Statistical inference for Cox processes*

Jesper Møller
Rasmus P. Waagepetersen

1 Introduction

This chapter is concerned with statistical inference for a large class of point process models, which were studied in a seminal paper by Cox (1955) under the name doubly stochastic Poisson processes, but are today usually called Cox processes. Much of the literature on Cox processes is concerned with point processes defined on the real line $\mathbb{R}$, but we pay attention to the general d-dimensional Euclidean case of $\mathbb{R}^d$, and in particular to the planar case $d = 2$ (which covers most cases of spatial applications). However, the theory does not make much use of the special properties of $\mathbb{R}^d$, and extensions to other state spaces are rather obvious. We discuss in some detail how various Cox process models can be constructed and simulated, study nonparametric as well as parametric analysis (with a particular emphasis on minimum contrast estimation), and relate the methods to simulated and real datasets of aggregated spatial point patterns. Further material on Cox processes can be found in the references mentioned in the sequel and in Grandell (1976), Diggle (1983), Daley & Vere-Jones (1988), Stoyan, Kendall & Mecke (1995), and the references therein.

To explain shortly what is meant by a Cox process, consider the spatial point patterns in Figure 1. As demonstrated in Section 3, each point pattern in Figure 1 is more aggregated than can be expected under a homogeneous Poisson process (the reference model in statistics for spatial point patterns). The aggregation is in fact caused by a realization $z = \{z(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}^2\}$ of an underlying nonnegative spatial process Z, which is shown in gray scale in

*To appear as a chapter in *Spatial Cluster Modelling*, eds. Andrew B. Lawson and David Denison, Chapman and Hall.

Figure 1. If the conditional distribution of a point process X given $Z = z$ is an inhomogeneous Poisson process with intensity function z, we call X a Cox process with random intensity surface Z (for details, see Section 3). Note that points of X are most likely to occur in areas where Z is large, cf. Figure 1. In many applications we can think of Z as an underlying "environmental" process. Aggregation in a spatial point process X may indeed be due to other sources, including (i) clustering of the points in X around the points of another point process C, and (ii) interaction between the points in X. For certain models, (i) is equivalent to a Cox process model (see Section 5.1). This is in fact the case for the upper left point pattern in Figure 1, where the cluster centers $C = \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_m\}$ are also shown. The case (ii) is not considered in this contribution, but we refer the interested reader to the literature on Markov point processes, see e.g. Møller (1999) and van Lieshout (2000).

General definitions and descriptions of Poisson and Cox processes are given in Sections 2–3, while Section 4 provides some background on non-parametric analysis. Section 5 concerns certain parametric models for Cox processes. Specifically, Section 5.1 considers the case where Neyman-Scott processes (Neyman & Scott 1958) are Cox processes, Section 5.2 deals with log Gaussian Cox processes (Coles & Jones 1991, Møller, Syversveen & Waagepetersen 1998), and Section 5.3 with shot noise G Cox processes (Brix 1999). The latter class of models include the Poisson/gamma processes (Wolpert & Ickstadt 1998). As explained in more detail later, the point patterns in Figure 1 are realizations of a certain Neyman-Scott process, a log Gaussian Cox process (LGCP), a shot-noise G Cox process (SNGCP), and a certain "logistic process", respectively.

In most applications with an aggregated point pattern modeled by a Cox process X, the underlying environmental process Z is unobserved. Further, only $X \cap W$ is observed, where W is a bounded region contained in the area where the points in X occur. In Section 6 we discuss various approaches to estimation in parametric models for Cox processes and focus in particular on minimum contrast estimation. In Section 7 we discuss how Z and $X \setminus W$ can be predicted under the various models from Section 5. Section 8 contains some concluding remarks.

The discussion in Sections 5–7 will be related to the dataset in Figure 2, which shows the positions of 378 weed plants (*Veronica* spp./ speedwell). This point pattern is a subset of a much larger dataset analyzed in Brix & Møller (2001) and Brix & Chadoeuf (2000) where several weed species at

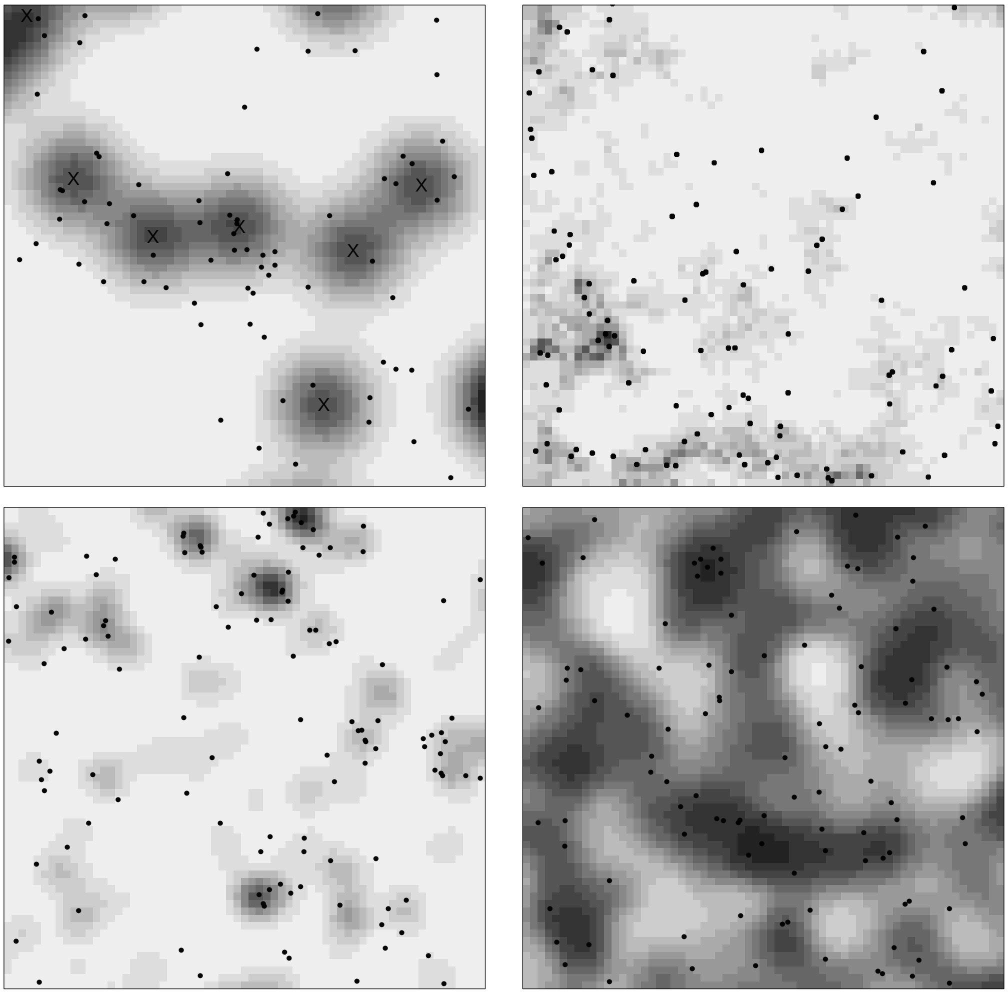

Figure 1: From left to right, top to bottom: realizations of Thomas, LGCP, SNGCP, and "logistic" processes, and associated random intensity surfaces (in gray scale). The crosses in the upper left plot show the cluster centers for the Thomas process. For more details, see Example 4 (Thomas), Example 5 (LGCP), Example 6 (SNGCP), and Example 2 and Example 5 ("logistic").

different sampling dates were considered. Note that we have rotated the design 90° in Figure 2. The 45 frames are of size 30×20 cm^2, and they are organized 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. The observation window W is given by the union of the 45 frames.

2 Poisson processes

This section surveys some fundamental concepts of point processes and Poisson processes, without going too much into measure theoretical details; for further details, we refer to Daley & Vere-Jones (1988) and Kingman (1993).

By a *point process* X in $\mathbb{R}^d$ we understand a random subset $X \subset \mathbb{R}^d$ which is locally finite, i.e. $X \cap A$ is finite for any bounded region $A \subset \mathbb{R}^d$. By a region we mean a Borel subset of $\mathbb{R}^d$, and measurability of X is equivalent to that

$$N(A) \equiv \text{card}(X \cap A)$$

is a random variable for each bounded region A.

Henceforth X denotes a point process in $\mathbb{R}^d$. Its distribution is determined by the joint distribution of $N(A_1), \ldots, N(A_n)$ for any disjoint regions $A_1, \ldots, A_n$ and any integer $n \geq 1$.

Now, let μ denote an arbitrary locally finite and diffuse measure defined on the regions in $\mathbb{R}^d$, i.e. $\mu(A) < \infty$ for bounded regions A, and $\mu(\{\boldsymbol{x}\}) = 0$ for all singleton sets $\{\boldsymbol{x}\} \subset \mathbb{R}^d$. We say that X is a *Poisson process* with *intensity measure* μ if the following two properties are satisfied:

(a) *independent scattering:* $N(A_1), \ldots, N(A_n)$ are independent for disjoint bounded regions $A_1, \ldots, A_n$ and integers $n \geq 2$;

(b) $N(A)$ is Poisson distributed with mean $\mu(A)$ for bounded regions A.

The properties (a)–(b) are easily seen to be equivalent to (b)–(c), where

(c) for any bounded region A with $\mu(A) > 0$ and any integer $n \geq 1$, conditionally on $N(A) = n$, the n points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ in $X \cap A$ are independent and each point has distribution $\mu(A \cap \cdot)/\mu(A)$.

