The Malliavin-Stein method for Hawkes functionals

Caroline Hillairet, Lorick Huang, Mahmoud Khabou, Anthony Réveillac

HAL Id: hal-03189614

https://hal.insa-toulouse.fr/hal-03189614

The Malliavin-Stein method for Hawkes functionals

Caroline Hillairet[*]   Lorick Huang[†]   Mahmoud Khabou[‡]   Anthony Réveillac[§]

April 4, 2021

Abstract

In this paper, following Nourdin-Peccati's methodology, we combine the Malliavin calculus and Stein's method to provide general bounds on the Wasserstein distance between functionals of a compound Hawkes process and a given Gaussian density. To achieve this, we rely on the Poisson embedding representation of an 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 the first Berry-Esséen bounds associated to Central Limit Theorems for the compound Hawkes process.

Keywords: Hawkes process; Malliavin's calculus; Stein's method; Limit Theorems; Berry-Esséen bounds

1   Introduction

Hawkes processes are an efficient generalisation of the Poisson processes 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 in a more accurate way, compared to inhomogeneous Poisson processes or Cox processes, self-exciting phenomena. Introduced by Hawkes in [12], 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, social network models, biology and population dynamics, finance or insurance. In finance, they are accurate to model for example credit risk contagion [10], order book or the microstructure noise's feature of financial markets [2]. In insurance also, some risks exhibit self-exciting features, as it is the case for cyber risk [4].

Naturally in all these applications, long term behaviour of the Hawkes process $H := (H_t)_{t \geq 0}$ (or of functionals of H) are 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 [5] for the first work in that direction for linear Hawkes processes with many extensions in the non-linear case).

The first main analysis of (central) limit theorems for Hawkes processes has been derived in [1]. For the sake of the presentation, we recall that a Hawkes process is a counting process $H := (H_t)_{t \geq 0}$ with stochastic intensity $\lambda := (\lambda_t)_{t \geq 0}$ given by

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

where $\mu \geq 0$ and $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ are two given parameters. Under an integrability assumption (which reads as Assumption 3.2 below), according to[1] [1, Theorem 2],

$$\left(\frac{H_{Tv} - \int_0^{Tv} \mathbb{E}[\lambda_s] ds}{\sqrt{T}} \right)_{v \in [0,1]} \xrightarrow[T \to +\infty]{\mathcal{L}-\mathcal{S}} (\tilde{\sigma} W_v)_{v \in [0,1]}, \tag{1.1}$$

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 μ and Φ and W a Brownian motion. This result is a consequence of the martingale limit theorem that can be found as [1, Lemma 7] (once again we present a particular case of this result)

$$\left(\frac{H_{Tv} - \int_0^{Tv} \lambda_s ds}{\sqrt{T}} \right)_{v \in [0,1]} \xrightarrow[T \to +\infty]{\mathcal{L}-\mathcal{S}} (\sigma W_v)_{v \in [0,1]}, \tag{1.2}$$

with $\sigma > 0$ an explicit constant depending only on μ and Φ (but still different from $\tilde{\sigma}$ in (1.1)). Naturally, limit theorems (1.1) and (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 [1].

Generalisations of limit theorems (1.1) and (1.2) have been obtained in [11], [30], [29], [16], [15], [14] 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 in a form of a Berry-Esséen bound [2] is not available in the literature. To be more specific, we are not aware of a Berry-Esséen bound for the 1-marginal convergence in law (by taking $v = 1$ in (1.2)):

$$\frac{H_T - \int_0^T \lambda_s ds}{\sqrt{T}} \xrightarrow[T \to +\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2). \tag{1.3}$$

Several authors (see [25], [3] or [28]) have used the Nourdin-Peccati methodology to provide Berry-Esséen-type bounds for the normal approximation of counting processes but no result

---

[1]Note that we consider a particular case of the results in [1] which hold for multidimensional Hawkes processes.

[2]Actually, Berry-Esséen bounds measure the CLT's rate of convergence by bounding the distance between cumulative distribution functions (Kolmogorov-Smirnov metric). Here we keep referring to the term Berry-Esséen for the quantification of the convergence in the Wasserstein distance.

in the literature allows one to quantify convergence (1.3). The closest result is the one of [28]; we refer to Remark 3.8 for a discussion on this matter. In this paper, we consider a compound Hawkes process $X := (X_t)_{t \geq 0}$, of the form

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

with $(Y_i)_{i \geq 1}$ a sequence of iid random variables with distribution ν (and independent of H) and we provide quantitative limit theorems for X in the spirit of (1.2)-(1.3). To this end, we adapt the (by now classical) Malliavin-Stein approach (initiated in [18] 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 3.4 and especially Relation (3.3)). More precisely, we make use of the Poisson embedding representation of the Hawkes process (initially introduced in [6]) which allows one to write the Hawkes process H together with its intensity process λ and the compound process X as the unique solution of an SDE driven by a Poisson measure N on $\mathbb{R}_+^2 \times \mathbb{R}$ (see Theorem 2.12 below) to derive a Malliavin calculus with respect to the Hawkes process. This completes the analysis initiated for a general Hawkes process in [13], 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 N) for which we adapt the Malliavin-Stein method of [18] (or more precisely in this Poisson context of [22]) 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 (1.3) for the Hawkes process H can be quantified in the Wasserstein distance denoted by d_W (see Theorem 3.10 and Theorem 3.12). For instance, in the famous cases of exponential and Erlang's kernels[3]

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

where $C_{\Phi,\nu}$ and σ^2, ϑ^2 are explicit constants and $m := \int_{\mathbb{R}} x\nu(dx)$ (see Theorem 3.12 for a precise statement). As Theorem 3.10 suggests, we were not able to prove the general $T^{-1/2}$ rate for any kernel Φ. 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 paper. We refer to Section 3.2.1 for a review of our results on quantitative counterparts of limit theorems (1.2)-(1.3) regarding the (compound) Hawkes process.

The paper is organised as follows. First we derive in Section 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 our companion paper [13]). The main element of the Malliavin-Stein method is also presented in this section. General bounds for Hawkes functionals are presented in Section 3.1 and the Berry-Esséen bounds associated to

---

[3]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 α, β are well-chosen positive constants.

generalisation for the Central Limit Theorem 1.2 are presented in Section 3.2, starting with a review of our quantitative limit theorems for the compound Hawkes process in Section 3.2.1. Technical lemmata including some results on exponential and Erlang kernels are contained in the Appendix (Section 4).

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 2.1. Then with this material at hand, we make precise in Section 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 2.3 the Malliavin derivative of the compound Hawkes process and give an integration by parts formula in Theorem 2.20 which completes the Mecke formula obtained in [13]. Finally, we recall the main elements of Stein's method in Section 2.4.

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

2.1 Elements of stochastic analysis on the Poisson space

Let ν be a Borelian 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 [13]. To this end, we need three variables for the Poisson measure : t for the jump times, x will stand for the size the jump (whose distribution ν) and θ will play the role of an auxiliary variable required for the representation itself (according to the thinning algorithm).

Let the space of configurations

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

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

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

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

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

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

Definition 2.1 (Shift operator). We define for (t, θ, x) in $\mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}$ the measurable maps

$$\varepsilon^+_{(t,\theta,x)} : \quad \Omega^N \quad \to \Omega^N$$
$$\omega \quad \mapsto \varepsilon^+_{(t,\theta,x)}(\omega),$$

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

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

Lemma 2.2. *Let $t \geq 0$ and F be an $\mathcal{F}^N_t$-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 2.3 (Malliavin derivative). For F in $L^2(\Omega, \mathcal{F}^N_\infty, \mathbb{P})$, 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 2.4. Our approach follows [23, 24]. In our setting, the measure ν 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 [26]. 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 ν (we refer to [17] or [24, Section 6.7] for more details) as our baseline process is the Poisson process N and not the compound process $\int xN(dt, d\theta, dx)$.

The following definition is a by-product of [23, Théorème 1] (see also [20]).

Definition 2.5. Let $\mathcal{I}$ be the sub-sigma field of $\mathcal{B}(\mathbb{R}^2_+ \times \mathbb{R}) \otimes \mathcal{F}^N$ of stochastic processes $Z := (Z_{(t,\theta,x)})_{(t,\theta,x) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}}$ 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 2.6. Let (t_0, θ_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 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 :

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

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

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

$$\delta^N(Z) := \int_{\mathbb{R}_+^2 \times \mathbb{R}} Z_{(t,\theta,x)} N(dt,d\theta,dx) - \int_{\mathbb{R}_+^2 \times \mathbb{R}} Z_{(t,\theta,x)} dt d\theta \nu(dx). \tag{2.1}$$

We conclude this section with the integration by parts formula on the Poisson space (see [23, Remarque 1])

Proposition 2.8 (See *e.g.* [23]). *Let F be in $L^2(\Omega, \mathcal{F}_\infty^N, \mathbb{P})$ and Z be in $\mathcal{S}$. We have that*

$$\mathbb{E}\left[F \delta^N(Z)\right] = \mathbb{E}\left[\int_{\mathbb{R}_+^2 \times \mathbb{R}} Z_{(t,\theta,x)} D_{(t,\theta,x)} F \, dt d\theta \nu(dx)\right]. \tag{2.2}$$

2.2 Representation of the compound Hawkes process

We first recall the definition of a Hawkes process.

Definition 2.9 (Standard Hawkes process, [12]). Let $\mu > 0$ and $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ be a bounded non-negative map with $\|\Phi\|_1 := \int_0^{+\infty} \Phi(u) du < 1$. A standard Hawkes process $H := (H_t)_{t \geq 0}$ with parameters μ and Φ is a counting process such that

(i) $H_0 = 0, \quad \mathbb{P} - a.s.,$

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

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

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

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

This definition can be generalized as follows.

Definition 2.10 (Generalized Hawkes process). Let $v \geq 0$, h^v be a $\mathcal{F}_v^N$-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 $\mathcal{F}_v^N$-measurable for any $t \geq v$, and $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ be a bounded non-negative map with $\|\Phi\|_1 < 1$. A Hawkes process on $[v, +\infty)$ with parameters μ^v, h^v and $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ is a ($\mathbb{F}^N$-adapted) counting process $H := (H_t)_{t \geq v}$ such that

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

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

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

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

$$\mathbb{E}\left[\mathbf{1}_A (H_t - H_s) | \mathcal{F}_v\right] = \mathbb{E}\left[\int_{(s,t]} \mathbf{1}_A \lambda_r dr \Big| \mathcal{F}_v\right].$$

Definition 2.11 (Compound Hawkes process). Let $\mu > 0$, $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ be a bounded non-negative map with $\|\Phi\|_1 < 1$ and $(Y_i)_{i \geq 1}$ be iid random variables, with common distribution ν, and independent of the Hawkes process H with parameters μ and Φ. We name compound Hawkes process $X := (X_t)_{t \geq 0}$ a stochastic process with representation :

$$X_t = \sum_{i=1}^{H_t} Y_i, \quad t \geq 0. \tag{2.3}$$

We now represent a Hawkes process (and a compound Hawkes process) as the unique solution to an SDE driven by N. This representation relies on the "Poisson embedding" (or "Thinning Algorithm" (see *e.g.* [6, 7, 8, 21] and references therein). The next result is an extension of [13, Theorem 3.3].

Theorem 2.12 (See [13]). *Let $\mu > 0$ and $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\|\Phi\|_1 < 1$. The SDE below admits a unique[4] solution (X, H, λ) with H (resp. λ) $\mathbb{F}^N$-adapted (resp. $\mathbb{F}^N$-predictable)*

$$\begin{cases} X_t = \displaystyle\int_{(0,t] \times \mathbb{R}_+ \times \mathbb{R}} x \mathbf{1}_{\{\theta \leq \lambda_s\}} N(ds, d\theta, dx), \quad t \geq 0, \\[3mm] H_t = \displaystyle\int_{(0,t] \times \mathbb{R}_+ \times \mathbb{R}} \mathbf{1}_{\{\theta \leq \lambda_s\}} N(ds, d\theta, dx), \quad t \geq 0, \\[3mm] \lambda_t = \mu + \int_{(0,t)} \Phi(t - u) dH_u, \quad t \geq 0. \end{cases} \tag{2.4}$$

In addition, (H, λ) is a Hawkes process in the sense of Definition 2.9. We set $\mathbb{F}^H := (\mathcal{F}_t^H)_{t \geq 0}$ (respectively $\mathbb{F}^X := (\mathcal{F}_t^X)_{t \geq 0}$) the natural filtration of H (respectively of X) and $\mathcal{F}_\infty^H := \lim_{t \to +\infty} \mathcal{F}_t^H$ (respectively $\mathcal{F}_\infty^X := \lim_{t \to +\infty} \mathcal{F}_t^X$). Obviously $\mathcal{F}_t^H \subset \mathcal{F}_t^X \subset \mathcal{F}_t^N$ as H is completely determined by the jump times of H which are exactly those of X.
Finally, X is a compound Hawkes process in the sense of Definition 2.11.

