CENTRAL LIMIT THEOREM FOR NONLINEAR HAWKES PROCESSES

LINGJIONG ZHU,* *New York University*

Abstract

The Hawkes process is a self-exciting point process with clustering effect whose intensity depends on its entire past history. It has wide applications in neuroscience, finance, and many other fields. In this paper we obtain a functional central limit theorem for the nonlinear Hawkes process. Under the same assumptions, we also obtain a Strassen's invariance principle, i.e. a functional law of the iterated logarithm.

Keywords: Central limit theorem; functional central limit theorem; point process; Hawkes process; self-exciting process

2010 Mathematics Subject Classification: Primary 60G55; 60F05

1. Introduction and main results

1.1. Introduction

The Hawkes process is a self-exciting simple point process first introduced by Hawkes [10]. The future evolution of a self-exciting point process is influenced by the timing of past events. The process is non-Markovian except for some very special cases. In other words, the Hawkes process depends on the entire past history and has a long memory. The Hawkes process has wide applications in neuroscience, seismology, genome analysis, finance, and many other fields. It has both self-exciting and clustering properties, which is very appealing to some financial applications. Errais *et al.* [9] stated the following.

> The collapse of Lehman Brothers brought the financial system to the brink of a breakdown. The dramatic repercussions point to the existence of feedback phenomena that are channeled through the complex web of informational and contractual relationships in the economy …. This and related episodes motivate the design of models of correlated default timing that incorporate the feedback phenomena that plague credit markets.

The self-exciting and clustering properties of the Hawkes process make it a viable candidate in modeling the correlated defaults and evaluating the credit derivatives in finance; see, for example, [8] and [9].

Most of the literature on Hawkes processes has been restricted to the linear case, which has an immigration-birth representation (see [11]). The stability, law of large numbers, central limit theorem, large deviations, Bartlett spectrum, etc. have all been studied and well understood. Almost all of the applications of the Hawkes process in the literature focus exclusively on the linear case. Because of the lack of immigration-birth representation and computational tractability, the nonlinear Hawkes process is much less studied. However, some efforts have already been made in this direction; see, for instance, [4], [15], and [16]. In this paper, we will

prove a functional central limit theorem for the nonlinear Hawkes process. Hopefully, in the future, nonlinear Hawkes processes will also be used in applications in various fields.

For a list of references on the theories and applications of the Hawkes process, we refer the reader to [6], [7], and [14].

1.2. Nonlinear Hawkes processes

Let N be a simple point process on $\mathbb{R}$, and let $\mathcal{F}_t^{-\infty} := \sigma(N(C), C \in \mathcal{B}(\mathbb{R}), C \subset (-\infty, t])$ be an increasing family of σ-algebras. Any nonnegative $\mathcal{F}_t^{-\infty}$-progressively measurable process λ_t with

$$\mathbb{E}[N(a, b] \mid \mathcal{F}_a^{-\infty}] = \mathbb{E}\left[\int_a^b \lambda_s \, \mathrm{d}s \,\middle|\, \mathcal{F}_a^{-\infty}\right]$$

almost surely (a.s.) for all intervals $(a, b]$ is called an $\mathcal{F}_t^{-\infty}$-intensity of N. We use the notation $N_t := N(0, t]$ to denote the number of points in the interval $(0, t]$.

A general Hawkes process is a simple point process N admitting an $\mathcal{F}_t^{-\infty}$-intensity

$$\lambda_t := \lambda\left(\int_{-\infty}^t h(t - s) N(\mathrm{d}s)\right), \tag{1}$$

where $\lambda(\cdot)\colon \mathbb{R}^+ \to \mathbb{R}^+$ is locally integrable and left continuous, $h(\cdot)\colon \mathbb{R}^+ \to \mathbb{R}^+$, and we always assume that $\|h\|_{L^1} = \int_0^\infty h(t) \, \mathrm{d}t < \infty$. In (1), $\int_{-\infty}^t h(t - s) N(\mathrm{d}s)$ stands for $\int_{(-\infty, t)} h(t - s) N(\mathrm{d}s) = \sum_{\tau < t} h(t - \tau)$, where the τ are the occurrences of the points before time t. In the literature, $h(\cdot)$ and $\lambda(\cdot)$ are usually referred to as the exciting function and the rate function, respectively. A Hawkes process is linear if $\lambda(\cdot)$ is linear and it is nonlinear otherwise.

Brémaud and Massoulié [4] proved that, under the assumption that $\lambda(\cdot)$ is α-Lipschitz with $\alpha\|h\|_{L^1} < 1$, there exists a unique stationary and ergodic version of the Hawkes process satisfying the dynamics (1). They also studied the stability of the nonlinear Hawkes process in great detail, including the existence, uniqueness, stability in distribution and in variation, etc. Later, Brémaud *et al.* [5] studied the rate of convergence of the nonlinear Hawkes process to its stationary version.

1.3. Limit theorems for Hawkes processes

When $\lambda(\cdot)$ is linear, say $\lambda(z) = \nu + z$ for some $\nu > 0$, and $\|h\|_{L^1} < 1$, the Hawkes process has a very nice immigration-birth representation; see, for example, [11]. For the linear Hawkes process, limit theorems are very well understood. There is the law of large numbers (see, for instance, [6] and [7]), i.e.

$$\frac{N_t}{t} \to \frac{\nu}{1 - \|h\|_{L^1}} \quad \text{as } t \to \infty \text{ a.s.}$$

Moreover, Bordenave and Torrisi [3] proved a large deviation principle for $(N_t/t \in \cdot)$ with rate function

$$I(x) = \begin{cases} x \log\left(\dfrac{x}{\nu + x\|h\|_{L^1}}\right) - x + x\|h\|_{L^1} + \nu & \text{if } x \in [0, \infty), \\ +\infty & \text{otherwise.} \end{cases}$$

Recently, Bacry *et al.* [1] proved a functional central limit theorem for the linear multivariate Hawkes process under certain assumptions. They included the linear Hawkes process as a special case and proved that

$$\frac{N_{\cdot t} - \cdot \mu t}{\sqrt{t}} \to \sigma B(\cdot) \quad \text{as } t \to \infty,$$

where $B(\cdot)$ is a standard Brownian motion. The convergence is the weak convergence on $D[0, 1]$, the space of càdlàg functions on $[0, 1]$, equipped with the Skorokhod topology. Here

$$\mu = \frac{\nu}{1 - \|h\|_{L^1}} \quad \text{and} \quad \sigma^2 = \frac{\nu}{(1 - \|h\|_{L^1})^3}.$$

Very recently, Karabash and Zhu [13] obtained a central limit theorem and large deviation principle for the linear Hawkes process with random marks. In a nutshell, the linear Hawkes process satisfies very nice limit theorems and the limits can be computed more or less explicitly.

