LIKELIHOODS AND PSEUDOLIKELIHOODS FOR MARKOV SPATIAL PROCESSES

A.J. BADDELEY

University of Western Australia

We study spatial random processes (mainly point processes in $\mathbb{R}^d$) which are defined to satisfy various spatial analogues of the Markov conditional independence property. We explore some issues in statistical inference for such models, including likelihood and pseudolikelihood methods, and identifiability.

AMS subject classifications: 60D05, 60G55, 62M30.
Keywords and phrases: Area-interaction process, Berman-Turner device, Conditional intensity, Directed Markov point processes, Efficiently estimable parameters, Exponential families, Gibbs processes, Identifiability, Markov point processes, Markov Chain Monte Carlo, Multiparameter counting processes, Ord's process, Pairwise interaction, Spatial clustering, Strauss process, Total variation distance, Widom-Rowlinson model.

1 Introduction

Markov point processes [75, 76] are a rich class of stochastic models for spatial patterns, with the virtue of being relatively tractable. They are defined to satisfy one of several spatial counterparts of the Markov conditional independence property. The likelihood takes a simple explicit form, apart from a difficult normalising factor. Indeed typically the likelihood is an exponential family, and the canonical sufficient statistic is often closely related to nonparametric spatial statistics. Typically each process is the equilibrium measure of an associated space-time Markov process; thus it is amenable to Markov Chain Monte Carlo simulation and bootstrap inference. Accordingly there is much current interest in exploring the potential applications of Markov point processes, which include spatial statistics, digital image analysis, and geostatistics.

The first half of this article is a condensed introduction to Markov point processes. The second half describes recent work by the author and collaborators (N.A. Cressie, N.I. Fisher, J. Møller, G. Nair, A. Särkkä and T.R. Turner) on finding new Markov models for different types of patterns, elaborating properties of these models, and performing statistical inference for spatial datasets using bootstrap, likelihood or pseudolikelihood methods.

2 Background

This section covers basic background about point process densities, Gibbs and Markov point processes, and conditional intensities.

2.1 Point process densities

See [18, 21] for definitions and background on point processes. In order that likelihoods may exist, we shall restrict attention to finite simple point processes whose distributions are absolutely continuous with respect to the distribution of the Poisson process.

Such a process may be visualised very easily as a random finite number of points at random locations in a space S. A realisation of the point process X is a finite unordered set of points,

$$\mathbf{x} = \{x_1, \ldots, x_n\}, \qquad x_i \in S, \quad n \geq 0$$

The space S in which the points lie is typically a subset of $\mathbb{R}^d$, but may be any Polish space. Let $\mathcal{X}$ be the space of all such realisations.

All the point process models X in this paper will be absolutely continuous with respect to the distribution of the Poisson point process [18, 43] with intensity measure ν on S where ν is a fixed, nonatomic, finite Borel measure. Then X has a *probability density* $f : \mathcal{X} \to [0, \infty]$ such that

$$\mathbb{P}\{X \in A\} = e^{-\nu(S)} \sum_{n=0}^{\infty} I_n(f, A) \tag{1}$$

for each $A \in \mathcal{F}$, where $I_0(f, A) = \mathbf{1}\{\emptyset \in A\} f(\emptyset)$ and for $n \geq 1$

$$I_n(f, A) = \frac{1}{n!} \int_S \cdots \int_S \mathbf{1}\{\{x_1, \ldots, x_n\} \in A\} f(\{x_1, \ldots, x_n\}) \, \mathrm{d}\nu(x_1) \ldots \mathrm{d}\nu(x_n).$$

In the simple case where S is a bounded subset of $\mathbb{R}^d$ and ν is the restriction to S of Lebesgue measure,

$$f(\{x_1, \ldots, x_n\}) \, \mathrm{d}x_1 \ldots \mathrm{d}x_n$$

(for distinct $x_1, \ldots, x_n \in S$) is the probability that the process consists of a point near each of the locations $x_1, \ldots, x_n$ and no other points.

Example 2.1 (Poisson process) Let $n(\mathbf{x})$ denote the number of points in a realisation $\mathbf{x} \in \mathcal{X}$. If $\alpha, \beta > 0$ are constants,

$$f(\mathbf{x}) = \alpha \beta^{n(\mathbf{x})}$$

is recognised from (1) as the density of the Poisson process with intensity measure $\beta\nu(\cdot)$, and the normalising constant α equals $\exp\{-(\beta - 1)\nu(S)\}$.

Example 2.2 For a function $\beta : S \to [0, \infty)$,

$$f(\mathbf{x}) = \alpha \prod_{i=1}^{n(\mathbf{x})} \beta(x_i) \tag{2}$$

is the density of the "inhomogeneous" Poisson process with intensity measure $\kappa(B) = \int_B \beta(u)\, \mathrm{d}\nu(u)$ on S and the normalising constant is

$$\alpha = \exp\left\{ -\int_S [\beta(u) - 1]\, \mathrm{d}\nu(u) \right\}.$$

2.2 Interpoint interactions

Definition 2.1 *A finite* Gibbs point process *is a finite simple point process with a density $f(\mathbf{x})$ satisfying the positivity condition*

$$f(\mathbf{x}) > 0 \quad \Rightarrow \quad f(\mathbf{y}) > 0 \text{ for all } \mathbf{y} \subset \mathbf{x}. \tag{3}$$

See [68, 81] and the excellent surveys by Ripley [74, 75]. By an application of the Möbius inversion formula or "inclusion-exclusion" [14, chap. 5,12] the density of any finite Gibbs point process can be written in the form

$$f(\mathbf{x}) = \exp\left\{ v_0 + \sum_{i=1}^{n(\mathbf{x})} v_1(x_i) + \sum_{i<j} v_2(x_i, x_j) + \sum_{i<j<k} v_3(x_i, x_j, x_k) + \ldots \right\} \tag{4}$$

where v_0 is constant and $v_k : S^k \to \mathbb{R} \cup \{-\infty\}$ are symmetric functions.

Thus the log likelihood of a particular configuration $\mathbf{x}$ is a sum of penalties incurred for the presence of each point $x_i \in \mathbf{x}$, for the interaction between each pair of points $x_i, x_j \in \mathbf{x}$, for the interaction between each triple of points $x_i, x_j, x_k \in \mathbf{x}$, and so on. The sum can be interpreted as the physical "potential energy" of the configuration. This interpretation is familiar in statistical physics [77, 69]; the individual functions v_k are called "interaction potentials".

Example 2.3 (Pairwise interaction) A *pairwise interaction process* on S has a density of the form

$$f(\mathbf{x}) = \alpha \prod_{i=1}^{n(\mathbf{x})} b(x_i) \prod_{i<j} h(x_i, x_j) \tag{5}$$

where $b : S \to \mathbb{R}_+$ is the 'activity' and $h : S \times S \to \mathbb{R}_+$ the 'interaction' function, and $\alpha > 0$ is the normalising constant.

The terms $b(x_i)$ in (5) influence the intensity and location of points, while the terms $h(x_i, x_j)$ introduce dependence ('interaction') between different points of the process X. Note that conditions must be imposed on b, h to ensure (5) is integrable. Typically α is not known explicitly; this is the 'partition function'.

Example 2.4 (Strauss process [82, 38]) This is a pairwise interaction process with constant activity $b(u) \equiv \beta$ and a 'threshold' interaction function

$$h(u,v) = \begin{cases} \gamma & \text{if } ||u-v|| \le r \\ 1 & \text{otherwise} \end{cases}$$

where $r > 0$ is a fixed interaction distance and $0 \le \gamma \le 1$ is the interaction parameter. Hence the probability density is

$$f(\mathbf{x}) = \alpha \beta^{n(\mathbf{x})} \gamma^{s(\mathbf{x})} \tag{6}$$

(taking $0^0 = 1$) where

$$s(\mathbf{x}) = \#\{(i,j) : i < j, \quad 0 < ||x_i - x_j|| \le r\}$$

is the number of unordered pairs of close points in $\mathbf{x}$. The Strauss process is well-defined for all $\gamma \in [0,1]$; the density is not integrable for $\gamma > 1$. If $\gamma = 1$, this reduces to the Poisson process with intensity $\beta\nu$. If $\gamma < 1$ the process exhibits "repulsion" or "inhibition" between points, since $s(x)$ tends to be smaller than under the Poisson model.

Example 2.5 (Hard core process) If $\gamma = 0$ the Strauss density (6) reduces to

$$f(\mathbf{x}) = \begin{cases} \alpha\beta^{n(\mathbf{x})} & \text{if } s(\mathbf{x}) = 0 \\ 0 & \text{if } s(\mathbf{x}) > 0 \end{cases} \tag{7}$$

called a *classical hard core process* with hard core diameter r. It is equivalent to a Poisson process with intensity $\beta\nu(\cdot)$ conditioned on the event that there are no points closer than r units apart.

Interpoint interactions of higher order also arise naturally.

