag}(m^1, \dots, m^d) \int_0^T \boldsymbol{\lambda}_t dt}{\sqrt{T}}$  and  $\mathbf{Y}_T = \frac{\mathbf{L}_T - \text{diag}(m^1, \dots, m^d) \int_0^T \mathbb{E}[\boldsymbol{\lambda}_t] dt}{\sqrt{T}}$ . Then we have the equality

$$\mathbf{Y}_T = J\mathbf{F}_T + \mathbf{R}_T, \quad (4.4.4)$$

where

$$J = \left( I_d - \text{diag}(m^1, \dots, m^d) B^{-1} A \right)^{-1}$$

and

$$\mathbf{R}_T = \text{diag}(m^1, \dots, m^d) \left( B - A \text{diag}(m^1, \dots, m^d) \right)^{-1} \frac{\mathbb{E}[\boldsymbol{\lambda}_T] - \boldsymbol{\lambda}_T}{\sqrt{T}}.$$

*Proof.* By taking the expected value of SDE (4.4.1) we have the system

$$\begin{cases} d\boldsymbol{\lambda}_t = B(\boldsymbol{\mu} - \boldsymbol{\lambda}_t)dt + A d\mathbf{L}_t, \\ d\mathbb{E}[\boldsymbol{\lambda}_t] = B(\boldsymbol{\mu} - \mathbb{E}[\boldsymbol{\lambda}_t])dt + A \text{diag}(m^1, \dots, m^d) \mathbb{E}[\boldsymbol{\lambda}_t] dt, \end{cases}$$

which yields after integrating with respect to time

$$\begin{cases} \boldsymbol{\lambda}_T - \boldsymbol{\mu} = B\boldsymbol{\mu}T - B \int_0^T \boldsymbol{\lambda}_t dt + A\mathbf{L}_T, \\ \mathbb{E}[\boldsymbol{\lambda}_T] - \boldsymbol{\mu} = B\boldsymbol{\mu}T - (B - A \text{diag}(m^1, \dots, m^d)) \int_0^T \mathbb{E}[\boldsymbol{\lambda}_t] dt. \end{cases}$$

In order to involve the quantities of interest  $\mathbf{F}_T$  and  $\mathbf{Y}_T$ , we state the fact that  $\text{diag}(m^1, \dots, m^d)$  is invertible (since  $\nu^1, \dots, \nu^d$  are supported by  $\mathbb{R}_+^*$ ), hence

$$\begin{cases} \boldsymbol{\lambda}_T - \boldsymbol{\mu} = B\boldsymbol{\mu}T - B \text{diag}(m^1, \dots, m^d)^{-1} \text{diag}(m^1, \dots, m^d) \int_0^T \boldsymbol{\lambda}_t dt + A\mathbf{L}_T, \\ \mathbb{E}[\boldsymbol{\lambda}_T] - \boldsymbol{\mu} = B\boldsymbol{\mu}T - (B \text{diag}(m^1, \dots, m^d)^{-1} - A) \text{diag}(m^1, \dots, m^d) \int_0^T \mathbb{E}[\boldsymbol{\lambda}_t] dt, \end{cases}$$

which yields by adding and subtracting  $\mathbf{L}_T$

$$\begin{cases} \boldsymbol{\lambda}_T - \boldsymbol{\mu} = B\boldsymbol{\mu}T + B \text{diag}(m^1, \dots, m^d)^{-1} \left( \sqrt{T} \mathbf{F}_T - \mathbf{L}_T \right) + A\mathbf{L}_T, \\ \mathbb{E}[\boldsymbol{\lambda}_T] - \boldsymbol{\mu} = B\boldsymbol{\mu}T + (B \text{diag}(m^1, \dots, m^d)^{-1} - A) \left( \sqrt{T} \mathbf{Y}_T - \mathbf{L}_T \right). \end{cases}$$

Subtracting the first equation from the second yields

$$\begin{aligned}\mathbb{E}[\boldsymbol{\lambda}_T] - \boldsymbol{\lambda}_T &= \left( B \operatorname{diag}(m^1, \dots, m^d)^{-1} - A \right) \sqrt{T} \mathbf{Y}_T - B \operatorname{diag}(m^1, \dots, m^d)^{-1} \sqrt{T} \mathbf{F}_T \\ &= \left( B - A \operatorname{diag}(m^1, \dots, m^d) \right) \operatorname{diag}(m^1, \dots, m^d)^{-1} \sqrt{T} \mathbf{Y}_T \\ &\quad - B \operatorname{diag}(m^1, \dots, m^d)^{-1} \sqrt{T} \mathbf{F}_T.\end{aligned}$$

Since  $(B - \operatorname{diag}(m^1, \dots, m^d)A)$  is invertible, we have

$$\begin{aligned}&\left( B - A \operatorname{diag}(m^1, \dots, m^d) \right)^{-1} \frac{\mathbb{E}[\boldsymbol{\lambda}_T] - \boldsymbol{\lambda}_T}{\sqrt{T}} \\ &= \operatorname{diag}(m^1, \dots, m^d)^{-1} \mathbf{Y}_T \\ &\quad - \left( B - A \operatorname{diag}(m^1, \dots, m^d) \right)^{-1} B \operatorname{diag}(m^1, \dots, m^d)^{-1} \mathbf{F}_T \\ &= \operatorname{diag}(m^1, \dots, m^d)^{-1} \mathbf{Y}_T \\ &\quad - \left( \operatorname{diag}(m^1, \dots, m^d) - \operatorname{diag}(m^1, \dots, m^d) B^{-1} A \operatorname{diag}(m^1, \dots, m^d) \right)^{-1} \mathbf{F}_T.\end{aligned}$$

Multiplying to the left by  $\operatorname{diag}(m^1, \dots, m^d)$  we get

$$\begin{aligned}&\operatorname{diag}(m^1, \dots, m^d) \left( B - A \operatorname{diag}(m^1, \dots, m^d) \right)^{-1} \frac{\mathbb{E}[\boldsymbol{\lambda}_T] - \boldsymbol{\lambda}_T}{\sqrt{T}} \\ &= \mathbf{Y}_T - \left( I_d - \operatorname{diag}(m^1, \dots, m^d) B^{-1} A \right)^{-1} \mathbf{F}_T.\end{aligned}$$

□

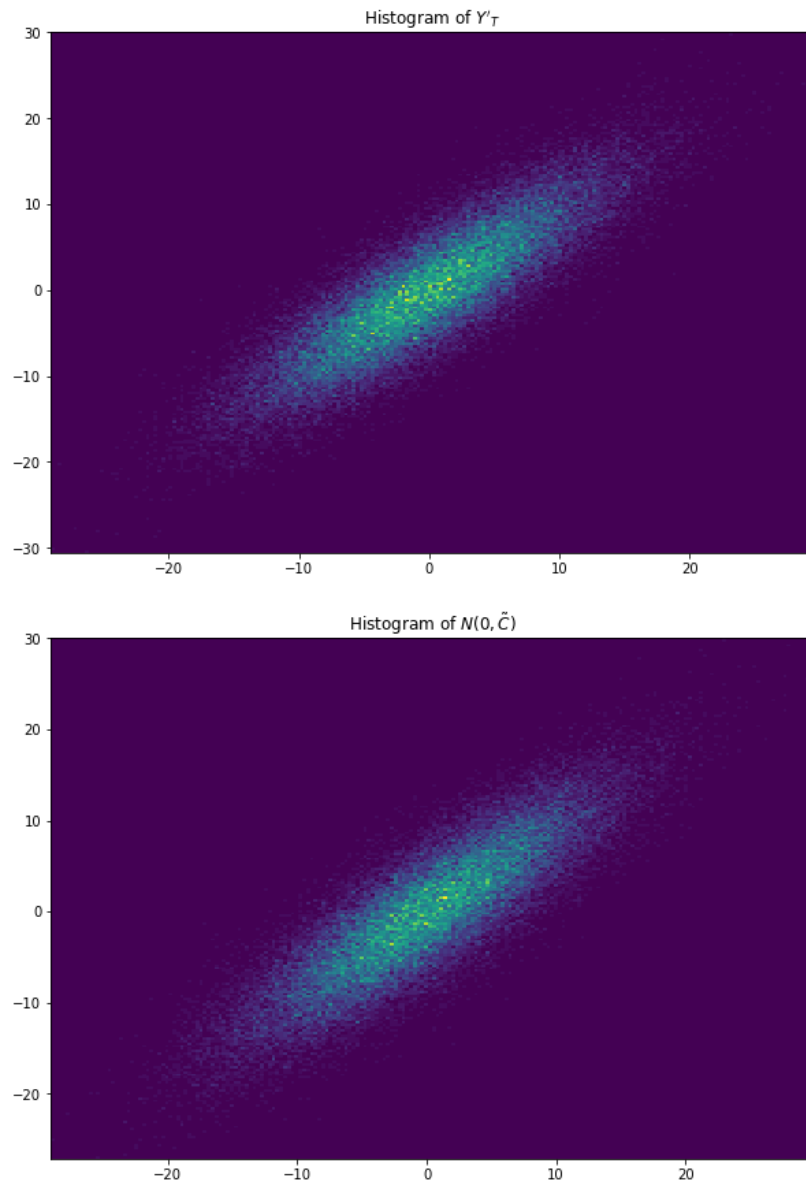


Figure 4.1 – 2 –  $D$  histograms of  $\mathbf{Y}'_T$  and  $\mathcal{N}(0, \tilde{C})$  for  $\beta = 4$ ,  $\boldsymbol{\mu} = (2, 3)^\top$  and  $T = 1000$  for  $n = 40000$  Monte Carlo samples.

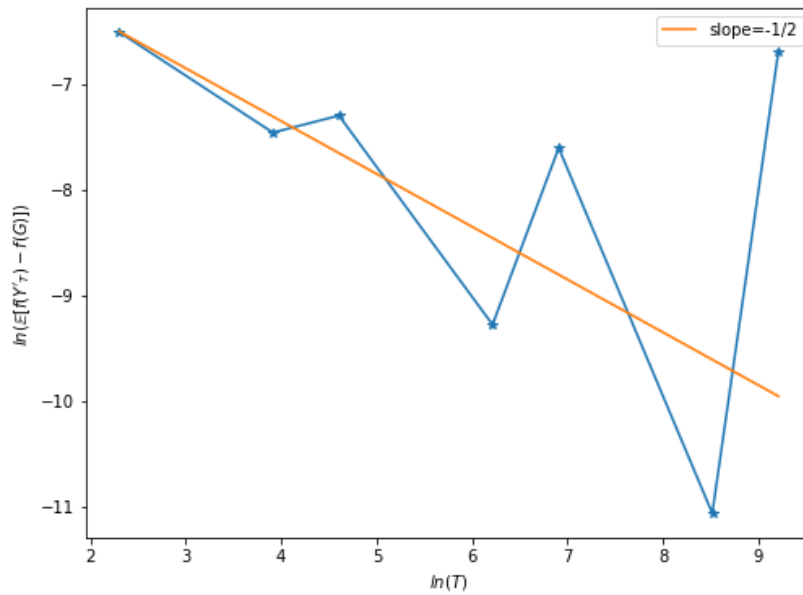


Figure 4.2 – Monte Carlo estimator  $\frac{1}{n} \sum_{k=1}^n f(\mathbf{Y}_T^k) - \mathbb{E}[f(\mathbf{G})]$  (in blue) for  $\beta = 6$ ,  $T = [10, 50, 100, 500, 1000, 5000, 10000]$  and  $n = 150000$ . Notice how for large  $T$ , the Monte Carlo estimator's error becomes stronger.

# The marked Hawkes process as a health insurance model

---

*This chapter corresponds to a work in progress with C. Hillairet, L. Huang and A. Réveillac.*

## Abstract

In this chapter, a more general *marked* Hawkes process is introduced. This process is used as a model of a health insurance policy, for which a central limit theorem and a large deviation principle are given.

Following [86] and [54], we also study the asymptotic infinite horizon ruin probability, with more detailed proofs.

## Contents

---

|            |  |            |
|------------|--|------------|
| <b>5.1</b> | <b>Introduction</b>  | <b>135</b> |
| <b>5.2</b> | <b>Quantitative CLT for the marked Hawkes process</b>          | <b>137</b> |
| <b>5.3</b> | <b>Large deviation principle</b>                               | <b>140</b> |
| 5.3.1      | The Laplace transform of the Poisson cluster                   | 141        |
| 5.3.2      | The Gärtner-Ellis theorem                                      | 143        |
| <b>5.4</b> | <b>The asymptotic ruin probability of the insurance policy</b> | <b>145</b> |
| 5.4.1      | Theoretical results  | 145        |
| 5.4.2      | A numerical example  | 147        |
| <b>5.5</b> | <b>Lemmata</b>   | <b>150</b> |

---

## 5.1 Introduction

In many real life applications, the random times of a given point process correspond to certain random marks that represent the main object of study. For instance, one can think of the magnitude of a given earthquake in seismology [71], the change in price after a meta-order [58], or the cost of a surgery for a health insurance company.

In this chapter, we focus on the last example. The health insurance company  $C$  has a client  $A$  who subscribes to a contract to hedge medical costs  $(X_i)_{i \in \mathbb{N}^*}$  at future



times  $(\tau_i)_{i \in \mathbb{N}^*}$ . The total loss of the company until time  $T$  takes the form

$$L_T = \sum_{i=1}^{N_T} g(X_i),$$

where  $N_T = \sum_{i \geq 1} \mathbb{1}_{\tau_i \leq T}$  is the ground process that counts the number of claims until  $T$ . The claims are supposed to be modulated by a function  $g$ , which in reality can be the effect of a cap on the amount the company reimburses, or a re-insurance contract that guarantees that  $C$  pays nothing above a fixed amount  $Q_g$ .

