3 December 2004

APPROXIMATE SIMULATION OF HAWKES PROCESSES

JESPER MØLLER,* *Aalborg University*

JAKOB G. RASMUSSEN,** *Aalborg University*

Abstract

This article concerns a simulation algorithm for unmarked and marked Hawkes processes. The algorithm suffers from edge effects but is much faster than the perfect simulation algorithm introduced in our previous work [12]. We derive various useful measures for the error committed when using the algorithm, and we discuss various empirical results for the algorithm compared with perfect simulations.

Keywords: Approximate simulation; edge effects; Hawkes process; marked Hawkes process; marked point process; perfect simulation; point process; Poisson cluster process; thinning

AMS 2000 Subject Classification: Primary 60G55
Secondary 68U20

1. Introduction

This paper concerns a useful simulation algorithm for unmarked and marked Hawkes processes [5, 6, 7, 8, 10]. Such processes are important in point process theory and its applications, cf., for example, p. 183 in [5]. Particularly, marked Hawkes processes have applications in seismology [9, 13, 14, 15] and neurophysiology [2, 4]. The algorithm in this paper suffers from edge effects but is of more practical importance than the perfect simulation algorithm introduced in our earlier work [12].

There are many ways to define a marked Hawkes process, but for our purpose it is most convenient to define it as a marked Poisson cluster process $X = \{(t_i, Z_i)\}$

* Postal address: Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, DK-9220 Aalborg, Denmark. Email address: jm@math.auc.dk

** Postal address: As above. Email address: jgr@math.auc.dk

with events (or times) $t_i \in \mathbb{R}$ and marks Z_i defined on an arbitrary (mark) space M equipped with a probability distribution Q. The cluster centres of X correspond to certain events called *immigrants* and the rest of the events are called *offspring*.

Definition 1. (*Hawkes process with unpredictable marks.*)

(a) The immigrants follow a Poisson process with a locally integrable intensity function $\mu(t)$, $t \in \mathbb{R}$.

(b) The marks associated to the immigrants are i.i.d. with distribution Q and independent of the immigrants.

(c) Each immigrant t_i generates a *cluster* C_i, which consists of marked events of generations of order $n = 0, 1, \ldots$ with the following *branching structure*: First we have (t_i, Z_i), which is said to be of generation zero. Recursively, given the $0, \ldots, n$ generations in C_i, each $(t_j, Z_j) \in C_i$ of generation n generates a Poisson process Φ_j of offspring of generation $n+1$ with intensity function $\gamma_j(t) = \gamma(t - t_j, Z_j)$, $t > t_j$. Here γ is a non-negative measurable function defined on $(0, \infty)$. We refer to Φ_j as an *offspring process*, and to γ_j and γ as *fertility rates*. Furthermore, the associated mark Z_k to any offspring $t_k \in \Phi_j$ has distribution Q and Z_k is independent of t_k and all (t_l, Z_l) with $t_l < t_k$. As in [5] we refer to this as the case of *unpredictable marks*.

(d) The clusters given the immigrants are independent.

(e) Finally, X consists of the union of all clusters.

Simulation procedures for Hawkes processes are needed for various reasons: Analytical results are rather limited due to the complex stochastic structure; statistical inference, especially model checking and prediction require simulations; displaying simulated realisations of specific model constructions provide a better understanding of the model. The general approach for simulating a (marked or unmarked) point process is to use a thinning algorithm such as Shedler-Lewis thinning algorithm or Ogata's modified thinning algorithm, see e.g. [5]. However, Definition 1 immediately leads to the following simulation algorithm, where $t_- \in [-\infty, 0]$ and $t_+ \in (0, \infty]$ are user-specified parameters, and the output is all marked points (t_i, Z_i) with $t_i \in [0, t_+)$.

Algorithm 1. The following steps (i)-(ii) generate a simulation of those marked events

$(t_i, Z_i) \in X$ with $0 \leq t_i < t_+$.

(i) Simulate the immigrants on $[t_-, t_+)$.

(ii) For each such immigrant t_i, simulate Z_i and those $(t_j, Z_j) \in C_i$ with $t_i < t_j < t_+$.

Usually in applications steps (i) and (ii) are easy because (a)–(c) in Definition 1 are straightforward. As discussed in Section 4.4, Algorithm 1 and many of our results apply or easily extend to the case where the immigrant process is non-Poisson.

