REGENERATIVE PROPERTIES OF THE LINEAR HAWKES PROCESS WITH UNBOUNDED MEMORY

Carl Graham

REGENERATIVE PROPERTIES OF THE LINEAR HAWKES PROCESS WITH UNBOUNDED MEMORY

By Carl Graham

École polytechnique, CNRS, IP Paris

We prove regenerative properties for the linear Hawkes process under minimal assumptions on the transfer function, which may have unbounded support. These results are applicable to sliding window statistical estimators. We exploit independence in the Poisson cluster point process decomposition, and the regeneration times are not stopping times for the Hawkes process. The regeneration time is interpreted as the renewal time at zero of a M/G/infinity queue, which yields a formula for its Laplace transform. When the transfer function admits some exponential moments, we stochastically dominate the cluster length by exponential random variables with parameters expressed in terms of these moments. This yields explicit bounds on the Laplace transform of the regeneration time in terms of simple integrals or special functions yielding an explicit negative upper-bound on its abscissa of convergence. These regenerative results allow, *e.g.*, to systematically derive long-time asymptotic results in view of statistical applications. This is illustrated on a concentration inequality previously obtained with coauthors.

1. Introduction.

1.1. *Background.* Hawkes [17] introduced the point process on the real line bearing his name in order to model earthquakes. Primary shocks arrive with a constant intensity and generate cascades of aftershocks, each shock generating direct aftershocks after durations given by an inhomogeneous Poisson process with a fixed intensity function in i.i.d. fashion. In mathematical terms, the conditional intensity of the point process of the instants of shocks is the sum of the primary shock arrival intensity and of time-shifts of the intensity function to the instants of past shocks. Thus, the intensity function must be non-negative, and this can only be used to model self-excitation effects.

Hawkes and Oakes [19] exploited this additive structure to provide a Poisson cluster point process decomposition of the Hawkes process. It allowed them to prove the existence of a stationary version under a sub-criticality

assumption on the cascade of aftershocks, and has since shown itself to be a powerful tool in the study of Hawkes processes.

Brémaud and Massoulié [6] generalized this to phenomena in which the response to the above sum is modulated by an excitation function. They studied a point process N satisfying an initial condition on $(-\infty, 0]$ and with conditional intensity on $(0, \infty)$ given for a transfer function $h : (0, \infty) \to \mathbb{R}$ and an excitation function $\phi : \mathbb{R} \to \mathbb{R}_+$ by

$$(1.1) \quad t \in (0, \infty) \mapsto \phi\left(\int_{(-\infty, t)} h(t - s)\, N(\mathrm{d}s)\right) \triangleq \phi\left(\sum_{s \in N, s < t} h(t - s)\right).$$

The process in Hawkes [17] corresponds to $h \geq 0$ and $\phi(x) = \lambda + \alpha x$ for $x \geq 0$ for constants $\lambda \geq 0$ and $\alpha \geq 0$, and this case is called a linear Hawkes process and the general case a nonlinear Hawkes process if needed.

If ϕ is non-decreasing then the positive values of h can be interpreted as self-excitation and the negative ones as self-inhibition, and ϕ modulates the response to the superposition of the effects of previous points. This allows to model phenomena featuring both self-excitation and self-inhibition, which appear in many applicative fields such as seismology [18, 27], finance [1, 2, 3, 21, 22], genetics [30], or neurosciences [8, 14, 28]. Already in [6] a neuron network was modeled by an interacting system of Hawkes processes, which provide interesting models such as those in [7, 13, 12, 14, 15].

Nonlinear Hawkes processes are much harder to analyze than linear ones, in particular no equivalent of the Poisson cluster point process decomposition is known. If ϕ is non-decreasing and h is non-negative then the natural monotonicity properties of the process help, but considering h taking negative values is much harder. This is obvious in [6], where the existence and sometimes the uniqueness and attractivity of a stationary version of the nonlinear Hawkes process is proved under various sets of assumptions.

An assumption which appears in many papers of the field and simplifies several proofs in [6] is that the response function ϕ is bounded, another one is that the transfer function h has bounded support and thus the Hawkes process has bounded memory. Both are restrictive for applications.

The mathematical analysis of the models and the development, calibration, and validation of efficient statistical tools is an important issue. For instance Reynaud-Bouret and Roy [29] obtained exponential concentration inequalities for ergodic theorems for the linear Hawkes process with bounded memory, in which necessarily $h \geq 0$ (pure self-excitation), using a coupling similar to what can be found in Berbee [5].

In case of transfer functions h which may take negative values (self-inhibition), similar results were obtained in Costa *et al.* [9] by coauthors

and myself under the bounded memory assumption that h has bounded support. The response function was taken to be $\phi : x \mapsto x^+$ for simplicity of exposition but could be generalized. The bounded memory assumption was crucial in representing the Hawkes process by a Markov process and proving that it has a positive recurrent state. This yielded long time limit theorems for functional statistical estimators of interest by deriving them from corresponding limit theorems for i.i.d. sequences. Notably exponential concentration inequalities were obtained from Bernstein's inequality using highly intricate renewal computations.

1.2. *Scope of the Paper.* A filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ satisfying the usual assumptions and supporting all required random elements is given. All processes will be adapted.

The present paper focuses on linear Hawkes process with unbounded memory, *i.e.*, with transfer functions with unbounded support, for which the techniques in [9] do not apply. Even though our main tool is the Poisson cluster point process decomposition, which generalizes to interacting systems and to marked linear Hawkes processes, we consider the one-dimensional case so as to focus on the main difficulties and ideas. The methods would generalize under adequate assumptions.

We aim in particular to provide long-time limit results applicable to an important class of functional statistical estimators involving a sliding window of fixed width A, and introduce the notion of A-regeneration which is natural in this context. The notation of the explicit dependence on A is often omitted in what follows.

The main contribution of the paper is to exhibit regeneration times for the linear Hawkes process and to study them in a precise fashion, under the minimal assumptions $\int h(t)\, dt < 1$ and $\int t h(t)\, dt < \infty$ and a mild condition on the initial condition ensuring that its influence eventually vanishes. The Laplace transform of the regeneration time is expressed in terms of the transfer function h. When $\int e^{\theta t} h(t)\, dt < \infty$ for some $\theta > 0$, the regeneration time is stochastically dominated by random variables with explicit Laplace transforms depending on a parameter computable in terms of the Laplace transform of h, which yields explicit exponential moment bounds.

The main idea is as follows. We construct the Hawkes process using $\mathcal{F}_0$-measurable random elements in order to take into account the influence on $(0, \infty)$ of the initial condition, and a $(\mathcal{F}_t)_{t \geq 0}$-Poisson cluster point process to add further points on $(0, \infty)$ similarly the decomposition of Hawkes and Oakes [19]. We then exhibit regeneration times by exploiting the $(\mathcal{F}_t)_{t \geq 0}$-strong Markov property of the Poisson cluster point process reflecting its in-

dependence properties. The regeneration times are stopping times for $(\mathcal{F}_t)_{t\geq 0}$ but not in general for the Hawkes process. Intuitively, the construction allows us to peek ahead of times t in a $\mathcal{F}_t$-measurable way and thus to be able to detect such times at which the past will not influence the future; [6, p.1570] writes "Note that an unbounded support for the function h makes the memory always infinite, and regenerative arguments do not come up naturally, although they may exist", and we found one.

Regenerative proprieties open up the toolbox of coupling techniques and of adapting results for i.i.d. sequences by partitioning the process into an initial delay and an independent i.i.d. sequence of cycles. In particular the proofs and results in [9] can be reproduced here without the bounded memory assumption that the transfer function h has bounded support, but here necessarily h is non-negative and covers self-excitation cases only whereas [9] allows for h taking negative values and covers self-inhibition cases.

Good bounds on the regeneration times are essential for applications. As in [9], we interpret the regeneration time as the renewal time at the empty state of a $M/G/\infty$ queue with service time given by the sum of the cluster length and the window width A, and use a formula of Takàcs [34, 35] expressing the Laplace transform of this renewal time in terms of the service time c.d.f. After this, [9] establishes a general result on exponential tails of $M/G/\infty$ renewal times at zero for service times with exponential tails, which they implement for the Hawkes process using the somewhat loose bound on the cluster length obtained in [29, Sect. 1.1] in terms of the exponential moments of h.

Here we seek better and more explicit bounds on the cluster length and renewal time. Surprisingly, studies of super-critical branching random walks abound, see [33, 20, 4, 24], $e.g.$, but there is very little recent work on the subcritical case. Möller and Rasmussen [26] obtained results based on a formula in Hawkes and Oakes [19] which allow us to stochastically dominate the cluster length by an exponential random variable with parameter computable in terms of the exponential moments of h.