When $\lambda(\cdot)$ is nonlinear, the usual immigration-birth representation no longer works and you may have to use some abstract theory to obtain limit theorems. Some progress has already been made for the nonlinear Hawkes process.

Brémaud and Massoulié's [4] stability result implies that, by the erogdic theorem,

$$\frac{N_t}{t} \to \mu := \mathbb{E}[N[0, 1]]$$

as $t \to \infty$, where $\mathbb{E}[N[0, 1]]$ is the mean of $N[0, 1]$ under the stationary and ergodic measure.

When $h(\cdot)$ is exponential (and $\lambda(\cdot)$ is nonlinear), the Hawkes process is Markovian and Zhu [15] obtained a large deviation principle for $(N_t/t \in \cdot)$ in this case. Zhu [15] also proved the large deviation principle for the case when $h(\cdot)$ is a sum of exponentials and used that as an approximation to recover the result for the linear case proved in [3].

For the most general $h(\cdot)$ and $\lambda(\cdot)$, Zhu [16] proved a process-level, i.e. level-3, large deviation principle for the Hawkes process and used the contraction principle to obtain a large deviation principle for $(N_t/t \in \cdot)$.

In this paper we will prove a functional central limit theorem and a functional law of the iterated logarithm for the nonlinear Hawkes process.

1.4. Main results

The following is the assumption we will use throughout this paper.

Assumption 1. *We assume that*

- $h(\cdot) \colon [0, \infty) \to \mathbb{R}^+$ *is a decreasing function and* $\int_0^\infty th(t)\, dt < \infty$,

- $\lambda(\cdot)$ *is positive and increasing and α-Lipschitz (i.e. $|\lambda(x) - \lambda(y)| \le \alpha|x - y|$ for any x, y) such that $\alpha\|h\|_{L^1} < 1$.*

Brémaud and Massoulié [4] proved that if $\lambda(\cdot)$ is α-Lipschitz with $\alpha\|h\|_{L^1} < 1$, there exists a unique stationary and ergodic Hawkes process satisfying the dynamics (1). Hence, under our Assumption 1 (which is slightly stronger than that in [4]), there exists a unique stationary and ergodic Hawkes process satisfying the dynamics (1).

Let $\mathbb{P}$ and $\mathbb{E}$ denote the probability measure and expectation for a stationary, ergodic Hawkes process, and let $\mathbb{P}(\cdot \mid \mathcal{F}_0^{-\infty})$ and $\mathbb{E}[\cdot \mid \mathcal{F}_0^{-\infty}]$ denote the conditional probability measure and conditional expectation for the Hawkes process given the past history.

The following theorems are the main results of this paper.

Theorem 1. *Under Assumption 1, let N be the stationary and ergodic nonlinear Hawkes process with dynamics (1). We have*

$$\frac{N_{\cdot t} - \cdot \mu t}{\sqrt{t}} \to \sigma B(\cdot) \quad \text{as } t \to \infty \tag{2}$$

with $B(\cdot)$ a standard Brownian motion and $0 < \sigma < \infty$, where

$$\sigma^2 := \mathbb{E}[(N[0, 1] - \mu)^2] + 2\sum_{j=1}^{\infty} \mathbb{E}[(N[0, 1] - \mu)(N[j, j + 1] - \mu)]. \tag{3}$$

The convergence in (2) is the weak convergence on $D[0, 1]$, the space of càdlàg functions on $[0, 1]$, equipped with the Skorokhod topology.

Remark 1. By a standard central limit theorem for martingales, i.e. Theorem 4 below, it is easy to see that

$$\frac{N_{\cdot t} - \int_0^{\cdot t} \lambda_s \, ds}{\sqrt{t}} \to \sqrt{\mu} B(\cdot) \quad \text{as } t \to \infty,$$

where $\mu = \mathbb{E}[N[0, 1]]$. In the linear case, say $\lambda(z) = \nu + z$, Bacry *et al.* [1] proved that σ^2 in (3) satisfies $\sigma^2 = \nu/(1 - \|h\|_{L^1})^3 > \mu = \nu/(1 - \|h\|_{L^1})$. This is not surprising because $N_{\cdot t} - \cdot \mu t$ 'should' have more fluctuations than $N_{\cdot t} - \int_0^{\cdot t} \lambda_s \, ds$. Therefore, we guess that, for nonlinear $\lambda(\cdot)$, σ^2 defined in (3) should also satisfy $\sigma^2 > \mu = \mathbb{E}[N[0, 1]]$. However, it might not be very easy to compute and say something about σ^2 in such a case.

In the classical case for a sequence of independent and identically distributed random variables X_i with mean 0 and variance 1, we have the central limit theorem $n^{-1/2}\sum_{i=1}^{n} X_i \to N(0, 1)$ as $n \to \infty$, and we also have $\sum_{i=1}^{n} X_i/\sqrt{n \log \log n} \to 0$ in probability as $n \to \infty$, but the convergence does not hold a.s. The law of the iterated logarithm says that

$$\limsup_{n \to \infty} \frac{\sum_{i=1}^{n} X_i}{\sqrt{n \log \log n}} = \sqrt{2} \quad \text{a.s.}$$

A functional version of the law of the iterated logarithm is called Strassen's invariance principle. It turns out that we also have a Strassen's invariance principle for nonlinear Hawkes processes under Assumption 1.

