Normal approximation of compound Hawkes functionals

Mahmoud Khabou[*] Nicolas Privault [†] Anthony Réveillac[‡]

January 16, 2023

Abstract

We derive quantitative bounds in the Wasserstein distance for the approximation of stochastic integrals with respect to Hawkes processes by a normally distributed random variable. In the case of deterministic and non-negative integrands, our estimates involve only the third moment of integrand in addition to a variance term using a square norm of the integrand. As a consequence, we are able to observe a "third moment phenomenon" in which the vanishing of the first cumulant can lead to faster convergence rates. Our results are also applied to compound Hawkes processes, and improve on the current literature where estimates may not converge to zero in large time, or have been obtained only for specific kernels such as the exponential or Erlang kernels.

Key words: Hawkes processes; Stein method; normal approximation; Malliavin calculus.

Mathematics Subject Classification (2020): 60G55, 60F05, 60G57, 60H07.

1 Introduction

Nourdin and Peccati [NP09] opened the way to a new methodology mixing Stein's method and the Malliavin calculus, to provide bounds on the distance between the distribution of a functional of a Gaussian field and a target Gaussian distribution. This

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

[†]School of Physical and Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, 637371 Singapre. Email: nprivault@ntu.edu.sg

[‡]INSA de Toulouse, IMT UMR CNRS 5219, Université de Toulouse, 135 avenue de Rangueil 31077 Toulouse Cedex 4 France. Email: anthony.reveillac@insa-toulouse.fr

analysis, which relies on a specific Gaussian structure, has been successfully transferred to the Gaussian approximation of Poisson functionals in [PSTU10]. Since then, several developments of the initial result of [PSTU10] have been obtained, for instance, Stein Gaussian approximation bounds have been obtained in [Pri18] in terms of the third cumulants for Poisson functionals expressed as the divergence of an adapted process with respect to an homogeneous Poisson process. Another development important for our analysis has been presented in [Tor16] for counting processes with stochastic intensity (including the Hawkes process) using the so-called Poisson imbedding representation, see[1] [BM96]. This technique, also known as the "Thinning Algorithm", allows one to represent a counting process and its intensity process as a solution to an SDE driven by an auxiliary Poisson random measure, and to adapted the general methodology of [PSTU10] to this framework. Following here the path of [Tor16] in the case of a linear Hawkes process H, i.e. a counting process with intensity process $\lambda := (\lambda_t)_{t\geq 0}$ given as

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

with $\mu > 0$, $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ and $\|\phi\|_1 < 1$, a specific Malliavin calculus for Hawkes processes have been developed in [HRR20, HHKR21], based on Relation (2.8) below on the simplification of the Malliavin integration by parts. As a consequence, a new bound has been obtained in [HHKR21] for the Gaussian approximation of Hawkes functionals.

In [BDHM13], a functional convergence result has been obtained for linear Hawkes processes, implying the (non-quantitative) convergence in distribution

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

see Lemma 7 therein, with $\sigma^2 := \mu/(1 - \|\phi\|_1)$.

In this paper, we propose to quantify this convergence in the Wasserstein distance. In Theorem 3.1 of [Tor16], see also Relations (5.2) and (5.4) therein, the estimates

$$d_W\left(\frac{H_T - \int_0^T \lambda_s ds}{\sqrt{T}}, \mathcal{N}(0,\sigma^2)\right) \leq B(T), \quad \text{with} \quad B(T) \geq \sqrt{\frac{8}{\pi}} \|\phi\|_1 \frac{2 - \|\phi\|_1}{1 - \|\phi\|_1}$$

[1]The term used in this reference is indeed "Poisson imbedding" whereas "Poisson embedding" refers to a specific technique in analysis based on Poisson-type PDEs.

have been derived, however they do not converge to 0 as T tends to $+\infty$, see Remark 3.5 below. Related bounds have been derived independently in [HHKR21, Theorem 3.4] for Hawkes processes, with in particular

$$d_W\left(\frac{H_T - \int_0^T \lambda_s ds}{\sqrt{T}}, \mathcal{N}(0,\sigma^2)\right) \le \frac{C_{\mu,\phi}}{\sqrt{T}} + R_T,$$

see Theorems 3.10 and 3.12 therein, where R_T is an additional term involving the Malliavin derivative of F_T, see also Proposition 2.10 below.

In case ϕ is the exponential kernel $\phi(x) := \alpha e^{-\beta x}$, $\alpha < \beta$, or the Erlang kernel $\phi(x) := \alpha x e^{-\beta x}$, $\alpha < \beta^2$, the remainder term R_T can be bounded to obtain more accurate bounds of the form

$$d_W\left(\frac{H_T - \int_0^T \lambda_s ds}{\sqrt{T}}, \mathcal{N}(0,\sigma^2)\right) \le \frac{\widetilde{C}_{\mu,\phi}}{\sqrt{T}}. \tag{1.1}$$

In this paper, we extend those results by deriving bounds of the form (1.1) for Hawkes processes with general kernel ϕ satisfying the condition $\|\phi\|_1 < 1$, see Theorem 3.4, where $C_{\mu,\phi} > 0$ is a constant and σ^2 an explicit asymptotic variance depending on μ, ϕ.

For this, in Theorem 3.1 we improve the bounds of [Tor16] and [HHKR21] for Hawkes functionals of the form $\int_0^\infty z(t)(dH_t - \lambda_t dt)$, where $z(t)$ is a deterministic function. In particular, in Theorem 3.1-(ii) we provide an estimate involving only the third moment of $\int_0^\infty z(t)(dH_t - \lambda_t dt)$ when $z(t)$ is deterministic and non-negative, and an estimate on a modified second moment of $z(t)$ where the Malliavin derivative is not involved, see (3.3) and the discussion in Remark 3.2. A discussion on the comparison of results and methods with the papers [Tor16] and [HHKR21] is presented in Remark 3.5.

In addition, these results are presented for a generalization of the Hawkes process, that is the compound Hawkes process S given by

$$S_t := \sum_{i=1}^{H_t} X_i, \quad t \ge 0,$$

where the random variables $(X_i)_i$ are independent and identically distributed (i.i.d.) square integrable random variables independent of H. Furthermore, in the spirit of

Theorem 3.7 in [CGS11], in Theorem 3.7 we obtain faster rates of convergence for the compound Hawkes process S when the third cumulant of the jump sizes on the random variables $(X_i)_i$ vanishes in the framework of the "third moment phenomenon", see Section 4.8 of [CGS11]. Finally, we provide an alternative version of our quantitative limit theorems as Theorem 3.8.

We proceed as follows. First, in Section 2 we present the main elements of stochastic analysis on the Poisson space and we recall the approach developed in [HRR20, HHKR21] regarding the linear Hawkes process. Main results are collected in Section 3. Proofs and technical lemmata are collected in Section 4.

2 Notations and preliminaries

For E a topological space, we set $\mathcal{B}(E)$ the σ-algebra of Borel sets. We denote by dt the Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

2.1 Stochastic analysis on the Poisson space

The notation and definitions stated in this section can be found for instance in [Pic96] or [Pri09]. Let ν be a Borel measure on $\mathbb{R}$ with $\nu(\mathbb{R}) = 1$ and $\nu(\{0\}) = 0$, and consider 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 $\mathcal{F}_\infty^N$ 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

$$N([0,t] \times [0,\theta] \times (-\infty, y])(\omega) := \omega([0,t] \times [0,\theta] \times (-\infty, x]), \quad (t,\theta,x) \in \mathbb{R}_+^2 \times \mathbb{R},$$

is an homogeneous Poisson process with intensity measure $dt \otimes d\theta \otimes \nu$, that is, for any $(t, \theta, x) \in [0,T] \times \mathbb{R}_+ \times \mathbb{R}$, $N([0,t] \times [0,\theta] \times (-\infty, x])$ is a Poisson random variable with intensity $\theta\, t\, \nu((-\infty, x])$. We also let $\mathbb{F}^N := (\mathcal{F}_t^N)_{t \geq 0}$ denote the natural history of N, that is

$$\mathcal{F}_t^N := \sigma(N(\mathcal{T} \times B),\ \mathcal{T} \subset \mathcal{B}([0,t]),\ B \in \mathcal{B}(\mathbb{R}_+ \times \mathbb{R})), \qquad t \geq 0,$$

hence $\mathcal{F}^N_\infty$ coincides with $\lim_{t\to+\infty}\mathcal{F}^N_t$. The expectation with respect to $\mathbb{P}^N$ is denoted by $\mathbb{E}[\,\cdot\,]$, and the conditional expectation knowing $\mathcal{F}^N_t$ is denoted by $\mathbb{E}_t[\,\cdot\,]$.

We also write $L^0\big(\Omega,\mathcal{F}^N_\infty,\mathbb{P}\big)$ for the space of $\mathcal{F}^N_\infty$-measurable random variables, and for any $p\geq 1$ we let $L^p\big(\Omega,\mathcal{F}^N_\infty,\mathbb{P}\big)$ the subspace made of $F\in L^0\big(\Omega,\mathcal{F}^N_\infty,\mathbb{P}\big)$ such that $\mathbb{E}[|F|^p]<+\infty$.

Next, we introduce the stochastic integral with respect to the Poisson point process N, which will be related to the gradient operator D defined below.

Definition 2.1. *We set*

$$\mathcal{P}^N := \left\{\rho := (\rho_{(t,\theta,x)})_{(t,\theta,x)\in\mathbb{R}^3_+}\ (\mathcal{F}^N_t)_{t\geq 0}\text{-}predictable\right\}.$$

For $\rho := (\rho_{(t,\theta,x)})_{(t,\theta,x)\in\mathbb{R}^3_+}\in\mathcal{P}^N_2$ *we define its divergence*

$$\delta^N(\rho) := \int_{\mathbb{R}^2_+\times\mathbb{R}} \rho_{(t,\theta,x)}\left(N(dt,d\theta,dx) - dtd\theta\nu(dx)\right),$$

which belongs to $L^2\big(\Omega,\mathcal{F}^N_\infty,\mathbb{P}\big)$ *if* $\mathbb{E}\left[\int_{\mathbb{R}^3_+}|\rho_{(t,\theta,x)}|^2 dtd\theta\nu(dx)\right]<+\infty$.

Next, we introduce the Malliavin derivative, first with respect to the Poisson point process N, using the adding point operator defined below. For any $A\in\mathcal{B}(\mathbb{R}_+\times\mathbb{R}_+\times\mathbb{R})$, we let

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

Definition 2.2 (Adding point operator)**.** *We define for* (t,θ,x) *in* $\mathbb{R}_+\times\mathbb{R}_+\times\mathbb{R}$ *the measurable maps*

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

where we let

$$(\varepsilon^+_{(t,\theta,x)}(\omega))(A) := \omega(A\setminus(t,\theta,x)) + \mathbf{1}_A(t,\theta,x), \quad A\in\mathcal{B}(\mathbb{R}_+\times\mathbb{R}_+\times\mathbb{R}).$$

We note that for any $\mathcal{F}^N_t$-measurable random variable F, $t\geq 0$, we have

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

for all $v>t$ and $(\theta,x)\in\mathbb{R}_+\times\mathbb{R}$.

Definition 2.3 (Malliavin derivative)**.** *For F in $L^0\big(\Omega, \mathcal{F}^N_\infty, \mathbb{P}\big)$, we define the Malliavin derivative DF of F as*

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

We conclude this section with the integration by parts formula on the Poisson space, see *e.g.* [Pic96] or [Pri09].

Proposition 2.4. *Let F be in $L^0\big(\Omega, \mathcal{F}^N_\infty, \mathbb{P}\big)$ and ρ be in $\mathcal{P}^N$. We have*

$$\mathbb{E}\big[F\delta^N(\rho)\big] = \mathbb{E}\left[\int_{\mathbb{R}^2_+\times\mathbb{R}} \rho_{(t,\theta,x)} D_{(t,\theta,x)}F dt d\theta \nu(dx)\right], \tag{2.1}$$

provided that $F\delta^N(\rho) \in L^p\big(\Omega, \mathcal{F}^N_\infty, \mathbb{P}\big)$ and $\mathbb{E}\left[\int_{\mathbb{R}^2_+\times\mathbb{R}} \rho_{(t,\theta,x)} D_{(t,\theta,x)}F dt d\theta \nu(dx)\right] < +\infty$.

Remark 2.5. *Let F in $L^0\big(\Omega, \mathcal{F}^N_\infty, \mathbb{P}\big)$. The definition of the Malliavin derivative together with the the relation*

$$a^2 - b^2 = (a-b)^2 + 2b(a-b), \quad \forall (a,b) \in \mathbb{R}^2,$$

entail that DF^2 rewrites as

$$D_{(t,\theta,x)}F^2 = F^2 \circ \varepsilon^+_{(t,\theta,x)} - F^2 = |D_{(t,\theta,x)}F|^2 + 2FD_{(t,\theta,x)}F, \quad (t,\theta,x) \in \mathbb{R}^2_+ \times \mathbb{R}. \tag{2.2}$$

2.2 Stochastic analysis for the compound Hawkes process

We first recall the definition of a Hawkes process.

