On the splitting and aggregating of Hawkes processes

BO LI AND GUODONG PANG

Abstract. We consider the random splitting and aggregating of Hawkes processes. We present the random splitting schemes using the direct approach for counting processes, as well as the the immigration-birth branching representations of Hawkes processes. From the second scheme, it is shown that random split Hawkes processes are again Hawkes. We discuss functional central limit theorems (FCLTs) for the scaled split processes from the different schemes. On the other hand, aggregating multivariate Hawkes processes may be not necessarily Hawkes. We identify a necessary and sufficient condition for the aggregated process to be Hawkes. We prove a FCLT for a multivariate Hawkes process under a random splitting and then aggregating scheme (under certain conditions, transforming into a Hawkes process of a different dimension).

1. Introduction

Hawkes processes were firstly introduced in [9, 10] as an extension of Poisson process, which has the so-called self-exciting effect, that is, the occurrence of an event will increase the probability of future events. They have been widely used to model various applications, for example, finance [2, 11], and internet traffic and queueing [14, 6, 8, 4]. There are extensive studies of Hawkes processes, including the exact distributional properties [12, 16] for exponential type of kernel functions, and limit theorems in both conventional scaling [1, 2, 13, 4] and large intensity scaling [8, 7, 15].

In this paper, we investigate the random splitting/sampling and aggregating/superposition of Hawkes processes. Random splitting/sampling of point processes has been an important topic in stochastic models. For example, one arrival stream of customers may require different types of services, and the same packet in communication networks may be simultaneously sent out on several outgoing links. Similarly, input into a service or communication system can come from aggregating several sources [17, 18]. See further discussions in [21, Chapter 9]. Splitting and aggregating of standard point processes are well understood in the literature [19, 20, 21].

For Hawkes processes, the Poisson branching representation [12], also known as the immigration-birth representation, has been a fundamental tool in their analysis. It says that for each individual (generation), the number of individuals/children produced by this individual over time, called the next generation, is a simple and conditional Poisson process with an intensity that is a functional of the counting process of his/her generation. It is known that random splitting/sampling of Poisson processes results in independent Poisson processes, each component with a rate equal to that of the original process multiplying the splitting probability. On the contrary, each sub-counting process resulting from the random splitting/sampling of a Hawkes process cannot be regarded as a Hawkes process itself, despite the conditional Poisson property for each generation. Intuitively, the jumping intensity of a sub-counting process depends on the history of the original Hawkes process, which requires information strictly larger than that provided by the sub-counting process. However, we show that the vector-valued splitting Hawkes process is a multidimensional Hawkes process (Propositions 2 and 8). It is also clear that the split processes are no longer independent, but the dependence structure of the split processes is not at all obvious.

We thus aim to understand the split processes of a Hawkes process and their dependence structure. We provide two representations of the (scaled) split processes, one directly using the original

Key words and phrases. Hawkes process, random splitting/sampling, aggregating/superposition, functional central limit theorem .

Hawkes process, and the other using the non-homogeneous conditional Poisson processes in each generation from the Poisson branching representation. We start from splitting a one-dimensional Hawkes process for the ease of exposition, and discuss how the splitting schemes work and show the equivalence of the limits in the FCLTs derived from them, which rely on the existing results in Chapter 9.5 of Whitt [21] and [1]. See the discussions in Section 3. We next aim to understand the aggregation/superposition of multivariate Hawkes process and show when the aggregated process can still be a Hawkes process (Section 4).

We then consider the scheme of first splitting and then aggregating of a multivariate Hawkes process, which may transform it into a point process of a different dimension (Section 5). We identify conditions under which the transformed process is again Hawkes. We prove a FCLT for the transformed process which is a Brownian motion limit with a surprisingly simple covariance function. We also discuss how that relates to the known result in the special case of being a Hawkes process.

1.1. **Notation.** All random variables and processes are defined in a common complete probability space $\big(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P}\big)$. Throughout the paper, $\mathbb{N}$ denotes the set of natural numbers. $\mathbb{R}(\mathbb{R}_+)$ denotes the space of real (nonnegative) number. Let $\mathbb{D} = \mathbb{D}(\mathbb{R}_+, \mathbb{R})$ denote $\mathbb{R}$-valued function space of all càdlàg functions on $\mathbb{R}_+$. $(\mathbb{D}, J_1)$ denotes space $\mathbb{D}$ equipped with Skorohod J_1 topology, see [3], which is complete and separable. $\mathbb{D}^n$ denotes an n-dimensional vector-valued càdlàg process endowed with the weak Skohorod J_1 topology [21], for which we write $(\mathbb{D}^n, J_1)$. $L^2(\mathbb{P})$ denotes the space of random variables with finite 2^{nd} moment. For integrable function $f : \mathbb{R} \to \mathbb{R}$, its L^1 norm is denoted by $\|f\|_1$. Notations $\to$ and $\Rightarrow$ mean convergence of real numbers and convergence in distribution, respectively. For a matrix $M = (M_{ij})_{i,j}$, we denote by $\text{ent}_{ij}M = M_{ij}$ its $(i,j)^{\text{th}}$ entry. $\text{row}_k M$ and $\text{col}_m M$ denotes its k^{th} row vector and the m^{th} column vector, respectively. M^{T} denotes the transpose of M. I denotes the identity matrix. e denotes the column vector of 1 with associated dimension. For a vector a, $\text{diag}(a)$ denotes the diagonal matrix with the elements of vector a on the main diagonal. $f * g(t) := \int_0^t f(t-s)g(s)ds$ denotes the convolution of f and g on $\mathbb{R}_+$. Additional notation is introduced in the paper whenever necessary.

2. Preliminaries on Hawkes processes

A d-dimensional Hawkes process, $N = \{N(t), t \geq 0\}$ with $N = \big(N_k\big)_{k=1,\dots,d}$, is formally defined as a $\mathbb{N}^d$-valued simple counting process with conditional intensity

$$\lambda_k(t) = \lambda_{k0} + \sum_{k'=1}^{d} \sum_{j=1}^{N_{k'}(t)} H_{kk'}(t - \tau_{kj}) = \lambda_{k0} + \sum_{k'=1}^{d} \int_0^t H_{kk'}(t-s) N_{k'}(ds), \quad t \geq 0, \tag{2.1}$$

for every $k = 1, \cdots, d$, where τ_{kj} is the j^{th} event time of N_k, and $\lambda_{k0} \geq 0$ is a constant called the baseline intensity of the k^{th} subprocess, and $H_{kk'} : \mathbb{R}_+ \to \mathbb{R}_+$ is called the mutually exciting function or the kernel function, also known as the cross-exciting function for $k \neq k'$ and the self-exciting function for $k = k'$.

Assumption A1. *For all $k, k' = 1, \cdots, d$, we have*

$$\int_0^\infty H_{kk'}(t)dt < \infty \tag{2.2}$$

and the spectral radius $\rho(\|H\|_1)$ of the matrix

$$\|H\|_1 := \Big(\int_0^\infty H_{kk'}(t)dt\Big)_{k,k'}$$

satisfies $\rho(\|H\|_1) < 1$.

