Chapter 4

The Bombing Model

4.1 Introduction

Many images found in microscopy, materials science and biology can be described by means of a random set. Perhaps the best known model is the *Boolean model* formalizing a configuration of independent, randomly placed particles. It is formed by replacing the points of a Poisson process by random closed sets. The points of the Poisson process are sometimes called the *germs*, the associated random sets the *grains* or *particles*. If the process is defined on the plane and the grains are discs, then it is also know as the 'bombing model'. Notwithstanding the strong independence assumptions, inference for Boolean models is far from trivial (Molchanov, 1997). The difficulty lies in the occlusion arising from the fact that only the union of all particles is observed, not the individual germ–grain pairs.

One may distinguish between two types of parameters of a Boolean model: aggregate (or macroscopic) parameters and individual (or microscopic) ones (Molchanov, 1997). Typical examples of aggregate parameters are the area fraction and the set-covariance, which can be easily estimated by their observed image counterparts. The resulting estimators are unbiased, and expressions for the variance can be obtained from Robbins' theorem (e.g., Stoyan et al. 1987). Under mild ergodicity assumptions they are strongly consistent (Molchanov, 1997) as the observation window expands to the entire plane. Aggregate functionals such as the contact distribution and pair correlation function are of interest when fitting the Boolean model to a data image. Usually, estimation is hampered by edge effects, but minus sampling ideas (Ripley, 1988, Stoyan et al. 1987) are generally applicable as are Horvitz–Thompson style estimators including the Kaplan–Meier (Baddeley and Gill, 1995) and Hanisch style estimators (Hanisch, 1984). Unbiased-

ness follows from the Campbell–Mecke theorem (Stoyan et al. 1987) and asymptotic results are available (Molchanov, 1997).

Individual parameters, including the intensity of the germ process, are much harder to estimate. Minimum contrast methods (Dupač, 1980, Serra, 1982) for the intensity are based on minimizing the distance between an estimated aggregate parameter (e.g. the contact distribution) and an approximation expressed in terms of the intensity. Some asymptotic results are available, but the expressions for the asymptotic variance are too complicated to be useful in practice. An alternative is the method of moments based on coverage fraction, mean area and boundary length, and the Euler–Poincaré characteristic. This method is computationally easy but leads to a biased estimator (Weil, 1988). In the tangent point approach, the Euler–Poincaré characteristic is replaced by the specific connectivity number, resulting in easier asymptotics (Molchanov and Stoyan, 1994). Further details can be found in (Molchanov, 1977) and the references therein.

In this chapter we shall take a likelihood based approach using Monte Carlo methods to perform the necessary computations. To do so, we need to be able to sample from the conditional distribution of a Boolean model given an observation of the union of its particles. In the next section we show that this distribution is straightforward if the grains are balls. However, due to an intractable normalizing constant, direct sampling is not possible. In section 4.3, we use 'coupling from the past' (Propp and Wilson, 1996) to design an algorithm yielding *exact* or *perfect* samples. We discussed coupling from the past in section 1.5.1 of this thesis. Also, we provide a simple modification of our CFTP algorithm which speeds it up considerably.

Section 4.4 is devoted to two approaches to maximum likelihood estimation through simulation. The first method is based on a Monte Carlo approximation of the likelihood ratio with respect to a fixed parameter value (Geyer, 1998). The other approach is a stochastic version of the EM-algorithm (StEM algorithm, cf. section 1.5.2) (Celeux and Diebolt, 1986). This iterative algorithm works as follows. We start by choosing some initial parameter value. At each E-step one or more samples are drawn from the conditional distribution, under the current parameter value, of the complete data given the observed data. These samples are used to estimate the conditional expectation of the likelihood, given the data. This estimated likelihood is then maximized in the M-step to obtain a new parameter value. Repeating these steps a sequence of parameter values is obtained. In fact, under certain conditions, this sequence is an ergodic Markov chain on the parameter space. The stochastic EM (StEM) estimator is defined to be a sample, or an average of samples, from the stationary distribution of this Markov chain. See (Nielsen, 1997a and b) for asymptotic results. A drawback of the algorithm is

that it is not clear when the chain has reached stationarity. However, we can make the algorithm 'perfect' by applying coupling from the past to obtain perfect samples from the 'StEM' chain's stationary distribution. This issue was explored in section 1.5.3.

In section 4.5 we present results of a modest simulation experiment we conducted to compare the two methods.

In section 4.6 we note that our simulation algorithm can be applied to sample from any point process for which the so-called Papangelou conditional intensity is bounded away from zero. In this sense, our work is complementary to an algorithm of Kendall and Møller (1999) to sample from point processes for which the Papangelou conditional intensity is bounded away from infinity. We also present a generalization of the algorithm of Kendall and Møller making it faster.

Between them, the two algorithms will enable us to sample from a very wide variety of point processes. Especially sampling from Boolean models under certain constraints could be of considerable practical use. A typical example comes from the oil industry (Chessa, 1995), where a Boolean model is employed to represent the reservoir geometry. If it is known from test drilling or geological surveying that the reservoir has certain characteristics, simulation studies into its further properties then amount to sampling from a Boolean model conditional on these characteristics.

4.2 The conditional Boolean model

Suppose a realization of a Boolean model is observed with the aim to estimate the intensity of the underlying point process of germs. Since the conditional distribution of the germs given the observed union of particles involves an intractable normalizing constant depending on the intensity parameter of interest, direct maximum likelihood estimation seems very hard. However, in certain situations, it is possible to sample from the conditional distribution, so that Monte Carlo-based maximum likelihood estimation is a viable alternative.

In this chapter, we focus our attention on the case where the particles are (random) balls. Then, the location of a germ is identified upon observation of any part of the boundary of its associated ball. The remaining, not directly identifiable germs turn out to be distributed as a Poisson process conditioned to satisfy a coverage condition. The following makes this claim more precise.

Definition 4.1. *A* Boolean model *of balls is defined as follows. Let X be a stationary Poisson process with intensity $\lambda > 0$ on $\mathbb{R}^d$, and $B = B(0,1)$ the*

d-dimensional closed unit ball centered at the origin. Then, writing $A \oplus B = \{a + b : a \in A,\ b \in B\}$*, the random set*

$$\mathcal{B}(X) = \cup_{x_i \in X}(x_i \oplus rB)$$

is a Boolean model of balls with radius $r > 0$ *on* $\mathbb{R}^d$*.*

For $d = 2$, the process of Definition 4.1 is sometimes referred to as the *bombing model*. The underlying points X are called the *germs*. The set B is called the *primary grain*. More general Boolean models are obtained by letting the germs be scattered according to a non-stationary Poisson process or letting the grains be arbitrary random closed sets.

Our goal in this chapter is to perform statistical inference for the intensity parameter λ based on an observation of $\mathcal{B}(X)$ in a compact sampling window W with non-empty interior. Because B is symmetric about the origin we have for all $x \in \mathbb{R}^2$

$$x \in W \oplus B \Leftrightarrow x \oplus B \cap W \neq \emptyset,$$

and hence

$$\mathcal{B}(X \cap (W \oplus rB)) \cap W = \mathcal{B}(X) \cap W$$

So inference may be based on the conditional distribution of the germ process $X \cap (W \oplus rB)$ given the data $Y = \mathcal{B}(X) \cap W$.

The unconditional distribution of $X \cap (W \oplus rB)$ is absolutely continuous with respect to that of a unit rate Poisson process on $(W \oplus rB)$ with Radon-Nikodym derivative at a configuration $\mathbf{x}$ given by $\exp\{(1 - \lambda)|W \oplus rB|\}\lambda^{n(\mathbf{x})}$. This density is defined on the space of all finite, unordered sets $\mathbf{x} = \{x_1, \ldots, x_n\}$ of points in $W \oplus rB$. Here $|W \oplus rB|$ denotes the area of the set $W \oplus rB$, and $n(\mathbf{x})$ is the number of points in configuration $\mathbf{x}$.

We return to the conditional distribution of $X \cap W \oplus rB$ given $Y = \mathcal{B}(X) \cap W$. Since the primary grains are balls, the location of a germ is identified whenever a part of its associated grain's boundary is exposed. Therefore, the conditional distribution of $X \cap (W \oplus rB)$ can be decomposed into a deterministic 'exposed boundary' part X^b and a stochastic 'interior' X^i of germs that cannot be uniquely identified. Indeed we write $X \cap (W \oplus rB) = X^i \cup X^b$. The conditional distribution given Y of the exposed boundary part X^b is of course degenerate at some configuration $\mathbf{x}^b$. The distribution of the unobserved germs X^i is only slightly more complicated.

Define

$$\mathcal{C} = Y \setminus \mathcal{B}(X^b) \tag{4.1}$$

$$\mathcal{D} = \{y \in W \oplus rB : (y \oplus rB) \cap W \subseteq Y\}. \tag{4.2}$$

In words, $\mathcal{C}$ is the part of Y which is not covered by exposed grains, and must therefore be covered by the interior grains. The set $\mathcal{D}$ describes the locations where interior points may fall such that their associated grains are not outside of Y.

Lemma 4.1. *The conditional distribution of X^i given Y is that of a Poisson process on $\mathcal{D}$ with intensity λ, conditional on coverage of $\mathcal{C}$.*

We write π_λ for this conditional distribution of X^i given Y. Its density is

$$f_\lambda(\mathbf{x}^i) = \frac{\mathbf{1}_{\{\mathbf{x}^i \text{covers } \mathcal{C}\}} e^{(1-\lambda)|\mathcal{D}|}\, \lambda^{n(\mathbf{x}^i)}}{P_\lambda(\mathcal{C} \text{ is covered})} \tag{4.3}$$

where $P_\lambda(\cdot)$ denotes the distribution of a Poisson process with intensity λ on $\mathcal{D}$.

