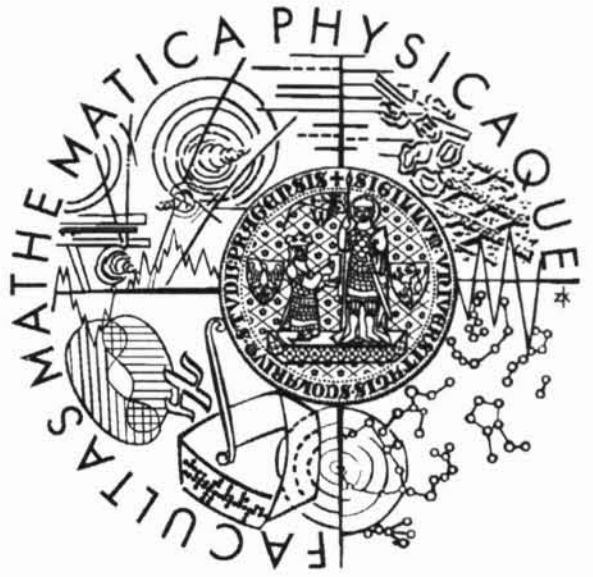
CHARLES UNIVERSITY IN PRAGUE
FACULTY OF MATHEMATICS AND PHYSICS



MODELLING AND STATISTICS OF SPATIAL POINT PROCESSES

DOCTORAL THESIS

Mgr. Michaela Prokešová

Supervisor: Prof. RNDr. Viktor Beneš, DrSc.

Field of study: m-4 Probability and Mathematical Statistics

Prague, 2005

Acknowledgements

I would like to thank my supervisor Professor Viktor Beneš for his help and support during my doctoral study and for careful reading of draft versions of the thesis.

During my PhD. studies I had the opportunity to visit two excellent foreign universities. I am very grateful to Professor Eva B. Vedel Jensen (University of Aarhus), Doctor Ute Hahn and Professor Lothar Heinrich (University of Augsburg) not only for their valuable suggestions and many fruitful discussions throughout the scientific research but also for making my stays abroad enjoyable. I also acknowledge the Marie Curie Fellowship Association and the German Academic Exchange Service DAAD which provided financial support for my foreign stays.

Declaration

I declare that this thesis has been written by myself and that all used literature is included in the list of references. I agree with using this thesis for study purposes.

Prague, 10th July 2005

Mgr. Michaela Prokešová

Contents

Chapter 1

Introduction

Spatial point processes are used to model random data sets consisting of positions in a two– or three– or k– dimensional region (point patterns). For more then 30 years spatial point processes have been a major area of research in spatial statistics, see. e.g. [40], [12], [47] and [30]. They have many applications in such different fields as astronomy, ecology, forestry, image analysis, medicine and epidemiology.

In the thesis we consider three different problems from modelling and statistics of finite inhomogeneous, stationary, and finally spatial-temporal point processes.

The most well known point process model is the Poisson point process with no interaction between the points. Many point patterns show dependence in form of inhibition or clustering (interaction), and in addition a trend in the intensity of the points (inhomogeneity) is often observed. The important class of Markov point processes is very useful in describing interaction [30]. However, statistical tools developed for Markov point processes models have mainly concerned homogeneous models. Four model classes allowing both for inhomogeneity and interaction were suggested and intensively studied in recent years. The models are all based on a homogeneous Markov point process model which is modified in order to introduce a trend. This can be done by introducing a non-constant first-order interaction term [48], [36], by applying a location dependent thinning [2] or by transformation of the homogeneous Markov point process [25], [35].

It turns out that the local geometry of each of these three types of inhomogeneous processes depends on the local intensity of the inhomogeneous process. For example, a transformed point process most often exhibits local anisotropy, depending on local intensity. However when concerning such important point patterns like cells in plant or animal tissues, where both the size and the cell number depend on the distance to the boundary of an organ, plant communities, where number density is governed by environmental conditions and the plants form similar patterns but with varying scale, or modern materials designed with structural inhomogeneity of particles of the same shape and locally varying size, we see that the local geometry (interaction structure) does not change with the location/intensity . Thus all three above mentioned models are inappropriate in such cases. Therefore the fourth alternative model class based on local scaling of the homogeneous template Markov point process has been suggested by Hahn et.al. in [18]. These processes have the property that locally they look like a scaled version of the template process. The local geometry is thus preserved.

Nevertheless statistical properties of this new and very realistic model still needed to be investigated. This is done in the first and largest part of the thesis.

The problem of the inference in locally scaled models is that the standard maximum likelihood estimation is infeasible (which is not an exception in complex point process models) thus an alternative method must be found. We propose a two-step procedure which uses a specific way of factorization of the likelihood. Thus we first estimate the scaling function by maximizing only the Poisson part of the likelihood (i.e. disregarding the interaction). Afterwards we estimate the interaction parameters of the template process with the scaling function assumed to be known and equal to the estimated value. The same two-step procedure can be applied to the maximum pseudolikelihood estimation. For the most feasible parametric case, when the scaling functions form an exponential family, explicit estimation equations for the first step are derived and conditions for existence and uniqueness of the estimate are formulated. Moreover it is shown that the first-step estimates of the scaling function are approximate moment estimates in the original model. Nevertheless because of the essential nonlinearity of the locally scaled point process model it is very hard to deduce further theoretic results. That is why a simulation study was carried out in order to further study the properties of the derived estimators. The results are very encouraging and especially they confirm the important fact that the value of the first-step estimate of the scaling function does not depend on the actual value of the interaction parameters of the template.

After having estimated the parameters of the model the next task is the validation of the model. Since the standard methods and statistics used for this purpose for other point process models are ineffective or even undefined for the locally scaled model, two of the classical methods are modified to suit the locally scaled model. Finally all of the developed methods are applied to the analysis of a real data set – a vegetation point pattern from the Australian desert.

In the second part of the thesis we move our attention from finite point processes to stationary point processes (i.e. point processes defined on whole $\mathbb{R}^k$ and invariant under translation) and from a specific model to a specific quantity σ^2 defined for (almost) any simple stationary point process. σ^2 denotes the asymptotic variance and it is indeed defined as (normalized) asymptotic variance

$$\sigma^2 = \lim_{n\to\infty} |W_n|\, \mathbb{E}(\widehat{\lambda}_n - \lambda)^2,$$

of the standard point process intensity estimator

$$\widehat{\lambda}_n = \frac{\Phi(W_n)}{|W_n|}.$$

Here $\{W_n\}$ is a sequence of sets (observation windows) growing to $\mathbb{R}^k$ and satisfying some regularity conditions and λ is the intensity of the point process Φ.

The quantity σ^2 is important because it is used for the construction of the confidence intervals for the intensity of the stationary point process and also as a normalizing term in goodness-of-fit tests for the K-function (or second moment measure) of stationary point processes (see e.g. [20]).

The asymptotic variance is estimated by a kernel type estimator. A whole class of them was defined in [21]. There was also shown their asymptotic unbiasedness and weak consistency. However neither their asymptotic optimality nor their behaviour on middle size windows was investigated in detail. This is done in the second part of the thesis.

The asymptotic behaviour of the estimators $\widehat{\sigma}_n^2$ of σ^2 can be classified in dependence on the behaviour of the so called second reduced factorial cumulant measure $\gamma_{red}^{(2)}$. If $\gamma_{red}^{(2)}$ has finite support it is possible to relax the conditions from the original theorem from [20] and get a new theorem about the asymptotic unbiasedness and weak consistency of $\widehat{\sigma}_n^2$. As a corollary of the proof we get the mean squared error (MSE) optimal bandwidth for this finite case and also for the two other cases when $|\gamma_{red}^{(2)}(\mathbb{R}^k \setminus B(\mathrm{o}, R))|$ decays exponentially or faster then exponentially as $R \to \infty$.

Moreover for the case of isotropic stationary simple processes we suggest a new estimator of σ^2 which has smaller variability then the general estimator. We show its asymptotic unbiasedness and weak consistency and derive its MSE-optimal bandwidths in dependence on $\gamma_{red}^{(2)}$.

Since the estimators of the quantity σ^2 are very much influenced by correlations among the points of the point process we usually need quite large samples for the asymptotic results to be applicable. To investigate how large is quite large in our case an extensive simulation study is carried out. The behaviour of the estimators of σ^2 with different kernels and different ranges of bandwidths is observed on observation windows of different sizes for several point processes including the reference stationary Poisson point process as well as different cluster and regular processes.

The third part of the thesis deals with temporal and spatial-temporal Cox (or more generally doubly stochastic) point processes. A Cox process Φ can be regarded as a mixture of Poisson processes with different intensities – first a random mechanism generates the (driving) intensity measure Λ and then conditionally on this Λ, Φ is a Poisson process with the intensity measure Λ. We suppose that Λ has a density with respect to the Lebesgue measure and denote it by λ. We consider the problem of filtration — the problem of estimating the unknown intensity function λ after having observed the point process Φ.

The problem of filtration is a classical problem and was studied by many authors e.g. [44], [14], [45], [28], [10], etc. Our situation is specific in that respect that we suppose the driving intensity function λ depends on the realization of an Ornstein-Uhlenbeck type stochastic process X.

Since we have a natural ordering on the time axis and usually we also get the data step by step as the time grows, not all at once, the analysis of the temporal and spatial-temporal processes is different from the general spatial case. Thus we do not derive an explicit formula for the estimator of λ or X on $[0, t)$, but we derive a differential equation for the estimate of $\lambda(t)$ or $X(t)$.

In the temporal case we proceed like in [45] and first find the characteristic form of the differential generator of X. Using this we derive a differential equation for the conditional distribution $\mathbb{P}[X(t) \mid (\Phi\,|_{[0,t)})]$ of X.

In the spatial-temporal case we have to use a different method formulated in [14]. First we derive a differential rule for a special stochastic process which is constructed from both X and Φ and using this rule we derive a differential equation for the conditional mean

$\mathbb{E}[X(t) \mid (\Phi \mid_{[0,t)\times\mathbb{R}^k})]$. Both the results are obtained not only for Cox processes but for the larger class of doubly stochastic analytic point processes (which is a generalization of Cox processes).

The organization of the thesis is as follows. Chapter 2 contains basic notation and several classical notions from geometry, measure and probability theory, as well as an introduction into Lévy processes and the Ornstein-Uhlenbeck type processes derived from them. Chapter 3 gives an overview of the theory of point processes including some basic information from statistics of point processes.

Chapter 4 deals with the development of statistical inference for the recently defined point process model – for the locally scaled point processes. The main results about the two-step estimation procedure are given in Theorem 4.2 (the factorization of the pseudolikelihood), Theorem 4.3 and 4.4 (the estimation equations for the scaling function and the existence and uniqueness conditions for the estimate in the exponential family case). The whole Chapter 4 comes from a joint work with Eva B. Vedel Jensen and Ute Hahn [38].

In Chapter 5 are investigated estimators of the asymptotic variance of stationary point processes. The main results are Theorem 5.5. about asymptotic unbiasedness and weak consistency of the estimator $\widehat{\sigma}_n^2$ in case of $\gamma_{red}^{(2)}$ with finite support and Lemma 5.7 which gives the MSE-optimal bandwidths for $\widehat{\sigma}_n^2$ in dependence on the behaviour of $\gamma_{red}^{(2)}$. Corresponding results for the newly defined isotropic estimator are given in Theorem 5.9 and Corollary 5.10. The chapter is closed with an extensive simulation study. Chapter 5 was motivated by a joint work with Lothar Heinrich [22].

The last Chapter 6 deals with the problem of filtration in temporal and spatial-temporal Cox processes driven by Ornstein-Uhlenbeck type processes thus extending the results of [45] and [14]. While in those papers the cases of Brownian diffusions and Poisson driven Markov processes were considered, our solution covers also the case where the background driving Lévy process is a jump process with infinite activity (but finite variation). Theorem 6.3 gives the differential equation for the conditional distribution of the driving process X in the temporal case. In Theorem 6.10 is derived the differential equation for the conditional mean value of the driving process X in the spatial-temporal case. Part of the results in Chapter 6 was presented in [8].

The results of this work are part of the grants GAČR 201/03/0946 and MSM 0021620839.

Chapter 2

Basic notation and preliminaries

2.1 Euclidean space and measures

Let $\mathbb{R}^k$ denote the k-dimensional Euclidean space with the Euclidean metric

$$\|x-y\| = \sqrt{(x_1-y_1)^2+\cdots+(x_k-y_k)^2},$$

$x=(x_1,\ldots,x_k)\in\mathbb{R}^k$, $y=(y_1,\ldots,y_k)\in\mathbb{R}^k$ and $\mathbb{R}_+=\{x\in\mathbb{R}\,:\,x>0\}$ the set of positive numbers. For a set $A\subseteq\mathbb{R}^k$ we shall denote $A^c=\mathbb{R}^k\setminus A$ the complement of A, $\operatorname{int} A$ the interior and ∂A the boundary of the set A.

The closed ball with center $x\in\mathbb{R}^k$ and of radius $r\geq 0$ is denoted by

$$B(x,r)=\{y\in\mathbb{R}^k : \|y-x\|\leq r\}.$$

Let $\mathrm{o}=(0,\ldots,0)\in\mathbb{R}^k$ be the origin.

Further for $u,v\in\mathbb{R}^k$ the scalar product is denoted by $\langle u,v\rangle$ and the vector product by $u\times v$.

Let $(E,\mathcal{E})$ be a measurable space. Under a signed measure we understand a σ-additive set function $\mu:\mathcal{E}\to\mathbb{R}\cup\{\infty\}$ satisfying $\mu(\emptyset)=0$. There exists a set $B\in\mathcal{E}$ such that $\mu(A\cap B)\geq 0$ and $\mu(A\setminus B)\leq 0$ for every $A\in\mathcal{E}$, see [27] (Theorem 2.8). The pair (B,B^c) is called a Hahn decomposition for μ. Define a measure μ^+ and a finite measure μ^- by

$$\mu^+(A)=\mu(A\cap B)\quad\text{and}\quad\mu^-(A)=-\mu(A\setminus B),\quad A\in\mathcal{E}.$$

The decomposition (called the Jordan decomposition) of $\mu=\mu^+-\mu^-$ into a positive variation μ^+ and a negative variation μ^- is independent of the choice of the Hahn decomposition. The measure $|\mu|=\mu^++\mu^-$ is the total variation of the signed measure μ.

By δ_x we denote the Dirac delta measure

$$\delta_x(A)=\begin{cases}1 & \text{if } x\in A,\\ 0 & \text{if } x\notin A.\end{cases}$$

We also use the notation $\mathbf{1}_A(x)=\delta_x(A)$ for the indicator function of the set A.

Definition 2.1 (diffuse measure) A measure μ on $(E, \mathcal{E})$ is called diffuse if it has no atoms, i.e. $\mu(\{x\}) = 0$ for all $x \in E$.

Definition 2.2 (measure image) For a measure μ on a measurable space $(E, \mathcal{E})$ and a measurable mapping $f : (E, \mathcal{E}) \to (\tilde{E}, \tilde{\mathcal{E}})$ let the image of the measure μ under the mapping f be a measure μf^{-1} on a measurable space $(\tilde{E}, \tilde{\mathcal{E}})$ defined by $\mu f^{-1}(A) = \mu(f^{-1}(A))$ for $A \in \tilde{\mathcal{E}}$.

If μ is a measure on $(E, \mathcal{E})$ and ν is a measure on $(\tilde{E}, \tilde{\mathcal{E}})$, the symbol $\mu \times \nu$ denotes the product measure on the product space $(E \times \tilde{E}, \mathcal{E} \otimes \tilde{\mathcal{E}})$, where $\mathcal{E} \otimes \tilde{\mathcal{E}}$ is the product σ-algebra.

Let the Borel σ-algebra on the space E be denoted by $\mathcal{B}(E)$ or simply by $\mathcal{B}$ when it is clear on which space it is defined.

Definition 2.3 (support of a measure) For a measure μ on a topological space E with Borel σ-algebra $\mathcal{B}$, we define the support by

$$\operatorname{supp} \mu = \bigcap \{F \text{ closed} : \mu(E \setminus F) = 0\}.$$

The k-dimensional Lebesgue measure of a Borel set B in $\mathbb{R}^k$ will be denoted by $|B|$. The volume of the unit ball $B(\mathrm{o}, 1)$ in $\mathbb{R}^k$ is

$$\omega_k = |B(\mathrm{o}, 1)| = \frac{\pi^{k/2}}{\Gamma(1 + k/2)}.$$

Definition 2.4 (Hausdorff measure) Let $d \in \{0, \ldots, k\}$ be fixed. The Hausdorff d-dimensional measure $\mathcal{H}^d$ in $\mathbb{R}^k$ is defined as

$$\mathcal{H}^d(A) = \lim_{\delta \to 0+} \inf_{A \subset \cup_i G_i, \operatorname{diam} G_i \le \delta} \sum_i \omega_d \left(\frac{\operatorname{diam} G_i}{2} \right)^d,$$

where $\operatorname{diam} G_i$ denotes the diameter of G_i and the infimum is taken over all at most countable coverings of A with (any) sets of diameters less or equal to δ.

The d-dimensional Hausdorff measure is sometimes also called the d-dimensional volume measure. For its basic properties we refer to [13]. In particular, $\mathcal{H}^0(B)$ is the cardinality (number of points) of B and $\mathcal{H}^k(B)$ is the k-dimensional Lebesgue measure of a Borel set B.

2.2 Probability theory and statistics

A measurable mapping X from a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ to a measurable space $(E, \mathcal{E})$ is called a random element of E. The distribution of X is the image of $\mathbb{P}$ under X. Equality in distribution (i.e. equality of distributions) of two random elements will be denoted by $\overset{D}{=}$.

By $\mathbb{E}X$ we denote the expectation of X. If $t(X)$ ($s(X)$ resp.) is some statistic of X and τ (σ resp.) some parameter of the distribution of X then $t(X)$ is called an unbiased estimator of τ if $\mathbb{E}(t(X)) = \tau$. $\frac{t(X)}{s(X)}$ is called a ratio-unbiased estimator of $\frac{\tau}{\sigma}$ if both t and s are unbiased estimators of τ and σ respectively.

The sequence of estimators $\{t_n(X)\}_{n\in\mathbb{N}}$ of τ is called asymptotically unbiased if $\lim_{n\to\infty} \mathbb{E}(t_n(X)) = \tau$. The sequence of estimators is strongly consistent if $t_n(X)$ converges to τ with probability 1. It is weakly consistent if $t_n(X)$ converges to τ in distribution.

For X a random variable or vector taking values in $\mathbb{R}^k$, $k \geq 1$ we denote by ψ its characteristic function

$$\psi(v) = \mathbb{E}[\mathrm{e}^{i\langle v, X\rangle}], \qquad v \in \mathbb{R}^k.$$

Definition 2.5 (probability kernel) Let $(S, \mathcal{S})$ and $(T, \mathcal{T})$ be two measurable spaces. A mapping $\mu : S \times \mathcal{T} \to \mathbb{R}_+$ is called a (probability) kernel from S to $\mathcal{T}$ if

(i) the function $\mu_s B = \mu(B \mid s)$ is $\mathcal{S}$-measurable in $s \in S$ for fixed $B \in \mathcal{T}$,

(ii) the function $\mu_s B = \mu(B \mid s)$ is a (probability) measure in $B \in \mathcal{T}$ for fixed $s \in S$.

An $\mathbb{R}^d$-valued stochastic process $X = \{X(t)\}_{t\geq 0}$ is a family of $\mathbb{R}^d$ valued random variables $X(t, \omega)$ with parameter $t \in [0, \infty)$, defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

Definition 2.6 (stochastically continuous process) A stochastic process $X = \{X(t)\}_{t\geq 0}$ is stochastically continuous (continuous in probability) if for any $t \in [0, \infty)$ and $\epsilon > 0$

$$\mathbb{P}[\|X(t+s) - X(t)\| > \epsilon] \to 0 \qquad \text{as} \quad s \to 0.$$

We will call history $\{\mathcal{F}(t), t \geq 0\}$ a collection of sub-σ-algebras of $\mathcal{A}$ which is non-decreasing (i.e. $\mathcal{F}(t) \subset \mathcal{F}(u), t < u$). A stochastic process $\{X(t), t \geq 0\}$ is called adapted to a history $\mathcal{F}(t)$ if each random variable $X(t)$ is $\mathcal{F}(t)$-measurable.

Definition 2.7 (martingale) Let $X = \{X(t)\}_{t\geq 0}$ be a stochastic process and denote $\{\mathcal{F}(t) = \sigma\{X(s); s \leq t\}, t \geq 0\}$ the history of X. If $\mathbb{E}[X(t) \mid \mathcal{F}(s)] = X(s)$ a.s. for any $s \leq t$, then X is a martingale.

We will need in the sequel the notion of an exponential family of probability distributions and some facts concerning parameter estimation in this case. For proofs and details we refer the reader to the classical book [5].

Definition 2.8 (exponential family) Let $\{\mathbb{P}_\theta\}$ be a family of probability distributions on a common measurable space $(\Omega, \mathcal{A})$ parametrized by $\theta \in \Theta \subseteq \mathbb{R}^d$. Suppose moreover that all the distributions have densities (with respect to some common non-zero measure μ) of the form

$$f_\theta(x) = a(\theta)b(x)\mathrm{e}^{\langle \alpha(\theta), \tau(x)\rangle}, \qquad x \in \Omega, \tag{2.1}$$

where $\tau(x) = \{\tau_1(x), \dots \tau_d(x)\}$ is a vector of some real-valued statistics of x, α some vector function of the parameter θ, b a non-negative function of x and $a(\theta)$ the normalizing constant. If d is the minimal integer for which the representation (2.1) is possible then $\{\mathbb{P}_\theta\}$ is called the d-dimensional exponential family of distributions and (2.1) is the minimal representation.

If $\alpha(\theta) = \theta$ in (2.1) then we have the canonical representation and the function τ is called the canonical statistics, θ the canonical parameter and

$$\{\theta \in \mathbb{R}^d : \int_\Omega f_\theta(x) < \infty\}, \tag{2.2}$$

is called the canonical parameter space. Let us moreover denote by C the common support of $\{\mathbb{P}_\theta\}$ and by S the convex support (i.e. the closed convex hull of C).

Definition 2.9 (regular exponential family) If Θ is open and equal to the set defined by (2.2) for the family $\{\mathbb{P}_\theta\}$ with the canonical representation

$$f_\theta(x) = a(\theta)\, b(x) \mathrm{e}^{\langle \theta, \tau(x) \rangle}, \qquad x \in \Omega,$$

then the exponential family $\{\mathbb{P}_\theta\}$ is called regular.

The maximum likelihood estimation in exponential families amounts to the moment estimation since the normal equations are then

$$\tau_i(y) = \mathbb{E}_\theta \tau_i, \qquad i = 1, \dots d,$$

where $y \in \Omega$ is the observation and $\mathbb{E}_\theta$ the mean value with respect to $\mathbb{P}_\theta$. For regular exponential families we have a well known theorem about existence and uniqueness of the maximum likelihood estimate

Theorem 2.1 *[[5] Corollary 9.6] Suppose $\{\mathbb{P}_\theta\}$ is regular. The maximum likelihood estimate exists if and only if $\tau \in$ int S, and then it is unique. Furthermore, the maximum likelihood estimator $\hat{\theta}$ is the one-to-one mapping of int S onto int Θ whose inverse is m, where $m(\theta) = \mathbb{E}_\theta \tau$.*

Sometimes the family of probability distributions $\{\mathbb{P}_{\theta,\psi}\}$ with parameters θ and ψ is not an exponential family, but it is exponential in θ conditionally on ψ fixed. Then ψ are called nuisance parameters. The estimation in such case is usually done on a grid of ψ values, for each of the values maximizing the profile likelihood (exponential) and then maximizing over the grid.

2.3 Lévy processes and processes of Ornstein-Uhlenbeck type

In this section we review the basic definitions and properties of the Lévy processes and the Ornstein-Uhlenbeck type processes derived from them. The general theory about Lévy

processes was driven from [42] and [43], and the theory about Ornstein-Uhlenbeck type processes derived from Lévy processes and some more specialized results were taken from [7] and [9].

Definition 2.10 (Lévy process) Let $Z = \{Z(t)\}_{t\geq 0}$ be a stochastic process taking values in $\mathbb{R}^d$ whose realizations are right continuous with limits from the left almost surely (rcll). Suppose that Z is stochastically continuous with stationary and independent increments and $Z(0) = 0$ a.s. Then Z is called a Lévy process.

If the increments have Gaussian distribution then we get the well known Brownian motion with continuous sample paths.

Definition 2.11 (cumulant transform) Let X be a random variable taking values in $\mathbb{R}^d$. The cumulant transform $C\{\cdot \ddagger X\}$ of X is defined by

$$C\{v \ddagger X\} = \log E[e^{i\langle v,\, X\rangle}], \qquad v \in \mathbb{R}^d.$$

A very convenient property holds for the cumulant transform of any Lévy process.

Theorem 2.2 *[[43] Theorems 1.1 and 1.3] Let $Z = \{Z(t)\}_{t\geq 0}$ be a Lévy process taking values in $\mathbb{R}^d$. Then for the cumulant transform of Z holds*

$$C\{v \ddagger Z(t)\} = t\, C\{v \ddagger Z(1)\}, \qquad \textit{for any } t \geq 0, v \in \mathbb{R}^d, \tag{2.3}$$

and $Z(1)$ has the Lévy-Khintchin representation

$$C\{v \ddagger Z(1)\} = i\,\langle a,\, v\rangle - \frac{1}{2}\,\langle v,\, Av\rangle + \int_{\mathbb{R}^d} \left(e^{i\langle v,\, x\rangle} - 1 - i\,\langle v,\, x\rangle\, \mathbf{1}\{\|x\| \leq 1\}\right) \mu(\mathrm{d}x), \tag{2.4}$$

where A is a symmetric nonnegative-definite $d \times d$ matrix, $a \in \mathbb{R}^d$, and μ is a measure on $\mathbb{R}^d$ satisfying $\mu(\{0\}) = 0$ and

$$\int_{\mathbb{R}^d} (\|x\|^2 \wedge 1)\mu(\mathrm{d}x) < \infty.$$

The triplet (a, A, μ) is unique.

Definition 2.12 (generating triplet, Lévy measure) We call the triplet (a, A, μ) from the previous theorem the generating triplet, μ is called the Lévy measure of the process $Z(t)$.

Each Lévy process can be decomposed into a deterministic, a Brownian diffusion and a pure jump part as the following theorem shows (it is a special case of Theorem 1.4 from [43]).

Theorem 2.3 *Let $Z(t)$ be a Lévy process defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ taking values in $\mathbb{R}^d$ with the generating triplet (a, A, μ). For any $G \in \mathcal{B}((0,\infty) \times \mathbb{R}^d)$ let $J_Z(G) = J_Z(G, \omega)$ be the number of jumps at time s with height $Z(s, \omega) - Z(s-, \omega)$ such*

that $(s, Z(s,\omega) - Z(s_-,\omega)) \in G$. *Then* $J_Z(G)$ *has Poisson distribution with mean* $\tilde{\mu}(G)$. *If* $G_1, \ldots, G_n$ *are disjoint, then* $J_Z(G_1), \ldots, J_Z(G_n)$ *are independent. We can define, a.s.,*

$$\begin{aligned} Z^1(t,\omega) &= \lim_{\epsilon \to 0} \int_{(0,t]\times\{\epsilon<\|x\|\le 1\}} \{x J_Z(\mathrm{d}(s,x),\omega) - x\tilde{\mu}(\mathrm{d}(s,x))\} \\ &\quad + \int_{(0,t]\times\{\|x\|>1\}} x J_Z(\mathrm{d}(s,x),\omega), \end{aligned} \tag{2.5}$$

where the convergence on the right-hand side is uniform in t *in any finite interval a.s. The process* $\{Z^1(t)\}$ *is a Lévy process with the generating triplet* $(0,0,\mu)$. *Let*

$$Z^2(t,\omega) = Z(t,\omega) - Z^1(t,\omega).$$

Then $\{Z^2(t)\}$ *is a Lévy process continuous in time a.s. with the generating triplet* $(a, A, 0)$. *The processes* $\{Z^1(t)\}$ *and* $\{Z^2(t)\}$ *are independent.*

It follows from the theorem, that every Lévy process can be represented as $Z(t) = Z^\epsilon(t) + R^\epsilon(t)$ such that $Z^\epsilon(t)$ is a process with finite number of jumps and $R^\epsilon(t)$ is a mean-zero square integrable martingale and $\mathrm{Var}(R^\epsilon(t)) \to 0$ as $\epsilon \to 0$ (see [9] Proposition 3.7).

The properties of J_Z in Theorem 2.3 show that it is a Poisson random measure on $R_+ \times \mathbb{R}^d$ with intensity measure $\tilde{\mu}(\mathrm{d}s, \mathrm{d}x) = \mathrm{d}s\mu(\mathrm{d}x)$. Thus for a deterministic measurable function f on $[0,t] \times \mathbb{R}^d$ the integral with respect to J_Z

$$\int_0^t \int_{\mathbb{R}^d} f(s,y) J_Z(\mathrm{d}s, \mathrm{d}y) = \sum_{n, t_n \in [0,t]} f(t_n, y_n),$$

is a stochastic process with jumps $y_n \in \mathbb{R}^d$ at times t_n, where $J_Z = \sum_{n\ge 1} \delta_{(t_n, y_n)}$.

We will consider the case $A = 0$ only. Then $\mu(\mathbb{R}^d) < \infty$ leads to a pure jump process with finitely many jumps on each finite time interval while when $\mu(\mathbb{R}^d) = \infty$ the jump times form a countable dense set in $\mathbb{R}_+$. Neverthless, even when $\mu(\mathbb{R}^d) = \infty$ the trajectories of Z have finite variation if and only if

$$\int_{\|x\|\le 1} \|x\| \mu(\mathrm{d}x) < \infty. \tag{2.6}$$

In the sequel we will always assume $A = 0$ and (2.6) for the process $Z(t)$. When $a = \int_{\|x\|\le 1} x\mu(\mathrm{d}x)$,

$$C\{v \ddagger Z(t)\} = t \int_{\mathbb{R}^d} (e^{i\langle v, x\rangle} - 1)\mu(\mathrm{d}x), \tag{2.7}$$

and

$$Z(t) = \int_{\mathbb{R}^d} \int_0^t x J_Z(\mathrm{d}(s,x)), \tag{2.8}$$

is a pure jump process with finite variation. If moreover $\mu(\mathbb{R}^d) < \infty$ then $Z(t)$ is a compound Poisson process, i.e. it has only finite number of jumps in any bounded interval and piecewise constant trajectories.

In Chapter 6 we will use Lévy and Ornstein-Uhlenbeck type processes as nonnegative intensities for Cox processes. That is why we will suppose in the rest that our Lévy process is of type (2.7) - i.e. with zero drift and zero Gaussian variance, and its Lévy measure is concentrated on $\mathbb{R}_+^d$. Thus the characteristic function of $Z(t)$ can be written as

$$E\left[e^{i\langle v,\, Z(t)\rangle}\right] = \exp\left(t\int_{\mathbb{R}_+^d}(e^{i\langle v,\, x\rangle}-1)\mu(\mathrm{d}x)\right), \tag{2.9}$$

where μ is the Lévy measure of $Z(1)$.

Now we can introduce the Ornstein-Uhlenbeck type (OU type) processes. Suppose first that $d = 1$.

Definition 2.13 (process of Ornstein-Uhlenbeck type, background driving Lévy process) Let $Z(t)$ be a one-dimensional Lévy process, $\gamma > 0$ and consider the stochastic differential equation for $X(t)$, $t \geq 0$

$$\mathrm{d}X(t) = -\gamma X(t)\mathrm{d}t + \mathrm{d}Z(\gamma t). \tag{2.10}$$

The stationary solution of (2.10) is called a process of Ornstein-Uhlenbeck type (OU type process). $Z(t)$ is called the background driving Lévy process (BDLP) for $X(t)$.

To be able to specify the conditions under which (2.10) has a desired solution we need one more definition.

Definition 2.14 (self-decomposable distribution) A random variable Y with characteristic function ψ has a self-decomposable distribution if for all $c \in (0,1)$ there exists a characteristic function ψ_c such that

$$\psi(v) = \psi(cv)\psi_c(v) \qquad \text{for all } v \in \mathbb{R}.$$

Theorem 2.4 *[[7] Theorem 1] Let ψ be the characteristic function of a random variable X. If X is self-decomposable then there is a stationary stochastic process $X(t)$ and a Lévy process $Z(t)$ such that $X(t) \stackrel{D}{=} X$ and*

$$X(t) = e^{-\gamma t}X(0) + \int_0^t e^{-\gamma(t-s)}\mathrm{d}Z(\gamma s), \tag{2.11}$$

for all $\gamma > 0$, thus $X(t)$ satisfies (2.10).

The process (2.11) is a unique stochastically continuous Markov process and it has a modification with right-continuous realizations with left limits. We will always work with this rcll modification of $X(t)$.

We can also start with the BDLP $Z(t)$ – there exists a sufficient condition on $Z(t)$ for the existence of a stationary solution of the equation (2.10). For the general case see e.g. [26] Theorem 3.6.6. Here we discuss the case of purely jump $Z(t)$ with some more details.

Lemma 2.5 *[[7] Lemma 1] Let $Z(t)$ be a Lévy process specified by (2.9) and assume that for its Lévy measure holds*

$$\int_1^\infty \log(x)\mu(\mathrm{d}x) < \infty. \tag{2.12}$$

Then there exists a unique solution of the equation (2.10) and $X(t)$ can be written as (2.11). For the cumulant transform of $X(t)$ holds

$$C\{\zeta \ddagger X(t)\} = \int_0^\infty C\{e^{-s}\zeta \ddagger Z(1)\}\mathrm{d}s. \tag{2.13}$$

If we moreover suppose that μ has a differentiable density w and we define a function u by

$$u(x) = \int_1^\infty w(vx)\mathrm{d}v, \tag{2.14}$$

then u is the Lévy density (density of the Lévy measure) of the marginal distribution of the process $X(t)$ and w can be computed from u by

$$w(x) = -u(x) - xu'(x), \tag{2.15}$$

where $u'(x)$ denotes the derivative of u.

For $d > 1$ let $Z(t)$ be the d-dimensional Lévy process with characteristic function given by (2.9) and suppose for simplicity that μ has the density w with respect to d-dimensional Lebesgue measure and denote by $w_i(x_i)$ the ith marginal of w, i.e.

$$w_i(x_i) = \int_{\mathbb{R}_+^{(d-1)}} w(x)\mathrm{d}x_1 \ldots \mathrm{d}x_{i-1}\mathrm{d}x_{i+1} \ldots \mathrm{d}x_d.$$

If each w_i satisfies condition (2.12) then we may (on account of Lemma 2.5) define the stationary processes $X_i(t)$ by

$$X_i(t) = e^{-\gamma t}X_i(0) + \int_0^t e^{-\gamma(t-s)}\mathrm{d}Z_i(\gamma s).$$

The vector process $X(t) = (X_1(t), \ldots, X_d(t))$ is then the solution of the vector stochastic differential equation

$$\mathrm{d}X(t) = -\gamma X(t) + \mathrm{d}Z(\gamma t), \tag{2.16}$$

where $\gamma > 0$, and it is a vector OU type process.

Let us end this section with two typical examples of OU type processes. Nevertheless to be able to verify the self-decomposability we will need the following lemma

Lemma 2.6 *[[9] Proposition 15.3] A distribution on $\mathbb{R}$ is self-decomposable if and only if it is infinitely divisible (i.e. for its characteristic function ψ and any $n \in \mathbb{N}$ there exists some characteristic function ψ_n such that $\psi = (\psi_n)^n$) and its Lévy measure μ has a density $u(x) = \frac{k(x)}{\|x\|}$, with k positive function, increasing on $(-\infty, 0)$ and decreasing on $(0, \infty)$.*

Example 2.1 The gamma OU process $X(t)$ has marginal probability density $\Gamma(\alpha, \nu)$, $\alpha > 0$, $\nu > 0$

$$p(x) = \frac{\alpha^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\alpha x}. \tag{2.17}$$

The Lévy density of the gamma distribution is

$$u(x) = \nu \frac{1}{x} e^{-\alpha x}, \qquad x \geq 0,$$

and obviously satisfies Lemma 2.6.
Now using the equation (2.15) from Lemma 2.5 we can evaluate

$$w(x) = \alpha\nu e^{-\alpha x}, \quad x \geq 0, \qquad w(x) = 0, \quad x < 0.$$

□

Example 2.2 The inverse Gaussian (IG) OU type process $X(t)$ has the marginal probability density

$$p\,(x) = \frac{\delta}{\sqrt{2\pi}} e^{-\delta\gamma} x^{-\frac{3}{2}} e^{-\frac{1}{2}(\frac{\delta^2}{x} + \frac{\gamma^2}{x})}, \qquad \delta > 0, \gamma \geq 0, \ x \geq 0$$

and the Lévy density

$$u(x) = \frac{1}{\sqrt{2\pi}} \delta\, x^{-\frac{3}{2}} e^{-\frac{\gamma^2}{2}}, \qquad x \geq 0.$$

Also here the condition of Lemma 2.6 is satisfied.
Using Lemma 2.5 we can compute the Lévy density of the corresponding Lévy process

$$w(x) = \frac{1}{\sqrt{2\pi}} \frac{\delta}{2} \left(\frac{1}{x} + \gamma^2 \right) \frac{1}{\sqrt{x}} e^{-\frac{\gamma^2 x}{2}}, \qquad x \geq 0.$$

□

Chapter 3

Point processes

This chapter contains a summary of basic definitions and theorems concerning point processes. The theory is driven from the books [10], [47] and [30]. Some basic facts about the statistical inference and standard estimation of basic characteristics of point processes are also given — the details can be found in [47], [32] and [30] if not stated otherwise.

3.1 Point processes

Let $(\mathcal{X}, \varrho)$ be a locally compact complete separable metric space. We denote by $\mathcal{B}(\mathcal{X})$ and $\mathcal{B}_0(\mathcal{X})$ the family of Borel subsets and bounded Borel subsets. Sometimes when it does not lead to a confusion we omit the space $\mathcal{X}$ and write shortly $\mathcal{B}$ or $\mathcal{B}_0$. In all the applications we will have either $\mathcal{X} = \mathbb{R}^k$ the k-dimensional Euclidean space or $\mathcal{X} \subset \mathbb{R}^k$.

Definition 3.1 (locally finite measure) A measure μ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is called locally finite if $\mu(B) < \infty$ for each $B \in \mathcal{B}_0$. The space of locally finite measures will be denoted by $\mathcal{M} = \mathcal{M}(\mathcal{X})$. Let $\mathfrak{M} = \mathfrak{M}(\mathcal{X})$ be the smallest σ-algebra on $\mathcal{M}$ which makes the mappings $\mu \mapsto \mu(B)$ measurable for all $B \in \mathcal{B}$. Further, denote by $\mathcal{N} = \mathcal{N}(\mathcal{X})$ the set of locally finite counting measures

$$\mathcal{N} = \{\mu \in \mathcal{M} : \mu(B) \in \mathbb{N} \cup \{0, \infty\},\ B \in \mathcal{B}\},$$

and the σ-algebra on it by $\mathfrak{N}$

$$\mathfrak{N} = \{M \cap \mathcal{N} : M \in \mathfrak{M}\}.$$

Definition 3.2 (random measure) A measurable mapping $\Psi : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathcal{M}, \mathfrak{M})$ is called a random measure. Its distribution $\mathbb{P}\,\Psi^{-1}$ will be denoted by $\mathbb{P}_\Psi$.

Definition 3.3 (point process, simple point process) A point process is defined as a measurable mapping $\Phi : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathcal{N}, \mathfrak{N})$. It has the distribution $\mathbb{P}_\Phi = \mathbb{P}\,\Phi^{-1}$. A point process Φ is called simple if $\mathbb{P}(\Phi \in \mathcal{N}^*) = 1$, where

$$\mathcal{N}^* = \{\varphi \in \mathcal{N} : \varphi(\{x\}) \leq 1 \text{ for all } x \in \mathcal{X}\}.$$

Point process is a random measure such that the support of its realization is almost surely a locally finite set. Let $\mathcal{F}_{lf}(\mathcal{X})$ be the measurable space of locally finite subsets of $\mathcal{X}$ ($\mathfrak{F}_{lf}$ is the smallest σ-algebra which makes the maps $\eta \mapsto \mathcal{H}^0(\eta \cap B)$ measurable for all $B \in \mathcal{B}$). Concerning the simple point processes $\varphi \in \mathcal{N}^*$ there is an isomorphism

$$\mathcal{I} : \varphi \mapsto \operatorname{supp} \varphi \tag{3.1}$$

(see [50] Theorem 3.1.2). Thus a simple point process can be as well regarded as a random locally finite subset of $\mathcal{X}$. In the following we will identify a simple point process with its support and will not distinguish between $x \in \Phi$ and $\Phi(\{x\}) = 1$.

Important characteristics of random measures and point processes are the moment and cumulant measures.

Definition 3.4 (intensity measure, intensity function) For a random measure Ψ a Borel measure Λ on $\mathcal{X}$

$$\Lambda(\cdot) = \mathbb{E}\Psi(\cdot) = \int_{\mathcal{M}} \psi(\cdot)\, \mathbb{P}_\Psi(\mathrm{d}\psi)$$

is called the intensity measure of Ψ.
For a random measure on $\mathbb{R}^d$ such that its intensity measure Λ is absolutely continuous with respect to the Lebesgue measure the derivative

$$\lambda(x) = \frac{\mathrm{d}\Lambda(\cdot)}{\mathrm{d}\,|\cdot|}\,(x),$$

is called the intensity function.

Definition 3.5 (d-th order moment measure) The d-th order moment measure of a random measure Ψ is defined as

$$\mu^{(d)}(\cdot) = \mathbb{E}\Psi^d(\cdot) = \int_{\mathcal{M}} \psi^d(\cdot)\, \mathbb{P}_\Psi(\mathrm{d}\psi),$$

where $\psi^d = \psi \times \cdots \times \psi$. Hence, $\mu^{(d)}$ is the measure on the product space $(\mathcal{X}^d, \mathcal{B}(\mathcal{X})^d)$. Specially $\mu^{(1)} = \Lambda$.

Definition 3.6 (d-th order factorial moment measure) Denote by $\mathcal{X}^{[d]} = \{(x_1, \ldots, x_d) \in \mathcal{X}^d : x_i \neq x_j \text{ for } i \neq j\}$ the set of d-tuples of distinct points from $\mathcal{X}$. The d-th order factorial moment measure of Ψ is defined by the relation

$$\alpha^{(d)}(\cdot) = \mathbb{E}\Psi^{[d]}(\cdot) = \int_{\mathcal{M}} \psi^{[d]}(\cdot)\, \mathbb{P}_\Psi(\mathrm{d}\psi),$$

where $\psi^{[d]} = \psi^d|_{\mathcal{X}^{[d]}}$ is the restriction of the product measure ψ^d to the space $\mathcal{X}^{[d]}$.

Obviously $\alpha^{(1)} = \mu^{(1)} = \Lambda$.

As the name suggests the (factorial) moment measures are connected to the (factorial) moments of the random variables $\Psi(B)$, $B \in \mathcal{B}$. It holds

$$\mu^{(d)}(B^d) = \mathbb{E}\Psi(B)^d, \quad B \in \mathcal{B},$$

and when Φ is a simple point process, also

$$\alpha^{(d)}(B^d) = \mathbb{E}\Phi(B)(\Phi(B)-1)\cdots(\Phi(B)-d+1), \quad B \in \mathcal{B}. \tag{3.2}$$

Thus for the variance of $\Phi(B)$ we have the formula

$$\operatorname{Var}\Phi(B) = \mu^{(2)}(B\times B) - \Lambda(B)^2 = \alpha^{(2)}(B\times B) + \Lambda(B) - \Lambda(B)^2, \quad B \in \mathcal{B}. \tag{3.3}$$

Definition 3.7 (d-th order (factorial) cumulant measure) The d-th order cumulant measure is given by

$$\gamma^d(B_1\times\cdots\times B_d) = \sum_{j=1}^{d}(-1)^{j-1}(j-1)!\sum_{\mathcal{T}\in\mathcal{P}_{j\,d}}\prod_{i=1}^{j}\mu^{(\mathcal{H}^0(S_i(\mathcal{T})))}(B_{i,1}\times\cdots\times B_{i,\mathcal{H}^0(S_i(\mathcal{T}))}), \tag{3.4}$$

where $B_1,\ldots,B_d \in \mathcal{B}$ and $\mathcal{T} \in \mathcal{P}_{j\,d}$ is the partition of the set $\{1,\ldots,d\}$ into j sets $S_1(\mathcal{T}),\ldots,S_j(\mathcal{T})$. When we replace in the equation the moment measures $\mu^{(\mathcal{H}^0(S_i(\mathcal{T})))}$ by the factorial moment measures $\alpha^{(\mathcal{H}^0(S_i(\mathcal{T})))}$, we get the d-th order factorial cumulant measure $\gamma^{(d)}$.

