Statistical inference for transformation inhomogeneous point processes

Linda Stougaard Nielsen and Eva B. Vedel Jensen
Laboratory for Computational Stochastics
University of Aarhus

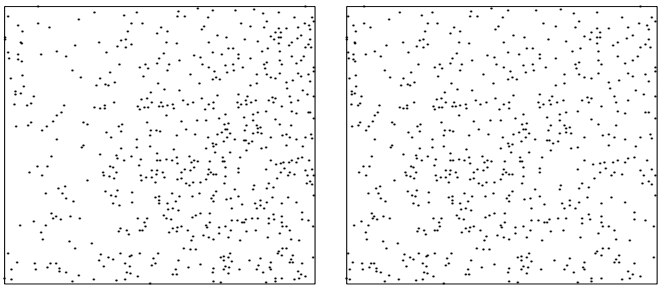
Abstract

Statistical inference for exponential inhomogeneous Markov point processes by transformation is discussed. It is argued that the inhomogeneity parameter can be estimated, using a partial likelihood based on an inhomogeneous Poisson point process. The inhomogeneity parameter can thereby be estimated without taking the interaction into account, which simplifies the statistical analysis considerably. Furthermore, an easily computable test for homogeneity is presented. Analysis of two data sets and simulation experiments support the results.

1. Introduction

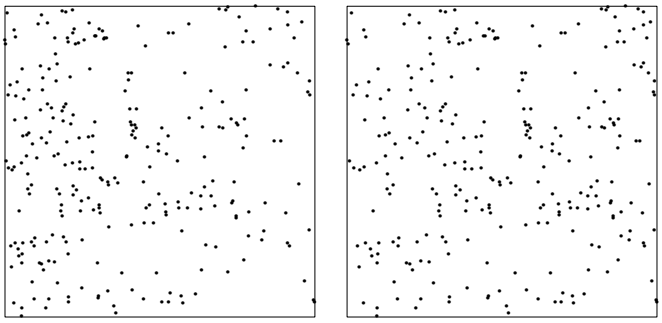
Various point process models allowing for both interaction and inhomogeneity have recently been suggested, cf. Baddeley et al. (2000), Brix and Møller (1998), Hahn et al. (2001), Jensen and Nielsen (2000, 2001) and Stoyan and Stoyan (1998). See also Ogata and Tanemura (1986). A majority of these model classes uses a homogeneous point process as starting point. The inhomogeneity is introduced by letting the first order interaction be location dependent or by applying a thinning, a transformation or a local scaling of the homogeneous process. The target is modelling of data such as the cell point pattern shown in Figure 1 (a) and the longleaf point pattern in Figure 2 (a). Both point patterns clearly have a trend along the first axis. Furthermore, there might be small-scale inhibition in the cell point pattern and clustering in the longleaf point pattern.

In the present paper we focus on parametric likelihood inference for inhomogeneous Markov point processes by transformation. This model class was introduced in Jensen and Nielsen (2000) and will henceforth be denoted TIM models (**T**ransformation, **I**nhomogeneous, **M**arkov). A parametrized TIM model is obtained by applying parametrized transformations $\{h_\theta : \theta \in \Theta\}$ to a homogeneous Markov point process with density parametrized by $\psi \in \Psi$, say. The inhomogeneity is introduced through the transformation while the interaction originates from the underlying homogeneous model. Accordingly, ψ is called the interaction parameter while θ is called the inhomogeneity



(a) Cell profile centres (b) Back-transf. cells

Figure 1: Point pattern y of cell profile centres (a) and the corresponding back-transformed point pattern $x = h_\theta^{-1}(y)$ (b). The transformation only affects the first coordinates of the points. The transformation parameter θ is the maximum likelihood estimate, $\hat{\theta}_0 = 1.3043$, under an inhomogeneous Poisson point process model. For details, see Section 3.4.



(a) Pine trees (b) Back-transf. pines

Figure 2: Point pattern y of positions of adult longleaf pine trees in a forest (a) and the corresponding back-transformed point pattern $x = h_\theta^{-1}(y)$ (b). The transformation parameter θ is the maximum likelihood estimate, $\hat{\theta}_0 = -1.38663$, under an inhomogeneous Poisson point process model.

parameter. However, in the transformed process both ψ and θ control the interaction. The transformed process is inhomogeneous in the intensity of the points as well as in the strength of the interactions among the points. In particular, the interaction range is shorter in areas where the concentration of points is high.

For any of the inhomogeneous models to be considered, there exists a subset $\Psi_0 \subseteq \Psi$ in which the model is Poisson. The transformation model with parameter space $\Psi_0 \times \Theta$ is a subclass consisting of inhomogeneous Poisson point processes, called the corresponding inhomogeneous Poisson model.

Likelihood inference for θ and ψ cannot be separated. Furthermore, for the important class of exponential inhomogeneous transformation models, the parameter θ appears as a nuisance parameter in an exponential family likelihood function, cf. Jensen and Nielsen (2000). Since the number of calculations involved in maximum likelihood estimation increases exponentially in the dimension of the nuisance parameters and maximum likelihood estimation is already quite involved in the homogeneous case, it is desirable to find an alternative estimate of θ.

In Figure 1 (b) and 2 (b) the cell and the longleaf point patterns have been transformed with an inverse transformation h_θ^{-1}, which is of simple exponential form and only affects the first coordinates of the points. The value of the inhomogeneity parameter θ is $\hat{\theta}_0$, the estimate based on the corresponding inhomogeneous Poisson model. Note that $\hat{\theta}_0$ is calculated under a model where the interaction is disregarded and only the inhomogeneity is taken into account. Motivated by the homogeneous appearance of these point patterns, the present paper is devoted to the study of the statistical properties of $\hat{\theta}_0$. In particular, it will be shown that $\hat{\theta}_0$ can be regarded as a moment estimator in the class of exponential inhomogeneous transformation models.

With this simplified estimation of the inhomogeneity parameter θ, the analysis of an exponential transformation model can be performed as follows. First $\hat{\theta}_0$ is computed, which is easy and very fast to do in practice. Next, ψ is estimated under the assumption that θ is known and equals $\hat{\theta}_0$. Using this two-step estimation procedure, we can first concentrate on finding the appropriate transformation without taking the interaction into account. Secondly, we can try to find a homogeneous model, using the back-transformed data $h_{\hat{\theta}_0}^{-1}(y)$ where y is the original inhomogeneous point pattern. For this purpose, well-studied tools can be used such as second order statistics F, G, J, K and the pair correlation function, see e.g. Diggle (1983), van Lieshout and Baddeley (1996) and Stoyan et al. (1995). The analysis of homogeneous Markov point process models is also a very thoroughly studied field, see e.g. Geyer (1999) and Baddeley and Turner (2000).

The rest of the paper is organised as follows. Basic terminology for point processes is given in Section 2. In Section 3 we briefly introduce the transformation models and the class of exponential transformations. In Section 4 we discuss full likelihood inference in transformation models. In Section 5 we investigate the statistical properties of $\hat{\theta}_0$. A simple test for homogeneity based on the Poisson model is given in Section 6. The focus is in Sections 3 to 6 on the class of exponential transformations, but it is pointed out when results apply to point processes in general. At the end of each of Sections 3 to 6, the results are illustrated by analysis of the cell or the longleaf point patterns. A supplementary simulation study is presented in Section 7. In Section 8, open questions and future work are discussed. The Appendix contains some prerequisite results for point processes with periodic boundary.

2. Point processes

2.1. Homogeneous point processes

In the present paper we consider finite point processes defined on a full-dimensional bounded subset $\mathcal{X}$ of $\mathbb{R}^m$. Below, we summarize the notation and concepts needed for such processes. A more detailed account can be found in Møller (1999) or van Lieshout (2000).

The state space for a finite point process on $\mathcal{X}$ is $\Omega_{\mathcal{X}}$, the set of finite subsets of $\mathcal{X}$. In what follows, we only consider point processes which have a density with respect to the Poisson point process on $\mathcal{X}$ with intensity measure λ_m, the Lebesgue measure in $\mathbb{R}^m$.

A point process defined on $\mathbb{R}^m$ is called homogeneous (or stationary) if its distribution is invariant under translation, cf. e.g. Stoyan et al. (1995) or van Lieshout (2000). For a point process defined on a bounded set, this concept can be modified as follows.

Definition 2.1 *Let X be a point process on a full-dimensional bounded set $\mathcal{X} \subseteq \mathbb{R}^m$ with density f with respect to the Poisson point process on $\mathcal{X}$ with intensity measure λ_m. We call X homogeneous if f is the restriction to $\Omega_{\mathcal{X}}$ of a function g defined on $\Omega_{\mathbb{R}^m}$ which is translation invariant, i.e. $g(x+c) = g(x)$ for all $c \in \mathbb{R}^m$ and $x \in \Omega_{\mathbb{R}^m}$. Here, $x + c = \{\eta + c : \eta \in x\}$.*

2.2. Markov point processes

The class of Markov point processes, see Ripley and Kelly (1977), will be used for modelling the interaction in a point pattern. Let $\sim$ be a reflexive and symmetric relation on $\mathcal{X}$. A point process is Markov w.r.t. $\sim$ if and only

if

$$f(x) = \prod_{z \subseteq x} \varphi(z), \quad x \in \Omega_{\mathcal{X}}, \tag{1}$$

where φ is a clique interaction function w.r.t. $\sim$. This means that if $\varphi(z) \neq 1$ then all pairs of points in the subset z are related with respect to $\sim$. Often, it is possible to extend the definition of φ to $\Omega_{\mathbb{R}^m}$. In that case, X is homogeneous if φ is translation invariant, cf. Definition 2.1.