Example 2.6 (Widom-Rowlinson process) Let S be a compact subset of $\mathbb{R}^2$. The Widom-Rowlinson penetrable sphere model [85], or 'area-interaction' process [7], has density

$$f(\mathbf{x}) = \alpha \beta^{n(\mathbf{x})} \gamma^{-A(\mathbf{x})} \tag{8}$$

where $\beta, \gamma > 0$ are parameters, $\alpha > 0$ is the normalizing constant, and $A(\mathbf{x})$ is the area of

$$U_r(\mathbf{x}) = \left(\bigcup_{i=1}^{n(\mathbf{x})} B(x_i; r) \right) \cap S,$$

where $B(x_i; r)$ is the disc of radius r centred at x_i. The density (8) is integrable for all values of $\gamma > 0$. The process produces clustered patterns

when $\gamma > 1$, ordered patterns when $0 < \gamma < 1$, and reduces to a Poisson process when $\gamma = 1$.

The Gibbs decomposition (4) of the density (8) can be computed explicitly by applying the inclusion-exclusion formula to the area of the union of the discs $B(x_i; r)$. Interaction terms of all orders are non-vanishing, i.e. the Widom-Rowlinson model has interactions of infinite order.

A simple but important relationship holds between a finite Gibbs point process and its conditional distributions.

Lemma 2.1 *Let X be a finite Gibbs point process on S with density f. Let $A \subset S$ be a compact subset. Then the conditional distribution of $X \cap A$ given $X \cap A^c$ is a finite Gibbs point process on A, with conditional density*

$$f_A(\mathbf{z} \mid \mathbf{y}) = \alpha(\mathbf{y}) f(\mathbf{z} \cup \mathbf{y}) \tag{9}$$

(with respect to the Poisson process on A whose intensity measure is the restriction of ν to A) for finite sets $\mathbf{z} \subset A$, $\mathbf{y} \subset A^c$, where $\alpha(\mathbf{y})$ is a normalising constant.

If f is expressed in terms of interaction potentials v_k as in (4), then the corresponding expression for f_A has interaction potentials

$$w_k(\mathbf{z}') = v_k(\mathbf{z}') + \sum_{\mathbf{y}' \subset \mathbf{y}} v_{n(\mathbf{y}' \cup \mathbf{z}')}(\mathbf{y}' \cup \mathbf{z}')$$

which is to say that interactions occur not only amongst the points of the configuration $\mathbf{z}$ but also between these random points and the 'fixed' points $\mathbf{y}$. (Note that the marginal distribution of $X \cap A$ does not satisfy a statement similar to Lemma 2.1.)

2.3 Conditional intensities

The (Papangelou) conditional intensity of a point process is the continuous space analogue of a certain conditional probability for discrete random fields.

The conditional intensity $\lambda(u; \mathbf{x})$ of X at a location $u \in S$ may be loosely interpreted as giving the conditional probability that X has a point at u given that the rest of the process coincides with $\mathbf{x}$:

$$\lambda(u; \mathbf{x}) = \lim_{\triangle \mathbf{u} \downarrow \{u\}} \frac{\mathbb{P}\{X \cap \triangle \mathbf{u} \neq \emptyset \mid X \setminus \triangle \mathbf{u} = \mathbf{x} \setminus \triangle \mathbf{u}\}}{\nu(\triangle \mathbf{u})}$$

where the limit is taken over decreasing open neighbourhoods $\triangle \mathbf{u}$ of $u \in S$.

Formally the conditional intensity is a Radon-Nikodým derivative defined to satisfy

$$\mathbb{E}\left[\sum_{x_i \in X \cap A} g(x_i, X)\right] = \int_A \mathbb{E}[\lambda(u; X) g(u, X)] \, \mathrm{d}\nu(u) \tag{10}$$

(the "Nguyen-Zessin formula") for all nonnegative bounded measurable functions $g : S \times \mathcal{X} \to \mathbb{R}_+$. See [37] for an informal introduction, or [28, 29, 36, 44] for details.

For any Gibbs process on W (see section 2.2) with density f, the conditional intensity at a point $u \in W$ equals

$$\lambda(u; \mathbf{x}) = \frac{f(\mathbf{x} \cup \{u\})}{f(\mathbf{x})} \tag{11}$$

if $u \notin \mathbf{x}$, while for $x_i \in \mathbf{x}$ we have $\lambda(x_i; \mathbf{x}) = f(\mathbf{x})/f(\mathbf{x} \setminus \{x_i\})$. In the statistical physics interpretation, $\log \lambda(u; \mathbf{x}) = \log f(\mathbf{x} \cup \{u\}) - \log f(\mathbf{x})$ is the energy required to add a new point u to an existing configuration $\mathbf{x}$.

For example, the inhomogeneous Poisson process with intensity function $\lambda(\cdot)$ has conditional intensity $\lambda(u; \mathbf{x}) = \lambda(u)$ at all points u. The general pairwise interaction process (5) has conditional intensity

$$\lambda(u; \mathbf{x}) = b(u) \prod_{\substack{i=1 \\ x_i \neq u}}^{n(\mathbf{x})} h(u, x_i). \tag{12}$$

Note that the intractable normalising constant in (5) has been eliminated in the conditional intensity. For this reason, inference based on the conditional intensity is typically easier than maximum likelihood.

2.4 Markov point processes

A Markov point process [76, 74, 75] is one in which interpoint interactions occur only between those points which are deemed to be 'neighbours'.

Example 2.7 Consider the pairwise interaction process (5) in $\mathbb{R}^2$. Assume the interaction function h has finite range $r > 0$, in the sense that $h(u, v) = 1$ whenever $||u - v|| > r$. Declare two points $u, v \in S$ to be *neighbours*, written $u \sim v$, if they are closer than r units apart:

$$u \sim v \text{ iff } ||u - v|| < r. \tag{13}$$

Then interactions occur only between neighbours, i.e. (5) becomes

$$f(\mathbf{x}) = \alpha \left[\prod_{i=1}^{n(\mathbf{x})} b(x_i) \right] \left[\prod_{x_i \sim x_j} h(x_i, x_j) \right] \tag{14}$$

where the second product is over all unordered pairs of *neighbouring* points. The conditional intensity (12) becomes

$$\lambda(u; \mathbf{x}) = b(u) \prod_{x_i \sim u} h(x_i, u) \tag{15}$$

where the product is over all *neighbours* of u in $\mathbf{x}$.

Note that in this example the conditional intensity (15) *depends only on u and on the neighbours of u in $\mathbf{x}$.* This important property signifies that interaction is "local".

Definition 2.2 (Ripley & Kelly [76]) *Let $\sim$ be a symmetric, reflexive relation on S. A* Markov point process *on S with respect to $\sim$ is a finite Gibbs point process whose conditional intensity $\lambda(u;\mathbf{x})$ depends only on u and $\{x_i \in \mathbf{x} : x_i \sim u\}$.*

For example, the inhomogeneous Poisson process (2) is a Markov point process with respect to any relation $\sim$ since $\lambda(u;\mathbf{x}) = \beta(u)$ depends only on u.

Example 2.8 For the Strauss process (Example 2.4)

$$\lambda(u,\mathbf{x}) = \beta\gamma^{t(u,\mathbf{x})} \tag{16}$$

where $t(u,\mathbf{x}) = s(\mathbf{x}\cup\{u\}) - s(\mathbf{x}) = \#\{x_i \in \mathbf{x} : 0 < ||x_i - u|| \le r\}$ is the number of points $x_i \in \mathbf{x}$ which are close to u, other than u itself. See Figure 1. Hence the Strauss process is Markov with respect to the relation $\sim$ of (13).

Figure 1. Illustration of conditional intensities. *Left:* Strauss process; *Right:* Widom-Rowlinson process. The conditional intensity of the Strauss process at a point u (○) depends on the number of existing points (•) of the configuration $\mathbf{x}$ which are closer than r units distant from u. In this illustration $t(u,\mathbf{x}) = 2$. The conditional intensity of the Widom-Rowlinson process at a point u (○) depends on the shaded area.

Example 2.9 For the Widom-Rowlinson process (Example 2.6)

$$\lambda(u;\mathbf{x}) = \beta\gamma^{T(u,\mathbf{x})} \tag{17}$$

where $T(u, \mathbf{x}) = A(\mathbf{x} \cup \{u\}) - A(\mathbf{x})$ is the area of the region $B(u; r) \setminus U_r(\mathbf{x})$. See Figure 1. Clearly $T(u, \mathbf{x})$ depends only on u and $\{x_i \in \mathbf{x} : ||u - x_i|| < 2r\}$. Thus, the Widom-Rowlinson process is Markov with respect to the relation (13) with r replaced by $2r$.

Definition 2.3 *Let $\sim$ be a symmetric relation on S. The* neighbourhood *of a set $A \subset S$ is*

$$\mathcal{N}(A) = \{u \in S : u \sim v \text{ for some } v \in A\}.$$

Definition 2.2 then states that a finite Gibbs point process is Markov if $\lambda(u; \mathbf{x})$ depends only on u and on $\mathcal{N}(u) \cap \mathbf{x}$.

The epithet 'Markov' for these processes is justified by the following conditional independence property.

