PERFECT SIMULATION OF HAWKES PROCESSES

JESPER MØLLER,* Aalborg University
JAKOB G. RASMUSSEN,** Aalborg University

Abstract

This article concerns a perfect simulation algorithm for unmarked and marked Hawkes processes. The usual straightforward simulation algorithm suffers from edge effects, whereas our perfect simulation algorithm does not. By viewing Hawkes processes as Poisson cluster processes and using their branching and conditional independence structure, useful approximations of the distribution function for the length of a cluster are derived. This is used to construct upper and lower processes for the perfect simulation algorithm. Examples of applications and empirical results are presented.

Keywords: Approximate simulation; dominated coupling from the past; edge effects; exact simulation; marked Hawkes process; marked point process; perfect simulation; point process; Poisson cluster process; thinning; upper and lower processes

AMS 2000 Subject Classification: Primary 60G55
Secondary 68U20

1. Introduction

Unmarked and marked Hawkes processes [10, 11, 12, 14] play a fundamental role for point process theory and its applications, cf., for example, [8], and they have important applications in seismology [13, 22, 23, 27] and neurophysiology [3, 7]. 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 $X = \{(t_i, Z_i)\}$ 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 \( X \) are given by certain events called *immigrants*, while the other 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* (see also Figure 1): 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 \( \Phi_j \) of offspring of generation \( n + 1 \) with intensity function \( \gamma_j(t) = \gamma(t - t_j, Z_j), t > t_j \). Here \( \gamma \) is a non-negative measurable function defined on \((0, \infty)\). We refer to \( \Phi_j \) as an *offspring process*, and to \( \gamma_j \) and \( \gamma \) 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 [8] 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.

The independence assumptions in (c) and (d) imply that we have i.i.d. marks. In the special case where \( \gamma(t, z) = \gamma(t) \) does not depend on its second argument (or if just \( P(\gamma(t, Z) = \gamma(t) \) for Lebesgue almost all \( t > 0 \) = 1 where \( Z \) denotes a generic mark), the events follow an unmarked Hawkes process. Apart from that case, the events and the marks are dependent processes. Another way of defining the process is as follows (see e.g. [8]): The marks are i.i.d. and the conditional intensity function \( \lambda(t) \) at time \( t \in \mathbb{R} \) for the events given the previous history \( \{(t_k, Z_k) : t_k < t\} \) is given by

\[
\lambda(t) = \mu(t) + \sum_{t_i < t} \gamma(t - t_i, Z_i).
\] (1)
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. [8]. However, Definition 1 immediately leads to the following simulation algorithm, where $t_\infty \in [-\infty, 0]$ 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 $t_i < t_j < t_+$.

In general Algorithm 1 suffers from edge effects, since clusters generated by immigrants before time $t_-$ may contain offspring in $[0, t_+)$. Algorithm 1 is investigated thoroughly in [18] where various measures for edge effects are introduced.

The objective in this paper is to construct a perfect (or exact) simulation algorithm. Perfect simulation has been a hot research topic since the seminal Propp-Wilson algorithm [24] appeared, but the application areas have so far been rather limited.
and many perfect simulation algorithms proposed in the literature are too slow for real applications. As demonstrated in [18] our perfect simulation algorithm can be practical and efficient. Moreover, apart from the advantage of not suffering from edge effects, our perfect simulation algorithm becomes useful to quantify edge effects caused by Algorithm 1, see [18].

The perfect simulation algorithm is derived using similar principles as in [6], but our algorithm requires the knowledge of the cumulative distribution function (c.d.f.) $F$ for the length of a cluster, and $F$ is unknown even for the simplest examples of Hawkes processes. By establishing certain monotonicity and convergence results we are able to approximate $F$ to any required precision, and more importantly to construct a dominating process and upper and lower processes in a similar fashion as in the dominated coupling from the past algorithm in [16].

The paper is organised as follows. Section 2 contains some preliminaries, including illuminating examples of Hawkes process models used throughout the paper to illustrate our results. Section 3 describes the perfect simulation algorithm, assuming that $F$ is known, while the above-mentioned convergence and monotonicity results are established in Section 4. Section 5 completes the perfect simulation algorithm, using dominated coupling from the past. Finally, Section 6 contains a discussion of our algorithms and results and how to extend these to more general settings.

2. Preliminaries and examples

2.1. The branching structure and self-similarity property of clusters

By Definition 1 we can view the marked Hawkes process $X = \{(t_i, Z_i)\}$ as a Poisson cluster process, with cluster centres given by the immigrants, and where the clusters given the immigrants are independent. This section describes a self-similarity property due to the specific branching structure within a cluster.

For events $t_i < t_j$, we say that $(t_j, Z_j)$ has ancestor $t_i$ of order $n \geq 1$ if there is a sequence $s_1, \ldots, s_n$ of offspring such that $s_n = t_j$ and $s_k \in \Phi(s_{k-1})$ for $k = 1, \ldots, n$, where $s_0 = t_i$. We say then that $t_j$ is an offspring of $n$-th generation with respect to $t_i$; for convenience we say that $t_i$ is of zeroth generation with respect to itself. Now, define the total offspring process $C_i$ as all $(t_j, Z_j)$ such that $t_j$ is an event of generation
n ∈ \mathbb{N}_0 \text{ with respect to } t_i \text{ (note that } (t_i, Z_i) \in C_i \text{). The clusters are defined as those } C_i \text{ where } t_i \text{ is an immigrant, cf. Definition 1.}

The total offspring processes have the same branching structure relative to their generating events. More precisely, since \( \gamma_i(t) = \gamma(t - t_i, Z_i) \) for any event \( t_i \), we see by Definition 1 that conditional on events \( t_i < t_j \), the translated total offspring processes \( C_i - t_i \equiv \{(t_l - t_i, Z_l) : (t_l, Z_l) \in C_i\} \) and \( C_j - t_j \equiv \{(t_l - t_j, Z_l) : (t_l, Z_l) \in C_j\} \) are identically distributed.

