Biometrika (2013), **xx**, x, *pp*. 1–16

Printed in Great Britain

Advance Access publication on xx xxxx 2013

Logistic regression for spatial Gibbs point processes

BY ADRIAN BADDELEY

School of Earth and Environment, University of Western Australia, 35 Stirling Highway, Nedlands Western Australia 6009, Australia

adrian.baddeley@uwa.edu.au

JEAN-FRANÇOIS COEURJOLLY

Laboratory Jean Kuntzmann, Grenoble Alpes University, 51 Rue des Mathmatiques, BP 53 - 38041 Grenoble Cedex, France

Jean-Francois.Coeurjolly@upmf-grenoble.fr

EGE RUBAK, RASMUS WAAGEPETERSEN

Department of Mathematical Sciences, Aalborg University, Fr. Bajers Vej 7G, DK-9220 Aalborg East, Denmark

rubak@math.aau.dk rw@math.aau.dk

SUMMARY

We propose a computationally efficient technique, based on logistic regression, for fitting Gibbs point process models to spatial point pattern data. The score of the logistic regression is an unbiased estimating function, and is closely related to the pseudolikelihood score. Implementation of our technique does not require numerical quadrature, and thus avoids a source of bias inherent in other methods. For stationary processes we prove that the parameter estimator is strongly consistent and asymptotically normal, and propose a variance estimator. We demonstrate the efficiency and practicability of the method on a real dataset and in a simulation study.

Some key words: estimating function, exponential family model, Georgii–Nguyen–Zessin formula, logistic regression, pseudolikelihood.

1. INTRODUCTION

Spatial Gibbs and Markov point processes are important classes of models for spatial dependence in point patterns (van Lieshout, 2000) with a broad range of applications (e.g. Harkness & Isham, 1983; Stoyan & Penttinen, 2000; Mattfeldt et al., 2007; Funwi-Gabga & Mateu, 2012). Popular options for parameter estimation include maximum likelihood (e.g. Ogata & Tanemura, 1981; Møller & Waagepetersen, 2004), maximum pseudolikelihood (e.g. Besag, 1977; Jensen & Møller, 1991; Baddeley & Turner, 2000; Billiot et al., 2008) and Takacs–Fiksel (e.g. Fiksel, 1984; Billiot, 1997; Coeurjolly et al., 2012) estimation. For all three methods, the associated estimating functions are unbiased.

However, practical implementations of these methods are typically biased, because the objective function or estimating function must be approximated. In the likelihood function the normalizing constant is not tractable and is typically approximated by stochastic methods such as Markov chain Monte Carlo simulation (Ripley, 1979; Huang & Ogata, 1999; Geyer, 1999; Møller & Waagepetersen, 2004). The score of the pseudolikelihood, and the Takacs–Fiksel estimating

function, involve an integral which must usually be approximated using numerical quadrature, which may introduce substantial bias.

The maximum pseudolikelihood and Takacs–Fiksel methods offer enormous savings in computation time over Markov chain Monte Carlo maximum likelihood estimation. Another advantage is that maximum pseudolikelihood can be implemented using standard software for generalized linear models, with the attendant benefits of numerical stability, computationally efficient optimisation, and flexible model specification.

One strategy for numerical approximation is to discretise the spatial domain onto a fine grid of pixels (Tukey, 1972) and to consider the random field of binary variables indicating presence or absence of points in each pixel. This is used extensively in Geographical Information Systems to fit spatial Poisson process models (Agterberg, 1974; Bonham-Carter, 1995; Baddeley et al., 2010; Warton & Shepherd, 2010). The discrete approximation to the Poisson process likelihood is a binomial regression with binary responses given by the presence/absence variables, which can be fitted using standard software. Approximation error can be controlled using a fine discretisation, but this leads to numerical instability and the failure of the delta-method approximation (Hauck & Donner, 1977), arising because the overwhelming majority of pixels do not contain a data point. In practice, this is avoided by using only a randomly-selected subset of the absence pixels. The pixel discretisation approach can be extended to form a pseudolikelihood for Gibbs processes (Clyde & Strauss, 1991), although this has not been widely used in practice. For a given choice of grid the binary random field pseudolikelihood again takes the form of a logistic regression likelihood. The spatial point process pseudolikelihood function may be viewed as a limit of binary random field pseudolikelihood functions (Besag, 1975, 1977; Besag et al., 1982; Clyde & Strauss, 1991).

Another popular strategy for numerical approximation is the sparse quadrature approximation pioneered by Berman & Turner (1992) for maximum likelihood estimation of spatial Poisson processes and extended to maximum pseudolikelihood estimation of Gibbs processes by Baddeley & Turner (2000). The approximate pseudolikelihood is equivalent to a Poisson regression likelihood which can be implemented using standard generalized linear model software. The sparse quadrature approximation involves a sum over the observed data points together with a set of dummy points. While it was originally envisaged that the dummy points might be generated at random (Berman & Turner, 1992; Baddeley & Turner, 2000), the standard software implementation in the `spatstat` package (Baddeley & Turner, 2005) generates a regular grid of them if none are provided by the user.

When unbiased estimating functions are approximated using deterministic numerical approximations, the resulting estimating functions are not in general unbiased, and it may be difficult to quantify the error due to the approximations. It can therefore be advantageous to replace deterministic numerical quadrature with Monte Carlo approximations which can provide both unbiased results and the possibility of quantifying the Monte Carlo error. Rathbun et al. (2007) and Waagepetersen (2007) introduced Monte Carlo approximation based on random dummy points for maximum likelihood estimation of Poisson processes and composite likelihood for Neyman–Scott point processes, respectively. The estimating function in Waagepetersen (2007), obtained with weights determined by the Dirichlet tessellation (Baddeley & Turner, 2000), takes the form of a conditional logistic regression, equivalent to the case-control conditional likelihood considered for epidemiological data in Diggle & Rowlingson (1994), and closely related to logistic regression in Geographical Information Systems where the absence pixels are subsampled (Bonham-Carter, 1995).

In this paper we introduce a logistic regression estimating function for the wide class of Gibbs point processes. This estimating function has several advantages. First, it is unbiased. Second,

since it takes the form of a logistic regression score, parameter estimates can easily be obtained using existing software for generalized linear models. Third, due to a decomposition of variance it is possible to quantify the proportion of variance due to using random dummy points, and to decide how many are needed in order to attain the desired accuracy. Fourth, the logistic regression estimating function typically requires fewer evaluations of the conditional intensity than methods based on deterministic numerical quadrature, leading to shorter computing times. It can be further motivated by its close relation to pseudolikelihood and to a time-invariance estimating (Baddeley, 2000) function obtained from Barker dynamics; see the Supplementary Material.

2. SPATIAL POINT PROCESSES

2·1. *General background and notation*

A point process X on $\mathcal{R}^d$ is a random subset of $\mathcal{R}^d$ which is locally finite, meaning that $X \cap W$ is almost surely finite for every bounded subset W of $\mathcal{R}^d$. We assume the point process is confined to a region $\Lambda \subseteq \mathcal{R}^d$; this can be bounded or unbounded depending on the application. The notation W will be reserved for a bounded subset of Λ. For a point pattern x we write $x_W = x \cap W$ for the subset of points falling in W, and $n(x)$ for the possibly infinite number of points in x. Finally, $|\cdot|$ will be used to denote, depending on the context, either the cardinality of a finite set or the volume of a bounded set or the supremum norm of a vector.

We assume X has an intensity function α. Then Campbell's Theorem holds (e.g. Møller & Waagepetersen, 2004):

$$E\left\{\sum_{u\in X} h(u)\right\} = \int_\Lambda h(u)\alpha(u)\,\mathrm{d}u \tag{1}$$

for any real function h defined on Λ such that $h\alpha$ is absolutely integrable.

A point process X has Papangelou conditional intensity $\lambda(u, X)$ if

$$E\left\{\sum_{u\in X} f(u, X\setminus u)\right\} = E\int f(u, X)\lambda(u, X)\mathrm{d}u \tag{2}$$

for all non-negative functions $f : \Lambda \times \Omega \to \mathcal{R}$ where Ω is the set of locally finite point configurations in Λ. Intuitively, $\lambda(u, X)\mathrm{d}u$ is the conditional probability that a point of X occurs in a small ball of volume $\mathrm{d}u$ around the location u, given the rest of the point process X. See Georgii (1976) for a general presentation.

