CHAPTER 1

PERFECT SAMPLING FOR POINT PROCESS CLUSTER MODELLING

Ian W. McKeague and Marc Loizeaux

1.1 Introduction

When disease incidence locations are observed in a region, there is often interest in studying whether there is clustering about landmarks representing possible centralized sources of the disease. In this article we study a Bayesian approach to the detection and estimation of such landmarks. Spatial point processes are used to specify both the observation process and the prior distribution of the landmarks. We develop a perfect sampling algorithm for the posterior distribution of landmarks under various conditions on the prior and likelihood.

Bayesian cluster models of the type we consider were introduced by Baddeley and van Lieshout (1993), primarily for applications in computer vision. The dissertation of van Lieshout (1995) focused on the special case of the Neyman–Scott cluster process, in which the observations arise from a superposition of inhomogeneous Poisson processes associated with each landmark (Neyman and Scott, 1972), and she applied it to the well known redwood seedling data used by Strauss (1975). Hurn (1998) applied the Baddeley–van Lieshout approach to the study of changes in the size and shape of living cells. Lawson and Clarke (1999) survey the statistical literature on disease clustering models.

Markov chain Monte Carlo (MCMC) techniques are indispensable for the application of point process models in statistics, see, e.g., the survey of Møller (1999). The typical MCMC sampler obtains draws that are at best only approximately from the target distribution, and are often plagued by convergence problems, even when a long "burn-in" period is used. Moreover, if independent draws are required, then every draw must be produced by a separate chain. However, using an algorithm developed by Kendall and Møller (2000), it is possible to sample perfectly from the posterior distribution in the Bayesian cluster model. Perfect samplers originate in the seminal work of Propp and Wilson (1996), whose coupling from the past (CFTP) algorithm delivers an *exact* draw from the target distribution. The most important practical advantage of perfect samplers over traditional MCMC schemes is that the need to assess convergence of the sampler is eliminated.

There are many examples of perfect samplers in the literature. Mira et al. (1999) apply perfect simulation to slice samplers for bounded target distributions; Casella et al. (1999) create perfect slice samplers for mixtures of exponential distributions and mixtures of normal distributions. For an example of perfect sampling of a conditioned Boolean model, see Cai and Kendall (1999). Häggström et al. (1997) obtain perfect samples from the area–interaction point process (attractive case) using auxiliary variables. Another version of perfect sampling called read-once CFTP, which runs the Markov chain forward in time and never restarts it at previous past times (thus avoiding the need to store random numbers for reuse), is given by Wilson (1999). For recent applications of perfect simulation in statistics, see Green and Murdoch (1999) and Møller and Nicholls (1999).

The Kendall–Møller version of perfect simulation was developed for spatial point processes that are locally stable. We show that the posterior distribution in the Bayesian cluster model is locally stable on its support, provided the prior is locally stable and the likelihood satisfies some mild conditions. In particular, this shows that the posterior density is *proper* (has unit total mass). The Kendall–Møller algorithm is computationally *feasible*, however, only when the locally stable point process is either attractive (favoring clustered patterns) or repulsive (discouraging clustered patterns), or a product of attractive and repulsive components. We examine the feasibility of the sampler in two special cases: the Neyman–Scott process and the pure silhouette model (in which only the support of the observations is determined by the landmarks).

Our approach is applied to data on leukemia counts in an eight county area of upstate New York during the years 1978–82. There is an extensive literature on the analysis of these data, the main goal being the detection of disease clusters; see, e.g., Ghosh et al. (1999), Ahrens et al. (1999), and Denison and Holmes (2001). The study area includes 11 inactive hazardous waste sites, and it is natural in any study to attempt to determine if any of these sites can be seen as a contributor to the incidence of leukemia in the area. We find evidence for an elevated leukemia incidence in the neighborhood of one of the sites.

The paper is organized as follows. Section 1.2 gives background material and describes the general model. In this section we also state the main result of the paper showing that the posterior is locally stable on its support. Section 1.4 examines several examples of the basic model, and addresses the question of whether perfect sampling is feasible in each case. Section 1.5 contains the disease clustering application, and a further application to the classic redwood seedlings data is given in Section 1.6.

1.2 Bayesian cluster model

1.2.1 Preliminaries

The point process notation used throughout the paper is standard, cf. Häggström et al. (1999), Baddeley and Møller (1989), or Møller (1999, 2001). For in-depth accounts of the theory of point processes, see Daley and Vere-Jones (1988) and van Lieshout (2000).

Let W be a compact subset of the plane representing the study region, and let $|\cdot|$ denote Lebesgue measure on W. A realization of a point process in W is a finite set of points $\mathbf{x} = \{x_1, x_2, \ldots, x_{n(\mathbf{x})}\} \subset W$, where $n(\mathbf{x})$ is the number of points in $\mathbf{x}$. If $n(\mathbf{x}) = 0$, write $\mathbf{x} = \emptyset$ for the empty configuration. Let Ω denote the exponential space (see Carter and Prenter, 1972) of all such finite point configurations in W, and furnish it with the σ-field $\mathcal{F}$ generated by sets of the form $\{\mathbf{x} : n(\mathbf{x} \cap B) = k\}$, where $B \in \mathcal{B}$, the Borel σ-field on W, and $k = 0, 1, 2, \ldots$.

Let $(W, \mathcal{B}, \lambda)$ be a measure space with $\lambda(W) < \infty$. The Poisson process Z on W with intensity measure λ is an Ω-valued random variable such that the number of points $n(Z)$ follows the Poisson distribution with mean $\lambda(W)$, and conditional on $n(Z) = m$, the m points are independent and identically distributed with distribution $\lambda(\cdot)/\lambda(W)$. If λ is Lebesgue measure $|\cdot|$, then we say Z is the unit rate Poisson process on W. All point processes considered here will have distributions that are absolutely continuous with respect to the distribution π of the unit rate Poisson process, and thus specified by a density with respect to π.

Two important point processes are the Strauss (1975) process and the area-interaction process of Baddeley and van Lieshout (1995). The Strauss process has unnormalized density $f(\mathbf{x}) = \beta^{n(\mathbf{x})}\gamma^{t(\mathbf{x})}$, where $\beta > 0$, $0 < \gamma \leq 1$ and $t(\mathbf{x})$ is the number of unordered pairs of points in $\mathbf{x}$ which are within a specified distance r of each other. The Strauss process only models pairwise interaction which is repulsive, i.e., the points in the process tend to separate from one another.

For the area-interaction process, the unnormalized density is given by $f(\mathbf{x}) = \beta^{n(\mathbf{x})}\gamma^{-\nu(U(\mathbf{x}))}$, where $\beta > 0$, $\gamma > 0$, ν is a totally finite, regular Borel measure on W, and $U(\mathbf{x})$ is the intersection of W with the union of balls of fixed radius r centered at points in $\mathbf{x}$. We can also write $U(\mathbf{x}) = (\mathbf{x} \oplus G) \cap W$, where $\oplus$ denotes Minkowski addition, and the grain G is the ball of radius r centered at the origin (more generally, G can be any compact subset of $\mathbb{R}^2$). Area-interaction processes can be used to model both clustered (attractive) and ordered (repulsive) patterns, and allow interactions of infinite order. The parameter γ controls the type of interaction; $\gamma > 1$ produces clustered patterns, $0 < \gamma < 1$ produces ordered patterns.

An unnormalized density f (or corresponding process X) is said to be locally stable if there is a constant $K > 0$ such that

$$f(\mathbf{x} \cup \{\xi\}) \leq K f(\mathbf{x})$$

for all $\mathbf{x} \in \Omega$, $\xi \in W \backslash \mathbf{x}$. As noted by Kendall and Møller (2000), local stability is equivalent to an upper bound on the Papangelou conditional intensity

$$q(\mathbf{x}, \xi) \equiv \frac{f(\mathbf{x} \cup \{\xi\})}{f(\mathbf{x})}, \ \mathbf{x} \in \Omega, \ \xi \in W \backslash \mathbf{x},$$

(with $c/0 = 0$ for $c \geq 0$) and the hereditary property, which states that if $\mathbf{x} \subset \mathbf{x}'$ then $f(\mathbf{x}') > 0$ implies $f(\mathbf{x}) > 0$. In addition, local stability implies $f(\emptyset) > 0$, since repeated use of the condition gives $f(\mathbf{x}) \leq K^{n(\mathbf{x})} f(\emptyset)$. Most point processes that have been suggested for modeling spatial point patterns are locally stable, including the Strauss and the area-interaction processes.

