Likelihood Based Inference and Diagnostics for Spatial Data Models

Rubak, Ege Holger

Publication date:
2010

Document Version
Early version, also known as pre-print

PhD Report Series - No. 18 - 2010

Likelihood Based Inference and Diagnostics for Spatial Data Models

Ege Rubak

AALBORG UNIVERSITY
Department of Mathematical Sciences

Likelihood Based Inference and Diagnostics for Spatial Data Models

Ege Rubak

Thesis submitted:	August 2010
Thesis defended:	October 15, 2010
PhD degree conferred:	November 10, 2010
PhD supervisor:	Professor Jesper Møller Department of Mathematical Sciences Aalborg University
PhD committee:	Associate Professor Kasper Berthelsen Department of Mathematical Sciences Aalborg University
	Associate Professor Jean-Francois Coeurjolly GIPSA-lab and Laboratory Jean Kuntzmann Grenoble University
	Professor Jens Ledet Jensen Department of Mathematical Sciences Aarhus University

DEPARTMENT OF MATHEMATICAL SCIENCES
Fredrik Bajers Vej 7 G
9220 Aalborg, Denmark
http://www.math.aau.dk

ISSN 1601-8346

Aalborg University

Department of Mathematical Sciences

PhD Thesis

Likelihood Based Inference and Diagnostics for Spatial Data Models

Ege Rubak

Preface

This thesis has been funded by the Danish Agency for Science, Technology and Innovation, grant 645-06-0528, *International PhD student.* The thesis is the result of work carried out at Aalborg University and during two long term visits to the University of Chicago and the University of Western Australia.

Chapter 1 gives a short introduction to some well known parts of the statistical theory for spatial data, which is the topic of this thesis. This material is the background for the subsequent Chapters 2-4 presenting the three journal papers constituting the main part of this PhD thesis:

Baddeley, A., E. Rubak and J. Møller (2010). Score, pseudo-score and residual diagnostics for goodness-of-fit of spatial point process models. Submitted to *Statistical Science.*

Møller, J. and E. Rubak (2010). A model for positively correlated count variables. *International Statistical Review* 78, 65-80.

Rubak, E., J. Møller and P. McCullagh (2010). Statistical Inference for a Class of Multivariate Negative Binomial Distributions. Submitted to *Bernoulli.*

The papers are presented in their journal form, and as a consequence the notation is not necessarily consistent between chapters and some material is presented in several chapters. However, this allows each chapter to be read independently of the others. Chapter 5 presents some work that has been initiated during the PhD study, but is not yet finished. The last pages of the thesis contain a complete list of all references used in the thesis.

Summary in English

The topic of this thesis is spatial statistics, which deals with data originating from different spatial locations. Spatial data are collected in many fields of science such as agriculture, astronomy, biology, epidemiology, and physics. There is a natural interest in statistical models and methodology for dealing with such diverse data. The thesis focuses on general models and methodology, which can potentially be applied in a range of different scientific settings rather than analysis of specific datasets.

As a preliminary a short review of spatial statistics is given. Spatial data discussed in the thesis are of two distinct types: *spatial point pattern data* composed of the random locations of an event of interest are analyzed using point process models, and *lattice data* composed of the values of a random quantity at fixed locations are analysed using random field models.

The fundamental point process model is the Poisson point process which is a model allowing for spatial variation in the abundance of points, but not for interaction between points. In the thesis it is reviewed how the score test has been applied to Poisson process models in the literature to test for a significant spatially varying abundance of points explained by a spatially varying covariate. In many applications the Poisson process is a too simple and unrealistic model, and in the thesis it is described how the score test can be applied in the case of much more general point processes with interactions between points. In classical statistics the score test is commonly used for model selection and for model validation. In the point process literature model validation is often based on comparing a functional summary statistic of the data with its expectation under the model. The thesis lends theoretical support to this procedure by showing it corresponds to a score test for goodness-of-fit of the model. Further goodness-of-fit diagnostics are derived based on the score test and score test approximations related to the point process residuals recently developed in the literature.

Lattice data can in general be of any type, but in this thesis only non-negative integer valued data are considered. This is a prominent example of lattice data, which e.g. appears when data correspond to counting something of interest at the fixed spatial locations. This could be cases of a disease counted within a number of administrative regions. Such data can be analyzed using the α-permanental random fields studied in this thesis. These provide a flexible model class for lattice data with positive associations between sites, meaning that sites close to each other tend to be alike. The probabilistic properties of α-permanental random fields are reviewed and some new results are given. In particular it is detailed how simulations from an α-permanental random field can be generated using a so-called Poisson-randomisation. While the likelihood of an α-permanental random field model can be expressed on closed form, it is usually very computer intensive to evaluate. To accommodate this problem it is shown how inference can be based either on approximate evaluation of the likelihood or on computationally simpler quantities such as the quasi-likelihood or composite likelihood.

The last part of the thesis describes two works in progress. First, a class of point process models which can be used to model the location and size of trees is introduced, and it is described how likelihood inference can be conducted. Second, so-called determinantal point processes are briefly reviewed and a scheme for Bayesian inference is discussed.

Summary in Danish

Denne afhandling omhandler rumlig statistik, hvilket er den gren af den statistiske teori, der omhandler data fra forskellige lokationer. Rumlige data kan stamme fra adskillige grene af videnskaben som f.eks. astronomi, biologi, epidemiologi, fysik og landbrugsvidenskab. Dette har skabt en naturlig interesse for statistiske modeller og metoder, der kan behandle sådanne forskelligartede data. Afhandlingen fokuserer på generelle modeller og metoder, der kan anvendes i mange forskellige videnskabelige sammenhænge, fremfor at analysere specifikke data.

Indledningsvis gennemgås nogle grundlæggende dele af den statistiske teori, der omhandler rumlige data. Afhandlingen beskæftiger sig med to forskellige typer rumlige data, hvor de *rumlige punktmønstre*, bestående af de stokastiske placeringer af en rumlig begivenhed, bliver analyseret ved hjælp af rumlige punktprocesmodeller, og de *rumlige gitter-data*, bestående af målingerne af en stokastisk størrelse på faste rumlige placeringer, bliver analyseret ved hjælp af modeller for såkaldte random fields.

Den grundlæggende punktprocesmodel er Poisson-punktprocessen, som er en model, der tillader rumlig variation i antallet af punkter, men som ikke tillader vekselvirkninger mellem punkterne. På baggrund af en litteraturgennemgang beskrives den såkaldte score-test, som bruges til at undersøge, om et punktmønster har en signifikant rumlig variation, der kan forklares ved hjælp af en rumlig kovariat. Dette har tidligere været baseret på, at punkterne antages at stamme fra en Poisson-punktproces, men dette er i mange tilfælde en urealistisk og for simpel model. I afhandlingen beskrives det, hvordan score-testet kan anvendes under mere generelle betingelser, hvor punktprocesmodellen inkluderer vekselvirkninger mellem punkterne. I klassisk statistik bliver score-testet ofte brugt til modelselektion og modelkontrol. I punktproceslitteraturen bliver modelkontrol ofte udført ved at sammenligne en funktionel statistisk observator for data med dens forventede værdi under modellen. Afhandlingen giver teoretisk støtte til denne fremgangsmåde, ved at vise hvordan det svarer til et score-test for modellens goodness-of-fit. Yderligere redskaber til undersøgelse af goodness-of-fit bliver udledt med udgangspunkt i score-testet og approksimationer til score-testet relateret til punktproces-residualerne, der for nyligt er beskrevet i litteraturen.

Gitter-data behandlet i denne afhandling består af ikke-negative heltal, også kaldt tælledata. Dette er et typisk eksempel på gitter-data, som f.eks. forekommer, når data svarer til en optælling af en given hændelse på fastsatte rumlige placeringer. Dette kunne være antallet af sygdomstilfælde inden for bestemte geografiske regioner. Sådanne data kan analyseres ved hjælp af de α-permanentale random fields beskrevet i afhandlingen. De udgør en fleksibel klasse af modeller for gitter-data med positive associationer mellem gitterplaceringerne, hvilket betyder at data fra nærtliggende placeringer har en tendens til at ligne hinanden. De sandsynlighedsteoretiske aspekter af α-permanente random fields gennemgås, og nogle nye resultater bliver udledt. Specielt udledes detaljerne for, hvordan simulationer fra et α-permanentalt random field kan genereres ved hjælp af en såkaldt Poisson-randomisering. Selvom likelihooden for et α-permanentalt random field kan udtrykkes på lukket form, kræver det normalt store mængder computer-beregninger at evaluere den. For at kompensere for dette problem bliver det vist, hvordan statistisk inferens enten kan baseres på approksimativ beregning af likelihooden eller

på beregningsmæssigt enklere størrelser som quasi-likelihood eller composite likelihood.

Afslutningsvis præsenteres to igangværende projekter. Først introduceres en klasse af punktprocesmodeller, som kan bruges til at beskrive placeringen og størrelsen af træer, og det beskrives, hvordan likelihood-baseret inferens kan håndteres for disse modeller. Dernæst gives en kort gennemgang af såkaldte determinante punktprocesser, og det diskuteres, hvordan bayesiansk inferens kan udføres.

Acknowledgements

First and foremost I wish to thank my supervisor Jesper Møller for sharing a vast knowledge of spatial statistics in general and spatial point processes in particular. Thank you for many fruitful discussions and ideas and for lending me support throughout the last three years.

I would like to thank Peter McCullagh and Adrian Baddeley for hosting my long term stays in Chicago and Perth. It has been a privilege to get to know you both personally and professionally, and our many discussions have been an invaluable part of much of the work in this thesis. Also thank you to the financial supporters of my visits: Augustinus Fonden, Cand. theol. Nicolai Michelsen og Valborg Michelsen, født Hjort' Fond, Niels Bohr Fond, and Oticon Fonden.

Thanks to my colleagues at Aalborg University for providing a pleasant and inspiring working environment. Thanks to Svante, Rasmus, Kasper and Jakob for informal but very useful discussions, and especially thanks to my good friend and office mate Torben. Our chats about programming, mathematics and statistics have been very useful, and I have particularly enjoyed our conversations about life outside the "PhD world" accompanied by an occasional beer.

My family and friends are thanked for their continuous support and for keeping me in their lives despite the long stays abroad. A special thanks goes to all of the visitors during the time in Chicago and Perth for prioritizing such a distant visit.

Finally, to my wife Julie: Thank you for moving to the other side of the world with me and for being there through both good and bad times. Thank you for all your love, friendship and continuous support, especially during these last months, where my absentmindedness has reached new heights. I could not imagine doing this without you, and I look forward to *dive* into our future adventures.

Aalborg, August 2010 Ege Rubak

A few errors have been corrected in this revised version.

Aalborg, November 2010 Ege Rubak

Contents

CHAPTER 1

Introduction

Spatial data are collected in a wide range of scientific areas as e.g. agriculture, astronomy, biology, epidemiology, and physics. Naturally, this gives rise to a great interest in the development of statistical models and methodology for this type of data. A thorough account of statistics for spatial data can be found in Cressie (1993) whereas the following only serves as a short introduction to the area of spatial statistics and some of its terminology. The introduction is by no means exhaustive and it should not be considered a summary of the most important topics in spatial statistics. Rather, the material is selected to allow a concise motivation of the subsequent development of models and methodology.

This thesis studies two important types of spatial data, which require separate methods of analysis: *point pattern data* consisting of the random locations of some event of interest, and *lattice data* consisting of measurements of some random quantity at fixed locations. The probabilistic and statistical frameworks treating these data types are respectively *spatial point processes* and *random fields on lattices*, as detailed further in Sections 1.1 and 1.2 below.

1.1 Spatial point processes

This thesis takes a non-technical approach to the theory of spatial point processes, and readers interested in the full mathematical generality and measure theoretical details are referred to Møller and Waagepetersen (2004) and Daley and Vere-Jones (2003, 2008), where most of the material presented below is discussed in further detail.

A spatial point pattern is a finite set $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ of points $x_i \in W$, where the observation window $W \subset \mathbb{R}^d$ has finite positive volume $|W|$. In all that follows $d = 2$ is assumed, but generalization to other d is possible. The notion of an observation window attached to the set of observed points is crucial, since the absence of points also conveys information about the process generating the point pattern.

In this non-technical approach a point process $\boldsymbol{X}$ in W is simply considered to be a random finite subset of W and the point pattern $\boldsymbol{x}$ is considered a realization of $\boldsymbol{X}$. More generally, a point process $\boldsymbol{Y}$ in a possibly unbounded space $S \subseteq \mathbb{R}^2$ is a locally finite random subset of S. That is, for any bounded set $A \subseteq S$ the restriction of $\boldsymbol{Y}$ to A denoted $\boldsymbol{Y}_A = \boldsymbol{Y} \cap A$ has finite cardinality $n(\boldsymbol{Y}_A) = |\boldsymbol{Y}_A|$. In some cases it is natural to assume that the point process $\boldsymbol{X}$ in W is the restriction $\boldsymbol{X} = \boldsymbol{Y} \cap W$ of a point process $\boldsymbol{Y}$ in a larger space $S \supset W$. The most typical example of this is that $\boldsymbol{Y}$ is a stationary process in $\mathbb{R}^2$, meaning that the distribution of $\boldsymbol{Y}$ is invariant under translation.

The most common first order characteristic of a point process $\boldsymbol{Y}$ in S is the intensity function, which is a non-negative function $\rho : S \to [0, \infty)$ that determines the expected number of points of $\boldsymbol{Y}_A$ for any bounded $A \subseteq S$ by

$$\mathbb{E}[n(\boldsymbol{Y}_A)] = \int_A \rho(u)\, \mathrm{d}u. \tag{1.1}$$

A point process with constant intensity $\rho(x) \equiv \rho$ is called homogeneous (or first order homogeneous).

A particularly important spatial point process model is the Poisson process, which is fully characterized by the intensity function, and Poisson(S, ρ) denotes a Poisson process in S with intensity function ρ. An important characterization is that if $\boldsymbol{Y} \sim$ Poisson(S, ρ) then $n(\boldsymbol{Y}_A)$ is Poisson distributed with mean given by (1.1) for any bounded $A \subseteq S$, and conditional on $n(\boldsymbol{Y}_A) = n$ the points are independently distributed on A with density proportional to ρ. Furthermore, $n(\boldsymbol{Y}_A)$ and $n(\boldsymbol{Y}_B)$ are independent for disjoint subsets $A, B \subseteq S$. A first order homogeneous Poisson process is sometimes referred to as complete spatial randomness (CSR) due to the aforementioned properties.

It is often the case that a point pattern $\boldsymbol{x}$ in W is not well modeled by a Poisson process on W, and more complicated spatial point process models allowing for interactions between the points are needed. A useful way of specifying a new model is by specifying a density f for $\boldsymbol{X}$ with respect to Poisson$(W, 1)$. The density is characterized by the property

$$\mathbb{E}[h(\boldsymbol{X})] = \mathbb{E}[h(\boldsymbol{Y}) f(\boldsymbol{Y})]$$

for all non-negative measurable functionals h, where $\boldsymbol{Y} \sim$ Poisson$(W, 1)$. Usually the density is only specified to be proportional to some non-negative integrable function, and evaluation of the normalizing constant is typically very difficult. In all that follows it is assumed that the density is hereditary such that

$$f(\boldsymbol{x}) > 0 \Rightarrow f(\boldsymbol{x}') > 0 \quad \text{when } \boldsymbol{x}' \subset \boldsymbol{x}.$$

The density can be used to introduce the Papangelou conditional intensity for $\boldsymbol{x} \subset W$ and $u \in W \setminus \boldsymbol{x}$

$$\lambda(u, \boldsymbol{x}) = \begin{cases} 0 & \text{when } f(\boldsymbol{x}) = 0 \\ f(\boldsymbol{x} \cup \{u\})/f(\boldsymbol{x}) & \text{otherwise.} \end{cases}$$

A point process is called attractive or clustered if

$$\lambda(u, \boldsymbol{x}') \leq \lambda(u, \boldsymbol{x}) \quad \text{when } \boldsymbol{x}' \subset \boldsymbol{x}$$

and regular, repulsive or inhibited if

$$\lambda(u, \boldsymbol{x}') \geq \lambda(u, \boldsymbol{x}) \quad \text{when } \boldsymbol{x}' \subset \boldsymbol{x}.$$

When a parametric point process model is proposed for a given dataset the first step of an analysis is usually to estimate the parameters of the model. Both Bayesian and likelihood based inference can in many cases be intractable analytically due to unknown normalizing constants. Therefore, computer intensive Markov Chain Monte Carlo (MCMC) methods are typically needed to estimate parameters in point process models. This fact severely hampers routine estimation of parameters by non-experts, since programming skills, theoretical knowledge and practical experience with MCMC methods are often needed. A less efficient approach than maximum likelihood estimation is to base the estimation on the pseudo-likelihood (Besag, 1978):

$$\mathsf{PL}(\theta) = \left[\prod_i \lambda_\theta(x_i, \boldsymbol{x})\right] \exp\left(-\int_W \lambda_\theta(u, \boldsymbol{x})\, \mathrm{d}u\right),$$

where θ is the vector of model parameters and $\lambda_\theta(u, \boldsymbol{x})$ is the Papangelou conditional intensity. The maximum pseudo-likelihood estimate (MPLE) can be found for a wide range of point process models using the `spatstat` package (Baddeley and Turner, 2005) for the statistical software R (R Development Core Team, 2009).

Once a point process model has been fitted to a dataset using Bayesian, likelihood, pseudo-likelihood or some other type of inference, it is important to be able to asses the goodness-of-fit of the model. A common procedure is to check that certain summaries of the data agree with what could be expected under the fitted model. A very simple example is that the observed number of points in the dataset should not be an extreme observation under the fitted model, but more generally functional summary statistics are used to capture certain characteristics of a point pattern. For example, given the point patten $\boldsymbol{x}$ in the observation window W, one may consider a functional summary statistic based on a weighted sum of inter-point distances

$$\hat{K}(r) = \sum_{i \neq j} w(x_i, x_j) \mathbb{I}\{\|x_i - x_j\| \leq r\},$$

where the weight for a specific pair of points $w(x_i, x_j)$ is allowed to depend on the total number of points n and the area of the observation window $|W|$. For appropriately chosen weights, $\hat{K}$ provides an unbiased estimate of Ripley's K-function (Ripley, 1976, 1977) for a stationary point process with intensity ρ. Informally, the K-function is defined such that the expected number of additional points of the process falling within distance r of a typical point of the

process is $\rho K(r)$.To asses the goodness-of-fit of a model using a functional summary statistic, an empirical estimate such as $\hat{K}$ is compared with its expected value under the model. However, apart from the case of a Poisson point process, evaluation of the expected value of most functional summary statistics is analytically intractable, and simulation methods are needed, which can be both time consuming and complex. In Chapter 2 it is suggested to use a computationally simple unbiased estimate of the mean value under the fitted model for informal model validation. The proposed method is closely related to the residuals of Baddeley et al. (2005), and it does not depend on the assumption of stationarity making it valid also for inhomogeneous models. It is also discussed how goodness-of-fit diagnostics based on functional summary statistics can be formally interpreted as a score test. Furthermore, Chapter 2 discusses the traditional use of score tests for Poisson point process models and shows how the score test is applicable to a wider class of models.

1.2 Random fields

Generally, a random field is a collection of random variables $\boldsymbol{X} = \{X_i : i \in S\}$, where the index set S is allowed to be uncountable. However, in this thesis we only consider random fields where S is a finite spatial lattice, meaning that S indexes a finite collection of spatial sites with associated neighbourhood information. The neighbourhood information may be a detailed geographical map giving the location of each site, or it may simply be a mathematical graph with a set of vertices representing the sites and a set of edges specifying which sites that are neighbours. This section only gives a short introduction to lattice data and random fields on lattices, and readers interested in a more detailed account are referred to Cressie (1993), where most of the material presented below can be found.

An example of lattice data is a digital image where each pixel may be taken as a spatial site, and the neighbourhood structure is given by a graph with edges between neighbouring pixels in the image. In this case the lattice is said to be regular.

Data on an irregular lattice are also common and often they appear as the result of spatially aggregated data. The areas of aggregation typically correspond to administrative regions for which the aggregated data are known, but where the spatial distribution of data within the region is unknown. In particular such data may originate from aggregation of a point pattern, which yields an aggregated dataset consisting of counts of events within each administrative region. In Chapter 4 this situation is exemplified by a dataset of aggregated counts of disease cases in 19 Danish municipalities. Ideally, a model for this data should be derived from a model for the underlying point process. Assuming an underlying Poisson point process model it is straightforward to derive the model for the aggregated counts. Due to the properties of the Poisson process detailed in Section 1.1, the counts will be independent and Poisson distributed. However, for non-Poisson models the distribution of the counts is typically intractable. An alternative approach is to ignore the underlying point process and directly specify a model for the lattice data, which is the approach considered in Chapter 4, and some further remarks on the aggregation problem are given therein (see also Richardson, 2003; Møller, 2003).

A common way of specifying a random field model is via the individual conditional distri-

butions $P_i(X_i \mid X_{-i})$, $i \in S$, where $X_{-i} = \{X_j : j \in S \setminus \{i\}\}$. For $i \in S$, let $N_i \subseteq S \setminus \{i\}$, be the neighbours of the i'th site based on the neighbourhood information associated with S, and let $X_{N_i} = \{X_j : j \in N_i\}$. Then, from a modeling viewpoint, it may be natural to assume a Markov property such that the conditional distribution of X_i only depends on X_{N_i}, i.e.

$$P_i(X_i \mid X_{-i}) = P_i(X_i \mid X_{N_i}). \tag{1.2}$$

It is emphasized that the set of neighbours N_i is not predetermined by the neighbourhood information attached to S, but rather it is a modeling choice based on this information. For example, if there is a map associated with S, a natural choice would be to define N_i as the sites within a certain distance of site i. Alternatively, if there is only a graph associated with S it would be natural to let N_i be the sites connected to i by no more than d edges, where d is another modeling choice. An important requirement of the choice of N_i is that it should be symmetric in the sense that if j is a neighbour of i, $j \in N_i$ then i must also be a neighbour of j, $i \in N_j$. A random field satisfying (1.2) above is called a Markov random field. This specification of a probability model in terms of individual conditional distributions is called a local characterization of the model, and care must be taken to ensure the existence of a joint distribution of $\boldsymbol{X}$ which is consistent with the specified conditional distributions. The admissible functional forms of the P_i's are completely characterized by the Hammersley-Clifford theorem (see e.g. Besag, 1974), which makes it possible to verify that a given local specification corresponds to a well-defined distribution. While the Hammersley-Clifford theorem details how a consistent joint distribution is obtained from the conditional distributions it turns out that the joint distribution involves an unknown normalizing constant which is usually not available on closed form. Since the likelihood of a Markov random field model depends on the unknown normalizing constant, likelihood inference is often intractable for these models, and alternative methods must be used.

A well-known example of a locally specified model is the auto-Poisson model which assumes that the conditional distribution of $X_i \mid X_{N_i}$ is Poisson with mean

$$\mu_i = \exp(\alpha_i + \sum_{j \in N_i} \beta_{ij} x_j),$$

where α_i and β_{ij} are parameters (Besag, 1974). A major drawback of the auto-Poisson model is that it is only well-defined for $\beta_{ij} \leq 0$. This implies that only non-positive associations between sites are allowed.

The discussion above clarifies that the locally specified auto-Poisson model cannot be applied to count data with positive associations between sites, which is a commonly expected feature in many applications. An entirely different approach to modeling is given by directly specifying the joint distribution of $\boldsymbol{X}$. In this approach, the modeling is still based on neighbourhood information. However, rather than specifying a conditional distribution one may for example model the joint mean and covariance structure of the random field taking the neighbourhood information into account. Chapter 3 provides a review of and some new results for a class of multivariate negative binomial distributions called α-permanental random fields. These are flexible models for count data with positive associations between sites, which are specified

directly in terms of the joint distribution of the counts. In Chapter 4 various approaches to statistical inference for these models are considered and exemplified.

CHAPTER 2

Score, pseudo-score and residual diagnostics for goodness-of-fit of spatial point process models

Publication details

Co-authors: Adrian Baddeley[‡], Jesper Møller[†]

[‡] *CSIRO*
[†] *Department of Mathematical Sciences, Aalborg University*

Journal: Statistical Science (submitted)

Abstract:
We develop new tools for formal inference and informal model validation in the analysis of spatial point pattern data. The score test is generalised to a 'pseudo-score' test derived from Besag's pseudo-likelihood, and to a class of diagnostics based on point process residuals. The results lend theoretical support to the established practice of using functional summary statistics such as Ripley's K-function, when testing for complete spatial randomness; and they provide new tools such as the *compensator* of the K-function for testing other fitted models. The results also support localisation methods such as the scan statistic and smoothed residual plots. Software for computing the diagnostics is provided.

Keywords:
compensator; functional summary statistics; model validation; point process residuals; pseudo-likelihood.

2.1 Introduction

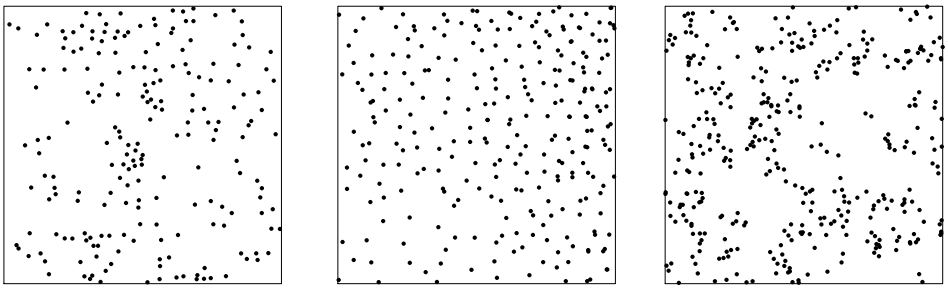
This paper develops new tools for formal inference and informal model validation in the analysis of spatial point pattern data. The score test statistic, based on the point process likelihood, is generalised to a 'pseudo-score' test statistic derived from Besag's pseudo-likelihood. The score and pseudo-score can be viewed as residuals, and further generalised to a class of residual diagnostics.

The likelihood score and the score test (Wald, 1941; Rao, 1948; Cox and Hinkley, 1974, pp 315 and 324) are used frequently in applied statistics to provide diagnostics for model selection and model validation (Atkinson, 1982; Cook and Weisberg, 1983; Pregibon, 1982; Chen, 1983; Wang, 1985). In spatial statistics, the score test is used mainly to support formal inference about covariate effects (Berman, 1986; Lawson, 1993; Waller et al., 1992) assuming the underlying point process is Poisson under both the null and alternative hypotheses. Our approach extends this to a much wider class of point processes, making it possible (for example) to check for covariate effects or localised hot-spots in a clustered point pattern.

Figure 2.1 shows three example datasets studied in the paper. Our techniques make it possible to check separately for 'inhomogeneity' (spatial variation in abundance of points) and 'interaction' (localised dependence between points) in these data.

Our approach also provides theoretical support for the established practice of using functional summary statistics such as Ripley's K-function (Ripley, 1976, 1977) to study clustering and inhibition between points. In one class of models, the score test statistic is equivalent to the empirical K-function, and the score test procedure is closely related to the customary goodness-of-fit procedure based on comparing the empirical K-function with its null expected value. Similar statements apply to the nearest neighbour distance distribution function G and the empty space function F.

For computational efficiency, especially in large datasets, the point process likelihood is often replaced by Besag's pseudo-likelihood (Besag, 1978). The resulting 'pseudo-score' is a possible surrogate for the likelihood score. In one model, the pseudo-score test statistic is



(a) (b) (c)

Figure 2.1: Point pattern datasets. (a) Japanese black pine seedlings and saplings in a 10×10 metre quadrat (Numata, 1961, 1964). Reprinted by kind permission of Professors M. Numata and Y. Ogata. (b) Simulated realisation of inhomogeneous Strauss process showing strong inhibition and spatial trend (Baddeley et al., 2005, Fig. 4b). (c) Simulated realisation of homogeneous Geyer saturation process showing moderately strong clustering without spatial trend (Baddeley et al., 2005, Fig. 4c).

equivalent to a *residual* version of the empirical K-function, yielding a new, efficient diagnostic for goodness-of-fit. However, in general, the interpretation of the pseudo-score test statistic is conceptually more complicated than that of the likelihood score test statistic, and hence difficult to employ as a diagnostic.

In classical settings the score test statistic is a weighted sum of residuals. Here the pseudo-score test statistic is a weighted point process residual in the sense of Baddeley et al. (2005, 2008). This suggests a simplification, in which the pseudo-score test statistic is replaced by another residual diagnostic that is easier to interpret and to compute.

In special cases this diagnostic is a residual version of one of the classical functional summary statistics K, G or F obtained by subtracting a *'compensator'* from the functional summary statistic. The compensator depends on the observed data and on the fitted model. For example, if the fitted model is the homogeneous Poisson process, then the compensator of $G(r)$ is $F(r)$, and the compensator of $K(r)$ is πr^2. This approach provides a new class of residual summary statistics that can be used as informal diagnostics for goodness-of-fit, for a wide range of point process models, in close analogy with current practice. The diagnostics apply under very general conditions, including the case of inhomogeneous point process models, where exploratory methods are underdeveloped or inapplicable. For instance, the compensator of $K(r)$ for an inhomogeneous non-Poisson model is illustrated in Figure 2.2.

Section 2.2 introduces basic definitions and assumptions. Section 2.3 describes the score test for a general point process model, and Section 2.4 develops the important case of Poisson point process models. Section 2.5 gives examples and technical tools for non-Poisson point process models. Section 2.6 develops the general theory for our diagnostic tools. Section 2.7 applies these tools to tests for first order trend and hotspots. Sections 2.8–2.11 develop diagnostics for interaction between points, based on pairwise distances, nearest neighbour distances and

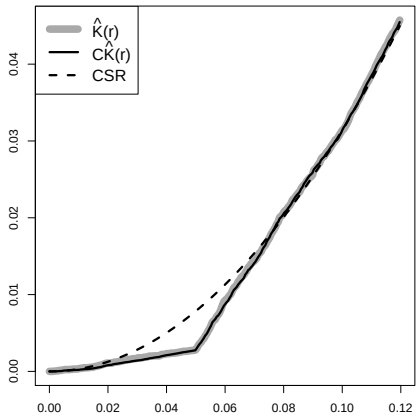


Figure 2.2: Empirical K-function (thick grey line) for the point pattern data in Figure 2.1b, compensator of the K-function (solid black line) for a model of the correct form, and expected K-function for a homogeneous Poisson process (dashed line).

empty space distances respectively. The tools are demonstrated on data in Sections 2.12–2.15. Further examples of diagnostics are given in Appendix 2.A. Appendices 2.B–2.E provide technical details.

2.2 Assumptions

2.2.1 Fundamentals

A spatial point pattern dataset is a finite set $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ of points $x_i \in W$, where the number of points $n(\boldsymbol{x}) = n \geq 0$ is not fixed in advance, and the domain of observation $W \subset \mathbb{R}^d$ is a fixed, known region of d-dimensional space with finite positive volume $|W|$. We take $d = 2$ but the results generalise easily to all dimensions.

A point process model assumes that $\boldsymbol{x}$ is a realisation of a finite point process $\boldsymbol{X}$ in W without multiple points. We can equivalently view $\boldsymbol{X}$ as a random finite subset of W. Much of the literature on spatial statistics assumes that $\boldsymbol{X}$ is the restriction $\boldsymbol{X} = \boldsymbol{Y} \cap W$ of a stationary point process $\boldsymbol{Y}$ on the entire space $\mathbb{R}^2$. We do not assume this; there is no assumption of stationarity, and some of the models considered here are intrinsically confined to the domain W. For further background material including measure theoretical details, see e.g. Møller and Waagepetersen (2004, Appendix B).

Write $\boldsymbol{X} \sim \text{Poisson}(W, \rho)$ if $\boldsymbol{X}$ follows the Poisson process on W with intensity function ρ, where we assume $\nu = \int_W \rho(u)\, \mathrm{d}u$ is finite. Then $n(\boldsymbol{X})$ is Poisson distributed with mean ν, and conditional on $n(\boldsymbol{X})$, the points in $\boldsymbol{X}$ are i.i.d. with density $\rho(u)/\nu$.

Every point process model considered here is assumed to have a probability density with respect to Poisson(W, 1), the unit rate Poisson process, under one of the following scenarios.

2.2.2 Unconditional case

In the *unconditional case* we assume $\boldsymbol{X}$ has a density f with respect to Poisson(W, 1). Then the density is characterised by the property

$$\mathbb{E}[h(\boldsymbol{X})] = \mathbb{E}[h(\boldsymbol{Y})f(\boldsymbol{Y})] \tag{2.1}$$

for all non-negative measurable functionals h, where $\boldsymbol{Y} \sim$ Poisson(W, 1). In particular the density of Poisson(W, ρ) is

$$f(\boldsymbol{x}) = \exp\left(\int_W (1-\rho(u))\,\mathrm{d}u\right)\prod_i \rho(x_i). \tag{2.2}$$

We assume that f is hereditary, i.e. $f(\boldsymbol{x}) > 0$ implies $f(\boldsymbol{y}) > 0$ for all finite $\boldsymbol{y} \subset \boldsymbol{x} \subset W$.

2.2.3 Conditional case

In the *conditional case*, we assume $\boldsymbol{X} = \boldsymbol{Y} \cap W$ where $\boldsymbol{Y}$ is a point process. Thus $\boldsymbol{X}$ may depend on unobserved points of $\boldsymbol{Y}$ lying outside W. The density of $\boldsymbol{X}$ may be unknown or intractable. Under suitable conditions (explained in Section 2.5.4) modelling and inference can be based on the conditional distribution of $\boldsymbol{X}^\circ = \boldsymbol{X} \cap W^\circ$ given $\boldsymbol{X}^+ = \boldsymbol{X} \cap W^+ = \boldsymbol{x}^+$, where $W^+ \subset W$ is a subregion, typically a region near the boundary of W, and only the points in $W^\circ = W \setminus W^+$ are treated as random. We assume that the conditional distribution of $\boldsymbol{X}^\circ = \boldsymbol{X} \cap W^\circ$ given $\boldsymbol{X}^+ = \boldsymbol{X} \cap W^+ = \boldsymbol{x}^+$ has an hereditary density $f(\boldsymbol{x}^\circ \mid \boldsymbol{x}^+)$ with respect to Poisson(W°, 1).

For ease of exposition, we focus mainly on the unconditional case, with occasional comments on the conditional case. For Poisson point process models, we always take $W = W^\circ$ so that the two cases agree.

2.3 Score test for point processes

In principle, any technique for likelihood-based inference is applicable to point process likelihoods. In practice, many likelihood-based computations require extensive Monte Carlo simulation (Geyer, 1999; Møller and Waagepetersen, 2004, 2007). To minimise such difficulties, when assessing the goodness-of-fit of a fitted point process model, it is natural to choose the score test which only requires computations for the null hypothesis (Wald, 1941; Rao, 1948).

Consider any parametric family of point process models for $\boldsymbol{X}$ with density f_θ indexed by a k-dimensional vector parameter $\theta \in \Theta \subseteq \mathbb{R}^k$. For a *simple* null hypothesis $H_0 : \theta = \theta_0$ where

$\theta_0 \in \Theta$ is fixed, the score test against any alternative $H_1 : \theta \in \Theta_1$, where $\Theta_1 \subseteq \Theta \setminus \{\theta_0\}$, is based on the score test statistic (Cox and Hinkley, 1974, p. 315)

$$T^2 = U(\theta_0)^{\mathrm{T}} I(\theta_0)^{-1} U(\theta_0). \tag{2.3}$$

Here $U(\theta) = \frac{\partial}{\partial\theta} \log f_\theta(\boldsymbol{x})$ and $I(\theta) = \mathbb{E}_\theta\left[U(\theta)U(\theta)^{\mathrm{T}}\right]$ are the score function and Fisher information respectively, and the expectation is with respect to f_θ. Here and throughout, we assume that the order of integration and differentation *with respect to* θ can be interchanged. Under suitable conditions, the null distribution of T^2 is χ^2 with k degrees of freedom. In the case $k = 1$ it may be informative to evaluate the signed square root

$$T = U(\theta_0)/\sqrt{I(\theta_0)} \tag{2.4}$$

which is asymptotically standard normally distributed under the same conditions.

For a *composite* null hypothesis $H_0 : \theta \in \Theta_0$ where $\Theta_0 \subset \Theta$ is an m-dimensional submanifold with $0 < m < k$, the score test statistic is defined in Cox and Hinkley (1974, p. 324). However, we shall not use this version of the score test, as it assumes differentiability of the likelihood with respect to nuisance parameters, which is not necessarily applicable here (as exemplified in Section 2.4.2).

In the sequel we often consider models of the form

$$f_{(\alpha,\beta)}(\boldsymbol{x}) = c(\alpha,\beta)h_\alpha(\boldsymbol{x})\exp(\beta S(\boldsymbol{x})) \tag{2.5}$$

where the parameter β and the statistic $S(\boldsymbol{x})$ are one dimensional, and the null hypothesis is $H_0 : \beta = 0$. For fixed α, this is a linear exponential family and (2.4) becomes

$$T(\alpha) = (S(\boldsymbol{x}) - \mathbb{E}_{(\alpha,0)}[S(\boldsymbol{x})]) / \sqrt{\mathbb{V}\mathrm{ar}_{(\alpha,0)}[S(\boldsymbol{x})]}.$$

In practice, when α is unknown, we replace α by its MLE under H_0 so that, with a slight abuse of notation, the signed square root of the score test statistic is approximated by

$$T = T(\hat{\alpha}) = (S(\boldsymbol{x}) - \mathbb{E}_{(\hat{\alpha},0)}[S(\boldsymbol{x})]) / \sqrt{\mathbb{V}\mathrm{ar}_{(\hat{\alpha},0)}[S(\boldsymbol{x})]}. \tag{2.6}$$

Under suitable conditions, T in (2.6) is asymptotically equivalent to T in (2.4), and so a standard Normal approximation may still apply.

2.4 Score test for Poisson processes

Application of the score test to Poisson point process models appears to originate with Cox (1972). Consider a parametric family of Poisson processes, Poisson(W, ρ_θ), where the intensity function is indexed by $\theta \in \Theta$. The score test statistic is (2.3) where

$$\begin{aligned} U(\theta) &= \sum_i \kappa_\theta(x_i) - \int_W \kappa_\theta(u)\rho_\theta(u)\,\mathrm{d}u \\ I(\theta) &= \int_W \kappa_\theta(u)\,\kappa_\theta(u)^{\mathrm{T}}\rho_\theta(u)\,\mathrm{d}u \end{aligned}$$

with $\kappa_\theta(u) = \frac{\partial}{\partial\theta}\log\rho_\theta(u)$. Asymptotic results are given in Kutoyants (1998); Rathbun and Cressie (1994).

2.4.1 Log-linear alternative

The score test is commonly used in spatial epidemiology to assess whether disease incidence depends on environmental exposure. As a particular case of (2.5), suppose the Poisson model has a log-linear intensity function

$$\rho_{(\alpha,\beta)}(u) = \exp(\alpha + \beta Z(u)) \tag{2.7}$$

where $Z(u), u \in W$ is a known, real-valued and non-constant covariate function, and α and β are real parameters. Cox (1972) noted that the uniformly most powerful test of $H_0 : \beta = 0$ (the homogeneous Poisson process) against $H_1 : \beta > 0$ is based on the statistic

$$S(\boldsymbol{x}) = \sum_i Z(x_i). \tag{2.8}$$

Recall that, for a point process $\boldsymbol{X}$ on W with intensity function ρ,

$$\mathbb{E}\left(\sum_{x_i \in \boldsymbol{X}} h(x_i)\right) = \int_W h(u)\rho(u)\,\mathrm{d}u \tag{2.9}$$

for any Borel function h such that the integral on the right hand side exists, and for Poisson(W, ρ),

$$\mathbb{V}\mathrm{ar}\left(\sum_{x_i \in \boldsymbol{X}} h(x_i)\right) = \int_W h(u)^2\rho(u)\,\mathrm{d}u \tag{2.10}$$

for any Borel function h such that the integral on the right hand side exists (Daley and Vere-Jones, 1988, p. 188). Hence the standardised version of (2.8) is

$$T = \left(S(\boldsymbol{x}) - \hat{\kappa}\int_W Z(u)\,\mathrm{d}u\right) \Big/ \sqrt{\hat{\kappa}\int_W Z(u)^2\,\mathrm{d}u} \tag{2.11}$$

where $\hat{\kappa} = n/|W|$ is the MLE of the intensity $\kappa = \exp(\alpha)$ under the null hypothesis. This is a direct application of the approximation (2.6) of the signed square root of the score test statistic.

Berman (1986) proposed several tests and diagnostics for spatial association between a point process $\boldsymbol{X}$ and a covariate function $Z(u)$. Berman's Z_1 test is equivalent to the Cox score test described above. Waller et al. (1992) and Lawson (1993) proposed tests for the dependence of disease incidence on environmental exposure, based on data giving point locations of disease cases. These are also applications of the score test. Berman conditioned on the number of points when making inference. This is in accordance with the observation that the statistic $n(\boldsymbol{x})$ is S-ancillary for β, while $S(\boldsymbol{x})$ is S-sufficient for β.

2.4.2 Threshold alternative and nuisance parameters

Consider the Poisson process with an intensity function of 'threshold' form,

$$\rho_{z,\kappa,\phi}(u) = \begin{cases} \kappa\exp(\phi) & \text{if } Z(u) \le z \\ \kappa & \text{if } Z(u) > z \end{cases}$$

where z is the threshold level. If z is fixed, this model is a special case of (2.7) with $Z(u)$ replaced by $\mathbb{I}\{Z(u) \le z\}$, and so (2.8) is replaced by

$$S(\boldsymbol{x}) = S(\boldsymbol{x}, z) = \sum_i \mathbb{I}\{Z(x_i) \le z\}$$

where $\mathbb{I}\{\cdot\}$ denotes the indicator function. By (2.11) the (approximate) score test of $H_0 : \phi = 0$ against $H_1 : \phi \neq 0$ is based on

$$T = T(z) = (S(\boldsymbol{x}, z) - \hat{\kappa} A(z)) / \sqrt{\hat{\kappa} A(z)}$$

where $A(z) = |\{u \in W : Z(u) \le z\}|$ is the area of the corresponding level set of Z.

If z is not fixed, then it plays the role of a nuisance parameter in the score test: the value of z affects inference about the canonical parameter ϕ, which is the parameter of primary interest in the score test. Note that the likelihood is not differentiable with respect to z.

In most applications of the score test, a nuisance parameter would be replaced by its MLE under the null hypothesis. However in this context, z is not identifiable under the null hypothesis. Several approaches to this problem have been proposed, including: replacing z by its MLE under the alternative (Conniffe, 2001), maximising $T(z)$ or $|T(z)|$ over z (Davies, 1977, 1987), and finding the maximum p-value of $T(z)$ or $|T(z)|$ over a confidence region for z under the alternative (Silvapulle, 1996).

These approaches appear to be inapplicable to the current context. While the null distribution of $T(z)$ is asymptotically $N(0, 1)$ for each fixed z as $\kappa \to \infty$, this convergence is not uniform in z. The null distribution of $S(\boldsymbol{x}, z)$ is Poisson with parameter $\kappa A(z)$; sample paths of $T(z)$ will be governed by Poisson behaviour where $A(z)$ is small.

In this paper, our approach is simply to plot the score test statistic as a function of the nuisance parameter. This turns the score test into a graphical exploratory tool, following the approach adopted in many other areas (Atkinson, 1982; Cook and Weisberg, 1983; Pregibon, 1982; Chen, 1983; Wang, 1985). A second style of plot based on $S(x, z) - \hat{\kappa} A(z)$ against z may be more appropriate visually. Such a plot is the lurking variable plot of Baddeley et al. (2005). Berman (1986) also proposed a plot of $S(\boldsymbol{x}, z)$ against z, together with a plot of $\hat{\kappa} A(z)$ against z, as a diagnostic for dependence on Z. This is related to the Kolmogorov-Smirnov test since, under H_0, the values $Y_i = Z(x_i)$ are i.i.d. with distribution function $\mathbb{P}(Y \le y) = A(y)/|W|$.

2.4.3 Hot spot alternative

Consider the Poisson process with intensity

$$\rho_{\kappa,\phi,v}(u) = \kappa \exp(\phi k(u - v)) \tag{2.12}$$

where k is a kernel (a probability density on $\mathbb{R}^2$), $\kappa > 0$ and ϕ are real parameters, and $v \in \mathbb{R}^2$ is a nuisance parameter. This process has a 'hot spot' of elevated intensity in the vicinity of

the location v. By (2.11) and (2.9)–(2.10) the score test of $H_0 : \phi = 0$ against $H_1 : \phi \neq 0$ is based on

$$T = T(v) = (S(\boldsymbol{x}, v) - \hat{\kappa} M_1(v))/\sqrt{\hat{\kappa} M_2(v)}$$

where

$$S(\boldsymbol{x}, v) = \sum_i k(x_i - v)$$

is the usual nonparametric kernel estimate of point process intensity (Diggle, 1985) evaluated at v without edge correction, and

$$M_i(v) = \int_W k(u - v)^i \, \mathrm{d}u, \quad i = 1, 2.$$

The numerator $S(\boldsymbol{x}, v) - \hat{\kappa} M_1(v)$ is the *smoothed residual field* (Baddeley et al., 2005) of the null model. In the special case where $k(u) \propto \mathbb{I}\{\|u\| \leq h\}$ is the uniform density on a disc of radius h, the maximum $\max_v T(v)$ is closely related to the *scan statistic* (Alm, 1988; Kulldorff, 1999).

