Quaderni di Statistica
Vol. 5, 2003

Refining pseudolikelihood estimates for point processes

Andrea Pallini
Dipartimento di Scienze Statistiche
Università di Bologna
E-mail: pallini@stat.unibo.it

Summary: Maximum pseudolikelihood estimates for Gibbs point processes (observed in a planar region A) are typically derived from a convenient approximation of the pseudolikelihood accomplished through numerical integration. Efficient estimates for Gibbs processes with loglinear conditional intensities can then be obtained by fitting a Poisson loglinear model as iteratively-reweighted least-squares estimates. We here study the local linear smoothing of all numerical contributions (from the observed region A) to maximum pseudolikelihood estimates as a way of further improving their statistical performance in terms of bias and variance (mean square error).

Key words: Berman-Turner device, Gibbs point process, local polynomial regression, pseudolikelihood, quadrature rule, smoothing.

1. Introduction

Computational methods for maximum pseudolikelihood estimates in Gibbs (Markov) point processes (Besag, 1977) are crucial in applications (cf. Diggle and Gratton, 1984, Diggle *et al.*, 1994, Goulard *et al.*, 1996, Huang and Ogata, 1999, 2002, and Mateu and Montes, 2001). These methods are typically based on numerical quadrature rules (cf. Davis and

Rabinowitz, 1984, chapters 1, 2, 5) over a planar (bounded) region, where the point configuration is observed.

The Berman-Turner device requires a preliminary random generation of artificial points, and then produce, together with the observed configuration of points of the process, maximum pseudolikelihood estimates as iteratively-reweighted least-squares estimates of a Poisson log-linear model (McCullagh and Nelder, 1989, chapter 6, and Chambers and Hastie, 1992, chapter 6). The original method is proposed in Berman and Turner (1992) for line and Poisson processes. Recently, extensions of the Berman-Turner device to general Gibbs (Markov) point processes have been proposed by Baddeley and Turner (2000a,b).

An interesting feature of the Berman-Turner device is the fact that maximum pseudolikelihood estimates can be calculated with well-known model-fitting statistical software (cf. Chambers and Hastie, 1992). Related work may include Diggle *et al.* (1994), Lindsey (1995), Assunção and Guttorp (1999) and Mateu and Montes (2001).

By applying the results presented in Preston (1976), chapter 6, we aim to identify asymptotics for these maximum pseudolikelihood estimates. In particular, we view the observed point configuration as a finite Gibbs lattice field. Following Green (1984), asymptotic variances can thus be derived as described by Jensen (1993), by considering an increasing domain asymptotics, which is theoretically valid under the Dobrushin uniqueness condition for Gibbs measure.

Here we wish to obtain an asymptotically more efficient version of the Berman-Turner device by smoothing all contributions to maximum pseudolikelihood estimates (from region A) with nonparametric techniques for local polynomial regression (cf. Fan and Gijbels, 1996, chapter 3, and Ruppert and Wand, 1994), and then integrating them formally.

2. Pseudolikelihood function

Let $x_A = \{x_1, \ldots, x_n\}$ be a point configuration in a bounded region $A \subset \mathbb{R}^2$, where $n = n(x_A) \geq 0$ is a random variable. We suppose that x_A is a restriction to A of a stationary (translation invariant) Gibbs (Markov) point process X .

Region A may be viewed as a bounded set in $\mathbb{R}^2$. Region A works as a sampling window within a larger region and x_A consists of a finite number of points x (with components $x^{(p)}$, $p = 1, 2$), generated by the point process X , which lay in A . We denote by a the area of region A.

We assume that the probability of X is

$$f(x_A) = f(x_A; \theta)\,, \tag{1}$$

with respect to the distribution of the Poisson process with intensity 1 on A , where $\theta \in \Theta \subset \mathbb{R}^q$. We also assume that $f(x_A;\theta) > 0$ implies that $f(x'_A;\theta) > 0$, for all point configurations $x'_A \subset x_A$. A broad class of Gibbs point processes (with unique Gibbs measure μ and finite interaction r) may be defined under such conditions; see Preston (1976), chapter 6, Baddeley and Møller (1989) and Baddeley and van Lieshout (1995).

The Papangelou intensity $\lambda_\theta(u; x_A)$ of the Gibbs point process X at a point $u \in A$ defines the pseudolikelihood function $PL(\theta; x_A)$ (Besag, 1977). From (1), the intensity $\lambda_\theta(u; x_A)$ can be obtained as

$$\lambda_\theta(u; x_A) = f(x_A \cup \{u\}) \,/\, f(x_A)\,,$$

$u \notin x_A$, or

$$\lambda_\theta(x_i; x_A) = f(x_A) \,/\, f(x_A \backslash \{x_i\})\,,$$

$x_i \in x_A$, and is the conditional probability that process X has a point at u or x_i given the rest of X in x_A .

In particular, for a subset $B \subseteq A$,

$$PL(\theta; x_A) = \left\{ \prod_{x_i \in B} \lambda_\theta(x_i; x_A) \right\} \exp\!\left(- \int_B \lambda_\theta(u; x_A)\, du \right). \tag{2}$$

See also Jensen (1993) and Barndorff-Nielsen *et al*. (1999), chapter 3.