Remark 2.13. The definition above is based on the following notation. Indeed, the rigorous definition of H is given as :

$$H(\mathcal{T}) := \int_{\mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}} \mathbf{1}_{\{s \in \mathcal{T}\}} \mathbf{1}_{\{\theta \leq \lambda_s\}} N(ds, d\theta, dx), \quad \mathcal{T} \in \mathcal{B}(\mathbb{R}_+),$$

$$X(\mathcal{T}) := \int_{\mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}} x \mathbf{1}_{\{s \in \mathcal{T}\}} \mathbf{1}_{\{\theta \leq \lambda_s\}} N(ds, d\theta, dx), \quad \mathcal{T} \in \mathcal{B}(\mathbb{R}_+).$$

Then by convention we set $H_t := H([0, t])$, $X_t := X([0, t])$ and $H_{t-} := H([0, t))$, $X_{t-} := X([0, t))$ for any $t > 0$, with $H_0 := X_0 := 0$.

Remark 2.14. 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 [13] and in that case one can just consider N to be a Poisson measure on $\mathbb{R}_+^2$.

---

[4]we refer to [13, Theorem 3.3] for a precise statement about uniqueness

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 [13]. The objective of this section is to derive an integration by parts formula for the Hawkes functionals as Theorem 2.20.

Lemma 2.15. *Let t and v in $\mathbb{R}_+$, (θ, x) and (θ_0, x_0) in $\mathbb{R}_+ \times \mathbb{R}$, it holds that :*

$$\mathbf{1}_{\{\theta \leq \lambda_t\}} (X_v \circ \varepsilon^+_{(t,\theta,x)}, H_v \circ \varepsilon^+_{(t,\theta,x)}, \lambda_v \circ \varepsilon^+_{(t,\theta,x)})_{v \geq 0}$$
$$= \mathbf{1}_{\{\theta_0 \leq \lambda_t\}} (X_v \circ \varepsilon^+_{(t,\theta_0,x)}, H_v \circ \varepsilon^+_{(t,\theta_0,x_0)}, \lambda_v \circ \varepsilon^+_{(t,\theta_0,x_0)})_{v \geq 0}.$$

Proof. The proof relies on the very definition of the shift 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 H at time t, but the value of this jump for H is equal to 1, and this for any θ such that $\theta \leq \lambda_t$. In other words : for $v \geq t$ on $\{\theta \leq \lambda_t\}$

$$H_v \circ \varepsilon^+_{(t,\theta_0,x_0)} = H_{t-} + \left(\int_{[t,v] \times \mathbb{R}_+ \times \mathbb{R}} \mathbf{1}_{\{\theta \leq \lambda_u\}} N(d\theta, du, dy) \right) \circ \varepsilon^+_{(t,\theta_0,x_0)}$$

$$= H_{t-} + \left(\int_{(t,v] \times \mathbb{R}_+ \times \mathbb{R}} \mathbf{1}_{\{\theta \leq \lambda_u\}} N(d\theta, du, dy) \right) \circ \varepsilon^+_{(t,\theta_0,x_0)}$$

$$= H_{t-} + \mathbf{1}_{\{\theta_0 \leq \lambda_t\}} + \int_{(t,v] \times \times \mathbb{R}} \int_{\mathbb{R}_+} \mathbf{1}_{\{\theta \leq \lambda_u \circ \varepsilon^+_{(t,\theta_0,x_0)}\}} N(d\theta, du, dy),$$

and $H_v \circ \varepsilon^+_{(t,\theta_0,x_0)} = H_v$ for any $v < t$. Note that it is part of the definition of the shift operator (see Definition 2.1) 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 $v \geq t$

$$X_v \circ \varepsilon^+_{(t,\theta_0,x_0)} = X_{t-} + \left(\int_{(t,v] \times \mathbb{R}_+ \times \mathbb{R}} y \mathbf{1}_{\{\theta \leq \lambda_u\}} N(d\theta, du, dy) \right) \circ \varepsilon^+_{(t,\theta_0,x_0)}$$

$$= X_{t-} + x_0 \mathbf{1}_{\{\theta_0 \leq \lambda_t\}} + \int_{(t,v] \times \mathbb{R}} \int_{\mathbb{R}_+} y \mathbf{1}_{\{\theta \leq \lambda_u \circ \varepsilon^+_{(t,\theta_0,x_0)}\}} N(d\theta, du, dy),$$

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 $v > t$,

$$\lambda_v \circ \varepsilon^+_{(t,\theta_0,x_0)} = \left(\mu + \int_{(0,t)} \Phi(v-u) dH_u + \int_{[t,v)} \Phi(v-u) dH_u \right) \circ \varepsilon^+_{(t,\theta_0,x_0)}$$

$$= \mu + \int_{(0,t)} \Phi(v-u) dH_u + \Phi(v-t) + \int_{(t,v)} \Phi(v-u) d(H_u \circ \varepsilon^+_{(t,\theta_0,x_0)}).$$

In other words, $(X \circ \varepsilon^+_{(v,\theta_0,x_0)}, H \circ \varepsilon^+_{(v,\theta_0,x_0)}, \lambda \circ \varepsilon^+_{(v,\theta_0,x_0)})$ solves the same (pathwise and in the SDE sense) equation for any θ_0 such that $\theta_0 \leq \lambda_t$. □

Proposition 2.16. *Let F be a $\mathcal{F}^X_\infty$-measurable random variable. Then for any $t \geq 0$, and $\forall \theta, \theta_0 \geq 0$, $\forall x \in \mathbb{R}$,*

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

Proof. This follows from Lemma 2.15 and the fact that any $\mathcal{F}_T^X$-measurable random variable F is a limit of a random variable of the form $\varphi((X_{t_1}, H_{t_1}, \lambda_{t_1}), \cdots, (X_{t_n}, H_{t_n}, \lambda_{t_n}))$, with : $0 \leq t_1 < \ldots < t_n$, $n \geq 1$ and φ a Borelian map from $\mathbb{R}^{3n}$ to $\mathbb{R}$. More precisely, H 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 H allows one to note for instance that the Malliavin derivative of a functional of H does not depend on the x variable for instance. $\square$

This motivates the introduction of the following notation.

Definition 2.17. Let $t \geq 0$. For a $\mathcal{F}_\infty^X$-measurable random variable F, we set

$$D_{(t,\lambda_t,x)}F := \mathbf{1}_{\{\theta \leq \lambda_t\}} D_{(t,\theta,x)}F, \quad \forall (\theta, x) \in \mathbb{R}_+ \times \mathbb{R}.$$

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 $D_{(t,\lambda_t,x)}F_s = F_s \circ \varepsilon_{(t,\lambda_t,x)}^+ - F_s$ and using Lemma 2.2.

Remark 2.18. Let F be a functional of (H, λ). Then for any $t \geq 0$, and $\forall \theta, \theta_0 \geq 0$, $\forall x, y \in \mathbb{R}$,

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

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

With this notation at hand, we determine the Malliavin derivative of the compound Hawkes process.

Proposition 2.19. *Let $t \geq 0$ and $x \in \mathbb{R}$. We have*

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

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

$$\begin{cases} \hat{H}_s^t = \displaystyle\int_{(t,s] \times \mathbb{R}_+ \times \mathbb{R}} \mathbf{1}_{\{\lambda_u \leq \theta \leq \lambda_u + \hat{\lambda}_u^t\}} N(du, d\theta, dy), & s \geq t, \\ \\ \hat{X}_s^t = \displaystyle\int_{(t,s] \times \mathbb{R}_+ \times \mathbb{R}} y \mathbf{1}_{\{\lambda_u \leq \theta \leq \lambda_u + \hat{\lambda}_u^t\}} N(du, d\theta, dy), & s \geq t, \qquad (2.5) \\ \\ \hat{\lambda}_s^t = \Phi(s - t) + \displaystyle\int_{(t,s)} \Phi(s - u) d\hat{H}_u^t, & s > t, \ \hat{\lambda}_t^t = 0. \end{cases}$$

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

Proof. Fix $t \geq 0$ and $x \in \mathbb{R}$. According to Definition 2.17, $(D_{(t,\lambda_t,x)}X, D_{(t,\lambda_t,x)}H, D_{(t,\lambda_t,x)}\lambda)$ is given as $D_{(t,\lambda_t,x)}X_s = X_s \circ \varepsilon_{(t,\lambda_t,x)}^+ - X_s$, $D_{(t,\lambda_t,x)}H_s = H_s \circ \varepsilon_{(t,\lambda_t,x)}^+ - H_s$, $D_{(t,\lambda_t,x)}\lambda_s = \lambda_s \circ \varepsilon_{(t,\lambda_t,x)}^+ - \lambda_s$. Besides, $D_{(t,\lambda_t,x)}X_s = D_{(t,\lambda_t,x)}H_s = 0$ for $s < t$ and $D_{(t,\lambda_t,x)}\lambda_s = 0$ for $s \leq t$. Let $s \geq t$, according to Lemma 2.15 we have that

$D_{(t,\lambda_t,x)}H_s$

$$= H_{t-} + 1 + \int_{(t,s]\times\mathbb{R}_+\times\mathbb{R}} \mathbf{1}_{\{\theta\leq\lambda_u\circ\varepsilon^+_{(t,\lambda_t,x)}\}} N(d\theta, du, dy) - \left(H_t + \int_{(t,s]\times\mathbb{R}_+\times\mathbb{R}} \mathbf{1}_{\{\theta\leq\lambda_u\}} N(d\theta, du, dy) \right)$$

$$= -(H_t - H_{t-}) + 1 + \int_{(t,s]\times\mathbb{R}_+\times\mathbb{R}} \mathbf{1}_{\{\lambda_u\leq\theta\leq\lambda_u\circ\varepsilon^+_{(t,\lambda_t,x)}\}} N(d\theta, du, dy)$$

$$= 1 + \int_{(t,s]\times\mathbb{R}_+\times\mathbb{R}} \mathbf{1}_{\{\lambda_u\leq\theta\leq\lambda_u\circ\varepsilon^+_{(t,\lambda_t,x)}\}} N(d\theta, du, dy), \quad \mathbb{P}-a.s..$$

$$D_{(t,\lambda_t,x)}X_s$$

$$= X_{t-} + x + \int_{(t,s]\times\mathbb{R}_+\times\mathbb{R}} \mathbf{1}_{\{\theta\leq\lambda_u\circ\varepsilon^+_{(t,\lambda_t,x)}\}} y N(d\theta, du, dy) - \left(X_t + \int_{(t,s]\times\mathbb{R}_+\times\mathbb{R}} \mathbf{1}_{\{\theta\leq\lambda_u\}} y N(d\theta, du, dy) \right)$$

$$= -(X_t - X_{t-}) + x + \int_{(t,s]\times\mathbb{R}_+\times\mathbb{R}} \mathbf{1}_{\{\lambda_u\leq\theta\leq\lambda_u\circ\varepsilon^+_{(t,\lambda_t,x)}\}} y N(d\theta, du, dy)$$

$$= x + \int_{(t,s]\times\mathbb{R}_+\times\mathbb{R}} \mathbf{1}_{\{\lambda_u\leq\theta\leq\lambda_u\circ\varepsilon^+_{(t,\lambda_t,x)}\}} y N(d\theta, du, dy), \quad \mathbb{P}-a.s..$$

Note that we ignore the possible jump at time t of H, as it is equal to 0 $\mathbb{P}$-a.s.. This does not lead to an issue when integrating in t as $\mathbb{P}\left[\{u,\ \Delta_u H \neq 0\} < +\infty\right] = 1$ (the same comment apply to X). More generally, the Malliavin derivative DF coincides $\mathbb{P}\otimes(dt\otimes d\theta\otimes\nu)$-a.e. with the operator $F\circ\varepsilon^+_{(t,\theta,x)} - F$ where it is understood that we exclude the possible jump of N at time t which is supported by a set of probability 0. Similarly, for $s > t$, we have that

$$D_{(t,\lambda_t,x)}\lambda_s = \Phi(s-t) + \int_{(t,s)} \Phi(s-u)d(H_u\circ\varepsilon^+_{(t,\lambda_t,x)} - dH_u)$$

$$= \Phi(s-t) + \int_{(t,s)} \Phi(s-u)d(D_{(t,\lambda_t,x)}H_u).$$

For $s = t$, $D_{(t,\lambda_t,x)}\lambda_s = 0$ as λ is $\mathbb{F}^H$-predictable. Writing $\lambda_u\circ\varepsilon^+_{(t,\lambda_t,x)} = \lambda_u + D_{(t,\lambda_t,x)}\lambda_u$ proves that $(D_{(t,\lambda_t,x)}X, D_{(t,\lambda_t,x)}H, D_{(t,\lambda_t,x)}\lambda)$ is solution to SDE (2.5) which admits a unique solution following [13, Theorem 3.3].