Proof The Janossy density (Daley and Vere–Jones, 1988) at any $\mathbf{x}^i \cup \mathbf{x}^b$ satisfying $\mathcal{B}(\mathbf{x}^i \cup \mathbf{x}^b) \cap W = Y$ is given by

$$\begin{aligned} \frac{e^{-|W \oplus rB|}}{(n(\mathbf{x}^i)+n(\mathbf{x}^b))!} p(\mathbf{x}^i \cup \mathbf{x}^b) &= \frac{e^{-|W \oplus rB|}}{(n(\mathbf{x}^i)+n(\mathbf{x}^b))!} e^{(1-\lambda)|W \oplus rB|} \lambda^{n(\mathbf{x}^i)+n(\mathbf{x}^b)} \\ &= \frac{e^{-\lambda|W \oplus rB|}}{(n(\mathbf{x}^i)+n(\mathbf{x}^b))!} \lambda^{n(\mathbf{x}^i)}\, \lambda^{n(\mathbf{x}^b)}. \end{aligned}$$

Hence, the Radon–Nikodym density is proportional to $\lambda^{n(\mathbf{x}_i)}$ for any $\mathbf{x}_i \subseteq \mathcal{D}$ covering $\mathcal{C}$. Since $\mathbf{x}_i \subseteq \mathcal{D}$, it is natural to replace the reference distribution of a unit rate Poisson process on $W \oplus rB$ by that of a unit rate Poisson process on $\mathcal{D}$. Noting that the probability of covering $\mathcal{C}$ by balls centered at the points of a Poisson process on $\mathcal{D}$ with intensity λ equals

$$P_\lambda(\mathcal{C} \text{ is covered}) = \sum_{n=0}^{\infty} \frac{e^{-\lambda|\mathcal{D}|}}{n!} \lambda^n \int_{\mathcal{D}} \cdots \int_{\mathcal{D}} \mathbf{1}_{\{\{e_1,\ldots,e_n\} \text{covers } \mathcal{C}\}}(e) \mathrm{d}e_1 \ldots \mathrm{d}e_n$$

it follows that the normalizing constant is $\frac{e^{(1-\lambda)|\mathcal{D}|}}{P_\lambda(\mathcal{C} \text{ is covered})}$. □

The covering probability $P_\lambda(\mathcal{C}$ is covered) is typically impossible to compute, see Hall (1988) on covering problems. Because of this, direct sampling from π_λ is also impossible. An alternative is of course rejection sampling: simply generating independent Poisson processes of intensity λ until one of them satisfies the covering condition. Unfortunately, depending on λ and $\mathcal{C}$ the probability of success, $P_\lambda(\mathcal{C}$ is covered), may be too small.

It is important for us to note that $P_\lambda(\mathcal{C}$ is covered) is strictly increasing in λ. This follows from the fact that a Poisson process of intensity $\kappa > \lambda$

can be written as the independent superposition of Poisson processes with intensity λ and $\kappa - \lambda$ respectively. So we see that if we choose κ large enough we can quite easily obtain a sample from π_κ. In the next section we use this fact to construct a method to sample from π_λ by thinning a sample from π_κ for an arbitrary $\kappa > \lambda$.

4.3 Coupling from the past for the conditional Boolean model

Recall that an ordinary stationary Poisson point process of intensity $\lambda > 0$ on a compact set S arises as the stationary distribution of a spatial birth-and-death process (Preston 1977). We start with any finite collection of points on S. Then each point is deleted after an exponential lifetime with rate 1. Also, after exponential waiting times with mean $1/\lambda$ new points are added which are uniformly distributed on S. The sequence of point configurations thus obtained converges weakly to a Poisson point process on S of intensity λ. Lantuéjoul (1997) demonstrates that similarly we can obtain a Poisson process point process, conditioned on a certain event $\mathcal{E}$ of positive probability, as the limit of a birth and death process. Informally, if whenever a point is added or deleted according to the scheme described above, we make sure never to enforce a transition violating $\mathcal{E}$, the stationary distribution of the resulting process is the conditional version of the Poisson process.

In our case, $S = \mathcal{D}$ and $\mathcal{E}$ is the event that $\mathcal{C}$ is covered. Let $X(t)$ be a birth and death process with birth rate λ and individual death rate 1. A birth is understood to be the addition of a uniform point on $\mathcal{D}$. Let $X(t^-)$ denote the configuration at time t^- just before t. If a point z_i is born at time t it is added: $X(t) = X(t^-) \cup \{x\}$. If $x \in X(t^-)$ dies at time t it is removed only if that does not cause part of $\mathcal{C}$ to become uncovered, that is if

$$B(x, r) \cap \mathcal{C} \subseteq \mathcal{B}(X(t^-) \setminus \{x\}) \cap \mathcal{C}. \tag{4.4}$$

If x cannot be removed, then it is granted an extra exponential lifetime after which its removal is reconsidered. Following the terminology in (Kendall and Thönnes, 1999) x is then called *perpetuated*.

Following Lantuéjoul (1997) it would not be difficult to show that the stationary distribution of $X(t)$ is indeed π_λ. However, $X(t)$ is *not* the birth and death process we shall use to sample from π_λ. We want to apply coupling from the past (section 1.5.1) to obtain perfect samples from π_λ but $X(t)$ does not appear amenable to this approach.

The difficulty is that for perfect sampling from the stationary distribution of a Markov chain it is convenient if the state-space admits a partial order

and there exist maximal and minimal elements majorizing and minorizing all other elements with respect to this partial order. The state space of $X(t)$ consists of all finite collections of points on $\mathcal{D}$. The obvious (partial) ordering of this space is the inclusion ordering, but then there is certainly no maximal state, because $\mathcal{D}$ itself is infinite.

We shall construct a different birth-and-death process on a state space consisting of all subsets of a finite (but random) set of points which also converges to π_λ. We shall be able to apply coupling from the past to this process.

It should be remarked that for the special case where $\mathcal{E}$ is the event that a certain finite collection of points is covered, a perfect sampling algorithm is available due to Kendall and Thönnes (1998). Unfortunately, their method does not seem to extend to our situation: coverage of a uncountable set. However, some of the ideas in (Kendall and Thönnes, 1998) play a role in our construction also.

As noted previously, rejection sampling from (4.3) is possible for large intensity parameters. Using this observation, the first step in our algorithm to sample from a given π_λ is to choose a $\kappa > \lambda$ and to generate a sample, say $D = \{z_1, z_2, \ldots, z_n\} \subset \mathcal{D}$, from π_κ. D will serve as maximal state. All configurations obtained when running the birth-and-death processes will be subsets of this finite set D. Now D is thinned, independently retaining each point with probability λ/κ.

The key result is the following.

Proposition 4.1. *Let D have distribution π_κ and define E_0 to be the independent thinning of D with retention probability λ/κ. Conditionally on coverage of $\mathcal{C}$, E_0 is distributed according to π_λ (cf. (4.3)).*

Proof Note that the conditional distribution of E_0 given coverage of $\mathcal{C}$ is concentrated almost surely on configurations of points of $\mathcal{D}$. For any such configuration $\mathbf{x}$, the Janossy density is given by

$$\begin{aligned} & j_n(\mathbf{x}) \\ &= \sum_{n=0}^{\infty} \frac{e^{-|\mathcal{D}|}}{n!} \int \ldots \int_{\mathcal{D}^n} f_\kappa(\mathbf{x} \cup \{y_1, \ldots, y_n\}) \left(\frac{\lambda}{\kappa}\right)^{n(\mathbf{x})} \left(1 - \frac{\lambda}{\kappa}\right)^n \mathrm{d}y_1 \ldots \mathrm{d}y_n \\ &= \sum_{n=0}^{\infty} \frac{e^{-|\mathcal{D}|}}{n!} \alpha_\kappa e^{(1-\kappa)|\mathcal{D}|} \lambda^{n(\mathbf{x})} \left[\kappa(1 - \frac{\lambda}{\kappa})\right]^n |\mathcal{D}|^n = \alpha_\kappa \lambda^{n(\mathbf{x})} e^{-\lambda|\mathcal{D}|}, \end{aligned}$$

where $\alpha_\kappa^{-1} = P_\kappa(\mathcal{C}$ is covered) is the probability that balls of radius r around the points of a Poisson process with intensity κ cover $\mathcal{C}$. The claim follows upon normalization and comparison to (4.3). □

We shall construct a coupling-from-the-past (CFTP) algorithm that selects a configuration Z between E_0 and D ($E_0 \subseteq Z \subseteq D$) such that Z is distributed according to π_λ.

First we construct a spatial birth-and-death process $E = \{E(t),\ t \geq 0\}$, such that $E(t) \subseteq D = \{z_1, z_2, \ldots, z_n\}$ for all t. We set $E(0) = E_0$. The dynamics of the process are such that each point is removed from $E(\cdot)$ after an exponentially distributed lifetime with mean 1. Whenever a point is removed, it is added again after an exponentially distributed waiting time of mean $(\kappa - \lambda)/lambda$. All life– and waiting times are independent of each other and everything else. Points $z_i \in E(0)$ are treated as if they were added at time 0, those $z_i \notin E(0)$ as removed at time 0.

Lemma 4.2. *Conditionally on D, the spatial birth-and-death process $E(t)$, $t \geq 0$, is irreducible, homogeneous, and positive recurrent. Moreover, $E(t)$ is in equilibrium and time-reversible.*

In particular $E(t)$ is distributed as $E(0) = E_0$ for all $t \leq 0$.

