This is to certify that the

dissertation entitled

Bayesian Modeling On
Inhomogeneous Point Patterns Via
Independent Increment Random Measures

presented by

Fanzhi Kong

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics

Major professor
Dennis Gilliland

Date August 5, 2002

# BAYESIAN MODELING ON INHOMOGENEOUS POINT PATTERNS VIA INDEPENDENT INCREMENT RANDOM MEASURES

By

Fanzhi Kong

A DISSERTATION

Submitted to

MICHIGAN STATE UNIVERSITY

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability

2002

# Abstract

BAYESIAN MODELING ON
INHOMOGENEOUS POINT PATTERNS VIA INDEPENDENT INCREMENT
RANDOM MEASURES

By

Fanzhi Kong

The main objective in this thesis is to investigate the inference and prediction of the intensity functions governing inhomogeneous point patterns. The intensity functions are modeled as random fields directed by independent increment random measures in a reference state space. Random fields are the foundation of spatial data analyses. One constructed model on point patterns is based on the Poisson generalized Gamma random fields. The generalized Gamma measures can be simulated by the Inverse Lévy Measure algorithm. For highly clustered point patterns, it is more appropriate to apply the geostatistical modeling approaches after the log transformation of count data. The related models are based on Wiener measure subordinated random fields, which are closely related to the ordinary random fields specified directly by mean and covariance functions.

Another concern is to make full use of collateral information obtained along with point patterns. The collateral data, treated as fixed covariates and covariate stochastic processes, are considered in the setting of joint modeling on the intensities. The involved modeling approaches follow both theoretical and simulation methods. The corresponding MCMC schemes are proposed for each data model. The actual implementation is demonstrated and performed for the joint model on counts and binary clipped random fields.

# Acknowledgements

My deepest regards and sincerest thanks go to my thesis advisors, Professors Dennis Gilliland and Oliver Schabenberger, for their many instructions and great enthusiasm support during this research and for their guidance through the early years of chaos and confusion. I sincerely appreciate Professor Gilliland's inspiration and extreme patience in the final stage of preparing this dissertation. My committee members Professors V. S. Mandrekar, Habib Salehi and Yimin Xiao provided the technical assistance that improved some results and eliminated errors; their help was invaluable.

Fuxia Cheng and Yichuan Xia shared with me their knowledge of second order analysis and provided many useful references and friendly encouragement. I am grateful to the other members of the Department of Statistics and Probability at Michigan State University who make the department such an exciting, friendly and inspiring place to study and work, in particular to Professor James Stapleton.

For financial support I thank Professors Oliver Schabenberger, Patrick Hart and Shengyang He, who mainly provided the research assistantship support for this project during my graduate studies. I should also mention that my graduate studies in the United States were supported in part by the Department of Crop and Soil Sciences and the DOE Plant Research Laboratory at Michigan State University.

East Lansing, Michigan, USA Fanzhi Kong

July 31, 2002

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1. Spatial Poisson-Generalized Gamma Random Fields

The study of inhomogeneous point patterns often involves the treatment of additional spatial data types. The joint analysis of point patterns and collateral information may improve the inference and prediction efficiency. The consideration is demanded because point patterns are practically most expensive and collateral data are often available along with point patterns. Thus, the methods of joint analysis of point patterns and their collateral data are highly recognized recently. The following section gives a short introduction to point patterns and other spatial data types considered in this thesis.

## 1.1 Point Patterns and Collateral Information

Consider a bounded subset $\mathbb{S}$ of the $d$-dimensional Euclidean space with $d \geq 1$. A ***point pattern***, $\{S_j; j = 1, \cdots, J\}$, is the set of locations of observed events from all sampling windows $\{A_i\}$ in $\mathbb{S}$.

Other spatial measurements $\{Y_i; i = 1, \cdots, n\}$ are taken from the corresponding locations $\{s_i\} \subset \mathbb{S}$. The locations $\{s_i\}$ are explicitly/implicitly associated with sampling windows $\{A_i\}$ for obtaining feasible measurements. Practical applications often involve those random variates $Y_i$ distributed on a regular lattice. Examples include image analysis, where locations represent pixels, and crop experiments, where they equate with plots in the field. Another example is a two-way table with responses that are the aggregate of independent row effects, column effects and higher order effects. In this context, the locations $\{s_i\}$ are called the ***positive integral coordinates*** or

the ***generic locations***. One may distinguish point patterns from other types of measurements by associating $Y_i$ with the sampling window $A_i$ with $s_i \in A_i \subset \mathbb{S}$. The sampling windows $\{A_i\}$ may fall far away from each other or may make contact with their neighbors. In the former case, $\{s_i\}$ can be assumed to have no distinguishable points in the space concerning the sizes of sampling windows, and in the latter case, $\{s_i\}$ together with $\{A_i\}$ form a partition of the study region. When the relative size of sampling windows can not be neglected, and sampling windows are disconnected, $\{s_i\}$ forms a ***network*** with features of the sampling windows as node attributes. The corresponding data $\{Y_i\}$ may be called ***geostatistical data***, ***lattice data*** and ***network data***, respectively.

The spatial data can also be categorized by the characteristics of observables and may be grouped into three categories: ***geostatistical data***, ***point pattern data*** and ***categorical data*** (or random sets[1]). Cressie (1993) classifies the general spatial data into three types, i.e., geostatistical data, lattice data, and point pattern data.

## 1.2 Thesis Synopsis

The choice of spatial models is dictated by the particular form of the spatial data. It is also influenced by the modeling objectives, which may be to summarize data, to predict dynamics, or to model mechanistically in order to advance understanding. In this thesis, the data representation and interpretation for different types of spatial data will be based on the Voronoi tessellations. The corresponding models will be constructed with the aid of their directed acyclic graphic representations. The direct approaches for inference and prediction in inhomogeneous models are generally not feasible. We will follow Monte Carlo and Markov Chain methods, and all the required

[1]The ***random set*** is the main subject of the spatial geometry.

full conditional distributions are derived or approximated for the proposed models.

The thesis develops several models on inhomogeneous point patterns in the presence of covariates and covariate stochastic processes. A chapter based synopsis of the involved models and their analyses follow.

The rest of this chapter concerns the Poisson generalized Gamma random fields. We establish properties of the generalized Gamma measures that are useful for modeling and inference for point patterns. We also discuss the incorporation of a covariate into the analysis.

Chapter 2 performs the simulation study for the joint model of count data and censored data. In Section 2.1, we continue the work on the two-stage model with a particular generalized Gamma measure and with a particular data structure (count and indicator variables). The data structure is more fully described in Section 2.2; it consists of count data on a subset of the cells of a regular lattice and indicator variables for counts in excess of a cutoff point on the other cells. In Section 2.3, we develop the full conditional distributions needed for an MCMC implementation. Section 2.4 gives the results of the MCMC implementation based on specific data values.

Chapter 3 studies modeling on point patterns with the intensity function as the transformed kernel mixture of Wiener measures, i.e., log Gaussian random fields. In Section 3.1, we introduce a model for log-count data that includes covariates and an additive error. In Section 3.2, we motivate a model for binary data derived from a clipped intensity function. The second order structure is discussed. In Section 3.3, we discuss inference for point patterns that associate with a covariate process.

Chapter 4 summarizes the main results in the thesis and discusses some issues

for further investigation.

### 1.2.1 Terminology and Notation

Some terminology and notation described here will be used through the thesis. The Bayesian approach is natural for the type of statistical problems treated in the thesis. The Bayesian approach to statistics uses probability measures, usually defined in terms of density functions, to quantify and assess unknown parameters and other unknown quantities, so it involves a direct application of probability theory. Densities and probability functions of random vectors $X$ and $Y$ will be denoted generically by square brackets in the context, so that the joint, conditional and marginal forms appear, respectively, as $[X, Y]$, $[X|Y]$, and $[Y]$. This notation is used by Gelfand et al. (1990a). The usual marginalization by integration or summation procedure will be denoted by forms such as $[X] = [X|Y][Y]$.

When applying the Monte Carlo Markov Chain methods, we will use the term 'MCMC' for simplicity. In deriving the conditional densities, the phrase "is proportional to" is denoted with the symbol '$\propto$'. For the log conditional densities, the involved additive constant will be denoted by a generic constant '$c_g$'. This convention is convenient because it reduces the amount of notation, and since these densities are seldom evaluated, it produces no confusion. The density or distribution of a random vector $X$ will denote by '$\sim$', interpreted as "$X$ is distributed as". The transpose of a vector or a matrix $X$ will be denoted as $X^T$.

In defining random measures the underlying probability space will be denoted by $(\Omega, \mathcal{F}, \mathbb{P})$. We will denote the Lebesgue measure on $d$-dimensional Euclidean space $\mathbb{R}^d$ by '$|\cdot|$'. Among others, '$GG$' is the shorthand for the generalized Gamma measures or distributions with arguments supplied in the context. Finally, the geostatistical

random fields, models, and data will be used to emphasize the continuous nature over a subset of Euclidean space $\mathbb{R}^d$.

## 1.3 Poisson-Generalized Gamma Random Field Models

In analyzing point patterns the unifying theme is to model its spatial uncertainty through its (first-order) intensity function. Point patterns are commonly modeled as a Poisson process $N(ds)$ over the state space $\mathbb{S}$ (and its Borel $\sigma$-algebra $\mathcal{S}$), which provides the independency structure for mathematical manipulation. The $N(ds)$ is an inhomogeneous ***Poisson process*** with mean measure $\Lambda$ if (i) for any $A \in \mathcal{S}$, $P(N(A) \in \{0, 1, \cdots\}) = 1$, and for any collection of disjoint sets $A_1, \cdots, A_n \in \mathcal{S}$, the random variable $N(A_1), \cdots, N(A_n)$ are independent; (ii) for all $s \in \mathbb{S}$,

$$\begin{aligned} P(N(ds) = 0) &= 1 - \Lambda(ds) + o(\Lambda(ds)) \\ P(N(ds) = 1) &= \Lambda(ds) + o(\Lambda(ds)) \\ P(N(ds) > 1) &= o(\Lambda(ds)) \end{aligned} \tag{1.1}$$

where $ds$ is an infinitesimal region located at $s \in \mathbb{S}$. From these two postulates, it has been shown that $N(A)$ has a Poisson distribution with mean measure $\Lambda(A)$, for all $A \in \mathcal{S}$, (e.g., pp13-14, Rogers, 1974),

$$P(N(A) = j) = \frac{\Lambda(A)^j e^{-\Lambda(A)}}{j\,!}, \quad j = 0, 1, 2, \cdots \tag{1.2}$$

Point patterns may present spatial correlation. The double stochastic process about to be described is a generic structure for creating spatial correlation. In this study, the intensity function $\Lambda(s)$ of a point pattern is modeled as a random field,

given by the stochastic integral,

$$\Lambda(s) := \int_{\mathbb{U}} k(s,u)M(du) \tag{1.3}$$

where $M(du)$ is an independent increment random measure on a reference state space $\mathbb{U}$ (also a bounded subset of the $d$-dimensional Euclidean space) with the control measure $\lambda(du)$ and the nonnegative deterministic kernel $k$ is such that

$$\sup_{s\in\mathbb{S}} k(s,u) \in \mathcal{L}_2(\mathbb{U},\mathcal{U},\lambda) \tag{1.4}$$

where $\mathcal{U}$ is the Borel $\sigma$-algebra of $\mathbb{U}$. The kernel mixture type of independent increment random measures provides a flexible family of models for inhomogeneous point patterns with a simple and interpretable structure. The specification of kernels often relies on the preliminary analysis of semi-variogram structures. The resulting models for point patterns are called ***Cox*** $M$ models.

The Cox $M$ models have been used to explain the interaction among the point events at small scales. Heisel et al. (1996) use one to quantify the small range covariance structure in a weeds application. Wolpert and Ickstadt (1998a, 1998b) have developed the theory of the Poisson Gamma random fields, that is, when $M(du)$ is a Gamma measure. Brix (1999) discusses the properties and the simulation of the generalized Gamma measures with constant parameter functions.

In this study, we consider the generalized Gamma measures (*GG-**measures*** for short), with possibly nonconstant parameter functions over $\mathbb{U}$. The family of *GG*-measures includes Gamma random measures, positive stable random measures and inverse Gaussian measures. The family of Cox *GG* models is a Poisson cluster process type and includes Poisson processes and ordinary Neyman-Scott processes. As such the family provides models which can be interpreted both as germ-grain models (Stoyan et al., 1995) and as overdispersion models for point patterns.

### 1.3.1 Generalized Gamma Random Measures

The Laplace functional for random measures plays a role similar to the Laplace transform for random variables. The distribution of a random measure $M(du)$ on a measurable space $(\mathbb{M}, \mathcal{M})$ is uniquely determined by its Laplace functional (Stoyan et al., 1995). The Lévy-Khinchin representation (Jacod and Shiryaev, 1987 and Wolpert and Ickstadt, 1998a) for a nonnegative independent increment random measure $M$ has a Laplace functional

$$Ee^{-M[f]} = \exp\left(-\phi[f] + \iint_{\mathbb{U}^*} (e^{-hf(u)} - 1)\ell(dudh)\right), \tag{1.5}$$

where $f(u)$ is any compactly-supported bounded measurable function, and the drift $\phi[f] = \int_{\mathbb{U}} f(u)\phi(du)$ with $\phi(du)$ a $\sigma$-finite positive measure on $\mathcal{U}$. The inhomogeneous ***Lévy measure*** $\ell(dudh)$ on $\mathbb{U}^* := \mathbb{U} \times \mathbb{R}^+$ satisfies the integrability condition

$$\iint_{B\times\mathbb{R}^+} (1 \wedge h)\ell(dudh) < \infty \tag{1.6}$$

for each compact $B \in \mathcal{U}$. Hence, the random measure $M$ is determined by the pair of measures $(\phi, \ell)$.

In the rest of this chapter and Chapter 2 we take $M(du)$ to be the generalized Gamma measure. It is the $GG(a(u), \lambda(du), b(u))$-measure specified by the Lévy measure,

$$\ell(dudh) = \frac{1}{\Gamma(1 - a(u))} h^{-a(u)-1} e^{-b(u)h} \lambda(du)dh, \tag{1.7}$$

where $a(u) < 1$ is the ***index*** parameter function, $b(u) > 0$ is the ***inverse scale*** parameter function, and the boundedly finite measure $\lambda(du)$ is the ***control measure*** on $(\mathbb{U}, \mathcal{U})$. By a boundedly finite measure, we mean it is finite on every bounded Borel set. A sufficient condition for the Lévy measure $\ell$ in (1.7) to satisfy the integrability condition in (1.6) is given by: for every compact $B \in \mathcal{U}$,

$$\int_B b(u)^{a(u)-1}\lambda(du) < \infty. \tag{1.8}$$

We take $a(u)$ and $b(u)$ to be continuous on $\mathbb{U}$ unless specified otherwise.

### The Generalized Gamma Distributions

The $GG$-measures are derived from the family of generalized gamma distributions, $GG(a, \lambda, b)$. This family is suggested by Tweedie (1984) and Hougaard (1986), and considered in various directions by Bar-Lev and Enis (1986), Jørgensen (1987), Aalen (1992), Hougaard et al. (1997), and Brix (1998). A $GG(a, \lambda, b)$-distribution is concentrated on $(0, \infty)$ and infinitely divisible in $\lambda$, which make it a natural basis for defining independent increment random measures. This also has an important implication in assigning independent random variables to disjoint subsets in the reference state space in modeling spatial data.

The family of $GG(a, \lambda, b)$-distributions is characterized by its Laplace transform,

$$L(t|a, \lambda, b) = \exp\left(-\frac{\lambda}{a}((b+t)^a - b^a)\right), \quad t \geq 0,$$

where $a \leq 1$, $\lambda > 0$ and $b \geq 0$. The family includes three subfamilies, separated by the index parameter $a$.

For $a = 0$, the family is defined only for $b > 0$ through the limit of its Laplace transform, $\lim_{a \to 0} L(t|a, \lambda, b)$. It consists of the Gamma distributions $GG(0, \lambda, b)$ with the shape parameter $\lambda$ and the inverse scale $b$.

For $a < 0$ and $b > 0$ the family is a convolution of a Poisson distributed number of Gamma random variables with shape $-a$ and inverse scale $b$ (Aalen, 1992), and its absolutely continuous component has density

$$q(h|a, \lambda, b) = \exp\left(bh + \frac{\lambda}{a}b^a\right)\frac{1}{h}\sum_{k=1}^{\infty}\frac{1}{k!\Gamma(-ka)}\left(-\frac{\lambda}{ah^a}\right)^k \tag{1.9}$$

on $(0, \infty)$; the point mass at zero is given by

$$q(0|a, \lambda, b) = \exp\left(\frac{\lambda b^a}{a}\right), \quad \lambda > 0.$$

For $0 < a < 1$, the family is the natural exponential family generated by the positive stable distributions with the stable index $a$; in particular, for $\lambda > 0$ and $b \geq 0$, the density is given by

$$q(h|a, \lambda, b) = -\exp\left(bh + \frac{\lambda}{a} b^a\right) \frac{1}{\pi h} \sum_{k=1}^{\infty} \frac{\Gamma(ka+1)}{k!} \left(-\frac{\lambda}{a h^a}\right)^k \sin(ak\pi). \qquad (1.10)$$

Its saddle point approximation is discussed by Hougaard (1986).

For a $GG(a, \lambda, b)$ distribution, all the moments exist for $b > 0$, and the $r$th cumulant, $r > 1$, is given by

$$\chi_r = \lambda(1-a)(2-a)\cdots(r-1-a)b^{a-r}.$$

The squared coefficient of variation (i.e., the ratio of the variance and the square of the mean) is $(1-a)/(\lambda b^a)$. The mean of the distribution is $\lambda b^{a-1}$ and the variance is $(1-a)\lambda b^{a-2}$. The conditional moments for the continuous part of the distribution in the case $a < 0$ was considered by Aalen (p955, 1992).

Aalen (1992) uses the powerful tool `Mathematica` to perform numerical studies of these densities and demonstrates the irregularly curved density surfaces for various ranges of parameters. The densities of $GG$-distributions are generally difficult to compute due to multi-modality except for special cases, such as the Gamma distribution $GG(0, \lambda, b)$ and the inverse Gaussian distribution $GG(1/2, \lambda, b)$. In statistical problems involving $GG$-distributions, it may be difficult to apply density computation oriented methods such as the maximum likelihood inference. In analyzing inhomogeneous point patterns with the Cox $M$ models, the MCMC method is the main tool for numerical implementation. This will be used Chapter 2.

**Poisson Processes Representations**

Any independent increment random measure $M(du)$ on $\mathbb{U}$ can be uniquely decomposed into the following form, (see Theorem 6.3VIII, Daley and Vere-Jones, 1988)

$$M(du) = \phi(du) + \int_0^\infty hN^*(dudh) + \sum_{j=1}^{\infty} W_j \delta_{u_j}(du). \tag{1.11}$$

Here the sequence $\{u_j\}$ enumerates the countable set of fixed atoms of $M$, $\delta_u$ is the Dirac measure in $u$, and $\{W_j\}$ is a sequence of mutually independent nonnegative random variables determining the masses at these atoms. $\phi(du)$ is a fixed diffuse boundedly finite measure on $\mathbb{U}$ and $N^*(dudh)$ is a Poisson process on $\mathbb{U}^*$, independent of $\{W_j\}$. The intensity measure $\ell$ of the process may be unbounded on sets of the form $B \times (0, \epsilon)$, but satisfies (1.6) for every bounded Borel set $B$, in which $\ell(\{u\} \times (0, \infty)) = 0$ for all $u \in B$.