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

$$\mathbb{E}\left[\hat{H}^t_{s_2} - \hat{H}^t_{s_1}|\mathcal{F}^H_{s_1}\right] = \mathbb{E}\left[\mathbb{E}\left[\hat{H}^t_{s_2} - \hat{H}^t_{s_1}|\mathcal{F}^N_{s_1}\right]|\mathcal{F}^H_{s_1}\right]$$

$$= \mathbb{E}\left[\int_{s_1}^{s_2} \mathbb{E}\left[\int_{\mathbb{R}_+} \mathbf{1}_{\{\lambda_u\leq\theta\leq\lambda_u+\hat{\lambda}^t_u\}}d\theta|\mathcal{F}^N_{s_1}\right] dt|\mathcal{F}^H_{s_1}\right]$$

$$= \int_{s_1}^{s_2} \mathbb{E}\left[\hat{\lambda}^t_u|\mathcal{F}^H_{s_1}\right] dt.$$

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

We conclude this section by re-writing the integration by parts formula (2.2) for the Hawkes process.

Theorem 2.20. *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 $\mathcal{S}$.

(ii) For any $\mathcal{F}_\infty^X$-measurable random variable F with $\mathbb{E}[|F|^2] < +\infty$,

$$\mathbb{E}\left[F\delta^N(Z\mathbf{1}_{\{\theta \leq \lambda_t\}})\right] = \mathbb{E}\left[\int_{\mathbb{R}_+ \times \mathbb{R}} \lambda_t Z_{(t,x)} D_{(t,\lambda_t,x)} F dt \nu(dx)\right], \qquad (2.6)$$

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, λ is $\mathbb{F}^H$-predictable. Hence, for any (t,θ,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 λ are predictable, our assumptions imply that

$$\mathbb{E}\left[\left(\int_{\mathbb{R}_+^2 \times \mathbb{R}} Z_{(t,x)}\mathcal{Z}_{(t,\theta)} N(dt,d\theta,dx)\right)^2\right]$$

$$\leq 2\left(\mathbb{E}\left[\left(\int_{\mathbb{R}_+^2 \times \mathbb{R}} Z_{(t,x)}\mathcal{Z}_{(t,\theta)}(N(dt,d\theta,dx) - dt d\theta \nu(dx))\right)^2\right]\right.$$

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

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

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

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, $\delta^N(Z\mathcal{Z})$ is well-defined and the IBP formula (2.2) is in force. It gives that

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

$$= \mathbb{E}\left[\int_0^{+\infty} \int_{\mathbb{R}} Z_{(t,x)} \int_{\mathbb{R}_+} \mathbf{1}_{\{\theta \leq \lambda_t\}} D_{(t,\theta,x)}F d\theta \nu(dx) dt\right]$$

$$= \mathbb{E}\left[\int_0^{+\infty} \int_{\mathbb{R}} Z_{(t,x)} \int_0^{\lambda_t} D_{(t,\lambda_t,x)}F d\theta \nu(dx) dt\right]$$

$$= \mathbb{E}\left[\int_0^{+\infty} \lambda_t \int_{\mathbb{R}} Z_{(t,x)} D_{(t,\lambda_t,x)}F \nu(dx) dt\right].$$

$\square$

2.4 Elements on Stein's method

Whereas Stein's method has been introduced by C. M. Stein in [27], the combination of the Malliavin calculus with Stein's method (and known as the Nourdin-Peccati's approach) has been initiated in [18] for the approximation of Gaussian functionals, extended in [22] for Poisson functionals (which is closer to our paper). We introduce in this section the original Stein's approach which allows one to derive Inequality (2.7) below. Then in the proof of Theorem 3.4 we will adapt [22] to transform the right-hand side of (2.7) into the right-hand side of (3.3) using the Malliavin calculus. We refer to [19] for a complete exposition of the original Stein method and of the Nourdin-Peccati approach.

Definition 2.21. Let F and G two random variables defined on some $(\Omega, \mathcal{F}_\infty^N, \mathbb{P})$. We recall that the Wasserstein distance between $\mathcal{L}_F$ and $\mathcal{L}_G$ (or simply between F and G) as :

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

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

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

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

Consider h in Lip. C. M. Stein proved in [27], that there exists a function f_h in $\mathcal{F}_W^0$ 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 :

$$\left| \mathbb{E}[h(F)] - \mathbb{E}[h(G)] \right| = \left| \mathbb{E}[\sigma^2 f_h'(F) - F f_h(F)] \right|.$$

Hence,
$$d_W(F, G) \leq \sup_{f \in \mathcal{F}_W^0} \left| \mathbb{E}[\sigma^2 f'(F) - F f(F)] \right|. \tag{2.7}$$

In addition, the right hand side is equal to 0 if and only if $F \sim \mathcal{N}(0, \sigma^2)$.

Remark 2.22. The original result proven by C. M. Stein in [27] was the existence of a function f_h in

$$\mathcal{F}_W := \left\{ f : \mathbb{R} \to \mathbb{R}, \text{ twice differentiable with } \|f'\|_\infty \leq 1, \|f''\|_\infty \leq 2 \right\}.$$

The sup on $\mathcal{F}_W^0$ in 2.7 coincides with a sup on $\mathcal{F}_W$ if F is centered. To see that, it is enough to replace $f \in \mathcal{F}_W$ by $f - f(0)$ and use the fact that $\mathbb{E}[f(0)F] = 0$.

3 Main results

We present in Section 3.1 a general bound on the Wasserstein distance between a given Hawkes functional and a centered Gaussian distribution. This bound is then applied in Section 3.2 to provide Berry-Esséen bounds for CLTs for compound Hawkes processes. We refer to Section 3.2.1 for a review of our results on quantitative counterparts of limit theorems (1.2)-(1.3) regarding the compound Hawkes process.

3.1 The general result

Recall that we consider the compound Hawkes process given as the unique solution to SDE (2.4)

$$
\begin{cases}
X_t = \int_{(0,t]\times\mathbb{R}_+\times\mathbb{R}} x\mathbf{1}_{\{\theta\leq\lambda_s\}}N(ds,d\theta,dx), \quad t\geq 0 \\[2mm]
H_t = \int_{(0,t]\times\mathbb{R}_+\times\mathbb{R}} \mathbf{1}_{\{\theta\leq\lambda_s\}}N(ds,d\theta,dx), \quad t\geq 0 \\[2mm]
\lambda_t = \mu + \int_{(0,t)} \Phi(t-u)dH_u, \quad t\geq 0.
\end{cases}
$$

Assumption 3.1. *Throughout this paper we recall that ν is a probability measure on $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ with $\nu(\{0\})=0$. In addition, we assume that :*

$$
m := \int_{\mathbb{R}} x\nu(dx) < +\infty, \quad \vartheta^2 := \int_{\mathbb{R}} x^2\nu(dx) < +\infty. \tag{3.1}
$$

Assumption 3.2. *The mapping Φ is such that*

$$
\|\Phi\|_1 = \int_0^{+\infty} \Phi(u)du < 1.
$$

This assumption is fundamental for our analysis as it allows us to set

$$
\psi := \sum_{n\geq 1} \phi^{(*n)}, \tag{3.2}
$$

with $\phi^{(*n)}$ the n-th convolution of ϕ with itself. We have that

$$
\int_0^{+\infty} \psi(t)dt = \int_0^{+\infty} \sum_{n\geq 1}\phi^{(*n)}(t)dt = \sum_{n\geq 1}\int_0^{+\infty}\phi^{(*n)}(t)dt = \sum_{n\geq 1}\|\phi\|_1^n = \frac{\|\phi\|_1}{1-\|\phi\|_1}.
$$

Assumption 3.3. *The mapping Φ is such that*

$$
\int_0^{+\infty} u\Phi(u)du < +\infty.
$$

Theorem 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) $F := \delta^N(Z\mathcal{Z}) = \int_{\mathbb{R}_+^2\times\mathbb{R}} Z_{(t,x)}\mathcal{Z}_{(t,\theta)}N(dt,d\theta,dx) - \int_{\mathbb{R}_+\times\mathbb{R}} Z_{(t,x)}\lambda_t dt\nu(dx).$

Let $\gamma > 0$. Then, letting $G \sim \mathcal{N}(0,\gamma^2)$,

$$
d_W(F,G) \leq \mathbb{E}\left[\left|\gamma^2 - \int_{\mathbb{R}_+\times\mathbb{R}} Z_{(t,x)}\lambda_t D_{(t,\lambda_t,x)}F dt\nu(dx)\right|\right] + \mathbb{E}\left[\int_{\mathbb{R}_+\times\mathbb{R}} |Z_{(t,x)}|\lambda_t \left|D_{(t,\lambda_t,x)}F\right|^2 dt\nu(dx)\right].
$$

$$
\tag{3.3}
$$

Proof. Note first of all that by Theorem 2.20, F is well-defined. Then, let f in $\mathcal{F}_W^0$ (see Definition 2.21). As $F = \delta^N(Z\mathcal{Z})$, the integration by parts formula (2.6) leads to

$$\mathbb{E}\left[f(F)F\right] = \mathbb{E}\left[f(F)\delta^N(Z\mathcal{Z})\right]$$
$$= \mathbb{E}\left[\int_{\mathbb{R}_+\times\mathbb{R}} Z_{(t,x)}\lambda_t D_{(t,\lambda_t,x)}f(F)dt\nu(dx)\right]$$
$$= \mathbb{E}\left[\int_{\mathbb{R}_+\times\mathbb{R}} Z_{(t,x)}\lambda_t\left(f(F\circ\varepsilon^+_{(t,\lambda_t,x)}) - f(F)\right)dt\nu(dx)\right].$$

Using a Taylor expansion, (with $\bar{F}$ be a random element between $F_T\circ\varepsilon^+_{(t,\lambda_t,x)}$ and F) we have

$$f(F\circ\varepsilon^+_{(t,\lambda_t,x)}) - f(F) = f'(F)D_{(t,\lambda_t,x)}F + \frac{1}{2}f''(\bar{F})\left|D_{(t,\lambda_t,x)}F\right|^2.$$

Hence,

$$\left|\mathbb{E}\left[\gamma^2 f'(F) - f(F)F\right]\right|$$
$$= \left|\mathbb{E}\left[f'(F)\left(\gamma^2 - \int_{\mathbb{R}_+\times\mathbb{R}} Z_{(t,x)}\lambda_t D_{(t,\lambda_t,x)}F dt\nu(dx)\right)\right] - \frac{1}{2}\mathbb{E}\left[\int_{\mathbb{R}_+\times\mathbb{R}} Z_{(t,x)}\lambda_t f''(\bar{F})\left|D_{(t,\lambda_t,x)}F\right|^2 dt\nu(dx)\right]\right|$$
$$\leq \|f'\|_\infty\mathbb{E}\left[\left|\gamma^2 - \int_{\mathbb{R}_+\times\mathbb{R}} Z_{(t,x)}\lambda_t D_{(t,\lambda_t,x)}F dt\nu(dx)\right|\right] + \frac{\|f''\|_\infty}{2}\mathbb{E}\left[\int_{\mathbb{R}_+\times\mathbb{R}} |Z_{(t,x)}|\lambda_t\left|D_{(t,\lambda_t,x)}F\right|^2 dt\nu(dx)\right]$$
$$\leq \mathbb{E}\left[\left|\gamma^2 - \int_{\mathbb{R}_+\times\mathbb{R}} Z_{(t,x)}\lambda_t D_{(t,\lambda_t,x)}F dt\right|\right] + \mathbb{E}\left[\int_{\mathbb{R}_+\times\mathbb{R}} |Z_{(t,x)}|\lambda_t\left|D_{(t,\lambda_t,x)}F\right|^2 dt\nu(dx)\right],$$

as f belongs to $\mathcal{F}_W^0$. $\square$

In particular, we obtain the following corollary.

