Logistic regression for spatial Gibbs point processes

BY ADRIAN BADDELEY
School of Earth and Environment, University of Western Australia, 35 Stirling Highway, Crawley, Perth, Western Australia 6009, Australia
adrian.baddeley@uwa.edu.au

JEAN-FRANÇOIS COEURJOLLY
Laboratory Jean Kuntzmann, Grenoble Alpes University, 51 Rue des Mathématiques, BP 53, 38041 Grenoble Cedex, France
Jean-Francois.Coeurjolly@upmf-grenoble.fr

EGE RUBAK AND RASMUS WAAGEPETERSEN
Department of Mathematical Sciences, Aalborg University, Fredrik 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; Möller & Waagepetersen, 2004), maximum pseudolikelihood (e.g., Besag, 1977; Jensen & Möller, 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
as Markov chain Monte Carlo simulation (Ripley, 1979; Geyer, 1999; Huang & Ogata, 1999; Møller & Waagepetersen, 2004). The score of the pseudolikelihood and the Takacs–Fiksel estimating function involve an integral which usually needs to be approximated using numerical quadrature, and this 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 optimization, and flexible model specification.

One strategy for numerical approximation is to discretize the spatial domain with a fine grid of pixels (Tukey, 1972) and consider the random field of binary variables indicating the presence or absence of points in each pixel. This approach 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 discretization, but this leads to numerical instability and failure of the delta-method approximation (Hauck & Donner, 1977), which occurs because the overwhelming majority of pixels do not contain a data point. In practice, this problem is avoided by using only a randomly selected subset of the absence pixels. The pixel discretization approach can be extended to form a pseudolikelihood for Gibbs processes (Clyde & Strauss, 1991), although this has not been widely adopted 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 can 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 would 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 such points 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 (2008) 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 (2008), 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 in Diggle & Rowlingson (1994) for epidemiological data, and is 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, owing to a decomposition of variance, it is possible to quantify the proportion of variance that is due to using random dummy points, and to decide how many such points 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 function (Baddeley, 2000) obtained from Barker dynamics; see the Supplementary Material.

2. Spatial point processes

2.1. General background and notation

A point process \( X \) on \( \mathbb{R}^d \) is a random subset of \( \mathbb{R}^d \) which is locally finite, meaning that \( X \cap W \) is almost surely finite for every bounded subset \( W \) of \( \mathbb{R}^d \). We assume that the point process is confined to a region \( \Lambda \subseteq \mathbb{R}^d \); this can be bounded or unbounded depending on the application. The notation \( W \) will be reserved for a bounded subset of \( \Lambda \). For a point pattern \( x \), we write \( x_W = x \cap W \) for the subset of points of \( x \) 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 that \( X \) has an intensity function \( \alpha \). Then Campbell’s theorem holds (see, e.g., Møller & Waagepetersen, 2004):

\[
E \left\{ \sum_{u \in X} h(u) \right\} = \int_{\Lambda} h(u) \alpha(u) \, du \tag{1}
\]

for any real function \( h \) defined on \( \Lambda \) such that \( h\alpha \) is absolutely integrable.

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

\[
E \left\{ \sum_{u \in X} f(u, X \setminus u) \right\} = E \left\{ \int f(u, X) \lambda(u, X) \, du \right\} \tag{2}
\]

for all nonnegative functions \( f : \Lambda \times \Omega \rightarrow \mathbb{R} \), where \( \Omega \) is the set of locally finite point configurations in \( \Lambda \). Intuitively, \( \lambda(u, X) \, du \) is the conditional probability that a point of \( X \) occurs in a small ball of volume \( du \) 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 \( \Omega_f \) be the set of finite point configurations in \( \Omega \), the distribution of \( X \) is specified in terms of a sufficient statistic \( t : \Omega_f \rightarrow \mathbb{R}^p \) for some \( p \geq 1 \), a parameter \( \theta \in \Theta \subseteq \mathbb{R}^p \) and a function \( H : \Omega_f \rightarrow [0, \infty) \) 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 \( \Lambda \) is bounded, \( X \) is a finite point process, and the
distribution of $X$ is specified in terms of its probability density

$$f(x) \propto H(x) \exp\{\theta^T t(x)\}$$

with respect to the homogeneous Poisson process on $\Lambda$ of unit rate.