Theorem 2. *Under Assumption 1, let N be the stationary and ergodic nonlinear Hawkes process with dynamics (1). Let $X_n := N[n - 1, n] - \mu$, $S_n := \sum_{i=1}^{n} X_i$, $s_n^2 := \mathbb{E}[S_n^2]$, $g(t) = \sup\{n : s_n^2 \le t\}$, and, for $t \in [0, 1]$, let $\eta_n(t)$ be the usual linear interpolation, i.e.*

$$\eta_n(t) = \frac{S_k + (s_n^2 t - s_k^2)(s_{k+1}^2 - s_k^2)^{-1} X_{k+1}}{\sqrt{2s_n^2 \log \log s_n^2}}, \qquad s_k^2 \le s_n^2 t \le s_{k+1}^2, \ k = 0, 1, \ldots, n - 1.$$

Then $g(e) < \infty$, $\{\eta_n, \ n > g(e)\}$ is relatively compact in $C[0, 1]$, the set of continuous functions on $[0, 1]$ is equipped with the uniform topology, and the set of limit points is the set of absolutely continuous functions $f(\cdot)$ on $[0, 1]$ such that $f(0) = 0$ and $\int_0^1 f'(t)^2 \, dt \le 1$.

2. Proof of Theorem 1

This section is devoted to the proof of Theorem 1. We use a standard central limit theorem, i.e. Theorem 3 below. In our proof, we need the fact that $\mathbb{E}[N[0, 1]^2] < \infty$, which is proved in Lemma 2 below. Lemma 2 is proved by proving a stronger result first, i.e. Lemma 1 below. We will also prove Lemma 3 below to guarantee that $\sigma > 0$, so that the central limit theorem is not degenerate.

Let us first quote the two necessary central limit theorems from [2]. In both Theorem 3 and Theorem 4, the filtrations are the natural ones, i.e. given a stochastic process $(X_n)_{n\in\mathbb{Z}}$, $\mathcal{F}_b^a := \sigma(X_n, \ a \le n \le b)$ for $-\infty \le a \le b \le \infty$.

Theorem 3. ([2, p. 197].) *Suppose that X_n, $n \in \mathbb{Z}$, is an ergodic stationary sequence such that $\mathbb{E}[X_n] = 0$ and*

$$\sum_{n \geq 1} \|\mathbb{E}[X_0 \mid \mathcal{F}_{-n}^{-\infty}]\|_2 < \infty,$$

where $\|Y\|_2 = (\mathbb{E}[Y^2])^{1/2}$. Let $S_n = X_1 + \cdots + X_n$. Then $S_{[n \cdot]}/\sqrt{n} \to \sigma B(\cdot)$ weakly, where the weak convergence is on $D[0, 1]$ equipped with the Skorokhod topology and $\sigma^2 = \mathbb{E}[X_0^2] + 2 \sum_{n=1}^{\infty} \mathbb{E}[X_0 X_n]$. The series converges absolutely.

Theorem 4. ([2, p. 196].) *Suppose that X_n, $n \in \mathbb{Z}$, is an erogdic, stationary sequence of square-integrable martingale differences, i.e. $\sigma^2 = \mathbb{E}[X_n^2] < \infty$, and let $\mathbb{E}[X_n \mid \mathcal{F}_{n-1}^{-\infty}] = 0$. Let $S_n = X_1 + \cdots + X_n$. Then $S_{[n \cdot]}/\sqrt{n} \to \sigma B(\cdot)$ weakly, where the weak convergence is on $D[0, 1]$ equipped with the Skorokhod topology.*

Now we are ready to prove our main result.

Proof of Theorem 1. Since in the stationary regime, $\mathbb{E}[N[n, n+1]] = \mathbb{E}[N[0, 1]]$ for any $n \in \mathbb{Z}$, let $\mathbb{E}[N[0, 1]] = \mu$. In order to apply Theorem 3, let us first prove that

$$\sum_{n=1}^{\infty} \{\mathbb{E}[(\mathbb{E}[N(n, n+1] - \mu \mid \mathcal{F}_0^{-\infty}])^2]\}^{1/2} < \infty. \tag{4}$$

Let $\mathbb{E}^{\omega_1^-}[N(n, n+1]]$ and $\mathbb{E}^{\omega_2^-}[N(n, n+1]]$ be two independent copies of $\mathbb{E}[N(n, n+1] \mid \mathcal{F}_0^{-\infty}]$. It is easy to check that

$$\frac{1}{2}\mathbb{E}[(\mathbb{E}^{\omega_1^-}[N(n, n+1]] - \mathbb{E}^{\omega_2^-}[N(n, n+1]])^2]$$
$$= \frac{1}{2}\mathbb{E}[\mathbb{E}^{\omega_1^-}[N(n, n+1]]^2] + \frac{1}{2}\mathbb{E}[\mathbb{E}^{\omega_2^-}[N(n, n+1]]^2]$$
$$\quad - \mathbb{E}[\mathbb{E}^{\omega_1^-}[N(n, n+1]]\mathbb{E}^{\omega_2^-}[N(n, n+1]]]$$
$$= \mathbb{E}[\mathbb{E}[N(n, n+1] \mid \mathcal{F}_0^{-\infty}]^2] - \mu^2$$
$$= \mathbb{E}[(\mathbb{E}[N(n, n+1] - \mu \mid \mathcal{F}_0^{-\infty}])^2].$$

Therefore, we have

$$\mathbb{E}[(\mathbb{E}[N(n, n+1] - \mu \mid \mathcal{F}_0^{-\infty}])^2]$$
$$= \frac{1}{2}\mathbb{E}[(\mathbb{E}^{\omega_1^-}[N(n, n+1]] - \mathbb{E}^{\omega_2^-}[N(n, n+1]])^2]$$
$$\leq \mathbb{E}[(\mathbb{E}^{\omega_1^-}[N(n, n+1]] - \mathbb{E}^{\varnothing}[N(n, n+1]])^2]$$
$$\quad + \mathbb{E}[(\mathbb{E}^{\omega_2^-}[N(n, n+1]] - \mathbb{E}^{\varnothing}[N(n, n+1]])^2]$$
$$= 2\mathbb{E}[(\mathbb{E}^{\omega_1^-}[N(n, n+1]] - \mathbb{E}^{\varnothing}[N(n, n+1]])^2],$$

where $\mathbb{E}^{\varnothing}[N(n, n+1]]$ denotes the expectation of the number of points in $(n, n+1]$ for the Hawkes process with the same dynamics (1) and empty history, i.e. $N(-\infty, 0] = 0$.

