Some non asymptotic tail estimates for Hawkes processes

Patricia Reynaud-Bouret   Emmanuel Roy

Abstract

We use the Poisson cluster process structure of a Hawkes process to derive non asymptotic estimates of the tail of the extinction time, of the coupling time or of the number of points per interval. This allows us to define a family of independent Hawkes processes; each of them approximating the initial process on a particular interval. Then we can easily derive exponential inequalities for Hawkes processes which can precise the ergodic theorem.

Introduction

The Hawkes processes have been introduced by Hawkes (1971). Since then they are especially applied to earthquake occurrences (Vere-Jones 1970), but have recently found applications to DNA modeling (Gusto & Schbath 2005). In particular, an assumption which was not very realistic for earthquakes is very reasonable in this framework: the support of the reproduction measure is known and bounded. The primary work is motivated by getting non asymptotic concentration inequalities for the Hawkes process, using intensively the bounded support assumption. Those concentration inequalities are fundamental to construct adaptive estimation procedure as the penalized model selection (Massart 2000, Reynaud-Bouret 2003). To do so, we study carefully in this paper the link between cluster length, extinction time and construction of an approximating family of independent processes. Doing the necessary computations, we find out that other possible assumptions are also giving nice estimates of those quantities. Those estimates allow us to give some non asymptotic

1991 *Mathematics Subject Classification :* 60G55.

Key words and phrases : Point processes, exponential inequalities, approximate simulation of a stationary Hawkes process.

answers to some problems studied by Brémaud, Nappo & Torrisi (2002) on approximate simulation. But first, let us start by presenting the model and giving the main notation.

A point process N is a countable random set of points on $\mathbb{R}$ without accumulation. In an equivalent way, N denotes the point measure, i.e. the sum of the Dirac measures in each point of N. Consequently, $N(\mathcal{A})$ is the number of points of N in $\mathcal{A}$, $N_{|\mathcal{A}}$ represents the points of N in $\mathcal{A}$; if N' is another point process, $N+N'$ is the set of points that are both in N and N'. The Hawkes process (Hawkes 1971) N^h is a point process whose intrinsic stochastic intensity is defined by:

$$\Lambda(t) = \lambda + \int_{-\infty}^{t^-} h(t-u)N(du) \tag{0.1}$$

where λ is a positive constant and h is a positive function with support in $\mathbb{R}_+$ such that $\int_0^{+\infty} h < 1$. We refer to Daley & Vere-Jones (1988) for the basic definitions of intensity and point process. We call h the reproduction function. The reproduction measure is $\mu(dt) = h(t)dt$, where dt represents the Lebesgue measure on the real line.

Hawkes & Oakes (1974) prove that N^h can be seen as a generalized branching process and admits a cluster structure. The structure is based on inductive constructions of the points of N^h on the real line, which can be interpreted, for a more visual approach, as births in different families. In this setup, the reproduction measure μ (with support in $\mathbb{R}_+$) is not necessarily absolutely continuous with respect to the Lebesgue measure. However, to avoid multiplicities on points (which would mean simultaneous births at the same date), we make the additional assumption that the measure is continuous.

The basic cluster process

Shortly speaking, considering the birth of an ancestor at time 0, the cluster associated to this ancestor is the set of births of all descendants of all generations of this ancestor, where the ancestor is included.

To fix the notation, let us consider an i.i.d. sequence $\{P_{i,j}\}_{(i,j)\in\mathbb{N}\times\mathbb{N}}$ of Poisson variables with parameter $p = \mu([0,\infty))$. Let us consider independently an i.i.d. sequence $\{X_{i,j,k}\}_{(i,j,k)\in\mathbb{N}\times\mathbb{N}\times\mathbb{N}}$ of positive variables with law given by μ/p. Let $m = \mathbb{E}(X_{i,j,k})$, $v = \mathrm{Var}(X_{i,j,k})$ and $\ell(t) = \log\left[\mathbb{E}(e^{tX_{i,j,k}})\right]$ if they exist.

We construct now the successive generations which constitute the Hawkes process. The 0th generation is given by the ancestor $\{0\}$. The number of births in this generation is $K_0 = 1$, the total number of births in the family until the 0th generation is $W_0 = 1$. The successive births in this generation are given by $\{X_1^0 = 0\}$.

By induction, let us assume that we have already constructed the $(n-1)$th generation, i.e. we know the following quantities: K_{n-1}, the number of births in the $(n-1)$th generation, W_{n-1}, the total number of births in the family until the $(n-1)$th generation with the addition of the successive births in the $(n-1)$th generation $\{X_1^{n-1}, \dots, X_{K_{n-1}}^{n-1}\}$.

Then the nth generation is constructed as follows:

- if $K_{n-1} = 0$ then the $(n-1)$th generation is empty and the nth generation does not exist. We set $K_n = 0$ and $W_n = W_{n-1}$.
- if $K_{n-1} > 0$ then
  - $K_n = P_{n,1} + \cdots + P_{n,K_{n-1}}$ is the number of births in the nth generation,
  - $W_n = W_{n-1} + K_n$ is the total number of births until the nth generation,
  - the births of the nth generations are given by
    $\{X_1^{n-1} + X_{n,1,1}\ ,\ \cdots\ ,\ X_1^{n-1} + X_{n,1,P_{n,1}}\}$
    which are the births of the children of the parent born at X_1^{n-1},
    $\{X_2^{n-1} + X_{n,2,1}\ ,\ \cdots\ ,\ X_2^{n-1} + X_{n,2,P_{n,2}}\}$
    which are the births of the children of the parent born at X_2^{n-1},
    $\vdots \quad \vdots \quad \vdots$
    $\{X_{K_{n-1}}^{n-1} + X_{n,K_{n-1},1}\ ,\ \cdots\ ,\ X_{K_{n-1}}^{n-1} + X_{n,K_{n-1},P_{n,K_{n-1}}}\}$
    which are the births of the children of the parent born at $X_{K_{n-1}}^{n-1}$.
    All these points are the births in the nth generation. We arrange them by increasing order to obtain $\{X_1^n, \ldots, X_{K_n}^n\}$, the successive births in the nth generation.