In the case where $\Lambda = \mathbb{R}^d$, the distribution is given by a so-called specification, which is a consistent family of conditional point process densities $\{f_W(\cdot | \cdot)\}_{W \subseteq \mathbb{R}^d}$ indexed by the bounded subsets of $\mathbb{R}^d$. For ease of exposition we assume a finite-range property, meaning that the conditional density of $X_W$ given $X_{\Lambda\setminus W}$ depends only on the points of $X_{\Lambda\setminus W}$ within distance $R$ from $W$, for some $0 \leq R < \infty$. For each $W$, we then define the conditional density as

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

where $W \oplus R = \{v \in \mathbb{R}^d : \inf_{u \in W} \|v - u\| \leq R\}$. Consistency means that whenever $W_1 \subset W_2$, one has $f_{W_1}(\cdot | x_{\Lambda\setminus W_1}) \propto f_{W_2}(\cdot \cup x_{\Lambda\setminus W_2} | x_{\Lambda\setminus W_2})$. The distribution of the point process $X$ is specified by the consistent family $\{f_W(\cdot | \cdot)\}_{W \subseteq \mathbb{R}^d}$ if $X_W | X_{\Lambda\setminus W} = x_{\Lambda\setminus W}$ has density $f_W(\cdot | x_{\Lambda\setminus W})$ 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., that $H(x \cup u) > 0$ implies $H(x) > 0$ for $u \in \Lambda$ and $x \in \Omega$, the Papangelou conditional intensity of $X$ exists and $\lambda_H(u, x) = \lambda(u \circ \theta, x)$ is equal to either of

$$H(u, x) \exp\{\theta^T t(u, x)\} \quad (\Lambda \text{ bounded}),$$

$$H(u, x_{B(u, R)}) \exp\{\theta^T t(u, x_{B(u, R)})\} \quad (\Lambda = \mathbb{R}^d),$$

where $H(u, x) = 1[H(x) > 0]H(x \cup u)/H(x), \ t(u, x) = t(x \cup u) - t(x)$ and $B(u, R)$ is the Euclidean ball centred at $u$ with radius $R$. Note that $\lambda_H$ 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 §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 $\mathbb{R}^d$. For example, $v$ might be the spatial location of a tree and $m$ its diameter at chest height. We write $\tilde{u} = v$ for the unmarked point corresponding to the marked point $u$. Let $\Lambda$ be a subset of $\mathbb{R}^d$ and let $\mathcal{M}$ be an arbitrary space, such as a countable set or a subset of $\mathbb{R}^k$ for some $k \geq 1$. A marked point process $Y$ on $S = \Lambda \times \mathcal{M}$ 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 $\Lambda$. The notation $\Omega$ will be used henceforth for the set of all locally finite marked point configurations. For a marked point configuration $y \in \Omega$, write $y_W = y \cap (W \times \mathcal{M})$, 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 $\mathbb{R}^d$ and $\mu$ is a probability distribution that serves as the reference measure on $\mathcal{M}$. For simplicity, we write $du = \mathcal{L}^d(\tilde{u}) \otimes \mu(dm)$ for a marked point $u = (\tilde{u}, m)$. Campbell’s formula (1) and the Georgii–Nguyen–Zessin formula (2) continue to hold when integration over $\mathbb{R}^d$ is replaced by integration over $S$ with respect to $\mathcal{L}^d \otimes \mu$. 


Remark 1 (Multi-type point processes). In the important special case of a multi-type point process where $\mathcal{M}$ is a finite set of $K$ elements, say, $\mu$ is typically the uniform distribution on $\mathcal{M}$. However, we could also have taken $\mu$ to be a counting measure, in which case the intensity function $\alpha_c$ 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 $S$ that is independent of $X$ and has positive intensity function $\rho$. For instance, $D$ could be a Poisson, binomial, or stratified binomial point process; see §4 for details. The proposed estimating function is

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

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

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

and

$$E \left\{ \sum_{u \in D_W} \frac{t(u, X) \lambda_\theta(u, X)}{\lambda_\theta(u, X) + \rho(u)} \right\} = \int_{W \times \mathcal{M}} \frac{\rho(u) t(u, X) \lambda_\theta(u, X)}{\lambda_\theta(u, X) + \rho(u)} \, du. \quad (6)$$

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 $l_{RLW}(X; \theta)$, where $l_{RL}$ stands for the logistic loglikelihood

$$l_{RLW}(X; \theta) = \sum_{u \in X_W} \log \left\{ \frac{\lambda_\theta(u, X \setminus u)}{\lambda_\theta(u, X \setminus u) + \rho(u)} \right\} + \sum_{u \in D_W} \log \left\{ \frac{\rho(u)}{\lambda_\theta(u, X) + \rho(u)} \right\}. \quad (7)$$