The Poisson process $N^*$ in (1.11) can be viewed as the generalized marked Poisson process on $\mathbb{U}$ with the marking space $\mathbb{R}^+$. In particular, the $GG$-measures can be represented as or approximated by the marked Poisson point processes.

For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a random measure $M$ is said to have a ***fixed atom*** $u$ if $\mathbb{P}(M(\{u\}) > 0) > 0$ and is ***diffuse*** if $M(\{u\}) = 0$ for every $u \in \mathbb{U}$ (p4, Karr, 1991). The following proposition shows that a $GG$-measure is almost surely purely atomic.

**Proposition 1.** *For a $GG$-measure $M$ with its Lévy measure $\ell$ given by* (1.7), *$M$ is almost surely purely atomic and has no fixed atoms if and only if $\lambda(du)$ is diffuse.*

*Proof.* The following proof can be extended to any random measure $M$ having the Laplace functional (1.5) without the drift $\phi$ and with its Lévy measure $\ell$ satisfying (1.6).

Let $N'$ be the Poisson process on $\mathbb{U}^*$ with the intensity measure $\ell(dudh)$ satisfying (1.6). Consider the random measure $M'$ on $\mathbb{U}$ given by

$$M'(du) = \int_0^\infty hN'(dudh)$$

which is almost surely purely atomic. For a bounded set $B \in \mathcal{U}$, the Laplace transform of $M'(B)$ is then, for $t \geq 0$,

$$\begin{aligned} E(e^{-tM'[I_B]}) &= E\exp\left(-\int_{\mathbb{U}}\int_0^\infty thI_B(u)N'(dudh)\right) \\ &= E\exp\left(-\int_0^\infty thN'(B\times dh)\right) \\ &= \exp\left(\int_0^\infty (e^{-th}-1)\ell(B\times dh)\right) \\ &= \exp\left(\int_{\mathbb{U}}\int_0^\infty (e^{-thI_B(u)}-1)\ell(dudh)\right) \end{aligned}$$

where the third equation follows from Campbell's theorem; see Kingman (p82, 1993) for a constructive proof. By taking limiting simple function approximations for all compactly-supported measurable functions $f$ on $\mathbb{U}$, it follows from the dominated convergence theorem and the integrability condition (1.6) that the Laplace functional of $M'$ is as follows:

$$\begin{aligned} E(e^{-M'[f]}) &= E\exp\left(-\int_{\mathbb{U}}\int_0^\infty f(u)hN'(dudh)\right) \\ &= \exp\left(\int_{\mathbb{U}}\int_0^\infty (e^{-hf(u)}-1)\ell(dudh)\right). \end{aligned}$$

Thus, $M$ and $M'$ have the same Laplace functional and consequently follow the same distribution. The fact that a $GG$-measure $M$ is almost surely purely atomic, and has no fixed atoms if and only if $\lambda(du)$ is diffuse follows from the result (1.11). □

Later we use the so-called ***Inverse Lévy Measure*** (ILM) algorithm[2] to construct a generalized marked Poisson process representation of $GG$-measures. Let

[2]Wolpert and Ickstadt (1998a) were the first to propose this algorithm.

$\Pi(du)$ be a Borel probability measure on $\mathcal{U}$ and assume that $\lambda(du)$ have a density $\lambda(u)$ with respect to the measure $\Pi(du)$. Let $\ell(u,h)$ denote the density of the measure $\ell(dudh)$ in (1.7) with respect to the measure $\Pi(du)dh$. Let $\{U_j\}$ be independent identically distributed from the probability distribution $\Pi(du)$, and $t_j \geq 0$ be the successive jump times of a unit rate Poisson process. Set

$$\tau_u(h) := \int_h^\infty \ell(u,x)dx \tag{1.12}$$

and, for each $j = 1, 2, \cdots$, define

$$H_j := \inf\{h \geq 0 : \tau_{U_j}(h) \leq t_j\}. \tag{1.13}$$

Then, for a compactly-supported bounded measurable function $f(u)$, the random measure given by

$$M[f] := \sum_{j<\infty} H_j f(U_j)$$

has the Laplace functional of the form (1.5) with its Lévy measure given by (1.7). In particular, for a bounded Borel $B$ in $\mathcal{U}$, we have

$$M(B) = \sum_{j<\infty} H_j \delta_{U_j}(B). \tag{1.14}$$

The number of terms in (1.14) depends on the index parameter function $a(u)$. The following proposition shows that the sum (1.14) consists of infinitely many terms when $0 \leq a(u) < 1$ on $\mathbb{U}$. For a random measure $M$, a Borel $B \in \mathcal{U}$ and $\epsilon > 0$ define $N_\epsilon(B)$ to be cardinality of the set $\{u \in B; M(\{u\}) \geq \epsilon\}$, i.e., the number of atoms for $M$ in $B$ with mass at least $\epsilon$ with $\epsilon > 0$. Denote by $S_M(B)$ the support of $M$ within $B$.

**Proposition 2.** *For any random measure $M(du)$ with its Lévy measure $\ell$, $N_\epsilon$ is a Poisson process on $(\mathbb{U},\mathcal{U})$ with finite intensity measure $\ell_\epsilon(\cdot) := \ell(\cdot \times [\epsilon,\infty))$. In particular, if $M(du)$ is a GG-measure with $a(u) < 0$ for all $u$ in $\mathbb{U}$, then, for any*

*compact $B$ in $\mathcal{U}$, the number of jumps in $S_M(B)$ is finite; for $0 \leq a(u) < 1$ for all $u$ in $\mathbb{U}$ and any Borel $B$ in $\mathcal{U}$, $S_M(B)$ is dense in the support of $\lambda$.*

*Proof.* For each $\epsilon > 0$ and with the notation in (1.11), $N_\epsilon$ can be written as $N_\epsilon(B) = N^*(B \times [\epsilon, \infty))$, which is a Poisson process with intensity $\ell_\epsilon$. Define $N_0$ as the almost sure limit of the process $N_\epsilon$ as $\epsilon \to 0$. Recall that $a(u)$ and $b(u)$ are continuous, they are bounded on a compact set $B$. If $a(u) < 0$ on $\mathbb{U}$ and $b(u) > 0$, then $a(u) \leq -\delta$ and $b(u) \geq \delta$ for some $\delta > 0$. Thus

$$\begin{aligned} \lim_{\epsilon \to 0} \ell_\epsilon(B) &= \lim_{\epsilon \to 0} \int_B \int_\epsilon^\infty \ell(dudh) \\ &= \int_B \frac{b(u)^{a(u)} \lambda(du)}{-a(u)} < \infty \end{aligned} \tag{1.15}$$

which implies that $N_0$ is a Poisson point process with finite intensity $\ell_0$. In general, $\ell_0$ in (1.15) is finite if $b(u)$ is bounded from above and $a(u)$ is bounded from zero. On the other hand, for any Borel $B$ in $\mathcal{U}$ with $0 \leq a(u) < 1$, $\ell_\epsilon(B) \to \infty$ as $\epsilon \to 0$, so that $N_0(S_M(B))$ is almost surely unbounded. □

The above proposition together with the marked Poisson process representation of *GG*-measures imply that the number of terms in (1.14) is finite on a compact $B$, on which $a(u) < 0$, with a Poisson intensity measure $\lambda(du)b(u)^{a(u)}/(-a(u))$ on $\mathbb{U}$ and a Gamma marking distribution of shape $-a(u)$ and inverse scale $b(u)$. When $0 \leq a(u) < 1$ the discrete measure $M$ has dense support on the support of $\lambda$ for any Borel set $B$. An approximate Poisson process representation can be simulated by the ILM algorithm. The choice of a finite number $J$ of atoms for the algorithm is discussed by Wolpert and Ickstadt (1998a) and by Brix (1999). The truncation bound given by

Brix (1999) could be applied in most applications. The ILM algorithm implements the sampling of atoms on the partially ordered sampling space, i.e., $\bigotimes_{j=1}^{J}(u_j, h_j)$ such that $\tau_{u_B}(h_j) \leq \tau_{u_{B+1}}(h_{j+1})$ for every $j < J$. When simulating atoms of the *GG*-measure with index $0 \leq a(u) < 1$, $J$ is used in tabulating $\tau^{-1}$, which has no closed form.

With the ILM algorithm, the joint distribution of the locations and marks of atoms is provided by the following theorem. Define

$$E_{a+1}(h) = \int_h^\infty x^{-a-1}e^{-x}dx \tag{1.16}$$

from which $E_1(h)$ is the exponential integral function (p220, Abramowitz and Stegun, 1964).

**Theorem 1.3.1.** *The joint distribution of locations and jump sizes* $\{(u_j, h_j)\}_{j \leq J} \subset \mathbb{U}^*$ *in the ILM algorithm has a density with respect to the product measure* $\prod_{j \leq J} \Pi(du_j)dh_j$ *proportional to*

$$\exp\left(-\frac{\lambda(u_J)b(u_J)^{a(u_A)}E_{a(u_A)+1}(b(u_J)h_J)}{\Gamma(1-a(u_J))}\right)$$
$$\times \prod_{j \leq J} \frac{\lambda(u_j)b(u_j)^{a(u_B)}}{\Gamma(1-a(u_j))} h_j^{-a(u_B)-1}e^{-b(u_B)h_B}$$

*on the set of* $\{(u_j, h_j)\}_{j \leq J}$ *for which the measure*

$$\tau_u(h) := \frac{\lambda(u)b(u)^{a(u)}}{\Gamma(1-a(u))}E_{a(u)+1}(b(u)h)$$

*satisfies* $\tau_{u_B}(h_j) \leq \tau_{u_{B+1}}(h_{j+1})$ *for every* $j < J$.

*Proof.* From the definition (1.16), for $b > 0$, by a change of variable,

$$E_{a+1}(bh) = b^{-a}\int_h^\infty x^{-a-1}e^{-bx}dx$$

It follows from the definition in (1.12),

$$\tau_u(h) = \frac{\lambda(u)}{\Gamma(1-a(u))} \int_h^\infty x^{-a(u)-1} e^{-b(u)x} dx$$
$$= \frac{\lambda(u) b(u)^{a(u)}}{\Gamma(1-a(u))} E_{a(u)+1}(b(u)h)$$

which is decreasing in $h$ and its inverse function $\tau_u^{-1}(t)$ exists for each fixed $u$. Thus, the greatest lower bound for $\inf_h \tau_u(h) = t$ is attained, and we can define

$$h := \tau_u^{-1}(t) = \frac{1}{b(u)} E_{a(u)+1}^{-1}\left(\frac{\Gamma(1-a(u))t}{\lambda(u)b(u)^{a(u)}}\right)$$

Consider a sequence $T_1, T_2, \cdots$ of independent exponential random variables, i.e., the waiting times for the successive jumps in the standard Poisson process. Then $H_j = \tau_{U_B}^{-1}(T_1 + \cdots + T_j)$ for each $j \leq J$. Thus the theorem follows by a change of variable. □

For a special case of the above theorem, where the index function of the *GG*-measure constant zero, see Corollary 2 to Theorem 2 in Wolpert and Ickstadt (1998a).

### 1.3.2 Intensity Models and Modeling Objectives

Generally, the intensity functions in (1.3) used for modeling point patterns are complex mechanisms with unknown parameter functions. The kernel and parameter functions are parameterized with problem specific unknown quantities $\theta$. We suppress the display of $\theta$ in $k^\theta(s,u)$, $a^\theta(u)$, $\lambda^\theta(du)$ and $b^\theta(u)$ for the rest of this chapter.

#### Covariates and Intensity Models

For some applications, covariates $x(s)$ are available to be included in the analysis, and the intensity model in (1.3) may be treated as a "baseline" intensity $\Lambda_0(s)$ of the

point patterns. The covariate factor $\beta(s)$ can be modulated by the Poisson intensity locally in locations, given by $\beta(s) = x^T(s)\boldsymbol{\beta}$ with the covariate coefficients $\boldsymbol{\beta}$. We adopt the multiplicative incorporation of collateral information into the analysis of intensities for a different class of spatial Poisson models. In this case, the intensity function $\Lambda(s)$ for point patterns may be factored into two factors, i.e.,

$$\Lambda(s) = \exp(\beta(s))\Lambda_0(s). \tag{1.17}$$

The multiplicative model with the exponential link has the merits of simplicity and positivity insured when further parameterizing the model (Diggle, 1990; Ickstadt and Wolpert, 1996). Often in applications, the covariates $x(s)$ are determined only in lattice cells of the state space $\mathbb{S}$, and one takes $x(s) = x(s_i)$ for all $s$ in the lattice cell with center $s_i$. In this case, on the cell with center $s_i$, $x(s) = x_i^T\boldsymbol{\beta}$. For the intensity model (1.17), the covariate factor $\beta(s)$ can be incorporated into the kernel $k(s,u) = \exp(\beta(s))k_0(s,u)$ with $k_0$ as the baseline kernel corresponding to the baseline intensity $\Lambda_0(s)$.

The mean intensity surface in (1.3) with a $GG$-measure $M(du)$ is given by

$$E\Lambda(s) = \int_{\mathbb{U}} k(s,u)b(u)^{a(u)-1}\lambda(du)$$

and the independence property of $GG$-measures leads to the covariance function given by

$$\gamma_\Lambda(s,s') = \int_{\mathbb{U}} k(s,u)k(s',u)(1-a(u))b(u)^{a(u)-2}\lambda(du) - E\Lambda(s)E\Lambda(s'). \tag{1.18}$$

This can also be derived by the Campbell theorem, see Brix (1999).

### Inference on Intensity Models

With the observed point patterns $\{S_n\}$, there is a general interest in the spatial correlation, which is specified through a parameterized kernel, say $k(s,u)$. Another

aspect of spatial dependence is the locality, which is explained by the distribution of a *GG*-measure $M_\theta(du)$ and covariate coefficients $\beta$. The inference on the intensity model is carried out through estimation of the involved parameter, estimation of the intensity at the observed locations, and prediction of the intensity at unobserved locations. The stochastic view to the inference problem is to treat the point patterns $\{S_n\}$ as part of one realization of the spatial Poisson process $N(ds)$ with the given intensity measure $\Lambda(ds)$, while the unobserved part (of the same realization) is predicted using the point patterns data, location-dependent covariates and correlation structure specified by the model. This has proved to be very useful as a surrogate for the analyst's lack of knowledge about the complex system generating $\{S_n\}$.

When modeling point patterns, uncertainty analysis is often of main interest. One concern is actually the spatial homogeneity, which involves invariant properties such as location-transformation, rotation, scaling, and location-dependent covariates. Another concern is spatial trends and spatial patterns, such as clustering and aggregation. The third concern lies with the spatial correlation. Traditionally, stationary random fields together with other special conditions are assumed; inference in spatial statistics is based on a combination of *ad hoc* nonparametric techniques, such as distance-based methods and second-order methods, kernel smoothing (Silverman, 1986), maximum-likelihood parameter estimates and associated approximations and simulation-based estimation and testing. With constant index and scale parameters, a *GG*-measure is stationary if and only if its control measure $\lambda(du)$ is proportional to the Lebesgue measure on $(\mathbb{U}, \mathcal{U})$; see Brix (1999).

The inference on inhomogeneous Cox *GG* models is difficult in part due to its numerical implementation. The emergence of MCMC algorithms (Gelfand and Smith, 1990a; Tierney, 1994; Gilks, Richardon and Spiegelhalter, 1996) has made it practicable, and increasingly common, to apply Bayesian statistical methods to spatial

analyses, where the full conditional distributions required for the Gibbs sampling approach are available. The posterior samples are obtained from the stabilized Markov chain in a MCMC scheme. One realization obtained by this simulated chain can be regarded as a dependent, approximate sample from the posterior, and various posterior inferences can be drawn from empirical data analysis of this sample. The estimates of parameters and prediction of $\Lambda(s)$ are obtained by the posterior sample means. The visual appreciation can be obtained from the posterior empirical distributions (histograms), which will be discussed in Chapter 2.

The implementation of MCMC schemes sometimes turns out to be difficult because of the difficulty in sampling from full conditional distributions. An important MCMC technique is data augmentation (Tanner and Wong, 1987). When the direct sampling from the conditional distributions for certain parameters is difficult, the data augmentation may be needed and will be demonstrated later in context. Another useful technique is the Metropolis-Hastings (MH) algorithm (see Smith and Roberts, 1993), which makes it possible to sample from complicated full conditional distributions. The MH scheme is not only straightforward to implement, but is also a general purpose method. This makes it suitable for handling any form of conditional distribution under general assumptions.

# Chapter 2. Joint Inference on Counts and Binary Random Fields

In this chapter, a simulation study is performed for the joint model of count data and binary clipping data. In Section 2.1, we continue the work on the two-stage Cox $M$ model with a Gamma measure (Wolpert et al., 1998a) and with a particular data structure (count and indicator variables). The data structure is more fully described in Section 2.2; it consists of count data on a subset of the cells of a regular lattice and indicator variables for counts in excess of a cutoff point on the other cells. In Section 2.3, we develop the full conditional distributions needed for an MCMC implementation. Section 2.4 gives the results of the MCMC implementation based on specific data values.

## 2.1 Model Specification and Representation

Consider a Poisson process $N(ds)$ on $\mathbb{S} \subset \mathbb{R}^2$ with the intensity defined as in (1.3), in which $M(du)$ is a $GG(0, \lambda(du), b(u))$ measure with parameter functions given by

$$\lambda(du) = \lambda\, du, \quad b(u) = \exp(b^T u/\theta^2)$$

where $\theta > 0$ and $b \in \mathbb{R}^2$. Take the kernel to be

$$k^\theta(s,u) = \frac{1}{\pi\theta^2} \exp\left(-\frac{1}{\theta^2}||s-u||^2\right), \tag{2.1}$$

which is a two-dimensional stationary Gaussian kernel, normalized by the condition, $\int_{\mathbb{R}^2} k = 1$, supporting a range of possible spatial correlations. With the above choices

the mean intensity function is given by

$$\begin{aligned} E\Lambda(s) &= \int_{\mathbb{U}} \frac{1}{\pi\theta^2} \exp\left(-\frac{1}{\theta^2}||s-u||^2\right) \frac{1}{b(u)} \lambda(du) \\ &= \int_{\mathbb{U}} \frac{1}{\pi\theta^2} \exp\left(-\frac{1}{\theta^2}\left(||s+\frac{b}{2}-u||^2\right)\right) du \\ &\quad \times \lambda \exp\left(\frac{1}{\theta^2}\left(b^T s + \frac{||b||^2}{4}\right)\right) \\ &= c\lambda \exp\left(\frac{1}{\theta^2}\left(b^T s + \frac{||b||^2}{4}\right)\right) \end{aligned}$$

where $c = c(\theta, \mathbb{U})$. It is seen that the mean function has the directional trend. For example, for $b = (b_1, 0)'$, the mean function has an "eastness" trend. The covariance function of the intensity random field $\Lambda(s)$ is given by

$$\begin{aligned} \gamma(s,s') &= \int_{\mathbb{U}} \frac{k(s,u)k(s',u)}{b^2(u)} \lambda(du) - E\Lambda(s)E\Lambda(s') \\ &= \frac{\lambda}{\sqrt{2}\pi\theta^2} \frac{\lambda}{\sqrt{2}\pi\theta^2} \exp\left(-\frac{1}{2\theta^2}\left(||s-s'||^2 - b^T(s+s') - ||b||^2\right)\right) \\ &\quad \times \int_{\mathbb{U}} \frac{1}{\pi\theta^2} \exp\left(-\frac{1}{\theta^2}||u - \frac{s+s'+b}{\sqrt{2}}||\right)^2 du - E\Lambda(s)E\Lambda(s') \qquad (2.2) \\ &= \frac{c\lambda}{\sqrt{2}\pi\theta^2} \exp\left(-\frac{1}{2\theta^2}\left(||s-s'||^2 - 2b^T(s+s') - ||b||^2\right)\right) \\ &\quad - c^2\lambda^2 \exp\left(\frac{1}{\theta^2}\left(b^T(s+s') + \frac{||b||^2}{2}\right)\right) \end{aligned}$$

where the covariance function consists of a stationary part and directional part.