Example 2.2 (Strauss process) A simple example of a homogeneous Markov point process is the Strauss process with density, cf. Strauss (1975),

$$f(x) = c(\beta, \gamma, r)^{-1} \beta^{n(x)} \gamma^{s_r(x)}, \quad x \in \Omega_{\mathcal{X}},$$

where $n(x)$ is the number of points in x, $s_r(x)$ is the number of point pairs in x with distance less than r, and $c(\beta, \gamma, r)^{-1}$ is the normalising constant. The parameters fulfil $\beta > 0$, $0 < \gamma \leq 1$, and $r > 0$.

The Strauss process is Markov w.r.t. the distance relation

$$\eta \sim \xi \Longleftrightarrow \|\eta - \xi\| < r. \tag{2}$$

2.3. Processes with periodic boundary

For a homogeneous point process defined on a bounded set, the distribution of the points at the boundary of $\mathcal{X}$ is typically slightly different from the distribution elsewhere. This phenomenon is known as edge effects. In the theoretical developments presented in Section 5 below, it is important to remove these edge effects. One way is to restrict attention to a set $\mathcal{X}$ which can tile $\mathbb{R}^m$ and modify the density of a homogeneous process such that it becomes $\mathcal{X}$-periodic.

Definition 2.3 *A bounded set $\mathcal{X} \subseteq \mathbb{R}^m$ is a fundamental region if there exists a sequence $\{z_j\} \subseteq \mathbb{R}^m$ such that*

$$\begin{aligned} &\cup_j \, (\mathcal{X} + z_j) = \mathbb{R}^m \\ &(\mathcal{X} + z_{j_1}) \cap (\mathcal{X} + z_{j_2}) = \emptyset \quad \textit{when} \quad j_1 \neq j_2 \\ &\{-z_j\} = \{z_j\} \end{aligned}$$

Example 2.4 Let $\mathcal{X} = [a_1, b_1) \times \cdots \times [a_m, b_m)$ be a rectangular box in $\mathbb{R}^m$. Then $\mathcal{X}$ is a fundamental region and the series

$$\{(j_1(b_1 - a_1), \ldots, j_m(b_m - a_m))\}_{(j_1, \ldots, j_m) \in \mathbb{Z} \times \cdots \times \mathbb{Z}}$$

fulfils the requirements in Definition 2.3.

Definition 2.5 *Let $\mathcal{X}$ be a fundamental region and let $\{z_j\}$ be a sequence fulfilling the requirements in Definition 2.3. A function $g : \Omega_{\mathbb{R}^m} \to \mathbb{R}$ is $\mathcal{X}$-periodic if for all i and j*

$$g(\{x_1, \ldots, x_{i-1}, x_i + z_j, x_{i+1}, \ldots, x_n\}) = g(\{x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n\}).$$

If $\mathcal{X}$ is an interval, one may imagine wrapping $\mathcal{X}$ around a circle, combining the two ends of the interval. A function g defined on all finite subsets of $\mathbb{R}$ is then $\mathcal{X}$–periodic if the values of g only depend on the wrapped point configuration. If the function g associated with a homogeneous process, cf. Definition 2.1, is $\mathcal{X}$–periodic, then X will be called circular. If $\mathcal{X}$ is a rectangle in $\mathbb{R}^2$, $\mathcal{X}$–periodicity involves folding $\mathcal{X}$ into a torus.

Using $\mathcal{X}$–periodicity, we can remove the edge effects, as shown in the proposition below. The proof of the proposition is deferred to the Appendix.

Proposition 2.6 *Let $\mathcal{X} \subseteq \mathbb{R}^m$ be a fundamental region and let X be a homogeneous point process on $\mathcal{X}$ as in Definition 2.1, where the function g is $\mathcal{X}-$periodic, cf. Definition 2.5. Furthermore, let us assume that $n(X) > 0$ almost surely.*

Then, a point Z chosen uniformly among the points in X, is uniformly distributed in $\mathcal{X}$.

As a consequence of Proposition 2.6, the mean value of averages over points in the homogeneous point process X does not depend on the interaction structure.

Corollary 2.7 *Let the situation be as in Proposition 2.6. Furthermore, let $q : \mathcal{X} \to \mathbb{R}^l$ and let U denote a uniform random variable in $\mathcal{X}$. Then,*

$$\mathbb{E}\left(\frac{1}{n(X)}\sum_{\eta\in X} q(\eta)\right) = \mathbb{E}\, q(U).$$

Proof. Let Z be chosen uniformly among the points in X. Then,

$$\mathbb{E}\,(q(Z)|X) = \frac{1}{n(X)}\sum_{\eta\in X} q(\eta).$$

Hence,

$$\mathbb{E}\left(\frac{1}{n(X)}\sum_{\eta\in X} q(\eta)\right) = \mathbb{E}\,\left(\mathbb{E}\,(q(Z)|X)\right) = \mathbb{E}\, q(Z) = \mathbb{E}\, q(U).$$

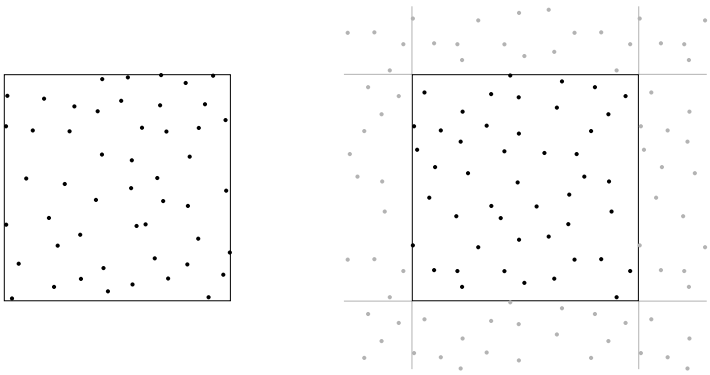


Figure 3: Realizations of the Strauss process on the unit square. To the left, the relation is the distance relation. To the right, the relation is the distance relation modified to be $\mathcal{X}$-periodic. To illustrate the periodicity, the point pattern has been translated with $\{(i,j)\}_{(i,j)\in\mathbb{Z}\times\mathbb{Z}}$ and the nearest neighbourhood is plotted in grey. In both point patterns, $\gamma = 0.01$, $\beta = 500$ and $r = 0.1$.

At the last equality sign we have used that Z is uniformly distributed in $\mathcal{X}$ as shown in Proposition 2.6. □

In the right hand-side of Figure 3, an $\mathcal{X}$–periodic version of the Strauss process is shown where $\mathcal{X}$ is the unit square. Instead of using the ordinary relation $\sim$, defined in (2), a modified relation is used

$$\eta \sim_{\mathrm{p}} \xi \iff \exists j_1, j_2 : \ \eta + z_{j_1} \sim \xi + z_{j_2}. \tag{3}$$

For comparison, a realization of the ordinary Strauss process is shown in the left hand-side of Figure 3. Notice that the edge-effects are removed when the relation is modified.

3. Inhomogeneous point processes by transformation

In this section, we summarize the important concepts from the theory of inhomogeneous point processes by transformation. For more details, see Jensen and Nielsen (2000).

3.1. Transformation of point processes

Let $h : \mathcal{X} \to \mathcal{Y}$ be a differentiable and bijective mapping between two full-dimensional bounded subsets of $\mathbb{R}^m$, and let X be a point process on $\mathcal{X}$ with density f_X. Then the transformed point process $Y = h(X)$ has density

$$f_Y(y) = f_X(h^{-1}(y))\mathrm{e}^{\lambda_m(\mathcal{Y})-\lambda_m(\mathcal{X})} \prod_{\eta\in y} Jh^{-1}(\eta), \quad y \in \Omega_{\mathcal{Y}}, \tag{4}$$

see Jensen and Nielsen (2000, Proposition 3.2). Here, Jh^{-1} is the Jacobian of the inverse transformation h^{-1}.

If Jh^{-1} is non-constant and X is a homogeneous point process, it follows that Y is an inhomogeneous point process, cf. Definition 2.1. Such processes are called transformation inhomogeneous.

