Non- and semi-parametric estimation of interaction in inhomogeneous point patterns

Adrian Baddeley* Jesper Møller† Rasmus Waagepetersen ‡

Abstract

We develop methods for analysing the 'interaction' or dependence between points in a spatial point pattern, when the pattern is spatially inhomogeneous. Completely non-parametric study of interactions is possible using an analogue of the K-function. Alternatively one may assume a semi-parametric model in which a (parametrically specified) homogeneous Markov point process is subjected to (non-parametric) inhomogeneous independent thinning. The effectiveness of these approaches is tested on datasets representing the positions of trees in forests.

Keywords: Kernel estimation; Markov chain Monte Carlo; Markov point process; Second-order methods; Semi-parametric inference; Spatial statistics; Summary statistics; Thinning.

1 Introduction

In the analysis of spatial point patterns, one of the main objectives is often to assess the strength of 'interaction' or dependence between the points. Methods for investigating interpoint interaction include non-parametric summaries such as the second moment function K (Ripley, 1977, Diggle, 1983) and explicit parametric models such as Markov point processes (Ripley and Kelly, 1977, Baddeley and Møller, 1989). However, the existing non-parametric summaries rely on the assumption that the point pattern is spatially homogeneous (stationary and/or isotropic). If an initial analysis suggests that the pattern has non-constant intensity, then current non-parametric techniques do not allow further analysis. The Markov point process literature also focuses mainly on stationary processes, with the exception of the work of Ogata and Tanemura (1986). Considerable attention is

*Department of Mathematics & Statistics, University of Western Australia, Nedlands WA 6907, Australia.

†Department of Mathematics, University of Aalborg, Fredrik Bajersvej 7E, DK-9220 Aalborg.

‡Biometry Research Unit, Department of Agricultural Systems, Danish Institute of Agricultural Sciences, P.O. Box 50, DK-8830 Tjele, email: `Rasmus.Waagepetersen@agrsci.dk` (address for correspondence).

now being given to inhomogeneous Markov and Gibbs processes (Baddeley and Turner, 2000, Nielsen and Jensen, 1999, Stoyan and Stoyan, 1998). As such models are analytically intractable, Markov chain Monte Carlo (MCMC) methods are indispensable for likelihood-based statistical inference (Geyer and Møller, 1994, Geyer, 1999, Møller, 1999).

Inhomogeneous point patterns may arise, for example, from observations of the positions of plants where the soil fertility exhibits spatial variation. If we assume that fertility is a random spatially varying field, but that conditional on the soil fertility (and possibly other environmental factors) the plant locations are independent, the appropriate model is a Cox process (Diggle, 1983, Stoyan *et al.*, 1995, Møller *et al.*, 1998, Brix and Møller, 1998). On the other hand, if there is dependence between plants (such as competition), it seems natural to modify the Cox model by adding interaction terms (Møller *et al.*, 1998); however, MCMC methods for such extended Cox process models are computationally very demanding.

In this paper we take a simpler approach, extending the second-order analysis of stationary processes (based on the K-function) to non-stationary processes. Firstly in section 2 we develop a non-stationary analogue K_{inhom} of the K-function. We propose a non-parametric estimator of K_{inhom} and demonstrate its usefulness on two standard point pattern datasets. It is also shown that K_{inhom} has both properties and weaknesses similar to those of K.

A key observation is that for stationary processes the K-function is invariant under independent thinning (independent random deletion/retention of each point of the process with a fixed retention probability p). A similar property holds for our K_{inhom} without the requirement of constant p. This suggests an approach which we explore in section 3. We construct a new model for inhomogeneous point patterns by taking a stationary Markov process and applying independent random thinning by an inhomogeneous thinning surface. We study semi-parametric inference for this model. Finally, in section 4 we discuss some open problems.

2 Non-parametric second-order analysis of inhomogeneous point patterns

Throughout this paper we restrict attention to planar point patterns (i.e. locally finite subsets of $\mathbb{R}^2$) observed through some bounded observation window $W \subset \mathbb{R}^2$ with area $|W| > 0$. Standard practice in the analysis of point patterns is first to form a non-parametric estimate of the first-order intensity $\lambda(s)$, $s \in \mathbb{R}^2$; then if this is deemed to be uniform, $\lambda(s) \equiv \lambda$, to investigate interpoint interactions by estimating various summary statistics such as the K-function or the empty space distribution function. These statistics are compared to their expected values for the homogeneous Poisson point process which serves as the null hypothesis of complete spatial randomness (i.e. absence of interaction), see e.g. Ripley (1977) and Diggle (1983).

Here we generalise this approach to inhomogeneous point patterns, using an extension of the K-function, defined in section 2.1.

We shall sometimes use Palm distributions (see Stoyan *et al.*, 1995, p. 119-121). Briefly, for a finite point configuration $y = \{y_1, \ldots, y_n\} \subset \mathbb{R}^2$ and a planar point process Y, the Palm distribution P_y of Y at y can be interpreted as the conditional distribution of Y given that $y \subset Y$. We let $\mathbb{E}_y$ denote expectation with respect to P_y, and set $P_s = P_{\{s\}}$ and $\mathbb{E}_s = \mathbb{E}_{\{s\}}$ for $s \in \mathbb{R}^2$.

2.1 K-function for inhomogeneous point patterns

In order to define the inhomogeneous K-function we need a few preliminaries. Assume that Y has a first-order intensity function $\lambda(s)$, $s \in \mathbb{R}^2$, i.e.

$$\mathbb{E}\,\mathrm{card}(Y \cap B) = \int_B \lambda(s)\,\mathrm{d}s \quad \text{for } B \in \mathcal{B}_0,$$

where $\mathcal{B}_0$ is the class of bounded Borel sets in $\mathbb{R}^2$. For any $A, B \in \mathcal{B}_0$ define

$$M(A,B) = \mathbb{E} \sum_{y_i \in Y \cap A} \sum_{y_j \in Y \cap B} \frac{1}{\lambda(y_i)\lambda(y_j)}$$

and assume this is finite for all $A, B \in \mathcal{B}_0$. Then M is the second moment measure of the random measure Ξ which puts mass $1/\lambda(y_i)$ on each point $y_i \in Y$, i.e.

$$\Xi = \sum_{y_i \in Y} \frac{1}{\lambda(y_i)} \delta_{y_i}.$$

Note that Ξ is well defined since $\lambda(y_i) > 0$ almost surely for all $y_i \in Y$.

Definition 1 *The point process Y is "*second-order intensity-reweighted stationary*" if the random measure Ξ is second-order stationary. Equivalently $M(A,B) = M(A+x, B+x)$ for all $x \in \mathbb{R}^2$ where $A + x$ denotes the translation of A by the vector x.*

A second-order stationary point process is also second-order intensity-reweighted stationary. A second-order intensity-reweighted stationary process must either have zero intensity almost everywhere or positive intensity almost everywhere. Examples of second-order intensity-reweighted stationary point processes include all inhomogeneous Poisson processes with an intensity function, and the independent thinning by a measurable thinning field of any stationary point process with finite intensity (see section 3).

Definition 2 *Let Y be a second-order intensity-reweighted stationary point process. Define the* inhomogeneous K-function *of Y by*

$$K_{\mathrm{inhom}}(t) = \frac{1}{|B|} \mathbb{E} \sum_{y_i \in Y \cap B} \sum_{y_j \in Y \setminus \{y_i\}} \frac{1(\|y_i - y_j\| \le t)}{\lambda(y_i)\lambda(y_j)}, \quad t \ge 0, \tag{1}$$

for any $B \in \mathcal{B}_0$ with $|B| > 0$, where $1(\cdot)$ denotes the indicator function, $|B|$ is the area (Lebesgue measure) of B, and we take $a/0 = 0$ for $a \ge 0$. This expression does not depend on the choice of B.

The inhomogeneous K-function is a generalisation of the usual K-function, since a stationary (or at least first-order stationary) point process has $\lambda(s) \equiv \lambda$ and the right hand side of (1) reduces to the definition of $K(t)$ in Ripley (1977). Analogously to the stationary case, we set $L_{\text{inhom}} = (K_{\text{inhom}}/\pi)^{1/2}$ (Besag, 1977; see also sections 2.2–2.3). For inhomogeneous Poisson processes with an intensity function, $L_{\text{inhom}}(t) \equiv t$.

Additionally $K_{\text{inhom}}(t)$ has an interpretation as a Palm expectation, similar to that for the stationary case:

$$K_{\text{inhom}}(t) = \mathbb{E}_s \sum_{y_i \in Y \setminus \{s\}} \frac{1(\|y_i - s\| \leq t)}{\lambda(y_i)} \tag{2}$$

for almost all $s \in \mathbb{R}^2$. This follows from the Campbell-Mecke formula.