Lemma 2.2 (Spatial Markov Property) *Let X be a Markov point process on S. Then the conditional distribution of $X \cap A$ given $X \cap A^c$ depends only on X in the neighbourhood $\mathcal{N}(A) \cap A^c$:*

$$P(X \cap A | X \cap A^c) = P(X \cap A | X \cap (\mathcal{N}(A) \cap A^c)).$$

In (4) we saw that a Gibbs density can be decomposed into interaction terms v_k of each order $k = 0, 1, 2, \ldots$ For the case of a Markov point process, the interaction term for a particular k-tuple of points is nontrivial only if all these points are neighbours. Grimmett [30] introduced the term "clique":

Definition 2.4 *Let $\sim$ be a symmetric, reflexive relation on S. A configuration $\mathbf{x} \in \mathcal{X}$ is a* clique *if all points of $\mathbf{x}$ are neighbours ($u \sim v$ for all $u, v \in \mathbf{x}$). A set containing 0 or 1 points is a clique.*

The following is analogous to the Hammersley-Clifford theorem for discrete Markov random fields [30, 16].

Theorem 2.1 (Ripley–Kelly [76]) *A finite simple point process with density f is a Markov point process iff its interaction potentials v_k satisfy $v_k(\mathbf{z}) = 0$ whenever $\mathbf{z}$ is not a clique.*

Equivalently, the density f is Markov iff

$$f(\mathbf{x}) = \prod_{\mathbf{z} \subseteq \mathbf{x}} \phi(\mathbf{z}) \tag{18}$$

where $\phi(\mathbf{z}) = \exp\{v_{n(\mathbf{z})}(\mathbf{z})\}$ is equal to 1 unless $\mathbf{z}$ is a clique.

For example, for the pairwise interaction density (5) the interaction potentials are $v_0 = \log \alpha$, $v_1(\{u\}) = \log b(u)$ for singletons, $v_2(\{u, v\}) = \log h(u, v)$ for two-point cliques, and $\varphi(\mathbf{y}) = 0$ for cliques $\mathbf{y}$ containing 3 or more points. The process is Markov iff $h(u, v) = 1$ for pairs of points u, v with $||u - v|| > r$.

2.5 Markov Chain Monte Carlo

Stochastic simulation of a finite Gibbs point process cannot in practice be performed by generic Monte Carlo techniques for sampling from a density such as the rejection method [32, 73]. For example, although the hard core process (Example 2.5) is the conditional distribution of a Poisson process given that no pair of points is closer than r units apart, the probability of this event for interesting cases is prohibitively small.

Instead, finite Gibbs point processes can be simulated using Markov Chain Monte Carlo (MCMC) techniques. Early examples are [55, 69, 72]; see the excellent reviews [26, 27]. In brief, these techniques involve running a Markov Chain (Y_t), in discrete or continuous time, with state space $\mathcal{X}$ (the space of all finite point patterns). The chain is designed to converge in distribution to the distribution of the point process X of interest, so that after a long run time the state of Y_t can be taken as a realisation of X. The chain must also be simple and quick to run. Typically the transitions or 'updates' of (Y_t) are simple operations such as the "birth" of a new point, $\mathbf{x} \mapsto \mathbf{x} \cup \{u\}$, where $\mathbf{x} \in \mathcal{X}, u \in S$; the "death" of an existing point, $\mathbf{x} \mapsto \mathbf{x} \setminus \{x_i\}$ where $x_i \in \mathbf{x}$; or the shifting of an existing point $x_i \in \mathbf{x}$ to a new location u.

To ensure that the stationary distribution of (Y_t) is the distribution π of X, it is sufficient and convenient to require that the transition kernel $P^t(x, A) = \mathbb{P}\{Y_t \in A \mid Y_0 = x\}$ be in 'detailed balance' with π,

$$\int_A P^t(x, B)\, \mathrm{d}\pi(x) = \int_B P^t(y, A)\, \mathrm{d}\pi(y) \tag{19}$$

for all $t > 0$. If, for example, the only possible transitions are instantaneous births $\mathbf{x} \mapsto \mathbf{x} \cup \{u\}$ at rate $b(\mathbf{x}, u)\, \mathrm{d}\nu(u)$ and instantaneous deaths $\mathbf{x} \mapsto \mathbf{x} \setminus \{x_i\}$ at rate $D(\mathbf{x}, x_i)$, then detailed balance is equivalent to $b(\mathbf{x}, u)/D(\mathbf{x} \cup \{u\}, u) = \lambda(u; \mathbf{x})$ whenever $f(\mathbf{x} \cup \{u\}) > 0$. This can be achieved by various schemes of Gibbs and Metropolis-Hastings type. If such a process (Y_t) exists (if the backwards equations have a unique solution) then it is irreducible and time reversible, and π is its unique equilibrium distribution [69].

The convergence of (Y_t) can be extremely slow, and is difficult to measure. This limitation was lifted recently following the work of Propp and Wilson [70] who developed a coupling algorithm for drawing exact simulations from the equilibrium distribution of a discrete state Markov chain. This idea has been adapted to some spatial birth-and-death processes to obtain exact simulation algorithms for certain finite Gibbs point processes [27, 40, 41, 31]. The virtues of exact simulation algorithms are that the output is guaranteed to have the correct distribution, and that the computation time is usually orders of magnitude smaller than that required for the convergence of Metropolis-Hastings algorithms.

3 Pseudolikelihood inference

This section describes Besag's concept of pseudolikelihood for point processes, and reports on recent work by the author and Rolf Turner [3] on fitting Gibbs/Markov point process models using pseudolikelihood.

3.1 Pseudolikelihood

Suppose we have data consisting of a spatial point pattern $\mathbf{x}$ observed in a bounded region W of $\mathbb{R}^d$. Thus $\mathbf{x} = \{x_1, \ldots, x_n\}$ where the number of points $n \geq 0$ is not fixed, and each x_i is a point in W. There may also be spatial covariates.

The aim is to fit to the data a finite Gibbs point process model with density $f_\theta(\mathbf{x})$ governed by a parameter θ ranging over $\Theta \subseteq \mathbb{R}^p$.

It is generally difficult to evaluate and maximise the likelihoods of point processes. The loglikelihood of the inhomogeneous Poisson process (2) includes an integral requiring iterative optimization methods. Even simple exponential family models such as the pairwise interaction processes (5) include a normalising constant which is an intractable function of θ. Methods for approximating $\alpha(\cdot)$ and maximising likelihood include functional expansions of $\alpha(\cdot)$, Monte Carlo integration, and analogues of E-M and stochastic approximation [27, 56, 57, 58, 59, 60, 63].

An alternative to the likelihood function is the *pseudolikelihood* [10, 11, 12, 35] which we describe here. See [22, 23, 24, 74, 75, 78, 83] for other applications.

Originally Besag [10, 11] defined the pseudolikelihood of a finite set of random variables $X_1, \ldots, X_n$ as the product of the conditional likelihoods of each individual X_i given the other variables $\{X_j : j \neq i\}$. This was extended [11, 12] to point processes, for which it can be viewed as an infinite product of infinitesimal conditional probabilities.

Besag [11] defined the *pseudolikelihood* of a point process with conditional intensity $\lambda_\theta(u; \mathbf{x})$ to be

$$\mathrm{PL}\,(\theta\,;\,\mathbf{x}) = \left[\prod_{i=1}^{n(\mathbf{x})} \lambda_\theta(x_i; \mathbf{x})\right] \exp\left\{-\int_S \lambda_\theta(u; \mathbf{x})\,\mathrm{d}u\right\} \tag{20}$$

Further theory was developed in [12, 34, 35].

If the process is Poisson the pseudolikelihood coincides with the likelihood (2) up to the factor $\exp(|S|)$. For a pairwise interaction process (5), the pseudolikelihood is

$$\begin{aligned} &\mathrm{PL}\,(\theta\,;\,\mathbf{x}) \\ &= \left[\prod_{i=1}^{n(\mathbf{x})} b_\theta(x_i)\right] \left[\prod_{i \neq j} h_\theta(x_i, x_j)\right] \exp\left\{-\int_S b_\theta(u) \prod_{i=1}^{n(\mathbf{x})} h_\theta(u, x_i)\,\mathrm{d}u\right\} \end{aligned} \tag{21}$$

in which the intractable normalising constant $\alpha(\theta)$ appearing in the likelihood (5) has been replaced by an exponential integral in (21) as if the process were Poisson.

For processes with 'weak interaction' in the sense that $\lambda_\theta(u;\mathbf{x})$ can be approximated well by a function of u only, the process is approximately Poisson and the pseudolikelihood is an approximation to the likelihood. Hence the maximum pseudolikelihood estimator should be efficient if interaction is weak. Folklore holds that it is inefficient for strong interactions.

For an exponential family model, the maximum pseudolikelihood normal equations $\partial/\partial\theta \log \mathrm{PL}\,(\theta;\mathbf{x}) = 0$ can be shown to be unbiased estimating equations using the Nguyen-Zessin formula (10). Diggle *et al* [22] showed in the stationary case that maximum pseudolikelihood is a special case of the Takacs-Fiksel method, itself an application of the method of moments [23, 24, 83]. These estimating equations can also be derived naturally from properties of the Markov chains used in MCMC methods [4].