To make the notation clearer, $X_{i,j,k}$ is the time that the jth parent in the $(i-1)$th generation has waited before giving birth to his kth child (the children are not ordered by age).

The sequence $(K_n)_{n\in\mathbb{N}}$ is a Galton-Watson process (Athreya & Ney 2004) from an initial population of one individual and with a Poisson distribution of parameter p as reproduction law. Since $p < 1$, the Galton-Watson process is sub-critical and the construction reaches an end almost surely, i.e. almost surely, there exists $\mathcal{N}$ such that $K_{\mathcal{N}} = 0$. The cluster is then given by $\cup_{n=0}^{\mathcal{N}}\{X_1^n, \ldots, X_{K_n}^n\}$. We denote this point process by N^c.

Hawkes process as Poisson cluster process

We are now considering the general case where numerous ancestors coexist and produce, independently of each others their own family. Let N^a be a Poisson process on $\mathbb{R}$ of intensity measure ν, which corresponds to the births of the different ancestors. Let us call the successive births of the ancestors $-\infty \le \cdots < T_{-1} < T_0 \le 0 < T_1 < \cdots \le +\infty$ where the eventual unnecessary points are rejected at infinity (this happens if there is a finite number of points).

Let us consider now independently an i.i.d. collection $\{N_n^c\}_{n\in\mathbb{Z}}$ of cluster processes constructed as previously according to the reproduction measure μ. Let us denote by $\{T_j^n, j \in \mathbb{N}\}$ the successive births in the cluster process N_n^c.

The Hawkes process N^h with ancestor measure ν and reproduction measure μ is given by $\cup_{n\in\mathbb{Z}} \cup_{j\in\mathbb{N}} \{T_n + T_j^n\}$, $T_n \in \mathbb{R}$. Heuristically, the points of N^h can be seen as the births in the different families: a family corresponding to one ancestor and all his progeny.

The case $\nu(dt) = \lambda dt$ corresponds to the stationary version of the Hawkes process. The intensity of N^h is given by (0.1) when $\nu(dt) = \lambda dt$ and $\mu(dt) = h(t)dt$ where dt is the Lebesgue measure on the real line.

When there is no possible confusion, N^h will always denote the Hawkes process with ancestor measure ν and reproduction measure μ. When several measures may coexist, we will denote the law of N^h, seen as a random variable on the point measures, by $\mathrm{H}(\nu,\mu)$.

A most important consequence of the Poisson cluster process structure of the Hawkes process is the superposition property (a straightforward consequence of (2.1)).

Proposition 0.1 (Superposition property)**.** *Let N_1^h and N_2^h be two independent Hawkes processes, respectively with distributions $\mathrm{H}(\nu_1,\mu)$ and $\mathrm{H}(\nu_2,\mu)$. Then $N^h = N_1^h + N_2^h$ is a Hawkes process with distribution $\mathrm{H}(\nu_1+\nu_2,\mu)$.*

In the first section, we intensively study the cluster process and we obtain some tail estimates for various quantities. In the second section, we apply these results to the Hawkes process. In the third section, we use the previous results to get a time cutting of the Hawkes process in approximating independent pieces and we apply this to get some non asymptotic estimates of the rate of convergence in the ergodic theorem.

1 Study of one cluster

1.1 Length of a cluster

Let N^c be a cluster process constructed as before. Let us denote by H the length of a cluster (i.e. the latest birth in the family), then H is given by

$$H = \sup_{j\leq K_n, n\in\mathbb{N}} X_j^n.$$

If H is quite naturally a.s. finite by construction, the question of integrability is not that clear. First of all, let us remark that if the $X_{i,j,k}$'s are not integrable, then of course H is not integrable, as soon as $p>0$. Now let us assume that the $X_{i,j,k}$'s are integrable. Let us define

$$U_n = \sum_{k=1}^{K_{n-1}} \sum_{j=1}^{P_{n,k}} X_{n,k,j}.$$

Clearly, U_1 is an upper bound of the latest birth in the first generation; U_1+U_2 is an upper bound of the latest birth until the second generation and by induction, $\mathcal{U} = \sum_{n=1}^{\infty} U_n$ is clearly an upper bound for H. By independence between the $X_{i,j,k}$ and the $P_{i,j}$, one has that $\mathbb{E}(U_n) = m\mathbb{E}(K_n)$. But, by induction (Athreya & Ney 2004), $\mathbb{E}(K_n) = p\mathbb{E}(K_{n-1}) = p^n$. Thus, $\mathbb{E}(\mathcal{U}) \leq m/(1-p) < \infty$, as soon as $p<1$. We can then easily get the following proposition.