The condition in (2.2) is also called the non-explosion criterion in [1, 5]. It is easy to calculate the mean of $N(t)$:

$$\mathbb{E}[N(t)] = \mathbb{E}\Big[\int_0^t \lambda(s)ds\Big] = \Big(\int_0^t \big(I + \varphi * I(s)\big)ds\Big) \cdot \lambda_0, \quad t \geq 0\,. \tag{2.3}$$

Here $\lambda_0 = \big(\lambda_{k0}\big)_k$ is the constant vector baseline intensity in (2.1), $\varphi * I(s)$ is a matrix with $\varphi_{kk'} * 1(s) = \int_0^s \varphi_{kk'}(s-u)\,du$ at its $(k,k')^{\text{th}}$ entry (we abuse notation of $1(\cdot)$ as a constant function equal to one), and φ is a $d \times d$ matrix defined as a $L^1(dt)$ limit of the following series

$$\varphi(t) = H(t) + H * H(t) + H * H * H(t) + \cdots, \tag{2.4}$$

where $F * G$ is defined as the matrix as each entry

$$\text{ent}_{ij}(F * G(t)) = \sum_k F_{ik} * G_{kj}(t) = \sum_k \int_0^t F_{ik}(t-s) G_{jk}(s)\,ds$$

for matrix valued functions F, G, see e.g., Theorem 2 in [1]. φ can also be understood as the renewal density of function H and satisfies (matrix) renewal equation

$$\varphi(t) = H(t) + \int_0^t H(t-s)\varphi(s)ds.$$

In addition, under Assumption A1,

$$I + \|\varphi\|_1 = (I - \|H\|_1)^{-1}. \tag{2.5}$$

The FLLN for the Hawkes process N reads [1, Theorem 1]: under Assumption A1,

$$\sup_{t\in[0,1]} \Big\| \bar{N}_T(t) - (I - \|H\|_1)^{-1} \cdot \lambda_0 t \Big\| \to 0 \quad \text{as} \quad T \to \infty, \tag{2.6}$$

almost surely and in $L^2(\mathbb{P})$, where $\bar{N}_T(t) := T^{-1}N(Tt)$. The FCLT for the Hawkes process N is stated as follows [1, Theorem 2]: under Assumption A1,

$$\hat{N}_T(t) := \sqrt{T}\big(\bar{N}_T(t) - \mathbb{E}[\bar{N}_T(t)]\big) \Rightarrow \hat{N}(t) \quad \text{in} \quad (\mathbb{D}^d,\, J_1), \tag{2.7}$$

as $T \to \infty$, where

$$\hat{N}(t) := (I - \|H\|_1)^{-1} \cdot \Sigma^{1/2} \cdot W, \tag{2.8}$$

W is a d-dimensional standard Brownian motion and

$$\Sigma = \text{diag}\big((I - \|H\|_1)^{-1} \cdot \lambda_0\big). \tag{2.9}$$

3. Splitting of one-dimensional Hawkes process

We now describe the random splitting mechanism of Hawkes processes. We first focus on splitting of one-dimensional Hawkes process for the sake of exposition. Splitting of d-dimensional Hawkes process can be derived similarly and will be discussed in Section 5 together with aggregation. We provide two representations of the split processes: the first using the method in Chapter 9.5 of Whitt [21], and the second using the immigration-birth branching representation. Note that for the case $d = 1$, Assumption A1 reduces to $\|H\|_1 \in (0, 1)$.

3.1. **The first representation.** Let N be the Hawkes process in (2.1) with $d=1$, and self-exciting function $H:\mathbb{R}_+\to\mathbb{R}_+$. Denoting by $\{\xi_j, j\geq 1\}$ the splitting variables, whenever $\xi_j=m$ the j^{th} individual occurring at τ_j is assigned to the m^{th} split process. Under this standard splitting, the process N splits into n sub-counting processes, denoted as $N^{(m)}$,

$$N^{(m)}(t)=\sum_{j=1}^{N(t)}\mathbf{1}(\xi_j=m)\quad\text{for every } m=1,2,\cdots,n, \text{ and } t\geq 0. \tag{3.1}$$

We assume that $\{\xi_j, j\geq 1\}$ is a sequence of i.i.d. variables, independent of N, with

$$\mathbb{P}\big(\xi_j=m\big)=p^{(m)}\quad\text{and}\quad\sum_{m=1}^{n}p^{(m)}=1. \tag{3.2}$$

By the independence between N and $\{\xi_j\}_j$, it is easy to see that

$$\mathbb{E}[N^{(m)}(t)]=p^{(m)}\mathbb{E}[N(t)]=\lambda_0\,p^{(m)}\int_0^t\big(1+\varphi*1(s)\big)ds. \tag{3.3}$$

(Here φ is a $\mathbb{R}_+$-valued function.)

We consider the scaled process indexed by T, and all the variables are marked with an additional subscript T, that is, N_T is a Hawkes process with intensity process $\lambda_T(\cdot)$ whose baseline intensity in (2.1) is $\lambda_{0,T}$ and the kernel function H stays the same. The splitting variables are denoted by $\{\xi_{j,T}, j\geq 1\}$ with distribution $\{p_T^{(m)}\}_m$.

Assumption A2. *Assume that, for some $\lambda_0, p^{(m)}>0$ with $\sum_{m=1}^n p^{(m)}=1$,*

$$\lambda_{0,T}\to\lambda_0\quad\textit{and}\quad p_T^{(m)}\to p^{(m)}\quad\textit{as}\quad T\to\infty.$$

Define for every $m=1,2,\cdots,d$,

$$\bar{N}_T^{(m)}(t):=\frac{1}{T}N_T^{(m)}(Tt)=\frac{1}{T}\sum_{j=1}^{N_T(Tt)}\mathbf{1}(\xi_{j,T}=m),$$

and the diffusion-scaled processes

$$\hat{N}_T^{(m)}(t):=\sqrt{T}\Big(\bar{N}_T^{(m)}(t)-\mathbb{E}\big[\bar{N}_T^{(m)}(t)\big]\Big)\quad\text{and}\quad\hat{S}_T^{(m)}(t):=\frac{1}{\sqrt{T}}\sum_{j=1}^{\lfloor Tt\rfloor}\Big(\mathbf{1}(\xi_{j,T}=m)-p_T^{(m)}\Big) \tag{3.4}$$

where $\lfloor t\rfloor$ represents the largest integer no larger than $t\in\mathbb{R}_+$. Then we have the first representation

$$\hat{N}_T^{(m)}(t)=\hat{S}_T^{(m)}(\bar{N}_T(t))+p_T^{(m)}\hat{N}_T(t). \tag{3.5}$$

Note that in this representation (3.5), the process $\hat{N}_T^{(m)}$ consists of two components, where $\hat{S}_T^{(m)}$ represents the oscillation from splitting scheme, and $p_T^{(m)}\hat{N}_T$ is the oscillation inherited from the original counting process and proportional to the splitting probability. Note that the processes $\hat{S}_T^{(m)}$ and $\hat{N}_T$ are independent, and in the limit, we see that the two components $\hat{S}_T^{(m)}\circ\bar{N}_T$ and $p_T^{(m)}\hat{N}_T$ converge to two independent processes as $T\to\infty$.

By applying [21, Theorem 9.5.1], under Assumptions A1 and A2, provided with the limit for $\bar{N}_T$ in (2.6) and the limit for $\hat{N}_T$ in (2.8), we obtain

Proposition 1. *Let $(\hat{N}_T^{(m)})_m$ be the diffusion-scaled process in (3.4). Assume that $\|H\|_1\in(0,1)$ and Assumption A2 hold. We have*

$$\big(\hat{N}_T^{(m)}\big)_m\Rightarrow\big(\hat{N}^{(m)}\big)_m\quad\textit{in}\quad(\mathbb{D}^n,J_1)\quad\textit{as}\quad T\to\infty,$$

where

$$\hat{N}^{(m)} = \frac{\lambda_0^{1/2}}{(1-\|H\|_1)^{1/2}}\hat{S}^{(m)} + p^{(m)}\frac{\lambda_0^{1/2}}{(1-\|H\|_1)^{3/2}}W \tag{3.6}$$

where $(\hat{S}^{(m)})_m$ *is an* n*-dimensional Brownian motion with covariance function*

$$\mathrm{Cov}\big(\hat{S}^{(m)}(t), \hat{S}^{(m')}(s)\big) = \big(-p^{(m)}p^{(m')} + p^{(m)}\delta_{mm'}\big)(t\wedge s),$$

and W *is a standard Brownian motion, independent of* $(\hat{S}^{(m)})_m$.

