Ann Inst Stat Math (2013) 65:387–412
DOI 10.1007/s10463-012-0376-7

Asymptotic Palm likelihood theory for stationary point processes

Michaela Prokešová · Eva B. Vedel Jensen

Received: 29 June 2011 / Revised: 4 June 2012 / Published online: 29 August 2012

Abstract In the present paper, we propose a Palm likelihood approach as a general estimating principle for stationary point processes in $\mathbf{R}^d$ for which the density of the second-order factorial moment measure is available in closed form or in an integral representation. Examples of such point processes include the Neyman–Scott processes and the log Gaussian Cox processes. The computations involved in determining the Palm likelihood estimator are simple. Conditions are provided under which the Palm likelihood estimator is strongly consistent and asymptotically normally distributed.

Keywords Asymptotic normality · Cluster processes · Consistency · Neyman–Scott processes · Log Gaussian Cox processes · Palm likelihood · Spatial point process · Strong mixing

1 Introduction

Estimation of parametric models for spatial point processes has been a very active research area in the last few years. Motivated by the need of analyzing always larger and more complicated data sets in a reasonably short time, several simulation-free estimation methods based on composite likelihood and/or estimating equations have been developed as alternatives to the computationally more demanding maximum likelihood and Bayesian methods; see Møller and Waagepetersen (2007) for a recent overview. Besides composite likelihood and estimating equations, there are

M. Prokešová (✉)
Department of Probability and Statistics, Faculty of Mathematics and Physics,
Charles University, Sokolovská 83, 186 75 Prague, Czech Republic
e-mail: prokesov@karlin.mff.cuni.cz

E. B. V. Jensen
Department of Mathematics, Centre for Stochastic Geometry and Advanced Bioimaging,
Aarhus University, Ny Munkegade 118, Building 1530, 8000 Aarhus C, Denmark

approximate Bayesian methods for latent Gaussian models including the log Gaussian Cox process (Rue et al. 2009).

In the present paper, we will focus on point process models for which the densities of the first- and second-order moment measures (and/or quantities derived from them—like the pair-correlation function g or the K-function) are available—either in closed form or in an integral representation. In particular, we will consider stationary cluster processes, Cox processes and related models.

One of the estimation methods that was first suggested for such processes is the minimum contrast method based either on the K-function or the g-function; see Diggle (2003); Møller et al. (1998); Møller and Waagepetersen (2003) and references therein. In addition to stationarity, this method requires that the point process X is isotropic. A parameter θ is estimated by minimizing the discrepancy measure

$$\int_0^R [\hat{K}^c(u) - K^c(u;\theta)]^2 \mathrm{d}u \qquad \text{or} \qquad \int_0^R [\hat{g}^c(u) - g^c(u;\theta)]^2 \mathrm{d}u$$

between the estimate $\hat{K}$ or $\hat{g}$ and its theoretical value $K(\cdot;\theta)$ or $g(\cdot;\theta)$, respectively. Note that $\hat{g}$ involves nonparametric density estimation. The user-specified constants c and R are used to control the sampling fluctuations in the estimates of K and g. These constants are usually chosen in some ad hoc manner. Asymptotic properties of these estimates have been derived in Guan and Sherman (2007) and Heinrich (1992).

An alternative estimation method based on maximization of the so-called Palm likelihood is suggested in Tanaka et al. (2008) for stationary cluster processes, including Neyman–Scott processes and related models. The Palm likelihood makes use of the process of differences

$$\{x - y : x \neq y \in X \cap W\},$$

where W is the observation window. This likelihood depends on the so-called Palm intensity function λ_0 which is the density of the second-order reduced factorial moment measure of the process X. In the original paper Tanaka et al. (2008), it was assumed that the point process X is an isotropic point process in $\mathbf{R}^2$, but Palm likelihood can actually be applied to any simple stationary point process in $\mathbf{R}^d$ with Palm intensity λ_0. Palm likelihood estimation has thereby a much wider applicability than originally anticipated in Tanaka et al. (2008).

The Palm likelihood estimation method belongs to the second-order moment estimation methods, since

$$\lambda^{(2)}(x,y) = \lambda\lambda_0(y-x), \qquad x, y \in \mathbf{R}^d,$$

where $\lambda^{(2)}$ is the density of the second-order factorial moment measure of the point process X. A number of related second-order methods based on a composite likelihood approach are available in the literature (Baddeley and Turner 2000; Guan 2006; Møller and Waagepetersen 2007; Waagepetersen 2007); see Lindsay (1988) for an introduction to composite likelihood. The Palm likelihood estimation method is closely related to the composite likelihood method suggested in Waagepetersen (2007); see Sect. 3

below. The Palm likelihood method is numerically simpler than the composite likelihood method in Guan (2006) because of a simpler form of the normalization term.

Concerning the asymptotic properties of Palm likelihood estimators, it was argued in Tanaka et al. (2008) that, for the considered cluster point processes, the process of differences is well approximated by a nonstationary Poisson point process with intensity $|X \cap W|\,\lambda_0$, because the process of differences can be regarded as a superposition of $|X \cap W|$ realizations of a point process with the distribution equal to the Palm distribution of the original process X—that means with the intensity λ_0. By a superposition theorem for $|X \cap W| \to \infty$, a convergence of the suitably normalized difference process to a Poisson process can be obtained (see Ogata and Katsura 1991 for the argument), which would imply the consistency of the obtained Palm likelihood estimates. A formal proof of consistency was, however, not provided.

The present paper fills this gap. We provide a proof of strong consistency and asymptotic normality of the maximum Palm likelihood estimator. Consistency is proved under the assumption of ergodicity of the point process X. The proof of asymptotic normality is provided under the additional assumptions that the process X is strongly mixing and the strong mixing coefficient decays sufficiently fast. Moreover for cluster processes, we derive some simple methods of checking sufficient conditions for the desired fast decay of the strong mixing coefficients.

The paper is organized as follows. We give the necessary notation and background information in Sect. 2 and introduce the Palm likelihood estimation procedure in detail in Sect. 3. This section contains the extension of the Palm likelihood from isotropic point processes to general stationary processes and a discussion of computational issues relating to the anisotropic case. In Sects. 4 and 5, the main results of the paper are presented— the strong consistency and the asymptotic normality of the Palm likelihood estimator. The obtained results are further exemplified in Sect. 6. Proofs are deferred to an Appendix.

2 Background

Let X denote a simple strictly stationary point process on $\mathbf{R}^d$. In the sequel, $\mathcal{B}^d$ is the Borel σ-algebra on $\mathbf{R}^d$, $|A|$ is the volume of the set $A \in \mathcal{B}^d$, ∂A its boundary and $|\partial A|$ the $(d-1)$-dimensional surface measure of ∂A, when it exists. Generally, we use $|\cdot|$ for the appropriate Hausdorff measure. The origin in $\mathbf{R}^d$ is denoted by o, $B(x, R)$ is the ball centered at $x \in \mathbf{R}^d$ with radius $R > 0$ and $\oplus$, $\ominus$ denotes Minkowski addition and substraction, respectively, with the convention that $A \oplus R = A \oplus B(o, R)$ and $A \ominus R = A \ominus B(o, R)$ for $R > 0$. The Euclidean norm of the vector x is denoted by $|x|$, for matrices we use also the Euclidean norm $|M| = (\text{trace}\,(M^T M))^{\frac{1}{2}}$ and I is the indicator function.