Proposition 1.1. *Assume that $0<p<1$. The length of the cluster, H, is integrable if and only if m is finite.*

But if we need a good estimate for the tail of H, we have to look closer. We already remarked that the sequence $(K_n)_{n\in\mathbb{N}}$ is a sub-critical Galton-Watson process. Consequently the Laplace transform of W_∞ exists and satisfies this well known equation (Istas 2000, Athreya & Ney 2004)

$$L_W(t) = t + p\left(e^{L_W(t)} - 1\right) \tag{1.1}$$

for all t such that $L_W(t) = \log[\mathbb{E}(e^{tW_\infty})]$ is finite. Let us denote $g_p(u) = u - p(e^u - 1)$ for all $u > 0$. Then it is easy to see that for all $0 \leq t \leq (p - \log p - 1)$,

$$L_W(t) = g_p^{-1}(t), \tag{1.2}$$

where g_p^{-1} is the reciprocal function of g_p and that if $t > (p - \log p - 1)$, $L_W(t)$ is infinite. We can now apply this to derive tail estimates for H.

Proposition 1.2.

- *If* $v < +\infty$ *then for all positive* x,

$$\mathbb{P}(H > x) \leq \frac{1}{x^2}\left(\frac{p}{1-p}v + \left(\frac{p}{(1-p)^3} + \frac{p^2}{1-p}\right)m^2\right).$$

- *If there exists an interval* I *such that for all* $t \in I$, $l(t) \leq p - \log p - 1$, *then for all positive* x,

$$\mathbb{P}(H > x) \leq \exp\left(-\sup_{t\in I}\left[xt + l(t) - g_p^{-1}(l(t))\right]\right).$$

In particular if there exists $t > 0$ *such that* $l(t) \leq p - \log p - 1$ *then*

$$\mathbb{P}(H > x) \leq \exp\left[-xt + 1 - p\right].$$

- *If* $\mathrm{Supp}(\mu) \subset [0, A]$, *then*

$$\forall x \geq 0, \quad \mathbb{P}(H > x) \leq \exp\left[-\frac{x}{A}(p - \log p - 1) + 1 - p\right]$$

Proof. Let $(Y_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. variables with law μ/p, independent of the Hawkes process. Then $\sum_{n=1}^{W_\infty - 1} Y_n$, with the Y_n's independent of W_∞, has the same law as $\mathcal{U}$ which is an upper bound for H. But for all $t \in I$, conditioning in W_∞:

$$\mathbb{E}\left[\exp\left(t\sum_{n=1}^{W_\infty-1} Y_n\right)\right] = \mathbb{E}\left[\exp(l(t)(W_\infty - 1))\right] = \exp(-l(t))\exp[g_p^{-1}(l(t))],$$

and differentiating L_W to get the moment of W_∞, one also gets that

$$\mathbb{E}\left[\left(\sum_{n=1}^{W_\infty-1} Y_n\right)^2\right] = \frac{p}{1-p}v + \left(\frac{p}{(1-p)^3} + \frac{p^2}{1-p}\right)m^2.$$

It is now sufficient to use Chebyshev's inequality to conclude the proof for the first two results. For the last result it is sufficient to note that $\ell(t) \leq tA$. ∎

1.2 Exponential decreasing of the number of points in the cluster

Now we would like to understand more precisely the distribution of the points of N^c. More precisely we would like to prove that $N^c([a;+\infty))$ is exponentially decreasing in a in some sense. The probability generating functional of a point process is a well known tool which is equivalent to the log-Laplace transform and which helps us here. For any bounded function f, let us define

$$L(f) = \log\left[\mathbb{E}\left(\exp\left[\int f(u)N^c(du)\right]\right)\right].$$

Then Daley & Vere-Jones (1988) gives for the Hawkes process that:

$$L(f) = f(0) + \int_0^\infty \left[e^{L(f(t+.))} - 1\right]\mu(dt),$$

where

$$L(f(t+.)) = \log\left[\mathbb{E}\left(\exp\left[\int f(t+u)N^c(du)\right]\right)\right].$$

Let $z > 0$ and $a \geq 0$ and let us apply this formula to $f = z\mathbb{1}_{[a;+\infty)}$. Then

$$U(a,z) = L(f) = \log\left[\mathbb{E}\left(e^{zN^c([a;+\infty))}\right)\right]$$

is the log-Laplace transform of the number of births after a. We are assuming in this section that $\mathrm{Supp}(\mu) \subset [0,A]$, then for all $a > 0$

$$U(a,z) = \int_0^A (e^{U(a-t,z)} - 1)\mu(dt). \tag{1.3}$$

Let us remark that the function $U(a,z)$ is decreasing in a, since the number of remaining births is decreasing. Moreover, $U(0,z)$ is the log-Laplace transform of W_∞. The previous computations give that for all $0 \leq z \leq (p - \log p - 1)$, $U(0,z) = g_p^{-1}(z)$. Moreover if we define $U(+,z) = \log\left[\mathbb{E}\left(e^{zN^c((0;+\infty))}\right)\right]$, since the ancestor is always in 0, this quantity satisfies for all $0 \leq z \leq (p - \log p - 1)$,

$$U(+,z) = g_p^{-1}(z) - z.$$

Hence, for all $0 < a < A$ and for all $0 \leq z \leq (p - \log p - 1)$,

$$U(a,z) \leq U(+,z) = g_p^{-1}(z) - z. \tag{1.4}$$

Let us prove by induction the following result which gives a sense to "the number of births after a is exponentially decreasing in a".

