J. Appl. Prob. **38,** 122–135 (2001)
Printed in Israel

HAWKES BRANCHING POINT PROCESSES WITHOUT ANCESTORS

PIERRE BRÉMAUD,* *CNRS, France, and École Polytechnique Fédérale de Lausanne*
LAURENT MASSOULIÉ,** *Microsoft Research*

Abstract

In this article, we prove the existence of critical Hawkes point processes with a finite average intensity, under a heavy-tail condition for the fertility rate which is related to a long-range dependence property. Criticality means that the fertility rate integrates to 1, and corresponds to the usual critical branching process, and, in the context of Hawkes point processes with a finite average intensity, it is equivalent to the absence of ancestors. We also prove an ergodic decomposition result for stationary critical Hawkes point processes as a mixture of critical Hawkes point processes, and we give conditions for weak convergence to stationarity of critical Hawkes point processes.

Keywords: Stochastic processes; point processes; Hawkes processes; spectral analysis; long-rang dependence; coupling

AMS 2000 Subject Classification: Primary 60G55; 62M15

1. Critical Hawkes processes

Hawkes processes are described in terms of *shot noise* or *filtered point processes*. To describe shot noise, start from a basic point process N on $\mathbb{R}$ with a sequence of points $\{T_n\}_{n\in\mathbb{Z}}$, and a non-negative function $h(t)$ defined on $\mathbb{R}$ such that $h(t) = 0$ if $t \leq 0$. The corresponding shot noise is the process

$$Z(t) = \sum_{n\in\mathbb{Z}} h(t - T_n)\mathbf{1}_{\{T_n \leq t\}} = \int_{(-\infty,t]} h(t-s)N(\mathrm{d}s).$$

The classical Hawkes process (see [9]) is a point process N on $\mathbb{R}$ with an $\mathcal{F}_t^N$-stochastic intensity, namely (see [3, 4])

$$\lambda(t) = \nu + \int_{(-\infty,t)} h(t-s)N(\mathrm{d}s) \tag{1}$$

for a positive constant ν and a non-negative function h as above.

Recall that a point process on $\mathbb{R}$ is a mapping from the basic probability space to $\mathcal{N}$, the canonical space of point processes, consisting of locally bounded measures μ on $\mathbb{R}$ with values in $\mathbb{N}\cup\{+\infty\}$, endowed with the smallest σ-field making the mappings $\mu \mapsto \mu(C)$ measurable for all Borel sets $C \subset \mathbb{R}$. For each $t \in \mathbb{R}$, $\mathcal{F}_t^N$ is the σ-field generated by the random variables

Received 2 November 1999; revision received 18 October 2000.
* Postal address: École Supérieure d'Électricité, Laboratoire des Signaux et Systèmes, Plateau du Moulon, 91192 Gif sur Yvette, France. Email address: bremaud@lss.supelec.fr
** Postal address: Microsoft Research, St George House, 1 Guildhall Street, Cambridge CB2 3NH, UK.

$N(C)$, where C is a Borel set contained in $(-\infty, t]$. We say that N admits the $\mathcal{F}_t^N$-intensity $\{\lambda(t)\}_{t\in\mathbb{R}}$, where $\{\lambda(t)\}_{t\in\mathbb{R}}$ is an almost surely locally integrable non-negative process adapted to the filtration $\{\mathcal{F}_t^N\}_{t\in\mathbb{R}}$, if for all $(a, b] \in \mathbb{R}$, all $A \in \mathcal{F}_a^N$,

$$\mathrm{E}[N(a, b]\mathbf{1}_A] = \mathrm{E}\left[\int_a^b \lambda(s)\,\mathrm{d}s\mathbf{1}_A\right] \tag{2}$$

(see [4]).

Under the condition

$$\int h = \int_0^\infty h(t)\,\mathrm{d}t < 1, \tag{3}$$

there exists a unique stationary distribution for N with the $\mathcal{F}_t^N$-stochastic intensity (1) and such that N admits a finite average intensity $\lambda = \mathrm{E}[\lambda(0)]$ (see [10]). Such a point process is interpreted in terms of branching as follows. Each mother of a colony gives birth to daughters. If a mother is born at time s, the birth dates of its daughters form a Poisson process with intensity $h(t-s)$. As soon as they are born, daughters become mothers. Some mothers, called the ancestors, appear in the colony by immigration, and their immigration (birth) dates form a homogeneous Poisson process with intensity ν. The resulting point process of birth (natural, or by immigration) dates is the Hawkes process with stochastic intensity (1). The stability condition (3) is then a natural one in branching process theory.

The reader will find more properties of the classical Hawkes model in [7].

For the time being, we note that, if a stationary Hawkes point process exists, its average intensity λ satisfies

$$\lambda = \nu + \lambda \int_0^\infty h(t)\,\mathrm{d}t,$$

and therefore, when $\nu > 0$, the stability condition (3) is also necessary if we want N to have a finite average intensity.

In this article, we are concerned with Hawkes branching point processes *without* ancestors, that is a stationary point process N with the stochastic intensity

$$\lambda(t) = \int_{(-\infty,t)} h(t-s)N(\mathrm{d}s). \tag{4}$$

We then have $\lambda = \lambda \int h$, and therefore, since we want a finite and non-null average intensity, necessarily

$$\int_0^\infty h(t)\,\mathrm{d}t = 1. \tag{5}$$

We are therefore in the critical case of branching process theory.

Before we start the study of critical Hawkes processes with a finite and non-null average intensity (and in particular, show their existence under specific conditions), we shall review a few facts on non-critical Hawkes processes that will be needed in the sequel.