This yields that the regeneration time is stochastically dominated by the renewal time at zero of the $M/G/\infty$ queue with service given by the sum of the dominating exponential random variable and of A, which yields explicit computable bounds. Notably, this yields explicit bounds on the Laplace transform of the regeneration time in terms of simple integrals or special functions appearing in like studies for M/M/infinity queues, which yield an explicit negative upper-bound on its abscissa of convergence.

These regenerative results allow to systematically derive long-time asymptotic results on the Hawkes process, in particular applicable to sliding window

statistical estimators. In this context, Costa *et al.* [9, Thm 1.5] thus establish exponential concentration inequalities using highly intricate computations which can be reproduced *verbatim* here. We demonstrate the usefulness of the precise explicit bounds we have obtained for the regeneration time on the simplified version [9, Cor. 1.6] of the concentration inequality and thus provide explicit non-asymptotic exponential bounds.

1.3. *Notation and Organization.* Section 2 builds from the Poisson cluster point process the processes which will be used to state and prove regeneration. Section 3 defines the notion of A-regeneration and then exhibits a sequence of regeneration times. Section 4 interprets the regeneration time in terms of a $M/G/\infty$ queue, uses this to obtain its Laplace transform, and stochastically dominates it to obtain explicit bounds. Section 5 discusses applications to limit theorems for a wide class of functional statistical estimators, and illustrates the explicit bounds on a concentration inequality. Appendix A discusses how to approximate other classes of useful statistical estimators by the previous class. Appendix B lists formulæ on the renewal times and busy periods of $M/G/\infty$ and $M/M/\infty$ queues.

When B is a Borel subset of $\mathbb{R}$, we denote by $\mathcal{N}(B)$ [resp. $\mathcal{N}_b(B)$] the space of boundedly finite [resp. finite] counting measures, in particular if B is bounded then $\mathcal{N}(B) = \mathcal{N}_b(B)$. There may be abuse of notation in which the restriction $\mu|_B \in \mathcal{N}(B)$ of $\mu \in \mathcal{N}(\mathbb{R})$ is identified with $\mu\mathbb{1}_B \in \mathcal{N}(\mathbb{R})$ and reciprocally in which $\mu \in \mathcal{N}(B)$ is identified with its extension to a measure on $\mathbb{R}$ using the null measure on $\mathbb{R} \setminus B$. The shift operator group $(S_t)_{t\in\mathbb{R}}$ satisfies (with $B - t = \{x - t : x \in B\}$, etc.)

$$(1.2) \quad \begin{cases} S_t : \mu \in \mathcal{N}(B) \mapsto S_t\mu \in \mathcal{N}(B-t)\,, \\ \displaystyle\int_{B-t} f(x)\, S_t\mu(\mathrm{d}x) \triangleq \int_B f(x-t)\,\mu(\mathrm{d}x)\,, \quad f \in \mathcal{B}_b(B-t)\,, \\ S_t\mu(C) = \mu(C+t)\,, \qquad\qquad\qquad\quad C \in \mathcal{N}(B-t)\,. \end{cases}$$

Point processes are considered as random variables with values in some specified $\mathcal{N}(B)$, sometimes with abuse of notation on $\mathcal{N}(\mathbb{R})$ with null extension outside B. They may be identified with the random set of their atoms since they always are simple, or with (appropriately started) counting processes in order to use results for classic càdlàg processes with values in a Polish space. Daley and Vere-Jones [10, 11] is a reference book on this topic, in which $\mathcal{N}(B)$ [resp. $\mathcal{N}_b(B)$] are denoted by $\mathcal{N}_B^{\#}$ [resp. $\mathcal{N}_B$].

2. Hawkes Process and Poisson Cluster Point Process. From now on we investigate the linear Hawkes Process satisfying (1.1) for linear (actually affine) excitation functions ϕ and (necessarily) non-negative transfer

functions h. We are specifically interested in the case that h has unbounded support and hence the process has unbounded memory. We often drop the adjective "linear" in the following.

DEFINITION 2.1 (Hawkes process). *Let a constant $\lambda \geq 0$, an integrable function $h : (0, \infty) \to \mathbb{R}_+$, and a $\mathcal{F}_0$-measurable boundedly finite point process N^{in} on $(-\infty, 0]$ be given. The point process N on $\mathbb{R}$ is a* (linear) *Hawkes process on $(0, \infty)$ with* initial condition N^{in}, base intensity λ, *and* transfer function h *if*

$$N|_{(-\infty, 0]} = N^{\mathrm{in}}$$

and the conditional intensity measure of $N|_{(0, \infty)}$ is absolutely continuous w.r.t. the Lebesgue measure with density

$$t \in (0, \infty) \mapsto \lambda + \int_{(-\infty, t)} h(t - u) \, N(\mathrm{d}u) \triangleq \lambda + \sum_{s \in N, s < t} h(t - s) \,.$$

Hawkes and Oakes [19] construct a stationary linear Hawkes process on $\mathbb{R}$ as the Poisson cluster point process of the arrival instants in the following immigration-branching process, in which all random events are independent: ancestors arrive on $\mathbb{R}$ as a Poisson process with intensity λ (immigration), and each individual arrived in the system has offspring after durations given by an inhomogeneous Poisson process of intensity function h (branching). Specifically, an individual arrived at time s has offsprings arriving on (s, ∞) as an inhomogeneous Poisson process of intensity function $t \mapsto h(t - s)$.

The number of offspring of an individual is Poisson of mean $\int h(t) \, \mathrm{d}t < \infty$ and the durations before their births are i.i.d. with density $h / \int h(t) \, \mathrm{d}t$. Along with most references, hereafter we make the sub-criticality assumption

$$(2.1) \qquad\qquad\qquad \int h(t) \, \mathrm{d}t < 1 \,.$$

We adapt this construction for our purposes. Using again the superposition property of Poisson point processes, $N|_{(0, \infty)}$ will be constructed as the sum of the Hawkes process with initial condition the null (empty) process, and the point process of the birth instants in $(0, \infty)$ of offspring of N^{in} and of all subsequent descendants of these offspring born in $(0, \infty)$. The latter is a Hawkes process with initial condition N^{in} and $\lambda = 0$.

Consider first the case of null initial condition $N^{\mathrm{in}} = 0$ under the assumption (2.1). Let $\mathbb{P}^{\mathrm{cl}}$ be the law on $\mathcal{N}_b(\mathbb{R}_+)$ of the cluster generated by an ancestor arriving at time 0 including 0 itself, and $\Gamma(\mathrm{d}t, \mathrm{d}\mu)$ a $(\mathcal{F}_t)_{t \geq 0}$-Poisson process on $(0, \infty) \times \mathcal{N}_b(\mathbb{R}_+)$ with intensity measure $\lambda \, \mathrm{d}t \otimes \mathbb{P}^{\mathrm{cl}}(\mathrm{d}\mu)$.

The process Γ induces a $(\mathcal{F}_t)_{t\geq 0}$-marked Poisson process of intensity λ with instants $(T_n)_{n\geq 1}$ independent of the marks $(\mu_n)_{n\geq 1}$ which are i.i.d. of law $\mathbb{P}^{\text{cl}}$. Using (1.2) and identifying measures on Borel subsets of $\mathbb{R}$ with their extension by the null measure when necessary, let

$$\begin{cases} \psi : \gamma \in \mathcal{N}((0,\infty) \times \mathcal{N}_b(\mathbb{R}_+)) \mapsto \psi(\gamma) = \int S_{-t}\mu\,\gamma(\mathrm{d}t,\mathrm{d}\mu) \in \mathcal{N}(\mathbb{R})\,, \\ \psi_t : \gamma \in \mathcal{N}((0,\infty) \times \mathcal{N}_b(\mathbb{R}_+)) \mapsto \psi_t(\gamma) = \psi(\mathbb{1}_{(0,t]\times\mathcal{N}_b(\mathbb{R}_+)}\gamma) \in \mathcal{N}(\mathbb{R})\,,\ t \geq 0\,. \end{cases}$$

Note that $\psi(\gamma) = \psi_t(\gamma) + \psi(\mathbb{1}_{(t,\infty)\times\mathcal{N}_b(\mathbb{R}_+)}\gamma)$. As in [19], the superposition properties of Poisson point processes imply that

$$\psi(\Gamma) \triangleq \int S_{-t}\mu\,\Gamma(\mathrm{d}t,\mathrm{d}\mu) \triangleq \sum_{n\geq 1} S_{-T_n}\mu_n$$

is a Hawkes process with initial condition the null point process on $(-\infty,0]$. For $t \geq 0$,

$$\psi_t(\Gamma) \triangleq \int S_{-s}\mu\,\mathbb{1}_{(0,t]}(s)\Gamma(\mathrm{d}s,\mathrm{d}\mu) \triangleq \sum_{n\geq 1} \mathbb{1}_{\{T_n\leq t\}}S_{-T_n}\mu_n$$

is $\mathcal{F}_t$-measurable and contains all the arrival instants of ancestors arrived up to time t and their descendance, and thus encodes the influence after time t of these ancestors.