2·2. *Gibbs point processes*

In this paper we consider inference for a finite or infinite Gibbs point process X of exponential family type. Letting Ω_f be the set of finite point configurations in Ω, the distribution of X is specified in terms of a sufficient statistic $t : \Omega_\mathrm{f} \to \mathcal{R}^p$ for some $p \geq 1$, a parameter $\theta \in \Theta \subseteq \mathcal{R}^p$ and a function $H : \Omega_\mathrm{f} \to [0, \infty)$ which serves as a baseline or reference density. Spatial covariates may be included in the components $t_1, \ldots, t_p$, of the sufficient statistic t, while a hard-core distance may be included in the baseline H. If Λ is bounded, X is a finite point process, and the distribution of X is specified in terms of its probability density

$$f(x) \propto H(x)e^{\theta^\top t(x)}$$

with respect to the homogeneous Poisson process on Λ of unit rate.

In the case $\Lambda = \mathcal{R}^d$, the distribution is given by a so-called specification which is a consistent family of conditional point process densities $\{f_W(\cdot \mid \cdot)\}_{W\subseteq\mathcal{R}^d}$ indexed by the bounded subsets

of $\mathcal{R}^d$. For ease of exposition we assume a finite range property, meaning that the conditional density of X_W given $X_{\Lambda\setminus W}$ only depends on the points of $X_{\Lambda\setminus W}$ within distance R from W for some $0 \le R < \infty$. For each W we then define the conditional density as

$$f_W(x_W \mid x_{\Lambda\setminus W}) \propto H(x_{W\oplus R})e^{\theta^\top t(x_{W\oplus R})},$$

where $W \oplus R = \{v \in \mathcal{R}^d : \inf_{u\in W} \|v-u\| \le R\}$. Consistency means that whenever $W_1 \subset W_2$ then $f_{W_1}(\cdot \mid x_{\Lambda\setminus W_1}) \propto f_{W_2}(\cdot \cup x_{W_2\setminus W_1} \mid x_{\Lambda\setminus W_2})$. The distribution of the point process X is specified by the consistent family $\{f_W(\cdot \mid \cdot)\}_{W\subseteq\mathcal{R}^d}$ if $X_W \mid X_{\Lambda\setminus W} = x_{\Lambda\setminus W}$ has density $f_W(\cdot \mid x_{\Lambda\setminus W})$ for each bounded subset W. For general discussion of conditions that ensure the existence of possibly infinite range Gibbs point processes we refer to Preston (1976), Georgii (1988) and Dereudre et al. (2012).

Assuming that H is hereditary, i.e., $H(x\cup u) > 0$ implies $H(x) > 0$ for $u \in \Lambda$ and $x \in \Omega$, the Papangelou conditional intensity of X exists and $\lambda_\theta(u,x)$ is equal to

$$H(u,x)e^{\theta^\top t(u,x)} \ \ (\Lambda \text{ bounded}) \quad \text{or} \quad H(u,x_{\mathcal{B}(u,R)})e^{\theta^\top t(u,x_{\mathcal{B}(u,R)})} \ \ (\Lambda = \mathcal{R}^d), \tag{3}$$

where $H(u,x) = 1\{H(x) > 0\}H(x\cup u)/H(x)$, $t(u,x) = t(x\cup u) - t(x)$ and $\mathcal{B}(u,R)$ is the Euclidean ball centered at u with radius R. Note that λ_θ is in one-to-one correspondence with f and with f_W up to a normalizing constant. Hence the distribution of a Gibbs point process can equivalently and often conveniently be specified in terms of the conditional intensity. Conditions that ensure the existence of a Gibbs point process corresponding to a given conditional intensity are discussed in Section 4 and stated in the Supplementary Material.

2·3. *Extension to marked point processes*

A marked point is a pair $u = (v,m)$ where v and m represent respectively the location and some other characteristic of an object observed in $\mathcal{R}^d$. For example v might be the spatial location of a tree, and m its diameter at breast height. We write $\mathring{u} = v$ for the unmarked point corresponding to the marked point u. Let Λ be a subset of $\mathcal{R}^d$ and let $\mathcal{M}$ be an arbitrary space; e.g., a countable set or a subset of $\mathcal{R}^k$ for some $k \ge 1$. A marked point process Y on $\mathcal{S} = \Lambda \times \mathcal{M}$ is a locally finite random subset of $\Lambda \times \mathcal{M}$ meaning that $Y \cap (W \times \mathcal{M})$ is finite whenever W is a bounded subset of Λ. The notation Ω is used henceforth for the set of all locally finite marked point configurations. For a marked point configuration $y \in \Omega$, write $y_W = y \cap (W \times \mathcal{M})$, that is the subset of marked points where the location part falls in W.

We equip $\Lambda \times \mathcal{M}$ with the product measure $\mathcal{L}^d \otimes \mu$ where $\mathcal{L}^d$ is the Lebesgue measure on $\mathcal{R}^d$ and μ is a probability distribution that serves as the reference measure on $\mathcal{M}$. For simplicity we write $\mathrm{d}u = \mathcal{L}^d(\mathrm{d}\mathring{u}) \otimes \mu(\mathrm{d}m)$ for a marked point $u = (\mathring{u}, m)$. Campbell's formula (1) and the Georgii–Nguyen–Zessin formula (2) continue to hold when integration over $\mathcal{R}^d$ is replaced by integration over $\mathcal{S}$ with respect to $\mathcal{L}^d \otimes \mu$.

Remark 1 *(Multitype point processes).* In the important special case of a multitype point process where $\mathcal{M}$ is a finite set of K elements, say, μ is typically the uniform distribution on $\mathcal{M}$. However, we could also have taken μ to be counting measure in which case the intensity function α_c with respect to Lebesgue-counting product measure would become $\alpha_c(u) = \alpha(u)/K$.

3. UNBIASED ESTIMATING FUNCTION FROM LOGISTIC REGRESSION

Let W denote the bounded observation window of X. Our estimating function involves a dummy point process D on $\mathcal{S}$ independent of X, with positive intensity function ρ. For instance D could be a Poisson, binomial, or stratified binomial point process; see Section 4 for details.

The proposed estimating function is

$$s_W(X, D; \theta) = \sum_{u \in X_W} \frac{\rho(u) t(u, X \setminus u)}{\lambda_\theta(u, X \setminus u) + \rho(u)} - \sum_{u \in D_W} \frac{t(u, X) \lambda_\theta(u, X)}{\lambda_\theta(u, X) + \rho(u)}. \tag{4}$$

By the Georgii–Nguyen–Zessin formula (2) for X and the Campbell formula (1) for D given X, respectively, we obtain

$$E\left\{ \sum_{u \in X_W} \frac{\rho(u) t(u, X \setminus u)}{\lambda_\theta(u, X \setminus u) + \rho(u)} \right\} = E\left\{ \int_{W \times \mathcal{M}} \frac{\rho(u) t(u, X) \lambda_\theta(u, X)}{\lambda_\theta(u, X) + \rho(u)} \mathrm{d}u \right\} \tag{5}$$

and

$$E\Big\{ \sum_{u \in D_W} \frac{t(u, X) \lambda_\theta(u, X)}{\lambda_\theta(u, X) + \rho(u)} \mid X \Big\} = \int_{W \times \mathcal{M}} \frac{\rho(u) t(u, X) \lambda_\theta(u, X)}{\lambda_\theta(u, X) + \rho(u)} \mathrm{d}u. \tag{6}$$

It follows that $s_W(X, D; \theta)$ is an unbiased estimating function where the expectation is taken over both X and D. The score (4) is the derivative of the function $\mathsf{LRL}_W(X; \theta)$, where LRL stands for the logistic log-likelihood

$$\mathsf{LRL}_W(X; \theta) = \sum_{u \in X_W} \log\left\{ \frac{\lambda_\theta(u, X \setminus u)}{\lambda_\theta(u, X \setminus u) + \rho(u)} \right\} + \sum_{u \in D_W} \log\left\{ \frac{\rho(u)}{\lambda_\theta(u, X) + \rho(u)} \right\}. \tag{7}$$

Since $\lambda_\theta(u, X) = \lambda_\theta(u, X \setminus u)$ for $u \notin X$, (7) is, conditional on $X \cup D$, formally equivalent to the log-likelihood function for Bernoulli trials $Y(u) = 1\{u \in X\}$, $u \in X \cup D$ with

$$\mathrm{pr}(Y(u) = 1) = \frac{\lambda_\theta(u, X \setminus u)}{\lambda_\theta(u, X \setminus u) + \rho(u)} = \frac{e^{\theta^\top t(u,X) + \log \widetilde{H}(u,X)}}{1 + e^{\theta^\top t(u,X) + \log \widetilde{H}(u,X)}},$$

where $\widetilde{H}(u, X) = H(u, X)/\rho(u)$. Thus (7) is precisely a logistic regression with offset term $\log \widetilde{H}(u, X)$. This connection has many advantages. Estimation can be implemented straightforwardly using standard software for generalized linear models. The loglikelihood is a concave function of θ, and conditions for existence and uniqueness of the maximum are well known (Silvapulle, 1981).

