

# PERFECT SIMULATION OF HAWKES PROCESSES

JESPER MØLLER \* \*\* AND

JAKOB G. RASMUSSEN, \* \*\*\* Aalborg University

## Abstract

Our objective is to construct 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 structures, 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. A tail-lightness condition turns out to be of importance for the applicability of the perfect simulation algorithm. Examples of applications and empirical results are presented.

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

2000 Mathematics Subject Classification: Primary 60G55

Secondary 68U20

## 1. Introduction

Unmarked and marked Hawkes processes [10], [11], [12], [14] play a fundamental role in point process theory and its applications (see, e.g. [8]), and they have applications in seismology [13], [22], [23], [27] and neurophysiology [7]. There are many ways to define a marked Hawkes process, but for our purposes 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 independent and identically distributed (i.i.d.) with distribution  $Q$ , and are independent of the immigrants.
- (c) Each immigrant  $t_i$  generates a *cluster*  $C_i$ , which consists of marked events of generations of order  $n = 0, 1, \dots$  with the following *branching structure* (see Figure 1). We first have  $(t_i, Z_i)$ , which is said to be of generation 0. Given the  $0, \dots, n$  generations in

Received 3 December 2004; revision received 13 April 2005.

\* Postal address: Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, DK-9220 Aalborg, Denmark.

\*\* Email address: jm@math.aau.dk

\*\*\* Email address: jgr@math.aau.dk

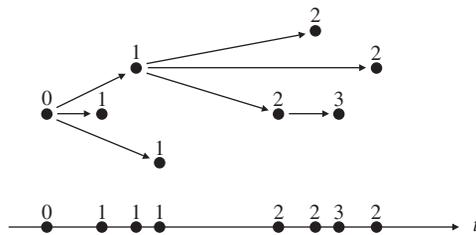


FIGURE 1: The branching structure of the various generations of events in a cluster (ignoring the marks) (top), and the events on the time axis (bottom).

$C_i$ , each  $(t_j, Z_j) \in C_i$  of generation  $n$  recursively generates a Poisson process  $\Phi_j$  of offspring of generation  $n + 1$  with intensity function  $\gamma_j(t) \equiv \gamma(t - t_j, Z_j), t > t_j$ . Here,  $\gamma$  is a nonnegative 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 mark  $Z_k$  associated 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) Given the immigrants, the clusters are independent.
- (e)  $X$  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

$$P(\gamma(t, Z) = \gamma(t) \text{ for Lebesgue almost all } t > 0) = 1,$$

where  $Z$  denotes a generic mark), the events follow an unmarked Hawkes process. Apart from in 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

$$\lambda(t) = \mu(t) + \sum_{t_i < t} \gamma(t - t_i, Z_i). \tag{1}$$

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; and displaying simulated realizations of specific model constructions provides a better understanding of the model. The general approach to simulating a (marked or unmarked) point process is to use a thinning algorithm such as the Shedler–Lewis thinning algorithm or Ogata’s modified thinning algorithm (see, e.g. [8]). However, Definition 1 immediately leads to the following approximate simulation algorithm, where  $t_- \in [-\infty, 0]$  and  $t_+ \in (0, \infty]$  are user-specified parameters and the output consists of all marked points  $(t_i, Z_i)$  with  $t_i \in [0, t_+)$ .

**Algorithm 1.** The following steps (i)–(ii) generate an approximate 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  $[0, t_+)$ . Brémaud *et al.* [5] studied the ‘rate of installation’, i.e. they considered a coupling of  $X$ , after time 0, to the output from Algorithm 1 when  $t_+ = \infty$ . Under a tail-lightness assumption (see the paragraph after Proposition 3, below) and other conditions, they established an exponentially decreasing bound for the probability  $P(t_-, \infty)$ , say, that  $X$ , after time 0, coincides with the output of the algorithm. Algorithm 1 was also investigated in the authors’ own work [18], where various measures for edge effects, including refined results for  $P(t_-, \infty)$ , were introduced.

Our 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 areas of application 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 can also be useful in quantifying the edge effects suffered by Algorithm 1 (see [18]).

The perfect simulation algorithm is derived using similar principles as in Brix and Kendall [6], but our algorithm is a nontrivial extension, since the Brix–Kendall algorithm requires the knowledge of the cumulative distribution function (CDF)  $F$  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 of [16]. Under a tail-lightness condition, our perfect simulation algorithm turns out to be feasible in applications, while, in the heavy-tailed case, we can at least say something about the approximate form of  $F$  (see Example 7).

The paper is organized as follows. Section 2 contains some preliminaries, including illuminating examples of Hawkes process models used throughout the paper to illustrate our results. In Section 3, we describe the perfect simulation algorithm, assuming that  $F$  is known, while the above-mentioned convergence and monotonicity results are established in Section 4. In Section 5, we complete the perfect simulation algorithm, using dominated coupling from the past. Finally, Section 6 contains a discussion of our algorithm and results, and suggestions on 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, where the clusters, given the immigrants, are independent. In this section, we describe a self-similarity property resulting from 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  $s_1, \dots, s_n$  of offspring such that  $s_n = t_j$  and  $s_k, k = 1, \dots, n$ , is one of the offspring of  $s_{k-1}$ , with  $s_0 = t_i$ . We then say 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 we define the *total offspring process*  $C_i$  as all those  $(t_j, Z_j)$  such that  $t_j$  is an event of generation  $n \in \mathbb{N}_0$  with respect to  $t_i$  (note that  $(t_i, Z_i) \in C_i$ ). The clusters are defined as those  $C_i$  for which  $t_i$  is an immigrant (see 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 the events  $t_i < t_j$ , the translated total offspring processes

$$C_i - t_i := \{(t_l - t_i, Z_l) : (t_l, Z_l) \in C_i\},$$

$$C_j - t_j := \{(t_l - t_j, Z_l) : (t_l, Z_l) \in C_j\}$$

are identically distributed.