Consider now a general initial condition $N^{\text{in}} = 0$, under the assumption (2.1) and the mild natural assumptions (the equality in the second following from Fubini's theorem)

(2.2a) $$\int th(t)\,\mathrm{d}t < \infty\,,$$

(2.2b) $$\int_{\{s\leq 0 < t\}} h(t-s)\,N^{\text{in}}(\mathrm{d}s)\,\mathrm{d}t = \int_0^\infty h(t)N^{\text{in}}([-t,0])\,\mathrm{d}t < \infty\,.$$

Let D_0 be a $\mathcal{F}_0$-measurable version of a Hawkes process satisfying Definition 2.1 with $\lambda = 0$. It is equal to N^{in} on $(-\infty,0]$, and on $(0,\infty)$ counts the arrival instants of the offspring of N^{in}, a point of which at $s \leq 0$ generates such offspring as an inhomogeneous Poisson process of intensity function $t \mapsto h(t-s)\mathbb{1}_{(0,\infty)}(t)$, and the descendants of these offspring generated according to the usual mechanism, all this using $\mathcal{F}_0$-measurable random elements. Since (2.1) and (2.2) hold, [6, Thm 1(c), Remarks 5,6] yields that D_0 is well-defined and eventually dies out in total variation. The superposition properties yield that the point process

$$N \triangleq D_0 + \psi(\Gamma) \triangleq D_0 + \int S_{-t}\mu\,\Gamma(\mathrm{d}t,\mathrm{d}\mu)$$

is a Hawkes process with initial condition N^{in}. For $t \geq 0$,

$$D_t \triangleq D_0 + \psi_t(\Gamma) \triangleq D_0 + \int S_{-s}\mu \, \mathbb{1}_{(0,t]}(s)\Gamma(\mathrm{d}s,\mathrm{d}\mu)$$

is $\mathcal{F}_t$-measurable and contains N^{in} as well as the arrival instants on $(0,\infty)$ of the offspring of N^{in} and of ancestors arrived up to time t and the descendants of all these. It thus encodes the influence after time t of the initial condition and of these ancestors. Note that D_t is not $\sigma(N|_{(-\infty,t]})$-measurable, but that $N|_{(-\infty,t]} = D_t|_{(-\infty,t]}$. We collect what we have achieved in the following.

THEOREM 2.2. *Assume that* (2.1) *and* (2.2) *hold. The point processes* $\psi(\Gamma)$ *and the* $\mathcal{F}_t$*-measurable* $\psi_t(\Gamma)$ *and* D_t *for* $t \geq 0$ *are well-defined, and*

$$(2.3) \qquad N \triangleq D_0 + \psi(\Gamma) = D_t + \psi(\mathbb{1}_{(t,\infty)\times\mathcal{N}_b(\mathbb{R}_+)}\Gamma)$$

is a Hawkes process with initial condition N^{in}.

PROOF. This follows from the above discussion and notably [6, Thm 1(c)], and the obvious $\psi(\gamma) = \psi_t(\gamma) + \psi(\mathbb{1}_{(t,\infty)\times\mathcal{N}_b(\mathbb{R}_+)}\gamma)$. □

3. Regenerative Results for Linear Hawkes Processes. Our results aim to be applicable to an important class of functional statistical estimators involving a sliding window of length A. For notational convenience the windows are of the form $(t-A,t]$ for $t \geq 0$, observation has started by time $-A$, and at the deterministic time $T > 0$ the estimators are of the form (see (1.2) for the shift operator S_t)

$$(3.1) \qquad \frac{1}{T}\int_0^T f(S_t(N|_{(t-A,t]}))\,\mathrm{d}t \triangleq \frac{1}{T}\int_0^T f((S_tN)|_{(-A,0]}))\,\mathrm{d}t\,,$$
$$f \in \mathcal{B}(\mathcal{N}((-A,0]))\,.$$

To consider observations starting at time 0 it suffices to shift the process or relabel time adequately. This useful class is widely studied as in [29, 9].

These estimators are absolutely continuous processes of T, and can be used to approximate in a well-controlled and precise way many estimators based on integrals with respect to N which are pure jump processes of T. For instance, a useful class is constituted of estimators of the form, for functions w in $\mathcal{B}(\mathbb{R})$ with support included in some $(-A,0]$,

$$(3.2) \qquad \frac{1}{T}\int_{(-A,T]\times(-A,T]} w(y-x)\,N(\mathrm{d}x)N(\mathrm{d}y)$$
$$= \frac{1}{T}\int_{\{-A<y\leq x\leq T\}} w(y-x)\,N(\mathrm{d}x)N(\mathrm{d}y)\,.$$

Appendix A shows how to approximate (3.2) by (3.1) for a well-chosen $f \triangleq f_w$ in a quite controllable way, which allows to apply results true for (3.1) for $f \triangleq f_w$ on (3.2). Regenerative techniques may be applied directly to estimators of the form (3.2), but we chose not to develop this. In a multi-dimensional setting where $N = (N^i : 1 \leq i \leq d)$ this class generalizes under the form $\left(\frac{1}{T} \int_{\{-A < y \leq x \leq T\}} w^{ij}(y-x) \, N^i(\mathrm{d}x) N^j(\mathrm{d}y) : 1 \leq i, j \leq d \right)$ and provides valuable information on the relations between components.

The following notion corresponds to classical regeneration for the càdlàg process $(S_t(N|_{(t-A,t]}), t \geq 0) \triangleq ((S_t N)|_{(-A,0]}), t \geq 0)$, which explains why the past is wide-sense and the future is strict in terms of N.

DEFINITION 3.1 (A-regeneration). *Let $0 \leq A < \infty$. An a.s. finite $(\mathcal{F}_t)_{t\geq 0}$-stopping time ρ is called an A-regeneration time if*

$$\mathcal{F}_\rho \text{ is independent of } S_\rho(N|_{(\rho-A,\infty)}) \triangleq (S_\rho N)|_{(-A,\infty)}$$

and the latter has same law as the restriction to $(-A, \infty)$ of a Hawkes process with null initial condition, i.e., as $\psi(\Gamma)|_{(-A,\infty)}$. Note that $\sigma(N|_{(-\infty,\rho]}) \subset \mathcal{F}_\rho$.

The event that the influence of the initial condition N^{in} and of ancestors arrived in $(0, \rho - A]$ has vanished in $(\rho - A, \infty)$ and that no ancestor has arrived in $(\rho - A, \rho]$ can be conveniently written $D_\rho((\rho - A, \infty)) = 0$.

THEOREM 3.2. *Assume that (2.1) and (2.2) hold. Let $0 \leq A < \infty$. Then an a.s. finite $(\mathcal{F}_t)_{t\geq 0}$-stopping time ρ is an A-regeneration time if and only if $D_\rho((\rho - A, \infty)) = 0$.*

PROOF. If $D_\rho((\rho - A, \infty)) = 0$ then, using (2.3),

$$S_\rho(N|_{(\rho-A,\infty)}) = S_\rho(\psi(\mathbb{1}_{(\rho,\infty)\times\mathcal{N}_b(\mathbb{R}_+)}\Gamma)|_{(\rho-A,\infty)})$$

and the strong Markov property of the $(\mathcal{F}_t)_{t\geq 0}$-Poisson point process Γ yields that ρ is an A-regeneration time. The converse is obvious. $\square$

We now state the main regenerative result of the paper.

