Simulation of stationary linear Hawkes processes: the subexponential case *

Technical report, September 2006

Giuditta Bravaccino † Pierre Brémaud ‡ Giovanna Nappo §

Abstract

Linear Hawkes processes form a class of branching point processes that is especially relevant to seismology and epidemiology. This work concerns the simulation of *stationary* linear Hawkes process when the exact sampling methods do not apply, in the case where the fertility rate of the process properly normalized is the density probability of a subexponential random variable (the known exact sampling method requires a light-tail assumption). This situation is precisely the one arising in the most popular model in seismology (the so-called ETA model, see the main text). Since exact sampling is not available, we have recourse to a simple and natural approximation, the accuracy of which is controlled by a unique parameter. The accuracy of such approximation is measured by the probability that the sample obtained is not typical. We give computable bounds for the accuracy that permit to assess the simulation.

1 The problem

A *linear marked Hawkes process* is one kind of marked point process. It is described by a sequence $\{(T_n\, Z_n)\}$, $n \in \mathbb{Z}$, where the sequence $\{Z_n\}$ of *marks* is a sequence of iid random variables with values in some arbitrary measurable space K, and the sequence of event times $\{T_n\}$ is associated to a point process (random counting measure) N by the formula

$$N(C) = \sum_{n\in\mathbb{Z}} \mathbf{1}_C(T_n), \qquad C \in \mathcal{B}(\mathbb{R}).$$

*Partially supported by MURST project 2005 - prot. C26A058977 - *Processi stocastici e applicazioni*

†Dipartimento di Metodi e Modelli Matematici – Università di Roma "La Sapienza", piazzale A. Moro 2, I 00185 Roma, Italy

‡School of Computer and Communication Sciences, EPFL, CH-1015 Lausanne, and INRIA/TREC, Département d'Informatique, École Normale Supérieure, 45 rue d'Ulm, 75005-Paris, France

§Dipartimento di Matematica - Università di Roma "La Sapienza", piazzale A. Moro 2, I 00185 Roma, Italy

Let $\{\mathcal{F}_t\}$, $t \in \mathbb{R}$, be the natural filtration of N. The (stochastic) $\mathcal{F}_t$-intensity of N has the form

$$\lambda(t) = \nu(t) + \sum_{n\in\mathbb{Z}} h(t - T_n, Z_n)\mathbf{1}_{\mathbb{R}}(T_n). \tag{1}$$

Here $\{\nu(t)\}_{t\in\mathbb{R}}$ is a given nonnegative locally integrable stochastic process and $h : \mathbb{R} \times L \to \mathbb{R}$ is a nonnegative measurable function such that

$$t < 0 \Rightarrow h(t, z) = 0\,.$$

Recall that the interpretation of $\mathcal{F}_t$-intensity is

$$\mathbb{E}[N(t, t + dt]\,|\mathcal{F}_t] = \lambda(t)dt\,.$$

Such processes are also called branching point processes (See below). The original Hawkes process in [9] has $\nu(t) = \nu > 0$, a deterministic constant, and $h(t, z) = h(t)$, independent of the mark argument. This process, called the *linear Hawkes process*, therefore has the stochastic intensity

$$\lambda(t) = \nu + \int_{(-\infty,t)} h(t - s)N(ds). \tag{2}$$

Hawkes processes arise as models in various domains of applied sciences, such as seismology, forestry, epidemiology (see for example, Daley and Vere-Jones [7]). Stationarity is an important feature for phenomena that have started "infinitely" far in the past, and therefore it is reasonable to consider stationary models since the stationary regime occurs (under some conditions) as the limit of transient regime (See Brémaud and Massoulié [3]).

The stationary version can be constructed from a stationary Poisson process $\overline{N}$ on $\mathbb{R}$, with average intensity ν, and a sequence of nonstationary linear Hawkes processes $\left\{N_0^{(n)}\right\}_{n\in\mathbb{N}}$ on $\mathbb{R}^+$, with intensity

$$\lambda_0^n(s) = h(s) + \int_{(0,s)} h(s - v)N_0^n(dv).$$

More precisely

$$N((0, t]) = \overline{N}((0, t]) + \sum_n N_0^{(n)}((0, t - S_n]),$$

where S_n are the event times of $\overline{N}$. Denoting by $T_{n,k}^{(0)}$ the event times of $N_0^{(n)}$, the stationary process N has event times

$$\left\{T_{n,k} = S_n + T_{n,k}^{(0)}\right\}.$$

The point processes $N_0^{(n)}$ also receive an interpretation as branching point processes (see below). Note that under the condition $\int_0^\infty h(t)dt < 1$ the branching point processes $N_0^{(n)}$ are subcritical, that is, have a finite number of points.