The ground process itself is assumed to be a *marked* Hawkes process, whose intensity  $\lambda$  follows the dynamics

$$\begin{aligned} \lambda_t &= \mu + \int_{[0,t)} \phi(t-s)b(X_{N_s})dN_s \\ &= \mu + \sum_{\tau_i < t} \phi(t-\tau_i)b(X_i), \end{aligned} \quad (5.1.1)$$

where  $\mu$  is a positive baseline intensity,  $\phi$  is a non-negative kernel and  $b$  a positive function. Throughout this chapter, the claims' amounts  $(X_i)_{i \in \mathbb{N}}$  are supposed to be *i.i.d* random variables of distribution  $\nu$ , but they are not independent from the ground process. In fact, like in the previous chapters, the marked Hawkes process is constructed by thinning from a 3 component Poisson measure  $\Pi(dt, d\theta, dx)$  whose intensity measure is  $dt \otimes d\theta \otimes \nu(dx)$

$$\begin{cases} L_t &= \int_{(0,t] \times \mathbb{R}_+^2} \mathbb{1}_{\theta \leq \lambda_s} g(x) \Pi(ds, d\theta, dx) \\ N_t &= \int_{(0,t] \times \mathbb{R}_+^2} \mathbb{1}_{\theta \leq \lambda_s} \Pi(ds, d\theta, dx) \\ \lambda_t &= \mu + \int_{(0,t) \times \mathbb{R}_+^2} \phi(t-s) \mathbb{1}_{\theta \leq \lambda_s} b(x) \Pi(ds, d\theta, dx) \end{cases} .$$

The marks  $X$  are not only influencing  $L$  by their presence as claims (as in chapter 3) but are also influencing the ground process  $N$  via their presence in  $\lambda$ . This influence is modulated by the function  $b$ , which in the health insurance context can be understood as follows:

Large claims correspond to serious health issues, *e.g.* serious surgeries, that tend to generate complications, which in turn mean more claims to the insurance company. For instance, one can choose  $b(x) = \mathbb{1}_{x \geq Q_b}$  where  $Q_b$  is a "seriousness level" below which a claim is not expected to trigger any other complications.

*Remark 5.1.1.* The compound Hawkes process introduced in chapter 4 is a special case of the marked Hawkes process, with both  $b$  and  $g$  are taken to be equal to the identity function.

For a similar work we refer to Swishchuk *et al.* [90] who study an insurance model whose claims constitute a Markov chain (thus not independent among themselves) while being independent from the ground process  $N$ .

In this chapter, we follow closely the methodology used by Karabash and Zhu [54] which is based on the seminal work [86] to prove large deviation results for their model. However, the model studied in [54] supposes that the marks that influence  $\lambda$  are completely independent from the claims, which makes it less realistic in the context of health insurance.

The chapter is organized as follows: In section 5.2 we prove a CLT on the total loss  $L$  with an exponential or an Erlang kernel. In section 5.3 the representation the linear marked Hawkes process as Poisson cluster processes is used to prove a large deviation principle (LDP) on the total loss. Finally, we give the asymptotics of the ruin probability of a health insurance policy in section 5.4, which we illustrate with a numerical example.

## 5.2 Quantitative CLT for the marked Hawkes process

In this section, we generalize the Wasserstein bound proven in chapters 3 and 4 to the marked Hawkes process. Throughout this section, we assume that  $g(X)$  and  $b(X)$  have a finite moment of order 3 and we set  $m_g = \mathbb{E}[g(X)]$  and  $m_b = \mathbb{E}[b(X)]$ . We also state the following two assumptions

**Assumption 5.2.1.** The self-excitation kernel  $\phi$  is assumed to be an exponential function

$$\phi(t) = \alpha e^{-\beta t},$$

or an Erlang function

$$\phi(t) = \alpha t e^{-\beta t},$$

for positive  $\alpha$  and  $\beta$ .

**Assumption 5.2.2.** The function  $b$  is assumed to be positive and

$$m_b \|\phi\|_1 = \int b(x) \nu(dx) \int \phi(t) dt < 1.$$

*Theorem 5.2.3.* Let  $L$  be a marked Hawkes process whose intensity  $\lambda$  follows the dynamics (5.1.1). Assume that Assumptions 5.2.1 and 5.2.2 are in force and that  $\mathbb{E}[g(X)^3] + \mathbb{E}[b(X)^3] < +\infty$ .

Set

$$\Gamma_T = \frac{L_T - \frac{m_g \mu}{1 - m_b \|\phi\|_1} T}{\sqrt{T}}.$$

Then, we have that

$$d_W(\Gamma_T, \mathcal{N}(0, \tilde{\sigma}^2)) = O\left(\frac{1}{\sqrt{T}}\right),$$

where

$$\tilde{\sigma}^2 = \frac{\mu \mathbb{E}\left[\left((1 - m_b \|\phi\|_1)g(X) + m_g \|\phi\|_1 b(X)\right)^2\right]}{(1 - m_b \|\phi\|_1)^3}.$$

*Proof.* We start by proving this result for the variable

$$V_T = \frac{L_T - m_g \int_0^T \mathbb{E}[\lambda_t] dt}{\sqrt{T}}.$$

We then introduce the auxiliary process  $\rho_t = \int_{(0,t] \times \mathbb{R}_+^2} b(x) \mathbf{1}_{\theta \leq \lambda_s} \Pi(ds, d\theta, dx)$  and its associated martingale  $M_t^b = \rho_t - m_b \int_0^t \lambda_s ds$ . We also introduce  $\chi^b = \sum_{n \geq 1} m_b \phi^{*n}$ . By artificially adding and subtracting  $\frac{m_g}{m_b} \rho_t$  we have that

$$\begin{aligned} \sqrt{T}V_T &= L_T - m_g \int_0^T \mathbb{E}[\lambda_t] dt \\ &= \left( L_T - \frac{m_g}{m_b} \rho_T \right) + \frac{m_g}{m_b} \left( \rho_T - m_b \int_0^T \mathbb{E}[\lambda_t] dt \right) \\ &= \left( L_T - \frac{m_g}{m_b} \rho_T \right) + \frac{m_g}{m_b} \left( M_T^b + \int_0^T \chi^b(s) M_{T-s}^b ds \right), \end{aligned}$$

where the last equality comes from Lemma 4 in [7].

Similarly for the normalized martingale  $F_T = \frac{L_T - m_g \int_0^T \lambda_t dt}{\sqrt{T}}$  we have that

$$\begin{aligned} \sqrt{T}F_T &= L_T - \frac{m_g}{m_b} \rho_T + \frac{m_g}{m_b} \rho_T - m_g \int_0^T \lambda_t dt \\ &= \left( L_T - \frac{m_g}{m_b} \rho_T \right) + \frac{m_g}{m_b} M_T^b. \end{aligned}$$

Thus, we have that

$$\begin{aligned} \sqrt{T} \left( V_T - \frac{F_T}{1 - m_b \|\phi\|_1} \right) &= \frac{-m_b \|\phi\|_1}{1 - m_b \|\phi\|_1} \left( L_T - \frac{m_g}{m_b} \rho_T \right) \\ &\quad + \frac{m_g}{m_b} \left( \frac{-m_b \|\phi\|_1}{1 - m_b \|\phi\|_1} M_T^b + \int_0^T \chi^b(s) M_{T-s}^b ds \right) \\ &= \frac{-m_b \|\phi\|_1}{1 - m_b \|\phi\|_1} \left( L_T - \frac{m_g}{m_b} \rho_T \right) \\ &\quad + \frac{m_g}{m_b} \left( - \int_0^{+\infty} \chi^b(s) ds M_T^b + \int_0^T \chi^b(s) M_{T-s}^b ds \right) \end{aligned}$$

because  $\int_0^{+\infty} \chi^b(s) ds = m_b \|\phi\|_1 / (1 - m_b \|\phi\|_1)$ , and after re-arranging the terms

$$V_T = \delta^\Pi \left( \left( \frac{\mathbf{1}_{t \leq T}}{\sqrt{T}} \mathbf{1}_{\theta \leq \lambda_t} \frac{g(x)(1 - m_b \|\phi\|_1) + m_g \|\phi\|_1 b(x)}{1 - m_b \|\phi\|_1} \right)_{(t, \theta, x)} \right) + \mathfrak{R}_T$$

where

$$\mathfrak{R}_T = \frac{1}{\sqrt{T}} \left( \int_0^T \chi^b(s) (M_{T-s}^b - M_T^b) ds - M_T^b \int_T^{+\infty} \chi^b(s) ds \right).$$

We start by bounding the remainder

$$\begin{aligned}\mathbb{E}[\mathfrak{R}_T^2] &\leq 2 \left( \frac{1}{T} \mathbb{E}[(M^b)_T^2] \left( \int_T^{+\infty} \chi^b(s) ds \right)^2 + \frac{1}{T} \mathbb{E} \left[ \left( \int_0^T \chi^b(s) (M_T^b - M_{T-s}^b) ds \right)^2 \right] \right) \\ &= 2(B_1 + B_2).\end{aligned}$$

Using the fact that the expected value of the square of a martingale is the expected value of its quadratic variation which in this case is the process' jumps we have that

$$\begin{aligned}B_1 &= \frac{1}{T} \mathbb{E} \left[ [M^b]_T \right] \left( \int_T^{+\infty} \chi^b(s) ds \right)^2 \\ &= \frac{1}{T} \mathbb{E} [\rho_T] \left( \int_T^{+\infty} \chi^b(s) ds \right)^2 \\ &= O(1) \left( \int_T^{+\infty} \chi^b(s) ds \right)^2.\end{aligned}$$

Hence, from the fact that  $\int u\phi(u)du < +\infty$  and the proof of [7, Lemma 5]

$$\begin{aligned}B_1 &\leq O(1) \left( \int_T^{+\infty} \frac{s}{T} \chi^b(s) ds \right)^2 \\ &\leq O\left(\frac{1}{T^2}\right) \left( \int_0^{+\infty} s \chi^b(s) ds \right)^2 \\ &= O\left(\frac{1}{T^2}\right).\end{aligned}$$

By expanding the square, the second term yields

$$\begin{aligned}B_2 &= \frac{1}{T} \mathbb{E} \left[ \left( \int_0^T \chi^b(s) (M_T^b - M_{T-s}^b) ds \right)^2 \right] \\ &= \frac{2}{T} \mathbb{E} \left[ \int_0^T \int_0^s \chi^b(s) (M_T^b - M_{T-s}^b) \chi^b(r) (M_T^b - M_{T-r}^b) dr ds \right] \\ &= \frac{2}{T} \int_0^T \int_0^s \chi^b(s) \chi^b(r) \mathbb{E} \left[ ((M_T^b)^2 - M_T^b M_{T-s}^b - M_T^b M_{T-r}^b + M_{T-r}^b M_{T-s}^b) \right] dr ds.\end{aligned}$$

Since for any  $u \leq v$ ,  $\mathbb{E}[M_u^b M_v^b] = \mathbb{E}[(M_u^b)^2] = \mathbb{E}[\rho_u]$ , we have that

$$\begin{aligned}B_2 &= \frac{2}{T} \int_0^T \int_0^s \chi^b(s) \chi^b(r) (\mathbb{E}[\rho_T] - \mathbb{E}[\rho_{T-s}] - \mathbb{E}[\rho_{T-r}] + \mathbb{E}[\rho_{T-s}]) dr ds \\ &= \frac{2}{T} \int_0^T \int_0^s \chi^b(s) \chi^b(r) (\mathbb{E}[\rho_T] - \mathbb{E}[\rho_{T-r}]) dr ds.\end{aligned}$$

In order to bound the integral, we use once again [7, Lemma 4] and Lemma 5.5.1 to obtain

$$\begin{aligned}\mathbb{E}[\rho_T] - \mathbb{E}[\rho_{T-r}] &= r\mu + \left( r \int_0^{T-r} \chi^b(u) du + \int_{T-r}^T \chi^b(u)(T-u) du \right) \mu \\ &= r\mu + \left( r \int_0^{T-r} \chi^b(u) du + \int_0^r \chi^b(T-u)u du \right) \mu \\ &\leq r\mu + \left( r\|\chi^b\|_1 + \int_0^r \chi^b(T-u)r du \right) \mu = O(r),\end{aligned}$$

and since  $\chi^b$  is positive

$$\begin{aligned}B_2 &\leq \frac{C}{T} \int_0^T \int_0^s \chi^b(s)\chi^b(r)r dr ds \\ &\leq \frac{C}{T} \|\chi^b\|_1 \int_0^T r\chi^b(r) dr \\ &= O\left(\frac{1}{T}\right)\end{aligned}$$

For the divergence term, we apply Lemma 5.5.2 with  $j(x) = \frac{g(x)(1-m_b\|\phi\|_1)+m_g\|\phi\|_1 b(x)}{1-m_b\|\phi\|_1}$  which yields using the proof of Theorem 3.3.13

$$d_W(V_T, \mathcal{N}(0, \tilde{\sigma}^2)) = O\left(\frac{1}{\sqrt{T}}\right).$$

Finally, using [7, Lemma 5]

$$\begin{aligned}V_T - \Gamma_T &= \frac{m_g}{\sqrt{T}} \left( \int_0^T \mathbb{E}[\lambda_t] dt - \frac{\mu T}{1 - m_b\|\phi\|_1} \right) \\ &= O\left(\frac{1}{\sqrt{T}}\right),\end{aligned}$$

which yields the desired result.  $\square$

*Remark 5.2.4.* Even though this result looks similar to Theorem 3.3.13 and Theorem 4.1.5, it has a fundamentally different proof. Since the action of the claims  $X$  on the intensity (via the function  $b$ ) is different from its impact on the total loss (via the function  $g$ ), which makes it impossible to connect the process  $V$  and the normalized martingale  $F$  by combining the SDEs satisfied by  $\lambda$ .

This means that it is possible to extend the result of Theorem 5.2.3 for Hawkes processes with general kernels if one has the bound on the distance between the normalized martingale and its Gaussian limit.

### 5.3 Large deviation principle

Another type of asymptotic analysis for this class of processes has been studied in the literature and takes the form of a *large deviation principle*. We start the section

with the following reminder: we say that a family of random variables  $(X_t)_{t \in \mathbb{R}_+}$  satisfies a large deviation principle (LDP) with rate function  $J$  and speed  $t$  if  $J$  is non-negative and lower semi continuous such that for any measurable set  $A$  we have

$$-\inf_{x \in A} J(x) \leq \liminf_{x \rightarrow +\infty} \frac{1}{t} \log \mathbb{P}(X_t \in A) \leq \limsup_{x \rightarrow +\infty} \frac{1}{t} \log \mathbb{P}(X_t \in A) \leq -\inf_{x \in A} J(x)$$

In order to prove an LDP for the loss process  $L$ , we follow the methodology of Bordenave and Torrisi [15] which consists of expressing the loss process  $L$  as a Poisson cluster. This can be done thanks to the fact that the ground process is a linear Hawkes process.

