4

An Introduction to Simulation-based Inference for Spatial Point Processes

Jesper Møller
Rasmus P. Waagepetersen

4.1 Introduction

Spatial point processes play a fundamental role in spatial statistics. In the simplest case they model "small" objects that may be identified by a map of points showing stores, towns, plants, nests, or cases of a disease observed in a two dimensional region or galaxies observed in a three dimensional region. The points may be decorated with marks (such as sizes or types) whereby marked point processes are obtained. The areas of applications are manifold: astronomy, geography, ecology, forestry, spatial epidemiology, image analysis, and many more. Currently spatial point processes is an active area of research, which probably will be of increasing importance for many new applications.

This chapter aims at collecting some of the recent theoretical advances and examples of applications in simulation-based inference for spatial point processes in a concise manner. We focus mainly on general ideas and methodology without going too much into technical details. Just about all the material, and more, will appear in Møller & Waagepetersen (2003). Sometimes the exposition will be biased towards our own work and interests, but most of the time we follow the mainstream of recent contributions in this research area. Particularly we focus on Cox processes and Markov point processes, which we find are the most useful models in data analysis. Moreover, we relax the stationary assumption, which seems very timely.

The chapter is organised as follows. Section 4.2 introduces two examples of applications which are used throughout this chapter. Section 4.3 describes in more detail what is meant by a spatial point process. Section 4.4 deals with Poisson point processes. Section 4.5 surveys explanatory analysis and model validation based on various kind of summary statistics. Sections 4.6–4.7 are the two main sections. Section 4.6 considers cluster processes and Cox processes, particularly log Gaussian Cox processes, and discusses simulation-based inference for clustered point patterns. Section 4.7 deals with different aspects of model construction, simulation (including perfect simulation), and inference for Markov point processes. Finally, Section 4.8 contains some concluding remarks.

4.2 Illustrating examples

Examples 1 and 2 below are used for illustrative purposes throughout this chapter. In Example 1 we consider what is later on called a simple point process, while the discs in Example 2 will later on be treated as a so-called marked point process.

4.2.1 Example 1: Weed plants

Figure 4.1 shows the position of 976 weed plants (*Trifolium* spp./clover) observed within 45 metal frames on a Danish barley field. This point pattern is a subset of a much larger dataset analysed in Brix & Møller (2001) and Brix & Chadoeuf (2000) where several weed species at different dates were considered. Note that we have rotated the design 90° in Figure 4.1. The 45 frames are of size 30×20 cm^2, and they are organised in 9 groups each containing 5 frames, where the vertical and horizontal distances between two neighbouring groups are 1 m and 1.5 m, respectively. The size of the experimental area is 7.5×5 m^2, where the longest side agrees with the ploughing direction. Note the trend in the point pattern: in general more weed plants occur in the upper frames (i.e. the frames to the left in Figure 4.1).

4.2.2 Example 2: Norwegian spruces

The Norwegian spruce data (Fiksel 1984, Penttinen, Stoyan & Henttonen 1992, Stoyan, Kendall & Mecke 1995, Goulard, Särkkä & Grabarnik 1996) is an example of a pattern of discs. The data are shown in Figure 4.2, where the centres of the 134 discs are the positions of the spruces observed in a rectangular window of size 56×38 m^2, and the radii are the stem diameters multiplied by 5. As discussed in Penttinen et al. (1992) and Goulard et al. (1996) the "influence zone" of a tree is about 5 times the stem diameter.

4.3 What is a spatial point process?

A formal answer to this question is given below by considering in Section 4.3.1 the special case of simple point processes in $\mathbb{R}^d$, in Section 4.3.2

FIGURE 4.1. Positions of weed plants when the design is rotated 90°.

some special cases of marked point processes which all can be related to the examples in Section 4.2, and in Section 4.3.3 a general setting and notation used throughout this text.

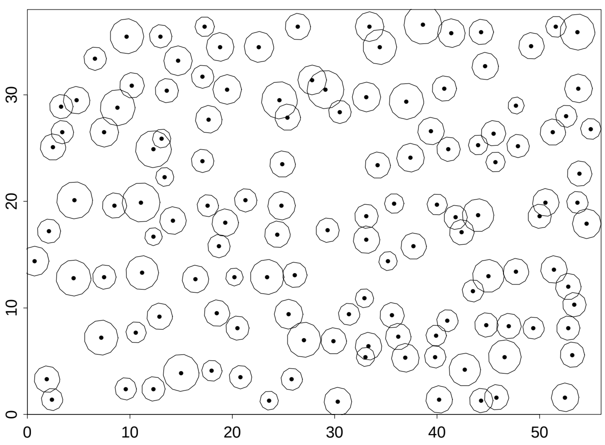

FIGURE 4.2. Positions of Norwegian spruces. The radii of the discs equal 5 times the stem diameters.

4.3.1 Simple point processes in $\mathbb{R}^d$

A simple point process in the d-dimensional Eucledian space $\mathbb{R}^d$ may be considered as a random countable set $X \subset \mathbb{R}^d$. What is meant here by random will be made precise in a moment. The elements in X are called *points*, and they may represent the occurrence of some event like the occurrence of a weed plant. Equivalently we may view X as a counting measure defined by $X(A) = n(X_A)$ for bounded Borel sets $A \subset \mathbb{R}^d$, where $n(X_A)$ denotes the cardinality of $X_A = X \cap A$. We here for simplicity restrict attention to the case of *simple point processes*, i.e. when there are no multiple points; a counting measure with multiplicities may in fact be considered as a special case of a marked point process as described below. Moreover, we exclude the case of accumulating points, i.e. $X(A) < \infty$ whenever A is bounded.

In practice $X = X_S$ will be concentrated on a bounded set $S \subset \mathbb{R}^d$, but for mathematical convenience, or if S is large or unknown, we may let $S = \mathbb{R}^d$. Usually in applications S is d-dimensional (we may formalize this by saying that S is topologically regular, i.e. it is equal to the closure of its interior), and most examples in the point process literature concerns the planar case $d = 2$. In other cases S may be a lower-dimensional manifold, e.g. the $(d-1)$-dimensional unit sphere. In practice we observe only X_W where $W \subseteq S$ is some bounded observation window like in Figures 4.1 and 4.2. If X is not concentrated on W, it may be important to take *boundary or edge effects* into account in the statistical analysis, as the unobserved points outside W may affect X_W.

4.3.2 Marked point processes in $\mathbb{R}^d$

Suppose now that Y is a simple point process in $\mathbb{R}^d$ and a random "mark" $m_\xi \in \mathcal{M}$ is attached to each point $\xi \in Y$. Then $X = \{(\xi, m_\xi) : \xi \in Y\}$ is called a *marked point process* in $\mathbb{R}^d$ with *mark space* $\mathcal{M}$. The marks may be dependent or not of each other and of Y; again what is meant by randomness is made precise below.

One simple example is a disc process as considered in Example 2 (Norwegian spruces), letting $\mathcal{M} = (0, \infty)$ and identifying (ξ, m_ξ) with the disc with centre ξ and radius m_ξ. Similarly, we obtain marked point processes for other kinds of geometric objects (line segments, ellipses, etc.), also called *germ-grain models* where ξ (the germ) specifies the location of the object m_ξ (the grain). See Section 3.6.3.

Another example is a *multivariate* or *multi-type point process*, where $\mathcal{M} = \{1, \dots, k\}$ and the marks specify k different types of points (e.g. different types of weed plants).

4.3.3 General setting and notation

Many readers (particularly those who are unfamiliar with measure theory) may prefer to skip the general setting described below. However, we recommend the reader to at least notice the meaning of the following notation which is used throughout this chapter.

Notation

We consider X to be a locally finite subset of a rather general metric space S with metric $d(\cdot, \cdot)$. Here locally finiteness means that $X(A) = n(X_A) < \infty$ for any bounded set $A \subseteq S$. The state space of X is denoted N and consists of all locally finite point configurations in S:

$$N = \{x \subset S : x(A) < \infty \text{ for all } A \in \mathcal{B}_0\}$$

where $\mathcal{B}_0$ denotes the class of bounded Borel sets contained in S. This state space is equipped with a suitable σ-algebra denoted $\mathcal{N}$ (see below). Elements of S and N are usually denoted $\xi, \eta, \ldots \in S$ and $x, y, \ldots \in N$. We abuse the notation and write $x \cup \xi$ for $x \cup \{\xi\}$ (with $x \in N$ and $\xi \in S \setminus x$), $y \setminus \eta$ for $y \setminus \{\eta\}$ (with $y \in N$ and $\eta \in S$), etc.

Measure theoretical details

Formally, we assume S to be a Polish space, i.e. a complete separable metric space; though this assumption may be weakened, it is commonly satisfied in applications and ensures the validity of some desirable properties as listed below. Moreover, we assume X to be defined on some underlying probability space $(\Omega, \mathcal{F}, P)$ so that measurability of X means that $X(A) : \Omega \to \mathbb{N}$ is a measurable function whenever $A \in \mathcal{B}_0$. In other words, we equip N with the smallest σ-algebra $\mathcal{N}$ containing all sets of the form $\{x \in N : x(A) = m\}$ with $A \in \mathcal{B}_0$ and $m \in \mathbb{N}_0$. It can be shown (Matheron 1975, Daley & Vere-Jones 1988) that

- N is a metric space, $\mathcal{N}$ is the corresponding Borel σ-algebra, and $\mathcal{N}$ is countably generated;
- the distribution of X is determined by the *void probabilities* as given by $P(X(A) = 0)$, $A \in \mathcal{B}_0$.

How is this related to simple and marked point processes considered so far?

For a simple point process in $\mathbb{R}^d$ we let $d(\cdot,\cdot)$ be the usual Eucledian metric or distance $\|\cdot\|$.

A marked point process with locations in $\mathbb{R}^d$ and mark space $\mathcal{M}$ can be considered as a point process defined on $S = \mathbb{R}^d \times \mathcal{M}$. Then S becomes a Polish space, if we assume the mark space $\mathcal{M}$ to be a Polish space with metric $d_{\mathcal{M}}$, and equip S with the metric

$$d((\xi_1, m_{\xi_1}), (\xi_2, m_{\xi_2})) = \max\{\|\xi_1 - \xi_2\|, d_{\mathcal{M}}(m_{\xi_1}, m_{\xi_1})\}.$$

This is also a natural metric in the sense that the Borel σ-algebra for S agrees with the product σ-algebra of the Borel sets in $\mathbb{R}^d$ and $\mathcal{M}$. If the marks are geometrical objects, we may use the formalism of stochastic geometry, considering $\mathcal{M}$ as the space of compact subsets of $\mathbb{R}^d$ equipped with the Hausdorff metric defined by

$$d_{\mathcal{M}}(A, B) = \inf\{r \geq 0 : A \subset B \oplus b(0,r),\, B \subset A \oplus b(0,r)\}$$

where $\oplus$ denotes Minkowski-addition, i.e. $A \oplus B = \{a + b : a \in A, b \in B\}$. Then M is a Polish space (Matheron 1975). For example for balls $A = b(0, r_1)$ and $B = b(0, r_2)$ with centre 0 and radii $0 < r_1 \leq r_2 < \infty$, $d_{\mathcal{M}}(A, B) = r_2 - r_1$.

4.4 Poisson point processes

Poisson point processes play a fundamental role, as they serve as a tractable model class for "no interaction" or "complete spatial randomness" in spatial point patterns, and as reference processes when comparing and constructing more advanced point process models. Furthermore, the initial step of a point process analysis often consists of looking for discrepancies with a Poisson model. This may point to alternative models as discussed later in Section 4.5.

General definitions and properties of Poisson processes are reviewed in Section 4.4.1, the particular case of Poisson processes in $\mathbb{R}^d$ is considered in Section 4.4.2, and Section 4.4.3 concerns marked Poisson processes in $\mathbb{R}^d$. Further material can be found in Daley & Vere-Jones (1988), Kingman (1993), and Stoyan et al. (1995).

4.4.1 Definitions and properties

Let μ be a locally finite and diffuse measure defined on the Borel sets in S, i.e. $\mu(A) < \infty$ for $A \in \mathcal{B}_0$ and μ has no mass at any point in S. We say that X is a *Poisson point process* on S with *intensity measure* μ, and write

$$X \sim \text{Poisson}(S, \mu),$$

if X satisfies the following properties:

- for any $A \in \mathcal{B}_0$, $X(A) \sim \text{po}(\mu(A))$, the Poisson distribution with mean $\mu(A)$;
- *independent scattering:* for any disjoint Borel sets $A_1, \ldots, A_n \subseteq S$ with an arbitrary $n \geq 2$, $X(A_1), \ldots, X(A_n)$ are independent.

The assumption that μ is locally finite and diffuse ensures that X is a locally finite point process with no multiple points. The independent scattering property explains the terminology of "no interaction" and "complete spatial randomness". In the definition above we can replace the independent scattering property by

- for any $A \in \mathcal{B}_0$ with $\mu(A) > 0$, conditional on $X(A) = n$, the n points in X_A are mutually independent with common distribution $\bar{\mu}(\cdot) = \mu(\cdot \cap A)/\mu(A)$; this is called a *binomial point process* of n points with distribution $\bar{\mu}$.

It is therefore not hard to verify that there exists a well-defined point process with these properties.

The simplest way of characterising a Poisson point process is by its void probabilities,

$$P(X(A) = 0) = \exp(-\mu(A)), \quad A \in \mathcal{B}_0.$$

A less well-known but very useful characterisation of a Poisson process is provided by the *Slivnyak-Mecke theorem* (Mecke 1967): $X \sim \text{Poisson}(S, \mu)$ if and only if, for any measurable function $h : N \times S \to [0, \infty)$,

$$\mathbb{E} \sum_{\xi \in X} h(X \setminus \xi, \xi) = \int \mathbb{E}h(X, \xi)\mu(\mathrm{d}\xi). \tag{4.1}$$

The extended Slivnyak-Mecke theorem is obtained by induction: we have that $X \sim \text{Poisson}(S, \mu)$ if and only if, for any $n \in \mathbb{N}$ and any measurable function $h : N \times S^n \to [0, \infty)$,

$$\mathbb{E} \sum_{\xi_1, \ldots, \xi_n \in X}^{\neq} h(X \setminus \{\xi_1, \ldots, \xi_n\}, \xi_1, \ldots, \xi_n) =$$
$$\int \cdots \int \mathbb{E} h(X, \xi_1, \ldots, \xi_n) \mu(\mathrm{d}\xi_1) \cdots \mu(\mathrm{d}\xi_n), \tag{4.2}$$

where the $\neq$ over the summation sign means that the n points $\xi_1, \ldots, \xi_n$ are all different.

The class of Poisson processes is closed under two basic operations for point processes:

- *superpositioning:* if $X_1 \sim \text{Poisson}(S, \mu_1)$ and $X_2 \sim \text{Poisson}(S, \mu_2)$ are independent, then $X_1 \cup X_2 \sim \text{Poisson}(S, \mu_1 + \mu_2)$;
- *independent thinning:* if $X \sim \text{Poisson}(S, \mu)$ and $R(\xi) \sim \text{Uniform}[0, 1]$, $\xi \in S$, are mutually independent, and $p : S \to [0, 1]$ is a measurable function, then $Z = \{\xi \in X : R(\xi) < p(\xi)\} \sim \text{Poisson}(S, \nu)$ with $\nu(A) = \int_A p(\xi)\mu(\mathrm{d}\xi)$.

These statements are easily verified by considering the void probabilities of the superposition $X_1 \cup X_2$ and the thinned process Z.

4.4.2 Poisson processes in $\mathbb{R}^d$

Suppose that $X \sim \text{Poisson}(\mathbb{R}^d, \mu)$. If μ is absolutely continuous with respect to the Lebesgue measure, then its density $\rho(\xi) = \mathrm{d}\mu(\xi)/\mathrm{d}\xi$ is called the *intensity function*. Often in statistical modelling of a Poisson point process, one specifies a parametric model for the intensity function, cf. Section 4.5.5. This may depend on covariate information as e.g. in Rathbun (1996).

In the particular case where $\rho(\cdot) = \rho$ is constant, X is said to be a *homogeneous Poisson point process with intensity* ρ. This is equivalent to assuming *stationarity* of X under translations, that is the distribution of $X + s = \{\xi + s : \xi \in X\}$ is the same as that of X for any $s \in \mathbb{R}^d$. A homogeneous Poisson point process is also *isotropic* as its distribution is invariant under rotations in $\mathbb{R}^d$. In the special case $d = 1$, the distances between successive points of a homogeneous Poisson point process are independent and exponentially distributed with mean $1/\rho$.

Consider a Poisson point process X with intensity function ρ. It is in general impossible to specify the density of X with respect to another Poisson process. For example, a homogeneous Poisson point process with intensity $\rho_1 > 0$ is absolutely continuous with respect to another homogeneous Poisson point process with intensity $\rho_2 > 0$ if and only if $\rho_1 = \rho_2$. If $W \subset \mathbb{R}^d$ has Lebesgue measure (or area/volume) $|W| \in (0, \infty)$ and $\int_W \rho(\xi)\mathrm{d}\xi < \infty$, then X_W has a *density*

$$f_W(x) = \exp\left(|W| - \int_W \rho(\xi)\mathrm{d}\xi\right) \prod_{\xi \in x} \rho(\xi), \quad x \subset W,\ x(W) < \infty, \quad (4.3)$$

with respect to the *standard Poisson point process* Poisson($\mathbb{R}^d$, Lebesgue).

Simulation of a homogeneous Poisson point process X with intensity $\rho > 0$ within a d-dimensional box $B = [0, a_1] \times \cdots \times [0, a_d]$ is straightforward: first generate $N \sim \mathrm{po}(\rho a_1 \cdots a_d)$ and next N independent and uniformly distributed points in B. Alternatively, we may use that

- the first coordinates $\xi^{(1)}$ of points $\xi = (\xi^{(1)}, \xi^{(2)}) \in X$ with $\xi^{(2)} \in [0, a_2] \times \cdots \times [0, a_d]$ form a homogeneous Poisson process on the real line with intensity $\rho a_2 \cdots a_d$,
- the remaining components $\xi^{(2)}$ of such points are independent and uniformly distributed on $[0, a_2] \times \cdots \times [0, a_d]$.

For simulation within a ball in $\mathbb{R}^d$, it is more convenient to make a shift to polar coordinates and use a radial simulation procedure (Quine & Watson 1984).

Combining this with independent thinning we obtain a simple simulation procedure for inhomogeneous Poisson processes with an intensity function $\rho(\cdot)$ which is bounded by a constant c on $B \subset \mathbb{R}^d$: generate a homogeneous Poisson process on B with intensity c, and let the retention probabilities be $p(\xi) = \rho(\xi)/c,\ \xi \in B$.

4.4.3 Marked Poisson processes in $\mathbb{R}^d$

Suppose that $X \sim \text{Poisson}(\mathbb{R}^d \times \mathcal{M}, \mu)$. Independence between points and marks in X is equivalent to that μ factorises into a product measure $\mu = \nu \times Q$ where ν is a locally finite measure in $\mathbb{R}^d$ and Q is a probability measure describing the common distribution of the marks; it is simply called the *mark distribution*. We shall not pay much attention to marked Poisson processes, but refer the interested reader to Stoyan et al. (1995) and the references therein.

4.5 Summary statistics

Exploratory analysis for spatial point patterns and the validation of fitted models are often based on non-parametric estimates of various summary statistics, cf. e.g. Ripley (1977), Stoyan et al. (1995), and Ohser & Mücklich (2000). In this section we confine ourselves to summary statistics for a single point pattern X observed in a bounded planar window $W \subset \mathbb{R}^2$

with Lebesgue measure $|W| > 0$. Extensions to replicated point patterns and to marked point processes are sometimes obvious; see Diggle, Lange & Beněs (1991), Diggle, Mateu & Clough (2000), Baddeley, Moyeed, Howard & Boyde (1993), Schlather (2001), and the references therein.