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 loglikelihood function for Bernoulli trials, $Y(u) = 1[u \in X]$ for $u \in X \cup D$, with

$$\text{pr}[Y(u) = 1] = \frac{\lambda_\theta(u, X \setminus u)}{\lambda_\theta(u, X \setminus u) + \rho(u)} = \frac{\exp[\theta^T t(u, X) + \log \tilde{H}(u, X)]}{1 + \exp[\theta^T t(u, X) + \log \tilde{H}(u, X)]}$$

where $\tilde{H}(u, X) = H(u, X)/\rho(u)$. Thus (7) is precisely a logistic regression with offset term log $\tilde{H}(u, X)$. This connection has many advantages. Estimation can be implemented straightforwardly using standard software for generalized linear models. The loglikelihood is a concave function of $\theta$, and conditions for existence and uniqueness of the maximum are well known (Silvapulle, 1981).
If we rearrange (4) as

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

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

\[ E \left\{ \sum_{u \in (X \cup D)_W} \frac{t(u, X \setminus u) \lambda_\theta(u, X \setminus u)}{\lambda_\theta(u, X \setminus u) + \rho(u)} \right\} = E \left\{ \int_{W \times \mathcal{M}} t(u, X) \lambda_\theta(u, X) \, du \right\}. \]

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 to both homogeneous and inhomogeneous Gibbs point processes defined on \( \Lambda = W \), even when \( W \subset \Lambda \) with \( \Lambda = \mathbb{R}^d \), since (5) and (6) are still valid in those 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 § 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 \( S = \mathbb{R}^d \times \mathcal{M} \), i.e., taking \( \Lambda = \mathbb{R}^d \), and we derive asymptotic properties for the logistic regression estimate. We assume that \( \lambda_\theta \) has finite interaction range \( R \geq 0 \), meaning that \( \lambda_\theta(u, x) = \lambda_\theta \{ u, x_{B(u, R)} \} \). We further assume that \( X \) is observed in a sequence of bounded observation windows \( W_n^+ \subset \mathbb{R}^d (n \geq 1) \). Under the assumption of finite range, a logistic regression estimate \( \hat{\theta}_n \) of \( \theta \) is obtained for each \( n \) by maximizing \( \ell_{RL, W_n}(X; \theta) \), where \( W_n = W_n^+ \ominus R \) is the erosion of \( W_n^+ \) by \( R \), i.e., \( W_n = \{ v \in W_n^+ : B(v, R) \subseteq W_n^+ \} \). Thus, we base inference on the conditional distribution of \( X_{W_n} \) given \( X_{\mathbb{R}^d \setminus W_n} \), where \( X \) is a point process on \( \mathbb{R}^d \) with conditional intensity \( \lambda_\theta \). 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 \( W_n \to \mathbb{R}^d \) as \( n \to \infty \).

Finite range together with the further technical assumptions stated in the Supplementary Material ensure the existence of a marked Gibbs point process on \( \mathbb{R}^d \) with the given conditional intensity (Bertin et al., 2008). These conditions are satisfied by a large class of models, including the Strauss process, its multi-scale and multi-type 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 \( \mu \). First, for the homogeneous marked Poisson process \( pO(\mathbb{R}^d, \rho) \), the locations constitute a homogeneous Poisson process. Second, for the marked binomial point process, we assume that \( \rho(W_n) \) is integer-valued. 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, with an abuse of notation we write \( D = \bigcup_{n=1}^\infty \{ D_n \} \sim B1(\mathbb{R}^d, \rho) \) and \( D_{W_n} = D_n \).

Finally, the marked stratified point process on \( S \) requires a more detailed definition.
Logistic regression for Gibbs processes

Definition 1. Let \( \mathbb{R}^d \) be decomposed as \( \bigcup_{k \in \mathbb{Z}^d} C_k \), where the cells \( C_k \) are disjoint cubes centred at \( k/\rho^{1/d} \) with volume \( 1/\rho \). For \( k \in \mathbb{Z}^d \), let \( U_k = (\hat{U}_k, M_k) \), where the random point \( \hat{U}_k \) is uniform on \( C_k \), \( M_k \sim \mu \) and all the \( U_k \) and \( M_k \) are independent. Then \( D = \bigcup_{k \in \mathbb{Z}^d} \{ U_k \} \) is referred to as a marked stratified binomial point process \( \text{sb}(\mathbb{R}^d, \rho) \) on \( S \).