Proof As $E(\cdot)$ is defined conditionally on $D_0 = \{z_1, z_2, \ldots, z_n\}$, at any time there are only a finite number of points. Clearly, the transition rates do not change in time, hence $E(\cdot)$ is homogeneous. Moreover, any state $\mathbf{x} \subseteq D$ can be reached from any other state $\mathbf{x}'$ (say) by successively deleting the points in $\mathbf{x}'$ followed by addition of the points in $\mathbf{x}$. Therefore, the birth-and-death process is well-defined and possesses a stationary distribution (Parzen, 1962, Chapter 7). Moreover, since the state space is finite, not all stationary probabilities can be zero, hence they are all positive and the $E(\cdot)$ is positive recurrent.

The process $E(\cdot)$ is in equilibrium and time reversible because it satisfies 'detailed balance' with respect to the distribution (given D) of $E(0) = E_0$

$$\frac{\lambda}{\kappa - \lambda} P(E_0 = \mathbf{x}|D) = P(E_0 = \mathbf{x} \cup \{z_i\}|D),$$

where $z_i \notin \mathbf{x} \subseteq D$. □

We now define a spatial birth-and-death process $Z(t)$, $t \geq 0$, as a conditional version of $E(\cdot)$ in the spirit of Lantuéjoul (1997). The equilibrium distribution of $Z(\cdot)$ will be π_λ.

Start with $Z(0)$ being any subset of D_0. The dynamics of the process are such that a point z_i *might be* removed from $Z(\cdot)$ after an exponentially distributed lifetime of mean 1. The point z_i is *actually* removed only if

$$B(z_i, r) \cap \mathcal{C} \subseteq \mathcal{B}(Z(t^-) \setminus \{z_i\}) \cap \mathcal{C}. \tag{4.5}$$

If z_i cannot be removed, then it is granted an extra exponential lifetime with mean 1 after which its removal is reconsidered. Whenever a point *is* removed, it is added again after an exponentially distributed waiting time of mean $(\kappa - \lambda)/\lambda$. All life– and waiting times are independent of each other and everything else. Points $z_i \in Z(0)$ are treated as if they were added at time 0, those $z_i \notin Z(0)$ as removed at time 0.

Proposition 4.2. *Conditionally on D, the spatial birth-and-death process $Z(t)$, $t \geq 0$, is homogeneous, and has a single positive recurrent class consisting of those subsets $\mathbf{z}$ of D for which $\mathcal{B}(\mathbf{z})$ covers $\mathcal{C}$. It tends in distribution to π_λ as $t \to \infty$.*

Proof We work conditionally on $D = \{z_1, z_2, \ldots, z_n\}$. First note that almost surely $\mathcal{B}(Z(t))$ will cover $\mathcal{C}$ for some $t \geq 0$. The transition mechanism ensures that $\mathcal{B}(Z(s)) \supseteq \mathcal{C}$ for all $s \geq t$. Moreover, the class C of configurations $\mathbf{z} \subseteq D$ whose associated Boolean model covers $\mathcal{C}$ is irreducible, since any state $\mathbf{z}$ can be reached from any other state $\mathbf{z\prime}$ by successively adding all points of $D \setminus \mathbf{z\prime}$, then deleting those of $D \setminus \mathbf{z}$. Thus $Z(\cdot)$ is well-defined, with a stationary distribution concentrated on C (Parzen, 1962). Again, since the state space is finite, not all stationary probabilities can be zero, hence they are all positive and the C is positive recurrent.

Consider a move at time t of $Z(\cdot)$ from a configuration $\mathbf{x} \cup z_i \subseteq D$ to $\mathbf{x}$, for which both $\mathcal{B}(\mathbf{x})$ and $\mathcal{B}(\mathbf{x} \cup z_i)$ cover $\mathcal{C}$ and $z_i \notin \mathbf{x}$. The birth rate of $Z(\cdot)$ is the same as the birth rate of $E(\cdot)$. The death rate for both perpetuated and non-perpetuated points is 1, as it is for E, provided the coverage condition is not violated. Thus, the detailed balance conditions for $E(\cdot)$ and $Z(\cdot)$ coincide on C, from which it follows that $Z(t)$ tends to the distribution of $E(0)$ restricted to C as $t \to \infty$. Using lemma 4.1 the result follows. □

We shall now describe how to apply coupling from the past to obtain a sample from the stationary distribution of $Z(\cdot)$. We already have a maximal state D.

Fix a time $-T < 0$. By lemma 4.2, $E(\cdot)$ is time-reversible, hence can easily be extended backwards from $E(0)$ until time $-T$. $E(\cdot)$ on the interval $[-T, 0]$ will be the minimal state.

It is straightforward to define a coupled process $Z_{-T}(t)$, $-T \leq t \leq 0$ which has the same dynamics as $Z(\cdot)$ while $E(t) \subseteq Z(t) \subseteq D$. First set $Z_{-T}(-T) = E(-T)$. Now with each $z_i \in D$ associate a unit rate Poisson process Ξ_i on the set $\{-T \leq t \leq 0 : z_i \notin E(t)\}$. These Ξ_i will govern possible deaths of perpetuated points. At the (forward) birth in the process $E(\cdot)$ at time t of a point z_i, add z_i to $Z(t^-)$ (if not already present). At the (forward)

death in the process $E(\cdot)$ at time t of a point z_i, remove z_i from $Z(t^-)$ but only if (cf. 4.5)

$$B(z_i, r) \cap \mathcal{C} \subseteq \mathcal{B}(Z_{-T}(t^-) \setminus \{z_i\}) \cap \mathcal{C}. \tag{4.6}$$

If z_i cannot be removed at time t, then its removal is reconsidered at the next event time after t of the process Ξ_i.

Lemma 4.3. *$Z_{-T}(0)$ tends in distribution to π_λ as $T \to \infty$.*

Proof $Z_{-T}(\cdot)$ has the same dynamics and hence the same stochastic properties as the process $Z(\cdot)$ considered in the beginning of this section. Hence, this lemma follows directly from Proposition 4.2. □

To check if $Z_{-T}(0)$ has reached equilibrium—so to say—we need processes $L_{-T}(\cdot)$ and $U_{-T}(\cdot)$ such that (cf. Kendall and Møller (1999) formulas (3.5) to (3.7))

1. ('sandwiching') $E(t) \subseteq L_{-T}(t) \subseteq U_{-T}(t) \subseteq D_0$ for all $-T \leq t \leq 0$;
2. ('funneling') $L_{-T}(t) \subseteq L_{-S}(t) \subseteq U_{-S}(t) \subseteq U_{-T}(t)$, for all $-S \leq -T \leq t \leq 0$;
3. ('coalescing') if, for some s, $L_{-T}(s) = U_{-T}(s)$ then $L_{-T}(t) = U_{-T}(t)$, for all $t \geq s$.

Moreover, we need that $L_{-T}(0) = U_{-T}(0)$ almost surely for finite (preferably small) T. If the above inclusions hold and the latter event occurs then we can conclude that $L_{-T}(0) = U_{-T}(0)$ is a perfect sample from π_λ. This is roughly the content of Theorem 4.1 below.

Now the construction of the processes $L_{-T}(\cdot)$ and $U_{-T}(\cdot)$ is very similar to that of $Z_{-T}(\cdot)$. The only difference is that removal (if present) of a point z_i at time t from $L_{-T}(t^-)$ is enforced only if

$$B(z_i, r) \cap \mathcal{C} \subseteq \mathcal{B}(U_{-T}(t^-) \setminus \{z_i\}) \cap \mathcal{C}. \tag{4.7}$$

and the same removal from $U_{-T}(t^-)$ is enforced only if

$$B(z_i, r) \cap \mathcal{C} \subseteq \mathcal{B}(L_{-T}(t^-) \setminus \{z_i\}) \cap \mathcal{C}. \tag{4.8}$$

The—at first perhaps mystifying—fact that these processes pay attention to each other is necessary to ensure the above inclusions. This 'cross-over' trick has also been applied by Häggström and Nelander (1998) and Kendall (1997).

Summarizing, we propose the following algorithm.

Algorithm 4.1. *Generate a random sample D from π_κ for some $\kappa > \lambda$ and delete each point independently with probability $1-(\lambda/\kappa)$ to obtain $E(0)$. Set $T = 1$ and write $[\frac{T}{2}]$ for the integer part of $\frac{T}{2}$.*

- *extend $E(\cdot)$ backwards on $[-T, -[\frac{T}{2}])$ with birth rate $\lambda/(\kappa-\lambda)$ and death rate 1;*
- *extend independent unit rate Poisson processes Ξ_i backwards on $\{-T \leq t < [\frac{T}{2}] : E_i(t) = 0\}$;*
- *set $L_{-T}(-T) = E(-T)$ and $U_{-T}(-T) = D$;*
- *at a birth transition $E(t^+) = E(t) \cup \{z_i\}$, add z_i to $L_{-T}(t)$ and $U_{-T}(t)$.*
- *at a death transition $E(t^+) = E(t) \setminus \{z_i\}$ or an event time of Ξ_i*
  - *delete (if present) z_i from $L_{-T}(t)$ provided that does not cause the Boolean model associated with $U_{-T}(t)$ to uncover part of $\mathcal{C}$, i.e. z_i may be deleted only if (4.8) holds;*
  - *delete (if present) z_i from $U_{-T}(t)$ provided that does not cause the Boolean model associated with $L_{-T}(t)$ to uncover part of $\mathcal{C}$, i.e. z_i may be deleted only if (4.7) holds;*
- *if $L_{-T}(0) = U_{-T}(0)$ exit; otherwise double T and repeat.*

Algorithm 4.1 is designed in so that the inclusion relations 1, 2 and 3 hold.