The conditional distribution in (c) is called a *binomial process*. Using (a)–(c) it is not hard to verify that the Poisson process exists.

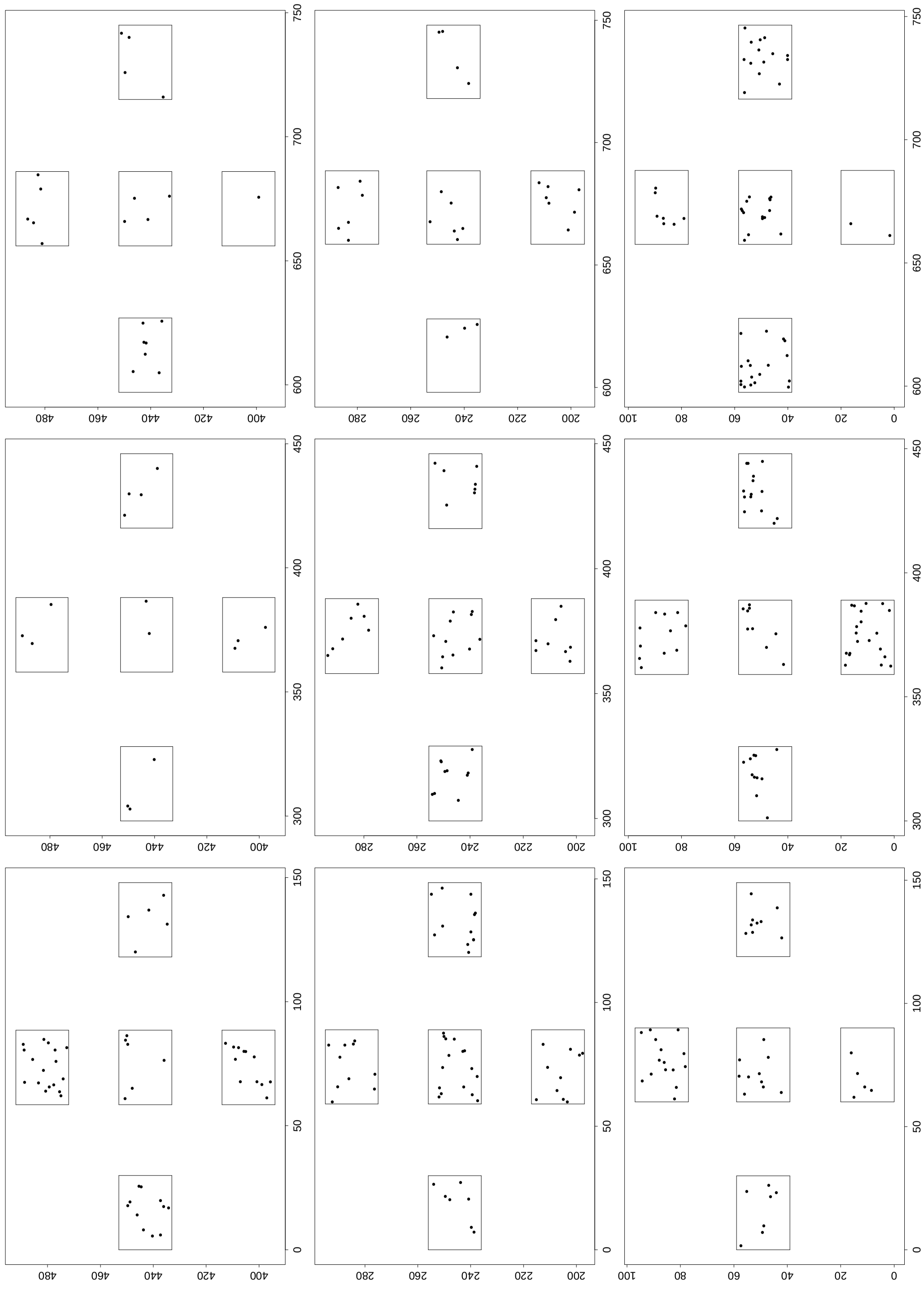

Figure 2: Positions of weed plants when the design is rotated 90°.

A point process is said to be *stationary* if its distribution is invariant under translations in $\mathbb{R}^d$, and *isotropic* if its distribution is invariant under rotations around the origin in $\mathbb{R}^d$. A Poisson process which is stationary is also isotropic, so it is called a *homogeneous* Poisson point process; otherwise it is said to be an *inhomogeneous* Poisson process.

Often μ has a density $\rho : \mathbb{R}^d \to [0, \infty)$ so that $\mu(A) = \int_A \rho(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$ for all regions A. Then ρ is called the *intensity function*. A homogeneous Poisson process has a constant intensity function, and this constant is simply called the *intensity*.

By (b)–(c) it is very easy to simulate a homogeneous Poisson process X on e.g. a rectangular or spherical region B. Denote ρ_{hom} the intensity of X, and imagine we have simulated X on B. Suppose we want to simulate another Poisson process X_{th} on a bounded region $A \subseteq B$, where X_{th} has an intensity function ρ_{th} which is bounded on A by ρ_{hom}. Then we obtain a simulation of $X_{\text{th}} \cap A$ by including/excluding the points from $X \cap A$ in $X_{\text{th}} \cap A$ independently of each other, so that a point $\boldsymbol{x} \in X \cap A$ is included in $X_{\text{th}} \cap A$ with probability $\pi(\boldsymbol{x}) = \rho_{\text{th}}(\boldsymbol{x})/\rho_{\text{hom}}$. This procedure is called *independent thinning*; it is extended in Example 2, Sections 3–4.

Finally, all moments of the counts $N(\cdot)$ for a Poisson process are easily obtained from (a)–(b). For instance,

$$\mathbb{E}N(A) = \mu(A) \quad \text{and} \quad \mathbb{C}\text{ov}(N(A), N(B)) = \mu(A \cap B) \tag{1}$$

for bounded regions A and B.

3 Cox processes

A natural extension of a Poisson process is to let μ be a realization of a random measure M so that the conditional distribution of X given $M = \mu$ follows a Poisson process with intensity measure μ. Then X is said to be a *Cox process driven by M*. This definition can be extended to both multivariate point processes and to space-time processes, see Diggle (1983), Møller, Syversveen & Waagepetersen (1998), Brix & Møller (2001), and Brix & Diggle (2001).

Example 1 A simple example of a Cox process is a *mixed Poisson process* with $M(A) = \int_A Z \mathrm{d}\boldsymbol{x}$ where Z is a positive random variable, i.e. $X|Z$ follows a homogeneous Poisson process with intensity Z. For example, if Z is gamma distributed, $N(A)$ follows a negative binomial distribution for bounded regions A. Note that $N(A)$ and $N(B)$ are correlated for disjoint bounded regions A and B (except in the trivial case where Z is almost surely constant, i.e. when we have a homogeneous Poisson process).

Example 2 Suppose that X is a Cox process driven by M, and $\Pi = \{\Pi(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}^d\} \subseteq [0,1]$ is a random process which is independent of (X, M). Let X_{th} denote the point process obtained by *random independent thinning* of the points in X with retention probabilities Π. More precisely, conditionally on $\Pi = \pi$, X_{th} is obtained by independent thinning of X where a point in X is retained with probability $\pi(\boldsymbol{x})$. Then X_{th} is a Cox process driven by $M_{\mathrm{th}}(A) = \int_A \Pi(\boldsymbol{x}) M(\mathrm{d}\boldsymbol{x})$. A simple example is shown in the lower right plot in Figure 1, where X is a homogenous Poisson process (i.e. M is proportional to Lebesgue measure), and Π follows a "logistic" process, i.e. $\log(\Pi/(1-\Pi))$ is a Gaussian process (see Example 5 in Section 5.2).

Cox processes are like inhomogeneous Poisson processes models for aggregated point patterns. Usually in applications M is unobserved, and so we cannot distinguish a Cox process X from its corresponding Poisson process $X|M$ when only one realization of $X \cap W$ is available (where W denotes the observation window). Which of the two models might be most appropriate, i.e. whether M should be random or "systematic"/deterministic, depends on

- prior knowledge and the scientific questions to be investigated: if e.g. one wants to investigate the dependence of certain covariates associated to M, these may be treated as systematic terms, while unobserved effects may be treated as random terms (for an example, see Benes, Bodlak, Møller & Waagepetersen 2001);
- the particular application: if it seems difficult to model an aggregated point pattern with a parametric class of inhomogeneous Poisson processes (e.g. a class of polynomial intensity functions), Cox process models such as those in Section 5 may allow more flexibility and/or a more parsimonious parametrization;
- another application is nonparametric Bayesian modelling: nonparametric Bayesian smoothing for the intensity surface of an inhomogeneous Poisson process is treated in Heikkinen & Arjas (1998); see also Remark 4 in Section 8.

Distributional properties of a Cox process X driven by M follow immediately by conditioning on M and exploiting the properties of the Poisson process $X|M$. For instance, by (1),

$$\mathbb{E}N(A) = \mathbb{E}M(A)$$

and

$$\mathbb{C}\mathrm{ov}(N(A), N(B)) = \mathbb{C}\mathrm{ov}(M(A), M(B)) + \mathbb{E}M(A \cap B)$$

for bounded regions A and B. Hence, $\mathbb{V}\mathrm{ar}(N(A)) = \mathbb{V}\mathrm{ar}(M(A)) + \mathbb{E}M(A) \geq \mathbb{E}N(A)$ with equality only when $M(A)$ is almost surely constant as in the Poisson case. In other words, a Cox process exhibits over dispersion when compared to a Poisson process.

