Lecture Notes in Statistics **185**

Edited by P. Bickel, P. Diggle, S. Fienberg, U. Gather,
I. Olkin, S. Zeger

Adrian Baddeley
Pablo Gregori
Jorge Mateu
Radu Stoica
Dietrich Stoyan
(Editors)

Case Studies in Spatial Point Process Modeling

With 107 Figures

Springer

Adrian Baddeley
Department of Mathematics
University of Western Australia
Nedlands 6907 Australia
adrian@maths.uwa.edu.au

Pablo Gregori
Department of Mathematics
Universitat Jaume 1 of Castellon
Castellon 12071 Spain
gregori@mat.uji.es

Jorge Mateu
Department of Mathematics
Universitat Jaume 1 of Castellon
Castellon 12071 Spain
mateu@mat.uji.es

Radu Stoica
INRA - Biometrie, Domaine St. Paul,
Site Agroparc
84914 Avignon, Cedex 9, France
Radu.Stoica@avignon.inra.fr

Dietrich Stoyan
Institut für Stochastik
Prüferstraße 9
TU Bergakademie Freiberg
D-09596 Freiberg
Germany
stoyan@orion.hrz.tu-freiberg.de

Library of Congress Control Number: 2005931123

ISBN-10: 0-387-28311-0
ISBN-13: 978-0387-28311-1

Printed on acid-free paper.

Camera ready copy provided by the editors.

Printed in the United States of America. (SBA)

9 8 7 6 5 4 3 2 1

springeronline.com

Preface

The week before Easter 2004 a conference on spatial point process modelling and its applications was held in Benicàssim (Castellón, Spain). The organizers targeted two aims. The first goal was to bring together most of the known people to guarantee the high scientific quality of the meeting to foster the theoretical and practical use of spatial point processes. The second one consisted of enabling young researchers to present their work and to obtain a valuable feed-back coming from the reknown specialists in the domain. The contributions of all the participants were published in the proceedings book of the conference.

The majority of the contributions in this book represents the reviewed version of the papers presented during the conference. In order to offer the reader a larger spectrum of this domain, authors that could not attend the conference were also invited to contribute.

The book is constituted by 16 chapters divided in three parts and gathering 44 authors coming from 13 different countries.

The first part of the volume – represented by its two first contributions – is dedicated to basic notions and tools for understanding and manipulating spatial point processes.

In the first contribution, D. Stoyan presents a general overview of the theoretical foundations for spatial point process. The author defines a point process and a marked point process, and describes the construction of the first and second order moment measures, which leads to the nowadays well known summary statistics such as the K-function, L-function or the pair-correlation function. The Poisson point process plays an important role, since in practice it is often used as null model for hypothesis testing and as reference model for the construction of realistic models for point patterns.

The second contribution, written by A.J. Baddeley and R. Turner, enters directly in the "flesh" of the problem presenting the concrete use of spatial point processes for modelling spatial point patterns, via the `spatstat` package – a software library for the `R` language. Four main points can be tackled by this package: basic manipulation of point patterns, exploratory data analysis, parametric model-fitting and simulation of spatial point processes. The very important issue of model validation is also addressed. The contribution contains also the necessary mathematical details and/or literature references in order to avoid the use of this software as a "black box". Two complete case studies are presented at the end of the contribution.

There is no serious practical application without a rigourous theoretical development. Therefore the second part of the book is more oriented towards theoretical and methodological advances in spatial point processes theory. Topics of this part of the book contain *analytical properties of the Poisson process* (presented in the contribution by S. Zuyev), *Bayesian analysis of Markov point processes* (by K. K. Berthelsen and J. Møller), *statistics for locally scaled point processes* (by M. Prokĕsová, U. Hahn and E. B. Vedel Jensen), *nonparametric testing of distribution functions in germ-grain models* (by Z. Pawlas and L. Heinrich), and *principal component analysis applied to point processes through a simulation study* (by J. Illian, E. Benson, J. Crawford and H. Staines). Remarkable is the fact, that almost all these contributions show direct applications of the presented development.

The third part of this volume is entirely dedicated to concrete, precise case studies, that are solved within the point processes theory. The presented applications are of big impact: *material science* (by F. Ballani), *human epidemiology* (by M. A. Martínez-Beneito *et al.*) , *social sciences* (by N.A.C. Cressie, O. Perrin and C. Thomas-Agnan), *animal epidemiology* (by Webster *et al.* and P.J. Diggle, S. J. Eglen and J. B. Troy), *biology* (by F. Fleischer *et al.* and by A. Stein and N. Georgiadis), and *seismology* (by J. Zhuang, Y. Ogata and D. Vere-Jones and by A. Veen and F.P. Schoenberg). In their contributions, the authors show skill and cleverness in using, combining and continuously evolving the point processes tools in order to answer the proposed questions.

