A review on perfect simulation in stochastic geometry

Jesper Møller
Department of Mathematical Sciences
University of Aalborg

ABSTRACT

We provide a review and a unified exposition of the Propp-Wilson algorithm and various other algorithms for making perfect simulations with a view to applications in stochastic geometry. Most examples of applications are for spatial point processes.

Key Words: coupling from the past (CFTP), ergodicity, falling leaves model, horizontal CFTP, locally stable point processes, noisy point processes, read-once algorithm, spatial birth and death processes, vertical CFTP, Widom-Rowlinson model.

1 Introduction

One of the most important and exciting recent developments in stochastic simulation is perfect simulation. Following the seminal work by Propp and Wilson (1996), many papers have proven that perfect simulation algorithms are particular useful in stochastic geometry, spatial statistics and statistical physics. It seems timely to review this development with a view to applications in stochastic geometry.

The aims of this paper are to provide such a review for readers with limited knowledge on perfect simulation, showing the mathematical details, and also to put things into a unified framework. From a mathematical view, the paper is self-contained, but in order to keep the paper within the limit of about 20 pages, no illustrative figures and empirical results are included (but the relevant references are provided). For the same reason I have chosen to focus on the Propp-Wilson algorithm, also called vertical coupling from the past (CFTP) in Section 3, and so-called horizontal CFTP in Section 4, also called dominated CFTP (Kendall and Møller, 2000) and coupling into and from the past (Wilson, 2000a). Section 2 provides some background material related to CFTP. Most examples of applications in Sections 2–4 are for finite spatial point processes. Other topics such as Fill's algorithm and extensions to infinite point processes are briefly discussed in Section 5.

2 CFTP and two examples

Proposition 1 below has some similarities with Theorem 3 in Propp and Wilson (1996) and Theorem 2.1 in Kendall and Møller (2000), but it is stated so that it applies for both the vertical and horizontal CFTP algorithms presented in Sections 3 and 4.

We consider a general setting with a given target distribution π defined on a state space E, and a discrete time process $\{X_t\}_{t=0}^{\infty}$ on E, which is defined by a so-called stochastic recursive sequence (SRS),

$$X_t = \varphi\left(X_{t-1}, R_t\right), \quad t = 1, 2, \dots. \tag{1}$$

Here the R_t are random variables and φ is a deterministic function, called the updating function. Under mild conditions, any discrete time homogeneous Markov chain can be represented as a SRS, where the R_t are IID (independent and identically distributed random variables); see Foss and Tweedie (1998) and the references therein. When making simulations, R_t is generated by a vector of pseudo-random numbers $V_t = (V_{t1}, \dots, V_{tN_t})$, where $N_t \in \mathbf{N}$ is either a constant or yet another pseudo-random number; see e.g. Example 2 below.

Moreover, we include negative times and let for any state $x \in E$ and times $s, t \in \mathbf{Z}$ with $s < t$,

$$X_s^t(x) = \varphi\Big(\cdots\varphi\Big(\varphi(x, R_{s+1}), R_{s+2}\Big)\cdots, R_t\Big)$$

denote the state of a process at time t when it is started in x at time s. In Section 4 we consider continuous-time jump processes $\tilde{X}_s(x) = \{\tilde{X}_s^t(x) : t \geq s\}$ with $\tilde{X}_s^s(x) = x$, and set $X_s^t(x) = \tilde{X}_s^{J_t(s,x)}$ for integers $s < t$, where $s \leq J_1(s,x) < J_2(s,x) < \dots$ is a random sequence containing the jump times of $\tilde{X}_s(x)$ (this sequence will only be needed for theoretical considerations — it is not used in the implementations of our perfect simulation algorithms). In order to unify the notation, for the discrete time setting used in Examples 1 and 2 and Section 3, we simply set $\tilde{X}_s^t(x) = X_s^t(x)$ for integers $s < t$. We require that π is the limiting distribution of $\tilde{X}_0^t(\hat{x})$, where in Section 4 we select a particular state $\hat{x}$, while in Examples 1 and 2 and in Section 3 an arbitrary state $\hat{x}$ can be chosen. Finally, we say that a random variable $T \in \mathbf{N}_0 \cup \{\infty\}$ is a stopping time with respect to $\mathcal{R}_- \equiv \{R_{-t}\}_{t=0}^{\infty}$, if for any $t \in \mathbf{N}_0$, the event $\{T \leq t\}$ is determined by $R_0, R_{-1}, \dots, R_{-t}$.

Proposition 1 *Assume that (i) the distribution of $\{R_t\}_{t \in \mathbf{Z}}$ is stationary in time, (ii) there exists a state $\hat{x} \in E$ so that $\tilde{X}_0^t(\hat{x})$ converges in distribution towards π as $t \to \infty$, and (iii) $T \geq 0$ is an almost surely finite stopping time with respect to $\mathcal{R}_-$ such that $X_{-t}^0(\hat{x}) = X_{-T}^0(\hat{x})$ whenever $t \geq T$. Then $X_{-T}^0(\hat{x}) \sim \pi$.*

Proof.

$$\mathsf{P}\left(X^0_{-T}(\hat{x}) \in F\right) = \mathsf{P}\left(\lim_{t\to\infty} \tilde{X}^0_{-t}(\hat{x}) \in F\right) = \lim_{t\to\infty} \mathsf{P}\left(\tilde{X}^0_{-t}(\hat{x}) \in F\right)$$

$$= \lim_{t\to\infty} \mathsf{P}\left(\tilde{X}^t_0(\hat{x}) \in F\right) = \pi(F)$$

where the monotone convergence theorem is used for obtaining the second equality and (i) is used for obtaining the third equality.

Generally speaking, by a CFTP algorithm we understand a way of determining a time $-t \leq -T$ and returning $X^0_{-t}(\hat{x}) \sim \pi$. Occasionally, T is referred to as the running time of the algorithm, though we should keep in mind that a more precise definition of "running time" may be appropriate in each specific algorithm. A CFTP algorithm is only perfect/exact in principle, as pseudo-random numbers are used in practice, and X^0_{-T} and T may be dependent, so long running times may cause a bias in the simulations, which are actually used. Still, for short, we call this perfect simulation.

Example 1: Falling leaves model Consider a closed set $S \subset \mathbf{R}^d$ and IID random closed sets $R_t \subset \mathbf{R}^d$, $t \in \mathbf{Z}$ so that with probability one, S will be covered by a finite number of the R_t's:

$$\mathsf{P}\Big(\exists t > 0 : S \subseteq R_1 \cup \ldots \cup R_t\Big) = 1.$$

Let E be the set of all closed subsets of S. Defining for $x \in E$ and closed subsets $R \subset \mathbf{R}^d$,

$$\varphi(x, R) = [(x \setminus R) \cup \partial R] \cap S,$$

where ∂R denotes the topological boundary of the set R, we obtain by (1) a Markov chain defined on E. If we think of the R_t as falling leaves on the ground (so $d = 2$) and there are no leaves at time $t = 0$, then $X^t_0(\emptyset)$ shows the boundaries of falling leaves when looking down at the region S at time $t > 0$. This model for falling leaves has been introduced by Matheron (1968,1975); see also Serra (1982), Jeulin (1997), Kendall and Thönnes (1999), and the illuminating applets at Wilfrid Kendall's homepage (http://www.warwick.ac.uk/statsdept/Staff/WSK/dead.html).

Since $X^t_0(x)$ does not depend on $x \in E$ whenever $t \geq \inf\{n > 0 : S \subseteq R_1 \cup \ldots \cup R_n\}$, the chain is easily seen to be uniformly ergodic (this is in fact verified in Section 3.1). In particular $X^t_0(x)$ has some limiting distribution π as $t \to \infty$, where π does not depend on the choice of $x \in E$. Though this may be a very complicated distribution, we can at least make perfect simulations from π: by Proposition 1, for any $\hat{x} \in E$, $X^0_{-T_{fl}}(\hat{x}) \sim \pi$ if we define

$$T_{fl} = \inf\left\{t \in \mathbf{N}_0 : S \subseteq R_0 \cup R_{-1} \cup \ldots \cup R_{-t}\right\}, \tag{2}$$

and $X^0_{-T_{fl}}(\hat{x})$ does not depend on $\hat{x}$.

Example 2: Widom-Rowlinson model Let $A \subset \mathbf{R}^d$ be a Borel set with positive and finite Lebesgue measure $|A|$. Denote the homogeneous Poisson point process defined on A and of rate $\beta > 0$ by Poisson(A, β). This means the following if $R \sim$ Poisson(A, β) is represented as a finite subset of A: the number $n(R)$ of points in R follows a Poisson distribution with mean $\beta|A|$; and conditionally on $n(R)$, the points in R are IID and each point follows a uniform distribution on A.