In many specific models for Cox processes, including those considered in Examples 1–2 and in Section 5, M is specified by a nonnegative spatial process $Z = \{Z(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}^d\}$ so that

$$M(A) = \int_A Z(\boldsymbol{x})\mathrm{d}\boldsymbol{x}. \tag{2}$$

Then we say that X is *driven by the random intensity surface* Z. In the sequel we restrict attention to Cox processes driven by a random intensity surface.

Simulation of X is in principle easy: if we have a simulation $z_A = \{z(\boldsymbol{x}) : \boldsymbol{x} \in A\}$ of Z restricted to a bounded region A, where z_A is bounded by a constant, then the simulation method at the end of Section 2 can be used to obtain a realization of $X \cap A | Z_A = z_A$. Note that by the independent scattering property (a) in Section 2, the conditional distribution $X \cap A|Z$ depends only on Z_A.

4 Summary statistics

The first and second order moments of the counts $N(\cdot)$ for a Cox process can be expressed in terms of two functions:

$$\rho(\boldsymbol{x}) = \mathbb{E}Z(\boldsymbol{x}) \quad \text{and} \quad g(\boldsymbol{x}_1, \boldsymbol{x}_2) = \mathbb{E}[Z(\boldsymbol{x}_1)Z(\boldsymbol{x}_2)]/[\rho(\boldsymbol{x}_1)\rho(\boldsymbol{x}_2)]$$

which are called the *intensity function* and the *pair correlation function*, respectively ("pair correlation function" is standard terminology, though it is somewhat misleading). The intensity and pair correlation functions can be defined for general point process models and not just Cox processes, see e.g. Stoyan et al. (1995). Intuitively, $\mathbb{E}[Z(\boldsymbol{x}_1)Z(\boldsymbol{x}_2)]\mathrm{d}\boldsymbol{x}_1\mathrm{d}\boldsymbol{x}_2$ is the probability for having a point from X in each of the infinitesimally small volumes $\mathrm{d}\boldsymbol{x}_1$ and $\mathrm{d}\boldsymbol{x}_2$. It is, however, more informative to work with the pair correlation function which is normalized with the "marginal probabilities" $\rho(\boldsymbol{x}_1)d\boldsymbol{x}_1$ and $\rho(\boldsymbol{x}_2)d\boldsymbol{x}_2$ for observing a point in $d\boldsymbol{x}_1$ and $d\boldsymbol{x}_2$, respectively. Another good reason is that g is invariant under random independent thinning of a point process; this is exemplified in Example 3 below for the particular case of a Cox process. Finally, $g = 1$ in the Poisson case.

Example 3 Consider again Example 2 with M given by (2), and let ρ and g denote the intensity and pair correlation functions of X. Then X_{th} is a Cox process with random intensity surface $Z_{\mathrm{th}}(\boldsymbol{x}) = \Pi(\boldsymbol{x})Z(\boldsymbol{x})$, intensity function $\rho_{\mathrm{th}}(\boldsymbol{x}) = \mathbb{E}\Pi(\boldsymbol{x})\rho(\boldsymbol{x})$, and pair correlation function $g_{\mathrm{th}} = g$.

The intensity and pair correlation functions can be estimated by nonparametric methods under rather general conditions, see Diggle (1985), Stoyan & Stoyan (1994, 2000), Baddeley, Møller & Waagepetersen (2000), Ohser & Mücklich (2000), and Møller & Waagepetersen (2001). When the pair correlation function exists (as it does for the Cox processes we consider), all we need to assume is that g is invariant under translations, i.e. $g(\boldsymbol{x}_1, \boldsymbol{x}_2) = g_0(\boldsymbol{x}_1 - \boldsymbol{x}_2)$. Then also a so-called *K-function* can be defined by

$$K(r) = \int_{\|\boldsymbol{x}\| \leq r} g_0(\boldsymbol{x})\mathrm{d}\boldsymbol{x}, \quad r \geq 0, \tag{3}$$

and it is easier to estimate K than g by nonparametric methods (Baddeley et al. 2000). For a stationary point process X, Ripley's K-function (Ripley 1976) agrees with K in (3), and $\rho K(r)$ can be interpreted as the expected number of further points within distance r from a typical point in X. Note that K and g_0 are in one-to-one correspondence if $g_0(\boldsymbol{x})$ depends only on the distance $\|\boldsymbol{x}\|$. Moreover, one often makes the transformation

$$L(r) = [K(r)/(\pi^{d/2}/\Gamma(1 + d/2))]^{1/d}$$

as $L(r) = r$ is the identity in the Poisson case.

Nonparametric estimators of ρ, g, K, and L are *summary statistics* of the first and second order properties of a spatial point process. These may be supplied with other summary statistics, including nonparametric estimators of the so-called F, G, and J functions which are based on interpoint distances for a stationary point process X. Briefly, for any $r > 0$, $1 - F(r)$ is the probability that X has no points within distance r from an arbitrary fixed point in $\mathbb{R}^d$, $1 - G(r)$ is the conditional probability that X has no further points within distance r from a typical point in X, and $J(r) = (1-G(r))/(1-F(r))$ (defined for $F(r) < 1$; see van Lieshout & Baddeley 1996). For a stationary Cox process,

$$1 - F(r) = \mathbb{E}\exp\left(- \int_{\|\boldsymbol{x}\| \le r} Z(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \right),$$

and

$$1 - G(r) = \mathbb{E}\left[\exp\left(- \int_{\|\boldsymbol{x}\| \le r} Z(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \right) Z(0) \right] \Big/ \rho.$$

In the special case of a homogeneous Poisson process,

$$1 - F(r) = 1 - G(r) = \exp\big(- \rho r^d \pi^{d/2} / \Gamma(1 + d/2) \big).$$

Further results are given in Section 5.1.

A plot of a summary statistic is often supplied with *envelopes* obtained by simulation of a point process under some specified model: Consider e.g. the L-function. Let $\hat{L}_0$ be a non-parametric estimator of L obtained from X observed within some window W, and let $\hat{L}_1, \ldots, \hat{L}_n$ be estimators obtained in the same way as $\hat{L}_0$ but from i.i.d. simulations $X_1, \ldots, X_n$ under the specified model for X. Then, for each distance r, we have that

$$\min_{1 \le i \le n} \hat{L}_i(r) \le \hat{L}_0 \le \max_{1 \le i \le n} \hat{L}_i(r) \tag{4}$$

with probability $(n - 1)/(n + 1)$ if X follows the specified model. We refer to the bounds in (4) as lower and upper envelopes. In our examples we choose $n = 39$ so that (4) specifies a 2.5% lower envelope and a 97.5% upper envelope.

The estimated L, g, and F functions for the weed data in Figure 2 are shown in Figure 3. The figure shows also the averages of these summary statistics and envelopes obtained from 39 simulations under a homogeneous Poisson process with expected number of points in W equal to the observed number of points. The plots for L and g clearly indicate aggregation, so the homogeneous Poisson process provides a poor fit. We do not consider

G due to problems with handling of edge effects in the special experimental design for the weed plants (see the discussion in Brix & Møller, 2001). The averages are close to the theoretical curves except for g where $\hat{g}(r)$ is biased upwards for small $r < 2.5$ cm (see the discussion p. 286 in Stoyan & Stoyan, 1994). The envelopes for g are rather wide for 25 cm $< r <$ 55 cm where few interpoint distances are observed. Similarly, the envelopes for L are wide for $r >$ 25 cm.

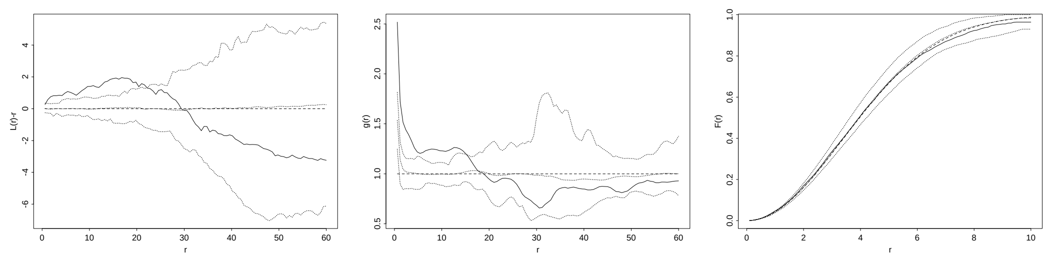

Figure 3: Summary statistics for positions of weed plants. Left: estimated $L(r) - r$ (solid line), envelopes, and average calculated from 39 simulations under the fitted homogeneous Poisson process (---), and theoretical value of $L(r) - r$ (– – –). Middle and right: as left but for g and F, respectively.

5 Parametric models of Cox processes

5.1 Neyman-Scott processes as Cox processes

A *Neyman-Scott process* (Neyman & Scott 1958) is a particular case of a *Poisson cluster process*. In this section we consider the case when Neyman-Scott processes become Cox processes.