Definition 2.6 (Standard Hawkes process, [Haw71])**.** *Let $\mu > 0$ and $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ be a bounded non-negative function with $\|\phi\|_1 := \int_0^\infty \phi(u)du < 1$. The standard Hawkes process $H := (H_t)_{t\geq 0}$ with parameters μ and ϕ is the counting process such that*

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

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

$$\lambda_t := \mu + \int_{(0,t)} \phi(t-s)dH_s, \quad t \geq 0, \tag{2.3}$$

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

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

Note that the stochastic integral in (2.3) is defined pathwise, i.e.

$$\int_{(0,t)} \phi(t-s)dH_s := \sum_{0<s<t} \phi(t-s)\mathbf{1}_{\{\Delta_s H=1\}},$$

where the sum is well defined and finite $\mathbb{P}$-a.s. for every t, where we used the notation $\Delta_s H := H_s - H_{s-}$. This definition can be generalized as follows.

Definition 2.7 (Compound Hawkes process)**.** *Consider a Hawkes process H with parameters $\mu > 0$ and $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ bounded non-negative with $\|\phi\|_1 < 1$. Given $(X_i)_{i\geq 1}$ an i.i.d sequence of random variables, independent of H, with common distribution ν, the process*

$$S_t := \sum_{i=1}^{H_t} X_i, \quad t \geq 0, \tag{2.4}$$

is called a compound Hawkes process.

Our approach uses the now classical construction of the Hawkes process by "thinning" or "Poisson embedding" as the unique solution to an SDE with respect to a Poisson random measure N, see *e.g.* [Oga81, DVJ88, BM96, CGMT20] and references therein. We refer to [HRR20, Theorem 3.3] for a precise statement on the uniqueness of solutions to the SDE (2.5). Here, we set $\mathbb{F}^H := (\mathcal{F}_t^H)_{t\geq 0}$ (respectively $\mathbb{F}^S := (\mathcal{F}_t^S)_{t\geq 0}$) the natural filtration of H (respectively of S) and $\mathcal{F}_\infty^H := \lim_{t\to+\infty} \mathcal{F}_t^H$ (respectively $\mathcal{F}_\infty^S := \lim_{t\to+\infty} \mathcal{F}_t^S$), and we have $\mathcal{F}_t^H \subset \mathcal{F}_t^S \subset \mathcal{F}_t^N$ as H is completely determined by the jump times of H, which coincide with those of S.

Theorem 2.8 (See Theorem 3.3 of [HRR20])**.** *Let $\mu > 0$ and $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\|\phi\|_1 < 1$. The system of stochastic differential equations*

$$\begin{cases} S_t = \int_{(0,t]\times\mathbb{R}_+\times\mathbb{R}} x\mathbf{1}_{\{\theta\leq\lambda_s\}} N(ds, d\theta, dx), \quad t \geq 0, \\ H_t = \int_{(0,t]\times\mathbb{R}_+\times\mathbb{R}} \mathbf{1}_{\{\theta\leq\lambda_s\}} N(ds, d\theta, dx), \quad t \geq 0, \\ \lambda_t = \mu + \int_{(0,t)} \phi(t-u)dH_u, \quad t \geq 0. \end{cases} \tag{2.5}$$

admits a unique solution (X, H, λ) with H (resp. λ) $\mathbb{F}^N$-adapted (resp. $\mathbb{F}^N$-predictable), where (H, λ) is a Hawkes process in the sense of Definition 2.6, and S is a compound Hawkes process in the sense of Definition 2.7.

We note that when $\nu(dx) = \delta_1(dx)$ equals the Dirac measure concentrated at $x = 1$, i.e. $X_i \equiv 1$ in (2.4), then $S \equiv H$.

Notation 2.9. *We let $\mathcal{Z} := (\mathcal{Z}_{(t,\theta)})_{(t,\theta)\in\mathbb{R}_+^2}$ denote the stochastic process defined as*

$$\mathcal{Z}_{(t,\theta)} := \mathbf{1}_{\{\theta \leq \lambda_t\}}, \quad (t,\theta) \in \mathbb{R}_+^2. \tag{2.6}$$

In this paper we consider stochastic integrals

$$\begin{aligned} F &= \delta^N(Z\mathcal{Z}) \\ &= \int_{\mathbb{R}_+^2\times\mathbb{R}} Z_{(t,x)}\mathcal{Z}_{(t,\theta)} \left(N(dt,d\theta,dx) - dtd\theta\nu(dx)\right) \\ &= \int_{\mathbb{R}_+\times\mathbb{R}} Z_{(t,x)} \left(dH_t - \lambda_t dt\nu(dx)\right) \end{aligned}$$

against the (compensated) Hawkes process, with $Z := (Z_{(t,x)})_{(t,x)\in\mathbb{R}_+\times\mathbb{R}}$ an element of $\mathcal{P}_2^N$. Most of our analysis will be carried out for a deterministic Z.

We now specify the Malliavin derivative and the integration by parts formula (2.1) for functionals of the Hawkes process (the Hawkes itself H or the compound Hawkes process X). For this, we note that by definition of H, any jump of N at an atom (t,θ,x) turns out to be a jump of H if and only if $\theta \in [0,\lambda_t]$, as stated in the next proposition.

Proposition 2.10 (Proposition 2.16 [HHKR21])**.** *For any $\mathcal{F}_\infty^H$-measurable random variable F we have*

$$D_{(t,0,x)}F = D_{(t,\theta,x)}F, \quad \theta \in [0,\lambda_t],\ t \geq 0,\ x \in \mathbb{R},\ \mathbb{P}\text{−}a.s..$$

In view of Proposition 2.10, for any $\mathcal{F}_\infty^H$-measurable random variable F, we set

$$D_{(t,x)}F := D_{(t,0,x)}F, \quad t \geq 0,\ x \in \mathbb{R}. \tag{2.7}$$

Next, we state an integration by part formula for functionals of Hawkes processes. We recall below the integration by parts formula obtained in a particular case of [HHKR21] for the Hawkes process.

Theorem 2.11. *(Theorem 2.20 in [HHKR21]) Let $\mathcal{Z} := (\mathcal{Z}_{(t,\theta)})_{(t,\theta)\in\mathbb{R}_+^2}$ be the stochastic process defined in (2.6) and $Z := (Z_{(t,x)})_{(t,x)\in\mathbb{R}_+\times\mathbb{R}}$ be a $\mathbb{F}^N$-predictable process* *satisfying*

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

Then, for any $F \in L^2\big(\Omega^N, \mathcal{F}^N_\infty, \mathbb{P}\big)$ *we have*

$$\mathbb{E}\left[F\delta^N(Z\boldsymbol{1}_{\{\theta\leq\lambda_\cdot\}})\right] = \mathbb{E}\left[\int_{\mathbb{R}_+\times\mathbb{R}} \lambda_t Z_{(t,x)} D_{(t,x)} F dt\nu(dx)\right]. \tag{2.8}$$

We conclude this section with a commutation property for the operators D and δ^N.

Lemma 2.12. *Let* $z := (z(t,x))_{(t,x)\in\mathbb{R}_+\times\mathbb{R}} \in L^2(\mathbb{R}_+\times\mathbb{R}, dt\otimes\nu)$ *and consider* $\mathcal{Z} := (\mathcal{Z}_{(t,\theta)})_{(t,\theta)\in\mathbb{R}^2_+}$ *given in* (2.6)*. We have*

$$D_{(t,x)}\delta^N(z\mathcal{Z}) = z(t,x) + \delta^N\big(z\widehat{\mathcal{Z}}^t\big), \quad t\geq 0,\ x\geq 0, \tag{2.9}$$

where

$$\widehat{\mathcal{Z}}^t_{(r,\theta)} := \boldsymbol{1}_{\{r>t\}}\boldsymbol{1}_{\{\lambda_r<\theta\leq\lambda_r\circ\varepsilon^+_{(t,0,1)}\}}, \quad r\in[t,+\infty),\ \theta\geq 0.$$

Proof. The commutation relation (2.9) can be derived in the ramework of the Malliavin calculus with respect to N. In particular, according to [Pri09, Proposition 4.1.4], for any $(t,\theta,x)\in\mathbb{R}^2_+\times\mathbb{R}$ we have

$$D_{(t,\theta,x)}\delta^N(z\mathcal{Z}) = z(t,x)\mathcal{Z}_{(t,\theta)} + \delta^N(D_{(t,\theta,x)}(z\mathcal{Z})).$$

By the definition (2.7) of $D_{(t,x)}$ and the fact that z is deterministic, we obtain

$$D_{(t,0,x)}\delta^N(z\mathcal{Z}) = z(t,x) + \delta^N(zD_{(t,0,x)}\mathcal{Z}).$$

In addition, as ν does not appear in the expression of (H,λ) (see (2.5)), we have

$$\begin{aligned} D_{(t,0,x)}\mathcal{Z}_{(r,\theta)} &= (\mathcal{Z}\circ\varepsilon_{(t,0,x)^+\}})_{(r,\theta)} - \mathcal{Z}_{(r,\theta)} \\ &= \mathbf{1}_{\{r>t\}}\big(\mathbf{1}_{\{\theta\leq\lambda_r\circ\varepsilon^+_{(t,0,x)}\}} - \mathbf{1}_{\{\theta\leq\lambda_r\}}\big) \\ &= \mathbf{1}_{\{r>t\}}\mathbf{1}_{\{\lambda_r<\theta\leq\lambda_r\circ\varepsilon^+_{(t,0,x)}\}} \\ &= \mathbf{1}_{\{r>t\}}\mathbf{1}_{\{\lambda_r<\theta\leq\lambda_r\circ\varepsilon^+_{(t,0,1)}\}}. \end{aligned}$$

where for the last equality, as remarked in the proof of Lemma 4.2, see also (4.3), $\lambda_r\circ\varepsilon^+_{(t,0,x)}$ does not depend on the value x which thus can be taken equal to $x=1$. □

3 Main results

3.1 A general estimate

In Theorem 3.1 we present our main estimate for functionals of the form

$$\begin{aligned} F &= \delta^N(z\mathcal{Z}) \\ &= \int_{\mathbb{R}_+^2 \times \mathbb{R}} z(t,x)\mathcal{Z}_{(t,\theta)} \left(N(dt,d\theta,dx) - dtd\theta\nu(dx)\right) \\ &= \int_{\mathbb{R}_+ \times \mathbb{R}} z(t,x) \left(dH_t - \lambda_t dt\nu(dx)\right), \end{aligned}$$

where

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

and $z(t,x)$ is a deterministic square-integrable function.

Theorem 3.1. *Let* $z := (z(t,x))_{(t,x)\in\mathbb{R}_+\times\mathbb{R}} \in L^2(\mathbb{R}_+ \times \mathbb{R}, dt \otimes \nu)$, $F := \delta^N(z\mathcal{Z})$ *and* $\mathcal{N}_{\gamma^2} \sim \mathcal{N}(0,\gamma^2)$ *with* $\gamma^2 > 0$. *It holds that:*

(i) $$d_W\big(F, \mathcal{N}_{\gamma^2}\big) \le \mathbb{E}\left[\left|\gamma^2 - \int_{\mathbb{R}_+\times\mathbb{R}} |z(t,x)|^2 \lambda_t dt\nu(dx)\right|\right] + \mathbb{E}\left[\int_{\mathbb{R}_+\times\mathbb{R}} |z(t,x)||D_{(t,x)}F|^2 \lambda_t dt\nu(dx)\right]. \quad (3.1)$$

(ii) If in addition, $\mathbb{E}[|F|^3] < +\infty$ *and* $z(t,x)$ *satisfies*

$$z(t,x) \ge 0, \text{ for } dt \otimes \nu \text{ almost every } (t,x), \quad (3.2)$$

then

$$d_W\big(F, \mathcal{N}_{\gamma^2}\big) \le \mathbb{E}\left[\left|\gamma^2 - \int_{\mathbb{R}_+\times\mathbb{R}} |z(t,x)|^2 \lambda_t dt\nu(dx)\right|\right] + \mathbb{E}\big[F^3\big]. \quad (3.3)$$

Remark 3.2. *We note that the term* $\mathbb{E}\big[F^3\big]$ *in* (3.3) *is also the third cumulant of the centered random variable* F, *see Section 4.8 of [CGS11] on the "third moment phenomenon". In particular, the vanishing of the first cumulant can lead to faster convergence rates, see, e.g. [Pri18, Pri19], [Dun21], and Theorem 3.7 below.*

3.2 Quantitative limit theorem for compound Hawkes processes

Throughout this section we consider

$$S_t = \sum_{i=1}^{H_t} X_i, \quad t \geq 0,$$

defined through the system (2.5), where $(X_i)_i$ is a sequence of independent and identically distributed random variables with common distribution ν, independent of the Hawkes process H with intensity λ given by (2.3). We will assume in addition that the kernel ϕ satisfies the following condition.

Assumption 3.3. *The function $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ is such that*

$$\|\phi\|_1 = \int_0^\infty \phi(u)du < 1, \quad \textit{and} \quad \int_0^\infty u\phi(u)du < +\infty.$$

Assumption 3.3 allows us to define

$$\psi := \sum_{n\geq 1} \phi^{(*n)}, \tag{3.4}$$

where $\phi^{(*n)}$ is the n-th convolution of ϕ with itself, with

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

In the remainder of this paper, C denotes a positive constant that depends only on μ, ϕ, ϑ and ν, and may change from place to place.

