A Palm space approach to non-linear Hawkes processes

Philippe Robert[*]   Gaëtan Vignoud[†]

Abstract

A Hawkes process on $\mathbb{R}$ is a point process whose intensity function at time t is a functional of its past activity before time t. It is defined by its activation function Φ and its memory function h. In this paper, the Hawkes property is expressed as an operator on the sub-space of non-negative sequences associated to distances between its points. By using the classical correspondence between a stationary point process and its Palm measure, we establish a characterization of the corresponding Palm measure as an invariant distribution of a Markovian kernel. We prove that if Φ is continuous and its growth rate is at most linear with a rate below some constant, then there exists a stationary Hawkes point process. The classical Lipschitz condition of the literature for an unbounded function Φ is relaxed. Our proofs rely on a combination of coupling methods, monotonicity properties of linear Hawkes processes and classical results on Palm distributions. An investigation of the Hawkes process starting from the null measure, the empty state, on $\mathbb{R}_-$ plays also an important role. The linear case of Hawkes and Oakes is revisited at this occasion.

If the memory function h is an exponential function, under a weak condition it is shown that there exists a unique stationary Hawkes point process. In this case, its Palm measure is expressed in terms of the invariant distribution of a one-dimensional Harris ergodic Markov chain. When the activation function is a polynomial Φ with degree > 1, there does not exist a stationary Hawkes process and if the Hawkes process starts from the empty state, a scaling result for the accumulation of its points is obtained.

---

[*]INRIA Paris, 2 rue Simone Iff, 75589 Paris Cedex 12, France.
  E-mail: Philippe.Robert@inria.fr, https://www-rocq.inria.fr/who/Philippe.Robert
[†]INRIA Paris, 2 rue Simone Iff, 75589 Paris Cedex 12, France. Center for Interdisciplinary Research in Biology (CIRB) - Collège de France (CNRS UMR 7241, INSERM U1050), 11 Place Marcelin Berthelot, 75005 Paris, France.
  E-mail: Gaetan.Vignoud@inria.fr

1 Introduction

A point process $\mathcal{N} = (t_n : n \geq 0)$ on $\mathbb{R}_+$ has the Hawkes property if conditionally on the points t_n, $n \geq 1$ of $\mathcal{N}$ up to time t, the rate at which a new jump occurs in the time interval $(t, t + \mathrm{d}t)$ is given by

$$\Phi\left(\sum_{n:t_n \leq t} h(t - t_n)\right) \mathrm{d}t = \Phi\left(\int_{-\infty}^{t} h(t - u)\mathcal{N}(\mathrm{d}u)\right) \mathrm{d}t,$$

This is described as a self-excitation property of the dynamic. It can be formulated as the fact that the process

$$\left(\mathcal{N}(0, t) - \int_0^t \Phi\left(\int_{-\infty}^s h(s - u)\mathcal{N}(\mathrm{d}u)\right)\mathrm{d}s\right)$$

is a local martingale with respect to a convenient filtration.

The parameters of the Hawkes process

a. The function h, the *memory function*. The quantity $h(t-s)$ gives the residual impact at time t of a jump which has occurred at time $s \leq t$. It is assumed that it is a non-increasing continuous function, such that the function $(th(t))$ is converging to 0 at infinity and

$$\alpha \stackrel{\text{def.}}{=} \int_0^{+\infty} h(u)\,\mathrm{d}u < +\infty. \tag{1.1}$$

b. The *activation function* Φ modulates the global impact of past jumps. It is assumed that it is continuous and

$$\beta \stackrel{\text{def.}}{=} \limsup_{t \to +\infty} \frac{\Phi(t)}{t} < +\infty. \tag{1.2}$$

This class of models has been used in numerous situations such as, mathematical finance, Bauwens and Hautsch [1], population dynamics, Boumezoued [2], biology, Reynaud-Bouret and Schbath [41], queueing systems, Daw and Pende [9], learning theory, Etesami et al. [13], or neurosciences, Gerhard et al. [15], ...

This work is initially motivated by our investigations of mathematical models of plastic synapticity in neural networks, see Robert and Vignoud [43, 44, 45].

From a mathematical point of view, these processes have generated, and still generate, a considerable interest. Pioneering works on the existence and uniqueness of stationary point process, i.e. when the distribution of the point process is invariant with respect to translation, are:

— Hawkes and Oakes [20] shows that when the activation function is affine, these processes can be represented by an age-dependent branching process, a special class of the so-called Crump-Mode-Jagers models. The branching property is an important feature in the analysis of these models.

— Kerstan [26] gives an existence and uniqueness result of a stationary point process in the more general context of a fixed point relation for the distribution of random measures. A contraction argument in a convenient functional setting is the key ingredient. The reference Brémaud and Massoulié [4] has considerably developed this method in the case of Hawkes processes.

In this setting, the main existence and uniqueness result for an unbounded activation function Φ and a general memory function h is obtained under the condition

that Φ is Lipschitz with coefficient L_Φ and that the relation $\alpha L_\Phi < 1$ holds, where α is defined by Relation (1.1). Note that $L_\Phi \geq \beta$, with β defined by Relation (1.2). We will see in this paper that the weaker conditions that Φ is continuous and that $\alpha\beta < 1$ are enough for the existence result.

Up to now, these are mainly the two main approaches to investigate the existence of stationary Hawkes processes. Renewal properties are also an important tool in the study of Hawkes processes. See Graham [16], Raad [39], and Costa et al. [7]. Appendix B is an attempt to present some aspects of the overwhelming literature of this domain in a table. Functional limit theorems like law of large numbers or central limit theorems and large deviations properties are also topics of interest for these processes.

In our framework, assuming that a point process is given on $\mathbb{R}_-$, the Hawkes property on $\mathbb{R}_+$ is expressed with a solution of a Stochastic Differential Equation (SDE) satisfied by its counting process. We reformulate this property in terms of the distance between the jump times of the point process. In the case of a stationary Hawkes point process, this gives a Markovian characterization of its Palm measure in terms of a positive recurrence property.

The natural state space to study Palm measures is a sub-space of non-negative sequences, of distances between the successive points of the point process. Unfortunately the appropriate state space is not complete as a metric space and, moreover, the Markov process does not have the Feller property so that the classical tools to prove positive recurrence do not seem to be available. See Hairer [17] for a quick presentation. Nevertheless, our main result, Theorem 6.2 shows that under appropriate, weak, conditions on Φ there exists such a Palm measure and therefore a stationary Hawkes process. It should be noted however that we have not been able to obtain a significant uniqueness result under these weak conditions. Contraction arguments, which are apparently not possible under our weak assumptions, may be required for that purpose. Nevertheless, an interesting coupling property, Corollary 6.5, is proved.

Our approach uses the Markov chain starting from the empty state, i.e. the solution of the SDE when the initial state is the Dirac measure at 0 (no activity on $\mathbb{R}_-$ except at 0). Proposition 4.5 shows that the points of the corresponding point process are lower-bounded by a functional of a simple random walk. A simple characterization of non-homogeneous point processes, see Proposition 3.2, plays an important role to derive this result. The existing literature of Hawkes processes relies more on a stochastic calculus approach, via a formulation in terms of previsible projections of stochastic intensities.

A second ingredient is a coupling with an linear Hawkes process, the Hawkes process of Hawkes and Oakes [20] when Φ is an affine function. Our analysis of linear Hawkes processes does not explicitly use the natural branching property of this model. Monotonicity properties and technical estimates give then the desired existence result under weak conditions. See the discussions of Section B of the appendix and at the beginning of Section 6.

When the memory function is exponential, an existence result of a stationary Hawkes process is obtained under an even weaker condition and an explicit representation of the Palm measure is obtained. We also analyze the transient case for which there are few studies in general. In this case, the non-decreasing sequence of points of the Hawkes dynamics blows-up, i.e. converges almost surely to a finite limit. The self-excitation property of the dynamic leads to an explosion in finite time. We derive a scaling result, see Theorem 7.9, for the sequence of point in terms of a Poisson process.

The paper is organized as follows. Section 2 introduces some basic definitions, Section 3 expresses the Hawkes property in terms of a stochastic differential equation and Section 4 gives a characterization of the Palm measure of a stationary Hawkes

process as an invariant measure of a Markov chain in the state space of non-negative sequences. Section 5 revisits a classical Hawkes process, when the activation function is affine. Section 6 gives the main existence of this paper with this approach and by using coupling techniques and some monotonicity properties. Section 7 investigates Hawkes process with an exponential memory function. Appendix A presents the main definitions and results concerning stationary point processes. Appendix B gives a quick review of the results and the methods for the existence and uniqueness of stationary Hawkes processes.

2 Definitions and notations

2.1 Probability space

It is assumed that on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is defined a Poisson point process $\mathcal{P}$ on $\mathbb{R}_+ \times \mathbb{R}$ with intensity measure $\mathrm{d}x \otimes \mathrm{d}y$. See Kingman [27], Last and Penrose [30] and Chapter 1 of Robert [42] for a brief account on Poisson processes.

Additionally $(\mathcal{F}_t)$ is a filtration such that, for $t \in \mathbb{R}$, $\mathcal{F}_t$ is the σ-field generated by the random variables $\mathcal{P}(A \times (s,t])$, where A is a Borelian set of $\mathbb{R}_+$ and $s \le t$. If λ and f are non-negative Borelian functions f on $\mathbb{R}$, we define

$$\int_{\mathbb{R}} f(s) \mathcal{P}((0, \lambda(s)], \mathrm{d}s) \stackrel{\mathrm{def.}}{=} \int_{\mathbb{R}_+ \times \mathbb{R}} f(s) \mathbb{1}_{\{v \le \lambda(s)\}} \mathcal{P}(\mathrm{d}v, \mathrm{d}s),$$

and it is assumed that the σ-field $\mathcal{F}$ of the probability space verifies

$$\bigcup_{t \ge 0} \mathcal{F}_t \subset \mathcal{F}.$$

The martingale and stopping time properties are understood with respect to this filtration.

If H is $\mathbb{R}$ or $\mathbb{R}_-$, we denote by $C_c(H)$ the space of continuous functions on H with compact support and $\mathcal{M}_p(H)$, the set of Radon point measures on H, that is positive Radon measures carried by points, for $m \in \mathcal{M}_p(H)$ then

$$m = \sum_{x \in S} \delta_x,$$

where δ_x is the Dirac measure at $x \in H$ and S is a countable subset of H with no limiting points in H. We may also represent m as a sequence $(x_n, n \in \mathbb{N})$ of points. If A is a subset of H, we denote by

$$m(A) = \int_A m(\mathrm{d}x) = \sum_{x \in S} \mathbb{1}_{\{x \in A\}},$$

the number of points of m in A. A point measure m is *simple* if $m(\{x\}) \in \{0, 1\}$, for all $x \in H$. The space $\mathcal{M}_p(H)$ is endowed with the topology of weak convergence.

2.2 State space of non-negative sequences

We refer to Appendix A for general definitions concerning point processes. We denote by $\mathcal{S}$ the sub-space of sequences of non-negative real numbers

$$\mathcal{S} \stackrel{\mathrm{def.}}{=} \left\{ x = (x_k) \in (\mathbb{R}_+ \cup \{+\infty\})^{\mathbb{N} \setminus \{0\}} : x_{k_0} = +\infty \Rightarrow x_k = +\infty, \, \forall k \ge k_0 \right\}. \qquad (2.1)$$

Correspondence between $\mathcal{S}$ and $\mathcal{M}_p(\mathbb{R}_-)$

A functional from $\mathcal{S}$ to the space of positive measures with a mass at 0 is introduced as follows, if $x = (x_k) \in \mathcal{S}$, the positive measure m_x on $\mathbb{R}_-$ is defined by

$$m_x = \delta_0 + \sum_{k=1}^{+\infty} \delta_{t_k}, \quad \text{with } t_k = -\sum_{i=1}^{k} x_i, \, k \geq 1, \tag{2.2}$$

with the convention that $\delta_{-\infty} \equiv 0$.

The measure m_x is a positive measure carried by points associated to $x \in \mathcal{S}$. Note that m_x is not necessarily a point measure, i.e it may not have Radon property, since we do not exclude the fact that the sequence (x_k) converges to 0 sufficiently fast so that the measure m_x may have a finite limiting point.

With this definition 0 is always a point of m_x and that the coordinates of x are the inter-arrivals times of m_x, in particular x_1 is the distance to the first point of m_x on the left of 0. The point measures with a finite number of points correspond to sequences (x_k) which are constant and equal to $+\infty$ after some finite index. In this case, if k_0 is the first index where $x_{k_0} = +\infty$, with a slight abuse of notation we will write it as a finite vector $x = (x_1, x_2, \ldots, x_{k_0-1}, +\infty)$ or $x = (x_1, x_2, \ldots, x_{k_0-1})$.

On $\mathcal{S}$, the distance, for $x = (x_k)$, $y = (y_k) \in \mathcal{S}$,

$$d(x, y) = \sum_{1}^{+\infty} \frac{1}{2^k} \, \min(|x_k - y_k|, 1), \tag{2.3}$$

with the convention, for $u \in \mathbb{R}_+$, $|u - \infty| = |\infty - u| = +\infty$ and $|\infty - \infty| = 0$.

An important subset of $\mathcal{S}$ is

$$\mathcal{S}_h \overset{\text{def.}}{=} \left\{ x = (x_k) \in \mathcal{S} : h(0) + \sum_{k=1}^{+\infty} h\left(\sum_{i=1}^{k} x_i\right) = \int_0^{+\infty} h(-u)\, m_x(\mathrm{d}u) < +\infty \right\}, \tag{2.4}$$

we have $x \in \mathcal{S}_h$ if and only if $(h(-u)) \in L_1(m_x)$. Throughout the paper, we will use the following convention, for $t \geq 0$ and $x \in \mathcal{S}_h$,

$$\sum_{k=0}^{+\infty} h\left(t + \sum_{i=1}^{k} x_i\right) = h(t) + \sum_{k=1}^{+\infty} h\left(t + \sum_{i=1}^{k} x_i\right). \tag{2.5}$$

Definition 2.1. *If $\Phi(0) > 0$, for $a \geq 0$ and $x \in \mathcal{S}_h$, we set*

$$\mathcal{T}(x, a) \overset{\text{def.}}{=} \inf_{t \geq 0} \left\{ \int_0^t \Phi\left(h(s) + \sum_{k \geq 1} h\left(s + \sum_{i=1}^{k} x_i\right) \right) \mathrm{d}s \geq a \right\} \tag{2.6}$$

$$= \inf_{t \geq 0} \left\{ \int_0^t \Phi\left(\int_{(\infty, s)} h(s - u)\, m_x(\mathrm{d}u) \right) \mathrm{d}s \geq a \right\},$$

and $\mathcal{T}(x, a) \overset{\text{def.}}{=} 0$ otherwise, i.e. if $x \in \mathcal{S} \backslash \mathcal{S}_h$, with the convention $h(+\infty) = 0$.

Lemma 2.2. *If $\Phi(0) > 0$ and $x \in \mathcal{S}_h$ then,*

$$\lim_{t \to +\infty} \int_0^t \Phi\left(\int_{(-\infty, s)} h(s - u)\, m_x(\mathrm{d}u) \right) \mathrm{d}s = +\infty.$$

In particular, for $x \in \mathcal{S}_h$ and $a \geq 0$, the variable $\mathcal{T}(x, a)$ is finite.

Proof. Note that, for any $t \geq 0$, the monotonicity property of h gives $h(t-u) \leq h(-u)$ and, since $x \in \mathcal{S}_h$,

$$\int_{(-\infty,0)} h(-u) m_x(\mathrm{d}u) < +\infty,$$

with Lebesgue's dominated convergence Theorem we obtain the identity

$$\lim_{t \to +\infty} \int_{(-\infty,0)} h(t-u) m_x(\mathrm{d}x) = 0. \tag{2.7}$$

Our lemma is proved since Φ is continuous with $\Phi(0) > 0$. $\qquad\square$

3 Hawkes SDEs

We first recall the classical definition of an Hawkes process, see Hawkes and Oakes [20].

Definition 3.1. *A* Hawkes process *is a point process $\mathcal{N}$ on $\mathbb{R}$ such that, for any $s \in \mathbb{R}$, the process*

$$\left(\mathcal{N}((s,t]) - \int_s^t \Phi\left(\int_{(-\infty,u)} h(u-x)\mathcal{N}(\mathrm{d}x) \right) \mathrm{d}u, t \geq s \right)$$

is a local martingale with respect to the filtration $(\mathcal{F}_t, t \geq s)$, where, for $t \in \mathbb{R}$, $\mathcal{F}_t$ is the σ-field containing the σ-field associated to the random variables $\mathcal{N}((u,v])$, $u \leq v \leq t$.