Lemma 4.4. *The processes $E(t)$, $L_{-T}(t)$, $U_{-T}(t)$ ($t \leq 0$) and D satisfy the inclusion relations 1, 2 and 3.*

Proof By definition, $E(-T) - L_{-T}(-T) \subseteq U_{-T}(-T) = D$, hence the sandwiching property holds for $t = -T$. Also D is a supset of all other sets. Since births in $E(\cdot)$ are reciprocated in $L_{-T}(\cdot)$ and $U_{-T}(\cdot)$, the inclusion relationship is preserved under birth transitions. Next, consider the death at some time $t \in [-T, 0]$, say $E(t^+) = E(t) \setminus \{z_i\}$ or an event time t of Ξ_i. Suppose that $E(t) \subseteq L_{-T}(t) \subseteq U_{-T}(t)$. Since Ξ_i is restricted to the set $\{t \leq 0 : E_i(t) = 0\}$, $E(t^+)$ is a subset of $L_{-T}(t^+)$ and $U_{-T}(t^+)$. Furthermore, if z_i dies in the upper process,

$$B(z_i, r) \cap \mathcal{C} \subseteq \mathcal{B}(L_{-T}(t) \setminus \{z_i\}) \cap C \subseteq \mathcal{B}(U_{-T}(t) \setminus \{z_i\}) \cap C$$

and consequently z_i also dies in the lower process.

Turning to the funneling property, we have to show that $L_{-T}(t) \subseteq L_{-S}(t)$ and that $U_{-S}(t) \subseteq U_{-T}(t)$. Now, by definition $E(-S) = L_{-S}(-S)$. Since the dynamics of algorithm 4.1 preserve inclusion, it follows that $L_{-T}(-T) = E(-T) \subseteq L_{-S}(-T)$ and, more generally, $L_{-T}(t) \subseteq L_{-S}(t)$ for any $t \geq -T$. Regarding the upper process, $U_{-T}(-T) = D \supseteq U_{-S}(-T)$. Applying once more the fact that the algorithm preserves the inclusion ordering yields $U_{-T}(t) \supseteq U_{-S}(t)$ for any $t \geq -T$.

Finally, suppose that the upper and lower processes meet at some time $s \leq 0$. Then, as they are coupled by the same E– and Ξ_i-processes, they proceed as one. □

We are now ready to state the main result of this section.

Theorem 4.1. *Algorithm 4.1 almost surely terminates in finite time; its output has density $f_\lambda(\cdot)$ (cf. (4.3)) with respect to a unit rate Boolean model on $\mathcal{D}$.*

Proof Note that $P(E(0) = D|D) = \left(\frac{\lambda}{\kappa}\right)^{n(D)}$, hence

$$P(E(0) = D) = E\left[\left(\frac{\lambda}{\kappa}\right)^{n(D)}\right]$$

where $n(D)$ is Poisson distributed with mean $\kappa|\mathcal{D}|$. Consequently, the event $\{E(0) = D\}$ has strictly positive probability. By lemma 4.1, $E(-T) = D$ will occur for some T almost surely. Hence, using Lemma 4.4, the algorithm terminates almost surely in finite time.

As stated by Lemma 4.3 $Z_{-T}(0)$ tends in distribution to π_λ as $T \to \infty$. Moreover, using the fact that Algorithm 4.1 preserves the inclusions ordering, $L_{-T}(t) \subseteq Z_{-T}(t) \subseteq U_{-T}(t)$ for all $-T \leq t \leq 0$.

Suppose T_0 is a (random) time such that $L_{-T_0}(0) = U_{-T_0}(0)$. It follows that $Z_{-T}(0) = L_{-T_0}(0) = U_{-T_0}(0)$ for all $-T \leq T_0$. Hence, $L_{-T_0}(0) = U_{-T_0}(0)$ is a sample from π_λ.

This completes the proof. □

To conclude this section we present a modification of algorithm 4.1, which not only speeds it up but also facilitates and even makes possible some interesting estimation methods. These estimation methods will be discussed in the following section.

First, we note that the algorithm typically takes a long time to terminate when the difference between the dominating pattern D and its thinning E_0 is large. For example, suppose we want to sample from π_1 while rejection

sampling only allows us to sample from π_{100} to obtain D. This is clearly a 'hard' problem. E_0 would on average contain only one hundredth of the points of D so the algorithm has many configurations between D and $E(0)$ to choose from—so to speak.

The idea is to first use algorithm 4.1 to reduce $D \sim \pi_{100}$ to a sample, say X_{99}, from π_{99}. Then use algorithm 4.1 again with $D = X_{99}$ to obtain a sample X_{98} from π_{98}. Repeating this, we obtain a *nested* sequence of samples $X_{99} \supseteq X_{98}, \cdots \supseteq X_1$ from $\pi_{99},\ \pi_{98}, \ldots, \pi_1$.

This will be much faster than trying to sample from π_1 straight away. In fact, for small steps of λ it might well happen that D and E_0 coincide, in which case there is clearly no need to run the algorithm. Also, it might happen that there is no pattern X with $E_0 \subseteq X \subset D$ which meets the coverage requirement. Again, there would be no need to run the algorithm at because the output must be D itself.

With an estimation method called 'perfect stochastic EM' (section 1.5.3) in mind, we now present an algorithm that produces a nested continuum of samples $\{X_\lambda \sim \pi_\lambda : \underline{\lambda} \leq \lambda \leq \overline{\lambda}\}$ for any two values $0 < \underline{\lambda} \leq \overline{\lambda}$. By 'nested' we mean that $X_\lambda \subset X_{\lambda'}$ if $\lambda \leq \lambda'$. The idea is to slowly thin an initial sample form $\pi_{\overline{\lambda}}$, removing one point at a time.

We suppose (without any loss of generality) that it is feasible to sample from $\pi_{\overline{\lambda}}$. Denote $\overline{\lambda} = \lambda(0)$. Let $X_{\lambda(0)} = \{x_1, \ldots, x_n\}$ be a sample from $\pi_{\lambda(0)}$. Now associate with each x_i a standard uniform random variable U_i^1; all the U_i^1 being independent. Define

$$E_\lambda^1 = \{x_i \in X_{\lambda(0)} : U_i^1 < \lambda/\lambda(0)\} \qquad \lambda \leq \lambda(1)$$

Note that for each λ, E_λ^1 is a thinning of $X_{\lambda(0)}$ with retention probability $\lambda/\lambda(0)$. Conceptually, we could—for every λ—run algorithm 4.1 with $D = X_{\lambda(0)}$ and $E_0 = E_\lambda^1$ to obtain a sample from π_λ.

Define $\lambda(1) = \lambda(0) \max_i U_i^1$. Then $E_\lambda^1 = X_{\lambda(0)}$ for all $\lambda(1) < \lambda \leq \lambda(0)$, while $E_{\lambda(1)}^1 = X_{\lambda(0)} \setminus \{x_i\}$ for some i.

Now set

$$X_\lambda = X_{\lambda(0)} \qquad \lambda(1) < \lambda \leq \lambda(0).$$

It is clear that these X_λ are distributed according to π_λ; application of algorithm 4.1 with $D = X_{\lambda(0)}$ and $E_0 = E_\lambda^1$ must output $X_{\lambda(0)}$ because $D = E_0$.

Next, run algorithm 4.1 with $D = X_{\lambda(0)}$ and $E_0 = E_{\lambda(1)}^1 = X_{\lambda(0)} \setminus \{x_i\}$ to obtain a sample $X_{\lambda(1)}$ from $\pi_{\lambda(1)}$. Note that when $E_{\lambda(1)}^1$ does not cover, the algorithm must output $X_{\lambda(0)}$ and hence need not be run.

Now repeat the entire procedure, starting with $X_{\lambda(1)}$ instead of $X_{\lambda(0)}$. We must associate new uniform random variables U_i^2 with the points of $X_{\lambda(1)}$ and

define new thinnings

$$E^2_\lambda = \{x_i \in X_{\lambda(1)} : U^2_i < \lambda/\lambda(1)\} \qquad \lambda \le \lambda(1)$$

Note that the distributions of E^1_λ and E^2_λ are the same for all $\lambda \le \lambda(1)$. Hence we are indeed starting afresh.

Repeating again and again in this manner, we obtain a sequence $\lambda(0) > \lambda(1) > \lambda(2) \dots$ and after n repetitions we have a nested continuum of samples $\{X_\lambda \sim \pi_\lambda : \lambda(n) \le \lambda \le \lambda(0) = \overline{\lambda}\}$.

Define N to be the smallest number such that $\lambda(N) < \underline{\lambda}$. It is not difficult to see that N is almost surely finite. Hence the above method provides, almost surely in finite time, nested samples from π_λ for all λ between $\underline{\lambda}$ and $\overline{\lambda}$.

4.4 Maximum likelihood

In general, direct maximum likelihood estimation of the intensity parameter in a Boolean model (Definition 4.1) seems very hard. In this section, we will describe two alternative techniques.