To further the work by Berman and Turner (2000a), we focus on Gibbs point processes with loglinear conditional intensity (cf. Baddelley and Møller, 1989, and Baddeley and van Lieshout, 1995). That is,

$$\lambda_\theta(u; x_A) = \exp(\theta^T S(u; x_A))\,, \tag{3}$$

where $S(u;x_A)$ is a vector of q spatial covariates defined at each point u in A. We assume that

$$\| \, S(u;x_A) \, \| \, \exp(\theta^T S(u;x_A))$$

is uniformly bounded in $u \in A$ and $\theta \in \Theta$, for each fixed x_A.

For simplicity, we take a square region A and $B = A$ in the above definitions of conditional intensity.

Normal equations

$$\partial \log(PL(\theta;x_A)) \, / \, \partial\theta = 0 \, ,$$

then become

$$\sum_{x_i \in A} S(x_i;x_A) = \int_A S(u;x_A) \, \exp(\theta^T S(u;x_A)) \, du \, , \tag{4}$$

with both sides equal under expectation.

The loglinear form of the conditional intensity makes $\log(PL(\theta;x_A))$ concave (cf. Pratt, 1981). If the parameter set Θ is convex, it also follows that maximum pseudolikelihood estimates exist at an interior point of Θ or on the convex boundary $\partial\Theta$ of Θ.

Solution to normal equations (4) requires numerical integration.

Example (Strauss process). This point process is a pairwise interaction process (cf. Baddeley and Møller, 1989). The Strauss process is a good model for ordered point configurations and defines the Poisson process as a specific case. Let r be the interaction radius. Let $n_r(x_A)$ be the number of pairs of distinct points in the region A, which lie within a distance r of one another. The conditional intensity is defined as

$$\lambda_{\beta,\gamma}(u;x_A) = \beta \, \gamma^{\tau_r(u;x_A)} \, , \tag{5}$$

where $\beta > 0$ and $0 \leq \gamma \leq 1$, $\tau_r(u;x_A)$ is the number of distinct points x_i in A, such that

$$0 < \| \, x_i - u \, \| \leq r \, .$$

The pseudolikelihood is loglinear,

$$\log(PL(\beta,\gamma;x_A)) \quad = \quad n(x_A) \log(\beta) + 2n_r(x_A) \log(\gamma)$$

$$- \quad \beta \left(\int_A \gamma^{\tau_r(u;x_A)} \, du \right) , \tag{6}$$

where vectors $\theta = (\log(\beta), \log(\gamma))^T$, and $S(u; x_A) = (1, \tau_r(u; x_A))^T$. Normal equations are of the form (4).

3. Estimates by quadrature rules

Extending the Berman-Turner device (Berman and Turner, 1992), Baddeley and Turner (2000a,b) suggests the preliminary random generation of a configuration of artificial points in the observed region A , to form (with the observed point configuration x_A of size n) a configuration of quadrature points

$$u_A = \{u_1, \dots, u_m\} ,$$

where the point configuration $x_A \subset u_A$ and $n < m$. See Figure 1.

For simplicity, we suppose that $u_A \backslash x_A$ is a restriction to A of a homogeneous Poisson point process U, independent of X . Size m of u_A is the value of the random variable $m = m(u_A)$, which depends on the intensity of the point process U in A . In this sense, we may say that extensions in Baddeley and Turner (2000a,b) require an appropriate superposition of point processes X and U .

The Berman-Turner device defines the integral in (2) as the finite Riemann sum

$$\int_A \lambda_\theta(u; x_A) \, du \; \approx \sum_{j=1}^{m} \lambda_\theta(u_j; x_A) \, \hat{w}_j \; , \tag{7}$$

where $\hat{w}_j > 0$ are the areas of the tiles partitioning the region A and containing one point u_j each, taken as quadrature weights summing to the area a of A (cf. Davis and Rabinowitz (1984), chapter 5).

Rectangular tiles (cf. Figure 1) may be used for speeding up the calculation of the Riemann sum in the right hand side of (7). In this case, the weight $\hat{w}_j$ may be obtained as the ratio of the area of the tile containing u_j to the number of quadrature points in the same tile, $j = 1, \dots, m$.

The log-pseudolikelihood $\log(PL(\theta; x_A))$, where $PL(\theta; x_A)$ is defined in (2), is then approximated by