The first order (factorial) cumulant measure again coincides with the intensity measure. The second order (factorial) cumulant measure is also called the (factorial) covariance measure:

$$\begin{aligned}\gamma^2(A\times B) &= \mu^{(2)}(A\times B) - \Lambda(A)\Lambda(B) = \mathbb{E}\Psi^2(A\times B) - \mathbb{E}\Psi(A)\mathbb{E}\Psi(B),\\ \gamma^{(2)}(A\times B) &= \alpha^{(2)}(A\times B) - \Lambda(A)\Lambda(B) = \mathbb{E}\Phi^{[2]}(A\times B) - \mathbb{E}\Psi(A)\mathbb{E}\Psi(B),\end{aligned}$$

where $A, B \in \mathcal{B}$. We can also express the variance of $\Phi(B)$ using the factorial cumulant measure

$$\operatorname{Var}\Phi(B) = \gamma^{(2)}(B\times B) + \Lambda(B). \tag{3.5}$$

Let us now concern integrals of functions on $\mathcal{X}$ with respect to the random measures and their expression using moment measures.

Definition 3.8 (Campbell measure) Let Ψ be a random measure on $\mathcal{X}$ with intensity measure Λ. The Campbell measure C associated to Ψ is a measure on $\mathcal{X}\times\mathcal{M}$ defined by

$$\int_{\mathcal{X}\times\mathcal{M}} f(x,\psi)C(\mathrm{d}(x,\psi))(\mathrm{d}\psi) = \int_{\mathcal{M}}\int_{\mathcal{X}} f(x,\psi)\psi(\mathrm{d}x)\mathbb{P}_\Psi(\mathrm{d}\psi),$$

where f is an arbitrary non-negative measurable function on $\mathcal{X}\times\mathcal{M}$.

Theorem 3.1 *(Campbell theorem) Let Ψ be a random measure on $\mathcal{X}$ with locally finite intensity Λ. Then*

$$\mathbb{E}\int_{\mathcal{X}} f(x)\Psi(\mathrm{d}x) = \int_{\mathcal{M}}\int_{\mathcal{X}} f(x)\psi(\mathrm{d}x)\mathbb{P}_\Psi = \int_{\mathcal{X}} f(x)\Lambda(\mathrm{d}x),$$

where f is an arbitrary non-negative measurable function on $\mathcal{X}$.

Corollary 3.2 *Let Φ be a simple point process with locally finite d-th order moment measure. For an arbitrary nonnegative measurable function f on $\mathcal{X}^d$ it holds*

$$\mathbb{E} \sum_{x_1,\ldots,x_d \in \Phi} f(x_1,\ldots,x_d) = \int_{\mathcal{X}^d} f(x_1,\ldots,x_d)\, \mu^{(d)}(\mathrm{d}x_1,\ldots,\mathrm{d}x_d)$$

and

$$\mathbb{E} \sum_{x_1,\ldots,x_d \in \Phi}^{\neq} f(x_1,\ldots,x_d) = \int_{\mathcal{X}^{[d]}} f(x_1,\ldots,x_d)\, \alpha^{(d)}(\mathrm{d}x_1,\ldots,\mathrm{d}x_d),$$

where $\sum^{\neq}$ indicates that the sum is taken over d-tuples of distinct points.

From the Campbell theorem we see that the Campbell measure is absolutely continuous with respect to the intensity Λ and this enables the definition of Palm kernels and distributions.

Theorem 3.3 *[[10] Proposition 12.1.IV] Let Ψ be a random measure on $\mathcal{X}$ with locally finite intensity measure Λ. Then there exists a probability kernel (Palm kernel) $x \to \mathbb{P}_x$ from $(\mathcal{X}, \mathcal{B})$ to $\mathcal{M}$ such that*

$$\int_{\mathcal{M}} \int_{\mathcal{X}} f(x,\psi)\psi(\mathrm{d}x)\mathbb{P}_\Psi(\mathrm{d}\psi) = \int_{\mathcal{X}} \int_{\mathcal{M}} f(x,\psi)\mathbb{P}_x(\mathrm{d}\psi)\Lambda(\mathrm{d}x),$$

where f is an arbitrary non-negative measurable function on $\mathcal{X} \times \mathcal{M}$.

Definition 3.9 (Palm distribution) The distribution $\mathbb{P}_x$ from Theorem 3.3 is called the Palm distribution of the random measure Ψ at the point x.

For a point process Φ the Palm distribution can be interpreted as the conditional distribution of the point process under the condition that Φ has a point in x.

Since for the support of the Campbell measure C of a point process Φ it holds $\operatorname{supp} C \subseteq \{(x,\varphi) \in \mathcal{X} \times \mathcal{N} : \varphi(\{x\}) \geq 1\}$ we can define the reduced Campbell measure and reduced Palm distributions.

Definition 3.10 (reduced Campbell measure) Let Φ be a point process on $\mathcal{X}$ with intensity measure Λ. The reduced Campbell measure $C^!$ associated to Φ is a measure on $\mathcal{X} \times \mathcal{N}$ defined by

$$\int_{\mathcal{X}\times\mathcal{N}} f(x,\varphi)C^!(\mathrm{d}(x,\varphi)) = \int_{\mathcal{N}} \int_{\mathcal{X}} f(x,\varphi - \delta_x)\varphi(\mathrm{d}x)\mathbb{P}_\Phi(\mathrm{d}\varphi),$$

where f is an arbitrary non-negative measurable function on $\mathcal{X} \times \mathcal{N}$.

Definition 3.11 (reduced Palm distribution) Let Φ be a point process on $\mathcal{X}$. Then we can define the reduced Palm distribution $\mathbb{P}_x^!$ of Φ by

$$\int_{\mathcal{M}} f(\varphi)\mathbb{P}_x^!(\mathrm{d}\varphi) = \int_{\mathcal{M}} f(\varphi + \delta_x)\mathbb{P}_x(\mathrm{d}\varphi).$$

Under a simple condition (which essentially precludes situations where the behaviour of the point process inside a set B is deterministically controlled by the behaviour outside B) the reduced Campbell measure $C^!$ can be disintegrated also with respect to $\mathbb{P}_\Phi$ and we obtain the Papangelou kernel.

Lemma 3.4 *Let Φ be a point process with distribution $\mathbb{P}_\Phi$ and suppose that for every $B \in \mathcal{B}_0$ holds*

$$\mathbb{P}[\Phi(B) = 0 | \Phi|_{B^c}] > 0 \quad a.s. \tag{3.6}$$

Then for every $B \in \mathcal{B}_0$ the measure $C^!(B \times \cdot)$ is absolutely continuous with respect to $\mathbb{P}_\Phi$.

Theorem 3.5 *[[10] Proposition 14.2.III] Let Φ be a simple point process on $\mathcal{X}$ with locally finite intensity and satisfying condition (3.6) for every $B \in \mathcal{B}_0$. Then there exists a kernel λ from $(\mathcal{M}(\mathcal{X}), \mathfrak{M}(\mathcal{X}), \mathbb{P}_\Phi)$ to $\mathcal{X}$ such that $\lambda(\cdot|\varphi)$ is a locally finite measure for $\mathbb{P}_\Phi$ almost all φ and*

$$\int_{\mathcal{X}\times\mathcal{N}} f(x,\varphi) C^!(\mathrm{d}(x,\varphi)) = \int_{\mathcal{N}} \int_{\mathcal{X}} f(x,\varphi) \lambda(\mathrm{d}x|\varphi) \mathbb{P}_\Phi(\mathrm{d}\varphi),$$

holds for an arbitrary non-negative measurable function f on $\mathcal{X} \times \mathcal{N}$.

Definition 3.12 (Papangelou conditional intensity) The kernel λ from Theorem 3.5 is called the Papangelou kernel or the (Papangelou) conditional intensity of the point process Φ.

In the following we will call the Papangelou conditional intensity also the density $\lambda(x|\varphi)$ of the kernel $\lambda(\mathrm{d}x|\varphi)$ with respect to a background measure on $(\mathcal{X}, \mathcal{B})$ (Lebesgue measure in the case $\mathcal{X} = \mathbb{R}^k$).

For temporal point procesess, i.e. point processes defined on $\mathcal{X} = \mathbb{R}_+$ it is possible to define a completely different notion of conditional intensity.

Definition 3.13 Let $\{\mathcal{F}(t), t \geq 0\}$ be a history, Φ an $\mathcal{F}(t)$-adapted temporal point process. If Φ is a simple point process, we will call a compensator the unique non-decreasing righ-continuous $\mathcal{F}(t)$-adapted process Y such that $\Phi([0,t]) - Y(t)$ is a martingal (see [23] for the existence and uniqueness of Y). Suppose now that there exists an integrable $\mathcal{F}(t)$-adapted process λ^* taking values in $\mathbb{R}_+$ with

$$\int_{t<u} \lambda^*(t)\mathrm{d}t = Y(u), \quad u \in \mathbb{R}_+.$$

Then λ^* is called $\mathcal{F}$-conditional intensity of Φ.

It moreover holds

$$\lambda^*(t) = \lim_{u \to 0} \frac{\mathbb{E}[\Phi([0,t+u]) - \Phi([0,t]) | \mathcal{F}(t)]}{u},$$

for each t a.s. (see [10]).

Let us end this section with the definition of stochastic counting integral. Let Φ be a simple point process on $\mathbb{R}_+ \times \mathbb{R}^k$, $\{\mathcal{F}(t) = \sigma(\Phi|_{[0,t)\times\mathbb{R}^k})\}$ a history of Φ and $g(t,r)$ a vector-valued $\mathcal{F}(t)$-adapted process, whose sample functions are a.s. continuous in r and left continuous in t. The stochastic counting integral with respect to Φ is defined as

$$\int_0^t \int_{\mathbb{R}^k} g(s,r)\,\Phi(\mathrm{d}s,\mathrm{d}r) = \sum_{i=1}^{\Phi([0,t)\times\mathbb{R}^k)} g(t_i,r_i), \tag{3.7}$$

where (t_i, r_i) are the points of Φ.

3.2 Stationary point processes on $\mathbb{R}^k$

In this section we consider the case $\mathcal{X} = \mathbb{R}^k$. Let us denote by t_x the translation

$$t_x : y \to (x+y), \quad x, y \in \mathbb{R}^k$$

and for $\mathcal{D} \subset \mathcal{N}$ let $\mathcal{D}t_x^{-1}$ be defined as $\{\nu t_x^{-1} : \nu \in \mathcal{D}\}$.

Definition 3.14 (stationary point process) A point process Φ on $\mathbb{R}^k$ is called (strictly) stationary if its distribution is invariant under translations, i.e.

$$\mathbb{P}_\Phi(\mathcal{D}t_x^{-1}) = \mathbb{P}_\Phi(\mathcal{D}), \qquad \text{for any} \quad \mathcal{D} \in \mathfrak{N},\ x \in \mathbb{R}^k.$$

The moment measures $\mu^{(d)}$ of a stationary point process Φ are invariant under diagonal shifts, i.e.

$$\mu^{(d)}((B_1 + x) \times \cdots \times (B_d + x)) = \mu^{(d)}(B_1 \times \cdots \times B_d),$$

for any $B_1, \ldots B_d \in \mathcal{B}_0$ and $x \in \mathbb{R}^k$. For the intensity measure it implies that Λ is translation invariant and if it is also locally finite then it must be a multiple of the k-dimensional Lebesgue measure.

Lemma 3.6 *If the stationary point process Φ on $\mathbb{R}^k$ has a locally finite intensity measure Λ then there exists a finite constant $\lambda \geq 0$ such that $\Lambda(\cdot) = \lambda\,|\cdot|$.*

Definition 3.15 (intensity of stationary point process) The constant λ from the previous lemma is called the intensity of the stationary point process Φ.

When we observe a realization of a stationary simple point process Φ on a set W (called usually the observation window) its intensity is usually estimated by

$$\widehat{\lambda} = \frac{\Phi(W)}{|W|}. \tag{3.8}$$

From the Theorem 3.1 follows that $\widehat{\lambda}$ is an unbiased estimator.

The higher order measures $\mu^{(d)}, \alpha^{(d)}, \gamma^d, \gamma^{(d)}$ of a stationary point process can be disintegrated with respect to $\lambda|\cdot|$ and corresponding reduced moment measures are obtained.

We will give the definition for the reduced factorial cumulant measure which will be used frequently in Chapter 5. The other reduced measures are defined analogously.

Definition 3.16 (reduced factorial cumulant measure) The reduced d-th order factorial cumulant measure is a unique signed measure $\gamma_{red}^{(d)}(\cdot)$ on $(\mathbb{R}^k)^{d-1}$ such that

$$\int_{(\mathbb{R}^k)^d} f(x_1,\dots,x_d)\gamma^{(d)}(\mathrm{d}x_1,\dots,\mathrm{d}x_d)$$
$$= \lambda \int_{\mathbb{R}^k}\int_{(\mathbb{R}^k)^{d-1}} f(x, x+y_1,\dots,x+y_{d-1})\,\gamma_{red}^{(d)}(\mathrm{d}y_1,\dots,\mathrm{d}y_{d-1})\,\mathrm{d}x,$$

for any nonnegative measurable function f on $\mathbb{R}^k$.

For the Palm (thus also the reduced Palm) distribution of the stationary point process holds

$$\mathbb{P}_x(\mathcal{D}t_x^{-1}) = \mathbb{P}_0(\mathcal{D}), \qquad \text{for any} \quad \mathcal{D}\in\mathfrak{N},\ x\in\mathbb{R}^k.$$

Moreover

$$\alpha_{red}^{(2)}(B) = \int_B \rho^{(2)}(x)\mathrm{d}x = \int_{\mathcal{N}} \varphi(B)\mathbb{P}_0^!(\mathrm{d}\varphi),$$

for any $B\in\mathcal{B}_0$. Thus $\alpha_{red}^{(2)}(B)$ is the mean number of points in $B\backslash\{\mathrm{o}\}$ under the Palm distribution. In the statistical literature on point processes the K-function is used instead.

Definition 3.17 (K-function of a stationary point process) Let φ be a stationary point process on $\mathbb{R}^k$ with locally finite second moment measure. The K-function is a function defined on $\mathcal{B}_0$ by

$$\lambda K(B) = \int_{\mathcal{N}} \varphi(B)\mathbb{P}_0^!(\mathrm{d}\varphi).$$

Definition 3.18 (isotropic point process) A point process Φ on $\mathbb{R}^k$ is called isotropic if its distribution is invariant under rotations.

For the special class of isotropic stationary point processes $\alpha^{(2)}$ is characterized by $K(r) = K(B(\mathrm{o},r))$ when we know it for all values $r>0$.

Definition 3.19 (pair-correlation function) Let Φ be an isotropic stationary point process with locally finite second moment measure. The pair-correlation function is defined by

$$g(r) = \frac{\mathrm{d}K(r)}{\mathrm{d}r}\,/(k\omega_k r^{k-1}),$$

where ω_k is the volume of $B(\mathrm{o},r)$.

The quantity $\lambda K(B(\mathrm{o},r))$ is the mean number of points of Φ within a sphere of radius r centered in a "typical" point of the process which is not itself counted.

The standard estimator of the K-function for a stationary (anisotropic) point process Φ on $\mathbb{R}^k$ observed on a window W is computed from

$$\widehat{\lambda^2 K(B)}) = \sum_{x,\,y\in\Phi}^{\neq} \frac{\mathbf{1}_B(x-y)\mathbf{1}_W(x)\mathbf{1}_W(y)}{|(W-x)\cap(W-y)|}, \tag{3.9}$$

which is valid for any $B \in \mathcal{B}_0$ such that $|W \cap (W - z)|$ is positive for all $z \in B$. It is an unbiased estimator of the quantity $\lambda^2 K(B))$.

In case of the stationary isotropic point process Φ on $\mathbb{R}^k$ we can use the unbiased estimator of $\lambda^2 K(r)$ given by (see [39], [37])

$$\widehat{\lambda^2 K(r)} = \sum_{x,y \in \Phi}^{\neq} \frac{\mathbf{1}(\|x-y\| \le r)\mathbf{1}_W(x)\mathbf{1}_W(y)}{|W^{(|x-y|)}|} k(x,y), \qquad \text{for} \quad 0 \le r \le r^*, \tag{3.10}$$

where

$$r^* = \sup\{r : |W^{(r)}| > 0\} \qquad \text{and} \qquad W^{(r)} = \{x \in W : \partial(B(x,r)) \cap W \neq \emptyset\},$$

and

$$k(x,y) = \frac{|\partial B(x, \|x-y\|)|}{|\partial B(x, \|x-y\|) \cap W|} \qquad \text{for} \quad x \neq y \in W, \ \|x-y\| \le r. \tag{3.11}$$

For the special case $k = 2$ is $k(x,y) = 2\pi/\alpha(x,y)$ where $\alpha(x,y)$ is the sum of all angles of the arcs in W of a circle with center x and radius $r = \|x-y\|$.

3.3 Examples of point processes

Poisson process

Poisson point process is the fundamental point process. It is also called a completely random process since it exhibits no interaction among the points of the process.

Definition 3.20 (Poisson point process) Let $\Lambda \in \mathcal{M}(\mathcal{X})$. A point process Π on $\mathcal{X}$ is called the Poisson point process with intensity measure Λ if it satisfies the following two conditions

(i) the random variables $\Pi(B_1), \ldots, \Pi(B_n)$ are independent for each $n \in \mathbb{N}$ and pairwise disjoint $B_1, \ldots, B_n \in \mathcal{B}$ satisfying $\Lambda(B_i) < \infty$ for $i = 1, \ldots, n$,

(ii) the random variable $\Pi(B)$ has a Poisson distribution with parameter $\Lambda(B)$ for all $B \in \mathcal{B}$ such that $\Lambda(B) < \infty$.

For a given Λ there exists a unique Poisson process with intensity measure Λ. The process is simple if and only if Λ is diffuse. In $\mathbb{R}^k$ the Poisson point process is stationary if and only if $\Lambda = \lambda|\cdot|$. Also the factorial and cumulant measures of the Poisson point process have simple form:

Proposition 3.7 *The d-th order factorial moment measure and d-th order factorial cumulant measure of the Poisson point process with intensity measure Λ satisfy*

$$\alpha^{(d)} \equiv \Lambda^d, \qquad \gamma^{(d)} \equiv 0 \qquad \textit{for any} \quad d \in \mathbb{N}, \ d \ge 2.$$

In the following we will usually denote the Poisson point process by Π and by Π_1 we will denote the unit rate Poisson process i.e. Poisson process with intensity measure equal to the Lebesgue measure.

Cox process

Cox processes belong to the class of doubly stochastic point processes — they are derived from Poisson point process by making the intensity measure Λ random.

Definition 3.21 (Cox point process) Let Λ be a random measure on $\mathcal{X}$ with distribution $\mathbb{Q}$ on $(\mathcal{M}, \mathfrak{M})$ and P_Λ be the distribution of the Poisson point process with intensity measure Λ. Then a point process Φ with distribution

$$\mathbb{Q}_\Phi(\cdot) = \int \mathbb{P}_\Lambda(\cdot)\, \mathbb{Q}(\mathrm{d}\Lambda),$$

is called a Cox process (or a doubly stochastic Poisson process) with driving measure Λ.

The Cox process is simple if $\mathbb{Q}$ is concentrated on the set of diffuse locally finite measures and it is stationary if the driving measure Λ is.

Proposition 3.8 *The factorial moment measures of a Cox process Φ coincide with the moment measures of its driving measure Λ*

$$\alpha_\Phi^{(d)} = \mathbb{E}\Lambda^{(d)}.$$

In particular, $\mathbb{E}\Phi(\cdot) = \mathbb{E}\Lambda(\cdot)$ *(i.e. the intensity measure of Φ coincides with the intensity measure of Λ) and* $\operatorname{Var}\Phi(\cdot) = \mathbb{E}\Lambda(\cdot) + \operatorname{Var}\Lambda(\cdot)$.

The ratio $\frac{\operatorname{Var}\Phi(B)}{\mathbb{E}\Phi(B)}$ indicates the strength of the dispersion (irregularity) of the point process. For a Poisson point process it is equal to 1. We see from the preceding proposition that the Cox process is overdispersed relative to the Poisson point process since $\frac{\operatorname{Var}\Phi(B)}{\mathbb{E}\Phi(B)} \geq 1$. The equality holds only for Λ deterministic – i.e. when Φ is a Poisson process again.

The Cox processes represent a class of more flexible models that the Poisson point process however still feasible thanks to the simplicity of the formulas for the moment measures.

In the case $\mathcal{X} = \mathbb{R}_+$ for the $\mathcal{F}(t)$-adapted Poisson process the $\mathcal{F}(t)$-conditional intensity (Definition 3.13) equals to the non-random intensity function ($\lambda^* = \lambda$). For the Cox process λ^* depends on the exact form of the history. If the history contains the full information on both the point process and the driving intensity λ then $\lambda^* = \lambda$ (random, cf. [29]).

Cluster processes

Cluster processes represent another frequently used class of point process models — they are derived by replacing the points of a parent point process by clusters of daughter

points. In the special case when the individual clusters have Poisson distribution, cluster processes can be also regarded as Cox processes.

Definition 3.22 (cluster point process) Let Φ_p be a point process on $\mathcal{X}$ and ζ_x, $x \in \mathcal{X}$, finite point processes (i.e. a point process with almost surely finite realizations). Then

$$\Phi(\cdot) = \int_{\mathcal{X}} \zeta_x(\cdot)\, \Phi_p(\mathrm{d}x)$$

is a cluster point process on $\mathcal{X}$.

Suppose now $\mathcal{X} = \mathbb{R}^k$. If we apply homogeneous independent clustering to a Poisson process we get Neyman-Scott processes.

Definition 3.23 (Neyman-Scott point process) The Neyman-Scott process is a cluster process such that Φ_p is a stationary Poisson point process on $\mathbb{R}^k$ and $(t_{-x}\zeta_x, x \in \mathbb{R}^k)$ are independent identically distributed, independent of Φ_p.

Under the conditions of the definition the Neyman-Scott process is also stationary and in case the scattering distribution ζ_x is isotropic so is Φ.

If we denote the distribution of the number of points in the respective cluster ζ_o by $\{p_n\}$, and the mean number of daughter points in a cluster by $\bar{c} = \sum_{n\in\mathbb{N}} n p_n$, and the intensity of the parent Poisson process by λ_p, then the intensity of the cluster process is

$$\lambda = \bar{c}\lambda_p.$$

In the isotropic case the K-function satisfies

$$K(r) = \omega_k r^k + \frac{1}{\bar{c}\lambda_p} \sum_{n=2}^{\infty} p_n n(n-1)\mathbf{F}(r) \qquad \text{for } r \geq 0, \tag{3.12}$$

where $\mathbf{F}$ is the distribution function of the distance between two random points from the same cluster.

Definition 3.24 (Matérn cluster process) Let Φ be a Neyman-Scott process such that $(t_{-x}\zeta_x, x \in \mathbb{R}^d)$ are Poisson point processes with the intensity measure

$$\frac{\mu}{|B(\mathrm{o}, r)|} |\cdot \cap B(\mathrm{o}, r)|, \qquad \mu > 0, \quad r > 0.$$

The point process Φ is called the Matérn cluster process.

Definition 3.25 (Thomas cluster process) Let Φ be a Neyman-Scott process such that $(t_{-x}\zeta_x, x \in \mathbb{R}^d)$ are Poisson point processes with the intensity function

$$\frac{\mu}{\int_{\mathbb{R}^k} f_N(y)\mathrm{d}y} f_N(y), \qquad \mu > 0, \quad y \in \mathbb{R}^k,$$

where f_N denotes the density of a symmetric normal distribution. The point process Φ is called the Matérn cluster process.

Both Matérn cluster process and Thomas cluster process have the mean number of points in the cluster equal to μ and the intensity

$$\lambda = \mu\, \lambda_p.$$

Hard-core processes

A hard-core point process is a point process in which any pair of points is forbidden to be closer than a hard-core distance h. Here we describe two hard-core processes suggested by B. Matérn which can be obtained by dependent thinning of a stationary Poisson point process. Thinned point processes are a special kind of cluster processes where the clusters are either δ_x or 0 (empty cluster – zero random measure a.s.).

Definition 3.26 (Matérn type I hard-core process) Let $h > 0$ be given. The Matérn hard-core process type I is a cluster process derived from a stationary Poisson point process Φ_p on $\mathbb{R}^k$ with clusters

$$\zeta_x = \begin{cases} \delta_x, & \text{if} \quad \Phi_p\left(B(x,h) \setminus \{x\}\right) = 0, \\ 0, & \text{if} \quad \Phi_p\left(B(x,h) \setminus \{x\}\right) > 0. \end{cases}$$

The intensity of the Matérn type I hard-core process is $\lambda_I = \lambda_p\, \mathrm{e}^{-\lambda_p\, \omega_k h^k}$.

Definition 3.27 (Matérn type II hard-core process) Let $h > 0$, Φ_p be a stationary Poisson point process on $\mathbb{R}^k$, and $\{U(x), x \in \mathbb{R}^k\}$ independent random variables uniformly distributed over the interval $(0, 1)$. The clusters ζ_x are defined by

$$\zeta_x = \begin{cases} \delta_x, & \text{if} \quad \Phi_p(\{y \in B(x,h) \setminus \{x\} : U(y) < U(x)\}) = 0, \\ 0, & \text{otherwise.} \end{cases}$$

Then the corresponding cluster process is called Matérn type II hard-core point process.

The intensity of the Matérn hard-core type II process is

$$\lambda_{II} = (1 - \mathrm{e}^{-\lambda_p\, \omega_k h^k})/\omega_k h^k,$$

and its pair correlation function in $\mathbb{R}^2$ (which is the special case we will use in the sequel) is

$$g(h) = \begin{cases} 0 & \text{if } h < r, \\ \frac{2G_r(h)(1-\exp(-\lambda_p \pi r^2)) - 2\pi r^2 (1-\exp(-\lambda_p G_r(h)))}{\pi r^2 G_r(h)(G_r(h) - \pi r^2)\lambda^2} & \text{if } r \le h < 2r, \\ 1 & \text{if } h > 2r, \end{cases} \tag{3.13}$$

where $G_r(h) = 2r^2(\pi - \arccos(\frac{h}{2r}) + \frac{r}{2}\sqrt{4r^2 - h^2})$.

Examples of realizations of cluster and Matérn type II processes can be found in the Section 5.7.

3.4 Finite point processes

Definition 3.28 (finite point process) Let Φ be a point process on $\mathcal{X}$ which takes values almost surely in $\mathcal{N}^{\#}(\mathcal{X})$ – the set of finite measures from $\mathcal{N}(\mathcal{X})$. Then it is measurable with respect to $\mathfrak{N}^{\#} = \{N \cap \mathcal{N}^{\#} : N \in \mathfrak{N}\}$ and we call it a finite point process.

A very frequent situation when finite point processes appear is the case $\mathcal{X} \in \mathcal{B}_0(\mathbb{R}^k)$, since any locally finite measure is finite on a bounded set in $\mathbb{R}^k$. This choice of $\mathcal{X}$ is often preferred to the other possibility $\mathcal{X} = \mathbb{R}^k$ especially when we want to define and estimate point process models with different sorts of interpoint interactions or inhomogeneities.

Let us denote by $\mathcal{X}^{\#}$ the set of all finite point configurations in $\mathcal{X}$ and the finite point configurations by bold small letters $\mathbf{x}, \mathbf{y}, \mathbf{z}$, etc. In the sequel we will work with simple finite point processes and thus we will use the convention based on the isomorphism (3.1). In the notation instead of the measure $\varphi = \mathbf{1}(\mathbf{x})$ we will use its support $\operatorname{supp}\varphi = \mathbf{x}$ and instead of $\mathcal{N}^{\#}(\mathcal{X}) \cap \mathcal{N}^{*}(\mathcal{X})$ we will use $\mathcal{X}^{\#}$. Because of this convention and also to distinguish the finite case from the general one we will denote the simple finite point processes on $\mathcal{X}$ by ordinary capital letters X, Y, Z, etc., instead of Φ.

The big advantage of simple finite point processes is that we can work directly with the density f of the point process. Usually the density is defined with respect to a Poisson point process Π with finite diffuse intensity measure Λ on $\mathcal{X}$.

$$\begin{aligned}\mathbb{P}_X(X \in F) &= \int_F f(\mathbf{x})\mathbb{P}_{\Pi}(\mathrm{d}\mathbf{x}) \qquad (3.14)\\ &= \sum_{n=0}^{\infty} \frac{\exp(-\Lambda(\mathcal{X}))}{n!} \int_{\mathcal{X}} \dots \int_{\mathcal{X}} \mathbf{1}[\{x_1, \dots, x_n\} \in F]\, f(\{x_1, \dots, x_n\})\, \Lambda^n(\mathrm{d}x_1 \dots \mathrm{d}x_n),\end{aligned}$$

for $F \in \mathfrak{N}^{\#} \cap \mathcal{N}^{*}$. The unit rate Poisson point process Π_1 is used in the case when $\mathcal{X} \in \mathcal{B}_0(\mathbb{R}^k)$.

For example the density with respect to Π_1 of a Poisson point process on $\mathcal{X} \in \mathcal{B}_0(\mathbb{R}^k)$ with intensity measure Λ absolutely continuous with respect to the Lebesgue measure (denoting by $\lambda(\cdot)$ the corresponding density of Λ) is

$$f(\mathbf{x}) = \exp(|\mathcal{X}| - \Lambda(\mathcal{X})) \prod_{x \in \mathbf{x}} \lambda(x). \qquad (3.15)$$

In case of finite point processes we can express the Papangelou conditional intensity (Definition 12) easily using the density of the point process.

Theorem 3.9 *[[30] Theorem 1.6] Let X be a finite point process specified by the density $f(\mathbf{x})$ with respect to the Poisson point process with diffuse finite intensity measure Λ. Then X has Papangelou conditional intensity*

$$\lambda(x \mid \mathbf{x}) = \frac{f(\mathbf{x} \cup \{x\})}{f(\mathbf{x})} \qquad \text{for} \quad x \notin \mathbf{x},\ \mathbf{x} \in \mathcal{X}^{\#}. \qquad (3.16)$$

Note that the assumption of the theorem implies that X is a simple point process.

For a Poisson point process with intensity function λ it holds, $\lambda(x \mid \mathbf{x}) = \lambda(x)$. Heuristically $\lambda(x \mid \mathbf{x})\mathrm{d}x$ can be interpreted as the conditional probability of X having a point in an infinitesimal region containing x of size $\mathrm{d}x$ given the rest of X is $\mathbf{x}$. Thus X is attractive (has attractive interactions between points) if

$$\lambda(x \mid \mathbf{x}) \leq \lambda(x \mid \mathbf{y}) \qquad \text{whenever} \quad \mathbf{x} \subset \mathbf{y}$$

and repulsive if the opposite inequality holds.

Sometimes we know the density f up to a normalizing constant i.e. we know some $f^*(\mathbf{x})$ which is equal to $Cf(\mathbf{x})$ for some $C > 0$ (we will denote this by $f \propto f^*$), and then it is useful to have conditions which ensure the integrability of f^*, i.e. that the point process specified by f^* is well defined.

Definition 3.29 (local and Ruelle stability, hereditarity) Let $h : \mathcal{X} \to [0, \infty)$ be a function satisfying $\int_{\mathcal{X}} h(x)\mathrm{d}x < \infty$. A function $f : \mathcal{X}^{\#} \to [0, \infty)$ is hereditary if

$$f(\mathbf{x}) > 0 \implies f(\mathbf{y}) > 0, \qquad \text{for all} \quad \mathbf{y} \subset \mathbf{x},\ \mathbf{x} \in \mathcal{X}^{\#},$$

is locally stable if

$$f(\mathbf{x} \cup x) \leq h(x)\, f(\mathbf{x}), \qquad \text{for any} \quad \mathbf{x} \in \mathcal{X}^{\#},$$

and it is Ruelle stable if

$$f(\mathbf{x}) \leq \alpha \prod_{x \in \mathbf{x}} h(x) \qquad \text{for any} \quad \mathbf{x} \in \mathcal{X}^{\#},$$

and some $\alpha > 0$.

Local stability implies Ruelle stability and hereditarity. Ruelle stability implies integrability of f with respect to Π_1 on $\mathcal{X}$.

3.5 Markov point processes

Markov point process are finite point processes which exhibit (attractive or repulsive) interactions among the points of the process. As the name suggests they also have some sort of spatial Markov property that facilitates the analysis and estimation of the process.

Let $\sim$ be a reflexive and symmetric relation on $\mathcal{X}$. We say that x and y are neighbours if $x \sim y$ and define the neighbourhood ∂A of a subset of $\mathcal{X}$ by

$$\partial A = \{x \in \mathcal{X} : x \sim a \text{ for some } a \in A\}.$$

Definition 3.30 (Markov point process) Let $\mathcal{X}$ be a complete separable metric space, $\Lambda(\cdot)$ a finite, non-atomic Borel measure on $\mathcal{X}$ and $\mathbb{P}_{\Pi}$ the distribution of a Poisson point process on $\mathcal{X}$ with intensity measure Λ.
Let X be a point process on $\mathcal{X}$ specified by its density f with respect to $\mathbb{P}_{\Pi}$. Then X is a Markov point process with respect to a symmetric, reflexive relation $\sim$ on $\mathcal{X}$ if for all $\mathbf{x} \in \mathcal{X}^{\#}$ such that $f(\mathbf{x}) > 0$ holds

(i) $f(\mathbf{y}) > 0$ for all $\mathbf{y} \subseteq \mathbf{x}$,

(ii) for all $x \in \mathcal{X}$, $\frac{f(\mathbf{x} \cup \{x\})}{f(\mathbf{x})}$ depends only on x and $\partial(\{x\} \cup \mathbf{x})$.

Theorem 3.10 *[[30] Theorem 2.1] Let X be a Markov point process with density $f(\cdot)$ on $\mathcal{X}$, and consider a Borel set $A \in \mathcal{B}(\mathcal{X})$. Then the conditional distribution of $X \cap A$ given $X \cap A^c$ depends only on X restricted to the neighbourhood*

$$\partial(A) \cap A^c = \{x \in \mathcal{X} \backslash A : x \sim a \text{ for some } a \in A\}.$$

Another very convenient property of Markov point processes is the possibility of factorization of the density function $f(\mathbf{x})$ into a product over the subconfigurations of $\mathbf{x}$.

Definition 3.31 (clique, interaction function) The set $\mathbf{x} \subset \mathcal{X}$ is called a clique if $x \sim y$ holds for all pairs $\{x, y\} \subset \mathbf{x}$. The function $\varphi : \mathcal{X}^{\#} \to [0, \infty)$ is called an interaction function if $\varphi(\mathbf{y}) = 1$ whenever $\mathbf{y}$ is not a clique.

The Hammersley-Clifford theorem (originates from [41]) provides the basic characterization of Markov point processes

Theorem 3.11 *[[32] Theorem 6.1] A point process density $f : \mathcal{X}^{\#} \to [0, \infty)$ is Markov with respect to the neighbourhood relation $\sim$ if and only if there is an interaction function φ such that*

$$f(\mathbf{x}) = \prod_{\mathbf{y} \subset \mathbf{x}} \varphi(\mathbf{y}), \qquad \mathbf{x} \in \mathcal{X}^{\#}.$$

Then for $f(\mathbf{x}) > 0$

$$\lambda(x \mid \mathbf{x}) = \prod_{\mathbf{y} \subset \mathbf{x}} \varphi(\mathbf{y} \cup \{x\}).$$

Examples of Markov point processes

Poisson point process with intensity function $\lambda(x)$ can be regarded as a Markov point process with interaction function (compare with equation (3.15))

$$\begin{aligned} \varphi(\emptyset) &= \exp(|\mathcal{X}| - \int_{\mathcal{X}} \lambda(x) \mathrm{d}x), \\ \varphi(\{x\}) &= \lambda(x), \\ \varphi(\mathbf{x}) &= 1, \qquad \text{otherwise.} \end{aligned}$$

So called distance-interaction point processes have density

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

where $D(\mathbf{y}) = \mathbf{y}$ if $\mathbf{y}$ is an empty set or a singleton, otherwise

$$D(\mathbf{y}) = \{\mathcal{H}^1([u,v]) : \{u,v\} \subset \mathbf{y}, u \neq v\},$$

denotes the set of all pairwise distances of points in $\mathbf{y}$, and $[u,v]$ is the line segment connecting the points u and v. Of course the integrability of the density must be checked for the individual choices of γ.

The oldest and easiest example of a distance-interaction process is the Strauss process introduced in [49].

Definition 3.32 (Strauss point process) Let $\mathcal{X} \in \mathcal{B}_0(\mathbb{R}^k)$ and X be the point process with the density

$$f(\mathbf{x}) = \alpha \prod_{x \in \mathbf{x}} \beta \prod_{y,z \subset \mathbf{x}, \|y-z\| \leq R} \gamma, \qquad \mathbf{x} \in \mathcal{X}^{\#}, \tag{3.17}$$

where $\beta, R > 0$, $\gamma \in [0,1]$ are parameters and $\alpha > 0$ is the normalizing constant. Then X is called the Strauss point process.

We will usually write the density of the Strauss point process in a shorter form

$$f(\mathbf{x}) = \alpha\, \beta^{n(\mathbf{x})} \gamma^{s(\mathbf{x})}, \tag{3.18}$$

where by $n(\mathbf{x})$ we denote the number of points in $\mathbf{x}$ and by $s(\mathbf{x})$ the number of R-close pairs in $\mathbf{x}$. The restriction on γ is needed for the density to be integrable and the process to be well defined, and it implies that the Strauss process is an inhibitive process. Special cases are for $\gamma = 1$ the Poisson process and for $\gamma = 0$ the hard-core process – here all the points have distance at least R from each other. See the example in Figure 3.1.

An example of a Markov point process with interaction function not defined by distances between the points is the area-interaction process introduced in [4].

Definition 3.33 (area-interaction point process) Let $\mathcal{X} \in \mathcal{B}_0(\mathbb{R}^k)$, $R, \beta, \gamma > 0$ and X let be the point process with the density

$$f(\mathbf{x}) \propto \beta^{n(\mathbf{x})} \gamma^{-\mathcal{H}^2(U_R(\mathbf{x}))}, \qquad \mathbf{x} \in \mathcal{X}^{\#}, \tag{3.19}$$

where $U_R(\mathbf{x}) = \bigcup_{x \in \mathbf{x}} B(x,R)$ is the union of balls with centers in points of $\mathbf{x}$ and radius $R > 0$. Then X is called the area-interaction point process.

The point patterns of the area-interaction process are slightly clustered for $\gamma > 1$ (see the example in Figure 3.1) and regular (i.e. showing inhibition) for $\gamma < 1$. The case $\gamma = 1$ is again a Poisson point process. Generalization of the area-interaction process are shot noise processes.

Definition 3.34 (shot noise process) Let $\mathcal{X} \in \mathcal{B}_0(\mathbb{R}^k)$, $R, \beta, \gamma > 0$ and let p be a function defined on non-negative integers $\mathbb{N}_0$ with $p(0) = 0$. The point process X with the density

$$f(\mathbf{x}) \propto \beta^{n(\mathbf{x})} \gamma^{-\int_{\mathbb{R}^k} p(C_{\mathbf{x}}(u))\mathcal{H}^k(\mathrm{d}u)}, \qquad \mathbf{x} \in \mathcal{X}^{\#},$$

where

$$C_{\mathbf{x}}(u) = \sum_{x \in \mathbf{x}} \mathbf{1}(u \in B(x,R)).$$

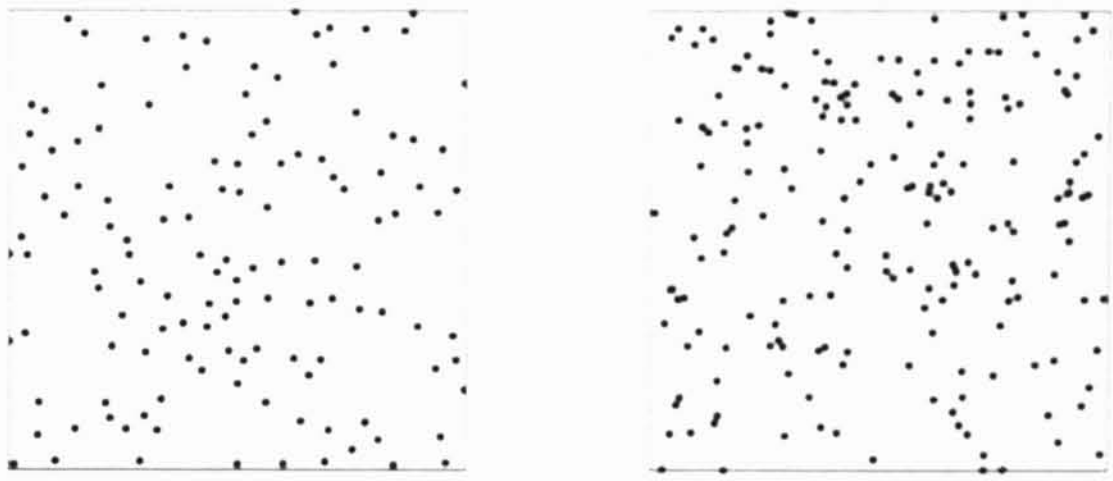


Figure 3.1: Left part: example of Strauss point process. Right part: example of area interaction process with $\gamma > 1$.

is called a shot noise point process.

The interaction functions of the shot noise process are

$$\begin{aligned} \varphi(\mathbf{y}) &= \beta\,\gamma^{-m(\mathbf{y})}, & n(\mathbf{y}) &= 1, \\ \varphi(\mathbf{y}) &= \gamma^{-m(\mathbf{y})}, & n(\mathbf{y}) &> 1, \end{aligned}$$

where

$$m(\mathbf{y}) = \mathcal{H}^k\left(\bigcap_{y\in\mathbf{y}} B(y,R)\right)\sum_{l=1}^{n(\mathbf{y})}\binom{n(\mathbf{y})}{l}(-1)^{n(\mathbf{y})-l}p\,(l).$$

If there exists $C > 0$ such that

$$|p\,(n)| < C\,n \qquad \text{for all } n \in \mathbb{N}_0,$$

then the density of the shot noise process is integrable and the process is well defined (for proof see [31]).

In the case of finite point processes we are not able to define stationarity, but still we need some notion describing the homogeneity property of the point process - i.e. that the point pattern does not show essential differences in different parts of $\mathcal{X}$ (like the point patterns in the Figure 3.1).

Definition 3.35 (homogeneous point process) Let X be a finite point process on $\mathcal{X} \subseteq \mathbb{R}^k$ specified by the density f with respect to the unit rate Poisson point process. Then we will call X homogeneous if there exists a function F on $\mathbb{R}^k$ which is translation invariant on $\mathbb{R}^k$ (i.e. $F(t_y(\mathbf{x}))) = F(\mathbf{x})$ for any $\mathbf{x} \in (\mathbb{R}^k)^{\#}$ and $y \in \mathbb{R}^k$) and $f = F$ on $\mathcal{X}$.

The definition of homogeneous Markov point processes can be extended to processes defined on $\mathbb{R}^k$, to so called Gibbs point process (see [32]). However we will not need this notion in our work.

