Point process models allowing for interaction and inhomogeneity

Ph.D. Thesis

Linda Stougaard Nielsen
Dept. of Mathematical Sciences, University of Aarhus, 2001

Preface

The following collection of papers constitutes my Ph.D. Thesis submitted to the Faculty of Science, University of Aarhus on August 17th 2001. The work has been done under supervision of Professor Eva B. Vedel Jensen.

The first paper is an introductory review where I try to give an overview of the seven accompanying, independently written papers. The papers introduce and treat two spatial point process models allowing for inhomogeneity and interaction. In the review, these two models are informally compared with two other models.

Five of the papers have been published in, submitted to, or will be submitted to scientific journals. The papers have been written in collaboration with Eva, Dr. Ute Hahn and Dr. Marie-Colette van Lishout, who have all given their consent to the inclusion of the papers in the thesis.

During my studies, for one year beginning April 1999, I visited the University of Western Australia. I enjoyed my stay very much, and am grateful to Professor Adrian Baddeley for inviting me. I would like to express my gratitude to all at the UWA Mathematical Department, especially Adrian and the others in the dynamic and diverse group around him, for making my stay very enjoyable and scientifically fruitful.

Dr. Ute Hahn was a member of Adrian's group during my stay in Australia. Later, she joined Eva and me for nine months in Aarhus where she was the main driver of the development of the local scaling model. I really enjoyed and learned a lot from working together with Ute, who is a person sharing my special way of thinking. She has become a dear friend, and was a valuable support who always had time for me.

If not for my supervisor Eva and her constant encouragement and support, scientifically as well as personal, this thesis would not have been possible. She always had time, ideas, advice, and patience to give me in amounts reaching far beyond what is expected of a supervisor.

I am also very grateful to the other staff and guests at the Statistical Department at University of Aarhus. They all contributed to making my daily life very pleasant.

Last but not least, I would like to thank my friends and my family for being there for me, believing in me, and encouraging me. Especially Klaus who kept me company during my stay in Australia and is the main reason that I myself believe in what I am doing.

Summary

Data sets consisting of positions in the plane or in space (point patterns) can be modelled using spatial point processes. Often the points show dependence in form of inhibition or clustering (interaction), and in addition a trend in the intensity of the points (inhomogeneity) is often observed. My thesis concerns the study of point process models that can describe interaction as well as inhomogeneity.

The behaviour of a point process is usually measured relative to the homogeneous Poisson point process. Under this model the points are independently distributed and the number of points in an area is proportional to the size of the area. Inhomogeneous Poisson point processes can be used to model point patterns where there is a trend but no interaction. The important class of Markov point processes is very useful for describing interaction. However, statistical tools developed for Markov point processes models have mainly concerned homogeneous models.

The first part of the thesis is a review briefly describing four model classes capable of modelling interaction as well as inhomogeneity. The models all build on a homogeneous Markov model which is modified in order to introduce a trend. For all models, the inhomogeneous Poisson point process plays an important role. The most basic way to obtain a trend is to change the reference process from a homogeneous Poisson point process to an inhomogeneous Poisson point process. In the first type of models, the trend is obtained in this way. In the three other model classes, the trend is obtained by position dependent thinning, transformation, and local scaling, respectively. These models can also be defined with the inhomogeneous Poisson point process as reference measure. The resulting point process models are essential different. The review concludes with an informal comparison of the models where their differences and individual strengths are emphasized, and suitable fields of applications are suggested. However, when the interaction is weak or the trend is small, then the differences between the four models fade, and they might serve as alternatives of each other.

Two of the model classes considered in the review are the transformation models and the local scaling models. The main part of my thesis has concerned the development of these two model classes and a study of the statistical properties of the transformation models. The work on the transformation models has been performed together with my supervisor, professor Eva B. Vedel Jensen. The local scaling models have been developed together with her, Dr. Ute Hahn (Institute of Mathematics, Augsburg University), and Dr. Marie-Colette van Lieshout (Centre for Mathematics and Computer Science, Amsterdam).

The main contributions to the thesis are 5 papers written for journal publishing. Three of these are on the transformation model, one on the local scaling model, and one is a review paper that presents and compares the statistical properties of the transformation models and the two model classes for inhomogeneity and interaction also described in the review section of the thesis.

A transformation model is a model for interaction and inhomogeneity that differs from earlier suggested existing models, since the interaction is position dependent. In one of the papers we argue that using the transformation model, the trend in a point pattern can be estimated without taking the interaction into account. Since the trend is introduced into the model by a transformation, it is therefore possible to transform the observed inhomogeneous point pattern into a homogeneous point pattern. Then well-known analysis tools for homogeneous point patterns can be utilized to describe the interaction.

A technical report is also attached to the thesis. The computational challenges, involved in analysing a particular planar point pattern using the transformation model, is described in details.

Since the trend is introduced into the transformation model by a deformation of a homogeneous point process, the interaction is of anisotropic nature. Interaction of this type is appealing if the point pattern has been formed by physical deformation, for instance under a growth process or under stretching/compression of the material containing the points. In other situations it is of interest to describe interaction that looks similar in different areas, but at different scale. This could be plants where the environmental conditions causes some areas to be sparsely covered and other areas more densely. In dense areas the plants stand closer than in sparse areas, so the point pattern looks locally like a scaled version of a regular homogeneous point pattern. The local scaling models are able to handle interaction of such type. As for the other three mentioned models for inhomogeneity and interaction, local scaling models are also obtained by modifying a homogeneous Markov point process in order to obtain the trend. However, rather than deforming the space where the points are living, (as is done for the transformation models), local scaling models are constructed by changing the measures by which distances, areas, and volumes are are calculated. Thus, the usual volume measures are replaced by locally scaled measures, obtained by using a scaling function. We prove that in areas where the scaling function is constant, the local scaling coincides with global scaling. Statistical properties of this model still need to be investigated. However, we have reasons to believe that non-parametric kernel estimation of the intensity can be used to estimate the scaling function.

Contents

Publications:

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

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

C 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., *Selected Proceedings of the Symposium on Inference for Stochastic Processes*. IMS Lecture Notes, volume 37: 297–318.

D Nielsen L. S. and Jensen, E. B. V. (2001). Statistical inference for transformation inhomogeneous point processes. *Research Report 12, Laboratory for Computational Stochastics, University of Aarhus*. Submitted.

E 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.*

F 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.*

Manuscripts:

G Nielsen L. S. and Jensen, E. B. V. (1999). On the existence of exponential inhomogeneous Markov point processes.

Introduction

My Ph.D. studies have been devoted to the study of point process models allowing for inhomogeneity and interaction. During my Master Thesis, Eva B. Vedel Jensen and I developed the TIM models (inhomogeneous Markov point processes by transformation). The papers [A], [B], and [D], the technical report [F], and the manuscript [G] introduce and explore the TIM models. In the paper [C], the TIM models and two other model classes allowing for inhomogeneity and interaction are compared. During the last year, Ute Hahn, Eva B. Vedel Jensen, Marie-Colette van Lieshout, and I, have worked on developing yet a fourth such model class, which is introduced in [E]. Thus, we have the following four model classes that can handle inhomogeneity as well as interaction (numbered as in [C]),

I:	1st order
II:	thinning
III:	transformation/TIM
IV:	local scaling

In this review we concentrate on presenting and comparing these four model classes. The statistical properties of models of type I, II, and III, were studied and compared in [C]. Type IV models have just recently been introduced, and statistical inference need to be developed. The main focus of this review will therefore be on studying the four model classes with emphasis on their differences, similarities, individual strengths and possible fields of applications. In particular we examine the shape of the influence zones (neighbourhoods) of points, since the shape has directly influence on the nature of the interaction between points in the patterns, and therefore also on the fields of application.

In Chapter 1 some basic concepts for point processes are presented. Next, type I and II models are shortly introduced. In Chapter 2, the type III models are presented and we motivate why we found it necessary to develop a new model class. In Chapter 3, the type IV models are defined and we consider some models with similar properties that came before the local scaling models. Finally, Chapter 4 summarizes the important features of the four model classes and compare them in an informal way with a view towards data analysis.

The appendix contains a proof and two examples. Only the most persistent reader is advised to read the appendix. The subjects considered here are of a technical nature and the results are solely presented for completeness.

Chapter 1

Type I and II

Point patterns with inter-point relations and position dependent features can be modelled using point processes. Inhomogeneous Poisson point processes are capable of modelling a trend in a point pattern, but given the number of points in the process, the points are independently distributed. Markov point processes are powerful when the aim is to model interaction between points. Homogeneous Markov point processes have been thoroughly studied over the years, see e.g. Ripley and Kelly (1977), Baddeley and Møller (1989), Geyer (1999), and van Lieshout (2000). This thesis explores four model classes that are able to describe both inhomogeneity and interaction in a point pattern. The four model classes all take their point of origin in the homogeneous Markov point process models, and modify them in different ways to introduce inhomogeneity. For all the models, the inhomogeneous Poisson point process plays an important role.

In this chapter we introduce the class of Markov point process models and the concept of homogeneity. A necessary condition that ensures homogeneity of a Markov point process is also given. Furthermore, we shortly present the inhomogeneous models of type I and II.

1.1 Homogeneity and Markov point processes

Let $\mathcal{X} \subseteq \mathbb{R}^m$ be a full-dimensional bounded set and let λ_m be the Lebesgue measure in $\mathbb{R}^m$. The state space for a finite point process on $\mathcal{X}$ is $\Omega_{\mathcal{X}}$, the set of finite subsets of $\mathcal{X}$. We will assume that a point process has a density with respect to the Poisson process on $\mathcal{X}$ with intensity measure λ_m. In Møller (1999) or van Lieshout (2000) a detailed description of point processes can be found.

Notice that in [A] and [C] $\mathcal{X} \subseteq \mathbb{R}^m$ is a k–dimensional manifold equipped with the Hausdorff measure, $k \leq m$. In many situations it is possible to extend to this set–up, but for simplicity we concentrate on full-dimensional sets in this review.

Typically a point process defined on the whole $\mathbb{R}^m$ is called homogeneous or stationary if its distribution is invariant under translation, cf. Stoyan et al. (1995) and van Lieshout (2000). In [D, Definition 2.1] an equivalent definition is given for a point process X on a bounded full-dimensional set having a density f. According to this definition, X is homogeneous if f is of the form

$$f(x) = 1(x \in \Omega_{\mathcal{X}})g(x), \tag{1.1}$$

where g is translation invariant and defined for all $x \in \Omega_{\mathbb{R}^m}$.

If $\mathcal{X}$ is a k–dimensional manifold in $\mathbb{R}^m$ with $k < m$, then (1.1) need to be modified. As an example, point processes on the unit sphere are studied in [A] and [C]. Translating points on the unit sphere correspond to rotating the sphere. For more complicated manifolds, there might also be straight forward modifications.

A Markov point process on $\mathcal{X}$ has a density of the form

$$f(x) = \prod_{z \subseteq x} \varphi(z), \quad x \in \Omega_{\mathcal{X}}, \tag{1.2}$$

cf. e.g. Ripley and Kelly (1977). Here φ is a clique interaction function with respect to a reflexive and symmetric relation $\sim$ between points in $\mathcal{X}$. We call $z \in \Omega_{\mathcal{X}}$ a clique if all points in z are related by $\sim$. An interaction function has the property, that if $\varphi(z) \neq 1$ then z is a clique.

The following proposition expresses homogeneity of a Markov point process in terms of the interaction function. The proof of the proposition can be found in Section A.1 of the appendix.

Proposition 1.1 *Suppose X is a Markov point process on $\mathcal{X}$.*

Then X is homogeneous in the sense of [D, Definition 1], if and only if there exists an interaction function for X of the form

$$\varphi(x) = 1(x \in \Omega_{\mathcal{X}})\varphi_0(x),$$

where $\varphi_0(x)$ is translation invariant and defined for all $x \in \Omega_{\mathbb{R}^m}$.

The neighbourhood of a point $\eta \in \mathcal{X}$ with respect to the relation $\sim$ is defined as the set of points in $\mathcal{X}$ that are related to η,

$$N(\eta) = \{\xi \in \mathcal{X} : \xi \sim \eta\} \tag{1.3}$$

If $\sim$ is defined for all pairs of points in $\mathbb{R}^m$ and is translation invariant,

$$\eta \sim \xi \Longleftrightarrow \eta + c \sim \xi + c, \quad \forall\, c \in \mathbb{R}^m,$$

then the neighbourhood of a point is the same no matter the location of the point, $N(\eta+c) = N(\eta)+c$. This property seems to be a natural requirement for a homogeneous Markov point process, but is not necessarily fulfilled. A simple counter example is the interaction function defined by $\varphi(\{\eta\}) = a$ and $\varphi(z) = 1$ for $n(z) \geq 2$. This is a translation invariant interaction function with respect to any reflexive relation. Thus, a Markov point process with this interaction function is homogeneous according to Proposition 1.1, but the relation can be of any form and is therefore not necessarily translation invariant. However, notice that the interaction in this particular example does not play any role (the process is a homogeneous Poisson point process).

Henceforth we will only consider homogeneous Markov point processes with translation invariant neighbourhood-relation. Notice that the most commonly used relation, the distance relation,

$$\eta \sim \xi \Longleftrightarrow \|\eta - \xi\| < r,$$

$r > 0$, is a translation invariant, reflexive, and symmetric relation defined on whole $\mathbb{R}^m$.

The following simple condition on two point sets $\{\eta, \xi\} \in \Omega_{\mathbb{R}^m}$ ensures that $\sim$ is translation invariant if X is homogeneous,

$$\varphi(\{\eta, \xi\}) \neq 1 \Longleftrightarrow \eta \sim \xi.$$

Notice that the implication from left to right is always fulfilled.

In what follows, we distinguish between first order interaction terms $\varphi(\eta) = \varphi(\{\eta\})$ and the higher order terms, $\varphi(z)$ where $n(z) \geq 2$. Since the relation is reflexive, one point sets are cliques and the density contains a first order term for each point in the process. The density $f(x)$ of a homogeneous Markov point process therefore contains the term $\beta^{n(x)}$, say, where $\beta > 0$.

1.2 Inhomogeneous Poisson reference process

Typically densities of point processes are defined with respect to the unit rate Poisson point process, the Poisson process with intensity measure λ_m. The unit rate Poisson point process is thereby used as reference process and inhomogeneity and interaction of a given point process is measured relative to the unit rate Poisson. Dealing with inhomogeneous models, it might be more obvious to use an inhomogeneous Poisson point process as reference.

In order to introduce inhomogeneity into a homogeneous model X, one option is therefore to replace the unit rate Poisson reference point process with an inhomogeneous Poisson point process. Suppose that X is a homogeneous point process on $\mathcal{X}$ with density $f_X^1(x) = 1(x \in \Omega_{\mathcal{X}})g(x)$ with respect to the unit rate Poisson point process. Here, following the notation of [E], the superscript refers to the reference process. Then, for any measurable function $\lambda : \mathcal{X} \to \mathbb{R}_+$ we can define an inhomogeneous point process Y on $\mathcal{X}$ by the density

$$f_Y^\lambda(y) = 1(y \in \Omega_{\mathcal{X}})g(y), \tag{1.4}$$

which is the density of X but now with respect to the inhomogeneous Poisson point process with intensity measure μ,

$$\mu(A) = \int_A \lambda(\eta)\lambda_m(d\eta), \quad A \in \mathcal{B}_m,$$

or equivalently, the inhomogeneous Poisson point process with intensity function λ. In the above formula, $\mathcal{B}_m$ denotes the Borel sets in $\mathbb{R}^m$.

The inhomogeneous Poisson point process with intensity function λ has density

$$\mathrm{e}^{-\int_{\mathcal{X}}(\lambda(\eta)-1)d\eta} \prod_{\eta \in y} \lambda(\eta),$$

with respect to the unit rate Poisson point process. Here $d\eta$ is short for $\lambda_m(d\eta)$. The density of Y can then be formulated with respect to the unit rate Poisson point process as follows,

$$f_Y^1(y) = \mathrm{e}^{-\int_{\mathcal{X}}(\lambda(\eta)-1)d\eta} \prod_{\eta \in y} \lambda(\eta) 1(y \in \Omega_{\mathcal{X}})g(y). \tag{1.5}$$

The considerations above will turn out useful when the connection between the models for inhomogeneity are discussed. We return to this in Chapter 4 after having presented and discussed the models I through IV one by one.

1.3 Type I: First order term

Let X be a homogeneous Markov point process with interaction function φ. The process Y defined by (1.4) then has density, cf. (1.5),

$$f_Y^1(y) = \alpha \prod_{\eta \in y} (\beta\lambda(\eta)) \prod_{z \subseteq y, n(z) \geq 2} \varphi(z), \quad y \in \Omega_{\mathcal{X}}, \tag{1.6}$$

where α is the normalising constant and $\beta = \varphi(\eta)$. This is a Markov point process with position dependent first order terms $\psi(\eta) = \beta\lambda(\eta)$, and translation invariant higher order terms $\psi(z) = \varphi(z)$. Models of this type have been studied by Ogata and Tanemura (1986) and Stoyan and Stoyan (1998).

The neighbourhood relation between the points in X, and thereby also in Y, is translation invariant. The intensity of the points varies over location. Such models may be natural for inhomogeneous structures where the interpoint relation does not depend on the position. This could for example be equally sized cells with a trend in the intensity. A hard-core model with constant interaction range but a position dependent term might be a good description in this particular situation.

If the density of type I models is expressed relative to the inhomogeneous Poisson point process with intensity function λ, then it equals the density of the homogeneous template Markov model. This means, that the inhomogeneity enters through the reference measure. However, it should be noted that the intensity in a type I inhomogeneous point pattern is not represented by λ. Thus, forcing more points into a type I point pattern with inhibition between the points will result in a more homogeneous point pattern since it is 'cheaper' (meaning that the density is higher) to have extra points in sparse regions where there are more room left for points than in dense regions. Thus, with sufficiently high point intensity, a type I model will appear homogeneous. This of course depends on the strength of the inhibition.

1.4 Type II: Thinning

Type II inhomogeneity is obtained by thinning a homogeneous Markov point process with a non-constant thinning function. Points are thereby deleted with a position dependent probability. This model class was introduced and studied by Baddeley et al. (2000).

Formally, let $p : \mathcal{X} \to [0;1]$ be a thinning function and let X be a homogeneous Markov point process. Then the process defined by

$$Y = \{\eta \in X : U_\eta \leq p(\eta)\}$$

is an inhomogeneous point process of type II. Here $U_\eta \in [0,1]$, $\eta \in X$, are independent and uniformly distributed random variables, generated independently of X.

A very appealing property of this model is that the intensity of Y is of the form

$$\lambda_Y(\eta) = \lambda_X(\eta)p(\eta), \tag{1.7}$$

where the intensity λ_X of X is approximately constant. Thus, in contrast to type I models, type II models do therefore not become homogeneous when the overall point intensity increases, since the trend is determined by p only.

The property (1.7) means that non-parametric statistical methods can be used to estimate the thinning probability in the model for a given observed point pattern, see e.g. Silverman (1986). A version of the K–function for type II inhomogeneous point patterns can also be constructed. See Baddeley et al. (2000) for more details.

The thinned point process Y is not Markov, but this might not be important.

Type II models can be used for modelling of patterns where a recent and natural thinning has occurred. This could be death in patterns consisting of animals or plants. Thus, consider an old forest of trees with equal environmental conditions all over. A homogeneous hard core model could describe this. Suppose now that a chemical factory outside the forest emits poisonous gases. The concentration of gases in the air rarefies as the distance to the factory increases. Trees are therefore more likely to die because of pollution the closer they are situated to the factory.

Chapter 2

Type III

The neighbourhoods of a type I process are the same no matter the location. However, the interaction structure changes with location, since the probability of finding another point in a neighbourhood is smaller in sparse areas simply because there are fewer points. Therefore points will interact more in areas with high point intensity than in areas with low point intensity. The same phenomenon is present in type II models, since in areas where many points are deleted, the interaction will weaken more than in areas where only few points are deleted.

Thus, the interaction structure of type I and II models is location dependent. However, many point patterns with inhomogeneity and interaction are not well described by type I and II. Environmental differences such as varying access to water and nutrition might cause plants to stand closer where the conditions are good. This suggests a model where the inter-point relation and thereby the neighbourhoods of the points are position dependent. In this chapter we will consider the type III model, which is an example of a model with this feature.

2.1 Type III: Transformation (TIM)

Inhomogeneous point processes by transformation was introduced in [A]. Such a process is obtained by applying a 1–1 transformation on a homogeneous point process. Let X be a homogeneous point process and let $h : \mathcal{X} \to \mathcal{Y}$ be a bijective differential mapping with non-constant Jacobian. Here both $\mathcal{X}$ and $\mathcal{Y}$ are full-dimensional subsets of $\mathbb{R}^m$. Then $Y = h(X)$ is an inhomogeneous point process.

If X is Markov with respect to the relation $\sim$, then Y is Markov with respect to the induced relation

$$\eta \approx \xi \Longleftrightarrow h^{-1}(\eta) \sim h^{-1}(\xi). \tag{2.1}$$

Notice that this relation is position dependent if h is non-linear. Transformation inhomogeneous Markov models are also called TIM models, cf. [D], and in this review they are referred to as type III models.

The density of the type III model becomes

$$f_Y^1(y) = \mathrm{e}^{-\int_{\mathcal{Y}}\left(Jh^{-1}(\eta)-1\right)d\eta^k} \prod_{\eta \in y} Jh^{-1}(\eta) \prod_{z \subseteq y} \varphi(h^{-1}(z)), \quad y \in \Omega_{\mathcal{Y}},$$

Here, Jh^{-1} is the Jacobian of the inverse transformation h^{-1}.

Notice that the transformation model can also be defined by the corresponding homogeneous density but now with respect to the inhomogeneous Poisson point process with intensity function Jh^{-1} and with x exchanged with $h^{-1}(y)$, see (1.5).

2.2 Features of type III models

The strength of the transformation models, in our opinion, lies in the fact that the inhomogeneity is obtained by a transformation; For an inhomogeneous point pattern y in $\mathcal{Y}$, suppose that we can find a transformation $h : \mathcal{X} \to \mathcal{Y}$ such that $x = h^{-1}(y)$ is a homogeneous point pattern. Well-known non-parametric and parametric analysis tools of homogeneous processes can then be utilized to model the interaction in x.

The challenge is to find such a transformation. The Papangelou conditional intensity, cf. e.g. van Lieshout (2000), of the transformed point process Y takes the form

$$\lambda_Y(\eta; y) = Jh^{-1}(\eta)\lambda_{\text{hom}}(h^{-1}(\eta); h^{-1}(y)), \quad \eta \in \mathcal{Y}, y \in \Omega_{\mathcal{Y}},$$

where λ_{hom} is the Papangelou conditional intensity of the corresponding homogeneous model. Thus, there are reasons to believe that the intensity function of Y is approximately proportional to the inverse Jacobian,

$$\lambda_Y(\eta) \propto Jh^{-1}(\eta).$$

Non-parametric methods can be used to estimate the intensity function of Y. Then finding the transformation is reduced to solving a differential equation. In essence, the interaction is ignored and the point pattern is treated as an inhomogeneous Poisson point pattern in order to estimate the trend. Such a procedure is common practice in the analysis of point patterns, cf. e.g. Baddeley et al. (2000).

In [D] this approach has been examined in a parametric likelihood setting. It is argued that the estimation of the transformation parameter θ can be based on the inhomogeneous Poisson reference process with intensity function $\lambda_\theta = Jh_\theta^{-1}$. Formally, suppose that the density of the corresponding homogeneous model is parametrized by $\psi \in \Psi$ and the transformation is parametrized by $\theta \in \Theta$. Then the likelihood function for (θ, ψ) decomposes as follows,

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

where $L_0(\theta; y)$ is the likelihood function for the inhomogeneous Poisson reference process with intensity function Jh_θ^{-1} and $L_{\mathrm{hom}}(\psi; x)$ is the likelihood function for X when x is observed. This decomposition has important consequences for the statistical inference.

In [D] we restrict attention to the important class of exponential type III models where $\mathcal{Y} = \mathcal{X}$ and the inverse Jacobian of the transformation is of exponential form

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

Here $\tau : \mathcal{X} \to \mathbb{R}^l$ and $\theta \in \Theta \subseteq \mathbb{R}^l$. The partial likelihood L_0 takes the form

$$L_0(\theta; y) = \beta(\theta)^{n(y)} \mathrm{e}^{\theta \cdot t(y)},$$

where $t(y) = \sum_{\eta \in y} \tau(\eta)$. The estimator of θ based on L_0 is denoted $\hat{\theta}_0$. It is argued in [D] that $\hat{\theta}_0$ can be used as estimate of the inhomogeneity parameter θ. It is easy and fast to compute $\hat{\theta}_0$. Next, ψ can be estimated using the homogeneous likelihood $L_{\mathrm{hom}}(\cdot; h_{\hat{\theta}_0}^{-1}(y))$.

The type III models are somewhat restrictive since both the inhomogeneous intensity and the inhomogeneous relation are determined by the transformation. Furthermore, the transformation is usually only defined on $\mathcal{X}$. Extending the space means therefore to change the transformation. This is of no importance if the observed point pattern lives on $\mathcal{X}$ only, but if $\mathcal{X}$ is an observation window of a larger space, then the restriction may seem awkward. For simple transformations this problem has, however, a satisfactory solution.

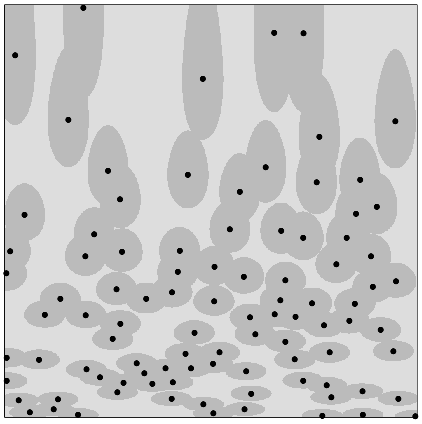
2.3 The shape of the neighbourhoods

The neighbourhoods of the induced relation are the original neighbourhoods transformed,

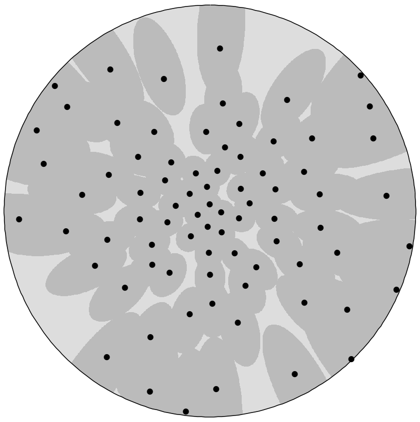
$$\begin{aligned} N(\eta;\approx) &= \{\xi : \xi \approx \eta\} = \{\xi : h^{-1}(\xi) \sim h^{-1}(\eta)\} \\ &= \{h(\xi) : \xi \sim h^{-1}(\eta)\} = h\left(\{\xi : \xi \sim h^{-1}(\eta)\}\right) = h\left(N(h^{-1}(\eta);\sim)\right). \end{aligned}$$

The induced neighbourhoods are usually anisotropic, i.e. non-spherical. As an example, in Figure 2.1 (a) a realization of a type III model with inhibition between the points and exponential increasing trend along the vertical axis has been plotted together with the induced neighbourhoods. Notice the anisotropic nature of the neighbourhoods. The neighbourhoods of the original homogeneous point pattern were all balls of the same size. The neighbourhoods are deformed in the following way. On horizontal lines, the intensity is constant. The first coordinates of the points are not transformed, and the widths of the neighbourhoods are therefore maintained and all identical. The trend is in the vertical direction and the heights of the neighbourhoods are therefore changed. In dense areas, the heights have been compressed whereas they have been elongated in sparse areas.

Neighbourhoods of this type are appealing if the point pattern has been



(a) Unit square



(b) Unit disc

Figure 2.1: Type III point patterns plotted together with their neighbourhoods. The first point pattern is taken from [C, Figure 5 (b) and (c)]. The point pattern is here plotted together with its neighbourhoods and the pattern has been rotated 90 degrees. The point pattern in (b) is from the same type III model as the point pattern in [C, Figure 1 (b)]. The parameters are $(\beta, \gamma, r, \theta) = (500, 0.01, 0.15, -3)$.

formed by deformation, for instance under a growth process or under stretching/compression of the material containing the points. (In such a setting, it is also natural that the space is bounded). An example, from Geology, is a layer of soil containing some particles. It is reasonable to assume that the degree of particle compression increases with depth, and that there is no trend parallel to the layer. Thus, the trend is perpendicular to the layer where there are natural boundaries. Since there is no trend parallel to the layer, it makes no difference if the space is bounded in this direction, and a part of the space can therefore be considered. Another example, from Biology, is surface covering layers, such as skin, membranes, and cartilage. Here, there is usually a trend perpendicular to the surface and none parallel to the surface (personal discussion with Hans Jørgen G. Gundersen, see also [B, Figure 1] where a data set of such type is shown). Thus, the space is bounded in the direction of the trend.

The neighbourhoods of the point pattern shown in Figure 2.1 (a) have shapes that are very different depending on the second coordinate. For other data sets, we may need a model where the variation in shape is more moderate. Consider the point pattern in Figure 2.1 (b). This is a realization of a type III model where the inhomogeneity is in the distance to the centre of the disc. Such models were studied in [A] and [C].

Let us compare the neighbourhoods of the two point patterns in Figure 2.1. In both examples let the height of a neighbourhood be measured in the direction of the trend and the widths be measured in the direction where the trend is constant. Thus, for the point pattern in the unit square, the heights are measured in the second coordinate direction and the widths in the first coordinate direction. For the point pattern on the unit disc, the heights are measured on the lines through the centre and the widths are measured on circles. In both point patterns the heights increases in the distance from the second axis and the centre, respectively. In the unit square, the widths are unchanged and therefore constant. On the unit disc, the widths are also increasing in the distance from the centre. Thus, the point pattern on the unit disc have neighbourhoods of more similar shapes than the point pattern in the unit square.

The reason for this difference is very simple. Transformations move the points. In the unit square, the points are moved from one horizontal line to another. On the disc, the points are moved from one circle to another. The lines in the unit square all have length 1. The circles have different radii. Thus, on the disc the points, and thereby the neighbourhoods, have to spread out when the points are moved to a circle with larger radius and contract when the new radius is smaller.

Notice that the neighbourhoods of the point pattern in the disc are still anisotropic. Furthermore, the local structure of a neighbourhood depend on global features. The behaviour described above, where the heights as well as the widths increase, depend on the transformation. If instead we would have increasing intensity in the distance from the centre, then the heights would decrease and the widths increase. This may be attractive for certain data sets, but a model where the shapes of the neighbourhoods depend on local factors and the neighbourhoods are of similar shape would be desirable in other situations. Thus, the type III model class can model position dependent interaction structure but does not cover all types of position dependent interaction structure. Other models for such interaction structures are therefore needed.

Chapter 3

Type IV

In this chapter, we discuss inhomogeneous point process models where the trend is arbitrary and the shape of the neighbourhoods depends only on local factors. The variation in shape is therefore more moderate. The main example is the class of type IV models, the local scaling models. Furthermore, we will consider some of the models that came before the type IV models.

3.1 Type IV: Local scaling

In the following we will give a short introduction to the type IV models, the local scaling models. For more details the reader is asked to consult [E], where local scaling is studied. A locally scaled point process has the property that the process locally looks like a scaled version of a homogeneous template process.

One of the concepts that plays a crucial role in the definition of these models is scale-invariance. A function $g(\cdot;\nu^*)$ defined on $\Omega_{\mathbb{R}^m}$, depending on a set of measures $\nu^* = (\nu^0, \ldots, \nu^m)$, is called scale-invariant if for all $x \in \Omega_{\mathbb{R}^m}$ and all $c > 0$,

$$g(cx;\nu_c^*) = g(x;\nu^*), \tag{3.1}$$

where $\nu_c^* = (\nu_c^0, \ldots, \nu_c^m)$ and ν_c^d is the scaled version of ν^d, $\nu_c^d(A) = \nu^d(c^{-1}A)$ for any set $A \in \mathcal{B}_d$ and $d = 0, \ldots, m$.

This concept is interesting because most homogeneous point processes have a density $f(\cdot)$ which is the restriction to $\Omega_{\mathcal{X}}$ of a scale-invariant function $g(\cdot;\nu^*)$, depending on ν^*, where ν^d is the d–dimensional Hausdorff measure in $\mathbb{R}^m$, $d = 0, \ldots, m$. Notice that then $\nu^m = \lambda_m$.

A locally scaled process is obtained by modifying a homogeneous template process X by replacing global scaling (3.1) by a scaling function $c : \mathbb{R}^m \to \mathbb{R}_+$

and globally scaled Hausdorff measures by locally scaled versions

$$\nu_c^d(A) = \int_A c(\eta)^{-d} \nu^d(du),$$

for any set $A \in \mathcal{B}_d$ and $d = 0, \ldots, m$. A locally scaled process X_c on $\mathcal{Y}$ with template process X has density of the form, cf. [E, Definition 3.3],

$$f_{X_c}^{\lambda}(y) \propto g(y; \nu_c^*), \quad y \in \Omega_{\mathcal{Y}},$$

where $\lambda(\eta) = c(\eta)^{-m}$.

One of the motivations for introducing the definition is that in case c is constant, then $f_{X_c}^{\lambda}$ is actually the density of the globally scaled process $X_c = cX$ on $\mathcal{Y} = c\mathcal{X}$. Note that global scaling can be obtained as a transformation $\eta \to c\eta$, whereas local scaling in general does not correspond to a transformation.

In [E], local scaling of two central Markov model classes, the distance-interaction processes and the shot noise processes, are discussed. It is shown that the Papangelou conditional intensity of the locally scaled process satisfies

$$\lambda_{X_c}(\eta|x) = \lambda_X \left(\frac{\eta}{c(\eta)} \middle| \frac{x}{c(\eta)} \right), \quad \eta \notin x,$$

if $c(\xi) = c(\eta)$ for all $\eta \in b(\eta, c(\eta)r)$. Here $r > 0$ is the interaction distance for X which, in both examples, is Markov with respect to the distance relation

$$\eta \sim \xi \iff \nu^1([\eta, \xi]) < r. \tag{3.2}$$

Here $[\eta, \xi]$ denotes the line segment between the two points and ν^1 is the length measure, $\nu^1([\eta, \xi]) = \|\eta - \xi\|$. The locally scaled processes thereby behave locally like a scaled version of the template process X.

Notice that type IV also covers non-Markovian point processes. However, in order to model interaction, we concentrate on the case where the template process is Markov. In [E, Appendix] it is shown, that then the locally scaled process is also Markov. In the general case, the relation appears very complicated. However, for the two mentioned model classes above, the relation becomes very nice as we will see in the following two sections where these two classes are studied.

3.1.1 Distance-interaction processes

A very useful class of homogeneous Markov point processes are the distance-interaction processes, characterized by the density

$$f_X^1(x) \propto \beta^{\nu^0(x)} \prod_{z \subseteq x, n(x) \geq 2} \varphi(D(z)), \quad x \in \Omega_{\mathcal{X}}, \tag{3.3}$$

where

$$D(z) = \{\nu^1([\eta, \xi]) : \{\eta, \xi\} \subseteq z\},$$

is the set of pairwise distances between points in z. Notice that ν^0 is the counting measure, $\nu^0(x) = n(x)$. The class of distance-interaction models cover the pairwise interaction models and the triplets process, cf. Baddeley and Turner (2000) or van Lieshout (2000), and Geyer (1999).

The right hand-side of (3.3) is of the form $g(x; \nu^*)$, which a is scale invariant function depending on the 0– and 1–dimensional Hausdorff measures. Accordingly, the locally scaled distance-interaction process has density of the form

$$f^{\lambda}_{X_c}(y) \propto g(y; \nu^*_c) = \beta^{\nu^0_c(y)} \prod_{z \subseteq y, n(y) \geq 2} \varphi(D_c(z)), \quad y \in \Omega_{\mathcal{Y}},$$

where $\nu^0_c(\cdot) = n(\cdot)$ and

$$D_c(z) = \{\nu^1_c([\eta, \xi]) : \{\eta, \xi\} \subseteq z\}$$

is the set of pairwise distances measured with the locally scaled length measure

$$\nu^1_c([\eta, \xi]) = \int_{[\eta,\xi]} c(u)^{-1} du = \|\eta - \xi\| \int_0^1 c\left(\eta + t(\xi - \eta)\right)^{-1} dt. \tag{3.4}$$

Since X is Markov with respect to the distance relation (3.2), then the locally scaled process is Markov with respect to the scaled relation

$$\eta \sim \xi \iff \nu^1_c([\eta, \xi]) < r. \tag{3.5}$$

3.1.2 Shot noise processes

A homogeneous shot noise process is Markov with respect to the relation

$$\eta \sim \xi \iff b(\eta, \tfrac{1}{2}r) \cap b(\xi, \tfrac{1}{2}r) \neq \emptyset.$$

Notice that this relation is the distance relation (3.2). The locally scaled shot noise process is Markov with respect to the relation

$$\eta \sim \xi \iff b_c(\eta, \tfrac{1}{2}r) \cap b_c(\xi, \tfrac{1}{2}r) \neq \emptyset,$$

which is of similar form as the homogeneous relation, but where the balls are replaced by scaled balls

$$b_c(\eta, d) = \{\xi \in \mathbb{R}^m : \nu^1_c([\xi, \eta]) < d\},$$

The shape of the scaled balls depend of course on the scaling function, but the balls are always star-shaped with respect to the centre-point. See [E] for more details.

3.1.3 Approximation of local scaling

Evaluating the function $g(x;\nu_c^*)$ often involves computation of complicated integrals. As an example, for the class of distance-interaction processes we need to compute the scaled lengths of line segments, that is integrals of the form (3.4). If c is of complicated form, then an approximation is useful. In [E, Section 6], approximations of the locally scaled distance-interaction processes and the shot noise processes are suggested. In [E, Figure 5] neighbourhoods corresponding to the relation (3.5) are plotted, and in [E, Figure 6] approximations of the same neighbourhoods are plotted. The approximation is in this case very close. For a given point configuration, the difference is hardly noticeable.

In the following two sections we run through other models which also have neighbourhoods of similar shape.

3.2 Type III related model

Consider the homogeneous pairwise interaction process with density

$$f_X^1(y) \propto \prod_{\eta\in y} \beta \prod_{\{\eta,\xi\}\subseteq y} \varphi(\|\eta-\xi\|).$$

The corresponding type III model has density

$$f_Y^\lambda(y) \propto \prod_{\eta\in y} \beta \prod_{\{\eta,\xi\}\subseteq y} \varphi(\|h^{-1}(\eta)-h^{-1}(\xi)\|), \tag{3.6}$$

where $\lambda = Jh^{-1}$.

In [C] the so-called type III related models were proposed, where the expression $h^{-1}(\eta)-h^{-1}(\xi)$ in (3.6) is replaced by $Jh^{-1}(\eta)^q Jh^{-1}(\xi)^q(\eta-\xi)$, $q\geq 0$. With this modification it is no longer needed to find a transformation satisfying the differential equation

$$\lambda(\eta) = Jh^{-1}(\eta)$$

where λ is known.

A type III related model is not a type III model. For $q = 1/(2m)$ it is however an approximation of a locally scaled pairwise interaction process with $c(\cdot) = [Jh^{-1}(\cdot)]^{-1/m}$. Thus the integral in (3.4) is approximated by

$$\begin{aligned}\int_0^1 c\left(\eta+t(\xi-\eta)\right)^{-1} dt &= \int_0^1 \left[Jh^{-1}(\eta+t(\xi-\eta))\right]^{1/m} dt \\ &\approx Jh^{-1}(\eta)^{1/(2m)} Jh^{-1}(\xi)^{1/(2m)}.\end{aligned}$$

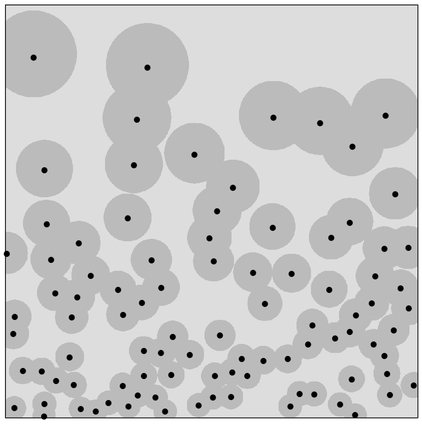


Figure 3.1: Type III related point pattern from [C, Figure 5 (b) and (c)]. The point pattern is plotted together with the neighbourhoods and have been rotated 90 degrees.

The point pattern in [C, Figure 5 (c)], shown in this review as Figure 3.1 with neighbourhoods, is a type III related Strauss model with $q = 1/(2m) = 1/4$. Notice that the neighbourhoods are of similar shape but different sizes. Compare with the corresponding type III model shown in Figure 2.1 (b).

3.3 Relation-inhomogeneity

Inspired by the way the inhomogeneity enters into a type III model, an early attempt to make a new model for inhomogeneous point processes with interaction, was based on modifying a homogeneous Markov point process with a density that only depends on the points through a translation invariant relation $\sim$. Examples are the Strauss process, see (A.4) in the appendix, and the triplets process. An inhomogeneous model is obtained by replacing $\sim$ with a position dependent relation $\overset{\star}{\sim}$ and change the reference measure to the distribution of an inhomogeneous Poisson point process with intensity function λ. We called such models *relation-inhomogeneous*. Obviously type III models are relation-inhomogeneous with $\lambda = Jh^{-1}$ and $\overset{\star}{\sim}$ as the induced relation (2.1). Thus, in order to create models with neighbourhoods that do not vary too much in shape, the challenge was to create suitable inhomogeneous relations.

In the following we consider two examples of position dependent relations that produces neighbourhoods that are roughly of similar shape. The first example was discussed in [G] and concerns the construction of an inhomogeneous point process with decreasing intensity in the distance from a planar curve $c = c(t)$, $t \in (0, 1)$. Such a trend was obtained for type III processes,

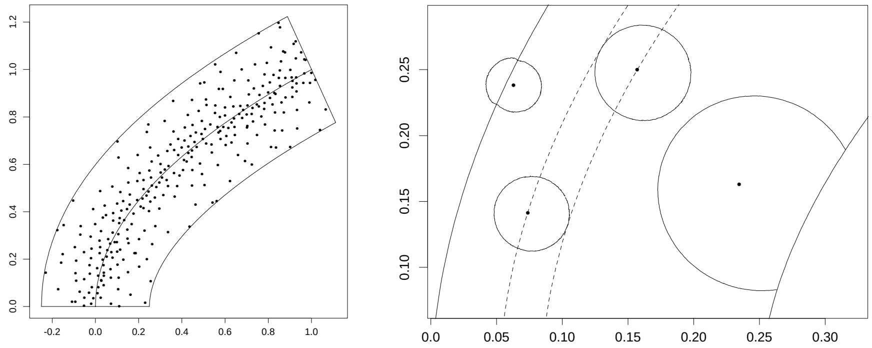


Figure 3.2: The set (3.7) with $c(t) = (t, \sqrt{t})$ and $a = 0.25$. The relation is given in Section A.2.1. Here R=0.005 and θ=10. To the left, a realization is shown where the model is a Strauss model with respect to the inhomogeneous relation. Here γ=0.05 and we have conditioned on the number of points, n=300. To the right, the indicated areas are the neighbourhoods of points in $\mathcal{X}$ for the indicated points. The dotted line illustrates the symmetry of the relation.

see [G, Figure 1]. The processes are defined on the set

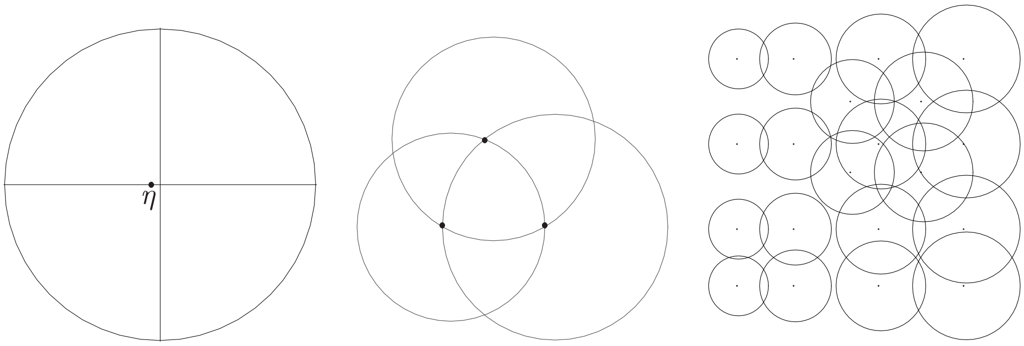
$$\mathcal{X} = \{\eta \in \mathbb{R}^2 : d(\eta, c) < a\}, \tag{3.7}$$

where $d(\eta, c)$ is the perpendicular distance from η to c, and $a > 0$, see [G]. The type III induced neighbourhoods are comparable to those in Figure 2.1 (b). Thus, when the heights of the neighbourhoods increase, then the widths are almost unchanged. As in Section 2.3, the height of a neighbourhood is measured in the direction of the trend (lines perpendicular to c), and the width is measured in the direction where there is no trend (curves in fixed perpendicular distance to c).

The point pattern shown in Figure 3.2 is a realization of a relation-inhomogeneous Strauss point process on $\mathcal{X}$. The neighbourhood relation on this set is created such that the neighbourhoods are convex sets and such that the distance from a point η to a point on the border of its neighbourhood $\xi \in \partial N(\eta)$ is increasing in $d(\xi, c)$. In this way we obtain decreasing point intensity in the perpendicular distance to the curve. Furthermore, this feature implies that the heights and the widths of the neighbourhoods increase similarly, see right hand-side of Figure 3.2. In Section A.2.1 of the appendix, the explicit form of the relation is given.

A more direct approach to obtain neighbourhoods of similar shape is to define the neighbourhoods and then define the relation via the neighbourhoods. Suppose that for each point $\eta \in \mathbb{R}^m$, a bounded set $N(\eta)$ is defined satisfying

$$\eta \in N(\eta) \quad \text{and} \quad \eta \in N(\xi) \Longleftrightarrow \xi \in N(\eta).$$



(a) $N(\eta)$ (b) Boundary points (c) Scaling

Figure 3.3: Circular neighbourhoods. In (a) a single neighbourhood is shown. In (b) and (c) neighbourhoods of different location have the same shape but different sizes. In (b) it is shown that points on the border determines other neighbourhoods successively. In (c) it is illustrated that the neighbourhoods give rise to a horizontal trend.

Then

$$\eta \overset{\star}{\sim} \xi \Longleftrightarrow \eta \in N(\xi).$$

is a reflexive and symmetric relation with neighbourhoods $N(\cdot)$.

It is easy to control the shape of the neighbourhoods when they are defined directly. One option is to let the neighbourhoods be scaled versions of each other,

$$N(\eta) = c(\eta)N(0) + \eta \text{ for all } \eta \in \mathcal{X}. \tag{3.8}$$

Notice that this implies that $c(0) = 0$. It seems natural to assume that the neighbourhoods fulfil

$$\eta \in \partial N(\xi) \Longleftrightarrow \xi \in \partial N(\eta), \tag{3.9}$$

see Figure 3.3 (b). Then the function c in (3.8) is uniquely defined. This can be seen using (3.9) and induction; For $\xi \in \partial N(0)$ we have that $0 \in \partial N(\eta)$. Knowing one point on $\partial N(\eta)$, the set $N(\eta)$ is known because its shape is known.

In Figure 3.3, the neighbourhoods are discs,

$$N(0) = b((\theta a_0, 0), a_0), \quad a_0 > 0, \quad -1 < \theta < 1.$$

Notice that the point is not the centre of the disc, but is moved a fraction θ away from the point along the first axis, see Figure 3.3 (a). The case $\theta = 0$ corresponds to the homogeneous distance relation. The neighbourhoods defined by (3.9) and (3.8) are uniquely determined by

$$N(\eta) = b(\eta + (\theta a(\eta_1), 0), a(\eta_1)). \tag{3.10}$$

where $a(0) = a_0$ and

$$a(u) = \frac{2\theta}{1-\theta^2} u + a_0, \quad u \in \mathbb{R}.$$

The neighbourhoods have increasing size in the first coordinate direction and constant size in the second order direction. Thus, this model give rise to a trend in the first coordinate direction, see Figure 3.3 (c). Notice that the larger the θ, the more inhomogeneity. The point pattern in Figure 2.1 (b) has a similar trend. The difference between these two models is quite obvious. The new model has neighbourhoods of the same shape.

It is interesting to notice that this example is a special case of the type III related model. The neighbourhood relation corresponding to (3.10) becomes

$$\eta \overset{\star}{\sim} \xi \iff d_2(\eta, \xi) \;<\; \lambda_{\theta,a_0}(\eta_1)^{1/2} \lambda_{\theta,a_0}(\xi_1)^{1/2} \tag{3.11}$$

where

$$\lambda_{\theta,a_0}(u) = a(u)\sqrt{1-\theta^2}.$$

The type III related model has relation

$$\eta \overset{\star}{\sim} \xi \iff Jh^{-1}(\eta)^q Jh^{-1}(\xi)^q \|\eta - \xi\| < r, \tag{3.12}$$

see Section 3.2. Thus, the relation (3.11) is of form (3.12) with

$$r \equiv 1, \quad q = 1/2 \quad \text{and} \quad Jh^{-1}(\eta) = \lambda_{a_0,\theta}(\eta_1)^{-1}.$$

Consequently, the circular neighbourhoods is an approximation of local scaling. A result like this is expected since the neighbourhoods actually *are* scaled version of each other. In Section A.2.2 of the appendix an extension of this example can be found.