Now suppose that $X^{(i)} \sim$ Poisson(S, β_i), $i = 1, 2$, are independent, $S \subset \mathbf{R}^d$ is closed with $0 < |S| < \infty$, and

$$\pi = \mathcal{D}\left(X^{(1)}, X^{(2)} | \operatorname{dist}(X^{(1)}, X^{(2)}) > \delta\right) \tag{3}$$

where $\operatorname{dist}(X^{(1)}, X^{(2)})$ denotes the shortest distance between a point from $X^{(1)}$ and a point from $X^{(2)}$, and β_1, β_2, δ are positive parameters. This is the Widom and Rowlinson (1970) model, which possesses many interesting properties as discussed in Häggström, van Lieshout and Møller (1999) and the references therein. The support of π is the set of all $(x^{(1)}, x^{(2)})$ where $x^{(1)}$ and $x^{(2)}$ are finite subsets of S with $\operatorname{dist}(x^{(1)}, x^{(2)}) > \delta$. However, it becomes convenient in Section 3.2, if we define the state space E as the set of all $(x^{(1)}, x^{(2)})$ where $x^{(1)}$ and $x^{(2)}$ are closed subsets of S.

For the purpose of making simulations from π, it seems natural to use a two-component Gibbs sampler, since

$$\mathcal{D}\left(X^{(1)} | X^{(2)}, \operatorname{dist}(X^{(1)}, X^{(2)}) > \delta\right) = \text{Poisson}\Big(S \setminus U_{X^{(2)}}\,,\, \beta_1\Big),$$

$$\mathcal{D}\left(X^{(2)} | X^{(1)}, \operatorname{dist}(X^{(1)}, X^{(2)}) > \delta\right) = \text{Poisson}\Big(S \setminus U_{X^{(1)}}\,,\, \beta_2\Big),$$

where for $x \subset S$, U_x denotes the union of balls of radius δ and centers $\xi \in x$. Using a systematic updating scheme, the two-component Gibbs sampler amounts to use a SRS with $R_t = (R_t^{(1)}, R_t^{(2)})$, where $R_t^{(i)} \sim$ Poisson(S, β_i), $i = 1, 2$, $t \in \mathbf{Z}$, are independent and the updating function φ is defined as follows: for $x^{(1)}, x^{(2)}, r^{(1)}, r^{(2)} \subset S$ and

$$\left(y^{(1)}, y^{(2)}\right) = \varphi\left((x^{(1)}, x^{(2)}), (r^{(1)}, r^{(2)})\right), \tag{4}$$

we have that

$$y^{(1)} = r^{(1)} \setminus U_{x^{(2)}}\,, \quad y^{(2)} = r^{(2)} \setminus U_{y^{(1)}}\,.$$

Clearly this sampler preserves π and it is easily shown to be uniformly ergodic due to the following simple facts: if $r^{(1)} = \emptyset$ then $y^{(1)} = \emptyset$ does not depend on $x^{(2)}$; similarly, if $r^{(2)} = \emptyset$ then $y^{(2)} = \emptyset$ does not depend on $y^{(1)}$; and both cases can happen in each update of the Gibbs sampler, since $\mathsf{P}(R_t^{(i)} = \emptyset) =$

$\exp(-\beta_i|S|) > 0$. Moreover, after updating the first component, the sampler stays within the support of π, no matter the choice of the initial state in E.

Due to these properties, an obvious choice for the stopping time used in Proposition 1 would be

$$T_{WR} = \inf\left\{t \in \mathbf{N}_0 : [R_{-t}^{(1)} = \emptyset] \text{ or } [R_{-t}^{(2)} = \emptyset \text{ and } t < 0]\right\}. \tag{5}$$

Indeed, for any $\hat{x} \in E$, $X_{-T_{WR}}^0(\hat{x}) \sim \pi$ and $X_{-T_{WR}}^0(\hat{x})$ does not depend on $\hat{x}$. However, except for possibly rather uninteresting cases with a very small value of β_1 or β_2, T_{WR} is expected to be extremely large. A much more efficient coupling time is specified in Section 3.2.

3 Vertical CFTP

In Examples 1 and 2 we noticed that the chain was uniformly ergodic. As discussed in Section 3.1, this property is in a sense a sufficient and necessary property for applying the Propp-Wilson algorithm which, for reasons which soon will become clear, henceforth is called vertical CFTP. Furthermore, Section 3.2 shows how monotonicity properties of the SRS make vertical CFTP feasible in practice. Finally, Wilson's read-once algorithm is discussed in Section 3.3.

Throughout Sections 3.1–3.3 the R_t's are assumed to be IID.

3.1 Vertical CFTP and uniform ergodicity

Propp and Wilson (1996) consider what Foss and Tweedie (1998) call the smallest "vertical backward coupling time", that is the first time before 0 for coalesence of all possible chains; denoting this by $-T_{PW}$ we have

$$T_{PW} = \inf\left\{t \in \mathbf{N}_0 : X_{-t}^0(x) = X_{-t}^0(y) \text{ for all } x, y \in E\right\}. \tag{6}$$

For example, for the falling leaves model, $T_{PW} = T_{fl}$ in (2), while for the Widom-Rowlinson model, $T_{PW} \leq T_{WR}$ in (5). In fact, typically $T_{PW} \ll T_{WR}$ unless the rates β_i are very small (see Figure 3 in Häggström, van Lieshout and Møller, 1999).

When does Proposition 1 apply if $T = T_{PW}$? Condition (i) in Proposition 1 is clearly satisfied. By the definition (6), T_{PW} is a stopping time with respect to $\mathcal{R}_-$, and $X_{-t}^0(x)$ does not depend on $(x,t) \in E \times \mathbf{N}_0$ when $-t \leq T_{PW}$. Moreover, as shown below, if $T_{PW} < \infty$ almost surely, the chain becomes uniformly ergodic, and so condition (ii) in Proposition 1 is satisfied. Hence, in order to verify the conditions of Proposition 1, we need only to verify that $\mathsf{P}(T_{PW} < \infty) = 1$:

Proposition 2 *If* $\mathsf{P}(T_{PW} < \infty) = 1$, *then* $X^0_{-t}(\hat{x}) \sim \pi$ *for any* $\hat{x} \in E$ *and* $t \geq T_{PW}$.

Indeed, as noticed in Propp and Wilson (1996), if the support of π is finite, then irreducibility implies that $\mathsf{P}(T_{PW} < \infty) = 1$. But in applications of stochastic geometry, the support of π is rarely finite as illustrated in Examples 1 and 2. However, we can often find another backwards stopping time T so that $T_{PW} \leq T$, where T is easily shown to be almost surely finite. One such example is $T = T_{WR}$ given by (5). Then by Proposition 2, we can make a perfect simulation from π, provided there is a feasible way for determining T_{PW} when an algorithm for vertical CFTP is implemented — see Section 3.2.

Now, how does the condition in Proposition 2 relate to uniform ergodicity? Recall that uniform ergodicity of a discrete time homogeneous Markov chain is equivalent to the existence of a time $n > 0$, a number $\epsilon > 0$, and a probability measure Q such that

$$\mathsf{P}\Big(X_0^n(x) \in \cdot\Big) \geq \epsilon Q(\cdot) \tag{7}$$

(Meyn and Tweedie, 1993, Theorem 16.0.2). Note that we in (7) do not require that the chain is defined in terms of a SRS, though we are still using the notation $X_0^n(x)$ for the state at time n when starting in x at time 0. Observe also that by time homogeneity,

$$T_{for} \equiv \inf\Big\{t \in \mathbf{N}_0 : X_0^t(x) = X_0^t(y) \text{ for all } x, y \in E\Big\} \sim T_{PW}. \tag{8}$$

Hence $T_{PW} < \infty$ almost surely if and only if there is a time $t \in \mathbf{N}$ so that $\mathsf{P}(C_t) > 0$ where $C_t = \{X_0^t(x) = X_0^t(y) \text{ for all } x, y \in E\}$ (the "only if part" follows immediately from (8), while the "if part" follows by considering the independent and equiprobable events $C_{it} = \{X^{it}_{(i-1)t}(x) = X^{it}_{(i-1)t}(y) \text{ for all } x, y \in E\}$, $i = 1, 2, \ldots$; in this section we use only the "if part", while the equivalence is used in Section 3.3).