The k-th order factorial moment measure $\alpha^{(k)}$ of the point process X is defined by the following equation

$$\int_{(\mathbf{R}^d)^k} f(u_1, \ldots, u_k)\, \alpha^{(k)}(\mathrm{d}(u_1, \ldots, u_k)) = E\left(\sum_{u_1,\ldots,u_k \in X}^{\neq} f(u_1, \ldots, u_k) \right) \tag{1}$$

for any non-negative, Borel measurable function f on $(\mathbf{R}^d)^k$, where the summation runs over k-tuples of distinct points of X. The kth-order factorial cumulant measure $\gamma^{(k)}$ of X is a locally finite signed measure on $[(\mathbf{R}^d)^k, \mathcal{B}^{dk}]$ which is formally connected with the measures $\alpha^{(1)}, \ldots, \alpha^{(k)}$ by

$$\gamma^{(k)}(\times_{i=1}^k A_i) = \sum_{j=1}^k (-1)^{j-1}(j-1)! \sum_{K_1 \cup \cdots \cup K_j = \{1,\ldots,k\}} \prod_{i=1}^j \alpha^{(\#K_i)} \left(\times_{k_i \in K_i} A_{k_i}\right)$$

for bounded $A_1, \ldots, A_k \in \mathcal{B}^d$, where the inner sum is taken over all partitions of the set $\{1, \ldots, k\}$ in disjoint non-empty subsets $K_1, \ldots, K_j$. In particular, $\alpha^{(1)}(A) = \gamma^{(1)}(A) = \lambda\,|A| = E|X \cap A|$ for $A \in \mathcal{B}^d$ where λ is called the intensity of X. For $k \geq 2$, we will assume in the sequel that the factorial moment measures have densities $\lambda^{(k)}$ with respect to the Lebesgue measure on $\mathbf{R}^{dk}$. These densities are called the k-th order product densities of X (or sometimes k-th order intensity functions of X).

Since, for any $k \geq 2$, $\alpha^{(k)}$ is invariant under diagonal shifts, there exists a corresponding reduced kth-order factorial moment measure $\alpha_{\text{red}}^{(k)}$ on $[(\mathbf{R}^d)^{k-1}, \mathcal{B}^{d(k-1)}]$ which is uniquely determined by the disintegration formula

$$\begin{aligned} &\int_{(\mathbf{R}^d)^k} f(u_1, \ldots, u_k)\alpha^{(k)}(\mathrm{d}(u_1, \ldots, u_k)) \\ &= \lambda \int_{\mathbf{R}^d} \int_{(\mathbf{R}^d)^{k-1}} f(u_1, u_2 + u_1, \ldots, u_k + u_1)\, \alpha_{\text{red}}^{(k)}(\mathrm{d}(u_2, \ldots, u_k))\, \mathrm{d}u_1, \end{aligned} \tag{2}$$

where f is as in (1). Similarly we may define the reduced kth-order factorial cumulant measure $\gamma_{\text{red}}^{(k)}$, which turns out to be a signed measure on $[(\mathbf{R}^d)^{k-1}, \mathcal{B}^{d(k-1)}]$.

For $k = 2$, the disintegration (2) implies that

$$\lambda^{(2)}(x, y) = \lambda\lambda_0(y - x), \qquad x, y \in \mathbf{R}^d,$$

where the function λ_0 is the density of $\alpha_{\text{red}}^{(2)}$. The function λ_0 is also called the conditional intensity or Palm intensity in the literature, since λ_0 is in fact the intensity function of the Palm distribution $\mathcal{P}_0$ of the original point process X. For a detailed introduction to these notions and their properties, we refer the reader to Daley and Vere-Jones (2003).

Two popular point process characteristics can be defined from λ_0, viz. the pair correlation function

$$g(x, y) = g(y - x) = \lambda_0(y - x)/\lambda, \quad x, y \in \mathbf{R}^d,$$

and the K-function defined by

$$\lambda\, K(R) = \int_{B(o,R)} \lambda_0(u)\mathrm{d}u = E\left[|X \cap B(o, R)\backslash\{o\}| \,|\, X \cap \{o\} \neq \emptyset\right].$$

Note that $\lambda K(R)$ can be interpreted as the mean number of further points from X in $B(x, R)$ centered at a typical point x of the point process X.

The Neyman–Scott process considered in Tanaka et al. (2008) can be constructed as follows. Let C be a stationary Poisson point process of intensity μ. This process is called the mother process. Each mother point $c \in C$ produces a random number M of daughter points with mean ν. The daughters around c are i.i.d. with density $k(c, \cdot) = h(\cdot - c)$. The set of daughters associated with the mother c is denoted by X_c. The Neyman–Scott process is then the union of the daughter clusters $X = \cup_{c \in C} X_c$. The intensity of X is $\mu\nu$. The Palm intensity of the Neyman–Scott process becomes

$$\lambda_0(z) = \mu\nu + \nu \int_{\mathbf{R}^d} h(u)h(z+u)\mathrm{d}u, \quad z \in \mathbf{R}^d; \tag{3}$$

see [Møller and Waagepetersen 2003, (5.8)]. If h is a Gaussian density, then (3) is in closed form. For other examples of usable kernels like the spherical or Matérn, see e.g. Jonsdottir et al. (2011). Further examples of Neyman–Scott processes are given in Tanaka et al. (2008).

Another class of point processes for which the Palm intensity can be obtained in closed form is the log Gaussian Cox processes (Møller et al. 1998). Here,

$$\lambda_0(z) = \exp(m + c(0)/2 + c(z)), \quad z \in \mathbf{R}^d,$$

where m and c are the mean and covariance function of the underlying stationary Gaussian random field, respectively.

3 Palm likelihood for stationary point processes

Let us assume that the parameter of interest of our point process model is the (vector) parameter θ and that the Palm intensity $\lambda_0(\cdot; \theta)$ is parametrized by θ. In the following, we will suppress θ in the notation if the dependence on θ is not important in the respective context.

In this section, we will discuss the Palm likelihood method that was introduced in Tanaka et al. (2008) for stationary isotropic Neyman–Scott processes and related models. The Palm log-likelihood function is for an arbitrary simple stationary point process in $\mathbf{R}^d$ defined by

$$\log L_P(\theta) = \sum_{\substack{x,y \in X \cap W \\ |x-y|<R}}^{\neq} \log \lambda_0(x-y; \theta) - |X \cap W| \int_{\mathbf{R}^d} I(|u| < R)\lambda_0(u; \theta)\mathrm{d}u, \tag{4}$$

where R is a chosen positive constant. The maximum Palm likelihood (MPL) estimator is obtained by maximizing $L_P(\theta)$.

Under the assumption that $L_P(\theta)$ is differentiable with respect to θ, the MPL-estimate is the solution to the following estimation equation

$$\frac{\mathrm{d}\ \log\ L_P(\theta)}{\mathrm{d}\theta} = 0. \tag{5}$$

The idea behind this estimation procedure is to use, instead of the original process X observed in the window W, the process of differences $Y = \{x - y : x \neq y \in X \cap W\}$. Note that the data used in the Palm likelihood are really only the difference process Y and the number of observed points $|X \cap W|$. For any fixed $x \in X$ let

$$Y_x = \{y - x : x \neq y \in X\}, \tag{6}$$

then the Palm log-likelihood in (4) is a sum (over $x \in X \cap W$) of Poisson log-likelihoods for the processes $Y_x \cap B(o, R)$, all assumed to have intensity function λ_0. Using the Poisson log-likelihoods implies that the higher-order interactions in the processes of differences are ignored. Furthermore, since the Poisson log-likelihoods are summed, the dependence among $Y_x,\ x \in X \cap W$ are ignored by treating them as independent replications.