1.2.2 Model specification

The observed point configuration which arises from the landmarks $\mathbf{x}$ will be denoted $\mathbf{y} = \{y_1, y_2, \ldots, y_{n(\mathbf{y})}\} \subset W$, and assumed to be non-empty. The prior and observation models are specified by point processes on W. The prior distribution of landmarks corresponds to a point process X having density $p_X(\mathbf{x})$ with respect to π.

The landmarks produce daughters $\mathbf{y}$, a certain proportion of which will be required to fall in a silhouette region $S(\mathbf{x}) \subset W$ given by the union of discs

$$S(\mathbf{x}) = \cup_{\xi \in \mathbf{x}} D(\xi, r_{\mathrm{sil}}) = (\mathbf{x} \oplus G_{\mathrm{sil}}) \cap W,$$

where $D(\xi, r_{\mathrm{sil}}) = \{x \in W : ||x - \xi|| \leq r_{\mathrm{sil}}\}$, $||\cdot||$ is Euclidean distance, and G_{sil} is the disk of radius r_{sil} centered at the origin. The random closed set $S(X)$ is known as a Boolean model, see, e.g., Stoyan et al. (1995, Ch. 3). With appropriate measurability conditions, our approach allows more general types of silhouettes $S(\mathbf{x})$, but we have restricted attention to Boolean silhouettes for simplicity. We have used different notation for $S(\mathbf{x})$ and $U(\mathbf{x})$, even though they are silhouettes of the same type, to avoid confusion later on.

The observation process is defined in terms of an unnormalized density $f(\cdot|\mathbf{x})$ conditioned on the event $\{\mathbf{y}\colon s(\mathbf{y}|\mathbf{x}) \geq c\}$, where $0 \leq c \leq 1$ and $s(\mathbf{y}|\mathbf{x}) = n(\mathbf{y} \cap S(\mathbf{x}))/n(\mathbf{y})$ is the proportion of points in $\mathbf{y}$ within the silhouette region (with $0/0 = 0$). Thus, for a given set of landmarks $\mathbf{x}$, the density of the observed point process Y with respect to π is

$$p_{Y|X=\mathbf{x}}(\mathbf{y}) = \alpha_Y(\mathbf{x}) f(\mathbf{y}|\mathbf{x}) 1\{s(\mathbf{y}|\mathbf{x}) \geq c\},$$

where

$$\alpha_Y(\mathbf{x}) = \left(\int_{\{\mathbf{v}: s(\mathbf{v}|\mathbf{x}) \geq c\}} f(\mathbf{v}|\mathbf{x})\, \pi(d\mathbf{v}) \right)^{-1}$$

is the normalizing constant. We assume that $f(\mathbf{y}|\mathbf{x})$ is jointly measurable in $\mathbf{x}$ and $\mathbf{y}$.

The silhouette portion of the observation process could be absorbed into $f(\mathbf{y}|\mathbf{x})$, as in Baddeley and van Lieshout's (1993) formulation, but we find it useful to isolate $1\{s(\mathbf{y}|\mathbf{x}) \geq c\}$ in this fashion in order to provide conditions under which perfect sampling is feasible. Our formulation of the model also provides a flexible class of Boolean models which are of independent interest. When $f(\mathbf{y}|\mathbf{x}) = f(\mathbf{y})$ and $c = 1$ we say the likelihood follows a *pure silhouette model*; this is similar to Baddeley and van Lieshout's (1993, p. 233) blur-free model.

From Bayes formula, the posterior density of X with respect to π is

$$p_{X|Y=\mathbf{y}}(\mathbf{x}) \propto f_{X|Y=\mathbf{y}}(\mathbf{x}) 1\{s(\mathbf{y}|\mathbf{x}) \geq c\}, \tag{1.1}$$

where the unnormalized density

$$f_{X|Y=\mathbf{y}}(\mathbf{x}) = \alpha_Y(\mathbf{x}) f(\mathbf{y}|\mathbf{x}) p_X(\mathbf{x}) \tag{1.2}$$

is called the unrestricted posterior.

Theorem 1.2.1 which follows provides sufficient conditions for the unrestricted posterior to be locally stable. We note in passing that the posterior itself is not hereditary (thus not locally stable) if $c > 0$, because $S(\emptyset) = \emptyset$ and $\mathbf{y} \neq \emptyset$. We assume local stability of the prior $p_X(\cdot)$ and of the likelihood $f(\mathbf{y}|\cdot)$ (for each fixed $\mathbf{y}$). In addition, we assume that the *family* of functions $\{f(\mathbf{y}|\cdot) : \ \mathbf{y} \in \Omega\}$ satisfies the following *local growth condition*: there exists a constant $L > 0$ such that

$$f(\mathbf{y}|\mathbf{x} \cup \{\xi\}) \geq L f(\mathbf{y}|\mathbf{x}) \tag{1.3}$$

for all $\mathbf{x} \in \Omega$, $\xi \in W \backslash \mathbf{x}$. We call this a 'local growth' condition because it implies that the Papangelou conditional intensity for $f(\mathbf{y}|\cdot)$ is bounded away from zero. As it turns out, $L = 1$ in the examples to follow.

Prior to stating our main theorem, for easy reference we state the following lemma giving the density of an inhomogeneous Poisson process with respect to the unit rate Poisson process. The proof can be found, e.g., in Daley and Vere-Jones (1988, p. 499).

Lemma 1.2.1 *Let π be the distribution of the unit rate Poisson process on the compact set $W \subset \mathbb{R}^2$, and let π^β be the distribution of the inhomogeneous Poisson process on W with intensity function $\beta(t)$. Then $\pi^\beta \ll \pi$ and*

$$\frac{d\pi^\beta}{d\pi}(\mathbf{x}) = \exp\left(\int_W (1 - \beta(t))\, dt \right) \prod_{j=1}^{n(\mathbf{x})} \beta(x_j).$$

Theorem 1.2.1 *Suppose $p_X(\cdot)$ and $f(\mathbf{y}|\cdot)$ (for each $\mathbf{y}$) are locally stable, and $\{f(\mathbf{y}|\cdot)\colon \mathbf{y} \in \Omega\}$ satisfies the local growth condition (1.3). Then the unrestricted posterior (1.2) is locally stable.*

Proof: We first show that density (1.1) is measurable. Since $p_X(\cdot)$ and $f(\mathbf{y}|\cdot)$ are measurable, we need only establish the measurability of $\alpha_Y(\cdot)$ and $\{\mathbf{x}\colon s(\mathbf{y}|\mathbf{x}) \geq c\}$.

First consider $\{\mathbf{x}\colon s(\mathbf{y}|\mathbf{x}) \geq c\}$ in the case $c = 1$. With $\mathbf{y} = \{y_1, y_2, \ldots, y_m\}$,

$$\begin{aligned}\{\mathbf{x}\colon s(\mathbf{y}|\mathbf{x}) \geq 1\} = \{\mathbf{x}\colon \mathbf{y} \subset S(\mathbf{x})\} &= \bigcap_{j=1}^{m} \{\mathbf{x}\colon n(\mathbf{x} \cap D(y_j, r)) \geq 1\} \\ &= \bigcap_{j=1}^{m} \bigcup_{k=1}^{\infty} \{\mathbf{x}\colon n(\mathbf{x} \cap D(y_j, r)) = k\},\end{aligned}$$

which belongs to $\mathcal{F}$. The general case reduces to the special case of $c = 1$ because $\{\mathbf{x}\colon s(\mathbf{y}|\mathbf{x}) \geq c\}$ is the finite union of sets of the form $\{\mathbf{x}\colon \mathbf{z} \subset S(\mathbf{x})\}$ with $\mathbf{z} \subset \mathbf{y}$ and $n(\mathbf{z}) \geq cn(\mathbf{y})$.