We hope the reader will enjoy reading the book and will find it instructive and inspiring for going a step further in this very open research field.

The Editors are grateful to all the authors that made possible finishing the book within an acceptable time scheduling. A word of thanks is given to Springer-Verlag and, in particular, to John Kimmel for creating the opportunity of making this project real.

Castellón (Spain)
May 2005

Adrian Baddeley
Pablo Gregori
Jorge Mateu
Radu Stoica
Dietrich Stoyan
Editors

Contents

List of Contributors

1. **Juan J. Abellán** Small Area Health Statistics Unit, Imperial College of London, UK
2. **Adrian Baddeley** School of Mathematics and Statistics, University of Western Australia, Crawley 6009 WA, Australia
3. **Felix Ballani** Institut für Stochastik, Technische Universität Bergakademie Freiberg, Prüferstr. 9, 09596 Freiberg, Germany
4. **Michael Beil** Department of Internal Medicine I, University Hospital Ulm, D-89070 Ulm, Germany
5. **Erica Benson** SIMBIOS, University of Abertay Dundee, Scotland, UK
6. **Kasper K. Berthelsen** Aalborg University, Department of Mathematical Sciences, F. Bajers Vej 7G, DK-9220 Aalborg, Denmark
7. **Helen E. Clough** Department of Veterinary Clinical Sciences, University of Liverpool, UK
8. **John Crawford** SIMBIOS, University of Abertay Dundee, Scotland, UK
9. **Noel A.C. Cressie** Department of Statistics, The Ohio State University, Columbus, OH 43210, USA
10. **Peter J. Diggle** Department of Statistics, Lancaster University, UK
11. **Stephen J. Eglen** University of Cambridge, UK
12. **José Fenollar** Centro de Salud Pública de Alcoi, Generalitat Valenciana, Spain
13. **Frank Fleischer** Department of Applied Information Processing and Department of Stochastics, University of Ulm, D-89069 Ulm, Germany
14. **Nigel P. French** Institute of Veterinary, Animal and Biomedical Sciences, Massey University, New Zealand.
15. **Nick Georgiadis** Mpala Research Centre, PO Box 555, Nanyuki, Kenya
16. **Robert B. Green** Department of Neuropathology, Veterinary Laboratories Agency, UK
17. **Ute Hahn** University of Augsburg, Department of Applied Stochastics, 86135 Augsburg, Germany
18. **Lothar Heinrich** University of Augsburg, Institute of Mathematics. Universitätsstr. 14, D-86135 Augsburg, Germany

19. **Janine Illian** SIMBIOS, University of Abertay Dundee, Scotland, UK
20. **Eva B. Vedel Jensen** University of Aarhus, The T.N. Thiele Centre of Applied Mathematics in Natural Science, Department of Mathematical Sciences, Ny Munkegade, 8600 Aarhus C, Denmark
21. **Guillermo Jorques** Centro de Salud Pública de Alcoi, Generalitat Valenciana, Spain
22. **Marian Kazda** Department of Systematic Botany and Ecology, University of Ulm, D-89069 Ulm, Germany
23. **Antonio López-Quílez** Departament d'Estadística i Investigació Operativa, Universitat de València, Spain
24. **Miguel A. Martínez-Beneito** Dirección General de Salud Pública, Generalitat Valenciana, Spain
25. **Jesper Møller** Aalborg University, Department of Mathematical Sciences, F. Bajers Vej 7G, DK-9220 Aalborg, Denmark
26. **Yosihiko Ogata** Institute of Statistical Mathematics, Japan
27. **Zbyněk Pawlas** Charles University, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics. Sokolovská 83, 186 75 Praha 8, Czech Republic
28. **Olivier Perrin** GREMAQ – Université Toulouse 1 (and LERNA-INRA), 21 allée de Brienne, 31000 Toulouse, France
29. **Michaela Prokešová** Charles University, Department of Probability, Sokolovská 83, 18675 Praha 8, Czech Republic
30. **Volker Schmidt** Department of Stochastics, University of Ulm, D-89069 Ulm, Germany
31. **Frederic Paik Schoenberg** UCLA Department of Statistics, 8125, Math Sciences Building, Box 951554, Los Angeles, CA 90095-1554, USA
32. **Harry Staines** SIMBIOS, University of Abertay Dundee, Scotland, UK
33. **Alfred Stein** International Institute for Geo-Information Science and Earth Observation (ITC), PO Box 6, 7500 AA Enschede, The Netherlands
34. **Dietrich Stoyan** Institut für Stochastik, Technische Universität Bergakademie Freiberg, Agricolastr. 1, 09596 Freiberg, Germany
35. **Christine Thomas-Agnan** GREMAQ – Université Toulouse 1 (and LSP – Université Toulouse 3), 21 allée de Brienne, 31000 Toulouse, France
36. **John B. Troy** Northwestern University, USA
37. **Rolf Turner** Department of Mathematics and Statistics, University of New Brunswick, Fredericton, N.B., E3B 5A3 Canada
38. **Hermelinda Vanaclocha** Dirección General de Salud Pública, Generalitat Valenciana, Spain
39. **Alejandro Veen** UCLA Department of Statistics, 8125 Math Sciences Building, Box 951554, Los Angeles, CA 90095-1554, USA
40. **David Vere-Jones** Victoria University of Wellington, New Zealand
41. **Simon Webster** Department of Veterinary Clinical Sciences, University of Liverpool, UK
42. **Jiancang Zhuang** Institute of Statistical Mathematics, Japan

