ON STABILITY OF HAWKES PROCESS.

DMYTRO KARABASH

Abstract. For Hawkes process, long-memory point process P with intensity $\lambda\big(g_0(t)+\sum_{\tau<t,\tau\in P} h(t-\tau)\big)$ at time t some existence and stability properties are observed. The main result is that under suitable conditions on parameters we show existence of unique invariant distribution of the process; the main difference with previous results is that Lipschitz condition of λ is not required. These methods also provide estimates on the difference at later times of the distributions of the process starting from two different initial conditions. Multi-type generalization is provided.

Date: first draft May 12, 2011; last revision October 28, 2011.
Partially supported National Science Foundation grants: DMS-0904701 and DARPA.

1. Introduction

1.1. **Hawkes Process.** This paper studies Hawkes Processes which are particular types of time-homogeneous self-exciting locally-finite long-memory point processes. Its realization is a discrete random subset of $\mathbb{R}$. The process is characterized by function $\sigma(t,\omega)$ that is the intensity of the point process at time t conditioned on the past history until time t. In other words the probability of k occurrences of the point process in an infinitesimal time interval $[t, t+\delta t]$ given the past history is $\frac{(\sigma(t,\omega)\delta t)^k}{k!} + o(\delta t^k)$. If we denote the random set of the point process by $D(\omega)$, then this statement can be written as

$$P\Big[\# \left|D(\omega) \cap [t, t+\delta]\right| = k | \mathcal{F}_t\Big] = \frac{(\sigma(t,\omega)\delta t)^k}{k!} + o(\delta t^k) \tag{1}$$

where $\mathcal{F}_t$ is the $\sigma-$field generated by $\{D(\omega) \cap I\}$ where I varies over intervals $I \subset (-\infty, t]$. Moreover

$$\sigma(t,\omega) = \sum_{\tau \in D_t} h(t-\tau)) \tag{2}$$

where $D_t := D \cap (-\infty, t)$ is $\mathcal{F}_t$-measurable. Here $\lambda(z)$ is a non-decreasing function of z on $[0,\infty)$ and $h \geq 0$ is a cadlag function of t on $[0,\infty)$ which is integrable and normalized via $\int_0^\infty h(s)ds = 1$. There is a built in invariance under time translation due to form of (2). Given the past history up to time t there is the *outlook function* from $\mathbb{R}_+$ to $\mathbb{R}_+$

$$g_t(s) = \sum_{\tau \in D_t} h(t+s-\tau) \tag{3}$$

which evolves continuously in time and encodes all information needed for future evolution of Hawkes process (but not necessarily all past information; for example in case $h(s) = e^{-s}$). The next occurrence time τ of the point process has the conditional distribution

$$\mathbb{P}[\tau \geq t | \mathcal{F}_0] = \exp\left[-\int_0^t \lambda(g_t(s))ds\right] \tag{4}$$

at which points $g_s(\cdot)$ jumps up by $h(\cdot)$. In section 2 semi-group for g process will be defined and more details given.

1.2. **Historical Perspective and References.** Point processes were first studied in [4] by Erlang in relevance to queueing theory and Hawkes process were first introduced in [6] to study self-exciting point processes. For further references please consult [2, 9]. Most of theoretical results on Hawkes process assume Lipschitz λ and this paper can be viewed as a stronger version of [1] and generalizes results of Bremaud and Massoulié. This paper is one of the first papers that considers non-Lipschitz λ; for a large deviations with non-Lipschitz λ please see [13]; for other large deviation results for Hawkes with linear λ can be found in [14, 15].

Hawkes processes are used to model many phenomena with examples ranging from queues and population to mutations and virus spread to defaults and minimal jumps in the markets to neuroactivity and social-networks; my personal favorite that makes this model interesting is creative thinking i.e. modeling of new ideas. Some of this modeling requires use of multi-dimensional Hawkes process. On the other hand this is written for one-dimensional Hawkes process. Yet this is only for purposes of clarity and results can be generalized to multi-dimensional self-exciting Hawkes process.