To show that $\alpha_Y(\cdot)$ is measurable, the π-λ theorem allows us to reduce to the case $f(\mathbf{v}|\mathbf{x}) = 1_C(\mathbf{v})1_D(\mathbf{x})$, where C and D are generating sets in $\mathcal{F}$. Thus we only need show the measurability of

$$\mathbf{x} \longmapsto \pi\{\mathbf{v}\colon n(\mathbf{v} \cap B) = k \text{ and } n(\mathbf{v} \cap S(\mathbf{x})) \geq cn(\mathbf{v})\},$$

where $B \in \mathcal{B}$ and $k \geq 0$. Since π is Poisson, the above function only depends on $\mathbf{x}$ through a continuous function of the areas of the disjoint regions $B^* \cap S(\mathbf{x})^*$, where $A^* = A$ or A^c. These areas are measurable functions of $\mathbf{x}$ by Lemma 2.1 of Baddeley and van Lieshout (1995).

To establish local stability of the unrestricted posterior it is enough to show that $\alpha_Y(\cdot)$ is locally stable; $p_X(\cdot)$ and $f(\mathbf{y}|\cdot)$ were assumed to be locally stable, and finite products of locally stable densities are locally stable. Given $\xi \in W \backslash \mathbf{x}$,

$$\begin{aligned}\alpha_Y(\mathbf{x} \cup \{\xi\})^{-1} &= \int_{\{\mathbf{v}: s(\mathbf{v}|\mathbf{x} \cup \{\xi\}) \geq c\}} f(\mathbf{v}|\mathbf{x} \cup \{\xi\})\, \pi(d\mathbf{v}) \\ &\geq L \int_{\{\mathbf{v}: s(\mathbf{v}|\mathbf{x} \cup \{\xi\}) \geq c\}} f(\mathbf{v}|\mathbf{x})\, \pi(d\mathbf{v}) \\ &\geq L \int_{\{\mathbf{v}: s(\mathbf{v}|\mathbf{x}) \geq c\}} f(\mathbf{v}|\mathbf{x})\, \pi(d\mathbf{v}) \\ &= L\alpha_Y(\mathbf{x})^{-1},\end{aligned}$$

where we have used the local growth condition (1.3) and the fact that $s(\mathbf{v}|\mathbf{x}) \leq s(\mathbf{v}|\mathbf{x} \cup \{\xi\})$. Thus $\alpha_Y(\cdot)$ is locally stable. □

It is convenient to introduce some notation for the bounds implied by the local stability and local growth conditions. Let K_{prior} be an upper bound on the Papangelou conditional intensity for $p_X(\mathbf{x})$, and let $K_{\text{lik}}(\mathbf{y})$ be an upper bound on the Papangelou conditional intensity for $f(\mathbf{y}|\cdot)$. Under the assumptions given in Theorem 1.2.1, since finite products of locally stable densities are locally stable, we find that

$$K_{\text{post}} = \frac{1}{L} K_{\text{lik}}(\mathbf{y}) K_{\text{prior}} \tag{1.4}$$

is an upper bound on the Papangelou conditional intensity for the unrestricted posterior density (1.2).

1.3 Sampling from the posterior

Clearly, once we develop a sampler for the unrestricted posterior (1.2), simple rejection sampling can be used to obtain samples from the posterior (1.1) itself. This is an efficient approach provided draws from the unrestricted posterior are not rejected too frequently.

As previously stated, MCMC methods are needed to sample from (1.2), and from point processes in general. The Hastings–Metropolis–Green sampler of Geyer and Møller (1994) is applicable and easy to implement. Under the conditions of Theorem 1.2.1, the unrestricted posterior is locally stable, so the Geyer–Møller algorithm is geometrically ergodic, see Geyer (1999). As explained in the Introduction, however, it would be preferable to utilize Kendall and Møller's (2000) algorithm if possible. This algorithm is not feasible in every situation, but when feasible has the advantage of producing perfect samples. Kendall (1998) introduced the algorithm initially for area-interaction processes. Kendall and Thönnes (1999) give a review of the method from the perspective of image processing. An alternative perfect sampler for area-interaction processes has been suggested by Häggström et al. (1999).

The Kendall–Møller algorithm uses the method of dominated coupling-from-the-past (CFTP) to produce perfect samples from a locally stable point process. The sampler is driven by a spatial birth-and-death process $\tilde{X}(t)$ having equilibrium distribution which agrees with the distribution of the target point process. Dominated CFTP is sometimes referred to as "horizontal CFTP", as opposed to the "vertical CFTP" originally introduced by Propp and Wilson (1996).

Vertical CFTP requires minimal and maximal elements with respect to some partial ordering of the state space; it is not applicable to point processes because a maximal element does not exist in Ω under set inclusion. Horizontal CFTP, however, only requires a minimal element $\emptyset$ which functions as an ergodic atom for a dominating process $D(t)$. Once the dominating process $D(t)$ hits the minimal element, subsequent evolution of $\tilde{X}(t)$ does not depend on the initial state, so $\tilde{X}(0)$ represents the endpoint of a path from the infinite past (coalescence), and thus a perfect sample from the target distribution. A further aspect of horizontal CFTP is the use of lower and upper processes that bound $\tilde{X}(t)$ more efficiently than $\emptyset$ and $D(t)$, and thus provide coalescence more rapidly. These various processes are defined below.

The spatial birth-and-death process $\tilde{X} = \{\tilde{X}(t) : t \in \mathbb{R}\}$ has birth rate $b(\mathbf{x}, \xi) = q(\mathbf{x}, \xi)$ given by the Papangelou conditional intensity corresponding to $f_{X|Y=\mathbf{y}}(\mathbf{x})$, and death rate $d(\mathbf{x}, \xi) = 1$. It is clear that the target density $f_{X|Y=\mathbf{y}}(\mathbf{x})$ and $\tilde{X}$ are in detailed balance, i.e.

$$f_{X|Y=\mathbf{y}}(\mathbf{x})b(\mathbf{x}, \xi) = f_{X|Y=\mathbf{y}}(\mathbf{x} \cup \xi)d(\mathbf{x}, \xi) > 0,$$

so $\tilde{X}(t)$ converges in distribution to X as $t \to \infty$, where X has density (1.2), see Kendall and Møller (2000). The dominating process $D(t)$ has birth rate K_{post} and death rate 1, so its equilibrium distribution is homogeneous Poisson with intensity K_{post}. Suppose that $\tilde{X}(t^-) = \mathbf{x} \subset \mathbf{y} = D(t^-)$ are the states of $\tilde{X}$ and D just before time t. If at time t a point ξ is added to D, then this same point is added to $\tilde{X}$ with probability $q(\mathbf{x}, \xi)/K_{\mathrm{post}}$, that is, if $M(t) \leq q(\mathbf{x}, \xi)/K_{\mathrm{post}}$ where $M(t)$ is an iid uniform(0,1) mark process. If at time t a point is deleted from D, then this same point is deleted from $\tilde{X}$.

For every $n > 0$, the upper process U_{-n} is initialized at time $-n$ with $\hat{\mathbf{x}} = D(-n)$. The initial state for L_{-n} is the empty configuration. Using the mark $M(t)$, a new point ξ is added to U_{-n} at time $-t$, $t < n$, with probability

$$\max\left\{\frac{q(\mathbf{x}, \xi)}{K_{post}} : L_{-n}^{-t^-}(\emptyset) \subseteq \mathbf{x} \subseteq U_{-n}^{-t^-}(\hat{\mathbf{x}})\right\},$$

and the same point ξ is added to L_{-n} with probability

$$\min\left\{\frac{q(\mathbf{x}, \xi)}{K_{post}} : L_{-n}^{-t^-}(\emptyset) \subseteq \mathbf{x} \subseteq U_{-n}^{-t^-}(\hat{\mathbf{x}})\right\},$$

where $L_{-n}^{-t^-}(\emptyset)$ and $U_{-n}^{-t^-}(\hat{\mathbf{x}})$ are the states of $L_{-n}(\emptyset)$ and $U_{-n}(\hat{\mathbf{x}})$ just before time $-t$.

The construction described then gives

$$L_{-n}^{-t}(\emptyset) \subseteq \tilde{X}(-t) \subseteq U_{-n}^{-t}(\hat{\mathbf{x}}) \subseteq D(-t),$$

and the funnelling property

$$L_{-n}^{-t} \leq L_{-m}^{-t} \leq U_{-m}^{-t} \leq U_{-n}^{-t}, \; m \geq n \geq t.$$

Since X is locally stable by Theorem 1.2.1, the empty configuration is an ergodic atom for $D(t)$. Thus it is simply a matter of finding n large enough so that $L^0_{-n}(\emptyset) = U^0_{-n}(\hat{\mathbf{x}})$, in which case $L^0_{-n}(\emptyset)$ is a perfect draw from the posterior distribution.