The event times S_n can be interpreted as the arrival times of "ancestors" whereas the event times $\{T_{n,k}\}_{k\geq 1}$ can be interpreted as the birth time of the "descendants" of the ancestor arrived at time S_n. (In a seismology context the ancestors and descendants are respectively the "main shocks" and the "after shocks")

Granted that a Poisson process and any of the *finite* $N_0^{(n)}$ processes are easy to construct (which is indeed the case[1]), there remains a major problem for simulating the restriction of the stationary version to positive times. Indeed it is easy to show that sufficiently far away in the past the ancestors have a lineage that is extinct at time 0, and therefore we need only to simulate a finite number of $N_0^{(n)}$, $n \leq -1$. But this number is random and cannot be in principle determined in advance since it depends on the whole past.

Recently Møller and Rasmussen ([13], [14]) have overcome this difficulty and given an exact sampling algorithm, as well as an approximate one ([12]). Their algorithms however depend crucially on a light-tail condition, that is (in the unmarked case for instance), the existence of a parameter $\theta > 0$ such that

$$\int_0^\infty e^{\theta t}\, h(t)\, dt \leq 1.$$

This forbids some applications of interest. For instance, a popular mathematical model in seismology (see Ogata [16] and [17]) uses for the fertility rate the so-called modified Omori function

$$h(t) = \frac{H}{(t+\kappa)^p}, \qquad \text{with } H,\ \kappa,\ p > 1 \text{ parameters.}$$

The above function $h(t)$ is proportional to the Pareto density function,

$$f_{Pareto}(t) = \frac{\alpha\kappa^\alpha}{(t+\kappa)^{\alpha+1}}, \qquad \text{with } \alpha = p - 1 > 0,\ .$$

Pareto distribution belongs to the general class of subexponential distributions, whose definition we recall.
Let $\{X_i\}_{i\in\mathbb{N}}$ be IID random variables with distribution function G such that $G(x) < 1$ for all $x > 0$. Denote by

$$\overline{G}(x) = 1 - G(x),$$

the tail of G, and by

$$\overline{G^{*n}}(x) = 1 - G^{*n}(x) = P(X_1 + \ldots + X_n > x)$$

the tail of the $n-$fold convolution of G.

Definition 1 (Subexponential distribution functions). *The subexponential class $\mathcal{S}$ consists of all distribution functions G which satisfy one of the following equivalent conditions:*

(a) $\lim_{x\to\infty} \frac{\overline{G^{*n}}(x)}{\overline{G}(x)} = n$ *for some (and then for all)* $n \geq 2$*;*

[1] The classical Lewis–Ogata algorithm ([15]), based on the thinning algorithm of Lewis and Shedler [10] for nonhomogeneous Poisson processes, works for point processes with *empty initial conditions*, such as $N_0^{(n)}$.

(b) $\lim_{x\to\infty} \frac{P(X_1+\ldots+X_n>x)}{P(max(X_1,\ldots,X_n)>x)} = 1$ *for some (and then for all)* $n \geq 2$.

The class of subexponential distributions was introduced by Chistyakov ([6]) who proved that the limit in (a) holds for all $n \geq 2$ if and only if it holds for $n = 2$. The main properties of this class can be found in Athreya and Ney [1], and more recently in Embrechts, Kluppelberg, and Mikosch [8].
Subexponential distributions are heavy tailed: more precisely it holds (see [8])

$$\lim_{t\to\infty} e^{\theta t}[\overline{G}(t)] = \infty, \qquad \text{for all } \theta > 0.$$

We recall for future reference the following basic result(see e.g. [1]) concerning subexponential distribution.

Lemma 1 (Kesten's lemma). *If* $G \in \mathcal{S}$, *then for every* $\epsilon > 0$ *there exists some positive constant* $K(\epsilon)$ *such that for all* $n \in \mathbb{N}$ *and* $t > 0$

$$\frac{\overline{G^{*n}}(t)}{\overline{G}(t)} \leq K(\epsilon)(1+\epsilon)^n.$$

A possible expression for the constant $K(\epsilon)$ is, for instance,

$$K(\epsilon) = [\epsilon(1 - G(T_\epsilon))]^{-1}, \tag{3}$$

where T_ϵ is defined as in (21).
One could also take, instead of $K(\epsilon)$, the smaller constant

$$\kappa(\epsilon) := \frac{1 + \epsilon\,\overline{G}\,(T_\epsilon)}{1+\epsilon}\, K(\epsilon). \tag{4}$$

Althhough the value of the constant $K(\epsilon)$ is not given explicitly in [1], the values suggested here can be easily deduced from the proof of the lemma in [1], and are given in [2].