If we rearrange (4) as

$$s_W(X, D; \theta) = \sum_{u \in X_W} t(u, X \setminus u) - \sum_{u \in (X \cup D)_W} \frac{t(u, X \setminus u) \lambda_\theta(u, X \setminus u)}{\lambda_\theta(u, X \setminus u) + \rho(u)} \tag{8}$$

and apply the Georgii–Nguyen–Zessin formula and (6) to the last term in (8), we obtain

$$E\left\{ \sum_{u \in (X \cup D)_W} \frac{t(u, X \setminus u) \lambda_\theta(u, X \setminus u)}{\lambda_\theta(u, X \setminus u) + \rho(u)} \right\} = E\left\{ \int_{W \times \mathcal{M}} t(u, X) \lambda_\theta(u, X) \mathrm{d}u \right\}. \tag{9}$$

Thus, if the last term in (8) is replaced by its integral compensator the score of the pseudolikelihood is obtained (Jensen & Møller, 1991). Hence our estimating function may be viewed as a Monte Carlo approximation of the pseudolikelihood score to which it converges in mean square when $\inf_{u \in W} \rho(u) \to \infty$.

The estimating equation (4) is applicable both for homogeneous and inhomogeneous Gibbs point processes defined on $\Lambda = W$ and even if $W \subset \Lambda$ with $\Lambda = \mathcal{R}^d$ since (5) and (6) are still valid in these cases. However in such a case the score (4) cannot be computed since it depends on

$X_{\Lambda\setminus W}$, and an edge correction such as the border correction should then be applied. This is done in Section 4 where we focus on stationary Gibbs models and stationary dummy point processes.

4. THEORETICAL RESULTS FOR STATIONARY MODELS

In this section we focus on exponential family models of stationary marked Gibbs point processes that are defined on $\mathcal{S} = \mathcal{R}^d \times \mathcal{M}$, i.e., taking $\Lambda = \mathcal{R}^d$, and we derive asymptotic properties for the logistic regression estimate. We assume that λ_θ has finite interaction range $R \geq 0$ meaning that $\lambda_\theta(u, x) = \lambda_\theta\{u, x_{\mathcal{B}(\mathring{u},R)}\}$. We further assume that X is observed in a sequence of bounded observation windows $W_n^+ \subset \mathcal{R}^d$, $n \geq 1$. Under the assumption of finite range, a logistic regression estimate $\widehat{\theta}_n$ of θ is obtained for each n by maximizing $\mathsf{LRL}_{W_n}(X;\theta)$, where $W_n = W_n^+ \ominus R$ is the erosion of W_n^+ by R: $W_n = \{v \in W_n^+ : \mathcal{B}(v, R) \subseteq W_n^+\}$. Thus we base inference on the conditional distribution of X_{W_n} given $X_{\mathcal{R}^d\setminus W_n}$ where X is a point process on $\mathcal{R}^d$ with conditional intensity λ_θ. This corresponds to using minus sampling to correct for edge effects (Miles, 1974). We assume that $(W_n)_{n\geq 1}$ is a sequence of increasing cubes such that $W_n \to \mathcal{R}^d$ as $n \to \infty$. Finite range together with the further technical assumptions in the Supplementary Material ensure the existence of a marked Gibbs point process on $\mathcal{R}^d$ with the given conditional intensity (Bertin et al., 2008). These conditions are satisfied by a large class of models including the Strauss process, its multiscale and multitype generalizations, Geyer's triplet process, the area-interaction process and Geyer's saturation process.

In the following we consider three different choices of the stationary marked dummy point process D of constant intensity $\rho > 0$. In all cases the marks are assigned independently of the locations of the points according to the reference mark distribution μ. First, for the homogeneous marked Poisson process $\mathcal{P}(\mathcal{R}^d, \rho)$ the locations constitute a homogeneous Poisson process. Second, for the marked binomial point process we assume that $\rho|W_n|$ is integer. The marked binomial point process D_n on W_n then consists of $\rho|W_n|$ independent and identically distributed random marked points with locations uniformly distributed in W_n. In the case of marked binomial dummy points, we abuse notation by writing $D = \cup_{n=1}^{\infty}\{D_n\} \sim \mathcal{B}(\mathcal{R}^d, \rho)$ and $D_{W_n} = D_n$. Finally, the marked stratified point process on $\mathcal{S}$ requires a more detailed definition:

DEFINITION 1. *Let $\mathcal{R}^d$ be decomposed as $\cup_{k\in\mathcal{Z}^d} C_k$ where the cells C_k are disjoint cubes centered at $k/\rho^{1/d}$ with volume $1/\rho$. For $k \in \mathcal{Z}^d$ let $U_k = (\mathring{U}_k, M_k)$ where the random point $\mathring{U}_k$ is uniform on C_k, $M_k \sim \mu$ and all U_k and M_k are independent. Then $D = \cup_{k\in\mathcal{Z}^d}\{U_k\}$ is referred to as a marked stratified binomial point process $\mathcal{SB}(\mathcal{R}^d, \rho)$ on $\mathcal{S}$.*

Let $\theta^\star$ denote the true parameter vector. The score $s_{W_n}(X, D; \theta^\star)$ evaluated at $\theta^\star$ is the sum of $T_{1,W_n}(X)$ and $T_{2,W_n}(X, D)$ where

$$T_{1,W_n}(X) = \sum_{u\in X_{W_n}} w_{\theta^\star}(u, X \setminus u) - \int_{W_n\times\mathcal{M}} w_{\theta^\star}(u, X)\lambda_{\theta^\star}(u, X)\mathrm{d}u, \tag{10}$$

$$T_{2,W_n}(X, D) = \int_{W_n\times\mathcal{M}} w_{\theta^\star}(u, X)\lambda_{\theta^\star}(u, X)\mathrm{d}u - \frac{1}{\rho}\sum_{u\in D_{W_n}} w_{\theta^\star}(u, X)\lambda_{\theta^\star}(u, X), \tag{11}$$

where for any $\theta \in \Theta$, $u \in \mathcal{S}$ and $x \in \Omega$, $w_\theta(u, x) = \rho\, t(u, x)/\{\lambda_\theta(u, x) + \rho\}$. Since T_{1,W_n} is a centered random vector depending only on X and since the expectation of T_{2,W_n} given X is zero, T_{1,W_n} and T_{2,W_n} are uncorrelated and can therefore be studied separately. Each component of the vector $T_{1,W_n}(X)$ is a special case of innovations for spatial Gibbs point processes introduced by Baddeley et al. (2005) with variances studied by Baddeley et al. (2008) and asymptotic results

provided by Coeurjolly & Lavancier (2013) and Coeurjolly & Rubak (2013). Based on these tools, we show in the Supplementary Material that

$$|W_n|^{-1/2} T_{1,W_n}(X) \to\ N(0, G_1) \tag{12}$$

in distribution, where G_1 is defined in Appendix 1.

Regarding the term T_{2,W_n}, conditional on X, a Lindeberg central limit theorem is available. Using this we show in the Supplementary Material that given X,

$$|W_n|^{-1/2} T_{2,W_n}(X, D) \to N(0, G_2) \tag{13}$$

in distribution, where

$$G_2 = \begin{cases} G_2^{\mathrm{p}} = \rho^{-1}\, E\{w_{\theta^\star}^{\lambda}(0^M, X) w_{\theta^\star}^{\lambda}(0^M, X)^\top\}, & D \sim \mathcal{P}(\mathcal{R}^d, \rho), \\ G_2^{\mathrm{b}} = \rho^{-1}\, \mathrm{var}\{w_{\theta^\star}^{\lambda}(0^M, X)\} = \frac{1}{\rho}\, \mathrm{var}\{w_{\theta^\star}^{\lambda}(U_0, X)\}, & D \sim \mathcal{B}(\mathcal{R}^d, \rho), \\ G_2^{\mathrm{sb}} = \rho^{-1}\, E[\mathrm{var}\{w_{\theta^\star}^{\lambda}(U_0, X) \mid X\}], & D \sim \mathcal{SB}(\mathcal{R}^d, \rho), \end{cases} \tag{14}$$

where for $\theta \in \Theta$, $u \in \mathcal{S}$ and $x \in \Omega$ we write $w_{\theta,j}^{\lambda}(u, x)$ for $w_{\theta,j}(u, x)\lambda_\theta(u, x)$, and U_0 is as in Definition 1. Here $0^M = (0, M)$ denotes a randomly-marked point at the origin in $\mathcal{R}^d$, where $M \sim \mu$. We can easily check that $G_2^{\mathrm{sb}} \le G_2^{\mathrm{b}} \le G_2^{\mathrm{p}}$ where for two square matrices A and B, $A \le B$ means that $B - A$ is a positive-semidefinite matrix. Therefore, among the three choices of random dummy points, the marked stratified point process seems to be the optimal choice.