1.3. **Example.** Simplest example comes from population control theory. Consider asexual population in certain region as a subset of time where each point in time represents either immigrant into region or birth inside the region; suppose there is no emigration and that immigrants are same as newborns. Further assume that rate of emigration is α and that number of children of each member of population is β with each child birth-date distribution being $h(t+X)dt$ where X is the birth-date of the parent. This corresponds to Hawkes process with $\lambda(z) = \alpha + \beta \cdot z$ and h as given. This is an example of Hawkes process where λ is linear and this also provides relation to Galton-Watson trees which happens to be an easy way of seeing many properties of Hawkes process.

2. Rigorous description of the process and statement of the main results

Let us define the process and briefly explain what the main results are. *Hawkes process is a random collection D of points in $\mathbb{R}_+ := [0,\infty)$ characterized by triplet of parameters (λ, h, g_0) where $\lambda, h, g_0 : [0,\infty) \to [0,\infty)$ with conditional Poisson intensity at time t given by*

$$\lambda\left(g_0(t) + \sum_{0\le s<t, s\in D} h(t-s)\right) \tag{5}$$

where the sum is over all previous points in $D_t := D \cap (-\infty, t)$. The function g_0 describes the initial condition and the functions λ and h describe the evolution of the process. We will denote the distribution of D by $\pi_g := \pi_g^{\lambda,h}$

It is convenient to set up the state space $\mathcal{G}$ of functions $g_0(\cdot)$ on $[0,\infty)$. Start from g_0 and consider random evolution determined by semi-group T_t, i.e. $g_{t+s} = T_t g_s$. Semi-group has two components: deterministic flow $(T_t g)(x) = g(t+x)$ for t up to stopping time τ at which point $(T_\tau g)(x) = g(\tau + x) + h(x)$ with the distribution of $\tau = \tau(g)$ given by

$$\mathbb{P}[\tau \ge t] = \exp\left[-\int_0^t g(t')dt'\right]. \tag{6}$$

After τ semi-group continues as before and new clock starts. Each time stopping time occurs there is a point in D. For any initial condition $g_0(\cdot)$, we have a random Markovian evolution $g_t(\cdot)$. This induces a probability measure P_{g_0} on the space of $\mathcal{G}$-processes. The quantity

$$z_t := g_t(0) = g(t) + \sum_{\tau\in D_t} h(\tau - t) \tag{7}$$

will be stationary and $\lambda(z_t)$ will be the intensity of the point process. The distribution of outlook-function at time s starting from initial condition g will be denotes $\mu_{g,s}$.

Remark. *In the introduction we assumed that $\int h(t)dt = 1$; this is without loss of generality. Indeed one can always achieve this if $\|h\|_1 = \int h(t)dt < \infty$ by observing that triples $(\lambda(z), h(t), g(t))$ and $(\lambda(z\|h\|_1), \frac{h(t)}{\|h\|_1}, \frac{g(t)}{\|h\|_1})$ produce the same Hawkes process with same distribution of D; on the other hand $\|h\|_1 = \infty$ would imply that*

$g_s(0)$ will be going to infinity a.s. and hence if we do not want process to blow up then λ must be bounded which is not natural in our model.

In this paper we observe several facts well known for attractive systems:

1. If $\lambda(z) \le A + Bz$, then the process is well defined for all times. If in addition $B < 1$, there is at least one stationary version.

2. If we start with $g_0 \equiv 0$, then the distribution of $g_t(\cdot)$ has a limit μ as $t \to \infty$ which is stationary, minimum and ergodic.

Then we present main result of this paper that strictly generalizes Bremaud-Massoulié theorem:

3. Under certain conditions on λ and h, for a wide class of g_0, the distribution of $g_t(\cdot)$ under P_{g_0} converges to the same limit as μ.

These results generalize to general multi-type self-exciting Hawkes process. This is described in section 6

3. TOOLS

The two main tools are coupling and addition of parent-child structure; these two tools allow one to get the first intuitions of the process rather quick since then process cannot be too different from the example we have given in subsection 1.3.