Next, let us estimate $\mathbb{E}^{\omega_1^-}[N(n, n+1]] - \mathbb{E}^{\varnothing}[N(n, n+1]]$. Here $\mathbb{E}^{\omega_1^-}[N(n, n+1]]$ is the expectation of the number of points in $(n, n+1]$ for the Hawkes process with intensity $\lambda_t = \lambda(\sum_{\{\tau : \tau \in \omega_1^- \cup \omega[0,t)\}} h(t - \tau))$. It is well defined for almost every ω_1^- under $\mathbb{P}$ because, under Assumption 1,

$$\mathbb{E}[\lambda_t] \leq \lambda(0) + \alpha \mathbb{E}\left[\int_{-\infty}^{t} h(t - s)N(\mathrm{d}s)\right] = \lambda(0) + \alpha\|h\|_{L^1}\mathbb{E}[N[0, 1]] < \infty,$$

which implies that $\lambda_t < \infty$, $\mathbb{P}$-a.s.

It is clear that $\mathbb{E}^{\omega_1^-}[N(n, n+1]] \geq \mathbb{E}^{\varnothing}[N(n, n+1]]$ a.s., so we can use a coupling method to estimate the difference. We will follow the ideas of Brémaud and Massoulié [4] and use the Poisson embedding method. Consider $(\Omega, \mathcal{F}, \mathcal{P})$, the canonical space of a point process on $\mathbb{R}^+ \times \mathbb{R}^+$ in which $\bar{N}$ is Poisson with intensity 1 under the probability measure $\mathcal{P}$. Then the Hawkes process N^0 with an empty past history and intensity λ_t^0 satisfies

$$\lambda_t^0 = \lambda\left(\int_{(0,t)} h(t-s)N^0(\mathrm{d}s)\right), \qquad t \in \mathbb{R}^+,$$

$$N^0(C) = \int_C \bar{N}(\mathrm{d}t \times [0, \lambda_t^0]), \qquad C \in \mathcal{B}(\mathbb{R}^+).$$

For $n \geq 1$, let us define recursively λ_t^n, D_n, and N^n as follows:

$$\lambda_t^n = \lambda\left(\int_{(0,t)} h(t-s)N^{n-1}(\mathrm{d}s) + \sum_{\tau \in \omega_1^-} h(t-\tau)\right), \qquad t \in \mathbb{R}^+, \tag{5}$$

$$D_n(C) = \int_C \bar{N}(\mathrm{d}t \times [\lambda_t^{n-1}, \lambda_t^n]), \qquad C \in \mathcal{B}(\mathbb{R}^+), \tag{6}$$

$$N^n(C) = N^{n-1}(C) + D_n(C), \qquad C \in \mathcal{B}(\mathbb{R}^+).$$

Following the arguments in Brémaud and Massoulié [4], we know that each λ_t^n is an $\mathcal{F}_t^{\bar{N}}$-intensity of N^n, where $\mathcal{F}_t^{\bar{N}}$ is the σ-algebra generated by $\bar{N}$ up to time t. By our Assumption 1, $\lambda(\cdot)$ is increasing, and it is clear that $\lambda^n(t)$ and $N^n(C)$ increase in n for all $t \in \mathbb{R}^+$ and $C \in \mathcal{B}(\mathbb{R}^+)$. Thus, D_n is well defined and, as $n \to \infty$, the limiting processes λ_t and N exist. N counts the number of points of $\bar{N}$ below the curve $t \mapsto \lambda_t$ and admits λ_t as an $\mathcal{F}_t^{\bar{N}}$-intensity. By the monotonicity properties of λ_t^n and N^n, we have

$$\lambda_t^n \leq \lambda\left(\int_{(0,t)} h(t-s)N(\mathrm{d}s) + \sum_{\tau \in \omega_1^-} h(t-\tau)\right),$$

$$\lambda_t \geq \lambda\left(\int_{(0,t)} h(t-s)N^n(\mathrm{d}s) + \sum_{\tau \in \omega_1^-} h(t-\tau)\right).$$

Letting $n \to \infty$ (which is valid since we assume that $\lambda(\cdot)$ is Lipschitz and, thus, continuous), we conclude that N and λ_t satisfy the dynamics (1). Therefore, with intensity λ_t, $N = N^0 + \sum_{i=1}^{\infty} D_i$ is the Hawkes process with past history ω_1^-.

We can then estimate the difference by noting that

$$\mathbb{E}^{\omega_1^-}[N(n, n+1]] - \mathbb{E}^{\varnothing}[N(n, n+1]] = \sum_{i=1}^{\infty} \mathbb{E}^{\mathcal{P}}[D_i(n, n+1]].$$

Here $\mathbb{E}^{\mathcal{P}}$ denotes the expectation with respect to $\mathcal{P}$, the probability measure on the canonical space that we defined earlier.

We have

$$\mathbb{E}^{\mathcal{P}}[D_1(n, n+1]] = \mathbb{E}^{\mathcal{P}}\left[\int_n^{n+1}(\lambda^1(t) - \lambda^0(t))\,\mathrm{d}t\right]$$

$$= \mathbb{E}^{\mathcal{P}}\left[\int_n^{n+1} \lambda\left(\sum_{\tau < t,\, \tau \in N^0 \cup \omega_1^-} h(t-\tau)\right) - \lambda\left(\sum_{\tau < t,\, \tau \in N^0 \cup \varnothing} h(t-\tau)\right)\mathrm{d}t\right]$$

$$\leq \alpha \int_n^{n+1} \sum_{\tau \in \omega_1^-} h(t-\tau)\,\mathrm{d}t, \tag{7}$$

where the first equality in (7) is due to the construction of D_1 in (6), the second equality in (7) is due to the definitions of λ^1 and λ^0 in (5), and, finally, the inequality in (7) is due to the fact that $\lambda(\cdot)$ is α-Lipschitz by Assumption 1. Similarly,

$$
\begin{aligned}
\mathbb{E}^{\mathcal{P}}[D_2(n, n+1]] &\leq \mathbb{E}^{\omega_1^-}\left[\alpha \int_n^{n+1} \sum_{\tau \in D_1, \tau < t} h(t-\tau)\,\mathrm{d}t\right] \\
&\leq \sum_{\tau \in \omega_1^-} \alpha^2 \int_n^{n+1} \int_0^t h(t-s)h(s-\tau)\,\mathrm{d}s\,\mathrm{d}t.
\end{aligned}
$$