In particular, conditional on the immigrants, the clusters relative to their cluster centres (the immigrants) are i.i.d. with distribution \( P \), say. Furthermore, conditional on the \( n \)-th generation events \( G_n \), say, in a cluster, the translated total offspring processes \( C_i - t_i \) with \( t_i \in G_n \) are i.i.d. with distribution \( P \). We refer to this last property as the i.i.d. self-similarity property of offspring processes or for short the self-similarity property. Note that the assumption of unpredictable marks is essential for these properties to hold.

2.2. A basic assumption and some terminology and notation

Let \( F \) denote the c.d.f. for the length \( L \) 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 \( \Phi(t_i) \), \( \bar{\nu} \equiv E\nu \), where

\[
\nu = \int_0^\infty \gamma(t, Z) \, dt
\]

is the total fertility rate of an offspring process and \( Z \) denotes a generic mark with distribution \( Q \). Henceforth we assume that

\[
0 < \bar{\nu} < 1.
\] (2)

The condition \( \bar{\nu} < 1 \) appears commonly in the literature on Hawkes processes, see e.g. [5], [8], and [14], and is essential for our convergence results in Section 4.2. It implies that

\[
F(0) = E\nu^{-\nu} > 0
\] (3)

where \( F(0) \) is the probability that a cluster has no offspring. It is equivalent to assuming that \( ES < \infty \), where \( S \) denotes the number of events in a cluster: By induction on
$n \in \mathbb{N}_0$, because of the branching and conditional independence structure of a cluster, $ar{\nu}^n$ is the mean number of generation $n$ events in a cluster, so

$$E\bar{S} = 1 + \bar{\nu} + \bar{\nu}^2 + \cdots = 1/(1 - \bar{\nu}) \quad (4)$$

if $\bar{\nu} < 1$, while $E\bar{S} = \infty$ otherwise.

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

$$F < 1. \quad (5)$$

Furthermore,

$$h(t) = E[\gamma(t, Z)/\nu], \quad t > 0, \quad (6)$$

and

$$\bar{h}(t) = E\gamma(t, Z)/\bar{\nu}, \quad t > 0, \quad (7)$$

are well-defined densities (with respect to Lebesgue measure). The density $\bar{h}$ will play a key role later in this paper; it can be interpreted as the normalized intensity function for the first generation of offspring in a cluster started at time $0$. Note that $h$ specifies the density of the distance $R$ from an arbitrary offspring to its nearest ancestor. In the sequel, since the clusters relative to their cluster centers are i.i.d. (Section 2.1), we assume without loss of generality that $L$, $R$ and $S$ are defined with respect to the same immigrant $t_0 = 0$, with mark $Z_0 = Z$.

Clearly, if $L > 0$ then $R > t$ implies $L > t$, so the distribution of $L$ has a thicker tail than that of $R$. The probability function for $S$ is given by $P(S = k) = P(S_{n+1} = k - 1|S_n = k)/k$, $k \in \mathbb{N}$, where $S_n$ denotes the number of events of $n$-th generation and $n \in \mathbb{N}$ is arbitrary (see [9] or Theorem 2.11.2 in [15]). Thus

$$P(S = k) = E\left[e^{-k\nu}(k\nu)^{k-1}/k!ight], \quad k \in \mathbb{N}. \quad (8)$$

### 2.3. Examples

Throughout the paper, we illustrate the results with the following cases.

**Example 1.** *(Unmarked process)* An unmarked Hawkes process with exponentially decaying fertility rate is given by

$$\bar{\nu} = \nu = \alpha, \quad \bar{h}(t) = h(t) = \beta e^{-\beta t},$$

$$h(t) = E[\gamma(t, Z)/\nu], \quad t > 0,$$
where $0 < \alpha < 1$ and $\beta > 0$ are parameters. Here $1/\beta$ is a scale parameter for both the distribution of $R$ and the distribution of $L$.

Figure 2 (at left) shows perfect simulations on $[0, 10]$ of this process when $\mu(t) = 1$ is constant, $\alpha = 0.9$, and $\beta = 10, 5, 2, 1$. By (4), we expect to see about 10 clusters (in total) and 100 events. The clusters of course become more visible as $\beta$ increases.

Figure 3 (at left) shows six simulations of clusters with $\alpha = 0.9$ (being an inverse scaling parameter, $\beta$ is irrelevant since we have omitted showing the scale to get comparable results for this example and the following two examples). All the clusters have been simulated conditional on $S > 1$ to avoid the frequent and rather uninteresting case containing only the immigrant. These few simulations indicate the general tendency that $L$ vary fairly much. □

![Figure 2: Left: Four perfect simulations on $[0, 10]$ of the unmarked Hawkes process (Example 1) with parameters $\alpha = 0.9$, $\mu = 1$, and $\beta = 10, 5, 2, 1$ (top to bottom). Random jitter has been added in the vertical direction to help distinguishing events located close together. Right: Three perfect simulations on $[0, 10]$ of the birth-death Hawkes process (Example 2) with parameters $\alpha = 0.9$, $\mu = 1$, and $\beta = 5, 2, 1$ (top to bottom), where the projections of the lines onto the horizontal axis show the size of the marks.](image)

**Example 2.** (Birth-death process) Consider a marked Hawkes process with

$$\gamma(t, Z) = \alpha 1[t \leq Z]/EZ,$$

where $0 < \alpha < 1$ is a parameter, $Z$ is a positive random variable, and $1[\cdot]$ 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 birth rate is

$$\lambda(t) = \mu(t) + (\alpha/EZ)\text{card}\{\{i : t_i < t \leq t_i + Z_i\}\}, \quad t \in \mathbb{R},$$

cf. (1). Moreover,

$$\nu = \alpha Z/EZ, \quad \bar{\nu} = \alpha, \quad h(t) = E(1[t \leq Z]/Z), \quad \bar{h}(t) = P(Z \geq t)/EZ.$$  

Since $\nu$ is random, the distribution of $S$ is more dispersed than in the unmarked case, cf. (8).

The special case where $\mu(t) = \mu$ is constant and $Z$ is exponentially distributed with mean $1/\beta$ is considered at page 136 in [5]. Then $X$ is a time-homogeneous Markov birth-death process with birth rate $\mu + \alpha \beta n$ and death rate $\beta n$, when $n$ is the number of living individuals. In this case $\bar{h}(t) = \beta e^{-\beta t}$ and $h(t) = \beta E_1(\beta t)$, where $E_1(s) = \int_s^\infty e^{-t}/t \, dt$ is the exponential integral function. As in Example 1, $1/\beta$ is a scale parameter for the distribution of $L$. As discussed in Example 8 in Section 5, the stationary distribution (i.e. the distribution of $X$ at any fixed time) is known up to proportionality and it is possible to simulate from this by rejection sampling.