2. Bartlett spectra of Hawkes processes

The power spectral measure μ_N of a stationary point process N is defined, when it exists, by the requirement that

$$\operatorname{var} \int_{\mathbb{R}} \phi(t)N(\mathrm{d}t) = \int_{\mathbb{R}} |\hat{\phi}(\omega)|^2 \mu_N(\mathrm{d}\omega) \tag{6}$$

for any function $\phi \in \mathbb{L}^1 \cap \mathbb{L}^2$, where $\hat{\phi}(\omega) := \int e^{i\omega t}\phi(t)\,dt$ is the Fourier transform of ϕ. In point process theory, this power spectral measure is called the Bartlett spectrum of N. It is intimately connected to the Bochner spectrum μ_X of a second-order stationary process $\{X(t)\}_{t\in\mathbb{R}}$, defined by

$$\operatorname{var}\int_{\mathbb{R}} \phi(t)X(t)\,dt = \int_{\mathbb{R}} |\hat{\phi}(\omega)|^2 \mu_X(d\omega)$$

for any function $\phi \in \mathbb{L}^1 \cap \mathbb{L}^2$. Formally, the Bartlett spectrum of N is the Bochner spectrum of the *generalized* second-order stationary process

$$X(t) = \sum_{n\in\mathbb{Z}} \delta_0(t - T_n)\mathbf{1}_{\{T_n \le t\}},$$

where δ_0 is the Dirac generalized function.

The Bartlett spectrum is well-defined if $E[N(C)^2] < \infty$ for all bounded Borel sets C [7]. In order to obtain equalities as in (6), we shall make use of the following fact (see [14, p. 41] for an alternative derivation of the Bartlett spectrum of N, see [7]). If $f \in \mathbb{L}^1 \cap \mathbb{L}^2$, then

$$\operatorname{var}\int_{\mathbb{R}} f(t)[N(dt) - \lambda(t)\,dt] = \int_{\mathbb{R}} |f(t)|^2 \lambda\,dt,$$

where $\lambda := E(N(0, 1])$ is the average intensity of N. In particular, for a stationary Hawkes process with a finite average intensity λ, and $f \in \mathbb{L}^1 \cap \mathbb{L}^2$,

$$\operatorname{var}\int [f(x) - \tilde{h} * f(x)]N(dx) = \frac{\lambda}{2\pi}\int |\hat{f}(\omega)|^2\,d\omega, \tag{7}$$

where $\tilde{h}(t) = h(-t)$, and $\tilde{h} * f(x) = \int \tilde{h}(x - t)f(t)\,dt$. From this it follows that, under the condition (3), the Bartlett power spectral measure μ_N of the point process N admits a density f_N (see [9]):

$$f_N(\omega) := \frac{\nu}{2\pi(1 - \int h)|1 - \hat{h}(\omega)|^2}.$$

The spectral density tends to a finite constant $f_N(0) = \nu/(2\pi(1 - \int h)^3)$ as ω tends to zero. This may be viewed as an indication of a short range dependence property (see [6] for a discussion of short- and long-range dependence in the context of time-series, and [8] for point processes): for instance, if the behaviour of $\operatorname{var} N(0, T]$ as $T \to \infty$ is investigated, using the expression $(e^{i\omega T} - 1)/(i\omega)$ of the Fourier transform of $\mathbf{1}_{[0,T]}$, it is easily seen that

$$\operatorname{var} N(0, T] \sim T f_N(0) \int_{\mathbb{R}} \frac{|e^{iu} - 1|^2}{u^2}\,du = 2\pi T f_N(0)$$

(the value of the integral is indeed 2π, as follows from Parseval's identity), which is qualitatively identical to the behaviour of the Poisson process, and the latter certainly deserves to be called short-range dependent, because it has independent increments. Thus, Hawkes processes for which the stability condition $\int h < 1$ hold are short-range dependent. However, at $\nu = 0$,

$$f_N(0) := \frac{\lambda}{2\pi|1 - \int h|^2},$$

and therefore, as we approach the critical case, $f_N(0)$ becomes increasingly large.

3. Existence of non-trivial critical Hawkes processes

The existence of critical Hawkes processes with a finite and non-null average intensity will be established. In the standard sub-critical case, the proof of existence is straightforward because there are obvious regeneration points, namely, the immigration dates of the ancestors, and as a matter of fact, the standard Hawkes point process is a Poisson cluster point process [10].

It is easy to find an example of critical dynamics (that is, with a stochastic intensity of type (4)) whose only solution with a finite average intensity is the trivial (empty) point process. It suffices to take a fertility rate function $h(t)$ of compact support, say, null outside $[0, a]$, where $a < \infty$. More generally, we have the following result.

Proposition 1. *Let N be a stationary, ergodic Hawkes point process without ancestors (i.e., such that $\nu = 0$), and assume that the fertility rate function h satisfies the short-range condition*

$$\int_0^\infty t h(t)\,\mathrm{d}t < \infty. \tag{8}$$

Then the average intensity $\lambda := \mathrm{E}N([0, 1))$ is either equal to 0 or $+\infty$, i.e., there do not exist non-trivial critical Hawkes processes with short-range fertility rate.

Proof. Write

$$\begin{aligned}
\mathrm{P}(N(\mathbb{R}^+) = 0) &= \mathrm{E}[\mathrm{P}(N(\mathbb{R}^+) = 0 \mid \mathcal{F}_0)] \\
&= \mathrm{E}\biggl[\exp\biggl\{-\int_0^\infty \mathrm{d}t \int_{(-\infty,0)} N(\mathrm{d}s)h(t-s)\biggr\}\biggr] \\
&\geq \exp\biggl(-\mathrm{E}\biggl\{\int_0^\infty \mathrm{d}t \int_{(-\infty,0)} N(\mathrm{d}s)h(t-s)\biggr\}\biggr) \\
&= \exp\biggl(-\lambda \int_0^\infty t h(t)\,\mathrm{d}t\biggr),
\end{aligned}$$

where the second equality is a direct application of a result of Jacod [11] (see also [5, Lemma 1]), the inequality follows from Jensen's inequality, and the last step is a simple application of Fubini's theorem. Assuming the average intensity λ is finite, under the short range hypothesis (8), this calculation yields $\mathrm{P}(N(\mathbb{R}^+) = 0) > 0$. Note that ergodicity of N implies that the event $\{N(\mathbb{R}^+) = 0\}$ has probability either 0 or 1; it thus follows that, with probability 1, $N(\mathbb{R}^+) = 0$, and as a consequence $\lambda = 0$.

The specific conditions that we impose on the fertility rate, besides (5), are

$$\sup_{t \geq 0} t^{1+\alpha} h(t) \leq R, \tag{9}$$

$$\lim_{t \to \infty} t^{1+\alpha} h(t) = r, \tag{10}$$

for finite constants $r, R > 0$ and where $\alpha \in (0, \frac{1}{2})$. We prove Theorem 1 at the end of this section.