In this section, we exploit the Poisson cluster representation to derive an implicit equation of the Laplace transform (or moment generating function, MGF for short) of the weight of a cluster, which then will be linked to the asymptotic logarithmic MGF of  $L$ . Finally, we establish an LDP using the Gärtner-Ellis theorem.

### 5.3.1 The Laplace transform of the Poisson cluster

From now on we assume the assumption 5.2.2 is in force. On the other hand, we only assume that

$$\int_0^{+\infty} u\phi(u)du < +\infty,$$

without  $\phi$  necessarily being exponential or Erlang.

Let  $L'$  be the Poisson cluster created as follows

1. Immigrants arrive according to a Poisson process of rate  $\mu$ . Each immigrant  $I_j$  arrives at time  $\tau_j^I$  has a mark  $X_j^I$  simulated according to the distribution  $\nu$ . The points  $(\tau_j^I, g(X_j^I))$  are added to  $L'$ .
2. Each immigrant  $I_j$  generates offspring according to a non-homogeneous Poisson process of intensity  $\phi(\cdot - \tau_j^I)b(X_j^I)$ . The finite offspring occurs at times  $\tau_j^{O_1}$  and have *i.i.d* marks  $X_j^{O_1}$ .
3.  $(\tau_j^{O_1}, g(X_j^{O_1}))$  are added to  $L'$ .
4. Each member of the first generation's offspring generates offspring according to the same rule, which is applied to all of its descendants.
5. The tree of the descendants of each immigrant (cluster) is finite since it is a sub critical Galton Watson tree.

Then  $(L'((0, t]))_{t \in \mathbb{R}_+}$  is a marked Hawkes process, whose ground process' intensity follows the dynamics (5.1.1)

$$\lambda_t = \mu + \sum_{\tau_i < t} \phi(t - \tau_i)b(X_i).$$

This construction is the key for obtaining the implicit expression of the asymptotic logarithmic MGF.

*Proposition 5.3.1.* Let  $L$  be a marked Hawkes process whose intensity solves equation (5.1.1) and assume that Assumption 5.2.2 is in force. Assume also that

$$\mathbb{E}[e^{\eta_1 b(X) + \eta_2 g(X)}] < +\infty, \quad (5.3.1)$$

for some positive  $\eta_1$  and  $\eta_2$ .

The limit of the loss' logarithmic MGF is

$$\lim_{t \rightarrow +\infty} \frac{1}{t} \log \mathbb{E}[e^{\eta L_t}] = \begin{cases} \mu (f(\eta) - 1), & \eta \leq \eta_c \\ +\infty, & \eta > \eta_c, \end{cases} \quad (5.3.2)$$

where  $f$  is the minimal solution of the equation

$$x = \int_{\mathbb{R}_+} e^{\eta g(a) + b(a) \|\phi\|_1 (x-1)} \nu(da), \quad (5.3.3)$$

and

$$\begin{cases} 1 &= \int_{\mathbb{R}_+} b(a) \|\phi\|_1 e^{\eta_c g(a) + b(a) \|\phi\|_1 (x_c - 1)} \nu(da) \\ x_c &= \int_{\mathbb{R}_+} e^{\eta_c g(a) + b(a) \|\phi\|_1 (x_c - 1)} \nu(da). \end{cases} \quad (5.3.4)$$

*Proof.* This proof follows closely the proof of Theorem 4 in [54] but with more details. We recall that the compound Hawkes process can be seen as a Poisson cluster, and thus if we call  $S_{\tau_j^I}$  the tree of marked descendants generated by an immigrant  $I_j$  and  $S_{\tau_j^I}(t)$  the sum of all marks (modulated by  $g$ ) that happened before  $t$ , we have that:

$$L_t = \sum_{\tau_j^I \leq t} S_{\tau_j^I}(t).$$

The Laplace transform becomes

$$\begin{aligned} \mathbb{E}[e^{\eta L_t}] &= \mathbb{E}[e^{\eta \sum_{\tau_j^I \leq t} S_{\tau_j^I}(t)}], \\ &= \sum_{k=0}^{+\infty} e^{-\mu t} \frac{(\mu t)^k}{k!} \mathbb{E} \left[ \mathbb{E} \left[ e^{\eta \sum_{j=1}^k S_{\tau_j^I}(t)} \mid \tau_1^I, \dots, \tau_k^I \right] \right], \end{aligned}$$

where the last equality comes from the fact that immigrants arrive according to a Poisson process of intensity  $\mu$ . Conditioned on the ancestors (immigrants) the trees are independent and identically distributed, but one has to take into account the fact that we only look at the descendants before  $t$ , hence:

$$\begin{aligned} \mathbb{E}[e^{\eta L_t}] &= \sum_{k=0}^{+\infty} e^{-\mu t} \frac{(\mu t)^k}{k!} \mathbb{E} \left[ \prod_{j=1}^k \mathbb{E} \left[ e^{\eta S_{\tau_j^I}(t)} \mid \tau_1^I, \dots, \tau_k^I \right] \right] \\ &= \sum_{k=0}^{+\infty} e^{-\mu t} \frac{(\mu t)^k}{k!} \mathbb{E} \left[ \prod_{j=1}^k \mathbb{E} \left[ e^{\eta S(t - \tau_j^I)} \mid \tau_1^I, \dots, \tau_k^I \right] \right], \end{aligned}$$

where  $S(t) = S_0(t)$  is a tree that has an ancestor at time 0.

Keeping in mind that knowing that the Poisson process (of immigrants) had  $k$  arrivals in  $[0, t]$  means that  $\tau_1^I, \dots, \tau_k^I$  are distributed as  $k$  ordered uniform variables on  $[0, t]$  (same goes for  $t - \tau_1, \dots, t - \tau_k$ ) we conclude that

$$\begin{aligned} \mathbb{E}[e^{\eta L_t}] &= \sum_{k=0}^{+\infty} e^{-\mu t} \frac{(\mu t)^k}{k!} \left( \int_0^t \mathbb{E}[e^{\eta S(x)}] dx \right)^k \frac{1}{t^k} \\ &= \exp \left( \mu \left( \int_0^t \mathbb{E}[e^{\eta S(x)}] - 1 dx \right) \right), \quad \text{by the change of variables } x = tz \\ &= \exp \left( \mu t \left( \int_0^1 \mathbb{E}[e^{\eta S(tz)}] - 1 dz \right) \right). \end{aligned}$$

And using the monotone convergence theorem:

$$\begin{aligned} \frac{1}{t} \log (\mathbb{E}[e^{\eta L_t}]) &= \mu \int_0^1 (\mathbb{E}[e^{\eta S(tz)}] - 1) dz \\ &\xrightarrow{t \rightarrow +\infty} \mu (\mathbb{E}[e^{\eta S}] - 1), \end{aligned}$$

where  $S$  is the sum of all the marks, modulated by  $g$ , of the descendants of a random immigrant. Finally, using Lemma 5.5.3, we have the implicit expression of  $f(\eta) = \mathbb{E}[e^{\eta S}]$ .  $\square$

We now conclude regarding the LDP on  $L$ , by linking the rate function with the limit logarithmic MGF.

### 5.3.2 The Gärtner-Ellis theorem

Unlike the paper of Karabash and Zhu [54] in which the authors use a tilting argument, we directly use the Gärtner-Ellis theorem (Theorem 2.3.6 in [29]) to prove the large deviation principle. It roughly says that the rate function is nothing but the dual of the asymptotic logarithmic MGF.

*Theorem 5.3.2.* Let  $L$  be a marked Hawkes process, and assume that all the assumptions of Proposition 5.3.1 are in force.

Set  $\Lambda(\eta) = \mu (f(\eta) - 1)$ , where  $f$  is defined in Proposition 5.3.1. Then  $(L_T/T)_{T>0}$  satisfies an LDP with rate function

$$\Lambda^*(x) = \sup_{\eta \leq \eta_c} (\eta x - \Lambda(\eta)) = \begin{cases} \eta^* x - \mu(x^* - 1) & x \geq 0 \\ +\infty & x < 0 \end{cases}, \quad (5.3.5)$$

where  $(\eta^*, x^*)$  solves the system

$$\begin{cases} \frac{x}{\mu} &= \int_{\mathbb{R}_+} \left( g(a) + b(a) \frac{x}{\mu} \|\phi\|_1 \right) e^{\eta^* g(a) + b(a) \|\phi\|_1 (x^* - 1)} \nu(da) \\ x^* &= \int_{\mathbb{R}_+} e^{\eta^* g(a) + b(a) \|\phi\|_1 (x^* - 1)} \nu(da) \end{cases}. \quad (5.3.6)$$

*Proof.* We prove the LDP from the limiting logarithmic MGF using the Gärtner-Ellis theorem [29]. To do so,  $\Lambda$  needs to be essentially smooth, *i.e.*



1. The interior of  $\Lambda$ 's definition domain is not empty ( $\mathcal{D}_\Lambda^\circ \neq \emptyset$ ).
2.  $\Lambda$  is differentiable throughout  $\mathcal{D}_\Lambda^\circ$ .
3.  $\Lambda$  is steep, *i.e.*  $\lim_{\eta \rightarrow \eta_c} \Lambda'(\eta) = +\infty$ .

For the first two properties, we recall that  $\Lambda(\eta) = \mu(f(\eta) - 1) = \mu(\mathbb{E}[e^{\eta S}] - 1)$  with  $S$  the sum of the marks of all the descendants of an immigrant and that is defined on  $(-\infty, \eta_c]$ . We now prove the steepness of  $f$  (consequently that of  $\Lambda$ ). First, we notice that

$$f'(\eta) = \mathbb{E}[S e^{\eta S}]$$

is increasing on  $(0, \eta_c]$ . Thus, either it is bounded and has a finite limit  $l$ , or it diverges to  $+\infty$ . In this proof we suggest a *reductio ad absurdum* and we suppose that  $f'(\eta_c) = l < +\infty$ . Stabile and Torrisi [86] suggest a more direct proof, because the explicit expression of the distribution of  $S$  for the simple (non marked) Hawkes process is available.

Differentiating the equation

$$f(\eta) = \int_{\mathbb{R}_+} e^{\eta c g(a) + b(a) \|\phi\|_1 (f(\eta) - 1)} \nu(da)$$

with respect to  $\eta$  and evaluating at  $\eta = \eta_c$  yields

$$\begin{aligned} f'(\eta_c) &= \int_{\mathbb{R}_+} (g(a) + b(a) \|\phi\|_1 f'(\eta_c)) e^{\eta_c a + b(a) \|\phi\|_1 (f(\eta_c) - 1)} \nu(da) \\ &= \int_{\mathbb{R}_+} g(a) e^{\eta_c g(a) + b(a) \|\phi\|_1 (f(\eta_c) - 1)} \nu(da) + f'(\eta_c) \int_{\mathbb{R}_+} b(a) \|\phi\|_1 e^{\eta_c g(a) + b(a) \|\phi\|_1 (f(\eta_c) - 1)} \nu(da) \end{aligned}$$

From the second equation in (5.3.4), we have that  $\int_{\mathbb{R}_+} b(a) \|\phi\|_1 e^{\eta_c g(a) + b(a) \|\phi\|_1 (f(\eta_c) - 1)} \nu(da) = 1$ , thus

$$l = \int_{\mathbb{R}_+} g(a) e^{\eta_c g(a) + b(a) \|\phi\|_1 (f(\eta_c) - 1)} \nu(da) + l$$

which is impossible for a finite  $l$  since the finitude of  $\eta_c$  ensures that the integral  $\int_{\mathbb{R}_+} g(a) e^{\eta_c a + b(a) \|\phi\|_1 (f(\eta_c) - 1)} \nu(da)$  is strictly positive. Hence

$$\lim_{\eta \rightarrow \eta_c} f'(\eta) = +\infty$$

and  $L_T/T$  satisfies a LDP with a function rate  $\Lambda^*$ , the Fenchel-Legendre transform of  $\Lambda$ .

To determine  $\Lambda^*$  we recall that  $\Lambda$  is twice differentiable on  $(-\infty, \eta_c)$ , thus

$$\frac{d^2}{d\eta^2} (\eta x - \Lambda(\eta)) = -\mu f''(\eta) = -\mu \mathbb{E}[S^2 e^{\eta S}] < 0.$$

To find the supremum we locate the critical points by solving, for a given  $x$  in  $\eta$

$$\begin{aligned} 0 &= x - \Lambda'(\eta^*) \\ &= x - \mu f'(\eta^*) \\ \frac{x}{\mu} &= f'(\eta^*). \end{aligned}$$

And the derivative of  $f$  evaluated at  $\eta^*$  satisfies the equation

$$\begin{aligned} f'(\eta^*) &= \int_{\mathbb{R}_+} (g(a) + b(a)\|\phi\|_1 f'(\eta^*)) e^{\eta^* g(a) + b(a)\|\phi\|_1 (f(\eta^*) - 1)} \nu(da) \\ \frac{x}{\mu} &= \int_{\mathbb{R}_+} \left( g(a) + b(a)\|\phi\|_1 \frac{x}{\mu} \right) e^{\eta^* g(a) + b(a)\|\phi\|_1 (f(\eta^*) - 1)} \nu(da) \end{aligned}$$

and  $x^* = f(\eta^*)$  solves the equation

$$x^* = \int_{\mathbb{R}_+} e^{\eta^* g(a) + b(a)\|\phi\|_1 (x^* - 1)} \nu(da)$$

which yields the result.  $\square$

## 5.4 The asymptotic ruin probability of the insurance policy

We now make use of the LDP to study the asymptotics of a risk process' ruin probability as the initial capital increases to infinity.

### 5.4.1 Theoretical results

We consider a risk model that describes the total capital of an insurance policy

$$R_t = u + \rho t - L_t, \quad (5.4.1)$$

where  $u$  denotes the initial capital,  $\rho$  is the continuous rate charged by the insurer and  $L_t$  is the sum of claims occurring randomly before time  $t$ . If the loss is of the form  $L_t = \sum_{k=1}^{N_t} X_k$  with  $N$  a Poisson process and the  $(X_i)_{i \in \mathbb{N}}$  are *i.i.d* independent from  $N$ , then the model is called the Cramér-Lundberg model and it has been thoroughly studied in the literature, *cf.* [44] for example.