Let \( \theta^* \) denote the true parameter vector. The score \( s_{W_n}(X, D; \theta^*) \) evaluated at \( \theta^* \) is the sum of \( T_1, W_n(X) \) and \( T_2, W_n(X, D) \) where

\[
T_1, W_n(X) = \sum_{u \in X_n} w_{0^*}(u, X \setminus u) - \int_{W_n \times \mathcal{M}} w_{0^*}(u, X) \lambda_{\theta^*}(u, X) \, du,
\]

\[
T_2, W_n(X, D) = \int_{W_n \times \mathcal{M}} w_{0^*}(u, X) \lambda_{\theta^*}(u, X) \, du - \frac{1}{\rho} \sum_{u \in D_n} w_{0^*}(u, X) \lambda_{\theta^*}(u, X), \tag{9}
\]

with \( w_{0^*}(u, x) = \rho \cdot t(u, x)/[\lambda_{\theta^*}(u, x) + \rho] \) for any \( \theta \in \Theta, u \in S \) and \( x \in \Omega \). Since \( T_1, W_n \) is a centroided 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 so can be studied separately. Each component of the vector \( T_1, W_n(X) \) is a special case of innovations for spatial Gibbs point processes, which were 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

\[
|W_n|^{-1/2} T_1, W_n(X) \rightarrow N(0, G_1) \tag{10}
\]

in distribution, where \( G_1 \) is defined in the Appendix.

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 \( X \),

\[
|W_n|^{-1/2} T_2, W_n(X, D) \rightarrow N(0, G_2) \tag{11}
\]

in distribution, with

\[
G_2 = \begin{cases} 
G_2^f = \rho^{-1} E[w_{0, \lambda^*}(0^M, X) w_{0^*}(0^M, X)^T], & D \sim \text{po}(\mathbb{R}^d, \rho); \\
G_2^f = \rho^{-1} \var[w_{0, \lambda^*}(0^M, X)] = \frac{1}{\rho} \var[w_{0^*}(U_0, X)], & D \sim \text{bi}(\mathbb{R}^d, \rho); \\
G_2^{sb} = \rho^{-1} E[\var[w_{0, \lambda^*}(U_0, X) \mid X]], & D \sim \text{sb}(\mathbb{R}^d, \rho) 
\end{cases} \tag{12}
\]

where, for \( \theta \in \Theta, u \in S \) and \( x \in \Omega \), we write \( w_{0, \lambda^*}(u, x) \) for \( w_{0^*}(u(x) \lambda_{\theta^*}(u, x) \), and \( U_0 \) is as in Definition 1. Here \( 0^M = (0, M) \) denotes a randomly marked point at the origin in \( \mathbb{R}^d \), where \( M \sim \mu \). We can easily check that \( G_2^{sb} \leq G_2^f \leq G_2^f \) where, for two square matrices \( A \) and \( B, A \leq B \) 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 \( n \rightarrow \infty \):

\[
-|W_n|^{-1} \frac{d}{d\theta^T} s_{W_n}(X, D; \theta^*) \rightarrow S = E \left\{ \frac{\rho t(0^M, X) t(0^M, X)^T}{\lambda_{\theta^*}(0^M, X) + \rho} \lambda_{\theta^*}(0^M, X) \right\}, \tag{13}
\]

where \( S \) is the sensitivity matrix.

Upon combining (10)–(13), we obtain the following main result, in which we denote by \( \hat{\theta} = \hat{\theta}_n(X, D) \) the logistic regression score estimate based on \( X_{W_n}^+ \).
Theorem 1. As \( n \to \infty \), \( \hat{\theta} \) is a strongly consistent estimate of \( \theta^* \). Assume that \( G_1 \) and \( G_2 \) are positive-definite matrices. Then \( |W_n|^{1/2}(\hat{\theta} - \theta^*) \) tends to a Gaussian distribution with covariance matrix \( \Sigma = S^{-1}(G_1 + G_2)S^{-1} \), which is consistently estimated by \( \hat{\Sigma} = \hat{S}^{-1}(\hat{G}_1 + \hat{G}_2)\hat{S}^{-1} \) where the matrices \( \hat{S}, \hat{G}_1 \) and \( \hat{G}_2 \) are as defined in the Appendix; in other words, 

\[
|W_n|^{1/2} \hat{\Sigma}^{-1/2}(\hat{\theta} - \theta^*) \to N(0, \text{I}_p) \text{ in distribution.}
\]

Remark 2 (Variance decomposition). Theorem 1 shows in particular that \( \text{var}(|W_n|^{1/2} \hat{\theta}) \) is the sum of \( \Sigma_1 = S^{-1}G_1S^{-1} \) and \( \Sigma_2 = S^{-1}G_2S^{-1} \) where \( G_1 = |W_n|^{-1} \text{var}(T_{1, W_n}(X; \theta^*)) \) and \( G_2 = |W_n|^{-1} \text{var}(T_{2, W_n}(X, D; \theta^*)) \). Equations (12) and (13) suggest that \( \Sigma_2 \) is approximately proportional to \( 1/\rho \). Furthermore, in the simulation studies of § 5, the estimated \( \Sigma_1 \) 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 \( \rho \) should be used in order to achieve a certain accuracy; see § 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 (R Development Core Team, 2014) 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 of ppm, in which \( n_d^2 \) specifies the expected number of dummy points in the case of Poisson or binomial dummy points while \( n_d \) specifies the grid dimensions in the case of stratified dummy points. Extending the rule of thumb used in ppm, we suggest taking \( \rho = 4n(X_W)/|W| \). 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 determining \( \rho \); see § 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 \( X \) whose distance from \( u \) is 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 10 000 simulations of the specified Strauss process and estimate \( \theta \) 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 obtained from the different estimation methods are shown in Fig. 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 logistic regression estimate is very close to that of the exact maximum pseudolikelihood estimate.
Logistic regression for Gibbs processes