4.4.1 MCMC approach

Suppose a Boolean model $\mathcal{B}(X)$ of discs with radius r is observed though a non-empty compact set W. We write $Y = \mathcal{B}(X) \cap W$. The goal is to estimate the intensity parameter λ of the underlying Poisson germ process X. Although the likelihood of Y is known, it involves a normalizing constant which we cannot compute. A solution is to use the approach in (Geyer and Thompson, 1992, Geyer, 1998) and consider the likelihood ratio

$$\frac{p(Y|\lambda)}{p(Y|\kappa)} = e^{(\kappa-\lambda)|W \oplus rB|} \frac{c(\lambda|Y)}{c(\kappa|Y)}$$

with respect to a fixed parameter $\kappa > 0$. Note that the normalizing constant $c(\lambda|Y) = E\left[\lambda^{n(X)} \mathbf{1}_{\{\mathcal{B}(X)\cap W = Y\}}\right]$ where the expectation is taken with respect to a unit rate Poisson process. Note that

$$\frac{c(\lambda|Y)}{c(\kappa|Y)} = E_\kappa \left[\left(\frac{\lambda}{\kappa} \right)^{n(X)} \middle| \mathcal{B}(X) = Y \right].$$

Hence the log likelihood ratio can be rewritten as

$$\ell(\lambda) = \log \frac{p(Y|\lambda)}{p(Y|\kappa)} = (\kappa - \lambda)|W \oplus rB| + \log E_\kappa \left[\left(\frac{\lambda}{\kappa} \right)^{n(X)} \middle| \mathcal{B}(X) \cap W = Y \right]. \tag{4.9}$$

The expectation can be estimated by an average over independent realizations of the conditional distribution of X given $\mathcal{B}(X) \cap W = Y$ (under parameter κ) (see section 4.3).

The approximation works well if the reference value κ is not too far from the true value. To try to make sure of this, we could use a pilot estimate as κ. Alternatively, we can take a sequence of different parameter values $\lambda < \lambda_1 <, \ldots, < \lambda_n < \kappa$. Noting that

$$\ell(\lambda) = \log \frac{p(Y|\lambda)}{p(Y|\lambda_1)} + \log \frac{p(Y|\lambda_1)}{p(Y|\lambda_2)} + \cdots + \log \frac{p(Y|\lambda_{n-1})}{p(Y|\lambda_n)} + \log \frac{p(Y|\lambda_n)}{p(Y|\kappa)}$$

we need samples from $\pi_{\lambda(i)}$ for $i = 1, 2, \ldots, n$. We could (perhaps should) produce these samples independently, but we can also start with a sample from $\pi_{\lambda(n)}$ and then successively work our way down. When sampling from $\pi_{\lambda(i)}$ we can then take the sample from $\pi_{\lambda(i+1)}$ as dominating state D.

4.4.2 EM

The EM-algorithm (Dempster et al. 1977) is an iterative technique to approximately solve the likelihood equations for missing data problems by alternating expectation and maximization steps. In general, let X be the unobserved (complete) data, and Y the observed data obtained by a many-to-one mapping $Y = Y(X)$ and assume that X has a density $p(\mathbf{x}; \lambda)$ depending on a parameter $\lambda > 0$. Starting from any initial value $\lambda(0)$, construct a sequence $\lambda(0), \lambda(1), \ldots$ by repeating the following.

Expectation-step compute as the conditional expectation under $\lambda(k)$ of the complete data log-likelihood, given the observed data

$$E_{\lambda(k)} \left[\log p(\lambda; X) \mid Y \right]; \tag{4.10}$$

Maximization-step find $\lambda(k+1)$ by maximizing (4.10) with respect to λ.

In our set-up, $Y = \mathcal{B}(X) \cap W$ and $p(\mathbf{x}; \lambda)$ is the Poisson(λ) density. The (complete data) log-likelihood is linear in the sufficient statistic $n(X)$. Hence the expectation and maximization steps amount to computing the conditional expectation under $\lambda(k)$ of the complete data maximum likelihood estimator given the observed data, that is

$$\lambda(k+1) = E_{\lambda(k)} \left[\frac{n(X)}{|W \oplus rB|} \mid Y \right] = E_{\lambda(k)} \left[\frac{n(X^i) + n(X^b)}{|W \oplus rB|} \mid Y \right] \tag{4.11}$$

We cannot compute (4.11), as required by the EM-algorithm. However, since we can sample from the conditional distribution of X given Y, we can use Monte Carlo methods to approximate it. Suppose that at the k-th iteration step we use m samples to do so. The average of these estimates $E_{\lambda(k)}[n(X)/|W \oplus rB| \mid Y]$. This estimate is the next parameter value, $\lambda(k+1)$. At the next iteration we need a new set of m independent samples under $\lambda(k+1)$. This algorithm is known as the Monte Carlo EM or, if $m = 1$, the stochastic EM-algorithm (StEM). We discussed this algorithm in section 1.5.2.

We now specialize to StEM, i.e. the case $m = 1$. Usually, the algorithm is run for a great number of steps to allow it to approach equilibrium (burn-in). After that, the chain is continued for many more iterations and its steps are averaged to bring down the variance of the estimator. Since the discovery of CFTP, the arbitrary burn-in is unsatisfactory. Here we demonstrate how we can use CFTP to generate a StEM chain that is actually guaranteed to be in equilibrium. We need one mild assumption, namely that the true λ is known to be less than some finite $\overline{\lambda}$. The complete data maximum likelihood estimator is slightly modified to become $n(X)/|W \oplus rB| \wedge \overline{\lambda}$.

We might as well (and do) assume that $\overline{\lambda}$ is so big that we can easily sample from $\pi_{\overline{\lambda}}$. Of course $\overline{\lambda}$ will be our 'maximal state'. As $n(X^i)$ is obviously bounded below by $|\mathcal{C}|/(\pi r^2)$, there is a very natural minimal state:

$$\underline{\lambda} = \frac{|\mathcal{C}|/(\pi r^2) + n(X^b)}{|W \oplus rB|}.$$

Thus we may restrict the parameter space to $[\underline{\lambda}, \overline{\lambda}]$.

Using the sampling scheme described at the end of section 4.3 we can obtain *nested* samples X_λ for *all* $\lambda \in [\underline{\lambda}, \overline{\lambda}]$. With such a collection of samples we can define a random transition maps (cf. section 1.5.1) $H : [\underline{\lambda}, \overline{\lambda}] \to [\underline{\lambda}, \overline{\lambda}]$ by

$$H(\lambda) = \frac{n(X_\lambda) + n(X^b)}{|W \oplus rB|} \wedge \overline{\lambda}.$$

Since the X_λ are nested, $n(X_\lambda)$ is ('surely') monotone increasing as a function of λ and hence so is H. Coupling-from-the-past works as follows here. Fix an integer $-T < 0$. For $t = -T+1, -T+2, \ldots, 0$ generate independent copies N_t of N. Define an upper chain $U_{-T} = \{U_{-T}(t) : t = -T, \ldots, 0\}$ by setting $U_{-T}(t) = H_t(U_{-T}(t-1))$ and $U_{-T}(-T) = \overline{\lambda}$. Similarly, define a lower chain L_{-T} by $L_{-T}(t) = H_t(L_{-T}(t-1))$ and $L_{-T}(-T) = \underline{\lambda}$.

At time 0, check if $U_{-T}(0) = L_{-T}(0)$. If so terminate; if not repeat, starting at time $-2T$ and re-using the $\{H_t : t = -T+1, \ldots, 0\}$.

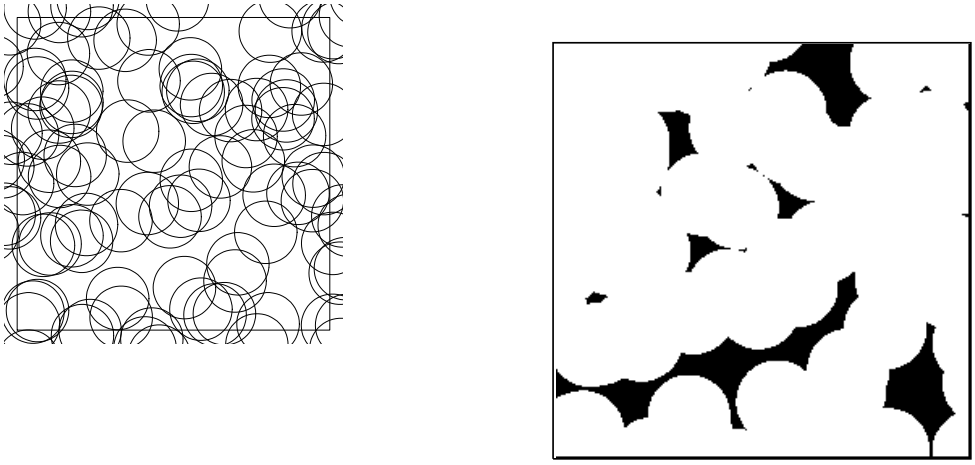

Figure 4.1: (a) Realization of a Poisson point process of intensity 75 on the set $[0,1]^2 \oplus rB$. There are 101 points. With each point a circle with radius 0.1 is associated. (b) The observed Boolean model $Y = \mathcal{B}(X) \cap [0,1]^2$

4.5 An example

In Figure 4.1 (a) we have a realization of a Boolean model of discs of radius .10 as observed through the set $[0,1]^2 \oplus (0.1)B$ (B is the unit disc). The intensity of the underlying germ process is 75. There were 101 points in the set $[0,1]^2 \oplus (0.1)B$, of which 56 could be identified from observing the Boolean model through $[0,1]^2$ as depicted in Figure 4.1 (b). Figure 4.2 shows the region $\mathcal{D}$ where the interior points 'live' and the region $\mathcal{C}$ to be covered.

We have applied the Monte Carlo likelihood ratio approach and the StEM algorithm to the data in Figure 4.1 (b). First, we estimated the log likelihood ratio (4.9) with respect to $\kappa = 80$ for $\lambda \in [50, 100]$. The estimate, shown in Figure 4.3, is based on 50 independent samples from π_{80}. To obtain such a sample we first generated a dominating pattern D. This was done by rejection sampling, where we gently increased the intensity until success. The first, rejection sampling attempts were done at intensity 80. After 50 failures, the intensity was increased by $10/|[0,1]^2 \oplus (0.1)B|$. Generally, success occurred at intensities around 110.

From figure 4.3 the log likelihood ratio is readily maximized. We find 69 as an estimate of the true intensity (75).