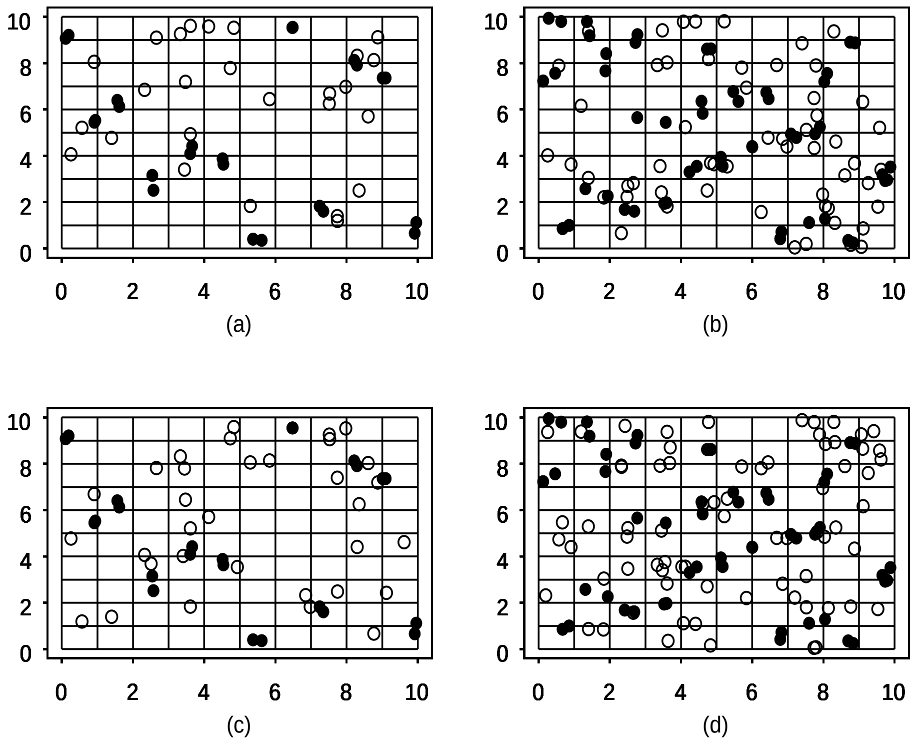

Figure 1. Simulated points (•) from moderately clustered point processes of Strauss type (with interaction radius $r = 0.5$) in a region A of size 10×10; number of points $n = 24$ in (a), $n = 52$ in (b), $n = 24$ in (c) and $n = 52$ in (d). Random artificial points (○) for integration rules, from a homogeneous Poisson process, with a set $W(l)$ of $l = 100$ quadrature tiles of size 1×1 in the same region A; number of artificial points $m - n = [1.1\, n]$ in (a) and (b), and $m - n = [1.4\, n]$ in (c) and (d), where $[u]$ denotes the integer not greater than real u.

$$\log(PL(\theta; u_A)) \approx \sum_{i=1}^{n} \log(\lambda_\theta(x_i; x_A)) - \sum_{j=1}^{m} \lambda_\theta(u_j; x_A)\, \hat{w}_j ,$$

which is equivalent to

$$\log(PL(\theta; u_A)) = \sum_{j=1}^{m} \hat{w}_j \{y_j \log(\lambda_j) - \lambda_j\} , \tag{8}$$

where

$$\lambda_j = \lambda_\theta(u_j; x_A) , \tag{9}$$

$$y_j = \hat{w}_j^{-1} z_j , \tag{10}$$

and $z_j = 1$, if $u_j \in x_A$, $z_j = 0$, if $u_j \in u_A \backslash x_A$.

Baddeley and Turner (2000a) point out that (8) is equivalent to the log-likelihood of independent Poisson variables Y_j with means λ_j , taken with weights $\hat{w}_j$ (cf. McCullagh and Nelder, 1989, chapter 6, and Chambers and Hastie, 1992, chapter 6).

Maximum pseudolikelihood estimates $\hat{\theta}$ of θ in (2), and in (3), can be obtained by fitting the Poisson loglinear model

$$\log(\lambda_j) = \theta^T v_j , \tag{11}$$

where λ_j are defined in (9), to responses y_j (given by (10)) and covariate values $v_j = S(u_j; x_A)$, with weights $\hat{w}_j$ defined in (7).

Example (Strauss process). Maximum pseudolikelihood estimates $\hat{\theta}$ can be obtained by fitting the model

$$\log(\lambda_j) = \theta_1 + \theta_2\, v_j ,$$

to y_j and $v_j = \tau_r(u_j; x_A)$, where λ_j is defined in (5), $\theta_1 = \log(\beta)$, $\theta_2 = \log(\gamma)$. Whereas $\gamma > 1$, concavity of the log-pseudolikelihood (6) implies that $\hat{\theta}_2 \equiv 0$.

Loglinear form (11) may also characterize soft core point process, point processes with step function interaction, Ord's point processes, area-interaction point processes and inhomogeneous models with spatial covariates (cf. Baddeley and van Lieshout, 1995, and Baddeley and Turner, 2000a).

Extensions in Baddeley and Turner (2000a) to the Berman-Turner device include maximum pseudolikelihood estimates $\hat{\theta}$ for multi-type and marked point processes (cf. Baddeley and Møller, 1989).

4. Refining estimates by local smoothing

Our refined Berman-Turner device approximates to the integral in (2) by formally integrating a nonparametric regression estimator fitted locally to $\lambda_\theta(u_j; x_A)$ and quadrature points u_j .

Let us consider a bivariate locally weighted regression (Scott, 1992, chapter 8, Ruppert and Wand, 1994, Wand and Jones, 1995, chapter 5, Fan and Gijbels, 1996, chapter 3, Hjort and Jones, 1996, and Loader, 1996), where the explanatory variables are observed as the components $u_j^{(1)}$ and $u_j^{(2)}$ of quadrature points u_j in the set u_A , which contains the configuration of points x_A .

Let H be a 2×2 symmetric positive definite matrix depending on m . We denote by

$$K_H(t) = |\, H \,|^{-1/2} \; K(H^{-1/2}\, t)$$

a bivariate kernel function, $t \in \mathbb{R}^2$.

For the kernel K and the sequence of bandwidth matrices $|\, H \,|^{1/2}$, we assume the regularity conditions (A1), (A2), (A3) and (A4) in Ruppert and Wand (1994). With (A3), in particular, we assume that $m^{-1} |\, H \,|$ and each entry of H tends to zero as $m \to \infty$, with H remaining symmetric and positive definite.

