Adv. Appl. Prob. **48,** 463–480 (2016)
*doi:*10.1017/apr.2016.10

POPULATION VIEWPOINT ON HAWKES PROCESSES

ALEXANDRE BOUMEZOUED,* *Paris 6 University*

Abstract

In this paper we focus on a class of linear Hawkes processes with general immigrants. These are counting processes with shot-noise intensity, including self-excited and externally excited patterns. For such processes, we introduce the concept of the age pyramid which evolves according to immigration and births. The virtue of this approach that combines an intensity process definition and a branching representation is that the population age pyramid keeps track of all past events. This is used to compute new distribution properties for a class of Hawkes processes with general immigrants which generalize the popular exponential fertility function. The pathwise construction of the Hawkes process and its underlying population is also given.

Keywords: Hawkes process; branching; immigration; age pyramid; nonstationarity; Laplace transform; thinning; Poisson point measure

2010 Mathematics Subject Classification: Primary 60G55
Secondary 60G57; 60G44; 60J80

1. Introduction

In this paper we investigate the link between some population dynamics models and a class of Hawkes processes. We are interested in processes whose behavior is modified by past events, which are self-excited and externally excited. The introduction of a self-excited process with shot-noise intensity is due to Hawkes (1971) and the famous Hawkes process has since been used in a variety of applications, including seismology, neuroscience, epidemiology, insurance and finance, to name but a few. The shot-noise intensity of the Hawkes process (N_t) is expressed as $\lambda_t = \mu + \sum_{T_n<t} \phi(t - T_n)$, where the T_n are the jump times of the Hawkes process N itself, $\mu > 0$, and ϕ is a nonnegative function. In the Hawkes model, when an event occurs at time T_n, the intensity grows by an amount $\phi(t - T_n)$: this models the self-exciting property. Also, for many modeling purposes, ϕ returns to 0 as t increases, so that the self-excitation vanishes after a long time. On the whole, each event excites the system as it increases its intensity, but this increase vanishes with time as it is natural to the model the fact that very old events have a negligible impact on the current behavior of the process. In the literature, several contributions focused on processes with self-exciting behavior and also some externally exciting component. To the best of our knowledge, the Hawkes process with general immigrants was introduced in Brémaud and Massoulié (2002), and specific forms can also be found in recent studies motivated by financial applications, such as Dassios and Zhao (2011), Wheatley *et al.* (2014), and Rambaldi *et al.* (2014), where external shocks, new arrivals, and contagion are crucial to model. In this paper we are interested in a class of Hawkes processes with general immigrants

Received 26 May 2015; revision received 12 July 2015.
* Current address: Milliman, 14 Rue Pergolèse, 75016 Paris, France.
Email address: alexandre.boumezoued@milliman.com

(see Brémaud and Massoulié (2002)), whose intensity is of the form

$$\lambda_t = \mu(t) + \sum_{T_n<t} \Phi_t(t - T_n, X_n) + \sum_{S_k<t} \Psi_t(t - S_k, Y_k).$$

In this model, the T_n are the jump times of N: if an event occurs for the system at time T_n, the intensity grows by an amount $\Phi_t(t - T_n, X_n)$, where X_n is some mark. This part models the self-exciting property. In parallel, external events occur at times S_k and excite the system of interest with some amount $\Psi_t(t - S_k, Y_k)$: this is the externally excited component.

Among the appealing properties of such models, one of them comes from the shot-noise form of the intensity. This is called the cluster (or branching) representation of the Hawkes process, and it is based on the following remark: if an event occurred at time T_n then $t - T_n$ is nothing but the 'age' of this event at time t. Following the seminal work of Hawkes (1971), Hawkes and Oakes (1974) proposed the cluster representation of the self-exciting process. They interpreted it as an immigration–birth process with age: they proved that under some stationarity conditions, it can be described as a branching Poisson process (also called a Poisson cluster). Also, in Dassios and Zhao (2011), a definition of a dynamic contagion process is given through its cluster representation. Until now, most studies on the Hawkes process recalled the immigration–birth representation as follows: immigrants arrive at times given by a Poisson process with intensity μ. Then each immigrant starts a new generation: it gives birth to new individuals with fertility function ϕ, each one giving birth with the same fertility function ϕ. This is often used as a definition for the Hawkes process, providing a good intuition on its behavior. The cluster representation of Hawkes and Oakes (1974) requires that the mean number of children per individual which is nothing but $\|\phi\| = \int_0^\infty \phi(a)\,\mathrm{d}a$ satisfies $\|\phi\| < 1$. In this paper we exhibit the immigration–birth dynamics underlying Hawkes processes with general immigrants which do not require the stationary assumption. The virtue of this approach that combines an intensity process definition and a branching representation is that the population age pyramid keeps track of all past events. This is used to compute new distribution properties for a class of linear Hawkes processes with general immigrants.

In the literature, the distribution properties of the Hawkes process have first been studied under stationary conditions. Hawkes (1971) addressed second-order stationary properties, whereas Adamopoulos (1975) derived the probability generating functional under stationarity by using the cluster representation of Hawkes and Oakes (1974). In this work, Adamopoulos (1975) expressed the probability generating function as a solution to some functional equation. Furthermore, Brémaud and Massoulié (2002) introduced the framework for studying moments of the stationary Hawkes process by means of the Bartlett spectrum. Let us also mention two recent studies of the distribution properties under stationarity. The moment generating function has been expressed in Saichev and Sornette (2011) as a solution to some transcendental equation. In addition, Jovanović *et al.* (2014) proposed a graphical way to derive closed-form solutions for cumulant densities, leading to the moments of the stationary Hawkes process. It is interesting to note that such recent contributions rely on the stationary branching representation of Hawkes and Oakes (1974). Recently, the computation of statistical properties has gained attention under nonstationarity, both for mathematical analysis and statistical estimation techniques. However, the recent studies in this framework only focus on exponential fertility rates $\phi(t) = \alpha e^{\beta t}$. The tool they rely on is the infinitesimal generator of the intensity process (λ_t) which is Markovian for such exponential fertility rates; see Oakes (1975). This includes the work of Errais *et al.* (2010), Aït-Sahalia *et al.* (2010), Dassios and Zhao (2011), and Da Fonseca and Zaatour

(2014). In this paper we generalize these studies in a natural direction for a wider class of Hawkes processes.

Scope of this paper. The aim of this paper is threefold. First, to introduce the concept of age pyramid for Hawkes processes with general immigrants. Second, to use this concept to compute new distribution properties for a class of fertility functions which generalize the popular exponential case. Finally, to give a pathwise representation of the general Hawkes processes and its underlying immigration–birth dynamics. We represent the population as a multi-type dynamics with ages, including immigration and births with mutations. Our population point of view is inspired by Bensusan *et al.* (2015) (see also Tran (2008)) and seems to reconcile the two definitions of Hawkes processes, through an intensity process or a branching dynamics.

The paper is organized as follows. In Section 2 we introduce the population point of view for Hawkes processes with general immigrants and study the dynamics of the age pyramid over time. In Section 3 we use this concept to compute the dynamics and Laplace transform of a class of Hawkes processes with general immigrants whose fertility functions generalize the popular exponential case. In Section 4 we present the pathwise construction of Hawkes processes with general immigrants and its underlying population. Finally, Section 5 details some results on the special case of standard Hawkes processes, including its Laplace transform and two first-order moments.

2. Population point of view

In this paper we focus on a class of counting processes named as Hawkes processes with general immigrants (see Brémaud and Massoulié (2002)), which is defined below through its intensity process. Existence and uniqueness issues will be discussed in Section 4. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space satisfying the usual conditions. Recall that the intensity process (λ_t) of a counting process (N_t) is the $(\mathcal{F}_t^N)$-predictable process such that $N_t - \int_0^t \lambda_s \, \mathrm{d}s$ is an $(\mathcal{F}_t^N)$-local martingale, where $(\mathcal{F}_t^N)$ denotes the canonical filtration of (N_t).

Definition 1. A Hawkes process with general immigrants is a counting process (N_t) whose intensity is given by

$$\lambda_t = \mu(t) + \sum_{T_n < t} \Phi_t(t - T_n, X_n) + \sum_{S_k < t} \Psi_t(t - S_k, Y_k), \tag{1}$$

where the T_n are the jump times of N, the S_k are those of a counting process with deterministic intensity $\rho(t)$, and the X_n (respectively Y_k) are real positive independent and identically distributed with distribution G (respectively H). The (S_k), (Y_k), and (X_n) are assumed to be independent of each other.