All the examples defined above are homogeneous. Inhomogeneous Markov point processes are still not used very often for modelling of real situations. A good review and

comparison of different ways of introducing inhomogeneity into a homogeneous Markov point process model can be found in [25].

Let us end this section with a brief comment about estimation of the model parameters. This is usually done using the maximum likelihood estimation. However in all but the simplest cases this cannot be done explicitly and computation intensive Monte Carlo Markov chain (MCMC) methods for the approximate maximum likelihood estimation are used (see [32], [30], [15]). A computationally less demanding alternative in the case of Markov point processes is the estimation using maximization of so called pseudolikelihood.

Definition 3.36 (pseudolikelihood) Let a finite simple point process X have density f_θ ($\theta \in \mathbb{R}^p$ parameter) with respect to the Poisson point process with intensity measure μ. The pseudolikelihood function $PL(\theta; \mathbf{x})$ based on a realization $\mathbf{x}$ of X observed in the window $W \subseteq \mathcal{X}$, is defined by

$$PL(\theta; \mathbf{x}) = \exp\left(-\int_W [\lambda_\theta(u \mid \mathbf{x}) - 1]\, \mu(\mathrm{d}u)\right) \prod_{x \in \mathbf{x} \cap W} \lambda_\theta(x \mid \mathbf{x} \backslash \{x\}) ,$$

where

$$\lambda_\theta(u \mid \mathbf{x}) = \frac{f_\theta(\mathbf{x} \cup \{u\})}{f_\theta(\mathbf{x})} , \quad u \notin \mathbf{x} ,$$

is the Papangelou conditional intensity associated with f_θ.

For increasing size of the observation window and under quite restrictive assumptions strong consistency (i.e. convergence with probability one) and asymptotic normality of the maximum likelihood estimates for homogeneous Markov point processes can be shown for some of the models. However there are still many open questions in this field. For details we refer the reader to [32]. More asymptotic results are available for the pseudolikelihood estimates for homogeneous Markov point processes with finite range of interaction and exponential family type densities. Under some assumptions the pseudolikelihood estimates are also strongly consistent and asymptotically normal, but they are not asymptotically efficient (see again [32]).

Chapter 4

Statistics for locally scaled point processes

4.1 Introduction

In recent years models for inhomogeneous spatial point processes have been studied quite intensively. There are several possibilities how to introduce inhomogeneity into the models - non-constant first order term, independent thinning or transformation of the point process, see [48], [2], [25].

The present paper deals with statistical analysis for inhomogeneous point processes that are obtained by local scaling — a recently suggested new class of inhomogeneous point processes with interactions ([18]). In these point processes, local geometry is constant, that is, subregions of the inhomogeneous process with different intensity appear to be scaled versions of the same homogeneous process. This property is characteristic of locally scaled point processes and it is not present in the other models for inhomogeneous point processes (see discussion in [18]).

Such patterns occur for example in vegetation of dry areas, as shown in Figure 4.1. Where water or other resources are short, plants grow sparsely and keep larger distances between individuals than in regions with better supply. Naturally there is no preference for a direction, and therefore the vegetation pattern is locally isotropic. Local scaling of an isotropic template process yields locally isotropic patterns in contrast to the transformation of an isotropic template process ([34]).

Similar locally scaled structures are found in arrangements of solid bodies with constant shape but location dependent size, such as the sinter filter discussed in [18] or in sponges with constant porosity but small pore size close to the surface and large pore size in the interior.

Locally scaled point processes are derived from a homogeneous template process which describes the interaction between points and is responsible for the local geometry of the resulting pattern. We will put the major focus on Markov template processes. Inhomogeneity is introduced through a location dependent function that gives the local scale. Fitting a model to a given pattern thus consists of finding the parameters inherited from the template and choosing an appropriate scaling function.

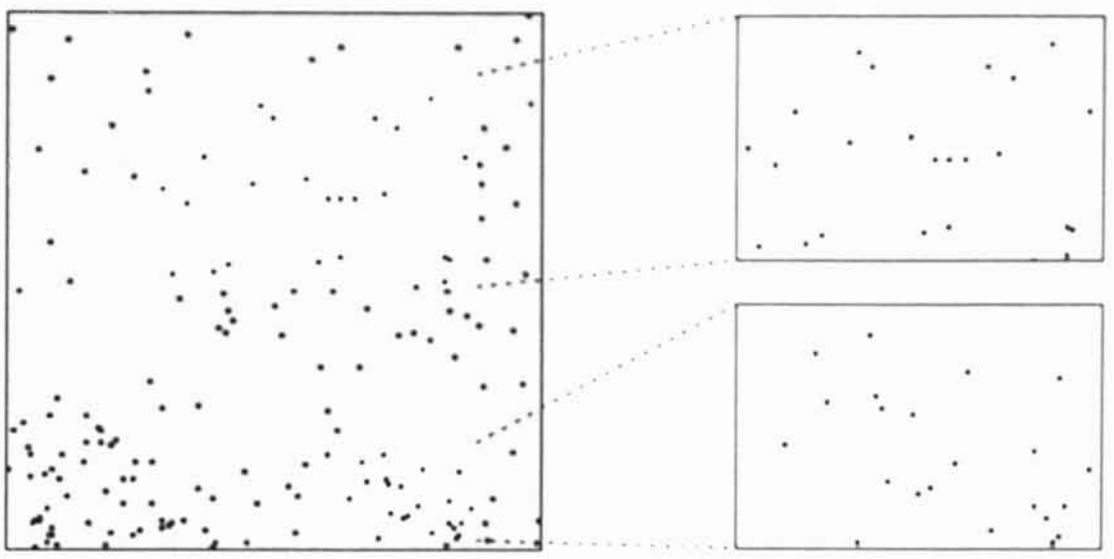


Figure 4.1: Left part: Map of 171 individuals of a *Scholtzia* aff. *involucrata* in Australian bush on a 220×220 m square. Right part: Two rectangular subregions with different intensity were rescaled such that they have the same number of individuals by unit area. Data from [1].

This can be achieved by several methods using likelihood or pseudolikelihood, simultaneous or two-step estimation. All the methods are discussed in the sequel and the results are supported by a simulation study. Thereafter we introduce two model validation methods suitable for locally scaled point processes and all the developed statistical tools are applied to the analysis of the vegetation point pattern from Figure 4.1.

4.2 Locally scaled point processes

In this section, we introduce the locally scaled point processes and discuss some of their basic properties. More details about locally scaled point processes can be found in [18].

At first we need several definitions.

Definition 4.1 (scale invariant function) Let $g(\mathbf{x}; \mu^*)$ be a real-valued measurable function defined on $(\mathbb{R}^k)^\#$, depending on a set $\mu^* = (\mu^1, \ldots, \mu^m)$ of measures on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. The function g is called scale invariant if for all $\mathbf{x} \in (\mathbb{R}^k)^\#$, and all $c > 0$

$$g(c\,\mathbf{x}; \mu_c^*) = g(\mathbf{x}; \mu^*),$$

where $\mu_c^* = (\mu_c^1, \ldots, \mu_c^m)$ and μ_c is a measure scaled from μ with a factor c, i.e. $\mu_c(A) = \mu(c^{-1}A)$, for all $A \in \mathcal{B}(\mathbb{R}^k)$.

Definition 4.2 (locally scaled volume measures) Let c be a positive Borel measurable function on $\mathbb{R}^k$. Then the locally scaled d-dimensional Hausdorff (volume) measure $\mathcal{H}_c^d$ is defined by

$$\mathcal{H}_c^d(A) = \int_A c(u)^{-d} \mathcal{H}^d(\mathrm{d}u),$$

for all $A \in \mathcal{B}(\mathbb{R}^k)$.

Let $\mathcal{H}^* = (\mathcal{H}^0, \ldots, \mathcal{H}^k)$ be the set of d-dimensional Hausdorff measures $\mathcal{H}^d$ in $\mathbb{R}^k$, $d = 0, 1, \ldots, k$. Then the classical homogeneous point processes have densities which are scale-invariant with respect to $\mathcal{H}^*$.

Let $\mathbb{B}^k$ denote the set of all full-dimensional (i.e. $\mathcal{H}^k(\mathcal{X}) > 0$) bounded subsets of $\mathbb{R}^k$. Let X be a finite point process, defined on a $\mathcal{X} \in \mathbb{B}^k$ and suppose that X has a density f_X with respect to the restriction of the unit rate Poisson point process Π_1 to $\mathcal{X}$.

X will serve as a template process. In order to construct a locally scaled version of X with scaling function $c : \mathbb{R}^k \to \mathbb{R}_+$, we replace the d-dimensional volume measures $\mathcal{H}^d$ in $\mathbb{R}^k$ with their locally scaled versions $\mathcal{H}^d_c$. In what follows, we assume that the scaling function c is bounded from below and from above, i.e.

$$0 < \underline{c} < c(u) < \bar{c}, \qquad u \in \mathbb{R}^k. \tag{4.1}$$

Definition 4.3 (locally scaled point process) Let X be a homogeneous point process on $\mathcal{X}$, with density f_X with respect to Π_1 of the form

$$f_X(\mathbf{x}) \propto g(\mathbf{x}; \mathcal{H}^*), \qquad \mathbf{x} \in \mathcal{X}^\#, \tag{4.2}$$

where g is scale invariant. Let c be a positive Borel measurable function on $\mathbb{R}^k$ and let Π_c be the Poisson point process with the locally scaled volume measure $\mathcal{H}^k_c$ as the intensity measure. Let $\mathcal{X}'$ be an arbitrary full-dimensional bounded subset of $\mathbb{R}^k$ and suppose that $g(\cdot; \mathcal{H}^*)$ is integrable on $(\mathcal{X}')^\#$ with respect to Π_c. A locally scaled point process X_c on $\mathcal{X}'$ with template X is defined by the following density with respect to Π_c.

$$f^{(c)}_{X_c}(\mathbf{x}) \propto g(\mathbf{x}; \mathcal{H}^*_c), \qquad \mathbf{x} \in (\mathcal{X}')^\#,$$

where the upper index $^{(c)}$ denotes that the density is taken with respect to Π_c.

Note that the density of X_c with respect to Π_1 is

$$f_{X_c}(\mathbf{x}) = \exp\left(- \int_{\mathcal{X}} [c(u)^{-k} - 1] \mathcal{H}^k(\mathrm{d}u) \right) \prod_{x \in \mathbf{x}} c(x)^{-k} \times f^{(c)}_{X_c}(\mathbf{x}), \quad \mathbf{x} \in (\mathcal{X}')^\#, \tag{4.3}$$

which follows from the formula (3.15) for the density of Π_c with respect to Π_1.

Let us show here several examples of the locally scaled point processes to make the rather technical definition more intelligible.

Example 4.3 According to the Definition 3.32 the Strauss process X with intensity parameter $\beta > 0$, interaction parameter $\gamma \in [0, 1]$ and interaction distance $R > 0$ is given by the density

$$f_X(\mathbf{x}) \propto \beta^{n(\mathbf{x})} \gamma^{s(\mathbf{x})}, \qquad \mathbf{x} \in \mathcal{X}^\#,$$

where $n(\mathbf{x})$ is the number of points in $\mathbf{x}$ and $s(\mathbf{x})$ is the number of R-close pairs. The density is of the form (4.2) with

$$g(\mathbf{x}; \mathcal{H}^*) = \beta^{\mathcal{H}^0(\mathbf{x})} \gamma^{\sum^{\neq}_{\{u,v\} \subset \mathbf{x}} \mathbf{1}\{\mathcal{H}^1([u,v]) \le R\}},$$

where the superscript $\neq$ in the summation indicates that u and v are different. This function is scale-invariant. The locally scaled Strauss process X_c has density with respect to Π_c of the form

$$f^{(c)}_{X_c}(\mathbf{x}) \propto \beta^{n(\mathbf{x})} \gamma^{s_c(\mathbf{x})}, \qquad \mathbf{x} \in (\mathcal{X}')^\#$$

where

$$s_c(\mathbf{x}) = \sum_{\{u,v\}\subseteq\mathbf{x}}^{\neq} \mathbf{1}\{\mathcal{H}_c^1([u,v]) \leq R\} \ .$$

Figure 4.2 shows locally scaled Strauss processes on $\mathcal{X} = [0,1]^2$ with scaling function of the exponential form

$$c_\theta(u) = \sqrt{\frac{1-\mathrm{e}^{-2\theta}}{2\theta}}\,\mathrm{e}^{\theta u_1}, \quad u = (u_1,u_2) \in \mathcal{X}, \quad \theta \in \mathbb{R}^1, \tag{4.4}$$

for four different values of the inhomogeneity parameter $\theta \in \{0.25, 0.5, 1, 1.5\}$. The normalisation $\sqrt{\frac{1-\mathrm{e}^{-2\theta}}{2\theta}}$ ensures that the four point patterns have approximately the same number of points (see Section 4.4.2 for details).

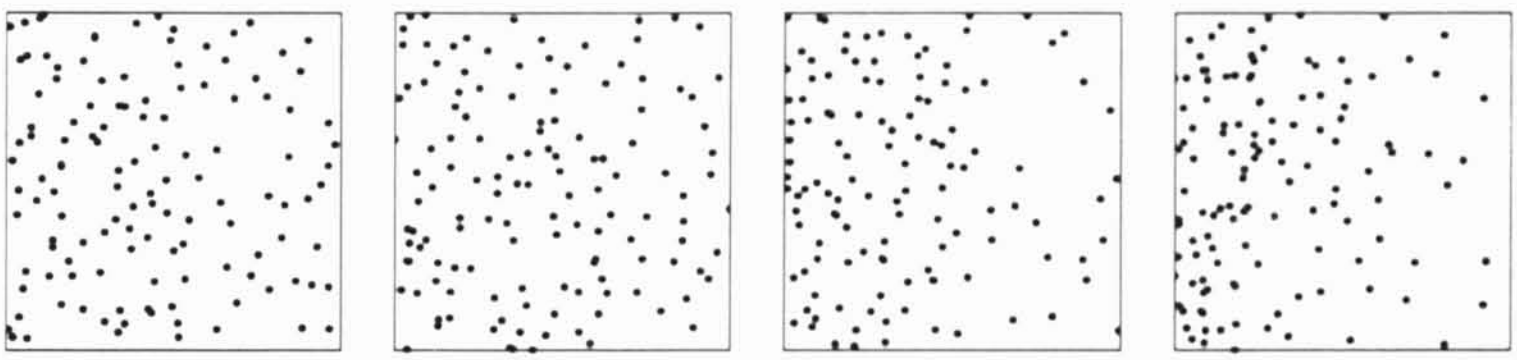


Figure 4.2: Simulation of the locally scaled Strauss processes on $[0,1]^2$ with the exponential scaling function (4.4) for $\theta \in \{0.25, 0.5, 1, 1.5\}$ (from left to right) and the template parameters $\beta = 250$, $\gamma = 0.3$ and $R = 0.05$.

□

Example 4.4 According to the Definition 3.33 the area-interaction point process with intensity parameter $\beta > 0$, interaction parameter $\gamma > 0$ and interaction distance $R > 0$ is given by the density

$$f(\mathbf{x}) \propto \beta^{n(\mathbf{x})}\gamma^{-\mathcal{H}^2(U_R(\mathbf{x}))}, \qquad \mathbf{x} \in \mathcal{X}^{\#},$$

where $U_R(\mathbf{x}) = \bigcup_{x\in\mathbf{x}} B(x,R)$ is the union of balls with centers in $\mathbf{x}$ and radius R. The density is again of the form (4.2) with scale invariant

$$g(\mathbf{x};\mathcal{H}^*) = \beta^{\mathcal{H}^0(\mathbf{x})}\gamma^{-\mathcal{H}^2(\cup_{x\in\mathbf{x}}\{v\in\mathcal{X}:\mathcal{H}^1([v,x])\leq R\})} \ .$$

The locally scaled area-interaction process has density with respect to Π_c of the form

$$f_{X_c}^{(c)}(\mathbf{x}) \propto \beta^{n(\mathbf{x})}\gamma^{-\mathcal{H}_c^2(U_{c,R}(\mathbf{x}))}, \qquad \mathbf{x} \in (\mathcal{X}')^{\#}$$

where $U_{c,R} = \bigcup_{x\in\mathbf{x}} B_c(x,R)$ and $B_c(x,R) = \{v \in \mathcal{X} : \mathcal{H}_c^1([v,x]) \leq R\}$ is the scaled ball. Figure 4.3 shows locally scaled area-interaction processes with the same scaling function (4.4) as in the Example 4.3. The value of the interaction parameter γ was chosen so that $\gamma^{-\pi R^2} \approx 0.1$ which causes the point patterns being visibly clustered.

□

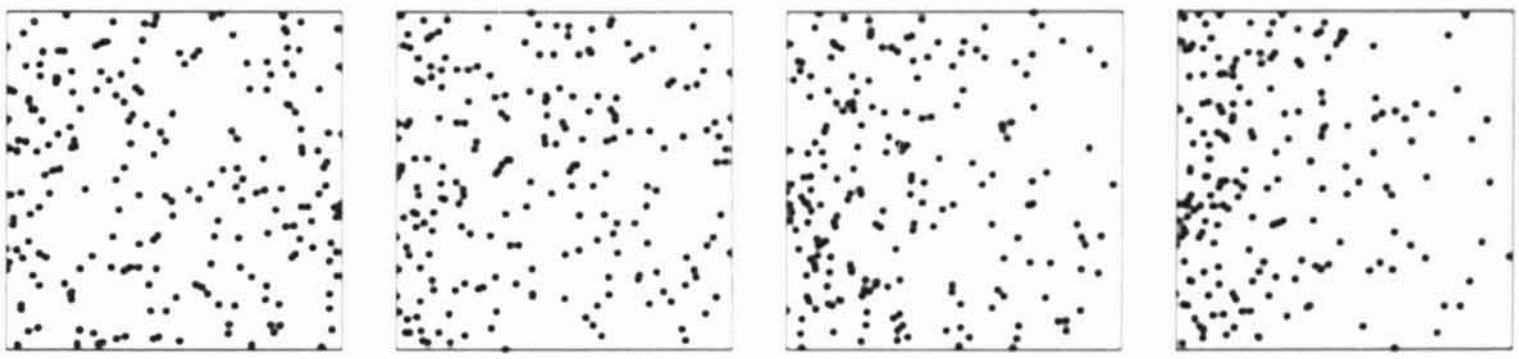


Figure 4.3: Simulation of the locally scaled area-interaction processes on $[0,1]^2$ with the exponential scaling function (4.4) for $\theta \in \{0.25, 0.5, 1, 1.5\}$ (from left to right) and the template parameters $\beta = 180$, $\gamma = 6.7 \cdot 10^{31}$ and $R = 0.1$.

The Strauss process and the area-interaction process are examples of point processes from two large classes of homogeneous template processes, viz. the distance-interaction processes and the shot noise weighted processes. For these two classes, it has been shown in [18] that the Papangelou conditional intensities $\lambda_{X_c}(x \mid \mathbf{x})$ of the locally scaled process X_c and $\lambda_X(x \mid \mathbf{x})$ of the template process X, respectively, satisfy

$$\lambda_{X_c}(x \mid \mathbf{x}) = c_0^{-k} \lambda_X(\frac{x}{c_0} \mid \frac{\mathbf{x}}{c_0}) , \tag{4.5}$$

if the scaling function c is constant and equal to c_0 in a scaled neighbourhood of x. It was explained in Section 3.4 that if we let $\mathrm{d}x$ be an infinitesimal region around x and $\mathcal{H}^k(\mathrm{d}x)$ the k-dimensional volume (Lebesgue measure) of $\mathrm{d}x$, then $\lambda(x \mid \mathbf{x})\mathcal{H}^k(\mathrm{d}x)$ can be interpreted as the conditional probability of finding a point from the process in $\mathrm{d}x$ given the configuration elsewhere is $\mathbf{x}$. Since the right-hand side of (4.5) is the conditional intensity of a globally scaled template process with scaling factor c_0, it is expected that the locally scaled process appears as a scaled version of the template process if the scaling function is slowly varying compared to the interaction radius. The development of further formal reasoning, supporting this statement, seems to be very hard.

It is also of interest to study the unconditional intensity function $\lambda_c(x)$, $x \in \mathcal{X}$, of the locally scaled process. Let us suppose that the template process X is homogeneous with intensity λ_0 (X may, for instance, be defined on a torus with periodic boundary conditions, see i.e. [35]). Then

$$\lambda_c(x) = c(x)^{-k} \lambda_0, \quad x \in \mathcal{X} , \tag{4.6}$$

holds if the template process is Poisson or the scaling function is constant. Also, (4.6) holds for any locally scaled distance-interaction process in $\mathbb{R}^1$.

Theorem 4.1 *Let X be a homogeneous distance-interaction process on an interval $I = [a, b]$, $a < b$ of $\mathbb{R}^1$ with periodic boundary conditions and let X_c be its locally scaled version derived using the scaling function c. Then it holds*

$$\lambda_c(x) = c(x)^{-1} \lambda_0, \quad x \in \mathcal{X},$$

where λ_0 is the intensity of the template process X and $\lambda_c(x)$ is the intensity of the locally scaled process X_c.

Proof. The distance-interaction process X has the density

$$f_X(\mathbf{x}) \propto \beta^{n(\mathbf{x})} \prod_{\mathbf{y} \subseteq_2 \mathbf{x}} \varphi(\{\mathcal{H}^1([u,v]) : \{u,v\} \subseteq \mathbf{y}, u \neq v\}),$$

where φ is the interaction function and $\subseteq_2$ indicates that $\mathbf{y}$ should have at least two elements. The density of X_c is then

$$f_{X_c}(\mathbf{x}) \propto \prod_{x \in \mathbf{x}} c(x)^{-1} \times \beta^{n(\mathbf{x})} \prod_{\mathbf{y} \subseteq_2 \mathbf{x}} \varphi(\{\mathcal{H}^1_c([u,v]) : \{u,v\} \subseteq \mathbf{y}, u \neq v\}).$$

Let us consider the transformation h of I onto I defined by

$$h^{-1}(x) = \int_a^x c(u)^{-1} \mathrm{d}u.$$

Then h is 1-1 differentiable because c is bounded from above and from below, and the density of X_c can be rewritten as

$$f_{X_c}(\mathbf{x}) \propto \prod_{x \in \mathbf{x}} Jh^{-1}(\mathbf{x}) \cdot f_X(h^{-1}(\mathbf{x})),$$

where Jh^{-1} is the Jacobian of the inverse transformation to h.
It follows that X_c is distributed as $h(X)$. In particular, for $A \in \mathcal{B}(I)$,

$$\mathbb{E}n(X_c \cap A) = \mathbb{E}n(X \cap h^{-1}(A)) = \int_{h^{-1}(A)} \lambda_0 \mathrm{d}x = \int_A c(u)^{-1} \lambda_0 \mathrm{d}u,$$

or

$$\lambda_c(u) = c(u)^{-1} \lambda_0.$$

□

The equality (4.6) is expected to hold approximately if the scaling function is slowly varying, compared to the interaction radius.

For statistical inference of locally scaled models, we will distinguish two cases. In fully parametric models, both the scaling function c and the homogeneous template process X are specified by a set of parameters. In semiparametric models only the template process is parametrically specified.

In the following, the parameters of the template process are denoted by ψ, and θ is the parameter of the scaling function (i.e. inhomogeneity parameter) in fully parametric models. The parameter space of a fully parametric model is $\Theta \times \Psi$, while, in semiparametric models, the scaling function can be any function in the space $\mathcal{C}^+$ of measurable positive functions, satisfying the regularity condition (4.1).

A particularly attractive parametric form of the scaling function is the exponential form

$$c_\theta(u) = \alpha(\theta)\, \mathrm{e}^{\langle \theta, \tau(u) \rangle}, \quad u \in \mathbb{R}^k, \tag{4.7}$$

where $\theta \in \Theta \subseteq \mathbb{R}^l$, $\alpha(\theta) \in \mathbb{R}_+$ and $\tau(u) \in \mathbb{R}^l$. A locally scaled model with an exponential scaling function is called an exponentially scaled model. Note that if $\tau(u) = u$, then scaled distances can be calculated explicitly. Using the coarea formula ([24] Corollary 2.4) we get

$$\begin{aligned}
\mathcal{H}^1_c([u,v]) &= \int_{[u,v]} \alpha(\theta)^{-1} \mathrm{e}^{-\langle \theta, x \rangle} \, \mathcal{H}^1(\mathrm{d}x) = \\
&= \int_0^1 \|v-u\| \, \alpha(\theta)^{-1} \mathrm{e}^{-\langle \theta, (u+t(v-u)) \rangle} \, \mathrm{d}t = \\
&= \|v-u\| \, \frac{[c_\theta(u)^{-1} - c_\theta(v)^{-1}]}{\theta \cdot (v-u)}, \qquad u, v \in \mathbb{R}^k .
\end{aligned}$$

4.3 Simultaneous maximum (pseudo)likelihood estimation of scaling function and template parameters

Equation (4.3) implies that in a fully parametric model, the likelihood factorizes as

$$L(\theta, \psi; \mathbf{x}) = L_0(\theta; \mathbf{x}) \times L_1(\theta, \psi; \mathbf{x}) , \tag{4.8}$$

where L_0 is the likelihood of an inhomogeneous Poisson point process Π_c with intensity measure $\mathcal{H}^k_c$, and $L_1(\theta, \psi; \mathbf{x}) = f^{(c)}_{X_c}(\mathbf{x}; \psi)$ is the density of the scaled process X_c with respect to Π_c. Recall that the scaling function is parametrized by θ, i.e. $c = c_\theta$.

Maximum likelihood estimation is most feasible in exponential families, since it amounts to moment estimation there. Most popular homogeneous Markov point process models are partially exponential, and the set ψ splits into two components – the nuisance parameters and the remaining parameters, that form exponential family parameters given the nuisance parameters. Since the likelihood in Markov point processes is known only up to the normalizing constant, one has to resort to MCMC methods for MLE, see e.g. [32] or [15]. Whilst moment estimation in these models can be done relatively precisely with affordable effort, estimation of the normalizing constant entails numerical pitfalls and should be avoided as much as possible. This suggests that MLE should be done on a grid of nuisance parameters, since given this component, the remaining parameters are exponential family parameters. In locally scaled processes, the inhomogeneity parameter also acts as a nuisance parameter.

Usually, the point process X_c is observed in a sampling window $W \subseteq \mathcal{X}$. In such cases, a conditional likelihood may be used, based on the conditional density of $X_c \cap W$ given $X_c \cap W^c = \mathbf{x}_{W^c}$ where $\mathbf{x}_{W^c}$ is a finite subset of W^c. Since

$$f_{X_c}(\cdot \mid \mathbf{x}_{W^c}) \propto f_{X_c}(\cdot \cup \mathbf{x}_{W^c}) ,$$

it follows from (4.3) that (4.8) still holds for the conditional likelihoods. This result is mainly of interest for locally scaled Markov point processes.

A less computationally demanding procedure then the maximum likelihood estimation is based on the pseudolikelihood function (Definition 3.36), see [3] and references therein. The following theorem shows that the same factorization like (4.8) holds also for the pseudolikelihood.

Theorem 4.2 *Let $PL_W(\theta, \psi; \mathbf{x})$ be the pseudolikelihood function for the density $f_{X_{c_\theta}}(\cdot; \psi)$ with respect to the unit rate Poisson point process and let $PL_{W,1}(\theta, \psi; \mathbf{x})$ be the pseudolikelihood for the density $f^{(c_\theta)}_{X_{c_\theta}}(\cdot; \psi)$ with respect to the Poisson point process with intensity measure $\mathcal{H}^k_{c_\theta}$, both of the processes observed on a window W. Then*

$$\lambda_{\theta,\psi}(u \mid \mathbf{x}) = c_\theta(u)^{-k}\, \lambda^{(c_\theta)}_{\theta,\psi}(u \mid \mathbf{x}) \,, \quad u \notin \mathbf{x} \,, \tag{4.9}$$

where $\lambda_{\theta,\psi}$ and $\lambda^{(c_\theta)}_{\theta,\psi}$ are the conditional intensities associated with $f_{X_{c_\theta}}$ and $f^{(c_\theta)}_{X_{c_\theta}}$, respectively. Moreover the following factorization of the pseudolikelihood holds

$$PL_W(\theta, \psi; \mathbf{x}) = L_0(\theta; \mathbf{x} \cap W) \times PL_{W,1}(\theta, \psi; \mathbf{x}) \,. \tag{4.10}$$

Proof. Because the scaling function c is bounded from above and from below, the intensity measure $\mathcal{H}^k_c$ is obviously diffuse and finite on $\mathcal{X} \in \mathbb{B}^k$. The first part of the assertion is then a direct consequence of Theorem 3.9 and equation (4.3).

$$\begin{aligned}
\lambda^{(c_\theta)}_{\theta,\psi}(u \mid \mathbf{x}) &= \frac{f^{(c_\theta)}_{X_{c_\theta}}(\mathbf{x} \cup \{u\})}{f^{(c_\theta)}_{X_{c_\theta}}(\mathbf{x})} \\
&= \frac{\exp\left(-\int_{\mathcal{X}} [c(u)^{-k} - 1]\mathcal{H}^k(\mathrm{d}u)\right) \prod_{x \in \mathbf{x} \cup \{u\}} c(x)^{-k} f_{X_{c_\theta}}(\mathbf{x} \cup \{u\})}{\exp\left(-\int_{\mathcal{X}} [c(u)^{-k} - 1]\mathcal{H}^k(\mathrm{d}u)\right) \prod_{x \in \mathbf{x}} c(x)^{-k} f_{X_{c_\theta}}(\mathbf{x})} \\
&= c_\theta(u)^{-k}\, \frac{f_{X_{c_\theta}}(\mathbf{x} \cup \{u\})}{f_{X_{c_\theta}}(\mathbf{x})} = c_\theta(u)^{-k}\, \lambda^{(c_\theta)}_{\theta,\psi}(u \mid \mathbf{x}).
\end{aligned}$$

The proof of (4.10) can be constructed as follows. From (4.3), we get

$$\begin{aligned}
&PL_W(\theta, \psi; \mathbf{x}) \\
&= \exp(-\int_W [\lambda_{\theta,\psi}(u \mid \mathbf{x}) - 1]\mathcal{H}^k(\mathrm{d}u)) \prod_{x \in \mathbf{x} \cap W} \lambda_{\theta,\psi}(x \mid \mathbf{x} \backslash \{x\}) \\
&= \exp(-\int_W [c_\theta(u)^{-k} - 1]\mathcal{H}^k(\mathrm{d}u)) \prod_{x \in \mathbf{x} \cap W} c_\theta(x)^{-k} \\
&\quad \times \exp(-\int_W [\lambda^{(c_\theta)}_{\theta,\psi}(u \mid \mathbf{x}) - 1] c_\theta(u)^{-k} \mathcal{H}^k(\mathrm{d}u)) \prod_{x \in \mathbf{x} \cap W} \lambda^{(c_\theta)}_{\theta,\psi}(x \mid \mathbf{x} \backslash \{x\}) \\
&= L_0(\theta; \mathbf{x}) \times PL_{W,1}(\theta, \psi; \mathbf{x}) \,.
\end{aligned}$$

□

Since the values of the scaled interaction statistics (e.g. $s_{c_\theta}(\mathbf{x})$ in the Strauss model) and subsequently the values of $\lambda^{(c_\theta)}_{\theta,\psi}(u \mid \mathbf{x})$ depend on the inhomogeneity parameter θ,

the latter is a nuisance parameter also in the pseudolikelihood estimation. This means we have to evaluate the profile pseudolikelihood on a grid of nuisance parameters similarly to the maximum likelihood approach. However, this is much less computational intensive in maximum pseudolikelihood estimation than in maximum likelihood estimation, since PL_1 can be calculated directly without having to estimate an unknown normalizing constant by simulation as it is the case with L_1.

4.4 Two step maximum likelihood estimation of scaling parameters prior to template parameters

The structural similarity of the full likelihood in locally scaled models and the full likelihood in transformation models for point processes suggests that partial likelihood inference as in the paper [35] will be successful also for locally scaled models. [35] estimated the inhomogeneity parameters by maximizing the Poisson part L_0 of the likelihood only, assuming no interaction in the model in this phase of the estimation. They chose an exponential model for the inhomogeneity function, since this largely simplifies calculations.

Below, this approach is followed for the locally scaled models. In Section 4.4.1, we find the maximum likelihood estimate $\widehat{\theta}_0$ of θ on the basis of L_0 and, in Section 4.4.2, it is shown that $\widehat{\theta}_0$ can be regarded as an approximate moment estimator. Estimation of the template parameters is considered in Section 4.4.3.

4.4.1 Estimation of scaling parameters, using the Poisson likelihood

We suppose that the scaling function is of the form

$$c(u) = \alpha\, \mathrm{e}^{\langle \theta,\, \tau(u) \rangle}, \qquad u \in \mathbb{R}^k\,, \tag{4.11}$$

where $\theta \in \Theta \subseteq \mathbb{R}^d$, $\alpha \in \mathbb{R}_+$ and τ is an $\mathbb{R}^d$-valued measurable function. In addition to the inhomogeneity parameter θ, the scaling function contains a global scaling parameter α. For the moment, these two parameters vary in a product set $\Theta \times \mathbb{R}_+$.

Then, the Poisson part of the likelihood of the process X_c, observed in a set W, is

$$L_0(\theta, \alpha; \mathbf{x} \cap W) = \exp\left(- \int_W (\alpha^{-k} \mathrm{e}^{-k\langle \theta,\, \tau(u) \rangle} - 1)\, \mathcal{H}^k(\mathrm{d}u) \right) \prod_{x \in \mathbf{x} \cap W} (\alpha^{-k} \mathrm{e}^{-k\langle \theta,\, \tau(x) \rangle})\,. \tag{4.12}$$

The log-likelihood becomes

$$\begin{aligned} l_0(\theta, \alpha; \mathbf{x} \cap W) = \quad & \int_W 1\, \mathcal{H}^k(\mathrm{d}u) - \int_W \alpha^{-k} \mathrm{e}^{-k\langle \theta,\, \tau(u) \rangle}\, \mathcal{H}^k(\mathrm{d}u) \; - \\ & -k\, n(\mathbf{x} \cap W) \ln \alpha + \sum_{x \in \mathbf{x} \cap W} (-k\, \langle \theta,\, \tau(x) \rangle)\,. \end{aligned}$$

Assume that $n(\mathbf{x} \cap W) > 0$ and $\|\tau(u)\| \, e^{\langle \theta, \tau(u) \rangle}$ is uniformly bounded in $u \in W$ and $\theta \in \Theta$. Then by differentiating we get $d + 1$ equations

$$\begin{aligned} k\alpha^{-k-1} \int_W e^{-k\langle \theta, \tau(u) \rangle} \, \mathcal{H}^k(\mathrm{d}u) &= k \, n(\mathbf{x} \cap W) \alpha^{-1} \\ \alpha^{-k} \int_W k\tau_i(u) e^{-k\langle \theta, \tau(u) \rangle} \, \mathcal{H}^k(\mathrm{d}u) &= k \sum_{x \in \mathbf{x} \cap W} \tau_i(x), \quad i = 1, \ldots, d \, . \end{aligned}$$

Dividing the last d equations by the first equation we get the vector equation

$$\frac{t(\mathbf{x} \cap W)}{n(\mathbf{x} \cap W)} = m(\widehat{\theta}_0) \, . \tag{4.13}$$

where $t(\mathbf{x} \cap W) = \sum_{x \in \mathbf{x} \cap W} \tau(x)$ and

$$m(\theta) = \frac{\int_W \tau(u) e^{-k\langle \theta, \tau(u) \rangle} \, \mathcal{H}^k(\mathrm{d}u)}{\int_W e^{-k\langle \theta, \tau(u) \rangle} \, \mathcal{H}^k(\mathrm{d}u)}, \qquad \theta \in \Theta. \tag{4.14}$$

Thus we estimate θ independently of α (the estimate $\widehat{\theta}_0$ of θ does not depend on the estimate of the constant α) and furthermore the estimate depends only on the statistic $t(\mathbf{x} \cap W)/n(\mathbf{x} \cap W)$. It turns out that we get exactly the same estimate of θ if we impose a normalization condition on c_θ. (As we shall see in Section 4.4.2, (4.15) appears to be a very natural condition.)

Theorem 4.3 *Let X_c be a locally scaled point process with the exponential scaling function (4.11) parametrized by θ. Suppose X is observed on the window W and impose on c_θ the following normalizing condition*

$$\int_W c_\theta(u)^{-k} \, \mathcal{H}^k(du) = \mathcal{H}^k(W) \, . \tag{4.15}$$

Then the maximum likelihood estimate $\widehat{\theta}_0$ of θ based on the Poisson likelihood (4.12) is the solution of the equation (4.13) and

$$\alpha = \alpha(\theta) = \left[\int_W e^{-k\langle \theta, \tau(u) \rangle} \mathcal{H}^k(du) \, / \, \mathcal{H}^k(W) \right]^{1/k} \, . \tag{4.16}$$

Proof. (4.16) follows directly from the condition (4.15) and under (4.16) the Poisson likelihood (4.12) takes the form

$$\begin{aligned} L_0(\theta; \mathbf{x} \cap W) = \quad & \exp\left(- \int_W (\alpha(\theta)^{-k} e^{-k\langle \theta, \tau(u) \rangle} - 1) \, \mathcal{H}^k(\mathrm{d}u) \right) \times \\ & \times \prod_{x \in \mathbf{x} \cap W} (\alpha(\theta)^{-k} e^{-k\langle \theta, \tau(x) \rangle}) \\ = \quad & \left(\alpha(\theta)^{-k} \exp\left(-k \left\langle \theta, \frac{t(\mathbf{x} \cap W)}{n(\mathbf{x} \cap W)} \right\rangle \right) \right)^{n(\mathbf{x} \cap W)} \, . \end{aligned}$$

Taking the logarithm and differentiating with respect to θ, we again get the vector equation (4.13). □

The existence and uniqueness of a solution $\widehat{\theta}_0$ to (4.13) have been studied in [35] in a closely related set up. The same type of arguments applies here. Using (4.16), it is seen that

$$\left\{ \frac{\alpha(\theta)^{-k}}{\mathcal{H}^k(W)} e^{-k\langle \theta, \tau(u) \rangle} : \theta \in \Theta \right\} \tag{4.17}$$

is an exponential family of densities on W, with respect to $\mathcal{H}^k$ and with the canonical parameter $-k\theta$. If the family is regular, then the function m in (4.14) is a bijection of $\{-k\theta : \theta \in \Theta\}$ on int S where S is the convex support of the family, cf. Section 2.2. Thus we get the following theorem.

Theorem 4.4 *Let X be a homogeneous template process defined on $\mathcal{X} \in \mathbb{B}^k$ and X_c a locally scaled point process on $W \in \mathbb{B}^k$ derived from X by the exponential scaling function (4.11) and suppose $n(X_c) > 0$ almost surely. Suppose that the densities (4.17) constitute a regular exponential family and denote by S the convex support of the family. Then the function m given by (4.14) is a bijection of $\Theta' = \{-k\theta : \theta \in \Theta\}$ on int S. For $t(\mathbf{x} \cap W)/n(\mathbf{x} \cap W) \in$ int S, there exists a unique solution $\widehat{\theta}_0$ of (4.13).*

For the special case $\tau(u) = u$ we can solve the estimation equation (4.13) explicitly.

Example 4.5 Let $\tau(u) = u$ and $W = [0,1]^k$. Then, $\Theta = \mathbb{R}^k$

$$\begin{aligned} c_\theta(u) &= \alpha(\theta) e^{\langle \theta, u \rangle}, \\ \alpha(\theta) &= \left(\prod_{i=1}^{k} \frac{1 - e^{-k\theta_i}}{k\theta_i} \right)^{1/k}, \end{aligned}$$

and $m(\theta) = (m_1(\theta), \ldots, m_k(\theta))$ where

$$m_i(\theta) = \frac{1 - e^{-k\theta_i} - k\theta_i e^{-k\theta_i}}{k\theta_i(1 - e^{-k\theta_i})}, \quad i = 1, \ldots, k.$$

□

4.4.2 Statistical properties of $\widehat{\theta}_0$

The estimator $\widehat{\theta}_0$ is the maximum likelihood estimator of θ if the template process is Poisson. It is also possible to give theoretical support to the use of $\widehat{\theta}_0$ for general template processes, as shown below.

Theorem 4.5 *Suppose that the intensity function of the locally scaled process X_{c_θ} satisfies*

$$\lambda_{\theta,c}(u) = c_\theta(u)^{-k}\lambda_{0c}, \qquad u \in W. \tag{4.18}$$

Then,

$$\frac{\mathbb{E}_{\theta,c}[t(X_{c_\theta} \cap W)]}{\mathbb{E}_{\theta,c}[n(X_{c_\theta} \cap W)]} = m(\theta) .$$

Proof. Using the Campbell theorem (Theorem 3.1) for X_{c_θ} we have

$$\mathbb{E}_{\theta,c} \sum_{x \in X_{c_\theta}} h(x) = \mathbb{E}h(x)X_{c_\theta}(\mathrm{d}x) = \int_{\mathcal{X}} h(x)\lambda_{\theta,c}(x)\mathcal{H}^k(\mathrm{d}x) .$$

Thus we get

$$\begin{aligned}\mathbb{E}_{\theta,c}[t(X_{c_\theta} \cap W)] &= \mathbb{E}_{\theta,c} \sum_{x \in X_{c_\theta}} \tau(x)\mathbf{1}_W(x) = \int_{\mathcal{X}} \tau(x)\mathbf{1}_W(x)\lambda_{\theta,c}(x)\mathcal{H}^k(\mathrm{d}x) \\ &= \lambda_{0c}\int_{\mathcal{X}} \tau(x)c_\theta(x)^{-k}\mathbf{1}_W\mathcal{H}^k(\mathrm{d}x) = \lambda_{0c}\int_W \tau(x)c_\theta(x)^{-k}\mathcal{H}^k(\mathrm{d}x).\end{aligned}$$

In particular,

$$\mathbb{E}_{\theta,c}[n(X_{c_\theta} \cap W)] = \lambda_{0c}\int_W c_\theta(x)^{-k}\mathcal{H}^k(\mathrm{d}x) . \tag{4.19}$$

The result now follows directly. □

If (4.18) holds, $\widehat{\theta}_0$ can thus be regarded as a moment estimator.

As mentioned in Section 4.2, the equation (4.18) holds if the template process is homogeneous and the scaling function is constant. More interestingly, (4.18) holds for a not necessarily constant scaling function for distance-interaction processes in $\mathbb{R}^1$, remind the Theorem 4.1 . Generally, equation (4.18) is expected to hold approximately if the scaling function varies slowly compared to the interaction radius.

4.4.3 Estimation of the template parameters

Having estimated the scaling parameter θ we can proceed by the estimation of the template process parameters. We will here concentrate on the case where the pseudolikelihood $PL_{W,1}(\widehat{\theta}_0, \psi; \mathbf{x})$ from the decomposition (4.10) is used. In the following we discuss the practical implementation of this method for the locally scaled models. We consider general parametric scaling functions.

Recall that the pseudolikelihood $PL_{W,1}(\theta, \psi; \mathbf{x})$ for the density $f^{(c_\theta)}_{X_{c_\theta}}(\,\cdot\,; \psi)$ with respect to the Poisson point process with intensity measure $\mathcal{H}^k_{c_\theta}$, based on observation in a window $W \subset \mathbb{R}^k$, is defined as follows

$$PL_{W,1}(\theta, \psi; \mathbf{x}) = \exp\left(-\int_W [\lambda^{(c_\theta)}_{\theta,\psi}(u \mid \mathbf{x}) - 1]\,\mathcal{H}^k_{c_\theta}(\mathrm{d}u)\right) \prod_{x \in \mathbf{x} \cap W} \lambda^{(c_\theta)}_{\theta,\psi}(x \mid \mathbf{x}\backslash\{x\}) . \tag{4.20}$$

In the second step of the two-step estimation procedure we fix the scaling parameter θ to $\widehat{\theta}_0$ and maximize $PL_{W,1}(\widehat{\theta}_0, \iota; \mathbf{x}) \exp(-\mathcal{H}^k_{c_{\theta_0}}(W))$ as a function of ι. This can be done in a way similar to the procedure used in the homogeneous case, cf. [3].