Let C be a homogeneous Poisson process of cluster centers with intensity $\kappa > 0$. Assume that conditionally on $C = \{\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots\}$, the clusters $X_1, X_2, \ldots$ are independent Poisson processes, where the intensity measure of $X_i,\ i \geq 1$, is given by

$$\mu_i(A) = \int_A \alpha f(\boldsymbol{x} - \boldsymbol{c}_i) \mathrm{d}\boldsymbol{x} \tag{5}$$

where $\alpha > 0$ is a parameter and f is a density function for a continuous random variable in $\mathbb{R}^d$. Then $X = \cup_i X_i$ is a Neyman-Scott process. It is also a Cox process with random intensity surface

$$Z(\boldsymbol{x}) = \sum_{\boldsymbol{c} \in C} \alpha f(\boldsymbol{x} - \boldsymbol{c}). \tag{6}$$

The process (6) is stationary. It is also isotropic if $f(\boldsymbol{x})$ only depends on $\|\boldsymbol{x}\|$. The intensity and the pair correlation function are given by

$$\rho = \alpha\kappa, \quad g_0(\boldsymbol{x}) = 1 + h(\boldsymbol{x})/\kappa, \tag{7}$$

where

$$h(\boldsymbol{x}) = \int f(\boldsymbol{x}_0) f(\boldsymbol{x} + \boldsymbol{x}_0) \mathrm{d}\boldsymbol{x}_0$$

is the density for the difference between two independent points which each have density f. Furthermore,

$$J(r) = \int f(\boldsymbol{x}_1) \exp\Big(- \alpha \int_{\|\boldsymbol{x}_2\| \le r} f(\boldsymbol{x}_1 + \boldsymbol{x}_2) \mathrm{d}\boldsymbol{x}_2 \Big) \mathrm{d}\boldsymbol{x}_1,$$

see Bartlett (1975) and Lieshout, M. N. M. van & Baddeley (1996). Hence, $F \le G$ and J is nonincreasing with range $[\exp(-\alpha), 1]$.

Example 4 Closed form expressions for g_0 are known for a few Neyman-Scott models. A *Thomas process* (Thomas 1949) has

$$f(\boldsymbol{x}) = \big(2\pi\omega^2\big)^{-d/2} \exp\big(- \|\boldsymbol{x}\|^2 / \big(2\omega^2\big)\big), \tag{9}$$

the density for $N_d(0, \omega^2 I_d)$, i.e. for d independent normally distributed variables with mean 0 and variance $\omega^2 > 0$. This process is isotropic with

$$g_0(\boldsymbol{x}) = 1 + \big(4\pi\omega^2\big)^{-d/2} \exp\big(- \|\boldsymbol{x}\|^2 / \big(4\omega^2\big)\big)/\kappa. \tag{10}$$

The K-function can be expressed in terms of the cumulative distribution function for a χ^2-distribution with d degrees of freedom, and for $d = 2$ we simply obtain

$$K(r) = \pi r^2 + [1 - \exp(-r^2/(4\omega^2))]/\kappa. \tag{11}$$

The upper left plot in Figure 1 shows a simulation of a Thomas process with $\kappa = 10$, $\alpha = 10$, and $\omega^2 = 0.1$.

Another mathematically tractable model is a *Matérn cluster process* (Matérn 1960, Matérn 1986), where f is the uniform density on a d-dimensional ball with center at the origin; see Santaló (1976) and Stoyan et al. (1995).

We can obviously modify the definition of a Neyman-Scott process and the results above in many ways, e.g. by replacing α and $f(\boldsymbol{x}-\boldsymbol{c}_i)$ in (5) by $\alpha(\boldsymbol{c}_i)$ and $f_{\boldsymbol{c}_i}(\boldsymbol{x})$ (using an obvious notation), or by replacing C in (6) by an arbitrary Poisson process — we consider such extensions in Section 5.3. In either case we obtain a Cox process.

Simulation of a Neyman-Scott process follows in principle straightforwardly from either (6) and its definition as a Cox process or from its construction as a Poisson cluster process. However, boundary effects play a role: if we wish to simulate X within a bounded region W, we first simulate C within an extended region $B_{\text{ext}} \supset W$ so that points in X associated to cluster centers in $C \setminus B_{\text{ext}}$ fall in W with a negligible probability.

5.2 Log Gaussian Cox processes

If $Y = \log Z$ is a Gaussian process, i.e. if any finite linear combination $\sum a_i Y(\boldsymbol{x}_i)$ follows a normal distribution, 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). The definition of an LGCP can easily be extended in a natural way to multivariate LGCPs (Møller et al. 1998) and to multivariate spatio-temporal LGCPs (Brix & Møller 2001, Brix & Diggle 2001).

It is necessary to impose weak conditions on the mean and covariance functions

$$m(\boldsymbol{x}) = \mathbb{E}Y(\boldsymbol{x}) \quad \text{and} \quad c(\boldsymbol{x}_1, \boldsymbol{x}_2) = \mathbb{C}\mathrm{ov}(Y(\boldsymbol{x}_1), Y(\boldsymbol{x}_2))$$

in order to get a well-defined and finite integral $\int_B Z(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$ for bounded regions B. For example, we may require that $\boldsymbol{x} \to Y(\boldsymbol{x})$ is almost surely continuous. Weak conditions ensuring this, and which are satisfied for the models of m and c used in practice, are given in Theorem 3.4.1 in Adler (1981) (or see Møller et al. 1998).

Example 5 The covariance function belongs to the *power exponential family* if

$$c(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sigma^2 \exp\big(-\|(\boldsymbol{x}_1 - \boldsymbol{x}_2)/\alpha\|^\delta\big), \quad 0 \le \delta \le 2, \tag{12}$$

where $\alpha > 0$ is a scale parameter for the correlation and $\sigma^2 > 0$ is the variance. The special case $\delta = 1$ is an exponential covariance function, and $\delta = 2$ a Gaussian covariance function. If m is continuous, then $\boldsymbol{x} \to Y(\boldsymbol{x})$ is almost surely continuous. For the LGCP in the upper right plot in Figure 1, $(\delta, \alpha) = (1, 0.14)$. The lower right plot in Figure 1 is for the logistic process in Example 2 with $(\delta, \alpha) = (2, 0.10)$. In both plots the Gaussian process has mean zero and variance one.

Due to the richness of possible mean and covariance functions, LGCPs are flexible models for aggregation 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. In the specific examples of applications considered in this paper, we let m be constant and c an exponential covariance function. Brix & Møller (2001) and Møller & Waagepetersen (2001) consider situations where m is a linear or polynomial function, and Benes, Bodlak, Møller & Waagepetersen (2001) consider a case where m depends on covariates.

The intensity and pair correlation functions of an LGCP are given by

$$\rho(\boldsymbol{x}) = \exp(m(\boldsymbol{x}) + c(\boldsymbol{x}, \boldsymbol{x})/2) \quad \text{and} \quad g(\boldsymbol{x}_1, \boldsymbol{x}_2) = \exp(c(\boldsymbol{x}_1, \boldsymbol{x}_2)). \tag{13}$$

Hence there is a one-to-one correspondence between (m, c) and (ρ, g), and the distribution is determined by (ρ, g). This makes parametric models easy to interpret. Moreover, stationarity respective isotropy of an LGCP is equivalent to stationarity respective isotropy of (m, c) or equivalently of (ρ, g).

The simple relationship (13) indicates that LGCPs are more tractable for mathematical analysis than Neyman-Scott processes; further results are given in Møller et al. (1998).

We now turn to simulation of an LGCP. In contrast to the case of Neyman-Scott processes, we do not have problems with boundary effects since the distribution of an LGCP restricted to a bounded region B depends only on the distribution of $Y_B = \{Y(\boldsymbol{x}) : \boldsymbol{x} \in B\}$. As 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 $Y(\boldsymbol{c}_i)$ within disjoint cells C_i. Here $B = \cup_{i \in I} C_i$ is a subdivision with finite index set I, and $\boldsymbol{c}_i \in C_i$ is a "center" point of C_i. So we actually consider how to simulate the Gaussian vector $\tilde{Y} = (\tilde{Y}_i)_{i \in I}$ where $\tilde{Y}_i = Y(\boldsymbol{c}_i)$.

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, $d = 2$, and B is rectangular: Let $I \subset B$ denote a rectangular grid which is embedded in a rectangular grid I_{ext}, which is wrapped on a torus. Then a block circulant matrix $K = \{K_{ij}\}_{i,j \in I_{\text{ext}}}$ can be constructed so that $\{K_{ij}\}_{(i,j) \in I}$ is the covariance matrix of $\tilde{Y}$. Since K is block circulant, it can easily be diagonalized by means of the two-dimensional discrete Fourier transform with associated matrix F_2 (see Wood & Chan 1994 and Section 6.1 in Møller et al. 1998). Suppose that 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

$$\tilde{Y}_{\text{ext}} = \Gamma Q + \mu_{\text{ext}} \tag{14}$$

where Γ follows a standard multivariate normal distribution, $Q = \bar{F}_2 \Lambda^{1/2} F_2$, Λ is the diagonal matrix of eigenvalues for K, and the restriction of μ_{ext} to I agrees with the mean of $\tilde{Y}$. Using the two-dimensional fast Fourier transform a fast simulation algorithm for $\tilde{Y}_{\text{ext}}$ and hence $\tilde{Y}$ is obtained. We use this method for the simulations in Figure 1 and in connection with the MCMC algorithm considered in Section 7.2.