THEOREM 3.3. *Assume that (2.1) and (2.2) hold. Let $0 \leq A < \infty$ and*

$\tau_0 \triangleq \tau_0^A \triangleq \inf\{t \geq 0 : D_t((t - A, \infty)) = 0\},$

$\tau_k \triangleq \tau_k^A \triangleq \inf\{t > \tau_{k-1}^A : D_t([t - A, \infty)) \neq 0, \ D_t((t - A, \infty)) = 0\}, \ k \geq 1.$

Then $(\tau_k)_{k\geq 0}$ is a sequence of A-regeneration times. Thus the delay and the cycles

$$N|_{(-A,\tau_0]}\,,\quad S_{\tau_{k-1}}(N|_{(\tau_{k-1}-A,\tau_k]}) \triangleq (S_{\tau_{k-1}}N)|_{(-A,\tau_k-\tau_{k-1}]}\,,\ k\geq 1\,,$$

are all independent and each cycle $(S_{\tau_{k-1}}N)|_{(-A,\tau_k-\tau_{k-1}]}$ has same law as $N|_{(-A,\tau_1]}$ when $N^{\mathrm{in}} = 0$, and thus as $\psi(\Gamma)|_{(-A,\tau]}$ for

$$\tau \triangleq \tau^A \triangleq \inf\{t > 0 : \psi_t(\Gamma)([t-A,\infty)) \neq 0\,,\ \psi_t(\Gamma)((t-A,\infty)) = 0\}\,.$$

Moreover τ and the $\tau_k - \tau_{k-1}$ are integrable, and the τ_k are integrable if τ_0 is integrable.

PROOF. Theorem 4.1 below yields that τ is integrable. If N^{in} is the null point process, then $\tau_0 = 0$ and the result follows from a recursive application of Theorem 3.2. For a general initial condition, $N \triangleq D_0 + \psi(\Gamma)$ in which D_0 is an a.s. finite point process and $\psi(\Gamma)$ has a null initial condition. (see Section 2). Let U denote the last point of D_0 and τ'_k the regeneration times of $\psi(\Gamma)$. Then $U < \infty$ and $\kappa \triangleq \inf\{k \geq 1 : U \leq \tau'_k\} < \infty$ and hence $\tau_0 \leq \tau'_{\kappa+1} < \infty$, a.s. $\square$

Simple conditions on h and N^{in} imply that τ_0^A is integrable. In the important case that N^{in} is stochastically dominated by the stationary Hawkes process it suffices that additionally $\int t^2 h(t)\,\mathrm{d}t < \infty$, see Theorem 5.2 below.

In order for these regenerative properties to be useful in applications, tight estimates on the A-regeneration times are required. This is quite crucial here since in general the τ_k are not stopping-times for N, nor even $\sigma(N)$-measurable since the observation of N cannot directly differentiate between ancestors and descendants. Nevertheless, developing statistical methods for τ is simpler than, *e.g.*, in the case of Nummelin splitting. In the bounded memory case, if h vanishes on (A,∞) then the $(\tau_k)_{k\geq 0}$ are the stopping times for N defined just before [9, Thm 3.6] which can be easily estimated.

4. Regeneration Time and Stochastic Domination. We hereafter assume that the transfer function h satisfies (2.1) and (2.2a). We consider $0 \leq A < \infty$ and explicitly mark the A-dependence for clarity, but it may be dropped when the context is clear. We give a series of results on $\tau \triangleq \tau^A$ culminating with a stochastic domination result providing good bounds in terms of the transfer function h. The evaluation of h is well understood and is a primary goal of any statistical study of the Hawkes process.

The cluster length L is a generic random variable with same law as the duration between the arrival instant of an ancestor and the last birth instant

of his descendants, *i.e.*, as the last point of the cluster under the law $\mathbb{P}^{\mathrm{cl}}$. Its cumulative distribution function (c.d.f.), with support $[0, \infty)$, is given by

$$F : x \in \mathbb{R} \mapsto F(x) = \mathbb{P}(L \leq x).$$

A crucial fact is that $\tau \triangleq \tau^A$ is the renewal time at state 0 of a $M/G/\infty$ queue with arrivals with intensity λ and generic service duration $L + A$ with c.d.f., with support $[A, \infty)$, given by

(4.1) $$F^A : x \in \mathbb{R} \mapsto F^A(x) = F(x - A).$$

As it is well known, see [19, p.502], since any individual in the branching process has a number of offspring with mean $\int h(t)\, \mathrm{d}t$ after durations of mean $\int t h(t)\, \mathrm{d}t / \int h(t)\, \mathrm{d}t$,

$$\int_0^\infty (1 - F^A(u))\, \mathrm{d}u = \mathbb{E}[L] + A \leq \frac{\int h(t)\, \mathrm{d}t}{1 - \int h(t)\, \mathrm{d}t} \frac{\int t h(t)\, \mathrm{d}t}{\int h(t)\, \mathrm{d}t} + A < \infty$$

and much better bounds are available. This yields the following.

THEOREM 4.1 (Takács). *Assume that* (2.1) *and* (2.2a) *hold. Let* $0 \leq A < \infty$. *The Laplace transform of* τ^A *is given by*

(4.2) $$\mathbb{E}\big[\mathrm{e}^{-s\tau^A}\big] = 1 - \frac{1}{\lambda + s} \left(\int_0^\infty \mathrm{e}^{-st - \lambda \int_0^t (1 - F^A(u))\, \mathrm{d}u}\, \mathrm{d}t \right)^{-1},$$

$$s \in \mathbb{C},\ \Re(s) > 0.$$

Moreover $\mathbb{E}[\tau^A] < \infty$, $\mathbb{E}[(\tau^A)^2] < \infty$ *if and only if* $\int t^2 h(t)\, \mathrm{d}t < \infty$, *and*

$$\mathbb{E}[\tau^A] = \frac{1}{\lambda} \mathrm{e}^{\lambda(\mathbb{E}[L] + A)} < \infty,$$

$$\mathbb{E}[(\tau^A)^2] = \frac{2}{\lambda} \mathrm{e}^{2\lambda(\mathbb{E}[L] + A)} \int_0^\infty \left(\mathrm{e}^{-\lambda \int_0^t (1 - F^A(u))\, \mathrm{d}u} - \mathrm{e}^{-\lambda(\mathbb{E}[L] + A)} \right) \mathrm{d}t$$

$$+ \frac{2}{\lambda^2} \mathrm{e}^{\lambda(\mathbb{E}[L] + A)}.$$

PROOF. See Takács [34, (37)–(39)] or [35, pp.210,211]. $\qquad\square$

Appendix B lists some expressions for (4.2) and the integral in this formula. The fact that τ^A is the sum of an exponential $\mathcal{E}(\lambda)$ idle period and an independent busy period β^A implies the product form $\mathbb{E}\big[\mathrm{e}^{-s\tau^A}\big] = \frac{\lambda}{\lambda + s} \mathbb{E}\big[\mathrm{e}^{-s\beta^A}\big]$

in (B.1) and the first equality in (B.2). Theorem 4.1 can be expressed in terms of F only using (4.1) which yields that

$$(4.3) \quad \int_0^t (1 - F^A(u))\, du = \mathbb{1}_{\{t \le A\}} t + \mathbb{1}_{\{t > A\}} \left(A + \int_0^{t-A} (1 - F(u))\, du \right),$$

$$(4.4) \quad \int_0^\infty e^{-st - \lambda \int_0^t (1 - F^A(u))\, du}\, dt$$

$$= \frac{1 - e^{-(\lambda + s)A}}{\lambda + s} + e^{-(\lambda + s)A} \int_0^\infty e^{-st - \lambda \int_0^t (1 - F(u))\, du}\, dt\,.$$

This yields relations between τ^A and τ^0 such as (B.3), β^A and β^0 such as (B.4), and formulæ for β^A and hence τ^A such as (B.5).

There is an apparent analytic singularity in the Laplace transform (4.2) at $s = 0$. It was removed under exponential decay assumptions on $1 - F^A$ in [9, Thm A.1] using integration by parts, analytic continuation and Laplace transform properties to provide a negative upper-bound for the abscissa of convergence. Here we will go further and use explicit computations. For $\theta > 0$ we denote the c.d.f. of the exponential law $\mathcal{E}(\theta)$ by

$$G^\theta : x \in \mathbb{R} \mapsto G^\theta(x) = (1 - e^{-\theta x})^+\,.$$

THEOREM 4.2 (Möller and Rasmussen).   *Assume that $\int h(t)\, dt < 1$. If $\theta > 0$ is such that $\int e^{\theta t} h(t)\, dt \le 1$ then $G^\theta \le F$. Then L is stochastically dominated by the exponential random variable $\mathcal{E}(\theta)$. Notably $\mathbb{E}[L] \le 1/\theta$.*

PROOF.   Follows from Möller and Rasmussen [26, Prop. 3, Lemma 2].   □

If $\int e^{\theta t} h(t)\, dt < \infty$ for some $\theta > 0$ then, using $\int h(t)\, dt < 1$ and monotone convergence,

$$(4.5) \quad \begin{cases} \theta^* \triangleq \sup\left\{ \theta > 0 : \int e^{\theta t} h(t)\, dt \le 1 \right\} > 0\,, \\[2mm] \forall \theta \in (0, \theta^*)\,, \ \int e^{\theta t} h(t)\, dt \le 1\,, \end{cases}$$

and if $1 \le \int e^{\theta t} h(t)\, dt < \infty$ for some $\theta > 0$ then $\int e^{\theta^* t} h(t)\, dt = 1$.