To illuminate the conditions under which an inhomogeneous K-function exists, assume that the product densities $\rho^{(1)}$ and $\rho^{(2)}$ of Y exist (see Stoyan *et al.*, 1995, p. 111). The first-order density $\rho^{(1)}(s)$ is just the intensity function $\lambda(s)$. The value $\rho^{(2)}(u,v)\,\mathrm{d}u\,\mathrm{d}v$ is the probability that Y has a point in each of two infinitesimally small discs with centers u, v and volumes $\mathrm{d}u$, $\mathrm{d}v$. Then by standard results,

$$M(A,B) = \int_A \int_B \rho^{(2)}(u,v) \frac{1}{\lambda(u)\lambda(v)}\,\mathrm{d}u\ \mathrm{d}v = \int_A \int_B g(u,v)\,\mathrm{d}u\ \mathrm{d}v$$

where

$$g(s_1,s_2) = \frac{\rho^{(2)}(s_1,s_2)}{\lambda(s_1)\lambda(s_2)}, \quad s_1, s_2 \in \mathbb{R}^2 \tag{3}$$

is the *pair correlation function* of Y.

If we assume that g is translation invariant, i.e. $g(s, s+h) = g_0(h)$ for some function $g_0 : \mathbb{R}^2 \to [0,\infty)$, then Y is second-order intensity-reweighted stationary, and the inhomogeneous K-function equals

$$K_{\text{inhom}}(t) = \int_{B(0,t)} g_0(h)\,\mathrm{d}h. \tag{4}$$

where for $s \in \mathbb{R}^2$ and $r \geq 0$, $B(s,r)$ denotes the disc with center s and radius r.

Thus, a third example of a second-order intensity-reweighted stationary point process is given by inhomogeneous log Gaussian Cox processes (see Brix and Møller, 1998, Møller *et al.*, 1998) with pair correlation function of the form $g(s_1,s_2) = \exp(c(s_1 - s_2))$ where $c(\cdot)$ is the translation invariant covariance function of a Gaussian field.

2.2 Estimation of K_{inhom}

For simplicity, we consider below the estimation of K_{inhom} when g_0 exists and is isotropic, i.e. when $g_0(h)$ depends only on $\|h\|$. However, using edge corrections as discussed in Stoyan and Stoyan (1994) and Stoyan *et al.* (1995) p. 134-137 or Baddeley (1999), the proposed estimators of K_{inhom} can easily be modified to the general case.

It is easily verified that a pointwise unbiased estimator of K_{inhom} is given by

$$\hat{K}_{\text{inhom}}(t) = \frac{1}{|W|} \sum_{y_i \in Y \cap W} \sum_{y_j \in Y \cap W \setminus \{y_i\}} \frac{w_{y_i,y_j} 1(\|y_i - y_j\| \le t)}{\lambda(y_i)\lambda(y_j)}, \quad 0 \le t < t^*, \tag{5}$$

where w_{y_i,y_j} is Ripley's (1977) edge correction factor (see also Stoyan *et al.*, 1995, Baddeley, 1999) and

$$t^* = \sup\{r \ge 0 : \ |\{s \in W : \partial B(s,r) \cap W \ne \emptyset\}| > 0\},$$

where $\partial B(s,r)$ denotes the boundary of $B(s,r)$.

An alternative, which we have not investigated further, is to estimate the pair correlation function g by

$$\hat{g}_0(a) = \frac{1}{2\pi a |W|} \sum_{y_i \in Y \cap W} \sum_{y_i \in Y \cap W \setminus \{y_i\}} w_{y_i,y_j} \frac{\kappa(a - \|y_i - y_j\|)}{\lambda(y_i)\lambda(y_j)}, \quad a > 0, \tag{6}$$

for some smoothing kernel κ. This estimate is similar to the usual non-parametric estimate for g_0 in the stationary case, see e.g. Stoyan and Stoyan (1994); it can easily be extended to an estimate for the cross pair correlation function of a bivariate point pattern, see Brix and Møller (1998).

In practice $\lambda(\cdot)$ is not known, so $\lambda(\cdot)$ in (5)–(6) must be replaced by an estimate. For this one may use the non-parametric kernel estimate given by

$$\hat{\lambda}_b(s) = \sum_{y_i \in Y \cap W} \kappa_b(s - y_i)/c_{W,b}(y_i), \quad s \in W, \tag{7}$$

where κ_b is yet another kernel with bandwidth b and

$$c_{W,b}(y_i) = \int_W \kappa_b(s - y_i) \mathrm{d}s$$

is an edge correction factor so that $\int_W \hat{\lambda}_b(s)\,\mathrm{d}s$ is an unbiased estimator of $\mathbb{E}\,\mathrm{card}(Y \cap W)$ (Diggle, 1985; the bandwidths used in (6) and (7) may be quite different). However, the estimate of K_{inhom} obtained in this way turns out to be severely biased downwards; see section 2.3. This appears to be due to positive bias in the kernel estimate $\hat{\lambda}_b$ when it is evaluated at data points $y_i \in Y$.

In order to study the bias of $\hat{\lambda}_b$ in more detail we assume that $\kappa_b(s_1 - s_2) = 0$ whenever $s_2 \notin B(s_1, b)$ and that there exists an s in W with $B(s, 2b) \subseteq W$ so that $c_{W,b}(v) = c_{W,b}(s)$ for all $v \in B(s,b)$. If we also assume that the pair correlation function g exists, then the bias of $\hat{\lambda}_b$ evaluated at a point $y_i \in Y$ located at s is

$$\mathbb{E}_s \hat{\lambda}_b(s) - \lambda(s) = \frac{\kappa_b(0)}{c_{W,b}(s)} + \int_W \frac{\kappa_b(s - v)}{c_{W,b}(s)} \left(g(s,v)\lambda(v) - \lambda(s)\right) \mathrm{d}v. \tag{8}$$

For an inhomogeneous Poisson process (where $g \equiv 1$) with a continuous intensity surface, the second term on the right hand side of (8) vanishes as b tends to zero, while the first term is always positive and typically increases as b tends to zero. For an inhomogeneous Poisson process it thus makes sense to replace $\hat{\lambda}_b$ by a modified estimate

$$\bar{\lambda}_b(s) = \sum_{y_i \in Y \cap W \setminus \{s\}} \kappa_b(s - y_i)/c_{W,b}(y_i), \quad s \in W, \tag{9}$$

whose bias consists only of the second term on the right hand side of (8). For a clustered point process with $g \geq 1$ we have that $|\mathbb{E}_s \bar{\lambda}_b(s) - \lambda(s)| \leq |\mathbb{E}_s \hat{\lambda}_b(s) - \lambda(s)|$ for b sufficiently small, so also in this case $\bar{\lambda}$ seems to provide a better estimate than $\hat{\lambda}$. For repulsive point processes with $g \leq 1$ the picture is less clear since the integral term in (8) now produces a negative bias when b tends to zero. Indeed, both estimators $\hat{\lambda}$ and $\bar{\lambda}$ may be essentially useless e.g. for a hard core process with $g(s, v) = 0$ for all $v \in B(s, b)$.

For the experiments in the next section we use the modified estimate (9). The experimental results show that for inhomogeneous Poisson processes, the estimator of L_{inhom} given by $\hat{L}_{\text{inhom}} = (\hat{K}_{\text{inhom}}/\pi)^{1/2}$ is severely biased downwards when $\hat{\lambda}$ is used, while better results are obtained if $\bar{\lambda}$ is used instead. Note that $\hat{\lambda}_b(u) = \bar{\lambda}_b(u)$ for $u \in W \setminus Y$, so if $\hat{\lambda}_b$ and $\bar{\lambda}_b$ are approximated by their values evaluated at a fixed grid of points, the two estimators of the intensity surface agree with probability 1.

In practice it is often difficult to make an objective distinction between large-scale variation given by $\lambda(\cdot)$ and variation due to interactions. This is especially the case if Y has attractive interactions, since clustering due to attractive interactions may be hard to distinguish from clustering due to peaks in the intensity surface. Some prior knowledge or assumptions concerning the smoothness of $\lambda(\cdot)$ are therefore required: the user needs to first settle on a suitable bandwidth b, so that the smoothness of $\bar{\lambda}_b$ is in accordance with prior beliefs, and subsequently K_{inhom} can be applied to study possible deviations from the Poisson hypothesis. If κ_b is not strictly positive and the chosen bandwidth b is small compared to the observed interpoint distances, it may happen that $\bar{\lambda}_b(y_i)$ takes the value zero for an observed point y_i. In such a case one may choose to estimate $K_{\text{inhom}}(t)$ only for $t \leq \sup\{R > 0 | B(0, R) \subseteq \{s \in \mathbb{R}^2 | \kappa_b(s) > 0\}\}$; or simply to use a strictly positive kernel instead.

Sometimes it may be relevant to use a parametric model for the intensity surface. Under the assumption of a Poisson process, such a parametric model can easily be fitted using standard software for generalized linear models, see Berman and Turner (1992) and Baddeley and Turner (2000)

2.3 Examples

Our two examples of point patterns are, firstly, the locations of adult longleaf pine trees in a $200m \times 200m$ region, see Platt *et al.* (1988), and secondly, a pattern of Japanese black pine seedlings and saplings (Numata, 1964) in which we restrict attention to the 88 'large' trees (defined as over 20 cm in height). For convenience, both datasets have been rescaled to the unit square. The intensity surfaces are estimated using the product kernel $\kappa_b(u) = e_b(u_1)e_b(u_2)$, $u = (u_1, u_2) \in [0, 1]^2$, where $e_b(a) = (0.75/b)(1 - (a/b)^2)1(|a| < b)$ is the Epanechnikov kernel. The bandwidth was taken to be $b = 0.24$ in the first example, and $b = 0.28$ in the second.