We now give the conditions on h that are assumed to hold throughout the present work.

H1 The function $h: \; t \mapsto h(t)$ is directly Riemann integrable and locally bounded on $\mathbb{R}^+$;

H2 It moreover satisfies the further conditions

$$\overline{h} = \int_0^\infty h(t)dt < 1, \tag{5}$$

$$\int_0^\infty t\, h(t)\, dt < \infty; \tag{6}$$

H3 The probability distribution with density

$$g(t) = \frac{h(t)}{\int_0^\infty h(s)\, ds} = \frac{h(t)}{\overline{h}}, \tag{7}$$

is subexponential. (We shall denote by X a nonnegative random variable with this distribution.)

Condition **H1** allows us to use the basic renewal theorem, while condition $\overline{h} < 1$ in **H2** guarantees the existence and uniqueness in law of the stationarity process N (see e.g. Brémaud and Massoulié [3]).

2 The Approximate Sample; Simple Bounds.

Since the available exact sampling algorithms do not apply to our situation, we have to resort to approximate sampling. Instead of looking for the last ancestor in the prehistory (before time 0) with an influence on the process after time 0, we restrict attention to only those in a fixed interval $[-a, 0)$ hoping that those in $(\infty, -a]$ have no descendents after time 0.
Therefore, we replace N by its approximation N_a defined to be the linear Hawkes process without ancestors before time $-a$. Having simulated N_a we hope that its restriction to $(0, \infty)$ coincides with the restriction to $(0, \infty)$ of N. For this to make sense we couple N_a to N by using the same point processes $\overline{N}$ and $N_0^{(n)}$, $n \in \mathbb{Z}$, or equivalently, the same sequences $\{S_n\}$ and $\{T_{n,k}^{(0)}\}$. The sequence of points associated to N_a is therefore

$$\left\{T_{n,k} = S_n + T_{n,k}^{(0)},\ n : S_n \geq -a\right\}.$$

Denote by $N|_C$ the restriction on the point process N to the set C. Defining the set

$$A_a = \{N_a|_{[0,\infty)} = N|_{[0,\infty)}\}, \tag{8}$$

the accuracy of the approximation is measured by $P(A_a)$ which should ideally be close to 1. More precisely, given $\delta > 0$, the aim is to find $a = a(\delta)$ such that

$$\mathbb{P}(A_a) \geq 1 - \delta, \tag{9}$$

and then, by the coupling inequality (see [11]),the total variation distance between the laws of $N_a|_{[0,\infty)}$ and of $N|_{[0,\infty)}$ is less or equal to 2δ.

Let now N_0 be a Hawkes process with the same distribution as $N_0^{(n)}$ and let T_E be its extinction time. Here is an exact expression of the quantity of interest $\mathbb{P}(A_a)$ in terms of the distribution of this extinction time:

$$F_E(t) = \mathbb{P}(T_E \leq t).$$

Lemma 2 ([2]).

$$\mathbb{P}(A_a) = \exp\Big\{-\nu \int_a^\infty \overline{F}_E(s)ds\Big\} = \exp\Big\{-\nu\mathbb{E}\big[(T_E - a)^+\big]\Big\}, \tag{10}$$

In particular, if the expectation $\mathbb{E}[T_E]$ *is finite*

$$\lim_{a\to\infty} \mathbb{P}(A_a) = 1.$$

Proof. The event A_a is realized if and only if for all n such that $S_n < -a$, one has $S_n + T_E^n < 0$, where the T_E^n's are the IID extinction times of the $N_0^{(n)}$'s. Therefore

$$\mathbb{P}(A_a) = \mathbb{E}[\mathbf{1}_{A_a}] = \mathbb{E}\left[\prod_{n:S_n<-a} \mathbf{1}_{(S_n+T_E^n<0)}\right],$$

then, by conditioning with respect to the Poisson process $\overline{N}$,

$$\begin{aligned}\mathbb{P}(A_a) &= \mathbb{E}\left[\mathbb{E}\left[\prod_{n:S_n<-a} \mathbf{1}_{(T_E^n<-S_n)}\Big|\overline{N}\right]\right] \\ &= \mathbb{E}\left[\prod_{n:S_n<-a} F_E(-S_n)\right] \\ &= \mathbb{E}\left[\exp\left(\sum_n \log\left(I_{(-\infty,-a)}(S_n)F_E(-S_n)\right)\right)\right].\end{aligned}$$

Then the exponential formula, applied for the function $\varphi(t) = -\log\left(I_{(-\infty,-a)}(t)F_E(-t)\right)$, implies (10). (Recall the exponential formula for a homogeneous Poisson process with intensity ν: for any measurable function φ with values in $\mathbb{R}^+ \cup \{\infty\}$

$$\mathbb{E}\left[\exp\left(-\int_{\mathbb{R}} \varphi(t)\overline{N}(dt)\right)\right] = \exp\left\{\int_{\mathbb{R}} \left(e^{-\varphi(t)} - 1\right)\nu dt\right\}.)$$

□

The analytic expression of $F_E(t)$, however, is not known; and therefore we have to work a little more. We shall naturally look for a function $L(t)$ integrable on $(0, \infty)$, such that

$$\overline{F}_E(t) \le L(t).$$

The following simple lemma appears in [4].