In this chapter we suppose that  $L_t = \sum_{k=1}^{N_t} g(X_k)$  where  $N$  is a marked Hawkes process whose intensity follows the dynamics (5.1.1). It is important for the insurer to know whether ruin (*i.e.* the capital goes below zero) occurs, that is why we introduce the ruin time given an initial capital  $u$

$$\tau_u = \inf_{t \geq 0} \{R_t \leq 0\}$$

as well as the infinite horizon ruin probability

$$\Psi(u) = \mathbb{P}[\tau_u < +\infty].$$

Generally, the explicit expression of  $\Psi(u)$  is not available, not even for the Cramér-Lundberg model. That is why, the asymptotic ruin probability is considered.

The first to solve this problem for a risk model with a simple (unmarked) Hawkes process claim arrivals and light-tailed claims were Stabile and Torrisi [86]. Their

result was extended to heavy-tailed claims by Zhu [95], and to marked Hawkes processes with claims independent from the marks by Karabash and Zhu [54].

In this chapter, the asymptotic ruin probability is extended to our marked Hawkes model, but we restrict ourselves to light-tailed distributions that satisfy (5.3.1).

By the law of large numbers, we have that

$$\lim_{t \rightarrow +\infty} \frac{R_t}{t} = \rho - \frac{\mu m_g}{1 - m_b \|\phi\|_1},$$

which means that ruin occurs almost surely if the right hand side is negative. Thus, we assume the following *net profit condition* (NPC for short)

$$\frac{\mu m_g}{1 - m_b \|\phi\|_1} < \rho.$$

We now state the result on the ruin probability  $\Psi(u)$  as the capital  $u$  grows to infinity

*Theorem 5.4.1.* Assume that all the assumptions in Proposition 5.3.1 as well as the NPC hold. Moreover, assume that

$$\rho < \frac{\mu(x_c - 1)}{\eta_c}, \quad (5.4.2)$$

where  $x_c$  and  $\eta_c$  are defined in Proposition 5.3.1, is in force. Then

$$\lim_{u \rightarrow +\infty} \frac{1}{u} \log \Psi(u) = -\omega,$$

where  $\omega$  is the unique positive solution of the equation  $\Lambda(\omega) = \rho\omega$ .

*Proof.* First, the inequality guarantees the existence and uniqueness of  $\omega$ . We set  $h(\eta) = \Lambda(\eta) - \rho\eta$ , defined on  $[0, \eta_c]$ .  $h$  is clearly convex,  $g(0) = \mu(f(0) - 1) - \rho = 0$  and

$$\begin{aligned} h'(0) &= \mu f'(0) - \rho \\ &= \mu \mathbb{E}[S] - \rho \\ &= \mu \frac{m_g}{1 - m_b \|\phi\|_1} - \rho < 0 \end{aligned}$$

by the NPC. For the critical value  $\eta_c$  we have that

$$\begin{aligned} h(\eta_c) &= \Lambda(\eta_c) - \rho\eta_c \\ &= \mu(x_c - 1) - \rho\eta_c > 0 \end{aligned}$$

if and only if inequality (5.4.2) is true, which is a necessary and sufficient condition for the existence and uniqueness of  $\omega$ .

The proof of this result is identical to that of Theorem 9 in [54] and is based on Theorem 2 in [43] which yields the desired result for discrete processes. In fact by considering  $Q_{n\epsilon}$  where

$$Q_t = L_t - \rho t$$

for a small positive  $\epsilon$  and defining  $\kappa_t(\eta) = \log \mathbb{E}[e^{\eta Q_t}]$  we have that  $\lim_{n \rightarrow +\infty} \frac{1}{n} \log \kappa_{n\epsilon}(\eta) = \epsilon(\Lambda(\eta) - \rho\eta)$ . Cf. [32] for more details on the ruin probabilities in continuous time.  $\square$

5.4.2 A numerical example

In this section, we examine an example where a semi-explicit formula exists. By semi-explicit we mean that the MGF of  $S$  is known up to the computation of Lambert’s function  $W$ .

Assume that the claims follow an exponential distribution of parameter  $\zeta$ , *i.e.*

$$\nu(da) = \mathbf{1}_{a \geq 0} \zeta e^{-\zeta a} da,$$

and those claims increase the intensity via the function

$$b(x) = \mathbf{1}_{x \geq Q_b}.$$

We assume that the insurer has a re-insurance contract that exonerates him from reimbursing the client beyond a threshold  $Q_g$ , or more precisely

$$g(x) = x \mathbf{1}_{0 \leq x \leq Q_g} + Q_g \mathbf{1}_{x \geq Q_g},$$

with  $Q_g \geq Q_b$ .

For these choices, the integral in equation (5.3.3) (where the unknown is  $x = f(\eta)$ ) can be computed which yields

$$x = c(\eta) + d(\eta)e^{\|\phi\|_1 x}, \tag{5.4.3}$$

where

$$\begin{cases} c(\eta) &= \zeta \frac{e^{(\eta-\zeta)Q_b} - 1}{\eta - \zeta} \\ d(\eta) &= \zeta e^{-\|\phi\|_1} \left( \frac{e^{(\eta-\zeta)Q_g} - e^{(\eta-\zeta)Q_b}}{\eta - \zeta} + \frac{e^{(\eta-\zeta)Q_g}}{\zeta} \right). \end{cases}$$

To find the solutions of equation (5.4.3) one must solve for the intersection between an exponential and an affine function. The intersection can be expressed in terms of Lambert’s function

$$f(\eta) = c(\eta) - \frac{W(-\|\phi\|_1 d(\eta)e^{\|\phi\|_1 c(\eta)})}{\|\phi\|_1}. \tag{5.4.4}$$

We now apply Theorem 5.4.1 to the risk process

$$R_t = u + \rho t - L_t$$

to express its asymptotic ruin probability. But first, one must make sure that it is feasible to charge a premium high enough to satisfy the NPC, but not too high that condition (5.4.2) no longer holds. To verify that these conditions are not mutually exclusive, we plot both  $\mu \frac{m_g}{1 - m_b \|\phi\|_1}$  and  $\mu \frac{x^* - 1}{\eta^*}$  as functions of  $\|\phi\|_1$ .

*Remark 5.4.2.* Figure 5.2 does not prove the compatibility of the NPC and inequality (5.4.2), not even for this particular model, because the lower and upper bound depend on many factors and not only  $\|\phi\|_1$ .

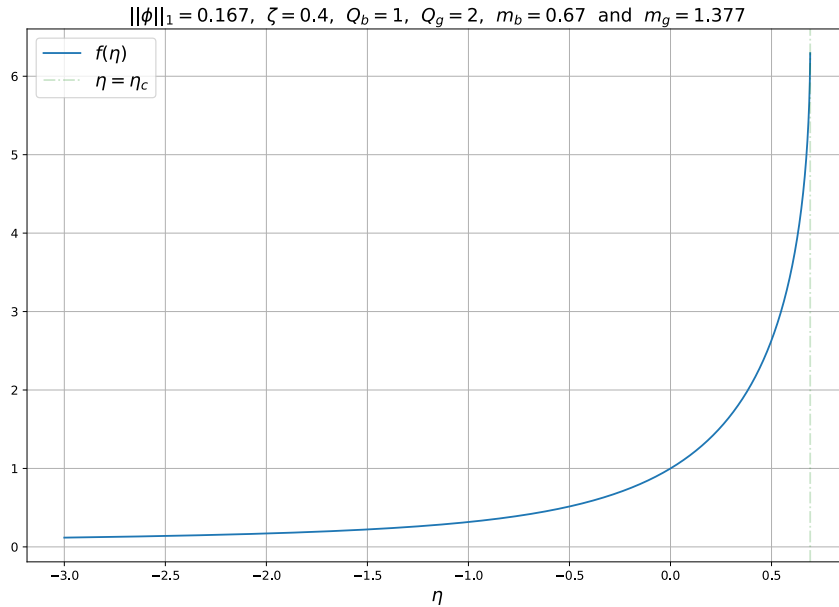


Figure 5.1 – An illustration of  $f(\eta) = \mathbb{E}[e^{\eta S}]$  for  $\eta \leq \eta_c$ . Notice how the curve has a vertical tangent at  $\eta_c$ .

For a kernel such that  $\|\phi\|_1 = 1/6$ , we chose the premium  $\rho = 20$  that satisfies both the NPC and condition (5.4.2) according to Figure 5.2. The equation  $\Lambda(\eta) = \rho\eta$  where  $\Lambda = \mu(f - 1)$  cannot be explicitly solved due to the presence of complex functions inside the Lambert function  $W$ . We hence solve it numerically as it is illustrated in Figure 5.3.

This means that, for a large initial capital  $u$ , one can roughly estimate that

$$\frac{1}{u} \log \Psi(u) \simeq -0.58.$$

For the sake of comparison, we run the same analysis on a risk model based on the classical compound Poisson process

$$R_t^p = u + \rho t - \sum_{i=1}^{N_t^p} g(X_i). \quad (5.4.5)$$

The claims are assumed to follow the distribution of  $g(X)$  where  $X$  has an exponential distribution with parameter  $\zeta$ . The ground point process is a homogeneous Poisson process  $N^p$  of intensity  $\lambda^p = \frac{\mu}{1 - m_b \|\phi\|_1}$ , which is exactly the stationary average intensity of the Hawkes process  $N$ , that counts the number of claims for the first model.

We are interested in evaluating this model's asymptotic ruin probability  $\Psi^p(u) =$

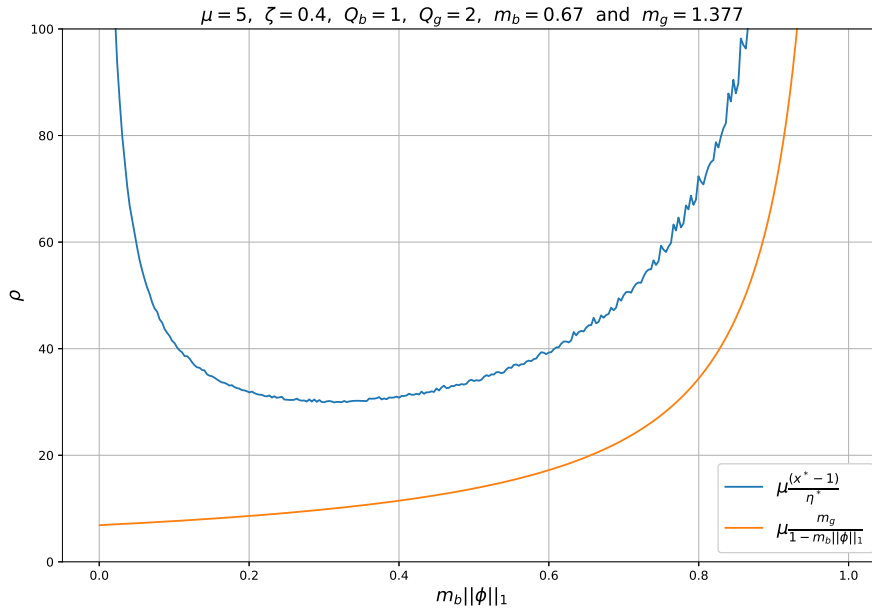


Figure 5.2 – For Theorem 5.4.1 to be applied, the premium  $\rho$  should be above the orange line and below the blue line.

$\mathbb{P}[\inf_{t \geq 0} R_t^p \leq 0]$ , when the initial capital  $u$  is large. Using the classical Cramér-Lundberg approximation for the compound Poisson model (*cf.* the book of Asmussen [3]), we have that, as soon as the net profit condition is satisfied

$$\lim_{u \rightarrow +\infty} \frac{1}{u} \Psi^p(u) = -\omega^p,$$

where  $\omega^p$  is the unique solution of the equation

$$\lambda^p (\alpha(\omega^p) - 1) = \rho \omega^p, \tag{5.4.6}$$

with

$$\alpha(\eta) = \mathbb{E}[e^{\eta g(X)}].$$

*Remark 5.4.3.* Notice that for the compound Poisson model, equation (5.4.6) has a unique solution if the NPC is respected. There is no upper bound on the premium, unlike the marked Hawkes model.

For the choice of function  $g$  made above, the moment generating function of  $g(X)$  can be explicitly computed

$$\alpha(\eta) = \mathbb{E}[e^{\eta g(X)}] = \zeta \frac{e^{(\eta-\zeta)Q_g} - 1}{\eta - \zeta} + e^{(\eta-\zeta)Q_g},$$

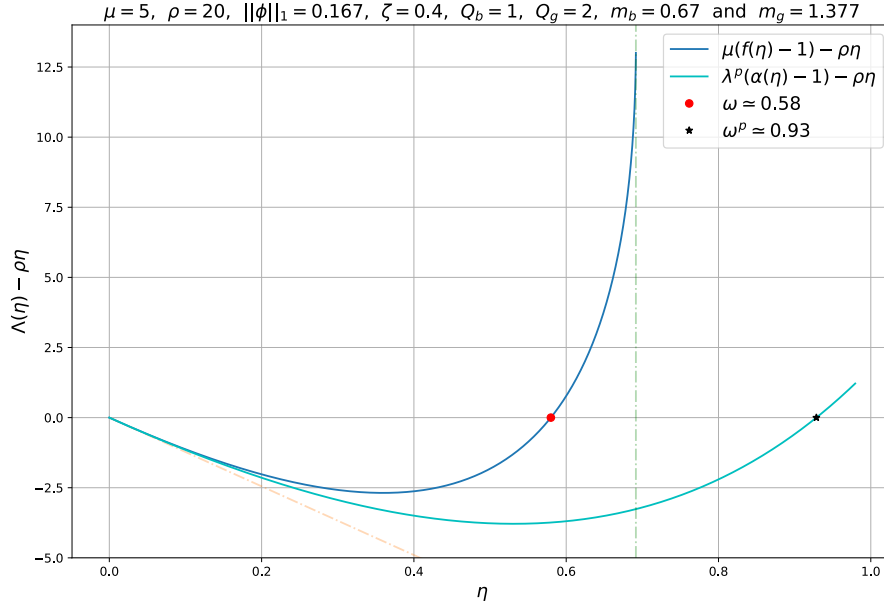


Figure 5.3 – The unique positive root of the equations  $\mu(f(\eta) - 1) = \rho\eta$  for the marked Hawkes model and  $\lambda^p(\alpha(\eta) - 1) = \rho\eta$  for the compound Poisson model. The fact that both curves have the same negative slope at  $\eta = 0$  shows that the same NPC is satisfied in both models. Notice how  $\omega^p$  is found beyond  $\eta_c$ , the limit value for the marked Hawkes model.

but equation (5.4.6) can only be solved numerically. The solution for the same parameters is  $\omega^p \simeq 0.93$ , as illustrated in Figure 5.3.