Consequently, on one hand, if $T_{PW} < \infty$ almost surely, the chain is uniformly ergodic: set $n = t+1$, $\epsilon = \mathsf{P}(C_t)$, and $Q(F) = \mathsf{P}\left(\varphi(Z, R_{t+1}) \in F \mid C_t\right)$ where Z denotes the common value of $X_0^t(x)$, $x \in E$, when the event C_t happens to occur. This shows the limitations of the applicability of vertical CFTP: it never works if we don't have uniform ergodicity. On the other hand, it can be shown that uniform ergodicity implies the existence of a SRS construction so that $T_{PW} < \infty$ almost surely (one such construction is provided by Nummelin's splitting technique if $n = 1$ in (7)); we refer to Foss and Tweedie (1998) for further details. Of course this result is purely theoretical as there exist infinitely many possible SRS's, and the art in practice is to pick one so that the chains coalesce quickly.

3.2 Monotonicity and anti-monotonicity

One concern with for example the case of the SRS construction used in Example 2 is that apparently there are uncountably many paths to take care of if we wish to determine T_{PW}. In contrast only one path is needed in Example 1!

However, as observed in Propp and Wilson (1996), this problem may be overcome if there is a partial ordering $\prec$ on E such that the updating function is monotone in its first argument,

$$\varphi(x,\cdot) \prec \varphi(y,\cdot) \quad \text{whenever } x \prec y, \tag{9}$$

and if there exist a unique minimum $\hat{0} \in E$ and a unique maximum $\hat{1} \in E$,

$$\forall x \in E: \; \hat{0} \prec x \prec \hat{1}.$$

Thereby T_{PW} is determined by only two paths "started in the infinite past". More precisely,

$$T_{PW} = \inf \left\{ n \in \mathbf{N}_0 : L^0_{-n} = U^0_{-n} \right\}, \tag{10}$$

where $L_{-n} = \{X^{-t}_{-n}(\hat{0}) : t = n, n-1, \ldots\}$ and $U_{-n} = \{X^{-t}_{-n}(\hat{1}) : t = n, n-1, \ldots\}$ are the lower and upper chains started at the minimal respective maximal state at time $-n$: set $L^{-n}_{-n} = \hat{0}$, $U^{-n}_{-n} = \hat{1}$, and

$$L^{-t}_{-n} = \varphi\left(L^{-t-1}_{-n}, R_{-t}\right), \; U^{-t}_{-n} = \varphi\left(U^{-t-1}_{-n}, R_{-t}\right), \; t = n-1, n-2, \ldots. \tag{11}$$

If the monotonicity property (9) is replaced by the anti-monotonicity property,

$$\varphi(y,\cdot) \prec \varphi(x,\cdot) \quad \text{whenever } x \prec y,$$

then using the cross-over trick introduced in Kendall (1998) we redefine

$$L^{-t}_{-n} = \varphi\left(U^{-t-1}_{-n}, R_{-t}\right), \; U^{-t}_{-n} = \varphi\left(L^{-t-1}_{-n}, R_{-t}\right), \; t = n-1, n-2, \ldots, \tag{12}$$

whereby (10) remains to be true. Notice that in both the monotone and the anti-monotone case we have the following sandwiching property,

$$L^{-t}_{-n} \prec X^{-t}_{-n}(x) \prec U^{-t}_{-n} \quad \text{for } t = n, n-1, \ldots \text{ and } n = 0, 1, \ldots, \tag{13}$$

and the funnelling property,

$$L^{-t}_{-n} \prec L^{-t}_{-m} \prec U^{-t}_{-m} \prec U^{-t}_{-n} \quad \text{for integers } m \geq n \geq t \text{ with } n \geq 0. \tag{14}$$

So putting things together with Proposition 2, we obtain the following proposition.

Proposition 3 *Assume that* $\mathsf{P}(T_{PW} < \infty) = 1$ *and there is a partial ordering with unique minimal and maximal states. Then in both the monotone and anti-monotone case,* $L^0_{-n} = X^0_{-T_{PW}} \sim \pi$ *whenever* $L^0_{-n} = U^0_{-n}$.

Thus, when the conditions of Proposition 3 hold, it becomes much simpler to determine T_{PW} and to implement a vertical CFTP algorithm: generate (L_{-n}, U_{-n}) further and further back in time for any strictly increasing sequence of n's, and then return $L^0_{-n} \sim \pi$ as soon as $L^0_{-n} = U^0_{-n}$. As argued in Propp and Wilson (1996), instead of using the sequence $n = 1, 2, 3, 4, \ldots$, it may be more efficient to use a doubling scheme $n = 1, 2, 4, 8, \ldots$; see also the discussion in Wilson (2000b). Notice that by (11) and (12), in the vertical CFTP algorithm we need to store each R_{-t} which has been used for determining a pair of lower and upper processes.

Example 2 (continued) A natural partial ordering is given by set-inclusion with respect to the two types of points: for closed subsets $x^{(i)}, y^{(i)} \subseteq S$, define

$$\left(x^{(1)}, x^{(2)}\right) \prec \left(y^{(1)}, y^{(2)}\right) \quad \text{if } x^{(i)} \subseteq y^{(i)},\ i = 1, 2. \tag{15}$$

Then

$$\hat{0} = (\emptyset, \emptyset), \quad \hat{1} = (S, S),$$

are unique minima and maxima, and the SRS construction for the two-component Gibbs sampler is seen to be anti-monotone, so vertical CFTP easily applies. Note that $\hat{0}$ belongs to the support of π (it is even an atom), while $\hat{1}$ does not (it is at this point the definition of E becomes convenient). Moreover we can in an obvious way extend the definition (3) to a Widom-Rowlinson model of $k \geq 2$ components, and the updating function (4) to an anti-monotone k-component Gibbs sampler with respect to an obvious extension of (15), so vertical CFTP applies for this case also.

There is another partial ordering which makes the two-component Gibbs sampler monotone: suppose now that

$$\left(x^{(1)}, x^{(2)}\right) \prec \left(y^{(1)}, y^{(2)}\right) \quad \text{if } x^{(1)} \subseteq y^{(1)},\ y^{(2)} \subseteq x^{(2)}, \tag{16}$$

in which case

$$\hat{0} = (\emptyset, S), \quad \hat{1} = (S, \emptyset), \tag{17}$$

are unique minima and maxima. Note that now neither $\hat{0}$ nor $\hat{1}$ is in the support of π, and (16) does not extend to the case of $k > 2$ components.

For a further discussion of perfect simulation using the monotone version above (but with (17) replaced by "quasi-minimal and quasi-maximal" states), including empirical results for the Widom-Rowlinson model, see Häggström, van Lieshout and Møller (1999).

3.3 Read-once algorithm

A natural question is if we instead of going backwards in time could generate perfect simulations by running forwards in time. For example, by (8), the "vertical forwards coupling time" T_{for} is distributed as the vertical backwards coupling time T_{PW}. However, in general $X_0^{T_{for}} \not\sim \pi$; a counterexample is provided by a random walk on $\{1, 2, 3\}$ defined by the SRS $\varphi(x, R_t) = x + R_t$, where the R_t are IID and uniformly distributed on $\{\pm 1\}$, and where we truncate $x + R_t$ at 1 and 3. Moreover, we have noticed in Section 3.2 the need of storing the R_{-t} which are reused in the lower and upper processes. Below we describe Wilson's (2000a) read-once algorithm, which runs forward in time, starting at time 0, and reading the R_t only once. As we shall see, it works whenever vertical CFTP does, and it can then be naturally used for producing IID samples from π.

For $m \in \mathbf{N}$ and $i \in \mathbf{Z}$, set $F_i(x) = X_{(i-1)m}^{im}(x)$. Then the "random maps" F_i, $i \in \mathbf{Z}$, are IID. As verified in Section 3.1, $T_{PW} < \infty$ almost surely if and only if $p \equiv \mathsf{P}(F_0 \text{ is a singleton})$ is strictly positive for m sufficiently large. Set $K_0 = 0$ and define recursively,

$$K_i = \inf\{k > K_{i-1} : F_k \text{ is a singleton}\}, \quad i = 1, 2 \ldots$$

$$K_i = \sup\{k < K_{i+1} : F_k \text{ is a singleton}\}, \quad i = -1, -2 \ldots$$

Finally, set $G_1 = F_{K_1}$, $G_i = F_{K_i - 1} \circ \ldots \circ F_{K_{i-1}}$ for $i \in \mathbf{Z} \setminus \{1\}$, and $\tau_i = K_i - 1 - K_{i-1}$ for $i \in \mathbf{Z}$, where $\circ$ denotes composition of mappings.