3.1. **Coupling.** The coupling can be explained in the following manner. If $g_1(t)$ and $g_2(t)$ are two initial conditions such that $g_1(s) \le g_2(s)$ for all s, then the two probability measures P_{g_1} and P_{g_2} can be coupled to provide a joint process P_{g_1,g_2} that satisfies

$$P_{g_1,g_2}[g_{1,t}(s) \le g_{2,t}(s)\ \forall s \ge 0] = 1 \tag{8}$$

and the set of times when $g_{1,t}(\cdot)$ jumps is almost surely a subset of jump times of $g_{2,t}(\cdot)$. The existence of the coupling can be proved by explicit construction. If $a(t)$, $b(t)$ are two intensities and if $a(t) \le b(t)$ for all t, one can construct a coupled Poisson process, such that the two processes jump together with rate $a(t)$ and the second one jumps by itself at rate $b(t) - a(t)$. In our case this is true till time τ of first jump. After the jump the relation $g_{1,\tau}(s) \le g_{2,\tau}(s)$ continues to hold. Inductively it holds through all the jumps and gives coupling.

Hence we have stochastic domination, that is if we start with two initial conditions one smaller than the other there is coupling for which this domination persists through time.

Analogously stochastic domination is possible for pair of processes with parameters (λ, h, g) and $(\widetilde{\lambda}, \widetilde{h}, \widetilde{g})$ whenever λ is non-decreasing, $\lambda \le \widetilde{\lambda}, h \le \widetilde{h}, g \le \widetilde{g}$; note here $\widetilde{h}$ is not normalized.

3.2. **Parent-Child Structure and Galton-Watson interpretation.** Adding the following Parent-Child Structure one would obtain a random forest structure embedded in time which is useful due to connection to Galton-Watson trees. Start with Hawkes process with parameters (λ, h, g_0) and let $D = \{\tau_1, \tau_2, \tau_3, ...\}$, where τ is an increasing sequence, i.e. $i < j \implies \tau_i < \tau_j$. Let us also denote $\lambda_0(z) = \lambda(z) - \lambda(0)$. Now to each τ_i we associate a parent element $p(\tau_i)$ from $D \cup \{-\infty\}$, where $-\infty$ represents having no parent and being a root node. Let $p(\tau_i)$ be identically distributed according to the following law:

$$(9)$$

$$P[p(\tau_i) = -\infty] = \frac{\lambda(0)}{\lambda(z_{\tau_i})} + \frac{\lambda_0(z_{\tau_i})}{\lambda(z_{\tau_i})} \frac{g(\tau_i)}{z_{\tau_i}} \tag{10}$$

$$(11)$$

$$P[p(\tau_i) = \tau_j] = 1_{j<i} \frac{\lambda_0(z_{\tau_i})}{\lambda(z_{\tau_i})} \frac{h(\tau_i - \tau_j)}{z_{\tau_i}}$$

Then for case $\lambda(z) = \alpha + \beta z$ roots have rate $\alpha + g(t)$ infinitesimal rate and each tree has a distribution of $Poisson(\beta)$-Galton-Watson tree and that

$$P[\tau_i - \tau_j > t | p(\tau_i) = \tau_j] = \int_t^\infty h(s)ds. \tag{12}$$

This allows for different view of the model which we call parent-child interpretation and in the case when λ is linear a Galton-Watson interpretation.

4. Results.

Proposition 1. Hawkes process is well defined whenever exist $A, B \geq 0$ s.t. for $\bar{\lambda}(z) = A + Bz$, $\lambda \leq \bar{\lambda}$ i.e.

$$\forall z \in \mathbb{R}, \qquad \lambda(z) \leq \bar{\lambda}(z). \tag{13}$$

Proposition 2. If we start with $g_0 = 0$, then $g_t(\cdot) \geq 0$ at time t and therefore for any function $f(x_1, \ldots, x_k)$ of k variables that is non decreasing in each x_i, and any k non-overlapping intervals $J_1, \ldots, J_k$

$$E^{P_0}\left[f(N(t + J_1), \ldots, N(t + J_k))\right] \tag{14}$$

as well as

$$E^{P_0}\left[f(g(t + s_1), \ldots, g(t + s_k))\right] \tag{15}$$

are increasing functions of t and have a limit defining a stationary point process, that can be extended to the entire line $(-\infty, \infty)$. Here $J_i + t$ is the interval J_i shifted by t. This process is minimal and is ergodic.