Theorem 3.4 (Quantitative limit theorem for the Hawkes process)**.** *Assume that $\mathbb{E}[X_1^2] < +\infty$ and that Assumption 3.3 holds, and set*

$$\gamma^2 := \mu \frac{\vartheta^2}{1-\|\phi\|_1} \quad \textit{and} \quad F_T := \frac{S_T - \mathbb{E}[X_1]\int_0^T \lambda_t dt}{\sqrt{T}}, \qquad T > 0.$$

Then, there exists $C > 0$ depending only on $\mu, \|\phi\|_1, \vartheta$, such that

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

Remark 3.5. *As noted in its proof, the above bound relies on the approach of [Tor16, HHKR21] to bound the Wasserstein distance between the distribution of* $F := \delta^N(z\mathcal{Z})$ *and* $\mathcal{N}_{\gamma^2} \sim \mathcal{N}(0, \gamma^2)$*. However, [Tor16, Theorem 3.1] only implies*

$$d_W(F_T, \mathcal{N}(0, \sigma^2)) \leq B(T), \quad \text{with } B(T) \geq \sqrt{\frac{8}{\pi}} \|\phi\|_1 \frac{2 - \|\phi\|_1}{1 - \|\phi\|_1},$$

which does not converge to zero as T *tends to* $+\infty$ *when* ϕ *only satisfies Assumption 3.3. The situation is similar in [HHKR21] which derives the bound*

$$d_W(F_T, \mathcal{N}(0, \sigma^2)) \leq \frac{C}{\sqrt{T}} + R_T, \tag{3.5}$$

where R_T *may not converge to* 0 *as* T *goes to* $+\infty$*, except when* ϕ *is an exponential or Erlang kernel, in which case we have* $R_T = O(T^{-1/2})$*, see Theorems 3.10 and 3.12 therein. Here, in the case of a deterministic integrand* z *we are able to remove* R_T *in* (3.5) *via a better estimate of*

$$\begin{aligned} \left|\mathbb{E}\left[\gamma^2 f'(F) - f(F)F\right]\right| \leq & \left|\mathbb{E}\left[f'(F)\left(\gamma^2 - \int_{\mathbb{R}_+ \times \mathbb{R}} z(t,x)\lambda_t D_{(t,x)} F dt \nu(dx)\right)\right]\right| \\ & + \frac{1}{2}\left|\mathbb{E}\left[\int_{\mathbb{R}_+ \times \mathbb{R}} z(t,x)\lambda_t f''\left(\overline{F}^{t,x}\right)|D_{(t,x)}F|^2 dt\nu(dx)\right]\right| \end{aligned}$$

in (4.11)*, uniformly in* $f \in \mathcal{F}_W$*. While in [HHKR21] the first term is controlled only for specific kernels* ϕ *with* z *a predictable process, in this paper, when* z *is deterministic we write*

$$\begin{aligned} & \left|\mathbb{E}\left[f'(F)\left(\gamma^2 - \int_{\mathbb{R}_+ \times \mathbb{R}} z(t,x)\lambda_t D_{(t,x)} F dt \nu(dx)\right)\right]\right| \\ & \leq \left|\mathbb{E}\left[f'(F)\left(\gamma^2 - \int_{\mathbb{R}_+ \times \mathbb{R}} |z(t,x)|^2 \lambda_t dt \nu(dx)\right)\right]\right| \\ & \leq \mathbb{E}\left[\left|\gamma^2 - \int_{\mathbb{R}_+ \times \mathbb{R}} |z(t,x)|^2 \lambda_t dt \nu(dx)\right|\right], \end{aligned}$$

see Lemma 4.1 and (4.12)*, and as a consequence we derive explicit convergence rates for compound Hawkes processes with general kernels* ϕ*. In addition, this approach yields third order cumulant-type estimates when* z *is deterministic, see Theorem 3.1-(ii), that were not obtained in [HHKR21].*

In Theorem 3.7 we provide a faster rate of convergence by considering a set of smoother test functions in the definition of the distance $d_{4,\infty}$ (see Notation A.1 in Appendix A) under additional conditions on the moments of the random variables X_i, as in the following quantitative central limit theorem given of [CGS11].

Proposition 3.6. *([CGS11, Corollary 4.4 in Section 4.8]) Assume that*

$$\mathbb{E}[X_1] = \mathbb{E}\big[X_1^3\big] = 0, \quad \mathbb{E}[X_1^2] = 1, \quad \text{and} \quad \mathbb{E}[X_1^4] = \int_{-\infty}^{\infty} x^4 \nu(dx) < +\infty.$$

Then, letting $W_n := n^{-1/2} \sum_{i=1}^n X_i$, $n \geq 1$, *we have*

$$d_{4,\infty}(W_n, \mathcal{N}(0,1)) \leq \frac{11 + \mathbb{E}[X_1^4]}{24n}.$$

The next result improves the rate of convergence $T^{-1/2}$ of Theorem 3.4 to T^{-1} under additional assumptions on the random sequence $(X_i)_i$.

Theorem 3.7. *Assume that Assumption 3.3 holds, that* $\mu \geq 0$, *and that*

$$\mathbb{E}[X_1] = \mathbb{E}\big[X_1^3\big] = 0, \quad \mathbb{E}[X_1^4] = \int_{-\infty}^{\infty} x^4 \nu(dx) < +\infty,$$

and let

$$\gamma^2 := \mu \frac{\vartheta^2}{1 - \|\phi\|_1} \quad \text{and} \quad \vartheta^2 := \mathbb{E}[X_1^2].$$

Then, there exists $C > 0$ *depending only on* $\mu, \|\phi\|_1, \vartheta$, *such that*

$$d_{4,\infty}(S_T/\sqrt{T}, \mathcal{N}(0,\gamma^2)) \leq \frac{C}{T}, \qquad T > 0.$$

Finally, as [HHKR21], we provide an alternative quantitative limit theorem by replacing the intensity process in the renormalisation with its asymptotic expectation. This result extends the Wasserstein bound of [HHKR21, Theorem 3.13] from simple Hawkes processes to compound Hawkes processes.

Theorem 3.8. *Assume that* $\mathbb{E}[X_1^2] < +\infty$ *and that Assumption 3.3 holds, and let*

$$\varpi := \mu \frac{\mathbb{E}[X_1]}{1 - \|\phi\|_1} \quad \text{and} \quad \Gamma_T := \frac{S_T - \varpi T}{\sqrt{T}}, \quad T > 0.$$

Then, there exists $C > 0$ *depending only on* $\mu, \|\phi\|_1, \vartheta$, *such that*

$$d_W(\Gamma_T, \mathcal{N}(0,\zeta^2)) \leq \frac{C}{\sqrt{T}}, \quad T > 0,$$

where

$$\zeta^2 := \mu \frac{\vartheta^2 + \|\phi\|_1(\vartheta^2 - (\mathbb{E}[X_1])^2)(\|\phi\|_1 - 2)}{(1 - \|\phi\|_1)^3}.$$

4 Proofs

4.1 Technical lemmata

Lemma 4.1. *Let* $z := (z(t,x))_{(t,x)\in\mathbb{R}_+\times\mathbb{R}} \in L^2(\mathbb{R}_+\times\mathbb{R}, dt\otimes\nu)$ *and consider* $\mathcal{Z}$ *defined in* (2.6)*. Then for any* $t \geq 0$ *and* $\varphi : \mathbb{R} \to \mathbb{R}$ *such that* $\varphi\big(\delta^N(z\mathcal{Z})\big)$ *belongs to* $L^2\big(\Omega^N, \mathbb{P}\big)$*, we have* $\mathbb{E}_t\big[\varphi\big(\delta^N(z\mathcal{Z})\big)\delta^N\big(z\widehat{\mathcal{Z}}^t\big)\big] = 0$.

Proof. Letting $F := \delta^N(z\mathcal{Z})$, we have

$$\mathbb{E}_t\big[\varphi(F)\delta^N\big(z\widehat{\mathcal{Z}}^t\big)\big] \quad = \quad \int_t^\infty \int_{\mathbb{R}_+\times\mathbb{R}} \mathbb{E}_t\big[z(s,y)\mathbf{1}_{\{\lambda_s<\theta\leq\lambda_s\circ\varepsilon^+_{(t,0,x)}\}} D_{(s,\theta,y)}\varphi(F)\big] d\theta\nu(dy)ds,$$

with, by definition,

$$\begin{aligned}
\mathbf{1}_{\{\lambda_s<\theta\}} D_{(s,\theta,y)}\varphi(F) &= \mathbf{1}_{\{\lambda_s<\theta\}} \Bigg[\varphi\Bigg(\Bigg(\int_{\mathbb{R}^2_+\times\mathbb{R}} z(t,x)\mathcal{Z}_{t,\rho}(N(dt,d\rho,dx) - dtd\rho\nu(dx))\Bigg)\circ\epsilon^+_{(s,\theta,y)}\Bigg) \\
&\quad -\varphi\Bigg(\int_{\mathbb{R}^2_+\times\mathbb{R}} z(t,x)\mathcal{Z}_{(t,\rho)}(N(dt,d\rho,dx) - dtd\rho\nu(dx))\Bigg)\Bigg] \qquad (4.1) \\
&= 0,
\end{aligned}$$

as $\mathbf{1}_{\{\lambda_s<\theta\}}\mathcal{Z}_{(s,\theta)} = 0$. More precisely, let

$$F_t := \int_{[0,t]\times\mathbb{R}_+\times\mathbb{R}} z(s,x)\mathcal{Z}_{s,\rho}(N(ds,d\rho,dx) - dsd\rho\nu(dx)), \qquad t \geq 0,$$

so that

$$F_\infty = \int_{\mathbb{R}^2_+\times\mathbb{R}} z(s,x)\mathcal{Z}_{s,\rho}(N(ds,d\rho,dx) - dsd\rho\nu(dx)),$$

and $(F_t, \lambda_t)_{t\in\mathbb{R}}$ is solution to the SDE

$$\begin{cases} F_t = \displaystyle\int_{(0,t]\times\mathbb{R}_+\times\mathbb{R}} z(s,x)\mathbf{1}_{\{\rho\leq\lambda_s\}}(N(ds,d\rho,dx) - dsd\rho\nu(dx)), \quad t \geq 0, \\ \lambda_t = \mu + \displaystyle\int_{(0,t)} \phi(t-u)dH_u, \quad t \geq 0. \end{cases}$$

Next, fix $(s,\theta,y) \in \mathbb{R}^2_+\times\mathbb{R}$. As $\mathbf{1}_{\{\lambda_s<\theta\}}\mathcal{Z}_{(s,\theta)} = 0$, we have

$$F_t\circ\varepsilon^+_{(s,\theta,y)} = F_t, \quad \lambda_t\circ\varepsilon^+_{(s,\theta,y)} = \lambda_t, \quad 0 \leq t < s,$$

and

$$\mathbf{1}_{\{\lambda_s<\theta\}}F_s\circ\varepsilon^+_{(s,\theta,y)} = \mathbf{1}_{\{\lambda_s<\theta\}}F_s, \quad \mathbf{1}_{\{\lambda_s<\theta\}}\lambda_{s^+}\circ\varepsilon^+_{(s,\theta,y)} = \mathbf{1}_{\{\lambda_s<\theta\}}\lambda_{s^+}.$$

As the process $\lambda_{s^+} \circ \varepsilon^+_{(s,\theta,y)}$ triggers the jumps of $F \circ \varepsilon^+_{(s,\theta,y)}$ and since it coincides with λ on $[0,s]$, the pair $(F \circ \varepsilon^+_{(s,\theta,y)}, \lambda \circ \varepsilon^+_{(s,\theta,y)})$ solves the same SDE as (F,λ) and thus coincides with it by uniqueness of the solution[2]; which yields the last equality in (4.1). $\square$

Lemma 4.2. *For all $(t,\eta,x) \in \mathbb{R}_+^2 \times \mathbb{R}$, it holds that:*

$$D_{(t,\eta,x)}\lambda_s \geq 0, \quad ds \otimes \nu - \; a.e..$$

Proof. The main proof idea relies on a comparison principle for the specific SDE involved, assessing that if intensity processes are ordered over the whole past (and not at some given time only), then this ordering between counting processes and intensities is preserved at future times. We recall that $D_{(t,\eta,x)}\lambda_s = \lambda_s \circ \epsilon^+_{(t,\eta,x)} - \lambda_s$. The conclusion follows from the definition of the intensity process λ which is the unique solution to an SDE with respect to N derived from (2.5):

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

Note that we can write for $s \geq t$,

$$\begin{aligned}\lambda_s = \mu &+ \int_{(0,t)\times\mathbb{R}_+\times\mathbb{R}} \phi(s-u)\mathbf{1}_{\{\theta\leq\lambda_u\}} N(du,d\theta,dx) \\ &+ \int_{(t,s)\times\mathbb{R}_+\times\mathbb{R}} \phi(s-u)\mathbf{1}_{\{\theta\leq\lambda_u\}} N(du,d\theta,dx),\end{aligned} \tag{4.2}$$