In this model, the T_n are the jump times of N: if an event occurs for the system at time T_n, the intensity grows by an amount $\Phi_t(t-T_n, X_n)$, where X_n is some mark. This part models the self-exciting property. In parallel, external events occur at times S_k and excite the system of interest with some amount $\Psi_t(t-S_k, Y_k)$: this is the externally excited component. The Hawkes process with general immigrants was introduced and studied under stationary conditions by Brémaud and Massoulié (2002). Due to their flexibility and natural interpretation, such models have gained recent attention for financial applications; see, e.g. Dassios and Zhao (2011), Wheatley *et al.* (2014), and Rambaldi *et al.* (2014). In particular, distribution properties of such processes have been investigated by Dassios and Zhao (2011) in the case $\Phi_t(a, x) = \Psi_t(a, x) = x\mathrm{e}^{-\delta a}$, in which framework the intensity process is Markovian. Our aim is to study the dynamics and

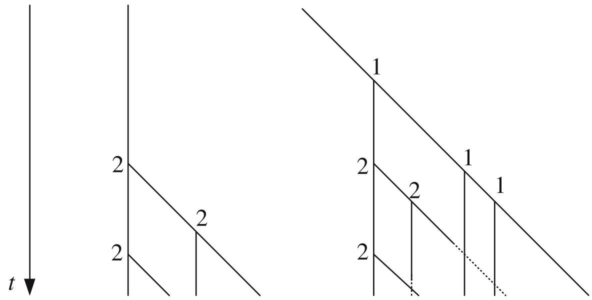

FIGURE 1: Population dynamics of the Hawkes process with general immigrants.

characterize the distribution of the nonstationary Hawkes process with general immigrants for a larger class of fertility functions, possibly time-dependent, which extends the previous work of Dassios and Zhao (2011) in this direction. To do this, we first represent it as a two-population immigration–birth dynamics with ages and characteristics.

Thanks to Definition 1, we obtain a representation of the intensity process. But in fact, the whole information on the dynamics is lost. Indeed, it is interesting to go back to the branching representation of Hawkes and Oakes (1974) to have in mind the underlying population dynamics. For the standard Hawkes process with intensity $\lambda_t = \mu + \sum_{T_n<t} \phi(t - T_n)$, we have the following interpretation. First, immigrants arrive according to a Poisson process with parameter μ, then each immigrant generates a cluster of descendants with the following rule: if an individual arrived or was born at some time T_n, it gives birth to new individuals with rate $\phi(t - T_n)$ at time t, where in fact $t - T_n$ is nothing but the *age* at time t of the individual. In the case of the Hawkes process with general immigrants (see Definition 1), the description of the dynamics is very similar, except that we have two populations: population (1) represents external shocks that occurred at times S_k, whereas population (2) represents internal shocks that occurred at times T_n. If an individual is born at time S_k (respectively T_n), we call the mark Y_k (respectively X_n) its *characteristic*. Then the immigration–birth dynamics can be described as follows.

(i) Let us first describe the population (1) of external shocks. It is made with immigrants that arrive in population (1) with rate $\rho(t)$; at arrival, they have age 0 and some characteristic x drawn with distribution H. Any individual, denoted (a, x), with age a and characteristic x at time t that belongs to population (1) gives birth with rate $\Psi_t(a, x)$. The newborn belongs to population (2); it has age 0 and some characteristic drawn with distribution G.

(ii) Let us now complete the description of population (2). In addition to births from population (1), the population (2) evolves according to two other kinds of events: immigration and internal birth. Immigrants arrive in population (2) with rate $\mu(t)$ with age 0 and a characteristic drawn with distribution G. Any individual (a, x) at time t that belongs to population (2) gives birth with rate $\Phi_t(a, x)$. The newborn also belongs to population (2); it has age 0 and some characteristic drawn with distribution G. In the end, the Hawkes process with general immigrants can be recovered as the size of population (2), therefore this construction can be seen as another definition of such a process. The dynamics are illustrated in Figure 1.

Since the immigration–birth mechanism is crucial to understand the behavior of the Hawkes dynamics, our aim now is to keep track of all ages and characteristics in each population (i), i =1 or 2. One way to address this issue is to count the number of individuals with age below $\bar{a} > 0$ and a characteristic in some set $[0, \bar{x}] \subset \mathcal{X}$ at time t, denoted $Z_t^{(i)}([0, \bar{a}], [0, \bar{x}])$. This can be computed for example for population (2) in the following way:

$$Z_t^{(2)}([0, \bar{a}], [0, \bar{x}]) = \sum_{T_n \le t} \mathbf{1}_{[0,\bar{a}]}(t - T_n)\mathbf{1}_{[0,\bar{x}]}(X_n).$$

This way, each population (i), $i = 1$ or 2, is represented at time t as a measure which puts a weight on the age and characteristic of each individual, denoted $Z_t^{(i)}(\mathrm{d}a, \mathrm{d}x)$. The two measures which we call the *age pyramid* are introduced in the following definition.

Definition 2. We denote *age pyramids at time t* by the following two measures:

$$Z_t^{(1)}(\mathrm{d}a, \mathrm{d}x) = \sum_{S_k \le t} \delta_{(t-S_k, Y_k)}(\mathrm{d}a, \mathrm{d}x), \qquad Z_t^{(2)}(\mathrm{d}a, \mathrm{d}x) = \sum_{T_n \le t} \delta_{(t-T_n, X_n)}(\mathrm{d}a, \mathrm{d}x). \quad (2)$$

The virtue of the measure representation is that one can compute time-dependent functions of the population age pyramid. Consider a function $f_t(a, x)$ depending on time, and also on ages and characteristics of individuals. This can be computed on the overall population with the following notation:

$$\langle Z_t^{(i)}, f_t\rangle = \int_{\mathbb{R}_+ \times \mathbb{R}_+} f_t(a, x) Z_t^{(i)}(\mathrm{d}a, \mathrm{d}x) \quad \text{for } i = 1 \text{ or } 2. \quad (3)$$

For example, the Hawkes process is $N_t^{(2)} = \langle Z_t^{(2)}, \mathbf{1}\rangle$, whereas the number of external shocks is $N_t^{(1)} = \langle Z_t^{(1)}, \mathbf{1}\rangle$. Also, the intensity λ_t of the Hawkes process $N_t^{(2)}$ given in (1) can be written as

$$\lambda_t = \mu(t) + \langle Z_{t-}^{(2)}, \Phi_t\rangle + \langle Z_{t-}^{(1)}, \Psi_t\rangle. \quad (4)$$

Viewed as a stochastic process, $(Z_t^{(1)}(\mathrm{d}a, \mathrm{d}x), Z_t^{(2)}(\mathrm{d}a, \mathrm{d}x))_{t \ge 0}$ is a (two-dimensional) measure-valued process. In fact, this age pyramid process is a Markov process; see Tran (2006). Note, however, that its differentiation in time is not straightforward; see Bensusan *et al.* (2015) and Lemma 1 below. The Markov property of the age pyramid process shows that all the information needed is contained in the population age structure. Let us mention the seminal point of view of Harris (1963), for whom

> it does seem intuitively plausible that we obtain a Markov process, in an extended sense, if we describe the state of the population at time t not simply by the number of objects present but by a list of the ages of all objects.

However, in practice this information is 'too large' to perform tractable computations. In the next section we illustrate how to identify some minimal components to add to the Hawkes process in order to make the dynamics Markovian. To do this, we first need to address the dynamics of the age pyramid. This is stated in the following lemma.