Corollary 3.5. *For each $T > 0$, consider $Z^T := (Z^T_{(t,x)})_{(t,x)\in\mathbb{R}_+\times\mathbb{R}}$ defined as :*

$$Z^T_{(t,x)} := x\alpha_t\mathbf{1}_{t\in[0,T]}, \quad t \geq 0, \ x \in \mathbb{R},$$

where $(\alpha_t)_{t\in[0,T]}$ is a $(\mathcal{F}_t^X)_{t\in[0,T]}$-predictable process. Let also :

$$F_T := \delta^N(Z^T\mathcal{Z}) = \int_{(0,T]\times\mathbb{R}_+\times\mathbb{R}} \alpha_s x\mathbf{1}_{\{\theta\leq\lambda_s\}}(N(ds,d\theta,dx) - dsd\theta\nu(dx))$$
$$= \int_{(0,T]}\alpha_t dX_t - m\int_0^T\alpha_t\lambda_t dt, \quad m := \int_\mathbb{R} x\nu(dx);$$

and $G \sim \mathcal{N}(0,\gamma^2)$ for any γ. Then

$$d_W(F_T, G) \leq \mathbb{E}\left[\left|\gamma^2 - \int_0^T\alpha_t\lambda_t\int_\mathbb{R} xD_{(t,\lambda_t,x)}F_T\nu(dx)dt\right|\right] + \mathbb{E}\left[\int_0^T|\alpha_t|\lambda_t\int_\mathbb{R}|x|\left|D_{(t,\lambda_t,x)}F_T\right|^2\nu(dx)dt\right].$$
$$(3.4)$$

Before going further we make the following remark showing that somehow the decomposition above is sharp.

Remark 3.6. Assume $\Phi \equiv 0$ so that H is an homogeneous Poisson process with intensity μ. Let $m = \int_{\mathbb{R}} y\nu(dy)$ and X is a compound Poisson process which can thus be represented as :

$$X_t = \sum_{i=1}^{H_T} Y_i,$$

with $(Y_i)_{i \geq 1}$ iid random variables independent of H with distribution ν. 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

$$\mathbb{E}\left[\left|\gamma^2 - \frac{1}{\sqrt{T}} \int_0^T \int_{\mathbb{R}} \lambda_t x D_{(t,\lambda_t,x)} F_T dt \nu(dx)\right|\right] = \left|\gamma^2 - \mu \vartheta^2\right| = 0.$$

Hence, the speed of convergence is completely contained in that case by the term

$$\frac{1}{\sqrt{T}} \mathbb{E}\left[\int_0^T \int_{\mathbb{R}} \lambda_t |x| \left|D_{(t,\lambda_t,x)} F_T\right|^2 dt \nu(dx)\right] = \frac{\mu \int_{\mathbb{R}} |x|^3 \nu(dx)}{\sqrt{T}}$$

and thus we recover a Berry-Esséen bound with central speed $T^{-1/2}$.

Remark 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's 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 δ 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 3.4 (and choosing $\nu(dx) = \delta_1(dx)$ for simplicity of notations), for a Hawkes functional F of the form $F = \delta^N(Z\mathcal{Z})$, by adapting the Nourdin-Peccati methodology to our framework (so by adapting [22]), one would obtain the inequality (once again using the notations of Theorem 3.4):

$$d_W(F,G) \leq \mathbb{E}\left[\left|\gamma^2 - \int_0^{+\infty} \lambda_t \left(-D_{(t,\lambda_t)} L^{-1} F\right) D_{(t,\lambda_t)} F dt\right|\right] + \mathbb{E}\left[\int_0^{+\infty} |D_{(t,\lambda_t)} L^{-1} F| \lambda_t \left|D_{(t,\lambda_t)} F\right|^2 dt\right].$$

$$(3.5)$$

However, as F is of the form $F = \delta^N(Z\mathcal{Z})$, bounds (3.3) and (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

$$\delta^N(\rho) = -\delta^N(-\rho) = -\delta^N(DL^{-1}F) = LL^{-1}F = F = \delta^N(Z\mathcal{Z}).$$

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 [25], [3] or [28]).

Remark 3.8. We would like to discuss here the work [28] which is the closest to ours. In [28], 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 [13] and fully exploited in this paper) that for a Hawkes process the impact of the operator $D_{(t,\theta,x)}$ on a Hawkes functional is the same regarding the value of θ (provided that $\theta \leq \lambda_t$) was not found. This remark is made clear in Proposition 2.16 and leads to a simplification of the integration by parts formula (see Relation (2.6)) for Hawkes functional and thus of the bound in Theorem 3.10 below. In [28], 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, when dealing with the particular case of a Hawkes process, plugging $u(t) := \frac{1}{\sqrt{T}}\mathbf{1}_{[0,T]}(t)$ in [28, Theorem 4.1] leads to an upper bound which does not converges to 0 as T goes to $+\infty$ (we chose the linear case with $\Phi(x) := \mu + x$ with the notations of [28, Theorem 4.1]). This might have been a motivation to consider an approximation of the Hawkes process by a counting process (δ_a in [28, p. 2120]) which is interesting but which does not cover strictly speaking Convergence (1.3).

3.2 Application to the Hawkes process

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 $X_t = H_t$, $\nu(dx) = \delta_1(dx)$ or equivalently $Y_i \equiv 1$ in (2.3)), it has been proved as [1, Lemma 7] that

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

(with W a Brownian motion and $\sigma^2 := \frac{\mu}{1 - \|\Phi\|_1}$) which in particular implies

$$\frac{H_T - \int_0^T \lambda_s ds}{\sqrt{T}} \xrightarrow[T \to +\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2). \tag{3.6}$$

16

In what follows, we provide for a compound Hawkes process $X := (X_t)_{t\in[0,T]}$,

$$X_t = \sum_{i=1}^{H_t} Y_i,$$

$((Y_i)_{i\geq 1}$ iid with common distribution ν and independent of H) the counterpart of the normal convergence (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 3.10, a general bound for the speed of convergence (with respect to T) of the convergence (3.6) in the generalized case of a compound Hawkes process.

2. We prove, as Theorem 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 (3.6) :

$$d_W\left(F_T, \mathcal{N}\left(0, \sigma^2\,\vartheta^2\right)\right) \leq \frac{C_{\alpha,\beta,\nu}}{\sqrt{T}};$$

3. We provide, as Theorem 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 [1, Theorem 2] as follows :

$$d_W\left(Y_T, \mathcal{N}(0, \tilde{\sigma}^2)\right) \leq \frac{\tilde{C}_{\alpha,\beta}}{\sqrt{T}},$$

where $Y_T := \frac{H_T - \int_0^T \mathbb{E}[\lambda_t] dt}{\sqrt{T}}$ and $\tilde{\sigma}^2 = \frac{\mu}{(1-\|\Phi\|_1)^3}$.

3.2.2 Quantitative Limit Theorems for compound Hawkes processes

In order to make precise some of the statements below we recall that the Malliavin derivative of X, H and λ involves the following parametrized system (see Proposition 2.19).

Notation 3.9. For fixed $t \geq 0$, we denote by $(\hat{X}_s^t, \hat{H}_s^t, \hat{\lambda}_s^t)_{s\geq t}$ the unique solution to the SDE

$$\begin{cases} \hat{H}_s^t = \int_{(t,s]\times\mathbb{R}_+\times\mathbb{R}} \mathbf{1}_{\{\lambda_u \leq \theta \leq \lambda_u + \hat{\lambda}_u^t\}} N(du, d\theta, dy), \quad s \geq t, \\[2mm] \hat{X}_s^t = \int_{(t,s]\times\mathbb{R}_+\times\mathbb{R}} y\mathbf{1}_{\{\lambda_u \leq \theta \leq \lambda_u + \hat{\lambda}_u^t\}} N(du, d\theta, dy), \quad s \geq t, \\[2mm] \hat{\lambda}_s^t = \Phi(s-t) + \int_{(t,s)} \Phi(s-u)d\hat{H}_u^t, \quad s > t, \; \hat{\lambda}_t^t = 0. \end{cases}$$

In addition we introduce different compensated martingale processes for the (shifted) Hawkes process and the (shifted) compound Hawkes process

$$M_s := X_s - m \int_0^s \lambda_u du, \quad s \in [0, T], \quad \hat{M}_s^t := \hat{X}_s^t - m \int_t^s \hat{\lambda}_u^t du, \quad s \in [t, T].$$

$$\mathcal{M}_s := H_s - \int_0^s \lambda_u du, \quad s \in [0, T], \quad \hat{\mathcal{M}}_s^t := \hat{H}_s^t - \int_t^s \hat{\lambda}_u^t du, \quad s \in [t, T].$$

Theorem 3.10. *Assume Assumptions 3.1, 3.2 and 3.3 are in force. For $T > 0$, let $Z_{(t,x)}^T := \frac{x}{\sqrt{T}}$, $t \in [0, T]$ and recall*

$$F_T = \delta^N(Z^T \mathcal{Z}^T) = \frac{X_T - m \int_0^T \lambda_t dt}{\sqrt{T}}.$$

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 Φ and ν 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 3.9.

Proof. To ease the presentation, we chose to provide here the main steps of the proof and to postpone in Section 4 the proof of general results concerning the convergence of moments of order 2. By Theorem 3.4 we have that

$$d_W(F_T, G) \leq A_1 + A_2$$

where

$$A_1 := \mathbb{E}\left[\left|\sigma^2 \vartheta^2 - \frac{1}{\sqrt{T}} \int_0^T \int_{\mathbb{R}} x \lambda_t D_{(t,\lambda_t,x)} F_T dt \nu(dx)\right|\right], \quad A_2 := \frac{1}{\sqrt{T}} \mathbb{E}\left[\int_0^T \int_{\mathbb{R}} \lambda_t |x| \left|D_{(t,\lambda_t,x)} F_T\right|^2 dt \nu(dx)\right].$$

Note that

$$\sigma^2 \vartheta^2 = \lim_{T \to +\infty} \mathbb{E}[|F_T|^2] = \lim_{T \to +\infty} \frac{1}{T} \mathbb{E}[|M_T|^2] = d \lim_{T \to +\infty} \frac{1}{T} \mathbb{E}[H_T] = d \lim_{T \to +\infty} \frac{1}{T} \int_0^T \mathbb{E}[\lambda_t] dt,$$

and we will quantify the speed of convergence. Before going further, recall that according to Proposition 2.19,

$$D_{(t,\lambda_t,x)} F_T = \frac{1}{\sqrt{T}} D_{(t,\lambda_t,x)} M_T = \frac{1}{\sqrt{T}} \left(x + \hat{M}_T^t\right) \tag{3.7}$$

where $(\hat{M}_s^t := \hat{X}_s^t - m \int_t^s \hat{\lambda}_u^t du)_{s \in [t,T]}$, and $(\hat{H}^t, \hat{\lambda}^t)$ is defined by (2.5) and recalled in Notation 3.9. We treat both terms A_1 and A_2 separately.

Term A_1
We have

$$A_1 = \mathbb{E}\left[\left|\sigma^2 \vartheta^2 - \frac{1}{\sqrt{T}} \int_0^T \int_{\mathbb{R}} x \lambda_t D_{(t,\lambda_t,x)} F_T dt \nu(dx)\right|\right]$$

18

$$= \mathbb{E}\left[\left|\sigma^2\vartheta^2 - \frac{\vartheta^2}{T}\int_0^T \lambda_t dt - \frac{1}{T}\int_0^T\int_{\mathbb{R}} x\lambda_t \hat{M}_T^t dt\nu(dx)\right|\right]$$

$$\le \left|\sigma^2\vartheta^2 - \frac{\vartheta^2}{T}\int_0^T \mathbb{E}[\lambda_t]dt\right| + \frac{\vartheta^2}{T}\mathbb{E}\left[\left|\int_0^T(\lambda_t - \mathbb{E}[\lambda_t])dt\right|\right] + \frac{|m|}{T}\mathbb{E}\left[\left|\int_0^T \lambda_t \hat{M}_T^t dt\right|\right]$$

$$=: A_{1,1} + \vartheta^2 A_{1,2} + |m|A_{1,3}.$$

By Lemma 4.1 (in Section 4),

$$A_{1,1} = O\left(\frac{1}{T}\right). \tag{3.8}$$

We turn to term $A_{1,2}$. According to the second line of [1, Lemma 4] one has

$$H_T - \mathbb{E}[H_T] = \mathcal{M}_T + \int_0^T \psi(T-s)\mathcal{M}_s ds,$$

where $\mathcal{M}_t = H_t - \int_0^t \lambda_s ds$. By subtracting H_T from the equation one gets

$$-\mathbb{E}[H_T] = -\int_0^T \lambda_s ds + \int_0^T \psi(T-s)\mathcal{M}_s ds,$$

thus since $\mathbb{E}[H_T] = \int_0^T \mathbb{E}[\lambda_s]ds$

$$\int_0^T \lambda_s - \mathbb{E}[\lambda_s]ds = \int_0^T \psi(T-s)\mathcal{M}_s ds.$$