Similarly, for $s \geq t \geq 0$ we have

$$\begin{aligned}\lambda_s \circ \epsilon^+_{(t,\eta,x)} = \mu &+ \int_{(0,t)\times\mathbb{R}_+\times\mathbb{R}} \phi(s-u)\mathbf{1}_{\{\theta\leq\lambda_u\}} N(du,d\theta,dx) \\ &+ \phi(t-s) + \int_{(t,s)\times\mathbb{R}_+\times\mathbb{R}} \phi(s-u)\mathbf{1}_{\{\theta\leq\lambda_u\circ\epsilon^+_{(t,\eta,x)}\}} N(du,d\theta,dx),\end{aligned} \tag{4.3}$$

and $\lambda_s \circ \epsilon^+_{(t,\eta,x)} = \lambda_s$ for all $0 \leq s \leq t$, $\eta \in \mathbb{R}_+$ and $x \in \mathbb{R}$. In addition, $\lambda_{t^+} \circ \epsilon^+_{(t,\eta,x)} = \lambda_{t^+} + \phi(0) \geq \lambda_{t^+}$. Note that $\lambda_s \circ \epsilon^+_{(t,\eta,x)}$ does not depend on x in Equation (4.3). Let now

$$\tau_1 := \inf\left\{s > t \,:\, \Delta_s H \circ \epsilon^+_{(t,\eta,x)} \neq 0\right\} \wedge \inf\left\{s > t \,:\, \Delta_s H \neq 0\right\},$$

[2]As F is a counting process and λ is deterministic between two jumps times of F, the uniqueness property is proved pathwise by considering the jump times of F which are completely determined by the Poisson point process N and the intensity process λ. We provide the details on the method of proof in the proof of Lemma 4.2 below where a comparison principle is derived.

where we use the notation $\Delta_s H := H_{s^+} - H_s$. So τ_1 is the first jump of the Hawkes process H or of the shifted Hawkes process $H \circ \epsilon^+_{(t,\eta,x)}$ after t. Hence, from (4.2)-(4.3) we have

$$\lambda_s \circ \epsilon^+_{(t,\eta,x)} \geq \lambda_s, \quad s \in [0, \tau_1).$$

Thus, at time τ_1 (note that since N is a Poisson point process we have $\tau_1 < +\infty$, $\mathbb{P}$-a.s.), N jumps at an atom (τ_1, θ_1, x_1) which, by the previous ordering between $\lambda \circ \epsilon^+_{(t,\eta,x)}$ and λ, imposes that τ_1 is either a common jump time of $H \circ \epsilon^+_{(t,\eta,x)}$ and H, or a jump time of $H \circ \epsilon^+_{(t,\eta,x)}$ only, but it cannot be a jump time for H and not for $H \circ \epsilon^+_{(t,\eta,x)}$. In both situations (common jump or only a jump for the shifted Hawkes process) we have

$$\lambda_s \circ \epsilon^+_{(t,\eta,x)} \geq \lambda_s, \quad s \in [0, \tau_2),$$

where

$$\tau_2 := \inf \left\{ s > \tau_1 \, : \, \Delta_s H \circ \epsilon^+_{(t,\eta,x)} \neq 0 \right\} \wedge \inf \left\{ s > t \, : \, \Delta_s H \neq 0 \right\}$$

is the next possible candidate jump time of H and $H \circ \epsilon^+_{(t,\eta,x)}$. Defining

$$\tau_{n+1} := \inf \left\{ s > \tau_n \, : \, \Delta_s H \circ \epsilon^+_{(t,\eta,x)} \neq 0 \right\} \wedge \inf \left\{ s > t \, : \, \Delta_s H \neq 0 \right\}, \quad n \geq 0,$$

and letting n go to $+\infty$ concludes the proof. □

In what follows, for ease of notation we recall the notation $\mathbb{E}_t[\cdot] := \mathbb{E}[\cdot | \mathcal{F}^N_t]$, $t \in \mathbb{R}_+$, and recall that by (4.3) we have $\lambda \circ \varepsilon^+_{(t,0,x)} = \lambda \circ \varepsilon^+_{(t,0,1)}$ for any $(t,x) \in \mathbb{R}_+ \times \mathbb{R}$.

Lemma 4.3. *For fixed any $0 \leq t < s \leq T$, let $\widehat{\lambda}^t_s := \lambda_s \circ \varepsilon^+_{(t,0,1)} - \lambda_s$, and*

$$\widehat{H}^t_s := D_{(t,x)} H_s - 1 = \int_{(t,s] \times \mathbb{R}_+ \times \mathbb{R}} \mathbf{1}_{\{\lambda_r < \theta \leq \lambda_r \circ \varepsilon^+_{(t,0,1)}\}} N(dr, d\theta, dx). \tag{4.4}$$

Then $\left(\widehat{H}^t, \widehat{\lambda}^t\right)$ is a generalized Hawkes process (with a baseline intensity $\phi(\cdot - t)$) in the sense that $\widehat{H}^t$ is a counting process starting at time t with stochastic intensity process $\widehat{\lambda}^t$. In addition, $\widehat{\lambda}^t$ satisfies

$$\widehat{\lambda}^t_s = \phi(s-t) + \int_{(t,s)} \phi(s-u) d\widehat{H}^t_u, \quad s > t, \; \widehat{\lambda}^t_t = 0, \tag{4.5}$$

and the following relations hold true:

(i) for any $p > 0$

$$\text{esssup}_{t \in [0,T]} \left(\mathbb{E}_t \left[\int_t^T \widehat{\lambda}^t_s ds \right] \right)^p \leq (\|\phi\|_1 (1 + \|\psi\|_1))^p, \quad \mathbb{P} - a.s.. \tag{4.6}$$

(ii) $\int_t^T \left(\widehat{\lambda}^t_s - \mathbb{E}_t\left[\widehat{\lambda}^t_s\right]\right) ds = \int_t^T \psi(T-s)\widehat{\mathcal{M}}^t_s ds$, $s \in (t,T]$, *where* $\widehat{\mathcal{M}}^t_s = \widehat{H}^t_s - \int_t^s \widehat{\lambda}^t_r dr$.

(iii) For any $p > 0$, *there exists* $C_p > 0$ *depending on* p *only and such that*

$$\text{esssup}_{t\in[0,T]}\mathbb{E}_t\left[\left|\int_t^T \widehat{\lambda}^t_s ds\right|^2\right] \leq C_p\|\psi\|_1^2 < +\infty. \tag{4.7}$$

Proof. We note that $\left(\widehat{H}^t, \widehat{\lambda}^t\right)$ is a generalized Hawkes process (with baseline intensity 0) in the sense that $\widehat{H}^t$ is a counting process starting at time t with stochastic intensity process λ^t. Indeed, the process $\widehat{\mathcal{M}}^t$ defined as

$$\begin{aligned}\widehat{\mathcal{M}}^t_s &:= \delta^N\left(\mathbf{1}_{\{t<r\leq s\}}\mathbf{1}_{\{\lambda_r<\theta\leq\lambda_r\circ\varepsilon^+_{(t,0,1)}\}}\right)\\ &= \int_{(t,s]\times\mathbb{R}_+\times\mathbb{R}} \mathbf{1}_{\{\lambda_r<\theta\leq\lambda_r\circ\varepsilon^+_{(t,0,1)}\}}\left(N(dr,d\theta,dx) - drd\theta\nu(dx)\right)\\ &= \widehat{H}^t_s - \int_t^s \widehat{\lambda}^t_r dr,\end{aligned}$$

is a $(\mathcal{F}^N_s)_{s\geq t}$-martingale by (4.4)-(4.5). We refer to [HHKR21, Proposition 2.19] for more details regarding the link between $\left(\widehat{H}^t, \widehat{\lambda}^t\right)$ and the Malliavin derivative of (H,λ).
(i) A direct computation leads to (see [HHKR21, Proof of Lemma 4.2])

$$\mathbb{E}_t\left[\int_t^T \widehat{\lambda}^t_s ds\right] = \int_t^T \phi(s-t)ds + \int_t^T\int_u^T \psi(T-u)ds\phi(u-t)du,$$

and in turns to (see [HHKR21, Lemma 4.2])

$$\mathbb{E}_t\left[\int_t^T \widehat{\lambda}^t_s ds\right] \leq \|\phi\|_1(1+\|\psi\|_1), \quad \mathbb{P}-a.s.,$$

which gives (4.6).
(ii) This result is similar to the one for a Hawkes process with constant baseline intensity μ. However, to make this paper self-contained we present the proof below. Recall first that from [BDHM13, Lemma 3], given a locally bounded map $h:\mathbb{R}_+\to\mathbb{R}$, the unique solution to equation

$$f(s) = h(s) + \int_0^s \phi(s-u)f(u)du,$$

is given by

$$f(s) = h(s) + \int_0^s \psi(s-u)h(u)du. \tag{4.8}$$

As $\widehat{\mathcal{M}}^t_s = \widehat{H}^t_s - \int_t^s \widehat{\lambda}^t_r dr$, we have

$$\widehat{\lambda}^t_s = \phi(s-t) + \int_t^s \phi(s-u) d\widehat{\mathcal{M}}^t_u + \int_t^s \phi(s-u)\widehat{\lambda}^t_u du.$$

Taking the conditional expectation $\mathbb{E}_t[\,\cdot\,]$, we get

$$\mathbb{E}_t\big[\widehat{\lambda}^t_s\big] = \phi(s-t) + \int_t^s \phi(s-u)\mathbb{E}_t\big[\widehat{\lambda}^t_u\big]du,$$

which leads to

$$\widehat{\lambda}^t_s - \mathbb{E}_t\big[\widehat{\lambda}^t_s\big] = \int_t^s \phi(s-u) d\widehat{\mathcal{M}}^t_u + \int_t^s \phi(s-u)\big(\widehat{\lambda}^t_u - \mathbb{E}_t\big[\widehat{\lambda}^t_u\big]\big)du.$$

This expression is true for any $s > t$, and can be extended to any $s > 0$ as follows:

$$\mathbf{1}_{\{s>t\}}\big(\widehat{\lambda}^t_s - \mathbb{E}_t\big[\widehat{\lambda}^t_s\big]\big) = \mathbf{1}_{\{s>t\}}\int_t^s \phi(s-u) d\widehat{\mathcal{M}}^t_u + \mathbf{1}_{\{s>t\}}\int_0^s \phi(s-u)\mathbf{1}_{\{u>t\}}\big(\widehat{\lambda}^t_u - \mathbb{E}\big[\widehat{\lambda}^t_u\big]\big)du.$$

Letting $f_t(s) := \mathbf{1}_{\{s>t\}}\big(\widehat{\lambda}^t_s - \mathbb{E}_t\big[\widehat{\lambda}^t_s\big]\big)$ being defined for any $s > 0$, and vanishing if $0 \le s \le t$, and $h_t(s) := \mathbf{1}_{\{s>t\}}\int_t^s \phi(s-u)d\widehat{\mathcal{M}}^t_u$, the above previous relation rewrites as

$$f_t(s) = h_t(s) + \mathbf{1}_{\{s>t\}}\int_0^s \phi(s-u) f_t(u)du.$$

As $\{s < t\}$ implies $\{u < t\}$ and so $f_t(u) = 0$, the indicator function can be removed and thus

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

Applying (4.8) we thus get $f_t(s) = h_t(s) + \int_0^s \psi(s-u)h_t(u)du$, which means

$$\mathbf{1}_{\{s>t\}}\big(\widehat{\lambda}^t_s - \mathbb{E}_t\big[\widehat{\lambda}^t_s\big]\big) = \mathbf{1}_{\{s>t\}}\int_t^s \phi(s-u) d\widehat{\mathcal{M}}^t_u + \int_0^s \psi(s-u)\mathbf{1}_{\{u>t\}}\int_t^u \phi(u-v)d\widehat{\mathcal{M}}^t_v du.$$