Sections 4.5.1–4.5.2 consider summary statistics related to the first and second order moments of the counts $X(A),\ A \in \mathcal{B}_0$, while summary statistics based on distribution functions for interpoint distances are treated in Section 4.5.3.

4.5.1 First order characteristics

Just as for Poisson point processes in $\mathbb{R}^d$, we define the following concepts. The *intensity measure* μ of X is given by $\mu(A) = \mathbb{E}X(A)$ for Borel sets $A \subseteq \mathbb{R}^d$. If μ is absolutely continuous with respect to the Lebesgue measure, its density $\rho(\xi) = \mathrm{d}\mu(\xi)/\mathrm{d}\xi$ is called the *intensity function*. Loosely speaking, $\rho(\xi)\mathrm{d}\xi$ is the probability for the occurrence of a point in an infinitesimally small ball with centre ξ and area $\mathrm{d}\xi$. If moreover $\rho(\xi) = \rho$ is constant, X is said to be *homogeneous or first order stationary with intensity* ρ; otherwise X is said to be *inhomogeneous*. Clearly, stationarity of X under translations implies homogeneity of X.

In the homogeneous case, a natural unbiased estimator is $\hat{\rho} = X(W)/|W|$. This is in fact the maximum likelihood estimator if X is a homogeneous Poisson process.

In the inhomogeneous case, a non-parametric kernel estimator of the intensity function (assuming this exists) is

$$\hat{\rho}(\xi) = \sum_{\eta \in X_W} k_1(\xi - \eta)/c_{W,k_1}(\eta), \quad \eta \in W, \tag{4.4}$$

where k_1 is a kernel (density function) and $c_{W,k_1}(\eta) = \int_W k_1(\xi - \eta)\mathrm{d}\xi$ is an *edge correction factor* so that $\int_W \hat{\rho}(\xi)\mathrm{d}\xi$ is an unbiased estimator of $\mu(W)$ (Diggle 1985). The estimator is usually very sensitive to the choice of bandwidth, while the choice of kernel function is less important.

Alternative estimators in connection to improved ratio-type estimators are discussed in Section 4.5.2, cf. (4.10) and (4.12).

4.5.2 Second order characteristics

Pair correlation, K, and L-functions

The so-called *second order factorial moment measure* is given by

$$\alpha^{(2)}(A \times B) = \mathbb{E} \sum_{\xi,\eta \in X}^{\neq} 1[\xi \in A, \eta \in B] = \mathbb{E}\big[X(A)X(B)\big] - \mu(A \cap B) \tag{4.5}$$

for Borel sets $A, B \subseteq \mathbb{R}^d$, where $1[\cdot]$ denotes indicator function. When X is Poisson with intensity measure μ, combining (4.2) and (4.5), we obtain that $\alpha^{(2)}(A \times B) = \mu(A)\mu(B)$. If $\alpha^{(2)}$ has a density $\rho^{(2)}(\xi, \eta)$ with respect to the Lebesgue measure on $\mathbb{R}^d \times \mathbb{R}^d$, this is called the *second order product density*; intuitively, $\rho^{(2)}(\xi, \eta)\mathrm{d}\xi\mathrm{d}\eta$ is the probability for observing a point in each of the infinitesimally small balls with centres ξ, η and areas $\mathrm{d}\xi, \mathrm{d}\eta$, respectively.

A widely used summary statistic (in spatial statistics and particularly astronomy and astrophysics, see e.g. Peebles (1974) and Kerscher (2000)), is the *pair correlation function* given by

$$g(\xi, \eta) = \rho^{(2)}(\xi, \eta)/(\rho(\xi)\rho(\eta))$$

provided the terms on the right hand side exist. For a Poisson process, we have that $g = 1$. In general, $g(\xi, \eta) > 1$ indicates *attraction* or *clustering*, and $g(\xi, \eta) < 1$ *repulsion* or *regularity* for points at locations ξ, η; this may in turn be due to certain latent processes (Section 4.6) or interaction between the points (Section 4.7). It is often assumed that $g(\xi, \eta) = g(\xi - \eta)$ is translation invariant; this is e.g. implied by stationarity of X under translations. It is convenient if $g(\xi, \eta) = g(\|\xi - \eta\|)$ depends only on the distance $\|\xi - \eta\|$; this is the case if X is both stationary and isotropic.

In the stationary case of X with finite intensity $\rho > 0$, there is a close relationship between g and *Ripley's K-function* (Ripley 1977) defined by

$$K(r) = \mathbb{E} \sum_{\xi \in X_W, \eta \in X}^{\neq} 1[\|\xi - \eta\| \leq r]/(\rho^2|W|). \tag{4.6}$$

It is easily seen that this definition does not depend on W. Further $\rho K(r)$ has an interpretation as the mean number of further points within distance r from a "typical" point in X. If $g(\xi, \eta) = g(\xi - \eta)$ exists and is translation invariant, then

$$K(r) = \int_{\|\xi\| \leq r} g(\xi)\mathrm{d}\xi, \quad r \geq 0.$$

Especially, for a homogeneous Poisson process,

$$K(r) = \omega_d r^d, \quad r \geq 0$$

where $\omega_d = \pi^{d/2}/\Gamma(1 + d/2)$ is the volume of a unit ball. One often considers the *L-function* given by $L = (K/\omega_d)^{1/d}$ instead of K, as $L(r) = r$ is the identity if X is a homogeneous Poisson process, and since this transformation is variance stabilising when K is estimated by non-parametric methods (Besag 1977b).

The K-function can be modified to directional K-functions for the anisotropic case (Stoyan & Stoyan 1994). It can also be extended to the inhomogeneous case (Baddeley, Møller & Waagepetersen 2000), where $K_{\mathrm{inhom}}(r)$

is defined as in (4.6) with ρ^2 replaced by $\rho(\xi)\rho(\eta)$, provided that $K_{\mathrm{inhom}}(r)$ does not depend on W. For example, this assumption is satisfied if $g(\xi,\eta) = g(\xi-\eta)$ exists and is translation invariant, in which case $K_{\mathrm{inhom}}(r) = \int_{\|\xi\|\le r} g(\xi)\mathrm{d}\xi$. It is also satisfied if X is obtained by independent thinning of a stationary point process. As in the homogeneous case, we often use $L_{\mathrm{inhom}} = (K_{\mathrm{inhom}}/\omega_d)^{1/d}$ instead of K_{inhom}. Note that $L_{\mathrm{inhom}}(r) = r$ in the Poisson case.

The summary statistics g and K_{inhom} are invariant under independent thinning. This also holds for K when all thinning probabilities $p(\xi)$ are equal. Furthermore, after independent thinning of a stationary point process, K_{inhom} for the thinned process agrees with K for the original process. These invariance properties can be exploited for semi-parametric inference, cf. Baddeley et al. (2000).

The summary statistics considered so far describe the second order properties of a spatial point process. It should be noticed that very different point process models can share the same first and second order properties as discussed in Baddeley & Silverman (1984) and Baddeley et al. (2000).

Non-parametric estimation

Non-parametric estimation of summary statistics is discussed in Stoyan & Stoyan (1994), Stoyan & Stoyan (2000), Ohser & Mücklich (2000), and the references therein. Such estimators may take boundary effects into consideration as demonstrated below.

One commonly used estimator of $\rho^2 K(r)$ is

$$\widehat{\rho^2 K(r)} = 2 \sum_{\{\xi,\eta\}\subseteq X_W}^{\neq} 1[\|\xi-\eta\| \le r]/|W_\xi \cap W_\eta| \tag{4.7}$$

where $W_\xi = \{\xi+\eta : \eta \in W\}$ denotes the translate of W by ξ. Because of stationarity, the estimator is unbiased when

$$|W \cap W_\xi| > 0 \text{ for all } \xi \in \mathbb{R}^d \text{ with } \|\xi\| \le r. \tag{4.8}$$

For example, if W is rectangular, it is by (4.8) required that r is smaller than the smallest side in W. If the pair correlation of X exists, then the estimator is still unbiased if

$$|W \cap W_\xi| > 0 \text{ for Lebesgue almost all } \xi \in \mathbb{R}^d \text{ with } \|\xi\| \le r. \tag{4.9}$$

In Section 4.5.5 we compare the conditions (4.8) and (4.9) for the special design in the weed plant example.

The estimator (4.7) may be combined with an estimator of ρ^2 to obtain an estimator of $K(r)$, but the best choice of combination depends on the particular model of X (Stoyan & Stoyan 2000). For example, for a homogeneous Poisson process X, it is recommended to use

$$\widehat{\rho^2} = X(W)(X(W)-1)/|W|^2 \tag{4.10}$$

which is an unbiased estimator of ρ^2. Then the combined estimator of K is biased but it is said to be *ratio-unbiased*.

A similar situation is noticed in Baddeley et al. (2000) for the inhomogeneous case, where $K_{\text{inhom}}(r)$ is estimated by

$$\widehat{K}_{\text{inhom}}(r) = 2 \sum_{\{\xi,\eta\}\subseteq X_W}^{\neq} 1[\|\xi - \eta\| \leq r]/\{|W_\xi \cap W_\eta|\bar{\rho}(\xi)\bar{\rho}(\eta)\}, \qquad (4.11)$$

and instead of the estimator $\hat{\rho}$ in (4.4),

$$\bar{\rho}(\xi) = \sum_{\eta\in X_W\setminus\xi} k_1(\xi - \eta)/c_{W,k_1}(\eta), \quad \eta \in W, \qquad (4.12)$$

is used. From (4.11) we obtain an estimator of L_{inhom}.

Non-parametric kernel estimators of g may be derived along similar lines when $g(\xi,\eta) = g(\|\xi - \eta\|)$ depends only on the distance. In the planar case $d = 2$, if $\widehat{\rho(\xi)\rho(\eta)}$ is an estimate of $\rho(\xi)\rho(\eta)$, we may estimate $g(r)$, $r > 0$, in analogy with (4.7) and (4.11) by

$$\hat{g}(r) = 2 \sum_{\{\xi,\eta\}\subseteq X_W}^{\neq} k_2(r - \|\xi - \eta\|) \Big/ \left[\pi r |W_\xi \cap W_\eta| \widehat{\rho(\xi)\rho(\eta)}\right] \qquad (4.13)$$

where k_2 is a symmetric kernel. Like in (4.4) the choice of bandwidth is important, and the estimator may be unreliable at small distances r as discussed later in Example 4.5.4. Alternative estimators are discussed in Stoyan & Stoyan (1994), Stoyan & Stoyan (2000), and Ohser & Mücklich (2000).

Plots of these estimators are often supplied with *envelopes* obtained by simulation of a specified model, for example, an estimated Poisson model; several examples are shown in the sequel. Let $\hat{T}_0$ be a non-parametric estimator of a summary statistic T computed from the data X, and let $\hat{T}_1, \ldots, \hat{T}_n$ be the same non-parametric estimator computed from i.i.d. simulations $X_1, \ldots, X_n$ under the specified model. For example, we could have $\hat{T}_0 = \hat{K}(r)$ for a given distance $r > 0$. We have that with probability (at least) $(n-1)/(n+1)$,

$$\min_{1\leq i\leq n} \hat{T}_i \leq \hat{T}_0 \leq \max_{1\leq i\leq n} \hat{T}_i \qquad (4.14)$$

if X follows the specified model (strictly speaking this has to be a simple hypothesis — the probability may differ from $(n-1)/(n+1)$ if the "specified model" is actually an estimated model — but this is usually ignored in applications). We refer to the bounds in (4.14) as lower and upper envelopes. In our examples we choose $n = 39$ so that (4.14) specifies a 2.5% lower envelope and a 97.5% upper envelope (for each fixed value of r if e.g. $\hat{T}_0 = \hat{K}(r)$).

Higher order intensities

Finally, we remark that higher order summary statistics can be introduced as well, but the corresponding non-parametric estimators may be less stable if the number of points observed is not sufficiently large; see Peebles & Groth (1975), Stoyan & Stoyan (1994), Møller, Syversveen & Waagepetersen (1998), and Schladitz & Baddeley (2000).

4.5.3 Nearest-neighbour and empty space functions

Consider again the stationary case with finite intensity $\rho > 0$. The *empty space function* (or spherical contact distribution function) F is the distribution function of the distance from the origin (or another fixed point in $\mathbb{R}^d$) to the nearest point in X, i.e.

$$F(r) = P(\inf_{\xi \in X} \|\xi\| \le r).$$

The *nearest-neighbour function* is defined by

$$G(r) = \mathbb{E} \sum_{\xi \in X_W} 1[\inf_{\eta \in X \setminus \xi} \|\xi - \eta\| \le r]/(\rho |W|)$$

and has the interpretation as the distribution function of the distance from a typical point in X to its nearest point in X. This definition does not depend on the choice of W. It is not obvious how to extend the definitions of F and G to the inhomogeneous case. For a homogeneous Poisson process, by (4.1), $F(r) = G(r) = 1 - \exp(-\rho\omega_d r^d)$. For other kind of models, closed form expressions of F and G are rarely known.

Van Lieshout and Baddeley (1996) suggest to consider the combined summary statistic

$$J(r) = (1 - G(r))/(1 - F(r)) \quad \text{for } F(r) < 1,$$

which is 1 for a homogeneous Poisson process, whilst values less (more) than 1 may be an indication of clustering (regularity) in X.

Non-parametric estimators of F and G (and thereby J) are easily derived using *minus sampling*: For each $r > 0$, let $W_{\ominus r} = \{\xi \in W : b(\xi, r) \subseteq W\}$ denote the set of points in W with a distance to the boundary of W which is greater than r. If $I_r \subset W_{\ominus r}$ is a finite grid of n_r points (chosen independently of X), we have the unbiased estimator

$$\hat{F}(r) = \sum_{\xi \in I_r} 1[\inf_{\eta \in X} \|\xi - \eta\| \le r]/n_r \tag{4.15}$$

and the ratio-unbiased estimator

$$\hat{G}(r) = \sum_{\xi \in X_{W_{\ominus r}}} 1[\inf_{\eta \in X \setminus \xi} \|\xi - \eta\| \le r]/(\hat{\rho}|W_{\ominus r}|). \tag{4.16}$$

Finally, envelopes may be simulated in the same way as described above for the L-function.

Alternative Kaplan-Meyer type estimators are discussed in Baddeley & Gill (1997).

4.5.4 Example 2: Norwegian spruces (continued)

In this section we consider only the positions of the spruces.

A non-parametric kernel estimate (4.4) of the intensity surface for the spruce positions is shown in Figure 4.3. A Gaussian kernel k_1 is used where the bandwidth is chosen subjectively in order to get a suitable trade off between smoothness of the estimate and level of detail in the estimate. There is no obvious trend in the estimated intensity surface, so we assume that the point pattern is a partially observed realization of a stationary point process.

A non-parametric estimate of $L(r) - r$, $r > 0$, (see Section 4.5.2), the estimates $\hat{G}$ (4.16) and $\hat{F}$ (4.15), the estimate of J obtained from $\hat{G}$ and $\hat{F}$, and a non-parametric estimate of the pair correlation are also shown in Figure 4.3. The pair correlation function is estimated using (4.13) with $\widehat{\rho(\xi)\rho(\eta)}$ given by (4.10), and an *Epanecnikov kernel*

$$k_2(r) = 1[|r| < b]3(1 - (r/b)^2)/(4b)$$

with bandwidth $b = 2$. For each summary statistic, 2.5% and 97.5% envelopes are calculated from 39 simulations of the fitted homogeneous Poisson process with intensity $\hat{\rho} = 134/(56 \times 38)$.

There is clear indication of regularity since the estimate of $L(r) - r$ is smaller than 0 and below the lower envelope for r up to around 8. Similarly, $\hat{G}$ is smaller than expected for a Poisson process, and $\hat{F}(r)$ is above the upper envelope for $r > 2$. Also the estimated J-function and pair correlation function suggest that the point pattern is regular. Note from the envelopes that the pair correlation estimate appears to be biased upwards for $0 < r < 1$ under the fitted Poisson model. So it does not seem advisable to interpret the small kink of the estimated $g(r)$ occurring for $0 < r < 1$.

4.5.5 Example 1: Weed plants (continued)

Brix & Møller (2001) observe a log linear trend for the intensity of the weed plants perpendicular to the ploughing direction, so they consider a parametric log linear model for the intensity function,

$$\log \rho(\xi; \theta) = \theta_1 + \theta_2 \xi_2, \quad \xi = (\xi_1, \xi_2) \in W,\ \theta = (\theta_1, \theta_2) \in \mathbb{R}^2, \qquad (4.17)$$

where W is the union of the 45 observation frames. This is supported by the fact that the humidity of the field seemed to have a gradient in that

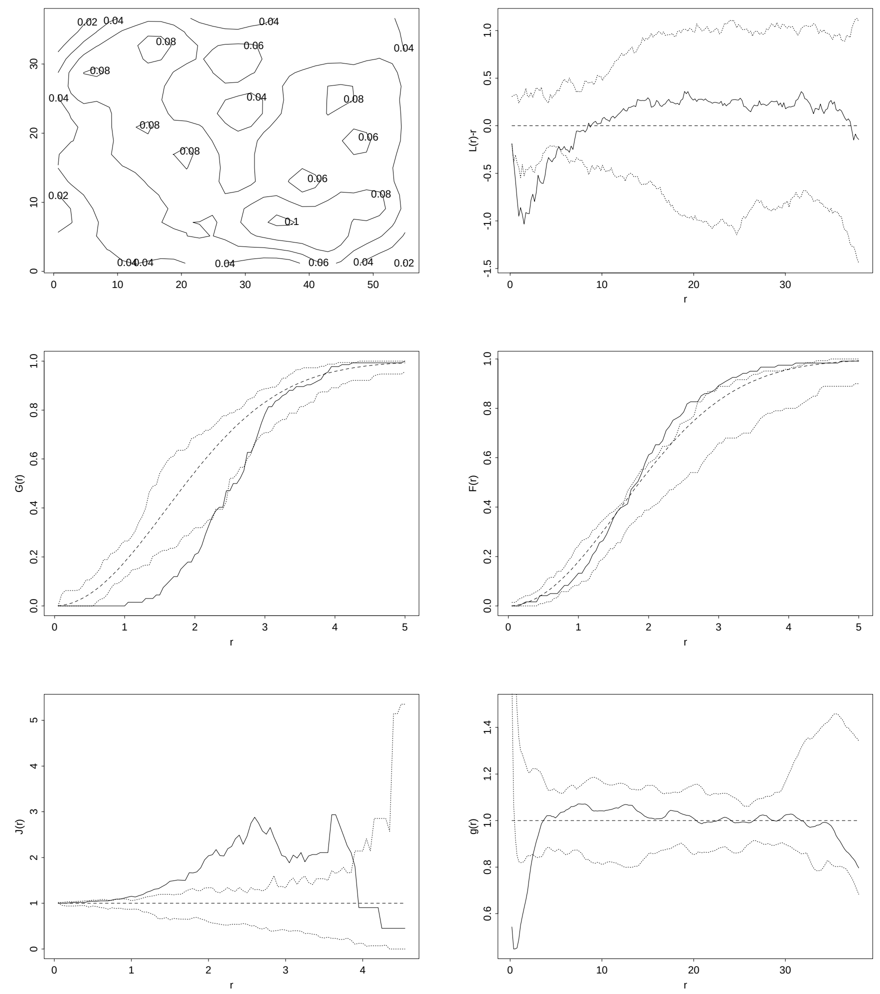

FIGURE 4.3. Summary statistics for positions of Norwegian spruces. Upper left: estimate of intensity surface. Upper right: estimated $L(r) - r$ (solid line) and envelopes calculated from 39 simulations under the fitted homogeneous Poisson process. Dashed line is the theoretical value of $L(r) - r$ for a Poisson process. Middle left, middle right, lower left, and lower right: as upper right but for G, F, J, and g.

direction; a fact that should stimulate the occurrence of *Trifolium* spp. In Brix & Møller (2001) the parameter θ is estimated in an ad hoc manner using a simple linear regression with dependent variables given by the weed counts in each frame.