43. **Óscar Zurriaga** Dirección General de Salud Pública, Generalitat Valenciana, Spain
44. **Sergei Zuyev** Department of Statistics and Modelling Science, University of Strathclyde, Glasgow, G1 1XH, UK

Fundamentals of Point Process Statistics

Dietrich Stoyan

Institut für Stochastik, Technische Universität Bergakademie Freiberg,
Agricolastr. 1, 09596 Freiberg, Germany, stoyan@orion.hrz.tu-freiberg.de

Summary. Point processes are mathematical models for irregular or random point patterns. A short introduction to the theory of point processes and their statistics, emphasizing connections between the presented theory and the use done by several authors and contributions appearing in this book is presented.

Key words: Marked point processes, Second-order characteristics, Spatial point processes overview, Statistical inference

1 Basic Notions of the Theory of Spatial Point Processes

The following text is a short introduction to the theory of point processes and their statistics, written mainly in order to make this book self-contained. For more information the reader is referred to the text [2] and the books [7, 16, 20, 21].

Point processes are mathematical models for irregular or random point patterns. The mathematical definition of a point process on $\mathbb{R}^d$ is as a random variable N taking values in the measurable space $[\mathbb{N}, \mathcal{N}]$, where $\mathbb{N}$ is the family of all sequences $\{x_n\}$ of points of $\mathbb{R}^d$ satisfying the local finiteness condition, which means that each bounded subset of $\mathbb{R}^d$ contains only a finite number of points. In this book only simple point processes are considered, i.e. $x_i \neq x_j$ if $i \neq j$.

The order of the points x_n is without interest, only the set $\{x_n\}$ matters. Thus the x_n are dummy variables and have no particular interpretation; for example x_1 need not be the point closest to the origin o.

The σ-algebra $\mathcal{N}$ is defined as the smallest σ-algebra of subsets of $\mathbb{N}$ to make measurable all mappings $\varphi \mapsto \varphi(B)$, for B running through the bounded Borel sets.

The reader should note that the term "process" does not imply a dynamic evolution over time and therefore the phrase "random point field" would be a more exact term; it is used in [21]. *Spatio-temporal point processes* explicitly involving temporal as well as spatial dispersion of points constitute a separate theory, see the paper by Zhuang et al. in this volume.

The *distribution* of a point process N is determined by the probabilities

$$\mathbf{P}(N \in Y) \qquad \text{for } \; Y \in \mathcal{N} .$$

The *finite-dimensional distributions* are of particular importance. These are probabilities of the form

$$\mathbf{P}(N(B_1) = n_1, \ldots, N(B_k) = n_k)$$

where $B_1, \ldots, B_k$ are bounded Borel sets and $n_1, \ldots, n_k$ non-negative integers. Here $N(B_i)$ is the number of points of N in B_i. The distribution of N on $[\mathbb{N}, \mathcal{N}]$ is uniquely determined by the system of all these values for $k = 1, 2, \ldots$. A still smaller subsystem is that of the *void probabilities*

$$v_B = \mathbf{P}(N(B) = 0) = \mathbf{P}(N \cap B = \emptyset)$$

for Borel sets B. Here N denotes the set of all points of the point process, the so-called support. If the point process is simple, as assumed here, then the distribution of N is already determined by the system of values of v_K as K ranges through the compact sets.