This means that

$$A_{1,2} \le \frac{1}{T}\int_0^T \psi(T-s)\mathbb{E}[|\mathcal{M}_s|]ds \le \frac{1}{T}\int_0^T \psi(T-s)\mathbb{E}[|\mathcal{M}_s|^2]^{1/2}ds \le \frac{1}{T}\int_0^T \psi(T-s)\mathbb{E}[H_s]^{1/2}ds.$$

According to the proof of [1, Lemma 5], with Assumption 3.3 that $\int_0^{+\infty} s\phi(s)ds < +\infty$, then $C := \int_0^{+\infty} s\psi(s)ds < +\infty$ and a fortiori $\forall p \in [0,1], \int_0^{+\infty} s^p\psi(s)ds < +\infty$. Following once again [1, Lemma 4],

$$\mathbb{E}[H_t] = \mu t + \int_0^t \psi(t-s)s ds = \mu t + \int_0^t \psi(s)(t-s)ds \le (\mu + \|\psi\|_1)t + C.$$

Hence, using $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$ we have the following inequality

$$\mathbb{E}[H_t]^{1/2} \le A\sqrt{t} + B.$$

Now the term $A_{1,2}$ becomes bounded by

$$A_{1,2} \le \frac{A}{T}\int_0^T \psi(T-s)\sqrt{s}ds + \frac{B}{T}\int_0^T \psi(T-s)ds,$$

$$\le \frac{A}{T}\int_0^T \psi(s)\sqrt{T-s}ds + \frac{B}{T}\|\psi\|_1,$$

$$\le \frac{A}{\sqrt{T}}\int_0^T \psi(s)\sqrt{\frac{T-s}{T}}ds + \frac{B}{T}\|\psi\|_1,$$

19

$$\leq \frac{A}{\sqrt{T}} \int_0^T \psi(s)ds + \frac{B}{T}\|\psi\|_1,$$

$$\leq (\frac{A}{\sqrt{T}} + \frac{B}{T})\|\psi\|_1,$$

leading to

$$A_{1,2} = O\left(\frac{1}{\sqrt{T}}\right). \tag{3.9}$$

Combining (3.8) and (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

$$
\begin{aligned}
A_2 &= \frac{1}{\sqrt{T}}\mathbb{E}\left[\int_0^T \int_{\mathbb{R}} \lambda_t |x| \left|D_{(t,\lambda_t,x)}F_T\right|^2 \nu(dx)dt\right]\\
&\leq \frac{2\int_{\mathbb{R}} |x|^3 \nu(dx)}{T^{3/2}}\mathbb{E}\left[\int_0^T \lambda_t dt\right] + \frac{2}{T^{3/2}}\mathbb{E}\left[\int_0^T \int_{\mathbb{R}} |x|\lambda_t \left|\hat{M}_T^t\right|^2 \nu(dx)dt\right]\\
&= \frac{2\int_{\mathbb{R}} |x|^3 \nu(dx)}{T^{3/2}}\mathbb{E}\left[\int_0^T \lambda_t dt\right] + \frac{2\int_{\mathbb{R}} |x|\nu(dx)}{T^{3/2}}\mathbb{E}\left[\int_0^T \lambda_t \left|\hat{M}_T^t\right|^2 dt\right]\\
&=: 2\left(\int_{\mathbb{R}} |x|^3 \nu(dx)A_{2,1} + \int_{\mathbb{R}} |x|\nu(dx)A_{2,2}\right).
\end{aligned}
$$

By Lemma 4.1, we immediately get that

$$A_{2,1} = O\left(\frac{1}{\sqrt{T}}\right). \tag{3.10}$$

Finally, for Term $A_{2,2}$, it holds that

$$
\begin{aligned}
A_{2,2} &= \frac{1}{T^{3/2}}\mathbb{E}\left[\int_0^T \lambda_t \left|\hat{M}_T^t\right|^2 dt\right]\\
&= \frac{1}{T^{3/2}}\int_0^T \mathbb{E}\left[\lambda_t \mathbb{E}\left[\left|\hat{M}_T^t\right|^2 |\mathcal{F}_t\right]\right] dt\\
&= \frac{1}{T^{3/2}}\int_0^T \mathbb{E}\left[\lambda_t \mathbb{E}\left[\int_t^T \int_{\mathbb{R}} \int_{\mathbb{R}_+} x^2 \mathbf{1}_{\{\theta \leq \hat{\lambda}_u^t\}} N(du, d\theta, dx)|\mathcal{F}_t\right]\right] dt\\
&= \frac{\vartheta^2}{T^{3/2}}\int_0^T \mathbb{E}\left[\lambda_t \int_t^T \mathbb{E}\left[\hat{\lambda}_s^t |\mathcal{F}_t\right] ds\right] dt
\end{aligned}
$$

where we used for the last equality the identity $\hat{H}_t^t = 0$. Using Lemma 4.2,

$$|A_{2,2}| \leq \vartheta^2 \frac{\|\phi\|_1(1 + \|\psi\|_1)}{T^{3/2}}\int_0^T \mathbb{E}[\lambda_t]dt,$$

and since $\int_0^T \mathbb{E}[\lambda_t]dt = \mathbb{E}[H_T] \leq AT + B$ where A and B are positive constants (see computations for Term $A_{1,1}$ above), we have

$$A_{2,1} = O\left(\frac{1}{\sqrt{T}}\right). \tag{3.11}$$

20

Combining (3.10) and (3.11), we get that

$$A_2 = O\left(\frac{1}{\sqrt{T}}\right).$$

$\square$

Remark 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 3.10, determining the speed of convergence for a general Hawkes process requires the speed of convergence for R_T. It is somehow embarrassing to admit that we were not able to deal with it for a general kernel Φ as it calls for a precise statement of the correlation between the original intensity λ and somehow the one of the shifted (more precisely of the Malliavin derivative) Hawkes martingale $\hat{M}$ (defined in Notation 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 research area (see *e.g.* [14]). However, in particular cases we can provide estimates on this quantity R_T as we will see in Theorem 3.12. It is also interesting to point out that this term is specific to the self-exciting feature of the intensity as in the Poisson case, that is when $\Phi \equiv 0$, we have that $R_T = 0$ for any T (see Remark 3.6).

Theorem 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 3.10 ($F_T = \frac{H_T - \int_0^T \lambda_t dt}{\sqrt{T}}$; $G \sim \mathcal{N}(0, \sigma^2 \vartheta^2)$), there exists in both cases a constant $C_{\alpha,\beta,\nu} > 0$ depending only on α, β and ν such that for any $T > 0$

$$d_W(F_T, G) \leq \frac{C_{\alpha,\beta,\nu}}{\sqrt{T}}.$$

Proof. Obviously, Assumptions (3.2)-(3.3) are in force. By Theorem 3.10, we need to estimate the quantity

$$R_T = \frac{1}{T}\mathbb{E}\left[\left|\int_0^T \lambda_t \hat{M}_T^t dt\right|\right] = \frac{1}{T}\mathbb{E}\left[\left|\int_0^T \lambda_t \left(D_{(t,\lambda_t)} M_T - 1\right) dt\right|\right],$$

where $\hat{M}^t$ is defined in Notation 3.9. In Lemma 4.7 we prove that $R_T = O(\frac{1}{\sqrt{T}})$. $\square$

3.2.3 Alternative quantitative Limit Theorem for the exponential and Erlang Hawkes processes

In this section we consider the Hawkes process H (alternatively one can set $\nu(dx) = \delta_1(dx)$ or $Y_i \equiv 1$ in Representation (2.3)). It has been proven in [1] that as T goes to infinity

$$Y_T \xrightarrow[T \to +\infty]{\mathcal{L}} \mathcal{N}(0, \tilde{\sigma}^2),$$

where $Y_T = \frac{H_T - \int_0^T \mathbb{E}[\lambda_t] dt}{\sqrt{T}}$ 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 3.13. *Set* $Y_T = \frac{H_T - \int_0^T \mathbb{E}[\lambda_t]dt}{\sqrt{T}}$ *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* α, β *such that for any* $T > 0$

$$d_W(Y_T, G) \leq \frac{C_{\alpha,\beta}}{\sqrt{T}}.$$

Proof. As we have shown in Lemma 4.8, 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 (2.7)

$$d_W\left(\frac{Y_T}{\gamma}, \frac{G}{\gamma}\right) \leq \sup_{f \in \mathcal{F}_W^0} \left| \mathbb{E}\left[\frac{\tilde{\sigma}^2}{\gamma^2} f'\left(\frac{Y_T}{\gamma}\right) - \frac{Y_T}{\gamma} f\left(\frac{Y_T}{\gamma}\right) \right] \right|,$$

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:

$$d_W\left(\frac{Y_T}{\gamma}, \frac{G}{\gamma}\right)$$

$$\leq \sup_{f \in \mathcal{F}_W^0} \left| \mathbb{E}\left[\frac{\tilde{\sigma}^2}{\gamma^2} f'(F_T + \mathfrak{R}_T) - (F_T + \mathfrak{R}_T) f(F_T + \mathfrak{R}_T) \right] \right|,$$

$$= \sup_{f \in \mathcal{F}_W^0} \left| \mathbb{E}\left[\frac{\tilde{\sigma}^2}{\gamma^2} (f'(F_T) + \mathfrak{R}_T f''(\bar{X})) - (F_T + \mathfrak{R}_T)(f(F_T) + \mathfrak{R}_T f'(X^*)) \right] \right|,$$

$$= \sup_{f \in \mathcal{F}_W^0} \left| \mathbb{E}\left[\frac{\tilde{\sigma}^2}{\gamma^2} f'(F_T) - F_T f(F_T) + \frac{\tilde{\sigma}^2}{\gamma^2} \mathfrak{R}_T f''(\bar{X}) - \mathfrak{R}_T f(F_T) - F_T \mathfrak{R}_T f'(X^*) - \mathfrak{R}_T^2 f'(X^*)) \right] \right|,$$

$$\leq \sup_{f \in \mathcal{F}_W^0} \left| \mathbb{E}\left[\frac{\tilde{\sigma}^2}{\gamma^2} f'(F_T) - F_T f(F_T) \right] \right|$$

$$+ \sup_{f \in \mathcal{F}_W^0} \mathbb{E}\left[\left| \frac{\tilde{\sigma}^2}{\gamma^2} R_T f''(\bar{X}) \right| + |\mathfrak{R}_T f(F_T)| + |F_T \mathfrak{R}_T f'(X^*)| + |\mathfrak{R}_T^2 f'(X^*)| \right].$$

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 3.12 we have proven that

$$\sup_{f \in \mathcal{F}_W^0} \left| \mathbb{E}\left[\sigma^2 f'(F_T) - F_T f(F_T) \right] \right| \leq O\left(\frac{1}{\sqrt{T}}\right).$$

22

And since f is in $\mathcal{F}_W^0$, the first and second derivatives are bounded and the Wasserstein distance is thus bounded by

$$d_W\left(\frac{Y_T}{\gamma}, \frac{G}{\gamma}\right) \leq C\left(\frac{1}{\sqrt{T}} + \mathbb{E}[|\mathfrak{R}_T|] + \sup_{f \in \mathcal{F}_W^0} \mathbb{E}[|\mathfrak{R}_T f(F_T)|] + \mathbb{E}[|\mathfrak{R}_T F_T|] + \mathbb{E}[|\mathfrak{R}_T^2|]\right)$$

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

$$\begin{aligned}
\mathbb{E}[\mathfrak{R}_T f(F_T)]^2 &\leq \mathbb{E}[\mathfrak{R}_T^2]\mathbb{E}[|f(F_T)|^2], \\
&\leq \mathbb{E}[\mathfrak{R}_T^2]\mathbb{E}[\|f'\|_\infty^2 |F_T|^2], \quad \text{using mean value equality and the fact that } f \in \mathcal{F}_W^0 \\
&\leq \mathbb{E}[\mathfrak{R}_T^2]\mathbb{E}[|F_T|^2].
\end{aligned}$$

In order to have an upper bound on $\mathbb{E}[|F_T|^2]$, we recall that $F_T = \frac{M_T}{\sqrt{T}}$ where $M_T = H_T - \int_0^T \lambda_t dt$ is a martingale. This means that $\mathbb{E}[|M_T|^2] = \mathbb{E}[H_T] = O(T)$ (the laste estimate can be found in [10]), 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

$$\begin{aligned}
\mathbb{E}[\mathfrak{R}_T^2] &= \mathbb{E}\left[\left(\frac{\mathbb{E}[\lambda_T] - \lambda_T}{\beta\sqrt{T}}\right)^2\right], \\
&= \frac{\mathbb{E}[(\lambda_T - \mathbb{E}[\lambda_T])^2]}{\beta^2 T}, \\
&= \frac{\text{Var}(\lambda_T)}{\beta^2 T}, \quad \text{since the second moment of } \lambda \text{ is bounded (according to Lemma 4.6),} \\
&= O\left(\frac{1}{T}\right).
\end{aligned}$$

Case 2 : the kernel is an Erlang function

In this case the squared remainder becomes bounded by:

$$\mathbb{E}[\mathfrak{R}_T^2] = \mathbb{E}\left[\left(\frac{\mathbb{E}[\lambda_T] - \lambda_T}{\beta\sqrt{T}} + \frac{\mathbb{E}[\xi_T] - \xi_T}{\beta^2\sqrt{T}}\right)^2\right],$$

$$\leq 2\mathbb{E}\left[\left(\frac{\mathbb{E}[\lambda_T] - \lambda_T}{\beta\sqrt{T}}\right)^2 + \left(\frac{\mathbb{E}[\xi_T] - \xi_T}{\beta^2\sqrt{T}}\right)^2\right], \quad \text{since } 2ab \leq a^2 + b^2,$$

$$\leq \frac{2\text{Var}(\lambda_T)}{\beta^2 T} + \frac{2\text{Var}(\xi_T)}{\beta^2 T},$$

$$\leq O\left(\frac{1}{T}\right) \quad \text{since the second moment of } \xi \text{ is bounded (according to the proof of Lemma 4.6).}$$

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

$$d_W(Y_T, G) \leq \frac{C_{\alpha,\beta}}{\sqrt{T}}.$$

$\square$

Conclusion : In this paper we have computed Berry-Esséen 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.