Chapter 4

Informal comparison of type I, II, III, and IV

The order in which the four models have been numbered and presented is historical. Models of type I have been around longest. Type II and III were developed simultaneously. Type IV is brand-new. In type II models, the local intensity is proportional to the thinning probability. Type III introduced position dependent neighbourhoods which could not be modelled using neither type I nor type II. Type IV introduced neighbourhoods only depending on local factors. This feature could not be modelled using type III. It is important to realize, that this does not mean that the model classes have become more and more general and therefore include each other successively. As an example, a point pattern with identical neighbourhoods can be modelled by type I but neither by type III nor type IV, since identical neighbourhoods in these two model classes imply no trend, hence homogeneity.

When the homogeneous template model is Poisson, then the four model classes coincide. Suppose that X is a homogeneous Poisson point process with intensity β. Then type I through IV models are all inhomogeneous Poisson process models with intensity functions

$$\begin{array}{ll} \text{I:} & \beta\lambda(\eta) \\ \text{II:} & \beta\, p(\eta) \\ \text{III:} & \beta\, Jh^{-1}(\eta) \\ \text{IV:} & \beta\, c(\eta)^{-m}. \end{array}$$

Let us compare the three Markov model classes, type I, III, and IV. They can all be thought of as having first order terms β and densities with respect to the inhomogeneous Poisson point process with intensity function λ, Jh^{-1}, and c^{-m}, respectively. Restrict attention to the distance-interaction models, see Section 3.1.1 of this review. Then the densities of type I, II, and IV are all of the form

$$f_Y^{\mu}(y) \propto \beta^{n(y)} \prod_{z \subseteq y, n(z) \geq 2} \varphi(D(z)),$$

where $D(z) = \{d(\eta, \xi) : \{\eta, \xi\} \subseteq z\}$ is the set of pair-wise modified distances, and,

$$\begin{array}{lll} \text{I:} & \mu = \lambda, & d(\eta,\xi) = \|\eta - \xi\| \\ \text{III:} & \mu = Jh^{-1}, & d(\eta,\xi) = \|h^{-1}(\eta) - h^{-1}(\xi)\| \\ \text{IV:} & \mu = c^{-m}, & d(\eta,\xi) = \|\eta - \xi\| \displaystyle\int_0^1 c(\eta + t(\xi - \eta))^{-1} dt. \end{array}$$

For type III or IV to coincide with type I, we have h is the identity or $c \equiv 1$, respectively. Thus, the homogeneous model.

For type III and IV to coincide, we must have

$$\|h^{-1}(\eta) - h^{-1}(\xi)\| = \|\eta - \xi\| \int_0^1 \left(Jh^{-1}(\eta + t(\xi - \eta))\right)^{1/m} dt.$$

This is fulfilled if $m = 1$, but not in general for $m > 2$. It might be a good approximation for some choices of h. However, it emphasizes that the type III models are not a subclass of the type IV models.

4.1 Differences

To get a better intuitive understanding of the differences between the four model classes, it is useful to consider a simple example; Let the homogeneous template model be the hard-core model with respect to the distance relation. Thus, realizations of this model are patterns consisting of equally sized non-overlapping balls and the number of ball centres in an area is proportional to the size of the area.

Modifying such a process using the type I approach, the shapes and sizes of the balls are maintained but the number of balls now changes with location. Thus, realizations of the type I hard-core model are patterns with equally sized non-overlapping balls with a trend. Forcing many balls into the pattern (high intensity of the template process) the pattern will appear more homogeneous.

In type II, balls in the template process are deleted with a position dependent probability. As for type I, we get patterns with non-overlapping balls of equal sizes and a trend. Forcing more points into the pattern, has a constant effect on the intensity since the trend is solely determined by the thinning probability.

The type III models is a deformation of the matrix containing the balls. In some areas the matrix is stretched. The balls are stretched correspondingly and the intensity of balls decreases since the size of the area increases but the number of balls is maintained. In other places the matrix is compressed, and here the balls are compressed and the intensity increases. We get point patterns with deformed balls and corresponding intensity.

For the type IV model, we scale the template locally with some scaling function. In areas with constant scaling factor, we therefore again find balls, but now with scaled radius. If $c < 1$, then the radii are smaller than in the template, and the intensity will correspondingly be higher for the same reason as in type III. If $c > 1$, the radii are larger and the intensity lower. In areas with non-constant scaling factor, the balls are deformed. Thus, as for type III point patterns, type IV point patterns consist of deformed balls with intensity changing correspondingly.

The difference between the type III and the type IV deformation becomes clearer if we let $\mathcal{X}$ be the unit square. Type III deforms $\mathcal{X}$ with the restriction of keeping the image deformation of $\mathcal{X}$ in the plane. Thus, a transformation works globally since if some areas are blown up, then other areas have to shrink in order to remain in the plane.

Type IV works as a local deformation. The difference to the type III is that we do not actually transform the points, instead we locally change the measures by which the geometry is measured. This can loosely be explained as follows. Divide the unit square into non-overlapping covering areas of infinitesimal sizes. The scaling function $c(\cdot)$ is approximately constant in each of these areas. Local scaling then corresponds to global scaling of each little area with local factor $\tilde{c}_i$. Thinking of this as a physical deformation of the unit square, then the image will in general no longer be contained in the plane, since the blowing up and shrinking does not necessarily fit together as for type III. However, rather than to consider the locally scaled process as an image of a homogeneous process where each area has been locally scaled, then the measure by which the geometry is measured, is locally changed. Thus, rather than to scale each little region with local factor $\tilde{c}_i$ and use the same ruler to measure length, then we have a ruler in each of the regions, which is a scaled version of the template ruler with factor $\tilde{c}_i$. Thus, in areas with a small ruler, $\tilde{c}_1$, neighbours are obtained at much smaller physical distance (the length measured with the template ruler) than in areas where the ruler is large, $\tilde{c}_2 > \tilde{c}_1$. However, the length measured with the local ruler is identical in the two areas.

Let us shortly summarize the nature of the four models.

I: Objects of same size are placed in $\mathcal{X}$ according to λ.

II: In a point pattern with equally sized objects and constant intensity objects are removed with local probability $1 - p$.

III: Matrix containing objects of equal size and constant intensity is deformed globally.

IV: Objects of different sizes are placed according to an intensity inverse proportional to their sizes.

4.2 Choosing between the models

The four models have their individual strengths in describing the following situations:

I: The occurrence of points varies over location but the influence zone of a point does not depend on the other points.

II: A natural thinning has occurred.

III: A physical deformation has occurred.

IV: The influence zones of the points have similar shape but different sizes. The intensity of the points is inversely proportional to the size of the influence zones.

In an ideal modelling situation one should therefore consider carefully what theoretically arguments lie behind the inhomogeneity. Does the inhomogeneity affect the intensity only or also the interaction? Is the inhomogeneity caused by a natural thinning or maybe a transformation? Is it reasonable that the geometry is identical at different scale as in type IV models? We might find point patterns that are not covered by any of the four model classes but maybe by a mixture. There might be one type of inhomogeneity in the interaction and another, completely different, in the intensity.

Considering point patterns where an actual neighbourhood is available, such as point patterns consisting of cells, balls, particles, and other physical objects, it seems natural to choose a model such that the neighbourhoods of the model coincide with the observed neighbourhoods. Typically such point patterns require a hard-core model or similar, since the observed objects most often do not intersect. Consider the point pattern in [E, Figure 1] consisting of particles in a metal sinter filter. Here a type IV model seems to be most attractive, since there is a trend in intensity as well as in size of the particles, and all particles are of similar shape.

If there are no physical neighbourhoods available, explanatory variables connected with the observed point pattern might be used as a guideline. As an example, for a point pattern consisting of trees, information of sunlight, water supply, and nutrition at each position might be used in the construction of local neighbourhoods.

However, it is also important to realize that the models might serve as approximations of each other in particular cases. An example is the point pattern in [B, Figure 1] consisting of cells in a membrane. This point pattern was modelled using a type III model. The estimated model is almost hard-core. Since the sizes of the cells are small relative to the number of cells and relative to the trend, models of type I, II, or IV would probably be good models as well. In this situation, it is hard to distinguish between the four model types because the neighbourhoods are very small.

In order to describe the differences between the models, we have mostly considered the hard-core model or models with very strong interaction (see Figure 2.1, 3.1, 3.2, and [E, Figure 2 and 3]). In order to model a point pattern where no physical neighbourhoods are available, weaker interaction might be needed. The weaker the interaction is, or the smaller the interaction range is, the more difficult it becomes to distinguish between the four models. In the Poisson case, which is the case where the interaction is weakest (there is none), or equivalently, the interaction range is smallest (equal to 0), then the four models coincide.

Thus, in point patterns where the interaction is weak or the trend is not very pronounced, the four models might be alternatives of each other. Therefore the choice of model should fall on the one that serves the purpose best for the particular point pattern and problem in question. Type I is convenient because the density is of exponential family form if the trend is exponential. Type II can easily be analyzed using non-parametric statistics. Type III is convenient since an observed inhomogeneous point pattern can be transformed into a homogeneous point pattern and the tools for such patterns can therefore be utilized. Type I, II, and IV allow modelling of an arbitrary trend. Statistical inference for type IV models has not yet been studied, but we have reasons to believe that this model, like type II, can be analyzed using non-parametric statistics, see [E, Discussion].

4.3 Subjects of future interest

There are still a lot of interesting issues that need to be investigated for the four models. First of all, statistical inference for the type IV model class need to be developed. Another subject of interest is to explore how close type I and II models are in non-extreme cases. Type II is particularly good for non-parametric inference and type I is suitable for parametric inference.

Hence, if the models are close, then an analysis of a point pattern could be based on both model classes. Type II could be used to form a non-parametric estimate of the trend and to determine the structure of the interaction. Then type I could be used to estimate a parametric model.

For the type III models, we have showed that the estimate $\hat{\theta}_0$ based on the inhomogeneous Poisson reference process is a moment estimator of θ. Further investigations of the variance and consistency of the estimator needs to be performed.

Doing statistical inference for spatial point processes in practice is still far away from being routine investigations. The density of a point process is not known explicitly due to the intractable normalising constant, and therefore iterative methods such as Markov chain Monte Carlo methods are needed. Performing an actual data analysis require a lot of programming and CPU time, see e.g. [F], where the maximum likelihood estimation of the parameters in a type III model is documented in details for a particular data point pattern. A lot of work is being done in order to implement methods for statistical inference for spatial point patterns, see e.g. Geyer (1999) and Venables and Ripley (1994).

A different approach is to base the inference on already existing methods from other areas of Statistics, and thereby use existing and reliable software. This idea seems appealing from a practical point of view. Berman and Turner (1992) and Baddeley and Turner (2000) use generalized linear models to do pseudo-likelihood analysis. Survival analysis can be utilized to deal with edge-effects in order to calculate non-parametric second order statistics, cf. Baddeley (1999). Similar methods might be used for the type III and type IV model classes. In both model classes, the point intensity is inversely proportional to the sizes of the neighbourhoods. Thus, in order to model the trend, non– and semi–parametric methods based on the first as well as higher order interactions could be useful. As an example, generalized linear models with position and nearest neighbour distances as input and response might be used to estimate the Jacobian and the scaling function, respectively.

Appendix

A.1 Proof of Proposition 1.1

In this appendix it is shown that a Markov point process X in a full-dimensional bounded set $\mathcal{X} \subseteq \mathbb{R}^m$ is homogeneous in the sense of [D, Definition 1] if and only if there exists an interaction function for X of the form

$$\varphi(x) = 1(x \in \Omega_{\mathcal{X}})\varphi_0(x), \tag{A.1}$$

where $\varphi_0(x)$ is translation invariant and defined for all $x \in \Omega_{\mathbb{R}^m}$.

Suppose X has interaction function of the form (A.1). Then the density of X is

$$f(x) = \prod_{z \subseteq x} 1(z \in \Omega_{\mathcal{X}})\varphi_0(z) = 1(x \in \Omega_{\mathcal{X}}) \prod_{z \subseteq x} \varphi_0(z), \tag{A.2}$$

which is of the form (1.1) and thereby X is homogeneous as defined in [D].

On the other hand, suppose that X is homogeneous. Let g be a function that fulfils (1.1). Now define φ_0 recursively,

$$\varphi_0(x) = \begin{cases} \psi(x) & x \in \Omega_{\mathcal{X}}, \\ g(x)/\prod_{z \subset x} \varphi_0(z), & x \in \Omega_{\mathbb{R}^m} \backslash \Omega_{\mathcal{X}}. \end{cases}$$

Here, $\cdot/0 = 0$. Clearly $\varphi(x) = 1(x \in \Omega_{\mathcal{X}})\varphi_0(x)$ is an interaction function for X since $\varphi(x) = \psi(x)$ for $x \in \Omega_{\mathcal{X}}$.

Now it remains to show that φ_0 is translation invariant.

Let $x \in \Omega_{\mathcal{X}}$. Then, since $z \in \Omega_{\mathcal{X}}$ for all $z \subset x$,

$$\psi(x) = \frac{\prod_{z \subseteq x} \psi(z)}{\prod_{z \subset x} \psi(z)} = \frac{g(x)}{\prod_{z \subset x} \varphi_0(z)}.$$

Thus, φ_0 is of similar form for all $x \in \Omega_{\mathbb{R}^m}$.

The translation invariance is shown using induction in $n(x)$. For $n(x) = 0$, $x = \emptyset \in \Omega_{\mathcal{X}}$. Thus, $\varphi_0(\emptyset) = \psi(\emptyset)$, which is trivially translation invariant.

Suppose that $n(x) \geq 1$ and that $\varphi_0(z)$ is translation invariant for all $z \subset x$. Let $c \in \mathbb{R}^m$. Then

$$\varphi_0(x+c) = \frac{g(x+c)}{\prod_{z \subset (x+c)} \varphi_0(z)} = \frac{g(x)}{\prod_{z \subset x} \varphi_0(z+c)} = \frac{g(x)}{\prod_{z \subset x} \varphi_0(z)} = \varphi_0(x).$$

Notice that this also holds if the denominator is 0. Suppose that $\varphi_0(z_0) = 0$ for a set $z_0 \subset x$. Because φ_0 is defined recursively, we have that there exists a set $y_0 \subseteq z_0$ such that $g(y_0) = 0$. Thus, $g(y_0 + c) = 0$ and thereby $\varphi_0(y_0 + c) = 0$. Now, since $y_0 \subset z_0$, we get that $\varphi_0(z_0 + c) = 0$. $\square$

A.2 Two inhomogeneous relations

In the following, we consider two examples of inhomogeneous relations, and examine their neighbourhoods. Both examples were inspired by the distance relation.

A.2.1 Curve set with exponential inhomogeneity

Let $c = c(t)$, $t \in (0, 1)$ be a continuous planar curve and let $d(\eta, c)$ be the perpendicular distance from η to c, see [G]. The curve set is defined by

$$\mathcal{X} = \{\eta \in \mathbb{R}^2 : d(\eta, c) < a\}.$$

For things to behave properly, we consider only sets such that for each $\eta \in \mathcal{X}$ there exists one and only one point on c which is closest. This regularity condition restricts the choice of curve and the choice of a.

Now, define a mapping $d_\theta : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ by

- If η_1 and η_2 lie on the same side of c, then

$$d_\theta(\eta_1, \eta_2) = \mathrm{e}^{\frac{1}{2}\theta(d_2(\eta_1,\eta_2) - d(\eta_1,c) - d(\eta_2,c))} l_\theta^{-1}(d_2(\eta_1, \eta_2)), \tag{A.3}$$

  where

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

- Otherwise,

$$d_\theta(\eta_1, \eta_2) = d_\theta(\eta_1, m) + d_\theta(\eta_2, m),$$

  where m is the point on c that crosses the line between η_1 and η_2.

Then a reflexive and symmetric relation on $\mathcal{X}$ is defined by

$$\eta \overset{\star}{\sim} \xi \iff d_\theta(\eta, \xi) < r,$$

where $r > 0$. The parameter $\theta \in \mathbb{R}$ is an inhomogeneity parameter. The function l_θ is borrowed from type III models, see [D, Example 3.1]. However, the relations do *not* coincide in the case where the curve is one of the axis. The neighbourhoods defined by this relation are of similar shapes.

In Figure 3.2, $c(t) = (t, \sqrt{t})$. The point pattern is a realization of the relation-inhomogeneous Strauss process which has density

$$f^\lambda(y) \propto \beta^{n(y)} \gamma^{s^\star(y)}, \quad \text{for } y \in \Omega_\mathcal{X}, \tag{A.4}$$

where

$$s^\star(y) = \sum_{i<j} 1(\eta_i \overset{\star}{\sim} \eta_j),$$

is the number of neighbours in y with respect to the inhomogeneous relation. For the point pattern in Figure 3.2, $\lambda \equiv 1$.

The relation depends only on the distances to c and inner-point distances. The situation is therefore equivalent when the local conditions are equivalent which ensures local behaviour. It can be shown that the regularity condition ensures that the neighbourhoods with respect to $\overset{\star}{\sim}$ are bounded and that the distance from $\eta \in \mathcal{X}$ to $\xi \in \partial N(\eta)$ is increasing in $d(\eta, c)$. Together with the regularity condition, we then have that the neighbourhoods are convex.

In [G] we constructed transformations for the curve set with the property that the intensity is constant on curves in fixed perpendicular distance to the main curve c. This is only approximate the case for processes based on d_θ and it depends on the curvature of c. The shape of the neighbourhoods depend on c and therefore the intensity does also.

The above example can be generalized. Let $d^\star : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}_+$ be a mapping with $d^\star(\eta, \eta) = 0$ and $d^\star(\eta, \xi) = d^\star(\eta, \xi)$. Then

$$\eta \overset{\star}{\sim} \xi \iff d^\star(\eta, \xi) < r, \tag{A.5}$$

where $r > 0$, defines a reflexive and symmetric relation on $\mathbb{R}^m$. The usual distance relation (3.2) is of this form with $d^\star(\eta, \xi) = \nu^1([\eta, \xi])$. Based on a homogeneous distance-interaction Markov process, the type III induced relation is of above form with $d^\star(\eta, \xi) = \|h^{-1}(\eta) - h^{-1}(\xi)\|$, and the type IV distance-interaction relation (3.4) is of this form with $d^\star(\eta, \xi) = \nu_c^1([\eta, \xi])$.

The relation (A.5) does not in general produce shape-homogeneous neighbourhoods.

A.2.2 Elliptic neighbourhood

Let us consider some extensions of the circular neighbourhood example introduced in Section 3.3. Let again $N(0)$ be a disc, but now with centre $(\theta_1 a_0, \theta_2 a_0)$. Then there is inhomogeneity along both coordinate axis.

The circular neighbourhood example can also easily be extended to ellipses. One way to do this is simply to rescale one of the axis by a constant.

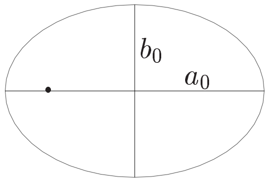


Figure A.1: Elliptic neighbourhood

In the circular case, the distance from 0 to $\xi \in \delta N(0)$ is not increasing in the first coordinate unless $\theta = 0$ which is the homogeneous case. For the example in Section A.2.1, the equivalent distance was increasing, and this might be a natural requirement. We have an option to ensure that for the elliptic neighbourhood. Suppose that

$$0 < b_0 \leq a_0 \quad \text{and} \quad 0 \leq \theta < 1,$$

where a_0 and b_0 are the dimensions of first and second axis of the ellipse, respectively, and the centre of the ellipse is $(\theta a_0, 0)$, see Figure A.1. Then the distance from 0 to $\xi \in \delta N(0)$ is increasing in ξ_1 iff

$$b_0 \geq a_0\sqrt{1-\theta}.$$

Notice that $b_0 = a_0 \iff \theta = 0$. Thus, if the neighbourhood is a disc, then only the homogeneous case fulfils the requirement.

Bibliography

Baddeley, A. (1999). Spatial sampling and censoring. In Barndorff-Nielsen, O. E., Kendall, W. S., and van Lieshout, M. N. M., editors, *Stochastic Geometry: Likelihood and Computation*, chapter 2, pages 37–78. Chapman and Hall/CRC, London.

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

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

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

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.

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 141–172. Chapman and Hall/CRC, London.

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.

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

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

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.

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

Venables, W. N. and Ripley, B. D. (1994). *Modern applied statistics with S-Plus*. Springer-Verlag, New York.

A

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

Inhomogeneous Markov point processes by transformation

Eva B. Vedel Jensen[1] and Linda Stougaard Nielsen
Laboratory for Computational Stochastics
Department of Mathematical Sciences
University of Aarhus
Ny Munkegade
DK–8000 Aarhus C
Denmark
Email: eva@imf.au.dk and lins@imf.au.dk.

Abstract

In the present paper, we construct parametrized models for point processes, allowing for both inhomogeneity and interaction. The inhomogeneity is obtained by applying parametrized transformations to homogeneous Markov point processes. An interesting model class, which can be constructed by this transformation approach, is that of exponential inhomogeneous Markov point processes. Statistical inference for such processes is discussed in some detail.

Keywords: Coarea formula; Hammersley-Clifford theorem; Hausdorff measure; Inhomogeneity; Interaction; Manifolds; Markov chain Monte Carlo; Markov point processes; Maximum likelihood estimation; Relation; Strauss process; Testing

1. Introduction

Models for spatial point processes, describing inhomogeneity as well as interaction between the points, have recently attracted considerable attention, cf. Baddeley and Turner (2000), Baddeley et al. (2000), Brix and Møller (1998) and Stoyan and Stoyan (1998). This appears to be a very natural step towards more realistic modelling, where both first and second order properties of the point pattern (like mean and variance in a univariate setting) are taken into account.

---

[1] To whom correspondence should be addressed.

At least two of these model classes can be derived from homogeneous Markov point processes. (In the present paper, we will add another of this type.) According to the Hammersley-Clifford theorem, cf. Ripley and Kelly (1977), such a process has a density with respect to a Poisson point process of the form

$$f(x) = \prod_{y \subseteq x} \varphi(y),$$

where $x = \{x_1, \ldots, x_n\}$, $x_i \in \mathcal{S}$ and $\mathcal{S}$ is a bounded Borel subset of $\mathbb{R}^m$, say. The integer n may be 0 to allow for $x = \emptyset$. The function φ is an interaction function, i.e. $\varphi(y) = 1$, unless all pairs of points in y are neighbours. The Markov point process is called homogeneous if φ is constant on all sets consisting of 1 point, cf. e.g. Ogata and Tanemura (1986) and Stoyan and Stoyan (1998).

The first type of models relates directly to the Hammersley-Clifford decomposition of the density and has been suggested among others by Ogata and Tanemura (1986). The idea is here to let the main effects in the decomposition (the interaction of 1–point sets) be non-constant. Under regularity conditions, an approximation to the likelihood function can be derived using methods from statistical physics. Stoyan and Stoyan (1998) have recently discussed this model in a forestry setting. See also Ripley (1990) and Møller et al. (1998).

Another type of models is constructed by using an independent thinning of a homogeneous Markov point process, cf. Baddeley et al. (2000). This is a well-known procedure for generating an inhomogeneous Poisson point process from a homogeneous one, cf. e.g. Stoyan et al. (1995). In Baddeley et al. (2000), semi-parametric inference of the thinned Markov point process is discussed.

In the present paper, yet another construction is suggested. The basic idea here is to introduce the inhomogeneity by applying a transformation to a homogeneous Markov point process. Inhomogeneous Poisson point processes as well as homogeneous Markov point processes can be included in such a model. The transformed point process is still a Markov point process with respect to the induced relation. This approach yields not only inhomogeneity in the intensity of the point process, but also inhomogeneity in the strengths of interactions among events. In particular, interactions are weaker among events in regions of high intensity than in regions of low intensity. This is appealing from an ecological perspective since competition among plants is weaker in regions where resources are abundant as compared to regions where resources are limited.

One of the useful properties of our model class is that the inhomogeneity and interaction can be separated. The statistical inference is based on the estimation of the transformation which can remove the inhomogeneity. After application of this transformation we are left with a homogeneous point pattern which can be analyzed by known tools, cf. e.g. Geyer (1999) and Møller (1999).

The idea of modelling inhomogeneity by transformations has been applied in other areas of spatial statistics; for instance for modelling the covariance structure of a non-stationary spatial process, cf. Sampson and Guttorp (1992) and Smith (1997). Related work can also be found in Monestiez et al. (1993), Meiring (1995), Perrin (1997) and references therein.

In Section 2, the basic concepts relating to Markov point processes are outlined. In Section 3, transformations of point processes are introduced and studied for Markov models. Parametrized transformations are considered in Section 4, resulting in models for point processes allowing for both inhomogeneity and interaction. An important particular case is that of the exponential inhomogeneous Markov point processes for which explicit expressions for the parametrized transformation can be found in the unit cube in $\mathbb{R}^m$ and on the unit sphere in $\mathbb{R}^3$. In Section 5, maximum likelihood estimation for the models described in Section 4 is discussed and the actual estimation procedure is applied to a simulated inhomogeneous point pattern on the unit sphere. In this section, tests for simple hypotheses are also derived. The final Section 6 discusses open questions and future work.

2. Markov point processes

In this section, we summarize some of the basic terminology for Markov point processes. A more detailed account of the notation and set-up can be found in Baddeley and Møller (1989) and Møller (1999).

Let $(\mathcal{S}, \mathcal{B}, \mu)$ be a measure space where $0 < \mu(\mathcal{S}) < \infty$ and $\mathcal{B}$ is separable and contains all singletons. Let Ω be the set of finite subsets of $\mathcal{S}$, equipped with the σ–field $\mathcal{F}$, as defined in Møller (1999). Then, a finite point process X is a measurable mapping defined on some probability space and taking values in $(\Omega, \mathcal{F})$. In what follows, it will be assumed that X has a density f with respect to the Poisson point process on $\mathcal{S}$ with intensity measure μ.

In order to define a Markov point process, we need a reflexive and symmetric relation $\sim$ on $\mathcal{S}$. Two points $\xi, \eta \in \mathcal{S}$ are called neighbours, if $\xi \sim \eta$. For $\eta \in \mathcal{S}$, the neighbourhood of η is the following set

$$\partial\eta = \{\xi \in \mathcal{S} : \xi \sim \eta\}.$$

A finite subset x of $\mathcal{S}$ is called a clique if all points in x are neighbours. A singleton is a clique by the requirement that $\sim$ is reflexive. By convention, the empty set is a clique. The set of cliques is denoted $\mathcal{C}$.

A finite point process X is said to be a Markov point process if, cf. Ripley and Kelly (1977),

(M1) $f(x) > 0 \Longrightarrow f(y) > 0$ for all $y \subseteq x$, $x \in \Omega$

(M2) if $f(x) > 0$, then

$$\lambda(\eta; x) = f(x \cup \{\eta\})/f(x), \quad \eta \in \mathcal{S},\ x \in \Omega,\ \eta \notin x$$

depends only on η and $\partial\eta \cap x$.

Note that $\lambda(\eta; x)$ can be regarded as the conditional 'intensity' of adding an extra point η to the point configuration x.

One way of introducing inhomogeneity into the model is to use a non-homogeneous intensity measure μ of the reference Poisson point process. This is equivalent to use a non-constant interaction function on singletons, but a homogeneous (Lebesgue, Hausdorff) intensity measure μ.

The Hammersley–Clifford theorem gives a factorization of a Markov density in terms of interactions which are only allowed between points in cliques.

Theorem 2.1 (Hammersley–Clifford) *A density f defines a Markov point process with respect to $\sim$ if and only if there exists a function $\varphi : \Omega \to [0, \infty)$, such that $\varphi(y) \neq 1$ implies that $y \in \mathcal{C}$, and such that*

$$f(x) = \prod_{y \subseteq x} \varphi(y)$$

for all $x \in \Omega$. The function φ is called the clique interaction function.

In the present paper, a Markov point process is called inhomogeneous if φ is non–constant on sets consisting of 1 point. Other definitions of inhomogeneity are of course possible, cf. Stoyan et al. (1995), but the definition given here suffices the purposes of our studies.

One of the most well-known homogeneous Markov point processes is the Strauss process, cf. Strauss (1975). If we let $n(x)$ be the number of elements in x, this process is characterized by the clique interaction function

$$\varphi(x) = \begin{cases} \alpha & \text{if } n(x) = 0 \\ \beta & \text{if } n(x) = 1 \\ \gamma & \text{if } n(x) = 2,\ x \in \mathcal{C} \\ 1 & \text{otherwise,} \end{cases}$$

such that

$$f(x) = \alpha\beta^{n(x)}\gamma^{s(x)}, \quad x \in \Omega,$$

where $s(x)$ is the number of neighbour pairs in x,

$$s(x) = \sum_{z \subseteq x} 1\left[n(z) = 2, z \in \mathcal{C}\right].$$

Note that $\alpha = \alpha(\beta, \gamma)$ is a function of $\beta, \gamma > 0$. Usually, it is also assumed that $\gamma \leq 1$.

3. Transformations of point processes

In this section, the attention is restricted to the case where $\mathcal{S}$ is a k–dimensional differentiable manifold $\mathcal{X} \subseteq \mathbb{R}^m$ and μ is the k–dimensional Hausdorff measure λ_m^k in $\mathbb{R}^m$, cf. e.g. Jensen (1998, Chapter 2) for a formal definition of Hausdorff measures. Intuitively, λ_m^k measures k–dimensional volume in $\mathbb{R}^m$. We will study smooth transformations of a point process X on $\mathcal{X}$. In Figure 1, an example of such a transformation is shown.

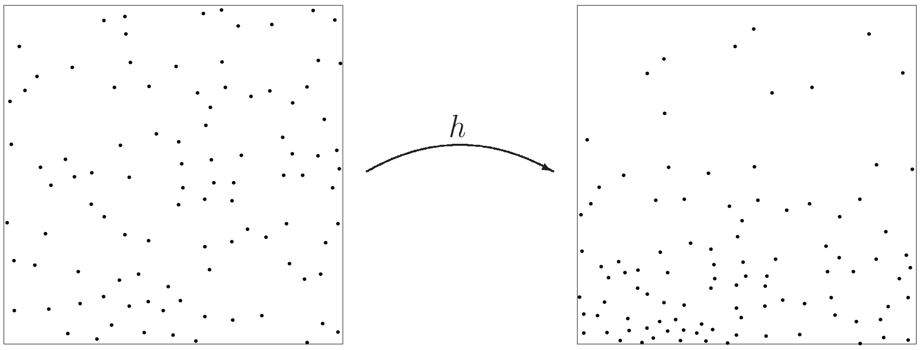


Figure 1: Adding inhomogeneity through transformation. The original point process is a conditional Strauss process on $[0, 1]^2$ with 100 points and $\gamma = 0.01$. Two points are neighbours if their mutual distance is less than $R = 0.05$.

If it is important to emphasize the containing space $\mathcal{X}$, the set of finite subsets of $\mathcal{X}$ is from now on denoted $\Omega_{\mathcal{X}}$ and the associated σ–field $\mathcal{F}_{\mathcal{X}}$. Likewise for other manifolds appearing below.

The coarea formula gives a useful transformation result for a mapping between two manifolds, cf. Jensen (1998, Theorem 2.1).

Lemma 3.1 (coarea formula) *Let $\mathcal{X} \subseteq \mathbb{R}^m$ and $\mathcal{Y} \subseteq \mathbb{R}^d$ be differentiable manifolds of dimension k. Let $h : \mathcal{X} \to \mathcal{Y}$ be a 1–1 differentiable mapping* *of $\mathcal{X}$ onto $\mathcal{Y}$. Then, there exists a function $Jh : \mathcal{X} \to [0, \infty)$, called the Jacobian, such that for any non–negative measurable function g on $\mathcal{X}$*

$$\int_{\mathcal{X}} g(x) Jh(x) dx^k = \int_{\mathcal{Y}} g(h^{-1}(y)) dy^k,$$

where dx^k and dy^k are short notations for $\lambda_m^k(dx)$ and $\lambda_d^k(dy)$, respectively.

Below, the coarea formula is used to find the density of a transformed point process $h(X) = \{h(\xi) : \xi \in X\}$.

Proposition 3.2 *Let $\mathcal{X}, \mathcal{Y}$ and h be as in Lemma 3.1. Furthermore, suppose that X is a point process on $\mathcal{X}$ with density f_X with respect to the Poisson point process on $\mathcal{X}$ with intensity measure λ_m^k. Then, $h(X)$ is the point process on $\mathcal{Y}$ with density, with respect to the Poisson point process on $\mathcal{Y}$ with intensity measure λ_d^k, of the form*

$$f_{h(X)}(y) = f_X(h^{-1}(y)) e^{\lambda_d^k(\mathcal{Y}) - \lambda_m^k(\mathcal{X})} \prod_{\eta \in y} Jh^{-1}(\eta), \quad y \in \Omega_{\mathcal{Y}}.$$

Proof. Let $F \in \mathcal{F}_{\mathcal{Y}}$. Using the well–known expansion of the distribution of the Poisson point process, cf. e.g. Møller (1999, Section 2), we get

$$\begin{aligned}
&\mathrm{P}(h(X) \in F) \\
&= \sum_{n=0}^{\infty} \mathrm{e}^{-\lambda_m^k(\mathcal{X})} \frac{1}{n!} \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} 1\left[\{h(x_1), \ldots, h(x_n)\} \in F\right] \\
&\qquad\qquad \times f_X(\{x_1, \ldots, x_n\}) dx_1^k \cdots dx_n^k \\
&\overset{(*)}{=} \sum_{n=0}^{\infty} \mathrm{e}^{-\lambda_m^k(\mathcal{X})} \frac{1}{n!} \int_{\mathcal{Y}} \cdots \int_{\mathcal{Y}} 1\left[\{y_1, \ldots, y_n\} \in F\right] \\
&\qquad\qquad \times f_X(\{h^{-1}(y_1), \ldots, h^{-1}(y_n)\}) \prod_{i=1}^{n} Jh^{-1}(y_i) dy_1^k \cdots dy_n^k.
\end{aligned}$$

At $(*)$, the coarea formula has been used on h^{-1}. For the term indexed by n, the formula has been used n times. The result now follows immediately. □

Next, the attention will be restricted to transformations of a point process X, which is Markov with respect to a relation $\sim$ on $\mathcal{X}$. In the corollary below it is shown that the transformed process is again Markov with respect to the induced relation. As will be apparent later in this paper, it is very important from a technical point of view to use the induced relation. It is also in many cases very natural, for instance in ecological applications, because the criterion for being neighbours in the transformed point pattern is more strict in regions where the transformation has attracted the points.

A

Corollary 3.3 *Let $\mathcal{X}, \mathcal{Y}$ and h be as in Lemma 3.1. Furthermore, suppose that X is a Markov point process with respect to $\sim$ such that*

$$f_X(x) = \prod_{y \subseteq x} \varphi(y), \quad x \in \Omega_{\mathcal{X}},$$

where φ is a clique interaction function. Then, $Y = h(X)$ is a Markov point process on $\mathcal{Y}$ with respect to the induced relation $\approx$, defined for $\eta_1, \eta_2 \in \mathcal{Y}$ by

$$\eta_1 \approx \eta_2 \Longleftrightarrow h^{-1}(\eta_1) \sim h^{-1}(\eta_2).$$

Furthermore, the density of Y is of the form

$$f_Y(y) = \prod_{z \subseteq y} \psi(z), \quad y \in \Omega_{\mathcal{Y}},$$

where ψ is the following clique interaction function

$$\psi(z) = \begin{cases} \varphi(\emptyset) e^{\lambda_d^k(\mathcal{Y}) - \lambda_m^k(\mathcal{X})} & \text{if } n(z) = 0 \\ \varphi(h^{-1}(\eta)) J h^{-1}(\eta) & \text{if } n(z) = 1, \ z = \{\eta\} \\ \varphi(h^{-1}(z)) & \text{otherwise.} \end{cases}$$

For the transformed point pattern it is worthwhile to notice that, except for the fact that all interactions are evaluated on the inversely transformed point pattern, only the main effects of the interaction function change. Furthermore, these values change with the same factor, as if a single point is transformed from one manifold to another.

Proof of Corollary 3.3. Most of the results of the corollary follows from Proposition 3.2. It only remains to verify that ψ is a clique interaction function with respect to $\approx$. Thus, let us suppose that $\psi(z) \neq 1$. We want to show that z is a clique with respect to $\approx$. Since every set with 0 and 1 points is a clique by convention, it suffices to consider the case where $n(z) \geq 2$. Then, $\psi(z) = \varphi(h^{-1}(z)$ and $h^{-1}(z)$ is a clique with respect to $\sim$. Accordingly, z is a clique with respect to $\approx$. $\square$

4. Exponential inhomogeneous Markov point processes

The transformation result from the previous section can be used to develop a new approach to inhomogeneity. Transforming homogeneous Markov point processes by a suitable bijective mapping, Markov point processes allowing for both interaction and inhomogeneity can be constructed. In what follows, we will restrict attention to the case where the transformation h maps $\mathcal{X}$ onto itself, that is $\mathcal{Y} = \mathcal{X}$.

Let X be a Markov point process on $\mathcal{X}$ with respect to a relation $\sim$ and with clique interaction function φ. Furthermore, let $g_\theta : \mathcal{X} \longrightarrow [0;\infty)$ be a parametrized model of the inhomogeneity where $\theta \in \Theta \subseteq \mathbb{R}^l$. Suppose that we can find, for each $\theta \in \Theta$, a differentiable 1–1 transformation h_θ of $\mathcal{X}$ onto $\mathcal{X}$ such that

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

Corollary 3.3 now gives that $Y = h_\theta(X)$ is a Markov point process on $\mathcal{X}$ with respect to the induced relation $\approx$ and with density

$$f_Y(y;\theta) = \prod_{\eta \in y} g_\theta(\eta) \prod_{z \subseteq h_\theta^{-1}(y)} \varphi(z), \quad y \in \Omega. \tag{2}$$

Notice that the inhomogeneity has been introduced by the transformation while the interaction has been inherited from the original homogeneous Markov point process. This has important consequences for the statistical inference, as shown in the next section.

Using the coarea formula, it is easily seen that (1) implies that

$$\int_{\mathcal{X}} g_\theta(\eta) d\eta^k = \lambda_m^k(\mathcal{X}), \quad \text{for all } \theta \in \Theta. \tag{3}$$

Therefore, $\{g_\theta : \theta \in \Theta\}$ can be regarded as a parametrized class of densities on $\mathcal{X}$ with respect to the uniform distribution on $\mathcal{X}$ (density $d\eta^k/\lambda_m^k(\mathcal{X})$). A natural and useful choice is an exponential family model of the inhomogeneity

$$g_\theta(\eta) = \beta(\theta) \mathrm{e}^{\theta \cdot \tau(\eta)}, \tag{4}$$

where $\cdot$ indicates inner product in $\mathbb{R}^l$ and $\tau : \mathcal{X} \longrightarrow \mathbb{R}^l$. Note that (3) then implies that

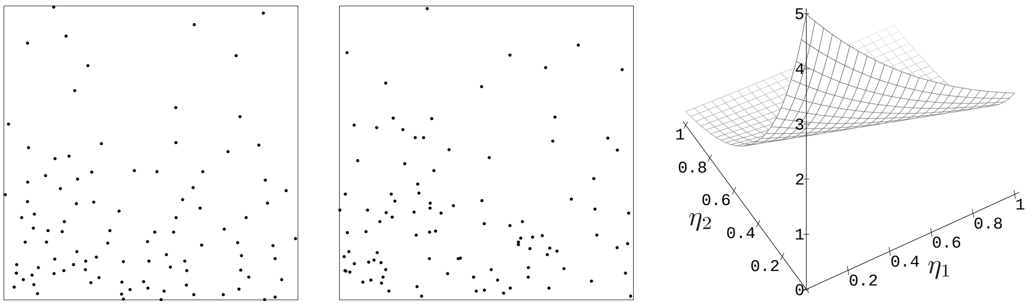
$$\beta(\theta) = \lambda_m^k(\mathcal{X}) / \int_{\mathcal{X}} \mathrm{e}^{\theta \cdot \tau(\eta)} d\eta^k.$$

A Markov point process with density (2) and g_θ given in (4) is called an *exponential inhomogeneous Markov point process.* Such a process has density

$$f_Y(y;\theta) = \beta(\theta)^{n(y)} \mathrm{e}^{\theta \cdot t(y)} \prod_{z \subseteq h_\theta^{-1}(y)} \varphi(z), \quad y \in \Omega, \tag{5}$$

where $t(y) = \sum_{\eta \in y} \tau(\eta)$.

The problem left is to find a bijective mapping h_θ that has inverse Jacobian g_θ. This problem is equivalent to that of solving a differential equation,



(a) $\gamma = 0.01$ (b) $\gamma = 1$ (c) Jacobian

Figure 2: Conditional simulations of the exponential inhomogeneous Strauss process on $[0,1]^2$ with 100 points, R=0.05 and γ as indicated. The inhomogeneity is introduced by the transformation (7) with $m = 2$ and $(\theta_1, \theta_2) = (-1, -3)$. The Jacobian of the inverse transformation is shown to the right. Another realization of the process with parameters as in (a) is shown in the right part of Figure 1.

which is not always an easy task. However, considering simple but still flexible types of inhomogeneity, it is possible to solve the equation. In the following, two such examples will be studied. Both examples are exponential inhomogeneous Markov point processes.

Example 4.1 (The unit cube in $\mathbb{R}^m$)
Let $\mathcal{X} = [0,1]^m$. Suppose that we are interested in adding independent exponential inhomogeneity on each axis. That is, for $\theta \in \mathbb{R}^m$, we consider

$$g_\theta(\eta) = \beta(\theta)\mathrm{e}^{\Sigma_{i=1}^m \theta_i \tau_i(\eta_i)}, \quad \eta \in [0,1]^m,$$

where $\tau_i : \mathbb{R} \to \mathbb{R}$ and η_i is the i'th coordinate of η. For this choice of g_θ, there is a unique solution h_θ to (1) among differentiable transformations of the form

$$h_\theta(\eta) = (h_{\theta 1}(\eta_1), \ldots, h_{\theta m}(\eta_m)), \quad \eta \in [0,1]^m,$$

where $h_{\theta i}$ is an increasing function of [0,1] onto itself. The uniqueness can be seen, using that then h_θ^{-1} is of a similar form and

$$Jh_\theta^{-1}(\eta) = \prod_{i=1}^m \frac{\partial h_{\theta i}^{-1}}{\partial \eta_i}(\eta_i), \quad \eta \in [0,1]^m.$$

For any $\theta \in \mathbb{R}^m$, the unique solution to (1) is given by

$$h_\theta^{-1}(\eta) = \left(\beta_1(\theta_1)\int_0^{\eta_1} \mathrm{e}^{\theta_1\tau_1(u)}du, \ldots, \beta_m(\theta_m)\int_0^{\eta_m} \mathrm{e}^{\theta_m\tau_m(u)}du\right), \ \eta \in [0,1]^m, \tag{6}$$

where

$$\beta_i(\theta_i) = \left(\int_0^1 \mathrm{e}^{\theta_i \tau_i(u)} du \right)^{-1}, \quad i = 1, \ldots, m,$$

and $\beta(\theta) = \prod_{i=1}^m \beta_i(\theta_i)$. Note that for $\theta = (0, \ldots, 0)$, h_θ is the identity mapping.

In particular, if $\tau_i(u) = u$ for all i, the integrals in (6) can be calculated explicitly and

$$h_\theta(\eta) = \left(\frac{1}{\theta_1} \log(1 + (\mathrm{e}^{\theta_1} - 1)\eta_1), \ldots, \frac{1}{\theta_m} \log(1 + (\mathrm{e}^{\theta_m} - 1)\eta_m) \right), \ \eta \in [0,1]^m. \tag{7}$$

In Figure 2, some simulated realizations of the exponential inhomogeneous Strauss process are shown. □

Example 4.2 (The unit sphere in $\mathbb{R}^3$)
Let $\mathcal{X} = S^2$, the unit sphere in $\mathbb{R}^3$, and let us consider exponential inhomogeneity which depends on $m \cdot \eta$ where $m \in S^2$ is fixed. So the aim is for $\theta \in \mathbb{R}$ to find a differentiable 1–1 mapping h_θ of S^2 onto S^2 with inverse Jacobian

$$J h_\theta^{-1}(\eta) = \beta(\theta) \mathrm{e}^{\theta \tau(m \cdot \eta)}, \quad \eta \in S^2, \tag{8}$$

where $\tau : \mathbb{R} \to \mathbb{R}$. Note that for $\tau(u) = u$, (8) is the density of the Fisher distribution in directional statistics whereas for $\tau(u) = u^2$, (8) is the density of the Dimroth–Watson distribution, cf. Mardia (1972).

Let us choose a coordinate system in $\mathbb{R}^3$ such that $m = (0, 0, 1)$. In the appendix, it is shown that there is a unique solution to (8) among differentiable 1–1 mappings h_θ of S^2 onto S^2 of the form

$$h_\theta(\eta_1, \eta_2, \eta_3) = \left(\frac{\sqrt{1 - r_\theta(\eta_3)^2}}{\sqrt{1 - \eta_3^2}} \eta_1, \frac{\sqrt{1 - r_\theta(\eta_3)^2}}{\sqrt{1 - \eta_3^2}} \eta_2, r_\theta(\eta_3) \right), \tag{9}$$

where r_θ is an increasing differentiable bijection on [-1,1]. Note that such a transformation only changes the angle between $\eta = (\eta_1, \eta_2, \eta_3)$ and $m = (0, 0, 1)$.

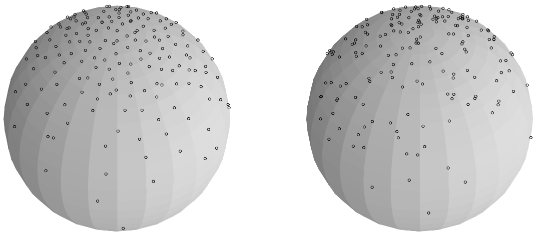
The solution is most easily expressed in terms of the inverse. We find, cf. the appendix, that the unique solution to (8) is given by

$$h_\theta^{-1}(\eta_1, \eta_2, \eta_3) = \left(\frac{\sqrt{1 - g_\theta(\eta_3)^2}}{\sqrt{1 - \eta_3^2}} \eta_1, \frac{\sqrt{1 - g_\theta(\eta_3)^2}}{\sqrt{1 - \eta_3^2}} \eta_2, g_\theta(\eta_3) \right),$$

where

$$g_\theta(u) = 1 - \int_u^1 \beta(\theta) \mathrm{e}^{\theta \tau(v)} dv, \quad -1 \le u \le 1,$$

A



(a) $\gamma = 0.02$ (b) $\gamma = 1$

Figure 3: Conditional simulations of the exponential inhomogeneous Strauss process on S^2 with 200 points, $R = 0.02$ and γ as indicated. The inhomogeneity is introduced by the transformation (9) with $\theta = 3$ and r_θ given in (10).

and

$$\beta(\theta) = \frac{2}{\int_{-1}^{1} \mathrm{e}^{\theta\tau(v)} dv}.$$

Note that for $\theta = 0$, h_θ is the identity mapping.

For τ equal to the identity mapping,

$$g_\theta(u) = 1 + \frac{\beta(\theta)}{\theta}\left(\mathrm{e}^{\theta u} - \mathrm{e}^{\theta}\right), \quad -1 \le u \le 1,$$

and

$$\beta(\theta) = \frac{2\theta}{\mathrm{e}^{\theta} - \mathrm{e}^{-\theta}}.$$

Hence,

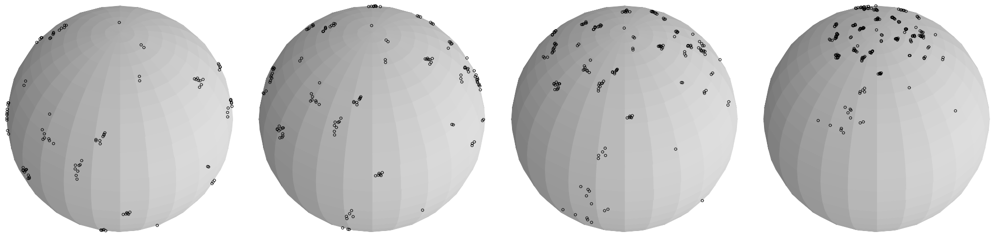
$$r_\theta(u) = \frac{1}{\theta}\log\left(\frac{u-1}{2}(\mathrm{e}^{\theta} - \mathrm{e}^{-\theta}) + \mathrm{e}^{\theta}\right), \quad -1 \le u \le 1. \tag{10}$$

In Figure 3, conditional simulations of the exponential inhomogeneous Strauss process are shown with $\tau(u) = u$ and $m = (0, 0, 1)$. The relation for the untransformed process is given by

$$\eta \sim \xi \iff d_3(\eta, \xi) < R, \tag{11}$$

where d_3 is the spatial distance. Note that the spatial distance is proportional to the geodesic distance.