Next, using Fubini's theorem, the fact that the stochastic integrals are defined pathwise, and the definition (3.4) of ψ, we have

$$\begin{aligned}
\int_t^T \big(\widehat{\lambda}^t_s - \mathbb{E}_t\big[\widehat{\lambda}^t_s\big]\big)ds &= \int_t^T\int_t^s \phi(s-u)d\widehat{\mathcal{M}}^t_u ds \\
&\quad + \int_t^T\int_0^s \psi(s-u)\mathbf{1}_{\{u>t\}}\int_t^u \phi(u-v)d\widehat{\mathcal{M}}^t_v du ds \\
&= \int_t^T\int_t^s \phi(s-u)d\widehat{\mathcal{M}}^t_u ds + \int_t^T\int_t^s\int_t^s \psi(s-u)\mathbf{1}_{\{u>v>t\}}\phi(u-v)d\widehat{\mathcal{M}}^t_v du ds
\end{aligned}$$

$$
\begin{aligned}
&= \int_t^T \int_t^s \phi(s-u)d\widehat{\mathcal{M}}_u^t ds + \int_t^T \int_t^s \int_v^s \psi(s-u)\phi(u-v)dud\widehat{\mathcal{M}}_v^t ds \\
&= \int_t^T \int_t^s \phi(s-u)d\widehat{\mathcal{M}}_u^t ds + \int_t^T \int_t^s \left(\int_0^{s-v} \psi(s-v-z)\phi(z)dz\right) d\widehat{\mathcal{M}}_v^t ds \\
&= \int_t^T \int_t^s \phi(s-u)d\widehat{\mathcal{M}}_u^t ds + \int_t^T \int_t^s (\psi * \phi)(s-v)d\widehat{\mathcal{M}}_v^t ds \\
&= \int_t^T \int_t^s \phi(s-u)d\widehat{\mathcal{M}}_u^t ds + \int_t^T \int_t^s \left(\sum_{n\geq 1} \phi^{*n} * \phi\right)(s-v)d\widehat{\mathcal{M}}_v^t ds \\
&= \int_t^T \int_t^s \phi(s-u)d\widehat{\mathcal{M}}_u^t ds + \int_t^T \int_t^s \left(\sum_{n\geq 2} \phi^{*n}\right)(s-v)d\widehat{\mathcal{M}}_v^t ds \\
&= \int_t^T \int_t^s \phi(s-u)d\widehat{\mathcal{M}}_u^t ds + \int_t^T \int_t^s (\psi(s-v) - \phi(s-v))\, d\widehat{\mathcal{M}}_v^t ds \\
&= \int_t^T \int_t^s \psi(s-v)d\widehat{\mathcal{M}}_v^t ds \\
&= \int_t^T \int_v^T \psi(s-v)dsd\widehat{\mathcal{M}}_v^t \\
&= \int_t^T \psi(T-v)\widehat{\mathcal{M}}_v^t dv,
\end{aligned}
$$

where we applied again Fubini's theorem, and integration by parts over $[t,T]$.
(iii) By the Burkholder-Davis-Gundy inequality, we have

$$
\text{esssup}_{\{(t,s)\,:\,0\leq t\leq s\leq T\}}\mathbb{E}_t\left[\left|\widehat{\mathcal{M}}_s^t\right|^2\right] \leq C\text{esssup}_{t\in[0,T]}\mathbb{E}_t\left[\widehat{H}_T^t\right] = C\text{esssup}_{t\in[0,T]}\mathbb{E}_t\left[\int_t^T \widehat{\lambda}_s^t ds\right]
$$

leading to

$$
\text{esssup}_{\{(t,s)\,:\,0\leq t\leq s\leq T\}}\mathbb{E}_t\left[\left|\widehat{\mathcal{M}}_s^t\right|^2\right] < +\infty, \quad \mathbb{P}-a.s..
$$

This relation shows that (iii) is a consequence of (ii). Indeed, if (ii) is satisfied, then Cauchy-Schwarz's inequality implies

$$
\begin{aligned}
\mathbb{E}_t\left[\left(\int_t^T \left(\widehat{\lambda}_s^t - \mathbb{E}_t\left[\widehat{\lambda}_s^t\right]\right)ds\right)^2\right] &= \mathbb{E}_t\left[\left(\int_t^T \psi(T-s)\widehat{\mathcal{M}}_s^t ds\right)^2\right] \\
&= 2\int_t^T \int_s^T \psi(T-s)\psi(T-u)\mathbb{E}_t\left[\widehat{\mathcal{M}}_s^t\widehat{\mathcal{M}}_u^t\right]duds \\
&\leq C\|\psi\|_1^2.
\end{aligned}
$$

Combining this estimate with (4.6) we get immediately that for any $p > 0$,

$$\text{esssup}_{t\in[0,T]} \left(\mathbb{E}_t \left[\left| \int_t^T \widehat{\lambda}_s^t ds \right|^2 \right] \right)^p \leq C < +\infty,$$

which proves (iii). □

Lemma 4.4. *Letting*

$$R_{t,T} := \int_{(t,T]\times\mathbb{R}_+\times\mathbb{R}} y \mathbf{1}_{\{\lambda_r < \theta \leq \lambda_r \circ \varepsilon^+_{(t,0,1)}\}} \left(N(dr, d\theta, dy) - dr d\theta \nu(dy) \right), \quad 0 \leq t \leq T,$$

we have

$$\text{esssup}_{t\in[0,T]} \mathbb{E}_t \left[|R_{t,T}|^3 \right] < \infty, \quad \mathbb{P} - a.s..$$

Proof. Throughout this proof, $C > 0$ denotes a positive constant that may change from line to line, and is independent of t and T considered below. By the Burkholder-Davis-Gundy inequality, and the Jensen inequality we have

$$\begin{aligned}
\mathbb{E}_t \left[|R_{t,T}|^3 \right] &\leq C \mathbb{E}_t \left[\left| \int_{(t,T]\times\mathbb{R}_+\times\mathbb{R}} y^2 \mathbf{1}_{\{\lambda_r < \theta \leq \lambda_r \circ \varepsilon^+_{(t,0,1)}\}} N(dr, d\theta, dy) \right|^{3/2} \right] \\
&\leq C \left(\mathbb{E}_t \left[\left| \int_{(t,T]\times\mathbb{R}_+\times\mathbb{R}} y^2 \mathbf{1}_{\{\lambda_r < \theta \leq \lambda_r \circ \varepsilon^+_{(t,0,1)}\}} N(dr, d\theta, dy) \right|^2 \right] \right)^{3/4} \\
&\leq C \left(\mathbb{E}_t \left[\left| \int_{(t,T]\times\mathbb{R}_+\times\mathbb{R}} y^2 \mathbf{1}_{\{\lambda_r < \theta \leq \lambda_r \circ \varepsilon^+_{(t,0,1)}\}} (N(dr, d\theta, dy) - dr d\theta \nu(dy)) \right|^2 \right] \right)^{3/4} \\
&\quad + C \left(\int_{-\infty}^{\infty} y^2 \nu(dy) \right)^{3/2} \left(\mathbb{E}_t \left[\left| \int_t^T \widehat{\lambda}_r^t dr \right|^2 \right] \right)^{3/4} \\
&= C \left(\int_{-\infty}^{\infty} y^4 \nu(dy) \right)^{3/4} \left(\mathbb{E}_t \left[\int_t^T \widehat{\lambda}_r^t dr \right] \right)^{3/4} + C \left(\int_{-\infty}^{\infty} y^2 \nu(dy) \right)^{3/2} \left(\mathbb{E}_t \left[\left| \int_t^T \widehat{\lambda}_r^t dr \right|^2 \right] \right)^{3/4},
\end{aligned}$$

and the conclusion follows from (4.6)-(4.7). □

Lemma 4.5. *For $T > 0$, set*

$$\mathfrak{R}_T := \frac{H_T - \int_0^T \mathbb{E}[\lambda_s] ds}{\sqrt{T}} - \frac{\mathcal{M}_T}{\sqrt{T}(1 - \|\phi\|_1)}, \quad \textit{and} \quad \mathcal{M}_T := H_T - \int_0^T \lambda_u du.$$

Under the assumptions of Theorem 3.8, there exists a constant $C > 0$ such that

$$\mathbb{E}[\mathfrak{R}_T^2] \leq \frac{C}{T}.$$

Proof. Using [BDHM13, Lemma 4], we have

$$\begin{aligned} Y_T &:= \frac{1}{\sqrt{T}}\left(H_T - \mathbb{E}[H_T]\right) \\ &= \frac{1}{\sqrt{T}}\left(H_T - \int_0^T \mathbb{E}[\lambda_t]dt\right) \\ &= \frac{1}{\sqrt{T}}\left(\mathcal{M}_T + \int_0^T \psi(T-s)\mathcal{M}_s ds\right), \end{aligned}$$

hence

$$\begin{aligned} Y_T - \frac{\mathcal{M}_T}{\sqrt{T}(1-\|\phi\|_1)} &= \frac{\mathcal{M}_T}{\sqrt{T}}\left(1 - \frac{1}{1-\|\phi\|_1}\right) + \frac{1}{\sqrt{T}}\int_0^T \psi(T-s)\mathcal{M}_s ds \\ &= -\frac{1}{\sqrt{T}}\int_0^{+\infty} \psi(s)\mathcal{M}_T ds + \frac{1}{\sqrt{T}}\int_0^T \psi(T-s)\mathcal{M}_s ds, \end{aligned}$$

because $\int_0^{+\infty} \psi(s)ds = \|\phi\|_1/(1-\|\phi\|_1)$. Thus, if we set

$$\mathfrak{R}_T = \frac{1}{\sqrt{T}}\left(\int_0^T \psi(s)(\mathcal{M}_{T-s} - \mathcal{M}_T)ds - \mathcal{M}_T\int_T^{+\infty}\psi(s)ds\right).$$

We have that

$$\begin{aligned} \mathbb{E}[\mathfrak{R}_T^2] &\leq 2\left(\frac{1}{T}\mathbb{E}[\mathcal{M}_T^2]\left(\int_T^{+\infty}\psi(s)ds\right)^2 + \frac{1}{T}\mathbb{E}\left[\left(\int_0^T \psi(s)(\mathcal{M}_T - \mathcal{M}_{T-s})ds\right)^2\right]\right) \\ &= 2(A_1 + A_2). \end{aligned}$$

Using the fact that the expected value of the square of a martingale is the expected value of its quadratic variation which in this case is the process jumps we have, using Assumption 3.3, that

$$\begin{aligned} A_1 &= \frac{1}{T}\mathbb{E}\left[[\mathcal{M}]_T\right]\left(\int_T^{+\infty}\psi(s)ds\right)^2 \\ &= \frac{1}{T}\mathbb{E}[H_T]\left(\int_T^{+\infty}\psi(s)ds\right)^2 \\ &= O(1)\left(\int_T^{+\infty}\psi(s)ds\right)^2 \\ &\leq O\left(\frac{1}{T^2}\right)\left(\int_0^{+\infty}s\psi(s)ds\right)^2 \\ &= O\left(\frac{1}{T^2}\right) \end{aligned}$$

as in the proof of [BDHM13, Lemma 5]. By expanding the square, the second term yields

$$
\begin{aligned}
A_2 &= \frac{1}{T}\mathbb{E}\left[\left(\int_0^T \psi(s)(\mathcal{M}_T - \mathcal{M}_{T-s})ds\right)^2\right] \\
&= \frac{2}{T}\mathbb{E}\left[\int_0^T \int_0^s \psi(s)(\mathcal{M}_T - \mathcal{M}_{T-s})\psi(r)(\mathcal{M}_T - \mathcal{M}_{T-r})drds\right] \\
&= \frac{2}{T}\int_0^T \int_0^s \psi(s)\psi(r)\mathbb{E}\left[(\mathcal{M}_T^2 - \mathcal{M}_T\mathcal{M}_{T-s} - \mathcal{M}_T\mathcal{M}_{T-r} + \mathcal{M}_{T-r}\mathcal{M}_{T-s})\right] drds.
\end{aligned}
$$

Since for any $a \leq b$, $\mathbb{E}[\mathcal{M}_a\mathcal{M}_b] = \mathbb{E}[\mathcal{M}_a^2] = \mathbb{E}[H_a]$, we have that

$$
\begin{aligned}
A_2 &= \frac{2}{T}\int_0^T \int_0^s \psi(s)\psi(r)(\mathbb{E}[H_T] - \mathbb{E}[H_{T-s}] - \mathbb{E}[H_{T-r}] + \mathbb{E}[H_{T-s}])drds \\
&= \frac{2}{T}\int_0^T \int_0^s \psi(s)\psi(r)\mathbb{E}[H_T - H_{T-r}]drds.
\end{aligned}
$$

In order to bound the integral, we use once again [BDHM13, Lemma 4] to obtain

$$
\begin{aligned}
\mathbb{E}[H_T - H_{T-r}] &= r\mu + \left(r\int_0^{T-r} \psi(u)du + \int_{T-r}^T \psi(u)(T-u)du\right)\mu \\
&= \mu r + \mu r\int_0^{T-r} \psi(u)du + \mu\int_0^r \psi(T-u)udu \\
&\leq \mu r + \mu r\|\phi\|_1 + \mu\int_0^r \psi(T-u)rdu \\
&\leq Cr,
\end{aligned}
$$

for some $C > 0$, and since ψ is nonnegative,

$$
\begin{aligned}
A_2 &\leq \frac{C}{T}\int_0^T \int_0^s \psi(s)\psi(r)rdrds \\
&\leq \frac{C}{T}\|\psi\|_1 \int_0^T r\psi(r)dr,
\end{aligned}
$$

which yields the desired result. □

4.2 Proof of Theorem 3.1

According to Stein's method, see Appendix A, the Wasserstein distance between F and $\mathcal{N}_{\gamma^2}$ can be bounded by

$$
d_W\big(F, \mathcal{N}_{\gamma^2}\big) \leq \sup_{f\in\mathcal{F}_W} \left|\mathbb{E}[\gamma^2 f'(F) - Ff(F)]\right|,
$$

see (A.1), where

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

In addition, the right hand side is equal to 0 if and only if $F \sim \mathcal{N}(0, \gamma^2)$.
(i) We follow the beginning the Nourdin-Peccati's methodology (see *e.g.* [PSTU10])and apply the integration by parts formula to $\mathbb{E}[Ff(F)]$ for f in $\mathcal{F}_W$. More precisely, according to [HHKR21, Proof of Theorem 3.4], and using the integration by parts formula for the Hawkes process, see (2.8), we have

$$\begin{aligned}
\mathbb{E}\left[f(F)F\right] &= \mathbb{E}\left[f(F)\delta^N(z\mathcal{Z})\right] \\
&= \mathbb{E}\left[\int_{\mathbb{R}_+\times\mathbb{R}} z(t,x)\lambda_t D_{(t,x)} f(F) dt\nu(dx)\right] \\
&= \mathbb{E}\left[\int_{\mathbb{R}_+\times} z(t,x)\lambda_t \left(f(F\circ\varepsilon^+_{(t,0,x)}) - f(F)\right) dt\nu(dx)\right].
\end{aligned}$$