Proposition 4 *If $\mathsf{P}(T_{PW} < \infty) = 1$, then the (G_i, τ_i) with $i \in \mathbf{Z} \setminus \{1\}$ are IID, where $G_i \sim \pi$ and τ_i follows a geometric distribution with mean $(1 - p)/p$.*

Proof. By Proposition 2, $G_0 \sim \pi$. Using that the random maps are IID, we obtain easily the following properties. The τ_i, $i \in \mathbf{Z}$, are IID, and each follows a geometric distribution with mean $(1 - p)/p$. Furthermore, conditionally on the τ_i, the random maps are mutually independent, where the conditional distribution of F_{K_i} is the same as the distribution of G_1 for $i \neq 0$, while for $j \notin \{K_i : i \neq 0\}$, the conditional distribution of F_j is the same as the conditional distribution of F_0 given that F_0 is not a singleton. Thus $(F_{K_i - 1}, \ldots, F_{K_{i-1}}, \tau_i)$, $i \in \mathbf{Z} \setminus \{1\}$, are IID, and so the (G_i, τ_i) with $i \in \mathbf{Z} \setminus \{1\}$ are IID.

As noticed in the proof, the τ_i (including τ_1) are IID. However, in general, $G_1 \not\sim \pi$. A counterexample is provided by the SRS (4) for the Widom-Rowlinson model: letting $m = 1$, we obtain that the first component of G_1 equals $\emptyset$, contradicting the fact that $\pi(\{\emptyset, \cdot\}) < 1$.

The read-once algorithm consists in generating $G_2, \ldots, G_j$ for a given integer $j \geq 2$. Notice that we can successively determine $G_2, \ldots, G_j$ from one path starting at $G_1 = F_{K_1}$, since $G_2 = F_{K_2-1} \circ \ldots \circ F_{K_1}$, while for $i = 3, 4, \ldots$, we have successively that $G_i = F_{K_i-1} \circ \ldots \circ F_{K_{i-1}} \circ G_{i-1}$. Here G_1 and $K_1, K_2, \ldots$ may be easily determined if we can bound the random maps by lower and upper processes like in the monotone and anti-monotone cases considered in Section 3.2. Wilson (2000a) recommends to choose m so that $p > 0.5$ or equivalently $\mathsf{E}(\tau_i) < 1$. Finally, we notice that Wilson (2000a) discusses a coupling method so that the read-once algorithm applies on locally stable point processes, as defined in Section 4.1.

4 Horizontal CFTP

As mentioned in Section 1, horizontal CFTP has other names in other papers; it is called so here in order to clarify the difference from vertical CFTP (we comment on this at the end of Section 4.1). The ideas behind horizontal CFTP are due to Kendall(1998); a general setting can be found in Kendall and Møller (2000). Section 4.1 shows the details in the case of using spatial birth and death processes with a locally stable equilibrium distribution. Section 4.2 concerns the case of a noisy point process model. Further examples are briefly discussed in Section 5.

4.1 Perfect simulation for locally stable point processes using spatial birth and death processes

This section is based on Kendall and Møller (2000). As the results will be used in Section 4.2, we give a detailed exposition.

We consider a general setting for finite point processes, where $\Lambda = \text{Poisson}(S, \kappa)$ is a Poisson point process defined on a space S and with intensity measure κ so that $0 < \kappa(S) < \infty$. For convenience we assume κ to be diffuse (i.e. non-atomic), whereby Λ is concentrated on the state space $E = \{x \subseteq S : n(x) < \infty\}$ of finite point configurations (but everything in the following easily extends to the case where κ is not diffuse); here $n(x)$ denotes the number of elements in the set x. So if $X \sim \Lambda$, then $n(X)$ follows a Poisson distribution with mean $\kappa(S)$, and conditionally on $n(X)$, the points in X are IID with distribution $\bar{\kappa}(\cdot) = \kappa(\cdot)/\kappa(S)$.

We restrict attention to point processes X with a target distribution π which is absolutely continuous with respect to Λ, and assume the following local stability condition: if $f = \mathrm{d}\pi/\mathrm{d}\Lambda$ denotes the density, there is a number $K > 0$ such that

$$f(x \cup \{\xi\}) \leq K f(x) \quad \text{for } x \in E,\ \xi \in S \setminus x. \tag{18}$$

In other words, defining the Papangelou conditional intensity,

$$\lambda(x,\xi) = \begin{cases} f(x \cup \{\xi\})/f(x) & \text{if } f(x) > 0,\ \xi \in S \setminus x \\ 0 & \text{otherwise} \end{cases}$$

we have assumed $\lambda \leq K$ and f to be hereditary, i.e. for any $x \in E$ and $\xi \in S \setminus x$, $f(x) > 0$ if $f(x \cup \{\xi\}) > 0$. Since everything in the sequel only depends on f through λ, we need only to know an explicit expression of f up to proportionality.

Consider first a spatial birth and death process $\tilde{X} = \{\tilde{X}_t : t \in \mathbf{R}\}$ with birth and death rates b and d. These are non-negative measurable functions defined on $E \times S$ such that, in any small time-interval $[t, t+\mathrm{d}t]$ and for any $B \in \mathcal{B}$, if we condition on $\tilde{X}_t = x$ and the process before time t, we have the following: (i) the probability for a birth $\tilde{X}_{t+\mathrm{d}t} = x \cup \{\xi\}$, with the newborn point $\xi \in B$, is $\int_B b(x,\xi)\kappa(\mathrm{d}\xi) \times \mathrm{d}t + \mathrm{O}(\mathrm{d}t)$; (ii) for any point $\eta \in x$, the probability for a death $\tilde{X}_{t+\mathrm{d}t} = x \setminus \{\eta\}$ is $d(x \setminus \{\eta\}, \eta)\mathrm{d}t + \mathrm{O}(\mathrm{d}t)$; (iii) the probability for more than one transition is $\mathrm{O}(\mathrm{d}t)$. The target density f and the spatial birth and death process $\tilde{X}$ are in detailed balance if

$$f(x)b(x,\xi) = f(x \cup \{\xi\})d(x,\xi) > 0 \quad \text{whenever } f(x \cup \{\xi\}) > 0, \tag{19}$$

in which case (18) ensures both the unique existence of $\tilde{X}$ and that $\tilde{X}_t$ converges in distribution towards π as $t \to \infty$, cf. Preston (1977) (alternatively, one can verify these properties directly using the coupling construction described below). In particular, (19) implies that the empty point configuration $\emptyset$ is an ergodic atom for $\tilde{X}$. In the sequel we set

$$b(x,\xi) = \lambda(x,\xi), \quad d \equiv 1,$$

whereby (19) holds. This is seemingly the simplest choice, but everything in the following can be extended to the general case of (19).

Consider next another spatial birth and death process $\tilde{D} = \{\tilde{D}_t : t \in \mathbf{R}\}$ with birth rate K and death rate 1. Then $\tilde{D}$ satisfies (19) when $f(x) \propto K^{n(x)}$, i.e. it has equilibrium distribution Poisson$(S, K\kappa)$, and $\emptyset$ is an ergodic atom. Let $J_1 < J_2 < \ldots$ denote the jump times of $\{\tilde{D}_t : t > 0\}$ and define the jump chain $\{D_t\}_{t=1}^{\infty} = \{\tilde{D}_{J_t}\}_{t=1}^{\infty}$; for convenience, set $D_0 = \tilde{D}_0$ and $J_0 = 0$. The key point is now that there is an explicit coupling of $\{\tilde{X}_t : t \geq 0\}$ and $\{\tilde{D}_t : t \geq 0\}$ so that the jump times of $\tilde{X}$ are included in $\{J_t\}$, and letting $\{X_t\}_1^{\infty} = \{\tilde{X}_{J_t}\}_1^{\infty}$ and $X_0 = \tilde{X}_0$, we have that $\tilde{D}$ dominates $\tilde{X}$ as

$$\tilde{X}_t \subseteq \tilde{D}_t, \quad t \geq 0. \tag{20}$$

Hence a natural partial ordering on E is $\subseteq$ (set inclusion). Notice that $\hat{0} = \emptyset$ is a unique minimum, but there exists no maximum.