Fig. 1. Boxplots of parameter estimates of $\theta_1$ and $\theta_2$ obtained from the different estimation methods for increasing values of $n_d$. Horizontal lines represent the true parameter values. From dark to light, the four shades of grey 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 over $n_d = \infty$ in both panels.

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 $\sigma$ of the estimator and the two contributions $\sigma_1$ and $\sigma_2$, 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 $\theta_1$ and $\theta_2$, respectively.

<table>
<thead>
<tr>
<th>$n_d$</th>
<th>$\sigma \times 100$</th>
<th>$\sigma_1 \times 100$</th>
<th>$\sigma_2 \times 100$</th>
<th>$(\sigma - \sigma_1)/\sigma_1$</th>
<th>$\sigma \times 100$</th>
<th>$\sigma_1 \times 100$</th>
<th>$\sigma_2 \times 100$</th>
<th>$(\sigma - \sigma_1)/\sigma_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>66.28</td>
<td>25</td>
<td>14</td>
<td>21</td>
<td>79.99</td>
</tr>
<tr>
<td>20</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>19.22</td>
<td>17</td>
<td>14</td>
<td>10</td>
<td>23.21</td>
</tr>
<tr>
<td>40</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>4.35</td>
<td>14</td>
<td>14</td>
<td>5</td>
<td>5.48</td>
</tr>
<tr>
<td>80</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0.64</td>
<td>14</td>
<td>14</td>
<td>2</td>
<td>0.84</td>
</tr>
<tr>
<td>160</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0.08</td>
<td>14</td>
<td>14</td>
<td>1</td>
<td>0.11</td>
</tr>
</tbody>
</table>

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 in the Supplementary Material shows that these 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 $\theta_2$, the increase in root mean squared error relative to the exact maximum pseudolikelihood estimate is just 0.9% or 0.07%, respectively, when stratified dummy points are used. From now on we consider only stratified dummy points.

As mentioned in Remark 2, the variance of the logistic regression estimator is a sum of terms $\Sigma_1$ and $\Sigma_2$, where $\Sigma_2$ 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 $\Sigma_1$ and $\Sigma_2$. 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 $\sigma_1^2$ and $\sigma_2^2$ are extracted from the diagonals of $\Sigma_1$ and $\Sigma_2$. 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 $\sigma_1$ quickly converges to a constant...
value as \( n_d \) increases. The reduction in variance as \( n_d \) increases thus mainly occurs for the \( \sigma_2^2 \) term. This justifies regarding \( \sigma_2^2 \) as the increase in variance over the maximum pseudolikelihood estimate variance that is due to the random dummy points. Note also that \( \sigma_2 \) 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, multi-scale Strauss processes, Geyer's saturation processes with saturation threshold 1, and multi-type Strauss processes with two types, i.e., \( \mathcal{M} = \{1, 2\} \). The latter two unmarked point processes are specified by conditional intensities of the form (3) with, respectively, \( t(u, X) = \{1, n_{R_1}(u, X), n_{R_2}(u, X) - n_{R_1}(u, X)\} \) for \( 0 \leq R_1 \leq R_2 < \infty \) and \( t(u, X) = \{1, \sum_{v \in X \cup u} 1[d(v, X \cup u) \leq R] - \sum_{v \in X} 1[d(v, X) \leq R]\} \) for \( 0 \leq R < \infty \), where \( d(v, X) \) denotes the distance from \( v \) to the nearest point in \( X \) without \( v \). The conditional intensity of the multi-type Strauss point process is also of the form (3) with

\[
\begin{align*}
t(\{\hat{u}, m\}, X) &= \left[\delta_{m,1}, \delta_{m,2}, n_{R_1}(\{\hat{u}, m\}, X^1)\delta_{m1}, n_{R_2}(\{\hat{u}, m\}, X^2)\delta_{m2}, n_{R_{12}}(\{\hat{u}, m\}, X^1)\delta_{m1} + n_{R_{12}}(\{\hat{u}, m\}, X^1)\delta_{m2}\right],
\end{align*}
\]