An alternative way of arriving at the Palm log-likelihood (4) is as follows. Consider $Y(R) = Y \cap B(o, R)$, a point process contained in $B(o, R)$. The intensity function of this point process can be derived as follows. Let A be a Borel subset of $B(o, R)$. Then,

$$E(|Y(R) \cap A|) = \int_W \int_W I(y - x \in A)\lambda\lambda_0(y - x)\mathrm{d}x\mathrm{d}y = \int_A \gamma_W(u)\lambda\lambda_0(u)\mathrm{d}u,$$

where $\gamma_W(u) = |W \cap (W + u)|$ is the set covariance of the window W; see Stoyan et al. (1995, p. 126) for further details. The point process $Y(R)$ has thus an intensity function concentrated on $B(o, R)$ of the form

$$\lambda_R(u) = \gamma_W(u)\lambda\lambda_0(u), \quad u \in B(o, R). \tag{7}$$

The Palm log-likelihood (4) can now be obtained by treating $Y(R)$ as an inhomogeneous Poisson process, replacing the intensity λ of the original point process X by the observed intensity $|X \cap W|/|W|$ and approximating $\gamma_W(u),\ u \in B(o, R)$, by $|W|$. This is a reasonable approximation for R, substantially smaller than the size of the observation window W.

As mentioned in the introduction, the Palm likelihood estimation method is closely related to the composite likelihood method suggested in Waagepetersen (2007). In fact, if we, in the last equation on p. 256 in Waagepetersen (2007), replace λ by $|X \cap W|/|W|$ and approximate $\gamma_W(u)$, $u \in B(o, R)$, by $|W|$, then the resulting estimating equation is equal to (4).

We can define a modified version of the Palm likelihood in which we consider only those points $x \in X$ for which $B(x, R) \subseteq W$, thus employing the inner region edge correction (minus sampling)

$$\log L_{PU}(\theta) = \sum_{\substack{x \in X \cap (W \ominus R) \\ y \in X, (y-x) \in B(o,R)}}^{\neq} \log \lambda_0(y-x;\theta) - |X \cap (W \ominus R)| \int_{B(o,R)} \lambda_0(u;\theta)\mathrm{d}u. \tag{8}$$

Then, the estimating (vector) equation $\frac{\mathrm{d} \log L_{PU}}{\mathrm{d}\theta} = 0$ is an unbiased estimating equation, since for $\tilde{Y} = \bigcup_{x \in W \ominus R} Y_x$ Eq. (7) becomes $\tilde{\lambda}_R(u) = |W \ominus R|\lambda\lambda_0(u)$, and consequently $E_{\theta_0}\left(\frac{\mathrm{d} \log L_{PU}(\theta)}{\mathrm{d}\theta}\Big|_{\theta=\theta_0}\right) = 0$, where E_{θ_0} denotes the mean value with respect to the distribution with the correct parameter value θ_0.

Since the difference between (4) and (8) is only in the employed edge correction (inner region or none edge correction), i.e. only in the way the points $x \in W \setminus (W \ominus R)$ are handled, for windows W large enough with respect to R the difference in the two estimates will be negligible. Under the assumptions on the sequences of observation windows that will be introduced in Sect. 4, the correctly normalized estimating equation $\frac{1}{|W|}\frac{\mathrm{d} \log L_P(\theta)}{\mathrm{d}\theta} = 0$ will be an asymptotically unbiased estimating equation.

Sometimes, in the literature another type of edge correction is used—namely the periodic boundary correction; see e.g. Illian et al. (2008, Sect. 4.2.2). It is applicable to rectangular observation windows only and also leads to a biased estimating equation in general.

From a practical point of view, it is important to have some guidelines for the choice of R. For a point process with interaction radius ρ, we have $\lambda_0(u) = \lambda$ for $|u| > \rho$. Consequently, the estimates of the interaction parameters we obtain by using $R > \rho$ cannot be better than those obtained by using ρ, since we gain no information by increasing R beyond ρ. For instance, for the Palm log-likelihood $\log L_P$ using $R > \rho$, we have

$$\log L_P(\theta) = \left[\sum_{\substack{x,y \in X \cap W \\ |x-y| < \rho}}^{\neq} \log \lambda_0(y-x;\theta) - |X \cap W| \int_{\mathbf{R}^d} I(|u| < \rho)\lambda_0(u;\theta)\mathrm{d}u\right] + f(\lambda, \rho, R),$$

where the first term is the Palm log-likelihood using $R = \rho$ and the last term depends on the parameters θ only through the intensity λ of the process. We therefore recommend that R should not be larger than an estimate of the range of interaction we get from the data (e.g. by an empirical pair-correlation function). It should also be noticed that the mean number of points in the difference process $\tilde{Y}$,

$$E|\tilde{Y}| = |W \ominus R|\,\lambda \int_{B(o,R)} \lambda_0(u)\mathrm{d}u,$$

is not necessarily a monotone function of R. For instance, for $d = 2$, $W = [0,1]^2$ and $\lambda_0(u) = \lambda[1 + \tau I(|u| < \rho)]$, we get for $R \le \min(\rho, \frac{1}{2})$

$$E|\tilde{Y}| = (1-2R)^2 \pi R^2 \lambda^2 (1+\tau).$$

This function increases to its maximum at $R = \frac{1}{4}$ and then decreases. A rule of thumb could therefore be to choose R no larger than one quarter of the minimal side length of the observation window W.

We would like to end this section by stressing that the assumption of isotropy, made in the original paper Tanaka et al. (2008) for computational reasons, is not necessary for the formulation or validity of the MPL estimation method. The Palm likelihood (8) is formulated for the Palm intensity $\lambda_0(u)$ as a function of the whole vector u, not just as a function of its length $|u|$.

Palm likelihood estimation is computationally tractable for Neyman–Scott processes with an anisotropic daughter distribution density $h(u)$. The simple and yet flexible example of a Gaussian density h in $\mathbf{R}^2$ with a general covariance matrix Σ is discussed in Sect. 6. The advantage of the Gaussian density is that the Palm intensity λ_0 computed by (3) has a closed form. Nevertheless, even if λ_0 is not available in closed form (the density h may, for instance, be the indicator of an ellipse), λ_0 can be computed numerically by (3) for any u. The maximization of the Palm likelihood (8) can then be implemented by e.g. the Nelder–Mead maximization algorithm (Nelder and Mead 1965).

Furthermore, the method is not restricted to Neyman–Scott processes and the other cluster processes discussed in Tanaka et al. (2008). The method can be used for estimation in any parametric model with an accessible form of the Palm intensity $\lambda_0(\cdot;\theta)$. One very important class of such processes not considered in Tanaka et al. (2008) is the log Gaussian Cox processes; see Møller et al. (1998). Palm likelihood estimation in log Gaussian Cox processes with exponential covariance is discussed in Sect. 6.

4 Strong consistency of MPLE

We will assume in the sequel that the point process model is parametrized by $\theta \in \Theta$, where Θ is a compact subset of $\mathbf{R}^q$ with non-empty interior. The true vector parameter θ_0 is assumed to be an interior point of Θ.