Therefore, the limit $\big(\hat{N}^{(m)}\big)_m$ *is an* n*-dimensional Brownian motion with covariance*

$$\mathrm{Cov}\big(\hat{N}^{(m)}(t), \hat{N}^{(m')}(s)\big) = (t\wedge s)\left(\frac{\lambda_0 p^{(m)}(\delta_{mm'} - p^{(m')})}{1-\|H\|_1} + \frac{\lambda_0 p^{(m)}p^{(m')}}{(1-\|H\|_1)^3}\right). \tag{3.7}$$

3.2. **The second representation.** Recall the immigration-birth branching representation with chronological levels for the Hawkes process N described in (2.1), which generalizes the one proposed by [12]. Basically, the points from N are categorized virtually into those from an exogenous arrival process which are called migrants and/or the first generation, and those generated from existing points called children, and where their chronological levels are obviously defined. Thus we rewrite

$$N(t) = \sum_{l\geq 1} N_l(t) \tag{3.8}$$

where $N_1(t) = \sup\{j\geq 1, \tau_{1j}\leq t\}$ is a Poisson process with parameter λ_0 representing the arrival rate of the immigrants, and τ_{1j} denotes the arrival time of the j^{th} immigrant. And for $l\geq 1$, $N_{l+1}(t) = \sup\{j\geq 1, \tau_{(l+1)j}\leq t\}$ representing the individuals of $(l+1)^{\text{th}}$ generation is a inhomogeneous Poisson process with intensity, by conditioning on $\mathscr{F}_l(t)$,

$$\lambda_l(t) = \sum_{j=1}^{N_l(t)} H(t-\tau_{lj}) = \int_0^t H(t-s)N_l(ds) \in \mathscr{G}_l(t) \tag{3.9}$$

where $\mathscr{G}_l(t) = \sigma\{N_l(s), s\leq t\}$ and $\mathscr{F}_l(t) = \bigvee_{1\leq l'\leq l}\mathscr{G}_{l'}(t)$, and where τ_{lj} denotes the birth time of the j^{th} children of the l^{th} generation which is produced by some individual in the $(l-1)^{\text{th}}$ generation. $\mathscr{G}_l(t)$ represents the information produced by the l^{th} generation, $\mathscr{F}_l(t)$ represents the information up to the l^{th} generation, and $\mathscr{F}_\infty(t) = \bigvee_{l\geq 1}\mathscr{F}_l(t)$ collects the information of all the generations up to time t, which includes not only the occurrence times of events but also the virtually defined generation information and hence is strictly larger than that generated by N itself. Under the non-explosion assumption (2.2), λ_l and N_l are finite valued and can be constructed pathwise. By conditioning and the additive property of the intensity for the independent counting processes, N in (3.8) is a simple counting process with conditional intensity in (2.1).

To describe the split processes with the branching generations, we further let $\{\xi_{lj}, j\geq 1\}$ be the i.i.d. splitting variables for the individuals from N_l, which have the same distributional properties as $\{\xi_j, j\geq 1\}$ in (3.1). Recall $p^{(m)}$ in (3.2). Let

$$N_l^{(m)}(t) := \sum_{j=1}^{N_l(t)} \mathbf{1}(\xi_{lj} = m), \quad \text{for every } m = 1, 2, \cdots, n. \tag{3.10}$$

Then $N^{(m)} = \sum_{l\geq 1} N_l^{(m)}$.

Proposition 2. *$\big(N^{(m)}\big)_m$ is an n-dimensional Hawkes process, where the baseline intensity vector is $\big(p^{(m)}\lambda_0\big)_m$ and mutual exciting matrix at (m,m')-entry is $p^{(m)}H$, that is, its intensity is given by*

$$\lambda^{(m)}(t) = p^{(m)}\lambda(t) = p^{(m)}\lambda_0 + \sum_{m'=1}^{n}\int_0^t \big(p^{(m)}H(t-s)\big)N^{(m')}(ds), \tag{3.11}$$

for every $m = 1, 2, \cdots, n$, where λ is the intensity for N in (2.1).

Remark 1. *We remark that although the split process $\big(N^{(m)}\big)_m$ is a multivariate Hawkes process, the cross-exciting function $H_{mm'}$ in the definition (2.1) takes a special form, $p^{(m)}H$, independent of m' in (3.11).*

Proof. We show that $\big(N^{(m)}\big)_m$ is a counting process with conditional jumping intensity $\big(\lambda^{(m)}\big)_m$ by evaluating its conditional jumping intensity.

For every l noice that given $\{\mathscr{F}_l(t)\}_{t\ge 0}$, $\big(N^{(m)}_{l+1}\big)_m$ is a Poisson process with intensity $\lambda^{(m)}_l = p^{(m)}\lambda_l$ and independent among m. Therefore, conditioning on $\mathscr{F}_\infty(t)$, the jumping intensity of $N^{(m)}$ at t is

$$\begin{aligned}\lambda^{(m)}(t) &:= \sum_{l\ge 1}\lambda^{(m)}_{l-1}(t) = p^{(m)}\lambda(t) = p^{(m)}\Big(\lambda_0 + \int_0^t H(t-s)N(ds)\Big)\\ &= p^{(m)}\lambda_0 + \sum_{m'=1}^{d}\int_0^t \big(p^{(m)}H(t-s)\big)N^{(m')}(ds)\end{aligned}$$

which is a process adapted to the natural filtration of $(N^{(m)})_m$. This proves Proposition 2. $\square$

Recall $\hat N^{(m)}_T$ defined in (3.4). By applying [1, Theorem 2] to Proposition 2, we obtain the following result.

Proposition 3. *Suppose that $\|H\|_1 \in (0,1)$ and Assumption A2 hold. We have*

$$\big(\hat N^{(m)}_T\big)_m \Rightarrow \big(\hat N^{(m)}\big)_m \quad in \quad (\mathbb{D}^n, J_1) \quad as \quad T\to\infty,$$

where $\big(\hat N^{(m)}\big)_m$ is a standard n-dimensional Brownian motion with covariance matrix

$$\big(I-\tilde\Xi\big)^{-1}\cdot\tilde\Sigma\cdot\big(I-\tilde\Xi^{\mathrm{T}}\big)^{-1} \tag{3.12}$$

where $\tilde\Xi = \big(p^{(m)}\|H\|_1\big)_{mm'}$ and $\tilde\Sigma = \mathrm{diag}(p)\dfrac{\lambda_0}{1-\|H\|_1}$.

Proposition 4. *The limits in Propositions 1 and 3 are equivalent in distribution.*

Proof. To check the equivalence it suffices to show that the covariance functions in (3.7) coincide with the matrix in (3.12).

By definition we have

$$\tilde\Xi = \big(p^{(m)}\|H\|_1\big)_{mm'} = \mathrm{diag}(p)\cdot\mathrm{ones}(n)\cdot\|H\|_1$$

where $\mathrm{ones}(n)$ denotes the n-dimensional square matrix with 1 for all its entries. Under Assumption A1, the spectral radius of $\tilde\Xi$ is $\|H\|_1\in(0,1)$. Therefore,

$$\tilde\Xi^j = \mathrm{diag}(p)\cdot\mathrm{ones}(n)\cdot\|H\|_1^j$$

where we use the fact $\sum_{m=1}^n p^{(m)} = 1$ in the identity. Thus,

$$\big(I-\tilde\Xi\big)^{-1} = I + \sum_{j\ge 1}\tilde\Xi^j = I + \mathrm{diag}(p)\cdot\mathrm{ones}(n)\cdot\frac{\|H\|_1}{1-\|H\|_1}$$

and from (2.8) and (2.9)

$$\tilde{\Sigma} = \mathrm{diag}\Big(\big(I - \tilde{\Xi}\big)^{-1} \big(p^{(m)}\lambda_0\big)_m \Big) = \mathrm{diag}(p) \frac{\lambda_0}{1 - \|H\|_1}.$$