By Taylor expansion, we have

$$f(F\circ\varepsilon^+_{(t,0,x)}) - f(F) = f'(F) D_{(t,x)}F + \frac{1}{2} f''\left(\overline{F}^{t,x}\right)|D_{(t,x)}F|^2, \tag{4.9}$$

where $\overline{F}^{t,x}$ denotes a random element between $F\circ\varepsilon^+_{(t,0,x)}$ and F. Hence we have

$$\begin{aligned}
\mathbb{E}\left[\gamma^2 f'(F) - f(F)F\right] = \mathbb{E}&\left[f'(F)\left(\gamma^2 - \int_{\mathbb{R}_+\times\mathbb{R}} z(t,x)\lambda_t D_{(t,x)}F dt\nu(dx)\right)\right] \\
&- \frac{1}{2}\mathbb{E}\left[\int_{\mathbb{R}_+\times\mathbb{R}} z(t,x)\lambda_t f''\left(\overline{F}^{t,x}\right)|D_{(t,x)}F|^2 dt\nu(dx)\right].
\end{aligned} \tag{4.10}$$

At this stage, we provide a different treatment of the first time by expanding $D_{(t,x)}F$ according to (2.9). Thus

$$\begin{aligned}
&\mathbb{E}\left[f'(F)\left(\gamma^2 - \int_{\mathbb{R}_+\times\mathbb{R}} z(t,x)\lambda_t D_{(t,x)}F dt\nu(dx)\right)\right] \\
&= \mathbb{E}\left[f'(F)\left(\gamma^2 - \int_{\mathbb{R}_+\times\mathbb{R}} |z(t,x)|^2\lambda_t dt\nu(dx)\right)\right] - \mathbb{E}\left[f'(F)\int_{\mathbb{R}_+\times\mathbb{R}} z(t,x)\lambda_t\delta^N\left(z\widehat{\mathcal{Z}}^t\right) dt\nu(dx)\right],
\end{aligned}$$

so that

$$\begin{aligned}
\mathbb{E}\left[\gamma^2 f'(F) - f(F)F\right] = \mathbb{E}&\left[f'(F)\left(\gamma^2 - \int_{\mathbb{R}_+\times\mathbb{R}} |z(t,x)|^2\lambda_t dt\nu(dx)\right)\right] \\
&- \frac{1}{2}\mathbb{E}\left[\int_{\mathbb{R}_+\times\mathbb{R}} z(t,x)\lambda_t f''\left(\overline{F}^{t,x}\right)|D_{(t,x)}F|^2 dt\nu(dx)\right]
\end{aligned} \tag{4.11}$$

$$-\int_{\mathbb{R}_+\times\mathbb{R}} z(t,x)\mathbb{E}\big[\lambda_t\mathbb{E}_t\big[f'(F)\delta^N\big(z\widehat{\mathcal{Z}}^t\big)\big]\big]dt\nu(dx).$$

We now compute the last term in (4.11). By Lemma 4.1, we get

$$\mathbb{E}_t\big[f'(F)\delta^N\big(z\widehat{\mathcal{Z}}^t\big)\big]=0,$$

and the estimate (4.11) reads

$$\begin{aligned}\mathbb{E}\left[\gamma^2 f'(F)-f(F)F\right]=&\mathbb{E}\left[f'(F)\left(\gamma^2-\int_{\mathbb{R}_+\times\mathbb{R}}|z(t,x)|^2\lambda_t dt\nu(dx)\right)\right]\\&-\frac{1}{2}\mathbb{E}\left[\int_{\mathbb{R}_+\times\mathbb{R}} z(t,x)\lambda_t f''\big(\overline{F}^{t,x}\big)|D_{(t,x)}F|^2 dt\nu(dx)\right],\end{aligned}\tag{4.12}$$

which in turn implies

$$\begin{aligned}&\big|\mathbb{E}\big[\gamma^2 f'(F)-f(F)F\big]\big|\leq\left|\mathbb{E}\left[f'(F)\left(\gamma^2-\int_{\mathbb{R}_+\times\mathbb{R}}|z(t,x)|^2\lambda_t dt\nu(dx)\right)\right]\right|\\&\quad+\frac{1}{2}\left|\mathbb{E}\left[\int_{\mathbb{R}_+\times\mathbb{R}} z(t,x)\lambda_t f''\big(\overline{F}^{t,x}\big)|D_{(t,x)}F|^2 dt\nu(dx)\right]\right|\\&\leq\|f'\|_\infty\mathbb{E}\left[\left|\gamma^2-\int_{\mathbb{R}_+\times\mathbb{R}}|z(t,x)|^2\lambda_t dt\nu(dx)\right|\right]+\frac{\|f''\|_\infty}{2}\mathbb{E}\left[\int_{\mathbb{R}_+\times\mathbb{R}}|z(t,x)||D_{(t,x)}F|^2\lambda_t dt\nu(dx)\right],\end{aligned}$$

which leads to (3.1) as f belongs to $\mathcal{F}_W$.

(ii) Using once again the integration by parts formula (2.8) for the Hawkes process and the fact that (by an elementary algebraic computation) that $D(F^2)=|DF|^2+2FDF$ (see Remark 2.5 and Relation (2.2)), we have

$$\begin{aligned}&\mathbb{E}\big[F^3\big]=\mathbb{E}\big[\delta^N(z\mathcal{Z})F^2\big]\\&=\int_{\mathbb{R}_+\times\mathbb{R}}\mathbb{E}\left[z(t,x)\lambda_t D_{(t,x)}(F^2)\right]dt\nu(dx)\\&=\int_{\mathbb{R}_+\times\mathbb{R}} z(t,x)\mathbb{E}\left[|D_{(t,x)}F|^2\lambda_t\right]dt\nu(dx)+2\int_{\mathbb{R}_+\times\mathbb{R}} z(t,x)\mathbb{E}\left[\lambda_t\, FD_{(t,x)}F\right]dt\nu(dx)\\&=:T_1+2T_2.\end{aligned}\tag{4.13}$$

Note that T_1 is exactly the second term in the right-hand side of (3.1), and thus the result follows if we show that $T_2\geq 0$. To this end we compute this term. Using Relation (2.9) and integration by parts, we can expand this term as follows:

$$T_2=\int_{\mathbb{R}_+\times\mathbb{R}} z(t,x)\mathbb{E}\left[\lambda_t\, FD_{(t,x)}F\right]dt\nu(dx)$$

$$= \int_{\mathbb{R}_+\times\mathbb{R}} |z(t,x)|^2 \mathbb{E}\left[\lambda_t F\right] dt\nu(dx) + \int_{\mathbb{R}_+\times\mathbb{R}} z(t,x)\mathbb{E}\big[\lambda_t F \delta^N\big(z\widehat{\mathcal{Z}}^t\big)\big] dt\nu(dx)$$
$$= \int_{\mathbb{R}_+\times\mathbb{R}} |z(t,x)|^2 \mathbb{E}\left[\lambda_t F\right] dt\nu(dx) + \int_{\mathbb{R}_+\times\mathbb{R}} z(t,x)\mathbb{E}\big[\lambda_t \mathbb{E}_t\big[F \delta^N\big(z\widehat{\mathcal{Z}}^t\big)\big]\big] dt\nu(dx)$$
$$= \int_{\mathbb{R}_+\times\mathbb{R}} |z(t,x)|^2 \mathbb{E}[\lambda_t F] dt\nu(dx),$$

where we applied Lemma 4.1. By definition of F, we get that

$$\begin{aligned}
T_2 &= \int_{\mathbb{R}_+\times\mathbb{R}} |z(t,x)|^2 \mathbb{E}[\lambda_t F] dt\nu(dx) \\
&= \int_{\mathbb{R}_+\times\mathbb{R}} |z(t,x)|^2 \mathbb{E}[\lambda_t \delta^N(z\mathcal{Z})] dt\nu(dx) \\
&= \int_{\mathbb{R}_+\times\mathbb{R}} |z(t,x)|^2 \mathbb{E}\left[\int_{[0,t)\times\mathbb{R}} z(s,y)\lambda_s D_{(s,y)}\lambda_t ds\nu(dy)\right] dt\nu(dx) \\
&= \int_{\mathbb{R}_+\times\mathbb{R}} \int_{[0,t)\times\mathbb{R}} z(s,y)|z(t,x)|^2 \mathbb{E}\left[\lambda_s D_{(s,y)}\lambda_t\right] ds\nu(dy) dt\nu(dx) \\
&\geq 0, \quad \mathbb{P}-a.s.,
\end{aligned}$$

where for the last inequality we used Lemma 4.2 and the Assumption (3.2) on $z(t,x)$. To summarize, we have shown that

$$\mathbb{E}\big[F^3\big] \geq \mathbb{E}\left[\int_{\mathbb{R}_+\times\mathbb{R}} |z(t,x)||D_{(t,x)}F|^2 \lambda_t dt\nu(dx)\right]. \tag{4.14}$$

4.3 Proof of Theorem 3.4

Let $T > 0$, and note that we have $F_T = \delta^N(z\mathcal{Z})$ with $z(t,x) := \mathbf{1}_{\{t\in[0,T]\}} x/\sqrt{T}$, $(t,x) \in \mathbb{R}_+ \times \mathbb{R}$. By Relation (3.1) in Theorem 3.1 we have

$$d_W\big(F, \mathcal{N}_{\gamma^2}\big) \leq \mathbb{E}\left[\left|\gamma^2 - \frac{\vartheta^2}{T}\int_0^T \lambda_t dt\right|\right] + \frac{1}{\sqrt{T}}\mathbb{E}\left[\int_{\mathbb{R}}\int_0^T |x||D_{(t,x)}F_T|^2 \lambda_t dt\nu(dx)\right]. \tag{4.15}$$

On the other hand, by [HHKR21, Estimates on Term A_2-Proof of Theorem 3.10] we have

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

for any ϕ satisfying Assumption 3.3. Regarding the first term in the right hand-side of (4.15), we have

$$\mathbb{E}\left[\left|\gamma^2 - \frac{\vartheta^2}{T}\int_0^T \lambda_t dt\right|\right] \leq \left|\gamma^2 - \frac{\vartheta^2}{T}\int_0^T \mathbb{E}[\lambda_t] dt\right| + \frac{\vartheta^2}{T}\mathbb{E}\left[\left|\int_0^T (\lambda_t - \mathbb{E}[\lambda_t]) dt\right|\right]$$

$$= O(T^{-1/2}),$$

since

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

by [HHKR21, Lemma 4.1] and [HHKR21, Estimate on Term $A_{1,2}$ - Proof of Theorem 3.10]. □

As $z(t,x)$ is non-negative, instead of (3.1) we could also have applied (3.3) which involves the quantity $\mathbb{E}[F_T^3]$, with

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

by (4.14), which recovers the convergence with decay rate $T^{-1/2}$ using the decomposition (4.13), in which the first term coincides with the bound in (3.1) and the second term is proved to decay as $T^{-1/2}$ using [HHKR21, Lemma 4.2].

