



Exponential moments for Hawkes processes under minimal assumptions

Théo Leblanc

► To cite this version:

Théo Leblanc. Exponential moments for Hawkes processes under minimal assumptions. *Electronic Communications in Probability*, 2024, 29, pp.1-11. [⟨10.1214/24-ECP625⟩](#). [⟨hal-04527359v2⟩](#)

HAL Id: hal-04527359

<https://hal.science/hal-04527359v2>

Submitted on 20 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Exponential moments for Hawkes processes under minimal assumptions

Théo Leblanc *

Abstract

We prove that the number of points of a stationary linear Hawkes process lying in any bounded subset of the real line has exponential moments, without any other assumption than the one needed for existence of such stationary process, namely the spectral radius of the matrix of \mathbb{L}^1 norms of interaction functions is smaller than one. The proof relies on a mass transport principle argument. We also specify the dependence of the bounds with respect to the base rates and the matrix of \mathbb{L}^1 norms of interaction functions defining the Hawkes process and give a functional version of the result.

Keywords: Hawkes processes; exponential moments; cluster representation.

MSC classifications: 60G55; 60E15; 60J85.

1 Introduction

1.1 Motivation

Hawkes processes, introduced by Hawkes in 1971 in [7], are modelling the continuous time arrivals of some random $0 - 1$ events when the arrivals depend on the past events. Hawkes processes can be self exciting or self inhibiting. This flexibility allows Hawkes processes to model a large variety of phenomena, from the time occurrences of earthquakes [11] to the activity of the neural network of the brain [5, 12], to cite but of few. Existence of stationary linear Hawkes processes, which only models excitation, is guaranteed under the well known condition that the spectral radius of the matrix of \mathbb{L}^1 norms of interaction functions is smaller than one [2]. This condition is expressed in [Assumption 1](#) in [Section 1.2](#). From a statistical point of view, having controls on the moments of Hawkes processes is key to construct robust estimation procedures of the interaction parameters of the process. Exponential moments of stationary Hawkes processes have been proved under conditions such as bounded memory interactions [6, 13] or decay assumptions on the interaction function in the univariate case [14]. But, up to our knowledge, there are no results about the existence of exponential moments for Hawkes processes under the minimal spectral radius condition that guarantees the existence of stationary Hawkes processes. In this paper we prove that a stationary Hawkes process has exponential moments under the spectral radius assumption, regardless of the tails of the interaction functions. We also provide quantitative bound of the exponential moments.

Along this paper we denote \mathbb{R} the set of real numbers, \mathbb{N} the set of integers: $\mathbb{N} = \{0, 1, 2, \dots\}$ and $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$.

*Université Paris Dauphine PSL, France. Email : theo.leblanc@dauphine.psl.eu

1.2 Setting

A *point process* on the real line \mathbb{R} is a random set of points $N \subset \mathbb{R}$ such that almost surely, N is locally finite. Given $\mathbf{M} = \llbracket 1, M \rrbracket$ with M a positive integer, a \mathbf{M} -*multivariate point process* on \mathbb{R} is the collection $(N^m)_{m \in \mathbf{M}}$ of random locally finite subsets of \mathbb{R} . Under mild conditions, a multivariate point process (N^1, \dots, N^M) is characterised by its stochastic intensity process $(\lambda^1, \dots, \lambda^M)$, defined by

$$\mathbb{P}(N^m \text{ has a point in } [t, t + dt] \mid \mathcal{F}_t) = \lambda_t^m dt, \quad m = 1, \dots, M,$$

where $(\mathcal{F}_t)_{t \in \mathbb{R}}$ is the filtration generated by (N^1, \dots, N^M) . See [1] for more precise statements. Hawkes processes are specific point processes. To define them, we consider $\boldsymbol{\mu} := (\mu_m)_{m \in \mathbf{M}} \subset \mathbb{R}_+^{\mathbf{M}}$ and $\mathbf{h} := (h_{m'}^m)_{m, m' \in \mathbf{M}}$, with $h_{m'}^m : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ measurable functions.

Definition 1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ a probabilistic space and a filtration $(\mathcal{F}_t)_{t \in \mathbb{R}}$. A Hawkes process with parameters $\mathbf{M}, \boldsymbol{\mu}, \mathbf{h}$ is a multivariate point process $\mathbf{N} := (N^m)_{m \in \mathbf{M}}$ such that

- \mathbf{N} is adapted to $(\mathcal{F}_t)_{t \in \mathbb{R}}$,
- almost surely, for $m \neq m'$ in \mathbf{M} , N^m has no common points with $N^{m'}$,
- \mathbf{N} has a predictable intensity $\boldsymbol{\lambda} := (\lambda^m)_{m \in \mathbf{M}}$ given by

$$\lambda_t^m = \mu_m + \sum_{m' \in \mathbf{M}} \int_{-\infty}^{t^-} h_{m'}^m(t-s) dN_s^{m'} \quad m \in \mathbf{M}, t \in \mathbb{R}. \quad (1)$$

In the previous definition, the constants μ_m are called *base rates*, modelling the spontaneous rates of arrivals of points. The functions $h_{m'}^m$ that take into account the influence of each point process on the others are called *interaction functions*.