Theorem 4.2 yields much tighter bounds on L than [29, Sect. 1.1], in which $p = \int h(t)\, dt$, W_∞ denotes the cardinal of the cluster, L is denoted by H and bounded by a sum of $W_\infty - 1$ random variables of density h/p (all independent), and it is assumed that $\ell(\theta) \triangleq \log(\int e^{\theta t} h(t)\, dt / p) \le p - \log p - 1$ to conclude that $\mathbb{P}(L > x) \le e^{1-p} e^{-\theta x}$. In this notation, Theorem 4.2 assumes that $\ell(\theta) \le -\log p$ to conclude that $\mathbb{P}(L > x) \le e^{-\theta x}$. Since

$-\log p = 1 - p + (p - \log p - 1) > p - \log p - 1 = (1-p)^2/2 + o_{p\to 1}((1-p)^2)$,
Theorem 4.2 provides much better exponents when p is close to 1, and the
absence of a prefactor $e^{1-p} > 1$ yields stochastic domination.

For $0 \leq A < \infty$ and $\theta > 0$, let $\tau^{\theta,A}$ denote the renewal time at 0 and
$\beta^{\theta,A}$ the busy period of a $M/G/\infty$ queue with arrivals at rate λ and service
time given by the sum of an exponential random variable $\mathcal{E}(\theta)$ and of A and
hence with c.d.f. $G^{\theta,A} : x \mapsto (1 - e^{-\theta(x-A)})^+$. We refer to Appendix B for
various expressions for their Laplace Transforms.

THEOREM 4.3. *Assume that $\int h(t)\, dt < 1$. Let $0 \leq A < \infty$. If $\theta > 0$
is such that $\int e^{\theta t} h(t)\, dt \leq 1$ then τ^A is stochastically dominated by $\tau^{\theta,A}$.
Theorem 4.1 and (4.1)-(4.3)-(4.4) remain true with τ^A, F^A, F, $\mathbb{E}[L]$ replaced
by $\tau^{\theta,A}$, $G^{\theta,A}$, G^θ, $1/\theta$, and the formulæ are made explicit by*

$$(4.6) \qquad \int_0^\infty e^{-st - \lambda \int_0^t (1 - G^\theta(u))\, du}\, dt = \frac{e^{-\frac{\lambda}{\theta}}}{\theta} \int_0^1 x^{\frac{s}{\theta} - 1} e^{\frac{\lambda}{\theta} x}\, dx$$

$$= \frac{1}{\theta} \int_0^1 (1-x)^{\frac{s}{\theta} - 1} e^{-\frac{\lambda}{\theta} x}\, dx\,.$$

*Then $\mathbb{E}\big[e^{-s\tau^A}\big] \leq \mathbb{E}\big[e^{-s\tau^{\theta,A}}\big] = \frac{\lambda}{\lambda+s}\mathbb{E}\big[e^{-s\beta^{\theta,A}}\big] < \infty$ for $0 \leq -s < \min(\lambda,\theta)$
and the Laplace transform $\mathbb{E}\big[e^{-s\tau^A}\big]$ converges in $\{\Re(s) > -\min(\lambda,\theta)\}$.*

PROOF. Theorem 4.2 implies that L is stochastically dominated by the ex-
ponential random variable $\mathcal{E}(\theta)$, and the stochastic domination of τ^A by $\tau^{\theta,A}$
is obtained by coupling the service durations. The rest uses again Takács [34,
(37)–(39)] or [35, pp.210,211],

$$\int_0^\infty e^{-st - \lambda \int_0^t (1 - G^\theta(u))\, du}\, dt = \int_0^\infty e^{-st - \frac{\lambda}{\theta}(1 - e^{-\theta t})}\, dt = \frac{e^{-\frac{\lambda}{\theta}}}{\theta} \int_0^1 x^{\frac{s}{\theta} - 1} e^{\frac{\lambda}{\theta} x}\, dx$$

and the change of variables $x \mapsto 1 - x$, as well as analytic continuation
(Rudin [32, Thm 10.18 p.208]) and Laplace transform properties (Widder [37,
Thm 5b p.58]). $\square$

There are many expressions for

$$\int_0^\infty e^{-st - \lambda \int_0^t (1 - G^\theta(u))\, du}\, dt\,, \quad \int_0^\infty e^{-st - \lambda \int_0^t (1 - G^{\theta,A}(u))\, du}\, dt\,,$$

and the Laplace transforms of $\tau^\theta \triangleq \tau^{\theta,0}$ and $\tau^{\theta,A}$ and of the correspond-
ing busy periods $\beta^\theta \triangleq \beta^{\theta,0}$ and $\beta^{\theta,A}$. These result from (4.1)-(4.3)-(4.4)
with F^A and F replaced by $G^{\theta,A}$ and G^θ and the extensive study of the

14 CARL GRAHAM

$M/M/\infty$ queue, which is the dominating queue in the case $A = 0$, notably by Guillemin and Simonian [16] in relation to Kummer's confluent hypergeometric function. Appendix B lists some material on this topic which would be useful in actual applications.

5. Applications. These regenerative proprieties enable to adapt classic and less classic limit theorems for i.i.d. random variables by decomposing the process into a delay and independent i.i.d. cycles, and to use coupling and other tools provided by regeneration such as found in the reference book Thorisson [36]. For instance we can now provide a proof of Brémaud and Massoulié [6, Thm 1(c)] using a general result for regenerative processes.

THEOREM 5.1 (Brémaud and Massoulié). *Assume that* (2.1) *and* (2.2) *hold. Then there exists a stationary version* N^* *of the Hawkes process on* $(0, \infty)$ *such that*

$$(S_t N)|_{(0,\infty)} \xrightarrow[t\to\infty]{\text{total variation}} N^* \,.$$

PROOF. The Hawkes process is regenerative, the simple fact that τ^A is spread out is left as an exercise (see [9, Sect. 4, Proof of Theorem 1.3 b)] for a solution) and the convergence statement follows from Thorisson [36, Thm 10.3.3 p.351]. □

Good controls on τ_0^A are important in practice. We say that a point process N' is stochastically dominated by a point process N'' if $N'(B)$ is stochastically dominated by $N''(B)$ for all Borel sets B. This is equivalent to the existence of a coupling such that $N' \le N''$.

THEOREM 5.2. *Assume that* (2.1) *and* (2.2a) *hold. Let* $0 \le A < \infty$. *Assume that* N^{in} *is stochastically dominated by a stationary version of the Hawkes process. Then* τ_0^A *is stochastically dominated by* $U[\tau^A]^*$, *where* U *is uniform on* $[0,1]$ *and* $[\tau^A]^*$ *is an independent length-biased version of* τ^A, *and for all nonnegative Borel* f,

$$\mathbb{E}[f(U[\tau^A]^*)] = \frac{1}{\mathbb{E}[\tau^A]}\mathbb{E}\left[\int_0^{\tau^A} f(u)\,du\right], \quad \mathbb{E}[\tau_0^A] \le \mathbb{E}[U[\tau^A]^*] = \frac{\mathbb{E}[(\tau^A)^2]}{2\mathbb{E}[\tau^A]} \,.$$

Thus τ_0^A *is a.s. finite, and* $\mathbb{E}[\tau_0^A] < \infty$ *if and only if* $\int t^2 h(t)\,dt < \infty$ *and then has a finite bound expressed from Theorem 4.2. If moreover there exists* $\theta > 0$ *such that* $\int e^{\theta t} h(t)\,dt \le 1$ *then* τ_0^A *is stochastically dominated by*

$U[\tau^{\theta,A}]^*$, where U is uniform on $[0,1]$ and $[\tau^{\theta,A}]^*$ is an independent length-biased version of $\tau^{\theta,A}$, and for all nonnegative Borel f,

$$\mathbb{E}[f(U[\tau^{\theta,A}]^*)] = \frac{1}{\mathbb{E}[\tau^{\theta,A}]}\mathbb{E}\left[\int_0^{\tau^{\theta,A}} f(u)\,du\right], \quad \mathbb{E}[\tau_0^A] \leq \frac{\mathbb{E}[(\tau^{\theta,A})^2]}{2\mathbb{E}[\tau^{\theta,A}]} < \infty,$$

with an explicit finite bound obtained using Theorem 4.2 with τ^A, $\mathbb{E}[L]$, and F^A replaced by $\tau^{\theta,A}$, $1/\theta$, and $G^{\theta,A}$.