Another possibility is to use the Choleski decomposition of the covariance matrix of $\tilde{Y}$, provided this covariance matrix is positive definite. This may be advantageous if the covariance function is not translation invariant or B is far from being rectangular. On the other hand, the Choleski decomposition is only practically applicable if the dimension of $\tilde{Y}$ is moderate.

5.3 Shot noise G Cox processes

Brix (1999) introduces *shot noise G Cox processes (SNGCP)* by smoothing a so-called G measure which is driving a so-called G Cox process, see Remark 2 in Section 8. We instead define directly a SNGCP as a Cox process X with

$$Z(\boldsymbol{x}) = \sum_j \gamma_j k(\boldsymbol{x}, \boldsymbol{c}_j) \tag{15}$$

where the notation means the following. For each $\boldsymbol{c} \in \mathbb{R}^p$, $k(\cdot, \boldsymbol{c})$ is a density function for a continuous random variable. Further, $\{(\boldsymbol{c}_j, \gamma_j)\}$ is a Poisson process defined on $\mathbb{R}^p \times [0, \infty)$ by the intensity measure

$$\nu(A \times B) = (\kappa(A)/\Gamma(1-\alpha)) \times \int_B \gamma^{-\alpha-1} \exp(-\tau\gamma)\mathrm{d}\gamma, \quad A \subseteq \mathbb{R}^p,\ B \subseteq [0, \infty). \tag{16}$$

Here $\alpha < 1$ and $\tau \geq 0$ are parameters with $\tau > 0$ if $\alpha \leq 0$, and κ is a locally finite measure.

This definition can obviously be modified in many ways. For instance, if $\alpha = 0$ and we redefine ν by

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

where τ is now a positive measurable function, we obtain a Poisson/gamma process as studied in Wolpert & Ickstadt (1998). Extensions to multivariate SNGCPs are also obvious (Brix 1999). In the sequel X denotes a SNGCP with Z and ν given by (15) and (16).

Assume that $\kappa(\cdot) = \kappa_0|\cdot|$ is proportional to Lebesgue measure. Then $\{\boldsymbol{c}_j\}$ is stationary. Assume also that the kernels $k(\boldsymbol{x}, \boldsymbol{c}) = k_0(\boldsymbol{x} - \boldsymbol{c})$, $\boldsymbol{x}, \boldsymbol{c} \in \mathbb{R}^p$, are given by a common kernel k_0. Then X is stationary, and the intensity and pair correlation functions exist for $\tau \neq 0$ and are given by

$$\rho = \kappa_0 \tau^{\alpha-1}, \quad g_0(\boldsymbol{x}) = 1 + \frac{1-\alpha}{\kappa_0 \tau^{\alpha}} \int k_0(\boldsymbol{x}_0) k_0(\boldsymbol{x} + \boldsymbol{x}_0) \mathrm{d}\boldsymbol{x}_0. \qquad (17)$$

These are of the same form as for a Neyman-Scott process, cf. (7), so we cannot distinguish between the two types of models by considering a non-parametric estimate of (ρ, g_0) only. If furthermore $k_0(\boldsymbol{x})$ depends only on the distance $\|\boldsymbol{x}\|$, then X is isotropic.

Example 6 If k_0 is given by the normal density f in (9) with variance ω^2, X is isotropic, and g and K are of the same form as for a Thomas process (replacing κ in (10) and (11) by $\tau^\alpha \kappa_0/(1-\alpha)$). In the lower left plot in Figure 1, $\alpha = 0$, $\kappa_0 = 50$, $\tau = 0.5$ and $\omega^2 = 0.001$.

The marginal distributions of the independent processes $\{\boldsymbol{c}_j\}$ and $\{\gamma_j\}$ depend much on α as described below.

The case $\alpha < 0$: Then $\{\boldsymbol{c}_j\}$ is a Poisson process with intensity measure $(\tau^\alpha/|\alpha|)\kappa$, and the γ_j are mutually independent and follow a common Gamma distribution $\Gamma(|\alpha|, \tau)$. Hence, X is a kind of modified Neyman-Scott process. Particularly, X can be simulated within a bounded region in much the same way as we would simulate a Neyman-Scott process, cf. Section 5.1.

The case $0 \leq \alpha < 1$: The situation is now less simple as $\{\boldsymbol{c}_j\}$ is not locally finite. For $\alpha = 0$, we have a Poisson/gamma model (Daley & Vere-Jones 1988, Wolpert & Ickstadt 1998). For simplicity, suppose that κ is concentrated on a region D with $0 < \kappa(D) < \infty$, and write $\{(\boldsymbol{c}_j, \gamma_j)\} = \{(\boldsymbol{c}_1, \gamma_1), (\boldsymbol{c}_2, \gamma_2), \ldots\}$ so that $\gamma_1 > \gamma_2 > \ldots > 0$. Define $q(t) = \nu(D \times [t, \infty))$

for $t > 0$. Then q is a strictly decreasing function which maps $(0, \infty)$ onto $(0, \infty)$, and $q(\gamma_1) < q(\gamma_2) < \ldots$ are the points of a homogeneous Poisson process of intensity 1 and restricted to $(0, \infty)$. Furthermore, $\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots$ are independent, and each $\boldsymbol{c}_j$ follows the probability measure $\kappa(\cdot)/\kappa(D)$. For simulation and inference, one approximates Z by

$$Z(\boldsymbol{x}) \approx Z^J(\boldsymbol{x}) = \sum_{j=1}^{J} \gamma_j k(\boldsymbol{x}, \boldsymbol{c}_j) \tag{18}$$

where $J < \infty$ is a "cut off". Brix (1999) and Wolpert & Ickstadt (1998) discuss how q^{-1} and the tail sum $\sum_{j>J} \gamma_j$ can be evaluated.

6 Estimation for parametric models of Cox processes

For specificity we discuss first estimation in the context of a Thomas process X with unknown parameters $\kappa > 0$, $\alpha > 0$, and $\sigma > 0$. We next turn to LGCPs and SNGCPs in Examples 8 and 9.

In most applications the process of cluster centers C is treated as missing data, and we have only observed $X \cap W = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, where W is a bounded observation window. Let $B_{\mathrm{ext}} \supset W$ be specified as in the end of Section 5.1, and redefine the random intensity surface (6) so that the sum over cluster centers C is replaced by a sum over $C \cap B_{\mathrm{ext}}$.

The *likelihood function* for $\theta = (\kappa, \alpha, \omega)$ is

$$L(\theta) = \mathbb{E}_\kappa \left[\exp\left(- \int_W Z(\boldsymbol{x}; C \cap B_{\mathrm{ext}}, \alpha, \omega) \mathrm{d}\boldsymbol{x} \right) \prod_{j=1}^{n} Z(\boldsymbol{x}_j; C \cap B_{\mathrm{ext}}, \alpha, \omega) \right] \tag{19}$$

where the mean is with respect to the Poisson process $C \cap B_{\mathrm{ext}}$, and

$$Z(\boldsymbol{x}; C \cap B_{\mathrm{ext}}, \alpha, \omega) = (\alpha/\omega^2) \sum_{\boldsymbol{c} \in C \cap B_{\mathrm{ext}}} \varphi((\boldsymbol{x} - \boldsymbol{c})/\omega)$$

where φ denotes the density for the d-dimensional standard normal distribution. If W is rectangular (or a disjoint union of rectangular sets as in Figure 2), the integral in (19) can be expressed in terms of the cumulative distribution function for $N(0,1)$. The likelihood (19) can further be estimated by *Markov chain Monte Carlo* (MCMC) methods: For a given value

$\theta_0 = (\kappa_0, \alpha_0, \omega_0)$ of θ, suppose $C^1 = \{\boldsymbol{c}_1^1, \ldots, \boldsymbol{c}_{m_1}^1\}, \ldots, C^k = \{\boldsymbol{c}_1^k, \ldots, \boldsymbol{c}_{m_k}^k\}$ is a Markov chain sample from the conditional distribution $C \cap B_{\mathrm{ext}} | X \cap W = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ (Section 7.1 describes how such a sample can be generated). The Monte Carlo approximation of $L(\theta)/L(\theta_0)$ is

$$\frac{1}{k}\sum_{i=1}^{k} \frac{\kappa^{m_i} \exp\big(-\int_W (\kappa + Z(\boldsymbol{x}; C^i, \alpha, \omega))\mathrm{d}\boldsymbol{x}\big) \prod_{j=1}^n Z(\boldsymbol{x}_j; C^i, \alpha, \omega)}{\kappa_0^{m_i} \exp\big(-\int_W (\kappa_0 + Z(\boldsymbol{x}; C^i, \alpha_0, \omega_0))\mathrm{d}\boldsymbol{x}\big) \prod_{j=1}^n Z(\boldsymbol{x}_j; C^i, \alpha_0, \omega_0)}.$$

Note that the approximation is based on importance sampling, so it only works for θ sufficiently close to θ_0. The generation and storing of $C^1, \ldots, C^k$ is further computationally rather demanding; see also Remark 1 in Section 8.