Theorem 1. *There exists a non-trivial stationary Hawkes point process without ancestors with finite intensity, provided the fertility rate integrates to 1 and satisfies the conditions (9) and (10), where $\alpha \in (0, \frac{1}{2})$.*

The proof of existence is non-constructive and uses weak convergence arguments. For this, we shall proceed as follows: for $\varepsilon \in (0, 1)$, consider the stationary Hawkes process with parameters $h^\varepsilon = (1-\varepsilon)h$, and $\nu^\varepsilon = \varepsilon\lambda$, for some fixed constant $\lambda > 0$. It is well defined, since the stability condition holds, and has average intensity λ for all $\varepsilon > 0$. As Proposition 1 shows, if we 'send the ancestors to infinity', their fertility rate must exhibit some long-tail property if we want a non-trivial limit process. The set of conditions chosen in this note is one of them.

Lemma 1. (i) *When (9) and (10) hold,*

$$\lim_{\omega\to 0}(\hat{h}(\omega) - 1)\omega^{-\alpha} = r\int_{\mathbb{R}+} \frac{e^{iu} - 1}{u^{\alpha+1}}\,du, \qquad 0 < \alpha < 1. \tag{11}$$

(ii) *If in addition $\alpha < \frac{1}{2}$, then the variances $V^\varepsilon = \operatorname{var} N^\varepsilon[0, T]$ are bounded as $\varepsilon \to 0$, and converge as $\varepsilon \to 0$ to the finite limit*

$$V^0 = \frac{\lambda}{2\pi}\int_{\mathbb{R}} \frac{|e^{i\omega T} - 1|^2}{\omega^2|1 - \hat{h}(\omega)|^2}\,d\omega.$$

Proof. (i) After a change of variables, the left-hand side of (11) reads

$$\int_{\mathbb{R}_+} \omega^{-1-\alpha}(e^{iu} - 1)h(u/\omega)\,du.$$

The modulus of the integrand is, in view of (9), less than $R|e^{iu} - 1|u^{-\alpha-1}$, which is integrable when $\alpha < 1$. We can thus apply Lebesgue's dominated convergence theorem to show that (11) follows from (10).

(ii) The spectral formula applied to N^ε with the function $f = \mathbf{1}_{[0,T]}$ yields

$$V^\varepsilon = \frac{\lambda}{2\pi}\int_{\mathbb{R}} \frac{|e^{i\omega T} - 1|^2}{\omega^2|1 - (1-\varepsilon)\hat{h}(\omega)|^2}\,d\omega.$$

The integrand is maximized for $1 - \varepsilon = \operatorname{Re}(\hat{h}(\omega))/|\hat{h}(\omega)|^2$. This ensures that the integrand is bounded above by

$$\frac{|e^{i\omega T} - 1|^2}{\omega^2} \times \frac{|\hat{h}(\omega)|^2}{|\operatorname{Im}\hat{h}(\omega)|^2},$$

where $\operatorname{Im} z$ denotes the imaginary part of the complex number z. In the neighbourhood of 0, say in some domain $|\omega| \le \omega_c$, this upper bound is integrable, since in view of (11) the second factor is of order $|\omega|^{-2\alpha}$, and we have assumed $\alpha < \frac{1}{2}$. In the complementary domain $|\omega| > \omega_c$, notice first that the function $\omega \mapsto |\hat{h}(\omega)|$ is continuous, goes to zero at infinity by the Riemann–Lebesgue lemma, and is therefore bounded above by some constant $\theta(\omega_c) > 0$. It is easily seen that necessarily, $\theta(\omega_c) < 1$. We thus have

$$|\omega| > \omega_c \Rightarrow |1 - (1-\varepsilon)\hat{h}(\omega)| \ge 1 - (1-\varepsilon)\theta(\omega_c) \ge 1 - \theta(\omega_c) > 0.$$

The integrand in the definition of V^ε is thus bounded above by some integrable function which does not depend on ε. The conclusions of the lemma then follow by dominated convergence.

Note that, in particular, at the limit, the power spectral density is a constant times $|\omega^{-2\alpha}|$ in the neighbourhood of the zero frequency, and the limit of var $N[0, T]$ is $O(T^{1+2\alpha})$ as $T \to \infty$. In this sense, the limit process exhibits long-range dependence. However, in order to make this argument rigorous, we shall have to prove that the limit of the variance is indeed the variance of the limit. This is done in Section 5.

Proof of Theorem 1. In view of Lemma 1, the family of random variables $N^{\varepsilon}[0, T]$ is tight. This ensures that the finite-dimensional distributions of the processes N^{ε} also constitute a tight family. As a consequence, by Theorem 9.1.VI of [7], the laws of the N^{ε} are tight. Consider, then, any limit distribution as $\varepsilon \to 0$. We first establish that this distribution is indeed that of a Hawkes process with parameters $\nu = 0$ and h. To see this, it is enough to prove (2) for all $a < b$, and all events A in some semiring generating $\mathcal{F}_a^N$. Take A of the form $\{N(C_1) = k_1, \dots, N(C_n) = k_n\}$ for integers $k_1, \dots, k_n$ and measurable subsets $C_1, \dots, C_n \subseteq (-\infty, a]$. In order to show that this formula holds true when N is the corresponding limiting process, we should apply the formula to N^{ε} and check that we can go to the limit $\varepsilon \to 0$ in the formula. But $N(a, b]\mathbf{1}_A$ is equal to $g(N(a, b], N(C_1), \dots, N(C_n))$ for some continuous function g (taking advantage of the fact that these random variables are integer-valued). Therefore, $g(N^{\varepsilon}(a, b], N^{\varepsilon}(C_1), \dots, N^{\varepsilon}(C_n))$ converges weakly to $g(N(a, b], N(C_1), \dots, N(C_n))$. To show that the limiting procedure is valid for the left-hand side, it suffices to show that the variables $g(N^{\varepsilon}(a, b], N^{\varepsilon}(C_1), \dots, N^{\varepsilon}(C_n))$ are uniformly integrable (see [2, Theorem 5.4]). This is true, since they are bounded by the random variables $N^{\varepsilon}(a, b]$ whose second moments are bounded, as shown in Lemma 1.