The asymptotics will be studied under an increasing domain setting assuming that we have a convex averaging sequence of windows $\{W_n\}_{n\in\mathbf{N}}$—i.e. that all the windows W_n are bounded convex sets, $W_n \subseteq W_{n+1}$ for all n and the inradii

$$\rho(W_n) = \sup\{\rho : W_n \text{ contains a ball of radius } \rho\}$$

converge to ∞ as $n \to \infty$; see Daley and Vere-Jones (2003, Chapter 10) for further details. We will assume that $|W_n| = \mathcal{O}(\rho(W_n)^d)$. This implies for the convex sets $\{W_n\}$ that $|\partial W_n| = \mathcal{O}(\rho(W_n)^{d-1})$ since according to Wills (1970) we have

$$\frac{|\partial W|}{|W|} \leq \frac{d}{\rho(W)},$$

if $W \subset \mathbf{R}^d$ is a convex set.

We will start by showing the consistency of the unbiased version of the log Palm likelihood L_{PU}. Let us write here in detail the score function $U(\theta) = \frac{1}{|W \ominus R|} \frac{\mathrm{d} \log L_{PU}(\theta)}{\mathrm{d}\theta}$ of the Palm likelihood

$$U(\theta) = \frac{1}{|W \ominus R|} \sum_{\substack{x \in X \cap (W \ominus R) \\ y \in X, (y-x) \in B(o,R)}}^{\neq} \frac{\mathrm{d}\,\lambda_0(y-x;\theta)}{\mathrm{d}\theta} \frac{1}{\lambda_0(y-x;\theta)} - \frac{|X \cap (W \ominus R)|}{|W \ominus R|} \frac{\mathrm{d} \int_{B(o,R)} \lambda_0(u;\theta)\mathrm{d}u}{\mathrm{d}\theta}. \tag{9}$$

(the score function computed from observations in the window W_n will be denoted by $U_n(\theta)$). We obtain

$$E_{\theta_0}\left(\sum_{\substack{x \in X \cap (W \ominus R) \\ y \in X, (y-x) \in B(o,R)}}^{\neq} \frac{\mathrm{d}\,\lambda_0(y-x;\theta)}{\mathrm{d}\theta} \frac{1}{\lambda_0(y-x;\theta)} \right) = \int_{W \ominus R} \lambda \int_{B(o,R)} \frac{\mathrm{d}\,\lambda_0(u;\theta)}{\mathrm{d}\theta} \frac{1}{\lambda_0(u;\theta)} \lambda_0(u;\theta_0)\mathrm{d}u\mathrm{d}x,$$

where E_{θ_0} denotes the mean value with respect to the distribution of the point process with $\theta = \theta_0$. We see that $E_{\theta_0} U(\theta_0) = 0$ for the true parameter value θ_0. Thus, the estimating equation $U_n(\theta) = 0$ is indeed unbiased. The Palm likelihood estimate obtained from this equation will be denoted by $\hat{\theta}_n$.

In the following theorem, the strong consistency of $\hat{\theta}_n$ is formulated. The proof of the theorem may be found in the Appendix.

Theorem 1 *Let X be a stationary ergodic point process observed in a convex averaging sequence $\{W_n\}_{n \in \mathbf{N}}$ of windows for which $|W_n| = \mathcal{O}(\rho(W_n)^d)$ holds. Assume that $E_{\theta_0} U_n(\theta) = 0$ only when $\theta = \theta_0$ and that $\frac{\mathrm{d}\,\lambda_0(u;\theta)}{\mathrm{d}\theta} \frac{1}{\lambda_0(u;\theta)}$ and $\frac{\mathrm{d}\,(\int_{B(o,R)} \lambda_0(u;\theta)du)}{\mathrm{d}\theta}$ are bounded and continuous* (*with respect to u and θ, and θ respectively*). *Then, $\hat{\theta}_n$ is a strongly consistent estimate of θ_0, i.e. $\hat{\theta}_n \to \theta_0$ P_{θ_0}-a.s.*

Remark 1 The boundedness and continuity conditions of Theorem 1 are satisfied if $\lambda_0(u;\theta)$ and $\frac{\mathrm{d}\,\lambda_0(u;\theta)}{\mathrm{d}\theta} \frac{1}{\lambda_0(u;\theta)}$ are bounded and continuous with respect to u and θ, and $\lambda_0(u;\theta)$ is bounded from 0 uniformly in u and θ. These conditions are easy to check and satisfied for a wide range of processes, including the Thomas process, all the generalizations of the Thomas process from Tanaka et al. (2008, Section 2.1), the anisotropic Thomas process from Sect. 6 and the Neyman–Scott processes with spherical or Matérn kernels. The assumption of boundedness of $\frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta}$ in u (even of $\lambda_0(u;\theta)$) is not fulfilled for the so-called inverse power type model from Tanaka et al. (2008) and Theorem 1 cannot be applied in this special case.

Let us now discuss the original Palm log-likelihood (4) without any included edge correction. The associated score function $\tilde{U}(\theta) = \frac{1}{|W|}\frac{\mathrm{d}\log L_P(\theta)}{\mathrm{d}\theta}$ takes the form

$$\tilde{U}(\theta) = \frac{1}{|W|} \sum_{\substack{x,y\in X\cap W\\ |x-y|<R}}^{\neq} \frac{\mathrm{d}\,\lambda_0(y-x;\theta)}{\mathrm{d}\theta} \frac{1}{\lambda_0(y-x;\theta)} - \frac{|X\cap W|}{|W|} \frac{\mathrm{d}\int_{B(o,R)}\lambda_0(u;\theta)\mathrm{d}u}{\mathrm{d}\theta}.$$

Let $\tilde{\theta}_n$ denote the estimate obtained from the estimating equation $\tilde{U}_n(\theta) = 0$. For $\theta = \theta_0$, we have

$$E_{\theta_0}\left(\sum_{\substack{x,y\in X\cap W\\ |x-y|<R}}^{\neq} \frac{\mathrm{d}\lambda_0(y-x;\theta)}{\mathrm{d}\theta} \frac{1}{\lambda_0(y-x;\theta)} \right) = \lambda \int_{B(o,R)} \gamma_W(u) \frac{\mathrm{d}\,\lambda_0(u;\theta)}{\mathrm{d}\theta}\bigg|_{\theta=\theta_0} \mathrm{d}u.$$

As a consequence, we do not have $E_{\theta_0}\tilde{U}_n(\theta_0) = 0$. Nevertheless since

$$\sup_{z\in B(o,R)} \left| \frac{\gamma_{W_n}(z)}{|W_n|} - 1 \right| \le \frac{R|\partial W_n|}{|W_n|} \le \frac{\mathrm{d}R}{\rho(W_n)}, \tag{10}$$

$\frac{\gamma_{W_n}(u)}{|W_n|}$ converges uniformly to 1 on the compact set $B(o, R)$ and therefore $E_{\theta_0}\tilde{U}_n(\theta_0) \to E_{\theta_0}U_n(\theta_0) = 0$ as $n \to \infty$. Therefore, the proof of the consistency of $\tilde{\theta}_n$ is analogous to that of $\hat{\theta}_n$.

Theorem 2 *Under the same assumptions as in Theorem 1, $\tilde{\theta}_n$ is a strongly consistent estimate of θ_0.*

The proof of Theorem 2 may be found in the Appendix.