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, H, λ) as (2.4) and $(\hat{X}^t, \hat{H}^t, \hat{\lambda}^t)$ describing the Malliavin derivative recalled in Notation 3.9.

4.1 General estimates

Throughout this section, we assume that Assumptions 3.1, 3.2 and 3.3 are in force.

Lemma 4.1. *For any $T > 0$, and recalling that $\sigma^2 = \frac{\mu}{1 - \|\phi\|_1}$,*

$$\left|\frac{\sigma^2}{\vartheta^2} - \frac{1}{T}\int_0^T \mathbb{E}[\lambda_t]dt\right| = O\left(\frac{1}{T}\right).$$

Proof. According to [1, Theorem 2], under Assumptions 3.2 and 3.3,

$$\mathbb{E}[H_t] = \int_0^t \mathbb{E}[\lambda_s]ds = \mu t + \mu \int_0^t \psi(t-s)sds,$$

where ψ is defined by (3.2). Recall that

$$\int_0^{+\infty} \psi(t)dt = \frac{\|\phi\|_1}{1 - \|\phi\|_1}.$$

As $\sigma^2 = \lim_{T \to +\infty} \frac{\mathbb{E}[H_T]}{T} = \lim_{T \to +\infty} \frac{\int_0^T \mathbb{E}[\lambda_t]dt}{T}$, the following computation gives that $\sigma^2 = \frac{\mu}{1 - \|\phi\|_1}$. Indeed,

$$\frac{1}{T}\int_0^T \mathbb{E}[\lambda_t]dt - \sigma^2 = \mu + \frac{\mu}{T}\int_0^T \psi(T - s)s\,ds - \sigma^2,$$

$$= \mu + \frac{\mu}{T}\int_0^T \psi(s)(T - s)ds - \sigma^2,$$

$$= \mu + \mu\int_0^T \psi(s)ds - \frac{\mu}{T}\int_0^T s\psi(s)ds - \sigma^2,$$

$$= \mu + \mu\int_0^{+\infty} \psi(s)ds - \mu\int_T^{+\infty} \psi(s)ds - \frac{\mu}{T}\int_0^T s\psi(s)ds - \sigma^2,$$

$$= \left(\frac{\mu}{1 - \|\phi\|_1} - \sigma^2\right) - \mu\int_T^{+\infty} \psi(s)ds - \frac{\mu}{T}\int_0^T s\psi(s)ds.$$

Following [1, Lemma 5], $\int_0^{+\infty} s\psi(s)ds < +\infty$ which entails that $\frac{\mu}{T}\int_0^T s\psi(s)ds = O\left(\frac{1}{T}\right)$. Concerning the other term, one can note that

$$\int_T^{+\infty} \psi(s)ds \le \int_T^{+\infty} \psi(s)\frac{s}{T}ds \le \frac{1}{T}\int_0^{+\infty} s\psi(s)ds.$$

These estimates conclude the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Lemma 4.2. *We make use of the notation of Proposition 2.19. Let $T > 0$ and $0 \le s \le t \le T$. The following estimate holds*

$$\int_t^T \mathbb{E}_t[\hat{\lambda}_s^t]ds \le \|\phi\|_1(1 + \|\psi\|_1).$$

Proof. Recall that for $t \ge 0$, $\mathbb{E}_t[\cdot]$ stands for the conditional expectation $\mathbb{E}[\cdot|\mathcal{F}_t^N]$. Taking the conditional expectation in Equation (2.5) for $0 \le t \le s$ 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 $t \ge 0$ set $f_t(s) := \mathbb{1}_{s \ge t}\mathbb{E}_t[\hat{\lambda}_s^t]$ and $\phi_t(s) := \mathbb{1}_{s \ge t}\phi(s - t)$, the last equation becomes

$$f_t(s) = \phi_t(s) + \int_0^s \phi(s - u)f_t(u)du.$$

A straightforward application of [1, Lemma 3] yields

$$f_t(s) = \phi_t(s) + \int_0^s \psi(s - u)\phi_t(u)du,$$

25

and by integrating between t and T one obtains

$$\int_t^T \mathbb{E}_t[\hat{\lambda}_s^t]ds = \int_t^T f_t(s)ds,$$

$$= \int_t^T \phi_t(s)ds + \int_t^T \int_0^s \psi(s-u)\phi_t(u)duds,$$

$$= \int_t^T \phi_t(s)ds + \int_t^T \int_t^s \psi(s-u)\phi_t(u)duds,$$

$$= \int_t^T \phi_t(s)ds + \int_t^T \int_t^T \mathbb{1}_{s\geq u}\psi(s-u)\phi_t(u)duds.$$

Since all the involved functions are positive, we use Fubini's theorem and exchange the integrals to get

$$\int_t^T \mathbb{E}_t[\hat{\lambda}_s^t]ds = \int_t^T \phi_t(s)ds + \int_t^T \int_u^T \psi(s-u)ds\phi_t(u)du$$

$$\leq \|\phi\|_1 + \int_t^T \|\psi\|_1\phi_t(u)du$$

$$\leq \|\phi\|_1(1+\|\psi\|_1).$$

$\square$

4.2 Specific estimates for the exponential and the Erlang's kernels

4.2.1 Definition and some properties of the exponential and the Erlang Hawkes processes

Definition 4.3 (Exponential Hawkes process). A counting process H as in Definition 2.9 is referred to

(i) an exponential Hawkes process if there exist $(\alpha, \beta) \in \mathbb{R}_+^2$ such that

$$\alpha < \beta \quad \text{and} \quad \Phi(u) := \alpha e^{-\beta u}, \quad u \in \mathbb{R}_+. \tag{4.1}$$

(ii) an Erlang Hawkes process if there exist $(\alpha, \beta) \in \mathbb{R}_+^2$ such that

$$\alpha < \beta^2 \quad \text{and} \quad \Phi(u) := \alpha u e^{-\beta u}, \quad u \in \mathbb{R}_+. \tag{4.2}$$

Remark 4.4. Obviously, the two kernels above ((4.1)-(4.2)) satisfy Assumptions 3.2 and (3.3).

Proposition 4.5. *Let H be a Hawkes process and let λ be its intensity.*

(i) *If H is an exponential Hawkes process, then $(\lambda_t)_{t\in\mathbb{R}_+}$ is a Markov process that satisfies the following Dynkin formula for each function $g \in \mathcal{C}^1$ and for each $t \leq T$:*

$$\mathbb{E}[g(\lambda_T)|\mathcal{F}_t] = g(\lambda_t) + \mathbb{E}\Big[\int_t^T (\mathcal{D}g)(\lambda_s)ds|\mathcal{F}_t\Big],$$

where $\mathcal{D}$ is the infinitesimal generator:

$$\mathcal{D}g(\lambda) := \beta(\mu-\lambda)g'(\lambda) + \lambda\big(g(\lambda+\alpha) - g(\lambda)\big),$$

whenever these expectations are finite.

(ii) If H is an Erlang Hawkes process, then $(\lambda_t, \xi_t)_{t \in \mathbb{R}_+}$ is a Markov process, where $\xi_t = \int_0^t \alpha e^{-\beta(t-s)} dH_s$ is an auxiliary process. In this case, $(\lambda_t, \xi_t)_{t \in \mathbb{R}_+}$ satisfies this Dynkin formula for each function $g \in \mathcal{C}^1$ and for each $t \leq T$

$$\mathbb{E}[g(\lambda_T, \xi_T) | \mathcal{F}_t] = g(\lambda_t, \xi_t) + \mathbb{E}\Big[\int_t^T (\mathcal{D}g)(\lambda_s, \xi_s) ds | \mathcal{F}_t\Big],$$

where $\mathcal{D}$ is the infinitesimal generator

$$\mathcal{D}g(\lambda, \xi) := \big(\xi + \beta(\mu - \lambda)\big)\partial_\lambda g(\lambda, \xi) - \beta\xi\partial_\xi g(\lambda, \xi) + \lambda\big(g(\lambda, \xi + \alpha) - g(\lambda, \xi)\big),$$

whenever these expectations are finite.

Proof. The exponential case comes from Proposition 2.1 in [10] and the Erlang case can be found in Proposition 1 in [9]. $\qquad\square$

Lemma 4.6. *Assume that H is an exponential or an Erlang Hawkes process (that is Condition (4.1) or (4.2) is in force for some parameters α, β). Then there is a positive constant C (depending only on the parameters α and β) such that for any $t \geq 0$*

$$\mathbb{E}[\lambda_t^2] \leq C.$$

Proof. We treat the two cases (4.1) or (4.2) separately.

- If $\Phi(u) = \alpha e^{-\beta u}$, $\alpha < \beta$

  According to [10], the intensity is a Markov process that satisfies the following Dynkin formula:
  $$\frac{d}{dt}\mathbb{E}[g(\lambda_t)] = \mathbb{E}[\beta(\mu - \lambda_t)g'(\lambda_t) + \lambda_t\big(g(\lambda_t + \alpha) - g(\lambda_t)\big)].$$

  This formula can be used to have the variance of the intensity λ_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 [9]) with an auxiliary process $(\xi_t)_{t \geq 0}$. The Dynkin formula for the vector process $(\lambda_t, \xi_t)_{t \geq 0}$ is

$$\frac{d}{dt}\mathbb{E}[g(\lambda_t, \xi_t)] = \mathbb{E}\Big[\big(\xi_t + \beta(\mu - \lambda_t)\big)\partial_\lambda g(\lambda_t, \xi_t) - \beta\xi_t\partial_\xi g(\lambda_t, \xi_t) + \lambda_t\big(g(\lambda_t, \xi_t + \alpha) - g(\lambda_t, \xi_t)\big)\Big].$$

After taking $g(\lambda, \xi) = \lambda^2$, $g(\lambda, \xi) = \xi^2$ and $g(\lambda, \xi) = \lambda\xi$, we obtain the following system

$$\begin{cases} \partial_t\mathbb{E}[\lambda_t^2] = -2\beta\mathbb{E}[\lambda_t^2] + 2\mathbb{E}[\lambda_t\xi_t] + 2\beta\mu\mathbb{E}[\lambda_t] \\ \partial_t\mathbb{E}[\xi_t^2] = -2\beta\mathbb{E}[\xi_t^2] + 2\alpha\mathbb{E}[\lambda_t\xi_t] + \alpha^2\mathbb{E}[\lambda_t] \\ \partial_t\mathbb{E}[\lambda_t\xi_t] = \alpha\mathbb{E}[\lambda_t^2] + \mathbb{E}[\xi_t^2] - 2\beta\mathbb{E}[\lambda_t\xi_t] + \beta\mu\mathbb{E}[\xi_t] \end{cases}. \qquad (4.3)$$

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

$$\mathbb{E}[\lambda_t^2] \leq C.$$

$\square$

Lemma 4.7. *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 $A_{1,3} = \frac{1}{T}\mathbb{E}\left[\left|\int_0^T \lambda_t \hat{M}_T^t dt\right|\right]$ and that $\hat{M}^t$ is defined in Notation 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 Φ. Then we make use of this estimate in the two particular cases of exponential and Erlang kernels (4.1)-(4.2).