If $\mathcal{N}$ is a Hawkes process, for any $s \in \mathbb{R}$ the dual predictable projection of the process of $(\mathcal{N}((s,t]), t \geq s)$ is almost surely,

$$\left(\int_s^t \Phi\left(\int_{(-\infty,u)} h(u-x)\mathcal{N}(\mathrm{d}x) \right) \mathrm{d}u, t \geq s \right),$$

see Theorem VI (21.7) of Rogers and Williams [46] for example. In the terminology of random measures, see Jacod [23], the *stochastic intensity* of $\mathcal{N}$ is

$$\left(\Phi\left(\int_{(-\infty,u)} h(u-x)\mathcal{N}(\mathrm{d}x) \right), u \in \mathbb{R} \right).$$

We now introduce a dynamical system extending a Radon measure on $\mathbb{R}_-$ into a measure on $\mathbb{R}$ exhibiting a Hawkes property on $\mathbb{R}_+$. The Markovian approach used in this paper relies heavily on this construction.

Proposition 3.2 (Hawkes SDE). *If $m \in \mathcal{M}_p(\mathbb{R}_-)$ is such that $(h(-u)) \in L_1(m)$, then there exists a unique positive random measure $\mathcal{N}_m$ on $\mathbb{R}$ such that $\mathcal{N}_m \equiv m$ on $\mathbb{R}_-$ and the counting measure $(\mathcal{N}_m((0,t]), t \geq 0)$ satisfies the stochastic differential equation*

$$\mathrm{d}\mathcal{N}_m((0,t]) = \mathcal{P}\left(\left(0, \Phi\left(\int_{(-\infty,t)} h(t-x)\mathcal{N}_m(\mathrm{d}x) \right) \right), \mathrm{d}t \right), \tag{3.1}$$

for all $t > 0$.

If $\Phi(0) > 0$, the points of $\mathcal{N}_m$ on $\mathbb{R}_+$ is a non-decreasing sequence of stopping times $(T_n, n \geq 1)$, such that if

$$E_n \overset{\mathrm{def.}}{=} \int_{T_{n-1}}^{T_n} \Phi\left(\int_{(-\infty,s)} h(s-x)\mathcal{N}_m(\mathrm{d}x) \right) \mathrm{d}s, \quad n \geq 1, \tag{3.2}$$

with the convention $T_0 = 0$, then $(E_n, n \geq 1)$ is an i.i.d. sequence of exponential random variables with parameter 1 and, for $n \geq 1$, the sequence $(E_k, k > n)$ is independent of $\mathcal{F}_{T_n}$.

Proof. We construct by induction the sequence of points (T_n) of $\mathcal{N}_m$. The first point is defined as

$$T_1 \stackrel{\text{def.}}{=} \inf \left\{ t \geq 0 : \int_{(0,t]} \mathcal{P}\left(\left(0, \Phi\left(\int_{(-\infty,0]} h(s-x)m(\mathrm{d}x) \right) \right), \mathrm{d}s \right) \neq 0 \right\}.$$

Let $\mathcal{R}$ be a Poisson process on $\mathbb{R}_+$ with intensity $\mathrm{d}x$, the relation

$$\left(\int_{(0,t]} \mathcal{P}\left(\left(0, \Phi\left(\int_{(-\infty,0]} h(s-u)m(\mathrm{d}u) \right) \right), \mathrm{d}s \right), t \geq 0 \right)$$
$$\stackrel{\text{dist.}}{=} \left(\mathcal{R}\left(\left(0, \int_{(0,t]} \Phi\left(\int_{(-\infty,0]} h(s-u)m(\mathrm{d}u) \right) \mathrm{d}s \right) \right), t \geq 0 \right) \quad (3.3)$$

between the processes of counting measures of two point processes is an easy consequence of the equality of their respective Laplace transforms. See Chapter 1 of Robert [42].

If E_1 is the first point of $\mathcal{R}$, we deduce therefore the identity

$$\int_{(0,T_1]} \Phi\left(\int_{(-\infty,0]} h(s-u)m(\mathrm{d}u) \right) \mathrm{d}s \stackrel{\text{dist.}}{=} E_1.$$

Using Definition 2.1, the random variable T_1 has therefore the same distribution as the random variable $\mathcal{T}(x, E_1)$ where x is the unique element of $\mathcal{S}_h$ such that $m = m_x$ and E_1 is an exponential random variable with parameter 1. It is in particular almost surely finite by Lemma 2.2.

By induction, for $n \geq 1$, if $T_1 \leq T_2 \leq \cdots \leq T_n < +\infty$, it is possible to define

$$T_{n+1} \stackrel{\text{def.}}{=} \inf \left\{ t \geq T_n : \right.$$
$$\left. \int_{(T_n,t]} \mathcal{P}\left(\Phi\left(\int_{(-\infty,0]} h(s-x)m(\mathrm{d}x) \right) + \sum_{k=1}^{n} h\left(s-T_k\right), \mathrm{d}s \right) \neq 0 \right\}, \quad (3.4)$$

The random variable T_{n+1} as defined by (3.4) is clearly a stopping time, it is almost surely finite with the same argument as for T_1. With the strong Markov property of $\mathcal{P}$ for the stopping time T_n and Relation (3.3), but conditionally on $\mathcal{F}_{T_n}$,

$$\left(\int_{(T_n,T_n+t]} \mathcal{P}\left(\Phi\left(\int_{(-\infty,0]} h(s-x)m(\mathrm{d}x) \right) + \sum_{k=1}^{n} h\left(s-T_k\right), \mathrm{d}s \right) \middle| \mathcal{F}_{T_n} \right)$$
$$\stackrel{\text{dist.}}{=} \left(\mathcal{R}\left(\left(0, \int_{T_n}^{T_n+t} \Phi\left(\int_{(-\infty,0]} h(s-u)m(\mathrm{d}u) + \sum_{k=1}^{n} h(s-T_k) \right) \mathrm{d}s \right) \right) \middle| \mathcal{F}_{T_n} \right),$$

hence, as for the case $n = 1$, if E_{n+1} is the distance between the nth and $(n+1)$th point of $\mathcal{R}$, it is an exponential random variable independent of $\mathcal{F}_{T_n}$ and

$$E_{n+1} = \int_{T_n}^{T_{n+1}} \Phi\left(\int_{(-\infty,0]} h(s-u)m(\mathrm{d}u) + \sum_{k=1}^{n} h(s-T_k) \right) \mathrm{d}s$$
$$= \int_{T_n}^{T_{n+1}} \Phi\left(\int_{(-\infty,s)} h(s-x)\,\mathcal{N}_m(\mathrm{d}x) \right) \mathrm{d}s.$$

The proposition is proved. $\qquad\square$

In the proof we have seen that the condition $\Phi(0) > 0$ in Lemma 2.2 guarantees that there is almost surely an infinite number of points in $\mathbb{R}_+$ for SDE (3.1).

It cannot be excluded nevertheless that this non-decreasing sequence of points may have a finite limit with positive probability and therefore that $\mathcal{N}_m$ may not be a point process. As it will be seen, it can indeed blow-up in finite time, i.e. the limit of its sequence (T_n) of points in $\mathbb{R}_+$ may be finite with positive probability. Proposition 4.5 gives a condition for which the sequence (T_n) converges almost surely to infinity when m is the null measure. Theorem 7.9 considers a case when there is such a blow-up and establishes a limiting result for the accumulation of these points.

However, when the sequence (T_n) is converging almost surely to infinity, $\mathcal{N}_m$ is a point process, and it is not difficult to see that the dual predictable projection of the process of $(\mathcal{N}_m((0, t]), t \geq 0)$ is

$$\left(\int_0^t \Phi \left(\int_{(-\infty, s)} h(s - x) \mathcal{N}_m(\mathrm{d}x) \right) \mathrm{d}s, t \geq 0 \right).$$

see Theorem VI (27.1) of Rogers and Williams [46]. Consequently, the stochastic intensity function of $\mathcal{N}_m$ on $\mathbb{R}_+$ is indeed

$$\left(\Phi \left(\int_{(-\infty, s)} h(s - x) \mathcal{N}_m(\mathrm{d}x) \right) \right).$$

Proposition 3.2 extends a point measure m on $\mathbb{R}_-$ to a point process on $\mathbb{R}$, it can be also seen as a dynamical system on point processes on $\mathbb{R}_-$ in the following way.

A dynamical system

If $m \in \mathcal{M}_p(\mathbb{R}_-)$ is such that $h(-x) \in L_1(m)$ and the non-negative measure $\mathcal{N}_m$ defined by Relation (3.1) is a point process, for $t \geq 0$, we introduce a (random) dynamical system $(T_t(m))$ in $\mathcal{M}_p(\mathbb{R}_-)$ as,

$$\int_{(-\infty, 0]} f(x) T_t(m)(\mathrm{d}x) = \int f(x) \theta_t(\mathcal{N}_m)(\mathrm{d}x) = \int_{(-\infty, t]} f(x - t) \mathcal{N}_m(\mathrm{d}x), \tag{3.5}$$

for any non-negative Borelian function on $\mathbb{R}_-$, $T_t(m)$ is the point process $\mathcal{N}_m$ seen from the point t.

A stationary point process is a distribution Q on $\mathcal{M}_p(\mathbb{R})$ which is invariant distribution for the group of transformations (θ_t). See Definition A.1 in Section A of the appendix. By Relation (3.1), it can be formulated as the existence of a distribution on $\mathcal{M}_p(\mathbb{R}_-)$ invariant by the operator T_t.

When $\Phi(0) = 0$, the null measure is clearly a solution of Relation (3.1). The next proposition shows that, under some mild condition, the null point process is the unique stationary Hawkes process in such a case.

Proposition 3.3. *Under the condition*

$$\int_{\mathbb{R}_+} t h(t) \, \mathrm{d}t < +\infty,$$

if for some $K \geq 0$, the non-negative function Φ satisfies the relation $\Phi(x) \leq Kx$, for all $x \geq 0$, then there does not exist an ergodic stationary Hawkes process $\mathcal{N}$ which is non-trivial, i.e. such that $\mathbb{P}(\mathcal{N} \not\equiv 0) > 0$.

See Definition A.1 for the stationarity and ergodicity properties, they are with respect to the flow of translation (θ_t). See Cornfeld et al. [6].

Proof. Assume that $\mathcal{N}$ is a non-trivial ergodic stationary Hawkes process. The Birkhoff-Khinchin ergodic theorem, see [6], gives the almost sure convergence

$$\lim_{t \to +\infty} \frac{\mathcal{N}((0,t])}{t} = \lambda \stackrel{\text{def.}}{=} \mathbb{E}(\mathcal{N}((0,1])) > 0,$$

since $\mathcal{N}$ is non-trivial, we obtain therefore that $\mathcal{N}(\mathbb{R}_+)$ is almost surely infinite. If T_1 is the first positive point of $\mathcal{N}$, the variable T_1 is in particular finite with probability 1.

As in the proof of Proposition 3.2, we have

$$\mathbb{P}(T_1 \geq t \mid \mathcal{F}_0) = \exp\left(-\int_0^t \Phi\left(\int_{(-\infty,0]} h(s-x)\mathcal{N}(\mathrm{d}x)\right)\mathrm{d}s\right).$$

The stationary property of the point process gives the relation

$$\mathbb{E}\left(\int_{\mathbb{R}} f(s)\mathcal{N}(\mathrm{d}s)\right) = \lambda \int_{\mathbb{R}} f(s)\,\mathrm{d}s,$$

for all $f \in \mathcal{C}_c(\mathbb{R})$. If

$$V \stackrel{\text{def.}}{=} \int_0^{+\infty} \Phi\left(\int_{(-\infty,0]} h(s-x)\mathcal{N}(\mathrm{d}x)\right)\mathrm{d}s,$$

we have, by Fubini's Theorem,

$$\mathbb{E}(V) \leq K\,\mathbb{E}\left(\int_0^{+\infty}\int_{(-\infty,0]} h(s-x)\mathcal{N}(\mathrm{d}x)\,\mathrm{d}s\right)$$

$$= \lambda K \int_0^{+\infty}\int_{(-\infty,0]} h(s-x)\,\mathrm{d}x\,\mathrm{d}s = \lambda K \int_0^{+\infty} sh(s)\,\mathrm{d}s < +\infty,$$

hence V is an integrable random variable, it is in particular almost surely finite. As a consequence we have that, almost surely, $\mathbb{P}(T_1 = +\infty|\mathcal{F}_0) > 0$. This is a contradiction. The proposition is proved. $\qquad\square$

4 A Markov chain formulation

In this section, we give another version of Proposition 3.2 in terms of a Markovian dependence on the state space of sequences.

4.1 An alternative formulation

We begin with a characterization of the Palm measure of a stationary Hawkes process.

For $m \in \mathcal{M}_p(\mathbb{R})$, if $(\tau_n(m), n \in \mathbb{Z}) \stackrel{\text{def.}}{=} (t_n(m) - t_{n-1}(m), n \in \mathbb{Z})$, where $(t_n(m))$ is given by Relation (A.1) of the Appendix. As a random variable on $\mathcal{M}_p(\mathbb{R})$, the sequence is also represented as $(\tau_n, n \in \mathbb{Z})$.

Proposition 4.1. *If $\Phi(0) > 0$ and if Q is the distribution on $\mathcal{M}_p(\mathbb{R})$ of a stationary Hawkes process associated to Φ and h then, under its associated Palm measure $\widehat{Q}$, the sequence of inter-arrivals $(\tau_n, n \in \mathbb{Z})$ is a stationary sequence. The sequence of random variables*

$$\left(\int_0^{\tau_{n+1}} \Phi\left(\sum_{k \leq n} h\left(s + \sum_{i=k+1}^{n} \tau_i\right)\right)\mathrm{d}s, \quad n \in \mathbb{Z}\right), \tag{4.1}$$

is i.i.d. with a common exponential distribution with parameter 1.

If there exists a stationary sequence $(\tau_n, n \in \mathbb{Z})$ satisfying Relation (4.1), then there exists a stationary Hawkes process associated to Φ and h.

From the point of view of Proposition 3.2 this proposition is quite intuitive, one has nevertheless to be careful, as always in this setting, since Proposition 3.2 is stated under the distribution Q.

Proof. For simplicity, in this proof we will use the canonical probability space $\Omega = \mathcal{M}_p(\mathbb{R})$ endowed with the weak topology and $\mathbb{E}_R$ denotes the expectation with respect to a probability distribution R on $\mathcal{M}_p(\mathbb{R})$.

The stationary property of the sequence $(\tau_n, n \in \mathbb{Z})$ under $\widehat{Q}$ is clear, by definition of $\widehat{Q}$. For $t \geq 0$, if f is a bounded Borelian function on $\mathcal{M}_p(\mathbb{R})$, we denote

$$\mathcal{E}_t \overset{\text{def.}}{=} \{m \in \mathcal{M}_p(\mathbb{R}) : m([-t,0]) \neq 0\}$$

then, Proposition 11.8, page 315 of Robert [42] shows that if $t \mapsto f(\theta_t(m))$ is Q-almost surely right continuous at $t = 0$, then the relation

$$\lim_{t \searrow 0} \mathbb{E}_Q\left(f \mid \mathcal{E}_t\right) = E_{\widehat{Q}}(f)$$

holds.

For $n \in \mathbb{Z}$ and $m \in \mathcal{M}_p(\mathbb{R})$, define

$$\Psi_n(m) \overset{\text{def.}}{=} \int_{t_n(m)}^{t_{n+1}(m)} \Phi\left(\int_{(-\infty, s]} h(s-x) m(\mathrm{d}x)\right) \mathrm{d}s,$$

since, Q-almost surely, $t_0(m) < 0 < t_1(m)$ and $t_n(m) < t_{n+1}(m)$, then for $t \geq 0$ such that $0 \leq t < t_1(m)$, we have

$$\Psi_n(\theta_t(m)) = \int_{t_n(m)-t}^{t_{n+1}(m)-t} \Phi\left(\int_{(-\infty, s+t]} h(s+t-x) m(\mathrm{d}x)\right) \mathrm{d}s = \Psi_n(m).$$

Let F be a continuous bounded Borelian function on $\mathbb{R}_+^n$ then

$$\lim_{t \searrow 0} \mathbb{E}_Q\left(F\left((\Psi_i, 1 \leq i \leq n)\right) \Big| \mathcal{E}_t\right) = \mathbb{E}_{\widehat{Q}}\left(F\left((\Psi_i, 1 \leq i \leq n)\right)\right).$$

Since the event $\mathcal{E}_t$ is $\mathcal{F}_0$-measurable and that for any $1 \leq k \leq n$, the random variable

$$m \longrightarrow \int_{t_k(m)}^{t_{k+1}(m)} \Phi\left(\int_{(-\infty, s]} h(s-x) m(\mathrm{d}x)\right) \mathrm{d}s$$

is independent of $\mathcal{F}_{t_k}$ and therefore of $\mathcal{F}_0$, by Proposition 3.2, by induction on k for example, we obtain that

$$\lim_{t \searrow 0} \mathbb{E}_Q\left(F\left((\Psi_i, 1 \leq i \leq n)\right) \Big| \mathcal{E}_t\right) = \mathbb{E}\left(F\left((E_i, 1 \leq i \leq n)\right)\right),$$

where $(E_i, 1 \leq i \leq n)$ are n independent random variables with a common exponential distribution with parameter 1. By gathering these results we obtain that

$$\mathbb{E}_{\widehat{Q}}\left(F\left((\Psi_i, 1 \leq i \leq n)\right)\right) = \mathbb{E}\left(F\left((E_i, 1 \leq i \leq n)\right)\right).$$

Under the probability distribution $\widehat{Q}$, the random variables Ψ_i, $1 \leq i \leq n$, are i.i.d. with a common exponential distribution with parameter 1. Relation (4.1) is a consequence of

the relation, for $m \in \mathcal{M}_p(\mathbb{R})$,

$$
\begin{aligned}
\Psi_n(m) &= \int_{t_n(m)}^{t_{n+1}(m)} \Phi\left(\sum_{k \leq n} h\left(s - t_k(m)\right)\right) \mathrm{d}s \\
&= \int_0^{\tau_{n+1}(m)} \Phi\left(\sum_{k \leq n} h\left(s + \sum_{i=k+1}^{n} \tau_i(m)\right)\right) \mathrm{d}s,
\end{aligned}
$$

hence $\Psi_{n+1}(m)$ is the variable $\Psi_n(m)$ associated to the sequence $(\tau_{n+1}) = \widehat{\theta}(\tau_n)$, see Appendix A. By invariance of $\widehat{Q}$ with respect to $\widehat{\theta}$, see Proposition 11.18 of Robert [42] for example, the sequence $(\Psi_n, n \in \mathbb{Z})$ is stationary under $\widehat{Q}$ and, therefore, it is i.i.d. with a common exponential distribution with parameter 1.