Ideally we should take $t_- = -\infty$, but in practice we need to determine t_- such that $\int_{t_-}^{0} \mu(t)\,\mathrm{d}t < \infty$. When $\int_{-\infty}^{t_-} \mu(t)\,\mathrm{d}t > 0$, Algorithm 1 suffers from edge effects, since clusters generated by immigrants before time t_- may contain offspring in $[0, t_+)$. The objective in this paper is to quantify these edge effects and to compare Algorithm 1 with the perfect simulation algorithm in [12].

The remainder of the paper is organised as follows. Section 2 contains some preliminaries. Section 3 contains some convergence results needed in this paper. In Section 4 various quantitative results for edge effects are introduced, and among other things we relate our results to those in Brémaud et al. [3] (which concerns approximate simulation of a stationary marked Hawkes process with unpredictable marks). Section 5 presents various examples of applications and empirical results for both Algorithm 1 and the perfect simulation algorithm in [12].

2. Preliminaries

Let F denote the c.d.f. (cumulative distribution function) for L, the length of a cluster, i.e. the time between the immigrant and the last event of the cluster. Consider the mean number of events in any offspring process Φ_i, $\bar{\nu} \equiv \mathrm{E}\nu$, where

$$\nu = \int_0^\infty \gamma(t, Z)\,\mathrm{d}t$$

is the total fertility rate of an offspring process and Z denotes a generic mark with distribution Q. We assume that

$$0 < \bar{\nu} < 1, \tag{1}$$

which among other places is needed in Proposition 1. This assumption is discussed in detail in [12]. Finally, let

$$\bar{h}(t) = \mathrm{E}\gamma(t, Z)/\bar{\nu}, \qquad t > 0, \tag{2}$$

which can be interpreted as the normalised intensity function for the first generation of offspring in a cluster started at time 0.

3. Approximations of F

It turns out that F is unknown even for very simple cases of Hawkes processes, cf. [12].

We first recall some convergence results from [12] and next establish a new useful result (Proposition 1) which provide useful approximations of F.

For $n \in \mathbb{N}_0$, let 1_n denote the c.d.f. for the length of a cluster when all events of generation $n+1, n+2, \ldots$ are removed. Clearly, 1_n is decreasing in n, $1_n \to F$ pointwise as $n \to \infty$, and

$$1_0(t) = 1, \qquad t \geq 0. \tag{3}$$

Let $\mathcal{C}$ denote the class of Borel functions $f : [0, \infty) \mapsto [0, 1]$. For $f \in \mathcal{C}$, define $\varphi(f) \in \mathcal{C}$ by

$$\varphi(f)(t) = \mathrm{E}\left[\exp\left(-\nu + \int_0^t f(t-s)\gamma(s, Z)\,\mathrm{d}s\right)\right], \qquad t \geq 0. \tag{4}$$

Then, as verified in [12] the assumption of unpredictable marks implies that

$$1_n = \varphi(1_{n-1}), \qquad n \in \mathbb{N}, \tag{5}$$

and

$$F = \varphi(F). \tag{6}$$

The recursion (5) provides a useful numerical approximation to F. As the integral in (4) with $f = 1_{n-1}$ quickly becomes difficult to evaluate analytically as n increases, we compute the integral numerically, using a quadrature rule.

Convergence with respect to the supremum norm of 1_n and certain other functions towards F is established in [12]. In this paper establishing convergence with respect to the L^1-norm becomes relevant. We let $\mathcal{C}_1$ denote the class of functions $f \in \mathcal{C}$ with $\|F - f\|_1 < \infty$, where $\|g\|_1 = \int_0^\infty |g(t)|\,\mathrm{d}t$ is the L^1-norm.