Define

$$\phi(N) := \int_a^b \mathrm{d}t \int_{(-\infty,t)} h(t-s)N(\mathrm{d}s)\mathbf{1}_A.$$

To prove that the limiting procedure is valid for the right-hand side, we have to show that the expectation of $\phi(N^{\varepsilon})$ tends to that of $\phi(N)$. By [7, Proposition 9.1.VIII], $\phi(N)$ is a continuous function of N for the appropriate topology on the canonical space of locally finite point processes that makes it a complete separable metric space (see [7, Chapter 9] for the details). Therefore, $\phi(N^{\varepsilon})$ converges weakly to $\phi(N)$, since N^{ε} converges weakly to N. It remains to show that the variables $\phi(N^{\varepsilon})$ are uniformly integrable. Denoting

$$Y^{\varepsilon} := \int_a^b \mathrm{d}t \int_{(-\infty,t)} h(t-s)N^{\varepsilon}(\mathrm{d}s)\mathbf{1}_A,$$

it is enough to show that the Y^{ε} have a bounded second moment. Applying the spectral formula (6), we obtain

$$\mathrm{E}(Y^{\varepsilon})^2 = [\lambda(b-a)]^2 + \int_{\mathbb{R}} |\hat{f}(\omega)|^2 f_{N^{\varepsilon}}(\omega)\,\mathrm{d}\omega,$$

where

$$f(s) := \int_{a-s}^{b-s} h(u)\,\mathrm{d}u.$$

The second moment of Y^{ε} is uniformly bounded when the function $|\hat{f}(\omega)|^2$ is integrable at infinity. Using the fact that f is differentiable with integrable derivative, it follows from the Riemann–Lebesgue lemma that $|\hat{f}(\omega)| = O(\omega^{-1})$ as $\omega \to \infty$, hence the uniform integrability of the Y^{ε}, with the same proof as that of Lemma 1(ii).

4. Coupling properties

We now address two intimately connected issues. The first is the behaviour of transient critical processes and, in particular, their convergence in some sense to stationarity. The second is the question of uniqueness. Note that, with respect to the latter issue, the sum of two independent critical Hawkes processes with the same fertility rate h is a critical Hawkes processes with the same fertility rate h. Thus, uniqueness can be expected to hold only if we add a constraint, say, to have a fixed average intensity.

We first treat the issue of convergence, using coupling methods. Consider two distinct point processes N_1, N_2 and assume that, for $i = 1, 2$, N_i admits on $\mathbb{R}^+$ the stochastic intensity

$$\lambda_i(t) = \nu_i(t) + \int_{(0,t)} h(t-s) N_i(\mathrm{d}s),$$

where ν_i is some non-negative, $\mathcal{F}_0$-measurable process. For instance, if these processes are defined on $\mathbb{R}^-$ as well, the particular form $\nu_i(t) = \int_{(-\infty,0]} h(t-s) N_i(\mathrm{d}s)$ corresponds to the classical Hawkes process dynamics as considered in this paper.

The following construction is given in [15] and [5] in the same context, and is a special case of the results of [12, Chapter 14, pp. 469–478]. By extending, if necessary, the probability space on which N_1 and N_2 are constructed, assume the existence of a homogeneous Poisson process $\bar{N}$ on $\mathbb{R}^2$, independent of $\mathcal{F}_0$. Then one can construct point processes $\hat{N}_i$, $i = 1, 2$, on $\mathbb{R}^+$ such that

$$\hat{N}_i(\mathrm{d}t) = \bar{N}(\mathrm{d}t \times [0, \hat{\lambda}_i(t)]),$$

where $\hat{\lambda}_i(t) = \nu_i(t) + \int_{(0,t)} h(t-s)\hat{N}_i(\mathrm{d}s)$. The resulting processes are such that the joint law of $(\mathcal{F}_0, N_i)$ coincides with that of $(\mathcal{F}_0, \hat{N}_i)$.

Note that this construction may yield improper (that is, not locally finite) point processes. However, if the stochastic intensities are locally integrable, the point processes are proper. To guarantee this, it suffices for $\nu(t)$ (for convenience we momentarily drop the indices) to be locally integrable. Indeed,

$$\lambda(t) \leq \nu(t) + \int_{(0,t)} h(t-s) N(\mathrm{d}s).$$

Taking expectations and iterating this equation, we obtain the following bound:

$$\mathrm{E}[\lambda(t) \mid \mathcal{F}_0] \leq \nu(t) + \sum_{n>0} \nu * h^{*n}(t).$$

Integrated between 0 and T, the right-hand side equals

$$\frac{\int_0^T \nu(t)\,\mathrm{d}t}{1 - \int_0^T h(t)\,\mathrm{d}t},$$

which is finite under our assumptions on the function h. Hence, $\mathrm{E}[\delta(t) \mid \mathcal{F}_0]$ is almost surely locally integrable.

We now give conditions under which the two shifted processes $S_t N_1$ and $S_t N_2$ become close in some sense as $t \to \infty$.

Lemma 2. *Assume that both ν_1 and ν_2 are locally integrable and that*

$$\lim_{t\to\infty} \frac{|\nu_1(t) - \nu_2(t)|}{\int_t^\infty h(s)\,\mathrm{d}s} = 1 \quad a.s. \tag{12}$$

For all $a < b$,

$$\lim_{t\to\infty} \mathrm{P}[\hat{N}_1 = \hat{N}_2 \text{ on } (t+a, t+b] \mid \mathcal{F}_0] = 0 \quad a.s. \tag{13}$$

This is a typical *coupling* result (see e.g. [13] for a general reference on this method, and [5, 15] for a very similar application).