Alternatively we here consider the log likelihood under the assumption that the weed plants form a Poisson process. By (4.3) the log likelihood is given by

$$\sum_{\eta \in x_W} \log \rho(\eta; \theta) - \log \int_W \rho(\xi; \theta) \mathrm{d}\xi \tag{4.18}$$

where x_W is the set of weed plant locations. As suggested in Berman & Turner (1992) and further investigated in Baddeley & Turner (2000), likelihoods of the form (4.18) can easily be maximised using standard software for generalised linear models, see also Section 4.7.2. Using this approach with the Splus routine `glm()` we obtain the maximum likelihood estimate $\hat{\theta} = (-4.10, 0.003)$ (which is close to the estimate $(-4.28, 0.003)$ obtained in Brix & Møller (2001)).

By replacing $\bar{\rho}(\cdot)$ in (4.11) with $\rho(\cdot; \hat{\theta})$ an estimate of K_{inhom} and thereby of L_{inhom} is obtained; denote this estimate by $\hat{L}_{\mathrm{inhom},\hat{\theta}}$. Figure 4.4 shows $\hat{L}_{\mathrm{inhom},\hat{\theta}}(r) - r$ which should be close to zero if the weed positions were Poisson. The dashed line shows the average of estimated $L_{\mathrm{inhom},\hat{\theta}}(r) - r$ functions calculated from 39 simulations under the fitted inhomogeneous Poisson process (i.e. when the intensity function is assumed to be known and given by $\rho(\cdot; \hat{\theta})$). It appears that $\hat{L}_{\mathrm{inhom},\hat{\theta}}$ is nearly unbiased under the fitted Poisson process. Furthermore, the upper/lower envelopes calculated from the 39 simulations of the fitted inhomogeneous Poisson process are rather constant when $r \leq 20$ cm, whilst they increase/decrease for larger values of r. This may be compared with the conditions (4.8) and (4.9) which require r to be less than 20 cm and $(30^2 + 60^2)^{1/2} \approx 67$ cm, respectively. Arguably, the envelopes may be too narrow since we are ignoring the variability of $\hat{\theta}$ when using the fitted Poisson model for the simulations.

A plot similar to that of the L-function but for the pair correlation function g is also shown in Figure 4.4. Here g is estimated by (4.13) with an Epanecnikov kernel with bandwidth 3 and $\hat{\rho}(\cdot)$ given by $\rho(\cdot; \hat{\theta})$. As in Section 4.5.4 the pair correlation function estimate is biased upwards for small distances under the fitted Poisson process.

By Figure 4.4, the weed data clearly exhibit clustering, since $\hat{L}_{\mathrm{inhom},\hat{\theta}}(r) - r$ takes positive values above the upper envelope when $r \leq 20$ cm. Similarly, the estimated $g(r)$ falls above the upper envelope for distances up to around 17. Note that the envelopes for $g(r)$ differ most for $20 < r < 40$ where few interpoint distances are observed due to the experimental design.

A close look at Figure 4.1 shows that the intensities of weed plants are higher in the third column of frames than in the first column of frames. This is not reflected by the log linear model (4.17), so an appropriate alternative might be a log third order polynomial model

$$\log \rho_3(\xi; \psi) = \psi_1 + \psi_2 \xi_2 + \psi_3 \xi_2^2 + \psi_4 \xi_2^3, \tag{4.19}$$

where $\xi = (\xi_1, \xi_2) \in W$ and $\psi = (\psi_1, \psi_2, \psi_3, \psi_4) \in \mathbb{R}^4$. The maximum likelihood estimate is

$$\hat{\psi} = (-3.670, -0.014, 9.342 \times 10^{-5}, -1.310 \times 10^{-7}).$$

Using $\rho_3(\cdot;\hat{\psi})$ in (4.11) we obtain an estimate $\hat{L}_{\mathrm{inhom},3,\hat{\psi}}(r) - r$. The plot (omitted) of $\hat{L}_{\mathrm{inhom},3,\hat{\psi}}(r) - r$ is qualitatively similar to the plot of the estimated $L(r) - r$ function in Figure 4.4: the clustering is less pronounced, but the Poisson model is still rejected.

As a more appropriate model, Brix & Møller (2001) consider a log Gaussian Cox process (see Section 4.6.2) where it is possible to model clustering due to environmental effects and the seed bank in the soil. In Section 4.6.3 we also consider a log Gaussian Cox process model for the weed plants.

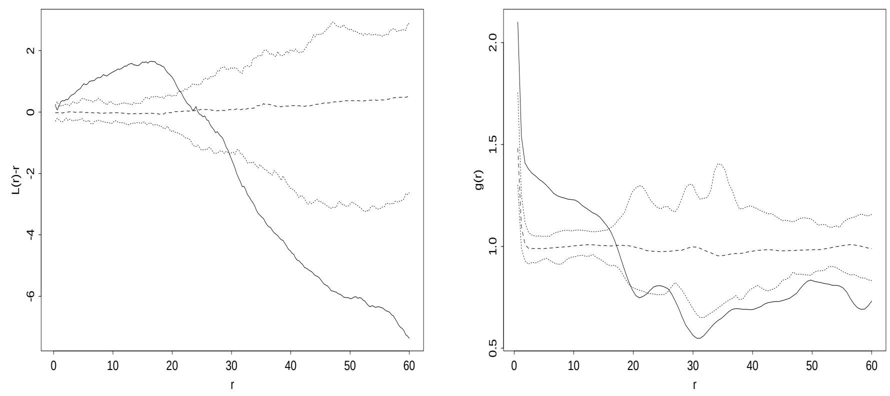

FIGURE 4.4. Left plot: solid line is $\hat{L}_{\mathrm{inhom},\hat{\theta}}(r) - r$ for weed plants; dashed horizontal line is average of $L_{\mathrm{inhom},\hat{\theta}}(r) - r$ functions computed from 39 simulations under the fitted Poisson model; dotted lines are 2.5% and 97.5% envelopes for $\hat{L}_{\mathrm{inhom},\hat{\theta}}(r) - r$ obtained from the 39 simulations. Right plot: as left plot, but for the pair correlation function.

4.6 Models and simulation-based inference for aggregated point patterns

Aggregation in a spatial point process may be caused by at least three factors: (i) spatial heterogeneity, e.g. due to some underlying "environmental" process, (ii) clustering of the points around the points of another point process, (iii) interaction between the points. In this section we concentrate mostly on (i) and partly on (ii), while (iii) is considered in Section 4.7.1. The relationship between (i) and (ii) is described in Section 4.6.1 which deals with Cox and cluster processes; further material on such processes can be found in Diggle (1983), Stoyan et al. (1995), and the references therein. In Section 4.6.2 we study the particular case of log Gaussian Cox processes. Other specific models such as shot noise Cox processes are discussed in Section 4.6.4.

4.6.1 Cox and cluster processes

Definitions and properties of Cox processes

A natural way to generalise the definition of a Poisson point process is to let Λ be a random, diffuse, and locally finite measure on S so that conditional on Λ, $X|\Lambda \sim \text{Poisson}(S, \Lambda)$. Thereby we obtain a *doubly stochastic Poisson point process* which is also called a *Cox process* with *driving measure* Λ. Specific constructions of Λ are considered in Sections 4.6.2 and 4.6.4, and in the following two simple examples.

The simplest non-trivial example of a Cox process is a *mixed Poisson process* in $\mathbb{R}^d$. This is obtained by letting $R > 0$ be a random variable and $X|R$ a homogeneous Poisson process in $\mathbb{R}^d$ with intensity R. For example, if R is gamma distributed, $X(A)$ follows a negative binomial distribution for $A \in \mathcal{B}_0$.

Another example is *random independent thinning of a Poisson process:* Suppose that conditional on a random field $\Pi = \{\Pi(\xi) : \xi \in S\}$ with $0 \le \Pi(\cdot) \le 1$, $X \sim \text{Poisson}(S, \mu)$ and $R(\xi) \sim \text{Uniform}[0, 1]$, $\xi \in S$, are mutually independent. Then $Z = \{\xi \in X : R(\xi) < \Pi(\xi)\}$ is a Cox process driven by the random measure given by $\Lambda(B) = \int_B \Pi(\xi)\mu(\mathrm{d}\xi)$ (provided this integral is well-defined for Borel sets $B \subseteq S$, and Λ is locally finite and diffuse).

By definition of a Cox process X and the properties of Poisson processes we obtain immediately the following general results. The void probabilities are given by

$$P(X(B) = 0) = \mathbb{E}\exp(-\Lambda(B)).$$

Furthermore, the intensity measure is given by

$$\mu(B) = \mathbb{E}X(B) = \mathbb{E}\Lambda(B),$$

and the second order factorial moment measure by

$$\alpha^{(2)}(A \times B) = \mathbb{E} \sum_{\xi,\eta \in X}^{\neq} 1[\xi \in A, \eta \in B] = \mathbb{E}[\Lambda(A)\Lambda(B)].$$

Combining these results, we see that

$$\mathbb{V}\text{ar}(X(B)) = \mathbb{V}\text{ar}(\Lambda(B)) + \mu(B).$$

Thus, compared to a Poisson process, a Cox process exhibits overdispersion.

In the particular case of $S = \mathbb{R}^d$, if Λ has a (random) density $\mathrm{d}\Lambda(\xi)/\mathrm{d}\xi = \lambda(\xi)$, then X has intensity function

$$\rho(\xi) = \mathbb{E}\lambda(\xi) \tag{4.20}$$

and second order product density

$$\rho^{(2)}(\xi, \eta) = \mathbb{E}[\lambda(\xi)\lambda(\eta)]. \tag{4.21}$$

Closed form expressions of these functions can sometimes be derived for specific models, cf. Sections 4.6.2 and 4.6.4.

Definitions and properties of cluster processes

Let Y be a point process defined on a space T, and for each $\xi \in Y$, let Z_ξ be a point process defined on a space S so that the superposition

$$X = \bigcup_{\kappa \in Y} Z_\kappa$$

is a simple and locally finite point process on S. Then X is called a *cluster process* with *mother process* Y and *clusters* or *daughters* Z_κ, $\kappa \in Y$. We may e.g. think of plants (the mothers) that spread seeds (the daughters). Usually in applications, $S = T \subseteq \mathbb{R}^d$ and the mother process is unobserved, so we are dealing with a missing data problem.

Certain cluster processes are special cases of Cox processes: If conditional on Y, the clusters are independent, each cluster $Z_\kappa \sim \text{Poisson}(S, \mu_\kappa)$, and the random measure given by

$$\Lambda(B) = \sum_{\kappa \in Y} \mu_\kappa(B), \quad B \subseteq S,$$

is locally finite and diffuse, then X is a Cox process driven by Λ. This follows simply by finding the void probabilities of X.

A particular important subclass of such models are *Neyman-Scott processes* X, where it is assumed that $S = \mathbb{R}^d$, each cluster Z_κ has an intensity function $\rho_\kappa(\xi)$, and $Y \sim \text{Poisson}(T, \mu)$ (slightly extending the definition of Neyman & Scott (1958)). Then Λ has density

$$\lambda(\xi) = \sum_{\kappa \in Y} \rho_\kappa(\xi)$$

with respect to Lebesgue measure, which combined with (4.20) and (4.21) give expressions for the intensity function and second order product density function of X,

$$\rho(\xi) = \mathbb{E} \sum_{\kappa \in Y} \rho_\kappa(\xi), \quad \rho^{(2)}(\xi, \eta) = \mathbb{E} \sum_{\kappa, \zeta \in Y} \rho_\kappa(\xi)\rho_\zeta(\eta).$$

Then by Slivnyak-Mecke (4.1)-(4.2),

$$\rho(\xi) = \int \rho_\kappa(\xi)\mathrm{d}\mu(\kappa)$$

and

$$\rho^{(2)}(\xi,\eta) = \int\int \rho_\kappa(\xi)\rho_\zeta(\eta)\mathrm{d}\mu(\kappa)\mathrm{d}\mu(\zeta) + \int \rho_\kappa(\xi)\rho_\kappa(\eta)\mathrm{d}\mu(\kappa).$$

These expressions can be further reduced when $T = \mathbb{R}^d$, Y is a homogeneous Poisson process, and $Z_\kappa - \kappa,\ \kappa \in Y$, (the clusters relative to their mother points) are i.i.d. and independent of Y; see e.g. Stoyan et al. (1995).

However, closed form expressions of $\rho^{(2)}$ and hence the pair correlation can only be derived for a few such Neyman-Scott processes, including a *Thomas process* X defined as follows: Y is a homogeneous Poisson process with intensity $\rho_Y > 0$, and $Z_\kappa - \kappa$ is a Poisson process where the number of points is po(α)-distributed and each point follows a d-dimensional normal distribution with mean 0 and covariance matrix $\sigma^2 I$. Then X is stationary and isotropic with intensity $\rho_X = \alpha\rho_Y$ and pair correlation function $g(\xi,\eta) = g(\|\xi-\eta\|)$ given by

$$g(r) = 1 + \exp(-r^2/(4\sigma^2))/[\lambda_Y(4\pi\sigma^2)^{d/2}], \quad r \geq 0.$$

Furthermore, in the planar case $d = 2$, Ripley's K-function is

$$K(r) = \pi r^2 + [1 - \exp(-r^2/(4\sigma^2))]/\lambda_Y.$$

Simulation procedures for Neyman-Scott processes on a bounded window B follow often straightforwardly from their definition as cluster processes or their construction as Cox processes: in either case, we first simulate Y and next $X_B|Y$. In order to avoid boundary effects the mother process usually must be simulated on an extended window $A \supset B$ so that daughter points from a mother outside A falls into B with a negligible probability; see Brix & Kendall (2002) and Møller (2002b).

4.6.2 Log Gaussian Cox processes

Definitions and properties

Suppose that $Y = \{Y(\xi) : \xi \in \mathbb{R}^d\}$ is a real-valued Gaussian process, i.e. any finite linear combination of the $Y(\xi)$ follows a normal distribution. If X is a Cox process on $\mathbb{R}^d$ driven by a random measure with density

$$\lambda(\xi) = \exp(Y(\xi))$$

with respect to Lebesgue measure, then X is said to be a *log Gaussian Cox process (LGCP)*. Such models have independently been introduced in astronomy by Coles & Jones (1991) and in statistics by Møller et al. (1998).

It is necessary to impose weak conditions on the mean function $m(\xi) = \mathbb{E}Y(\xi)$ and covariance function $c(\xi, \eta) = \mathbb{C}\mathrm{ov}(Y(\xi), Y(\eta))$ in order to get a well-defined and finite integral $\int_B \exp(Y(\xi))\mathrm{d}\xi$ for bounded Borel sets $B \subset \mathbb{R}^d$. For example, we may require that $\xi \to Y(\xi)$ is almost surely continuous. This is the case if m and c are continuous functions, and for some $0 < C < \infty$ and some $\epsilon > 0$,

$$c(\xi, \xi) + c(\eta, \eta) - 2c(\xi, \eta) \leq C/(-\log \|\xi - \eta\|)^{1+\epsilon} \tag{4.22}$$

whenever $\|\xi - \eta\| < 1$, cf. Theorem 3.4.1 in Adler (1981). These are fairly weak conditions which are usually satisfied for the models used in practice.

The definition of an LGCP can easily be extended in a natural way to multivariate LGCPs as shown in Møller et al. (1998) and to multivariate spatio-temporal LGCP as studied in Brix & Møller (2001). LGCPs are flexible models for clustering as demonstrated in Møller et al. (1998), where examples of covariance functions together with simulated realizations of LGCPs and their underlying Gaussian processes are shown. Certain Thomas processes may in practice be difficult to distinguish from LGCPs with a Gaussian covariance function $c(\xi, \eta) = \sigma^2 \exp(-(\|\xi - \eta\|/\alpha)^2)$, where $\sigma^2 > 0$ and $\alpha > 0$ are parameters (Møller et al. 1998).

The intensity function and pair correlation function of an LGCP are simply given by

$$\rho(\xi) = \exp(m(\xi) + c(\xi, \xi)/2), \quad g(\xi, \eta) = \exp(c(\xi, \eta)). \tag{4.23}$$

Hence, if $c(\xi, \eta) = c(\xi - \eta)$ is translation invariant, K_{inhom} exists. Further theoretical results for the intensities of an LGCP (including third and higher order properties) can be found in Møller et al. (1998). These results are in general of a different and much simpler form than for other Cox processes.

By (4.23) there is a one-to-one correspondence between c and g and between (m, c) and (ρ, g). Consequently, the distribution of an LGCP is uniquely determined by its first- and second-order properties as given by the intensity and pair correlation functions. This makes parametric models easy to interpret and simple methods for parameter estimation and model checking become available as discussed in Møller et al. (1998).

Finally, notice that there is no problem with edge effects as the distribution of an LGCP restricted to a bounded subset is known.

Simulation of LGCPs

Unconditional simulation: Below we describe shortly how to simulate an LGCP X and its underlying Gaussian process Y when both are restricted to a bounded region B. Since $X_B|Y$ is simply a Poisson process with intensity function $\exp(Y)$ on B, we restrict attention to simulation from $Y_B = \{Y(\xi) : \xi \in B\}$. As the infinitely dimensional process Y_B does not in general have a finite representation in a computer, we approximate Y_B by a random step function with constant value $Y(c_i)$ within disjoint cells C_i, where $B = \cup_{i \in I} C_i$, I is a finite index set, and $c_i \in C_i$ is a "centre" point of C_i. So we actually consider how to simulate the Gaussian vector $\tilde{Y} = (\tilde{Y}_i)_{i \in I}$ where $\tilde{Y}_i = Y(c_i)$.

Suppose for the moment that B is rectangular, say $B = [0,1[^2$, and let $I \subset B$ denote a regular grid. As discussed in Møller et al. (1998), there is an efficient way of simulating $\tilde{Y}$ when $c(\xi, \eta) = c(\xi - \eta)$ is invariant under translations. Briefly, I is embedded in a rectangular grid I_{ext}, which is wrapped on a torus, and a block circulant matrix $K = \{K_{ij}\}_{i,j \in I_{\text{ext}}}$ is constructed so that the submatrix $\{K_{ij}\}_{(i,j) \in I}$ is the covariance matrix of $\tilde{Y}$. Since K is block circulant, it can easily be diagonalised by means of the two-dimensional discrete Fourier transform with associated matrix F_2 (see Section 6.1 in Møller et al. (1998) and Wood & Chan (1994)). Suppose that K is positive semi-definite (i.e. K has non-negative eigenvalues). Then we can extend $\tilde{Y} = (\tilde{Y}_{(i,j)})_{(i,j) \in I}$ to a larger Gaussian field $\tilde{Y}_{\text{ext}} = (\tilde{Y}_{(i,j)})_{(i,j) \in I_{\text{ext}}}$ with covariance matrix K: set

$$\tilde{Y}_{\text{ext}} = \Gamma Q + \mu_{\text{ext}} \tag{4.24}$$

where Γ follows a standard multivariate normal distribution, the restriction of μ_{ext} to I agrees with the mean of $\tilde{Y}$, and $Q = \bar{F}_2 \tilde{\Lambda}^{1/2} F_2$, where $\tilde{\Lambda}$ is the diagonal matrix of eigenvalues for K, and $\bar{F}_2$ is the complex conjugate of F_2. Using the two-dimensional *fast Fourier transform* a fast simulation algorithm for $\tilde{Y}_{\text{ext}}$ and hence $\tilde{Y}$ is obtained.