The following almost sure convergence is also proved to hold as $n \to \infty$

$$-|W_n|^{-1} \frac{\mathrm{d}}{\mathrm{d}\theta^\top} s_{W_n}(X, D; \theta^\star) \to S = E\left\{ \frac{\rho t(0^M, X) t(0^M, X)^\top}{\lambda_{\theta^\star}(0^M, X) + \rho} \lambda_{\theta^\star}(0^M, X) \right\} \tag{15}$$

where S is the sensitivity matrix.

Combining (12)-(15) we obtain the following main result, in which we denote by $\widehat{\theta} = \widehat{\theta}_n(X, D)$ the logistic regression score estimate based on $X_{W_n^+}$.

THEOREM 1. *As $n \to \infty$, $\widehat{\theta}$ is a strongly consistent estimate of $\theta^\star$. Assume that G_1 and G_2 are positive-definite matrices, then $|W_n|^{1/2}(\widehat{\theta} - \theta^\star)$ tends to a Gaussian distribution with covariance matrix $\Sigma = S^{-1}(G_1 + G_2)S^{-1}$ which is consistently estimated by $\widehat{\Sigma} = \widehat{S}^{-1}(\widehat{G}_1 + \widehat{G}_2)\widehat{S}^{-1}$ where the matrices $\widehat{S}$, $\widehat{G}_1$ and $\widehat{G}_2$ are defined in Appendices 1 and 2. In other words, $|W_n|^{1/2}\widehat{\Sigma}^{-1/2}(\widehat{\theta} - \theta^\star) \to N(0, I_p)$ in distribution.*

Remark 2 *(Variance decomposition).* Theorem 1 shows in particular that $\mathrm{var}(|W_n|^{1/2}\widehat{\theta})$ is the sum of $\Sigma_1 = S^{-1}G_1S^{-1}$ and $\Sigma_2 = S^{-1}G_2S^{-1}$ where $G_1 = |W_n|^{-1}\,\mathrm{var}\{T_{1,W_n}(X; \theta^\star)\}$ and $G_2 = |W_n|^{-1}\,\mathrm{var}\{T_{2,W_n}(X, D; \theta^\star)\}$. Equations (14)–(15) suggest that Σ_2 is approximately proportional to $1/\rho$. Furthermore, in the simulation studies in Section 5 the estimated Σ_1 is close to the covariance matrix of the maximum pseudolikelihood estimate. We can thus quantify the increase in estimation variance due to the use of the random dummy points D relative to the variance of the exact maximum pseudolikelihood estimate. This also allows us to determine how large a ρ should be used in order to achieve a certain accuracy; see Section 5·5.

Remark 3 *(Ergodicity).* Theorem 1 requires neither the assumption that X be ergodic nor that the Gibbs point process be uniquely determined by the specification.

5. SIMULATION STUDIES AND DATA EXAMPLE

5·1. *Implementation*

Our estimating function is implemented in the `R` package `spatstat` as an option of the function `ppm`. We specify the expected number of dummy points in W using a parameter n_d. By default `ppm` uses a deterministic grid of dummy points where a one-dimensional n_d specifies the number of grid points in each spatial direction. We have implemented the logistic regression estimate as an option for `ppm` and then n_d^2 specifies the expected number of dummy points in case of Poisson or binomial dummy points while n_d specifies the grid dimensions in case of stratified dummy points. Extending the rule of thumb used in `ppm`, we suggest using $\rho = 4n(X_W)/|W|$. In our simulation studies this usually resulted in moderate additional variance due to the random dummy points. Moreover, this choice can be used as a starting point for a data driven approach to determine ρ; see Section 5·5.

5·2. *Comparison of logistic likelihood and pseudolikelihood estimation*

We generate simulations of a stationary unmarked Strauss process in $W = [0,1]^2$ specified by a conditional intensity of the form (3) with $t(u, X) = \{1, n_R(u, X)\}$ and $\theta = (\theta_1, \theta_2)$ where $n_R(u, X)$ is the number of neighbouring points in X of distance from u less than or equal to R. The parameter values used for the simulations are $\theta_1 = \log 1000$, $\theta_2 = \log 0.5$ and $R = 0.01$. The interaction distance R is treated as a known parameter. We generate 10000 simulations of the specified Strauss process and estimate θ using Poisson, binomial or stratified dummy points as well as with default `ppm` and with n_d equal to 10, 20, 40, 80 or 160. The empirical intensity for the simulated patterns is 871. In the particular case of a stationary Strauss process it is also possible to compute the exact maximum pseudolikelihood estimate (Baddeley & Turner, 2000, 2013) which we also consider for comparison.

Boxplots of the parameter estimates for the different estimation methods are shown in Figure 1. The default `ppm` estimate is strongly biased even with $n_d = 80$ while the logistic regression estimate is essentially unbiased for all n_d. With n_d equal to 80 or 160 the variance of the

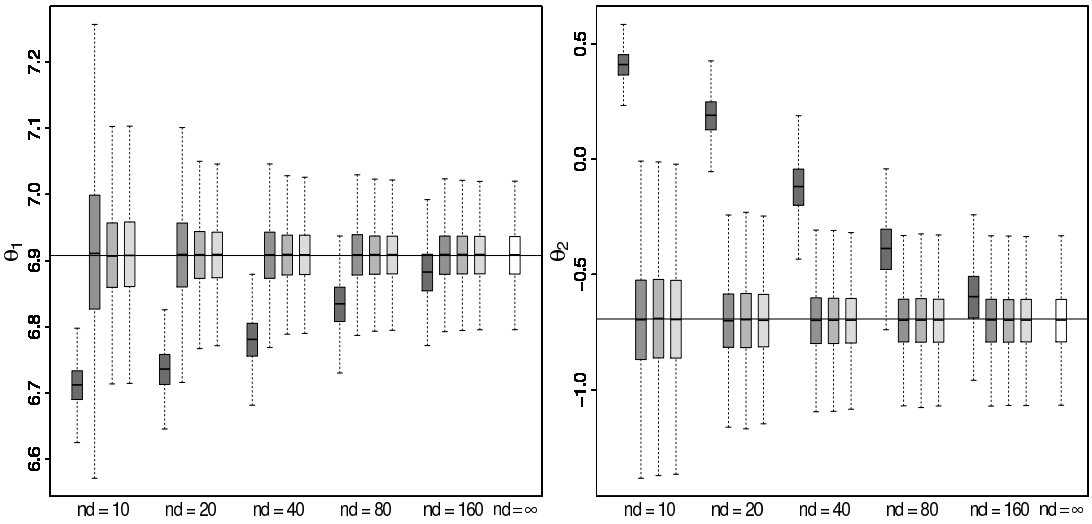

Fig. 1. Boxplots of parameter estimates for increasing values of n_d for the different estimation methods. Horizontal lines show true parameter values. The four greyscales from dark to light represent default ppm and the logistic regression estimates with Poisson, binomial and stratified binomial dummy points. For comparison the exact maximum pseudolikelihood estimate is included as the rightmost box labelled $n_d = \infty$ in both subfigures.