Jensen and Møller [35] proved that for Gibbs point processes with exponential family likelihoods, the pseudolikelihood is log-concave and the maximum pseudolikelihood estimator is consistent as $S \nearrow \mathbb{R}^d$, under suitable conditions. Jensen and Künsch [34] proved the MPLE is asymptotically normal for stationary pairwise interaction processes, under suitable conditions (see (C1) and (C2) of [34]). There may be room for considerable generalisation, since the latter results impose strong constraints on the interaction potential which are not needed for the case of discrete random fields [17].

The pseudolikelihood of a point process is analogous to the pseudolikelihood of a discrete (Markov) random field as defined in [10]. Indeed [11, 12] certain classes of point processes can be obtained as the a.s. limit of a sequence of Markov random fields defined on discrete lattices whose spacing tends to zero; the pseudolikelihood function of the Markov random field converges pointwise to the pseudolikelihood of the point process. Recent applications include [78].

3.2 Computational device for maximum pseudolikelihood

In [3] we proposed a computational device for obtaining approximate maximum pseudolikelihood estimates. The method is an adaptation of a technique of Berman and Turner [9]. Related ideas have been explored by Lindsey [48, 49, 50, 51].

Approximating the integral in (20) by a finite sum using any quadrature rule, we may approximate the log pseudolikelihood

$$\log \mathrm{PL}\,(\theta\,;\,\mathbf{x}) \approx \sum_{i=1}^{n(\mathbf{x})} \log \lambda_\theta(x_i\,;\,\mathbf{x}) - \sum_{j=1}^{m} \lambda_\theta(u_j\,;\,\mathbf{x})\, w_j \tag{22}$$

where u_j, $j = 1, \ldots, m$ are points in W and $w_j > 0$ are quadrature weights. Note that if the list of points $\{u_j, j = 1, \ldots, m\}$ *includes all the data points* $\{x_i, i = 1, \ldots, n\}$, then we can rewrite (22) as

$$\log \mathrm{PL}\,(\theta\,;\,\mathbf{x}) \approx \sum_{j=1}^{m} (y_j \log \lambda_j - \lambda_j)\; w_j \tag{23}$$

where $\lambda_j = \lambda_\theta(u_j)$ and $y_j = z_j/w_j$, where

$$z_j = \begin{cases} 1 & \text{if } u_j \text{ is a data point, } u_j \in \{x_1, \ldots, x_n\} \\ 0 & \text{if } u_j \text{ is a dummy point, } u_j \notin \{x_1, \ldots, x_n\}. \end{cases} \tag{24}$$

The right side of (23), for fixed $\mathbf{x}$, is formally equivalent to the log likelihood of independent Poisson variables $Y_k \sim \mathrm{Poisson}(\lambda_k)$ taken with weights w_k.

The expression (23) can therefore be maximised using standard software for fitting Generalised Linear Models [53]. This makes it possible to fit rapidly a wide variety of Gibbs point process models incorporating effects such as spatial trend, dependence on covariates, interpoint interaction, and mark information.

4 Identifiability

It is relatively straightforward to construct Markov point process models since there is an explicit characterisation of their densities (Theorem 2.1). The interaction potentials may be chosen virtually at will, subject to the requirement that the density be integrable. However, the behaviour of the resulting process is difficult to determine. In particular it is not clear whether the resulting process will be distinguishable from the Poisson process and whether the parameters will be identifiable.

This is important in the case of the Widom-Rowlinson process. Simulated realizations of both the repulsive and attractive cases displayed in [7] and [40] do not seem to differ markedly from Poisson patterns. This has been further investigated by A. Särkkä and the author [79].

Recall that the Widom-Rowlinson density (8) involves the area $A(\mathbf{x})$ of the union of discs of radius r centred at the points $x_i \in \mathbf{x}$, intersected with S. If r is small, then under the reference Poisson process, there is a high probability that these discs do not overlap, so that $A(X)$ is equal to $n(X) \cdot \pi r^2$ with high probability. Thus

$$f_{\beta,\gamma}(X) = \alpha \delta^{n(X)}$$

with high probability under the reference Poisson process, where $\delta = \beta\gamma^{-\pi r^2}$. Thus, when r is small, the Widom-Rowlinson process is approximately Poisson with intensity δ. The parameters β, γ are not identifiable, only the derived parameter δ.

Alternatively, if r is large, then either $\mathbf{x}$ is empty or the discs cover the whole domain S, so that

$$f_{\beta,\gamma}(X) = \begin{cases} \alpha'\beta^{n(X)} & \text{if } n(X) > 0 \\ 0 & \text{if } n(X) = 0 \end{cases}$$

with high probability under the Poisson process. Thus when r is large, the Widom-Rowlinson process is approximately a mixture of a Poisson process with intensity β and the process which is a.s. empty. The parameter γ is not identifiable.

This is an instance of the general fact that in a 2-parameter exponential family

$$f_{\theta_1,\theta_2}(x) = \exp(\theta_1 S(x) + \theta_2 T(x)),$$

if S and T are linearly dependent statistics under the reference distribution, then the model degenerates to a 1-parameter or 0-parameter family and the parameters are not identifiable.

For a general exponential family $f_\theta(x) = c\exp(\theta^T B(x))$, where θ and $B(x)$ are p-dimensional, θ is efficiently estimable iff μ lies in the convex hull of the support of the distribution of $B(X)$ under the reference distribution ($\theta = 0$). Geyer [27] has made very similar comments in relation to Monte Carlo maximum likelihood methods for Gibbs point processes.

We have investigated this aspect of identifiability for the Widom-Rowlinson process by simulation. Figure 4 shows scatterplots of the empirical distribution of $(n(X), A(X))$ under the Poisson process, for various values of r. The first and last plots, for $r = 0.02$ and $r = 0.12$, confirm the predictions that for small and large r values, respectively, the statistics $n(X), A(X)$ are linearly dependent. In the middle of the range, $r \approx 0.08$, the statistics appear to be linearly independent.

Since the Widom-Rowlinson process degenerates to a Poisson process or a Poisson/empty mixture in cases of linear dependence, the question is whether for some values of r the process is distinguishable from a Poisson process.

We investigated this by computing the total variation distance between the Widom-Rowlinson process and a Poisson process with equal intensity. Let P and Q be any probability distributions having densities f and g (respectively) with respect to some reference measure μ. The total variation distance [71, sections 1.3–1.4] is

$$||Q - P|| = \sup_{B \in \mathcal{B}} |Q(B) - P(B)| = \frac{1}{2}\mathbb{E}_P \left| \frac{g(X)}{f(X)} - 1 \right|. \tag{25}$$

Now let Q be the distribution of the Widom-Rowlinson process with parameters β, γ and P the distribution of the Poisson process with intensity λ.

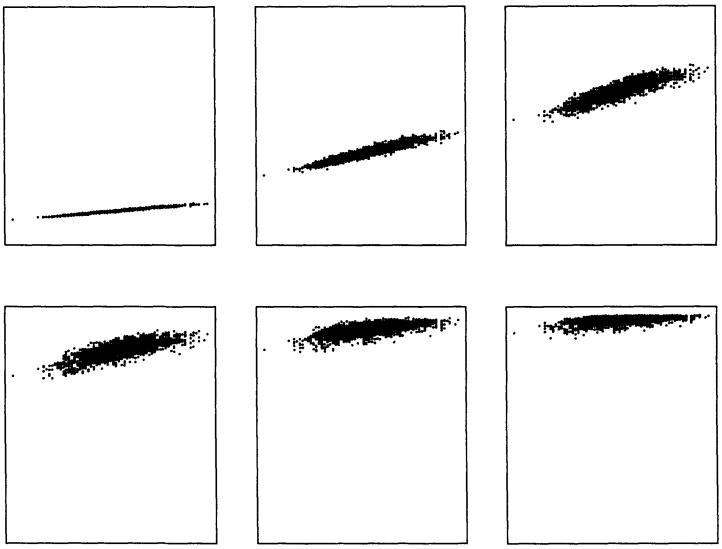

Figure 2. Scatterplots of the observed values of the sufficient statistics $n(\mathbf{x})$ and $A(\mathbf{x})$ of the Widom-Rowlinson process, for various values of r, generated by the Poisson point process with $\beta = 100$. All panels show $(n(X), A(X))$ on the same scale. The values of r are (top row, left to right) 0.2, 0.4, 0.6, (bottom row, left to right) 0.8, 1.0 and 1.2.