PROOF. Thorisson [36, Thm 10.2.1 p.341] and uniqueness in law of the stationary version of the Hawkes process on $\mathbb{R}$ imply that if N^{in} is a stationary version of the Hawkes process on $(-\infty, 0]$ then τ_0^A has same law as $U[\tau^A]^*$. A monotony argument on the $M/G/\infty$ queue yields that if an initial condition is stochastically dominated by a second one then the corresponding τ_0^A are in the same stochastic order. The result on $\mathbb{E}[\tau_0^A] < \infty$ follows from Theorem 4.1. The statement about stochastic domination by $U[\tau^{\theta,A}]^*$ follows from another monotony argument on the $M/G/\infty$ queue. $\square$

Long-time limit results for functional statistical estimators such as (3.1) are important for applications. The following is classic in this context. A Borel function on $\mathcal{N}((-A,0])$ is said to be locally bounded if it is uniformly bounded on $\{\nu \in \mathcal{N}((-A,0]) : \nu((-A,0]) \leq n\}$ for each $n \geq 1$.

THEOREM 5.3 (Pointwise ergodic theorem). *Assume that* (2.1) *and* (2.2) *hold. Let* $0 \leq A < \infty$. *If* f *is a locally bounded Borel function on* $\mathcal{N}((-A,0])$ *which is nonnegative or* π^A-*integrable (see below), then*

$$\frac{1}{T}\int_0^T f(S_t(N|_{(t-A,t]}))\,dt \xrightarrow[T\to\infty]{\text{a.s.}} \pi^A f \triangleq \frac{1}{\mathbb{E}[\tau^A]}\mathbb{E}\left[\int_0^{\tau^A} f(S_t(\psi(\Gamma)|_{(t-A,t]}))\,dt\right].$$

PROOF. This is a simple consequence of the renewal reward theorem. See [9, Sect. 4, Proof of Theorem 1.3 a)] with the notation $X_t \triangleq S_t(N|_{(t-A,t]})$ for the classic proof. $\square$

The following central limit theorem is not as commonly found, and its proof is considerably more involved.

THEOREM 5.4 (Central limit theorem). *Assume that* (2.1) *and* (2.2) *hold. Let* $0 \leq A < \infty$. *If* f *is a locally bounded Borel function on* $\mathcal{N}((-A,0])$ *which is* π^A-*integrable and such that*

$$(5.1) \qquad \sigma^2(f) \triangleq \frac{1}{\mathbb{E}[\tau^A]}\mathbb{E}\left[\left(\int_0^{\tau^A}\left(f(S_t(N|_{(t-A,t]})) - \pi_A f\right)dt\right)^2\right] < \infty,$$

then

$$\sqrt{T}\left(\frac{1}{T}\int_0^T f(S_t(N|_{(t-A,t]}))\,\mathrm{d}t - \pi^A f\right)\xrightarrow[T\to\infty]{\text{in law}}\mathcal{N}(0,\sigma^2(f))\,.$$

PROOF. See [9, Sect. 4, Proof of Theorem 1.4] with the notation $X_t \triangleq S_t(N|_{(t-A,t]})$. $\qquad\square$

Many other long-time limit results for i.i.d. sequences can be adapted to the functional statistic estimators such as (3.1), notably to establish and quantify asymptotic and non-asymptotic convergence rates and confidence intervals for the pointwise ergodic Theorem 5.3.

The main goal and achievement of Costa *et al.* [9, Thm 1.5, Cor. 1.6] is to provide non-asymptotic concentration inequalities for Theorem 5.3 in a set-up allowing for self-inhibition in which h can take negative values. Regeneration is proved in a way requiring that h has support bounded by A in order to consider an auxiliary Markov process with values in $\mathcal{N}((-A,0])$. The proof of the main result [9, Thm 1.5] uses regeneration techniques and Bernstein's inequality [25, Cor. 2.10 p.25, (2.17)–(2.18) p.24]. It was quite intricate and required to finely control a number of terms, and applied Bernstein's inequality to diverse deviations above and below the mean in order to get good bounds. Its corollary [9, Cor. 1.6] provides a less precise but more tractable statement.

The proofs in [9, Sect. 4] can be reproduced here *verbatim*, with the notation $X_t \triangleq S_t(N|_{(t-A,t]})$, for linear Hawkes processes with nonnegative transfer functions having possibly an unbounded support, and hence in a context with unbounded memory. The elements allowing to state [9, Thm 1.5] are quite long, and we refer the interested reader directly to it. We satisfy ourselves with a transcription of [9, Cor. 1.6] showing how the tight bounds we have obtained come into play. In order to better understand it, see Theorem 4.3 and its reference to Theorem 4.1 and the discussion after Theorem 4.2 and notably (4.5).

THEOREM 5.5.   *For simplicity let $N^{\mathrm{in}} = 0$. Assume that* (2.1) *and* (2.2) *hold, and that*

$$\theta^* \triangleq \sup\left\{\theta > 0 : \int \mathrm{e}^{\theta t}h(t)\,\mathrm{d}t \le 1\right\} > 0\,.$$

Then $\mathbb{E}[\tau^A] = \frac{1}{\lambda}\mathrm{e}^{\lambda(\mathbb{E}[L]+A)} < \infty$. Choose $0 < \alpha < \min(\lambda,\theta^*)$, which is such that $\mathbb{E}\big[\mathrm{e}^{\alpha\tau^A}\big] < \infty$, and $-\infty < a < b < \infty$. Let

$$v \triangleq \frac{2(b-a)^2}{\alpha^2}\left\lfloor\frac{T}{\mathbb{E}[\tau^A]}\right\rfloor\mathbb{E}\big[\mathrm{e}^{\alpha\tau^A}\big]\mathrm{e}^{\alpha\mathbb{E}[\tau^A]}\,,\quad c \triangleq \frac{|b-a|}{\alpha}\,.$$

If f is a Borel function on $\mathcal{N}((-A,0])$ with values in $[a,b]$ then, for $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{1}{T}\int_0^T f(S_t(N|_{(t-A,t]}))\,\mathrm{d}t - \pi_A f\right| \geq \varepsilon\right)$$

$$\leq 4\exp\left(-\frac{\left(T\varepsilon - |b-a|\mathbb{E}[\tau^A]\right)^2}{4\left(2v + c(T\varepsilon - |b-a|\mathbb{E}[\tau^A])\right)}\right)$$

or equivalently, for any η in $(0,1)$ there exists $\varepsilon_\eta > 0$ such that

$$\varepsilon_\eta \triangleq \frac{1}{T}\left(|b-a|\mathbb{E}[\tau^A] - 2c\log(\eta/4) + \sqrt{4c^2\log^2(\eta/4) - 8v\log(\eta/4)}\right),$$

$$\mathbb{P}\left(\left|\frac{1}{T}\int_0^T f(S_t(N|_{(t-A,t]}))\,\mathrm{d}t - \pi_A f\right| \geq \varepsilon_\eta\right) \leq \eta.$$

Moreover $\mathbb{E}\left[e^{\alpha\tau^A}\right] \leq \mathbb{E}\left[e^{\alpha\tau^{\theta,A}}\right]$ for $\alpha < \theta < \min(\lambda, \theta^*)$ and the r.h.s. has explicit forms given in Theorem 4.3 and Appendix B.