Lemma 1. *For each function* $f: (t, x, a) \mapsto f_t(a, x)$ *differentiable in t and a, the dynamics of the process* $\langle Z_t^{(i)}, f_t\rangle$ *for* $i = 1$ *or* 2 *is given by*

$$\mathrm{d}\langle Z_t^{(i)}, f_t\rangle = \int_{\mathbb{R}_+} f_t(0, x) N^{(i)}(\mathrm{d}t, \mathrm{d}x) + \langle Z_t^{(i)}, (\partial_a + \partial_t) f_t\rangle \,\mathrm{d}t, \quad (5)$$

where the point measures $N^{(1)}$ *and* $N^{(2)}$ *are given by*

$$N^{(1)}(\mathrm{d}t, \mathrm{d}x) = \sum_{k\geq 1} \delta_{(S_k, Y_k)}(\mathrm{d}t, \mathrm{d}x), \qquad N^{(2)}(\mathrm{d}t, \mathrm{d}x) = \sum_{n\geq 1} \delta_{(T_n, X_n)}(\mathrm{d}t, \mathrm{d}x).$$

Proof. Let us first remark that, by (3),

$$\langle Z_t^{(i)}, f_t \rangle = \int_{(0,t]\times\mathbb{R}_+} f_t(t-s, x) N^{(i)}(\mathrm{d}s, \mathrm{d}x). \tag{6}$$

Then write between s and t, $f_t(t-s, x) = f_s(0, x) + \int_s^t (\partial_a + \partial_u) f_u(u-s, x)\,\mathrm{d}u$ and use it in (6) to obtain

$$\begin{aligned}\langle Z_t^{(i)}, f_t \rangle &= \int_{(0,t]\times\mathbb{R}_+} f_s(0, x) N^{(i)}(\mathrm{d}s, \mathrm{d}x) \\ &\quad + \int_{(0,t]\times\mathbb{R}_+} \bigg(\int_s^t (\partial_a + \partial_u) f_u(u-s, x)\,\mathrm{d}u \bigg) N^{(i)}(\mathrm{d}s, \mathrm{d}x).\end{aligned}$$

By Fubini's theorem, the last term of the sum can be expressed as

$$\int_0^t \bigg(\int_{(0,u]\times\mathbb{R}_+} (\partial_a + \partial_u) f_u(u-s, x) N^{(i)}(\mathrm{d}s, \mathrm{d}x) \bigg)\,\mathrm{d}u,$$

and, by (6), this is equal to $\int_0^t \langle Z_u, (\partial_a + \partial_u) f_u \rangle\,\mathrm{d}u$. This concludes the proof. □

The decomposition (5) is classical in the field of measure-valued population dynamics; see Tran (2008) and Bensusan *et al.* (2015). The first term refers to the pure jump part of arrivals of individuals with age 0, whereas the second term of transport-type illustrates the ageing phenomenon (all ages are translated along the time axis), as well as the time component. The fact that the drift part depends on both $\langle Z_t^{(i)}, \partial_a f_t \rangle$ and $\langle Z_t^{(i)}, \partial_t f_t \rangle$ is the starting point of our results derived in the next section. Let us remark that as a particular case, taking $\Phi_t(a, x) = \phi(a)$ and $\Psi_t(a, x) = 0$, this shows why the intensity process $\lambda_t = \mu + \langle Z_{t-}^{(2)}, \phi \rangle$ is Markovian in the case where the fertility function is exponential (see Oakes (1975)); that is, $\phi(a) = \alpha \mathrm{e}^{\beta a}$. In this case, $\phi' = \beta\phi$, and (5) leads to the differential form $\mathrm{d}\langle Z_t^{(2)}, \phi \rangle = \alpha\,\mathrm{d}N_t^{(2)} + \beta \langle Z_t^{(2)}, \phi \rangle\,\mathrm{d}t$, where we recall that $N_t^{(2)} = \langle Z_t^{(2)}, \mathbf{1} \rangle$ is nothing but the Hawkes process itself. Note that $\mathrm{d}N_t^{(2)}$ only depends on the past of (λ_t) by means of the current value λ_t, which proves the Markov property.

3. The exponential case generalized

3.1. Assumptions on the fertility rates

In the following, we introduce the assumptions allowing us to recover a finite-dimensional Markovian dynamics.

Assumption 1. (i) *The birth rates* Φ *and* Ψ *are nonnegative and satisfy* $\Phi_t(a, x) = v(t)\phi(a, x)$ *and* $\Psi_t(a, x) = w(t)\psi(a, x)$*, where*

$$\phi^{(n)}(a, x) = c_{-1} + \sum_{k=0}^{n-1} c_k \phi^{(k)}(a, x), \qquad v^{(p)}(t) = d_{-1}(t) + \sum_{l=0}^{p-1} d_l(t) v^{(l)}(t)$$

https://doi.org/10.1017/apr.2016.10 Published online by Cambridge University Press

with $n, p \geq 1$ *and initial conditions* $\phi^{(k)}(0, x) = \phi_0^{(k)}(x)$*, and*

$$\psi^{(m)}(a, x) = r_{-1} + \sum_{k=0}^{m-1} r_k \psi^{(k)}(a, x), \qquad w^{(q)}(t) = k_{-1}(t) + \sum_{l=0}^{q-1} k_l(t) w^{(l)}(t)$$

with $m, q \geq 1$ *and initial conditions* $\psi^{(k)}(0, x) = \psi_0^{(k)}(x)$*. Note that we use the notation* $f^{(k)}(a, x) = \partial_a^k f(a, x)$*.*

(ii) *The maps* $(d_l)_{-1 \leq l \leq p-1}$ *and* $(k_l)_{-1 \leq l \leq q-1}$ *are continuous.*

Assumption 1 defines a wide class of self and externally exciting fertility functions of the form $\Phi_t(a, x) = v(t)\phi(a, x)$. Let us first focus on the time-independent part and introduce $F(a, x)$ such that $F = (1, \phi, \dots, \phi^{(n-1)})^T$. Then $F' = C(c)F$, where the function $C(\cdot)$ which transforms a vector c into a matrix $C(c)$ is defined as

$$C(c) = \begin{pmatrix} 0 & 0 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ c_{-1} & c_0 & \dots & c_{n-2} & c_{n-1} \end{pmatrix}. \tag{7}$$

In particular, if the polynomial $P(y) = y^n - \sum_{k=0}^{n-1} c_k y^k$ is split with distinct roots $y_1, \dots, y_p$ and corresponding multiplicities $n_1, \dots, n_p$, then ϕ can be written up to some constant as $\sum_{i=1}^{p} P_i(x, a) \mathrm{e}^{y_i a}$, where P_i is a polynomial in a with degree at most $n_i - 1$ whose coefficients may depend on x. This includes, for example, the framework of Dassios and Zhao (2011), where $\Phi_t(a, x) = \Psi_t(a, x) = x\mathrm{e}^{-\delta a}$. This is also a sufficiently large set of functions to approximate any fertility function outside of the range of Assumption 1. As an example, the power-law kernel is of importance for many applications. In the context of earthquakes, the Omori law describes the epidemic-type aftershock model: it corresponds to a specific form $\phi(a) \sim K/a^{1+\epsilon}$. Also in the field of financial microstructure, recent studies (see, e.g. Hardiman *et al.* (2013)) found that high-frequency financial activity is better described by a Hawkes process with power-law kernel rather than exponential. The power-law kernel with cut-off can be approximated as in Hardiman *et al.* (2013) up to a constant by the smooth function

$$\phi(a) = \sum_{i=0}^{M-1} \frac{\mathrm{e}^{-a/(\tau_0 m^i)}}{(\tau_0 m^i)^{1+\epsilon}} - S\mathrm{e}^{-a/(\tau_0 m^{-1})},$$

where S is such that $\phi(0) = 0$. In general, one can use approximation theory to construct a sequence of fertility functions which tends to the original one. As a result, this constructs a sequence of Hawkes processes that approximate the original Hawkes process. As we also allow for time-dependency, such birth rates Φ and Ψ that satisfy Assumption 1 are also useful to define nonstationary Hawkes processes, and in particular to include seasonality. As an example, one can simply think of a kernel of the form $\cos^2(\alpha t)\phi(a, x)$, where $v(t) = \cos^2(\alpha t)$ satisfies $v'' = 4\alpha^2(1 - v)$. Note that in Assumption 1, coefficients in the equation for the time-dependent part are allowed to vary with time, therefore a wide variety of time dependence structures can be included in the model.

3.2. Age pyramid dynamics

Let us go back to the dynamics of the age pyramid over time. The key property that will allow us to compute distribution properties is that the population enables us to identify the components to add to the Hawkes process and its intensity to make the dynamics Markovian. This is stated in the following proposition.