Let X be the homogeneous Poisson point process with intensity measure $c\lambda_m$ where $c > 0$. Then Y has density

$$\begin{aligned} f_Y(y) &= \mathrm{e}^{-(c-1)\lambda_m(\mathcal{X})} c^{n(y)} \mathrm{e}^{\lambda_m(\mathcal{Y})-\lambda_m(\mathcal{X})} \prod_{\eta\in y} Jh^{-1}(\eta) \\ &\overset{(*)}{=} \mathrm{e}^{-\int_{\mathcal{Y}}(c\,Jh^{-1}(\eta)-1)d\eta} \prod_{\eta\in y} \left(c\,Jh^{-1}(\eta)\right), \end{aligned}$$

which is the density of an inhomogeneous Poisson point process with intensity function $c\,Jh^{-1}(\eta)$. At $(*)$ we have used the fact that

$$\int_{\mathcal{Y}} Jh^{-1}(\eta)d\eta = \lambda_m(\mathcal{X}). \tag{5}$$

3.2. TIM models

The transformation result from the previous section can be used to develop models for inhomogeneous point processes Y on $\mathcal{Y}$. Let X be a homogeneous Markov point process with respect to a relation $\sim$ and with density f parametrized by $\psi \in \Psi$. Furthermore, let $g_\theta : \mathcal{Y} \to [0, \infty)$ be a parametrized model of the inhomogeneity where $\theta \in \Theta \subseteq \mathbb{R}^l$. Suppose that we can find for each θ a differentiable and bijective transformation $h_\theta : \mathcal{X} \to \mathcal{Y}$ such that

$$Jh_\theta^{-1}(\eta) = g_\theta(\eta), \quad \eta \in \mathcal{Y}. \tag{6}$$

Then, $Y = h_\theta(X)$ is a Markov point process with respect to the induced relation $\approx$ given by

$$\eta \approx \xi \iff h_\theta^{-1}(\eta) \sim h_\theta^{-1}(\xi), \tag{7}$$

cf. Jensen and Nielsen (2000, Corollary 3.3). We can think of $\approx$ as an inhomogeneous version of $\sim$.

The model for Y is called a TIM (transformation inhomogeneous Markov) model. The inhomogeneity in the model is induced by $\{g_\theta; \theta \in \Theta\}$. The first order terms in the density of Y are non-constant and proportional to $g_\theta(\eta)$, cf. (4) and (6), and the relation is inhomogeneous and determined by the solution to the differential equation (6), cf. (7).

3.3. Exponential transformations

Let us restrict attention to the case where $\mathcal{Y} = \mathcal{X}$. We will call a transformation $h_\theta : \mathcal{X} \to \mathcal{X}$ *exponential* when its inverse Jacobian is of exponential form

$$Jh_\theta^{-1}(\eta) = \beta(\theta)\mathrm{e}^{\theta\cdot\tau(\eta)}, \quad \eta \in \mathcal{X}, \tag{8}$$

where $\tau : \mathcal{X} \to \mathbb{R}^l$ and $\theta \in \Theta \subseteq \mathbb{R}^l$. Using (5) we get,

$$\beta(\theta) = \lambda_m(\mathcal{X})\alpha(\theta), \quad \text{where} \quad \alpha(\theta)^{-1} = \int_{\mathcal{X}} \mathrm{e}^{\theta\cdot\tau(u)}du. \tag{9}$$

Notice that $\theta = 0$ is the case of the identity transformation, $h_0(\eta) = \eta$.

By combining (1), (4), and (8), we get in the particular case where X is a Markov point process,

$$f_Y(y;\theta) = \beta(\theta)^{n(y)}\mathrm{e}^{\theta\cdot t(y)} \prod_{z\subseteq h_\theta^{-1}(y)} \varphi(z), \quad \text{where} \quad t(y) = \sum_{\eta\in y} \tau(\eta).$$

The model for Y is called an exponential inhomogeneous Markov model.

Example 3.1 (1–dimensional exponential transformations)
Let $I \subseteq \mathbb{R}$ be a bounded interval. There exists a unique differentiable and increasing mapping h_θ of I onto itself, $\theta \in \mathbb{R}^l$, such that

$$\frac{d}{du}h_\theta^{-1}(u) \propto \mathrm{e}^{\theta\cdot\tau(u)}.$$

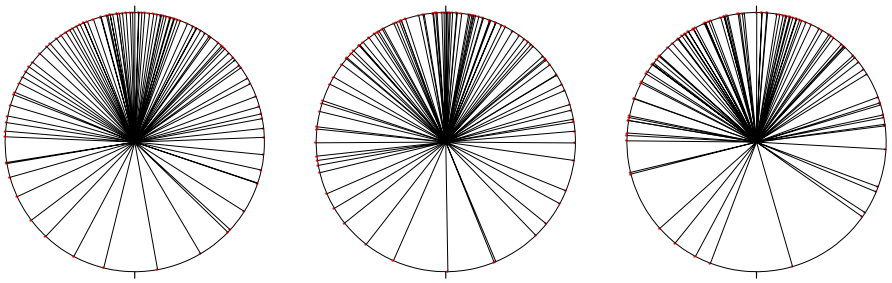
In particular, when $\tau(u) = u$ and $I = (0;a)$ we obtain the *simple exponential transformation* given by

$$h_\theta^{-1}(u) = a\frac{\mathrm{e}^{\theta\, u} - 1}{\mathrm{e}^{\theta\, a} - 1},$$

$\theta \in \mathbb{R}$. Another example is the *symmetric exponential transformation* with $\tau(u) = |u|$ and $I = (-a;a)$. It is closely related to the simple transformation and takes the form

$$h_\theta^{-1}(u) = \mathrm{sign}(u)\, a\, \frac{\mathrm{e}^{\theta\,|u|} - 1}{\mathrm{e}^{\theta\, a} - 1}.$$

Since $-a \to -a$, $a \to a$ and $d/du\ h_\theta^{-1}(a) = d/du\ h_\theta^{-1}(-a)$, we can think of this transformation as a differentiable mapping of a circle into itself. In Figure 4 we have used this transformation on the circular Strauss point process on $(-\pi;\pi)$ with modified relation (3) where $\sim$ is the usual distance relation. To emphasize the periodicity, the point patterns are plotted on the unit circle where the points $-\pi$ and π are identified. This example will be used in all the simulation experiments we will present in this paper and the process will be denoted symmetric exponential inhomogeneous circular Strauss.



(a) $\gamma = 0.1$ (b) $\gamma = 0.5$ (c) $\gamma = 1$

Figure 4: Symmetric exponential inhomogeneous circular Strauss. Fixed number of points $n = 100$. Circular distance relation with $r = 0.05$, and γ as indicated. Symmetric exponential transformation with $\theta = 1$, the mark on the bottom indicates the points $-\pi = \pi$.

3.4. Data analysis, part 1

The cell data shown in Figure 1 (a) is from a tissue section of the mucous membrane in the stomach of a healthy rat. The data have earlier been analysed in Nielsen (2000). The original image of the section has been converted into points marking the centres of the cell profiles. It is known that the variation in size is small for these cells. We only consider a small window of the original data and scale it such that $\mathcal{X} = [0, 1] \times [0, 0.89]$. There is a trend in the cell intensity perpendicular to the stomach wall, and the section has been taken along this trend. In Figure 1 (a), the trend is along the first axis.

In Figure 2 (a) the points mark 271 adult longleaf pine trees observed in a $200 \times 200m^2$ area of a forest, which for convenience is rescaled to the unit square. The data set was first studied in Platt et al. (1988).

One possible model for the data sets is a transformation model with coordinate-wise transformation $h_\theta(\eta_1, \eta_2) = (h_{\theta_1}(\eta_1), h_{\theta_2}(\eta_2))$ where h_{θ_1} is a simple exponential transformation parametrized by $\theta_1 \in \mathbb{R}$, cf. Example 3.1, and h_{θ_2} is chosen as the identity. If the underlying homogeneous process is Poisson, the point patterns are regarded as realizations from inhomogeneous Poisson point processes. As mentioned in the introduction, the maximum likelihood estimate of θ under this model is very easy to calculate and will be denoted $\hat{\theta}_0$. See also Section 5.1 below.

In Figure 1 (b) and 2 (b) the two data sets have been back-transformed using $h^{-1}_{\hat{\theta}_0}$. Both data sets appear to be homogeneous. Summary statistics calculated for the back-transformed data sets show that the pine data is almost Poisson with a small tendency of clustering. The cell data show significant small scale inhibition. In Nielsen (2000) a TIM model with the

Strauss process as underlying homogeneous interaction model was used with success. We might also choose to model the clustering in the pine data. One of the models introduced by Geyer (1999) might be applicable.

In the sections to follow we will discuss likelihood inference for transformation models and only use the cell data for illustrations.

4. Full likelihood inference

Let $h_\theta : \mathcal{X} \to \mathcal{Y}$ be a differentiable and bijective mapping parametrized by $\theta \in \Theta$. Let X be a homogeneous point process with density $f_X(\cdot;\psi)$ parametrized by $\psi \in \Psi$. Let $Y = h_\theta(X)$ and let y be an observed point pattern. We want to estimate θ and ψ in the transformation model class for Y.

The likelihood function for (θ, ψ) decomposes as follows, cf. (4),

$$L(\theta, \psi; y) = \; L_0(\theta; y) L_{\hom}(\psi; h_\theta^{-1}(y)), \tag{10}$$

where $L_0(\theta; y)$ is the likelihood function for the inhomogeneous Poisson point process with intensity function equal to Jh_θ^{-1} and $L_{\hom}(\psi; x) = f_X(x;\psi)$ is the likelihood function for the corresponding homogeneous model when x is observed.

Since the inhomogeneity parameter θ enters into both parts of the likelihood decomposition (10), traditional likelihood inference on θ cannot be restricted to L_0.