### Point Pattern Voronoi Tessellations

In spatial data analysis, Voronoi tessellation provides a way of translating the state space into spatial attributes, $\theta$, for instance, which can be in turn characterized through parameter functions or unknown quantities for modeling inference. *Voronoi* ***tessellation*** is a method of partitioning a (state) space into convex polygonal regions; divisions of the plane are most often discussed. The Voronoi cell of a spatial object is the set of all points in the state space closer to the object than to any other

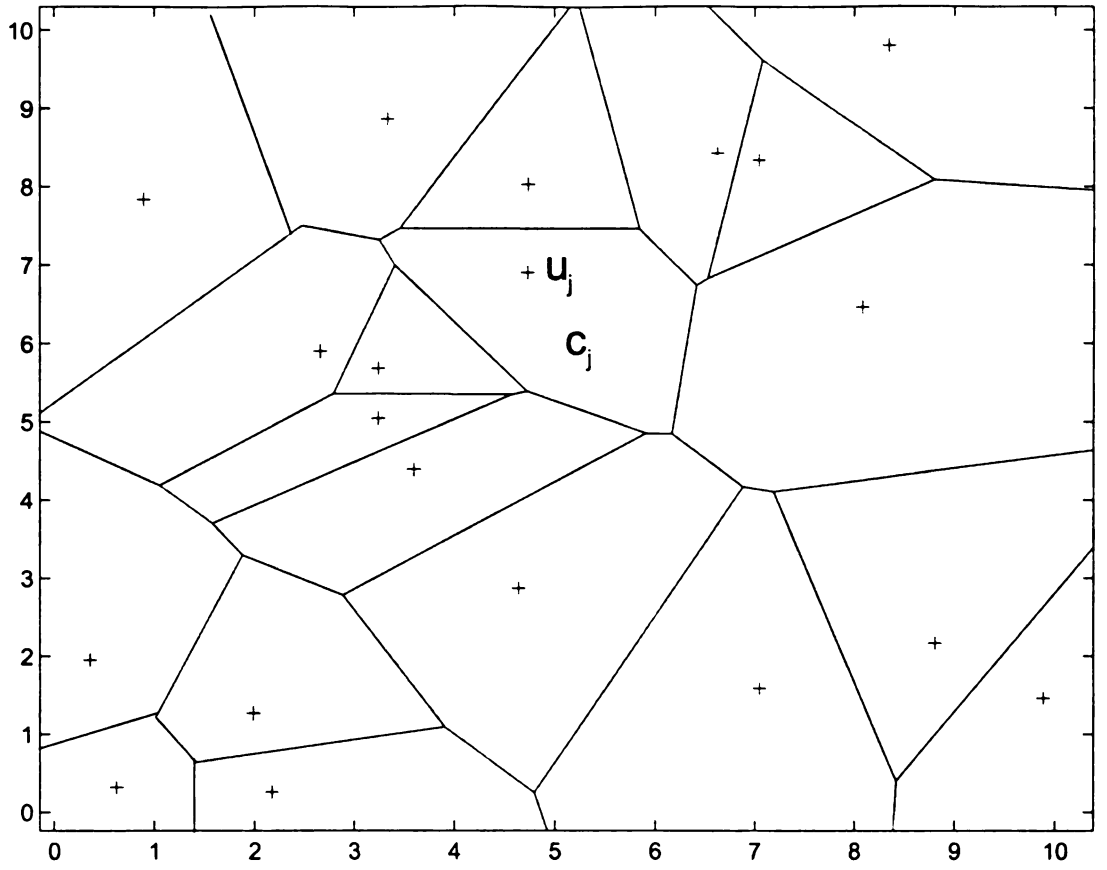


Figure 2.1: Homogeneous Poisson Voronoi tessellation

object in the set. The set of Voronoi cells for a set of spatial objects, called *Voronoi* ***diagram*** (also known as Dirichlet or Thiessen polygons), provides the partition of a set of spatial objects according to its spatial structure. For a theoretical treatment of Voronoi tessellations and mathematical proofs concerning their properties refer to Møller (1994) and Okabe et al. (2000). The latter follows an application oriented approach. In a two-dimensional Euclidean space, Voronoi cells are convex polygons, whose edges are half-way between two points in $\mathbb{U}$ and perpendicular to the line segment joining them. Green and Sibson (1978) provide an efficient computer algorithm for constructing a Voronoi tessellation from a given set of points. Okabe et al. (2000) compare the Green-Sibson to alternative construction methods.

The Voronoi tessellation generated from a random point pattern is a random partition of the state space (Figure 2.1). Miles (1972) gives an excellent survey for the case where a Poisson process generates the Voronoi tessellation. For a homogeneous planar Poisson-Voronoi tessellation of rate $\lambda$, the expected number of sides (or edges) in an arbitrarily-chosen cell (e.g., the cell containing the origin) is 6, and its expected area is $1/\lambda$. The connection between the intensity and tessellation cell attributes thus provides a way of representing a point pattern. This can be extended to inhomogeneous point processes, in which the intensity measure is characterized through the parameter functions. Probably the most useful application of the Voronoi model lies in compilation of intensity surface maps from point data. As the area of cells estimates the inverse value of the local intensity, assigning $1/\lambda$ to the point location as local intensities, the map can be easily compiled with the aid of any kind of (either deterministic or stochastic) interpolation method.

We will use the lattice or gridding tessellation of point patterns, where the corresponding spatial models are then specified based on count data instead of point pattern data directly. We shall take a state space $\mathbb{S}$ of the rectangular shape with length $I_1 w$ and width $I_2 w$, where $w$ is the length of the lattice cell. The number $I = I_1 \times I_2$ is the total number of lattice cells. A cell is referred to as $C_i = C_{(i_1,i_2)}$. The center of a lattice cell $C_i$ is then $((i_1, i_2) - 1/2)w$ used for necessary approximation, and the area of a lattice cell is $w^2$. Each sampling window $A_i$ can be considered as the union of $C_i$s with generic location at $s_i$, and $A = \cup_i A_i$. Similarly, a reference space $\mathbb{U}$ is defined in an analogous way indexed by $J$ with $\mathbb{S} \subset \mathbb{U}$ for correcting the edge effect and with generic locations $u_j$. In general the sampling windows may be any shape or size. We shall take the sampling windows as the union of neighboring square lattice cells (or quadrants) for convenience.

**Data Representation and Approximation**

Tessellation is an effective way of approximating the sample paths of a stochastic process on the state space. However, the accuracy of the approximation highly depends on the smoothness of the sample paths. The smoothness of parametric functions is problem specific. For deterministic kernels, $k$, Marcus and Rosiński (2000) derived explicit and easily verifiable sufficient conditions for the almost sure continuity of the stochastic integrals (1.3) directed by the $GG$-measures. Tessellation is used to provide data driven approximation of parameter functions or unknown quantities in the spatial models. For smooth parameter functions, we may integrate parameter functions over tessellation cells or evaluate parameter functions at any point in the tessellation cell.

In this study, we follow the generic (or cell center) location approximation for the smoothing parameter functions $b(u)$ and random intensity function $\Lambda(s)$. Based on the above gridding tessellation scheme, we have $\lambda_j = \lambda w^2$, $b_j = \exp(u^T b/\theta; u \in C_j) \approx \exp(u_j^T b/\theta)$, $k_{ij}^\theta \approx k^\theta(s_i, u_j)$. With $\boldsymbol{\theta} = (\theta, \lambda_j, b_j) \sim \pi(\boldsymbol{\theta})$, a lattice model is given by

$$
\begin{aligned}
[M_j|\boldsymbol{\theta}] &\sim \prod_j GG(0, \lambda_j^{\boldsymbol{\theta}}, b_j^{\boldsymbol{\theta}}) \\
\Lambda_i &= \sum_j k_{ij}^\theta M_j ds_i \qquad (2.3) \\
[N_i|\boldsymbol{\theta}, \{M_j\}] &\sim \prod_i \mathrm{Poi}(\Lambda_i).
\end{aligned}
$$

With known counts $\{N_i\}$, Ickstadt et al. (1996) show that the features (including mean vectors and covariance matrices) of the joint distribution of random variables

$\{M_j\}$ and $\{N_i\}$ conditional on $\boldsymbol{\theta} \in \Theta$, can be studied from their Laplace exponents, i.e., the negative logarithms of their Laplace transforms,

$$
\begin{aligned}
-\log E_\theta(e^{-\sum_B f_B M_B}) &= \sum_j \log(1 + f_j/b_j)\lambda_j \\
-\log E_\theta(e^{-\sum_Y f_Y N_Y} | M_j) &= \sum_{ij} (1 - e^{-f_B}) k_{ij} M_j \\
-\log E_\theta(e^{-\sum_Y f_Y N_Y}) &= \sum_j \log\left(1 + \sum_i (1 - e^{-f_B}) k_{ij}/b_j\right) \\
&\quad \times \lambda_j
\end{aligned}
\tag{2.4}
$$

## 2.2 Counts and Binary Data

In the rest of this chapter, we illustrate inference for partially censored data on a lattice. The data consists of counts $\mathbf{N} = \{N_i\}$ for each lattice cell within sampling windows, and binary variables $\mathbf{Z} = \{Z_i\}$, where

$$Z_i = I\{N_i^z \geq n_0\}.$$

Here $n_0$ is the level of censoring and the unobserved counts $N_i^z$ are assumed to be on the same path as the observed counts $N_i$.

In simulation of Section 2.3, the exact locations of the events are known. The counts in lattice cells are represented by independent Poisson random variables $\mathbf{N}^+ = \{N_i\} \cup \{N_{i'}^z\}$ with $N_{i'}^z$ denoting the counts in a binary observation cell $C_{i'}$. The Poisson means are a linear combination (approximation) in (2.3) of unobserved

independent Gamma-distributed random measure $\mathbf{M} = \{M_j\}_{j \in J}$ with uncertain non-negative coefficients $K^\theta = \{k_{ij}^\theta\}$. Uncertainty about the parameters of the Gamma distribution of the measures $M_j \sim GG(0, \lambda_j, b_j)$, and about the values or collateral dependence of the coefficients $k_{ij}^\theta$ is modeled through dependence of $k_{ij}^\theta$, $\lambda_j = \lambda_j^\theta$ and $b_j = b_j^\theta$ on an unknown parameter $\theta \in \Theta$, regarded as a random variable taking values in $\Theta \subset \mathbb{R}^p$ for some $p$, with intensity $\pi(\theta)$ with respect to Lebesgue measure $d\theta$. Dependence on $\theta$ is indicated by sub- or super-scripts.

Given the random measure $M$, the $\mathbf{N}^+$ are conditionally independent Poisson random variables with means $\mathbf{\Lambda} = \{\Lambda_i\}$. It follows from (2.4) that unconditionally the $\{N_i\}$ are dependent, distributed as sums of independent negative-binomial random variates. Their means and covariances are given by

$$\begin{aligned} E_\theta N_i &= \sum_j k_{ij}^\theta \lambda_j / b_j \\ \operatorname{cov}(N_i, N_{i'}) &= \sum_j (\delta_{i'}^i + k_{ij}^\theta / b_j)(k_{i'j}^\theta / b_j)\lambda_j \end{aligned} \tag{2.5}$$

where $\delta_{i'}^i$ is the Kronecker symbol, 1 if $i = i'$ and 0 otherwise. From (2.5) we see that $N_i$ and $N_{i'}$ are uncorrelated (and in fact independent) only if $k_{ij}^\theta k_{i'j}^\theta = 0$ for all $j$and each $i \neq i'$. That is, $k_{i\cdot}^\theta$ and $k_{i'\cdot}^\theta$ are orthogonal in the sequence space $\ell^2(J, \lambda_\cdot^\theta / (b_\cdot^\theta)^2)$.

## 2.3 Joint Inference on Counts and Binary Random Fields

### 2.3.1 Graphical Representation

The specification of spatial models may become complicated when introducing many parameters, especially in the presence of data augmentation and collateral stochastic processes. Gilks et al. (1996) develop a graphical representation, called ***directed acyclic graph*** (DAG), to describe the conditional independence for model components. The DAG representation of the model under consideration in this Chapter is shown in Figure 2.2. The statement "$\boldsymbol{\theta}$ influences $M$" describes $\boldsymbol{\theta}$ as being a parent of $M$, as shown by the arrow that goes from $\boldsymbol{\theta}$ to $M$. The DAG makes statements about a sequence of conditional distributions. Each node is independent of all its other ancestors given its parents. Types of nodes can be differentiated (Gilks et al., 1996). We shall use squares about $\mathbf{N}$ and $\mathbf{Z}$, and circles about $\boldsymbol{\theta}$ and $M$, for observed and unknown quantities, respectively. In addition, the dotted circles about $\mathbf{N}^*$ and $V$, denote augmented data nodes. The model dependence structure and assumptions can be readily derived from the directed acyclic graph. The implementation of a MCMC scheme will accordingly visit each node in a DAG. We leave out the implicit functional relationship among unknown quantities for simplicity in a DAG scheme, although they are important in deriving full conditional distributions. Several DAGs related intensity models of spatial point processes are given in the next chapter.

The implementation of MCMC inference below requires samples from the full conditional distribution of the random measure $\mathbf{M}$, given $\boldsymbol{\theta}$ and the observed counts $\mathbf{N} = \mathbf{n}$, a mixture distribution from which direct sampling is cumbersome. Therefore,

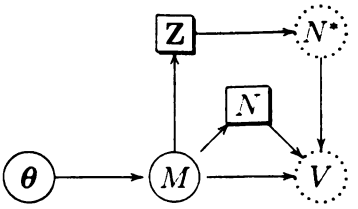


Figure 2.2: DAG of counts and binary random fields

we introduce latent "augmentation" variables $V = \{V_{ij}\}_{i,\, j \in I \times J}$, which can be sampled easily and which lead to a mixture-model representation of model (2.3) and to the conjugate (Gamma) full conditional distribution for **M**. More specifically, conditional on $\mathbf{M} = \mathbf{m}$, $\mathbf{N} = \mathbf{n}$ and $\mathbf{Z} = \mathbf{z}$, where $Z_i$ is the indicator variable, $n_0$ is assumed to be a constant. Also the augmented counts $\{N_i^z\}$ corresponding to binaries variables $\{Z_i\}$ are conditionally independent Poisson random variables with means $\{\Lambda_i\}$. Set $\Lambda_{ij} = k_{ij} m_j$, $\Lambda_{1i} = \sum_j \Lambda_{ij}$, $p_{ij} = \frac{\Lambda_{YB}}{\Lambda_{1Y}}$ and let $\{V_{ij}\}$ have independent Multinomial distributions $MN(n_i^+, \mathbf{p}_i)$. We might regard the random variates $V_{ij}$ as the portions of counts $n_i$ attributable to the $j$th measure, $j \in J$ in some applications.

### 2.3.2 MCMC Scheme

We now study the (analytically intractable) posterior distribution of uncertain quantities $\theta$ and **M** by simulating steps from an ergodic Markov chain with the posterior as its stationary distribution. The joint distribution of all uncertain quantities are given by the following theorem

**Theorem 2.3.1.** *The joint distribution of* $\boldsymbol{\theta}$, **M**, $V$, **N**, $\mathbf{N}^z$, *and* **Z** *has a density function with respect to the product of counting measure for discrete and Lebesgue*

*measure for continuous variables, given by*

$$[\boldsymbol{\theta},\mathbf{M}, V, \mathbf{N}, \mathbf{N}^z, \mathbf{Z}] = \pi(\boldsymbol{\theta}) \times \left(\prod_i \left(\frac{z_i}{p_i^z} + \frac{1-z_i}{1-p_i^z}\right)\right) \left(\prod_{j\in J} \frac{m_j^{\lambda_B + v_{2B}-1} b_j^{\lambda_B}}{\Gamma(\lambda_j)}\right) \times \left(\prod_{ij} \frac{k_{ij}^{v_{YB}}}{v_{ij}!}\right) \exp\left(-\sum_j (b_j + k_{2j}) m_j\right)$$

*where* $p_i^z = \mathbb{P}(N_i^z \geq n_0)$, $v_{2j} = \sum_i v_{ij}$ *and* $k_{2j} = \sum_i k_{ij}$.

*Proof.* With the adopted notations, and from the directed graph presentation, we have the following factorization under the considered model,

$$[\boldsymbol{\theta}, \mathbf{M}, V, \mathbf{N}, \mathbf{N}^z, \mathbf{Z}] = [\boldsymbol{\theta}][\mathbf{M}|\boldsymbol{\theta}][V|\boldsymbol{\theta}, \mathbf{M}, \mathbf{N}, \mathbf{N}^z, \mathbf{Z}, n_0][\mathbf{N}|\boldsymbol{\theta}, \mathbf{M}]$$
$$[\mathbf{N}^z|\boldsymbol{\theta}, \mathbf{M}, \mathbf{Z}, n_0][\mathbf{Z}|\boldsymbol{\theta}, \mathbf{M}, n_0]$$

Note that $[V|\boldsymbol{\theta}, \mathbf{M}, \mathbf{N}, \mathbf{N}^z, \mathbf{Z}] = [V|\boldsymbol{\theta}, \mathbf{M}, \mathbf{N}, \mathbf{N}^z]$, in which $\mathbf{Z}$ is redundant, is the product of multinomial distributions. $[\mathbf{N}|\boldsymbol{\theta}, \mathbf{M}]$ and $[\mathbf{N}^z|\boldsymbol{\theta}, \mathbf{M}]$ are products of Poisson distributions. $[\mathbf{Z}|\boldsymbol{\theta}, \mathbf{M}]$ can be evaluated by computing the tail probabilities of a Poisson distribution. Direct calculation will lead to the result. □

In the above theorem, we consider $n_0$ to be a fixed value. In practice, this may be estimated by the available count data and collateral information.