In particular, conditional on the immigrants, the clusters relative to their cluster centres (the immigrants) are i.i.d. with distribution  $\mathbf{P}$ , say. Furthermore, conditional on a cluster's  $n$ th generation events  $\mathcal{G}_n$ , say, the translated total offspring processes  $C_j - t_j$  with  $t_j \in \mathcal{G}_n$  are i.i.d. with distribution  $\mathbf{P}$ . 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 CDF 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  $\Phi(t_i)$ ,  $\bar{\nu} := E \nu$ , 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

$$0 < \bar{\nu} < 1. \tag{2}$$

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

$$F(0) = E e^{-\nu} > 0,$$

where  $F(0)$  is the probability that a cluster has no offspring. It is equivalent to assuming that  $E S < \infty$ , where  $S$  denotes the number of events in a cluster: by induction on  $n = 0, 1, \dots$ , because of the branching and conditional independence structure of each cluster,  $\bar{\nu}^n$  is the mean number of generation  $n$  events in a cluster, meaning that

$$E S = 1 + \bar{\nu} + \bar{\nu}^2 + \dots = \frac{1}{1 - \bar{\nu}} \tag{3}$$

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

The other condition,  $\bar{\nu} > 0$ , excludes the trivial case in which there are almost surely no offspring. It is readily seen to be equivalent to

$$F < 1. \tag{4}$$

Furthermore,

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

and

$$\bar{h}(t) = E[\gamma(t, Z)/\bar{\nu}], \quad t > 0, \tag{5}$$

are well-defined densities (with respect to the 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 centres, are i.i.d. (see 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 that  $L > t$ , meaning the distribution of  $L$  has a thicker tail than that of  $R$ . The probability function for  $S$  is given by

$$P(S = k) = P(S_{n+1} = k - 1 \mid S_n = k)/k, \quad k \in \mathbb{N},$$

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 of [15]). Thus,

$$P(S = k) = E[e^{-kv} (kv)^{k-1}/k!], \quad k \in \mathbb{N}. \tag{6}$$

### 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

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

where  $\alpha$ ,  $0 < \alpha < 1$  and  $\beta > 0$  are parameters. Here,  $1/\beta$  is a scale parameter for both the distribution of  $R$  and the distribution of  $L$ .

The left-hand panel of Figure 2 shows perfect simulations of this process on  $[0, 10]$  when  $\mu(t) = 1$  is constant,  $\alpha = 0.9$ , and  $\beta = 10, 5, 2, 1$ . By (3), we expect to see about 10 clusters (in total) and 100 events. The clusters of course become more visible as  $\beta$  increases.

The left-hand panel of Figure 3 shows six simulations of clusters with  $\alpha = 0.9$ . Here,  $\alpha$  is an inverse scaling parameter;  $\beta$  is irrelevant since, to obtain comparable results for this example and the following two examples, we have omitted showing the scale. 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 of  $L$  to vary widely.

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

$$\gamma(t, Z) = \frac{\alpha \mathbf{1}(t \leq Z)}{E Z},$$

where  $\alpha$ ,  $0 < \alpha < 1$ , is a parameter,  $Z$  is a positive random variable, and  $\mathbf{1}(\cdot)$  denotes the indicator function. Then  $X$  can be viewed as a birth–death process with birth at time  $t_i$  and survival time  $Z_i$  for the  $i$ th individual. The birth rate is

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

(cf. (1)). Moreover,

$$v = \frac{\alpha Z}{E Z}, \quad \bar{v} = \alpha, \quad h(t) = E \left[ \frac{\mathbf{1}(t \leq Z)}{Z} \right], \quad \bar{h}(t) = \frac{P(Z \geq t)}{E Z}.$$

Since  $v$  is random, the distribution of  $S$  is more dispersed than in the unmarked case (cf. (6)).

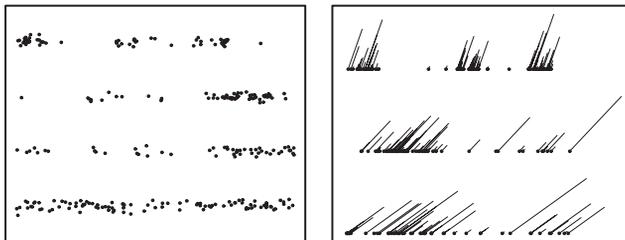


FIGURE 2: On the left, we display four perfect simulations on  $[0, 10]$  of the unmarked Hawkes process (Example 1) with parameters  $\alpha = 0.9, \mu = 1,$  and  $\beta = 10, 5, 2, 1$  (top to bottom). Random jitter has been added in the vertical direction to help distinguish events located close together. On the right, we display 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.