Proposition 1.3. *Assume that* $\mathrm{Supp}(\mu) \subset [0,A]$. *For all* $a > 0$, *let* $k = \lfloor a/A \rfloor$. *Then for all* $0 \leq z \leq (p - \log p - 1)$,

$$U(a,z) \leq \left(g_p^{-1}(z) - z\right)e^{-kz}. \tag{1.5}$$

Proof. We already checked this fact for $k = 0$. Let us assume that the second inequality holds for k and let us prove it for $k+1$. As $U(a,z)$ is decreasing in a, one has that $U(a,z) \leq U((k+1)A,z)$. Applying (1.3) and (1.5), since μ is continuous, one has for all $0 \leq z \leq (p - \log p - 1)$,

$$U((k+1)A,z) \leq p\left(\exp\left[(g_p^{-1}(z) - z)e^{-kz}\right] - 1\right).$$

But for all $a \leq 1$ and $x \geq 0$,

$$e^{ax} - 1 \leq a(e^x - 1). \tag{1.6}$$

Moreover one has $g_p^{-1}(z) \geq z$, since their inverses are in the inverse order. Consequently for all $0 \leq z \leq (p - \log p - 1)$,

$$U((k+1)A,z) \leq pe^{-kz}\left(\exp\left[g_p^{-1}(z) - z\right] - 1\right).$$

But

$$e^{g_p^{-1}(z)} = 1 + \frac{g_p^{-1}(z) - z}{p}.$$

Hence for all $0 \leq z \leq (p - \log p - 1)$,

$$\begin{aligned} U((k+1)A,z) &\leq pe^{-kz}\left(\left[1 + \frac{g_p^{-1}(z) - z}{p}\right]e^{-z} - 1\right) \\ &\leq (g_p^{-1}(z) - z)e^{-(k+1)z} + pe^{-kz}(e^{-z} - 1). \end{aligned}$$

Since the last term is negative, this completes the proof. ■

2 Consequences for the Hawkes process

Let us now look at the consequences for the Hawkes process of these results.

2.1 Application to the number of points per interval

One of the first applications is really straightforward. It is based on the link between the different probability generating functionals. Let us define for all bounded functions f,

$$\mathcal{L}(f) = \log\left[\mathbb{E}\left(\exp\left[\int f(u)N^h(du)\right]\right)\right].$$

Then Vere-Jones (1970) proves that

$$\mathcal{L}(f) = \int_{-\infty}^{+\infty}\left(e^{L(f(t+.))} - 1\right)\nu(dt). \tag{2.1}$$

Let $z > 0$ and $T > 0$. Let us apply this formula to $f = z\mathbb{1}_{[0,T]}$. Then $\mathcal{L}(f)$ is the log-Laplace transform of the number of points of the Hawkes process between 0 and T. But $L(f(t+.))$ is the log-Laplace transform of the number of births of the cluster N^c between $-t$ and $T-t$. Consequently

- if $t > T$, $L(f(t+.)) = 0$,
- if $T \geq t \geq 0$, $L(f(t+.))$ can be upper bounded by the log-Laplace transform of W_∞, i.e. $U(0,z)$.
- if $0 > t$, $L(f(t+.))$ can be upper bounded by the log-Laplace transform of the number of births of the cluster N^c after $-t$, i.e. $U(-t,z)$.

This leads to

$$\mathcal{L}(f) \leq \int_{-\infty}^{0} \left(e^{U(-t,z)} - 1\right) d\nu_t + \left(\int_0^T d\nu_t\right)\left(e^{U(0,z)} - 1\right).$$

If we assume that the ancestors are "uniformly" distributed, one can prove the following fact.

Proposition 2.1. *Let us assume that* $\nu(dt) = \lambda dt$ *and* $\mathrm{Supp}(\mu) \subset [0,A]$. *Let* $0 \leq z \leq (p - \log p - 1)$ *and* $T > 0$. *Then*

$$\log\left[\mathbb{E}\left(e^{zN^h([0,T])}\right)\right] \leq \lambda T \ell_0(z) + \lambda A \ell_1(z)$$

where $\ell_0(z) = e^{g_p^{-1}(z)} - 1$ *and* $\ell_1(z) = \dfrac{e^{g_p^{-1}(z)-z}}{1-e^{-z}} - 1$. *Moreover for all integer* n

$$\mathbb{P}(N^h([0,T]) \geq n) \leq \exp\left[-nz + \lambda T \ell_0(z) + \lambda A \ell_1(z)\right]. \tag{2.2}$$

Proof. We know that $U(0,z) = g_p^{-1}(z)$. Now let us split the integral into pieces of length A and use the fact that $U(a,z)$ is decreasing in a. This gives

$$\log\left[\mathbb{E}\left(e^{zN^h([0,T])}\right)\right] \leq \lambda T\left[e^{g_p^{-1}(z)} - 1\right] + \sum_{k=0}^{\infty} \int_{kA}^{(k+1)A} \lambda(e^{U(t,z)} - 1)dt.$$

Let us apply Proposition 1.3. This gives, using (1.6),

$$\begin{aligned}\log\left[\mathbb{E}\left(e^{zN^h([0,T])}\right)\right] &\leq \lambda T\left[e^{g_p^{-1}(z)} - 1\right] + \sum_{k=0}^{\infty} \lambda A\left(\exp\left[(g_p^{-1}(z) - z)e^{-kz}\right] - 1\right) \\ &\leq \lambda T\left[e^{g_p^{-1}(z)} - 1\right] + \sum_{k=0}^{\infty} \lambda A e^{-kz}\left(\exp[g_p^{-1}(z) - z] - 1\right).\end{aligned}$$

This easily concludes the proof. ∎

2.2 Application to the extinction time

Another important quantity on the Hawkes process is the extinction time T_e. Let us define a Hawkes process N^h with reproduction measure μ and ancestor measure $\nu = \lambda \mathbb{1}_{\mathbb{R}_-} dt$. i.e. the ancestors appear homogeneously before 0 but not after. The latest birth in this process is the extinction time T_e. How fast does $\mathbb{P}(T_e > a)$ decrease in a ?