Proposition 1. *Let us define for* $-1 \le k \le n-1$ *and* $-1 \le l \le p-1$, $X_t^{k,l} := \langle Z_t^{(2)}, \partial_a^k \partial_t^l \Phi_t \rangle$ *and for* $-1 \le k \le m-1$ *and* $-1 \le l \le q-1$, $Y_t^{k,l} := \langle Z_t^{(1)}, \partial_a^k \partial_t^l \Psi_t \rangle$. *Let us also define the two matrices*

$$M_t^{(2)} = (X_t^{k,l})_{-1 \le k \le n-1, -1 \le l \le p-1}, \qquad M_t^{(1)} = (Y_t^{k,l})_{-1 \le k \le m-1, -1 \le l \le q-1}.$$

(i) *Let us denote by* $D^\top$ *the transpose of a given matrix* D. *The processes* $M^{(1)}$ *and* $M^{(2)}$ *follow the dynamics*

$$\mathrm{d}M_t^{(i)} = \int_{\mathbb{R}_+} W^{(i)}(t,x) N^{(i)}(\mathrm{d}t, \mathrm{d}x) + (C^{(i)} M_t^{(i)} + M_t^{(i)} D_t^{(i)\top}), \tag{8}$$

where

$$\begin{gathered} W_{k,l}^{(1)}(t,x) = w^{(l)}(t) \psi_0^{(k)}(x) \quad \text{for } -1 \le k \le m-1,\ -1 \le l \le q-1, \\ W_{k,l}^{(2)}(t,x) = v^{(l)}(t) \phi_0^{(k)}(x) \quad \text{for } -1 \le k \le n-1,\ -1 \le l \le p-1, \\ C^{(1)} = C(r), \qquad C^{(2)} = C(c), \qquad D_t^{(1)} = C(k(t)), \qquad D_t^{(2)} = C(d(t)), \end{gathered}$$

where $C(\cdot)$ *is defined by (7).*

(ii) *As a consequence of (8),* $(M_t^{(1)}, M_t^{(2)})_{t \ge 0}$ *is a Markov process.*

Proof. We focus on the dynamics of the $X^{k,l}$, the problem being the same for the $Y^{k,l}$. From Lemma 1, for $0 \le k \le n-2$ and $0 \le l \le p-2$,

$$\mathrm{d}X_t^{k,l} = v^{(l)}(t) \int_{\mathbb{R}_+} \phi_0^{(k)}(x) N^{(2)}(\mathrm{d}t, \mathrm{d}x) + (X_t^{k+1,l} + X_t^{k,l+1})\,\mathrm{d}t. \tag{9}$$

From Assumption 1, $X_t^{n,l} = \sum_{k=-1}^{n-1} c_k X_t^{k,l}$ and $X_t^{k,n} = \sum_{l=-1}^{p-1} d_l(t) X_t^{k,l}$. This shows that for $0 \le l \le p-2$ and $0 \le k \le n-2$,

$$\mathrm{d}X_t^{n-1,l} = v^{(l)}(t) \int_{\mathbb{R}_+} \phi_0^{(n-1)}(x) N^{(2)}(\mathrm{d}t, \mathrm{d}x) + \left(\sum_{k=-1}^{n-1} c_k X_t^{k,l} + X_t^{n-1,l+1} \right) \mathrm{d}t, \tag{10}$$

$$\mathrm{d}X_t^{k,p-1} = v^{(p-1)}(t) \int_{\mathbb{R}_+} \phi_0^{(k)}(x) N^{(2)}(\mathrm{d}t, \mathrm{d}x) + \left(X_t^{k+1,p-1} + \sum_{l=-1}^{p-1} d_l(t) X_t^{k,l} \right) \mathrm{d}t, \tag{11}$$

$$\begin{aligned} \mathrm{d}X_t^{n-1,p-1} &= v^{(p-1)}(t) \int_{\mathbb{R}_+} \phi_0^{(n-1)}(x) N^{(2)}(\mathrm{d}t, \mathrm{d}x) \\ &\quad + \left(\sum_{k=-1}^{n-1} c_k X_t^{k,p-1} + \sum_{l=-1}^{p-1} d_l(t) X_t^{n-1,l} \right) \mathrm{d}t. \end{aligned} \tag{12}$$

https://doi.org/10.1017/apr.2016.10 Published online by Cambridge University Press

In addition, from Lemma 1 for $0 \le l \le p-2$ and $0 \le k \le n-2$, we have

$$\mathrm{d}X_t^{-1,l} = v^{(l)}(t)\,\mathrm{d}N_t^{(2)} + X_t^{-1,l+1}\,\mathrm{d}t, \tag{13}$$

$$\mathrm{d}X_t^{k,-1} = \int_{\mathbb{R}_+} \phi_0^{(k)}(x) N^{(2)}(\mathrm{d}t, \mathrm{d}x) + X_t^{k+1,-1}\,\mathrm{d}t. \tag{14}$$

Finally, by Assumption 1 again, we obtain the following two equations:

$$\mathrm{d}X_t^{-1,p-1} = v^{(p-1)}(t)\,\mathrm{d}N_t^{(2)} + \Bigg(\sum_{l=-1}^{p-1} d_l(t) X_t^{-1,l}\Bigg)\mathrm{d}t, \tag{15}$$

$$\mathrm{d}X_t^{n-1,-1} = \int_{\mathbb{R}_+} \phi_0^{(n-1)}(x) N^{(2)}(\mathrm{d}t, \mathrm{d}x) + \Bigg(\sum_{k=-1}^{n-1} c_k X_t^{k,-1}\Bigg)\mathrm{d}t. \tag{16}$$

From (9)–(16), we then deduce (8). The proof of Proposition 1(ii) follows immediately. □

3.3. Laplace transform

Here we highlight an exponential martingale which allows us to compute the Laplace transform of the whole dynamics. This is the main result of our paper. To ensure tractability of the Laplace transform, we also state the following assumption.

Assumption 2. *For each $\lambda > 0$, $\int_{\mathbb{R}_+} \exp(\lambda \max_{0\le k\le n-1} \phi_0^{(k)}(x)) G(x)\,\mathrm{d}x < +\infty$.*

Our main result is stated below. Note that the trace of the matrix $u^\top M$ given by $\mathrm{tr}(u^\top M) = \sum_{k,l} u_{k,l} M_{k,l}$ computes a linear combination of the components of a given matrix M, and recall that $u^\top$ denotes the transposition of the matrix u.

Theorem 1. *Denote by $\mathcal{F}^M$ the filtration of $(M^{(1)}, M^{(2)})$. Under Assumption 1, we have the following.*

(i) *For any deterministic and differentiable matrix-valued $(A_t^{(1)})$ and $(A_t^{(2)})$ with derivatives $(\mathring{A}_t^{(1)})$ and $(\mathring{A}_t^{(2)})$, the following process is an $\mathcal{F}^M$-martingale:*

$$\begin{aligned}\exp\Bigg\{&\sum_{i=1}^{2} \mathrm{tr}(A_t^{(i)} M_t^{(i)}) - \int_0^t \mathrm{tr}(A_s^{(i)} C^{(i)} M_s^{(i)} + A_s^{(i)} M_s^{(i)} D_s^{(i)\top} + \mathring{A}_s^{(i)} M_s^{(i)})\,\mathrm{d}s \\ &- \int_0^t \int_{\mathbb{R}_+} (\mathrm{e}^{\mathrm{tr}(A_s^{(1)} W^{(1)}(s,x))} - 1)\rho(s) H(x)\,\mathrm{d}x\,\mathrm{d}s \\ &- \int_0^t \int_{\mathbb{R}_+} (\mathrm{e}^{\mathrm{tr}(A_s^{(2)} W^{(2)}(s,x))} - 1)(\mu(s) + M_s^{(1)}[0,0] + M_s^{(2)}[0,0]) G(x)\,\mathrm{d}x\,\mathrm{d}s\Bigg\}.\end{aligned} \tag{17}$$

(ii) *For each of the matrices u and v with dimensions $(n+1)(p+1)$ and $(m+1)(q+1)$, respectively, the joint Laplace transform can be expressed as*

$$\begin{aligned}&\mathbb{E}[\exp(\mathrm{tr}(u^\top M_t^{(1)} + v^\top M_t^{(2)}))] \\ &\quad = \exp\Bigg\{\int_0^t \int_{\mathbb{R}_+} (\mathrm{e}^{\mathrm{tr}(A_s^{(1)} W^{(1)}(s,x))} - 1)\rho(s) H(x)\,\mathrm{d}x\,\mathrm{d}s \\ &\qquad\qquad + \int_0^t \int_{\mathbb{R}_+} (\mathrm{e}^{\mathrm{tr}(A_s^{(2)} W^{(2)}(s,x))} - 1)\mu(s) G(x)\,\mathrm{d}x\,\mathrm{d}s\Bigg\},\end{aligned}$$

where, for $i \in \{1, 2\}$,

$$\mathring{A}_t^{(i)} + A_t^{(i)} C^{(i)} + D_t^{(i)\top} A_t^{(i)} = \bigg\{ \int_{\mathbb{R}_+} (1 - \mathrm{e}^{\mathrm{tr}(A_t^{(2)} W^{(2)}(t,x))}) G(x)\,\mathrm{d}x \bigg\} K, \tag{18}$$

with terminal conditions

$$A_T^{(1)} = u^\top, \qquad A_T^{(2)} = v^\top, \tag{19}$$

where the matrices K and J are given by

$$K = J^\top J \quad \text{and} \quad J = (0, 1, 0, \dots, 0), \tag{20}$$

respectively. Moreover a solution to (18) and (19) of class $\mathcal{C}^1$ exists and is unique provided that Assumption 2 is satisfied.