A consistent nonparametric regression estimator (cf. Wand and Jones, 1995, chapter 5) is the weighted linear combination of the $\lambda_\theta(u_j; x_A)$

defined as

$$\hat{g}(t) = m^{-1}\,\{K_H(0)\}^{-1}\sum_{j=1}^{m}\lambda_\theta(u_j;x_A)\,K_H(t-u_j)\,. \qquad (12)$$

Nonparametric fits of (12) can be improved by analogous locally linear and locally quadratic regression estimators (cf. Ruppert and Wand, 1994).

A refined log-pseudolikelihood $\log(PL(\theta;x_A))$ can be defined as

$$\begin{aligned}\log(PL_{ref}(\theta;u_A)) &\approx n^{-1}\sum_{i=1}^{n}\log(\lambda_\theta(x_i;x_A)) \\ &- \hat{\omega}\sum_{j=1}^{m}\lambda_\theta(u_j;x_A)\,, \end{aligned}\qquad (13)$$

which is

$$\log(PL_{ref}(\theta;u_A)) = \hat{\omega}\sum_{j=1}^{m}\{y_j\log(\lambda_j)-\lambda_j\}\,, \qquad (14)$$

where

$$\lambda_j = \lambda_\theta(u_j;x_A)\,, \qquad (15)$$

$$\hat{\omega} = m^{-1}\,\{K_H(0)\}^{-1}\int_A K_H(t-u_j)\,dt\,, \qquad (16)$$

$$H = a^{1/2}\begin{pmatrix} h_{11}^2 & h_{12}^2 \\ h_{21}^2 & h_{22}^2 \end{pmatrix} n^{-1/5}\,, \qquad (17)$$

where a is the area of region A,

$$y_j = \hat{\omega}^{-1}\,n^{-1}\,z_j\,, \qquad (18)$$

where $z_j = 1$, if $u_j \in x_A$, $z_j = 0$, if $u_j \in u_A \backslash x_A$.

Refined maximum pseudolikelihood estimates $\hat{\theta}_{ref}$ of θ in (2) can be calculated by fitting the Poisson loglinear model

$$\log(\lambda_j) = \theta^T v_j\,, \qquad (19)$$

where λ_j are defined in (15), to responses y_j (given by (18)) and covariate values $v_j = S(u_j; x_A)$, with weights $\hat{\omega}$ defined in (16).

Bandwidth selectors for multivariate locally weighted regression estimators are studied in Ruppert *et al.* (1995) and might be modified (cf. Hjort and Jones, 1996) to improve the statistical efficiency of the first term in log-pseudolikelihood (13).

Refined maximum pseudolikelihood estimates $\hat{\theta}_{ref}$ for multi-type and marked point processes can be introduced and studied as well, starting from Baddeley and Turner (2000a) and (14).

5. Asymptotics

Preston (1976), chapter 6, shows that a point process given by the density (1) can be viewed as a Gibbs random field, and vice versa.

Let $\{X_v\, ,\ v \in \mathbb{Z}^2\}$ be a Gibbs random field. Without loss of generality, let e be a real value, product of two components $e^{(1)}$ and $e^{(2)}$, where $e = e^{(1)} e^{(2)}$, $e^{(1)} = e^{(2)}$, and $e^{(1)} \geq r$. For every $z \in \mathbb{Z}^2$, define

$$A_v = \{\, t \in \mathbb{R}^2 \,:\, e^{(p)}(2\, z^{(p)} - 1)/2 \leq\, t^{(p)} <\, e^{(p)}(2\, z^{(p)} + 1)/2 \,\}\,, \quad (20)$$

where $p = 1, 2$. Square tiles $\{A_v\}$ partition $\mathbb{R}^2$, while e determines their common area.

We can transform a Gibbs point process into the Gibbs field $\{X_v\, \}$, by taking the point configuration x_{A_v} in the tile A_v as the v-site variable,

$$X_v = x_{A_v}\,. \quad (21)$$

More precisely, the tile A_v in (21) is given by (20), and site v of X_v is the 2-dimensional vector of the barycentric coordinates $v^{(1)}$ and $v^{(2)}$ of A_v , for every $v \in \mathbb{Z}^2$.

Here we study asymptotics of iteratively-reweighted least-squares estimates (cf. Green, 1984), without pursuing the topic of numerical efficiency of Riemann sums in log-pseudolikelihoods (8) and (14). Mathematical theory for numerical errors and their bounds may be found in Davis and Rabinowitz (1984), chapters 2 and 5.