**Corollary 1.** *The conditional distribution of the augmented data vector* $(V, \mathbf{N}^z)$ *given the observed counts* $\mathbf{n}$ *and* $\mathbf{z}$*, the signal measure* $\mathbf{M}$*, and the parameter* $\boldsymbol{\theta}$ *is the*

*product of multinomials with parameters* $n_i$ *and* $\mathbf{p}_i$, *independent for* $i \in I$ *and* $[\mathbf{N}^z = \mathbf{n}^z|\boldsymbol{\theta}, \mathbf{M}, \mathbf{z}]$, *i.e.*,

$$[V = v, \mathbf{N}^z = \mathbf{n}^z|\mathbf{n}, \mathbf{z}, \mathbf{m}, \boldsymbol{\theta}] =$$

$$\prod_i MN(n_i^+, \mathbf{p}_i) \left( \frac{z_i}{p_i^z} + \frac{1 - z_i}{1 - p_i^z} \right) \frac{\Lambda_i^{n_Y^E}}{n_i!} \exp(-\Lambda_i)$$

**Corollary 2.** *The conditional joint distribution of* $\mathbf{M}$ *given* $\mathbf{N}$, $\mathbf{Z}$, $V$, *and* $\boldsymbol{\theta}$ *is again gamma component, independent for* $j \in J$ *with parameters* $\lambda_j + v_{2j}$ *and* $b_j + k_{2j}$, *i.e.*,

$$[\mathbf{M}|\boldsymbol{\theta}, v, \mathbf{n}, \mathbf{z}] = \prod_j GG(0, \lambda_j + v_{2j}, b_j + k_{2j})$$

**Corollary 3.** *The conditional density of* $\boldsymbol{\theta}$ *given* $\{n_i\}$, $\{z_i\}$, $\{v_{ij}\}$, *and* $\{m_j\}$ *is given by*

$$[\boldsymbol{\theta}|\mathbf{M}, V, \mathbf{N}] \propto \pi(\boldsymbol{\theta})$$

$$\left( \prod_{j \in J} \frac{m_j^{\lambda_B + v_{2B} - 1} b_j^{\lambda_B}}{\Gamma(\lambda_j)} \right) \left( \prod_{ij} \frac{k_{ij}^{v_{YB}}}{v_{ij}!} \right)$$

$$\exp\left( - \sum_j (b_j + k_{2j}) m_j \right)$$

*Proof.* Immediate from Theorem 2.3.1 and the definition of $\{V_{ij}\}_{i,j}$. □

The above corollaries lead directly to the following MCMC scheme. Given a prior density $\pi(\theta)$ on $\Theta$, a parametric kernel $\{k_{ij}^{\theta}\}$, a transition probability kernel

$Q(\theta, d\theta') = q(\theta, \theta')d\theta'$, parameters $\{\lambda_j\}$ and $\{b_j\}$ for the Gamma distributions of the random measures $\{M_j\}$, and initial points $\theta^0 \in \Theta$, and nonnegative integers $\{v_{ij}\}$, we can generate successive points starting at $t = 0$ as follows:

**Step I:** Gibbs step to update the random measure. Given $\{n_i\}$, $\{z_i\}$, $\{v_{ij}^t\}$, and $\theta^t$,

- set $\lambda_j^{t+1} = \lambda_j^{\theta^S} + v_{2j}^t$ and $b_j^{t+1} = b_j^{\theta^S} + k_{2j}^{\theta^S}$;
- generate independent $M_j^{t+1} = m_j^{t+1}$ with $GG(0, \lambda_j^{t+1}, b_j^{t+1})$;
- set $\Lambda_{ij}^{t+1} = k_{ij}^{\theta^S} m_j^{t+1}$, $\Lambda_{1i}^{t+1} = \sum_{j \in J} \Lambda_{ij}^{t+1}$, and $p_{ij}^{t+1} = \Lambda_{ij}^{t+1} / \Lambda_{1i}^{t+1}$

**Step II:** Gibbs step to update the augmentation points. Given $\{n_i\}$, $\{z_i\}$, $\{m_j^{t+1}\}$, and $\theta^t$,

- generate independent random variates $N_i^z$ with probability

$$\left(\frac{z_i}{p_i^z} + \frac{1 - z_i}{1 - p_i^z}\right) \frac{\Lambda_i^{n_Y^E}}{n_i!} \exp(-\Lambda_i) \tag{2.6}$$

  for each $i$w here the indicator random variable $Z_i$ is observed;

- generate $\{V_{ij}^{t+1}\}_j \sim MN(n_i^+, \mathbf{p}_i^{t+1})$ independent for $i \in I$.

**Step III:** Metropolis-Hastings step to update the parameter $\theta$. Given $\{n_i\}$, $\{v_{ij}^{t+1}\}$, $\{m_j^{t+1}\}$, and $\theta^t$

- generate a new candidate $\theta' \sim Q(\theta^t, d\theta')$;

- calculate the acceptance probability

$$P' = \frac{\pi(\theta')q(\theta',\theta^t)}{\pi(\theta^t)q(\theta^t,\theta')} \prod_{ij} \left( \frac{k_{ij}^{\theta'}}{k_{ij}^{\theta^S}} \right)^{v_{YB}^{S+1}} \times \prod_j \frac{(m_j^{t+1})^{\lambda_B^{I'}+v_{2B}^{S+1}-1}(b_j^{\theta'})^{\lambda_B^{I'}}}{(m_j^{t+1})^{\lambda_B^{I^S}+v_{2B}^{S+1}-1}(b_j^{\theta^S})^{\lambda_B^{I^S}}} \frac{\Gamma(\lambda_j^{\theta^S})}{\Gamma(\lambda_j^{\theta'})} \times \exp\left( -\sum_j (b_j^{\theta'} - b_j^{\theta^S} + k_{2j}^{\theta'} - k_{2j}^{\theta^S}) m_j^{t+1} \right) \tag{2.7}$$

- generate $\theta^{t+1} = \theta'$ with probability $(1 \wedge P')$ and $= \theta^t$ otherwise.

**Step IV:** set $t \leftarrow\!\rho\, t+1$ and return to **Step I**.

## 2.4 Simulation and Results

### 2.4.1 Simulation Setup

Consider the model (2.3) with the state space, $[0.2, 1] \times [0.2, 1]$, and the reference state space, $[0, 1.2] \times [0, 1.2]$. The width of the lattice cell (quadrant) is set to be 0.2. The control measure is proportional to Lebesgue measure with the proportion coefficient $\lambda = 1000$, the parameter $\theta = 0.4$ for modeling the circular correlation structure, and the spatial trend parameters $b^T = (b_1, b_2) = (1, 0)$ for modeling "eastness" in this case. One run with true parameters is performed to get the raw count data for all 16 lattice cells. We set $n_0 = 10$ to be equal to the average of raw counts. Six cells

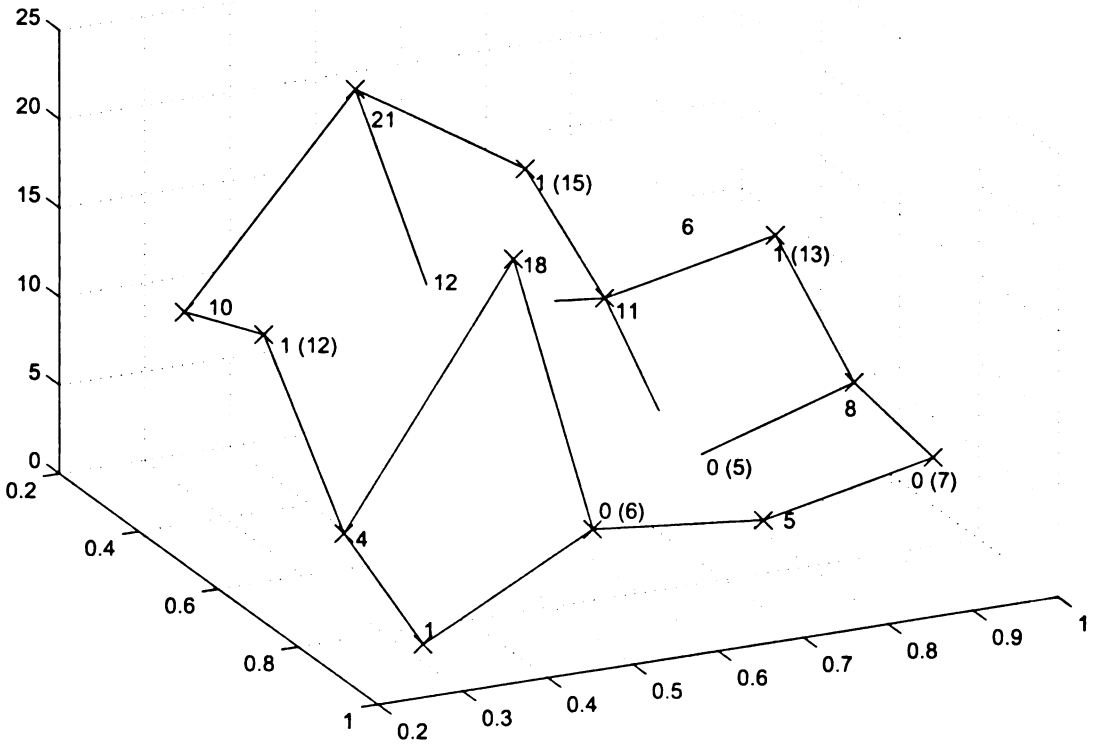


Figure 2.3: Empirical counts plot

are randomly selected, and their counts are compared with $n_0$ to get the indicator variables, $Z_i = I\{N_i \geq n_0\}$. Hence, the simulated data consist of six $0-1$ indicator values and raw counts for the rest of the 10 cells. The outcome of the simulated data is shown in Figure 2.3, in which the true counts for indicators are shown in parentheses.

The sensibility of model parameters in modeling this data is studied by comparing the three models shown in Table 2.1. Model I has all three parameters unspecified and to be estimated. It is most general in the sense of catching all the directional covariance structure that is implied in (2.2). Model II is a wrong model in this case in that it takes $b_1 = 0$ even though the data were generated from a model with $b_1 = 1$,

| Model | Parameters | Estimates |
|---|---|---|
| True Model | $(0.4, 1, 0)$ | |
| Model I | $(\theta, b_1, b_2)$ | $(0.93, 0.89, -1.00)$ |
| Model II | $(\theta, 0, b_2)$ | $(1.61, —, -0.91)$ |
| Model III | $(\theta, b_1, 0)$ | $(1.77, -3.10, —)$ |

Table 2.1: True model and other models

that is, the "eastness" is expected. Finally, Model III is the correct model in that $b_2 = 0$ is assumed and is expected to give better estimates and prediction. The values in the first column of the table is the posterior sample mean corresponding to each model.

For each model we have chosen independent standard normal prior distributions for each parameters $\theta$, $b_1$ and $b_2$ and independent $GG(0, \lambda_j, b_j)$ distributions for $M_j$ with $\lambda_j \equiv \lambda w^2$ constant in all $j$ and $b_j$ as defined before. To perform the Metropolis-Hastings step, the transition distributions for $b_1$ and $b_2$ are also normal with standard deviations $\psi_1$ and $\psi_2$. For the parameter $\theta$, first a transformation is imposed such that $\xi = \log\theta$, then $\xi'_{t+1}$ is generated from the proposal density $q_\xi(\,\cdot\,|\xi_t) = N(\xi_t, \psi_3^2)$, and the value is accepted proportional to

$$\frac{q_\theta(\theta'|\theta)}{q_\theta(\theta|\theta')} = \frac{\theta}{\theta'} \exp\left(-\frac{1}{2\psi_3^2}\left(\log(\theta'/\theta) + \log(\theta/\theta')\right)\right).$$

The tuning scale parameters $\psi_i$ control the rate of acceptance $P'$ in (2.7) of the Metropolis-Hastings step, and a judicious selection is needed for efficiency of the algorithm. If $\psi_i$'s are too small, the rate of acceptance $P'$ will be high, but the steps such as $\xi'_{t+1} - \xi_t$ will in general be small. Such a chain will move very slowly through the support of the posterior distribution. On the other hand, if $\psi_i$'s are too large, the proposal distribution will generate large steps, often from the body to the tails of the posterior distribution, so the rate of acceptance will be small. Such a chain will frequently not move, producing slow mixing again. In both cases it would

take a large number of iterations for the chain to move through the support of the posterior distribution, making the algorithm inefficient.

A rule-of-thumb used by many practitioners is to select (by experimentation) values of $\psi_i$'s giving an empirical rate of acceptance of about $0.4 - 0.5$. This rule has some theoretical support for the case when the target and proposal distributions are both normal (Gelman et al., 1996). In this study, for model I, we use $\psi_i = 1$ for all $i = 1, 2, 3$, while for model II and III, $\psi_i = 2$, and the empirical acceptance rate was about 0.40. Implementation for all models was based on 100,000 MCMC steps, and the posterior sampling was started at 90,000 and successive samples were taken at every other 100 runs. The Monte Carlo iteration was realized by a `C++` program. The judgement as to how close an iterative algorithm is to convergence after a finite number of iterations is difficult. Some discussion is found in Kass et al. (1997).

### 2.4.2 Posterior Analysis

One of our simulation goals is to predict intensities $\{\Lambda_i\}$ for those lattice cells without known count data. The predictions of the intensities are obtained by the corresponding posterior sample means of size 100. The predictions for Model I are shown in Figure 2.4. The smooth prediction surface for intensities is curved instead of showing a pure "eastness" trend. This is because of the presence of the parameter $b_2$, which measures the "northness" trend. In fact, Model I can explain all directional covariance structure if present. However, the other models did not improve the prediction surfaces in the sense of catching the covariance structure during the experimentation.

The following figures are plotted using the function `plotdens()` from `MatLab`'s statistics toolbox. Figures 2.5, 2.6 and 2.7 show the prior (solid curves), true value

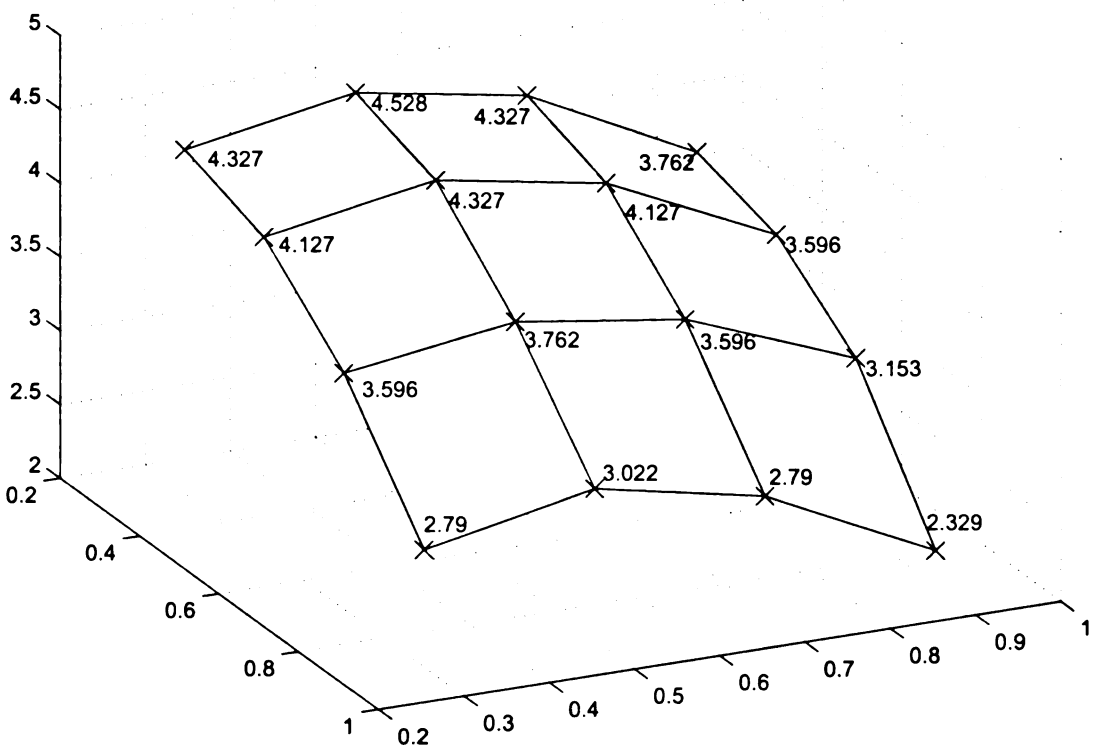


Figure 2.4: Posterior means for intensities over lattice cells

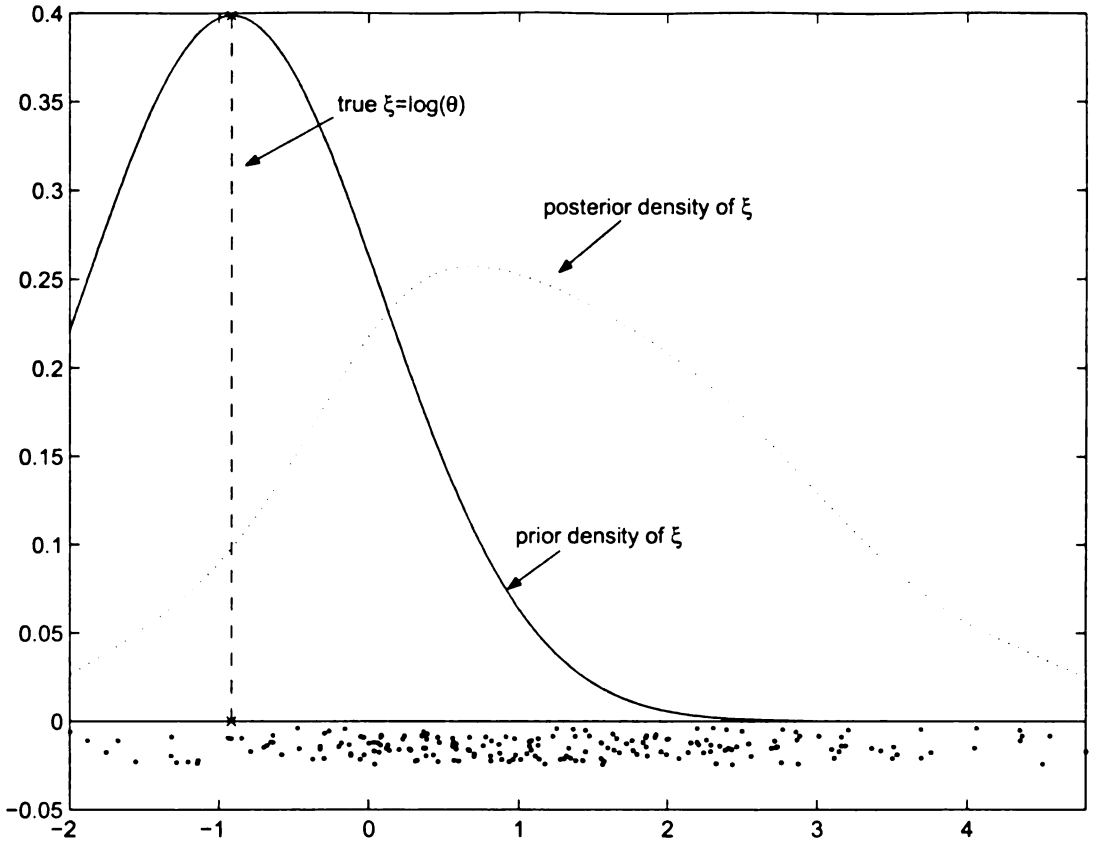


Figure 2.5: Posterior density for log-$\theta$

(dashed vertical lines) and the estimated posterior density function (dotted curves) for $\xi = \log\theta$, $b_1$ and $b_2$, respectively. The posterior means are listed in Table 2.1. The estimated densities for $\xi$ and $b_1$ suggest that the systematic trends should be included in the analysis. The estimated density for $b_2$ tends to be flat and suggests a weak "northness" trend that is consistent with the true model.