We partition W into a finite number of cells C_i, each containing one dummy point u_i, $i = 1, \ldots, l$. The union of the dummy points and the points of the observed pattern is denoted $\{u_j : j = 1, \ldots, m\}$. Furthermore let $C_{i(j)}$ be the unique cell containing u_j, $j = 1, \ldots, m$, with dummy point $u_{i(j)}$. Then we approximate the integral in the pseudolikelihood by

$$\int_W \lambda^{c_{\theta_0}}_{\theta_0,\iota}(u \mid \mathbf{x}) \mathcal{H}^k_{c_{\theta_0}}(\mathrm{d}u) \approx \sum_{j=1}^{m} \lambda^{c_{\theta_0}}_{\theta_0,\iota}(u_j \mid \mathbf{x} \backslash \{u_j\})\, w_j \,,$$

where

$$w_j = \frac{\mathcal{H}^k(C_{i(j)})}{c_{\widehat{\theta}_0}(u_{i(j)})^k} \frac{1}{(1 + n(\mathbf{x} \cap C_{i(j)}))} \approx \mathcal{H}^k_{c_{\theta_0}}(C_{i(j)}) \frac{1}{(1 + n(\mathbf{x} \cap C_{i(j)}))} \,. \tag{4.21}$$

Here, $n(\mathbf{x} \cap C_{i(j)})$ is the total number of the observed points in the cell $C_{i(j)}$.

$\mathcal{H}^k(C_{i(j)}) \,/\, c_{\widehat{\theta}_0}(u_{i(j)})^k$ approximates $\mathcal{H}^k_{c_{\theta_0}}(C_{i(j)})$ if the cells $C_{i(j)}$ are sufficiently small, such that the scaling function c is approximately constant in $C_{i(j)}$. Let us denote

$$\lambda^{c_{\theta_0}}_{\theta_0,\iota}(u_j \mid \mathbf{x} \backslash \{u_j\})$$

by λ_j, $j = 1, \ldots, m$. The pseudolikelihood can then be approximated as a weighted likelihood of independent Poisson variables y_j with means λ_j and weights w_j

$$\log(PL_{W,1}(\widehat{\theta}_0, \iota; \mathbf{x}) \exp(-\mathcal{H}^k_{c_{\theta_0}}(W))) \approx \sum_{j=1}^{m} (y_j \log \lambda_j - \lambda_j)\, w_j \,, \tag{4.22}$$

$$y_j = \frac{1}{w_j} \mathbf{1}\{u_j \in \mathbf{x}\} \,, \quad j = 1, \ldots, m \,. \tag{4.23}$$

When the conditional intensity $\lambda^{c_{\theta_0}}_{\theta_0,\iota}$ is of exponential family form, (4.22) can easily be maximized, using standard software for generalized linear models.

4.5 Simulation study

Since it is very hard to deduce further theoretic results for the two-step estimation procedure proposed in Section 4.4, a simulation study was carried out in order to further investigate the properties of the derived estimators.

The simulation experiment concerns the exponentially scaled Strauss point process with the scaling function

$$c_\theta(u) = \sqrt{\frac{1 - \mathrm{e}^{-2\theta}}{2\theta}}\, \mathrm{e}^{\theta u_1} \,, \quad u = (u_1, u_2) \in \mathbb{R}^2 \,, \tag{4.24}$$

observed on the unit square $W = [0,1]^2$. We used four different values of the inhomogeneity parameter $\theta \in \{0.25, 0.5, 1, 1.5\}$. Note that $\theta = 1.5$ represents quite strong inhomogeneity, compare with Figure 4.2.

For the template Strauss process we fixed the interaction radius R to 0.05 and used a dense set of γ-values in $\{0.01, 0.02, \ldots, 1.00\}$. For β, we used the two values of 250 and 100 to investigate the influence of the total intensity.

For each combination of the parameters, 1000 point patterns were generated using MCMC and the distribution of θ_0 was approximated by the empirical distribution from the 1000 realisations.

All figures relating to the simulation results are collected at the end of this chapter.

In Figure 4.9, the empirical mean values of the statistic $\frac{t(\mathbf{x})}{n(\mathbf{x})}$ are plotted. Full drawn lines correspond to $\beta = 250$, the dashed ones to $\beta = 100$. In each plot the mean value of the statistic under the Poisson model (i.e. $m(\theta)$) is indicated by the dashed horizontal line. We can see that apart from the fluctuations caused by the approximation with empirical average over realisations the mean of $\frac{t(\mathbf{x})}{n(\mathbf{x})}$ is constant and equal to $m(\theta)$ independently of the parameters of the template process.

In Figure 4.10, the estimated standard deviations of $\frac{t(\mathbf{x})}{n(\mathbf{x})}$ are plotted. We get a very slight dependence on the interaction parameter γ and as expected big difference in variance for different values of β. Both effects can be explained by the mean number of observed points which strongly influences the stability of the ratio statistic $\frac{t(\mathbf{x})}{n(\mathbf{x})}$.

Next we study the distribution of $\widehat{\theta}_0$. Since the function m defined by (4.14) is concave and $\frac{t(\mathbf{x})}{n(\mathbf{x})}$ was found to be approximately unbiased, $\widehat{\theta}_0$ tends to overestimate θ. This can be seen in Figure 4.11 for $\theta = 1.5$ and 1, however the relative bias is not larger than 1% and again it does not depend on the interaction parameter γ. The 95% envelopes for $\widehat{\theta}_0$ are also shown in Figure 4.11 and for reasonably high number of observed points (i.e. $\beta = 250$) the inhomogeneity is reliably detected by $\widehat{\theta}_0$. Notice for example that for $\theta = 1$, 95% of the estimates $\widehat{\theta}_0$ falls into the interval $[0.75, 1.25]$ and even for $\theta = 0.25$ – an inhomogeneity often hardly recognizable from the realizations, 95% of the $\widehat{\theta}_0$ estimates are larger that zero.

For sake of completeness the plots of the empirical standard deviation of $\widehat{\theta}_0$ are provided in Figure 4.12. Similar conclusions as for the statistic $\frac{t(\mathbf{x})}{n(\mathbf{x})}$ can be made here. Moreover since $\frac{t(\mathbf{x})}{n(\mathbf{x})}$ is approximately unbiased and independent of γ and m can be approximated closely by a linear function on the intervals given by the 95% envelopes for $\widehat{\theta}_0$ we can expect that the variance of $\widehat{\theta}_0$ can be approximated by

$$\mathrm{Var}_{\theta,\gamma}(\widehat{\theta}_0) \approx \frac{1}{m'(\theta)^2} \mathrm{Var}_{\theta,\gamma} \left(\frac{t(X_c)}{n(X_c)} \right). \tag{4.25}$$

In Figure 4.13 both the standard deviations of the estimate of $\widehat{\theta}_0$ (full drawn line) and the approximation using the statistic $\frac{t(\mathbf{x})}{n(\mathbf{x})}$ (circles) are plotted and we see that the approximation is very precise.

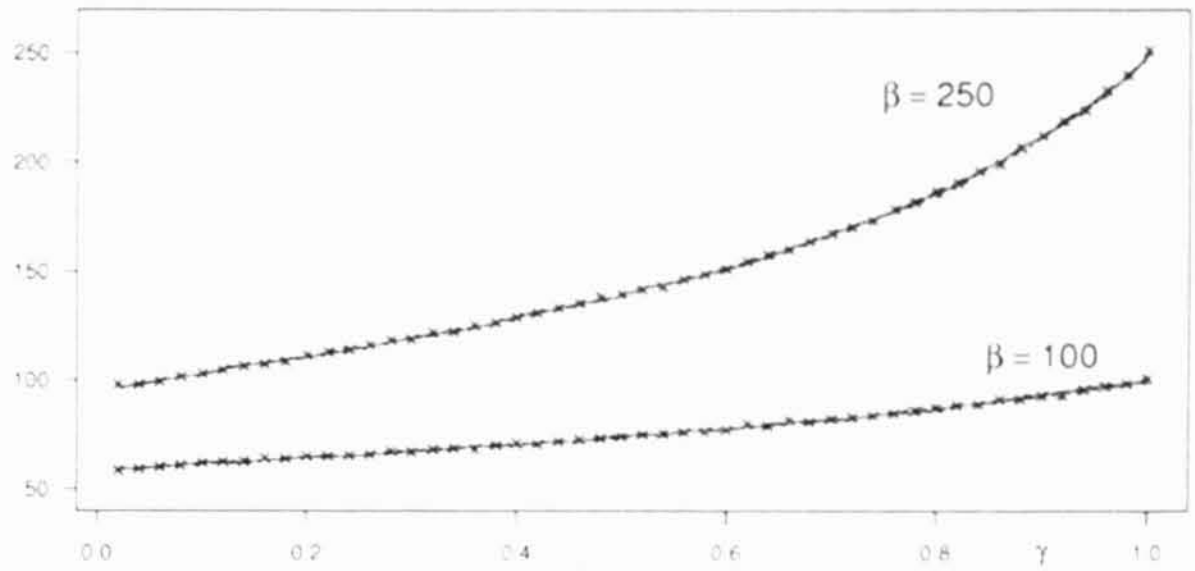


Figure 4.4: Comparison of the intensities $\mathbb{E}_{\theta,\gamma}\, n(X_{c\theta} \cap W)$ (estimated by averages over 200 realizations) on the unit square W for the exponentially scaled Strauss point process $X_{c\theta}$ with different $\theta \in \{0.25, 0.5, 1, 1.5\}$ (*dashed, chain-dotted, full and dotted curves*) and the intensity of the template processes $\lambda_{0\gamma}\mathcal{H}^k(W)$ (*crosses*) in dependence on the template parameter γ. Results are showed for two different values of the template parameter $\beta = 100$ and 250 and the same $R = 0.05$ for all the processes. Since the differences among all the four curves and the crosses are hardly distinguishable, we got a perfect agreement with equation (4.26).

Note that since the scaling function has been normalized as in (4.15), (4.18) implies that

$$\mathbb{E}_{\theta,\gamma}\, n(X_{c\theta} \cap W) = \lambda_{0\gamma}\mathcal{H}^k(W)\,, \tag{4.26}$$

i.e. the mean number of points in W does not depend on the inhomogeneity parameter θ. Since (4.18) does not hold exactly for the Strauss process, we investigated whether (4.26) holds approximately, using the simulated data. The approximation is excellent in this example, cf. Figure 4.4.

Let us next study the estimation of the template parameters. The density of the Strauss process is of exponential family form with one nuisance parameter R – the interaction radius (see Example 4.3). Thus $\psi = (\beta, \gamma, R)$ and

$$\log \lambda_j = \log \beta + s_c(u_j; \mathbf{x}) \log \gamma\,,$$

where

$$s_c(u_j; \mathbf{x}) = \sum_{x \in \mathbf{x}\setminus\{u_j\}} \mathbf{1}\{\mathcal{H}^1_c([u_j, x]) \le R\}\,.$$

To find the estimate of ψ we have to compute and compare the profile pseudolikelihood

$$\overline{PL}_{W,1}(R) = \max_{\beta,\gamma} PL_{W,1}(\widehat{\theta}_0, \beta, \gamma, R; \mathbf{x})$$

on a grid of values of R. We let β_R and γ_R be the values of β and γ at which

$$PL_{W,1}(\widehat{\theta}_0, \cdot, \cdot, R; \mathbf{x})$$

is maximal (the subscript R indicates the dependence on R).

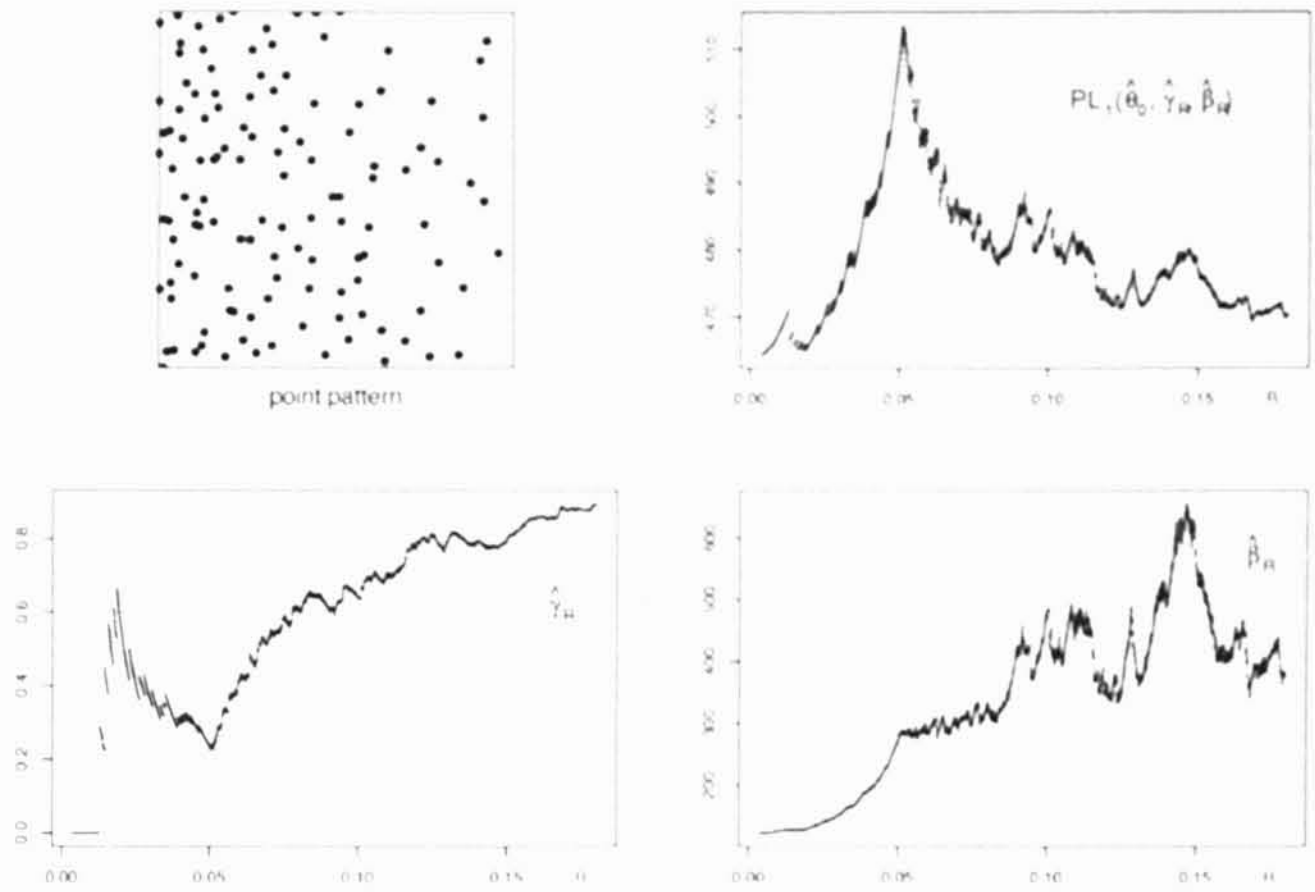


Figure 4.5: Profile pseudolikelihood estimation of the template parameters β, γ, R of a simulated exponentially scaled Strauss process on $[0,1]^2$ without any border correction. The first picture shows the data $\mathbf{x}$. In the three graphs, the profile pseudolikelihood and the corresponding estimates $\hat{\beta}_R$ and $\hat{\gamma}_R$ are plotted as functions of R. The final estimates $\hat{R} = 0.0508$, $\hat{\beta} = 283$, $\hat{\gamma} = 0.23$ are indicated by the dotted lines. The true values are $R = 0.05$, $\beta = 250$, $\gamma = 0.25$.

In Figures 4.5 and 4.6 we illustrate the procedure on a simulated exponentially scaled Strauss point pattern with the scaling function (4.24) and parameters

$$\theta = 1, \ \beta = 250, \ \gamma = 0.25, \ R = 0.05, \ W = [0,1]^2.$$

The parameter θ has been fixed to the correct value and a regular grid of 100×100 dummy points was used.

In the plots presented in Figures 4.5 and 4.6, the profile pseudolikelihood and the estimates $\hat{\beta}_R$ and $\hat{\gamma}_R$ are plotted as functions of the nuisance parameter R. The jaggedness of the plots is due to the discontinuity of the interpoint distance function s_c as a function of R. In Figure 4.5 we used no border correction (the pseudolikelihood (4.20) with $\mathbf{x}$ replaced by $\mathbf{x} \cap W$) while in Figure 4.6 we used a border correction of $\mathcal{H}_c^1 = 0.05$ (the psedolikelihood (4.20) with W replaced by an irregular observation window $\widetilde{W} = \{u \in W : \mathcal{H}_c^1(u, \partial W) > 0.05\}$, where ∂W denotes the boundary of W).

The obtained estimates of ψ are in good agreement with the true values, especially the estimate of the interaction radius is very precise. It is also important that the estimates with and without border correction do not differ substantially (which is probably caused by the sufficiently large number of observed points in W).

The results concerning pseudolikelihood estimation were confirmed in repeated simulation experiments.

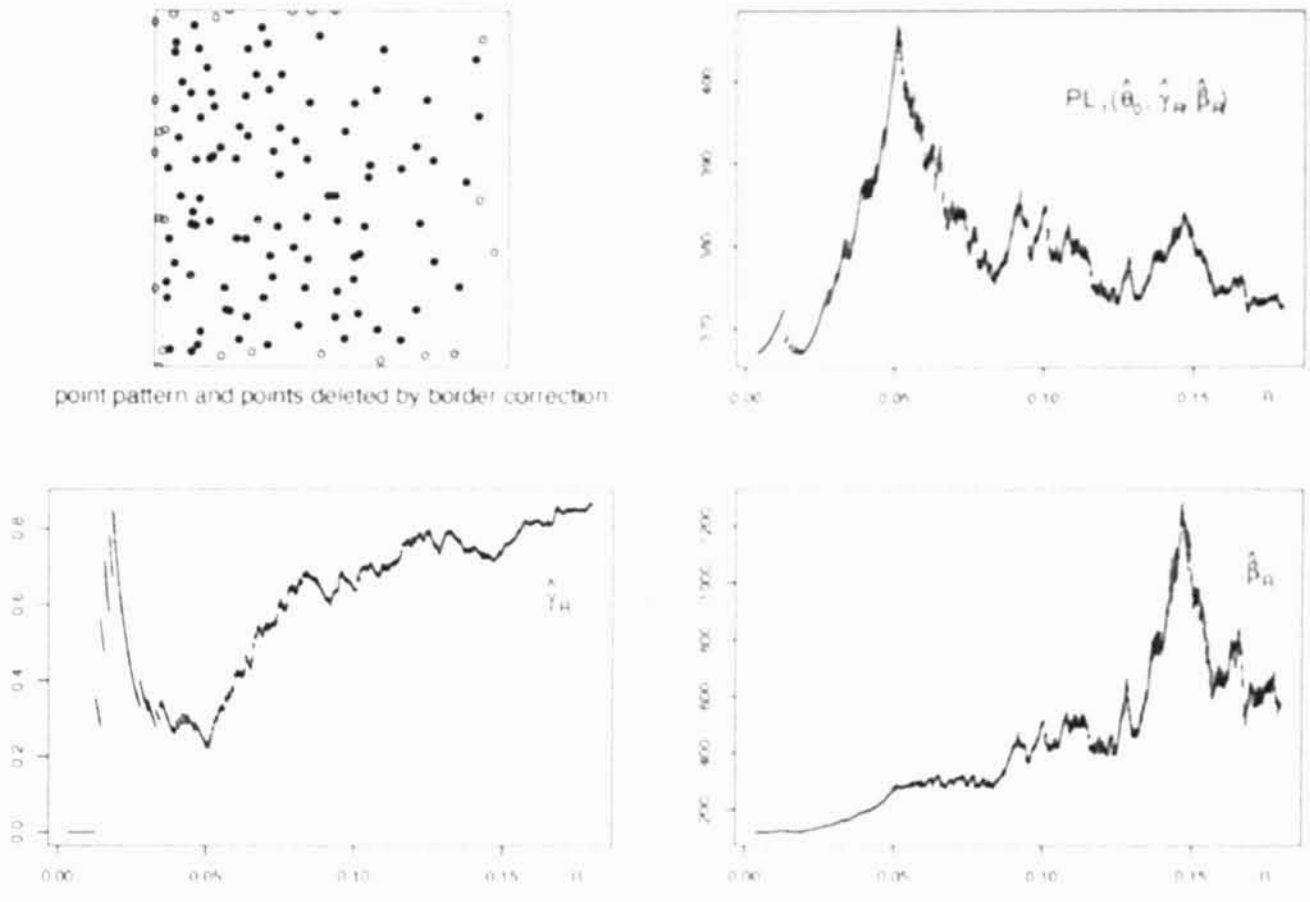


Figure 4.6: Profile pseudolikelihood estimation of the template parameters β, γ, R of a simulated exponentially scaled Strauss process on $W = [0,1]^2$ with border correction $\mathcal{H}_c^1 = 0.05$. The first picture shows the same data used in Figure 4.5. The full circles are the data points used for the estimation. The three graphs are constructed as in Figure 4.5. The obtained estimates are $\hat{R} = 0.0508$, $\hat{\beta} = 282$, $\hat{\gamma} = 0.22$. The true values are $R = 0.05$, $\beta = 250$, $\gamma = 0.25$.

4.6 Two step inference where scaling function is estimated using other (non ML) methods

Going one step further, one could also estimate c in some other way from the local intensity

$$\lambda_c(u) = \mathbb{E}\, \lambda_c(u \mid X_c) \,,$$

$u \in \mathcal{X}$, of the locally scaled process X_c, using the approximate relation

$$\lambda_c(u) \approx c(u)^{-k} \lambda_0 \,, \tag{4.27}$$

where λ_0 is the intensity of the template process. In order to estimate the scaling function, we could use an estimate $\hat{\lambda}_c(u)$ of the local intensity, and set

$$\hat{c}(u) = \left[C \hat{\lambda}_c(u) \right]^{-1/k} \,,$$

where $C = 1/\lambda_0$ is some constant that can be arbitrarily fixed. For convenience, one may choose $C = 1$.

If, in a parametric setting, $\hat{\lambda}_c(u)$ is the maximum likelihood estimator of the intensity of an inhomogeneous Poisson process, then $\hat{c}(u)$ is the same partial MLE as the one based on L_0. On the other hand, $\lambda_c(u)$ can also be estimated nonparametrically, for example, using kernel methods or Voronoi tessellations (see e.g. [2]). Or it can be estimated parametrically by other methods than maximum likelihood, e.g. regression methods.

4.7 Model validation

Since the two-step estimation procedure, suggested in Section 4.4, can only be justified theoretically in special cases, it is particularly important to develop effective procedures for model validation. This is the topic of the present section. For non-Poisson point processes only little is known about the theoretical distribution of characteristics that can be used for model validation. Therefore tests are usually simulation based.

In the case of homogeneous point processes, the probably most popular tests are based on second order summary statistics such as Ripley's K-function, see e.g. [12] and [32]. In [2] the definition of the K-function and other second order characteristics are extended to a certain type of inhomogeneous point processes, so called second order intensity reweighted stationary processes. This class of processes comprises inhomogeneous processes obtained by independent thinning, but does not include locally scaled processes (apart from the Poisson process).

In [17] is considered another so-called Q^2 statistic for tests against the Poisson process hypotheses, which simply spoken amounts to a goodness-of-fit test for the frequency distribution of number of neighbours in an r-neighbourhood. In this section we investigate how K-functions and Q^2-statistics can be adapted to locally scaled inhomogeneous point processes.

4.7.1 The K-function

Let us remind here the Definition 3.17 which tells us that the K-function for a stationary point process X is defined as the expected number of points in $X \setminus \{o\}$, given $o \in X$, in a ball of radius r around o, divided by the intensity λ_0 of X

$$K_0(r) = \frac{\mathbb{E}_0^! n(X \cap B(o,r) \setminus \{o\})}{\lambda_0} .$$

(In the case of finite processes with interaction, we assume that the domain of X is large enough that $X \cap B(o,r)$ is virtually free of boundary effects.) By the Campbell theorem, a ratio-unbiased estimator of $K_0(r)$ is given by

$$\widehat{K}_0(W,r) = \frac{1}{\lambda_0 n(\mathbf{x} \cap W)} \sum_{x \in \mathbf{x} \cap W} \sum_{y \in \mathbf{x} \setminus \{x\}} \mathbf{1}\{\mathcal{H}^1([x,y]) \le r\} ,$$

where $\mathbf{x}$ is an observed point pattern from the stationary point process X. If instead a locally scaled point pattern $\mathbf{x}$ is observed, we suggest to use a locally scaled analogue of $\widehat{K}_0(W,r)$, viz.

$$\widetilde{K}_0(W,r) = \frac{1}{\lambda_0 n(\mathbf{x} \cap W)} \sum_{x \in \mathbf{x} \cap W} \sum_{y \in \mathbf{x} \setminus \{x\}} \mathbf{1}\{\mathcal{H}_c^1([x,y]) \le r\} . \tag{4.28}$$

Note that $\widetilde{K}_0(W,r)$ is ratio-unbiased for $K_0(r)$ if c is constant. Furthermore, $\widetilde{K}_0(W,r)$ is ratio-unbiased also for processes in $\mathbb{R}^1$.

Theorem 4.6 *$\widetilde{K}_0(W,r)$ is ratio-unbiased for general scaling functions and distance interaction point processes defined on an interval I of $\mathbb{R}^1$.*

Proof. In $\mathbb{R}^1$ a locally scaled distance-interaction process X_c has the same distribution as $h(X)$ where h is a 1-1 differentiable transformation of I onto I with $(h^{-1})' = c^{-1}$ (see Theorem 4.1). Therefore, we have

$$\begin{aligned}
&\mathbb{E}\Big(\sum_{x\in X_c\cap W}\sum_{y\in X_c\setminus\{x\}} \mathbf{1}\{\mathcal{H}^1_c([x,y]) < r\}\Big)\\
&= \mathbb{E}\Big(\sum_{x\in h(X)\cap W}\sum_{y\in h(X)\setminus\{x\}} \mathbf{1}\{\mathcal{H}^1([h^{-1}(x),h^{-1}(y)] \le r\}\Big)\\
&= \mathbb{E}\Big(\sum_{x\in X\cap h^{-1}(W)}\sum_{y\in X\setminus\{x\}} \mathbf{1}\{\mathcal{H}^1([x,y]) < r\}\Big)\,.
\end{aligned}$$

Accordingly, the ratio-unbiasedness of $\widetilde{K}_0$ follows from the ratio-unbiasedness of $\widehat{K}_0$. □

Generally, $\widetilde{K}_0$ is expected to be (approximately) ratio-unbiased if r is small such that c varies little in the scaled neighbourhood. In any case, one should use simulations of the scaled null hypothesis model, not only of the template, for model validation.

A further simplification is accomplished by applying $\mathcal{H}^1_c([x,y]) \approx 2/(c(x)+c(y)) \times \mathcal{H}^1([x,y])$, which was introduced for distance-interaction processes in [18]. The corresponding statistic

$$\widetilde{K}_0(W,r) = \frac{1}{\lambda_0 n(\mathbf{x}\cap W)}\sum_{x\in\mathbf{x}\cap W}\sum_{y\in\mathbf{x}\setminus\{x\}} \mathbf{1}\{\mathcal{H}^1([x,y]) \le \tfrac{1}{2}(c(x)+c(y))r\}\,, \tag{4.29}$$

is particularly useful if c is estimated nonparametrically, because it requires evaluation of c only in the data points.

In practical situations, both λ_0 and c have to be estimated from the data. As discussed in the preceeding sections, the estimation of c cannot be separated from the estimation of λ_0. Since the template is unique only up to a constant scale factor which determines λ_0, the scaling function c is unique only up to a constant as well.

We suggest to normalize c such that $\mathcal{H}^k_c(W) = \mathcal{H}^k(W)$, see (4.15). Thus, we set

$$\hat{\lambda}_0 := n(\mathbf{x}\cap W)/\mathcal{H}^k(W),$$

since

$$\mathbb{E}n(X_c\cap W) = \int_W \lambda_c(x)\,\mathcal{H}^k(\mathrm{d}x) \approx \int_W \lambda_0 c(x)^{-k}\,\mathcal{H}^k(\mathrm{d}x) = \lambda_0\,\mathcal{H}^k(W).$$

4.7.2 The Q^2 statistic

The Q^2-statistic proposed in [17] is (in the simplest case) based on the numbers $M_\ell(W, r)$ of points in W with ℓ r-close neighbours, $\ell = 0, 1, \ldots, q$. For a homogeneous Poisson point process, the expectation μ and the covariance matrix Σ of the vector $M = (M_0, M_1, \ldots, M_q)^\top$ can be calculated (see [17]). A finite range dependency argument is used to show that the statistic

$$Q^2 = (M - \mu)^\top \Sigma^{-1} (M - \mu)$$

(squared Mahalanobis distance) is asymptotically χ^2-distributed for increasing size of the observation window W. By simulation experiments, [17] showed that Q^2 discriminates well between patterns from a mixed cluster and regular point process and the Poisson process.

Since μ and Σ can also be calculated for an inhomogeneous Poisson point process, it would be possible to use the same Q^2-statistic also for tests of inhomogeneous Poisson processes. However, the expected number of neighbours in a ball of radius r around a point x would depend on the local intensity $\lambda(x)$. Hence, inhomogeneity introduces much extra variation to M which would largely cut down the diagnostic value of Q^2.

This effect can be avoided by adjusting r to the local intensity. We propose to replace the Euclidean neighbour distance by the locally scaled neighbour distance. In an inhomogeneous Poisson point process with intensity $\lambda_c(x) = c(x)^k \lambda_0$, the number of r-scaled-close neighbours of a point x is Poisson distributed with parameter $\lambda_0 \mathcal{H}^k_c(B_c(x, r))$. Since

$$\frac{\mathcal{H}^k_c(B_c(x, r))}{\mathcal{H}^k(B(x, r))} \to 1 \quad \text{as} \quad r \to 0 ,$$

the distribution of r-scaled-close neighbours does hardly depend on the location for small r, and is close to the distribution of r-close neighbour number in the homogeneous case.

The local scaling analogue of M_ℓ is

$$M_{\ell\,\mathrm{inhom}}(W, r) = \sum_{x \in \mathbf{x} \cap W} \mathbf{1}\{n(B_c(x, r) \cap \mathbf{x} \setminus \{x\}) = \ell\} . \tag{4.30}$$

Since calculation of μ and Σ is feasible only for the Poisson point process with slowly varying scaling function, we suggest to do simulation tests. This would allow to test any hypothesis. While any distance between observed and expected neighbour number distribution can be used, we still recommend to use the statistics Q^2, however to replace μ and Σ with estimates obtained by simulation. Note that the simulations for estimating μ and Σ are not to be reused for the test.

4.8 Data analysis

The map shown in Figure 4.1 was recorded in Australian heath. This vegetation is subjected to regular fires, the study area having been last burnt ten years before the collection

of the data [1]. The species under study, *Scholtzia* aff. *involucrata*, is long lived and regenerates from root stock after fire yielding daughter plants that stand close together. Furthermore, seed germinates after fire, with young plants coming up within a distance of at most 2m of the parent plant. However, only very few seedlings survive the dry summer. These facts explain the slight clustering observable in the point pattern of plant locations. The heterogeneous intensity is likely to be due to soil mosaic, affecting mostly the seedlings that are very sensitive to shortage of water.

We therefore need to model attractive interaction among the plants. The exponentially scaled area-interaction model appears to be a good candidate because the area of a location dependent neighbourhood around each plant enters explicitly into the model density.

We used the two-step fitting procedure. For convenience we rescaled the data to the unit square $W = [0,1]^2$. As the pattern exhibits obvious inhomogeneity in the vertical direction but appears quite homogeneous in the horizontal direction we used an exponential scaling function of the form

$$c(u) = \sqrt{\frac{1-e^{-2\theta}}{2\theta}}\, e^{\theta u_2}, \quad u = (u_1, u_2) \in \mathbb{R}^2. \tag{4.31}$$

Based on $L_0(\theta; \mathbf{x} \cap W)$ we obtained the following estimate of θ

$$\widehat{\theta}_0 = 1.0839,$$

with $\alpha(\widehat{\theta}_0) = 0.6391$, see (4.16).

Secondly, we maximized the pseudolikelihood $PL_{W,1}(\theta_0, \psi; \mathbf{x})$ with θ_0 fixed. The density of the area-interaction process is of an exponential family form with one nuisance parameter R – the interaction radius. Like for the Strauss process, $\psi = (\beta, \gamma, R)$ and for the estimation we use the same weights as in (4.21) and

$$\begin{aligned} \log \lambda_j &= \log \beta - \mathcal{H}^2_c(U_{c,R}(u_j; \mathbf{x})) \log \gamma, \\ U_{c,R}(u_j; \mathbf{x}) &= \{y \in W : \mathcal{H}^1_c([y, u_j]) \le R, \mathcal{H}^1_c([y, \mathbf{x} \backslash \{u_j\}]) > R\}. \end{aligned}$$

We used a grid of 100×100 dummy points which were equidistant in the horizontal direction and $\mathcal{H}^1_{c_{\widehat{\theta}_0}}$–equidistant in the vertical direction (actually this means that the dummy points were $\mathcal{H}^1_{c_{\widehat{\theta}_0}}$–equidistant in both directions – compare with (4.31)).

We maximized the profile pseudolikelihood on a grid of R-values. The main problem is the computation of the scaled volumes $\mathcal{H}^2_c(U_{c,R}(u_j; \mathbf{x}))$ for all the points u_j, $j = 1, \ldots, m$. This can be done only approximately. To approximate these scaled volumes with a reasonable precision it is necessary to compute the scaled distances from the points $\{u_j, j = 1, \ldots, m\}$ to each point in a very fine grid of points in W. This job is computationally quite demanding.

The approximate profile pseudolikelihood $PL_{W,1}(\widehat{\theta}_0, \psi; \mathbf{x})$ was computed with border correction $\mathcal{H}^1_c = 0.05$. This degree of border correction was chosen as a compromise between minimizing the bias caused by missing unobserved points and not excluding too many observed points from the estimation (with the chosen border correction one fourth

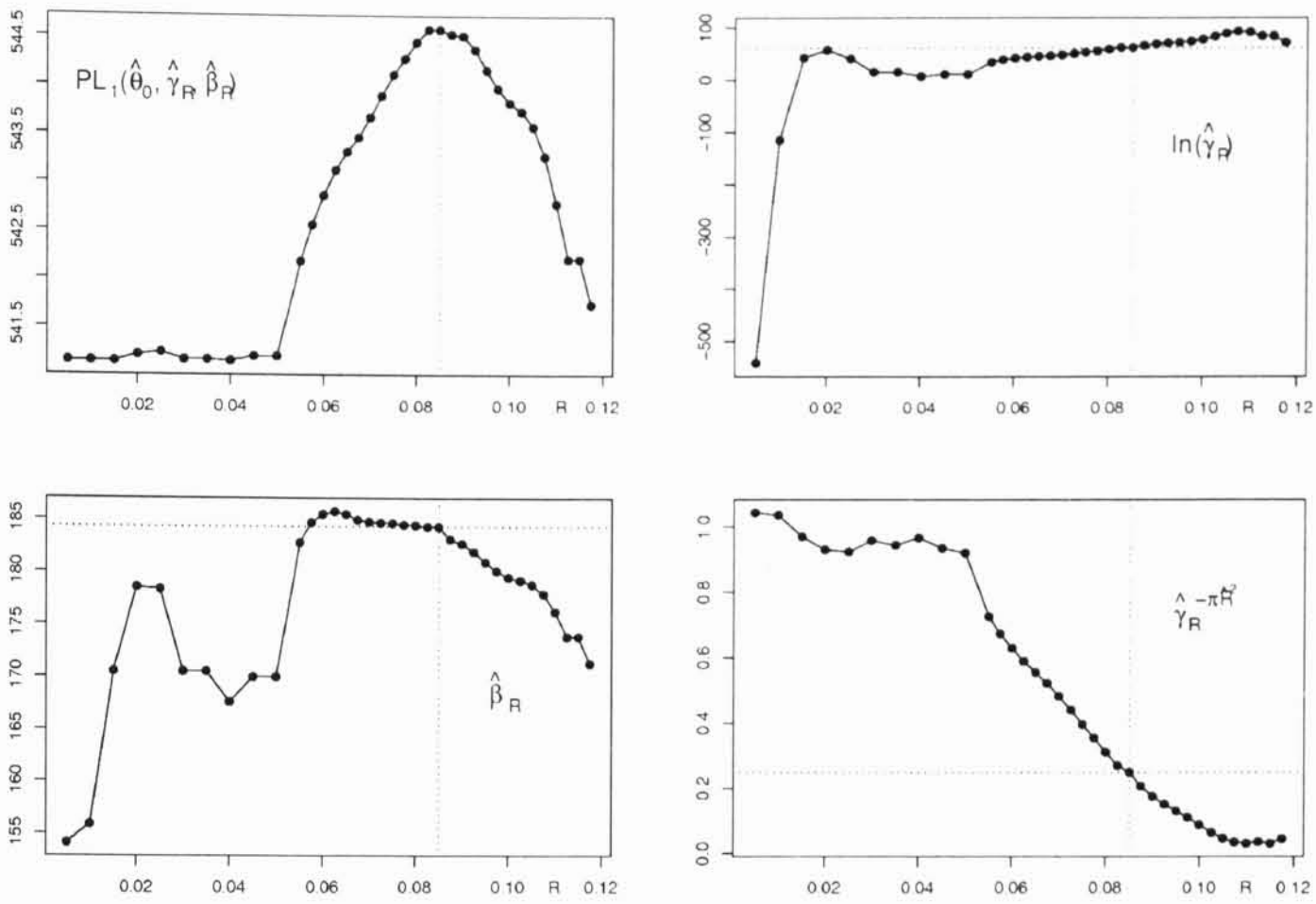


Figure 4.7: Pseudolikelihood estimation of the template parameters β, γ, R of the exponentially scaled area-interaction process for the plant data from Figure 4.1 rescaled to $[0,1]^2$ with border correction $\mathcal{H}_c^1 = 0.05$. The graphs show the profile pseudolikelihood and the corresponding estimates $\hat{\beta}_R$ and $\hat{\gamma}_R$ as functions of R. The last graph of $\widehat{\gamma}^{-\pi\hat{R}^2}$ shows the strength of the attractive interaction. The resulting estimates are $\hat{R} = 0.085$, $\hat{\beta} = 184$, $\hat{\gamma} = 3.99 \cdot 10^{26}$.

of the points was not used in the estimation). The profile pseudolikelihood and estimates of the parameters as functions of R are plotted in Figure 4.7. Note that the curves are smoother than in the case of the Strauss process because now the interaction function is continuous as a function of R. We obtained the following values

$$\widehat{R} = 0.085,\ \widehat{\beta} = 184,\ \widehat{\gamma} = 3.99 \cdot 10^{26},\ \widehat{\gamma}^{-\pi\widehat{R}^2} = 0.25. \tag{4.32}$$

The value of $\widehat{\gamma}^{-\pi\widehat{R}^2}$ is included because it gives a better impression of the strength of the interaction, since this is actually the term which appears in the template density. The fit indicates a slightly clustered point pattern as we expected.

For model validation we used the $\widetilde{K}_0(\widetilde{W}, r)$ and Q^2 statistics from Section 4.7. Figure 4.8 shows the locally scaled estimate $\widetilde{K}_0(\widetilde{W}, r)$ with $\widetilde{W} = \{u \in W : \mathcal{H}_c^1(u, \partial W) > 0.05\}$ (full-drawn line) together with the empirical mean and 95% envelopes for $\widetilde{K}_0(\widetilde{W}, r)$ calculated from 399 simulations under the fitted exponentially scaled area-interaction model (dashed lines). The locally scaled estimate $\widetilde{K}_0(\widetilde{W}, r)$ for the plant data lies inside the envelopes of the fitted area-interaction model.

Next we tested the locally scaled Poisson hypothesis (H_0: the observed point pattern is a realization of a locally scaled Poisson process) on the plant data. We used the

$$Q_{\mathrm{P}}^2 = (M_{\mathrm{inhom}} - \mu_P)^\top \Sigma_P^{-1} (M_{\mathrm{inhom}} - \mu_P)$$

statistic with $r = 0.05$ and $M_{\mathrm{inhom}} = (M_{0\mathrm{inhom}}, \ldots, M_{6\mathrm{inhom}})$ defined by (4.30). The subscript P indicates that in the formula for Q^2 we use as μ and Σ the mean μ_{P} and the

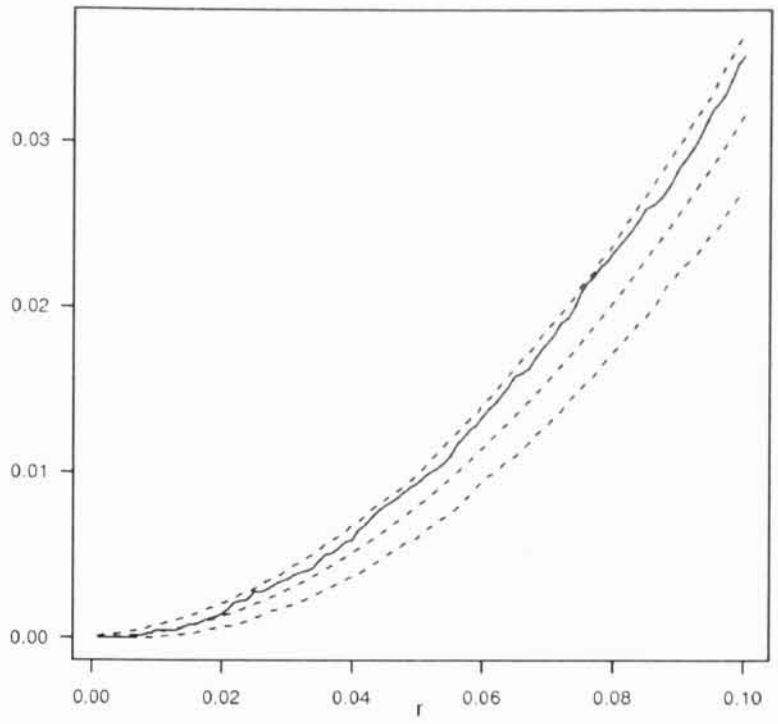


Figure 4.8: The estimate $\widetilde{K}_0$ for the plant data (full drawn line) and mean and 95% envelopes for $\widetilde{K}_0$ for the exponentially scaled area-interaction model (dashed lines).

covariance matrix Σ_{P} of M_{inhom} for the fitted locally scaled Poisson model with $\theta = \widehat{\theta}_0$. The values of μ_{P} and Σ_{P} were estimated from 8000 simulated realizations of the fitted locally scaled Poisson model.

The simulation test (using 399 realizations of the hypothesis locally scaled Poisson model with $\theta = \widehat{\theta}_0$) gives the p-value of 0.05. Thus the plant data is not very well described by the Poisson model.

Then we used the Q^2_{A} statistic (i.e. the mean value μ_{A} and covariance matrix Σ_{A} of M_{inhom} are computed for the fitted exponentially scaled area-interaction model) for testing of the fitted locally scaled area-interaction model (H_0: the observed point pattern is a realization of a locally scaled area-interaction point process). The test gave the p-value of 0.106, thus like with the K-function we do not reject the the locally scaled area-interaction model.