Of course the choice $\kappa = 80$ as reference value for the log likelihood ratio is quite arbitrary. Instead, we could have used some pilot estimate.

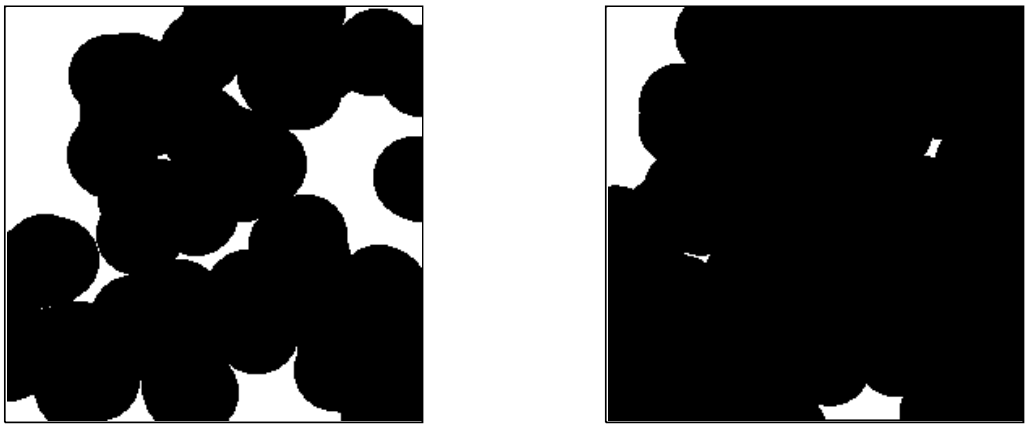

Figure 4.2: (a) The set $\mathcal{D}$ where the interior points 'live'. (b) the set $\mathcal{C}$ which must be covered by the Boolean model associated with the interior points.

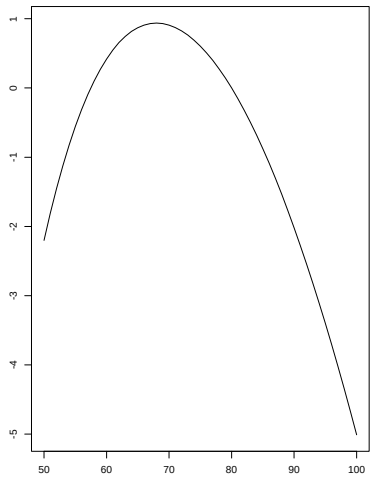

Figure 4.3: Estimated log likelihood ratio as a function of λ with respect to $\kappa = 80$.

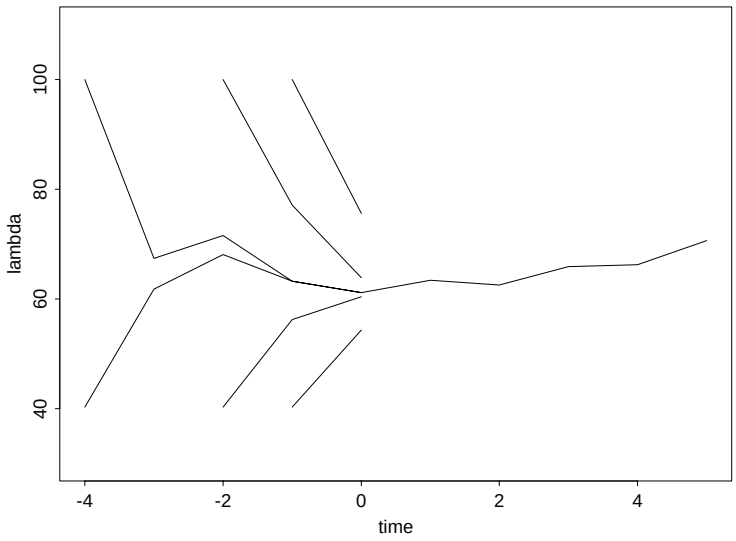

Figure 4.4: The upper and lower processes in the perfect stochastic EM algorithm, starting at times -1, -2 and -4. After the upper and lower processes have coalesced, the (then stationary) chain is continued.

A run of the perfect StEM algorithm is shown in figure 4.4. We see upper and lower processes $U_{-1}(t)$ and $L_{-1}(t)$ and $U_{-2}(t)$ and $L_{-2}(t)$ which have not met at time zero, and processes $U_{-4}(t)$ and $L_{-4}(t)$ which have. We have chosen to start the upper processes at $\overline{\lambda} = 100$, pretending that we know for a fact that the true intensity is below 100. Note how quickly the algorithm has terminated. The value at time zero (after termination) is 61. We could now continue to run the StEM chain starting at 61 at time zero and average the result to bring down the variance.

We should mention that it took us longer to generate Figure 4.4 than Figure 4.3. Although fewer samples were needed for the former, they were of lower intensity (40) than the samples required for the latter (80).

An advantage of the likelihood ratio approach is that it allows us to do much more than just estimate the intensity. Indeed, it is easy to estimate the expectation and variance of any function ϕ of the complete data. We use importance sampling ideas (Geyer, 1994). To this end, 50 independent samples were generated from the conditional distribution of interior points with at intensities $\kappa_i = 60.0 + 10i$ $(i = 0, 1, 2, 3)$. Hence, in all, 200 independent

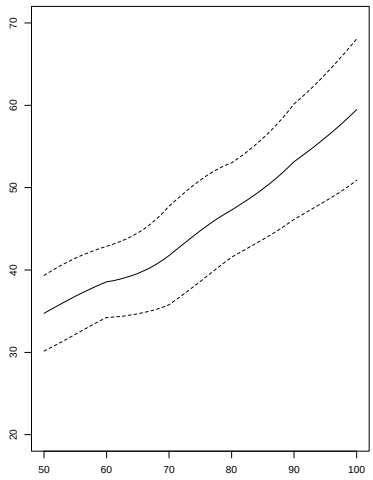

Figure 4.5: The solid line denotes a Monte Carlo estimate of $E_\lambda n(X^i)$ as a function of λ. The dashed lines show the (pointwise) estimated standard deviation.

samples were generated. An estimate for $E_\lambda \phi$ given by

$$s\,\frac{\sum_{j=1}^{50}\phi(X_{i,j})(\frac{\lambda}{\kappa_i})^{n(X_{i,j})}}{\sum_{j=1}^{50}(\frac{\lambda}{\kappa_i})^{n(X_{i,j})}} + (1-s)\,\frac{\sum_{j=1}^{50}\phi(X_{i+1,j})(\frac{\lambda}{\kappa_{i+1}})^{n(X_{i=1,j})}}{\sum_{j=1}^{50}(\frac{\lambda}{\kappa_{i+1}})^{n(X_{i=1,j})}}$$

if $\lambda = s\kappa_i + (1-s)\kappa_{i+1}, 0 \le s < 1$, and where $X_{i,j}$ is the jth sample from π_{κ_i}.

In our case, the expectation of $n(X^i)$ may be the most interesting. In Figure 4.5 we provide Monte Carlo estimates of the expected number of interior points under π_λ, as a function of λ, together with the (pointwise) standard deviation envelopes.

4.6 Sampling from general point processes

In a recent paper Kendall and Møller (1999) presented two perfect sampling algorithms for so-called hereditary (or more precisely: locally stable) point processes. One based on spatial birth-and-death processes and the other a Metropolis–Hastings algorithm. In this section we show that our algorithm 4.1 will generally allow us to sample from what one might call anti-hereditary point processes. We show that the modification we applied to algorithm 4.1 can also be used in this more general setup. Finally, we show that there

exists a similar modification of the birth-and-death algorithm of Kendall and Møller.

Following the notation and setup of Kendall and Møller (1999), let $(S, \mathcal{B}, \lambda)$ be a measure space. For us, λ will always be the uniform measure on S. Consider the 'Carter–Prenter exponential space' Ω of all finite collections of elements of S with its natural sigma-algebra $\mathcal{F}$ generated by sets $\{x \in \Omega : n(x \cap B) = n\}$ $(B \in \mathcal{B})$.

In particular we can take S to be a compact set in $\mathbb{R}^2$ and let $\mathrm{d}\lambda = \lambda \mathrm{d}s$, where $\mathrm{d}s$ is Lebesgue measure on S. Let Po_λ denote the probability measure on $(\Omega, \mathcal{F})$ corresponding to a Poisson process of intensity λ on S. The Poisson process of intensity λ is of course absolutely continuous with respect to the Poisson process of intensity 1. The density is $p_\lambda : \Omega \to [0, \infty)$ given by

$$p_\lambda(x) = e^{(1-S)|S|} \lambda^{n(x)} \tag{4.12}$$

Consider a point process X on S, whose distribution π_λ on $(\Omega, \mathcal{F})$ is absolutely continuous with respect to Po_λ with density $f : \Omega \to [0, \infty)$

$$\mathrm{d}\pi_\lambda = f \mathrm{dPo}_\lambda \tag{4.13}$$

In the previous sections we considered the special case $S = \mathcal{D}$ and

$$f(x) = \frac{\mathbf{1}_{\{x \text{ covers } \mathcal{C}\}}(x)}{\int \mathbf{1}_{\{x \text{ covers } \mathcal{C}\}}(x) \mathrm{dPo}_\lambda(x)}.$$

Defining

$$f_\lambda(x) = f(x) e^{(1-\lambda)|S|} \lambda^{n(x)} \tag{4.14}$$

we have

$$\mathrm{d}\pi_\lambda = f_\lambda \mathrm{dPo}_1.$$

Now consider the problem of obtaining a sample from π_λ. If f is bounded by, say, M we can apply *rejection sampling*. First we take a sample $X \sim \mathrm{Po}_\lambda$ and a uniform random variable U on $[0, M]$. We 'accept' X as a sample from π_λ if $U \leq f(X)$ and otherwise we repeat the procedure. Unfortunately, it often happens that the acceptance probability is so small that this sampling method would take forever. We shall now investigate various other approaches.