Figure 2 (at right) shows three perfect simulations in the Markov case on $[0, 10]$ with $\mu = 1$, $\alpha = 0.9$, and $\beta = 5, 2, 1$, where the marks are indicated by line segments of different lengths. Figure 3 (at middle) shows six simulations of clusters (with
marks excluded) with $\alpha = 0.9$ simulated conditional on $S > 1$. These simulations slightly indicate that $L$ is more dispersed than in Example 1, since the marks introduce additional variation in the cluster lengths. In fact the coefficient of variation estimated from 10000 perfect simulations is 1.92 for Example 1 and 2.85 for the present case. \[\]

**Example 3.** (A heavy-tailed distribution for $L$) Suppose that

$$\gamma(t, Z) = \alpha Ze^{-t},$$

where $\alpha \in (0,1)$ is a parameter and $Z$ is exponentially distributed with mean $1/\beta$. Then $\tilde{\nu} = \nu = \alpha$ is constant, so the distribution of $S$ is the same as in the unmarked case, cf. (8). Further,

$$h(t) = \tilde{h}(t) = \beta/(t + \beta)^2$$

specifies a Pareto density. This is a heavy-tailed distribution as is has infinite Laplace transform ($\mathcal{L}(\theta) = Ee^{\theta R} = \infty$ for all $\theta > 0$). Moreover it has infinite moments ($E(R^p) = \infty$ for all $p \geq 1$). Consequently, $L$ also has a heavy-tailed distribution with infinite moments and infinite Laplace transform. Note that $\beta$ is a scale parameter for the distribution of $L$.

Figure 3 (at right) shows six simulations of clusters with $\alpha = 0.9$ and $\beta = 1$. These indicate that $L$ is much more dispersed than in Examples 1 and 2 (in fact the dispersion is infinite in the present case). \[\]

**3. Perfect Simulation**

Assuming for the moment that $F$ (the c.d.f. for the length of a cluster) 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 [6]; see also [17] and [21].

**Algorithm 2.** Let $I_1$ be the point process of immigrants on $[0, t_+)$, and let $I_2$ be the point process of immigrants $t_i < 0$ such that $\{(t_j, Z_j) \in C_i : t_j \in [0, \infty]\} \neq \emptyset$.

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

2. For each $t_i \in I_1$, simulate $Z_i$ and those $(t_j, Z_j) \in C_i$ with $t_i < t_j < t_+$. 


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

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

5. The output is all marked points from (1), (2), and (4).

Remark 1. In steps (1) and (2) of Algorithm 2, we use Algorithm 1 (with $t_0 = 0$). In step (4), 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 $Z_i$ from $Q$ and the successive generations of offspring $t_j$ to $t_i$ (together with their marks $Z_j$) until there is at least one event of $C_i$ after time 0.

The key point is how to simulate $I_2$ in step (3), since this requires the knowledge of $F$, which is unknown on closed from (Remark 3, Section 4.1). In Section 4 we address this problem, and in Section 5 we construct an algorithm for simulating $I_2$.

In practice we must require that $I_2$ is (almost surely) finite or equivalently that

$$
\int_{-\infty}^0 (1 - F(-t))\mu(t)\,dt < \infty. \tag{9}
$$

In the case where $\mu(t)$ is bounded, (9) is satisfied if $\sup_{t\geq 0} \mu(t) EL < \infty$. A condition for finiteness of EL is established in Lemma 1 and Remark 2 below. □

Proposition 1. The output of Algorithm 2 follows the distribution of the marked Hawkes process.

Proof. The immigrant process minus $I_1 \cup I_2$ generates clusters with no events in $[0, t_+)$. Since $I_1$ is the immigrants on $[0, t_+)$, it follows directly that $I_1$ is a Poisson process with intensity $\lambda_1(t) = \mu(t)$ on $[0, t_+)$. Since $I_2$ is those immigrants on $(-\infty, 0)$ with offspring after 0, $I_2$ can be viewed as an independent thinning of the immigrant process with retention probability $p(t) = 1 - F(-t)$, and thus $I_2$ is a Poisson process with intensity $\lambda_2(t) = (1 - F(-t))\mu(t)$. Since $I_1$ and $I_2$ are independent, it follows from Section 2.1 that $\{C_i : t_i \in I_1\}$ and $\{C_i : t_i \in I_2\}$ are independent. Viewing the marked Hawkes process as a Poisson cluster process, it follows from Remark 1 that the clusters are generated in the right way in (2) and (4) when we only want to sample those marked points $(t_j, Z_j)$ with $t_j \in [0, t_+)$. Thus Algorithm 2 is correct. □
Using a notation as in Section 2.2, the following lemma generalises and sharpens a result in [14] about the mean length of a cluster.

**Lemma 1.** We have that

\[
\frac{1}{\mathbb{E} e^{-\nu}} \mathbb{E} \left[ (1 - e^{-\nu}) \mathbb{E} [R[Z]] \right] \leq EL \leq \frac{\bar{\nu}}{1 - \bar{\nu}} \mathbb{E} \hat{R}. \tag{10}
\]

**Proof.** Consider a cluster starting with an immigrant at time \( t_0 = 0 \), with mark \( Z_0 = Z \), cf. Section 2.1. For \( t_j \in \mathcal{G}_1 \), let \( R_j \) denote the distance from \( t_j \) to 0, and \( L_j \) the length of the total offspring \( C_j \) process started by \( t_j \). Then \( L = \max \{ R_j + L_j : t_j \in \mathcal{G}_1 \} \), so if we condition on \( Z \), and let \( R_j;Z \) be distributed as \( R_j \) conditional on \( Z \), then

\[
 EL = \mathbb{E} E[L[Z]] = \mathbb{E} \left[ \sum_{i=1}^{\infty} \frac{e^{-\nu \nu^i}}{i!} \mathbb{E} \left[ \max \{ R_{ij,Z} + L_j : j = 1, \ldots, i \} \right] \right]. \tag{11}
\]