Iteratively, we have, for any $k \in \mathbb{N}$,

$$
\begin{aligned}
\mathbb{E}^{\mathcal{P}}[D_k(n, n+1]] &\leq \sum_{\tau \in \omega_1^-} \alpha^k \int_n^{n+1} \int_0^{t_k} \cdots \int_0^{t_2} h(t_k - t_{k-1})h(t_{k-1} - t_{k-2}) \cdots \\
&\qquad\qquad\qquad\qquad\qquad\qquad \times h(t_2 - t_1)h(t_1 - \tau)\,\mathrm{d}t_1 \cdots \mathrm{d}t_k \\
&=: \sum_{\tau \in \omega_1^-} K_k(n, \tau).
\end{aligned}
$$

Now let $K(n, \tau) := \sum_{k=1}^{\infty} K_k(n, \tau)$. Then

$$
\begin{aligned}
\mathbb{E}[(\mathbb{E}^{\omega_1^-}&[N(n, n+1]] - \mathbb{E}^{\varnothing}[N(n, n+1]])^2] \\
&\leq \mathbb{E}\left[\left(\sum_{\tau \in \omega_1^-} K(n, \tau)\right)^2\right] \\
&\leq \mathbb{E}\left[\sum_{i, j \leq 0} K(n, i)K(n, j)N[i, i+1]N[j, j+1]\right] \\
&= \sum_{i, j \leq 0} K(n, i)K(n, j)\mathbb{E}[N[i, i+1]N[j, j+1]] \\
&\leq \sum_{i, j \leq 0} K(n, i)K(n, j)\frac{1}{2}\{\mathbb{E}[N[i, i+1]^2] + \mathbb{E}[N[j, j+1]^2]\} \\
&= \mathbb{E}[N[0, 1]^2]\left(\sum_{i \leq 0} K(n, i)\right)^2.
\end{aligned}
$$

Here, $\mathbb{E}[N[0, 1]^2] < \infty$ by Lemma 2. Therefore, we have

$$
\begin{aligned}
\sum_{n=1}^{\infty}&\{\mathbb{E}[(\mathbb{E}[N(n, n+1] - \mu \mid \mathcal{F}_0^{-\infty}])^2]\}^{1/2} \\
&\leq \sqrt{2\mathbb{E}[N[0, 1]^2]}\sum_{n=1}^{\infty}\sum_{i=-\infty}^{0} K(n, i) \\
&\leq \sqrt{2\mathbb{E}[N[0, 1]^2]}\sum_{k=1}^{\infty} \alpha^k \int_0^{\infty} \int_0^{t_k} \cdots \int_0^{t_2} \int_{-\infty}^0 h(t_k - t_{k-1})h(t_{k-1} - t_{k-2}) \cdots \\
&\qquad\qquad\qquad\qquad\qquad\qquad \times h(t_2 - t_1)h(t_1 - s)\,\mathrm{d}s\,\mathrm{d}t_1 \cdots \mathrm{d}t_k.
\end{aligned}
$$

Let $H(t) := \int_t^\infty h(s)\, \mathrm{d}s$. It is easy to check that $\int_0^\infty H(t)\, \mathrm{d}t = \int_0^\infty t h(t)\, \mathrm{d}t < \infty$ by Assumption 1. We have

$$\alpha^k \int_0^\infty \int_0^{t_k} \cdots \int_0^{t_2} \int_{-\infty}^0 h(t_k - t_{k-1}) h(t_{k-1} - t_{k-2}) \cdots h(t_2 - t_1) h(t_1 - s)\, \mathrm{d}s\, \mathrm{d}t_1 \cdots \mathrm{d}t_k$$

$$= \alpha^k \int_0^\infty \int_0^{t_k} \cdots \int_0^{t_2} h(t_k - t_{k-1}) h(t_{k-1} - t_{k-2}) \cdots h(t_2 - t_1) H(t_1)\, \mathrm{d}t_1 \cdots \mathrm{d}t_k$$

$$= \alpha^k \int_0^\infty \cdots \int_{t_{k-2}}^\infty \int_{t_{k-1}}^\infty h(t_k - t_{k-1})\, \mathrm{d}t_k h(t_{k-1} - t_{k-2})\, \mathrm{d}t_{k-1} \cdots H(t_1)\, \mathrm{d}t_1$$

$$= \alpha^k \|h\|_{L^1}^{k-1} \int_0^\infty H(t_1)\, \mathrm{d}t_1$$

$$= \alpha^k \|h\|_{L^1}^{k-1} \int_0^\infty t h(t)\, \mathrm{d}t.$$

Since $\alpha \|h\|_{L^1} < 1$, we conclude that

$$\sum_{n=1}^\infty \{\mathbb{E}[(\mathbb{E}[N(n, n+1] - \mu \mid \mathcal{F}_0^{-\infty}])^2]\}^{1/2} \leq \sum_{k=1}^\infty \sqrt{2\mathbb{E}[N[0,1]^2]} \alpha^k \|h\|_{L^1}^{k-1} \int_0^\infty t h(t)\, \mathrm{d}t$$

$$= \sqrt{2\mathbb{E}[N[0,1]^2]} \frac{\alpha}{1 - \alpha \|h\|_{L^1}} \int_0^\infty t h(t)\, \mathrm{d}t$$

$$< \infty.$$

Hence, by Theorem 3 we have

$$\frac{N_{[\cdot t]} - \mu[\cdot t]}{\sqrt{t}} \to \sigma B(\cdot) \quad \text{as } t \to \infty,$$

where

$$\sigma^2 = \mathbb{E}[(N[0,1] - \mu)^2] + 2 \sum_{j=1}^\infty \mathbb{E}[(N[0,1] - \mu)(N[j, j+1] - \mu)] < \infty. \tag{8}$$