Figure 1 shows the longleaf pines data together with the kernel estimate $\bar{\lambda}_b$ of the intensity surface and the estimate of L_{inhom} obtained using $\bar{\lambda}_b$. The envelopes in the right hand plot of Figure 1 are obtained as in Diggle (1983) from simulated realisations of the inhomogeneous Poisson process with intensity surface equal to the estimated surface shown in the middle plot of Figure 1. We used 39 simulations so that $\hat{L}_{\text{inhom}}(t)$ is expected to be within the envelopes with 95% probability for each $t > 0$ under the hypothesis that the observed data is generated by an inhomogeneous Poisson process. The analysis suggests there is attraction between points. We have also estimated L_{inhom} using a parametric model $\log(\lambda((s_1, s_2); \psi)) = \psi_1 + \psi_2 s_1^2 + \psi_3 s_1^3$ for the intensity surface and got similar results (figure omitted).

Figure 2 shows the analogous plots for the large Japanese black pines. There is a slight hint of repulsion. The Japanese black pines data have previously been modelled by both homogeneous and inhomogeneous Gibbs or Markov point processes, see Ogata and Tanemura (1985), Ogata and Tanemura (1986), and Goulard *et al.* (1996). Knowledge about the environment in which the plants grow seems to be crucial in practice to make an informed choice between the homogeneous and the inhomogeneous setup.

For the inhomogeneous Poisson process with intensity surface as in Figure 1, Figure 3 shows the bias of the estimator of L_{inhom} obtained with the usual kernel estimator $\hat{\lambda}_b$ (upper plots) and with the modified kernel estimator $\bar{\lambda}_b$ (lower plots) for values of $b = 0.18$, 0.24 and 0.30. The estimate of L_{inhom} is substantially biased downwards for all three values of b when $\hat{\lambda}_b$ is used while the bias is of a small magnitude when $\bar{\lambda}_{0.24}$ or $\bar{\lambda}_{0.30}$ is used.

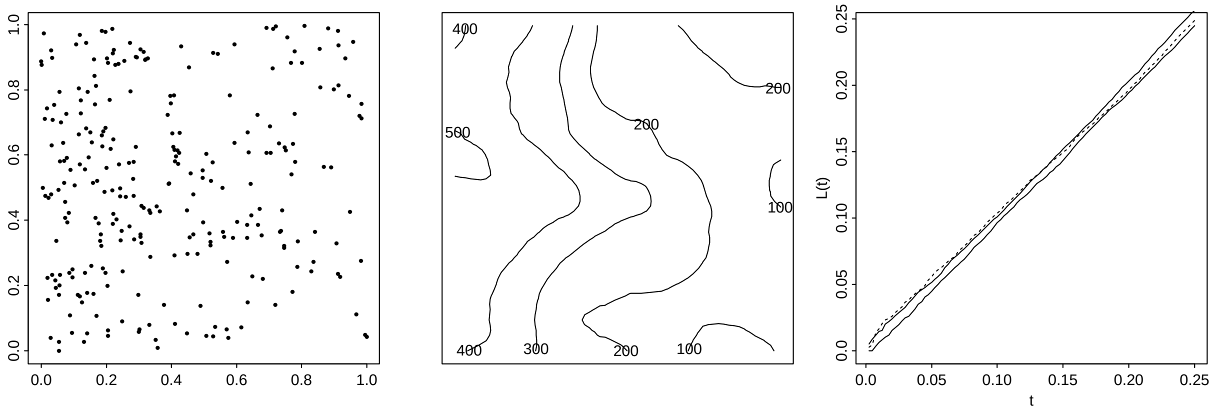

Figure 1: Longleaf pines data. *Left:* the locations of the 271 trees. *Middle:* non-parametric kernel estimate of the intensity surface. *Right:* estimate of L_{inhom} from the data (dotted) and envelopes of simulated estimates (solid) computed from 39 simulated realisations of an inhomogeneous Poisson process with intensity equal to the fitted intensity shown in the middle plot.

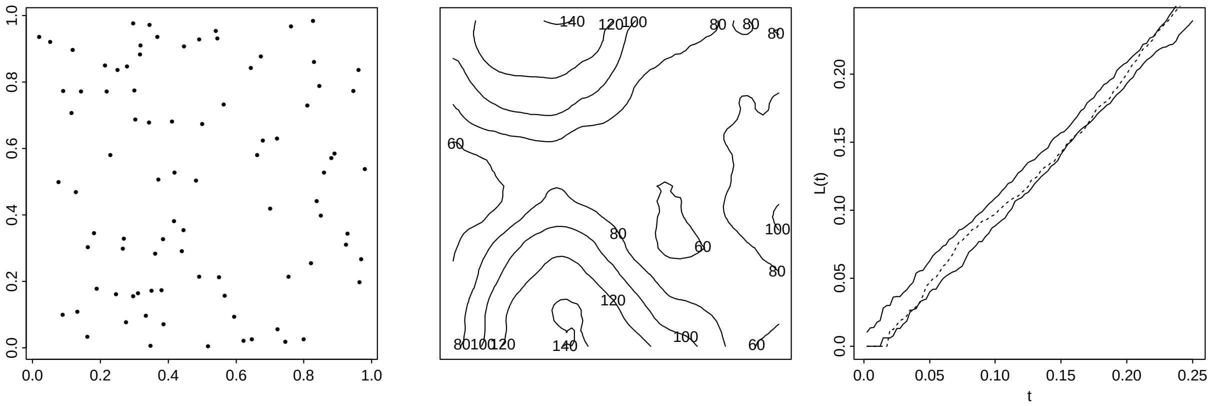

Figure 2: Japanese black pines data. *Left:* the locations of the 88 large trees. *Middle:* non-parametric kernel estimate of the intensity surface. *Right:* estimate of L_{inhom} from the data (dotted) and envelopes of simulated estimates (solid) computed from 39 simulated realisations of an inhomogeneous Poisson process with intensity equal to the fitted intensity shown in the middle plot.

2.4 Counterexamples

To emphasise the fact that the K-function does not uniquely characterise a point process, Baddeley and Silverman (1984) gave an example of a non-Poisson stationary process whose K function is identical to that of the homogeneous Poisson process. Here we modify that counterexample to apply to the inhomogeneous K-function. Specifically, we will show that for any given intensity function $\lambda(s)$, there exists a point process X which is not a Poisson process and whose inhomogeneous K-function is identical to that of the inhomogeneous Poisson process with intensity function $\lambda(s)$.

First note that it is easy to construct non-negative integer random variables N such that $\mathbb{E}N = \mathrm{Var}N = \mu$, for any given $\mu > 0$, which nevertheless have a distribution which is markedly different from the Poisson distribution. One such construction is as follows. For any $0 < a \leq 1$ the variable N_a which takes values $0, 1$ and 10 with probabilities $1-a+a^2/10$, $a - a^2/9$ and $a^2/90$ respectively, has mean and variance equal to a. Hence for any $\mu > 0$, if we let $m = [\mu]$ be the largest integer smaller than μ and $a = \mu - [\mu]$ the fractional part of μ, the random variable

$$N = N_a + N_1^{(1)} + N_1^{(2)} + \ldots + N_1^{(m)}$$

has mean and variance equal to μ, where the summands are independent and $N_1^{(1)}, N_1^{(2)}, \ldots,$ $N_1^{(m)}$ are distributed as N_1.

Now let $\lambda : \mathbb{R}^2 \to [0, \infty)$ be any locally integrable function. Partition $\mathbb{R}^2$ into disjoint cells $C_i,\ i = 1, 2, \ldots$, of finite positive area. For each i, let