We keep the notation given in the introduction and we define H_n the length of the cluster N_n^c. Then $T_e = \sup_{n \in \mathbb{Z}_-}\{T_n + H_n\}$. So one can easily compute $\mathbb{P}(T_e \leq a)$ for any positive a. By conditioning with respect to the ancestors and using Proposition 1.1, one gets the following result, which seems to be known for a while (see for instance Møller & Rasmussen 2004*a*).

Proposition 2.2. *Let* $0 < p < 1$. *For all* $a \geq 0$, *one has*

$$\mathbb{P}(T_e \leq a) = \exp\left(-\lambda \int_a^{+\infty} \mathbb{P}(H > x)dx\right).$$

Moreover, the extinction time, T_e, *is finite if and only if the reproduction measure,* μ, *satisfies* $\int_0^{+\infty} t\mu(dt) < \infty$.

Since we have now good estimates for the cluster length, we get the following bounds under various assumptions, simply using Chebyshev's inequality.

Proposition 2.3.

- *Assume that* $\left(\frac{p}{1-p}v + \left(\frac{p}{(1-p)^3} + \frac{p^2}{1-p}\right)m^2\right) = c$ *is finite, then*

$$\mathbb{P}\left(T_e > a\right) \leq 1 - \exp\left[-\lambda \min\left(2\sqrt{c} - a, \frac{c}{a}\right)\right] \leq \lambda \min\left(2\sqrt{c} - a, \frac{c}{a}\right)$$

- *Assume that there exists* $t > 0$ *such that* $l\left(t\right) \leq p - \log p - 1$, *then*

$$\mathbb{P}\left(T_e > a\right) \leq 1 - \exp\left[-\frac{\lambda}{t}e^{-at+1-p}\right] \leq \frac{\lambda}{t}e^{-at+1-p}$$

- *Assume that* $\mathrm{Supp}(\mu) \subset [0, A]$, *then*

$$\mathbb{P}\left(T_e > a\right) \leq 1 - \exp\left[-\frac{\lambda A e^{-\frac{a}{A}(p-\log p-1)+1-p}}{p - \log p - 1}\right] \leq \frac{\lambda A e}{p - \log p - 1}e^{-\frac{a}{A}(p-\log p-1)}. \tag{2.3}$$

2.3 Superposition property and approximate simulation

As it has been said in the introduction, Hawkes processes model a lot of different problems. It is so natural to look for theoretical validation of simulation procedures. To simulate a stationary Hawkes process on $\mathbb{R}_+$ (that is, the restriction of $\mathrm{H}(\lambda dt, \mu)$ to $\mathbb{R}_+$), it is classical to use the superposition property (Proposition 0.1): a stationary Hawkes process is the independent superposition of $\mathrm{H}(\lambda \mathbb{1}_{\mathbb{R}_-} dt, \mu)$ and $\mathrm{H}(\lambda \mathbb{1}_{\mathbb{R}_+} dt, \mu)$. This means that we have to simulate first a Hawkes process with ancestors after time 0, which is easy, and then make the correct adjustment by artificially adding, independently, the points coming from ancestors born before time 0, that is, points coming from the the restriction of $\mathrm{H}(\lambda \mathbb{1}_{\mathbb{R}_-} dt, \mu)$ to $\mathbb{R}_+$. But to create these points, one needs, a priori, the knowledge of the whole past. However, we know they are a.s. in finite number if and only if $\int_0^{\infty} t\mu\left(dt\right) < +\infty$ by Proposition 2.2 (this result can also be found in (Møller & Rasmussen 2004*b*)). Under this assumption, it is not surprising we will get a good approximation of the restriction of $\mathrm{H}(\lambda \mathbb{1}_{\mathbb{R}_-} dt, \mu)$ to $\mathbb{R}_+$ by using the restriction of $\mathrm{H}(\lambda \mathbb{1}_{[-a,0)} dt, \mu)$ to $\mathbb{R}_+$ for a large a. Finally, putting things together, we can approximate a stationary Hawkes process on $\mathbb{R}_+$ by looking at the restriction of $\mathrm{H}(\lambda \mathbb{1}_{[-a,+\infty)} dt, \mu)$ to $\mathbb{R}_+$. We can see that doing this, the error is easy to evaluate non asymptotically by means of the variation distance which, here, is less than $\mathbb{P}(T_e > a)$ where T_e still denotes the extinction time of the previous section. Proposition 2.3 then gives some explicit and non asymptotic values in various useful cases. This answers a question asked to the second author by Brémaud who previously, together with Nappo and Torrisi in (Bremaud et al. 2002) gave some asymptotic results for this error. In particular, they give in the exponential unmarked case, an asymptotic exponential rate of decreasing for the exinction time (see Proposition 2.3, result 2) which is larger than ours. It seems to us that the results of Proposition 2.3 are probably non sharp, but are giving answers in a non asymptotic way, that can be really useful in practice. The question of approximate and perfect simulation has also been considered by Møller & Rasmussen (2004*a*)(see also Møller & Rasmussen 2004*b*), however their setup is quite different and makes the comparison with our results very difficult.

3 Applications of the superposition property

3.1 Construction of approximating i.i.d. sequences

A Poisson process N^p is said to be completely independent, that is for instance, $N^p_{|\mathcal{A}}$, the set of points of N^p in $\mathcal{A}$, is independent of $N^p_{|\mathcal{B}}$, the set of points of N^p in $\mathcal{B}$ as soon as $\mathcal{A}$ and $\mathcal{B}$ are disjoint. For N^h, a Hawkes process with distribution $\mathrm{H}(\lambda dt, \mu)$, despite of a hidden independent structure explained earlier, the clusters overlap each others and such independence cannot happen. Nevertheless by looking at distant intervals we are very close to independence.