4.4 Proof of Theorem 3.7

Let $T > 0$. As $m = \int_{-\infty}^{\infty} x\nu(dx) = \mathbb{E}[X_1] = 0$, $S_T = \delta^N(z\mathcal{Z})$ with $z(t,x) = x\mathbf{1}_{\{t\in[0,T]\}}/\sqrt{T}$, using (A.1) in Appendix A it holds that :

$$\|S_T - \mathcal{N}(0,\gamma^2)\|_{4,\infty} \leq \sup_{f\in\mathcal{F}_W^4} \left|\mathbb{E}\left[\gamma^2 f'(S_T) - S_T f(S_T)\right]\right|,$$

with $\gamma^2 = \vartheta^2\mu/(1-\|\phi\|_1)$, where $\vartheta^2 = \mathbb{E}[X_1^2]$. Following the lines of the proof of Theorem 3.1 and using a Taylor expansion of order 3 for $D_{(t,x)}f(S_T)$, we have

$$f\left(S_T \circ \varepsilon^+_{(t,0,x)}\right) - f(S_T) = f'(S_T)D_{(t,x)}S_T + \frac{1}{2}f''(S_T)|D_{(t,x)}S_T|^2 + \frac{1}{6}f^{(3)}\left(\overline{S}^{t,x}\right)(D_{(t,x)}S_T)^3,$$

where $\overline{S}^{t,x}$ denotes a random element between S_T and $S_T \circ \varepsilon^+_{(t,0,x)}$. Relation (4.12) then becomes

$$\begin{aligned}\mathbb{E}\left[\gamma^2 f'(S_T) - f(S_T)S_T\right] = \mathbb{E}&\left[f'(S_T)\left(\gamma^2 - \int_{\mathbb{R}_+\times\mathbb{R}} z(t,x)D_{(t,x)}S_T\lambda_t dt\nu(dx)\right)\right] \\ &-\frac{1}{2}\mathbb{E}\left[\int_{\mathbb{R}_+\times\mathbb{R}} z(t,x)\lambda_t f''(S_T)|D_{(t,x)}S_T|^2 dt\nu(dx)\right] \\ &-\frac{1}{6}\mathbb{E}\left[\int_{\mathbb{R}_+\times\mathbb{R}} z(t,x)\lambda_t f^{(3)}\left(\overline{S}^{t,x}\right)(D_{(t,x)}S_T)^3 dt\nu(dx)\right]\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\left[f'(S_T)\left(\gamma^2 - \int_{\mathbb{R}_+\times\mathbb{R}} |z(t,x)|^2 \lambda_t dt\nu(dx)\right)\right] \\
&\quad -\frac{1}{2}\mathbb{E}\left[\int_{\mathbb{R}_+\times\mathbb{R}} z(t,x)\lambda_t f''(S_T)|D_{(t,x)}S_T|^2 dt\nu(dx)\right] \\
&\quad -\frac{1}{6}\mathbb{E}\left[\int_{\mathbb{R}_+\times\mathbb{R}} z(t,x)\lambda_t f^{(3)}\big(\overline{S}^{t,x}\big)(D_{(t,x)}S_T)^3 dt\nu(dx)\right],
\end{aligned}
$$

where we used Relation (2.9) for the Malliavin derivative of S_T, i.e.

$$
D_{(t,x)}S_T = z(t,x) + \delta^N\big(z\widehat{\mathcal{Z}}^t\big), \tag{4.16}
$$

and Lemma 4.1. It is important to notice that the quantity $\delta^N\big(z\widehat{\mathcal{Z}}^t\big)$ is independent of x as $\lambda_r \circ \varepsilon^+_{(t,0,x)}$ is (see 4.3), and thus we have

$$
\begin{aligned}
\delta^N\big(z\widehat{\mathcal{Z}}^t\big) &= \frac{1}{\sqrt{T}}\int_{(t,T]\times\mathbb{R}_+\times\mathbb{R}} y\mathbf{1}_{\{\lambda_r<\theta\le\lambda_r\circ\varepsilon^+_{(t,0,x)}\}}(N(dr,d\theta,dy) - drd\theta\nu(dy)) \\
&= \frac{1}{\sqrt{T}}\int_{(t,T]\times\mathbb{R}_+\times\mathbb{R}} y\mathbf{1}_{\{\lambda_r<\theta\le\lambda_r\circ\varepsilon^+_{(t,0,1)}\}}(N(dr,d\theta,dy) - drd\theta\nu(dy)).
\end{aligned}
$$

Using the definition of $z(t,x)$, we get

$$
\begin{aligned}
&\mathbb{E}\left[\gamma^2 f'(S_T) - f(S_T)S_T\right] = \vartheta^2\mathbb{E}\left[f'(S_T)\left(\frac{\mu}{1-\|\phi\|_1} - \frac{1}{T}\int_0^T \lambda_t dt\right)\right] \\
&\quad -\frac{1}{2\sqrt{T}}\mathbb{E}\left[\int_0^T\int_{-\infty}^{\infty} x\lambda_t f''(S_T)|D_{(t,x)}S_T|^2 dt\nu(dx)\right] \\
&\quad -\frac{1}{6\sqrt{T}}\mathbb{E}\left[\int_0^T\int_{-\infty}^{\infty} x\lambda_t f^{(3)}\big(\overline{S}^{t,x}\big)(D_{(t,x)}S_T)^3 dt\nu(dx)\right] \\
&= \vartheta^2\mathbb{E}\left[f'(S_T)\left(\frac{\mu}{1-\|\phi\|_1} - \frac{1}{T}\int_0^T \lambda_t dt\right)\right] - \frac{1}{2T^{3/2}}\int_{-\infty}^{\infty} x^3\nu(dx)\mathbb{E}\left[\int_0^T \lambda_t f''(S_T)dt\right] \\
&\quad -\frac{\int_{-\infty}^{\infty} x\nu(dx)}{2\sqrt{T}}\mathbb{E}\left[\int_0^T \lambda_t f''(S_T)\big|\delta^N\big(z\widehat{\mathcal{Z}}^t\big)\big|^2 dt\right] - \frac{\vartheta^2}{T}\mathbb{E}\left[\int_0^T \lambda_t\mathbb{E}_t\big[f''(S_T)\delta^N\big(z\widehat{\mathcal{Z}}^t\big)\big]dt\right] \\
&\quad -\frac{1}{6\sqrt{T}}\mathbb{E}\left[\int_0^T\int_{-\infty}^{\infty} x\lambda_t f^{(3)}\big(\overline{S}^{t,x}\big)(D_{(t,x)}S_T)^3\nu(dx)dt\right] \\
&= \vartheta^2\mathbb{E}\left[f'(S_T)\left(\frac{\mu}{1-\|\phi\|_1} - \frac{1}{T}\int_0^T \lambda_t dt\right)\right] \\
&\quad -\frac{1}{6\sqrt{T}}\mathbb{E}\left[\int_0^T\int_{-\infty}^{\infty} x\lambda_t f^{(3)}\big(\overline{S}^{t,x}\big)(D_{(t,x)}S_T)^3\nu(dx)dt\right],
\end{aligned}
$$

where we used once again Lemma 4.1 and the fact that $\int_{-\infty}^{\infty} x\nu(dx) = \int_{-\infty}^{\infty} x^3\nu(dx) = 0$. Hence, we have

$$
\begin{aligned}
\left|\mathbb{E}\left[\gamma^2 f'(S_T) - f(S_T)S_T\right]\right| &\le \vartheta^2 \|f'\|_\infty \left|\frac{\mu}{1-\|\phi\|_1} - \frac{1}{T}\int_0^T \mathbb{E}[\lambda_t]dt\right| \\
&+\frac{1}{T}\left|\mathbb{E}\left[f'(S_T)\int_0^T(\lambda_t - \mathbb{E}[\lambda_t])dt\right]\right| + \frac{2\|f^{(3)}\|_\infty \int_{-\infty}^{\infty} x^4\nu(dx)}{3T^2}\mathbb{E}\left[\int_0^T \lambda_t dt\right] \\
&+\frac{2\|f^{(3)}\|_\infty \int_{-\infty}^{\infty} |x|\nu(dx)}{3\sqrt{T}}\mathbb{E}\left[\int_0^T \lambda_t \left|\delta^N\left(z\widehat{\mathcal{Z}}^t\right)\right|^3 dt\right] \\
&= \vartheta^2 \|f'\|_\infty \left|\frac{\mu}{1-\|\phi\|_1} - \frac{1}{T}\int_0^T \mathbb{E}[\lambda_t]dt\right| + \frac{1}{T}\left|\mathbb{E}\left[f'(S_T)\int_0^T(\lambda_t - \mathbb{E}[\lambda_t])dt\right]\right| \\
&+\frac{2\|f^{(3)}\|_\infty \int_{-\infty}^{\infty} x^4\nu(dx)}{3T^2}\mathbb{E}\left[\int_0^T \lambda_t dt\right] + \frac{2\|f^{(3)}\|_\infty \int_{-\infty}^{\infty} |x|\nu(dx)}{3T^2}\mathbb{E}\left[\int_0^T |R_{t,T}|^3\lambda_t dt\right] \\
&=: A_1 + A_2 + A_3 + A_4, \qquad (4.17)
\end{aligned}
$$

where we set

$$
R_{t,T} := \sqrt{T}\delta^N\left(z\widehat{\mathcal{Z}}^t\right) = \int_{(t,T]\times\mathbb{R}_+\times\mathbb{R}} y\mathbf{1}_{\{\lambda_r < \theta \le \lambda_r \circ \varepsilon^+_{(t,0,1)}\}} \left(N(dr,d\theta,dy) - drd\theta\nu(dy)\right).
$$

We now treat the above three terms separately.

First, note that by [HHKR21, Lemma 4.1] we have $A_1 = O(T^{-1})$. In addition, as $\mathbb{E}\left[\int_0^T \lambda_t dt\right] = O(T)$ and since $\mathbb{E}_t\left[|R_{t,T}|^3\right]$ is bounded uniformly in t, T by Lemma 4.4, we have $A_3 + A_4 = O(T^{-1})$. It remains to deal with the term A_2. Using [BDHM13, Relation (14)] and [HHKR21, Proof of Theorem 3.10, Term $A_{1,2}$], we get that

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

where $\mathcal{M}_s := H_s - \int_0^s \lambda_u du$. Hence, using (4.9), we have

$$
\begin{aligned}
\mathbb{E}\left[f'(S_T)\int_0^T(\lambda_t - \mathbb{E}[\lambda_t])dt\right] &= \mathbb{E}\left[f'(S_T)\int_0^T \psi(T-s)\mathcal{M}_s ds\right] \\
&= \int_0^T \psi(T-s)\mathbb{E}\left[f'(S_T)\mathcal{M}_s\right]ds \\
&= \int_0^T \psi(T-s)\mathbb{E}\left[f'(S_T)\delta^N(\mathcal{Z}1_{\{\cdot \le s\}})\right]ds \\
&= \int_0^T \psi(T-s)\int_0^s\int_{\mathbb{R}} \mathbb{E}\left[\lambda_u D_{(u,x)} f'(S_T)\right]\nu(dx)duds
\end{aligned}
$$

$$
\begin{aligned}
&= \int_0^T \psi(T-s)\int_0^s \int_{\mathbb{R}} \mathbb{E}\left[\lambda_u f''(S_T) D_{(u,x)} S_T\right] \nu(dx) du ds \\
&\quad + \frac{1}{2}\int_0^T \psi(T-s)\int_0^s \int_{\mathbb{R}} \mathbb{E}\left[\lambda_u f^{(3)}\big(\overline{S}^{u,x}\big)|D_{(u,x)} S_T|^2\right] \nu(dx) du ds \\
&= \frac{1}{\sqrt{T}}\int_{\mathbb{R}} x\nu(dx)\int_0^T \psi(T-s)\int_0^s \mathbb{E}\left[\lambda_u f''(S_T)\right] du ds \\
&\quad + \int_0^T \psi(T-s)\int_0^s \int_{\mathbb{R}} \mathbb{E}\big[\lambda_u \mathbb{E}_u\big[f''(S_T)\delta^N\big(z\widehat{\mathcal{Z}}^u\big)\big]\big]\nu(dx) du ds \\
&\quad + \frac{1}{2}\int_0^T \psi(T-s)\int_0^s \int_{\mathbb{R}} \mathbb{E}\left[\lambda_u f^{(3)}\big(\overline{S}^{u,x}\big)|D_{(u,x)} S_T|^2\right] \nu(dx) du ds \\
&= \frac{1}{2}\int_0^T \psi(T-s)\int_0^s \int_{\mathbb{R}} \mathbb{E}\left[\lambda_u f^{(3)}\big(\overline{S}^{u,x}\big)|D_{(u,x)} S_T|^2\right] \nu(dx) du ds,
\end{aligned}
$$

as $\int_{\mathbb{R}} x\nu(dx) = \mathbb{E}[X_1] = 0$ and $\mathbb{E}_u\big[f''(S_T)\delta^N\big(z\widehat{\mathcal{Z}}^u\big)\big] = 0$ by Lemma 4.1, where $\overline{S}^{u,x}$ denotes a random element between S_T and $S_T \circ \varepsilon^+_{(u,0,x)}$. Hence, by (4.16) we have

$$
\begin{aligned}
&\mathbb{E}\left[f'(S_T)\int_0^T (\lambda_t - \mathbb{E}[\lambda_t])dt\right] \\
&= \frac{1}{2}\int_0^T \psi(T-s)\int_0^s \int_{\mathbb{R}} \mathbb{E}\big[\lambda_u f^{(3)}\big(\overline{S}^{u,x}\big)|D_{(u,x)} S_T|^2\big]\nu(dx) du ds \\
&\leq \frac{1}{2}\|f^{(3)}\|_\infty \int_0^T \psi(T-s)\int_0^s \int_{\mathbb{R}} \mathbb{E}\big[\lambda_u |D_{(u,x)} S_T|^2\big]\nu(dx) du ds \\
&\leq \frac{1}{T}\|f^{(3)}\|_\infty \|\psi\|_1 \int_{\mathbb{R}} x^2\nu(dx)\int_0^T \mathbb{E}\left[\lambda_u\right] du \\
&\quad + \|f^{(3)}\|_\infty \int_0^T \psi(T-s)\int_0^T \mathbb{E}\big[\lambda_u \mathbb{E}_u\big[\big|\delta^N\big(z\widehat{\mathcal{Z}}^u\big)\big|^2\big]\big] du ds.
\end{aligned}
$$