A computationally easier alternative for parameter estimation is *minimum contrast estimation* (Diggle 1983, Møller et al. 1998): The closed form expressions (10) and (11) may be compared with the nonparametric summary statistics $\hat{g}$ and $\hat{K}$ obtained assuming only stationarity and isotropy of X (Section 4). If for example $d = 2$ and the K-function is used, a minimum contrast estimate $(\hat{\kappa}, \hat{\omega})$ is chosen to minimize

$$\int_{a_1}^{a_2} \big\{\big(\hat{K}(r) - \pi r^2\big) - \big(\big[1 - \exp(-r^2/(4\omega^2))\big]/\kappa\big)\big\}^2 \mathrm{d}r \qquad (20)$$

where $0 \le a_1 < a_2$ are user specified parameters (see Example 7 below). Setting

$$A\big(\omega^2\big) = \int_{a_1}^{a_2} \big[1 - \exp(-r^2/(4\omega^2))\big]^2 \mathrm{d}r$$

and

$$B\big(\omega^2\big) = \int_{a_1}^{a_2} \big[(\hat{K}(r) - \pi r^2)(1 - \exp(-r^2/(4\omega^2)))\big] \mathrm{d}r,$$

then

$$\hat{\omega}^2 = \arg\max \big[B\big(\omega^2\big)^2 / A\big(\omega^2\big)\big], \quad \hat{\kappa} = \big[A\big(\hat{\omega}^2\big)/B\big(\hat{\omega}^2\big)\big].$$

Inserting this into the equation $\rho = \alpha\kappa$ and using the natural estimate $\hat{\rho} = n/|W|$ where $|W|$ denotes the area of W, we obtain finally the estimate

$$\hat{\alpha} = \hat{\rho}/\hat{\kappa}.$$

Diggle (1983) suggests the alternative contrast

$$\int_{a_1}^{a_2} \big\{\big(\hat{K}(r)\big)^b - \big(\pi r^2 + \big[1 - \exp(-r^2/(4\omega^2))\big]/\kappa\big)^b\big\}^2 \mathrm{d}r$$

where $b > 0$ is a third user-specified parameter recommended to be between 0.25 and 0.5. This approach requires numerical minimization with respect to κ and ω jointly. Brix (1999) reports that minimum contrast estimation for SNGCPs is numerically more stable if g is used instead of K. Minimum contrast estimation for LGCPs and space-time LGCPs is considered in Møller et al. (1998), Brix & Møller (2001), and Brix & Diggle (2001); see also Example 8 below.

Example 7 (Thomas) For the weed data and certain choices of a_1 and a_2, the contrast (20) did not have a well-defined minimum. Numerically more stable results were obtained by using the g-function given by (10), i.e. by considering the contrast obtained by replacing $\hat{K}(r) - \pi r^2$ and $\left[1 - \exp(-r^2/(4\omega^2))\right]/\kappa$ in (20) by $\hat{g}(r) - 1$ and $\exp(-r^2/(4\omega^2)/(4\pi\omega^2\kappa)$, respectively. The middle plot in Figure 3 suggests to use $a_1 = 2.5$ cm and $a_2 = 25$ cm due to the bias of $\hat{g}(r)$ for $r < 2.5$ cm and the large variability of $\hat{g}(r)$ for $r > 25$ cm. With these values of a_1 and a_2 the estimates $\hat{\kappa}$=0.005, $\hat{\omega}^2 = 51.0$, and $\hat{\alpha} = 3.05$ are obtained.

Example 8 (LGCP) Turning next to a stationary LGCP, the mean function m is equal to a real constant ξ, say. Let $c = c_{\alpha,\sigma}$ be the exponential covariance function with positive parameters (α, σ) as in (12) with $\delta = 1$. For similar reasons as for the Thomas process, likelihood estimation of $\theta = (\xi, \alpha, \sigma^2)$ is computationally demanding, and minimum contrast estimation is a simpler method. Because of the simple relationship (13), the minimum contrast estimate $(\hat{\alpha}, \hat{\sigma}^2)$ is chosen to minimize

$$\int_{a_1}^{a_2} \left\{\hat{g}(r)^b - \left(\sigma^2 \exp(-r/\alpha)\right)^b\right\}^2 \mathrm{d}r$$

where $b > 0$. As for the Thomas process we can easily find $(\hat{\alpha}, \hat{\sigma}^2)$ and combine this with (13) and the estimate for the intensity to obtain the equation $n/|W| = \exp(\hat{\xi} + \hat{\sigma}^2/2)$ for the estimate of ξ. Using the same values of a_1 and a_2 as in Example 5 and letting $b = 1$, we obtain the minimum contrast estimates $\hat{\xi} = -4.5$, $\hat{\sigma}^2 = 0.45$, and $\hat{\alpha} = 9.5$.

Example 9 (SNGCP) Consider the SNGCP from Example 6. Using minimum contrast estimation as for the Thomas process we obtain $\hat{\kappa}_0$=0.005, $\hat{\omega}^2 = 51.0$, and $\hat{\tau} = 0.33$.

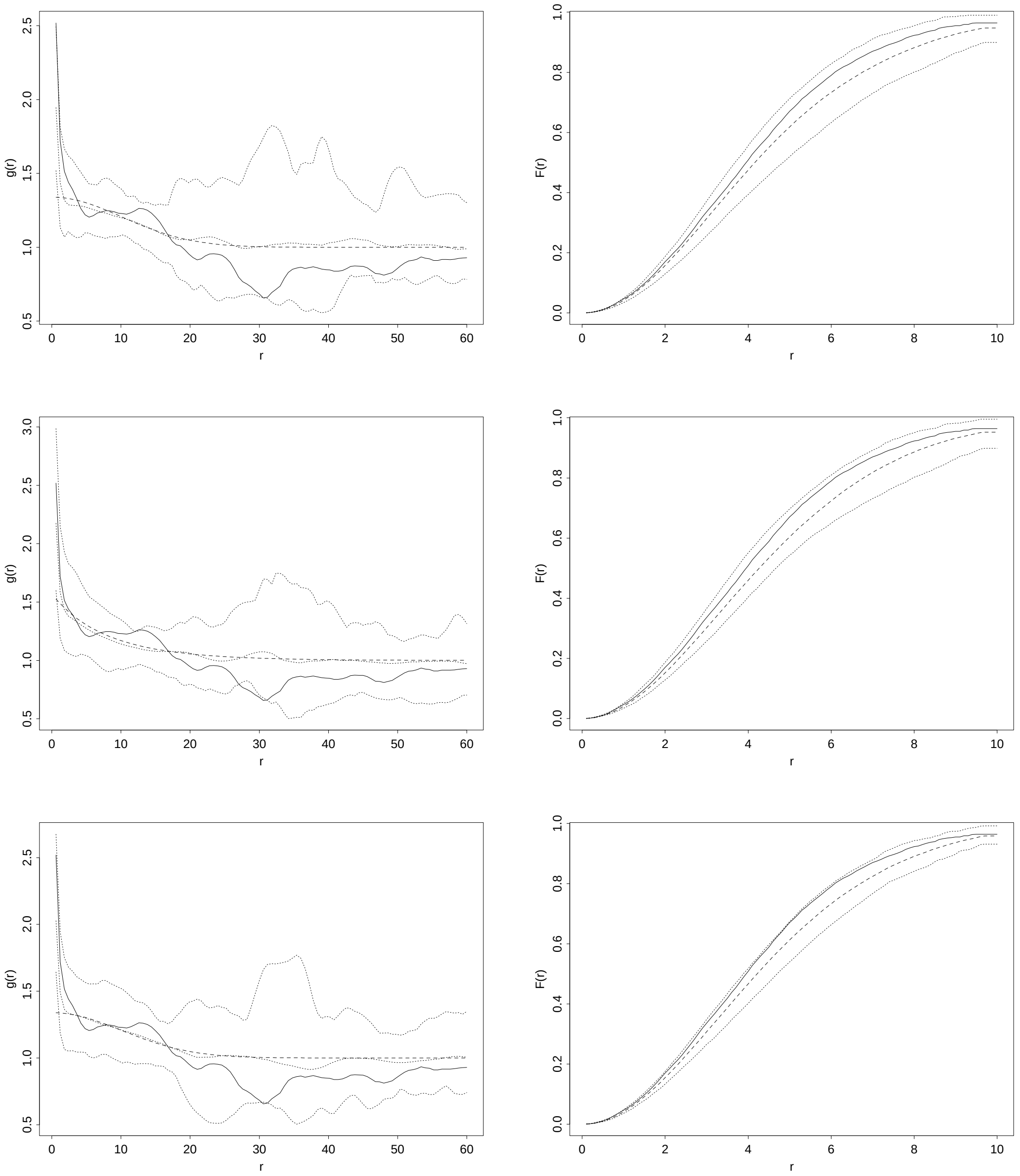

Figure 4: Upper left: $\hat{g}(r)$ for the weed data (solid line), envelopes and average calculated from 39 simulations under the fitted Thomas process (---), and theoretical value of $g(r)$ for fitted Thomas process ($------$). Upper right: estimated $F(r)$ (solid line), and envelopes and average calculated from 39 simulations under the fitted Thomas process (- - -). Middle plots: as upper plots but for LGCP (Example 8). Lower plots: as upper plots but for SNGCP (Example 9).

A comparison of the non-parametric estimates $\hat{g}$ and $\hat{F}$ with the envelopes calculated under the various fitted models in Examples 7–9 (see Figure 4) does not reveal obvious inconsistencies between the data and any of the fitted models, except perhaps for the SNGCP where $\hat{F}(r)$ coincide with the upper envelope for a range of distances r. The bias of $\hat{g}$ near zero makes it difficult to make inference about the behaviour of the pair correlation near zero. For a LGCP it is for example difficult to infer whether an exponential or a "Gaussian" covariance function should be used for the Gaussian field Y. See also Remark 1, Section 8.

