A note on a perfect simulation algorithm for marked Hawkes processes*

Jesper Møller and Jakob Gulddahl Rasmussen
Aalborg University, Department of Mathematical Sciences,
F. Bajers Vej 7G, DK-9220 Aalborg, Denmark.

Abstract

The usual straightforward simulation algorithm for (marked or unmarked) Hawkes processes suffers from edge effect. In this note we describe a perfect simulation algorithm which is partly derived as in Brix and Kendall [1] and partly using upper and lower processes as in the Propp-Wilson algorithm [7], or rather as in the dominated CFTP algorithm by Kendall and Møller [3]. Various monotonicity properties and approximations of the cumulative distribution function for the length of a so-called cluster in a marked Hawkes process play an important role.

Keywords: Marked Hawkes process; perfect simulation; thinning; upper and lower processes.

1 Introduction

Marked Hawkes processes play a fundamental role for point process theory and its applications, and they have have important applications in seismology and neurophysiology, see [2] and the references therein. This paper concerns a perfect simulation algorithm for marked Hawkes processes. For proofs and further details we refer to [5], which also considers the usual straightforward algorithm which suffers from edge effects.

For mathematical convenience we restrict attention to marked Hawkes processes with unpredictable marks. Such a process $X = \{(t_i, Z_i)\} \subset \mathbb{R} \times M$ is defined as follows, where M denotes an arbitrary mark space (equipped with a σ-field).

- Each event (or time) t_i is of one of two types: an immigrant or an offspring. The immigrants follow a Poisson point process, with an intensity function $\mu(t)$ on $\mathbb{R}$, which is locally integrable.

* Work partly based on [5]. 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.

- The process has unpredictable marks in the sense that conditional on any event t_i and the previous history $\{(t_k, Z_k) : t_k < t_i\}$, Z_i follows Q, where Q denotes a probability distribution on M which is independently of t_i and the previous history.

- If we condition on Z_i, then independently of the previous history, (t_i, Z_i) generates a Poisson point process $\Phi(t_i)$ of offspring on (t_i, ∞), with intensity function $\gamma_i(t) = \gamma(t - t_i, Z_i),\ t > t_i$, where γ is a non-negative measurable function on $(0, \infty) \times M$ (for convenience we suppress in the notation that $\Phi(t_i)$ may depend on Z_i).

Outline: Section 2 considers the distribution function F for the length of a so-called cluster in a marked Hawkes process. Since F is not known on closed form, certain approximations of F and monotonicity properties are discussed in Section 3. This is then used for the perfect simulation algorithm described in Section 4.

2 The distribution of the length of a cluster

A marked Hawkes process is a Poisson cluster process, with cluster centres given by the immigrant process. Here a cluster associated to an immigrant t_i consists of (t_i, Z_i), all offspring (t_j, Z_j) generated by (t_i, Z_i), all offspring (t_k, Z_k) generated by (t_j, Z_j), and so on. The clusters are i.i.d. relative to their associated immigrants. In this section we consider the c.d.f. F for the length of a cluster.

Consider the mean number of points in any offspring process,

$$\bar{\nu} \equiv \mathrm{E}\nu \quad \text{where} \quad \nu = \int_0^\infty \gamma(t, Z)\,\mathrm{d}t$$

and Z denotes a generic mark with distribution Q. Henceforth we assume that

$$0 < \bar{\nu} < 1.$$

The condition $\bar{\nu} < 1$ appears commonly in the literature on Hawkes processes, and it is equivalent to that the mean number of events in a cluster is finite. The other condition $\bar{\nu} > 0$ excludes the trivial case where there are almost surely no offspring. It is readily seen to be equivalent to

$$F < 1.$$

For $n \in \mathbb{N}_0$, let F_n denote the c.d.f. for the length of a cluster when all events of generation $n+1, n+2, \ldots$ are removed. Then F_n is non-increasing in n, and $F_n \to F$

pointwise as $n \to \infty$. Obviously,

$$F_0(t) = 1 \text{ for } t \geq 0.$$

Furthermore, for any Borel function $f : [0, \infty) \mapsto \mathbb{R}$, define another Borel function $\varphi(f)$ by

$$\varphi(f)(t) = \mathrm{E}\left[\exp\left(-\nu + \int_0^t f(t-s)\gamma(s, Z)\,\mathrm{d}s\right)\right], \qquad t \geq 0. \tag{1}$$

Proposition 1 We have that

$$F_n = \varphi(F_{n-1}), \qquad n \in \mathbb{N}, \tag{2}$$

and

$$F = \varphi(F). \tag{3}$$

We have been unsuccessful in using (3) to obtain a closed form expression for F even for simple choices of γ. However, the recursion (2) can be very useful to obtain a numerical approximation to F, though the integral in (1) with $f = F_{n-1}$ quickly becomes difficult to evaluate analytically as n increases. Therefore we compute the integral in (1) numerically, using a quadrature rule.