Proposition 2 provides stationary distribution; in particular stationary distribution of outlook-function. We will call this distribution μ_0.

Theorem 3. If in addition we also have:

$$\lambda(z) \leq \bar{\lambda}(z) := a + bz, \qquad b < 1 \tag{16}$$

$$\sup_{x \in \mathbb{R}_+} (\lambda(x + s) - \lambda(x)) \leq \phi(s) \tag{17}$$

for some concave increasing ϕ satisfying

$$\int_0^\infty \phi(H(s))ds < \infty, \tag{18}$$

where

$$H(s) = \int_s^\infty h(t)dt. \tag{19}$$

Then starting from any initial condition g in

$$\mathcal{C} := \left\{ g \in C(\mathbb{R}_+) : \int_0^\infty \phi(g(s))ds < \infty \right\} \tag{20}$$

the distribution will converge

$$\mu_{g,s} \xrightarrow[s\to\infty]{} \mu_0 \tag{21}$$

Furthermore proof provides coupling with the process starting from 0 initial condition that differs, almost surely, in only finitely many points. Hence μ_0 is unique invariant measure supported on the space $\mathcal{C}$ that is $\mu_0(\mathcal{C}) = 1$.

Remark. *This generalizes* [1] *and now does not require Lipschitz condition for* λ *and gives result for Lipschitz* λ *as corollary. Furthermore the restrictions for* ϕ *stated are taken to be stricter than necessary. With addition of some technicalities one can take* ϕ *to be:*

$$\int \lambda(g(0) + x) - \lambda(g(0))d\mu_0(g). \tag{22}$$

5. Proofs.

Proof of Proposition 1: Consider first Hawkes process with rate $\bar{\lambda} = A + Bz$; by coupling this process is dominating and by Galton-Watson interpretation it is well defined for all times. Hence dominated process is well defined for all times. Now if $B < 1$ then dominating process is bounded in expectation (since the expected number of points in each tree is finite); then so is dominated process and the following standard argument gives us some invariant distribution: pick any inition condition g_0, for example $g_0 = 0$, then invariant π is obtained as a converging subsequence of Cesaro summations of future measures

$$\pi := \lim_{t_k\to\infty} \frac{1}{t_k}\int_0^{t_k} \pi_{s,g_0} ds \tag{23}$$

where $\pi_{s,g}$ represents distribution of the future at time s starting from initial condition g_0 at time 0. □

Proof of Proposition 2: We can now bound our process by Galton-Watson forest with $\beta < 1$ for which expected size of each tree is finite and hence we have finite density which tells us that all test functions are bounded and hence for non-decreasing f, the following limit is strictly increasing:

$$\lim_{t\to\infty} E^{P_0}\left[f(N(t+J_1), \ldots, N(t+J_k))\right] < \infty. \tag{24}$$

This gives us weak convergence. The obtained invariant π is minimum; since if we have any other invariant $\widetilde{\pi}$ then we can consider initial condition chosen randomly corresponding to measure $\widetilde{\pi}$; but then we have point-wise domination and by coupling we see that $\widetilde{\pi}$ dominates π. Hence π is also extremal invariant measure and hence ergodic. □

Before we start with proof of theorem 3 let us prove the following lemma.