While the above graphical outputs are very suggestive of which models fit well and which do not, a more objective and more quantitative approach, using so called ***Bayes factors***, can be applied (Ickstadt, et al., 1996). This was not done in our example.

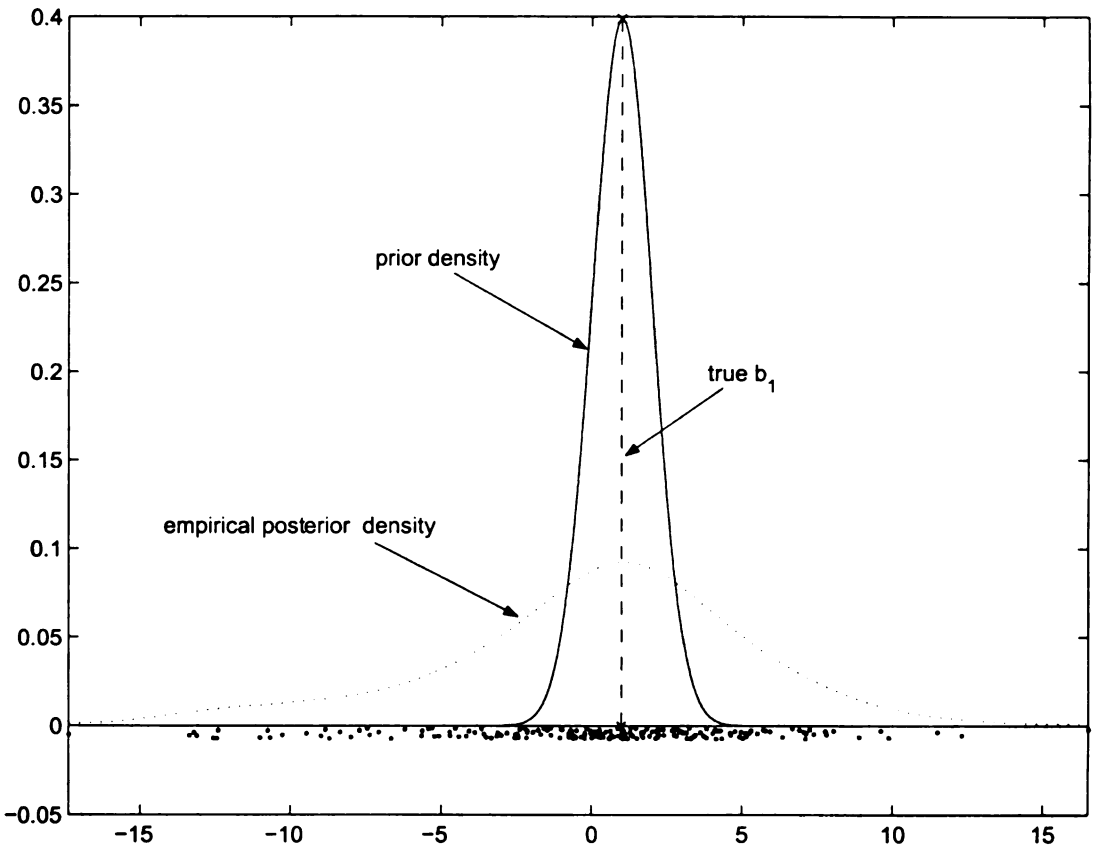


Figure 2.6: Posterior density for $b_1$

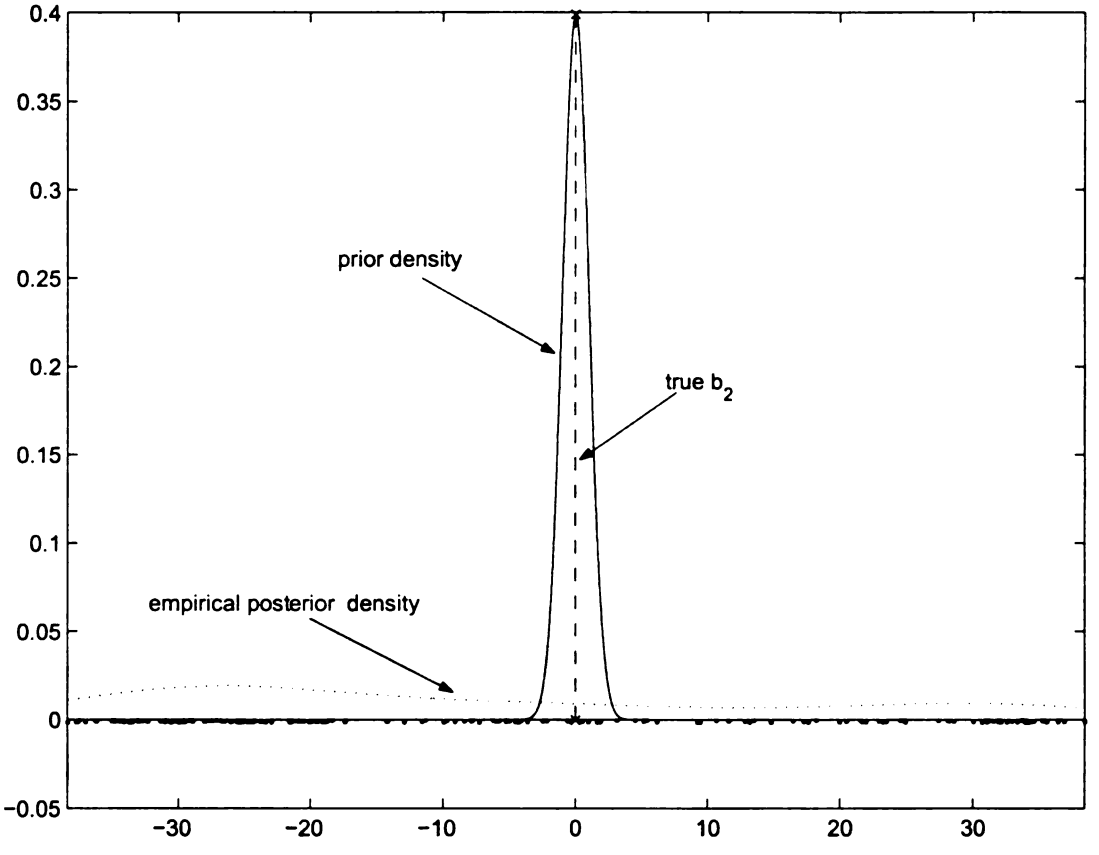


Figure 2.7: Posterior density for $b_2$

# Chapter 3. Models Based on Log Gaussian Random Fields

In this chapter we continue to focus on developing models based on inhomogeneous point patterns. For highly clustered point patterns, it may be more suitable to perform the geostatistical analysis after an appropriate transformation of the count data. The imposed transformation is convenient for factoring the intensity models, and is required to achieve the Gaussian regularity for geostatistical models. This commonly happens to be the log-transformation. More general transformation methods are discussed by Martinez (1997). Geostatistical models for log-counts tend to explain large scale variation rather than small scale variation in point patterns (Heisel et al., 1996).

In the geostatistical modeling setting, a log intensity model is suggested by the intensity model (1.17) and is given by

$$\begin{aligned} \eta(s) &= \beta(s) + \xi(s) \\ \Lambda(s) &= \exp(\beta(s))\Lambda_0(s) = \exp(\eta(s)) \end{aligned} \tag{3.1}$$

where $\xi(s)$ is a Gaussian random field on $\mathbb{S}$. The direct specification of covariance function of $\xi(s)$ is convenient under stationarity. For non-stationary covariance function of $\xi(s)$, we shall instead consider a second order random measure $W$ with control measure $\lambda$ on $\mathcal{U}$. Its subordinated random field $\xi(s)$ is given by

$$\xi(s) := \int_{\mathbb{U}} k(s,u)W(du), \quad s \in \mathbb{S} \tag{3.2}$$

then $\xi(s)$ is diffuse and is well defined on $\mathbb{S}$; see page 10 for the concept of diffuse measure. Clearly, $\xi$ is a centered second order random field, i.e., $E\xi(s) = 0$, and its covariance structure can be specified by choosing appropriate kernels and parameter

functions defining random measures. It is more convenient to work with the kernels $k(s, \cdot)$ rather than directly with the covariance function $\gamma(s, s')$, in which it might be difficult to ensure symmetry and positive definiteness for all $s$ and $s'$. Under the condition (1.4), the specification of $\xi$ in (3.2) is valid with arbitrary choice of kernels and the derived the covariance function $\gamma(s, s')$ of $\xi$ is positive definite (Higdon et al., 1999). Under the suggested transformation in (3.1), the relation between $\Lambda_0(s)$ and $\xi(s)$ is defined through the exponential link function, which is chosen so that the intensity is nonnegative. The covariate factor $\beta(s)$ is an additive term in the transformed model, and in this setting, $\xi(s)$ may also be interpreted as a covariate process to point patterns.

## 3.1 Models for Log Counts

### 3.1.1 Model Specification

Consider the log-count data, denoted by $\mathbf{Y} = (Y(s_1), \cdots, Y(s_n))'$ at the generic locations $s_1, \cdots, s_n$. In this section we discuss the model

$$Y_i = x_i^T \boldsymbol{\beta} + \xi_i + \epsilon_i \tag{3.3}$$

where $\xi_i = \xi(s_i)$ is from an underlying spatial stochastic process and $\epsilon_i = \epsilon(s_i)$ is from a spatial error process. In spatial data analysis, the spatial variation includes two sources, i.e., the spatial uncertainty and measurement error. Measuring heterogeneity is sometimes critical to the success of an error modeling system, especially in the case where the measurements have underlying spatial patterns. Spatial processes $\xi(s)$ and $\epsilon(s)$ are assumed to be independent, and the $Y_i$'s are thus conditionally independent given $\xi(s)$.

In the sequel, $\epsilon_1, \cdots, \epsilon_n$ are assumed to be i.i.d normal random variables with mean zero and variance $\sigma^2$, and $\xi(s)$ is a Gaussian random field with zero mean, i.e., $E\xi(s) = 0$, and with covariance function $\gamma_\theta(s, s') = \text{cov}(\xi(s), \xi(s'))$ depending on a parameter $\theta \in \Theta$. We let $\boldsymbol{\gamma}_s := (\gamma(s_1, s), \cdots, \gamma(s_n, s))^T$ and let $\Gamma = \{\gamma(s_i, s_j)\}$ denote the $n \times n$ covariance matrix at generic locations. Although not demanded, it is often assumed that $\xi(s)$ has a constant variance $\tau^2$, i.e.,

$$\gamma_\theta(s, s') = \tau^2 \gamma^*_{\theta^*}(s, s') \tag{3.4}$$

where $\gamma^*$ is the correlation function specified by $\theta^*$ with $\theta = (\tau, \theta^*)$. The mean for $Y_i$ is

$$\mu(s_i) = EY(s_i) = E(E(Y(s_i)|\xi)) = \beta(s_i) = x_i^T \boldsymbol{\beta} \tag{3.5}$$

where $x_i = x(s_i)$ is the $p$-dimensional covariate vector. The covariance matrix for $\mathbf{Y}$ is given by

$$\gamma^y(s_i, s_j) = \begin{cases} \sigma^2 + \gamma(s_i, s_j), & \text{for } i = j; \\ \gamma(s_i, s_j), & \text{for } i \neq j. \end{cases}$$

and is conveniently written as

$$\text{cov}(\mathbf{Y}) = \sigma^2(I + \Gamma/\sigma^2) = \sigma^2 \Sigma$$

The additive measurement error $\sigma^2$ is the variance of the conditional random field $Y$, analogous to a ***nugget effect*** in the spatial statistics literature. Also let $\boldsymbol{\sigma}_s := (\gamma^y(s_1, s), \cdots \gamma^y(s_n, s))'$ for any $s \in \mathbb{S}$ and denote the $n \times p$ covariate matrix by $X$.

### 3.1.2 MCMC Inference

In the inhomogeneous spatial models including the non-Gaussian cases, the main difficulty in performing likelihood-based inference and prediction is the shortage of multivariate families of distributions as flexible and operationally convenient as the Gaussian distribution. In most cases, those densities are analytically intractable. The most widely implemented methodology for coping with non-Gaussian problems is trans-Gaussian Kriging and lognormal Kriging (Cressie, 1993), which consist of applying standard Gaussian methods after a marginal non-linear transformation.

When applying log Gaussian models (3.3) for point pattern data analysis, the error term is often omitted (Diggle and Tawn, 1998). In this section we introduce the Bayesian analysis of the log Gaussian model with the error term present. Let $\boldsymbol{\xi} = (\xi(s_1), \cdots, \xi(s_n))'$ and let $\boldsymbol{\xi}^* = (\xi^*(s_{10}), \cdots, \xi^*(s_{m0}))'$ for the corresponding values of $\xi(s)$ at locations $s_{i0}$ for which predictions are required. We are then working on the extended model $[\boldsymbol{\xi}, \boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{Y}, \sigma, \boldsymbol{\xi}^*]$ instead of the model $[\boldsymbol{\xi}, \boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{Y}, \sigma]$, where $\boldsymbol{\theta}$ denotes the parameter vector in the covariance structure $\Gamma_{\boldsymbol{\theta}}$ of $\xi(s)$, in particular, under constant variance in (3.4), $\boldsymbol{\theta}$ includes the variance $\tau^2$, and any further parameters in the specification of the correlation structure of $\boldsymbol{\xi}$, and $\boldsymbol{\beta}$ consists of all the regression parameters. Our objective is to use MCMC methods to estimate the model parameters, $\boldsymbol{\theta}$, $\boldsymbol{\beta}$ and $\sigma$, and to generate samples from the conditional distribution of $(\boldsymbol{\xi}, \boldsymbol{\xi}^*)$ given $\mathbf{Y}$ under the stated assumptions. This allows us to obtain predictions for any functional $L$ of interest associated with this conditional distribution, while making proper allowance for uncertainty in the parameter estimates. Under a standard MCMC scheme we need to generate samples from the posterior distribution of $[(\boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma)|\mathbf{Y}]$ for inference and from the posterior of $[\boldsymbol{\xi}^*|(\boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\beta}, \mathbf{Y}, \sigma)]$ for prediction.

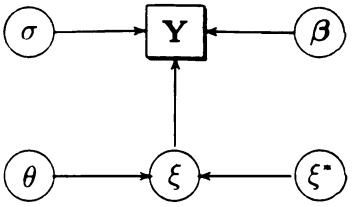


Figure 3.1: DAG of log Gaussian models

The implementation of the MCMC scheme requires sampling from the conditional distributions: $[\boldsymbol{\theta}|\boldsymbol{\xi},\boldsymbol{\beta},\mathbf{Y},\sigma]$, $[\boldsymbol{\beta}|\boldsymbol{\theta},\boldsymbol{\xi},\mathbf{Y},\sigma]$, and $[\boldsymbol{\xi}_i|\boldsymbol{\theta},\boldsymbol{\xi}_{-i},\boldsymbol{\beta},\mathbf{Y},\sigma]$, where $\boldsymbol{\xi}_{-i} = \{(\xi_j); j \neq i\}$. A graphical representation of the dependence structure of the model is shown in Figure 3.1, the model can be factorized into

$$[\boldsymbol{\theta},\boldsymbol{\xi},\boldsymbol{\beta},\mathbf{Y},\sigma,\boldsymbol{\xi}^*] = [\boldsymbol{\theta}][\boldsymbol{\xi}|\boldsymbol{\theta}][\boldsymbol{\beta}][\sigma][\mathbf{Y}|\boldsymbol{\theta},\boldsymbol{\xi},\boldsymbol{\beta},\sigma][\boldsymbol{\xi}^*|\boldsymbol{\theta},\boldsymbol{\xi},\boldsymbol{\beta},\mathbf{Y},\sigma] \tag{3.6}$$

The first objective is inference about model parameters. We can simplify this structure by dropping the node $\boldsymbol{\xi}^*$. It only requires visiting the nodes $\boldsymbol{\theta}$, $\boldsymbol{\xi}$, $\boldsymbol{\beta}$ and $\sigma$ in Figure 3.1. Given $\boldsymbol{\xi}$, it is clear that $\boldsymbol{\theta}$ is conditionally independent of $\mathbf{Y}$, and that $\boldsymbol{\theta}$ is conditionally independent of $\boldsymbol{\beta}$ and $\sigma$. So for inference, a single cycle of the MCMC algorithm involves first updating $[\boldsymbol{\theta}|\boldsymbol{\xi}]$, then $[\boldsymbol{\xi}|\boldsymbol{\theta},\boldsymbol{\beta},\mathbf{Y},\sigma]$, $[\boldsymbol{\beta}|\boldsymbol{\xi},\mathbf{Y},\sigma]$ and $[\sigma|\boldsymbol{\xi},\boldsymbol{\beta},\mathbf{Y}]$.

Under the model (3.3), $[\mathbf{Y}|\boldsymbol{\theta},\boldsymbol{\xi},\boldsymbol{\beta},\sigma]$ is the product of normal densities with respective means $\eta_i = x_i^T\boldsymbol{\beta} + \xi_i$ and variance $\sigma^2$, say,

$$[\mathbf{Y}|\boldsymbol{\theta},\boldsymbol{\xi},\boldsymbol{\beta},\sigma] = [\mathbf{Y}|\boldsymbol{\xi},\boldsymbol{\beta},\sigma] = \varphi(\mathbf{y}|\boldsymbol{\xi},\boldsymbol{\beta},\sigma) \tag{3.7}$$

and

$$[\boldsymbol{\theta}|\boldsymbol{\xi},\boldsymbol{\beta},\mathbf{Y},\sigma] = [\boldsymbol{\theta}|\boldsymbol{\xi}] = [\boldsymbol{\xi}|\boldsymbol{\theta}][\boldsymbol{\theta}] = \varphi(\boldsymbol{\xi}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{3.8}$$

and

$$[\xi_i|\boldsymbol{\xi}_{-i},\boldsymbol{\beta},\mathbf{Y},\sigma] = [\mathbf{Y}|\boldsymbol{\xi},\boldsymbol{\beta},\sigma][\xi_i|\boldsymbol{\theta},\boldsymbol{\xi}_{-i}] \tag{3.9}$$

where $\varphi(\boldsymbol{\xi}|\boldsymbol{\theta})$ in (3.8) is a multivariate normal density. As a consequence, $[\xi_i|\boldsymbol{\theta}, \boldsymbol{\xi}_{-i}]$ in equation (3.9) has a univariate Gaussian distribution. Finally,

$$[\boldsymbol{\beta}|\boldsymbol{\theta}, \boldsymbol{\xi}, \mathbf{Y}, \sigma] = [\boldsymbol{\beta}|\boldsymbol{\xi}, \mathbf{Y}, \sigma] \propto [\mathbf{Y}|\boldsymbol{\xi}, \boldsymbol{\beta}, \sigma][\boldsymbol{\beta}] \tag{3.10}$$

and

$$[\sigma|\boldsymbol{\theta}, \boldsymbol{\xi}, \mathbf{Y}, \boldsymbol{\beta}] = [\sigma|\boldsymbol{\xi}, \mathbf{Y}, \boldsymbol{\beta}] \propto \ [\mathbf{Y}|\boldsymbol{\xi}, \boldsymbol{\beta}, \sigma][\sigma] \tag{3.11}$$

In practice, any prior distributions $[\boldsymbol{\theta}]$, $[\boldsymbol{\beta}]$ and $\sigma$ will make MCMC work.