Now we assume that there exists a stationary sequence $\tau = (\tau_n, n \in \mathbb{Z})$ of integrable random variables satisfying Relation (4.1). Recall that m_τ is the point process defined by Relation (2.2). Using a similar proof as in Proposition 3.2, we can show that m_τ satisfies an Hawkes SDE (3.1) on $\mathbb{R}$. For $u \in \mathbb{R}$, the same property clearly holds for $\theta_u(m_\tau)$, the point process m_τ translated at u, see Definition (A.2). For $K > 0$, let U_K be an independent uniform random variable on $[-K, K]$, the point process $\theta_{U_K}(m_\tau)$ satisfies the Hawkes property. With the same method used for the proof of Proposition 11.2 of [42], we obtain that, as K goes to infinity, $\theta_{U_K}(m_\tau)$ converges in distribution to a stationary point process whose Palm measure is given by the distribution τ. It is not difficult to show that the Hawkes property is preserved in the limit. The proposition is proved. □

4.2 A Markov chain on $\mathcal{S}$

The previous proposition has highlighted the importance of the Palm space, and therefore led us to develop a second formulation of the Hawkes property using a Markovian kernel.

Definition 4.2. *The sequence of random variables $(\mathcal{X}_n^x)$ with initial point $\mathcal{X}_0^x = x \in \mathcal{S}$ is defined by induction as follows, for $n \geq 0$,*

$$
\mathcal{X}_{n+1}^x = (X_{n+1}^x, \mathcal{X}_n^x) \tag{4.2}
$$

where X_{n+1}^x is defined by the relation

$$
\int_0^{X_{n+1}^x} \Phi\left(\int_{\mathbb{R}_-} h(T_n + s - u)m(\mathrm{d}u) + \sum_{k=1}^{n} h\left(s + \sum_{i=k+1}^{n} X_i^x\right)\right) \mathrm{d}s = E_{n+1}, \tag{4.3}
$$

with $T_n = X_1^x + \cdots + X_n^x$ and (E_k) is an i.i.d. sequence of exponentially distributed random variable with parameter 1.

The associated Markovian kernel is denoted by $\mathcal{K}$

$$
\int_{\mathcal{S}} f(y)\mathcal{K}(x, \mathrm{d}y) = \mathbb{E}_x(f(\mathcal{X}_1^x)), \tag{4.4}
$$

for a non-negative Borelian function f on $\mathcal{S}$.

The element $\mathcal{X}_{n+1}^x$ is obtained by shifting $\mathcal{X}_n^x$ and adding $X_{n+1}^x = \mathcal{T}(\mathcal{X}_n^x, E_{n+1})$ at the beginning of the sequence. $\mathcal{T}$ is the operator defined by Relation (2.6).

The sequence $(\mathcal{X}_n^x)$ clearly has the Markov property.

Proposition 4.3 (The Markov chain $(\mathcal{X}_n^x)$ and Hawkes SDEs). *If $\Phi(0) > 0$ and $m \in \mathcal{M}_p(\mathbb{R}_-)$ is such that $m(\{0\}) > 0$, then the distribution of $\mathcal{N}_m$, the solution of Relation (3.1) can be expressed as*

$$\mathcal{N}_m \overset{\text{dist.}}{=} m + \sum_{n \geq 1} \delta_{T_n},$$

with $T_0 = 0$ and, for $n \geq 1$, $T_{n+1} - T_n = X_{n+1}^x$, where the sequence (X_n^x) is defined by Relation (4.3).

The process $(\mathcal{X}_n^x) = ((X_n^x, \ldots, X_1^x, \mathcal{X}_0^x))$ is the Markov chain with transition kernel $\mathcal{K}(\cdot, \cdot)$ of Relation (4.4) and initial point

$$\mathcal{X}_0^x = (s_{n+1} - s_n), \ \text{if } m = (s_n, n \geq 0)),$$

with the convention $\cdots \leq s_n \leq s_{n-1} \leq \cdots \leq s_1 \leq s_0 = 0$.

In (4.3), recall the convention of Relation (2.5), for $s \geq 0$ and $n \geq 2$,

$$\sum_{k=1}^{n-1} h\left(s + \sum_{i=k+1}^{n-1} X_i^x\right) = h(s) + \sum_{k=1}^{n-2} h\left(s + \sum_{i=k+1}^{n-1} X_i^x\right).$$

Proof. This is a straightforward consequence that Relation (4.3) is a rewriting of Relation (3.2) of Proposition 3.2. $\qquad\square$

We can now state the main result concerning the relation between the Hawkes SDE and the Markov transition kernel $\mathcal{K}$.

Proposition 4.4. *The Markov chain associated to transition kernel $\mathcal{K}$ of Definition (4.4) has an invariant distribution on $\mathcal{S}_h$ if and only if there exists a stationary Hawkes process associated to the functions Φ and h.*

The distribution of the sequence of inter-arrival times of a stationary Hawkes process associated to the functions Φ and h is an invariant measure for the Markov chain associated to transition kernel $\mathcal{K}$.

Proof. If there exists a stationary Hawkes point process (t_n), Proposition 4.1 shows that the distribution of $(t_{n+1} - t_n)$ under its Palm distribution is an invariant distribution of the Markov chain $(\mathcal{X}_n)$. Conversely, if the Markov chain associated to transition kernel $\mathcal{K}$ has an invariant distribution, then one can construct a stationary version of the Markov chain $\mathcal{X} \overset{\text{def.}}{=} (\mathcal{X}_n)$, in particular $\mathcal{X}$ satisfies Relation (4.1). Proposition 4.1 shows then that there exists a stationary Hawkes point process in this case. $\qquad\square$

4.3 Markov chains starting from the empty state

We define Markov chains $(\mathcal{X}_n^0)$ with transition kernel $\mathcal{K}$ defined by Relation (4.4) when the initial state empty, i.e. it is the constant sequence equal to $+\infty$, i.e. $\mathcal{X}_0^0 = (+\infty)$. This initial state corresponds to the case of a point process with a point at 0 only. The sequence $(\mathcal{X}_n^0)$ is in fact associated to the point process $\mathcal{N}_{\delta_0}$ of Proposition 3.2.

The Markov chain can be defined by, for $n \geq 1$,

$$\mathcal{X}_n^0 = (X_n^0, X_{n-1}^0, X_{n-2}^0, \ldots, X_2^0, X_1^0, +\infty),$$

where the sequence (X_n^0) is defined by Relation (4.3) with $m = \delta_0$.

Proposition 4.5. *If $\Phi(0) > 0$ and α and β are defined by Relations (1.1) and (1.2), there exist $\nu > 0$ and $\beta_0 > \beta$ such that, almost surely, for all $n \geq 1$,*

$$\sum_{k=1}^{n} (E_k - \alpha\beta_0) \leq \nu \sum_{k=1}^{n} X_k^0 - \beta_0 \sum_{k=1}^{n} \overline{H}\left(\sum_{j=k}^{n} X_j^0\right), \tag{4.5}$$

with

$$\overline{H}(x) = \int_x^{+\infty} h(s)\, \mathrm{d}s,$$

If $\alpha\beta < 1$, the random measure $\mathcal{N}_{\delta_0}$ solution of Relation (3.1) of Proposition 3.2 is almost surely a point process.

Proof. For any $\beta_0 > \beta$, there exists some $\nu > 0$ such that Relation $\Phi(x) \leq \nu + \beta_0 x$, holds for all $x \geq 0$. Equation (4.3) gives, for $n \geq 1$,

$$E_n = \int_0^{X_n^0} \Phi\left(\sum_{k=1}^n h\left(s + \sum_{i=k}^{n-1} X_i^0\right)\right)\mathrm{d}s$$

$$\leq \nu X_n^0 + \beta_0 \int_0^{X_n^0} \sum_{k=1}^n h\left(s + \sum_{i=k}^{n-1} X_i^0\right)\mathrm{d}s, \quad (4.6)$$

and, therefore, with the definition of $\overline{H}$,

$$E_n \leq \nu X_n^0 + \beta_0 \int_0^{X_n^0} h(s)\, \mathrm{d}s + \beta_0 \int_0^{X_n^0} \sum_{k=1}^{n-1} h\left(s + \sum_{i=k}^{n-1} X_i^0\right)\mathrm{d}s,$$

$$E_n - \alpha\beta_0 \leq \nu X_n^0 - \beta_0\overline{H}(X_n^0) + \beta_0 \int_{X_{n-1}^0}^{X_{n-1}^0 + X_n^0} \sum_{k=1}^{n-1} h\left(s + \sum_{i=k}^{n-2} X_i^0\right)\mathrm{d}s.$$

By using Inequality (4.6) for the index $n-1$ and by adding these relations, we get

$$\sum_{k=n-1}^n (E_k - \alpha\beta_0) \leq \nu\left[X_n^0 + X_{n-1}^0\right] - \beta_0\left[\overline{H}(X_n^0) + \overline{H}(X_n^0 + X_{n-1}^0)\right]$$

$$+ \beta_0 \int_{X_{n-2}^0}^{X_{n-2}^0 + X_{n-1}^0 + X_n^0} \sum_{k=0}^{n-2} h\left(s + \sum_{i=k+1}^{n-3} X_i^0\right)\mathrm{d}s.$$

By proceeding by induction, we finally get the relation, for $0 \leq p < n$,

$$\sum_{k=n-p}^n (E_k - \alpha\beta_0) \leq \nu \sum_{k=n-p}^n X_k^0 - \beta_0 \sum_{k=n-p}^n \overline{H}\left(\sum_{j=k}^n X_j^0\right)$$

$$+ \beta_0 \int_{X_{n-p-1}^0}^{X_{n-p-1}^0 + \cdots + X_n^0} \sum_{k=1}^{n-p-1} h\left(s + \sum_{i=k}^{n-p-2} X_i^0\right)\mathrm{d}s,$$

for $p = n - 1$ it gives Relation (4.5).

If $\alpha\beta < 1$, β_0 can be chosen so that $\alpha\beta_0 < 1$ for some $\nu > 0$, hence, by the law of large numbers

$$\liminf_{n\to+\infty} \frac{1}{n}\sum_{k=1}^n X_k^0 \geq \frac{1 - \alpha\beta_0}{\nu} > 0,$$

holds almost surely, which implies that $\mathcal{N}_{\delta_0}$ is almost surely a point process. The proposition is proved. □

5 Affine activation function

In this section, we present quickly in a self-contained way a well-known existence result for a stationary Hawkes process when the activation function is affine,

$$\Phi_\beta(x) = \nu + \beta x, \quad x \geq 0, \quad (5.1)$$

where ν, $\beta > 0$. This is the model investigated in the classical reference Hawkes and Oakes [20]. The results are presented in our Markovian setting $(\mathcal{X}_n)$. It is an important ingredient in the proof of existence of stationary Hawkes processes of Section 6.

The associated Hawkes process can be interpreted in terms of birth instants of a branching process: births of external individuals occur according to a Poisson process with rate ν, each individual with age $t > 0$ gives birth to a new individual at rate $\beta h(t)$, the quantity $\alpha\beta$ turns out to be the average number of births due to an individual. Recall that α is the integral of h on $\mathbb{R}_+$, see Relation (1.1). This branching characterization of the model will not be explicitly used in our arguments.

We denote by $\mathcal{Y}_n^0 = (Y_n^0, Y_{n-1}^0, \ldots, Y_1^0, +\infty)$ the Markov chain of Definition 4.2 for the function $\Phi = \Phi_\beta$, whose initial state is empty, i.e. $\mathcal{Y}_0^0 = (+\infty)$. For any $n \geq 1$, we have

$$\int_0^{Y_n^0} \Phi\left(\sum_{k=1}^n h\left(s + \sum_{i=k+1}^{n-1} Y_i^0\right)\right) \mathrm{d}s = E_n, \tag{5.2}$$

where (E_n) is an i.i.d. sequence of exponential random variables with parameter 1.

Proposition 5.1. *Under the condition $\alpha\beta < 1$, the sequence $(\mathcal{Y}_n^0)$ is converging in distribution, for the topology induced by d of Relation (2.3), to a stationary sequence $\mathcal{Y}^\infty = (Y_1^\infty, Y_2^\infty, \ldots, Y_n^\infty, \ldots)$ such that*

$$\left(\int_0^{Y_n^\infty} \left(\Phi_\beta\left(\sum_{i=n}^{+\infty} h\left(u + \sum_{j=n+1}^i Y_j^\infty\right)\right)\right) \mathrm{d}u\right),$$

is an i.i.d. sequence of exponential random variables with parameter 1 and

$$\mathbb{E}(Y_1^\infty) = \frac{1 - \alpha\beta}{\nu},$$

furthermore, if

$$(I_n) \stackrel{\text{def.}}{=} \left(\sum_{k=1}^n h\left(\sum_{i=2}^k Y_i^n\right)\right) \text{ and } I \stackrel{\text{def.}}{=} \sum_{k=1}^{+\infty} h\left(\sum_{i=2}^k Y_i^\infty\right),$$

the sequence (I_n) converges in distribution to the integrable random variable I.

The proposition gives in particular that $Y_1^\infty \stackrel{\text{dist.}}{=} \mathcal{T}((Y_n^\infty, n \geq 2), E_1)$. The variable $\nu + \beta I$ is the intensity function of the stationary Hawkes process just after one of its point, i.e. for its Palm distribution.

As it can be seen, compared to $\mathcal{Y}^\infty$, the numbering of the coordinates of $\mathcal{Y}_n^0$ is reversed. This is mainly for a notational convenience in fact. The important fact is that the sums inside the function Φ in both cases are depending on the past.

Proof. The proof is done in several steps.

Step 1. Backward Coupling.

The idea is of using a backward coupling idea. See Chapter 22 of Levin et al. [31] for a general presentation and Loynes [34] for an early use of this method.