To obtain the upper inequality observe that

\[
 EL \leq \mathbb{E} \left[ \sum_{i=1}^{\infty} \frac{e^{-\nu \nu^i}}{i!} \mathbb{E} \left[ \sum_{j=1}^{i} (R_{j,Z} + L_j) \right] \right] = \mathbb{E} [\nu \mathbb{E} [R[Z]]] + \bar{\nu} EL,
\]

where we have used that the \( L_j \) are identically distributed and has the same distribution as \( L \) because of the self-similarity property (Section 2.1), and that the \( R_j \) are identically distributed when conditioned on \( Z \). Hence

\[
 EL \leq \frac{1}{1 - \bar{\nu}} \mathbb{E} [\nu \mathbb{E} [R[Z]]] = \frac{1}{1 - \bar{\nu}} \mathbb{E} \left[ \int_0^{\infty} s \gamma(s, Z) ds \right] = \frac{\bar{\nu}}{1 - \bar{\nu}} \mathbb{E} \hat{R},
\]

which verifies the upper inequality. Finally, by (11),

\[
 EL \geq \mathbb{E} \left[ \sum_{i=1}^{\infty} \frac{e^{-\nu \nu^i}}{i!} (\mathbb{E} [R[Z] + EL] \right] = \mathbb{E} [(1 - e^{-\nu}) \mathbb{E} [R[Z]]] + \mathbb{E} [1 - e^{-\nu}] EL,
\]

which reduces to the lower inequality. \( \square \)

**Remark 2.** If either \( \nu \) or \( \gamma/\nu \) is independent of \( Z \) (in other words, either the number or the locations of offspring in an offspring process are independent of the mark associated to the generic event), then it is easily proven that \( \bar{h} = h \) and thus (10) reduces to

\[
 \left( \frac{1}{\mathbb{E} e^{-\nu}} - 1 \right) \mathbb{E} \hat{R} \leq EL \leq \frac{\bar{\nu}}{1 - \bar{\nu}} \mathbb{E} \hat{R}.
\]

Consequently, \( EL < \infty \) if and only if \( \mathbb{E} \hat{R} < \infty \). This immediately shows that \( EL < \infty \) in Example 1 and \( EL = \infty \) in Example 3. In Example 2 when \( Z \) is exponentially
distributed with mean $1/\beta$, (10) becomes $\alpha(\alpha+2)/(2(\alpha+1)\beta) \leq EL \leq \alpha/(\beta(1-\alpha))$, so in this case $EL < \infty$. Not surprisingly, apart from small values of $\alpha \in (0,1)$, the bounds are rather poor and of little use except in establishing finiteness of $EL$. □

4. The distribution of the length of a cluster

In this section we derive various distributional results concerning the length $L$ of a cluster. The results are needed in Section 5 to complete step (3) in Algorithm 2; however, many of the results are also of own interest.

4.1. An integral equation for $F$

Below in Proposition 2 an integral equation for $F$ is derived, and it is discussed how to approximate $F$ by numerical methods, using a certain recursion. Proposition 2 is a generalisation of Theorem 5 in [14], which is proved using void probabilities obtained from a general result for the probability generating functional for an unmarked Hawkes process. However, as [8] point out, the probability generating functional for the marked Hawkes process is difficult to obtain. We give a direct proof based on void probabilities.

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 (it becomes clear in Section 4.2 why we use the notation $1_n$). Clearly, $1_n$ is decreasing in $n$, $1_n \to F$ pointwise as $n \to \infty$, and

$$1_0(t) = 1, \quad t \geq 0. \quad (12)$$

Furthermore, 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

$$\varphi(f)(t) = \mathbb{E} \left[ \exp \left( -\nu + \int_0^t f(t-s)\gamma(s, Z) \, ds \right) \right], \quad t \geq 0. \quad (13)$$

**Proposition 2.** We have that

$$1_n = \varphi(1_{n-1}), \quad n \in \mathbb{N}, \quad (14)$$

and

$$F = \varphi(F). \quad (15)$$

**Proof.** As in the proof of Lemma 1, we can consider a cluster started at time $t_0 = 0$ with associated marks $Z_0 = Z$. For fixed $t \geq 0$ and $n \in \mathbb{N}$, split $\Phi(0)$ into three point
processes $\Phi_1$, $\Phi_2$, $\Phi_3$: $\Phi_1$ consists of those first generation offspring $t_i \in \Phi(0) \cap [0, t)$ which do not generate events of generation $n - 1$ or lower with respect to $t_i$ on $[t, \infty)$; $\Phi_2 = (\Phi(0) \cap [0, t)) \setminus \Phi_1$ consists of the remaining first generation offspring on $[0, t)$; and $\Phi_3 = \Phi(0) \cap [t, \infty)$ are the first generation offspring on $[t, \infty)$. Conditional on $Z$, we have that $\Phi_1$, $\Phi_2$, and $\Phi_3$ are independent Poisson processes with intensity functions $\lambda_1(s) = \gamma(s, Z)F_{n-1}(t-s)$ on $[0, t)$, $\lambda_2(s) = \gamma(s, Z)(1 - F_{n-1}(t-s))$ on $[0, t)$, and $\lambda_3(s) = \gamma(s, Z)$ on $[t, \infty)$, respectively. This follows by an independent thinning argument, since conditional on $G_n$ (the $n$-th generation of offspring in $C_0$), the processes $C_j - t_j$ with $t_j \in G_n$ are i.i.d. and distributed as $C_0$ (this is the self-similarity property from Section 2.1). Consequently,

$$1_n(t) = E[P(\Phi_2 = \emptyset|Z)P(\Phi_3 = \emptyset|Z)]$$

$$= E \exp \left(- \int_0^t \lambda_2(s, Z) \, ds - \int_t^\infty \lambda_3(s, Z) \, ds \right)$$

which reduces to (14). Taking the limit as $n \to \infty$ on both sides of (14), we obtain (15) by monotone convergence, since $1_n(t) \leq 1_{n-1}(t)$ for all $t \geq 0$ and $n \in \mathbb{N}$. 