Proposition 1. *With respect to the L^1-norm, φ is a contraction on $\mathcal{C}_1$, that is, for all $f, g \in \mathcal{C}_1$ and $n \in \mathbb{N}$, we have that $f_n, g_n \in \mathcal{C}_1$ and*

$$\|\varphi(f) - \varphi(g)\|_1 \leq \bar{\nu}\|f - g\|_1. \tag{7}$$

Furthermore, F is the unique fixpoint,

$$\|F - f_n\|_1 \to 0 \qquad \text{as } n \to \infty, \tag{8}$$

and if either $f \leq \varphi(f)$ or $f \geq \varphi(f)$, then f_n increases respectively decreases towards F with a geometric rate:

$$\|F - f_n\|_1 \leq \frac{\bar{\nu}^n}{1 - \bar{\nu}}\|\varphi(f) - f\|_1. \tag{9}$$

Proof. Let $f, g \in \mathcal{C}_1$. Recall that by the mean value theorem (e.g. Theorem 5.11 in [1]), for any real numbers x and y, $\mathrm{e}^x - \mathrm{e}^y = (x - y)\mathrm{e}^{z(x,y)}$, where $z(x, y)$ is a real number between x and y. Thus by (4),

$$\|\varphi(f) - \varphi(g)\|_1 = \int_0^\infty \left| \mathrm{E}\left[\mathrm{e}^{-\nu} \mathrm{e}^{c(t,f,g)} \int_0^t (f(t-s) - g(t-s))\gamma(s, Z)\mathrm{d}s \right] \right| \mathrm{d}t \tag{10}$$

where $c(t, f, g)$ is a random variable between $\int_0^t f(t-s)\gamma(s, Z)\,\mathrm{d}s$ and $\int_0^t g(t-s)\gamma(s, Z)\,\mathrm{d}s$. Since $f, g \leq 1$, we obtain $\mathrm{e}^{c(t,f,g)} \leq \mathrm{e}^\nu$, cf. (1). Consequently,

$$\|\varphi(f) - \varphi(g)\|_1 \leq \int_0^\infty \left| \mathrm{E}\left[\int_0^t (f(t-s) - g(t-s))\gamma(s, Z)\,\mathrm{d}s \right] \right| \mathrm{d}t \tag{11}$$

$$\leq \mathrm{E}\left[\int_0^\infty \int_0^\infty |f(u) - g(u)|\,\mathrm{d}u\,\gamma(s, Z)\,\mathrm{d}s \right] = \bar{\nu}\,\|f - g\|_1 \tag{12}$$

where in the latter inequality we have used first the triangle inequality, next Fubini's theorem, and finally a simple transformation. Thereby (7) is verified. The remaining part is verified along similar lines as in the proof of Theorem 1 in [12] (with the minor observations that F is the unique fixpoint because of (8), and that we use monotone convergence when establishing (9)).□

Remark 1. The following observation motivates why we restrict attention to the class $\mathcal{C}_1$ in Proposition 1, at least when considering functions $f \in \mathcal{C}$ such that $f \leq F$: For such functions f convergence fails as

$$\|F - f\|_1 = \infty \;\Rightarrow\; \|F - f_n\|_1 = \infty, \qquad n \in \mathbb{N}. \tag{13}$$

To verify this, consider two non-negative Borel functions $f \le g$ defined on $[0,\infty)$. Then as in (10)–(12), but now observing that $c(t,f,g)$ is between 0 and ν,

$$\|\varphi(f)-\varphi(g)\|_1 \ge \mathrm{E}\left[\int_0^\infty \int_0^\infty (g(u)-f(u))\mathrm{e}^{-\nu}\gamma(s,Z)\,\mathrm{d}s\,\mathrm{d}u\right] = \|f-g\|_1 \mathrm{E}[\nu \mathrm{e}^{-\nu}].$$

By (1), $\mathrm{E}[\nu\mathrm{e}^{-\nu}] > 0$, and so letting $g = F$, we obtain (13) when $n = 1$, whereby (13) follows by induction.

As noted the sequence $f_n = 1_n$ decreases towards F pointwise. In order to obtain L^1-convergence by Proposition 1 we need $1_0 \in \mathcal{C}_1$, that is, $\mathrm{E}L = \|1-F\|_1$ is finite. A sufficient and necessary condition for this is given in Lemma 1 in [12].