7 Prediction

As in Section 6 suppose that a realization $x = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ of $X \cap W$ is observed, where W is a bounded observation window. Given a bounded region B, one may be interested in predicting $Z_B = \{Z(\boldsymbol{x})\}_{\boldsymbol{x} \in B}$ or $X \cap (B \setminus W)$ or, more generally, in computing the conditional distributions of Z_B and $X \cap (B \setminus W)$ given $X \cap W = x$. In the sequel we mainly focus on the conditional distribution of Z_B, since $X \cap (B \setminus W)$ is conditionally independent of $X \cap B$ given Z_B, and $X \cap (B \setminus W)$ is simply an inhomogeneous Poisson process given Z_B. The conditional distribution of Z_B is in general not analytically tractable, so MCMC methods become useful for computing characteristics of the conditional distribution. In the following we discuss MCMC algorithms for the model classes considered in Section 5. For background knowledge on MCMC algorithms (particularly Metropolis-Hastings algorithms), see e.g. Gilks, Richardson & Spiegelhalter (1996) and Robert & Casella (1999).

7.1 Conditional simulation for Neyman-Scott processes

For a Neyman-Scott process (Section 5.1), Z_B is determined by the process of cluster centers. We approximate this by the process of cluster centers falling in a sufficiently large region B_{ext} which contains B, and ignore cluster centers outside B_{ext}. The conditional distribution of the cluster centers on B_{ext} then has a density with respect to the homogeneous Poisson process of intensity 1 and restricted to B_{ext}. The density is given by

$$p(c|x) \propto \kappa^{\text{card}(c)} \exp\left(- \int_W Z(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \right) \prod_{\boldsymbol{c} \in c} Z(\boldsymbol{c})$$

for finite point configurations $c \subset B_{\text{ext}}$, where $Z(\boldsymbol{x})$ is given by (6) but with C replaced by c. The conditional distribution can be simulated using the MCMC algorithm studied in Geyer & Møller (1994). Prediction and fully Bayesian inference for Neyman-Scott processes viewed as cluster processes has been considered by e.g. Lawson (1993), Baddeley & van Lieshout (1993), and Van Lieshout & Baddeley (1995).

7.2 Conditional simulation for LGCPs

Consider now an LGCP with $Y_B = \log Z_B$ where we use a notation as in Section 5.2. 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 (14). Omitting an additive constant depending on x only, the log conditional density of Γ given x is

$$-\|\gamma\|^2/2 + \sum_{i\in I}(\tilde{y}_i n_i - A_i \exp(\tilde{y}_i)) \tag{21}$$

where, in accordance with (14), $(\tilde{y}_i)_{i\in I_{\text{ext}}} = \gamma Q + \mu_{\text{ext}}$, $n_i = \text{card}(x \cap C_i)$, and $A_i = |C_i|$ if $C_i \subset W$ and $A_i = 0$ otherwise. The gradient of (21) 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 and Tweedie 1996) and earlier in the physics literature by Rossky, Doll & Friedman (1978). This is a Metropolis-Hastings algorithm with proposal distribution $N_d(\gamma + (h/2)\nabla(\gamma), hI_d)$ where $h > 0$ is a user-specified proposal variance. 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 (2000) and Christensen & Waagepetersen (2001).

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

$$\nabla_{\text{trun}}(\gamma) = -\gamma + \big(n_i - \min\{H, A_i \exp(\tilde{y}_i)\}\big)_{i\in I_{\text{ext}}} Q^{\mathsf{T}} \tag{22}$$

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

Benes et al. (2001) consider a fully Bayesian approach for a LGCP where the truncated Langevin-Hastings algorithm is extended with updates of the model parameters. Benes et al. (2001) also discuss convergence of the posterior for the discretized LGCP when the cell sizes tends to zero.

Example 10 Let W be the union of the five frames in the lower right corner in Figure 2, let B be the smallest rectangle containing these five frames, and define the discretized Gaussian field $\tilde{Y}$ on a 60×40 rectangular grid I on B. The left plot in Figure 5 shows a prediction of $\tilde{Z} = \exp(\tilde{Y})$ given by the conditional mean $\mathbb{E}(\tilde{Z}|x)$ of $\tilde{Z}$ using the parameter estimates from Example 8. The minimum and maximal values of $\mathbb{E}(\tilde{Z}|x)$ are 0.01 and 0.03, respectively.

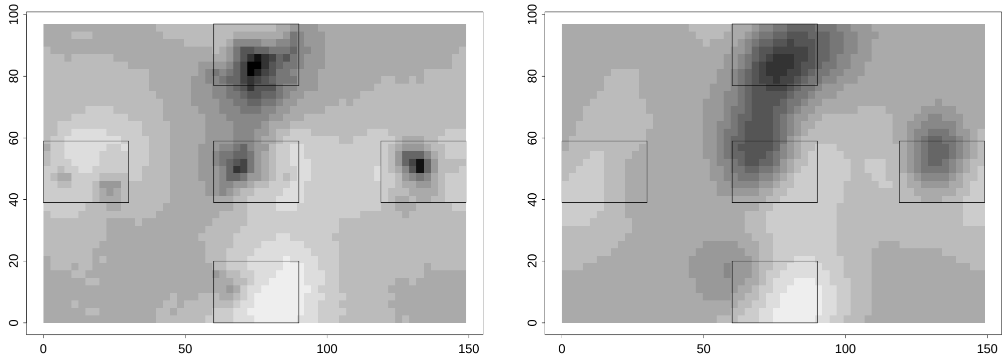

Figure 5: Left: conditional mean of $\tilde{Z}$ under an LGCP model. Right: conditional mean of Z_B^J under a SNGCP model. The same gray scales are used in the two plots.

7.3 Conditional simulation for shot-noise G Cox processes

Conditional simulation for a SNGCP with $\alpha < 0$ follows much the same way as in Section 7.1, so we let $0 \leq \alpha < 1$ in the sequel. In applications involving MCMC we consider the approximation (18) where D is typically equal to an extended region $B_{\text{ext}} \supset B$ as in Section 7.1. Let $\psi_j = q(\gamma_j)$, $j = 1, \ldots, J$. By (18), conditional simulation of Z^J is given by conditional simulation of the vector $(\boldsymbol{c}_1, \psi_1, \ldots, \boldsymbol{c}_J, \psi_J)$. Assuming that the measure κ has a density k_κ with respect to Lebesgue measure, $(\boldsymbol{c}_1, \psi_1, \ldots, \boldsymbol{c}_J, \psi_J)$ has density proportional to $\exp(-\psi_J) \prod_{i=1}^J k_\kappa(\boldsymbol{c}_i)$ for $(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_J) \in D^J$ and $0 < \psi_1 < \ldots < \psi_J$. Note that the conditional density

$$p(x|\boldsymbol{c}_1, \psi_1, \ldots, \boldsymbol{c}_J, \psi_J) \propto \exp\left(- \int_W Z^J(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \right) \prod_{j=1}^n Z^J(x_j)$$

does not depend on the order of $(\boldsymbol{c}_j, \psi_j)$, $j = 1, \ldots, J$. So we can forget the constraint $0 < \psi_1 < \ldots < \psi_J$, and simply consider the posterior obtained with the prior density proportional to $\exp(-\psi_{\max}) \prod_{i=1}^J k_\kappa(\boldsymbol{c}_i)$ where $\psi_{\max} = \max\{\psi_1, \ldots, \psi_J\}$. The density of $(\boldsymbol{c}_1, \psi_1, \ldots, \boldsymbol{c}_J, \psi_J)$ given x is then proportional to

$$\left[\exp\left(- \int_W Z^J(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \right) \prod_{j=1}^n Z^J(x_j) \right] \left[\exp(-\psi_{\max}) \prod_{i=1}^J k_\kappa(\boldsymbol{c}_i) \right].$$

For the special case of shot-noise Poisson-gamma models, Wolpert & Ickstadt (1998) use a certain data augmentation scheme and a Gibbs sampler, but we consider a more simple approach where we just apply a standard random scan single-site Metropolis algorithm. A proposal for our Metropolis algorithm is obtained by picking a j in $\{1, \ldots, J\}$ uniformly at random, and then replacing $(\boldsymbol{c}_j, \psi_j)$ by $(\boldsymbol{c}_j', \psi_j')$ generated from the uniform distribution on $B_{\text{ext}} \times]c_j - s_c, c_j + s_c[$, where $s_c > 0$ is a user-specified parameter. In the implementation of the algorithm it is helpful to introduce an auxiliary variable U which takes the value j if $\psi_{\max} = \psi_j$, $j = 1, \ldots, J$, and thus keeps track of the maximal ψ_j. The variable U can further be used to monitor convergence for the Markov chain, since U follows the uniform distribution on $\{1, \ldots, J\}$ when the chain is in equilibrium.

The algorithm can straightforwardly be extended to accomodate a fully Bayesian analysis with priors also on the model parameters. Fully Bayesian analysis of Poisson-gamma shot-noise processes is considered in e.g. Wolpert & Ickstadt (1998) and Best, Ickstadt & Wolpert (2000).

Example 11 Let W and $B = [0, 150] \times [0, 100]$ be as in Example 10, let $B_{\text{ext}} = [-40, 190] \times [-40, 140]$, and consider the fitted SNGCP from Example 9. The right plot in Figure 5 shows the posterior mean of $(Z^J(\boldsymbol{x}))_{\boldsymbol{x}\in B}$ obtained with $J = 2000$, using a sample obtained by subsampling each 10^4 state of a Markov chain of length 10^7 generated by the Metropolis algorithm. The minimal and maximal values of $\mathbb{E}(Z^J(\boldsymbol{x})|x)$, $\boldsymbol{x} \in B$, are the same as in Example 10, but the predicted intensity surface is more smooth under the SNGCP.