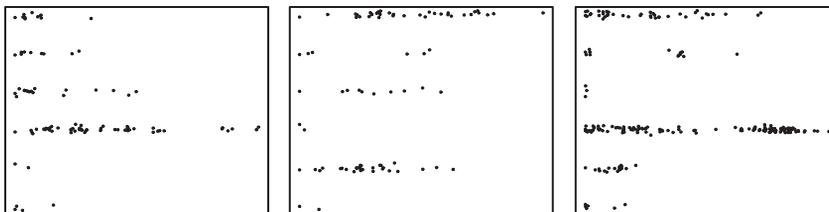


FIGURE 3: On the left, we display six simulations of clusters started at 0 and conditioned on  $S > 1$  in the unmarked case with  $\alpha = 0.9$ . In the centre, we display the same simulations, in the birth–death case, and, on the right, in the heavy-tailed case. Different scalings are used in the three cases.

The special case in which  $\mu(t) = \mu$  is constant and  $Z$  is exponentially distributed with mean  $1/\beta$  was considered in [5, p. 136]. In this case,  $X$  is a time-homogeneous Markov birth–death process with birth rate  $\mu + \alpha\beta n$  and death rate  $\beta n$ , where  $n$  is the number of living individuals. Furthermore,  $\bar{h}(t) = \beta e^{-\beta t}$  and  $h(t) = \beta E_1(\beta t)$ , where

$$E_1(s) = \int_s^\infty \frac{e^{-t}}{t} dt$$

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, below, the stationary distribution (i.e. the distribution of  $X$  at any fixed time) is known up to a constant of proportionality, and it is possible to simulate from this by rejection sampling.

The right-hand panel of Figure 2 shows three perfect simulations on  $[0, 10]$  in the Markov case with  $\mu = 1, \alpha = 0.9,$  and  $\beta = 5, 2, 1,$  where the marks are indicated by line segments of different lengths. The centre panel of Figure 3 shows six simulations of clusters (with marks excluded) with  $\alpha = 0.9,$  simulated conditional on  $S > 1$ . These simulations indicate that  $L$  is slightly more dispersed than in Example 1, since the marks introduce additional variation in the cluster lengths. In fact, the coefficient of variation estimated from 10 000 perfect simulations is 1.92 in Example 1 and 2.85 in the present case.

**Example 3.** (*Heavy-tailed distribution for  $L$ .*) Suppose that  $\gamma(t, Z) = \alpha Z e^{-tZ}$ , where  $\alpha \in (0, 1)$  is a parameter and  $Z$  is exponentially distributed with mean  $1/\beta$ . Then  $\bar{v} = v = \alpha$  is constant, meaning that the distribution of  $S$  is the same as in the unmarked case (cf. (6)). Furthermore,

$$h(t) = \bar{h}(t) = \frac{\beta}{(t + \beta)^2}$$

specifies a Pareto density. This is a heavy-tailed distribution, as it has infinite Laplace transform ( $\mathcal{L}(\theta) = E e^{\theta R} = \infty$  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  $L$ .

The right-hand panel of Figure 3 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 CDF for the length of a cluster) is known, the following algorithm for perfect simulation of the marked Hawkes process is similar to the algorithm for 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 the event that  $\{(t_j, Z_j) \in C_i : t_j \in [0, \infty)\} \neq \emptyset$ .
5. The output is all marked points from steps 1, 2, and 4.

**Remark 1.** In steps 1 and 2 of Algorithm 2, we use Algorithm 1 (with  $t_- = 0$ ). 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 knowledge of  $F$ , which is unknown in closed form (see Remark 3, below). 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

$$\int_{-\infty}^0 (1 - F(-t))\mu(t) dt < \infty. \tag{7}$$

In the case that  $\mu(t)$  is bounded, (7) is satisfied if  $\sup_{t \geq 0} \mu(t) E L < \infty$ . A condition for the finiteness of  $E L$  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  $I_1$  consists of 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  $I_2$  consists of 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 steps 2 and 4 of Algorithm 2 when we only want to sample those marked points  $(t_j, Z_j)$  with  $t_j \in [0, t_+)$ . Thus, Algorithm 2 produces realizations from the distribution of the marked Hawkes process.

Using the notation of Section 2.2, the following lemma generalizes and sharpens a result of [14] about the mean length of a cluster.

**Lemma 1.** *We have*

$$\frac{1}{E e^{-\nu}} E[(1 - e^{-\nu}) E[R \mid Z]] \leq E L \leq \frac{\bar{\nu}}{1 - \bar{\nu}} E \bar{R}. \tag{8}$$

*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 the event  $Z = z$ , then

$$E L = E[E[L \mid Z]] = E\left[\sum_{i=1}^{\infty} \frac{e^{-\nu} \nu^i}{i!} E[\max\{R_{j,Z} + L_j : j = 1, \dots, i\}]\right]. \tag{9}$$

To obtain the upper inequality, observe that

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

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

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

which verifies the upper inequality. Finally, by (9),

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

which reduces to the lower inequality.