The MCMC algorithm consists of the following steps. Initial values need to be chosen for $\boldsymbol{\theta}$, $\boldsymbol{\beta}$ and $\boldsymbol{\xi}$ to start the algorithm. Arbitrary starting values for $\boldsymbol{\theta}$,$\boldsymbol{\beta}$ , $\sigma$ and $\boldsymbol{\xi}$ may be chosen in the range specified by the prior. However, assuming $\boldsymbol{\xi} = 0$, our model reduces to a standard generalized linear model and the corresponding estimate of $b$ from this model may be used as an initial value, $\boldsymbol{\beta}^0$. We then set the starting value, $\xi_i^0$, for each $\xi_i$ by equating $\eta_i$ to $y_i$ for each $i$ in the link function defined to give $\xi_i^0 = y_i - x_i^T\boldsymbol{\beta}^0$. $\sigma^0$ can be set to the sample standard deviation.

In a MCMC cycle, we first update all the components of the parameter vector $\boldsymbol{\theta}$ through a MH step to accept a new state $\boldsymbol{\theta}'$ chosen uniformly from the parameter space specified by the prior, with acceptance probability

$$\Delta(\boldsymbol{\theta}, \boldsymbol{\theta}') = \min\left\{\frac{p(\boldsymbol{\xi}|\boldsymbol{\theta}')}{p(\boldsymbol{\xi}|\boldsymbol{\theta})}, 1\right\}$$

The second step is to update the signal $\boldsymbol{\xi}$. Choose a new value, $\xi_i'$, for the $i$th component of $\boldsymbol{\xi}$, from the transition probability function $q(\xi_i, \xi_i') = [\xi'|\boldsymbol{\xi}, \boldsymbol{\theta}]$, and accept $\xi_i'$ with probability

$$\Delta(\xi_i, \xi_i') = \min\left\{\frac{\varphi(y_i|\xi_i', \boldsymbol{\beta})}{\varphi(y_i|\xi_i, \boldsymbol{\beta})}, 1\right\}.$$

Repeat the sub-steps for all $i = 1, \cdots, n$. The third step is to update the elements of the regression parameter $\boldsymbol{\beta}$. Choose a new value $\boldsymbol{\beta}'$ from some appropriate density

$q(\boldsymbol{\beta}, \boldsymbol{\beta}') = [\boldsymbol{\beta}'|\boldsymbol{\beta}]$. Accept $\boldsymbol{\beta}'$ with probability

$$\Delta(\boldsymbol{\beta}, \boldsymbol{\beta}') = \min\left\{\frac{\varphi(\mathbf{y}|\boldsymbol{\xi}, \boldsymbol{\beta}', \sigma)q(\boldsymbol{\beta}', \boldsymbol{\beta})}{\varphi(\mathbf{y}|\boldsymbol{\xi}, \boldsymbol{\beta}, \sigma)q(\boldsymbol{\beta}, \boldsymbol{\beta}')}, 1\right\}.$$

The fourth step is to update the parameter $\sigma$. Choose a new value $\sigma'$ from some appropriate density $q(\sigma, \sigma') = [\sigma'|\sigma]$. Accept $\sigma'$ with probability

$$\Delta(\sigma, \sigma') = \min\left\{\frac{\varphi(\mathbf{y}|\boldsymbol{\xi}, \boldsymbol{\beta}, \sigma')q(\sigma', \sigma)}{\varphi(\mathbf{y}|\boldsymbol{\xi}, \boldsymbol{\beta}, \sigma)q(\sigma, \sigma')}, 1\right\}.$$

The choice of the transition kernel $q$ is problem-specific. We iterate the above steps until the chain is judged to have reached its equilibrium distribution. At this point we draw a random sample from the multivariate Gaussian distribution of $[\boldsymbol{\xi}^*|\boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\beta}, \mathbf{Y}, \sigma]$, where $(\boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma)$ are the values generated in previous updating steps. This step reduces to direct simulation from the Gaussian distribution of $[\boldsymbol{\xi}^*|\boldsymbol{\theta}, \boldsymbol{\xi}]$, since our model implies that $\boldsymbol{\xi}^*$ is conditionally independent of $\mathbf{Y}$, $\boldsymbol{\beta}$ and $\sigma$, given $\boldsymbol{\xi}$. Specifically, it follows from the assumptions that

$$[\boldsymbol{\xi}^*|\boldsymbol{\theta}, \boldsymbol{\xi}] \sim N_m(\Gamma_*^T\Gamma^{-1}\boldsymbol{\xi}, \Gamma_{**} - \Gamma_*^T\Gamma^{-1}\Gamma_*)$$

where $\Gamma = \text{var}(\boldsymbol{\xi})$, $\Gamma_* = \text{cov}(\boldsymbol{\xi}, \boldsymbol{\xi}^*)$ and $\Gamma_{**} = \text{var}(\boldsymbol{\xi}^*)$.

We then cycle over all the above steps as many times as necessary to obtain the required number of realizations from the distribution of $[(\boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\beta}, \sigma)|\mathbf{Y}]$ and of $[\boldsymbol{\xi}^*|\boldsymbol{\theta}, \boldsymbol{\xi}]$. After convergence we sample every $r$th realization of the chain. Increasing the value of $r$ reduces the serial correlation in the resulting output-sample. The last step is then only necessary at every $r$th cycle.

## 3.2 Clipped Intensity Models

This thesis has been motivated in part by a research project concerning the prediction and inference of deoxynivalerol (DON vomitoxin) intensity in wheat scab from truckloads and wheat fields; see Schabenberger, Gregoire and Kong (1999). Unfortunately, the developed models were not successfully applied to the DON assessment due to lack of information about marking distributions. One issue raised in that research is that the measurement of DON in ppm cannot be obtained when the DON exceeds the threshold for the assay techniques. This leads to be the multinomial mapping problem in the geostatistical modeling. This type of spatial data has recently received large attention in terms of "hotspots" and "excursion set" (Gregoire, Schabenberger and Kong, 2000), where a spatial point process is highly clustered with certain attributes. Other examples where this sort of situation occurs are numerous: a geologic formation composed of three rock types, a contaminated geographic region with a subregion determined by the locations where the contaminant concentration surpasses a safety level, or below a limit of detection.

Now we consider a binary random field $\{Z(s); s \in \mathbb{S}\}$, which assumes the values 0 or 1 for every location $s$. Furthermore, we assume that the binary random field $Z(s)$ come from clipping an underlying intensity surface at some given cutoff point $c'$, i.e.,

$$Z(s) = I\{\Lambda(s) \geq c'\}. \tag{3.12}$$

Suppose that we have $n$ observations $\mathbf{Z} = (Z_1, \cdots, Z_n)$ based on (3.12) and a single realization of the random field $\Lambda(s)$ given by (3.1), where $s_1, \cdots, s_n$ are known sampling locations in $\mathbb{S}$. Based on $\mathbf{Z}$ and our prior knowledge on parameters from the random field model, we want to predict the unobserved random vector $\mathbf{Z}^* = (Z(s_{01}), \cdots, Z(s_{0m}))$, where $\mathbf{Z}^*$ comes from the same realization of $\Lambda(s)$ as

the data vector $\mathbf{Z}$, and $s_{01}, \cdots, s_{0m}$ are known locations in $\mathbb{S}$. To apply geostatistical analysis methods, the equivalent log transform on intensity to (3.12) is considered, namely,

$$Z(s) = I\{\eta(s) \geq c\} \tag{3.13}$$

where $\eta(s) = x^T(s)\boldsymbol{\beta} + \xi(s)$ as defined before. Since the binary data contain no information about variance, we assume that $\{\xi(s); s \in \mathbb{S}\}$ is an underlying unobserved random field with the form (3.1), with $E\xi(s) = 0$ and the covariance structure as in (3.4).

Under the normality assumption for $\xi(s)$, the connection between the two random fields (3.12) and (3.13) provides an indirect way to model the family of finite dimensional distributions of a binary random field, enabling one to perform likelihood-based inference. It also allows for an 'exact' sampling-based predictive inference in the resulting binary random field by exploiting the mathematical tractability of random field $\xi(s)$. The likelihood of the parameters $(\boldsymbol{\beta}, \tau, \theta^*, c)$, based on the binary data $\mathbf{z}$, is given by

$$\begin{aligned} L(\boldsymbol{\beta}, \tau, \theta^*, c|\mathbf{z}) &= \int_{A_1} \cdots \int_{A_n} \frac{1}{(2\pi\tau^2)^{n/2}} |\Gamma_\theta^*|^{-1/2} \\ &\quad \times \exp\left(-\frac{1}{2\tau^2}(\boldsymbol{\eta} - X^T\boldsymbol{\beta})^T \Gamma_\theta^{*-1}(\boldsymbol{\eta} - X^T\boldsymbol{\beta})\right) d\boldsymbol{\eta} \\ &= \int_{A_1'} \cdots \int_{A_n'} \frac{1}{(2\pi\tau^2)^{n/2}} |\Gamma_\theta^*|^{-1/2} \exp\left(-\frac{1}{2\tau^2}\boldsymbol{\xi}^T \Gamma_\theta^{*-1}\boldsymbol{\xi}\right) d\boldsymbol{\xi} \end{aligned} \tag{3.14}$$

where $\boldsymbol{\xi} = (\xi(s_1), \cdots, \xi(s_n))'$, $\boldsymbol{\eta} = X^T\boldsymbol{\beta} + \boldsymbol{\xi}$, and

$$A_i = A(z_i) = \begin{cases} (-\infty, c), & \text{if } z_i = 0; \\ [c, \infty), & \text{if } z_i = 1. \end{cases} \qquad \text{for } i = 1, \cdots, n$$

$$A_i' = A'(z_i) = \begin{cases} (-\infty, c - x_i^T\boldsymbol{\beta}), & \text{if } z_i = 0; \\ [c - x_i^T\boldsymbol{\beta}, \infty), & \text{if } z_i = 1. \end{cases} \qquad \text{for } i = 1, \cdots, n$$

It is apparent from (3.14) that a direct Bayesian analysis of this model is intractable, even when using MCMC, because the likelihood of the model parameters is given in terms of high-dimensional integrals, which are difficult to evaluate. Notice from (3.14) that under this full model, the parameters are not identified. If $X_{n\times p}$ with $p = 1$, then for any constant $a_1 > 0$, $a_2 \in \mathbb{R}$, the parameter vector $(a_1\beta - a_2, a_1^2\tau^2, \theta, a_1 c - a_2)$ all have the same likelihood, and in particular the binary data contain no information about $\tau^2$. To avoid this situation, we fix the parameter $\tau^2 = 1$ and $c = 0$ in the rest of this section so that our unknown model parameters are $(\boldsymbol{\theta}, \boldsymbol{\beta})$.

### 3.2.1 Second Order Structure of Binary Random Fields

In the study of the mean and covariance structure of the resulting binary random field it is helpful to understand the global behavior of the binary realizations and their relation with the parameters of the underlying random field $\xi(\cdot)$. Given $(\theta, \boldsymbol{\beta})$ we have

$$E(Z(s)) = \Phi(x^T(s)\boldsymbol{\beta})$$
$$\text{var}(Z(s)) = \Phi(x^T(s)\boldsymbol{\beta})(1 - \Phi(x^T(s)\boldsymbol{\beta}))$$

where $\Phi$ is the cdf of the standard normal distribution, so the mean and variance of the binary random field are controlled only by the parameter $\boldsymbol{\beta}$ as well as the location dependent covariates. Likewise, conditional on $(\theta, \boldsymbol{\beta})$, the correlation function of the

binary random field is given by

$$\gamma_z^*(s,s') = \int_{-x^N(s)\boldsymbol{\beta}}^{\infty}\int_{-x^N(s')\boldsymbol{\beta}}^{\infty} \varphi_2(w_1,w_2;\gamma^*(s,s'))dw_1dw_2$$
$$/(\sqrt{\Phi(x^T(s)\boldsymbol{\beta})(1-\Phi(x^T(s)\boldsymbol{\beta}))}$$
$$\sqrt{\Phi(x^T(s')\boldsymbol{\beta})(1-\Phi(x^T(s')\boldsymbol{\beta}))}\ )$$
$$-\sqrt{\frac{\Phi(x^T(s)\boldsymbol{\beta})\Phi(x^T(s')\boldsymbol{\beta})}{(1-\Phi(x^T(s)\boldsymbol{\beta}))(1-\Phi(x^T(s')\boldsymbol{\beta}))}} \tag{3.15}$$

where $\varphi_2(w_1,w_2;\gamma)$ is the bivariate normal density function with zero mean, unit variance and correlation $\gamma^*$. Note that the correlation structure in (3.15) of the binary random field depends on both $\theta$and $\boldsymbol{\beta}$, as well as on the covariates, even though the correlation structure of the underlying field $\xi(s)$ is independent of $\boldsymbol{\beta}$ and only depends on $\theta$.

Suppose that $\xi(\cdot)$ has a constant mean, say $\beta(s) = \beta$. We assume this is so in the remainder of this subsection. Using the formula given in Kedem (p35, 1980), the correlation function (3.15) reduces to

$$\gamma_z^*(s,s') = \frac{1}{2\pi\Phi(\beta)(1-\Phi(\beta))}\int_0^{\gamma^*(s,s')} \frac{\exp\left(-\beta^2/(1+t)\right)}{(1-t^2)^{1/2}}dt \tag{3.16}$$

This provides a strictly increasing correspondence between $\gamma^*(s,s')$ and $\gamma_z^*(s,s')$. For any fixed value of $\gamma^*(s,s')$, $\gamma_z^*(s,s')$ is an even function of $\beta$, having its maximum at $\beta = 0$, and decreasing to zero in absolute value as $|\beta|$ increases (Ochi and Prentice,

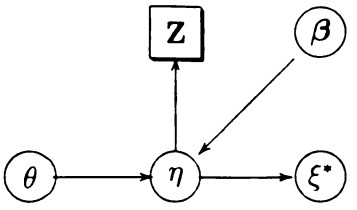


Figure 3.2: DAG of clipped intensity models

1984). For the case $\beta = 0$, the equation (3.16) reduces to $(2/\pi)\sin^{-1}(\gamma^*(s,s'))$, so for any $\beta$ and $\gamma^*(s,s')$ we have $|\gamma_z^*(s,s')| \leq |\gamma^*(s,s')|$. Therefore, the correlation structure induced on the binary field is always weaker than that of the underlying Gaussian field, and extremely so if $|\beta|$ is large, i.e., if $P(Z(s) = 1)$ is close to 0 or 1. In these extreme cases the realizations of the binary random field will be 'almost constant,' either 0 or 1.

### 3.2.2 MCMC Inference

The direct Bayesian analysis of the clipped Gaussian random field is intractable. We again use 'data augmentation' and work with the extended model $[\theta, \boldsymbol{\beta}, \boldsymbol{\eta}, \mathbf{Z}, \boldsymbol{\eta}^*, \mathbf{Z}^*]$ that explicitly includes the unobserved latent vectors $\boldsymbol{\xi}$ and $\boldsymbol{\xi}^*$. In doing so, we take advantage of the mathematical tractability of the underlying Gaussian random field and, by using MCMC methods we are able to: 1) obtain posterior inference of the model parameters, and 2) obtain a posterior sample of the latent vectors $\boldsymbol{\xi}$ and $\boldsymbol{\xi}^*$, which provide the needed probabilistic bridge between $\mathbf{Z}$ and $\mathbf{Z}^*$, allowing us to predict the former from the latter.

The model assumption with its graphic representation in Figure 3.2 is defined

as follows:

$$[\theta, \boldsymbol{\beta}, \boldsymbol{\eta}, \mathbf{Z}, \boldsymbol{\eta}^*] = [\boldsymbol{\theta}][\boldsymbol{\beta}][\boldsymbol{\eta}|\boldsymbol{\theta}, \boldsymbol{\beta}][\mathbf{Z}|\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}][\boldsymbol{\eta}^*|\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}] \tag{3.17}$$

To identify the full conditional distribution for each component of the vector $(\boldsymbol{\xi}, \boldsymbol{\beta}, \boldsymbol{\theta})$, we need to pick up all the unknown factors in (3.17) depending on the component in question. In doing so, and using

$$[\boldsymbol{\eta}|\boldsymbol{\theta}, \boldsymbol{\beta}] = \prod_{i=1}^{n} [\eta_i|\boldsymbol{\eta}_{-i}, \boldsymbol{\theta}, \boldsymbol{\beta}][\boldsymbol{\eta}_{-i}|\boldsymbol{\theta}, \boldsymbol{\beta}]$$

we have for each component of the latent vector $\boldsymbol{\eta}$ or $\boldsymbol{\xi}$ that

$$\begin{aligned}
[\eta_i|\boldsymbol{\eta}_{-i}, \boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{z}] &= [\eta_i|\boldsymbol{\eta}_{-i}, \boldsymbol{\theta}, \boldsymbol{\beta}, z_i] \\
&= N(x_i^T\boldsymbol{\beta} + \boldsymbol{\gamma}_i^T\Gamma_{-i}^{-1}(\boldsymbol{\eta}_{-i} - X_{-i}\boldsymbol{\beta}), 1 - \boldsymbol{\gamma}_i^T\Gamma_{-i}^{-1}\boldsymbol{\gamma}_i) \\
&\quad \times I\{\eta_i \in A_i\} \\
[\xi_i|\boldsymbol{\xi}_{-i}, \boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{z}] &= N(\boldsymbol{\gamma}_i^T\Gamma_{-i}^{-1}\boldsymbol{\xi}_{-i}, 1 - \boldsymbol{\gamma}_i^T\Gamma_{-i}^{-1}\boldsymbol{\gamma}_i) \\
&\quad \times I\{\xi_i \in A_i'\}
\end{aligned}$$

where $\boldsymbol{\eta}_{-i} = \{(\eta_j); j \neq i\}$ and $\boldsymbol{\xi}_{-i} = \{(\xi_j); j \neq i\}$, and the matrix $\Gamma_{-i}$ is $\Gamma$ with the $i$th column and row removed, $\boldsymbol{\gamma}_i^T$ is the $i$th row of $\Gamma$, and the matrix $X_{-i}$i s $X$ with the $i$th row removed, $i = 1, \cdots, n$. The above full conditional distribution is a Normal distribution truncated to $A_i$ or $A'$.