Proof. We begin by highlighting the exponential martingale (17). Let us denote

$$\langle N^{(i)}, H \rangle_t = \int_0^t \int_{\mathbb{R}_+} H(s, x) N^{(i)}(\mathrm{d}s,\ \mathrm{d}x).$$

For deterministic $\alpha(t, x)$ and $\beta(t, x)$, then by the classical exponential formula the following process is a martingale:

$$\exp\bigg\{ \langle N^{(1)}, \alpha \rangle_t + \langle N^{(2)}, \beta \rangle_t - \int_0^t \int_{\mathbb{R}_+} (\mathrm{e}^{\alpha(s,x)} - 1) \rho(s) H(x)\,\mathrm{d}x\,\mathrm{d}s \\ - \int_0^t \int_{\mathbb{R}_+} (\mathrm{e}^{\beta(s,x)} - 1)(\mu(s) + \langle Z_{s-}^{(1)}, \Psi_s \rangle + \langle Z_{s-}^{(2)}, \Phi_s \rangle) G(x)\,\mathrm{d}x\,\mathrm{d}s \bigg\}. \tag{21}$$

The aim now is to compute the joint Laplace transform of the processes $M_t^{(1)}$ and $M_t^{(2)}$. This amounts to computing $\mathbb{E}[\mathrm{e}^{\mathrm{tr}(u^\top \cdot M_t^{(1)} + v^\top \cdot M_t^{(2)})}]$, since $\mathrm{tr}(u^\top M) = \sum_{k,l} u_{k,l} M_{k,l}$. Let us consider the two (deterministic) processes $A_t^{(1)}$ and $A_t^{(2)}$ with sizes $(m+1)(q+1)$ and $(n+1)(p+1)$, respectively. By integration by parts, $\mathrm{d}(A_t^{(i)} M_t^{(i)}) = A_t^{(i)}\,\mathrm{d}M_t^{(i)} + \mathring{A}_t^{(i)} M_t^{(i)}\,\mathrm{d}t$. From (8), we obtain the dynamics

$$\mathrm{d}\,\mathrm{tr}(A_t^{(i)} M_t^{(i)}) = \int_{\mathbb{R}_+} \mathrm{tr}(A_t^{(i)} W^{(i)}(t, x)) N^{(i)}(\mathrm{d}t, \mathrm{d}x) \\ + \mathrm{tr}(C^{(i)} M_t^{(i)} + M_t^{(i)} D_t^{(i)\top} + \mathring{A}_t^{(i)} M_t^{(i)})\,\mathrm{d}t.$$

Let us now use (21) with $\alpha(t, x) = \mathrm{tr}(A_t^{(1)} W^{(1)}(t, x))$ and $\beta(t, x) = \mathrm{tr}(A_t^{(2)} W^{(2)}(t, x))$ to obtain (17). To obtain the Laplace transform, it remains to make the random part of the integrand in (17) vanish. To do this, let us first identify the term in $M^{(1)}$ to find the linear equation (18) for $i = 1$. In addition, the term in $M^{(2)}$ leads to (18) for $i = 2$. If we set terminal conditions (19), we obtain the Laplace transform in Theorem 1(ii) by the martingale property of (17). To conclude on the existence and uniqueness, we use the Cauchy–Lipschitz theorem. To show that a solution of class $\mathcal{C}^1$ to (18) exists and is unique, it is sufficient to prove that the map $(Y, t) \mapsto \int_{\mathbb{R}_+} \mathrm{e}^{\mathrm{tr}(Y W^{(2)}(t,x))} G(x)\,\mathrm{d}x$ is of class $\mathcal{C}^1$. Since the integrand is $\mathcal{C}^1$ by Assumptions 1(i) and 1(ii), it is sufficient to prove that its gradient, given by

$$(\mathrm{e}^{\mathrm{tr}(Y W^{(2)}(t,x))} W^{(2)}(t, x)^T, \mathrm{tr}(Y \partial_t W^{(2)}(t, x)) \mathrm{e}^{\mathrm{tr}(Y W^{(2)}(t,x))}), \tag{22}$$

is locally bounded by some quantity that is independent of Y and t, and is integrable with respect to G. Let us use some localization argument, and define the set

$$B(0, r) = \{A \operatorname{Re}(n+1) \times (p+1) \text{ matrix such that} \|A\|_\infty \leq r\},$$

where $r > 0$ and $\|A\|_\infty = \max_{-1 \leq i \leq n-1} \sum_{j=-1}^{p-1} |A_{i,j}|$. Now, for $(Y, t) \in B(0, r) \times [0, T]$, we obtain

$$\begin{aligned}
\exp(\operatorname{tr}(Y W^{(2)}(t, x))) &\leq \exp\biggl(\sum_{i=-1}^{n-1} \sum_{k=-1}^{p-1} |Y_{i,k}||W_{k,i}^{(2)}(t, x)|\biggr) \\
&\leq \exp\biggl((n+1) \max_{-1 \leq i \leq n-1} \sum_{k=-1}^{p-1} |Y_{i,k}||W_{k,i}^{(2)}(t, x)|\biggr) \\
&\leq \exp\Bigl(r(n+1) \max_{-1 \leq l \leq p-1} \sup_{t \in [0,T]} |v^{(l)}(t)| \max_{-1 \leq k \leq n-1} |\phi_0^{(k)}(x)|\Bigr),
\end{aligned}$$

where the last inequality uses the fact that $Y \in B(0, r)$. As for the first component of (22),

$$|W_{k,l}^{(2)}(t, x)| \leq |\phi_0^{(k)}(x)| \sup_{t \in [0,T]} |v^{(l)}(t)|,$$

and for the second component, we have (since $Y \in B(0, r)$)

$$|\operatorname{tr}(Y \partial_t W^{(2)}(t, x))| \leq r(n+1) \max_{0 \leq l \leq p} \sup_{t \in [0,T]} |v^{(l)}(t)| \max_{-1 \leq k \leq n-1} |\phi_0^{(k)}(x)|.$$

This concludes the proof by the use of Assumptions 1 and 2. $\square$

4. Pathwise representation of the Hawkes population

Definition 1 uses a classical formulation to define a counting process with its own intensity. However, it does not keep track of the branching population and also does not give a concrete pathwise representation. Also, the definition in terms of an immigration–birth process (see Section 2) is intuitive and gives more information through the age pyramid. The aim of this section is to discuss the pathwise representation of the age pyramid process with its own intensity by means of Poisson point measures. This approach allows us both to keep track of the age pyramid (branching population) and to represent it as a process with its own intensity in a pathwise way. In this way, we are able to reconcile the two standard definitions of the Hawkes process, through a counting process or a branching dynamics. Let us describe the thinning construction of a general random point measure on $\mathbb{R}_+ \times E$, say $\Gamma(\mathrm{d}s, \mathrm{d}y) = \sum_{n \geq 1} \delta_{(T_n, Y_n)}(\mathrm{d}s, \mathrm{d}y)$, where $(E, \mathcal{E})$ is some measurable space. Assume that its intensity measure $\gamma(\mathrm{d}s, \mathrm{d}y)$ admits a density: $\gamma(\mathrm{d}s, \mathrm{d}y) = \gamma(s, y)\, \mathrm{d}s \mu(\mathrm{d}y)$. In this model, events occur with intensity $s \mapsto \int_{x \in E} \gamma(s, x) \mu(\mathrm{d}x)$, and if a birth occurs at time T_n, then the characteristics Y_n of the newborn are drawn with distribution $\gamma(T_n, y)\mu(\mathrm{d}y)/\int_{x \in E} \gamma(T_n, x)\mu(\mathrm{d}x)$. Let $Q(\mathrm{d}s, \mathrm{d}y, \mathrm{d}\theta)$ be a Poisson point measure on $\mathbb{R}_+ \times E \times \mathbb{R}_+$ with intensity measure $\mathrm{d}s \mu(\mathrm{d}y)\, \mathrm{d}\theta$; see, e.g. Çınlar (2011) for a definition. Denote by $(\mathcal{F}_t^Q)$ the canonical filtration generated by Q, and introduce $P(\mathcal{F}_t^Q)$ the predictable σ-field associated with $\mathcal{F}_t^Q$. We further assume that $\gamma(t, y)$ is $P(\mathcal{F}_t^Q) \times \mathcal{E}$-measurable and also that $\int_0^t \int_E \gamma(s, y)\, \mathrm{d}s \mu(\mathrm{d}y) < +\infty$ almost surely. Now, define $\Gamma(\mathrm{d}s, \mathrm{d}y) = \int_{\mathbb{R}_+} \mathbf{1}_{[0, \gamma(s,y)]}(\theta) Q(\mathrm{d}s, \mathrm{d}y, \mathrm{d}\theta)$. This clearly defines a point

https://doi.org/10.1017/apr.2016.10 Published online by Cambridge University Press

measure and the martingale property of Q ensures that the random point measure $\Gamma(\mathrm{d}s, \mathrm{d}y)$ has intensity measure $\gamma(s, y)\,\mathrm{d}s\mu(\mathrm{d}y)$. Such a construction can be found in Massoulié (1998); we refer the reader to this paper for more details.