**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, (8) reduces to

$$\left(\frac{1}{E e^{-\nu}} - 1\right) E R \leq E L \leq \frac{\bar{\nu}}{1 - \bar{\nu}} E R.$$

Consequently,  $E L < \infty$  if and only if  $E R < \infty$ . This immediately shows that  $E L < \infty$  in Example 1 and  $E L = \infty$  in Example 3. In Example 2, when  $Z$  is exponentially distributed with mean  $1/\beta$ , (8) becomes

$$\frac{\alpha(\alpha + 2)}{2(\alpha + 1)\beta} \leq E L \leq \frac{\alpha}{\beta(1 - \alpha)},$$

so in this case  $E L < \infty$ . Not surprisingly, apart from for small values of  $\alpha \in (0, 1)$ , the bounds are rather poor and of little use except in establishing the finiteness of  $E L$ .

#### 4. The distribution for 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 of Algorithm 2; however, many of the results are also of independent interest.

##### 4.1. An integral equation for $F$

Below, in Proposition 2, an integral equation for  $F$  is derived, and we discuss how to approximate  $F$  by numerical methods, using a certain recursion. Proposition 2 is a generalization of Theorem 5 of [14], which was proved using void probabilities obtained from a general result for the probability-generating functional for an unmarked Hawkes process. However, as was pointed out in [8], 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 CDF for the length of a cluster when all events of generations  $n + 1, n + 2, \dots$  are removed (it will become clear in Section 4.2 why we use the notation ‘ $1_n$ ’). Clearly,  $1_n$  is decreasing in  $n$ ,  $1_n \rightarrow F$  pointwise as  $n \rightarrow \infty$ , and

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

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

$$\varphi(f)(t) = E \left[ \exp \left( -\nu + \int_0^t f(t-s)\gamma(s, Z) ds \right) \right], \quad t \geq 0. \tag{10}$$

**Proposition 2.** *We have*

$$1_n = \varphi(1_{n-1}), \quad n \in \mathbb{N}, \tag{11}$$

and

$$F = \varphi(F). \tag{12}$$

*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 a fixed  $t \geq 0$  and  $n \in \mathbb{N}$ , split  $\Phi(0)$  into three point processes  $\Phi_1, \Phi_2$ , and  $\Phi_3$ : the process  $\Phi_1$  consists of those first generation offspring  $t_i \in \Phi(0) \cap [0, t)$  that 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  $\Phi_3 = \Phi(0) \cap [t, \infty)$  consists of the first generation offspring on  $[t, \infty)$ . Conditional on  $Z$ ,  $\Phi_1$ ,  $\Phi_2$ , and  $\Phi_3$  are independent Poisson processes with intensity functions  $\lambda_1(s) = \gamma(s, Z)1_{n-1}(t - s)$  on  $[0, t]$ ,  $\lambda_2(s) = \gamma(s, Z)(1 - 1_{n-1}(t - s))$  on  $[0, t]$ , and  $\lambda_3(s) = \gamma(s, Z)$  on  $[t, \infty)$ , respectively. This follows by an independent thinning argument since, conditional on  $\mathcal{G}_n$  (the  $n$ th generation of offspring in  $C_0$ ), the processes  $C_j - t_j$  with  $t_j \in \mathcal{G}_n$  are i.i.d. and distributed as  $C_0$  (this is the self-similarity property from Section 2.1). Consequently,

$$1_n(t) = E[P(\Phi_2 = \emptyset \mid Z) P(\Phi_3 = \emptyset \mid Z)] \\ = E \exp\left(-\int_0^t \lambda_2(s, Z) ds - \int_t^\infty \lambda_3(s, Z) ds\right),$$

which reduces to (11). Taking the limit as  $n \rightarrow \infty$  on both sides of (11), we obtain (12) 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 (12) to obtain a closed form expression for  $F$  even for simple choices of  $\gamma$ . Fortunately, the recursion (11) provides a useful numerical approximation to  $F$ . As the integral in (10), 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 (12) 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 approximations of  $F$  other than  $1_n$  exist, and their rates of convergence may be geometric with respect to different norms. Notice that certain monotonicity properties are fulfilled by  $\varphi$ , where, for functions  $f: [0, \infty) \rightarrow [0, 1]$ , we recursively define  $\varphi^{[0]}(f) = f$  and  $\varphi^{[n]}(f) = \varphi(\varphi^{[n-1]}(f))$ ,  $n = 1, 2, \dots$ , and set  $f_n = \varphi^{[n]}(f)$ ,  $n = 0, 1, \dots$ . Note that  $F_n = F$  for all  $n \in \mathbb{N}_0$ . As  $1_n = \varphi^{[n]}(1)$  is decreasing towards the CDF  $F$ , cases in which  $G$  is a CDF and  $G_n$  increases to  $F$  are of particular interest.

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

$$f \leq g \Rightarrow f_n \leq g_n, \quad n \in \mathbb{N}, \tag{13}$$

$$f \leq \varphi(f) \Rightarrow f_n \text{ is nondecreasing in } n, \tag{14}$$

$$f \geq \varphi(f) \Rightarrow f_n \text{ is nonincreasing in } n. \tag{15}$$

*Proof.* Equation (13) follows immediately from (10) when  $n = 1$ , and then by induction in the remaining cases. Equations (14) and (15) follow from (13).