$$\mu_i = \int_{C_i} \lambda(s)\,\mathrm{d}s > 0$$

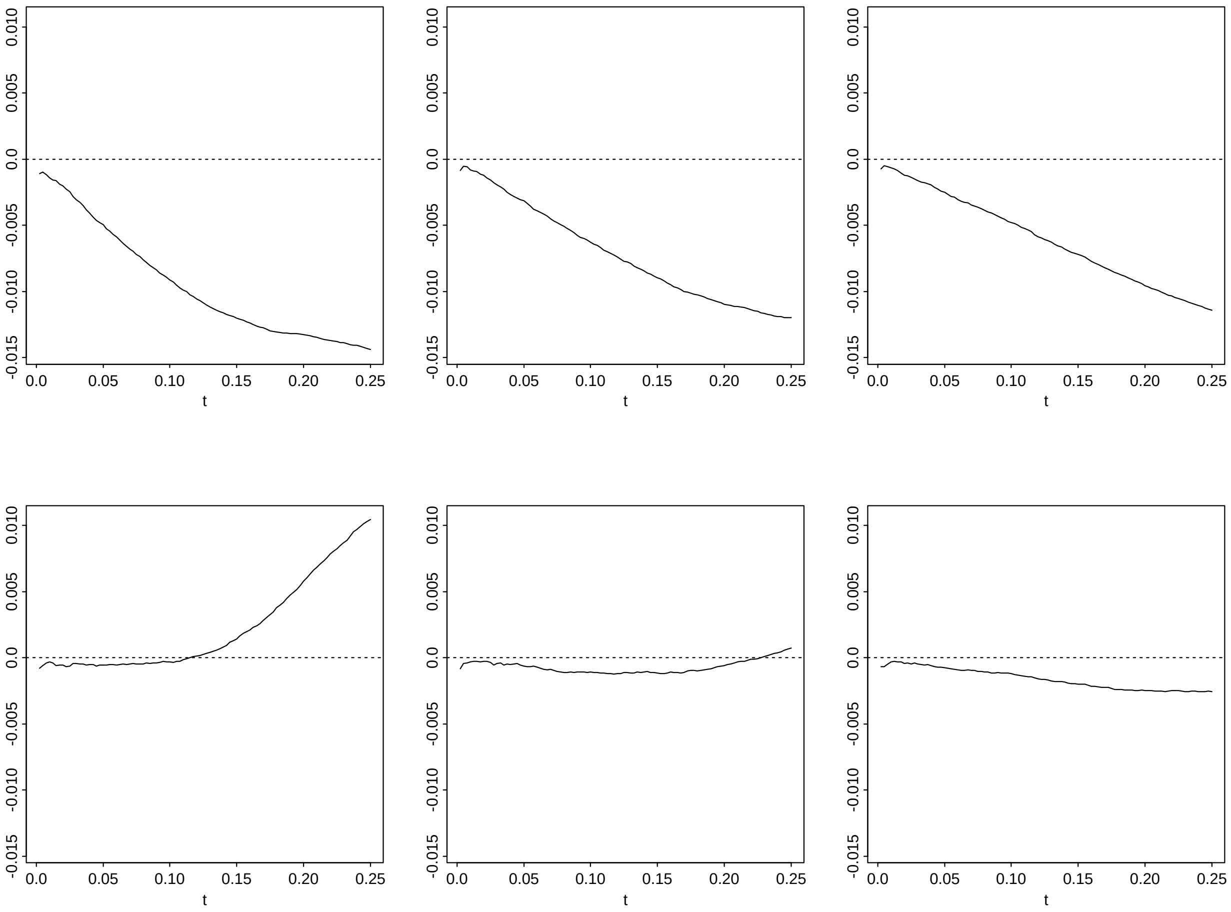

Figure 3: Each plot shows the mean of $L_{\text{inhom}}(t) - t$ estimated from 200 simulations of the inhomogeneous Poisson process with intensity surface shown in Figure 1. *Upper row:* Estimates obtained using $\hat{\lambda}_b$ and values of b equal to 0.18, 0.24 and 0.30 (left to right). *Lower row:* As upper row but using $\bar{\lambda}_b$.

and let N_i be independent random variables constructed as above (setting $N_i = 0$ if $\mu_i = 0$) such that $\mathbb{E}\, N_i = \mathrm{Var}(N_i) = \mu_i$. Given $N_i = n_i$, let n_i i.i.d. random points be placed in C_i with probability density $f_i(s) = \mu_i^{-1}\lambda(s)$, $s \in C_i$ and zero otherwise. The points in C_i are independent of points in other cells C_j, $j \neq i$. All points taken together form the counterexample process X.

Consider the second-order reduced moment measure $\alpha^{(2)}$ for X,

$$\alpha^{(2)}(B_1, B_2) = \mathbb{E} \sum_{y_i \in X} \sum_{y_j \in X \setminus \{y_i\}} 1(y_i \in B_1, y_j \in B_2)\,,$$

for $B_1, B_2 \in \mathcal{B}_0$. Writing

$$B_m = \bigcup_k (B_m \cap C_k)\,, \quad m = 1, 2\,,$$

it is easy to show by construction of X that

$$\alpha^{(2)}(B_1, B_2) = \sum_{k,l} \nu_{k,1}\nu_{l,2} = \sum_k \nu_{k,1} \sum_l \nu_{l,2} \,,$$

where $\nu_{k,m} = \int_{B_m \cap C_k} \lambda(s)\,\mathrm{d}s$. It follows that $\alpha^{(2)}$ has density $\rho^{(2)}(u,v) = \lambda(u)\,\lambda(v)$ with respect to the Lebesgue measure on $\mathbb{R}^4$. Hence by equations (3) and (4), X has inhomogeneous K-function

$$K_{\text{inhom}}(t) = \int_{B(0,t)} 1\,\mathrm{d}s = \pi t^2$$

identical to that of the inhomogeneous Poisson process with intensity function λ.

3 Thinned Markov point processes

3.1 Introduction

The class of *Markov point processes* (Ripley and Kelly, 1977, Baddeley and Møller, 1989) allows for parametric modelling of clustering or repulsion between points. In the sequel we consider Markov point processes on $\mathbb{R}^2$ as specified below.

We suppose that an 'interaction function' ϕ is defined for all point configurations $x \subset \mathbb{R}^2$ and that ϕ has finite interaction radius $r > 0$. Specifically, ϕ is a non-negative measurable function with $\phi(x) = 1$ whenever $\|x_i - x_j\| > r$ for some distinct points $x_i, x_j \in x$, and $\phi(x) > 0$ implies that $\phi(z) > 0$ for any $z \subset x$.

For disjoint point patterns y and x with $\mathrm{card}(y) < \infty$, define

$$\lambda(y \mid x) = \prod_{z \subseteq y \cup x \,:\, z \cap y \neq \emptyset} \phi(z) \qquad (10)$$

if $\phi(x) > 0$, and $\lambda(y \mid x) = 0$ otherwise. This will be the *Papangelou conditional intensity* of the point process at x. Note that $\lambda(y \mid x)$ depends on x only through $\{x_i \in x | \exists\, y_j \in y : \|x_i - y_j\| \le r\}$ provided $\phi(x) > 0$. For $B \in \mathcal{B}_0$, let μ_B denote the distribution of the unit rate Poisson point process on B.

A Markov point process X with interaction function ϕ is specified by a set of conditional densities $\{f_B(\cdot|\cdot)\}_{B \in \mathcal{B}_0}$, where

$$f_B(x \cap B \mid x \setminus B) \propto \lambda(x \cap B \mid x \setminus B) \qquad (11)$$

considered as function of $x \cap B$ is assumed to be a well-defined density with respect to μ_B if $\phi(x \setminus B) > 0$. Additional conditions on ϕ are, of course, required in order to ensure that X exists and further conditions may be added to ensure that its distribution is uniquely defined (Preston, 1976, Georgii, 1988).

A standard example of an inhibitive Markov point process is the Strauss process. For parameters $\theta_1 \in \mathbb{R}$ and $\theta_2 \in [-\infty, 0]$, the interaction function is given by $\phi(\{x_i\}) = \exp(\theta_1)$,

$\phi(\{x_i, x_j\}) = \exp(\theta_2)$ if $\|x_i - x_j\| \leq r$ and $\phi(\{x_i, x_j\}) = 1$ otherwise, while $\phi(x) = 1$ if $\text{card}(x) > 2$. The right hand side of (11) is integrable with respect to μ_B provided $\theta_2 \leq 0$. As ϕ is translation invariant, it can be shown that for all $(\theta_1, \theta_2) \in \mathbb{R} \times [-\infty, 0]$, there exists at least one stationary Markov point process with interaction function ϕ.

Models for inhomogeneous point patterns can be derived from Markov point processes in various ways. Inhomogeneous Markov point processes may be obtained by letting the first-order terms $\phi(\{x_i\})$ depend on x_i. For second-order terms, ϕ is usually chosen to be translation invariant, while $\phi = 1$ for higher-order terms, see Ogata and Tanemura (1986), and Stoyan and Stoyan (1998). Nielsen and Jensen (1999) consider another interesting approach where smooth transformations are applied to homogeneous Markov point processes in order to obtain models for inhomogeneous point patterns.

In the following we propose a third approach where an independent thinning is applied to a stationary Markov point process X to obtain a point process $Y = \{x_i \in X : U_i \leq p(x_i)\}$. Here $p : \mathbb{R}^2 \to [0, 1]$ is a measurable function and the U_i are i.i.d. random variables uniformly distributed on $[0, 1]$ which are generated independently of X. As in section 2, we suppose that Y is observed within a bounded window W. For later purposes, note that the conditional probability of observing a point pattern $Y \cap W = y$ given $X = x$ is

$$p(y \mid x) = 1(y \subseteq x) \prod_{x_i \in y} p(x_i) \prod_{x_i \in (x \cap W) \setminus y} (1 - p(x_i)). \tag{12}$$

Further, the intensity of Y is given by

$$\lambda(s) = \rho\, p(s), \quad s \in \mathbb{R}^2, \tag{13}$$

and its second-order product density by

$$\rho^{(2)}(s_1, s_2) = \lambda(s_1)\lambda(s_2) g_X(s_1 - s_2), \quad s_1, s_2 \in \mathbb{R}^2,$$

where ρ and $g_X(\cdot)$ denote the intensity and pair correlation function of X. Hence, the pair correlation function of Y equals g_X, so the K_{inhom}-function of Y is equal to the K-function for X.