2.5 Non-Poisson models

The remainder of the paper deals with the case where the alternative (and perhaps also the null) is not a Poisson process. Key examples are stated in Section 2.5.1. Non-Poisson models require additional tools including the conditional intensity (Section 2.5.2) and pseudo-likelihood (Section 2.5.3).

2.5.1 Point process models with interaction

We shall frequently consider densities of the form

$$f(\boldsymbol{x}) = c \left[\prod_i \lambda(x_i) \right] \exp(\phi V(\boldsymbol{x})) \tag{2.13}$$

where c is a normalising constant, the first order term λ is a non-negative function, ϕ is a real interaction parameter, and $V(\boldsymbol{x})$ is a real non-additive function which specifies the interaction between the points. We refer to V as the interaction potential. In general, apart from the Poisson density (2.2) corresponding to the case $\phi = 0$, the normalising constant is not expressible in closed form.

Often the definition of V can be extended to all finite point patterns in $\mathbb{R}^2$ so as to be invariant under rigid motions (translations and rotations). Then the model for $\boldsymbol{X}$ is said to be homogeneous if λ is constant on W, and inhomogeneous otherwise.

Let

$$d(u, \boldsymbol{x}) = \min_j \|u - x_j\|$$

denote the distance from a location u to its nearest neighbour in the point configuration $\boldsymbol{x}$. For $n(\boldsymbol{x}) = n \geq 1$ and $i = 1, \ldots, n$, define

$$\boldsymbol{x}_{-i} = \boldsymbol{x} \setminus \{x_i\}.$$

In many places in this paper we consider the following three motion-invariant interaction potentials $V(\boldsymbol{x}) = V(\boldsymbol{x}, r)$ depending on a parameter $r > 0$ which specifies the range of interaction. The *Strauss process* (Strauss, 1975) has interaction potential

$$V_S(\boldsymbol{x}, r) = \sum_{i<j} \mathbb{I}\{\|x_i - x_j\| \leq r\} \tag{2.14}$$

the number of r-close pairs of points in $\boldsymbol{x}$; the *Geyer saturation model* (Geyer, 1999) with saturation threshold 1 has interaction potential

$$V_G(\boldsymbol{x}, r) = \sum_i \mathbb{I}\{d(x_i, \boldsymbol{x}_{-i}) \leq r\} \tag{2.15}$$

the number of points in $\boldsymbol{x}$ whose nearest neighbour is closer than r units; and the Widom-Rowlinson penetrable sphere model (Widom and Rowlinson, 1970) or *area-interaction process* (Baddeley and van Lieshout, 1995) has interaction potential

$$V_A(\boldsymbol{x}, r) = -|W \cap \bigcup_i B(x_i, r)| \tag{2.16}$$

the negative area of W intersected with the union of balls $B(x_i, r)$ of radius r centred at the points of $\boldsymbol{x}$. Each of these densities favours spatial clustering (positive association) when $\phi > 0$ and spatial inhibition (negative association) when $\phi < 0$. The Geyer and area-interaction models are well-defined point processes for any value of ϕ (Baddeley and van Lieshout, 1995; Geyer, 1999), but the Strauss density is integrable only when $\phi \leq 0$ (Kelly and Ripley, 1976).

2.5.2 Conditional intensity

Consider a parametric model for a point process $\boldsymbol{X}$ in $\mathbb{R}^2$, with parameter $\theta \in \Theta$. Papangelou (1974) defined the *conditional intensity* of $\boldsymbol{X}$ as a non-negative stochastic process $\lambda_\theta(u, \boldsymbol{X})$ indexed by locations $u \in \mathbb{R}^2$ and characterised by the property that

$$\mathbb{E}_\theta \left[\sum_{x_i \in \boldsymbol{X}} h(x_i, \boldsymbol{X} \setminus \{x_i\}) \right] = \mathbb{E}_\theta \left[\int_{\mathbb{R}^2} h(u, \boldsymbol{X}) \lambda_\theta(u, \boldsymbol{X}) \, \mathrm{d}u \right] \tag{2.17}$$

for all measurable functions h such that the left or right hand side exists. Equation (2.17) is known as the *Georgii-Nguyen-Zessin (GNZ) formula* (Georgii, 1976; Kallenberg, 1978, 1984; Nguyen and Zessin, 1979); see also Section 6.4.1 in Møller and Waagepetersen (2004). Adapting a term from stochastic process theory, we will call the random integral on the right side of (2.17) the *(Papangelou) compensator* of the random sum on the left side.

Consider a finite point process $\boldsymbol{X}$ in W. In the unconditional case (Section 2.2.2) we assume $\boldsymbol{X}$ has density $f_\theta(\boldsymbol{x})$ which is hereditary for all $\theta \in \Theta$. We may simply define

$$\lambda_\theta(u, \boldsymbol{x}) = f_\theta(\boldsymbol{x} \cup \{u\}) / f_\theta(\boldsymbol{x}) \tag{2.18}$$

for all locations $u \in W$ and point configurations $\boldsymbol{x} \subset W$ such that $u \notin \boldsymbol{x}$. Here we take $0/0 = 0$. For $x_i \in \boldsymbol{x}$ we set $\lambda_\theta(x_i, \boldsymbol{x}) = \lambda_\theta(x_i, \boldsymbol{x}_{-i})$, and for $u \notin W$ we set $\lambda_\theta(u, \boldsymbol{x}) = 0$. Then it may be verified directly from (2.1) that (2.17) holds, so that (2.18) is the Papangelou conditional intensity of $\boldsymbol{X}$. Note that the normalising constant of f_θ cancels in (2.18). For a Poisson process, it follows from (2.2) and (2.18) that the conditional intensity is equivalent to the intensity function of the process.

In the conditional case (Section 2.2.3) we assume that the conditional distribution of $\boldsymbol{X}^\circ = \boldsymbol{X} \cap W^\circ$ given $\boldsymbol{X}^+ = \boldsymbol{X} \cap W^+ = \boldsymbol{x}^+$ has an hereditary density $f_\theta(\boldsymbol{x}^\circ \mid \boldsymbol{x}^+)$ with respect to Poisson$(W^\circ, 1)$, for all $\theta \in \Theta$. Then define

$$\lambda_\theta(u, \boldsymbol{x}^\circ \mid \boldsymbol{x}^+) = f_\theta(\boldsymbol{x}^\circ \cup \{u\} \mid \boldsymbol{x}^+) / f_\theta(\boldsymbol{x}^\circ \setminus \{u\} \mid \boldsymbol{x}^+) \tag{2.19}$$

if $u \in W^\circ$, and zero otherwise. It can similarly be verified that this is the Papangelou conditional intensity of the conditional distribution of $\boldsymbol{X}^\circ$ given $\boldsymbol{X}^+ = \boldsymbol{x}^+$.

It is convenient to rewrite (2.18) in the form

$$\lambda_\theta(u, \boldsymbol{x}) = \exp(\Delta_u \log f(\boldsymbol{x}))$$

where Δ is the one-point difference operator

$$\Delta_u h(\boldsymbol{x}) = h(\boldsymbol{x} \cup \{u\}) - h(\boldsymbol{x} \setminus \{u\}). \tag{2.20}$$

Note the Poincaré inequality for the Poisson process $\boldsymbol{X}$

$$\mathbb{V}\mathrm{ar}[h(\boldsymbol{X})] \le \mathbb{E} \int_W [\Delta_u h(\boldsymbol{X})]^2 \rho(u)\, \mathrm{d}u \tag{2.21}$$

holding for all measurable functionals h such that the right hand side is finite; see Last and Penrose (2010); Wu (2000).

2.5.3 Pseudo-likelihood and pseudo-score

To avoid computational problems with point process likelihoods, Besag (1978) introduced the *pseudo-likelihood* function

$$\mathsf{PL}(\theta) = \left[\prod_i \lambda_\theta(x_i, \boldsymbol{x})\right] \exp\left(-\int_W \lambda_\theta(u, \boldsymbol{x})\, \mathrm{d}u\right). \tag{2.22}$$

This is of the same functional form as the likelihood function of a Poisson process (2.2), but has the conditional intensity in place of the Poisson intensity. The corresponding *pseudo-score*

$$\mathsf{PU}(\theta) = \frac{\partial}{\partial\theta} \log \mathsf{PL}(\theta) = \sum_i \frac{\partial}{\partial\theta} \log \lambda_\theta(x_i, \boldsymbol{x}) - \int_W \frac{\partial}{\partial\theta} \lambda_\theta(u, \boldsymbol{x})\, \mathrm{d}u \tag{2.23}$$

is an unbiased estimating function (i.e. $PU(\theta)$ has zero-mean) by virtue of (2.17).

The pseudo-likelihood function can also be defined in the conditional case (Jensen and Møller, 1991). In (2.22) the product is instead over points $x_i \in \boldsymbol{x}^\circ$ and the integral is instead over W°; in (2.23) the sum is instead over points $x_i \in \boldsymbol{x}^\circ$ and the integral is instead over W°; and in both places $\boldsymbol{x} = \boldsymbol{x}^\circ \cup \boldsymbol{x}^+$. The conditional intensity $\lambda_\theta(u, \boldsymbol{x})$ must also be replaced by $\lambda_\theta(u, \boldsymbol{x}^\circ \mid \boldsymbol{x}^+)$.

2.5.4 Markov point processes

For a point process $\boldsymbol{X}$ constructed as $\boldsymbol{X} = \boldsymbol{Y} \cap W$ where $\boldsymbol{Y}$ is a point process in $\mathbb{R}^2$, the density and conditional intensity of $\boldsymbol{X}$ may not be available in simple form. Progress can be made if $\boldsymbol{Y}$ is a *Markov point process* of interaction range $R < \infty$, see Georgii (1976); Nguyen and Zessin (1979); Ripley and Kelly (1977); van Lieshout (2000); Møller and Waagepetersen (2004, Sect. 6.4.1). Briefly, this means that the conditional intensity $\lambda_\theta(u, \boldsymbol{Y})$ of $\boldsymbol{Y}$ satisfies $\lambda_\theta(u, \boldsymbol{Y}) = \lambda_\theta(u, \boldsymbol{Y} \cap B(u, R))$ where $B(u, R)$ is the ball of radius R centred at u. Define the erosion of W by distance R

$$W_{\ominus R} = \{u \in W : B(u, R) \subset W\}$$

and assume this has non-zero area. Let $B = W \setminus W_{\ominus R}$ be the border region. The process satisfies a spatial Markov property: the processes $\boldsymbol{Y} \cap W_{\ominus R}$ and $\boldsymbol{Y} \cap W^c$ are conditionally independent given $\boldsymbol{Y} \cap B$.

In this situation we shall invoke the conditional case with $W^\circ = W_{\ominus R}$ and $W^+ = W \setminus W^\circ$. The conditional distribution of $\boldsymbol{X} \cap W^\circ$ given $\boldsymbol{X} \cap W^+ = \boldsymbol{x}^+$ has Papangelou conditional intensity

$$\lambda_\theta(u, \boldsymbol{x}^\circ \mid \boldsymbol{x}^+) = \begin{cases} \lambda_\theta(u, \boldsymbol{x}^\circ \cup \boldsymbol{x}^+) & \text{if } u \in W^\circ \\ 0 & \text{otherwise.} \end{cases} \tag{2.24}$$

Thus *the unconditional and conditional versions of a Markov point process have the same Papangelou conditional intensity* at locations in W°.

For $\boldsymbol{x}^\circ = \{x_1, \ldots, x_{n^\circ}\}$, the conditional probability density becomes

$$f_\theta(\boldsymbol{x}^\circ \mid \boldsymbol{x}^+) = c_\theta(\boldsymbol{x}^+) \lambda_\theta(x_1, \boldsymbol{x}^\circ) \prod_{i=2}^{n^\circ} \lambda_\theta(x_i, \{x_1, \ldots, x_{i-1}\} \cup \boldsymbol{x}^+)$$

if $n^\circ > 0$, and $f_\theta(\emptyset \mid \boldsymbol{x}^+) = c_\theta(\boldsymbol{x}^+)$, where $\emptyset$ denotes the empty configuration, and the inverse normalising constant $c_\theta(\boldsymbol{x}^+)$ depends only on $\boldsymbol{x}^+$.

For example, instead of (2.13) we now consider

$$f(\boldsymbol{x}^\circ \mid \boldsymbol{x}^+) = c(\boldsymbol{x}^+) \left[\prod_{i=1}^{n^\circ} \lambda(x_i) \right] \exp\left(\phi V(\boldsymbol{x}^\circ \cup \boldsymbol{x}^+)\right)$$

assuming $V(\boldsymbol{y})$ is defined for all finite $\boldsymbol{y} \subset \mathbb{R}^2$ such that for any $u \in \mathbb{R}^2 \setminus \boldsymbol{y}$, $\Delta_u V(\boldsymbol{y})$ depends only on u and $\boldsymbol{y} \cap B(u, R)$. This condition is satisfied by the interaction potentials (2.14)-(2.16); note that the range of interaction is $R = r$ for the Strauss process, and $R = 2r$ for both the Geyer and the area-interaction models.

2.6 Score, pseudoscore and residual diagnostics

This section develops the general theory for our diagnostic tools.

By (2.6) in Section 2.3 it is clear that comparison of a summary statistic $S(\boldsymbol{x})$ to its predicted value $\mathbb{E}S(\boldsymbol{X})$ under a null model, is effectively equivalent to the score test under an exponential family model where $S(\boldsymbol{x})$ is the canonical sufficient statistic. Similarly, the use of a *functional* summary statistic $S(\boldsymbol{x}, z)$, depending on a function argument z, is related to the score test under an exponential family model *where z is a nuisance parameter* and $S(\boldsymbol{x}, z)$ is the canonical sufficient statistic for fixed z. In this section we construct the corresponding exponential family models, apply the score test, and propose surrogates for the score test statistic.

2.6.1 Models

Let $f_\theta(\boldsymbol{x})$ be the density of any point process $\boldsymbol{X}$ on W governed by a parameter θ. Let $S(\boldsymbol{x}, z)$ be a functional summary statistic of the point pattern dataset $\boldsymbol{x}$, with function argument z belonging to any space.

Consider the *extended model* with density

$$f_{\theta,\phi,z}(\boldsymbol{x}) = c_{\theta,\phi,z}\, f_\theta(\boldsymbol{x}) \exp(\phi S(\boldsymbol{x}, z)) \tag{2.25}$$

where ϕ is a real parameter, and $c_{\theta,\phi,z}$ is the normalising constant. The density is well-defined provided

$$M(\theta, \phi, z) = \mathbb{E}\left[f_\theta(\boldsymbol{Y}) \exp(\phi S(\boldsymbol{Y}, z))\right] < \infty$$

where $\boldsymbol{Y} \sim \text{Poisson}(W, 1)$. The extended model is constructed by 'exponential tilting' of the original model by the statistic S. By (2.6), for fixed θ and z, assuming differentiability of M with respect to ϕ in a neighbourhood of $\phi = 0$, the signed root of the score test statistic is approximated by

$$T = (S(\boldsymbol{x}, z) - \mathbb{E}_{\hat\theta}[S(\boldsymbol{X}, z)]) \,/\, \sqrt{\mathbb{V}\text{ar}_{\hat\theta}[S(\boldsymbol{X}, z)]} \tag{2.26}$$

where $\hat\theta$ is the MLE under the null model, and the expectation and variance are with respect to the null model with density $f_{\hat\theta}$.

Insight into the qualitative behaviour of the extended model (2.25) can be obtained by studying the *perturbing model*

$$g_{\phi,z}(\boldsymbol{x}) = k_{\phi,z} \exp(\phi S(\boldsymbol{x}, z)), \tag{2.27}$$

provided this is a well-defined density with respect to Poisson$(W, 1)$, where $k_{\phi,z}$ is the normalising constant. When the null hypothesis is a homogeneous Poisson process, the extended model is identical to the perturbing model, up to a change in the first order term. In general, the extended model is a qualitative hybrid between the null and perturbing models.

In this context the score test is equivalent to naive comparison of the observed and null-expected values of the functional summary statistic S. The test statistic T in (2.26) may be difficult to evaluate; typically, apart from Poisson models, the moments (particularly the variance) of S would not be available in closed form. The null distribution of T would also

typically be unknown. Hence, implementation of the score test would typically require moment approximation and simulation from the null model, which in both cases may be computationally expensive. Various approximations for the score or the score test statistic can be constructed, as discussed in the sequel.

2.6.2 Pseudo-score of extended model

The extended model (2.25) is an exponential family with respect to ϕ, having conditional intensity

$$\kappa_{\theta,\phi,z}(u, \boldsymbol{x}) = \lambda_\theta(u, \boldsymbol{x}) \exp\left(\phi \Delta_u S(\boldsymbol{x}, z)\right)$$

where $\lambda_\theta(u, \boldsymbol{x})$ is the conditional intensity of the null model. The pseudo-score function with respect to ϕ, evaluated at $\phi = 0$, is

$$\mathsf{PU}(\theta, z) = \sum_i \Delta_{x_i} S(\boldsymbol{x}, z) - \int_W \Delta_u S(\boldsymbol{x}, z) \lambda_\theta(u, \boldsymbol{x}) \, \mathrm{d}u$$

where the first term

$$\Sigma\Delta S(\boldsymbol{x}, z) = \sum_i \Delta_{x_i} S(\boldsymbol{x}, z) \tag{2.28}$$

will be called the *pseudo-sum* of S. If $\hat{\theta}$ is the maximum pseudo-likelihood estimate (MPLE) under H_0, the second term with θ replaced by $\hat{\theta}$ becomes

$$\mathrm{C}\Delta S(\boldsymbol{x}, z) = \int_W \Delta_u S(\boldsymbol{x}, z) \lambda_{\hat{\theta}}(u, \boldsymbol{x}) \, \mathrm{d}u \tag{2.29}$$

and will be called the *(estimated) pseudo-compensator* of S. We call

$$\mathrm{R}\Delta S(\boldsymbol{x}, z) = \mathsf{PU}(\hat{\theta}, z) = \Sigma\Delta S(\boldsymbol{x}, z) - \mathrm{C}\Delta S(\boldsymbol{x}, z) \tag{2.30}$$

the *pseudo-residual* since it is a weighted residual in the sense of Baddeley et al. (2005).

The pseudo-residual serves as a surrogate for the numerator in the score test statistic (2.26). For the denominator, we need the variance of the pseudo-residual. Appendix 2.B gives an exact formula (2.65) for the variance of the pseudo-score $\mathsf{PU}(\theta, z)$, which can serve as an approximation to the variance of the pseudo-residual $\mathrm{R}\Delta S(\boldsymbol{x}, z)$. The leading term in this approximation is

$$\mathrm{C}^2\Delta S(\boldsymbol{x}, z) = \int_W [\Delta_u S(\boldsymbol{x}, z)]^2 \lambda_{\hat{\theta}}(u, \boldsymbol{x}) \, \mathrm{d}u \tag{2.31}$$

which we shall call the *Poincaré pseudo-variance* because of its similarity to the Poincaré upper bound in (2.21). We propose to use the square root of (2.31) as a surrogate for the denominator in (2.26). This yields a *'standardised' pseudo-residual*

$$\mathrm{T}\Delta S(\boldsymbol{x}, z) = \mathrm{R}\Delta S(\boldsymbol{x}, z) / \sqrt{\mathrm{C}^2\Delta S(\boldsymbol{x}, z)}. \tag{2.32}$$

We emphasise that this quantity is not guaranteed to have zero mean and unit variance (even approximately) under the null hypothesis. It is a computationally efficient surrogate for the score test statistic; its null distribution must be investigated by other means.

The pseudo-sum (2.28) can be regarded as a functional summary statistic for the data in its own right. Its definition depends only on the choice of the statistic S, and it may have a meaningful interpretation as a non-parametric estimator of a property of the point process. The pseudo-compensator (2.29) might also be regarded as a functional summary statistic, but its definition involves the null model. If the null model is true we may expect the pseudo-residual to be approximately zero. Sections 2.9-2.11 and Appendix 2.A study particular instances of pseudo-residual diagnostics based on (2.28)-(2.30).

In the conditional case, the Papangelou conditional intensity $\lambda_{\hat\theta}(u, \boldsymbol{x})$ must be replaced by $\lambda_{\hat\theta}(u, \boldsymbol{x}^\circ \mid \boldsymbol{x}^+)$ given in (2.19) or (2.24). The integral in the definition of the pseudo-compensator (2.29) must be restricted to the domain W°, and the summation over data points in (2.28) must be restricted to points $x_i \in W^\circ$, i.e. to summation over points of $\boldsymbol{x}^\circ$.

2.6.3 Residuals

A simpler surrogate for the score test is available when the canonical sufficient statistic S of the perturbing model is naturally expressible as a sum of local contributions

$$S(\boldsymbol{x}, z) = \sum_i s(x_i, \boldsymbol{x}_{-i}, z). \tag{2.33}$$

Note that any statistic can be decomposed in this way unless some restriction is imposed on s; such a decomposition is not necessarily unique. We call the decomposition 'natural' if $s(u, \boldsymbol{x}, z)$ only depends on points of $\boldsymbol{x}$ that are close to u, as demonstrated in the examples in Sections 2.9, 2.10 and 2.11 and in Appendix 2.A.

Consider a null model with conditional intensity $\lambda_\theta(u, \boldsymbol{x})$. Following Baddeley et al. (2005) define the (s-weighted) innovation by

$$\mathrm{I}\, S(\boldsymbol{x}, r) = S(\boldsymbol{x}, z) - \int_W s(u, \boldsymbol{x}, z)\lambda_\theta(u, \boldsymbol{x})\, \mathrm{d}u \tag{2.34}$$

which by the GNZ formula (2.17) has mean zero under the null model. In practice we replace θ by an estimate $\hat\theta$ (e.g. the MPLE) and consider the *(s-weighted) residual*

$$\mathrm{R}\, S(\boldsymbol{x}, z) = S(\boldsymbol{x}, z) - \int_W s(u, \boldsymbol{x}, z)\lambda_{\hat\theta}(u, \boldsymbol{x})\, \mathrm{d}u. \tag{2.35}$$

The residual shares many properties of the score function and can serve as a computationally efficient surrogate for the score. The data-dependent integral

$$\mathrm{C}\, S(\boldsymbol{x}, z) = \int_W s(u, \boldsymbol{x}, z)\lambda_{\hat\theta}(u, \boldsymbol{x})\, \mathrm{d}u \tag{2.36}$$

is the *(estimated) Papangelou compensator* of S. By the general variance formula (2.64) and by analogy with (2.31) we propose to use the *Poincaré variance*

$$\mathrm{C}^2 S(\boldsymbol{x}, z) = \int_W s(u, \boldsymbol{x}, z)^2 \lambda_{\hat{\theta}}(u, \boldsymbol{x}) \, \mathrm{d}u \tag{2.37}$$

as a surrogate for the variance of $\mathrm{R}S(\boldsymbol{x}, z)$, and thereby obtain a 'standardised' residual

$$\mathrm{T}S(\boldsymbol{x}, z) = \mathrm{R}S(\boldsymbol{x}, z) / \sqrt{\mathrm{C}^2 S(\boldsymbol{x}, z)}.$$

Once again $\mathrm{T}S(\boldsymbol{x}, z)$ is not exactly standardised, and its null distribution must be investigated by other means.

In the conditional case, the integral in the definition of the compensator (2.36) must be restricted to the domain W°, and the summation over data points in (2.33) must be restricted to points $x_i \in W^\circ$, i.e. to summation over points of $\boldsymbol{x}^\circ$.

2.7 Diagnostics for first order trend

Consider any null model with density $f_\theta(\boldsymbol{x})$ and conditional intensity $\lambda_\theta(u, \boldsymbol{x})$. By analogy with Section 2.4 we consider alternatives of the form (2.25) where

$$S(\boldsymbol{x}, z) = \sum_i s(x_i, z)$$

for some function s. The perturbing model (2.27) is a Poisson process with intensity $\exp(\phi s(\cdot, z))$ where z is a nuisance parameter. The score test is a test for the presence of an (extra) first order trend. The pseudo-score and residual diagnostics are both equal to

$$\mathrm{R}S(\boldsymbol{x}, z) = \sum_i s(x_i, z) - \int_W s(u, z) \lambda_{\hat{\theta}}(u, \boldsymbol{x}) \, \mathrm{d}u. \tag{2.38}$$

This is the s-weighted residual described in Baddeley et al. (2005). The variance of (2.38) can be estimated by simulation, or approximated by the Poincaré variance (2.37).

If Z is a real-valued covariate function on W then we may take $s(u, z) = \mathbb{I}\{Z(u) \le z\}$ for $z \in \mathbb{R}$, corresponding to a threshold effect (cf. Section 2.4.2). A plot of (2.38) against z was called a *lurking variable plot* in Baddeley et al. (2005).

If $s(u, z) = k(u - z)$ for $z \in \mathbb{R}^2$, where k is a density function on $\mathbb{R}^2$, then

$$\mathrm{R}S(\boldsymbol{x}, z) = \sum_i k(x_i - z) - \int_W k(u - z) \lambda_{\hat{\theta}}(u, \boldsymbol{x}) \, \mathrm{d}u$$

which was dubbed the *smoothed residual field* in Baddeley et al. (2005). Examples of application of these techniques have been discussed extensively in Baddeley et al. (2005).

2.8 Interpoint interaction

In the remainder of the paper we concentrate on diagnostics for interpoint interaction.

2.8.1 Classical summary statistics

Following Ripley's influential paper (Ripley, 1977) it is standard practice, when investigating association or dependence between points in a spatial point pattern, to evaluate functional summary statistics such as the K-function, and to compare graphically the empirical summaries and theoretical predicted values under a suitable model, often a stationary Poisson process ('Complete Spatial Randomness', CSR) (Ripley, 1977; Cressie, 1991; Diggle, 2003).

The three most popular functional summary statistics for spatial point processes are Ripley's K-function, the nearest neighbour distance distribution function G, and the empty space function (spherical contact distance distribution function) F. Definitions of K, G and F and their estimators can be seen in Baddeley (1999); Cressie (1991); Diggle (2003); Møller and Waagepetersen (2004). Simple empirical estimators of these functions are of the form

$$\hat{K}(r) = \hat{K}_{\boldsymbol{x}}(r) = \frac{1}{\widehat{\rho^2}(\boldsymbol{x})|W|} \sum_{i \neq j} e_K(x_i, x_j) \mathbb{I}\{\|x_i - x_j\| \leq r\} \tag{2.39}$$

$$\hat{G}(r) = \hat{G}_{\boldsymbol{x}}(r) = \frac{1}{n(\boldsymbol{x})} \sum_{i} e_G(x_i, \boldsymbol{x}_{-i}, r) \mathbb{I}\{d(x_i, \boldsymbol{x}_{-i}) \leq r\} \tag{2.40}$$

$$\hat{F}(r) = \hat{F}_{\boldsymbol{x}}(r) = \frac{1}{|W|} \int_W e_F(u, r) \mathbb{I}\{d(u, \boldsymbol{x}) \leq r\} \, \mathrm{d}u \tag{2.41}$$

where $e_K(u, v)$, $e_G(u, \boldsymbol{x}, r)$ and $e_F(u, r)$ are edge correction weights, and typically $\widehat{\rho^2}(\boldsymbol{x}) = n(\boldsymbol{x})(n(\boldsymbol{x}) - 1)/|W|^2$.

2.8.2 Score test approach

The classical approach fits naturally into the scheme of Section 2.6. In order to test for dependence between points, we choose a perturbing model that exhibits dependence. Three interesting examples of perturbing models are the Strauss process, the Geyer saturation model with saturation threshold 1 and the area-interaction process, with interaction potentials $V_S(\boldsymbol{x}, r)$, $V_G(\boldsymbol{x}, r)$ and $V_A(\boldsymbol{x}, r)$ given in (2.14)-(2.16). The nuisance parameter $r \geq 0$ determines the range of interaction. Although the Strauss density is integrable only when $\phi \leq 0$, a Strauss hybrid (between f_θ and the Strauss density) may be well-defined for some $\phi > 0$ so that the extended model may support alternatives that are clustered relative to the null, as originally intended by Strauss (Strauss, 1975).

The potentials of these three models are closely related to the summary statistics $\hat{K}, \hat{G}$ and $\hat{F}$

in (2.39)–(2.41). Ignoring the edge correction weights $e(\cdot)$ we have

$$\hat{K}_{\boldsymbol{x}}(r) \approx \frac{2|W|}{n(\boldsymbol{x})\,(n(\boldsymbol{x})-1)} V_S(\boldsymbol{x}, r) \tag{2.42}$$

$$\hat{G}_{\boldsymbol{x}}(r) \approx \frac{1}{n(\boldsymbol{x})} V_G(\boldsymbol{x}, r) \tag{2.43}$$

$$\hat{F}_{\boldsymbol{x}}(r) \approx -\frac{1}{|W|} V_A(\boldsymbol{x}, r). \tag{2.44}$$

To draw the closest possible connection with the score test, instead of choosing the Strauss, Geyer or area-interaction process as the perturbing model, we shall take the perturbing model to be defined through (2.27) where S is one of the statistics $\hat{K}$, $\hat{G}$ or $\hat{F}$. We call these the *(perturbing) $\hat{K}$-model*, *$\hat{G}$-model* and *$\hat{F}$-model* respectively. The score test is then precisely equivalent to comparing $\hat{K}$, $\hat{G}$ or $\hat{F}$ with its predicted expectation using (2.6).

Essentially $\hat{K}$, $\hat{G}$, $\hat{F}$ are re-normalised versions of V_S, V_G, V_A as shown in (2.42)–(2.44). In the case of $\hat{F}$ the renormalisation is not data-dependent, so the $\hat{F}$-model is virtually an area-interaction model, ignoring edge correction. For $\hat{K}$, the renormalisation depends only on $n(\boldsymbol{x})$, and so conditionally on $n(\boldsymbol{x}) = n$, the $\hat{K}$-model and the Strauss process are approximately equivalent. Similarly for $\hat{G}$, the normalisation also depends only on $n(\boldsymbol{x})$, so conditionally on $n(\boldsymbol{x}) = n$, the $\hat{G}$-model and Geyer saturation process are approximately equivalent. If we follow the recommendation of Ripley (1977) to condition on n when testing for interaction, this implies that the use of the K, G or F-function is approximately equivalent to the score test of CSR against a Strauss, Geyer or area-interaction alternative, respectively.

When the null hypothesis is CSR, we saw that the extended model (2.25) is identical to the perturbing model, up to a change in intensity, so that the use of the $\hat{K}$-function is equivalent to testing the null hypothesis of CSR against the alternative of a $\hat{K}$-model. Similarly for $\hat{G}$ and $\hat{F}$. For a more general null hypothesis, the use of the $\hat{K}$-function, for example, corresponds to adopting an alternative hypothesis that is a hybrid between the fitted model and a $\hat{K}$-model.

Note that if the edge correction weight $e_K(u, v)$ is uniformly bounded, the $\hat{K}$-model is integrable for all values of ϕ, avoiding a difficulty with the Strauss process (Kelly and Ripley, 1976).

Computation of the score test statistic (2.26) requires estimation or approximation of the null variance of $\hat{K}(r)$, $\hat{G}(r)$ or $\hat{F}(r)$. A wide variety of approximations is available when the null hypothesis is CSR (Ripley, 1988; Diggle, 2003). For other null hypotheses, simulation estimates would typically be used. A central limit theorem is available for $\hat{K}(r)$, $\hat{G}(r)$ and $\hat{F}(r)$, see e.g. Baddeley (1980); Heinrich (1988b,a); Jolivet (1980); Ripley (1988). However, convergence is not uniform in r, and the normal approximation will be poor for small values of r. Instead Ripley (1976) developed an exact Monte Carlo test (Barnard, 1963; Hope, 1968) based on simulation envelopes of the summary statistic under the null hypothesis.

In the following sections we develop the residual and pseudo-residual diagnostics corresponding to this approach.

2.9 Residual diagnostics for interaction using pairwise distances

This section develops residual (2.35) and pseudo-residual (2.30) diagnostics derived from a summary statistic S which is a sum of contributions depending on pairwise distances.

2.9.1 Residual based on perturbing Strauss model

General derivation

Consider any statistic of the general 'pairwise interaction' form

$$S(\boldsymbol{x}, r) = \sum_{i<j} q(\{x_i, x_j\}, r). \tag{2.45}$$

This can be decomposed in the local form (2.33) with

$$s(u, \boldsymbol{x}, r) = \frac{1}{2} \sum_i q(\{x_i, u\}, r), \quad u \notin \boldsymbol{x}.$$

Hence

$$\Delta_{x_i} S(\boldsymbol{x}, r) = 2s(x_i, \boldsymbol{x}_{-i}, r) \quad \text{and} \quad \Delta_u S(\boldsymbol{x}, r) = 2s(u, \boldsymbol{x}, r), \quad u \notin \boldsymbol{x}.$$

Consequently the pseudo-residual and the pseudo-compensator are just twice the residual and the Papangelou compensator:

$$\Sigma\Delta S(\boldsymbol{x}, r) = 2S(\boldsymbol{x}, r) = \sum_{i \neq j} q(\{x_i, x_j\}, r) \tag{2.46}$$

$$\mathrm{C}\Delta S(\boldsymbol{x}, r) = 2\,\mathrm{C}\,S(\boldsymbol{x}, r) = \int_W \sum_i q(\{x_i, u\}, r) \lambda_{\hat{\theta}}(u, \boldsymbol{x})\, \mathrm{d}u \tag{2.47}$$

$$\mathrm{R}\Delta S(\boldsymbol{x}, z) = 2\,\mathrm{R}\,S(\boldsymbol{x}, r) = 2S(\boldsymbol{x}, r) - 2\,\mathrm{C}\,S(\boldsymbol{x}, r). \tag{2.48}$$

Residual of Strauss potential

The Strauss interaction potential V_S of (2.14) is of the general form (2.45) with $q(\{x_i, x_j\}, r) = \mathbb{I}\{\|x_i - x_j\| \leq r\}$. Hence V_S can be decomposed in the form (2.33) with $s(u, \boldsymbol{x}, r) = \frac{1}{2} t(u, \boldsymbol{x}, r)$ where

$$t(u, \boldsymbol{x}, r) = \sum_i \mathbb{I}\{\|u - x_i\| \leq r\}, \quad u \notin \boldsymbol{x}.$$

Hence the Papangelou compensator of V_S is

$$\mathrm{C}\, V_S(\boldsymbol{x}, r) = \frac{1}{2} \int_W t(u, \boldsymbol{x}, r) \lambda_{\hat{\theta}}(u, \boldsymbol{x})\, \mathrm{d}u. \tag{2.49}$$

Case of CSR

If the null model is CSR with intensity ρ estimated by $\hat{\rho} = n(\boldsymbol{x})/|W|$ (the MLE, which agrees with the MPLE in this case), the Papangelou compensator (2.49) becomes

$$\mathrm{C}\, V_S(\boldsymbol{x}, r) = \frac{\hat{\rho}}{2} \int_W \sum_i \mathbb{I}\{\|u - x_i\| \le r\}\, \mathrm{d}u = \frac{\hat{\rho}}{2} \sum_i |W \cap B(x_i, r)|.$$

Ignoring edge effects we have $|W \cap B(x_i, r)| \approx \pi r^2$ and, applying (2.42), the residual is approximately

$$\mathrm{R}\, V_S(\boldsymbol{x}, r) \approx \frac{n(\boldsymbol{x})^2}{2|W|} \left[\hat{K}_{\boldsymbol{x}}(r) - \pi r^2\right]. \tag{2.50}$$

The term in brackets is a commonly-used measure of departure from CSR, and is a sensible diagnostic because $K(r) = \pi r^2$ under CSR. The Poincaré variance (2.37) is

$$\mathrm{C}^2\, V_S(\boldsymbol{x}, r) = \frac{n(\boldsymbol{x})}{4|W|} \int_W t(u, \boldsymbol{x}, r)^2\, \mathrm{d}u$$

while the exact variance formula (2.64) yields

$$\begin{aligned} \mathbb{V}\mathrm{ar}\,[\mathrm{R}\, V_S(\boldsymbol{X}, r)] &\approx \mathbb{V}\mathrm{ar}\,[\mathrm{I}\, V_S(\boldsymbol{X}, r)] \\ &= \frac{\rho}{4} \int_W \mathbb{E}\left[t(u, \boldsymbol{X}, r)^2\right] \mathrm{d}u + \frac{\rho^2}{4} \int_W \int_W \mathbb{I}\{\|u - v\| \le r\}\, \mathrm{d}u\, \mathrm{d}v. \end{aligned}$$

Now $Y = t(u, \boldsymbol{X}, r)$ is Poisson distributed with mean $\mu = \rho|B(u, r) \cap W|$ so that $\mathbb{E}(Y^2) = \mu + \mu^2$. For $u \in W_{\ominus r}$ we have $\mu = \rho\pi r^2$, so ignoring edge effects

$$\mathbb{V}\mathrm{ar}\,[\mathrm{R}\, V_S(\boldsymbol{X}, r)] \approx \frac{\rho^2}{2}|W|\pi r^2 + \frac{\rho^3}{4}|W|\pi^2 r^4.$$

This has similar functional form to expressions for the variance of $\hat{K}$ under CSR obtained using the methods of U-statistics (Lotwick and Silverman, 1982; Chetwynd and Diggle, 1998; Ripley, 1988), summarised in Diggle (2003, p. 51 ff.). For small r, we have $t(u, \boldsymbol{x}, r) \in \{0, 1\}$ so that

$$\begin{aligned} \mathrm{C}^2\, V_S(\boldsymbol{x}, r) &\approx \frac{n(\boldsymbol{x})^2}{4|W|}\pi r^2 \\ \mathbb{V}\mathrm{ar}\,[\mathrm{R}\, V_S(\boldsymbol{X}, r)] &\approx \frac{\rho^2}{2}|W|\pi r^2 \end{aligned}$$

so that $\mathrm{C}^2\, V_S(\boldsymbol{x}, r)$ is a substantial underestimate (by a factor of approximately 2) of the true variance. Thus a test based on referring $\mathrm{T}\, V_S(\boldsymbol{x}, r)$ to a standard normal distribution may be expected to be conservative for small r.

2.9.2 Residual based on perturbing $\hat{K}$-model

Assuming $\widehat{\rho^2}(\boldsymbol{x}) = \widehat{\rho^2}(n(\boldsymbol{x}))$ depends only on $n(\boldsymbol{x})$, the empirical K-function (2.39) can also be expressed as a sum of local contributions $\hat{K}_{\boldsymbol{x}}(r) = \sum_i k(x_i, \boldsymbol{x}_{-i}, r)$ with

$$k(u, \boldsymbol{x}, r) = \frac{t^w(u, \boldsymbol{x}, r)}{\widehat{\rho^2}(n(\boldsymbol{x}) + 1)|W|}, \quad u \notin \boldsymbol{x}$$

where

$$t^w(u, \boldsymbol{x}, r) = \sum_j e_K(u, x_j) \mathbb{I}\{\|u - x_j\| \le r\}$$

is a weighted count of the points of $\boldsymbol{x}$ that are r-close to the location u. Hence the compensator of the $\hat{K}$-function is

$$\mathrm{C}\,\hat{K}_{\boldsymbol{x}}(r) = \frac{1}{\widehat{\rho^2}(n(\boldsymbol{x}) + 1)|W|} \int_W t^w(u, \boldsymbol{x}, r) \lambda_{\hat{\theta}}(u, \boldsymbol{x})\, \mathrm{d}u. \tag{2.51}$$

Assume the edge correction weight $e_K(u, v) = e_K(v, u)$ is symmetric; e.g. this is satisfied by the Ohser-Stoyan edge correction weight (Ohser and Stoyan, 1981; Ohser, 1983) given by $e_K(u, v) = 1/|W_u \cap W_v|$ where $W_u = \{u + v : v \in W\}$, but not by Ripley's (Ripley, 1976) isotropic correction weight. Then the increment is, for $u \notin \boldsymbol{x}$,

$$\Delta_u \hat{K}_{\boldsymbol{x}}(r) = \frac{\widehat{\rho^2}(\boldsymbol{x}) - \widehat{\rho^2}(\boldsymbol{x} \cup \{u\})}{\widehat{\rho^2}(\boldsymbol{x} \cup \{u\})} \hat{K}_{\boldsymbol{x}}(r) + \frac{2t^w(u, \boldsymbol{x}, r)}{\widehat{\rho^2}(\boldsymbol{x} \cup \{u\})|W|}$$

and when $x_i \in \boldsymbol{x}$

$$\Delta_{x_i} \hat{K}_{\boldsymbol{x}}(r) = \frac{\widehat{\rho^2}(\boldsymbol{x}_{-i}) - \widehat{\rho^2}(\boldsymbol{x})}{\widehat{\rho^2}(\boldsymbol{x}_{-i})} \hat{K}_{\boldsymbol{x}}(r) + \frac{2t^w(x_i, \boldsymbol{x}_{-i}, r)}{\widehat{\rho^2}(\boldsymbol{x}_{-i})|W|}.$$

Assuming the standard estimator $\widehat{\rho^2}(\boldsymbol{x}) = n(n-1)/|W|^2$ with $n = n(\boldsymbol{x})$, the pseudo-sum is seen to be zero, so the pseudo-residual is apart from the sign equal to the pseudo-compensator, which becomes

$$\mathrm{C}\Delta\, \hat{K}_{\boldsymbol{x}}(r) = 2\,\mathrm{C}\,\hat{K}_{\boldsymbol{x}}(r) - \left[\frac{2}{n-2} \int_W \lambda_{\hat{\theta}}(u, \boldsymbol{x})\, \mathrm{d}u\right] \hat{K}_{\boldsymbol{x}}(r)$$

where $\mathrm{C}\,\hat{K}_{\boldsymbol{x}}(r)$ is given by (2.51). So if the null model is CSR and the intensity is estimated by $n/|W|$, the pseudo-residual is approximately $2[\hat{K}_{\boldsymbol{x}}(r) - \mathrm{C}\,\hat{K}_{\boldsymbol{x}}(r)]$, and hence it is equivalent to the residual approximated by (2.50). This is also the conclusion in the more general case of a null model with an activity parameter κ, i.e. where the conditional intensity factorises as

$$\lambda_\theta(u, \boldsymbol{x}) = \kappa \xi_\beta(u, \boldsymbol{x})$$

where $\theta = (\kappa, \beta)$ and $\xi_\beta(\cdot)$ is a conditional intensity, since the pseudo-likelihood equations then imply that $n = \int_W \lambda_{\hat{\theta}}(u, \boldsymbol{x})\, \mathrm{d}u$.

In conclusion, the residual diagnostics obtained from the perturbing Strauss and $\hat{K}$-models are very similar, the major difference being the data-dependent normalisation of the $\hat{K}$-function; similarly for pseudo-residual diagnostics which may be effectively equivalent to the residual diagnostics. In practice, the popularity of the K-function seems to justify using the residual diagnostics based on the perturbing $\hat{K}$-model. Furthermore, due to the familarity of the K-function we often choose to plot the compensator(s) of the fitted model(s) in a plot with the empirical K-function rather than the residual(s) for the fitted model.

2.9.3 Edge correction in conditional case

In the conditional case, the conditional intensity $\lambda_{\hat{\theta}}(u, \boldsymbol{x})$ is known only at locations $u \in W^{\circ}$. The diagnostics must be modified accordingly, by restricting the domain of summation and integration to W°. Appropriate modifications are discussed in Appendices 2.C–2.E.

2.10 Residual diagnostics for interaction using nearest neighbour distances

This section develops residual and pseudo-residual diagnostics derived from summary statistics based on nearest neighbour distances.

2.10.1 Residual based on perturbing Geyer model

The Geyer interaction potential $V_G(\boldsymbol{x}, r)$ given by (2.15) is clearly a sum of local statistics (2.33), and its compensator is

$$\mathrm{C}\, V_G(\boldsymbol{x}, r) = \int_W \mathbb{I}\{d(u, \boldsymbol{x}) \le r\} \lambda_{\hat{\theta}}(u, \boldsymbol{x})\, \mathrm{d}u.$$

The Poincaré variance is equal to the compensator in this case. Ignoring edge effects, $V_G(\boldsymbol{x}, r)$ is approximately $n(\boldsymbol{x})\hat{G}_{\boldsymbol{x}}(r)$, cf. (2.40).

If the null model is CSR with estimated intensity $\hat{\kappa} = n(\boldsymbol{x})/|W|$, then

$$\mathrm{C}\, V_G(\boldsymbol{x}, r) = \hat{\kappa} |W \cap \bigcup_i B(x_i, r)|;$$

ignoring edge effects, this is approximately $\hat{\kappa}|W|\hat{F}(r)$, cf. (2.41). Thus the residual diagnostic is approximately $n(\boldsymbol{x})(\hat{G}(r) - \hat{F}(r))$. This is a reasonable diagnostic for departure from CSR, since $F \equiv G$ under CSR. This argument lends support to Diggle's (Diggle, 1979, eq. (5.7)) proposal to judge departure from CSR using the quantity $\sup |\hat{G} - \hat{F}|$.

This example illustrates the important point that the compensator of a functional summary statistic S should not be regarded as an alternative parametric estimator of the same quantity that S is intended to estimate. In the example just given, under CSR the compensator of $\hat{G}$ is

approximately $\hat{F}$, a qualitatively different and in some sense 'opposite' summary of the point pattern.