We notice that  $-\omega^p < -\omega$ , which in turn means (asymptotically)

$$\Psi^p(u) \leq \Psi(u).$$

This is probably because, even if the average intensity and claims are identical, one underestimates the risk by ignoring the clustering property of the Hawkes process.

## 5.5 Lemmata

*Lemma 5.5.1.* Assume that assumption 5.2.2 are in force. Set  $m_b = \mathbb{E}[b(X)]$  and

$$M_t^b = \rho_t - m_b \int_0^t \lambda_s ds = \int_{(0,t] \times \mathbb{R}_+^2} (\mathbf{1}_{\theta \leq \lambda_s} b(x)) \Pi(ds, d\theta, dx) - ds d\theta \nu(dx).$$

Then we have that

$$\mathbb{E}[\lambda_t] = \mu \left( 1 + \int_0^t \chi^b(t-s) ds \right)$$

and

$$\int_0^t (\lambda_s - \mathbb{E}[\lambda_s]) ds = \int_0^t \chi^b(t-s) M_s^b ds,$$

where  $\chi^b = \sum_{n \geq 1} (m_b \phi)^{*n}$ .

*Proof.* We recall that

$$\begin{aligned} \lambda_t &= \mu + \int_{[0,t)} \phi(t-s) b(X_{N_s}) dN_s \\ &= \mu + \int_{[0,t)} \phi(t-s) \int_{\mathbb{R}_+^2} b(x) \mathbf{1}_{\theta \leq \lambda_s} \Pi(ds, d\theta, dx), \end{aligned}$$

which yields after noticing

$$\int_{\mathbb{R}_+^2} b(x) \mathbf{1}_{\theta \leq \lambda_s} \Pi(ds, d\theta, dx) = dM_s^b + m_b \lambda_s ds,$$

that

$$\lambda_t = \mu + \int_0^t \phi(t-s) dM_s^b + \int_0^t m_b \phi(t-s) \lambda_s ds.$$

The rest follows from Theorem 1 in [7], since  $\|m_b \phi\|_1 = m_b \|\phi\|_1 < 1$  by assumption 5.2.2.  $\square$

*Lemma 5.5.2.* Let  $j$  be a function such that  $\mathbb{E}[j(X)^3] < +\infty$  and assume that assumptions 5.2.1 and 5.2.2 are in force. Set  $\vartheta_j^2 = \mathbb{E}[j(X)^2]$  and  $\sigma_j^2 = \frac{\vartheta_j^2 \mu}{1 - m_b \|\phi\|_1}$ . For any  $T > 0$  let

$$J_T = \delta^\Pi \left( \left( \frac{j(x) \mathbf{1}_{t \leq T} \mathbf{1}_{\theta \leq \lambda_t}}{\sqrt{T}} \right)_{(t,\theta,x)} \right)$$

be the normalized martingale of the loss process. Then there is a constant  $K$  independent from  $T$  such that

$$d_W(J_T, \mathcal{N}(0, \sigma_j^2)) \leq \frac{K}{\sqrt{T}},$$

where  $\mathcal{N}(0, \sigma^2)$  is the centered Gaussian distribution with variance  $\sigma^2$ .

*Proof.* We proceed in a similar manner to chapter 3. Using Theorem 3.3.10

$$d_W(F_T, \mathcal{N}(0, \sigma_j^2)) \leq A_1 + A_2$$

where

$$\begin{aligned} A_1 &:= \mathbb{E} \left[ \left| \sigma^2 - \frac{1}{\sqrt{T}} \int_0^T \int_{\mathbb{R}_+} j(x) \lambda_t D_{(t,0,x)} F_T dt \nu(dx) \right|^2 \right], \\ A_2 &:= \frac{1}{\sqrt{T}} \mathbb{E} \left[ \int_0^T \int_{\mathbb{R}_+} \lambda_t j(x) |D_{(t,0,x)} F_T|^2 dt \nu(dx) \right]. \end{aligned}$$



Proceeding in the same manner yields the decompositions

$$\begin{aligned} A_1 &= \mathbb{E} \left[ \left| \sigma^2 - \frac{\vartheta_j^2}{T} \int_0^T \lambda_t dt - \frac{1}{T} \int_0^T \int_{\mathbb{R}_+} g(x) \lambda_t \hat{M}_T^{t,x} dt \nu(dx) \right|^2 \right] \\ &\leq \left| \sigma^2 - \frac{\vartheta_j^2}{T} \int_0^T \mathbb{E}[\lambda_t] dt \right|^2 + \frac{\vartheta_j^2}{T} \mathbb{E} \left[ \left| \int_0^T (\lambda_t - \mathbb{E}[\lambda_t]) dt \right|^2 \right] + \frac{1}{T} \mathbb{E} \left[ \left| \int_0^T \int_{\mathbb{R}_+} \lambda_t \hat{M}_T^{t,x} \nu(dx) dt \right|^2 \right] \\ &=: A_{1,1} + \vartheta^2 A_{1,2} + A_{1,3}, \end{aligned}$$

and

$$\begin{aligned} A_2 &= \frac{1}{\sqrt{T}} \mathbb{E} \left[ \int_0^T \int_{\mathbb{R}_+} \lambda_t j(x) |D_{(t,\lambda_t,x)} F_T|^2 \nu(dx) dt \right] \\ &\leq \frac{2 \int_{\mathbb{R}_+} j(x)^3 \nu(dx)}{T^{3/2}} \mathbb{E} \left[ \int_0^T \lambda_t dt \right] + \frac{2}{T^{3/2}} \mathbb{E} \left[ \int_0^T \int_{\mathbb{R}_+} j(x) \lambda_t |\hat{M}_T^{t,x}|^2 \nu(dx) dt \right] \\ &=: 2 (\mathbb{E}[j(X)^3] A_{2,1} + A_{2,2}), \end{aligned}$$

where  $\hat{M}_T^{t,x} = D_{(t,0,x)} M_T - j(x)$ ,  $M$  being the martingale  $M_T = \sqrt{T} J_T$ .

Lemma 5.5.1 yields directly that

$$A_{1,1} + A_{1,2} + A_{2,1} = O\left(\frac{1}{\sqrt{T}}\right).$$

The bound on term  $A_{2,2}$  is obtained using the fact that  $\mathbb{E} \left[ |\hat{M}_T^{t,x}|^2 \right] = \mathbb{E}[j(X)^2] \mathbb{E} \left[ \int_t^T \hat{\lambda}_s^{t,x} ds \right]$  where

$$\mathbb{E}[\hat{\lambda}_s^{t,x}] = b(x) \phi(s-t) + \int_t^s m_b \phi(u-t) \mathbb{E}[\hat{\lambda}_u^{t,x}] du.$$

For the last term  $A_{1,3}$ , we proceed like Theorem 4.3.3 and Lemma 4.4.5  $\square$

*Lemma 5.5.3.* Let  $L'$  be the Poisson cluster defined in 5.3.1 and assume that assumption 5.2.2 is in force. Assume also that

$$\mathbb{E}[e^{\eta_1 b(X) + \eta_2 g(X)}] < +\infty,$$

for some positive  $\eta_1$  and  $\eta_2$ .

Let  $S$  be the sum of all the marks, modulated by  $g$ , of the descendants of a given immigrant, including itself. Then the MGF of  $S$

$$f(\eta) = \mathbb{E}[e^{\eta S}]$$

is the minimal solution of the equation

$$x = \int_{\mathbb{R}_+} e^{\eta g(a) + b(a) \|\phi\|_1 (x-1)} \nu(da)$$

which exists whenever  $\eta \leq \eta_c$  where  $\eta_c$  solves the system

$$\begin{cases} 1 &= \int_{\mathbb{R}_+} b(a) \|\phi\|_1 e^{\eta_c g(a) + b(a) \|\phi\|_1 (x_c-1)} \nu(da) \\ x_c &= \int_{\mathbb{R}_+} e^{\eta_c g(a) + b(a) \|\phi\|_1 (x_c-1)} \nu(da) \end{cases},$$

with  $x_c = f(\eta_c)$ .

*Proof.* Let  $\kappa$  be the random number of direct descendants of a given immigrant. Knowing the immigrant's mark  $X_1^I$ , we have that

$$\kappa \sim \text{Pois}(b(X_1^I)\|\phi\|_1).$$

Set  $S^1, \dots, S^\kappa$  the sums of marks (modulated by  $g$ ) of the trees of each descendant. Then one has

$$S = g(X_1^I) + \sum_{i=1}^{\kappa} S^i.$$

Let  $\eta$  be a parameter, one has:

$$\begin{aligned} f(\eta) &= \mathbb{E}[e^{\eta S}] = \mathbb{E}\left[e^{\eta g(X_1^I) + \eta \sum_{i=1}^{\kappa} S^i}\right] \\ &= \int_{\mathbb{R}_+} e^{\eta g(a)} \mathbb{E}\left[e^{\eta \sum_{i=1}^{\kappa} S^i}\right] \nu(da) \\ &= \int_{\mathbb{R}_+} e^{\eta g(a)} \sum_{k=1}^{+\infty} \mathbb{E}\left[e^{\eta \sum_{i=1}^k S^i}\right] \mathbb{P}(\kappa = k) \nu(da) \\ &= \int_{\mathbb{R}_+} e^{\eta g(a)} \sum_{k=1}^{+\infty} \mathbb{E}\left[e^{\eta S}\right]^k \mathbb{P}(\kappa = k) \nu(da) \quad \text{because } S^i \text{ are i.i.d } \sim S \\ &= \int_{\mathbb{R}_+} e^{\eta g(a)} \sum_{k=1}^{+\infty} f(\eta)^k \frac{(b(a)\|\phi\|_1)^k}{k!} e^{-b(a)\|\phi\|_1} \nu(da) \\ &= \int_{\mathbb{R}_+} e^{\eta g(a) + b(a)\|\phi\|_1} (f(\eta) - 1) \nu(da). \end{aligned}$$

Thus,  $f(\eta)$  is a solution of the equation

$$G_\eta(x) = 0,$$

where  $G_\eta(x) = \int_{\mathbb{R}_+} e^{\eta g(a) + b(a)\|\phi\|_1(x-1)} \nu(da) - x$ .

Note that for any  $\eta$ , the function  $G_\eta$  is convex and  $G_\eta(x) \rightarrow +\infty$  as  $x \rightarrow +\infty$ .  $G$  is also continuous in both of its arguments.

**We start with  $\eta = 0$ :** In this case,  $G_0(1) = 0$  and

$$\begin{aligned} G'_0(1) &= \int_{\mathbb{R}_+} e^{b(a)\|\phi\|_1(1-1)} b(a)\|\phi\|_1 \nu(da) - 1, \\ &= \mathbb{E}[b(X)]\|\phi\|_1 - 1, \\ &< 0, \quad \text{because of Assumption 5.2.2.} \end{aligned}$$

Which proves that  $G_0$  vanishes at exactly two points:  $x = 1$  and another point  $x' > 1$ . Of course, since  $f(0) = 1$ , one must take the minimal solution into account. **As  $\eta$  increases slightly:**  $G_\eta(1)$  increases from zero (continuously) and  $G'_\eta(1)$  as well. This means that the equation  $G_\eta(x) = 0$  will still have two solutions. By continuity,  $f(\eta)$  is the minimal one.

Since  $G_\eta(1) \rightarrow +\infty$  and  $G'_\eta(1) \rightarrow +\infty$  as  $\eta$  grows to infinity, this means that after

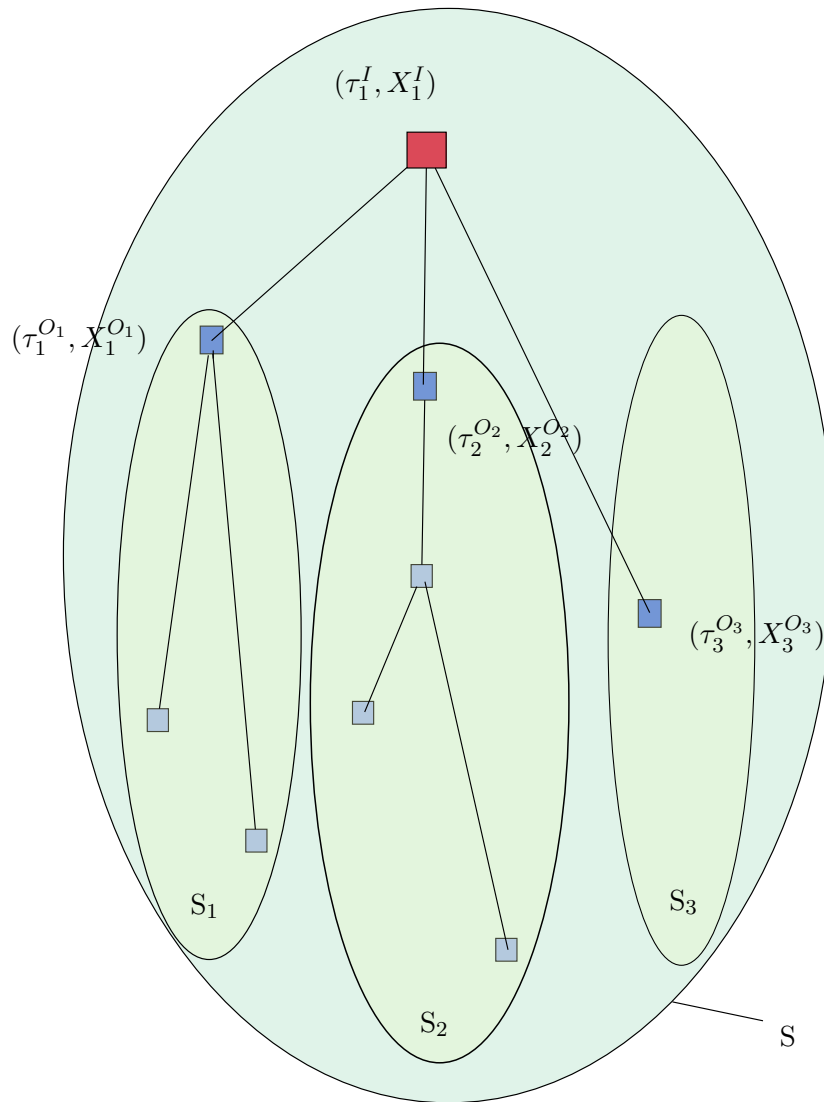


Figure 5.4 – On this figure,  $\kappa = 3$ . Note that by construction,  $S$  exhibits a fractal behaviour: the  $S^i$  have the same distribution as the whole tree  $S$ .