where \( \delta_{jk} \) 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 have, respectively, \( \theta_2 = \log 0.8 \) and \( \theta_2 = \log 0.2 \); two multi-scale Strauss processes with \( R_1 = 0.05, R_2 = 0.1 \) and \( \theta_1 = \log 100 \), where models M1 and M2 have, respectively, \( (\theta_2, \theta_3) = (\log 0.2, \log 0.8) \) 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 have, respectively, \( \theta_2 = \log 0.2 \); and \( \theta_2 = \log 1.2 \) and \( \theta_2 = \log 0.2 \); and two multi-type Strauss processes with \( R_1 = R_2 = R_{12} = 0.05 \) and \( \theta_1 = \theta_2 = \log 50 \), where models MS1 and MS2 have, respectively, \( (\theta_3, \theta_4, \theta_5) = (0.5, 0.5, 0.5) \) and \( (\theta_3, \theta_4, \theta_5) = (0.8, 0.8, 0.2) \). For all models we use relatively small values of \( \theta_1 \), as well as of \( \theta_2 \) 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 multi-scale Strauss process and \( \max(R_1, R_2, R_{12}) \) for the multi-type 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 \( \{\theta : \|W\|^{1/2} \Sigma^{-1/2} (\hat{\theta} - \theta) \|^2 \leq 2 \chi^2_{0.95}(p)\} \). 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
Logistic regression for Gibbs processes

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 side-length \( \ell \). The \( n/|W| \) column shows the average empirical intensities for the models. The results are based on 2000 realizations from each of the models.

| Model | \( n/|W| \) | \( n_d = 20 \) | \( n_d = 40 \) | \( n_d = 80 \) | \( n_d = 20 \) | \( n_d = 40 \) | \( n_d = 80 \) |
|-------|------------|------------|------------|------------|------------|------------|------------|
| 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         |

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, closeness to the nominal level does not appear to depend on \( n_d \).

5.4. Data example

We consider the mucous membrane data shown in Fig. 1.3 of Møller & Waagepetersen (2004); our analyses are motivated by Examples 9.3 and 9.5 in that book. 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 multi-type Strauss process with log conditional intensity \( \log \lambda(u, X) = q_m(y, \theta) + \theta_{11} n_R(u, X) \) 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 \( \theta \in \mathbb{R}^{11} \) consists of the 10 polynomial regression coefficients and the interaction parameter \( \theta_{11} \leq 0 \). The polynomials depend only 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 take \( R = 0.008 \) 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, generated by sampling uniformly 1 or 2 as marks, with each mark being independent of all other variables. In spatstat, multi-type point processes are specified with respect to a counting measure on the mark space. To comply with this and following Remark 1, in our implementation we specify the dummy point intensity as \( \rho = n_d^2/0.7 \), where for this dataset we 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 employ a parametric bootstrap based on 1000 simulations generated under the fitted model, where we still take \( n_d = 60 \) 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 at six equidistant \( y \) values in the range \([0, 0.7]\). Figure 2(b) shows the estimated polynomials without the constant term, along with bootstrap confidence intervals at the selected set of \( y \) values. The plot indicates that the two trends are significantly different.
Table 3. Real-data example: logistic regression estimates with $n_d = 60$ of the two polynomials $q_1$ and $q_2$ at $y = 0.1$ or 0.6 and of the interaction parameter $\theta_{11}$, along with estimated standard deviations and relative percentage increases for $n_d = 60$ and $n_d = 120$, based on 1000 simulations from the fitted model.

<table>
<thead>
<tr>
<th></th>
<th>$n_d = 60$</th>
<th></th>
<th></th>
<th></th>
<th>$n_d = 120$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$q_1(0.1)$</td>
<td>$6.00$</td>
<td>$0.20$</td>
<td>$0.19$</td>
<td>$0.05$</td>
<td>$3.52$</td>
<td>$0.19$</td>
<td>$0.19$</td>
</tr>
<tr>
<td>$q_1(0.6)$</td>
<td>$3.53$</td>
<td>$0.75$</td>
<td>$0.75$</td>
<td>$0.06$</td>
<td>$0.35$</td>
<td>$0.75$</td>
<td>$0.75$</td>
</tr>
<tr>
<td>$q_2(0.1)$</td>
<td>$7.80$</td>
<td>$0.09$</td>
<td>$0.08$</td>
<td>$0.05$</td>
<td>$16.07$</td>
<td>$0.08$</td>
<td>$0.08$</td>
</tr>
<tr>
<td>$q_2(0.6)$</td>
<td>$7.14$</td>
<td>$0.10$</td>
<td>$0.09$</td>
<td>$0.04$</td>
<td>$10.71$</td>
<td>$0.09$</td>
<td>$0.09$</td>
</tr>
<tr>
<td>$\theta_{11}$</td>
<td>$-2.59$</td>
<td>$0.34$</td>
<td>$0.34$</td>
<td>$0.05$</td>
<td>$1.05$</td>
<td>$0.34$</td>
<td>$0.34$</td>
</tr>
</tbody>
</table>