In [2], Brémaud and Massoulié proved in particular that such a linear Hawkes process, in the sense of Definition 1 exists and is stationary under the following assumption.

Assumption 1. The spectral radius of the matrix $\mathbf{H} = (\|h_{m'}^m\|_1)_{m', m \in \mathbf{M}}$ is strictly smaller than 1, ie $\text{SpR}(\mathbf{H}) < 1$.

As explained in the next section, this assumption is sufficient to have exponential moments.

1.3 Main result

Given a Hawkes process \mathbf{N} and $B \subset \mathbb{R}$ we are interested in the number of points lying in B :

$$N(B) = \sum_{m \in \mathbf{M}} N^m(B)$$

where $N^m(B) := \text{Card}(N^m \cap B)$. To state our main result we need the following definition.

Definition 2. Let A a square matrix. We say that A satisfies the property $\mathbf{Ge}(r, K)$ with $r, K \geq 0$ and denote $A \in \mathbf{Ge}(r, K)$ if

$$\forall n \in \mathbb{N}^*, \quad \|A^n\|_\infty \leq Kr^n,$$

where and $\|\cdot\|_\infty$ is the operator norm associated to the norm $\|\cdot\|_\infty$.

The well known Gelfand's Theorem states that for any square matrix A , for any norm $\|\cdot\|$, we have $\|A^n\|^{1/n} \xrightarrow{n \rightarrow \infty} \text{SpR}(A)$. Thus, for any $r > \text{SpR}(A)$ there exists $K < \infty$ such that $A \in \mathbf{Ge}(r, K)$. It follows that under Assumption 1, there exist r and K with $\text{SpR}(\mathbf{H}) < r < 1$ and $0 < K < \infty$ such that $\mathbf{H} \in \mathbf{Ge}(r, K)$. Now we can state the main result of this paper.

Theorem 1. Suppose that [Assumption 1](#) holds, ie the spectral radius of $\mathbf{H} = (\|h_{m'}^m\|_1)_{m',m}$ is strictly less than 1. Then there exists $\xi > 0$ depending only on \mathbf{H} such that for all $B \subset \mathbb{R}$ bounded, we have

$$\mathbb{E} \left[e^{\xi N(B)} \right] < \infty.$$

More precisely, if $B \subset [s, s + L)$ for some $s \in \mathbb{R}$ and $L > 0$, and if $\mathbf{H} \in \mathbf{Ge}(r, K)$ with $0 < r < 1$ and $0 \leq K < \infty$ then we have

$$\mathbb{E} \left[e^{u \xi_{r,K} N(B)} \right] \leq \exp \left(\left[\left(\frac{1+r}{2r} \right)^u - 1 \right] L \sum_{m \in M} \mu_m \right), \quad \text{for all } u \in [0, 1] \quad (2)$$

with

$$\xi_{r,K} = \frac{\log \left(\frac{1+r}{2r} \right)}{1 + \frac{2K}{1-r}}.$$

Some remarks are in order.

The bound given in (2) can be rewritten as follows:

$$\mathbb{E} \left[e^{t N(B)} \right] \leq \exp \left(\left[e^{t C_{r,K}} - 1 \right] L \sum_{m \in M} \mu_m \right), \quad 0 \leq t \leq \xi_{r,K} \quad (3)$$

with $C_{r,K} = 1 + \frac{2K}{1-r}$. Thus, the bound for $\mathbb{E} [e^{t N(B)}]$ is increasing with K and r , which is the expected behaviour.

In the univariate case, ie $\text{Card}(\mathbf{M}) = 1$, the matrix \mathbf{H} is in fact a scalar with value $\alpha = \|h\|_1$ where h is the (only one) interaction function. Since $\mathbf{H}^n = (\alpha^n)$ it is straightforward that $\mathbf{H} \in \mathbf{Ge}(\alpha, 1)$. If $\mathbf{H} = 0$, meaning that there are no interactions and the Hawkes process consists of independent homogeneous Poisson processes with rates μ_m for $m \in \mathbf{M}$, then we have $\mathbf{H} \in \mathbf{Ge}(r, 0)$ for all $r \in (0, 1)$. In this case, the exponential moments of $N([0, L])$ can be explicitly calculated and compared to the bound (2) as follows:

$$\exp \left((e^t - 1) L \sum_{m \in M} \mu_m \right) = \mathbb{E} \left[e^{t N([0, L])} \right] \underset{\text{by (2)}}{\leq} \exp \left((e^t - 1) L \sum_{m \in M} \mu_m \right) \quad t \geq 0.$$

Indeed, $C_{r,0} = 1$ and $\xi_{r,0} \xrightarrow{r \rightarrow 0^+} +\infty$. Thus, in this particular case, the bound (2) becomes an equality.