**Remark 3.** As illustrated in the following example, we have been unsuccessful in using (15) to obtain a closed form expression for $F$ even for simple choices of $\gamma$. Fortunately, the recursion (14) provides a useful numerical approximation to $F$. As the integral in (13) with $f = 1_{n-1}$ quickly becomes difficult to evaluate analytically as $n$ increases, we compute the integral numerically, using a quadrature rule. 

**Example 4.** (Unmarked process) Consider Example 1 with $\beta = 1$. Then (15) is equivalent to

$$\int_0^t F(s)e^s \, ds = \frac{e^t}{\alpha} \ln(e^\alpha F(t))$$

which is not analytically solvable. 

**4.2. Monotonicity properties and convergence results**

As established in Theorem 1 below, many other approximations of $F$ than $1_n$ exist, and the rate of convergence may be geometric with respect to different norms. First we notice that certain monotonicity properties are fulfilled by $\varphi$, where we for functions $f : [0, \infty) \mapsto [0, 1]$ set $f_0 = \varphi^{[0]}(f) = f$ and define recursively $f_n = \varphi^{[n]}(f) = \varphi(f_{n-1})$, $n \in \mathbb{N}$. 

N. Note that $F_n = F$ for all $n \in \mathbb{N}_0$. As $1_n = \varphi^{(n)}(1)$ is decreasing towards the c.d.f. $F$, cases where $G$ is a c.d.f. and $G_n$ increases to $F$ are of particular interest.

**Lemma 2.** For any $f, g \in \mathcal{C}$, we have that

\[ f \leq g \Rightarrow f_n \leq g_n, \quad n \in \mathbb{N}, \tag{16} \]

\[ f \leq \varphi(f) \Rightarrow f_n \text{ is non-decreasing in } n, \tag{17} \]

\[ f \geq \varphi(f) \Rightarrow f_n \text{ is non-increasing in } n. \tag{18} \]

**Proof.** We obtain immediately (16) from (13) when $n = 1$, whereby (16) follows by induction. Thereby (17) and (18) follow. □

**Theorem 1.** With respect to the supremum norm $\|f\|_\infty = \sup_{t \geq 0} |f(t)|$, $\varphi$ is a contraction on $\mathcal{C}$, that is, for all $f, g \in \mathcal{C}$ and $n \in \mathbb{N}$, we have that $f_n, g_n \in \mathcal{C}$ and

\[ \|\varphi(f) - \varphi(g)\|_\infty \leq \bar{\nu}\|f - g\|_\infty. \tag{19} \]

Further, $F$ is the unique fixedpoint,

\[ \|F - f_n\|_\infty \to 0 \quad \text{as } n \to \infty, \tag{20} \]

and

\[ \|F - f_n\|_\infty \leq \frac{\bar{\nu}^n}{1 - \bar{\nu}} \|\varphi(f) - f\|_\infty, \tag{21} \]

where $\|\varphi(f) - f\|_\infty \leq 1$. Furthermore, if $f \leq \varphi(f)$ (or $f \geq \varphi(f)$), then $f_n$ converges to $F$ from below (above).

**Proof.** Let $f, g \in \mathcal{C}$. Recall that by the mean value theorem (e.g. Theorem 5.11 in [1]), for any real numbers $x$ and $y$, $e^x - e^y = (x - y)e^{z(x,y)}$, where $z(x,y)$ is a real number between $x$ and $y$. Thus by (13),

\[ \|\varphi(f) - \varphi(g)\|_\infty = \sup_{t \geq 0} \left| \mathbb{E} \left[ e^{-\nu t} e^{c(t,f,g)} \int_0^t (f(s) - g(s)) \gamma(s,Z) \, ds \right] \right| \tag{22} \]

where $c(t,f,g)$ is random variable between $\int_0^t f(s) \gamma(s,Z) \, ds$ and $\int_0^t g(s) \gamma(s,Z) \, ds$. Since $f, g \leq 1$, we obtain $e^{c(t,f,g)} \leq e^\nu$, cf. (2). Consequently,

\[ \|\varphi(f) - \varphi(g)\|_\infty \leq \sup_{t \geq 0} \left| \mathbb{E} \left[ \int_0^t (f(s) - g(s)) \gamma(s,Z) \, ds \right] \right| \leq \mathbb{E} \left[ \int_0^\infty \|f - g\|_\infty \gamma(s,Z) \, ds \right] = \bar{\nu} \|f - g\|_\infty. \]
Thereby (19) is verified. Since $\mathcal{C}$ is complete (see e.g. Theorem 3.11 in [26]), it follows from the fixpoint theorem for contractions (see e.g. Theorem 4.48 in [1]) that the contraction has a unique fixpoint. By (15), this is $F$.

Since $f \in \mathcal{C}$ implies $\varphi(f) \in \mathcal{C}$, we get by induction that $f_n \in \mathcal{C}$. Hence, using (15), (19) and induction,

$$
\|f_n - F\|_\infty = \|\varphi(f_{n-1}) - \varphi(F)\|_\infty \leq \bar{\nu}\|f_{n-1} - F\|_\infty \leq \bar{\nu}^n\|f - F\|_\infty,
$$

(23)

for $n \in \mathbb{N}$. Since $\bar{\nu} < 1$, (20) is obtained.

For similar reasons as in (23),

$$
\|f_n - f_{n-1}\|_\infty \leq \bar{\nu}^{n-1}\|f_1 - f\|_\infty, \quad n \in \mathbb{N}.
$$

(24)

Further, by (20),

$$
\|F - f\|_\infty = \lim_{m \to \infty} \|f_m - f\|_\infty.
$$

So by the triangle inequality and (24),

$$
\|F - f\|_\infty \leq \lim_{m \to \infty} (\|f_1 - f\|_\infty + \|f_2 - f_1\|_\infty + \cdots + \|f_m - f_{m-1}\|_\infty)
\leq \lim_{m \to \infty} \|f_1 - f\|_\infty (1 + \bar{\nu} + \cdots + \bar{\nu}^{m-1}) = \|f_1 - f\|_\infty / (1 - \bar{\nu}),
$$

cf. (2). Combining this with (23), we obtain (21). Finally, if $f \leq \varphi(f)$ (or $f \geq \varphi(f)$), then by (17) (or (18)) and (20), $f_n$ converges from below (or above). $\square$

Similar results to those in Theorem 1 but for the $L^1$-norm are established in [18].