Then

$$||Q - P|| = \frac{1}{2}\,\mathbb{E}_P \left| \frac{\alpha(\beta,\gamma)}{\alpha(\lambda,1)} \left(\frac{\beta}{\lambda}\right)^{n(X)} \cdot \gamma^{-A(X)} - 1 \right|$$

which can be estimated by simulation as follows. First we generate an adequate number of simulated realisations from Q to estimate the intensity of the Widom-Rowlinson process. Setting λ equal to this estimated intensity, we generate an adequate number of simulated realisations $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)}$ of P, the Poisson process with intensity λ. We estimate the ratio of normalising constants $\alpha(\beta,\gamma)/\alpha(\lambda,1)$ following [27, 63] by

$$\frac{\sum_{i=1}^{m} \lambda^{n(\mathbf{x}^{(i)})}}{\sum_{i=1}^{m} \beta^{n(\mathbf{x}^{(i)})} \gamma^{-A(\mathbf{x}^{(i)})}}$$

and estimate the total variation distance $||Q - P||$ by

$$\frac{1}{2m} \sum_{i=1}^{m} \left| \frac{\alpha(\beta,\gamma)}{\alpha(\lambda,1)} \left(\frac{\beta}{\lambda}\right)^{n(x^{(i)})} \cdot \gamma^{-A(x^{(i)})} - 1 \right|.$$

Figure 3 shows estimates of total variation distance plotted against disc radius r. They refer to the Widom-Rowlinson process in the unit square

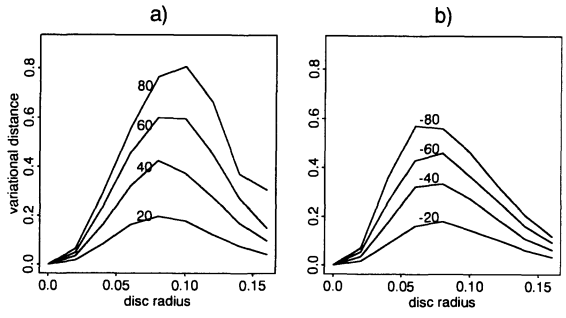

Figure 3. Total variation distance between some a) attractive and b) repulsive area-interaction processes and the corresponding Poisson process with the same intensity; $\beta = 100$ and the labels 20, 40, ... refer to the values of $\log \gamma$.

with $\beta = 100$ and $\log \gamma = -80, -60, \ldots, 60, 80$ for various values of r, and the Poisson process with equal intensity. The curves confirm the prediction that for small r and for large r the Widom-Rowlinson process degenerates to a Poisson process. However for moderate values of r they show that the Widom-Rowlinson process is distinguishable from the Poisson process. We also calculated the Hellinger and Kullback-Leibler distances which yield very similar results. These distance curves can be used to determine reasonable values for r for data analysis. For further details see [79].

It is common in statistical physics to construct 'phase diagrams' which partition the parameter space into regions where the process exhibits different 'regimes' of behaviour. The graphs in Figure 3 could be reinterpreted as a three-dimensional surface of total variation values for each point in (γ, r) space; and a level set of this surface could be regarded as a phase diagram. See also [54].

5 'Dynamic' Markov point processes

In spatial statistics there is considerable interest in developing new point process models, because the existing repertoire is thought to be too narrow and unrealistic for many applications. In particular, clustered point pattern data are thought to be very difficult to model by Markov point processes.

An extension of Markov point processes was proposed in [6]. Recall that the interpoint interactions in a Markov point process occur between all points x_i, x_j of the configuration $\mathbf{x}$ that are 'close' in a predefined sense ($x_i \sim x_j$). In [6] this is extended by allowing the definition of 'close' to depend on the configuration $\mathbf{x}$. This allows interactions to occur only between those points

which are close in the context of the configuration – for example, nearest neighbours — and for interaction to occur at any distance.

For example, consider a renewal process in one-dimensional time. The intervals between successive points of the process are i.i.d. random variables. The density of this process on $[0,T]$ is of the form

$$f(\mathbf{x}) = v(x_1) \prod_{i=2}^{n(\mathbf{x})} w(x_{i-1}, x_i) \tag{26}$$

where $x_1 < x_2 < \ldots < x_n$ are the points of $\mathbf{x}$. While the form of (26) is very similar to a pairwise interaction density (5), this is not a Markov point process with respect to any nontrivial fixed relation $\sim$ on $\mathbb{R}$. Inter-point interactions occur only between *consecutive* pairs of points, and occur regardless of the distance between them.

Again, Ord [62] suggested that for two-dimensional point pattern data in geography, an improvement on the Strauss process would be a model taking account of the size of the 'territory' of each point as expressed by its Dirichlet cell:

$$f(\mathbf{x}) = \alpha\beta^n \prod_i g(\text{ area of } C(x_i|\mathbf{x})) \tag{27}$$

where $C(x_i|\mathbf{x})$ is the Dirichlet or Voronoi cell associated with point x_i in the configuration $\mathbf{x}$,

$$C(x_i \mid \mathbf{x}) = \{u \in \mathbb{R}^2 \,:\, ||u - x_i|| = \min_j ||u - x_j||\}$$

see [61]. Sibson [80] and others have suggested using the areas of Dirichlet-Voronoi cells and the associated neighbour distances as statistics for point pattern data.

Processes such as (27) do not have the Markov property as in Definition 2.2 because $C(x_i|\mathbf{x})$ depends on x_i and its Dirichlet neighbours (those x_j such that $C(x_j \mid \mathbf{x})$ has a common edge with $C(x_i \mid \mathbf{x})$) regardless of how far away these neighbours are.

In [6] we defined a Markov property analogous to the Ripley-Kelly definition 2.2, except that the concept of "neighbourhood" now depends on context and is denoted $\underset{\mathbf{x}}{\sim}$.

Definition 5.1 *Assume that for each $\mathbf{x} \in \mathcal{X}$ there is given a symmetric, reflexive relation $\underset{\mathbf{x}}{\sim}$ defined on $\mathbf{x}$. If points $u, v \in \mathbf{x}$ are related under $\underset{\mathbf{x}}{\sim}$ we write $u \underset{\mathbf{x}}{\sim} v$ and say that u and v are* $\mathbf{x}$-neighbours. *The* $\mathbf{x}$-neighbourhood *of a subset $\mathbf{y} \subseteq \mathbf{x}$ is*

$$\mathcal{N}(\mathbf{y}|\mathbf{x}) = \{u \,:\, u \underset{\mathbf{x}}{\sim} v \text{ for some } v \in \mathbf{y}\}.$$

The subset $\mathbf{y} \subseteq \mathbf{x}$ *is a* clique in $\mathbf{x}$ *if all members of* $\mathbf{y}$ *are* $\mathbf{x}$*-neighbours* ($u \underset{\mathbf{x}}{\sim} v$ *for all* $u, v \in \mathbf{y}$). *A set containing 0 or 1 members is a clique.*

Example 5.1 (Renewal process) In one dimension, given the points $\mathbf{x} = \{x_1, \ldots, x_n\} \subset \mathbb{R}$ with $x_1 < x_2 < \ldots < x_n$, define $x_i \underset{\mathbf{x}}{\sim} x_{i+1}$ for $i = 1, \ldots, n-1$ and let no other relations hold. That is, each point is a neighbour of the next and the previous elements in the sequence.

Example 5.2 (Delaunay neighbours) In $\mathbb{R}^2$ define $x_i \underset{\mathbf{x}}{\sim} x_j$ iff x_i, x_j are neighbors in the Delaunay tessellation generated by $\mathbf{x}$. That is, x_i, x_j are neighbours iff their Dirichlet cells $C(x_i \mid \mathbf{x})$, $C(x_j \mid \mathbf{x})$ have a common edge.

Example 5.3 (Connected components) In $\mathbb{R}^2$ let $\sim$ be any fixed relation such as (13), and define $\underset{\mathbf{x}}{\sim}$ to be the transitive closure of $\sim$ on $\mathbf{x}$. Thus, $x_i \underset{\mathbf{x}}{\sim} x_j$ iff

$$x_i \sim y_1 \sim y_2 \sim \ldots \sim y_m \sim x_j$$

for some $y_k \in \mathbf{x}$. Two points are $\underset{\mathbf{x}}{\sim}$-neighbours if they belong to the same connected component of the graph induced by $\sim$ on $\mathbf{x}$.

Next we generalise Definition 2.2 of Markov point processes to the case of non-constant relations $\underset{\mathbf{x}}{\sim}$. This should at least embrace functions of the form

$$f(x) = \alpha \beta^n \prod_{x_i \underset{\mathbf{x}}{\sim} x_j} g(||x_i - x_j||).$$

However since $\underset{\mathbf{x}}{\sim}$ now depends on x,

$$\frac{f(x \cup u)}{f(x)} = \beta \prod_{x_i \underset{\mathbf{x} \cup u}{\sim} u} g(||x_i - u||) \frac{\prod_{x_i \underset{\mathbf{x} \cup u}{\sim} x_j} g(||x_i - x_j||)}{\prod_{x_i \underset{\mathbf{x}}{\sim} x_j} g(||x_i - x_j||)};$$

the new terms arise because some pairs $x_i, x_j \in x$ may be neighbours with respect to $\underset{\mathbf{x}}{\sim}$ but not $\underset{\mathbf{x} \cup u}{\sim}$, or vice versa. If $f(x \cup u)/f(x)$ is to depend only on "local" information, then $\underset{\mathbf{x} \cup u}{\sim}$ must only differ from $\underset{\mathbf{x}}{\sim}$ in a "neighbourhood" of u. Thus we need to impose conditions on $\underset{\mathbf{x}}{\sim}$.