We then have

$$\begin{aligned} &\big(I - \tilde{\Xi}\big)^{-1} \cdot \tilde{\Sigma} \cdot \big(I - \tilde{\Xi}^{\mathrm{T}}\big)^{-1} \\ &= \Big(I + \mathrm{diag}(p) \cdot \mathrm{ones}(n) \cdot \frac{\|H\|_1}{1 - \|H\|_1} \Big) \mathrm{diag}\big(\frac{\lambda_0 p^{(m)}}{1 - \|H\|_1}\big) \Big(I + \mathrm{ones}(n) \cdot \frac{\|H\|_1}{1 - \|H\|_1} \cdot \mathrm{diag}(p)^{\mathrm{T}} \Big) \\ &= \frac{\lambda_0}{1 - \|H\|_1} \mathrm{diag}(p) + \lambda_0 \frac{2\|H\|_1 - \|H\|_1^2}{(1 - \|H\|_1)^3} p \cdot p^{\mathrm{T}}, \end{aligned}$$

which coincides with the matrix in (3.7). □

3.2.1. *Another perspective via decomposed processes.* We have from [1, Lemma 4] that

$$N^{(m)}(t) - \mathbb{E}\big[N^{(m)}(t)\big] = X^{(m)}(t) + p^{(m)} Y(t)$$

where $X^{(m)}(t) = N^{(m)}(t) - \int_0^t \lambda^{(m)}(s)ds$ and

$$X(t) = \sum_{m=1}^{n} X^{(m)}(t) = N(t) - \int_0^t \lambda(s)ds \quad \text{and} \quad Y(t) = \int_0^t \varphi(t-s) X(s) ds.$$

In addition to the process $N_l^{(m)}$ in the proof of Proposition 2, we further define

$$X_l^{(m)}(t) = N_l^{(m)}(t) - \int_0^t \lambda_{l-1}^{(m)}(s)ds = N_l^{(m)}(t) - p^{(m)} \int_0^t \lambda_{l-1}(s)ds\,.$$

Then we also have $X^{(m)} = \sum_{l\geq 1} X_l^{(m)}$. By the conditional independence of $X_l^{(m)}$'s for $m = 1, \ldots, n$ and zero covariance for different l and l' by definition, we obtain immediately the following covariance function of $X^{(m)}$: for $m, m' = 1, \ldots, n$ and $t, s \geq 0$,

$$\begin{aligned} \mathrm{Cov}\Big(X^{(m)}(t), X^{(m')}(s) \Big) &= \mathbb{E}\Big[\sum_{l\geq 1} X_l^{(m)}(t) \sum_{l'\geq 1} X_{l'}^{(m')}(s) \Big] \\ = \delta_{mm'} \sum_{l\geq 1} \mathbb{E}\Big[X_l^{(m)}(t) X_l^{(m)}(s) \Big] &= \delta_{mm'} \sum_{l\geq 1} \mathbb{E}\Big[\int_0^{t\wedge s} \lambda_{l-1}^{(m)}(u) du \Big] \\ = \delta_{mm'} p^{(m)} \sum_{l\geq 1} \int_0^{t\wedge s} H^{(l-1)} * \lambda_0(u) du &= \delta_{mm'} p^{(m)} \lambda_0 \int_0^{t\wedge s} \big(1 + \varphi * 1(u)\big) du, \end{aligned}$$

where $\delta_{mm'} = 1$ if $m = m'$ and $\delta_{mm'} = 0$ if $m \neq m'$. Notice that $\big(X^{(m)}\big)_m$ is a martingale with respect to the natural filtration $\big\{\sigma\big\{\big(N^{(m)}(s)\big)_m, s \leq t\big\}, t \geq 0\big\}$ and its subprocesses $X_l^{(m)}$ do not jump at the same time.

Again we use the same scaling for the processes and quantities indexed by T. With the representations in the proposition, we can define $\big(N_T^{(m)}\big)_m$ and write

$$\hat{N}_T^{(m)}(t) := \frac{1}{\sqrt{T}} \Big(N_T^{(m)}(Tt) - \mathbb{E}\big[N_T^{(m)}(Tt)\big] \Big) = \hat{X}_T^{(m)}(t) + p_T^{(m)} \hat{Y}_T(t), \quad t \geq 0, \tag{3.13}$$

where for every $m = 1, 2, \cdots, n$ and $l \geq 1$

$$\hat{X}_{l,T}^{(m)}(t) = \frac{1}{\sqrt{T}} X_{l,T}^{(m)}(Tt) = \frac{1}{\sqrt{T}} \Big(N_{l,T}^{(m)}(t) - \int_0^t \lambda_{l-1,T}^{(m)}(s) ds \Big)$$

with

$$\hat{X}_T^{(m)}(t) = \sum_{l\geq 1} \hat{X}_{l,T}^{(m)}(t), \quad \hat{X}_T(t) = \sum_{m=1}^n \hat{X}_{l,T}^{(m)}(t) \quad \text{and} \quad \hat{Y}_T(t) = \int_0^t \varphi(T(t-s))\hat{X}_T(s)ds. \tag{3.14}$$

Then we have from the immigration-birth representation that

$$\mathbb{E}\Big[\bar{N}_T^{(m)}(t)\Big] = \lambda_{0,T} p_T^{(m)} \cdot \int_0^t \big(1 + \varphi * 1(Ts)\big)ds, \tag{3.15}$$

$$\mathrm{Cov}\Big(\hat{X}_T^{(m)}(t), \hat{X}_T^{(m')}(s)\Big) = \delta_{mm'}\lambda_{0,T} p_T^{(m)} \cdot \int_0^{t\wedge s} \big(1 + \varphi * 1(Tu)\big)du. \tag{3.16}$$

It is clear that given $\|H\|_1 \in (0,1)$ and Assumption A2, by (2.5), we obtain this covariance converges as $T \to \infty$ to

$$\delta_{mm'}\lambda_0 p^{(m)}(1 + \|\varphi\|_1)(t\wedge s) = \delta_{mm'}\lambda_0 p^{(m)}(1 - \|H\|_1)^{-1}(t \wedge s).$$

We observe that the two components in the expression of $\hat{N}_T^{(m)}$ in (3.13) are intrinsically correlated, which appear to be more complicated than the first expression in (3.5), but the martingale convergence method as in [1] can be applied and result in the following proposition (proof details are omitted for brevity).

Proposition 5. *Under the conditions of $\|H\|_1 \in (0,1)$ and Assumption A2, $\big(\hat{X}_T^{(m)}\big)_m \Rightarrow \big(\hat{X}^{(m)}\big)_m$ in $(\mathbb{D}^n, J_1)$ as $T \to \infty$, where $\big(\hat{X}^{(m)}\big)_m$ is an n-dimensional Brownian motion with covariance function*

$$\mathrm{Cov}\Big(\hat{X}^{(m)}(t), \hat{X}^{(m')}(s)\Big) = (t \wedge s) \cdot \delta_{mm'} \frac{\lambda_0 p^{(m)}}{1 - \|H\|_1}.$$

Thus, $\hat{Y}_T \Rightarrow \hat{Y}$ in $(\mathbb{D}, J_1)$ as $T \to \infty$, where

$$\hat{Y} = \frac{\|H\|_1}{1 - \|H\|_1}\hat{X} = \frac{\|H\|_1}{1 - \|H\|_1}\sum_{m=1}^n \hat{X}^{(m)}.$$

As a consequence, the process $\big(\hat{N}^{(m)}\big)_m$ can also be represented as

$$\hat{N}^{(m)} = \frac{\lambda_0^{1/2}}{(1 - \|H\|_1)^{1/2}}\Big(\sqrt{p^{(m)}}W^{(m)}\Big) + p^{(m)}\frac{\lambda_0^{1/2}\|H\|_1}{(1 - \|H\|_1)^{3/2}}W \tag{3.17}$$

where $\big(W^{(m)}\big)_m$ is a standard n-dimensional Brownian motion, and $W = \sum_{m=1}^n \sqrt{p^{(m)}}W^{(m)}$.

Remark 2. *We observe that we can rewrite the expression of the limit $\hat{N}^{(m)}$ in (3.17) as*

$$\hat{N}^{(m)} = \frac{\lambda_0^{1/2}}{(1 - \|H\|_1)^{1/2}}\hat{S}^{(m)} + p^{(m)}\frac{\lambda_0^{1/2}}{(1 - \|H\|_1)^{3/2}}W, \tag{3.18}$$

where $\hat{S}^{(m)} := \sqrt{p^{(m)}}W^{(m)} - p^{(m)}W$. It can be checked by direct calculations that the two components $(\hat{S}^{(m)})_m$ and W in the second expression in (3.18) are independent. Moreover, for every m and m',

$$\begin{aligned}\mathrm{Cov}\big(\hat{S}^{(m)}(t), \hat{S}^{(m')}(s)\big) &= \sqrt{p^{(m)}p^{(m')}}\mathbb{E}\big[W^{(m)}(t)W^{(m')}(s)\big] - p^{(m)}p^{(m')}\mathbb{E}\big[W^{(m)}(t)W^{(m)}(s)\big] \\ &\quad - p^{(m)}p^{(m')}\mathbb{E}\big[W^{(m')}(t)W^{(m')}(s)\big] + p^{(m)}p^{(m')}\mathbb{E}\big[W(t)W(s)\big] \\ &= p^{(m)}\big(\delta_{mm'} - p^{(m')}\big)(t \wedge s),\end{aligned}$$

which is exactly the identity in (3.6). This also proves the equivalence of the limit in (3.17) and that in Proposition 1, and thus, equivalence with that in Proposition 3.