The choice between the three approaches will depend on the scientific context. For the thinned Markov point process, the thinning probability may be interpreted, for example, as the probability of survival of a plant or the probability of observing an animal in a wildlife population survey. It is natural to assume that such probabilities are location-dependent. The thinned Markov point process setup has some advantages over the others, since summary statistics may be defined and estimated as discussed in section 2. One possibly less appealing property of the thinned Markov point process model is that it is non-Markovian (see counterexample 2 in Baddeley *et al.*, 1995) except in the case where $\phi(x) = 1$ for $\text{card}(x) > 1$, i.e. when Y is an inhomogeneous Poisson process. However, as discussed in more detail in section 3.2, for likelihood inference based on MCMC methods, we need only to sample from the distribution of X and the conditional distribution of X given Y, and these distributions are both Markov. Moreover, the simple correspondence

(13) between the intensity surface of the thinned point process and the thinning surface enables us to propose a semi-parametric approach for the statistical analysis of a thinned Markov point process, see section 3.2. The semi-parametric approach is useful in many real applications where it is not always easy to propose a parsimonious parametric model for the large scale variation determined by the thinning surface p. The discussion at the end of section 2.2 concerning confounding also applies in the context of semi-parametric inference, especially if X has attractive interactions.

A general problem for all three approaches is related to modelling of positive correlation between points by interactions of a Markov point process, since there seems to be a lack of Markov models for clustering which are both flexible and biologically interpretable, see Møller (1999) — the area-interaction process (Baddeley and Van Lieshout, 1995, Särkkä and Baddeley, 1998) or the interacting neighbour point processes in Grabarnik and Särkkä (1999) are possible exceptions. Repulsion due to effects such as competition between plants can however most naturally be modelled by a Markov point process. For the Strauss process, for example, the interaction distance r may describe the size of the "influence" zone of a tree, and θ_2 determines the "strength" of the influence. See also the discussion in Stoyan and Stoyan (1998).

3.2 Semi-parametric inference for a thinned Markov point process

Let X be a stationary Markov process and Y the process obtained by independent random thinning of X according to a thinning field $p : \mathbb{R}^2 \to [0, 1]$. Suppose that the distribution of X is specified by a parametric model, depending on a parameter $\theta \in \Theta \subseteq \mathbb{R}^d$, $d \geq 1$, with (translation invariant) interaction function $\phi(\cdot;\theta)$, Papangelou conditional intensity $\lambda(\cdot|\cdot;\theta)$, and nth order product density $\rho^{(n)}(\cdot;\theta)$. To begin with, we shall assume that the intensity surface $\lambda(\cdot)$ is known inside the observation window W. For a given $\theta \in \Theta$ satisfying $\rho(\theta) \geq \sup_{s\in\mathbb{R}^2} \lambda(s)$ where $\rho(\theta)$ is the intensity of X under θ, it follows from (12) that $p(\cdot) = p(\cdot;\theta)$ is given by

$$p(s;\theta) = \frac{\lambda(s)}{\rho(\theta)}, \quad s \in W. \tag{15}$$

We also assume that the interaction radius r is known (estimation of r is discussed in section 3.2.3). The likelihood of θ is given by the following result:

Proposition 1 *Let y be an observation of the thinned process $Y \cap W$. The likelihood of θ given the data $Y \cap W = y$, i.e. the density with respect to the unit rate Poisson process on W, is*

$$L(\theta) = \mathbb{E}_\theta\left[p(y \mid y \cup X;\theta)\lambda(y \mid X;\theta)\right] = \rho^{(\mathrm{card}(y))}(y;\theta)\, \mathbb{E}_{y,\theta}\left[p(y \mid X;\theta)\right] \tag{16}$$

where $\mathbb{E}_\theta$ and $\mathbb{E}_{y,\theta}$ denote expectation with respect to the distribution of X and the Palm distribution of X at y, respectively.

Remarks: For the first expectation in (16), $\lambda(y \mid X;\theta)$ is well-defined with probability 1, since the stationarity of X implies that $X \cap y = \emptyset$ almost surely. The likelihood (16) has the usual structure of a missing data likelihood: heuristically, $\rho^{(\mathrm{card}(y))}(y;\theta) =$ "$P_\theta(y \subseteq X)$" and $\mathbb{E}_{y,\theta} p(y \mid X) =$ "$\mathbb{E}_\theta[p(y \mid X) \mid y \subseteq X]$" so that $L(\theta) =$ "$\mathbb{E}_\theta[p(y \mid X)]$".

Proof. We suppress dependence on the parameter θ. Let μ denote the unit rate Poisson process on W. Since the superposition $\Phi = \Phi_1 \cup \Phi_2$ of two independent unit rate Poisson processes Φ_1 and Φ_2 is a Poisson process with intensity 2, and the conditional distribution of Φ_1 given Φ is uniform on the set consisting of the 2^Φ subsets of Φ, we have that

$$\int \sum_{y \subseteq x} h(x \setminus y, y)\mu(dx) = \exp(|W|) \int \int h(z,y)\mu(dz)\mu(dy) \tag{17}$$

for non-negative measurable functions h. Using (12) and (17) we find that

$$\begin{aligned} P(Y \cap W \in F) &= \mathbb{E}\left[\int f_W(x \mid X \setminus W) \sum_{y \subseteq x:\, y \in F} p(y \mid x)\mu(dx)\right] \\ &= \exp(|W|)\mathbb{E}\left[\int_{y \in F} \int f_W(y \cup z \mid X \setminus W) p(y \mid y \cup z)\mu(dz)\ \mu(dy)\right] \end{aligned}$$

where F is any measurable event in the canonical probability space. Hence, the density $f_W(y)$ of $Y \cap W$ with respect to μ is given by

$$\begin{aligned} f_W(y)\exp(-|W|) &= \mathbb{E}\left[\int p(y \mid y \cup z) f_W(y \cup z \mid X \setminus W)\ \mu(dz)\right] \\ &= \mathbb{E}\left[\int p(y \mid y \cup z)\lambda(y \mid z \cup (X \setminus W)) f_W(z \mid X \setminus W)\ \mu(dz)\right] \\ &= \mathbb{E}[\mathbb{E}\{p(y \mid y \cup (X \cap W))\lambda(y \mid X) \mid X \setminus W\}] \\ &= \mathbb{E}[p(y \mid y \cup (X \cap W))\lambda(y \mid X)] \end{aligned}$$

where we have used (10)–(11) to obtain the second equality. Thereby the first equality in (16) is proved. The second equality in (16) is a straightforward extension of the Nguyen and Zessin (1979) equality.

In section 3.2.1 the mean value with respect to the stationary process X in (16) will be approximated by the mean value with respect to a finite point process $\tilde{X}$ defined on an extended rectangular window $\tilde{W} \supset W$ such that $\tilde{X}$ is "circulant stationary", i.e. stationary under toroidal translations of W. The density of $\tilde{X}$ with respect to $\mu_{\tilde{W}}$ is

$$\tilde{f}(x;\theta) = c(\theta)\tilde{\lambda}(x;\theta) = c(\theta) \prod_{\emptyset \neq z \subseteq x} \tilde{\phi}(z;\theta) \tag{18}$$

for finite point configurations $x \subset \tilde{W}$, where $c(\theta)$ is the normalizing constant and the interaction function $\tilde{\phi}(\cdot;\theta)$ is obtained from $\phi(\cdot;\theta)$ in a straightforward manner when $\tilde{W}$ is wrapped on a torus. Note that $\tilde{\lambda}(x;\theta)$ is the expression (10) for the "Papangelou conditional intensity of x given $\emptyset$", with respect to $\tilde{\phi}(\cdot;\theta)$. We assume that the mean value in (16) is well approximated when $\tilde{W}$ is chosen sufficiently large.

In practice $\lambda(\cdot)$ is unknown and we propose therefore a semi-parametric approach where $\lambda(\cdot)$ in (15) is replaced by the non-parametric estimate $\hat{\lambda}_b(\cdot)$ evaluated at a fine grid of points (see section 2.2). For a given $\theta \in \Theta$ satisfying $\tilde{\rho}(\theta) \geq \sup_{s\in W} \hat{\lambda}_b(s)$ where $\tilde{\rho}(\theta)$ is the intensity of $\tilde{X}$ under θ, the natural estimate of $p(\cdot)$ is $\hat{p}(s;\theta) = \hat{\lambda}_b(s)/\tilde{\rho}(\theta)$, $s \in W$. Inserting this into (12) gives an estimate $\hat{p}(y \mid x;\theta)$ for $p(y \mid x)$.

The approximate semi-parametric likelihood is

$$\tilde{L}(\theta) = \int \hat{p}(y \mid y \cup z;\theta)\tilde{f}(y \cup z;\theta)\mu_{\tilde{W}}(dz) = \int \hat{p}(y \mid x;\theta)\tilde{f}(x;\theta)\mu_{\tilde{W},y}(dx)\,, \qquad (19)$$

where $\mu_{\tilde{W},y}$ is the Palm distribution of $\mu_{\tilde{W}}$ at y, which by Slivnyak's theorem is simply the distribution of $Z \cup y$ when $Z \sim \mu_{\tilde{W}}$.