Let us assume that the reproduction measure (Proposition 2.2) is such that the extinction time is almost surely finite. Our aim is to build an independent sequence $\left\{M^x_q\right\}_{q\in\mathbb{N}}$ such that M^x_q has the distribution of $H(\lambda dt, \mu)$ restricted to $[2qx - a, 2qx + x)$, for $0 < a < x$ and the variation distance between the distribution of M^x_q and $N_{|[2qx-a,2qx+x)}$ is controlled. The form of the interval (( ] or [ ], etc...) has no impact since there is a.s. no point of the process at a given site (this is a consequence of stationarity which implies that the measure that counts the mean number of points on Borel sets is indeed a multiple of Lebesgue measure and thus, non atomic). Let $\left\{N^h_{q,n}\right\}_{(q,n)\in\mathbb{N}\times\mathbb{Z}}$ be independent Hawkes processes $\mathrm{H}(\lambda \mathbb{1}_{[-x+2nx,x+2nx)}dt, \mu)$ which means that the ancestors appears homogeneously only on the interval $[-x + 2nx, x + 2nx)$. We now form the following point processes:

$$N^h := \sum_{n=-\infty}^{n=+\infty} N^h_{0,n}, \text{ and for all } q \geq 1,\ N^h_q := \sum_{n=-\infty}^{n=q-1} N^h_{q,n} + N^h_{0,q}.$$

It is clear, from the superposition property (Proposition 0.1) that, for each $q \geq 1$, N^h_q is a Hawkes process with distribution $\mathrm{H}(\lambda \mathbb{1}_{(-\infty,2qx+x)}dt, \mu)$ and that N^h a Hawkes process with distribution $\mathrm{H}(\lambda dt, \mu)$. It is also clear that all the N^h_q's are independent, for $q \geq 1$. We now take M^x_q to be $N^h_q{}_{|[2qx-a,2qx+x)}$, the points of N^h_q in $[2qx - a, 2qx + x)$. It is clear from the construction that the M^x_q's are independent and that they all have the stationary distribution $\mathrm{H}(\lambda dt, \mu)$ restricted to an interval of length $x + a$.

Let $q \geq 1$. Let $S = N^h_{0,q\ |[2qx-a,2qx+x)}$,

$$S_1 = \sum_{n=-\infty}^{n=q-1} N^h_{q,n\ |[2qx-a,2qx+x)} \text{ and } S'_1 = \sum_{n=-\infty}^{n=q-1} N^h_{0,n\ |[2qx-a,2qx+x)}.$$

To evaluate the variation distance between M^x_q and $N^h_{|(2qx-a,2qx+x]}$, we can write $M^x_q = S + S_1$ and $N^h_{|(2qx-a,2qx+x]} = S + S'_1$. We have for all measurable subset $\mathcal{A}$ of the set of point measures:

$$\begin{aligned}\left|\mathbb{P}\left(M^x_q \in \mathcal{A}\right) - \mathbb{P}\left(N^h_{|[2qx-a,2qx+x)} \in \mathcal{A}\right)\right| &= \left|\mathbb{E}\left[1_{\{S+S_1\in\mathcal{A}\}} - 1_{\left\{S+S'_1\in\mathcal{A}\right\}}\right]\right| \\ &= \left|\mathbb{E}\left[\left(1_{\{S+S_1\in\mathcal{A}\}} - 1_{\left\{S+S'_1\in\mathcal{A}\right\}}\right)\left(1-\left(1_{S_1=\emptyset}1_{S'_1=\emptyset}\right)\right)\right]\right| \\ &\leq \mathbb{E}\left[\left(1-\left(1_{S_1=\emptyset}1_{S'_1=\emptyset}\right)\right)\right] = \left(1-\mathbb{P}\left[S_1=\emptyset\right]^2\right).\end{aligned}$$

Now we can remark that $S_1 = \emptyset$ if $\sum_{n=-\infty}^{q-1} N^h_{q,n}$ is extinct before $2qx - a$. By stationarity, this probability is larger than $\mathbb{P}\left(T_e \leq x - a\right)$. Consequently the variation distance is less than $\left[1 - \mathbb{P}\left(T_e \leq x - a\right)^2\right]$. It is then very easy to prove the following result.

Proposition 3.1. *Let* $0 < a < x$. *Let* N^h *be a Hawkes process with distribution* $\mathrm{H}(\lambda dt, \nu)$. *There exists an i.i.d. sequence* M^x_q *of Hawkes processes with distribution* $\mathrm{H}(\lambda dt, \nu)$ *restricted to* $[2qx - a, 2qx + x)$ *such that for all* q, *the variation distance between* M^x_q *and* $N^h_{|[2qx-a,2qx+x)}$ *is less than* $2\mathbb{P}(T_e > x - a)$ *as soon as the extinction time* T_e *of* N^h *is an almost surely finite random variable.*

3.2 Example of application

Let f be a measurable function of $N^h_{|[-a,0)}$. For instance, the intensity Λ of the process in 0 is a possible f with $a = A$, if $\mathrm{Supp}(\mu) = \mathrm{Supp}(hdt) \subset [0, A]$ (see (0.1)). Let $\{\theta_s\}_{s\in\mathbb{R}}$ be the flow induced by the stationarity of the Hawkes process. This implies that for instance if $f = \Lambda(0)$, $f \circ \theta_s = \Lambda(s)$ is the intensity in s. The Hawkes process is ergodic since it is a Poisson cluster process (p. 347 of Daley & Vere-Jones 1988), this means that for $f \in L^1$

$$\frac{1}{T}\int_0^T f \circ \theta_s ds \to_{T\to\infty} \mathbb{E}(f) \text{ a.s.}$$

We are interested in this subsection in bounding from above quantities such as:

$$\mathbb{P}\left(\frac{1}{T}\int_0^T f \circ \theta_s ds \geq \mathbb{E}(f) + u\right), \tag{3.1}$$

for any positive u, in order to get a "non asymptotic ergodic theorem".