4. Aggregating Hawkes processes

Splitting a Hawkes process results a multivariate Hawkes process with special exciting function. However, the aggregation of a multivariate Hawkes process is not necessarily a Hawkes process. In the following, we identify a necessary and sufficient condition for the aggregated multivariate Hawkes process to be a one-dimensional Hawkes process.

Let $N = (N_k)_k$ be a multivariate Hawkes process with conditional intensity in (2.1). The associated aggregating process of N is denoted as

$$A(t) = \sum_{k=1}^{d} N_k(t). \tag{4.1}$$

Then, similar to the proof of Proposition 2, by conditioning on the natural filtration of $(N_k)_k$, the conditional jumping intensity is given by

$$\lambda_A(t) = \sum_{k=1}^{d} \lambda_k(t) = \sum_{k=1}^{d} \lambda_{k0} + \sum_{k'=1}^{d} \int_0^t \Big(\sum_{k=1}^{d} H_{kk'}(t-s) \Big) N_{k'}(ds),$$

which is a process adapted to the natural filtration of $\big(N_k\big)_k$. Note that the filtration generated by the vector process N is strictly larger than that of A. Thus, to make λ_A above adapted to the natural filtration of A, an immediate observation is the following property.

Proposition 6. *$A = \{A(t), t \geq 0\}$ is a one-dimensional Hawkes process if and only if $\tilde{H} := \sum_{k=1}^{d} H_{kk'}$ is a function independent of k', under which the conditional intensity for A is*

$$\lambda_A(t) = \sum_{k=1}^{d} \lambda_{k0} + \int_0^t \tilde{H}(t-s) A(ds) = e^{\mathrm{T}} \cdot \lambda_0 + \int_0^t \tilde{H}(t-s) A(ds).$$

Remark 3. *As a very special case, suppose that N_k's are independent one-dimensional Hawkes processes with conditional intensity $\lambda_k(t) = \lambda_{k0} + \int_0^t H_k(t-s) N_k(ds)$. Then $\lambda_A(t) = \sum_{k=1}^{d} \lambda_{k0} + \sum_{k=1}^{d} \int_0^t H_k(t-s) N_k(ds)$. Thus, the aggregating process A is Hawkes with conditional intensity $\lambda_A(t) = e^{\mathrm{T}} \cdot \lambda_0 + \int_0^t H(t-s) A(ds)$ only if $H_k \equiv H$.*

Next we give another example to illustrate the property. Let N be a two-dimensional Hawkes process with positive baseline intensities $\lambda_{10}, \lambda_{20} > 0$ and with positive mutually exciting functions $(H_{ij})_{ij}$. Then the associated aggregating process A is a Hawkes process if and only if $H_{11} + H_{21} = H_{12} + H_{22}$.

Let $\hat{A}_T$ be the scaled aggregating process of the Hawkes process N,

$$\hat{A}_T(t) = \sqrt{T}\big(A(Tt) - \mathbb{E}\big[A(Tt)\big]\big) = e^{\mathrm{T}} \cdot \hat{N}_T(t).$$

Proposition 7. *Under Assumption A1, we have*

$$\hat{A}_T \Rightarrow \hat{A} = e^{\mathrm{T}} \cdot \big(I - \|H\|_1\big)^{-1} \cdot \Sigma^{1/2} \cdot W. \tag{4.2}$$

where W is a standard Brownian motion and $\Sigma = \mathrm{diag}\big((I - \|H\|_1)^{-1}\lambda_0\big)$ is the diagonal matrix in (2.9).

If, in addition, $\tilde{H} := \sum_{k=1}^{d} H_{kk'}$ is a function independent of k', then $\|\tilde{H}\|_1 \in (0,1)$ and $\hat{A}$ is a Brownian motion with diffusion coefficient $\sum_{k=1}^{d} \lambda_{k0}\big(1 - \|\tilde{H}\|_1\big)^{-3}$.

Proof. The limit result for $\hat{A}_T$ in (4.2) is a direct consequence of (2.7) for $\hat{N}_T$ and the continuous mapping theorem, which is a 1-dimensional Brownian motion with diffusion coefficient

$$e^{\mathrm{T}} \cdot \big(I - \|H\|_1\big)^{-1} \cdot \Sigma \cdot \big(I - \|H\|_1^T\big)^{-1} \cdot e.$$

For the special case that $\sum_k H_{kk'}$ is independent of k', since it is a Hawkes process, we can directly apply [1, Theorem 2] to the one-dimensional Hawkes process A with the second expression of $\lambda_A(t)$ in Proposition 6 and obtain the diffusion coefficient as stated. We next verify that the expression can be also obtained from that in (4.2). We have $e^{\mathrm{T}} \cdot H(t) = e^{\mathrm{T}} \tilde{H}(t)$ and

$$e^{\mathrm{T}} \cdot \|H\|_1 = e^{\mathrm{T}} \|\tilde{H}\|_1,$$

which shows that $\|\tilde{H}\|_1$ and e^{T} is the left eigenvalue and the associated eigenvector of kernel matrix $\|H\|_1$. Therefore, $\|\tilde{H}\|_1 \in (0,1)$ by Assumption A1,

$$e^{\mathrm{T}} \cdot \|H\|_1^j = e^{\mathrm{T}} \big(\|\tilde{H}\|_1\big)^j \quad \text{and} \quad e^{\mathrm{T}} \cdot \big(I - \|H\|_1\big)^{-1} = e^{\mathrm{T}} \big(1 - \|\tilde{H}\|_1\big)^{-1}$$

and the diffusion coefficient for $\hat{A}$ is

$$\begin{aligned} & e^{\mathrm{T}} \cdot (I - \|H\|_1)^{-1} \cdot \Sigma \cdot (I - \|H\|_1^{\mathrm{T}})^{-1} \cdot e = e^{\mathrm{T}} \cdot \Sigma \cdot e \big(1 - \|\tilde{H}\|_1\big)^{-2} \\ &= e^{\mathrm{T}} \cdot \big(I - \|H\|_1\big)^{-1} \cdot \lambda_0 \big(1 - \|\tilde{H}\|_1\big)^{-2} = e^{\mathrm{T}} \cdot \lambda_0 \big(1 - \|\tilde{H}\|_1\big)^{-3}. \end{aligned}$$

Here we make use of the fact that $\mathrm{diag}(u) \cdot e = u$ for a vector u. □

5. Splitting and aggregating multivariate Hawkes processes

Now, we consider randomly splitting and aggregating a multivariate Hawkes process $N = \big(N_k\big)_k$ in (2.1). Let $\{\xi_{kj}\}_{k,j}$ be the splitting variables, whenever $\xi_{kj} = m$, the j^{th} individual of the k^{th} component counting process N_k occurring at τ_{kj} is assigned to the m^{th} sub-counting process. Then, the $\mathbb{N}^d$-valued Hawkes process $\big(N_k\big)_k$ splits into an $\mathbb{N}^{d\times n}$-valued process $\big(N_k^{(m)}\big)_{k,m}$ and

$$N_k^{(m)}(t) = \sum_{j=1}^{N_k(t)} \mathbf{1}(\xi_{kj} = m) \quad \text{for every } k, m,$$

where $\{\xi_{kj}\}_j$ are i.i.d. variables, independent of N, with distribution

$$\mathbb{P}\big(\xi_{kj} = m\big) = p_k^{(m)} \quad \text{and} \quad \sum_{m=1}^{n} p_k^{(m)} = 1 \quad \text{for every } k, m.$$

We consider the following aggregated process:

$$A^{(m)}(t) = \sum_{k=1}^{d} N_k^{(m)}(t), \quad t \geq 0.$$

By the splitting and then aggregating procedure, we have transformed a d-dimensional Hawkes process into an n-dimensional counting process.