The dependency on M , the cardinal of \mathbf{M} , is hidden in the constant K of $\mathbf{Ge}(r, K)$. Indeed, the operator norm $\|\cdot\|_\infty$ takes into account the dimension of the matrix in the following sense: for a square matrix A of size $d \geq 1$, we have

$$\|A\|_\infty = \max_{i \in [1, d]} \sum_{j=1}^d |A_{ij}|.$$

Remark 1. The constant $\xi_{r,K}$, which comes from [Lemma 2](#) which states exponential moments of Galton Watson processes with Poisson offspring measure, may be not optimal. In particular, in the simplest univariate case of a Galton Watson process \mathcal{T} with offspring measure $\mathcal{Pois}(\alpha)$ where $0 \leq \alpha < 1$, if we denote $g(\xi) := \log \mathbb{E} [e^{\xi \text{Card}(\mathcal{T})}]$ then for $\xi > 0$, $g(\xi) < \infty$ if and only if the equation $x = \xi + \alpha(e^x - 1)$ has a positive solution for x and in this case $g(\xi)$ is given by the first positive

solution (on x) of $x = \xi + \alpha(e^x - 1)$. Given α , finding the largest ξ such that there exists a positive solution of the above equation cannot be solved with the usual functions. Thus the optimal $\xi_{r,K}$ can not be expressed in closed form. However, the above reasoning shows (with the proof of [Theorem 1](#)) that for each matrix \mathbf{H} such that $\text{SpR}(\mathbf{H}) < 1$, there exists a $\xi_{\mathbf{H}} > 0$ satisfying $\mathbb{E}[e^{\xi_{\mathbf{H}} N(B)}] < \infty$ for all B bounded and $\mathbb{E}[e^{\xi N(B)}] = \infty$ for any $\xi > \xi_{\mathbf{H}}$ and B non-zero interval.

1.4 Functional result

Let f a non negative real function and define

$$N(f) := \sum_{m \in \mathbf{M}} \sum_{x \in N^m} f(x).$$

The following result states exponential moments of $N(f)$.

Theorem 2. *Let f a non negative real function, and $T > 0$. Suppose that $\mathbf{H} \in \mathbf{Ge}(r, K)$ with $0 < r < 1$ and $0 \leq K < \infty$. Then we have*

$$\mathbb{E}[e^{\xi N(f)}] \leq \exp\left(\left[e^{\xi \|f\|_{1,T}^{\infty} C_{r,K}} - 1\right] T \sum_{m \in \mathbf{M}} \mu_m\right), \quad 0 \leq \xi \leq \xi_{r,K} / \|f\|_{1,T}^{\infty}$$

where $\xi_{r,K}$ and $C_{r,K}$ are defined in [Theorem 1](#) and

$$\|f\|_{1,T}^{\infty} = \sup_{t \in \mathbb{R}} \sum_{n \in \mathbb{Z}} f(t + nT).$$

[Theorem 2](#) is an extension of [Theorem 1](#). Indeed to recover the classical exponential moments, one can take $f = \mathbf{1}_B$ and $T = L$.

The $\|\cdot\|_{1,T}^{\infty}$ is similar to a discrete L^1 norm for non negative functions with discretization scale T .

The proof of [Theorem 2](#) is similar to the proof of [Theorem 1](#) and is based on the same arguments. Thus we only give the key details of the proof.

Before giving the proof of [Theorem 1](#), we recall in the next section the cluster representation of Hawkes processes, which is crucial to prove our result.

2 Cluster representation

The *cluster representation*, introduced by Hawkes and Oakes in 1974 [\[8\]](#), is a powerful tool to study Hawkes processes. In a few words, the cluster representation means that, under [Assumption 1](#), a Hawkes process as introduced in [Definition 1](#) can be seen as a branching Poisson process on homogeneous Poisson processes with rates μ_m for $m \in \mathbf{M}$. Points generated by the homogeneous Poisson processes are called immigrants points. Points generated by the branching Poisson processes are called offspring points. Each immigrant point may generate new points, called its children, due to the self exciting term of [\(1\)](#). Then each child may also have some children and so on. These successive arrivals of offspring points, with its genealogy, is called a cluster. In the sequel we define more formally the clusters.

We use the Ulam Harris Neveu notation to label the individuals of trees. Let $\mathcal{U} = \cup_{n \in \mathbb{N}} (\mathbb{N}^*)^n$ with $(\mathbb{N}^*)^0 = \{\emptyset\}$. A $\mathcal{Pois}(\mathbf{h})$ -cluster with root of type $m_0 \in \mathbf{M}$ and born at time $t_0 \in \mathbb{R}$ is defined by $((u, \text{tp}(u), \text{bd}(u)))_{u \in \mathcal{T}}$ where

- $\mathcal{T} \subset \mathcal{U}$ is a random tree, for $u \in \mathcal{T}$, $\text{tp}(u) \in \mathbf{M}$ is the type of u and $\text{bd}(u) \in \mathbb{R}$ is the birth date of u .