To construct a sequence f_n which increases towards F in the L^1-norm, it suffices to find $f \in \mathcal{C}_1$ such that $\varphi \le \varphi(f)$. Methods for finding a c.d.f. G with $G \le \varphi(G)$ are discussed in [12] (see in particular Proposition 3 in [12]), in which case $G \le F$ (see Theorem 1 in [12]). Note that if $G \le F$ is a c.d.f. and $\|1-F\|_1 < \infty$, then G needs to have a finite mean, since $\|1-G\|_1 = \|F-G\|_1 + \|1-F\|_1$. □

4. Edge effects

Let $N(t_-,t_+)$ denote the number of missing events when using Algorithm 1. In this section we consider the mean number of missing offspring, $\mathrm{E}(t_-,t_+) \equiv \mathrm{E}N(t_-,t_+)$, and the probability of having any missing offspring, $\mathrm{P}(t_-,t_+) \equiv \mathrm{P}(N(t_-,t_+) > 0)$. Furthermore, we relate these to the total variation distance between simulations and the target distribution.

4.1. The mean number of missing offspring

Consider a cluster $C_0 = \{(s_i, Z_i)\}$ started at time $t_0 = 0$. This has conditional intensity function

$$\lambda_0(t) = \gamma(t, Z_0) + \sum_{0<s_i<t} \gamma(t-s_i, Z_i), \qquad t \ge 0, \tag{14}$$

and unpredictable marks with distribution Q. For $t > 0$, let $\lambda(t) = \mathrm{E}\lambda_0(t)$ be the intensity function of the offspring in C_0, and $\bar{\gamma}(t) = \mathrm{E}\gamma(t,Z) = \bar{\nu}\bar{h}(t)$ be the intensity function of the first generation of offspring in C_0. The following proposition expresses $\mathrm{E}(t_-,t_+)$ and $\lambda(t)$ in terms of μ and $\bar{\gamma}$.

Proposition 2. *We have that*

$$\lambda(t) = \sum_{n=1}^{\infty} \bar{\gamma}^{*n}(t) = \sum_{n=1}^{\infty} \bar{\nu}^n \bar{h}^{*n}(t), \qquad t \geq 0, \tag{15}$$

where $*n$ *denotes convolution* n *times, and*

$$\mathrm{E}(t_-, t_+) = \int_{-\infty}^{t_-} \left(\int_{-t}^{t_+ - t} \lambda(s)\,\mathrm{d}s \right) \mu(t)\,\mathrm{d}t. \tag{16}$$

Proof. We claim that $\rho_n = \bar{\gamma}^{*n}$ is the intensity function of $\mathcal{G}_n$, the n-th generation of offspring in the cluster C_0: This is clearly true for $n = 1$, and so by induction

$$\begin{aligned}\rho_{n+1}(t) &= \mathrm{E} \sum_{s_i \in \mathcal{G}_n} \gamma(t - s_i, Z_i) = \mathrm{E} \sum_{s_i \in \mathcal{G}_n} \mathrm{E}[\gamma(t - s_i, Z_i)|s_i] = \mathrm{E} \sum_{s_i \in \mathcal{G}_n} \bar{\gamma}(t - s_i) \\ &= \int_0^t \rho_n(s) \bar{\gamma}(t - s)\,\mathrm{d}s = \bar{\gamma}^{*(n+1)}\end{aligned}$$

where we have used Campbell's theorem in the second last equality and the induction hypothesis in the last equality. Thereby (15) follows. Finally, if I denotes the Poisson process of immigrants,

$$\begin{aligned}\mathrm{E}(t_-, t_+) &= \mathrm{E} \sum_{t_i \in I} \sum_{s \in C_i} \mathbf{1}[t_i < t_-,\, 0 \leq s < t_+] = \mathrm{E} \sum_{t_i \in I:\, t_i < t_-} \mathrm{E}\left[\sum_{s \in C_i} \mathbf{1}[0 \leq s < t_+] \middle| t_i \right] \\ &= \mathrm{E} \sum_{t_i \in I:\, t_i < t_-} \int_{-t_i}^{t_+ - t_i} \lambda(u)\,\mathrm{d}u\end{aligned}$$

which reduces to (16) by Campbell's theorem. □

Remark 2. It follows immediately that

$$\rho = \mu + \mu * \lambda \tag{17}$$

is the intensity function of all events. When quantifying edge effects it is natural to consider $\mathrm{E}(t_-, t_+)/\mathrm{E}(t_+)$, where the expected number of events on $[0, t_+]$

$$\mathrm{E}(t_+) = \int_0^{t_+} \rho(t)\,\mathrm{d}t$$

is the expected number of events on $[0, t_+]$. □

4.2. The probability of having any missing offspring

Obviously, $\mathrm{P}(t_-, t_+)$ is an increasing function of $t_+ \in (0, \infty]$. Proposition 3 gives an expression and upper and lower bounds for $\mathrm{P}(t_-, \infty)$.