Before we proceed to study the properties of the split and aggregated processes, we discuss some potential applications. This often occurs when the demands require re-categorization in order for them to be processed. In an insurance company, demands may arrive as home or automobile insurances as a bivariate Hawkes process, while each type may split into claims and new personal/commercial services, which will be processed by the corresponding service departments. In a remanufacturing facility, different products may arrive as a multivariate Hawkes process, and then given the different component reprocessing needs, they will then be split and aggregated in order to be reprocessed at the corresponding machines. Similarly, in a data center, jobs may arrive as a multivariate Hawkes process, while they must be regrouped in order to be processed at the separate parallel servers due to computational requirements/constraints.

Similar to Propositions 2 and 6, the Hawkes property preserved for the splitting process indexed by (k,m), and the aggregated process $A^{(m)}$ is a Hawkes process under certain conditions.

Proposition 8. *The following properties hold:*

(i) $\big(N_k^{(m)}\big)_{k,m}$ *is an* $\mathbb{N}^{d\times n}$*-valued Hawkes process with conditional intensity*

$$\lambda_k^{(m)}(t) = p_k^{(m)}\lambda_k(t) = p_k^{(m)}\lambda_{k0} + \sum_{k',m'}\int_0^t \Big(p_k^{(m)}H_{kk'}(t-s)\Big)N_{k'}^{(m')}(ds)$$

where $(\lambda_k)_k$ *is the intensity for* N *in* (2.1).

(ii) $\big(A^{(m)}\big)_m$ *is an* n*-dimensional Hawkes process if and only if* $\tilde{H}^{(m)} := \sum_{k=1}^d p_k^{(m)}H_{kk'}$ *is a function independent of* $k' \in \mathcal{L}^{(m)} := \{k, p_k^{(m)} > 0\}$*, under which the conditional intensity of* $\big(A^{(m)}\big)_m$ *is given by*

$$\lambda_A^{(m)}(t) = \Big(\sum_{k=1}^d p_k^{(m)}\lambda_{k0}\Big) + \sum_{m'=1}^n \int_0^t \tilde{H}^{(m)}(t-s)A^{(m')}(ds).$$

Remark 4. *For a* d*-dimensional Hawkes process, it is always assumed to be irreducible, that is,* $\mathbb{E}\big(N_k(\infty)\big) > 0$ *for every* k*. The independency of* $\tilde{H}^{(m)}$ *in Proposition 6 is required for every* k'*. However, in the splitting and aggregating case,* $\mathbb{P}\big(N_k^{(m)}(\infty) = 0\big) = 1$ *for those* $p_k^{(m)} = 0$ *(that is,* $k \notin \mathcal{L}^{(m)}$*), the necessity and sufficiency of the condition can be checked by evaluating the intensity for the second occurrence time of* $A^{(m)}$ *in Proposition 8(i), which should remain unchanged for different* k'*'s.*

For example, let $d = n$ *and* $\{m_k\}_k$ *be a permutation of* $\{1, 2, \cdots, d\}$ *with* $p_k^{(m_k)} = 1$*. Then we have from this assumption that* $N_k^{(m)}(\infty) = 0$ *for* $m \neq m_k$ *and* $A^{(m_k)} = N_k$ *for every* k*. This means* $(A^{(m)})_m$ *defines a rotated Hawkes process which switches the positions of individuals. It is of course Hawkes, however,*

$$\sum_{k=1}^d p_k^{(m_{k_0})}H_{kk'} = p_{k_0}^{(m_{k_0})}H_{k_0k'} = H_{k_0k'}$$

depends on k' *for every* k_0*.*

We consider the scaled process indexed by T, and all the variables are marked with additional subscripts T, that is, $N_T = \big(N_{k,T}\big)_k$ is a Hawkes process whose baseline intensity is $\lambda_{0,T} = (\lambda_{k0,T})_k \in \mathbb{R}_+^d$ and the kernel matrix function $H \in \mathbb{R}_+^{d\times d}$ stays the same. The splitting variables $\{(\xi_{kj,T})_k\}_j$ have subscripts T and distribution matrix $(p_{k,T}^{(m)})_{k,m} \in \mathbb{R}_+^{d\times n}$, which results in the splitting process $\big(N_{k,T}^{(m)}\big)_{k,m}$. The average process and the diffusion-scaled process are defined by

$$\bar{N}_T(t) = \Big(\frac{1}{T}N_{k,T}^{(m)}(Tt)\Big)_{k,m} \quad \text{and} \quad \hat{N}_T(t) = \Big(\hat{N}_{k,T}^{(m)}(t)\Big)_{k,m} = \sqrt{T}\Big(\bar{N}_T(t) - \mathbb{E}\big[\bar{N}_T(t)\big]\Big).$$

Assumption A3. *Assume that for some* $\lambda_{k0}, p_k^{(m)} \geq 0$ *with* $\sum_{m=1}^n p_k^{(m)} = 1$ *for every* k*,*

$$\lambda_{k0,T} \to \lambda_{k0} \quad \textit{and} \quad p_{k,T}^{(m)} \to p_k^{(m)} \quad \textit{as} \quad T \to \infty.$$

We are interested in the FCLT of its aggregating process $(A_T^{(m)})_m$, defined by

$$\hat{A}_T^{(m)} = \sum_{k=1}^d \hat{N}_{k,T}^{(m)}.$$

Theorem 1. *Suppose Assumptions A1 and A3 hold. We have*

$$\big(\hat{A}_T^{(m)}\big)_m \Rightarrow \big(\hat{A}^{(m)}\big)_m \quad \textit{in } (\mathbb{D}^n, J_1)$$

where $(\hat{A}^m)_m$ *is a* n*-dimensional Brownian motion with covariance function*

$$\mathrm{Cov}\big(\hat{A}^{(m)}(t), \hat{A}^{(m')}(s)\big) = (t \wedge s) \times \Big(\mathrm{diag}\big(p^{\mathrm{T}} \cdot \Sigma \cdot e\big) - p^{\mathrm{T}} \cdot \Sigma \cdot p + p^{\mathrm{T}} \cdot (I - \|H\|_1)^{-1} \cdot \Sigma \cdot (I - \|H\|_1^{\mathrm{T}})^{-1} \cdot p\Big).$$

where $\Sigma = \mathrm{diag}\big((I - \|H\|_1)^{-1}\lambda_0\big)$ *is the diagonal matrix in* (2.8)*.*

Proof. Applying [1, Lemma 4] to $\big(N_{k,T}^{(m)}\big)_{k,m}$ we have the representation

$$\hat{N}_{k,T}^{(m)}(t) = \hat{X}_{k,T}^{(m)}(t) + p_{k,T}^{(m)}\hat{Y}_{k,T}(t) \tag{5.1}$$

where

$$\hat{X}_{k,T}^{(m)}(t) = \frac{1}{\sqrt{T}}\Big(N_{k,T}^{(m)}(Tt) - \int_0^{Tt} \lambda_{k,T}^{(m)}(s)ds\Big),$$

$$\hat{X}_T(t) = \big(\hat{X}_{k,T}(t)\big)_k = \sum_{m=1}^n \big(\hat{X}_{k,T}^{(m)}(t)\big)_k \quad \text{and} \quad \hat{Y}_{k,T}(t) = \int_0^t \mathrm{row}_k\big(\varphi(T(t-s))\big)\hat{X}_T(s)ds,$$

and where φ is the matrix function defined in (2.4). Moreover, similar to (3.15) and (3.16),

$$\mathbb{E}\big[\bar{N}_{k,T}^{(m)}(t)\big] = p_{k,T}^{(m)} \cdot \mathrm{ent}_k\Big(\int_0^t (I + \varphi * 1(Ts))ds \cdot \lambda_{0,T}\Big),$$

$$\mathrm{Cov}\Big(\hat{X}_{k,T}^{(m)}(t), \hat{X}_{k',T}^{(m')}(s)\Big) = \delta_{kk'}\delta_{mm'}p_{k,T}^{(m)} \cdot \mathrm{ent}_k\Big(\int_0^t (I + \varphi * 1(Ts))ds \cdot \lambda_{0,T}\Big).$$