Let B be a convex compact Borel set in $\mathbb{R}^d$ with o being an inner point of B. The *contact distribution function* H_B with respect to the *test set* B is defined by

$$H_B(r) = 1 - \mathbf{P}(N(rB)) = 0) \qquad \text{for } r \geq 0 . \tag{1}$$

In the special case of $rB = b(o, r) =$ sphere of radius r centred at o the contact distribution function is denoted as $F(r)$ or $H_s(r)$ and called the *spherical contact distribution function* or *empty space distribution function*. It can be interpreted as the distribution function of the random distance from the origin of $\mathbb{R}^d$ to the closest point of N. The function $H_B(r)$ is of a similar nature, but the metric is given by B.

A point process N is said to be *stationary* if its characteristics are invariant under translation: the processes $N = \{x_n\}$ and $N_x = \{x_n + x\}$ have the same distribution for all x in $\mathbb{R}^d$. So

$$\mathbf{P}(N \in Y) = \mathbf{P}(N_x \in Y) \tag{2}$$

for all Y in $\mathcal{N}$ and all x in $\mathbb{R}^d$. If we put $Y_x = \{\varphi \in \mathbb{N} : \varphi_{-x} \in Y\}$ for $Y \in \mathcal{N}$ then equation (2) can be rewritten as

$$\mathbf{P}(N \in Y) = \mathbf{P}(N \in Y_{-x}) .$$

The notion of *isotropy* is entirely analogous: N is isotropic if its characteristics are invariant under rotation. Stationarity and isotropy together yield *motion-invariance*. The assumption of stationarity simplifies drastically the statistics of point patterns and therefore many papers in this book assume at least stationarity.

The *intensity measure* Λ of N is a characteristic analogous to the mean of a real-valued random variable. Its definition is

$$\Lambda(B) = \mathbf{E}\,(N(B)) \qquad \text{for Borel } B\,. \tag{3}$$

So $\Lambda(B)$ is the mean number of points in B. If N is stationary then the intensity measure simplifies; it is a multiple of Lebesgue measure ν_d, i.e.

$$\Lambda(B) = \lambda \nu_d(B) \tag{4}$$

for some (possibly infinite) non-negative constant λ, which is called the *intensity* of N it can be interpreted as the mean number of points of N per unit volume.

The point-related counterpart to $F(r)$ or $H_s(r)$ in the stationary case is the *nearest neighbour distance distribution function* $G(r)$ or $D(r)$, i.e. the d.f. of the distance from the typical point of N to its nearest neighbour.

Note that the application of $F(r)$ and $G(r)$ in the characterization of point processes is different. This is particularly important for cluster processes. In such cases $G(r)$ mainly describes distributional aspects in the clusters, while $F(r)$ characterizes particularly the empty space between the clusters. This different behaviour also explains the success of the *J-function* introduced by [12] defined as

$$J(r) = \frac{1 - G(r)}{1 - F(r)} \quad \text{for} \quad r \geq 0. \tag{5}$$

2 Marked Point Processes

A point process is made into a *marked point process* by attaching a characteristic (the *mark*) to each point of the process. Thus a marked point process on $\mathbb{R}^d$ is a random sequence $M = \{[x_n; m_n]\}$ from which the points x_n together constitute a point process (not marked) in $\mathbb{R}^d$ and the m_n are the marks corresponding to the x_n. The marks m_n may have a complicated structure. They belong to a given *space of marks* $\mathbb{M}$ which is assumed to be a Polish space. The Borel σ-algebra of $\mathbb{M}$ is denoted by $\mathcal{M}$. Specific examples or marked points are:

- For x the centre of a particle, m the volume of the particle;
- For x the position of a tree, m the stem diameter of the tree;
- For x the centre of an atom, m the type of the atom;
- For x the location (suitably defined) of a convex compact set, m the centred (shifted to origin) set itself.

Point process statistics often uses constructed marks. Examples are:

- $m =$ distance to the nearest neighbour of x;
- $m =$ number of points within distance r from x.

The marks can be continuous variables, as in the first two examples, indicators of types as in the third example (in which case the terms “multivariate point process” or “multitype point process” are often used, in the case of two marks the term “bivariate point processes”) or actually very complicated indeed, as in the last example which occurs in the marked point process interpretation of germ-grain models (see [20]).