In Figure 4, a realization of the continuum random cluster process with tendency of clustering, cf. Møller (1999), is transformed with three different transformation parameters. The original untransformed process is shown in Figure 4 (a). The relation is the one from (11). □



(a) $\theta = 0$ (b) $\theta = 1$ (c) $\theta = 3$ (d) $\theta = 10$

Figure 4: Conditional simulations of the exponential inhomogeneous continuum random cluster process on S^2 with 200 points, R=0.05 and $\gamma = 1000$. As in Figure 3, the inhomogeneity is introduced by the transformation (9), here with θ as indicated.

5. Statistical inference for exponential inhomogeneous Markov point processes

An exponential inhomogeneous Markov point process has a density of the form (5). Note that the Jacobian part

$$\beta(\theta)^{n(y)} \mathrm{e}^{\theta \cdot t(y)}$$

of (5) is the density of $h_\theta(X)$ when X is a homogeneous Poisson point process on $\mathcal{X}$ with intensity measure λ_m^k.

We will mainly discuss statistical inference conditional on $n(Y) = n$, the observed number of points. Using that h_θ is 1–1 so that $n(Y) = n(h_\theta(X)) = n(X)$, it is easy to see that the conditional density of Y given $n(Y) = n$ is of the form

$$f_n(y;\theta) = \beta(\theta)^n \mathrm{e}^{\theta \cdot t(y)} \prod_{z \subseteq h_\theta^{-1}(y)} \psi(z), \quad n(y) = n,$$

where

$$\psi(z) = \begin{cases} \varphi(z) & \text{if } n(z) > 0 \\ \frac{\varphi(\emptyset)}{\mathrm{P}(n(X)=n)} & \text{if } n(z) = 0. \end{cases}$$

Let us suppose that the interaction function ψ can be parametrized by some parameter $\gamma \in \Gamma \subseteq \mathbb{R}^p$. Let $L(\theta, \gamma; y)$ be the conditional likelihood function based on the inhomogeneous data y. Furthermore, let

$$L_0(\theta; y) = \beta(\theta)^n \mathrm{e}^{\theta \cdot t(y)}$$

be the conditional likelihood function of θ based on y, when disregarding the interaction, and let

$$L_1(\gamma; x) = \prod_{z \subseteq x} \psi(z; \gamma)$$

be the likelihood of γ, when observing x in the homogeneous model. Then,

$$L(\theta, \gamma; y) = L_0(\theta; y) L_1(\gamma; h_\theta^{-1}(y)).$$

This decomposition of the likelihood function has important consequences for the statistical inference. In particular, for each fixed θ, the maximum of L with respect to γ can be found, using an algorithm developed for the homogeneous case with data $h_\theta^{-1}(y)$. Further results concerning estimation can be obtained if we choose a specific model for the interaction.

Let us assume that

$$L_1(\gamma; x) = \alpha_n(\gamma)\gamma^{u(x)}, \quad n(x) = n, \ \gamma > 0, \tag{12}$$

such that the interaction model is a regular exponential family of order 1, cf. Barndorff-Nielsen (1978). There are three interesting special cases of this model:

- Strauss model (Strauss (1975))

$$u(x) = s(x) = \sum_{z \subseteq x} 1(n(z) = 2, z \in \mathcal{C})$$

- Continuum random cluster model (Møller (1999))

$$u(x) = -c(x)$$

  where $c(x)$ is the number of path-connected components in x.

- Area–interaction model (Baddeley and van Lieshout (1995))

$$u(x) = -\lambda_m(\cup_{\eta \in x} B(\eta, R))$$

  where $B(\eta, R)) \subseteq \mathbb{R}^m$ is a ball with centre η and radius R, and $\lambda_m = \lambda_m^m$ is the Lebesgue measure in $\mathbb{R}^m$.

The reason for using a minus in the two latter models is that then the interaction parameter γ has the same qualitative interpretation in all three models: $\gamma < 1$ corresponds to inhibition, $\gamma = 1$ to independence and $\gamma > 1$ to clustering.

The continuum random cluster model is not a Markov model in Ripley-Kelly sense, but a nearest–neighbour Markov model, cf. Baddeley and Møller (1989), but this distinction is not important here when discussing likelihood inference.

If a maximum likelihood estimate $(\hat{\theta}, \hat{\gamma})$ exists of (θ, γ) under the interaction model (12), then $u(h_{\hat{\theta}}^{-1}(y)) \in \operatorname{int} C$ where C is the convex support of $u(X)$, cf. Barndorff-Nielsen (1978, p. 151). It therefore suffices to restrict attention to transformations in

$$\Theta^* = \{\theta \in \Theta : u(h_\theta^{-1}(y)) \in \operatorname{int} C\}.$$

For $\theta \in \Theta^*$, there is a unique $\gamma = \gamma(\theta)$ for which $L(\theta, \cdot; y)$ attains its maximum, viz. the unique solution of

$$\mathrm{E}_{\gamma(\theta)} u(X) = u(h_\theta^{-1}(y)),$$

cf. Barndorff-Nielsen (1978, p. 152). This solution can be found using Markov chain Monte Carlo simulations, cf. e.g. Geyer (1999) and Møller (1999). An example is given later in this section.

The next step is to evaluate the partially maximized likelihood function for $\theta \in \Theta^*$,

$$\overline{L}(\theta; y) = \max_{\gamma > 0} L(\theta, \gamma; y) = L_0(\theta; y) \alpha_n(\gamma(\theta)) \gamma(\theta)^{u(h_\theta^{-1}(y))}. \tag{13}$$

This step also requires Markov chain Monte Carlo simulation since the normalizing constant $\alpha_n(\gamma(\theta))$ is not known explicitly. In order to get a stable calculation, it is very important to evaluate a ratio of likelihoods instead of a likelihood directly, cf. e.g. Geyer (1999). Typically, we want to determine, up to a constant, $\overline{L}(\theta; y)$ at a grid of θ–values. For this purpose, it suffices, since $L_0(\theta; y)$ can be calculated directly, to calculate for pairs of neighbour grid points $\theta, \tilde{\theta} \in \Theta^*$,

$$\begin{aligned} &\log \left(\frac{\overline{L}(\theta; y)/L_0(\theta; y)}{\overline{L}(\tilde{\theta}; y)/L_0(\tilde{\theta}; y)} \right) \qquad (14) \\ &= \left(u(h_\theta^{-1}(y)) - u(h_{\tilde{\theta}}^{-1}(y)) \right) \log \gamma(\tilde{\theta}) + \log \mathrm{E}_{\gamma(\theta)} \left(\frac{\gamma(\tilde{\theta})}{\gamma(\theta)} \right)^{u(X) - u(h_\theta^{-1}(y))}. \end{aligned}$$

The mean-value at the right hand side of this formula can be evaluated using Markov chain Monte Carlo simulations.

Being able to calculate the partially maximized likelihood function $\overline{L}$, a maximum of $\overline{L}$ can be searched for. Note that if $\overline{L}(\cdot; y)$ is maximal at $\hat{\theta}$, then $L(\cdot, \cdot; y)$ is maximal at $(\hat{\theta}, \gamma(\hat{\theta}))$.

A

For the Strauss model and the continuum random cluster model, $u(X)$ is discrete and the function $\theta \to \gamma(\theta)$ is actually a step function. Let the support of $u(X)$ be

$$S = \{u_-, u_- + 1, \ldots, u_+\}$$

such that $\operatorname{int} C = \operatorname{int}[u_-, u_+] = (u_-, u_+)$ and let for $i \in S$

$$\Theta_i = \{\theta \in \Theta : u(h_\theta^{-1}(y)) = i\}.$$

Then,

$$\Theta^* = \bigcup_{i=u_-+1}^{u_+-1} \Theta_i,$$

and for $\theta \in \Theta_i$, $\gamma(\theta) = \gamma_i$, say. The partially maximized likelihood function becomes

$$\overline{L}(\theta; y) = L_0(\theta; y)\alpha_n(\gamma_i)\gamma_i^i, \quad \theta \in \Theta_i,$$

$i = u_- + 1, \ldots, u_+ - 1$. Accordingly, in the subregion Θ_i, $\overline{L}(\theta; y)$ will be a rescaling of $L_0(\theta; y)$ with factor $\alpha_n(\gamma_i)\gamma_i^i$. Therefore, $\overline{L}(\cdot; y)$ is not continuous at $\theta \in \partial\Theta_i \cap \partial\Theta_{i'}$, $i \neq i'$, and traditional iterative procedures such as Newton–Raphson do not seem to be appropriate for seeking a maximum of $\overline{L}$. Instead, tabulating $\overline{L}$ in a reduced parameter set $\Theta_{\mathrm{red}} \subseteq \Theta^*$, is a better idea, when $u(X)$ is discrete. In order to be able to disregard parameter values outside Θ_{red}, this reduced set should have the property that for any $\theta \in \Theta^* \backslash \Theta_{\mathrm{red}}$ there exists $\theta' \in \Theta_{\mathrm{red}}$ such that

$$\overline{L}(\theta; y) \leq \overline{L}(\theta'; y).$$

The first procedure for reducing the parameter set is based on the proportionality of $\overline{L}$ and L_0 in the subregions Θ_i. Note that Θ_i does not need to be connected. Since $L_0(\cdot; y)$ is the likelihood function for an exponential family model, it is log–concave and thereby unimodal. Let $\hat{\theta}_0$ be the maximum likelihood estimate of θ based on $L_0(\cdot; y)$. (If L_0 is the likelihood for a regular exponential family, then $\hat{\theta}_0$ exists and is unique if and only if $t(y)$ lies in the interior of its convex support). If θ is one–dimensional, all points in Θ_i can be excluded except for points in the two sets

$$\Theta_{i-} = \{\max\{\theta \in \Theta_i : \theta \leq \hat{\theta}_0\}\}$$

and

$$\Theta_{i+} = \{\min\{\theta \in \Theta_i : \theta \geq \hat{\theta}_0\}\}.$$

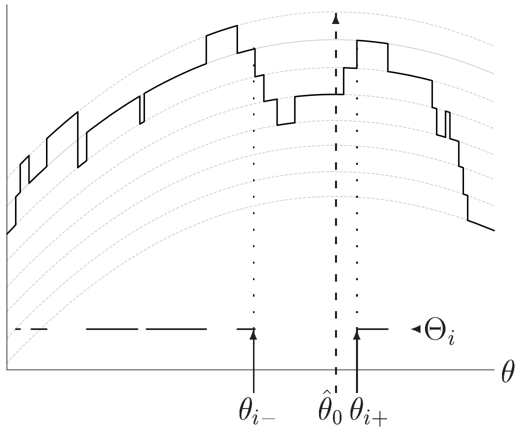


Figure 5: The partial log-likelihood function $\log \overline{L}$. The parallel curves are translations of $\log L_0(\theta; y)$. The intervals marked horizontally constitute Θ_i for the value of i corresponding to the second curve from the top.

Each of these two sets consists of at most one point. See also the illustration in Figure 5 where $\Theta_{i-} = \{\theta_{i-}\}$ and $\Theta_{i+} = \{\theta_{i+}\}$. So as Θ_{red}, any set including

$$\bigcup_{i=u_-+1}^{u_+-1} \Theta_{i-} \cup \Theta_{i+},$$

can be used. Note that Θ_{i-} and Θ_{i+} can be found by tabulating the function

$$\theta \to u(h_\theta^{-1}(y))$$

and using the determined value of $\hat{\theta}_0$.

Similar procedures must be possible if the dimension of θ is larger than 1.

The second procedure for reducing the parameter set requires that the only interaction considered is inhibition, i.e. $\gamma \leq 1$. Such a restriction may be quite natural, however, since we expect that it is going to be difficult to distinguish between inhomogeneity and clustering. For $\theta \in \Theta^*$, let $\widetilde{\gamma}(\theta)$ be the unique $\gamma \in (0, 1]$ maximizing $L(\theta, \cdot; y)$. Note that

$$\widetilde{\gamma}(\theta) = \begin{cases} \gamma(\theta) & \gamma(\theta) \leq 1 \\ 1 & \gamma(\theta) > 1. \end{cases}$$

Furthermore, let $\widetilde{L}(\theta; y) = \max_{\gamma \in (0,1]} L(\theta, \gamma; y)$. Then, we have

$\boxed{\textbf{A}}$

Proposition 5.1 *Let $\hat{\theta}_0$ be the maximum likelihood estimate of θ based on L_0. Suppose that $\hat{\theta}_0 \in \Theta^*$. Then, for any $\theta \in \Theta^*$,*

$$u(h_\theta^{-1}(y)) \geq u(h_{\hat{\theta}_0}^{-1}(y)) \Longrightarrow \widetilde{L}(\theta; y) \leq \widetilde{L}(\hat{\theta}_0; y).$$

Proof. Since $\hat{\theta}_0$ is the maximum likelihood estimate of θ based on $L_0(\cdot; y)$,

$$\beta(\theta)^n \mathrm{e}^{\theta \cdot t(y)} \leq \beta(\hat{\theta}_0)^n \mathrm{e}^{\hat{\theta}_0 \cdot t(y)}.$$

Therefore, since $\gamma^{u(h_\theta^{-1}(y))} \leq \gamma^{u(h_{\hat{\theta}_0}^{-1}(y))}$ for $\gamma \leq 1$,

$$L(\theta, \gamma; y) \leq L(\hat{\theta}_0, \gamma; y), \quad \gamma \leq 1,$$

and accordingly, the corresponding relation holds for the partially maximized likelihood function $\widetilde{L}$. □

According to Proposition 5.1, when seeking a maximum of $\widetilde{L}$, it is enough to search in

$$\{\theta \in \Theta^* : u(h_\theta^{-1}(y)) \leq u(h_{\hat{\theta}_0}^{-1}(y))\}.$$

As will be seen in the example below, this may result in a drastic reduction in the number of θ–values at which $\widetilde{L}$ has to be evaluated.

Example 5.2 (An application of the estimation procedure)
In this example we will show how the estimation procedure can be carried out, using a simulated point pattern y which is a realization of an exponential inhomogeneous Strauss process on the unit sphere, cf. the right part of Figure 6. The density is given by

$$f_n(y) = \left(\frac{2\theta}{\mathrm{e}^\theta - \mathrm{e}^{-\theta}}\right)^n \mathrm{e}^{\theta \sum_{i=1}^n y_{i3}} \alpha_n(\gamma) \gamma^{s(h_\theta^{-1}(y))}, \tag{15}$$

where $y = \{y_1, \ldots, y_n\}$ and $y_i = (y_{i1}, y_{i2}, y_{i3})$, $i = 1, \ldots, n$. The transformation h_θ is given in (9) and (10) and the relation $\sim$ is given in (11).

The aim is to estimate (θ, γ) on the basis of the inhomogeneous data set y. We will assume that $\gamma \leq 1$. As suggested earlier in this section, the estimation is based on a tabulation of the partially maximized likelihood function $\widetilde{L}(\theta; y)$ in a reduced region Θ_{red} which is known to contain $\hat{\theta}$.

In order to construct Θ_{red}, we first determine $\hat{\theta}_0$. For the example, $\hat{\theta}_0 = 5.24$ whereas the true value is $\theta = 5$. Secondly, the mapping $\theta \to s(h_\theta^{-1}(y))$ is tabulated, cf. Figure 7, and using this information we find

$$\begin{aligned}
&\{\theta : s(h_\theta^{-1}(y)) \leq s(h_{\hat{\theta}_0}^{-1}(y))\} \\
&= \{\theta : s(h_\theta^{-1}(y)) \in \{110, 111, 112, 113\}\} \\
&\subseteq [4.45, 5.50]
\end{aligned}$$

Then, according to Proposition 5.1, $\hat{\theta} \in [4.45, 5.50]$. Furthermore, we know that θ–values with the same value of $s(h_\theta^{-1}(y))$ lie in the same Θ_i–region and here we can restrict attention to θ–values closest to $\hat{\theta}_0$. In the example, this means that $\hat{\theta} \in \{5.02, 5.06, 5.08, 5.24, 5.27, 5.30\}$, cf. Figure 7. As Θ_{red}, we can take any set containing these 6 values. Since $\widetilde{L}(\theta; y)$ is evaluated by calculating ratios at pairs of close θ–values, cf. (14) with $\overline{L}$ replaced by $\widetilde{L}$ and $\gamma(\theta)$ by $\tilde{\gamma}(\theta)$, we take

$$\Theta_{\text{red}} = \{5.02, 5.03, \ldots, 5.30\}.$$

The next step is to determine $\theta \to \gamma(\theta)$. Using Markov chain Monte Carlo simulations (MCMC), the mapping

$$\gamma \longrightarrow \mathrm{E}_\gamma s(X) \tag{16}$$

is tabulated on a coarse grid of γ–values in $(0, 1]$. The mean-value in (16) is calculated in the homogeneous model. The function (16) is tabulated once more on a finer reduced grid $\Gamma_{\text{red}} \subseteq (0, 1]$ of γ–values. This reduced set is chosen such that

$$\{s(h_\theta^{-1}(y)) : \theta \in \Theta_{\text{red}}\} \subseteq \{\mathrm{E}_\gamma(s(X)) : \gamma \in \Gamma_{\text{red}}\},$$

as we are only interested in these particular values of neighbour pairs.

Regression analysis gives the relation

$$\log(\mathrm{E}_\gamma(s(X))) = \alpha + \beta \log(\gamma), \quad \gamma \in \Gamma_{\text{red}},$$

and replacing $\mathrm{E}_\gamma(s(X))$ with $s(h_\theta^{-1}(y))$ we get the following approximation

$$\tilde{\gamma}(\theta) = \exp\left(\frac{\log(s(h_\theta^{-1}(y)) - \alpha}{\beta}\right), \quad \theta \in \Theta_{\text{red}}.$$

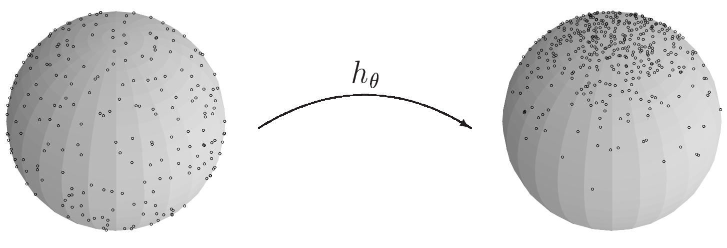


Figure 6: Transformation into the particular realization of the exponential inhomogeneous Strauss process studied in Example 5.2. Here $n = 400$, $\gamma = 0.5$, $R = 0.1$ and $\theta = 5$.

A

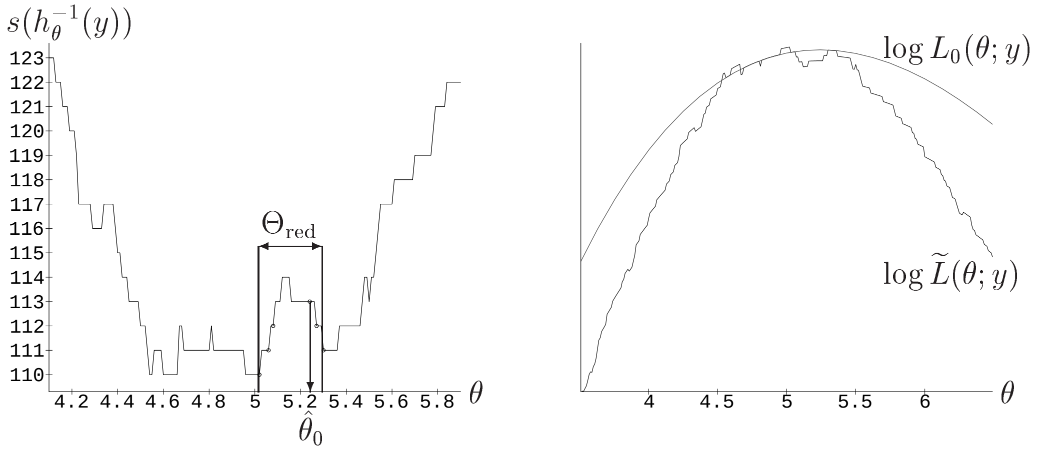


Figure 7: The number of neighbours in the inversely transformed process and the likelihood functions with and without the information about the interaction.

In the studied example, $\alpha = 5.254$ and $\beta = 0.751$. Note that $\gamma(\theta)$ only depends on $s(h_\theta^{-1}(y))$ which is discrete and constant in Θ_i–regions. Therefore, these two functions jump at the same time.

Now we are ready to find the partially maximized likelihood function. If we let

$$\Theta_{\text{red}} = \{\theta_0, \theta_1, \dots, \theta_d\},$$

$\widetilde{L}(\theta; y)$, $\theta \in \Theta_{\text{red}}$, can be determined up to a constant by using the recursive formula

$$\log\left(\frac{\widetilde{L}(\theta_l; y)/L_0(\theta_l; y)}{\widetilde{L}(\theta_0; y)/L_0(\theta_0; y)}\right) = \sum_{j=1}^{l} \log\left(\frac{\widetilde{L}(\theta_j; y)/L_0(\theta_j; y)}{\widetilde{L}(\theta_{j-1}; y)/L_0(\theta_{j-1}; y)}\right),$$

for $l = 1, \dots, d$. The terms in the sum can be found by MCMC, using (14). Denoting the above expression by Δ_l, we have that

$$\log \widetilde{L}(\theta_l; y) = \log L_0(\theta_l; y) + \Delta_l + \text{constant}.$$

Note that Δ_l is constant on Θ_i.

In Figure 7, $\log \widetilde{L}(\theta; y)$ is shown for a larger range of θ–values than Θ_{red}, in order to get a general impression of the function. Note that this function is in fact the same as the one shown in Figure 5. We find $(\hat{\theta}, \hat{\gamma}) = (\hat{\theta}, \gamma(\hat{\theta})) = (5.02, 0.48)$. $\square$

Let us finally discuss two types of tests. Let us first consider the hypothesis of no interaction $H_0 : \gamma = \gamma_0$, where $\gamma_0 = 1$. Since $L_1(\gamma_0; \cdot) \equiv 1$, we find

$$\mathrm{Q} = \frac{L(\hat{\theta}_0, \gamma_0; y)}{L(\hat{\theta}, \hat{\gamma}; y)} = \frac{L_0(\hat{\theta}_0; y)}{L_0(\hat{\theta}; y)} \frac{L_1(\gamma_0; h_{\hat{\theta}}^{-1}(y))}{L_1(\hat{\gamma}; h_{\hat{\theta}}^{-1}(y))} = \mathrm{Q}_0 \cdot \mathrm{Q}_1,$$

where Q_0 represents a comparison between $\hat{\theta}_0$ and $\hat{\theta}$, while Q_1 is a test for $\gamma = \gamma_0$ on the basis of $x = h_{\hat{\theta}}^{-1}(y)$. Note that

$$\mathrm{Q}_1 = \left[\mathrm{E}_{\hat{\gamma}} \left(\frac{\gamma_0}{\hat{\gamma}} \right)^{u(X)-u(x)} \right]^{-1}$$

and can be calculated, using Markov chain Monte Carlo simulation.

It is also of interest to test for homogeneity. Let us suppose that this corresponds to $H_0 : \theta = \theta_0$, where $\theta_0 = 0$ and h_{θ_0} is the identity. Furthermore, let $\hat{\gamma}_0$ be the estimate of γ under homogeneity. Then,

$$\begin{aligned} \mathrm{Q} &= \frac{L(\theta_0, \hat{\gamma}_0; y)}{L(\hat{\theta}, \hat{\gamma}; y)} = \frac{L_0(\theta_0; y)}{L_0(\hat{\theta}; y)} \frac{L_1(\hat{\gamma}_0; y)}{L_1(\hat{\gamma}; h_{\hat{\theta}}^{-1}(y))} \\ &= \frac{L_0(\theta_0; y)}{L_0(\hat{\theta}; y)} \frac{L_1(\hat{\gamma}_0; h_{\hat{\theta}}^{-1}(y))}{L_1(\hat{\gamma}; h_{\hat{\theta}}^{-1}(y))} \frac{L_1(\hat{\gamma}_0; y)}{L_1(\hat{\gamma}_0; h_{\hat{\theta}}^{-1}(y))} \\ &= \widetilde{\mathrm{Q}}_0 \cdot \widetilde{\mathrm{Q}}_1 \cdot \hat{\gamma}_0^{u(y)-u(h_{\hat{\theta}}^{-1}(y))}, \end{aligned}$$

say. The intermediate ratio $\widetilde{\mathrm{Q}}_1$ can be calculated using Markov chain Monte Carlo.

Large values of the test statistics Q are critical. Since their distributions are not known under the respective null hypotheses, simulations are needed in order to evaluate the observed values.

6. Remarks, open questions and related work

The basic idea of the present paper is that of introducing inhomogeneity by transformation. Observing an inhomogeneous point pattern, the problem is then to construct the inverse transformation which can compensate for the inhomogeneity. Similar approaches can be found in a number of related areas. In addition to the examples presented in the introduction, one could mention that Baddeley and van Lieshout (1995, p. 605) discuss the possibility to let the balls, appearing in the area-interaction model, depend on a parameter θ.

Note that our model may be extended, such that densities of the following form are considered

$$f_Y(\{y_1, \ldots, y_n\}) \propto \prod_i \psi(h_{\theta_1}^{-1}(y_i)) \prod_{i<j} \varphi(h_{\theta_2}^{-1}(y_i, y_j)) \cdots .$$

(Personal communication with Jesper Møller.) The approach by Ogata and Tanemura (1986) is then obtained by letting h_{θ_2} be the identity, while our transformation approach corresponds to $h_{\theta_1} = h_{\theta_2}$ and

$$\psi(z) = \varphi(z)/Jh(z).$$

Note also that the model suggested by Ogata and Tanemura (1986) can be used to describe a system of hard discs of fixed diameters with location dependent intensity. Such a system cannot be described by the transformation set–up discussed in the present paper.

The differential equation (1) makes some restrictions on the kind of inhomogeneity which can be described by the transformation approach. It will be of interest to characterize the class of point processes that can be described by this approach.

From a practical point of view, it is going to be important to use concomitant environmental variables to explain the inhomogeneity. In fact, the exponential inhomogeneity as described by (4) is identical to the one considered in Rathbun (1996) if

$$\tau(\eta) = (\tau_1(\eta), \ldots, \tau_l(\eta))$$

is a vector of explanatory variables evaluated at η. Rathbun (1996) considers, however, only the Poisson case. It is included in our future plans to analyze concrete data sets of this type where also the interaction is taken into account.

The estimation procedure developed in Section 5 worked surprisingly well on the simulated example in Example 5.2. The main reason was that the mapping $\theta \to s(h_\theta^{-1}(y))$ was first essentially decreasing and then essentially increasing with a minimum near $\hat{\theta}_0$. Whether this is true in more generality needs to be investigated.

Maximum likelihood estimation is somewhat involved and it is therefore of interest to investigate alternative procedures such as pseudo-likelihood estimation. See Baddeley and Turner (2000) for its implementation in cases where both inhomogeneity and interaction is present. An even simpler procedure would be to use $\hat{\theta}_0$ as estimate of the inhomogeneity parameter. In the example, this appeared to work well. Such a procedure would be justified if the distribution of $t(Y)$ does not depend very much on θ. If so, the extensive work on estimating intensity functions in inhomogeneous Poisson models could then also be applied. These estimation problems are currently under investigation, cf. Nielsen and Jensen (2001).

On the theoretical side it still remains to find conditions that ensure the existence and uniqueness of the maximum likelihood estimates and to develop an asymptotic distribution theory for the maximum likelihood estimators as well as for the likelihood ratio tests.

Appendix

In this appendix, we show that for $\theta \in \mathbb{R}$

$$Jh_\theta^{-1}(\eta) = \beta(\theta)\mathrm{e}^{\theta\tau(m\cdot\eta)}, \quad \eta \in S^2, \tag{A-1}$$

has a unique solution among mappings of the form

$$h_\theta^{-1}(\eta_1,\eta_2,\eta_3) = \left(\frac{\sqrt{1-g_\theta(\eta_3)^2}}{\sqrt{1-\eta_3^2}}\eta_1, \frac{\sqrt{1-g_\theta(\eta_3)^2}}{\sqrt{1-\eta_3^2}}\eta_2, g_\theta(\eta_3)\right), \tag{A-2}$$

where g_θ is an increasing differentiable function of [-1,1] onto [-1,1]. This result is used in Example 4.2.

Let

$$\begin{aligned} p : [0,\pi) \times [0,2\pi) &\longrightarrow S^2 \\ (\omega_1,\omega_2) &\longrightarrow (\sin(\omega_1)\cos(\omega_2), \sin(\omega_1)\sin(\omega_2), \cos(\omega_1)). \end{aligned}$$

be the polar coordinate mapping. Functions of the form (A-2) can then equivalently be described as

$$h_\theta^{-1} = p \circ k_\theta \circ p^{-1}, \tag{A-3}$$

where $k_\theta(\omega_1,\omega_2) = (l_\theta(\omega_1),\omega_2)$ and l_θ is an increasing differentiable bijection on the interval $[0,\pi)$. Here $\circ$ denotes the composition of mappings.

Below, we show the following result

$$Jh_\theta^{-1}(p(\omega_1,\omega_2)) = Jk_\theta(\omega_1,\omega_2)\frac{\sin l_\theta(\omega_1)}{\sin\omega_1} = l'_\theta(\omega_1)\frac{\sin l_\theta(\omega_1)}{\sin\omega_1}. \tag{A-4}$$

In order to prove the first equality, we use the coarea formula and get for an arbitrary function f on S^2 that

$$\begin{aligned} &\int_{S^2} f(h_\theta(\eta))d\eta^2 \\ &= \int_{S^2} f(p \circ k_\theta^{-1} \circ p^{-1}(\eta))d\eta^2 \\ &= \int_0^\pi \int_0^{2\pi} f(p \circ k_\theta^{-1}(\omega_1,\omega_2))\sin\omega_1 d\omega_2 d\omega_1 \\ &= \int_0^\pi \int_0^{2\pi} f(p(\omega_1,\omega_2))Jk_\theta(\omega_1,\omega_2)\sin(l_\theta(\omega_1))d\omega_2 d\omega_1 \\ &= \int_{S^2} f(\eta)Jk_\theta(p^{-1}(\eta))\frac{\sin((k_\theta \circ p^{-1}(\eta))_1)}{\sin(p^{-1}(\eta)_1)}d\eta^2. \end{aligned}$$

From these results, the first equality of (A-4) follows. The next equality follows from the fact that k_θ is a bijection on a subset of $\mathbb{R}^2$ of full dimension and

$$Dk_\theta(\omega_1, \omega_2) = \begin{bmatrix} l'_\theta(\omega_1) & 0 \\ 0 & 1 \end{bmatrix}.$$

Combining (A-1) and (A-4), we find

$$\beta(\theta)\mathrm{e}^{\theta\tau(\cos\omega_1)} = l'_\theta(\omega_1)\frac{\sin l_\theta(\omega_1)}{\sin\omega_1},$$

or

$$(\cos l_\theta(\omega_1))' = -\beta(\theta)\sin(\omega_1)\mathrm{e}^{\theta\tau(\cos\omega_1)}.$$

This equation has the unique solution, among increasing bijections on $[0, \pi)$,

$$\cos l_\theta(\omega_1) = 1 - \int_{\cos(\omega_1)}^{1} \beta(\theta)\mathrm{e}^{\theta\tau(u)}du = g_\theta(\cos\omega_1),$$

say, where

$$\beta(\theta) = \frac{2}{\int_{-1}^{1}\mathrm{e}^{\theta\tau(u)}du}.$$

Therefore, according to (A-3), h_θ^{-1} is of the form stated in (A-2).

Acknowledgement

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. *Statistica 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.

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

Barndorff-Nielsen, O. E. (1978). *Information and Exponential Families in Statistical Theory.* Wiley, Chichester.

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.

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.

Jensen, E. B. V. (1998). *Local Stereology.* World Scientific, Singapore.

Mardia, K. V. (1972). *Statistics of Directional Data.* Academic Press, London.

Meiring, W. (1995). *Estimation of Heterogeneous Space-Time Covariance.* PhD thesis, University of Washington.

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.

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

Monestiez, P., Sampson, P. D., and Guttorp, P. (1993). Modelling of heterogeneous spatial correlation structure by spatial deformation. *Cahiers de Géostatistique*, 3:1–12.

Nielsen, L. S. and Jensen, E. B. V. (2001). Statistical inference for transformation inhomogeneous point processes. *Research Report 12, Laboratory for Computational Stochastics, University of Aarhus.* Submitted.

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.* PhD thesis, Université de Paris I Panthéon-Sorbonne.

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

Ripley, B. D. (1990). Gibbsian interaction models. In *Spatial Statistics: Past, Present and Future*, pages 3–28. Image, New York.

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

Sampson, P. D. and Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *J. Amer. Statist. Assoc.*, 87:108–119.

Smith, R. L. (1997). Detecting signals in climatological data. *Proceedings of 51st Session of the International Statistical Institute, Book 1*, pages 211–214.

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

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.

B

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

MODELLING THE POSITION OF CELL PROFILES ALLOWING FOR BOTH INHOMOGENEITY AND INTERACTION

LINDA STOUGAARD NIELSEN

Laboratory for Computational Stochastics, Department of Mathematical Sciences, University of Aarhus, Denmark. Email: lins@imf.au.dk.

(Accepted October 6, 2000)

ABSTRACT

It is of interest to consider models for point processes that allow for interaction between the points as well as for inhomogeneity in the intensity of the points. Markov point process models are very useful to describe point interaction and can also be used to describe inhomogeneity. A particular type of inhomogeneous Markov point processes obtained by transforming a homogeneous Markov point process will be considered. The position of cell profiles in a 2D section of the mucous membrane in the stomach of a rat will be examined using this model.

Keywords: Inhomogeneity, Interaction, Likelihood, Markov point process, Poisson point process, Strauss point process.

INTRODUCTION

The position of the centres of cells in tissue can be described by a spatial point process, where each cell centre is represented by a point. A point process is a stochastic model for point patterns that describes the properties of the point pattern: Trend in the intensity of the points (inhomogeneity) and interaction between the points.

In surface covering layers (epithelium) such as skin, mucous membranes and cartilage, a gradient can usually be observed in the direction perpendicular to the surface. In planes parallel to the surface homogeneity can be assumed. Thus, it is of interest to consider models for point processes that allow for inhomogeneity in a given direction. Furthermore, such models must be able to describe interaction between points. Between the cell centres there is a natural inhibition since a cell has an extend and two cell centres therefore have to be a certain distance apart. Considering cell division, inhibition on a small scale and clustering on a larger scale might be observed.

Typically we do not observe the full 3 dimensional data but only a 2 dimensional section where the trend is along one of the two dimensions. In this paper we will consider a 2 dimensional section of epithelium and model the position of cell profiles using a point process with a directional gradient and small scale inhibition.

Markov point process models are flexible and tractable models for describing interaction. Homogeneous Markov point processes are well studied and tools for statistical inference well developed, see e.g. Diggle (1983), Geyer (1999) and Baddeley and Turner (2000). Inhomogeneous Markov point processes have been much less considered.

The model proposed by Jensen and Nielsen (2000) is an inhomogeneous Markov point process model, where the inhomogeneity is obtained by applying a bijective differentiable mapping to a homogeneous Markov point process. The interaction is then inherited from the homogeneous Markov model. In cases of inhibition between the points there is strong experimental evidence that the statistical inference can be separated into the statistical inference for a inhomogeneous Poisson point process (model with no interaction) and a homogeneous Markov point process, cf. Jensen and Nielsen (2000). However, this issue still needs a thorough investigation. In this paper we will use this type of statistical inference and will see that it gives a good description of the epithelium data.

MATERIAL AND METHODS

The image in Figure 1 was supplied by Thomas F. Bendtsen. The image shows a 2D section of the mucous membrane of the stomach of a healthy rat, where the darker spots are the cell profiles. In the top of the image we have the cavity of the stomach, and in the bottom of the image the muscle tissue begins. In this paper the profiles only in the marked region are studied. In right hand-side of Figure 1 the position of the nuclei in this region are shown. For convenience, the window is scaled into the window $[0,893]\times[0,1]$. In what follows, we find a model that describe this point pattern.

As mentioned in the introduction there is an obvious directional trend in the point pattern and small scale inhibition since two cells cannot overlap. In the remains of this paper, the point pattern will be analyzed using the class of Markov point processes by transformation, which will be introduced below.

Let X denote a 2 dimensional homogeneous Markov point process, and let γ be the parameter that describes the point interaction. In Markov point processes, interactions are only allowed between neighbour points. A typical example of a Markov point process is the Strauss point process, cf. Strauss (1975), where two points are neighbours if their distance is less than some constant $R>0$. The probability density of the point process is controlled by $\gamma^{s(x)}$, where $s(x)$ is the number of neighbours in a given point pattern x. Then, if $0<\gamma<1$, the tendency of points rejecting each other increases the smaller γ is. A value of γ equal to 0 is the hard-core model where no pair of points have distance less than R, and $\gamma=1$ is the Poisson model, the model with no interaction at all.

Let $h_\theta(\eta,\xi)=(\eta,l_\theta(\xi))$ be a differentiable bijective parametrized mapping. Then, according to Jensen and Nielsen (2000), $Y=h_\theta(X)$ is a Markov point process, inhomogeneous in the direction of the second coordinate, and with density

$$f_Y(y)=f_X(h_\theta^{-1}(y))\prod_{i=1}^{n(y)}(l_\theta^{-1})'(y_{i2}), \tag{1}$$

where f_X is the density of X, y_{i2} is the second coordinate of the i'th point y_i of y and $n(y)$ is the number of points in y. As f_X controls the interaction between the points, the interaction between points in the transformed point process Y is inherited from the homogeneous model X. The differential term $(l_\theta^{-1})'$ controls the inhomogeneity. For more detail, see Jensen and Nielsen (2000). Note that the first coordinates are not transformed, and therefore the model (1) only describes inhomogeneity in the second coordinate direction.

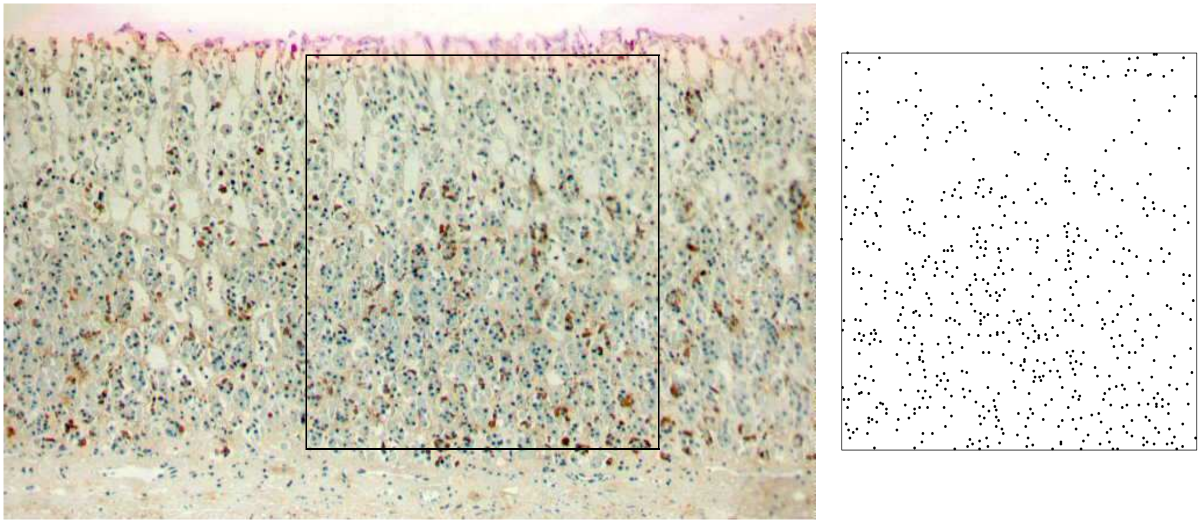


Figure 1: *Image by Thomas F. Bendtsen showing a 2D section of the mucous membrane in the stomach of a healthy rat. To the right, the positions of the nuclei of the cell profiles in the marked region of the image are plotted. This is the point pattern considered in the present paper.*

In applications such as the point pattern from Figure 1, we have an inhomogeneous point pattern which we want to describe by a point process model. Using the transformation approach, we assume that the point pattern in question, say y, is a realization of a transformation-inhomogeneous point process Y. The aim is then to find a transformation such that the back-transformed point pattern $x = h_\theta^{-1}(y)$ is a homogeneous point pattern. Finally, well-known theory for homogeneous point patterns can be applied in order to find a model X describing x.

The problem of finding a suitable transformation is not simple. The density (1) can be split into the density for a homogeneous point process and that of an inhomogeneous Poisson point process. However, the inhomogeneity parameter θ enters into both parts, and so maximum likelihood estimation of θ is not based on the inhomogeneous Poisson part alone. In practice however (simulation studies), the estimate $\hat{\theta}_0$ based on the inhomogeneous Poisson point process seems to be very good to describe the inhomogeneity. Thus, $\hat{\theta}_0$ will here be used as an estimate of θ.

In order to analyze a homogeneous point pattern we first consider summary statistics in order to find the characteristics of the particular point pattern such that a suitable parametric model can be proposed, see e.g. Diggle (1983). Then the parameters in the proposed Markov point process model are estimated using likelihood or pseudo-likelihood theory, see Geyer (1999) and Baddeley and Turner (2000). Notice that it is also possible to use likelihood and pseudo-likelihood theory in the full inhomogeneous model. This is however much more complicated, see Jensen and Nielsen (2000).

RESULTS

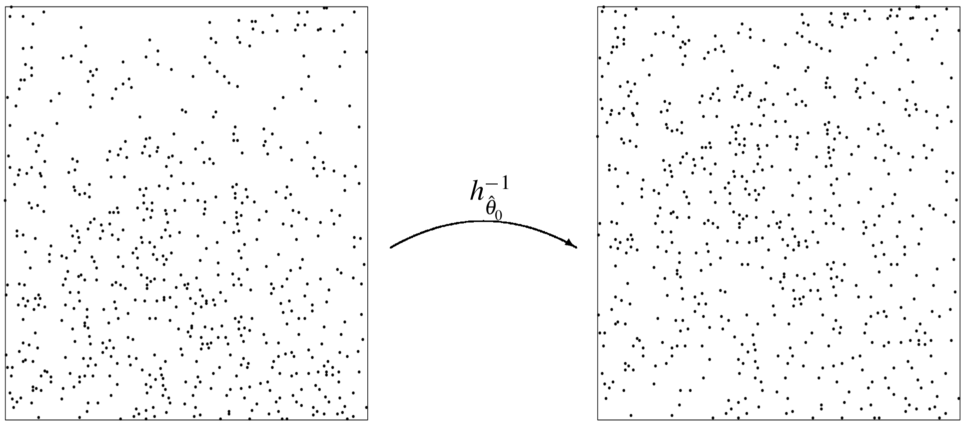
To describe the inhomogeneity in the second coordinate direction of the point pattern from Figure 1, we will consider exponential inhomogeneity, i.e.

$$(l_\theta^{-1})'(\eta) \propto \mathrm{e}^{\theta\eta}.$$

The following mapping $l_\theta : [0,1] \to [0,1]$ has the above property,

$$l_\theta(\eta) = \frac{1}{\theta} \log\left(\eta(\mathrm{e}^\theta - 1) + 1\right).$$

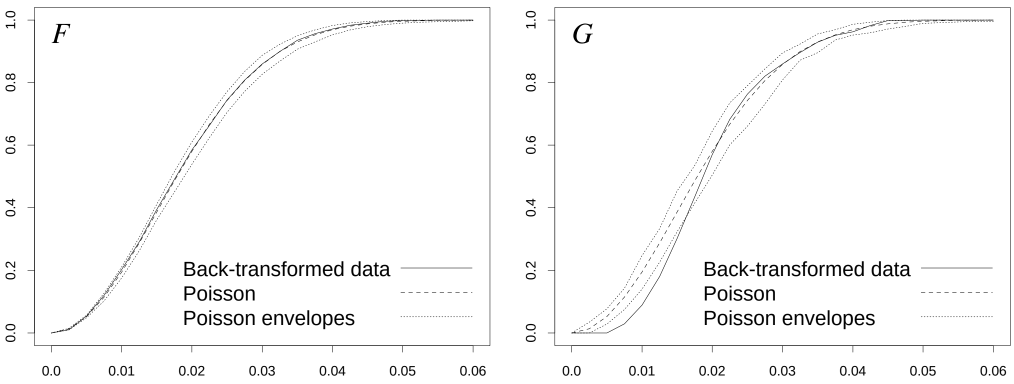
In Figure 2 (b) the back-transformed data point pattern is shown, where the inhomogeneity parameter used in the transformation is



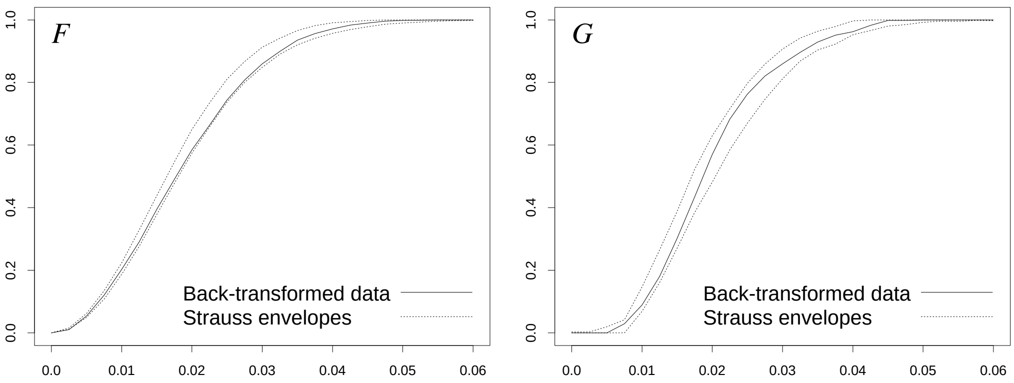
(a) Centres of cell profiles

(b) Back-transformed centres

Figure 2: *Point pattern y from Figure 1 of cell profile centres (a) and the corresponding back-transformed point pattern $x = h_\theta^{-1}(y)$ (b). The transformation parameter is the maximum likelihood estimate in the inhomogeneous Poisson point process model,* $\hat{\theta}_0 = 1.3043$.



(a) Back-transformed data vs Poisson model



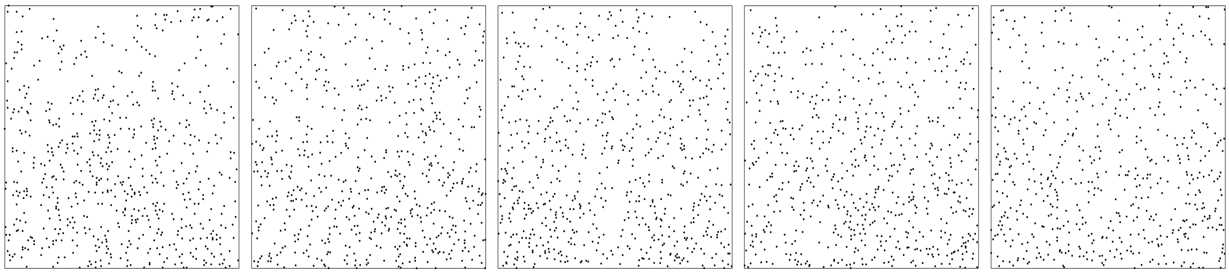
(b) Back-transformed data vs regular Strauss model

Figure 3: *Empty space function F and nearest neighbour function G for the back-transformed data point pattern versus the Poisson point process (a) and the Strauss point process with parameters $\beta = 760$, $\gamma = 0.09$ and $R = 0.007$ (b). Envelopes are the pointwise maxima and minima of 39 simulated processes.*

the maximum likelihood estimate $\hat{\theta}_0$ in the inhomogeneous Poisson model. Notice that this point pattern appears to be homogeneous.

To determine the interaction structure of a homogeneous point pattern, we compare the point pattern with a homogeneous Poisson point process. This is typically done by comparing the F, G, J, K and pair-correlation functions. See e.g. Diggle (1983), Stoyan *et al.* (1995) and van Lieshout and Baddeley (1996) for more details on these functions. In what follows, I have considered only the F and G functions, but the conclusions would be the same using any of the other functions. In the first plot of Figure 3 (a) the F function for the back-transformed data and for the homogeneous Poisson point process are plotted. The F-function can be thought of as a statistic that measures the amount of empty space in the point pattern. As seen, the back-transformed point pattern has the same amount of empty space as the Poisson point process. In the second plot of Figure 3 (a), the G function is shown. This function is the probability function for the nearest neighbour distances in the point pattern. As seen, the back-transformed point pattern has less of the small nearest neighbour distances than the Poisson point process, which is a sign of small scale inhibition. This is of course due to the fact that we only consider the centres of the cell profiles.

A Strauss point process is a useful model for



(a) Data

(b) Transformed simulations of regular Strauss point processes

Figure 4: *Plot of the data points (a) and four simulations from the exponential inhomogeneous Strauss point process (b).*

describing inhibition. In Figure 3 (b) the data is compared with a Strauss point process with parameters $\beta = 760$, $\gamma = 0.09$ and $R = 0.007$. Due to the very small distance R, the amount of empty space is close to that of the Poisson point process. The amount of very small nearest neighbour distances is reduced. From the two plots we conclude that the back-transformed data point pattern can be assumed to be a realization of the Strauss point process with parameters as above. Hence, the data point pattern can be regarded a realization of the exponential inhomogeneous Strauss point process with density

$$f_Y(y) = \alpha(\beta,\gamma)\beta^{n(y)}\gamma^{s(h_\theta^{-1}(y))}\left(\frac{\theta}{e^\theta - 1}\right)^{n(y)} \prod_{i=1}^{n(y)} e^{\theta y_{i1}}, \quad (2)$$

where $\theta = 1.3043$, $\beta = 760$, $\gamma = 0.09$ and $R = 0.007$. In Figure 4 the data is plotted together with four simulations from this point process.