We are now ready to construct the age pyramid processes of Definition 2. Let us introduce two independent Poisson point measures $Q^{(1)}(\mathrm{d}t, \mathrm{d}x, \mathrm{d}\theta)$ and $Q^{(2)}(\mathrm{d}t, \mathrm{d}x, \mathrm{d}\theta)$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ (enlarged if necessary) with same intensity measure $\mathrm{d}s\,\mathrm{d}x\,\mathrm{d}\theta$ on $\mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+$. The construction of the first population is immediate since its intensity does not depend on it. Let us define

$$Z_t^{(1)}(\mathrm{d}a, \mathrm{d}x) = \int_{(0,t]} \int_{\mathbb{R}_+ \times \mathbb{R}_+} \mathbf{1}_{[0, \rho(s) H(x)]}(\theta) \delta_{(t-s,x)}(\mathrm{d}a, \mathrm{d}x) Q^{(1)}(\mathrm{d}s, \mathrm{d}x, \mathrm{d}\theta).$$

As for the second population (whose size is the Hawkes process), the intensity is given as a particular form of the process itself; see (4). Therefore, the idea is to define the population underlying the Hawkes process as the solution to the following stochastic equation, often called a *thinning problem*:

$$\begin{aligned} Z_t^{(2)}(\mathrm{d}a, \mathrm{d}x) = \int_{(0,t]} \int_{\mathbb{R}_+ \times \mathbb{R}_+} & \mathbf{1}_{[0, (\mu(s) + \langle Z_{s-}^{(2)}, \Phi_s \rangle + \langle Z_{s-}^{(1)}, \Psi_s \rangle) G(x)]} \\ & \times (\theta) \delta_{(t-s,x)}(\mathrm{d}a, \mathrm{d}x) Q^{(2)}(\mathrm{d}s, \mathrm{d}x, \mathrm{d}\theta). \end{aligned}$$

Such representations are used in the field of stochastic population dynamics for populations with ages and/or characteristics; see, in particular, Fournier and Méléard (2004), Tran (2008), and Bensusan *et al.* (2015). This formulation establishes the link between the Hawkes process and the field of stochastic population dynamics. To further investigate this link seems to be a promising direction for future research.

Remark 1. General results about existence and uniqueness for the Hawkes process (even nonlinear) as the solution of a thinning problem can be found in Brémaud and Massoulié (1996) and Massoulié (1998) (see also Delattre *et al.* (2014), and the books of Daley and Vere-Jones (2008) and Çınlar (2011)). The thinning method to represent a counting process as the solution of a stochastic equation is in fact classical. This general mathematical representation goes back to Kerstan (1964) and Grigelionis (1971). One often refers to the thinning algorithms that have been proposed by Lewis and Shedler (1979) and Ogata (1981), which are very useful in order to perform numerical simulations for quite complex intensity processes.

5. The special case of standard Hawkes process

5.1. Assumption and dynamics

This section focuses on the special case of the standard Hawkes process (N_t) with intensity process $\lambda_t = \mu + \int_{(0,t)} \phi(t-s)\,\mathrm{d}N_s$. Let us denote by $Z_t(\mathrm{d}a) = \sum_{T_n < t} \delta_{t-T_n}(\mathrm{d}a)$ the associated age pyramid (see Definition 2). Below we express Assumption 1 in this context, as well as the associated dynamics (see Proposition 1).

Assumption 3. *The map $a \in \mathbb{R}_+ \mapsto \phi(a)$ is nonnegative, of class $\mathcal{C}^n(\mathbb{R}_+)$, and there exists $c = (c_{-1}, \ldots, c_{n-1}) \in \mathbb{R}^{n+1}$ such that ϕ satisfies $\phi^{(n)} = c_{-1} + \sum_{k=0}^{n-1} c_k \phi^{(k)}$ with initial conditions $\phi^{(k)}(0) = m_k$ for $0 \le k \le n-1$.*

https://doi.org/10.1017/apr.2016.10 Published online by Cambridge University Press

Proposition 2. *Under Assumption 3, the process* $X_t = (\langle Z_t, 1\rangle, \langle Z_t, \phi\rangle, \dots, \langle Z_t, \phi^{(n-1)}\rangle)^T$ *satisfies the dynamics*

$$X_t = N_t m + \int_0^t C X_s \, \mathrm{d}s, \tag{23}$$

where

$$m = (1, m_0, \dots, m_{n-1})^T, \tag{24}$$

and the matrix $C = C(c)$ *is given in (7). In particular, X is a Markov process.*

5.2. Laplace transform

We express below the Laplace transform of the standard Hawkes process, both in the direct form of Theorem 1 and also in terms of a single function.

Proposition 3. *Let us work under Assumption 3.*

(i) *For any* $(n+1)$ *real vector* v,

$$\mathbb{E}[\exp(v \cdot X_T)] = \exp\biggl(-\mu \int_0^T (1 - \mathrm{e}^{A_s \cdot m}) \, \mathrm{d}s\biggr), \tag{25}$$

where the vector map A satisfies the following nonlinear differential equation:

$$C^\top A_t + A_t' + (\mathrm{e}^{A_t \cdot m} - 1) J = 0 \tag{26}$$

with terminal condition $A_T = v$. *Here,* $v \cdot X_T$ *denotes the scalar product between* v *and* X_T, $C^\top$ *is the transpose of the matrix C, and J is defined in (20).*

(ii) *The Laplace transform of the Hawkes process is given for each real* θ *by*

$$\mathbb{E}[\exp(\theta N_T)] = \exp\biggl\{-\mu\biggl((-1)^n G^{(n)}(0) + \sum_{k=0}^{n-1}(-1)^{k+1} c_k G^{(k)}(0)\biggr)\biggr\},$$

where G satisfies the nonlinear differential equation: for each $0 \le t \le T$,

$$(-1)^{n-1} G^{(n+1)}(t) + \sum_{k=0}^{n-1}(-1)^k c_k G^{(k+1)}(t) + \exp\biggl(\theta - c_{-1} G(t) + \sum_{k=0}^{n-1} b_k G^{(k+1)}(t)\biggr) - 1 = 0$$

with terminal conditions $G^{(k)}(T) = 0$ *for* $0 \le k \le n$, *and for* $0 \le k \le n-1$, $b_k = (-1)^k (m_{n-1-k} - \sum_{l=k+1}^{n-1} m_{n-1-l} c_{n-l+k})$.

Numerical example. Before giving the proof of Proposition 3, we illustrate it numerically for the computation of the generating functional $\mathbb{E}[u^{N_T}]$ (the survival probability at time T of a system which survives with probability u at each shock) as well as quantities as $\mathbb{P}(N_T = k) = (1/k!)\partial_u^k \mathbb{E}[u^{N_T}]|_{u=0}$ (the probability to obtain exactly k shocks until time T). Setting $u = \mathrm{e}^\theta$, an explicit discretization scheme has been used to solve the nonlinear differential equation satisfied by G, and the differentiation step for the derivatives of the generating functional has been chosen carefully. The results for the two critical cases $\phi(a) = \mathrm{e}^{-a}$ (case 1) and $\phi(a) = a\mathrm{e}^{-a}$ (case 2)

https://doi.org/10.1017/apr.2016.10 Published online by Cambridge University Press

are described in Tables 1 and 2 to three significative figures. Note that even if the mean number of children per individual is 1 in each case, the results are different due to the shape of each birth rate ϕ. This promotes the use of many kernels, beyond the exponential case. To conclude this numerical experiment, we emphasize that the computation of $\mathbb{P}(N_t = k)$ for higher values of k will require more stable numerical differentiation methods, and are therefore beyond the scope of this paper.