Note the inherent difficulty in calculating the probability of adding a point to U at time $-t$ (and similarly to L): in the general case it is necessary to evaluate $q(\mathbf{x},\xi)/K_{post}$ for every $\mathbf{x}$ such that $L_{-n}^{-t^-}(\emptyset) \subseteq \mathbf{x} \subseteq U_{-n}^{-t^-}(\hat{\mathbf{x}})$. These probabilities are easy to calculate, however, if the locally stable point process X is either attractive:

$$q(\mathbf{x},\xi) \leq q(\mathbf{x}',\xi) \text{ whenever } \xi \notin \mathbf{x}' \text{ and } \mathbf{x} \subset \mathbf{x}',$$

or repulsive:

$$q(\mathbf{x},\xi) \geq q(\mathbf{x}',\xi) \text{ whenever } \xi \notin \mathbf{x}' \text{ and } \mathbf{x} \subset \mathbf{x}'.$$

1.4 Specialized examples

In this Section we examine various choices of the prior and likelihood which satisfy Theorem 1.2.1. In addition, we describe conditions under which the posterior is either attractive or repulsive, so that perfect simulation is feasible. A small dataset is used to illustrate the viability of perfect simulation in each case. In Section 1.4.1 the Neyman–Scott model is described in detail (see also Chapter **??**). Section 1.4.2 provides several cases of the pure silhouette model.

1.4.1 Neyman–Scott model

In this section we consider the Neyman–Scott model in which the observation process Y is the superposition of $n(\mathbf{x})$ independent inhomogeneous Poisson processes Z_{x_i} and a background Poisson noise process of intensity $\epsilon > 0$. The intensity $h(\cdot|x_i)$ of Z_{x_i} is specified parametrically, and the prior $p_X(\mathbf{x})$ is assumed to be locally stable.

This fits into our general framework by taking $c = 0$ and $f(\mathbf{y}|\mathbf{x}) = \prod_{j=1}^{n(\mathbf{y})} \mu(y_j|\mathbf{x})$, where

$$\mu(t|\mathbf{x}) = \epsilon + \sum_{i=1}^{n(\mathbf{x})} h(t|x_i)$$

is the conditional intensity at t of Y given $\mathbf{x}$. We now check the relevant conditions of Theorem 1.2.1, assuming that $h(t|\cdot)$ is bounded for each $t \in W$.

To show that $f(\mathbf{y}|\cdot)$ is locally stable, note that for $\xi \in W \backslash \mathbf{x}$

$$\begin{aligned} f(\mathbf{y}|\mathbf{x} \cup \{\xi\}) &= \prod_{j=1}^{n(\mathbf{y})} \left(\epsilon + \sum_{i=1}^{n(\mathbf{x})} h(y_j|x_i) + h(y_j|\xi) \right) \\ &\leq \prod_{j=1}^{n(\mathbf{y})} \left(\epsilon \left(1 + \frac{h(y_j|\xi)}{\epsilon} \right) + \sum_{i=1}^{n(\mathbf{x})} h(y_j|x_i) \right) \\ &\leq K_{\text{lik}}(\mathbf{y}) f(\mathbf{y}|\mathbf{x}), \end{aligned}$$

where

$$K_{\text{lik}}(\mathbf{y}) = \sup_{\xi \in W} \prod_{j=1}^{n(\mathbf{y})} \left(1 + \frac{h(y_j|\xi)}{\epsilon}\right) < \infty.$$

To check the local growth condition, note that for $\xi \in W \backslash \mathbf{x}$

$$f(\mathbf{y}|\mathbf{x} \cup \{\xi\}) = \prod_{j=1}^{n(\mathbf{y})} (\mu(y_j|\mathbf{x}) + h(y_j|\xi)) \geq f(\mathbf{y}|\mathbf{x}),$$

uniformly in $\mathbf{y}$, and we can use $L = 1$. Thus the conditions of Theorem 1.2.1 are satisfied, so the posterior is locally stable.

As previously noted, the Kendall–Møller algorithm is feasible if the posterior is either attractive or repulsive, or factors into attractive/repulsive components. The Papangelou conditional intensity corresponding to $f(\mathbf{y}|\cdot)$ is

$$\frac{f(\mathbf{y}|\mathbf{x} \cup \{\xi\})}{f(\mathbf{y}|\mathbf{x})} = \prod_{j=1}^{n(\mathbf{y})} \left(1 + \frac{h(y_j|\xi)}{\epsilon + \sum_{i=1}^{n(\mathbf{x})} h(y_j|x_i)}\right),$$

for $\xi \in W \backslash \mathbf{x}$, which is clearly decreasing in $\mathbf{x}$, thus repulsive.

By Lemma 1.2.1 (with intensity $\beta(t) = \mu(t|\mathbf{x})$), we have

$$\alpha_Y(\mathbf{x}) = \exp\left\{\int_W (1 - \mu(t|\mathbf{x}))\, dt\right\}.$$

Therefore, the Papangelou conditional intensity corresponding to $\alpha_Y(\cdot)$ is

$$\frac{\alpha_Y(\mathbf{x} \cup \{\xi\})}{\alpha_Y(\mathbf{x})} = \exp\left\{-\int_W h(t|\xi)\, dt\right\},$$

for $\xi \in W \backslash \mathbf{x}$, which does not depend on $\mathbf{x}$. We conclude that the posterior is repulsive whenever the prior is repulsive, and sampling from the posterior is possible if the prior is either repulsive or attractive.

Figure 1.1 shows an artificial dataset in the unit square consisting of two clusters of three observations each and one isolated observation. Figure 1.2 gives an illustration of perfect sampling for the Neyman–Scott model based on these data. We used a Strauss prior that produces empty configurations of landmarks approximately 50% of the time.

For the likelihood we used the Thomas intensity model (radially symmetric Gaussian)

$$h(t|x) = \frac{\kappa}{2\pi\sigma^2} e^{-||t-x||^2/2\sigma^2}, \tag{1.5}$$

where $\kappa, \sigma > 0$. Note that the contribution of a particular landmark x to the intensity rate at x is exactly $\kappa^* = \kappa/(2\pi\sigma^2)$, the maximum value of $h(t|x)$. We will refer to κ^* as the (Thomas) landmark weight, and find it convenient to express κ^* as a multiple of the background intensity rate

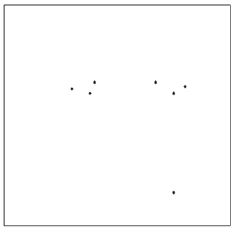

Figure 1.1 *Artificial dataset used to illustrate the examples in Section 1.4.*

ϵ. Large values of κ^* increase the influence of the landmarks, while small values increase the influence of the background noise.

The Kendall–Møller algorithm was used to sample from the posterior; $K_{\text{prior}} = \beta_X$. The algorithm provided 10,000 samples on a Silicon Graphics workstation in less than 1 minute for the prior and about 3 minutes for the largest value of κ^*.

1.4.2 Pure silhouette models

In this section we look at some special cases of the pure silhouette model: $f(\mathbf{y}|\mathbf{x}) = f(\mathbf{y})$ and $c = 1$. In this case, $\{s(\mathbf{y}|\mathbf{x}) \geq c\} = \{\mathbf{y} \subset S(\mathbf{x})\}$, and the landmarks are now unidentifiable in a frequentist sense because the likelihood only depends on the landmarks through their silhouette: for $\xi \in W \backslash \mathbf{x}$ such that $S(\{\xi\}) \subset S(\mathbf{x})$ (that is, $S(\mathbf{x} \cup \{\xi\}) = S(\mathbf{x})$), we have $p_{Y|X=\mathbf{x}}(\cdot) = p_{Y|X=\mathbf{x}\cup\{\xi\}}(\cdot)$.

The conditions of Theorem 1.2.1 are satisfied provided the prior is locally stable, and then $K_{\text{post}} = K_{\text{prior}}$. The posterior now only depends on the data through the restriction $\mathbf{y} \subset S(\mathbf{x})$ because $f(\mathbf{y})$ is absorbed into the normalizing constant:

$$p_{X|Y=\mathbf{y}}(\mathbf{x}) \propto \alpha_Y(\mathbf{x}) p_X(\mathbf{x}) 1\{\mathbf{y} \subset S(\mathbf{x})\}.$$

Thus it is possible to generate draws from the unrestricted posterior before seeing the data. The computationally expensive MCMC simulations can be carried out ahead of time; once the data becomes available, we just select those draws with silhouettes covering the data.