Definition 5.2 *For a dynamic relation* $\underset{\mathbf{x}}{\sim}$ *define the* clique indicator function

$$\chi(\mathbf{y}|\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{y} \text{ is a clique in } \mathbf{x} \\ 0 & \text{if not.} \end{cases}$$

Define the following conditions on $\underset{\mathbf{x}}{\sim}$. *Let* $\mathbf{y} \subset \mathbf{z} \in \mathcal{X}$ *and* $u, v \in S$ *with* $u, v \notin \mathbf{z}$ *and* $\mathbf{z} \cup \{u, v\} \in \mathcal{X}$.

$(C1)$ $\chi(\mathbf{y}|\mathbf{z}) \neq \chi(\mathbf{y}|\mathbf{z} \cup \{u\})$ *implies* $\mathbf{y} \subset \mathcal{N}(u|\mathbf{z} \cup \{u\})$;

$(C2)$ *if* $u \not\underset{\mathbf{x}}{\sim} v$ *where* $\mathbf{x} = \mathbf{z} \cup \{u, v\}$, *then* $\chi(\mathbf{y}|\mathbf{z} \cup \{u\}) + \chi(\mathbf{y}|\mathbf{z} \cup \{v\}) = \chi(\mathbf{y}|\mathbf{z}) + \chi(\mathbf{y}|\mathbf{x})$.

Note particularly the strength of the conclusion in $(C1)$, i.e. "*all* points of y are neighbours of u". It is shown in [6] that Examples 5.1–5.3 and several other examples satisfy (C1) and (C2).

Definition 5.3 *A finite Gibbs point process with density f is a* dynamic Markov *point process with respect to $\underset{\mathbf{x}}{\sim}$ if the conditional intensity $\lambda(u; \mathbf{x})$ depends only on u, $\mathcal{N}(u|x \cup u)$ and on the restrictions of the relations $\underset{\mathbf{x}}{\sim}$, $\underset{\mathbf{x} \cup u}{\sim}$ to $\mathcal{N}(u|x \cup u)$.*

In words, the conditional intensity depends only on the added point u, on the neighbours of u (after its addition), and on information whether the addition of u has altered neighbourhood relations between any of these points.

If the relation $\underset{\mathbf{x}}{\sim}$ does not depend on x, then this is equivalent to the original Ripley-Kelly definition of a Markov point process, Definition 2.2.

Dynamic Markov point processes with respect to the connected component relation (Example 5.3) are studied further in [5, 15, 84]. This class includes many cluster processes, and is closed under superposition.

The following result generalises the Hammersley-Clifford theorem for Markov point processes (Theorem 2.1).

Theorem 5.1 *Let $\underset{\mathbf{x}}{\sim}$ be a system of neighbour relations satisfying (C1)–(C2). Then a finite Gibbs point process with density f is a dynamic Markov point process with respect to $\underset{\mathbf{x}}{\sim}$ if and only if*

$$f(\mathbf{x}) = \prod_{\mathbf{y} \subset \mathbf{x}} \varphi(\mathbf{y})^{\chi(\mathbf{y}|\mathbf{x})} \tag{28}$$

(taking $0^0 = 0$) for all $\mathbf{x} \in \mathcal{X}$, where $\varphi : \mathcal{X} \to [0, \infty)$ satisfies

$(I1)$ *if $\varphi(\mathbf{x}) > 0$ then $\varphi(\mathbf{y}) > 0$ for all $\mathbf{y} \subset \mathbf{x}$;*

$(I2)$ *if $\varphi(\mathbf{x}) > 0$ and $\varphi(\mathcal{N}(u|\mathbf{x} \cup \{u\})) > 0$, then $\varphi(\mathbf{x} \cup \{u\}) > 0$.*

When $f(\mathbf{x}) > 0$ the decomposition (28) reduces to

$$f(\mathbf{x}) = \prod_{\text{cliques}} \varphi(\mathbf{y}) \tag{29}$$

where the product is over all $\mathbf{y} \subset \mathbf{x}$ with $\chi(\mathbf{y}|\mathbf{x}) = 1$.

This larger class of dynamic Markov point processes is amenable to MCMC techniques, with some increase in complexity of the algorithms. Kendall [39] proved an analogue of the spatial Markov property (Lemma 2.2). Monte Carlo maximum likelihood or maximum pseudolikelihood techniques can be applied. Baddeley and Turner [3] use pseudolikelihood to fit Ord's process (27) to point pattern data.

6 Directed Markov point processes

The simulation of Markov point processes is still computationally expensive, despite recent advances. This is a bottleneck for many applications, and also retards the development of our mathematical understanding of such processes. This is in contrast to Markov processes in one-dimensional time, which are relatively easy to simulate using the natural ordering of the real line.

One strategy for avoiding the bottleneck is to modify the interpoint interactions in a spatial Markov point processes so that they respect a partial order. This is explored in our recent papers [2, 20] for one special case. Consider point processes in the unit square $S = [0,1]^2$ in $\mathbb{R}^2$, and define a partial order $\preceq$ on S by

$$(u,v) \preceq (u',v') \text{ iff } u \leq u' \text{ and } v \leq v'.$$

Interpoint interactions will occur only between points that are related in this partial order.

We might call such processes *directed Markov point processes*. They are analogous to the directed Markov random fields (Markov mesh models) on a discrete lattice, studied by Abend et al. [1], Pickard [64, 65, 66, 67], Lauritzen, Spiegelhalter et al. [46, 45], Cressie [19, 52, 47] and others [25, 42, 8]. Directed Markov random fields can be simulated exactly in a single pass over the lattice, in close analogy with the simulation of Markov chains in one-dimensional time.

Naively one could try to construct directed Markov point processes by writing down probability densities $f(\mathbf{x})$ which are products of terms associated with subsets $\mathbf{y} \subseteq \mathbf{x}$ of points that are related in the partial order, or by similarly constraining the factorisation (18). However, this turns out to be incorrect; such densities do not have the desired dependence properties.

Instead one should modify the spatial Markov property (Lemma 2.2) and the conditional intensity (section 2.3) and derive the necessary form of $f(\mathbf{x})$. This approach is pursued in [2]. Another way to construct specific point process models of this directed type is to take the limit of a sequence of directed Markov fields defined on increasingly finer discrete lattices. This approach is explored in a few special cases in [20].

For the partial order considered here, general results are already known from the theory of counting processes in multidimensional time [33]. A finite point process X in S is equivalent to a counting process $(N_z, z \in S)$ where N_z counts the number of points of X dominated by z. For $z = (z_1, z_2) \in S$ let $D(z) = [0, z_1] \times [0, z_2] = \{u \in S : u \preceq z\}$ be the set of points dominated by z, and $D^*(z) = ([0, z_1] \times [0, 1]) \cup ([0, 1] \times [0, z_2]) = \{u \in S : z \not\preceq u\}$ the set of points that do not dominate z. See Figure 4. Let $\mathcal{F}_z, \mathcal{F}_z^*$ be the σ-fields of events generated by $X \cap D(z)$ and $X \cap D^*(z)$ respectively.

Figure 4. The regions $D(z)$ (left) and $D^*(z)$ (right).

Then $(N_z, z \in S)$ has a *directed conditional intensity* $\lambda^+(u, \mathbf{x})$ if

$$N_z - \int_{D(z)} \lambda^+(u, X)\, \mathrm{d}u \tag{30}$$

is a 'strong martingale' [33] with respect to the filtration $\mathcal{F}^* = (\mathcal{F}_z^*, z \in S)$. Equivalently

$$\mathbb{E} \sum_{x_i \in X} g(x_i, X) = \mathbb{E} \int_S g(u, X)\, \lambda^+(u, X)\, \mathrm{d}u \tag{31}$$

for all $\mathcal{F}^*$-predictable functions $g(u, X)$. This is very similar in form to the Nguyen-Zessin identity (10) which defines the "undirected" Papangelou conditional intensity.

Our paper [2] proves equivalences between several different versions of the spatial Markov property, one being the condition known as (F4) in [33], and others being similar to the conclusion of Lemma 2.2 or to the one-dimensional Markov property.

The following results are known from multiparameter counting process theory [33, Theorems 1.3, 2.1] and [13]. Suppose λ^+ exists and $\int_S \lambda^+(u; \mathbf{x})\, \mathrm{d}u$ is bounded above, uniformly in $\mathbf{x}$. Then X has a probability density f which satisfies the Mazziotto-Szpirglas exponential formula

$$f(\mathbf{x}) = \left[\prod_{x_i \in \mathbf{x}} \lambda^+(x_i; \mathbf{x}) \right] \exp\left\{ \int_S \left[1 - \lambda^+(u; \mathbf{x}) \right] \mathrm{d}u \right\} \tag{32}$$

Furthermore, a similar expression holds for the probability density f_z of the subprocess

$$X \cap D(z) = \{x_i \in X : x_i \preceq z\}$$

with respect to the Poisson process on $D(z)$.

Conversely if $\lambda^+(u; X)$ is any positive, predictable process such that $\int_S \lambda^+(u; \mathbf{x})\, du$ is bounded above uniformly in $\mathbf{x}$, then the function f constructed by (32) is a probability density for a point process absolutely continuous with respect to the unit Poisson process, satisfying the same measurability properties [33, Theorem 1.3, p. 275].