Note that, except for the small scale inhibition caused by non-overlapping cells, the data point pattern is very close to being an inhomogeneous Poisson point process, cf. Figure 3 (a). Thereby the estimate $\hat{\theta}_0$ based on the inhomogeneous Poisson model is a very accurate estimate.

DISCUSSION

It appears as if there is also a gradient in the first coordinate of the data point pattern. The maximum likelihood estimate based on the inhomogeneous Poisson point process is $\hat{\theta}_{02} = -0.27159$. The estimate is very small and it could be interesting to perform a statistical test: $\theta_2 = 0$. More interesting would be to compare gradients within various sections from the same rat. Probably the gradient in the first coordinate can be explained as variation, while the gradient in the second direction is significant. It could also be interesting to compare different rats and to compare healthy rats with rats treated with anti-acid drugs.

Furthermore, I have not distinguished between the type of cells in the mucous membrane. There is much interest in a particular type of cells, the ECL cells, and it could be interesting to test whether these cells are uniformly distributed among all the cells. This is the case if the gradient for the ECL cells is the same as the gradient for all the cells.

Note that, in this paper we have modelled a 2D section of a 3D structure. Supposing that we have the full 3D data set, the same model could be used where now X is a 3D homogeneous point process, the transformation is $h_\theta(\eta,\xi,\nu) = (\eta, l_\theta(\xi), \nu)$ and the inhomogeneous model is still of the form (1). However, we do not have information about the third dimension, and this will most often be the situation in reality. Conclusions about the model for the 3D data based on the model obtained from a 2D section is not straightforward. Methods used in Stereology may be useful. See Stoyan *et al.* (1995, Section 11.6.2) and Hahn *et al.* (1999).

Finally, it should be noticed that the transformation model only describes the *position* of the cell profile centres and not the whole image. Transforming the image would give homogeneous distributed positions but inhomogeneous distributed profile sizes. The image should be described by a marked point process, see e.g. Baddeley and Møller (1989), with profile positions modelled by (2) and marks of the points being the sizes and shapes of the cell profiles.

A preliminary report was presented at the Xth International Congress for Stereology, Melbourne, Australia, 1-4 November 1999.

ACKNOWLEDGEMENTS

I would like to thank Thomas F. Bendtsen for letting me use his data and Hans Jørgen G. Gundersen for bringing data to my attention. Furthermore, I thank both Thomas and Hans Jørgen for their effort in making me understand some of the biology behind the image.

I am also very grateful to my supervisor Eva B. Vedel Jensen, to Ute Hahn and to the anonymous referee for correcting errors and improving the paper.

REFERENCES

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

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

Diggle PJ (1983). Statistical analysis of spatial point patterns. Academic Press.

Geyer CJ (1999). Likelihood inference for spatial processes. In: Barndorff-Nielsen OE, Kendall WS, van Lieshout MNM, eds. Stochastic Geometry, Likelihood and Computation. Chapman and Hall/CRC: Boca Raton.

Hahn U, Micheletti A, Pohlink R, Stoyan D, Wendrock H (1999). Stereological analysis and modelling of gradient structures. J Microsc 195:113–124.

Jensen EBV, Nielsen LS (2000). Inhomogeneous Markov point processes by transformation. Bernoulli 6:761–782.

Stoyan D, Kendall WS, Mecke J (1995). Stochastic geometry and its Applications. Chichester: Wiley & sons, 2nd edition.

Strauss DJ (1975). A model for clustering. Biometrika 62:467–475.

van Lieshout MNM, Baddeley AJ (1996). A nonparametric measure of spatial interaction in point patterns. Statistica Neerlandica 50:344–361.

C

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., *Selected Proceedings of the Symposium on Inference for Stochastic Processes*. IMS Lecture Notes, volume 37: 297–318.

Institute of Mathematical Statistics

LECTURE NOTES — MONOGRAPH SERIES

A review on inhomogeneous Markov point processes

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

ABSTRACT

Recent models for inhomogeneous spatial point processes with interaction are reviewed. The focus is on models derived from homogeneous Markov point processes. For some of the models, the interaction is location dependent. A new type of transformation related model with this property is also suggested. The statistical inference based on likelihood and pseudolikelihood is discussed for the different models. In particular, it is shown that for transformation models, the pseudolikelihood function can be decomposed in a similar fashion as the likelihood function.

Keywords: Cox point processes, Gibbs point processes, inhomogeneity, interaction, likelihood, Markov point processes, Papangelou conditional intensity, Poisson point processes, pseudolikelihood, thinning, transformation

1 Introduction

In recent years, models for inhomogeneous point processes with interaction have been suggested by several authors. We will in the present paper concentrate on three ways of introducing inhomogeneity into a Markov model, i.e. inhomogeneity induced by a non-constant first-order interaction (Stoyan and Stoyan (1998); see also Ogata and Tanemura (1986)), by thinning of a homogeneous Markov point process (Baddeley et al. (2000)) and by transformation of a homogeneous Markov point process (Jensen and Nielsen (2000)). The aim is to give a unified exposition of these models in order to be able to assess their relative merits and point to research problems that remain to be solved in this area.

We restrict attention to finite point processes. For any of the three point process models to be considered, the inhomogeneity may be described by a function λ defined on the same set as the points. In the case where the inhomogeneous point process is Poisson, λ is the ordinary intensity function. In addition to the point pattern, explanatory variables may be observed at each point, for the purpose of explaining the inhomogeneity. Such information may be included in any of the models. The interaction specified in the models may or may not be location dependent.

In Section 2, inhomogeneous Poisson and Cox point processes are considered. In Section 3.1, a short summary of homogeneous Markov point processes is given, followed in Section 3.2 by a formal description of the three types of inhomogeneous point processes derived from homogeneous Markov point processes. In Section 3.3, parametric specification of the inhomogeneity is discussed, while in Section 3.4 parametric statistical inference is outlined for each of the three model types. Section 4 contains a discussion and some considerations concerning future research.

2 Poisson and related point processes

We will throughout the paper consider point processes on a $k-$dimensional manifold $\mathcal{X}$ in R^m. Often, $\mathcal{X}$ will be full-dimensional such that $k = m$. (Formally, we will call $\mathcal{X}$ full-dimensional if $\mathcal{X}$ is a regular compact, that is a non-empty compact subset of R^m which is the closure of its interior.) But $\mathcal{X}$ may for instance also be a planar curve or a spatial surface. We will assume that $0 < \lambda_m^k(\mathcal{X}) < \infty$ where λ_m^k is the $k-$dimensional volume measure in R^m (Hausdorff measure). We let $\mathcal{B}_0$ be the bounded Borel subsets of $\mathcal{X}$.

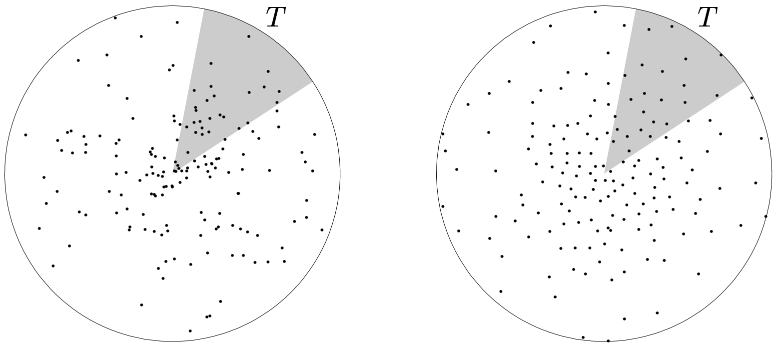
It is easy to introduce inhomogeneity within the class of Poisson point processes. Let μ be a locally finite, non-atomic measure on $\mathcal{X}$ with density λ with respect to λ_m^k and let $n(\cdot)$ be the number of elements in $\cdot$. A point process X on $\mathcal{X}$ (a random finite subset of $\mathcal{X}$) is then said to be a Poisson point process with *intensity function* λ if

- $\forall A \in \mathcal{B}_0 : n(X \cap A) \sim Po(\int_A \lambda(u) du^k)$
- $\forall A_1, \ldots, A_s \in \mathcal{B}_0$ disjoint : $n(X \cap A_1), \ldots, n(X \cap A_s)$ independent

We have here used the short notation du^k for $\lambda_m^k(du)$. It can be shown that the first requirement is enough, cf. e.g. Stoyan et al. (1995).

If λ is constant the Poisson point process is said to be homogeneous, otherwise the process is inhomogeneous. A homogeneous Poisson point process is often used as a reference (null) model. The reason is the following result

- Let X be a homogeneous Poisson point process on $\mathcal{X}$ and let $A \in \mathcal{B}_0$. Then, given $n(X \cap A) = n, X \cap A$ is a binomial process with n points, i.e. $X \cap A$ is distributed as $\{X_1, \ldots, X_n\}$ where $X_1, \ldots, X_n$ are independent and identically uniformly distributed on A.



(a) Poisson (b) Strauss

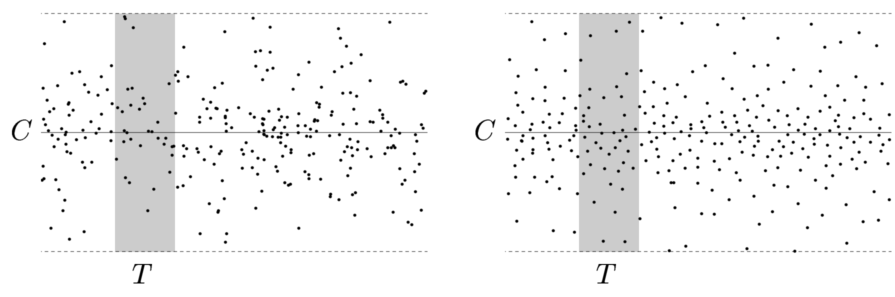
Figure 1: Realizations of inhomogeneous point processes in the unit circular disc with inhomogeneity function $\lambda(\eta) \propto \mathrm{e}^{\theta d_C(\eta)}$, where $d_C(\eta)$ is the distance from η to C, the centre of the disc. The point pattern to the left is inhomogeneous Poisson, i.e. no interaction between points, whereas the point pattern to the right is inhomogeneous Strauss with $\gamma = 0.01$ and therefore shows inhibition between points. For details, see Section 3.3 and Appendix III. The number of points in the Poisson process have been chosen to equal the number of points in the Strauss process. For both processes, the distribution of the points in the shaded sampling window T remains the same if T is rotated around C.

The independence property of the homogeneous Poisson point process ensures that there is no interaction in the process, the uniformity means that the process is homogeneous.

The inhomogeneity of a point process may depend on explanatory variables. One simple geometric example is an intensity function of the form

$$\lambda(\eta) = g(d_C(\eta)), \quad \eta \in \mathcal{X},$$

where $d_C(\eta)$ is the distance from η to a reference structure C. For $m = k = 2$, the reference structure may be a point or a planar curve. In Figure 1, C is a point (centre of a circle) while, in Figure 2, C is a straight line (centre of a linear band). For $m = k = 3$, the reference structure may be a point, a spatial curve or a spatial surface. See also Berman (1986) and references therein. Points lying *on* curves in two or three dimensions or points lying *on* spatial surfaces may also show inhomogeneity. In Figure 3 and 4, point processes on the unit circle S^1 and unit sphere S^2 are shown. (Points are represented as directions in Figure 3.) In any of the Figures 1 to 4, Poisson point processes are shown to the left while corresponding processes with



(a) Poisson (b) Strauss

Figure 2: Realizations of inhomogeneous point processes in the unit band $\{\eta \in \mathbb{R}^2 : d_C(\eta) < 1\}$, where $d_C(\eta)$ is the distance from η to C, the full-drawn horizontal line. The inhomogeneity function is as in Figure 1 and likewise the right hand-side point pattern is inhomogeneous Strauss and the left is inhomogeneous Poisson with the same number of points. The distribution of points in the shaded sampling window T remains the same under horizontal translations of T.

inhibition between points are shown to the right (for details, see Section 3.3).

Statistical inference for inhomogeneous Poisson processes with a parametrically specified intensity function can be performed as follows. Let X be an inhomogeneous Poisson point process with intensity function $\lambda_\theta, \theta \in \Theta \subseteq R^l$. Then, the likelihood function of θ with respect to the homogeneous Poisson point process with intensity 1 takes the form

$$L_0(\theta; x) = \exp(-\int_{\mathcal{X}} [\lambda_\theta(u) - 1] du^k) \prod_{\eta \in x} \lambda_\theta(\eta). \tag{1}$$

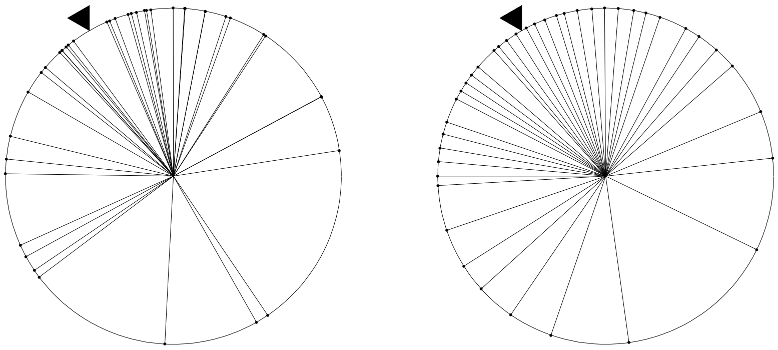
We use index 0 in this likelihood because later it enters into more complicated likelihoods. In Berman and Turner (1992), log-linear models for λ_θ are discussed,

$$\lambda_\theta(\eta) \propto \mathrm{e}^{\theta \cdot \tau(\eta)}, \quad \eta \in \mathcal{X},$$

where $\tau(\eta) = (\tau_1(\eta), \ldots, \tau_l(\eta))$ is a list of explanatory variables evaluated at η and $\cdot$ indicates Euclidean inner product. After approximation of the integral by a finite sum, the likelihood takes the same analytical form as the likelihood of a generalized linear model with Poisson responses and standard software can be used to analyze the model. See also Rathbun (1996).

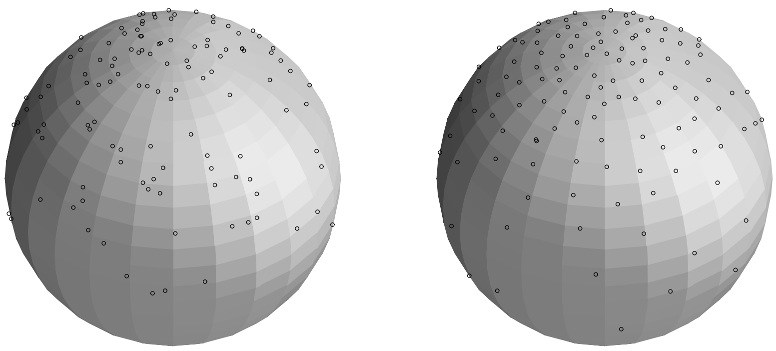
Alternatively, the intensity function can be estimated non-parametrically, using kernel estimation (Silverman (1986)) or a Bayesian method (Heikkinen and Arjas (1998)).

C



(a) Poisson (b) Strauss

Figure 3: Realizations of inhomogeneous point processes on the unit circle S^1. The situation is the same as in Figure 1 except that $d_C(\eta)$ is the distance along the circle from η to the point $C = (\cos(2\pi/3), \sin(2\pi/3))$ marked with an arrow. The points are shown as directions.



(a) Poisson (b) Strauss

Figure 4: Realizations of inhomogeneous point processes on the unit sphere S^2. The situation is the same as in Figure 1 except that $d_C(\eta)$ is the geodesic distance from η to C which is the north pole.

As a generalization one may consider inhomogeneous Cox processes, i.e. inhomogeneous doubly stochastic Poisson point processes. The definition of a Cox process is as follows. Let Λ be a random intensity function on $\mathcal{X}$. Then, X is a Cox process if, given $\Lambda = \lambda$, X is a Poisson point process with intensity function λ, cf. Stoyan et al. (1995, p. 154).

In Møller et al. (1998) and Brix and Møller (1998), log Gaussian Cox processes are discussed, i.e. Cox processes for which $\lambda = e^Y$ and $Y = \{Y_s\}_{s \in \mathcal{X}}$ is a Gaussian field. Inhomogeneity is introduced by letting the mean-value of Y_s depend on s, see also Møller (1999a). *Clustered* inhomogeneous point patterns may be modelled by this process and this appears to be a natural model if the aggregation is due to stochastic environmental heterogeneity. This type of model has in Brix and Møller (1998) been used to describe the spatio-temporal development of two types of weeds in an organic barley field.

3 Markov point processes

If one wants to describe inhibition in addition to clustering then the class of Markov point processes is useful, cf. Ripley and Kelly (1977), Baddeley and Møller (1989), the recent monograph van Lieshout (2000) and references therein. Let us start by recalling a few preliminaries for Markov point processes.

3.1 Homogeneous Markov point processes

Let $\sim$ be a reflexive and symmetric relation on $\mathcal{X}$. Two points $\xi, \eta \in \mathcal{X}$ are called neighbours if $\xi \sim \eta$. A finite subset x of $\mathcal{X}$ is called a *clique* if all points of x are neighbours. By convention, sets of 0 and 1 points are cliques. The set of cliques is denoted $\mathcal{C}$.

If $\mathcal{X} \subseteq R^m$ is full-dimensional then the relation induced by Euclidean distance is often used. If $\mathcal{X}$ is a planar curve, distances along the curve may be more natural. If $\mathcal{X}$ is the unit sphere S^{m-1}, such that the observed points in fact are directions, then geodesic distance is natural.

Markov point processes are characterized by the Hammersley-Clifford theorem, cf. Ripley and Kelly (1977). This theorem states that a point process X on $\mathcal{X}$, with density f with respect to the homogeneous Poisson point process with intensity 1, is a Markov point process iff

$$f(x) = \prod_{y \subseteq x} \varphi(y), \tag{2}$$

where $\varphi \geq 0$ is an interaction function with respect to $\sim$, i.e. $\varphi(x) = 1$ unless $x \in \mathcal{C}$. Note that then the Papangelou conditional intensity

$$\lambda(\eta; x) = \frac{f(x \cup \eta)}{f(x)}, \quad \eta \in \mathcal{X} \setminus x,$$

depends only on those points in x which are neighbours of η. Here, $x \cup \eta$ is short for $x \cup \{\eta\}$.

Let φ_k be the restriction of φ to subsets consisting of k points. A pairwise interaction process is then a process for which $\varphi_k \equiv 1$ for $k > 2$. The famous Strauss process (Strauss (1975) and Kelly and Ripley (1976)) is the pairwise interaction process with

$$\varphi_k(x) = \begin{cases} \alpha & k = 0 \\ \beta & k = 1 \\ \gamma & k = 2,\ x \in \mathcal{C}. \end{cases}$$

The density of the Strauss process becomes, cf. (2),

$$f(x) = \alpha\beta^{n(x)}\gamma^{s(x)},$$

where $s(x)$ is the number of neighbour pairs in x. If $\|\cdot\|$ denotes the usual Euclidean distance and the neighbourhood relation is given by

$$\eta \sim \xi \iff ||\eta - \xi|| < R,$$

then the process is called a Strauss process with interaction radius R.

If $\mathcal{X} \subseteq R^m$ is full-dimensional, a Markov point process X on $\mathcal{X}$ is said to be homogeneous if φ is translation invariant, cf. e.g. Stoyan and Stoyan (1998) and Baddeley et al. (2000). (We assume that φ is defined on all finite subsets of R^m.) Other definitions of homogeneity are of course possible, cf. Jensen and Nielsen (2000). Note that translation invariance implies that φ_1 is constant and for $k > 1$, $\varphi_k(y)$ only depends on the relative positions of the k points in y. A homogeneous pairwise interaction process has a density of the form

$$f(x) = \alpha\beta^{n(x)} \prod_{\{\eta,\xi\}\subseteq x}^{\neq} u(\eta - \xi), \tag{3}$$

where $\neq$ indicates that η and ξ are different. For lower dimensional manifolds $\mathcal{X}$, homogeneity may be defined in terms of invariance under other types of transformations. For instance, for $\mathcal{X} = S^{m-1}$ a natural set of transformations are the rotations. Recall that the group $O(m)$ of rotations consists of $m \times m$ real matrices

$$O(m) = \{A | AA^T = A^T A = I_m\}.$$

A homogeneous, with respect to this choice, pairwise interaction process on S^{m-1} has a density of the form

$$f(x) = \alpha\beta^{n(x)} \prod_{\{\eta,\xi\}\subseteq x}^{\neq} u(\eta \cdot \xi).$$

Recall that $\eta \cdot \xi = \cos\theta$, where θ is the angle between η and ξ.

3.2 Introducing inhomogeneity

Throughout this section, X is a homogeneous Markov point process with respect to $\sim$ on $\mathcal{X}$ and density

$$f_X(x) = \prod_{y \subseteq x} \varphi(y). \tag{4}$$

Note that φ_1 is then constant. Below, we describe three ways of introducing inhomogeneity into the model. The resulting inhomogeneous point process is denoted by Y and is a point process on a $k-$dimensional manifold $\mathcal{Y}$ in R^d, say. For the first two ways of constructing inhomogeneity, $\mathcal{Y} = \mathcal{X}$, i.e. the homogeneous process and the associated inhomogeneous process are defined on the same space.

The inhomogeneity is described by a function $\lambda(\eta)$, $\eta \in \mathcal{Y}$, which we will call the inhomogeneity function. Common to each of the three constructions is the feature that if X is a homogeneous Poisson point process, then the associated inhomogeneous point process Y is Poisson with intensity function proportional to λ. The three inhomogeneous models are therefore extensions of the inhomogeneous Poisson model.

An obvious way of introducing inhomogeneity is by making the first-order interaction non-constant. We will call this type I inhomogeneity. The associated inhomogeneous Markov point process has then a density of the form

$$f_Y(y) \propto \prod_{\eta \in y} \lambda(\eta) \prod_{z \subseteq y} \varphi(z). \tag{5}$$

This type of model is natural if the interaction does not depend on the local intensity of points. This set-up has been studied in Ogata and Tanemura (1986), Stoyan and Stoyan (1998) and Baddeley and Turner (2000), among others. In Ogata and Tanemura (1986), $\log \lambda(\eta)$ is a polynomium in Cartesian coordinates, while in Stoyan and Stoyan (1998), a piecewise (region-wise) constant function is studied. It is also interesting to note that in the hierarchical point process models described in Högmander and Särkkä (1999), densities of the form (5) appear.

Type II inhomogeneity is obtained by using an independent inhomogeneous thinning of the homogeneous Markov point process. Let us suppose that the inhomogeneity function $\lambda(\eta)$, $\eta \in \mathcal{Y}$, is bounded by $\lambda_{\max}$, and let $p(\eta) = \lambda(\eta)/\lambda_{\max}$, $\eta \in \mathcal{Y}$. The inhomogeneous process is then obtained by thinning with p,

$$Y = \{x_i \in X : U_i \leq p(x_i)\},$$

where $\{U_i\}$ is a sequence of independent and identically uniformly distributed random variables in $[0,1]$, independent of X. In Baddeley et al. (2000), this approach is suggested and studied in detail. According to them this model is natural if p can be interpreted as the probability of survival of a plant or the probability of observing an animal in a wildlife population. A possibly less appealing property of the thinned Markov process is that it is non-Markovian except if X is Poisson. However, this does not complicate the likelihood inference, and an extension of Ripley's $K-$function applies, cf. Baddeley et al. (2000).

A third way of introducing inhomogeneity is by applying a 1-1 transformation on a homogeneous Markov point process, cf. Jensen and Nielsen (2000). This is type III inhomogeneity. The idea of using transformations to introduce inhomogeneity has also been used for the modelling of the covariance structure of a non-stationary spatial process, cf. Perrin (1997) and references therein.

Let $h : \mathcal{X} \to \mathcal{Y}$ be a 1-1 differentiable mapping. We consider on $\mathcal{Y}$ the induced relation

$$\eta_1 \approx \eta_2 \Longleftrightarrow h^{-1}(\eta_1) \sim h^{-1}(\eta_2), \quad \eta_1, \eta_2 \in \mathcal{Y}.$$

Using the induced relation, interactions become location dependent. Note that we in this case have inhomogeneity both in the intensity and the strength of the interaction. The transformation approach may be extended by using a series of transformations.

It can be shown, cf. Jensen and Nielsen (2000, Corollary 3.3), that

$$h(X) = \{h(\xi) : \xi \in X\}$$

is Markov with respect to $\approx$ on $\mathcal{Y}$ and has the density

$$f_Y(y) = \exp(-\int_{\mathcal{Y}} [\lambda(\eta) - 1] d\eta^k) \prod_{\eta \in y} \lambda(\eta) \prod_{z \subseteq y} \varphi(h^{-1}(z)), \tag{6}$$

where $\lambda(\eta) = Jh^{-1}(\eta)$, the Jacobian of the inverse transformation h^{-1}. This transformation result can be proved by the coarea formula in geometric measure theory, cf. Jensen (1998).

Note that if X is Poisson, then the last product in (6) is of the form

$$\exp(-(\beta - 1)\lambda_m^k(\mathcal{X}))\beta^{n(y)},$$

and therefore Y is an inhomogeneous Poisson point process with intensity function $\beta\lambda(\cdot)$.

It is not always easy to find an appropriate transformation which introduces an inhomogeneity of a given form. (The problem to be solved is to find h such that $Jh^{-1} = \lambda$ where λ is a given inhomogeneity function.) It is therefore useful to construct approximate transformation models with the same qualitative properties as the original transformation models. Let us suppose that $d = m = k$, i.e. the manifolds $\mathcal{X}$ and $\mathcal{Y}$ are full-dimensional. Furthermore, suppose that the original process is a homogeneous pairwise interaction process, cf. (3). Then, the density of the transformed point process is, cf. (6),

$$f_Y(y) \propto \beta^{n(y)} \prod_{\eta \in y} \lambda(\eta) \prod_{\{\eta,\xi\} \subseteq y}^{\neq} u(h^{-1}(\eta) - h^{-1}(\xi)).$$

Recalling that for a transformation model $Jh^{-1} = \lambda$, an obvious way of avoiding to construct the transformation is to replace

$$h^{-1}(\eta) - h^{-1}(\xi) \tag{7}$$

by an expression of the form

$$\lambda(\eta)^{\nu} \cdot \lambda(\xi)^{\nu}(\eta - \xi), \tag{8}$$

where $\nu \geq 0$ is some suitably chosen power. The density of the transformation related model becomes

$$f_Y(y) \propto \beta^{n(y)} \prod_{\eta \in y} \lambda(\eta) \prod_{\{\eta,\xi\} \subseteq y}^{\neq} u(\lambda(\eta)^{\nu} \lambda(\xi)^{\nu}(\eta - \xi)). \tag{9}$$

This type of model has also been considered in Baddeley and Turner (2000). Note that for point processes on the real line ($k = 1$), (8) can for η and ξ close be regarded as an approximation to (7), if $\nu = 1/2$. This will generally not be the case for $k > 1$.

3.3 Exponential inhomogeneity

The inhomogeneity function may be modelled parametrically or non-parametrically or both. If no prior knowledge is available about the inhomogeneity, non-parametric modelling may be useful, at least initially. With knowledge of the inhomogeneity (e.g. monotone decreasing in a known direction) then it can be worthwhile to consider parametrically modelled inhomogeneity such as that of exponential form

$$\lambda_{\theta}(\eta) = \alpha(\theta) \mathrm{e}^{\theta \cdot \tau(\eta)}, \quad \eta \in \mathcal{Y},$$

where $\theta \in \Theta \subseteq R^l$ and $\tau(\eta) \in R^l$.

Let us concentrate on a comparison between type I and type III exponential inhomogeneity. If the inhomogeneity is of type I, then the density of the inhomogeneous point process takes the form, cf. (5),

$$f_Y(y; \theta) \propto \alpha(\theta)^{n(y)} \mathrm{e}^{\theta \cdot t(y)} \prod_{z \subseteq y} \varphi(z), \tag{10}$$

C

where $t(y) = \sum_{\eta \in y} \tau(\eta)$. Note that if the homogeneous Markov point process is an exponential family model then the associated inhomogeneous model is too. In particular, if the homogeneous model is a Strauss model then

$$f_Y(y; \theta, \beta, \gamma) \propto \mathrm{e}^{\theta \cdot t(y)} (\alpha(\theta)\beta)^{n(y)} \gamma^{s(y)}. \tag{11}$$

This is a nice three parameter exponential family model.

If instead the transformation approach is used, we need to find a parametrized class of transformations $h_\theta, \theta \in \Theta$ such that

$$J h_\theta^{-1}(\eta) = \alpha(\theta) \mathrm{e}^{\theta \cdot \tau(\eta)}, \quad \eta \in \mathcal{Y}. \tag{12}$$

Let us give a fairly general geometric example where this problem has a simple solution.

Example 3.1 The example concerns inhomogeneity for point patterns in R^d which depends on the distance d_C to a $p-$dimensional linear subspace C in R^d, $p = 0, 1, \ldots, d-1$. We will define the transformation h_θ on the whole set $\{\eta \in R^d : d_C(\eta) \leq 1\}$ and let $\tau(\eta)$ only depend on the distance of η to C, i.e. $\tau(\eta) = \tilde{\tau}(d_C(\eta))$, say. The cases $(d,p) = (2,0)$ and $(2,1)$ with $\tilde{\tau}$ the identity are illustrated in Figure 1 and 2, respectively.

Then, (12) has a unique solution among transformations of the form

$$h_\theta(\eta) = p_C(\eta) + g_\theta(d_C(\eta)) \frac{\eta - p_C(\eta)}{d_C(\eta)}, \tag{13}$$

where $p_C(\eta)$ is the orthogonal projection onto C and g_θ is an increasing function of $[0,1]$ onto itself. The solution is given by, cf. Appendix I,

$$g_\theta^{-1}(r) = \left[\frac{\int_0^r u^{d-p-1} \mathrm{e}^{\theta \cdot \tilde{\tau}(u)} du}{\int_0^1 u^{d-p-1} \mathrm{e}^{\theta \cdot \tilde{\tau}(u)} du} \right]^{1/(d-p)}. \tag{14}$$

It follows also from Appendix I that for h_θ defined by (13) and (14),

$$J h_\theta^{-1}(\eta) = \alpha(\theta) \mathrm{e}^{\theta \cdot \tilde{\tau}(d_C(\eta))},$$

where

$$\alpha(\theta) = \left[(d-p) \int_0^1 u^{d-p-1} \mathrm{e}^{\theta \cdot \tilde{\tau}(u)} du \right]^{-1}.$$

For $p > 0$, this model may be used locally also in the case where C is curved. □

Likewise, it is possible to construct transformations on S^1 or S^2 for the case of exponential inhomogeneity with $\tau(\eta) = d_C(\eta)$, where C is a point on S^1 or S^2 and d_C is the geodesic distance to C, cf. Jensen and Nielsen

(2000). Illustrations are given in Figure 3 and 4. For general functions τ, the construction of an appropriate set of transformations may be difficult. An example is the inhomogeneity of the hickory tree data from Stoyan and Stoyan (1998, Figure 1).

The density of a type III exponential inhomogeneous point process becomes

$$f_Y(y;\theta) \propto \alpha(\theta)^{n(y)} \mathrm{e}^{\theta \cdot t(y)} \prod_{z \subseteq y} \varphi(h_\theta^{-1}(z)),$$

compare with (10). In particular, if the homogeneous model is a Strauss model then

$$f_Y(y;\theta,\beta,\gamma) \propto \mathrm{e}^{\theta \cdot t(y)} (\alpha(\theta)\beta)^{n(y)} \gamma^{s(h_\theta^{-1}(y))}. \tag{15}$$

Note that in contrast to type I inhomogeneity, θ enters as a nuisance parameter in an exponential family density.

The simulations in Figure 1 to 4, right-hand sides, are from (15). Parameter values and other details of the simulations are given in Appendix III.

The transformation related approach yields densities of the form

$$f_Y(y;\theta) \propto \mathrm{e}^{\theta \cdot t(y)} (\alpha(\theta)\beta)^{n(y)} \prod_{\{\eta,\xi\} \subseteq y}^{\neq} u(\alpha(\theta)^{2\nu} \mathrm{e}^{\nu(\tau(\eta)+\tau(\xi)) \cdot \theta} [\eta - \xi]).$$

In particular, in the Strauss case we get

$$f_Y(y;\theta,\beta,\gamma) \propto \mathrm{e}^{\theta \cdot t(y)} (\alpha(\theta)\beta)^{n(y)} \gamma^{s_\theta(y)}, \tag{16}$$

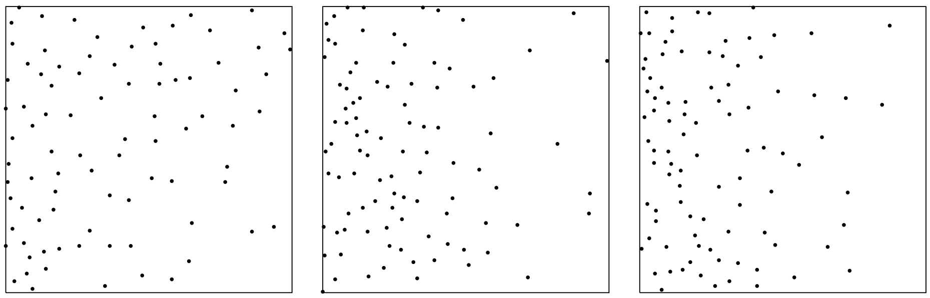
where $s_\theta(y)$ is the number of $\sim_\theta$ $-$neighbours of y. If two points η, ξ are related in the homogeneous Strauss process when $||\eta - \xi|| < R$, then the relation $\sim_\theta$ is defined by

$$\eta \sim_\theta \xi \Leftrightarrow \alpha(\theta)^{2\nu} \mathrm{e}^{\nu(\tau(\eta)+\tau(\xi)) \cdot \theta} \|\eta - \xi\| < R.$$

The three models (11), (15) and (16) are compared by simulation in Figure 5. Note that the type I process appears somewhat more homogeneous than the other processes, because the relation is for this process not location dependent. This feature becomes more pronounced if more points are forced into the point patterns by increasing β.

Furthermore, the intensity in the type I point process appears to be lower than the intensity in the other two point processes. These two are, however, similar both regarding the relation and the point intensity. Thus, the inhomogeneity function and the parameters from the associated homogeneous process play a quite different role in the actual point intensity and point interaction in the different types of inhomogeneous models. These issues have of course to be examined in more detail.

C



(a) Type I (b) Type III (c) Type III related

Figure 5: Realizations of inhomogeneous point processes on the unit square. The densities used are, from left to right, (11), (15) and (16), respectively, all with $\tau(\eta_1, \eta_2) = \eta_1$. The parameter values used are $\beta = 200$, $\gamma = 0.01$, R(interaction radius)$= 0.05$ and $\theta = -3$. In (16), the exponent is $\nu = 1/4$. The number of observed points are, from left to right, $n(y) = 87, 95$ and 93, respectively.

3.4 Parametric statistical inference

3.4.1 Likelihood inference

If a Markov point process is observed in a sampling window $T \subseteq \mathcal{X}$, cf. Figure 1 and 2, the conditional density of the point pattern observed in T, given the remaining points, may be used for inference. As in Baddeley et al. (2000), let for disjoint point patterns y and x

$$\chi(y \mid x) = \prod_{z \subseteq y \cup x : z \cap y \neq \emptyset} \varphi(z). \tag{17}$$

For a homogeneous Markov point process with density (4), the conditional density, with respect to the homogeneous Poisson point process on T with intensity 1, is then of the form, cf. e.g. Møller (1999b, formula (14)),

$$f(x_T \mid x_{T^c}) \propto \chi(x_T \mid x_{\partial T}),$$

where $x_T = x \cap T$, T^c is the complement of T and $\partial T = \{\xi \in T^c | \exists \eta \in T : \eta \sim \xi\}$. Note that this density depends on x_{T^c} only via $x_{\partial T}$.

Let us suppose that the interaction function φ can be parametrized by $\psi \in \Psi$. Then, the likelihood function based on observation of the homogeneous process in T becomes

$$L_T(\psi; x) = c_T(\psi; x_{\partial T}) \chi(x_T; \psi \mid x_{\partial T}), \tag{18}$$

where $c_T(\psi; x_{\partial T})$ is a normalizing constant and $\chi(\cdot; \psi|\cdot)$ is defined as in (17) with φ parametrized by ψ.

In the Strauss case, we get

$$L_T(\beta, \gamma; x) = c_T(\beta, \gamma; x_{\partial T})\beta^{n(x_T)}\gamma^{s(x_T)+s(x_T; x_{\partial T})}, \tag{19}$$

where for disjoint point patterns x and y

$$s(x; y) = \sum_{\eta \in x} \sum_{\xi \in y} 1[\eta \sim \xi].$$

Likelihood inference based on these likelihood functions requires Markov chain Monte Carlo (MCMC) approximations, since the normalizing constant is not known explicitly, cf. Geyer (1999) and Møller (1999b).

Let us now compare likelihood inference for the two Markovian inhomogeneous processes, viz. type I and III. Likelihood inference for the type II case, based on MCMC techniques for missing data problems, has been discussed in detail in Baddeley et al. (2000).

For type I inhomogeneity, the conditional density is given by

$$f_Y(y_T \mid y_{T^c}) \propto \prod_{\eta \in y_T} \lambda(\eta) \cdot \chi(y_T \mid y_{\partial T}),$$

where χ refers to the homogeneous process. If $\lambda_\theta(\eta) = \alpha(\theta)\mathrm{e}^{\theta \cdot \tau(\eta)}$, and φ is parametrized by ψ, we get

$$L_T(\theta, \psi; y) = \tilde{c}_T(\theta, \psi; y_{\partial T})\mathrm{e}^{\theta \cdot t(y_T)}\chi(y_T; \psi \mid y_{\partial T}).$$

In particular, if the homogeneous process is a Strauss process we have

$$L_T(\theta, \beta, \gamma; y) = \tilde{c}_T(\theta, \beta, \gamma; y_{\partial T})\mathrm{e}^{\theta \cdot t(y_T)}\beta^{n(y_T)}\gamma^{s(y_T)+s(y_T; y_{\partial T})}.$$

Again, MCMC is required for the analysis.

Statistical inference in the case of type III inhomogeneity is based on the following conditional density, derived from (6),

$$f_Y(y_T \mid y_{T^c}) \propto \prod_{\eta \in y_T} \lambda(\eta) \cdot \chi(h^{-1}(y_T) \mid h^{-1}(y_{\partial T})).$$

If h_θ is chosen such that $Jh_\theta^{-1} = \lambda_\theta$ and φ is parametrized by ψ, we get, cf. Appendix II,

$$L_T(\theta, \psi; y) = L_0(\theta; y_T)L_{h_\theta^{-1}(T)}(\psi; h_\theta^{-1}(y)), \tag{20}$$

where $L_0(\cdot; y_T)$ is the likelihood (1) of an inhomogeneous Poisson point process with intensity function λ_θ and $L_{h_\theta^{-1}(T)}(\cdot; h_\theta^{-1}(y))$ is the likelihood function (18) for the homogeneous Markov point process with observation $h_\theta^{-1}(y)$, observed in $h_\theta^{-1}(T)$. Recall that (18) reduces to (19) in the Strauss case. Note that for the transformations derived in Example 3.1 and windows T as shown in Figure 1 and 2, $h_\theta^{-1}(T)$ does not depend on θ. In fact, $h_\theta(T) = T$.

Likelihood inference is simpler for type I than for type III models since the inhomogeneity parameter is a nuisance parameter in an exponential family model in the latter case. However, simulation studies indicate that the estimate $\hat{\theta}_0$ of θ based on $L_0(\cdot; y_T)$, cf. (20), is close to $\hat{\theta}$. In that case, the interaction parameter ψ can be estimated on the basis of

$$L_{h_{\hat{\theta}_0}^{-1}(T)}(\cdot; h_{\hat{\theta}_0}^{-1}(y))$$

and the analysis will be no more complicated than the analysis of a homogeneous Markov point process.

3.4.2 Pseudolikelihood inference

A less demanding inference procedure is based on the pseudolikelihood function which is the likelihood function for a Poisson point process with intensity function equal to the Papangelou conditional intensity of the process, cf. e.g. Besag (1975) and Jensen and Møller (1991). Recently, pseudolikelihood inference has been discussed by Baddeley and Turner (2000).

If the homogeneous process is parametrized by ψ the pseudolikelihood function based on observation in T becomes

$$PL_T(\psi; x) = \exp(-\int_T [\lambda_\psi(u; x) - 1] du^k)[\prod_{\eta \in x_T} \lambda_\psi(\eta; x \backslash \eta)]$$

where $x \setminus \eta$ means $x \setminus \{\eta\}$ and

$$\lambda_\psi(\eta; x) = \frac{f_X(x \cup \eta; \psi)}{f_X(x; \psi)}, \quad \eta \notin x.$$

If the homogeneous process is a Strauss process, then

$$\lambda_{\beta,\gamma}(\eta; x) = \beta\gamma^{s(\eta;x)}, \quad \eta \notin x,$$

and

$$PL_T(\beta, \gamma; x) = \exp(-\int_T [\beta\gamma^{s(u;x)} - 1] du^k) \beta^{n(x_T)} \gamma^{2s(x_T) + s(x_T; x_{\partial T})}.$$

Compared to likelihood inference the 'normalizing' constant is much simpler.

Let us now look at pseudolikelihood inference for the two Markovian inhomogeneous processes. For type I inhomogeneity of exponential form, the pseudolikelihood function takes the form

$$PL_T(\theta,\psi;y) = \exp(-\int_T [e^{\theta\cdot\tau(u)}\lambda_\psi(u;y)-1]du^k)e^{\theta\cdot t(y_T)}[\prod_{\eta\in y_T}\lambda_\psi(\eta;y\backslash\eta)].$$

In the Strauss case, we get

$$\begin{aligned} PL_T(\theta,\beta,\gamma;y) &= \exp(-\int_T [\beta e^{\theta\cdot\tau(u)}\gamma^{s(u;y)}-1]du^k) \\ &\quad \times e^{\theta\cdot t(y_T)}\beta^{n(y_T)}\gamma^{2s(y_T)+s(y_T;y_{\partial T})}. \end{aligned}$$

Note that for fixed θ and γ the maximum pseudolikehood estimate of β is known explicitly. This is an example where the Papangelou conditional intensity is of log-linear form and the analysis suggested by Baddeley and Turner (2000) can therefore be used. Using the approximations described in Berman and Turner (1992) and Baddeley and Turner (2000) one should, however, be careful when choosing the dummy points involved in the approximation.

In the case of type III inhomogeneity, it can be shown, cf. the Appendix II,

$$PL_T(\theta,\psi;y) = L_0(\theta;y_T)PL_{h_\theta^{-1}(T)}(\psi;h_\theta^{-1}(y)). \tag{21}$$

The pseudolikelihood function thus decomposes as the likelihood function, cf. (20). Pseudolikelihood inference for type III processes appears to be more complicated but again it is expected that the inference can be split into two parts.

4 Discussion

In the present paper, we have discussed three types of inhomogeneous point processes, derived from homogeneous Markov point processes. It is of course also of interest to study how inhomogeneity can be introduced into other of the classical classes of point processes. For instance, one may consider inhomogeneous Neyman-Scott point processes (the Poisson point process of the mothers is inhomogeneous), inhomogeneous Matérn hard-core processes (the unthinned Poisson point process is inhomogeneous), inhomogeneous simple sequential inhibition point processes (the size of the circular region around each point depends on the position of the point) and inhomogeneous Gibbs processes (e.g. transformations of homogeneous Gibbs processes). See Clausen et al. (2000).

The emphasis has in the present paper been on parametrically modelled inhomogeneity. This is a new approach for type II processes. Dually, it will

C

also be of interest to study non-parametric estimation of the transformation involved in type III models.

Summary statistics like the $K-$, $F-$ and $G-$functions have been developed for the initial study of the interaction in homogeneous point processes and for checking of models for homogeneous point processes, cf. Stoyan et al. (1995). In Baddeley et al. (2000), an analogue of the $K-$function is suggested for the inhomogeneous case. For a type II process, this analogue has the nice property of being identical to the $K-$function of the unthinned process. It still remains, however, to find versions of the $F-$ and $G-$functions that can be used in the inhomogeneous case. For type III processes, an alternative is to estimate the transformation, either parametrically or non-parametrically, and then use the traditional summary statistics for the homogeneous case on the inversely transformed point pattern.

Type III processes have the special feature that the neighbourhood relation induced by the transformation is location dependent. The relation is generally not isotropic in the sense that relationship only depends on the distance between the points. Another quite promising idea is to introduce inhomogeneity in Markov point processes by location dependent scaling. This is a topic for future research.

5 Acknowledgements

The authors want to thank Ute Hahn, Jesper Møller and Aila Särkkä for fruitful discussions while preparing this paper. This research has been supported by the Centre for Mathematical Physics and Stochastics (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. *Statistica 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.

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

Berman, M. (1986). Testing for spatial association between a point process and another stochastic process. *Appl. Statist.*, 35:54–62.

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

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

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.

Clausen, W. H. O., Rasmussen, H. H., and Rasmussen, M. H. (2000). *Inhomogeneous Point Processes.* Master Thesis, Department of Mathematical Sciences, Aalborg University.

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.

Heikkinen, J. and Arjas, E. (1998). Non-parametric Bayesian estimation of spatial Poisson intensity. *Scand. J. Statist.*, 25:435–450.

Högmander, H. and Särkkä, A. (1999). Multitype spatial point patterns with hierarchical interactions. *Biometrics*, 55:1051–1058.

Jensen, E. B. V. (1998). *Local Stereology.* World Scientific, Singapore.

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

Jensen, J. L. and Møller, J. (1991). Pseudolikelihood for exponential family models of spatial point processes. *The Annals of Applied Probability*, 1:445–461.

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

Møller, J. (1999a). *Aspects of Spatial Statistics, Stochastic Geometry and Markov Chain Monte Carlo.* Doctoral Thesis, Aalborg University.

Møller, J. (1999b). 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 141–172. Chapman and Hall/CRC, London.

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

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.

$\boxed{\text{C}}$

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.

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

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

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

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.

Appendix I

Let us start by finding Jh_θ^{-1} in the case $p = 0$ where (13) reduces to

$$h_\theta(\eta) = g_\theta(\|\eta\|)\frac{\eta}{\|\eta\|}.$$

Let $B_d(O,1)$ be the unit ball in R^d. Using polar decomposition twice we get for an arbitrary function f on $B_d(O,1)$,

$$\begin{aligned}
&\int_{B_d(O,1)} f(h_\theta(\eta))d\eta^d \\
&\quad = \int_{S^{d-1}}\int_0^1 f(g_\theta(t)\omega)t^{d-1}dtd\omega^{d-1} \\
&\quad = \int_{S^{d-1}}\int_0^1 f(u\omega)(g_\theta^{-1}(u))^{d-1}(g_\theta^{-1})'(u)dud\omega^{d-1} \\
&\quad = \int_{B_d(O,1)} f(\eta)(g_\theta^{-1}(\|\eta\|))^{d-1}(g_\theta^{-1})'(\|\eta\|)\|\eta\|^{-(d-1)}d\eta^d.
\end{aligned}$$

Therefore,

$$Jh_\theta^{-1}(\eta) = (g_\theta^{-1}(\|\eta\|))^{d-1}(g_\theta^{-1})'(\|\eta\|)\|\eta\|^{-(d-1)} \text{ for } p = 0. \tag{22}$$

This result can now be used to find Jh_θ^{-1} for general p. Let

$$T_d(O,1) = \{\eta \in R^d : d_C(\eta) \leq 1\}.$$

For an arbitrary function f on $T_d(O,1)$ we then get

$$\begin{aligned}
&\int_{T_d(O,1)} f(h_\theta(\eta))d\eta^d \\
&\quad = \int_{C^\perp}\int_C f(x + g_\theta(\|y\|)\frac{y}{\|y\|})1\{\|y\| \leq 1\}dx^p dy^{d-p} \\
&\quad = \int_{C^\perp}\int_C f(x+y)1\{\|y\| \leq 1\}(g_\theta^{-1}(\|y\|))^{d-p-1}(g_\theta^{-1})'(\|y\|)\|y\|^{-(d-p-1)}dx^p dy^{d-p} \\
&\quad = \int_{T_d(O,1)} f(\eta)(g_\theta^{-1}(d_C(\eta)))^{d-p-1}(g_\theta^{-1})'(d_C(\eta))d_C(\eta)^{-(d-p-1)}d\eta^d,
\end{aligned}$$

where we at the second equality sign have used (22). It follows that

$$Jh_\theta^{-1}(\eta) = (g_\theta^{-1}(d_C(\eta)))^{d-p-1}(g_\theta^{-1})'(d_C(\eta))d_C(\eta)^{-(d-p-1)}.$$

Since we also have

$$Jh_\theta^{-1}(\eta) = \alpha(\theta)\mathrm{e}^{\theta\cdot\tilde{\tau}(d_C(\eta))},$$

g_θ must satisfy

$$\frac{1}{d-p}[(g_\theta^{-1}(u))^{d-p}]' = \alpha(\theta)u^{d-p-1}\mathrm{e}^{\theta\cdot\tilde{\tau}(u)}, u \in [0,1]. \tag{23}$$

Since g_θ is increasing 1-1 of [0,1] onto itself, the unique solution of (23) is (14).