Specifically, let $\epsilon_{t1}, \ldots, \epsilon_{t4},\ t \in \mathbf{Z}$ be mutually independent, where each of $\epsilon_{t1}, \epsilon_{t3}, \epsilon_{t4}$ is uniformly distributed on $[0,1]$, while $\epsilon_{t2} \sim \bar{\kappa}$. Then a SRS construction for D is given by

$$D_t = D_{t-1} \cup \{\xi_t\} \setminus \{\eta_t\}, \quad t = 1, 2, \ldots \tag{21}$$

where ξ_t and η_t are defined as follows. A birth happens if

$$\epsilon_{t1} < K\kappa(S)/[K\kappa(S) + n(D_{t-1})],$$

in which case $\xi_t = \epsilon_{t2}$ and $\{\eta_t\} = \emptyset$; otherwise a death happens, in which case $\{\xi_t\} = \emptyset$ and $\eta_t = \eta\,(D_{t-1}, \epsilon_{t3})$ is a uniformly selected point from D_{t-1}. Suppose $X_0 = \emptyset$ and define recursively for $t = 1, 2, \ldots$,

$$X_t = \begin{cases} X_{t-1} \cup \{\xi_t\} & \text{if } \{\xi_t\} \neq \emptyset \text{ and } \epsilon_{t4} < \lambda\,(X_{t-1}, \xi_t)\,/K \\ X_{t-1} \setminus \{\eta_t\} & \text{if } \{\xi_t\} = \emptyset \\ X_{t-1} & \text{otherwise.} \end{cases} \tag{22}$$

Finally, consider the exponentially distributed waiting times $J_t - J_{t-1} \sim \mathrm{Exp}(K\kappa(S) + n(D_{t-1})),\ t = 1, 2, \ldots$, whereby $\{\tilde{D}_t : t \geq 0\}$ and $\{\tilde{X}_t : t \geq 0\}$ are obtained. Then clearly (20) is satisfied, and it can be straightforwardly verified that $\{\tilde{X}_t : t \geq 0\}$ and $\{D_t : t \geq 0\}$ define two spatial birth and death processes with death rate 1 and birth rates λ and K, respectively. So setting $\hat{x} = \emptyset$, condition (ii) in Proposition 1 is satisfied.

In order to see that the other conditions of Proposition 1 are satisfied, imagine we first draw ϵ_{04} and $D_0 \sim \mathrm{Poisson}(S, K\kappa)$ so that the dominating process is in equilibrium. Next, imagine we simulate (D_t, ϵ_{t4}) forwards in time $t = 1, 2, \ldots$ and (D_{-t}, ϵ_{-t4}) backwards in time $-t = -1, -2, \ldots$ (by reversibility, this is an easy task). Finally, imagine we start "target chains" in $\hat{x} = \emptyset$: for $n = 0, 1, \ldots$, use (22) to obtain $X_{-n}(\emptyset) = \{X_{-n}^{-t}(\emptyset) : t = n, n-1 \ldots\}$. By Proposition 1, if $T \geq 0$ is an almost surely finite stopping time with respect to $\mathcal{R}_-$ such that $X_{-n}^{0}(\emptyset) = X_{-T}^{0}(\emptyset)$ whenever $n \geq T$, then $X_{-T}^{0}(\emptyset) \sim \pi$. One such stopping time is

$$T_\emptyset = \inf\{t \in \mathbf{N}_0 : D_{-t} = \emptyset\},$$

where $T_\emptyset < \infty$ almost surely, since $\emptyset$ is an ergodic atom. But like the stopping time (5) for the Widom-Rowlinson model, $T_\emptyset$ can be extremely large.

However, we can bound $X_{-n}(\emptyset)$ by the following lower and upper processes: set $L_{-n}^{-n} = \emptyset$, $U_{-n}^{-n} = D_{-n}$, and for $t = n-1, n-2, \ldots$,

$$L_{-n}^{-t} = \begin{cases} L_{-n}^{-t-1} \cup \{\xi_{-t}\} & \text{if } \{\xi_{-t}\} \neq \emptyset \text{ and } \epsilon_{-t4} < \alpha_L\left(L_{-n}^{-t-1}, U_{-n}^{-t-1}, \xi_{-t}\right) \\ L_{-n}^{-t-1} \setminus \{\eta_{-t}\} & \text{if } \{\xi_{-t}\} = \emptyset \\ L_{-n}^{-t-1} & \text{otherwise} \end{cases}$$

$$U_{-n}^{-t} = \begin{cases} U_{-n}^{-t-1} \cup \{\xi_{-t}\} & \text{if } \{\xi_{-t}\} \neq \emptyset \text{ and } \epsilon_{-t4} < \alpha_U\left(L_{-n}^{-t-1}, U_{-n}^{-t-1}, \xi_{-t}\right) \\ U_{-n}^{-t-1} \setminus \{\eta_{-t}\} & \text{if } \{\xi_{-t}\} = \emptyset \\ U_{-n}^{-t-1} & \text{otherwise} \end{cases}$$

where

$$\alpha_L\left(L_{-n}^{-t-1}, U_{-n}^{-t-1}, \xi_{-t}\right) = \min\left\{\lambda(x, \xi_{-t})/K : L_{-n}^{-t-1} \subseteq x \subseteq U_{-n}^{-t-1}\right\}, \quad (23)$$

$$\alpha_U\left(L_{-n}^{-t-1}, U_{-n}^{-t-1}, \xi_{-t}\right) = \max\left\{\lambda(x, \xi_{-t})/K : L_{-n}^{-t-1} \subseteq x \subseteq U_{-n}^{-t-1}\right\}. \quad (24)$$

Notice that the same $\xi_{-t}, \eta_{-t}, \epsilon_{-t4}$ are used for generating all $(L_{-n}^{-t}, U_{-n}^{-t})$ with $-n \leq -t$. Thereby we obtain the following sandwiching property

$$L_{-n}^{-t} \subseteq X_{-n}^{-t}(\emptyset) \subseteq U_{-n}^{-t} \subseteq D_{-t}, \quad \text{for } t = n, n-1, \ldots \text{ and } n = 0, 1, \ldots. \quad (25)$$

Observe also that the funnelling property (14) is satisfied. Finally,

$$T_{hor} = \inf\left\{n \in \mathbf{N}_0 : L^0_{-n} = U^0_{-n}\right\}$$

provides a stopping time which is often applicable (see below). Indeed $T_{hor} \leq T_\emptyset$ and typically $T_{hor} \ll T_\emptyset$.

Proposition 5 *For a locally stable point process with distribution π and lower and upper processes as defined above, $T_{hor} < \infty$ almost surely, and if $D_0 \sim Poisson(S, K\kappa)$ then $L^0_{-n} = X^0_{-T_{hor}}(\emptyset) \sim \pi$ whenever $L^0_{-n} = U^0_{-n}$.*

Proof. This follows immediately from Proposition 1 with $\hat{x} = \emptyset$, using the above-mentioned sandwiching and funnelling properties.

A CFTP algorithm based on generating lower and upper processes may now be implemented in the same way as the vertical CFTP algorithm in Section 3.2, except that we start by drawing ϵ_{04} and $D_0 \sim \text{Poisson}(S, K\kappa)$ and generate, as long as needed, (D_{-t}, R_{-t4}) further and further back in time, i.e. until $L^0_{-n} = U^0_{-n}$. We call this for horizontal CFTP, since we have a horizontal coupling between the paths in (25), and in contrast to vertical CFTP, we cannot use an arbitrary initial state $\hat{x}$ for the target chains; in fact we can only use $\hat{x} = \emptyset$.

In practice, it may only be feasible to determine (23) and (24) if $\lambda(x, \xi)$ considered as a function of x is either increasing (the so-called attractive case) or decreasing (the so-called repulsive case) with respect to the partial ordering $\subseteq$ — or, at least, if $\lambda(x, \xi)$ factorizes into terms which are increasing or decreasing in x, we may modify (23) and (24) in an obvious way so that the compuations become feasible. Fortunately, in most applications, the model is either attractive or repulsive.

4.2 Perfect simulation for a noisy point process

So far we have mainly for illustrative purposes considered rather simple examples of perfect simulation algorithms. More complicated algorithms are usually needed in "real applications" as exemplified in this section, where we consider an interesting noisy point process model studied in Lund and Rudemo (2000) and Lund et al. (1999), and give an alternative and shorter description of a slightly improved version of a CFTP algorithm introduced in Lund and Thönnes (2000) (the difference is explained at the end). Particularly, we demonstrate how the results in Section 4.1 can be applied.

The noisy point process model can briefly be described as follows. Let $A \subset \mathbf{R}^d$ be a Borel set with $0 < |A| < \infty$ and let $\mu = \text{Poisson}(A, 1)$. Suppose X is a point process with density f with respect to μ, which is subject to the following four operations.

(I) Independent thinning, where each point $\xi \in X$ is retained with probability p; here $0 < p < 1$ is a given parameter.

(II) If $\xi \in X$ is retained, then it is translated by a vector $R(\xi)$ with density ϕ with respect to Lebesgue measure on $\mathbf{R}^d$; here the vectors $R(\eta),\ \eta \in \mathbf{R}^d$ are IID and independent of the retained points in X.

(III) The retained displaced points $Z(\xi) = \xi + R(\xi)$, which are outside A, are censored.

(IV) Let $U \subseteq X$ denote the points which are either thinned in (I) or censored in (III) after the displacement in (II). Set $V = X \setminus U$, $Z = \{Z(\xi) : \xi \in V\}$, and $Q = \{(Z(\xi), R(\xi)) : \xi \in V\}$. Assume that $W \sim \text{Poisson}(A, \alpha)$ is independent of (U, Q), where $\alpha > 0$ is a parameter, and that only a realization of $Y = Z \cup W$ is observed.