**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  $f_n, g_n \in \mathcal{C}$  and*

$$\|\varphi(f) - \varphi(g)\|_\infty \leq \bar{\nu} \|f - g\|_\infty. \tag{16}$$

Furthermore,  $F$  is the unique fixpoint, i.e.

$$\|F - f_n\|_\infty \rightarrow 0 \text{ as } n \rightarrow \infty \tag{17}$$

and

$$\|F - f_n\|_\infty \leq \frac{\bar{\nu}^n}{1 - \bar{\nu}} \|\varphi(f) - f\|_\infty, \tag{18}$$

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 or, respectively, above.

*Proof.* Let  $f, g \in \mathcal{C}$ . Recall that, by the mean value theorem (e.g. Theorem 5.11 of [1]), for any real numbers  $x$  and  $y$ , we have  $e^x - e^y = (x - y)e^{z(x,y)}$ , where  $z(x, y)$  is a real number between  $x$  and  $y$ . Thus, by (10),

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

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

$$\begin{aligned} \|\varphi(f) - \varphi(g)\|_\infty &\leq \sup_{t \geq 0} \left| \mathbb{E} \left[ \int_0^t (f(t-s) - g(t-s)) \gamma(s, Z) ds \right] \right| \\ &\leq \mathbb{E} \left[ \int_0^\infty \|f - g\|_\infty \gamma(s, Z) ds \right] \\ &= \bar{\nu} \|f - g\|_\infty. \end{aligned}$$

Thereby, (16) is verified. Since  $\mathcal{C}$  is complete (see, e.g. Theorem 3.11 of [26]), it follows, from the fixpoint theorem for contractions (see, e.g. Theorem 4.48 of [1]), that the contraction has a unique fixpoint: by (12), this is  $F$ .

Since  $f \in \mathcal{C}$  implies that  $\varphi(f) \in \mathcal{C}$ , we find that  $f_n \in \mathcal{C}$  by induction. Hence, using (12), (16), and induction, we have

$$\|f_n - F\|_\infty = \|\varphi(f_{n-1}) - \varphi(F)\|_\infty \leq \bar{\nu} \|f_{n-1} - F\|_\infty \leq \bar{\nu}^n \|f - F\|_\infty \tag{19}$$

for  $n \in \mathbb{N}$ . Since  $\bar{\nu} < 1$ , we recover (17).

Similarly to (19), we have

$$\|f_n - f_{n-1}\|_\infty \leq \bar{\nu}^{n-1} \|f_1 - f\|_\infty, \quad n \in \mathbb{N}. \tag{20}$$

Furthermore, by (17), we have

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

So, by the triangle inequality and (20), we have

$$\begin{aligned} \|F - f\|_\infty &\leq \lim_{m \rightarrow \infty} (\|f_1 - f\|_\infty + \|f_2 - f_1\|_\infty + \dots + \|f_m - f_{m-1}\|_\infty) \\ &\leq \lim_{m \rightarrow \infty} \|f_1 - f\|_\infty (1 + \bar{\nu} + \dots + \bar{\nu}^{m-1}) \\ &= \frac{\|f_1 - f\|_\infty}{1 - \bar{\nu}} \end{aligned}$$

(see (2)). Combining this with (19), we obtain (18). Finally, if  $f \leq \varphi(f)$  or  $f \geq \varphi(f)$  then by (14) or, respectively, (15) and (17),  $f_n$  converges from below or, respectively, above.

Similar results to those of Theorem 1, but for the  $L^1$ -norm, were established in [18]. The following remark and proposition show how to find upper and lower bounds on  $F$  in many cases.

**Remark 4.** Consider a function  $f \in \mathcal{C}$ . The conditions  $f \leq \varphi(f)$  and  $f \geq \varphi(f)$  are satisfied in the extreme cases  $f = 0$  and  $f = 1$ , respectively. The upper bound  $f = 1$  is useful in the following sections, but the lower bound  $f = 0$  is too small a function for our purposes; if we require that  $E L < \infty$  (cf. Remark 1) then  $f = 0$  cannot be used (in fact we use only  $f = 0$  when producing the right-hand plot in Figure 4, below). To obtain a more useful lower bound, observe that  $f \leq \varphi(f)$  implies  $f \leq F < 1$  (cf. (4) and Theorem 1). If  $f < 1$  then a sufficient condition for  $f \leq \varphi(f)$  is

$$\frac{1}{\bar{v}} \geq \frac{\int_0^t (1 - f(t - s))\bar{h}(s) ds + \int_t^\infty \bar{h}(s) ds}{1 - f(t)}, \quad t \geq 0. \tag{21}$$

This follows readily from (5) and (10), using the fact that  $e^x \geq 1 + x$ .