5 Asymptotic normality of MPLE

We will now show the asymptotic normality of the MPLE under the assumption that the point process X is strongly mixing (cf. Heinrich 2012). Recall that for two σ-algebras $\mathcal{F}_1, \mathcal{F}_2$ defined on the same probability space the strong mixing coefficient is defined by

$$\alpha(\mathcal{F}_1, \mathcal{F}_2) = \sup\{|P(A_1\cap A_2) - P(A_1)P(A_2)| : A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}.$$

For a stationary point process X, the strong mixing coefficient $\alpha(p; k)$ quantifies the dependence between the behavior of the point process on sets of volume at most p separated by a distance larger than or equal to k. Thus for a point process X and $p, k \ge 0$, we define

$$\alpha(p;k) = \sup\{\alpha(\mathcal{F}^X(A), \mathcal{F}^X(B)) : d(A,B) \ge k, |A| \le p, |B| \le p\}, \tag{11}$$

where $\mathcal{F}^X(A)$ denotes the σ-algebra generated by $X \cap A$ and the supremum is taken over all measurable subsets $A,\ B$ in $\mathcal{B}^d$.

We will assume that

$$\sup_{p\geq 0} \frac{\alpha(p;k)}{\max(p,1)} = \mathcal{O}(k^{-\epsilon}) \quad \text{for some} \quad \epsilon > d. \tag{12}$$

For the class of log Gaussian Cox processes (Møller et al. 1998), this mixing condition is implied by the mixing condition for the driving field, which have been treated in the literature; see Doukhan (1994). For instance, the condition (12) is satisfied if the correlation function of the underlying Gaussian field decays at a polynomial rate faster then $d + \epsilon$ and has a spectral density which is bounded below. This follows from Doukhan (1994, Corollary 2). A concrete example of such correlation functions often used in practice is the class of Matérn correlation functions (including also the exponential correlation function), see e.g. Stein (1999, Section 2.7).

Concerning the Neyman–Scott processes, (12) is obviously satisfied for Neyman–Scott processes with a kernel density $k(c,\cdot)$ with bounded support (e.g. the Matérn cluster process with the spherical kernel k). In the following lemma, we show that (12) is also satisfied if the density has polynomially decaying tails of order $d + \epsilon$. This condition is satisfied for all the processes considered in Tanaka et al. (2008), the anisotropic Thomas process from Sect. 6 as well as Neyman–Scott processes with Matérn kernels. The proof of the lemma can be found in the Appendix.

Lemma 1 *Let X be a Neyman–Scott process with mother intensity μ and mean number ν of daughter points in a cluster. Let the daughter points around a mother point at the location c be distributed according to the kernel density $k(c,\cdot)$. If there exists a function h such that $k(c,x) = h(x-c)$ and $h(v) = \mathcal{O}(|v|^{-\epsilon-d})$ as $|v| \to \infty$, then* $\frac{\alpha(p;k)}{\max(p,1)} \leq \mathcal{O}(k^{-\epsilon})$.

Remark 2 In the literature on point processes, see e.g. Guan (2006); Heagerty and Lumley (2000); Politis and Sherman (2001), an alternative weaker version of the strong mixing coefficient is sometimes used

$$\alpha(p;k) = \sup\{\alpha(\mathcal{F}^X(A), \mathcal{F}^X(B)) : A = B + x, d(A,B) \geq k, |A| = |B| \leq p\},$$

$p > 0$, where the supremum is taken over all compact, convex sets A and all $x \in \mathbf{R}^d$. This version of the strong mixing coefficient has been inspired by the strong mixing coefficient used in the classical paper Rosenblatt (1956). In the proof of the asymptotic normality below, we follow the methods of Guan (2006) (described in detail in Guan et al. 2007), based on the blocking technique presented in Ibragimov and Linnik (1971). In our proof, we need to use the mixing coefficient for two sets A,B where A is a union of disjoint cubes (see the proof in the Appendix) and as such definitely not convex. Furthermore, it is not possible to find an x such that $B \subseteq A + x$ and the distance between the sets A and $A + x$ is larger than the desired value. Thus, the more general version (11) of the strong mixing coefficient must be used. This problem was not fully acknowledged in the papers Guan (2006) and Guan et al. (2007).

The Neyman–Scott and log Gaussian Cox processes with suitably mixing driving field (as described above) satisfy (12) for either definition of the mixing coefficient $\alpha(p; k)$. Thus, from a practical point of view, the definition (11) is not restrictive.

For the asymptotic normality of the MPL estimate, we will further assume a mild moment condition on $U_n(\theta_0)$:

$$\sup_{n\in\mathbf{N}} E_{\theta_0}(|\sqrt{|W_n \ominus R|}|U_n(\theta_0)|^q) < C_q < \infty \text{ for some } q > 2. \tag{13}$$

This condition is slightly stronger than the existence of the standardized asymptotic variances of $U_n(\theta)$; see also Guan (2006, p. 1505). It is satisfied for example for the processes which have the first six reduced cumulant moment measures of finite total variation provided $\frac{\mathrm{d}\,\lambda_0(x;\theta)}{\mathrm{d}\theta}\frac{1}{\lambda_0(x;\theta)}$ is bounded for $x \in B(o, R)$ and $\theta \in \Theta$. In particular, the class of Brillinger-mixing processes (i.e. processes for which the reduced cumulant moment measures of all orders have finite total variation) obviously fulfill this condition. The Brillinger-mixing processes include among others Neyman–Scott processes for which the distribution of the size (i.e. number of points) of the cluster has finite moments of all orders, as it is the case for all examples in Tanaka et al. (2008). For further examples and discussions of Brillinger-mixing, see Heinrich (1988).

Below, we present the theorem concerning asymptotic normality of the Palm likelihood estimator. The proof of the theorem can be found in the Appendix.

Theorem 3 *Assume that the conditions of Theorem 1 are satisfied and moreover that (12) and (13) hold. Then the reduced factorial cumulant measures of X up to fourth order have finite total variation and*

$$\sup_{\substack{u\in B(o,R)\\ |\theta_1-\theta_2|<\delta}} \left| \frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\mathrm{d}\,\lambda_0(u;\theta)}{\mathrm{d}\theta}\frac{1}{\lambda_0(u;\theta)}\right)\bigg|_{\theta=\theta_1} - \frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta}\frac{1}{\lambda_0(u;\theta)}\right)\bigg|_{\theta=\theta_2}\right|$$
$$\to 0 \quad as\, \delta \to 0 \tag{14}$$

and

$$\sup_{|\theta_1-\theta_2|<\delta} \left| \frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\left(\int_{B(o,R)} \lambda_0(u;\theta)du\right)\bigg|_{\theta=\theta_1} - \frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\left(\int_{B(o,R)} \lambda_0(u;\theta)du\right)\bigg|_{\theta=\theta_2}\right|$$
$$\to 0 \quad as\, \delta \to 0. \tag{15}$$

Then, $\lim_{n\to\infty} |W_n \ominus R| Var_{\theta_0}(U_n(\theta)) = \Sigma(\theta)$ *exists and does not depend on the convex averaging sequence* W_n *and* $\sqrt{|W_n \ominus R|}(\hat{\theta}_n - \theta_0)$ *converges to a normal distribution with zero mean vector and covariance matrix* $M^{-1}\Sigma(\theta_0)M^{-1}$ *where*

$$M = \lambda \int_{B(o,R)} \frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta}\frac{1}{\lambda_0(u;\theta)}\right)\bigg|_{\theta=\theta_0} \lambda_0(u;\theta_0)du$$
$$- \lambda \frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\left(\int_{B(o,R)} \lambda_0(u;\theta)du\right)\bigg|_{\theta=\theta_0}. \tag{16}$$

Remark 3 By similar reasoning as in Remark 1, we get that if $\frac{d^2\lambda_0(u;\theta)}{d\theta^2}$ is bounded and continuous in both u and θ, then (14) and (15) are fulfilled. This simplified boundedness and continuity condition is satisfied in particular by all the processes from Remark 1.