We have observed that the interaction potential V_G of the Geyer saturation model is closely related to $\hat{G}$. However, the pseudo-residual associated to V_G is a more complicated statistic, since a straightforward calculation shows that the pseudo-sum is

$$\Sigma\Delta\, V_G(\boldsymbol{x}, r) = V_G(\boldsymbol{x}, r) + \sum_i \sum_{j:\, j\neq i} \mathbb{I}\{\|x_i - x_j\| \leq r \text{ and } d(x_j, \boldsymbol{x}_{-i}) > r\},$$

and the pseudo-compensator is

$$\mathrm{C}\Delta\, V_G(\boldsymbol{x}, r) = \int_W \mathbb{I}\{d(u, \boldsymbol{x}) \leq r\}\lambda_{\hat{\theta}}(u, \boldsymbol{x})\,\mathrm{d}u + \sum_i \mathbb{I}\{d(x_i, \boldsymbol{x}_{-i}) > r\} \int_W \mathbb{I}\{\|u - x_i\| \leq r\}\lambda_{\hat{\theta}}(u, \boldsymbol{x})\,\mathrm{d}u.$$

2.10.2 Residual based on perturbing $\hat{G}$-model

The empirical G-function (2.40) can be written

$$\hat{G}_{\boldsymbol{x}}(r) = \sum_i g(x_i, \boldsymbol{x}_{-i}, r) \tag{2.52}$$

where

$$g(u, \boldsymbol{x}, r) = \frac{1}{n(\boldsymbol{x}) + 1} e_G(u, \boldsymbol{x}, r)\mathbb{I}\{d(u, \boldsymbol{x}) \leq r\}, \quad u \notin \boldsymbol{x} \tag{2.53}$$

so that the Papangelou compensator of the empirical G-function is

$$\mathrm{C}\,\hat{G}_{\boldsymbol{x}}(r) = \int_W g(u, \boldsymbol{x}, r)\lambda_{\hat{\theta}}(u, \boldsymbol{x})\,\mathrm{d}u = \frac{1}{n(\boldsymbol{x}) + 1} \int_{W \cap \bigcup_i B(x_i, r)} e_G(u, \boldsymbol{x}, r)\lambda_{\hat{\theta}}(u, \boldsymbol{x})\,\mathrm{d}u.$$

The residual diagnostics obtained from the Geyer and $\hat{G}$-models are very similar, and we choose to use the diagnostic based on the popular $\hat{G}$-function. As with the K-function we typically use the compensator(s) of the fitted model(s) rather than the residual(s), to visually maintain the close connection to the empirical G-function.

The expressions for the pseudo-sum and pseudo-compensator of $\hat{G}$ are not of simple form, and we refrain from explicitly writing out these expressions. For both the $\hat{G}$- and Geyer models, the pseudo-sum and pseudo-compensator are not directly related to a well-known summary statistic. We prefer to plot the pseudo-residual rather than the pseudo-sum and pseudo-compensator(s).

2.11 Diagnostics for interaction based on empty space distances

2.11.1 Pseudo-residual based on perturbing area-interaction model

When the perturbing model is the area-interaction process, it is convenient to re-parametrise the density, such that the canonical sufficient statistic V_A given in (2.16) is re-defined as

$$V_A(\boldsymbol{x}, r) = \frac{1}{|W|}|W \cap \bigcup_i B(x_i, r)|.$$

This summary statistic is not naturally expressed as a sum of contributions from each point as in (2.33), so we shall only construct the pseudo-residual. Let

$$U(\boldsymbol{x}, r) = W \cap \bigcup_i B(x_i, r).$$

The increment

$$\Delta_u V_A(\boldsymbol{x}, r) = \frac{1}{|W|}\left(|U(\boldsymbol{x} \cup \{u\}, r)| - |U(\boldsymbol{x}, r)|\right), \quad u \notin \boldsymbol{x}$$

can be thought of as 'unclaimed space' — the proportion of space around the location u that is not "claimed" by the points of $\boldsymbol{x}$. The pseudo-sum

$$\Sigma\Delta\, V_A(\boldsymbol{x}, r) = \sum_i \Delta_{x_i} V_A(\boldsymbol{x}, r)$$

is the proportion of the window that has 'single coverage' — the proportion of locations in W that are covered by exactly one of the balls $B(x_i, r)$. This can be used in its own right as a functional summary statistic, and it corresponds to a raw (i.e. not edge corrected) empirical estimate of a summary function $F_1(r)$ defined by

$$F_1(r) = \mathbb{P}\left(\#\{x \in \boldsymbol{X} | d(u, x) \leq r\} = 1\right),$$

for any stationary point process $\boldsymbol{X}$, where $u \in \mathbb{R}^2$ is arbitrary. Under CSR with intensity ρ we have

$$\mathbb{E}F_1(r) = \rho\pi r^2 \exp(-\rho\pi r^2).$$

This summary statistic does not appear to be treated in the literature, and it may be of interest to study it separately, but we refrain from a more detailed study here.

The pseudo-compensator corresponding to this pseudo-sum is

$$\mathrm{C}\Delta\, V_A(\boldsymbol{x}, r) = \int_W \Delta_u V_A(\boldsymbol{x}, r)\lambda_{\hat{\theta}}(u, \boldsymbol{x})\,\mathrm{d}u.$$

This integral does not have a particularly simple interpretation even when the null model is CSR.

2.11.2 Pseudo-residual based on perturbing $\hat{F}$-model

Alternatively one could use a standard empirical estimator $\hat{F}$ of the empty space function F as the summary statistic in the pseudo-residual. The pseudo-sum associated with the perturbing $\hat{F}$-model is

$$\Sigma\Delta\, \hat{F}_{\boldsymbol{x}}(r) = n(\boldsymbol{x})\hat{F}_{\boldsymbol{x}}(r) - \sum_i \hat{F}_{\boldsymbol{x}_{-i}}(r),$$

with pseudo-compensator

$$\mathrm{C}\Delta\, \hat{F}_{\boldsymbol{x}}(r) = \int_W (\hat{F}_{\boldsymbol{x}\cup\{u\}}(r) - \hat{F}_{\boldsymbol{x}}(r))\lambda_{\hat{\theta}}(u, \boldsymbol{x})\, \mathrm{d}u.$$

Ignoring edge correction weights, $\hat{F}_{\boldsymbol{x}\cup\{u\}}(r) - \hat{F}_{\boldsymbol{x}}(r)$ is approximately equal to $\Delta_u V_A(\boldsymbol{x}, r)$, so the pseudo-sum and pseudo-compensator associated with the perturbing $\hat{F}$-model are approximately equal to the pseudo-sum and pseudo-compensator associated with the perturbing area-interaction model. Here, we usually prefer graphics using the pseudo-compensator(s) and the pseudo-sum since this has an intuitive interpretation as explained above.

2.12 Test case: Trend with inhibition

In Sections 2.12–2.14 we demonstrate the diagnostics on the point pattern datasets shown in Figure 2.1. This section concerns the synthetic point pattern in Figure 2.1b.

2.12.1 Data and models

Figure 2.1b shows a simulated realisation of the inhomogeneous Strauss process with first order term $\lambda(x, y) = 200\exp(2x + 2y + 3x^2)$, interaction range $R = 0.05$, interaction parameter $\gamma = \exp(\phi) = 0.1$ and W equal to the unit square, see (2.13) and (2.14). This is an example of extremely strong inhibition (negative association) between neighbouring points, combined with a spatial trend. Since it is easy to recognise spatial trend in the data, (either visually or using existing tools such as kernel smoothing (Diggle, 1985)) the main challenge here is to detect the inhibition after accounting for the trend.

We fitted four point process models to the data in Figure 2.1b. They were (A) a homogeneous Poisson process (CSR); (B) an inhomogeneous Poisson process with the correct form of the first order term, i.e. with intensity

$$\rho(x, y) = \exp(\beta_0 + \beta_1 x + \beta_2 y + \beta_3 x^2) \tag{2.54}$$

where $\beta_0, \ldots, \beta_3$ are real parameters; (C) a homogeneous Strauss process with the correct interaction range $R = 0.05$; and (D) a process of the correct form, i.e. inhomogeneous Strauss with the correct interaction range $R = 0.05$ and the correct form of the first order potential (2.54).

2.12.2 Software implementation

The diagnostics defined in Sections 2.9–2.11 were implemented in the R language, and will be publicly available in the `spatstat` library (Baddeley and Turner, 2005). Unless otherwise stated, models were fitted by approximate maximum pseudo-likelihood using the algorithm of Baddeley and Turner (2000) with the default quadrature scheme in `spatstat`, having an $m \times m$ grid of dummy points where $m = \max(25, 10[1 + 2\sqrt{n(\boldsymbol{x})}/10])$ was equal to 40 for most of our examples. Integrals over the domain W were approximated by finite sums over the quadrature points.

Some models were refitted using a finer grid of dummy points, usually 80×80. The software also supports Huang-Ogata (Huang and Ogata, 1999) one-step approximate maximum likelihood.

2.12.3 Application of $\hat{K}$ diagnostics

Diagnostics for correct model

First we fitted a point process model of the correct form (D). The fitted parameter values were $\hat{\beta} = (5.6, -0.46, 3.35, 2.05)$ and $\hat{\gamma} = 0.217$ using the coarse grid of dummy points, and $\hat{\beta} = (5.6, -0.64, 4.06, 2.44)$ and $\hat{\gamma} = 0.170$ using the finer grid of dummy points, as against the true values $\beta = (5.29, 2, 2, 3)$ and $\gamma = 0.1$.

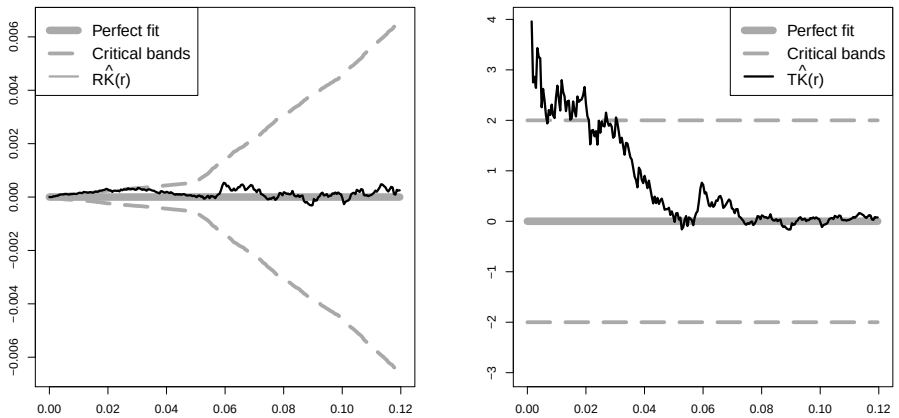


Figure 2.3: Residual diagnostics based on pairwise distances, for a model of the correct form fitted to the data in Figure 2.1b. (a) residual $\hat{K}$-function and two-standard-deviation limits under the fitted model of the correct form. (b) standardised residual $\hat{K}$-function under the fitted model of the correct form.

Figure 2.2 in Section 2.1 shows $\hat{K}$ along with its compensator for the fitted model, together with the theoretical K-function under CSR. The empirical K-function and its compensator coincide very closely, suggesting correctly that the model is a good fit. Figure 2.3a shows the residual $\hat{K}$-function and the two-standard-deviation limits, where the surrogate standard deviation is the square root of (2.37). Figure 2.3b shows the corresponding standardised residual $\hat{K}$-function obtained by dividing by the surrogate standard deviation.

Although this model is of the correct form, the standardised residual exceeds 2 for small values of r. This is consistent with the prediction in Section 2.9.1 that the test would be conservative for small r. For very small r there are small-sample effects so that a normal approximation to the null distribution of the standardised residual is inappropriate.

Formal significance interpretation of the critical bands is limited, because the null distribution of the standardised residual is not known exactly, and the values ± 2 are approximate *pointwise* critical values, i.e. critical values for the score test based on fixed r. The usual problems of multiple testing arise when the test statistic is considered as a function of r: see Diggle (2003, p. 14).

Comparison of competing models

Figure 2.4a shows the empirical K-function and its compensator for each of the models (A)–(D) in Section 2.12.1. Figure 2.4b shows the corresponding residual plots, and Figure 2.4c the standardised residuals. A positive or negative value of the residual suggests that the data are more clustered or more inhibited, respectively, than the model. The clear inference is that the Poisson models (A) and (B) fail to capture interpoint inhibition at range $r \approx 0.05$, while the homogeneous Strauss model (C) is less clustered than the data at very large scales, suggesting that it fails to capture spatial trend. The correct model (D) is judged to be a good fit.

The interpretation of this example requires some caution, because the residual $\hat{K}$-function of the fitted Strauss models (C) and (D) is constrained to be approximately zero at $r = R = 0.05$. The maximum pseudo-likelihood fitting algorithm solves an estimating equation that is approximately equivalent to this constraint, because of (2.42).

It is debatable which of the presentations in Figure 2.4 is more effective at revealing lack-of-fit. A compensator plot such as Figure 2.4a seems best at capturing the main differences between competing models. It is particularly useful for recognising a gross lack-of-fit. A residual plot such as Figure 2.4b seems better for making finer comparisons of goodness-of-fit, for example, assessing models with slightly different ranges of interaction. A standardised residual plot such as Figure 2.4c tends to be highly irregular for small values of r, due to discretisation effects in the computation and the inherent nondifferentiability of the empirical statistic. In difficult cases we may apply smoothing to the standardised residual.

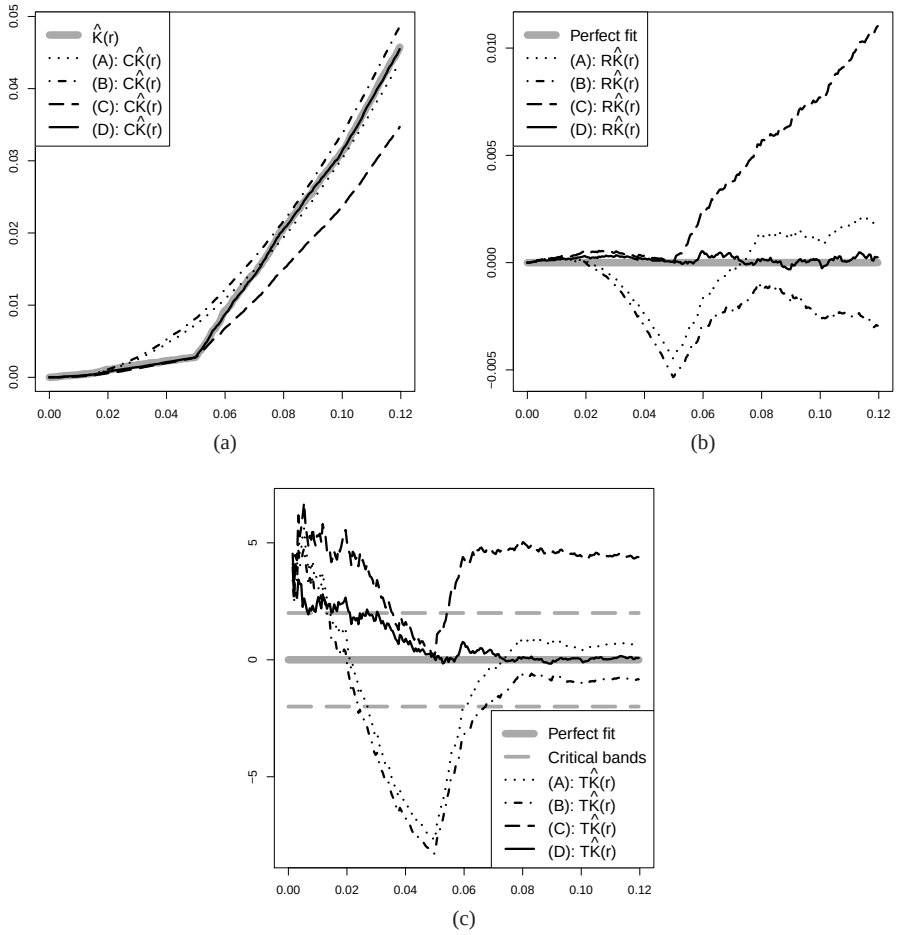


Figure 2.4: Goodness-of-fit diagnostics based on pairwise distances, for each of the models (A)–(D) fitted to the data in Figure 2.1b. (a) $\hat{K}$ and its compensator under each model. (b) residual $\hat{K}$-function (empirical minus compensator) under each model. (c) standardised residual $\hat{K}$-function under each model.

2.12.4 Application of $\hat{G}$ diagnostics

Diagnostics for correct model

Consider again the model of the correct form (D). The residual and compensator of the empirical nearest neighbour function $\hat{G}$ for the fitted model are shown in Figure 2.5. The residual plot suggests a marginal lack-of-fit for $r < 0.025$. This may be correct, since the fitted model parameters (Section 2.12.3) are marginally poor estimates of the true values, in particular of the interaction parameter. This was not reflected so strongly in the $\hat{K}$ diagnostics. This suggests that the residual of $\hat{G}$ may be particularly sensitive to lack-of-fit of interaction.

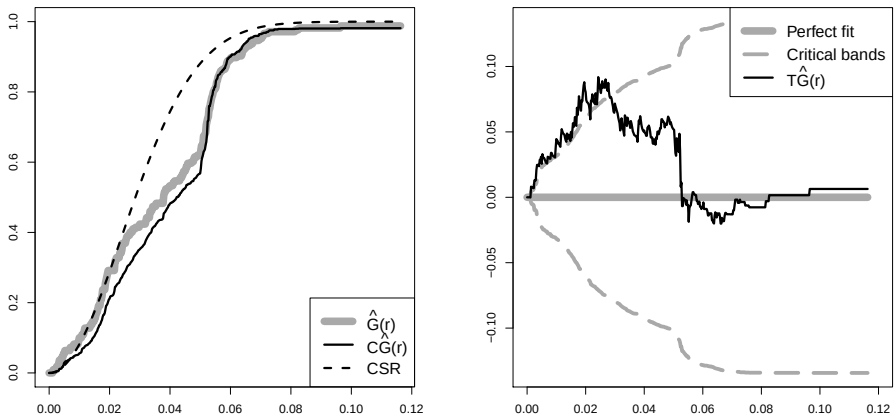


Figure 2.5: Residual diagnostics obtained from the perturbing $\hat{G}$-model when the data pattern is a realisation of an inhomogeneous Strauss process. (a) $\hat{G}$ and its compensator under a fitted model of the correct form, and theoretical G-function for a Poisson process. (b) residual $\hat{G}$-function and two-standard-deviation limits under the fitted model of the correct form.

Comparison of competing models

For each of the four models, Figure 2.6a shows $\hat{G}$ and its Papangelou compensator. This clearly shows that the Poisson models (A) and (B) fail to capture interpoint inhibition in the data. The Strauss models (C) and (D) appear virtually equivalent in Figure 2.6a.

Figure 2.6b shows the standardised residual of $\hat{G}$, and Figure 2.6c the pseudo-residual of V_G (i.e. the pseudo-residual based on the pertubing Geyer model), with spline smoothing applied to both plots. The Strauss models (C) and (D) appear virtually equivalent in Figure 2.6c. The standardised residual plot Figure 2.6b correctly suggests a slight lack of fit for model (C) while model (D) is judged to be a reasonable fit.

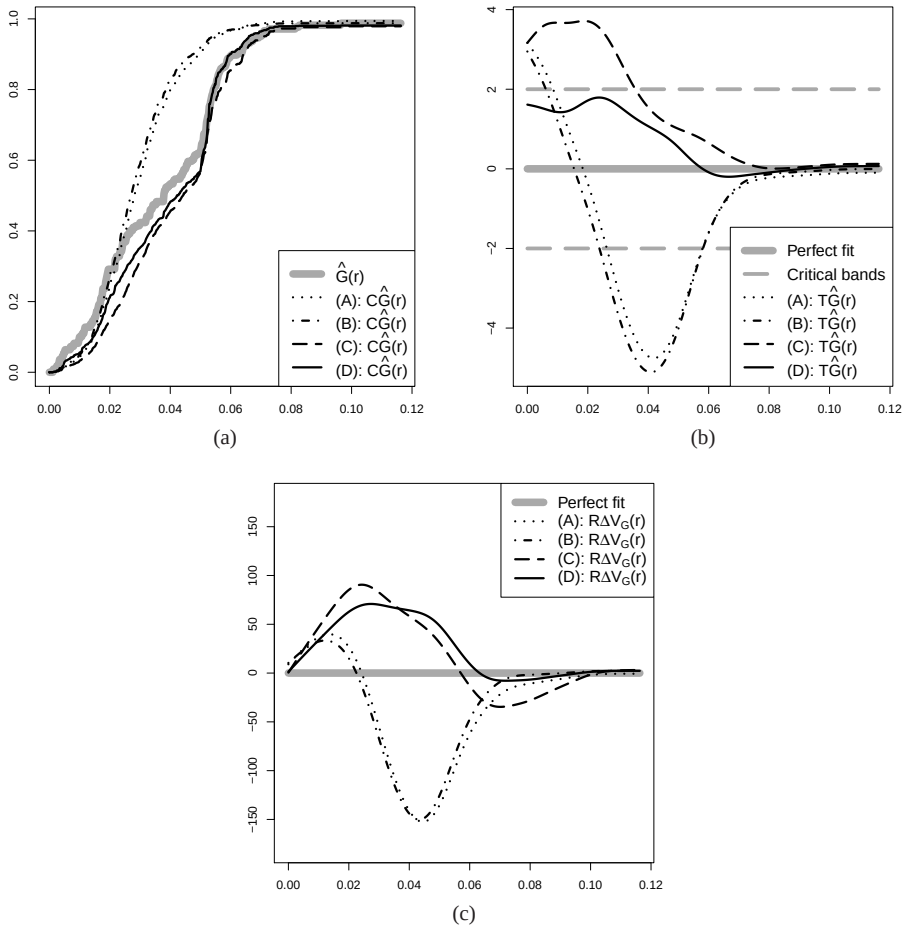


Figure 2.6: Diagnostics based on nearest neighbour distances, for the models (A)–(D) fitted to the data in Figure 2.1b. (a) compensator for $\hat{G}$. (b) smoothed standardised residual of $\hat{G}$. (c) smoothed pseudo-residual derived from a perturbing Geyer model.

2.12.5 Application of $\hat{F}$ diagnostics

Figure 2.7 shows the pseudo-residual diagnostics based on empty space distances. Both diagnostics clearly show models (A)–(B) are poor fits to data. However, in Figure 2.7a it is hard to decide which of the models (C)–(D) provide a better fit. Despite the close connection between the area-interaction process and the $\hat{F}$-model, the diagnostic in Figure 2.7b based on the $\hat{F}$-model performs better in this particular example and correctly shows (D) is the best fit to data. In both cases it is noticed that the pseudo-sum has a much higher peak than the pseudo-compensators for the Poisson models (A)–(B), correctly suggesting that these models do not capture the strength of inhibition present in the data.

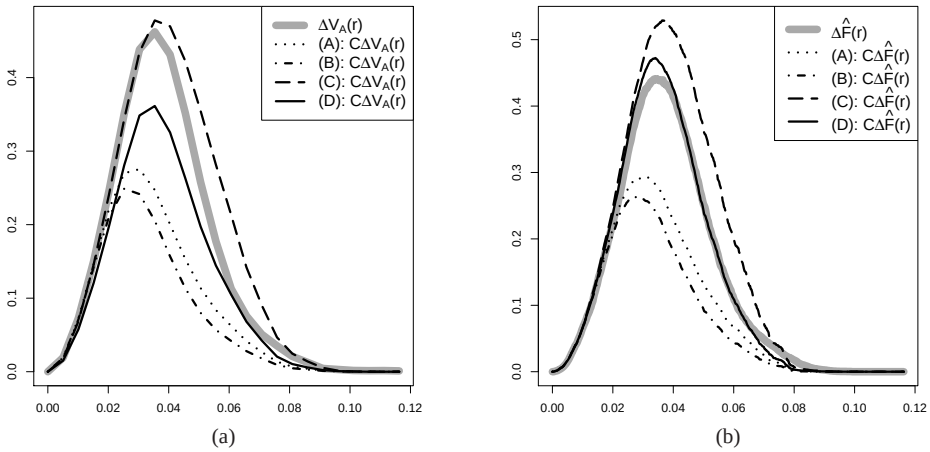


Figure 2.7: Pseudo-sum and pseudo-compensators for the models (A)–(D) fitted to the data in Figure 2.1b when the perturbing model is (a) the area-interaction process (null fitted on a fine grid) and (b) the $\hat{F}$-model (null fitted on a coarse grid).

2.13 Test case: Clustering without trend

2.13.1 Data and models

Figure 2.1c is a realisation of a homogeneous Geyer saturation process (Geyer, 1999) on the unit square, with first order term $\lambda = \exp(4)$, saturation threshold $s = 4.5$ and interaction parameters $r = 0.05$ and $\gamma = \exp(0.4) \approx 1.5$, i.e. the density is

$$f(\boldsymbol{x}) \propto \exp(n(\boldsymbol{x}) \log \lambda + V_{G,s}(\boldsymbol{x}, r) \log \gamma) \tag{2.55}$$

where

$$V_{G,s}(\boldsymbol{x}, r) = \sum_i \min\left\{ s, \sum_{j:\, j\neq i} \mathbb{I}\{\|x_i - x_j\| \leq r\} \right\}.$$

This is an example of moderately strong clustering (with interaction range $R = 2r = 0.1$) without trend. The main challenge here is to correctly identify the range and type of interaction.

We fitted three point process models to the data: (E) a homogeneous Poisson process (CSR); (F) a homogeneous area-interaction process with disc radius $r = 0.05$; (G) a homogeneous Geyer saturation process of the correct form, with interaction parameter $r = 0.05$ and saturation threshold $s = 4.5$ while the parameters λ and γ in (2.55) are unknown. The parameter estimates for (G) were $\log \hat{\lambda} = 4.12$ and $\log \hat{\gamma} = 0.38$.

2.13.2 Application of $\hat{K}$ diagnostics

A plot (not shown) of the $\hat{K}$-function and its compensator, under each of the three models (E)–(G), demonstrates clearly that the homogeneous Poisson model (E) is a poor fit, but does not discriminate between the other models.

Figure 2.8 shows the residual $\hat{K}$ and the smoothed standardised residual $\hat{K}$ for the three models. These diagnostics show that the homogeneous Poisson model (E) is a poor fit, with a positive residual suggesting correctly that the data are more clustered than the Poisson process. The plots suggests that both models (F) and (G) are considerably better fits to the data than a Poisson model. They show that (G) is a better fit than (F) over a range of r values, and suggest that (G) captures the correct form of the interaction.

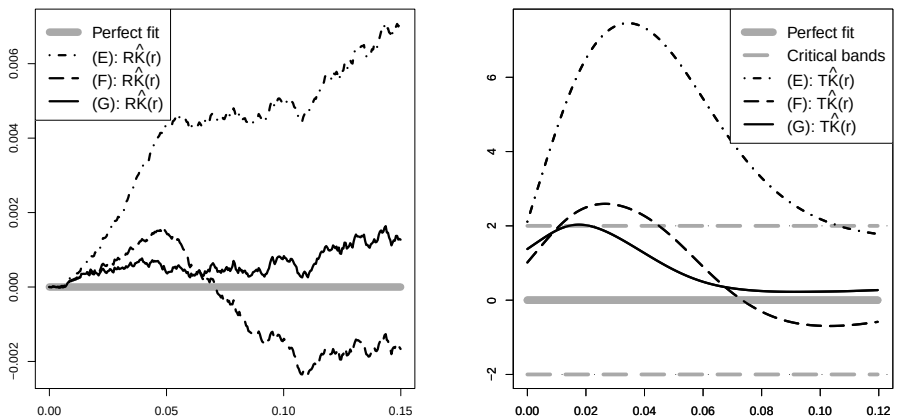
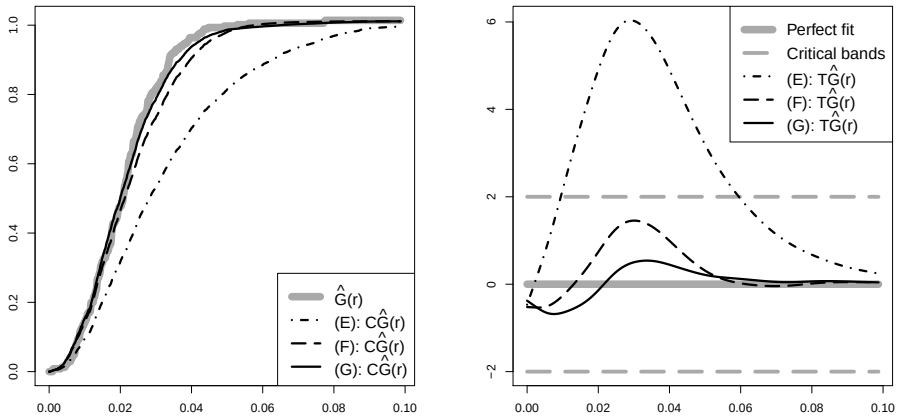


Figure 2.8: Goodness-of-fit diagnostics based on pairwise distances for each of the models (E)–(G) fitted to the data in Figure 2.1c. (a) residual $\hat{K}$; (b) smoothed standardised residual $\hat{K}$.

2.13.3 Application of $\hat{G}$ diagnostics



(a) (b)

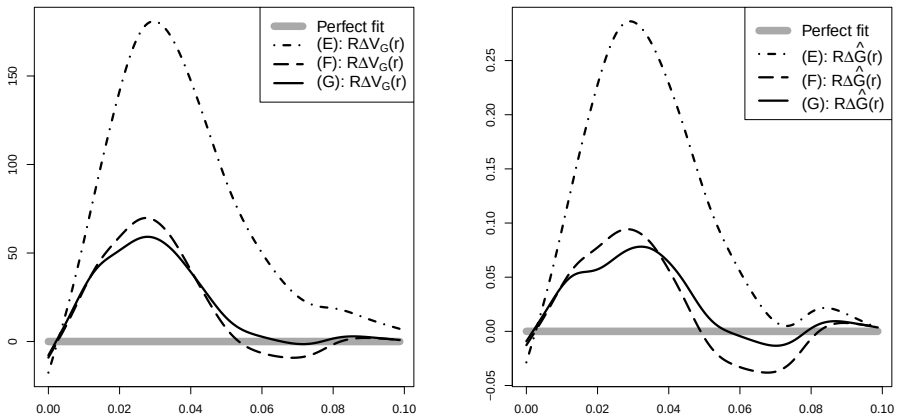
Figure 2.9: Goodness-of-fit diagnostics based on nearest neighbour distances for each of the models (E)–(G) fitted to the data in Figure 2.1c. (a) $\hat{G}$ and its compensator under each model; (b) smoothed standardised residual $\hat{G}$.

Figure 2.9 shows $\hat{G}$ and its compensator, and the corresponding residuals and standardised residuals, for each of the models (E)–(G) fitted to the clustered point pattern in Figure 2.1c. The conclusions obtained from Figure 2.9a are the same as those in Section 2.13.2 based on $\hat{K}$ and its compensator. Figure 2.10 shows the smoothed pseudo-residual diagnostics based on the nearest neighbour distances. The message from these diagnostics is very similar to that from Figure 2.9.

Models (F) and (G) have the same range of interaction $R = 0.1$. Comparing Figures 2.8 and 2.9 we might conclude that the $\hat{G}$-compensator appears less sensitive to the *form* of interaction than the $\hat{K}$-compensator. Other experiments suggest that $\hat{G}$ is more sensitive than $\hat{K}$ to discrepancies in the *range* of interaction.

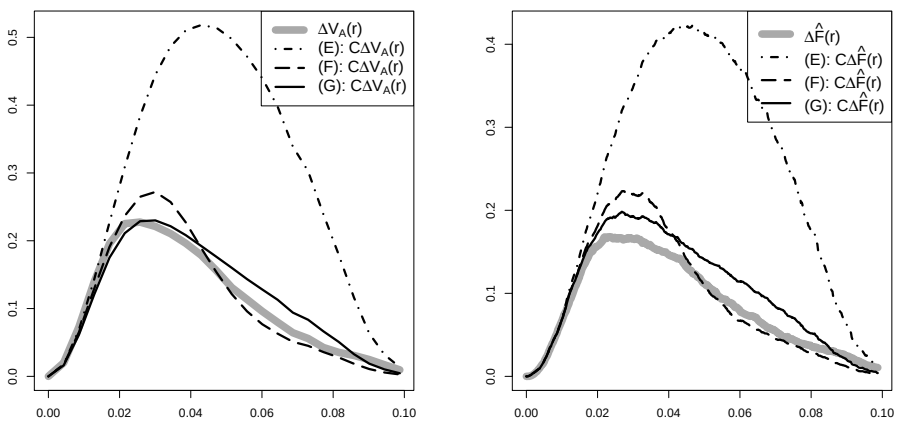
2.13.4 Application of $\hat{F}$ diagnostics

Figure 2.11 shows the pseudo-residual diagnostics based on the empty space distances, for the three models fitted to the clustered point pattern in Figure 2.1c. In this case diagnostics based on the area-interaction process and the $\hat{F}$-model are very similar, as expected due to the close connection between the two diagnostics. Here it is very noticeable that the pseudo-compensator for the Poisson model has a higher peak than the pseudo-sum, which correctly indicates that the data is more clustered than a Poisson process.



(a) (b)

Figure 2.10: Smoothed pseudo-residuals for each of the models (E)–(G) fitted to the clustered point pattern in Figure 2.1c when the perturbing model is (a) Geyer saturation model with saturation 1, and (b) the $\hat{G}$-model.



(a) (b)

Figure 2.11: Pseudo-sum and pseudo-compensators for the models (E)–(G) fitted to the clustered point pattern in Figure 2.1c when the perturbing model is (a) area-interaction process and (b) the $\hat{F}$-model.

2.14 Test case: Japanese pines

2.14.1 Data and models

Figure 2.1a shows the locations of seedlings and saplings of Japanese black pine, studied by Numata (1961, 1964) and analysed extensively by Ogata and Tanemura (1981, 1986). In their definitive analysis (Ogata and Tanemura, 1986) the fitted model was an inhomogeneous 'soft core' pairwise interaction process with log-cubic first order term $\lambda_\beta(x,y) = \exp(P_\beta(x,y))$, where P_β is a cubic polynomial in x and y with coefficient vector β, and density

$$f_{(\beta,\sigma^2)}(\boldsymbol{x}) = c_{(\beta,\sigma^2)} \exp\left(\sum_i P_\beta(x_i) - \sum_{i<j} \left(\sigma^4 / \|x_i - x_j\|^4 \right) \right) \tag{2.56}$$

where σ^2 is a positive parameter.

Here we evaluate the goodness-of-fit of three models: (H) an inhomogeneous Poisson process with log-cubic intensity; (I) a homogeneous soft core pairwise interaction process, i.e. when $P_\beta(x,y)$ in (2.56) is replaced by a real parameter; (J) the Ogata-Tanemura model (2.56). For more detail on the dataset and the fitted inhomogeneous soft core model, see Ogata and Tanemura (1986); Baddeley et al. (2005).

A complication in this case is that the soft core process (2.56) is not Markov, since the pair potential $c(u,v) = \exp(-\sigma^4/\|u-v\|^4)$ is always positive. Nevertheless, since this function decays rapidly, it seems reasonable to apply the residual and pseudo-residual diagnostics, using a cutoff distance R such that $|\log c(u,v)| \le \epsilon$ when $\|u-v\| \le R$, for a specified tolerance ϵ. The cutoff depends on the fitted parameter value σ^2. We chose $\epsilon = 0.0002$ yielding $R = 1$. Estimated interaction parameters were $\hat{\sigma}^2 = 0.11$ for model (I) and $\hat{\sigma}^2 = 0.12$ for model (J).

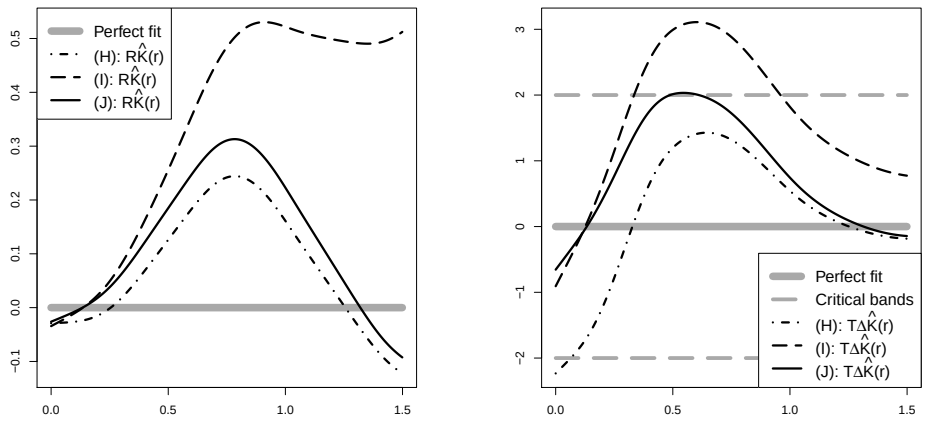
2.14.2 Application of $\hat{K}$ diagnostics

A plot (not shown) of $\hat{K}$ and its compensator for each of the models (H)–(J) suggests that the homogeneous soft core model (I) is inadequate, while the inhomogeneous models (H) and (J) are reasonably good fits to the data. However it does not discriminate between the models (H) and (J).

Figure 2.12 shows smoothed version of the residual and standardised residual of $\hat{K}$ for each model. The Ogata-Tanemura model (J) is judged to be the best fit.

2.14.3 Application of $\hat{G}$ diagnostics

Finally, for each of the models (H)–(J) fitted to the Japanese pines data in Figure 2.1a, Figure 2.13a shows $\hat{G}$ and its compensator. The conclusions are the same as those based on $\hat{K}$ shown in Figure 2.12. Figure 2.14 shows the pseudo-residuals when using either a perturbing Geyer model (Figure 2.14a) or a perturbing $\hat{G}$-model (Figure 2.14b). Figures 2.14a-2.14b tell



(a) (b)

Figure 2.12: Goodness-of-fit diagnostics based on pairwise distances for each of the models (H)–(J) fitted to the Japanese pines data in Figure 2.1a. (a) smoothed residual $\hat{K}$; (b) smoothed standardised residual $\hat{K}$.

almost the same story: the inhomogeneous Poisson model (H) provides the worst fit, while it is difficult to discriminate between the fit for the soft core models (I) and (J). In conclusion, considering Figures 2.12, 2.13 and 2.14, the Ogata-Tanemura model (J) provides the best fit.

2.14.4 Application of $\hat{F}$ diagnostics

Finally, the empty space pseudo-residual diagnostics are shown in Figure 2.15 for the Japanese Pines data in Figure 2.1a. This gives a clear indication that the Ogata-Tanemura model (J) is the best fit to the data, and the data pattern appears to be too regular compared to the Poisson model (H) and not regular enough for the homogeneous softcore model (I).

2.15 Summary of test cases

In this section we discuss which of the diagnostics we prefer to use based on their behaviour for the three test cases in Sections 2.12-2.14.

Typically the various diagnostics supplement each other well, and it is recommended to use more than one diagnostic when judging goodness-of-fit. Compensator and pseudo-compensator plots are informative for gaining an overall picture of goodness-of-fit, and tend to make it easy to recognize a poor fit when comparing competing models. To compare models which fit closely, it may be more informative to use (standardised) residuals or pseudo-residuals. We prefer to use the standardised residuals, but it is important not to over-interpret the significance of departure from zero.

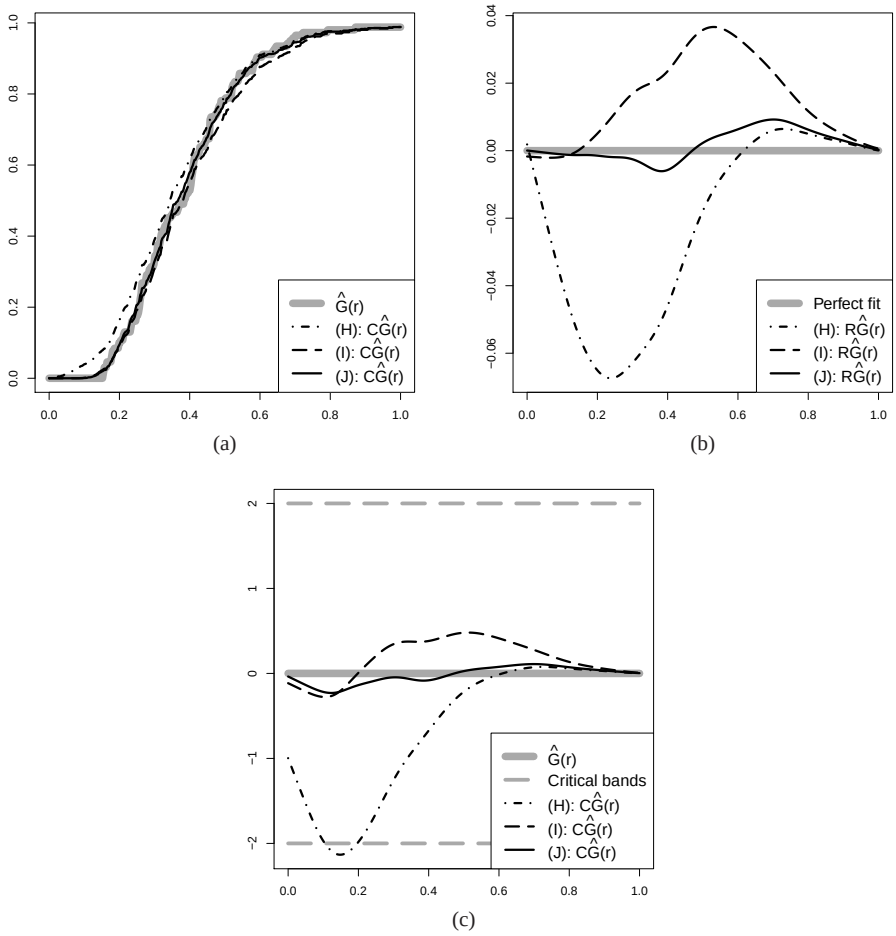


Figure 2.13: Goodness-of-fit diagnostics based on nearest neighbour distances for each of the models (H)–(J) fitted to the Japanese pines data in Figure 2.1a. (a) $\hat{G}$ and its compensator; (b) smoothed residual $\hat{G}$; (c) smoothed standardised residual $\hat{G}$.

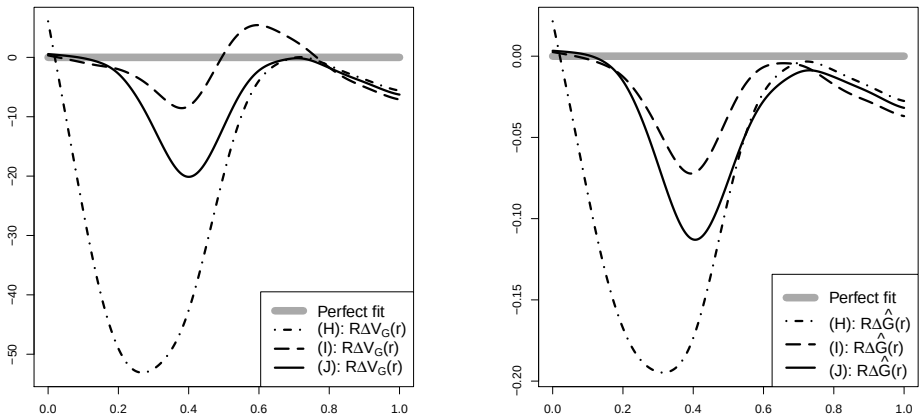


Figure 2.14: Smoothed pseudo-residuals for each of the models (H)–(J) fitted to the Japanese pines data in Figure 2.1a when the perturbing model is (a) Geyer saturation model with saturation 1 (null fitted on a fine grid) and (b) the $\hat{G}$-model.

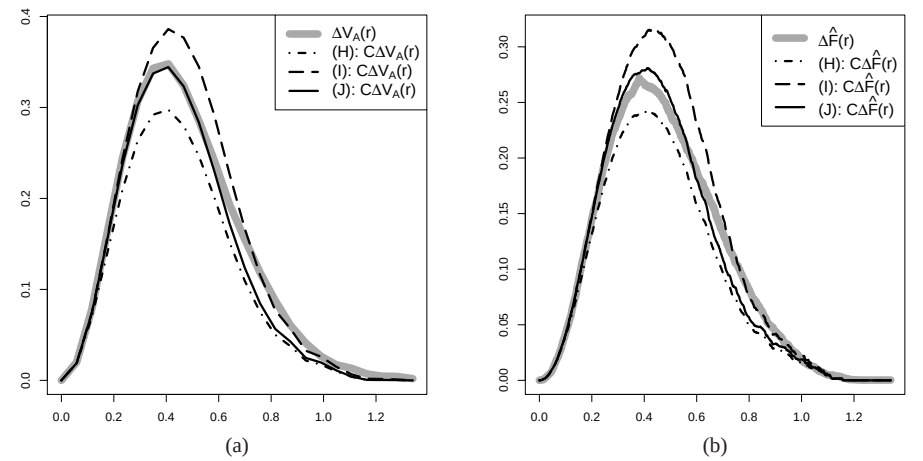


Figure 2.15: Pseudo-sum and pseudo-compensators for the models (H)–(J) fitted to the real data pattern in Figure 2.1a when the perturbing model is (a) area-interaction process and (b) the $\hat{F}$-model.