Appendix II

In order to derive (20), we need to find the constant $c_T(\theta,\psi;y_{\partial T})$ in the expression for the conditional density

$$f_Y(y_T;\theta,\psi|y_{T^c}) = c_T(\theta,\psi;y_{\partial T})[\prod_{\eta\in y_T}\lambda_\theta(\eta)]\chi(h_\theta^{-1}(y_T);\psi|h_\theta^{-1}(y_{\partial T})).$$

Since this is a density with respect to the Poisson point process on T with intensity measure λ_d^k we get, using that $\lambda_\theta(\eta) = Jh_\theta^{-1}(\eta)$ and the well-known expansion of the distribution of the Poisson point process, cf. Møller (1999b,

$\boxed{\text{C}}$

Section 2),

$$
\begin{aligned}
c_T(\theta, \psi; y_{\partial T})^{-1} \\
&= \mathrm{e}^{-\lambda_d^k(T)} \sum_{n=0}^{\infty} \frac{1}{n!} \int_T \cdots \int_T \prod_{i=1}^{n} J h_\theta^{-1}(y_i) \\
&\qquad \times \chi(\{h_\theta^{-1}(y_1), \ldots, h_\theta^{-1}(y_n)\}; \psi | h_\theta^{-1}(y_{\partial T})) dy_1^k \cdots dy_n^k \\
&= \mathrm{e}^{-\lambda_d^k(T)} \sum_{n=0}^{\infty} \frac{1}{n!} \int_{h_\theta^{-1}(T)} \cdots \int_{h_\theta^{-1}(T)} \chi(\{x_1, \ldots, x_n\}; \psi | h_\theta^{-1}(y_{\partial T})) dx_1^k \cdots dx_n^k \\
&= \mathrm{e}^{-\lambda_d^k(T)} \mathrm{e}^{\lambda_m^k(h_\theta^{-1}(T))} c_{h_\theta^{-1}(T)}(\psi; h_\theta^{-1}(y_{\partial T}))^{-1}.
\end{aligned}
$$

The result now follows by noting that

$$
\lambda_m^k(h_\theta^{-1}(T)) = \int_{h_\theta^{-1}(T)} d\xi^k = \int_T J h_\theta^{-1}(\eta) d\eta^k = \int_T \lambda_\theta(\eta) d\eta^k.
$$

The proof of (21) is obtained as follows. The Papangelou conditional intensity of the transformed process becomes, cf. (6),

$$
\frac{f_Y(y \cup \eta; \theta, \psi)}{f_Y(y; \theta, \psi)} = \lambda_\theta(\eta) \lambda_\psi(h_\theta^{-1}(\eta); h_\theta^{-1}(y)), \eta \notin y,
$$

where λ_ψ is the Papangelou conditional intensity of the untransformed process. The pseudolikelihood function of the transformed process therefore becomes

$$
\begin{aligned}
PL_T(\theta, \psi; y) \\
&= \exp(-\int_T [\lambda_\theta(u) \lambda_\psi(h_\theta^{-1}(u); h_\theta^{-1}(y)) - 1] du^k) \\
&\quad \times \prod_{\eta \in y_T} [\lambda_\theta(\eta) \lambda_\psi(h_\theta^{-1}(\eta); h_\theta^{-1}(y \setminus \eta))] \\
&= [\prod_{\eta \in y_T} \lambda_\theta(\eta)] \exp(-\int_T [\lambda_\theta(u) - 1] du^k) \\
&\quad \times \exp(-\int_T \lambda_\theta(u) [\lambda_\psi(h_\theta^{-1}(u); h_\theta^{-1}(y)) - 1] du^k) \\
&\quad \times \prod_{\eta \in y_T} \lambda_\psi(h_\theta^{-1}(\eta); h_\theta^{-1}(y \setminus \eta)).
\end{aligned}
$$

The result is now obtained by noting that

$$
\int_T \lambda_\theta(u) [\lambda_\psi(h_\theta^{-1}(u); h_\theta^{-1}(y)) - 1] du^k = \int_{h_\theta^{-1}(T)} [\lambda_\psi(v; h_\theta^{-1}(y)) - 1] dv^k.
$$

Appendix III

Simulations from the inhomogeneous Strauss point process (15) are shown in the right hand-sides of Figure 1 to 4. In Table 1 the model parameters and the resulting number of points in the simulated point patterns are given. Note, however, that the number of points in Figure 2, $n(x) = 355$, is for a 33% wider rectangle. A larger area was used to avoid edge problems.

The point patterns shown in Figure 1 to 4, right, and the three point patterns in Figure 5 have been simulated using Metropolis-Hastings birth-death algorithm with 500000 iterations, cf. e.g. Møller (1999b).

Figure	β	γ	R	$n(y)$	θ
1	1000	0.01	0.1	163	-3
2	400	0.01	0.1	355	-3
3	70	0.01	0.1	40	-1
4	100	0.01	0.2	143	-2

Table 1: Parameters used for simulation and the resulting number of points for the point patterns in the right-hand sides of Figure 1 to 4.

Nielsen L. S. and Jensen, E. B. V. (2001).
Statistical inference for transformation inhomogeneous point processes. *Research Report 12, Laboratory for Computational Stochastics, University of Aarhus*. Submitted.

Statistical inference for transformation inhomogeneous point processes

LINDA STOUGAARD NIELSEN AND EVA B. VEDEL JENSEN
Laboratory for Computational Stochastics
University of Aarhus

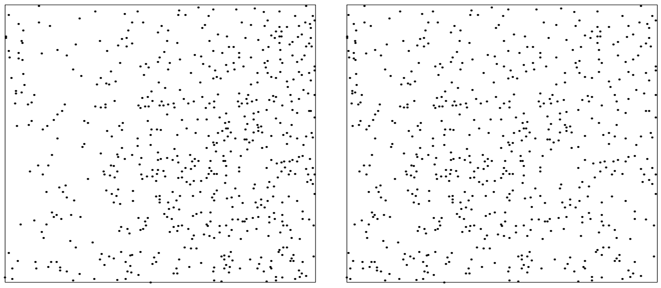
Abstract

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

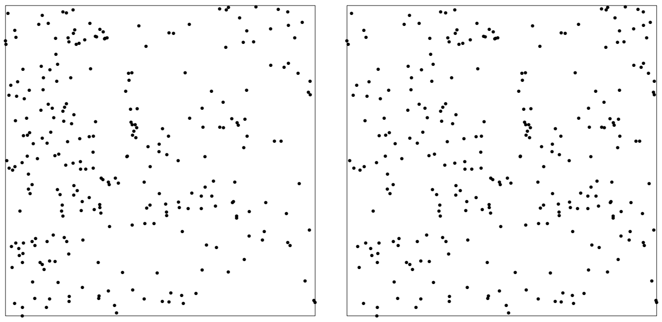
Various point process models allowing for both interaction and inhomogeneity have recently been suggested, cf. Baddeley et al. (2000), Brix and Møller (1998), Hahn et al. (2001), Jensen and Nielsen (2000, 2001) and Stoyan and Stoyan (1998). See also Ogata and Tanemura (1986). A majority of these model 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

$\boxed{\mathbf{D}}$

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

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

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*

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

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,

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

D

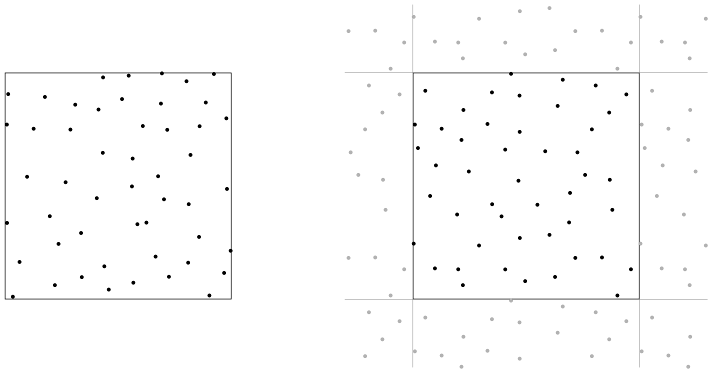


Figure 3: Realizations of the Strauss process on the unit square. To the left, the relation is the distance relation. To the right, the relation is the distance relation modified to be $\mathcal{X}$-periodic. To illustrate the periodicity, the point pattern has been translated with $\{(i,j)\}_{(i,j)\in\mathbb{Z}\times\mathbb{Z}}$ 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. $\square$

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

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

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).

$\boxed{\mathbf{D}}$

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

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

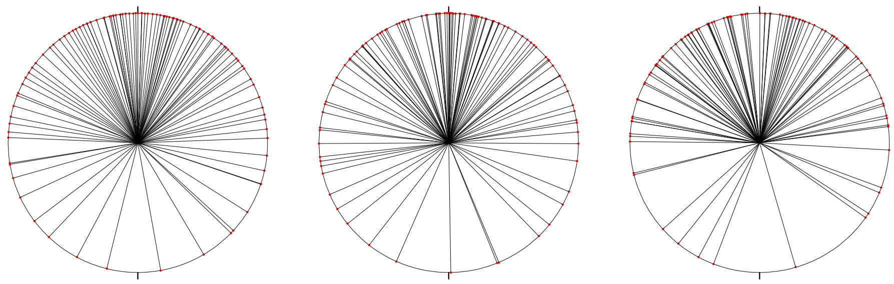
In particular, when $\tau(u) = u$ and $I = (0;a)$ we obtain the *simple exponential transformation* given by

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

$\theta \in \mathbb{R}$. 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),

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

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

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

where $L_{\text{hom}}(\psi_1, \psi_2; h_\theta^{-1}(y))$ 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, $L_{\text{hom}}(\beta, \gamma, r; h_\theta^{-1}(y))$ is of exponential family form and its maximum with respect to (β, γ) is attained as the unique solution to the likelihood equations

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

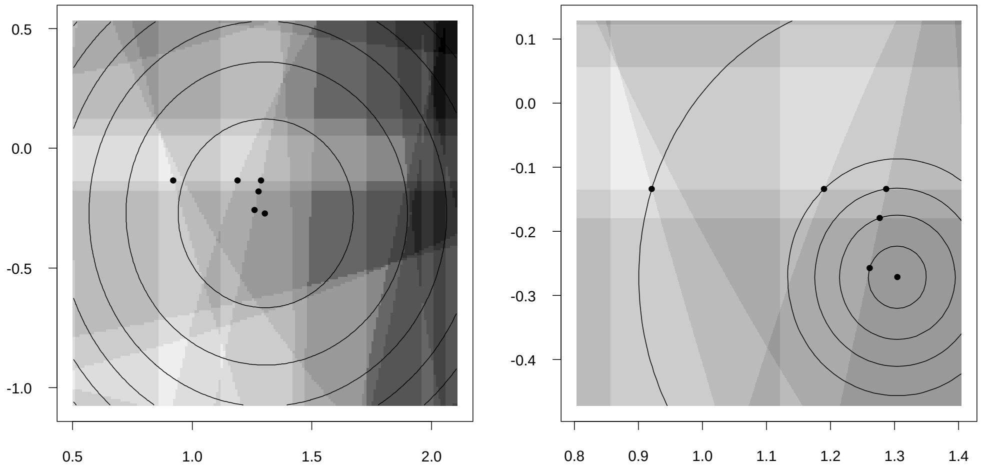
where X is a Strauss process with parameters (β, γ, r). The solution is denoted $(\hat\beta(\theta, r), \hat\gamma(\theta, r))$. The profile likelihood with nuisance parameter (θ, r) becomes

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

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

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

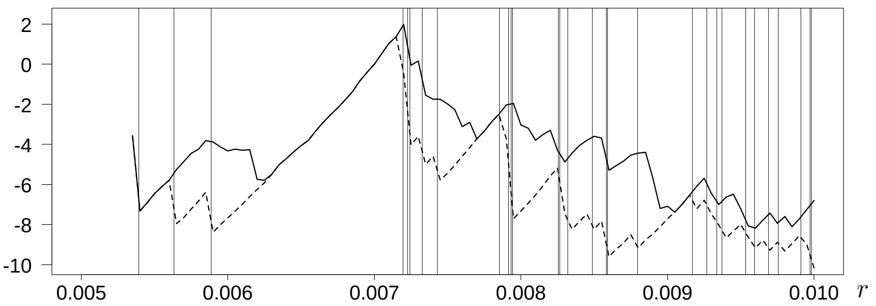


Figure 6: The profile log-likelihood function $\log \overline{L}(r; y)$ is plotted (in full) for the cell data. The stippled line is the partial profile log-likelihood when $\theta = \hat{\theta}_0$ is fixed. For details, see the text.

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.

D

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 $t(y)/n(y) \in int C$, 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,

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

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),

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

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

D

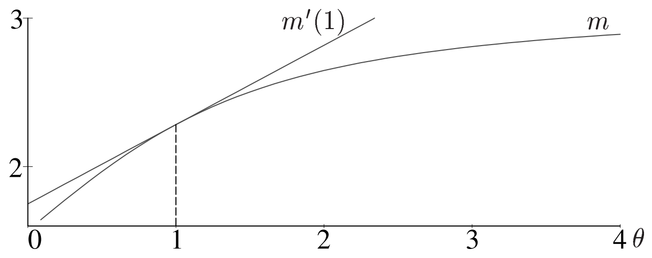


Figure 7: The function m given in (18).

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 $L_{\text{hom}}(\beta, \gamma, r; h_{\hat{\theta}_0}^{-1}(y))$ 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

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

where

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

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

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

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,

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

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

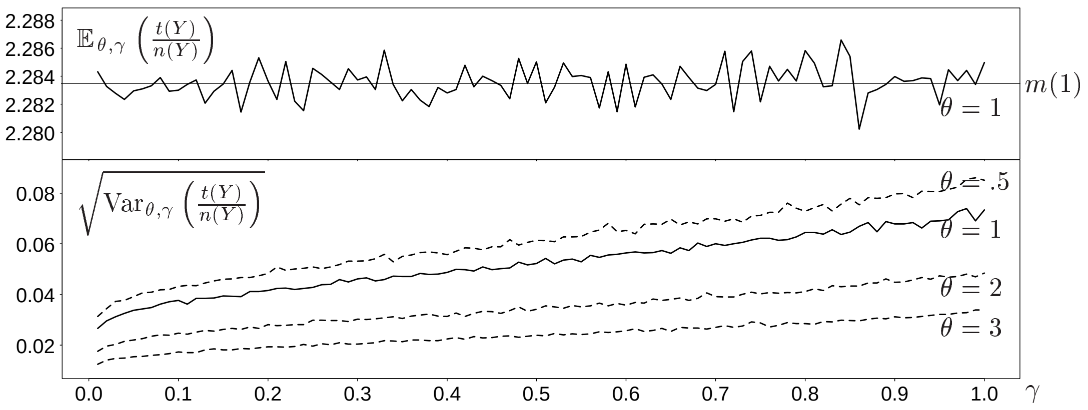


Figure 8: Simulation experiment examining the mean and standard deviation of $\frac{t(Y)}{n(Y)}$ for varying γ, and θ as indicated.

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 $t(Y)/n(Y)$ shown in 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 and the linear approximation therefore is better.

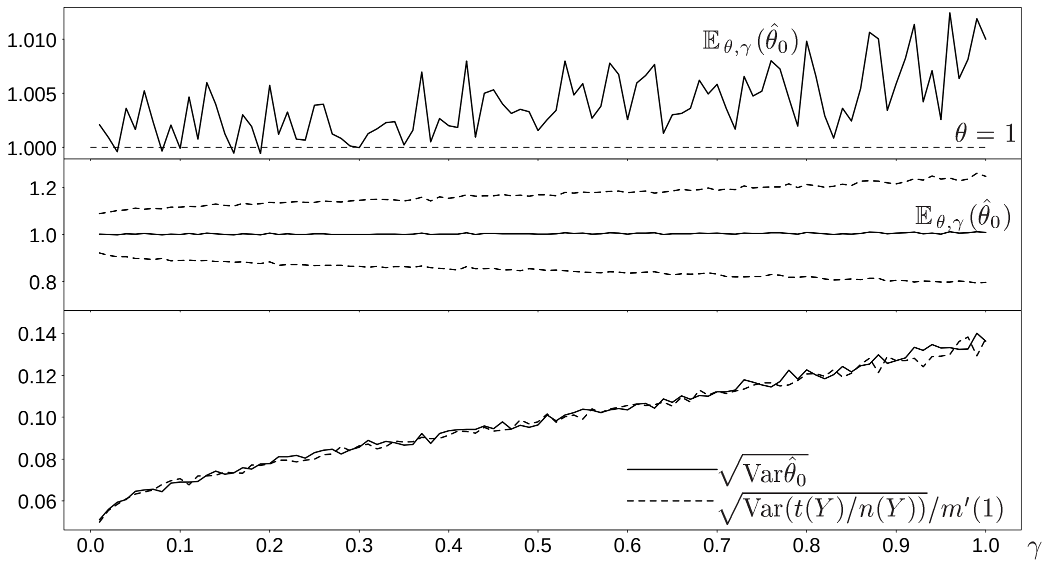


Figure 9: Simulation experiment examining the mean and standard deviation of $\hat{\theta}_0$ for varying γ, and $\theta = 1$. In the middle plot the mean is plotted together with 95% envelopes. In the lower plot the approximation (21) is examined.

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.

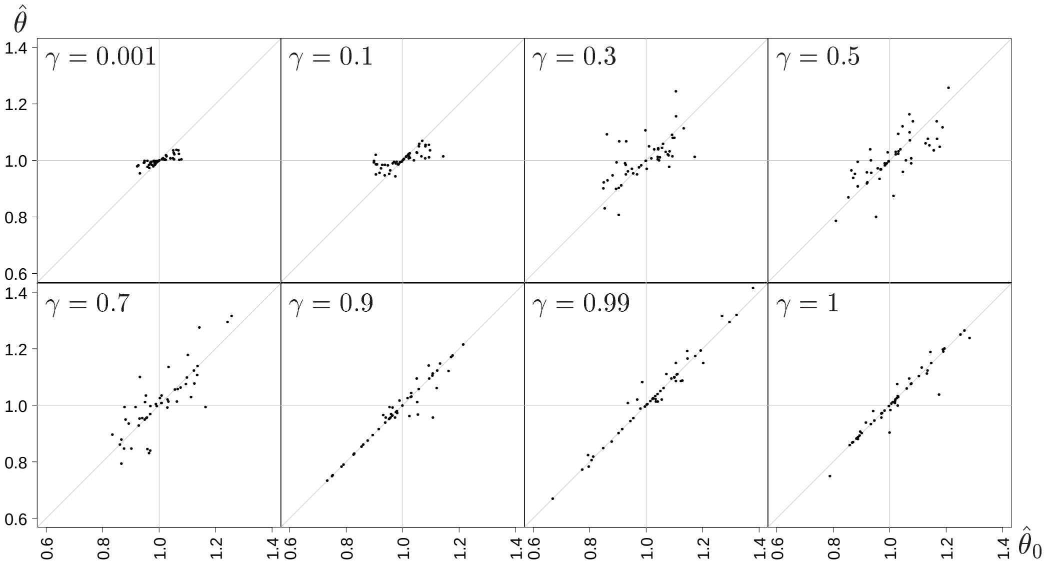


Figure 10: Comparison of $\hat{\theta}$ and $\hat{\theta}_0$. Simulation experiment with γ as indicated in each of the eight plots, and $\theta = 1$.

$\boxed{\mathbf{D}}$

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.*

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*, volume 37 of *IMS Lecture Notes*, pages 297–318.

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 141–172. 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.

E

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*.

Inhomogeneous spatial point processes by location dependent scaling

UTE HAHN[1]
EVA B. VEDEL JENSEN[1]
MARIE-COLETTE VAN LIESHOUT[2]
LINDA STOUGAARD NIELSEN[1]

Abstract

A new class of models for inhomogeneous spatial point processes is introduced. A locally scaled point process is a modification of a homogeneous point process, obtained by replacing volume measures with local analogues defined by a scaling function. If the scaling function is constant, then local scaling coincides with global scaling. The new approach is particularly appealing for modelling inhomogeneous Markov point processes. We show that locally the Papangelou conditional intensity of the new process behaves as that of a global scaling of the homogeneous process. The classes of distance-interaction and shot noise Markov point processes are discussed in detail. Approximations are suggested that simplify calculation of the density e.g. in simulation.

1. Introduction

Point patterns with non homogeneous intensity are observed quite frequently in nature and technology. For example, the number of trees per unit area in a forest depends on environmental conditions and therefore maps showing location of trees usually look inhomogeneous. In plant and animal tissue, cell size and, correspondingly, cell number often depend on the distance to the boundary of an organ. Many modern materials are designed with structural inhomogeneity, imitating natural structures in order to improve functional properties. An example is the bronze sinter filter shown in Figure 1. It consists of almost spherical bronze particles, the diameters of which decrease along an axis that marks the filtering direction. Since the particles are densely packed, the number of particles per unit volume increases as the diameters decrease. This is also observable on sections parallel to the directions of inhomogeneity: The centres of the particle section profiles form an inhomogeneous point pattern.

[1]Laboratory for Computational Stochastics, University of Aarhus
[2]Centre for Mathematics and Computer Science, Amsterdam

Hahn et al. (2001)

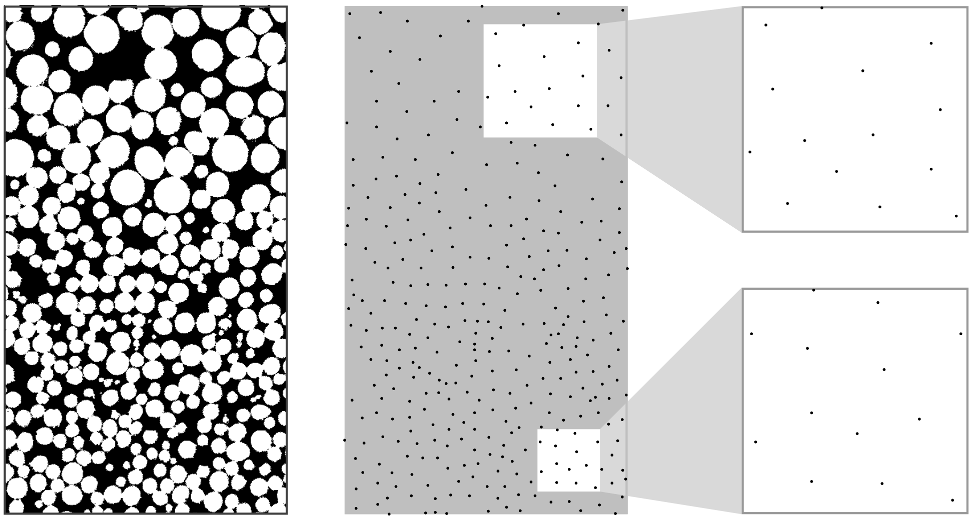


Figure 1: Left part: Section of a bronze sinter filter with a gradient in particle size and number. Right part: Centres of the particle profiles. Two enlargements from top and bottom part, containing about the same number of points, show similar geometry.

While it is easy to model inhomogeneous point patterns with independently positioned points by inhomogeneous Poisson point processes, situations as shown in Figure 1 require more sophisticated approaches. This pattern is characterized by repulsive interaction between the points due to the fact that it results from a packing of spheres. The packing is of similar volume fraction, and similar geometry in regions with larger and with smaller sphere diameters. Therefore, regions with large sphere diameters look like scaled versions of regions with small diameters and vice versa. A similar effect can often be observed in nature, e.g. in plant communities where number density is governed by environmental conditions. For example desert plants tend to form regular patterns with varying scale, such that distances between plants are smaller in densely covered regions. Such point patterns shall be called *homogeneous up to a local scale factor*.

In recent years, various models have been suggested for inhomogeneous point processes with interaction. Since Markov point processes are very popular for modelling interaction in homogeneous point patterns it is natural that they are used as starting points for inhomogeneous models. The survey by Jensen and Nielsen (2001) discusses three ways of introducing inhomogeneity into a Markov model. As explained in Section 2, homogeneous Markov point processes are defined by a density with respect to the unit rate Poisson process. A straightforward idea is therefore to define an inhomogeneous process by the same density (up to a constant factor) but with respect to an inhomogeneous Poisson process (Stoyan and Stoyan, 1998; Ogata and Tanemura,

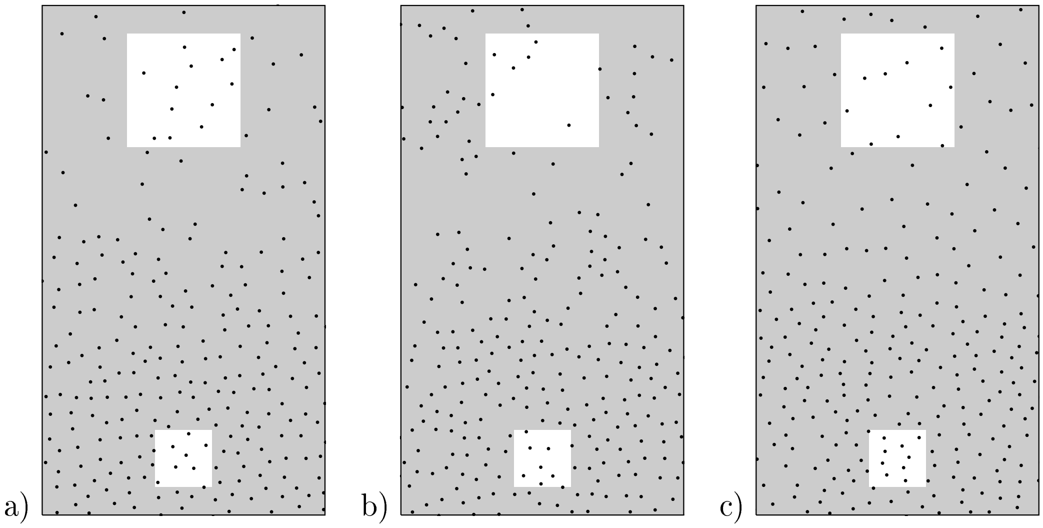


Figure 2: Inhomogeneous hard core point patterns obtained by a) defining the density as with respect to an inhomogeneous Poisson point process, b) inhomogeneous independent thinning, c) transformation of coordinates. Note that dense and sparse regions differ in geometry.

1986). Inhomogeneity can also be obtained by location dependent thinning (Baddeley et al., 2000), or by transformation of a homogeneous Markov point process (Jensen and Nielsen, 2000).

In all three cases, local geometry of the point pattern changes with intensity. This is illustrated in Figure 2, which shows realizations of inhomogeneous hard core point processes obtained by the three methods. In order to obtain patterns that are homogeneous up to a scale factor, interaction has to be adapted to intensity. However this is not the case in the first approach where interaction does not depend on the location, cf. Figure 2 (a). Thinning on the other hand destroys interaction structure in regions with small number of points. This leads to a Poisson like appearance of sparse regions, see Figure 2 (b). Transformation of coordinates finally not only introduces inhomogeneity but also local anisotropy, as shown in Figure 2 (c). Therefore these three approaches are not suitable for modelling situations as given in Figure 1.

In the present paper we propose a model for point processes that are homogeneous up to a local scale factor. As in the previous three approaches, the inhomogeneous model is obtained by modifying a homogeneous "template" process that yields the interaction. The idea is that inhomogeneity is obtained by scaling the template process with a location dependent scaling factor. A large scaling factor hereby results in low intensity and large interaction distances, whereas a small scaling factor yields high intensity and small

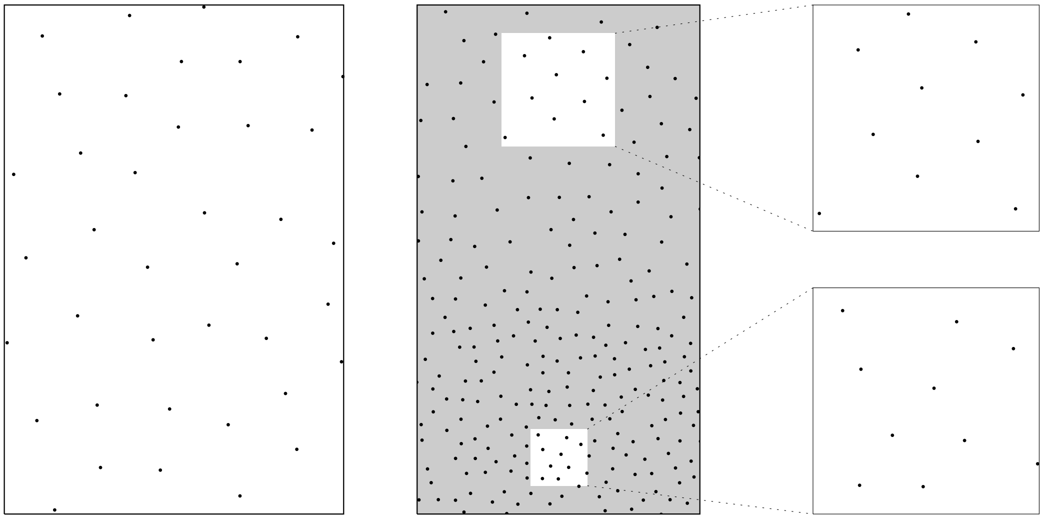


Figure 3: Homogeneous template hard core process (left part) and inhomogeneous process obtained by local scaling. Enlargements from dense and sparse regions of the inhomogeneous pattern look similar to the template pattern.

interaction range. In regions with constant scaling factor the point process should locally behave like a scaled version of the template, see Figure 3.

The method and results presented in this paper are applicable to practically any template process that is given by a density with respect to a homogeneous Poisson point process; however the main emphasis will be on Markov point processes. The definition of Markov point processes and other prerequisites are recalled in Section 2.

The local scaling model is introduced in Section 3. Calculating the density function of a point process for a given point pattern usually implies evaluating distances, areas, etc. The proposed local scaling model changes the way such quantities are measured according to a location dependent scaling function.

Sections 4 and 5 give a closer look on the important classes of distance-interaction and shot noise processes. In particular we show that locally scaled Markov point processes are again Markov. Some useful approximations of local scaling, which simplify the calculations (e.g. in simulation) are presented in Section 6.

The paper concludes with a critical discussion.

2. Prerequisites

Let $\mathcal{K}$ denote the set of all full-dimensional bounded subsets of $\mathbb{R}^k$. We consider finite point processes X on sets $\mathcal{X} \in \mathcal{K}$. A point process X on $\mathcal{X}$ is a random variable taking values in $\Omega_{\mathcal{X}}$, the set of all finite subsets

$\mathbf{x} = \{x_1, \cdots, x_n\}$ of $\mathcal{X}$.

We will concentrate on point processes X that have a density f_X with respect to the restriction of the unit rate Poisson point process Π to $\mathcal{X}$. Following Nielsen and Jensen (2001, Definition 2.1), a point process X on $\mathcal{X}$ is called *homogeneous* if f_X is the restriction to $\mathcal{X}$ of a translation invariant function defined on all finite subsets of $\mathbb{R}^k$.

Markov point processes in the sense of Ripley and Kelly (1977) are particularly useful for modelling point patterns with interaction. They are defined with respect to a symmetric and reflexive relation $\sim$ on $\mathcal{X}$. Two points $x_1, x_2 \in \mathcal{X}$ are said to be neighbours if $x_1 \sim x_2$, and a finite subset $\mathbf{x} \subset \mathcal{X}$ is called a clique if all points in $\mathbf{x}$ are neighbours. Note that sets consisting of 0 or 1 points are cliques.

Following the Hammersley-Clifford theorem, a Ripley-Kelly Markov point process X has a density f_X with respect to the unit rate Poisson point process of the form

$$f_X(\mathbf{x}) = \prod_{\mathbf{y} \subseteq \mathbf{x}} \varphi(\mathbf{y}), \quad \mathbf{x} \in \Omega_{\mathcal{X}}, \tag{1}$$

where φ is an interaction function, i.e. $\varphi(\mathbf{y}) = 1$ when the set $\mathbf{y}$ is not a clique. We will assume that the interaction function φ is defined on *all* finite subsets of $\mathbb{R}^k$. A Markov point process X is thereby homogeneous if φ is translation invariant.

For a Markov point process, the Papangelou conditional intensity

$$\lambda(x \mid \mathbf{x}) = \begin{cases} \frac{f_X(\mathbf{x} \cup \{x\})}{f_X(\mathbf{x})}, & f_X(\mathbf{x}) > 0, \\ 0 & \text{otherwise}, \end{cases} \qquad x \notin \mathbf{x},$$

depends only on those points in $\mathbf{x}$ which are neighbours of x. If we let $\mathrm{d}x$ be an infinitesimal region around x and $\nu^k(\mathrm{d}x)$ the k-dimensional volume (Lebesgue measure) of $\mathrm{d}x$, then $\lambda(x \mid \mathbf{x})\nu^k(\mathrm{d}x)$ can be interpreted as the conditional probability of finding a point from the process in $\mathrm{d}x$ given the configuration elsewhere is $\mathbf{x}$, cf. e.g. van Lieshout (2000).

Before defining local scaling of point processes let us consider global scaling as a transformation of coordinates. Global scaling with a constant factor $c > 0$ maps a point process X on $\mathcal{X}$ to a process $X_c = cX$ on

$$\mathcal{X}_c = c\mathcal{X} = \{x : x/c \in \mathcal{X}\}.$$

The unit rate Poisson point process Π on $\mathcal{X}$ with intensity measure ν^k is transformed into a Poisson point process Π_c on $\mathcal{X}_c$ with intensity measure $c^{-k}\nu^k$.

Let f_X be the density of the original process with respect to Π. Then the scaled process X_c has density

$$f^{(c)}_{X_c}(\mathbf{x}) = f_X(\mathbf{x}/c) \tag{2}$$

with respect to Π_c, cf. e.g. Hoffmann-Jørgensen (1994, p. 171). (The superscript (c) in $f^{(c)}_{X_c}$ is used to indicate that the density is with respect to Π_c.) The conditional intensity associated with $f^{(c)}_{X_c}$ is

$$\lambda^{(c)}_c(x \mid \mathbf{x}) = \lambda\Big(\frac{x}{c} \,\Big|\, \frac{\mathbf{x}}{c}\Big). \tag{3}$$

The density of X_c with respect to Π is

$$f_{X_c}(\mathbf{x}) = e^{(1-c^{-k})\nu^k(\mathcal{X}_c)} c^{-kn(\mathbf{x})} f^{(c)}_{X_c}(\mathbf{x}),$$

where $n(\mathbf{x})$ is the number of points in $\mathbf{x}$.

3. Local scaling of homogeneous point processes

In this section, we give a general definition of a locally scaled version of a homogeneous template process.

The concept of scale invariance plays a crucial role in the definition. This concept relates to global scaling with a constant factor $c > 0$. Note that under scaling with factor c, a measure μ on $(\mathbb{R}^k, \mathcal{B}_k)$ is transformed into μ_c where $\mu_c(A) = \mu(c^{-1}A), A \in \mathcal{B}_k$.

Definition 1 *Let $g(\mathbf{x}; \mu^*)$ be a real-valued measurable function defined on $\Omega_{\mathbb{R}^k}$, depending on a set $\mu^* = (\mu^1, \cdots, \mu^m)$ of measures on $(\mathbb{R}^k, \mathcal{B}_k)$. The function g is called* scale invariant *if for all $\mathbf{x} \in \Omega_{\mathbb{R}^k}$ and all $c > 0$*

$$g(c\mathbf{x}; \mu^*_c) = g(\mathbf{x}; \mu^*), \tag{4}$$

where $\mu^*_c = (\mu^1_c, \cdots, \mu^m_c)$.

As will be apparent from the sections to follow, most homogeneous point processes defined in the literature have a density which is the restriction to $\Omega_{\mathcal{X}}$ of a scale-invariant function $g(\cdot, \nu^*)$, where $\nu^* = (\nu^0, \cdots, \nu^k)$ and ν^d is d-dimensional volume (Hausdorff) measure in $\mathbb{R}^k$, $d = 0, 1, \cdots, k$. Recall that ν^0 is the counting measure, and ν^1 the length measure in $\mathbb{R}^k$. Note also that $\nu^d_c(A) = \nu^d(c^{-1}A) = c^{-d}\nu^d(A)$, $A \in \mathcal{B}_k$.

Under local scaling, the constant scaling factor c is replaced by a non constant location dependent scaling function $c : \mathbb{R}^k \to \mathbb{R}_+$. The globally scaled measures ν^d_c can easily be extended to this case.

$\boxed{\mathbf{E}}$

Definition 2 *Let c be a positive and Borel measurable function on $\mathbb{R}^k$. Then the* (locally) scaled d-dimensional volume measure ν_c^d *is defined by*

$$\nu_c^d(A) = \int_A c(u)^{-d} \nu^d(du), \tag{5}$$

for all $A \in \mathcal{B}_k$.

In the following, we will assume that c is bounded from below and above, i.e. there exist $0 < \underline{c} \leq \overline{c} < \infty$ such that

$$\underline{c} \leq c(u) \leq \overline{c}, \;\; u \in \mathbb{R}^k.$$

This assumption implies in particular that $\nu_c^d(A) < \infty$ whenever $\nu^d(A) < \infty$.

We can now present the definition of locally scaled point processes.

Definition 3 *Let X be a homogeneous point process on $\mathcal{X}$, with density f_X with respect to Π of the form*

$$f_X(\mathbf{x}) \propto g(\mathbf{x}; \nu^*), \;\; \mathbf{x} \in \Omega_{\mathcal{X}},$$

where g is scale invariant. Let Π_c be the Poisson point process with the locally scaled volume measure ν_c^k as intensity measure. Let $\mathcal{Y}$ be arbitrary and suppose that $g(\cdot; \nu_c^*)$ is integrable on $\Omega_{\mathcal{Y}}$ with respect to Π_c. A locally scaled point process *X_c on $\mathcal{Y}$ with template X is defined by the following density with respect to Π_c,*

$$f_{X_c}^{(c)}(\mathbf{x}) \propto g(\mathbf{x}; \nu_c^*), \quad \mathbf{x} \in \Omega_{\mathcal{Y}}, \tag{6}$$

where ν_c^* is the set of locally scaled volume measures.

If $c : \mathbb{R}^k \to \mathbb{R}_+$ is constant, $c(u) \equiv c$, say, then the density with respect to Π_c of the scaled process on $\mathcal{Y} = c\mathcal{X}$ becomes

$$f_{X_c}^{(c)}(\mathbf{x}) \propto g(\mathbf{x}; \nu_c^*) = g(c^{-1}\mathbf{x}; \nu^*), \quad \mathbf{x} \in \Omega_{\mathcal{Y}}.$$

Local scaling with a constant scaling function is thereby equivalent to global scaling. In the general case where c is non constant, local scaling does not necessarily correspond to a mapping. Therefore there is no natural choice of $\mathcal{Y}$ which is related to $\mathcal{X}$, and the set $\mathcal{Y}$ can be arbitrary. In particular, we may choose $\mathcal{Y} = \mathcal{X}$. Note that the density of the locally scaled process X_c with respect to the unit rate Poisson point process Π is,

$$f_{X_c}(\mathbf{x}) \propto \prod_{x \in \mathbf{x}} c(x)^{-k} \cdot f_{X_c}^{(c)}(\mathbf{x}). \tag{7}$$

In Sections 4 and 5 below we discuss local scaling of two general Markov model classes, distance-interaction processes and shot noise processes. Conditions on the scaling function which ensure integrability of $g(\cdot, \nu_c^*)$ will be given. The locally scaled processes are Markov with respect to a suitably chosen relation $\sim_c$, see also the Appendix. Finally, it will be shown that the Papangelou conditional intensity of the locally scaled process,

$$\lambda_c^{(c)}(x \mid \mathbf{x}) = \begin{cases} \dfrac{f_{X_c}^{(c)}(\mathbf{x} \cup \{x\})}{f_{X_c}^{(c)}(\mathbf{x})}, & f_{X_c}^{(c)}(\mathbf{x}) > 0, \\ 0 & \text{otherwise}, \end{cases} \tag{8}$$

satisfies a local analogue of (3),

$$\lambda_c^{(c)}(x \mid \mathbf{x}) = \lambda\Big(\frac{x}{c(x)} \,\Big|\, \frac{\mathbf{x}}{c(x)}\Big), \tag{9}$$

if c is constant in a $\sim_c$-neighbourhood of x. The locally scaled processes thereby behave locally like a scaled version of the template process.

4. Distance-interaction processes

Consider the important class of *distance-interaction processes* X with densities of the form

$$f_X(\mathbf{x}) = \prod_{\mathbf{y} \subseteq \mathbf{x}} \varphi\big(D(\mathbf{y})\big), \tag{10}$$

where $D(\mathbf{y}) = \mathbf{y}$ if $n(\mathbf{y}) < 2$, and for $n(\mathbf{y}) \geq 2$

$$D(\mathbf{y}) = \{\nu^1([u,v]) : \{u,v\} \subseteq \mathbf{y},\ u \neq v\}$$

denotes the set of all pairwise distances of points in $\mathbf{y}$. Here $[u,v]$ is the line segment between the two points and ν^1 is the usual distance measure. Notice also that ν^0 is the counting measure, $\nu^0(\mathbf{y}) = n(\mathbf{y}) = \nu_c^0(\mathbf{y})$. Examples of distance-interaction processes are pairwise interaction processes and the triplets process (Geyer, 1999).

We will assume that $\varphi(\{x\}) = \beta$ and that $\varphi\big(D(\mathbf{y})\big) = 1$ for $n(\mathbf{y}) \geq 2$ unless $\nu^1([u,v]) \leq r$ for all $\{u,v\} \subseteq \mathbf{y}$. The process X is thereby homogeneous and Markov with respect to the relation

$$u \sim v \Longleftrightarrow \nu^1([u,v]) \leq r.$$

According to (10),

$$f_X(\mathbf{x}) \propto g(\mathbf{x}; \nu^*), \quad \mathbf{x} \in \Omega_{\mathcal{X}},$$

where

$$g(\mathbf{x};\nu^*) = \beta^{\nu^0(\mathbf{x})} \prod_{\mathbf{y}\subseteq_2 \mathbf{x}} \varphi\big(\{\nu^1([u,v]) : \{u,v\} \subseteq \mathbf{y}, u \neq v\}\big)$$

and $\mathbf{y} \subseteq_2 \mathbf{x}$ is short for $\{\mathbf{y} \subseteq \mathbf{x} : \nu^0(\mathbf{y}) \geq 2\}$. The function g is clearly scale invariant since for any constant $c > 0$

$$\begin{aligned}
&\prod_{\mathbf{y}\subseteq_2 c\mathbf{x}} \varphi\big(\{\nu_c^1([u,v]) : \{u,v\} \subseteq \mathbf{y}\}\big) \\
&= \prod_{c^{-1}\mathbf{y}\subseteq_2 \mathbf{x}} \varphi\big(\{\nu^1([c^{-1}u, c^{-1}v]) : \{u,v\} \subseteq \mathbf{y}\}\big) \\
&= \prod_{\mathbf{y}\subseteq_2 \mathbf{x}} \varphi\big(\{\nu^1([u,v]) : \{u,v\} \subseteq \mathbf{y}\}\big).
\end{aligned}$$

If $\varphi\big(D(\mathbf{y})\big) \leq 1$ for $n(\mathbf{y}) \geq 2$, then X is repulsive since each clique $\mathbf{y} \subseteq \mathbf{x}$ contributes a penalty $\varphi\big(D(\mathbf{y})\big)$ to the density. In this case,

$$g(\mathbf{x};\nu_c^*) \leq \beta^{n(\mathbf{x})},$$

$c : \mathbb{R}^k \to \mathbb{R}_+$. Hence, $g(\cdot;\nu_c^*)$ is integrable on $\Omega_{\mathcal{Y}}$ for any $\mathcal{Y} \in \mathcal{K}$, and locally scaled versions of such processes do exist. Otherwise, integrability has to be proofed case by case and may require certain restrictions on the scaling function c. The locally scaled process has a density of the form

$$f_{X_c}^{(c)}(\mathbf{x}) \propto \beta^{n(\mathbf{x})} \prod_{\mathbf{y}\subseteq_2 \mathbf{x}} \varphi\big(D_c(\mathbf{y})\big), \quad \mathbf{x} \in \Omega_{\mathcal{Y}},$$

where

$$D_c(\mathbf{y}) = \{\nu_c^1([u,v]) : \{u,v\} \subseteq \mathbf{y} \;\; u \neq v\}.$$

Example 1 (Strauss process). A Strauss process X on $\mathcal{X} \subseteq \mathbb{R}^k$ with intensity parameter $\beta > 0$, interaction parameter $\gamma \in [0,1]$ and interaction distance r has density

$$f_X(\mathbf{x}) \propto \beta^{n(\mathbf{x})}\gamma^{s(\mathbf{x})}, \quad s(\mathbf{x}) = \sum_{\{u,v\}\subseteq \mathbf{x}}^{\neq} \mathbf{1}\left(\nu^1([u,v]) \leq r\right), \quad \mathbf{x} \in \Omega_{\mathcal{X}},$$

where $s(\mathbf{x})$ is the number of r-close pairs in $\mathbf{x}$ (Strauss, 1975). (The superscript $\neq$ in the summation indicates that u and v are different.) For $\gamma = 0$ we obtain the hard core process, for $\gamma = 1$ the Poisson process with intensity β. The locally scaled Strauss process has density

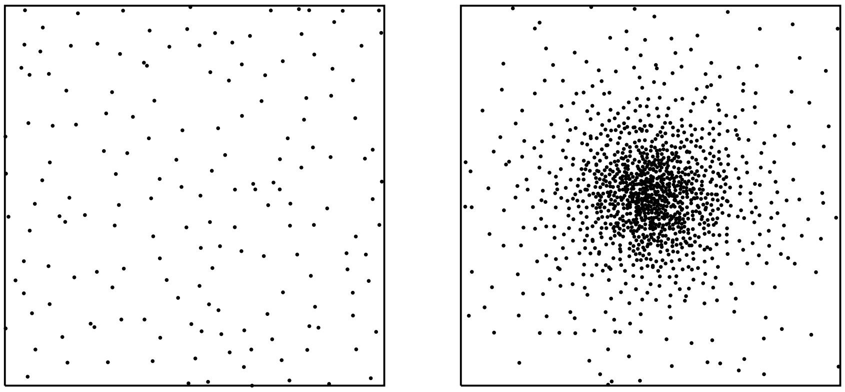


Figure 4: Left: Homogeneous template Strauss process X on $\mathcal{X} = [-1,1]^2$ with parameters $\beta = 200, \gamma = 0.1, R = 0.1$. Right: Inhomogeneous Strauss process X_c on $\mathcal{Y} = [-1,1]^2$ obtained by local scaling of X with $c(u) = 0.1 + \|u\|^2$.

$$f_{X_c}^{(c)}(\mathbf{x}) \propto \beta^{n(\mathbf{x})}\gamma^{s_c(\mathbf{x})}, \quad s_c(\mathbf{x}) = \sum_{\{u,v\}\subseteq \mathbf{x}}^{\neq} \mathbf{1}\left(\nu_c^1([u,v]) \leq r\right), \quad \mathbf{x} \in \Omega_{\mathcal{Y}}.$$

Figure 4 shows a realization; for details see the figure caption. $\square$

The locally scaled process is Markov with respect to the relation

$$u \sim v \Longleftrightarrow \nu_c^1([u,v]) \leq r.$$

While the neighbourhood $\partial(x)$ of a point in the homogeneous and isotropic template X is naturally ball shaped,

$$\partial(x) = \{y : \nu^1([x,y]) \leq r\} = b(x,r)$$

the shape of the neighbourhood

$$\partial_c(x) = \{y : \nu_c^1([x,y]) \leq r\} = b_c(x,r)$$

in the locally scaled process depends on the scaling function c. It is not necessarily convex but always star-shaped, cf. Figure 5. A neighbourhood $\partial_c(u)$ is called star-shaped if it is star-shaped with respect to u, which means that $v \in \partial_c(u) \Longrightarrow [v,u] \subseteq \partial_c(u)$.

In regions where c is constant, the neighbourhood $\partial_c(x)$ is similar to the neighbourhood in the template, cf. Figure 5. More precisely, we have

Proposition 1 *If $c(u) = c(x)$ for all $u \in b(x, c(x)r)$, then*

$$b_c(x,r) = b(x, c(x)r).$$

$\boxed{\mathbf{E}}$

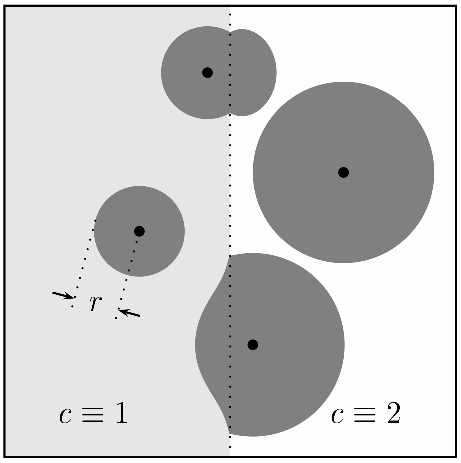


Figure 5: Neighbourhoods of four points in a locally scaled distance-interaction process, with scale factor $c \equiv 1$ in the left half and $c \equiv 2$ in the right half of the domain.