Lemma 3. *Let $\lambda_0(t)$ be the intensity of the Hawkes process N_0, and define*

$$\overline{\lambda}_0(t) = \mathbb{E}[\lambda_0(t)].$$

Then

$$\mathbb{P}(T^E > t) \le \int_t^\infty \overline{\lambda}_0(s)ds. \tag{11}$$

Moreover, $\overline{\lambda}_0(t)$ satisfies the defective renewal equation

$$\overline{\lambda}_0(t) = h(t) + \overline{h}\int_0^t \overline{\lambda}_0(t-s)dG(s), \tag{12}$$

and

$$\bar{\lambda}_0(t) = \left(h * \sum_{i\ge 0} \overline{h}^i G^{*i}\right)(t). \tag{13}$$

Inequality (11) allows to take

$$L(t) = \int_t^\infty \overline{\lambda}_0(s)ds$$

However this is not satisfactory, since the expression (13) of $\overline{\lambda}_0(t)$ is not explicit enough. The previous result allows to state the following propositions, which give the upper bounds for $\overline{F}_E$, all of them stated under the same hypotheses of Proposition **??**.

3 Towards practical bounds.

Proposition 4 (Extinction time 2 [5]). *For all K such that*

$$K > \frac{\overline{h}}{(1-\overline{h})^2}$$

there exists $\overline{t} = \overline{t}(K)$ such that for all $t \geq \overline{t}$

$$\mathbb{P}(T^E > t) = \overline{F}_E(t) \leq K\,\overline{G}(t) = K\mathbb{P}(X > t). \tag{14}$$

The proof is reproduced in [2].

The proof of the following proposition is based on Lemma 3 and on Kesten's Lemma.

Proposition 5 (Extinction time 3, [2]). *For all $\epsilon \in \left(0, \frac{1-\overline{h}}{\overline{h}}\right)$ and for all $t > 0$*

$$\mathbb{P}(T^E > t) \leq \frac{K(\epsilon)\overline{h}(1+\epsilon)}{1-\overline{h}(1+\epsilon)}\,\overline{G}(t) = \frac{K(\epsilon)\overline{h}(1+\epsilon)}{1-\overline{h}(1+\epsilon)}\,\mathbb{P}(X > t), \tag{15}$$

where $K(\epsilon) = [\epsilon\,\overline{G}(T_\epsilon)]^{-1}$ as in Lemma 1, and where T_ϵ satisfies condition (21).

Proof. Kesten's Lemma assures that

$$\overline{G}^{*n}(t) \leq K(\epsilon)\overline{G}(t)(1+\epsilon)^n \;\forall t; \tag{16}$$

so, because of (13)

$$\int_t^\infty \overline{\lambda}(s)ds = \sum_{i=0}^{\infty} \overline{h}^{i+1}\overline{G}^{*(i+1)}(t), \tag{17}$$

it holds

$$\mathbb{P}(T_n^E > t) \leq \frac{K(\epsilon)\overline{h}(1+\epsilon)}{1-\overline{h}(1+\epsilon)}\overline{G}(t), \tag{18}$$

for every ϵ s.t. $\overline{h}(1+\epsilon) < 1$, that is $\epsilon < (1-\overline{h})/\overline{h}$. It only remains to note that $K(\epsilon)$ is taken as in formula (3). □

Finally 6 and 7 are the consequence of Lemma 2 and of the above Propositions 4 and 5, respectively.

Remark 1. *The quantity a mentioned in Proposition 4 has to be taken bigger or equal to $t^* = t^*(\epsilon)$ or $\tilde{t} = \tilde{t}(\epsilon)$, respectively. Though t^* and $\tilde{t}$ can be given easily in terms of $\overline{t}(K)$ and $t_0(K)$, these values seem hard to be computed, either explicitly or numerically. On the other hand, this computational problem does not show up when using the bound given in Proposition 5, the bound being valid for all t; however also this bound has a drawback, due to the facts that the explicit computation of T_ϵ is necessary and that*

$$\lim_{\epsilon\to 0} K(\epsilon) = \lim_{\epsilon\to 0} \left[\epsilon\left(1 - G\left(T_\epsilon\right)\right)\right]^{-1} = \infty.$$

Summarizing, to accomplish completely the purpose of this work by means of the previous upper bounds, one need to compute T_ϵ, for a properly chosen value of ϵ: the value of ϵ has to be taken small in such a way that $[1-\overline{h}(1+\epsilon)]^{-1}$ is not too large, but not too small in such a way that $K(\epsilon)$ is not too large.