Suppose now that the observation process is inhomogeneous Poisson with intensity $\beta_Y(t)$ on the silhouette, so $f(\mathbf{y}) = \prod_{j=1}^{n(\mathbf{y})} \beta_Y(y_j)$. If the prior is homogeneous Poisson then the unrestricted posterior turns out to be an area-interaction process, as stated in the following theorem.

Theorem 1.4.1 *In the pure silhouette model, if the prior is homogeneous Poisson with parameter β_X, and the observation process is Poisson with intensity $\beta_Y(\cdot)$, then the unrestricted posterior is an area-interaction process*

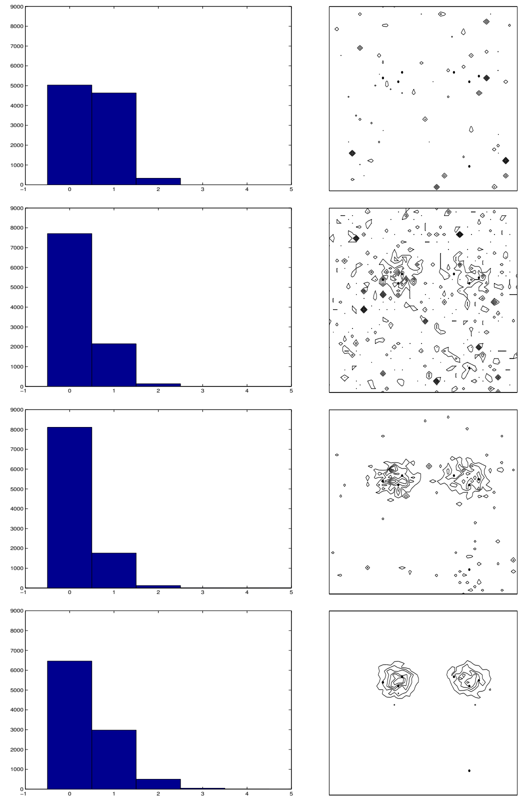

Figure 1.2 *Features of the prior and posterior distributions using the Neyman–Scott model with a Thomas intensity,* $\epsilon = 0.1$ *and* $\sigma = 0.07$*, and a Strauss prior,* $\beta_X = 0.1$*,* $r = 0.1$*, and* $\gamma_X = 0.1$ *The dataset consists of seven points in the unit square. Left: histogram of the number of landmarks. Right: contour plot of the intensity. Top: the Strauss prior.* $\epsilon = 0.1$ *and* $\sigma = 0.07$*. Row 2: the posterior with* $\kappa^* = \epsilon$*. Row 3: the posterior with* $\kappa^* = 2\epsilon$*. Row 4: the posterior with* $\kappa^* = 5\epsilon$*. In each case,* $10{,}000$ *exact samples were drawn using the Kendall–Møller algorithm.*

with $\nu(\cdot) = \int_{\cdot} \beta_Y(t)\,dt$ *and parameters* $\beta = \beta_X$*,* $\gamma = e$*:*

$$p_{X|Y=\mathbf{y}}(\mathbf{x}) \propto \beta_X^{n(\mathbf{x})} e^{-\nu(S(\mathbf{x}))} 1\{\mathbf{y} \subset S(\mathbf{x})\}.$$

Proof: Note that

$$\alpha_Y(\mathbf{x}) = \left[\int_{\{\mathbf{v}:\mathbf{v}\subset S(\mathbf{x})\}} \prod_{j=1}^{n(\mathbf{v})} \beta_Y(v_j)\, \pi(d\mathbf{v}) \right]^{-1}$$

$$= \left[\int_W \prod_{j=1}^{n(\mathbf{v})} \beta_Y(v_j) 1\{v_j \in S(\mathbf{x})\}\, \pi(d\mathbf{v}) \right]^{-1}.$$

Then, applying Lemma 1.2.1 to the intensity $\beta(t) = \beta_Y(t) 1\{t \in S(\mathbf{x})\}$,

$$\alpha_Y(\mathbf{x}) = \exp\left\{ |W| - \int_{S(\mathbf{x})} \beta_Y(t)\, dt \right\}.$$

Since $\gamma_X = 1$ and $e^{|W|}$ can be absorbed into the normalizing constant, the result follows by substitution in (1.1). □

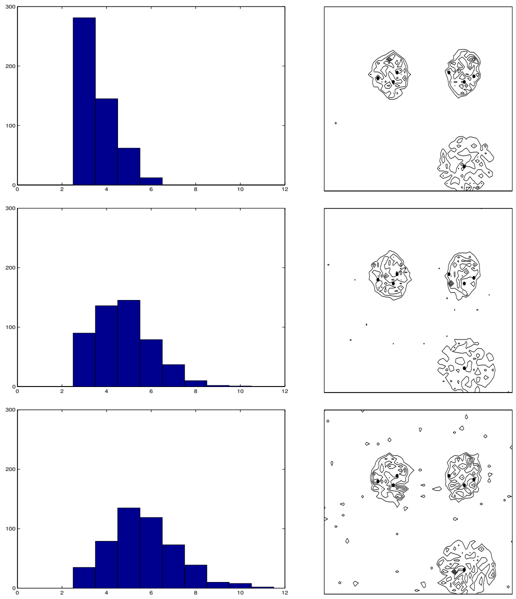

Figure 1.3 *Features of the posterior using the pure silhouette model,* $r_{\text{sil}} = 0.15$, *and a Poisson prior. The observation model is homogeneous Poisson, with* $\beta_Y = 3$. *Top row:* $\beta_X = 0.693$, *so that the prior probability of the empty configuration is* 0.5. *Middle row:* $\beta_X = 2$. *Bottom row:* $\beta_X = 3$. *In each case 500 exact samples were drawn using the Kendall–Møller algorithm.*

Thus the unrestricted posterior is attractive ($\gamma > 1$) in this case. For a homogeneous Poisson likelihood with intensity β_Y, the above result reduces to

$$p_{X|Y=\mathbf{y}}(\mathbf{x}) \propto \beta_X^{n(\mathbf{x})} (e^{\beta_Y})^{-|S(\mathbf{x})|} 1\{\mathbf{y} \subset S(\mathbf{x})\}. \qquad (1.6)$$

To implement perfect sampling for (1.6), use $K_{\text{post}} = K_{\text{prior}} = \beta_X$. Figure

1.3 illustrates this example with the same data as Figure 1.2. For this figure we varied the prior intensity β_X, while keeping all other hyperparameters constant. In the top row the prior probability of the empty configuration is 0.5 ($\beta_X = 0.693$). The Kendall–Møller algorithm along with rejection sampling provided 500 exact samples in about 72 hours on a Silicon Graphics workstation. The sampling in this case took much longer than for the Neyman–Scott model (even for a reduced number of samples) due to an acceptance rate in the rejection sampling step of approximately $1/44,000$; this was caused by the difficulty of covering the data by the silhouette of the landmarks. The results are markedly different from the Neyman–Scott model (Figure 1.2): three distinct circular clusters are readily apparent.

Now suppose that the prior is an arbitrary locally stable point process (and the observation process is still inhomogeneous Poisson with intensity $\beta_Y(t)$). From the expression for $\alpha_Y(\cdot)$ in the proof of Theorem 1.4.1, we see that the Papangelou conditional intensity corresponding to $\alpha_Y(\cdot)$ is attractive:

$$\frac{\alpha_Y(\mathbf{x} \cup \{\xi\})}{\alpha_Y(\mathbf{x})} = \exp\left(-\int_{S(\{\xi\})\setminus S(\mathbf{x})} \beta_Y(t)\,dt\right)$$

for $\xi \in W\setminus\mathbf{x}$, which is increasing in $\mathbf{x}$. We conclude that the unrestricted posterior is attractive if the prior is attractive. Perfect sampling is feasible for either an attractive or a repulsive prior.