Another possibility is to use the *Cholesky decomposition* of the covariance matrix of $\tilde{Y}$, provided this covariance matrix is positive definite. This may be advantageous if c is not translation invariant or B is far from being rectangular, see Section 4.6.3. On the other hand, the Cholesky decomposition is only practically applicable if the dimension of $\tilde{Y}$ is moderate. We can still refer to (4.24) when the Cholesky decomposition is used, letting now I_{ext}, $\tilde{Y}_{\text{ext}}$, K, and Q be specified as follows: $I_{\text{ext}} = I$, $\tilde{Y}_{\text{ext}} = \tilde{Y}$, K is the covariance matrix of $\tilde{Y}$, and Q denotes the upper-triangular matrix obtained from the Cholesky decomposition.

Conditional simulation: Now, suppose that we have observed a point pattern $X_W = x$ within a window $W \subseteq B$. When making conditional simulations of (Y_B, X_A) given X_W, we may simulate first from $Y_B | X_W$ and next from $X_A | (Y_B, X_W)$. The latter conditional distribution is just a Poisson process with intensity function $\exp(Y_A)$ on A, so we restrict attention to simulation of $Y_B | X_W$ below. Note that if we wish to make further conditional simulations within a region D which is disjoint to B, we may first make simulations from $Y_D | (Y_B, X_B)$, which is Gaussian and does not depend on X_B, and next from the Poisson process $X_D | Y_D$ with intensity function $\exp(Y_D)$ on D.

Approximate simulations of $Y_B|X_W = x$ can be obtained from simulations of $\tilde{Y}|X_W = x$ which in turn can be obtained from simulations of $\Gamma|X_W = x$ using the transformation (4.24). Omitting an additive constant depending on x only, the log conditional density of Γ given x is

$$-\|\gamma\|^2/2 + \sum_{i \in I} (\tilde{y}_i n_i - A_i \exp(\tilde{y}_i)) \tag{4.25}$$

where, in accordance with (4.24), $(\tilde{y}_i)_{i \in I_{\text{ext}}} = \gamma Q + \mu_{\text{ext}}$, $n_i = x(C_i)$, and $A_i = |C_i \cap W|$. Note that (4.25) is formally equivalent to the conditional density of the random effects given observations n_i, $i \in I$, in a generalised linear mixed model with Poisson error distribution, cf. Section 2.9.1. The gradient of (4.25) becomes

$$\nabla(\gamma) = -\gamma + \big(n_i - A_i \exp(\tilde{y}_i)\big)_{i \in I_{\text{ext}}} Q^{\mathsf{T}},$$

and differentiating once more the conditional density of Γ given x is seen to be strictly log-concave.

For simulation from $\Gamma|X_W = x$, Møller et al. (1998) use a *Langevin-Hastings algorithm* or *Metropolis-adjusted Langevin algorithm* as introduced in the statistical community by Besag (1994) (see also Roberts & Tweedie (1996) and Section 1.3.1) and earlier in the physics literature by Rossky, Doll & Friedman (1978). This is a Metropolis-Hastings algorithm with collective updating inspired by the definition of a Langevin diffusion. If γ is the current state generated by the Langevin-Hastings algorithm, then we first propose a new state generated from a multivariate normal distribution with mean $\gamma + (h/2)\nabla(\gamma)$ and covariance matrix hI_d, where I_d is the $d \times d$ identity matrix and $h > 0$ is a user-specified parameter. Secondly we accept or reject the proposal in accordance to the Hastings ratio. Theoretical results in Roberts & Rosenthal (1998) and Breyer & Roberts (2000) suggest that one should tune h to obtain acceptance rates around 0.57. The use of the gradient in the proposal distribution may lead to much better convergence properties when compared to the standard alternative of a random walk Metropolis algorithm, see Christensen, Møller & Waagepetersen (2001) and Christensen & Waagepetersen (2002). See also Sections 1.3.1 and 2.9.1.

A truncated version of the Langevin-Hastings algorithm is obtained by replacing the gradient $\nabla(\gamma)$ by

$$\nabla_{\text{trun}}(\gamma) = -\gamma + \big(n_i - \min\{H, A_i \exp(\tilde{y}_i)\}\big)_{i \in I_{\text{ext}}} Q^{\mathsf{T}} \tag{4.26}$$

where $H > 0$ is a user-specified parameter which can e.g. be taken to be twice the maximal n_i, $i \in I$. We still tune h so that the acceptance rate is about 0.57. As shown in Møller et al. (1998) the *truncated Langevin-Hastings algorithm* is geometrically ergodic.

The algorithms described above can easily be generalised to the case of multivariate LGCPs, cf. Møller et al. (1998).

Bayesian inference for LGCPs

Møller et al. (1998) consider an empirical Bayesian approach to *prediction* of Y_B given $X_W = x$, where a parametric model has been chosen for the mean and covariance functions and the parameters are estimated by a so-called minimum contrast method. This estimation method depends on certain user-specified parameters, and the uncertainty of the estimated parameters are not taken into account, so arguably the variation of $Y_B|X_W = x$ may be underestimated. In the sequel we consider an alternative *fully Bayesian approach* with hyper priors on the parameters and discuss how we can make MCMC simulations of the full posterior given $X_W = x$. Again the random step function approximation of Y_B is used, and we consider the posterior of Γ from which the posterior of $\tilde{Y}$ can be computed.

Specifically, assume that the mean function m restricted to B is a linear function $m(\xi) = \theta z(\xi)^\mathsf{T}$ of a p-dimensional covariate $z(\xi)$, and the covariance function is of the form $c(\xi, \eta) = \sigma^2 r(\|\xi - \eta\|/\alpha))$ where $\sigma > 0$ is the standard deviation of $Y(\xi)$ and $\alpha > 0$ is a scale parameter for the correlation. Such a situation is considered in Benes, Bodlak, Møller & Waagepetersen (2002) and in Section 4.6.3 below. We impose independent hyper priors p_1, p_2, and p_3 on θ, σ, and $\kappa = \log \alpha$, respectively. The posterior density is thus

$$\begin{aligned}\pi(\gamma, \theta, \kappa, \sigma|x) \propto\, & p_1(\theta)p_2(\sigma)p_3(\kappa) \\ & \times \exp\Big(- \|\gamma\|^2/2 + \sum_{i \in I} (\tilde{y}_i n_i - A_i \exp(\tilde{y}_i)) \Big), \qquad (4.27)\end{aligned}$$

cf. (4.25). Note that $\tilde{y}$ is a function of $(\gamma, \theta, \kappa, \sigma)$.

As in Christensen & Waagepetersen (2002), Christensen et al. (2001), and Benes et al. (2002) we use a hybrid algorithm (also called Metropolis-within-Gibbs, cf. Section 1.6.1) with a systematic MCMC updating scheme for the full conditional distribution of each of the four parameters $\gamma, \theta, \kappa, \sigma$. For γ we use a truncated Langevin-Hastings algorithm where the gradient by (4.27) is still given by (4.26). For θ we use also a truncated Langevin-Hastings algorithm with gradient

$$\nabla_{\text{trun}}(\theta) = \big(n_i - \min\{H, A_i \exp(\tilde{y}_i)\}\big)_{i \in I_{\text{ext}}} D + \frac{\partial}{\partial \theta} \log p_1(\theta)$$

where D is the design matrix with rows given by $z(c_i)$, $i \in I$ (as $n_i = A_i = 0$ for cells outside W, we need only to specify D for centre points contained in W). The same truncation constant H is used in the two Langevin-Hastings algorithms, and as before the variances in the proposal distributions are tuned so that acceptance rates about 0.57 are obtained. Finally, random walk Metropolis updates are used for κ and $\log \sigma$, respectively, with acceptance rates around the optimal value 0.23 (Roberts, Gelman & Gilks 1997). See also Section 1.5.3.

4.6.3 Example 1: Weed plants (continued)

For the weed data, B is the union of the 45 observation frames and we use a discretisation of the Gaussian field where each of the 45 frames is subdivided into 6 quadratic cells of side length 10 cm. The covariance matrix of the 270 dimensional vector $\tilde{Y}$ is decomposed using the Cholesky decomposition. The covariate vector is $z(\xi) = (z_1(\xi), z_2(\xi)) = (1, \xi_2)$ for $\xi = (\xi_1, \xi_2) \in B$, and we use the exponential correlation function $r(\|\xi\|) = \exp(-\|\xi\|)$. The hyper priors are chosen to be $p_1(\theta) \propto 1$, $\theta \in \mathbb{R}^2$, $p_2(\sigma) \propto \exp(-10^{-5}/\sigma)/\sigma$, $\sigma > 0$, and $p_3(\kappa) \propto 1$, $\log 0.75 < \kappa < \log 75$. The improper prior p_1 is completely flat, and the improper p_2 yields an essentially flat prior for $\log \sigma$ on $(0, \infty)$. The limits for the log uniform prior p_3 were chosen subjectively in order to accommodate a reasonable range of strengths of correlation. For the discretised LGCP one can check as in Christensen & Waagepetersen (2002) that these priors yield a proper posterior but strictly speaking we do not know whether a proper posterior is also obtained for the original LGCP. One may therefore consider the possibility of restricting the supports of θ and σ to large but bounded regions.

The hybrid algorithm described at the end of Section 4.6.2 is used for the computations. In order to improve mixing of the Markov chain we use a *reparameterisation* (see Section 1.3.2) where $(z_2(c_i))_{i \in I}$ is normalised to have zero mean and maximum absolute value equal to one. The posterior distributions shown in Figure 4.5 are computed from time series obtained by subsampling each 10th of 200,000 scans of the hybrid algorithm; here a scan means an update of each of γ, θ, σ, κ in a systematic order. Plots of the time series (omitted) suggest that equilibrium is attained after around 400 scans and according to estimated autocorrelations, the states in the time series are uncorrelated for lags greater than 30. The posterior means

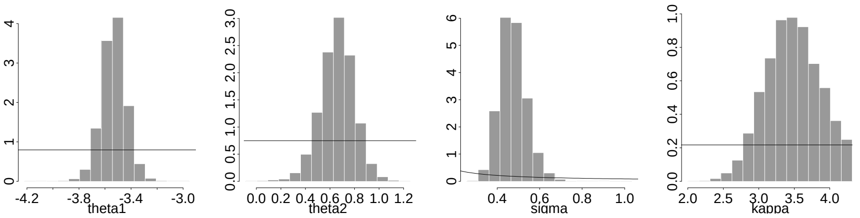

FIGURE 4.5. Posterior distributions of θ_1, θ_2, σ, and κ. The solid lines indicate the prior distributions.

of θ_1 and θ_2 are -4.11 and 0.003, i.e. very close to the maximum likelihood estimates for the inhomogeneous Poisson process considered in Section 4.5.5. The posterior means of σ and κ are 0.47 and 3.48. The value $\kappa = 3.48$ yields correlations 0.40 and 0.05 at distances 30 cm and 100 cm, respectively. Note that the posterior for κ is very sensitive to the choice of prior — one can in fact verify that the posterior support is equal to the prior support. The posterior mean of $\tilde{Y}$ is shown in Figure 4.6 where large posterior means of $\tilde{Y}_i$ coincide with cells C_i containing many weed plants.

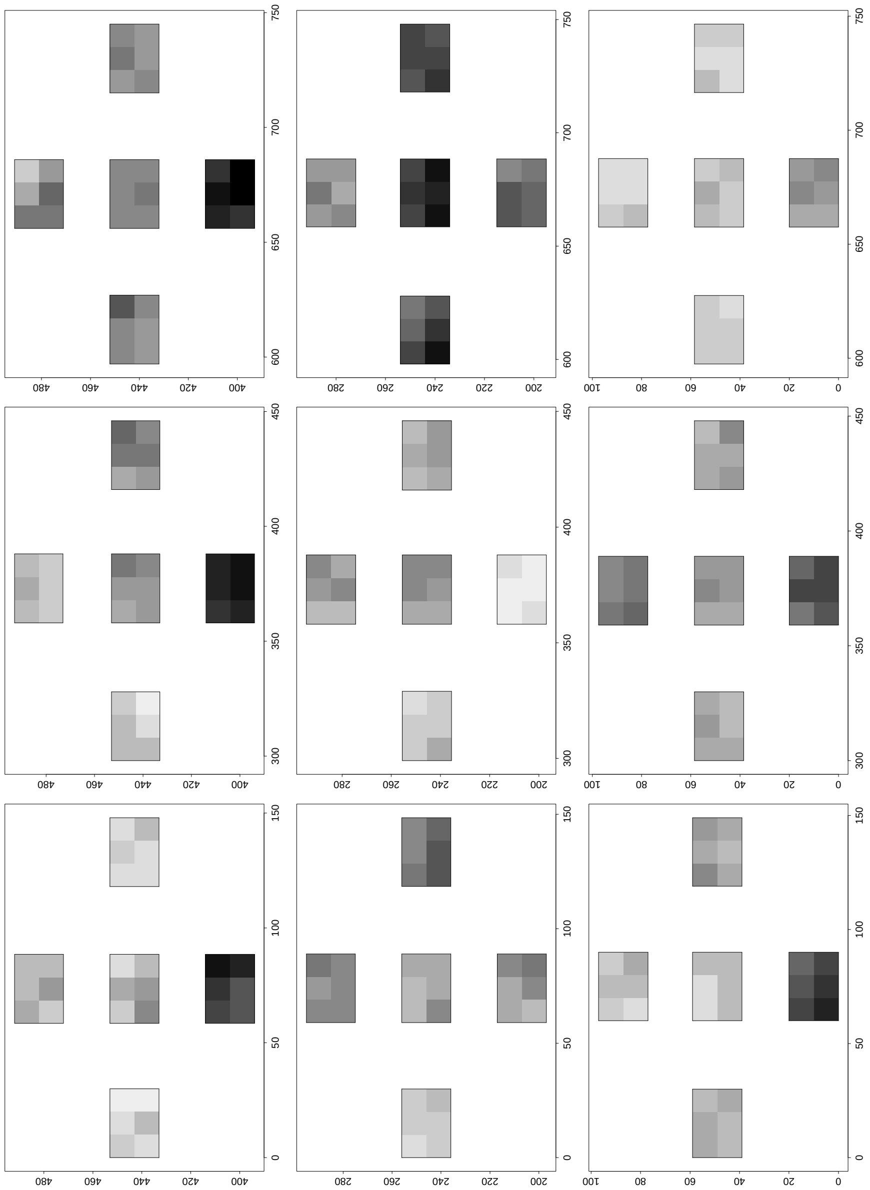

FIGURE 4.6. Posterior means of $\tilde{Y}$. Values range between -0.60 (light grey) and 0.97 (dark grey).

Figure 4.7 is similar to Figure 4.4. It shows $\hat{L}_{\text{inhom},\hat{\theta}}(r) - r$ and $\hat{g}(r)$ from Section 4.5.5 but now with envelopes calculated from simulations under the posterior predictive distribution (Gelfand 1996) for the LGCP. Each such simulation consists in first generating $(\theta, \sigma, \kappa, Y_B)$ from the posterior and next $X_W|(\theta, \sigma, \kappa, Y_B)$ as discussed in Section 4.6.2. For the first step, approximately independent posterior realizations of $(\theta, \sigma, \kappa, Y_B)$ are sampled from the MCMC sample. Clustering among the weed plants is ac-

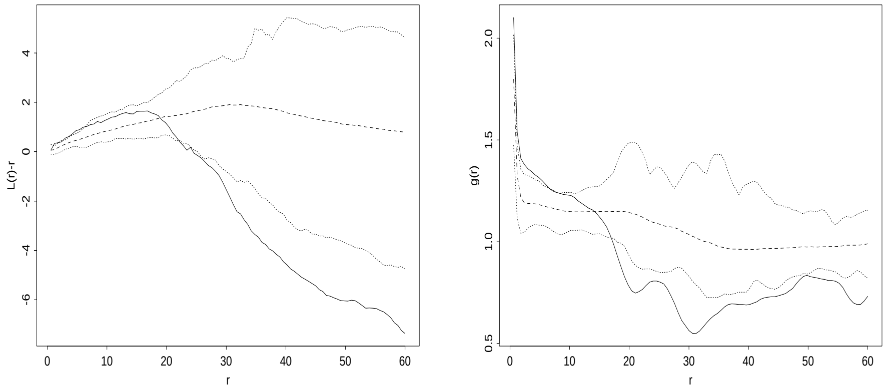

FIGURE 4.7. Left plot: solid line is $\hat{L}_{\text{inhom},\hat{\theta}}(r) - r$ for weed plants (see Section 4.5.5); dashed horizontal line is average of $L_{\text{inhom},\hat{\theta}}(r) - r$ functions computed from 39 simulations under the posterior predictive distribution; dotted lines are 2.5% and 97.5% envelopes for $\hat{L}_{\text{inhom},\hat{\theta}}(r) - r$ obtained from the 39 simulations. Right plot: as left plot, but for the pair correlation function.

commodated by the LGCP model but both summary statistics fall below the envelopes for distances greater than 20 cm. As noted in Section 4.5.5 the linear model for m may be too inflexible. Perhaps it is necessary to use a model for m which allows for rapid changes in the intensity as e.g. between the second and third row in Figure 4.1. In Brix & Møller (2001) only distances up to 20 cm were considered in the model checking.

4.6.4 Other specific models for Cox processes

So far we have concentrated much on LGCPs. Below we describe briefly two other interesting classes of Cox processes which can be used for non-parametric Bayesian modelling: the Heikkinen & Arjas (1998) model and *shot noise G Cox processes (SNGCP)* (Brix 1999). We introduce these models by specifying their random intensity function λ on a bounded region $R \subseteq \mathbb{R}^2$.

In Heikkinen & Arjas (1998) $\lambda(\xi) = \sum_k \lambda_k \mathbf{1}_{A_k}(\xi)$ where $\{A_k\}$ is the Voronoi tessellation generated by a point process of nuclei $\{y_k\} \subset R$, i.e. A_k is the set of points in R closer to y_k than to any other nuclei. The nuclei follow a homogeneous Poisson process restricted to R, and conditional on

$\{y_k\}$, $\{\log \lambda_k\}$ is modelled by a conditional autoregression (Besag 1974).

The construction of a SNGCP is a bit more complicated. Now, $\lambda(\xi) = \sum_j k(\xi, u_j)\gamma_j$ where k is a kernel (for simplicity we assume that $k(\cdot, u)$ is a density function for a continuous random variable), and $\{(u_j, \gamma_j)\} \subset E \times [0, \infty)$ where E is a given planar region (a more general setting is considered in Wolpert & Ickstadt, 1998). Typically in applications, $E = R$, or in order to reduce edge effects, $R \subset E$ where E is much larger than R. Further, $\{(u_j, \gamma_j)\}$ is a Poisson process with intensity measure

$$\nu(A \times B) = (\alpha(A)/\Gamma(1-\kappa)) \times \int_B \gamma^{-\kappa-1} \exp(-\tau\gamma)\mathrm{d}\gamma, \quad A \subseteq E,\ B \subseteq [0, \infty),$$

where $\kappa < 1$ and $\tau \geq 0$ are parameters with $\tau > 0$ if $\kappa \leq 0$, and α is a locally finite measure (or, as in Brix, 1999, a nonnegative and nonzero Radon measure). If $\kappa < 0$, we obtain a kind of modified Neyman-Scott process as $\{u_j\}$ is a Poisson process with intensity measure $(\tau/|\kappa|)\alpha$, and $\{u_j\}$ is independent of the "marks" $\{\gamma_j\}$, which in turn are mutually independent and follow a common Gamma distribution $\Gamma(|\kappa|, \tau)$ (in a usual Neyman-Scott process, all marks are equal and deterministic). The situation is less simple for $\kappa \geq 0$ as $\{u_j\}$ is not locally finite. For $\kappa = 0$, we have a Poisson/gamma model (Daley & Vere-Jones 1988, Wolpert & Ickstadt 1998). As noticed in Wolpert & Ickstadt (1998) we may extend the model by replacing the parameter τ with a positive function $\tau(u), u \in E$, and redefining

$$\nu(A \times B) = (1/\Gamma(1-\kappa)) \int_A \int_B \gamma^{-\kappa-1} \exp(-\tau(u)\gamma)\alpha(\mathrm{d}u)\mathrm{d}\gamma.$$

The models are reviewed and compared with LGCPs in Møller (2002a) and Møller & Waagepetersen (2002). A general class of shot noise Cox processes, including SNGCPs, is studied in Møller (2002b).