We define a sequence of random variables $(\widetilde{\mathcal{Y}}_n^0) = (\widetilde{Y}_1^n, \ldots, \widetilde{Y}_n^n, +\infty)$ by induction as follows, recall that (E_n) is an i.i.d. of random variables with an exponential distribution with parameter 1,

$$\begin{cases} \widetilde{Y}_n^n = \mathcal{T}((+\infty), E_n), \\ \widetilde{Y}_{n-k}^n = \mathcal{T}\left(\left(\widetilde{Y}_{n-k+1}^n, \ldots, \widetilde{Y}_n^n, +\infty\right), E_{n-k}\right), \quad 1 \leq k \leq n, \end{cases}$$

in particular,

$$\widetilde{Y}_1^n = \mathcal{T}\left(\left(\widetilde{Y}_2^n, \ldots, \widetilde{Y}_n^n, +\infty\right), E_1\right).$$

It is easily checked that

$$\left(\widetilde{Y}_1^n, \ldots, \widetilde{Y}_n^n, +\infty\right) \overset{\text{dist.}}{=} \left(\widetilde{Y}_2^{n+1}, \ldots, \widetilde{Y}_{n+1}^{n+1}, +\infty\right) \tag{5.3}$$

and that

$$\widetilde{\mathcal{Y}}_n^0 \overset{\text{dist.}}{=} \mathcal{Y}_n^0, \text{ since } (E_1, E_2, \ldots, E_n) \overset{\text{dist.}}{=} (E_n, E_{n-1}, \ldots, E_1). \tag{5.4}$$

For any $n \geq 1$ and $1 \leq k \leq n$, we show that $\widetilde{Y}_k^{n+1} \leq \widetilde{Y}_k^n$ holds. This is done by induction. Since h is non-negative and non-increasing, we have

$$E_n = \int_0^{\widetilde{Y}_n^n} (\nu + \beta h(u)) \, \mathrm{d}u \leq \int_0^{\widetilde{Y}_n^n} \left(\nu + \beta\left(h(u) + h\left(u + \widetilde{Y}_{n+1}^{n+1}\right)\right)\right) \mathrm{d}u,$$

and, since,

$$E_{n+1} = \int_0^{\widetilde{Y}_n^n} \left(\nu + \beta\left(h(u) + h\left(u + \widetilde{Y}_{n+1}^{n+1}\right)\right)\right) \mathrm{d}u,$$

this implies that $\widetilde{Y}_n^n \geq \widetilde{Y}_n^{n+1}$ holds. If the relation $\widetilde{Y}_i^n \geq \widetilde{Y}_i^{n+1}$ holds for any $k+1 \leq i \leq n$, then

$$E_k = \int_0^{\widetilde{Y}_k^n} \left(\nu + \beta \sum_{i=k}^n h\left(u + \sum_{j=k+1}^i \widetilde{Y}_j^n\right)\right) \mathrm{d}u$$

$$\leq \int_0^{\widetilde{Y}_k^n} \left(\nu + \beta \sum_{i=k}^n h\left(u + \sum_{j=k+1}^i \widetilde{Y}_j^{n+1}\right)\right) \mathrm{d}u, \quad (5.5)$$

which gives the relation $\widetilde{Y}_k^n \geq \widetilde{Y}_k^{n+1}$. We can now define, for $k \geq 1$,

$$\widetilde{Y}_k^\infty = \lim_{n \to +\infty} \widetilde{Y}_k^n.$$

Relation (5.3) shows that the sequence $(\widetilde{Y}_k^\infty)$ is stationary and Relation (5.4) gives the desired convergence in distribution for $(\mathcal{Y}_n^0)$.

Step 2. Convergence of the intensities (I_n). Proposition 4.5 shows that, for all $1 \leq i \leq n$,

$$\sum_{j=2}^i \widetilde{Y}_j^n \geq S_i \overset{\text{def.}}{=} \frac{1}{\nu} \sum_{j=2}^i (E_j - \alpha\beta). \tag{5.6}$$

The assumption on $\alpha\beta$ shows that, almost surely (S_i) is converging to infinity. If $a^+ = \max(a, 0)$, then

$$\sum_{i=2}^n h\left(\sum_{j=2}^i \widetilde{Y}_j^n\right) \leq \sum_{i=2}^n h\left(S_i^+\right).$$

Since h is non-increasing converging to 0 at infinity, there exists a non-negative finite Radon measure μ on $\mathbb{R}_+$ such that $h(x) = \mu((x, +\infty))$, with Fubini's Theorem, we obtain

$$\mathbb{E}\left(\sum_{i=1}^{+\infty} h\left(S_i^+\right)\right) = \int_0^{+\infty} \mathbb{E}\left(\mathcal{N}_+(0, x)\right) \mu(\mathrm{d}x), \tag{5.7}$$

where $\mathcal{N}_+$ is the counting measure of the sequence (S_n), i.e. for $x \geq 0$,

$$\mathcal{N}_+(0, x) \overset{\text{def.}}{=} \sum_{i=2}^{+\infty} \mathbb{1}_{\{x \geq S_i\}}. \tag{5.8}$$

The Laplace transform of $\mathcal{N}_+$ on $\mathbb{R}_+$ is given by, for $\xi \geq 0$ sufficiently small so that $\nu \exp(\alpha\beta\xi/\nu) < (\nu + \xi)$, we have

$$\omega(\xi) \overset{\text{def.}}{=} \mathbb{E}\left(\sum_{n \geq 1} e^{-\xi S_n}\right) = \frac{\nu + \xi}{\nu + \xi - \nu \exp(\alpha\beta\xi/\nu)}$$

Hence,

$$\lim_{\xi \to 0} \xi\omega(\xi) = \frac{\nu}{1 - \alpha\beta},$$

A Tauberian theorem, see Theorem 2 of Chapter XIII of Feller [14], implies that

$$\lim_{x \to +\infty} \frac{\mathbb{E}(\mathcal{N}_+(0, x))}{x} = \frac{\nu}{1 - \alpha\beta},$$

hence there exist D_1 and $D_2 \geq 0$ such that for all $x \geq 0$,

$$\mathbb{E}(\mathcal{N}_+(0, x)) \leq D_1 + D_2 x. \tag{5.9}$$

Hence, with Relation (5.7)

$$\mathbb{E}\left(\sum_{i=1}^{+\infty} h\left(S_i^+\right)\right) \leq D_1 h(0) + D_2 \int_0^{+\infty} x\mu(\mathrm{d}x)$$

$$\leq D_1 h(0) + D_2 \int_0^{+\infty} \mu(x, +\infty)\,\mathrm{d}x = D_1 h(0) + D_2 \alpha, \tag{5.10}$$

with the integration by parts formula.

For $p \geq 1$ and $n \geq p$,

$$\sum_{i=p}^{n} h\left(\sum_{j=2}^{i} \widetilde{Y}_j^n\right) \leq \sum_{i=p}^{n} h\left(S_i^+\right) \leq \sum_{i=p}^{+\infty} h\left(S_i^+\right),$$

this last term converges almost surely to 0 as p gets large, hence we obtain the almost sure convergence

$$\lim_{n \to +\infty} \sum_{i=1}^{n} h\left(\sum_{j=2}^{i} \widetilde{Y}_j^n\right) = \sum_{i=1}^{+\infty} h\left(\sum_{j=2}^{i} \widetilde{Y}_j^\infty\right) \leq \sum_{i=1}^{+\infty} h\left(S_i^+\right) < +\infty, a.s.$$

by Relations (5.6) and (5.9).

Step 3. Relation (5.5) with $k = 1$ gives

$$E_1 = \int_0^{\widetilde{Y}_1^1} \mathbb{1}_{\{u \leq \widetilde{Y}_1^n\}} \left(\nu + \beta \sum_{i=1}^{n} h\left(u + \sum_{j=2}^{i} \widetilde{Y}_j^n\right)\right)\,\mathrm{d}u,$$

since

$$\mathbb{1}_{\{u \leq \widetilde{Y}_1^n\}} \left(\nu + \beta \sum_{i=1}^{n} h\left(u + \sum_{j=2}^{i} \widetilde{Y}_j^n\right)\right) \leq \nu + \beta \sum_{i=1}^{n} h\left(\sum_{j=2}^{i} \widetilde{Y}_j^n\right)$$

$$\leq \nu + \beta \sum_{i=1}^{+\infty} h\left(S_i^+\right) < +\infty,$$

with the same argument as in the proof of the convergence (I_n), and with Lebesgue's dominated convergence theorem, we obtain the relation

$$E_1 = \int_0^{\widetilde{Y}_1^\infty} \left(\nu + \beta \sum_{i=1}^{+\infty} h \left(u + \sum_{j=2}^i \widetilde{Y}_j^\infty \right) \right) \mathrm{d}u.$$

With a standard argument, via Kolmogorov's extension Theorem, see Section VII-38 of Halmos [18], there exists a stationary sequence $(\widetilde{Z}_n^\infty, n \in \mathbb{Z})$ of random variables such that $(\widetilde{Z}_n^\infty, n \geq p) \overset{\text{dist.}}{=} (\widetilde{Y}_n^\infty, n \geq 1)$, holds for all $p \in \mathbb{Z}$. After integration of the last relation, with a change of variable, we obtain

$$1 - \nu \mathbb{E}\left(Y_1^\infty\right) = \beta \int_0^{+\infty} h\left(u\right) \sum_{i=1}^{+\infty} \mathbb{P}\left(\sum_{j=2}^i \widetilde{Y}_j^\infty \leq u \leq \sum_{j=1}^i \widetilde{Y}_j^\infty \right) \mathrm{d}u.$$

For $i \geq 1$, we have

$$\mathbb{P}\left(\sum_{j=2}^i \widetilde{Y}_j^\infty \leq u \leq \sum_{j=1}^i \widetilde{Y}_j^\infty \right) = \mathbb{P}\left(\sum_{j=-i+1}^{-1} \widetilde{Z}_j^\infty \leq u \leq \sum_{j=-i}^{-1} \widetilde{Z}_j^\infty \right)$$

$$= \mathbb{P}\left(\sum_{j=1}^{i-1} \widetilde{Z}_{-j}^\infty \leq u \leq \sum_{j=1}^i \widetilde{Z}_{-j}^\infty \right),$$

hence

$$1 - \nu \mathbb{E}\left(Y_1^\infty\right) = \beta \mathbb{P}\left(\int_0^{+\infty} h\left(u\right) \sum_{i=1}^{+\infty} \mathbb{1}_{\{\sum_{j=1}^{i-1} \widetilde{Z}_{-j}^\infty \leq u \leq \sum_{j=1}^i \widetilde{Z}_{-j}^\infty\}} \right) \mathrm{d}u = \beta \alpha,$$

by Fubini's Theorem. The proposition is proved. $\qquad \square$

Proposition 5.2. *The Markov chain with transition kernel $\mathcal{K}$ on $\mathcal{S}_h$ has a unique invariant measure.*

Proof. Let Λ a probability distribution on $\mathcal{S}_h$ invariant by $\mathcal{K}$, and $\mathcal{Z} = (Z_{-p}, p \geq 0)$ a random variable on $\mathcal{S}$ with distribution Λ. We construct the sample paths of Markov chains starting $(\mathcal{Z}_n) = (Z_n, Z_{n-1}, \ldots, Z_1) \cdot (Z_{-p}, p \geq 0)$ and $(\mathcal{Y}_n^0)$, from initial states $\mathcal{Z}$ and $(+\infty)$ respectively, in the following way, for $n \geq 1$,

$$\int_0^{Z_n} \left(\nu + \beta \left(\sum_{i=0}^n h\left(s + \sum_{j=i}^{n-1} Z_j \right) + \sum_{i=0}^{+\infty} h\left(s + \sum_{j=0}^i Z_{-j} \right) \right) \right) \mathrm{d}s = E_n,$$

and

$$\int_0^{Y_n} \left(\nu + \beta \sum_{i=0}^n h\left(s + \sum_{j=i}^{n-1} Y_j \right) \right) \mathrm{d}s = E_n.$$

It is easily seen by induction that, almost surely, for all $n \geq 1$, the relation $Z_n \leq Y_n^0$ holds, hence, since $\mathcal{Z}_n \overset{\text{dist.}}{=} \mathcal{Z}$, for any $k \geq 1$ and $n \geq k$,

$$(Z_0, \ldots, Z_{-k}) \leq_{\text{st}} (Y_n^0, \ldots, Y_{n-k}^0),$$

where $\leq_{st}$ denotes the stochastic order with respect to the order of the coordinates. Proposition 5.1 gives therefore

$$(Z_0, \ldots, Z_{-k}) \leq_{\text{st}} (Y_0^\infty, \ldots, Y_k^\infty).$$

Since (Z_n) is invariant for $\mathcal{K}$, the relation $\mathcal{T}(\mathcal{Z}, E_0) \stackrel{\text{dist.}}{=} Z_1$ holds and, as in the proof of Proposition 5.1 the relation $\mathbb{E}(Z_1) = (1 - \alpha\beta)/\nu$ also holds, and therefore, for all $k \geq 1$,

$$(Z_0, \ldots, Z_k) \stackrel{\text{dist.}}{=} (Y_0^\infty, \ldots, Y_k^\infty).$$

The proposition is proved. $\qquad\square$

6 Stationary Hawkes processes

We start with a coupling result between a general non-linear Hawkes process and the Hawkes process of Hawkes and Oakes [20]. The idea of the coupling with the "classical" Hawkes process is also mentioned in Karabash [25] but in a space of functions. The existence result of this reference seems to be incomplete since the topology used is not clearly defined and an important continuity property, the Feller property in fact, is apparently missing. This is the main technical difficulty of the proof of Theorem 6.2.

6.1 A monotonicity property

Throughout this section, (E_n) denotes an i.i.d. sequence of exponential random variables with parameter 1.

Let m be a point measure on $\mathbb{R}_-$, with Definition 4.2, the sequences $(X_n^m, n \geq 1)$ and $(Y_n^m, n \geq 1)$ are defined by induction by, for $n \geq 1$,

$$\int_0^{X_n^m} \Phi\left(\int_{-\infty}^0 h\left(s + \sum_{i=1}^{n-1} X_i^m - x\right) m(\mathrm{d}x) + \sum_{k=0}^{n-1} h\left(s + \sum_{i=n-k}^{n-1} X_i^m\right)\right) \mathrm{d}s = E_n, \quad (6.1)$$

$$\int_0^{Y_n^m} \Phi_{\beta_0}\left(\int_{-\infty}^0 h\left(s + \sum_{i=1}^{n-1} Y_i^m - x\right) m(\mathrm{d}x) + \sum_{k=0}^{n-1} h\left(s + \sum_{i=n-k}^{n-1} Y_i^m\right)\right) \mathrm{d}s = E_n, \quad (6.2)$$

where β_0 is, for the moment, a positive constant and $\Phi_{\beta_0}(x) \stackrel{\text{def.}}{=} \nu + \beta_0 x$ for $x \in \mathbb{R}$.

The affine function Φ_{β_0} will be used as follows. If β defined by Relation (1.2) is such that $\alpha\beta < 1$, then there exists some $\beta_0 > \beta$ such that $\alpha\beta_0 < 1$ and a constant ν such that

Proposition 6.1 (Coupling). *If ν and $\beta_0 > 0$ are such that the relation $\Phi(x) \leq \Phi_{\beta_0}(x)$, for all $x \geq 0$, with Φ_{β_0} defined by Relation (5.1), and if m be a point measure on $\mathbb{R}_-$ with*

$$\int_{-\infty}^0 h(-x)m(\mathrm{d}x) < +\infty,$$

then, for $\beta_0 > \beta$ there exists $\nu > 0$ such that, almost surely, the sequence (Y_n^m) defined by Relation (6.2) satisfies the relation $Y_n^m \leq X_n^m$, for all $n \geq 1$.

Proof. The proof is straightforward. If, by induction, the relation holds up to index $n - 1 > 0$. For $t \geq 0$,

$$\int_0^t \Phi\left(\int_{-\infty}^0 h\left(s + \sum_{i=1}^{n-1} X_i^m - x\right) m(\mathrm{d}x) + \sum_{k=0}^{n-1} h\left(s + \sum_{i=n-k}^{n-1} X_i^m\right)\right) \mathrm{d}s$$

$$\leq \int_0^t \Phi_{\beta_0}\left(\int_{-\infty}^0 h\left(s + \sum_{i=1}^{n-1} X_i^m - x\right) m(\mathrm{d}x) + \sum_{k=0}^{n-1} h\left(s + \sum_{i=n-k}^{n-1} X_i^m\right)\right) \mathrm{d}s$$

$$\leq \int_0^t \Phi_{\beta_0}\left(\int_{-\infty}^0 h\left(s + \sum_{i=1}^{n-1} Y_i^m - x\right) m(\mathrm{d}x) + \sum_{k=0}^{n-1} h\left(s + \sum_{i=n-k}^{n-1} Y_i^m\right)\right) \mathrm{d}s,$$

by the monotonicity properties of h and Φ_{β_0}. Hence we deduce that the relation $Y_n^m \leq X_n^m$ holds. $\qquad\square$

6.2 Existence theorem

We can now formulate our main existence result.

Theorem 6.2. *If the activation function function Φ is continuous with $\Phi(0) > 0$ and the quantities*

$$\alpha = \int_0^{+\infty} h(t)\,\mathrm{d}t \text{ and } \beta = \limsup_{y \to +\infty} \frac{\Phi(y)}{y}$$

are such that $\alpha\beta < 1$, then there exists a stationary Hawkes point process.

Proof. The assumptions imply that there exist $\nu > 0$ and $\beta_0 \geq 0$ such that $\alpha\beta_0 < 1$ and $\Phi(x) \leq \Phi_{\beta_0}(x)$ for all $x \geq 0$, where Φ_{β_0} is defined by Relation (5.1).

The Markov chain $(\mathcal{X}_n^0)$, resp. $(\mathcal{Y}_n^0)$, associated to Φ, resp. to Φ_{β_0}, with initial point $(+\infty)$, is denoted by

$$\mathcal{X}_n^0 = (X_n^0, X_{n-1}^0, \ldots, X_1^0, +\infty) \text{ and } \mathcal{Y}_n^0 = (Y_n^0, Y_{n-1}^0, \ldots, Y_1^0, +\infty),$$

see Definition 4.2.