Next, by the Itô isometry, see *e.g.*, [Pri09, Proposition 6.5.4] and (4.6), for some constant $C > 0$ it holds that

$$
\begin{aligned}
\mathbb{E}_u\big[\big|\delta^N\big(\widehat{\mathcal{Z}}^u\big)\big|^2\big] &= \frac{1}{T}\int_{\mathbb{R}} x^2\nu(dx)\int_u^T \int_0^{+\infty} \mathbb{E}_u\big[\big|\mathbf{1}_{\{\lambda_r < \theta \leq \lambda_r \circ \varepsilon^+_{(t,0,1)}\}}\big|^2\big] d\theta ds \\
&= \frac{1}{T}\int_{\mathbb{R}} x^2\nu(dx)\int_u^T \mathbb{E}_u\big[\widehat{\lambda}^u_s\big] ds \\
&\leq \frac{C}{T}, \qquad u \in [0,T].
\end{aligned}
$$

Thus, for some constant $C > 0$ we have

$$
\operatorname{esssup}_{T>0}\left|\mathbb{E}\left[f'(S_T)\int_0^T (\lambda_t - \mathbb{E}[\lambda_t])dt\right]\right| \leq C\|f^{(3)}\|_\infty \|\psi\|_1 \int_{\mathbb{R}} x^2\nu(dx)\, \sup_{T>0}\int_0^T \frac{\mathbb{E}[\lambda_u]}{T} du
$$

as $(\mathbb{E}[\lambda_u])_{u\in\mathbb{R}_+}$ is continuous by [BDHM13, Lemma 3] and $\lim_{T\to+\infty} T^{-1}\int_0^T \mathbb{E}[\lambda_u]\,du = \mu/(1-\|\phi\|_1)$, which shows that $A_2 = O(T^{-1})$ and concludes the proof.

4.5 Proof of Theorem 3.8

We start by proving the upper bound on the Wasserstein distance between the distribution of

$$V_T := \frac{S_T - m\int_0^T \mathbb{E}[\lambda_s]ds}{\sqrt{T}}$$

and $\mathcal{N}(0,\zeta^2)$. For this, we consider the normalized martingale

$$F_t := \frac{S_t - m\int_0^t \lambda_s ds}{\sqrt{t}}$$

associated to the compound process S, we write

$$\begin{aligned} F_T &= \frac{1}{\sqrt{T}}\left(S_T - m\int_0^T \lambda_s ds\right) \\ &= \frac{1}{\sqrt{T}}(S_T - mH_T) + m\frac{\mathcal{M}_T}{\sqrt{T}}, \end{aligned}$$

where $\mathcal{M}_T := H_T - \int_0^T \lambda_u du$. Similarly, we have

$$\begin{aligned} V_T &= \frac{1}{\sqrt{T}}\left(S_T - m\int_0^T \mathbb{E}[\lambda_s]ds\right) \\ &= \frac{1}{\sqrt{T}}(S_T - mH_T) + mY_T, \end{aligned}$$

with $Y_T := (H_T - \int_0^T \mathbb{E}[\lambda_s]ds)/\sqrt{T}$. Thus,

$$(1-\|\phi\|_1)V_T - F_T = -\|\phi\|_1\frac{S_T - mH_T}{\sqrt{T}} + m(1-\|\phi\|_1)\mathfrak{R}_T,$$

where we let

$$\mathfrak{R}_T := Y_T - \frac{\mathcal{M}_T}{\sqrt{T}(1-\|\phi\|_1)}.$$

Next, we note that

$$\delta^N\left(((x-m)\mathbb{1}_{t\le T})_{(t,x)\in(\mathbb{R}_+\times\mathbb{R})}\mathcal{Z}\right) = \int_0^T\int_{\mathbb{R}_+\times\mathbb{R}} (x-m)\mathbf{1}_{\{\theta\le\lambda_t\}}\left(N(dt,d\theta,dx) - dtd\theta\nu(dx)\right)$$

$$= \int_0^T\int_{\mathbb{R}_+\times\mathbb{R}} (x-m)\mathbf{1}_{\{\theta\le\lambda_t\}}N(dt,d\theta,dx) - \int_0^T\int_{\mathbb{R}_+\times\mathbb{R}} x\mathbf{1}_{\{\theta\le\lambda_t\}}dtd\theta\nu(dx)$$

$$+ m\int_0^T \int_{\mathbb{R}_+\times\mathbb{R}} \mathbf{1}_{\{\theta\le\lambda_t\}} dt d\theta \nu(dx)$$
$$= S_T - mH_T.$$

Hence, using the fact that F is also written as a divergence, we have

$$(1-\|\phi\|_1)V_T = \delta^N\big((z(t,x))_{(t,x)\in(\mathbb{R}_+\times\mathbb{R})}\mathcal{Z}\big) + m(1-\|\phi\|_1)\mathfrak{R}_T,$$

where

$$z(t,x) := \frac{x+(m-x)\|\phi\|_1}{\sqrt{T}}\mathbb{1}_{[0,T]}(t).$$

We now proceed in the same manner as the proof of Theorem 3.4, only by replacing x by $(1-\|\phi\|_1)x + \|\phi\|_1 m$. This proves that $\delta^N\big((z(t,x))_{(t,x)\in(\mathbb{R}_+\times\mathbb{R})}\mathcal{Z}\big)$ converges to a centered Gaussian random variable of variance

$$\begin{aligned}
&\frac{\mu}{1-\|\phi\|_1}\int_{\mathbb{R}}((1-\|\phi\|_1)x + \|\phi\|_1 m)^2\nu(dx)\\
&= \frac{\mu}{1-\|\phi\|_1}\int_{\mathbb{R}}((1-\|\phi\|_1)^2x^2 + 2(1-\|\phi\|_1)\|\phi\|_1 xm + \|\phi\|_1^2 m^2)\nu(dx)\\
&= \frac{\mu}{1-\|\phi\|_1}\big((1-2\|\phi\|_1+\|\phi\|_1^2)\vartheta^2 + 2(1-\|\phi\|_1)\|\phi\|_1 m^2 + \|\phi\|_1^2 m^2\big)\\
&= \frac{\mu}{1-\|\phi\|_1}\big(\vartheta^2 + \|\phi\|_1(\vartheta^2-m^2)(\|\phi\|_1-2)\big)
\end{aligned}$$

where the Wasserstein distance between the two variables is bounded by $O(T^{-1/2})$. By proceeding as in the proof of Theorem 3.13 in [HHKR21] it is enough to show that $\mathbb{E}[\mathfrak{R}_T^2] = O(T^{-1})$, which is done in Lemma 4.5, to obtain

$$d_W\big(V_T, \mathcal{N}(0,\zeta^2)\big) \le \frac{C}{\sqrt{T}}.$$

Finally, by [BDHM13, Lemma 5], we have

$$V_T - \Gamma_T = \frac{m}{\sqrt{T}}\left(\int_0^T \mathbb{E}[\lambda_t]dt - \frac{\mu T}{1-\|\phi\|_1}\right) = O(T^{-1/2}),$$

which yields the desired result.

A Elements of Stein's method

In this section we describe elements of Stein's method, which was has been introduced by C.M. Stein in [Ste72], that are relevant to our analysis and to the derivation of bounds of the form (A.1). We let $h^{(i)}$, $i \ge 1$, denotes the ith derivative of h.

Definition A.1. *In what follows $\mathcal{H}$ denotes one of the following Hilbert spaces :*

1. $\mathcal{H}_W := \{h : \mathbb{R} \to \mathbb{R}$ *differentiable a.e. with* $\|h'\|_\infty \leq 1\}$,
2. $\mathcal{H}_{4,\infty} := \{h : \mathbb{R} \to \mathbb{R}$ *four times differentiable a.e. with* $\max_{1\leq i\leq 4} \|h^{(i)}\|_\infty \leq 1\}$.

Given F and G two random variables on a probability space $(\Omega, \mathcal{F}_\infty^N, \mathbb{P})$, we let

$$d_{\mathcal{H}}(F,G) := \sup_{h\in\mathcal{H}} |\mathbb{E}[h(F) - h(G)]|$$

denote the distance (with respect to the class of test functions $\mathcal{H}$) between the laws $\mathcal{L}_F$ and $\mathcal{L}_G$ of F and G. In addition,

1. *if $\mathcal{H} = \mathcal{H}_W$ we write d_W for $d_{\mathcal{H}_W}$ and corresponds to the Wasserstein distance;*
2. *if $\mathcal{H} = \mathcal{H}_{4,\infty}$ we write $d_{4,\infty}$ for $d_{\mathcal{H}_{4,\infty}}$.*

We also set

1. $\mathcal{F}_W := \{f : \mathbb{R} \to \mathbb{R}$ *twice differentiable map* , $\|f'\|_\infty \leq 1$, $\|f''\|_\infty \leq 2\}$,
2. $\mathcal{F}_{4,\infty} := \{f : \mathbb{R} \to \mathbb{R}$ *four times differentiable map* , $\|f^{(i)}\|_\infty \leq 1/i,\, i = 1,2,3,4\}$,

and

$$\mathcal{F}_{\mathcal{H}} := \begin{cases} \mathcal{F}_W \text{ if } \mathcal{H} = \mathcal{H}_W, \\ \mathcal{F}_{4,\infty} \text{ if } \mathcal{H} = \mathcal{H}_{4,\infty}. \end{cases}$$

Let $\mathcal{N}_{\sigma^2} \sim \mathcal{N}(0,\sigma^2)$, let $\mathcal{H}$ be one of the spaces in 1.-2. above, and let h in $\mathcal{H}$. C.M. Stein proved in [Ste72] (see also [CGS11, Lemma 2.6 and Section 4.8] for the $\mathcal{H}_{4,\infty}$ distance), that there exists a function f_h in $\mathcal{F}_W$ solution to the functional Stein equation

$$h(x) - \mathbb{E}[h(\mathcal{N}_{\sigma^2})] = \sigma^2 f_h'(x) - x f_h(x), \quad x \in \mathbb{R}.$$

For F a centered random variable, plugging F in this equation and taking expectations, we get

$$|\mathbb{E}[h(F) - h(\mathcal{N}_{\sigma^2})]| = \left|\mathbb{E}[\sigma^2 f_h'(F) - F f_h(F)]\right|,$$

which yields

$$d_{\mathcal{H}}(F, \mathcal{N}_{\sigma^2}) \leq \sup_{f\in\mathcal{F}_{\mathcal{H}}} \left|\mathbb{E}[\sigma^2 f'(F) - F f(F)]\right|. \tag{A.1}$$

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

References

[BDHM13] 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.

[BM96] P. Brémaud and L. Massoulié. Stability of nonlinear Hawkes processes. *Annals of Probability*, 24(3):1563–1588, 1996.

[CGMT20] 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.

[CGS11] L. H.Y. Chen, L. Goldstein, and Q.-M. Shao. *Normal Approximation by Stein's Method.* Springer Berlin Heidelberg, 2011.

[DVJ88] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes.* Springer. New York, 1988.

[Dun21] N.T. Dung. *Rates of Convergence in the Central Limit Theorem for Nonlinear Statistics Under Relaxed Moment Conditions. Acta Math Vietnam*, 2021.

[Haw71] A. G. Hawkes. Spectra of some self-excited and mutually exciting point process. *Biometrika*, 1:83–90, 1971.

[HHKR21] C. Hillairet, L. Huang, M. Khabou, and A. Réveillac. The Malliavin-Stein method for Hawkes functionals. *Preprint*, 2021.

[HRR20] C. Hillairet, A. Réveillac, and M. Rosenbaum. An expansion formula for Hawkes processes and application to cyber-insurance derivatives. Preprint, 2020.

[NP09] I. Nourdin and G. Peccati. Stein's method on Wiener chaos. *Probab. Theory Related Fields*, 145:75–118, 2009.

[Oga81] Y. Ogata. On Lewis' simulation method for point processes. *IEEE Transactions of Information Theory*, 27:23–31, 1981.

[Pic96] J. Picard. Formules de dualité sur l'espace de Poisson. *Ann. Inst. H. Poincaré Probab. Statist.*, 32(4):509–548, 1996.

[Pri09] N. Privault. *Stochastic analysis in discrete and continuous settings with normal martingales*, volume 1982 of *Lecture Notes in Mathematics.* Springer-Verlag, Berlin, 2009.

[Pri18] N. Privault. Stein approximation for multidimensional poisson random measures by third cumulant expansions. *ALEA - Latin American Journal of Probability and Mathematical Statistics*, 15:1141–1161, 2018.

[Pri19] N. Privault. Third cumulant Stein approximation for Poisson stochastic integrals. *J. Theoret. Probab.*, 32:1461–1481, 2019.

[PSTU10] 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.

[Ste72] 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.

[Tor16] G. L. Torrisi. Gaussian approximation of nonlinear Hawkes processes. *The Annals of Applied Probability*, 26(4):2106–2140, 2016.