Proof: It is straightforward to show that $v \in b(x, c(x)r) \Rightarrow v \in b_c(x, r)$ and $v \notin b(x, c(x)r) \Rightarrow v \notin b_c(x, r)$. $\square$

Locally scaled distance-interaction processes have the desired property that in regions where c is constant the process behaves like a scaled version of the template process.

Proposition 2 *Suppose $c(u) = c(x)$ for all $u \in b(x, c(x)r)$. Then*

$$\lambda_c^{(c)}(x \mid \mathbf{x}) = \lambda\left(\frac{x}{c(x)} \;\middle|\; \frac{\mathbf{x}}{c(x)}\right).$$

Proof: First notice that the assumption implies that $b(x, c(x)r) = b_c(x, r)$.

The conditional intensity of the locally scaled process is of the form

$$\lambda_c^{(c)}(x \mid \mathbf{x}) = \beta \prod_{\mathbf{y} \subseteq_1 \mathbf{x}} \varphi\left(D_c(\mathbf{y} \cup x)\right),$$

where $\mathbf{y} \subseteq_1 \mathbf{x}$ is short for $\{\mathbf{y} \subseteq \mathbf{x} : n(\mathbf{y}) \geq 1\}$. Let $\mathbf{y} \subseteq_1 \mathbf{x}$, and suppose first that $\mathbf{y} \subseteq b(x, c(x)r)$. Then

$$\nu_c^1([u, v]) = c(x)^{-1}\nu^1([u, v]) = \nu^1\left(\left[\frac{u}{c(x)}, \frac{v}{c(x)}\right]\right),$$

for any $u, v \in \mathbf{y} \cup x$, and therefore

$$D_c(\mathbf{y} \cup x) = D\left(\frac{\mathbf{y}}{c(x)} \cup \frac{x}{c(x)}\right).$$

On the other hand, suppose that there exists $u \in \mathbf{y}$ such that $u \notin b(x, c(x)r)$. Thus, $\nu_c^1([u,x]) > r$, and therefore

$$\varphi\left(D_c(\mathbf{y} \cup x)\right) = 1 = \varphi\left(D\left(\frac{\mathbf{y}}{c(x)} \cup \frac{x}{c(x)}\right)\right).$$

It follows that

$$\lambda_c^{(c)}(x \mid \mathbf{x}) = \beta \prod_{\mathbf{y} \subseteq_1 \mathbf{x}} \varphi\left(D\left(\frac{\mathbf{y}}{c(x)} \cup \frac{x}{c(x)}\right)\right) = \lambda\left(\frac{x}{c(x)} \,\Big|\, \frac{\mathbf{x}}{c(x)}\right).$$

$\square$

5. Shot noise processes

Shot noise processes, such as the area-interaction process, are based on geometric quantities other than pairwise distances. Write

$$C_{\mathbf{x}}(u) = \sum_{x \in \mathbf{x}} \mathbf{1}\left(u \in b(x,r)\right)$$

for the template coverage function, then a shot noise weighted point process with potential function $p(\cdot)$ is defined by

$$f_X(\mathbf{x}) \propto \beta^{n(\mathbf{x})} \gamma^{-\int p(C_{\mathbf{x}}(u))du}, \quad \mathbf{x} \in \Omega_{\mathcal{X}}, \tag{11}$$

where $\gamma > 0$ and p is a measurable function on the non-negative integers $\mathbb{N}_0$ with $p(0) = 0$. Area interaction corresponds to $p(n) = \mathbf{1}\,(n \geq 1)$.

The interaction functions of a shot noise process are

$$\varphi(\mathbf{y}) = \beta\gamma^{m(\mathbf{y})}, \quad n(\mathbf{y}) = 1,$$
$$\varphi(\mathbf{y}) = \gamma^{m(\mathbf{y})}, \quad n(\mathbf{y}) > 1,$$

where

$$m(\mathbf{y}) = \nu^k\left(\bigcap_{y \in \mathbf{y}} b(y,r)\right) \sum_{l=1}^{n(\mathbf{y})} \binom{n(\mathbf{y})}{l} (-1)^{n(\mathbf{y})-l} p(l).$$

Compare with Theorem 3.3 in van Lieshout and Molchanov (1998). The homogeneous shot noise processes are Markov with respect to the overlapping relation

$$u \sim v \iff b(u,r) \cap b(v,r) \neq \emptyset \quad \iff \quad \|u - v\| \leq 2r.$$

$\boxed{\mathbf{E}}$

Thus, the neighbourhood of a point is a ball with radius $R = 2r$.

It is straightforward to show that

$$g(\mathbf{x}; \nu^*) = \beta^{\nu^0(\mathbf{x})} \gamma^{-\int p\left(\sum_{x \in \mathbf{x}} \mathbf{1}\left(\nu^1([x,u]) \le r\right)\right) \nu^k(du)}$$

is scale invariant. The locally scaled shot noise process has density

$$f_{X_c}^{(c)}(\mathbf{x}) \propto g(\mathbf{x}, \nu_c^*) = \beta^{n(\mathbf{x})} \gamma^{-\int p(C_{c,\mathbf{x}}(u)) \nu_c^k(du)}, \quad \mathbf{x} \in \Omega_{\mathcal{Y}}, \tag{12}$$

with scaled coverage function

$$C_{c,\mathbf{x}}(u) = \sum_{x \in \mathbf{x}} \mathbf{1}\left(u \in b_c(x, r)\right).$$

Van Lieshout and Molchanov (1998) show that (11) is integrable if there exits some constant $0 < C < \infty$ with

$$|p(n)| \le Cn \quad \forall\, n \in \mathbb{N}_0. \tag{13}$$

Proposition 3 *Under condition* (13), *$g(\cdot; \nu_c^*)$ is integrable on $\Omega_{\mathcal{Y}}$ for any $\mathcal{Y} \in \mathcal{K}$ and hence the locally scaled process defined by* (12) *does exist.*

Proof: We show that there exists $0 < M < \infty$ such that

$$g(\mathbf{x}; \nu_c^*) \le M^{n(\mathbf{x})}, \quad \mathbf{x} \in \Omega_{\mathcal{Y}}.$$

This is fulfilled if

$$\left| \int p(C_{c,\mathbf{x}}(u)) \nu_c^k(du) \right| \le M' n(\mathbf{x}) \tag{14}$$

for some $0 < M' < \infty$ and all $\mathbf{x} \in \Omega_{\mathcal{Y}}$.

Let $S_C(\mathbf{x})$ denote the support of $C_{c,\mathbf{x}}$. As $C_{c,\mathbf{x}}(u) \le n(\mathbf{x})$, we have

$$\left| \int p(C_{c,\mathbf{x}}(u)) \nu_c^k(du) \right| \le C n(\mathbf{x}) \nu_c^k\left(S_C(\mathbf{x})\right)$$

with C as in (13). Since

$$S_C(\mathbf{x}) = \bigcup_{x \in \mathbf{x}} b_c(x, r) \subseteq \mathcal{Y} \oplus b(0, \bar{c}r),$$

where $\oplus$ denotes Minkowski addition, (14) holds with

$$M' = C\, \underline{c}^{-k} \nu^k\left(\mathcal{Y} \oplus b(0, \bar{c}r)\right).$$

$\square$

The locally scaled shot noise process has interaction functions

$$\varphi_c(\mathbf{y}) = \beta\gamma^{m_c(\mathbf{y})}, \quad n(\mathbf{y}) = 1,$$
$$\varphi_c(\mathbf{y}) = \gamma^{m_c(\mathbf{y})}, \quad n(\mathbf{y}) > 1,$$

where

$$m_c(\mathbf{y}) = \nu_c^k\left(\bigcap_{y\in\mathbf{y}} b_c(y,r)\right)\sum_{l=1}^{n(\mathbf{y})}\binom{n(\mathbf{y})}{l}(-1)^{n(\mathbf{y})-l}p(l).$$

It follows that X_c is Markov with respect to the relation

$$\begin{aligned} u \sim_c v \quad &\Longleftrightarrow \quad b_c(u,r)\cap b_c(v,r) \neq \emptyset \\ &\Longleftrightarrow \quad \exists w : \nu_c^1([u,w]) \leq r \wedge \nu_c^1([w,v]) \leq r. \end{aligned} \tag{15}$$

The neighbourhood of a point u is

$$\partial_c(u) = \bigcup_{w\in b_c(u,r)} b_c(w,r), \tag{16}$$

which in general is not ball shaped. As the triangular inequality does not necessarily hold for scaled distances defined by ν_c^1, it is possible that two points are neighbours in X_c although their scaled distance may be larger than $2r$. However, in analogy with the results obtained for the distance interaction processes, the following proposition holds.

Proposition 4 *For a shot noise process,*

$$c(u) = \tilde{c}\ \forall\ u \in b(x, 2\tilde{c}r) \quad \Longrightarrow \quad \partial_c(x) = b(x, 2\tilde{c}r).$$

Proof: Consider $w \in b_c(x,r)$. Then $b(w,\tilde{c}r) \subset b(x,2\tilde{c}r)$ and thus $c(u) = \tilde{c}$ for all $u \in b(w,\tilde{c}r)$. Therefore Proposition 1 yields

$$b_c(x,r) = b(x,\tilde{c}r) \text{ and } b_c(w,r) = b(w,\tilde{c}r),$$

hence

$$\partial_c(x) = \bigcup_{w\in b_c(x,r)} b_c(w,r) = \bigcup_{w\in b(x,\tilde{c}r)} b(w,\tilde{c}r) = b(x,2\tilde{c}r).$$

$\square$

If the scaling function is constant in a neighbourhood of a point x, the conditional intensity of a locally scaled shot noise process again behaves like under global scaling.

$\boxed{\mathbf{E}}$

Proposition 5 *If $c(u) = c(x)$ for all $u \in b(x, 2c(x)r)$, then*

$$\lambda_c^{(c)}(x \mid \mathbf{x}) = \lambda\Big(\frac{x}{c(x)} \,\Big|\, \frac{\mathbf{x}}{c(x)}\Big).$$

Proof: Since the conditional intensity can be written as a product of interactions, $\lambda(x \mid \mathbf{x}) = \prod_{\mathbf{y} \subseteq \mathbf{x} \cap \partial(x)} \varphi(\mathbf{y} \cup \{x\})$, we only need to show that

$$\prod_{\mathbf{y} \subseteq \mathbf{x} \cap \partial_c(x)} \varphi_c(\mathbf{y} \cup \{x\}) = \prod_{\frac{\mathbf{y}}{c(x)} \subseteq \frac{\mathbf{x}}{c(x)} \cap \partial(\frac{x}{c(x)})} \varphi\Big(\frac{\mathbf{y} \cup \{x\}}{c(x)}\Big).$$

This is fulfilled if

$$m_c(\mathbf{y} \cup \{x\}) = m\left(\frac{\mathbf{y} \cup \{x\}}{c(x)}\right) \quad \text{for finite } \mathbf{y} \subset \partial_c(x),\ \mathbf{y} \neq \emptyset,$$

i.e., if

$$\nu_c^k\Big(\bigcap_{z \in (\mathbf{y} \cup \{x\}) \cap \partial_c(x)} b_c(z, r)\Big) = \nu^k\Big(\bigcap_{z \in \frac{(\mathbf{y} \cup \{x\}) \cap \partial_c(x)}{c(x)}} b(z, r)\Big). \tag{17}$$

By the assumption and with similar arguments as in the proof of Proposition 1,

$$b_c(z, r) \cap \partial_c(x) = b(z, c(x)r) \cap \partial_c(x) \quad \text{for any } z \in \partial_c(x),$$

and thus

$$b_c(z, r) \cap b_c(x, r) = b(z, c(x)r) \cap b(x, c(x)r).$$

Therefore,

$$\bigcup_{z \in (\mathbf{y} \cup x) \cap \partial_c(x)} b_c(z, r) = \bigcup_{z \in (\mathbf{y} \cup x) \cap \partial_c(x)} b(z, c(x)r) = \bigcup_{z \in \frac{(\mathbf{y} \cup x) \cap \partial_c(x)}{c(x)}} c(x) b(z, r),$$

which immediately leads to (17) □

6. Approximation of local scaling

For simulation of locally scaled Markov point processes using e.g. the Metropolis-Hastings algorithm, one has to evaluate expressions of the form $g(\mathbf{x}; \nu_c^*)$. This usually involves integration with respect to scaled d-dimensional volume measures ν_c^d. In locally scaled distance-interaction processes introduced in Section 4, for example, we deal with scaled distances

$$\nu_c^1([u, v]) = \int_{[u,v]} c(w)^{-1} dw = \|u - v\| \underbrace{\int_0^1 c\,(u + t(v - u))^{-1}\, dt}_{\overline{c^{-1}}(u,v)}. \tag{18}$$

The function $\overline{c^{-1}}(u,v)$ is the integral mean of the inverse scaling function $w \to 1/c(w)$ on the segment $[u,v]$.

For certain scaling functions c, such integrals can be expressed explicitly. However, if one strives to design programs that handle arbitrary scaling functions, one would have to resort to numeric algorithms. Time consuming calculations can be avoided by defining *approximately scaled* processes that require only pointwise evaluation of the scaling function.

Markov point processes comprise a large variety of models that are essentially different to each other e.g. with respect to the order of interaction, or to the dimensionality of volume measures involved in the definition of their density. There is no best recipe for approximate local scaling of all possible models. However we can cover a large range of models with the approaches proposed below for distance-interaction processes, finite order interaction processes, and shot noise processes.

6.1. Local scaling by c-averaging for distance-interaction processes

Locally scaled distance-interaction processes are based on the neighbour relation $u \sim_c v \iff \nu_c^1([u,v]) \leq r$. They require only the calculation of scaled pairwise distances

$$\nu_c^1([u,v]) = \|u-v\|\overline{c^{-1}}(u,v),$$

cf. (18). A natural idea is to replace the integral mean $\overline{c^{-1}}(u,v)$ by a simpler mean $\widehat{c^{-1}}(u,v)$ of the inverse scaling factors $c(u)^{-1}$ and $c(v)^{-1}$. We propose to use the harmonic mean

$$\widehat{c^{-1}}(u,v) = \frac{2}{c(u)+c(v)},$$

which leads to the approximated neighbour relation

$$u \sim_{\hat{c}} v \quad \iff \quad \|u-v\| \leq \frac{1}{2}\big(c(u)+c(v)\big)r. \tag{19}$$

This relation allows for a nice geometric interpretation. Two points u, v are neighbours iff the balls $b(u, \frac{1}{2}c(u)r)$ and $b(v, \frac{1}{2}c(v)r)$ overlap. Note that (19) actually means that the scaling function c itself is locally replaced by the arithmetic mean $\frac{1}{2}\big(c(u)+c(v)\big)$. Therefore we call this approach *local scaling by c-averaging.*

In proper locally scaled distance-interaction processes, neighbourhoods are always star shaped, i.e.

$$u \sim_c v \;\Rightarrow\; u \sim (1-t)u + tv \quad \text{for all } t \in [0,1]. \tag{20}$$

E

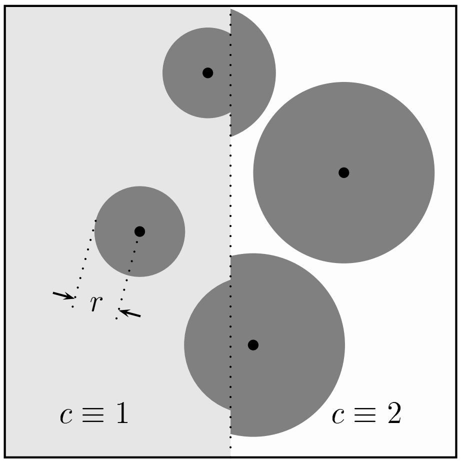


Figure 6: Neighbourhoods of four points in an approximately locally scaled distance-interaction process, obtained by c-averaging with scale factor $c \equiv 1$ in the left half and $c \equiv 2$ in the right half of the domain. Compare with Figure 5, where the true neighbourhoods of the same four points are shown.

Conversely, neighbourhoods in distance-interaction processes obtained by c-averaging are not necessarily star shaped, as the following example shows.

Example 2. Consider a distance-interaction process on $[0,1]^2$, obtained by c-averaging with $c(u) = 1$ for $u \in (-\infty, 0.5] \times \mathbb{R}$ and otherwise, $c(u) = 2$, which is the same situation as in Figure 5. However now, the neighbourhood of a point $u \in (0.5 - R, 0.5] \times \mathbb{R}$ is no longer star shaped, see Figure 6. □

A relatively weak Lipschitz condition on the scaling function ensures that the inhomogeneous point process has star shaped neighbourhoods.

Proposition 6 *Let X_c be an inhomogeneous distance-interaction process on $\mathcal{Y}$ obtained by local c-averaging from a homogeneous process with interaction distance r using the scaling function c. Then X_c has star shaped neighbourhoods if*

$$|c(u) - c(v)| \leq \frac{2}{r}\|u - v\| \quad \textit{for all } u, v \in \mathcal{Y}. \tag{21}$$

Proof: Let $u, v \in \mathcal{Y}$ be related with respect to the approximated relation, $u \sim_{\hat{c}} u$, and let $w \in [u, v]$ be any point on the line segment between u and v. Then we need to show that $u \sim_{\hat{c}} w$ under the Lipschitz condition (21). This is trivially fulfilled if $c(w) \geq c(v)$. Otherwise, by (21),

$$c(v) - c(w) \leq \frac{2}{r}\|v - w\|.$$

Recall that under c-averaging,

$$u \sim_{\hat{c}} v \Longleftrightarrow \|u-v\| \leq \tfrac{1}{2}\big(c(u)+c(v)\big)r \Longleftrightarrow c(u)+c(v) \geq \frac{2}{r}\|u-v\|.$$

Subtracting the two inequalities, we find that

$$c(u)+c(w) \geq \frac{2}{r}\big(\|u-v\|-\|v-w\|\big) = \frac{2}{r}\,\|u-w\|.$$

Thus $u \sim_{\hat{c}} w$. $\square$

6.2. Local scaling by φ-averaging for finite-order interaction processes

The ideas presented above are not directly applicable in cases where the density function involves more complicated features than just pairwise distances. In order to restrict the evaluation of c nevertheless to the points in the pattern $\mathbf{x} = \{x_1, \cdots, x_n\}$, one could resort to interpreting local scaling as an average of global scalings, with scaling factors $c(x_1), \cdots, c(x_n)$, i.e. defining

$$f^{(c)}_{X_c}(\mathbf{x}) = \text{Average}\big(f_X(\mathbf{x}/c(x_1)), \cdots, f_X(\mathbf{x}/c(x_n))\big).$$

This would however invalidate the paradigm of locally defined interaction, since interaction in a subset $\mathbf{y} \subset \mathbf{x}$ would then also be modified by scale factors taken from points outside $\mathbf{y}$, in $\mathbf{x} \setminus \mathbf{y}$.

In cases where the homogeneous process is Markov with interactions of finite order k, i.e. where $\varphi(\mathbf{y}) \equiv 1$ if $n(\mathbf{y}) > k$, a feasible concept of local averaging is therefore based on averaging interaction functions. We propose to define m-th order interaction by the geometric mean

$$\hat{\varphi}_c(\mathbf{y}) = \left(\prod_{y \in \mathbf{y}} \varphi\big(\mathbf{y}/c(y)\big)\right)^{1/m}, \quad n(\mathbf{y}) = m \geq 2. \tag{22}$$

Thus we obtain the density $\hat{f}_{X_c}$ *by local φ-averaging* as

$$\hat{f}^{(c)}_{X_c}(\mathbf{x}) \propto \prod_{\mathbf{y} \subseteq \mathbf{x}} \hat{\varphi}_c(\mathbf{y}),$$

or

$$\hat{f}_{X_c}(\mathbf{x}) = \prod_{x \in \mathbf{x}} c(x)^{-1} \prod_{\mathbf{y} \subseteq \mathbf{x}} \hat{\varphi}_c(\mathbf{y}).$$

Using the geometric mean in (22) is motivated by the fact that interaction functions are usually of the form

$$\varphi(\mathbf{y}) = \exp(p(\mathbf{y})),$$

where $p(\cdot)$ is the so-called potential function. This notion stems from statistical physics, where Markov point processes were first described as Gibbs processes. Taking the geometric mean of φ means taking the arithmetic mean of the potential function,

$$\hat{\varphi}_c(\mathbf{y}) = \exp\left(\frac{1}{m}\sum_{y\in\mathbf{y}} p\big(\mathbf{y}/c(y)\big)\right), \quad n(\mathbf{y}) = m \geq 2.$$

Of course, this concept can also be applied to distance-interaction processes. The following example illustrates the difference between c-averaging and φ-averaging in local scaling of a Strauss process.

Example 1 (Strauss process, continued, continued). The density of an approximately locally scaled Strauss process takes the form

$$\hat{f}_{X_c}(\mathbf{x}) \propto (c(x)^{-k}\beta)^{n(x)}\gamma^{\hat{s}_c(\mathbf{x})},$$

where the definition of $\hat{s}_c(\mathbf{x})$ depends on the type of approximation used.

For local scaling by c-averaging, $\hat{s}_c(\mathbf{x})$ is the number of neighbour pairs in $\mathbf{x}$ defined by the relation

$$u \sim_{\hat{c}} v \iff \|u-v\| \leq \frac{1}{2}\big(c(u)+c(v)\big)r.$$

For local scaling by φ-averaging, $\hat{s}_c(\mathbf{x})$ is calculated from the number of directed neighbours given by the relation

$$u \rightsquigarrow v \iff \|u-v\| \leq c(u)r.$$

Here,

$$\hat{s}_c(\mathbf{x}) = \sum_{\{u,v\}\subseteq\mathbf{x}}^{\neq} \frac{1}{2}\big(\mathbf{1}\,(u \rightsquigarrow v) + \mathbf{1}\,(v \rightsquigarrow u)\big)$$

is the number of directed neighbours divided by two. $\square$

6.3. Local scaling by influence zones for shot noise processes

Shot noise processes as defined in Section 5 require the evaluation of the coverage function

$$C_{c,\mathbf{x}}(u) = \sum_{x\in\mathbf{x}} \mathbf{1}\,(u \in b_c(x,r))$$

which gives the number of "influence zones" $b_c(x,r)$ covering a point u. In Proposition 1 we saw that $b_c(x,r) = b(x,\tilde{c}r)$ if $c(u) = \tilde{c} = c(x)$ for all

$u \in b(x, \tilde{c}r) = b(x, c(x)r)$. Assuming that c does not vary very much in $b(x, c(x)r)$, we can use this result to approximate the influence zones by

$$b_c(x, r) \approx b(x, c(x)r)$$

and thus obtain the coverage function

$$\hat{C}_{c,\mathbf{x}}(u) = \sum_{x \in \mathbf{x}} \mathbf{1}\left(u \in b(x, c(x)r)\right).$$

Calculating the density function

$$\hat{f}^{(c)}_{X_c}(\mathbf{x}) \propto \beta^{n(\mathbf{x})} \gamma^{-\int p(\hat{C}_{c,\mathbf{x}}(u))\nu^k_c(du)}$$

still requires integration with respect to the locally scaled measure ν^k_c. But even when dealing with homogeneous processes, the integral $\int p(C_{c,\mathbf{x}}(u))du$ is usually approximated by grid methods. Once the coverage function $\hat{C}_{c,\mathbf{x}}(\cdot)$ is known, evaluating

$$\int_{\mathbb{R}^k} p(\hat{C}_{c,\mathbf{x}}(u))\nu^k_c(du) = \int_{\mathbb{R}^k} p(\hat{C}_{c,\mathbf{x}}(u))c(u)^{-k}\nu^k(du)$$

is therefore no bigger a problem than evaluating the corresponding integral in a homogeneous (template) setting.

7. Discussion

Inhomogeneity in natural structures may be caused by very different mechanisms. Correspondingly, there is a myriad of ways to define inhomogeneous models. Therefore some restrictions have to be introduced that replace the usual homogeneity condition. The four models for inhomogeneous point processes described in the introduction stand for four different situations. In the first model, the interaction between points is independent of location. In the second model, inhomogeneity results from a (physical) location dependent thinning, and in the third from (physical) deformation of the matrix. In this paper we have presented a new approach to constructing models for patterns that are homogeneous up to a local scale factor. Such point processes may describe packings of spheres with diameters that vary with location, cf. Figure 1, or situations where both intensity and interaction are governed by the same external factor, such as desert plant communities that are ruled by water supply.

When it comes to choosing an appropriate model for a given situation, there will sometimes be prior information about the physical genesis of the patterns that strongly suggests one of the approaches. In general however it will be necessary to define criteria for the best choice which then can be used to develop model tests. A test on the local scaling assumption could exploit scale invariant local geometric properties, as e.g. shape factors of corresponding Voronoi cells.

The (local) intensity $\lambda_c(x)$ of a globally scaled point process is proportional to the intensity λ of the template, $\lambda_c(x) = c^{-k}\lambda(x/c) \approx c^{-k}\lambda$, since the intensity of a homogeneous template is approximately constant. Analogously, the intensity of a locally scaled process is (approximately)

$$\lambda_c(x) \approx c(x)^{-k}\lambda. \tag{23}$$

This allows firstly to model practically any inhomogeneous intensity and, secondly, to retrieve the scaling function (up to a proportionality constant) from a given or estimated density. In this aspect the scaling function plays a similar role as the survival probability of the thinning model.

Once the scaling function has been estimated, it can be used to subsequently fit the parameters of the template process and thus to complete the model specification. A similar approach has been followed by Nielsen and Jensen (2001) for fitting the transformation model. Furthermore possible empirical relations between estimated scaling function and explanatory variables such as water supply in the desert vegetation case can be used for prediction purposes.

Future work will concentrate on validating the approximation (23) as well as on development of model tests and other statistical methods.

Appendix: Local scaling of Markov point processes, general case

It is easy to show that a homogeneous Markov point process X with scale invariant density function has scale invariant interaction functions $\varphi(\mathbf{y}) = h(\mathbf{y};\nu^*)$. Then a locally scaled version X_c on $\mathcal{Y}$ has interaction function

$$\varphi_c(\mathbf{y}) = h(\mathbf{y};\nu_c^*). \tag{24}$$

It is indeed possible to define a neighbour relation $\sim_c$ such that φ_c is a proper interaction function. Hence the locally scaled process is Markov, too. The following definition which may look a little awkward at the first glance is in fact consistent with the simple neighbour relations derived earlier for locally scaled distance-interaction and shot noise processes.

Definition 4 *Consider a function φ_c on $\Omega_\mathcal{Y}$ as given by* (24). *Two points u and v are called* neighbours with respect to the scaled relation $\sim_c$*, iff there* *exists a finite point configuration* $\mathbf{y} \in \Omega_{\mathcal{Y}}$ *such that*

$$\varphi_c(\mathbf{y} \cup \{u, v\}) \neq 1.$$

Proposition 7 *Let* X_c *be a locally scaled point process with density*

$$f_{X_c}^{(c)}(\mathbf{x}) = \prod_{\mathbf{y} \subseteq \mathbf{x}} \varphi_c(\mathbf{y}).$$

Then X_c *is Markov with respect to the neighbour relation* $\sim_c$ *given in Definition 4.*

Proof: The proposition follows from the Hammersley-Clifford theorem, if the function φ_c is a proper interaction function with respect to the relation $\sim_c$, i.e. if $\varphi_c(\mathbf{y}) \neq 1$ implies that $\mathbf{y}$ is a $\sim_c$ clique. To see this, consider a set $\mathbf{y}$ with $n(\mathbf{y}) \geq 2$ and $\varphi_c(\mathbf{y}) \neq 1$. Then, by definition, any two points u and v in $\mathbf{y}$ are neighbours with respect to $\sim_c$, as $\varphi_c(\mathbf{y}^* \cup \{u, v\}) \neq 1$ with $\mathbf{y}^* = \mathbf{y} \setminus \{u, v\} \in \Omega_{\mathcal{Y}}$. $\square$

References

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

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.

Hoffmann-Jørgensen, J. (1994). *Probability with a view toward statistics. Vol. I.* Chapman & Hall, New York.

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*, volume 37 of *IMS Lecture Notes*, pages 297–318.

Nielsen, L. S. and Jensen, E. B. V. (2001). Statistical inference for transformation inhomogeneous point processes. *Research Report 12, Laboratory for Computational Stochastics, University of Aarhus.* Submitted.

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.

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

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 Molchanov, I. S. (1998). Shot noise weighted processes: a new family of spatial point processes. *Comm. Statist. Stochastic Models*, 14(3):715–734.

F

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*.

On statistical inference in TIM models - a study of the cell data

Linda Stougaard Nielsen
Laboratory for Computational Stochastics
University of Aarhus

This report can be considered as a supplement to the paper Nielsen and Jensen (2001). All methods regarding MLE for the cell data will be documented in details.

The report falls in two parts. In Section 1 the data and the model is introduced and the techniques used to analyse the model are presented. At the end of Section 1 an overview of an algorithm for MLE is given.

Section 2 deals with the analysis of the data set and the particular results and experiences achieved. We use the methods described in Section 1 along with standard statistical methods. Furthermore, various tests are presented.

1 Background: The data and the model

1.1. The data

The cell data set is a map of the cells in a 2 dimensional section of the mucous membrane in the stomach of a healthy rat. There is a trend in the cell intensity, and the section is taken such that the trend follows the first axis. The original image of cells has been converted into points marking the centres of the cells sections. See Nielsen (2000) and Nielsen and Jensen (2001) for more detail.

In the following we let $\mathcal{X} = [0;1] \times [0;0.89]$ denote the observation window and Ω be the set of all finite point patterns in $\mathcal{X}$. The observed point pattern will be denoted y. We will model the point pattern using a point process allowing for inhomogeneity as well as interaction in the form of inhibition.

1.2. The model

The point pattern is analyzed using a TIM model, see Jensen and Nielsen (2000) and Nielsen and Jensen (2001). In short, let X be a homogeneous Markov point process in $\mathcal{X}$ parametrized by $\psi \in \Psi$ and $h_\theta : \mathcal{X} \to \mathcal{X}$ a differentiable and bijective mapping parametrized by $\theta \in \Theta$. Then $Y =$

$h_\theta(X)$ is an inhomogeneous point process model, a TIM model, and the likelihood function for (ψ, θ) decomposes as follows,

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

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

The Strauss point process is useful to describe inhibition in a point pattern and is used as the underlying homogeneous point process X for the cell data. Thus, $\psi = (\beta, \gamma, r)$, $\Psi = B \times G \times R = (0; \infty) \times [0; 1] \times (0; \infty)$ and

$$L_{\text{hom}}(\beta, \gamma, r; x) = c(\beta, \gamma, r)^{-1} \beta^{n(x)} \gamma^{s_r(x)},$$

where $c(\beta, \gamma, r) = \int_\Omega \beta^{n(x)} \gamma^{s_r(x)} \Pi(dx)$ is the normalising constant, $n(x)$ is the number of points in x and $s_r(x)$ is the number of point pairs with distance less than r. Notice that $\gamma = 0$ is the hard-core model.

From the sampling we know that there is a trend in the first coordinate direction. There should be none in the second. However initially we will allow for inhomogeneity along both axis and use a coordinate-wise transformation, where the coordinates are transformed independently using the simple transformation, cf. Nielsen and Jensen (2001). Thus

$$L_0(\theta; y) = \left(\frac{\theta_1}{e^{\theta_1} - 1} \frac{a\theta_2}{e^{a\theta_2} - 1} \right)^{n(y)} e^{\sum_{\eta \in y} \theta \cdot \eta}.$$

Here $a = 0.893175$, $\theta = (\theta_1, \theta_2) \in \Theta = \mathbb{R}^2 \backslash 0$. If $\theta_i = 0$ then the corresponding part of L_0 becomes 1. Note that this is the case of homogeneity along the i'th axis. Consequently the i'th coordinates do not enter into L_0. The inverse transformation is of the form

$$h_\theta(u_1, u_2)^{-1} = \left(\frac{e^{\theta_1 u_1} - 1}{e^{\theta_1} - 1}, a \frac{e^{\theta_2 u_2} - 1}{e^{a\theta_2} - 1} \right), \quad (u_1, u_2) \in \mathcal{X}.$$

for all $\theta \in \Theta$ except $\theta_i = 0$, $i = 1, 2$, in which case the i'th coordinate of h_θ is the identity mapping.

1.3. MLE

Exponential family theory can be used in the maximization of the likelihood function (1). For fixed (r, θ) the function $L_{\text{hom}}(\beta, \gamma, r; h_\theta^{-1}(y))$ is of exponential family form. The likelihood equations are

$$\begin{array}{lllll} \mathbb{E}_{\beta,\gamma,r} n(X) & = & n(x) & = & n(y) \\ \mathbb{E}_{\beta,\gamma,r} s_r(X) & = & s_r(x) & = & s_r(h_\theta^{-1}(y)), \end{array} \tag{2}$$

$\boxed{\mathbf{F}}$

where X is a homogeneous Strauss point process. The solution is denoted $(\hat{\beta}(r,\theta), \hat{\gamma}(r,\theta))$.

The profile likelihood with nuisance parameter (r,θ) becomes

$$\begin{aligned}\overline{L}(r,\theta;y) &= L_0(\theta;y) \max_{\beta,\gamma} L_{\text{hom}}(\beta,\gamma,r;h_\theta^{-1}(y)) \\ &= L_0(\theta;y)\, L_{\text{hom}}(\hat{\beta}(r,\theta),\hat{\gamma}(r,\theta),r;h_\theta^{-1}(y)). \qquad (3)\end{aligned}$$

Hence, for values of (r,θ) in a grid we solve the equations (2) and compute (3) up to a constant. The grid value maximising the profile likelihood will be denoted $(\hat{r},\hat{\theta})$ and the MLE is then $(\hat{\beta},\hat{\gamma},\hat{r},\hat{\theta}) = (\hat{\beta}(\hat{r},\hat{\theta}),\hat{\gamma}(\hat{r},\hat{\theta}),\hat{r},\hat{\theta})$.

1.4. Finding $T(r)$, a reduced set of θ-values for each $r \in R$

First we show a general result that quite drastically reduces the set of θ-values to be considered.

Corollary 1 *Given $r \in R$. Let $\Theta_k^r = \{\theta \in \Theta : s_r(h_\theta^{-1}(y)) = k\}$ for all $k \geq 0$ and $s_0(r) = s_r(h_{\hat{\theta}_0}^{-1}(y))$.*

Then,

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

is a finite set with at most $s_0(r)+1$ elements and

$$\hat{\theta} \in \bigcup_{r\in R} T(r).$$

Proof. To show the last part of the corollary, we show that for fixed r, $\overline{L}$ is maximized for $\theta \in T(r)$. Then consequently, the MLE $\hat{\theta}$ is contained in the union of $T(r)$.

Thus, fix $r \in R$. Suppose that $\theta, \theta^* \in \Theta_k^r$ for some $k \geq 0$. Then

$$(\hat{\beta}(\theta,r),\hat{\gamma}(\theta,r)) = (\hat{\beta}(\theta^*,r),\hat{\gamma}(\theta^*,r)),$$

see (2). From (3) we now get,

$$L_0(\theta;y) < L_0(\theta^*;y) \Longrightarrow \overline{L}(r,\theta;y) < \overline{L}(r,\theta^*;y).$$

Since $\Theta = \cup_{k=0}^{\infty}\Theta_k^r$, $\overline{L}$ is then maximized in the set $\cup_{k=0}^{\infty}A_k$, where

$$A_k := \{\theta \in \Theta_k^r : L_0(\theta;y) \geq L_0(\theta^*;y),\ \text{for all } \theta^* \in \Theta_k^r\}.$$

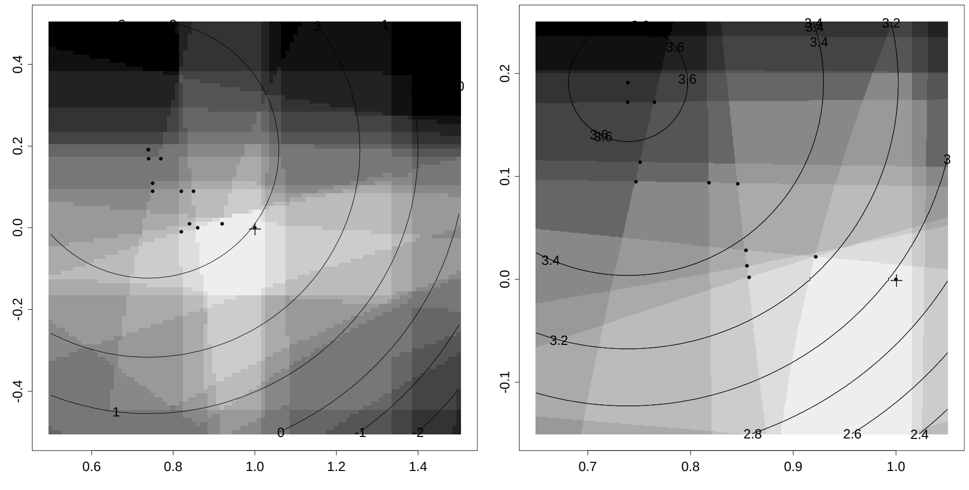
Now we need to find an upper limit for the union. In Jensen and Nielsen (2000, Proposition 5.1) it is shown, that

$$s_r(h_\theta^{-1}(y)) > s_r(h_{\hat{\theta}_0}^{-1}(y)) \Longrightarrow L(\beta,\gamma,r,\theta;y) < L(\beta,\gamma,r,\hat{\theta}_0;y).$$

Hence, we only need to look for the maximum in the sets $\Theta_0,\ldots,\Theta_{s_0(r)}$, and therefore $\overline{L}$ is maximized in $T(r)$.

Now it only remains to show that $T(r)$ has at most $s_0(r)+1$ elements. This is straightforward since $\Theta_k^r = \emptyset \Longrightarrow A_k = \emptyset$. If $\Theta_k^r \neq \emptyset$, then A_k will with probability 1 contain exactly one element because L_0 is strictly monotone. $\square$

Figure 1 illustrates the result in Corollary 1. The point pattern is a simulated realization from the exponential inhomogeneous Strauss point process with parameters $(\beta,\gamma,r,\theta_1,\theta_2) = (700, 0.1, 0.05, 1, 0)$. In the plot, $r = 0.05$, the true value. The plot shows the regions Θ_k^r in different grey tones. Level curves of L_0 are also plotted and from the corollary we know that we are only interested in one θ value in each of the areas Θ_k^r, namely the one with highest L_0 value. These values will always lie on the edge of the Θ_k^r regions, and with probability 1 there will be only one value in each Θ_k^r with maximal L_0. The set $T(r)$ is shown as dots.



(a) Coarse θ-grid (b) Fine θ-grid

Figure 1: Illustration of $T(r)$ for a simulated exponential inhomogeneous Strauss point pattern with $r = 0.05$, the true value. The colours correspond to different Θ_k^r regions, the lighter the colour the smaller the k. The curves are contours of L_0. The true $\theta = (1, 0)$ is plotted as a plus. The tabbing between grid points is 0.01 and 0.001 in the left and right picture, respectively.

$\boxed{\mathbf{F}}$

1.5. Calculating the profile likelihood

Because of the intractable normalising constant, the likelihood function is not known explicitly. Therefore we have to estimate the values up to a constant and use importance sampling for this purpose.

Following Geyer (1999), let X be a point process with density

$$f(x) = c(\psi)^{-1} g_\psi(x), \quad x \in \Omega,$$

with respect to the unit rate Poisson point process Π. Here g_ψ is the explicitly known part of the density, and $c(\psi) = \int_\Omega g_\psi(x)\Pi(dx)$ is the normalising constant. Then,

$$\frac{c(\psi_0)}{c(\psi)} = \mathbb{E}_\psi \left(\frac{g_{\psi_0}(X)}{g_\psi(X)} \right). \tag{4}$$

This mean can be approximated by the sample mean over realizations from the process with parameter ψ which is called importance sampling. However, such an estimate is only good when ψ is sufficiently close to ψ_0.

We want to compute

$$\log \frac{\overline{L}(r, \theta; y)}{\overline{L}(r_0, \hat{\theta}_0; y)} \tag{5}$$

for $r \in R$ and $\theta \in T(r)$. We cannot trust that $(\hat{\beta}(r, \theta), \hat{\gamma}(r, \theta), r, \theta)$ is close enough to $(\hat{\beta}(r_0, \hat{\theta}_0), \hat{\gamma}(r_0, \hat{\theta}_0), r_0, \hat{\theta}_0)$ for using importance sampling. Therefore the ratio will be calculated in terms. Make a path,

$$(r_i, \underbrace{\theta^{(t)}) \to (r_i, \theta^{(t-1)}) \to \cdots \to (}_{\text{inner-plane path}} r_i, \underbrace{\theta^{(0)}) \to (r_{i-1}, \theta^{(0)}) \to \cdots \to (}_{\text{between-planes path}} r_0, \theta^{(0)}),$$

where $(r_i, \theta^{(t)}) = (r, \theta)$ and $\theta^{(0)} = \hat{\theta}_0$. As indicated, the paths are called the inner-plane and the between-planes path. The background for this notation is that it makes sense to first investigate what is going on in the Θ-plane since we have a reduced $T(r)$ for each $r \in R$. See also Figure 2.

Suppose that the path is constructed such that two consecutive parameters (including the (β, γ)–estimate) in the path are close enough to make importance sampling. Then (5) can be computed as follows,

$$\log \frac{\overline{L}(r_i, \theta^{(t)}; y)}{\overline{L}(r_0, \theta^{(0)}; y)} = \underbrace{\sum_{j=1}^{t} \log \frac{\overline{L}(r_i, \theta^{(j)}; y)}{\overline{L}(r_i, \theta^{(j-1)}; y)}}_{\text{inner-plane ratio}} + \underbrace{\sum_{j=1}^{i} \log \frac{\overline{L}(r_j, \theta^{(0)}; y)}{\overline{L}(r_{j-1}, \theta^{(0)}; y)}}_{\text{between-planes ratio}}. \tag{6}$$

In the next two sections we will consider the between-planes and the inner-plane ratios, respectively, and see how the corresponding paths can be constructed to ensure that consecutive parameters are sufficiently close.

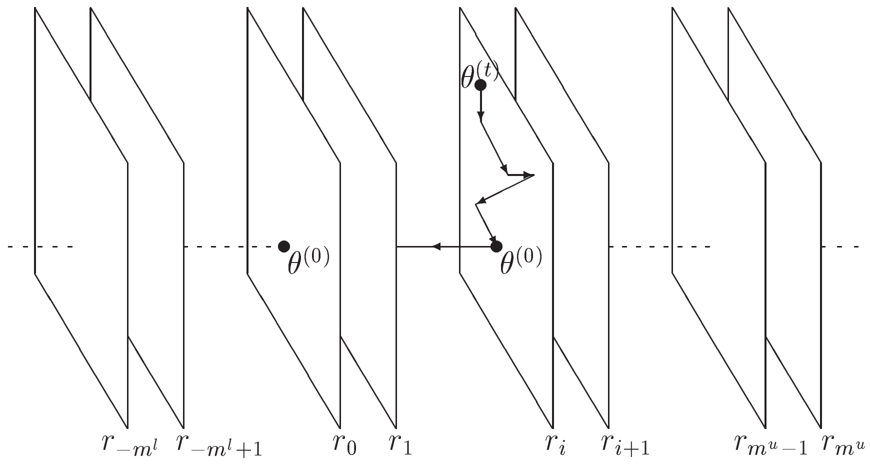


Figure 2: Illustration of the successive calculation of the loglikelihood ratio (6): First calculate the inner-plane ratio from $(\theta^{(t)}, r_i)$ to $(\theta^{(0)}, r_i)$. Then calculate the between-planes ratio from $(\theta^{(0)}, r_i)$ to $(\theta^{(0)}, r_0)$.

1.6. The between-planes ratio

Fix θ. The between-planes ratio then takes the form

$$\log \frac{\overline{L}(r, \theta; y)}{\overline{L}(r_0, \theta; y)} = \log \frac{\overline{L}_{\text{hom}}(r)}{\overline{L}_{\text{hom}}(r_0)}$$

where $\overline{L}_{\text{hom}}(r) := L_{\text{hom}}(\hat{\beta}(r,\theta), \hat{\gamma}(r,\theta), r; h_\theta^{-1}(y))$, is the homogeneous profile likelihood.

Let $d_1 < d_2 < \cdots < d_m$, $m = n(x)(n(x)-1)/2$, be the distances between pairs of points in $x = h_\theta^{-1}(y)$. Then

$$R = (0; d_1] \cup (d_1; d_2] \cup \cdots \cup (d_m; \infty),$$

is a partition of $R = (0; \infty)$ into disjoint intervals. Let $d_0 = 0$ and $d_{m+1} = \infty$. For $r \in (d_i; d_{i+1}]$, $i \in \{0, \ldots, m\}$, we have that $s_r(x) = i$.

Let (β, γ) be fixed. The homogeneous likelihood function $L_{\text{hom}}(\beta, \gamma, \cdot; x)$ is continuous and increasing on intervals $(d_i; d_{i+1}]$. For $r_1 < r_2$, where $r_1, r_2 \in (d_i; d_{i+1}]$ we then have,

$$\begin{aligned} \overline{L}_{\text{hom}}(r_1) &= L_{\text{hom}}(\hat{\beta}(r_1,\theta), \hat{\gamma}(r_1,\theta), r_1; x) \leq L_{\text{hom}}(\hat{\beta}(r_1,\theta), \hat{\gamma}(r_1,\theta), r_2; x) \\ &\leq L_{\text{hom}}(\hat{\beta}(r_2,\theta), \hat{\gamma}(r_2,\theta), r_2; x) = \overline{L}_{\text{hom}}(r_2). \end{aligned}$$

Thus, $\overline{L}_{\text{hom}}(\cdot)$, and thereby also the between-planes ratio, is continuous and increasing on the intervals $(d_i; d_{i+1}]$. Furthermore,

$$\lim_{\epsilon \to 0} \overline{L}_{\text{hom}}(d_i + \epsilon) = \hat{\gamma}(d_i) \overline{L}_{\text{hom}}(d_i),$$

for $i = 2, 3, \ldots, m$. Consequently the function $\log \overline{L}_{\text{hom}}(\cdot)$ have negative jumps in $\{d_2, \ldots, d_m\}$ of size $\log(\hat{\gamma}(d_i))$. In d_1 there is also a negative jump.

Let

$$\mathbb{G}_r = \{r_{-m^l}, r_{-m^l+1}, \ldots, r_{-1}, r_0, r_1, \ldots, r_{m^u-1}, r_{m^u}\},$$

be a grid of close and increasing values of $r \in R$. Between two grid-values, the corresponding number of neighbours are either equal or increases by 1. Thus, it is reasonable to believe that also the corresponding β- and γ-estimates are close and use of importance sampling can be justified.

The between-planes ratio are computed as in (6) where the sum is changed to $-\sum_{j=i+1}^{0}$ if $i < 0$. We need to compute ratios of the form

$$\Delta(j, j-1) = \log \frac{\overline{L}_{\text{hom}}(r_j)}{\overline{L}_{\text{hom}}(r_{j-1})}.$$

Introduce the notation,

$$\gamma_j^\theta = \hat{\gamma}(r_j, \theta), \quad \beta_j^\theta = \hat{\beta}(r_j, \theta), \quad \text{and} \quad s_j^\theta = s_{r_j}(h_\theta^{-1}(y)).$$

Let r_k be the smallest grid value above d_1, and suppose that r_0 is chosen such that $s_0^\theta > 0$. Thus, $k \leq 0$. To compute the between-planes ratio in r_i, the procedure is as follows.

For $i > 0$:	Compute $\Delta(j, j-1)$ for $j = 1, \ldots, i$. Note that $s_j^\theta \geq s_{j-1}^\theta > 0$.
For $i = 0$:	The between-planes ratio is 0.
For $k \leq i < 0$:	Compute $\Delta(j, j-1)$ for $j = l+1, \ldots, 0$. Note that $s_j^\theta \geq s_{j-1}^\theta > 0$.
For $i = k-1$:	As for $k \leq i < 0$. Compute in addition $\Delta(k, k-1)$. Note that $s_k^\theta > 0$ but $s_{k-1}^\theta = 0$.

The between-planes ratios for $i < k-1$ need not to be computed since we are only searching for the maximum of the profile log-likelihood. The reason is that $s_i^\theta = 0$ and therefore $T(r)$ contains only one point. Consequently the inner-plane ratio is 0, which implies that the profile log-likelihood equals the between-planes ratio, cf. (6). Recall that the between-planes ratio is increasing in $(0; d_1]$.