One interesting choice of prior is the area-interaction process with parameters β_X, γ_X, $\nu =$ Lebesgue measure, and interaction radius r taken to coincide with the silhouette radius. Then the unrestricted posterior density (cf. (1.6)) is an area-interaction process with parameters $\beta = \beta_X$ and $\gamma = \gamma_X e^{\beta_Y}$:

$$f_{X|Y=\mathbf{y}}(\mathbf{x}) = \beta_X^{n(\mathbf{x})} (\gamma_X e^{\beta_Y})^{-|U(\mathbf{x})|}. \tag{1.7}$$

The unrestricted posterior is then repulsive if $\beta_Y \leq -\log(\gamma_X)$ and attractive if $\beta_Y \geq -\log(\gamma_X)$.

Beyond the examples we have considered, it may be difficult to check the attractive or repulsive properties; in that case, a computationally feasible approach to perfect sampling for the posterior does not appear to be readily available. For the general cluster model with $0 < c < 1$, the limits of integration for $\alpha_Y(\mathbf{x})$ are difficult to handle. If $c = 0$ this integral is over the entire window W. If $c = 1$ this integral is over the silhouette $S(\mathbf{x})$. But for the general case no such simplification seems available, and the resulting Papangelou conditional intensity is intractable. A fully Bayesian treatment in which the various hyperparameters are assigned priors may also be difficult to implement using perfect sampling.

1.5 Leukemia incidence in upstate New York

The study region is comprised of 790 census tracts in an eight county region of upstate New York; it includes the cities of Syracuse in the north and Binghamton in the south. The 1980 U.S. census reported a total of 1,057,673 residents. Leukemia incidence was recorded by the New York Department of Health during the years 1978–82, see Waller et al. (1992, 1994). The data are available from the Statlib archive (`lib.stat.cmu.edu`). In some instances a case could not be associated with a unique census tract, resulting in many fractional counts. Our approach does not accommodate this type of data, so we follow Ghosh et al. (1999) and group the 790 census tracts into 281 blocks in order to identify most of the cases with a specific block. Less than 10% of all cases could not be identified with a specific block, and such cases are excluded from our analysis.

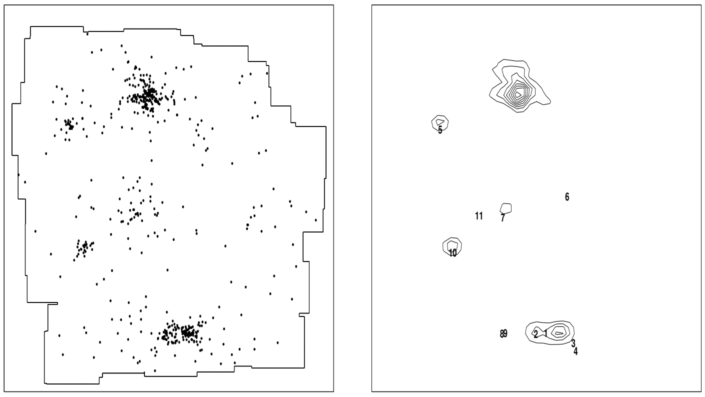

Figure 1.4 *Left: Locations of* 552 *leukemia cases in upstate New York, along with an approximate outline of the eight county study region. The rectangular region is* 1×1.2 *square units. Right: contour plot of the population density* $\lambda(t)$, *and the locations of the* 11 *hazardous waste sites.*

The locations of the centroids of the census blocks are available, but precise locations of the leukemia cases are not. Our methods require these precise locations, so we randomly dispersed the cases throughout their corresponding census blocks. An approximation to the census blocks was achieved using the Voronoi tesselation (Green and Sibson, 1978) with the Euclidean distance metric: each point in the region was identified with the closest centroid. All the cases identified with a particular centroid were then randomly dispersed in the block identified with that centroid; if there was exactly one case in a block we placed it at the centroid. Five independent randomizations of the case locations (within census blocks) are used in our study; see the left panel of Figure 1.4 for one such example. This provides an ad-hoc sensitivity analysis for our approach. The approximate study region is also indicated in the left panel of Figure 1.4, but to avoid edge effects we consider the enlarged rectangular area (roughly 145×174 square kilometers) as the support for the landmarks.

Our analysis is based on the Neyman–Scott model with the leukemia intensity rate specified by

$$\mu(t|\mathbf{x}) = \lambda(t)\left(\epsilon + \sum_{i=1}^{n(\mathbf{x})} h(t|x_i)\right),$$

where $\lambda(t)$ adjusts for population density, $h(t|x)$ is the Thomas intensity (1.5), and $\epsilon > 0$ is a background leukemia incidence rate. Our earlier treatment of the Neyman–Scott model extends without change to this form of the model because $\lambda(t)$ does not depend on $\mathbf{x}$. We use a Strauss prior for the landmarks $\mathbf{x}$. For $\lambda(t)$ we use a smoothed version of the population density based on the 1980 U.S. census (extrapolated to the large rectangular region), see the right panel of Figure 1.4; this plot also gives the locations of the 11 inactive hazardous waste sites suspected of causing elevated leukemia incidence rates. The original dimensions (kilometers) of the rectangular region in Figure 1.4 are divided by 145 for the purpose of this analysis, resulting in a rectangular region which is 1×1.2 square units.

For the interaction radius in the Strauss prior we use $r = 0.1$ (approximately 14.5 kilometers), and the interaction parameter is taken as $\gamma_X = 0.1$. The choice of these parameter values is rather arbitrary, but recall from the toy example (Section 1.4.1) that varying these values seemed to have little effect on the posterior distribution. In the Neyman-Scott model we specify $\epsilon = 5.2 \times 10^{-4}$, which is the average leukemia incidence rate of 5.2 cases per ten thousand residents of the study region. For the standard deviation of the Thomas intensity we choose $\sigma = .01$ (approximately 1.45 km.). These hyperparameter values will remain fixed throughout our analysis of the New York leukemia data.

In the context of the leukemia data and toxic waste sites the choice of the prior hyperparameters is of great importance. If we believe that the incidence locations are simply a function of population density, and are not influenced by any underlying sources (such as the waste sites), then the prior should place high probability on the empty configuration. This is certainly a reasonable (and perhaps conservative) assumption. Should the posterior indicate the existence of one or more landmarks (i.e. place low probability on the empty configuration), then one would have a strong argument in favor of clustering. It may also be reasonable to place high prior probability on the existence of one landmark, but to posit the existence of more than one landmark is probably stacking the deck a bit too much. It should be noted that the locations of the eleven toxic waste sites are not a part of our model.

In our initial analysis we choose a prior which places high probability on the empty configuration: $\beta_X = 0.113$. Figure 1.5 is based on the data shown in the left panel of Figure 1.4, and shows the effect of varying the landmark weight κ^* on the number of posterior points produced. The number of landmarks given by the prior distribution is included for reference. For each figure we used a Silicon Graphics workstation to produce 1000 samples using the Kendall–Møller algorithm. (It took less than one minute to produce the samples from the prior, and just over 4.5 hours for $\kappa^* = 0.23\epsilon$.) It is clear from this figure that increasing the landmark weight has little effect on the number of posterior landmarks. In fact, the posterior probability of the empty configuration is approximately 80% for each value of κ^*, even though the prior probability of the empty configuration is about 50%.

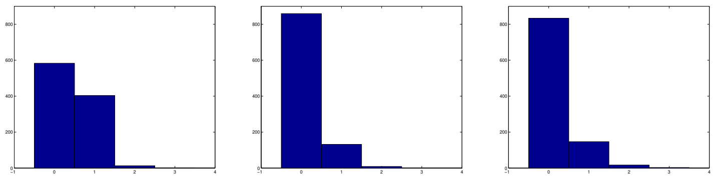

Figure 1.5 *Histograms of the number of points in the prior and two posterior distributions. Each plot is based on* 1000 *exact samples. The prior is Strauss, with* $r = 0.1$, $\beta_X = 0.113$ *and* $\gamma_X = 0.1$. *The likelihood is Neyman-Scott with a Thomas intensity,* $\epsilon = 5.2 \times 10^{-4}$, $\sigma = 0.01$. *Prior distribution; posterior with* $\kappa^* = 0.1\epsilon$*;* $\kappa^* = 0.23\epsilon$.