Sampling from the truncated distribution can in principle be done using rejection sampling, but this is inefficient due to the large proportion of rejections. To draw an observation from $N(\mu, \sigma^2)$ distribution, truncated to $(a, b)$ with $a, b \in \mathbb{R}$, the

inversion CDF methods can be applied (Devroye, 1986). If $U \sim Unif(0,1)$, the following statistic $T$, where

$$T := \mu + \sigma\Phi^{-1}\left(\Phi\left(\frac{a-\mu}{\sigma}\right) + U\left(\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)\right)\right),$$

has the desired distribution. The ideas are easily extended to give a general account of Bayesian analysis for order-restrictive parameters (Gelfand et al., 1990a). A mixed rejection algorithm for sampling from univariate normal distribution truncated to intervals is given in Geweke (1991).

Under the present framework, first regarding our prior specification, $\boldsymbol{\beta}$ and $\theta$ are assumed to be independent, with $\boldsymbol{\beta} \sim N_p(\boldsymbol{\beta}_0, B_0^{-1})$, where $\boldsymbol{\beta}_0$ and $B_0$ are chosen to reflect prior information about $P(Z(s) = 1)$. In most cases a vague prior distribution may be obtained by setting $\boldsymbol{\beta}_0 = 0$ and $B_0 = \tau_0 I_p$ with $\tau_0$ small, say 0.05, which assures that the inferences are mainly driven by the likelihood. The prior on $\theta$ would depend on the correlation function $\gamma^*(s, s')$. Suppose that the prior distribution for the model parameters is given by

$$\log[\boldsymbol{\beta}, \theta] = -\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T B_0(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \log[\theta] + c_g.$$

Then, for the full conditional distribution of $\boldsymbol{\beta}$, again from Figure 3.2, we have

$$\begin{aligned}[\boldsymbol{\beta}|\boldsymbol{\theta}, \boldsymbol{\eta}, \mathbf{z}] = [\boldsymbol{\beta}|\boldsymbol{\theta}, \boldsymbol{\eta}] &= [\boldsymbol{\eta}|\boldsymbol{\theta}, \boldsymbol{\beta}][\boldsymbol{\beta}] \\ &\propto [\boldsymbol{\eta} - X^T\boldsymbol{\beta}|\boldsymbol{\theta}, \boldsymbol{\beta}][\boldsymbol{\beta}] = [\boldsymbol{\xi}|\boldsymbol{\theta}, \boldsymbol{\beta}][\boldsymbol{\beta}]\end{aligned}$$

since conditional on $\boldsymbol{\eta}$ the data $\mathbf{z}$ are redundant. Similarly, for the full conditional distribution for $\boldsymbol{\theta}$,

$$[\boldsymbol{\theta}|\boldsymbol{\beta}, \boldsymbol{\eta}, \mathbf{z}] \propto [\boldsymbol{\xi}|\boldsymbol{\theta}, \boldsymbol{\beta}][\boldsymbol{\theta}].$$

The last two full conditional distributions are of non-standard form and are difficult to directly sample from, so one has to perform a Metropolis-Hastings step. For example, given the current value $\theta^t$, a candidate $\theta'$ for the next iteration is generated from a

proposal distribution $q(\theta'|\theta^t)$ and this candidate is accepted as the next iterate with probability

$$\Delta(\theta^t, \theta') = \min\left\{\frac{[\theta'|\boldsymbol{\beta}, \xi]q(\theta^t|\theta')}{[\theta^t|\boldsymbol{\beta}, \xi]q(\theta'|\theta^t)}, 1\right\}$$

If the candidate $\theta'$ is accepted, the next iterate becomes $\theta^{t+1} = \theta'$; otherwise it is set to the current state $\theta^t$. The choice of proposal distribution depends on the choice of correlation function $\gamma_\theta(s, s')$ and so is problem specific. The MH algorithm also works for updating $\boldsymbol{\beta}$.

Starting with initial value of $(\boldsymbol{\eta}^0, \boldsymbol{\theta}^0, \boldsymbol{\beta}^0)$, one cycle (iteration) of this MCMC algorithm consists of successively (and in that order) sampling from the full conditional distributions (4.9), and then performing the MH step described above. The generated sequence, of length $m$ say, is a trajectory of a Markov chain having $[\boldsymbol{\eta}, \boldsymbol{\theta}, \boldsymbol{\beta}|\mathbf{z}]$ as its limiting distribution. Therefore, after discarding an initial burn-in of the chain, say $r$ iterations, long enough to surpass the transient stage of the chain, the subsequent iterations are a dependent sample $\{(\boldsymbol{\eta}^t, \boldsymbol{\theta}^t, \boldsymbol{\beta}^t); t = r+1, \cdots, m\}$ approximately distributed as $[\boldsymbol{\eta}, \boldsymbol{\theta}, \boldsymbol{\beta}|\mathbf{z}]$. This sample can be used to make inferences about the model

parameters and the latent vector $\boldsymbol{\eta}$. The MCMC algorithm is used to obtain approximate samples from the posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\beta})$, which is given by

$$
\begin{aligned}
[\boldsymbol{\eta}, \boldsymbol{\theta}, \boldsymbol{\beta} | \mathbf{Z}] &= [\boldsymbol{\theta}][\boldsymbol{\beta}][\boldsymbol{\eta} | \boldsymbol{\theta}, \boldsymbol{\beta}][\mathbf{Z} | \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}] \\
&\propto [\boldsymbol{\theta}][\boldsymbol{\beta}] |\Gamma|^{-1/2} \\
&\quad \times \exp\left( -\frac{1}{2} (\boldsymbol{\eta} - X^T \boldsymbol{\beta})^T \Gamma^{-1} (\boldsymbol{\eta} - X^T \boldsymbol{\beta}) \right) \\
&\quad \times \prod_{i=1}^{n} I\{\eta_i \in A_i\} \\
[\boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\beta} | \mathbf{Z}] &\propto [\boldsymbol{\theta}][\boldsymbol{\beta}] |\Gamma|^{-1/2} \exp\left( -\frac{1}{2} \boldsymbol{\xi}^T \Gamma^{-1} \boldsymbol{\xi} \right) \prod_{i=1}^{n} I\{\xi_i \in A_i'\}
\end{aligned}
\tag{3.18}
$$

## 3.3 Joint Inference of Point Patterns and Geostatistical Random Fields

In this section we consider the point patterns that associate with the covariate stochastic processes. When the data sets of various types are closely related to the same topic of interest, the joint modeling of those data sets may lead to more efficient analysis. This is possible, because we assume all the data sets are associated with or depend on a spatial process (a random field $\xi(s)$). Then we can make prediction and inference based on this random field. The joint modeling makes full use of the data, such as the location information in the point pattern data. Point pattern data associates with the random field $\xi(s)$ through the intensity function, and the

categorical data associates with the random field through clipping. Consequently, we can perform the joint analysis of point patterns and other types of spatial data. We can also measure their association and dependence.

Assume that an observed point pattern $\{S_j\}$ of $N(ds)$ with intensity function $\Lambda(s)$, and observations $\{Y_i\}$ of a covariate stochastic process $\xi(s)$ are both available over the same region $\mathbb{S}$. The covariate random field is modeled as in (3.3), and the point process may depend on the covariate process $\xi$ through a functional link defined to be

$$\log \Lambda(s) = \zeta + \alpha\eta(s) \tag{3.19}$$

where $\eta(s)$ is defined in (3.1) as $\eta(s) = \beta(s) + \xi(s)$.

### 3.3.1 Association between Point Patterns and Covariate Processes

Assume that the local intensity of a spatial inhomogeneous Poisson process $N(ds)$ depends on the realization of another spatial stochastic process $\xi(s)$. The extent of the dependence of the point process on the covariate process $\xi(s)$ is determined by the unknown regression coefficient $\alpha$ in (3.19). We examine the dependence of the intensity $\Lambda(s)$ on the level of the covariate process $\xi(s)$ using a model of the form (3.19).

The data consist of a partial realization of a point process, $\{S_j, J\}$, and measurements $\{Y_i\ ; i = 1, \cdots, n\}$ of a spatial process, $\xi(s)$ on a region $\mathbb{S}$. In the case where $\xi(s)$ has no effect on the intensity of the point process, $\alpha = 0$. Conditional on $\eta(s)$ and on the observed point pattern, $\{S_j, J\}$, the locations of the points $S_j$ from a subset $\mathbb{S}$ of the study region are independently and identically distributed

with probability density function proportional to $\Lambda(s)$. Given $(\beta, \xi)$, the conditional log-likelihood equation for $(\alpha, \zeta)$ in (3.19) is

$$l(\alpha, \zeta|\eta) = -\int_A \exp(\zeta + \alpha\eta(s))ds + \sum_j (\zeta + \alpha\eta_j) \quad (3.20)$$

where $\eta_j = \eta(S_j)$. Inference for $\alpha$ and $\zeta$ can be based on Cox (1972), provided that the true value of the $\xi(s)$ throughout $A$ is known. When $\eta$ is estimated, the conditional likelihood estimates of $\hat{\alpha}$ and its variance derived from Cox (1972) will be invalid, as they do not allow for prediction or measurement error. However, we may estimate $(\beta, \xi)$ by Kriging estimates or other generic geostatistical methods described in the previous sections.

Currie (1998) develops a test for association between realizations of a point pattern and a covariate process $\xi(s)$ on the study region. The test is also valid with the presence of covariates $\beta(s)$. The test is motivated by the score test. Measurements $Y_i$ of this process are assumed to follow the model given in Equation (3.3). The intensity of the point process depend on $\xi$ and $\beta(s)$ jointly according to (3.19). Under this model, the test for association is a test of the null hypothesis: $H_0 : \alpha = 0$. Up to an additive constant, the conditional log-likelihood for the intensity $\Lambda(s)$ of the point process given $\eta(s)$ is given in (3.20). The parameter $\zeta$ is a nuisance parameter, so we maximize over it. Differentiating the expression with respect to $\zeta$, and then setting the expression equal to zero gives the expression for $\zeta$

$$e^{\zeta} = \frac{J}{\int_A \exp(\alpha\eta(s))ds}.$$

Differentiating this equation with respect to $\alpha$, and substituting for $\zeta$, the derivative of the log-likelihood with respect to $\alpha$ is

$$\frac{\partial l}{\partial \alpha} = \sum_j \eta_j - \frac{J\int_A \eta(s)\exp(\alpha\eta(s))ds}{\int_A \exp(\alpha\eta(s))ds}. \quad (3.21)$$

For testing the null hypothesis, $\alpha = 0$, consider the statistic

$$U_0 = \sum_j \eta_j - \frac{J}{|A|}\int_A \eta(s)ds$$

$$= \sum_j (x_j^T\boldsymbol{\beta} + \xi_j) - \frac{J}{|A|}\int_A (x^T(s)\boldsymbol{\beta} + \xi(s))ds$$

This is the score statistic for testing the association between $\Lambda(s)$ and $\xi(s)$, the derivative of the log-likelihood evaluated at the null hypothesis. If we had exact ascertainment of $\xi(s)$ throughout $\mathbb{S}$, then we could calculate the mean and variance of the distribution of this test statistic under the null hypothesis of $\alpha = 0$, and hence carry out the test of the null hypothesis. Note that calculating $U_0$ only involves the evaluation of (3.21) at $\alpha = 0$. Thus we avoid the need for calculation of stochastic integrals any more complex than $\int_A \eta(s)ds$.

In practice we do not observe $\eta(s)$ directly but only through observation $Y_i$ on $\mathbb{S}$. From these measurements, we obtain the conditional expectation of $\eta(s)$ throughout $\mathbb{S}$. The $Y_i$'s are conditionally independent given $\eta$, according to the model (3.3). We may use an estimator (e.g., Kriging, see Diggle et al., 1998) for $\eta$. Substituting $\hat{\eta}$ for $\eta$ in the expression, we obtain the approximate test statistic

$$\hat{U}_0 = \sigma^{-2}\sum_{j=1}^{J}\left(X^T\boldsymbol{\beta} + \boldsymbol{\gamma}_j^T\Sigma^{-1}(\mathbf{Y} - X^T\boldsymbol{\beta})\right)$$

$$- \frac{J}{\sigma^2|A|}\int_A \left(X^T\boldsymbol{\beta} + \boldsymbol{\gamma}_s^T\Sigma^{-1}(\mathbf{Y} - X^T\boldsymbol{\beta})\right)ds$$

where $\boldsymbol{\gamma}_j = \boldsymbol{\gamma}_{s_B} = \{\text{COV}(\xi_j, Y_i)\}_{i=1}^n = \{\text{COV}(\xi_j, \xi_i)\}_{i=1}^n$ is the column vector of length $n$. Under the assumption that $\xi(s)$ is Gaussian process, the distribution of $\hat{U}_0$ is Gaussian. To find the expected value of this statistic under the null

hypothesis, we take the expectation of $\hat{U}_0$w ith respect to $\mathbf{Y}$, which is zero, that is

$$\begin{aligned}\sigma^2 E_{\mathbf{Y}}(\hat{U}_0) &= \sum_{j=1}^{J} E_{\mathbf{Y}}\left(X^T\boldsymbol{\beta} + \boldsymbol{\gamma}_j^T\Sigma^{-1}(\mathbf{Y} - X^T\boldsymbol{\beta})\right) \\ &\quad - \frac{J}{|A|}\int_A E_{\mathbf{Y}}\left(X^T\boldsymbol{\beta} + \boldsymbol{\gamma}_s^T\Sigma^{-1}(\mathbf{Y} - X^T\boldsymbol{\beta})\right) ds \\ &= 0\end{aligned}$$

Thus, the approximate score statistic $\hat{U}_0$ has expectation zero under the null hypothesis $\alpha = 0$. The variance of $\hat{U}_0$ with respect to the measurements $\mathbf{Y}$ is given by

$$\begin{aligned}\sigma^2 \operatorname{var}_{\mathbf{Y}}(\hat{U}_0) &= \sum_j \boldsymbol{\gamma}_j^T\Sigma^{-1}\boldsymbol{\gamma}_j + \sum_{j=1}^{J}\sum_{k\neq j}\boldsymbol{\gamma}_j^T\Sigma^{-1}\boldsymbol{\gamma}_k \\ &\quad - \frac{J}{|A|}\sum_j \int_A \boldsymbol{\gamma}_j^T\Sigma^{-1}\boldsymbol{\gamma}_s ds \\ &\quad + \frac{J^2}{|A|^2}\int_A\int_A \boldsymbol{\gamma}_s^T\Sigma^{-1}\boldsymbol{\gamma}_{s'} ds ds'.\end{aligned}$$

Thus, given locations of events of a point processes, $\{S_j\}$ on $A$, the measurements $\{Y_i\}$ of the underlying continuous process $\{\eta(s)\}$ are used to calculate $\hat{U}_0$. Then conditioning on the locations of the process points, we use expressions for the mean and variance of this statistics under the null hypothesis, to obtain a $p$-value for the data.

Currie (1998) examines the performance of this test on a selection of simulated data sets for a simple model. The test performs well, in the sense that it is able to identify association between the point process intensity and the covariate process $\xi(s)$

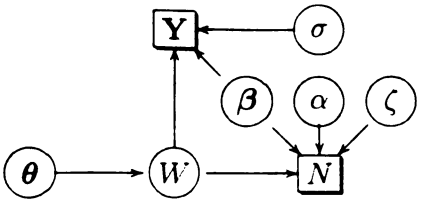


Figure 3.3: DAG of point patterns and geostatistical models

when the association is not evident to the eye. Currie also investigates the effect of estimating the covariance parameters of the covariate process, rather than using their true values. The replacement of the true covariance parameters by their maximum likelihood estimates has only a small effect on the null distribution of the approximate score statistic and on the power of the test.

### 3.3.2 MCMC Inference on Joint Models

Consider when both a point pattern $\{S_j\}$ and a geostatistical process $Y(s)$ depend on the realization of $\xi(s)$ and on the parameters $\sigma$, $\boldsymbol{\beta}$, and $\boldsymbol{\theta}$. The dependence structure for the quantities in (3.19) can be referred to its graphic representation in Figure 3.3. The graphical model also aids the factorization of the joint distribution of all the quantities in our graph into the product of the full conditionals. The analysis for factorization is to take the product over all nodes of the conditional density of that node given its parents. With the given DAG, the joint distribution of our model quantities can be factored as follows:

$$[\boldsymbol{\theta}, W, \boldsymbol{\beta}, \alpha, \zeta, N, \sigma, Y] = [\boldsymbol{\theta}][W|\boldsymbol{\theta}][\boldsymbol{\beta}][\sigma][Y|\boldsymbol{\theta}, W, \boldsymbol{\beta}, \sigma][\alpha, \zeta][N|\boldsymbol{\theta}, W, \boldsymbol{\beta}, \alpha, \zeta] \quad (3.22)$$

Model specification is involved in the distributional forms for each of the factors in (3.22). The covariate process $\xi(s)$ is most commonly modeled as a stationary

Gaussian random field with a partially uncertain covariance structure. Some notable references in this area include Besag (1974), Ripley (1981), Cressie and Chan (1989), Clayton and Kaldor (1987), Isaaks and Srivastava (1990), Besag, York and Mollié (1991), Cressie (1993), Bernardinelli et al. (1997), and Møller, Syverseveen and Waagepetersen (1998). Ideally, the distributional assumptions about the underlying properties of our model should be sufficiently flexible to accommodate a wide range of behavior of the random variates they describe. In this study, we are focusing on the model in (3.2) with the inhomogeneous intensity. Modeling on the combined spatial data with the kernel mixture type is best implemented through tessellation of the reference state $\mathbb{U}$. Consider a tessellation of $J$ cells in $\mathbb{U}$ with cell centers $\{u_j\}$. The $W(du; u_j \in du)$ are assumed to be independent normal random variates with mean zero and variance depending on some parameters. Approximately, the Gaussian random field $\xi(s)$ is constructed through the kernel mixture of Wiener process $W(du)$ on $\mathbb{U}$, given by,

$$\boldsymbol{\xi} = K^{\boldsymbol{\theta}}\mathbf{W}^{\boldsymbol{\theta}} \tag{3.23}$$

where $\boldsymbol{\xi} = (\xi(s_1), \cdots, \xi(s_n))$ with $s_1, \cdots, s_n$i n $\mathbb{S}$, and the elements of the random vector $\mathbf{W} = \{W_j\}_{j \in J}$ are independent normally distributed with the variance subject to smooth variation in parameters $\theta$ and cell attributes $\lambda^{\theta}(du)$. Note that $K = \{k_{ij}^{\boldsymbol{\theta}}\}$ with $k_{ij}^{\theta} = k^{\theta}(s_i, u_j)$, and the covariance matrix of $\boldsymbol{\xi}$ is then $\Gamma = KK^T$. Thus, the density of $\boldsymbol{\xi}$ is given by

$$p_{\boldsymbol{\xi}}(\mathbf{z}) = (2\pi)^{-n/2}|\Gamma|^{-1/2}\exp\left(-\frac{1}{2}\mathbf{z}^T\Gamma^{-1}\mathbf{z}\right).$$

The density of $\eta(s) = x^T(s)\boldsymbol{\beta} + \xi(s)$ is then

$$p_{\eta}(\mathbf{z}) = (2\pi)^{-n/2}|\Gamma|^{-1/2}\exp\left(-\frac{1}{2}(\mathbf{z} - X^T\boldsymbol{\beta})^T\Gamma^{-1}(\mathbf{z} - X^T\boldsymbol{\beta})\right).$$

Other model distributional assumptions are given by

$$[N(ds)|\xi] \sim \text{Poi}(\Lambda(ds))$$

$$[Y(s)|\xi] \sim N(\eta(s), \sigma^2).$$

When applying the MCMC method to simulate from the joint posterior density function given in (3.22), we may employ independent vague priors on the model parameters in the absence of genuine prior information, for example,

$$[\boldsymbol{\theta}]\sim\ \pi(\theta),\ [\boldsymbol{\beta}]\sim\ N_p(\boldsymbol{\beta}_0, B),\ [\alpha]\sim\ N(\alpha_0, \sigma_\alpha^2),\ [\zeta]\sim\ N(\zeta_0, \sigma_\zeta^2),\ [\sigma]\sim\ \text{gam}(a_\sigma, b_\sigma)$$

Other steps for generating the chain are to update the other unknown quantities conditional on the data, $\{S_j\}$ and $\mathbf{Y}$, and visiting each of the other nodes in Figure 3.3 in turn. The full conditionals of the quantities $\theta$, $\boldsymbol{\beta}$, $\alpha$, $\zeta$, $\sigma$ and $W$ are derived as follows

The log of the conditional density for $\boldsymbol{\theta}$ is derived based on the Bayesian relation, $[\boldsymbol{\theta}|W, \boldsymbol{\beta}, \alpha, \zeta, N, \sigma, \mathbf{Y}] = [\boldsymbol{\theta}|W] \propto\ [W|\boldsymbol{\theta}][\boldsymbol{\theta}]$ then, up to an additive constant, we have

$$\log[\theta_i|\boldsymbol{\theta}_{-i}, \boldsymbol{\xi}] = -\log|\Gamma_{\boldsymbol{\theta}}| - \frac{1}{2}\boldsymbol{\xi}^T\Gamma_{\boldsymbol{\theta}}^{-1}\boldsymbol{\xi} + c_g$$

where $\boldsymbol{\xi}$ is the vector of the $\{\boldsymbol{\xi}_i\}_{i\in I}$. We may employ a symmetric random walk proposal on the feasible values of $\theta_i$'s, so the update steps for each of the $\theta_i$ reduce to MH steps.