Table 1. *Decomposition of variance for the logistic estimator using stratified dummy points with increasing values of n_d. The columns show the standard deviation of the estimator σ and the two contributions σ_1 and σ_2 as well as the percentage increase of the standard deviation due to random dummy points. For the exact maximum pseudolikelihood estimate the standard deviations are 0.04 and 0.14 for θ_1 and θ_2.*

	θ_1				θ_2			
n_d	σ×100	σ_1×100	σ_2×100	$(\sigma-\sigma_1)/\sigma_1$	σ×100	σ_1×100	σ_2×100	$(\sigma-\sigma_1)/\sigma_1$
10	7	4	6	66.28	25	14	21	79.99
20	5	4	3	19.22	17	14	10	23.21
40	4	4	1	4.35	14	14	5	5.48
80	4	4	0	0.64	14	14	2	0.84
160	4	4	0	0.08	14	14	1	0.11

logistic regression estimate is very close to that of the exact maximum pseudolikelihood estimate. For small values of n_d the variance for default `ppm` is much smaller than for the logistic regression estimate. On the other hand, the table of root mean squared errors provided in the Supplementary Material shows that they are always largest for the `ppm` estimate and that for each n_d, the lowest estimation variance is obtained with stratified dummy points. With n_d equal to 80 or 160 and considering θ_2, the increase in root mean squared error relative to the exact maximum pseudolikelihood estimate is just 0.9 %, respectively 0.07 %, when stratified dummy points are used. In the remainder we only consider stratified dummy points.

As mentioned in Remark 2, the variance of the logistic regression estimator is a sum of terms Σ_1 and Σ_2 where Σ_2 is due to the random dummy points. To investigate this we consider 500 simulations from the Strauss model and for each simulation we refit the model 10 times using independent realizations of the dummy process. A one-way analysis of variance then partitions the total estimation variance into Σ_1 and Σ_2. Results from the analysis of variance are given in Table 1. Here we use the generic notation $\sigma^2 = \sigma_1^2 + \sigma_2^2$ for the variance of a univariate parameter where σ_1^2 and σ_2^2 are extracted from the diagonals of Σ_1 and Σ_2. For n_d greater than or equal to 40 the relative increase in estimation standard deviation $(\sigma-\sigma_1)/\sigma_1$ due to using random dummy points is less than 5.5 %. For both parameters the standard error σ_1 quickly converges to a constant value as n_d increases. The reduction in variance as n_d increases thus mainly occurs for the σ_2^2 term. This justifies regarding σ_2^2 as the increase in variance additional to the maximum pseudolikelihood estimate variance due to the random dummy points. Note also that σ_2 is approximately halved each time n_d is doubled.

The logistic regression estimate outperforms the default ppm method and is competitive with the exact maximum pseudolikelihood estimate for the Strauss process. The same conclusion is valid for more sophisticated examples, such as the ones presented in the next section, for which the exact maximum pseudolikelihood estimate is unavailable.

5·3. *Coverage rates of approximate confidence regions*

In this section we study finite sample coverage properties of approximate confidence intervals based on the asymptotic normality demonstrated in Theorem 1. Simulations are generated from Strauss processes, multiscale Strauss processes, Geyer's saturation processes with saturation threshold 1 and multitype Strauss processes with two types, i.e., $\mathcal{M} = \{1, 2\}$. The two latter unmarked point processes are specified by conditional intensities of the form (3) with respectively $t(u, X) = \{1, n_{R_1}(u, X), n_{R_2}(u, X) - n_{R_1}(u, X)\}$ with $0 \le R_1 \le R_2 < \infty$ and $t(u, X) = [1, \sum_{v\in X\cup u} 1\{d(v, X \cup u) \le R\} - \sum_{v\in X} 1\{d(v, X) \le R\}]$ where $d(v, X)$ denotes

Table 2. *Empirical coverage rates using a nominal level of 95% for the logistic regression estimator with stratified dummy points and increasing values of n_d when W is a square with sidelength ℓ. The first column contains the average empirical intensities for the models. The results are based on 2000 realizations from each of the models.*

		$\ell = 1$			$\ell = 2$		
	$n/\lvert W\rvert$	$n_d = 20$	$n_d = 40$	$n_d = 80$	$n_d = 20$	$n_d = 40$	$n_d = 80$
S1	88	96	94	95	95	94	94
S2	65	95	95	95	96	95	95
M1	53	94	95	95	95	95	96
M2	41	93	93	94	94	95	94
G1	56	95	94	95	94	94	94
G2	45	95	94	95	95	93	95
MS1	74	94	95	95	94	95	94
MS2	74	94	93	93	94	94	94

the distance from v to the nearest point in X without v and with $0 \leq R < \infty$. The conditional intensity of the multitype Strauss point process is also of the form (3) with $t\{(\mathring{u}, m), X\} = [\delta_{m,1}, \delta_{m,2}, n_{R_{11}}\{(\mathring{u}, m), X^1\}\delta_{m1}, n_{R_{22}}\{(\mathring{u}, m), X^2\}\delta_{m2}, n_{R_{12}}\{(\mathring{u}, m), X^2\}\delta_{m1} + n_{R_{12}}\{(\mathring{u}, m), X^1\}\delta_{m2}]$ where δ_{jk} is equal to 1 when $j = k$ and 0 otherwise, X^j consists of the points in X with mark $j \in \{1, 2\}$ and $0 < R_1, R_2, R_{12} < \infty$.

More specifically we consider two Strauss processes with $R = 0.05$ and $\theta_1 = \log 100$, where models S1 and S2 respectively have $\theta_2 = \log 0.8$ and $\theta_2 = \log 0.2$, two multiscale Strauss processes with $R_1 = 0.05$, $R_2 = 0.1$, and $\theta_1 = \log 100$, where models M1 and M2 respectively have $(\theta_2, \theta_3) = (\log 0.2, \log 0.8)$, and $(\theta_2, \theta_3) = (\log 0.8, \log 0.2)$, two Geyer saturation processes with saturation parameter $s = 1$, $R = 0.05$ and $\theta_1 = \log 50$, where models G1 and G2 respectively have $\theta_2 = \log 1.2$ and $\theta_2 = \log 0.8$ and two multitype Strauss processes with $R_1 = R_2 = R_{12} = 0.05$ and $\theta_1 = \theta_2 = \log 50$ where models MS1 and MS2 respectively have $(\theta_3, \theta_4, \theta_5) = (0.5, 0.5, 0.5)$ and $(\theta_3, \theta_4, \theta_5) = (0.8, 0.8, 0.2)$. For all models we use relatively small values of θ_1, as well as θ_2 for MS1 and MS2, to illustrate that the asymptotic results can be applied even for small point patterns.

For all the models the observation window is $W^+ = [-R, \ell + R]^2$, $\ell = 1, 2$, where R is the interaction range of each model equal to R_2 for the multiscale Strauss process and to $\max(R_1, R_2, R_{12})$ for the multitype Strauss process. Due to edge effects, the simulations of X_{W^+} are not realizations of stationary processes. To obtain approximate realizations of stationary processes we use the `spatstat` default settings and simulate a finite process on W^+ expanded by a border of size $2R$ and consider the restriction to W^+. For each simulation we obtain parameter estimates using stratified dummy points with $n_d = 20$, $n_d = 40$ and $n_d = 80$. Subsequently we record whether or not the estimate falls within the approximate 95 % ellipsoidal confidence region $\{\theta : \||W|^{1/2}\widehat{\Sigma}^{-1/2}(\widehat{\theta} - \theta)\|^2 \leq \chi^2_{0.95}(p)\}$. The results given in Table 2 show that the coverage rates are in general close to the nominal 95 % for all the models. Model M2 is one exception where the coverage rates are consistently too low when $\ell = 1$ suggesting that there are too few points to rely on asymptotic results. This agrees with the fact that M2 has the lowest empirical intensity. The estimated Monte Carlo errors are of the order 0.5 %–1 % so the remaining deviations from the nominal 95 % are not worrying. As expected, the closeness to the nominal level does not appear to depend on n_d.