3.2.1 Monte Carlo approximation of the semi-parametric likelihood

The approximate semi-parametric likelihood (19) is not known in closed form, but may be estimated by using Markov chain Monte Carlo methods as described in Gelfand and Carlin (1991) and Geyer (1994). To do so, choose a fixed parameter point $\psi \in \Theta$, and write

$$\frac{\tilde{L}(\theta)}{\tilde{L}(\psi)} = \frac{c(\theta)}{c(\psi)} \int \frac{\hat{p}(y \mid x;\theta)\tilde{\lambda}(x;\theta)}{\hat{p}(y \mid x;\psi)\tilde{\lambda}(x;\psi)} \hat{f}(x \mid y;\psi)\mu_{\tilde{W},y}(dx), \qquad (20)$$

where

$$\hat{f}(x \mid y;\psi) = \hat{p}(y \mid x;\psi)\tilde{f}(x;\psi)/\tilde{L}(\psi) \propto \hat{p}(y \mid x;\psi)\tilde{\lambda}(x;\psi) \qquad (21)$$

is the conditional density with respect to $\mu_{\tilde{W},y}$ of X given $Y = y$ when p is replaced by $\hat{p}(\cdot;\psi)$. A Monte Carlo approximation of the integral in (20) may thus be obtained from simulations of $\tilde{f}(\cdot \mid y;\psi)$. The unknown ratio of normalizing constants in (20) must typically also be replaced by a Monte Carlo approximation.

The likelihood (19) may be highly multimodal since a thinned point process with a high underlying intensity and small retention probabilities may produce realizations which are hard to distinguish from realizations of a thinned point process where the underlying intensity is small and the retention probabilities high. As noted in section 3.2.2 below, if $\tilde{f}$ is e.g. the density of a Strauss process, then the likelihood (19) is in particular constant for parameter values θ with $\theta_2 = 0$ (i.e. the Poisson case). To deal with the possible multimodality we choose a range of values $\alpha_i \geq \sup_{s\in W} \hat{\lambda}_b(s)$, $i = 1,\ldots,m$, and maximize $\tilde{L}(\theta)$ for each fixed value of $\tilde{\rho}(\theta) = \alpha_i\,, i = 1,\ldots,m$. The corresponding estimates are denoted $\hat{\theta}(\alpha_i)\,, i = 1,\ldots,m$. In certain applications one may take $m = 1$; it may, for instance, be known that $\sup_{s\in W} p(s) = 1$ in which case one can take $\alpha_1 = \sup_{s\in W} \hat{\lambda}_b(s)$ so that $\sup_{s\in W} \hat{p}(s;\theta) = 1$.

Note that for θ and ψ with $\tilde{\rho}(\theta) = \tilde{\rho}(\psi) = \alpha$, the expression (20) reduces to

$$H_\alpha(\theta, \psi) = \frac{c(\theta)}{c(\psi)} \int \frac{\tilde{\lambda}(x;\theta)}{\tilde{\lambda}(x;\psi)} \hat{f}_\alpha(x \mid y;\psi)\mu_{\tilde{W},y}(dx), \tag{22}$$

where $\hat{f}_\alpha(x \mid y;\psi) \propto \hat{p}_\alpha(y \mid x)\tilde{\lambda}(x;\psi)$ with $\hat{p}_\alpha(y \mid x)$ given by (12) when $p(\cdot)$ is replaced by $\hat{p}_\alpha(\cdot) = \hat{\lambda}_b(\cdot)/\alpha$.

The Monte Carlo approximation of (22) is

$$\log H_\alpha(\theta, \psi) \approx \log \sum_{i=1}^{n} \frac{\tilde{\lambda}(X_i^*;\theta)}{\tilde{\lambda}(X_i^*;\psi)} - \log \sum_{i=1}^{n} \frac{\tilde{\lambda}(X_i;\theta)}{\tilde{\lambda}(X_i;\psi)}, \tag{23}$$

where $(X_i^*)_{i=1}^n$ is a sample from $\hat{f}_\alpha(\cdot \mid y;\psi)$ and $(X_i)_{i=1}^n$ is a sample from $\tilde{f}(\cdot;\psi)$. These samples can be generated by the Metropolis-Hastings algorithm studied in Geyer and Møller (1994), Geyer (1999), and Møller (1999).

3.2.2 The exponential family case

In this section we restrict attention to the case where $\{\tilde{f}_\theta \mid \theta \in \Theta\}$ is an exponential family of densities so that

$$\tilde{f}_\theta(x) \propto \exp\left(\theta_1 n(x) + \theta_2 s(x)^\mathsf{T}\right),$$

where $n(x)$ is the number of points in x and the statistic $s(x) = (s_1(x), \ldots, s_{d-1}(x))$ determines the interaction structure of the point process given by $\tilde{f}_\theta$ ('T' denotes matrix transpose). Hence for the MCMC approximation (23) we only need to store the values of the sufficient statistics $(n(X_i), s(X_i))$ and $(n(X_i^*), s(X_i^*))$, $i = 1, \ldots, n$. For example, for the Strauss process,

$$s(x) = \sum_{\{x_i, x_j\} \subseteq x : x_i \neq x_j} 1(\|x_i - x_j\|_{\text{torus}} \leq r),$$

where $\|\cdot\|_{\text{torus}}$ denotes distance on the torus. Note that $\tilde{f}_\theta$ is the density of a Poisson process with intensity $\exp(\theta_1)$ whenever $\theta_2 = 0$. It follows easily from (19) that $\tilde{L}(\theta)$ is constant for all θ with $\theta_2 = 0$. Since $\partial\tilde{\rho}(\theta)/\partial\theta_1 = \mathbb{V}\mathrm{ar}_\theta n(\tilde{X})$, there is a one-to-one correspondence between (θ_1, θ_2) and $(\tilde{\rho}(\theta), \theta_2)$ when $\mathbb{V}\mathrm{ar}_\theta n(\tilde{X}) > 0$. Henceforth we assume that $\mathbb{V}\mathrm{ar}_\theta n(\tilde{X}) > 0$.

Assume further that $\tilde{\rho}(\theta) = \tilde{\rho}(\psi) = \alpha$ is fixed. Then $\theta_1 = \theta_1(\theta_2)$ becomes a function of θ_2 and by the implicit function theorem,

$$\frac{\partial\theta_1}{\partial\theta_2} = -\frac{\mathbb{C}\mathrm{ov}_\theta(n(\tilde{X}), s(\tilde{X}))}{\mathbb{V}\mathrm{ar}_\theta n(\tilde{X})}.$$

Further,

$$\frac{\partial \log H_\alpha(\theta,\psi)}{\partial \theta_2} = \frac{\partial \log H_\alpha(\theta,\psi)}{\partial \theta}\frac{\partial \theta^\mathsf{T}}{\partial \theta_2} =$$

$$\left[\mathbb{E}_\theta(n(\tilde{X}) \mid y) - \mathbb{E}_\theta n(\tilde{X}) \quad \mathbb{E}_\theta(s(\tilde{X}) \mid y) - \mathbb{E}_\theta s(\tilde{X})\right] \begin{bmatrix} -\frac{\mathbb{C}\mathrm{ov}_\theta(n(\tilde{X}),s(\tilde{X}))}{\mathbb{V}\mathrm{ar}_\theta n(\tilde{X})} \\ I_{d-1} \end{bmatrix}, \quad (24)$$

where I_k is the $k \times k$ identity matrix and the conditional expectations are with respect to $\hat{f}_\alpha(\cdot \mid y;\theta)$. Similarly, the Hessian $\frac{\partial^2 \log H_\alpha(\theta,\psi)}{\partial\theta_2 \partial\theta_2^\mathsf{T}}$ can also be expressed in terms of moments of $n(\tilde{X})$ and $s(\tilde{X})$. We can then estimate θ_2 by maximizing $\log H_\alpha(\theta,\psi)$ using Newton-Raphson, see section 3.3. The moments $\mathbb{E}_\theta n(\tilde{X})^l s(\tilde{X})^j$ which occur in the gradient and Hessian are replaced by Monte Carlo estimates given by

$$\mathbb{E}_\theta n(\tilde{X})^l s(\tilde{X})^j \approx \{\sum_{i=1}^n n(X_i)^l s(X_i)^j \frac{\tilde{\lambda}(X_i;\theta)}{\tilde{\lambda}(X_i;\psi)}\} / \{\sum_{i=1}^n \frac{\tilde{\lambda}(X_i;\theta)}{\tilde{\lambda}(X_i;\psi)}\}\,. \quad (25)$$

The conditional moments $\mathbb{E}_\theta(n(\tilde{X})^l s(\tilde{X})^j \mid y)$ are estimated similarly but with X_i replaced by X_i^*. In order to estimate the moments for a given value θ_2 we determine first $\theta_1(\theta_2)$ by solving $\tilde{\rho}(\theta) = \alpha$ numerically with respect to θ_1. Again we use Newton-Raphson and a Monte Carlo estimate of $\tilde{\rho}(\theta) = \mathbb{E}_\theta n(\tilde{X})/|\tilde{W}|$.