From the relation, $[\boldsymbol{\beta}|\boldsymbol{\theta}, W, \alpha, \zeta, N, \sigma, \mathbf{Y}] \propto [\mathbf{Y}|\boldsymbol{\theta}, W, \sigma][N|\boldsymbol{\theta}, W, \alpha, \zeta][\boldsymbol{\beta}]$, the log of the conditional density of $\boldsymbol{\beta}$ is, up to an additive constant, given by

$$\begin{aligned}\log[\boldsymbol{\beta}|\boldsymbol{\xi}, N] = &-\frac{1}{2\sigma^2}\sum_i(Y_i - \eta_i)^2 \\ &- \int \exp(\zeta + \alpha\eta(s))ds \\ &+ \alpha\sum_j \eta_j + \log[\boldsymbol{\beta}] + c_g\end{aligned}$$

where $\eta_i = \eta(s_i)$ and $\eta_j = \eta(S_j)$.

For the full conditional density of $\zeta$, from the relation $[\zeta|\boldsymbol{\theta}, W, \boldsymbol{\beta}, \alpha, N, \sigma, \mathbf{Y}] \propto [N|\boldsymbol{\theta}, W, \alpha, \zeta][\zeta]$, we have

$$\log[\zeta|\boldsymbol{\theta}, W, \boldsymbol{\beta}, \alpha, N, \sigma, \mathbf{Y}] = -\int (\exp(\zeta + \alpha\eta(s))ds + J\zeta + \log[\zeta] + c_g.$$

The full conditional density of $\alpha$, follows $[\alpha|\boldsymbol{\theta}, W, \boldsymbol{\beta}, \zeta, N, \sigma, \mathbf{Y}] \propto [N|\boldsymbol{\theta}, W, \alpha, \zeta][\alpha]$ and is given by

$$\begin{aligned}
&[\alpha|\boldsymbol{\theta}, W, \boldsymbol{\beta}, \zeta, N, \sigma, \mathbf{Y}] \\
&\quad = -\int (\exp(\zeta + \alpha\eta(s))ds + \alpha \sum_j \eta_j + \log[\alpha].
\end{aligned}$$

With the model, $[\sigma|\boldsymbol{\theta}, W, \boldsymbol{\beta}, \alpha, \zeta, N, \sigma, \mathbf{Y}] \propto [\mathbf{Y}|\boldsymbol{\theta}, W, \boldsymbol{\beta}, \sigma][\sigma]$. Up to an additive constant, the log density of the conditional density of $\sigma$ is

$$\log[\sigma|\boldsymbol{\xi}, \mathbf{Y}] = n \log \sigma - \frac{1}{2\sigma^2} \sum_i (Y_i - \eta_i)^2 + \log[\sigma] + c_g.$$

Based on $[\boldsymbol{\xi}|\boldsymbol{\theta}, W, \boldsymbol{\beta}, \alpha, \zeta, N, \sigma, \mathbf{Y}, \sigma] \propto [\mathbf{Y}|\boldsymbol{\theta}, W, \boldsymbol{\beta}, \sigma][N|\boldsymbol{\theta}, W, \alpha, \zeta][\boldsymbol{\xi}|\boldsymbol{\theta}]$, the log density of covariate random vector $\boldsymbol{\xi}$ is given by

$$\begin{aligned}
\log[\boldsymbol{\xi}|\boldsymbol{\theta}, N, \boldsymbol{\beta}, \mathbf{Y}] = &-\frac{1}{2\sigma^2} \sum_i (Y_i - \eta_i)^2 \\
&- \int \exp(\zeta + \alpha\eta(s))ds + \alpha \sum_j \eta_j \\
&- \frac{1}{2} \boldsymbol{\eta}^T \Gamma_\theta^{-1} \boldsymbol{\eta} + c_g.
\end{aligned}$$

From equation (3.23), $K$ is constant in $\mathbf{W}$ and $\boldsymbol{\xi}$, the log of the full conditional density of $\mathbf{W}$ is given by the following, up to an additive constant,

$$\begin{aligned}\log[\mathbf{W}|\boldsymbol{\theta}, N, \boldsymbol{\beta}, \mathbf{Y}] = &-\frac{1}{2\sigma^2}\sum_i (Y_i - \eta_i)^2 \\ &- \exp(\zeta)\sum_i ds_i \exp(\alpha\eta_i) \\ &+ \alpha\sum_i n_i\eta_i - \frac{1}{2}\mathbf{w}^T(KK^T)^{-1}\mathbf{w} + c_g.\end{aligned}$$

Then we can use an independent random walk proposal for the candidate $\mathbf{W}$ with centered normal density on $W_j$. As such, the update step reduces to a MH step in the MCMC algorithm.

# Chapter 4. Summary of Results and Further Investigation

## 4.1 Summary of Results

The thesis concerns the Cox $M$ modeling approach to inhomogeneous point patterns with the random intensity functions directed by independent increment random measures in a reference state space. The idea of kernel mixture representations is relatively new in spatial statistics. It has the potential to model non-stationary processes, such as the Poisson/Gamma mixture models by Wolpert and Ickstadt (1998a) and Binary/Gaussian models as in Section 3.2.

In Chapter 2, the proposed model on point patterns is based on the Poisson generalized Gamma random fields. The generalized Gamma measures can be simulated by the Inverse Lévy Measure algorithm. The tessellation representation of a point pattern creates a lattice structure, which is the vehicle to tackle the problem of spatial prediction. This is worked out in detail for an example where the data are counts from a subset of lattice cells and binary variables on the complement. The implementation is performed with MCMC for this type of model on counts and binary clipped random fields.

For highly clustered point patterns, it is more appropriate to apply the geostatistical modeling approaches after the log transformation of count data. The related models are based on Wiener measure subordinated random fields, which are closely related to the ordinary random fields specified directly by mean and covariance functions. The proposed models in Chapter 3 are intended to make full use of collateral

information obtained along with point patterns. The collateral data, treated as fixed covariates and covariate stochastic processes, are considered in the setting of joint modeling on the intensities. The corresponding MCMC schemes are proposed for each data model.

The probability structure in the binary clipped models (in Section 3.2) that contain latent Gaussian processes may be of great interest to spatial statisticians, but very little progress has been made in the past in deriving likelihoods and in studying the covariance structures in such processes. One may think of a myriad of examples where the two types of processes are linked. For example, a smooth fertility trend in soil will affect both the emergence of weed plants (a point process) as well as the crop yield surface (a geostatistical process).

## 4.2 Further Investigation

Wolpert and Ickstadt (1998a) propose a data augmentation scheme for the Cox $M$ models when $M$ is a Gamma measure. The data augmentation method does not work with nonzero index parameter function $a(u)$. For the discrete case in Chapter 2, a product multinomial augmentation scheme works in general for the spatial Poisson-generalized Gamma random fields. A new data augmentation scheme may be needed in order to implement a MCMC inference on point patterns. The method is computationally intensive and time consuming in practice. However, it is demanded in that it utilizes the point patterns.

A point pattern may be obtained by thinning of another one, or by the superposing of others, or by the clipping of geostatistical random fields. The limiting point patterns may result in geostatistical random fields. Weak convergence of random

fields has been studied largely using mixing conditions, see Yoshihara (1992). The explicit conditions for $GG$-measures to be approximated by geostatistical random fields are still under investigation, especially on the index set where $a(u) \neq 0$ on $\mathbb{U}$. The classic asymptotic normality theory for a $GG$ distribution may suggest the conditions for weak convergence of point patterns. For a $GG(a, \lambda, b)$ random variable $M$, the standardized variable is asymptotically normal, i.e.,

$$W = \frac{M - \lambda b^{a-1}}{\sqrt{\lambda(1-a)b^{a-2}}} \Rightarrow N(0,1) \quad \text{as } \frac{1-a}{\lambda b^a} \downarrow 0$$

where the convergence is along the decreasing sequences of the squared coefficient of variation. The proof comes from taking the limit of the Laplace transforms; see also Jørgensen (1987). Note that when $a \to -\infty$ the distribution of $bM/(1-a)$ converges to a Poisson distribution with mean $\lambda b^a/(1-a)$. This follows by applying a limit argument to the Laplace transform; see also Aalen (1992). Then the property of asymptotic normality of a $GG(a, \lambda, b)$ may suggest that, for a large inverse scale $b(u)$ when $a(u) > 0$ or small $b(u)$ when $a(u) < 0$, the limiting random field may tend to be a Gaussian random field. The re-weighted $GG$-measures with $a(u) \neq 0$ can not be constructed under the current augmentation method. In application, this occurs mostly for highly clustered point patterns; it might be more appropriate for applying geostatistical methods as discussed in Chapter 3.

When testing of association of a point pattern and geostatistical random fields in Section 3.3 the hypothesis being tested is $H_0 : \alpha = 0$. In case you reject, we plan to study whether the direction of the association can be gleaned from the test statistic. It would be interesting to know whether the latent process inhibits one process while exacerbating the other. One might imagine that through a site-specific application (spatially varied) of herbicide one can reduce the number of weeds (suppress the point pattern) and increase the crop yield (accelerate the yield process).

# Bibliography


AALEN, O. (1992). Modeling heterogeneity in survival analysis be the compound poisson distribution. *The Annals of Applied Probability* **2** 951–972.

ABRAMOWITZ, M. and STEGUN, I. A. (1972). *Handbook of Mathematical Functions*. Dover, New York.

BAR-LEV, S. and ENIS, P. (1986). Reproducibility and natural exponential families with power variance functions. *Annals of Statistics* **14** 1507–1522.

BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice system (with discussion). *J. R. Statist. Soc. B* **35** 192–236.

BESAG, J. YORK, J. and MOLLIÉ, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion). *Ann. Inst. Statist. Math.* **43** 1–59.

BRIX, A. (1998). Spatial point processes in weed science. Technical report Department of Mathematics and Statistics Lancaster University, Lancaster, United Kingdom.

BRIX, A. (1999). Generalized gamma measures and shot-noise cox processes. *Advances in Applied Probability* **31**.

CLAYTON, D. G. and KALDOR, J. (1987). Empirical bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* **43** 671–81.

COX, D. (1972). The statistical analysis of dependencies in point processes. In Lewis, P., editor, *Stochastic point processes: statistical analysis, theory, and applications*. Wiley.

CRESSIE, N. A. C. (1993). *Statistics for Spatial Data*. John Wiley & Sons.

CRESSIE, N. A. C. and CHAN, N. H. (1989). Spatial modeling of regional variables. *J. Amer. Statist. Assoc.* **84** 393–401.

CURRIE, J. E. (1998). *Joint Modelling of Point Process and Geostatistical Measurement Data*. Ph.D. thesis Lancaster University.

DALEY, D. J. and VERE-JONES, D. (1988). *Introduction to the Theory of Point Processes*. Springer, New York.

DEVROYE, L. (1986). *Non-uniform random variate generation*. Springer Verlag: New York.

DIGGLE, P. J., MOYEED, R. A. and TAWN, J. A. (1998). Model-based geostatistics. *Applied Statistics* **47(3)** 299–350.

DIGGLE, P. J. (1990). A point process modelling approach to raised incidence of a rare phenomenon in the vincity of a prespecified point. *Journal of the Royal Statistical Society (Series A)* **153** 349–362.

GELFAND, A. E. and SMITH, A. F. M. (1990a). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85** 398–409.

GELFAND, A. E., HILLS, S. E., RACINE-POON, A. and SMITH, A. F. M. (1990b). Illustration of bayesian inference in normal data models using gibbs sampling. *Journal of the American Statistical Association* **85** 972–985.

GELMAN, A., ROBERTS, G. O. and GILKS, W. R. (1996). Efficient metropolis jumping rules. In J. M. Bernardo, J. O. Berger, A. P. D. and Smith, A. F. M., editors, *Bayesian Statistics* 5. Oxford University Press 599–607.

GEWEKE, J. (1991). Efficient simulation from the multivariate normal and student-$t$ distributions subject to linear constraints. In *Computing Science and Statistics: Proceedings of the 23rd Symposium on the interface*. Ammerican Statistical Association Alexandria, VA 571–578.

GILKS, W. R., RICHARDSON, S. and SPIEGELHALTER, D. J. (1996). *Markov Chain Monte Carlo in Practice*. New York: Chapman and Hall.

GILKS, W. and ROBERTS, G. (1996). *Markov chain Monte Carlo in practice* chapter Strategies for improving MCMC. Chapman & Hall 89–110.

GREEN, P. and SIBSON, R. (1978). Computing dirichlet tesselations in the plane. *Computer Journal* **21** 168–173.

GREGOIRE, T. G., SCHABENBERGER, O. and KONG, F. Z. (2000). Prediction from an integrated regression equation: a forestry application. *Biometrics* **56** 414–9.

HEISEL, T., ANDREASEN, C. and ERSBØLL, A. K. (1996). Annual weed distributions can be mapped with kriging. *Weed Research* **36** 325–337.

HIGDON, D., SWALL, J. and KERN, J. (1999). Non-stationary spatial modeling. In Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M., editors, *Bayesian Statistics: 6* 6. Oxford University Press 761–768.

HOUGAARD, P., LEE, M.-L. and WHITMORE, G. (1997). Analysis of overdispersed count data by mixtures of poisson variables and poisson processes. *Biometrics* **53** 1225–1238.

HOUGAARD, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika* **73** 387–396.

ICKSTADT, K. and WOLPERT, R. L. (1996). Spatial correlation or spatial variation? a comparison of gamma/poisson hierarchical models. Technical Report 01 Institute of Statistics and Decision Sciences, Duke University, Durham NC 27708-0251.

JACOD, J. and SHIRYAEV, A. N. (1987). *Limit Theorems for Stochastic Processes.* volume 288 of *Grundlehren der mathematischen Wissenschaften*. Berlin: Springer-Verlag.

JØRGENSEN, B. (1987). Exponential dispersion models. *Journal of the Royal Statistical Society, Series B* **49** 127–162.

KARR, A. F. (1991). *Point Processes and Their Statistical Inference*. Markcel Dekker second edition.

KASS, R. E., CARLIN, B. P., GELMAN, A. and NEAL, R. M. (1997). Mcmc in practice: A rountable discussion. Technical report Carnegie Mellon University.

KEDEM, B. (1980). *Binary Times Series*. New York: Marcel Dekker.

KINGMAN, J. F. C. (1993). *Poisson Processes*. Oxford Science Publications.

MARCUS, M. B. and ROSIŃSKI (2000). Continuity of stochastic integrals with respect to infinitely divisible random measures. Department of Mathematics, The City College of CUNY, New York, NY 10031, mbmarcus@earthlink.net.

MARTINEZ, V. D. O. (1997). *Prediction in Some Classes of Non-Gaussian Random Fields*. Ph.D. thesis University of Maryland at College Park.

MILES, R. (1972). The random division of space. *Suppl. Adv. Appl. Prob.* 243–266.

MØLLER, J. SYVERSVEEN, A. R. and WAAGEPETERSEN, R. P. (1998). Log gaussian cox process. *Scand. J. Statist.* **25**.

MØLLER, J. (1994). *Lectures on Random Voronoi Tessellations*. Springer.

OKABE, A., BOOTS, B., SUGIHARA, K. and CHIU, S. (2000). *Spatial Tessellations*. Wiley, Chichester.

RIPLEY, B. D. (1981). *Spatial Statistics*. John Wiley and Sons, New York.

ROGERS, A. (1974). *Statistical analysis of spatial dispersion*. Pion, London.

SCHABENBERGER, O., GREGOIRE, T. G. and KONG, F. Z. (1999). Statistical features of wheat scab sampling protocols. random volumes and inhomogeneous point processes. Spring meetings of the Eastern North American Region of the international biometric society.

SILVERMAN, B. W. (1986). *Density estimation for statistics and data analysis*. Chapman and Hall: London.

SMITH, A. F. M. and ROBERTS, G. O. (1993). Bayesian computation via the gibbs sampler and related markov chain monte carlo methods. *J. R. Statist. Soc. B*, **55,** 3–23.

STOYAN, D., K. W. and MECKE, J. (1995). *Stochastic Geometry and its Applications*. Wiley, Chichester second edition.

TANNER, M. A. and WONG, W. H. (1987). The calculation of posterior distribution by data augmentation. *J. Am. Statist. Assoc.* **82** 528–50.

TIERNEY, L. (1994). Markov chains for exploring posterior distributions (with discussion and a rejoinder by the author). *Ann. Statist.* **22** 1701–1762.

TWEEDIE, M. (1984). An index which distinguishes between some important exponential families. In Ghosh, J. and Roy, J., editors, *Statistics: Applications and new directions*. Proceedings of the Indian Statistical Institute golden jubilee international conference 579–604.

WOLPERT, R. and ICKSTADT, K. (1998a). Poisson/gamma random field models for spatial statistics. *Biometrika* **85** 251–267.

WOLPERT, R. L. and ICKSTADT, K. (1998b). *Simulation of Lévy Random Fields. Practical Nonparametric and Semiparametric Bayesian Statistics*. New York: Springer-Verlag dipak dey, peter müller and debajyoti sinha (editors) edition.

YOSHIHARA, K. (1992). *Weakly Dependent Stochastic Sequences and Their Applications* volume VOL. VII Generalized partial-sum processes. Sanseido Co., Ltd. Misaki-cho.