Let $T > 0$, $k \in \mathbb{N}$ and $x > 0$ such that $T = 2kx$. Let us assume now that f has zero mean for care of simplicity. First let us remark by stationarity that:

$$\mathbb{P}\left(\frac{1}{T}\int_0^T f \circ \theta_s ds \geq u\right) \leq 2\mathbb{P}\left(\sum_{q=0}^{k-1}\int_{2qx}^{2qx+x} f \circ \theta_s ds \geq \frac{uT}{2}\right)$$

But $G_q = \int_{2qx}^{2qx+x} f \circ \theta_s ds$ is a measurable function of the points of N^h appearing in $[2qx - a, 2qx + x)$, denoted by $\mathfrak{F}(N_{|[2qx-a,2qx+x)})$. Let us now pick a sequence $\left\{M_q^x\right\}_{0,\dots,(k-1)}$ of i.i.d. stationary Hawkes processes restricted to an interval of length $a + x$ and let $F_q = \mathfrak{F}(M_q^x)$. We have consequently constructed an i.i.d sequence $\{F_q\}_{0,\dots,(k-1)}$ with the same law as the G_q's. Moreover, by Proposition 3.1, the sequence $\left\{M_q^x\right\}_{0,\dots,(k-1)}$ can be chosen such that $\mathbb{P}(F_q \neq G_q)$ is less $2\mathbb{P}(T_e > x - a)$. By using Proposition 3.1, one gets

$$\mathbb{P}\left(\frac{1}{T}\int_0^T f \circ \theta_s ds \geq u\right) \leq 2\left[\mathbb{P}\left(\frac{1}{k}\sum_{q=0}^{k-1} F_q \geq ux\right) + \mathbb{P}(\exists q, F_q \neq G_q)\right].$$

This leads to the following result.

Theorem 3.2. *Let N^h be a stationary Hawkes process with distribution $\mathrm{H}(\lambda dt, \mu)$. Let $T, a > 0$ and k a positive integer such that $0 < a < T/2k$. Let f be a measurable function of $N^h_{|[-a,0)}$ with zero mean and θ_s be the flow induced by N^h.*
Then there exists an i.i.d. sequence F_k with distribution $\int_0^{T/2k} f \circ \theta_s ds$ such that

$$\mathbb{P}\left(\frac{1}{T}\int_0^T f \circ \theta_s ds \geq u\right) \leq 2\mathbb{P}\left(\frac{1}{k}\sum_{q=0}^{k-1} F_q \geq \frac{uT}{2k}\right) + 4k\mathbb{P}\left(T_e > \frac{T}{2k} - a\right),$$

where T_e is the extinction time of a Hawkes process with law $\mathrm{H}(\lambda \mathbb{1}_{\mathbb{R}_-} dt, \mu)$.

Now to get precise estimates, we need extra assumptions. Here are just a few examples of the possible applications of our construction.

Proposition 3.3. *Let N^h be a stationary Hawkes process with distribution $\mathrm{H}(\lambda dt, \mu)$ such that $\mathrm{Supp}(\mu) \subset [0, A]$. Let θ_s be the flow induced by N^h.*
Let $a > 0$ and f be a measurable function of $N^h_{|[-a,0)}$ with zero mean. Let $u, T > 0$ such that

$$a \leq A(u + \log T)/(p - \log p - 1) \text{ and } 4A(u + \log T) \leq T(p - \log p - 1).$$

Then with probability larger than $1 - \left(2 + \frac{\lambda e}{u + \log T}\right) e^{-u}$,

1. *(Hoeffding) if there exist $B, b > 0$ such that $B \geq f \geq b$,*
$$\frac{1}{T}\int_0^T f \circ \theta_s ds \leq (B - b)\sqrt{\frac{4A(u + \log T)u}{T(p - \log p - 1)}}$$
2. *(Bernstein) if there exist $V, C > 0$ such that $\forall n \geq 2, \mathbb{E}(f^n) \leq \frac{n!}{2} V C^{n-2}$,*
$$\frac{1}{T}\int_0^T f \circ \theta_s ds \leq \sqrt{\frac{16VA(u + \log T)u}{T(p - \log p - 1)}} + \frac{8CAu(u + \log T)}{T(p - \log p - 1)}$$
3. *(Weak Bernstein) if there exists $V, B > 0$ such that $V \geq \mathbb{E}(f^2)$ and $-B \leq f \leq B$,*
$$\frac{1}{T}\int_0^T f \circ \theta_s ds \leq \sqrt{\frac{16VA(u + \log T)u}{T(p - \log p - 1)}} + \frac{8BAu(u + \log T)}{3T(p - \log p - 1)}.$$

Proof. First we need to apply Hoeffding or Bernstein inequalities (Massart 2005) to the first term in Proposition 3.1. It remains then to bound the extinction time using Equation (2.3). The only remaining problem is then to choose k. With the assumption on T and u there exists always an integer k such that $a \leq T/4k$ and

$$\frac{T(p - \log p - 1)}{8A(u + \log T)} \leq k \leq \frac{T(p - \log p - 1)}{4A(u + \log T)},$$

which concludes the proof. ■

First let us remark that the conditions on T are fulfilled as soon as T is large enough. One can also see that, as usual, under the same assumptions the "weak Bernstein" inequality is sharper than the "Hoeffding inequality". The construction and proof of these time cutting and application to concentration inequalities is mainly inspired by the work of Baraud, Comte & Viennet (2001) on autoregressive sequences. In particular the $\log T$ factor seems to be, by analogy, a weak loss with respect to the independent case.