By Lemma 3, $\sigma > 0$. Now, finally, for any $\varepsilon > 0$ and sufficiently large t,

$$\mathbb{P}\left(\sup_{0 \leq s \leq 1} \left| \frac{N_{[st]} - \mu[st]}{\sqrt{t}} - \frac{N_{st} - \mu st}{\sqrt{t}} \right| > \varepsilon \right)$$

$$= \mathbb{P}\left(\sup_{0 \leq s \leq 1} |(N_{[st]} - N_{st}) + \mu(st - [st])| > \varepsilon \sqrt{t} \right)$$

$$\leq \mathbb{P}\left(\sup_{0 \leq s \leq 1} |N_{[st]} - N_{st}| + \mu > \varepsilon \sqrt{t} \right)$$

$$\leq \mathbb{P}\left(\max_{0 \leq k \leq [t],\, k \in \mathbb{Z}} N[k, k+1] > \varepsilon \sqrt{t} - \mu \right)$$

$$\leq ([t] + 1)\mathbb{P}(N[0,1] > \varepsilon \sqrt{t} - \mu)$$

$$\leq \frac{[t] + 1}{(\varepsilon \sqrt{t} - \mu)^2} \int_{N[0,1] > \varepsilon \sqrt{t} - \mu} N[0,1]^2\, \mathrm{d}\mathbb{P}$$

$$\to 0$$

as $t \to \infty$ by Lemma 2. Hence, we conclude that $(N_{\cdot t} - \cdot \mu t)/\sqrt{t} \to \sigma B(\cdot)$ as $t \to \infty$.

The following lemma is used to prove Lemma 2 below.

Lemma 1. *There exists some $\theta > 0$ such that $\sup_{t \geq 0} \mathbb{E}^{\varnothing}[\exp\{\int_0^t \theta h(t-s)N(\mathrm{d}s)\}] < \infty$.*

Proof. Note first that, for any bounded deterministic function $f(\cdot)$,

$$\exp\left\{\int_0^t f(s)N(\mathrm{d}s) - \int_0^t (\mathrm{e}^{f(s)} - 1)\lambda(s)\,\mathrm{d}s\right\}$$

is a martingale. Therefore, using the Lipschitz assumption of $\lambda(\cdot)$, i.e. $\lambda(z) \leq \lambda(0) + \alpha z$, and applying Hölder's inequality, for $1/p + 1/q = 1$, we have

$$\mathbb{E}^{\varnothing}\left[\exp\left\{\int_0^t \theta h(t-s)N(\mathrm{d}s)\right\}\right]$$
$$= \mathbb{E}^{\varnothing}\left[\exp\left\{\int_0^t \theta h(t-s)N(\mathrm{d}s) - \frac{1}{p}\int_0^t (\mathrm{e}^{p\theta h(t-s)} - 1)\lambda(s)\,\mathrm{d}s\right.\right.$$
$$\left.\left. + \frac{1}{p}\int_0^t (\mathrm{e}^{p\theta h(t-s)} - 1)\lambda(s)\,\mathrm{d}s\right\}\right]$$
$$\leq \mathbb{E}^{\varnothing}\left[\exp\left\{\frac{q}{p}\int_0^t (\mathrm{e}^{p\theta h(t-s)} - 1)\lambda(s)\,\mathrm{d}s\right\}\right]^{1/q}$$
$$\leq \mathbb{E}^{\varnothing}\left[\exp\left\{\frac{q}{p}\int_0^t (\mathrm{e}^{p\theta h(t-s)} - 1)(\lambda(0) + \alpha\int_0^s h(s-u)N(\mathrm{d}u))\,\mathrm{d}s\right\}\right]^{1/q}$$
$$\leq \mathbb{E}^{\varnothing}\left[\exp\left\{\int_0^t \frac{q}{p}(\mathrm{e}^{p\theta h(t-s)} - 1)\alpha\int_0^s h(s-u)N(\mathrm{d}u)\,\mathrm{d}s\right\}\right]^{1/q}$$
$$\times \exp\left\{\frac{1}{p}\int_0^\infty (\mathrm{e}^{p\theta h(s)} - 1)\lambda(0)\,\mathrm{d}s\right\}.$$

Let $C(t) = \int_0^t q p^{-1}(\mathrm{e}^{p\theta h(t-s)} - 1)\alpha\,\mathrm{d}s$. Then, for any $t \in [0, T]$,

$$\mathbb{E}^{\varnothing}\left[\exp\left\{\int_0^t \frac{q}{p}(\mathrm{e}^{p\theta h(t-s)} - 1)\alpha\int_0^s h(s-u)N(\mathrm{d}u)\,\mathrm{d}s\right\}\right]$$
$$= \mathbb{E}^{\varnothing}\left[\exp\left\{\frac{1}{C(t)}\int_0^t \frac{q}{p}(\mathrm{e}^{p\theta h(t-s)} - 1)\alpha C(t)\int_0^s h(s-u)N(\mathrm{d}u)\,\mathrm{d}s\right\}\right]$$
$$\leq \mathbb{E}^{\varnothing}\left[\frac{1}{C(t)}\int_0^t \frac{q}{p}(\mathrm{e}^{p\theta h(t-s)} - 1)\alpha \exp\left\{C(t)\int_0^s h(s-u)N(\mathrm{d}u)\right\}\mathrm{d}s\right]$$
$$\leq \sup_{0 \leq s \leq T} \mathbb{E}^{\varnothing}\left[\exp\left\{C(\infty)\int_0^s h(s-u)N(\mathrm{d}u)\right\}\right], \tag{9}$$

where in the first inequality in (9) we used Jensen's inequality since $x \mapsto \mathrm{e}^x$ is convex and $(C(t))^{-1}\int_0^t q p^{-1}(\mathrm{e}^{p\theta h(t-s)} - 1)\alpha\,\mathrm{d}s = 1$, and in the second inequality in (9) we used the facts that $C(t) \leq C(\infty)$ and, again, $(C(t))^{-1}\int_0^t q p^{-1}(\mathrm{e}^{p\theta h(t-s)} - 1)\alpha\,\mathrm{d}s = 1$. Now choose $q > 1$ so small that $q\alpha\|h\|_{L^1} < 1$. Once p and q are fixed, choose $\theta > 0$ so small that

$$C(\infty) = \int_0^\infty \frac{q}{p}(\mathrm{e}^{p\theta h(s)} - 1)\alpha\,\mathrm{d}s < \theta.$$