- $((u, \text{tp}(u)))_{u \in \mathcal{T}}$ is a $\mathcal{Pois}(\mathbf{H})$ -Galton Watson tree with root \emptyset of type m_0 (ie $\text{tp}(\emptyset) = m_0$). It means that independently for each point $u \in \mathcal{T}$, if we denote $X(u, \text{tp}(u), m)$ the number of children of u with type $m \in \mathbf{M}$, then for $m \in \mathbf{M}$ the random variables $X(u, \text{tp}(u), m)$ are independent and $X(u, \text{tp}(u), m) \sim \mathcal{Pois}(\mathbf{H}_{\text{tp}(u), m})$ for all $m \in \mathbf{M}$.
- Conditionally to $((u, \text{tp}(u)))_{u \in \mathcal{T}}$, the birth dates are distributed as follows. For the root of \mathcal{T} we have $\text{bd}(\emptyset) = t_0$. Then, independently, for any $u, v \in \mathcal{T}$ with v a child of u , we have

$$\text{bd}(v) - \text{bd}(u) \sim \frac{h_{\text{tp}(u)}^{\text{tp}(v)}(s)}{\int h_{\text{tp}(u)}^{\text{tp}(v)}(t) dt} \mathbf{1}_{s>0} ds.$$

Finally, let π^m be independent Poisson processes on \mathbb{R} with rates μ_m for $m \in \mathbf{M}$. They are the *immigrant points*. Conditionally to $(\pi^m)_{m \in \mathbf{M}}$, let G_x^m be independent $\mathcal{Pois}(\mathbf{h})$ -clusters with root of type m and born at time x for $m \in \mathbf{M}$ and $x \in \pi^m$. The following result is known [8] and provides the cluster representation for Hawkes processes.

Proposition 1. *Under Assumption 1, if $(N^m)_{m \in \mathbf{M}}$ is a Hawkes process from Definition 1, then we have the following equality in distribution*

$$(N^m)_{m \in \mathbf{M}} = (\{s \in \mathbb{R} \mid \exists m' \in \mathbf{M}, \exists x \in \pi^{m'}, \exists u \in \mathcal{U} \text{ such that } (u, m, s) \in G_x^{m'}\})_{m \in \mathbf{M}}$$

where the immigrants $(\pi^m)_{m \in \mathbf{M}}$ are independent Poisson point processes with rates $\boldsymbol{\mu}$. Equivalently, in distribution, N^m is the collection of all the birth dates of any points of type m of any clusters generated by the immigrants points.

Figure 1 illustrates the cluster representation.

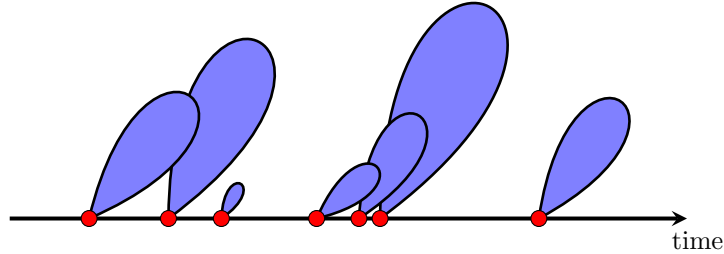


Figure 1: Illustration of the cluster representation. The red dots are the immigrant points. They arrive according to the immigrant rate, being μ_m for type $m \in \mathbf{M}$, and give birth to clusters, represented by the blue shapes. In the end, only the birth dates and the types are relevant for Hawkes processes, which can be interpreted as a projection of the blue shapes on the time axis, in the sense that the tree structure is forgotten

3 Proof of Theorem 1

In the sequel since Assumption 1 is satisfied we fix $r \in (0, 1)$ and $K \geq 0$ such that $\mathbf{H} \in \mathbf{Ge}(r, K)$. Thanks to Proposition 1 we can express a Hawkes process in terms of clusters.