Proposition 3. *We have that*

$$\mathrm{P}(t_-,\infty) = 1 - \exp\left(-\int_{-\infty}^{t_-}(1-F(-t))\mu(t)\,\mathrm{d}t\right). \tag{18}$$

Further, for any $f \in \mathcal{C}_1$ such that $f \le \varphi(f)$, we have an upper bound,

$$\mathrm{P}(t_-,\infty) \le 1 - \exp\left(-\int_{-\infty}^{t_-}(1-f_n(-t))\mu(t)\,\mathrm{d}t\right), \tag{19}$$

which is a decreasing function of n, and a lower bound

$$\mathrm{P}(t_-,\infty) \ge 1 - \exp\left(-\int_{-\infty}^{t_-}(1-1_n(-t))\mu(t)\,\mathrm{d}t\right), \tag{20}$$

which is a increasing function of n.

Proof. Let I_{t_-} be the point process of immigrants $t_i < t_-$ with $\{(t_j, Z_j) \in C_i : t_j \ge 0\} \neq \emptyset$. Then I_{t_-} is a Poisson process with intensity function $\lambda_{t_-}(t) = (1-F(-t))\mu(t)$ on $(-\infty, t_-)$, since we can view I_{t_-} as an independent thinning of the immigrant process on $(-\infty, t_-)$, with retention probabilities $p(t) = 1 - F(-t)$, $t < t_-$. Hence, since $\mathrm{P}(t_-,\infty)$ equals the probability that $I_{t_-} \neq \emptyset$, we obtain (18). Thereby (19) and (20) follows from (18) and Proposition 1. □

Remark 3. Proposition 1 ensures that the upper bound in (19) and the lower bound in (20) converge monotoneously to $\mathrm{P}(t_-,\infty)$ provided e.g. that μ is bounded and $\mathrm{E}L < \infty$, cf. Remark 1. □

4.3. The total variation distance between simulations and the target distribution

Recently, Brémaud et al. [3] derived related results to Propositions 2 and 3 when $\mu(t)$ is constant and $t_+ = \infty$. Proposition 4 below generalises their results to the situation in the present paper where $\mu(t)$ is not necessarily constant and t_+ may be finite. Moreover, our proof is much simpler.

We let $\tilde{X}$ be another marked Hawkes process obtained from X by removing all clusters C_i with immigrants $t_i < t_-$. Furthermore, we let Y and $\tilde{Y}$ denote the restriction of X and $\tilde{X}$ to the marked events on $[0, t_+)$, and denote their distributions by $\pi(t_-, t_+)$ and $\tilde{\pi}(t_-, t_+)$. Thus the output of Algorithm 1 follows $\tilde{\pi}(t_-, t_+)$, which approximates the target distribution $\pi(t_-, t_+)$.

Proposition 4. *Let* $\|\cdot\|_{\mathrm{TV}}$ *denote the total variation distance, then*

$$\|\pi(t_-,t_+) - \tilde{\pi}(t_-,t_+)\|_{\mathrm{TV}} \leq \mathrm{P}(t_-,t_+) \leq \mathrm{E}(t_-,t_+). \tag{21}$$

Proof. By the construction of $\tilde{Y}$, we have that $\tilde{Y} \subseteq Y$. The first inequality then follows immediately from the coupling inequality (see e.g. [11]), while the second inequality is trivially satisfied. □

Remark 4. In contrast to the first upper bound in (21) the second upper bound does not depend on knowing F or any approximation of F, cf. Propositions 2 and 3. □

4.4. Extensions and open problems

It would be of practical importance to extend our results to the case of predictable marks. Proposition 4 is still true if the conditional intensity function for X is larger than or equal to the conditional intensity function for $\tilde{X}$; this follows by a thinning argument, cf. [5]. However, this observation seems of little use, since the assumption of unpredictable marks is essential in the proofs of (15) in Proposition 2 and (19)–(20) in Proposition 3. Moreover, though (18) in Proposition 3 remains true, it is expected to be of limited use, since F is expected to be of a more complicated form in the case of predictable marks.