This implies that, for any $t \in [0, T]$,

$$
\mathbb{E}^{\varnothing}\left[\exp\left\{\int_0^t \theta h(t - s) N(\mathrm{d}s)\right\}\right]
$$

$$
\leq \sup_{0 \leq s \leq T} \mathbb{E}^{\varnothing}\left[\exp\left\{\theta \int_0^s h(s - u) N(\mathrm{d}u)\right\}\right]^{1/q} \exp\left\{\frac{1}{p}\int_0^\infty (\mathrm{e}^{p\theta h(s)} - 1)\lambda(0)\,\mathrm{d}s\right\}.
$$

Hence, we conclude that, for any $T > 0$,

$$
\sup_{0 \leq t \leq T} \mathbb{E}^{\varnothing}\left[\exp\left\{\theta \int_0^t h(t - s) N(\mathrm{d}s)\right\}\right] \leq \exp\left\{\int_0^\infty (\mathrm{e}^{p\theta h(s)} - 1)\lambda(0)\,\mathrm{d}s\right\} < \infty.
$$

Lemma 2. *There exists some $\theta > 0$ such that $\mathbb{E}[\mathrm{e}^{\theta N[0,1]}] < \infty$. Hence, $\mathbb{E}[N[0, 1]^2] < \infty$.*

Proof. By Assumption 1, $h(\cdot)$ is positive and decreasing. Thus, $\delta = \inf_{t \in [0,1]} h(t) > 0$. Hence,

$$
\mathbb{E}^{\varnothing}[\mathrm{e}^{\theta N[t-1,t]}] \leq \mathbb{E}^{\varnothing}\left[\exp\left\{\frac{\theta}{\delta}\int_0^t h(t - s) N(\mathrm{d}s)\right\}\right].
$$

By Lemma 1, we can choose $\theta > 0$ so small that

$$
\limsup_{t \to \infty} \mathbb{E}^{\varnothing}[\mathrm{e}^{\theta N[t-1,t]}] < \infty.
$$

Finally, $\mathbb{E}[\mathrm{e}^{\theta N[0,1]}] \leq \liminf_{t \to \infty} \mathbb{E}^{\varnothing}[\mathrm{e}^{\theta N[t-1,t]}] < \infty$. $\qquad \square$

It is intuitively clear that $\sigma > 0$. However, we still require a proof.

Lemma 3. *It holds that $\sigma > 0$, where σ is defined in (8).*

Proof. Let $\eta_n = \sum_{j=n}^\infty \mathbb{E}[N(j, j + 1] - \mu \mid \mathcal{F}_{n+1}^{-\infty}]$, where $\mu = \mathbb{E}[N[0, 1]]$. Here η_n is well defined because we proved (4). To see this, note that

$$
\|\eta_n\|_2 = \left\|\sum_{j=n}^\infty \mathbb{E}[N(j, j + 1] - \mu \mid \mathcal{F}_{n+1}^{-\infty}]\right\|_2 \leq \sum_{j=n}^\infty \|\mathbb{E}[N(j, j + 1] - \mu \mid \mathcal{F}_{n+1}^{-\infty}]\|_2 < \infty,
$$

by (4). Also, it is easy to check that

$$
\mathbb{E}[\eta_{n+1} - \eta_n + N(n, n + 1] - \mu \mid \mathcal{F}_{n+1}^{-\infty}]
$$

$$
= \mathbb{E}\left[\sum_{j=n+1}^\infty \mathbb{E}[N(j, j + 1] - \mu \mid \mathcal{F}_{n+2}^{-\infty}] \,\middle|\, \mathcal{F}_{n+1}^{-\infty}\right]
$$

$$
- \mathbb{E}\left[\sum_{j=n}^\infty \mathbb{E}[N(j, j + 1] - \mu \mid \mathcal{F}_{n+1}^{-\infty}] \,\middle|\, \mathcal{F}_{n+1}^{-\infty}\right] + N(n, n + 1] - \mu
$$

$$
= \sum_{j=n+1}^\infty \mathbb{E}[N(j, j + 1] - \mu \mid \mathcal{F}_{n+1}^{-\infty}] - \sum_{j=n+1}^\infty \mathbb{E}[N(j, j + 1] - \mu \mid \mathcal{F}_{n+1}^{-\infty}]
$$

$$
- N(n, n + 1] + \mu + N(n, n + 1] - \mu
$$

$$
= 0.
$$

Let $Y_n = \eta_{n-1} - \eta_{n-2} + N(n-2, n-1] - \mu$. Then Y_n is an ergodic, stationary sequence such that $\mathbb{E}[Y_n \mid \mathcal{F}_{n-1}^{-\infty}] = 0$. By (4), $\mathbb{E}[Y_n^2] < \infty$ and, by Theorem 4, $S'_{[n \cdot]}/\sqrt{n} \to \sigma' B(\cdot)$, where $S'_n = \sum_{j=1}^n Y_j$. It is clear that $\sigma = \sigma' < \infty$ since, for any $\varepsilon > 0$,

$$
\mathbb{P}\left(\max_{1 \le k \le [n],\, k \in \mathbb{Z}} \frac{1}{\sqrt{n}} \sum_{j=1}^k (\eta_{j-1} - \eta_{j-2}) > \varepsilon\right)
$$

$$
= \mathbb{P}\left(\max_{1 \le k \le [n],\, k \in \mathbb{Z}} (\eta_{k-1} - \eta_{-1}) > \varepsilon \sqrt{n}\right)
$$