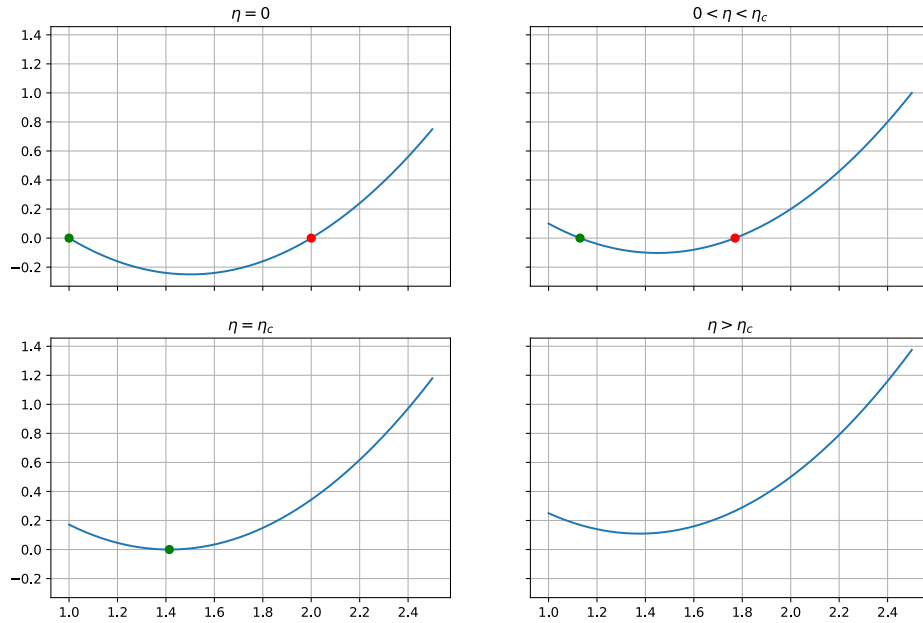


Figure 5.5 – An example of  $G_\eta(x)$  as  $\eta$  increases.

a certain critical value  $\eta_c$  the equation has no solutions and  $f(\eta)$  no longer exists.

**Critical value  $\eta = \eta_c$  :** At the critical value  $\eta_c$  the two roots fuse in one root  $f(\eta_c) = x_c$  that satisfies

$$G_{\eta_c}(x_c) = G'_{\eta_c}(x_c) = 0.$$

In this case  $(x_c, \eta_c)$  solve the system

$$\begin{cases} \int_{\mathbb{R}_+} e^{\eta_c g(a) + b(a)} \|\phi\|_1(x_c - 1) \nu(da) & = x_c \\ \int_{\mathbb{R}_+} b(a) \|\phi\|_1 e^{\eta_c g(a) + b(a)} \|\phi\|_1(x_c - 1) \nu(da) & = 1 \end{cases}.$$

This proof is illustrated on Figure 5.5. □



# Conclusions and perspectives

---

In the previous sections, we proved a number of novel results on Hawkes processes: the approximation of the nonlinear Hawkes process by time-series, a quantitative bound on the CLT satisfied by the linear Hawkes process in the univariate and the multi-variate settings as well as the asymptotic ruin probability of a health insurance policy driven by a marked Hawkes process.

We discuss here some perspectives that are implied by these results and we give potential hints to some open questions.

## 6.1 The approximation of the nonlinear Hawkes process by time series

*This has been studied in Chapter 2.*

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

The logical continuation would be to extend the same approximation result for the discretization of a Hawkes process with a general kernel  $\phi$ , *i.e.*

$$P_n \sim \text{Pois} \left( h\psi \left( \mu + \sum_{k=1}^{n-1} \phi(h(t_n - t_k))P_k \right) \right).$$

This means however that the standard Markov techniques are no longer exploitable and that one should look for an alternative method. A first guess that comes to mind is the classical tightness argument, combined with the convergence of finite-dimensional distributions.

Another major enhancement of the approximation result is to prove it for a general non bounded jump rate  $\psi$ . This should be merely a technical problem, as the result is valid if we chose to model the jumps with a Bernoulli distribution (instead of a Poisson), which is practically indistinguishable from our model as the time step  $h$  goes to zero.

A parametric calibration method has been studied numerically but without theoretical guarantees on the convergence of the estimated parameters to the ground

truth parameters. The experiments show that the calibrated parameters are close to the actual kernels despite the (reasonably small) Gaussian perturbations and the bias introduced by the discretization. This method can be enhanced for instance by adding a penalisation.

Hidden Markov chains techniques can be fruitful in the statistical calibration of this model as well.

## 6.2 The Malliavin-Setin method and the quantitative CLT for Hawkes processes

*This has been studied in Chapters 3 and 4.*

By relying on the Poisson embedding of general point processes, we managed to develop a Malliavin calculus on functionals of the Hawkes process. This was then combined with Stein's method to prove, for the first time, that the distance between the normalized Hawkes process and its Gaussian limit is of order  $O(1/\sqrt{T})$ . Even though the result was proven solely for two kernels (exponential and Erlang), it should be possible to extend it for the wider Erlang class of kernel, because they preserve the same Markov structure. However, computations are expected to be way more tedious.

We also have a reasonable hope that the same bound can be proven for a more general class of kernels, if the martingale term  $\hat{M}^t$  can be tamed without using Markov techniques.

When it comes to the nonlinear Hawkes processes, things become quickly more complicated as we do not have explicit formulae for the first moments, but it would be interesting to find out what error one commits when approximating the normalized process with its implicitly given average.

Another interesting perspective would be the study of the functional version of the approximation result. In fact, it is known that given a linear Hawkes process  $N$ , the normalized  $(N_{vT})_{v \in [0,1]}$  converges to a Brownian motion  $(W_v)_{v \in [0,1]}$ , but the speed of convergence is still unknown.

One might be tempted to generalize the result of the multi-marginals proven in chapter 4 to the functional setting, however we are pretty skeptical. The reason of this skepticism is that the dependence of the constant hidden in  $O(1/\sqrt{T})$  on the number of marginals remains unknown.

From a more practical point of view, quantifying the bound between the Hawkes process and a "close" Poisson process would be an important contribution. Indeed, as we saw on the numerical example of chapter 5, one does not estimate the risk of ruin correctly when replacing the marked Hawkes process by a compound Poisson of the same average intensity.

### 6.3 The risk process driven by a marked Hawkes process

*This has been studied in Chapter 5.*

Following closely the existing literature, we proved a large deviation principle on an insurance model that is based on a marked Hawkes process, for light-tailed claims. Similar results have already been proven for heavy-tailed claims.

Nevertheless, some questions on the current results naturally arise. For instance, in the proof of the ruin probability's asymptotic limit we assume that the premium  $\rho$  satisfies

$$\mu \frac{m_g}{1 - m_b \|\phi\|_1} < \rho < \mu \frac{x_c - 1}{\eta_c},$$

without giving a guarantee that the two extremal terms are in the correct order. The inequality is verified on some simple examples, however it is not possible to tell for sure in the general case because of the implicit nature of  $x_c$  and  $\eta_c$ .

Another natural question is the behaviour of the ruin probability if one charges an "excessive" premium that is larger than  $\mu(x_c - 1)/\eta_c$ . The ruin probability is expected to vanish at a different rate in this case.

From a practical point of view, obtaining approximations on the finite horizon probabilities is very interesting, as most people do not live forever, nor do their insurance policies. One can also make the model more realistic by taking a basket of (eventually correlated) policies and study their ruin probabilities. This would very likely call for the introduction of novel simulation and approximation methods.





# Bibliography

- [1] Massil Achab, Emmanuel Bacry, Stéphane Gaïffas, Iacopo Mastromatteo, and Jean-François Muzy. “Uncovering causality from multivariate Hawkes integrated cumulants”. In: *J. Mach. Learn. Res.* 18 (2017), Paper No. 192, 28. ISSN: 1532-4435.
- [2] Abdulhamid A. Alzaid and Mohamed Al-Osh. “An integer-valued  $p$ th-order autoregressive structure (INAR( $p$ )) process”. In: *J. Appl. Probab.* 27.2 (1990), pp. 314–324. ISSN: 0021-9002. DOI: [10.2307/3214650](https://doi.org/10.2307/3214650).
- [3] Søren Asmussen. *Ruin probabilities*. Vol. 2. Advanced Series on Statistical Science & Applied Probability. World Scientific Publishing Co., Inc., River Edge, NJ, 2000, pp. xii+385. ISBN: 981-02-2293-9. DOI: [10.1142/9789812779311](https://doi.org/10.1142/9789812779311).
- [4] Grigelionis B. “On representation of integer-valued random measures by means of stochastic integrals with respect to the Poisson measure”. In: *Lithuanian Mathematical Journal* 11.1 (1971), pp. 93–108. DOI: [10.15388/LMJ.1971.20963](https://doi.org/10.15388/LMJ.1971.20963).
- [5] Emmanuel Bacry, Martin Bompain, Stéphane Gaïffas, and Jean-François Muzy. “Sparse and low-rank multivariate Hawkes processes”. In: *J. Mach. Learn. Res.* 21 (2020), Paper No. 50, 32. ISSN: 1532-4435.
- [6] Emmanuel Bacry, Sylvain Delattre, Marc Hoffmann, and Jean-François Muzy. “Modelling microstructure noise with mutually exciting point processes”. In: *Quant. Finance* 13.1 (2013), pp. 65–77. ISSN: 1469-7688. DOI: [10.1080/14697688.2011.647054](https://doi.org/10.1080/14697688.2011.647054).
- [7] Emmanuel Bacry, Sylvain Delattre, Marc Hoffmann, and Jean-François Muzy. “Some limit theorems for Hawkes processes and application to financial statistics”. In: *Stochastic Process. Appl.* 123.7 (2013), pp. 2475–2499. ISSN: 0304-4149. DOI: [10.1016/j.spa.2013.04.007](https://doi.org/10.1016/j.spa.2013.04.007).
- [8] Emmanuel Bacry, Iacopo Mastromatteo, and Jean-François Muzy. “Hawkes processes in finance”. In: *Market Microstructure and Liquidity* 1.01 (2015), p. 1550005.
- [9] Emmanuel Bacry and Jean-François Muzy. “First- and second-order statistics characterization of Hawkes processes and non-parametric estimation”. In: *IEEE Trans. Inform. Theory* 62.4 (2016), pp. 2184–2202. ISSN: 0018-9448. DOI: [10.1109/TIT.2016.2533397](https://doi.org/10.1109/TIT.2016.2533397).
- [10] Henry Berbee. “Chains with infinite connections: uniqueness and Markov representation”. In: *Probab. Theory Related Fields* 76.2 (1987), pp. 243–253. ISSN: 0178-8051. DOI: [10.1007/BF00319986](https://doi.org/10.1007/BF00319986).
- [11] Eustache Besançon. “Speed of convergence of diffusion approximations”. PhD thesis. <https://tel.archives-ouvertes.fr/tel-03128781>: Télécom ParisTech, 2020.

- 
- [12] Eustache Besançon, Laure Coutin, Laurent Decreusefond, and Pascal Moyal. *Diffusive limits of Lipschitz functionals of Poisson measures*. 2021. DOI: [10.48550/ARXIV.2107.05339](https://doi.org/10.48550/ARXIV.2107.05339).
- [13] Yannick Bessy-Roland, Alexandre Boumezoued, and Caroline Hillairet. “Multivariate Hawkes process for cyber insurance”. In: *Annals of Actuarial Science* (2020).
- [14] Patrick Billingsley. *Convergence of probability measures*. Second. Wiley Series in Probability and Statistics: Probability and Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1999, pp. x+277. ISBN: 0-471-19745-9. DOI: [10.1002/9780470316962](https://doi.org/10.1002/9780470316962).
- [15] Charles Bordenave and Giovanni Luca Torrisi. “Large deviations of Poisson cluster processes”. In: *Stoch. Models* 23.4 (2007), pp. 593–625. ISSN: 1532-6349. DOI: [10.1080/15326340701645959](https://doi.org/10.1080/15326340701645959).
- [16] Björn Böttcher, René Schilling, and Jian Wang. *Lévy matters. III*. Vol. 2099. Lecture Notes in Mathematics. Lévy-type processes: construction, approximation and sample path properties, With a short biography of Paul Lévy by Jean Jacod, Lévy Matters. Springer, Cham, 2013, pp. xviii+199. ISBN: 978-3-319-02683-1; 978-3-319-02684-8. DOI: [10.1007/978-3-319-02684-8](https://doi.org/10.1007/978-3-319-02684-8).
- [17] Pierre Brémaud. *Point processes and queues*. Springer Series in Statistics. Martingale dynamics. Springer-Verlag, New York-Berlin, 1981, pp. xviii+354. ISBN: 0-387-90536-7.
- [18] Pierre Brémaud and Laurent Massoulié. “Stability of nonlinear Hawkes processes”. In: *Ann. Probab.* 24.3 (1996), pp. 1563–1588. ISSN: 0091-1798. DOI: [10.1214/aop/1065725193](https://doi.org/10.1214/aop/1065725193).
- [19] Nicolas Brunel. “Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons”. In: *Journal of computational neuroscience* 8.3 (2000), pp. 183–208.
- [20] Sourav Chatterjee and Elizabeth Meckes. “Multivariate normal approximation using exchangeable pairs”. In: *ALEA Lat. Am. J. Probab. Math. Stat.* 4 (2008), pp. 257–283.
- [21] Julien Chevallier. “Modelling large neural networks via Hawkes processes”. Theses. COMUE Université Côte d’Azur (2015 - 2019), 2016.
- [22] Manon Costa, Carl Graham, Laurence Marsalle, and Viet Chi Tran. “Renewal in Hawkes processes with self-excitation and inhibition”. In: *Adv. in Appl. Probab.* 52.3 (2020), pp. 879–915. ISSN: 0001-8678. DOI: [10.1017/apr.2020.19](https://doi.org/10.1017/apr.2020.19).
- [23] David R. Cox. “Some statistical methods connected with series of events”. In: *J. Roy. Statist. Soc. Ser. B* 17 (1955), 129–157; discussion, 157–164. ISSN: 0035-9246.