Remark 4 The matrix $\Sigma(\theta)$ can be expressed as a sum of mixed integrals of $\frac{d\lambda_0(u;\theta)}{d\theta}\frac{1}{\lambda_0(u;\theta)}$ and $I_{B(o,R)}(u)$ with respect to the reduced factorial cumulant measures $\gamma_{\mathrm{red}}^{(k)}$, $k = 2, 3, 4$ and the Lebesgue measure. The higher-order moment measures $\gamma_{\mathrm{red}}^{(3)}$ and $\gamma_{\mathrm{red}}^{(4)}$ are typically not available in a feasible form and thus the theoretical expression for the variance matrix $\Sigma(\theta_0)$ seems to be of limited practical use. For evaluating the efficiency of the estimator $\hat{\theta}_n$, an estimate of the variance matrix $\Sigma(\theta_0)$ can be determined by means of simulation. We can produce independent realizations X_l, $l \in \{1, \ldots, N\}$ from the fitted model and approximate $\Sigma(\theta_0)$ by the sample variance matrix of the score functions $U_n(\hat{\theta}_n)$ computed for each of the replications X_l. If the original data are large enough, we can use subsampling methods for the estimation of $\Sigma(\theta_0)$ (see e.g. Heagerty and Lumley 2000; Politis and Sherman 2001 for further information).

Remark 5 An alternative proof of asymptotic normality could be based on the central limit theorem from Bolthausen (1982) for stationary α-mixing random fields. The required mixing assumptions would be more restrictive in this case.

Let us finish this section with a discussion of the original log Palm likelihood (4) without any included edge correction. The estimate $\tilde{\theta}_n$ is derived from the estimating equation $\tilde{U}(\theta) = \frac{1}{|W|}\frac{d \log L_P(\theta)}{d\theta} = 0$. It follows from the proof of Theorem 3 that if $\sqrt{|W_n|}\,\tilde{U}_n(\theta_0)$ converges in distribution to $N(a(\theta_0), Q(\theta_0))$, then $\sqrt{|W_n|}(\tilde{\theta}_n - \theta_0)$ converges to $N(M^{-1}a(\theta_0), M^{-1}Q(\theta_0)M^{-1})$ with the same M as in Theorem 3. By repeating the proof of Theorem 3 step by step, it moreover follows that $Q(\theta_0) = \Sigma(\theta_0)$, i.e. the asymptotic variance of $\sqrt{|W_n \ominus R|}\,U_n(\theta_0)$ and $\sqrt{|W_n|}\,\tilde{U}_n(\theta_0)$ is the same and that $\sqrt{|W_n|}(\tilde{U}_n(\tilde{\theta}_n) - E_{\theta_0}\tilde{U}_n(\tilde{\theta}_n))$ converges in distribution to $N(0, \Sigma(\theta_0))$.

However, it follows from the discussion at the end of Sect. 4 that the bias of $\tilde{U}_n(\theta_0)$ is of order $\mathcal{O}(\frac{\partial W_n}{|W_n|}) = \mathcal{O}(\rho_n^{-1})$, which is too large for convergence to 0 when multiplied by the normalization term $\sqrt{|W_n|} = \mathcal{O}(\rho_n^{d/2})$. Thus, for $\tilde{\theta}_n$, we cannot establish a result of the type presented in Theorem 3.

6 Examples

This section discusses in detail the Palm likelihood estimation procedure for two examples not considered in Tanaka et al. (2008).

6.1 Example 1: anisotropic Thomas process

The (isotropic) modified Thomas process (Thomas 1949) belongs to the class of Neyman–Scott processes introduced at the end of Sect. 2. The daughter points are distributed according to the bivariate zero mean Gaussian density h_{σ^2} for independent components with the same variance σ^2, i.e. $h_{\sigma^2}(u) = \frac{1}{2\pi\sigma^2}\exp(-\frac{|u|^2}{2\sigma^2}),\ u \in \mathbf{R}^2$. Thus, the Palm intensity is given by

$$\lambda_0(u;\theta) = \mu\nu + \frac{\nu}{4\pi\sigma^2}\exp\left(-\frac{|u|^2}{4\sigma^2}\right).$$

We can obtain the anisotropic version of the modified Thomas process by using a general bivariate zero mean Gaussian density h_Σ with covariance matrix Σ for the distribution of the daughter points, i.e. $h_\Sigma(u) = \frac{1}{2\pi\sqrt{\det\Sigma}}\exp(-\frac{u^T\Sigma^{-1}u}{2})$. The Palm intensity has still a closed form

$$\lambda_0(u;\theta) = \mu\nu + \frac{\nu}{4\pi\sqrt{\det\Sigma}}\exp\left(-\frac{u^T\Sigma^{-1}u}{4}\right),$$

and the corresponding Palm log likelihood function can be expressed as

$$\log L_{PU}(\theta) = \sum_{\substack{x\in X\cap(W\ominus R)\\ y\in X\\ (y-x)\in B(o,R)}}^{\neq} \log\left(\mu\nu + \frac{\nu}{4\pi\sqrt{\det\Sigma}}\exp\left(-\frac{(y-x)^T\Sigma^{-1}(y-x)}{4}\right)\right)$$
$$-|X\cap(W\ominus R)|\nu\left(\mu\pi R^2 + \int_{B(o,R)}\frac{1}{4\pi\sqrt{\det\Sigma}}\exp\left(-\frac{u^T\Sigma^{-1}u}{4}\right)\mathrm{d}u\right). \quad (17)$$

If we let $\Sigma = \begin{pmatrix}\sigma_1^2 & \sigma_1\sigma_2\rho\\ \sigma_1\sigma_2\rho & \sigma_2^2\end{pmatrix}$, the anisotropic Thomas process can be parametrized by $\theta = (\nu,\mu,\sigma_1,\sigma_2,\rho)$. If we choose the parameter space $\Theta = \prod_{i=1}^4[a_i,b_i] \times [-a_5,a_5]$ with $0 < a_i < b_i < \infty$, $i = 1,2,3,4$, $0 < a_5 < 1$, then the continuity and boundedness assumptions of Theorems 1 and 3 are satisfied. From a practical point of view, the restriction of the parameter space to the compact set Θ is not a problem since we can always choose a_i and b_i appropriately so that Θ covers all values of the parameters which are reasonable for a particular application. Further, the process is Brillinger mixing and satisfies the assumptions of Lemma 1 for any $\epsilon > 2$—thus also the assumptions (12) and (13) from Theorem 3 are satisfied.

By differentiation of $\log L_{PU}(\theta)$ with respect to ν, we get the estimate

$$\hat{\nu} = \frac{N}{\left(\mu\pi R^2 + \int_{B(o,R)}\frac{1}{4\pi\sqrt{\det\Sigma}}\exp\left(-\frac{u^T\Sigma^{-1}u}{4}\right)\mathrm{d}u\right)|X\cap(W\ominus R)|},$$

where N is the number of pairs of points (x,y) satisfying $x \in X\cap(W\ominus R)$, $y \in X$, $0 < |x-y| < R$, and R is the chosen positive constant used as upper limit of the distance between pairs of points in the Palm likelihood. Thus, we only need to maximize L_{PU} over $(\mu,\sigma_1,\sigma_2,\rho)$.