The function  $f$  in (21) is closest to  $F$  when  $f$  is a CDF  $G$  and we have equality in (21). Equivalently,  $G$  satisfies the renewal equation

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

which has the unique solution

$$G(t) = 1 - \bar{v} + \sum_{n=1}^\infty (1 - \bar{v})\bar{v}^n \int_0^t \bar{h}^{*n}(s) ds, \quad t \geq 0, \tag{22}$$

where  $*n$  denotes  $n$ -times convolution (see Theorem IV2.4 of [2]). In other words,  $G$  is the CDF of  $\bar{R}_1 + \dots + \bar{R}_K$  (setting  $\bar{R}_1 + \dots + \bar{R}_K = 0$  if  $K = 0$ ), where  $K, \bar{R}_1, \bar{R}_2, \dots$  are independent random variables, each  $\bar{R}_i$  has density  $\bar{h}$ , and  $K$  has a geometric density  $(1 - \bar{v})\bar{v}^n$ . Interestingly, this geometric density is equal to  $E S_n / E S$  (see (3)).

The next proposition shows that, in many situations,  $G \leq \varphi(G)$  when  $G$  is an exponential CDF with a sufficiently large mean. In such cases,  $F$  has no heavier tails than such an exponential distribution.

Denote by

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

the Laplace transform of  $\bar{h}$ .

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

*Proof.* Upon inserting  $f = G$  into the right-hand side of (21), 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  $t > 0$ , (21) is satisfied if and only if  $\mathcal{L}(\theta) \leq 1/\bar{v}$ .

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  $\theta > 0$ . The condition  $\mathcal{L}(\theta) \leq 1/\bar{v}$  is equivalent to the tail-lightness condition [5, Equation (2.1)].

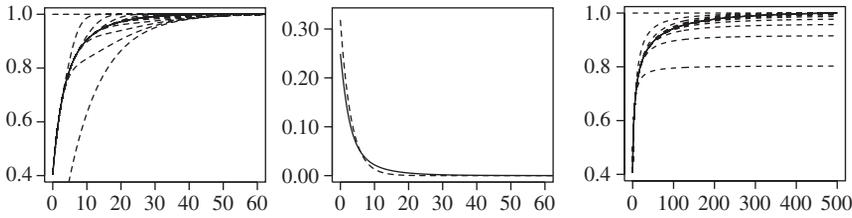


FIGURE 4: On the left, we display plots of  $1_n$  and  $G_n$  for  $n = 0, 5, \dots, 50$  in the unmarked case with  $\alpha = 0.9$  and  $\beta = 1$  (see Example 5);  $1_{50}$  and  $G_{50}$  are drawn solid to illustrate the approximate form of  $F$ , whereas the other curves are dashed. In the centre, we display plots of the density  $\frac{1}{2}[1'_n/(1 - 1_n(0)) + G'_n/(1 - G_n(0))]$  when  $n = 50$  (solid) and the exponential density with the same mean (dashed). On the right, we display the same plots as on the left, for Example 7 with  $\alpha = 0.9$  and  $\beta = 1$ , using  $1_n$  and  $0_n$  as approximations of  $F$ .

### 4.3. Examples

In Examples 5 and 6 below, we let

$$G(t) = 1 - e^{-\theta t}, \quad t \geq 0, \tag{23}$$

be the exponential CDF 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’,  $\theta = \mathcal{L}^{-1}(1/\bar{v}) = \beta(1 - \alpha)$ , (23) becomes the CDF for  $R$  times  $ES$ , which is easily seen to be the same as the CDF in (22).

The left-hand panel of Figure 4 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 CDF and  $G_{n+1} \geq G_n$ , we find that  $G_n$  is also a CDF. The centre panel of Figure 4 shows the density  $F'(t)/(1 - F(0))$  ( $t > 0$ ) approximated by

$$\frac{1}{2} \left( \frac{1'_n(t)}{1 - 1_n(0)} + \frac{G'_n(t)}{1 - G_n(0)} \right)$$

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.

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

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

where  $\mathcal{L}_Z(\theta) = \mathbb{E} e^{\theta Z}$  is the Laplace transform of  $Z$ . 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

$$\frac{1}{2} \left( \frac{1'_n}{1 - 1_n(0)} + \frac{G'_n}{1 - G_n(0)} \right)$$

for  $n = 0, 5, \dots, 50$  and  $(\alpha, \beta) = (0.9, 1)$  are similar to those in the centre and right-hand panels of Figure 4, and are therefore omitted.

**Example 7.** (*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 CDF in (22) is not known in 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, log-normal, or log-gamma distributions). Nonetheless, it is still possible to get an idea of what  $F$  looks like: the right-hand panel of Figure 4 shows  $1_n$  and  $0_n := \varphi^{[n]}(0)$  for  $n = 0, 5, \dots, 50$  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 by the plots and verified in [18],  $\lim_{t \rightarrow 0} G_n(t) < 1$  when  $G = 0$ , meaning that  $G_n$  is not a CDF.

### 5. Simulation of $I_2$

To complete the perfect simulation algorithm (Algorithm 2), 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 the spirit of the dominated-coupling-from-the-past algorithm [16].

Suppose that  $f \in \mathcal{C}$  is in closed form, with  $f \leq \varphi(f)$ , and that (7) is satisfied when we replace  $F$  by  $f$  (situations in which these requirements are fulfilled are considered in Sections 3, 4.2, and 4.3). Particularly, if  $\mu$  is constant and  $f$  is a CDF, (7) implies 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

$$\lambda_n^u(t) = (1 - f_n(-t))\mu(t) \quad \text{and} \quad \lambda_n^l(t) = (1 - 1_n(-t))\mu(t),$$

respectively. By Theorem 1,  $\lambda_n^u$  is nonincreasing and  $\lambda_n^l$  is nondecreasing 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/funnelling property (see [16]):

$$\emptyset = L_0 \subseteq L_1 \subseteq L_2 \subseteq \dots \subseteq I_2 \subseteq \dots \subseteq U_2 \subseteq U_1 \subseteq U_0. \tag{24}$$

The details are given by the following algorithm.