Let $W(l) \subset \mathbb{Z}^2$ be the set of sites of cardinality l, such that the square region A may be written as

$$A = \bigcup_{v \in W(l)} A_v \,. \tag{22}$$

See Figure 1, for a set $W(l)$ with $l = 100$ tiles. Log-pseudolikelihoods (8) and (14), being based on Riemann sums, have the linear form

$$\log(PL(\theta; u_A)) = \sum_{v \in W(l)} \sum_{j=1}^{m_v} \log(PL(\theta; u_j))\,, \tag{23}$$

where $m_v = m(A_v)$ denotes the random number of quadrature points u_j in the tile A_v. In particular, we denote by $\hat{w}_{vj}$, y_{vj} and z_{vj}, and λ_{vj} the weights, responses and predictors in the tile A_v, for every $v \in W(l) \subset \mathbb{Z}^2$. Recall that in (8) weights $\hat{w}_j$ are positive and vary with tiles, $j = 1, \ldots, m$. Let $\hat{w} = max_{\,1 \le j \le m}\,(\hat{w}_j)$. From (23), it follows that

$$\begin{aligned} \sum_{j=1}^{m_v} \log(PL(\theta; u_j)) &= \sum_{j=1}^{m_v} \hat{w}_{vj}\,\{\hat{w}_{vj}^{-1} z_{vj} \log(\lambda_{vj}) - \lambda_{vj}\} \\ &\ge \hat{w} \sum_{j=1}^{m_v} \{y_{vj} \log(\lambda_{vj}) - \lambda_{vj}\}\,, \end{aligned} \tag{24}$$

where $y_{vj} = \hat{w}^{-1} z_{vj}$ in log-pseudolikelihood (8), and

$$\sum_{j=1}^{m_v} \log(PL(\theta; u_j)) = \hat{\omega} \sum_{j=1}^{m_v} \{y_{vj} \log(\lambda_{vj}) - \lambda_{vj}\}\,, \tag{25}$$

where $y_{vj} = \hat{\omega}^{-1} n^{-1} z_{vj}$ in log-pseudolikelihood (14). From (24) and (25), it is easy to see that the first and second derivatives of log-pseudolikelihoods (8) and (14) can be written as sums over $W(l)$.

We assume that $W(l)$ expands to $\mathbb{Z}^2$ in all directions, as does the disc D of radius $e^{(1)} l/2$ contained in the region A, where $D \subset W(l)$, as $l \to \infty$. Recall definition (22) for A. We use $l \Rightarrow \infty$, to indicate this increasing domain asymptotics. In log-pseudolikelihoods (8) and (14), stationarity of points and artificial points in u_A, in the superposition of point processes X and U, implies that $n \to \infty$, $m \to \infty$, and $m - n \to \infty$, as $l \Rightarrow \infty$.

We assume the Dobrushin uniqueness condition for Gibbs measure μ. It follows a mixing decay of spatial correlations for the Gibbs field $\{X_v\}$ (cf. Jensen, 1993, and Pallini, 2000), as $l \Rightarrow \infty$.

We denote by S_{vrj} the r-th component of the vector $S(u_j; x_{A_v})$ of spatial covariates defined at u_j in the tile A_v. From (3) and (24), we define

$$\begin{aligned} \gamma_{vrs} &= -\,\hat{w}\,(\partial^2 \,/\, \partial\theta_r \partial\theta_s)\,\Big[\sum_{j=1}^{m_v} \{\hat{w}^{-1} z_{vj} \log(\lambda_{vj}) - \lambda_{vj}\}\Big]\,, \\ &= \hat{w} \sum_{j=1}^{m_v} \{\partial^2 \lambda_{vj} / \partial \log(\lambda_{vj})^2\}\, S_{vrj}\, S_{vsj}\,, \end{aligned} \tag{26}$$

where $r, s = 1, \ldots, q$, and

$$\hat{\Gamma} = \sum_{v \in W(l)} \begin{pmatrix} E(\gamma_{v11}) & \cdots & E(\gamma_{v1q}) \\ \vdots & & \vdots \\ E(\gamma_{vq1}) & \cdots & E(\gamma_{vqq}) \end{pmatrix}\,, \tag{27}$$

where expectation is taken with respect to the density (1). From (3) and (25), we define

$$\gamma_{vrs} = \hat{\omega} \sum_{j=1}^{m_v} \{\partial^2 \lambda_{vj} / \partial \log(\lambda_{vj})^2\}\, S_{vrj}\, S_{vsj}\,,$$

where $r, s = 1, \ldots, q$, and

$$\hat{\Gamma}_{ref} = \sum_{v \in W(l)} \begin{pmatrix} E(\gamma_{v11}) & \cdots & E(\gamma_{v1q}) \\ \vdots & & \vdots \\ E(\gamma_{vq1}) & \cdots & E(\gamma_{vqq}) \end{pmatrix}\,. \tag{28}$$

In (27) and (28), we assume that

$$\hat{w} \xrightarrow{p} w\,,$$

$$\hat{\omega} \xrightarrow{p} \omega\,,$$

as $l \Rightarrow \infty$, respectively, where $w > 0$ and $\omega > 0$ are scalars. The asymptotic covariance matrix Γ^{-1}, is obtained from (27) and (28) by convergence in probability,

$$l^{-1}\,\hat{\Gamma}^{-1} \xrightarrow{p} w^{-1}\,\Gamma^{-1}\,,$$

$$l^{-1}\,\hat{\Gamma}^{-1}_{ref} \xrightarrow{p} \omega^{-1}\,\Gamma^{-1}\,,$$

as $l \Rightarrow \infty$. The asymptotic covariance matrices of responses Y_j in log-pseudolikelihoods (8) and (14) are $Var(Y) = w^{-1}\,\Lambda$ and $Var(Y) = \omega^{-1}\,\Lambda$, as $l \Rightarrow \infty$, respectively, and Λ is a specified matrix. According to the Central Limit Theorem 4.1 in Jensen (1993), under the Dobrushin uniqueness condition for Gibbs measure μ, it is seen that

$$l^{-1/2}\,(\hat{\theta} - \theta) \xrightarrow{d} N_q(0\,,\, w^{-1}\,\Gamma^{-1})\,, \tag{29}$$

$$l^{-1/2}\,(\hat{\theta}_{ref} - \theta) \xrightarrow{d} N_q(0\,,\, \omega^{-1}\,\Gamma^{-1})\,, \tag{30}$$

as $l \Rightarrow \infty$.