Therefore, we have

$$\Big(\big(\hat{X}_{k,T}^{(m)}\big)_{k,m}, \big(\hat{Y}_{k,T}\big)_k\Big) \Rightarrow \Big(\big(\hat{X}_k^{(m)}\big)_{k,m}, \big(\hat{Y}_k\big)_k\Big)$$

where for some standard Brownian motion $(W_k^{(m)})_{k,m}$

$$\hat{X}_k^{(m)} = \sqrt{p_k^{(m)}\Sigma_{kk}}W_k^{(m)}, \quad \hat{Y}_k = \mathrm{row}_k\big(\|\varphi\|_1\big)\sum_{m=1}^n \Big(\hat{X}_{k'}^{(m)}\Big)_{k'} \quad \text{and} \quad \hat{N}_k^{(m)} = \hat{X}_k^{(m)} + p_k^{(m)}\hat{Y}_k.$$

A direct calculation shows that

$$\mathrm{Cov}\Big(\hat{N}_k^{(m)}(t), \hat{N}_{k'}^{(m')}(s)\Big) = (t \wedge s) \times \Big(\delta_{kk'}\Sigma_{kk}p_k^{(m)}\big(\delta_{mm'} - p_{k'}^{(m')}\big) + p_k^{(m)}p_{k'}^{(m')}\mathrm{ent}_{kk'}\big((I - \|H\|_1)^{-1} \cdot \Sigma \cdot (I - \|H\|_1^{\mathrm{T}})^{-1}\big)\Big).$$

Summing over k and k' gives the covariance function for $\big(\hat{A}^{(m)}\big)_m$. □

Remark 5. *From the proof, we can have a second representation for* $\big(\hat{N}_k^{(m)}\big)_{k,m}$*:*

$$\hat{N}_k^{(m)} = \Sigma_{kk}^{1/2}\hat{S}_k^{(m)} + p_k^{(m)}\mathrm{row}_k\big((I - \|H\|_1)^{-1}\big)\Sigma^{1/2}W \tag{5.2}$$

where $\big(W_k^{(m)}\big)_{k,m}$ *is the standard Brownian motions in the proof, and*

$$W_k = \sum_{m=1}^n \sqrt{p_k^{(m)}}W_k^{(m)}, \quad \hat{S}_k^{(m)} = \sqrt{p_k^{(m)}}W_k^{(m)} - p_k^{(m)}W_k \quad \textit{and} \quad W = (W_k)_k \in \mathbb{R}^d.$$

Then, $\big(W_k\big)$ *is an* n*-dimensional standard Brownian motion, and* $\big\{(\hat{S}_k^{(m)})_m, k = 1, 2, \cdots, d\big\}$ *is a sequence of Brownian motions independent to* $(W_k)_k$ *and over* k*, and*

$$\mathrm{Cov}\big(\hat{S}_k^{(m)}(t), \hat{S}_k^{(m')}(s)\big) = (t \wedge s) \times p_k^{(m)}\big(\delta_{mm'} - p_k^{(m')}\big).$$

Under the conditions of Proposition 8(ii), $(A_T^{(m)})_m$ becomes a multivariate Hawkes process. Thus, [1, Theorem 2] can be applied directly and we can have a second representation for its covariances.

Proposition 9. *Suppose $\tilde{H}_T^{(m)} = \sum_{k=1}^d p_{k,T}^{(m)} H_{kk'}$ is independent of k', and Assumptions A1 and A3 hold, then the covariance function for $\hat{A} = \big(\hat{A}^{(m)}\big)_m$ in Theorem 1 can also be represented as*

$$\big(I - \tilde{\Xi}\big)^{-1} \cdot \mathrm{diag}\Big(\big(I - \tilde{\Xi}\big)^{-1} \cdot p^{\mathrm{T}} \cdot \lambda_0\Big) \cdot \big(I - \tilde{\Xi}^{\mathrm{T}}\big)^{-1}, \tag{5.3}$$

where $\tilde{\Xi} = p^{\mathrm{T}} \cdot \mathrm{col}_1\big(\|H\|_1\big) \cdot e_n^{\mathrm{T}}$, and e_n is the n-dimensional column vector of ones.

Proof. By assumption, $(A_T^{(m)})_m$ is a Hawkes process with conditional intensity

$$\lambda_{A,T}^{(m)}(t) = \Big(\sum_{k=1}^d p_{k,T}^{(m)} \lambda_{k0,T}\Big) + \sum_{m'=1}^n \int_0^t \tilde{H}_T^{(m)}(t-s) A_T^{(m')}(ds).$$

The representation for covariance functions in (5.3) follows from (2.7) from [1, Theorem 2]. In the following, we show that (5.3) coincides with the covariance in Theorem 1 in this special case.

Denoting by $\nu = p^{\mathrm{T}} \cdot \mathrm{col}_1\big(\|H\|_1\big)$ and $c_0 = \nu^{\mathrm{T}} \cdot e_n$, we have

$$\tilde{\Xi} = \nu \cdot e_n^{\mathrm{T}}, \quad p^{\mathrm{T}} \cdot \|H\|_1 = \nu \cdot e_d^{\mathrm{T}} \quad \text{and} \quad e_d^{\mathrm{T}} \cdot \|H\|_1 = c_0 \cdot e_d^{\mathrm{T}} \tag{5.4}$$

by the fact $p \cdot e_n = e_d$. Thus, c_0 and e_d^{T} is the right-eigenvalue and the associated eigenvector of $\|H\|_1$, and $c_0 \in (0,1)$ by Assumption A1. We futher have for $\tilde{\Xi}$,

$$\tilde{\Xi}^j = \nu \cdot e_n^{\mathrm{T}} \big(e_n^{\mathrm{T}} \cdot \nu\big)^{j-1} = c_0^{j-1} \cdot \tilde{\Xi} \quad \text{and} \quad \big(I - \tilde{\Xi}\big)^{-1} = I + \frac{1}{1-c_0} \nu \cdot e_n^{\mathrm{T}}$$

and

$$\big(I - \tilde{\Xi}\big)^{-1} \cdot p^{\mathrm{T}} \cdot \lambda_0 = p^{\mathrm{T}} \cdot \lambda_0 + \frac{1}{1-c_0} \nu \cdot e_n^{\mathrm{T}} \cdot p^{\mathrm{T}} \cdot \lambda_0 = p^{\mathrm{T}} \cdot \lambda_0 + \frac{e_d^{\mathrm{T}} \cdot \lambda_0}{1-c_0} \nu.$$

Therefore, we have

$$\begin{aligned}
&\big(I - \tilde{\Xi}\big)^{-1} \cdot \mathrm{diag}\Big(\big(I - \tilde{\Xi}\big)^{-1} \cdot p^{\mathrm{T}} \cdot \lambda_0\Big) \cdot \big(I - \tilde{\Xi}^{\mathrm{T}}\big)^{-1} \\
&= \mathrm{diag}\Big(p^{\mathrm{T}} \cdot \lambda_0 + \frac{e_d^{\mathrm{T}} \cdot \lambda_0}{1-c_0} \nu\Big) + \frac{1}{(1-c_0)^2} \nu \cdot e_n^{\mathrm{T}} \cdot \Big(p^{\mathrm{T}} \cdot \lambda_0 \cdot \nu^{\mathrm{T}} + \frac{e_d^{\mathrm{T}} \cdot \lambda_0}{1-c_0} \nu \cdot \nu^{\mathrm{T}}\Big) \\
&\quad + \frac{1}{1-c_0} \Big(p^{\mathrm{T}} \cdot \lambda_0 \cdot \nu^{\mathrm{T}} + \nu \cdot \lambda^{\mathrm{T}} \cdot p + 2 \frac{e_d^{\mathrm{T}} \cdot \lambda_0}{1-c_0} \nu \cdot \nu^{\mathrm{T}}\Big) \\
&= \mathrm{diag}\Big(p^{\mathrm{T}} \cdot \lambda_0 + \frac{e_d^{\mathrm{T}} \cdot \lambda_0}{1-c_0} \nu\Big) + \frac{1}{1-c_0} \Big(p^{\mathrm{T}} \cdot \lambda_0 \cdot \nu^{\mathrm{T}} + \nu \cdot \lambda_0^{\mathrm{T}} \cdot p\Big) \\
&\quad + \big(\nu \cdot \nu^{\mathrm{T}}\big) \Big(\frac{2 e_d^{\mathrm{T}} \cdot \lambda_0}{(1-c_0)^2} + \frac{e_d^{\mathrm{T}} \cdot \lambda_0}{(1-c_0)^3}\Big),
\end{aligned} \tag{5.5}$$

where we make use of the fact that $\mathrm{diag}(u) \cdot e = u$ for vector u in the calculation.