We proceed in the spirit of Kryloff and Bogoliouboff [29], let $\overline{\mathbb{R}}_+ = \mathbb{R}_+ \cup \{+\infty\}$ and Q_n the distribution on $\overline{\mathbb{R}}_+^{\mathbb{N}}$ given by

$$Q_n(F) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}\left(F(X_i^0, X_{i-1}^0, \ldots, X_{i-k+1}^0)\right),$$

if F is a Borelian function on $\overline{\mathbb{R}}_+^{\mathbb{N}}$ depending only on the first k coordinates, with the convention that $X_i^0 = +\infty$ for $i \leq 0$. The space $\overline{\mathbb{R}}_+^{\mathbb{N}}$ is endowed with the topology defined by the distance (2.3).

The sequence of probability distributions (Q_n) on the compact space $\overline{\mathbb{R}}_+^{\mathbb{N}}$ is clearly tight. Let (Q_{n_p}) be convergent sequence whose limit is Q_∞. We now prove that Q_∞ is an invariant distribution for the Markov chain $(\mathcal{X}_n^0)$. Due to definition of (Q_n), it is clear that the probability Q_∞ on $\overline{\mathbb{R}}_+^{\mathbb{N}}$ is invariant by the shift operator $(x_i) \to (x_{i+1})$.

Let $\mathcal{Z}_p = (Z_1^p, Z_2^p, \ldots)$ be a random variable on $\overline{\mathbb{R}}_+^{\mathbb{N}}$ with distribution Q_{n_p} and E_0 an independent exponential random variable with parameter 1, $\mathcal{Z}_\infty = (Z_1^\infty, Z_2^\infty, \ldots)$ is a random variable with distribution Q_∞.

If F_1 is an exponential random variable independent of $\mathcal{Z}_p$, by definition of (Q_{n_p}) we have, for the convergence in distribution,

$$\lim_{p \to +\infty} (\mathcal{T}(\mathcal{Z}_p, F_1), Z_1^p, Z_2^p, \ldots) = \mathcal{Z}_\infty,$$

If we prove that, for the convergence in distribution,

$$\lim_{p \to +\infty} \mathcal{T}(\mathcal{Z}_p, E_0) = \mathcal{T}(\mathcal{Z}_\infty, E_0), \tag{6.3}$$

the probability distribution Q_∞ will be an invariant distribution for the Markov chain $(\mathcal{X}_n)$ and therefore that there exists a stationary Hawkes process by Proposition 4.1.

We fix $T > 0$, $p \geq 1$, we have, by definition of Q_{n_p}

$$\mathbb{P}(\mathcal{T}(\mathcal{Z}_p, E_0) \geq T) = \frac{1}{n_p} \sum_{i=1}^{n_p} \mathbb{E}\left(\exp\left(-\int_0^T \Phi\left(\sum_{k=0}^i h\left(s + \sum_{j=i+1-k}^i X_j^0\right)\right)\mathrm{d}s\right)\right). \tag{6.4}$$

Our proof is carried out in four steps.

Step 0: an integrability property of $\mathcal{Z}_\infty$.

We use the notations of Section 6.1. With Proposition 6.1, we can construct a coupling such that $Y_n^0 \leq X_n^0$ holds almost surely for all $n \geq K > 0$.

$$\sum_{k=0}^{K} h \left(\sum_{j=i+1-k}^{i} X_j^0 \right) \leq \sum_{k=0}^{i} h \left(\sum_{j=i+1-k}^{i} Y_j^0 \right),$$

we get,

$$\frac{1}{n_p} \sum_{i=K}^{n_p} \sum_{k=0}^{K} h \left(\sum_{j=i+1-k}^{i} X_j^0 \right) \leq \frac{1}{n_p} \sum_{i=K}^{n_p} \sum_{k=0}^{i} h \left(\sum_{j=i+1-k}^{i} Y_j^0 \right),$$

and, by taking the expected value, this gives the relation

$$\frac{1}{n_p} \sum_{i=K}^{n_p} \mathbb{E} \left(\sum_{k=0}^{K} h \left(\sum_{j=i+1-k}^{i} X_j^0 \right) \right) \leq \frac{1}{n_p} \sum_{i=K}^{n_p} \mathbb{E} \left(\sum_{k=0}^{i} h \left(\sum_{j=i+1-k}^{i} Y_j^0 \right) \right).$$

As p goes to infinity, the right-hand side is converging to $E(I)$, where I is the random variable defined in Proposition 5.1. The left-hand side can be expressed as

$$\mathbb{E} \left(\sum_{k=0}^{K} h \left(\sum_{j=1}^{k} Z_j^p \right) \right),$$

and this term is converging to

$$\mathbb{E} \left(\sum_{k=0}^{K} h \left(\sum_{j=1}^{k} Z_j^\infty \right) \right).$$

We have thus obtained that,

$$\mathbb{E} \left(\sum_{k=0}^{K} h \left(\sum_{j=1}^{k} Z_j^\infty \right) \right) \leq \mathbb{E}(I) < +\infty,$$

by letting K go to infinity, this gives the relation

$$\mathbb{E} \left(\sum_{k=0}^{+\infty} h \left(\sum_{j=1}^{k} Z_j^\infty \right) \right) < +\infty, \tag{6.5}$$

Step 1: A truncation argument.
For $i > 0$, by using the monotonicity property of h, we obtain the relations, for $s \geq 0$,

$$\left\{ \sum_{k=0}^{i} h \left(s + \sum_{j=i+1-k}^{i} X_j^0 \right) \geq C \right\} \subset \left\{ \sum_{k=0}^{i} h \left(\sum_{j=i+1-k}^{i} X_j^0 \right) \geq C \right\}$$

hence, with the couplings of Propositions 5.1 and 6.1 and the notations of its proof, and Relation (5.4), we obtain

$$\mathbb{P} \left(\inf \left(\sum_{k=0}^{i} h \left(s + \sum_{j=i+1-k}^{i} X_j^0 \right) : s \geq 0 \right) \geq C \right) \leq \mathbb{P} \left(\sum_{k=0}^{i} h \left(\sum_{j=i+1-k}^{i} X_j^0 \right) \geq C \right)$$

$$\leq \mathbb{P} \left(\sum_{k=0}^{i} h \left(\sum_{j=i+1-k}^{i} Y_j^0 \right) \geq C \right) = \mathbb{P} \left(\sum_{k=0}^{i} h \left(\sum_{j=1}^{k} \widetilde{Y}_j^i \right) \geq C \right)$$

$$\leq \mathbb{P} \left(\sum_{k=0}^{+\infty} h \left(\sum_{j=1}^{k} \widetilde{Y}_j^\infty \right) \geq C \right).$$

For $\varepsilon > 0$, Proposition 5.1 gives therefore the existence of $C_0 > 0$ such that

$$\sup_{i \geq 1} \mathbb{P} \left(\sup_{s \geq 0} \sum_{k=0}^{i} h \left(s + \sum_{j=i+1-k}^{i} X_j^0 \right) \geq C_0 \right) \leq \varepsilon. \tag{6.6}$$

For $\delta > 0$, the function Φ being uniformly continuous $[0, C_0]$, there exists some η_0 such that $|\Phi(x) - \Phi(y)| \leq \delta$, for all elements x and y of $[0, C_0]$ such that $|x - y| \leq \eta_0$.

Step 2.

For $s \geq 0$, $i > K \geq 0$, with the same arguments, we get

$$\left\{ \left| \Phi \left(\sum_{k=0}^{i} h \left(s + \sum_{j=i+1-k}^{i} X_j^0 \right) \right) - \Phi \left(\sum_{k=0}^{K} h \left(s + \sum_{j=i+1-k}^{i} X_j^0 \right) \right) \right| > \delta \right\}$$

$$\subset \left\{ \sup_{s \geq 0} \sum_{k=0}^{i} h \left(s + \sum_{j=i+1-k}^{i} X_j^0 \right) \geq C_0 \right\} \bigcup \left\{ \sum_{k=K+1}^{i} h \left(s + \sum_{j=i+1-k}^{i} X_j^0 \right) > \eta_0 \right\}$$

$$\subset \left\{ \sup_{s \geq 0} \sum_{k=0}^{i} h \left(\sum_{j=i+1-k}^{i} X_j^0 \right) \geq C_0 \right\} \bigcup \left\{ \sum_{k=K+1}^{i} h \left(\sum_{j=i+1-k}^{i} X_j^0 \right) > \eta_0 \right\},$$

and, as before,

$$\mathbb{P} \left(\sum_{k=K+1}^{i} h \left(\sum_{j=i+1-k}^{i} X_j^0 \right) > \eta_0 \right) \leq \mathbb{P} \left(\sum_{k=K+1}^{i} h \left(\sum_{j=i+1-k}^{i} Y_j^0 \right) > \eta_0 \right)$$

$$= \mathbb{P} \left(\sum_{k=K+1}^{i} h \left(\sum_{j=1}^{k} \widetilde{Y}_j^i \right) > \eta_0. \right) \leq \mathbb{P} \left(\sum_{k=K+1}^{i} h \left(\sum_{j=1}^{k} \widetilde{Y}_j^\infty \right) > \eta_0. \right)$$

$$\leq \mathbb{P} \left(\sum_{k=K+1}^{+\infty} h \left(\sum_{j=1}^{k} \widetilde{Y}_j^\infty \right) > \eta_0. \right).$$

The random variable I being almost surely finite, Proposition 5.1, gives that there exists K_0 such that

$$\mathbb{P} \left(\sum_{k=K_0+1}^{+\infty} h \left(\sum_{j=1}^{k} \widetilde{Y}_j^\infty \right) > \eta_0 \right) \leq \varepsilon.$$

By gathering these results, with Relation (6.6), we obtain

$$\sup_{i \geq K_0} \mathbb{P} \left(\sup_{s \geq 0} \left| \Phi \left(\sum_{k=0}^{i} h \left(s + \sum_{j=i+1-k}^{i} X_j^0 \right) \right) - \Phi \left(\sum_{k=0}^{K_0} h \left(s + \sum_{j=i+1-k}^{i} X_j^0 \right) \right) \right| > \delta \right) \leq 2\varepsilon,$$

hence

$$\lim_{K \to +\infty} \sup_{i \geq K} \left| \mathbb{E} \left(\exp \left(- \int_0^T \Phi \left(\sum_{k=0}^{i} h \left(s + \sum_{j=i+1-k}^{i} X_j^0 \right) \right) \mathrm{d}s \right) \right) \right.$$

$$\left. - \mathbb{E} \left(\exp \left(- \int_0^T \Phi \left(\sum_{k=0}^{K} h \left(s + \sum_{j=i+1-k}^{i} X_j^0 \right) \right) \mathrm{d}s \right) \right) \right| = 0. \tag{6.7}$$

Step 3.

Relations (6.4) and (6.7) show that, for $\varepsilon > 0$, there exists K_0 such that the relation

$$\left| \mathbb{P}(\mathcal{T}(\mathcal{Z}_p, E_0) \geq T) - \frac{1}{n_p} \sum_{i=K_0}^{n_p} \mathbb{E}\left(\exp\left(-\int_0^T \Phi\left(\sum_{k=0}^{K_0} h\left(s + \sum_{j=i+1-k}^{i} X_j^0 \right) \right) \mathrm{d}s \right) \right) \right| \leq \varepsilon$$

(6.8)

holds for all $p \geq 1$. Relation (6.5) shows that one can choose a constant K_0 sufficiently large so that

$$\left| \mathbb{E}\left(\exp\left(-\int_0^T \Phi\left(\sum_{k=0}^{K_0} h\left(s + \sum_{j=-k}^{1} Z_j^\infty \right) \right) \mathrm{d}s \right) \right) - \right.$$
$$\left. \mathbb{E}\left(\exp\left(-\int_0^T \Phi\left(\sum_{k=0}^{+\infty} h\left(s + \sum_{j=1}^{k} Z_j^\infty \right) \right) \mathrm{d}s \right) \right) \right| \leq \varepsilon.$$

With the convergence of the sequence (Q_{n_p}), we have

$$\lim_{p \to +\infty} \frac{1}{n_p} \sum_{i=K_0}^{n_p} \mathbb{E}\left(\exp\left(-\int_0^T \Phi\left(\sum_{k=0}^{K_0} h\left(s + \sum_{j=i+1-k}^{i} X_j^0 \right) \right) \mathrm{d}s \right) \right)$$
$$= \mathbb{E}\left(\exp\left(-\int_0^T \Phi\left(\sum_{k=0}^{K_0} h\left(s + \sum_{j=1}^{k} Z_j^\infty \right) \right) \mathrm{d}s \right) \right).$$

Note that the K_0 first terms in the expression of Q_{n_p} can be arbitrarily small as p gets large. With Relation (6.8), by letting ε go to 0, we obtain Relation (6.3). The theorem is proved. $\qquad \square$

6.3 Coupling properties

We have not been able to obtain an analogue, sufficiently strong, result for the uniqueness of a stationary Hawkes point process. Without a contracting scheme available, one of the few possibilities is of using a coupling argument of the trajectories of the Markov chain $(\mathcal{X}_n^z)$. In this section, we investigate the coupling of the Markov chain $(\mathcal{X}_n^z)$ with the Markov chain starting from the empty state.

We assume that the activation function Φ has the Lipschitz property,

$$|\Phi(x) - \Phi(y)| \leq L_\Phi |x - y|, \quad \forall x, y \geq 0,$$

(6.9)

for some constant $L_\Phi > 0$. Let f be a non-negative Borelian function on $\mathbb{R}_+$, and τ_f be the random variable such that the random variable

$$\int_0^{\tau_f} f(u) \, \mathrm{d}u$$

has an exponential distribution with parameter 1, i.e. for $t \geq 0$,

$$\mathbb{P}(\tau_f \geq t) = \exp\left(-\int_0^t f(u) \, \mathrm{d}u \right).$$

Lemma 6.3. *If f and g are non-negative integrable functions on $\mathbb{R}_+$, then there exists a coupling of τ_f and τ_g, such that*

$$\mathbb{P}(\tau_f \neq \tau_g) \leq \int_{\mathbb{R}_+} |f(u) - g(u)| \, \mathrm{d}u.$$

Proof. Define $h_f = (f - g)^+$, $h_g = (g - f)^+$, we can take the random variables τ_{h_f}, τ_{h_g} and $\tau_{f \wedge g}$ independent, then it is easily checked that

$$T_f = \tau_{h_f} \wedge \tau_{f \wedge g} \overset{\text{dist.}}{=} \tau_f \text{ and } T_g = \tau_{h_g} \wedge \tau_{f \wedge g} \overset{\text{dist.}}{=} \tau_g,$$

and

$$\mathbb{P}(T_f = T_g) \geq \mathbb{P}(\tau_{h_f} \wedge \tau_{h_g} \geq \tau_{f \wedge g}) = \mathbb{P}(\tau_{h_f + h_g} \geq \tau_{f \wedge g}) = \mathbb{P}(\tau_{|f-g|} \geq \tau_{f \wedge g}),$$

hence, we obtain,

$$\mathbb{P}(T_f \neq T_g) \leq \mathbb{E}\left(\exp\left(-\int_0^{\tau_{|f-g|}} f(u) \wedge g(u) \, \mathrm{d}u \right) \right)$$

$$= \int_0^{+\infty} (|f(v) - g(v)|) \exp\left(-\int_0^v f(u) \wedge g(u) + |f(v) - g(v)| \, \mathrm{d}u \right) \mathrm{d}v$$

$$\leq \int_0^{+\infty} |f(v) - g(v)| \, \mathrm{d}v.$$

The lemma is proved. $\qquad\square$

The next proposition gives an upper bound on the probability that the Markov chain $(\mathcal{X}_n^z)$ do not couple right from the beginning with the Markov chain starting from the empty state.