APPENDIX A: APPROXIMATION OF JUMP-TYPE ESTIMATORS

Estimators of the form (3.1) are absolutely continuous processes of T. We now show how these estimators allow to well approximate an estimator of the form (3.2), which is a pure-jump process of T. Recall that by hypothesis $w = w\mathbb{1}_{[-A',0]}$ for some $A' < A$ and let

$$f_w : \mu \in \mathcal{N}((-A,0]) \mapsto \int_{\{-A < y \leq x \leq 0\}} \frac{w(y-x)}{y-x+A}\,\mu(\mathrm{d}x)\mu(\mathrm{d}y)$$

which is well-defined as a finite sum. Using (1.2) and $x - t - (y - t) = x - y$ and the Fubini theorem for bounded functions yields

$$\int_0^T f_w((S_tN)|_{(-A,0]}))\,\mathrm{d}t$$

$$= \int_0^T \int_{\{-A < y \leq x \leq 0\}} \frac{w(y-x)}{y-x+A}\,(S_tN)(\mathrm{d}x)(S_tN)(\mathrm{d}y)\,\mathrm{d}t$$

$$= \int_0^T \int_{\{t-A < y \leq x \leq t\}} \frac{w(y-x)}{y-x+A}\,N(\mathrm{d}x)N(\mathrm{d}y)\,\mathrm{d}t$$

$$= \int_{\{-A < y \leq x \leq T\}} \int_{\{x \leq t < (y+A)\wedge T\}} \frac{w(y-x)}{y-x+A}\,\mathrm{d}t\,N(\mathrm{d}x)N(\mathrm{d}y)$$

$$= \int_{\{-A < y \leq x \leq T\}} ((y+A)\wedge T - x)\frac{w(y-x)}{y-x+A}\,N(\mathrm{d}x)N(\mathrm{d}y)$$

and thus

$$\int_0^T f_w((S_t N)|_{(-A,0]})\, \mathrm{d}t$$

$$= \int_{\{-A < y \le x \le T,\, y \le T-A\}} w(y-x)\, N(\mathrm{d}x)N(\mathrm{d}y)$$

$$+ \int_{\{T-A < y \le x \le T\}} \frac{T-x}{y-x+A} w(y-x)\, N(\mathrm{d}x)N(\mathrm{d}y)\,.$$

Hence

$$\text{(A.1)} \quad \frac{1}{T}\int_{\{-A < x \le y \le T\}} w(x-y)\, N(\mathrm{d}x)N(\mathrm{d}y) - \frac{1}{T}\int_0^T f_w((S_t N)|_{(-A,0]})\, \mathrm{d}t$$

$$= \frac{1}{T}\int_{\{T-A < y \le x \le T\}} \frac{y-T+A}{y-x+A} w(y-x)\, N(\mathrm{d}x)N(\mathrm{d}y)$$

which can be easily controlled, for instance if $\|w\|_\infty < \infty$ then

$$\left| \frac{1}{T}\int_{\{T-A < y \le x \le T\}} \frac{y-T+A}{y-x+A} w(y-x)\, N(\mathrm{d}x)N(\mathrm{d}y) \right|$$

$$\le \frac{\|w\|_\infty}{2T}[N((T-A,T])^2 + N((T-A,T])]$$

and if $w \ge 0$ then

$$\frac{1}{T}\int_0^T f_w((S_t N)|_{(-A,0]})\, \mathrm{d}t \le \frac{1}{T}\int_{\{-A < x \le y \le T\}} w(x-y)\, N(\mathrm{d}x)N(\mathrm{d}y)$$

$$\le \frac{1}{T}\int_0^{T+A} f_w((S_t N)|_{(-A,0]})\, \mathrm{d}t$$

which can be used for signed w using $w = w^+ - w^-$ and linearity.

There are still practical problems to solve. In order to control f_w in terms of w we use the fact that $A' < A$ and

$$\frac{|w(y-x)|}{y-x+A} = \frac{|w(y-x)|\mathbb{1}_{[-A',0]}(y-x)}{y-x+A} \le \frac{|w(y-x)|\mathbb{1}_{[-A',0]}(y-x)}{A-A'}\,.$$

Since A may be chosen as large as desired this yields nice bounds, for instance if $A \ge A' + 1$ then

$$\frac{|w(y-x)|}{y-x+A} \le |w(y-x)|\,.$$

The second point is that f_w is not an a.s. bounded function of μ, but one can control the probability of having an abnormally large number of jumps

of N on any $(t - A, t]$ for $0 \leq t \leq T$ and use this to obtain in terms of $\|w^+\|_\infty < \infty$ and $\|w^-\|_\infty < \infty$ and well-chosen $-\infty < a < b < \infty$ that

$$\frac{1}{T} \int_0^T f_w((S_t N)|_{(-A,0]}) \, \mathrm{d}t = \frac{1}{T} \int_0^T (a \vee f_w \wedge b)((S_t N)|_{(-A,0]}) \, \mathrm{d}t$$

with given probability arbitrarily close to 1 and then use Theorem 5.5 or like results.

APPENDIX B: RENEWAL TIMES AND BUSY PERIODS

In the sequel $0 \leq A < \infty$ and $s \in \mathbb{C}$ and $\Re(s) > 0$ except if stated otherwise. The product form

$$(\text{B.1}) \quad \mathbb{E}\big[\mathrm{e}^{-s\tau^A}\big] = \frac{\lambda}{\lambda + s} \mathbb{E}\big[\mathrm{e}^{-s\beta^A}\big]$$
$$= \frac{\lambda}{\lambda + s}\left(\frac{\lambda + s}{\lambda} - \frac{1}{\lambda}\left(\int_0^\infty \mathrm{e}^{-st - \lambda \int_0^t (1 - F^A(u))\,\mathrm{d}u}\,\mathrm{d}t\right)^{-1}\right)$$

yields that the Laplace transform of the busy period can be written

$$(\text{B.2}) \quad \mathbb{E}\big[\mathrm{e}^{-s\beta^A}\big] = \frac{\lambda + s}{\lambda} - \frac{1}{\lambda}\left(\int_0^\infty \mathrm{e}^{-st - \lambda \int_0^t (1 - F^A(u))\,\mathrm{d}u}\,\mathrm{d}t\right)^{-1}$$
$$= \frac{\int_0^\infty F^A(t)\mathrm{e}^{-st - \lambda \int_0^t (1 - F^A(u))\,\mathrm{d}u}\,\mathrm{d}t}{\int_0^\infty \mathrm{e}^{-st - \lambda \int_0^t (1 - F^A(u))\,\mathrm{d}u}\,\mathrm{d}t}$$

using integration by parts to obtain that (since $\Re(s) > 0$)

$$\int_0^\infty s\mathrm{e}^{-st - \lambda \int_0^t (1 - F^A(u))\,\mathrm{d}u}\,\mathrm{d}t = 1 - \lambda \int_0^\infty (1 - F^A(t))\mathrm{e}^{-st - \lambda \int_0^t (1 - F^A(u))\,\mathrm{d}u}\,\mathrm{d}t.$$

Moreover (4.1)-(4.3)-(4.4) allow to concentrate on the case $A = 0$, and provide relations such as

$$(\text{B.3}) \quad \mathbb{E}\big[\mathrm{e}^{-s\tau^A}\big] = 1 - \mathrm{e}^{(\lambda+s)A}\left(\mathrm{e}^{(\lambda+s)A} - 1 + \big(1 - \mathbb{E}\big[\mathrm{e}^{-s\tau^0}\big]\big)^{-1}\right)^{-1},$$

$$(\text{B.4}) \quad \mathbb{E}\big[\mathrm{e}^{-s\beta^A}\big]$$
$$= \frac{\lambda + s}{\lambda} - \frac{1}{\lambda}\left(\frac{1 - \mathrm{e}^{-(\lambda+s)A}}{\lambda + s} + \mathrm{e}^{-(\lambda+s)A}\left(\lambda + s - \lambda\mathbb{E}\big[\mathrm{e}^{-s\beta^0}\big]\right)^{-1}\right)^{-1},$$

and formulæ such as

$$(\text{B.5}) \quad \mathbb{E}\big[\mathrm{e}^{-s\beta^A}\big] = \frac{(\lambda + s)\int_0^\infty F(t)\mathrm{e}^{-st - \lambda \int_0^t (1 - F(u))\,\mathrm{d}u}\,\mathrm{d}t}{\mathrm{e}^{(\lambda+s)A} - 1 + (\lambda + s)\int_0^\infty \mathrm{e}^{-st - \lambda \int_0^t (1 - F(u))\,\mathrm{d}u}\,\mathrm{d}t}$$
$$= 1 - \frac{\mathrm{e}^{(\lambda+s)A} - 1}{\mathrm{e}^{(\lambda+s)A} - 1 + (\lambda + s)\int_0^\infty \mathrm{e}^{-st - \lambda \int_0^t (1 - F(u))\,\mathrm{d}u}\,\mathrm{d}t}.$$