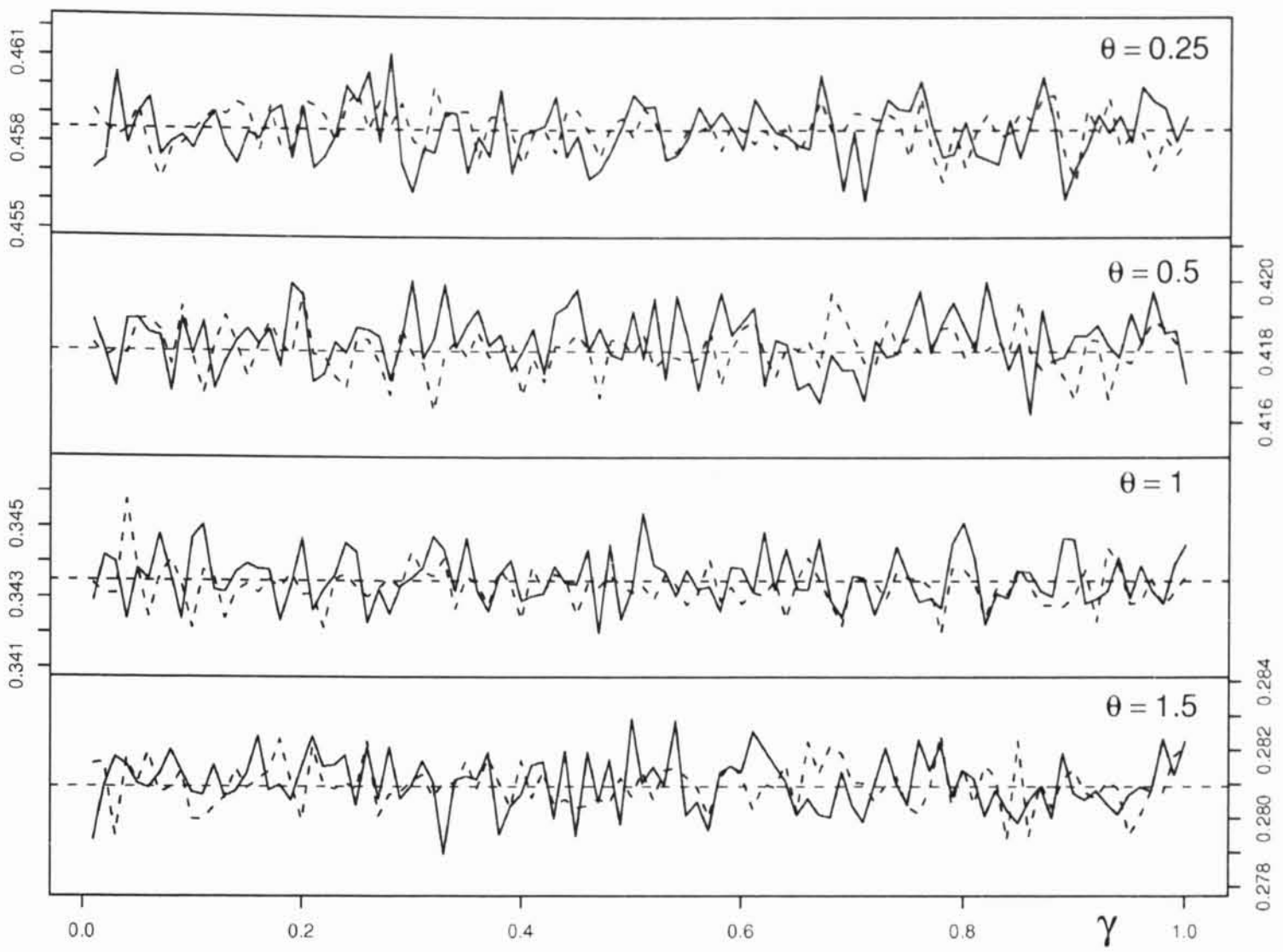


Figure 4.9: Empirical mean values of the canonical statistic $\frac{t(\mathbf{x})}{n(\mathbf{x})}$ for four different values of the inhomogeneity parameter θ (values are indicated in the plots) and for template parameter $\beta = 250$ (full drawn lines) and $\beta = 100$ (dashed lines) as functions of the template parameter γ.

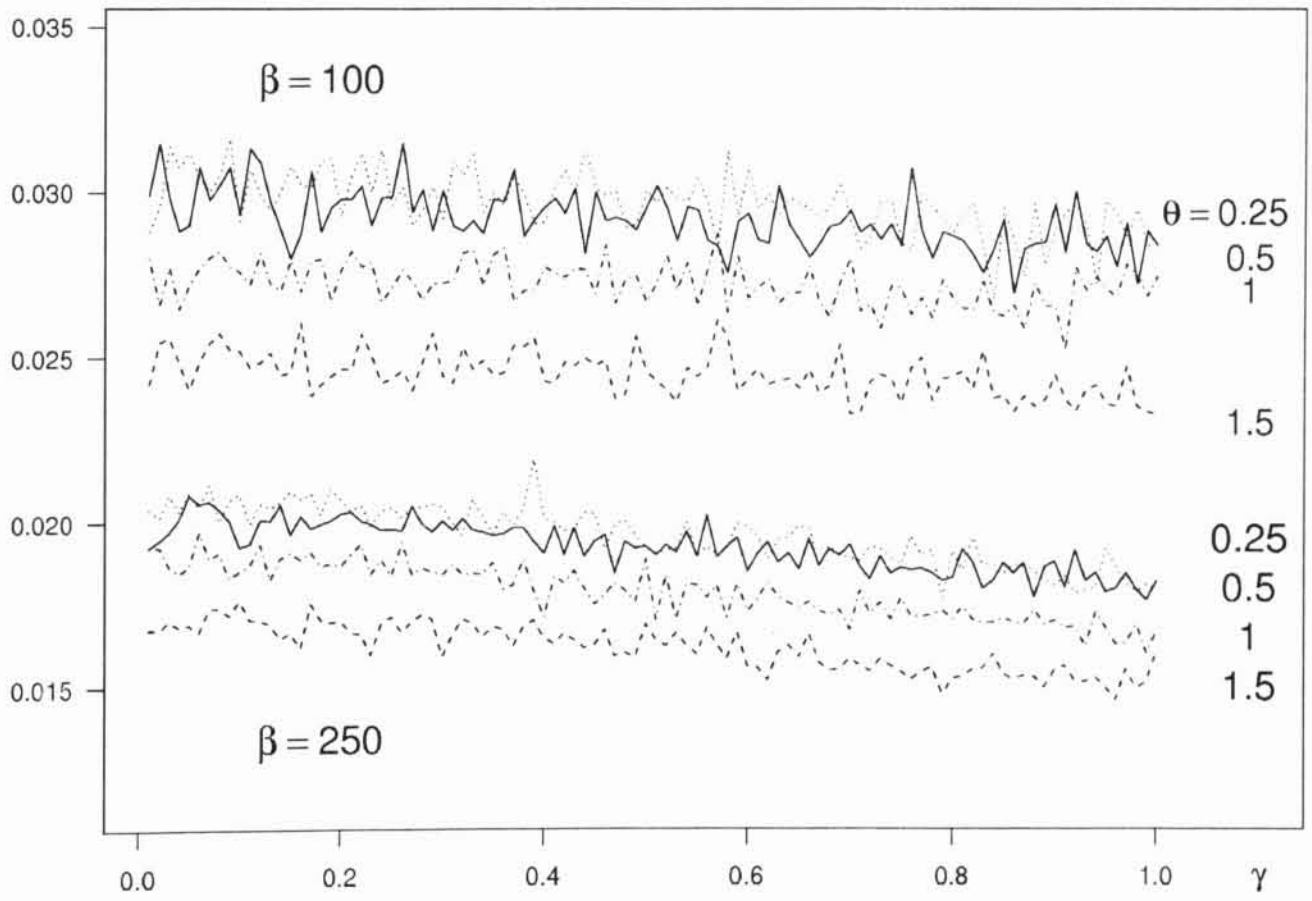


Figure 4.10: Empirical standard deviations of the canonical statistic $\frac{t(\mathbf{x})}{n(\mathbf{x})}$ for the indicated β and θ values, as a function of the interaction parameter γ. Full drawn lines correspond to the inhomogeneity parameter $\theta = 0.5$.

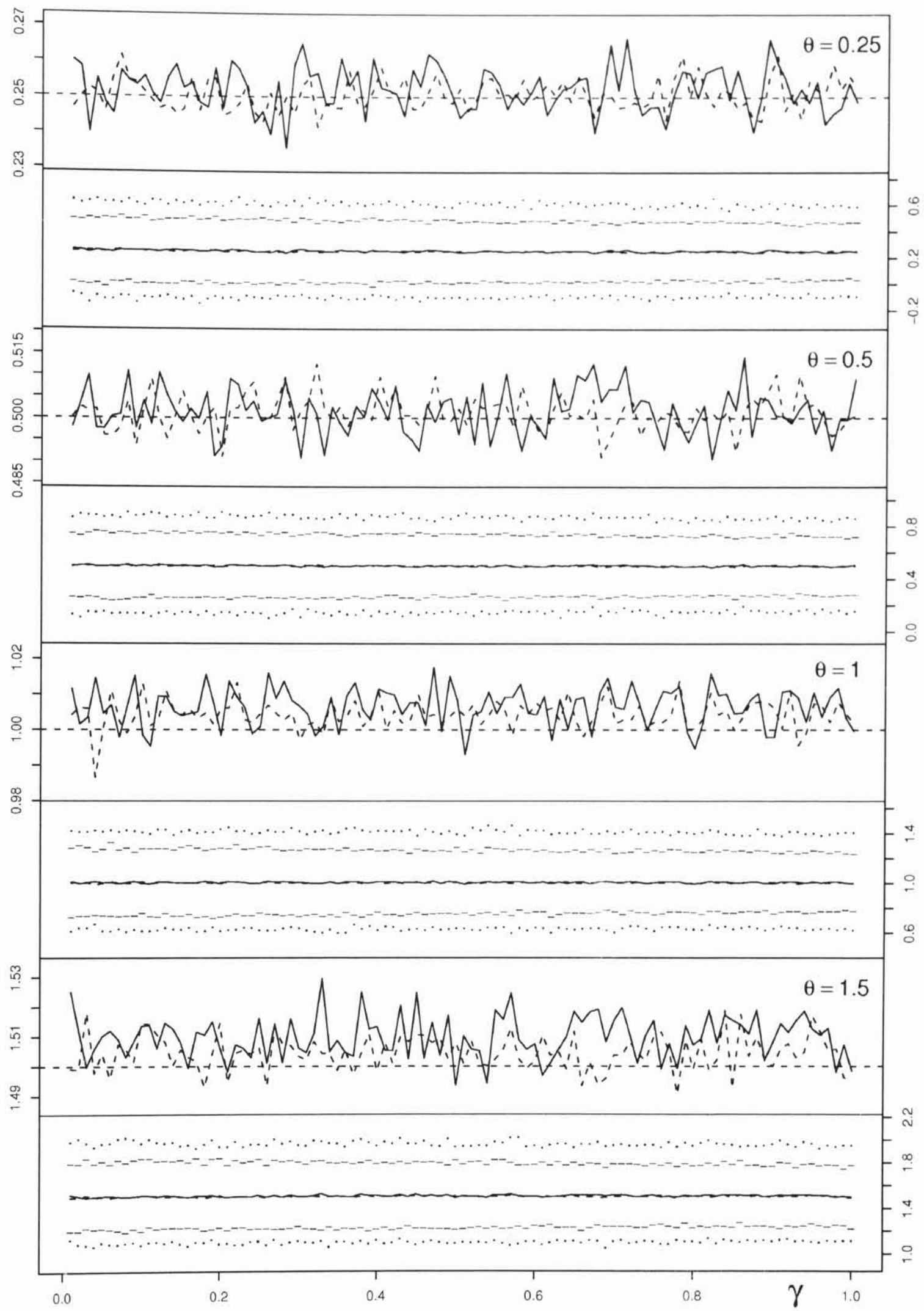


Figure 4.11: Empirical mean values and 95% envelopes for the estimator $\widehat{\theta_0}$ for four different values of the inhomogeneity parameter θ (values are indicated in the plots) and for template parameter $\beta = 250$ (*full drawn lines*, resp. *dashed lines* for envelopes) and $\beta = 100$ (*dashed lines*, resp. *dotted lines* for envelopes), as a function of the template parameter γ. The central lines in the envelope plots are the empirical means again.

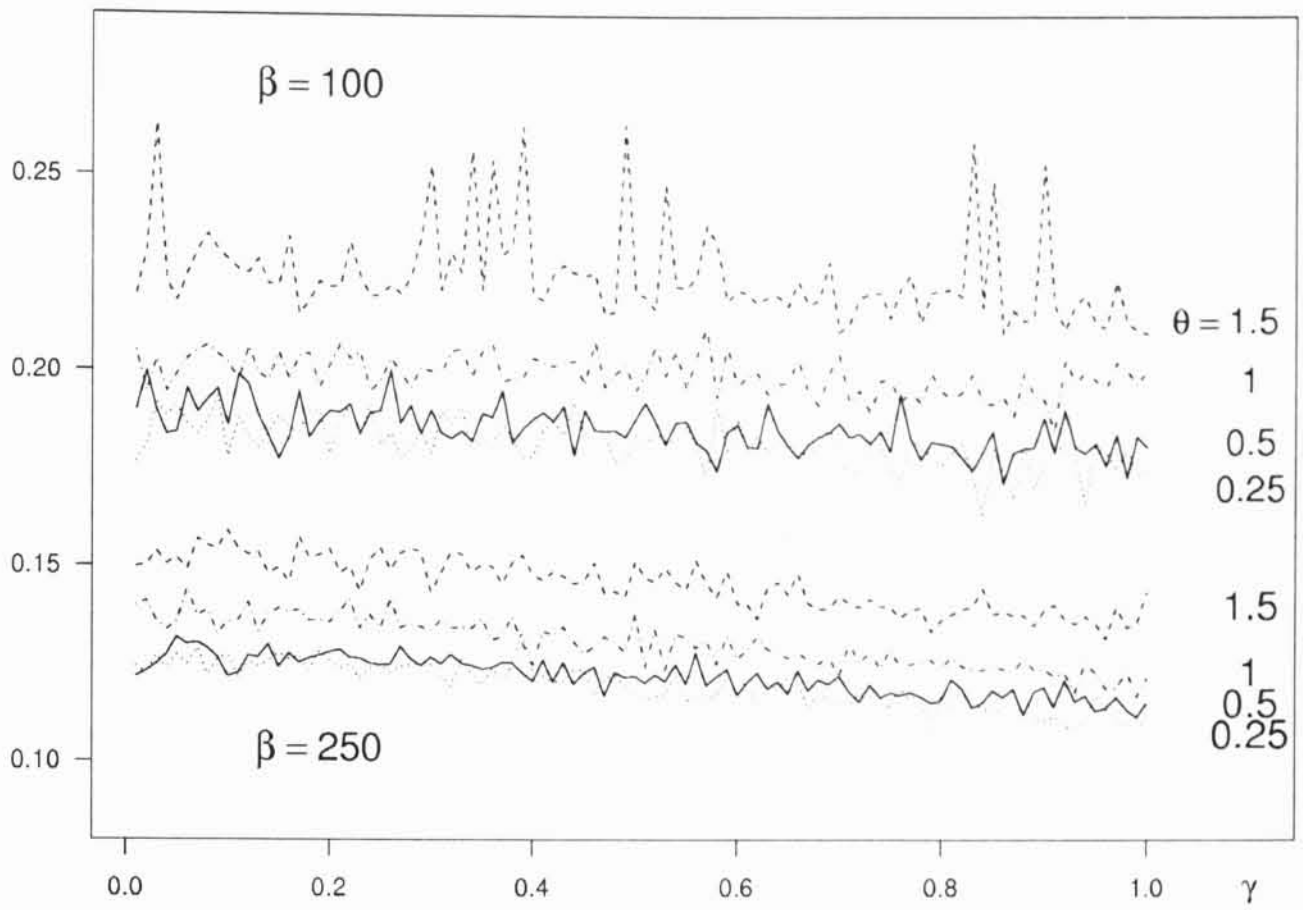


Figure 4.12: Empirical standard deviations of the estimator $\widehat{\theta}_0$ for the indicated values of β and θ as functions of the interaction parameter γ. The full drawn line corresponds to the inhomogeneity parameter $\theta = 0.5$.

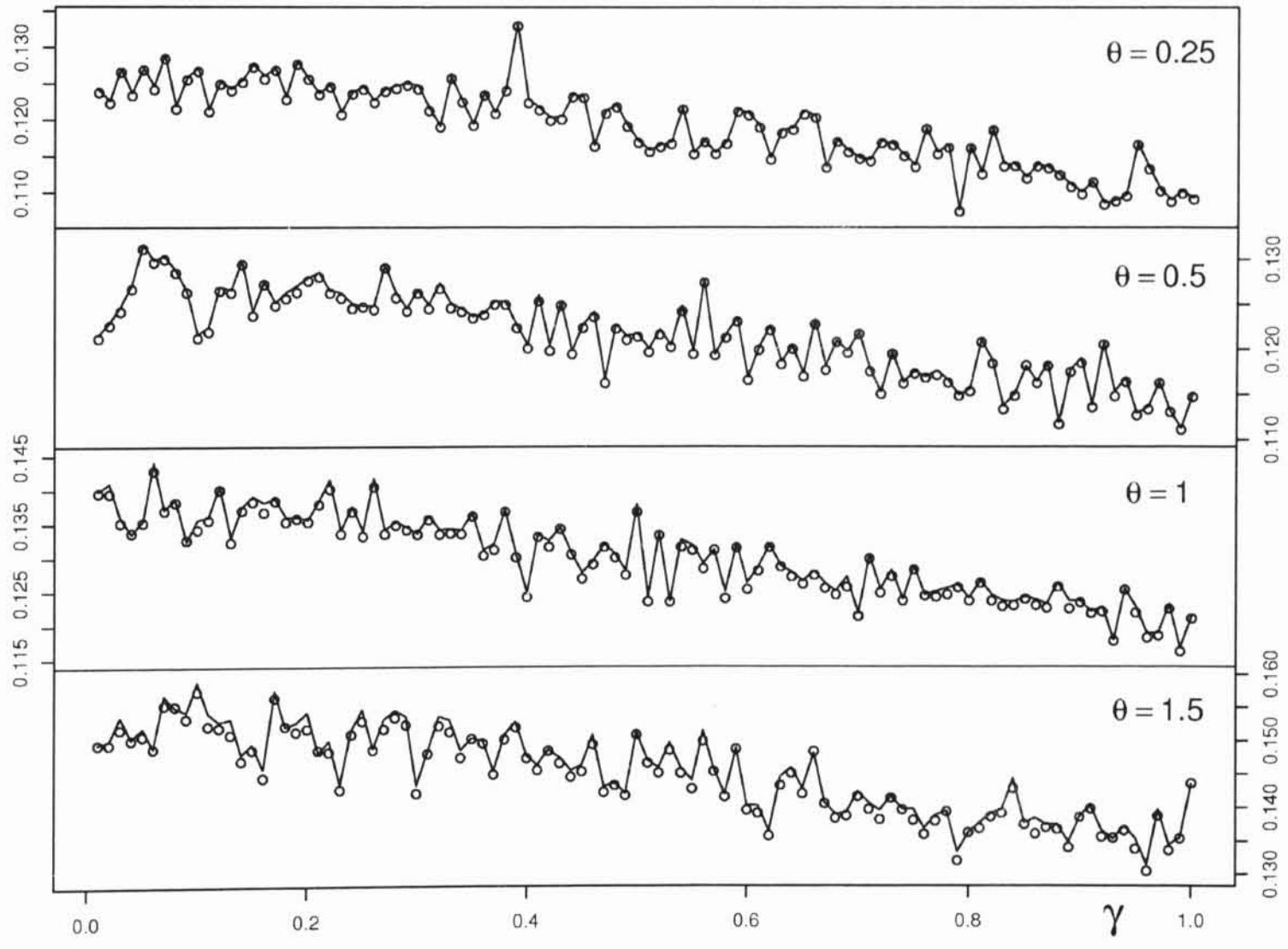


Figure 4.13: Comparison of the standard deviation of the estimator $\widehat{\theta}_0$ (full drawn lines) with the approximation (4.25) using the standard deviation of the canonical statistic $\frac{t(\mathbf{x})}{n(\mathbf{x})}$ (circles). In this case β was 250.

Chapter 5

Estimators of asymptotic variance of stationary point processes

5.1 Introduction

This chapter is devoted to the examination of asymptotic and middle-sample properties of estimators of a quantity σ^2 called the asymptotic variance.

Asymptotic variance σ^2 is defined for simple stationary point processes with intensity $\lambda > 0$ and it is indeed the asymptotic variance of the standard intensity estimator $\widehat{\lambda}$ defined by (3.8). Thus let throughout this chapter Φ denote a simple stationary point process on $\mathbb{R}^k$ with intensity $0 < \lambda < \infty$. The asymptotic variance is defined by

$$\sigma^2 = \lim_{n\to\infty} |W_n| \, \mathbb{E}(\widehat{\lambda}_n - \lambda)^2, \tag{5.1}$$

where $\{W_n\}$ is a sequence of sets (observation windows) growing to $\mathbb{R}^k$ and satisfying some regularity conditions (to be defined later), and $\widehat{\lambda}_n$ are corresponding intensity estimators (3.8) computed for W_n.

The quantity σ^2 is used for constructing the confidence intervals for the intensity of the stationary point process and also as a normalizing term in goodness-of-fit tests for the K-function (or second moment measure) of stationary point processes (see e.g. [20]).

A class of kernel type estimators of σ^2 is defined e.g. in [21] and their asymptotic unbiasedness and weak consistency is shown. However there are different choices of possible kernels or bandwidths for the estimators and neither their asymptotic optimality nor their behaviour on middle size windows was investigated in detail. We will concentrate on these two questions in the present chapter. Moreover in Section 5.5 we will define and analyze a new estimator which can be used for the class of isotropic stationary point processes.

5.2 Preliminaries

At first let us remind the relation between the dth-order factorial moment measure $\alpha^{(d)}$ and the dth-order factorial cumulant measure $\gamma^{(d)}$ defined in Definition 3.7. Equation (3.4)

states the recursive definition for any d. We give here the explicit equations for the first three nontrivial factorial cumulant measures because we will need them in the sequel:

$$\begin{aligned}
\gamma^{(2)}(B_1 \times B_2) &= \alpha^{(2)}(B_1 \times B_2) - \alpha^{(1)}(B_1)\alpha^{(1)}(B_2), && (5.2)\\
\gamma^{(3)}(B_1 \times B_2 \times B_3) &= \alpha^{(3)}(B_1 \times B_2 \times B_3) - \alpha^{(1)}(B_1)\alpha^{(2)}(B_2 \times B_3) && (5.3)\\
&\quad -\alpha^{(1)}(B_2)\gamma^{(2)}(B_1 \times B_3) - \alpha^{(1)}(B_3)\gamma^{(2)}(B_1 \times B_2)\\
\gamma^{(4)}(B_1 \times B_2 \times B_3 \times B_4) &= \alpha^{(4)}(B_1 \times B_2 \times B_3 \times B_4) && (5.4)\\
&\quad -\alpha^{(1)}(B_1)\gamma^{(3)}(B_2 \times B_3 \times B_4) - \alpha^{(1)}(B_2)\gamma^{(3)}(B_1 \times B_3 \times B_4)\\
&\quad -\alpha^{(1)}(B_3)\gamma^{(3)}(B_1 \times B_2 \times B_4) - \alpha^{(1)}(B_4)\gamma^{(3)}(B_1 \times B_2 \times B_3)\\
&\quad -\gamma^{(2)}(B_1 \times B_2)\alpha^{(2)}(B_3 \times B_4) - \gamma^{(2)}(B_1 \times B_3)\alpha^{(2)}(B_2 \times B_4)\\
&\quad -\gamma^{(2)}(B_1 \times B_4)\alpha^{(2)}(B_2 \times B_3).
\end{aligned}$$

For the stationary point process Φ moreover the reduced factorial moment measures and reduced factorial cumulant measures (Definition 3.16) are defined. We will often use the expression for the asymptotic variance using the second order reduced factorial cumulant measure (for general proof of the equation see [19])

$$\sigma^2 = \lambda(1 + \gamma^{(2)}_{red}(\mathbb{R}^k)). \tag{5.5}$$

In the sequel we will also need the following lemma.

Lemma 5.1 *Let Φ be a stationary simple point process with intensity λ. Let f be a nonnegative symmetric (i.e. $f(x,y) = f(y,x)$ for any $x, y \in \mathbb{R}^k$) measurable function on $\mathbb{R}^k \times \mathbb{R}^k$. Then*

$$\mathbb{E} \sum_{x,y \in \Phi}^{\neq} f(x,y) = \lambda \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} f(x, x+y)\gamma^{(2)}_{red}(\mathrm{d}y)\,\mathrm{d}x + \lambda^2 \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} f(x,y)\,\mathrm{d}x\,\mathrm{d}y, \tag{5.6}$$

and

$$\begin{aligned}
\mathrm{Var} \sum_{x,y \in \Phi}^{\neq} f(x,y) &= 2\lambda \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} f(x, x+y)^2 \gamma^{(2)}_{red}(\mathrm{d}y)\mathrm{d}x && (5.7)\\
&\quad +2\lambda^2 \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} f(x,y)^2 \mathrm{d}y\mathrm{d}x\\
&\quad + 4\lambda \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} f(x, x+y) f(x+y, x+u)\gamma^{(3)}_{red}(\mathrm{d}y, \mathrm{d}u)\mathrm{d}x\\
&\quad + 4\lambda^2 \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} f(x,y)[2f(y, y+u) + f(y, x+u)]\gamma^{(2)}_{red}(\mathrm{d}u)\mathrm{d}y\mathrm{d}x\\
&\quad + 4\lambda^3 \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} f(x,y) f(y,u)\mathrm{d}u\mathrm{d}y\mathrm{d}x\\
&\quad + \lambda \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} f(x, x+y) f(x+u, x+v)\gamma^{(4)}_{red}(\mathrm{d}y, \mathrm{d}u, \mathrm{d}v)\mathrm{d}x\\
&\quad + 4\lambda^2 \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} f(x,y) f(y+u, y+v)\gamma^{(3)}_{red}(\mathrm{d}u, \mathrm{d}v)\mathrm{d}y\mathrm{d}x
\end{aligned}$$

(5.8)

$$+2\lambda^2 \int_{\mathbb{R}^k}\int_{\mathbb{R}^k}\int_{\mathbb{R}^k}\int_{\mathbb{R}^k} f(x,y)f(x+u,y+v)\gamma_{red}^{(2)}(\mathrm{d}u)\gamma_{red}^{(2)}(\mathrm{d}v)\mathrm{d}y\mathrm{d}x$$
$$+4\lambda^3 \int_{\mathbb{R}^k}\int_{\mathbb{R}^k}\int_{\mathbb{R}^k}\int_{\mathbb{R}^k} f(x,y)f(x+u,v)\gamma_{red}^{(2)}(\mathrm{d}u)\mathrm{d}v\mathrm{d}y\mathrm{d}x.$$

Proof. From Corollary 3.2 and equation (5.2) we have

$$\begin{aligned}\mathbb{E}\sum_{x,y\in\Phi}^{\neq} f(x,y) &= \int_{\mathbb{R}^k}\int_{\mathbb{R}^k} f(x,y)\alpha^{(2)}(\mathrm{d}x,\mathrm{d}y)\\ &= \int_{\mathbb{R}^k}\int_{\mathbb{R}^k} f(x,y)\gamma^{(2)}(\mathrm{d}x,\mathrm{d}y)+\int_{\mathbb{R}^k}\int_{\mathbb{R}^k} f(x,y)\Lambda(\mathrm{d}x)\Lambda(\mathrm{d}y),\end{aligned}$$

and using stationarity of Φ and the definition of the reduced factorial cumulant measure $\gamma_{red}^{(2)}$ we get equation (5.6).

From Corollary 3.2, equations (5.2) — (5.4) and the following observation

$$\left(\sum_{x,y\in\Phi}^{\neq} f(x,y)\right)^2 = 2\sum_{x,y\in\Phi}^{\neq} f(x,y)^2 + 4\sum_{x,y,u\in\Phi}^{\neq} f(x,y)f(y,u) + \sum_{x,y,u,v\in\Phi}^{\neq} f(x,y)f(u,v),$$

we get subsequently

$$\begin{aligned}\mathrm{Var}\sum_{x,y\in\Phi}^{\neq} f(x,y) &= \mathbb{E}\left(\sum_{x,y\in\Phi}^{\neq} f(x,y)\right)^2 - \left(\mathbb{E}\sum_{x,y\in\Phi}^{\neq} f(x,y)\right)^2\\ &= 2\int_{\mathbb{R}^k}\int f(x,y)\,\alpha^{(2)}(\mathrm{d}x,\mathrm{d}y)+4\int_{\mathbb{R}^k}\int_{\mathbb{R}^k}\int f(x,y)f(y,u)\,\alpha^{(3)}(\mathrm{d}x,\mathrm{d}y,\mathrm{d}u)\\ &\quad+\int_{\mathbb{R}^k}\int_{\mathbb{R}^k}\int_{\mathbb{R}^k}\int f(x,y)f(u,v)\,\left[\alpha^{(4)}(\mathrm{d}x,\mathrm{d}y,\mathrm{d}u,\mathrm{d}v)-\alpha^{(2)}(\mathrm{d}x,\mathrm{d}y)\alpha^{(2)}(\mathrm{d}u,\mathrm{d}v)\right]\\ &= 2\int_{\mathbb{R}^k}\int f(x,y)\,[\gamma^{(2)}(\mathrm{d}x,\mathrm{d}y)+\lambda^2\,\mathrm{d}x\,\mathrm{d}y]\\ &\quad+4\int_{\mathbb{R}^k}\int_{\mathbb{R}^k}\int f(x,y)f(y,u)\,\left[\gamma^{(3)}(\mathrm{d}x,\mathrm{d}y,\mathrm{d}u)\right.\\ &\quad\left.+\,2\lambda\,\gamma^{(2)}(\mathrm{d}y,\mathrm{d}u)\,dx+\lambda\,\gamma^{(2)}(\mathrm{d}x,\mathrm{d}u)\,\mathrm{d}y+\lambda^3\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}u\right]\\ &\quad+\int_{\mathbb{R}^k}\int_{\mathbb{R}^k}\int_{\mathbb{R}^k}\int f(x,y)f(u,v)\,\left[\gamma^{(4)}(\mathrm{d}x,\mathrm{d}y,\mathrm{d}u,\mathrm{d}v)+4\lambda\,\gamma^{(3)}(\mathrm{d}y,\mathrm{d}u,\mathrm{d}v)\,\mathrm{d}x\right.\\ &\quad\left.+\,2\gamma^{(2)}(\mathrm{d}x,\mathrm{d}u)\gamma^{(2)}(\mathrm{d}y,\mathrm{d}v)+4\lambda^2\,\gamma^{(2)}(\mathrm{d}x,\mathrm{d}u)\,\mathrm{d}y\,\mathrm{d}v\right].\end{aligned}$$

Equation (5.7) follows from the definition of the reduced factorial cumulant measures. □

In the rest of this chapter we will also use two symbols o and $\mathcal{O}$ for comparing the magnitudes of two (deterministic) sequences $\{a_n\}$, $\{c_n\}$ as $n\to\infty$.

$a_n = o(c_n)$ will denote $\lim_{n\to\infty}\frac{a_n}{c_n}=0$,

and $a_n=\mathcal{O}(c_n)$ means for us that for some $k, K>0$

$$k<\liminf_{n\to\infty}\left|\frac{a_n}{c_n}\right|\le\limsup_{n\to\infty}\left|\frac{a_n}{c_n}\right|<K.$$

5.3 The estimator

In order to be able to define a good estimator of σ^2 and to study its asymptotic behaviour we will need the following assumptions on the sequence of observation windows (or sampling regions), kernels, bandwidths and the process Φ:

(A0) The sampling regions are $W_1 = [-1,1]^k$, $W_n = nW_1, n > 1$.

(A1) The kernel function $w : \mathbb{R}^k \to \mathbb{R}^1$ is bounded, symmetric, non-negative, with support in W_1 and $\lim_{|t|\to 0} w(t) = w(0) = 1$.

(A2) Put $w_n(t) = w(t/(b_n n))$ and

$$\mathrm{w}_n = \int_{\mathbb{R}^k} w_n(t)\mathrm{d}t = (b_n n)^k \int_{\mathbb{R}^k} w(t)\mathrm{d}t = (b_n n)^k \mathrm{w},$$

where $\{b_n\}$ is a sequence of positive numbers such that $b_n \leq b_1 = 1$ and $b_n \downarrow 0$.

(A3) For the bandwidth $b_n n$ it holds $b_n n \to \infty$.

(A4) The second order reduced factorial cumulant measure $\gamma_{red}^{(2)}$ is of bounded variation, i.e.

$$C_2 = \int_{\mathbb{R}^k} |\gamma_{red}^{(2)}(\mathrm{d}x)| < \infty.$$

(A5) The reduced factorial cumulant measures $\gamma_{red}^{(i)}$, $i = 3, 4$ are of bounded variation, i.e., $C_3 = \int_{(\mathbb{R}^k)^2} |\gamma_{red}^{(3)}(\mathrm{d}(x,y))| < \infty$ and $C_4 = \int_{(\mathbb{R}^k)^3} |\gamma_{red}^{(4)}(\mathrm{d}(x,y,z))| < \infty$, and the sequence $\{b_n\}$ satisfies $b_n^2 n \to 0$.

Note that the assumptions (A4) and (A5) imply that $\alpha^{(3)}$ and $\alpha^{(4)}$ are bounded on $\mathcal{B}_0$.

The estimators $\widehat{\sigma}_n^2$ are defined as follows

$$\widehat{\sigma}_n^2 = \sum_{x,y\in\Phi} \frac{w_n(y-x)\mathbf{1}_{W_n}(x)\mathbf{1}_{W_n}(y)}{|(W_n - x)\cap(W_n - y)|} - \mathrm{w}_n \left(\frac{\Phi(W_n)}{|W_n|}\right)^2. \tag{5.9}$$

For sake of completeness let us show here the proof of relation (5.5) under our assumptions.

Lemma 5.2 *Under assumptions (A0) and (A4) it holds* $\sigma^2 = \lambda(1 + \gamma_{red}^{(2)}(\mathbb{R}^k))$.

Proof. Using (3.5) and the definition of $\gamma_{red}^{(2)}$ we get subsequently

$$\begin{aligned}
|W_n|\mathbb{E}(\widehat{\lambda}_n - \lambda)^2 &= \frac{\mathbb{E}(\Phi(W_n) - \lambda|W_n|)^2}{|W_n|} = \frac{\mathrm{Var}(\Phi(W_n))}{|W_n|} = \frac{\gamma^{(2)}(W_n \times W_n) + \Lambda(W_n)}{|W_n|} \\
&= \frac{1}{|W_n|}\,\lambda \int_{W_n}\int_{\mathbb{R}^k} \mathbf{1}_{(W_n - x)}\,\gamma_{red}^{(2)}(\mathrm{d}z)\,\mathrm{d}x \;+\; \lambda \\
&= \lambda\left(\int_{\mathbb{R}^k} \frac{|W_n \cap (W_n - z)|}{|W_n|}\,\gamma_{red}^{(2)}(\mathrm{d}z) \;+\; 1\right).
\end{aligned}$$

Finally because $\frac{|W_n \cap (W_n - z)|}{|W_n|}$ converges monotonically to 1 for each z and $\gamma_{red}^{(2)}$ has finite total variation we get (5.5) from monotone convergence theorem. □

5.4 Asymptotic results and optimal estimators of the asymptotic variance

The following two propositions are proved in [21].

Theorem 5.3 *Under the assumptions (A0) – (A4), the sequence of estimators* $\{\widehat{\sigma}_n^2\}$ *is asymptotically unbiased for* σ^2.

Theorem 5.4 *Under assumptions (A0) – (A3) and (A5) it holds*

$$\mathbb{E}|\widehat{\sigma}_n^2 - \sigma^2|^2 \to 0 \qquad as \quad n \to \infty. \tag{5.10}$$

Thus the estimators $\{\widehat{\sigma}_n^2\}$ are asymptotically unbiased and weakly consistent. However the speed of the convergence $\widehat{\sigma}_n^2 \to \sigma^2$ depends on the shape of the chosen kernel w and on $\{b_n\}$. We want to find an asymptotically optimal form of $\{b_n\}$ with respect to the mean squared error

$$MSE(\widehat{\sigma}_n^2) = \mathbb{E}(\sigma^2 - \widehat{\sigma}_n^2)^2.$$

Using the decomposition of MSE into the variance and the squared bias

$$MSE(\widehat{\sigma}_n^2) = \mathrm{Var}(\widehat{\sigma}_n^2) + (\mathbb{E}\widehat{\sigma}_n^2 - \sigma^2)^2,$$

we will show in the following that

$$MSE(\widehat{\sigma}_n^2) = \mathcal{O}\left(\int_{\mathbb{R}^k} (w_n(x) - 1)\gamma_{red}^{(2)}(\mathrm{d}x)\right)^2 + \mathcal{O}(b_n^k) + \mathcal{O}((b_n^2 n)^k) + \mathcal{O}(n^{-k}).$$

Thus the optimal bandwidth depends on the behaviour of the integral in the first term which depends on the particular shape of w and $\gamma_{red}^{(2)}$.

To be able to do a more detailed analysis of the MSE asymptotics we will assume a special form of w namely

$$w(x) = \mathbf{1}(\|x\| \leq 1), \qquad x \in \mathbb{R}^k.$$

w is then equal to ω_k the volume of the k-dimensional unit ball. (We will see from the simulation study, that besides being suitable for a theoretical analysis this kernel also gives the best results on the middle-size observation windows.) Further we distinguish three different types of the tail behaviour of $\gamma_{red}^{(2)}$ – decaying exponentially, decaying more quickly than exponentially and $\gamma_{red}^{(2)}$ with bounded support.

Let us start with the last case of supp $(\gamma_{red}^{(2)}) \in \mathcal{B}_0$ because this enables us to relax the conditions for the consistency a little bit — we do not need the condition $b_n\, n \to \infty$.

Theorem 5.5 *Suppose that for the stationary point process Φ hold (A0)–(A2), (A4) and (A5) and the measure $\gamma^{(2)}_{red}$ has bounded support in $\mathbb{R}^k$. Suppose moreover that*

$$\text{(A6)} \quad \text{supp}\,(\gamma^{(2)}_{red}) \subseteq B(\mathrm{o}, L), \qquad L = \liminf_{n\to\infty} b_n\, n, \qquad L \in \mathbb{R}^+,$$

is satisfied. Then for $w(x) = \mathbf{1}(\|x\| \le 1)$ the sequence of estimators $\{\widehat{\sigma}^2_n\}$ is asymptotically unbiased for σ^2 and

$$\mathbb{E}|\widehat{\sigma}^2_n - \sigma^2|^2 \to 0 \qquad as \quad n \to \infty.$$

The MSE-asymptotically optimal bandwidth is

$$b^*_n = \frac{c^*}{n}, \qquad for \qquad c^* = \inf\{c : \text{supp}\,(\gamma^{(2)}_{red}) \subseteq B(\mathrm{o}, c)\}.$$

Proof. To show the asymptotic unbiasedness let us first decompose $\widehat{\sigma}^2_n$ as

$$\begin{aligned}\widehat{\sigma}^2_n &= \frac{\Phi(W_n)}{|W_n|} + \left(\sum_{x,\,y\in\Phi}^{\neq} \frac{w_n(y-x)\mathbf{1}_{W_n}(x)\mathbf{1}_{W_n}(y)}{|(W_n - x)\cap(W_n - y)|} - \mathrm{w}_n\lambda^2\right) \\ &\quad + \mathrm{w}_n\left(\lambda^2 - \left(\frac{\Phi(W_n)}{|W_n|}\right)^2\right).\end{aligned} \tag{5.11}$$

Corollary 3.2, Lemma 5.1 and (3.2) imply

$$\begin{aligned}\mathbb{E}\widehat{\sigma}^2_n &= \lambda + \lambda\int_{\mathbb{R}^k}\int_{\mathbb{R}^k} \frac{w_n(y)\mathbf{1}_{W_n}(x)\mathbf{1}_{W_n}(x+y)}{|(W_n - x)\cap(W_n - x - y)|}\gamma^{(2)}_{red}(\mathrm{d}y)\mathrm{d}x \\ &\quad +\lambda^2\int_{\mathbb{R}^k}\int_{\mathbb{R}^k} \frac{w_n(y-x)\mathbf{1}_{W_n}(x)\mathbf{1}_{W_n}(y)}{|(W_n - x)\cap(W_n - y)|}\,\mathrm{d}x\,\mathrm{d}y - \mathrm{w}_n\lambda^2 \\ &\quad +\mathrm{w}_n\lambda^2 - \frac{\mathrm{w}_n}{|W_n|^2}(\Lambda(W_n)^2 + \gamma^{(2)}(W_n\times W_n) + \Lambda(W_n)) \\ &= \lambda + \lambda\int_{\mathbb{R}^k} w_n(y)\left(\int_{\mathbb{R}^k}\frac{\mathbf{1}_{W_n}(x)\mathbf{1}_{W_n}(x+y)}{|(W_n)\cap(W_n - y)|}\mathrm{d}x\right)\gamma^{(2)}_{red}(\mathrm{d}y) \\ &\quad +\lambda^2\int_{\mathbb{R}^k} w_n(z)\left(\int_{\mathbb{R}^k}\frac{\mathbf{1}_{W_n}(x)\mathbf{1}_{W_n}(x+z)}{|(W_n)\cap(W_n - z)|}\,\mathrm{d}x\right)\mathrm{d}z - \mathrm{w}_n\lambda^2 \\ &\quad -\frac{\mathrm{w}_n}{|W_n|}\left(\lambda\gamma^{(2)}_{red}(W_n) + \lambda\right) \\ &= \lambda + \lambda\int_{\mathbb{R}^k} w_n(y)\gamma^{(2)}_{red}(\mathrm{d}y) - \frac{(b_n\,n)^k\mathrm{w}}{(2n)^k}\left(\lambda + \lambda\gamma^{(2)}_{red}(W_n)\right).\end{aligned}$$

Thus together with equation (5.5) we have

$$|\mathbb{E}\widehat{\sigma}^2_n - \sigma^2| = \left|\lambda\int_{\mathbb{R}^k}(w_n(y)-1)\gamma^{(2)}_{red}(\mathrm{d}y) - \left(\frac{b_n}{2}\right)^k \mathrm{w}(\lambda + \lambda\gamma^{(2)}_{red}(W_n))\right| \tag{5.12}$$

$$\le \lambda\left|\int_{\mathbb{R}^k}(w_n(y)-1)\gamma^{(2)}_{red}(\mathrm{d}y)\right| + \left(\frac{b_n}{2}\right)^k \mathrm{w}(\lambda + \lambda C_2). \tag{5.13}$$

When $w(x) = \mathbf{1}(\|x\| \le 1)$ and $n\,b_n \ge L$ we get from assumption (A6)

$$\mathbb{E}\widehat{\sigma}_n^2 - \sigma^2 = \mathcal{O}(b_n^k), \tag{5.14}$$

which converges to 0 as $n \to \infty$.

To show that $\mathrm{Var}(\widehat{\sigma}_n^2) \to 0$ let us introduce a new decomposition of $\widehat{\sigma}_n^2$

$$\widehat{\sigma}_n^2 = \sum_{x,\,y\in\Phi}^{\neq} f_1(x,y) - \sum_{x,\,y\in\Phi}^{\neq} \mathrm{w}(n\,b_n)^k f_2(x,y) + \frac{\Phi(W_n)}{|W_n|}\left(1 - \frac{\mathrm{w}(n\,b_n)^k}{|W_n|}\right) = T_1 + T_2 + T_3, \tag{5.15}$$

where

$$f_1(x,y) = \frac{\mathbf{1}_{B(\mathrm{o},n\,b_n)}(x-y)\mathbf{1}_{W_n}(x)\mathbf{1}_{W_n}(y)}{|(W_n - x)\cap(W_n - y)|} \qquad \text{and} \qquad f_2(x,y) = \frac{\mathbf{1}_{W_n}(x)\mathbf{1}_{W_n}(y)}{|W_n|^2}.$$

Note that we have already used our form of $w(x) = \mathbf{1}(\|x\| \le 1)$.