On the other hand, for the covariance function in Theorem 1 in this special case, we have

$$\begin{aligned}
p^{\mathrm{T}} \cdot \big(I - \|H\|_1\big)^{-1} &= p^{\mathrm{T}} + \sum_{j \geq 1} p^{\mathrm{T}} \cdot \|H\|_1^j = p^{\mathrm{T}} + \sum_{j \geq 1} \nu \cdot e_d^{\mathrm{T}} \cdot \|H\|_1^{j-1} \\
&= p^{\mathrm{T}} + \frac{1}{1-c_0} \, \nu \cdot e_d^{\mathrm{T}},
\end{aligned} \tag{5.6}$$

where the fact from (5.4) is used. Thus, by the definition $\Sigma \cdot e_d = (I - \|H\|_1)^{-1} \cdot \lambda_0$, we have

$$\begin{aligned} p^{\mathrm{T}} \cdot \Sigma \cdot e_d &= p^{\mathrm{T}} \cdot \big(I - \|H\|_1\big)^{-1} \cdot \lambda_0 = p^{\mathrm{T}} \cdot \lambda_0 + \nu \frac{e_d^{\mathrm{T}} \cdot \lambda_0}{1 - c_0}, \\ p^{\mathrm{T}} \cdot \Sigma \cdot e_d \cdot \nu^{\mathrm{T}} &= p^{\mathrm{T}} \cdot \lambda_0 \cdot \nu^{\mathrm{T}} + \frac{e_d^{\mathrm{T}} \cdot \lambda_0}{1 - c_0} \big(\nu \cdot \nu^{\mathrm{T}}\big), \\ e_d^{\mathrm{T}} \cdot \Sigma \cdot e_d &= e_d^{\mathrm{T}} \cdot \big(I - \|H\|_1\big)^{-1} \cdot \lambda_0 = \frac{e_d^{\mathrm{T}} \cdot \lambda_0}{1 - c_0}, \end{aligned} \tag{5.7}$$

where the fact $e_d^{\mathrm{T}} \cdot \|H\|_1 = c_0 \cdot e_d^{\mathrm{T}}$ is used.

Therefore, by applying the identities in (5.6) and (5.7), we obtain

$$\begin{aligned} &\mathrm{diag}\big(p^{\mathrm{T}} \cdot \Sigma \cdot e\big) + p^{\mathrm{T}} \cdot (I - \|H\|_1)^{-1} \cdot \Sigma \cdot (I - \|H\|_1^{\mathrm{T}})^{-1} \cdot p - p^{\mathrm{T}} \cdot \Sigma \cdot p \\ &= \mathrm{diag}\big(p^{\mathrm{T}} \cdot \Sigma \cdot e\big) + \frac{1}{1 - c_0}\Big(p^{\mathrm{T}} \cdot \Sigma \cdot e_d \cdot \nu^{\mathrm{T}} + \nu \cdot e_d \cdot \Sigma \cdot p\Big) + \frac{1}{(1 - c_0)^2}\Big(\nu \cdot e_d^{\mathrm{T}} \cdot \Sigma \cdot e_d \cdot \nu^{\mathrm{T}}\Big) \\ &= \mathrm{diag}\Big(p^{\mathrm{T}} \cdot \lambda_0 + \nu \frac{e_d^{\mathrm{T}} \cdot \lambda_0}{1 - c_0}\Big) + \frac{1}{1 - c_0}\Big(p^{\mathrm{T}} \cdot \lambda_0 \cdot \nu^{\mathrm{T}} + \nu \cdot \lambda_0^{\mathrm{T}} \cdot p + \nu \cdot \nu^{\mathrm{T}} \frac{2 e_d^{\mathrm{T}} \cdot \lambda_0}{(1 - c_0)}\Big) + \nu \cdot \nu^{\mathrm{T}} \Big(\frac{e_d^{\mathrm{T}} \cdot \lambda_0}{(1 - c_0)^3}\Big). \end{aligned}$$

Comparing with the expression in (5.5), we conclude the equivalence. □

References

[1] E. Bacry, S. Delattre, M. Hoffmann, and J. F. Muzy. Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Processes and their Applications*, 123(7):2475–2499, 2013.

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

[3] P. Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. Wiley, New York, NY. 2nd edition, 1999. A Wiley-Interscience Publication.

[4] X. Chen. Perfect sampling of Hawkes processes and queues with Hawkes arrivals. *Stochastic Systems*, 11(3), 264–283, 2020.

[5] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes Volume I: Elementary Theory and Methods*. Springer, New York, 2nd edition, 2003.

[6] A. Daw and J. Pender. Queues driven by Hawkes processes. *Stochastic Systems*, 8(3):192–229, 2018.

[7] Fuqing Gao and Lingjiong Zhu. Some asymptotic results for nonlinear Hawkes processes. *Stochastic Processes and their Applications*, 128(12):4051–4077, 2018.

[8] X. Gao and L. Zhu. Functional central limit theorems for stationary Hawkes processes and application to infinite-server queues. *Queueing Systems*, 90(1-2):161–206, 2018.

[9] A. G. Hawkes. Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 33(3):438–443, 1971.

[10] A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

[11] A. G. Hawkes. Hawkes processes and their applications to finance: a review. *Quantitative Finance*, 18(2):193–198, 2017.

[12] A. G. Hawkes and D. Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11(3):493–503, 1974.

[13] D. Karabash and L. Zhu. Limit theorems for marked Hawkes processes with application to a risk model. *Stochastic Models*, 31(3):433–451, 2015.

[14] D. T. Koops, M. Saxena, O. J. Boxma, and M. Mandjes. Infinite-server queues with Hawkes input. *Journal of Applied Probability*, 55(3):920–943, 2018.

[15] B. Li and G. Pang. Functional limit theorems for nonstationary marked Hawkes processes in the high intensity regime. *Stochastic Processes and their Applications*, Vol. 143, 285–339, 2021.

[16] D. Oakes. The Markovian self-exciting process. *Journal of Applied Probability*, 12(1):69–77, 1975.

[17] K. Sriram and W. Whitt. Characterizing superposition arrival processes and the performance of multiplexers for voice and data. *Proceedings of IEEE Global Telecommunications Conference, New Orleans*, 2:25.4.1-7, 1985.

[18] K. Sriram and W. Whitt. Characterizing superposition arrival processes in packet multiplexers for voice and data. *IEEE Journal on Selected Areas in Communications*, SAC-4 (6):833–846, 1986.

[19] W. Whitt. Limits for the superposition of m-dimensional point processes. *Journal of Applied Probability*, 9(2):462–465, 1972.

[20] W. Whitt. Queues with superposition arrival processes in heavy traffic. *Stochastic Processes and Their Applications*, 21(1):81–91, 1985.

[21] W. Whitt. *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues*. Springer New York, 2002.

School of Mathematics and LPMC, Nankai University, Tianjin, 300071 China
Email address: libo@nankai.edu.cn

Department of Computational Applied Mathematics and Operations Research, George R. Brown College of Engineering, Rice University, Houston, TX 77005
Email address: gdpang@rice.edu