4.7 Models and simulation-based inference for Markov point processes

Markov or Gibbs point processes arose in statistical physics for the description of large interacting particle systems, see e.g. Ruelle (1969), Georgii (1976), Preston (1976), and Nguyen & Zessin (1979). Van Lieshout (2000) provides a recent account of the state of the art of Markov point processes in spatial statistics; see also the reviews in Ripley (1977), Baddeley & Møller (1989), and Stoyan et al. (1995). In Section 4.7.1 we concentrate on the case of a finite point process specified by a density with respect to a Poisson point process so that a local Markov property is satisfied. Pseudo likelihood estimation for Markov processes is considered in Sections 4.7.2–4.7.3, and maximum likelihood inference based on MCMC in Sections 4.7.4–4.7.5. Bayesian analysis is surveyed in Section 4.7.6. Simulation procedures for Markov point processes using Metropolis-Hastings algorithms and spatial birth-death processes are considered in Sections 4.7.7–4.7.8, while Section 4.7.9 concerns perfect simulation.

4.7.1 Definitions and properties

In the remaining part of this chapter we consider a finite point process X which is absolutely continuous with respect to Poisson(S,μ) where $\mu(S) < \infty$. Its density is denoted f. The point process may be extended to a larger region, and f may depend on points outside S, for example in order to take care of edge effects. However, we usually suppress such dependences and write $f(x)$ for $x \in N_\mathrm{f}$, where $N_\mathrm{f} = \{x \subseteq S : n(x) < \infty\}$ denotes the set of finite point configurations in S. We equip N_f with the σ-algebra $\mathcal{N}_\mathrm{f} = \{F \in \mathcal{N} : F \subseteq N_\mathrm{f}\}$. So by definition of a finite Poisson process, for events $F \in \mathcal{N}_\mathrm{f}$,

$$P(X \in F) = \sum_{n=0}^{\infty} \exp(-\mu(S))/n! \int \cdots \int 1[\{x_1,\ldots,x_n\} \in F] \\ f(\{x_1,\ldots,x_n\})\mu(\mathrm{d}x_1)\cdots\mu(\mathrm{d}x_n) \qquad (4.28)$$

where the term for $n = 0$ is read as $\exp(-\mu(S))1[\emptyset \in F]f(\emptyset)$. An example of such a density is given by the density (4.3) for a Poisson process, but in the following we shall construct much more interesting models exhibiting interactions between the points. Often f is assumed to be *hereditary*, that is,

$$f(x) > 0 \Rightarrow f(y) > 0 \quad \text{for } y \subset x. \qquad (4.29)$$

This amounts to the positivity condition in the Hammersley-Clifford theorem for Markov random fields, see e.g. Besag (1974) and Section 3.2.

In many applications we have a *pairwise interaction point process*,

$$f(x) \propto \prod_{\xi \in x} \phi(\xi) \prod_{\{\xi,\eta\} \subseteq x} \phi(\{\xi,\eta\}) \qquad (4.30)$$

where ϕ is a non-negative function for which the right hand side is integrable with respect to Poisson(S,μ). A standard example is the *Strauss process* (Strauss 1975), where

$$\phi(\xi) = \beta \text{ and } \phi(\{\xi,\eta\}) = \gamma^{1[d(\xi,\eta)\le R]}, \qquad (4.31)$$

setting $0^0 = 1$. Here $\beta > 0$, $0 \le \gamma \le 1$, and $R > 0$ are parameters (if $\gamma > 1$ we do not in general have integrability, cf. Kelly & Ripley, 1976). If $\gamma = 1$ we obtain $X \sim$ Poisson$(S,\beta\mu)$, while for $\gamma < 1$ there is repulsion between R-close pairs of points in X. The special case where $\gamma = 0$ is called a *hard core point process* with *hard core* R as balls of diameter R and with centres at the points in X are not allowed to overlap. The Strauss process can be extended to a *multiscale point process* (Penttinen 1984) where

$$\phi(\{\xi,\eta\}) = \gamma_i \quad \text{if } R_{i-1} < d(\xi,\eta) \le R_i \tag{4.32}$$

with $R_0 = 0 < R_1 < \ldots < R_k = \infty$, $\gamma_k = 1$, and $k \ge 2$; integrability is ensured if $\gamma_1 = 0$ and $\gamma_2 \ge 0, \ldots, \gamma_{k-1} \ge 0$, or if $0 < \gamma_1 \le 1$ and $0 \le \gamma_2 \le 1, \ldots, 0 \le \gamma_{k-1} \le 1$. A collection of other examples of pairwise interaction point processes can be found in van Lieshout (2000).

A fundamental characteristic is the *Papangelou conditional intensity* defined by

$$\lambda^*(x,\xi) = f(x \cup \xi)/f(x), \quad x \in N_\mathrm{f},\ \xi \in S \setminus x, \tag{4.33}$$

taking $a/0 = 0$ for $a \ge 0$ (Kallenberg 1984). For example, for the pairwise interaction point process (4.30),

$$\lambda^*(x,\xi) = \phi(\xi) \prod_{\eta \in x} \phi(\{\xi,\eta\}).$$

Heuristically, $\lambda^*(x,\xi)\mu(\mathrm{d}\xi)$ can be interpreted as the conditional probability of X having a point in an infinitesimal region containing ξ and of size $\mu(\mathrm{d}\xi)$ given the rest of X is x. If f is hereditary, then there is a one-to-one correspondence between f and λ^*. Often in applications $f \propto h$ is only specified up to proportionality, but $\lambda^*(x,\xi) = h(x \cup \xi)/h(x)$ does not depend on the normalising constant.

Integrability of a given function h with respect to Poisson(S,μ) may be implied by stability conditions in terms of λ^*. *Local stability* means that λ^* is uniformly bounded and f is hereditary; this implies integrability. A weaker condition for integrability is *Ruelle stability* (Ruelle 1969) meaning that $h(x) \le \alpha\beta^{n(x)}$ for some positive constants α, β and all $x \in N_\mathrm{f}$. As shown later in this section, local stability also plays an important role in simulation algorithms. Local stability is satisfied by many point process models (Geyer 1999, Kendall & Møller 2000). One example, where Ruelle but not local stability is satisfied, is a Lennard-Jones model (Ruelle 1969); this is a pairwise interaction point process (4.30) with $\phi(\xi) = \beta > 0$ constant and $\log \phi(\{\xi,\eta\}) = ar^6 - br^{12}$ for $r = ||\xi - \eta||$, where $a > 0$ and $b > 0$ are parameters.

The role of the Papangelou conditional intensity is similar to that of the local characteristics of a Markov random field when defining local Markov properties. Let $\sim$ be an arbitrary symmetric relation on S, let for the moment $f : N_\mathrm{f} \to [0,\infty)$ denote any function, and define λ^* as in (4.33). If f is hereditary and if for any $x \in N_\mathrm{f}$ and $\xi \in S \setminus x$, $\lambda^*(x,\xi)$ depends on x only through the *neighbours* $\eta \in x$ to ξ, i.e. those $\eta \in x$ with $\xi \sim \eta$, then f is said to be a *Markov function.* By the *Hammersley-Clifford-Ripley-Kelly theorem* (Ripley & Kelly 1977), f is Markov if and only if f is of the form

$$f(x) = \prod_{y \subseteq x} \phi(y), \quad x \in N_\mathrm{f}, \tag{4.34}$$

where ϕ is a so-called *interaction function*, i.e. a function $\phi : N_{\mathrm{f}} \to [0, \infty)$ with the property that $\phi(y) = 1$ if there are two distinct points in y which are not neighbours. If especially f is a density with respect to Poisson(S, μ), we have a *Markov density function* with normalising constant $\phi(\emptyset)$. Then $X \sim f$ is said to be a *Markov point process*. Combining (4.28) and (4.34) we easily obtain a *spatial Markov property*: for Borel sets $A, B \subset S$ so that no point in A is a neighbour to any point in B, X_A and X_B are conditionally independent given X_C where $C = S \setminus (A \cup B)$.

If f is hereditary and for all $x \in N_{\mathrm{f}}$, all $\xi \in S \backslash x$, and some finite $R > 0$ we have that $\lambda^*(x, \xi) = \lambda^*(x \cap b(\xi, R), \xi)$ where $b(\xi, R)$ denotes the closed ball with centre ξ and radius R, then f is said to be of *finite interaction range* R. Then f is obviously Markov with respect to the *finite range neighbour relation* given by $\xi \sim \eta$ if and only if $d(\xi, \eta) \leq R$. So Strauss and multiscale point processes are Markov. For the Norwegian spruce data in Figure 4.2, where a disc $b(\xi, m_\xi)$ specifies the influence zone of a tree located at ξ, it is natural to consider a Markov model with $\sim$ defined by

$$(\xi, m_\xi) \sim (\eta, m_\eta) \Leftrightarrow b(\xi, m_\xi) \cap b(\eta, m_\eta) \neq \emptyset, \tag{4.35}$$

i.e. trees are only allowed to interact when their influence zones overlap. In Section 4.7.5 we consider a Markov model for the spruces with respect to the relation (4.35).

Many other Markov models can be constructed by specifying different kinds of relations and interaction functions, using (4.34) and checking of course for integrability in each case. In fact pairwise interaction point processes are useful models for regularity/inhibition/repulsion but not so much for clustering/attraction. Models for both types of interactions may be constructed by allowing higher order interaction terms, see e.g. Baddeley & van Lieshout (1995), Geyer (1999), and Møller (1999). Moreover, the concept of a Markov function can be extended in different ways, whereby even more flexible models are obtained, see Baddeley & Møller (1989), van Lieshout (2000), and the references therein.

By allowing f to depend on points outside S, it is possible to extend Markov point processes to infinite Gibbs point processes, but questions like existence, uniqueness or not (phase transition behaviour), and stationarity may be hard to answer; we refer the interested reader to Preston (1976) and Georgii (1988) for the mathematical details.

The Papangelou conditional intensity will also play a key role in the sequel concerning statistical inference and simulation procedures for finite point processes.

4.7.2 Pseudo likelihood

Consider a parametric model $f_\theta \propto h_\theta$, $\theta \in \Theta$, for the density of a spatial point process X with respect to $\nu = $ Poisson(S, μ). In general, apart from the Poisson case, the normalising constant

$$
\begin{aligned}
Z_\theta &= \int h_\theta(x)\nu(\mathrm{d}x) && (4.36)\\
&= \sum_{n=0}^{\infty} \exp(-\mu(S))/n! \int \cdots \int h_\theta(\{x_1,\ldots,x_n\})\mu(\mathrm{d}x_1)\cdots\mu(\mathrm{d}x_n)
\end{aligned}
$$

cannot be evaluated explicitly. In order to avoid this problem, Besag (1977a) extended the definition of the pseudo likelihood function for Markov random fields (see Besag 1975 and Section 3.4.3) to the Strauss process by an approximation given by an auto-Poisson lattice process (Besag, Milne & Zachary 1982). Based on this derivation a general expression of the pseudo likelihood for point processes is stated in Ripley (1988). The pseudo likelihood for point processes is derived by a direct argument in Jensen & Møller (1991) as follows.

Suppose that each density f_θ is hereditary and Ruelle stable. Let $T \subseteq S$ be an arbitrary Borel set. For $x \in N_{\mathrm{f}}$, define the *pseudo likelihood* on T by

$$PL_T(\theta;x) = \exp(-\mu(T)) \lim_{i\to\infty} \prod_{j=1}^{m_i} f_\theta(x_{A_{ij}}|x_{S\setminus A_{ij}}),$$

where $\{A_{ij} : j = 1,\ldots,m_i\}$, $i = 1,2,\ldots$, are nested subdivisions of T such that $m_i \to \infty$ and $m_i[\max_{1\le j\le m_i} \mu(A_{ij})]^2 \to 0$ as $i \to \infty$. Further, $f_\theta(x_{A_{ij}}|x_{S\setminus A_{ij}})$ is a conditional density for $X_{A_{ij}}$ given that $X_{S\setminus A_{ij}} = x_{S\setminus A_{ij}}$:

$$f_\theta(x_{A_{ij}}|x_{S\setminus A_{ij}}) = \frac{f_\theta(x_{A_{ij}} \cup x_{S\setminus A_{ij}})}{\int f_\theta(y \cup x_{S\setminus A_{ij}})\mathrm{d}\mu_{A_{ij}}(y)}$$

if the denominator is strictly positive, and $f_\theta(x_{A_{ij}}|x_{S\setminus A_{ij}}) = 0$ otherwise; here $\mu_{A_{ij}}$ denotes the restriction of μ to A_{ij}. By Theorem 2.2 in Jensen & Møller (1991), for μ almost all $x \in N_{\mathrm{f}}$ the pseudo likelihood on T is well-defined and given by

$$PL_T(\theta;x) = \exp\left(-\int_T \lambda_\theta^*(x,\xi)\mathrm{d}\mu(\xi)\right) \prod_{\xi\in x_T} \lambda_\theta^*(x\setminus\{\xi\},\xi) \qquad (4.37)$$

where λ_θ^* is the Papangelou conditional intensity associated to f_θ. Note that the pseudo likelihood function is unaltered whether λ_θ^* is based on f_θ or on the conditional density $f_\theta(x_T|x_{S\setminus T})$. Usually in applications, either $T = S$ or $T = \{\xi \in S : b(\xi,R) \subseteq W\}$ in order to reduce edge effects when $X_W = x$ is observed within a window $W \subseteq S$ and all densities f_θ have finite interaction range R.

The *maximum pseudo likelihood estimate (MPLE)* is found by maximising (4.37). For certain models, consistency and asymptotic normality of the MPLE are established in Jensen & Møller (1991), Jensen & Künsch

(1994), Mase (1995), and Mase (1999). Intuitively, as the pseudo likelihood only depends on the local dependence structure, global information may better be taken into account when using the likelihood function. In fact the asymptotic variance of the MPLE can be much larger than for the MLE (the maximum likelihood estimator), in particular for spatial point processes with high dependence. Note that (4.37) agrees with the likelihood function for a Poisson process when $\lambda^*_\theta(x,\xi)$ depends only on ξ, so for point processes with weak interaction the MPLE and MLE may be expected to be close.

Assume now that the density f_θ belongs to an exponential family,

$$f_\theta(x) = b(x)\exp(\theta \cdot t(x))/Z_\theta, \quad x \in N_{\mathrm{f}},\ \theta \in \Theta, \tag{4.38}$$

where $\Theta = \{\theta \in \mathbb{R}^p : \int b(x)\exp(\theta \cdot t(x))\nu(\mathrm{d}x) < \infty\}$, $\cdot$ is the usual inner product, $b : N_{\mathrm{f}} \to [0,\infty)$ is hereditary, and $t : N_{\mathrm{f}} \to \mathbb{R}^p$. Then

$$\lambda^*_\theta(x,\xi) = b(x,\xi)\exp(\theta \cdot t(x,\xi)) \tag{4.39}$$

where $b(x,\xi) = b(x \cup \xi)/b(x)$ and $t(x,\xi) = t(x \cup \xi) - t(x)$. Proposition 2.3 in Jensen & Møller (1991) states that $PL_T(\theta;x)$ is log concave, and gives a condition for strict log concavity. Hence, if the MPLE exists and belongs to the interior of Θ, it is the solution to the pseudo likelihood equation $(\mathrm{d}/\mathrm{d}\theta)\log PL_T(\theta;x) = 0$, which is equivalent to

$$\int_T b(x,\xi)t(x,\xi)\exp(\theta \cdot t(x,\xi))\mathrm{d}\mu(\xi) = \sum_{\xi \in x_T} t(x \setminus \{\xi\},\xi). \tag{4.40}$$

Maximum pseudo likelihood estimation based on (4.37) is computationally equivalent to maximum likelihood estimation in an inhomogeneous Poisson process. In practice the integrals in (4.37) and (4.40) may be approximated by numerical methods. As noticed in Berman & Turner (1992) and Baddeley & Turner (2000), standard software such as Splus for fitting generalised linear models can be used to provide an approximate MPLE as follows.

Consider again the case (4.39), partition T into cells C_i, and let $c_i \in C_i$ denote a given "centre point". Let u_j, $j = 1,\dots,m$, denote a list of these centre points and the points in x_T. Then the integral in (4.37) is approximated by

$$\int_T \lambda^*_\theta(x,\xi)\mathrm{d}\mu(\xi) \approx \sum_{j=1}^m \lambda^*_\theta(x \setminus u_j, u_j)w_j, \tag{4.41}$$

where $w_j = \mu(C_i)/(1 + x(C_i))$ if $u_j \in C_i$. Note that the approximation (4.41) involves a "discontinuity error", since for $u_j \in x$, $\lambda^*_\theta(x \setminus u_j, u_j)$ is in general not equal to the limit of $\lambda^*_\theta(x,\xi)$ as $\xi \to u_j$, cf. the discussion in

Baddeley & Turner (2000). The advantage of including x_T in the sum in (4.41) is that we obtain

$$\log PL_T(\theta; x) \approx \sum_{j=1}^{m} (y_j \log \lambda_j^* - \lambda_j^*) w_j, \tag{4.42}$$

where $y_j = 1[u_j \in x]/w_j$ and $\lambda_j^* = \lambda_\theta^*(x \setminus u_j, u_j)$. The right side of (4.42) is formally equivalent to the log likelihood of independent Poisson variables y_j with means λ_j^* taken with weights w_j. If $b(x, u_j) > 0$, $j = 1, \ldots, m$, then (4.42) can easily be maximised using standard software for generalised linear models (taking $\log b(x, u_j)$ as an offset term). Moreover, if $b(\cdot,\cdot) = 1$, then we have a log linear model

$$\log \lambda_j^* = \theta \cdot t(x \setminus u_j, u_j).$$

4.7.3 Example 2: Norwegian spruces (continued)

For the point pattern of spruce locations we consider a multiscale process (4.32) with $k = 5$ and $R_i = 1.1 \times i, 0 < \gamma_i \leq 1, i = 1, \ldots, 4$. The minimal interaction range is thus less or equal to 4.4, a value suggested by the estimated summary statistics in Figure 4.3. We use the approximation (4.41) with $T = [0, 56] \times [0, 38]$ partitioned into 56×38 quadratic cells C_i, $i = (k, l) \in \{0, \ldots, 55\} \times \{0, \ldots, 37\}$, each of unit area and with centre points $c_{(k,l)} = ((k + 0.5)/56, (l + 0.5)/38)$. We have a log linear model with $\theta = (\log\beta, \log\gamma_1, \ldots, \log\gamma_4) \in \mathbb{R} \times (-\infty, 0]^4$ and $t(x, u_j) = (1, s[j, 1], \ldots, s[j, 4])$, where $s[j, i]$ denotes the number of points $\xi \in x \setminus u_j$ with $R_{i-1} < \|\xi - u_j\| \leq R_i$, $i = 1, \ldots, 4$. Using the Splus routine `glm()` with the call

```
glm(y~s[,1]+...+s[,4],family=poisson(link=log),weights=w),
```

and with y, w, and s constructed as above, the estimates -0.84, -3.35, -1.38, -0.62, -0.15 for $\log\beta$, $\log\gamma_1$, …, $\log\gamma_4$ are obtained (for a comparison with maximum likelihood estimates, see Section 4.7.5).