- [24] José Da Fonseca and Riadh Zaatour. “Hawkes Process: Fast Calibration, Application to Trade Clustering, and Diffusive Limit”. In: *Journal of Futures Markets* 34.6 (2014), pp. 548–579. DOI: <https://doi.org/10.1002/fut.21644>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/fut.21644>.
- [25] Daryl J. Daley and David Vere-Jones. *An introduction to the theory of point processes. Vol. I*. Second. Probability and its Applications (New York). Elementary theory and methods. Springer-Verlag, New York, 2003, pp. xxii+469. ISBN: 0-387-95541-0.
- [26] Daryl J. Daley and David Vere-Jones. *An introduction to the theory of point processes. Vol. II*. Second. Probability and its Applications (New York). General theory and structure. Springer, New York, 2008, pp. xviii+573. ISBN: 978-0-387-21337-8. DOI: [10.1007/978-0-387-49835-5](https://doi.org/10.1007/978-0-387-49835-5).
- [27] Angelos Dassios and Hongbiao Zhao. “Efficient simulation of clustering jumps with CIR intensity”. In: *Oper. Res.* 65.6 (2017), pp. 1494–1515. ISSN: 0030-364X. DOI: [10.1287/opre.2017.1640](https://doi.org/10.1287/opre.2017.1640).
- [28] Angelos Dassios and Hongbiao Zhao. “Exact simulation of Hawkes process with exponentially decaying intensity”. In: *Electron. Commun. Probab.* 18 (2013), no. 62, 13. DOI: [10.1214/ECP.v18-2717](https://doi.org/10.1214/ECP.v18-2717).
- [29] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*. Second. Vol. 38. Applications of Mathematics (New York). Springer-Verlag, New York, 1998, pp. xvi+396. ISBN: 0-387-98406-2. DOI: [10.1007/978-1-4612-5320-4](https://doi.org/10.1007/978-1-4612-5320-4).
- [30] Aline Duarte, Eva Löcherbach, and Guilherme Ost. “Stability, convergence to equilibrium and simulation of non-linear Hawkes processes with memory kernels given by the sum of Erlang kernels”. In: *ESAIM Probab. Stat.* 23 (2019), pp. 770–796. ISSN: 1292-8100. DOI: [10.1051/ps/2019005](https://doi.org/10.1051/ps/2019005).
- [31] Darrell Duffie, Jun Pan, and Kenneth Singleton. “Transform analysis and asset pricing for affine jump-diffusions”. In: *Econometrica* 68.6 (2000), pp. 1343–1376. ISSN: 0012-9682. DOI: [10.1111/1468-0262.00164](https://doi.org/10.1111/1468-0262.00164).
- [32] Nick G. Duffield and Neil O’Connell. “Large deviations and overflow probabilities for the general single-server queue, with applications”. In: *Math. Proc. Cambridge Philos. Soc.* 118.2 (1995), pp. 363–374. ISSN: 0305-0041. DOI: [10.1017/S0305004100073709](https://doi.org/10.1017/S0305004100073709).
- [33] Michael Eichler, Rainer Dahlhaus, and Johannes Dueck. “Graphical modeling for multivariate Hawkes processes with nonparametric link functions”. In: *J. Time Series Anal.* 38.2 (2017), pp. 225–242. ISSN: 0143-9782. DOI: [10.1111/jtsa.12213](https://doi.org/10.1111/jtsa.12213).
- [34] Paul Embrechts, Thomas Liniger, and Lu Lin. “Multivariate Hawkes processes: an application to financial data”. In: *J. Appl. Probab.* 48A. New frontiers in applied probability: a Festschrift for Søren Asmussen (2011), pp. 367–378. ISSN: 0021-9002. DOI: [10.1239/jap/1318940477](https://doi.org/10.1239/jap/1318940477).

- [35] Eymen Errais, Kay Giesecke, and Lisa R. Goldberg. “Affine point processes and portfolio credit risk”. In: *SIAM J. Financial Math.* 1.1 (2010), pp. 642–665. DOI: [10.1137/090771272](https://doi.org/10.1137/090771272).
- [36] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes: Characterization and convergence*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986, pp. x+534. ISBN: 0-471-08186-8. DOI: [10.1002/9780470316658](https://doi.org/10.1002/9780470316658).
- [37] Eric Fournié, Jean-Michel Lasry, Jérôme Lebuchoux, Pierre-Louis Lions, and Nizar Touzi. “Applications of Malliavin calculus to Monte Carlo methods in finance”. In: *Finance Stoch.* 3.4 (1999), pp. 391–412. ISSN: 0949-2984. DOI: [10.1007/s007800050068](https://doi.org/10.1007/s007800050068).
- [38] Noufel Frikha, Arturo Kohatsu-Higa, and Libo Li. “Integration by parts formula for killed processes: a point of view from approximation theory”. In: *Electron. J. Probab.* 24 (2019), Paper No. 95, 44. DOI: [10.1214/19-ejp352](https://doi.org/10.1214/19-ejp352).
- [39] Noufel Frikha and Libo Li. “Well-posedness and approximation of some one-dimensional Lévy-driven non-linear SDEs”. In: *Stochastic Process. Appl.* 132 (2021), pp. 76–107. ISSN: 0304-4149. DOI: [10.1016/j.spa.2020.10.002](https://doi.org/10.1016/j.spa.2020.10.002).
- [40] Antonio Galbis and Enrique Jordá. “Composition operators on the Schwartz space”. In: *Rev. Mat. Iberoam.* 34.1 (2018), pp. 397–412. ISSN: 0213-2230. DOI: [10.4171/RMI/989](https://doi.org/10.4171/RMI/989).
- [41] Xuefeng Gao and Lingjiong Zhu. “Functional central limit theorems for stationary Hawkes processes and application to infinite-server queues”. In: *Queueing Syst.* 90.1-2 (2018), pp. 161–206. ISSN: 0257-0130. DOI: [10.1007/s11134-018-9570-5](https://doi.org/10.1007/s11134-018-9570-5).
- [42] Kay Giesecke and Baeho Kim. “Estimating tranche spreads by loss process simulation”. In: *2007 Winter Simulation Conference*. 2007, pp. 967–975. DOI: [10.1109/WSC.2007.4419693](https://doi.org/10.1109/WSC.2007.4419693).
- [43] Peter W. Glynn and Ward Whitt. “Logarithmic asymptotics for steady-state tail probabilities in a single-server queue”. In: vol. 31A. *Studies in applied probability*. 1994, pp. 131–156. DOI: [10.2307/3214953](https://doi.org/10.2307/3214953).
- [44] Jan Grandell. *Aspects of risk theory*. Springer Series in Statistics: Probability and its Applications. Springer-Verlag, New York, 1991, pp. x+175. ISBN: 0-387-97368-0. DOI: [10.1007/978-1-4613-9058-9](https://doi.org/10.1007/978-1-4613-9058-9).
- [45] Hélène Halconruy. “Malliavin calculus and Dirichlet structures for independent random variables”. Theses. Institut Polytechnique de Paris, 2020.
- [46] Alan G. Hawkes. “Spectra of some self-exciting and mutually exciting point processes”. In: *Biometrika* 58 (1971), pp. 83–90. ISSN: 0006-3444. DOI: [10.1093/biomet/58.1.83](https://doi.org/10.1093/biomet/58.1.83).
- [47] Alan G. Hawkes and David Oakes. “A cluster process representation of a self-exciting process”. In: *J. Appl. Probability* 11 (1974), pp. 493–503. ISSN: 0021-9002. DOI: [10.2307/3212693](https://doi.org/10.2307/3212693).

- [48] Caroline Hillairet, Anthony Réveillac, and Mathieu Rosenbaum. *An expansion formula for Hawkes processes and application to cyber-insurance derivatives*. Preprint. 2020.
- [49] Ulrich Horst and Wei Xu. “A scaling limit for limit order books driven by Hawkes processes”. In: *SIAM J. Financial Math.* 10.2 (2019), pp. 350–393. DOI: [10.1137/17M1148682](https://doi.org/10.1137/17M1148682).
- [50] Ulrich Horst and Wei Xu. “Functional limit theorems for marked Hawkes point measures”. In: *Stochastic Process. Appl.* 134 (2021), pp. 94–131. ISSN: 0304-4149. DOI: [10.1016/j.spa.2020.12.002](https://doi.org/10.1016/j.spa.2020.12.002).
- [51] Jean Jacod and Albert N. Shiryaev. *Limit theorems for stochastic processes*. Second. Vol. 288. Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 2003, pp. xx+661. ISBN: 3-540-43932-3. DOI: [10.1007/978-3-662-05265-5](https://doi.org/10.1007/978-3-662-05265-5).
- [52] Thibault Jaisson and Mathieu Rosenbaum. “Limit theorems for nearly unstable Hawkes processes”. In: *Ann. Appl. Probab.* 25.2 (2015), pp. 600–631. ISSN: 1050-5164. DOI: [10.1214/14-AAP1005](https://doi.org/10.1214/14-AAP1005).
- [53] David W. Kammler. “Approximation with sums of exponentials in  $L_p[0, \infty)$ ”. In: *J. Approximation Theory* 16.4 (1976), pp. 384–408. ISSN: 0021-9045. DOI: [10.1016/0021-9045\(76\)90071-x](https://doi.org/10.1016/0021-9045(76)90071-x).
- [54] Dmytro Karabash and Lingjiong Zhu. “Limit theorems for marked Hawkes processes with application to a risk model”. In: *Stoch. Models* 31.3 (2015), pp. 433–451. ISSN: 1532-6349. DOI: [10.1080/15326349.2015.1024868](https://doi.org/10.1080/15326349.2015.1024868).
- [55] Matthias Kirchner. “An estimation procedure for the Hawkes process”. In: *Quant. Finance* 17.4 (2017), pp. 571–595. ISSN: 1469-7688. DOI: [10.1080/14697688.2016.1211312](https://doi.org/10.1080/14697688.2016.1211312).
- [56] Matthias Kirchner. “Hawkes and INAR( $\infty$ ) processes”. In: *Stochastic Process. Appl.* 126.8 (2016), pp. 2494–2525. ISSN: 0304-4149. DOI: [10.1016/j.spa.2016.02.008](https://doi.org/10.1016/j.spa.2016.02.008).
- [57] Lawrence A. Klimko and Paul I. Nelson. “On conditional least squares estimation for stochastic processes”. In: *Ann. Statist.* 6.3 (1978), pp. 629–642. ISSN: 0090-5364.
- [58] Kyungsub Lee and Byoung Ki Seo. “Marked Hawkes process modeling of price dynamics and volatility estimation”. In: *Journal of Empirical Finance* 40 (2017), pp. 174–200. ISSN: 0927-5398. DOI: <https://doi.org/10.1016/j.jempfin.2016.08.004>.
- [59] Remi Lemonnier and Nicolas Vayatis. “Nonparametric markovian learning of triggering kernels for mutually exciting and mutually inhibiting multivariate hawkes processes”. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer. 2014, pp. 161–176.

- [60] Peter A. W. Lewis and Gerald S. Shedler. “Simulation of nonhomogeneous Poisson processes by thinning”. In: *Naval Res. Logist. Quart.* 26.3 (1979), pp. 403–413. ISSN: 0028-1441. DOI: [10.1002/nav.3800260304](https://doi.org/10.1002/nav.3800260304).
- [61] Peter A. W. Lewis and Gerald S. Shedler. “Simulation of nonhomogeneous Poisson processes with log linear rate function”. In: *Biometrika* 63.3 (1976), pp. 501–505. ISSN: 0006-3444. DOI: [10.1093/biomet/63.3.501](https://doi.org/10.1093/biomet/63.3.501).
- [62] Thomas Josef Verfasser Liniger. *Multivariate Hawkes processes*. Zürich, 2009.
- [63] Eva Löcherbach. “Spiking neurons: interacting Hawkes processes, mean field limits and oscillations”. In: *Journées MAS 2016 de la SMAI—Phénomènes complexes et hétérogènes*. Vol. 60. ESAIM Proc. Surveys. EDP Sci., Les Ulis, 2017, pp. 90–103. DOI: [10.1051/proc/201760090](https://doi.org/10.1051/proc/201760090).
- [64] Rüdiger Murr. “Characterization of infinite divisibility by duality formulas. Application to Lévy processes and random measures”. In: *Stochastic Process. Appl.* 123.5 (2013), pp. 1729–1749. ISSN: 0304-4149. DOI: [10.1016/j.spa.2012.12.012](https://doi.org/10.1016/j.spa.2012.12.012).
- [65] Ivan Nourdin and Giovanni Peccati. *Normal approximations with Malliavin calculus*. Vol. 192. Cambridge Tracts in Mathematics. From Stein’s method to universality. Cambridge University Press, Cambridge, 2012, pp. xiv+239. ISBN: 978-1-107-01777-1. DOI: [10.1017/CB09781139084659](https://doi.org/10.1017/CB09781139084659).
- [66] Ivan Nourdin and Giovanni Peccati. “Stein’s method on Wiener chaos”. In: *Probab. Theory Related Fields* 145.1-2 (2009), pp. 75–118. ISSN: 0178-8051. DOI: [10.1007/s00440-008-0162-x](https://doi.org/10.1007/s00440-008-0162-x).
- [67] Ivan Nourdin, Giovanni Peccati, and Anthony Réveillac. “Multivariate normal approximation using Stein’s method and Malliavin calculus”. In: *Ann. Inst. Henri Poincaré Probab. Stat.* 46.1 (2010), pp. 45–58. ISSN: 0246-0203. DOI: [10.1214/08-AIHP308](https://doi.org/10.1214/08-AIHP308).
- [68] David Nualart and Josep Vives. “Anticipative calculus for the Poisson process based on the Fock space”. In: *Séminaire de Probabilités, XXIV, 1988/89*. Vol. 1426. Lecture Notes in Math. Springer, Berlin, 1990, pp. 154–165.
- [69] Yosihiko Ogata. “On Lewis’ simulation method for point processes”. In: *IEEE Transactions on Information Theory* 27.1 (1981), pp. 23–31. DOI: [10.1109/TIT.1981.1056305](https://doi.org/10.1109/TIT.1981.1056305).
- [70] Yosihiko Ogata. “Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes”. In: *Journal of the American Statistical Association* 83.401 (1988), pp. 9–27. ISSN: 01621459.
- [71] Yosihiko Ogata. “Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes”. In: *Journal of the American Statistical Association* 83.401 (1988), pp. 9–27. DOI: [10.1080/01621459.1988.10478560](https://doi.org/10.1080/01621459.1988.10478560). eprint: <https://www.tandfonline.com/doi/pdf/10.1080/01621459.1988.10478560>.