Our target distribution π is the conditional distribution of (U, Q) given $Y = y$. From this we can obtain the conditional distribution of X given $Y = y$, which in the above-mentioned papers is considered as the posterior distribution of primarily interest. In order to apply the results in Section 4.1, we let π be defined on an augmented state space

$$E = \{(u, q) :\ u \subset (A \setminus y),\ n(u) < \infty,\ q \subset y \times \mathbf{R}^d,\ n(q) < \infty\},$$

i.e. we include marked point configurations $q = \{(z_1, r_1), \ldots, (z_n, r_n)\} \subset y \times \mathbf{R}^d$ where some of the points $z_1, \ldots, z_n \in y$ can be equal, though Q under π is concentrated on the set M consisting of those q where $z_1, \ldots, z_n$ are pairwise distinct and $z_i - r_i \in A,\ i = 1, \ldots, n$. Finally, for a given point $\xi \in X$,

$$h(\xi) \equiv \mathsf{P}(\xi \in U | \xi \in X) = p \int_{r \notin A} \phi(r - \xi) \mathrm{d}r + 1 - p$$

is the probability that the point is either thinned in (I) or censored in (III) after the displacement in (II), and we define for finite point configurations

$u \subset A$,

$$H(\emptyset) = 1, \quad H(u) = \prod_{\xi \in u} h(u) \quad \text{for } u \neq \emptyset.$$

Proposition 6 *Let*

$$\Lambda' = \text{Poisson}(A \setminus y, \kappa_1) \times \text{Poisson}(y \times \mathbf{R}^d, \kappa_2) \tag{26}$$

where κ_1 denotes Lebesgue measure restricted to $A \setminus y$ and κ_2 is defined by

$$\kappa_2(B \times C) = \sum_{\eta \in y} \mathbf{1}[\eta \in B] \times \int_C p\phi(r)/\alpha \mathrm{d}r$$

for $B \subseteq y$ and Borel sets $C \subseteq \mathbf{R}^d$. Assume that $g(y) > 0$, where g is the density of Y with respect to μ, see (30) in Appendix A. Then π has a density with respect to Λ',

$$(\mathrm{d}\pi/\mathrm{d}\Lambda')(u, q) = c(y)H(u)f(u \cup v)\mathbf{1}[q \in M], \quad (u, q) \in E \tag{27}$$

where we let $v = \{v_1, \ldots, v_n\}$ be specified by $q = \{(v_1 + r_1, r_1), \ldots, (v_n + r_n, r_n)\}$, and where $c(y)$ depends on y only.

Proof. See Appendix A.

In the sequel we assume that f is locally stable, and let its Papangelou conditional intensity λ_f, say, be bounded by the constant K, cf. Section 4.1. Then $f(\emptyset) > 0$, and so by (31) in Appendix A, the condition $g(y) > 0$ is satisfied. Owing to the factorization in (26) and the fact that (U, Q) is in one-to-one correspondence with $U \cup Q$, we can then put things into the framework of Section 4.1: the density of $U \cup Q | Y = y$ with respect to $\Lambda = \text{Poisson}(S, \kappa)$, where $S = (A \setminus y) \cup (y \times \mathbf{R}^d)$ and $\kappa(B) = \kappa_1(B \cap A) + \kappa_2(B \cap (y \times \mathbf{R}^d))$, is given by (27). This density is clearly hereditary and its Papangelou conditional density λ is given by

$$\lambda(u \cup q, \xi) = h(\xi)\lambda_f(u \cup v, \xi)\mathbf{1}[q \in M] \quad \text{if } \xi \in (A \setminus y) \setminus u,$$

$$\lambda(u \cup q, (\eta, r)) = \lambda_f(u \cup v, \eta)\mathbf{1}[q \cup \{(\eta, r)\} \in M] \quad \text{if } (\eta, r) \in (y \times \mathbf{R}^d) \setminus q,$$

so $\lambda \leq K$ as $\lambda_f \leq K$ and $h \leq 1$, i.e. local stability is satisfied.

Therefore we can make perfect simulations from $U \cup Q | Y = y$ as described in Section 4.1. In order to avoid any confusion with the notation used in Section 4.1, let us rename ξ_t and η_t from (21) by χ_t and ζ_t, respectively. Note that conditionally on a birth $\{\chi_t\} \neq \emptyset$, we have that $\chi_t \in A \setminus y$ with probability $\kappa_1(A \setminus y)/\kappa(S) = |A|/(|A| + n(y)p/\alpha)$, in which case χ_t follows a uniform distribution on $A \setminus y$; otherwise $\chi_t = (\eta_t, r_t)$, where η_t is uniformly selected from y, independently of $r_t \sim \phi$. Moreover, conditionally on D_{t-1} and a death $\{\zeta_t\} \neq \emptyset$, we have that ζ_t is uniformly selected from D_{t-1}. Finally, λ considered as a function of its first argument splits into a product of an indicator function, which is decreasing, and λ_f, which is typically decreasing or increasing, in which case the computations become feasible, cf. the last paragraph of Section 4.1. But as the state space for $\tilde{D}_t$ is much larger than that for $\tilde{X}_t$, this perfect simulation procedure can be rather inefficient as demonstrated in Lund and Thönnes (2000).

However, a better dominating jump process $\tilde{D}^*_t$, say, can be constructed so that

$$\tilde{X}_t \subseteq \tilde{D}^*_t \subseteq \tilde{D}_t, \quad t \geq 0. \tag{28}$$

Assuming $\tilde{X}_0 = \emptyset$ and $\tilde{D}^*_0 \subseteq \tilde{D}_0$, then the jump times of $\tilde{D}^*$ and $\tilde{D}$ are the same. Furthermore, in the corresponding two jump chains D^* and D, we are using the same births χ_t, and the same deaths ζ_t whenever $\zeta_t \in (A \setminus y)$. However, if a death $\zeta_t = (\eta_t, r_t) \in y \times \mathbf{R}^d$ happens, then as X_t does not contain η_t, we set

$$D^*_t = D^*_{t-1} \setminus \{(\eta, r) \in D^*_{t-1} : \eta = \eta_t\},$$

whereby (28) is seen to hold.

Note that we can write $\tilde{D}^*_t = \tilde{U}_t \cup \{(\xi, \tilde{S}_t(\xi)) : \xi \in y, \tilde{S}_t(\xi) \neq \emptyset\}$, where $\{\tilde{U}_t : t \geq 0\}$, $\{\tilde{S}_t(\xi) : t \geq 0\}$, $\xi \in y$ are mutually independent jump processes. Here $\{\tilde{U}_t : t \geq 0\}$ is a spatial birth and death process on $A \setminus y$ with birth rate K and death rate 1, so this is reversible with invariant distribution equal to Poisson$(A \setminus y, K)$. Further, $\tilde{N}_t(\xi) \equiv n(\tilde{S}_t(\xi))$ has generator $G = (g_{m,n})$ given by $g_{m,m+1} = Kp/\alpha$ for $m = 0, 1, \ldots$, $g_{m,0} = 1$ for $m = 1, 2, \ldots$, and $g_{m,n} = 0$ otherwise for off-diagonal elements; recall that the generator for a jump process with discrete state space has all row sums equal to 0 (see e.g. Norris, 1997). As its invariant density p_n satisfies $\sum p_m g_{m,n} = 0$, $n = 0, 1, \ldots$, we find that $p_n \propto (\alpha/(\alpha + Kp))^n$ specifies a geometric equilibrium distribution. Furthermore, conditionally on that a birth happens in $\tilde{S}(\xi)$ at time t, the newborn point $r_t \sim \phi$, and it is independent of the other points in $\tilde{S}_t(\xi)$ and the previous history of the process. So the equilibrium distribution of $\tilde{S}_t(\xi)|\tilde{N}_t(\xi)$ is simply a binomial process of IID points, each following the density ϕ. Therefore $D^*_0 = \tilde{D}^*_0$ can easily be started in equilibrium. Moreover, though $\{\tilde{S}_t(\xi) : t \geq 0\}$ is not reversible, it can be easily simulated backwards in time: if $G' = (g'_{m,n})$ is the generator of the reversed process of $\tilde{N}_t(\xi)$, then $g'_{m+1,m} = 1 + Kp/\alpha$ for $m = 0, 1, \ldots$, $g'_{0,m} = (\alpha/(\alpha + Kp))^m$ for $m = 1, 2, \ldots$, and $g'_{m,n} = 0$ otherwise for off-diagonal elements. This follows by solving the equations $p_m g_{m,m+1} = p_{m+1} g'_{m+1,m}$ and $p_{m+1} g_{m+1,0} = p_0 g'_{0,m+1}$ for $m = 0, 1, \ldots$.