A biologically more interesting model is obtained by treating the spruce data as a marked point pattern where the stem diameters are used in the modelling. This approach is considered in Goulard et al. (1996) who discuss pseudo likelihood inference for marked point processes with the spruce data as one of the examples. We further discuss a marked point process approach in Section 4.7.5.

4.7.4 Likelihood inference

Consider again a parametric model of densities $f_\theta \propto h_\theta$, $\theta \in \Theta$, with respect to Poisson(S, μ), and where a closed expression for the normalising constant Z_θ given by (4.36) is not avaliable. In this section we discuss how to find the *maximum likelihood estimator (MLE)* $\hat{\theta}$ and the *likelihood ratio statistic* for hypothesis testing using MCMC methods. For simplicity we assume that the support $\{x : h_\theta(x) > 0\}$ does not depend on $\theta \in \Theta$. Furthermore, $\mathbb{E}_\theta$ denotes expectation with respect to $X \sim f_\theta$.

Assume that a realization $X = x$ is observed. In the exponential family case (4.38), $\hat{\theta}$ is the solution to the likelihood equation $\mathbb{E}_\theta t(X) = t(x)$. This suggest to approximate $\mathbb{E}_\theta t(X)$ by Monte Carlo methods, e.g. combined with Newton-Raphson (Penttinen 1984) or the EM-algorithm or stochastic approximation/gradient methods. Geyer (1999) advocates the use of other methods based on *importance sampling* as described below; see Geyer & Thompson (1992), Geyer & Møller (1994), Gu & Zhu (2001), and the references therein; see also Sections 1.5.4 and 3.4. Suppose that $\psi \in \Theta$ is an initial guess of $\hat{\theta}$, e.g. the MPLE. Then

$$Z_\theta / Z_\psi = \mathbb{E}_\psi \left[h_\theta(X) / h_\psi(X) \right] \tag{4.43}$$

can be estimated by a sample $X_1, X_2, \ldots$ from an irreducible Markov chain with invariant density f_ψ, see Section 4.7.7. Hence the logarithm of the likelihood ratio

$$f_\theta(x)/f_\psi(x) = \big(h_\theta(x)/h_\psi(x)\big) / \big(Z_\theta / Z_\psi\big)$$

is approximated by

$$l_n(\theta) = \log \left[\frac{h_\theta(x)}{h_\psi(x)} \right] - \log \left[\frac{1}{n} \sum_{i=1}^{n} \frac{h_\theta(X_i)}{h_\psi(X_i)} \right]. \tag{4.44}$$

For fixed ψ we can consider (4.44) as an approximation of the log *likelihood function* from which we may obtain an *approximate MLE* $\hat{\theta}_n$. Defining the importance weights

$$w_{\theta,\psi,n}(x) = \frac{h_\theta(x)/h_\psi(x)}{\sum_{i=1}^{n} h_\theta(X_i)/h_\psi(X_i)}$$

and for any function $k : N_{\mathrm{f}} \to \mathbb{R}^p$,

$$\mathbb{E}_{\theta,\psi,n} k(X) = \sum_{i=1}^{n} k(X_i) w_{\theta,\psi,n}(X_i),$$

we obtain approximate score functions, etc. by replacing exact expectations by Monte Carlo expectations. For example, in the exponential family case (4.38) the *score function* is approximated by

$$\nabla l_n(\theta) = t(x) - \mathbb{E}_{\theta,\psi,n} t(X), \tag{4.45}$$

and the *Fisher information* by

$$-\nabla^2 l_n(\theta) = \mathbb{V}\mathrm{ar}_{\theta,\psi,n} t(X), \tag{4.46}$$

and $l_n(\theta)$ is concave so that Newton-Raphson is feasible. Note that $\hat{\theta}_n$ is a function of both x and $X_1, \dots, X_n$. Asymptotic normality of the Monte Carlo error $\sqrt{n}(\hat{\theta}_n - \hat{\theta})$ as $n \to \infty$ is established in Geyer (1994).

The approximations (4.44)–(4.46) are only useful for θ sufficiently close to ψ. When ψ is not close to $\hat{\theta}$, Geyer & Thompson (1992) propose to use an iterative procedure with a "trust region", but one should be particular careful if the likelihood function is multi-modal.

A natural requirement is that $l_n(\theta)$ has finite variance. This is the case if the chain is geometrically ergodic (Section 1.4.3) and $\mathbb{E}_\psi |h_\theta(X)/h_\psi(X)|^{2+\epsilon} < \infty$ for some $\epsilon > 0$ (Theorem 1 in Chan & Geyer (1994)); or if just $\mathbb{E}_\psi (h_\theta(X)/\ h_\psi(X))^2 < \infty$ provided the chain is reversible (Corollary 3 in Roberts & Rosenthal (1997)). For example, for the Strauss process (4.31), if $\theta = (\beta, \gamma, R)$ and $\psi = (\beta', \gamma', R')$ with $\beta, \beta', R, R' > 0$ and $\gamma, \gamma' \in [0, 1]$, then $\mathbb{E}_\psi |h_\theta(X)/h_\psi(X)|^2 < \infty$ if and only if $\gamma \le \sqrt{\gamma'}$.

In order to estimate Z_θ / Z_ψ and hence the log likelihood ratios when θ and ψ are far apart umbrella sampling and the method of reverse logistic regression have been proposed, see Geyer (1991) and Geyer (1999). It is however our experience that these methods are often numerically unstable due to large variances for ratios of unnormalised densities. We turn therefore now to another technique called *path sampling*. The advantages of using this approach over the importance sampling approach is discussed in Gelman & Meng (1998).

Briefly, path sampling works as follows (see also Section 1.7.2). Suppose that $\Theta \subseteq \mathbb{R}^p$, $\theta \to \log h_\theta(X)$ is differentiable for $\theta \in \Theta$, and $(\mathrm{d}/\mathrm{d}\theta) h_\theta(X)$ is locally dominated integrable along a continuous differentiable path $\theta(s) \in \Theta,\ 0 \le s \le 1$ where $\psi = \theta(0)$ and $\theta = \theta(1)$. Letting

$$V_\theta(X) = (\mathrm{d}/\mathrm{d}\theta) \log h_\theta(X) \quad \text{and} \quad \theta'(s) = \mathrm{d}\theta(s)/\mathrm{d}s,$$

the identity

$$\log\left(Z_\theta / Z_\psi\right) = \int_0^1 \mathbb{E}_{\theta(s)} V_{\theta(s)}(X) \theta'(s)^\top \mathrm{d}s \tag{4.47}$$

is straightforwardly derived. Note that for many exponential family models (4.38), Monte Carlo estimation is more stable for $\mathbb{E}_{\theta(s)} V_{\theta(s)}(X) = \mathbb{E}_{\theta(s)} t(X)$ in (4.47) than for $\mathbb{E}_\psi [h_\theta(X)/h_\psi(X)] = \mathbb{E}_\psi \exp((\theta - \psi) \cdot t(X))$ in (4.43). The right hand side in (4.47) can be approximated by a Riemann sum using a discrete grid of points $\theta(s_i),\ i = 1, \dots, m$, generating independent Markov chains $X_t^i,\ i = 1, \dots, m$, with invariant densities $f_{\theta(s_i)}$, and estimating $\mathbb{E}_{\theta(s_i)} V_{\theta(s_i)}(X)$ by Monte Carlo; Berthelsen & Møller (2001b) combine this with independent runs of the chains, starting with a perfect simulation (see Section 4.7.9) for each chain. Alternatively, a Markov chain (X_t, S_t) defined on $E \times [0, 1]$ may be used; see Gelman & Meng (1998) for details.

Often in applications we can choose ψ or θ so that Z_ψ or Z_θ is known. For example, for the Strauss process (4.31), when (β, R) is fixed and $\Theta = \{\gamma \in (0, 1]\}$, we may choose $\theta = 1$ so that Z_θ is the normalising constant of a Poisson process. Then by (4.47), for $0 < \psi < 1$,

$$\log\big(Z_1/Z_\psi\big) = \int_\psi^1 \mathbb{E}_\gamma \sum_{\{\xi,\eta\}\subseteq x} 1[d(\xi,\eta) \le R]/\gamma \,\mathrm{d}\gamma. \tag{4.48}$$

Thereby an estimate of $\log Z_\psi$ is obtained. Repeating this for different values of (β, R), the entire likelihood surface, the likelihood ratio statistic for a specified hypothesis (e.g. that $\gamma = 1$), etc., can be approximated. For details, see Berthelsen & Møller (2001b) who also determine the distribution of the approximate likelihood ratio statistic by making further perfect and independent simulations.

Similar methods apply for *missing data situations.* Suppose that only $X_W = x$ is observed within a window $W \subseteq S$. Let $V = S\setminus W$, $Y = X_W$, and $Z = X_V$ (which is unobserved). Recall that if $X \sim \nu = \text{Poisson}(S,\mu)$, then $Y \sim \nu_W = \text{Poisson}(W,\mu_W)$ and $Z \sim \nu_V = \text{Poisson}(V,\mu_V)$ are independent, where μ_A denotes the restriction of μ to A. So if $X \sim f_\theta$, then Y has density

$$f_{\theta,W}(x) = \int f_\theta(x \cup z)\nu_V(\mathrm{d}z)$$

with respect to ν_W. Note that $Z_\theta(x) = \int h_\theta(x \cup z)\nu_V(\mathrm{d}z)$ is the normalising constant of the conditional density $f_{\theta,V}(z|x) \propto h_\theta(x \cup z)$ with respect to ν_V. Consequently, the logarithm of the likelihood ratio

$$f_{\theta,W}(x)/f_{\psi,W}(x) = \big[Z_\theta(x)/Z_\psi(x)\big]/\big[Z_\theta/Z_\psi\big]$$

can be approximated by

$$l_{n,W}(\theta) = \log\left[\frac{1}{n}\sum_{i=1}^n \frac{h_\theta(x \cup Z_i)}{h_\psi(x \cup Z_i)}\right] - \log\left[\frac{1}{n}\sum_{i=1}^n \frac{h_\theta(X_i)}{h_\psi(X_i)}\right] \tag{4.49}$$

where $Z_1, Z_2, \ldots$ is a sample from an irreducible Markov chain with invariant density $f_{\theta,V}(z|x)$, and $X_1, X_2, \ldots$ is a chain as in (4.44). Alternatively, path sampling can be used for estimating each of the terms $\log\big[Z_\theta/Z_\psi\big]$ and $\log\big[Z_\theta(x)/Z_\psi(x)\big]$.

4.7.5 Example 2: Norwegian spruces (continued)

For the spruce data we consider two different models: first a multiscale model as in Section 4.7.3 where we ignore the stem diameters and, second, a biologically more realistic model where overlap of the influence zones (see Section 4.2.2 and Figure 4.2) of the trees is penalised.

For the multiscale process the likelihood is maximised using Newton-Raphson with the score function and Fisher information approximated by (4.45) and (4.46), respectively. We apply a trust region procedure where ψ and θ initially are taken equal to the maximum likelihood estimate under the Poisson model (i.e. with $\log \gamma_i = 0$, $i = 1, \ldots, 4$). When the output $\tilde{\theta}$ of a Newton-Raphson iteration falls outside the trust region $\prod_{i=1}^{5}[\psi_i - 0.05, \psi_i + 0.05]$ new approximations (4.45) and (4.46) are calculated with the previous ψ value replaced by $\tilde{\theta}$. The Newton-Raphson procedure converges to the estimate $\hat{\theta} = (-0.38, -3.69, -1.49, -0.71, -0.30)$ for which the repulsion is stronger than for the pseudo likelihood estimate obtained in Section 4.7.3. We also maximise the likelihood under the null hypothesis $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma$, that is for the Strauss process with $R = 4.4$, and obtain the estimate $(-1.27, -0.44)$ for $(\log \beta, \log \gamma)$. The left plot in Figure 4.8 shows the interaction functions corresponding to $\hat{\theta}$ and the estimate under the null hypothesis.

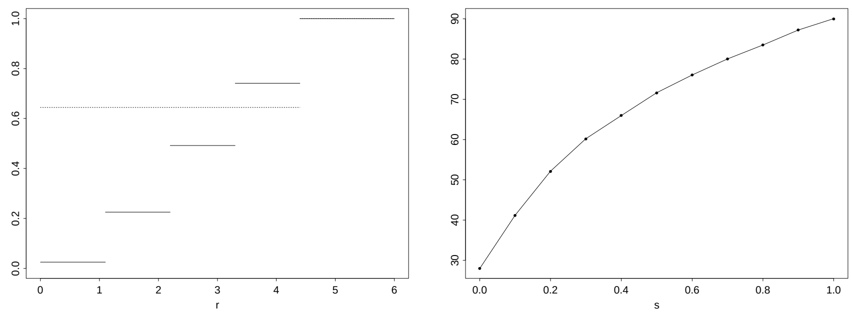

FIGURE 4.8. Left: estimated multiscale interaction function $\phi(\{\xi, \eta\})$ (see (4.32)) plotted as a function of distance $r = \|\xi - \eta\|$; solid line is for $\hat{\theta}$ and dotted is for the estimate under the null hypothesis (the Strauss model). Right: Monte Carlo estimates of $E_{\theta(k/10)} V_{\theta(k/10)}(X)\theta'(k/10)^{\mathsf{T}}$, $k = 0, \ldots, 10$ and curve corresponding to trapezoidal approximation.

Using path sampling we finally compute the log likelihood ratio statistic for the null hypothesis. Letting $\theta_0 = (-1.27, -0.44, -0.44, -0.44, -0.44)$ and $\theta_1 = \hat{\theta}$, we use the path $\theta(s) = \theta_0 + (\theta_1 - \theta_0)s$. The integral (4.47) is approximated as follows: for each of 11 quadrature points $\theta(k/10)$, $k = 0, \ldots, 10$, the integrand values $\mathbb{E}_{\theta(k/10)} V_{\theta(k/10)}(X)\theta'(k/10)^{\mathsf{T}}$ are replaced by Monte Carlo estimates and the integral with respect to s is finally approximated using the trapezoidal rule. The right plot in Figure 4.8 shows the Monte Carlo estimates of $\mathbb{E}_{\theta(k/10)} V_{\theta(k/10)}(X)\theta'(k/10)^{\mathsf{T}}$ together with the trapezoidal approximation. The log ratio $\log Z(\theta_1)/Z(\theta_0)$ is equal to 68 and the value of -2 times the log likelihood ratio statistic is 44 which is highly significant according to standard asymptotic results for the likelihood ratio statistic. Alternatively one may consider a parametric bootstrap where the observed log likelihood ratio statistic is compared with the distribution of the log likelihood ratio statistic under the fitted Strauss model. Specifically we compute -2 times the log likelihood ratio statistic from 99 simulations under the fitted Strauss model (proceeding exactly as for the observed data), obtaining values between 0.06 and 15.66, so the bootstrap also provides strong evidence against the Strauss model.

Neither the multiscale nor the Strauss model are very satisfactory from a biological point of view. A more realistic model can be obtained if the influence zones are included in the modelling so that low probability is assigned to point patterns with large area of overlaps between the influence zones. Let m_{lo} and m_{up} denote the minimal and maximal observed radii for the influence zones. For marked points (x_1, m_1) and (x_2, m_2) in $S = [0,56] \times [0,38] \times [m_{\mathrm{lo}}, m_{\mathrm{up}}]$ we define interaction functions ϕ_1 and ϕ_2 by

$$\phi_1((x_i, m_i)) = \beta_k \quad \text{if } k(m_{\mathrm{up}} - m_{\mathrm{lo}})/6 < m_i - m_{\mathrm{lo}} \leq (k+1)(m_{\mathrm{up}} - m_{\mathrm{lo}})/6$$

where $\beta_k > 0$ and $k = 0, \ldots, 5$, and

$$\log \phi_2(\{(x_1, m_1), (x_2, m_2)\}) = |b(x_1, m_1) \cap b(x_2, m_2)| \log \gamma$$

where $0 < \gamma \leq 1$. The function ϕ_1 allows modelling of the intensities of points with different values of the marks and ϕ_2 models the degree of repulsion in the point pattern. The marked point process is finally given by the pairwise interaction density

$$f(\{(x_1, m_1), \ldots, (x_n, m_n)\}) \propto \prod_{i=1}^{n} \phi_1((x_i, m_i)) \prod_{i<j} \phi_2(\{(x_i, m_i), (x_j, m_j)\})$$

with respect to the standard Poisson point process on $[0,56] \times [0,38] \times [m_{\mathrm{lo}}, m_{\mathrm{up}}]$. Proceeding as for the multiscale process using Newton-Raphson and path sampling we compute maximum likelihood estimates and the likelihood ratio statistic for the null hypothesis $\beta_0 = \beta_1 = \cdots = \beta_5 = \beta$. The maximum likelihood estimates are $(-1.23, -0.34, 0.53, -0.40, -0.80, -0.67, -1.10)$ for $(\log \beta_0, \ldots, \log \beta_5, \log \gamma)$ in the full model and $(-0.33, -1.07)$ for $(\log \beta, \log \gamma)$ under the reduced model. The log likelihood ratio statistic is -25 which is highly significant according to both standard asymptotics and a parametric bootstrap.

We conclude by giving some computational details. The samples used in the Newton-Raphson optimisation for Monte Carlo estimation of the score function and Fisher information were of length 5000 and obtained by subsampling each 200th state of a Metropolis-Hastings chain generated as described in Section 4.7.7. The Monte Carlo estimates of the integrands $\mathbb{E}_{\theta(k/10)} V_{\theta(k/10)}(X) \theta'(k/10)^{\mathsf{T}}$ in the path sampling procedure were computed from samples of length 1000 also obtained by subsampling each 200th state of a Metropolis-Hastings chain. The influence of the sample lengths and subsampling intervals on the Monte Carlo error of the various Monte Carlo estimates is a subject that deserves further study. The simulations for the parametric bootstrap were obtained by subsampling each 1000th state of a Metropolis-Hastings chain to obtain approximately uncorrelated simulations. A more appropriate approach would be to generate independent samples using perfect simulation, see Section 4.7.9.

4.7.6 Bayesian inference

Below we shortly comment on some of the rather few Bayesian contributions for Markov point processes which so far have been published.

Suppose we are extending the situation considered at the beginning of the previous section to a Bayesian setting with a prior on θ. The posterior distribution for $\theta|X = x$ is complicated by the fact that the normalising constant Z_θ in the likelihood term is usually unknown. Heikkinen & Penttinen (1999) suggest a Bayesian smoothing technique for estimation in pairwise interaction processes, where the likelihood function is approximated by the multiscale point process (4.32) having a large number of fixed change points $R_1, \ldots, R_{k-1}$. For convenience, they condition on the observed number $n(x)$ of points. A Gaussian Markov chain prior for $\theta = (\log \gamma_1, \ldots, \log \gamma_{k-1})$ is chosen so that large differences $|\log \gamma_i - \log \gamma_{i-1}|$ are penalised. As the full posterior analysis is considered to be too demanding, they concentrate on finding the posterior mode, using ideas from MCMC MLE as given in Penttinen (1984) and Geyer & Thompson (1992). Berthelsen & Møller (2001a) consider a similar situation, without conditioning on $n(x)$ but imposing a prior on k and $R_1, \ldots, R_{k-1}$, and finding the normalising constant of the likelihood term by path sampling so that a full Bayesian analysis is possible.