Proof. The point process Δ defined by $\Delta(\mathrm{d}t) = |\hat{N}_1(\mathrm{d}t) - \hat{N}_2(\mathrm{d}t)|$ admits on $\mathbb{R}^+$ the stochastic intensity (see [15] and [5])

$$\delta(t) = \left|\nu_1(t) - \nu_2(t) + \int_{(0,t)} h(t-s)(\hat{N}_1(\mathrm{d}s) - \hat{N}_2(\mathrm{d}s))\right|.$$

In particular,

$$\delta(t) \le |\nu_1(t) - \nu_2(t)| + \int_{(0,t)} h(t-s)\Delta(\mathrm{d}s).$$

Taking expectations and iterating this equation, we obtain the following bound:

$$\mathrm{E}[\delta(t) \mid \mathcal{F}_0] \le |\nu_1 - \nu_2|(t) + \sum_{n>0} |\nu_1 - \nu_2| * h^{*n}(t). \tag{14}$$

Fix $\varepsilon > 0$, and select $T > 0$ so that for $u \ge T$, $|\nu_1 - \nu_2|(u) \le \varepsilon \int_u^\infty h(s)\,\mathrm{d}s$ (such T exists almost surely for all $\varepsilon > 0$ by the assumption at (12)). Then for $u \ge T$, (14) implies that

$$\mathrm{E}[\delta(u) \mid \mathcal{F}_0] \le \varepsilon \sum_{n\ge 0} \int_0^u \mathrm{d}s\, h^{*n}(u-s) \int_s^\infty h(v)\,\mathrm{d}v + \sum_{n>0} \int_0^T h^{*n}(u-s)|\nu_1 - \nu_2|(s)\,\mathrm{d}s.$$

As h integrates to 1, the generic term in the first summation reads

$$\int_0^u h^{*n}(v)\,\mathrm{d}v - \int_0^u h^{*n}(v)\,\mathrm{d}v \int_0^{u-v} h(s)\,\mathrm{d}s.$$

A change of variable argument shows that the second term there equals $\int_0^u h^{*(n+1)}(v)\,\mathrm{d}v$. This enables us to cancel terms in this first summation, which then reduces exactly to ε. Integrating this inequality between $t+a$ and $t+b$ yields

$$\int_{t+a}^{t+b} \mathrm{E}[\delta(u) \mid \mathcal{F}_0]\,\mathrm{d}u \le (b-a)\varepsilon + \int_0^T |\nu_1 - \nu_2|(s)\,\mathrm{d}s \left\{\sum_{n>0} \int_{t+a}^{t+b} h^{*n}(u-s)\,\mathrm{d}u\right\}.$$

Denoting by U the renewal measure associated with the probability density h, the second term in this right-hand side is bounded above by

$$\int_0^T |\nu_1 - \nu_2|(s)\,\mathrm{d}s\, U([t+a-s, t+b-s]).$$

It thus follows by dominated convergence that $\int_{(t+a,t+b)} \mathrm{E}[\delta(u) \mid \mathcal{F}_0]\,\mathrm{d}u$ goes to 0 as $t \to \infty$. Indeed, $U([t+a-s, t+b-s])$ tends to $0 = 1/\int t h(t)\,\mathrm{d}t$ by the renewal theorem (see, for instance, [13, pp. 76–77] for the extension of the standard renewal theorem to the present situation, where the inter-renewal times have infinite expectation $\int th(t)\,\mathrm{d}t$), and is bounded by the finite quantity $U([0, b-a])$ as a simple argument shows. The conclusion of the lemma follows by noticing that

$$\mathrm{E}[|\hat{N}_1 - \hat{N}_2|(t+a, t+b] \mid \mathcal{F}_0] = \int_{(t+a,t+b)} \mathrm{E}[\delta(u) \mid \mathcal{F}_0]\,\mathrm{d}u,$$

which goes to zero as $t \to \infty$, by the previous analysis. Therefore,

$$\lim_{t\to\infty} \mathrm{E}[|\hat{N}_1 - \hat{N}_2|(t+a, t+b] \mid \mathcal{F}_0] = 0 \quad \text{a.s.}$$

from which (13) follows.

Lemma 3. *Consider two point processes N_i, $i = 1, 2$ on $\mathbb{R}$, admitting on $\mathbb{R}^+$ the stochastic intensities*

$$\lambda_i(t) = \int_{(-\infty,t)} h(t-s)N_i(\mathrm{d}s).$$

Assume, in addition, that for some $\lambda > 0$,

$$\lim_{t\to\infty} \frac{N_i[-t, 0]}{t} = \lambda \quad \textit{a.s.}, \quad i = 1, 2. \tag{15}$$

Then, by setting $\nu_i(t) = \int_{(-\infty,0]} h(t-s)N_i(\mathrm{d}s)$, both ν_i are locally integrable, and the following equivalences hold

$$\nu_i(t) \sim \lambda \int_t^\infty h(s)\,\mathrm{d}s \sim \alpha^{-1}\lambda r t^{-\alpha} \quad \textit{as } t \to \infty,$$

with r as at (10). In particular, the assumptions of Lemma 2 are all satisfied.

Proof. For fixed $t > 0$, we have

$$\int_0^t \nu_i(s)\,\mathrm{d}s = \int_{(-\infty,0]} N_i(\mathrm{d}u) \int_{-u}^{t-u} h(s)\,\mathrm{d}s.$$

Split the integral of the right-hand side according to whether $u \geq -1$ or $u < -1$ (the choice -1 is arbitrary). The first integral so obtained is then bounded above by $N_i[-1, 0] \int_0^\infty h(s)\,\mathrm{d}s$, which is finite. In view of (9), the second integral is less than

$$\int_{(-\infty,-1)} N_i(\mathrm{d}u) Rt(-u)^{-1-\alpha}.$$

Replace $(-u)^{-1-\alpha}$ by $\int_{-u}^\infty (1+\alpha)z^{-2-\alpha}\,\mathrm{d}z$ and apply Fubini's theorem to see that this expression equals

$$Rt(1+\alpha) \int_1^\infty z^{-2-\alpha} N_i[-z, -1)\,\mathrm{d}z.$$

In view of (15), the integrand is equivalent to $\lambda z^{-1-\alpha}$ as $z \to \infty$, so that the integral is finite. Thus, ν_i is locally integrable.

Write

$$\nu_i(t) = \int_{(-\infty,-1]} h(t-s)N_i(\mathrm{d}s) + \int_{(-1,0]} h(t-s)N_i(\mathrm{d}s).$$

The second term of the right-hand side is, for large t, of order $t^{-1-\alpha}$. Fix $\varepsilon > 0$, and consider T such that for $t \geq T$ (cf. (10)), $r^{-1}t^{1+\alpha}h(t) \in (1-\varepsilon, 1+\varepsilon)$. For $t \geq T$, the first term is equal to

$$C \int_{(-\infty,-1]} r(t-s)^{-1-\alpha} N_i(\mathrm{d}s) = r(1+\alpha) \int_{t+1}^{\infty} u^{-2-\alpha}\,\mathrm{d}u\, N_i[t-u, -1],$$

for some constant C in $(1-\varepsilon, 1+\varepsilon)$. Because of the assumption (15), this last term is equivalent to

$$Cr(\alpha+1)\lambda \int_{t+1}^{\infty} u^{-2-\alpha}(u-t)\,\mathrm{d}u,$$

itself equivalent to $C\alpha^{-1}r\lambda t^{-\alpha}$. Recalling that ε is arbitrary, this latter term is equivalent to $\nu_i(t)$ as $t \to \infty$, and hence of order $\int_t^\infty h(s)\,\mathrm{d}s$.