The following observations may also be of practical relevance.

Algorithm 1 applies for a non-Poisson immigrant process, e.g. a Markov or Cox process provided it is feasible to simulate the immigrants on $[t_-,t_+)$. Furthermore, Proposition 2 remains true for any immigrant process with intensity function μ. Finally, Proposition 3 partly relies on the immigrants being a Poisson process: for instance, if now μ is a random intensity function and the immigrant process is a Cox process driven by μ, then (18)–(20) should be modified by taking the mean of the expressions on the right hand sides.

5. Examples and comparison with perfect simulation

Illustrative examples of specific unmarked and marked Hawkes processes (with plots showing perfect simulations) are given in [12]. In this section we consider the same examples of models and demonstrate the use and limitations of our results in Section 4.

We also demonstrate the practical differences between Algorithm 1 and the perfect simulation algorithm in [12].

5.1. An unmarked Hawkes process model

The events and marks of X are independent if and only if $\gamma(t,z) = \gamma(t)$ does not depend on the mark z (for almost all z) in which case the events form an unmarked Hawkes process. In this section we consider an unmarked Hawkes process with exponentially decaying fertility rate given by $\gamma(t) = \alpha\beta \mathrm{e}^{-\beta t}$, where $0 < \alpha < 1$ and $\beta > 0$ are parameters.

Note that $1/\beta$ is a scale parameter for the distribution of L, $\bar{\nu} = \nu\alpha$, and $\bar{h} = \beta \mathrm{e}^{-\beta t}$. Hence $\bar{h}^{*n}$ is the density for a gamma distribution with shape parameter n and inverse scale parameter β. Using (15), we obtain $\lambda(t) = \alpha\beta \mathrm{e}^{(\alpha-1)\beta t}$. Inserting this into (16), assuming that $t_- > -\infty$ and $\mu(t) = \delta \mathrm{e}^{\kappa t}$ where $\delta > 0$ and $\kappa > (\alpha-1)\beta$ are parameters, we obtain that

$$\mathrm{E}(t_-, t_+) = \frac{\alpha\delta}{(1-\alpha)((1-\alpha)\beta + \kappa)}(1 - \mathrm{e}^{(\alpha-1)\beta t_+})\mathrm{e}^{((1-\alpha)\beta+\kappa)t_-}.$$

Here the restriction on κ is equivalent to that ρ is finite, in which case $\rho(t) = \delta \mathrm{e}^{\kappa t}(\kappa + \beta)/(\kappa + (1-\alpha)\beta)$, cf. (17).

Figure 1 shows $\mathrm{E}(t_-, t_+)/\mathrm{E}(t_+)$ as a function of $-t_- \geq 0$ in the case $\alpha = 0.9$, $\delta = \beta = 1$, $t_+ = 10$, and for different values of κ. As expected numerically smaller values of t_- are needed as κ increases. For $\kappa \geq 0$, effectively perfect simulation are produced when $t_- = -50$.

Let $f(t) = 1 - \mathrm{e}^{-\theta t}$ be the c.d.f. for an exponential distribution with parameter $\theta = \beta(1-\alpha)$. As verified in [12], $f \leq \varphi(f)$, and so the bounds of $\mathrm{P}(t_-, \infty)$ in Proposition 3 hold. Figure 2 shows these bounds when $\alpha = 0.9$, $\beta = \delta = 1$ and $\kappa = 0$ (i.e. $\mu = 1$), and $n = 0, 7, \ldots, 70$. The convergence of the bounds to $\mathrm{P}(t_-, t_+)$ is clearly visible, and for $n = 70$ both bounds are practically equal. Also the plot reveals that for the present choice of parameters, the probability for having one or more missing events is effectively 0 for $t_- = -50$.