There is a particular feature of marked point processes: Euclidean motions of marked point processes are defined as transforms which move the points but leave the marks unchanged. So M_x, the translate of M by x, is given by

$$M_x = \{[x_1 + x; m_1], [x_2 + x; m_2], \ldots\} .$$

Rotations act on marked point processes by rotating the points but *not* altering the marks.

A marked point process M is said to be *stationary* if for all x the translated process M_x has the same distribution as M. It is *motion-invariant* if for all Euclidean motions m the process mM has the same distribution as M.

The definition of the *intensity measure* Λ of a marked point process M is analogous to that of the intensity measure of M when M is interpreted as a non-marked point process:

$$\Lambda(B \times L) = \mathbf{E}\left(M(B \times L)\right) .$$

When M is stationary

$$\Lambda = \lambda \times \nu_d \times P_M, \tag{6}$$

where P_M denotes the mark distribution.

3 The Second-order Moment Measure

In the classical theory of random variables the moments (particularly mean and variance) are important tools of statistics. Point process theory has analogues to these. However, numerical means and variances must be replaced by the more complicated moment measures.

The second-order factorial moment measure of the point process N is the measure $\alpha^{(2)}$ on $\mathbb{R}^{2d}$ defined by

$$\int\limits_{\mathbb{R}^{2d}} f(x_1, x_2)\ \alpha^{(2)}(\mathrm{d}(x_1, x_2)) = \mathbf{E}\left(\sum_{x_1, x_2 \in N}^{\neq} f(x_1, x_2)\right) \tag{7}$$

where f is any non-negative measurable function on $\mathbb{R}^{2d}$. The sum in (7) is extended over all pairs of different points; this is indicated by the symbol $\sum^{\neq}$.

It is

$$\mathbf{E}\left(N(B_1)N(B_2)\right) = \alpha^{(2)}(B_1 \times B_2) + \Lambda(B_1 \cap B_2)$$

and

$$\mathbf{var}\ (N(B)) \ = \ \alpha^{(2)}(B \times B) + \Lambda(B) - (\Lambda(B))^2\,.$$

If N is stationary then $\alpha^{(2)}$ is translation invariant in an extended sense:

$$\alpha^{(2)}(B_1 \times B_2) = \alpha^{(2)}((B_1 + x) \times (B_2 + x))$$

for all x in $\mathbb{R}^d$.

Suppose that $\alpha^{(2)}$ is locally finite and absolutely continuous with respect to Lebesgue measure ν_{2d}. Then $\alpha^{(2)}$ has a density $\varrho^{(2)}$, the *second-order product density*:

$$\alpha^{(2)}(B_1 \times B_2) = \int_{B_1}\int_{B_2} \varrho^{(2)}(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2\,. \tag{8}$$

Moreover, for any non-negative bounded measurable function f

$$\mathbf{E}\left(\sum_{x_1,x_2\in N}^{\neq} f(x_1, x_2)\right) = \int\int f(x_1, x_2)\varrho^{(2)}(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2\,.$$

The product density has an intuitive interpretation, which probably accounts for its historical precedence over the product measure and the K-function introduced below. (Note that there are also n^{th} order product densities and moment measures.) Suppose that C_1 and C_2 are disjoint spheres with centres x_1 and x_2 and infinitesimal volumes $\mathrm{d}V_1$ and $\mathrm{d}V_2$. Then $\varrho^{(2)}(x_1, x_2)\mathrm{d}V_1\mathrm{d}V_2$ is the probability that there is each a point of N in C_1 and C_2. If N is stationary then $\varrho^{(2)}$ depends only on the difference of its arguments and if furthermore N is motion-invariant then it depends only on the distance r between x_1 and x_2 and it is simply written as $\varrho^{(2)}(r)$. The *pair correlation function* $g(r)$ results by normalization:

$$g(r) = \varrho^{(2)}(r)/\lambda^2. \tag{9}$$

Without using the product density $\varrho^{(2)}$, the second factorial moment measure can be expressed by the *second reduced moment measure* $\mathcal{K}$ as

$$\begin{aligned}\alpha^{(2)}(B_1 \times B_2) &= \lambda^2 \int_{B_1} \mathcal{K}(B_2 - x)\mathrm{d}x \\ &= \lambda^2 \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbf{1}_{B_1}(x)\mathbf{1}_{B_2}(x+h)\mathcal{K}(\mathrm{d}h)\mathrm{d}x\,. \end{aligned}\tag{10}$$