Refined maximum pseudolikelihood estimates $\hat{\theta}_{ref}$ from the observed region A may be preferred to corresponding maximum pseudolikelihood estimates $\hat{\theta}$, if

$$\omega^{-1} \leq w^{-1}\,, \tag{31}$$

as $l \Rightarrow \infty$. Asymptotic result (31) is also valid with appropriate values for sizes m and n in the configuration of quadrature points u_A in the region A, and a sufficiently large number l of tiles $\{A_v\}$ given by (20) in A.

Linear form (23) is determined by the linearity of the Riemann sums, which are calculated in (8) and (14), for approximating the integral in pseudolikelihood (2). Considering stationarity of point processes X and U, observe that the numerical accuracy of Riemann sums in (8) and (14) may improve (cf. Davis and Rabinowitz, 1984, chapters 2 and 5), if the intensity of U increases, as $W(l)$ expands to $\mathbb{Z}^2$.

Results (29) and (30) basically originate from asymptotics for iteratively-reweighted least-squares estimates with independent and identically distributed observations. See Green (1984). A more general formulation of (27) and (28) should be based on the shift operator over the space of configurations of points in $\mathbb{R}^2$, as described in Jensen (1993).

Comparison with other stationary point processes U for generating the configuration u_A of $m - n$ artificial points in the observed region A of the point process X remains of interest.

The asymptotic behavior of maximum pseudolikelihood estimates $\hat{\theta}$ and $\hat{\theta}_{ref}$ in multi-type and marked point processes (cf. Baddeley and Turner, 2000a) requires other theoretical details.

6. Simulation study

In this section, we report on a simulation experiment conducted to compare and analyze the performance of our refined maximum pseudo-likelihood estimates $\hat{\theta}_{ref}$.

We studied a product kernel $K_H(t)$, $t \in \mathbb{R}^2$, defined by

$$K(t) = k(t^{(1)})\, k(t^{(2)})\,, \tag{32}$$

with univariate Gaussian kernels $k(t^{(p)})$, where $p = 1, 2$, with the bandwidth matrix $H = diag(\, h_1^2\,,\, h_2^2\,)$, where $h_1 = h_2 = h$. In particular, condition (A3) in Ruppert and Wand (1994) became $m^{-1}h^4 \rightarrow 0$ and $h^2 \rightarrow 0$, as $l \Rightarrow \infty$.

We considered moderately clustered Strauss-like processes with $\gamma = 1.2$ ($\theta_2 \equiv 0$, $\theta_1 = n/a$, where $n = n(x_A)$ in a region A of area $a = 10 \times 10$) and interaction radius $r = 0.5$, simulated (by the alternating birth-death technique of Ripley, 1977) from the conditional intensity $\lambda_{\beta,\gamma}(u; x_A)$, which determines (6). We studied configurations x_A of n points, with $n = 16, 20, 24, 28, 32, 36, 40,\ 44, 48$ and 52, with simulations of artificial points of size $m - n = [1.1\, n]$, $m - n = [1.2\, n]$, $m - n = [1.3\, n]$ and $m - n = [1.4\, n]$, where $[u]$ denotes the integer not greater than real u.

Monte Carlo bias and variance of maximum pseudolikelihood estimates $\hat{\theta}$ and Monte Carlo bias and variance of refined maximum pseudo-likelihood estimates $\hat{\theta}_{ref}$ were obtained from 500 independent repetitions of the same simulation trial.