Lemma. *For any Hawkes process with some stationary distribution* μ *of outlook-function then*

$$\mathbb{E}_\mu[g(s)] = \mathbb{E}_\mu[g(0)]H(s). \tag{25}$$

Proof. By stationarity

$$\mathbb{E}_\mu[g(s)] = \mathbb{E}_\mu\left[\int_s^\infty g(0)h(t)dt\right] = \mathbb{E}_\mu[g(0)]\int_s^\infty h(t)dt = \mathbb{E}[g(0)]H(s). \tag{26}$$

□

Proof of Theorem 3: We use recurrence argument. It is enough to show that exists some class $\mathcal{C}'\subset\mathcal{C}$ such that:

i) $f \in \mathcal{C}'$ implies that π_f and π_0 have non-trivial intersection

$$I(\pi_f, \pi_0) := 1 - d_{TV}(\pi_f, \pi_0) > 0 \tag{27}$$

ii) that process is $\mathcal{C}'$-recurrent.

iii) the class $\mu(\mathcal{C}) = 1$

We are going to choose $\mathcal{C}'$ later. First observe that by applying Jensen and concavity of ϕ we obtain:

$$\begin{aligned}
I(\pi_f, \pi_0) &:= 1 - d_{TV}(\pi_f, \pi_0) && (28)\\
&= \mathbb{E}_0\left[\exp\left(-\int_0^\infty \lambda(z_t + f(t)) - \lambda(z_t)dt\right)\right] && (29)\\
&\geq \exp\left(-\mathbb{E}_0\left[\int_0^\infty \lambda(z_t + f(t)) - \lambda(z_t)dt\right]\right) && (30)\\
&\geq \exp\left(-\mathbb{E}_0\left[\int_0^\infty \phi(f(t))dt\right]\right) && (31)\\
&= \exp\left(-\int_0^\infty \phi(f(t))dt\right) && (32)
\end{aligned}$$

Which suggests us to take

$$\mathcal{C}' := \left\{g \in C(\mathbb{R}_+) : \int_0^\infty \phi(g(s))ds < C\right\} \tag{33}$$

and to satisfy $\mathcal{C}'$-recurrent condition we will choose C as follows. We know that for any consider starting from $\mathcal{C}$ but with $\bar{\lambda}$. Then applying domination, Jensen, Lemma and concavity of ϕ we obtain

$$\begin{aligned}
\limsup_{t\to\infty} \mathbb{E}_g^{\lambda}\left[\int_0^\infty \phi(g_t(s))ds\right] &\leq \limsup_{t\to\infty} \mathbb{E}_g^{\bar{\lambda}}\left[\int_0^\infty \phi(g_t(s))ds\right] && (34)\\
&\leq \lim_{t\to\infty} \mathbb{E}_0^{\bar{\lambda}}\left[\int_0^\infty \phi(g_t(s))ds\right] && (35)\\
&\leq \lim_{t\to\infty} \int_0^\infty \phi(\mathbb{E}_0^{\bar{\lambda}}\left[g_t(s)\right])ds && (36)\\
&\leq \lim_{t\to\infty} \int_0^\infty \phi(\mathbb{E}_0^{\bar{\lambda}}[g_t(0)]H(s))ds && (37)\\
&\leq \lim_{t\to\infty} \mathbb{E}_0^{\bar{\lambda}}[g_t(0)]\left[\int_0^\infty \phi(H(s))ds\right] && (38)
\end{aligned}$$

where the last expression is finite since $\lim_{t\to\infty} \mathbb{E}_0^{\bar{\lambda}}[g_t(0)]$ is finite by Galton-Watson interpretation and $\int_0^\infty \phi(H(s))ds < \infty$ by (18) and hence setting C to be (38) guarantees recurrence to the class and we are done. □

6. Multi-type and Other Generalizations.

This section generalizes theorem 3 to multi-type and also allows to include δ functions into shocks h. The proofs are direct extensions of ideas presented; in this paper for clarity and brevity purposes proofs were presented for single-type process with h being a function rather than distribution; we will not reprove statement of theorem 3 for these generalization. Yet we present here the conditions for which theorem 3 will hold. The key players then are as before λ and h but they are now more complicated.