Combining the various upper bounds for the survival function $\overline{F}_E(t)$ of Propositions 4 and 5) we obtain our main results.

Proposition 6. *For every $\epsilon > 0$, there exists a $t^* = t^*(\epsilon)$, such that for $a \geq t^*$*

$$\mathbb{P}(A_a^c) \leq 1 - \exp\left\{-\nu\left(\frac{\overline{h}}{(1-\overline{h})^2} + \epsilon\right)\mathbb{E}\left[(X-a)^+\right]\right\}. \tag{19}$$

Proposition 7. *For every $a > 0$, and for all $\epsilon \in \left(0, \frac{\overline{h}}{1-\overline{h}}\right)$*

$$\mathbb{P}(A_a^c) \leq 1 - \exp\left\{-\nu\,\frac{1}{\epsilon\,\overline{G}(T_\epsilon)}\,\frac{\overline{h}(1+\epsilon)}{1-\overline{h}(1+\epsilon)}\,\mathbb{E}\left[(X-a)^+\right]\right\}, \tag{20}$$

where T_ϵ is such that

$$\sup_{t \geq T_\epsilon} \frac{\overline{G^{*2}}(t)}{\overline{G}(t)} \leq 2 + \epsilon. \tag{21}$$

Remark 2. *First of all it has to be noticed that the upper bounds (19) and (20) for $\mathbb{P}(A_a^c)$ give useful information only if they are finite quantities. For the last two bounds this is assured by condition (6), which is equivalent to require that the first moment of X is finite.*

As already said the aim is to compute the quantity $a = a(\delta)$ mentioned above. In this respect Proposition 6 presents a drawback, in that the computation of $\tilde{t}$ and t^* seems not feasible. On the other hand from Proposition 7 we know that in order to determine an a such that $\mathbb{P}(A_a^c) \leq \delta$, it is sufficient to take $a = a_\epsilon(\delta)$ such that

$$\mathbb{E}\left[(X-a)^+\right] \leq \frac{-\log(1-\delta)}{\nu}\,\frac{\overline{G}(T_\epsilon)}{M(\epsilon)}, \tag{22}$$

where

$$M(\epsilon) = \frac{\overline{h}(1+\epsilon)}{\epsilon\,\left(1-\overline{h}(1+\epsilon)\right)}, \tag{23}$$

and in order to use the previous upper bound we only need to find a T_ϵ satisfying (21). This can be done numerically in some examples, while in the Pareto case T_ϵ can be explicitly computed, allowing to obtain the following result.

Proposition 8. *Assume $G(t)$ to be a Pareto distribution, with $\alpha > 1$. Then for any $\epsilon \in (0, \frac{\overline{h}}{1-\overline{h}}) \cap (0, 4(2^\alpha - 1)]$, inequality (9) holds with*

$$a = a_\epsilon(\delta) = \widehat{a}_\epsilon\left(\kappa, \alpha, \nu, \overline{h}\right)\,\frac{1}{|log\,(1-\delta)|^{\frac{1}{\alpha-1}}} - \kappa \tag{24}$$

where

$$\widehat{a}_\epsilon\left(\kappa,\alpha,\nu,\overline{h}\right)=\left[\frac{\kappa^\alpha\,\nu\,M(\epsilon)}{(\alpha-1)}\left(1+\frac{1}{\theta_\epsilon}\left[\left(\frac{2}{\epsilon}\right)^{1/\alpha}\frac{1}{\theta_\epsilon}-1\right]\right)^\alpha\right]^{\frac{1}{\alpha-1}}, \tag{25}$$

with $M(\epsilon)$ *given by* (23), *and* $\theta_\epsilon=1-\left(\frac{4}{4+\epsilon}\right)^{\frac{1}{\alpha}}$.

This result is proved in Section 4, and has been used in some simulations for the Pareto case, for a particular choice of $\overline{\epsilon}$, minimizing $M(\epsilon)$. With this choice $\overline{\epsilon}$, $\overline{M}=M(\overline{\epsilon})$ and $\overline{\theta}=\theta_{\overline{\epsilon}}$ are explicit functions of $\overline{h}$ and α, given respectively, by (31), (32) and (33), with $\gamma=1$. The quantity a, can be thought also as a function of $\overline{h}$, and clearly converges to infinity as $\overline{h}$ converges to 1.