We consider two cases:

hereditary There is a constant $K > 0$ such that

$$f(x \cup \{\xi\}) \leq K f(x), \qquad \forall x \in \Omega,\ \xi \in S\ (\xi \notin x) \tag{4.15}$$

To be precise, the hereditary property usually (Kendall and Møller 1999, Geyer 1999) means: if $x \subseteq y$ then $f(x) > 0$ implies $f(y) > 0$. Condition (4.15) is really a 'local stability condition' which *implies* the hereditary property. The ratio $f(x \cup \{\xi\})/f(x)$ is sometimes called the 'Papangelou conditional intensity' and is denoted $\ell^*(x;\xi)$. Perfect sampling for locally stable point processes was considered by Kendall and Møller (1999).

anti-hereditary There is a constant $K > 0$ such that

$$f(x) \leq K f(x \cup \{\xi\}), \qquad \forall x \in \Omega,\ \xi \in S\ (\xi \notin x) \tag{4.16}$$

Straightforward rejection sampling is feasible in the hereditary case when λ is small. In fact, it works trivially for $\lambda = 0$ when we would (almost) always obtain the empty configuration. In the anti-hereditary case rejection sampling typically is feasible when λ is large. Therefore we shall assume throughout that we can easily obtain a sample from π_κ when we choose κ large enough.

4.6.1 spatial birth-and-death processes

We now discuss so-called spatial birth-and-death (b&d) processes as they were introduced by Preston (1975). We closely follow a brief review in Kendall and Møller (1999).

A spatial birth-and-death process X_t $(t \geq 0)$ is a random process taking its values in Ω, the collection of finite sets of points of S. The process is Markov and (hardly surprising) it has two kinds of transitions: births (addition of a point) and deaths (deletion of a point). Now suppose $X(\cdot)$ has 'birth rate' b and death 'rate' d. These are non-negative functions on $\Omega \times S$, while b must satisfy $B(x) = \int b(x,\xi)\mathrm{d}\lambda(\xi) < \infty$. The birth rate $b(x,\xi)$ determines the rate at which $X(\cdot)$ jumps from a state x to $x \cup \{\xi\}$ and the death rate $d(x,\xi)$ determines the rate at which it jumps from $x \cup \{\xi\}$ to x. More precisely, the dynamics of $X(\cdot)$ are as follows.

Suppose that $X_t = x = \{x_1, \ldots, x_n\}$. Let $E_0, E_1, \ldots, E_n$ denote independent exponential random variables with respective means $1/B(x), 1/d(x \setminus \{x_1\}, x_1), \ldots, 1/d(x \setminus \{x_n\}, x_n)$. The first transition after time t happens at time $t + E$, where E is the minimum of $E_0, \ldots, E_n$. If $E = E_0$ then the transition is a birth and a point ξ is added. The point ξ is selected according to the density $b(x,\xi)/B(x)$. If $E = E_i$, $i \neq 0$, the transition is a death and the point x_i is removed.

The following important lemma is from Preston (1975)

Lemma 4.5. *If there exists a density $g : \Omega \to [0, \infty)$ satisfying the so-called* detailed balance equations

$$g(x)b(x,\xi) = g(x \cup \{\xi\})d(x,\xi) > 0, \text{ whenever } g(x \cup \{\xi\}) > 0, \tag{4.17}$$

then X_t is time reversible and it has g as density of its unique stationary distribution.

Suppose that we have—as we do—a density f and wish to construct a b&d process with f as its equilibrium, then we can take $b(x,\xi) = f(x \cup \{\xi\})/f(x)$ and $d(x,\xi) = 1$ or alternatively $b(x,\xi) = 1$ and $d(x,\xi) = f(x)/f(x \cup \{\xi\})$. Of course these birth and death rates should be well-defined, and in this light conditions (4.15) and (4.16) above are not surprising.

4.6.2 the anti-hereditary case

Suppose we have a sample from an anti-hereditary point process with distribution π_κ. In this section we show how to obtain a sample from π_λ for any $\lambda \leq \kappa$ by means of thinning. This section generalizes our earlier work on sampling from the bombing model under a covering condition. The following lemma is a generalization of Proposition 4.1.

Lemma 4.6. *Suppose condition (4.16) holds. Let $\kappa > \lambda > 0$. Suppose Y is a sample from π_κ and let Y' be an independent thinning of Y with retention probability λ/κ. Then*

$$\int_y \frac{f(x)}{f(x \cup y)} P(Y' = x | Y = x \cup y)\pi_\kappa(x \cup \mathrm{d}y) = e^{-|S|} f_\lambda(x) \tag{4.18}$$

Proof By condition (4.16) we have that $f(x)/f(x \cup y)$ is bounded. Now it follows from (4.14) that

$$f_\kappa(x \cup y) = f(x \cup y)e^{(1-\kappa)|S|}\kappa^{n(x)+n(y)}$$

and

$$f(x) = f_\lambda(x)e^{(\lambda-1)|S|}\lambda^{-n(x)}.$$

Hence

$$\begin{aligned}
&\int_y \frac{f(x)}{f(x \cup y)} P(Y' = x | Y = x \cup y) \pi_\kappa(x \cup \mathrm{d}y) \\
&= \sum_{n=0}^{\infty} \frac{e^{-|S|}}{n!} \int \cdots \int_{S^n} \frac{f(x)}{f(x \cup \{y_1, \ldots, y_n\})} \left(\frac{\lambda}{\kappa}\right)^{n(x)} \left(1 - \frac{\lambda}{\kappa}\right)^n \\
&\times f_\kappa(x \cup \{y_1, \ldots, y_n\}) \mathrm{d}y_1 \ldots \mathrm{d}y_n \\
&= \sum_{n=0}^{\infty} \frac{e^{-|S|}}{n!} \int \cdots \int_{S^n} \frac{f_\lambda(x)}{f(x \cup \{y_1, \ldots, y_n\})} e^{(\lambda-1)|S|} \lambda^{-n(x)} \left(\frac{\lambda}{\kappa}\right)^{n(x)} \left(1 - \frac{\lambda}{\kappa}\right)^n \\
&\times f(x \cup \{y_1, \ldots, y_n\}) e^{(1-\kappa)|S|} \kappa^{n(x)+n} \mathrm{d}y_1 \ldots \mathrm{d}y_n \\
&= f_\lambda(x) e^{-|S|} e^{(\lambda-\kappa)|S|} \sum_{n=0}^{\infty} \frac{1}{n!} \kappa^n \left(1 - \frac{\lambda}{\kappa}\right)^n |S|^n \\
&= f_\lambda(x) e^{-|S|}
\end{aligned}$$

□

Suppose we have a sample $D = y = \{y_1, \ldots, y_n\}$ from π_κ and require a sample from π_λ for some $\lambda < \kappa$. Consider a spatial b&d process, $X(t)$, $t \geq 0$, on the powerset of y (denoted $\mathcal{P}(y)$) with birthrate $b(x, \xi) = 1_{\{\xi \in y \setminus x\}} \lambda/(\kappa - \lambda)$ and death rate $d(x, \xi) = f(x)/f(x \cup \{\xi\})$. Let D' denote an independent thinning of D with retention probability λ/κ. We have the following identity, for all $x \subset x \cup \{\xi\} \subseteq y$

$$\frac{\lambda}{\kappa - \lambda} \frac{f(x)}{f(y)} P(D' = x | D = y) = \frac{f(x)}{f(x \cup \{\xi\})} \frac{f(x \cup \{\xi\})}{f(y)} P(D' = x \cup \{\xi\} | D = y),$$

In other words, $X(t)$ satisfies detailed balance with respect to

$$\frac{f(x)}{f(y)} P(D' = x | D = y)$$

which is therefore the stationary distribution of $X(t)$, conditionally on $D = y$. Since D is distributed according to π_κ it follows from lemma 4.6 that *unconditionally* the stationary distribution of $X(t)$ is π_λ.

Now consider another spatial birth and death process, $E(t)$, $t \geq 0$, on $\mathcal{P}(y)$ with birth rate $b(x, \xi) = 1_{\{\xi \in y \setminus x\}} \lambda/(\kappa - \lambda)$ and death rate K. Let E_0 be an independent thinning of D with retention probability

$$\frac{\lambda/(K(\kappa - \lambda))}{1 + \lambda/(K(\kappa - \lambda))}.$$

This time we have, for all $x \subset x \cup \{\xi\} \subseteq y$

$$\frac{\lambda}{\kappa - \lambda} P(E_0 = x | D = y) = KP(E_0 = x \cup \{\xi\} | D = y),$$

We see that $E(t)$ satisfies detailed balance with respect to the conditional distribution of E_0 given D. It follows that $E(t)$ is time-reversible and in equilibrium if we set $E(0) = E_0$.

$X(t)$ and $E(t)$ have identical birth rates while by (4.16) the death rate of $E(t)$ is greater than that of $X(t)$. Setting $X(0) = E(0)$ it is possible to generate coupled realizations of $X(t)$ and $E(t)$ such that $E(t) \subseteq X(t)$ for all t. This works as follows. Suppose $X(t) = x$ and $E(t) = e$, $e \subseteq x \subseteq y$. Then associate with every point of $x_i \in x$ an exponential lifetime L_i with mean $1/K$ and a uniform random variable U_i on $[0, K]$. Associate with all points $y_i \in y \setminus e$ an exponential waiting time W_i with mean $(\kappa - \lambda)/\lambda$. The first transition after time t takes place at time $t + m$ where $m = \min(L_1, \ldots, L_{n(x)}, W_1, \ldots, W_{n(y)-n(e)})$. If $m = L_i$ then the point x_i is removed (if present) from the E process and x_i is removed from the X process, *but only if* $U_i \leq f(x \setminus \{x_i\})/f(x)$. If $m = W_i$ the point e_i is added to the E process and (if not already present) to the X process.