Using the formula $\mathrm{Var}(\Phi(W_n)) = \Lambda(W_n) + \gamma^{(2)}(W_n \times W_n) \le \lambda|W_n|(1 + C_2)$, we have

$$\mathrm{Var}(T_3) = \mathcal{O}\left(\frac{1}{|W_n|}\right) = \mathcal{O}(n^{-k}). \tag{5.16}$$

From Lemma 5.1 we get

$$\begin{aligned}
\mathrm{Var}(T_2) \;\le\; & \frac{(\mathrm{w}(n\,b_n)^k)^2}{|W_n|^4}\left(2\lambda C_2|W_n| + 2\lambda^2|W_n|^2 + 4\lambda|W_n|C_3 + 4\lambda^2|W_n|^2 3C_2\right. \\
& \left. +\; 4\lambda^3|W_n|^3 + \lambda|W_n|C_4 + 4\lambda^2|W_n|C_3 + 2\lambda^2|W_n|^2C_2^2 + 4\lambda^3|W_n|^3C_2\right) \\
=\; & \mathcal{O}\left(\frac{(n\,b_n)^{2k}}{|W_n|}\right) = \mathcal{O}(b_n^{2k}n^k).
\end{aligned} \tag{5.17}$$

Since $(1 - b_n)W_n \subseteq W_n \cap (W_n - z)$ for $z \in b_n\,W_n \supseteq B(\mathrm{o}, n\,b_n)$

$$f_1(x,y) \le \frac{1}{(1-b_n)|W_n|}\mathbf{1}_{W_n}(x)\mathbf{1}_{W_n}(y)\mathbf{1}_{B(\mathrm{o},n\,b_n)}(x-y) \qquad \text{for} \quad b_n < 1. \tag{5.18}$$

Using this inequality it follows from Lemma 5.1

$$\begin{aligned}
\mathrm{Var}(T_1) \;\le\; & \frac{1}{(1-b_n)^2|W_n|^2}\left(2\lambda C_2|W_n| + 2\lambda^2|W_n|(n\,b_n)^k\mathrm{w} + 4\lambda|W_n|C_3\right. \\
& +\; 4\lambda^2 3|W_n|(n\,b_n)^k\mathrm{w}C_2 + 4\lambda^3|W_n|(n\,b_n)^{2k}\mathrm{w}^2 + \lambda|W_n|C_4 \\
& +\; 4\lambda^2|W_n|(n\,b_n)^k\mathrm{w}C_3 + 2\lambda^2|W_n|(n\,b_n)^k\mathrm{w}C_2^2 \\
& \left. +\; 4\lambda^3|W_n|(n\,b_n)^{2k}\mathrm{w}^2C_2\right) \\
=\; & \mathcal{O}(n^{-k}) + \mathcal{O}(b_n^k) + \mathcal{O}(n^k b_n^{2k}).
\end{aligned} \tag{5.19}$$

Under assumptions (A2) and (A5) all $\mathrm{Var}(T_i), i \in \{1,2,3\}$ converge to 0 as $n \to \infty$. Thus $\mathbb{E}|\widehat{\sigma}_n^2 - \sigma^2|^2 \to 0$.

Combining estimates (5.14), (5.16), (5.17) and (5.19) with assumption (A6) we get

$$MSE(\widehat{\sigma}_n^2) = \mathcal{O}(b_n^k) + \mathcal{O}(n^{-k}) + \mathcal{O}(b_n^{2k} n^k) \quad \text{for} \quad b_n n \geq c^*,$$

which is minimal for $(b_n n)^k = (c^*)^k$. □

As the corollary of the proof of Theorem 5.5 we get the upper bound for the $MSE(\widehat{\sigma}_n^2)$ for the processes satisfying the assumptions (A0) – (A5) as well.

Corollary 5.6 *Let Φ be a stationary point process satisfying assumptions (A4) and (A5) and let also the assumptions (A0) – (A2) be satisfied. Then*

$$MSE(\widehat{\sigma}_n^2) = \mathcal{O}\left(\int_{\mathbb{R}^k} (w_n(x) - 1)\gamma_{red}^{(2)}(\mathrm{d}x)\right)^2 + \mathcal{O}(b_n^k) + \mathcal{O}((b_n^2 n)^k) + \mathcal{O}(n^{-k}). \qquad (5.20)$$

Now it is easy to find the asymptotically MSE-optimal bandwidth for the kernel $w(x) = \mathbf{1}(\|x\| \leq 1)$ for processes with exponentially and more quickly than exponentially decaying $\gamma_{red}^{(2)}$.

Lemma 5.7 *Suppose that the assumptions (A0) – (A3) and (A5) are fulfilled and let $w(x) = \mathbf{1}(\|x\| \leq 1)$. If*

$$|\gamma_{red}^{(2)}(\mathbb{R}^k \setminus B(\mathrm{o}, R))| = \mathcal{O}(R^{-\epsilon d}) \quad \text{as } R \to \infty \quad \text{for some } \epsilon > 0 \qquad (5.21)$$

then

$$MSE(\widehat{\sigma}_n^2) = \mathcal{O}((b_n n)^{-2\epsilon k}) + \mathcal{O}((b_n^2 n)^k), \quad n \to \infty,$$

and the asymptotically MSE-optimal bandwidth is

$$b_n^* = c\, n^{-\frac{1+2\epsilon}{2+2\epsilon}} \quad \text{for some constant } c > 0.$$

If

$$|\gamma_{red}^{(2)}(\mathbb{R}^k \setminus B(\mathrm{o}, R))| = o(R^{-\epsilon k}) \quad \text{as } R \to \infty \quad \text{for all } \epsilon > 0, \qquad (5.22)$$

the asymptotically MSE-optimal bandwidth has the form

$$b_n^* = \frac{h(n)}{n}$$

for some function h satisfying $h(n) \to \infty$ and $h(n) = o(n^\epsilon)$ for all $\epsilon > 0$ as $n \to \infty$.

Proof. The first assertion follows immediately from substitution into (5.20) for $R = n\, b_n$. The second assertion follows from the first assertion and the condition (A3). □

In the special case of an isotropic point process Φ conditions (5.21) and (5.22) can be checked using the pair correlation function g.

Lemma 5.8 *Let Φ be a stationary isotropic point process with the pair correlation function $g(r), r \in \mathbb{R}^+$.*
(i) If $g(r) = 1 + \mathcal{O}(r^{-k-\epsilon k})$ as $r \to \infty$ for some $\epsilon > 0$, then the condition (5.21) is fulfilled.
(ii) If $g(r) = 1 + \mathcal{O}(r^{\alpha} \exp(-p\, r^{\beta}))$ as $r \to \infty$ for some $\alpha, \beta, p > 0$, then the condition (5.22) is fulfilled.

Proof. Using definitions of the K-function and the pair correlation function g for an isotropic process we can write

$$\begin{aligned}\gamma^{(2)}_{red}(B(\mathrm{o}, R)) &= \alpha^{(2)}_{red}(B(\mathrm{o}, R)) - \lambda|B(\mathrm{o}, R)| = \lambda K(R) - \lambda|B(\mathrm{o}, R)| \\ &= \lambda \int_0^R g(r) k\omega_k r^{k-1} \mathrm{d}r - \lambda \int_0^R 1\, k\omega_k r^{k-1} \mathrm{d}r,\end{aligned}$$

where ω_k is the volume of the k-dimensional unit ball.

Thus

$$\gamma^{(2)}_{red}(\mathbb{R}^k \backslash B(\mathrm{o}, R)) = \lambda \int_R^{\infty} (g(r) - 1) k\omega_k r^{k-1} \mathrm{d}r,$$

and in case (i)

$$|\gamma^{(2)}_{red}(\mathbb{R}^k \backslash B(\mathrm{o}, R))| = \lambda \int_R^{\infty} \mathcal{O}(r^{-k-\epsilon k}) k\omega_k r^{k-1} \mathrm{d}r = \int_R^{\infty} \mathcal{O}(r^{-1-\epsilon k})\, \mathrm{d}r = \mathcal{O}(R^{-\epsilon k})$$

satisfies condition (5.21).

In case (ii) using integration by parts in the third equality we get

$$\begin{aligned}|\gamma^{(2)}_{red}(\mathbb{R}^k \backslash B(\mathrm{o}, R))| &= \lambda \int_R^{\infty} \mathcal{O}(r^{\alpha} \exp(-p r^{\beta})) k\omega_k r^{k-1} \mathrm{d}r = \int_R^{\infty} \mathcal{O}(r^{k-1+\alpha} \exp(-p\, r^{\beta})) \mathrm{d}r \\ &= \mathcal{O}(R^{k+\alpha-\beta} \exp(-pR^{\beta})) + o\left(\int_R^{\infty} \mathcal{O}(r^{k-1+\alpha} \exp(-p\, r^{\beta})) \mathrm{d}r \right) \\ &= \mathcal{O}(R^{k+\alpha-\beta} \exp(-pR^{\beta})) = o\,(R^{-\epsilon k}) \qquad \text{for all} \quad \epsilon > 0,\end{aligned}$$

and the condition (5.22) is fulfilled. □

5.5 New estimator $\tilde{\sigma}^2_n$ for stationary isotropic point processes

If we rewrite the estimator $\widehat{\sigma}^2_n$ with the kernel $w(x) = \mathbf{1}(\|x\| \le 1)$ in the following way

$$\widehat{\sigma}^2_n = \frac{\Phi(W_n)}{|W_n|} + \sum_{x,\, y \in \Phi}^{\neq} \frac{w(\frac{y-x}{n\, b_n}) \mathbf{1}_{W_n}(x) \mathbf{1}_{W_n}(y)}{|(W_n - x) \cap (W_n - y)|} - \omega_n \left(\frac{\Phi(W_n)}{|W_n|} \right)^2, \tag{5.23}$$

we can see that the middle term

$$\sum_{x,\, y \in \Phi}^{\neq} \frac{\mathbf{1}(\|x - y\| \le n\, b_n) \mathbf{1}_{W_n}(x) \mathbf{1}_{W_n}(y)}{|(W_n - x) \cap (W_n - y)|}, \tag{5.24}$$

is equal to the standard rigid motion-invariant estimate (3.9) of $\lambda^2 K(B(\mathrm{o}, n\,b_n))$.

This suggests that we may improve the estimator $\widehat{\sigma}_n^2$ using the facts that are known about different ratio-unbiased estimators of the K-function. For stationary isotropic processes usually the estimator (3.10) is used. Because in our case

$$n\,b_n \le n \le \mathrm{radius}(W_n),$$

and W_n is convex we have $W_n^{(r)} = W_n$, and

$$\widehat{\lambda^2 K(r)} = \sum_{x,y\in\Phi}^{\neq} \frac{\mathbf{1}(\|x-y\| \le r)\mathbf{1}_{W_n}(x)\mathbf{1}_{W_n}(y)}{|W_n|}\, k(x,y), \qquad \text{for} \quad 0 \le r \le n, \tag{5.25}$$

where

$$k(x,y) = \frac{|\partial B(x, \|x-y\|)|}{|\partial B(x, \|x-y\|) \cap W_n|} \qquad \text{for} \quad x \neq y \in W_n,\ \|x-y\| \le n. \tag{5.26}$$

The isotropic estimator (5.25) is believed to have similar properties like the rigid motion-invariant estimator (5.24) for $n\,b_n$ small in comparison with the size of the window W_n. For $n\,b_n$ large it should have smaller variance then the rigid motion invariant estimator (see [46]).

Inspired by (5.25) for isotropic processes Φ we can define a new estimator of σ^2

$$\tilde{\sigma}_n^2 = \sum_{x,y\in\,\Phi} \frac{\mathbf{1}(\|x-y\| \le n\,b_n)\mathbf{1}_{W_n}(x)\mathbf{1}_{W_n}(y)}{|W_n|}\, k(x,y) \;-\; |B(\mathrm{o}, n\,b_n)| \left(\frac{\Phi(W_n)}{|W_n|}\right)^2, \tag{5.27}$$

where $k(x,y)$ is defined by (5.26) for $x \neq y$ and $k(x,x) = 1$.

Note that the isotropic estimator $\tilde{\sigma}_n^2$ cannot be written in the form (5.9) because the function $k(x,y)$ depends also on the window W_n.

Since all $W_n = n[-1,1]^k$ it holds

$$k(x,y) \le 2^k \qquad \text{for all} \quad x, y, \in W_n,\ \|x-y\| \le n. \tag{5.28}$$

In the special case of $k = 2$ we can compute

$$\begin{aligned} k(x,y) \;=\; & 2\pi - \left(\arccos\left(\frac{n-|x_1|}{\|x-y\|}\right) + \arccos\left(\frac{n-|x_2|}{\|x-y\|}\right)\right. \\ & \left. + \min\left\{\frac{\pi}{2}, \left(\arccos\left(\frac{n-|x_1|}{\|x-y\|}\right) + \arccos\left(\frac{n-|x_2|}{\|x-y\|}\right)\right)\right\}\right), \end{aligned}$$

where $x = (x_1, x_2)$, $y = (y_1, y_2) \in W_n$ and $\|x-y\| \le n$.

Theorem 5.9 *Let Φ be a stationary isotropic point process on $\mathbb{R}^k$. Under the assumptions (A0) - (A4), the sequence of estimators $\{\tilde{\sigma}_n^2\}$ is asymptotically unbiased for σ^2 and if (A5) or (A6) is fulfilled it moreover holds*

$$\mathbb{E}|\tilde{\sigma}_n^2 - \sigma^2|^2 \to 0 \qquad as \quad n \to \infty. \tag{5.29}$$

Proof. Analogous to the proof of Theorem 5.5. For the proof of unbiasedness we use the fact

$$\mathbb{E} \sum_{x,y\in\Phi}^{\neq} \frac{\mathbf{1}(\|x-y\| \le n\, b_n)\mathbf{1}_{W_n}(x)\mathbf{1}_{W_n}(y)}{|W_n|}\, k(x,y) = \lambda^2 K(B(\mathrm{o}, n\, b_n))$$

$$= \lambda \gamma^{(2)}_{red}(B(\mathrm{o}, n\, b_n)) + \lambda^2 |B(\mathrm{o}, n\, b_n)|,$$

thus equation (5.12) holds and from (A4) we have the asymptotic unbiasedness of $\tilde{\sigma}^2_n$.

Also the proof of L_2 convergence (5.29) is the same, only instead of the bound (5.18) we use the bound

$$f_1(x,y) \le \frac{2^k}{|W_n|}\, \mathbf{1}_{W_n}(x)\mathbf{1}_{W_n}(y)\mathbf{1}_{B(\mathrm{o},nb_n))}(x-y),$$

that we get from (5.28). □

It is easy to see that the MSE-asymptotically optimal bandwidths for the isotropic estimator $\tilde{\sigma}^2_n$ are the same as for the estimator $\widehat{\sigma}^2_n$ with cylinder weight function $w(x) = \mathbf{1}(\|x\| \le 1)$.

Corollary 5.10 *Let Φ be a stationary isotropic point process on $\mathbb{R}^k$ and suppose that the assumptions (A0) – (A3) are fulfilled.*
If (A6) holds then the MSE-asymptotically optimal bandwidth for $\tilde{\sigma}^2_n$ is

$$b^*_n = \frac{c^*}{n}, \qquad \text{for} \qquad c^* = \inf\{c : \mathrm{supp}\ (\gamma^{(2)}_{red}) \subseteq B(\mathrm{o}, c)\}.$$

If (A5) holds and

$$|\gamma^{(2)}_{red}(\mathbb{R}^k \setminus B(\mathrm{o}, R))| = \mathcal{O}(R^{-\epsilon k}) \quad as\ R \to \infty \quad for\ some\ \epsilon > 0 \tag{5.30}$$

then the MSE-asymptotically optimal bandwidth for $\tilde{\sigma}^2_n$ is

$$b^*_n = c\, n^{-\frac{1+2\epsilon}{2+2\epsilon}} \quad for\ some\ constant\ c > 0.$$

If (A5) holds and

$$|\gamma^{(2)}_{red}(\mathbb{R}^k \setminus B(\mathrm{o}, R))| = o(R^{-\epsilon k}) \quad as\ R \to \infty \quad for\ all\ \epsilon > 0, \tag{5.31}$$

then the MSE-asymptotically optimal bandwidth for $\tilde{\sigma}^2_n$ has the form

$$b^*_n = \frac{h(n)}{n}$$

for some function h satisfying $h(n) \to \infty$ and $h(n) = o(n^\epsilon)$ for all $\epsilon > 0$ as $n \to \infty$.

The advantage of $\tilde{\sigma}^2_n$ is that the correction weights $k(x,y)$ are larger than 1 not for all x, y far from each other, but only when they are near the boundary of the observation window. On the other hand $\tilde{\sigma}^2_n$ should have approximately the same bias as $\widehat{\sigma}^2_n$ with the cylinder kernel.

5.6 Behaviour of the estimators on middle-size windows in $\mathbb{R}^2$ – simulation study

Returning once again to the decomposition (5.23) we can see that the middle term is responsible for the main part of the variability of the estimator $\widehat{\sigma}_n^2$, especially when the value of b_n is near to 1 because then the correction weights

$$1 \leq \frac{|W_n|}{|(W_n - x) \cap (W_n - y)|} \leq \frac{1}{(1 - b_n)^2} \qquad \text{for} \quad (x - y) \in [-n\, b_n, n\, b_n]^k, \tag{5.32}$$

can acquire much larger values than 1. For small windows the estimator can even have negative values.

In the following we want to study the behaviour of the estimator on middle-size windows (in $\mathbb{R}^2$) by which we mean at least 60 observed points in the window and b_n at most $1/\sqrt{2}$.

The variability of the middle term depends also on the form of the chosen kernel function w. For functions like

$$w(x) = \max(1 - \|x\|^2, 0), \tag{5.33}$$

(let us call it the half-ball kernel) or

$$w(x) = \max(1 - \|x\|, 0), \tag{5.34}$$

(the cone kernel), the larger correction weights (5.32) are balanced by smaller values of the kernel function for larger values of $x - y$. On the other hand for these kernel functions the estimate $\widehat{\sigma}_n^2$ can be biased more easily than for

$$w(x) = \mathbf{1}(\|x\| \leq 1),$$

(the cylinder kernel), because the bias is determined by

$$\int_{\mathbb{R}^k} (w_n(x) - 1)\gamma_{red}^{(2)}(\mathrm{d}x).$$

To see how the estimators of σ^2 behave on middle-size windows for different specific stationary point processes we made a simulation study for processes in $\mathbb{R}^2$. We simulated the processes on windows of 3 different sizes, namely on

$$W_5 = [-5, 5]^2, \quad W_{10} = [-10, 10]^2, \quad W_{20} = [-20, 20]^2$$

and from 100 realizations of the same point process we computed the squared bias, the variance and the MSE for the four estimators:

- $\widehat{\sigma}_n^2$ with the cone kernel,
- $\widehat{\sigma}_n^2$ with the half-ball kernel,
- $\widehat{\sigma}_n^2$ with the cylinder kernel
- the isotropic estimator $\tilde{\sigma}_n^2$.

For each of the estimators we used a scale of bandwidth values

$$n\,b_n \in [1,3], \qquad n\,b_n \in [1,5], \qquad n\,b_n \in [1,6],$$

on windows W_5, W_{10} and W_{20} respectively.

The procedure was done for 6 different point processes fulfilling condition (A5), all of them were isotropic and five of them had intensity $\lambda = 1$ (i.e. the mean number of observed points was 100, 400 and 1600 respectively) the last process had a slightly smaller intensity. In Section5.7 there are examples of realizations of the processes on each of the three windows. The following processes were used :

1. Poisson point process with intensity $\lambda = 1$.

   Here $\gamma_{red}^{(2)} \equiv 0$ and $\sigma^2 = \lambda$.

2. Matérn cluster point process with intensity $\lambda = 1$, mean number of points in the cluster $\mu = 5$ and radius $r = 1/2$.

   $\gamma_{red}^{(2)}$ has bounded support and it holds

   $$\gamma_{red}^{(2)}(\mathbb{R}^2 \backslash B(\mathrm{o}, 2r)) = \gamma_{red}^{(2)}(\mathbb{R}^2 \backslash B(\mathrm{o}, 1)) = 0,$$

   and from the formula (3.12) for the K-function we get

   $$K(R) = |B(\mathrm{o}, R)| + \mu, \qquad \text{for } R > 2r.$$

   Thus together we can derive

   $$\begin{aligned} \sigma^2 &= \lambda(1 + \gamma_{red}^{(2)}(\mathbb{R}^2)) = \lambda(1 + (\alpha_{red}^{(2)} - \lambda|\cdot|)(\mathbb{R}^2)) \\ &= \lambda(1 + \lambda(\lim_{R\to\infty}(K(B(\mathrm{o}, R)) - |B(\mathrm{o}, R)|))) = \lambda(1+\mu) = 6. \end{aligned}$$

3. Matérn cluster point process with intensity $\lambda = 1$, mean number of points in the cluster $\mu = 5$ and radius $r = 1$.

   Thus in comparison with the preceding point process we have less compact clusters of points.

   $$\gamma_{red}^{(2)}(\mathbb{R}^2 \backslash B(\mathrm{o}, 2)) = 0, \qquad \sigma^2 = 6.$$

4. Thomas cluster point process with intensity $\lambda = 1$, mean number of points in the cluster $\mu = 5$ and variance $v = 1/2$.

   Here $\gamma_{red}^{(2)}$ does not have bounded support and from (3.12) we have

   $$\lim_{R\to\infty}(K(R) - |B(\mathrm{o}, R)|) = \mu$$

   That means we get the same asymptotic variance as for the Matérn cluster process

   $$\sigma^2 = \lambda(1+\mu) = 6.$$

5. Thomas cluster point process with intensity $\lambda = 1$, mean number of points in the cluster $\mu = 5$ and variance $v = 1$.

   Again this process has less compact clusters than the preceding point process, $\gamma_{red}^{(2)}$ does not have bounded support and the asymptotic variance $\sigma^2 = 6$.

6. Matérn type II hard-core point process with $r = 1/2$ and $\lambda_p = 1$.

   The intensity of the process is $\lambda = (1 - \exp(-\lambda_p \pi r^2))/(\pi r^2)$ and we see from (3.13) that the pair correlation function $g = 1$ for $R \geq 2r$. Thus $\gamma_{red}^{(2)}$ has bounded support and using the formula $K(R) = \int_0^R g(x) k \omega_k x^{k-1} \mathrm{d}x$ we have

$$\begin{aligned} \lambda &\doteq 0.693, \\ \gamma_{red}^{(2)}(B(\mathrm{o}, 2r)) &\doteq -0.494, \\ \gamma_{red}^{(2)}(\mathbb{R}^2 \setminus B(\mathrm{o}, 2r)) &= 0, \\ \sigma^2 &\doteq 0.350. \end{aligned}$$

The resulting graphs can be found in Section 5.7. There are shown the empirical MSE, variance and squared bias of the estimators as functions of the bandwidth $n\, b_n$.

To be able to compare the behaviour of the estimators for the different processes we used in the graphs the relative MSE

$$\text{rel MSE} = \text{MSE}(\widehat{\sigma}_n^2)/(\sigma^2)^2,$$

relative squared bias and relative variance.

From the figures we can see good performance of the estimators for Poisson point process since there is no problem with the bias in this case. For Matern II point process we get similar results because supp $(\gamma_{red}^{(2)}) \subseteq B(\mathrm{o}, 1)$ thus the estimates $\tilde{\sigma}_n^2$ and $\widehat{\sigma}_n^2$ with the cylinder kernel are unbiased even for $n\, b_n = 1$ and the variability of the estimators is smaller because of the regularity of the point pattern.

We get a different picture with the cluster processes, since $diam(\text{supp } (\gamma_{red}^{(2)}))$ here is bigger or even infinite and bias is a problem. The smallest value of MSE we get in all cases for $\widehat{\sigma}_n^2$ with cylinder kernel and for the isotropic estimator $\tilde{\sigma}_n^2$ with a suitable bandwidth. These two estimators behave almost identically on W_{10} and W_{20} for smaller values of $n\, b_n$, for larger values (larger than the MSE-optimal bandwidth) of the bandwidth the isotropic estimator $\tilde{\sigma}_n^2$ has smaller variance. The bias–variance trade-off in dependence on the bandwidth can be seen well for the cluster processes and it is also evident that the two other estimators $\widehat{\sigma}_n^2$ with the cylinder and cone kernel are not good.

Looking at the optimal obtained values of rel MSE for cluster processes on W_5 and W_{10} which are between 0.05 and 0.15 we can see nevertheless that it is impossible to estimate σ^2 well by such sizes of the observation window. The estimators behave reasonably well only for big windows like W_{20} (which means approximately 1600 observed points in this case).

5.7 Figures

In the following figures are shown the results of the simulation study on the asymptotic variance of stationary point processes.

However to get a better idea about the different point processes used in the simulation study, examples of realizations of the five nontrivial processes (i.e. excluding the Poisson point process) are shown on the first page. The processes are shown on windows of all three sizes to be able to observe the behaviour of the point processes on different scales (note that in each case it is actually only one realization simulated on W_{20} and then the central area is magnified).

Each row of the figures corresponds to one type of point process which is indicated (included the value of the distinguishing parameter, if needed) above the three figures. The figures in each row are from left to right realizations of the corresponding point process on W_5, W_{10} and W_{20}. The number of observed points is given in the parentheses after the indication of the point process.

On the subsequent four pages are shown the graphs of the relative MSE (y-axis)

$$\text{rel MSE} = \text{MSE}(\widehat{\sigma}_n^2)/(\sigma^2)^2,$$

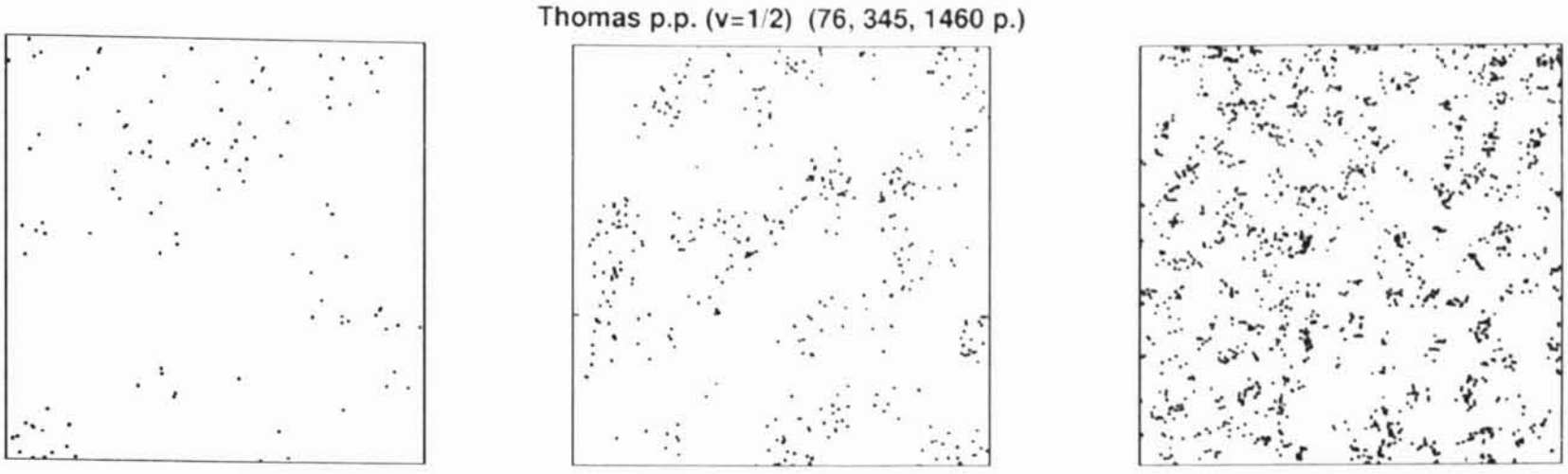
of the different estimators of the asymptotic variance σ^2 as functions of the value of the bandwidth $n\,b_n$ (x-axis). The value of the empirical relative MSE (full-drawn line) is moreover decomposed into the relative squared bias $(\mathbb{E}(\widehat{\sigma}_n^2) - \sigma^2)^2/(\sigma^2)^2$ (dotted line) and the relative variance $\text{Var}(\widehat{\sigma}_n^2)/(\sigma^2)^2$ (dashed line) part. Thus the development of the bias–variance trade-off can be observed easily.

The different estimators are distinguished by colours:

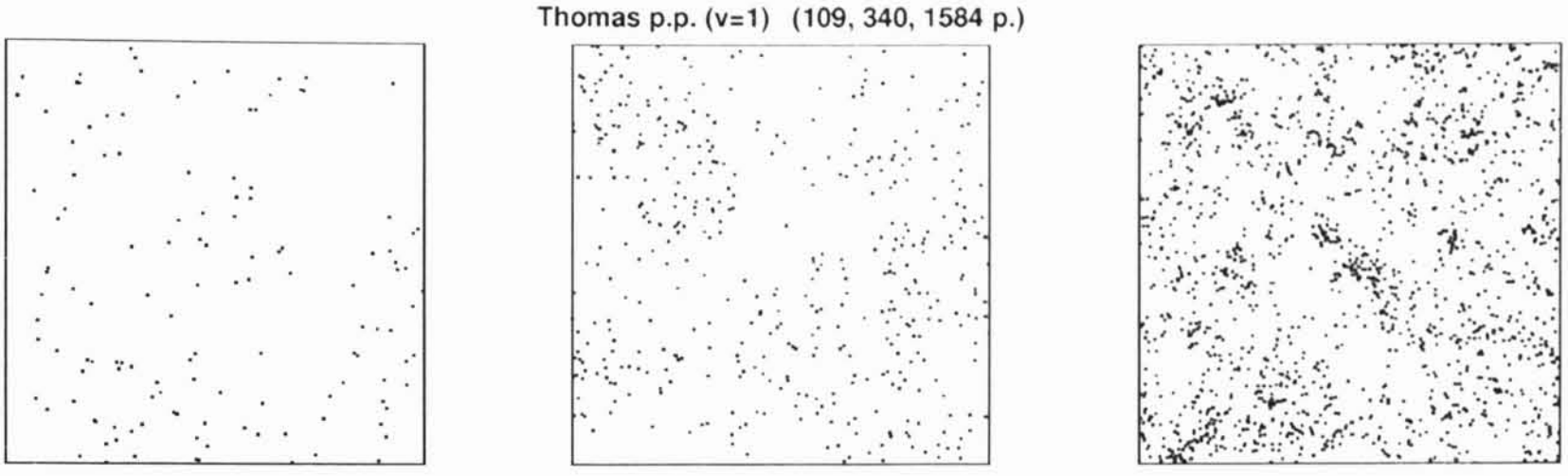
black lines	...	$\widehat{\sigma}_n^2$ with the cylinder kernel
red lines	...	$\widehat{\sigma}_n^2$ with the half-ball kernel
green lines	...	$\widehat{\sigma}_n^2$ with the cone kernel
blue lines	...	$\tilde{\sigma}_n^2$ (isotropic estimator)

In all figures the type of the point process and the size of the observation window are indicated in the upper margin. First all graphs for $W_5 = [-5, 5]^2$ are shown then $W_{10} = [-10, 10]^2$ and $W_n = [-20, 20]^2$. Moreover for the window $W_{20} = [-20, 20]^2$ are shown the more detailed graphs for the shorter range of $n\,b_n \in [1, 3]$.

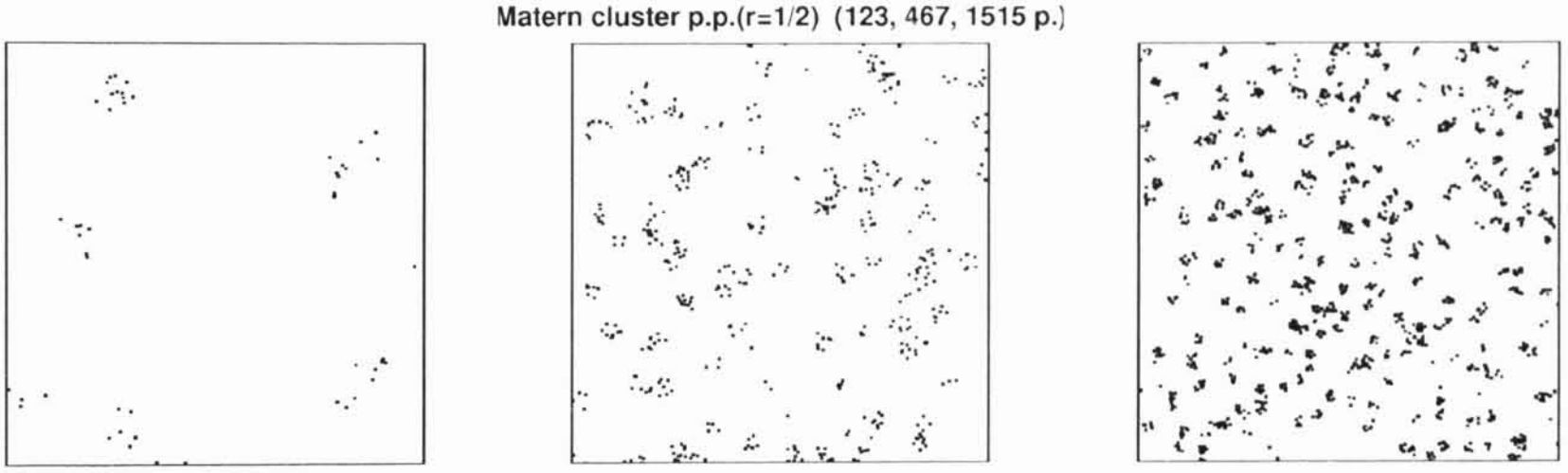
Thomas p.p. (v=1/2) (76, 345, 1460 p.)



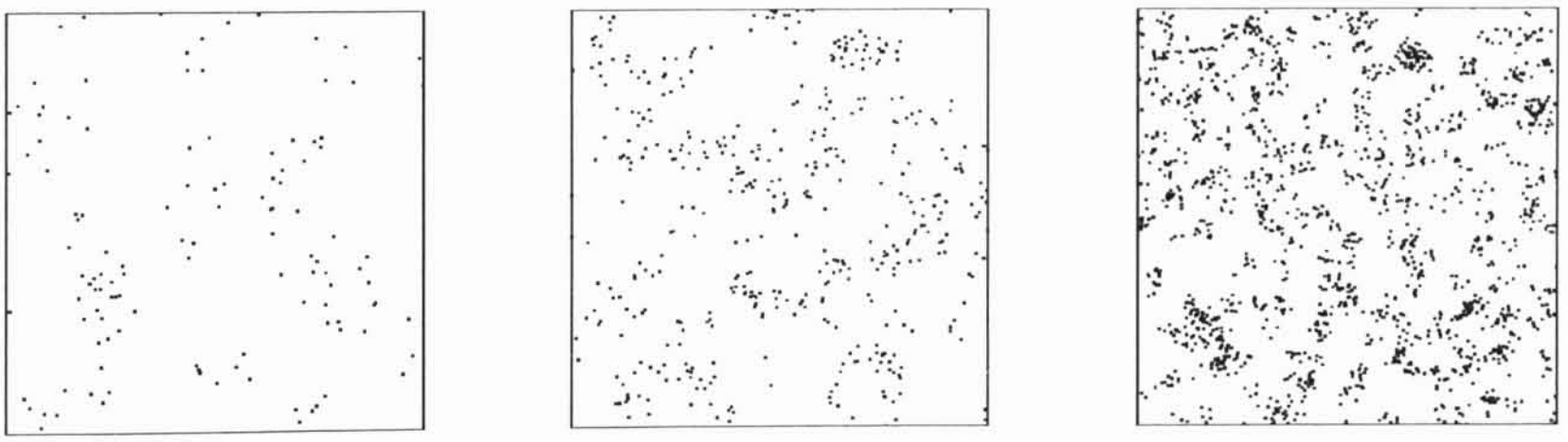
Thomas p.p. (v=1) (109, 340, 1584 p.)



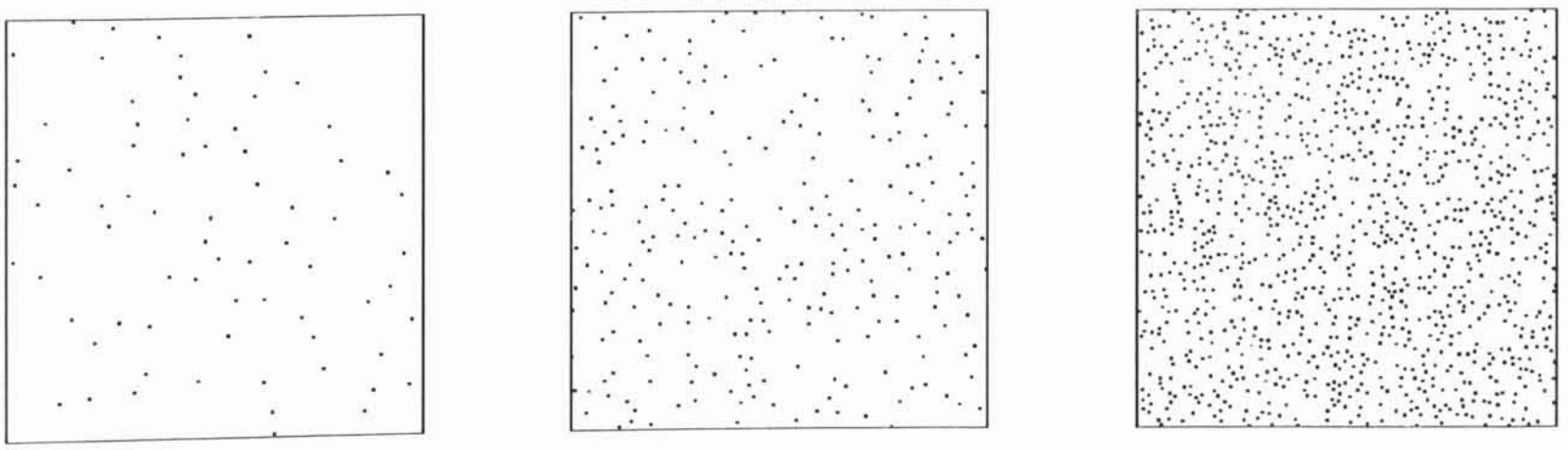
Matern cluster p.p.(r=1/2) (123, 467, 1515 p.)



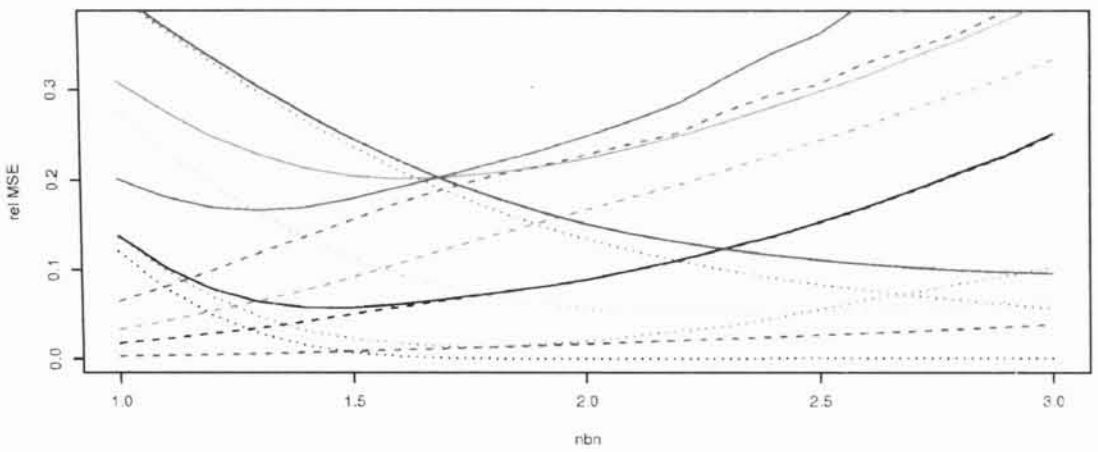
Matern cluster p.p. (r=1) (86, 396, 1591 p.)



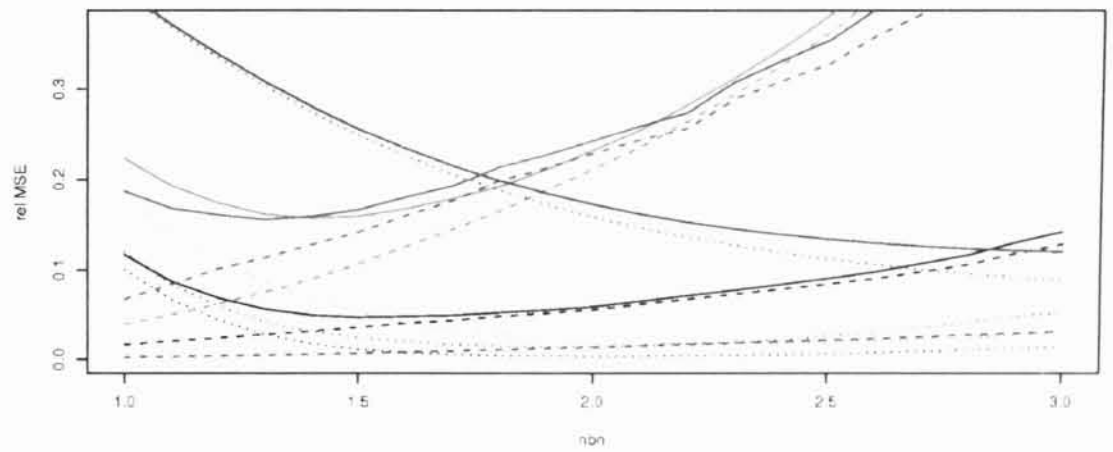
MaternII p.p. (66, 281, 1117 p.)