4 An application to seismology

In this section we consider the case of Burr distributions, or more precisely the case when $h(t)=\overline{h}g(t)$, with

$$\overline{G}(x)=\left(\frac{\kappa}{\kappa+x^\gamma}\right)^\alpha=\left(1+\left(\frac{x}{\chi}\right)^\gamma\right)^{-\alpha} \tag{26}$$

where α, κ, γ are strictly positive, and $\chi=\kappa^{1/\gamma}$, so that

$$g(t)=\frac{d}{dt}\left(1-\left(\frac{\kappa}{\kappa+t^\gamma}\right)^\alpha\right)=\frac{\alpha\gamma\kappa^\alpha\,t^{\gamma-1}}{(\kappa+t^\gamma)^{\alpha+1}},$$

We start by applying the result of Proposition 7 in order to find a more explicit upper bound for $\mathbb{P}(A_a^c)$ (see Proposition 10). In the case of Pareto distributions, i.e. when $\gamma=1$, this bound immediately leads to the announced value $a_\epsilon(\delta)$ given in (24) of Proposition 8. Successively the criterion used in the simulation in order to choose a particular value of ϵ is briefly discussed at the end of this section.

As a first step, for any $\epsilon\in\left(0,(1-\overline{h})/\overline{h}\right)$, a constant T_ϵ satisfying (21) of Proposition 7 is computed. In order to find the value of T_ϵ explicitly a preliminary result is needed. This result is used in the proof of Lemma 1.3.1 of [8] to show that the class of distribution functions $F(x)$, with $x^\alpha\,\overline{F}(x)$ a slowly varying function is closed with respect to convolution.

Proposition 9. *Let* $G(t)$ *be a subexponential distribution. Define*

$$\psi(\theta):=\sup_{x\geq 0}\frac{\overline{G}(\theta x)}{\overline{G}(x)}.$$

Then for any T, *and for any* $\theta\in(0,\frac{1}{2})$,

$$\sup_{x\geq T}\frac{\overline{G^{*2}}(x)}{\overline{G}(x)}\leq 2\,\psi(1-\theta)+\psi(\theta)\,\overline{G}(\theta\,T).$$

Proof. It makes use essentially of this fact: if X_1 and X_2 are two independent random variable, with values in $[0,\infty)$, then for all $\theta \in (0, \frac{1}{2})$

$$\begin{aligned}\mathbb{P}(X_1 + X_2 > x) \le \mathbb{P}(X_1 > (1-\theta)x) + \mathbb{P}(X_2 > (1-\theta)x) \\ + \mathbb{P}(X_1 > \theta x)\mathbb{P}(X_2 > \theta x)\end{aligned}$$

or equivalently,

$$\overline{F_1 * F_2}(x) \le \overline{F}_1((1-\theta)x) + \overline{F}_2((1-\theta)x) + \overline{F}_1(\theta x)\overline{F}_2(\theta x),$$

where $F_1(x)$ and $F_2(x)$ are the distribution functions of X_1 and X_2 respectively (see also [2] for the details). □

Proposition 10. *Assume that $G(x)$ belongs to the class of Burr distributions given in* (26).

Then, for every $\epsilon \in \big(0, (1-\overline{h})/\overline{h}\big) \cap \big(0, 4(2^{\alpha\,\gamma}-1)\big]$ and $a \in (0,\infty)$, it holds

$$\mathbb{P}(A_a^c) \le 1 - \exp\left\{-\nu\, \tfrac{J(\epsilon)}{M(\epsilon)}\, \mathbb{E}\left[(X-a)^+\right]\right\}, \tag{27}$$

where $J(\epsilon) = \left(1 + \left(\frac{1}{\theta_\epsilon}\right)^\gamma \left[\left(\frac{2}{\epsilon}\right)^{1/\alpha}\left(\frac{1}{\theta_\epsilon}\right)^\gamma - 1\right]\right)^\alpha$, *$M(\epsilon)$ as in* (23), *and* $\theta_\epsilon = 1 - \left(\frac{4}{4+\epsilon}\right)^{\frac{1}{\alpha\,\gamma}}$.