Incidentally the undirected conditional intensity $\lambda(u; \mathbf{x})$ is related to the directed conditional intensity $\lambda^+(u; \mathbf{x})$ through (32),

$$\begin{aligned} \lambda(u; \mathbf{x}) &= \lambda^+(u, \mathbf{x} \cup \{u\}) \prod_{i=1}^{n} \frac{\lambda^+(x_i, \mathbf{x} \cup \{u\})}{\lambda^+(x_i, \mathbf{x})} \\ &\quad \times \exp\left(- \int_S \left[\lambda^+(v, \mathbf{x} \cup \{u\}) - \lambda^+(v, \mathbf{x}) \right] dv \right) \end{aligned} \tag{33}$$

Note the similarity of (32) to the expressions for the likelihood of a Poisson process (2) and the pseudolikelihood of a finite Gibbs process (20). It differs from the likelihood of a general finite Gibbs process (4) in that (32) contains an integral over S. The latter effectively replaces the intractable normalising constant in (4).

For example if λ^+ is a deterministic function, $\lambda^+(u, X) = g(u)$, we obtain the inhomogeneous Poisson process with intensity function g.

Example 6.1 ('Directed Hard Core') By analogy with the Hard Core process (Example 2.5), let

$$\lambda^+(z, \mathbf{x}) = \begin{cases} 1 & \text{if } \|z - x_i\| \geq r \text{ for all } x_i \preceq z \\ 0 & \text{otherwise} \end{cases}$$

where $r > 0$ is a fixed parameter, the interaction distance. The function $f(\mathbf{x})$ resulting via (32) is

$$f(\mathbf{x}) = I(\mathbf{x})\, e^{|S| - |U(\mathbf{x})|}$$

where $I(\mathbf{x})$ is the indicator function

$$I(\mathbf{x}) = \begin{cases} 1 & \text{if } \|x_j - x_i\| \geq r \text{ for all } x_i, x_j \in \mathbf{x} \text{ such that } x_i \preceq x_j \\ 0 & \text{otherwise} \end{cases}$$

and $|U(\mathbf{x})|$ is the area of the region

$$\begin{aligned} U(\mathbf{x}) &= \{u \in S : \|u - x_i\| > r \text{ for all } x_i \in \mathbf{x} \text{ such that } x_i \preceq u\} \\ &= S \setminus \bigcup_i C_r(x_i) \end{aligned}$$

where $C_r(z) = b(z, r) \setminus R^*(z)$. See Figure 5.

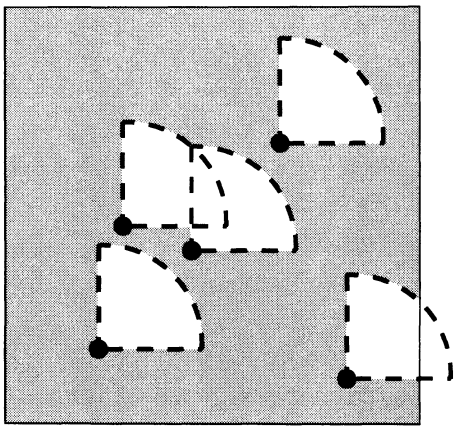

Figure 5. A typical realisation of the directed hard core process in Example 6.1. Filled dots: points $x_i \in \mathbf{x}$. Dotted circular sectors: forbidden regions $C_r(x_i)$. Shaded area: permitted region $U(\mathbf{x})$.

Example 6.2 ('Directed Strauss') By analogy with the Strauss process set

$$g(z, \mathbf{x}) = \gamma^{s(z,\mathbf{x})}$$

where $0 \leq \gamma \leq 1$ and

$$s(z, \mathbf{x}) = \#\{x_i \in \mathbf{x} : x_i \preceq z \text{ and } ||x_i - z|| < r\}$$

is the number of points of $\mathbf{x}$ which are closer to z than a fixed distance $r > 0$ and which precede z in the partial order. The case $\gamma = 0$ reduces to the previous Example.

One of the chief advantages of directed Markov random fields in the discrete case, as expounded by Pickard [64, 65, 66, 67] and others, is that they can be simulated directly in a single sweep of the index set. Each value X_v is drawn from the conditional distribution given the already generated values $\{X_u : u \prec v\}$.

Similarly, Monte Carlo simulation of directed Markov point processes is much simpler than for their undirected counterparts. Under mild conditions, a directed Markov point process can be obtained from a Poisson process by a random (i.e. data-dependent) multidimensional time change. This can be interpreted to give a simple algorithm for generating a realisation of the desired process in a single sweep of the spatial domain.

Because they are easy to simulate, directed Markov processes have numerous potential uses. They might be used as reference distributions for importance sampling, or as proposal distributions for simulating Markov point processes (either by the rejection method or for Metropolis-Hastings

algorithms). They might serve as approximations to (undirected) Markov point processes in some cases.

It would also be of interest to generalise the partial order $\preceq$ and indeed to allow dynamic directed graphs (partial orders which depend on the configuration $\mathbf{x}$).

Acknowledgements. I thank the referees for helpful advice.

REFERENCES

[1] K. Abend, T.J. Harley, and L.N. Kanal. Classification of binary random patterns. *IEEE Transactions on Information Theory*, IT–11:538–544, 1965.

[2] A. Baddeley, N. Cressie, and G. Nair. Directed markov point processes. Manuscript in preparation, 1999.

[3] A. Baddeley and R. Turner. Practical maximum pseudolikelihood for spatial point patterns. *Australian and New Zealand Journal of Statistics*, 2000. To appear.

[4] A.J. Baddeley. Time-invariance estimating equations. *Bernoulli*. To appear.

[5] A.J. Baddeley, M.N.M. van Lieshout, and J. Møller. Markov properties of cluster processes. *Advances in Applied Probability*, 28:346–355, 1996.

[6] A.J. Baddeley and J. Møller. Nearest-neighbour Markov point processes and random sets. *International Statistical Review*, 57:89–121, 1989.

[7] A.J. Baddeley and M.N.M van Lieshout. Area-interaction point processes. *Ann. Inst. Statist. Math.*, 47:601–619, 1995.

[8] D. Basu and C.A.B. Pereira. Conditional independence in statistics. *Sānkhyā series A*, 45:324–337, 1983.

[9] M. Berman and T.R. Turner. Approximating point process likelihoods with GLIM. *Applied Statistics*, 41:31–38, 1992.

[10] J. Besag. Statistical analysis of non-lattice data. *The Statistician*, 24:179–195, 1975.

[11] J. Besag. Some methods of statistical analysis for spatial data. *Bulletin of the International Statistical Institute*, 47:77 – 92, 1977.

[12] J. Besag, R. Milne, and S. Zachary. Point process limits of lattice processes. *Journal of Applied Probability*, 19:210–216, 1982.

[13] T.C. Brown, B.G. Ivanoff, and N.C. Weber. Poisson convergence in two dimensions with application to row and column exchangeable arrays. *Stochastic Processes and their Applications*, 23:307–318, 1986.

[14] P.J. Cameron. *Combinatorics: topics, techniques, algorithms.* Cambridge University Press, 1994.

[15] Y.C. Chin and A.J. Baddeley. On connected component Markov point processes. *Advances in Applied Probability*, 31:279–282, 1999.

[16] Peter Clifford. Markov random fields in statistics. In *Disorder in physical systems*, pages 19–32. Oxford Univ. Press, New York, 1990.

[17] F. Comets and M. Janžura. A central limit theorem for conditionally centred random fields with an application to Markov fields. *J. Appl. Prob.*, 35:608–621, 1998.

[18] D. R. Cox and V. Isham. *Point processes.* Chapman and Hall, London, 1980.

[19] N. Cressie and J.L. Davidson. Image analysis with Partially Ordered Markov Models. Preprint 94-15, Department of Statistics, Iowa State University, June/October 1994.

[20] N.A.C. Cressie, J. Zhu, A.J. Baddeley, and M.G. Nair. Directed Markov point processes as limits of partially ordered markov models. *Methodology and Computing in Applied Probability*, 2000. To appear.

[21] D.J. Daley and D. Vere-Jones. *An introduction to the theory of point processes.* Springer Verlag, New York, 1988.

[22] P.J. Diggle, T. Fiksel, P. Grabarnik, Y. Ogata, D. Stoyan, and M. Tanemura. On parameter estimation for pairwise interaction processes. *International Statistical Review*, 62:99–117, 1994.

[23] T. Fiksel. Estimation of parameterized pair potentials of marked and non-marked Gibbsian point processes. *Elektronische Informationsverarbeitung u. Kybernetika*, 20:270–278, 1984.

[24] T. Fiksel. Estimation of interaction potentials of Gibbsian point processes. *Statistics*, 19:77–86, 1988.

[25] M. Frydenberg. The chain graph Markov property. *Scandinavian Journal of Statistics*, 17:333–353, 1990.

[26] Charles J. Geyer and Jesper Møller. Simulation procedures and likelihood inference for spatial point processes. *Scand. J. Statist.*, 21(4):359–373, 1994.