The previous lemmas allow us to establish the following result.

Theorem 2. *Suppose that the fertility rate h satisfies (5), (9) and (10). The following properties then hold:*

(i) *For all $\lambda > 0$, there exists at most one distribution for a stationary Hawkes point process with average intensity λ and such that $\lim_{t\to\infty} N[-t,0]/t = \lambda$ almost surely; let $\pi_\lambda(\cdot)$ denote this distribution when it exists.*

(ii) *The distribution π_λ is ergodic and mixing.*

(iii) *Any stationary Hawkes point process such that $\Lambda := \lim_{t\to\infty} N[-t,0]/t$ is almost surely finite, admits the ergodic decomposition*

$$\mathrm{P}(N \in \cdot) = \int \pi_\lambda(\cdot)\mathrm{P}(\Lambda \in \mathrm{d}\lambda).$$

(iv) *Any point process N on $\mathbb{R}^+$ with a stochastic intensity*

$$\lambda(t) = \nu_i(t) + \int_{(0,t)} h(t-s)N(\mathrm{d}s),$$

where $\nu_i(t) \sim \Lambda \int_t^\infty h(s)\,\mathrm{d}s$ as $t \to \infty$ for an almost surely finite random variable Λ such that the distribution $\pi(\Lambda)$ exists almost surely, satisfies the weak convergence property

$$\mathrm{P}(S_t N \in \cdot) \xrightarrow{\mathrm{w}} \int \pi_\lambda(\cdot)\mathrm{P}(\Lambda \in \mathrm{d}\lambda).$$

Proof. (i) Let N_1 and N_2 be two stationary Hawkes point processes satisfying

$$\lim_{t\to\infty} N_i[-t,0]/t = \lambda$$

almost surely. Let $\hat{N}_1$ and $\hat{N}_2$ be as in the coupling lemma (Lemma 2),

$$\lim_{t\to\infty} \mathrm{P}[\hat{N}_1 = \hat{N}_2 \text{ on } (t+a, t+b] \mid \mathcal{F}_0] = 1 \quad \text{a.s.}$$

From this it follows that the finite-dimensional distributions of N_1 and N_2 are the same, hence their distributions coincide: there is at most one distribution for a stationary Hawkes point process N such that $\lim_{t\to\infty} N_i[-t,0]/t = \lambda$ almost surely, and we denote it by π_λ when it exists.

(ii) Consider some arbitrary finite interval $[a, b]$, and let A, B be two measurable subsets of the space $\mathcal{N}$ of point processes involving only the restriction of the point process to the interval $[a, b]$. Mixing holds if for all such $a < b$, A and B,

$$\lim_{t\to\infty} \mathrm{P}(N \in A, S_t N \in B) = \mathrm{P}(N \in A)\mathrm{P}(N \in B).$$

Consider two independent copies N_i of N, and apply the construction of Lemma 2 to them, starting not from time 0 but from time b. It then holds that

$$\begin{aligned}\mathrm{P}(&N \in A, S_t N \in B)\\ &= \mathrm{P}(\hat{N}_1 \in A, S_t\hat{N}_2 \in B) + \{\mathrm{P}(\hat{N}_1 \in A, S_t\hat{N}_1 \in B) - \mathrm{P}(\hat{N}_1 \in A, S_t\hat{N}_2 \in B)\}.\end{aligned}$$

Note that $\hat{N}_1$ coincides with N_1 on $(-\infty, b]$, while $\hat{N}_2$ is constructed from N_2 and $\bar{N}$ only, and is thus independent of the event $\{\hat{N}_1 \in A\}$. The first term in the previous equation is then exactly equal to $\mathrm{P}(N \in A)\mathrm{P}(N \in B)$. The term in curly brackets goes to 0 as $t \to \infty$, because its modulus is less than the variation distance between the distributions of $\hat{N}_1$ and $\hat{N}_2$ when restricted to $[t + a, t + b]$, and by Lemma 2 the latter goes to zero as $t \to \infty$.

(iii) By assumption, $\lim_{t\to\infty} N[-t, 0]/t = \Lambda$, which is an almost surely finite random variable (existence of the almost surely limit Λ is implied by Birkhoff's ergodic theorem; its finiteness needs to be assumed, though). It is easy to show that the distribution of N conditionally on $\Lambda = l$ is still the distribution of a stationary Hawkes point process, under which $\lim_{t\to\infty} N[-t, 0]/t = l$ almost surely. It then follows from (i) that the distribution of N conditionally on $\Lambda = l$ must coincide with π_l.

(iv) The final property follows from the coupling lemma in an obvious way.

5. Second-order properties

The existence of stationary critical Hawkes processes N_λ with average intensity λ for any $\lambda > 0$ has been established, but at this point their second order structure is yet to be analysed. This in turn will enable us to prove the existence of an ergodic Hawkes point process with average intensity λ for any given $\lambda > 0$ (see Theorem 3 below), thus complementing the results of Theorem 2, which established uniqueness but left open the question of existence for every $\lambda > 0$. As mentioned just after the proof of Lemma 1, we need to show that the variance of the limit is equal to the limit of the variance. One route is to prove uniform integrability for the second moments; we take another.

One can observe that for all $g \in C_0^K$ (where C_0^K denotes the space of continuous functions with compact support), viewing $\int g(x)N_\lambda(\mathrm{d}x)$ as the weak limit of the $\int g(x)N^\varepsilon(\mathrm{d}x)$ when ε follows a sequence decreasing to 0 along which the N^ε converge weakly, Fatou's lemma implies that

$$\operatorname{var}\int g(x)N_\lambda(\mathrm{d}x) \le \frac{\lambda}{2\pi}\int \frac{|\hat{g}(\omega)|^2}{|1-\hat{h}(\omega)|^2}\,\mathrm{d}\omega. \tag{16}$$

Lemma 4 provides an inequality in the converse direction.