Step 1 : a general estimate
In the following lines we provide an estimate for term $A_{1,3}$ for a general kernel Φ enjoying Assumptions 3.2 and 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

$$
\begin{aligned}
\mathbb{E}\left[\left|\int_0^T \lambda_t \hat{M}_T^t dt\right|^2\right] &= 2\mathbb{E}\left[\int_0^T \int_0^t \lambda_t \lambda_s \hat{M}_T^t \hat{M}_T^s\right] ds dt \\
&= 2\int_0^T \int_0^t \mathbb{E}[\lambda_t \lambda_s \mathbb{E}_t[\hat{M}_T^t \hat{M}_T^s]] ds dt \\
&= 2\int_0^T \int_0^t \mathbb{E}[\lambda_t \lambda_s \mathbb{E}_t[\hat{M}_T^t (\hat{M}_T^s - \hat{M}_t^s)]] ds dt \\
&= 2\int_0^T \int_0^t \mathbb{E}[\lambda_t \lambda_s \mathbb{E}_t\left[[\hat{M}^t, (\hat{M}^s - \hat{M}_t^s)]_T\right]] ds dt,
\end{aligned}
$$

where we have used the fact that for any t, $\hat{M}^t$ is a martingale with $\hat{M}_t^t = 0$. Hence

$$
\begin{aligned}
&\mathbb{E}_t[[\hat{M}^t, (\hat{M}^s - \hat{M}_t^s)]_T] \\
&= \mathbb{E}_t\left[\sum_{t<u\leq T} \Delta_u \hat{M}^t \Delta_u \hat{M}^s\right] \\
&= \mathbb{E}_t\left[\int_{(t,T]\times E} x\mathbf{1}_{\{\theta \leq \hat{\lambda}_u^t\}} N(du, d\theta, dx) \int_{(t,T]\times E} x\mathbf{1}_{\{\theta \leq \hat{\lambda}_u^s\}} N(du, d\theta, dx)\right]
\end{aligned}
$$

$$= \mathbb{E}_t \left[\int_{(t,T] \times E} x^2 \mathbf{1}_{\{\theta \le \hat{\lambda}_u^t\}} \mathbf{1}_{\{\theta \le \hat{\lambda}_u^s\}} du d\theta \nu(dx) \right]$$

$$= \vartheta^2 \, \mathbb{E}_t \left[\int_t^T \min(\hat{\lambda}_u^t, \hat{\lambda}_u^s) du \right].$$

The last entity can be bounded as follows

$$\mathbb{E}_t[[\hat{M}^t, (\hat{M}^s - \hat{M}_t^s)]_T] \le \vartheta^2 \int_t^T \min(\mathbb{E}_t[\hat{\lambda}_u^t], \mathbb{E}_t[\hat{\lambda}_u^s]) du.$$

Since the intensity λ of a Hawkes process is positive, multiplying by it preserves the inequality, leading to

$$A_{1,3} \le \frac{1}{T} \mathbb{E} \left[\left| \int_0^T \lambda_t \hat{M}_T^t dt \right|^2 \right]^{\frac{1}{2}}$$

$$\le \vartheta^2 \frac{\sqrt{2}}{\sqrt{T}} \left(\frac{1}{T} \int_0^T \int_0^t \mathbb{E} \left[\lambda_t \lambda_s \int_t^T \min(\mathbb{E}_t[\hat{\lambda}_u^t], \mathbb{E}_t[\hat{\lambda}_u^s]) du \right] ds dt \right)^{1/2} \tag{4.4}$$

$$=: \vartheta^2 \frac{\sqrt{2}}{\sqrt{T}} \left(\frac{1}{T} I_T \right)^{1/2}, \tag{4.5}$$

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 (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 λ and $\hat{\lambda}$ are known explicitly. In fact, given a starting time t, $\hat{\lambda}^t$ is defined for $u \ge t$ and it satisfies the following SDE (using once again Notation 3.9)

$$d\hat{\lambda}_u^t = (\alpha - \beta)\hat{\lambda}_u^t du + \alpha d\mathcal{M}_u^t,$$

with $\mathcal{M}_u^t = \hat{H}_u^t - \int_t^u \hat{\lambda}_s^t ds$, 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 α, β that may differ from line to line)

$$I_T \le C \int_0^T \int_0^t \mathbb{E}[\lambda_t \lambda_s \int_t^T \min(\alpha, \hat{\lambda}_t^s) e^{(\alpha-\beta)(u-t)} du] ds dt,$$

$$= C \int_0^T \int_0^t \mathbb{E}[\lambda_t \lambda_s \min(\alpha, \hat{\lambda}_t^s)] \int_t^T e^{(\alpha-\beta)(u-t)} du ds dt,$$

$$= C \int_0^T \int_0^t \mathbb{E}[\lambda_t \lambda_s \min(\alpha, \hat{\lambda}_t^s)] ds (1 - e^{(\alpha-\beta)(T-t)}) dt,$$

$$= C \int_0^T \int_0^t \mathbb{E}[\lambda_s \mathbb{E}_s[\lambda_t \min(\alpha, \hat{\lambda}_t^s)]] ds (1 - e^{(\alpha-\beta)(T-t)}) dt,$$

$$\leq C \int_0^T \int_0^t \mathbb{E}[\lambda_s \mathbb{E}_s[\lambda_t \hat{\lambda}_t^s]] ds (1 - e^{(\alpha-\beta)(T-t)}) dt.$$

Since $\hat{\lambda}_t^s$ starts at s, it is independent from $\mathcal{F}_s^N$ 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 4.6 entail :

$$
\begin{aligned}
I_T &\leq C \int_0^T \int_0^t \mathbb{E}[\lambda_s \mathbb{E}_s[\lambda_t]] e^{(\alpha-\beta)(t-s)} ds (1 - e^{(\alpha-\beta)(T-t)}) dt, \\
&= C \int_0^T \int_0^t \mathbb{E}[\lambda_s \lambda_t] e^{(\alpha-\beta)(t-s)} ds (1 - e^{(\alpha-\beta)(T-t)}) dt, \\
&\leq C \int_0^T \int_0^t \mathbb{E}[\lambda_s^2]^{\frac{1}{2}} \mathbb{E}[\lambda_t^2]^{\frac{1}{2}} e^{(\alpha-\beta)(t-s)} ds (1 - e^{(\alpha-\beta)(T-t)}) dt, \quad \text{Cauchy-Schwarz} \\
&\leq C \int_0^T \int_0^t e^{(\alpha-\beta)(t-s)} ds (1 - e^{(\alpha-\beta)(T-t)}) dt, \quad \text{using Lemma 4.6} \\
&= C \int_0^T (1 - e^{(\alpha-\beta)t})(1 - e^{(\alpha-\beta)(T-t)}) dt, \\
&= C \int_0^T 1 - e^{(\alpha-\beta)(T-t)} - e^{(\alpha-\beta)t} + e^{(\alpha-\beta)T} dt, \\
&\leq C(T + e^{(\alpha-\beta)T} - 1 + T e^{(\alpha-\beta)T}) \quad \text{since } \alpha - \beta < 0, \\
&\leq C \cdot T.
\end{aligned}
$$

Finally using (4.4) we conclude that $\quad A_{1,3} = O\left(\frac{1}{\sqrt{T}}\right).$

Step 3 : the Erlang case
We assume an Erlang kernel (4.2) $\Phi(u) = \alpha u e^{-\beta u}$ with $0 < \alpha < \beta^2$.
Even though the intensity λ is no longer a Markov process, it is possible to "Markovize" it by taking an auxiliary process ξ into account.
In the case of the shifted vanishing process $\hat{\lambda}^s$, $\hat{\xi}_u^s = \alpha e^{-\beta(u-s)} + \alpha \int_s^u e^{-\beta(u-v)} d\hat{H}_v^s$, and the process $(\hat{\lambda}^s, \hat{\xi}^s)$ follows the SDE:

$$
\begin{cases}
d\hat{\lambda}_u^s = -\beta \hat{\lambda}_u^s du + \hat{\xi}_u^s du \\
d\hat{\xi}_u^s = -\beta \hat{\xi}_u^s du + \alpha d\hat{H}_u^s.
\end{cases}
$$

This yields after solving the system

$$
\begin{aligned}
\begin{pmatrix} \hat{\lambda}_u^s \\ \hat{\xi}_u^s \end{pmatrix} = \frac{\sqrt{\alpha}}{2} &\left(\left(\int_s^u e^{(\sqrt{\alpha}-\beta)(s-v)} d\hat{\mathcal{M}}_v^s + 1 \right) e^{(\sqrt{\alpha}-\beta)(u-s)} \begin{pmatrix} 1 \\ \sqrt{\alpha} \end{pmatrix} \right. \\
&\left. - \left(\int_s^u e^{-(\sqrt{\alpha}+\beta)(s-v)} d\hat{\mathcal{M}}_v^s + 1 \right) e^{-(\sqrt{\alpha}+\beta)(u-s)} \begin{pmatrix} 1 \\ -\sqrt{\alpha} \end{pmatrix} \right).
\end{aligned}
$$

And finally

$$
\begin{pmatrix} \mathbb{E}_t[\hat{\lambda}_u^s] \\ \mathbb{E}_t[\hat{\xi}_u^s] \end{pmatrix} = \frac{1}{2}\left(\hat{\lambda}_t^s + \frac{\hat{\xi}_t^s}{\sqrt{\alpha}}\right) e^{(\sqrt{\alpha}-\beta)(u-t)} \begin{pmatrix} 1 \\ \sqrt{\alpha} \end{pmatrix} + \frac{1}{2}\left(\hat{\lambda}_t^s - \frac{\hat{\xi}_t^s}{\sqrt{\alpha}}\right) e^{-(\sqrt{\alpha}+\beta)(u-t)} \begin{pmatrix} 1 \\ -\sqrt{\alpha} \end{pmatrix}. \quad (4.6)
$$

Once again we make use of the general estimate (4.4) and estimate the quantity

$$I_T = \int_0^T \int_0^t \mathbb{E}[\lambda_t \lambda_s \int_t^T \min(\mathbb{E}_t[\hat{\lambda}_u^t], \mathbb{E}_t[\hat{\lambda}_u^s]) du] ds dt,$$

30

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.

$$
\begin{aligned}
I_T \leq & C \int_0^T \int_0^t \mathbb{E}[\lambda_t \lambda_s \int_t^T \frac{1}{2}(\hat{\lambda}_t^s + \frac{\hat{\xi}_t^s}{\sqrt{\alpha}}) e^{(\sqrt{\alpha}-\beta)(u-t)} + \frac{1}{2}(\hat{\lambda}_t^s - \frac{\hat{\xi}_t^s}{\sqrt{\alpha}}) e^{-(\sqrt{\alpha}+\beta)(u-t)} du ds dt, \\
= & C \int_0^T \int_0^t \mathbb{E}[\lambda_t \lambda_s (\hat{\lambda}_t^s + \frac{\hat{\xi}_t^s}{\sqrt{\alpha}})] \int_t^T e^{(\sqrt{\alpha}-\beta)(u-t)} du ds dt \\
& + C \int_0^T \int_0^t \mathbb{E}[\lambda_t \lambda_s (\hat{\lambda}_t^s - \frac{\hat{\xi}_t^s}{\sqrt{\alpha}})] \int_t^T e^{-(\sqrt{\alpha}+\beta)(u-t)} du ds dt, \\
\leq & C \int_0^T (1 - e^{(\sqrt{\alpha}-\beta)(T-t)}) \int_0^t \mathbb{E}[\lambda_t \lambda_s (\hat{\lambda}_t^s + \frac{\hat{\xi}_t^s}{\sqrt{\alpha}})] ds dt \\
& + C \int_0^T (1 - e^{-(\sqrt{\alpha}+\beta)(T-t)}) \int_0^t \mathbb{E}[\lambda_t \lambda_s (\hat{\lambda}_t^s - \frac{\hat{\xi}_t^s}{\sqrt{\alpha}})] ds dt, \\
:= & C(I_T^+ + I_T^-).
\end{aligned}
$$

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 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 4.6, 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

$$
\begin{aligned}
I_T^+ & \leq \int_0^T (1 - e^{(\sqrt{\alpha}-\beta)(T-t)}) \int_0^t e^{(\sqrt{\alpha}-\beta)(t-s)} ds dt, \\
& \leq \int_0^T (1 - e^{(\sqrt{\alpha}-\beta)(T-t)})(1 - e^{(\sqrt{\alpha}-\beta)t}) dt,
\end{aligned}
$$