4.1. Exponential family densities and profile likelihood

Likelihood inference for point processes is tractable when the density is of exponential family form, see e.g. Geyer (1999). However, most homogeneous point process models studied do not have density of exponential form, but have the property that the parameter ψ can be split into two components (ψ_1, ψ_2) such that the density is of exponential form for fixed ψ_2,

$$f_X(x; (\psi_1, \psi_2)) = c(\psi_1, \psi_2)^{-1} \mathrm{e}^{\phi(\psi_1) \cdot t(x; \psi_2)}, \quad x \in \Omega_{\mathcal{X}}. \tag{11}$$

An example is the Strauss process with distance relation determined by r where $\psi_1 = (\beta, \gamma)$, $\psi_2 = r$, $\phi(\beta, \gamma) = (\log(\beta), \log(\gamma))$ and $t(x; r) = (n(x), s_r(x))$, see also Example 2.2.

If X has density of the form (11), then the density of Y contains the term

$$\mathrm{e}^{\phi(\psi_1) \cdot t(h_\theta^{-1}(y); \psi_2)},$$

which is not an exponential family term unless ψ_2 as well as θ are fixed. It does not make sense to introduce new homogeneous models such that θ enters

into the density of Y as an exponential family parameter. The reason is, that the very idea of the transformation models is to create inhomogeneous models based on the homogeneous models broadly studied, and to take advantage of the already well developed tools for statistical inference for homogeneous models.

However, the properties of exponential families can still be utilized in the analysis of a transformation model. When X has density of the form (11), the profile likelihood with nuisance parameter (θ, ψ_2) becomes

$$\overline{L}(\theta, \psi_2; y) = L_0(\theta; y) \max_{\psi_1} L_{\hom}(\psi_1, \psi_2; h_\theta^{-1}(y)),$$

where $L_{\hom}(\psi_1, \psi_2; h_\theta^{-1}(y))$ is of exponential family form for fixed (θ, ψ_2).

4.2. Data analysis, part 2

Let us illustrate the problems involved in full likelihood inference by the cell data. We use the Strauss process as the underlying homogeneous process. The inhomogeneity is described by a coordinate-wise transformation $h_\theta(\eta_1, \eta_2) = (h_{\theta_1}(\eta_1), h_{\theta_2}(\eta_2))$, $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$, where both coordinate mappings are simple exponential, cf. Example 3.1. The nuisance parameter (θ, r) is 3–dimensional. In Jensen and Nielsen (2000, Example 5.2), a simpler example based on simulated data was studied. Here the nuisance parameter was 1–dimensional.

For fixed θ and r, $L_{\hom}(\beta, \gamma, r; h_\theta^{-1}(y))$ is of exponential family form and its maximum with respect to (β, γ) is attained as the unique solution to the likelihood equations

$$\begin{array}{rcl} \mathbb{E}_{\beta,\gamma,r}\, n(X) & = & n(y) \\ \mathbb{E}_{\beta,\gamma,r}\, s_r(X) & = & s_r(h_\theta^{-1}(y)), \end{array} \tag{12}$$

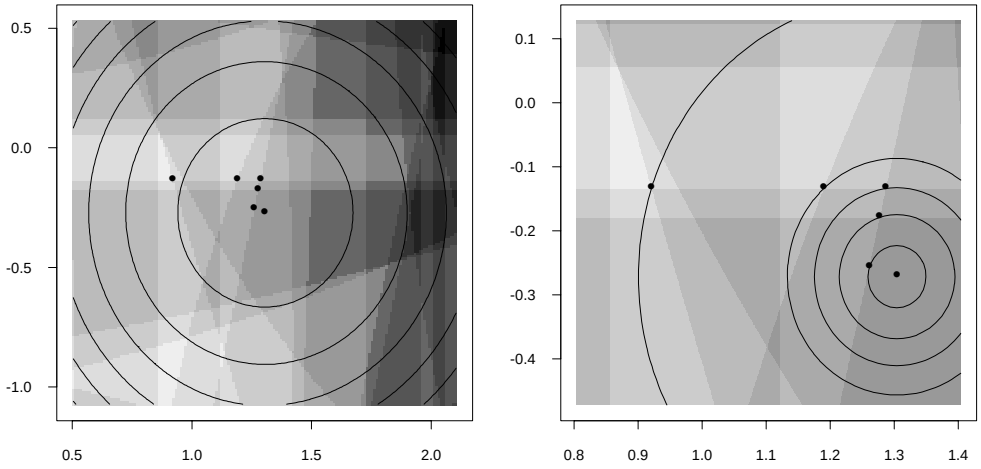
where X is a Strauss process with parameters (β, γ, r). The solution is denoted $(\hat{\beta}(\theta, r), \hat{\gamma}(\theta, r))$. The profile likelihood with nuisance parameter (θ, r) becomes

$$\overline{L}(\theta, r; y) = L_0(\theta; y)\, L_{\hom}(\hat{\beta}(\theta, r), \hat{\gamma}(\theta, r), r; h_\theta^{-1}(y)). \tag{13}$$

Hence, for values of (θ, r) in a grid we solve the equations (12) and compute (13) up to a constant. The grid value maximising the profile likelihood will be denoted $(\hat{\theta}, \hat{r})$ and the maximum likelihood estimate is then $(\hat{\theta}, \hat{r}, \hat{\beta}, \hat{\gamma}) = (\hat{\theta}, \hat{r}, \hat{\beta}(\hat{\theta}, \hat{r}), \hat{\gamma}(\hat{\theta}, \hat{r}))$.

The set of $\theta-$values to be considered can be reduced quite drastically. If for $k = 0, 1, \ldots$

$$\Theta_k(r) = \{\theta \in \Theta : s_r(h_\theta^{-1}(y)) = k\},$$



(a) Coarse θ-grid (b) Fine θ-grid

Figure 5: Illustration of $T(r)$ for the cell data with $r = 0.0087$. The tabbing between grid points is 0.01 and 0.001 in the left and right figure, respectively. The grey values correspond to the different $\Theta_k(r)$ regions, the lighter the grey value the smaller the k. The curves are level curves of L_0.

then, according to Jensen and Nielsen (2000, Proposition 5.1), $\overline{L}(\theta, r; y)$ attains its maximum for fixed r for $\theta \in T(r)$ where

$$T(r) = \bigcup_{k=0}^{s_0(r)} \{\theta \in \Theta_k(r) : L_0(\theta; y) \geq L_0(\theta^*; y), \text{ for all } \theta^* \in \Theta_k(r)\},$$

and $s_0(r) = s_r(h_{\hat{\theta}_0}^{-1}(y))$. The set $T(r)$ is a finite set with at most $s_0(r) + 1$ elements. An illustration is shown for the cell data in Figure 5.

It therefore suffices to tabulate the profile likelihood function $\overline{L}(\theta, r; y)$ in

$$\{(\theta, r) : \theta \in T(r),\ r \in R\},$$

where R is the chosen grid of $r-$values. This can be done using multiple bridge sampling, cf. Gelman and Meng (1998). Details are provided in the technical report Nielsen (2001). In Figure 6, the profile log-likelihood function $\log \overline{L}(r; y)$ is plotted for the cell data where

$$\overline{L}(r; y) = \max_{\theta \in T(r)} \overline{L}(\theta, r; y)$$

and

$$R = \{0.00500, 0.00505, \ldots, 0.01000\}.$$

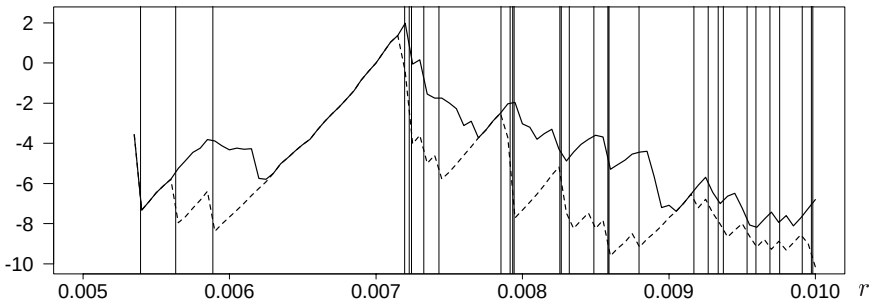


Figure 6: The profile log-likelihood function $\log \overline{L}(r; y)$ is plotted (in full) for the cell data. The stippled line is the partial profile log-likelihood when $\theta = \hat{\theta}_0$ is fixed. For details, see the text.

Notice that $\overline{L}(\cdot; y)$ is multi-modal. For $r \in R$ the maximum likelihood estimates are

$$(\hat{\theta}_1, \hat{\theta}_2, \hat{r}, \hat{\beta}, \hat{\gamma}) = (1.304, -0.275, 0.0072, 767.6, 0.08149)$$

The choice of the grid R has been based on analysis of the J–function and the profile pseudo-likelihood of the back-transformed data set from Figure 1 (b) which appears homogeneous, cf. van Lieshout and Baddeley (1996) and Baddeley and Turner (2000). These analyses point to a value of r about $r_0 = 0.007$. The nearest neighbour distances are very small in the back-transformed cell point pattern. The smallest observed nearest neighbour distance is 0.00539. The 25%, 50%, 75%, and 100% quantiles are 0.01370, 0.01869, 0.02463, and 0.04806, respectively.