Without loss of generality, due to stationarity, we can suppose that $B = [0, L]$. Let $\xi = u\xi_{r,K}$ with $u \in [0, 1]$ and $\xi_{r,K}$ defined in Theorem 1. Let $(\pi^m)_{m \in \mathbf{M}}$ be the immigrant points. For $x \in \pi^m$ let G_x^m be the cluster associated to x . For any cluster G and $A \subset \mathbb{R}$ we define

$$G(A) = \text{Card}(\{s \in A \mid \exists (u, m) \in \mathcal{U} \times \mathbf{M}, (u, m, s) \in G\}).$$

Then we have from [Proposition 1](#)

$$N(B) = \sum_{m \in \mathbf{M}} \sum_{x \in \pi^m} G_x^m([0, L]).$$

Until (5) the proof is an application of [3] Example 8.3(c). For sake of completeness we give the few lines of calculation since it contains a step which is crucial for the proof. Since conditionally to the immigrants the clusters are independent, we have

$$\mathbb{E} \left[e^{\xi N(B)} \right] = \mathbb{E} \left[\prod_{m \in \mathbf{M}} \prod_{x \in \pi^m} \mathbb{E} \left[e^{\xi G_x^m([0, L])} \mid (\pi^m)_{m \in \mathbf{M}} \right] \right].$$

Our proof relies on the following change of paradigm, which will lead to a mass transport principle [9], see [Figure 2](#). For $m \in \mathbf{M}$, let G^m a random variable with the law of a cluster with root of type m and born at $t = 0$. Then, since the clusters are invariant by time translation, we have for $m \in \mathbf{M}$ and $x \in \pi^m$,

$$\mathbb{E} \left[e^{\xi G_x^m([0, L])} \mid (\pi^m)_{m \in \mathbf{M}} \right] = \mathbb{E} \left[e^{\xi G^m([-x, -x+L])} \right] =: g(-x, m, \xi, L). \quad (4)$$

The left hand side of the formula represents the quantity that immigrant x gives to $B = [0, L]$. The right hand side represents what $B - x$ receives from an immigrant at $t = 0$, see [Figure 2](#).

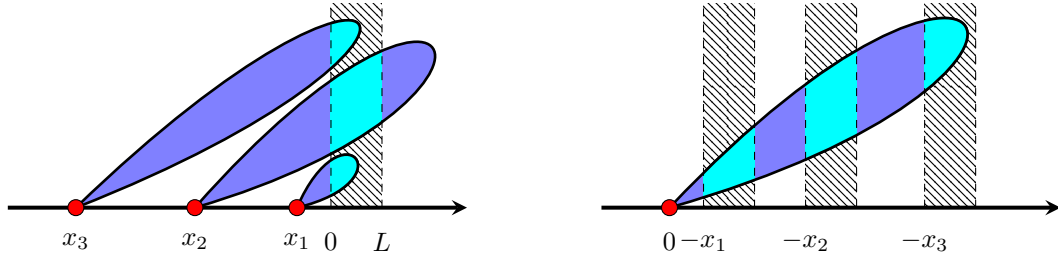


Figure 2: Illustration of the mass transport principle. In expectation, both side are equal.

We can now use Campbell Theorem for Poisson processes (see [10], section 3.2):

$$\begin{aligned} \mathbb{E} \left[e^{\xi N(B)} \right] &= \mathbb{E} \left[\prod_{m \in \mathbf{M}} \prod_{x \in \pi^m} g(-x, m, \xi, L) \right] \\ &= \exp \left(\sum_{m \in \mathbf{M}} \int_{-\infty}^L [g(-t, m, \xi, L) - 1] \mu_m dt \right) \\ &= \exp \left(\sum_{m \in \mathbf{M}} \int_{-L}^{\infty} [g(t, m, \xi, L) - 1] \mu_m dt \right). \end{aligned} \quad (5)$$

It remains to prove that $L^{-1} \int_{-L}^{\infty} [g(t, m, \xi, L) - 1] dt$ is finite and more precisely, bounded by $((1 + r)/2r)^u - 1$. For this purpose, in the literature, it is classical to use assumptions on the tail of the interaction functions, such as bounded support assumptions, to control the integrand for each t to prove that the integral is convergent, see [6]. Here we will prove the bound on the integral regardless of the tail of the interaction functions and show that the relevant quantity is only the matrix \mathbf{H} of the L^1 norms of the interaction functions. The following lemma is key and is made possible due to the mass transportation equality.

Lemma 1. Let $L > 0$ and a $\mathcal{Pois}(\mathbf{h})$ -cluster G^m from $m \in \mathbf{M}$ and started at $t = 0$. Then, for any $s \in [-L, 0)$ and any $\xi \geq 0$,

$$\sum_{n=0}^{\infty} \left(\mathbb{E} \left[e^{\xi G^m([s+nL, s+(n+1)L])} \right] - 1 \right) \leq \mathbb{E} \left[e^{\xi G^m([0, \infty))} \right] - 1.$$

Proof of Lemma 1. To prove this lemma, one just has to remark the following inequality. For any sequence y_1, y_2, \dots such that $y_n \geq 1$ for $n \geq 1$, we have

$$\sum_{n=1}^{\infty} (y_n - 1) \leq -1 + \prod_{n=1}^{\infty} y_n.$$

Indeed, it suffices to denote $x_n = y_n - 1 \geq 0$ and expand the product.