Proof of Proposition 3. We prove only (ii), the proof of (i) being a direct adaptation of that of Theorem 1. Let us denote $A_t = (A_{-1}(t), \dots, A_{n-1}(t))$ and identify the terms in (26), leading to

$$c_{-1}A_{n-1}(t) + A'_{-1}(t) = 0, \tag{27}$$

$$A'_0(t) + c_0 A_{n-1}(t) + \mathrm{e}^{A_t \cdot m} - 1 = 0. \tag{28}$$

As for $1 \le k \le n-1$, we have $A_{k-1}(t) + c_k A_{n-1}(t) + A'_k(t) = 0$, whose recursive computation provides, for $0 \le k \le n-1$,

$$A_k(t) = (-1)^{n-1-k} A_{n-1}^{(n-1-k)}(t) + \sum_{l=1}^{n-1-k} (-1)^l c_{k+l} A_{n-1}^{(l-1)}(t). \tag{29}$$

We deduce that

$$A'_0(t) = (-1)^{n-1} A_{n-1}^{(n)}(t) + \sum_{k=1}^{n-1} (-1)^k c_k A_{n-1}^{(k)}(t). \tag{30}$$

Let us introduce the function $G(t) = \int_T^t A_{n-1}(s)\,\mathrm{d}s$ and choose $A_{-1}(t) = \theta - c_{-1}G(t)$ that satisfies (27). Now, substitute (29) and (30) into (28) to obtain the following nonlinear ordinary differential equation for G:

$$\begin{aligned}
&(-1)^{n-1} G^{(n+1)}(t) + \sum_{k=0}^{n-1} (-1)^k c_k G^{(k+1)}(t) \\
&\quad + \exp\biggl(\theta - c_{-1}G(t) + m_{n-1}G'(t) \\
&\qquad + \sum_{k=0}^{n-2} m_k \biggl[(-1)^{n-1-k} G^{(n-k)}(t) + \sum_{l=1}^{n-1-k} (-1)^l c_{k+l} G^{(l)}(t)\biggr]\biggr) - 1 \\
&\quad = 0.
\end{aligned} \tag{31}$$

Let us simplify the sum in the exponential. By changing variable k into $n-1-k$, it can be expressed as

$$\sum_{k=1}^{n-1} m_{n-1-k}(-1)^k G^{(k+1)}(t) + \sum_{k=1}^{n-1}\sum_{l=1}^{k} (-1)^l m_{n-1-k} c_{n-1-k+l} G^{(l)}(t).$$

Then exchanging the sums leads to

$$\sum_{k=1}^{n-1} m_{n-1-k}(-1)^k G^{(k+1)}(t) + \sum_{l=1}^{n-1} (-1)^l \biggl(\sum_{k=l}^{n-1} m_{n-1-k} c_{n-1-k+l}\biggr) G^{(l)}(t).$$

TABLE 1: Computed values of $\mathbb{E}[u^{N_T}]$ with $\mu = 0.15$ and $T = 5$.

u	0.1	0.3	0.5	0.7	0.9
Case 1: $\phi(a) = \mathrm{e}^{-a}$	0.490	0.532	0.588	0.672	0.828
Case 2: $\phi(a) = a\mathrm{e}^{-a}$	0.494	0.546	0.615	0.714	0.874

TABLE 2: Computed values of $\mathbb{P}(N_T = k)$ with $\mu = 0.15$ and $T = 5$.

k	0	1	2	3	4
Case 1: $\phi(a) = \mathrm{e}^{-a}$	0.472	0.165	0.0894	0.0577	0.0407
Case 2: $\phi(a) = a\mathrm{e}^{-a}$	0.472	0.203	0.1130	0.0700	0.0451

Finally, by setting $l \leftarrow l + 1$ and exchanging notations k and l, (31) can be written as

$$(-1)^{n-1}G^{(n+1)}(t) + \sum_{k=0}^{n-1}(-1)^k c_k G^{(k+1)}(t) + \exp\biggl(\theta - c_{-1}G(t) + \sum_{k=0}^{n-1} b_k G^{(k+1)}(t)\biggr) - 1 = 0,$$

where, for $0 \le k \le n-1$, $b_k = (-1)^k(m_{n-1-k} - \sum_{l=k+1}^{n-1} m_{n-1-l}c_{n-l+k})$.

Now, let us use (25) with (28) to obtain

$$\begin{aligned}\mathbb{E}[\exp(v \cdot X_T)] &= \exp\biggl(-\mu \int_0^T (A_0^{'}(t) + c_0 A_{n-1}(t)\biggr)\,\mathrm{d}t), \\ &= \exp\biggl(-\mu\biggl((-1)^{n-1}(G^{(n)}(T) - G^{(n)}(0)) + \sum_{k=0}^{n-1}(-1)^k c_k (G^{(k)}(T) - G^{(k)}(0))\biggr)\biggr),\end{aligned}$$

where the last equality comes from (30). Let us set, for $0 \le k \le n-1$, $A_k(T) = 0$. One can show by (29) that the previous conditions are equivalent to the terminal values $G^{(k)}(T) = 0$ for $1 \le k \le n-1$. Note that by definition of G, we also have $G(T) = 0$. □

5.3. Moments

On the particular case of the standard Hawkes process, we illustrate how to compute first- and second-order moments explicitly.

First-order moments. The differential system of (23) is linear and allows us to propose a straightforward differential equation for the first-order moments. We also perform explicit computations for small dimensions $n = 1$ and $n = 2$.

Proposition 4. *Under Assumption 3, the vector map $u(t) := \mathbb{E}[X_t]$ is a solution to*

$$u'(t) = \mu m + Au(t), \tag{32}$$

where the $(n + 1) \times (n + 1)$ matrix A is given by

$$A = C + mJ, \tag{33}$$

where C, m, and J are given in (7), (24), and (20), respectively.

https://doi.org/10.1017/apr.2016.10 Published online by Cambridge University Press

Proof. Let us use the martingale property of the compensated counting process to obtain $\mathbb{E}[N_t] = \int_0^t (\mu + \mathbb{E}[\langle Z_s, \phi\rangle])\,\mathrm{d}s$. Now, let us take expectation in (23) and use the previous equation to obtain (32). This concludes the proof. □

Equation (32) allows us to obtain explicit equations for the expected number of events. We derive such results for the popular exponential case $\phi(a) = \mathrm{e}^{-ca}$ (see also Dassios and Zhao (2011)) and also for the birth rate $\phi(a) = \alpha^2 a\mathrm{e}^{-\beta a}$. This case can be useful for a variety of applications to model a smooth delay at excitation. We remark on the different behavior of the first moment, in particular in the critical case. For the two examples given below, the computations are left to the reader.

Corollary 1. *For the Hawkes process with* $\phi(a) = \mathrm{e}^{-ca}$,

$$\mathbb{E}[N_t] = \begin{cases} \mu\Big(t + \dfrac{t^2}{2}\Big) & \text{if } c = 1,\\ \dfrac{\mu}{1-c}\Big(\dfrac{\mathrm{e}^{(1-c)t} - 1}{1-c} - ct\Big) & \text{if } c \neq 1. \end{cases}$$

Corollary 2. *For the Hawkes process with* $\phi(a) = \alpha^2 a\mathrm{e}^{-\beta a}$,

$$\mathbb{E}[N_t] = \begin{cases} \dfrac{\mu}{8\beta}(1 - \mathrm{e}^{-2\beta t}) + \dfrac{3\mu}{4}t + \dfrac{\beta\mu}{4}t^2 & \text{if } \alpha = \beta,\\ \dfrac{\mu\beta^2}{\beta^2 - \alpha^2}t + \dfrac{\alpha\mu}{2}\Big(\dfrac{\mathrm{e}^{(\alpha-\beta)t} - 1}{(\alpha-\beta)^2} - \dfrac{\mathrm{e}^{-(\alpha+\beta)t} - 1}{(\alpha+\beta)^2}\Big) & \text{if } \alpha \neq \beta. \end{cases}$$

Second-order moments. We now derive the dynamics of the matrix $V_t := X_t X_t^\top$ with $X_t = (N_t, \langle Z_t, \phi\rangle, \dots, \langle Z_t, \phi^{(n-1)}\rangle)^T$. As a consequence, we represent the variance-covariance matrix of the process (X_t) as the solution to a linear ordinary differential equation. Our method is based on differential calculus with the finite variation process (X_t) with dynamics (23) and could be extended to higher moments.