Based on the test cases here, it is not clear whether diagnostics based on pairwise distances, nearest neighbour distances, or empty space distances are preferable. However, for each of these we prefer to work with compensators and residuals rather than pseudo-compensators and pseudo-residuals when possible (i.e. it is only necessary to use pseudo-versions for diagnostics based on empty space distances). For instance, for the first test case (Section 2.12) the best compensator plot is that in Figure 2.4a based on pairwise distances ($\hat{K}$ and $\mathrm{C}\,\hat{K}$) which makes it easy to identify the correct model. On the other hand in this test case the best residual type plot is that in Figure 2.6b based on nearest neighbour distances ($\mathrm{T}\,\hat{G}$) where the correct model is the only one within the critical bands. However, in the third test case (Section 2.14) the best compensator plot is one of the plots in Figure 2.15 with pseudo-compensators based on empty space distances ($\Sigma\Delta\, V_A$ and $\mathrm{C}\Delta\, V_A$ respectively $\Sigma\Delta\,\hat{F}$ and $\mathrm{C}\Delta\,\hat{F}$) which clearly indicates which model is correct.

In the first and third test cases (Sections 2.12 and 2.14), which both involve inhomogeneous models, it is clear that $\hat{K}$ and its compensator are more sensitive to lack of fit in the first order term than $\hat{G}$ and its compensator (compare e.g. the results for the homogeneous model (C) in Figures 2.4a and 2.6a). It is our general experience that diagnostics based on $\hat{K}$ are particularly well suited to assess the presence of interaction and to identify the general form of interaction. Diagnostics based on $\hat{K}$ and in particular on $\hat{G}$ seem to be good for assessing the range of interaction.

Finally, it is worth mentioning the computational difference between the various diagnostics (timed on a 2.5 GHz laptop). The calculations for $\hat{K}$ and $\mathrm{C}\,\hat{K}$ used in Figure 2.2 are carried out in approximately five seconds whereas the corresponding calculations for $\hat{G}$ and $\mathrm{C}\,\hat{G}$ only take a fraction of a second. For e.g. $\Sigma\Delta\,\hat{F}$ and $\mathrm{C}\Delta\,\hat{F}$ the calculations take about 45 seconds.

Acknowledgments

This paper has benefited from very fruitful discussions with Professor Rasmus P. Waagepetersen. The research was supported by The University of Western Australia, the Danish Natural Science Research Council (grants `272-06-0442` and `09-072331`, *Point process modelling and statistical inference*), the Danish Agency for Science, Technology and Innovation (grant `645-06-0528`, *International PhD student*) and by the Centre for Stochastic Geometry and Advanced Bioimaging, funded by a grant from the Villum Foundation.

Bibliography

Alm, S. (1988). Approximation and simulation of the distributions of scan statistics for Poisson processes in higher dimensions. *Extremes 1*(1), 111–126.

Atkinson, A. (1982). Regression diagnostics, transformations and constructed variables (with discussion). *J. R. Statist. Soc. Ser. B Stat. Methodol. 44*, 1–36.

Baddeley, A. (1980). A limit theorem for statistics of spatial data. *Adv. in Appl. Probab. 12*, 447–461.

Baddeley, A. (1993). Stereology and survey sampling theory. *Bull. Int. Statist. Inst. 50, book 2*, 435–449.

Baddeley, A. (1999). Spatial sampling and censoring. In O. Barndorff-Nielsen, W. Kendall, and M. van Lieshout (Eds.), *Stochastic Geometry: Likelihood and Computation*, Chapter 2, pp. 37–78. London: Chapman and Hall.

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

Baddeley, A., J. Møller, and A. Pakes (2008). Properties of residuals for spatial point processes. *Ann. Inst. Statist. Math. 60*, 627–649.

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

Baddeley, A. and R. Turner (2005). Spatstat: an R package for analyzing spatial point patterns. *J. Statist. Software 12*(6), 1–42. URL: `www.jstatsoft.org`, ISSN: 1548-7660.

Baddeley, A., R. Turner, J. Møller, and M. Hazelton (2005). Residual analysis for spatial point processes (with discussion). *J. R. Statist. Soc. Ser. B Stat. Methodol. 67*(5), 617–666.

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

Barnard, G. (1963). Contribution to discussion of "The spectral analysis of point processes" by M.S. Bartlett. *J. R. Statist. Soc. Ser. B Stat. Methodol. 25*, 294.

Berman, M. (1986). Testing for spatial association between a point process and another stochastic process. *J. Roy. Statist. Soc. Ser. C. 35*, 54–62.

Besag, J. (1978). Some methods of statistical analysis for spatial data. *Bull. Int. Statist. Inst. 44*, 77–92.

Chen, C. (1983). Score tests for regression models. *J. Amer. Statist. Assoc. 78*, 158–161.

Chetwynd, A. and P. Diggle (1998). On estimating the reduced second moment measure of a stationary spatial point process. *Aust. N. Z. J. Stat. 40*, 11–15.

Conniffe, D. (2001). Score tests when a nuisance parameter is unidentified under the null hypothesis. *J. Statist. Plann. Inference 97*, 67–83.

Cook, R. and S. Weisberg (1983). Diagnostics for heteroscedasticity in regression. *Biometrika 70*, 1–10.

Cordy, C. (1993). An extension of the Horvitz-Thompson theorem to point sampling from a continuous universe. *Statist. Probab. Lett. 18*, 353–362.

Cox, D. (1972). The statistical analysis of dependencies in point processes. In P. Lewis (Ed.), *Stochastic Point Processes*, pp. 55–66. New York: Wiley.

Cox, D. and D. Hinkley (1974). *Theoretical statistics*. London: Chapman and Hall.

Cressie, N. (1991). *Statistics for Spatial Data*. New York: John Wiley and Sons.

Daley, D. and D. Vere-Jones (1988). *An Introduction to the Theory of Point Processes*. New York: Springer Verlag.

Davies, R. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika 64*, 247–254.

Davies, R. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika 74*, 33–43.

Diggle, P. (1979). On parameter estimation and goodness-of-fit testing for spatial point patterns. *Biometrika 35*, 87–101.

Diggle, P. (1985). A kernel method for smoothing point process data. *J. Roy. Statist. Soc. Ser. C. 34*, 138–147.

Diggle, P. (2003). *Statistical Analysis of Spatial Point Patterns* (Second ed.). London: Hodder Arnold.

Georgii, H.-O. (1976). Canonical and grand canonical Gibbs states for continuum systems. *Comm. Math. Phys. 48*, 31–51.

Geyer, C. (1999). Likelihood inference for spatial point processes. In O. Barndorff-Nielsen, W. Kendall, and M. van Lieshout (Eds.), *Stochastic Geometry: Likelihood and Computation*, Number 80 in Monographs on Statistics and Applied Probability, Chapter 3, pp. 79–140. Boca Raton, Florida: Chapman and Hall / CRC.

Hanisch, K. (1984). Some remarks on estimators of the distribution function of nearest neighbour distance in stationary spatial point patterns. *Math. Operationsforsch. Statist. Ser. Statist. 15*, 409–412.

Heinrich, L. (1988a). Asymptotic behaviour of an empirical nearest-neighbour distance function for stationary Poisson cluster processes. *Math. Nachr. 136*, 131–148.

Heinrich, L. (1988b). Asymptotic Gaussianity of some estimators for reduced factorial moment measures and product densities of stationary Poisson cluster processes. *Statistics 19*, 87–106.

Hope, A. (1968). A simplified Monte Carlo significance test procedure. *J. R. Statist. Soc. Ser. B Stat. Methodol. 30*, 582–598.

Huang, F. and Y. Ogata (1999). Improvements of the maximum pseudo-likelihood estimators in various spatial statistical models. *J. Comput. Graph. Statist. 8*(3), 510–530.

Illian, J., A. Penttinen, H. Stoyan, and D. Stoyan (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. Chichester: John Wiley and Sons.

Jensen, J. and J. Møller (1991). Pseudolikelihood for exponential family models of spatial point processes. *Ann. Appl. Probab. 1*, 445–461.

Jolivet, E. (1980). Central limit theorem and convergence of empirical processes for stationary point processes. In P. Bastfai and J. Tomko (Eds.), *Point processes and queueing problems*, pp. 117–161. Amsterdam: North-Holland.

Kallenberg, O. (1978). On conditional intensities of point processes. *Z. Wahrscheinlichkeit. 41*, 205–220.

Kallenberg, O. (1984). An informal guide to the theory of conditioning in point processes. *Int. Stat. Rev. 52*, 151–164.

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

Kulldorff, M. (1999). Spatial scan statistics: models, calculations, and applications. In J. Glaz and N. Balakrishnan (Eds.), *Recent Advances on Scan Statistics*, pp. 303–322. Boston: Birkhauser.

Kutoyants, Y. (1998). *Statistical Inference for Spatial Poisson Processes*. Number 134 in Lecture Notes in Statistics. New York: Springer.

Last, G. and R. Penrose (2010). Poisson process Fock space representation, chaos expansion and covariance inequalities. Manuscript in preparation.

Lawson, A. (1993). On the analysis of mortality events around a prespecified fixed point. *J. Roy. Statist. Soc. Ser. A 156*(3), 363–377.

Lotwick, H. W. and B. W. Silverman (1982). Methods for analysing spatial processes of several types of points. *J. R. Statist. Soc. Ser. B Stat. Methodol. 44*, 406–413.

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

Møller, J. and R. Waagepetersen (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Boca Raton: Chapman and Hall/CRC.

Møller, J. and R. P. Waagepetersen (2007). Modern spatial point process modelling and inference (with discussion). *Scand. J. Statist. 34*, 643–711.

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

Numata, M. (1961). Forest vegetation in the vicinity of Choshi — coastal flora and vegetation at Choshi, Chiba prefecture, IV (in Japanese). *Bull. Choshi Mar. Lab.* (3), 28–48. Chiba University.

Numata, M. (1964). Forest vegetation, particularly pine stands in the vicinity of Choshi — flora and vegetation in Choshi, Chiba prefecture, VI (in Japanese). *Bull. Choshi Mar. Lab.* (6), 27–37. Chiba University.

Ogata, Y. and M. Tanemura (1981). Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure. *Ann. Inst. Statist. Math. B 33*, 315–338.

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

Ohser, J. (1983). On estimators for the reduced second moment measure of point processes. *Math. Operationsforsch. Statist. Ser. Statist. 14*, 63–71.

Ohser, J. and D. Stoyan (1981). On the second-order and orientation analysis of planar stationary point processes. *Biom. J. 23*, 523–533.

Papangelou, F. (1974). The conditional intensity of general point processes and an application to line processes. *Z. Wahrscheinlichkeit. 28*, 207–226.

Pregibon, D. (1982). Score tests in GLIM with applications. In *GLIM 82: Proceedings of the International Conference on Generalized Linear Models*, Number 14 in Lecture Notes in Statistics, New York. Springer.

Rao, C. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Math. Proc. Cambridge Philos. Soc. 44*, 50–57.

Rathbun, S. and N. Cressie (1994). Asymptotic properties of estimators of the parameters of spatial inhomogeneous Poisson point processes. *Adv. in Appl. Probab. 26*, 122–154.

Ripley, B. (1976). The second-order analysis of stationary point processes. *J. Appl. Probab. 13*, 255–266.

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

Ripley, B. (1988). *Statistical Inference for Spatial Processes*. Cambridge University Press.

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

Schladitz, K. and A. Baddeley (2000). A third order point process characteristic. *Scand. J. Statist. 27*, 657–671.

Silvapulle, M. (1996). A test in the presence of nuisance parameters. *J. Amer. Statist. Assoc. 91*, 1690–1693.

Stillinger, F., D. Stillinger, S. Torquato, T. Truskett, and P. Debenedetti (2000). Triangle distribution and equation of state for classical rigid disks. *J. Stat. Phys. 100*, 49–72.

Stoyan, D., W. Kendall, and J. Mecke (1987). *Stochastic Geometry and its Applications*. Chichester: John Wiley and Sons.

Stoyan, D. and H. Stoyan (1995). *Fractals, Random Shapes and Point Fields*. Chichester: John Wiley and Sons.

Strauss, D. (1975). A model for clustering. *Biometrika 63*, 467–475.

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

Wald, A. (1941). Some examples of asymptotically most powerful tests. *Ann. Math. Statist. 12*, 396–408.

Waller, L., B. Turnbull, L. Clark, and P. Nasca (1992). Chronic disease surveillance and testing of clustering of disease and exposure: Application to leukaemia incidence and TCE-contaminated dumpsites in upstate New York. *Environmetrics 3*, 281–300.

Wang, P. (1985). Adding a variable in generalized linear models. *Technometrics 27*, 273–276.

Widom, B. and J. Rowlinson (1970). New model for the study of liquid-vapor phase transitions. *J. Chem. Phys. 52*, 1670–1684.

Wu, L. (2000). A new modified logarithmic Sobolev inequality for Poisson point processes and several applications. *Probab. Theory Related Fields 118*, 427–438.

2.A Further diagnostics

In this appendix we present other diagnostics which we have not implemented in software. The examples are therefore not accompanied by experimental results.

2.A.1 Third and higher order functional summary statistics

While the intensity and K-function are frequently-used summaries for the first and second order moment properties of a spatial point process, third and higher order summaries have been less used, though various such summaries have been suggested in e.g. Schladitz and Baddeley (2000); Møller et al. (1998); Stillinger et al. (2000); Stoyan and Stoyan (1995).

Statistic of order k

For a functional summary statistic of k-th order, say

$$S(\boldsymbol{x}, r) = \sum_{\{x_{i_1},\ldots,x_{i_k}\}\subseteq \boldsymbol{x}} q(\{x_{i_1}, \ldots, x_{i_k}\}, r) \tag{2.57}$$

we obtain

$$\Sigma\Delta S(\boldsymbol{x}, r) = k!S(\boldsymbol{x}, r) \ = k! \sum_{\{x_{i_1},\ldots,x_{i_k}\}\subseteq \boldsymbol{x}} q(\{x_{i_1}, \ldots, x_{i_k}\}, r) \tag{2.58}$$

$$\mathrm{C}\Delta S(\boldsymbol{x}, r) = k!\,\mathrm{C}S(\boldsymbol{x}, r) = (k-1)! \int_W \lambda_{\hat{\theta}}(u, \boldsymbol{x}) \sum_{\{x_{i_1},\ldots,x_{i_{k-1}}\}\subseteq \boldsymbol{x}} q(\{x_{i_1}, \ldots, x_{i_{k-1}}, u\}, r)\,\mathrm{d}u \tag{2.59}$$

$$\mathsf{PU}(\hat{\theta}, r) = k!\,\mathrm{R}S(\boldsymbol{x}, r) = k!S(\boldsymbol{x}, r) - k!\,\mathrm{C}S(\boldsymbol{x}, r) \tag{2.60}$$

where $i_1, i_2, \ldots$ are pairwise distinct in the sums in (2.58)-(2.59). So in this case again, pseudo-residual diagnostics are equivalent to those based on residuals.

Third order example

For a stationary and isotropic point process (i.e., when the distribution of $\boldsymbol{X}$ is invariant under translations and rotations), the intensity and K-function of the process completely determine its first and second order moment properties. However, even in this case, the simplest description of third order moments depends on a three-dimensional vector specified from triplets (x_i, x_j, x_k) of points from $\boldsymbol{X}$ such as the lengths and angle between the vectors $x_i - x_j$ and $x_j - x_k$. This is often considered too complex, and instead one considers a certain one-dimensional property of the triangle $T(x_i, x_j, x_k)$ as exemplified below, where $L(x_i, x_j, x_k)$ denotes the largest side in $T(x_i, x_j, x_k)$.

Let the canonical sufficient statistic of the perturbing density (2.27) be

$$S(\boldsymbol{x}, r) = V_T(\boldsymbol{x}, r) = \sum_{i<j<k} \mathbb{I}(L(x_i, x_j, x_k) \le r). \tag{2.61}$$

The perturbing model is a special case of the *triplet interaction point process* studied in Geyer (1999). It is also a special case of (2.57) with

$$q(\{x_i, x_j, x_k\}, r) = \mathbb{I}(L(x_i, x_j, x_k) \le r)$$

and so residual and pseudo-residual diagnostics are equivalent and given by (2.58)-(2.60).

2.A.2 Tessellation functional summary statistics

Some authors have suggested the use of tessellation methods for characterizing spatial point processes; see Illian et al. (2008) and the references therein. A planar tessellation is a subdivision of planar region such as W or the entire plane $\mathbb{R}^2$.

For example, consider the Dirichlet tessellation of W generated by $\boldsymbol{x}$, that is, the tessellation with cells

$$C(x_i|\boldsymbol{x}) = \{u \in W | \, \|u - x_i\| \le \|u - x_j\| \text{ for all } x_j \text{ in } \boldsymbol{x}\}, \quad i = 1, \ldots, n.$$

Suppose the canonical sufficient statistic of the perturbing density (2.27) is

$$S(\boldsymbol{x}, r) = V_O(\boldsymbol{x}, r) = \sum_i \mathbb{I}(|C(x_i|\boldsymbol{x})| \le r). \tag{2.62}$$

This is a sum of local contributions as in (2.33), although not of local statistics in the sense mentioned in Section 2.6.3, since $\mathbb{I}(|C(x_i|\boldsymbol{x})| \le r)$ depends on those points in $\boldsymbol{x}_{-i}$ which are Dirichlet neighbours to x_i and such points may of course not be r-close to x_i (unless r is larger than the diameter of W). We call this perturbing model for a *soft Ord's process*; Ord's process as defined in Baddeley and Møller (1989) is the limiting case $\phi \to -\infty$ in (2.27), i.e. when r is the lower bound on the size of cells. Since $V_O(\boldsymbol{x}) \le n(\boldsymbol{x})$, the perturbing model is well-defined for all $\phi \in \mathbb{R}$.

Let $\sim_{\boldsymbol{x}}$ denote the Dirichlet neighbour relation for the points in $\boldsymbol{x}$, that is, $x_i \sim_{\boldsymbol{x}} x_j$ if $C(x_i|\boldsymbol{x}) \cap C(x_j|\boldsymbol{x}) \ne \emptyset$. Note that $x_i \sim_{\boldsymbol{x}} x_i$. Now,

$$\Delta_u S(\boldsymbol{x}, r) = \mathbb{I}(|C(u|\boldsymbol{x} \cup \{u\})| \le r) + \sum_{v \ne u:\, v \sim_{\boldsymbol{x} \cup \{u\}} u} [\mathbb{I}(|C(v|\boldsymbol{x} \cup \{u\})| \le r) - \mathbb{I}(|C(v|\boldsymbol{x} \setminus \{u\})| \le r)] \tag{2.63}$$

depends not only on the points in $\boldsymbol{x}$ which are Dirichlet neighbours to u (with respect to $\sim_{\boldsymbol{x} \cup \{u\}}$) but also on the Dirichlet neighbours to those points (with respect to $\sim_{\boldsymbol{x} \cup \{u\}}$ or with respect to $\sim_{\boldsymbol{x} \setminus \{u\}}$). In other words, if we define the iterated Dirichlet neighbour relation by that $x_i \sim^2_{\boldsymbol{x}} x_j$ if there exists some x_k such that $x_i \sim_{\boldsymbol{x}} x_k$ and $x_j \sim_{\boldsymbol{x}} x_k$, then $t(u, \boldsymbol{x})$ depends on those points in $\boldsymbol{x}$ which are iterated Dirichlet neighbours to u with respect to $\sim_{\boldsymbol{x} \cup \{u\}}$ or with respect to $\sim_{\boldsymbol{x} \setminus \{u\}}$. The pseudo-sum associated to the soft Ord's process is

$$\Sigma\Delta\, V_O(\boldsymbol{x}, r) = V_O(\boldsymbol{x}, r) + \sum_i \sum_{j \ne i:\, x_j \sim_{\boldsymbol{x}} x_i} \left[\mathbb{I}(|C(x_j|\boldsymbol{x})| \le r) - \mathbb{I}(|C(x_j|\boldsymbol{x}_{-i})| \le r))\right]$$

and from (2.29) and (2.63) we obtain the pseudo-compensator. From (2.36) and (2.62), we obtain the Papangelou compensator

$$\mathrm{C}\, V_O(\boldsymbol{x}, r) = \int_W \mathbb{I}(|C(u|\boldsymbol{x} \cup \{u\})| \leq r)\lambda_{\hat{\theta}}(u, \boldsymbol{x})\, \mathrm{d}u.$$

Many other examples of tessellation characteristics may be of interest. For example, often the Delaunay tessellation is used instead of the Dirichlet tessellation. This is the dual tessellation to the Dirichlet tessellation, where the Delaunay cells generated by $\boldsymbol{x}$ are given by those triangles $T(x_i, x_j, x_k)$ such that the disc containing x_i, x_j, x_k in its boundary does not contain any further points from $\boldsymbol{x}$ (strictly speaking we need to assume a regularity condition, namely that $\boldsymbol{x}$ has to be in general quadratic position; for such details, see Baddeley and Møller (1989). For instance, the summary statistic $t(\boldsymbol{x}, r)$ given by the number of Delaunay cells $T(x_i, x_j, x_k)$ with $L(x_i, x_j, x_k) \leq r$ is related to (2.61) but concerns only the maximal cliques of Dirichlet neighbours (assuming again the general quadratic position condition). The corresponding perturbing model has to the best of our knowledge not been studied in the literature.

2.B Variance formulae

This Appendix concerns the variance of diagnostic quantities of the form

$$\begin{aligned} I &= \sum_i h(x_i, \boldsymbol{X}_{-i}) - \int_W h(u, \boldsymbol{x})\lambda_\theta(u, \boldsymbol{X})\, \mathrm{d}u \\ R &= \sum_i h(x_i, \boldsymbol{X}_{-i}) - \int_W h(u, \boldsymbol{x})\lambda_{\hat{\theta}}(u, \boldsymbol{X})\, \mathrm{d}u \end{aligned}$$

where $h(\cdot)$ is a functional for which these quantities are almost surely finite, $\boldsymbol{X}$ is a point process on W with conditional intensity $\lambda_\theta(u, \boldsymbol{X})$ and $\hat{\theta}$ is an estimate of θ (e.g. the MPLE).

2.B.1 General identity

Exact formulae for the variance of the innovation I and residual R are given in Baddeley et al. (2008). Expressions for $\mathbb{V}\mathrm{ar}[R]$ are unwieldy (Baddeley et al., 2008, Sect. 6), but to a first approximation we may ignore the effect of estimating θ and consider the variance of I. Suppressing the dependence on θ, this is (Baddeley et al., 2008, Prop. 4)

$$\mathbb{V}\mathrm{ar}[I] = \int_W \mathbb{E}\left[h(u, \boldsymbol{X})^2\lambda(u, \boldsymbol{X})\right] \mathrm{d}u + \int_W \int_W \mathbb{E}\left[A(u, v, \boldsymbol{X}) + B(u, v, \boldsymbol{X})\right] \mathrm{d}u\, \mathrm{d}v \qquad (2.64)$$

where

$$\begin{aligned} A(u, v, \boldsymbol{X}) &= \Delta_u h(v, \boldsymbol{X})\, \Delta_v h(u, \boldsymbol{X})\lambda_2(u, v, \boldsymbol{X}) \\ B(u, v, \boldsymbol{X}) &= h(u, \boldsymbol{X})h(v, \boldsymbol{X})\left\{\lambda_2(u, v, \boldsymbol{X}) - \lambda(u, \boldsymbol{X})\lambda(v, \boldsymbol{X})\right\} \end{aligned}$$

where $\lambda_2(u, v, \boldsymbol{x}) = \lambda(u, \boldsymbol{x})\lambda(v, \boldsymbol{x} \cup \{u\})$ is the second order conditional intensity. Note that for a Poisson process $B(u, v, \boldsymbol{X})$ is identically zero since $\lambda(u, \boldsymbol{X}) = \lambda(u)$.

2.B.2 Pseudo-score

Let $S(\boldsymbol{x}, z)$ be a functional summary statistic with function argument z, and take $h(u, \boldsymbol{X}) = \Delta_u S(\boldsymbol{x}, z)$. Then the innovation I is the pseudo-score (2.23), and the variance formula (2.64) becomes

$$\begin{aligned}\mathbb{V}\mathrm{ar}\,[\mathsf{PU}(\theta)] = &\int_W \mathbb{E}\left[(\Delta_u S(\boldsymbol{X}, z))^2\, \lambda(u, \boldsymbol{X})\right] \mathrm{d}u \\ &+ \int_W \int_W \mathbb{E}\left[(\Delta_u \Delta_v S(\boldsymbol{X}, z))^2\, \lambda_2(u, v, \boldsymbol{X})\right] \mathrm{d}u\, \mathrm{d}v \\ &+ \int_W \int_W \mathbb{E}\left[\Delta_u S(\boldsymbol{x}, z)\, \Delta_v S(\boldsymbol{x}, z)\, \{\lambda_2(u, v, \boldsymbol{X}) - \lambda(u, \boldsymbol{X})\lambda(v, \boldsymbol{X})\}\right] \mathrm{d}u\, \mathrm{d}v \quad (2.65)\end{aligned}$$

where for $u \neq v$ and $\{u, v\} \cap \boldsymbol{x} = \emptyset$,

$$\Delta_u \Delta_v S(\boldsymbol{x}, z) = \Delta_v \Delta_u S(\boldsymbol{x}, z) = S(\boldsymbol{x} \cup \{u, v\}, z) - S(\boldsymbol{x} \cup \{u\}, z) - S(\boldsymbol{x} \cup \{v\}, z) + S(\boldsymbol{x}, z).$$

2.C Modified edge corrections

Appendices 2.C–2.E describe modifications to the standard edge corrected estimators of $K(r)$ and $G(r)$ that are required in the conditional case (Section 2.2.3) because the Papangelou conditional intensity $\lambda(u, \boldsymbol{x})$ can or should only be evaluated at locations $u \in W^\circ$ where $W^\circ \subset W$. The corresponding compensators are also given.

Assume the point process is Markov and we are in the conditional case as described in Section 2.5.4. Consider an empirical functional statistic of the form

$$S_W(\boldsymbol{x}, r) = \sum_{x_i \in \boldsymbol{x}} s_W(x_i, \boldsymbol{x} \setminus \{x_i\}, r) \quad (2.66)$$

with compensator (in the unconditional case)

$$\mathrm{C}S_W(\boldsymbol{x}, r) = \int_W s_W(u, \boldsymbol{x}, r) \lambda_{\hat{\theta}}(u, \boldsymbol{x})\, \mathrm{d}u.$$

We explore two different strategies for modifying the edge correction.

In the *restriction approach*, we replace W by W° and $\boldsymbol{x}$ by $\boldsymbol{x}^\circ = \boldsymbol{x} \cap W^\circ$ yielding

$$\begin{aligned} S_{W^\circ}(\boldsymbol{x}, r) &= \sum_{x_i \in \boldsymbol{x}^\circ} s_{W^\circ}(x_i, \boldsymbol{x}^\circ \setminus \{x_i\}, r) \quad (2.67) \\ \mathrm{C}S_{W^\circ}(\boldsymbol{x}, r) &= \int_{W^\circ} s_{W^\circ}(u, \boldsymbol{x}^\circ, r) \lambda_{\hat{\theta}}(u, \boldsymbol{x}^\circ \mid \boldsymbol{x}^+)\, \mathrm{d}u. \end{aligned}$$

In this approach, data points in the boundary region W^+ are ignored in the calculation of the empirical statistic S. The boundary configuration $\boldsymbol{x}^+ = \boldsymbol{x} \cap W^+$ contributes only to the

estimation of θ and the calculation of the conditional intensity $\lambda_{\hat\theta}(\cdot,\cdot \mid \boldsymbol{x}^+)$. This has the advantage that the modified empirical statistic (2.67) is identical to the standard statistic S computed on the subdomain W°; it can be computed using existing software, and requires no new theoretical justification.

The disadvantage of the restriction approach is that we lose information by discarding some of the data. In the *reweighting approach* we retain the boundary points and compute

$$\begin{aligned} S_{W^\circ,W}(\boldsymbol{x},r) &= \sum_{x_i\in\boldsymbol{x}^\circ} s_{W^\circ,W}(x_i,\boldsymbol{x}\setminus\{x_i\},r) \\ \mathrm{C}S_{W^\circ,W}(\boldsymbol{x},r) &= \int_{W^\circ} s_{W^\circ,W}(u,\boldsymbol{x},r)\lambda_{\hat\theta}(u,\boldsymbol{x}^\circ \mid \boldsymbol{x}^+)\,\mathrm{d}u \end{aligned}$$

where $s_{W^\circ,W}(\cdot)$ is a modified version of $s_W(\cdot)$. Thus, boundary points contribute to the computation of the modified summary statistic $S_{W^\circ,W}$ and its compensator. The modification is designed so that $S_{W^\circ,W}$ has properties analogous to S_W.

The K-function and G-function of a point process $\boldsymbol{Y}$ in $\mathbb{R}^2$ is defined (Ripley, 1976, 1977) under the assumption that $\boldsymbol{Y}$ is second-order stationary. The standard estimators $\hat K_W(r)$ respectively $\hat G_{\boldsymbol{x}}(r)$ of the K-function and G-function are designed to be approximately pointwise unbiased estimators of $K(r)$ respectively $G(r)$ when applied to $\boldsymbol{X} = \boldsymbol{Y}\cap W$.

We do not necessarily assume stationarity, but when constructing modified summary statistics $\hat K_{W^\circ,W}(r)$ and $\hat G_{W^\circ,W}(r)$, we shall require that they are also approximately pointwise unbiased estimators of $K(r)$ respectively $G(r)$ when $\boldsymbol{Y}$ is stationary. This greatly simplifies the interpretation of plots of $\hat K_{W^\circ,W}(r)$ and $\hat G_{W^\circ,W}(r)$ and their compensators.

2.D Modified edge corrections for the K-function

2.D.1 Horvitz-Thompson estimators

The most common nonparametric estimators of the K-function (Ripley, 1976; Ohser, 1983; Baddeley, 1999) are continuous Horvitz-Thompson type estimators (Baddeley, 1993; Cordy, 1993) of the form

$$\hat K(r) = \hat K_W(r) = \frac{1}{\widehat{\rho^2}(\boldsymbol{x})|W|}\sum_{i\neq j} e_W(x_i,x_j)\mathbb{I}\{\|x_i-x_j\|\le r\}. \tag{2.68}$$

Here $\widehat{\rho^2} = \widehat{\rho^2}(\boldsymbol{x})$ should be an approximately unbiased estimator of the squared intensity ρ^2 for stationary processes. Usually $\widehat{\rho^2}(\boldsymbol{x}) = n(n-1)/|W|^2$ where $n = n(\boldsymbol{x})$.

The term $e_W(u,v)$ is an edge correction weight, depending on the geometry of W, designed so that the double sum in (2.68), say $\hat Y(r) = \widehat{\rho^2}(\boldsymbol{x})|W|\hat K(r)$, is an unbiased estimator of $Y(r) = \rho^2|W|K(r)$. Popular examples are the Ohser-Stoyan translation edge correction with

$$e_W(u,v) = e_W^{\mathrm{trans}}(u,v) = \frac{|W|}{|W\cap(W+(u-v))|} \tag{2.69}$$

and Ripley's isotropic correction with

$$e_W(u,v) = e_W^{\mathrm{iso}}(u,v) = \frac{2\pi\|u-v\|}{\mathsf{length}(\partial B(u,\|u-v\|)\cap W)}. \tag{2.70}$$

Estimators of the form (2.68) satisfy the local decomposition (2.66) where

$$s_W(u,\boldsymbol{x},r) = \frac{1}{\widehat{\rho^2}(\boldsymbol{x}\cup\{u\})|W|}\sum_j e_W(u,x_j)\mathbb{I}\{\|u-x_j\|\le r\}, \quad u\notin\boldsymbol{x}.$$

Now we wish to modify (2.68) so that the outer summation is restricted to data points x_i in $W^\circ \subset W$, while retaining the property of unbiasedness for stationary and isotropic point processes.

The *restriction estimator* is

$$\hat{K}_{W^\circ}(r) = \frac{1}{\widehat{\rho^2}(\boldsymbol{x}^\circ)|W^\circ|}\sum_{x_i\in\boldsymbol{x}^\circ}\sum_{x_j\in\boldsymbol{x}^\circ_{-i}} e_{W^\circ}(x_i,x_j)\mathbb{I}\{\|x_i-x_j\|\le r\} \tag{2.71}$$

where the edge correction weight is given by (2.69) or (2.70) with W replaced by W°.

A more efficient alternative is to replace (2.68) by the *reweighting estimator*

$$\hat{K}_{W^\circ,W}(r) = \frac{1}{\widehat{\rho^2}(\boldsymbol{x})|W^\circ|}\sum_{x_i\in\boldsymbol{x}^\circ}\sum_{x_j\in\boldsymbol{x}_{-i}} e_{W^\circ,W}(x_i,x_j)\mathbb{I}\{\|x_i-x_j\|\le r\} \tag{2.72}$$

where $e_{W^\circ,W}(u,v)$ is a modified version of $e_W(\cdot)$ constructed so that the double sum in (2.72) is unbiased for $Y(r)$. Compared to the restriction estimator (2.71), the reweighting estimator (2.72) contains additional contributions from point pairs (x_i,x_j) where $x_i\in\boldsymbol{x}^\circ$ and $x_j\in\boldsymbol{x}^+$.

The modified edge correction factor $e_{W^\circ,W}(\cdot)$ for (2.72) is the Horvitz-Thompson weight (Baddeley, 1999) in an appropriate sampling context. Ripley's (Ripley, 1976, 1977) isotropic correction (2.70) is derived assuming isotropy, by Palm conditioning on the location of the first point x_i, and determining the probability that x_j would be observed inside W after a random rotation about x_i. Since the constraint on x_j is unchanged, no modification of the edge correction weight is required, and we take $e_{W^\circ,W}(\cdot) = e_W(\cdot)$ as in (2.70). Note however that the denominator in (2.72) is changed from $|W|$ to $|W^\circ|$.

The Ohser-Stoyan (Ohser and Stoyan, 1981) translation correction (2.69) is derived by considering two-point sets (x_i,x_j) sampled under the constraint that both x_i and x_j are inside W. Under the modified constraint that $x_i\in W^\circ$ and $x_j\in W$, the appropriate edge correction weight is

$$e_{W^\circ,W}(u,v) = e_{W^\circ,W}(u-v) = \frac{|W\cap(W^\circ+(u-v))|}{|W^\circ|}$$

so that $1/e_{W^\circ,W}(z)$ is the fraction of locations u in W° such that $u+z\in W$.

2.D.2 Border correction

A slightly different creature is the border corrected estimator (using usual intensity estimator $\hat{\rho} = n(\boldsymbol{x})/|W|$)

$$\hat{K}_W(r) = \frac{|W|}{n(\boldsymbol{x})n(\boldsymbol{x} \cap W_{\ominus r})} \sum_{x_i \in \boldsymbol{x}} \sum_{x_j \in \boldsymbol{x}_{-i}} \mathbb{I}\{x_i \in W_{\ominus r}\}\mathbb{I}\{\|x_i - x_j\| \le r\}$$

with compensator (in the unconditional case)

$$\mathrm{C}\,\hat{K}_W(r) = \int_{W_{\ominus r}} \frac{|W| \sum_{x_j \in \boldsymbol{x}} \mathbb{I}\{\|u - x_j\| \le r\}}{(n(\boldsymbol{x}) + 1)(n(\boldsymbol{x} \cap W_{\ominus r}) + 1)} \lambda_{\hat{\theta}}(u, \boldsymbol{x}^\circ \mid \boldsymbol{x}^+)\,\mathrm{d}u.$$

The restriction estimator is

$$\hat{K}_{W^\circ}(r) = \frac{|W^\circ|}{n(\boldsymbol{x}^\circ)n(\boldsymbol{x} \cap W^\circ_{\ominus r})} \sum_{x_i \in \boldsymbol{x}^\circ} \sum_{x_j \in \boldsymbol{x}^\circ_{-i}} \mathbb{I}\{x_i \in W^\circ_{\ominus r}\}\mathbb{I}\{\|x_i - x_j\| \le r\}$$

and the compensator is

$$\mathrm{C}\,\hat{K}_{W^\circ}(r) = \int_{W^\circ_{\ominus r}} \frac{|W^\circ| \sum_{x_j \in \boldsymbol{x}^\circ} \mathbb{I}\{\|u - x_j\| \le r\}}{(n(\boldsymbol{x}^\circ) + 1)(n(\boldsymbol{x} \cap W^\circ_{\ominus r}) + 1)} \lambda_{\hat{\theta}}(u, \boldsymbol{x}^\circ \mid \boldsymbol{x}^+)\,\mathrm{d}u.$$

Typically, $W^\circ = W_{\ominus R}$, so $W^\circ_{\ominus r}$ is equal to $W_{\ominus(R+r)}$.

The *reweighting estimator* is

$$\hat{K}_{W^\circ,W}(r) = \frac{|W|}{n(\boldsymbol{x})n(\boldsymbol{x}^\circ \cap W_{\ominus r})} \sum_{x_i \in \boldsymbol{x}^\circ} \sum_{x_j \in \boldsymbol{x}_{-i}} \mathbb{I}\{x_i \in W_{\ominus r}\}\mathbb{I}\{\|x_i - x_j\| \le r\}$$

and the compensator is

$$\mathrm{C}\,\hat{K}_{W^\circ,W}(r) = \int_{W^\circ \cap W_{\ominus r}} \frac{|W| \sum_{x_j \in \boldsymbol{x}} \mathbb{I}\{\|u - x_j\| \le r\}}{(n(\boldsymbol{x}) + 1)(n(\boldsymbol{x}^\circ \cap W_{\ominus r}) + 1)} \lambda_{\hat{\theta}}(u, \boldsymbol{x}^\circ \mid \boldsymbol{x}^+)\,\mathrm{d}u.$$

Usually $W^\circ = W_{\ominus R}$, so $W^\circ \cap W_{\ominus r}$ is equal to $W_{\ominus \max(R,r)}$. From this we conclude that when using border correction we should always use the reweighting estimator since the restriction estimator discards additional information and neither the implementation nor the interpretation is easier.

2.E Modified edge corrections for nearest neighbour function G

2.E.1 Hanisch estimators

Hanisch (1984) considered estimators for $G(r)$ of the form $\hat{G}_W(r) = \hat{D}_{\boldsymbol{x}}(r)/\hat{\rho}$, where $\hat{\rho}$ is some estimator of the intensity ρ, and

$$\hat{D}_{\boldsymbol{x}}(r) = \sum_{x_i \in \boldsymbol{x}} \frac{\mathbb{I}\{x_i \in W_{\ominus d_i}\}\mathbb{I}\{d_i \le r\}}{|W_{\ominus d_i}|} \tag{2.73}$$

where $d_i = d(x_i, \boldsymbol{x} \setminus \{x_i\})$ is the nearest neighbour distance for x_i. If $\hat{\rho}$ were replaced by ρ then $\hat{G}_W(r)$ would be an unbiased, Horvitz-Thompson estimator of $G(r)$. See Stoyan et al. (1987, pp. 128–129); Baddeley (1999).

Hanisch's recommended estimator D_4 is the one in which $\hat{\rho}$ is taken to be

$$\hat{D}_{\boldsymbol{x}}(\infty) = \sum_{x_i \in \boldsymbol{x}} \frac{\mathbb{I}\{x_i \in W_{\ominus d_i}\}}{|W_{\ominus d_i}|}.$$

This is sensible because $\hat{D}_{\boldsymbol{x}}(\infty)$ is an unbiased estimator of ρ and is positively correlated with $\hat{D}_{\boldsymbol{x}}(r)$. The resulting estimator $\hat{G}_W(r)$ can be decomposed in the form (2.66) where

$$s_W(u, \boldsymbol{x}, r) = \frac{\mathbb{I}\{u \in W_{\ominus d(u,\boldsymbol{x})}\}\mathbb{I}\{d(u, \boldsymbol{x}) \le r\}}{\hat{D}_{\boldsymbol{x}\cup\{u\}}(\infty)|W_{\ominus d(u,\boldsymbol{x})}|}$$

for $u \notin \boldsymbol{x}$, where $d(u, \boldsymbol{x})$ is the ('empty space') distance from location u to the nearest point of $\boldsymbol{x}$. Hence the corresponding compensator is

$$\mathrm{C}\,\hat{G}_W(r) = \int_W \frac{\mathbb{I}\{u \in W_{\ominus d(u,\boldsymbol{x})}\}\mathbb{I}\{d(u, \boldsymbol{x}) \le r\}}{\hat{D}_{\boldsymbol{x}\cup\{u\}}(\infty)|W_{\ominus d(u,\boldsymbol{x})}|} \lambda_{\hat{\theta}}(u, \boldsymbol{x})\,\mathrm{d}u$$

This is difficult to evaluate, since the denominator of the integrand involves a summation over all data points: $D_{\boldsymbol{x}\cup\{u\}}(\infty)$ is not related in a simple way to $D_{\boldsymbol{x}}(\infty)$.

Instead, we choose $\hat{\rho}$ to be the conventional estimator $\hat{\rho} = n(\boldsymbol{x})/|W|$. Then

$$\hat{G}_W(r) = \frac{|W|}{n(\boldsymbol{x})}\hat{D}_{\boldsymbol{x}}(r)$$

which can be decomposed in the form (2.66) with

$$s_W(u, \boldsymbol{x}, r) = \frac{|W|}{n(\boldsymbol{x}) + 1} \frac{\mathbb{I}\{u \in W_{\ominus d(u,\boldsymbol{x})}\}\mathbb{I}\{d(u, \boldsymbol{x}) \le r\}}{|W_{\ominus d(u,\boldsymbol{x})}|}$$

for $u \notin \boldsymbol{x}$, so that the compensator is

$$\mathrm{C}\,\hat{G}_W(r) = \frac{|W|}{n(\boldsymbol{x}) + 1} \int_W \frac{\mathbb{I}\{u \in W_{\ominus d(u,\boldsymbol{x})}\}\mathbb{I}\{d(u, \boldsymbol{x}) \le r\}}{|W_{\ominus d(u,\boldsymbol{x})}|} \lambda_{\hat{\theta}}(u, \boldsymbol{x})\,\mathrm{d}u. \tag{2.74}$$

In the *restriction estimator* we exclude the boundary points and take $d_i^\circ = d(x_i, \boldsymbol{x}^\circ_{-i})$, effectively replacing the dataset $\boldsymbol{x}$ by its restriction $\boldsymbol{x}^\circ = \boldsymbol{x} \cap W^\circ$.

$$\hat{G}_{W^\circ}(r) = \frac{|W^\circ|}{n(\boldsymbol{x}^\circ)} \sum_{x_i \in \boldsymbol{x}^\circ} \frac{\mathbb{I}\{x_i \in W^\circ_{\ominus d_i^\circ}\}\mathbb{I}\{d_i^\circ \le r\}}{|W^\circ_{\ominus d_i^\circ}|}$$

The compensator is (2.74) but computed for the point pattern $\boldsymbol{x}^\circ$ in the window W°:

$$\mathrm{C}\,\hat{G}_{W^\circ}(r) = \frac{|W^\circ|}{n(\boldsymbol{x}^\circ) + 1} \int_{W^\circ} \frac{\mathbb{I}\{u \in W^\circ_{\ominus d(u,\boldsymbol{x}^\circ)}\}\mathbb{I}\{d(u, \boldsymbol{x}^\circ) \le r\}}{|W^\circ_{\ominus d(u,\boldsymbol{x}^\circ)}|} \lambda_{\hat{\theta}}(u, \boldsymbol{x}^\circ \mid \boldsymbol{x}^+)\,\mathrm{d}u.$$

In the usual case $W^\circ = W_{\ominus R}$, we have $W^\circ_{\ominus d} = W_{\ominus(R+d)}$.

In the *reweighting estimator* we take $d_i = d(x_i, \boldsymbol{x} \setminus \{x_i\})$. To retain the Horvitz-Thompson property we must replace the weights $1/|W_{\ominus d_i}|$ in (2.73) by $1/|W^\circ \cap W_{\ominus d_i}|$. Thus the modified statistics are

$$\hat{G}_{W^\circ,W}(r) = \frac{|W|}{n(\boldsymbol{x})} \sum_{x_i \in \boldsymbol{x}^\circ} \frac{\mathbb{I}\{x_i \in W_{\ominus d_i}\}\mathbb{I}\{d_i \le r\}}{|W^\circ \cap W_{\ominus d_i}|} \tag{2.75}$$

and

$$\mathrm{C}\,\hat{G}_{W^\circ,W}(r) = \frac{|W|}{n(\boldsymbol{x}) + 1} \int_{W^\circ} \frac{\mathbb{I}\{u \in W_{\ominus d(u,\boldsymbol{x})}\}\mathbb{I}\{d(u,\boldsymbol{x}) \le r\}}{|W^\circ \cap W_{\ominus d(u,\boldsymbol{x})}|} \lambda_{\hat{\theta}}(u, \boldsymbol{x}^\circ \mid \boldsymbol{x}^+)\,\mathrm{d}u. \tag{2.76}$$

In the usual case where $W^\circ = W_{\ominus R}$ we have $W^\circ \cap W_{\ominus d_i} = W_{\ominus \max(R,d_i)}$.