To be precise, the profile log-likelihood function plotted in Figure 6 is

$$\log \overline{L}(r; y) - \log \overline{L}(\hat{\theta}_0, r_0; y) = \log \frac{\overline{L}(\hat{\theta}_0, r; y)}{\overline{L}(\hat{\theta}_0, r_0; y)} + \max_{\theta \in T(r)} \log \frac{\overline{L}(\theta, r; y)}{\overline{L}(\hat{\theta}_0, r; y)}. \quad (14)$$

The first term on the right hand-side of (14) is a partial profile log-likelihood for $\theta = \hat{\theta}_0$ fixed. The partial likelihood is increasing in intervals between the inner-point distances in the back-transformed point pattern and has downwards jumps at the inner-point distances. These distances are marked as vertical lines in Figure 6 and the partial likelihood is the stippled line. The full and the partial log-likelihood are identical below the smallest inner-point distance 0.00539, since $T(r) = \{\hat{\theta}_0\}$ for $r \leq 0.00539$. Therefore, in order to find the maximum of $\overline{L}(\cdot; y)$, the partial log-likelihood need only to be evaluated in the largest r grid value below the smallest inner-point distance.

5. Partial likelihood inference for θ

In this section we investigate the statistical properties of the estimator $\hat{\theta}_0$ obtained by maximising the Poisson likelihood L_0 in (10). We restrict attention to exponential transformations as introduced in Section 3.3 and we assume that $\mathcal{Y} = \mathcal{X}$.

5.1. Existence and uniqueness of $\hat{\theta}_0$
The inhomogeneous Poisson likelihood takes the form

$$L_0(\theta; y) = \lambda_m(\mathcal{X})^{n(y)} \alpha(\theta)^{n(y)} \mathrm{e}^{\theta \cdot t(y)} = \lambda_m(\mathcal{X})^{n(y)} \left(\alpha(\theta) \mathrm{e}^{\theta \cdot \frac{t(y)}{n(y)}} \right)^{n(y)},$$

see (8) and (9). Recall that $t(y) = \sum_{\eta \in y} \tau(\eta)$. If $n(y) > 0$, then the likelihood equation for θ based on L_0 is

$$\frac{t(y)}{n(y)} = -\frac{\frac{d}{d\theta}\alpha(\theta)}{\alpha(\theta)} =: m(\theta).$$

Existence and uniqueness of $\hat{\theta}_0$ follow from the theory of exponential families, cf. e.g. Barndorff-Nielsen (1978, Corollary 9.6). The results are formulated in the proposition below.

Proposition 5.1 *Let X be a homogeneous point process on a bounded set $\mathcal{X} \subseteq \mathbb{R}^m$ and suppose that $n(X) > 0$ almost surely. Let $Y = h_\theta(X)$ where h_θ, $\theta \in \Theta$, is an exponential transformation. Suppose that the densities $\{\alpha(\theta) e^{\theta \cdot \tau(\cdot)} : \theta \in \Theta\}$ constitute a regular exponential family. Then, if C is the convex support of the family, then m is a bijection of Θ onto $\mathrm{int} C$. For $t(y)/n(y) \in \mathrm{int} C$, $\hat{\theta}_0$ exists and is given by*

$$\hat{\theta}_0 = m^{-1} \left(\frac{t(y)}{n(y)} \right). \tag{15}$$

Note that m can be calculated as

$$m(\theta) = \frac{\int_{\mathcal{X}} \tau(u) \mathrm{e}^{\theta \cdot \tau(u)} du}{\int_{\mathcal{X}} \mathrm{e}^{\theta \cdot \tau(u)} du}. \tag{16}$$

5.2. Statistical properties of $\hat{\theta}_0$
Below, we show under regularity conditions that $\hat{\theta}_0$ is a moment estimator of θ based on $t(Y)/n(Y)$.

Proposition 5.2 *Let the situation be as in Proposition 5.1. Furthermore, let $\mathcal{X} \subseteq \mathbb{R}^m$ be a fundamental region, cf. Definition 2.3, and let X be homogeneous as in Definition 2.1 where the function g is $\mathcal{X}$–periodic, cf. Definition 2.5. Let the density be parametrized by $\psi \in \Psi$.*

Then,

$$\mathbb{E}_{\theta,\psi}\left(\frac{t(Y)}{n(Y)}\right) = m(\theta).$$

For $t(y)/n(y) \in \mathrm{int}C$, the estimator $\hat{\theta}_0$ is the unique θ, satisfying

$$\mathbb{E}_{\theta}\left(\frac{t(Y)}{n(Y)}\right) = \frac{t(y)}{n(y)},$$

where y is the observed point pattern and $\mathbb{E}_{\theta}$ indicates mean value under (θ,ψ) for an arbitrary ψ.

Proof. We use Corollary 2.7 with $q = \tau \circ h_\theta$, and let again U denote a uniformly distributed random variable in $\mathcal{X}$. Then,

$$\begin{aligned}\mathbb{E}_{\theta,\psi}\left(\frac{t(Y)}{n(Y)}\right) &= \mathbb{E}_{\theta,\psi}\frac{1}{n(X)}\sum_{\eta\in X}\tau(h_\theta(\eta)) = \mathbb{E}\,\tau\left(h_\theta(U)\right) = \int_{\mathcal{X}}\tau\left(h_\theta(u)\right)\frac{du}{\lambda_m(\mathcal{X})}\\ &= \int_{\mathcal{X}}\tau(u)Jh_\theta^{-1}(u)\frac{du}{\lambda_m(\mathcal{X})} = \int_{\mathcal{X}}\tau(u)\alpha(\theta)\mathrm{e}^{\theta\cdot\tau(u)}du = m(\theta).\end{aligned}$$

We have used (8), (9) and (16). The conclusion follows from (15). □

If the variance of $t(Y)/n(Y)$ is not too large then m^{-1} will appear linear and $\hat{\theta}_0$ is approximately unbiased, since, cf. (15),

$$\mathbb{E}_{\theta,\psi}\hat{\theta}_0 = \mathbb{E}_{\theta,\psi}m^{-1}\left(\frac{t(Y)}{n(Y)}\right) \approx m^{-1}\left(\mathbb{E}_{\theta,\psi}\frac{t(Y)}{n(Y)}\right) = \theta.$$

Suppose that θ is 1–dimensional and that m is concave. Then m^{-1} is convex, and from Jensens inequality we get,

$$\mathbb{E}_{\theta,\psi}\hat{\theta}_0 = \mathbb{E}_{\theta,\psi}m^{-1}(m(\hat{\theta}_0)) \geq m^{-1}\left(\mathbb{E}_{\theta,\psi}m(\hat{\theta}_0)\right) = m^{-1}(m(\theta)) = \theta. \qquad (17)$$

Thus, $\hat{\theta}_0$ is a positively biased estimator of θ under this condition. The size of the bias depends, of course, on the variance of $t(Y)/n(Y)$.

Example 5.3 For both transformations introduced in Example 3.1, we get, using e.g. (16),

$$m(\theta) = a\frac{\mathrm{e}^{\theta a}}{\mathrm{e}^{\theta a}-1} - \frac{1}{\theta}. \qquad (18)$$

This function is concave for $\theta > 0$, see Figure 7.

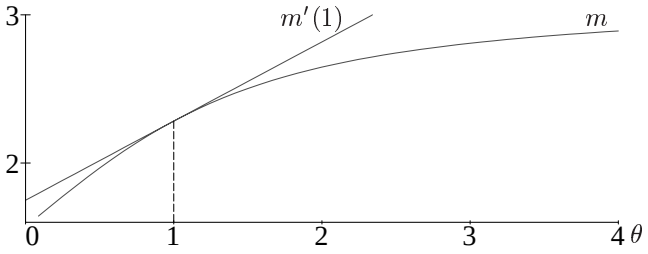


Figure 7: The function m given in (18).

5.3. Data analysis, part 3

The cell data set is just a small window of the original data set, which is a long band stretching far along the second coordinate over and under the observation window. Thus, there are no edge effects in the lower and upper part of the cell data set. In the left and right hand-sides of the window, there might be some edge-effects. However, since the interaction range is very small, we can safely ignore this.

Thereby we can assume that the conditions in Proposition 5.2 are fulfilled, and $\hat{\theta}_0$ can be regarded as a moment estimator of θ. We get,

$$\hat{\theta}_0 = (1.304, -0.272).$$

The estimates of (r, β, γ) based on $L_{\mathrm{hom}}(\beta, \gamma, r; h^{-1}_{\hat{\theta}_0}(y))$ becomes

$$(\hat{r}_0, \hat{\beta}_0, \hat{\gamma}_0) = (0.00715, 766.0, 0.08398).$$

Notice that the determination of this maximum likelihood estimate only involves techniques from the analysis of a homogeneous process.

The estimate obtained under the partial analysis based on the moment estimator is very close to that obtained using the full likelihood, cf. Section 4.2. In Figure 6 the full profile log-likelihood is plotted as a full line and the stippled line is the partial profile log-likelihood

$$\log \overline{L}_{\mathrm{hom}}(r; y) - \log \overline{L}_{\mathrm{hom}}(r_0; y) = \log \frac{\overline{L}(\hat{\theta}_0, r; y)}{\overline{L}(\hat{\theta}_0, r_0; y)},$$

where

$$\log \overline{L}_{\mathrm{hom}}(r; y) = \log L_{\mathrm{hom}}(\hat{\beta}(\hat{\theta}_0, r), \hat{\gamma}(\hat{\theta}_0, r), r; h^{-1}_{\hat{\theta}_0}(y))$$

is the homogeneous profile log-likelihood for $\theta = \hat{\theta}_0$ fixed. The partial profile log-likelihood is calculated as a part of the full profile loglikelihood, see (14). The two functions have parallel behaviour.