Figure 1.6 is similar to Figure 1.5, except that we use $\beta_X = 0.5$. The histogram for the prior shows that most of the prior weight is placed on configurations comprised of a single landmark. Nevertheless, the posterior distributions still strongly favor the empty configuration. As the landmark weight κ^* increases there is a slight movement towards configurations with at least one landmark. At large values of κ^*, limitations in computing power come into play. It took 146 hours to complete the simulation with $\kappa^* = 0.23\epsilon$. Initial runs with $\kappa^* = 0.25\epsilon$ indicate that the time required for a complete run is prohibitive, and the resulting histograms are virtually identical to that for $\kappa^* = 0.23\epsilon$. It is unfortunate that our exploration of this data is limited in this manner. Based on the results seen in the toy example of Section 1.4.1 and in the redwood seedling example of the next section, it seems likely that larger values of κ^* would result in a significant increase in the number of posterior landmarks.

Figure 1.7 gives the posterior intensity contour plots when $\beta_X = 0.5$, for three values of κ^*. As κ^* increases, the posterior landmarks concentrate in two small areas (close to Syracuse and Binghamton) even though the

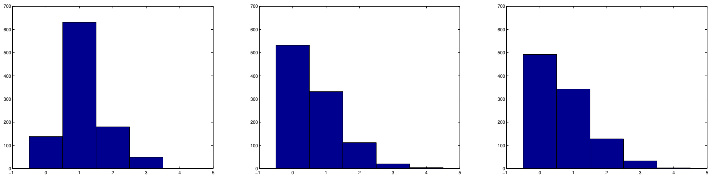

Figure 1.6 *Same as Figure 1.5 except that* $\beta_X = 0.5$.

number of posterior landmarks does not change significantly (see Figure 1.6). Figure 1.8 show the relationship of the hazardous waste site locations to areas of high posterior landmark intensity. Note that several of the waste sites (1, 2, 3, and 4) are located close to areas of high intensity.

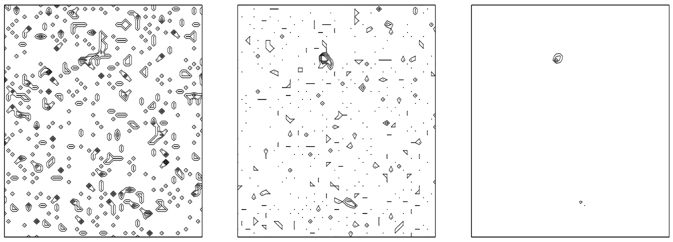

Figure 1.7 *Intensity contour plots for posterior distributions with* $\kappa^* = 0.1\epsilon$, $\kappa^* = 0.15\epsilon$, *and* $\kappa^* = 0.2\epsilon$. *Each plot is based on* 1000 *exact samples. The prior is Strauss, with interaction radius* $r = 0.1$, $\beta_X = 0.5$ *and* $\gamma_X = 0.1$. *The likelihood is Neyman-Scott with a Thomas intensity,* $\epsilon = 5.2 \times 10^{-4}$, $\sigma = 0.01$.

To assess the significance of an elevated leukemia rate in the neighborhood of a given site, we compare the 'observed' with the 'expected' posterior landmark distribution. The relevant null hypothesis here is that the leukemia cases form an inhomogeneous Poisson process with intensity $\rho\lambda(t)$, where ρ is the average leukemia rate throughout the study region. To sample from the null distribution, we generate an artificial dataset using independent Poisson counts for each census tract, then analyze the artificial data in the same way as the original data. For the artificial data, histograms of the number of posterior landmarks (not shown) look almost identical to those in Figure 1.6, and intensity plots (also not shown) like those shown in Figure 1.7.

In Figure 1.9 we compare the observed and expected posterior probabilities of at least one landmark within a given distance (0–7.25 kms) of site 1 (with the same hyperparameters as in Figure 1.6). The 20 dotted lines correspond to samples from the null distribution (20 different artificial

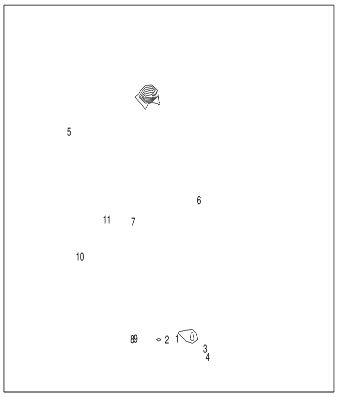

Figure 1.8 *Posterior intensity map for the leukemia data, Neyman–Scott model with* $\epsilon = 5.2 \times 10^{-4}$, $\sigma = 0.01$, $\kappa^* = 0.23\epsilon$, *and a Strauss prior with interaction radius* $r = 0.1$, $\beta_X = 0.5$ *and* $\gamma_X = 0.1$. *Locations of the* 11 *hazardous waste sites are included.*

datasets), and the 5 solid lines correspond to the data (five random dispersions of the leukemia cases throughout their corresponding census blocks). For this site, it is clear that increasing the landmark weight increases the disparity between the real and the simulated datasets. For $\kappa^* \geq 2.0$, these plots seem to provide some evidence of elevated leukemia rates in the neighborhood of site 1. (It should be noted that the vertical scales are different for each of the plots.)

We repeated the comparison shown in Figure 1.9 for each of the 11 hazardous waste sites, and found no evidence of elevated leukemia rates in the neighborhood of any site other than site 1. The complete set of comparisons can be seen in Loizeaux and McKeague (2001).

Ghosh et al. (1999) found little evidence of elevated leukemia rates in the proximity of any of the waste sites. They applied a hierarchical Bayes generalized linear model to estimate the leukemia incidence rate locally, and included the hazardous waste site locations in their model (in particular, using the inverse distance to the nearest hazardous waste site as a covariate). Ahrens et al. (1999) found increased rates at three of the sites (1, 5 and 7). Denison and Holmes (2001) found significant evidence of increased leukemia rates in the neighborhoods of sites 1 and 2. Their Bayesian partition model assumes that the region can be split into local areas in which the leukemia counts come from the same distribution. The Voronoi tesselation (Green and Sibson, 1978) is used to define these areas. Our results are mixed. On the one hand, the posterior distributions place

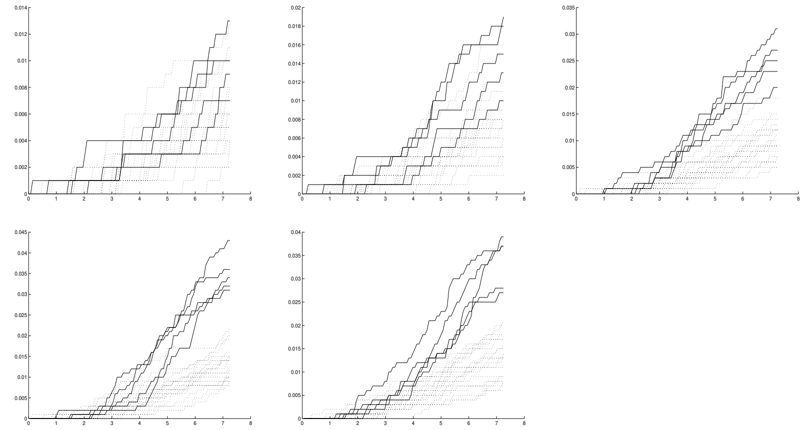

Figure 1.9 *Posterior observed (solid lines) and expected (dotted lines) probabilities of at least one landmark within a given distance (in kms.) of site 1. Neyman–Scott model with* $\epsilon = 5.2 \times 10^{-4}$, $\sigma = 0.01$, *and a Strauss prior with interaction radius* $r = 0.1$, $\beta_X = 0.5$ *and* $\gamma_X = 0.1$. *Top row:* $\kappa^* = 0.1\epsilon$; $\kappa^* = 0.15\epsilon$; $\kappa^* = 0.2\epsilon$. *Bottom row:* $\kappa^* = 0.22\epsilon$; $\kappa^* = 0.23\epsilon$. *Each plot is based on* 1000 *exact samples.*

approximately 50% of their mass on the empty configuration. But on the other hand, a comparison of the actual data with artificial datasets based only on the population density shows a marked difference at site 1.

1.6 Redwood seedlings data

These data have been analyzed by Strauss (1975), Ripley (1977), Diggle (1978), and van Lieshout (1995), among others. Our results are shown in Figure 1.10.