Optionally we may also replace $|W|/n(\boldsymbol{x})$ in (2.75) by $|W^\circ|/n(\boldsymbol{x} \cap W^\circ)$, and correspondingly replace $|W|/(n(\boldsymbol{x}) + 1)$ in (2.76) by $|W^\circ|/(n(\boldsymbol{x} \cap W^\circ) + 1)$.

2.E.2 Border correction

The classical border correction estimate of G is

$$\hat{G}_W(r) = \frac{1}{n(\boldsymbol{x} \cap W_{\ominus r})} \sum_{x_i \in \boldsymbol{x}} \mathbb{I}\{x_i \in W_{\ominus r}\}\mathbb{I}\{d(x_i, \boldsymbol{x}_{-i}) \le r\} \tag{2.77}$$

with compensator (in the unconditional case)

$$\mathrm{C}\,\hat{G}_W(r) = \frac{1}{1 + n(\boldsymbol{x} \cap W_{\ominus r})} \int_{W_{\ominus r}} \mathbb{I}\{d(u,\boldsymbol{x}) \le r\}\lambda_{\hat{\theta}}(u,\boldsymbol{x})\,\mathrm{d}u. \tag{2.78}$$

In the conditional case, the Papangelou conditional intensity $\lambda_{\hat{\theta}}(u,\boldsymbol{x})$ must be replaced by $\lambda_{\hat{\theta}}(u, \boldsymbol{x}^\circ \mid \boldsymbol{x}^+)$ given in (2.24). The *restriction estimator* is obtained by replacing W by W° and $\boldsymbol{x}$ by $\boldsymbol{x}^\circ$ in (2.77)–(2.78) yielding

$$\begin{aligned}
\hat{G}_{W^\circ}(r) &= \frac{1}{n(\boldsymbol{x} \cap W^\circ_{\ominus r})} \sum_{x_i \in \boldsymbol{x}^\circ} \mathbb{I}\{x_i \in W^\circ_{\ominus r}\}\mathbb{I}\{d(x_i, \boldsymbol{x}^\circ_{-i}) \le r\} \\
\mathrm{C}\,\hat{G}_{W^\circ}(r) &= \frac{1}{1 + n(\boldsymbol{x} \cap W^\circ_{\ominus r})} \int_{W^\circ_{\ominus r}} \mathbb{I}\{d(u,\boldsymbol{x}^\circ) \le r\}\lambda_{\hat{\theta}}(u, \boldsymbol{x}^\circ \mid \boldsymbol{x}^+)\,\mathrm{d}u
\end{aligned}$$

Typically $W^\circ = W_{\ominus R}$ so that $W^\circ_{\ominus r} = W_{\ominus(R+r)}$. The *reweighting estimator* is obtained by restricting x_i and u in (2.77)–(2.78) to lie in W°, yielding

$$\begin{aligned}
\hat{G}_{W^\circ,W}(r) &= \frac{1}{n(\boldsymbol{x}^\circ \cap W_{\ominus r})} \sum_{x_i \in \boldsymbol{x}^\circ} \mathbb{I}\{x_i \in W_{\ominus r}\}\mathbb{I}\{d(x_i, \boldsymbol{x}_{-i}) \le r\} \\
\mathrm{C}\,\hat{G}_{W^\circ,W}(r) &= \frac{1}{1 + n(\boldsymbol{x}^\circ \cap W_{\ominus r})} \int_{W^\circ \cap W_{\ominus r}} \mathbb{I}\{d(u,\boldsymbol{x}) \le r\}\lambda_{\hat{\theta}}(u, \boldsymbol{x}^\circ \mid \boldsymbol{x}^+)\,\mathrm{d}u.
\end{aligned}$$

In the usual case where $W^\circ = W_{\ominus R}$ we have $W^\circ \cap W_{\ominus r} = W_{\ominus \max(R,r)}$.

In the same way as for the border corrected estimate for the K-function we always choose to use the reweighting estimator rather than the restriction estimator since there are no disadvantages connected with this.

The border corrected estimator $\hat{G}(r)$ is well known for having relatively poor performance and sample properties. It is not guaranteed to be a monotonically increasing function of r, and its bias and variance are generally greater than those of the Horvitz-Thompson style estimators. The main reason for choosing the border corrected estimator is its computational efficiency in large datasets. We may expect similar considerations to apply to its compensator.

CHAPTER 3

A Model for Positively Correlated Count Variables

Publication details

Co-authors: Jesper Møller[†]

[†] *Department of Mathematical Sciences, Aalborg University*

Journal: International Statistical Review, 78 (2010) 65-80

Abstract:
An α-permanental random field is briefly speaking a model for a collection of non-negative integer valued random variables with positive associations. Though such models possess many appealing probabilistic properties, many statisticians seem unaware of α-permanental random fields and their potential applications. The purpose of this paper is to summarize useful probabilistic results, study stochastic constructions and simulation techniques, and discuss some examples of α-permanental random fields. This should provide a useful basis for discussing the statistical aspects in future work.

Keywords:
α-determinant; α-permanent; covariance; doubly stochastic construction; negative binomial distribution; simulation; Poisson randomization.

3.1 Introduction

For any real number α and any $n \times n$ matrix A with entries $A_{i,j}$, the α-determinant of A is defined as (Shirai, 2007)

$$\det_\alpha A = \sum_{\sigma \in \mathcal{S}_n} \alpha^{n-c(\sigma)} A_{1,\sigma(1)} A_{2,\sigma(2)} \cdots A_{n,\sigma(n)}, \tag{3.1}$$

where $\mathcal{S}_n$ is the set of all permutations of $1, \ldots, n$, and $c(\sigma)$ denotes the number of cycles in σ. If $\alpha = -1$ we obtain the usual determinant, which can be easily calculated, and for $\alpha = 1$, (3.1) is called the permanent of A. For $\alpha \neq -1$ the computation of $\det_\alpha A$ is believed to be intractable (for $\alpha = 1$, the computation is know to be intractable in the sense it is #P-complete, see Valiant (1979)). However, in Kou and McCullagh (2009) an algorithm for approximating the α-determinant for $\alpha > 0$ is given, which may help overcome the difficulties of calculating the α-determinant in applied work. Notice that some authors prefer to work with the related α-permanent $|A|_\alpha = \alpha^n \det_{1/\alpha} A$, but in the present paper the α-determinant is used.

The α-determinant plays an important role in the study of permanental (or boson) point processes (where $\alpha = 1$) and determinantal (or fermion) point processes (where $\alpha = -1$) as introduced by (Macchi, 1971, 1975) and their extensions to α-permanental ($\alpha > 0$) and α-determinantal ($\alpha < 0$) point processes, which have received much research interest in probability theory in recent years (Shirai and Takahashi, 2003a,b; Georgii and Yoo, 2005; Hough et al., 2006; McCullagh and Møller, 2006). The focus of the present paper is on α-permanental point process models, i.e. when $\alpha > 0$. We consider the simplest setting, namely when such point processes can be identified by a collection of of discrete non-negative random variables $\boldsymbol{N} = (N_s;\ s \in S)$, which are indexed by a finite set $S = \{s_1, \ldots, s_m\}$. In applications the indices s_i typically correspond to distinct spatial locations in $\mathbb{R}$ or $\mathbb{R}^2$ referred to as sites. In the terminology of spatial statistics (see, e.g., Cressie (1993)), $\boldsymbol{N}$ is then a *random field*, and therefore α-permanental point processes in this setup are called α-permanental random fields.

Such random fields are used to model multivariate count data, with spatial dependence between the counts. A simple example of an α-permanental random fields is obtained by a doubly stochastic construction as follows. First, introduce auxiliary real random variables Y_s

associated with the sites $s \in S$, and assume that $\boldsymbol{Y} = (Y_{s_1}, \ldots, Y_{s_m})$ follows a zero-mean multivariate Gaussian distribution with covariance matrix C. Second, conditional on $\boldsymbol{Y}$, independently for each site $s \in S$, let N_s be Poisson distributed with mean Y_s^2. Then $\boldsymbol{N} = (N_s;\ s \in S)$ is an α-permanental random field with $\alpha = 2$, as will be described in more detail in Section 3.4.1. The class of α-permanental random fields is much more general and cannot be constructed in such a simple way, but this example illustrates which type of data the model class can be used for.

To the best of our knowledge, the statistical and computational aspects of these models have so far mainly been unexplored, and many statisticians may be unaware of the models many appealing properties and potential applications. The present paper should provide a useful basis for discussing the statistical aspects of α-permanental random fields, and it is based partly on the above-mentioned references and the seminal work by Griffiths (1984), Griffiths and Milne (1987), and in particular Vere-Jones (1997), and partly on some new results of our own.

The remainder of this paper is organized as follows. Section 3.2 introduces some notation, discusses the definition and existence of α-permanental random fields, and presents three specific examples of model types, which are also discussed in the subsequent sections. Section 3.3 reviews various useful properties of α-permanental random fields. Section 3.4 considers stochastic constructions and simulation of α-permanental random fields.

3.2 Preliminaries

3.2.1 Definition and notation

Let $S = \{s_1, \ldots, s_m\}$ be an arbitrary finite set and $\boldsymbol{N} = (N_s, s \in S)$ a collection of non-negative integer-valued random variables. This will be an α-permanental random field with parameter (α, C) if for all $\boldsymbol{z} = (z_s; s \in S)$ with $|z_s| \leq 1$, $s \in S$, the probability generating function for $\boldsymbol{N}$,

$$\varphi(\boldsymbol{z}) = \varphi(z_s; s \in S) = \mathrm{E} \prod_{s \in S} z_s^{N_s}$$

is of a particular form specified below. Here α is a positive number and $C : S \times S \to \mathbb{R}$ is a function which satisfy certain conditions such the random field exists; these conditions are also discussed below.

Since S is finite, the function C can be identified with a real $m \times m$ matrix, also denoted C. Whether C is considered a function or a matrix will be clear from the context, and the two representations are used interchangeably throughout the paper. Notationally we write $C_{i,j} = C(s_i, s_j)$. Furthermore, I denotes the identity matrix, $|A|$ is the determinant of a square matrix A, and we take $0^0 = 1$.

Definition 1. We say that $\boldsymbol{N} = (N_s, s \in S)$ is an α-permanental random field with parameter (α, C) if

$$\varphi(\boldsymbol{z}) = |I + \alpha(I - Z)C|^{-1/\alpha} \tag{3.2}$$

where Z denotes the diagonal matrix with $(z_s; s \in S)$ on the diagonal. We then write $\boldsymbol{N} \sim \text{per}(\alpha, C)$.

In accordance with the references given at the very beginning of Section 3.1, we call it an α-permanental random field, while a so-called α-determinantal random field appears if α is negative. The reason for the names of these models are partly explained by the close connection between α-determinants and α-permanents and the fact that density and moment expressions are given in terms of α-determinants or α-permanents, see Section 3.3.

If $\boldsymbol{N} \sim \text{per}(\alpha, C)$ then $I + \alpha C$ is necessarily non-singular (otherwise (3.2) would not be well-defined for $\boldsymbol{z} = \boldsymbol{0}$), and we can define the matrix

$$\tilde{C} = \alpha C(I + \alpha C)^{-1} = I - (I + \alpha C)^{-1}. \tag{3.3}$$

Using this parametrization we can write (3.2) as

$$\varphi(\boldsymbol{z}) = \left[|I - \tilde{C}|/|I - Z\tilde{C}|\right]^{1/\alpha}. \tag{3.4}$$

On the other hand, if (3.4) is a probability generating function then $I - \tilde{C}$ is necessarily non-singular, and setting

$$C = \frac{1}{\alpha}\tilde{C}(I - \tilde{C})^{-1} \tag{3.5}$$

we obtain (3.2). Consequently, we can equally well parametrize $\text{per}(\alpha, C)$ by $(\alpha, \tilde{C})$.

For notational convenience we sometimes write i for s_i. Using the Schur decomposition of C (Golub and Van Loan, 1996), the relation between the eigenvalues λ_i of C and the eigenvalues $\tilde{\lambda}_i$ of $\tilde{C}$ is seen to be

$$\lambda_i = \frac{\tilde{\lambda}_i}{\alpha - \alpha\tilde{\lambda}_i}, \quad \tilde{\lambda}_i = \frac{\alpha\lambda_i}{1 + \alpha\lambda_i}, \quad i = 1, \ldots, m. \tag{3.6}$$

We let $\|\lambda_i\|$ denote the modulus of λ_i and define the spectral norm of C as

$$\|C\| = \max\{\|\lambda_1\|, \ldots, \|\lambda_m\|\} \quad \text{(and similarly for } \tilde{C}\text{)}.$$

Finally, a useful expansion for $\|z_s\| \leq 1$, $s \in S$, is

$$-\log|I - Z\tilde{C}| = \sum_{n=1}^{\infty} \text{tr}\left\{\left(Z\tilde{C}\right)^n\right\}/n \quad \text{if } \|\tilde{C}\| < 1. \tag{3.7}$$

See e.g. Goulden and Jackson (1983).

3.2.2 Existence of the α-permanental random field

By Definition 1, $\text{per}(\alpha, C)$ exists if and only if (3.2) (or equivalently (3.4)) is a proper probability generating function. It is clear that this is not the case for all (α, C). The problem of

characterizing the set of (α, C) such that (3.2) is a proper probability generating function is treated in detail in Vere-Jones (1997), but no easily verifiable necessary and sufficient condition is known. There are however some known sufficient conditions expressed either through (α, C) or $(\alpha, \tilde{C})$, and the two most important sufficient conditions for the present exposition are the following.

Condition I: C is a covariance matrix and $\alpha \in \left(0, \frac{2}{m-1}\right) \cup \left\{\frac{2}{m-1}, \frac{2}{m-2}, \ldots, 1, 2\right\}$.

Condition II: $\tilde{C}$ has non-negative entries and $\|\tilde{C}\| < 1$.

Condition I is a minor extension of the corresponding result in Vere-Jones (1997), and it can be found in e.g. Shirai (2007). It is related to the double stochastic construction of the α-permanental random field described in Section 1 and Section 3.4.1. The sufficiency of Condition II is an immediate consequence of (3.19) in Section 3.3.3, where the density of the α-permanental random field is expressed using α-determinants of $\tilde{C}$. Note that α can be any positive number under Condition II.

One important necessary condition C must satisfy is

$$C(s, s) \geq 0 \quad \text{for all } s \in S. \tag{3.8}$$

This follows later from equation (3.10).

3.2.3 Examples

In this section, our running examples of α-permanental random field models are introduced.

Example I

Let $C = \kappa Q$, where $\kappa > 0$ and Q is a projection of rank $r > 0$. In this special case $\boldsymbol{N}$ satisfies many striking and unusual properties, and we refer therefore to it as the special α-permanental random field. In this case it turns out that C and $\tilde{C}$ are proportional. More specifically $\tilde{C} = \frac{\alpha\kappa}{1+\alpha\kappa} Q$, which is verified in the following. From (3.3) we need only verify that

$$\frac{\alpha\kappa}{1+\alpha\kappa} Q = \alpha\kappa Q(I + \alpha\kappa Q)^{-1}.$$

This is equivalent to

$$\alpha\kappa Q + (\alpha\kappa Q)^2 = \alpha\kappa Q + \alpha^2\kappa^2 Q,$$

which clearly is true since Q is a projection and consequently idempotent.

Example II

If C has rank one it can be written on the form, $C_{i,j} = a_i b_j$, $i, j = 1, \ldots, m$, for some real vectors $(a_1, \ldots, a_m)$ and $(b_1, \ldots, b_m)$. Assume that C is of this form with $\sum_{i=1}^m C_{i,i} > 0$. The matrix $A := (I - Z)C$ appearing in (3.2) has (i, j)'th entry $(1 - z_i)a_i b_j$. If A is a non-zero matrix, i.e. $z_i \neq 1$ for all $i = 1, \ldots, m$, then A has rank one and eigenvalue $\sum_{i=1}^m (1 - z_i)C_{i,i}$ with corresponding eigenvector $((1 - z_1)a_1, \ldots, (1 - z_m)a_m)^\top$. Consequently, by (3.2),

$$\varphi(\boldsymbol{z}) = \left(1 + \alpha \left[\sum_{i=1}^m (1 - z_i)C_{i,i}\right]\right)^{-1/\alpha}.$$

It follows that the distribution of $\boldsymbol{N}$ depends only on C through the diagonal elements. Consequently, we may without loss of generality assume C to be a positive definite symmetric matrix with non-negative entries of the form $C_{i,j} = \sqrt{c_i c_j}$ for some non-zero vector $\boldsymbol{c} = (c_1, \ldots, c_m)$, $c_i \geq 0$, $i = 1, \ldots, m$. Then the only non-zero eigenvalue of C is $\kappa := \sum_{i=1}^m c_i = \sum_{i=1}^m C_{i,i}$, and it is a special α-permanental random field as discussed in Example I with $Q = \frac{1}{\kappa} C$.

Remark. *Example II it was sufficient to let C be symmetric, but this is not in general possible for an α-permanental random field where C has rank higher than one. Take e.g. $\boldsymbol{N} \sim per(\alpha, C)$ with C a non-symmetric matrix such that the α-permanental random field is well-defined. Then a corresponding random field parametrized by a symmetric matrix C' would have to be given by $C'_{i,j} = \sqrt{C_{i,j} C_{j,i}}$ for the covariances to be the same, but the distribution is in general not the same using C and C' since the corresponding α-determinants (and thereby the factorial moments as considered in Section 3.3.2) differ when the rank is higher than one.*

Example III

In this example, we consider a model for an α-permanental random field in the case where $S = \{s_1, \ldots, s_m\}$ is a finite number of sites on the real line with $s_1 < \cdots < s_m$. First a slight modification of the double stochastic construction of $\boldsymbol{N} = (N_s; s \in S)$ as described in Section 1 and Section 3.4.1 (Method II) is considered, where we require that $\alpha = 2/k$ for some $k \in \mathbb{N}$. Furthermore, for each $s \in S$, let $z(s) = (z_0(s), z_1(s), \ldots, z_p(s))$ be given covariates for N_s, where we let $z_0(s) = 1$ for all $s \in S$ such that β_0 introduced below has the interpretation of an intercept on the log-scale. Let a random mean field $\boldsymbol{M} = (M_s; s \in S)$ be modeled as $M_s = \exp(\beta^\top z(s))(Y_{1,s}^2 + \cdots + Y_{k,s}^2)$, where $\boldsymbol{Y}_1 = (Y_{1,s}; s \in S), \ldots, \boldsymbol{Y}_k = (Y_{k,s}; s \in S)$ are independent zero-mean Gaussian random fields with the exponential covariance matrix $\mathrm{Cov}(Y_{i,s}, Y_{i,t}) = \rho^{|s-t|}$, $0 < \rho < 1$. Suppose that $\boldsymbol{N}$ conditioned on $\boldsymbol{M}$ consists of mutually independent Poisson random variables N_s with mean M_s, $s \in S$. Then $\boldsymbol{N} \sim \mathrm{per}(\alpha, C)$, where

$$C_{i,j} = C(s_i, s_j) = \exp\left(\beta^\top (z(s_i) + z(s_j))/2\right) \rho^{|s_i - s_j|}. \tag{3.9}$$

Using this construction the model is at least well-defined for $\alpha = 2/k$, $k \in \mathbb{N}$, but the following proposition extends the model to all $\alpha > 0$.

Proposition 1. Let $S = \{s_1, \ldots, s_m\}$, $s_1 < \cdots < s_m$, $0 < \rho < 1$, and $\alpha > 0$. If C is given by (3.9) then all entries of $\tilde{C} = \alpha C(I + \alpha C)^{-1}$ are non-negative and per(α, C) is thus well-defined.

Proof. We have $C = DBD$, where D is a diagonal matrix with $D_{i,i} = \exp(\beta^\top z(s_i)/2)$, $i = 1, \ldots, m$, and B is the matrix with entries $B_{i,j} = \rho^{|s_i - s_j|}$. Using a notation as in Appendix, B is a Green's matrix with $a_i = \rho^{-|s_i - s_1|}$ and $b_i = \rho^{|s_i - s_1|}$. Thus, if the inverse $B^{-1} = T$ exists, T is tridiagonal, and it is straightforward to verify that the matrix T given in the following is indeed the inverse of B. The diagonal elements are

$$T_{i,i} = \frac{1 - \rho^{2|s_{i+1} - s_{i-1}|}}{(1 - \rho^{2|s_{i+1} - s_i|})(1 - \rho^{2|s_i - s_{i-1}|})}, \quad i = 1, \ldots, m$$

where we define $s_0 = s_{m+1} = \infty$, such that $\rho^{2|s_1 - s_0|} = \rho^{2|s_2 - s_0|} = \rho^{2|s_{m+1} - s_m|} = \rho^{2|s_{m+1} - s_{m-1}|} = 0$. The non-zero off-diagonal elements are

$$T_{i,i+1} = T_{i+1,i} = \frac{-\rho^{|s_{i+1} - s_i|}}{1 - \rho^{2|s_{i+1} - s_i|}}, \quad i = 1, \ldots, m-1.$$

Now,

$$\tilde{C} = (I + (\alpha C)^{-1})^{-1} = (I + \alpha^{-1} D^{-1} B^{-1} D^{-1})^{-1} = \alpha D(\alpha D^2 + T)^{-1} D,$$

where the first equality follows by the Woodbury formula (Golub and Van Loan, 1996) since C is non-singular. Clearly the matrix αD^2 is diagonal and positive definite. The sum of positive definite matrices is positive definite, so $(\alpha D^2 + T)$ is a symmetric positive definite tridiagonal matrix with non-positive off-diagonal elements. Lemma 1 in Appendix implies that all elements of $(\alpha D^2 + T)^{-1}$ are non-negative, and the result follows. □

Remark. *Condition II also requires $\|\tilde{C}\| < 1$ for the α-permanental random field to be well-defined. However, from* (3.6) *this is clearly true in the present example where C is a covariance matrix and hence has non-negative real eigenvalues.*

Figure 3.1 is inspired by a dataset that fits into this setup (counts of clover leaves in 200 squares of size 5×5 cm along a 10 m transect line, see Augustin et al. (2006)), where the data can be viewed as a one-dimensional random field consisting of 200 sites on the real line with positive association expected between the counts due to clustering of clovers in patches. Figure 3.1 shows four different simulated datasets of this type using different values of α and ρ. Since no covariates are available the only other parameter in the model is β_0, which controls the mean value $\mathrm{E}N_1 = \cdots = \mathrm{E}N_{200} = \exp(\beta_0)$ (as shown in Section 3.3.2 the mean is given by the diagonal elements of C). For different values of (α, ρ), permanental random fields were simulated using the Poisson randomization described in Section 3.4.2, where $\beta_0 = \log(1.28)$ is fixed so that the random fields have the same mean as the data from Augustin et al. (2006).

3.3 Properties of α-permanental random fields

This section reviews various useful properties of α-permanental random fields. We will need various matrices formed from C and $\tilde{C}$. We introduce these for C in the following, while the

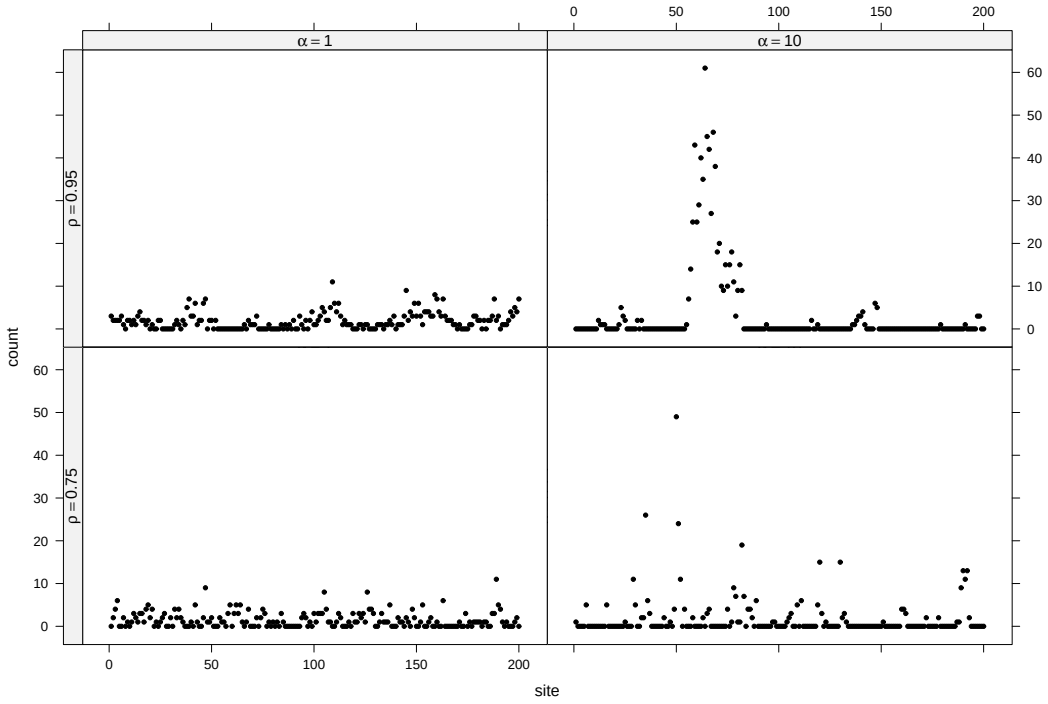


Figure 3.1: Realizations of the random field of Example III for different values of α and ρ.

analogous definitions for $\tilde{C}$ simply are obtained by replacing C with $\tilde{C}$. For any multi-set $T = \{t_1, \ldots, t_n\}$, $t_i \in S$ we let C_T denote the $n \times n$ matrix with (i, j)'th entry $C(t_i, t_j)$. If T is of the special form

$$T = \{\underbrace{s_1, \ldots, s_1}_{n_{s_1}}, \ldots, \underbrace{s_i, \ldots, s_i}_{n_{s_i}}, \ldots, \underbrace{s_m, \ldots, s_m}_{n_{s_m}}\},$$

for non-negative integers $\boldsymbol{n} = (n_s; s \in S)$ with $n_\star = \sum_{s \in S} n_s > 0$ we also write $C_T = C[\boldsymbol{n}]$, and we define $\det_\alpha C[\mathbf{0}] = 1$.

3.3.1 Relation to the negative binomial distribution

Let $\boldsymbol{N} \sim \mathrm{per}(\alpha, C)$. From the form of (3.2) it is clear that for any $S' \subset S$ the subfield $\boldsymbol{N}_{S'} = (N_s; s \in S')$ is also a α-permanental random field; $\boldsymbol{N}_{S'} \sim \mathrm{per}(\alpha, C_{S'})$. Particularly, the probability generating functions of the one dimensional marginals N_s, $s \in S$ are of the form $(1+\alpha(1-z)C(s,s))^{-1/\alpha}$. Hence, if $b^-(\kappa, \pi)$ denotes the negative binomial distribution with

parameters $\kappa > 0$ and $0 \le \pi < 1$, and probability density function

$$\frac{\Gamma(n+\kappa)}{n!\Gamma(\kappa)}\pi^n(1-\pi)^\kappa, \quad n = 0, 1, \ldots,$$

we see that

$$N_s \sim b^-\left(\frac{1}{\alpha}, \frac{\alpha C(s,s)}{1+\alpha C(s,s)}\right). \tag{3.10}$$

Consider the sum $N_\star = \sum_{s\in S} N_s$. By (3.2), the probability generating function of $N_\star$ is

$$\varphi_\star(z) = |I + \alpha(1-z)C|^{-1/\alpha}. \tag{3.11}$$

Rewriting in terms of the eigenvalues of C, (3.11) yields

$$\begin{aligned}\varphi_\star(z) = &\prod_{i:\,\lambda_i\in\mathbb{R}} (1+\alpha(1-z)\lambda_i)^{-1/\alpha} \\ &\times \prod_{i:\,\lambda_i\notin\mathbb{R}} \left(1 + 2\alpha(1-z)\mathrm{Re}(\lambda_i) + \alpha^2(1-z)^2\|\lambda_i\|^2\right)^{-1/(2\alpha)}.\end{aligned} \tag{3.12}$$

Hence, if $1/\alpha$ is an integer, the distribution of $N_\star$ is of matrix geometrical form, see Asmussen and O'Cinneide (1998) and the references therein. If C only has real eigenvalues $\lambda_i \ge 0$, $i = 1, \ldots, m$, then (3.12) implies that

$$N_\star \sim b^-\left(\frac{1}{\alpha}, \frac{\alpha\lambda_1}{1+\alpha\lambda_1}\right) \star \ldots \star b^-\left(\frac{1}{\alpha}, \frac{\alpha\lambda_m}{1+\alpha\lambda_m}\right). \tag{3.13}$$

A well-known property for 'zero-states' of the negative binomial distribution can be generalized as follows concerning the probability

$$\varphi_\star(0) = \mathrm{P}(N_s = 0 \text{ for all } s \in S).$$

From (3.12) follows that

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\log\varphi_\star(0) = \frac{2m_1+m_2}{2\alpha^2} - \frac{1}{2}\sum_{i:\,\lambda_i\notin\mathbb{R}}\|\lambda_i\|^2$$

where m_1 respective m_2 denote the number of real respective non-real eigenvalues λ_i, $i = 1, \ldots, m$. Thus, if C has only real eigenvalues, $\varphi_\star(0)$ is an increasing function of α, and $\varphi_\star(0) \to 1$ for $\alpha \to \infty$.

3.3.2 Moments

For non-negative integers a and b, let $a^{(0)} = 1$, and $a^{(b)} = a!/(a-b)! = a(a-1)\cdots(a-b+1)$ if $a \ge b > 0$. The factorial moments are given by

$$\mathrm{E}\prod_{s\in S} N_s^{(n_s)} = \det{}_\alpha C[\boldsymbol{n}] \tag{3.14}$$

for non-negative integers $(n_s; s \in S)$. This can be obtained by expanding out the powers of $(z_s - 1)$ in (3.2), cf. Vere-Jones (1997) and Shirai and Takahashi (2003a). Note that (3.14) implies that (α, C) is such that $\det_\alpha C[\boldsymbol{n}] \geq 0$ for all non-negative integers $(n_s; s \in S)$.

In general, only the lower dimensional moments are computationally tractable. The first and second order moments are given by

$$\mathrm{E}N_s = C(s,s), \quad \mathrm{Var}N_s = C(s,s) + \alpha C(s,s)^2, \quad \mathrm{Cov}(N_s, N_t) = \alpha C(s,t)C(t,s), \quad \text{if } s \neq t. \tag{3.15}$$

By (3.15) it is clear that

$$N_s = 0 \text{ (almost surely)} \qquad \text{if and only if} \qquad C(s,s) = 0. \tag{3.16}$$

If $C(s,s) > 0$, we obtain from (3.15) the well-known property of the negative binomial distribution that N_s is over-dispersed. Moreover, (3.15) implies that $\mathrm{Cov}(N_s, N_t) \geq 0$, cf. Vere-Jones (1997). Note that if C is symmetric and non-negative, there is a one-to-one correspondence between (α, C) and the moments given by (3.15).

If C is a covariance function, consider its correlation function

$$R(s,t) = C(s,t)/\left[C(s,s)C(t,t)\right]^{1/2}, \quad s,t \in S \tag{3.17}$$

where we take $R(s,t) = 0$ if $C(s,s) = 0$ or $C(t,t) = 0$. Then by (3.15), the correlation between N_s and N_t is

$$\mathrm{Corr}(N_s, N_t) = \alpha R(s,t)^2 \left[\frac{C(s,s)C(t,t)}{(1+\alpha C(s,s))(1+\alpha C(t,t))}\right]^{1/2}, \quad s,t \in S. \tag{3.18}$$

The right hand side in (3.18) is an increasing function of α, and it tends to $R(s,t)^2$ as $\alpha \to \infty$.

3.3.3 Probability density function

The probability density function of an α-permanental random field can be expressed using α-determinants of $\tilde{C}$ as follows, see Vere-Jones (1997). For any non-negative integers $\boldsymbol{n} = (n_s; s \in S)$ with $n_\star = \sum_{s\in S} n_s$,

$$\mathrm{P}(\boldsymbol{N} = \boldsymbol{n}) = |I - \tilde{C}|^{1/\alpha} \alpha^{-n_\star} \det_\alpha \tilde{C}[\boldsymbol{n}] \Big/ \prod_{s\in S} n_s!\,. \tag{3.19}$$

This can be obtained by expanding out the powers of z_s in (3.4).

As described in Section 3.3.1 the marginal distribution of any N_s and possibly also of the 'margin' $N_\star$ are related to the negative binomial distribution. However, even the joint distribution of two random variables N_s and N_t is in general complicated, cf. the discussion in Griffiths and Milne (1987).

3.3.4 Independence

Independence properties of infinite divisible α-permanental random fields have been studied in Griffiths and Milne (1987), and their results are summarized here with slight generalizations.

Suppose that $S = T \cup U$ where T and U are disjoint and non-empty. Recall that the subfields $\boldsymbol{N}_T$ and $\boldsymbol{N}_U$ are independent if and only if the probability generating function $\varphi(z_s; s \in S)$ of $\boldsymbol{N}$ is a product of two functions, one of $(z_s; s \in T)$ and one of $(z_s; s \in U)$.

It follows immediately from (3.2) that $\boldsymbol{N}_T$ and $\boldsymbol{N}_U$ are independent

$$\text{if } C(t,u) = C(u,t) = 0 \text{ whenever } t \in T \text{ and } u \in U. \tag{3.20}$$

If C is symmetric, then by (3.15), $\mathrm{Cov}(N_t, N_u) = \alpha C(t,u)^2$, and so $\boldsymbol{N}_T$ and $\boldsymbol{N}_U$ are independent

$$\text{if and only if } C(t,u) = 0 \text{ whenever } t \in \mathrm{T} \text{ and } u \in U. \tag{3.21}$$

The property of C in (3.20)-(3.21) means that if we order the elements in S so that the elements of T come before those of U, then C restricted to $T \cup U$ is block-diagonal with respect to the partition given by T and U. If C is not symmetric, it is possible that $\mathrm{Cov}(N_s, N_t) = C(t,u)C(u,t)$ is zero even if N_t and N_u are not independent, and we can not in general replace 'if' in (3.20) by 'if and only if'.

Furthermore, we can replace C by $\tilde{C}$ everywhere in (3.20)-(3.21). This follows by similar arguments as above but using (3.4). In addition, assume that the eigenvalues of $\tilde{C}$ are bounded strictly in modulus by one, and define a directed graph $G(\tilde{C})$ with vertex set S and edges $\langle s_i, s_j \rangle$ if $s_i \neq s_j$ and $\tilde{C}(s_i, s_j) \neq 0$. Then $\boldsymbol{N}_T$ and $\boldsymbol{N}_U$ are independent

$$\begin{aligned}&\text{if and only if every directed circuit in } G(\tilde{C}) \text{ contains vertices of either } T \text{ or } U,\\ &\text{but not both.}\end{aligned} \tag{3.22}$$

This follows by combining (3.4) and (3.7), using similar arguments as in the proof of Theorem 3 in Griffiths and Milne (1987).

3.3.5 Thinning

Let $0 \leq \pi_s \leq 1$, $s \in S$, be given numbers, and consider a random field $\boldsymbol{N}^{\mathrm{th}} = (N_s^{\mathrm{th}}; s \in S)$ so that conditional on $\boldsymbol{N}$, the N_s^{th} are mutually independent and $N_s^{\mathrm{th}} \sim b(N_s, \pi_s)$. We say that $\boldsymbol{N}^{\mathrm{th}}$ is obtained by an independent thinning of $\boldsymbol{N}$ with retention probabilities π_s, $s \in S$. Define

$$C_{s,t}^{\mathrm{th}} = \sqrt{\pi_s \pi_t} C_{s,t}, \quad s,t \in S. \tag{3.23}$$

It follows immediately from (3.2) that

$$\boldsymbol{N}^{\mathrm{th}} \sim \mathrm{per}(\alpha, C^{\mathrm{th}}). \tag{3.24}$$

Suppose that C is a covariance matrix. Then C^{th} given by (3.23) is also a covariance matrix, and $\boldsymbol{N}$ and $\boldsymbol{N}^{\mathrm{th}}$ share the same correlation matrix R given by (3.17). By (3.18) we have $0 \leq \mathrm{Corr}(N_s^{\mathrm{th}}, N_t^{\mathrm{th}}) \leq \mathrm{Corr}(N_s, N_t)$, where $\mathrm{Corr}(N_s^{\mathrm{th}}, N_t^{\mathrm{th}})$ is an increasing function of π_s and of π_t.

3.3.6 Convolution

By (3.2), for any $\alpha_1 > 0$ and $\alpha_2 > 0$,

$$\mathrm{per}\left(\alpha_1, \frac{\alpha_2}{\alpha_1+\alpha_2}C\right) \star \mathrm{per}\left(\alpha_2, \frac{\alpha_1}{\alpha_1+\alpha_2}C\right) = \mathrm{per}\left(\left(\frac{1}{\alpha_1}+\frac{1}{\alpha_2}\right)^{-1}, C\right)$$

provided of course that the two first α-permanental random fields exist. In particular,

$$\mathrm{per}(\alpha, C) = \mathrm{per}(\alpha n, C/n)^{\star n}$$

for any $n \in \mathbb{N}$ such that $\mathrm{per}(\alpha n, C/n)$ exists, where $\star n$ denotes convolution n times.

3.3.7 Examples

Example I (continued)

Let the situation be as in Section 3.2.3. Since C has r non-zero eigenvalues which are all equal to κ, (3.13) reduces to

$$N_\star \sim b^-\left(r/\alpha, \alpha\kappa/(1+\alpha\kappa)\right).$$

Further, $\tilde{C} = (\alpha\kappa/(1+\alpha\kappa))Q$, and we obtain from (3.14) and (3.19) that the expressions for the factorial moments and the probability density function are closely related, since

$$\mathrm{E}\prod_{s\in S} N_s^{(n_s)} = \kappa^{n_\star}\det{}_\alpha Q[\boldsymbol{n}], \quad \mathrm{P}(\boldsymbol{N}=\boldsymbol{n}) = \frac{\kappa^{n_\star}\det{}_\alpha Q[\boldsymbol{n}]}{(1+\alpha\kappa)^{n_\star + r/\alpha}\prod_{s\in S} n_s!} \tag{3.25}$$

where $n_\star = \sum_{s\in S} n_s$.

Example II (continued)

Let the situation be as in Section 3.2.3. From (3.13) we have

$$N_\star \sim b^-\left(\frac{1}{\alpha}, \frac{\alpha\kappa}{1+\alpha\kappa}\right). \tag{3.26}$$

By differentiation of the probability generating function it is straightforward to find the probability of $\boldsymbol{N} = \boldsymbol{n}$ for any vector of non-negative integers $\boldsymbol{n} = (n_1, \ldots, n_m)$ with $\sum_{i=1}^m n_i = n_\star$

$$p(\boldsymbol{n}) = \frac{\Gamma(\frac{1}{\alpha}+n_\star)}{\Gamma(\frac{1}{\alpha})}\left(1+\alpha\kappa\right)^{-n_\star-\frac{1}{\alpha}}\prod_{i=1}^m \frac{c_i^{n_i}}{n_i!}.$$

Combining this with (3.26) yields

$$p(\boldsymbol{n}|n_\star) = n_\star!\prod_{i=1}^m \frac{1}{n_i!}\left(\frac{c_i}{\kappa}\right)^{n_i} \tag{3.27}$$

such that $\boldsymbol{N}|n_\star$ is multinomial with event probabilities $\frac{c_1}{\kappa}, \ldots, \frac{c_m}{\kappa}$.

In this setup the random field is parametrized by the mean $(\mathrm{E}N_1, \ldots, \mathrm{E}N_m) = (c_1, \ldots, c_m)$, and using the fact that $N_\star$ follows a negative binomial distribution and that $\boldsymbol{N}|N_\star$ is multinomial makes a two step simulation scheme straightforward. The correlation between N_i and N_j is

$$\mathrm{Corr}(N_i, N_j) = \sqrt{\frac{c_i}{1/\alpha + c_i} \frac{c_j}{1/\alpha + c_j}}$$

so sites with a large mean is more strongly correlated to all other sites than a site with a smaller mean. If $\boldsymbol{N}$ is homogeneous in the sense that $\mathrm{c}_1 = \cdots = c_m = c$ the correlation between the counts at any two sites is $\mathrm{Corr}(N_i, N_j) = \alpha c/(1 + \alpha c)$. Furthermore, as is the case for α-permanental random fields in general, correlation grows with α as well.

Figure 3.2 shows four realizations of such a homogeneous random field with $c = 100$ and $\alpha = 1$. The figure exemplifies how the correlation in this model effectively results in very little variation within a realization of the random field compared to the large variation between realizations. Based on 1000 simulations the average of the empirical variance within each realization was 15.4 compared to the marginal variance $\mathrm{Var}(N_i) = 110$, $i = 1, \ldots, 2500$. While this model is mathematically tractable it seems to be of less interest in applications due to low flexibility, and in spatial applications the model is unaffected by usual neighborhood relations based on distances since correlation structures only depend on the mean values at any given given pair of sites.

3.4 Stochastic constructions and simulation

In this section we discuss stochastic constructions and simulation algorithms for the α-permanental random field $\boldsymbol{N}$. To exclude the trivial case where $N_s = 0$ for all $s \in S$, we assume that C has rank $r > 0$. Furthermore, we assume $m > 1$, since $\boldsymbol{N}$ just follows a negative binomial distribution if $m = 1$.

3.4.1 Doubly stochastic construction

Assume that $\boldsymbol{G} = (G_s; s \in S)$ is a random field of non-negative real random variables with Laplace transform (or moment generating function) of the form

$$\mathrm{E} \exp\left(\sum_{s \in S} G_s z_s \right) = |I - \alpha Z C|^{-1/\alpha} \tag{3.28}$$

for $z_s \in [-1, 1]$, $s \in S$, where Z is the diagonal matrix with diagonal $(z_s; s \in S)$. This is a multivariate extension of the gamma distribution, where all one-dimensional marginals are gamma-distributed, but it is an open question to establish necessary and sufficient conditions on (α, C) for (3.28) to be a Laplace transform of some distribution on $[0, \infty)^m$, see Krishnamoorthy and Parthasarathy (1951) and Vere-Jones (1997). Suppose that $\boldsymbol{N}$ conditioned on

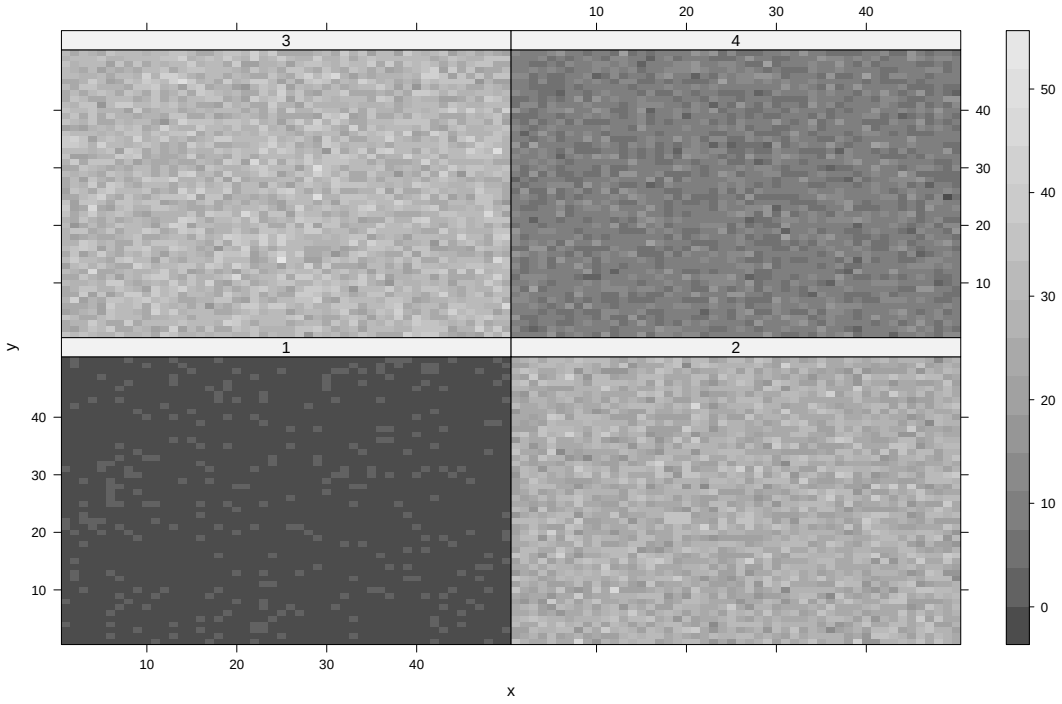


Figure 3.2: Four independent realizations of the random field of Example II on a 50×50 grid with $c_1 = \cdots = c_{2500} = 10$ and $\alpha = 1$.

$\boldsymbol{G}$ consists of mutually independent Poisson random variables N_s with mean G_s, $s \in S$. It is immediately verified that (3.2) is satisfied, so $\boldsymbol{N} \sim \text{per}(\alpha, C)$, cf. Vere-Jones (1997).

By this doubly stochastic construction, if we can generate $\boldsymbol{G}$, we can straightforwardly generate $\boldsymbol{N}$. Below two different constructions of $\boldsymbol{G}$ are described.

Method I: Assume Condition I (Section 3.2.2) is satisfied. Generate a $m \times m$ Wishart matrix K with $2/\alpha$ degrees of freedom and mean C. If $G_{s_i} = K_{i,i}$ then $\boldsymbol{G}$ has moment generating function (3.28). Simulation of Wishart distributed matrices is described in e.g. Johnson (1987).