We can determine $N(t_-, t_+)$, or at least its distribution, from the perfect simulation algorithm in [12]. Figure 3 shows one minus the corresponding empirical distribution function based on 10000 perfect simulations when $\alpha = 0.9$, $\beta = \delta = 1$, $\kappa = 0$, $t_+ = 10$,

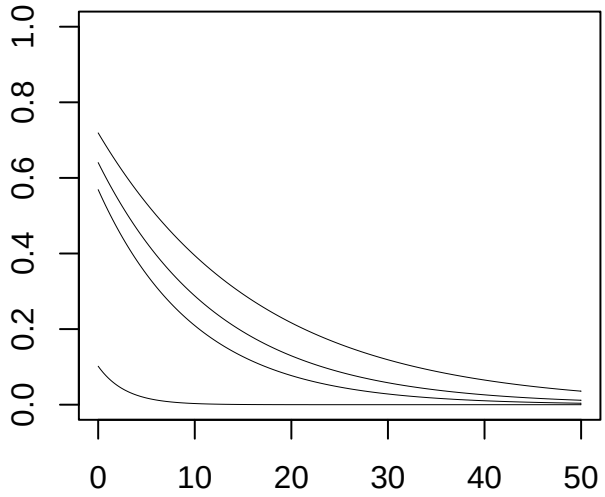

FIGURE 1: Plot of $\mathrm{E}(t_-, t_+)/\mathrm{E}(t_+)$ versus $-t_-$ for the unmarked case with parameters $\alpha = 0.9$, $\delta = \beta = 1$, $t_+ = 10$, and $\kappa = -0.04, -0.02, 0, 0.25$ (top to bottom).

and $t_- = 0, -10$, or -50. In each of the three cases, since $\mathrm{E}(t_+) = 100$, the number of missing events in the case $t_- = 0$ is substantially reduced, but still too large, when $t_- = -10$, while edge effects are practically non-existent for $t_- = -50$.

Comparing Figures 1–3 when for example $\alpha = 0.9$, $\beta = \delta = 1$, $\kappa = 0$, $t_+ = 10$, and $t_- = -50$, Algorithm 1 and the perfect simulation algorithm from [12] are effectively producing identical results. Algorithm 1 uses roughly one-thousandth of a second for each simulation in our implementation, while the perfect simulation algorithm uses one-tenth of a second.

5.2. A marked Hawkes process model with birth and death transitions

Consider a marked Hawkes process with

$$\gamma(t, z) = \alpha \mathbf{1}[t \leq z]/\mathrm{E}Z,$$

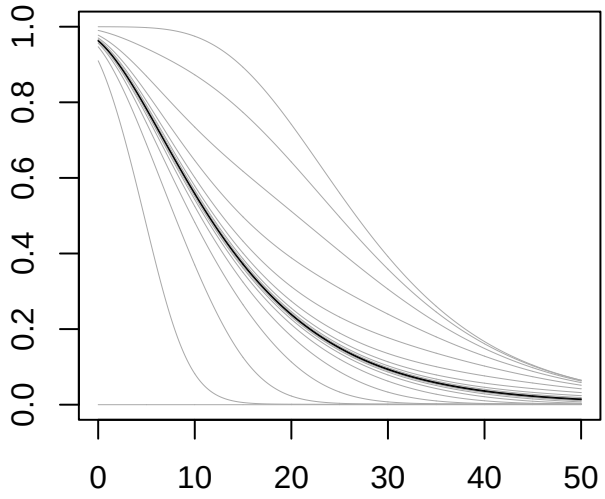

FIGURE 2: Upper and lower bounds (19) and (20) of $P(t_-, t_+)$ versus $-t_-$ in the unmarked case with $\alpha = 0.9$, $\mu = \beta = 1$, $t_+ = \infty$, and $n = 0, 7, \ldots, 70$. The bounds using $n = 70$ are shown in black to illustrate the approximate form of $\mathrm{P}(t_-, t_+)$, whereas the rest are shown in gray.

where $0 < \alpha < 1$ is a parameter, Z is a positive random variable with distribution Q, and $\mathbf{1}[\cdot]$ denotes the indicator function. Then X can be viewed as a birth and death process, with birth at time t_i and survival time Z_i of the i'th individual.

The special case where $\mu(t) = \mu$ is constant and Z is exponentially distributed with mean $1/\beta$ is considered at page 136 in [3]. Since $\bar{h}(t) = \beta \mathrm{e}^{-\beta t}$ is the same function as in Section 5.1, $\mathrm{E}(t_-, t_+)$ is also the same as in Section 5.1. Further, a plot of $P(t_-, t_+)$ (omitted here) is similar to Figure 2 (when using the same parameters). Also a plot of the empirical distribution function of $N(t_-, t_+)$ (omitted here) is similar to Figure 3.