**Algorithm 3.** (*Simulation of  $I_2$ .*)

1. Generate a realization  $\{(t_1, Z_1), \dots, (t_k, Z_k)\}$  of  $U_0$ , where  $t_1 < \dots < t_k$ .
2. If  $U_0 = \emptyset$  then return  $I_2 = \emptyset$  and stop; otherwise, generate independent uniform numbers  $W_1, \dots, W_k$  on  $[0, 1]$  (independently of  $U_0$ ) and set  $n = 1$ .
3. For  $j = 1, \dots, k$ , assign  $(t_j, Z_j)$  to  $L_n$  or  $U_n$  if  $W_j \lambda_0^u(t_j) \leq \lambda_n^l(t_j)$  or, respectively,  $W_j \lambda_0^u(t_j) \leq \lambda_n^u(t_j)$ .
4. If  $U_n = L_n$  then return  $I_2 = L_n$  and stop; otherwise, 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 that, regardless of whether  $U_0 = \emptyset$  in step 2 or  $U_n = L_n$  in step 4, we continue to generate  $(U_1, L_1), (U_2, L_2)$ , etc. Furthermore, add an extra step: for  $j = 1, \dots, k$ , assign  $(t_j, Z_j)$  to  $I_2$  if and only if

$$W_j \lambda_n^u(t_j) \leq \lambda_2(t_j).$$

Then clearly, because of the convergence properties of  $\lambda_n^u$  and  $\lambda_n^l$  (see the discussion above), (24) is satisfied and, conditional on  $t_1, \dots, t_k$ ,

$$\begin{aligned} \mathbb{P}(L_n \neq U_n \text{ for all } n \in \mathbb{N}_0) &\leq \sum_{j=1}^k \lim_{n \rightarrow \infty} \mathbb{P}(W_j \lambda_0^u(t_j) \leq \lambda_n^u(t_j), W_j \lambda_0^u(t_j) > \lambda_n^l(t_j)) \\ &= \sum_{j=1}^k \mathbb{P}(\lambda_2(t_j) < W_j \lambda_0^u(t_j) \leq \lambda_2(t_j)) = 0. \end{aligned}$$

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

**Remark 5.** We compute  $1_n$  and  $f_n$  numerically, using a quadrature rule (see 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  $n = 0, 1, 2, \dots$  until Algorithm 3 terminates, there is no advantage in using a doubling scheme for  $n$ , as in the Propp–Wilson algorithm [24].

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

$$\sum_0^\infty \pi'_n = \sum_0^m \frac{\pi_n}{\pi_0} + \frac{a}{1-a} \frac{\pi_m}{\pi_0}.$$

Then, since  $\pi'_n \geq \pi_n/\pi_0$ , we can sample  $N$  from  $\pi_n$  by rejection sampling (see [25]). Furthermore, conditional on  $N = n$ , we generate  $n$  independent marks  $Z'_1, \dots, Z'_n$  that 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, it is a lot faster than our perfect simulation algorithm (roughly 1200 times as fast in the case  $\alpha = 0.9, \beta = \mu = 1$ , and  $t_+ = 10$ ), since it exploits the fact that we know the stationary distribution in this special case.

## 6. Extensions and open problems

Except in 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  $F$  (see Example 7).

For applications such as those in seismology [23], extensions of our results and algorithms to the heavy-tailed cases are important. The epidemic-type aftershock sequences (ETAS model) [22], used for modelling 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]), can be applied to 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 nonlinear Hawkes processes [3], [8] would also be interesting. However, things again become complicated, since a nonlinear 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. Let  $\Psi_1$  have unpredictable marks, with  $Z_i^1 \sim \text{Exp}(\beta)$ , and let  $\Psi_2$  have predictable marks, with  $Z_i^2 \sim \text{Exp}(\beta + 1/Z_A^2)$ , where  $Z_A^2$  is the mark of the first-order ancestor of  $t_i$ . Both models have  $\gamma(t, Z) = \alpha\beta \mathbf{1}(t < Z)$ , with the same  $\alpha$  and  $\beta$ , and they also have the same  $\mu(t)$ . The model  $\Psi_2$  has the intuitive appeal that long-lived individuals have long-lived offspring. Note that the intensity of  $\Psi_1$  dominates the intensity of  $\Psi_2$  if the marks are simulated such that  $Z_i^1 > Z_i^2$ .

To simulate  $\Psi_2$ , we first simulate  $\Psi_1$  using Algorithm 3, with the modifications that we associate both marks  $Z_i^1$  and  $Z_i^2$  to the event  $t_i$ , and we keep all events from the algorithm whether they fall in or before  $[0, t_+)$ . Each marked event  $(t_j, Z_j^1)$  is then included in  $\Psi_2$  with retention probability

$$\frac{\mu(t) + \alpha\beta \sum_{t_i < t_j, t_i \in \Psi_2} \mathbf{1}(t_j - t_i < Z_i^2)}{\mu(t) + \alpha\beta \sum_{t_i < t_j} \mathbf{1}(t_j - t_i < Z_i^1)},$$

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 correct process  $\Psi_2$ .