The approximations (23) and (25) are only accurate for θ in a neighbourhood around ψ. Suppose that there exists an $\epsilon > 0$ such that $\psi = (\theta_1(\psi_2),\psi_2)$ works well as an importance sampling parameter for estimation of moments with respect to both $\tilde{f}(\cdot;\theta)$ and $\hat{f}_\alpha(\cdot \mid y;\theta)$ whenever $\|\theta_2 - \psi_2\| < \epsilon$ and $\theta = (\theta_1(\theta_2),\theta_2)$. Assume also that the Newton-Raphson procedure is initialized in θ_2^0 where $\psi = (\theta_1(\theta_2^0),\theta_2^0)$ is used as the importance sampling parameter. Let θ_2^1 denote the output of the Newton-Raphson update. If $\|\theta_2^1 - \psi_2\| < \epsilon$, we reuse the two importance samples generated for ψ when calculating the next update θ_2^2. Otherwise new importance samples are generated with ψ replaced by $(\theta_1(\theta_2^1),\theta_2^1)$ as the new importance sampling parameter.

Suppose that $\hat{\theta}_2$ is the estimate obtained by the Newton-Raphson procedure for given α. In order to calculate the log likelihood of $\hat{\theta}_2$ we generate importance samples for a range of importance sampling parameters ψ_2^i, $i = 0,\dots,m$, with $\psi_2^0 = 0$ and $\|\psi_2^{i+1} - \psi_2^i\| < \epsilon$, $i = 0,\dots,m$, where $\psi_2^{m+1} = \hat{\theta}_2$. Let $\psi^i = (\theta_1(\psi_2^i),\psi_2^i)$. The log likelihood of $\hat{\theta}(\alpha) = (\theta_1(\hat{\theta}_2),\hat{\theta}_2)$ is

$$\log H_\alpha(\hat{\theta}(\alpha),\psi^0) = \log H_\alpha(\hat{\theta}(\alpha),\psi^m) + \sum_{i=0}^{m-1} \log H_\alpha(\psi^{i+1},\psi^i)\,, \quad (26)$$

which can be calculated using the approximation (23). The approximation (26) may not work well if $\hat{\theta}(\alpha)$ and ψ^0 are far apart since Monte Carlo errors for each term in the sum may accumulate. Alternatively one may use umbrella sampling or the method of reverse logistic regression as discussed in Geyer (1998).

Recall that $\tilde{L}(\psi^0) = \tilde{L}((\log(\alpha),0))$ is constant for all α, so $\log H_\alpha(\hat{\theta}(\alpha),\psi^0)$ can be considered as a log likelihood ratio statistic for the Poisson hypothesis $\theta_1 \in \mathbb{R}$ and $\theta_2 = 0$. If a bootstrap is used for hypothesis testing concerning the parameter θ_2, it is computationally demanding to use the likelihood ratio statistic if the required number $m+1$ of importance samples is high. The Wald statistic given by

$$\left[\theta_2 \frac{\partial^2 H_\alpha(\theta,\psi)}{\partial\theta_2 \partial\theta_2^\mathsf{T}} \theta_2^\mathsf{T}\right]_{\theta=\hat{\theta}(\alpha)} \tag{27}$$

is a computationally less expensive alternative which can simply be evaluated in the last step of the parameter estimation procedure.

3.2.3 Estimation of the interaction radius and model validation

The interaction distance r can of course also be included in the likelihood estimation but the Newton-Raphson procedure described in section 3.2.2 is not applicable for r, since r is not an exponential family parameter. Instead one may maximize (22) for a range of values of r, whereby estimates $\hat{\theta}(\alpha, r)$ are obtained and then choose the value of r for which $\hat{\theta}(\alpha, r)$ has the maximal likelihood. Since L_{inhom} for Y is equal to L for the Markov point process, a plot of the estimated L_{inhom} may indicate the range of r values which should be considered.

Finally, the estimated model may be checked by calculating envelopes for L_{inhom} (see section 2.3 and section 3.3) or by Monte Carlo tests for goodness of fit (see e.g. Diggle, 1983).

3.3 Example

In accordance with the discussion in the end of section 3.1 we shall refrain from fitting a thinned Markov model to the longleaf pines data (Figure 1, section 2.3) for which it is not obvious how to relate the apparent clustering to interactions of a Markov point process. Instead we will fit a thinned Strauss process to the large Japanese black pines data (Figure 2).

We use the same bandwidth $b = 0.28$ as in section 2.3 and let the extended window $\tilde{W} = [0, 1.2]^2$. The right plot in Figure 2 suggests that $0.02 \leq r \leq 0.05$. Estimates $\hat{\theta}(\alpha_i, r_j)$ obtained for a range of values of α and the interaction distance r are shown in Table 1. Here $\alpha_i = \sup_s \hat{\lambda}_b(s)/c_i$ where c_i varies between 1.00 and 0.60. For each value of α and r, the Newton-Raphson maximization procedure with $\epsilon = 0.025$ was initialized in $\psi^0(\alpha) = (\log(\alpha), 0)$ (the estimate under the Poisson assumption). The values of the maximized Monte Carlo approximation of the log likelihood (26) are also given in Table 1. The maximized likelihood is essentially constant as a function of α, but has a maximum for $r = 0.04$.

It is not guaranteed that the trust region maximization procedure will reach a global maximum. In our experiments we tried other initial values than $\psi^0(\alpha)$ but this did not affect the value of the estimates. We also tried larger values of $\epsilon = 0.05$ and $\epsilon = 0.10$. The value $\epsilon = 0.05$ worked well but the maximization procedure failed to converge for

$\log \alpha = 5.47$ when $\epsilon = 0.10$ was used. Markov chain Monte Carlo samples of length 8000 were obtained by equispaced subsampling of chains corresponding to 1.6 million basic updates of the Geyer and Møller (1994) algorithm.

$\sup_s \hat{p}_\alpha(s)$	$\log \alpha$	r	$\hat{\theta}_1(\alpha, r)$	$\hat{\theta}_2(\alpha, r)$	$\log H_\alpha$
1.00	4.96	0.02	5.08	-0.57	1.61
1.00	4.96	0.03	5.19	-0.41	2.30
1.00	4.96	0.04	5.34	-0.36	3.11
1.00	4.96	0.05	5.27	-0.16	1.09
0.90	5.06	0.02	5.19	-0.54	1.58
0.90	5.06	0.03	5.32	-0.40	2.24
0.90	5.06	0.04	5.49	-0.36	3.11
0.90	5.06	0.05	5.40	-0.16	1.08
0.80	5.18	0.02	5.33	-0.54	1.58
0.80	5.18	0.03	5.46	-0.40	2.21
0.80	5.18	0.04	5.67	-0.36	3.00
0.80	5.18	0.05	5.57	-0.16	1.07
0.70	5.31	0.02	5.49	-0.56	1.58
0.70	5.31	0.03	5.64	-0.39	2.23
0.70	5.31	0.04	5.88	-0.37	3.04
0.70	5.31	0.05	5.77	-0.16	1.02
0.60	5.47	0.02	5.67	-0.56	1.51
0.60	5.47	0.03	5.85	-0.39	2.16
0.60	5.47	0.04	6.18	-0.40	2.92
0.60	5.47	0.05	5.95	-0.15	1.07

Table 1: Estimates of θ and corresponding log likelihoods for various values of α and r. The rightmost column shows $\log H_\alpha(\hat{\theta}(\alpha, r), \psi^0(\alpha))$.

In the following we consider inference for θ and p taking r to be fixed at $r_0 = 0.04$ and setting α equal to $\alpha_0 = \sup_s \hat{\lambda}_b(s)$. This means that $\sup_s \hat{p}_{\alpha_0}(s) = 1$. We do not know if we can rely on asymptotic normality for the parameter estimate and use instead a semi-parametric bootstrap for inference concerning θ. That is, a number (in this case 99) of independent simulations is generated from the Poisson model given by $\psi(\alpha_0)$ and $\hat{p}_{\alpha_0}(\cdot)$. The intensity surface is reestimated for each simulated Poisson pattern and an estimate of θ is obtained with the Newton-Raphson procedure for $r = r_0$ and α equal to the supremum of the intensity surface estimate for the simulated point pattern. Figure 4 shows a scatter-plot of the parameters estimated from the simulations together with the estimate $\hat{\theta}(\alpha_0, r_0)$ from the data.