The renewal time $\tau^{\theta,A}$ and busy period $\beta^{\theta,A}$ satisfy the same in which F^A and F are replaced respectively by $G^{\theta,A}$ and G^θ, made quite explicit by

$$\int_0^\infty \mathrm{e}^{-st-\lambda \int_0^t (1-G^{\theta,A}(u))\,\mathrm{d}u}\,\mathrm{d}t$$
$$= \frac{1 - \mathrm{e}^{-(\lambda+s)A}}{\lambda + s} + \mathrm{e}^{-(\lambda+s)A} \int_0^\infty \mathrm{e}^{-st-\lambda \int_0^t (1-G^\theta(u))\,\mathrm{d}u}\,\mathrm{d}t$$

and (4.6). The resulting functions of s are analytic together with their inverses in the half-space $\{s \in \mathbb{C} : \Re(s) > -\theta\}$. Similar computations on the last expression in (B.2), with F^A replaced by G^θ, yield the classic form

$$\mathbb{E}\big[\mathrm{e}^{-s\beta^{\theta,0}}\big] = \frac{\int_0^1 x^{\frac{s}{\theta}-1}(1-x)\mathrm{e}^{\frac{\lambda}{\theta}x}\,\mathrm{d}x}{\int_0^1 x^{\frac{s}{\theta}-1}\mathrm{e}^{\frac{\lambda}{\theta}x}\,\mathrm{d}x} = \frac{\int_0^1 (1-x)^{\frac{s}{\theta}-1}x\mathrm{e}^{-\frac{\lambda}{\theta}x}\,\mathrm{d}x}{\int_0^1 (1-x)^{\frac{s}{\theta}-1}\mathrm{e}^{-\frac{\lambda}{\theta}x}\,\mathrm{d}x}$$

under which this Laplace transform can be obtained using martingale techniques, see Robert [31, Sect. 6.3].

Guillemin and Simonian [16, Sect. 4] have thoroughly investigated this form in relation with Kummer's confluent hypergeometric function. See Lebedev [23, Sect. 9.11] for the integral representation of this function that allows to express (4.6) and $\mathbb{E}\big[\mathrm{e}^{-s\beta^{\theta,0}}\big]$ in terms of it, and [23, Sect. 9.9–9.10] for some of its its properties.

AKNOWLEDGMENTS

REFERENCES

[1] BACRY, E., DELATTRE, S., HOFFMANN, M. and MUZY, J. F. (2013). Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Process. Appl.* **123** 2475–2499. MR3054533

[2] BACRY, E., DELATTRE, S., HOFFMANN, M. and MUZY, J. F. (2013). Modelling microstructure noise with mutually exciting point processes. *Quant. Finance* **13** 65–77. MR3005350

[3] BACRY, E. and MUZY, J. F. (2016). First- and second-order statistics characterization of Hawkes processes and non-parametric estimation. *IEEE Trans. Inform. Theory* **62** 2184–2202. MR3480107

[4] BARRAL, J., HU, Y. and MADAULE, T. (2018). The minimum of a branching random walk outside the boundary case. *Bernoulli* **24** 801–841. MR3706777

[5] BERBEE, H. C. P. (1979). *Random walks with stationary increments and renewal theory. Mathematical Centre Tracts* **112**. Mathematisch Centrum, Amsterdam. MR547109

[6] Brémaud, P. and Massoulié, L. (1996). Stability of nonlinear Hawkes processes. *Ann. Probab.* **24** 1563–1588. MR1411506

[7] Chevallier, J. (2017). Mean-field limit of generalized Hawkes processes. *Stochastic Process. Appl.* **127** 3870–3912. MR3718099

[8] Chevallier, J., Cáceres, M. J., Doumic, M. and Reynaud-Bouret, P. (2015). Microscopic approach of a time elapsed neural model. *Math. Models Methods Appl. Sci.* **25** 2669–2719. MR3411353

[9] Costa, M., Graham, C., Marsalle, L. and Tran, V. C. (2018). Renewal in Hawkes Processes with Self-Excitation and Inhibition. Preprint, arXiv:1801.04645.

[10] Daley, D. J. and Vere-Jones, D. (2003). *An introduction to the theory of point processes. Vol. I*, second ed. *Probability and its Applications (New York)*. Springer-Verlag, New York Elementary theory and methods. MR1950431

[11] Daley, D. J. and Vere-Jones, D. (2008). *An introduction to the theory of point processes. Vol. II*, second ed. *Probability and its Applications (New York)*. Springer, New York General theory and structure. MR2371524

[12] Delattre, S. and Fournier, N. (2016). Statistical inference versus mean field limit for Hawkes processes. *Electron. J. Stat.* **10** 1223–1295. MR3499526

[13] Delattre, S., Fournier, N. and Hoffmann, M. (2016). Hawkes processes on large networks. *Ann. Appl. Probab.* **26** 216–261. MR3449317

[14] Ditlevsen, S. and Löcherbach, E. (2017). Multi-class oscillating systems of interacting neurons. *Stochastic Process. Appl.* **127** 1840–1869. MR3646433

[15] Duarte, A., Löcherbach, E. and Ost, G. (2016). Stability, convergence to equilibrium and simulation of non-linear Hawkes processes with memory kernels given by the sum of Erlang kernels. Preprint, arXiv:1610.03300.

[16] Guillemin, F. and Simonian, A. (1995). Transient characteristics of an $M/M/\infty$ system. *Adv. in Appl. Probab.* **27** 862–888. MR1341889

[17] Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58** 83–90. MR0278410

[18] Hawkes, A. G. and Adamopoulos, L. (1973). Cluster Models for Earthquakes: Regional Comparisons. *Bull. Int. Statist. Inst.* **45** 454–461.

[19] Hawkes, A. G. and Oakes, D. (1974). A cluster process representation of a self-exciting process. *J. Appl. Probability* **11** 493–503. MR0378093

[20] Hu, Y. (2016). How big is the minimum of a branching random walk? *Ann. Inst. Henri Poincaré Probab. Stat.* **52** 233–260. MR3449302

[21] Jaisson, T. and Rosenbaum, M. (2015). Limit theorems for nearly unstable Hawkes processes. *Ann. Appl. Probab.* **25** 600–631. MR3313750

[22] Jaisson, T. and Rosenbaum, M. (2016). Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes. *Ann. Appl. Probab.* **26** 2860–2882. MR3563196

[23] Lebedev, N. N. (1965). *Special functions and their applications. Revised English edition. Translated and edited by Richard A. Silverman.* Prentice-Hall, Inc., Englewood Cliffs, N.J. MR0174795

[24] Mallein, B. (2018). Genealogy of the extremal process of the branching random walk. *ALEA Lat. Am. J. Probab. Math. Stat.* **15** 1065–1087. MR3852245

[25] Massart, P. (2007). *Concentration inequalities and model selection. Lecture Notes in Mathematics* **1896**. Springer, Berlin Lectures from the 33rd Summer School on

Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard. MR2319879

[26] Möller, J. and Rasmussen, J. G. (2005). Perfect simulation of Hawkes processes. *Adv. in Appl. Probab.* **37** 629–646. MR2156552

[27] Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of American Statistical Association* **83** 9–27.

[28] Reynaud-Bouret, P., Rivoirard, V. and Tuleau-Malot, C. (2013). Inference of functional connectivity in Neurosciences via Hawkes processes. In *1st IEEE Global Conference on Signal and Information Processing*.

[29] Reynaud-Bouret, P. and Roy, E. (2006). Some non asymptotic tail estimates for Hawkes processes. *Bull. Belg. Math. Soc. Simon Stevin* **13** 883–896. MR2293215

[30] Reynaud-Bouret, P. and Schbath, S. (2010). Adaptive estimation for Hawkes processes; application to genome analysis. *Ann. Statist.* **38** 2781–2822. MR2722456

[31] Robert, P. (2003). *Stochastic networks and queues*, French ed. *Applications of Mathematics (New York)* **52**. Springer-Verlag, Berlin Stochastic Modelling and Applied Probability. MR1996883

[32] Rudin, W. (1987). *Real and complex analysis*, Third ed. McGraw-Hill Book Co., New York. MR924157 (88k:00002)

[33] Shi, Z. (2015). *Branching random walks. Lecture Notes in Mathematics* **2151**. Springer, Cham 42nd Saint Flour Probability Summer School 2012. MR3444654

[34] Takács, L. (1956). On a probability problem arising in the theory of counters. *Proc. Cambridge Philos. Soc.* **52** 488–498. MR0081585

[35] Takács, L. (1962). *Introduction to the theory of queues. University Texts in the Mathematical Sciences.* Oxford University Press, New York. MR0133880

[36] Thorisson, H. (2000). *Coupling, stationarity, and regeneration. Probability and its Applications (New York).* Springer-Verlag, New York. MR1741181

[37] Widder, D. V. (1941). *The Laplace Transform. Princeton Mathematical Series, v. 6.* Princeton University Press, Princeton, N. J. MR0005923

CMAP, École polytechnique
91128 Palaiseau, France
E-mail: carl.graham@polytechnique.edu