Table 3. *First column: logistic regresssion estimates with $n_d = 60$ of the two polynomials at $y = 0.1, 0.6$ and of the interaction parameter. Remaining columns: estimated standard deviations and relative increases in percent with $n_d = 60$ or $n_d = 120$ based on 1000 simulations from the fitted model. See the text for further details.*

		$n_d = 60$				$n_d = 120$			
	estimate	σ	σ_1	σ_2	$(\sigma - \sigma_1)/\sigma_1$	σ	σ_1	σ_2	$(\sigma - \sigma_1)/\sigma_1$
$q_1(0.1)$	6.00	0.20	0.19	0.05	3.52	0.19	0.19	0.02	0.81
$q_1(0.6)$	3.53	0.75	0.75	0.06	0.35	0.75	0.75	0.03	0.09
$q_2(0.1)$	7.80	0.09	0.08	0.05	16.07	0.08	0.08	0.02	3.80
$q_2(0.6)$	7.14	0.10	0.09	0.04	10.71	0.09	0.09	0.02	2.65
θ_{11}	-2.59	0.34	0.34	0.05	1.05	0.34	0.34	0.02	0.20

5·4. *Data example*

We consider the mucous membrane data shown in Figure 1.3 in Møller & Waagepetersen (2004); our analyses are motivated by Examples 9.3 and 9.5 therein. The dataset used is a subset of the `mucosa` dataset available in `spatstat` and consists of the locations of two types of cells in an observation window $W = [0, 1] \times [0, 0.7]$. There are 87 points of type 1 and 806 points of type 2. We fit an inhomogeneous multitype Strauss process with log conditional intensity $\log \lambda_\theta(u, X) = q_m(y, \theta) + \theta_{11} n_R(u, X)$ where $u = (x, y, m)$, $m = 1$ or 2 denotes the cell type, $q_m(y, \theta)$, $m = 1, 2$, are fourth order polynomials in y with coefficients depending on the type of points and $\theta \in \mathcal{R}^{11}$ consists of the 10 polynomial regression coefficients and the interaction parameter $\theta_{11} \leq 0$. The polynomials only depend on y since the point pattern is assumed to be homogeneous in the x-direction. As before $n_R(u, X)$ denotes the total number of neighbouring points and we use $R = 0.008$ as in Møller & Waagepetersen (2004). One question of interest is whether the conditional intensities of the two types of points share the same large scale polynomial trends.

In this case we use a marked stratified point process, which is generated by sampling uniformly 1 or 2 as marks where each mark is independent of all other variables. In `spatstat` multitype point processes are specified with respect to counting measure on the mark space. To comply with this and following Remark 1 we specify the dummy point intensity as 0.5 times $\rho = n_d^2/0.7$ in our implementation where we for this dataset use $n_d = 60$ according to the rule of thumb. To obtain standard deviations and confidence intervals for the fitted polynomials and the interaction parameter we use a parametric bootstrap based on 1000 simulations generated under the fitted model, where we still use $n_d = 60$ when estimating parameters for each simulation. Furthermore, to enable empirical decomposition of estimation variance we use two replications of the dummy point process for each simulated dataset.

The estimated coefficients of the fourth order polynomials vary considerably but the resulting polynomials do not. We therefore chose to focus on values of the polynomial for six equidistant y values in the range $[0, 0.7]$. Figure 2 shows the estimated polynomials without the constant term as well as bootstrap confidence intervals at the selected set of y values. This plot indicates that the two trends are significantly different.

The first column in Table 3 shows the estimated values of the polynomials for $y = 0.1, 0.6$ and the estimate of the interaction parameter. A more extensive table is provided in the Supplementary Material. The three next columns show the bootstrap estimates of the standard deviation σ of the parameter estimates, the standard deviation σ_1 due to T_1, the standard deviation due to

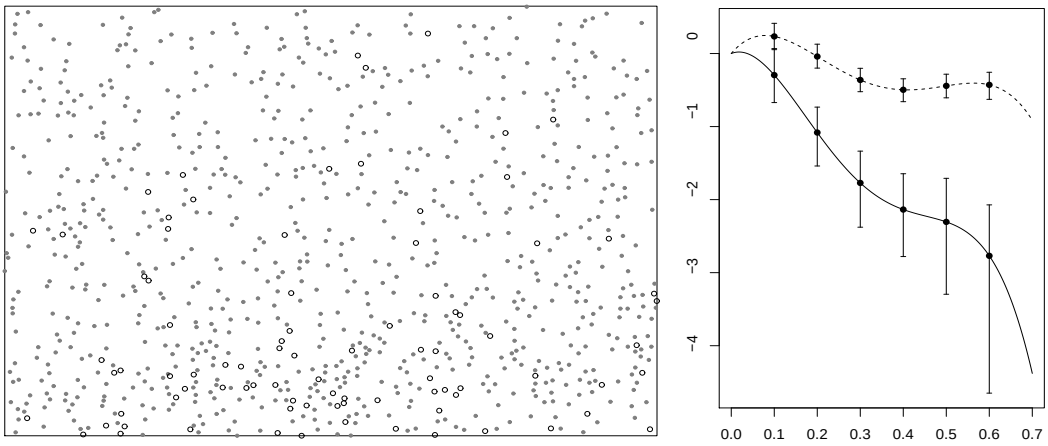

Fig. 2. Left: mucous membrane data with type 1 and type 2 cells indicated by open circles and grey dots, respectively. Right: Fitted fourth order polynomials without the constant term. Bootstrap confidence intervals are given at selected values. Solid and dashed lines correspond to type 1 and type 2 cells, respectively.

T_2 and the relative increase $(\sigma - \sigma_1)/\sigma_1$ in standard deviation due to T_2. Comparing the fitted polynomials, the smallest standard deviations are obtained for the more abundant type 2 cells. For this reason also the largest relative increases, which are at most 16 %, in estimation standard error due to the random dummy points are obtained for the type 2 cells. We also applied $n_d = 120$ and this brings the maximal relative increase in standard deviation down to 4 %, see the four columns labeled $n_d = 120$ in Table 3.

To test the hypothesis of equal polynomials we fitted the null model with common coefficients of the non-constant terms of the fourth order polynomials. We then calculated $-2 \log Q$ for each of 1000 simulations under the fitted null model where Q is the ratio of the likelihoods for the logistic regressions corresponding to the null model and the original model. The 1000 values of this test statistic were between 0.1 and 16.3. The observed value of 28.6 is thus highly significant.

The main drawback of using the default `ppm` is that it is not a priori clear how large n_d must be used to avoid severe bias (Baddeley & Turner, 2013) and sometimes the required value of n_d may even be computationally prohibitive. In the present example we obtained reliable results for the logistic estimator with $n_d = 60$ and the estimation for 1000 datasets took 2 minutes. For default `ppm` we need $n_d = 120$ to avoid strong bias and in this case the 1000 estimations required over 20 minutes of computing time.

5·5. *Data-driven determination of ρ*

As mentioned in Remark 2, the variance σ^2 of a parameter estimate is the sum of a term σ_1^2 which is roughly constant as a function of ρ and a term σ_2^2 which is roughly proportional to $1/\rho$, $\sigma_2^2 = s_2^2/\rho$, say. For a given choice of ρ, for example using the rule of thumb, our asymptotic results provide estimates $\hat{\sigma}_1^2$ and $\hat{s}_2^2$ of these quantities. To find a ρ_p so that the dummy point additional variance s_2^2/ρ_p is less than a specified fraction p of σ_1^2, we may determine ρ_p as $\rho_p = \hat{s}_2^2/(\hat{\sigma}^2 p)$. This relation can also be used to determine $p_\rho = \hat{s}_2^2/(\hat{\sigma}^2 \rho)$ for a given ρ. In practice we may rewrite these relations in terms of standard deviations such that p gives the relative increase of the standard deviation, and thereby of the confidence interval length, due to

random dummy points. We illustrate this approach in the Supplementary Material for the mucous membrane data.

6. FURTHER PERSPECTIVES

Our theoretical results only cover finite range stationary Gibbs point processes but, based on practical examples, e.g., for the mucous membrane data in the Supplementary Material, we believe that similar asymptotic results are also valid for infinite range and non-stationary Gibbs point processes which depend on spatial covariates. Our method requires that the covariates are observed at each random dummy point, but in practice they are often observed on a fixed regular grid, in which case we are not aware of a central limit theorem for the Monte Carlo approximation error. For increasingly fine grids our estimating function nevertheless converges to the pseudolikelihood score, and our confidence intervals for stratified dummy points might provide a conservative assessment of the parameter uncertainty when evaluated using the fixed grid dummy points.

We have assumed a known interaction range R. In practice, R is often estimated by maximizing a profile pseudolikelihood over a grid but the theoretical properties of this procedure are not well studied. The fast computation for our method with moderate n_d is advantageous for evaluating the profile logistic regression likelihood at a large number of R-values.

Acknowledgements

Professor Antonietta Mira drew our attention to the connection with Barker dynamics. This research was conducted when J.-F. Coeurjolly was a Visiting Professor at Aalborg University. Adrian Baddeley was supported by the Australian Research Council and by CSIRO Mathematics, Informatics and Statistics. Ege Rubak and Rasmus Waagepetersen were supported by the Danish Natural Science Research Council, Danish Council for Independent Research Natural Sciences, and the Centre for Stochastic Geometry and Advanced Bioimaging, funded by the Villum Foundation. J.-F. Coeurjolly, Rasmus Waagepetersen and Ege Rubak were also supported by the Institut Français du Danemark.