Now, the point is that we can obtain a horizontal CFTP algorithm along the same lines as in Section 4.1, but replacing D with D^*: as described in detail above, D^* can easily be started in equilibrium at time 0 and simulated backwards in time — without simulating D — and we set $U_{-n}^{-n} = D^*_{-n}$, and let the points which are going to be added or deleted at time $-t$ be given by $D^*_{-t} \setminus D^*_{-t-1}$ and $D^*_{-t-1} \setminus D^*_{-t}$, respectively. As suggested in Lund and Thönnes (2000), it may be convenient to replace the "acceptance probabilities" for births as given by (23) and (24) by lower respective upper bounds, which do not depend on evaluating the function h, but depend on the properties of λ_f only: as $0 < 1 - p \leq h \leq 1$ and the function $\mathbf{1}[q \in M]$ is decreasing, we can redefine

$$\alpha_L\left(L_{-n}^{-t-1}, U_{-n}^{-t-1}, \chi_{-t}\right) = \mathbf{1}[U_{-n}^{-t-1} \cap (y \times \mathbf{R}^d) \in M](1-p)/K \times$$

$$\min\left\{\lambda_f((u \cup v), \chi_{-t}) : L_{-n}^{-t-1} \subseteq (u,q) \subseteq U_{-n}^{-t-1}\right\} \quad \text{if } \chi_{-t} \in A \setminus y,$$

$$\alpha_L\left(L_{-n}^{-t-1}, U_{-n}^{-t-1}, \chi_{-t}\right) = \mathbf{1}[(U_{-n}^{-t-1} \cup \{\chi_{-t}\}) \cap (y \times \mathbf{R}^d) \in M]/K \times$$

$$\min\left\{\lambda_f((u \cup v), \chi_{-t}) : L_{-n}^{-t-1} \subseteq (u,q) \subseteq U_{-n}^{-t-1}\right\} \quad \text{if } \chi_{-t} \in y \times \mathbf{R}^d,$$

where we let v be determined by $(u,q) \in E$; similarly, redefine

$$\alpha_U\left(L_{-n}^{-t-1}, U_{-n}^{-t-1}, \chi_{-t}\right) = \mathbf{1}[L_{-n}^{-t-1} \cap (y \times \mathbf{R}^d) \in M](1-p)/K \times$$

$$\max\left\{\lambda_f((u \cup v), \chi_{-t}) : L_{-n}^{-t-1} \subseteq (u,q) \subseteq U_{-n}^{-t-1}\right\} \quad \text{if } \chi_{-t} \in A \setminus y,$$

$$\alpha_U\left(L_{-n}^{-t-1}, U_{-n}^{-t-1}, \chi_{-t}\right) = \mathbf{1}[(U_{-n}^{-t-1} \cup \{\chi_{-t}\}) \cap (y \times \mathbf{R}^d) \in M]/K \times$$

$$\min\left\{\lambda_f((u \cup v), \chi_{-t}) : L_{-n}^{-t-1} \subseteq (u,q) \subseteq U_{-n}^{-t-1}\right\} \quad \text{if } \chi_{-t} \in y \times \mathbf{R}^d.$$

Proposition 7 *Let the situation be as described above. Then* $T^*_{hor} = \inf\{n \geq 0 : L^0_{-n} = U^0_{-n}\}$ *is almost surely finite, and if* D^*_0 *is started in equilibrium then* $L^0_{-n} \sim \pi$ *whenever* $L^0_{-n} = U^0_{-n}$.

Proof. This follows in a similar way as the proof of Proposition 5, recalling the possibility of coupling D^* with D so that $T^*_{hor} \leq T_{hor}$.

As promised we now compare this with the perfect simulation algorithm in Lund and Thönnes (2000, Section 6.3); apparently, they include the jump times, but since this is not needed, let us just consider the jump chains and use our notation. Lund and Thönnes do not start by simulating D^*_0 in equilibrium, but for each time $-n$, before simulating (L_{-n}, U_{-n}) started at time $-n$, they start D^* further back at time $\min\{-T_n(\xi) : \xi \in y\}$, where $-T_n(\xi) = \sup\{-t \leq -n : N_{-t}(\xi) = 0\}$; here $N(\xi)$ is the jump process of $\tilde{N}(\xi)$. So the running time T_{LT} of their algorithm is related to ours by that $T_{LT} \sim T^*_{hor} + \epsilon$, where ϵ is distributed as the maximum of $n(y)$ IID random variables, each following the geometric distribution with density p_n. For the particular application considered in Lund and Thönnes (2000), possibly ϵ is rather small. They report that typically $T_{LT} \ll T_{hor}$, where T_{hor} is the running time for horizontal CFTP based on Section 4.1, i.e. before the better dominating jump process $\tilde{D}^*$ is introduced.

5 Concluding remarks and further reading

Remark 1 (Fill's algorithm) Fill (1998) introduces a clever form of rejection sampling, assuming a finite state space and a monotone setting with unique minimal and maximal states. Applications and extensions of Fill's algorithm can be found in Fill et al. (2000) and the references therein. The advantage of Fill's algorithm compared to CFTP is that it is interruptible in the sense that the output is independent on the running time (like in any rejection sampler). The disadvantages may be problems with storage and that it seems more limited for applications than CFTP. As regards applications of Fill's algorithm in stochastic geometry, Thönnes (1999) considers the special case of the Widom-Rowlinson model (using the anti-monotone setting, this can be extended to the case of k-components, see Example 2, Section 3.2), while Møller and Schladitz (1999) consider more general spatial point processes approximated by lattice processes.

Remark 2 (other horizontal CFTP algorithms) As noticed, the algorithm in Section 4.1 is only practical if the Papangelou conditional intensity $\lambda(x, \xi)$ is increasing or decreasing in its first argument; and this is often the case in applications of stochastic geometry. In contrast the horizontal CFTP algorithm in Fernándes et al. (1999) depends only on λ through its "interaction radius"; for example, if $S \subset \mathbf{R}^d$, then $\inf\{r > 0 : \lambda(x, \xi) = 1, \mathrm{dist}(x, \xi) = r\}$ is the interaction radius at the point $\xi \in S$). The algorithm consists in finding "clans of ancestors" in the dominating spatial birth and death process, when this is simulated backwards in time from time 0 until there are no more ancestors, and then obtain a target process by thinning as in (22); so no lower and upper processes are needed. For the running time T_F of the algorithm (when jump times are ignored), we have that $T_F \geq T_{hor}$. This may not be a fair comparison if we are using a doubling scheme for the algorithm in Section 4.1, and the efficiency of the algorithms depend much on how close λ is to its upper bound K used in the dominating process, and also on the range of the interaction radius.

There are other horizontal CFTP algorithms: Metropolis-Hastings algorithms for locally stable point processes are studied in Kendall and Møller (2000), and the use of spatial jump processes for more general classes of spatial point processes is studied in Berthelsen and Møller (2000). It is noticed in Kendall and Møller (2000) that the considered Metropolis-Hastings algorithm is geometrically ergodic, but in general it is not uniformly ergodic — recall that uniform ergodicity is a necessary condition for vertical CFTP to work.

Remark 3 (infinitely many points) Often one considers point processes with infinitely many points contained in an "infinite volume" such as $\mathbf{R}^d$. In order to avoid edge-effects, a perfect sample within a bounded region may be achieved by extending simulations both backwards in time and in space (Kendall 1997; Fernándes et al. 1999). This is sometimes possible, for example if λ is sufficiently close to K and the interaction radius is sufficiently small. Such coupling constructions may be of great theoretical interest, but in my opinion they remain so far unpractical for applications of real interest.

Remark 4 (statistical applications) Section 4.2 provides one example of a Bayesian application of horizontal CFTP using the results in Section 4.1; another is given in Loizeaux and McKeague (2000).

Van Zwet (1999) relates horizontal CFTP to likelihood based inference for a conditional Boolean model. Here a lower dominating process is used, while the upper dominating process is the same for all times; see also Kendall and Thönnes (1999). It would be interesting to study more complicated parametric models like germ grain models of interacting geometrical objects, where possibly perfect simulated tempering (Møller and Nicholls, 1999) could be applied for finding the maximum likelihood estimate.

Acknowledgement Kasper K. Berthelsen, Katja Ickstadt, and Rasmus P. Waagepetersen have commented on an earlier version of this paper. This research has been supported by the Centre for Mathematical Physics and Stochastics (MaPhySto), funded by a grant from the Danish National Research Foundation.