Method II: Assume Condition I is satisfied and $\alpha = 2/k$ for some $k \in \mathbb{N}$. Generate independent zero-mean Gaussian random fields $\mathbf{Y}_1 = (Y_{1,s}; s \in S), \ldots, \mathbf{Y}_k = (Y_{k,s}; s \in S)$ with covariance function C/k. If $G_s = Y_{1,s}^2 + \cdots + Y_{k,s}^2$, $s \in S$, then $\boldsymbol{G}$ has moment generating function (3.28). Various simulation methods for Gaussian random fields are implemented in the `R` package `RandomFields` by Martin Schlather. See also Lantuejoul (2002), and the references therein.

Method I corresponds to the extension given in Shirai (2007), and the simpler Method II has also been considered in Vere-Jones (1997).

3.4.2 Poisson randomization

In the sequel, it seems more natural to work with $\tilde{C}$ rather than C, where we assume that Condition II (Section 3.2.2) is satisfied. The assumption that $\tilde{C} \geq 0$ ensures that the right hand sides in the density expressions (3.29)-(3.30) and (3.32) below are non-negative. The α-permanental field $\boldsymbol{N}$ can then be constructed by the following five steps of a Poisson randomization (a similar construction for spatial point processes was introduced in McCullagh and Møller (2006)).

1. For $n \in \mathbb{N}$, define a probability density function by

$$p_n(t_1, \ldots, t_n) = \frac{1}{\operatorname{tr}(\tilde{C}^n)} \prod_{i=1}^{n} \tilde{C}(t_i, t_{i+1}), \quad (t_1, \ldots, t_n) \in S^n, \tag{3.29}$$

   where $t_{n+1} = t_1$. Using the Schur decomposition of $\tilde{C}$ (Golub and Van Loan, 1996), we obtain the normalizing constant $\operatorname{tr}(\tilde{C}^n) = \sum_{i=1}^{m} \tilde{\lambda}_i^n$ of this density. It can be viewed as a Markov random field defined on the graph with vertices $1, \ldots, n$ and edges $\langle i, i+1 \rangle$, $i = 1, \ldots, n$, with the turn-around edge $\langle n, n+1 \rangle = \{n, 1\}$. It reduces to the "Ising model on the ring" if $S = \{s, t\}$ and $\tilde{C}(s, s) = \tilde{C}(t, t)$.

2. Define a random variable W with probability density function

$$p_W(n) = \frac{\operatorname{tr}(\tilde{C}^n)}{Dn}, \quad n \in \mathbb{N}, \tag{3.30}$$

   where

$$D = -\log |I - \tilde{C}|. \tag{3.31}$$

   If the eigenvalues of $\tilde{C}$ are real with $0 \leq \tilde{\lambda}_i < 1$, then $D = -\sum_{i=1}^{m} \log(1 - \tilde{\lambda}_i)$ and W follows a mixture of logarithmic distributions with parameters $\tilde{\lambda}_i$, $i = 1, \ldots, m$, where the i'th logarithmic distribution has weight $-\log(1 - \tilde{\lambda}_i)/D$ in the mixture distribution.

3. Consider an ordered point process $(R_1, \ldots, R_W)$, where conditioned on $W = n$, $(R_1, \ldots, R_n)$ follows (3.29). Thus $(R_1, \ldots, R_W)$ takes values in the countable set $\cup_{n=1}^{\infty} S^n$, and its probability density function $p(t_1, \ldots, t_n) = p_W(n) p_n(t_1, \ldots, t_n)$ is

$$p(t_1, \ldots, t_n) = \frac{1}{nD} \prod_{i=1}^{n} \tilde{C}(t_i, t_{i+1}), \quad (t_1, \ldots, t_n) \in S^n, \ n \geq 1. \tag{3.32}$$

   Moreover, define a random field $\boldsymbol{M} = (M_s; s \in S)$ with $M_s = \sum_{j=1}^{W} \mathbb{I}[R_j = s]$. We call $\boldsymbol{M}$ a cluster and each R_i, $i = 1, \ldots, W$, a point of the cluster, i.e. M_s counts how many points in the cluster are equal to s.

4. Let V be a Poisson random variable with mean D/α, and conditioned on $V = n$, if $n > 0$, let $\boldsymbol{M}^{(1)}, \ldots, \boldsymbol{M}^{(n)}$ be mutually independent copies of $\boldsymbol{M}$. These clusters are generated by corresponding mutually independent ordered point processes $(R_1^{(1)}, \ldots, R_{W_1}^{(1)})$, $(R_1^{(2)}, \ldots, R_{W_2}^{(2)})$, ..., which are independent of V.
5. The Poisson randomization is given by the random field $\boldsymbol{N} = (N_s; s \in S)$ with

$$N_s = \sum_{i=1}^{V} M_s^{(i)}$$

counting how many points in all the V clusters are equal to s (setting $N_s = 0$ if $V = 0$).

The validity of this Poisson randomization is stated and proven below.

Proposition 2. Let Condition II be satisfied. Then the random field $\boldsymbol{N}$ given by the Poisson randomization 1.–5. has a probability generating function of the form (3.4), i.e. $\boldsymbol{N} \sim \mathrm{per}(\alpha, C)$.

Proof. The proof in McCullagh and Møller (2006) of the validation of the Poisson randomization is based on density calculations. Below we give an alternative, short, and simple proof based on the probability generating function.

Let $z_s \in [-1, 1]$, $s \in S$. By the construction of $\boldsymbol{N}$ in the Poisson randomization, and by first conditioning on V, and next using that V is Poisson distributed with mean D/α, we obtain

$$\mathrm{E} \prod_{s \in S} z_s^{N_s} = \mathrm{E}\Big[\Big(\mathrm{E} \prod_{s \in S} z_s^{M_s}\Big)^V\Big] = \exp\Big[\frac{D}{\alpha}\Big(\mathrm{E} \prod_{s \in S} z_s^{M_s} - 1\Big)\Big]. \tag{3.33}$$

By the construction of $\mathbf{M}$ and (3.32),

$$\begin{aligned} \mathrm{E} \prod_{s \in S} z_s^{M_s} &= \sum_{n=1}^{\infty} \sum_{(t_1,\ldots,t_n) \in S^n} \prod_{s \in S} z_s^{\sum_{j=1}^n \mathbb{I}[t_j = s]} p(t_1, \ldots, t_n) \\ &= \frac{1}{D} \sum_{n=1}^{\infty} \sum_{(t_1,\ldots,t_n) \in S^n} \frac{1}{n} \prod_{j=1}^{n} z_{t_j} \tilde{C}(t_j, t_{j+1}) \\ &= \frac{1}{D}(-\log|I - Z\tilde{C}|) \end{aligned} \tag{3.34}$$

where the last identity follows from (3.7). Combining (3.31) and (3.33)-(3.34) yields

$$\mathrm{E} \prod_{s \in S} z_s^{N_s} = \exp\Big[\frac{1}{\alpha}\Big(-\log|I - Z\tilde{C}| + \log|I - \tilde{C}|\Big)\Big] = \Big(|I - \tilde{C}|/|I - Z\tilde{C}|\Big)^{1/\alpha}$$

which agrees with the probability generating function (3.2). □

Incidentally, if $C' = \alpha C$ is fixed, then $\boldsymbol{N}|(N_{S\star} > 0)$ can be seen to converge in distribution to $\boldsymbol{M}$ as $\alpha \to \infty$, cf. McCullagh and Møller (2006).

Remark. *The requirement of Condition II to be satisfied can be replaced by only requiring the permanental random field to be infinitely divisible (which is implied by Condition II). Infinite divisibility has been characterized by Griffiths and Milne (1987). It implies both* $\|\tilde{C}\| < 1$ *and that all cyclic products formed using* $\tilde{C}$ *are non-negative. The latter property ensures the density* (3.32) *is well-defined.*

3.4.3 Simulation of the Poisson randomization

Let the situation be as in Section 3.4.2. Simulation of a realization from the Poisson randomization is straightforward if we know how to make a simulation of a cluster as given in steps 1.-2. This can be done by first generating a realization $W = n$ from (3.30), and then use the following sequential simulation scheme. From (3.29) follows by induction that for any $n \in \mathbb{N}$,

$$p_{n-i}(t_1, \ldots, t_{n-i}) = \frac{1}{\mathrm{tr}(\tilde{C}^n)} \tilde{C}^{i+1}(t_{n-i}, t_1) \prod_{j=1}^{n-i-1} \tilde{C}(t_j, t_{j+1}), \quad i = 0, 1, \ldots, n-1,$$

where we set $\prod_{j=1}^{n-i-1} \cdots = 1$ if $i = n-1$. Hence, first we draw t_1 from the probability density function

$$p_1(t_1) \propto \tilde{C}^n(t_1, t_1)$$

and second, successively for $i = 2, \ldots, n$, since $t_i|(t_1, \ldots, t_{i-1}) \sim t_i|(t_1, t_{i-1})$, we draw t_i from the conditional probability density function

$$p_{i|1,i-1}(t_i|t_1, t_{i-1}) \propto \tilde{C}^{n-i+1}(t_i, t_1)\tilde{C}(t_{i-1}, t_i).$$

3.4.4 Examples

Example I (continued)

Let the situation be as in Section 3.2.3. If Q has non-negative entries, the procedure for simulation of a cluster (Section 3.4.3) simplifies, since $\tilde{C}^i = (\alpha\kappa/(1+\alpha\kappa))^i Q$ for any $i \in \mathbb{N}$, and the conditional probability density functions

$$p_{i|1,i-1}(t_i|t_1, t_{i-1}) \propto Q(t_i, t_1)Q(t_{i-1}, t_i), \quad i = 2, \ldots, n,$$

are of the same form.

Example III (continued)

Let the situation be as in Section 3.2.3. Table 3.1 summarizes some characteristics for each of the simulated models in Figure 3.1. Here the correlation between neighboring sites is straightforward to calculate, and for the real data the empirical estimate is reported. Further, V is the number of clusters in a simulation, and from both its mean and its four simulated values

Table 3.1: Parameter values and characteristics of the four simulated random fields considered in Example III. The bottom two rows are observed quantities for the specific simulation whereas the other values are calculated theoretically. The right most column shows the empirical mean and lag 1 autocorrelation of one of the real data sets from Augustin et al. (2006).

Simulation	1	2	3	4	Real data
(α, ρ)	$(1, 0.75)$	$(1, 0.95)$	$(10, 0.75)$	$(10, 0.95)$	-
$\mathrm{E}(N_s)$	1.28	1.28	1.28	1.28	1.28
$\mathrm{Corr}(N_{s_i}, N_{s_{i+1}})$	0.316	0.507	0.522	0.837	0.508
$\mathrm{E}(V)$	119	63	39	21	-
$P(W = 1)$	0.627	0.563	0.408	0.475	-
$P(W \leq 2)$	0.793	0.706	0.575	0.623	-
$P(W \leq 10)$	0.980	0.919	0.869	0.849	-
$P(W \leq 100)$	1.000	0.999	0.994	0.975	-
V	130	62	45	20	-
$\overline{W}$	2.29	5.85	8.18	34.9	-

it is clear that realizations of V tend to be higher for smaller values of α and ρ. On the other hand, realizations of W, which denotes the size of a cluster, tends to be larger for larger values of α and ρ. This gives an intuitive understanding of how the dependence structure is created in the Poisson randomization: Large values of α lead to a small number of very large clusters, and large values of ρ makes the correlation within the cluster high, such that a few close sites are sampled many times in a cluster. Simulations 1 and 2 ($\alpha = 1$) were also done using the double stochastic construction of Section 3.4.1 to compare simulation time of the two algorithms. In the Poisson randomization the most computer intensive part is calculating all the necessary powers of $\tilde{C}$ used both to simulate the cluster length W and in the simulation of a cluster, cf. Sections 3.4.2-3.4.3. After this initialization repeated simulations of the random field are faster and $1,000$ simulations only take about 20 times longer to generate as the first simulation alone. It is however much faster to use the double stochastic scheme, which for $1,000$ simulations took only $1/30$ of the corresponding simulation time for the Poisson randomization.

Acknowledgments

Supported by the Danish Natural Science Research Council, grant 272-06-0442, 'Point process modelling and statistical inference', and by the Danish Agency for Science, Technology and Innovation, grant 645-06-0528, 'International Ph.D.-student'. We are grateful to Professor David Vere-Jones for helpful comments.

Bibliography

Asmussen, S. and C. A. O'Cinneide (1998). Matrix-exponential distributions. In S. Kotz, C. Read, and D. Banks (Eds.), *Encyclopedia of Statistical Sciences, Supplementary Volume*, New York, pp. 435–440. John Wiley & Sons.

Augustin, N. H., J. McNicol, and C. A. Marriott (2006). Using the truncated auto-Poisson model for spatially correlated counts of vegetation. *J. Agric. Biol. Environ. Stat. 11*, 1–23.

Cressie, N. (1993). *Statistics for Spatial Data* (second ed.). New York: John Wiley and Sons.

Georgii, H.-O. and H. J. Yoo (2005). Conditional intensity and Gibbsianness of determinantal point processes. *J. Stat. Phys. 118*, 617–666.

Golub, G. H. and C. F. Van Loan (1996). *Matrix Computations*. Baltimore: Johns Hopkins.

Goulden, I. P. and D. M. Jackson (1983). *Combinatorial Enumeration*. New York: John Wiley & Sons.

Griffiths, R. C. (1984). Characterization of infinitely divisible multivariate gamma distributions. *J. Multivariate Anal. 15*, 13–20.

Griffiths, R. C. and R. K. Milne (1987). A class of infinitely divisible multivariate negative binomial distributions. *J. Multivariate Anal. 22*, 13–23.

Hough, J. B., M. Krishnapur, Y. Peres, and B. Viràg (2006). Determinantal processes and independence. *Probab. Surv. 3*, 206–229.

Johnson, M. E. (1987). *Multivariate Statistical Simulation*. New York: John Wiley & Sons, New York.

Karlin, S. (1968). *Total Positivity*. Stanford: Stanford University Press.

Kou, S. C. and P. McCullagh (2009). Approximating the α-permanent. *Biometrika 96*, 635–644.

Krishnamoorthy, A. S. and M. Parthasarathy (1951). A multivariate gamma-type distribution. *Ann. Math. Statist. 22*, 549–557.

Lantuejoul, C. (2002). *Geostatistical Simulation: Models and Algorithms*. Berlin: Springer-Verlag.

Macchi, O. (1971). Stochastic point processes and multicoincidences. *IEEE Trans. Inform. Theory 17*, 2–7.

Macchi, O. (1975). The coincidence approach to stochastic point processes. *Adv. in Appl. Probab. 7*, 83–122.

McCullagh, P. and J. Møller (2006). The permanental process. *Adv. in Appl. Probab. 38*, 873–888.

Shirai, T. (2007). Remarks on the positivity of α-determinants. *Kyushu J. Math. 61*, 169–189.

Shirai, T. and Y. Takahashi (2003a). Random point fields associated with certain Fredholm determinants I: fermion, Poisson and boson point processes. *J. Func. Anal. 205*, 414–463.

Shirai, T. and Y. Takahashi (2003b). Random point fields associated with certain Fredholm determinants II: fermion shifts and their ergodic and Gibbs properties. *Ann. Probab. 31*, 1533–1564.

Valiant, L. G. (1979). The complexity of cumputing the permanent. *Theoret. Comput. Sci. 8*(2), 189–201.

Vere-Jones, D. (1997). Alpha-permanents and their applications to multivariate gamma, negative binomial and ordinary binomial distributions. *New Zealand J. Math. 26*, 125–149.

3.A Green's matrices and tridiagonal matrices

We will need some results on Green's matrices and tridiagonal matrices (sometimes called Jacobi matrices). The results presented here are either from Karlin (1968) or direct consequences of results herein.

A Green's matrix is a symmetric $n \times n$ matrix G with $G_{ij} = a_{\min(i,j)} b_{\max(i,j)}$ for some $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{R}$. If G is invertible, then it is a Green's matrix if and only if the inverse $T = G^{-1}$ is symmetric and tridiagonal.

For any $n \times n$ matrix A and any $\{i_1, \ldots, i_m\} \subseteq \{1, \ldots, n\}$ we introduce the minor of A, $m_A(i_1, \ldots, i_m)$, as the determinant of the matrix obtained from A by deleting all other rows and columns than $i_1, \ldots, i_m$. If a symmetric tridiagonal matrix T is positive definite, any minor of T is positive.

The (i, j)'th element of the inverse T^{-1} is given as the following (due to symmetry we only need to specify the elements with $i \leq j$). If $i = j$, then

$$T_{i,i}^{-1} = \frac{1}{|T|} m_T(1, \ldots, i-1, i+1, \ldots, n).$$

If $i < j$, then

$$T_{i,j}^{-1} = \frac{(-1)^{j+i}}{|T|} m_T(1, \ldots, i-1) T_{i,i+1} T_{i+1,i+2} \cdots T_{j-1,j} m_T(j+1, \ldots, n).$$

Consequently, a sufficient condition for all elements of T^{-1} to be non-negative is that the off-diagonal elements are non-positive and T is positive definite. This result is summarized in the following lemma.

Lemma 1. Let T be a symmetric tridiagonal matrix. If T is positive definite and $T_{i,j} \leq 0$ for all $i \neq j$, then $T_{i,j}^{-1} \geq 0$ for all i, j.

CHAPTER 4

Statistical Inference for a Class of Multivariate Negative Binomial Distributions

Publication details

Co-authors: Jesper Møller[†], Peter McCullagh[‡]

[†] *Department of Mathematical Sciences, Aalborg University*
[‡] *Department of Statistics, University of Chicago*

Journal: Bernoulli, submitted

Abstract:
This paper considers statistical inference procedures for a class of models for positively correlated count variables called α-permanental random fields, and which can be viewed as a family of multivariate negative binomial distributions. Their appealing probabilistic properties have earlier been studied in the literature, while this is the first statistical paper on α-permanental random fields. The focus is on maximum likelihood estimation, maximum quasi-likelihood estimation and on maximum composite likelihood estimation based on uni- and bivariate distributions. Furthermore, new results for α-permanents and for a bivariate α-permanental random field are presented.

Keywords:
α-permanent; α-permanental random field; composite likelihood; doubly stochastic construction; maximum likelihood; quasi-likelihood.

4.1 Introduction

Møller and Rubak (2010) provided a review of a class of models for positively correlated count variables $\boldsymbol{N} = (N_1, \ldots, N_m)$, which possess a number of appealing properties. This model class was referred to as α-permanental random fields, since it is a special case of the class of general α-permanental point processes which have been the subject of much research interest in recent years, see Macchi (1971, 1975), Shirai and Takahashi (2003a,b), Georgii and Yoo (2005), and McCullagh and Møller (2006). As each count variable N_i follows a negative binomial distribution, an α-permanental random field may be referred to as a multivariate negative binomial distribution. The probabilistic properties of these multivariate distributions have been studied in detail in Griffiths and Milne (1987), Vere-Jones (1997), and Møller and Rubak (2010), but to the best of our knowledge no statistical inference based on the models have been conducted. In this paper we develop statistical inference procedures using the full likelihood, quasi-likelihood or composite likelihoods.

Section 4.2 introduces the notation and provides the necessary background material. Section 4.3 describes the inferential procedures, and Section 4.4 illustrates their use for analyzing two different data sets. Technicalities are deferred to Appendix 4.A, which, among other things, establishes a new result concerning the joint density of any two count variables (N_i, N_j).

4.2 The α-permanental random field

This section contains a very brief introduction to the necessary background material about the α-permanental random field. We mainly follow the notation and terminology of Møller and Rubak (2010), and further details can be found therein.

We start by recalling the definition of the α-permanent of an $n \times n$ matrix A with entries $A_{i,j}$,

$$\mathrm{per}_\alpha(A) = \sum_{\sigma \in \mathcal{S}_n} \alpha^{c(\sigma)} A_{1,\sigma(1)} A_{2,\sigma(2)} \cdots A_{n,\sigma(n)},$$

where $\mathcal{S}_n$ is the set of all permutations of $1,\ldots,n$, and $c(\sigma)$ denotes the number of cycles in σ. In a more general setup, it may be convenient to work with the related α-determinant $\det_\alpha(A) = \alpha^n \mathrm{per}_{1/\alpha}(A)$ as in Møller and Rubak (2010), but it is not necessary here. In general the α-permanent is very expensive computationally, and apart from a few special cases it can only be approximated (see Appendix 4.A for details).

The distribution of an α-permanental random field $\boldsymbol{N} = (N_1,\ldots,N_m)$ is specified by a positive real parameter α and a real $m \times m$ matrix C, and we write $\boldsymbol{N} \sim \mathrm{per}(\alpha, C)$. Throughout this paper we assume that the matrix

$$\tilde{C} = \alpha C(I + \alpha C)^{-1} \tag{4.1}$$

exists. As discussed below, further restrictions need to be satisfied by (α, C) or by $(\alpha, \tilde{C})$ to ensure the existence of the distribution $\mathrm{per}(\alpha, C)$. Then, for

$$\boldsymbol{n} = (n_1,\ldots,n_m) \in \{0,1,\ldots\}^m, \quad n_\star = \sum_{i=1}^m n_i,$$

the probability function is given by

$$p(\boldsymbol{n}) = \frac{|I - \tilde{C}|^{1/\alpha}}{\prod_{i=1}^m n_i!} \mathrm{per}_{1/\alpha}(\tilde{C}[\boldsymbol{n}]), \tag{4.2}$$

where $\tilde{C}[\boldsymbol{n}]$ is the $n_\star \times n_\star$ block matrix obtained from $\tilde{C}$ by repeating the i'th index n_i times (cf. Section 4.A.2 for further details). Marginally each N_i follows a negative binomial distribution with mean $\mathrm{E}N_i = C_{i,i}$ and variance $\mathrm{Var}N_i = C_{i,i} + \alpha C_{i,i}^2$, $i = 1,\ldots,m$. Furthermore, $\mathrm{Cov}(N_i, N_j) = \alpha C_{i,j} C_{j,i} \geq 0$ for $i \neq j$, so all correlations are non-negative. The parameter α influences both the amount of over-dispersion and the strength of correlation between variables. In particular these decrease as α tends to zero and the limiting distribution is Poisson with independent components regardless of the matrix C. No combination of parameters (α, C) exists such that the components of $\boldsymbol{N}$ are Poisson variables with positive correlation. However, over-dispersion without correlation is possible, in which case the components are independent negative binomial variables. In other words, the α-permanental model is such that, if there is correlation among the counts, over-dispersion will also be present. The over-dispersion factor for each N_i is $1 + \alpha \mathrm{E}(N_i)$.

In this paper we mainly consider the case where the following doubly stochastic construction applies: First, let $\boldsymbol{X} = (X_1,\ldots,X_m)$ follow a certain multivariate gamma distribution denoted $\Gamma_m(\alpha, C)$, where Proposition 4.5 in Vere-Jones (1997) gives a sufficient and necessary condition for the existence of this multivariate gamma distribution, but the following sufficient condition (C1) is simpler to use:

(C1) C is a covariance matrix and $\alpha \in \left(0, \frac{2}{m-1}\right] \cup \left\{\frac{2}{m-2}, \frac{2}{m-3}, \ldots, 1, 2\right\}$.

Under (C1), $\boldsymbol{X}$ is distributed as the diagonal of a Wishart matrix with $2/\alpha$ degrees of freedom and mean C, so X_i is gamma distributed with $\mathrm{E}X_i = C_{i,i}$ and $\mathrm{Cov}(X_i, X_j) = \alpha C_{i,j} C_{j,i}$ (Møller and Rubak, 2010, Section 4.1). Second, conditionally on $\boldsymbol{X}$, let the N_i's be independent Poisson random variables with $\mathrm{E}(N_i \mid X_i) = X_i$.

Under the doubly stochastic scheme, for $k = 1, 2, \ldots$ and given an observation of $\boldsymbol{N} = \boldsymbol{n}$, the Bayes estimate of the k'th moment of the unknown mean X_i is

$$\begin{aligned}
\mathrm{E}(X_i^k \mid \boldsymbol{n}) &= \frac{1}{p(\boldsymbol{n})} \int_{\mathbb{R}_+^m} x_i^k p(\boldsymbol{n} \mid \boldsymbol{x}) p(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \\
&= (n_i + 1) \cdots (n_i + k) \frac{1}{p(\boldsymbol{n})} \int_{\mathbb{R}_+^m} \frac{x_i^{n_i+k}}{(n_i+k)!} e^{-x_i} \prod_{j \neq i} \Big[\frac{x_j^{n_j}}{n_j!} e^{-x_j} \Big] p(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \\
&= (n_i + 1) \cdots (n_i + k) \frac{p(\boldsymbol{n}_i^k)}{p(\boldsymbol{n})} \\
&= \frac{\mathrm{per}_{1/\alpha}(\tilde{C}[\boldsymbol{n}_i^k])}{\mathrm{per}_{1/\alpha}(\tilde{C}[\boldsymbol{n}])},
\end{aligned} \tag{4.3}$$

where $\boldsymbol{n} = (n_1, \ldots, n_m)$ and $\boldsymbol{n}_i^k = (n_1, \ldots, n_{i-1}, n_i + k, n_{i+1}, \ldots, n_m)$. Furthermore, if D is a diagonal matrix with diagonal entries $D_{i,i} = \sqrt{a_i}$ where $a_i \geq 0$, then

$$(X_1, \ldots, X_m) \sim \Gamma_m(\alpha, C) \;\Rightarrow\; (a_1 X_1, \ldots, a_m X_m) \sim \Gamma_m(\alpha, DCD). \tag{4.4}$$

As noted in Vere-Jones (1997) the doubly stochastic construction is not necessary for the existence of the α-permanental random field: there are (α, C) such that $\mathrm{per}(\alpha, C)$ exists, but a corresponding gamma random field $\Gamma_m(\alpha, C)$ does not exist. Another sufficient condition for the existence of $\mathrm{per}(\alpha, C)$ is

(C2) $\tilde{C}$ has non-negative entries and all eigenvalues have modulus less than 1

(Vere-Jones (1997); Møller and Rubak (2010)). When (C1) is satisfied, simulation of first $\boldsymbol{X}$ and second $\boldsymbol{N}$ is easily done by the doubly stochastic construction. If (C2) but not (C1) is satisfied, a Poisson randomization can be used for simulation (Møller and Rubak, 2010, Section 4.2).

4.3 Inference

4.3.1 Full likelihood

Given a realization $\boldsymbol{n}$ of an α-permanental random field with a parametric model for the matrix $C = C_\psi$, where ψ is a real d-dimensional parameter, note that $\tilde{C} = \tilde{C}_\theta$ depends on $\theta = (\alpha, \psi)$, cf. (4.1). In principle, we can evaluate the log-likelihood

$$\ell(\alpha, \psi; \boldsymbol{n}) = \frac{1}{\alpha} \log |I - \tilde{C}_\theta| + \log \mathrm{per}_{1/\alpha}(\tilde{C}_\theta[\boldsymbol{n}]) \tag{4.5}$$

on a grid of (α, ψ) in order to obtain the maximum likelihood estimate (MLE) $(\hat{\alpha}, \hat{\psi})$ (provided it exists). Further, for each grid point (α, ψ), we have access to the log-likelihood ratio $\lambda(\alpha, \psi) = 2(\ell(\hat{\alpha}, \hat{\psi}) - \ell(\alpha, \psi))$, which may be compared with quantiles of the χ^2_{d+1} distribution to find approximate confidence regions.

However, as mentioned previously and discussed in Appendix 4.A, exact calculation of the α-permanent is usually not tractable, and in fact even approximate calculation may be computationally expensive. Furthermore, the grid evaluation requires some knowledge of the range of (α, ψ) values to include in the grid. Therefore we study composite likelihoods which both serve as a computationally simple method for inference in its own right and can be used for initializing the grid evaluation of the full likelihood.

4.3.2 Composite likelihood

Composite likelihoods have been extensively studied in many connections, see e.g. Lindsay (1988) and Cox and Reid (2004). Here we outline how composite likelihood methods can be used for the α-permanental random field model, using either the univariate or the bivariate distributions.

Given an observation $\boldsymbol{n}$ and a parametric model as in Section 4.3.1, we define the *first-order composite log-likelihood* by

$$\ell^1(\theta) = \sum_{i=1}^{m} \log p_i(n_i \,|\, \theta), \tag{4.6}$$

where p_i is the marginal probability function for N_i. It corresponds to the log-likelihood for m independent negative binomial random variables. In this case likelihood inference can be done using an iterative Newton-Raphson procedure and efficient software implementations are readily available (Venables and Ripley, 2002, Section 7.4). Depending on the parametric model for C, some parameters may be unidentifiable using this procedure, since only the diagonal elements of C enter in the first-order composite log-likelihood, as exemplified in Section 4.4.1. Due to the computational simplicity of this composite log-likelihood, it is well suited for initialization of the parameters in more complicated methods.

In a similar manner as above, we define the *pairwise composite log-likelihood* by

$$\ell^2(\theta) = \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \log p_{i,j}(n_i, n_j \,|\, \theta), \tag{4.7}$$

where $p_{i,j}$ denotes the bivariate probability function for (N_i, N_j). These bivariate distributions have been thought to be quite complicated, cf. the discussion in Griffiths and Milne (1987), and previously it was not possible to use these bivariate distributions in practice. However, in Appendix 4.A.2 we give a computationally simple formula for calculation of the relevant α-permanent. The resulting bivariate probability function is

$$p_{i,j}(n_i, n_j) = \frac{\left(\frac{a_i}{b}\right)^{n_i} \left(\frac{a_j}{b}\right)^{n_j} \Gamma(\frac{1}{\alpha} + n_i)\Gamma(\frac{1}{\alpha} + n_j)}{b^{\frac{1}{\alpha}} n_i! n_j! \Gamma(\frac{1}{\alpha})\Gamma(\frac{1}{\alpha})} \sum_{k=0}^{n_i \wedge n_j} \binom{n_i}{k}\binom{n_j}{k} \frac{k!\Gamma(\frac{1}{\alpha})}{\Gamma(\frac{1}{\alpha} + k)} c^{2k}, \tag{4.8}$$

where

$$a_i = \alpha^2(C_{i,i}C_{j,j} - C_{i,j}^2) + \alpha C_{i,i}, \quad a_j = \alpha^2(C_{i,i}C_{j,j} - C_{i,j}^2) + \alpha C_{j,j},$$
$$b = \alpha^2(C_{i,i}C_{j,j} - C_{i,j}^2) + \alpha(C_{i,i} + C_{j,j}) + 1, \quad c = \frac{\alpha C_{i,j}}{\sqrt{a_i a_j}}.$$

This makes it practically feasible to implement the pairwise composite log-likelihood for statistical inference.

In many applications there is a distance function or neighbourhood structure attached to the domain, or index set, of the field. For example, when modeling spatial regions some regions will share a boundary and will be called neighbours. In this way there will also be a natural notion of higher order neighbours, such that regions not sharing a boundary but with a common neighbour are second order neighbours etc. The part of the pairwise composite log-likelihood (4.7) corresponding to contributions from k'th order neighbours is denoted

$$\ell_k^2(\theta) = \sum_{(i,j)\in\mathcal{P}(k)} \log p_{i,j}(n_i, n_j \,|\, \theta),$$

where $\mathcal{P}(k)$ denotes the set of distinct pairs (i, j) that are k'th order neighbours. It may then be interesting to use the *k'th order pairwise composite log-likelihood*

$$\ell_{<k}^2(\theta) = \sum_{l=1}^{k} \ell_l^2(\theta).$$

Note that the pairwise composite log-likelihood defined in (4.7) corresponds to including neighbours of all orders and we may write $\ell^2(\theta) = \ell_{<\infty}^2(\theta)$.

4.3.3 Quasi-likelihood

As an alternative to composite likelihood inference based on low dimensional marginal distributions as above we may consider inference based on low order moments. For an α-permanental random field the factorial moments are given by α-permanents and are especially tractable for low orders (see Vere-Jones, 1997; Møller and Rubak, 2010).

The quasi-likelihood as introduced by Wedderburn (1974) has been widely used in the literature and has a well developed asymptotic theory (cf. McCullagh, 1983). In the following we detail how to apply quasi-likelihood methods for α-permanental random fields, and only briefly recall the necessary general results.

As an initial step we construct a vector

$$\boldsymbol{Y} = (N_i, N_i(N_i - 1), N_{ij})_{\{i=1,\ldots,m\,;\, i<j\le m\}},$$

and denote the length of $\boldsymbol{Y}$ by n. We do not necessarily include products of all pairs of counts N_iN_j with $j > i$; we may only consider a subset based on neighbour relations. Note, that the

mean $\mu = \mu(\theta) = \mathrm{E}_\theta(\boldsymbol{Y})$ and the covariance matrix $\Sigma = \Sigma_\theta = \mathrm{Cov}_\theta(\boldsymbol{Y})$ can be expressed in terms of factorial moments of order at most 4, which are easily evaluated analytically.

Let $D = D_\theta$ be the $n \times d$ derivative matrix with entries $D_{ir} = \partial \mu_i / \partial \theta_r$. Then the quasi-likelihood estimating function for θ is

$$\boldsymbol{U}(\theta; \boldsymbol{Y}) = D_\theta^\top \Sigma_\theta^{-1} (\boldsymbol{Y} - \mu(\theta)),$$

which has zero expectation and covariance matrix $V = V_\theta = \mathrm{Cov}(\boldsymbol{U}) = D^\top \Sigma^{-1} D$. The quasi-likelihood estimator $\hat{\theta} \equiv \hat{\theta}(\boldsymbol{Y})$ is the root of the vector equation $\boldsymbol{U}(\hat{\theta}) = \boldsymbol{0}$, which can be found iteratively. Using a modified Newton-Raphson scheme the current estimate $\hat{\theta}^{(i)}$ is updated to

$$\hat{\theta}^{(i+1)} = \hat{\theta}^{(i)} + V_{\hat{\theta}^{(i)}}^{-1} \boldsymbol{U}(\hat{\theta}^{(i)}; \boldsymbol{Y}).$$

The iterative procedure is stopped once the estimate has converged within a specified tolerance. Under regularity conditions the quasi-likelihood estimator is asymptotically Gaussian with covariance matrix V^{-1}, which is calculated in each step of the iterative procedure. This allows us to attach an asymptotic variance V_θ^{-1} to the quasi-likelihood estimate.

4.4 Examples

4.4.1 One dimensional example

Figure 4.1 shows counts of clover leaves in 200 squares of size 5×5 cm along a 10 m transect line as detailed in Augustin et al. (2006). This data can be viewed as a realization of a one-dimensional random field consisting of 200 sites on the real line, with positive association expected between the counts due to the multiplicity of leaves per plant and the clustering of plants in patches. We model the leaf counts $\boldsymbol{N} = (N_1, \ldots, N_{200})$ as $\boldsymbol{N} \sim \mathrm{per}(\alpha, C)$, where $C_{i,j} = \kappa \rho^{|i-j|}$ with $0 \le \rho \le 1$ and $\kappa \ge 0$. Then condition (C2) is satisfied (Møller and Rubak, 2010, Proposition 1). Furthermore, by arguments similar to those used in the proof of Proposition 1 in Møller and Rubak (2010), it can be shown that for all $\alpha > 0$, C satisfies a regularity condition (Vere-Jones, 1997, Proposition 4.5) implying the existence of $\boldsymbol{X} \sim \Gamma_m(\alpha, C)$ so that $\mathrm{per}(\alpha, C)$ has a doubly stochastic construction, cf. Section 4.2. Therefore, it makes sense to calculate the Bayes estimate $\mathrm{E}_\theta(\boldsymbol{X} \mid \boldsymbol{n})$ of the conditional intensity for all positive α. The Bayes estimate for the model using the MLE as found below is superimposed as a line in Figure 4.1.

As an initial step in the parameter estimation we use the first-order composite log-likelihood. Notice, since $\ell^1(\alpha, \kappa, \rho)$ is independent of ρ, it is not possible to estimate this parameter using ℓ^1. Using the iterative Newton-Raphson procedure of Venables and Ripley (2002), the estimate of $\log(\kappa)$ is 0.247 ± 0.257 and the estimate of $1/\alpha$ is 0.396 ± 0.141, where both estimates are quoted plus/minus two standard errors. The point estimates correspond to $\hat{\kappa} = 1.28$ and $\hat{\alpha} = 2.5$. For a grid of parameter values evaluation of the full log-likelihood yielded the MLE $(\hat{\alpha}, \hat{\kappa}, \hat{\rho}) = (2.3, 1.28, 0.860)$. A three dimensional approximate 95% confidence region can be found by calculating the likelihood ratio $\lambda(\alpha, \kappa, \rho) = 2(\ell(\hat{\alpha}, \hat{\kappa}, \hat{\rho}) - \ell(\alpha, \kappa, \rho))$ for all points

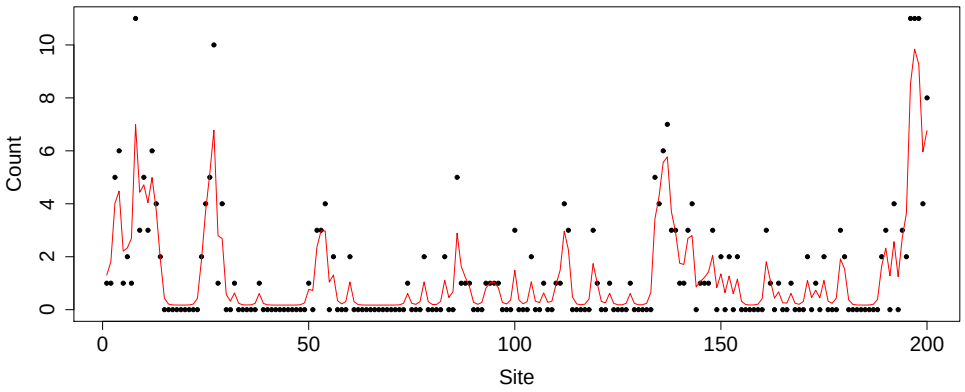


Figure 4.1: Counts of clover leaves in 200 square patches with Bayes estimate of the random mean field superimposed as a line.

of the parameter grid and compare with the 95th percentile of the χ^2_3-distribution. Marginal confidence intervals are $(1.4, 4.4)$ for α, $(0.7, 4.7)$ for κ, and $(0.8, 0.95)$ for ρ. To visualize the confidence region in two dimensions Figure 4.2a shows a contour plot of $\lambda(\alpha, \hat{\kappa}, \rho)$ as a function of (α, ρ) with κ fixed at the MLE $\hat{\kappa} = 1.28$. The contours are based on the 50th, 95th, and 99th percentile of the χ^2_3-distribution. Figures 4.2b-4.2d are similar contour plots based on ℓ^2_1, ℓ^2_9 and $\ell^2_{<\infty}$ with κ fixed at $\hat{\kappa} = 1.28$. In these plots the contours are no longer related to any confidence regions. It is clear that ℓ^2_1 in Figure 4.2b determines ρ quite well, and the higher order neighbour pairs do not contain much information about ρ. A plot of the empirical autocorrelation function (not shown) reveals that it is negative for neighbours of order 9, which explains the shape of the contour plot in Figure 4.2c, where the maximum is at $\rho = 0$. The pairwise composite log-likelihood with neighbours of all orders is a sum of many composite log-likelihoods, where ρ is poorly determined for the majority of them, which causes the shape of the contour plot in Figure 4.2d. However, the point estimates of the parameters other than ρ do not change much when inference is based on $\ell^2_{<k}$ for growing k. Based on $\ell^2_{<2}$ the estimates are $(\hat{\alpha}, \hat{\kappa}, \hat{\rho}) = (2.5, 1.28, 0.860)$ whereas $\ell^2_{<\infty}$ yields the point estimates $(\hat{\alpha}, \hat{\kappa}, \hat{\rho}) = (2.5, 1.28, 0.855)$.

Using the modified Newton-Raphson scheme described in Section 4.3.3 the quasi-likelihood estimates (with corresponding two standard errors) are found to be $\hat{\alpha} = 2.2 \pm 1.5$, $\hat{\kappa} = 1.35 \pm 0.7$ and $\hat{\rho} = 0.85 \pm 0.16$ when only first order neighbours are used. The quasi-likelihood estimates only change slightly when higher order neighbours are used, and they are not quoted here.

The full likelihood calculations have been carried out using the Monte Carlo (MC) importance sampling algorithm of Kou and McCullagh (2009), which provides an estimate of both the α-permanent in (4.5) and the standard error of this estimate. We used 10^5 samples, giving an average relative error (ratio of the standard error and the estimate) of 0.077. As noted in Kou and McCullagh (2009), their algorithm is especially well suited for estimating ratios of α-permanents as required in the Bayes estimate (4.3). For the calculation used for obtaining

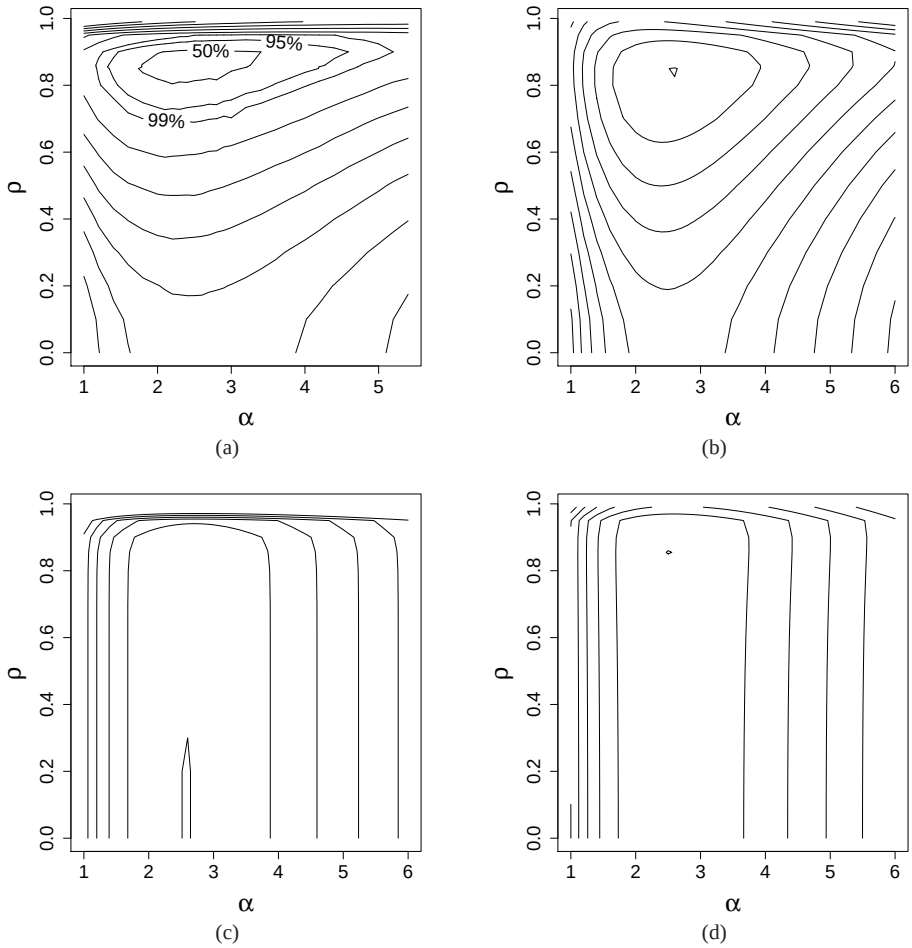


Figure 4.2: Contour plot of (a) the full log-likelihood, $\ell(\theta)$, compared with contour plots of the pairwise composite log-likelihood with (b) first order neighbours only, $\ell^2_1(\theta)$; (c) ninth order neighbours only, $\ell^2_9(\theta)$; (d) neighbours of all orders, $\ell^2_{<\infty}(\theta)$. For all the plots κ is fixed at the MLE $\hat{\kappa} = 1.28$.

Figure 4.1, 10^4 MC samples were sufficient.

It is possible to perform model validation based on simulation using the Poisson randomization (Møller and Rubak, 2010, Section 4.2). We simulated 100 realizations from the model using the MLE, and checked some properties of the data against the simulated realizations. A characteristic feature for the data is the large number of zeros overall and the apparent clustering of the zeros. For example, the average total number of zeros in the simulated realizations was 111 with the first and third quartile at 103 respectively 119, while data has 114 zeros. The largest cluster of zeros in data is 13 where the simulated realizations have an average of 12 with the first and third quartile at 10 respectively 15. Besides the simulation based validation we also checked empirical first and second order moments of data with the theoretical moments of the fitted model, and they also revealed a very good fit.

In conclusion any of the proposed estimation methods provide good point estimates, but in particular the composite likelihood based approach including neighbours of all orders has a big information loss about the correlation parameter ρ. When it is computationally feasible, as it was the case here, using the full likelihood is preferred.

4.4.2 Disease mapping

Choo and Walker (2008) presented a so-called multivariate Poisson-Gamma (MPG) model to investigate the spatial variations of cases $\boldsymbol{n} = (n_1, \ldots, n_m)$ of testis cancer in the $m = 19$ municipalities of the county of Frederiksborg in Denmark, where corresponding expected values $\boldsymbol{e} = (e_1, \ldots, e_m)$ based on the population and age structures are treated as covariates. For illustrative purposes, we present another approach using α-permanental random fields and leading to the perhaps surprising conclusion that there is little evidence in these data of either overdispersion or spatial correlation.