[27] C.J. Geyer. Likelihood inference for spatial point processes. In O.E. Barndorff-Nielsen, W.S. Kendall, and M.N.M. van Lieshout, editors, *Stochastic Geometry: Likelihood and Computation*, number 80 in Monographs on Statistics and Applied Probability, chapter 3, pages 79–140. Chapman and Hall / CRC, Boca Raton, 1999.

[28] E. Glötzl. Bemerkungen zu einer Arbeit von O.K. Kozlov. *Mathematische Nachrichten*, 94:277–289, 1980.

[29] E. Glötzl. Lokale Energien und Potentiale für Punktprozessen. *Mathematische Nachrichten*, 96:195–206, 1980.

[30] G.R. Grimmett. A theorem about random fields. *J. London Math. Soc.*, 5:81–84, 1973.

[31] O. Häggström, M.N.M. van Lieshout, and J. Møller. Characterisation results and Markov Chain Monte Carlo algorithms including exact simulation for some spatial point processes. *Bernoulli*, 1998. To appear.

[32] J.M. Hammersley and D.C. Handscomb. *Monte Carlo methods.* Methuen, London, 1964.

[33] B.G. Ivanoff and E. Merzbach. Intensity-based inference for planar Poisson processes. *Journal of Multivariate Analysis*, 32:269–281, 1990.

[34] J.L. Jensen and H.R. Künsch. On asymptotic normality of pseudo likelihood estimates for pairwise interaction processes. *Annals of the Institute of Statistical Mathematics*, 46:475–486, 1994.

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

[36] O. Kallenberg. *Random measures.* Akademie Verlag/Academic Press, Berlin/New York, third edition, 1983.

[37] O. Kallenberg. An informal guide to the theory of conditioning in point processes. *International Statistical Review*, 52:151–164, 1984.

[38] F.P. Kelly and B.D. Ripley. On Strauss's model for clustering. *Biometrika*, 63:357–360, 1976.

[39] W.S. Kendall. A spatial Markov property for nearest-neighbour Markov point processes. *Journal of Applied Probability*, 28:767–778, 1990.

[40] W.S. Kendall. Perfect simulation for the area-interaction point process. Technical Report 292, Department of Statistics, University of Warwick, Coventry, 1996.

[41] W.S. Kendall and J. Møller. Perfect metropolis-hastings simulation of locally stable point processes. Research Report 347, Department of Statistics, University of Warwick, 1999.

[42] H. Kiiveri, T.P. Speed, and J.B. Carlin. Recursive causal models. *Journal of the Australian Mathematical Society, series A*, 36:30–52, 1984.

[43] J.F.C. Kingman. *Poisson processes.* Oxford University Press, 1993.

[44] O.K. Kozlov. Gibbsian description of point random fields. *Theory of probability and its applications*, 21:339–355, 1976.

[45] S.L. Lauritzen, A.P. Dawid, B.N. Larsen, and H.-G. Leimer. Independence properties of directed Markov fields. *Networks*, 20:491–505, 1990.

[46] S.L. Lauritzen and D.J. Spiegelhalter. Local computations with probabilities on graphical structures and their applications to expert systems (with discussion). *Journal of the Royal Statistical Society, series B*, 50:157–224, 1988.

[47] J. Lee, M.S. Kaiser, and N. Cressie. Multiway dependence in exponential family conditional distributions. Preprint 96–17, Department of Statistics, Iowa State University, July 1996.

[48] J.K. Lindsey. *The analysis of stochastic processes using GLIM.* Springer, Berlin, 1992.

[49] J.K. Lindsey. Fitting parametric counting processes by using linear models. *Applied Statistics*, 44:201–212, 1995.

[50] J.K. Lindsey. Fitting bivariate intensity functions, with an application to modelling delays in reporting acquired immune deficiency syndrome. *Journal of the Royal Statistical Society, Series A*, 159:125–131, 1996.

[51] J.K. Lindsey and G. Mersch. Fitting and comparing probability distributions with log linear models. *Computational Statistics and Data Analysis*, 13:373–384, 1992.

[52] C. Liu and N. Cressie. On the equivalence between symmetric Ising models and some partially ordered Markov models for binary images. Preprint 95–9, Department of Statistics, Iowa State University, April 1995.

[53] P. McCullagh and J.A. Nelder. *Generalized Linear Models.* Chapman and Hall, second edition, 1989.

[54] K. Mecke. *Integralgeometrie in der Statistischen Physik.* Number 25 in Reihe Physik. Verlag Harri Deutsch, 1993.

[55] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equations of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087–1092, 1953.

[56] R.A. Moyeed and A.J. Baddeley. Stochastic approximation for the MLE of a spatial point process. *Scandinavian Journal of Statistics*, 18:39–50, 1991.

[57] Y. Ogata and M. Tanemura. Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure. *Annals of the Institute of Statistical Mathematics*, B 33:315–338, 1981.

[58] Y. Ogata and M. Tanemura. Likelihood analysis of spatial point patterns. *Journal of the Royal Statistical Society, series B*, 46:496–518, 1984.

[59] Y. Ogata and M. Tanemura. Estimation of interaction potentials of marked spatial point processes through the maximum likelihood method. *Biometrics*, 41:421–433, 1985.

[60] Y. Ogata and M. Tanemura. Likelihood estimation of interaction potentials and external fields of inhomogeneous spatial point patterns. In I.S. Francis, B.J.F. Manly, and F.C. Lam, editors, *Pacific Statistical Congress*, pages 150–154. Elsevier, 1986.

[61] A. Okabe, B. Boots, and K. Sugihara. *Spatial tessellations: concepts and applications of Voronoi diagrams.* Wiley series in probability and mathematical statistics. John Wiley and Sons, Chichester, England; New York, 1992.

[62] J.K. Ord. Contribution to discussion of the paper by b.d. ripley. *Journal of the Royal Statistical Society, series B*, 39, 1997.

[63] A. Penttinen. *Modelling interaction in spatial point patterns: parameter estimation by the maximum likelihood method.* Number 7 in Jyväskylä

Studies in Computer Science, Economics and Statistics. University of Jyväskylä, 1984.

[64] D.K. Pickard. A curious binary lattice process. *Journal of Applied Probability*, 14:717–731, 1977.

[65] D.K. Pickard. Unilateral Markov fields. *Advances in Applied Probability*, 12:655–671, 1980.

[66] D.K. Pickard. Statistical inference on discrete Markov fields. *Advances in Applied Probability*, 17:253–254, 1985.

[67] D.K. Pickard. Inference for discrete Markov random fields: the simplest non-trivial case. *Journal of the American Statistical Association*, 82:90–96, 1987.

[68] C.J. Preston. *Random fields.* Springer Verlag, Berlin-Heidelberg-New York, 1976.

[69] C.J. Preston. Spatial birth-and-death processes. *Bulletin of the International Statistical Institute*, 46:371 - 391, 1977.

[70] J.G. Propp and D.B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9:223–252, 1996.

[71] R.-D. Reiss. *A Course on Point Processes.* Springer-Verlag, New York-Berlin-Heidelberg, 1993.

[72] B.D. Ripley. Simulating spatial patterns: dependent samples from a multivariate density. *Applied Statistics*, 28:109–112, 1979.

[73] B.D. Ripley. *Stochastic simulation.* John Wiley and Sons, New York, 1987.

[74] B.D. Ripley. *Statistical inference for spatial processes.* Cambridge University Press, 1988.

[75] B.D. Ripley. Gibbsian interaction models. In D.A. Griffiths, editor, *Spatial statistics: past, present and future*, pages 1–19. Image, New York, 1989.

[76] B.D. Ripley and F.P. Kelly. Markov point processes. *Journal of the London Mathematical Society*, 15:188–192, 1977.

[77] D. Ruelle. *Statistical mechanics.* John Wiley and Sons, New York, 1969.

[78] A. Särkkä. *Modelling interaction in spatial point patterns: parameter estimation by the maximum likelihood method.* Number 22 in Jyväskylä Studies in Computer Science, Economics and Statistics. University of Jyväskylä, 1993.

[79] A. Särkkä and A. Baddeley. Parameter estimation of area-interaction processes. Manuscript in preparation, 1999.

[80] Robin Sibson. The Dirichlet tessellation as an aid in data analysis. *Scand. J. Statist.*, 7(1):14–20, 1980.

[81] D. Stoyan, W.S. Kendall, and J. Mecke. *Stochastic Geometry and its Applications.* John Wiley and Sons, Chichester, second edition, 1995.

[82] D.J. Strauss. A model for clustering. *Biometrika*, 63:467–475, 1975.

[83] R. Takacs. Estimator for the pair potential of a Gibbsian point process. *Statistics*, 17:429–433, 1986.

[84] Håkon Tjelmeland and Julian Besag. Markov random fields with higher-order interactions. *Scand. J. Statist.*, 25(3):415–433, 1998.

[85] B. Widom and J.S. Rowlinson. A new model for the study of liquid-vapor phase transitions. *Journal of Chemical Physics*, 52:1670–1684, 1970.

Department of Mathematics and Statistics
University of Western Australia
Nedlands WA 6907
Australia
adrian@maths.uwa.edu.au