The following remark and proposition show how to find upper and lower bounds of $F$ in many cases.

**Remark 4.** Consider a function $f \in \mathcal{C}$. The condition $f \leq \varphi(f)$ or $f \geq \varphi(f)$ is satisfied for the extreme cases $f = 0$ or $f = 1$. The upper bound $f = 1$ is useful in the following sections, but the lower bound $f = 0$ is a too small function for our purposes; if we require that $EL < \infty$, cf. Remarks 1 (in fact we use only $f = 0$ when producing the right plot in Figure 4). To obtain a more useful lower bound, observe that $f \leq \varphi(f)$ implies $f \leq F < 1$, cf. (5) and Theorem 1. If $f < 1$, then a sufficient condition for $f \leq \varphi(f)$ is

$$
\frac{1}{\bar{\nu}} \geq \int_0^t (1 - f(t - s))h(s)\,ds + \int_t^\infty h(s)\,ds, \quad t \geq 0.
$$

(25)
This follows readily from (7) and (13), using that $e^x \geq 1 + x$.

The case where $f$ in (25) is closest to $F$ happens when $f$ is a c.d.f. $G$ and we have equality in (25). Equivalently, $G$ satisfies the renewal equation

$$G(t) = 1 - \bar{\nu} + \bar{\nu} \int_0^t G(t-s)\bar{h}(s)\,ds, \quad t \geq 0,$$

which has the unique solution

$$G(t) = 1 - \bar{\nu} + \sum_{n=1}^{\infty} (1 - \bar{\nu})\bar{\nu}^n \int_0^t \bar{h}^{*n}(s)\,ds, \quad t \geq 0, \quad (26)$$

where $*n$ denotes convolution $n$ times, cf. Theorem IV2.4 in [2]. In other words, $G$ is the c.d.f. of $\bar{R}_1 + \cdots + \bar{R}_K$ (setting $\bar{R}_1 + \cdots + \bar{R}_K = 0$ if $K = 0$), where $K, \bar{R}_1, \bar{R}_2, \ldots$ are independent random variables, each $\bar{R}_i$ has density $\bar{h}$, and $K$ follows a geometric density $(1 - \bar{\nu})\bar{\nu}^n$. Interestingly, this geometric density is equal to $\text{E}S_n/\text{ES}$, cf. (4).

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. In such cases $F$ has no heavier tails than such an exponential distribution. □

Denote by

$$\mathcal{L}(\theta) = \int_0^\infty e^{\theta t}h(t)\,dt, \quad \theta \in \mathbb{R},$$

the Laplace transform of $h$.

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

**Proof.** Inserting $f = G$ into the right side of (25) we obtain

$$\int_0^t e^{\theta s}\bar{h}(s)\,ds + e^{\theta t} \int_t^\infty \bar{h}(s)\,ds.$$

Since this is an increasing function of $t > 0$, (25) is satisfied if and only if $\mathcal{L}(\theta) \leq 1/\bar{\nu}$. □

Note that Proposition 3 always applies for sufficiently small $\theta > 0$ except in the case where $\bar{h}$ is heavy-tailed in the sense that $\mathcal{L}(\theta) = \infty$ for all $\theta > 0$.

**4.3. Examples**

For Examples 5 and 6 below, we let

$$G(t) = 1 - e^{-\theta t}, \quad t \geq 0,$$  \quad (27)
be the exponential c.d.f. with parameter $\theta > 0$.

**Example 5.** *(Unmarked process)* For the case in Example 1, $\mathcal{L}(\theta) = \beta/(\beta-\theta)$ if $\theta < \beta$, and $\mathcal{L}(\theta) = \infty$ otherwise. Interestingly, for “the best choice” $\theta = \mathcal{L}^{-1}(1/\bar{\nu}) = \beta(1-\alpha)$, (27) becomes the c.d.f. for $R$ times $E_S$, which is easily seen to be the same as the c.d.f. in (26).

Figure 4 (at left) shows $1_n$ and $G_n$ when $\theta = \beta(1-\alpha)$ and $(\alpha, \beta) = (0.9, 1)$. The convergence of $1_n$ and $G_n$ (with respect to $\| \cdot \|_\infty$) and the approximate form of $F$ are clearly visible. Since $G$ is a c.d.f. and $G_{n+1} \geq G_n$, we have that $G_n$ is a c.d.f. Figure 4 (at middle) shows the density $F'(t)/(1 - F(0)) \ (t > 0)$ approximated by $[1_n'(t)/(1-1_n(0)) + G_n'(t)/(1 - G_n(0))]/2$ when $n = 50$ (in which case $1_n'(t)/(1-1_n(0))$ and $G_n'(t)/(1 - G_n(0))$ are effectively equal). As shown in the plot, the density is close to the exponential density with the same mean, but the tail is slightly thicker. $\Box$

**Example 6.** *(Birth-death process)* For the case in Example 2,

$$\mathcal{L}(\theta) = E \int_0^Z e^{\theta s}/E Z \ ds = \frac{\mathcal{L}_Z(\theta) - 1}{\theta E Z}$$

where $\mathcal{L}_Z(\theta) = E e^{\theta Z}$ is the Laplace transform for $Z$. In the special case where $Z$ is exponentially distributed with mean $1/\beta$, $\mathcal{L}(\theta) = \mathcal{L}_Z(\theta) = \beta/(\beta - \theta)$ is of the same
form as in Example 5. Plots of $1_n, G_n$, and \([1_n/(1 - 1_n(0)) + G'_n/(1 - G_n(0))]/2\) for $n = 0, 5, \ldots, 50$ and $(\alpha, \beta) = (0.9, 1)$ are similar to those in Figure 4 (at right and middle) and are therefore omitted. \(\Box\)

**Example 7.** (A heavy-tailed distribution for $L$) For the case in Example 3, Proposition 3 does not apply as $\mathcal{L}(\theta) = \infty$ for all $\theta > 0$. The c.d.f. in (26) is not known on closed form, since the convolutions are not tractable (in fact this is the case when $\bar{h}$ specifies any known heavy-tailed distribution, including the Pareto, Weibull, lognormal or loggamma distribution). Nonetheless, it is still possible to get an idea of what $F$ looks like: Figure 4 (at right) shows $1_n$ and $0_n$ for $n = 0, 5, \ldots, 50$ in the case $(\alpha, \beta) = (0.9, 1)$. As in Examples 5 and 6, the convergence of $1_n$ and $G_n$ (where now $G = 0$) and the approximate form of $F$ are clearly visible. However, as indicated in Figure 4 (at right) and verified in [18], $\lim_{t \to 0} G_n(t) < 1$ when $G = 0$, so $G_n$ is not a c.d.f. \(\Box\)

## 5. Simulation of $I_2$

To complete the perfect simulation algorithm (Algorithm 2 in Section 3), we need a useful way of simulating $I_2$. Our procedure is based on a dominating process and the use of coupled upper and lower processes in a similar spirit as in the dominated coupling from the past algorithm [16].