The parameters of interest are the incidence ratios γ_i, $i = 1, \ldots, m$, which indicate whether municipality i has an over-representation of testis cancer ($\gamma_i > 1$) or not ($0 < \gamma_i \leq 1$). Specifically, conditional on $\Gamma = (\gamma_1, \ldots, \gamma_m)$, we assume the data is a realization of independently Poisson distributed counts N_i with $\mathrm{E}(N_i \,|\, \Gamma) = \gamma_i e_i$, $i = 1, \ldots, m$. The raw estimates are given by $\hat{\gamma}_i = n_i / e_i$, which agree with the MLE if Γ is a deterministic parameter vector. However, typically Γ would be treated as a random field with spatial dependence, cf. Choo and Walker (2008) and the references therein.

Before proceeding any further, some general remarks about modeling of this type of spatial epidemiological data are needed. In principle, each count N_i can be viewed as the aggregation over an area A_i of an underlying point process specifying the domestic location of each individual diagnosed with the disease. It would be natural to specify a Cox point process model for this underlying data process, where the random intensity at location x, $\gamma(x)$ has mean $e(x)$, which is the known age-adjusted population density at x. Then, conditional on γ the counts N_i are independent Poisson variables with mean $\int_{A_i} \gamma(x)\,\mathrm{d}x$. The distributional properties of this integral are usually intractable, and it is a well-known unsolved problem in the literature to specify a point process model where inference based on aggregated count data is tractable (see

Richardson, 2003; Møller, 2003). A common approach, which we follow here, is simply to specify a model directly in terms of the aggregated data without considering a consistent underlying point process model. However, an important point to be derived from this discussion is that the model should respect geographic integrity, namely that the marginal distribution for a subset of the aggregated data should belong to the same class.

We assume $\Gamma \sim \Gamma_m(\alpha, R)$ where R is a correlation matrix. This ensures that $\mathrm{E}(N_i) = \mathrm{E}(\gamma_i e_i) = e_i$, as one may naturally require. Let $C = DRD$ with D diagonal and $D_{i,i} = \sqrt{e_i}$. We consider a doubly stochastic construction as in Section 4.2 with $\boldsymbol{X} = D\Gamma D \sim \Gamma_m(\alpha, C)$ and $\boldsymbol{N} \sim \mathrm{per}(\alpha, C)$, cf. (4.4). Moreover, assuming that $\alpha \in \left(0, \frac{2}{m}\right] \cup \left\{\frac{2}{m-1}, \frac{2}{m-2}, \ldots, 2\right\}$, condition (C1) is satisfied, and so the model is well defined.

The final stage of the model is to specify the off diagonal entries of R which determine the correlation structure of the model. A natural approach is to use a neighbourhood relation when specifying R, and we assume that

$$R_{i,j} = \begin{cases} \rho & \text{if } i \sim j \\ 0 & \text{otherwise,} \end{cases} \tag{4.9}$$

where for the present data, $i \sim j$ indicates that municipalities i and j share a border. Care must be taken to ensure R is indeed semi-definite; we realized empirically that R is only semi-definite if $0 \leq \rho \leq \rho_c$, where $\rho_c \leq 1$ is a critical value depending on the neighbourhood structure. The critical value can be approximated before any inference is carried out, e.g. by using a spectral decomposition, which for the data at hand revealed $\rho_c \approx 0.416$.

In the special case $\rho = 0$, the model reduces to m independent negative binomial random variables, and so the full log-likelihood is equivalent to the first-order composite log-likelihood. For this model α is the only parameter, and it is straightforward to find the Bayes estimates

$$\mathrm{E}(\gamma_i \mid \boldsymbol{n}) = \frac{1 + \hat{\alpha} n_i}{1 + \hat{\alpha} e_i} \quad i = 1, \ldots, m.$$

The MLE of $1/\alpha$ is 36.2 ± 69.2 leading to the point estimate $\hat{\alpha} = 0.0277$. The large value of twice the standard error indicates that a negative binomial model is not necessary and a likelihood ratio test against the simpler Poisson null model is performed. The negative binomial model has $-2\ell(\hat{\alpha}) = 107.66$ whereas the Poisson model (corresponding to $\alpha = 0$) has $-2\ell(0) = 105.44$, and the likelihood ratio test yields a p-value of about 14%. Similarly, the standard Pearson χ^2 test for over-dispersion yields the test statistic of $\sum(n_i - e_i)^2/e_i = 25.5$ on 18 degrees of freedom, for a p-value of about 11%. In other words, there is little evidence of either over-dispersion or spatial correlation.

If ρ is included as a parameter in the model, either full, quasi- or pairwise composite likelihood inference can be used. However, in this example the MC importance sampling algorithm of Kou and McCullagh (2009) used to estimate the α-permanent performs poorly; even for a very large number of MC samples (10^8) the standard error of the estimate is relatively large. On the other hand, both quasi- and pairwise composite likelihood inference is fast and does not require any approximation (apart from the inherent surrogate nature of these methods).

For the quasi-likelihood iterative scheme ρ quickly approaches zero at which point the covariance matrix V becomes singular, so no standard errors can be given. However, α stabilizes at 0.027 ± 0.064 making it clear that $\alpha = 0$ is well within two standard errors of the estimate. Figure 4.3a shows a contour plot of the pairwise composite log-likelihood based on first order

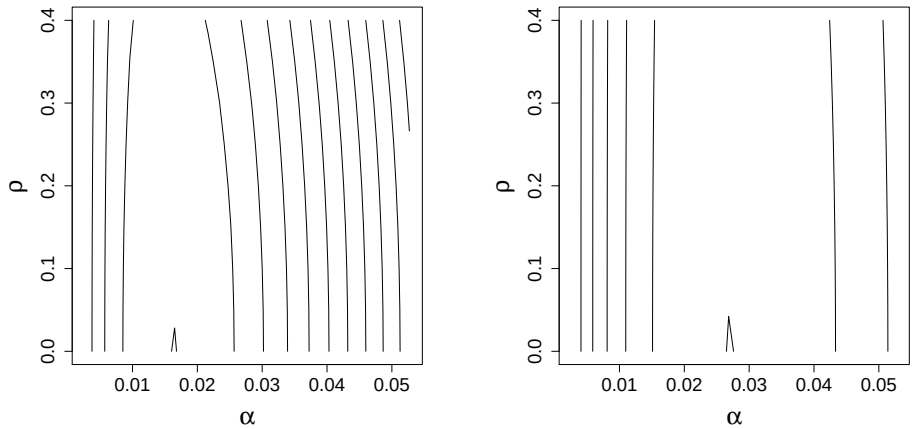


Figure 4.3: Contour plots based on pairwise composite log-likelihood using (a) first order neighbours only (b) neighbours of all orders.

neighbours only, whereas the contour plot in Figure 4.3b is based on neighbours of all orders. Notice that in both cases the correlation parameter ρ is poorly determined and the maximal composite likelihood value is attained at $\rho = 0$ confirming the findings of the quasi-likelihood method. Furthermore, it appears that Figure 4.3b contains less information about ρ than Figure 4.3a. This is explained by the fact that ρ only enters in bivariate distributions of directly neighbouring sites, and all the terms of $\ell^2_{<\infty}$ not appearing in ℓ^2_1 are independent of ρ. The estimate of α is respectively 0.0165 and 0.0268 when using ℓ^2_1 and $\ell^2_{<\infty}$. Thus, it seems preferable to use $\ell^2_{<\infty}$ to estimate α since it yields an estimate close to the MLE for $\rho = 0$.

For this dataset the main interest is in estimating the incidence ratios γ_i, which is done by calculating the Bayes estimates $\mathrm{E}_{\hat{\theta}}(\gamma_i \mid \boldsymbol{n})$ under the fitted model. Table 4.1 lists these estimates for each model as well as the estimates for the MPG model in Choo and Walker (2008). The model with $\rho = \rho_c$ is included for illustrative purposes and for both this model and the independent negative binomial model with $\rho = 0$ the value of α is fixed at 0.0277. The table reveals that estimates based on the MPG model are close to estimates based on the the independent negative binomial model lending further support to the findings that a complex model is unnecessary for this particular dataset. In conclusion, it appears that it suffices to use the model with no spatial dependence between incidence ratios, which was not touched upon by Choo and Walker (2008).

Table 4.1: Bayes estimates of the incidence ratios for the two α-permanental models with $\rho = \rho_c$ and $\rho = 0$ compared with raw Poisson estimates and MPG estimates of Choo and Walker (2008).

	n_i	e_i	raw	$\rho = 0$	$\rho = \rho_c$	MPG
Allerød	18	17.61	1.02	1.01	0.97	1.01
Birkerød	17	18.20	0.93	0.98	1.01	1.00
Farum	14	13.65	1.03	1.01	1.02	0.99
Fredensborg-Humlebæk	14	14.29	0.98	0.99	0.97	0.93
Frederikssund	21	13.17	1.59	1.16	1.17	1.14
Frederiksværk	14	14.63	0.96	0.99	0.99	0.98
Græsted-Gilleleje	13	12.38	1.05	1.01	0.98	0.93
Helsinge	8	13.66	0.59	0.89	0.89	0.86
Helsingør	31	47.18	0.66	0.81	0.81	0.73
Hillerød	28	27.23	1.03	1.01	1.00	0.98
Hundested	8	6.44	1.24	1.04	1.22	1.03
Hørsholm	28	17.04	1.64	1.21	1.03	1.23
Jægerspris	4	6.05	0.66	0.95	0.98	0.97
Karlebo	12	13.78	0.87	0.96	1.01	0.99
Skibby	6	4.57	1.31	1.03	1.10	1.09
Skævinge	6	4.28	1.40	1.04	0.98	1.02
Slangerup	3	6.44	0.47	0.92	1.05	0.95
Stenløse	13	10.47	1.24	1.05	0.95	1.05
Ølstykke	14	10.93	1.28	1.06	1.06	1.11

To calculate the Bayes estimates for the model with $\rho = \rho_c$, ratios of α-permanents are again needed, but this poses no significant problem, since the MC importance sampling algorithm estimates these well even though the individual α-permanents are difficult to estimate.

4.5 Discussion

For the dataset of counts of clover leaves in Section 4.4.1 the α-permanental random field model with an exponential covariance matrix provides a good fit. Estimation based on both the full, quasi- and pairwise composite likelihood gives similar point estimates, but the shape of the pairwise composite likelihood is sensitive to the choice of neighbourhood order included in the model. This adds the disadvantage of having to choose the neighbourhood order when using the pairwise composite likelihood, while the quasi-likelihood appears to be less sensitive to this choice. In the analysis of this dataset, it is noticeable that the Bayes estimate of the random mean field in Figure 4.1 is spiky, which may be caused by the choice of covariance model. An immediate advantage of using the exponential covariance model is that the α-permanental model is well defined for all values of $\alpha \geq 0$. For a general covariance model the largest generally admissible value of α is 2. However, it may be possible to find covariance models allowing for $\alpha > 2$ as it was the case for the exponential covariance model. Alterna-

tively, it may be possible to obtain a good fit with α fixed at 2 using an alternative covariance model of e.g. polynomial type, which would be expected to yield a smoother Bayes estimate of the random mean field.

The dataset of testis cancer cases in Section 4.4.2 illustrates a simple yet important fact: There is little point in using a complicated model with over-dispersion and spatial dependence if the data shows evidence of neither. However, the example still allows us to illustrate the potential use of the α-permanental model for disease mapping.

Acknowledgments

Supported by the NSF, grant no. DMS-0906592, by the Danish Natural Science Research Council, grants 272-06-0442 and 09-072331('Point process modelling and statistical inference'), by the Danish Agency for Science, Technology and Innovation, grant 645-06-0528, 'International PhD student', and by Centre for Stochastic Geometry and Advanced Bioimaging, funded by a grant from the Villum Foundation.

Bibliography

Augustin, N. H., J. McNicol, and C. A. Marriott (2006). Using the truncated auto-Poisson model for spatially correlated counts of vegetation. *J. Agric. Biol. Environ. Stat. 11*, 1–23.

Choo, L. and S. G. Walker (2008). A new approach to investigating spatial variations of disease. *J. Roy. Statist. Soc. Ser. A 171*, 395–405.

Cox, D. R. and N. Reid (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika 91*, 729–737.

Georgii, H.-O. and H. J. Yoo (2005). Conditional intensity and Gibbsianness of determinantal point processes. *J. Stat. Phys. 118*, 617–666.

Griffiths, R. C. and R. K. Milne (1987). A class of infinitely divisible multivariate negative binomial distributions. *J. Multivariate Anal. 22*, 13–23.

Hammersley, J. M. and D. C. Handscomb (1964). *Monte Carlo Methods*. London: Methuen.

Kou, S. C. and P. McCullagh (2009). Approximating the α-permanent. *Biometrika 96*, 635–644.

Lindsay, B. G. (1988). Composite likelihood methods. In N. U. Prabhu (Ed.), *Statistical Inference from Stochastic Processes*, Providence, pp. 221–239. American Mathematical Society.

Macchi, O. (1971). Stochastic point processes and multicoincidences. *IEEE Trans. Inform. Theory 17*, 2–7.

Macchi, O. (1975). The coincidence approach to stochastic point processes. *Adv. in Appl. Probab. 7*, 83–122.

Maybee, J. and J. Quirk (1969). Qualitative problems in matrix theory. *SIAM Rev. 11*, 30–51.

McCullagh, P. (1983). Quasi-likelihood functions. *Ann. Statist. 11*, 59–67.

McCullagh, P. and J. Møller (2006). The permanental process. *Adv. in Appl. Probab. 38*, 873–888.

Minc, H. (1978). *Permanents*. Reading, MA: Addison-Wesley.

Møller, J. (2003). A comparison of spatial point process models in epidemiological applications. In P. Green, N. Hjort, and S. Richardson (Eds.), *Highly Structured Stochastic Systems*, pp. 264–268. Oxford: Oxford University Press.

Møller, J. and E. Rubak (2010). A model for positively correlated count variables. *Int. Stat. Rev. 78*(1), 65–80.

Richardson, S. (2003). Spatial models in epidemiological applications. In P. Green, N. Hjort, and S. Richardson (Eds.), *Highly Structured Stochastic Systems*, pp. 237–259. Oxford: Oxford University Press.

Shirai, T. and Y. Takahashi (2003a). Random point fields associated with certain Fredholm determinants I: fermion, Poisson and boson point processes. *J. Func. Anal. 205*, 414–463.

Shirai, T. and Y. Takahashi (2003b). Random point fields associated with certain Fredholm determinants II: fermion shifts and their ergodic and Gibbs properties. *Ann. Probab. 31*, 1533–1564.

Sweet, R. A. (1969). A recursive relation for the determinant of a pentadiagonal matrix. *Communications of the ACM 12*, 330–332.

Venables, W. N. and B. D. Ripley (2002). *Modern Applied Statistics with S*. New York: Springer.

Vere-Jones, D. (1997). Alpha-permanents and their applications to multivariate gamma, negative binomial and ordinary binomial distributions. *New Zealand J. Math. 26*, 125–149.

Wedderburn, R. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika 61*, 439–447.

4.A Evaluating α-permanents

In this appendix we both present some general results for α-permanents (Appendix 4.A.1) and some results on simple patterned matrices (Appendices 4.A.2-4.A.3) as well as illustrate how an existing algorithm for approximating α-permanents in some cases may be improved (Appendix 4.A.4).

4.A.1 Preliminary results

Here we give a few general results for α-permanents, which we will need later.

Expansion by sums of cyclic products

For any positive integer n let $I_n = (1, \ldots, n)$, and let I_0 denote the "empty subsequence". Given a positive integer $m \leq n$, let $I = (i_1, \ldots, i_m)$ be an ordered subsequence of I_n meaning that $1 \leq i_1 < \cdots < i_m \leq n$, and let $I^c = (j_1, \ldots, j_{n-m})$ denote the complementary subsequence so that $\{i_1, \ldots, i_m\}$ and $\{j_1, \ldots, j_{n-m}\}$ are disjoint with union $\{1, \ldots, n\}$. For any such I we let $\mathcal{I}(r, I)$ denote the class of ordered subsequences of I of length $r \geq 0$ using the convention $\mathcal{I}(0, I) = \{I_0\}$.

For any $n \times n$ matrix A, we define A_I as the $m \times m$ submatrix of A with (k, l)'th entry A_{i_k, i_l}. Furthermore, we let $\text{cyp}(A_I)$ denote the sum of cyclic products of length $|I| = m$ formed from A_I. Thus, $\text{cyp}(A_I)$ is a sum over $(m-1)!$ terms, and if e.g. $m = 3$ we have

$$\text{cyp}(A_I) = A_{i_1,i_2} A_{i_2,i_3} A_{i_3,i_1} + A_{i_1,i_3} A_{i_3,i_2} A_{i_2,i_1}.$$

Maybee and Quirk (1969) provides the following formula for calculating the determinant of a $n \times n$ matrix A.

Theorem 1. For $n > 1$ and any fixed $I \in \mathcal{I}(n-1, \{1, \ldots, n\})$,

$$|A| = A_{I^c, I^c} |A_I| + \sum_{r=0}^{n-2} (-1)^{n-1-r} \sum_{J \in \mathcal{I}(r, I)} |A_J| \text{cyp}(A_{J^c}),$$

where we define $|A_\emptyset| = 1$.

This result extends straightforwardly to α-permanents.

Corollary 1. For all $\alpha \in \mathbb{R}$, $n > 1$ and any fixed $I \in \mathcal{I}(n-1, \{1, \ldots, n\})$,

$$\text{per}_\alpha(A) = \alpha A_{I^c, I^c} \text{per}_\alpha(A_I) + \sum_{r=0}^{n-2} \sum_{J \in \mathcal{I}(r, I)} \alpha \text{per}_\alpha(A_J) \text{cyp}(A_{J^c}). \tag{4.10}$$

where we define $\text{per}_\alpha(A_\emptyset) = 1$.

Proof. From Theorem 1 we know that the right hand side of (4.10) has all the $n!$ terms of the form $A_{1,\sigma(1)} \cdots A_{n,\sigma(n)}$ and we only need to verify each term is weighted correctly. The first term on the right hand side of (4.10) is

$$\alpha A_{I^c,I^c} \mathrm{per}_\alpha(A_I) = \alpha A_{I^c,I^c} \sum_{\sigma \in \mathcal{S}_{n-1}} \alpha^{c(\sigma)} \prod_{i=1}^{n-1} (A_I)_{i,\sigma(i)},$$

and since A_{I^c,I^c} introduces a new cycle to all terms the weighting with $\alpha^{c(\sigma)+1}$ is correct. The rest of the terms are

$$\sum_{r=0}^{n-2} \sum_{J \in \mathcal{I}(r,I)} \alpha \mathrm{per}_\alpha(A_J) \mathrm{cyp}(A_{J^c}) = \sum_{r=0}^{n-2} \sum_{J \in \mathcal{I}(r,I)} \alpha \sum_{\sigma \in \mathcal{S}_r} \alpha^{c(\sigma)} \prod_{i=1}^{r} (A_J)_{i,\sigma(i)} \mathrm{cyp}(A_{J^c})$$

and since again exactly one new cycle is introduced by the cyclic product the weight is correct. □

Expansion by cofactors

Let A be a $n \times n$ matrix. By isolating a given element $A_{r,s}$ of $\mathrm{per}_\alpha(A)$ it is obvious that the coefficient of $A_{r,s}$ depends only on the elements of the reduced matrix of order $n-1$ with row r and column s deleted. However, the coefficient is in general not the α-permanent of the reduced matrix, and Vere-Jones (1997) remarks that no simple cofactor expansion of $\mathrm{per}_\alpha(A)$ is known. However, in the following we give a cofactor expansion by a slight modification of the recipe for determinants. The (r,s) minor is a square matrix $A^{r,s}$ of order $n-1$ obtained from A in two steps as follows: First switch rows r and s; then delete column s and row s (row r from the original matrix). The switching of rows is not a part of the standard definition of a minor, but it is needed for $\alpha \neq \pm 1$ to maintain the cycle structure, and is done prior to deletion to avoid ambiguity about labelling. If $r = s$, the first step is nugatory; otherwise if $r \neq s$ the symmetrically opposed component $A_{s,r}$ occurs on the diagonal of $A^{r,s}$, and every other element on the diagonal of the minor also occurs on the diagonal of A. The components of the cofactor matrix $\mathrm{cof}_\alpha(A)$ are defined as

$$\mathrm{cof}_\alpha(A)_{r,s} = \begin{cases} \alpha \mathrm{per}_\alpha(A^{r,s}) & r = s \\ \mathrm{per}_\alpha(A^{r,s}) & \text{otherwise.} \end{cases}$$

On row r, the cofactor expansion of the α-permanent is

$$\begin{aligned} \mathrm{per}_\alpha(A) &= \alpha A_{r,r} \mathrm{per}_\alpha(A^{r,r}) + \sum_{s \neq r} A_{r,s} \mathrm{per}_\alpha(A^{r,s}) \\ &= \sum_{s=1}^{n} A_{r,s} \ \mathrm{cof}_\alpha(A)_{r,s}. \end{aligned} \tag{4.11}$$

Although the definition of a minor has been modified to suit the general case, for $\alpha = -1$ this reduces to the standard cofactor expansion of a determinant.

4.A.2 Block matrices

Evaluating multivariate negative binomial probabilities involves the α-permanent of a block matrix, which are studied in this appendix. For any $m \times m$ matrix A and non-negative integers $\boldsymbol{n} = (n_1, \ldots, n_m)$, let $n_\star = \sum_{i=1}^m n_i$ and define the block matrix $A[\boldsymbol{n}]$ as the $n_\star \times n_\star$ matrix obtained from A by repeating index i n_i times. For example, if $m = 4$ and $\boldsymbol{n} = (2, 0, 1, 3)$

$$A[\boldsymbol{n}] = A[(2,0,1,3)] = \begin{bmatrix} A_{1,1} & A_{1,1} & A_{1,3} & A_{1,4} & A_{1,4} & A_{1,4} \\ A_{1,1} & A_{1,1} & A_{1,3} & A_{1,4} & A_{1,4} & A_{1,4} \\ A_{3,1} & A_{3,1} & A_{3,3} & A_{3,4} & A_{3,4} & A_{3,4} \\ A_{4,1} & A_{4,1} & A_{4,3} & A_{4,4} & A_{4,4} & A_{4,4} \\ A_{4,1} & A_{4,1} & A_{4,3} & A_{4,4} & A_{4,4} & A_{4,4} \\ A_{4,1} & A_{4,1} & A_{4,3} & A_{4,4} & A_{4,4} & A_{4,4} \end{bmatrix}.$$

We call A the generating matrix, $\boldsymbol{n}$ the block sizes, and $A[\boldsymbol{n}]$ a m-dimensional block matrix.

One-dimensional block matrices

When A is a 1×1 matrix with element a and the block size is n we have $A[n] = a\mathbf{1}_n$, where $\mathbf{1}_n$ is the $n \times n$ matrix whose elements are all one. This matrix has α-permanent

$$\mathrm{per}_\alpha(\mathbf{1}_n) = \alpha^{\uparrow n} = \alpha(\alpha+1)\cdots(\alpha+n-1),$$

called the ascending factorial function. Furthermore, $\mathrm{per}_\alpha(A[n]) = a^n \alpha^{\uparrow n}$.

Block-diagonal matrices

For a general block-diagonal matrix it is easy to verify that the α-permanent is the product of the α-permanent of the blocks. The special block diagonal matrix with constant blocks can be written as $A[\boldsymbol{n}]$, where the generator A is a diagonal matrix with diagonal $(a_1, \ldots, a_m)$, and in this case we have

$$\mathrm{per}_\alpha(A[\boldsymbol{n}]) = \prod_{i=1}^m \mathrm{per}_\alpha(a_i \mathbf{1}_{n_i}) = \prod_{i=1}^m a_i^{n_i} \alpha^{\uparrow n_i}.$$

Two-dimensional block matrix

For two-dimensional block matrices we have the following result allowing efficient calculation of the α-permanent.

Proposition 1. Let A be a 2×2 matrix and define

$$\rho = \frac{A_{1,2}A_{2,1}}{A_{1,1}A_{2,2}}.$$

Then

$$\mathrm{per}_\alpha(A[n_1,n_2]) = A_{1,1}^{n_1}A_{2,2}^{n_2}\alpha^{\uparrow n_1}\alpha^{\uparrow n_2}\sum_{j=0}^{n_1\wedge n_2}\frac{n_1^{\downarrow j}\,n_2^{\downarrow j}\,\rho^j}{j!\;\alpha^{\uparrow j}}, \tag{4.12}$$

where we define $\alpha^{\uparrow n_1}\alpha^{\uparrow n_2}/\alpha^{\uparrow j} = 0$ when both the numerator and denominator is zero, $n^{\downarrow j} = n(n-1)\cdots(n-j+1)$, $n^{\downarrow 0} = 1$ and $\alpha^{\uparrow 0} = 1$. Thus, $n^{\downarrow n} = 1^{\uparrow n} = n!$ and $\mathrm{per}_\alpha(A[0,0]) = 1$.

Proof. As a preliminary, we note the following property of the ascending factorial function:

$$\sum_{r=0}^{n}\alpha^{\uparrow(k+r)}/r! = \frac{\alpha^{\uparrow(n+k+1)}}{n!\,(\alpha+k)} \tag{4.13}$$

for non-negative integer k such that $\alpha+k \neq 0$. If $k=0$ the sum is $(\alpha+1)^{\uparrow n}/n!$, which is readily established by induction on n. The result for general k then follows from $\alpha^{\uparrow(k+r)} = \alpha^{\uparrow k}(\alpha+k)^{\uparrow r}$.

Any 2×2 matrix A can be factorized as

$$A = DRD = \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix}\begin{bmatrix} 1 & \rho_1 \\ \rho_2 & 1 \end{bmatrix}\begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix},$$

where d_1, d_2, ρ_1, ρ_2 are the (possibly complex) numbers satisfying $d_1^2 = A_{1,1}$, $d_2^2 = A_{2,2}$, $\rho_1 = A_{1,2}/(d_1d_2)$, and $\rho_2 = A_{2,1}/(d_1d_2)$. Then it can be verified that

$$\mathrm{per}_\alpha(A[n_1,n_2]) = A_{1,1}^{n_1}A_{2,2}^{n_2}\mathrm{per}_\alpha(R[n_1,n_2]), \tag{4.14}$$

and therefore to prove (4.12) it is sufficient to show that

$$\mathrm{per}_\alpha(R[n_1,n_2]) = \alpha^{\uparrow n_1}\alpha^{\uparrow n_2}\sum_{j=0}^{n_1\wedge n_2}\frac{n_1^{\downarrow j}\,n_2^{\downarrow j}\,\rho^j}{j!\;\alpha^{\uparrow j}} \tag{4.15}$$

with $\rho = \rho_1\rho_2$.

For $n_2 > 0$, let $S[n_1,n_2]$ be the matrix obtained from $R[n_1,n_2]$ by replacing the first row by the last row. Then S is square but asymmetric, and the cofactor expansion gives the bivariate permanental recurrence relation

$$\begin{aligned}\mathrm{per}_\alpha(R[n_1+1,n_2]) &= (\alpha+n_1)\mathrm{per}_\alpha(R[n_1,n_2]) + n_2\rho_{12}\mathrm{per}_\alpha(S[n_1+1,n_2-1]),\\ \mathrm{per}_\alpha(S[n_1+1,n_2]) &= (\alpha+n_1)\rho_{21}\mathrm{per}_\alpha(R[n_1,n_2]) + n_2\mathrm{per}_\alpha(S[n_1+1,n_2-1]).\end{aligned}$$

For successively smaller values of n_2, repeated substitution of the second expression into the first eliminates $\mathrm{per}_\alpha(S[...])$, giving the linear recurrence relations

$$\mathrm{per}_\alpha(R[n_1+1,n_2]) = (\alpha+n_1)\mathrm{per}_\alpha(R[n_1,n_2]) + \rho(\alpha+n_1)\sum_{i=1}^{n_2}n_2^{\downarrow i}\mathrm{per}_\alpha(R[n_1,n_2-i]), \tag{4.16}$$

one equation for each $n_1, n_2 \geq 0$. These equations are linearly independent of full rank, and have a unique solution for any given boundary value $\mathrm{per}_\alpha(R[0,0])$. It follows immediately that $\mathrm{per}_\alpha(R[n,0]) = \alpha^{\uparrow n}\mathrm{per}_\alpha(R[0,0])$, so the desired boundary value is one.

On the assumption that $n_2 \leq n_1$, we now show that (4.15) is a solution of the system of linear equations (4.16). We start by noticing from (4.16) that $\alpha^{\uparrow n_1}$ is a common factor in all terms of $\mathrm{per}_\alpha(R[n_1, n_2])$ implying $\mathrm{per}_\alpha(R[n_1, n_2]) = 0$ when α is a non-positive integer bigger than $-n_2$. This proves the claim for these values of α and in what follows we consider all other values of α. After substituting (4.15), the right side of (4.16) becomes

$$\begin{aligned} & \alpha^{\uparrow n_1+1}\alpha^{\uparrow n_2}\sum_{j=0}^{n_2}\frac{n_1^{\downarrow j}\,n_2^{\downarrow j}}{j!}\frac{\rho^j}{\alpha^{\uparrow j}} + \rho\alpha^{\uparrow n_1+1}\sum_{i=1}^{n_2} n_2^{\downarrow i}\alpha^{\uparrow n_2-i}\sum_{j=0}^{n_2-i}\frac{n_1^{\downarrow j}\,(n_2-i)^{\downarrow j}}{j!}\frac{\rho^j}{\alpha^{\uparrow j}} \\ = \quad & \alpha^{\uparrow n_1+1}\alpha^{\uparrow n_2}\sum_{j=0}^{n_2}\frac{n_1^{\downarrow j}\,n_2^{\downarrow j}}{j!}\frac{\rho^j}{\alpha^{\uparrow j}} + \rho\alpha^{\uparrow n_1+1}\sum_{j=0}^{n_2-1}\binom{n_1}{j}\frac{\rho^j}{\alpha^{\uparrow j}}\sum_{i=1}^{n_2-j} n_2^{\downarrow i}\alpha^{\uparrow n_2-i}(n_2-i)^{\downarrow j}. \end{aligned}$$

On account of the ascending factorial identity (4.13), the final sum reduces to

$$\sum_{i=1}^{n_2-j} n_2^{\downarrow i}\alpha^{\uparrow n_2-i}(n_2-i)^{\downarrow j} = \alpha^{\uparrow n_2}n_2!/((\alpha+j)(n_2-j-1)!) = \alpha^{\uparrow n_2}n_2^{\downarrow j+1}/(\alpha+j),$$

which simplifies the preceding expression to

$$\begin{aligned} & \alpha^{\uparrow n_1+1}\alpha^{\uparrow n_2}\sum_{j=0}^{n_2}\binom{n_1}{j}\frac{n_2^{\downarrow j}\rho^j}{\alpha^{\uparrow j}} + \rho\alpha^{\uparrow n_1+1}\alpha^{\uparrow n_2}\sum_{j=0}^{n_2-1}\binom{n_1}{j}\frac{n_2^{\downarrow j+1}\rho^j}{(\alpha+j)\alpha^{\uparrow j}} \\ = \quad & \alpha^{\uparrow n_1+1}\alpha^{\uparrow n_2}\sum_{j=0}^{n_2}\binom{n_1}{j}\frac{n_2^{\downarrow j}\rho^j}{\alpha^{\uparrow j}} + \alpha^{\uparrow n_1+1}\alpha^{\uparrow n_2}\sum_{j=0}^{n_2-1}\binom{n_1}{j}\frac{n_2^{\downarrow j+1}\rho^{j+1}}{\alpha^{\uparrow j+1}} \\ = \quad & \alpha^{\uparrow n_1+1}\alpha^{\uparrow n_2}\sum_{j=0}^{n_2}\binom{n_1}{j}\frac{n_2^{\downarrow j}\rho^j}{\alpha^{\uparrow j}} + \alpha^{\uparrow n_1+1}\alpha^{\uparrow n_2}\sum_{j=1}^{n_2}\binom{n_1}{j-1}\frac{n_2^{\downarrow j}\rho^j}{\alpha^{\uparrow j}} \\ = \quad & \alpha^{\uparrow n_1+1}\alpha^{\uparrow n_2}\sum_{j=0}^{n_2}\binom{n_1+1}{j}\frac{n_2^{\downarrow j}\rho^j}{\alpha^{\uparrow j}}, \end{aligned}$$

showing that (4.15) satisfies the permanental recurrence equations (4.16). Since the recurrence equations are linear, any solution satisfying the desired boundary condition $\mathrm{per}_\alpha(A[0,0]) = 1$ is necessarily unique. □

Proposition 1 can be combined with the result on block-diagonal matrices (Section 4.A.2) to calculate the α-permanent of a block-diagonal matrix where each block possibly is a two-dimensional block matrix.

The ordinary permanent ($\alpha = 1$)

For a general m-dimensional block matrix we have the following result in the special case $\alpha = 1$, for which the α-permanent reduces to the ordinary permanent (Minc, 1978).

Proposition 2. Let $A[\boldsymbol{n}]$ be a m-dimensional block matrix with block sizes $\boldsymbol{n} = (n_1, \ldots, n_m)$. Further, let $T_{\boldsymbol{n}}$ denote the set of all two way tables $k = \{k_{ij}\}_{i,j=1,\ldots,m}$, $k_{ij} \in \mathbb{N}_0$ with row and column totals $n_1, \ldots, n_m$. Then

$$\operatorname{per}_1(A[\boldsymbol{n}]) = \sum_{T_{\boldsymbol{n}}} \prod_{i=1}^{m} (n_i!)^2 \prod_{i,j=1}^{m} \frac{A_{i,j}^{k_{ij}}}{k_{ij}!}$$

Proof. By definition,

$$\operatorname{per}_1(A[\boldsymbol{n}]) = \sum_{\sigma \in \mathcal{S}_{n_\star}} A[\boldsymbol{n}]_{1,\sigma(1)} \cdots A[\boldsymbol{n}]_{n_\star,\sigma(n_\star)}.$$

In each term of the sum the i'th row index must occur exactly n_i times and the j'th column index must occur exactly n_j times. This makes it clear that each term in the sum is of the form

$$\prod_{i=1}^{m} \prod_{j=1}^{m} A_{i,j}^{k_{ij}}, \quad \text{where} \quad \sum_{j=1}^{m} k_{ij} = n_i \quad \text{and} \quad \sum_{i=1}^{m} k_{ij} = n_j. \tag{4.17}$$

The question is how many times each term of this form occurs in the sum over all permutations. First k_{11} $A_{1,1}$'s must be chosen from $A[\boldsymbol{n}]$, which can be done in

$$\frac{n_1 n_1 \cdot (n_1 - 1)(n_1 - 1) \cdots (n_1 - k_{11}+1)(n_1 - k_{11} + 1)}{k_{11}!}$$

ways. When these are chosen we must choose k_{12} $A_{1,2}$'s, which can be done in

$$\frac{(n_1 - k_{11})n_2 \cdot (n_1 - k_{11} - 1)(n_2 - 1) \cdots (n_1 - k_{11} - k_{12} + 1)(n_2 - k_{12} + 1)}{k_{12}!}$$

ways. We can continue in this fashion and finally find the number of ways to choose the k_{1m} $A_{1,m}$'s. Then we can start a new row and find that the k_{21} $A_{2,1}$'s can be chosen in

$$\frac{n_2(n_1 - k_{11}) \cdot (n_2 - 1)(n_1 - k_{11} - 1) \cdots (n_2 - k_{21} + 1)(n_1 - k_{11} - k_{21} + 1)}{k_{21}!}$$

ways. Continuing in this fashion we see that for $i, j = 1, \ldots, m$ the number of ways to choose the k_{ij} A_{ij}'s is

$$\frac{(n_i - k_{i1} - \cdots - k_{i,j-1})(n_j - k_{1j} - \cdots - k_{i-1,j}) \cdots (n_i - k_{i1} - \cdots - k_{ij} + 1)(n_j - k_{1j} - \cdots - k_{ij} + 1)}{k_{ij}!}. \tag{4.18}$$

To find the coefficient for the term in (4.17) we need to take the product over $i, j = 1, \ldots, m$ of (4.18). The product of the numerators simplifies considerably and we end up with

$$\frac{\prod_{i=1}^{m} (n_i!)^2}{\prod_{i=1}^{m} \prod_{j=1}^{m} k_{ij}!},$$

and so the result follows. □

In the two-dimensional case Proposition 2 extends to $\alpha > 0$ as detailed in Section 4.A.2. It is plausible that this is also the case in the m-dimensional case. More precisely, if we as in the proof of Proposition 1 consider a $m \times m$ matrix R with unit diagonal we conjecture

$$\mathrm{per}_\alpha(R[\boldsymbol{n}]) = \prod_{i=1}^{m}(n_i!\alpha^{\uparrow n_i}) \sum_{T_{\boldsymbol{n}}} \prod_{i,j=1}^{m} \frac{R_{i,j}^{k_{ij}}}{k_{ij}!} P(k,\alpha)$$

where P is a (hopefully simple) rational function in k and α necessarily satisfying $P(k,1) = 1$. However, we have not yet been able to establish the correct form for P even though some promising patterns have been found in low dimensional examples.

4.A.3 Penta-diagonal matrices

This section generalizes the efficient algorithm of Sweet (1969) for computing the determinant of a penta-diagonal matrix to the more general case of the α-permanent. The development follows the same lines as Sweet (1969).

Let A be a $n \times n$ penta-diagonal matrix (i.e. $A_{i,j} = 0$ for $|i-j| > 2$) and let $n > 3$. By applying Corollary 1 with $I = I_{n-1} = (1,\ldots,n-1)$ we have

$$\mathrm{per}_\alpha(A) = \alpha A_{n,n}\mathrm{per}_\alpha(A_{I_{n-1}}) + \sum_{r=0}^{n-2} \sum_{J\in\mathcal{I}(r,I_{n-1})} \alpha\mathrm{per}_\alpha(A_J)\mathrm{cyp}(A_{J^c}).$$

Note that for $J \in \mathcal{I}(r, I_{n-1})$ the subsequence J^c always contains n. If $|J^c| \geq 3$ (i.e. $r \leq n-3$) the only subsequences J^c giving rise to non-zero cyclic products are of the form $J^c = I_r^c = (r+1,\ldots,n)$. This can be seen by considering a subsequence of the form $\tilde{J} = (r,\ldots,r+j-1,r+j+1,\ldots,n)$. In order to make a cyclic product $A_{i_1,i_2}\cdots A_{i_r,i_1}$ non-zero using $\tilde{J}$ as index set, we have to choose the element $A_{r+j-1,r+j+1}$, but then we have no way of connecting the upper and lower end of the index set without having a difference of more than two in the indices leading to one of the elements being zero.

When $r = n-2$ such that $|J^c| = 2$ the only non-zero two-cycles clearly arise when $J^c = (n-1,n)$ and $J^c = (n-2,n)$. Consequently we have

$$\begin{aligned}
&\sum_{r=0}^{n-2} \sum_{J\in\mathcal{I}(r,I_{n-1})} \alpha\mathrm{per}_\alpha(A_J)\mathrm{cyp}(A_{J^c}) \\
&\quad = \alpha A_{n-1,n}A_{n,n-1}\mathrm{per}_\alpha(A_{I_{n-2}}) + \alpha A_{n-2,n}A_{n,n-2}\mathrm{per}_\alpha(A_{(1,\ldots,n-3,n-1)}) \\
&\qquad + \sum_{r=0}^{n-3} \alpha\mathrm{per}_\alpha(A_{I_r})\mathrm{cyp}(A_{I_r^c}).
\end{aligned}$$

If we consider $\mathrm{per}_\alpha(A_{(1,\ldots,n-3,n-1)})$ similar arguments as above yield

$$\mathrm{per}_\alpha(A_{(1,\ldots,n-3,n-1)}) = \alpha A_{n-1,n-1}\mathrm{per}_\alpha(A_{I_{n-3}}) + \alpha A_{n-3,n-1}A_{n-1,n-3}\mathrm{per}_\alpha(A_{I_{n-4}}).$$

Finally we need to analyze the sum of cyclic products $\mathrm{cyp}(A_{I_r^c})$ when $0 \le r \le n-3$. In this case when $n-r$ is even the only two non-zero terms in the sum are

$$\mathrm{cyp}(r,n) = A_{r+1,r+2}A_{r+2,r+4}\cdots A_{n-2,n}A_{n,n-1}A_{n-1,n-3}\cdots A_{r+5,r+3}A_{r+3,r+1},$$

$$\mathrm{cyp}^t(r,n) = A_{r+1,r+3}A_{r+3,r+5}\cdots A_{n-3,n-1}A_{n-1,n}A_{n,n-2}\cdots A_{r+4,r+2}A_{r+2,r+1},$$

and when $n-r$ is odd

$$\mathrm{cyp}(r,n) = A_{r+1,r+2}A_{r+2,r+4}\cdots A_{n-3,n-1}A_{n-1,n}A_{n,n-2}\cdots A_{r+5,r+3}A_{r+3,r+1},$$

$$\mathrm{cyp}^t(r,n) = A_{r+1,r+3}A_{r+3,r+5}\cdots A_{n-2,n}A_{n,n-1}A_{n-1,n-3}\cdots A_{r+4,r+2}A_{r+2,r+1}.$$

To ease the notation we let

$$\begin{aligned} a_i &= A_{i,i}, \quad i = 1,\ldots,n \\ b_i &= A_{i,i+1}A_{i+1,i}, \quad i = 1,\ldots,n-1 \\ \beta_i &= A_{i,i+2}A_{i+2,i}, \quad i = 1,\ldots,n-2 \\ p_i^\alpha &= \mathrm{per}_\alpha(A_{I_i}), \quad i = 0,\ldots,n, \end{aligned}$$

when stating the formula in the following corollary.

Corollary 2. Let A be a $n \times n$ penta-diagonal matrix with $n > 3$. Then, using the notation from above,

$$p_n^\alpha = \alpha a_n p_{n-1}^\alpha + \alpha b_{n-1}p_{n-2}^\alpha + \alpha^2\beta_{n-2}a_{n-1}p_{n-3}^\alpha + \alpha^2\beta_{n-2}\beta_{n-3}p_{n-4}^\alpha + \sum_{r=0}^{n-3}\alpha p_r^\alpha(\mathrm{cyp}(r,n)+\mathrm{cyp}^t(r,n)).$$

This gives an easy way to recursively calculate the α-permanent of a penta-diagonal matrix, and if we also assume that $b_i \neq 0$, $i = 1,\ldots,n-1$, we can simplify the calculations further. This follows the exact same lines as for the regular determinant in Sweet (1969), and we leave out the details in the following. The key idea is that the cyclic products of length greater than three can be written in terms of shorter cyclic products. Using the notation

$$\begin{aligned} c_i &= A_{i,i+1}A_{i+1,i+2}A_{i+2,i}, \\ c_i^t &= A_{i,i+2}A_{i+2,i+1}A_{i+1,i}, \end{aligned}$$

for $i = 1,\ldots,n-2$ we have that

$$\mathrm{cyp}(r,n) = \begin{cases} \frac{c_{r+1}c_{r+2}^t\cdots c_{n-3}c_{n-2}^t}{b_2b_3\cdots b_{n-2}} & \text{for } n-r \text{ even} \\ \frac{c_{r+1}c_{r+2}^t\cdots c_{n-4}c_{n-3}^tc_{n-2}}{b_2b_3\cdots b_{n-2}} & \text{for } n-r \text{ odd.} \end{cases}$$

The recursive algorithm for calculating the α-permanent is then:

Set $p_{-1} = 0$, $p_0 = 1$, $p_1 = \alpha a_{1,1}$, $p_2 = \alpha^2 a_{1,1}a_{2,2} + \alpha b_1$, $\epsilon_{-1} = e_{-1} = 0$, and compute

$$\begin{aligned}
\rho_{k-2} &= a_{k-1}p_{k-3} + \beta_{k-3}p_{k-4},\\
\epsilon_{k-3} &= p_{k-3} + \frac{c^t_{k-3}}{b_{k-2}}e_{k-4},\\
e_{k-3} &= p_{k-3} + \frac{c_{k-3}}{b_{k-2}}\epsilon_{k-4},\\
p_k &= \alpha a_k p_{k-1} + \alpha b_{k-1}p_{k-2} + \alpha^2\beta_{k-2}\rho_{k-2} + \alpha c_{k-2}\epsilon_{k-3} + \alpha c^t_{k-2}e_{k-3}.
\end{aligned}$$

4.A.4 Approximating the α-permanent

As mentioned previously, the exact calculation of the α-permanent is in general computationally intractable apart from the special cases treated in the previous sections, but it can be approximated using the importance sampling scheme of Kou and McCullagh (2009). Using this method approximation of the α-permanent in e.g. the log-likelihood (4.5) is feasible for datasets with a moderate total number of counts $n_\star$ (of the order a couple of hundreds). In the following we will discuss how the introduction of a control variate (see Hammersley and Handscomb (1964)) potentially can improve the performance of the algorithm.