From this observation, we have

$$\begin{aligned} \sum_{n=0}^{\infty} \left(\mathbb{E} \left[e^{\xi G^m([s+nL, s+(n+1)L])} \right] - 1 \right) &= \mathbb{E} \left[\sum_{n=0}^{\infty} \left(e^{\xi G^m([s+nL, s+(n+1)L])} - 1 \right) \right] \\ &\leq \mathbb{E} \left[\prod_{n=0}^{\infty} e^{\xi G^m([s+nL, s+(n+1)L])} \right] - 1 \\ &= \mathbb{E} \left[e^{\xi G^m([0, \infty))} \right] - 1. \end{aligned}$$

This is the desired result. \square

Going back to the proof of Theorem 1. Let us rewrite the integral as follows

$$\int_{-L}^{\infty} [g(t, m, \xi, L) - 1] dt = \int_{-L}^0 \sum_{n=0}^{\infty} [g(t + nL, m, \xi, L) - 1] dt. \quad (6)$$

From Lemma 1 we have that

$$\sum_{n=0}^{\infty} [g(t + nL, m, \xi, L) - 1] \leq \mathbb{E} \left[e^{\xi G^m([0, \infty))} \right] - 1,$$

thus it follows that

$$\mathbb{E} \left[e^{\xi N(B)} \right] \leq \exp \left(L \sum_{m \in M} \mu_m \mathbb{E} \left[e^{\xi G^m([0, \infty))} \right] - 1 \right). \quad (7)$$

It remains to control $\mathbb{E} \left[e^{t G^m([0, \infty))} \right]$ with $t \geq 0$. At this point, the time embedding of G^m is no more relevant since we have the equality in law

$$G^m([0, \infty)) = \text{Card}(\mathcal{T}) \quad (8)$$

where $((u, \text{tp}(u)))_{u \in \mathcal{T}}$ is a $\mathcal{Pois}(\mathbf{H})$ -Galton Watson tree with root of type m . The next lemma is the last piece of the proof.

Lemma 2. Let $((u, \text{tp}(u)))_{u \in \mathcal{T}}$ be a $\mathcal{Pois}(\mathbf{H})$ -Galton Watson tree with root of any type. Suppose that $\mathbf{H} \in \mathbf{Ge}(r, K)$ with $r \in (0, 1)$ and $K \geq 0$, which means that

$$\|\mathbf{H}^n\|_{\infty} \leq Kr^n, \quad n \in \mathbb{N}^*.$$

Then we have,

$$\log \left(\mathbb{E} \left[\exp \left(t \text{Card}(\mathcal{T}) \right) \right] \right) \leq t \left[1 + \frac{2K}{1-r} \right] = \frac{t}{\xi_{r,K}} \log \left(\frac{1+r}{2r} \right)$$

for all $0 \leq t \leq \xi_{r,K}$ with

$$\xi_{r,K} = \frac{\log\left(\frac{1+r}{2r}\right)}{1 + \frac{2K}{1-r}}.$$

The proof of this result is postponed at the end of this paper.

Thus, up to [Lemma 2](#), the proof is complete. Indeed from [Lemma 2](#) and (8), we have for $t = \xi = u\xi_{r,K}$,

$$\mathbb{E}\left[e^{\xi G^m([0,\infty))}\right] \leq \left(\frac{1+r}{2r}\right)^u. \quad (9)$$

Finally, injecting (9) in (7) concludes the proof of [Theorem 1](#).

3.1 Proof of [Lemma 2](#)

It remains to prove [Lemma 2](#). In [4], the authors proved a similar result, stating that a $\mathcal{Pois}(\mathbf{H})$ -Galton Watson tree has exponential moments if the spectral norm of \mathbf{H} is smaller than one. [Lemma 2](#) is stronger since we only need the spectral radius of \mathbf{H} to be smaller than one (which is equivalent to require that $\mathbf{H} \in \mathbf{Ge}(r, K)$ with some $r < 1$). The proof is a bit more technical since we have to deal with the constant K , possibly $K > 1$, from the property $\mathbf{Ge}(r, K)$.

Proof of [Lemma 2](#). Let for $n \in \mathbb{N}$ and $m \in \mathbf{M}$,

$$g_n^m(t) = \log\left(\mathbb{E}\left[e^{tZ_n^m}\right]\right)$$

where Z_n^m is the cardinal a $\mathcal{Pois}(\mathbf{H})$ -Galton Watson tree with a root of type $m \in \mathbf{M}$ and clipped at generation n .

The classical branching property of Galton Watson trees states the following. Given any point $u \in \mathcal{T}$, define the sub-tree $\mathcal{T}_{\geq u}$ as the sub-tree of \mathcal{T} that contains u and all its offspring (children, grandchildren and so on). Then we have, conditionally to $(u \in \mathcal{T})$ and the type of u , the random variables $((v, \text{tp}(v)))_{v \in \mathcal{T}_{\geq u}}$ and $((v, \text{tp}(v)))_{v \in (\mathcal{T} \setminus \mathcal{T}_{\geq u}) \cup \{u\}}$ are independent and $((v, \text{tp}(v)))_{v \in \mathcal{T}_{\geq u}}$ is distributed as a $\mathcal{Pois}(\mathbf{H})$ -Galton Watson tree with root u of type $\text{tp}(u) \in \mathbf{M}$.