Supplementary material

Supplementary material available at Biometrika online includes proofs of the results, additional tables and figures related to the data analysis, and explains how the proposed estimating function is related to a time-invariance estimating function obtained from Barker dynamics.

APPENDIX 1

Definitions of G_1 A_1, A_2, A_3 and S and their estimates

The matrix G_1 equals $\sum_{i=1}^{3} A_i(w_{\theta^\star}, w_{\theta^\star})$ where for $i = 1, 2, 3$ the $p \times p$ matrices $A_i(g, h)$ for two functions $g, h : \mathcal{S} \times \Omega \to \mathcal{R}^p$ are given by

$$A_1(g,h) = E\left\{g(0^M, X)h(0^M, X)^\top \lambda_{\theta^\star}(0^M, X)\right\},$$

$$A_2(g,h) = E\left[\int_{\mathcal{B}(0,R)\times\mathcal{M}} g(0^M, X)h(v, X)^\top \{\lambda_{\theta^\star}(0^M, X)\lambda_{\theta^\star}(v, X) - \lambda_{\theta^\star}(\{0^M, v\}, X)\}\mathrm{d}v\right],$$

$$A_3(g,h) = E\left\{\int_{B(0,R)\times\mathcal{M}} \Delta_v g(0^M, X)\Delta_{0^M} h(v, X)^\top \lambda_{\theta^\star}(\{0^M, v\}, X)\mathrm{d}v\right\},$$

where $0^M = (0, M)$ with $M \sim \mu$ and where for $\theta \in \Theta$, $u, v \in \mathcal{S}$,

$$\lambda_\theta(\{u,v\}, X) = \lambda_\theta(u, X \cup v)\lambda_\theta(v, X) = \lambda_\theta(v, X \cup u)\lambda_\theta(u, X)$$
$$\Delta_v g(u, X) = g(u, X \cup v) - g(u, X).$$

From (15) the sensitivity matrix can be rewritten as

$$S = \frac{1}{\rho} A_1\big(w_{\theta^\star}\sqrt{\lambda_{\theta^\star}} + \rho, w_{\theta^\star}\sqrt{\lambda_{\theta^\star}} + \rho\big).$$

We also define for two functions $g, h : \mathcal{S} \times \Omega \to \mathcal{R}^p$ the computationally fast empirical estimates (Coeurjolly & Rubak, 2013) of $A_i(g,h)$ for $i = 1, \ldots, 3$ by

$$\widehat{A}_1(X, D, g, h) = \frac{1}{|W_n|} \sum_{u \in (X \cup D)_{W_n}} g(u, X \setminus u) h(u, X \setminus u)^\top \frac{\lambda_{\widehat{\theta}}(u, X \setminus u)}{\lambda_{\widehat{\theta}}(u, X \setminus u) + \rho}, \tag{A1}$$

$$\widehat{A}_2(X, g, h) = \frac{1}{|W_n|} \sum_{\substack{u,v \in X_{W_n} \\ u \neq v, \|\check{u} - \check{v}\| \leq R}} g(u, X \setminus \{u,v\}) h(v, X \setminus \{u,v\})^\top \times \left\{ \frac{\lambda_{\widehat{\theta}}(u, X \setminus \{u,v\})\lambda_{\widehat{\theta}}(v, X \setminus \{u,v\})}{\lambda_{\widehat{\theta}}(\{u,v\}, X \setminus \{u,v\})} - 1 \right\}, \tag{A2}$$

$$\widehat{A}_3(X, g, h) = \frac{1}{|W_n|} \sum_{\substack{u,v \in X_{W_n} \\ u \neq v, \|\check{u} - \check{v}\| \leq R}} \Delta_v g(u, X \setminus \{u,v\}) \Delta_u h(v, X \setminus \{u,v\})^\top. \tag{A3}$$

The matrices $\widehat{S}$ and $\widehat{G}_1$ are defined by

$$\widehat{S} = \frac{1}{\rho} \widehat{A}_1\big(X, D, w_{\widehat{\theta}}\sqrt{\lambda_{\widehat{\theta}}} + \rho, w_{\widehat{\theta}}\sqrt{\lambda_{\widehat{\theta}}} + \rho\big), \tag{A4}$$

$$\widehat{G}_1 = \widehat{A}_1(X, D, w_{\widehat{\theta}}, w_{\widehat{\theta}}) + \widehat{A}_2(X, w_{\widehat{\theta}}, w_{\widehat{\theta}}) + \widehat{A}_3(X, w_{\widehat{\theta}}, w_{\widehat{\theta}}). \tag{A5}$$

Remark 1 *(On the definition of $\widehat{S}$).* If we had followed the strategy proposed in Coeurjolly & Rubak (2013), the estimate of $A_1(w_{\theta^\star}, w_{\theta^\star})$ would have been based only on X. We include the dummy point pattern D to get a more accurate estimate. No new numerical computations are required since $\widehat{S}$, using (A1), depends only on the quantities $t_j(u, X \setminus u)$ for $j = 1, \ldots, p$ and $u \in (X \cup D)_{W_n}$ which have already been stored when computing the estimate $\widehat{\theta}$. The estimates (A2) and (A3) involve second order characteristics which have not been computed before and are therefore defined using only the data point pattern X.

APPENDIX 2
Definition of $\widehat{G}_2$

According to the dummy point process D considered, the matrix G_2 is consistently estimated as follows:

1. if $D \sim \mathcal{P}(\mathcal{R}^d, \rho)$ the estimate $\widehat{G}_2^{\mathrm{p}}$ is defined by

$$\frac{1}{\rho} \widehat{A}_1(X, D, w_{\widehat{\theta}}\sqrt{\lambda_{\widehat{\theta}}}, w_{\widehat{\theta}}\sqrt{\lambda_{\widehat{\theta}}}); \tag{B1}$$

2. if $D \sim \mathcal{B}(\mathcal{R}^d, \rho)$ the estimate $\widehat{G}_2^{\mathrm{b}}$ is defined by

$$\frac{1}{\rho}\left\{ \kappa_n \widehat{A}_1(X, D, w_{\widehat{\theta}}\sqrt{\lambda_{\widehat{\theta}}}, w_{\widehat{\theta}}\sqrt{\lambda_{\widehat{\theta}}}) - \widehat{A}_1(X, D, w_{\widehat{\theta}}\sqrt{\lambda_{\widehat{\theta}}}, \sqrt{\lambda_{\widehat{\theta}}}) \widehat{A}_1(X, D, w_{\widehat{\theta}}\sqrt{\lambda_{\widehat{\theta}}}, \sqrt{\lambda_{\widehat{\theta}}})^T \right\} \tag{B2}$$

where $\kappa_n = |W_n|^{-1} \sum_{u \in (X \cup D)_{W_n}} (\lambda_{\widehat{\theta}}(u, X \setminus u) + \rho)^{-1}$;

3. if $D \sim \mathcal{SB}(\mathcal{R}^d, \rho)$ the estimate $\widehat{G}_2^{\mathrm{sb}}$ is defined by

$$\frac{1}{2\rho^2 |W_n|} \sum_{\substack{\ell \in \mathcal{Z}^d : \\ C_\ell \cap W_n \neq \emptyset}} \left\{ w_{\widehat{\theta}}^{\lambda}(U_\ell, X) - w_{\widehat{\theta}}^{\lambda}(U'_\ell, X) \right\} \left\{ w_{\widehat{\theta}}^{\lambda}(U_\ell, X) - w_{\widehat{\theta}}^{\lambda}(U'_\ell, X) \right\}^{\top} \tag{B3}$$

where $D' = \cup_{k \in \mathcal{Z}^d} \{U'_k\}$ is a marked stratified point process independent of D.

Remark 1. The variable κ_n in (B2) converges to 1 as $n \to \infty$. It has been introduced to ensure that the estimate $\widehat{G}_2^{\mathrm{b}}$ is a positive-semidefinite matrix. By definition, $\widehat{G}_2^{\mathrm{p}}$ and $\widehat{G}_2^{\mathrm{sb}}$ also fulfill this property.