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 given in the Supplementary Material. Also shown are the bootstrap estimates of the standard deviation $\sigma$ of the parameter estimates, the standard deviation $\sigma_1$ due to $T_1$, the standard deviation $\sigma_2$ due to $T_2$, 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, the largest relative increases in estimation standard error due to the random dummy points, around 16%, are obtained for the type 2 cells as well. We also used $n_d = 120$, and this brings the maximal relative increase in standard deviation down to about 4%; see the four columns under $n_d = 120$ in Table 3.

To test the hypothesis of equal polynomials, we fitted the null model with common coefficients of the nonconstant terms in 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 default ppm is that it is not a priori clear how large an $n_d$ must be used to avoid severe bias (Baddeley & Turner, 2013); sometimes the required value of $n_d$ may even be computationally prohibitive. In this 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 would 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 \( \rho \)

As mentioned in Remark 2, the variance \( \sigma^2 \) of a parameter estimate is the sum of a term \( \sigma_1^2 \) which is roughly constant as a function of \( \rho \) and a term \( \sigma_2^2 \) which is roughly proportional to \( 1/\rho \), \( \sigma_2^2 = \hat{s}_2^2 / \rho \), say. For a given choice of \( \rho \), 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 \( \rho_p \) such that the dummy point additional variance \( s_2^2 / \rho_p \) is less than a specified fraction \( p \) of \( \sigma_1^2 \), we can determine \( \rho_p \) as \( \rho_p = \hat{s}_2^2 / (\hat{\sigma}_2^2 p) \). This relation can also be used to determine \( p_p = \hat{s}_2^2 / (\hat{\sigma}_2^2 \rho) \) for a given \( \rho \). In practice, we may rewrite these relations in terms of standard deviations so that \( p \) gives the relative increase of the standard deviation, and hence 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 cover only finite-range stationary Gibbs point processes, but based on practical examples, such as the analysis of the mucous membrane data in the Supplementary Material, we believe that similar asymptotic results are also valid for infinite-range and nonstationary Gibbs point processes which depend on spatial covariates. Our method requires that the covariates be 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.

Acknowledgement

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, the 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 an explanation of how the proposed
estimating function is related to a time-invariance estimating function obtained from Barker dynamics.

**Appendix**

**Definitions of G₁, A₁, A₂, A₃, S and their estimates**

The matrix $G₁$ equals $\sum_{i=1}^{3} A_i(w_{i\theta}, w_{i\theta'})$ where, for $i = 1, 2, 3$, the $p \times p$ matrices $A_i(g, h)$ for two functions $g, h : S \times \Omega \rightarrow \mathbb{R}^p$ are given by

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

$$A_2(g, h) = E \left[ \int_{B(0, R) \times M} g(0^M, X)h(v, X)^T \left\{ \lambda_{0^M}(0^M, X)\lambda_{0^M}(v, X) - \lambda_{0^M}([0^M, v], X) \right\} dv \right],$$

$$A_3(g, h) = E \left[ \int_{B(0, R) \times M} \Delta_v g(0^M, X)\Delta_v h(v, X)^T \lambda_{0^M}([0^M, v], X) dv \right].$$

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

$$\lambda_\theta([u, v], X) = \lambda_\theta(u, X \cup v)\lambda_\theta(v, X) = \lambda_\theta(v, X \cup u)\lambda_\theta(u, X),$$

$$\Delta_v g(u, X) = g(u, X \cup v) - g(u, X).$$

From (13), the sensitivity matrix can be rewritten as

$$S = \frac{1}{\rho} A_1 \left( w_{i\theta}, \sqrt{\lambda_{0^M}} + \rho, w_{i\theta'}, \sqrt{\lambda_{0^M}} + \rho \right).$$