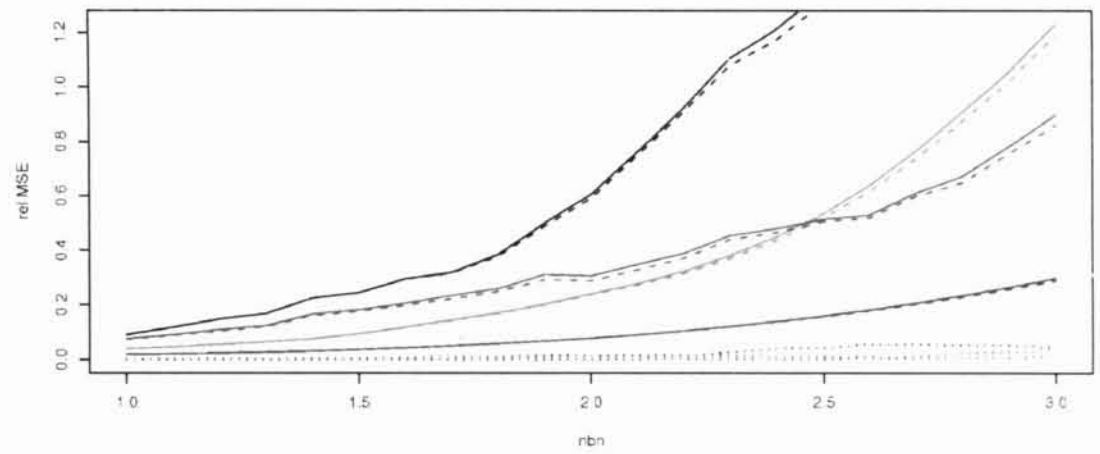
Matern cluster p.p. (r=1) on W_5



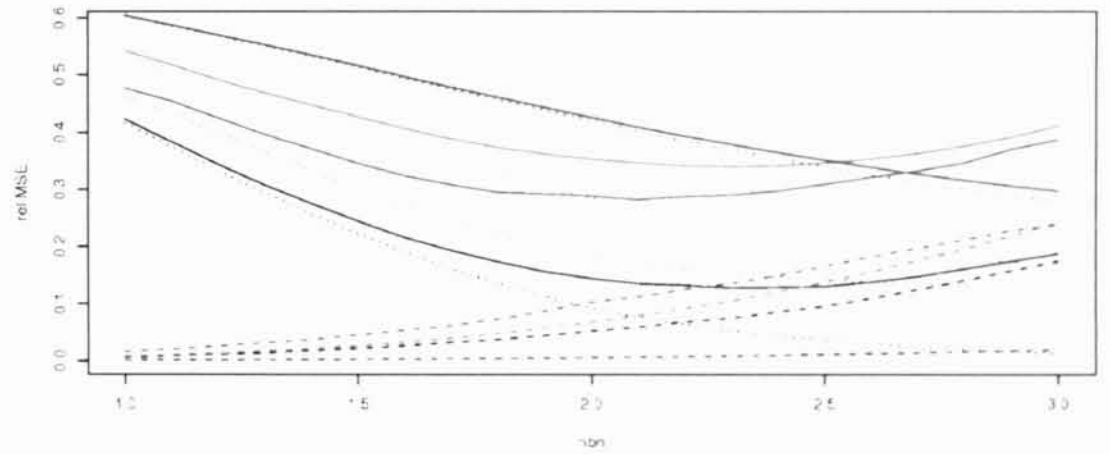
Thomas p.p. (v=1/2) on W_5



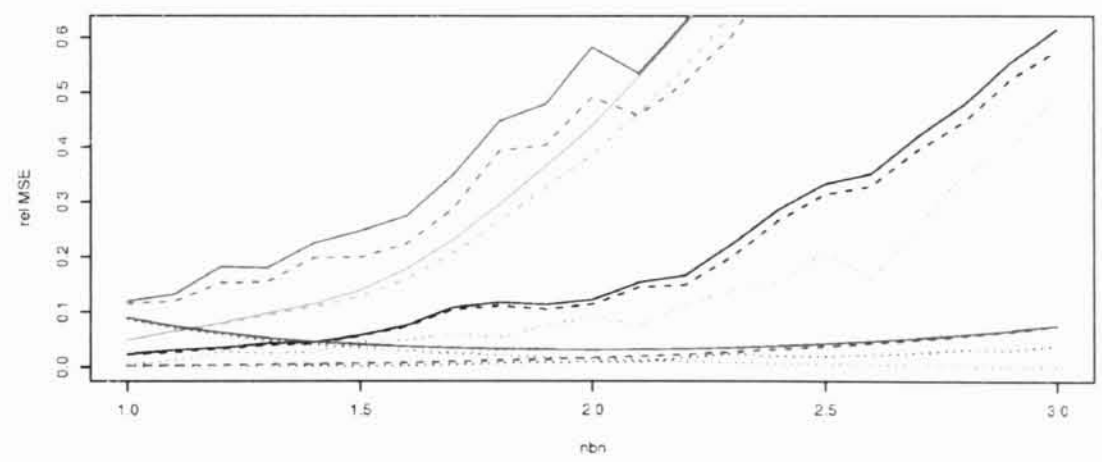
Poisson p.p. on W_5



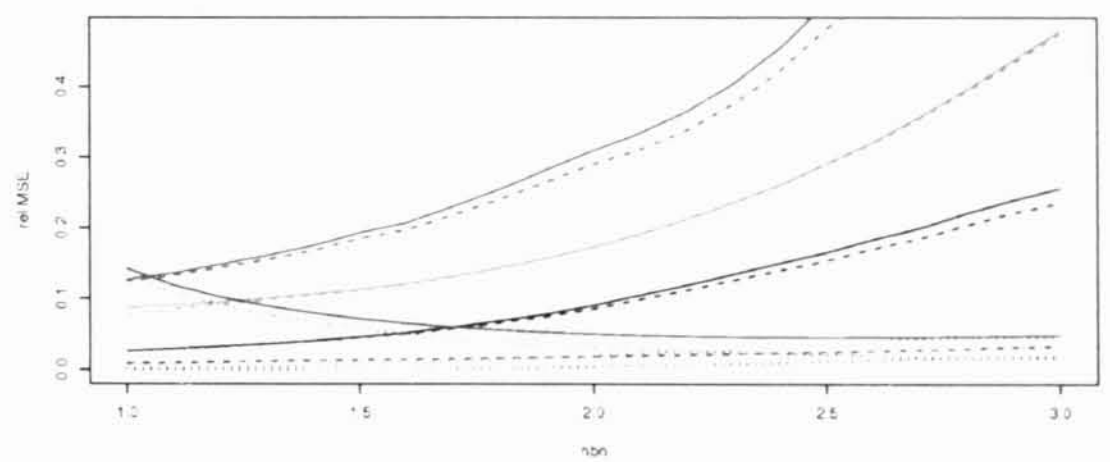
Thomas p.p. (v=1) on W_5



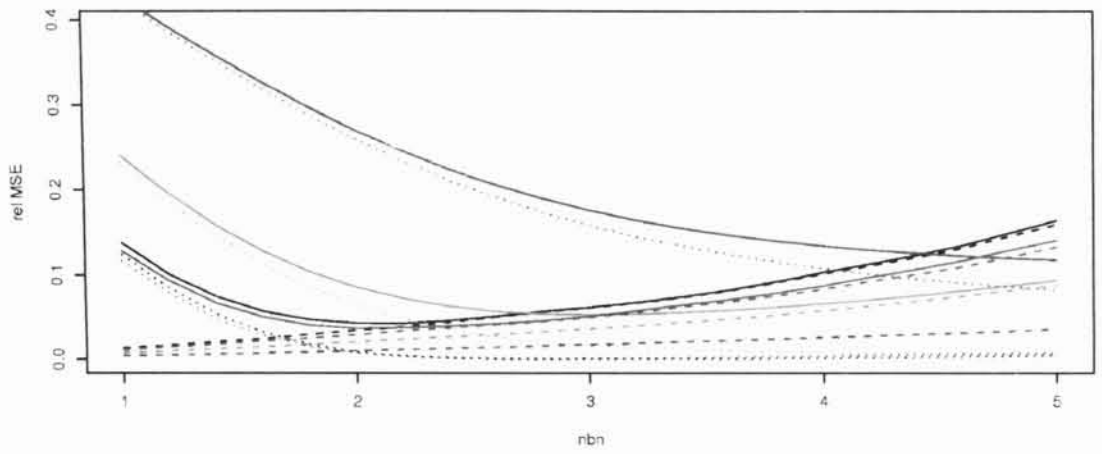
Matern II p.p. on W_5



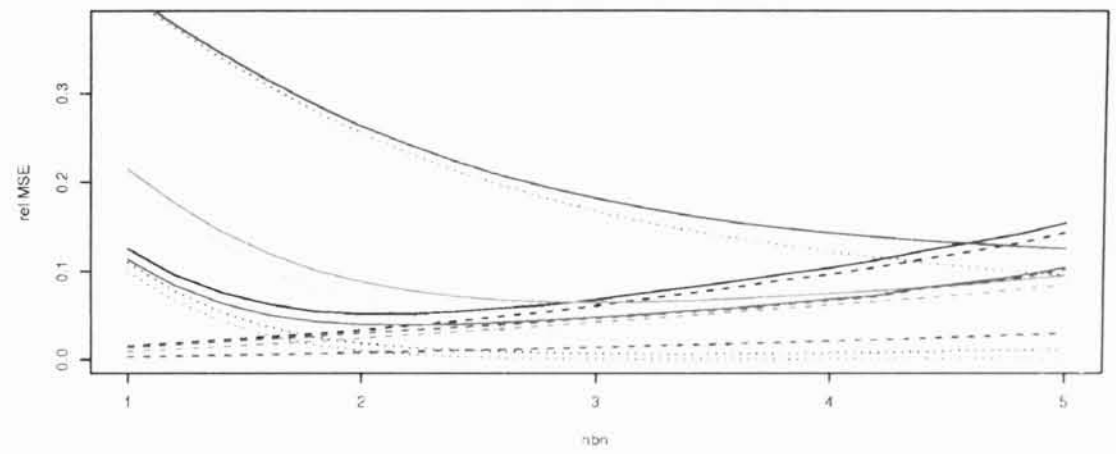
Matern cluster p.p. (r=1/2) on W_5



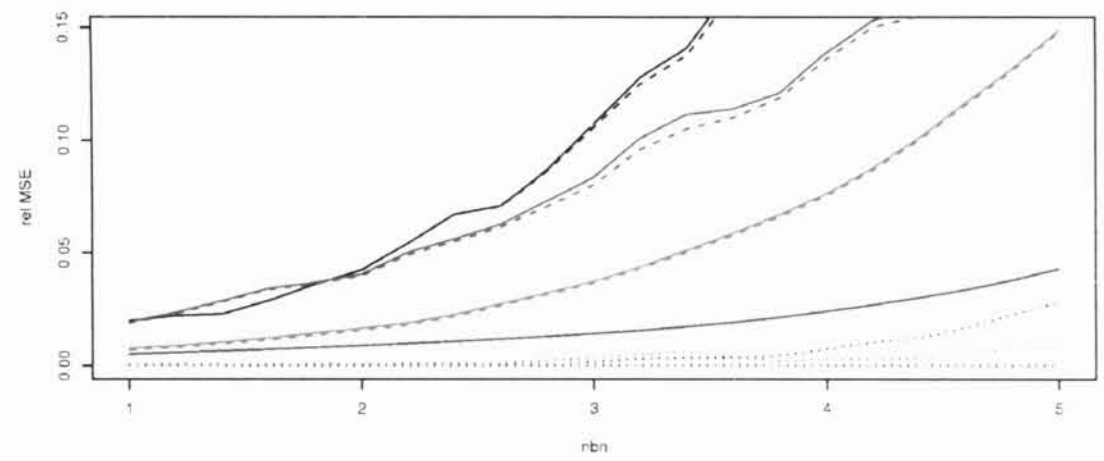
Matern cluster p.p. (r=1) on W_10



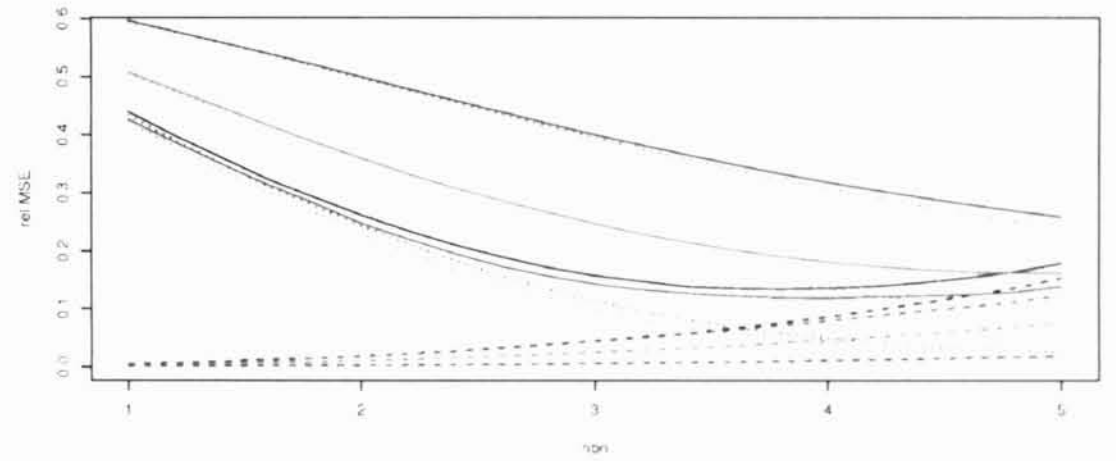
Thomas p.p. (v=1/2) on W_10



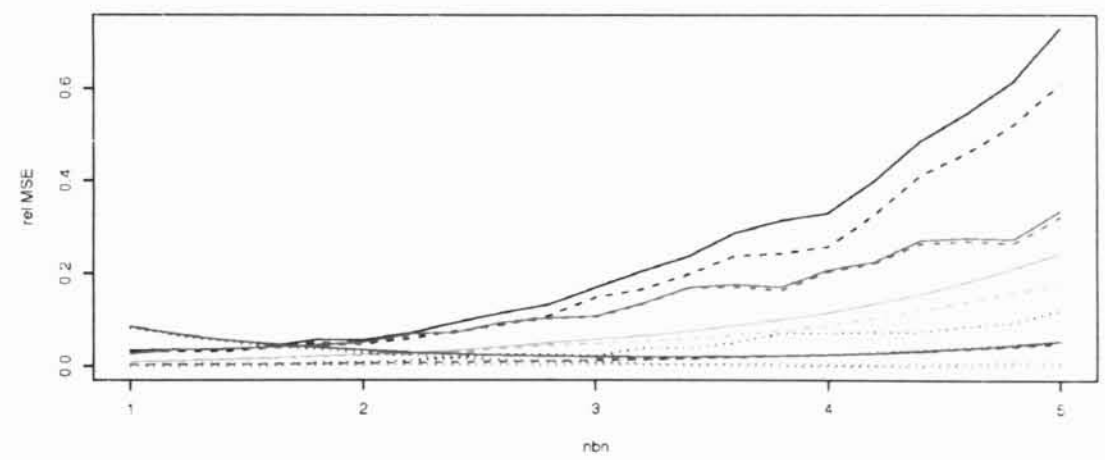
Poisson p.p. on W_10



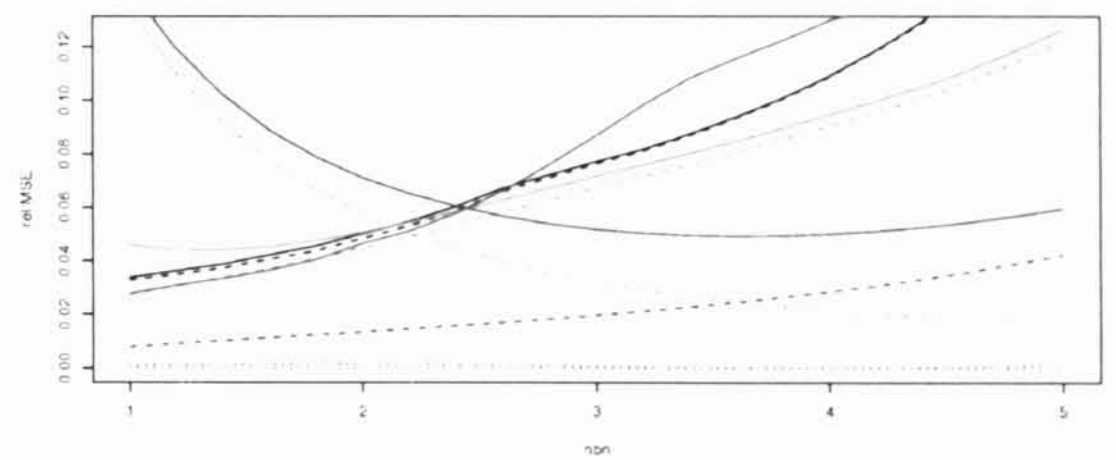
Thomas p.p. (v=1) on W_10



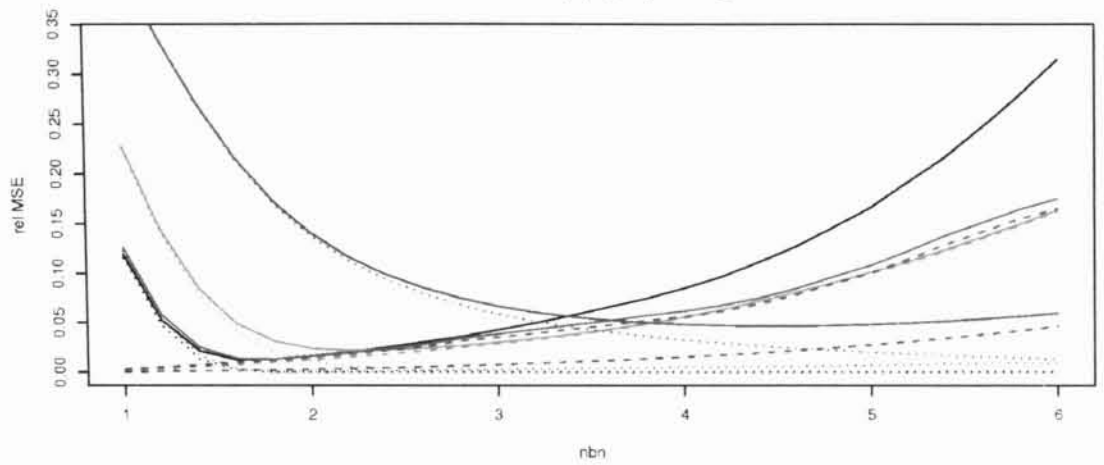
Matern II p.p. on W_10



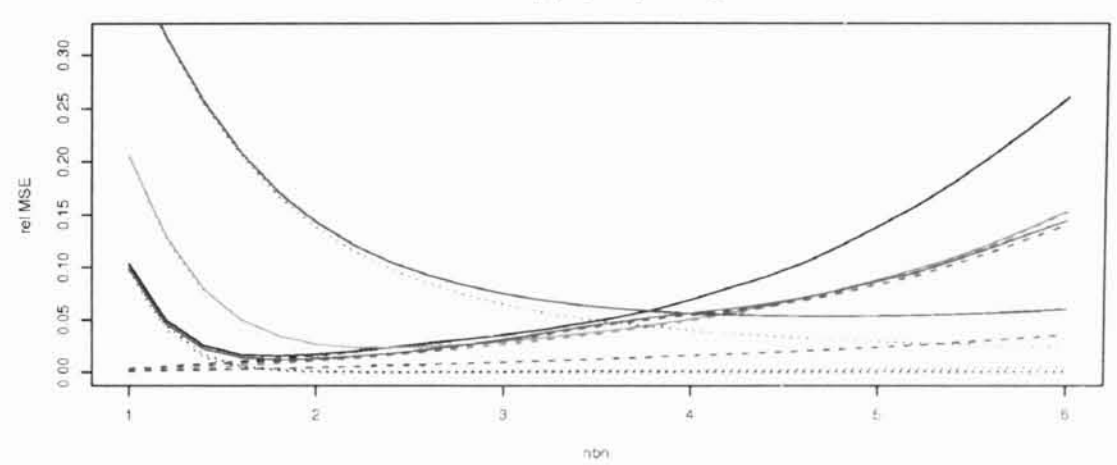
Matern cluster p.p. (r=1/2) on W_10



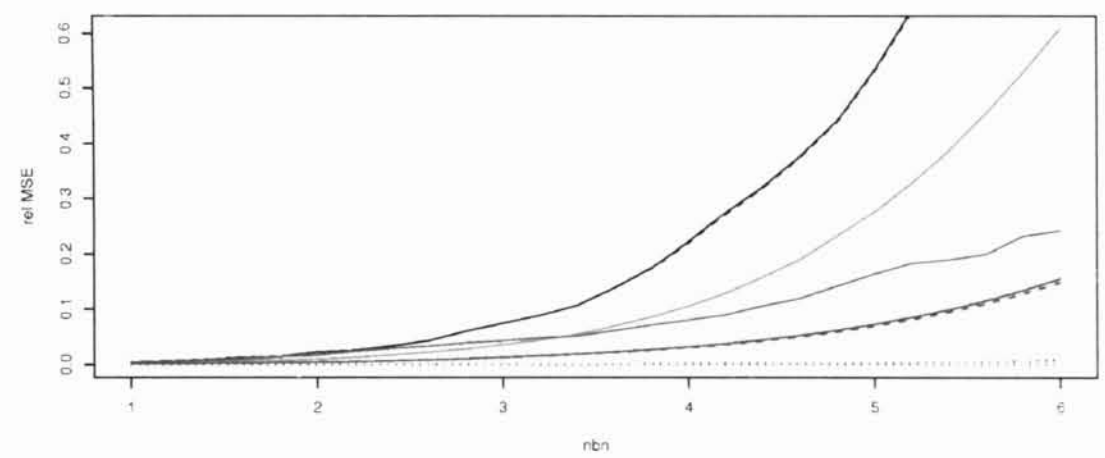
Matern cluster p.p. (r=1) on W_20



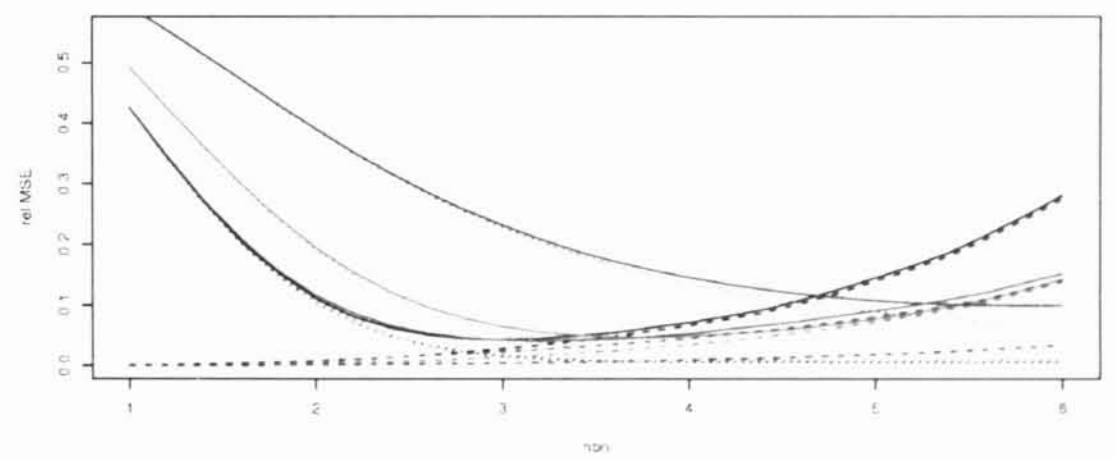
Thomas p.p. (v=1/2) on W_20



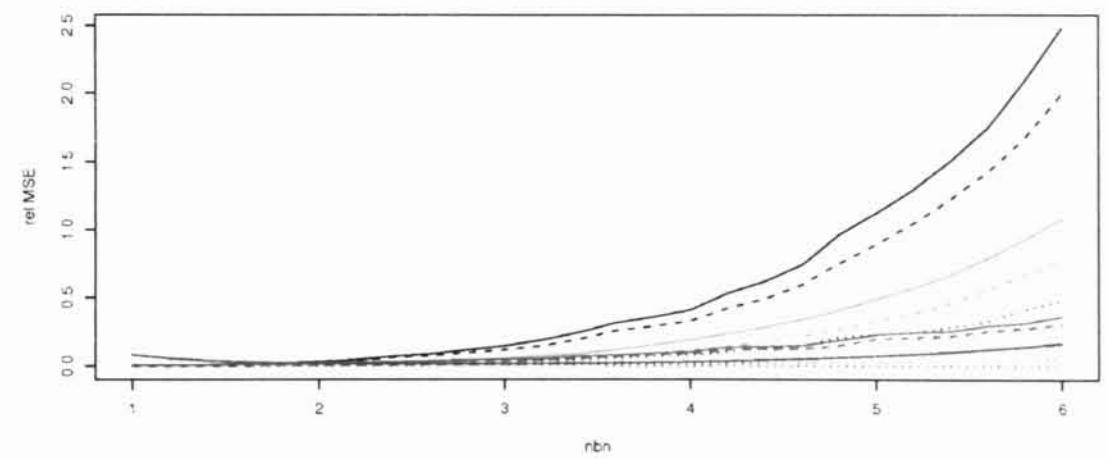
Poisson p.p. on W_20



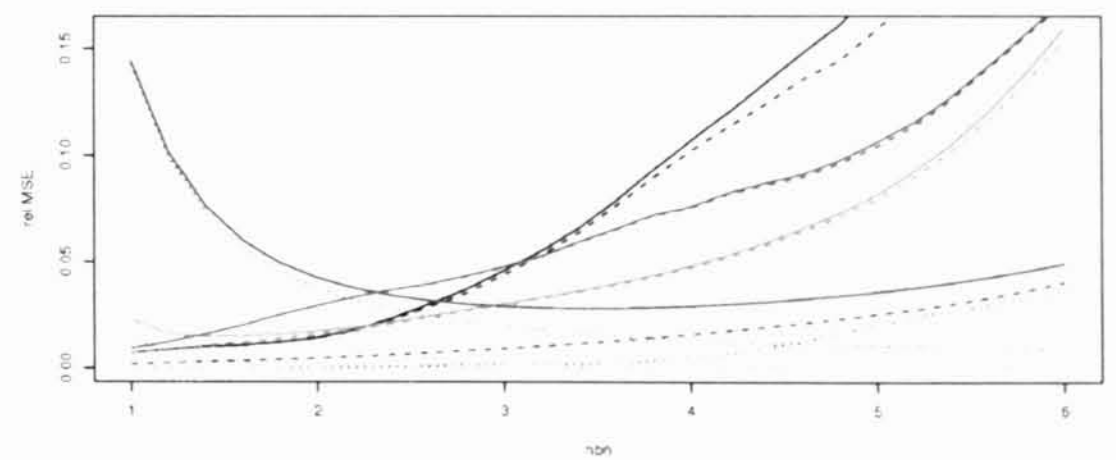
Thomas p.p. (v=1) on W_20



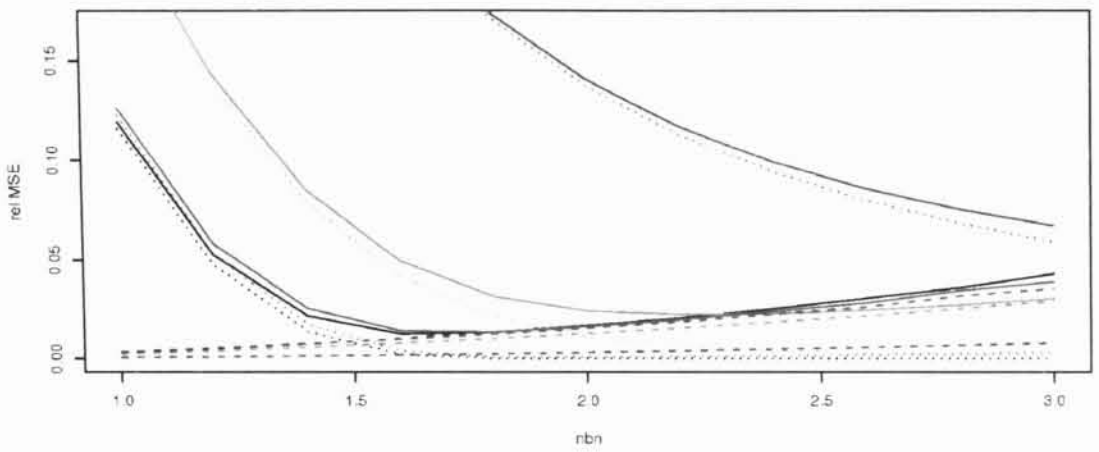
Matern II p.p. on W_20



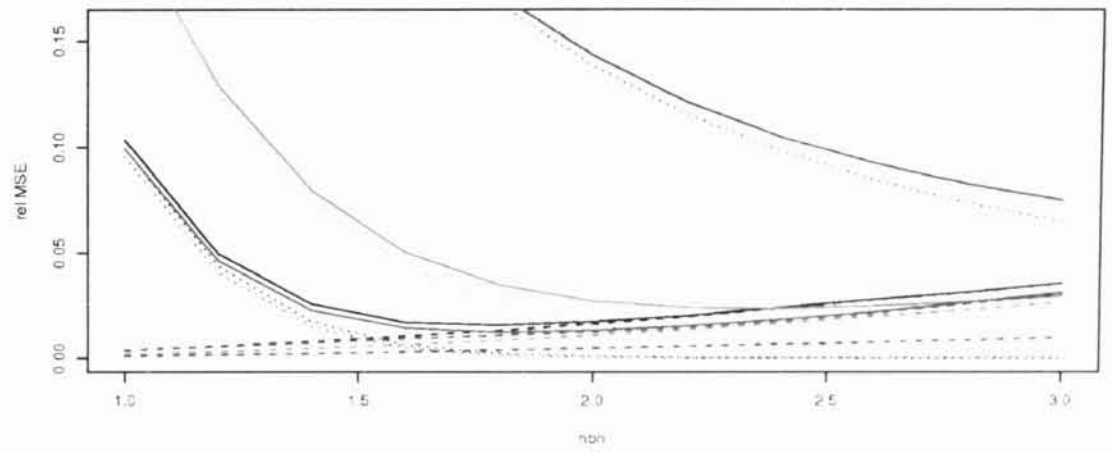
Matern cluster p.p. (r=1/2) on W_20



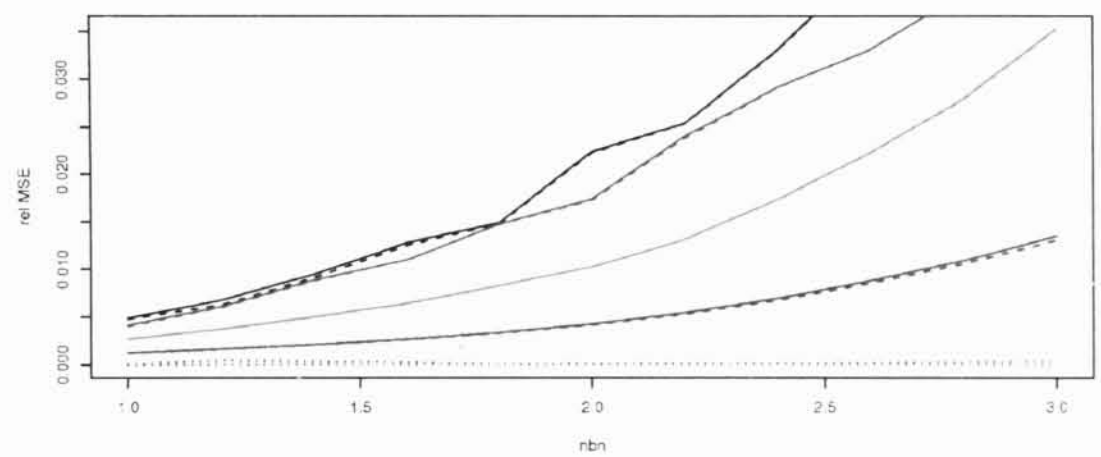
Matern cluster p.p. (r=1) on W_20



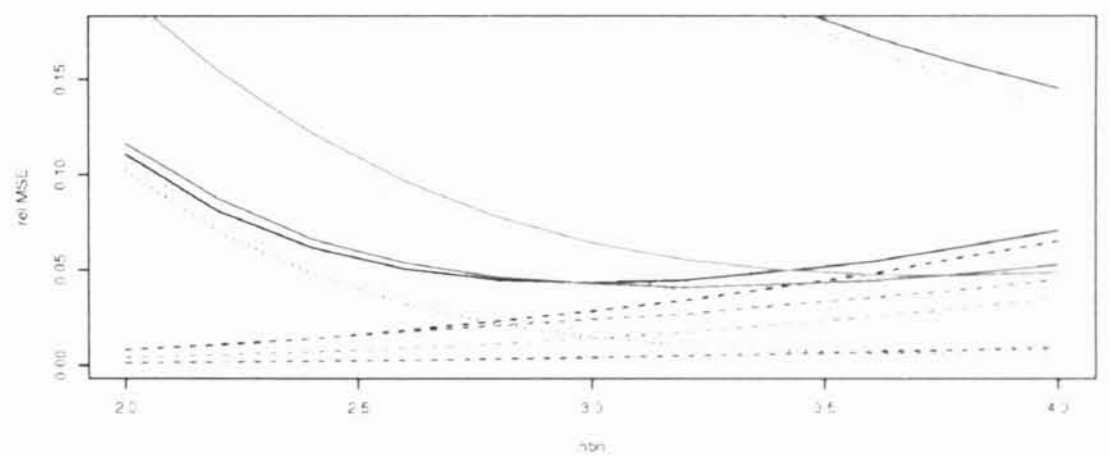
Thomas p.p. (v=1/2) on W_20



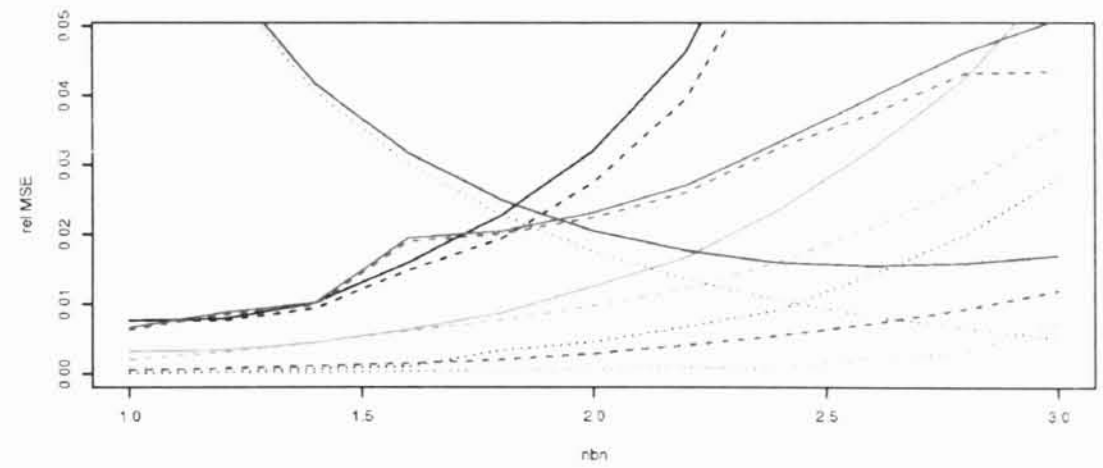
Poisson p.p. on W_20



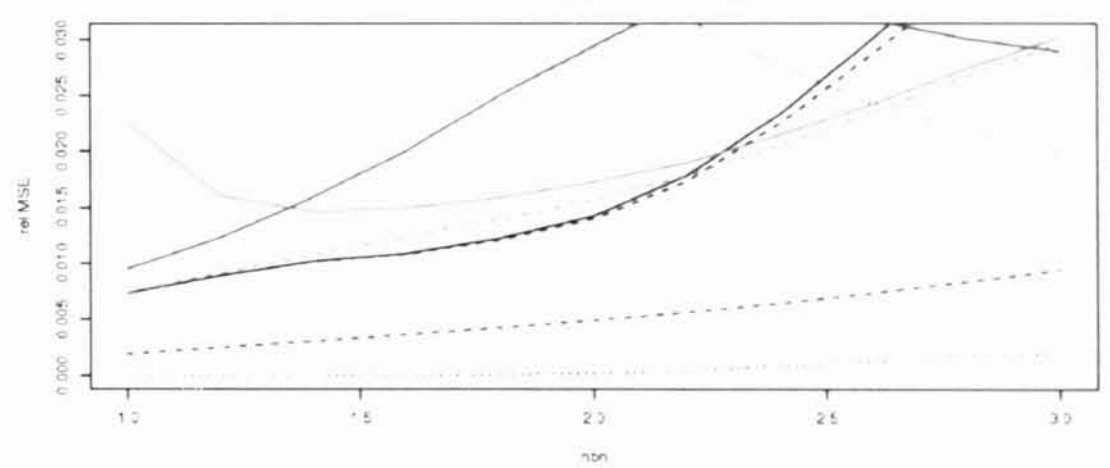
Thomas p.p. (v=1) on W_20



Matern II p.p. on W_20



Matern cluster p.p. (r=1/2) on W_20



Chapter 6

Filtration problem for Cox processes driven by Ornstein-Uhlenbeck type processes

6.1 Introduction

In this chapter we consider the problem of filtration for temporal and spatial-temporal Cox point processes driven by OU type processes.

Let Φ be a Cox point process on $\mathbb{R}_+$ or $\mathbb{R}_+ \times \mathbb{R}^k$ and suppose that the random measure Λ which is driving the Cox process (recall the Definition 3.21) is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}_+$ or $\mathbb{R}_+ \times \mathbb{R}^k$ respectively, and let us denote the corresponding (random) density function by λ. We will call λ the driving intensity of the Cox process. The problem of filtration is to estimate the unknown driving intensity λ after having observed the random point process Φ on the interval $[0, t)$ or on $[0, t) \times \mathbb{R}^k$ respectively.

The filtration problem is a classical problem which was studied by many authors (c.f. [44], [14], [45], [28], [10], etc.). We will concentrate on the situation, when the driving intensity depends on the realization of an OU type stochastic process X. We will adopt two different attitudes to the temporal and spatial-temporal case.

In the temporal case we derive a differential equation for the conditional distribution $\mathbb{P}[X(t) \mid (\Phi\ |_{[0,t)})]$ of X using the characteristic form of the differential generator, thus extending the results of [45]. In the spatial-temporal case we will work with a larger class of doubly stochastic analytic point processes defined in [14]. For such processes driven by an OU type stochastic process X we will derive a differential equation for the conditional mean $\mathbb{E}[X(t) \mid (\Phi\ |_{[0,t)\times\mathbb{R}^k})]$. Because under some minimal assumptions on the driving intensity λ the Cox processes belong to the class of analytic doubly stochastic point processes, we get the result also for the spatial-temporal OU Cox processes.

6.2 Filtration problem for temporal Cox processes

Let Φ be a Cox process on $\mathbb{R}_+$ with the driving intensity λ. We will make the following assumptions on λ

(L1) $\{X(t)\}_{t\geq 0}$ is an $\mathbb{R}^d$ valued stochastic process and the driving intensity of Φ has the form

$$\lambda(t) = \lambda(t, X(t)),$$

where $\lambda : \mathbb{R}_+ \times \mathbb{R}^d \mapsto \mathbb{R}_+$ is a positive function

(L2) It holds

$$\mathbb{E}(\lambda(t, X(t)) < \infty, \qquad \text{for all } t \in \mathbb{R}_+.$$

Definition 6.1 (counting process of Φ) Let Φ be a simple point process on $\mathbb{R}_+$. Let us define the counting process $\{N(t)\}_{t\geq 0}$ of Φ by

$$N(t) = \Phi([0, t]), \qquad t \in \mathbb{R}_+.$$

It is a nondecreasing rcll random process.

In the filtration problem our task is to find the MSE optimal estimator $\widehat{\lambda}$ of $\lambda(t, X(t))$ given $\{N(s), 0 \leq s < t\}$, i.e. the random function $\widehat{\lambda}$ that minimizes

$$\mathbb{E}[|\lambda(t, X(t)) - \widehat{\lambda}|^2 \mid N(s), 0 \leq s < t] = \mathbb{E}[|\lambda(t, X(t)) - \widehat{\lambda}|^2 \mid (\Phi \mid_{[0,t)})].$$

The solution is the conditional expectation of the driving intensity

$$\widehat{\lambda}(t) = \mathbb{E}[\lambda(t, X(t)) \mid N(s),\, 0 \leq s < t]. \tag{6.1}$$

Thus our problem reduces to the evaluation of (6.1).

To shorten the notation we will denote by hat all the conditional characteristics given the sample path of N, i.e.

$$\widehat{\mathbb{E}}[\,.\,] = \mathbb{E}[.\mid N(s),\, 0 \leq s < t].$$

For example for the conditional characteristic function of $X(t)$ we write

$$\widehat{\psi}_t(v) = \widehat{\mathbb{E}}[e^{i\langle v, X(t)\rangle}].$$

Definition 6.2 (characteristic form of the differential generator) Suppose $X(t)$ is a Markov process and there exists a nonnegative function $g_t(v, X(t))$ with finite mean such that for all $\triangle t > 0$ it holds

$$\frac{1}{\triangle t}\left|\mathbb{E}\left[e^{i\langle v, \triangle X(t)\rangle} - 1 \mid X(t)\right]\right| \leq g_t(v, X(t)), \tag{6.2}$$

where $\triangle X(t) = X(t + \triangle t) - X(t)$.

The characteristic form for the differential generator of $X(t)$ is defined as

$$\Psi_t(v \mid X(t)) = \lim_{\triangle t \downarrow 0} \frac{1}{\triangle t} \mathbb{E}\left[e^{i\langle v,\, \triangle X(t)\rangle} - 1 \mid X(t)\right], \tag{6.3}$$

if the right hand side exists.

In [45] the following theorem is proved.

Theorem 6.1 *[[45] Theorem 7.4.2] Suppose (L1) and (L2) are satisfied for the $\mathbb{R}^d$-valued driving process $\{X(t)\}_{t\geq 0}$ which is Markov stochastically continuous and suppose that the driving intensity $\lambda(t, X(t))$ is left continuous. Then the following differential equation holds for the conditional characteristic function of $X(t)$*

$$\mathrm{d}\widehat{\psi}_t(v) = \widehat{\mathbb{E}}\left[e^{i\langle v,\, X(t)\rangle}\Psi_t(v \mid X(t))\right]\mathrm{d}t + \tag{6.4}$$

$$+\widehat{\mathbb{E}}\left[e^{i\langle v,\, X(t)\rangle}(\lambda(t, X(t)) - \widehat{\lambda}(t))\right]\frac{1}{\widehat{\lambda}(t)}(\mathrm{d}N(t) - \widehat{\lambda}(t)\mathrm{d}t),$$

$$\widehat{\psi}_0(v) = \mathbb{E}\, e^{i\langle v,\, X(0)\rangle}.$$

Inverse Fourier transform of (6.4) leads to a differential equation for the conditional distribution of $X(t)$ given $\{N(s),\ 0 \leq s < t\}$. This equation can be principally solved numerically, cf.[45]. The desired estimate of the driving intensity $\widehat{\lambda}(t)$ is obtained by evaluating the expectation of $\lambda(t, X(t))$ with respect to this distribution. Our aim is not the numerical solution but the extension of the class of models for the theoretical study of the filtration problem.

6.3 Filtration problem for temporal OU Cox process

Let us now return to the filtration problem for a Cox point process Φ driven by an OU type vector process $X(t)$. Firstly we derive the characteristic form of the differential operator for such $X(t)$.

Lemma 6.2 *Let $X(t)$ be an OU type d-dimensional process given as a solution of (2.16) with $\gamma > 0$ and with the background driving Lévy process $Z(t)$ satisfying equation (2.9). If $X(t)$ has finite mean then the characteristic form for the differential generator for $X(t)$ is*

$$\Psi_t(v \mid X(t)) = -i\gamma\,\langle v,\, X(t)\rangle + \gamma\int_{\mathbb{R}^d}(e^{i\langle v,\, x\rangle} - 1)\mu(\mathrm{d}x). \tag{6.5}$$

Proof. For the Markov process $X(t)$ holds (2.11)

$$X(t+u) = e^{-\gamma u}X(t) + \int_t^{t+u} e^{-\gamma(t+u-s)}\mathrm{d}Z(\gamma s),$$

thus for $\triangle t > 0$

$$\triangle X(t) = X(t+\triangle t) - X(t) \; = X(t)(e^{-\gamma \triangle t} - 1) + \int_t^{t+\triangle t} e^{-\gamma(t+\triangle t - s)} \mathrm{d}Z(\gamma s),$$

where the first term on the right is for a fixed value of $X(t)$ nonrandom $\mathcal{O}(\triangle t)$ and it follows

$$\lim_{\triangle t \downarrow 0} \frac{1}{\triangle t} \mathbb{E}[e^{i\langle v, \triangle X(t)\rangle} - 1 | X(t)] \tag{6.6}$$

$$= \; -\gamma i \langle v, X(t)\rangle + \lim_{\triangle t \downarrow 0} \frac{1}{\triangle t} \mathbb{E}\left[\exp\left(i\left\langle v, \int_t^{t+\triangle t} e^{-\gamma(t+\triangle t - s)} \mathrm{d}Z(\gamma s)\right\rangle\right) - 1 | X(t)\right],$$

if the limit on the right hand side exists.