6. Testing hypotheses on θ

Suppose that we want to test the hypothesis $\theta \in \Theta_H \subseteq \Theta$. A simple example is a test for homogeneity in the class of exponential transformations where $\Theta_H = \{0\}$ since h_0 is the identity, see also Jensen and Nielsen (2000). Let $(\hat{\theta}_H, \hat{\psi}_H)$ and $(\hat{\theta}, \hat{\psi})$ be the maximum likelihood estimates under the hypothesis and the model, respectively. Then, the likelihood ratio test for H takes the form

$$Q = \frac{L(\hat{\theta}_H, \hat{\psi}_H; y)}{L(\hat{\theta}, \hat{\psi}; y)} = \frac{L_0(\hat{\theta}_H; y)}{L_0(\hat{\theta}; y)} \cdot \frac{L_{\hom}(\hat{\psi}_H; h_{\hat{\theta}_H}^{-1}(y))}{L_{\hom}(\hat{\psi}; h_{\hat{\theta}}^{-1}(y))} = Q_0 \cdot Q_1, \tag{19}$$

say. The ratio Q_0 is explicitly known and therefore easy to calculate, whereas Q_1 has to be calculated using MCMC.

The distribution of Q under H is usually not known. One option is to simulate the distribution of Q. But, as we have seen in Section 4, calculating just one value of the maximum likelihood estimate under the full model is rather time consuming and therefore the simulation of, say, 1000 values of Q may be an overwhelming task.

Another option is to evaluate Q in a $\chi^2(d)$ distribution where d is the difference between the dimensions of Θ and Θ_H. However, we do not have theoretical support for this procedure.

6.1. Poisson based test statistic

A simple alternative to the test statistic (19) is to use the likelihood ratio test statistic under the corresponding Poisson model,

$$Q_0^{\mathrm{p}} = \frac{L_0(\hat{\theta}_0^H; y)}{L_0(\hat{\theta}_0; y)},$$

where $\hat{\theta}_0^H$ and $\hat{\theta}_0$ are the maximum likelihood estimates based on $L_0(\theta; y)$ under the hypothesis and the model, respectively. Compare with Q_0 in (19). This suggestion is motivated by the encouraging results concerning the estimation of the inhomogeneity parameter θ without taking the interaction into account.

In particular, let $\mathcal{X} = I_1 \times I_2$ be the product set of two intervals and consider the coordinate-wise transformation $h_\theta(\eta_1, \eta_2) = (h_{\theta_1}(\eta_1), h_{\theta_2}(\eta_2))$, where both coordinate mappings are 1–dimensional exponential, cf. Example 3.1. Then a test for homogeneity of the second coordinates is a test of the hypothesis $\Theta_H = \mathbb{R} \times \{0\}$. Under the Poisson assumption, the estimate of θ_1 is the same under H as under the general model. Thereby the Poisson

based test statistic becomes

$$\begin{aligned} -2\log Q_0^{\mathrm{p}} &= -2\log\frac{L_0((\hat{\theta}_{01},0);y)}{L_0((\hat{\theta}_{01},\hat{\theta}_{02});y)} \\ &= 2n(y)\left[\log\lambda_1(I_2) - \log\int_{I_2} e^{\hat{\theta}_{02}\cdot\tau(u)}du + \hat{\theta}_{02}\cdot\frac{t_2(y)}{n(y)}\right], \quad (20) \end{aligned}$$

where $t_2(y)$ denotes the sum of the second coordinates of the data set y. Notice that the first coordinates and $\hat{\theta}_{01}$ do not enter into the test statistic.

The distribution of Q_0^{p} can easily be simulated. The simulations can be made under the homogeneous interaction model with parameter $\hat{\psi}_0$, which is the maximum likelihood estimate of ψ based on $h^{-1}_{(\hat{\theta}_{01},0)}(y)$. Notice that the value of $\hat{\theta}_{01}$ plays an indirect role.

6.2. Data analysis, part 4

In the cell data, it is of interest to test for homogeneity in the second coordinates. The Poisson based test statistic (20) is $-2\log Q_0^{\mathrm{p}} = 3.02$. The maximum likelihood estimate of the interaction parameter based on $h^{-1}_{(\hat{\theta}_{01},0)}(y)$ is $(\beta,\gamma,R) = (765.1, 0.10969, 0.0072)$ and the exact (simulated) test probability becomes 5.65%. This is based on 2000 realizations from the homogeneous Strauss process. For comparison, the test probability based on 2000 homogeneous Poisson processes is 7.4% and the test probability based on the $\chi^2(1)$ distribution is 8.22%. A similar test for homogeneity in the first coordinates gives $-2\log Q_0^{\mathrm{p}} = 83.9$ corresponding to test probability 0.

It is also possible to calculate the value of the test statistic (19) for homogeneity in the second coordinates, based on the full likelihood function. We get $-2\log Q = 7.23$. Evaluating in a $\chi^2(1)$ distribution, the test probability becomes 0.72% and is thereby somewhat smaller than the one obtained using (20). Notice that also (20) gives a small test probability although we expect homogeneity in the second coordinates. This is probably due to the large number (617) of points in the cell data.

For comparison, testing for homogeneity in the second coordinates of the pine data from Figure 2 (271 points), we get $Q_0^{\mathrm{p}} = 0.6337$ with exact (simulated Poisson) and $\chi^2(1)$ test probabilities of 42.67% and 42.60%, respectively.

7. Simulation experiments

In this section we will present some simulation experiments that illustrate the theory in the previous sections and point to some interesting results regarding the variance of $\hat{\theta}_0$.

The simulation experiments concern the TIM model from Example 3.1, illustrated in Figure 4, where the underlying homogeneous process is the circular Strauss process on $(-\pi;\pi)$. The transformation is symmetric exponential. Furthermore, we have conditioned on $n(x) = 100$ and supposed that r is known and equal to the true value. The inhomogeneity parameter θ as well as the interaction parameter ψ is 1–dimensional, $\psi = \gamma \in (0;1]$. Recall that $\gamma = 1$ is the Poisson point process, the process without interaction, and as γ decreases, the degree of inhibition increases until the hard core process is reached for $\gamma = 0$ where no points can lie closer than r apart. Notice that when $\gamma = 1$, $\hat{\theta}_0$ is the ordinary maximum likelihood estimate.

7.1. Distribution of $t(Y)/n(Y)$

First, we have examined the mean and standard deviation of $t(Y)/n(Y)$. These quantities have been approximated by the sample mean and sample standard deviation over 2000 realizations from the model.

In the upper plot of Figure 8, the sample mean is plotted for $\theta = 1$ and $\gamma = 0.01, 0.02, \ldots, 1.00$. In the lower plot, the sample standard deviation is shown. The four curves represent 4 different values of the transformation parameter $\theta = 0.5, 1, 2, 3$.

In the upper plot, we recognize the result from Proposition 5.2: the mean of $t(Y)/n(Y)$ is constant and equal to $m(\theta)$. The fluctuations come from the approximation. The lower plots indicate that the standard deviation increases with γ. Hence, the more inhibition, the smaller standard deviation. But that also corresponds with the fact, that the more inhibition, the more a small change in the transformation parameter will influence the relative number of neighbours. And therefore the estimated transformation parameter is

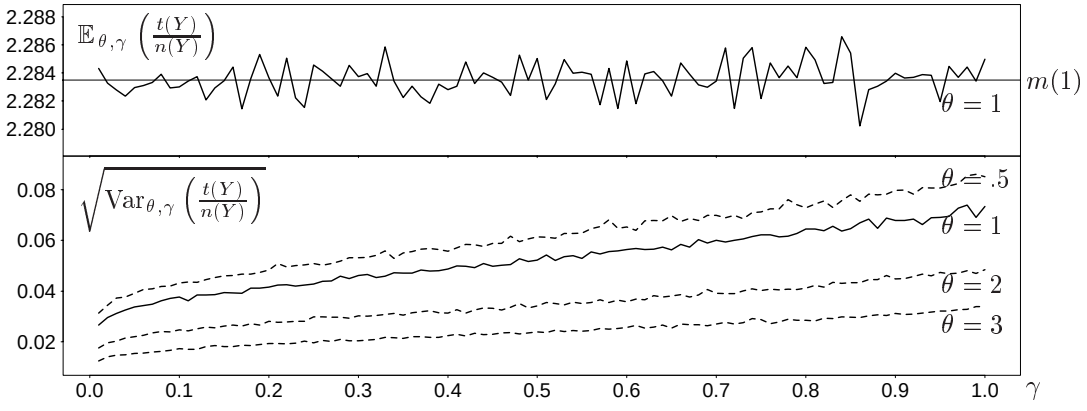


Figure 8: Simulation experiment examining the mean and standard deviation of $\frac{t(Y)}{n(Y)}$ for varying γ, and θ as indicated.

not allowed to vary very much. The same is the case when θ increases.