Proposition 6.4. *For $z \in \mathcal{S}_h$, if Φ is Lipschitz with constant L_Φ, there exists a coupling of the Markov chains $(\mathcal{X}_n^0)$ and $(\mathcal{X}_n^z)$ of Definition 4.2, and $D_1, D_2 \in \mathbb{R}_+$ such that*

$$\mathbb{P}\left(\exists n \geq 1, X_n^0 \neq X_n^z \right)$$

$$\leq L_\Phi \left(D_1 \int_{\mathbb{R}_-} \int_0^{+\infty} h(s - u) \, \mathrm{d}s \, m_z(\mathrm{d}u) + D_2 \int_{\mathbb{R}_-} \int_0^{+\infty} s h(s - u) \, \mathrm{d}s \, m_z(\mathrm{d}u) \right). \quad (6.10)$$

Proof. Since the random variables X_1^0 and X_1^z can be expressed as τ_{f_0} and τ_{f_y}, with

$$f_0(s) = \Phi(h(s)) \text{ and } f_y(s) = \Phi\left(h(s) + \int_0^{+\infty} h(s - u) m_z(\mathrm{d}u) \right),$$

Lemma 6.3 shows that

$$P(X_1^0 \neq X_1^z) \leq \int_0^{+\infty} \left| \Phi(h(s)) - \Phi\left(\int_{\mathbb{R}_-} h(s - u) m_z(\mathrm{d}u) + h(s) \right) \right| \mathrm{d}s$$

$$\leq L_\Phi \int_0^{+\infty} \int_{\mathbb{R}_-} h(s - u) m_z(\mathrm{d}u) \, \mathrm{d}s.$$

By induction, we obtain that, for $n \geq 1$,

$$P(X_{n+1}^0 \neq X_{n+1}^z | X_k^0 = X_k^z, \forall 1 \leq k \leq n)$$

$$\leq \int_0^{+\infty} \left| \Phi\left(\sum_{k=1}^{n+1} h\left(s + \sum_{i=k}^n X_i^z \right) \right) \right.$$

$$\left. - \Phi\left(\int_{\mathbb{R}_-} h\left(s + \sum_{i=1}^n X_i^0 - u \right) m_z(\mathrm{d}u) + \sum_{k=1}^{n+1} h\left(s + \sum_{i=k}^n X_i^0 \right) \right) \right| \mathrm{d}s$$

$$\leq L_\Phi \int_0^{+\infty} \int_{\mathbb{R}_-} h\left(s + \sum_{i=1}^n X_i^0 - u \right) m_z(\mathrm{d}u) \, \mathrm{d}s.$$

With Proposition 4.5, there exists some β_0 such that $\alpha\beta_0 < 1$ and

$$\sum_{i=1}^{n} X_i^0 \geq S_n^+, \text{ with } S_n \stackrel{\text{def.}}{=} \sum_{i=1}^{n} E_i - \alpha\beta_0,$$

hence

$$\mathbb{P}\left(\exists n \geq 1, X_n^0 \neq X_n^z\right) \leq L_\Phi \sum_{n=0}^{+\infty} \int_0^{+\infty} \int_{\mathbb{R}_-} \mathbb{E}\left(h\left(s + \sum_{i=1}^{n} X_i^0 - u\right)\right) m_z(\mathrm{d}u)\,\mathrm{d}s$$

$$\leq L_\Phi \sum_{n=0}^{+\infty} \int_0^{+\infty} \int_{\mathbb{R}_-} \mathbb{E}\left(h\left(s + S_n^+ - u\right)\right) m_z(\mathrm{d}u)\,\mathrm{d}s.$$

For $u > 0$, in the same way as in the proof of Proposition 5.1, with Fubini's Theorem, we obtain

$$\sum_{n=0}^{+\infty} \int_0^{+\infty} \mathbb{E}\left(h\left(s + S_n^+ + u\right)\right)\mathrm{d}s = \sum_{n=0}^{+\infty} \int_0^{+\infty} \mathbb{E}\left(h\left(s + u\right) \mathbb{1}_{\{s \geq S_n^+\}}\right)\mathrm{d}s$$

$$= \int_0^{+\infty} h\left(s + u\right) \mathbb{E}\left(\mathcal{N}_+([0, s])\right)\mathrm{d}s,$$

where $\mathcal{N}_+$ is defined by Relation (5.8), with Inequality (5.10) we obtain the desired estimate. $\qquad\square$

Corollary 6.5. *For $x_0 = (z_i, i \geq 1) \in \mathcal{S}_h$, if Φ is Lipschitz with constant L_Φ and*

$$\int_{\mathbb{R}_-} \int_0^{+\infty} s\,h\left(s - u\right)\mathrm{d}s\, m_{x_0}(\mathrm{d}u) < +\infty,$$

then there exists $K_0 > 0$ such that if $a \geq K_0$ and $x_1(a) = (a, z_2, z_3, \ldots)$, then

$$\mathbb{P}\left(X_n^0 = X_n^{x_1(a)}, \forall n \geq 1\right) > \frac{1}{2}.$$

Proof. It is enough to note that, by monotonicity of h, the right-hand side of Relation (6.10) goes to 0 if z_1 is replaced by a quantity a sufficiently large and, in particular less that $1/2$ if $a \geq K_0$, for some $K_0 > 0$. $\qquad\square$

7 The case of exponential memory

In this section we assume that the function h associated to the memory of previous jumps is exponentially decreasing.

$$h(u) = \exp(-u/\alpha),$$

for some $\alpha > 0$. In this case, the past activity of the Hawkes process can be encoded by a one-dimensional Markov process. One of the early analyses is Oakes [38]. Duarte, Aline et al. [11] considers a more general model for which h is the density of the sum of identically exponential random variables. The approach is of describing the Hawkes process in terms of a multi-dimensional Markov process, to encode the past activity of the Hawkes process.

In this section, we give a existence and uniqueness result of the stationary Hawkes chain with a weaker condition than the classical relation $\alpha L < 1$, where L is the Lipschitz constant associated to Φ. The result is obtained by using the Markov process of Section 4. At the same time an explicit representation of the distribution of the corresponding Palm

measure in terms of the invariant distribution of a one-dimensional Markov chain is obtained. We conclude this section with non-Lipschitz activation functions Φ for which the solution of the Hawkes SDE blows-up in finite time. A limit result gives a scaling description of how accumulation of the points of Hawkes process occurs in this case.

Proposition 7.1. *Let $x \in \mathcal{S}_h$, m_x defined by Relation (2.2) and $\mathcal{N}_{m_x} = (T_n)$ of Proposition 3.2, then*

$$(Z_n) \stackrel{\text{def.}}{=} \left(\int_{(-\infty, T_n]} h(T_n - s) \mathcal{N}_{m_x}(\mathrm{d}s) \right)$$
$$= \int_{(-\infty, 0]} \exp\left(-(T_n - u)/\alpha \right) m_x(\mathrm{d}u) + \sum_{k=1}^{n} \exp\left(-(T_n - T_k)/\alpha \right)$$

is a Markov chain on $(1, +\infty)$ such that, for $n \geq 0$,

$$Z_0 = \int_{(-\infty, 0]} e^{s/\alpha} m_x(\mathrm{d}s), \text{ and } Z_{n+1} = 1 + e^{-X_{n+1}/\alpha} Z_n, \tag{7.1}$$

where $X_{n+1} = T_{n+1} - T_n$ is the unique solution of the equation

$$\int_0^{X_{n+1}} \Phi\left(e^{-s/\alpha} Z_n \right) \mathrm{d}s = E_{n+1}, \tag{7.2}$$

where (E_n) are i.i.d. random variables with an exponential distribution with parameter 1.

Proof.

$$Z_{n+1} = \int_{(-\infty, T_{n+1}]} \exp(-(T_{n+1} - s)/\alpha) \mathcal{N}_{m_x}(\mathrm{d}s)$$
$$= 1 + \exp(-(T_{n+1} - T_n)/\alpha) \int_{(-\infty, T_n]} \exp(-(T_n - s)/\alpha) \mathcal{N}_{m_x}(\mathrm{d}s)$$
$$= 1 + \exp(-X_{n+1}/\alpha) Z_n,$$

and Relation (4.3) gives that

$$E_{n+1} = \int_{T_n}^{T_{n+1}} \Phi\left(\int_{\mathbb{R}_-} h(s - u) m_x(\mathrm{d}u) + \sum_{k=1}^{n} h\left(s - T_k \right) \right) \mathrm{d}s$$
$$= \int_0^{T_{n+1} - T_n} \Phi\left(e^{-s/\alpha} \left(\int_{\mathbb{R}_-} h(T_n - u) m_x(\mathrm{d}u) + \sum_{k=1}^{n} h\left(T_n - T_k \right) \right) \right) \mathrm{d}s$$
$$= \int_0^{X_{n+1}} \Phi\left(e^{-s/\alpha} Z_n \right) \mathrm{d}s$$

is an exponentially distributed random variable with parameter 1 and that the sequence (E_n) is i.i.d. $\qquad \square$

Proposition 7.2 (Harris ergodicity). *If Φ satisfies Condition A-2, and if $\alpha \beta_e < 1$, where*

$$\beta_e \stackrel{\text{def.}}{=} \limsup_{u \to +\infty} \int_{u-1}^{u} \frac{\Phi(s)}{s} \mathrm{d}s \tag{7.3}$$

then the sequence (Z_n) is a Harris Markov chain on $[1, +\infty)$.

For a general introduction on Harris Markov chains, see Nummelin [37].

Proof. The proof will be in two steps. We will first prove, via a Lyapunov function, that the set $[1, K]$ is recurrent, for some $K > 0$. Then, we will show that the subset $[1, 2]$ is also recurrent and *small*, see [37] for example. Proposition 5.10 of [37] gives then that the Markov chain (Z_n) is Harris ergodic.

We first exhibit a Lyapunov function for this Markov chain. Equation (7.2) can be rewritten as, if $Z_0 = z_0$,

$$E_1 = \int_0^{X_1} \phi\left(e^{-s/\alpha} z_0\right) \mathrm{d}s = \alpha \int_{z_0 e^{-X_1/\alpha}}^{z_0} \frac{\Phi(u)}{u} \mathrm{d}u = \alpha \int_{Z_1 - 1}^{z_0} \frac{\Phi(u)}{u} \mathrm{d}u. \qquad (7.4)$$

Let, for $y > 1$,

$$F(y) = \int_1^{y-1} \frac{\Phi(u)}{u} \mathrm{d}u.$$

Relation (7.2) give the identity

$$\int_{z_0}^{z_0 \exp(-X_1/\alpha)} \frac{\Phi(u)}{u} \mathrm{d}u = \frac{E_0}{\alpha},$$

and consequently,

$$\mathbb{E}_{z_0}(F(Z_1) - F(z_0)) = \int_{z_0 - 1}^{z_0} \frac{\Phi(u)}{u} \mathrm{d}u - \frac{1}{\alpha},$$

hence there exist $\eta > 0$ and $K > 0$ such that if $z_0 \geq K$ then

$$\mathbb{E}_{z_0}(F(Z_1)) - \mathbb{E}(F(z_0)) \leq \left(\beta_e + \eta - \frac{1}{\alpha}\right) < 0.$$

The function F is a Lyapunov function for the Markov chain (Z_n). The interval $[1, K]$ is a *recurrent set* for the Markov chain. See Theorem 8.6 of Robert [42].

By using Relation (7.4), we obtain, for $z_0 \in [1, K]$,

$$\mathbb{P}_{z_0}(Z_1 < 2) = \mathbb{P}\left(\frac{E_1}{\alpha} > \int_1^{z_0} \frac{\Phi(u)}{u} \mathrm{d}u\right) = \exp\left(-\alpha \int_1^{z_0} \frac{\Phi(u)}{u} \mathrm{d}u\right)$$

$$\geq \exp\left(-\alpha \int_1^K \frac{\Phi(u)}{u} \mathrm{d}u\right) > 0,$$

the interval $[1, 2]$ is a recurrent set for the Markov chain (Z_n).

For $0 < t \leq z_0$, the relation

$$\mathbb{P}(Z_1 - 1 \leq t) = \exp\left(-\alpha \int_t^{z_0} \frac{\Phi(u)}{u} \mathrm{d}u\right)$$

gives that the density of $Z_1 - 1$ is given by, for $z_0 \leq K$,

$$\alpha \frac{\Phi(t)}{t} \exp\left(-\alpha \int_t^{z_0} \frac{\Phi(u)}{u} \mathrm{d}u\right) \geq \alpha \frac{\Phi(t)}{t} \exp\left(-\alpha \int_t^K \frac{\Phi(u)}{u} \mathrm{d}u\right).$$

There is a positive lower bound independent of $z_0 \leq K$. We can now use the same argument as in example of Section 4.3.3 page 98 of Meyn and Tweedie [35] to prove that $[1, 2]$ is a small set. The proposition is proved. $\qquad \square$

Definition 7.3. *For $z > 1$ and $y > 0$, we define $G_\Phi(z, y)$ by the relation*

$$\int_0^{G_\Phi(z,y)} \Phi\left(e^{-s/\alpha} z\right) \mathrm{d}s = y. \qquad (7.5)$$

Theorem 7.4. *[Invariant distribution of $(\mathcal{X}_n)$] If $\alpha\beta_e < 1$, then for any $x \in \mathcal{S}_h$, the Markov chain $(\mathcal{X}_n)$ of Definition 4.2 converges in distribution to the law of*

$$\left(G_\Phi \left(Z^*_{-n}, \alpha \int_{Z^*_{-n+1}-1}^{Z^*_{-n}} \frac{\Phi(u)}{u} \, \mathrm{d}u \right), n \geq 1 \right)$$

where G_Φ is defined by Relation (7.5) and $(Z^*_n, n \in \mathbb{Z})$ is the stationary version of the Harris Markov chain (Z_n) of Proposition 7.1.

Proof. For $n \geq 1$, with the above notation, by definition $X_{n+1} = G_\Phi(Z_n, E_{n+1})$, i.e.

$$E_{n+1} = \alpha \int_{Z_{n+1}-1}^{Z_n} \frac{\Phi(u)}{u} \, \mathrm{d}u.$$

For $n \geq k \geq 1$, $(X_n, X_{n-1}, \ldots, X_{n-k+1})$ are the k-first coordinates of $\mathcal{X}_n$, they can be expressed as

$$(G_\Phi(Z_{n-1}, E_n), G_\Phi(Z_{n-2}, E_{n-1}), \ldots, G_\Phi(Z_{n-k}, E_{n-k+1}))$$
$$= \left(G_\Phi \left(Z_{n-1}, \alpha \int_{Z_n-1}^{Z_{n-1}} \frac{\Phi(u)}{u} \, \mathrm{d}u \right), \ldots G_\Phi \left(Z_{n-k-1}, \alpha \int_{Z_{n-k}-1}^{Z_{n-k-1}} \frac{\Phi(u)}{u} \, \mathrm{d}u \right) \right).$$

The Harris ergodicity of (Z_n) implies that the random variable $(Z_n, Z_{n-1}, \ldots, Z_{n-k})$ is converging in distribution to $(Z^*_0, Z^*_{-1}, \ldots, Z^*_{-k})$. The mapping $(z, y) \mapsto G_\Phi(z, y)$ is continuous, the continuous mapping theorem concludes the proof of our result. $\qquad\square$

Theorem 7.4 shows that in the case of exponential memory, the invariant distribution of $(\mathcal{X}_n)$ can be expressed in terms of a one-dimensional stationary Markov chain. The following corollary rephrases this result in terms of Hawkes processes. This is a direct application of Proposition 4.4.

Corollary 7.5. *If Φ is a continuous function such that $\Phi(0) > 0$ and $\alpha\beta_e < 1$, where β_e is defined by Relation (7.3), then there exists a unique stationary Hawkes process.*

Note that the condition $\alpha\beta_e < 1$ is weaker than the classical conditions of the literature: Φ Lipschitz with Lipschitz constant β such that $\alpha\beta < 1$.

Transient Hawkes processes

From now on, we assume a polynomial behavior for Φ so that, for $x > 0$,

$$\Phi(x) = (\nu + \beta x)^\gamma,$$

where ν, β and γ are positive real numbers.

Theorem 7.4 shows that $(\mathcal{X}_n)$ is converging in distribution for all α, ν and β when $\gamma < 1$, and, when $\gamma = 1$, the convergence occurs if $\alpha\beta < 1$.

Proposition 7.6. *If $\Phi(u) = (\nu + \beta u)^\gamma$ and if $\gamma > 1$ with $\beta, \nu > 0$, then the Markov process (Z_n) is transient.*

Proof. From Relation (7.4), we have

$$E_1 = \alpha \int_{Z_1-1}^{z_0} \frac{\Phi(u)}{u} \, \mathrm{d}u,$$

where E_1 is an exponentially distributed random variable with parameter 1. Let, for $u \geq 0$, $\Phi_p(u) \stackrel{\text{def.}}{=} (\beta u)^\gamma$ and, on the event $\{\gamma E_1 < \alpha\beta^\gamma z_0^\gamma\}$, we define the variable $Z_{1,p}$ such that

$$E_1 = \alpha \int_{Z_{1,p}-1}^{z_0} \frac{\Phi_p(u)}{u} \, \mathrm{d}u = \frac{\alpha\beta^\gamma}{\gamma} \left(z_0^\gamma - (Z_{1,p}-1)^\gamma \right),$$

then, since $\Phi_p \leq \Phi$, we have $z_0 - Z_{1,p} + 1 \geq z_0 - Z_1 + 1 \geq 0$ and $E_2 \stackrel{\text{def.}}{=} \gamma E_1/(\alpha\beta^\gamma)$, then

$$z_0 - Z_{1,p} + 1 = z_0 \left(1 - \left(1 - \frac{E_2}{z_0^\gamma} \right)^{1/\gamma} \right).$$

The elementary inequality

$$1 - (1-h)^{1/\gamma} \leq \frac{2}{\gamma}h, \quad 0 \leq h \leq 1/2,$$

gives the relation, for $a \geq 1$

$$\mathbb{E}_{z_0} \left((z_0 - Z_1 + 1)^a \right) \leq z_0^a \mathbb{P}(2\gamma E_1 \geq \alpha\beta^\gamma z_0^\gamma) + \mathbb{E} \left((z_0 - Z_{1,p} + 1)^a \mathbb{1}_{\{2\gamma E_1 \leq \alpha\beta^\gamma z_0^\gamma\}} \right)$$
$$\leq z_0^a \exp \left(-\frac{\alpha\beta^\gamma}{2\gamma} z_0^\gamma \right) + \frac{2^a}{\alpha^a \beta^{a\gamma}} \frac{\mathbb{E}\left(E_1^a\right)}{z_0^{a(\gamma-1)}}.$$

Since $\gamma > 1$, we deduce that

$$\lim_{z_0 \to +\infty} \mathbb{E}_{z_0}\left(Z_1 - z_0\right) = 1 \text{ and } \sup_{z_0 \geq 1} \mathbb{E}_{z_0}\left((Z_1 - z_0)^2\right) < +\infty. \tag{7.6}$$

Theorem 8.10 of Robert [42] shows that the Markov chain is transient. Strictly speaking Theorem 8.10 is for a Markov chain with a countable state space, nevertheless a glance at the proof of this result shows that it is also valid in our setting. The proposition is proved. $\square$

Relation (7.2) of Proposition 7.1 gives that the Hawkes Point Process (T_n) of Proposition 3.2 is such that

$$E_{n+1} \geq \beta^\gamma Z_n^\gamma \int_0^{T_{n+1}-T_n} e^{-s\gamma/\alpha}\,\mathrm{d}s = \frac{\alpha}{\gamma}\beta^\gamma Z_n^\gamma \left(1 - e^{-\gamma(T_{n+1}-T_n)/\alpha} \right), \tag{7.7}$$

where (E_n) is an i.i.d. sequence of exponential random variables with parameter 1. Under the assumptions of Proposition 7.6 the sequence (Z_n) is converging in distribution to infinity and, with the last relation, the relation

$$\lim_{n \to +\infty} T_{n+1} - T_n = 0$$

holds for the convergence in law. This result suggests that the points (T_n) are closer and closer asymptotically. We investigate this aspect in the rest of the section.