We also define, for two functions $g, h : S \times \Omega \rightarrow \mathbb{R}^p$, the computationally fast empirical estimates (Coeurjolly & Rubak, 2013) of $A_i(g, h)$ for $i = 1, 2, 3$:

$$\hat{A}_1(X, D, g, h) = \frac{1}{|W_n|} \sum_{u \in (X \cup D)_{w_u}} g(u, X \setminus u)h(u, X \setminus u)^T \frac{\lambda_{0^M}(u, X \setminus u)}{\lambda_{0^M}(u, X \setminus u) + \rho}, \quad (A1)$$

$$\hat{A}_2(X, g, h) = \frac{1}{|W_n|} \sum_{u, v \in X_{w_u}} g(u, X \setminus \{u, v\})h(v, X \setminus \{u, v\})^T$$

$$\times \left\{ \frac{\lambda_{0^M}(u, X \setminus \{u, v\})\lambda_{0^M}(v, X \setminus \{u, v\})}{\lambda_{0^M}(\{u, v\}, X \setminus \{u, v\})} - 1 \right\}, \quad (A2)$$

$$\hat{A}_3(X, g, h) = \frac{1}{|W_n|} \sum_{u, v \in X_{w_u}} \Delta_v g(u, X \setminus \{u, v\})\Delta_v h(v, X \setminus \{u, v\})^T. \quad (A3)$$

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

$$\hat{S} = \frac{1}{\rho} \hat{A}_1(X, D, w_{i\theta}, \sqrt{\lambda_{0^M}} + \rho, w_{i\theta'}, \sqrt{\lambda_{0^M}} + \rho),$$

$$\hat{G}_1 = \hat{A}_1(X, D, w_{i\theta}, w_{i\theta'}) + \hat{A}_2(X, w_{i\theta}, w_{i\theta'}) + \hat{A}_3(X, w_{i\theta}, w_{i\theta'}).$$

**Remark A1** (Definition of $\hat{S}$). If we had followed the strategy proposed in Coeurjolly & Rubak (2013), the estimate of $A_1(w_{i\theta}, w_{i\theta'})$ 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 $\hat{S}$, using (A1), depends only on the quantities $t_j(u, X \setminus u)$ for $j = 1, \ldots, p$ and $u \in (X \cup D)_{w_u}$, which have already been stored when computing the estimate $\hat{\theta}$. 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 $X$. 
Definition of $\hat{G}_2$

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

(i) if $D \sim \nu \Omega(\mathbb{R}^d, \rho)$, the estimate $\hat{G}_2^\nu$ is defined by

$$\frac{1}{\rho} \hat{A}_1(X, D, w_\delta \sqrt{\lambda_\delta}, w_\delta \sqrt{\lambda_\delta})$$

(ii) if $D \sim \mu \Omega(\mathbb{R}^d, \rho)$, the estimate $\hat{G}_2^\mu$ is defined by

$$\frac{1}{\rho} \left\{ \kappa_n \hat{A}_1(X, D, w_\delta \sqrt{\lambda_\delta}, w_\delta \sqrt{\lambda_\delta}) - \hat{A}_1(X, D, w_\delta \sqrt{\lambda_\delta}, \sqrt{\lambda_\delta}) \hat{A}_1(X, D, w_\delta \sqrt{\lambda_\delta}, \sqrt{\lambda_\delta})^T \right\}$$

where $\kappa_n = |W_n|^{-1} \sum_{u \in (X \cup D) \backslash u} (\lambda(u, X \backslash u) + \rho)^{-1}$;

(iii) if $D \sim \sigma \Omega(\mathbb{R}^d, \rho)$, the estimate $\hat{G}_2^\sigma$ is defined by

$$\frac{1}{2\rho^2|W_n|} \sum_{\ell \in \mathbb{Z}^d} \left\{ w_\delta(U_{\ell}, X) - w_\delta(U_{\ell}', X) \right\} \left\{ w_\delta(U_{\ell}, X) - w_\delta(U_{\ell}', X) \right\}^T$$

where $D' = \bigcup_{k \in \mathbb{Z}^d} \{U_k\}$ is a marked stratified point process independent of $D$.

Remark A2. The variable $\kappa_n$ in (A4) converges to 1 as $n \to \infty$. It has been introduced to ensure that the estimate $\hat{G}_2^\nu$ is a positive-semidefinite matrix. By definition, $\hat{G}_2^\nu$ and $\hat{G}_2^\mu$ also satisfy this property.

Remark A3. Following Remark A1, we have proposed, in the Poisson and binomial cases, estimates of $G_2$ based on $X \cup D$. As for $\hat{S}$, the estimates $\hat{G}_2^\nu$ and $\hat{G}_2^\mu$ do not involve new numerical computations. The estimate of $G_2^\sigma$ is more awkward to handle and requires an extra dummy point process $D'$. As pairs of points are involved in (A5), we could not include the data points without adding second-order characteristics computations for $X$, and this has not yet been investigated.

References


A. BADDELEY, J.-F. COEURJOLLY, E. RUBAK AND R. WAAGEPETersen


[Received January 2013. Revised October 2013]