Lemma 4. *Suppose that the fertility rate h satisfies (5), (9), and (10). For all $g \in \mathbb{L}^1 \cap \mathbb{L}^2$, and every stationary Hawkes process N such that $\Lambda := \lim_{t\to\infty} N[-t, 0]/t$ almost surely finite,*

$$\operatorname{var}\int g(x)N(\mathrm{d}x) \ge \operatorname{var}(\Lambda)\left(\int g(x)\,\mathrm{d}x\right)^2 + \frac{\mathrm{E}(\Lambda)}{2\pi}\int \frac{|\hat{g}(\omega)|^2}{|1-\hat{h}(\omega)|^2}\,\mathrm{d}\omega. \tag{17}$$

Proof. It suffices to establish (17) in the case where Λ reduces to a constant, since the general formula follows from that case by using the conditional variance formula. We thus assume that $\Lambda = \lambda$ almost surely for some constant $\lambda > 0$. If $\int g(x)N(\mathrm{d}x)$ has infinite variance, the lemma holds trivially. Hence, assume that $\operatorname{var}\int g(x)N(\mathrm{d}x)$ is finite. Applying (7) to the function

$$f := g + g * \tilde{h} + \cdots + g * \tilde{h}^{*n}$$

yields

$$\operatorname{var}\int [g(x) - \tilde{h}^{*(n+1)} * g(x)]N(\mathrm{d}x) = \frac{\lambda}{2\pi}\int |\hat{g}(\omega)|^2 \left|\sum_{k=0}^{n} \hat{h}(\omega)^k\right|^2 \mathrm{d}\omega.$$

As is shown below, $\operatorname{var}\int g * \tilde{h}^{*n}(x)N(\mathrm{d}x) \to 0$ as $n \to \infty$. Expanding the left-hand side of the previous expression, and letting $n \to \infty$, we thus see that (17) holds with equality instead of inequality, under the assumption that $\Lambda = \lambda$ almost surely, and $\operatorname{var}\int g(x)N(\mathrm{d}x)$ is finite.

We now show that $\operatorname{var}\int g * \tilde{h}^{*n}(x)N(\mathrm{d}x)$ converges to zero as $n \to \infty$. Consider the process $X_t^n := \int(\tilde{h}^{*n} * g)(x+t)N(\mathrm{d}x)$. Note that one process is obtained from the previous one by convolutional filtering:

$$X_t^{n+1} = \int h(s)X_{t+s}^n\,\mathrm{d}s. \tag{18}$$

Let $C^0(u)$ denote $\operatorname{cov}(X_t^0, X_{t+u}^0)$. Note that $C^0(u)$ is bounded above by $\operatorname{var} X_0^0$, which is finite by assumption. Let us now show that the process X^1 admits a spectral measure μ. By Bochner's theorem (see e.g. [16]), its existence is ensured provided the process is second-order stationary, and $\mathbb{L}^2$-continuous. The process is in fact stationary ergodic and, hence, second-order stationary. In order to prove $\mathbb{L}^2$-continuity, write

$$\begin{aligned}\operatorname{var}(X_y^1 - X_0^1) &= \operatorname{cov}\left(\int\left(h(s-y) - h(s)\right)X_s^0\,\mathrm{d}s, \int(h(t-y) - h(t))X_t^0\,\mathrm{d}t\right)\\ &= \int\int C^0(t-s)(h(s-y) - h(s))(h(t-y) - h(t))\,\mathrm{d}s\mathrm{d}t.\end{aligned}$$

The absolute value of the last term is smaller than $C_0^0(\int |h(t-y) - h(t)|\,\mathrm{d}t)^2$, which goes to zero as $y \to 0$, since h is integrable. In view of (18), X^n also admits a spectral measure, given by $\mu(\mathrm{d}\omega)|\hat{h}(\omega)|^{2n-2}$. Its integral is equal to the variance of $\int g * \tilde{h}^{*n}(x)N(\mathrm{d}x)$, which we want to prove goes to zero. Since $\hat{h}$ is continuous and $|\hat{h}(\omega)| < 1$ for all $\omega \neq 0$, this convergence holds if we can prove that $\mu(\{0\}) = 0$. We use the following lemma, whose proof we defer momentarily.

Lemma 5. *Let $\{X_n\}_{n>0}$ be a sequence of centred random variables, converging to* 0 *in probability as $n \to \infty$. Assume further that for all $n > 0$, $X_n \leq_{\mathrm{cx}} Y$ where Y is a random variable with finite variance, and $\leq_{\mathrm{cx}}$ denotes the convex stochastic ordering (see e.g. [1]). Then* $\lim_{n\to\infty}\operatorname{var}(X_n) = 0$.

Let $X_n := (1/n)\int_0^n X_t^1\,\mathrm{d}t - \mathrm{E}(X_0^1)$. By Birkhoff's ergodic theorem, X_n converges almost surely, and thus *a fortiori* in probability, to 0 as $n \to \infty$. Also, taking $Y := X_0^1 - \mathrm{E}(X_0^1)$, it holds that $X_n \leq_{\mathrm{cx}} Y$, and Y has finite variance. Lemma 5 thus applies, and so $\lim_{n\to\infty}\operatorname{var}(X_n) = 0$. By the spectral formula, we have

$$\operatorname{var}(X_n) = \int\left|\frac{\mathrm{e}^{\mathrm{i}\omega n} - 1}{\mathrm{i}\omega n}\right|^2 \mu(\mathrm{d}\omega).$$

From this it follows that $\lim_{n\to\infty}\operatorname{var}(X_n) = \mu(\{0\})$, and thus necessarily $\mu(\{0\}) = 0$.