Applying the branching property to all the points of the first generation (the children of the root) leads, for all $n \geq 0$, to the following equality in distribution

$$Z_{n+1}^m = 1 + \sum_{m' \in \mathbf{M}} \sum_{k=1}^{X_{m,m'}} Z_n^{m',k}$$

where $X_{m,m'} \sim \mathcal{Pois}(\mathbf{H}_{m,m'})$ and $Z_n^{m',k} \sim Z_n^{m'}$ and the random variables $X_{m,m'}$, $Z_n^{m',k}$ for $m' \in \mathbf{M}$, $k \geq 1$ are all independent. Thus we obtain that for all $m \in \mathbf{M}$,

$$g_{n+1}^m(t) = t + \sum_{m' \in \mathbf{M}} \mathbf{H}_{mm'}(e^{g_n^{m'}(t)} - 1).$$

One also has that

$$g_0^m(t) = t.$$

Define $u_n^m(t) := g_{n+1}^m(t) - g_n^m(t) \geq 0$ and $u_n(t)$ the vector with coordinates $u_n^m(t)$ for $m \in \mathbf{M}$. Then

$$u_0^m(t) = (e^t - 1) \sum_{m'} \mathbf{H}_{mm'},$$

which can be written

$$u_0(t) = (e^t - 1)\mathbf{H}\mathbf{1}_M,$$

where $\mathbf{1}_M$ is the vector with only ones. Using the relation between g_n and g_{n+1} gives

$$\begin{aligned} u_{n+1}^m(t) &= \sum_{m' \in M} \mathbf{H}_{mm'} (e^{g_{n+1}^{m'}(t)} - e^{g_n^{m'}(t)}) \\ &\leq \sum_{m' \in M} \mathbf{H}_{mm'} e^{g_{n+1}^{m'}(t)} (g_{n+1}^{m'}(t) - g_n^{m'}(t)) \\ &\leq \exp(\|g_{n+1}(t)\|_\infty) \sum_{m' \in M} \mathbf{H}_{mm'} u_n^{m'}(t) \\ &\leq \exp\left(t + \sum_{j=0}^n \|u_j(t)\|_\infty\right) \sum_{m' \in M} \mathbf{H}_{mm'} u_n^{m'}(t), \end{aligned}$$

which becomes

$$u_{n+1}(t) \leq \exp\left(t + \sum_{j=0}^n \|u_j(t)\|_\infty\right) \times \mathbf{H} u_n(t).$$

Unfolding the previous equation gives

$$\begin{aligned} u_n(t) &\leq \exp\left(\sum_{i=1}^n \left[t + \sum_{j=0}^{i-1} \|u_j(t)\|_\infty\right]\right) \times \mathbf{H}^n u_0(t) \\ &= (e^t - 1) \exp\left(\sum_{i=1}^n \left[t + \sum_{j=0}^{i-1} \|u_j(t)\|_\infty\right]\right) \times \mathbf{H}^{n+1} \mathbf{1}_M. \end{aligned}$$

Let $0 \leq t \leq \log(1/r)$ and constants $K_1 > 0$, $r < r_1 < 1$ that will be fixed later. We will prove by induction on n that $\|u_n(t)\|_\infty \leq tK_1r_1^n$ if t is small enough (the precise condition is given at the end).

$n = 0$: Since $t \leq \log(1/r)$, we have $e^t - 1 \leq t/r$ and by $\mathbf{Ge}(r, K)$ we have $\|\mathbf{H}\mathbf{1}_M\|_\infty \leq Kr^1 \times 1 = Kr$. Thus if $K_1 \geq K$ it works.

$0, 1, \dots, n-1 \rightarrow n$: In this case we have on the one hand

$$\sum_{i=1}^n \left[t + \sum_{j=0}^{i-1} \|u_j(t)\|_\infty\right] \leq nt \left(1 + \frac{K_1}{1-r_1}\right).$$

On the other hand by $\mathbf{Ge}(r, K)$,

$$\|\mathbf{H}^{n+1} \mathbf{1}_M\|_\infty \leq Kr^{n+1}.$$

Thus, since $e^t - 1 \leq t/r$,

$$\|u_n(t)\|_\infty \leq tK \exp\left(n \left[\log(r) + t \left(1 + \frac{K_1}{1-r_1}\right)\right]\right).$$

The proof of the induction is complete if

$$\begin{cases} K_1 \geq K \\ \log(r) + t \left(1 + \frac{K_1}{1-r_1}\right) \leq \log(r_1) \Leftrightarrow t \leq \frac{\log(r_1/r)}{1 + \frac{K_1}{1-r_1}} \end{cases}$$