31

$$\leq \int_0^T 1 - e^{(\sqrt{\alpha}-\beta)t} - e^{(\sqrt{\alpha}-\beta)(T-t)} + e^{(\sqrt{\alpha}-\beta)T}dt,$$
$$= O(T).$$

Following the same lines we get for $|I_T^-|$:

$$I_T^- = \int_0^T (1 - e^{-(\sqrt{\alpha}+\beta)(T-t)}) \int_0^t \mathbb{E}[\lambda_s \mathbb{E}_s[\lambda_t(\hat{\lambda}_t^s - \frac{\hat{\xi}_t^s}{\sqrt{\alpha}})]]dsdt,$$
$$= \int_0^T (1 - e^{-(\sqrt{\alpha}+\beta)(T-t)}) \int_0^t \mathbb{E}[\lambda_s \mathbb{E}_s[\lambda_t]\mathbb{E}[\hat{\lambda}_t^s - \frac{\hat{\xi}_t^s}{\sqrt{\alpha}}]]dsdt,$$
$$= -\int_0^T (1 - e^{-(\sqrt{\alpha}+\beta)(T-t)}) \int_0^t \mathbb{E}[\lambda_t \lambda_s]\sqrt{\alpha}e^{-(\sqrt{\alpha}+\beta)(t-s)}dsdt.$$

So $|I_T^-| = O(T)$, and since $I_T \leq I_T^+ + |I_T^-|$, we conclude using once again (4.4) that

$$A_{1,3} = O\left(\frac{1}{\sqrt{T}}\right).$$

$\square$

Lemma 4.8. *Set* $Y_T = \frac{H_T - \int_0^T \mathbb{E}[\lambda_t]dt}{\sqrt{T}}$ *and* $F_T = \frac{H_T - \int_0^T \lambda_t dt}{\sqrt{T}}$.
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

$$\gamma = \frac{1}{1 - \|\Phi\|_1} = \begin{cases} \frac{1}{1-\alpha/\beta}, & if \quad \Phi(u) = \alpha e^{-\beta u} \\ \frac{1}{1-\alpha/\beta^2}, & if \quad \Phi(u) = \alpha u e^{-\beta u} \end{cases}$$

and

$$\mathfrak{R}_T = \begin{cases} \frac{\mathbb{E}[\lambda_T]-\lambda_T}{\beta\sqrt{T}}, & if \quad \Phi(u) = \alpha e^{-\beta u} \\ \frac{\mathbb{E}[\lambda_T]-\lambda_T}{\beta\sqrt{T}} + \frac{\mathbb{E}[\xi_T]-\xi_T}{\beta^2\sqrt{T}}, & if \quad \Phi(u) = \alpha u e^{-\beta u} \end{cases}$$

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:

$$d\lambda_t = \beta(\mu - \lambda_t)dt + \alpha dH_t.$$

We start by integrating the SDE between 0 and T:

$$\int_0^T d\lambda_t = \int_0^T \beta(\mu - \lambda_t)dt + \alpha dH_t,$$

$$\lambda_T - \mu = \beta\mu T - \beta \int_0^T \lambda_t dt + \alpha H_T.$$

After rearranging the terms, F_T can be put under the form:

$$F_T = \frac{\lambda_T - \mu + (\beta - \alpha)H_T - \beta\mu T}{\beta\sqrt{T}}.$$

When it comes to Y_T, we start by taking the expected value of the SDE:

$$d\mathbb{E}[\lambda_t] = \beta(\mu - \mathbb{E}[\lambda_t])dt + \alpha d\mathbb{E}[H_t],$$
$$= \beta(\mu - \mathbb{E}[\lambda_t])dt + \alpha\mathbb{E}[\lambda_t]dt,$$

and after integrating with respect to time:

$$\mathbb{E}[\lambda_T] - \mu = \beta\mu T + (\alpha - \beta) \int_0^T \mathbb{E}[\lambda_t]dt,$$

$$= \beta\mu T + (\alpha - \beta)(\int_0^T \mathbb{E}[\lambda_t]dt - H_T) + (\alpha - \beta)H_T.$$

And finally:

$$Y_T = \frac{\mathbb{E}[\lambda_T] - \mu + (\beta - \alpha)H_T - \beta\mu T}{(\beta - \alpha)\sqrt{T}}.$$

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 $R_T = \frac{\mathbb{E}[\lambda_T]-\lambda_T}{\beta\sqrt{T}}$.

Case 2 : the kernel is an Erlang function
In this case $\|\Phi\|_1 = \frac{\alpha}{\beta^2}$ and the vector (λ_t, ξ_t) where $\xi_t = \int_{[0,t)} \alpha e^{-\beta(t-s)}dH_s$ follows satisfies the following SDE:

$$\begin{cases} d\lambda_t = \xi_t dt + \beta(\mu - \lambda_t)dt, \\ \\ d\xi_t = -\beta\xi_t dt + \alpha dH_T. \end{cases}$$

We integrate the system:

$$\begin{cases} \lambda_T - \mu = \int_0^T \xi_t dt + \beta\mu T - \beta \int_0^T \lambda_t dt, \\ \\ \xi_T = -\beta \int_0^T \xi_t dt + \alpha H_T. \end{cases}$$

We eliminate $\int_0^T \xi_t dt$ in the system to obtain:

$$\lambda_T - \mu = \frac{\alpha}{\beta}H_T - \frac{1}{\beta}\xi_T + \mu\beta T - \beta \int_0^T \lambda_t dt,$$

$$= \frac{\alpha}{\beta} H_T - \frac{1}{\beta} \xi_T + \mu \beta T + \beta \sqrt{T} F_T - \beta H_T,$$

and after re-arranging the terms:

$$\beta \sqrt{T} F_T = \lambda_T - \mu + \frac{\beta^2 - \alpha}{\beta} H_T + \frac{1}{\beta} \xi_T - \mu \beta T.$$

When it comes to Y_T we take the expected value of the SDE:

$$\begin{cases} d\mathbb{E}[\lambda_t] = \big(\mathbb{E}[\xi_t] + \beta(\mu - \mathbb{E}[\lambda_t])\big)dt, \\ d\mathbb{E}[\xi_t] = -\beta\mathbb{E}[\xi_t]dt + \alpha\mathbb{E}[\lambda_t]dt. \end{cases}$$

After taking the integral and eliminating $\int_0^T \mathbb{E}[\xi_t]dt$ in the system:

$$\mathbb{E}[\lambda_T] - \mu = -\frac{1}{\beta}\mathbb{E}[\xi_T] + \frac{\alpha}{\beta}\int_0^T \mathbb{E}[\lambda_t]dt + \beta\mu T - \beta\int_0^T \mathbb{E}[\lambda_t]dt,$$

$$= -\frac{1}{\beta}\mathbb{E}[\xi_T] + \beta\mu T + (\frac{\alpha}{\beta} - \beta)\int_0^T \mathbb{E}[\lambda_t]dt,$$

$$= -\frac{1}{\beta}\mathbb{E}[\xi_T] + \beta\mu T + \frac{\beta^2 - \alpha}{\beta}\sqrt{T}Y_T + \frac{\alpha - \beta^2}{\beta}H_T,$$

which yields after re-arranging the terms:

$$\frac{\beta^2 - \alpha}{\beta}\sqrt{T}Y_T = \mathbb{E}[\lambda_T] - \mu + \frac{\beta^2 - \alpha}{\beta}H_T + \frac{1}{\beta}\mathbb{E}[\xi_T] - \mu\beta T.$$

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 $R_T = \frac{\mathbb{E}[\lambda_T] - \lambda_T}{\beta\sqrt{T}} + \frac{\mathbb{E}[\xi_T] - \xi_T}{\beta^2\sqrt{T}}$. $\qquad\qquad \square$

References

[1] E. Bacry, S. Delattre, M. Hoffmann, and J.F. Muzy. Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Processes and their Applications*, 123(7):2475 – 2499, 2013. A Special Issue on the Occasion of the 2013 International Year of Statistics.

[2] E. Bacry, I. Mastromatteo, and J.F. Muzy. Hawkes processes in finance. *Market Microstructure and Liquidity*, 1(01):1550005, 2015.

[3] E. Besancon. *Speed of convergence of diffusion approximations*. PhD thesis, Télécom ParisTech, https://tel.archives-ouvertes.fr/tel-03128781, 2020.

[4] Y. Bessy-Roland, A. Boumezoued, and C. Hillairet. Multivariate hawkes process for cyber insurance. *Annals of Actuarial Science*, 2020.

[5] C. Bordenave and G. L. Torrisi. Large deviations of Poisson cluster processes. *Stoch. Models*, 23(593–625), 2007.

[6] P. Brémaud and L. Massoulié. Stability of nonlinear Hawkes processes. *Annals of Probability*, 24(3):1563–1588, 1996.

[7] M. Costa, C. Graham, L. Marsalle, and V. C. Tran. Renewal in hawkes processes with self-excitation and inhibition. *Advances in Applied Probability*, 52(3):879–915, 2020.

[8] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes*. Springer. New York, 1988.

[9] A. Duarte, E. Löcherbach, and G. Ost. Stability, convergence to equilibrium and simulation of non-linear Hawkes processes with memory kernels given by the sum of Erlang kernels. *ESAIM: Probability and Statistics*, 23:770–796, 2019.

[10] E. Errais, K. Giesecke, and L. R. Goldberg. Affine point processes and portfolio credit risk. *SIAM Journal on Financial Mathematics*, 1(1):642–665, 2010.

[11] X Gao and L. Zhu. Functional central limit theorems for stationary Hawkes processes and application to infinite-server queues. *Queueing Syst.*, 90:161–206, 2018.

[12] A. G. Hawkes. Spectra of some self-excited and mutually exciting point process. *Biometrika*, 1:83–90, 1971.

[13] C. Hillairet, A. Réveillac, and M. Rosenbaum. An expansion formula for Hawkes processes and application to cyber-insurance derivatives. Preprint., 2020.

[14] U. Horst and W. Xu. Functional limit theorems for marked Hawkes point measures. *arXiv preprint arXiv:1908.06703*, 2019.

[15] U. Horst and W. Xu. A scaling limit for limit order books driven by Hawkes processes. *SIAM J. Financial Mathematics*, 10(2):350–393, 2019.

[16] T. Jaisson and M. Rosenbaum. Limit theorems for nearly unstable Hawkes processes. *The Annals of Applied Probability*, 25(2):600–631, 2015.

[17] R. Murr. Characterization of infinite divisibility by duality formulas. application to Lévy processes and random measures. *Stochastic Processes and their Applications*, 123:1729–1749, 2013.

[18] I. Nourdin and G. Peccati. Stein's method on Wiener chaos Stein's method on Wiener chaos. *Probab. Theory Related Fields*, 145:75–118, 2009.

[19] I. Nourdin and G. Peccati. *Normal approximations with Malliavin calculus*, volume 192 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2012. From Stein's method to universality.

[20] D. Nualart and J. Vives. Anticipative calculus for the Poisson process based on the fock space. *Séminaire de probabilités de Strasbourg*, 24:154–165, 1990.

[21] Y. Ogata. On Lewis' simulation method for point processes. *IEEE Transactions of Information Theory*, 27:23–31, 1981.

[22] G. Peccati, J. L. Solé, M. S. Taqqu, and F. Utzet. Stein's method and Normal approximation of Poisson functionals. *The Annals of Probability*, 38(2):443–478, 2010.

[23] J. Picard. Formules de dualité sur l'espace de Poisson. *Ann. Inst. H. Poincaré Probab. Statist.*, 32(4):509–548, 1996.

[24] N. Privault. *Stochastic analysis in discrete and continuous settings with normal martingales*, volume 1982 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2009.

[25] N. Privault. Stein approximation for multidimensional Poisson random measures by third cumulant expansions. Preprint., 06 2018.

[26] J. L. Solé, F. Utzet, and J. Vives. Canonical Lévy process and Malliavin calculus. *Stochastic Processes and their Applications*, 117:165–187, 2007.

[27] C. M. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. 6th Berkeley Sympos. math. Statist. Probab., Univ. Calif. 1970*, 2(583–602), 1972.

[28] G. L. Torrisi. Gaussian approximation of nonlinear Hawkes processes. *The Annals of Applied Probability*, 26(4):2106–2140, 2016.

[29] L. Zhu. Central limit theorems for non-linear Hawkes processes. *J. Appl. Probab.*, 50(760–771), 2013.

[30] L. Zhu. Limit theorems for a Cox-Ingersoll-Ross process with Hawkes jumps. *J. Appl. Probab.*, 51:699–712, 2014.