When for example $\alpha = 0.9$, $\beta = \mu = 1$, $t_+ = 10$, and $t_- = -50$, Algorithm 1 uses roughly one-five hundredth of a second for each simulation, and the perfect simulation algorithm uses just under three seconds. As in the unmarked case both algorithms are

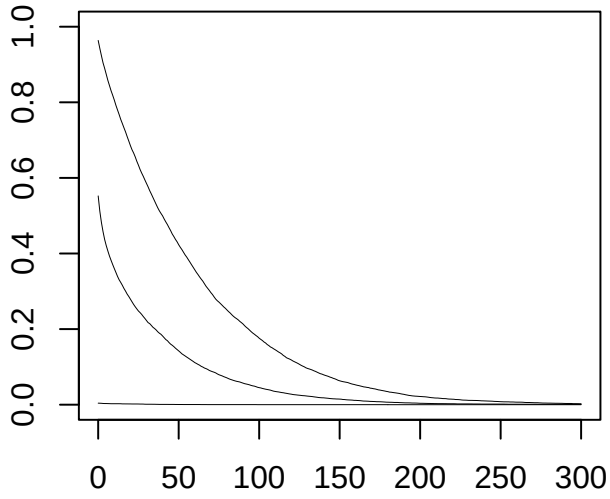

FIGURE 3: One minus the empirical distribution function for $N(t_-, t_+)$ in the unmarked case with $\alpha = 0.9$, $\beta = \delta = 1$, $\kappa = 0$, $t_+ = 10$, and $t_- = 0, -10, -50$ (top to bottom).

feasible, but the difference is much more clear in the present case.

5.3. A heavy-tailed distribution for L

We conclude by observing that heavy-tailed cases of the distribution of L are problematic. For instance, suppose that

$$\gamma(t, z) = \alpha z \mathrm{e}^{-tz},$$

where $\alpha \in (0, 1)$ is a parameter, and let Q be the exponential distribution with mean $1/\beta$. As argued in [12], $\bar{h}(t) = \beta/(t + \beta)^2$ is a Pareto density and L has a heavy-tailed distribution with infinite moments and infinite Laplace transform. As $\mathrm{E}L = \infty$, Proposition 1 and hence Proposition 3 seem of rather limited use, cf. Remark 1. Proposition 2 is also not applicable, since λ is not known on closed form, cf. Example 7 in [12]. It is a challenging open problem to handle such heavy-tailed cases.

Acknowledgements

The research of Jesper Møller was supported by the Danish Natural Science Research Council and the Network in Mathematical Physics and Stochastics (MaPhySto), funded by grants from the Danish National Research Foundation.

References

[1] Apostol, T. M. (1974). *Mathematical Analysis*. Addison-Wesley, Reading.

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

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

[4] Chornoboy, E. S., Schramm, L. P. and Karr, A. F. (2002). Maximum likelihood identification of neural point process systems. *Adv. Appl. Prob.* **34,** 267–280.

[5] Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods* 2nd ed. Springer, New York.

[6] Hawkes, A. G. (1971). Point spectra of some mutually exciting point processes. *J. Roy. Statist. Soc. Ser. B* **33,** 438–443.

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

[8] Hawkes, A. G. (1972). Spectra of some mutually exciting point processes with associated variables. In *Stochastic Point Processes*. ed. P. A. W. Lewis. Wiley, New York, pp. 261–271.

[9] Hawkes, A. G. and Adamopoulos, L. (1973). Cluster models for earthquakes – regional comparisons. *Bull. Int. Statist. Inst.* **45,** 454–461.

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

[11] Lindvall, T. (1992). *Lectures on the Coupling Method*. Wiley, New York.

[12] Møller, J. and Rasmussen, J. G. (2004). Perfect simulation of Hawkes processes. Research report R-2004-18, Department of Mathematical Sciences, Aalborg University. Available at http://www.math.aau.dk/ jm.

[13] Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Amer. Statist. Assoc.* **83,** 9–27.

[14] Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Ann. Inst. Statist. Math.* **50,** 379–402.

[15] Vere-Jones, D. and Ozaki, T. (1982). Some examples of statistical inference applied to earthquake data. *Ann. Inst. Statist. Math.* **34,** 189–207.