Proposition 5. *Let us introduce the matrix* $V_t := X_t X_t^\top$, *where* $X_t^\top$ *denotes the transpose of* X_t. *Then the matrix* V_t *satisfies the dynamics*

$$\mathrm{d}V_t = \mathrm{d}N_t(X_{t-}m^\top + mX_{t-}^\top + mm^\top) + \mathrm{d}t(V_t C^\top + CV_t).$$

In particular, the variance-covariance matrix $v(t) := \mathbb{E}[V_t]$ *satisfies*

$$v'(t) = v(t)A^\top + Av(t) + \mu(mm^\top + u(t)m^\top + mu^\top(t)) + Ju(t)mm^\top, \tag{34}$$

where $u(t)$ *is a solution to (32) and the matrix* A *is defined in (33).*

Proof. Denote $X_t = (X_t^{[-1]}, X_t^{[0]}, \dots, X_t^{[n-1]})$. Integration by parts leads to, for $-1 \le l, k \le n-1$, $\mathrm{d}(X_t^{[k]}X_t^{[l]}) = X_{t-}^{[k]}\mathrm{d}X_t^{[l]} + X_{t-}^{[l]}\mathrm{d}X_t^{[k]} + m_k m_l \mathrm{d}N_t$. The previous equation shows that $\mathrm{d}V_t = X_{t-}\mathrm{d}X_t^\top + (\mathrm{d}X_t)X_{t-}^\top + \mathrm{d}N_t \cdot mm^\top$. By Proposition 2, we obtain $\mathrm{d}V_t = \mathrm{d}N_t(X_{t-}m^\top + mX_{t-}^\top + mm^\top) + \mathrm{d}t(V_t C^\top + CV_t)$. Now, take expectation in the previous equation to obtain

$$v'(t) = \mathbb{E}[(\mu + X_t^{[0]})X_t]m^\top + m\mathbb{E}[(\mu + X_t^{[0]})X_t^\top] + (\mu + \mathbb{E}[X_t^{[0]}])mm^\top + v(t)C^\top + Cv(t).$$

Finally since $X_t^{[0]}X_t = V_t J^\top$, see (20), the previous equation reduces to (34). □

We give explicit equations for $\phi(a) = \mathrm{e}^{-ca}$ and at a higher order for the critical case $\phi(a) = \beta^2 a \mathrm{e}^{-\beta a}$. Computations are based on (34) and are left to the reader.

Corollary 3. *For the Hawkes process with* $\phi(a) = \mathrm{e}^{-ca}$,

$$\operatorname{var}(N_t) = \begin{cases} \mu t\left(1 + \frac{3}{2}t + \frac{2}{3}t^2 + \frac{1}{12}t^3\right) & \text{if } c = 1, \\ \frac{\mu}{(1-c)^3}\left[\frac{1-c/2}{1-c}\mathrm{e}^{2(1-c)t} + \left(\frac{3c^2-1}{1-c} - 2ct\right)\mathrm{e}^{(1-c)t} - c^3 t \right. & \\ \qquad \left. + \frac{c(1/2-3c)}{1-c}\right] & \text{if } c \neq 1. \end{cases}$$

Corollary 4. *For the Hawkes process with* $\phi(a) = \beta^2 a \mathrm{e}^{-\beta a}$,

$$\operatorname{var}(\lambda_t) = \beta\mu\left(-\frac{7}{128} + \frac{3\beta}{32}t + \frac{\beta^2}{16}t^2 + \frac{1-\beta t}{8}\mathrm{e}^{-2\beta t} - \frac{9}{128}\mathrm{e}^{-4\beta t}\right).$$

6. Conclusion

We introduced the concept of age pyramid for Hawkes processes with general immigrants. The virtue of this approach is to keep track of all past events. This allows tractable computations for the Hawkes process with general immigrants whose fertility functions are time dependent generalizations of the popular exponential case, providing natural extensions of the existing results in this direction. In addition, we illustrated the pathwise construction of the Hawkes dynamics and its underlying population process. On the whole, our approach seems to reconcile two definitions of Hawkes processes, through an intensity process or a branching dynamics. This framework appears to be a promising direction for further research. As an example, the large population asymptotics in the field of measure-valued population dynamics could give further insights on the macroscopic behavior of Hawkes processes.

Acknowledgements

The author is grateful to his supervisor Nicole El Karoui for her help to improve the results and the whole paper. The author thanks Mathieu Rosenbaum, Thibault Jaisson, and Monique Jeanblanc for fruitful discussions and enlightening comments, as well as the anonymous referee for valuable advice.

References

Adamopoulos, L. (1975). Some counting and interval properties of the mutually-exciting processes. *J. Appl. Prob.* **12,** 78–86.

Aït-Sahalia, Y., Cacho-Diaz, J. and Laeven, R. J. A. (2010). Modeling financial contagion using mutually exciting jump processes. Working Paper 15850, National Bureau of Economic Research.

Bensusan, H., Boumezoued, A., El Karoui, N. and Loisel, S. (2015). Bridging the gap from microsimulation practice to population models: a survey. (Preliminary version as Chapter 2 of A. Boumezoued Doctoral thesis.) Available at https://tel.archives-ouvertes.fr/tel-01307921/document

Brémaud, P. and Massoulié, L. (1996). Stability of nonlinear Hawkes processes. *Ann. Prob.* **24,** 1563–1588.

Brémaud, P. and Massoulié, L. (2002). Power spectra of general shot noises and Hawkes point processes with a random excitation. *Adv. Appl. Prob.* **34,** 205–222.

Çinlar, E. (2011). *Probability and Stochastics* (Graduate Texts Math. **261**). Springer, New York.

Da Fonseca, J. and Zaatour, R. (2014). Hawkes process: fast calibration, application to trade clustering, and diffusive limit. *J. Futures Markets* **34,** 548–579.

Daley, D. J. and Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes*, Vol. II, *General Theory and Structure*, 2nd edn. Springer, New York.

Dassios, A. and Zhao, H. (2011). A dynamic contagion process. *Adv. Appl. Prob.* **43,** 814–846.

Delattre, S., Fournier, N. and Hoffmann, M. (2014). High dimensional Hawkes processes. Preprint. Available at http://arxiv.org/abs/1403.5764.

Errais, E., Giesecke, K. and Goldberg, L. (2010). Affine point processes and portfolio credit risk. *SIAM J. Financial Math.* **1,** 642–665.

Fournier, N. and Méléard, S. (2004). A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *Ann. Appl. Prob.* **14,** 1880–1919.

Grigelionis, B. (1971). The representation of integer-valued random measures as stochastic integrals over the Poisson measure. *Litovsk. Mat. Sb.* **11,** 93–108.

Hardiman, S. J., Bercot, N. and Bouchaud, J.-P. (2013). Critical reflexivity in financial markets: a Hawkes process analysis. Preprint. Available at http://arxiv.org/abs/1302.1405.

Harris, T. E. (1963). *The Theory of Branching Processes*. Springer, Berlin.

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

Hawkes, A. G. and Oakes, D. (1974). A cluster process representation of a self-exciting process. *J. Appl. Prob.* **11,** 493–503.

Jovanović, S., Hertz, J. and Rotter, S. (2014). Cumulants of Hawkes point processes. Preprint. Available at http://arxiv.org/abs/1409.5353.

Kerstan, J. (1964). Teilprozesse Poissonscher prozesse. In *Trans. Third Prague Conf. Information Theory, Statist. Decision Functions, Random Processes* (Liblice, 1962), Publishing House of the Czech Academy of Sciences, Prague, pp. 377–403.

Lewis, P. A. and Shedler, W. G. S. (1979). Simulation of nonhomogeneous Poisson processes by thinning. *Naval Res. Logistics Quart.* **26,** 403–413.

Massoulié, L. (1998). Stability results for a general class of interacting point processes dynamics, and applications. *Stoch. Process. Appl.* **75,** 1–30.

Oakes, D. (1975). The Markovian self-exciting process. *J. Appl. Prob.* **12,** 69–77.

Ogata, Y. (1981). On Lewis' simulation method for point processes. *IEEE Trans. Inf. Theory* **27,** 23–31.

Rambaldi, M., Pennesi, P. and Lillo, F. (2014). Modeling FX market activity around macroeconomic news: a Hawkes process approach. Preprint. Available at http://arxiv.org/abs/1405.6047.

Saichev, A. I. and Sornette, D. (2011). Generating functions and stability study of multivariate self-excited epidemic processes. *Europ. Physical J. B* **83,** 271–282.

Tran, V. C. (2006). Modèles particulaires stochastiques pour des problèmes d'évolution adaptative et pour l'approximation de solutions statistiques. Doctoral Thesis, Université Paris X – Nanterre. Available at http://tel.archives-ouvertes.fr/tel-00125100.

Tran, V. C. (2008). Large population limit and time behaviour of a stochastic particle model describing an age-structured population. *ESAIM Prob. Statist.* **12,** 345–386.

Wheatley, S., Filimonov, V. and Sornette, D. (2014). Estimation of the Hawkes process with renewal immigration using the EM algorithm. Preprint. Available at http://arxiv.org/abs/1407.7118.