Appendix A

We start by verifying Proposition 6. By (I)-(III) in Section 4.2, for finite point configurations $u \subseteq x \subset A$ and $v = x \setminus u = \{v_1, \ldots, v_n\}$,

$$\mathsf{P}(U = u, Q \in G | X = x) = H(u) \int \cdots \int \mathbf{1}[q \in G, z \subset A] \prod_1^n (p\phi(r_i)) \, \mathrm{d}r_1 \cdots \mathrm{d}r_n$$

where $q = \{q_1, \ldots, q_n\}$ and $z = \{z_1, \ldots, z_n\}$ are given by $z_i = v_i + r_i$ and $q_i = (z_i, r_i)$, and where the n-fold integral is read as $\mathbf{1}[\emptyset \in G]$ if $n = 0$. Using the definition of μ and Fubini's theorem, it is easily seen that

$$\int \sum_{n=0}^{\infty} \int \cdots \int \sum_{\{v_1,\ldots,v_n\}\subseteq x} k(u,q)\,\mathrm{d}r_1 \cdots \mathrm{d}r_n\,\mu(\mathrm{d}x)$$
$$= \int \sum_{n=0}^{\infty} \frac{1}{n!} \int \cdots \int k(u,q)\,\mathrm{d}q_1 \cdots \mathrm{d}q_n\,\mu(\mathrm{d}u) \tag{29}$$

for integrable functions k. Hence

$$\mathsf{P}(U \in F, Q \in G) =$$
$$\int_F H(u) \sum_{n=0}^{\infty} \frac{p^n}{n!} \int \cdots \int \mathbf{1}[q \in G, v \subset A] f(u \cup v)\phi(r_1)\mathrm{d}q_1 \cdots \phi(r_n)\mathrm{d}q_n\,\mu(\mathrm{d}u).$$

Combining this with (IV) in Section 4.2, we obtain that

$$\mathsf{P}(U \in F, Q \in G, Y \in N) =$$
$$\exp(|A| - \alpha|A|) \int_F H(u) \int \alpha^{n(w)} \sum_{n=0}^{\infty} \frac{p^n}{n!} \int \cdots \int \mathbf{1}[q \in G, v \subset A, z \cup w \in N]$$
$$f(u \cup v)\phi(r_1)\mathrm{d}q_1 \cdots \phi(r_n)\mathrm{d}q_n\,\mu(\mathrm{d}w)\,\mu(\mathrm{d}u) =$$
$$\exp(|A| - \alpha|A|) \int_F H(u) \int_N \alpha^{n(y)} \sum_{n=0}^{\infty} (p/\alpha)^n \int \cdots \int \sum_{\{z_1,\ldots,z_n\}\subseteq y}$$
$$\mathbf{1}[q \in G, v \subset A] f(u \cup v)\phi(r_1)\mathrm{d}r_1 \cdots \phi(r_n)\mathrm{d}r_n\,\mu(\mathrm{d}y)\,\mu(\mathrm{d}u)$$

where the second identify is obtained from (29), replacing u, x, v with w, y, z, respectively. Thereby Proposition 6 follows with

$$g(y) = \exp(|A| - \alpha|A|)\alpha^{n(y)} \int H(u) \sum_{n=0}^{\infty} (p/\alpha)^n \times$$
$$\int \cdots \int \sum_{\{z_1,\ldots,z_n\}\subseteq y} f(u \cup v)\phi(r_1)\mathrm{d}r_1 \cdots \phi(r_n)\mathrm{d}r_n\,\mu(\mathrm{d}y)\,\mu(\mathrm{d}u) \tag{30}$$

and $c(y) \propto g(y)$.

We conclude with various comments.

The condition $g(y) > 0$ is satisfied if $f(\emptyset) > 0$: considering the case where $u = v = \emptyset$ and $n = 0$ in (30), we obtain that

$$g(y) \geq \exp(|A| - \alpha|A|)\alpha^{n(y)} f(\emptyset). \tag{31}$$

Proposition 6 can easily be extended to cases where the parameter p in (I) is allowed to depend on the location of a point in X (so-called inhomogeneous thinning), ϕ in (II) is allowed to depend on the location of a retained point, W has a density with respect to μ, and we put a prior distribution on

(p, α, ϕ). However, as an explicit expression for the normalizing constant of f is usually not known in applications, it will usually not be feasible to introduce a prior on f.

Incidentally, a simplified proof of Theorem 1 in Lund and Rudemo (2000), which concerns the expression for the conditional density of Y given $X = x$, can easily be obtained along similar lines as in the proof above.

References

Berthelsen, K.K. and Møller, J. (2000). Spatial jump processes and perfect simulation for spatial point processes. In preparation.

Fernández, R., Ferrari, P.A. and Garcia, N.L. (1999). Perfect simulation for interacting point processes, loss networks and Ising models. Avaliable at http://xxx.lanl.gov/abs/math.PR/9911162

Fill, J. (1998). An interruptible algorithm for exact sampling via Markov chains. *Ann. Appl. Prob.*, **8**, 131–162.

Fill, J.A., Machida, M., Murdoch, D.J. and Rosenthal, J.S. (2000). Extensions of Fill's perfect rejection sampling algorithm to general chains. *Random Structures and Algorithms*, **16**. To appear.

Foss, S.G. and Tweedie, R.L. (1998). Perfect simulation and backward coupling. *Stoch. Models*, **14**, 187–203.

Häggström, O., van Lieshout, M.N.M. and Møller, J. (1999). Characterization results and Markov chain Monte Carlo algorithms including exact simulation for some spatial point processes. *Bernoulli*, **5**, 641–659.

Jeulin, D. (1997). Dead leaves model: from space tesellation to random functions. *Advances in Theory and Applications of Random Sets*, D. Jeulin, ed., World Scientific Press, Singapore, 137–156.

Kendall, W.S. (1997). Perfect simulation for spatial point processes. *Proc. ISI* 51^{st} *session, Istanbul (August 1997)*, **3**, 163–166.

Kendall, W.S. (1998). Perfect simulation for the area-interaction point process. *Probability Towards 2000*, L. Accardi and C.C. Heyde, eds., Springer Verlag, New York, 218–234.

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**. To appear.

Kendall, W.S. and Thönnes, E. (1999). Perfect simulation in stochastic geometry. *Pattern Recognition*, **32**, 1569–1586.

Loizeaux, M.A. and McKeague, I.W. (2000). Bayesian inference for spatial point processes via perfect simulation. Available at http://stat.fsu.edu/~mckeague/ps/index.html

Lund, J., Penttinen, A. and Rudemo, M. (1999). Bayesian analysis of spatial point patterns from noisy observations. Preprint 1999:57, Department of Mathematics, Chalmers University of Technology and Gothenburg University. Availiable at http://www.math.chalmers.se/Stat/Research/Preprints/

Lund, J. and Rudemo, M. (2000). Models for point processes observed with noise. *Biometrika*, **87**, 235–249.

Lund, J. and Thönnes, E. (2000). Perfect simulation of point patterns from noisy observations. Research Report 366, Department of Statistics, University of Warwick.

Matheron, G. (1968). Schéma booléen séquentiel de partition aléatoire. Internal Report N-83, CMM, Fontainebleau.

Matheron, G. (1975). *Random Sets and Integral Geometry*, Wiley, New York.

Meyn, S.P. and Tweedie, R.L. (1993). *Markov Chains and Stochastic Stability*, Springer Verlag, New York.

Møller, J. and Nicholls, G. (1999). Perfect simulation for sample-based inference. Research Report R-99-2011, Department of Mathematical Sciences, Aalborg University.

Møller, J. and Schladitz, K. (1999). Extensions of Fill's algorithm for perfect simulation. *J. Roy. Statist. Soc.*, **B 61**, 955–969.

Norris, J.R. (1997). *Markov Chains*, Cambridge University Press, Cambridge.

Preston, C.J. (1977). Spatial birth-and-death processes. *Bull. Inst. Internat. Statist.*, **46**, 371–391.

Propp, J.G. and Wilson, D.B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, **9**, 223–252.

Serra, J. (1982). *Image Analysis and Mathematical Morphology*, Academic Press, London.

Thönnes, E. (1999). Perfect simulation of some point processes for the impatient user. *Adv. Appl. Prob.*, **31**, 69–87.

Wilson, D.B. (2000a). How to couple from the past using a read-once source of randomness. *Random Structures and Algorithms*, **16**, 85–113.

Wilson, D.B. (2000b). Layered multishift coupling for use in perfect sampling algorithms (with a primer to CFTP). *Monte Carlo Methods*, N. Madras, ed., Fields Institute Communications, **26**, 141–176.

van Zwet, E.W. (1999). Likelihood devices in spatial statistics. PhD thesis, Universiteit Utrecht, Faculteit Wiskunde en Informatica.