7.2. Distribution of $\hat{\theta}_0$

Next, we study the distribution of $\hat{\theta}_0$. As known from (17), $\hat{\theta}_0$ overestimates θ since m is concave, see Figure 7. This is also found in the upper plot in Figure 9 where the sample mean of $\hat{\theta}_0$ is plotted for γ between 0 and 1. However, considering the scale on the second axis the bias is very small. As before, the fluctuations in the mean are most likely to come from the random approximation of the mean. Still we can see an increase of the bias with increasing γ, compare with the variance of $t(Y)/n(Y)$, see Figure 8. In the middle plot of Figure 9, the mean and 95% envelopes are plotted for the 2000 samples of $\hat{\theta}_0$. The envelopes are the stippled lines. Notice that the values of $\hat{\theta}_0$ in 95% of the cases fall in the interval $[0.75; 1.25]$. From Figure 7 we see that in this interval m can be approximated by a straight line. Using (15), we then expect that the variance of $\hat{\theta}_0$ can be approximated by

$$\mathrm{Var}_{\theta,\psi}(\hat{\theta}_0) \approx \frac{1}{m'(\theta)^2}\mathrm{Var}_{\theta,\psi}\left(\frac{t(Y)}{n(Y)}\right). \tag{21}$$

In the lower plot of Figure 9 both the left-hand side and the right-hand side of (21) are plotted, and we see that the approximation is very good especially for small γ for which the envelopes are more narrow.

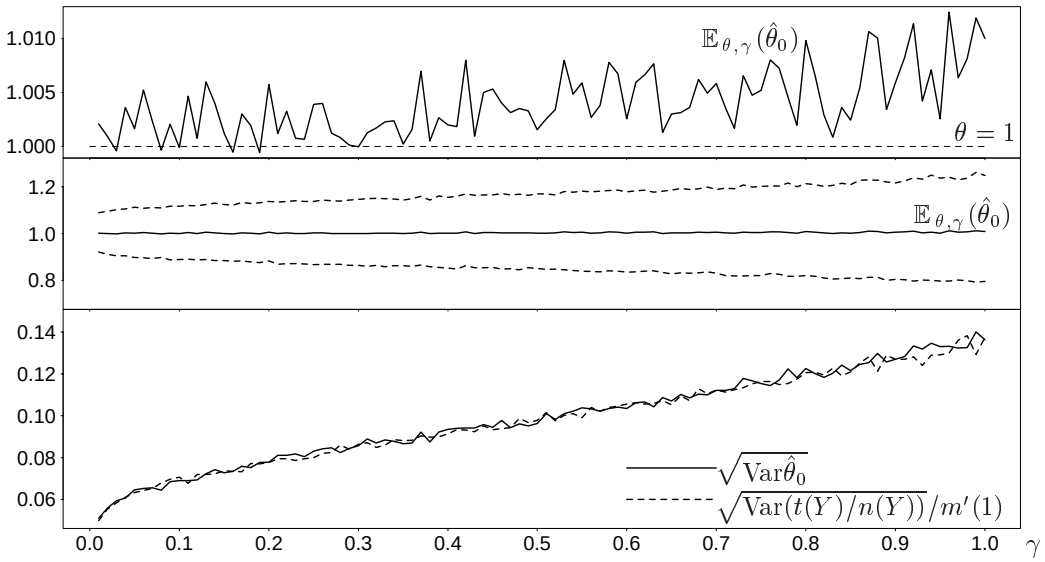


Figure 9: Simulation experiment examining the mean and standard deviation of $\hat{\theta}_0$ for varying γ, and $\theta = 1$. In the middle plot the mean is plotted together with 95% envelopes. In the lower plot the approximation (21) is examined.

Finally, note that both the bias and the variance of $\hat{\theta}_0$ decrease with γ so $\hat{\theta}_0$ has better statistical properties for small γ.

7.3. Comparison of $\hat{\theta}_0$ and $\hat{\theta}$

Finally, we will study the simultaneous distribution of $(\hat{\theta}_0, \hat{\theta})$ where $\hat{\theta}$ is the maximum likelihood estimator based on the full likelihood (10).

In the eight plots in Figure 10, the interaction parameter γ varies between 0 and 1. In each plot we have simulated 50 point patterns from the model and plotted $(\hat{\theta}_0, \hat{\theta})$. The true value of θ is 1, which is marked by the horizontal and perpendicular lines. The diagonal lines correspond to $\hat{\theta}_0 = \hat{\theta}$.

First notice that $\hat{\theta}_0$ and $\hat{\theta}$ are very similar. This is more pronounced for γ close to 1, where $\hat{\theta}_0$ is the true maximum likelihood estimate. The more interaction, the more information is lost. Secondly, the variance of $\hat{\theta}_0$ and $\hat{\theta}$ increases with γ. This is intuitively clear, since the smaller the γ, the less variation in the position of the points. Thirdly, for γ very close to 0, $(\hat{\theta}_0, \hat{\theta})$ satisfies

$$\hat{\theta}_0 \geq \hat{\theta} \geq \theta \quad \text{or} \quad \theta \leq \hat{\theta} \leq \hat{\theta}_0,$$

i.e. $\hat{\theta}$ is closer to the true value than $\hat{\theta}_0$. The effect is however not that impressive for larger values of γ. It is important to remember that the parameters are estimated under the true model with r fixed to the true value. This is an artificial restriction which never occurs with a real data set, and of course $\hat{\theta}$ gains from this.

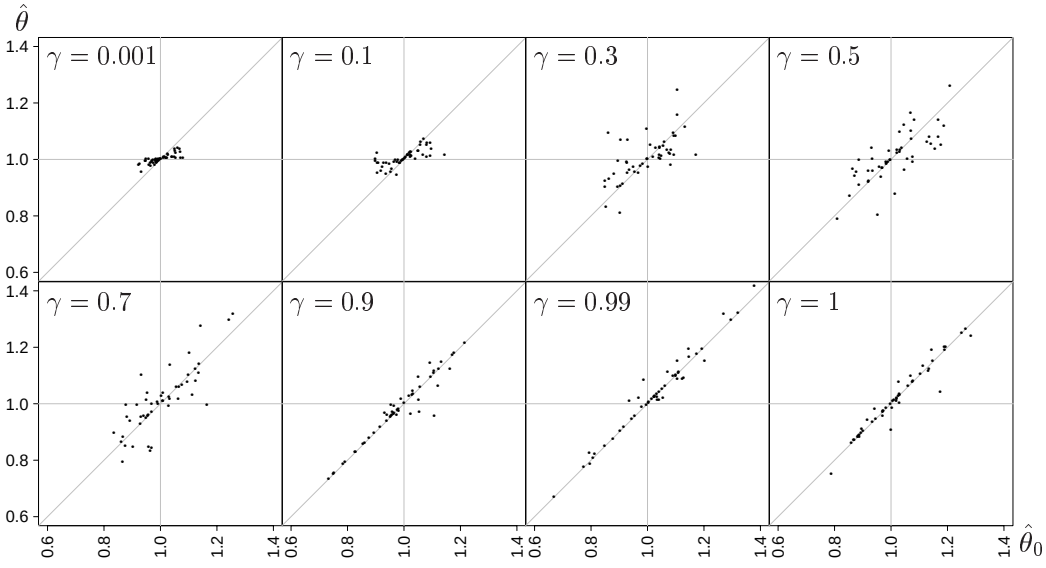


Figure 10: Comparison of $\hat{\theta}$ and $\hat{\theta}_0$. Simulation experiment with γ as indicated in each of the eight plots, and $\theta = 1$.

Thus, these simulation studies indicate that $\hat{\theta}_0$ is close to $\hat{\theta}$. Furthermore, the statistical properties of both estimators seem to be better the smaller the γ.

8. Discussion

The present paper concerns statistical inference for transformation inhomogeneous point processes. The idea of modelling inhomogeneity by transformation has also been applied in other areas of spatial statistics, cf. e.g. Perrin (1997). Transformation models for point processes yield not only inhomogeneity in the intensity of points but also in the strengths of interactions among points. For point processes defined in $\mathbb{R}^m, m \geq 2$, the neighbourhoods induced by the transformation will, except in trivial cases, be anisotropic (non-circular). This might just be what one wants if the point pattern has actually been formed by deformation. Otherwise transformation models may be used as approximations in cases where a precise model of the neighbourhoods is not important. In particular, we stress that the analysis of the cell data by a transformation model has been included in the paper for illustrative purposes. If it is important to use isotropic neighbourhoods then the inhomogeneous Markov point processes by location dependent scaling, presented in Hahn et al. (2001), may be a good alternative. For circular point processes on $\mathbb{R}$, anisotropy is not a question at all and the transformation models may generally be very useful models for describing correlated directions.

In the present paper it is argued that the inhomogeneity parameter θ can be estimated, using the partial likelihood L_0 based on an inhomogeneous Poisson point process. The statistical analysis can thereby be simplified. Most importantly, the analysis can be made into a two-step procedure where first the inhomogeneity parameter is estimated and then the back-transformed point pattern is analysed as a homogeneous point process. The likelihood analysis of a homogeneous process is already time-consuming, especially if the estimation of the interaction range r is taken into account. In our analysis this step was the real time-killer.