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 nontrivial (i.e. not almost surely empty), it is necessary that  $\bar{\nu} = 1$ , which means that any dominating Hawkes process has  $\bar{\nu} \geq 1$  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 that  $n_i$  is still conditionally

independent of  $t_i$  and the previous history. A particularly simple case occurs when  $n_i$  is either 1 or 0, and

$$\bar{p} = E[P(n_i = 1 \mid Z_i)]$$

is assumed to be strictly between 0 and 1 (here  $\bar{p}$  plays a role similar to that of  $\bar{v}$ , introduced in Section 4). We then redefine  $\varphi$  by

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

where, now,

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

Since  $\varphi$  is now linear, the situation is much simpler. For example,  $F$  is given by  $G$  in (22) (with  $\bar{v}$  replaced by  $\bar{p}$ ).

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 that it is feasible to simulate the immigrants on  $[t_-, t_+)$ . However, the perfect simulation algorithm relies heavily 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 (see [19] and [20]).

### Acknowledgements

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

- [1] APOSTOL, T. M. (1974). *Mathematical Analysis*. Addison-Wesley, Reading, MA.
- [2] ASMUSSEN, S. (1987). *Applied Probability and Queues*. John Wiley, Chichester.
- [3] BRÉMAUD, P. AND MASSOULIÉ, L. (1996). Stability of nonlinear Hawkes processes. *Ann. Probab.* **24**, 1563–1588.
- [4] BRÉMAUD, P. AND MASSOULIÉ, L. (2001). Hawkes branching point processes without ancestors. *J. Appl. Prob.* **38**, 122–135.
- [5] BRÉMAUD, P., NAPPO, G. AND TORRISI, G. (2002). Rate of convergence to equilibrium of marked Hawkes processes. *J. Appl. Prob.* **39**, 123–136.
- [6] BRIX, A. AND KENDALL, W. S. (2002). Simulation of cluster point processes without edge effects. *Adv. Appl. Prob.* **34**, 267–280.
- [7] CHORNOBOY, E. S., SCHRAMM, L. P. AND KARR, A. F. (1988). Maximum likelihood identification of neural point process systems. *Biol. Cybernet.* **59**, 265–275.
- [8] DALEY, D. J. AND VERE-JONES, D. (2003). *An Introduction to The Theory of Point Processes*, Vol. 1, *Elementary Theory and Methods*, 2nd edn. Springer, New York.
- [9] DWASS, M. (1969). The total progeny in a branching process and a related random walk. *J. Appl. Prob.* **6**, 682–686.
- [10] HAWKES, A. G. (1971). Point spectra of some mutually exciting point processes. *J. R. Statist. Soc. B* **33**, 438–443.
- [11] HAWKES, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58**, 83–90.
- [12] HAWKES, A. G. (1972). Spectra of some mutually exciting point processes with associated variables. In *Stochastic Point Processes*, ed. P. A. W. Lewis, John Wiley, New York, pp. 261–271.
- [13] HAWKES, A. G. AND ADAMOPOULOS, L. (1973). Cluster models for earthquakes – regional comparisons. *Bull. Internat. Statist. Inst.* **45**, 454–461.

- [14] HAWKES, A. G. AND OAKES, D. (1974). A cluster representation of a self-exciting process. *J. Appl. Prob.* **11**, 493–503.
- [15] JAGERS, P. (1975). *Branching Processes with Biological Applications*. John Wiley, London.
- [16] KENDALL, W. S. AND MØLLER, J. (2000). Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Adv. Appl. Prob.* **32**, 844–865.
- [17] MØLLER, J. (2003). Shot noise Cox processes. *Adv. Appl. Prob.* **35**, 614–640.
- [18] MØLLER, J. AND RASMUSSEN, J. G. (2005). Approximate simulation of Hawkes processes. Submitted.
- [19] MØLLER, J. AND TORRISI, G. L. (2005). Generalised shot noise Cox processes. *Adv. Appl. Prob.* **37**, 48–74.
- [20] MØLLER, J. AND TORRISI, G. L. (2005). Perfect and approximate simulation of spatial Hawkes processes. In preparation.
- [21] MØLLER, J. AND WAAGEPETERSEN, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall, Boca Raton, FL.
- [22] OGATA, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Amer. Statist. Assoc.* **83**, 9–27.
- [23] OGATA, Y. (1998). Space-time point-process models for earthquake occurrences. *Ann. Inst. Statist. Math.* **50**, 379–402.
- [24] PROPP, J. G. AND WILSON, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures Algorithms* **9**, 223–252.
- [25] RIPLEY, B. D. (1987). *Stochastic Simulation*. John Wiley, New York.
- [26] RUDIN, W. (1987). *Real and Complex Analysis*. McGraw-Hill, New York.
- [27] VERE-JONES, D. AND OZAKI, T. (1982). Some examples of statistical inference applied to earthquake data. *Ann. Inst. Statist. Math.* **34**, 189–207.