We now explain how we can apply coupling-from-the-past (CFTP) to obtain a perfect sample from the π_λ, the stationary distribution of $X(t)$. Fix a time $-T < 0$ in the past. Since the E process is time reversible, we can extend it from $E(0) = E_0$ into the past until time $-T$. Now define a process X_{-T} on $[-T, 0]$ starting at $X_{-T}(-T) = E(-T)$. We can arrange it so that X_{-T} is a birth and death process with the same dynamics as X, while $E(t) \subseteq X_{-T}(t) \subseteq D$ for all $t \in [-T, 0]$. Because X_{-T} has the same dynamics as X, it follows that $X_{-T}(0)$ tends in distribution to π_λ as $T \to \infty$.

To check if $X_{-T}(0)$ has reached equilibrium—so to say—we need a 'lower bound process' $L_{-T}(\cdot)$ with $L_{-T}(-T) = E(-T)$ and an 'upper bound process' $U_{-T}(\cdot)$ with $U_{-T}(-T) = D$ such that (cf. Kendall and Møller (1999) formulas (3.5) to (3.7))

1. ('sandwiching') $E(t) \subseteq L_{-T}(t) \subseteq X_{-T}(t) \subseteq U_{-T}(t) \subseteq D$ for all $-T \leq t \leq 0$;

2. ('funneling') $L_{-T}(t) \subseteq L_{-S}(t) \subseteq X_{-S}(t) \subseteq U_{-S}(t) \subseteq U_{-T}(t)$, for all $-S \leq -T \leq t \leq 0$;

3. ('coalescing') if, for some s, $L_{-T}(s) = U_{-T}(s)$ then $L_{-T}(t) = U_{-T}(t)$, for all $t \geq s$.

Moreover, we need that the event $\{L_{-T}(0) = U_{-T}(0)\}$ almost surely occurs for finite T. If the above 3 inclusion properties hold and the latter event

occurs then we can conclude that $L_{-T}(0) = U_{-T}(0)$ is a perfect sample from π_λ. This is roughly the content of Theorem 3.1 of Kendall and Møller (1999). The difference is that they consider a varying 'maximal state' $D(t)$ but a fixed minimal state $E(t) = \emptyset$, whereas we consider a fixed maximal state $D(t) = D$, but a varying minimal state.

In practice, the CFTP algorithm would consist of successively choosing increasing T and running $E(t)$, $L_{-T}(t)$ and $U_{-T}(t)$ for $t \in [-T, 0]$ until $L_{-T}(0) = U_{-T}(0)$.

The algorithm takes a long time to terminate when the difference between D and $E(0)$ is large. This occurs typically if the difference between κ and λ is large. Therefore it is a good idea to choose a decreasing sequence $\kappa = \lambda(0) > \lambda(1) > \cdots > \lambda(n) = \lambda$ and successively sample from $\pi_{\lambda(0)}, \pi_{\lambda(1)}, \ldots, \pi_{\lambda(n)}$. At each step we can use the sample obtained in the previous step as dominating pattern D. In fact, in section 4.3 we showed how to choose the $\lambda(i)$ in such a way that the difference between D and $E(0)$ is never more than a single point. This allowed us to obtain samples from π_λ for *all* λ in a given interval.

4.6.3 the hereditary case

Recall our definition of an hereditary point process at (4.15). Kendall and Møller (1999) give two algorithms to obtain samples from a hereditary point process with distribution π_λ, one of which is based on birth-and-death processes. They construct a b&d process $X(t)$ with π_λ as its stationary distribution and a dominating process $D(t)$ which is in equilibrium and time reversible. In fact, $D(t)$ is at all times t a spatial Poisson process of intensity $K\lambda$. Kendall and Møller give a coupling of $D(t)$ and $X(t)$ such that $X(t) \subseteq D(t)$ for all t. They then demonstrate a coupling-from-the-past algorithm with $D(t)$ ($t \in [-T, 0]$) as a (variable) maximal state and $E(t) = \emptyset$ as a trivial minimal state. The sample from π_λ which the algorithm produces lies between $D(0)$ and $E = \emptyset$. When λ is large the difference between $D(0)$ and E will typically be large. If this is the case, then the algorithm will take very long to terminate because it has many different configurations to choose from.

We give a generalization of the work of Kendall and Møller which allows us to take as a fixed minimal state a sample $E \sim \pi_\kappa$ for an arbitrary $\kappa < \lambda$. The maximal state $D(t)$ will be a birth-and-death process which has the superposition of E and a Poisson process of intensity $K(\lambda - \kappa)$ as its equilibrium distribution. Noting that the empty set is trivially a sample from π_0 we see that our generalization reduces to the Kendall and Møller algorithm when we take $\kappa = 0$.

The use of the generalization is that it will allow us to choose an increasing

sequence $0 = \lambda(0) < \lambda(1) < \cdots < \lambda(n) = \lambda$ and successively sample from $\pi_{\lambda(0)}, \pi_{\lambda(1)}, \ldots, \pi_{\lambda(n)}$. At each step we can use the sample obtained in the previous step as minimal pattern E to which we add a Poisson point process of intensity $K(\lambda(i+1) - \lambda(i))$ to obtain the maximal pattern $D(0)$. If we choose the steps from $\lambda(i)$ to $\lambda(i+1)$ small enough, then the difference between $D(0)$ and E will always be small too.

The next lemma shows how we can obtain a sample from π_λ by adding points to a sample from π_κ for an arbitrary $\kappa < \lambda$.

Lemma 4.7. *Suppose condition (4.15) holds. Let $0 \le \kappa < \lambda$. Then for all $x \in \Omega$*

$$\sum_{y \subseteq x} \frac{f(x)}{f(y)} f_\kappa(y) p_{\lambda-\kappa}(x \setminus y) = f_\lambda(x). \tag{4.19}$$

Proof This lemma basically follows from the superposition property of Poisson processes. By condition (4.15) we have that $f(x)/f(y)$ is bounded. Now it follows from (4.14) that

$$f_\kappa(y) = f(y) e^{(1-\kappa)|S|} \kappa^{n(y)}$$

and

$$f(x) = f_\lambda(x) e^{(\lambda-1)|S|} \lambda^{-n(x)}.$$

Hence

$$\begin{aligned}
&\sum_{y \subseteq x} \frac{f(x)}{f(y)} p_{\lambda-\kappa}(x \setminus y) f_\kappa(y) \\
&= \sum_{y \subseteq x} \frac{f_\lambda(x)}{f(y)} e^{(\lambda-1)|S|} \lambda^{-n(x)} e^{-(\lambda-\kappa)|S|} (\lambda - \kappa)^{n(x \setminus y)} f(y) e^{(1-\kappa)|S|} \kappa^{n(y)} \\
&= f_\lambda(x) \lambda^{-n(x)} \sum_{y \subseteq x} (\lambda - \kappa)^{n(x)-n(y)} \kappa^{n(y)} \\
&= f_\lambda(x).
\end{aligned}$$

□

Suppose we have a sample $E = y = \{y_1, \ldots, y_n\}$ from π_κ for some $\kappa < \lambda$. Now consider a spatial birth-and-death process $X(t)$, $t \ge 0$ on the space $\{y \cup x : x \in \Omega\}$ with birth rate $b(y \cup x, \xi) = (\lambda - \kappa) f(y \cup x \cup \{\xi\}) / f(y \cup x)$ and death rate $d(y \cup x, \xi) = 1$. One easily checks that

$$(\lambda - \kappa) \frac{f(y \cup x \cup \{\xi\})}{f(y \cup x)} \frac{f(y \cup x)}{f(y)} p_{\lambda-\kappa}(x) = 1 \frac{f(y \cup x \cup \{\xi\})}{f(y)} p_{\lambda-\kappa}(x \cup \{\xi\}).$$

So $X(t)$ satisfies detailed balance with respect to

$$\frac{f(y \cup x)}{f(y)} p_{\lambda-\kappa}(x),$$

which is therefore the density of the stationary distribution of $X(t)$, conditionally on $E = y$. Hence, by lemma 4.7 the unconditional stationary distribution of $X(t)$ is π_λ.

Consider another b&d process, $D(t)$, $t \geq 0$, on $\{y \cup x : x \in \Omega\}$ with birthrate $b(y \cup x, \xi) = K(\lambda - \kappa)$ and death rate $d(y \cup x, \xi) = 1$. Let D_0 be the union of y and be a sample from $\mathrm{Po}_{K(\lambda-\kappa)}$. We have for all $x \in \Omega$

$$K(\lambda - \kappa) p_{K(\lambda-\kappa)}(x) = 1 p_{K(\lambda-\kappa)}(x \cup \{\xi\}).$$

Hence if we set $D(0) = D_0$ then $D(t)$ is time reversible and in equilibrium.

$X(t)$ and $D(t)$ have the same death rates, while by (4.15) the birth rate of $D(t)$ is greater than the birth rate of $X(t)$. Setting $X(0) = D(0)$ it is not difficult to couple $D(t)$ and $X(t)$ such that $X(t) \subseteq D(t)$ for all t. This could be done much like we indicated in the previous subsection for the anti-hereditary case. In fact, a detailed description of the coupling is given in Kendall and Møller (1999) for the particular—but in no sense special—case $\kappa = 0$. They also explain how to apply CFTP to obtain a perfect sample from the stationary distribution of $X(t)$.