3 Convergence and monotonicity results

Many other approximations of F than F_n exist, and the rate of convergence may be geometric with respect to different norms, cf. [5]. Particularly, we have the following convergence and monotonicity results, where we define $\|f\|_1 = \int_0^\infty |f(t)|\,\mathrm{d}t$ and $\|f\|_\infty = \sup|f(t)|$ for any Borel function $f : [0, \infty) \mapsto \mathbb{R}$.

Proposition 2 Let $f : [0, \infty) \mapsto [0, 1]$ be a Borel function. If $f \leq \varphi(f)$ (or $f \geq \varphi(f)$), then f_n converges monotonously to F from below (above) with respect to the supremum norm, and also with respect to the L_1-norm provided $\|F - f\|_1 < \infty$. The rate of convergence is $\bar{\nu}^n \|\varphi(f) - f\|_m / (1 - \bar{\nu})$ with $m = 1$ or $m = \infty$.

Proposition 2 applies for $f = F_0$, since $F_0 \geq \varphi(F_0)$. The next proposition shows that in many situations $G \leq \varphi(G)$ when G is an exponential c.d.f. with a sufficiently large mean. Denote the Laplace transform of $\bar{h}(t) = \mathrm{E}[\gamma(t, Z)]/\bar{\nu}$ by

$$\mathcal{L}(\theta) = \int_0^\infty \mathrm{e}^{\theta t}\bar{h}(t)\,\mathrm{d}t, \qquad \theta \in \mathbb{R}.$$

Proposition 3 If $G(t) = 1 - \mathrm{e}^{-\theta t}$ for $t \geq 0$, where $\theta > 0$ and $\mathcal{L}(\theta) \leq 1/\bar{\nu}$, then $G \leq \varphi(G)$.

Example 1 Consider an unmarked Hawkes process (i.e. $\gamma(t, Z) = \gamma(t)$ does not depend on Z) with

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

where $0 < \alpha < 1$ and $\beta > 0$ are parameters. Then $\bar{\nu} = \nu = \alpha$, $\mathcal{L}(\theta) = \beta/(\beta - \theta)$ if $\theta < \beta$, and $\mathcal{L}(\theta) = \infty$ otherwise. Hence $\mathcal{L}(\theta) \leq 1/\bar{\nu}$ is equivalent to $\theta \leq \beta(1 - \alpha)$, and so Proposition 3 applies.

Example 2 Consider a marked Hawkes process with

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

where $\alpha > 0$ is a parameter, Z is exponentially distributed with inverse mean $\beta > 0$, 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. Further, $\nu = \alpha\beta Z$ and $\bar{\nu} = \alpha$. Furthermore, $\mathcal{L}(\theta) = \mathcal{L}_Z(\theta) = \beta/(\beta - \theta)$ is of the same form as in Example 1.

4 Perfect simulation algorithm

Assuming for the moment that F is known, the following algorithm for perfect simulation of the marked Hawkes process is similar to the simulation of Poisson cluster processes without edge effects given in Brix and Kendall [1]; see also [4, 6].

Assume that we want to make a perfect simulation of the marked Hawkes process restricted to events on an interval $[0, t_+]$ where $0 < t_+ \leq \infty$. We let I_1 be the immigrants on $[0, t_+)$, and let I_2 be the immigrants $t_i < 0$ such that $\{(t_j, Z_j) \in C_i : t_j \in [0, \infty)\} \neq \emptyset$. The algorithm works as follows.

(a) Simulate I_1 as a Poisson process with intensity function $\lambda_1(t) = \mu(t)$ on $[0, t_+)$.

(b) For each $t_i \in I_1$, simulate first Z_i from Q and next those $(t_j, Z_j) \in C_i$ with $t_i < t_j \leq t_+$.

(c) Simulate I_2 as a Poisson process with intensity function $\lambda_2(t) = \mu(t)(1 - F(-t))$ on $(-\infty, 0)$.

(d) For each $t_i \in I_2$, simulate Z_i from Q and simulate $\{(t_j, Z_j) \in C_i : t_j \in [0, t_+)\}$ conditional on that $\{(t_j, Z_j) \in C_i : t_j \in [0, \infty)\} \neq \emptyset$.

(e) The output is all marked points from (a)–(d).

Proposition 4 The output follows the distribution of the marked Hawkes process.