It is expected that Proposition 5.2 can be used to prove a consistency result for $\hat{\theta}_0$. Let us suppose that θ is a 1–dimensional parameter. Then, from (15) it follows that

$$\hat{\theta}_0 = \theta + \frac{1}{m'(\tilde{\theta})} \left(\frac{t(y)}{n(y)} - m(\theta) \right),$$

where $\mid \tilde{\theta} - \theta \mid \leq \mid \hat{\theta}_0 - \theta \mid$. Proposition 5.2 implies that $m(\theta)$ is the mean value of $t(Y)/n(Y)$. Thus, if $\mathrm{Var}_{\theta,\psi} t(Y)/n(Y)$ tends to 0 when $\mathcal{X}$ is expanding to

$\mathbb{R}^m$, then $\hat{\theta}_0$ will converge in probability to θ. The details of this reasoning have still to be fully investigated.

9. Appendix: Proof of Proposition 2.6

Let $\mathcal{X} \subseteq \mathbb{R}^m$ be a fundamental region (Definition 2.3). Let X be a homogeneous point process in $\mathcal{X}$ (Definition 2.1) with density f_X which is the restriction to $\Omega_{\mathcal{X}}$ of a translation invariant and $\mathcal{X}$-periodic (Definition 2.5) function g_X defined on $\Omega_{\mathbb{R}^m}$. Suppose also that $n(X) > 0$ almost surely. Under these conditions we will in this appendix show that a point Z chosen uniformly among the points in X is uniformly distributed in $\mathcal{X}$.

First we derive a formula for the density of Z. Let F be a measurable subset of $\mathcal{X}$, and let Π be the distribution of the unit rate Poisson point process on $\mathbb{R}^m$. Using $f_X(\emptyset) = 0$, and the well-known expansion of the distribution of the Poisson point process, see e.g. Møller (1999, Section 2), we get

$$\begin{aligned}
&\mathrm{P}(Z \in F) \\
&= \int_{\Omega_{\mathcal{X}}} P(Z \in F | X = x) f_X(x) \Pi(dx) \\
&= \int_{\Omega_{\mathcal{X}}} \frac{1}{n(x)} \sum_{\eta \in x} 1(\eta \in F) f_X(x) \Pi(dx) \\
&= \sum_{n=1}^{\infty} \mathrm{e}^{-\lambda_m(\mathcal{X})} \frac{1}{n!} \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} \frac{1}{n} \sum_{i=1}^{n} 1(x_i \in F) f_X(\{x_1, \ldots, x_n\}) dx_n \cdots dx_1 \\
&\overset{(a)}{=} \sum_{n=1}^{\infty} \mathrm{e}^{-\lambda_m(\mathcal{X})} \frac{1}{n!} \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} 1(x_1 \in F) f_X(\{x_1, \ldots, x_n\}) dx_n \cdots dx_1 \\
&= \int_F \left(\sum_{n=1}^{\infty} \mathrm{e}^{-\lambda_m(\mathcal{X})} \frac{1}{n!} \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} f_X(\{z, x_2, \ldots, x_n\}) dx_n \cdots dx_2 \right) dz
\end{aligned}$$

In (a) we have interchanged the inner sum and the integrals and used that all the terms in the new inner sum is equal to, say, the first one.

The density of Z is the expression in brackets which can be rewritten as

$$f_Z(z) = 1(z \in \mathcal{X}) \sum_{n=1}^{\infty} \mathrm{e}^{-\lambda_m(\mathcal{X})} \frac{1}{n!} \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} g_X(\{z, x_2, \ldots, x_n\}) dx_n \cdots dx_2.$$

Denote the sum $g_Z(z)$. The function g_Z is defined on $\mathbb{R}^m$.

In order to show that Z is uniformly distributed in $\mathcal{X}$, we show that g_Z is constant, i.e. that $g_Z(z + c) = g_Z(z)$ for any $c \in \mathbb{R}^m$. This is done

by showing that all terms in the sum are constant. For $n = 1$, the term is $\exp(-\lambda_m(\mathcal{X}))g_X(\{z\})$, which is constant since g_X is translation invariant. For $n \geq 2$,

$$\begin{aligned}
&\int_{\mathcal{X}} \cdots \int_{\mathcal{X}} g_X(\{z+c, x_2, \ldots, x_n\}) dx_n \cdots dx_2 \\
&\overset{(a)}{=} \int_{\mathcal{X}+c} \cdots \int_{\mathcal{X}+c} g_X(\{z+c, x_2, \ldots, x_n\}) dx_n \cdots dx_2 \\
&= \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} g_X(\{z+c, x_2+c, \ldots, x_n+c\}) dx_n \cdots dx_2 \\
&\overset{(b)}{=} \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} g_X(\{z, x_2, \ldots, x_n\}) dx_n \; cdots dx_2.
\end{aligned}$$

In (b) we have used that g_X is translation invariant. Now it only remains to show (a). Since the order of the integrals can be interchanged, similar results are to be shown for each integral. For the i'th integral, $2 \leq i \leq n$, we have,

$$\begin{aligned}
&\int_{\mathcal{X}+c} g_X(\{z+c, x_2, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n\}) dx_i \\
&\overset{(c)}{=} \sum_j \int_{(\mathcal{X}+c)\cap(\mathcal{X}+z_j)} g_X(\{z+c, x_2, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n\}) dx_i \\
&= \sum_j \int_{(\mathcal{X}+c-z_j)\cap\mathcal{X}} g_X(\{z+c, x_2, \ldots, x_{i-1}, x_i+z_j, x_{i+1}, \ldots, x_n\}) dx_i \\
&\overset{(d)}{=} \sum_j \int_{(\mathcal{X}+c-z_j)\cap\mathcal{X}} g_X(\{z+c, x_2, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n\}) dx_i \\
&\overset{(e)}{=} \int_{\mathcal{X}} g_X(\{z+c, x_2, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n\}) dx_i.
\end{aligned}$$

At (c) we have used that $\mathcal{X}$ is a fundamental region and at (d) we have used that g_X is $\mathcal{X}$-periodic. Finally, at (e) we have used that $\cup_j(\mathcal{X}+c-z_j) = \mathbb{R}^m$.

Acknowledgement

We would like to thank Ute Hahn for fruitful discussions. This work was supported in part by MaPhySto, funded by a grant from the Danish National Research Foundation.

References

Baddeley, A., Møller, J., and Waagepetersen, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Sta-*

tistica Neerlandica, 54:329–350.

Baddeley, A. and Turner, R. (2000). Practical maximum pseudolikelihood for spatial point patterns (with discussion). *Aust. N. Z. J. Statist.*, 42(3):283–322.

Barndorff-Nielsen, O. E. (1978). *Information and exponential families in statistical theory*. John Wiley & Sons Ltd., Chichester. Wiley Series in Probability and Mathematical Statistics.

Brix, A. and Møller, J. (1998). Space-time multitype log Gaussian Cox processes with a view to modelling weed data. *Research Report R–98–2012, Department of Mathematical Sciences, Aalborg University*. To appear in Scand. J. Statist., 2001.

Diggle, P. J. (1983). *Statistical analysis of spatial point patterns*. Academic Press.

Gelman, A. and Meng, X.-L. (1998). Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statist. Sci.*, 13(2):163–185.

Geyer, C. J. (1999). Likelihood inference for spatial point processes. In Barndorff-Nielsen, O. E., Kendall, W. S., and van Lieshout, M. N. M., editors, *Stochastic Geometry: Likelihood and Computation*, chapter 3, pages 79–140. Chapman and Hall/CRC, London.

Hahn, U., Jensen, E. B. V., van Lieshout, M. N. M., and Nielsen, L. S. (2001). Inhomogeneous Markov point processes by location dependent scaling. *Research Report 16, Laboratory for Computational Stochastics, University of Aarhus*. In preparation.

Jensen, E. B. V. and Nielsen, L. S. (2000). Inhomogeneous Markov point processes by transformation. *Bernoulli*, 6:761–782.

Jensen, E. B. V. and Nielsen, L. S. (2001). A review on inhomogeneous spatial point processes. In Basawa, I. V., Heyde, C. C., and Taylor, R. L., editors, *Selected Proceedings of the Symposium on Inference for Stochastic Processes*. IMS Lecture notes. To appear.

Møller, J. (1999). Markov chain Monte Carlo and spatial point processes. In Barndorff-Nielsen, O. E., Kendall, W. S., and van Lieshout, M. N. M., editors, *Stochastic Geometry: Likelihood and Computation*, chapter 4, pages 79–140. Chapman and Hall/CRC, London.

Nielsen, L. S. (2000). Modelling the position of cell profiles allowing for both inhomogeneity and interaction. *Image Analysis and Stereology*, 19(3):183–187.

Nielsen, L. S. (2001). On statistical inference in TIM models - a study of the cell data. *Technical report 1, Laboratory for Computational Stochastics, University of Aarhus.*

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

Perrin, O. (1997). *Modèle de Covariance d'un Processus Non-Stationaire par Déformation de l'Espace et Statistique.* Ph.D Thesis, Université de Paris I Panthéon-Sorbonne.

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.

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

Stoyan, D., Kendall, W. S., and Mecke, J. (1995). *Stochastic Geometry and its Statistical Applications.* Wiley, Chichester, second edition.

Stoyan, D. and Stoyan, H. (1998). Non-homogeneous Gibbs process models for forestry – a case study. *Biometrical Journal*, 40:521–531.

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

van Lieshout, M. N. M. (2000). *Markov point processes and their applications.* World Scientific.

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