If the estimate $\hat{\theta}_2(\alpha_0, r_0)$ is used as a test statistic for the hypothesis $\theta_2 = 0$ against the hypothesis $\theta_2 < 0$, a Monte Carlo p-value of 0.10 is obtained. The Monte Carlo p-values based on the likelihood ratio statistic (26) and the Wald statistic (27) are 0.04 and 0.02,

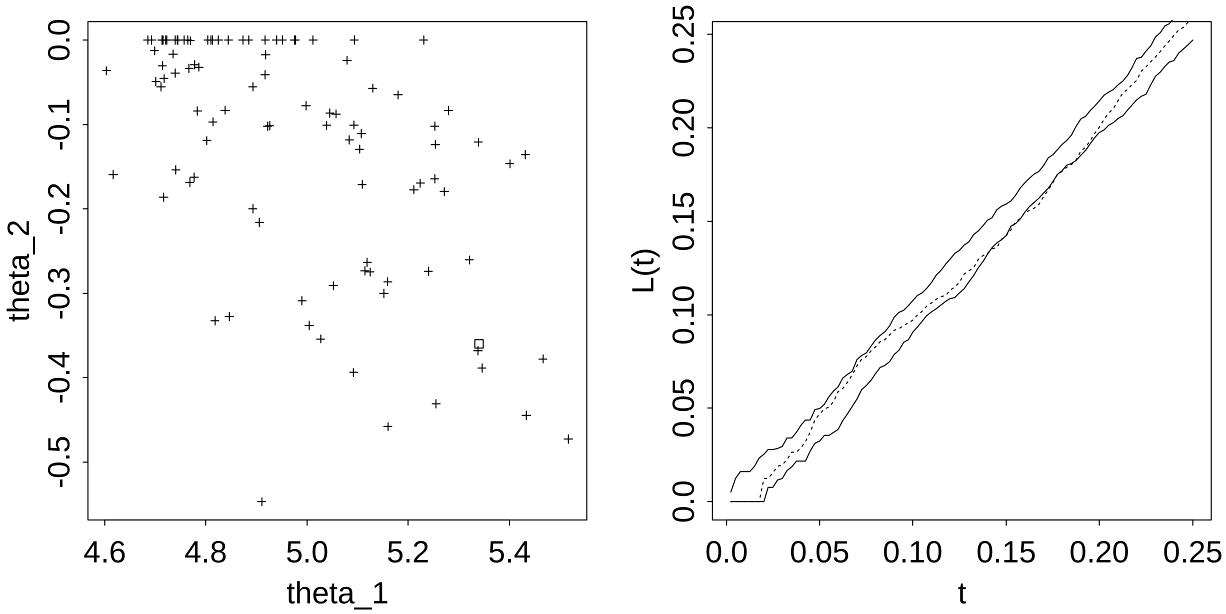

Figure 4: *Left:* Parameter values estimated from 99 simulations of the fitted Poisson process. The square shows the estimate obtained from the data. *Right:* L_{inhom} estimated from data, and envelopes calculated from 39 simulations of fitted thinned Strauss process.

respectively. Note that the Monte Carlo p-values are quite variable when they are based on just 99 simulations. Goulard *et al.* (1996) also fit a repulsive model to the large Japanese pines, but they do not report on a test for no interaction.

The right plot in Figure 4 shows L_{inhom} estimated from the data together with envelopes calculated from simulations of the thinned Strauss process. Compared to Figure 2, the repulsive interaction seems to improve the fit, although in both Figures 2 and 4 the estimated L_{inhom} is slightly below the lower envelope for distances around 0.15.

4 Discussion

We conclude this paper by discussing some open problems for the statistical analysis of inhomogeneous point patterns.

In section 2 we introduced and studied the use of the inhomogeneous K-function. It is also possible to define analogues to the empty space distribution function and the nearest-neighbour distribution function. Let for any $t \geq 0$ and $s \in \mathbb{R}^2$, $r(s,t) \geq 0$ be determined by

$$t = \int_{B(s,r(s,t))} \lambda(u) du \,. \tag{28}$$

If λ is strictly positive, then $r(s,t)$ is uniquely determined by (28). Set $d(s,Y) = \inf\{\|s - y_i\| : y_i \in Y\}$. Then, for $t \geq 0$, the analogue of the empty space distribution function is

$$F_s(t) = P(d(s,Y) \leq r(s,t)),$$

and the 'nearest-neighbour' distribution function is

$$G_s(t) = P_s(d(s,Y) \le r(s,t))\,.$$

When Y is a Poisson process, $F_s(t)$ and $G_s(t)$ do not depend on s and are both equal to $\exp(-t)$. Under the assumption of a Poisson process these summary statistics can therefore be estimated and compared to their theoretical values in order to investigate possible departures from the Poisson model. However, this requires that we are able to compute the function $r(s,t)$ which may be difficult in practice, and this function depends on a three dimensional argument.

We expect that the K_{inhom}-function and the semi-parametric approach will prove to be useful in many cases of applications as it may be difficult to model an inhomogeneous point pattern by a simple parametric model (one exception is the linear trend found in one of the point patterns analysed in Brix and Møller, 1998). Problems with bias depending on the choice of bandwidth may on the other hand occur due to the non-parametric kernel estimation of the intensity surface. For the semi-parametric method one advantage compared to the non-parametric approach based on K_{inhom} is that it is not required to estimate the intensity surface in points belonging to the point pattern data. One thereby avoids the additional bias (see (8)) due to possible spatial dependence in the data.

For the large Japanese black pines modelled by a thinned Strauss process (section 3.3) we observed that the maximized likelihood function appeared constant as a function of $\tilde{\rho}(\theta)$. One may ask if this is a feature of other exponential family models as considered in section 2.2.

Acknowledgement: This research was supported by the Australian Research Council, by the European Union's network "Statistical and Computational Methods for the Analysis of Spatial Data. ERB-FMRX-CT96-0096", by MaPhySto – Centre for Mathematical Physics and Stochastics, funded by a grant from The Danish National Research Foundation – and by the Danish Informatics Network in the Agricultural Sciences, funded by a grant from the Danish Research Councils.

References

Baddeley, A. J. (1999). Spatial sampling and censoring. In: *Stochastic Geometry: Likelihood and Computation* (eds. O. E. Barndorff-Nielsen, W. S. Kendall and M. N. M. van Lieshout), Chapman and Hall/CRC, London, 37–78.

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

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

Baddeley, A. J. and Turner, R. (2000). Practical maximum pseudolikelihood for spatial point patterns. *Australian and New Zealand Journal of Statistics* To appear.

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

Baddeley, A. J., Van Lieshout, M. N. M. and Møller, J. (1995). Markov properties of cluster processes. *Adv. Appl. Prob. (SGSA)* **28**, 346–355.

Berman, M. and Turner, R. T. (1992). Approximating point process likelihoods using GLIM. *Appl. Statist.* **41**, 31–38.

Besag, J. E. (1977). Discussion contribution to Ripley (1977). *J. R. Statist. Soc.* **39**, 193–195.

Brix, A. and Møller, J. (1998). Space-time multi type log Gaussian Cox processes with a view to modeling weed data. *Research Report R-98-2012*, Department of Mathematics, Aalborg University, submitted.

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.

Gelfand, A. E. and Carlin, B. P. (1991). Maximum likelihood estimation for constrained or missing data models. *Research Report 91-002*, Division of Biostatistics, University of Minnesota.

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

Geyer, C. J. (1994). On the convergence of Monte Carlo maximum likelihood calculations. *J. R. Statist. Soc.* **56**, 261–274.

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

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

Goulard, M., Grabarnik, P. and Särkkä, A. (1996). Parameter estimation for marked Gibbs point processes through the maximum pseudo-likelihood method. *Scand. J. Statist.* **23**, 365–379.

Grabarnik, P. and Särkkä, A. (1999). Interacting neighbour point processes. Submitted.

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

Møller, J., Syversveen, A.-R. and Waagepetersen, R. (1998). Log Gaussian Cox processes. *Scand. J. Statist.* **25**, 451–482.

Nguyen, X. X. and Zessin, H. (1979). Integral and differential characterizations of the Gibbs point process. *Math. Nachr.* **88**, 105–115.

Nielsen, L. S. and Jensen, E. B. V. (1999). Inhomogeneous Markov point processes by transformation. *Bernouilli* To appear.

Numata, M. (1964). Forest vegetation, particularly pine stems in the vicinity of Choshi-flora and vegetation at Choshi, Chiba prefecture IV. *Bull. Choshi Marine Laboratory* (6), 27–37.

Ogata, Y. and Tanemura, M. (1985). Estimation of interaction potentials of marked spatial point patterns through the maximum likelihood method. *Biometrics* **41**, 421–433.

Ogata, Y. and Tanemura, M. (1986). Likelihood estimation of interaction potentials and external fields of inhomogeneous spatial point patterns. In: *Proc. Pacific Statistical Congress-1985* (eds. I. S. Francis, B. F. J. Manly and F. C. Lam), 150–154.

Platt, W. J., Evans, G. W. and Rathbun, S. L. (1988). The population dynamics of a long-lived Conifer (*Pinus palustris*). *The American Naturalist* **131**, 491–525.

Preston, C. (1976). *Random Fields*. Springer-Verlag, Berlin.

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

Ripley, B. D. and Kelly, F. P. (1977). Markov point processes. *J. Lond. Math. Soc.* **15**, 188–192.

Särkkä, A. and Baddeley, A. J. (1998). Parameter estimation and identifiability for the Widom-Rowlinson process. In preparation.

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

Stoyan, D. and Stoyan, H. (1998). Non-homogeneous Gibbs Process Models for Forestry - A Case Study. *Biometrical Journal* **40**, 521–531.

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