To approximate $\mathrm{per}_\alpha(A)$ for a given $n \times n$ matrix A the algorithm uses permutations $\sigma_1, \ldots, \sigma_N \in \mathcal{S}_n$ independently sampled from a certain probability distribution $f(\sigma)$ on $\mathcal{S}_n$ as detailed in Kou and McCullagh (2009). The unbiased estimate is then

$$X = g(\sigma_1, \ldots, \sigma_N; A, \alpha) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{f(\sigma_i)}\alpha^{n-c(\sigma_i)}A_{1,\sigma_i(1)}A_{2,\sigma_i(2)}\cdots A_{n,\sigma_i(n)}.$$

Now let A' approximate A in some sense and have a form such that $\mathrm{per}_\alpha(A')$ can be calculated efficiently (e.g. a block-diagonal or penta-diagonal approximation as detailed in Sections 4.A.2-4.A.3). Then we use the same set of permutations to form the zero mean random variable $Y = g(\sigma_1, \ldots, \sigma_N; A', \alpha) - \mathrm{per}_\alpha(A')$, and introduce the control variate corrected unbiased estimate of $\mathrm{per}_\alpha(A)$ as $Z = X - \beta Y$. Notice that

$$\sigma_Z^2 = \sigma_X^2 + \beta^2\sigma_Y^2 - 2\beta\rho\sigma_X\sigma_Y,$$

where $\sigma_X^2 = \mathrm{Var}(X)$, $\sigma_Y^2 = \mathrm{Var}(Y)$, $\sigma_Z^2 = \mathrm{Var}(Z)$ and $\rho = \mathrm{Corr}(X, Y)$. Hereby, the optimal value of β minimizing the variance of Z is

$$\hat{\beta} = \rho\frac{\sigma_X}{\sigma_Y}, \tag{4.19}$$

which changes the variation in the estimate of $\mathrm{per}_\alpha(A)$ by a factor $\sigma_Z^2/\sigma_X^2 = 1 - \rho^2$. In Hammersley and Handscomb (1964) the suboptimal fixed value of $\beta = 1$ is used, but we prefer the optimal value (4.19), which only requires the additional calculation of an estimate of ρ. We exemplify the use of control variates in what follows below.

Example using control variates

Consider a multivariate negative binomial distribution of dimension $m = 10$ parametrized by $\alpha = 1$ and C with entries $C_{i,j} = \kappa\rho^{|i-j|}$, where $\kappa = 2$ and $\rho = 0.5$. The probability of observing any given outcome $\boldsymbol{n}$ is given by (4.2), which depends on the α-permanent of the block matrix $\tilde{C}[\boldsymbol{n}]$. We have approximated this α-permanent for different outcomes $\boldsymbol{n}$ using either a penta-diagonal control variate or a block-diagonal control variate. The penta-diagonal matrix is obtained simply by truncating $\tilde{C}[\boldsymbol{n}]$ to be penta-diagonal (i.e. all entries not on the diagonal or the two first super- or sub-diagonals are set to zero). The block-diagonal matrix is obtained by only retaining the five two-dimensional block matrices of sizes $n_{2i-1} + n_{2i}$, $i = 1, \ldots, 5$ along the diagonal of $\tilde{C}[\boldsymbol{n}]$ and setting all other entries to zero. Table 4.2 shows the estimated probability for three different outcomes plus/minus two standard errors. Results are shown for both types of control variates as well as with no control variate using 500 MC samples.

		(1,1,1,1,1,1,1,1,1,1)	(1,3,1,3,1,3,1,3,1,3)	(3,3,3,3,3,3,3,3,3,3)
none	$\beta = 0$	$37.31 \pm 1.34 \times 10^{-8}$	$38.85 \pm 3.87 \times 10^{-10}$	$13.88 \pm 2.21 \times 10^{-11}$
block	$\beta = 1$	$37.70 \pm 1.04 \times 10^{-8}$	$40.19 \pm 2.40 \times 10^{-10}$	$14.48 \pm 1.94 \times 10^{-11}$
	$\beta = \hat{\beta}$	$37.55 \pm 0.75 \times 10^{-8}$	$39.93 \pm 2.28 \times 10^{-10}$	$14.39 \pm 1.93 \times 10^{-11}$
penta	$\beta = 1$	$38.16 \pm 0.02 \times 10^{-8}$	$40.81 \pm 2.45 \times 10^{-10}$	$13.29 \pm 2.00 \times 10^{-11}$
	$\beta = \hat{\beta}$	$38.16 \pm 0.02 \times 10^{-8}$	$40.41 \pm 2.32 \times 10^{-10}$	$13.37 \pm 1.99 \times 10^{-11}$

Table 4.2: Comparison of control variates

CHAPTER 5

Work in progress

This chapter contains joint work with Jesper Møller and Adrian Baddeley, which has been initiated during the PhD study, but at this time is not developed enough to be published.

5.1 A model class for spatial point patterns with real marks

This section contains work on a model class for spatial point patterns with real marks, which we mainly have studied for their possible use in the modeling of forest stands. A typical forest stand dataset consists of the location of trees in a given observation area and usually also a list of marks associated with each tree. Examples of such marks are the height of the tree, the diameter at breast height (DBH) of the tree, the species of the tree, etc. Here we focus on datasets with one non-negative real mark which can be taken to express the size of the tree, whether it be the height, the DBH, or some other measure of tree size. For such a dataset we model the distribution of trees conditionally on the location and size of all bigger trees using a modification of the *conditional intensity function*, which is commonly used when the marks correspond to time.

5.1.1 Conditionally specified models

In this section we firstly review some theory for specifying a point process model through the conditional intensity function λ^*, which can be found in more detail in Daley and Vere-Jones

(2003). Secondly, we study how this theory can be used for modeling forest stand data with marks corresponding to tree size.

We consider a general setting of a finite marked point process $\boldsymbol{Y}$ with locations in a planar Borel set $W \subset \mathbb{R}^2$ of finite area $|W| > 0$ and non-negative real marks. A realization of $\boldsymbol{Y}$ is denoted $\boldsymbol{y} = \{y_1, \ldots, y_n\} = \{(x_1, m_1), \ldots, (x_n, m_n)\}$, where x_i and m_i respectively denote the location and mark of the marked point y_i, $i = 1, \ldots, n$. At first we assume a classical setup where the mark m_i correspond to the time of observation of the location x_i.

The conditional intensity function for $\boldsymbol{Y}$ is a function $\lambda^* : W \times \mathbb{R}_+$, which heuristically has the meaning

$$\lambda^*(x, m)\,\mathrm{d}x\,\mathrm{d}m = \mathrm{E}(N(\mathrm{d}x \times \mathrm{d}m)|\mathcal{H}_m(\boldsymbol{y})),$$

where $\mathcal{H}_m(\boldsymbol{y}) = \{(y_i) \in \boldsymbol{y} : m_i < m\}$ is the history of the process, and $N(\cdot)$ counts the number of points of the process falling in the specified space-time region. If we need to stress the history of the process in the conditional intensity we may also write $\lambda^*(x, m) = \lambda^*(x, m|\mathcal{H}_m(\boldsymbol{y}))$. The conditional intensity function determines the point process completely and therefore a parametric point process model can be constructed by specifying a parametric model for the conditional intensity function. The essential assumption for this type of model to be reasonable is that the distribution of new points is determined from the distribution of the points in the history. When the marks correspond to time, the causal direction of time makes the validity of this model type obvious. However, this conditioning may also be appropriate in other circumstances. Here we exemplify this situation by modeling the position and size (DBH) of trees. In this setup it may be a reasonable approximation to model the distribution of trees with a given DBH conditional on all trees with a larger DBH. This implies the history should be modified to $\mathcal{H}_m(\boldsymbol{y}) = \{y_i \in \boldsymbol{y} : m_i > m\}$ and corresponding minor changes have to be made to the theoretical development of models specified via the conditional intensity, which we will not detail here.

The process is typically only observed for marks in a bounded interval $M = (M_{\min}, M_{\max})$, $0 < M_{\min} < M_{\max} < \infty$. E.g. for marks corresponding to time we only observe the point process over some finite period of time. The likelihood of a model with conditional intensity λ^*_θ parametrized via θ given a realization $\boldsymbol{y} \subset W \times M$ is

$$L(\theta|\boldsymbol{y}) = \exp\left(-\int_{M\times W} \lambda^*(x, m)\ \mathrm{d}x\ \mathrm{d}m\right) \prod_{i=1}^{n} \lambda^*(x_i, m_i).$$

However, for some datasets we do not have information about M, and the endpoints are then parameters that have to be estimated from data. The parameter space for M is denoted $\Pi = \{(a, b) : 0 < a < b < \infty\}$ and the parameter space for θ is denoted Θ. We assume the parameters are variation independent such that the combined parameter space is the product space and $\Theta \times \Pi$. The highest likelihood is then obtained by minimizing the size of the interval M, i.e. we estimate the endpoints by the minimal and maximal observed mark.

To conduct statistical inference for conditionally specified models we need a parametric model for the conditional intensity. We consider a log-linear model of the form

$$\lambda^*(x, m) = \exp[\xi f(x) + \eta g(m) + \theta h^*(x, m)],$$

where the functions f, g, h^* (and their corresponding parameters) are allowed to be multidimensional. The functions f and g respectively model the effect of location and mark independently of the history, whereas h^* is allowed to depend on the history of the process, and models the effect of the history on the current location and mark. The structure of the likelihood and the log-linear form of λ^* for known f, g, h^* allows us to use the Berman-Turner device (Berman and Turner, 1992; Baddeley and Turner, 2000) to find the maximum likelihood estimate $(\hat{\xi}, \hat{\eta}, \hat{\theta})$.

5.1.2 Example

Here we study pairwise interaction processes, meaning that $h^*(x,m) = (h_1^*(x,m), \ldots, h_k^*(x,m))$ is assumed to be of the form

$$h_j^*(x,m) = \sum_{i:m_i<m} \phi_j(\|x - x_i\|, m, m_i)$$

for all $j = 1, \ldots, k$, where k is the dimension of h^*.

For forest stand data with the DBH as marks we consider a pairwise interaction model where h^* is 3 dimensional with

$$\begin{aligned}
\phi_1(\|x - x_i\|, m, m_i) &= \|x - x_i\|^\alpha \\
\phi_2(\|x - x_i\|, m, m_i) &= (m_i - m)^\beta \\
\phi_3(\|x - x_i\|, m, m_i) &= \|x - x_i\|^\alpha (m_i - m)^\beta,
\end{aligned}$$

where $\alpha, \beta \in \mathbb{R}$ are parameters to be estimated. In this example we will assume a spatially homogeneous model by setting $\xi = 0$, and we assume $g(m) = (g_1(m), \ldots, g_{n_g}(m))$ is of the form

$$g_j(m) = \mathbb{I}(M_j \le m < M_{j+1}),$$

where $M_1, \ldots, M_{n_g+1}$ is a set of break points covering the mark range. Furthermore, we introduce the hardcore condition $\mathbb{I}[m + m_i < 2\|x_i - x\| \text{ for all } i : m_i > m]$ to avoid physical impossibilities meaning that the discs $b(x_i, m_i/2)$ are not allowed to overlap. That is we use the model

$$\begin{aligned}
\lambda^*(x, m; \mathcal{H}_m(\boldsymbol{y})) &= \mathbb{I}[m + m_i < 2\|x_i - x\| \text{ for all } i : m_i > m] \quad \times \\
&\exp\left(\eta g(m) - \theta_1 \sum_{i:m_i<m} \|x - x_i\|^\alpha - \theta_2 \sum_{i:m_i<m} (m_i - m)^\beta - \theta_3 \sum_{i:m_i<m} \|x - x_i\|^\alpha (m_i - m)^\beta\right)
\end{aligned}$$

It is noted that the Berman-Turner device only can find MLEs for η, θ for fixed α, β, which can be viewed as nuisance parameters in this context. These must estimated by some other method.

5.1.3 Future work

Simulation from the model suggest that it is possible to produce marked point patterns, which resemble forest stand data well, and inference based on the Berman-Turner device appears to work well for simulated datasets. However, it still remains to investigate how the model works on real datasets, and possibly come up with other interaction functions allowing for more complicated interactions between trees.

5.2 Determinantal point processes

This section contains work on determinantal point processes, which have been extensively studied from a purely probabilistic point of view, but to our knowledge the literature contains no treatment of the statistical aspects of these processes. An excellent survey of the probabilistic properties can be found in Hough et al. (2006), and in the following we only give a short introduction to the processes, and outline how Markov Chain Monte Carlo (MCMC) based Bayesian inference can be carried out.

5.2.1 Definitions and some basic properties

Let $S \subset \mathbb{R}^d$ be a Borel set with Lebesgue measure $|S| \in (0, \infty)$. Assume $C : S \times S \to \mathbb{C}$ is given by

$$C(x, y) = \sum_{k=1}^{\infty} \lambda_k \phi_k(x) \overline{\phi_k(y)}$$

for real $\lambda_k \in [0, 1]$ such that $\sum \lambda_k < \infty$ and orthonormalized functions $\phi_k : S \times S \to \mathbb{C}$, meaning

$$\int_S \phi_k(x) \overline{\phi_l(x)} \, \mathrm{d}x = \begin{cases} 1 & \text{if } k = l \\ 0 & \text{if } k \neq l, \end{cases}$$

for $k = 1, 2, \ldots$.

Let $[C](x_1, \ldots, x_n)$ denote the $n \times n$ matrix with (i, j)'th entry $C(x_i, x_j)$. Then a point process X on S is called a *determinantal point process* with kernel C if the n'th order product density function for X is given by

$$\rho^{(n)}(x_1, \ldots, x_n) = \det[C](x_1, \ldots, x_n), \quad (x_1, \ldots, x_n) \in S^n, \tag{5.1}$$

for all $n \in \mathbb{N}$. This is denoted $X \sim \text{determinantal}(C)$.

Note that the intensity function for a determinantal point process is

$$\rho(x) = C(x, x) = \sum_{k=1}^{\infty} \lambda_k |\phi_k(x)|^2, \quad x \in S.$$

Hence, if $|\phi_k(x)| \equiv 1$ as e.g. for the Fourier basis on a rectangular region, the intensity is constant $\rho(x) \equiv \sum_{k=1}^{\infty} \lambda_k$. Furthermore, the pair-correlation function is

$$g(x,y) = 1 - \frac{|C(x,y)|^2}{C(x,x)C(y,y)},$$

making it clear that a determinantal point process models repulsive interactions.

It follows immediately from (5.1) that if $X_0 \sim$ determinantal(C_0) and X is obtained as an independent thinning of X_0 with retention probabilities $p(x)$, $x \in S$, then $X \sim$ determinantal(C) with $C(x,y) = \sqrt{p(x)}C_0(x,y)\sqrt{p(y)}$.

For $k = 1, 2, \ldots$, let B_k be independent Bernoulli variables with mean λ_k. Define the random orthogonal projection kernel $K : S \times S \to \mathbb{C}$ by

$$K(x,y) = \sum_{k=1}^{\infty} B_k \phi_k(x)\overline{\phi_k(y)}.$$

Then

$$\text{determinantal}(K) \sim \text{determinantal}(C), \tag{5.2}$$

in the sense that if we first generate the independent Bernoulli variables, and second independently generate a determinantal point process with kernel K, then the resulting point process is determinantal with kernel C. Note that $\sum \lambda_k < \infty$ implies $\sum B_k < \infty$ with probability one. Thus, to simulate an arbitrary determinantal point process it is sufficient to be able to simulate a determinantal point process where the kernel defines an orthogonal projection of finite rank $n \in \mathbb{N}$. An algorithm for producing such a simulation and a proof of its validity is given in Hough et al. (2006) in a very general setup. In the following we explain and prove the algorithm using mainly linear algebra, which may make it accessible for a wider range of statisticians.

A projection kernel is of the form

$$K(x,y) = \sum_{k=1}^{n} \phi_k(x)\overline{\phi_k(y)} = \boldsymbol{v}(x)\boldsymbol{v}(y)^*$$

where we let $\boldsymbol{v}(x) = (\phi_1(x), \ldots, \phi_n(x))$ for all $x \in S$ and * denotes the conjugate transpose of a vector or matrix. A realization of determinantal(K) is generated by the following procedure.

Algorithm 1. For $i = n, \ldots, 1$:

- Let V_i be the $(n-i)\times n$ matrix with rows $\boldsymbol{v}(x_n), \ldots, \boldsymbol{v}(x_{i+1})$ and define the orthogonal projection matrices

  $$Q_i = V_i^*(V_i V_i^*)^{-1} V_i \quad \text{and} \quad P_i = I_n - Q_i,$$

  where we take Q_n as the $n\times n$ zero matrix.

- Generate x_i from the distribution on S with density

$$p_i(x) = \frac{1}{i}\boldsymbol{v}(x)P_i\boldsymbol{v}(x)^*. \tag{5.3}$$

Note that for $i < n$, we suppress in the notation that $p_i(x) = p_i(x|x_n, \ldots, x_{i+1})$, $Q_i = Q_i(x_n, \ldots, x_{i+1})$, and $P_i = P_i(x_n, \ldots, x_{i+1})$ depend on the previously generated points $x_n, \ldots, x_{i+1}$.

Proposition 1. The random vector $(x_1, \ldots, x_n)$ generated by Algorithm 1 is distributed as a random numbering of the points of determinantal(K).

Proof. First, we show that if P_i is an orthogonal projection of rank i, then

$$p_i(x) = \frac{1}{i}\boldsymbol{v}(x)P_i\boldsymbol{v}(x)^*$$

is indeed a proper probability density. Since P_i is an orthogonal projection of rank i, we have $P_i = U\Lambda_i U^*$, where U is unitary with row vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ and Λ_i is diagonal with the first i diagonal elements equal to one and the rest zero. Then

$$p_i(x) = \frac{1}{i}\boldsymbol{v}(x)U\Lambda_i U^*\boldsymbol{v}(x)^* = \frac{1}{i}\sum_{j=1}^{i}\boldsymbol{v}(x)\boldsymbol{u}_j^*\boldsymbol{u}_j\boldsymbol{v}(x)^* = \frac{1}{i}\sum_{j=1}^{i}|\boldsymbol{v}(x)\boldsymbol{u}_j^*|^2,$$

which is non-negative for all $x \in S$. For each term in the last sum we have

$$\begin{aligned}\int_S |\boldsymbol{v}(x)\boldsymbol{u}_j^*|^2\,\mathrm{d}x &= \int_S \sum_{k=1}^{n}\sum_{l=1}^{n} u_{kj}\phi_k(x)\overline{u}_{lj}\overline{\phi_l(x)}\,\mathrm{d}x = \sum_{k=1}^{n}\sum_{l=1}^{n} u_{kj}\overline{u}_{lj}\int_S \phi_k(x)\overline{\phi_l(x)}\,\mathrm{d}x \\ &= \sum_{k=1}^{n}|u_{kj}|^2\int_S|\phi_k(x)|^2\,\mathrm{d}x = \|\boldsymbol{u}_j\|^2 = 1.\end{aligned}$$

Thereby $\int_S p_i(x)\,\mathrm{d}x = 1$, and so the assertion is verified.

Next, we show that P_i is almost surely an orthogonal projection of rank i. Clearly, this is true for $P_n = I_n$. For $i = n-1, \ldots, 1$, it is clear from the definition that Q_i is the orthogonal projection onto $H_i = \mathrm{span}\{\boldsymbol{v}(x_{i+1}), \ldots, \boldsymbol{v}(x_n)\}$, and P_i is the orthogonal projection onto $H_i^\perp$; for later use, let $H_n = \{0\}$ and $H_n^\perp = \mathbb{R}^n$. Since $\mathrm{rank}(P_i) = \dim(H_i^\perp) = n - \dim(H_i)$, we need to show for $i < n$ that the $n-i$ vectors spanning H_i are linearly independent. Linear dependence would only be introduced if we at the i'th step generate $x' \in \{x|\boldsymbol{v}(x) \in H_i\}$, but then

$$p_i(x') = \frac{1}{i}\boldsymbol{v}(x')P_i\boldsymbol{v}(x')^* = 0,$$

since P_i is the orthogonal projection onto $H_i^\perp$. Thus $\dim(H_i) = n - i$ with probability one.

The density of the random vector $(X_1, \ldots, X_n)$ which the algorithm produces a realization of is derived in the following. Note that

$$p_i(x_i) = \frac{1}{i}\|\boldsymbol{v}(x_i)P_i\|^2,$$

and therefore

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p_i(x_i) = \frac{1}{n!} \prod_{i=1}^{n} \|\boldsymbol{v}(x_i)P_i\|^2.$$

The i'th term in the last product is the squared length of $\boldsymbol{v}(x_i)$'s projection onto $H_i^{\perp}$. This product is exactly the square of the volume of the parallelepiped determined by the vectors $\boldsymbol{v}(x_1), \ldots, \boldsymbol{v}(x_n)$. It is well known that the squared volume of the parallelepiped can be calculated as the determinant of the Gram matrix G with entries

$$G_{ij} = \boldsymbol{v}(x_i)\boldsymbol{v}(x_j)^* = K(x_i, x_j).$$

Thus,

$$p(x_1, \ldots, x_n) = \frac{1}{n!} \det[K](x_1, \ldots, x_n). \tag{5.4}$$

Viewed as a point process $\{X_1, \ldots, X_n\}$, the number of points is fixed and equal to n, and hence by definition of $\rho^{(n)}$,

$$\rho^{(n)}(x_1, \ldots, x_n) = n!p(x_1, \ldots, x_n) = \det[K](x_1, \ldots, x_n).$$

Consequently, the point process is determinantal with kernel K. □

As noticed this simulation procedure always produces n (the rank of the projection) points. This implies that

$$n(X) \sim \sum_{k=1}^{\infty} B_k,$$

for independent Bernoulli variables B_k as defined above. In particular,

$$\mathrm{E}[n(X)] = \sum_{k=1}^{\infty} \lambda_k, \quad \mathrm{Var}[n(X)] = \sum_{k=1}^{\infty} \lambda_k(1 - \lambda_k).$$

5.2.2 Modeling and inference

Let $\phi_k : S \times S \to \mathbb{C}$, $k = 1, 2, \ldots$ be orthonormalized functions as in Section 5.2.1, and let $\lambda_k = \lambda(k; \theta)$ be determined by a (possibly multidimensional) parameter θ. These choices are of course of great importance in regards to which types of determinantal point processes that can be modeled with this approach. Once these have been chosen the main interest is to make inference about the parameter θ, since it completely determines the process.

In the following, we assume $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ is an observed point pattern, and use a Bayesian approach based on MCMC, where the unobserved Bernoulli variables $\boldsymbol{B} = \{B_k\}_{k \in \mathbb{N}}$ are included in the posterior density

$$p(\theta, \boldsymbol{B}|\boldsymbol{x}) \propto p(x_1, \ldots, x_n|\theta, \boldsymbol{B})p(\boldsymbol{B}|\theta)p(\theta).$$

Note that the ordering of the n points plays no role. Furthermore, we make the natural assumption that $p(x_1, \ldots, x_n|\theta, \boldsymbol{B}) = p(x_1, \ldots, x_n|\boldsymbol{B})$ does not depend on θ.

The posterior density in only non-zero when $\sum B_k = n$. We introduce $\boldsymbol{k} = \{k_1, \ldots, k_n\}$ to denote the B_k's that are non-zero. Then the posterior density becomes

$$p(\theta, \boldsymbol{k}|\boldsymbol{x}) \propto p(\boldsymbol{x}|\boldsymbol{k})p(\boldsymbol{k}|\theta)p(\theta).$$

By (5.2),

$$p(\boldsymbol{k}|\theta) = \prod_{k\in\boldsymbol{k}} \lambda_k \prod_{k\notin\boldsymbol{k}} (1-\lambda_k),$$

and by (5.4),

$$p(x_1, \ldots, x_n|\boldsymbol{k}) = \det[K_{\boldsymbol{k}}](x_1, \ldots, x_n),$$

where

$$K_{\boldsymbol{k}}(x, y) = \sum_{k\in\boldsymbol{k}} \phi_k(k)\overline{\phi_k(y)}.$$

From a computational point of view it is worth noting that

$$\det[K_{\boldsymbol{k}}](x_1, \ldots, x_n) = |\det A_{\boldsymbol{k}}(x_1, \ldots, x_n)|^2,$$

where $A_{\boldsymbol{k}}(x_1, \ldots, x_n)$ is the $n\times n$ matrix with (i, j)'th entry $\phi_{k_i}(x_j)$.

To sample from the posterior we construct a Metropolis-within-Gibbs (or hybrid Metropolis-Hastings) algorithm with equilibrium distribution $p(\theta, \boldsymbol{k}|\boldsymbol{x})$. The scheme for proposing updates of θ and $\boldsymbol{k}$ must be decided based on the specific choice of parametric model $\lambda(k; \theta)$ and orthonormal functions ϕ_k, $k = 1, 2, \ldots$.

In the following, we assume an update of $\boldsymbol{k}$ is proposed with probability $p_{\boldsymbol{k}}$ and an update of θ is proposed with probability $1 - p_{\boldsymbol{k}}$. For the update of $\boldsymbol{k}$ we assume an update of each k_i, $i = 1, \ldots, n$ is proposed sequentially. The Hastings ratio H_i related to a proposed update $k_i \to k_i'$ from the proposal density $q_{\boldsymbol{k}}(k_i'; k_i)$ is

$$H_i = \frac{|\det A_{\boldsymbol{k}'}(x_1, \ldots, x_n)|^2}{|\det A_{\boldsymbol{k}}(x_1, \ldots, x_n)|^2} \frac{\lambda(k_i'; \theta)(1-\lambda(k_i; \theta))}{\lambda(k_i; \theta)(1-\lambda(k_i'; \theta))} \frac{q(k_i; k_i')}{q(k_i'; k_i)},$$

where $\boldsymbol{k}' = \boldsymbol{k}^i \cup \{k_i'\}$ with $\boldsymbol{k}^i = \boldsymbol{k} \setminus \{k_i\}$.The Hastings ratio H_θ related to a proposed update $\theta \to \theta'$ from the proposal density $q_\theta(\theta'; \theta)$ is

$$H_\theta = \frac{p(\boldsymbol{k}|\theta')}{p(\boldsymbol{k}|\theta)} \frac{p(\theta')}{p(\theta)} \frac{q(\theta; \theta')}{q(\theta'; \theta)}.$$

5.2.3 Example

Suppose $S = [0, A] \times [0, B]$. For ϕ_k we will use the Fourier functions on S, i.e., all places above we replace the index $k = 1, 2, \ldots$ by the index $k = (k_1, k_2) \in \{0, 1, \ldots\}^2$, and for all $x = (x_1, x_2) \in S$,

$$\phi_{k_1,k_2}(x) = \tfrac{1}{\sqrt{AB}} \exp[2\pi i(k_1 \tfrac{x_1}{A} + k_2 \tfrac{x_2}{B})].$$

For any parametric model $\lambda(k;\theta) = \lambda(k_1, k_2; \theta)$ such that

$$\sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \lambda(k_1, k_2; \theta) < \infty,$$

the process $X \sim \text{determinantal}(C_\theta)$, with

$$C_\theta(x, y) = \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \lambda(k_1, k_2; \theta) \phi_{k_1,k_2}(x) \overline{\phi_{k_1,k_2}(y)}$$

is well-defined.

For practical reasons we would often consider a truncated model such that $\lambda(k_1, k_2; \theta) = 0$ for $(k_1, k_2) \in \{(k_1, k_2) | k_1 \geq N_1 \vee k_2 \geq N_2\}$ where N_1 and N_2 are fixed. To simulate such a process we first generate $B_{k_1,k_2} \sim \text{Bernoulli}(\lambda(k_1, k_2; \theta))$ for $k_1 = 0, \ldots, N_1 - 1$ and $k_2 = 0, \ldots, N_2 - 1$ and then generate the point pattern $\boldsymbol{x}$ using Algorithm 1.

The algorithm requires that we can generate $x_i \in S$ with density (5.3). In the present example we have $\boldsymbol{v}(x)\boldsymbol{v}(x)^* = \|\boldsymbol{v}(x)\|^2 = \frac{n}{AB}$ for all $x \in S$, and the density is given by

$$p_i(x) = \frac{1}{i} \left(\frac{n}{AB} - \|\boldsymbol{v}(x) Q_i\|^2 \right).$$

Consequently, first x_n is drawn from the uniform distribution on S. Then for $i = n-1, \ldots, 1$, x_i is drawn from a non-uniform distribution with a density that attains the maximal value $p_i(x) = \frac{n}{i} \frac{1}{AB}$ on the set $\{x | \boldsymbol{v}(x) \in H_i^\perp\}$ and the minimal value $p_i(x) = 0$ on the set $\{x | \boldsymbol{v}(x) \in H_i\}$, where as in the proof of Proposition 1 we let $H_i = \text{span}\{\boldsymbol{v}(x_{i+1}), \ldots, \boldsymbol{v}(x_n)\}$.

Since we have $p_i(x) \leq \frac{n}{i} \frac{1}{AB}$ for all $x \in S$, we can use rejection sampling to generate a realization x_i with density p_i is the following way. First generate u from the uniform distribution on $(0, 1)$ and x from the uniform distribution on S. If $u < AB \frac{i}{n} p_i(x)$ then x is retained as a realization $x_i = x$ from p_i otherwise new realizations u, x are generated until a proposal is accepted. Notice that the acceptance probability is i/n, making it clear that it becomes progressively harder to generate a realization from p_i as i decreases from n to 1. This is very much in line with the inhibited nature of the determinantal point process.

A simple example of a parametric model is

$$\lambda(k_1, k_2; \theta) = \rho_1^{k_1} \rho_2^{k_2},$$

where $\theta = (\rho_1, \rho_2) \in (0, 1)^2$, and to make the inference likelihood driven the prior $p(\theta) = 1$ can be used.

5.2.4 Future work

Initial experiments with inference for data simulated from the model described in 5.2.3 suggests that the Bayesian MCMC approach works, but much more thorough investigations are

needed to evaluate the method properly. A main issue that needs to be addressed is the problem of slow mixing of the MCMC scheme, which has been observed in some of the initial experiments. Furthermore, it remains to investigate both the flexibility of determinantal point process models and which type of real data sets they are appropriate models for.

Bibliography

Alm, S. (1988). Approximation and simulation of the distributions of scan statistics for Poisson processes in higher dimensions. *Extremes 1*(1), 111–126.

Asmussen, S. and C. A. O'Cinneide (1998). Matrix-exponential distributions. In S. Kotz, C. Read, and D. Banks (Eds.), *Encyclopedia of Statistical Sciences, Supplementary Volume*, New York, pp. 435–440. John Wiley & Sons.

Atkinson, A. (1982). Regression diagnostics, transformations and constructed variables (with discussion). *J. R. Statist. Soc. Ser. B Stat. Methodol. 44*, 1–36.

Augustin, N. H., J. McNicol, and C. A. Marriott (2006). Using the truncated auto-Poisson model for spatially correlated counts of vegetation. *J. Agric. Biol. Environ. Stat. 11*, 1–23.

Baddeley, A. (1980). A limit theorem for statistics of spatial data. *Adv. in Appl. Probab. 12*, 447–461.

Baddeley, A. (1993). Stereology and survey sampling theory. *Bull. Int. Statist. Inst. 50, book 2*, 435–449.

Baddeley, A. (1999). Spatial sampling and censoring. In O. Barndorff-Nielsen, W. Kendall, and M. van Lieshout (Eds.), *Stochastic Geometry: Likelihood and Computation*, Chapter 2, pp. 37–78. London: Chapman and Hall.

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

Baddeley, A., J. Møller, and A. Pakes (2008). Properties of residuals for spatial point processes. *Ann. Inst. Statist. Math. 60*, 627–649.

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

Baddeley, A. and R. Turner (2005). Spatstat: an R package for analyzing spatial point patterns. *J. Statist. Software 12*(6), 1–42. URL: `www.jstatsoft.org`, ISSN: 1548-7660.

Baddeley, A., R. Turner, J. Møller, and M. Hazelton (2005). Residual analysis for spatial point processes (with discussion). *J. R. Statist. Soc. Ser. B Stat. Methodol. 67*(5), 617–666.

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

Barnard, G. (1963). Contribution to discussion of "The spectral analysis of point processes" by M.S. Bartlett. *J. R. Statist. Soc. Ser. B Stat. Methodol. 25*, 294.

Berman, M. (1986). Testing for spatial association between a point process and another stochastic process. *J. Roy. Statist. Soc. Ser. C. 35*, 54–62.

Berman, M. and T. Turner (1992). Approximating point process likelihoods with GLIM. *J. Roy. Statist. Soc. Ser. C. 41*, 31–38.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. R. Statist. Soc. Ser. B Stat. Methodol. 36*, 192–236.

Besag, J. (1978). Some methods of statistical analysis for spatial data. *Bull. Int. Statist. Inst. 44*, 77–92.

Chen, C. (1983). Score tests for regression models. *J. Amer. Statist. Assoc. 78*, 158–161.

Chetwynd, A. and P. Diggle (1998). On estimating the reduced second moment measure of a stationary spatial point process. *Aust. N. Z. J. Stat. 40*, 11–15.

Choo, L. and S. G. Walker (2008). A new approach to investigating spatial variations of disease. *J. Roy. Statist. Soc. Ser. A 171*, 395–405.

Conniffe, D. (2001). Score tests when a nuisance parameter is unidentified under the null hypothesis. *J. Statist. Plann. Inference 97*, 67–83.

Cook, R. and S. Weisberg (1983). Diagnostics for heteroscedasticity in regression. *Biometrika 70*, 1–10.

Cordy, C. (1993). An extension of the Horvitz-Thompson theorem to point sampling from a continuous universe. *Statist. Probab. Lett. 18*, 353–362.

Cox, D. (1972). The statistical analysis of dependencies in point processes. In P. Lewis (Ed.), *Stochastic Point Processes*, pp. 55–66. New York: Wiley.

Cox, D. and D. Hinkley (1974). *Theoretical statistics*. London: Chapman and Hall.

Cox, D. R. and N. Reid (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika 91*, 729–737.

Cressie, N. (1991). *Statistics for Spatial Data*. New York: John Wiley and Sons.

Cressie, N. (1993). *Statistics for Spatial Data* (second ed.). New York: John Wiley and Sons.

Daley, D. and D. Vere-Jones (1988). *An Introduction to the Theory of Point Processes*. New York: Springer Verlag.

Daley, D. and D. Vere-Jones (2003). *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods* (second ed.). New York: Springer Verlag.

Daley, D. J. and D. Vere-Jones (2008). *An Introduction to the Theory of Point Processes. Volume II: General Theory and Structure* (second ed.). New York: Springer Verlag.

Davies, R. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika 64*, 247–254.

Davies, R. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika 74*, 33–43.

Diggle, P. (1979). On parameter estimation and goodness-of-fit testing for spatial point patterns. *Biometrika 35*, 87–101.

Diggle, P. (1985). A kernel method for smoothing point process data. *J. Roy. Statist. Soc. Ser. C. 34*, 138–147.

Diggle, P. (2003). *Statistical Analysis of Spatial Point Patterns* (Second ed.). London: Hodder Arnold.

Georgii, H.-O. (1976). Canonical and grand canonical Gibbs states for continuum systems. *Comm. Math. Phys. 48*, 31–51.

Georgii, H.-O. and H. J. Yoo (2005). Conditional intensity and Gibbsianness of determinantal point processes. *J. Stat. Phys. 118*, 617–666.

Geyer, C. (1999). Likelihood inference for spatial point processes. In O. Barndorff-Nielsen, W. Kendall, and M. van Lieshout (Eds.), *Stochastic Geometry: Likelihood and Computation*, Number 80 in Monographs on Statistics and Applied Probability, Chapter 3, pp. 79–140. Boca Raton, Florida: Chapman and Hall / CRC.

Golub, G. H. and C. F. Van Loan (1996). *Matrix Computations*. Baltimore: Johns Hopkins.

Goulden, I. P. and D. M. Jackson (1983). *Combinatorial Enumeration*. New York: John Wiley & Sons.

Griffiths, R. C. (1984). Characterization of infinitely divisible multivariate gamma distributions. *J. Multivariate Anal. 15*, 13–20.

Griffiths, R. C. and R. K. Milne (1987). A class of infinitely divisible multivariate negative binomial distributions. *J. Multivariate Anal. 22*, 13–23.

Hammersley, J. M. and D. C. Handscomb (1964). *Monte Carlo Methods*. London: Methuen.

Hanisch, K. (1984). Some remarks on estimators of the distribution function of nearest neighbour distance in stationary spatial point patterns. *Math. Operationsforsch. Statist. Ser. Statist. 15*, 409–412.

Heinrich, L. (1988a). Asymptotic behaviour of an empirical nearest-neighbour distance function for stationary Poisson cluster processes. *Math. Nachr. 136*, 131–148.

Heinrich, L. (1988b). Asymptotic Gaussianity of some estimators for reduced factorial moment measures and product densities of stationary Poisson cluster processes. *Statistics 19*, 87–106.

Hope, A. (1968). A simplified Monte Carlo significance test procedure. *J. R. Statist. Soc. Ser. B Stat. Methodol. 30*, 582–598.

Hough, J. B., M. Krishnapur, Y. Peres, and B. Viràg (2006). Determinantal processes and independence. *Probab. Surv. 3*, 206–229.

Huang, F. and Y. Ogata (1999). Improvements of the maximum pseudo-likelihood estimators in various spatial statistical models. *J. Comput. Graph. Statist. 8*(3), 510–530.

Illian, J., A. Penttinen, H. Stoyan, and D. Stoyan (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. Chichester: John Wiley and Sons.

Jensen, J. and J. Møller (1991). Pseudolikelihood for exponential family models of spatial point processes. *Ann. Appl. Probab. 1*, 445–461.

Johnson, M. E. (1987). *Multivariate Statistical Simulation*. New York: John Wiley & Sons, New York.

Jolivet, E. (1980). Central limit theorem and convergence of empirical processes for stationary point processes. In P. Bastfai and J. Tomko (Eds.), *Point processes and queueing problems*, pp. 117–161. Amsterdam: North-Holland.

Kallenberg, O. (1978). On conditional intensities of point processes. *Z. Wahrscheinlichkeit. 41*, 205–220.

Kallenberg, O. (1984). An informal guide to the theory of conditioning in point processes. *Int. Stat. Rev. 52*, 151–164.

Karlin, S. (1968). *Total Positivity*. Stanford: Stanford University Press.

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

Kou, S. C. and P. McCullagh (2009). Approximating the α-permanent. *Biometrika 96*, 635–644.

Krishnamoorthy, A. S. and M. Parthasarathy (1951). A multivariate gamma-type distribution. *Ann. Math. Statist. 22*, 549–557.

Kulldorff, M. (1999). Spatial scan statistics: models, calculations, and applications. In J. Glaz and N. Balakrishnan (Eds.), *Recent Advances on Scan Statistics*, pp. 303–322. Boston: Birkhauser.

Kutoyants, Y. (1998). *Statistical Inference for Spatial Poisson Processes*. Number 134 in Lecture Notes in Statistics. New York: Springer.

Lantuejoul, C. (2002). *Geostatistical Simulation: Models and Algorithms*. Berlin: Springer-Verlag.

Last, G. and R. Penrose (2010). Poisson process Fock space representation, chaos expansion and covariance inequalities. Manuscript in preparation.

Lawson, A. (1993). On the analysis of mortality events around a prespecified fixed point. *J. Roy. Statist. Soc. Ser. A 156*(3), 363–377.

Lindsay, B. G. (1988). Composite likelihood methods. In N. U. Prabhu (Ed.), *Statistical Inference from Stochastic Processes*, Providence, pp. 221–239. American Mathematical Society.

Lotwick, H. W. and B. W. Silverman (1982). Methods for analysing spatial processes of several types of points. *J. R. Statist. Soc. Ser. B Stat. Methodol. 44*, 406–413.

Macchi, O. (1971). Stochastic point processes and multicoincidences. *IEEE Trans. Inform. Theory 17*, 2–7.

Macchi, O. (1975). The coincidence approach to stochastic point processes. *Adv. in Appl. Probab. 7*, 83–122.

Maybee, J. and J. Quirk (1969). Qualitative problems in matrix theory. *SIAM Rev. 11*, 30–51.

McCullagh, P. (1983). Quasi-likelihood functions. *Ann. Statist. 11*, 59–67.

McCullagh, P. and J. Møller (2006). The permanental process. *Adv. in Appl. Probab. 38*, 873–888.

Minc, H. (1978). *Permanents*. Reading, MA: Addison-Wesley.

Møller, J. (2003). A comparison of spatial point process models in epidemiological applications. In P. Green, N. Hjort, and S. Richardson (Eds.), *Highly Structured Stochastic Systems*, pp. 264–268. Oxford: Oxford University Press.

Møller, J. and E. Rubak (2010). A model for positively correlated count variables. *Int. Stat. Rev. 78*(1), 65–80.

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

Møller, J. and R. Waagepetersen (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Boca Raton: Chapman and Hall/CRC.

Møller, J. and R. P. Waagepetersen (2007). Modern spatial point process modelling and inference (with discussion). *Scand. J. Statist. 34*, 643–711.

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

Numata, M. (1961). Forest vegetation in the vicinity of Choshi — coastal flora and vegetation at Choshi, Chiba prefecture, IV (in Japanese). *Bull. Choshi Mar. Lab.* (3), 28–48. Chiba University.

Numata, M. (1964). Forest vegetation, particularly pine stands in the vicinity of Choshi — flora and vegetation in Choshi, Chiba prefecture, VI (in Japanese). *Bull. Choshi Mar. Lab.* (6), 27–37. Chiba University.

Ogata, Y. and M. Tanemura (1981). Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure. *Ann. Inst. Statist. Math. B 33*, 315–338.

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

Ohser, J. (1983). On estimators for the reduced second moment measure of point processes. *Math. Operationsforsch. Statist. Ser. Statist. 14*, 63–71.

Ohser, J. and D. Stoyan (1981). On the second-order and orientation analysis of planar stationary point processes. *Biom. J. 23*, 523–533.

Papangelou, F. (1974). The conditional intensity of general point processes and an application to line processes. *Z. Wahrscheinlichkeit. 28*, 207–226.

Pregibon, D. (1982). Score tests in GLIM with applications. In *GLIM 82: Proceedings of the International Conference on Generalized Linear Models*, Number 14 in Lecture Notes in Statistics, New York. Springer.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Rao, C. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Math. Proc. Cambridge Philos. Soc. 44*, 50–57.

Rathbun, S. and N. Cressie (1994). Asymptotic properties of estimators of the parameters of spatial inhomogeneous Poisson point processes. *Adv. in Appl. Probab. 26*, 122–154.

Richardson, S. (2003). Spatial models in epidemiological applications. In P. Green, N. Hjort, and S. Richardson (Eds.), *Highly Structured Stochastic Systems*, pp. 237–259. Oxford: Oxford University Press.

Ripley, B. (1976). The second-order analysis of stationary point processes. *J. Appl. Probab. 13*, 255–266.

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

Ripley, B. (1988). *Statistical Inference for Spatial Processes*. Cambridge University Press.

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

Schladitz, K. and A. Baddeley (2000). A third order point process characteristic. *Scand. J. Statist. 27*, 657–671.

Shirai, T. (2007). Remarks on the positivity of α-determinants. *Kyushu J. Math. 61*, 169–189.

Shirai, T. and Y. Takahashi (2003a). Random point fields associated with certain Fredholm determinants I: fermion, Poisson and boson point processes. *J. Func. Anal. 205*, 414–463.

Shirai, T. and Y. Takahashi (2003b). Random point fields associated with certain Fredholm determinants II: fermion shifts and their ergodic and Gibbs properties. *Ann. Probab. 31*, 1533–1564.

Silvapulle, M. (1996). A test in the presence of nuisance parameters. *J. Amer. Statist. Assoc. 91*, 1690–1693.

Stillinger, F., D. Stillinger, S. Torquato, T. Truskett, and P. Debenedetti (2000). Triangle distribution and equation of state for classical rigid disks. *J. Stat. Phys. 100*, 49–72.

Stoyan, D., W. Kendall, and J. Mecke (1987). *Stochastic Geometry and its Applications*. Chichester: John Wiley and Sons.

Stoyan, D. and H. Stoyan (1995). *Fractals, Random Shapes and Point Fields*. Chichester: John Wiley and Sons.

Strauss, D. (1975). A model for clustering. *Biometrika 63*, 467–475.

Sweet, R. A. (1969). A recursive relation for the determinant of a pentadiagonal matrix. *Communications of the ACM 12*, 330–332.

Valiant, L. G. (1979). The complexity of cumputing the permanent. *Theoret. Comput. Sci. 8*(2), 189–201.

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

Venables, W. N. and B. D. Ripley (2002). *Modern Applied Statistics with S*. New York: Springer.

Vere-Jones, D. (1997). Alpha-permanents and their applications to multivariate gamma, negative binomial and ordinary binomial distributions. *New Zealand J. Math. 26*, 125–149.

Wald, A. (1941). Some examples of asymptotically most powerful tests. *Ann. Math. Statist. 12*, 396–408.

Waller, L., B. Turnbull, L. Clark, and P. Nasca (1992). Chronic disease surveillance and testing of clustering of disease and exposure: Application to leukaemia incidence and TCE-contaminated dumpsites in upstate New York. *Environmetrics 3*, 281–300.

Wang, P. (1985). Adding a variable in generalized linear models. *Technometrics 27*, 273–276.

Wedderburn, R. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika 61*, 439–447.

Widom, B. and J. Rowlinson (1970). New model for the study of liquid-vapor phase transitions. *J. Chem. Phys. 52*, 1670–1684.

Wu, L. (2000). A new modified logarithmic Sobolev inequality for Poisson point processes and several applications. *Probab. Theory Related Fields 118*, 427–438.