We now study the asymptotic behavior of (Z_n) in the transient case. We start with a technical lemma.

Lemma 7.7. *If $\gamma \geq 2$ then, for any $\delta > 0$,*

$$\sup_{z_0 > 1} \mathbb{E}_{z_0}\left(e^{\delta|Z_1 - z_0|}\right) < +\infty. \tag{7.8}$$

Proof. From Relation (7.4), we get, for $z_0 \geq 1$ and $1 \leq t \leq z_0$,

$$\mathbb{P}_{z_0}\left(z_0 - Z_1 + 1 \geq t\right) = \mathbb{P}\left(E_1 \geq \alpha \int_{z_0-t}^{z_0} \frac{\Phi(u)}{u}\right)$$
$$\leq \exp\left(-bz_0^\gamma \left(1 - \left(1 - \frac{t}{z_0} \right)^\gamma \right) \right), \quad (7.9)$$

with $b \stackrel{\text{def.}}{=} \alpha\beta^\gamma/\gamma$. Note that, by Definition (7.1) of Z_1, $z_0 - Z_1 + 1 \geq 0$, hence, for $\delta > 0$,

$$\mathbb{E}_{z_0}\left(e^{\delta|Z_1 - z_0 - 1|}\right) - 1 = \delta \int_0^{z_0} e^{\delta t} \mathbb{P}_{z_0}\left(z_0 - Z_1 + 1 \geq t\right) \mathrm{d}t$$

$$\leq \delta \int_0^{z_0} \exp\left(\delta t - bz_0^\gamma\left(1 - \left(1 - \frac{t}{z_0}\right)^\gamma\right)\right) \mathrm{d}t$$

$$= \delta z_0 \int_0^1 \exp\left(\delta u z_0 - bz_0^\gamma\left(1 - (1 - u)^\gamma\right)\right) \mathrm{d}u.$$

When z_0 is sufficiently large, we split the integral into two terms,

$$z_0 \int_{1-(1-2\delta/(bz_0^{\gamma-1}))^{1/\gamma}}^1 \exp\left(\delta u z_0 - bz_0^\gamma\left(1 - (1 - u)^\gamma\right)\right) \mathrm{d}u \leq z_0 \exp\left(-\delta z_0\right),$$

and

$$z_0 \int_0^{1-(1-2\delta/(bz_0^{\gamma-1}))^{1/\gamma}} \exp\left(\delta u z_0 - bz_0^\gamma\left(1 - (1 - u)^\gamma\right)\right) \mathrm{d}u$$

$$\leq z_0 \left(1 - \left(1 - 2\frac{\delta}{bz_0^{\gamma-1}}\right)^{1/\gamma}\right) \exp\left(\delta z_0\left(1 - (1 - 2\delta/(bz_0^{\gamma-1}))^{1/\gamma}\right)\right).$$

The lemma is proved. □

Proposition 7.8. *If $\gamma \geq 2$, then, almost surely,*

$$\lim_{n \to +\infty} \frac{Z_n}{n} = 1.$$

Proof. This is a consequence of Relations (7.6) and (7.8). Theorem 8.11 of Robert [42] shows that, almost surely, the relation

$$\liminf_{n \to +\infty} \frac{Z_n}{n} \geq 1$$

holds. Condition b) of Theorem 8.11 of Robert [42] follows from Lemma 7.7 and Condition c) of this proposition is replaced in this context by the fact that if $z_0 < K_0$, then there exists $n_0 \geq 1$ such that $\mathbb{P}_{z_0}(Z_{n_0} \geq K_0) > 0$.

Notice that $Z_n \leq n + z_0$ holds for all $n \geq 1$, we get therefore that almost surely

$$\lim_{n \to +\infty} \frac{Z_n}{n} = 1.$$

The proposition is proved. □

For $x \in \mathcal{S}_h$ and m_x defined by Relation (2.2), if $\mathcal{N}_{m_x} = (T_n)$ the point process of Proposition 3.2, we know that the sequence $(T_{n+1} - T_n)$ is converging in distribution to 0. The following proposition gives a much more detailed description of the accumulation of points:

For $n \geq 1$, the point process $(T_n - T_k, 1 \leq k \leq n)$, the point process seen from the nth point, scaled by the factor n^γ converges in distribution to a Poisson point process.

Theorem 7.9 (Asymptotic behavior of points of a transient Hawkes process)**.** *Assume that $\Phi(u) = (\nu + \beta u)^\gamma$ with $\gamma \geq 2$ and β, $\nu > 0$, if $x \in \mathcal{S}_h$, m_x defined by Relation (2.2) and $\mathcal{N}_{m_x} = (T_n)$ the point process of Proposition 3.2, then the point process*

$$(n^\gamma(T_n - T_k), 1 \leq k < n)$$

converges in distribution to a Poisson process with rate β^γ.

Proof. Define $\mathcal{P}_n = (n^\gamma(T_n - T_k), 1 \leq k < n)$. Relation (7.2) and the representation of (T_n) in terms of the Markov chain (Z_n) of Proposition 7.1 give the identity

$$E_{n+1} = \int_0^{X_{n+1}} \left(\nu + \beta Z_n e^{-s/\alpha}\right)^\gamma \mathrm{d}s, \tag{7.10}$$

where (E_n) is an i.i.d. sequence of exponential random variables with parameter 1 and, as before, $X_{n+1} = T_{n+1} - T_n$.

For any $\delta > 0$ and $n \geq 1$, Relation (7.10) gives the inequality

$$\mathbb{P}(X_{n+1} \geq \delta) \leq \mathbb{P}\left(\frac{\alpha}{\gamma}\beta^\gamma\left(1 - e^{-\delta\gamma/\alpha}\right)Z_n^\gamma \leq E_{n+1}\right). \tag{7.11}$$

On the event $\mathcal{E}_{n,\delta}$,

$$\mathcal{E}_{n,\delta} \stackrel{\text{def.}}{=} \left\{X_{n+1} \leq \delta, \frac{\nu}{\beta}e^{\delta/\alpha} \leq Z_n\right\}.$$

Relation (7.10) shows that

$$\beta^\gamma e^{-\delta\gamma/\alpha}Z_n^\gamma X_{n+1} \leq E_{n+1},$$

and therefore that the sequence of random variables $(Z_n^\gamma X_{n+1})$ is therefore tight.

The elementary relation

$$(1 + h)^\gamma - 1 \leq C_1 h, \quad 0 \leq h \leq 1,$$

with $C_1 = \gamma 2^{\gamma-1}$ and Relation (7.10) give the following inequality

$$0 \leq E_{n+1} - \beta^\gamma Z_n^\gamma \int_0^{X_{n+1}} e^{-s\gamma/\alpha}\,\mathrm{d}s$$

$$\leq C_1\nu\left(\beta Z_n\right)^{\gamma-1}\int_0^{X_{n+1}} e^{-(\gamma-1)s/\alpha}\,\mathrm{d}s \leq C_1\nu\beta^{\gamma-1}X_{n+1}Z_n^{\gamma-1},$$

on the event $\mathcal{E}_{n,\delta}$. Using the fact that $(Z_n^\gamma X_{n+1})$ is tight and the almost sure convergence of (Z_n/n) to 1 of Proposition 7.8 shows that the random variables $(X_{n+1}Z_n^{\gamma-1})$ is converging in distribution to 0. For a sufficiently large n, Relation (7.11) shows that the probability of the event $\mathcal{E}_{n,\delta}$ is arbitrarily close to 1, we obtain that the sequence of random variables

$$(Y_n) \stackrel{\text{def.}}{=} \left(\beta^\gamma Z_n^\gamma \int_0^{X_{n+1}} e^{-s\gamma/\alpha}\,\mathrm{d}s\right)$$

is converging in law to an exponential distribution with parameter 1.

On the event $\mathcal{E}_{n,\delta}$, we have

$$Y_n \leq \beta^\gamma Z_n^\gamma X_{n+1} \leq e^{\delta\gamma/\alpha}Y_n,$$

hence, for $x \geq 0$,

$$\mathbb{P}(\beta^\gamma Z_n^\gamma X_{n+1} \geq x) \leq \mathbb{P}\left(e^{\delta\gamma/\alpha}Y_n \geq x\right) + \mathbb{P}\left(\mathcal{E}_{n,\delta}^c\right),$$

hence

$$\limsup_{n \to +\infty} \mathbb{P}\left(\beta^\gamma Z_n^\gamma X_{n+1} \geq x\right) \leq \exp\left(-xe^{-\delta\gamma/\alpha}\right),$$

and, by letting δ go to 0, we obtain

$$\limsup_{n \to +\infty} \mathbb{P}\left(\beta^\gamma Z_n^\gamma X_{n+1} \geq x\right) \leq e^{-x}.$$

An analogue lower bound holds for the $\liminf$, hence $(\beta^\gamma Z_n^\gamma X_{n+1})$ is converging in distribution to an exponential distribution with parameter 1. By using the almost sure convergence of (Z_n/n) to 1, the same property holds for $(\beta^\gamma n^\gamma X_{n+1})$.

For $p \geq 1$, the same argument may be used to prove that the sequence of random variables $(\beta^\gamma n^\gamma X_{n-k}, 0 \leq k \leq p-1)$ is converging in distribution to the product of p exponential distributions with parameter 1. The key argument is again the almost sure convergence of (Z_n/n) to 1.

Final step. Let f be a continuous function on $\mathbb{R}_+$ with compact support included in $[0, K_0]$ for some $K_0 > 0$. The relations $Z_n \leq n + z_0$ and (7.10) give the inequality, for all $n \geq 1$,

$$E_{n+1} \leq (\nu + \beta(n + z_0))^\gamma X_{n+1} \leq C_0 (n+1)^\gamma X_{n+1},$$

for some constant C_0 independent of n.

For $p \leq n$, we thus have

$$\mathbb{P}(n^\gamma(T_n - T_{n-p}) \leq K_0) \leq \mathbb{P}\left(\sum_{k=n-p+1}^{n} k^\gamma X_k \leq K_0\right) \leq \mathbb{P}\left(\frac{1}{C_0}\sum_{k=1}^{p} E_k \leq K_0\right).$$

We denote $\mathcal{N}_{\beta^\gamma}$ a Poisson process with rate β^γ, with associated exponential random variables $(E_n^{\beta^\gamma})$.

The last inequality gives that $\mathcal{P}_n$ is stochastically dominated by $\mathcal{N}_{C_0 + \beta^\gamma}$, a Poisson process with rate $C_0 + \beta^\gamma$, in the sense that

$$\mathbb{P}(\mathcal{P}_n((0, K_0]) \geq p) \leq \mathbb{P}(\mathcal{N}_{C_0 + \beta^\gamma}((0, K_0]) \geq p), \text{ for } p \leq n. \tag{7.12}$$

Clearly $\mathcal{N}_{\beta^\gamma}$ is also stochastically dominated by $\mathcal{N}_{C_0 + \beta^\gamma}$.

$$\left|\mathbb{E}\left(\exp\left(-\int f(u)\mathcal{P}_n(u)\right)\right) - \mathbb{E}\left(\exp\left(-\int f(u)\mathcal{N}_{\beta^\gamma}(u)\right)\right)\right|$$
$$\leq \left|\mathbb{E}\left(\exp\left(-\sum_{k=n-p+1}^{n} f(n^\gamma(T_n - T_{n-k}))\right)\right) - \mathbb{E}\left(\exp\left(-\sum_{k=n-p+1}^{n} f\left(E_k^{\beta^\gamma}\right)\right)\right)\right|$$
$$+ 2\mathbb{P}(\mathcal{N}_{C_0 + \beta^\gamma}((0, K_0]) \geq p) + 2\mathbb{P}(\mathcal{N}_{\beta^\gamma}((0, K_0]) \geq p),$$

with Relation (7.12). By using the convergence in distribution of the random variables $(\beta^\gamma n^\gamma X_{n-k}, 0 \leq k \leq p-1)$, the first term of the right-hand side of this inequality can be made arbitrarily small for n sufficiently large. We can take an integer p independent of n sufficiently large such that the last two terms can be made arbitrarily small.

We have thus proved that

$$\lim_{n \to +\infty} \mathbb{E}\left(\exp\left(-\int_{\mathbb{R}_+} f(u)\mathcal{P}_n(u)\right)\right) = \mathbb{E}\left(\exp\left(-\int_{\mathbb{R}_+} f(u)\mathcal{N}_{\beta^\gamma}(u)\right)\right)$$

holds for all continuous functions with compact support. We can use Theorem 3.2.6 of Dawson [10] to conclude the proof of the proposition. $\qquad\square$

A General results on point processes

Point processes

We recall the notations on point measures and the associated random variables, the point processes used throughout this paper. See Neveu [36], Chapter 1 and 11 of Robert [42], and Dawson [10] for a general introduction on random measures.

Stationary point processes

If m is a simple point measure on $\mathbb{R}$, the points of m are enumerated by an increasing sequence $(t_k(m), k \in \mathbb{Z})$, numbered so that the relations

$$t_{-1}(m) < t_0(m) \le 0 < t_1(m) < \cdots \tag{A.1}$$

hold, with the convention that $t_k(m) = +\infty$ if there are less than $k \ge 1$ points of m in $\mathbb{R}_+$, and similarly on $\mathbb{R}_-$. The flow of translation operators (θ_t) on $\mathcal{M}_p(\mathbb{R})$ is defined by, for $t \in \mathbb{R}$ and $m \in \mathcal{M}_p(\mathbb{R})$,

$$\int_{\mathbb{R}} f(s)\theta_t(m)(\mathrm{d}s) \stackrel{\text{def.}}{=} \int_{\mathbb{R}} f(s-t)m(\mathrm{d}s), \tag{A.2}$$

for any non-negative Borelian function f on $\mathbb{R}$.

A distribution on $\mathcal{M}_p(H)$, an element of the set $\mathcal{P}(\mathcal{M}_p(H))$, is defined as *a point process* on H.

Definition A.1 (Stationarity). *A point process $\mathcal{N}$ on $\mathbb{R}$ is* stationary *with intensity $\lambda > 0$, if the random variable $\mathcal{N}([0,1])$ is integrable and $\mathbb{E}(\mathcal{N}([0,1])) = \lambda$, and if its distribution is invariant by translation, i.e. for $t \in \mathbb{R}$, $\theta_t(\mathcal{N}) \stackrel{\text{dist.}}{=} \mathcal{N}$, where θ_t is the translation operator defined by Relation* (A.2)*.*

Palm space of point processes

The set $\mathcal{M}_p^0(\mathbb{R})$ is a subset of elements m of $\mathcal{M}_p(R)$ such that $m(0) \ne 0$. If $m \in \mathcal{M}_p^0(\mathbb{R})$ then m can be represented either by the non-decreasing sequence $(t_k, k \in \mathbb{Z})$ of its points, or by the sequence

$$x = (x_k) = (t_{-k} - t_{-k-1}, k \ge 0)$$

of increments between them, with the convention that $t_0 = 0$.