Proof. In this case, the function ψ can be computed explicitly, and as $\psi(\theta) = \theta^{-\alpha\,\gamma}$, ψ is continuous, as well as $\overline{G}$, and so the upper bound of the previous Proposition holds for all $\theta \le 1/2$:

$$\sup_{x\ge T} \frac{\overline{G^{*2}}(x)}{\overline{G}(x)} \le 2\,(1-\theta)^{-\alpha\,\gamma} + \frac{1}{\theta^{\alpha\,\gamma}} \left(1 + \left(\frac{\theta\,T}{\chi}\right)^\gamma\right)^{-\alpha}.$$

As a consequence, for any $\epsilon > 0$, one can take $\theta = \theta_\epsilon \le 1/2$ such that

$$2\,(1-\theta)^{-\alpha\,\gamma} = 2 + \frac{\epsilon}{2}, \tag{28}$$

and thereafter one can take $T = T_\epsilon$ in such a way that

$$\psi(\theta_\epsilon)\,\overline{G}(\theta_\epsilon T) = \theta_\epsilon^{-\alpha\,\gamma} \left(1 + \left(\frac{\theta_\epsilon\,T}{\chi}\right)^\gamma\right)^{-\alpha} = \frac{\epsilon}{2}. \tag{29}$$

Therefore (28) holds by taking $\theta = \theta_\epsilon$ as given in the Proposition, which is less or equal to $1/2$ if $\epsilon \le 4\big(2^{\alpha\,\gamma}-1\big)$; finally (29) holds by taking $T_\epsilon = \frac{\chi}{\theta_\epsilon}\left[\left(\frac{2}{\epsilon}\right)^{1/\alpha}\left(\frac{1}{\theta_\epsilon}\right)^\gamma - 1\right]^{1/\gamma}$. Then by (20), it holds (27), taking into account that in this case

$$\overline{G}(T_\epsilon) = J(\epsilon)^{-1}. \tag{30}$$

□

In the Burr example a particular bound can be found for every fixed ϵ, but how to choose the best ϵ?

This problem corresponds to minimizing the r.h.s. of (27) with respect to ϵ, or equivalently to maximizing the r.h.s. of (22), i.e. it corresponds to minimizing the function $M(\epsilon)/\overline{G}(T_\epsilon)$, where $\overline{G}(T_\epsilon)$ is given in (30).

Minimizing this function analytically is a heavy task, and so an alternative way to proceed is: first minimize $M(\epsilon)$ on the admissible values of ϵ, in order to choose $\epsilon = \overline{\epsilon}$, and then give the bound taking $T_\epsilon = T_{\overline{\epsilon}}$. This procedure leads to another bound as follows.

A first attempt to optimization of this latter upper bound can consist in finding the value of $\epsilon \in (0, (1-\overline{h})/\overline{h})$ minimizing the value of $M(\epsilon)$. This procedure would have the advantage of being the same for every subexponential distribution. However, if we want to take into account the upper bound obtained in the Burr case, also condition $\epsilon \leq \epsilon_0 = \epsilon_0(\alpha\,\gamma) = 4(2^{\alpha\,\gamma}-1)$ is needed, and therefore one has to minimize $M(\epsilon)$ over $\epsilon \in \big(0, \epsilon_0 \wedge [(1-\overline{h})/\overline{h}]\big)$. Some computations (see [2]) give the minimizing value

$$\overline{\epsilon} = \Big(\frac{1}{\sqrt{\overline{h}}} - 1\Big) \wedge \epsilon_0, \tag{31}$$

and, denoting

$$h_0 := h_0(\alpha\,\gamma) = \frac{1}{(2^{2+\alpha\,\gamma}-3)},$$

the minimizing constant

$$\overline{M} = \min_{0<\epsilon\leq\epsilon_0\wedge[(1-\overline{h})/\overline{h}]} M(\epsilon) = M(\overline{\epsilon}) = \begin{cases} \frac{\overline{h}}{\left(1-\sqrt{\overline{h}}\right)^2}, & \text{if } \overline{h} \geq h_0^2, \\ \frac{\overline{h}}{(1-h_0)\left(1-\frac{\overline{h}}{h_0}\right)} & \text{otherwise.} \end{cases} \tag{32}$$

Finally

$$\overline{\theta} = \theta_{\overline{\epsilon}} = \begin{cases} 1-\left(1-\frac{1-\sqrt{\overline{h}}}{3\sqrt{\overline{h}}+1}\right)^{\frac{1}{\alpha\,\gamma}}, & \text{for } \overline{h} \geq h_0^2 \\ \theta_{\epsilon_0} = \frac{1}{2}, & \text{for } \overline{h} < h_0^2. \end{cases} \tag{33}$$

Note that $\overline{\epsilon}$, $\overline{M}$ and $\overline{\theta}$ are explicit functions of $\overline{h}$ and of the product $\alpha\,\gamma$.