Suppose that $f \in \mathcal{C}$ is on a closed form, $f \leq \varphi(f)$, and (9) is satisfied when we replace $F$ by $f$ (situations where these requirements are fulfilled are considered in Sections 3, 4.2 and 4.3). Particularly, if $\mu$ is constant and $f$ is a c.d.f., (9) means that $f$ has a finite mean. Now, for $n \in \mathbb{N}_0$, let $U_n$ and $L_n$ denote Poisson processes on $(-\infty, 0)$ with intensity functions $\lambda^n(t) = (1 - f_n(-t))\mu(t)$ and $\lambda'_n(t) = (1 - 1_n(-t))\mu(t)$, respectively. By Theorem 1, $\lambda^n_n$ is non-increasing and $\lambda'_n$ is non-decreasing in $n$, and they both converge to $\lambda_2$ (geometrically fast with respect to the supremum norm). Consequently, we can use independent thinning to obtain the following sandwiching/funneling property, cf. [16]:

$$\emptyset = L_0 \subseteq L_1 \subseteq L_2 \subseteq \cdots \subseteq I_2 \subseteq \cdots \subseteq U_2 \subseteq U_1 \subseteq U_0.$$  \(28\)

The details are given by the following algorithm.
Algorithm 3. Simulation of $I_2$:

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

2. If $U_0 = \emptyset$, then return $I_2 = \emptyset$ and stop, else generate independent uniform numbers $W_1, \ldots, W_k$ on $[0, 1]$ (independently of $U_0$), and set $n = 1$.

3. For $j = 1, \ldots, k$, assign $(t_j, Z_j)$ to $L_n$ respective $U_n$ if $W_j \lambda^u_0(t_j) \leq \lambda^u_n(t_j)$ respective $W_j \lambda^u_0(t_j) \leq \lambda^u_n(t_j)$.

4. If $U_n = L_n$, then return $I_2 = L_n$ and stop, else increase $n$ by 1 and repeat steps (3)–(4).

Proposition 4. Algorithm 3 works correctly and terminates almost surely within finite time.

Proof. To see this, imagine no matter if $U_0 = \emptyset$ in step (2) or $U_n = L_n$ in step (4), we continue to generate $(U_1, L_1)$, $(U_2, L_2)$, etc. Further, add an extra step: For $j = 1, \ldots, k$, assign $(t_j, Z_j)$ to $I_2$ if and only if $W_j \lambda^u_0(t_j) \leq \lambda^u_2(t_j)$. Then clearly, because of the convergence properties of $\lambda^u_n$ and $\lambda^u_n$ (see the discussion above), (28) is satisfied and conditional on $t_1, \ldots, t_k$,

$$P(L_n \neq U_n \ \forall n \in \mathbb{N}_0) \leq \sum_{j=1}^{k} \lim_{n \to \infty} P(W_j \lambda^u_0(t_j) \leq \lambda^u_n(t_j), W_j \lambda^u_0(t_j) > \lambda^u_n(t_j))$$

$$= \sum_{j=1}^{k} P(\lambda^u_2(t_j) < W_j \lambda^u_0(t_j) \leq \lambda^u_2(t_j)) = 0.$$ 

Thus almost surely Algorithm 3 terminates within finite time and the output equals $I_2$. $\square$

Remark 5. We compute $1_n$ and $f_n$ numerically, using a quadrature rule, cf. Remark 3. After step (1) in Algorithm 3, we let the last quadrature point be given by $-t_1$ (since we do not need to calculate $1_n(t)$ and $f_n(t)$ for $t > -t_1$). Since we have to calculate $1_n$ and $f_n$ recursively for all $n = 0, 1, 2, \ldots$ until Algorithm 3 terminates, there is no advantage in using a doubling scheme for $n$ like in the Propp-Wilson algorithm [24]. $\square$

Example 8. (Birth-death process) We have checked our computer code for Algorithms 2 and 3 by comparing with results produced by another perfect simulation
algorithm: Consider the case in Example 2 when $\mu(t) = \mu$ is constant and $Z$ is exponentially distributed with mean $1/\beta$. If $N$ denotes the number of events alive at time 0, we have the following detailed balance condition for its equilibrium density $\pi_n$:

$$\pi_n(\mu + \alpha \beta n) = \pi_{n+1} \beta (n+1), \quad n \in \mathbb{N}_0.$$ 

This density is well-defined, since $\lim_{n \to \infty} \pi_{n+1}/\pi_n = \alpha < 1$. Now, choose $m \in \mathbb{N}_0$ and $\epsilon \geq 0$ such that $a = \alpha + \epsilon < 1$ and $\pi_{n+1}/\pi_n \leq a$ whenever $n \geq m$. If $\mu \leq \alpha \beta$, we can take $\epsilon = m = 0$; otherwise we can use $m = (\mu - \alpha \beta)/(\beta \epsilon)$ for some $\epsilon > 0$. Define an unnormalised density $\pi'_n$, $n \in \mathbb{N}_0$, by $\pi'_n = \pi_n/\pi_0$ if $n \leq m$, and $\pi'_n = a^{n-m} \pi_m/\pi_0$ otherwise. We can easily sample from $\pi'_n$ by inversion, cf. [25], since we can calculate

$$\sum_{0}^{\infty} \pi'_n = \sum_{0}^{m} \pi_n/\pi_0 + \frac{a}{1-a} \pi_m/\pi_0.$$