Consider that points now have a type from E. Hawkes process is now defined by functional h from E to space of measures on $\mathbb{R}_+ \times E$ and functional $\kappa : E \times (\mathbb{R}_+ \cup \{\infty\}) \to \mathbb{R}_+$ continuous at ∞ which represent what happens to δ masses (which in our case become Poisson variables with corresponding rate); then let $\lambda_a(z) = z\kappa_a(z)$. Then the doubly infinitesimal rate at time t of type a is given by

$$\alpha(da) + \int_E \lambda_{da} \left(\sum_{\tau_i < t} h_{a_i}(dt - \tau_i, da) \right) \tag{39}$$

where a_i is a type of τ_i.

Then μ_0 still exists whenever the process does not blow up and Theorem 3 holds whenever for some

$$\phi(a, x) \geq \mathbb{E}_{\mu_0}[\lambda_a(z_a + x) - \lambda_a(z_a)] \tag{40}$$

convex in x and

$$\int_E \int_{\mathbb{R}_+} \lambda_a(H(da, dt)) < \infty \tag{41}$$

where H is defined by

$$H(da, dt) = \int_g \int_{db \in E} \int_{s=t}^{\infty} h(ds, db, da) \lambda(db, g_a(0)) \mu_0(dg) \tag{42}$$

The μ_0 would exist whenever spectrum of the operator from E to E defined by

$$\nu(a, db) = \|h_a(\cdot, db)\|_1 \lim_{z \to \infty} \kappa_a(z) \tag{43}$$

is supported on $\{c \in \mathbb{C} : |c| \leq 1\}$.

Acknowledgments.

I would like to thank my adviser S. R. Srinivasa Varadhan for giving this topic to me and Lingjiong Zhu. He not only gave us this rich and interesting topic, but also provided guidance in motivation, formulation of results and clear view of the subject. I also thank Lingjiong Zhu for helpful discussions and review of my paper.

The author is supported by NSF DMS-0904701 and DARPA grants.

References

[1] Brémaud, P., and L. Massoulié, *Stability of Nonlinear Hawkes Processes*, The Annals of Probability, Vol.24, No.3, 1563-1588, 1996

[2] Cox, D. R. and V. Isham, *Point Processes*, Chapman and Hall, 1980

[3] Daley, D. J. and D. Vere-Jones, *An Introduction to the Theory of Point Processes*, Volume I and II, Springer, Second Edition, 2003

[4] Erlang, A.K. *The Theory of Probabilities and Telephone Conversations*, Nyt Tidsskrift for Matematik B, vol 20, 1909.

[5] Hairer, M., *Convergence of Markov Processes*, Lecture Notes, University of Warwick, available at `http://www.hairer.org/notes/Convergence.pdf`, August 2010

[6] Hawkes, A. G., *Spectra of Some Self-Exciting and Mutually Exciting Point Processes*, Biometrika 58, 1, p83, 1971

[7] Hawkes, A. G. and D. Oakes, *A Cluster Process Representation of a Self-Exciting Process*, J. Appl. P. rob.1 1,4 93-503, 1974

[8] Jagers, P., *Branching Processes with Biological Applications*, John Wiley, London, 1975

[9] Liniger, T., *Multivariate Hawkes Processes*, PhD thesis, ETH, 2009

[10] Oakes, D., *The Markovian Self-Exciting Process*, Journal of Applied Probability, Vol. 12, No. 1, Mar., 1975

[11] Saichev, A. and Sornette, D., *Hierarchy of Temporal Responses of Multivariate Self-Excited Epidemic Processes*, arXiv:1101.1611v1.

[12] Stabile, G. and G. L. Torrisi, *Risk Processes with Non-Stationary Hawkes Arrivals*, Methodol. Comput. Appl. Prob. 12:415-429, 2010

[13] Zhu, L., *Large Deviations for General Markovian Hawkes Processes* , preprint, arXiv:1108.2432.

[14] Zhu, L., *Process-Level Large Deviations for General Hawkes Processes*, preprint, arXiv:1108.2431.

[15] Karabash, D. and Zhu, L., *Large Deviations for Marked Hawkes Processes* , preprint, arXiv:.