An operator $\widehat{\theta}$ on $\mathcal{M}_p^0(\mathbb{R})$ is defined by,

$$\widehat{\theta}(m) \stackrel{\text{def.}}{=} \theta_{t_1(m)}(m)\mathbb{1}_{\{t_1(m) < +\infty\}}, \tag{A.3}$$

for $m \in \mathcal{M}_p^0(\mathbb{R})$, where $t_1(m)$ and (θ_t) by defined respectively by Relations (A.1) and (A.2). A simple point process of $\mathcal{M}_p^0(\mathbb{R})$ can be identified to its sequence of inter-arrival times and it is easily seen that the relation

$$\Big((t_{k+1} - t_k)(\widehat{\theta}(m)), k \in \mathbb{Z} \Big) = \Big((t_{k+2} - t_{k+1})(m), k \in \mathbb{Z} \Big),$$

holds.

The mapping $\widehat{\theta}$ is the classical shift operator on sequences. If $m = (t_n, n \in \mathbb{Z})$ is in $\mathcal{M}_p^0(\mathbb{R})$ and $x = (x_n) = (t_n - t_{n-1}, n \in \mathbb{Z})$, then $m = m_x$ and $\widehat{\theta}(m) = m_{\bar{x}}$, with $\bar{x} = (x_{n+1})$ if $m(\mathbb{R}_-) = m(\mathbb{R}_+) = +\infty$.

Equivalence between stationary point processes and Palm measure

We now recall some classical results on stationary point processes on $\mathbb{R}$. A stationary simple point process with intensity $\lambda > 0$ can be equivalently defined by either by

    a. a distribution Q on $\mathcal{M}_p(\mathbb{R})$ which is invariant for the continuous flow of translations (θ_t);

    b. a distribution $\widehat{Q}$ on $\mathcal{M}_p^0(\mathbb{R})$ called the *Palm measure* of Q which is invariant for the operator $\widehat{\theta}$.

When a) is given, a distribution $\widehat{Q}$ on $\mathcal{M}_p^0(\mathbb{R})$ is constructed via Mecke's Formula so that property b) holds. See Chapter II of Neveu [36] or Proposition 11.6 of Robert [42].

If b) holds, i.e. if $\widehat{Q}$ is given, then the construction of Q is done with a fundamental construction of ergodic theory, it is the *special flow* associated to the operator $\widehat{\theta}$ and the function $m \mapsto t_1(m)$ on $\mathcal{M}_p^0$. See Chapter 11 of Cornfeld et al. [6], and Chapter 10 of Robert [42]. The distribution Q is expressed as

$$\int_{\mathcal{M}_p(\mathbb{R})} F(m)\, Q(\mathrm{d}m) = \lambda \int_{\mathcal{M}_p^0(\mathbb{R})} \int_0^{t_1(m)} F(\theta_s(m))\, \mathrm{d}s\, \widehat{Q}(\mathrm{d}m), \tag{A.4}$$

for any non-negative Borelian function on $\mathcal{M}_p(\mathbb{R})$. The probability distribution $\widehat{Q}$ is simply determined by a distribution of the sequence of inter-arrival times which is invariant by the shift operator. The space $(\mathcal{M}_p^0(\mathbb{R}), \widehat{\theta}, \widehat{Q})$ can be seen as a probability space whose elements are positive sequences. It is sometimes called the Palm space of Q.

B   Hawkes processes: a quick review

Hawkes processes have been introduced by Hawkes in 1974 in Hawkes [19] as a class of point processes $\mathcal{N}$, whose stochastic intensity $(\lambda(t))$ depends on previous jumps, i.e through,

$$(\lambda(t)) = \left(\Phi\left(\int_{(-\infty, t)} h(t-s)\mathcal{N}(\mathrm{d}s) \right) \right).$$

The first Hawkes processes investigated were restricted to affine activation function Φ of the form,

$$\Phi(x) = \nu + \beta x,$$

which have a nice representation in terms of age-dependent branching processes, see Lewis [32], Vere-Jones [47], and Daley and Vere-Jones [8]. The condition of existence and uniqueness of stationary Hawkes process in this case is

$$\beta \int_0^{+\infty} h(s)\, \mathrm{d}s < 1,$$

see Hawkes and Oakes [20].

The special case where $\nu = 0$ was investigated in Brémaud and Massoulié [3], where a particular interest was dedicated to the critical Hawkes process,

$$\beta \int_0^{+\infty} h(s)\, \mathrm{d}s = 1.$$

The same critical Hawkes process, with general immigration rate ν is investigated in Kirchner [28]. A more precise study of the statistics of stationary Hawkes processes is developed in Jovanović [24] using a Poisson cluster process representation. The addition of an external jump process to the linearized Hawkes process, in view of applications in neuroscience example, has been considered in Boumezoued [2].

Hawkes processes with exponential functions $h(x) = \exp(-x/\alpha)$ have attracted a particular interest because the associated counting process has the Markovian property, see Oakes [38].

Having an affine activation function Φ is very helpful because of the corresponding branching process representation, However, when studying auto-inhibiting point processes, non-linear activation functions Φ are natural candidates to use. In this setting the investigation of sufficient conditions for the existence of stationary versions is more

Table 1: Existence of stationary Hawkes processes.

Branching Processes	
References	Hawkes and Oakes [20], Lewis [32], and Vere-Jones [47]
Assumptions 1	
a. $\Phi(x) = \nu + \beta x$ with $\nu > 0$ and $\beta \geq 0$	Oakes [38], and Errais et al. [12]
b. $h(x) = \exp(-x/\alpha)$	
Assumptions 2	
a. $\Phi(x) = \nu + \beta x$ with $\nu > 0$ and $\beta \geq 0$	Hawkes [19], Hawkes and Oakes [20], Brémaud
b. $h : \mathbb{R}_+ \to \mathbb{R}_+$	and Massoulié [3], Jovanović [24], Boumezoued [2], and Kirchner [28]
Analytical Methods	
References	Kerstan [26]
Assumptions 3	
a. $\Phi : \mathbb{R} \to \mathbb{R}_+$ Lipschitz	Brémaud and Massoulié [4], Chen et al. [5], and
b. $h : \mathbb{R}_+ \to \mathbb{R}$ general	Karabash [25]
Renewal Properties/Markov Processes	
General References	Lindvall [33], Nummelin [37], and Hairer [17]
Assumptions 3	
a. $\Phi : \mathbb{R} \to \mathbb{R}_+$ general	Brémaud and Massoulié [4], Hodara [21], Raad
b. $h : \mathbb{R}_+ \to \mathbb{R}$ general	et al. [40], Graham [16], Raad [39], and Costa et al. [7]
Assumptions 4	
a. $\Phi : \mathbb{R} \to \mathbb{R}_+$ general	Duarte, Aline et al. [11]
b. $h(x) = \exp(-x/\alpha)$	

delicate. Most proofs in this domain are based on the functional relation defining stationary Hawkes processes that can be expressed as a fixed point equation which can be solved through a Picard scheme, see Kerstan [26] for one of the pioneering papers on this subject. Brémaud and Massoulié [4] has developed this approach when Φ is unbounded and supposed to be β-Lipschitz, with the following condition,

$$\beta \int_0^{+\infty} |h(s)| \, \mathrm{d}s < 1.$$

Thinning techniques have been applied to the case of a bounded Lipschitz function Φ in the same reference, where the condition

$$\int_0^{+\infty} s|h(s)| \, \mathrm{d}s < +\infty$$

is sufficient to prove the existence and uniqueness of the stationary version of the Hawkes process.

In Brémaud and Massoulié [4], renewal theory is used to investigate Hawkes processes, with finite memory. This approach has been extended in Graham [16], Raad [39], and Costa et al. [7]. It is also possible to limit the influence of intensity rate to the last jump of the Hawkes process as done in Hodara and Löcherbach [22]. A recent study (Raad et al. [40]) has added a refractory effect to prevent explosion in the study of non-linear Hawkes processes.

To go further, it is interesting to consider Hawkes processes as Markov process in a general state space, either using counting processes (Duarte, Aline et al. [11]) or Markov

theory in the state of càdlàg functions (Karabash [25]). Coupling methods (Lindvall [33]) and general Markov theory (Nummelin [37], and Hairer [17]) are natural tools in this setting.

References

[1] Luc Bauwens and Nikolaus Hautsch. Modelling financial high frequency data using point processes. In: Ed. by Thomas Mikosch, Jens-Peter Kreiß, Richard A. Davis, and Torben Gustav Andersen. *Handbook of Financial Time Series*, pp. 953–979. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009. MR2848661

[2] Alexandre Boumezoued. Population viewpoint on Hawkes processes. *Advances in Applied Probability* 48.2 (June 2016), 463–480. MR3511771

[3] Pierre Brémaud and Laurent Massoulié. Hawkes branching point processes without ancestors. *Journal of Applied Probability* 38.1 (2001), 122–135. MR1816118

[4] Pierre Brémaud and Laurent Massoulié. Stability of nonlinear Hawkes processes. *The Annals of Probability* 24.3 (1996), 1563–1588. MR1411506

[5] Shizhe Chen, Ali Shojaie, Eric Shea-Brown, and Daniela Witten. *The multivariate Hawkes process in high dimensions: beyond mutual excitation*. arXiv:1707.04928 [stat] (July 2017).

[6] I. P. Cornfeld, S. V. Fomin, and Y. G. Sinaĭ. *Ergodic Theory*. Grundlehren der mathematischen Wissenschaften. Springer-Verlag, New York, 1982. MR0832433

[7] Manon Costa, Carl Graham, Laurence Marsalle, and Viet Chi Tran. Renewal in Hawkes processes with self-excitation and inhibition. *Advances in Applied Probability* 52.3 (Sept. 2020), 879–915. MR4153592

[8] Daryl J. Daley and David Vere-Jones. *An Introduction to the Theory of Point Processes. Volume 1: Elementary Theory and Methods*, 2nd edn. Probability and its Applications. Springer, Nov. 2003. MR1950431

[9] Andrew Daw and Jamol Pender. Queues driven by Hawkes processes. *Stochastic Systems* 8.3 (2018), 192–229. MR3863281

[10] Donald A. Dawson. Measure-valued Markov processes. In: *École d'Été de Probabilités de Saint-Flour XXI—1991*. Lecture Notes in Math., vol. 1541, pp. 1–260. Springer, Berlin, Nov. 1993. MR1242575

[11] Aline Duarte, Eva Löcherbach, and Guilherme Ost. Stability, convergence to equilibrium and simulation of non-linear Hawkes processes with memory kernels given by the sum of Erlang kernels. *ESAIM: PS* 23 (2019), 770–796. MR4044609

[12] E. Errais, K. Giesecke, and L. Goldberg. Affine point processes and portfolio credit risk. *SIAM Journal on Financial Mathematics* 1.1 (Jan. 2010), 642–665. MR2719785

[13] Jalal Etesami, Negar Kiyavash, Kun Zhang, and Kushagra Singhal. Learning network of multivariate Hawkes processes: a time series approach. In: *Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence, UAI'16*, pp. 162–171. AUAI Press, Arlington, Virginia, United States, 2016.

[14] W. Feller. *An Introduction to Probability Theory and Its Applications, Vol. II*, 2nd edn. John Wiley & Sons Ltd, New York, 1971. MR0270403

[15] Felipe Gerhard, Moritz Deger, and Wilson Truccolo. On the stability and dynamics of stochastic spiking neuron models: Nonlinear Hawkes process and point process GLMs. *PLOS Computational Biology* 13.2 (Feb. 2017), e1005390.

[16] Carl Graham. Regenerative properties of the linear Hawkes process with unbounded memory. *The Annals of Applied Probability* 31.6 (2021), 2844–2863. MR4350975

[17] Martin Hairer. *Convergence of Markov processes*, 2010.

[18] Paul R. Halmos. *Measure Theory*. D. Van Nostrand Co., Inc., New York, N.Y., 1950, xi+304 pp. MR0033869

[19] Alan G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58.1 (1971), 83–90. MR0278410

[20] Alan G. Hawkes and David Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability* 11.3 (1974), 493–503. MR0378093

[21] Pierre Hodara. Systèmes de neurones en interactions: modélisation probabiliste et estimation, PhD thesis, 2016.

[22] Pierre Hodara and Eva Löcherbach. Hawkes processes with variable length memory and an infinite number of components. *Advances in Applied Probability* 49.1 (2017), 84–107. MR3631217

[23] Jean Jacod. *Calcul Stochastique et Problèmes de Martingales*. Lecture Notes in Mathematics, vol. 714. Springer, Berlin, 1979, x+539 pp. MR0542115

[24] Stojan Jovanović. Cumulants of Hawkes point processes. *Physical Review E* 91.4 (2015).

[25] Dmytro Karabash. *On stability of Hawkes process*. arXiv:1201.1573 [math] (Jan. 2012).

[26] J. Kerstan. Teilprozesse Poissonscher Prozesse. In: *Trans. Third Prague Conf. Information Theory, Statist. Decision Functions, Random Processes*, pp. 377–403. House Czech. Acad. Sci., Liblice, Jan. 1964. MR0166826

[27] J. F. C. Kingman. *Poisson Processes*. Clarendon Press, Dec. 1992. MR1207584

[28] Matthias Kirchner. *Critical cluster cascades*, 2022. arXiv:2208.08383 [math.PR]. MR4583781

[29] Nicolas Kryloff and Nicolas Bogoliouboff. La théorie générale de la mesure dans son application à l'étude des systèmes dynamiques de la mécanique non linéaire. *Annals of Mathematics* 38.1 (1937), 65–113. MR1503326

[30] Günter Last and Mathew Penrose. *Lectures on the Poisson Process*. Cambridge University Press, Oct. 2017. MR3791470

[31] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, Providence, RI, 2009. MR2466937

[32] Peter A. W. Lewis. A branching Poisson process model for the analysis of computer failure patterns. *Journal of the Royal Statistical Society. Series B (Methodological)* 26.3 (1964), 398–456. MR0174090

[33] Torgny Lindvall. *Lectures on the Coupling Method*. Courier Corporation, Jan. 2002. MR1924231

[34] R. M. Loynes. The stability of queues with non independent inter-arrival and service times. *Proc. Cambridge Ph. Soc.* 58 (1962), 497–520. MR0141170

[35] S. Meyn and R. Tweedie. *Markov Chains and Stochastic Stability*. Communications and Control Engineering Series. Springer, 1993. MR1287609

[36] Jacques Neveu. Processus ponctuels. In: Ed. by P.-L. Hennequin. *École d'Été de Probabilités de Saint-Flour*. Lecture Notes in Math., vol. 598, pp. 249–445. Springer-Verlag, Berlin, 1977. MR0474493

[37] Esa Nummelin. *General Irreducible Markov Chains and Non-Negative Operators*. Cambridge University Press, June 2004. MR0776608

[38] David Oakes. The Markovian self-exciting process. *Journal of Applied Probability* 12.1 (1975), 69–77. MR0362522

[39] Mads Bonde Raad. Renewal time points for Hawkes processes, June 2019.

[40] Mads Bonde Raad, Susanne Ditlevsen, and Eva Löcherbach. Stability and mean-field limits of age dependent Hawkes processes. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 56.3 (2020), 1958–1990. MR4116713

[41] Patricia Reynaud-Bouret and Sophie Schbath. Adaptive estimation for Hawkes processes; application to genome analysis. *Annals of Statistics* 38.5 (Nov. 2010), 2781–2822. MR2722456

[42] Philippe Robert. *Stochastic Networks and Queues*. Stochastic Modelling and Applied Probability. Springer-Verlag, Berlin Heidelberg, 2003. MR1996883

[43] Philippe Robert and Gaëtan Vignoud. Averaging principles for Markovian models of plasticity. *Journal of Statistical Physics* 183.3 (June 2021), 47–90. MR4270776

[44] Philippe Robert and Gaëtan Vignoud. Stochastic models of neural plasticity. *SIAM Journal on Applied Mathematics* 81.5 (Sept. 2021), 1821–1846. MR4309869

[45] Philippe Robert and Gaëtan Vignoud. Stochastic models of neural synaptic plasticity: a scaling approach. *SIAM Journal on Applied Mathematics* 81.6 (Nov. 2021), 2362–2386. MR4335635

[46] L. C. G. Rogers and David Williams. *Diffusions, Markov Processes and Martingales: Volume 2, Itô Calculus*. Cambridge University Press, Sept. 2000. MR1780932

[47] D. Vere-Jones. Stochastic models for earthquake occurrence. *Journal of the Royal Statistical Society. Series B (Methodological)* 32.1 (1970), 1–62. MR0272087

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards

- Free for authors, free for readers

- Quick publication (no backlog)

- Secure publication (LOCKSS[1])

- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4], ProjectEuclid[5]

- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)

- Submit your best articles to EJP-ECP

- Choose EJP-ECP over for-profit journals

---

[1] LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[2] EJMS: Electronic Journal Management System: https://vtex.lt/services/ejms-peer-review/
[3] IMS: Institute of Mathematical Statistics http://www.imstat.org/
[4] BS: Bernoulli Society http://www.bernoulli-society.org/
[5] Project Euclid: https://projecteuclid.org/
[6] IMS Open Access Fund: https://imstat.org/shop/donation/