Statements about ET, where $-T$ denotes the first point in I_2 are also given in [5]. Steps (a) and (b) are straightforward. Since F is unknown in practice, the crucial point is how to simulate I_2 in step (c)—at the end of this section we solve that problem. In step (d), it is not obvious how to construct an elegant approach ensuring that at least one point will fall after 0. Instead we use a simple rejection sampler: we repeatedly simulate the cluster of an immigrant until the cluster contains an offspring after time 0.

In practice we must require that I_1 and I_2 are (almost surely) finite or equivalently that $\int_0^{t_+} \mu(t)\,\mathrm{d}t < \infty$ and

$$\int_{-\infty}^{0} (1 - F(-t))\mu(t)\,\mathrm{d}t < \infty. \tag{4}$$

Clearly, since μ is assumed to be locally integrable, he first condition is satisfied if $t_+ < \infty$. In the stationary case where $\mu(t)$ is constant, (4) reduces to the condition that F has finite mean. We have the following results, where L is a random variable with c.d.f. F and R is a random variable with density $h(t) = \mathrm{E}[\gamma(t, Z)/\nu],\ t \geq 0$.

Proposition 5 We have that

$$(1 - \mathrm{E}\mathrm{e}^{-\nu})\mathrm{E}R \leq \mathrm{E}L \leq \frac{\bar{\nu}}{1-\bar{\nu}}\mathrm{E}R,$$

so E$L < \infty$ if and only if E$R < \infty$.

Our simulation algorithm for I_2 is based on the use of coupled upper and lower processes in a way somewhat similar as in the Propp-Wilson algorithm [7] or rather as in the dominated CFTP algorithm in Kendall and Møller [3]. Suppose that $f : [0, \infty) \mapsto [0, 1]$ is a Borel function on a closed/known form such that $f \leq \varphi(f)$ and (4) is satisfied when we replace F by f. For $n \in \mathbb{N}_0$, let U_n and L_n denote Poisson processes on $(-\infty, 0)$ with intensity functions $\lambda_n^u(t) = \mu(t)(1 - f_n(-t))$ and $\lambda_n^l(t) = \mu(t)(1 - F_n(-t))$, respectively. By Corollary 3, λ_n^u is non-increasing and λ_n^l is non-decreasing in n, and they both converge to λ_2 (geometrically fast with respect to the supremum norm). Consequently, we can use independent thinning to obtain the following sandwiching/funnelling property, cf. [3]:

$$\emptyset = L_0 \subseteq L_1 \subseteq L_2 \subseteq \cdots \subseteq I_2 \subseteq \cdots \subseteq U_2 \subseteq U_1 \subseteq U_0.$$

The details are given by the following steps.

(i) Generate a realisation $\{(t_1, Z_1), \ldots, (t_k, Z_k)\}$ of U_0, where $t_1 < \ldots < t_k$.

(ii) If $U_0 = \emptyset$, then return $I_2 = \emptyset$ and stop, else generate independent uniform numbers $R_1, \ldots, R_k$ on $[0, 1]$ (independently of U_0), and set $n = 1$.

(iii) For $j = 1, \dots, k$, assign (t_j, Z_j) to L_n respective U_n if $R_j\lambda_0^u(t_j) \leq \lambda_n^l(t_j)$ respective $R_j\lambda_0^u(t_j) \leq \lambda_n^u(t_j)$.

(vi) If $U_n = L_n$, then return $I_2 = L_n$ and stop, else increase n by 1 and go to step (iii).

Proposition 6 This works correctly and terminates almost surely within finite time.

As noted earlier, we compute F_n and f_n numerically, using a quadrature point. After step (i) in Algorithm 3, we let the last quadrature point be given by $-t_1$ (since we do not need to calculate $F_n(t)$ and $f_n(t)$ for $t > -t_1$). Since we have to calculate F_n and f_n recursively for all $n = 0, 1, 2, \dots$ until Algorithm 3 terminates, there is no advantage in using a doubling scheme for n like in the Propp-Wilson algorithm.

References

[1] A. Brix and W.S. Kendall. Simulation of cluster point processes without edge effects. *Advances in Applied Probability*, 34: 267-280, 2002.

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

[3] W. S. Kendall and J. Møller. Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Advances in Applied Probability*, 32: 844–865, 2000.

[4] J. Møller. Shot noise Cox processes. *Advances in Applied Probability*, 35:614–640, 2003.

[5] J. Møller and J. G. Rasmussen. Perfect and approximate simulation of Hawkes processes. In preparation.

[6] J. Møller and R. P. Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes*. Chapman & Hall, Boca Raton, Florida, 2004.

[7] J. Propp and D. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9:223–252, 1996.