Lund, Penttinen & Rudemo (1999) consider a situation where an unobserved point process X is degraded by independent thinning, random displacement, a simple censoring mechanism, and independent superpositioning with a Poisson process of "ghost points"; this is related to aerial photographs of trees disturbed by the image analysis process, cf. Lund & Rudemo (2000). A known pairwise interaction prior on X is imposed in Lund et al. (1999), so its normalising constant is unimportant when dealing with the posterior distribution for X. Perfect simulation for this posterior is discussed in Lund & Thönnes (2000) and Møller (2001).

Finally, we mention in passing that Bayesian cluster models, using a locally stable prior density for the mother points, are studied in Baddeley & van Lieshout (1993), van Lieshout (2000), and Loizeaux & McKeague (2001). Here the parameters for the prior density for the mother points are assumed to be known, and local stability of the posterior density may be established. In particular, Loizeaux & McKeague (2001) discuss perfect simulation for the posterior distribution, and applies this to data on cases of leukaemia.

4.7.7 Metropolis-Hastings algorithms

In this and the following sections we concentrate on how to make simulations from a finite spatial point process X with a density f with respect to Poisson(S, μ) where $0 < \mu(S) < \infty$. Conditional simulation given that the number of points $n(X) = n$ is fixed may be done by any standard algorithm for updating n components $(X_1, \ldots, X_n)$ with density $f_n(x_1, \ldots, x_n) \propto$

$f(\{x_1, \ldots, x_n\})$ with respect to the product measure $\mu \times \cdots \times \mu$ (n times), e.g. by the classical *Metropolis algorithm* (Metropolis, Rosenbluth, Rosenbluth, Teller & Teller 1953) or by *Gibbs sampling* (Ripley 1979); see the survey in Møller (1999). Incidentally the purpose in Metropolis et al. (1953) was to simulate a hard core point process, and also the Gibbs sampler, which is now widely used in statistics, was first introduced in spatial statistics (and in statistical physics) in connection to random fields and spatial point processes.

We follow Geyer & Møller (1994) and define a *Metropolis-Hastings* chain $X_0, X_1, \ldots$ as follows. For technical reasons, assume that f is hereditary, cf. (4.29). Define the state space E of the chain as the set of finite point configurations which are feasible with respect to f, i.e. $E = \{x \in N_{\mathrm{f}} : f(x) > 0\}$. Let $\bar{\mu} = \mu/\mu(S)$ denote the normalisation of the measure μ. Now, for $X_{i-1} = x \in E$, with probability $1/2$ we propose to add a point $\xi \sim \bar{\mu}$ to x, and else we generate a uniformly selected point $\eta \in x$ and propose to delete η from x (if $x = \emptyset$ we set $\eta = \emptyset$). In the former case we return $X_i = x \cup \xi$ with probability $\min\{1, r(x, \xi)\}$ where $r(x, \xi) = \lambda^*(x, \xi)\mu(S)/(n(x)+1)$, and we retain $X_i = x$ otherwise. In the latter case we return $X_i = x \setminus \eta$ with probability $\min\{1, 1/r(x \setminus \eta, \eta)\}$ (setting this to 1 if $x = \emptyset$), and we retain $X_i = x$ otherwise.

The algorithm provides a simple example of Peter Green's reversible jump MCMC algorithm (Green 1995, Waagepetersen & Sorensen 2001). Its theoretical properties are studied in Geyer & Møller (1994), Geyer (1999), and Møller (1999). The chain is straightforwardly seen to be reversible with respect to f. Since f is hereditary, the state $\emptyset$ can be reached with probability 1 within a finite number of steps from any other state $x \in E$. It thereby follows that the chain is irreducible, and that f specifies the unique invariant distribution. As it can stay in $\emptyset$ with a positive probability, it is aperiodic. Assuming local stability, geometrical ergodicity of the chain can be established, and it becomes uniformly ergodic if and only if $f(x) = 0$ whenever $n(x)$ is sufficiently large, cf. the above-mentioned references.

The algorithm can obviously be modified by using other kinds of proposals for the addition or deletion of a point, and by incorporating the possibility of making a “fixed dimension move” as mentioned at the beginning of this section, cf. Geyer & Møller (1994). Such modifications may improve the mixing properties of the chain, but one should keep in mind that extra programming will be needed and the CPU time for each transition will usually be increased. The algorithm may also be combined with auxiliary variable techniques. For example, it is combined with *simulated tempering* in Mase, Møller, Stoyan, Waagepetersen & Döge (2001) in order to make simulations of hard core point processes with a high density of points feasible.

4.7.8 Simulations based on spatial birth-death processes

Let the situation be as at the beginning of the previous section. Preston (1977) notices that under suitable conditions, (approximate) realizations of $X \sim f$ may be obtained by running a *spatial birth-death process* $Y = \{Y_t : t \geq 0\}$ for a sufficient long time (briefly, Y is a continuous time Markov chain where each transition consists in either adding a new point or deleting an existing point). In this section we consider a *coupling construction* for the simplest case, which becomes useful for making simulations; this construction is also used in Section 4.7.9 for making perfect simulations.

Assume again that local stability is satisfied and let $K \geq \lambda^*$ denote an upper bound on the Papangelou conditional intensity. We start by describing how a spatial birth-death process $D = \{D_t : t \geq 0\}$ with equilibrium distribution Poisson$(S, K\mu)$ can easily be generated. This is next used to generate the abovementioned process Y by a thinning procedure so that $D_t \supseteq Y_t$ for all $t \geq 0$, provided $D_0 \supseteq Y_0$; we say that D *dominates* Y.

Generation of D: Given an initial state $D_0 \in N_{\mathrm{f}}$, it is straightforward to generate successive transition times and new states for D as follows. Suppose that we have generated D_s for $0 \leq s \leq t$, and $D_t = x = \{x_1, \ldots, x_n\}$. Let $\tau_0, \tau_1, \ldots, \tau_n$ be independent and exponentially distributed with means $1/(K\mu(S)), 1, \ldots, 1$, respectively. Then the waiting time until the next transition in D is given by $\tau = \min\{\tau_0, \tau_1, \ldots, \tau_n\} \sim \mathrm{Exp}(K\mu(S) + n)$. If $\tau = \tau_0$, then we have a birth: generate a point $\xi \sim \bar{\mu} = \mu/\mu(S)$ and set $D_{t+\tau} = D_t \cup \xi$. If instead $\tau = \tau_i$ with $1 \leq i \leq n$, then we have a death: set $D_{t+\tau} = D_t \setminus x_i$.

It can be shown that no matter which initial state $D_0 \in N_{\mathrm{f}}$ is used, D converges towards its equilibrium distribution given by $\nu = \mathrm{Poisson}(S, K\mu)$: for any $F \in \mathcal{N}_f$, $P(D_t \in F|D_0) \to \nu(F)$ as $t \to \infty$. Moreover, D regenerates each time $D_t = \emptyset$ and $D_{t-} \neq \emptyset$ (where $D_{t-} = \lim_{s \uparrow t} D_s$), and with probability 1, this happens infinitely often.

Generation of Y: Given initial states $Y_0 \in E$ and $D_0 \in N_{\mathrm{f}}$ with $Y_0 \subseteq D_0$, and letting the transition times of Y be included in the transition times of D, it is straightforward to generate successive transition times and new states for Y by thinning from D as follows. Suppose that we have generated $Y_s \subseteq D_s$ for $0 \leq s \leq t$, where every $Y_s \in E$, and $D_t = x = \{x_1, \ldots, x_n\}$ and $Y_t = y \subseteq x$. Using a notation as above, in the case of a birth $D_{t+\tau} = D_t \cup \xi$, we let $Y_{t+\tau} = Y_t \cup \xi$ with probability $\lambda^*(y, \xi)/K$, and retain $Y_{t+\tau} = Y_t$ otherwise. In the case of a death $D_{t+\tau} = D_t \setminus x_i$, we set $Y_{t+\tau} = Y_t \setminus x_i$; so Y is unchanged at time $t + \tau$ if $x_i \notin Y_t$.

It can be shown that for all initial states $(D_0, Y_0) \in N_{\mathrm{f}} \times E$ with $D_0 \supseteq Y_0$, Y converges towards its equilibrium density f as $t \to \infty$. Further, Y regenerates each time $Y_t = \emptyset$ and $Y_{t-} \neq \emptyset$, and with probability 1, this happens infinitely often. Hence by the renewal theorem, for any measurable function $k : E \to [0, \infty)$,

$$\frac{1}{t}\int_0^t k(Y_s)\mathrm{d}s \to \mathbb{E}k(X) \quad \text{as } t \to \infty \tag{4.50}$$

almost surely. Finally, D and Y are each reversible, but (D, Y) is in general not reversible. See Preston (1977) and Møller (1989) for further details, and Berthelsen & Møller (2002) for extensions to more general cases of spatial birth-death processes and other kinds of spatial jump processes.

If Y is generated on a finite time interval $[0, t]$, (4.50) may be used for estimating expectations. However, Metropolis-Hastings simulations as described in Section 4.7.7 seem more popular in practice, possibly due to their simplicity. Spatial birth-death processes have advantages for perfect simulation as will be demonstrated in the following section.

4.7.9 Perfect simulation

Since the seminal paper by Propp & Wilson (1996), *perfect or exact simulation* has been an intensive area of research. The term "perfect simulation" rather than "exact simulation" has been introduced in Kendall (1998) to emphasise that the output of the algorithms are only exact up to deficiencies in the pseudo random number generator applied in the computer implementation of the algorithm, and since very long runs may be omitted due to time constraints, which possibly is causing a bias in the output, cf. the discussion in Kendall & Møller (2000).

Perfect simulation techniques seem particular applicable for many spatial point process models, see e.g. Kendall (1998), Häggström, van Lieshout & Møller (1999), Thönnes (1999), Kendall & Møller (2000), Berthelsen & Møller (2001a); see also the surveys Møller (2001), Berthelsen & Møller (2001b), and the references therein. In this section we follow Kendall (1998) and Kendall & Møller (2000) and show how the coupling construction of the spatial birth-death processes Y and D introduced in Section 4.7.8 can be used for making perfect simulations from a locally stable density f.

Recall that $\{D_t : t \geq 0\}$ is reversible with invariant distribution $\nu =$ Poisson$(S, K\mu)$. Hence we can easily start in equilibrium $D_0 \sim \nu$, and extend the process backwards in time to obtain $\{D_t : t \leq 0\}$, using the same procedure as for forwards simulations of D. Let $\ldots, T_{-2} < T_{-1} < T_1 < T_2 < \ldots$ denote the times where $D_t = \emptyset$ and $D_{t-} \neq \emptyset$, such that $T_{-1} \leq 0 < T_1$. As D regenerates at these time instances, the cycles of D

$$\ldots,\ \{D_t : T_{-2} \leq t < T_{-1}\},\ \{D_t : T_{-1} \leq t < T_1\},\ \{D_t : T_1 \leq t < T_2\}, \ldots$$

are i.i.d. Imagine that we generate Y within each cycle of D, setting first $Y_{T_i} = \emptyset$ for $i \in \mathbb{Z} \setminus \{0\}$ as D dominates Y, and then using the forwards thinning procedure described in Section 4.7.8. Then $\{(D_t, Y_t) : -\infty < t <$

$\infty\}$ is a continuous time stationary process. Consequently, for any fixed time t, we have that $Y_t \sim f$.

This means that a perfect simulation $Y_0 \sim f$ can be obtained by first simulating $\{D_t : 0 \geq t \geq T_{-1}\}$ backwards in time, starting in equilibrium at time 0, and then generate $\{Y_t : T_{-1} \leq t \leq 0\}$ forwards in time by the thinning procedure. For this we actually only need to generate the jump chain of $\{D_t : 0 \geq t \geq T_{-1}\}$, i.e. the states D_t where a backwards transition occurs, since this contains the jump chain of $\{Y_t : T_{-1} \leq t \leq 0\}$. However, T_{-1} can be infeasible large, cf. Berthelsen & Møller (2001b), so alternative algorithms as described below are used in practice.

One possibility is to use *upper and lower processes* defined as follows. Let again $D_0 \sim \nu$, denote $Z_{-1}, Z_{-2}, \dots$ the jump chain of $\{D_t : t < 0\}$ when considered backwards in time, and denote $\dots, W_{-2}, W_{-1}$ the states of $\{Y_t : t < 0\}$ at the times where the jumps of $\{D_t : t < 0\}$ occur when considered forwards in time. Note that $D_0 = Z_{-1}$, $Y_0 = W_{-1}$, and the jump chain of $\{Y_t : t < 0\}$ agrees with the jumps of $\dots, W_{-2}, W_{-1}$. For $n = 1, 2, 3 \dots$, we define below the upper process $U^n = \{U_t^n : t = -n, \dots, -1\}$ and the lower process $L^n = \{L_t^n : t = -n, \dots, -1\}$ so that

$$L_t^n = U_t^n \;\Rightarrow\; L_s^n = U_s^n \quad \text{for } s = t, \dots, -1, \tag{4.51}$$

and

$$L_t^n \subseteq W_t \subseteq U_t^n \subseteq Z_t. \tag{4.52}$$

The *coalescence property* (4.51) and the *sandwiching property* (4.52) imply that if $L_t^n = U_t^n$ for some $-n \leq t \leq 0$, then by induction $L_s^n = W_s = U_s^n$ for $s = t, \dots, -1$, and so $U_{-1}^n = W_{-1} = Y_0 \sim f$. Hence,

$$L_t^n = U_t^n \text{ for some } -n \leq t \leq 0 \;\Rightarrow\; U_{-1}^n \sim f.$$

We now consider the coupling construction for Z and W, and thereby realize how to extend this to a coupling construction for upper and lower processes satisfying (4.51) and (4.52). Let $R_{-1}, R_{-2}, \dots$ be independent and uniformly distributed on $[0,1]$, and independent of $Z_{-1}, Z_{-2}, \dots$. For each n, since we only know that $\emptyset \subseteq W_{-n} \subseteq Z_{-n}$, we set first $U_{-n}^n = Z_{-n}$ and $L_{-n}^n = \emptyset$. Then we iterate as follows for $t = -n+1, \dots, -1$: If a death happens so that $Z_t = Z_{t-1} \setminus \eta$, then $W_t = W_{t-1} \setminus \eta$, and so we set $U_t^n = U_{t-1}^n \setminus \eta$ and $L_t^n = L_{t-1}^n \setminus \eta$. If instead a birth $Z_t = Z_{t-1} \cup \xi$ happens, then $W_t = W_{t-1} \cup \xi$ if $R_t < \lambda^*(W_{t-1}, \xi)/K$, while $W_t = W_{t-1}$ is unchanged otherwise; so we set

$$U_t^n = U_{t-1}^n \cup \xi \quad \text{if } R_t < \max\{\lambda^*(x,\xi)/K : L_{t-1}^n \subseteq x \subseteq U_{t-1}^n\} \tag{4.53}$$

and $U_t^n = U_{t-1}^n$ otherwise, and set

$$L_t^n = L_{t-1}^n \cup \xi \quad \text{if } R_t < \min\{\lambda^*(x,\xi)/K : L_{t-1}^n \subseteq x \subseteq U_{t-1}^n\} \tag{4.54}$$

and $L_t^n = L_{t-1}^n$ otherwise.

By induction, (4.51) and (4.52) are satisfied, and L^n and U^n are seen to be the maximal respective minimal lower and upper processes with these properties. Hence, if

$$M = \inf\{n > 0 : L_{-1}^n = U_{-1}^n\}$$

denotes the first time a pair of lower and upper processes coalesce, we have that $U_{-1}^M \sim f$. We call M the *coalescence time*.

By induction we have the following *funnelling property*,

$$L_t^{n-1} \subseteq L_t^n \subseteq U_t^n \subseteq U_t^{n-1}, \quad -n \geq t > 0. \tag{4.55}$$

So for $n = 1, 2, 3, \ldots$, we may generate pairs of upper and lower processes as described above, until $L_{-1}^n = U_{-1}^n$, and then return $U_{-1}^n = U_{-1}^M$. It is, however, usually more efficient to use a *doubling scheme*: for $n = 1$ generate Z_{-1}, R_{-1}, U^1, L^1, and for $n = 2, 4, 8, \ldots$, generate $Z_{-n}, R_{-n}, \ldots, Z_{-1-n/2}$, $R_{-1-n/2}, U^n, L^n$, until $U_{-1}^n = L_{-1}^n$; if N denotes the first such n where this happens, then return $U_{-1}^N \sim f$. This follows from the fact that $U_{-1}^N = U_{-1}^M$, since $M \leq N$ and because of (4.55). As noticed in Propp & Wilson (1996), $N \leq 4M$. As observed in Berthelsen & Møller (2001b), the doubling scheme can be improved slightly by using the scheme $n = m, m+1, m+2, m+4, m+8, \ldots$, where $-m$ denotes the first time a point in Z_{-1} is born.

It may be time consuming to find the maximum and minimum in (4.53) and (4.54) unless the Papangelou conditional intensity satisfies certain *monotonicity properties*: we say that f is *attractive* if $\lambda^*(x, \xi) \leq \lambda^*(y, \xi)$ whenever $x \subset y$; and *repulsive* if $\lambda^*(x, \xi) \geq \lambda^*(y, \xi)$ whenever $x \subset y$; notice that in both cases we easily obtain the maximum and minimum in (4.53) and (4.54). As a matter of fact many point process models satisfy one of these conditions. For instance, the Strauss process (4.31) is repulsive.

Fernández, Ferrari & Garcia (1999) introduce another perfect simulation algorithm based on spatial birth-death processes, but without using upper and lower processes, and with no requirement of monotonicity properties. The algorithm is reviewed and compared with the one using upper and lower processes in Berthelsen & Møller (2001b). In general, if f is attractive or repulsive, the algorithm described in this section is the most efficient.

4.8 Further reading and concluding remarks

Textbooks and review articles on different aspects of spatial point processes include Matérn (1960, 1986), Bartlett (1963, 1964), Kerstan, Matthes & Mecke (1974), Kallenberg (1975), Matheron (1975), Ripley (1977, 1981, 1988), Diggle (1983), Penttinen (1984), Daley & Vere-Jones (1988), Baddeley & Møller (1989), Mecke, Schneider, Stoyan & Weil (1990), Karr (1991), Cressie (1993), Reiss (1993), Stoyan & Stoyan (1994), Stoyan et al. (1995), Geyer (1999), Møller (1999), van Lieshout (2000), and Ohser & Mücklich (2000).

Examples 1 and 2 concern a single planar pattern; many other examples of simple point processes and marked point processes can be found in the abovementioned books and review papers. Replicated point patterns in 2 and 3 dimensions are discussed in Diggle et al. (1991), Baddeley et al. (1993), and Diggle et al. (2000).

Much of the more classical literature on spatial point processes deal with non-parametric methods for stationary point processes. The focus has changed over the years to exploiting more and more flexible and complex parametric statistical models which are analysed using a Bayesian or likelihood approach by means of MCMC methods as exemplified in this chapter. Most of the literature deal with homogeneous point patterns, but we have also treated the inhomogeneous case in some detail. Jensen & Nielsen (2001) provide a review on some recent developments in this direction.

We expect an increasing need for developing models and tools for analysing inhomogeneous point patterns as new technology such as geographical information systems makes huge amounts of spatial point process data available. Such patterns may exhibit both large-scale aggregation and small-scale inhibition, and simulations may allow to model this using Cox and Markov point processes as building blocks in e.g. a nonparametric Bayesian setting, modifying the approach in Heikkinen & Penttinen (1999) and Berthelsen & Møller (2001a).

We have demonstrated that MCMC methods provide indispensable tools for analysing spatial point process models. Perfect simulation is obviously appealing for many reasons, not at least in connection to simulation-based inference. Its advantages and limitations for statistical inference for spatial point processes are discussed in Berthelsen & Møller (2001a).