Then, since $\pi'_n \geq \pi_n/\pi_0$, we can sample $N$ from $\pi_n$ by rejection sampling, cf. [25]. Furthermore, conditional on $N = n$, we generate $n$ independent marks $Z'_1, \ldots, Z'_n$ which are exponentially distributed with mean $1/\beta$ (here we exploit the memoryless property of the exponential distribution). Finally, we simulate the marked Hawkes process with events in $(0, t+]$, using the conditional intensity

$$\lambda'(t) = \mu + \alpha \beta \left( \sum_{i=1}^{n} \mathbf{1}[t < Z'_i] + \sum_{0 < t_i < t} \mathbf{1}[t < t_i + Z'_i] \right).$$

We have implemented this algorithm for comparison with our algorithm. Not surprisingly this algorithm is a lot faster than our perfect simulation algorithm (roughly 1200 times as fast in the case $\alpha = 0.9$, $\beta = 1$, and $t_+ = 10$), since it exploits the fact that we know the stationary distribution in this special case. □

6. Extensions and open problems

Except for the heavy-tailed case, our perfect simulation algorithm is feasible in the examples we have considered. However, simulation of the heavy-tailed cases is an unsolved problem. In these cases, we can only say something about the approximate form of $F$, cf. Example 7.

For applications such as in seismology [23], extensions of our results and algorithms to the heavy-tailed cases are important. The ETAS-model (epidemic type aftershock
sequences) from [22] used for modeling times and magnitudes of earthquakes is a heavy-tailed marked Hawkes process. Its spatio-temporal extension, which also includes the locations of the earthquakes (see [23]), furthermore includes the problem of predictable marks (the location of an aftershock depends on the location of the earthquake that causes it). This problem is easily solved, though, since the times and magnitudes are independent of the locations and can be simulated without worrying about these. This, of course, still leaves the unsolved problem of the heavy tails.

Extensions to non-linear Hawkes processes [3, 8] would also be interesting. Again things become complicated, since a non-linear Hawkes process is not even a Poisson cluster process.

Simulations of Hawkes processes with predictable marks can in some cases be obtained by using a thinning algorithm, if it is possible to dominate the Hawkes process with predictable marks by a Hawkes process with unpredictable marks. We illustrate the procedure with a simple birth-death example.

**Example 9. (Birth-death process)** Consider two birth-death Hawkes processes as defined in Example 2: \( \Psi_1 \) with unpredictable marks, \( Z_1^i \sim \text{Exp}(\beta) \), and \( \Psi_2 \) with predictable marks, \( Z_2^i \sim \text{Exp}(\beta+1/Z_{\Delta_0}) \), where \( Z_{\Delta_0} \) is the mark of the first order ancestor of \( t_i \). Both models have \( \gamma(t, Z) = \alpha/\beta 1[t < Z] \) with the same \( \alpha \) and \( \beta \) and they also have the same \( \mu(t) \). The model \( \Psi_2 \) has the intuitive appeal that long-living individuals have long-living offspring. Note that the intensity of \( \Psi_1 \) dominates the intensity of \( \Psi_2 \) if the marks are simulated such that \( Z_1^i > Z_2^i \).

To simulate \( \Psi_2 \), we first simulate \( \Psi_1 \) using Algorithm 3 with the modifications that we associate both marks \( Z_1^i \) and \( Z_2^i \) to the event \( t_i \), and we keep all events from the algorithm whether they fall in or before \([0, t_+)]\). Each marked event \((t_j, Z_j^j)\) is then included in \( \Psi_2 \) with retention probability

\[
\frac{\mu(t) + \alpha \beta \sum_{t_k < t_j, t_k \in \Psi_2} 1[t_j - t_k < Z_k^j]}{\mu(t) + \alpha \beta \sum_{t_k < t_j} 1[t_j - t_k < Z_k^j]}
\]

and the final output is all marked events from \( \Psi_2 \) falling in \([0, t_+)]\). It is easily proven that these retention probabilities result in the right process \( \Psi_2 \). \( \square \)

Another process that would be interesting to obtain by thinning is the Hawkes process without immigrants considered in [4]; this process has \( \mu(t) = 0 \) for all \( t \).
However, for this to be non-trivial (i.e. not almost surely empty), it is necessary that $\tilde{\nu} = 1$, which means that any dominating Hawkes process has $\tilde{\nu} \geq 1$ and thus cannot be simulated by Algorithm 3.

Many of our results and algorithms can be modified if we slightly extend the definition in Section 1 of a marked Hawkes process as follows: For any event $t_i$ with associated mark $Z_i$, let $n_i$ denote the number of (first generation) offspring generated by $(t_i, Z_i)$, and suppose that $n_i$ conditional on $Z_i$ is not necessarily Poisson distributed, but $n_i$ is still conditionally independent of $t_i$ and the previous history. A particular simple case occurs when $n_i$ is either 1 or 0, where $\tilde{p} = \text{EP}(n_i = 1|Z_i)$ is assumed to be strictly between 0 and 1 (here $\tilde{p}$ plays a similar role as $\tilde{\nu}$ introduced in Section 4). Then we redefine $\varphi$ by

$$\varphi(f)(t) = 1 - \tilde{p} + \tilde{p} \int_0^t f(t-s)\bar{h}(s) \, ds$$

where now

$$\bar{h}(s) = \text{E}(p(Z)h(s, Z))/\tilde{p}.$$

Since $\varphi$ now is linear, the situation is much simpler. For example, $F$ is given by $G$ in (26) (with $\varphi$ replaced by $\tilde{\varphi}$).

Another extension of practical relevance is to consider a non-Poisson immigrant process, e.g. a Markov or Cox process. The results in Section 4 do not depend on the choice of immigrant process, and the straightforward simulation algorithm (Algorithm 1) applies provided it is feasible to simulate the immigrants on $[t_-, t_+]$. However, the perfect simulation algorithm relies much on the assumption that the immigrant process is Poisson.

Finally, we notice that it would be interesting to extend our ideas to spatial Hawkes processes, cf. [19] and [20].

**Acknowledgement**

We are grateful to Søren Asmussen, Svend Berntsen, Horia Cornean, Martin Jacobsen, Arne Jensen, and Giovanni Luca Torrisi for helpful discussions. 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