If $s_{j-1}^\theta > 0$, then,

$$\begin{aligned}\Delta(j, j-1) &= (s_j^\theta - s_{j-1}^\theta)\log(\gamma_j^\theta) \\ &\quad - \log \mathbb{E}_{\beta_{j-1}^\theta, \gamma_{j-1}^\theta, r_{j-1}} \left(\left(\frac{\beta_j^\theta}{\beta_{j-1}^\theta} \right)^{n(X)-617} \frac{(\gamma_j^\theta)^{s_{r_j}(X) - s_{j-1}^\theta}}{(\gamma_{j-1}^\theta)^{s_{r_{j-1}}(X) - s_{j-1}^\theta}} \right).\end{aligned}$$

If $s^\theta_{j-1} = 0$ and $s^\theta_j > 0$, then,

$$\begin{aligned}\Delta(j, j-1) &= s^\theta_j \log(\gamma^\theta_j) \\ &\quad + \log \mathbb{E}_{\beta^\theta_j, \gamma^\theta_j, r_j} \left(1\left(s_{r_{j-1}}(X) = 0\right) \left(\frac{\beta^\theta_{j-1}}{\beta^\theta_j}\right)^{n(X)-617} (\gamma^\theta_j)^{-s_{r_j}(X)} \right).\end{aligned}$$

Notice that in both cases, the simulations are performed under the Strauss model. Furthermore the formulas are constructed such that the means are close to 1. This makes the simulation of the means more stable.

1.7. The inner-plane ratio

Fix, $r \in R$. We only need to consider $\theta \in T(r)$. First notice that $\hat{\theta}_0 \in T(r)$ with highest number of neighbours. Now order $T(r)$ and fill up empty holes such that we get a set $T^*(r) = \{\theta^{(0)}, \ldots, \theta^{(m)}\} \supset T(r)$ that fulfills

$$s^{(i-1)} = s^{(i)} + 1, \quad i = 1, \ldots, m, \tag{7}$$

where $s^{(i)} = s(h^{-1}_{\theta^{(i)}}(y))$.

For the Strauss process, the inner-plane ratio is,

$$\sum_{j=1}^{t} \log \frac{\overline{L}(r, \theta^{(j)}; y)}{\overline{L}(r, \theta^{(j-1)}; y)} = \sigma_t + \log L_0(\theta^{(t)}; y) - \log L_0(\theta^{(0)}; y),$$

where

$$\sigma_0 = 0 \quad \text{and} \quad \sigma_j = \sigma_{j-1} + \Delta(j, j-1), \quad j = 1, ..., m,$$

and $\Delta(j, j-1)$ is calculated as follows. Let

$$\gamma^j_r = \hat{\gamma}(r, \theta^{(j)}), \quad \beta^j_r = \hat{\beta}(r, \theta^{(j)}), \quad \text{and} \quad s^j_r = s_r(h^{-1}_{\theta^{(j)}}(y)) = s^{(j)}.$$

Then, if $s^j_r > 0$, we also have that $s^{j-1}_r > 0$ and

$$\begin{aligned}\Delta(j, j-1) &= (s^j_r - s^{j-1}_r) \log(\gamma^{j-1}_r) \\ &\quad - \log \mathbb{E}_{\beta^{j-1}_r, \gamma^{j-1}_r, r} \left(\left(\frac{\beta^j_r}{\beta^{j-1}_r}\right)^{n(X)-617} \left(\frac{\gamma^j_r}{\gamma^{j-1}_r}\right)^{s_r(X)-s^j_r} \right).\end{aligned}$$

If $s^j_r = 0$, then $s^{j-1}_r > 0$ and

$$\begin{aligned}\Delta(j, j-1) &= s^{j-1}_r \log(\gamma^{j-1}_r) \\ &\quad - \log \mathbb{E}_{\beta^{j-1}_r, \gamma^{j-1}_r, r} \left(\left(\frac{\beta^j_r}{\beta^{j-1}_r}\right)^{n(X)-617} 1\left(s_r(X) = 0\right) \right).\end{aligned}$$

$\boxed{\mathbf{F}}$

Notice that in both situations the simulations are performed under a Strauss model.

Importance sampling is used to estimate $\Delta(j, j-1)$. We have made sure that s_r^j and s_r^{j-1} are as close as possible, namely to differ by exactly 1. Thereby the estimates (β_r^j, γ_r^j) and $(\beta_r^{j-1}, \gamma_r^{j-1})$ also are as close as possible since they only depend on s_r^j and s_r^{j-1}, respectively, cf. (2).

The actual values of $\theta^{(j)}$ and $\theta^{(j-1)}$ do not enter into $\Delta(j, j-1)$, they only enter through s_r^j and s_r^{j-1}. Thereby it does not make sense to make a finer grid of θ-values on the path to calculate the inner-plane ratios, since $\Delta(j, j-1)$ is only positive when $s_r^j \neq s_r^{j-1}$. Furthermore it doesn't make any impact whether the θ-values on the path are infinitely close or far apart as long as the corresponding s values differ by exactly 1.

1.8. Algorithm for MLE in the full TIM model

Step 0: Make a grid $\mathbb{G}_r \times \mathbb{G}_1 \times \mathbb{G}_2$ of the nuisance parameter (r, θ_1, θ_2).

Step 1: Find $T(r)$ for each $r \in \mathbb{G}_r$.

Now we have reduced the number of grid points for the nuisance parameters to only consider the set

$$\mathbb{G} = \{(r, \theta) : r \in \mathbb{G}_r \text{ and } \theta \in T(r)\}. \tag{8}$$

Thus, the dimension of the grids $\mathbb{G}_i$ was only important in step number 1. Therefore these grids can be as fine and large as we have energy for, as they are only used this once.

The next two steps deal with estimating β and γ for each $(r, \theta) \in \mathbb{G}$.

Step 2: Find a (β, γ)-grid in which to simulate $\mathbb{E}\, n(X)$ and $\mathbb{E}\, s_r(X)$. This grid has to change in r. We find a way to move the grids and make the simulations in each grid point.

Step 3: Estimate $(\hat{\beta}(r, \theta), \hat{\gamma}(r, \theta))$ using multivariate regression based on the simulations from step 2.

Step 4: Fix $r \in \mathbb{G}_r$. Calculate the inner-plane ratio.

We keep only the value $\hat{\theta}(r) \in T(r)$ that maximizes the inner-plane ratio since they are all to be added to the same between-planes ratio to get the corresponding profile likelihood ratio. Thus, we restrict attention to the set

$$\mathbb{G}_r^{\hat{\theta}} = \{(r, \hat{\theta}(r)) : r \in \mathbb{G}_r\}. \tag{9}$$

Step 5: Fix $\theta = \hat{\theta}_0$. Calculate the between-planes ratio.

Step 6: The final maximization.

2. Analysis of the cell data

Step 0: Grid for the nuisance parameter

First we make a fine grid for the nuisance parameters. The grids need to be around the true values of the parameters. Let $\hat{\theta}_0$ be the MLE based on $L_0(\theta; y)$, $\hat{\theta}_0 = (1.304, -0.272)$. We know that $\hat{\theta}_0$ is close to $\hat{\theta}$ and therefore the grids $\mathbb{G}_1$ and $\mathbb{G}_2$ are centered around $\hat{\theta}_0$ with 400 grid values to each side and step size 0.001.

In Nielsen (2000), we supposed that $r = 0.007$. This estimate was based on the J–function and on profile pseudo-likelihood analysis for the point pattern $h_{\hat{\theta}_0}^{-1}(y)$, see the Appendix. Here we have chosen the r-grid to

$$\mathbb{G}_r = \{0.00500, 0.00505, \ldots, 0.01000\}.$$

Step 1: Finding $T(r)$ for each $r \in \mathbb{G}_r$

In Figure 3 we have plotted the regions Θ_k^r and the set $T(r)$ for $r = 0.0087$. See Section 1.4 and Figure 1 for details. Let

$$S(r) = \{s_r(h_\theta^{-1}(y)) : \theta \in T(r)\}.$$

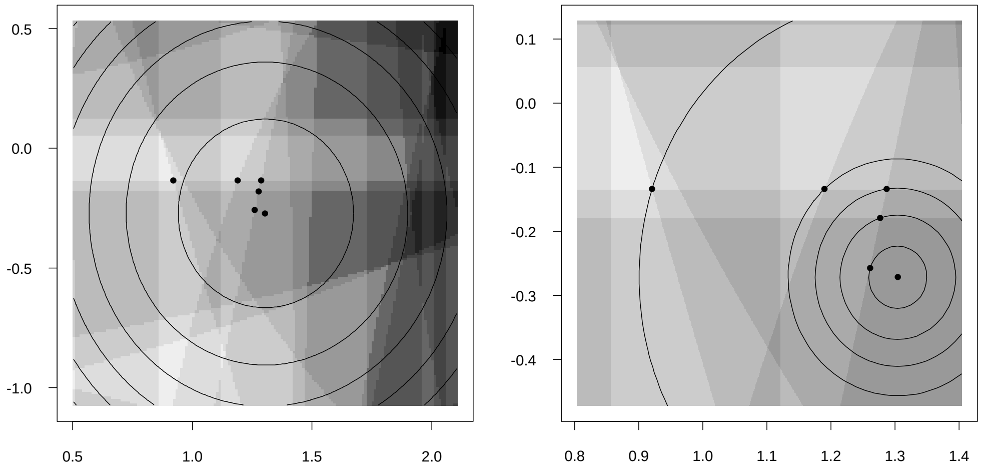
There is a 1 to 1 correspondence between the elements in $S(r)$ and the elements in $T(r)$. Furthermore, $S(r) \subset \{0, 1, 2, \ldots\}$.

In Figure 4 the sets $S(r)$ are plotted for the cell data for each value of $r \in \mathbb{G}_r$. We see that the number of elements in each set is very small, at most 6. The nuisance parameter grid $\mathbb{G}_r \times \mathbb{G}_1 \times \mathbb{G}_2$ contains $101 * 801 * 801$ grid points. Henceforth we only need to consider the set $\mathbb{G}$ given in (8), which contains only 271 elements. Thus, we see that the dimension of $\mathbb{G}_1 \times \mathbb{G}_2$ is indeed not very critical. The dimension of $\mathbb{G}_r$ is on the other hand very critical since this set will not be reduced and therefore effects the run-time of all the following steps.

Step 2: Finding a (β, γ)-grid in which to simulate $\mathbb{E}\, n(X)$ and $\mathbb{E}\, s_r(X)$

Next step of the estimation procedure is to find $\hat{\beta}(r, \theta)$ and $\hat{\gamma}(r, \theta)$ for each of the 271 values of $(r, \theta) \in \mathbb{G}$. Thus we need to estimate $\mathbb{E}_{\beta,\gamma,r}\, n(X)$ and

F



(a) Coarse θ-grid

(b) Fine θ-grid

Figure 3: 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 picture, respectively.

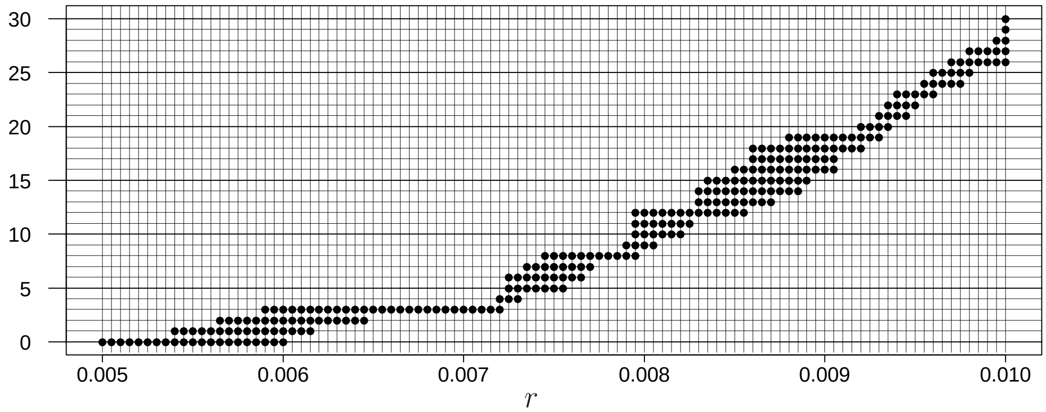


Figure 4: The sets $S(r)$ plotted for each value of r in the grid.

$\mathbb{E}_{\beta,\gamma,r} s_r(X)$ and solve the likelihood equations (2) with $n(x) = 617$ and $s(x) \in S(r)$, $r \in \mathbb{G}_r$.

As usual we approximate the means by the sample means,

$$\begin{array}{lclcl} \mathbb{E}_{\beta,\gamma,r} n(X) & \approx & \frac{1}{m}\sum_{i=1}^{m} n(x_i) & =: & \hat{\mathbb{E}} n \\ \mathbb{E}_{\beta,\gamma,r} s_r(X) & \approx & \frac{1}{m}\sum_{i=1}^{m} s_r(x_i) & =: & \hat{\mathbb{E}} s, \end{array} \tag{10}$$

where $x_1, \ldots, x_m$ are realizations from the Strauss process with parameters (β, γ, r).

Fix $r \in \mathbb{G}_r$ and let $\mathbb{G}^r_\beta \times \mathbb{G}^r_\gamma$ be a grid of (β, γ) values in which to simulate the mean number of points and neighbours. We let $\mathbb{G}_\beta^r \times \mathbb{G}_\gamma^r$ vary in r such that the corresponding values of $(\hat{\mathbb{E}n}, \hat{\mathbb{E}s})$ are around the values $(617, s)$, $s \in S(r)$. The goal of step 2 is therefore to find values $(\beta_c(r), \gamma_c(r))$ that can be used as centres in the grids $\mathbb{G}_\beta^r \times \mathbb{G}_\gamma^r$.

By trial and error we found usable centres, see also Figure 5,

$$\begin{aligned} \gamma_c(r) &= \begin{cases} -0.19 + 40 * r, & r \in [0.0055, 0.007] \\ -0.8883333 + 140 * r, & r \in [0.007, 0.008] \\ -0.4282802 + 83.22355 * r, & r \in [0.008, 0.01] \end{cases} \\ \beta_c(r) &= \begin{cases} 642.5 + 17500 * r, & r \in [0.0055, 0.007] \\ 730.1667 + 5000 * r, & r \in [0.007, 0.008] \\ 693.6249 + 9613.009 * r, & r \in [0.008, 0.01] \end{cases} \end{aligned} \tag{11}$$

The size of $\mathbb{G}_\beta^r \times \mathbb{G}_\gamma^r$ were small since the centres (11) are very good. In Figure 6 the simulations for six different values of r are shown. The grid points of $\mathbb{G}_\beta^r \times \mathbb{G}_\gamma^r$ are shown as dots. For β we used 5 grid points to each side of $\beta_c(r)$ with equal spacing 5. For γ we used 6 grid points to each side of γ. The grid was not rectangular since we tried to keep $\hat{\mathbb{E}n}$ close to 617, meaning that there was no reason to simulate the cases when β and γ were both small or large simultaneously. For small values of $\gamma_c(r)$, the grid evenly covered the interval $(0; \gamma_c(r))$, see e.g. Figure 6 (b). For larger values, the step size was 0.001. For $r = 0.0059, \ldots, 0.0062$ we had some problems with $\hat{\mathbb{E}s}$ not being large enough, so we used 10 grid points above $\gamma_c(r)$ instead, see Figure 6 (b).

The plots in Figure 6 also show the results of the simulations. In each grid point we calculated $\hat{\mathbb{E}n}$ and $\hat{\mathbb{E}s}$ using (10). A local regression model were fitted for both the mean number of points and the mean number of neighbours using the `Splus` commando `loess`. In each of the plots, the dotted curve is the level curve for the local regression model for $\hat{\mathbb{E}n}$ corresponding to $\hat{\mathbb{E}n} = 617$ and the full curves are level curves corresponding to $\hat{\mathbb{E}s} = s$, where s are integers as indicated on the plots. For each r, the intersections between the dotted curve and the full curves cover the set $\{(617, s) : s \in S(r)\}$.

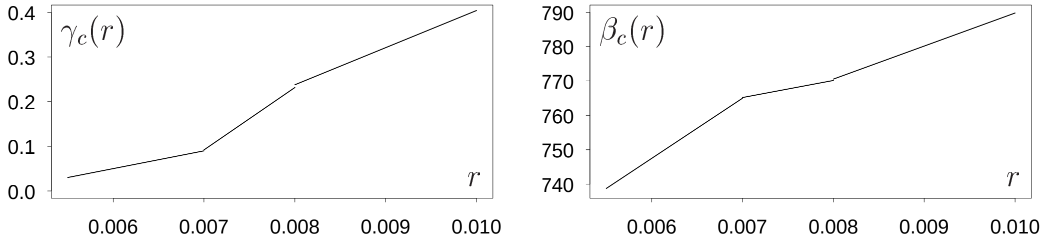
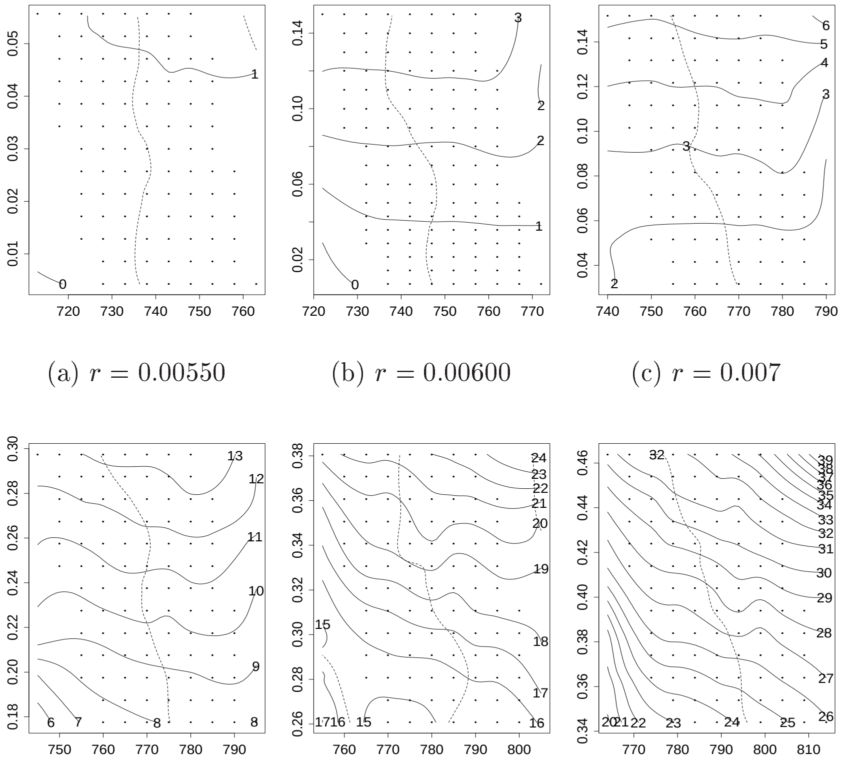


Figure 5: Grid midpoints of γ and β for each $r \in R$.

$\boxed{\mathbf{F}}$



(a) $r = 0.00550$ (b) $r = 0.00600$ (c) $r = 0.007$

(d) $r = 0.008$ (e) $r = 0.009$ (f) $r = 0.010$

Figure 6: In each plot, r is fixed as indicated. The dots are grid points (β, γ) centered around the value given in (11). $\hat{\mathbb{E}}nX$ and $\hat{\mathbb{E}}\,sX$ have been evaluated in each grid point and two surfaces, $\hat{n}$ and $\hat{s}$, have been estimated. The dotted curve is the level curve $\hat{n}(\beta, \gamma) = 617$ and the full curves are level curves $\hat{s}(\beta, \gamma) = s$, $s \in \mathbb{Z}$.

Step 3: Estimating $(\hat{\beta}(r, \theta), \hat{\gamma}(r, \theta))$

In this step we find $(\hat{\beta}(r, \theta), \hat{\gamma}(r, \theta))$ and check that the likelihood equations (2) are solved satisfactory.

We use the estimations $(\hat{\mathbb{E}n}, \hat{\mathbb{E}s})$ from step number 2, shown in Figure 6. The intersections of the curves could be used as estimates for $(\hat{\beta}(r, \theta), \hat{\gamma}(r, \theta))$. However, dealing with random estimates, we choose not to trust the estimates to much. Instead we use a linear regression model, and find the intersections of the level lines. The model is

$$Y = XB + U,$$

where X is the input, a $m \times 3$ matrix with rows $(\beta_i, \gamma_i, 1)$ for each of the m grid points of $\mathbb{G}^r_\beta \times \mathbb{G}^r_\gamma$. Y is the response, a $m \times 2$ matrix with rows

$(\hat{\mathbb{E}n_i}, \hat{\mathbb{E}s_i})$. The matrix B with dimensions 3×2 is the unknown regression parameters. The matrix U is the stochastic error.

From Mardia et al. (1994, Chapter 6) we find the MLE of B, based on normal distributed error,

$$\hat{B} = (X'X)^{-1}X'Y,$$

which gives,

$$\begin{aligned}\hat{\mathbb{E}n} &= \hat{B}_{11} * \beta + \hat{B}_{21} * \gamma + \hat{B}_{31}\\ \hat{\mathbb{E}s} &= \hat{B}_{12} * \beta + \hat{B}_{22} * \gamma + \hat{B}_{32},\end{aligned}$$

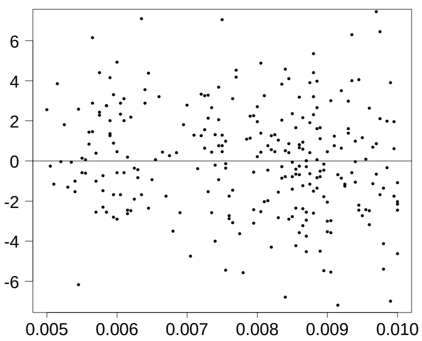
where $\hat{B} = \{\hat{B}_{ij}\}$. The level lines $\hat{\mathbb{E}n} = 617$ and $\hat{\mathbb{E}s} = s$ are linear,

$$\beta(s) = a_\beta + b_\beta * \gamma(s), \quad \text{and} \quad \gamma(s) = a_\gamma + b_\gamma * s \tag{12}$$

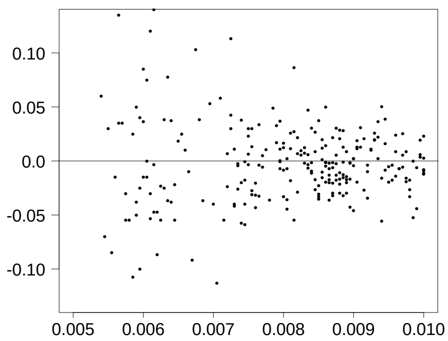
where

$$\begin{aligned} a_\beta &= \frac{617 - \hat{B}_{31}}{\hat{B}_{11}}, & b_\beta &= -\frac{\hat{B}_{21}}{\hat{B}_{11}}\\ a_\gamma &= \frac{1}{\hat{B}_{12} * b_\beta + \hat{B}_{22}}, & b_\gamma &= -(b_\beta * \hat{B}_{12} + \hat{B}_{32}) * a_\gamma.\end{aligned}$$

For each $r \in R$ we have a regression model as described above. Then, for each $s \in S(r)$ with corresponding $\theta \in T(r)$ we use (12) to compute $\hat{\gamma}(r,\theta) = \gamma(s)$ and $\hat{\beta}(r,\theta) = \beta(s)$.



(a) $\hat{\mathbb{E}}\, nX - 617$



(b) $(\hat{\mathbb{E}}\, sX - s)/s$

Figure 7: For each $s \in S(r)$ we have found $(\hat{\beta}, \hat{\gamma})$. Here we have calculated $\hat{\mathbb{E}n}$ and $\hat{\mathbb{E}s}$ based on each of these estimates and plotted the indicated residuals against the corresponding r.

$\boxed{\textbf{F}}$

In Figure 7 we have checked that the estimates of β and γ actually solves the likelihood equations (2). For each (r, s), $s \in S(r)$, $r \in \mathbb{G}_r$, we have simulated the mean number of points and neighbours using (10) with parameters $(r, \beta(s), \gamma(s))$. Then we have plotted $\hat{\mathbb{E}n} - 617$ and $(\hat{\mathbb{E}s} - s)/s$ to show the (relative) deviation. We see, that both plots spread around 0. For small values of r we see larger deviation of the mean number of points, which is expected since s is very small. Still, the deviation is less than 15 percent. All in all, the estimates are reliable. We have also tried to log-transform the input and response, but the untransformed data gave the finest overall result.

Step 4: The inner-plane ratio

For the cell data, (7) is fulfilled, see Figure 4, so $T^*(r) = T(r)$ for all $r \in \mathbb{G}_r$.

For each $r \in \mathbb{G}_r$, the inner-plane ratios for the data has been calculated for each $\theta \in T(r)$. As explained earlier, attention can now be restricted to the set (9) since only that value $\hat{\theta}(r) \in T(r)$ which maximizes the inner-plane ratio is a possible candidate for the MLE. In Figure 8 (a) we see a plot of the maximal inner-plane ratios joined by full lines. The dotted lines is a plot of the corresponding ratios $\log(L_0(r, \hat{\theta}(r))/L_0(r, \hat{\theta}_0))$. Notice that the two curves are 0 only simultaneously and this is in the cases when $\hat{\theta}_0$ is the maximum: Given $r \in \mathbb{G}_r$. Then,

$$\hat{\theta}(r) = \hat{\theta}_0 \iff \log \frac{L_0(\hat{\theta}(r); y)}{L_0(\hat{\theta}_0; y)} = 0 \iff \log \frac{\overline{L}(r, \hat{\theta}(r); y)}{\overline{L}(r, \hat{\theta}_0; y)} = 0$$

$$\hat{\theta}(r) \neq \hat{\theta}_0 \iff \log \frac{L_0(\hat{\theta}(r); y)}{L_0(\hat{\theta}_0; y)} < 0 \iff \log \frac{\overline{L}(r, \hat{\theta}(r); y)}{\overline{L}(r, \hat{\theta}_0; y)} > 0$$

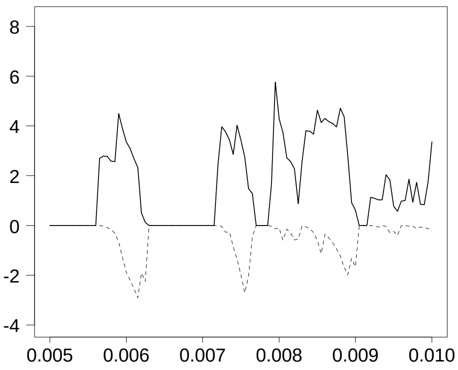
Step 5: The between-planes ratio

Let $r_0 = 0.007$ and

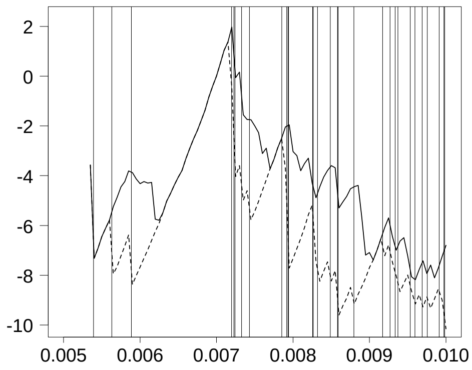
$$\mathbb{G}_r = \{r_{-40}, r_{-39}, \ldots, r_{-1}, r_0, r_1, \ldots, r_{59}, r_{60}\}.$$

For any r_i, $i \geq 0$ and any $\theta^{(t)} \in T(r_i)$, we compute the profile likelihood ratio using (6). If $i < 0$, the second sum in (6) is changed to $-\sum_{j=i+1}^{0}$.

In Figure 8 (b) the between-planes ratios are plotted as the dotted lines. The profile likelihood ratio in $(r, \hat{\theta}(r))$ is then the sum of the inner-plane and between-planes ratios, see (6). The profile likelihood ratios are plotted as full lines in Figure 8 (b).



(a) Inner-plane (—) and $\log L_0$ (- -)



(b) Full (—) and partial (- -)

Figure 8: For each $r \in \mathbb{G}_r$, the indicated loglikelihood ratios have been evaluated, see (6). Only the maximal inner-plane ratios are plotted in (a) and the dotted lines are the ratios $\log(L_0(r,\theta)/L_0(r,\hat{\theta}_0))$ for the corresponding value of θ. Note that the between-planes ratio equals the partial likelihood.

The vertical lines marks the inter-point distances $\{d_1, d_2, \ldots, d_m\}$. Note that the between-planes ratio is increasing (and continuous) between two such values and have a down-wards jump crossing the lines, see Section 1.6.

Furthermore, notice that in the interval $(0; d_1]$ up to the first line marking d_1, only the last between-planes ratios has been calculated. The other between-planes ratios in this interval are all smaller, and so are the profile log-likelihood ratios since the inner-plane ratios are all 0.

Step 6: The final maximization

The profile likelihood (3) is maximized in $r = 0.0072$. Thereby the full MLE is

$$(\hat{\beta}, \hat{\gamma}, \hat{r}, \hat{\theta}_1, \hat{\theta}_2) = (767.6, 0.08149, 0.00720, 1.304, -0.275)$$

Since $\theta^{(0)} = \hat{\theta}_0$, we also have the partial profile likelihood where θ is fixed to $\hat{\theta}_0$. This is namely the between-planes ratios. Thus, the partial MLE is

$$(\hat{\beta}, \hat{\gamma}, \hat{r}, \hat{\theta}_{01}, \hat{\theta}_{02}) = (766.0, 0.08398, 0.00715, 1.304, -0.272).$$

The between-planes ratio, is increasing on intervals $(d_i; d_{i+1}]$. Thus, the MLE for the partial analysis is explicitly known, $\hat{r} = d_i = 0.00719678$, since this is the only d_i between 0.00715 and the next grid-point 0.00720.

Test for homogeneity

From the way data is sampled, we believe that $\theta_2 = 0$, and will now test this. The test statistic can be calculated as 2(A − B − C), where

$$\mathrm{A} = \log \frac{\overline{L}(\hat{r}, (\hat{\theta}_1, \hat{\theta}_2); y)}{\overline{L}(r_0, (\hat{\theta}_{01}, \hat{\theta}_{02}); y)}, \quad \mathrm{B} = \log \frac{\overline{L}(\hat{r}, (\hat{\theta}_1, 0); y)}{\overline{L}(r_0, (\hat{\theta}_{01}, 0); y)},$$
$$\mathrm{C} = \log \frac{\overline{L}(r_0, (\hat{\theta}_{01}, 0); y)}{\overline{L}(r_0, (\hat{\theta}_{01}, \hat{\theta}_{02}); y)}.$$

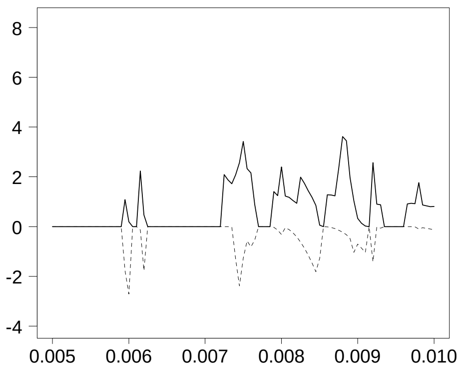
We have A=1.97957. This is the maximum value of the full profile likelihood from Figure 8 (b). To get the value B, we make the full analysis for $\theta_2 = 0$ fixed. In Figure 9 the resulting likelihood ratios are plotted. The full MLE becomes

$$(\hat{\beta}, \hat{\gamma}, \hat{r}, \hat{\theta}_1) = (767.0, 0.10895, 0.00725, 1.315)$$

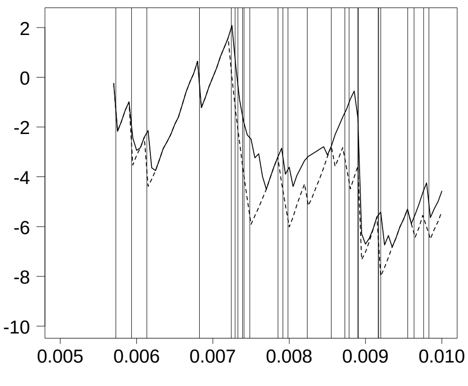
and the partial MLE

$$(\hat{\beta}, \hat{\gamma}, \hat{r}, \hat{\theta}_{01}) = (765.1, 0.10969, 0.00720, 1.304).$$

Again we can find the exact partial MLE as the smallest inter-point distance for the point pattern $h^{-1}_{(\hat{\theta}_{01},0)}(y)$ larger than 0.00720. We get $\hat{r} = 0.00724218$. Furthermore, B=2.08809.



(a) Inner-plane (—) and log L_0 (- -)



(b) Full (—) and partial (- -)

Figure 9: The analysis where $\theta_2 = 0$. In (a) we have L_0 ratios and the inner-plane ratios. In (b) the profile likelihood ratios for the full analysis and the partial analysis $\theta_1 = \hat{\theta}_{01}$ are plotted. See Figure 8 for more detail.

The value of C can be calculated as an inner-plane ratio, $r = r_0$, with only one term in the sum since $s_{r_0}(h_{\hat{\theta}_0}^{-1}(y)) = 3$ and $s_{r_0}(h_{(\hat{\theta}_{01},0)}^{-1}(y)) = 4$. We get C$= -3.72176$.

Thus, the test statistic is $2(\mathrm{A}-\mathrm{B}-\mathrm{C}) = 7.23$. Evaluating the test statistic in a $\chi^2(1)$ distribution, the test probability is 0.72%. From a theoretically point of view we had expected that the tests for $\theta_2 = 0$ could be accepted. The reason we get such a low test probability probably because the number of data points is very large.

Remarks on the practical part

Technical details
All programs are coded in `C++` using GNU CC version 2.96 compiler. The runtimes (elapsed times) mentioned below are CPU times measured on a machine equipped with a 1.2 GHz AMD Athlon Thunderbird processor and 256 Mb RAM running under Red Hat Linux 7.1.

Step 1 (1276 minutes ≈ 21 hours)
For each grid point $(r, \theta_1, \theta_2) \in \mathbb{G}_r \times \mathbb{G}_1 \times \mathbb{G}_2$ we had to calculate $s_r(h_\theta^{-1}(y))$, the number of neighbours in the back-transformed cell point pattern. Here the size of G_r made almost no impact, since the number of neighbours could be counted simultaneously for all r. For each θ we computed the distance d between every pair of points in the back-transformed point pattern. Then if $r > d$, the corresponding $r \in \mathbb{G}_r$ were counted 1 up. Thus, the run-time was proportional to the dimension of the grid $\mathbb{G}_1 \times \mathbb{G}_2$. We chose however to make this grid very fine, as this grid was to be used in this step only. The dimension of the grid was 801×801. In average it took 0.12 second per θ grid point.

One value of $(\hat{\mathbb{E}n}, \hat{\mathbb{E}s})$
In steps 2–5 the computer intensive work concerned estimating means either to solve the likelihood equations (2) or to compute normalising constants (4). The estimation of means are based on sample averages, see e.g. (10). We used $m = 200$, and simulated the point patterns using a Metropolis-Hastings birth-death algorithm. The point patterns were sampled from the same Markov chain with spacing 1000 and initial burn-in 10000, see e.g. Geyer (1999) and Møller (1999). One such mean estimation will henceforth be called a run.

Every run take more or less same time for varying parameter values, since the point patterns all contain around 617 points, and since we don't need to worry about running short on RAM. One run took in average 76.541 seconds. In step 5 we also had to count the number of neighbours under a different value of r for each of the 200 point patterns. This added 22.138 seconds to each run. All other operations had no considerable impact on the runtimes. Hence, in steps 2 to 5, the runtimes of the programs are approximately proportional to the number of runs (times we have to estimate a mean),

$$\text{estimated runtime in minutes} = a * \text{no. of runs}, \tag{13}$$

where $a = 76.541/60$ in steps 2–4 and $a = 98.679/60$ in step 5. In Table 1 the estimated runtimes based on (13) are given for the steps 2 to 5 together with the measured runtimes.

Step no.	no. of runs	estimated time	measured time
2	10397	13263	n.a.
3	271	346	348
4	170	217	216
5	93	153	153

Table 1: Estimated and measured runtimes in minutes.

Step 2 (9.2 days)
This step was the real time-killer, both concerning human time and computer run-time. It took really a long time to fiddle around to find the grid midpoints (10). Then, for each of the 101 $r \in \mathbb{G}_r$ and each of the 101-129 grid points in $\mathbb{G}_\beta^r \times \mathbb{G}_\gamma^r$, in all 10397 different parameter values, we had to estimate $(\hat{\mathbb{E}} n, \hat{\mathbb{E}} s)$.

No measured run-time available since the programs were run parallel on machines of different types.

Step 3 (5.8 hours)
The multiple regression for all $r \in \mathbb{G}_r$ took 0.08 second. Second part of step 3 is to check all 271 estimates. This is the runtime given in Table 1.

Step 4 (3.6 hours)
In this step we estimate normalising constant ratios. For each $r \in \mathbb{G}_r$ we make ratios between the different parameter-values in $S(r)$. Thus, there are 271-101=170 grid points in which to estimate.

Step 5 (2.7 hours)
As step 4. In all there are 100 possible ratios $\Delta(j, j-1)$, but 7 of them does not need to be computed, cf. Section 1.6. Thus, this step involves 93 runs.

The tests

First we made a full analysis in the case where $\theta_2 = 0$ was fixed. In this case step 1 took 95 seconds. Notice that the size of the θ–grid is now 801, where it was 801^2 before, and 1276 minutes divided by 801 is 95.6 seconds. The grid $\mathbb{G}$ contains now 193 different values, and (7) is still fulfilled. Step 2 was not run, the results simply reused. Step 3 took 0.07 second. Step 4 was estimated to take $93 * 76.541/60 = 119$ minutes and took 119 minutes. Step 5 involved 86 runs (14 ratios dismissed) and was estimated to, and took, 141 minutes. To calculate the last missing loglikelihood ratio C involved 1 run and took 76.04 seconds.

Comments on test probabilities

We could have tried to simulate the test probabilities. Then we should have simulated, say, 100 point patterns and run them through the same estimation procedure as the observed point pattern in the case where $\theta_2 = 0$ is fixed. Step 1 through 3 takes no considerable time. Step 4 and 5 took 260 minutes for the cell data. To make step 4 faster, a more narrow grid of r values could be considered. Step 5 could be speeded up since we only need to compute the between-plane ratio in those r grid values where the profile log-likelihood can obtain its maximum. This set consist of the inner-point distances in $h^{-1}_{(\hat{\theta}_{01},0)}(y_i)$, and those grid values where the inner-plane ratio found in step 4 is positive. Then the full maximum likelihood estimation could be reduced to take, say, 200 minutes, and we would have a test probability after 14 days of CPU time. Using parallel programming, the 14 days can be reduced with a factor equal to the number of CPUs, and in a couple of years the machines are probably so powerful that the 14 days are reduced to a couple of hours or even minutes.

Hence, the worrying thing about making inference in the full TIM model, is not the CPU time, but the human time it takes to construct and implement the algorithms. With very powerful machines, many of the considerations, such as reducing the grid of relevant θ–values or to adjust the (β, γ)–grid to r, can of course be skipped. But still it takes some effort to get the programs implemented and to make sure that the results produced are reliable.

In Nielsen and Jensen (2001) we have suggested to use a Poisson based test instead, where the interaction model is not used at all. Such a test is very fast and easy to implement and to compute. However, the distribution of the test statistic is at present not known, and if we want to simulate a test probability this involves finding the MLE for the data based on an interaction model. The advantage is that we use $\hat{\theta}_0$ as estimate for θ, which reduces the complexity of the maximum likelihood estimation considerably since the model is homogeneous.

The exact (simulated) test probability given in Nielsen and Jensen (2001) was based on 2000 realizations from the Strauss model and took 24.9 minutes to compute. Based on realizations from the Poisson model, it took 0.26 seconds to compute a test probability.

Appendix: On estimating r for the cell data

In Nielsen (2000) the cell data set was analysed. Here we used $\hat{\theta}_0$ as estimator for θ and an estimate of the interaction range r was obtained using the J–function and profile pseudo-likelihood analysis for the homogeneous back-transformed point pattern $x = h_{\hat{\theta}_0}^{-1}(y)$. The relevant plots are shown in this appendix. All calculations have been done using Adrian Baddeley and Rolf Turners program `spatstat`, cf. Baddeley and Turner (2000).

J–function

The empty space function F and the nearest neighbour function G, cf. e.g. Diggle (1983), for the back-transformed cell data are shown in Nielsen (2000, Figure 3). These functions are compared with the corresponding functions for the homogeneous Poisson process in order to evaluate the interaction structure of the point pattern in question. On basis of these functions we concluded that there is inhibition between the points in the cell point pattern.

Another non-parametric statistic that can reveal the interaction structure, is the J–function introduced by van Lieshout and Baddeley (1996),

$$J(d) = \frac{1 - G(d)}{1 - F(d)}, \quad d \geq 0.$$

For the Poisson process, $J \equiv 1$ and for clustered and regular point patterns, $J < 1$ and $J > 1$, respectively. In Figure 10 the empirical J–function (Kaplan-Meyer) for the cell data is plotted together with envelopes (pointwise maxima and minima) based on 39 simulated Poisson point patterns. This plot supports that there is inhibition in the back-transformed cell point pattern.

The variance of the J–function increases in d. This effect can be seen in the range of the Poisson envelopes. Therefore, the empirical J–function for the cell data is not to be trusted for large values of d.

From the J–function we get a rough estimate of r. In a model such as the Strauss process, where the interaction is of finite range r, the J–function is increasing in $(0; r)$ and constant for $d > r$. The cell data does not show such behaviour. However, choosing $r = 0.007$, then the theoretical Strauss J–function will increase until $d = r$, and then be constant which in average will represent the empirical J–function. Therefore we choose $r = 0.007$ as a rough estimator of the interaction range in the Strauss model. A plot of the J–function for the Strauss process with parameters $(\beta, \gamma, r) = (760, 0.09, 0.007)$ as found in Nielsen (2000), follows the empirical J–function in $(0; r)$.

Profile pseudo-likelihood

Baddeley and Turner (2000) presents a fast and efficient method to compute the pseudo-likelihood function for a point pattern based on software for generalized linear models. The method requires that the parameters enter into the density in a certain way. For a Strauss model (β, γ) enters as an exponential family parameter, which is all right, but the interaction range r does not enter in a way such that it is covered by the method. Therefore we make profile pseudo-likelihood analysis with r as nuisance parameter. In Figure 11 the maximum value of the pseudo-likelihood has been plotted for the back-transformed cell data for each value of r in a large grid. The corresponding maximum pseudo-likelihood estimates of β and γ are also plotted. In the left hand-side of Figure 11 the pseudo-log-likelihood has been evaluated based on a dummy point grid of size 30×30, and in the right hand-side the size of the dummy grid is 50×50. The accuracy of the pseudo-likelihood depends heavily on the size of the dummy point grid, and the result using the small grid is obviously wrong since the profile pseudo-likelihood is maximized for $\gamma > 1$ which is in conflict with the point pattern being regular. For the larger grid, the profile pseudo-likelihood is maximized in $r = 0.007$, corresponding to $(\beta, \gamma) = (721.2, 0.2312)$.

We did not try an even finer dummy grid, since the computer ran out of space. However, we know that $\gamma < 1$, which corresponds to a small value of

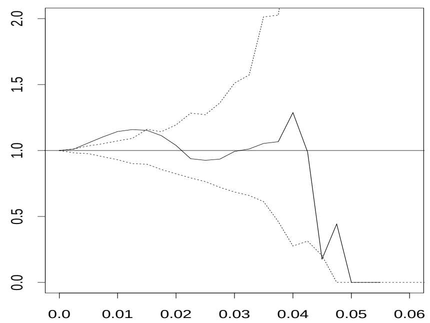


Figure 10: J–function for the back-transformed cell data with envelopes for Poisson process.

F

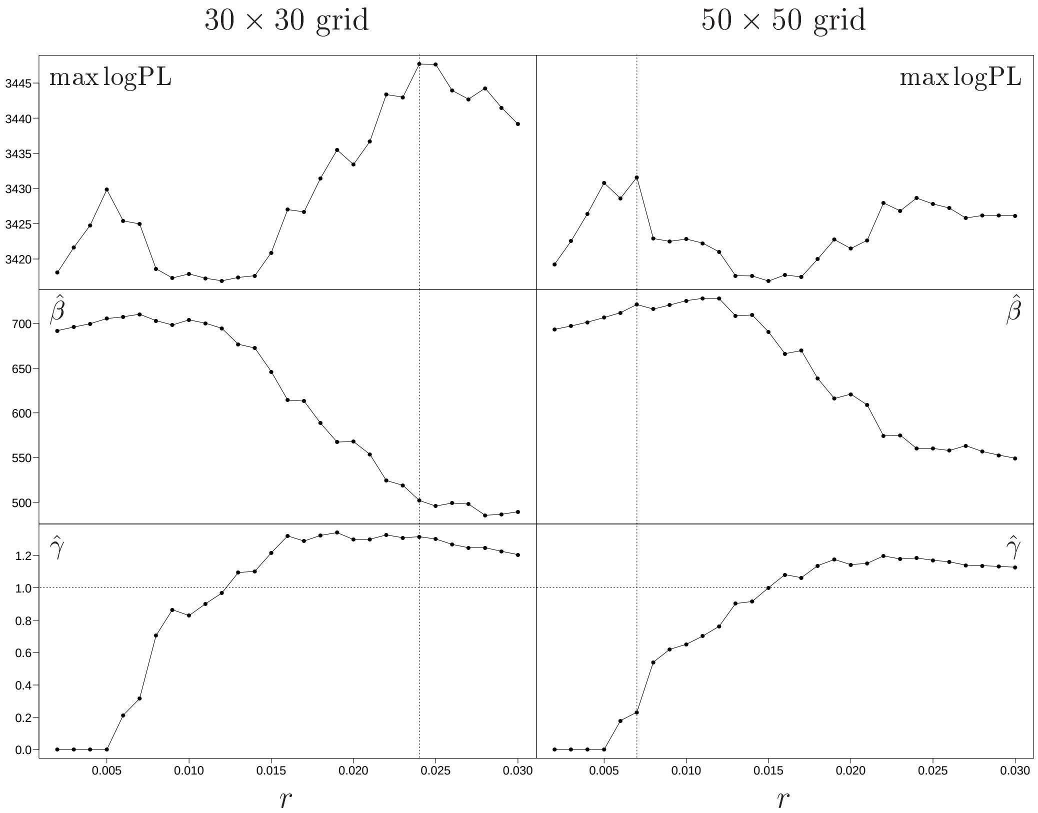


Figure 11: Profile pseudo log-likelihood for the back-transformed cell data for two different dummy point grid sizes. The model is the Strauss model. For various values of r, (β, γ) has been estimated by maximising the pseudo log-likelihood.

r. For small values of r, the value of the pseudo-likelihood does not change very much using a 50×50 grid rather than a 30×30 grid, the error seems to be when r is large. Therefore we tend to believe that the pseudo-likelihood is correct for small values of r, and that $r = 0.007$ therefore is a good estimate of the interaction range.

The estimates of β and γ may however change more using an even finer grid. Baddeley and Turner (2000) also analysed a regular point pattern using the Strauss model. On page 301 they note that in their analysis, a finer dummy grid always led to a smaller value of γ and a larger value of β (for fixed r). We got the same result going from the 30×30 to the 50×50 grid. Thus, we conclude that $\beta > 721.2$, $\gamma < 0.2312$, and $r \approx 0.007$.

Furthermore, Baddeley and Turner (2000, p. 302) note that also edge-correction can be important for the accuracy of the pseudo-likelihood. However, for the back-transformed cell data, the interaction range is very small compared to the window size, and therefore edge-correction might not have large impact in this particular case. However, it could be interesting to investigate this. For the analysis shown in Figure 11 we have used the translate edge-correction, c.f. e.g. Baddeley (1999).

References

Baddeley, A. (1999). Spatial sampling and censoring. In Barndorff-Nielsen, O. E., Kendall, W. S., and van Lieshout, M. N. M., editors, *Stochastic Geometry: Likelihood and Computation*, chapter 2, pages 37–78. Chapman and Hall/CRC, London.

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

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

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.

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

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1994). *Multivariate Analysis.* Academic Press Limited.

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 141–172. 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. and Jensen, E. B. V. (2001). Statistical inference for transformation inhomogeneous point processes. *Research Report 12, Laboratory for Computational Stochastics, University of Aarhus.* Submitted.

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

G

Nielsen L. S. and Jensen, E. B. V. (1999).
On the existence of exponential inhomogeneous Markov point processes. Manuscript.

On the existence of exponential inhomogeneous Markov point processes

Linda Stougaard Nielsen and Eva B. Vedel Jensen
University of Aarhus[1]

Abstract

In the present paper the transformation model for inhomogeneous point processes introduced in Jensen and Nielsen (2000) is examined further. We restrict attention to 2-dimensional manifolds that can be parametrized by a product set. For such sets and for a given type of exponential inhomogeneity the transformation is found in closed form. A criteria for preserving the 'ordering' in a point pattern under the transformation is also presented. As an example we study point processes in a region around a smooth planar curve where the inhomogeneity depends on the distance to the curve.