The prior hyperparameters have been chosen to produce empty landmark configurations with probability 50%. For $\kappa^* = 0.5\epsilon$ the posterior shows a clear shift towards the empty configuration, but this tendency is reversed as κ^* is increased to ϵ. In addition, the intensity plots indicate that the posterior landmarks start to collect around the data as κ^* increases.

Acknowledgements. This research was partially supported by NSA Grant MDA904-99-1-0070 and NSF Grant 9971784. Equipment support was provided under ARO Grant DAAG55-98-1-0102 and NSF Grant 9871196.

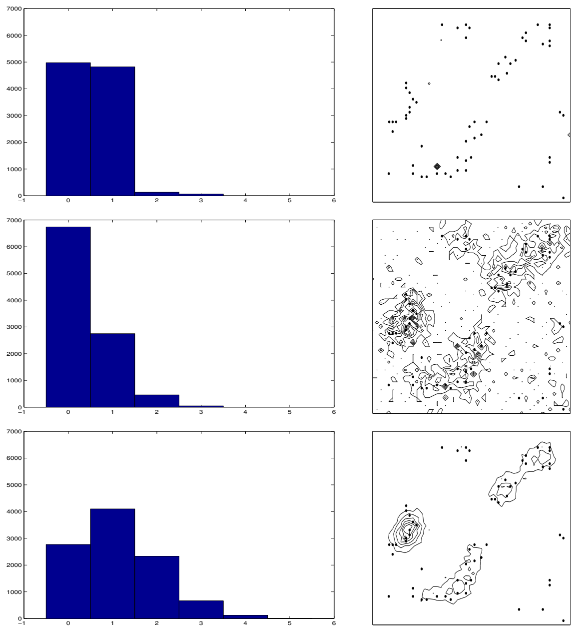

Figure 1.10 *Features of the prior and two posterior distributions for the redwood seedling data using the Neyman–Scott model, $\epsilon = 0.1$ and $\sigma = 0.07$. Top row: the Strauss prior, with interaction radius $r = 0.2$, $\beta_X = 0.124$, and $\gamma_X = 0.1$. Middle row: posterior with $\kappa^* = 0.5\epsilon$. Bottom row: posterior with $\kappa^* = \epsilon$. Left: histogram of the number of landmarks. Right: contour plot of the intensity. In each case 10,000 exact samples were drawn using the Kendall–Møller algorithm.*

References

Ahrens, C., Altman, N., Casella, G., Eaton, M., Hwang, J. T. G., Staudenmayer, J. and Stefanscu, C. (1999). Leukemia clusters and TCE waste sites in upstate New York: How adding covariates changes the story. Preprint.

Baddeley, A. J. and van Lieshout, M. N. M. (1993). Stochastic geometry models in high-level vision. In K. V. Mardia and G. K. Kanji, editors, *Advances in Applied Statistics*, Statistics and Images: 1, 231–256, Carfax Publishing.

Baddeley, A. J. and van Lieshout, M. N. M. (1995). Area-interaction point processes. *Ann. Inst. Statist. Math.* 47, 601–619.

Baddeley, A. J. and Møller, J. (1989). Nearest-neighbor Markov point processes and random sets. *Int. Statist. Rev.* 57, 89–121.

Cai, Y. and Kendall, W. S. (1999). Perfect implementation of simulation for conditioned Boolean model via correlated Poisson random variables. `http://www.warwick.ac.uk/statsdept/Staff/WSK/ppt.html`

Carter, D. S. and Prenter, P. M., (1972). Exponential spaces and counting processes. *Z. Wahr. verw. Geb.* 21, 1–19.

Casella, G., Mengersen, K. L., Robert, C. P. and Titterington, D. M. (1999). Perfect slice samplers for mixtures of distributions. `http://www.ceremade.dauphine.fr/~xian/publications.html`

Daley, D. J. and Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes.* Springer, New York.

Denison, D. G. T. and Holmes, C. C. (2001). Bayesian partitioning for estimating disease risk. *Biometrics* 57, 143–149.

Diggle, P. J. (1978). On parameter estimation for spatial point processes. *J. Roy. Statist. Soc. Ser. B* 40, 178–181.

Diggle, P. J. (1983). *Statistical Analysis of Spatial Point Patterns.* Academic Press

Geyer, C. J. (1999). Likelihood inference for spatial point processes. In O. E. Barndorff–Nielsen, W. S. Kendall, and M. N. M. van Lieshout, editors, *Stochastic Geometry: Likelihood and Computation*, pp. 79–140. Chapman and Hall.

Geyer, C. J. and Møller, J. (1994). Simulation and likelihood inference for spatial point processes. *Scand. J. Statist.* 21, 359–373.

Ghosh, M., Natarajan, K., Waller, L. and Kim, D. (1999). Hierarchical Bayes GLMs for the analysis of spatial data: An application to disease mapping. *J. Statist. Planning Inference* 75, 305–318.

Green, P. J. and Murdoch, D. J. (1999). Exact sampling for Bayesian inference: towards general purpose algorithms. In J. M. Bernardo et al., editors, *Bayesian Statistics 6*, Oxford University Press. Presented as an invited paper at the 6th Valencia International Meeting on Bayesian Statistics, Alcossebre, Spain, June 1998.

Green, P. J. and Sibson, R. (1978). Computing Dirichlet tesselations in the plane. *Computer Journal* 21, 168–173.

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.

Hurn, M. (1998). Confocal flourescence microscopy of leaf cells: an application of Bayesian image analysis. *Appl. Statist.* 47, 361–377.

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

Kendall, W. S. and Møller, J. (2000). Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Adv. in Appl. Probab.* 32, 844–865.

Kendall, W. S. and Thönnes, E. (1999). Perfect simulation in stochastic geometry. `http://www.warwick.ac.uk/statsdept/Staff/WSK/papers/323.ps.gz`.

Lawson, A. B. and Clark, A. (1999). Markov chain Monte Carlo for putative sources of hazard and general clustering. In A. B. Lawson, D. Böhning, A. Biggeri, E. Lesaffre, J.-F. Viel, editors, *Disease Mapping and Risk Assessment for Public Health.* Wiley.

Loizeaux, M. and McKeague, I. W. (2001). Perfect sampling for posterior landmark distributions with an application to the detection of disease clusters. *IMS Lecture Notes - Monograph Series*, Volume 37, 321–331.

Møller, J. (1999). Markov chain Monte Carlo and spatial point processes. In O. E. Barndorff–Nielsen, W. S. Kendall, and M. N. M. van Lieshout, editors, *Stochastic Geometry: Likelihood and Computation*, pp. 141–172. Chapman and Hall.

Møller, J. (2001). A review of perfect simulation in stochastic geometry. *IMS Lecture Notes - Monograph Series*, Volume 37, 333–355.

Møller, J. and Nicholls, G. K. (1999). Perfect simulation for sample-based inference.
`http://www.math.auckland.ac.nz/~nicholls /linkfiles/papers/perfect.sim.temp1.ps.gz`.

Neyman, J. and Scott, E. L. (1972). Processes of clustering and applications. In P. A. W. Lewis, editor, *Stochastic Point Processes*, Wiley, New York, pp. 641–681.

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

Ripley, B. D. (1977). Modelling spatial patterns (with discussion). *J. Roy. Statist. Soc. Ser. B* 39, 172–212.

Stoyan, D., Kendall, W. S. and Mecke, J. (1995). *Stochastic Geometry and its Applications*, 2nd ed., Wiley, Chichester.

Strauss, D.J. (1975). A model for clustering. *Biometrika* 62, 467–475.

Van Lieshout, M. N. M. (1995). *Stochastic Geometry Models in Image Analysis and Spatial Statistics.* CWI Tract 108. Stichting Mathematisch Centrum, Amsterdam.

Van Lieshout, M. N. M. (2000). *Markov Point Processes and their Applications.* Imperial College Press, London.

Waller, L., Turnbull, B., Clark, L. and Nasca, P. (1992). Chronic disease surveillance and testing of clustering of disease and exposure: application to leukemia incidence and TCE-contaminated dump sites in upstate New York. *Environmetrics* 3, 281–300.

Waller, L., Turnbull, B., Clark, L. and Nasca, P. (1994). Spatial pattern analysis to detect rare disease clusters. In N. Lange, L. Ryan, L. Billard, D. Brillinger, L. Conquest, J. Greenhouse, editors, pp. 3–23, *Case Studies in Biometry*. Wiley.

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