Remark 2. Following Remark 1 of Appendix 1, we have proposed, in the Poisson and binomial cases, estimates of G_2 based on $X \cup D$. As for $\widehat{S}$, the estimates $\widehat{G}_2^{\mathrm{p}}$ and $\widehat{G}_2^{\mathrm{b}}$ do not involve new numerical computations. The estimate of G_2^{sb} is more awkward to handle and requires an extra dummy point process D'. As pairs of points are involved in (B3), we could not include the data points without adding second order characteristics computations for X and this has not been investigated.

REFERENCES

AGTERBERG, F. P. (1974). Automatic contouring of geological maps to detect target areas for mineral exploration. *Journal of the International Association for Mathematical Geology* **6**, 373–395.

BADDELEY, A. (2000). Time-invariance estimating equations. *Bernoulli* **6**, 783–808.

BADDELEY, A., BERMAN, M., FISHER, N. I., HARDEGEN, A., MILNE, R. K., SCHUHMACHER, D., SHAH, R. & TURNER, R. (2010). Spatial logistic regression and change-of-support for Poisson point processes. *Electronic Journal of Statistics* **4**, 1151–1201.

BADDELEY, A., MØLLER, J. & PAKES, A. G. (2008). Properties of residuals for spatial point processes. *Annals of the Institute of Statistical Mathematics* **60**, 627–649.

BADDELEY, A. & TURNER, R. (2000). Practical maximum pseudolikelihood for spatial point patterns (with discussion). *Australian and New Zealand Journal of Statistics* **42**, 283–322.

BADDELEY, A. & TURNER, R. (2005). Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software* **12**, 1–42.

BADDELEY, A. & TURNER, R. (2013). Bias correction for parameter estimates of spatial point process models. *Journal of Statistical Computation & Simulation* In Press.

BADDELEY, A. J., TURNER, R., MØLLER, J. & HAZELTON, M. (2005). Residual analysis for spatial point processes (with discussion). *Journal of the Royal Statistical Society, Series B* **67**, 617–666.

BERMAN, M. & TURNER, R. (1992). Approximating point process likelihoods with GLIM. *Applied Statistics* **41**, 31–38.

BERTIN, E., BILLIOT, J.-M. & DROUILHET, R. (2008). R-local Delaunay inhibition model. *Journal of Statistical Physics* **132**, 649–667.

BESAG, J. (1977). Some methods of statistical analysis for spatial data. *Bulletin of the International Statistical Institute* **47**, 77–92.

BESAG, J., MILNE, R. K. & ZACHARY, S. (1982). Point process limits of lattice processes. *Journal of Applied Probability* **19**, 210–216.

BESAG, J. E. (1975). Statistical analysis of non-lattice data. *The Statistician* **24**, 179–195.

BILLIOT, J.-M. (1997). Asymptotic properties of Takacs-Fiksel estimation method for Gibbs point processes. *Statistics* **30**, 68–89.

BILLIOT, J.-M., COEURJOLLY, J.-F. & DROUILHET, R. (2008). Maximum pseudolikelihood estimator for exponential family models of marked Gibbs point processes. *Electronic Journal of Statistics* **2**, 234–264.

BONHAM-CARTER, G. (1995). *Geographic Information Systems for geoscientists: modelling with GIS*. No. 13 in Computer Methods in the Geosciences. Kidlington, Oxford, UK: Pergamon Press/ Elsevier.

CLYDE, M. & STRAUSS, D. (1991). Logistic regression for spatial pair-potential models. In *Spatial Statistics and Imaging*, A. Possolo, ed., vol. 20 of *Lecture Notes - Monograph series*, chap. II. Institute of Mathematical Statistics, pp. 14–30. ISBN 0-940600-27-7.

COEURJOLLY, J.-F., DEREUDRE, D., DROUILHET, R. & LAVANCIER, F. (2012). Takacs-Fiksel method for stationary marked Gibbs point processes. *Scandinavian Journal of Statistics* **49**, 416–443.

COEURJOLLY, J.-F. & LAVANCIER, F. (2013). Residuals for stationary marked Gibbs point processes. *Journal of the Royal Statistical Society, Series B* **75**, 247–276.

COEURJOLLY, J.-F. & RUBAK, E. (2013). Fast covariance estimation for innovations computed from a spatial Gibbs point process. *To appear in Scandinavian Journal of Statistics* .

DEREUDRE, D., DROUILHET, R. & GEORGII, H. O. (2012). Existence of Gibbsian point processes with geometry-dependent interactions. *Probability Theory and Related Fields* **153**, 643–670.

DIGGLE, P. J. & ROWLINGSON, B. (1994). A conditional approach to point process modelling of elevated risk. *Journal of the Royal Statistical Society, series A (Statistics in Society)* **157**, 433–440.

FIKSEL, T. (1984). Estimation of parameterized pair potentials of marked and nonmarked Gibbsian point processes. *Elektronische Informationsverarbeitung und Kybernetik* **20**, 270–278.

FUNWI-GABGA, N. & MATEU, J. (2012). Understanding the nesting spatial behaviour of gorillas in the Kagwene Sanctuary, Cameroon. *Stochastic Environmental Research and Risk Assessment* **26**, 793–811.

GEORGII, H. O. (1976). Canonical and grand canonical Gibbs states for continuum systems. *Communications in Mathematical Physics* **48**, 31–51.

GEORGII, H.-O. (1988). *Gibbs Measures and Phase Transitions*. Berlin: Walter de Gruyter.

GEYER, C. J. (1999). Likelihood inference for spatial point processes. In *Stochastic Geometry: Likelihood and Computation*, O. E. Barndorff-Nielsen, W. S. Kendall & M. N. M. van Lieshout, eds., no. 80 in Monographs on Statistics and Applied Probability, chap. 3. Boca Raton, Florida: Chapman and Hall/CRC, pp. 79–140.

HARKNESS, R. & ISHAM, V. (1983). A bivariate spatial point pattern of ants' nests. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **32**, pp. 293–303.

HAUCK, JR., W. & DONNER, A. (1977). Wald's test as applied to hypotheses in logit analysis. *Journal of the American Statistical Association* **72**, 851–853.

HUANG, F. & OGATA, Y. (1999). Improvements of the maximum pseudo-likelihood estimators in various spatial statistical models. *Journal of Computational and Graphical Statistics* **8**, 510–530.

JENSEN, J. L. & MØLLER, J. (1991). Pseudolikelihood for exponential family models of spatial point processes. *Annals of Applied Probability* **1**, 445–461.

MATTFELDT, T., ECKEL, S., FLEISCHER, F. & SCHMIDT, V. (2007). Statistical modelling of the geometry of planar sections of prostatic capillaries on the basis of stationary strauss hard-core processes. *Journal of Microscopy* **228**, 272–281.

MILES, R. E. (1974). On the elimination of edge effects in planar sampling. In *Stochastic Geometry: A Tribute to the Memory of Rollo Davidson*, E. F. Harding & D. G. Kendall, eds. Chichester: John Wiley and Sons, Ltd, pp. 228–47.

MØLLER, J. & WAAGEPETERSEN, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Boca Raton: Chapman and Hall/CRC.

OGATA, Y. & TANEMURA, M. (1981). Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure. *Annals of the Institute of Statistical Mathematics* **33**, 315–338.

PRESTON, C. (1976). *Random fields*. No. 534 in Lecture Notes in Mathematics. Berlin: Springer-Verlag.

RATHBUN, S. L., SHIFFMAN, S. & GWALTNEY, C. J. (2007). Modelling the effects of partially observed covariates on Poisson process intensity. *Biometrika* **94**, 153–165.

RIPLEY, B. D. (1979). Simulating spatial patterns: dependent samples from a multivariate density. *Applied Statistics* **28**, 109–112.

SILVAPULLE, M. (1981). On the existence and uniqueness of the maximum likelihood estimates for the binomial response models. *Journal of the Royal Statistical Society, series B* **43**, 310–313.

STOYAN, D. & PENTTINEN, A. (2000). Recent applications of point process methods in forestry statistics. *Statistical Science* **15**, 61–78.

TUKEY, J. W. (1972). Discussion of paper by F. P. Agterberg and S. C. Robinson. *Bulletin of the International Statistical Institute* **44**, 596. Proceedings, 38th Congress, International Statistical Institute.

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

WAAGEPETERSEN, R. (2007). Estimating functions for inhomogeneous spatial point processes with incomplete covariate data. *Biometrika* **95**, 351–363.

WARTON, D. I. & SHEPHERD, L. C. (2010). Poisson point process models solve the "pseudo-absence problem" for presence-only data in ecology. *Annals of Applied Statistics* **4**, 1383–1402.

[*Received xxxx* 2013. *Revised xxxx* 2013]