The integral in (17) is not in general available in a closed form. In such a case, the numerical approximation of the integral together with the Nelder–Mead maximization algorithm (Nelder and Mead 1965) provides a computationally feasible procedure for finding the estimates. If instead the composite likelihood method in Guan (2006) is used, the parameter ν cannot be identified and it has to be estimated separately, e.g. from the intensity λ of the point process by $\hat{\nu} = \frac{\hat{\lambda}}{\mu} = \frac{|X \cap W|}{\mu|W|}$ with μ replaced by an estimate obtained from the composite likelihood.

6.2 Example 2: log Gaussian Cox process

Let us consider the stationary log Gaussian Cox process in $\mathbf{R}^2$ (as described at the end of Sect. 2) driven by a stationary Gaussian field with the mean value $m \in \mathbf{R}$ and the exponential covariance function $c(u, v) = \sigma^2 \exp(-\beta|u - v|)$ with $\sigma^2, \beta > 0$. The model is parametrized by $\theta = (m, \sigma^2, \beta) \in \Theta$ and the corresponding Palm intensity is given by

$$\lambda_0(u; \theta) = \exp\left(m + \sigma^2/2 + \sigma^2 \exp(-\beta|u|)\right).$$

If we choose $\Theta = \prod_{i=1}^{3}[a_i, b_i]$ with $-\infty < a_1 < b_1 < \infty, 0 < a_i < b_i < \infty, i = 2, 3,$ then all the boundedness and continuity assumptions of Theorems 1 and 3 are satisfied as well as conditions (12) and (13) from Theorem 3. By differentiation of the log Palm likelihood, we get the following estimate for m

$$\hat{m} = \log\left(\frac{N}{|X \cap (W \ominus R)| \int_{B(o,R)} \exp(\sigma^2 \mathrm{e}^{-\beta|u|})\mathrm{d}u}\right) - \frac{\sigma^2}{2}$$

where N is the number of pairs of points (x, y) satisfying $x \in X \cap (W \ominus R)$, $y \in X$, $0 < |x - y| < R$. The estimate of the parameters (σ^2, β) has to be found again by numerical maximization of $\log L_{PU}$ using, e.g. the Nelder–Mead method.

7 Appendix

This Appendix contains a proof of Theorem 1, Theorem 2, Lemma 1 and Theorem 3. First, we need to prove two lemmas.

For $\epsilon > 0$, let $R_\epsilon = \{|\theta - \theta_0| < \epsilon\}$ and

$$\omega_n(\delta) = \sup_{|\theta_1 - \theta_2| < \delta} \{|U_n(\theta_1) - E_{\theta_0} U_n(\theta_1) - U_n(\theta_2) + E_{\theta_0} U_n(\theta_2)|\}$$

be the modulus of continuity of $U_n(\theta) - E_{\theta_0} U_n(\theta)$.

For the proof of Theorem 1, we will use the ideas about consistency of estimating equations introduced in Crowder (1986). However, since all the results in that paper apply to weak convergence only, we have to prove modified versions of the relevant lemmas which can be used for proving the P_{θ_0}-a.s. convergence. Lemma 2 is a stronger version of Crowder (1986, Theorem 3.1).

Lemma 2 *Let Θ be compact. Suppose that the following conditions are satisfied for any $\epsilon > 0$*

$$\inf_{\Theta \setminus R_\epsilon} |E_{\theta_0} U_n(\theta)| \geq C_\epsilon \text{ for some } C_\epsilon > 0 \text{ and all } n > N \text{ for some fixed } N > 0, \tag{18}$$

$$\sup_{\Theta} |U_n(\theta) - E_{\theta_0} U_n(\theta)| \to 0 \quad P_{\theta_0} - a.s., \tag{19}$$

then $\hat{\theta}_n \to \theta_0$ P_{θ_0}-a.s.

Proof It suffices to show that for all $\epsilon > 0$

$$P_{\theta_0}\left(\exists n_0 \, \forall n \geq n_0 : \{\theta \in \Theta : U_n(\theta) = 0\} \subseteq R_\epsilon\right) = 1.$$

Let

$$S_{mn} = \{\theta \in \Theta : |E_{\theta_0} U_n(\theta)| \leq m\},$$

where $m > 0$. From (18) we find for $m < C_\epsilon$ and $n > N$ that $S_{mn} \subseteq R_\epsilon$. Let us choose such an $m < C_\epsilon$. It therefore suffices to show that for all $\epsilon > 0$

$$P_{\theta_0}\left(\exists n_0 \, \forall n \geq n_0 : \{\theta \in \Theta : U_n(\theta) = 0\} \subseteq S_{mn}\right) = 1. \tag{20}$$

For θ with $U_n(\theta) = 0$, we have

$$|E_{\theta_0} U_n(\theta)| \leq |E_{\theta_0} U_n(\theta) - U_n(\theta)| + U_n(\theta) \leq \sup_{\Theta} |U_n(\theta) - E_{\theta_0} U_n(\theta)|.$$

Using (19), (20) follows. □

Moreover for checking the assumption (19), we derive a stronger version of Crowder (1986, Lemma 3.2).

Lemma 3 *Let Θ be compact and assume that*

$$|U_n(\theta) - E_{\theta_0} U_n(\theta)| \to 0 \; P_{\theta_0} - a.s. \text{ for any } \theta \in \Theta, \tag{21}$$

then there exists a sequence $\{\epsilon_k\}_{k \in \mathbf{N}}$, $\epsilon_k \underset{[k \to \infty]}{\longrightarrow} 0$, such that

$$P_{\theta_0}\left(\limsup_{n \to \infty} \omega_n\left(\frac{1}{k}\right) \geq \epsilon_k\right) = 0 \text{ for each } k \in \mathbf{N}. \tag{22}$$

Then $\sup_\Theta |U_n(\theta) - E_{\theta_0} U_n(\theta)| \to 0$ P_{θ_0}-a.s.

Proof We want to show that for all $\delta > 0$

$$P_{\theta_0}\Big(\exists n_0\, \forall n \geq n_0 : \sup_{\Theta} |U_n(\theta) - E_{\theta_0} U_n(\theta)| < \delta\Big) = 1.$$

The proof is by a standard covering argument (see e.g. Guyon 1995). For a fixed $\delta > 0$, we choose k such that $\delta > 2\epsilon_k$ and cover the compact space Θ with N balls $B(\theta_i, \frac{1}{k})$, $i = 1, \ldots, N$. Let $D = \{\limsup_{n\to\infty} \omega_n(\frac{1}{k}) < \epsilon_k\}$. Then, by (22), $P_{\theta_0}(D) = 1$. In other words, P_{θ_0}-a.s. there exists an n_{01} such that for $n \geq n_{01}$,

$$\omega_n\left(\frac{1}{k}\right) < \epsilon_k < \frac{\delta}{2}.$$

Furthermore for $\theta \in B(\theta_i, \frac{1}{k})$, we have

$$\begin{aligned} |U_n(\theta) - E_{\theta_0} U_n(\theta)| &\leq |U_n(\theta) - E_{\theta_0} U_n(\theta) - U_n(\theta_i) + E_{\theta_0} U_n(\theta_i)| \\ &\quad + |U_n(\theta_i) - E_{\theta_0} U_n(\theta_i)| \\ &\leq \omega_n\left(\frac{1}{k}\right) + \max_{i=1,\ldots,N} \{|U_n(\theta_i) - E_{\theta_0} U_n(\theta_i)|\}. \end{aligned}$$