Denote $\mathcal{Z}_t$ the σ-algebra generated by $\{Z(\gamma s), s \le t\}$ and $\psi_{Z(1)}$ the characteristic function of $Z(1)$. Then

$$\begin{aligned}
& \mathbb{E}\left[\exp\left(i\left\langle v, \int_t^{t+\triangle t} e^{-\gamma(t+\triangle t - s)} \mathrm{d}Z(\gamma s)\right\rangle\right) - 1 | X(t)\right] \\
= \; & \mathbb{E}\left[\exp\left(i\left\langle v, \int_t^{t+\triangle t} e^{-\gamma(t+\triangle t - s)} \mathrm{d}Z(\gamma s)\right\rangle\right) - 1 | \mathcal{Z}_t\right] \\
= \; & \mathbb{E}\left[\exp\left(i\left\langle v, \int_0^{\triangle t} e^{-\gamma(\triangle t - s)} \mathrm{d}Z(\gamma s)\right\rangle\right) - 1\right] \\
= \; & \exp\left(\int_0^{\triangle t} \log(\psi_{Z(1)}(e^{-\gamma(\triangle t - s)} v))\, \mathrm{d}(\gamma s)\right) - 1.
\end{aligned}$$

In the second equality we used the fact that $Z(t)$ has independent increments. As the integrand is a continuous function of v and $(\triangle t - s)$, the integral $\int_0^{\triangle t} \log(\psi_{Z(1)}(e^{-\gamma(\triangle t - s)} v))\mathrm{d}(\gamma s)$ is also $\mathcal{O}(\triangle t)$ and it follows that for $X(t)$ with finite mean the condition (6.2) is fulfilled. From continuity of $\log(\psi_{Z(1)})$ we get further

$$\lim_{\triangle t \downarrow 0} \frac{1}{\triangle t} \left(\exp\left(\int_0^{\triangle t} \log(\psi_{Z(1)}(e^{-\gamma(\triangle t - s)} v))\mathrm{d}(\gamma s)\right) - 1\right) \tag{6.7}$$

$$= \lim_{\triangle t \downarrow 0} \frac{1}{\triangle t} \int_0^{\triangle t} \log(\psi_{Z(1)}(e^{-\gamma(\triangle t - s)} v))\mathrm{d}(\gamma s) \; = \; \gamma \log(\psi_{Z(1)}(v)).$$

Combining (6.6) and (6.7) and substituting for $\psi_{Z(1)}$ we get the assertion of the lemma. □

Now we are ready to derive the differential equation for the probability density of $\widehat{X}(t)$.

Theorem 6.3 *Under the assumptions of Theorem 6.1 and Lemma 6.2 let the Lévy measure μ have density w with respect to the Lebesgue measure. If $\mu(\mathbb{R}^d) < \infty$ the conditional probability density of $X(t)$ given $\{N(s), 0 \le s < t\}$, satisfies*

$$\begin{aligned}
\mathrm{d}\widehat{p}_t(x) \; = \; & \gamma\left(\widehat{p}_t(x)(1 - \mu(\mathbb{R}^d)) + (\widehat{p}_t * w)(x) + \left\langle x, \frac{d\widehat{p}_t(x)}{dx}\right\rangle\right)\mathrm{d}t \\
& + \widehat{p}_t(x)\left(\lambda^*(t,x) - \widehat{\lambda}^*(t)\right)\frac{1}{\widehat{\lambda}^*(t)}\left(\mathrm{d}N(t) - \widehat{\lambda}^*(t)\mathrm{d}t\right),
\end{aligned} \tag{6.8}$$

where $\widehat{p}_t * w$ *denotes the convolution of* $\widehat{p}_t$ *and* w.
Generally (also for $\mu(\mathbb{R}^d) = \infty$*) it holds*

$$\begin{aligned} \mathrm{d}\widehat{p}_t(x) &= \gamma \left(F^{-1}(\widehat{\psi}_t \cdot \log(\psi_{Z(1)})) + \widehat{p}_t(x) + \left\langle x, \frac{d\widehat{p}_t(x)}{dx} \right\rangle \right) \mathrm{d}t \\ &\quad + \widehat{p}_t(x) \left(\lambda(t,x) - \widehat{\lambda}(t) \right) \frac{1}{\widehat{\lambda}(t)} \left(\mathrm{d}N(t) - \widehat{\lambda}(t)\mathrm{d}t \right), \end{aligned} \tag{6.9}$$

where F^{-1} *denotes the inverse Fourier transform.*

Proof. From (6.5) and Theorem 6.1 it follows

$$\begin{aligned} \mathrm{d}\widehat{\psi}_t(v) &= \widehat{\mathbb{E}} \left[e^{i\langle v, X(t)\rangle} \left(\gamma \int_{\mathbb{R}^d} (e^{i\langle v, x\rangle} - 1)\mu(dx) - i\gamma \langle v, X(t)\rangle \right) \right] \mathrm{d}t \\ &\quad + \widehat{\mathbb{E}} \left[e^{i\langle v, X(t)\rangle} \left(\lambda(t, X(t)) - \widehat{\lambda}(t) \right) \right] \frac{1}{\widehat{\lambda}(t)} \left(\mathrm{d}N(t) - \widehat{\lambda}(t)\mathrm{d}t \right), \\ \widehat{\psi}_0(v) &= \mathbb{E} \left[e^{i\langle v, X(0)\rangle} \right]. \end{aligned} \tag{6.10}$$

The inverse Fourier transform F^{-1} gives

$$\begin{aligned} \mathrm{d}\widehat{p}_t(x) &= \gamma F^{-1} \left(\widehat{\mathbb{E}}[\mathrm{e}^{i\langle v, X(t)\rangle}] \cdot \int_{\mathbb{R}^d} e^{i\langle v, x\rangle} w(x)\mathrm{d}x \right) - \gamma \mu(\mathbb{R}^d)\widehat{p}_t(x) \\ &\quad + \gamma F^{-1} \left(\widehat{\mathbb{E}}[-i \langle v, X(t)\rangle \mathrm{e}^{i\langle v, X(t)\rangle}] \right) \mathrm{d}t \\ &\quad + \widehat{p}_t(x) \left(\lambda(t,x) - \widehat{\lambda}(t) \right) \frac{1}{\widehat{\lambda}(t)} \left(\mathrm{d}N(t) - \widehat{\lambda}(t)\mathrm{d}t \right). \end{aligned}$$

Now using the fact that the Fourier transform of a convolution yields a product of Fourier transforms and the equality

$$F^{-1} \left(\widehat{\mathbb{E}}[-i \langle v, X(t)\rangle \mathrm{e}^{i\langle v, X(t)\rangle}] \right) = \left\langle x, \frac{d\widehat{p}_t(x)}{dx} \right\rangle + \widehat{p}_t(x), \tag{6.11}$$

which we got using integration by parts, we get the desired equation (6.8).

Alternatively, when we cannot divide the first term in (6.10) the inverse Fourier transform gives

$$\begin{aligned} \mathrm{d}\widehat{p}_t(x) &= \gamma F^{-1} \left(\widehat{\mathbb{E}}[\mathrm{e}^{i\langle v, X(t)\rangle}] \cdot \log(\psi_{Z(1)}) \right) \mathrm{d}t + \gamma F^{-1} \left(\widehat{\mathbb{E}}[-i \langle v, X(t)\rangle \mathrm{e}^{i\langle v, X(t)\rangle}] \right) \mathrm{d}t \\ &\quad + \widehat{p}_t(x) \left(\lambda(t,x) - \widehat{\lambda}(t) \right) \frac{1}{\widehat{\lambda}(t)} \left(\mathrm{d}N(t) - \widehat{\lambda}(t)\mathrm{d}t \right), \end{aligned}$$

and once again using (6.11) we get equation (6.9). □

In the following examples we demonstrate both situations from the Theorem 6.3 for $d = 1$.

Example 6.6 In the Example 2.1 we considered the gamma OU process $X(t)$. We got the density w of the Lévy measure μ of $Z(1)$

$$w(x) = \alpha\nu e^{-\alpha x}, \quad x \geq 0, \qquad w(x) = 0, \quad x < 0.$$

Since

$$\int_0^\infty e^{-(\alpha - iy)x} \mathrm{d}x = \frac{1}{\alpha - iy}$$

holds for $\alpha > 0$, we get from Lemma 6.2

$$\Psi_t(v \mid X(t)) = \nu\gamma \left(\frac{\alpha}{\alpha - iv} - 1 \right) - iv\gamma X(t). \tag{6.12}$$

When we check finiteness of the Lévy measure of $Z(1)$ we see

$$\int_{\mathbb{R}} \alpha\nu e^{-\alpha x} \mathrm{d}x = \nu < \infty,$$

and we can use Theorem 6.3 and by plugging (6.12) in the equation (6.9) we get for the gamma OU type process

$$\begin{aligned} \mathrm{d}\widehat{p}_t(x) \;=\; & -\nu\gamma \left(\widehat{p}_t(x) - (\widehat{p}_t * \mathcal{E}_\alpha)(x)\right)\right) \mathrm{d}t \; + \\ & + \gamma \left(\widehat{p}_t(x) + x \frac{d\widehat{p}_t(x)}{dx} \right) \mathrm{d}t + \widehat{p}_t(x) \left(\lambda(t,x) - \widehat{\lambda}(t) \right) \frac{1}{\widehat{\lambda}(t)} \left(\mathrm{d}N(t) - \widehat{\lambda}(t)\mathrm{d}t \right), \end{aligned} \tag{6.13}$$

where $\mathcal{E}_\alpha$ is the density of the exponential distribution with parameter α. □

Example 6.7 In the Example 2.2 we considered the inverse Gaussian OU type process $X(t)$. We got the Lévy density of the Lévy measure μ of $Z(1)$

$$w(x) = \frac{1}{\sqrt{2\pi}} \frac{\delta}{2} \left(\frac{1}{x} + \gamma^2 \right) \frac{1}{\sqrt{x}} e^{-\frac{\gamma^2 x}{2}}, \qquad x \geq 0.$$

Here we can see that $\int_{\mathbb{R}} w(x)\mathrm{d}x = \infty$ and thus equation (6.9) from Theorem 6.3 applies and we have to use directly the characteristic functions of $X(t)$ and $Z(1)$ and inverse Fourier transform. □

6.4 Spatial-temporal doubly stochastic analytic point processes

Spatial-temporal point processes can be defined in several ways([10]). We follow the approach of [14] based on the notion of the analytic point process and an alternative definition of the conditional intensity (which was defined in Section 3 using the martingale approach). Further doubly stochastic spatial-temporal point processes are defined and the problem of filtration is studied.

A spatial-temporal point process is a point process Φ defined on $\mathcal{X} = [0, \infty) \times \mathbb{R}^k$. In this section we will identify $(\Omega, \mathcal{F}) = (\mathcal{N}, \mathfrak{N})$, with Φ being the identity map from $(\Omega, \mathcal{F}, \mathbb{P})$ to $\mathcal{N}^*$ (i.e. Φ is simple) and

$$\mathbb{E}\Phi([0,t) \times \mathbb{R}^k) < \infty \qquad \text{for any} \quad t > 0. \tag{6.14}$$

The symbol $N(t),\ t > 0$ will now denote the random variable equal to $\Phi([0,t) \times \mathbb{R}^k)$, $N(0) = 0$.

Since we have a natural ordering on the time axis it is possible to use an attitude different from the general spatial point process case. We identify the realizations $\Phi(\omega), \omega \in \Omega$ as

$$\Phi(\omega) = \{(t_1, r_1), (t_2, r_2), \dots\},$$

where $0 \le t_1 \le t_2 \le \dots$ are the times of events and $r_i \in \mathbb{R}^k$ their locations. For $j = 1, 2, \dots$ let $U_j = \{(t_1, r_1), (t_2, r_2), \dots, (t_j, r_j)\}$ and $\mathcal{F}_j \subset \mathcal{F}$ be the $\sigma-$algebra generated by U_j. Further let $\mathcal{F}_t = \sigma\{\Phi\ |_{[0,t)\times\mathbb{R}^k}\} \subset \mathfrak{N}$ be the $\sigma-$algebra generated by the past of the process up to time t, $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

Definition 6.3 (analytic point process) A spatial-temporal point process Φ is analytic if the following conditions hold:

a) $\mathbb{P}(N(t) < \infty) = 1$ for all $t \ge 0$ finite.

b) The measure $P_j(Q) = \mathbb{P}(U_j \in Q),\ Q \in \mathcal{B}(\mathbb{R}^{(k+1)j})$ is absolutely continuous w.r.t. Lebesgue measure on $\mathbb{R}^{(k+1)j},\ j = 1, 2, \dots$

c) The conditional distribution

$$F_{j+1}(t \mid \mathcal{F}_j) = \mathbb{P}(t_{j+1} < t \mid t_1, r_1, \dots, t_j, r_j),$$

$j = 0, 1, 2, \dots$ satisfies $F_{j+1}(t \mid \mathcal{F}_j) < 1$ for all finite t a.s.

According to the condition b) of Definition 6.3 there is a density $f_j[(t_1, r_1), \dots, (t_j, r_j)]$ of the first j points of Φ. The conditional density

$$f_{j+1}(t, r \mid \mathcal{F}_j) = \frac{f_{j+1}[(t_1, r_1), \dots, (t_j, r_j), (t, r)]}{f_j[(t_1, r_1), \dots, (t_j, r_j)]}, \quad j \ge 1, \qquad f_1(t, r \mid \mathcal{F}_0) = f_1(t, r),$$

enables to define

$$g_j(t, r; \omega) = \begin{cases} 0, & t_0 \le t < t_j \\ f_{j+1}(t, r \mid \mathcal{F}_j)[1 - \int_{t_j}^t \int_{\mathbb{R}^k} f_{j+1}(s, q \mid \mathcal{F}_j)\, \mathrm{d}q\, \mathrm{d}s]^{-1}. \end{cases}$$

Definition 6.4 (conditional intensity) The conditional intensity of an analytic spatial-temporal point process is defined by

$$\lambda^*(t, r) = g_{N(t)}(t, r).$$

Under the condition

$$\mathbb{E}\left[\int_0^t \int_{\mathbb{R}^k} \lambda^*(u,v)\mathrm{d}v\mathrm{d}u\right]^2 < \infty \tag{6.15}$$

for any $B \in \mathcal{B}(\mathbb{R}^k)$ and $0 \le t < u < \infty$ it holds a.s. (cf. (3.7))

$$\mathbb{E}[\Phi([t,u) \times B) \mid \mathcal{F}_t] = \mathbb{E}\left[\int_t^u \int_B \lambda^*(u,v)\mathrm{d}v\mathrm{d}u \mid \mathcal{F}_t\right].$$

Also, the likelihood of a realization of the process observed up to time t is expressed by means of the conditional intensity as

$$L_t(\omega) = \Pi_{i=1}^{N(t)} \lambda^*(t_i, r_i) \exp\left[-\int_0^t \int_{\mathbb{R}^k} \lambda^*(u,v)\mathrm{d}v\,\mathrm{d}u\right],$$

assuming the product to be equal to 1 if $N(t) = 0$.

Lemma 6.4 *Suppose that for the intensity Λ of a spatial-temporal Poisson point process Π holds (6.14) and that Λ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}_+ \times \mathbb{R}^k$. Then Π is analytic and its conditional intensity is equal to the intensity function $\lambda^* = \lambda$.*

Proof. Condition a) from the Definition 6.3 follows from (6.14), b) and c) from the absolute continuity of Λ with respect to the Lebesgue measure on $\mathcal{X}$. $\lambda^* = \lambda$ follows from (6.15) and the definition of the Poisson point process. □

Now we proceed to the doubly stochastic processes.

Let $(\Omega^*, \mathcal{S}, P^*)$ be another probability space and

$$P : \Omega^* \times \mathcal{F} \mapsto [0,1]$$

a probability kernel such that for each $\omega^* \in \Omega^*$ $P(\omega^*, .)$ is the distribution of a spatial-temporal point process, denoted Φ^*.

Definition 6.5 (doubly stochastic point process) A doubly stochastic (spatial-temporal) point process is a process with distribution $\bar{P}(A) = P'(A \times \Omega^*),\ A \in \mathcal{F}$, where

$$P'(A \times S) = \int_S P(\omega^*, A) P^*(\mathrm{d}\omega^*),\ S \in \mathcal{S}.$$

If for each $\omega^* \in \Omega^*$ $P(\omega^*, .)$ is the distribution of a Poisson point process on $\mathcal{X}$, the doubly stochastic process with distribution $\bar{P}$ is a Cox process.

This is a natural notion of a Cox process in our setting of the product probability space $\Omega \times \Omega^*$, c.f. Definition 3.21 on a simple space.

Theorem 6.5 *[[14] Theorem 3] Let Φ be a doubly stochastic spatial-temporal point process such that Φ^* is analytic for each $\omega^* \in \Omega^*$ and the corresponding conditional intensities $\lambda^*(t, r; \omega, \omega^*)$ are jointly measurable in the arguments t, r, ω and ω^* and*

$$\int_{\Omega \times \Omega^*} \left[\int_0^t \int_{\mathbb{R}^d} \lambda^*(s, q; \omega, \omega^*) \mathrm{d}q\mathrm{d}s \right] P'(\mathrm{d}\omega \times \mathrm{d}\omega^*) < \infty. \tag{6.16}$$

Then (i) Φ is analytic,
(ii) the conditional intensity $\widehat{\lambda}^*(t, r; \omega)$ of the point process Φ is

$$\widehat{\lambda}^*(t, r; \omega) = \mathbb{E}'[\lambda^*(t, r; \omega, \omega^*) | \mathcal{F}_t \times \mathcal{S}_0],$$

where $\mathcal{S}_0 = \{\emptyset, \Omega^*\}$ is the trivial $\sigma-$algebra and $\mathbb{E}'$ the expectation w.r.t. P'.

Combining the theorem with Lemma 6.4 a direct consequence for the Cox processes is

Corollary 6.6 *Let Φ be a Cox process on $\mathcal{X}$ satisfying (6.14), such that its driving measure $\Lambda(\omega^*)$ is absolutely continuous with respect to the Lebesgue measure on $\mathcal{X}$ a.s. Then Φ is an analytic doubly stochastic point process.*

The likelihood of a general doubly stochastic process is equal to

$$\bar{L}_t(\omega) = \Pi_{i=1}^{N(t)} \widehat{\lambda}^*(t_i, r_i) \exp\left[- \int_0^t \int_{\mathbb{R}^k} \widehat{\lambda}^*(u, v) \, \mathrm{d}v \, \mathrm{d}u \right].$$

6.5 Filtration for spatial-temporal doubly stochastic analytic processes driven by OU type processes

In the above setting used for doubly stochastic spatial-temporal point processes, the filtration problem means doing inference about the value of the unobserved ω^* if we observe the realization $\{(t_1, r_1), \ldots, (t_{N(t)}, r_{N(t)})\}$ of Φ in $[0, t) \times \mathbb{R}^k$. We can express the conditional probability

$$P^*(S \mid \mathcal{F}_t)(\omega) = P'(\Omega \times S \mid \mathcal{F}_t \times \mathcal{S}_0)(\omega, \omega^*), \qquad S \in \mathcal{S},$$

using the likelihood of Φ.

Theorem 6.7 *[[14] Theorem 5] For a doubly stochastic process Φ satisfying the conditions of Theorem 6.5 it holds for all $S \in \mathcal{S}$ and $t \in [0, \infty)$*

$$P^*(S \mid \mathcal{F}_t)(\omega) = \frac{1}{\bar{L}_t(\omega)} \int_S L_t(\omega, \omega^*) P^*(d\omega^*),$$

$\bar{P}$ a.s. on $\mathcal{F}_t$. Here $L_t(\omega, \omega^*)$ are likelihoods for Φ^* given ω^*.

Corollary 6.8 *[[14]] Let Φ be a doubly stochastic analytic point process satisfying conditions of Theorem 6.5 and suppose that the conditional intensities $\lambda^*(t,r,\omega,\omega^*)$ are left continuous in t and continuous in r. For an $(\mathcal{F}_t \times \mathcal{S})$-adapted spatial-temporal vector random field $Y(t,r;\omega,\omega^*)$ defined on $[0,\infty)\times\mathbb{R}^k\times\Omega\times\Omega^*$ denote $\widehat{Y}(t,r) = \mathbb{E}'[Y(t,r;\omega,\omega^*)|\mathcal{F}_t\times\mathcal{S}_0]$. It holds*

$$\widehat{Y}(t,r) = \mathbb{E}^*(Y(t,r)\exp(\xi(t))),$$

where $\mathbb{E}^*$ is the mean value with respect to P^* and

$$\xi(t) = -\int_0^t\int_{\mathbb{R}^k}(\lambda^*(s,r) - \widehat{\lambda}^*(s,r))\,\mathrm{d}r\mathrm{d}s + \int_0^t\int_{\mathbb{R}^k}\log\frac{\lambda^*(s,r)}{\widehat{\lambda}^*(s,r)}\,\Phi(\mathrm{d}s,\mathrm{d}r). \tag{6.17}$$

Now we have at our disposal all we need to approach the problem of filtration for doubly stochastic point processes driven by OU type temporal stochastic processes. Thus in the rest of this chapter we suppose that the probability space $(\Omega^*,\mathcal{S},P^*)$ corresponds to the probability space of an OU type process $\{X(t)\}$.

We are interested in the filtered estimate $\widehat{X}(t) = \mathbb{E}'[X(t,\omega,\omega^*)|\mathcal{F}_t\times\mathcal{S}_0]$. To be able to derive the differential equation for $\widehat{X}$ we need the Itó differential formula for a special type of vector process. Stochastic integral with respect to the Poisson random measure is understood in the sense of [9], Subsection 8.1.4.

Lemma 6.9 *Let Φ be a doubly stochastic point process from Theorem 6.5. Let for the vector process $\zeta(t) = (\zeta^1(t),\ldots,\zeta^m(t))$ hold*

$$\mathrm{d}\zeta(t;\omega,\omega^*) = \alpha(t;\omega,\omega^*)\mathrm{d}t + C\,\mathrm{d}Z(\gamma t;\omega,\omega^*) + \int_{\mathbb{R}^k}\delta(t,r;\omega,\omega^*)\Phi((\mathrm{d}t,\mathrm{d}r);\omega,\omega^*), \tag{6.18}$$

where α,δ are random m-dimensional vectors, C is an $m\times d$ matrix of real numbers, and $\{Z(t)\}$ is a d-dimensional Lévy process with the characteristic function given by (2.9). We assume that α, δ have sample paths which are left continuous in t and continuous in r and are $(\mathcal{F}_t\times\mathcal{S}_t)$-adapted, where $\mathcal{S}_t = \sigma\{Z(s);s\le t\}\subset\mathcal{S}$. Let η be a $C^1(\mathbb{R}^m)$ scalar function. Then η satisfies the following stochastic differential equation a.s.

$$\begin{aligned}\mathrm{d}\eta(\zeta(t)) &= \left\langle\alpha(t),\frac{\partial\eta}{\partial\zeta}(t)\right\rangle\mathrm{d}t + \int_{\mathbb{R}^k}(\eta(\zeta(t_-)+\delta(t,r)) - \eta(\zeta(t_-)))\,\Phi(\mathrm{d}t,\mathrm{d}r)\\ &\quad + \int_{\mathbb{R}^d}(\eta(\zeta(t_-)+Cy) - (\eta(\zeta(t_-))))\,J_Z(\mathrm{d}\gamma t,\mathrm{d}y),\end{aligned} \tag{6.19}$$

where $\zeta(t_-) = \lim_{s\uparrow t}\zeta(s)$ and J_Z is a Poisson random measure satisfying (2.8).

Proof. We need to show that

$$\begin{aligned}\eta(\zeta(t)) - \eta(\zeta(0)) &= \int_0^t\left\langle\alpha(t),\frac{\partial\eta}{\partial\zeta}(t)\right\rangle\mathrm{d}t\\ &\quad + \int_0^t\int_{\mathbb{R}^k}(\eta(\zeta(t_-)+\delta(t,r)) - \eta(\zeta(t_-)))\,\Phi(\mathrm{d}t,\mathrm{d}r)\\ &\quad + \int_0^t\int_{\mathbb{R}^d}(\eta(\zeta(t_-)+Cy) - (\eta(\zeta(t_-))))\,J_Z(\mathrm{d}\gamma t,\mathrm{d}y),\end{aligned} \tag{6.20}$$

holds a.s.

Let us first suppose that the Lévy process Z has finite Lévy measure μ, i.e. $\{Z(t)\}$ has only finite number of jumps on every interval $[0,t)$ a.s. Then we can proceed like in the classical proof of the Itó formula in [16], cf. [14]. We consider a sequence of partitions $\mathcal{T}_j = \{\tau_{1,j}, \ldots, \tau_{j+1,j}\}$ of the interval $[0,t)$ defined for $j \in \mathbb{N}$ by

$$\tau_{i,j} = \frac{i-1}{j}\, t, \qquad i = 1, 2 \ldots j+1.$$

For any such partition we can write

$$\eta((\zeta(t)) - \eta(\zeta(0)) = \sum_{i=1}^{j} \eta(\zeta(\tau_{i+1,j})) - \eta(\zeta(\tau_{i,j})) = \sum_{i=1}^{J} \eta(\zeta(\tau_{i,j}) + \triangle\zeta(\tau_{i,j})) - \eta(\zeta(\tau_{i,j})),$$

where

$$\triangle\zeta(\tau_{i,j}) = \zeta(\tau_{i+1,j}) - \zeta(\tau_{i,j}) = \triangle A_{i,j} + \triangle B_{i,j} + \triangle D_{i,j},$$

and

$$\begin{aligned}
\triangle A_{i,j} &= \int_{\tau_{i,j}}^{\tau_{i+1,j}} \alpha(s;\omega,\omega^*)\mathrm{d}s, \\
\triangle B_{i,j} &= C\int_{\tau_{i,j}}^{\tau_{i+1,j}} \mathrm{d}Z(\gamma s;\omega,\omega^*), \\
\triangle D_{i,j} &= \int_{\tau_{i,j}}^{\tau_{i+1,j}} \int_{\mathbb{R}^k} \delta(t,r;\omega,\omega^*)\Phi((\mathrm{d}t,\mathrm{d}r);\omega,\omega^*).
\end{aligned}$$

Let us write $\eta((\zeta(t)) - \eta(\zeta(0)) = \sum_{A,j} + \sum_{B,j} + \sum_{D,j}$ where

$$\begin{aligned}
\sum_{A,j} &= \sum_{i=1}^{j} \eta(\zeta(\tau_{i,j}) + \triangle\zeta(\tau_{i,j})) - \eta(\zeta(\tau_{i,j}) + \triangle B_{i,j} + \triangle D_{i,j}), \\
\sum_{B,j} &= \sum_{i=1}^{j} \eta(\zeta(\tau_{i,j}) + \triangle B_{i,j} + \triangle D_{i,j}) - \eta(\zeta(\tau_{i,j}) + \triangle D_{i,j}), \\
\sum_{D,j} &= \sum_{i=1}^{j} \eta(\zeta(\tau_{i,j}) + \triangle D_{i,j}) - \eta(\zeta(\tau_{i,j})).
\end{aligned}$$

From the mean-value theorem we have

$$\lim_{j\to\infty} \sum_{A,j} = \int_0^t \left\langle \alpha(t),\, \frac{\partial\eta}{\partial\zeta}(t) \right\rangle \mathrm{d}t. \tag{6.21}$$

Because Φ is an analytic doubly stochastic point process it has only finite number of points in $[0,t)\times\mathbb{R}^k$ thus $\int_0^t \int_{\mathbb{R}^k} \mathrm{d}\Phi(\mathrm{d}t,\mathrm{d}r)$ has only finite number of jumps. The same holds for

$$C\int_0^t \mathrm{d}Z(\gamma s) = C\int_0^t \int_{\mathbb{R}^d} y J_Z(d(\gamma s, y))$$

when the Lévy measure μ of Z is finite on $\mathbb{R}^d$. Moreover the jumps of Φ and J_Z arise at different times. Thus for big enough $m(\omega, \omega^*)$ there is at most one jump of Φ and J_Z in each $[\tau_{i,j}, \tau_{i+1,j})$ for any $j \geq m(\omega, \omega^*)$ and

$$\sum\nolimits_{D,j} = \sum_{l=1}^{\Phi([0,t)\times\mathbb{R}^k;\omega,\omega^*)} \eta(\zeta(\tau_{l,j}) + \delta(t_l, r_l)) - \eta(\zeta(\tau_{l,j}))$$

$$\sum\nolimits_{B,j} = \sum_{q=1}^{\int_0^t \mathbf{1}[\int_{\mathbb{R}^d} J_Z(\gamma s, \mathrm{d}y)\neq 0]\mathrm{d}s} \eta(\zeta(\tau_{q,j}) + Cy_q) - \eta(\zeta(\tau_{q,j})).$$

Here (t_l, r_l), $(\gamma t_q, y_q)$ are the occurrence points of Φ, J_Z, respectively and the intervals $[\tau_{l,j}, \tau_{l+1,j})$ and $[\tau_{q,j}, \tau_{q+1,j})$ are the elements of the partition containing the points t_l, t_q respectively. From the continuity of η and the existence of left limits for ζ it follows

$$\begin{aligned}\lim_{j\to\infty} \sum\nolimits_{D,j} &= \sum_{l=1}^{\Phi([0,t)\times\mathbb{R}^k;\omega,\omega^*)} \eta(\zeta(t_{l-})) + \delta(t_l, r_l) - \eta(\zeta(t_{l-})) \\ &= \int_0^t \int_{\mathbb{R}^k} (\eta(\zeta(t_-) + \delta(t, r)) - \eta(\zeta(t_-)))\, \Phi(\mathrm{d}t, \mathrm{d}r) \qquad (6.22)\end{aligned}$$

$$\begin{aligned}\lim_{j\to\infty} \sum\nolimits_{B,j} &= \sum_{q=1}^{\int_0^t \mathbf{1}[\int_{\mathbb{R}^d} J_Z(\gamma s, \mathrm{d}y)\neq 0]\mathrm{d}s} \eta(\zeta(t_{q-})) + Cy_q) - \eta(\zeta(t_{q-})) \\ &= \int_0^t \int_{\mathbb{R}^d} (\eta(\zeta(t_-) + Cy) - (\eta(\zeta(t_-))))\, J_Z(\mathrm{d}\gamma t, \mathrm{d}y), \qquad (6.23)\end{aligned}$$

almost surely. By combining (6.21), (6.23) and (6.22) we get (6.20).

For general Lévy processes $\{Z(t)\}$ we know from the remark behind Theorem 2.3 that every Lévy process can be decomposed as $Z(t) = Z^\epsilon(t) + R^\epsilon(t)$, where $Z^\epsilon(t)$ is a Lévy process with finite number of jumps in any bounded interval and $R^\epsilon(t)$ is a mean-zero square integrable martingale with $\mathrm{Var}(R^\epsilon(t)) \to 0$ as $\epsilon \to 0$. Denote by ζ^ϵ the process defined by (6.18) but with the Lévy process $\{Z^\epsilon(t)\}$ instead of $\{Z(t)\}$.

Suppose that η and its first derivative is bounded by a constant K. Then

$$|(\eta(\zeta(t_-) + Cy) - (\eta(\zeta(t_-)))| \leq KCy$$

thus the right hand side of (6.20) is finite since $Z(t) = \int_0^t \int_{\mathbb{R}^d} y J_Z(\mathrm{d}(y, s))$ is. Moreover

$$|\eta(\zeta(t)) - \eta(\zeta^\epsilon(t))|^2 \leq K^2 (R^\epsilon(t))^2.$$

Thus

$$\lim_{\epsilon\to 0} \eta(\zeta^\epsilon(t)) = \eta(\zeta(t)),$$

in $L^2(\mathbb{P})$. But the equation (6.20) holds for ζ^ϵ and taking the limits on both sides we get the equality (6.20) also for $Z(t)$.

A general Lévy process fulfilling the assumptions of Lemma 6.9 is of finite variation thus if we define the sets $A_M = \{(\omega, \omega^*) : \zeta(s; \omega, \omega^*) \leq M \text{ for all } s \leq t\}$, then $A_M \to (\Omega \times \Omega^*)$ as $M \to \infty$. But η is bounded with bounded first derivatives on A_M and (6.20) holds on A_M. Thus taking limit $M \to \infty$ the validity of (6.20) follows for any ζ satisfying the assumptions of the theorem. □

Now we are ready to derive the differential equation for the posterior mean $\hat{X}$ of $X(t)$.

Theorem 6.10 *Let Φ be a doubly stochastic point process from Theorem 6.5 driven by an OU type d-dimensional process $X(t)$ given by (2.16) with $\gamma > 0$ and with the Lévy process $Z(t)$ satisfying equation (2.9). Suppose that the conditional intensities $\lambda^*(t, r)$ are left continuous in t and continuous in r. Then the conditional mean $\widehat{X}(t)$ satisfies*

$$\begin{aligned} \mathrm{d}\widehat{X}(t) &= -\gamma \widehat{X}(t)\mathrm{d}t - \int_{\mathbb{R}^d} (\widehat{\mathbb{E}}[X(t_-)\lambda^*(t,r)] - \widehat{X}(t_-)\widehat{\lambda^*}(t,r))\mathrm{d}r\mathrm{d}t \qquad (6.24) \\ &\quad + \int_{\mathbb{R}^k} (\widehat{\mathbb{E}}[X(t_-)\lambda^*(t,r)] - \widehat{X}(t_-)\widehat{\lambda^*}(t,r)) \frac{1}{\widehat{\lambda^*}(t,r)} \Phi(\mathrm{d}t, \mathrm{d}r) + \mathrm{d}\widehat{Z}(\gamma t), \end{aligned}$$

Proof. Let $\zeta(t) = [X(t), \xi(t)]$ be a vector process with $(d+1)$ components. Combining (2.16) and (6.17) we get the stochastic differential equation for ζ

$$\mathrm{d}\zeta(t) = \begin{pmatrix} -\gamma X(t) \\ -\int_{\mathbb{R}^k} (\lambda^* - \widehat{\lambda^*})\mathrm{d}r \end{pmatrix} \mathrm{d}t + \begin{pmatrix} 0 \\ \int_{\mathbb{R}^k} \log \frac{\lambda^*}{\widehat{\lambda^*}} \Phi(\mathrm{d}t, \mathrm{d}r) \end{pmatrix} + \begin{pmatrix} \mathrm{d}Z(\gamma t) \\ 0 \end{pmatrix}.$$

For $i = 1, \ldots, d$ let

$$\eta_i(\zeta(t)) = X_i(t) \exp(\xi(t)).$$

Then η_i are continuously differentiable functions and we have

$$\frac{\partial \eta_i}{\partial \zeta_j} = \begin{cases} \mathrm{e}^{\xi(t)} & i = j \\ 0 & j \neq i, d+1 \\ \eta_i & j = d+1 \end{cases}$$

Thus we can use Lemma 6.9 for ζ and each η_i with

$$\begin{aligned} \alpha(t) &= [-\gamma X(t), -\int_{\mathbb{R}^k} (\lambda^* - \widehat{\lambda^*})\mathrm{d}r], \\ \delta(t,r) &= \left[0, \ldots, 0, \log \frac{\lambda^*}{\widehat{\lambda^*}}\right], \\ C &= \begin{pmatrix} \mathbf{1} \\ \mathbf{0} \end{pmatrix}, \end{aligned}$$

where $\mathbf{1}$ is $d \times d$ identity matrix and $\mathbf{0}$ is $1 \times d$ matrix of zeros. Using the equalities

$$\eta_i(\zeta(t_-) + Cy) - (\eta_i(\zeta(t_-))) = (X(t) + y)\mathrm{e}^{\xi(t_-)} - X(t)\mathrm{e}^{\xi(t_-)} = y\mathrm{e}^{\xi(t_-)},$$

and

$$\eta_i(\zeta(t_-)+\delta(t,r))-\eta_i(\zeta(t_-))=X_i(t_-)\mathrm{e}^{\xi(t\ )+\log\frac{\lambda^*}{\widehat{\lambda^*}}}-X_i(t_-)\mathrm{e}^{\xi(t_-)}=\eta_i(\zeta(t_-))\left(\frac{\lambda^*}{\widehat{\lambda^*}}-1\right),$$

we obtain

$$\begin{aligned}\mathrm{d}\eta_i(\zeta(t)) &= -\gamma X_i(t)\mathrm{e}^{\xi(t-)}\mathrm{d}t-\eta_i(\zeta(t_-))\int_{\mathbb{R}^k}(\lambda^*-\widehat{\lambda^*})\,\mathrm{d}r\mathrm{d}t \qquad (6.25)\\ &\quad +\eta_i(\zeta(t_-))\int_{\mathbb{R}^k}(\lambda^*-\widehat{\lambda^*})\frac{1}{\widehat{\lambda^*}}\,\Phi(\mathrm{d}t\times\mathrm{d}r)+\int_{\mathbb{R}^d}ye^{\xi(t)}J_Z(\mathrm{d}\gamma t\times\mathrm{d}y).\end{aligned}$$

Now from Lemma 6.8 we have $\widehat{X}_i(t)=\mathbb{E}^*(\eta_i(\zeta(t)))$ and from Fubini's theorem

$$\mathrm{d}\widehat{X}_i(t)=\mathbb{E}^*(\mathrm{d}\eta_i(\zeta(t))).$$

Therefore taking expectations on both sides of (6.25) and writing

$$\mathbb{E}^*\left[\int_{\mathbb{R}^d}ye^{\xi(t)}J_Z(\mathrm{d}\gamma t\times\mathrm{d}y)\right]=\mathbb{E}^*\left[e^{\xi(t)}\int_{\mathbb{R}^d}yJ_Z(\mathrm{d}\gamma t\times\mathrm{d}y)\right]=\mathrm{d}\widehat{Z}(\gamma t),$$

we obtain the equation (6.24). □

Corollary 6.11 *Let Φ be a spatial-temporal Cox process from Corollary 6.6 driven by an OU type d-dimensional process $X(t)$ from the Theorem 6.10. Suppose that the conditional intensities $\lambda^*(t,r)$ are left continuous in t and continuous in r. Then the conditional mean $\widehat{X}(t)$ satisfies equation (6.24).*

Bibliography

[1] Armstrong, P. (1991). *Species patterning in heath vegetation of the Northern Sandplain.* Honours Thesis, Murdoch University, Western Australia.

[2] Baddeley, A. J., Møller, J., and Waagepetersen, R. (2000). Non- and semiparametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54(3): 329–350.

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

[4] Baddeley, A. J. and van Lieshout, M. N. M. (1995). Area-interaction point processes. *Ann. Inst. Statist. Math.*, 47: 601–619.

[5] Barndorff-Nielsen, O. E. (1978). *Information and Exponential Families in Statistical Theory.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Ltd., Chichester.

[6] Barndorff-Nielsen O., Schmiegel J. (2003). *Lévy based tempo-spatial modelling; with applications to turbulence.* Preprint 2003/20, MaPhySto, Aarhus University.

[7] Barndorff-Nielsen O., Shephard N. (2001). Non-Gaussian Ornstein-Uhlenbeck based models and some of their uses in financial economics. *JRSS B*, 63: 167-241.

[8] Beneš, V. and Prokešová, M. (2005). Nonlinear filtration in doubly stochastic point processes. In *Proceedings of the 4th International Conference Aplimat 2005* (ed. Kováčová, M.), SjF STU Bratislava, 415-420.

[9] Cont, R. and Tankov, P. (2004). *Financial Modelling with Jump Processes.* Chapman and Hall/CRC, Boca Ranton.

[10] Daley, D. J. and Vere-Jones, D. (1988). *An introduction to the theory of point processes.* Springer Verlag, New York.

[11] Daley, D. J. and Vere-Jones, D. (2003). *An introduction to the theory of point processes, Vol I: Elementary Theory and Methods*, 2nd ed., Springer Verlag, New York.

[12] Diggle, P. J. (1983). *Statistical Analysis of Point Patterns.* Chapman and Hall/CRC, Boca Ranton.

[13] Federer, H. (1969). *Geometric Measure Theory* Springer Verlag, New York.

[14] Fishman P.M., Snyder D. (1976). The statistical analysis of space-time point processes. *IEEE Transactions on Information Theory*, 22: 257-274.

[15] Geyer, C. (1999). Likelihood inference for spatial point patterns. In *Stochastic Geometry – Likelihood and Computation*, (eds. Barndorff-Nielsen, O. E., Kendall, W. S. and van Lieshout, M. N. M.), Chapman & Hall, London, pp. 79–140.

[16] Gihman, I. and Dorogoncev, A. (1965). On stability of solutions of stochastic differential equations. *Ukrain. Mat. Z.*, 6: 229 250.

[17] Grabarnik, P. and Chiu, S. N. (2002). Goodness-of-fit test for complete spatial randomness against mixtures of regular and clustered spatial point processes. *Biometrika*, 89(2): 411–421.

[18] Hahn, U., Jensen, E. B. V., van Lieshout, M. N. M., and Nielsen, L. S. (2003). Inhomogeneous spatial point processes by location-dependent scaling. *Adv. Appl. Prob.*, 35: 319–336.

[19] Heinrich, L. (1988). Asymptotic Gaussianity of some estimators for reduced factorial moment measures and product densities of stationary Poisson cluster processes. *Statistics*, 19: 87–106.

[20] Heinrich, L. (1991). Goodness-of-fit tests for the second moment function of a stationary multidimensional Poisson Process. *Statistics*, 22: 245 268.

[21] Heinrich, L. (1994). Normal approximation for some mean-value estimates of absolutely regular tesselations. *Mathematical Methods of Statistics*, 3(1): 1–24.

[22] Heinrich, L. and Prokešová, M. (2005). On estimating the asymptotic variance. *(tentative title), in preparation.*

[23] Jacod, J. (1975). Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 31:235–253.

[24] Jensen, E. B. V. (1998). *Local Stereology.* World Scientific, London.

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

[26] Jurek, Z. J. and Mason, J. D. (1993). *Operator-limit Distributions in Probability Theory.* John Wiley & Sons Ltd., New York.

[27] Kallenberg, O. (2002). *Foundations of Modern Probability.* Springer Verlag, New York, second edition.

[28] Karr A.F. (1983). State estimation for Cox processes on general spaces. *Stochastic Processes and their Applications*, 14: 209–232.

[29] Karr A.F. (1986). *Point Processes and Their Statistical Inference*, Marcel Dekker, Inc., New York.

[30] Lieshout, M. N. M. van (2000). *Markov point processes and their applications*. Imperial College Press, London.

[31] Lieshout, M. N. M. van, and Molchanov I. (1998). Shot noise weighted processes: A new family of spatial point processes. *J. Commun. Stat., Stochastic Models*, 14: 715–734.

[32] Møller, J., and Waagepetersen, R. (2003). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton.

[33] Nguyen, X. X. and Zessin, H. (1976). Punktprozesse mit Wechselwirkung. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 37: 91 126.

[34] Nielsen, L. S. (2001). *Point process models allowing for interaction and inhomogeneity*. PhD Thesis, Department of Mathematical Sciences, University od Aarhus.

[35] Nielsen, L. S. and Jensen, E. B. V. (2003). Statistical inference for transformation inhomogeneous point processes. *Scand. J. Statist.*, 31: 131 142.

[36] Ogata, Y. and Tanemura, M. (1986). Likelihood estimation of interaction potentials and external fields of inhomogeneous spatial point patterns. In *Proc. Pacific Statistical Congress – 1985*, (eds. Francis, I. S., Manly, B. F. J., and Lam, F. C.), 150–154. Amsterdam, Elsevier. *J. London Math. Soc.*, 15:188–192.

[37] Ohser, J. (1983). On estimators for the reduced second-moment measure of point processes. *Math. Operationsf. Statist., Ser. Statistics*, 14: 63 71.

[38] Prokešová, M., Hahn,U., Jensen, E.B.V. (2005). Statistics for Locally Scaled Point Processes. In *Case Studies in Spatial Point Process Models,* (eds. Baddeley, A., Gregori, P., Mateu, J., Stoica, R. and Stoyan, D.), Springer-Verlag, Berlin. *(to appear).*

[39] Ripley, B. D. (1976). The second order analysis of stationary point processes. *J. Appl. Prob.*, 13: 255–266.

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

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

[42] Sato K. (1999). *Lévy processes and infinitely divisible distributions.* Cambridge University Press.

[43] Sato K. (2001). Basic results on Lévy processes. In *Lévy Processes – Theory and Applications*, (eds. Barndorff-Nielsen O., Mikosch T., Resnick S.), Birkäuser, Boston.

[44] Snyder D.L. (1972). Filtering and detection for doubly stochastic Poisson processes. *IEEE Trans. Inform. Theory*, 18: 91–102.

[45] Snyder D., Miller M. (1991). *Random Point Processes in Time and Space.* Springer Verlag, New York.

[46] Stein, M. L. (1993). Asymptotically optimal estimation for the reduced second moment measure of point processes. *Biometrika*, 80(2): 443–449.

[47] Stoyan, D., Kendall, W. S., Mecke, J. (1995). *Stochastic Geometry and its Applications.* John Wiley & Sons Ltd., Chichester, second edition.

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

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

[50] Schneider, R., Weil, W. (2000). *Stochastische Geometrie.* B.G.Teubner Stuttgart, Leipzig.