Keywords: Differential equation; Inhomogeneity; Interaction; Jacobian; Manifold; Parametrization; Relation; Smooth curve; Strauss process; Transformed point process

1. Introduction

Let $\mathcal{X}$ be a k–dimensional differentiable manifold. We are interested in finding a bijective mapping $h_\theta : \mathcal{X} \to \mathcal{X}$, depending on a parameter θ, that transforms a homogeneous point process on $\mathcal{X}$ into an inhomogeneous point process, such that the inhomogeneity is of exponential form. From Jensen and Nielsen (2000, Proposition 3.2) we have that, if X is a point process on $\mathcal{X}$ with density f_X, then $h_\theta(X)$ is a point process on $\mathcal{X}$ with density

$$f_{h_\theta(X)}(x) = f_X(h_\theta^{-1}(x)) \prod_{\eta \in x} Jh_\theta^{-1}(\eta), \quad \text{for } x \in \Omega_{\mathcal{X}}. \tag{1}$$

Here $\Omega_{\mathcal{X}}$ is the set of finite subsets of $\mathcal{X}$ and Jf denotes the Jacobian of the mapping f, see Jensen (1998).

[1]Postal address: Laboratory for Computational Stochastics, Department of Mathematical Sciences, University of Aarhus, Ny Munkegade, DK–8000 Aarhus C, Denmark. Email addresses: lins@imf.au.dk and eva@imf.au.dk.

The first term of the density (1) is the homogeneous density evaluated at the inversely transformed point pattern. The Jacobian term of the density is the term that controls the inhomogeneity of the point process. Thus, we are searching for a mapping h_θ that is a solution of the differential equation

$$Jh_\theta^{-1}(\eta) = \beta(\theta)\mathrm{e}^{\theta\cdot\epsilon(\eta)}. \tag{2}$$

Here the inhomogeneity parameter θ has dimension l, say, and $\epsilon : \mathcal{X} \to \mathbb{R}^l$ is some mapping. Below (2) is called the intensity of the transformed process, although we are aware that this term traditionally has another formal definition. Notice that for fixed $\epsilon(\eta)$ the intensity (2) of the inhomogeneous process is constant. Hence, for $K \in \mathbb{R}^l$, the sets $\{\eta \in \mathcal{X} : \epsilon(\eta) = K\}$ are level curves for the intensity of the inhomogeneous process.

In the next section we will restrict attention to 2–dimensional manifolds that can be parametrized by a product set. In Section 3 we will examine the two examples from Jensen and Nielsen (2000) and see that they both are special cases of the set-up presented in Section 2. In Section 4 we will introduce another more complicated example of a 2–dimensional manifold parametrized by a product set, and for this example discuss solutions of the differential equation (2). This example will reappear several times in this paper. In Section 5 we introduce a class of transformations, the so-called lexicographic transformations. Inside this particular class of transformations it is possible to find a unique solution to the differential equation (2). In Section 6, the example from Section 4 is reviewed in the light of the transformation results presented in Section 5. In Section 7 we discuss the importance of having a transformation that preserves the visual order of the points. A criteria for the preservation of visual order under the lexicographic transformations is presented.

2. Two-dimensional manifolds parametrized by a product set

In this paper we will concentrate on 2–dimensional manifolds that can be parametrized by a product of two intervals. By a rescaling and translation, we can choose the parametrization intervals to be the unit interval $[0;1]$. But in concrete examples, other intervals may be more natural to consider. Hence, let p be a differentiable bijective mapping from a product set of intervals to $\mathcal{X}$,

$$p : \mathcal{I}_1 \times \mathcal{I}_2 \longrightarrow \mathcal{X},$$

where $\mathcal{I}_i = [a_i, b_i]$ for $-\infty < a_i < b_i < \infty$, $i = 1, 2$. Henceforth we shall refer to $p^{-1}(\eta)$ as the geodesic coordinates of a point $\eta \in \mathcal{X}$. The above mapping is a parametrization of $\mathcal{X}$, and the transformation mapping is then of the form

$$h_\theta = p \circ k_\theta \circ p^{-1}, \tag{3}$$

where

$$k_\theta : \mathcal{I}_1 \times \mathcal{I}_2 \longrightarrow \mathcal{I}_1 \times \mathcal{I}_2$$

is the corresponding bijective mapping on the parametrization set. The Jacobian of the transformation mapping can be rewritten in the following manner

$$\begin{aligned} Jh_\theta^{-1}(\eta) &= Jp \circ k_\theta^{-1} \circ p^{-1}(\eta) \\ &= Jp(k_\theta^{-1} \circ p^{-1}(\eta))Jk_\theta^{-1}(p^{-1}(\eta))Jp^{-1}(\eta) \\ &= \frac{Jp(k_\theta^{-1} \circ p^{-1}(\eta))}{Jp(p^{-1}(\eta))} Jk_\theta^{-1}(p^{-1}(\eta)). \end{aligned}$$

Let $\tau = \epsilon \circ p$. The intensity level curves are then, expressed in geodesic coordinates, of the form

$$\{(\omega_1, \omega_2) \in \mathcal{I}_1 \times \mathcal{I}_2 : \tau(\omega_1, \omega_2) = K\}, \quad K \in \mathbb{R}^l.$$

The differential equation (2) to be solved becomes a differential equation in k_θ,

$$\frac{Jp(k_\theta^{-1}(\omega_1, \omega_2))}{Jp(\omega_1, \omega_2)} Jk_\theta^{-1}(\omega_1, \omega_2) = \beta(\theta)\mathrm{e}^{\theta \cdot \tau(\omega_1, \omega_2)}. \tag{4}$$

The general aim is to find the set of solutions to the above-mentioned differential equation. This problem will not be solved in the present paper, but we will later introduce a sub-class of transformations for which a unique explicit solution to the differential equation can be found.

3. Two well-known examples

In Jensen and Nielsen (2000) we have already seen two examples of the set-up described in Section 2. Even though they seemed quite different, they share some important characteristics. This will be explored further in the next section, see e.g. Example 4.1. In the following we will shortly present the two examples.

Example 3.1 (The unit square)
In Jensen and Nielsen (2000, Example 4.1), we studied exponential inhomogeneity in the unit square. Here $\mathcal{I}_i = [0;1]$, $i = 1, 2$. The parametrization mapping p was the identity mapping and the inhomogeneity mapping was of the form

$$\tau(\omega_1, \omega_2) = (\tau_1(\omega_1), \tau_2(\omega_2)),$$

where $\tau_i : \mathbb{R} \to \mathbb{R}$. Hence $\theta = (\theta_1, \theta_2)$ was 2–dimensional.

This is indeed a model of the form described in Section 2. The differential equation (4) becomes

$$Jk_\theta^{-1}(\omega_1, \omega_2) = \beta(\theta)\mathrm{e}^{\theta_1\tau_1(\omega_1)}\mathrm{e}^{\theta_2\tau_2(\omega_2)}. \tag{5}$$

Using the following rewriting of the Jacobian

$$\begin{aligned} &Jk_\theta^{-1}(\omega_1, \omega_2) \\ &= \left| \frac{d}{d\omega_1} k_\theta^{-1}(\omega_1, \omega_2)_1 \frac{d}{d\omega_2} k_\theta^{-1}(\omega_1, \omega_2)_2 - \frac{d}{d\omega_1} k_\theta^{-1}(\omega_1, \omega_2)_2 \frac{d}{d\omega_2} k_\theta^{-1}(\omega_1, \omega_2)_1 \right|, \end{aligned} \tag{6}$$

it is easy to see that the differential equation (5) has a unique solution among functions of the form

$$k_\theta^{-1}(\omega_1, \omega_2) = \left(l_{\theta 1}^{-1}(\omega_1), l_{\theta 2}^{-1}(\omega_2)\right),$$

where $l_{\theta i}$, $i = 1, 2$, are increasing bijections on $[0; 1]$, viz.

$$l_{\theta i}^{-1}(\omega_i) = \beta_i(\theta_i) \int_0^{\omega_i} \mathrm{e}^{\theta_i \tau_i(u)} du, \quad i = 1, 2.$$

Here $\beta_1(\theta_1)\beta_2(\theta_2) = \beta(\theta)$, and $\beta_i(\theta_i)$ are chosen such that the mappings $l_{\theta i}(\omega_i)$, $i = 1, 2$, are bijections on $[0; 1]$.

In this example an intensity level curve is a product set. If τ is strictly monotone in each coordinate, then an intensity level curve consists of one single point.

Another way of handling this example is to consider two independent mappings, one in each coordinate. Hence,

$$\tau_1(\omega_1, \omega_2) = \tau_1(\omega_1) \quad \text{and} \quad \tau_2(\omega_1, \omega_2) = \tau_2(\omega_2),$$

which gives the mappings

$$k_{\theta 1}(\omega_1, \omega_2) = (l_{\theta 1}(\omega_1), \omega_2) \quad \text{and} \quad k_{\theta 2}(\omega_1, \omega_2) = (\omega_1, l_{\theta 2}(\omega_2)).$$

For the first mapping an intensity level curve is of the form $S_1 \times \mathcal{I}_2$. Here $S_1 \subseteq \mathcal{I}_1$ is a set where $\tau_1(\omega_1) = K$ for all $\omega_1 \in S_1$ and $\tau_1(\omega_1) \neq K$ for all $\omega_1 \in \mathcal{I}_1 \backslash S_1$. If τ_1 is strictly monotone, then lines parallel to the second coordinate axis are the intensity level curves. The mapping move the points from one line parallel to the second coordinate axis to another along lines parallel to the first coordinate axis.

Similarly for the second mapping. $\square$

$\boxed{\mathbf{G}}$

Example 3.2 (The unit sphere)
The other example that was examined in Jensen and Nielsen (2000) concerned the unit sphere. Here $\mathcal{I}_1 = [0;\pi)$, $\mathcal{I}_2 = [0;2\pi)$ and p was the polar coordinate mapping, see Jensen and Nielsen (2000, Appendix). Furthermore, τ was some function of the first coordinate, $\tau(\omega_1,\omega_2) = \tau(\omega_1)$. Hence, the inhomogeneity parameter θ was 1–dimensional.

Under these circumstances, using (6) and the fact that $Jp(\omega_1,\omega_2) = \sin(\omega_1)$ the differential equation (4) can be formulated as

$$\sin(k_\theta^{-1}(\omega_1,\omega_2)_1)\left|\frac{d}{d\omega_1}k_\theta^{-1}(\omega_1,\omega_2)_1\frac{d}{d\omega_2}k_\theta^{-1}(\omega_1,\omega_2)_2 - \frac{d}{d\omega_2}k_\theta^{-1}(\omega_1,\omega_2)_1\frac{d}{d\omega_1}k_\theta^{-1}(\omega_1,\omega_2)_2\right| = \beta(\theta)\mathrm{e}^{\theta\tau_1(\omega_1)}\sin(\omega_1). \quad (7)$$

The right hand-side of the equality does not depend on ω_2, and therefore the left hand-side can of course not depend on ω_2 either. Let us suppose

$$k_\theta^{-1}(\omega_1,\omega_2) = \left(l_{\theta 1}^{-1}(\omega_1),\omega_2\right).$$

Then the differential equation (7) can be formulated as

$$-\frac{d}{d\omega_1}\cos(l_{\theta 1}^{-1}(\omega_1)) = \beta(\theta)\mathrm{e}^{\theta\tau_1(\omega_1)}\sin(\omega_1).$$

By integrating on both sides of the equality, we get an explicit expression for $\cos(l_{\theta 1}^{-1}(\omega_1))$.

Here the intensity is constant on lines parallel to the second coordinate axis in the parametrization set. Thus, the intensity is constant on the latitudes of the sphere. Notice that the transformation moves the points along the longitudinales from one latitude to another. □

4. Perpendicular transformation in the curve set

One of the features that made the solution of the differential equation in the examples from the previous section relative simple was, that using the parametrization mapping, the problem could be translated into simpler problems where the intensity was constant on lines parallel to one of the coordinate axis. Thus, if this axis was the second axis, then the inhomogeneity function was of the form

$$\tau(\omega_1,\omega_2) = \tau(\omega_1). \quad (8)$$

In both examples a solution of the differential equation (4) was a mapping that moved the points from one line parallel to the second coordinate axis to another along lines parallel to the first coordinate axis. Thus,

$$k_\theta(\omega_1, \omega_2) = (l_\theta(\omega_1), \omega_2), \tag{9}$$

where l_θ is an increasing bijective mapping on $\mathcal{I}_1$.

It is interesting to find out whether such a transformation in general is a solution to the differential equation (4) when the inhomogeneity is of the form (8), or whether the two examples from Section 3 had some nice properties that made things work out fine. Therefore we will in this section take a look at another example.

The set we will study will henceforth be referred to as the curve set. Let $\mathcal{S} \subseteq \mathbb{R}^3$ be a 2–dimensional manifold equipped with a metric. The curve set is defined as the subset $\mathcal{X} \subseteq \mathcal{S}$ consisting of the points lying within distance $a > o$ from a smooth curve $c \subseteq \mathcal{S}$. Let c be defined as

$$c(t) = (c_1(t), c_2(t)), \quad t \in [0; 1].$$

Hence, the curve set is defined by

$$\mathcal{X} = \{\eta \in \mathcal{S} : d(\eta, c) < a\},$$

for some $a > 0$, where $d(\eta, c)$ is the distance from a point $\eta \in \mathcal{S}$ to the curve.

In the following we will restrict attention to planar curves $c \subseteq \mathbb{R}^2$. Then the distance from a point in $\mathbb{R}^2$ to the curve will be defined as follows. For a point $\eta \in \mathbb{R}^2$ there exists at least one point $\xi \in c$ with tangent vector perpendicular to the line through ξ and η. Let $\xi_1, \ldots, \xi_l$ be the points on c with this property, then we define $d(\eta, c) = \min_{i=1,\ldots,l} d_2(\eta, \xi_i)$, where d_2 denotes the Euclidian distance in $\mathbb{R}^2$. Note, that for η close enough to c relative to the curvature of c, $l = 1$.

The planar curve set can be parametrized by the mapping $p : [-a; a] \times [0; 1] \to \mathcal{X}$,

$$\begin{aligned} p(\omega_1, \omega_2) &= c(\omega_2) - \frac{\widehat{c'(\omega_2)}}{||c'(\omega_2)||}\omega_1 \\ &= \left(c_1(\omega_2) + \frac{c_2'(\omega_2)\omega_1}{\sqrt{c_1'(\omega_2)^2 + c_2'(\omega_2)^2}}, c_2(\omega_2) - \frac{c_1'(\omega_2)\omega_1}{\sqrt{c_1'(\omega_2)^2 + c_2'(\omega_2)^2}} \right). \end{aligned}$$

Here $|| \cdot ||$ denotes the length of a vector. The first coordinate ω_1 in the parametrization set is the signed distance from the point $p(\omega_1, \omega_2)$ to the curve, i.e.

$$\omega_1 = sign * d\left(p(\omega_1, \omega_2), c\right),$$

where $sign \in \{-1, 1\}$ indicates on which side of the curve the point lies. When the inhomogeneity is of the form (8), it only depends on the signed distance from the curve. Thereby the intensity is constant on curves with fixed perpendicular distance from the curve.

Suppose that

$$a < \frac{1}{\max_{\omega_2 \in \mathcal{I}_2} |\kappa(\omega_2)|},$$

where κ denotes the curvature of the curve c. Then it can be shown, that each point in $\mathcal{X}$ is perpendicular to exactly one point on c. Hence p is bijective.

Let us derive the explicit form of the differential equation (4) in this setting. The derivatives of the parametrization mapping has the form

$$v_i = \begin{pmatrix} \frac{d}{d\omega_i} p(\omega_1, \omega_2)_1 \\ \frac{d}{d\omega_i} p(\omega_1, \omega_2)_2 \end{pmatrix}, \quad i = 1, 2.$$

It can easily be shown that v_1 and v_2 are perpendicular and that $||v_1|| = 1$. Thus, the Jacobian for the parametrization mapping is

$$Jp(\omega_1, \omega_2) = ||v_2|| = ||c'(\omega_2)|| \left(1 + \kappa(\omega_2)\omega_1\right),$$

and the differential equation (4) is then

$$\beta(\theta) e^{\theta \cdot \tau(\omega_1, \omega_2)} = \frac{||c'(k_\theta^{-1}(\omega_1, \omega_2)_2)||(1 + \kappa(k_\theta^{-1}(\omega_1, \omega_2)_2) k_\theta^{-1}(\omega_1, \omega_2)_1)}{||c'(\omega_2)||(1 + \kappa(\omega_2)\omega_1)} Jk_\theta^{-1}(\omega_1, \omega_2). \quad (10)$$

Before we study the curve set further, we will argue that the unit square and the unit sphere are special cases of the curve set.

Example 4.1 (The unit square and unit sphere)
In Example 3.1 we noted, that the transformation studied here could be divided into two transformations. The set-up from above can be directly transferred to each of these transformations. Let, for the first transformation, the main curve c be the line $x_1 = 1/2$. The lines parallel to the main curve has constant distance to c. The curvature and the tangent vector are constant, so (10) becomes the differential equation $Jk_\theta^{-1}(\omega_1, \omega_2) = \beta(\theta) e^{\theta_1 \tau_1(\omega_1)}$, as we saw in Section 3, see (5).

In Example 3.2 we did not consider a planar curve set but the 2–dimensional sphere in $\mathbb{R}^3$, and therefore the set-up from above does not cover this example. However, the sphere can be regarded as a curve set where the main curve is a latitude infinitely close to the north pole. Hence, each latitude has constant distance to the main curve. As in the unit square example, we have that the curvature and the length of the tangent vector are constant. This does nicely follow the fact that the Jacobian does not depend on the second geodesic coordinate. □

Example 4.2 (Simple perpendicular transformation)
In this example, we will examine whether a transformation of the form (9) can be a solution of the differential equation (10) when the inhomogeneity is of the form (8).

In this situation the differential equation can be rewritten into the following equation

$$(1 + \kappa(\omega_2)\omega_1)\,\beta(\theta)\mathrm{e}^{\theta\cdot\tau(\omega_1)} = \big(1 + \kappa(\omega_2) l_\theta^{-1}(\omega_1)\big)\,\frac{d}{d\omega_1} l_\theta^{-1}(\omega_1).$$

This implies that

$$\kappa(\omega_2) = \frac{\beta(\theta)\mathrm{e}^{\theta\tau(\omega_1)} - \frac{d}{d\omega_1} l_\theta^{-1}(\omega_1)}{l_\theta^{-1}(\omega_1)\frac{d}{d\omega_1} l_\theta^{-1}(u) - \omega_1\beta(\theta)\mathrm{e}^{\theta\tau(\omega_1)}}. \tag{11}$$

But the left hand-side of the equation depends on ω_2 only, while the right hand-side depends on ω_1. This is a contradiction unless both sides are equal to some constant. If the curvature is constant, then the intensity lines are either straight lines or (part of) circles. This was the situation in both cases mentioned in Example 4.1.

So, the differential equation does not have a solution among functions of the form (9) unless the main curve has constant curvature. It is actually quite obvious why this is so. Consider two pieces of the curve with the same length, one with positive curvature and the other with negative. If we transform a homogeneous point process with same intensity in the two areas along perpendicular lines such that the signed distance to the main curve increases, we get an intensity which does not only depend on the signed distance to the main curve because the points in the concave area are spread out, while the points in the convex area are are squeezed together.

The problem with the contradictory equation (11) is not solved by transforming also the second parameter. In this situation, we get an equation of the same form but now with left hand-side $\kappa(d_\theta(\omega_1, \omega_2))$, where d_θ is the transformed second coordinate. If we then integrate with respect to ω_2 on both sides of the equality (11), the right hand-side is equal to 0. The left hand-side is equal to 0 if either $\kappa \equiv 0$, i.e. the curve is a straight line, if κ is an even function, or if $d_\theta(\omega_1, \omega_2)$ does not depend on ω_2. The situation where the curve is a straight line have already been examined, see Example 4.1. The examples where κ is even and the example where d_θ do not depend on the second geodesic coordinate are not interesting in a search for a general solution for any curve c. $\square$

Example 4.3 (Perpendicular transformation)
Let us expand the transformation class from Example 4.2. If l_θ also depends on ω_2, the above equation (11) makes sense. Hence, let

$$k_\theta(\omega_1, \omega_2) = (l_\theta(\omega_1; \omega_2), \omega_2), \tag{12}$$

where $l_\theta(\cdot; \omega_2)$ is an increasing differentiable bijective mapping on $[-a; a]$ for every choice of $\omega_2 \in [0; 1]$. Notice, that the transformation still follows the perpendicular distance lines, but now the transformation along these lines is different for each line. Therefore such a transformation will be called a perpendicular transformation, and the transformation (9) from the previous example will be called a simple perpendicular transformation.

The differential equation becomes

$$(1 + \kappa(\omega_2)\omega_1)\, \beta(\theta) \mathrm{e}^{\theta \cdot \tau(\omega_1)} = \left(1 + \kappa(\omega_2) l_\theta^{-1}(\omega_1; \omega_2)\right) \frac{d}{d\omega_1} l_\theta^{-1}(\omega_1; \omega_2).$$

Integrating on both sides with respect to ω_1 in the interval $[-a; a]$, we get the following expression for the constant $\beta(\theta)$

$$\beta(\theta) = \frac{2a}{\int_{-a}^{a} \mathrm{e}^{\theta\tau(\omega_1)} d\omega_1 + \kappa(\omega_2) \int_{-a}^{a} \omega_1 \mathrm{e}^{\theta\tau(\omega_1)} d\omega_1}.$$

Suppose that κ is not constant. As $\beta(\theta)$ does not depend on the parameter ω_2, we have that

$$\int_{-a}^{a} \omega_1 \mathrm{e}^{\theta\tau(\omega_1)} d\omega_1 \equiv 0.$$

This is true for every choice of a, which implies that τ is an even function,

$$\tau(-\omega_1) = \tau(\omega_1).$$

Thus, the differential equation has a solution among functions of the form (12) only if the inhomogeneity depends on the absolute distance from the main curve, and not as previously assumed, the signed distance.

At first, in the light of the discussion in Example 4.2, it does not seem reasonable that the perpendicular transformation should work out just because we allow the transformation of the first geodesic coordinate to depend on the second coordinate. The problems with points being squeezed in concave areas and being spread out in convex areas are expected to carry over. But the reason is, that the transformation now is allowed to adjust locally. Given a piece of the main curve, then it is easy to show that the two pieces of curve in perpendicular distance K, $0 \leq K < a$, from the main curve has total length equal to two times the length of the corresponding piece of the main curve. Thereby the length of the two level curves with same intensity is always equal to twice the length of the main curve.

The solution of the differential equation (10) will be postponed to Section 6, see Example 6.1. $\square$

In this section we have studied perpendicular transformations. In Example 4.2 we saw that we had to allow the transformation of the first geodesic coordinate to depend on the second one, and in Example 4.3 we saw that only even inhomogeneity functions τ were covered by this approach if the second geodesic parameter stayed untransformed. Hence, if we want to cover other inhomogeneity functions than the even functions, we have to allow the second coordinate to be transformed. This means that the transformation is no longer perpendicular.

5. Lexicographic transformation

With reference to the discussion in the previous section, we will consider a transformation

$$k_\theta(\omega_1, \omega_2) = (l_\theta(\omega_1; d_\theta(\omega_2)), d_\theta(\omega_2)), \tag{13}$$

where $d_\theta : \mathcal{I}_2 \to \mathcal{I}_2$ is an increasing differentiable bijective mapping and $l_\theta(\cdot; \omega_2) : \mathcal{I}_1 \to \mathcal{I}_1$ is an increasing differentiable bijective mapping for all $\omega_2 \in \mathcal{I}_2$. Notice that k_θ is increasing with respect to the lexicographic ordering

$$(\omega_1, \omega_2) \preceq (\tilde{\omega}_1, \tilde{\omega}_2) \Leftrightarrow (\omega_2 < \tilde{\omega}_2) \text{ or } (\omega_2 = \tilde{\omega}_2 \text{ and } \omega_1 \leq \tilde{\omega}_1).$$

For this reason, the transformation is called lexicographic.

Let us now find explicit expressions for the transformations d_θ and l_θ that solves the differential equation (4). We have that

$$k_\theta^{-1}(\omega_1, \omega_2) = (l_\theta^{-1}(\omega_1; \omega_2), d_\theta^{-1}(\omega_2)).$$

Using (6) the Jacobian is

$$Jk_\theta^{-1}(\omega_1, \omega_2) = \frac{d}{d\omega_1} l_\theta^{-1}(\omega_1; \omega_2) \frac{d}{d\omega_2} d_\theta^{-1}(\omega_2).$$

Therefore the differential equation (4) can be rewritten as

$$\frac{Jp(l_\theta^{-1}(\omega_1; \omega_2), d_\theta^{-1}(\omega_2))}{Jp(\omega_1, \omega_2)} \frac{d}{d\omega_2} d_\theta^{-1}(\omega_2) \frac{d}{d\omega_1} l_\theta^{-1}(\omega_1; \omega_2) = \beta(\theta) \mathrm{e}^{\theta \cdot \tau(\omega_1, \omega_2)}. \tag{14}$$

This differential equation is possible to solve. We have the following result.

$\boxed{\mathbf{G}}$

Proposition 5.1 *Let*

$$\psi_\theta(u_2) = \int_{a_2}^{u_2}\int_{a_1}^{b_1} e^{\theta\cdot\tau(\omega_1,\omega_2)} Jp(\omega_1,\omega_2)d\omega_1 d\omega_2, \quad u_2 \in [a_2; b_2],$$

and for $u_2 \in [a_2; b_2]$ fixed, let

$$\varphi_\theta(u_1; u_2) = \int_{a_1}^{u_1} e^{\theta\cdot\tau(\omega_1,u_2)} Jp(\omega_1,u_2)d\omega_1, \quad u_1 \in [a_1; b_1].$$

With the inhomogeneity constant of the form

$$\beta(\theta) = \frac{\psi_0(b_2)}{\psi_\theta(b_2)},$$

there exists a unique solution to the differential equation (14) *in the class of lexicographic transformations. The solution is given by the two coordinate functions in* (13)

$$d_\theta(u_2) = \psi_\theta^{-1}\left(\frac{\psi_\theta(b_2)}{\psi_0(b_2)}\psi_0(u_2)\right) \tag{15}$$

and

$$l_\theta(u_1; d_\theta(u_2)) = \varphi_\theta^{-1}\left(\frac{\varphi_\theta(b_1; d_\theta(u_2))}{\varphi_0(b_1; u_2)}\varphi_0(u_1; u_2)\ ;\ d_\theta(u_2)\right). \tag{16}$$

Remark 5.2
The functions $\psi_\theta(\cdot)$ and $\varphi_\theta(\cdot; u_2)$, $u_2 \in [a_2; b_2]$, are both increasing functions. Furthermore, $\psi_\theta(a_2) = 0$ and $\varphi_\theta(a_1; u_2) = 0$. Thereby both $d_\theta(\cdot)$ and $l_\theta(\cdot; u_2)$ are well-defined increasing bijective mappings on $[a_2; b_2]$ and $[a_1; b_1]$, respectively.

Notice that we have to know both the original and the transformed geodesic second coordinate in order to transform the first geodesic coordinate, see (16). To transform the geodesic coordinates, we just need to derive some values of integrals involving the Jacobian of the parametrization mapping. If we know the Jacobian, this is a simple problem that can be solved using numerical computing.

Finally, notice the similarity between the expressions for d_θ and l_θ in (15) and (16), and that

$$\psi_\theta(u_2) = \int_{a_2}^{u_2} \varphi_\theta(b_1; \omega_2)d\omega_2. \qquad \square$$

Proof of Proposition 5.1 Let $u_1 \in [a_1; b_1]$ and $u_2 \in [a_2; b_2]$. Integrating the above version of the differential equation (14), we get

$$\begin{aligned}
&\int_{a_2}^{u_2}\int_{a_1}^{u_1} \beta(\theta)\mathrm{e}^{\theta\cdot\tau(\omega_1,\omega_2)} Jp(\omega_1,\omega_2)d\omega_1 d\omega_2 \\
&\quad= \int_{a_2}^{u_2}\int_{a_1}^{u_1} Jp(l_\theta^{-1}(\omega_1;\omega_2), d_\theta^{-1}(\omega_2))\frac{d}{d\omega_2}d_\theta^{-1}(\omega_2)\frac{d}{d\omega_1}l_\theta^{-1}(\omega_1;\omega_2)d\omega_1 d\omega_2 \\
&\quad= \int_{a_2}^{d_\theta^{-1}(u_2)}\int_{a_1}^{l_\theta^{-1}(u_1;d_\theta(\omega_2))} Jp(\omega_1,\omega_2)d\omega_1 d\omega_2. \qquad (17)
\end{aligned}$$

Here we have used that because both d_θ and $l_\theta(\cdot;\omega_2)$ are increasing and surjective,

$$d_\theta(a_2) = a_2 \quad \text{and} \quad l_\theta(a_1;\omega_2) = a_1,\ \omega_2 \in [a_2; b_2].$$

For the same reasons we also have

$$d_\theta(b_2) = b_2 \quad \text{and} \quad l_\theta(b_1;\omega_2) = b_1,\ \omega_2 \in [a_2; b_2].$$

Let $u_1 = b_1$, then the equation (17) is of the form

$$\beta(\theta)\psi_\theta(u_2) = \psi_0(d_\theta^{-1}(u_2)). \qquad (18)$$

With $\beta(\theta)$ as in the proposition, $\beta(\theta)\psi_\theta(\cdot)$ and $\psi_0(d_\theta^{-1}(\cdot))$ are increasing surjective mappings from $[a_2; b_2]$ onto $[0; \psi_0(b_2)]$. Hence there exists a unique solution d_θ to (18) with the required properties.

Consider the equation (17) again, and differentiate on both sides with respect to u_2. Then we get the following equation,

$$\beta(\theta)\varphi_\theta(u_1;u_2) = \frac{1}{d'_\theta(d_\theta^{-1}(u_2))}\varphi_0(l_\theta^{-1}(u_1;u_2); d_\theta^{-1}(u_2)). \qquad (19)$$

From this equation we can derive an expression for d'_θ by letting $u_1 = b_1$,

$$d'_\theta(u_2) = \frac{\varphi_0(b_1;u_2)}{\varphi_\theta(b_1; d_\theta(u_2))}\frac{1}{\beta(\theta)}.$$

For each $u_2 \in [a_2; b_2]$, $\beta(\theta)\varphi_\theta(\cdot; u_2)$ and $\varphi_0(l_\theta^{-1}(\cdot;u_2); d_\theta^{-1}(u_2))/d'_\theta(d_\theta^{-1}(u_2))$ are increasing and surjective mappings from $[a_1; b_1]$ onto $[0; \beta(\theta)\varphi_\theta(b_1; u_2)]$ and so there exists a unique solution l_θ.

As the coordinate mappings given by (15) and (16) in the proposition are solutions to the equations (18) and (19), respectively, these coordinate mappings are the unique solution to the differential equation (14). □

$\boxed{\mathbf{G}}$

6. The curve set – revisited

Let us again study the curve set from Section 4 where the inhomogeneity was of the form (8). In Section 4 we discovered, that the transformation of the first geodesic coordinate had to depend on the second geodesic coordinate. Using a transformation of the form (12), we could only cover inhomogeneity functions τ that were even functions. The transformation (12) is a subclass of the class of lexicographic transformations where d_θ is the identity function. In the example below, we will consider the case when τ is an even function, using the lexicographic transformation approach. We will see, that then d_θ will always be the identity mapping.

Example 6.1 (Inhomogeneity in the distance from the main line)
Let τ be an even function, $\tau(-\omega_1) = \tau(\omega_1)$, or equivalently

$$\tau(\omega_1) = \tau(|\omega_1|).$$

Hence, the inhomogeneity depends only on the distance from the main curve.

Using the transformation approach from Section 5, we get

$$\begin{aligned}
\psi_0(u_2) &= 2a \int_0^{u_2} ||c'(\omega_2)||d\omega_2, \\
\psi_\theta(u_2) &= 2 \int_0^{a} \mathrm{e}^{\theta\tau(\omega_1)} d\omega_1 \int_0^{u_2} ||c'(\omega_2)||d\omega_2 \quad \text{for } \theta \neq 0.
\end{aligned}$$

Due to the fact that $d_\theta(1) = 1$, we have from (18) that $\beta(\theta)\psi_\theta(1) = \psi_0(1)$. From the expressions above this implies that $\beta(\theta) \int_0^a \mathrm{e}^{\theta\tau(\omega_1)} d\omega_1 = a$. Hence $\beta(\theta)\psi_\theta(u_2) = \psi_0(u_2)$ for all $u_2 \in [0; 1]$. Hereby d_θ is the identity function. Now we are in the situation from Example 4.3. Notice that we now have shown, that using the transformation approach from Section 5, then τ is even if and only if d_θ is the identity function.

Let us now find the first geodesic transformation. We have that

$$\varphi_\theta(u_1; u_2) = ||c'(u_2)|| \left(\int_{-a}^{u_1} \mathrm{e}^{\theta\tau(|\omega_1|)} d\omega_1 + \kappa(u_2) \int_{-a}^{u_1} \omega_1 \mathrm{e}^{\theta\tau(|\omega_1|)} d\omega_1 \right).$$

This function does only depend on the second coordinate through the curvature and the length of the tangent vector of c in that particular point. When the main curve is either a straight line or (part of) a circle, then the curvature and the length of the tangent vector are constant, and φ_θ does not depend on the second coordinate, and so does neither l_θ. This agrees with the results from Example 4.1 and 4.2.

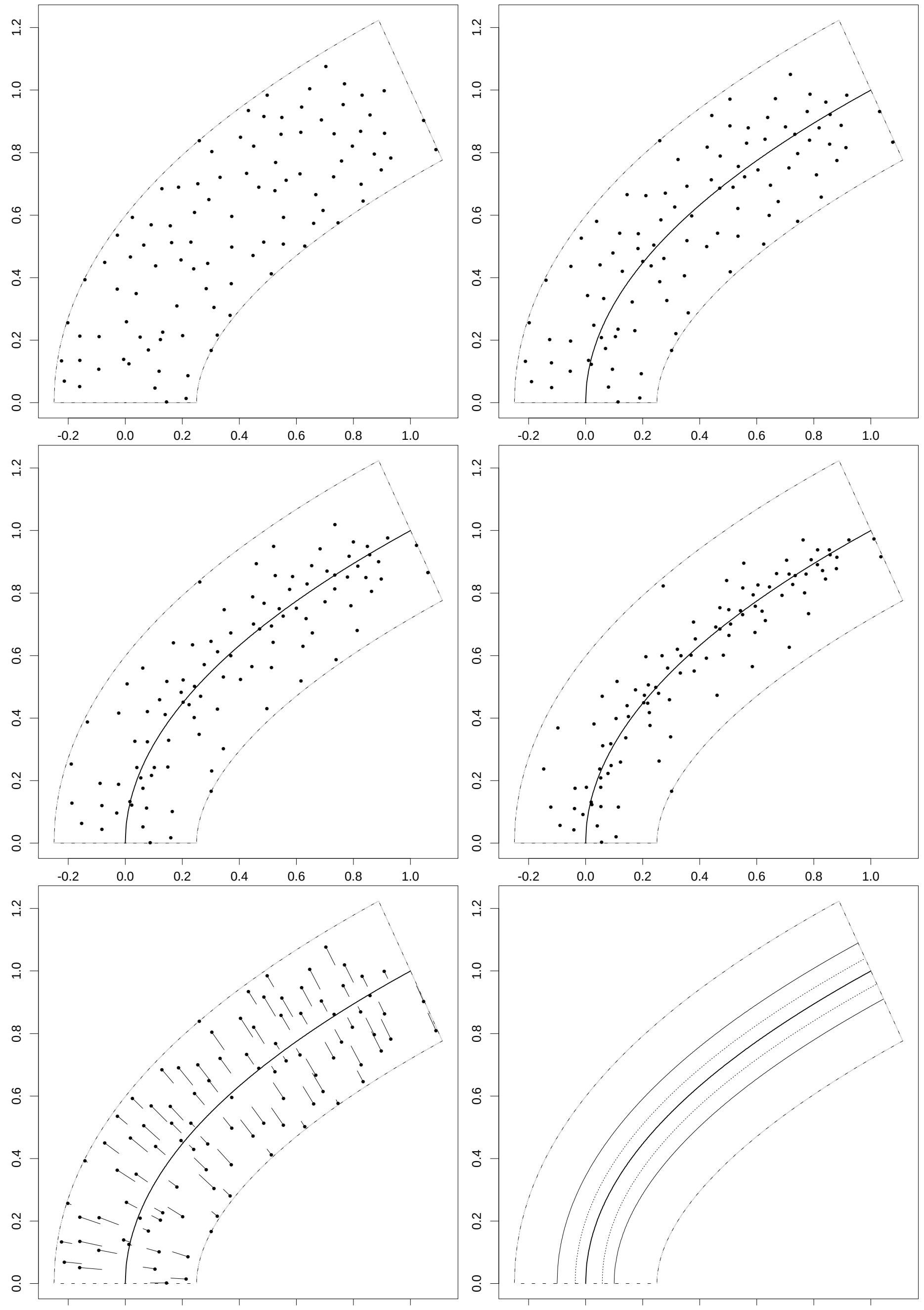


Figure 1: A homogeneous Strauss process in the square-root set with 100 points and parameters $\gamma = 0.01$ and $R = 0.05$ is transformed using the perpendicular transform. The dashed curve (- - -) is the boundary of the square-root set, and the solid line is the main curve. The first plot is the homogeneous process, the second, third and forth plots are the resulting inhomogeneous process with parameters θ =-5,-10 and -20, respectively. In the two last plot the behaviour of the transformation is illustrated for $\theta = -10$. In the fifth plot we see lines indicating to where each point is transformed. In the last plot, the solid curves are curves with the same distance to the main curve, the dotted curves $(\cdots)$ show to where these curves are transformed.

$\boxed{\mathbf{G}}$

Let us in particular consider the curve $c(t) = (t, \sqrt{t})$, $a = 0.25$, and inhomogeneity defined by

$$\tau(\omega_1) = |\omega_1|.$$

Then it is possible to derive an explicit expression for the transformation l_θ. In Figure 1 a realization of a repulsive Strauss process living in this square-root set is being transformed with three different transformation parameters, $\theta = -5, -10, -20$, respectively. In the fifth plot we can see that the transformation is perpendicular. In the last plot it is illustrated that the transformation of the first geodesic coordinate, the signed distance to the main curve, depends on more than the first geodesic coordinate. □

This example shows, as indicated earlier, that with τ not even, then we have to transform the second geodesic coordinate. In this situation ψ_θ and ψ_0 are not proportional, and so we do not have that d_θ is the identity function, which also would be in contradiction with the results in Example 4.3 and 6.1. Let us therefore in the following consider an example where τ is uneven.

Example 6.2 (Inhomogeneity in the distance from the boundary line) In this example we will consider exponential inhomogeneity in the distance from the lower right border-line of the curve set. Such type of one-sided inhomogeneity could for example be relevant if we want to construct a point process model on the bank of a river. Thus, let

$$\tau(\omega_1, \omega_2) = a - \omega_1.$$

In Figure 2 we again consider the square-root introduced in Example 6.1, and we also reuse the point pattern from Figure 1. In the third plot in Figure 2 we see that the transformation is not perpendicular. Notice, that it seems as if the point process preserves the visual ordering of the points. In the two last plots this is clearly illustrated. □

7. Preserving the visual ordering of the points

When transforming point processes with interaction, it would be preferable that the 'ordering' of the points does not change under the transformation. This is expected to be the case for moderate inhomogeneity.

One possibility is to require that the geodesic coordinate transformations $k_{\theta 1}$ and $k_{\theta 2}$ are increasing functions, cf. (3). This is not fulfilled for lexicographic transformations, however.

Such a transformation is increasing in the second geodesic coordinate. The transformation is also increasing in the first geodesic coordinate when the

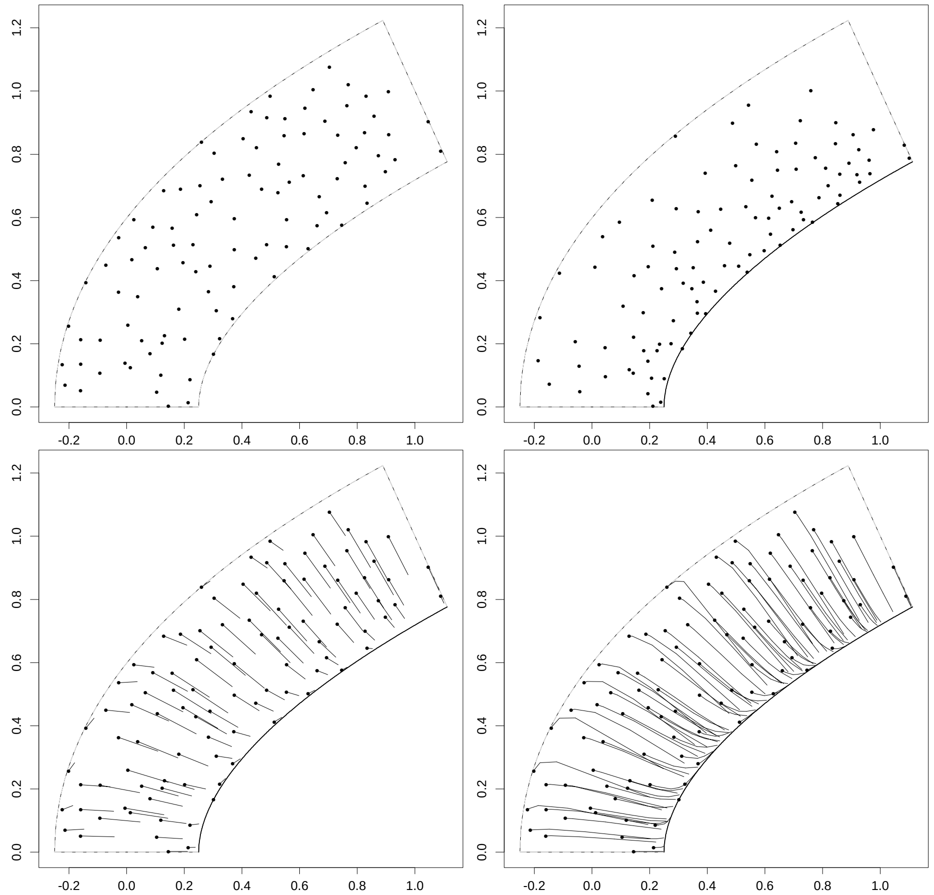


Figure 2: A homogeneous Strauss process in the square-root set is transformed using the lexicographic transform with parameter $\theta = -5$. The dashed curve (- - -) is the boundary of the square-root set, and the solid line is the main curve. The first plot is the homogeneous process, which is the realization from Figure 1, first plot. The second plot is the resulting inhomogeneous process. The third plot shows lines indicating to where each point is transformed. The forth plot is the same as the third plot, here the lines between the transformations with parameters $\theta = -1, -5, -10, -50$ are plotted.

second coordinate is fixed, but in general the first geodesic coordinate is not increasing. However, from a visual inspection of point patterns transformed with a lexicographic transformation, it seems as if the point patterns maintain the ordering of the points anyway, see Figure 1 and 2.

Below we will give another definition of 'ordering' which includes the above definition. The new definition is dependent on the realization instead of the transformation.

For $\omega, \widetilde{\omega} \in \mathcal{I}_1 \times \mathcal{I}_2$, let $R(\omega, \widetilde{\omega}) \subseteq \mathcal{I}_1 \times \mathcal{I}_2$ be the smallest closed rectangle containing ω and $\widetilde{\omega}$. Let $y \in \Omega_{\mathcal{I}_1 \times \mathcal{I}_2}$ be a point pattern. Then $\omega, \widetilde{\omega} \in y$ are

$\boxed{\mathbf{G}}$

rectangular neighbours iff

$$R(\omega,\widetilde{\omega}) \cap y \backslash \{\omega,\widetilde{\omega}\} = \emptyset.$$

A point pattern $x \in \Omega_{\mathcal{X}}$ is said to preserve the order under a transformation of the form (3) if the following condition is fulfilled for $\omega, \widetilde{\omega} \in y = p^{-1}(x)$

$$\begin{array}{l} \omega \text{ and } \widetilde{\omega} \text{ are rectangular neighbours in } y \\ \Updownarrow \\ k_\theta(\omega) \text{ and } k_\theta(\widetilde{\omega}) \text{ are rectangular neighbours in } k_\theta(y). \end{array}$$

The above condition is fulfilled for any point pattern if the transformation is increasing in each of it's coordinates. Below, we give a condition that ensures that this is also true for lexicographic transformation with parameter θ.

Proposition 7.1 *Let*

$$\Delta_\theta = \max_{\omega_1 \in \mathcal{I}_1} \left| \max_{\omega_2 \in \mathcal{I}_2} l_\theta^{-1}(\omega_1;\omega_2) - \min_{\omega_2 \in \mathcal{I}_2} l_\theta^{-1}(\omega_1;\omega_2) \right|.$$

Given two points $\omega, \widetilde{\omega} \in y$. Suppose that

$$|\omega_1 - \widetilde{\omega}_1| > \Delta_\theta. \tag{20}$$

Then, ω and $\widetilde{\omega}$ are rectangular neighbours in y iff $k_\theta(\omega)$ and $k_\theta(\widetilde{\omega})$ are rectangular neighbours in $k_\theta(y)$.

Remark 7.2
If (20) is fulfilled for all the points in y, then the order of the points in x is preserved under the lexicographic transformation with parameter θ. If there is inhibition in the original untransformed point process, then the condition (20) is more likely to be fulfilled for all the points in y.

Consider the band $[\omega_1;\omega_1+\Delta_\theta]\times\mathcal{I}_2$, for $\omega_1 \in [a_1; b_1-\Delta_\theta]$. If a point pattern x has two points with geodesic coordinates within such a band, these two points are allowed to exchange the order of the first geodesic coordinate. As the transformation function l_θ is continuous, the points have almost identical first geodesic coordinate after a transformation if Δ_θ is small relative to a. Hence, if p does not behave too wildly, it is practically impossible to discover visually that some points have exchanged in the original point pattern x, especially if the points lie very far apart in the second geodesic coordinate. So even though the condition (20) is not fulfilled for all the points in y, the order is somewhat preserved under the transformation (3) if Δ_θ is small enough. This will be examined further in Example 7.3. □

Proof of Proposition 7.1 Suppose that ω and $\widetilde{\omega}$ are rectangular neighbours in y. For simplicity, suppose that

$$\omega_1 < \widetilde{\omega}_1 \quad \text{and} \quad \omega_2 < \widetilde{\omega}_2.$$

Since d_θ is increasing, the only thing that can disturb this neighbour relationship under the transformation k_θ is a point $\breve{\omega} \in y$, satisfying

$$\omega_2 < \breve{\omega}_2 < \widetilde{\omega}_2.$$

Let us for simplicity suppose that

$$\widetilde{\omega}_1 < \breve{\omega}_1.$$

We will then show that

$$l_\theta(\widetilde{\omega}_1; d_\theta(\widetilde{\omega}_2)) < l_\theta(\breve{\omega}_1; d_\theta(\breve{\omega}_2)).$$

Thus, suppose that we instead have the relation

$$l_\theta(\widetilde{\omega}_1, d_\theta(\widetilde{\omega}_2)) > l_\theta(\breve{\omega}_1, d_\theta(\breve{\omega}_2)).$$

Then because $l_\theta(\cdot; d_\theta(\breve{\omega}_2))$ is increasing

$$\begin{aligned} \breve{\omega}_1 &\in \left(\widetilde{\omega}_1; l_\theta^{-1}(l_\theta(\widetilde{\omega}_1; d_\theta(\widetilde{\omega}_2)); d_\theta(\breve{\omega}_2))\right) \\ &= \left(l_\theta^{-1}(v; d_\theta(\widetilde{\omega}_2)); l_\theta^{-1}(v; \breve{\omega}_2))\right), \end{aligned}$$

where $v = l_\theta(\widetilde{\omega}_1; d_\theta(\widetilde{\omega}_2))$, and accordingly

$$\breve{\omega}_1 - \widetilde{\omega}_1 < \Delta_\theta,$$

which is a contradiction.

Using the same kind of reasoning, it may easily be shown that if $k_\theta(\omega)$ and $k_\theta(\widetilde{\omega})$ are rectangular neighbours in y, then ω and $\widetilde{\omega}$ are rectangular neighbours in y. □

Example 7.3 (Inhomogeneity in the distance from the border line) Let us again take a look at Example 6.2. In Figure 2 we saw a realization of a transformation of the homogeneous Strauss process in the curve set with the graph of the square-root function as main line. Here there was an striking visual preservation of the order of the points. However, requirement (20) is not fulfilled for all the points in the geodesic point process y. In Table 1 the expression Δ_θ from Proposition 7.1 has been calculated for the transformation

θ	Δ_θ		
1	0.000233921	6	0
5	0.00255696	31	1
10	0.00814604	55	0
50	0.0409686	82	1

Table 1: An examination of the ordering for transformations of the realization of a Strauss process from Figure 2. The first column are the transformation parameters, the second column are the width of the swopping bands given in Proposition 7.1. The third column are the number of pairs of points in the realization that swoppes first geodesic coordinates, and the last column are the number of point-pairs from column three with distance less than $R = 0.05$.

in the square-root set for various values of θ. As we see, when $\theta = -5$ as in Figure 2, then only points with first geodesic coordinate distance less than 0.0026 can change order in the first geodesic coordinate, and this is impossible to discover visually. It is interesting to notice, that most of the points that swop the first geodesic coordinate lie very far away from each other. Actually none or only one pair of points in the original realization in Figure 2 with distance less than R is swopped when $\theta = -1, -5, -10, -50$. $\square$

References

Jensen, E. B. V. (1998). *Local Stereology.* World Scientific, Singapore.

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