In the case of the Pareto distribution, i.e. when $\gamma = 1$,

$$\mathbb{E}\left[(X-a)^+\right] = \frac{\kappa^\alpha}{\alpha-1}\,\frac{1}{(\kappa+a)^{\alpha-1}} \tag{34}$$

and therefore, by taking into account (22) and (30), and by means of the previous Proposition, an explicit value for $a = a_\epsilon(\delta)$ can be computed. It turns out that this value of a is given by (24), and the proof of Proposition 8 is accomplished. Moreover, when taking $\overline{\epsilon}$ as in (31), then the explicit expression of $\widehat{a}_{\overline{\epsilon}}$ is

$$\widehat{a}_{\overline{\epsilon}}\left(\kappa,\alpha,\nu,\overline{h}\right) = \left[\frac{\kappa^\alpha\,\nu\,\overline{M}}{(\alpha-1)}\left(1+\frac{1}{\overline{\theta}}\left[\left(\frac{2}{\overline{\epsilon}}\right)^{1/\alpha}\frac{1}{\overline{\theta}}-1\right]\right)^{\alpha}\right]^{\frac{1}{\alpha-1}}$$

i.e. for $\overline{h} \geq h_0^2$

$$= \left[\frac{\kappa^\alpha\,\nu\,\overline{h}}{\left(\alpha-1\right)\left(1-\sqrt{\overline{h}}\right)^2}\right]^{\frac{1}{\alpha-1}} \left[1+\frac{1}{1-\left(\frac{4\sqrt{\overline{h}}}{3\sqrt{\overline{h}}+1}\right)^{1/\alpha}}\left(\frac{\left(\frac{2\sqrt{\overline{h}}}{1-\sqrt{\overline{h}}}\right)^{1/\alpha}}{1-\left(\frac{4\sqrt{\overline{h}}}{3\sqrt{\overline{h}}+1}\right)^{1/\alpha}}-1\right)\right]^{\frac{\alpha}{\alpha-1}},$$

and for $\overline{h} < h_0^2$

$$= \left[\frac{\kappa^\alpha\,\nu\,\overline{h}}{\left(\alpha-1\right)\left(1-h_0\right)\left(1-\frac{\overline{h}}{h_0}\right)}\right]^{\frac{1}{\alpha-1}} \left[1+2\left(\left(\frac{2}{\epsilon_0}\right)^{1/\alpha}2-1\right)\right]^{\frac{\alpha}{\alpha-1}}.$$

References

[1] Athreya, K. B., and Ney, P. E. *Branching processes.* Springer-Verlag, New York, 1972. Die Grundlehren der mathematischen Wissenschaften, Band 196.

[2] Bravaccino, G. *Simulation of stationary linear Hawkes processes.* PhD thesis, Università di Roma "La Sapienza", 2004. http://www.dmmm.uniroma1.it/dottorato/bravaccino.pdf.

[3] Brémaud, P., and Massoulié, L. Stability of nonlinear Hawkes processes. *Ann. Probab. 24*, 3 (1996), 1563–1588.

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

[5] Brémaud, P., Nappo, G., and Torrisi, G. L. A note on the rate of extinction of marked Hawkes processes. Private communication, 2003.

[6] Čistjakov, V. P. A theorem on sums of independent positive random variables and its applications to branching random processes. *Teor. Verojatnost. i Primenen 9* (1964), 710–718.

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

[8] Embrechts, P., Klüppelberg, C., and Mikosch, T. *Modelling extremal events*, vol. 33 of *Applications of Mathematics (New York).* Springer-Verlag, Berlin, 1997. For insurance and finance.

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

[10] Lewis, P. A. W., and Shedler, G. S. Simulation of nonhomogeneous Poisson processes by thinning. *Naval Res. Logist. Quart. 26*, 3 (1979), 403–413.

[11] Lindvall, T. *Lectures on the coupling method.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1992. A Wiley-Interscience Publication.

[12] Møller, J., and Rasmussen, J. G. Approximate simulation of Hawkes processes. MaPhySto Research Report n.28; http://www.maphysto.dk/publications2/MPS-RR/2004/28.pdf, 2004.

[13] Møller, J., and Rasmussen, J. G. A note on a perfect simulation algorithm for marked hawkes processes. In *In Spatial point process modelling and its applications; E ditors: A. Baddeley, P. Gregori, J. Mateu, R. Stoica and D. Stoyan.* Publicacions de la Universitat Jaume I, 2004, pp. 187–192.

[14] Møller, J., and Rasmussen, J. G. Perfect simulation of Hawkes processes. *Adv. in Appl. Probab.* (2005). (to appear).

[15] Ogata, Y. On Lewis' simulation method for point processes. *IEEE Trans. Inf. Theory 27* (1981), 23–31.

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

[17] Ogata, Y. Exploratory analysis of earthquake clusters by likelihood-based trigger models. *J. Appl. Probab. 38A* (2001), 202–212. Probability, statistics and seismology.