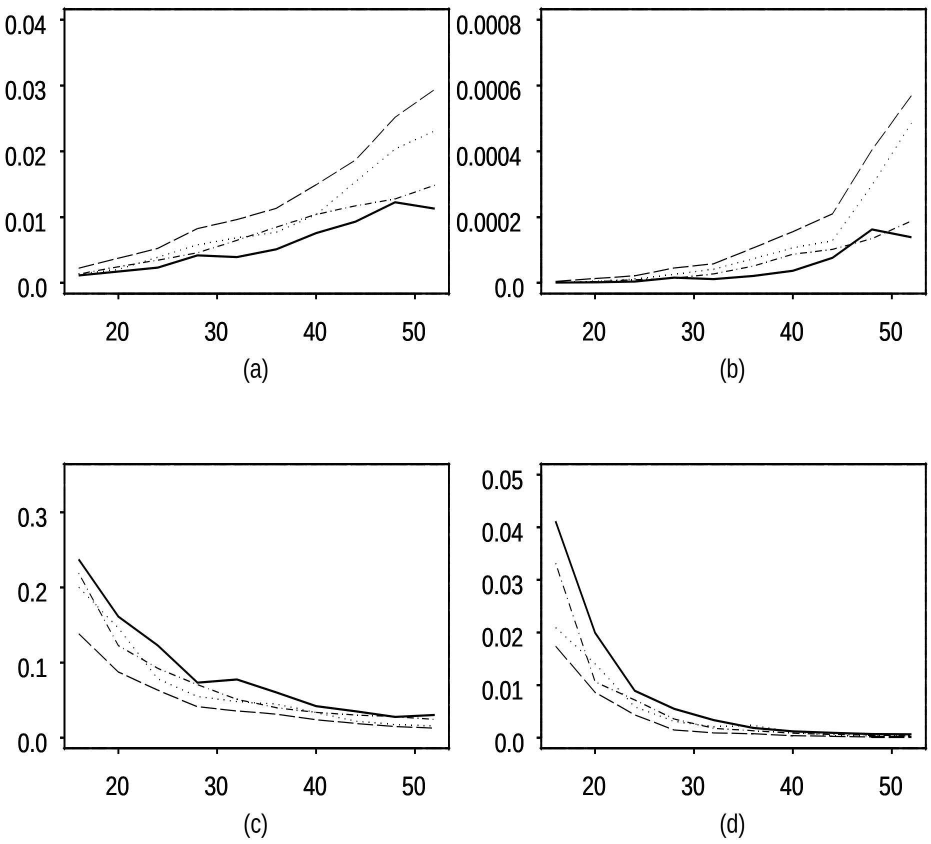

Figure 2. Comparison of maximum pseudolikelihood estimates $\hat{\theta}_{ref}$ and $\hat{\theta}$ for θ in point processes of Strauss type, for 10 values of n (horizontal axes); $Bias(\hat{\theta}_{ref\,1})/Bias(\hat{\theta}_1)$ in (a), and $Var(\hat{\theta}_{ref\,1})/Var(\hat{\theta}_1)$ in (b), and $Bias(\hat{\theta}_{ref\,2})/Bias(\hat{\theta}_2)$ in (c), and $Var(\hat{\theta}_{ref\,2})/Var(\hat{\theta}_2)$ in (d), where $n = 16, 20, 24, 28, 32, 36, 40, 44, 48, 52$ in $m - n = [1.1\,n]$ (solid), $m - n = [1.2\,n]$ ($\cdots$), $m - n = [1.3\,n]$ ($- \cdot -$) and $m - n = [1.4\,n]$ ($- - -$), where $[u]$ denotes the integer not greater than real u.

Maximum pseudolikelihood estimates $\hat{\theta}$ were calculated by partitioning the square region A into $l = 100$ quadrature tiles $\{A_v\}$ of common area $e = a/l$, where $a = 10 \times 10$, with quadrature weights $\hat{w}_j = e/m(A_v)$, $j = 1, \ldots, m$, where the tile A_v contains the j th quadrature point u_j. See Figure 1.

Observe that in the refined log-pseudolikelihood (14) the integral is invariant with respect to the bandwidth matrix $H = diag(h^2, h^2)$, which defines the product kernel (32). Refined log-pseudolikelihood (14) was always implemented with the value

$$|\,H\,|^{-1/2} = a^{-1/2}\,(2101.53)^{-1}\,n^{1/5}\,,$$

where area $a = 10 \times 10$.

For calculating iteratively-reweighted least-squares estimates $\hat{\theta}$ and $\hat{\theta}_{ref}$, we applied the S-PLUS command

$$glm\,(\,y \sim v\,, family = poisson\,,\, link = \log\,,\, weights = t\,)$$

(cf. Chambers and Hastie, 1992, chapter 6) with

$$t = (\hat{w}_1, \ldots, \hat{w}_m)^T\,,$$

$$t = (\hat{\omega}, \ldots, \hat{\omega})^T\,,$$

respectively, where

$$y = (y_1, \ldots, y_m)^T\,,$$

$$v = (\tau_r(u_1; u_A), \ldots, \tau_r(u_m; u_A))^T\,.$$

Edge corrections (cf. Barndorff-Nielsen *et al.*, 1999, chapter 3) for the square region A were never implemented.

Figure 2 shows that $\hat{\theta}_{ref}$ outperforms $\hat{\theta}$ in estimating θ, for all sizes m from $m - n = [1.1\,n]$ to $m - n = [1.4\,n]$; this conclusion might be expected from the asymptotic comparison (31) of covariance matrices.

For $m - n < [0.95\,n]$, $\hat{\theta}$ seems to be numerically hard for good computer facilities (HP Pentium III, 700 Mhz); with the present simulation experiment, in particular, the Monte Carlo variance of $\hat{\theta}_1$ may be meaningless (cf. Deng and Paul, 2000). Estimate $\hat{\theta}_{ref\,2}$ gains efficiency with respect to $\hat{\theta}_2$, with a simultaneous increase in the variability of refined estimates $\hat{\theta}_{ref\,1}$ for θ_1; this effect may be explained in part by the orthogonality between the variance of responses Y_j and the estimates of components θ_1 and θ_2 in θ. The bias of components $\hat{\theta}_{ref\,1}$ and $\hat{\theta}_{ref\,2}$ in $\hat{\theta}_{ref}$ behaves similarly.

Different results for iteratively-reweighted least-squares estimates $\hat{\theta}$ and $\hat{\theta}_{ref}$ may be obtained, by defining weights $\hat{w}_j$ and $\hat{\omega}_j$ in terms of Dirichlet, Johnson-Mehl, Voronoi or centroidal Voronoi tessellations (cf. Barndorff-Nielsen *et al.*, 1999, chapter 2, Du *et al.*, 1999, Okabe *et al.*, 2000, and Baddeley and Turner, 2000a) in region A. In this way, we have only one quadrature point in each tessellation, and we may produce better numerical approximations of Riemann integrals.