Finally, we would like to give a nice estimate for an unbounded function f, which naturally appears: the intensity. First let us suppose that the reproduction measure $\mu(du)$ is given by $h(u)du$. Then the intensity of N^h with distribution $\mathrm{H}(\lambda dt, \mu)$ is given by (0.1). Let us assume that h has support in $[0, A]$ and that h is bounded by a positive constant H. Let us first remark that $f = \Lambda(0) \leq \lambda + HN^h((-A, 0])$. So bounding the intensity $\Lambda(s) = f \circ \theta_s$ can be done if we bound the number of points per interval of length A.

Let $K = \lceil (T + A)/A \rceil$. Let $\mathcal{N}$ be a positive number and

$$\Omega = \left\{ \forall k \in \{0, \ldots, K-1\}, N^h((-A + kA, kA]) \leq \mathcal{N} \right\}.$$

Then by Proposition 2.2, $\mathbb{P}(\Omega^c) \leq K e^{\lambda A[\ell_0(p - \log p - 1) + \ell_1(p - \log p - 1)]} e^{-\mathcal{N}(p - \log p - 1)}$.

Now let us apply Proposition 3.3 (Weak Bernstein) to $f = \Lambda(0) \wedge M - \mathbb{E}(\Lambda(0) \wedge M)$, where $M = \lambda + 2H\mathcal{N}$. As on Ω, $f = \Lambda(0)$, we get the following result.

Proposition 3.4.
Let N^h be a Hawkes process with distribution $\mathrm{H}(\lambda dt, h(t)dt)$ where h the reproduction function is bounded by H and has a support included in $[0, A]$. Let Λ be its intensity given by (0.1). Let $u > 0$. There exists a $T_0 > 0$ depending on A, u and p such that for all $T \geq T_0$, with probability larger than $1 - \left(3 + \frac{\lambda e}{u + \log T}\right) e^{-u}$,

$$\frac{1}{T}\int_0^T \Lambda(s)ds \leq \mathbb{E}(\Lambda(0)) + \sqrt{\frac{16\mathbb{E}(\Lambda(0)^2)A(u + \log T)u}{T(p - \log p - 1)}} + \frac{8u(\lambda + 2H\mathcal{N})(u + \log T)}{3T(p - \log p - 1)}$$

where

$$\mathcal{N} = \frac{\lambda A[\ell_0(p - \log p - 1) + \ell_1(p - \log p - 1)] + \log T + u}{p - \log p - 1}.$$

In view of the ergodic theorem, this result explains very precisely and non asymptotically, how far $\frac{1}{T}\int_0^T \Lambda(s)ds$ is from its expectation. This result gives a non asymptotic answer to a question asked to us by P. Brémaud on the existence of a C.L.T. for those quantities.

References

Athreya, K. B. & Ney, P. E. (2004), *Branching processes*, Dover Publications Inc. Reprint of the 1972 original [Springer, New York].

Baraud, Y., Comte, F. & Viennet, G. (2001), 'Adaptive estimation in autoregression or β-mixing regression via model selection', *Ann. Statist.* **29**(3), 839–875.

Brémaud, P., Nappo, G. & Torrisi, G. L. (2002), 'Rate of convergence to equilibrium of marked Hawkes processes', *J. Appl. Probab.* **39**(1), 123–136.

Daley, D. J. & Vere-Jones, D. (1988), *An introduction to the theory of point processes*, Springer Series in Statistics, Springer-Verlag, New York.

Gusto, G. & Schbath, S. (2005), F.A.D.O. : a statistical method to detect favored or avoided distances between occurrences of motifs using the Hawkes model. submitted.

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

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

Istas, J. (2000), *Introduction aux modélisations mathématiques pour les sciences du vivant. (Introduction to mathematical modeling for life sciences).*, Springer.

Massart, P. (2000), 'Some applications of concentration inequalities', *Ann. Fac. Sci. Toulouse Math. (6)* **9**(2), 245–303. Probability theory.

Massart, P. (2005), *Lectures on probability theory and statistics*, Lecture Notes in Mathematics, Springer-Verlag, http://www.math.u-psud.fr/ massart/. Lectures from the 33nd Probability Summer School held in Saint-Flour, July 7–24, 2003, Edited by Jean Picard, *to appear.*

Møller, J. & Rasmussen, J. G. (2004*a*), Approximate simulation of Hawkes processes. Methodology and Computing in Applied Probability, 8, 53 - 64.

Møller, J. & Rasmussen, J. G. (2004*b*), Perfect simulation of Hawkes processes, Advances in Applied Probability, 37, 629 - 646.

Reynaud-Bouret, P. (2003), 'Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities', *P.T.R.F.* **126**(1), 103–153.

Vere-Jones, D. (1970), 'Stochastic models for earthquake occurrence', *J. Roy. Statist. Soc. Ser. B* **32**, 1–62.

Patricia Reynaud-Bouret
DMA - ENS (Equipe Probabilités et Statistiques) Paris,
45 rue d'Ulm,
75230 Paris Cedex 05,
Patricia.Reynaud@ens.fr

Emmanuel Roy
DI - ENS (Equipe TREC) Paris
45 rue d'Ulm, 75230 Paris Cedex 05,
Emmanuel.Roy@ens.fr