- [72] Tohru Ozaki. “Maximum likelihood estimation of Hawkes’ self-exciting point processes”. In: *Ann. Inst. Statist. Math.* 31.1 (1979), pp. 145–155. ISSN: 0020-3157. DOI: [10.1007/BF02480272](https://doi.org/10.1007/BF02480272).
- [73] Giovanni Peccati, Josep L. Solé, Murad S. Taqqu, and Frederic Utzet. “Stein’s method and normal approximation of Poisson functionals”. In: *Ann. Probab.* 38.2 (2010), pp. 443–478. ISSN: 0091-1798. DOI: [10.1214/09-AOP477](https://doi.org/10.1214/09-AOP477).
- [74] Giovanni Peccati and Cengbo Zheng. “Multi-dimensional Gaussian fluctuations on the Poisson space”. In: *Electron. J. Probab.* 15 (2010), no. 48, 1487–1527. DOI: [10.1214/EJP.v15-813](https://doi.org/10.1214/EJP.v15-813).
- [75] Jean Picard. “Formules de dualité sur l’espace de Poisson”. In: *Ann. Inst. H. Poincaré Probab. Statist.* 32.4 (1996), pp. 509–548. ISSN: 0246-0203.
- [76] Nicolas Privault. “Stein normal approximation for multidimensional Poisson random measures by third cumulant expansions”. In: *ALEA Lat. Am. J. Probab. Math. Stat.* 15.2 (2018), pp. 1141–1161. DOI: [10.30757/alea.v15-42](https://doi.org/10.30757/alea.v15-42).
- [77] Nicolas Privault. *Stochastic analysis in discrete and continuous settings with normal martingales*. Vol. 1982. Lecture Notes in Mathematics. Springer-Verlag, Berlin, 2009, pp. x+310. ISBN: 978-3-642-02379-8. DOI: [10.1007/978-3-642-02380-4](https://doi.org/10.1007/978-3-642-02380-4).
- [78] Mads Bonde Raad, Susanne Ditlevsen, and Eva Löcherbach. “Stability and mean-field limits of age dependent Hawkes processes”. In: *Ann. Inst. Henri Poincaré Probab. Stat.* 56.3 (2020), pp. 1958–1990. ISSN: 0246-0203. DOI: [10.1214/19-AIHP1023](https://doi.org/10.1214/19-AIHP1023).
- [79] Mads Bonde Raad, Susanne Ditlevsen, and Eva Löcherbach. “Stability and mean-field limits of age dependent Hawkes processes”. In: *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques* 56.3 (2020), pp. 1958–1990. DOI: [10.1214/19-AIHP1023](https://doi.org/10.1214/19-AIHP1023).
- [80] Patricia Reynaud-Bouret, Vincent Rivoirard, Franck Grammont, and Christine Tuleau-Malot. “Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis”. In: *J. Math. Neurosci.* 4 (2014), Art. 3, 41. DOI: [10.1186/2190-8567-4-3](https://doi.org/10.1186/2190-8567-4-3).
- [81] Patricia Reynaud-Bouret, Vincent Rivoirard, and Christine Tuleau-Malot. “Inference of functional connectivity in Neurosciences via Hawkes processes”. In: *2013 IEEE Global Conference on Signal and Information Processing*. 2013, pp. 317–320. DOI: [10.1109/GlobalSIP.2013.6736879](https://doi.org/10.1109/GlobalSIP.2013.6736879).
- [82] Patricia Reynaud-Bouret and Sophie Schbath. “Adaptive estimation for Hawkes processes; application to genome analysis”. In: *Ann. Statist.* 38.5 (2010), pp. 2781–2822. ISSN: 0090-5364. DOI: [10.1214/10-AOS806](https://doi.org/10.1214/10-AOS806).
- [83] Youngsoo Seol. “Limit theorems for discrete Hawkes processes”. In: *Statist. Probab. Lett.* 99 (2015), pp. 223–229. ISSN: 0167-7152. DOI: [10.1016/j.spl.2015.01.023](https://doi.org/10.1016/j.spl.2015.01.023).



- [84] Leigh Shlomovich, Edward AK Cohen, and Niall Adams. “A Parameter Estimation Method for Multivariate Aggregated Hawkes Processes”. In: *arXiv preprint arXiv:2108.12357* (2021).
- [85] Josep Lluís Solé, Frederic Utzet, and Josep Vives. “Canonical Lévy process and Malliavin calculus”. In: *Stochastic Process. Appl.* 117.2 (2007), pp. 165–187. ISSN: 0304-4149. DOI: [10.1016/j.spa.2006.06.006](https://doi.org/10.1016/j.spa.2006.06.006).
- [86] Gabriele Stabile and Giovanni Luca Torrisi. “Risk processes with non-stationary Hawkes claims arrivals”. In: *Methodol. Comput. Appl. Probab.* 12.3 (2010), pp. 415–429. ISSN: 1387-5841. DOI: [10.1007/s11009-008-9110-6](https://doi.org/10.1007/s11009-008-9110-6).
- [87] Charles Stein. “A bound for the error in the normal approximation to the distribution of a sum of dependent random variables”. In: *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory*. 1972, pp. 583–602.
- [88] Michael JW Stokesbury, Steven LH Teo, Andrew Seitz, Ronald K O’Dor, and Barbara A Block. “Movement of Atlantic bluefin tuna (*Thunnus thynnus*) as determined by satellite tagging experiments initiated off New England”. In: *Canadian Journal of Fisheries and Aquatic Sciences* 61.10 (2004), pp. 1976–1987. DOI: [10.1139/f04-130](https://doi.org/10.1139/f04-130). eprint: <https://doi.org/10.1139/f04-130>.
- [89] Deborah Sulem, Vincent Rivoirard, and Judith Rousseau. “Bayesian estimation of nonlinear Hawkes process”. In: *arXiv preprint arXiv:2103.17164* (2021).
- [90] Anatoliy Swishchuk, Rudi Zagst, and Gabriela Zeller. “Hawkes processes in insurance: risk model, application to empirical data and optimal investment”. In: *Insurance Math. Econom.* 101 (2021), pp. 107–124. ISSN: 0167-6687. DOI: [10.1016/j.insmatheco.2020.12.005](https://doi.org/10.1016/j.insmatheco.2020.12.005).
- [91] Giovanni Luca Torrisi. “Gaussian approximation of nonlinear Hawkes processes”. In: *Ann. Appl. Probab.* 26.4 (2016), pp. 2106–2140. ISSN: 1050-5164. DOI: [10.1214/15-AAP1141](https://doi.org/10.1214/15-AAP1141).
- [92] Lingjiong Zhu. “Central limit theorem for nonlinear Hawkes processes”. In: *J. Appl. Probab.* 50.3 (2013), pp. 760–771. ISSN: 0021-9002. DOI: [10.1239/jap/1378401234](https://doi.org/10.1239/jap/1378401234).
- [93] Lingjiong Zhu. “Large deviations for Markovian nonlinear Hawkes processes”. In: *Ann. Appl. Probab.* 25.2 (2015), pp. 548–581. ISSN: 1050-5164. DOI: [10.1214/14-AAP1003](https://doi.org/10.1214/14-AAP1003).
- [94] Lingjiong Zhu. “Limit theorems for a Cox-Ingersoll-Ross process with Hawkes jumps”. In: *J. Appl. Probab.* 51.3 (2014), pp. 699–712. ISSN: 0021-9002. DOI: [10.1239/jap/1409932668](https://doi.org/10.1239/jap/1409932668).
- [95] Lingjiong Zhu. “Ruin probabilities for risk processes with non-stationary arrivals and subexponential claims”. In: *Insurance Math. Econom.* 53.3 (2013), pp. 544–550. ISSN: 0167-6687. DOI: [10.1016/j.insmatheco.2013.08.008](https://doi.org/10.1016/j.insmatheco.2013.08.008).

## Sur l'approximation des processus de Hawkes: des séries temporelles aux théorèmes centraux limites quantitatifs

**Résumé:** Dans cette thèse, nous étudions une classe de processus ponctuels auto-régressifs, celle des processus de Hawkes. Ces processus ont des applications dans plusieurs domaines comme la neuroscience, les assurances et les réseaux sociaux. Nous nous concentrons sur l'approximation de ces processus sur la droite réelle ou bien en temps long.

Le chapitre 2 porte sur l'approximation des processus de Hawkes non linéaires, éventuellement avec une intensité perturbée par un bruit gaussien. Le processus non linéaire a l'avantage de permettre de modéliser l'auto-inhibition. Comme les processus dont les noyaux sont des fonctions d'Erlang ont une structure markovienne, on propose une chaîne de Markov d'approximation. On montre la convergence de cette chaîne vers le processus en temps continu quand le pas de temps tend vers zéro, étendant ainsi l'approximation de Kirchner (2016) qui est valable uniquement pour les processus linéaires et dont la démonstration nécessite des techniques différentes. En guise d'application, on calibre numériquement le processus en temps continu en ayant recours à une régression sur son approximation discrète.

Le chapitre 3, qui est une collaboration avec Caroline Hillairet, Lorick Huang et Anthony Réveillac, est dédié à l'application de la méthode de Malliavin-Stein aux processus de Hawkes linéaires. On y démontre des bornes sur la distance de Wasserstein entre les fonctionnelles de Hawkes et une gaussienne quelconque. En considérant les processus de Hawkes comme un amincissement d'une mesure de Poisson générale, on développe un calcul de Malliavin spécifique à ces processus. Ce résultat est appliqué ensuite pour obtenir, pour la première fois, la vitesse de convergence du processus de Hawkes normalisé vers sa limite gaussienne pour une classe spéciale de noyaux. Au chapitre 4, on étend les opérateurs de Malliavin définis dans le chapitre 3 aux dimensions supérieures, afin de travailler avec les processus de Hawkes linéaires multivariés. En mesurant la distance entre le processus de Hawkes et sa limite gaussienne multidimensionnelle avec une distance de Wasserstein adéquate, on généralise la borne obtenue au chapitre précédent. Par suite, on l'étend aux processus composés évalués à différentes marginales de temps.

Le chapitre 5, plus applicatif, généralise le travail de Karabash et Zhu (2015) à une police d'assurances dont les sinistres arrivent selon un processus de Hawkes linéaire marqué. On y démontre un théorème central limite quantitatif, ainsi qu'un principe de grandes déviations (PGD) en donnant des preuves détaillées. Le PGD est utilisé par la suite pour quantifier la probabilité de ruine asymptotique de la police. Finalement, on illustre la probabilité de ruine sur un exemple numérique et on la compare au modèle classique de Cramér-Lundberg.

## On the approximation of Hawkes processes: From time series to quantitative central limit theorems

**Abstract:** In this thesis we study Hawkes processes, which are a class of auto-regressive point processes that have applications in many fields such as neuroscience, insurance and social media. We focus on the approximation of these processes, either on the whole time axis or when the time horizon goes to infinity. In Chapter 2 we propose a time series model to approximate nonlinear Hawkes processes on a time grid. Despite being less explicit than its linear counterpart, the nonlinear Hawkes process allows for auto-inhibition and for the Gaussian perturbation of its intensity. By using the Markov structure of Erlang kernels, we suggest a hidden Markov chain that approximates the Hawkes process with the aforementioned kernels. We then show that the approximation actually converges in the Skorokhod topology to the time-continuous process. This result is an extension of the work of Kirchner (2016) that is only available for linear processes and whose proof uses different techniques. Finally, we numerically calibrate a realisation of the multivariate Hawkes process by running a regression on its discrete-time approximation.

---

Chapter 3 is based on a joint work with Caroline Hillairet, Lorick Huang and Anthony Réveillac. It is dedicated to the application of the Malliavin-Stein method to the linear Hawkes process. In this Chapter, we provide general bounds on the Wasserstein distance between the law of functionals of a Hawkes process and the one of a Gaussian random variable. This is achieved by relying on the Poisson embedding the Hawkes process, which allows us to develop a specific Malliavin calculus. We apply the general bound to centered and normalized versions to obtain, for the first time, a speed of convergence of the Hawkes process towards its Gaussian limit. In Chapter 4 the operators are extended to higher dimensions in order to adapt the results to linear compound multivariate Hawkes process. By choosing an adequate Wasserstein distance between the Hawkes process and its multivariate Gaussian limit, we extend the bound established in Chapter 3. The result is then extended to vectors formed by the compound process, evaluated at different time marginals.

The last chapter follows closely the paper of Karabash and Zhu (2015) by dealing with an insurance policy driven by a marked linear Hawkes process. We provide a quantitative central limit theorem as well as a large deviations principle (LDP) with their detailed proofs. The LDP is then used to quantify the asymptotic ruin probability of the insurance policy. We finally illustrate the result on a numerical example which we then compare to the classical Cramér-Lundberg model.