Thus we set $K_1 = K$ and arbitrarily choose $r_1 = (1+r)/2$. Thus the above bound for t is exactly $\xi_{r,K}$, leading to, for all $t \leq \xi_{r,K}$, for all $n \geq 0$, we have $\|u_n(t)\|_\infty \leq tKr_1^n$. Going back to $g_n(t)$ leads to

$$\forall m \in \mathbf{M}, \forall n \in \mathbb{N}, \forall t \leq \xi_{r,K}, g_n^m(t) \leq t \left(1 + \frac{K}{1-r_1}\right).$$

Letting $n \rightarrow \infty$ gives the intended result by monotone convergence theorem. \square

3.2 Proof of Theorem 2

By following the steps of the proof of Theorem 1 one obtains

$$\mathbb{E} \left[e^{\xi N(f)} \right] = \exp \left(\sum_{m \in \mathbf{M}} \int_{-\infty}^{\infty} [g(t, m, \xi, f) - 1] \mu_m dt \right)$$

where

$$g(t, m, \xi, f) = \mathbb{E} \left[e^{\xi G^m(f \circ \tau_t)} \right]$$

and G^m is a cluster born at time $t = 0$ and with root of type m and $\tau_t(x) = x + t$, $x \in \mathbb{R}$. Then we discretize the integral at scale T ,

$$\int_{-\infty}^{\infty} [g(t, m, \xi, f) - 1] dt = \int_0^T \sum_{n \in \mathbb{Z}} [g(t + nT, m, \xi, f) - 1] dt.$$

The same argument as the one for the proof of Lemma 1 leads to

$$\sum_{n \in \mathbb{Z}} [g(t + nT, m, \xi, f) - 1] \leq \mathbb{E} \left[e^{\xi G^m(\sum_{n \in \mathbb{Z}} f \circ \tau_{t+nT})} \right] - 1.$$

Since $\sum_{n \in \mathbb{Z}} f \circ \tau_{t+nT} \leq |f|_{1,T}^\infty$, it follows that

$$\mathbb{E} \left[e^{\xi N(f)} \right] \leq \exp \left(T \sum_{m \in \mathbf{M}} \mu_m \mathbb{E} \left[e^{\xi |f|_{1,T}^\infty G^m(\mathbb{R})} - 1 \right] \right).$$

Lemma 2 applied with $t = \xi |f|_{1,T}^\infty$ concludes the proof of Theorem 2.

References

- [1] Pierre Brémaud. *Point Processes and Queues*. Springer New York, 1981.
- [2] Pierre Brémaud and Laurent Massoulié. Stability of nonlinear hawkes processes. *The Annals of Probability*, pages 1563–1588, 1996.
- [3] Daryl J Daley, David Vere-Jones, et al. *An introduction to the theory of point processes: volume I: elementary theory and methods*. Springer, 2003.
- [4] Sophie Donnet, Vincent Rivoirard, and Judith Rousseau. Supplement to “nonparametric bayesian estimation for multivariate hawkes processes”. *The Annals of Statistics*, 48(5):2698 – 2727, 2020.
- [5] Antonio Galves and Eva Löcherbach. Infinite systems of interacting chains with memory of variable length—a stochastic model for biological neural nets. *Journal of Statistical Physics*, 151:896–921, 2013.
- [6] Niels Richard Hansen, Patricia Reynaud-Bouret, and Vincent Rivoirard. Lasso and probabilistic inequalities for multivariate point processes. *Bernoulli*, 21(1):83–143, 2015.

- [7] Alan G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58:83–90, 1971.
- [8] Alan G Hawkes and David Oakes. A cluster process representation of a self-exciting process. *Journal of applied probability*, 11(3):493–503, 1974.
- [9] Ali Khezeli. Unimodular random measured metric spaces and palm theory on them. 2023.
- [10] J F C Kingman. *Poisson Processes: Oxford Studies In Probability*. 3. Oxford University Press, 12 1992.
- [11] Yosihiko Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical association*, 83(401):9–27, 1988.
- [12] 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):1958–1990, 2020.
- [13] Patricia Reynaud-Bouret and Emmanuel Roy. Some non asymptotic tail estimates for hawkes processes. *Bulletin of the Belgian Mathematical Society-Simon Stevin*, 13(5):883–896, 2007.
- [14] François Roueff, Rainer von Sachs, and Laure Sansonnet. Locally stationary hawkes processes. *Stochastic Processes and their Applications*, 126(6):1710–1743, 2016.

Acknowledgments. I am grateful to my PhD supervisors, Vincent Rivoirard and Patricia Reynaud Bouret, for their advice and helpful comments. I would also like to thank the referee for his/her comments and pertinent questions, which have enabled me to significantly improve this paper.