8 Discussion

Remark 1 (Likelihood-based inference) Concerning parameter estimation we have focused on minimum contrast estimation which is advantageous for computational reasons. This is somewhat out of line with modern spatial statistics where likelihood-based methods (either in a Bayesian or frequentist framework) attracts increasing attention. In general one must expect minimum contrast estimation to be less efficient than likelihood-based estimation. The minimum contrast estimates can also be very sensitive to the choice of user-specified parameters.

If we consider LGCPs and SNGCPs with $0 \leq \alpha < 1$, then two problems appear in connection with likelihood-based inference. First, in order to apply MCMC methods, we need to approximate the processes either by discretizing the Gaussian field for a LGCP or by using the truncated sum (18). Second, in order to get a good approximation, we need to use either a fine grid I or a large truncation value J. Conditional simulation for the approximate processes thereby becomes computationally very demanding. Also the storage of samples for Monte Carlo estimation of the likelihood may be problematic. For Neyman-Scott processes and SNGCPs with $\alpha < 0$ and a moderate value of κ, the conditional distribution will in contrast typically be of moderate dimension, and there is no need to introduce an approximation (apart from possible edge effects when the kernel density f does not have bounded support).

Remark 2 (Definition of a SNGCP) In Section 5.3 we gave a direct definition of a SNGCP. Below we briefly discuss how such processes have been introduced in Brix (1999).

Brix (1999) defines first a G measure M. This is a random measure so that $M(A_1), \ldots, M(A_n)$ are independent for disjoint bounded regions $A_1, \ldots, A_n$, and each $M(A_i)$ follows a so-called G distribution which is parameterized in a certain way by $(\alpha, \kappa(A_i), \tau)$. It turns out that a G measure is a locally finite measure of the form $M(A) = \sum_j \gamma_j \mathbf{1}\{\boldsymbol{c}_j \in A\}$ where $\{(\boldsymbol{c}_j, \gamma_j)\}$ is defined as in Section 5.3. Secondly, since M is purely atomic, Brix views a G Cox process as a random measure Φ: conditionally on M, we have that Φ is a random measure so that $\Phi(A_1), \ldots, \Phi(A_n)$ are independent for disjoint regions $A_1, \ldots, A_n$, and each $\Phi(A_i)$ is Poisson distributed with mean $M(A_i)$. Finally, in order to obtain some smoothing, Brix defines a SNGCP by extending the definition of M to $M(A) = \sum_j \gamma_j K(A, \boldsymbol{c}_j)$ where each $K(\cdot, \boldsymbol{c}_j)$ is an arbitrary probability measure. When $K(\cdot, \boldsymbol{c}_j)$ is given by a density function $k(\cdot, \boldsymbol{c}_j)$ as in Section 5.3, we obtain that Φ is now a nonatomic measure, which can be identified with our SNGCP X given by (15).

G Cox processes have nice properties under aggregation due to the simple form of the distribution of aggregated counts $\Phi(A_1), \ldots, \Phi(A_n)$. For example under a Poisson-gamma model (Wolpert & Ickstadt 1998), the counts are negative binomial distributed. However, one should note that these simple properties are not valid for SNGCPs, due to the smoothing by the kernel $k(\cdot, \cdot)$. See also the discussion in Richardson (2001) and Møller (2001).

Remark 3 (Extension of SNGCPs) One may note that both SNGCPs and Neyman-Scott processes are special cases of Cox processes driven by a random intensity surface of the form

$$Z(\boldsymbol{x}) = \sum_j \lambda_j k(x, \boldsymbol{c}_j) \tag{23}$$

where $\{(\boldsymbol{c}_j, \lambda_j)\}$ is a Poisson process with an intensity measure of the form $\nu(\mathrm{d}(\boldsymbol{c}, \lambda)) = \kappa(\mathrm{d}\boldsymbol{c})\zeta(\mathrm{d}\lambda)$ (i.e. a product measure). Here ν is assumed to be locally finite, but the marginal process $\{\boldsymbol{c}_j\}$ may not be locally finite, since it is possible that $\zeta([0, \infty)) = \infty$, cf. the case of a SNGCP with $0 \leq \alpha \leq 1$.

Many properties for Neyman-Scott processes and SNGCPs follow from general results for the extended model. For example, if $k(\boldsymbol{x}, \boldsymbol{c}) = k_0(\boldsymbol{x} - \boldsymbol{c})$ and $\kappa(\cdot) = \kappa_0 |\cdot|$ where $\kappa_0 > 0$ is a parameter, the process is stationary with intensity $\rho = \kappa_0 \int \lambda \zeta(\mathrm{d}\lambda)$. Assuming $\int \lambda^2 \zeta(\mathrm{d}\lambda) < \infty$, the pair correlation exists and is given by

$$g_0(\boldsymbol{x}) = 1 + (\kappa_0/\rho)^2 \int \lambda^2 \zeta(\mathrm{d}\lambda) \int k_0(\boldsymbol{x}_0) k_0(\boldsymbol{x} + \boldsymbol{x}_0) \mathrm{d}\boldsymbol{x}_0.$$

Equations (7) and (17) are special cases of this result.

As in Møller (2001) we suggest that more attention should be drawn to "modified" Neyman-Scott models of the type (23) with $\{\boldsymbol{c}_j\}$ a locally finite

Poisson process. One can thereby utilize the additional flexibility offered by the random coefficients λ_j and still work with a locally finite conditional distribution in connection with Bayesian inference or maximum likelihood estimation.

Remark 4 (Comparison of models) The focus in this paper has been on inferential and computational aspects of various parametric models for Cox processes. The advantages and disadvantages of LGCPs, SNGCPs, and the Heikkinen & Arjas (1998) model for inhomogeneous Poisson processes are also discussed in Richardson (2001) and Møller (2001). Which model is most appropriate depends of course on prior knowledge and the purpose of the statistical analysis. LGCPs provide easily interpretable Cox process models with some flexibility in modelling aggregated point patterns when the aggregation is due to an environmental heterogeneity. Even more flexibility may be obtained by using SNGCPs and their extensions (Remark 3), and such models seem natural when the aggregation is due to clustering around a process of cluster centers. They seem also appropriate for nonparametric Bayesian modelling (Wolpert & Ickstadt 1998, Richardson 2001, Møller 2001).

References

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

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

Bartlett, M. S. (1975). *The Statistical Analysis of Spatial Pattern*, Chapman and Hall, London.

Benes, V., Bodlak, K., Møller, J. & Waagepetersen, R. P. (2001). Bayesian analysis of log Gaussian Cox process models for disease mapping. In preparation.

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

Best, N. G., Ickstadt, K. & Wolpert, R. (2000). Spatial Poisson regression for health and exposure data measured at disparate resolutions, *Journal of the American Statistical Association* **95**: 1076–1088.

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. & Diggle, P. J. (2001). Spatio-temporal prediction for log-Gaussian Cox processes, *Journal of the Royal Statistical Society* **B 63**. To appear.

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.

Christensen, O. F., Møller, J. & Waagepetersen, R. P. (2000). Geometric ergodicity of Metropolis-Hastings algorithms for conditional simulation in generalised linear mixed models, *Technical Report R-00-2010*, Department of Mathematical Sciences, Aalborg University. *Methodology and Computation in Applied Probability.* To appear.

Christensen, O. F. & Waagepetersen, R. (2001). Bayesian prediction of spatial count data using generalised linear mixed models. Submitted.

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

Cox, D. R. (1955). Some statistical models related with series of events, *Journal of the Royal Statistical Society* **B 17**: 129–164.

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.

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

Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London.

Grandell, J. (1976). *Doubly Stochastic Poisson Processes*, Springer-Verlag, Berlin.

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

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

Lawson, A. B. (1993). Discussion contribution, *Journal of the Royal Statistical Society* **B 55**: 61–62.

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

Lieshout, M. N. M. van & Baddeley, A. J. (1995). Markov chain Monte Carlo methods for clustering of image features, *Proceedings of the 5th IEE International Conference on Image Processing and its Applications*, Vol. 410 of *IEE Conference Publication*, IEE Press, London, pp. 241–245.

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

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

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

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

Møller, J. & Waagepetersen, R. P. (2001). Simulation based inference for spatial point processes, *in* M. B. Hansen & J. Møller (eds), *Spatial Statistics and Computational Methods*, Lecture Notes in Statistics, Springer-Verlag. To appear.

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

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

Richardson, S. (2001). Spatial models in epidemiological applications, *in* P. J. Green, N. L. Hjort & S. Richardson (eds), *Highly Structured Stochastic Systems*, Oxford University Press, Oxford. To appear.

Ripley, B. D. (1976). The second-order analysis of stationary point processes, *Journal of Applied Probability* **13**: 255–266.

Robert, C. P. & Casella, G. (1999). *Monte Carlo statistical methods*, Springer-Verlag, New York.

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.

Santaló, L. (1976). *Integral Geometry and Geometric Probability*, Addison–Wesley, Reading, MA.

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.

Thomas, M. (1949). A generalization of Poisson's binomial limit for use in ecology, *Biometrika* **36**: 18–25.

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.