Existing codes and *software* may be used in order to save time. Geyer (1999) contains a short description and the address to the code he uses for MCMC likelihood analysis of certain point process models. Adrian Baddeley and Rolf Turner have written spatstat, a contributed library in S-PLUS and R for the statistical analysis of spatial point patterns (available at http://www.maths.uwa.edu.au/∼adrian/spatstat/). They provide also references to other spatial statistical software.

However, it is necessary to write new code or modify existing code for many models. The code for a sampler and e.g. MCMC likelihood or Bayesian analysis may be written in C (to speed things up), while a high-level statistical language such as S or R may be used for graphical procedures.

Acknowledgments: Kasper K. Berthelsen and Jakob Gulddahl Rasmussen are acknowledged for their useful comments.

4.9 References

Adler, R. (1981). *The Geometry of Random Fields*, Wiley, New York.

Baddeley, A. & Gill, R. D. (1997). Kaplan-Meier estimators of distance distributions for spatial point processes, *Annals of Statistics* **25**: 263–292.

Baddeley, A. J., Moyeed, R. A., Howard, C. V. & Boyde, A. (1993). Analysis of a three-dimensional point patttern with applications, *Applied Statistics* **42**: 641–668.

Baddeley, A. J. & van Lieshout, M. N. M. (1993). Stochastic geometry models in high-level vision, *in* K. V. Mardia & G. K. Kanji (eds), *Statistics and Images, Advances in Applied Statistics, a supplement to the Journal of Applied Statistics*, Vol. 20, Carfax Publishing, Abingdon, chapter 11, pp. 235–256.

Baddeley, A. J. & van Lieshout, M. N. M. (1995). Area-interaction point processes, *Annals of the Institute of Statistical Mathematics* **46**: 601–619.

Baddeley, A. & Møller, J. (1989). Nearest-neighbour Markov point processes and random sets, *International Statistical Review* **2**: 89–121.

Baddeley, A., Møller, J. & Waagepetersen, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns, *Statistica Neerlandica* **54**: 329–350.

Baddeley, A. & Silverman, B. W. (1984). A cautionary example for the use of second-order methods for analysing point patterns, *Biometrics* **40**: 1089–1094.

Baddeley, A. & Turner, R. (2000). Practical maximum pseudolikelihood for spatial point patterns, *Australian and New Zealand Journal of Statistics* **42**: 283–322.

Bartlett, M. S. (1963). The spectral analysis of point processes, *Journal of the Royal Statistical Society Series B* **29**: 264–296.

Bartlett, M. S. (1964). The spectral analysis of two-dimensional point processes, *Biometrika* **51**: 299–311.

Benes, V., Bodlak, K., Møller, J. & Waagepetersen, R. P. (2002). Bayesian analysis of log Gaussian Cox process models for disease mapping, *Technical Report R-02-2001*, Department of Mathematical Sciences, Aalborg University.

Berman, M. & Turner, T. R. (1992). Approximating point process likelihoods with GLIM, *Applied Statistics* **41**: 31–38.

Berthelsen, K. K. & Møller, J. (2001a). Perfect simulation and inference for spatial point processes, *Technical Report R-01-2009*, Department of Mathematical Sciences, Aalborg University. Conditionally accepted for publication in the *Scandinavian Journal of Statistics.*

Berthelsen, K. K. & Møller, J. (2001b). A primer on perfect simulation for spatial point processes, *Technical Report R-01-2026*, Department of Mathematical Sciences, Aalborg University. To appear in *Bulletin of the Brazilian Mathematical Society,* **33**, 2003.

Berthelsen, K. K. & Møller, J. (2002). Spatial jump processes and perfect simulation, *in* K. Mecke & D. Stoyan (eds), *Morphology of Condensed Matter*, Lecture Notes in Physics, Springer-Verlag. To appear.

Besag, J. (1977a). Some methods of statistical analysis for spatial data, *Bulletin of the Institute of International Statistics* **47**: 77–92.

Besag, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion), *Journal of the Royal Statistical Society Series B* **36**: 192–236.

Besag, J. E. (1975). Statistical analysis of non-lattice data, *The Statistician* **24**: 179–195.

Besag, J. E. (1977b). Discussion on the paper by Ripley (1977), *Journal of the Royal Statistical Society Series B* **39**: 193–195.

Besag, J. E. (1994). Discussion on the paper by Grenander and Miller, *Journal of the Royal Statistical Society Series B* **56**: 591–592.

Besag, J., Milne, R. & Zachary, S. (1982). Point process limits of lattice processes, *Journal of Applied Probability* **19**: 210–216.

Breyer, L. A. & Roberts, G. O. (2000). From Metropolis to diffusions: Gibbs states and optimal scaling, *Stochastic Processes and their Applications* **90**: 181–206.

Brix, A. (1999). Generalized gamma measures and shot-noise Cox processes, *Advances in Applied Probability* **31**: 929–953.

Brix, A. & Chadoeuf, J. (2000). Spatio-temporal modeling of weeds and shot-noise G Cox processes. Submitted.

Brix, A. & Kendall, W. S. (2002). Simulation of cluster point processes without edge effects, *Advances in Applied Probability* **34**: 267–280.

Brix, A. & Møller, J. (2001). Space-time multitype log Gaussian Cox processes with a view to modelling weed data, *Scandinavian Journal of Statistics* **28**: 471–488.

Chan, K. S. & Geyer, C. J. (1994). Discussion of the paper 'Markov chains for exploring posterior distributions' by Luke Tierney, *Annals of Statistics* **22**: 1747–1747.

Christensen, O. F., Møller, J. & Waagepetersen, R. P. (2001). Geometric ergodicity of Metropolis-Hastings algorithms for conditional simulation in generalised linear mixed models, *Methodology and Computing in Applied Probability* **3**: 309–327.

Christensen, O. F. & Waagepetersen, R. (2002). Bayesian prediction of spatial count data using generalised linear mixed models, *Biometrics* **58**: 280–286.

Coles, P. & Jones, B. (1991). A lognormal model for the cosmological mass distribution, *Monthly Notices of the Royal Astronomical Society* **248**: 1–13.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*, second edn, Wiley, New York.

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

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

Diggle, P. J. (1985). A kernel method for smoothing point process data, *Applied Statistics* **34**: 138–147.

Diggle, P. J., Lange, N. & Beněs, F. (1991). Analysis of variance for replicated spatial point patterns in clinical neuroanatomy, *Journal of the American Statistical Association* **86**: 618–625.

Diggle, P. J., Mateu, L. & Clough, H. E. (2000). A comparison between parametric and non-parametric approaches to the analysis of replicated spatial point patterns, *Advances of Applied Probability* **32**: 331–343.

Fernández, R., Ferrari, P. A. & Garcia, N. L. (1999). Perfect simulation for interacting point processes, loss networks and Ising models. Manuscript.

Fiksel, T. (1984). Estimation of parameterized pair potentials of marked and nonmarked Gibbsian point processes, *Elektronische Informationsverarbeitung und Kypernetik* **20**: 270–278.

Gelfand, A. E. (1996). Model determination using sampling-based methods, *in* W. R. Gilks, S. Richardson & D. J. Spiegelhalter (eds), *Markov chain Monte Carlo in Practice*, Chapman and Hall, London, pp. 145–161.

Gelman, A. & Meng, X.-L. (1998). Simulating normalizing constants: from importance sampling to bridge sampling to path sampling, *Statistical Science* **13**: 163–185.

Georgii, H.-O. (1976). Canonical and grand canonical Gibbs states for continuum systems, *Communications of Mathematical Physics* **48**: 31–51.

Georgii, H.-O. (1988). *Gibbs Measures and Phase Transition*, Walter de Gruyter, Berlin.

Geyer, C. J. (1991). Markov chain Monte Carlo maximum likelihood, *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, pp. 156–163.

Geyer, C. J. (1994). On the convergence of Monte Carlo maximum likelihood calculations, *Journal of the Royal Society of Statistics Series B* **56**: 261–274.

Geyer, C. J. (1999). Likelihood inference for spatial point processes, *in* O. E. Barndorff-Nielsen, W. S. Kendall & M. N. M. van Lieshout (eds), *Stochastic Geometry: Likelihood and Computation*, Chapman and Hall/CRC, London, Boca Raton, pp. 79–140.

Geyer, C. J. & Møller, J. (1994). Simulation procedures and likelihood inference for spatial point processes, *Scandinavian Journal of Statistics* **21**: 359–373.

Geyer, C. J. & Thompson, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data, *Journal of the Royal Society of Statistics Series B* **54**: 657–699.

Goulard, M., Särkkä, A. & Grabarnik, P. (1996). Parameter estimation for marked Gibbs point processes through the maximum pseudo-likelihood method, *Scandinavian Journal of Statistics* **23**: 365–379.

Green, P. J. (1995). Reversible jump MCMC computation and Bayesian model determination, *Biometrika* **82**: 711–732.

Gu, M. G. & Zhu, H.-T. (2001). Maximum likelihood estimation for spatial models by Markov chain Monte Carlo stochastic approximation, *Journal of the Royal Statistical Society Series B* **63**: 339–355.

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

Heikkinen, J. & Arjas, E. (1998). Non-parametric Bayesian estimation of a spatial Poisson intensity, *Scandinavian Journal of Statistics* **25**: 435–450.

Heikkinen, J. & Penttinen, A. (1999). Bayesian smoothing in the estimation of the pair potential function of Gibbs point processes, *Bernoulli* **5**: 1119–1136.

Jensen, E. B. V. & Nielsen, L. S. (2001). A review on inhomogeneous spatial point processes, *in* I. V. Basawa, C. C. Heyde & R. L. Taylor (eds), *Selected Proceedings of the Symposium on Inference for Stochastic Processes*, Vol. 37, IMS Lecture Notes & Monographs Series, Beachwood, Ohio, pp. 297–318.

Jensen, J. L. & Künsch, H. R. (1994). On asymptotic normality of pseudo likelihood estimates for pairwise interaction processes, *Annals of the Institute of Statistical Mathematics* **46**: 475–486.

Jensen, J. L. & Møller, J. (1991). Pseudolikelihood for exponential family models of spatial point processes, *Annals of Applied Probability* **3**: 445–461.

Kallenberg, O. (1975). *Random Measures*, Akadamie-Verlag, Berlin.

Kallenberg, O. (1984). An informal guide to the theory of conditioning in point processes, *International Statistical Review* **52**: 151–164.

Karr, A. F. (1991). *Point Processes and Their Statistical Inference*, Marcel Dekker, New York.

Kelly, F. P. & Ripley, B. D. (1976). A note on Strauss' model for clustering, *Biometrika* **63**: 357–360.

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

Kendall, W. S. & Møller, J. (2000). Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes, *Advances in Applied Probability* **32**: 844–865.

Kerscher, M. (2000). Statistical analysis of large-scale structure in the Universe, *in* K. R. Mecke & D. Stoyan (eds), *Statistical Physics and Spatial Statistics*, Lecture Notes in Physics, Springer, Berlin, pp. 36–71.

Kerstan, J., Matthes, K. & Mecke, J. (1974). *Unbegrenzt teilbare Punktprozesse*, Akademie-Verlag, Berlin.

Kingman, J. F. C. (1993). *Poisson Processes*, Clarendon Press, Oxford.

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

Lieshout, M. N. M. van & Baddeley, A. J. (1996). A nonparametric measure of spatial interaction in point patterns, *Statistica Neerlandica* **50**: 344–361.

Loizeaux, M. A. & McKeague, I. W. (2001). Perfect sampling for posterior landmark distributions with an application to the detection of disease clusters, *in* I. V. Basawa, C. C. Heyde & R. L. Taylor (eds), *Selected Proceedings of the Symposium on Inference for Stochastic Processes*, Vol. 37, IMS Lecture Notes & Monographs Series, Beachwood, Ohio, pp. 321–331.

Lund, J., Penttinen, A. & Rudemo, M. (1999). Bayesian analysis of spatial point patterns from noisy observations. Available at http://www.math.chalmers.se/Stat/ Research/Preprints/.

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

Lund, J. & Thönnes, E. (2000). Perfect simulation for point processes given noisy observations. Research Report 366, Department of Statistics, University of Warwick.

Mase, S. (1995). Consistency of the maximum pseudo-likelihood estimator of continuous state space Gibbs processes, *Annals of Applied Probability* **5**: 603–612.

Mase, S. (1999). Marked Gibbs processes and asymptotic normality of maximum pseudo-likelihood estimators, *Mathematische Nachrichten* **209**: 151–169.

Mase, S., Møller, J., Stoyan, D., Waagepetersen, R. P. & Döge, G. (2001). Packing densities and simulated tempering for hard core Gibbs point processes, *Annals of the Institute of Statistical Mathematics* **53**: 661–680.

Matérn, B. (1960). Spatial Variation. Meddelanden från Statens Skogforskningsinstitut, Band 49, No. 5.

Matérn, B. (1986). *Spatial Variation*, Lecture Notes in Statistics. Springer-Verlag, Berlin.

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

Mecke, J. (1967). Stationäre zufällige Maße auf lokalkompakten Abelschen Gruppen, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **9**: 36–58.

Mecke, J., Schneider, R. G., Stoyan, D. & Weil, W. R. R. (1990). *Stochastische Geometrie*, Birkhäuser Verlag, Basel.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). Equations of state calculations by fast computing machines, *Journal of Chemical Physics* **21**: 1087–1092.

Møller, J. (1989). On the rate of convergence of spatial birth-and-death processes, *Annals of the Institute of Statistical Mathematics* **3**: 565–581.

Møller, J. (1999). Markov chain Monte Carlo and spatial point processes, *in* O. E. Barndorff-Nielsen, W. S. Kendall & M. N. M. van Lieshout (eds), *Stochastic Geometry: Likelihood and Computations*, number 80 in *Monographs on Statistics and Applied Probability*, Chapman and Hall/CRC, Boca Raton, pp. 141–172.

Møller, J. (2001). A review of perfect simulation in stochastic geometry, *in* I. V. Basawa, C. C. Heyde & R. L. Taylor (eds), *Selected Proceedings of the Symposium on Inference for Stochastic Processes*, Vol. 37, IMS Lecture Notes & Monographs Series, Beachwood, Ohio, pp. 333–355.

Møller, J. (2002a). A comparison of spatial point process models in epidemiological applications, *in* P. J. Green, N. L. Hjort & S. Richardson (eds), *Highly Structured Stochastic Systems*, Oxford University Press, Oxford. To appear.

Møller, J. (2002b). Shot noise Cox processes, *Technical Report R-02-2009*, Department of Mathematical Sciences, Aalborg University.

Møller, J., Syversveen, A. R. & Waagepetersen, R. P. (1998). Log Gaussian Cox processes, *Scandinavian Journal of Statistics* **25**: 451–482.

Møller, J. & Waagepetersen, R. P. (2002). Statistical inference for Cox processes, *in* A. B. Lawson & D. Denison (eds), *Spatial Cluster Modelling*, Chapman and Hall/CRC, Boca Raton.

Møller, J. & Waagepetersen, R. P. (2003). *Statistical Inference and Simulation for Spatial Point Processes*, Chapman and Hall/CRC, Boca Raton. In preparation.

Neyman, J. & Scott, E. L. (1958). Statistical approach to problems of cosmology, *Journal of the Royal Statistical Society Series B* **20**: 1–43.

Nguyen, X. X. & Zessin, H. (1979). Integral and differential characterizations of Gibbs processes, *Mathematische Nachrichten* **88**: 105–115.

Ohser, J. & Mücklich, F. (2000). *Statistical Analysis of Microstructures in Materials Science*, Wiley, New York.

Peebles, P. J. E. (1974). The nature of the distribution of galaxies, *Astronomy and Astrophysics* **32**: 197–202.

Peebles, P. J. E. & Groth, E. J. (1975). Statistical analysis of extragalactic objects. V. Three-point correlation function for the galaxy distribution in the Zwicky catalog, *Astrophysical Journal* **196**: 1–11.

Penttinen, A. (1984). *Modelling Interaction in Spatial Point Patterns: Parameter Estimation by the Maximum Likelihood Method*, Number 7 in Jyväskylä Studies in Computer Science, Economics, and Statistics.

Penttinen, A., Stoyan, D. & Henttonen, H. M. (1992). Marked point processes in forest statistics, *Forest Science* **38**: 806–824.

Preston, C. (1976). *Random Fields*, Lecture Notes in Mathematics, 534. Springer-Verlag, Berlin-Heidelberg.

Preston, C. J. (1977). Spatial birth-and-death processes, *Bulletin of the International Statistical Institute* **46**: 371–391.

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

Quine, M. P. & Watson, D. F. (1984). Radial simulation of n-dimensional Poisson processes, *Journal of Applied Probability* **21**: 548–557.

Rathbun, S. L. (1996). Estimation of Poisson intensity using partially observed concomitant variables, *Biometrics* **52**: 226–242.

Reiss, R.-D. (1993). *A Course on Point Processes*, Springer Verlag, New York.

Ripley, B. D. (1977). Modelling spatial patterns (with discussion), *Journal of the Royal Statistical Society Series B* **39**: 172–212.

Ripley, B. D. (1979). Simulating spatial patterns: dependent samples from a multivariate density. Algorithm AS 137, *Applied Statistics* **28**: 109–112.

Ripley, B. D. (1981). *Spatial Statistics*, Wiley, New York.

Ripley, B. D. (1988). *Statistical Inference for Spatial Processes*, Cambridge University Press, Cambridge.

Ripley, B. D. & Kelly, F. P. (1977). Markov point processes, *Journal of the London Mathematical Society* **15**: 188–192.

Roberts, G. O., Gelman, A. & Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms, *Annals of Applied Probability* **7**: 110–120.

Roberts, G. O. & Rosenthal, J. S. (1997). Geometric ergodicity and hybrid Markov chains, *Electronic Communications in Probability* **2**: 13–25.

Roberts, G. O. & Rosenthal, J. S. (1998). Optimal scaling of discrete approximations to Langevin diffusions, *Journal of the Royal Statistical Society Series B* **60**: 255–268.

Roberts, G. O. & Tweedie, R. L. (1996). Exponential convergence of Langevin diffusions and their discrete approximations, *Bernoulli* **2**: 341–363.

Rossky, P. J., Doll, J. D. & Friedman, H. L. (1978). Brownian dynamics as smart Monte Carlo simulation, *Journal of Chemical Physics* **69**: 4628–4633.

Ruelle, D. (1969). *Statistical Mechanics: Rigorous Results*, W.A. Benjamin, Reading, Massachusetts.

Schladitz, K. & Baddeley, A. J. (2000). A third-order point process characteristic, *Scandinavian Journal of Statistics* **27**: 657–671.

Schlather, M. (2001). On the second-order characteristics of marked point processes, *Bernoulli* **7**: 99–117.

Stoyan, D., Kendall, W. S. & Mecke, J. (1995). *Stochastic Geometry and Its Applications*, second edn, Wiley, Chichester.

Stoyan, D. & Stoyan, H. (1994). *Fractals, Random Shapes and Point Fields*, Wiley, Chichester.

Stoyan, D. & Stoyan, H. (2000). Improving ratio estimators of second order point process characteristics, *Scandinavian Journal of Statistics* **27**: 641–656.

Strauss, D. J. (1975). A model for clustering, *Biometrika* **63**: 467–475.

Thönnes, E. (1999). Perfect simulation of some point processes for the impatient user, *Advances in Applied Probability* **31**: 69–87.

Waagepetersen, R. & Sorensen, S. (2001). A tutorial on reversible jump MCMC with a view toward applications in QTL-mapping, *International Statistical Review* **69**(1): 49–61.

Wolpert, R. L. & Ickstadt, K. (1998). Poisson/gamma random field models for spatial statistics, *Biometrika* **85**: 251–267.

Wood, A. T. A. & Chan, G. (1994). Simulation of stationary Gaussian processes in $[0, 1]^d$, *Journal of Computational and Graphical Statistics* **3**: 409–432.