Proof of Lemma 5. Suppose that $\mathrm{var}(X_n)$ does not converge to 0; by considering a subsequence, assume without loss of generality that $\mathrm{var}(X_n)$ is bounded from below by $\sigma^2 > 0$. Using convergence in probability of X_n to 0, for all $\delta > 0$, there exists $n(\delta)$ such that

$$\mathrm{P}(X_n^2 > \tfrac{1}{3}\sigma^2) \le \delta, \qquad n \ge n(\delta).$$

Write $\mathrm{var}(X_n)$ as $\int_0^\infty \mathrm{P}(X_n^2 > t)\,\mathrm{d}t$, and split this integral in three intervals

$$[0, \tfrac{1}{3}\sigma^2), \qquad [\tfrac{1}{3}\sigma^2, \tfrac{1}{3}\sigma^2(1+\delta^{-1})), \qquad [\tfrac{1}{3}\sigma^2(1+\delta^{-1}), +\infty).$$

Bounding $\mathrm{P}(X_n^2 > t)$ from above by 1 on the first interval, and by $\mathrm{P}(X_n^2 > \frac{1}{3}\sigma^2)$ on the second, we obtain:

$$\int_{(\sigma^2/3)(1+\delta^{-1})}^{\infty} \mathrm{P}(X_n^2 > t)\,\mathrm{d}t \ge \sigma^2 - \tfrac{1}{3}\sigma^2 - \tfrac{1}{3}\sigma^2 = \tfrac{1}{3}\sigma^2, \qquad n \ge n(\delta).$$

This in turn is equivalent to

$$\mathrm{E}(X_n^2 - \tfrac{1}{3}\sigma^2(1+\delta^{-1}))^+ \ge \tfrac{1}{3}\sigma^2, \qquad n \ge n(\delta).$$

However, by assumption, the left-hand side is bounded above by $\mathrm{E}(Y^2 - \frac{1}{3}\sigma^2(1+\delta^{-1}))^+$, since the function $x \mapsto (x^2 - t)^+$ is convex for all t. Since δ can be taken arbitrarily small, this implies that $\lim_{t\to\infty} \mathrm{E}(Y^2 - t)^+ \ge \frac{1}{3}\sigma^2 > 0$, which is not compatible with the assumption that $\mathrm{var}(Y)$ is finite, hence a contradiction.

As a corollary to Lemma 4, we obtain the following result, thus complementing Theorem 2 which left open the question of existence of the distributions π_λ.

Theorem 3. *Under the assumptions of Lemma 4, there exists an ergodic distribution π_λ for every $\lambda > 0$. Also, formula (17) holds with equality, i.e., for all $g \in \mathbb{L}^1 \cap \mathbb{L}^2$, and every stationary Hawkes process N such that $\Lambda := \lim_{t\to\infty} N[-t,0]/t$ almost surely finite,*

$$\mathrm{var} \int g(x) N(\mathrm{d}x) = \mathrm{var}(\Lambda)\biggl(\int g(x)\,\mathrm{d}x\biggr)^2 + \frac{\mathrm{E}(\Lambda)}{2\pi} \int \frac{|\hat{g}(\omega)|^2}{|1-\hat{h}(\omega)|^2}\,\mathrm{d}\omega. \tag{19}$$

Proof. Combining the lower bound (17) of Lemma 4 with the upper bound (16) we had for N_λ, enables us to deduce that for $g \in C_0^K$ that $\mathrm{var}(\Lambda) = 0$, where $\Lambda := \lim_{t\to\infty} N_\lambda([-t,0])/t$. Thus, by Theorem 2(i), N_λ must be distributed as π_λ, which thus exists (note that, since any adherent distribution to the sequence $\{N^\varepsilon\}$ has to be distributed as π_λ, $\{N^\varepsilon\}$ is not only tight but also weakly convergent). This combination of lower and upper bounds also shows that

$$\mathrm{var} \int g(x) N_\lambda(\mathrm{d}x) = \frac{\lambda}{2\pi} \int \frac{|\hat{g}(\omega)|^2}{|1-\hat{h}(\omega)|^2}\,\mathrm{d}\omega.$$

Thus, (19) holds when N is distributed as π_λ (and for $g \in C_K^0$); the general case follows from the conditional variance formula (and by density of C_K^0).

Acknowledgements

Part of this article was presented at the meeting on Applied Probability held in the Mathematical Institute at Oberwolfach in November 1998. The authors wish to express their gratitude to this institution for encouraging and supporting such events. They are also grateful to the reviewers for their constructive comments, which greatly helped improve the presentation of this article.

References

[1] BACCELLI, F. AND BRÉMAUD, P. (1994). *Elements of Queueing Theory*. Springer, New York.
[2] BILLINGSLEY, P. (1968). *Weak Convergence of Probability Measures*. John Wiley, New York.
[3] BRANDT, A. AND LAST, G. (1995). *Marked Point Processes on the Real Line. The Dynamic Approach*. Springer, New York.
[4] BRÉMAUD, P. (1981). *Point Processes and Queues*. Springer, New York.
[5] BRÉMAUD, P. AND MASSOULIÉ, L. (1996). Stability of nonlinear Hawkes processes. *Ann. Prob.* **24**, 1563–1588.
[6] COX, D. R. (1984). Long range dependence: a review. In *Statistics: an Appraisal* (Proc. 50th Anniv. Conf. Iowa State Statist. Lab.), eds H. A. David and H. T. David. Iowa State University Press.
[7] DALEY, D. J. AND VERE-JONES, D. (1988). *An Introduction to the Theory of Point Processes*. Springer, New York.
[8] DALEY, D. J. AND VESILO, R. (1997). Long range dependence of point processes, with queueing examples. *Stoch. Proc. Appl.* **70**, 265–282.
[9] HAWKES, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58**, 83–90.
[10] HAWKES, A. G. AND OAKES, D. (1974). A cluster representation of a self-exciting point process. *J. Appl. Prob.* **11**, 493–503.
[11] JACOD, J. (1975). Multivariate point processes: predictable projections, Radon–Nikodým derivatives, representation of martingales. *Z. Wahrscheinlichkeitsth.* **31**, 235–253.
[12] JACOD, J. (1979). *Calcul Stochastique et Problèmes de Martingales*. Springer, Berlin.
[13] LINDVALL, T. (1992). *Lectures on the Coupling Method*. John Wiley, New York.
[14] MASSOULIÉ, L. (1995). Stabilité, simulation et optimisation de systèmes à événements discrets. Doctoral Thesis, Université Paris-Sud.
[15] MASSOULIÉ, L. (1998). Stability results for a general class of interacting point processes dynamics, and applications. *Stoch. Proc. Appl.* **75**, 1–30.
[16] ROZANOV, YU. A. (1967). *Stationary Random Processes*. Holden-Day, San Francisco.