References

Assunção, R. and Guttorp, P. (1999), Robustness for inhomogeneous Poisson point processes, *Annals of the Institute of Statistical Mathematics*, 51, 657-678.

Baddeley, A.J. and Møller, J. (1989), Nearest-neighbour Markov point processes and random sets, *International Statistical Review*, 57, 89-121.

Baddeley, A. and Turner, R. (2000a), Practical maximum pseudolikelihood for spatial point patterns (with discussion), *Australian and New Zealand Journal of Statistics*, 42, 283-322.

Baddeley, A. and Turner, R. (2000b), SPATSTAT, an Splus library for spatial statistics, *http://www.maths.uwa.edu.au/∼adrian/spatstat.html* .

Baddeley, A. and van Lieshout, M.N.M. (1995), Area-interaction point processes, *Annals of the Institute of Statistical Mathematics*, 47, 601-619.

Barndorff-Nielsen, O.E., Kendall, W.S. and van Lieshout, M.N.M. (eds.) (1999), *Stochastic Geometry: Likelihood and Computation*, London: Chapman and Hall.

Berman, M. and Turner, T.R. (1992), Approximating point process likelihoods with GLIM, *Applied Statistics*, 41, 31-38.

Besag, J. (1977), Some methods of statistical analysis for spatial data, *Bulletin of the Institute of International Statistics*, 47, 77-92.

Chambers, J.M. and Hastie, T.J. (1992), *Statistical Models in S*, Pacific Grove (California): Wadsworth and Brooks/Cole.

Davis, P.J. and Rabinowitz, P. (1984), *Methods of Numerical Integration*, 2nd edn., New York: Wiley and Sons.

Deng, D. and Paul, S.R. (2000), Score tests for zero inflation in generalized linear models, *Canadian Journal of Statistics*, 27, 563-570.

Diggle, P.J. and Gratton, R.J. (1984), Monte Carlo methods of inference for implicit statistical models (with discussion), *Journal of the Royal Statistical Society B*, 46, 193-227.

Diggle, P.J., Fiksel, T., Grabarnik, P., Ogata, Y., Stoyan, D. and Tanemura, M. (1994), On parameter estimation for pairwise interaction processes, *International Statistical Review*, 62, 99-117.

Du, Q., Faber, V., and Gunzburger, M. (1999), Centroidal Voronoi tessellations: applications and algorithms, *SIAM Review*, 41, 637-676.

Fan, J. and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, London: Chapman and Hall.

Goulard, M., Särkkä, A. and Grabarnik, P. (1996), Parameter estimation for marked Gibbs point processes through the maximum pseudo-likelihood method, *Scandinavian Journal of Statistics*, 23, 365-379.

Green, P.J. (1984), Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives (with discussion), *Journal of the Royal Statistical Society B*, 46, 149-192.

Hjort, N.L. and Jones, M.C. (1996), Locally parametric nonparametric density estimation, *Annals of Statistics*, 24, 1619-1647.

Huang, F. and Ogata, Y. (1999), Improvements of the maximum pseudo-likelihood estimators in various spatial statistical models, *Journal of Computational Graphics and Statistics*, 8, 510-530.

Huang, F. and Ogata, Y. (2002), Generalized pseudo-likelihood estimates for Markov random fields on lattice, *Annals of the Institute of Statistical Mathematics*, 54, 1-18.

Jensen, J.L. (1993), Asymptotic normality of estimates in spatial point processes, *Scandinavian Journal of Statistics*, 20, 97-109.

Lindsey, J.K. (1995), Fitting parametric counting processes by using log-

linear models, *Applied Statistics*, 44, 201-212.
Loader, C.R. (1996), Local likelihood density estimation, *Annals of Statistics*, 24, 1602-1618.
Mateu, J. and Montes, F. (2001), Likelihood inference for Gibbs processes in the analysis of spatial point patterns, *International Statistical Review*, 69, 81-104.
McCullagh, P. and Nelder, J.A. (1989), *Generalized Linear Models*, 2nd edn., London: Chapman and Hall.
Okabe, A., Boots, B., Sugihara, K. and Chiu, S.N. (2000), *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, 2nd edn., New York: Wiley and Sons.
Pallini, A. (2000), Resampling configurations of points through coding schemes, *Journal of the Italian Statistical Society*, 9, 1-24.
Pratt, J.W. (1981), Concavity of the log likelihood, *Journal of the American Statistical Association*, 76, 103-106.
Preston, C. (1976), *Random Fields*, Lectures Notes in Mathematics 534, Springer-Verlag, Berlin.
Ripley, B.D. (1977), Modelling spatial patterns (with discussion), *Journal of the Royal Statistical Society B*, 39, 172-212.
Ruppert, D., Sheather, S.J. and Wand, M.P. (1995), An effective bandwidth selector for local least squares regression, *Journal of the American Statistical Association*, 90, 1257-1270.
Ruppert, D. and Wand, M.P. (1994), Multivariate locally weighted least squares regression, *Annals of Statistics*, 22, 1346-1370.
Scott, D.W. (1992), *Multivariate Density Estimation*, New York: Wiley and Sons.
Wand, M.P. and Jones, M.C. (1995), *Kernel Smoothing*, London: Chapman and Hall.