$$
\le \mathbb{P}\left(\left\{\max_{1 \le k \le [n],\, k \in \mathbb{Z}} |\eta_{k-1}| > \frac{\varepsilon \sqrt{n}}{2}\right\} \cup \left\{|\eta_{-1}| > \frac{\varepsilon \sqrt{n}}{2}\right\}\right)
$$

$$
\le \sum_{k=1}^{[n]} \mathbb{P}\left(|\eta_{k-1}| > \frac{\varepsilon \sqrt{n}}{2}\right) + \mathbb{P}\left(|\eta_{-1}| > \frac{\varepsilon \sqrt{n}}{2}\right)
$$

$$
= ([n] + 1)\mathbb{P}\left(|\eta_{-1}| > \frac{\varepsilon \sqrt{n}}{2}\right)
$$

$$
\le \frac{4([n] + 1)}{\varepsilon^2 n} \int_{|\eta_{-1}| > \varepsilon \sqrt{n}/2} |\eta_{-1}|^2 \, d\mathbb{P}
$$

$$
\to 0 \quad \text{as } n \to \infty,
$$

where we used the stationarity of $\mathbb{P}$, Chebychev's inequality, and (4).

Now, it becomes clear that

$$
\sigma^2 = (\sigma')^2
$$

$$
= \mathbb{E}[Y_1^2]
$$

$$
= \mathbb{E}(\eta_0 - \eta_{-1} + N(-1, 0] - \mu)^2
$$

$$
= \mathbb{E}\left(\sum_{j=0}^{\infty} \mathbb{E}[N(j, j+1] - \mu \mid \mathcal{F}_1^{-\infty}] - \sum_{j=0}^{\infty} \mathbb{E}[N(j, j+1] - \mu \mid \mathcal{F}_0^{-\infty}]\right)^2.
$$

Consider $D = \{\omega \colon \omega^- \ne \varnothing,\ \omega(0, 1] = \varnothing\}$. Note that $\mathbb{P}(\omega^- = \varnothing) = 0$. By Jensen's inequality and Assumption 1, we have

$$
\mathbb{P}(D) = \int \mathbb{P}^{\omega^-}(N(0, 1] = 0)\mathbb{P}(d\omega^-)
$$

$$
= \mathbb{E}\left[\exp\left\{-\int_0^1 \lambda\left(\sum_{\tau \in \omega^-} h(t - \tau)\right) dt\right\}\right]
$$

$$
\ge \exp\left\{-\mathbb{E}\int_0^1 \lambda\left(\sum_{\tau \in \omega^-} h(t - \tau)\right) dt\right\}
$$

$$
\ge \exp\left\{-\lambda(0) - \alpha \mathbb{E}\int_0^1 \sum_{\tau \in \omega^-} h(t - \tau) \, dt\right\}
$$

$$
\ge \exp\{-\lambda(0) - \alpha \mathbb{E}[N[0, 1]] \cdot \|h\|_{L^1}\}
$$

$$
> 0.
$$

It is clear that, given the event D,

$$\sum_{j=0}^{\infty} \mathbb{E}[N(j, j+1) - \mu \mid \mathcal{F}_1^{-\infty}] < \sum_{j=0}^{\infty} \mathbb{E}[N(j, j+1) - \mu \mid \mathcal{F}_0^{-\infty}].$$

Therefore,

$$\mathbb{P}\left(\sum_{j=0}^{\infty} \mathbb{E}[N(j, j+1) - \mu \mid \mathcal{F}_1^{-\infty}] \neq \sum_{j=0}^{\infty} \mathbb{E}[N(j, j+1) - \mu \mid \mathcal{F}_0^{-\infty}]\right) > 0,$$

which implies that $\sigma > 0$.

Proof of Theorem 2. By [12], Strassen's invariance principle holds if we have (4) and $\sigma > 0$.

Acknowledgements

References

[1] BACRY, E., DELATTRE, S., HOFFMANN, M. AND MUZY, J. F. (2012). Scaling limits for Hawkes processes and application to financial statistics. Available at http://arxiv.org/abs/1202.0842v1.

[2] BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd edn. John Wiley, New York.

[3] BORDENAVE, C. AND TORRISI, G. L. (2007). Large deviations of Poisson cluster processes. *Stoch. Models* **23**, 593–625.

[4] BRÉMAUD, P. AND MASSOULIÉ, L. (1996). Stability of nonlinear Hawkes processes. *Ann. Prob.* **24**, 1563–1588.

[5] BRÉMAUD, P., NAPPO, G. AND TORRISI, G. L. (2002). Rate of convergence to equilibrium of marked Hawkes processes. *J. Appl. Prob.* **39**, 123–136.

[6] DALEY, D. J. AND VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes*, Vol. I , 2nd edn. Springer, New York.

[7] DALEY, D. J. AND VERE-JONES, D. (2008). *An Introduction to the Theory of Point Processes*, Vol. II , 2nd edn. Springer, New York.

[8] DASSIOS, A. AND ZHAO, H. (2011). A dynamic contagion process. *Adv. Appl. Prob.* **43**, 814–846.

[9] ERRAIS, E., GIESECKE, K. AND GOLDBERG, L. R. (2010). Affine point processes and portfolio credit risk. *SIAM J. Financial Math.* **1**, 642–665.

[10] HAWKES, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58**, 83–90.

[11] HAWKES, A. G. AND OAKES, D. (1974). A cluster process representation of a self-exciting process. *J. Appl. Prob.* **11**, 493–503.

[12] HEYDE, C. C. AND SCOTT, D. J. (1973). Invariance principles for the law of the iterated logarithm for martingales and processes with stationary increments. *Ann. Prob.* **1**, 428–436.

[13] KARABASH, D. AND ZHU, L. (2013). Limit theorems for marked Hawkes processes with application to a risk model. Available at http://arxiv.org/abs/1211.4039v2.

[14] LINIGER, T. (2009). *Multivariate Hawkes Processes*. Doctoral Thesis, ETH.

[15] ZHU, L. (2012). Large deviations for Markovian nonlinear Hawkes Processes. Available at http://arxiv.org/abs/1108.2432v2.

[16] ZHU, L. (2013). Process-level large deviations for nonlinear Hawkes point processes. To appear in *Ann. Inst. H. Poincaré Prob. Statist.*