It follows that

$$\sup_{\Theta} |U_n(\theta) - E_{\theta_0} U_n(\theta)| \leq \omega_n\left(\frac{1}{k}\right) + \max_{i=1,\ldots,N} \{|U_n(\theta_i) - E_{\theta_0} U_n(\theta_i)|\}.$$

Using (21), P_{θ_0}-a.s. there exists an n_{02} such that for $n \geq n_{02}$

$$\max_{i=1,\ldots,N} \{|U_n(\theta_i) - E_{\theta_0} U_n(\theta_i)|\} < \frac{\delta}{2}.$$

The result of the lemma now follows by choosing $n_0 = \max\{n_{01}, n_{02}\}$. □

Proof (Proof of Theorem 1) It suffices to show that the conditions of Lemma 2 are satisfied. First notice that

$$E_{\theta_0} U_n(\theta) = \lambda\left(\int_{B(o,R)} \frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta} \frac{1}{\lambda_0(u;\theta)} \lambda_0(u;\theta_0)\mathrm{d}u - \frac{\mathrm{d}\int_{B(o,R)} \lambda_0(u;\theta)\mathrm{d}u}{\mathrm{d}\theta}\right)$$

is bounded and continuous (even uniformly continuous on the compact set Θ) with respect to θ from the assumptions and it does not depend on the observation window W_n. Thus from the assumptions that Θ is compact and $|E_{\theta_0} U_n(\theta)| = 0$ only for θ_0, we get that (18) holds, in fact for all n.

To show (19), we will use Lemma 3. From the ergodicity of the point process X and the form of $U_n(\theta)$ (being just a sum over pairs of points closer than R of some continuous, bounded function over the convex averaging sequence of $\{W_n\}$), it follows that $U_n(\theta) \to E_{\theta_0} U_n(\theta)$ almost surely for any fixed θ (see e.g. Daley and Vere-Jones 2003, pp. 335–338). Thus, (21) holds true.

For (22), we observe that for any $\epsilon > 0$ there exists $\delta > 0$ such that

$$\sup_{\substack{u\in B(o,R)\\ |\theta_1-\theta_2|<\delta}} \left| \frac{\mathrm{d}\lambda_0(u;\theta_1)}{\mathrm{d}\theta} \frac{1}{\lambda_0(u;\theta_1)} - \frac{\mathrm{d}\lambda_0(u;\theta_2)}{\mathrm{d}\theta} \frac{1}{\lambda_0(u;\theta_2)} \right| < \frac{\epsilon}{6\lambda^2 K(R;\theta_0)}, \tag{23}$$

where $K(R;\theta_0) = \frac{1}{\lambda} \int_{B(o,R)} \lambda_0(u;\theta_0) du$ is the K-function of X, and

$$\sup_{|\theta_1-\theta_2|<\delta} \left| \frac{\mathrm{d} \int_{B(o,R)} \lambda_0(u;\theta_1)\mathrm{d}u}{\mathrm{d}\theta} - \frac{\mathrm{d} \int_{B(o,R)} \lambda_0(u;\theta_2)\mathrm{d}u}{\mathrm{d}\theta} \right| < \frac{\epsilon}{6\lambda}, \tag{24}$$

due to the continuity assumptions on $\frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta}\frac{1}{\lambda_0(u;\theta)}$ and $\frac{\mathrm{d}(\int_{B(o,R)} \lambda_0(u;\theta)\mathrm{d}u)}{\mathrm{d}\theta}$. It follows that we can construct a sequence $\{\epsilon_k\}$ such that (23) and (24) hold for every $k \in \mathbf{N}$ when we take $\delta = \frac{1}{k}$ and $\epsilon = \epsilon_k$. Moreover since $E_{\theta_0} U_n(\theta)$ is uniformly continuous on Θ, we can modify the sequence $\{\epsilon_k\}$ in such a way that

$$\sup_{|\theta_1-\theta_2|<\frac{1}{k}} \left| E_{\theta_0} U_n(\theta_1) - E_{\theta_0} U_n(\theta_2) \right| < \frac{\epsilon_k}{3}, \tag{25}$$

holds true for all k (and any n).

Let

$$A_n = \frac{1}{|W_n \ominus R|} \sum_{\substack{x\in X\cap(W_n\ominus R)\\ y\in X,(y-x)\in B(o,R)}}^{\neq} 1.$$

Then, for every k and n holds

$$\begin{aligned} &\sup_{|\theta_1-\theta_2|<\frac{1}{k}} \left| U_n(\theta_1) - E_{\theta_0} U_n(\theta_1) - U_n(\theta_2) + E_{\theta_0} U_n(\theta_2) \right| \\ &< \frac{\epsilon_k}{6\lambda^2 K(R;\theta_0)} A_n + \frac{\epsilon_k}{6\lambda} \frac{|X \cap (W_n \ominus R)|}{|W_n \ominus R|} \\ &\quad + \sup_{|\theta_1-\theta_2|<\frac{1}{k}} \left| E_{\theta_0} U_n(\theta_1) - E_{\theta_0} U_n(\theta_2) \right|. \end{aligned}$$

Thus, for any k, we have

$$\begin{aligned} &P_{\theta_0} \limsup_{n\to\infty} \left(\omega_n(\frac{1}{k}) \geq \epsilon_k \right) \\ &\leq P_{\theta_0} \left(\limsup_{n\to\infty} A_n > 2\lambda^2 K(R,\theta_0) \right) \end{aligned}$$

$$+P_{\theta_0}\left(\limsup_{n\to\infty}\frac{|X\cap(W_n\ominus R)|}{|W_n\ominus R|}>2\lambda\right)$$
$$+P_{\theta_0}\left(\limsup_{n\to\infty}\sup_{|\theta_1-\theta_2|<\frac{1}{k}}|E_{\theta_0}U_n(\theta_1)-E_{\theta_0}U_n(\theta_2)|>\frac{\epsilon_k}{3}\right)=0.$$

The first two terms are equal to 0 because $\frac{|X\cap(W\ominus R)|}{|W\ominus R|}$ converges to λ and A_n converges to $\lambda^2 K(R;\theta_0)$ almost surely from the ergodicity of the process X. The third term is equal to 0 from (25). Thus, (22) holds true and from Lemma 3 we obtain (19), which completes the proof. □

Proof (Proof of Theorem 2) The proof is analogous to the proof of Theorem 1. We need to show that the conditions (18) and (19) are fulfilled for $\tilde{U}_n(\theta)$ instead of $U_n(\theta)$.

First, observe that

$$\begin{aligned}
&\left|E_{\theta_0}U_n(\theta)-E_{\theta_0}\tilde{U}_n(\theta)\right| \\
&\quad\le \lambda\int_{B(o,R)}\left|\frac{\gamma_{W_n}(u)}{|W_n|}-1\right|\left|\frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta}\frac{1}{\lambda_0(u;\theta)}\right|\lambda_0(u;\theta_0)\mathrm{d}u \\
&\quad\le \lambda K(R;\theta_0)\max_{u\in B(o,R),\theta\in\Theta}\left|\frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta}\frac{1}{\lambda_0(u;\theta)}\right|\sup_{z\in B(o,R)}\left|\frac{\gamma_{W_n}(z)}{|W_n|}-1\right| \\
&\quad< C_\epsilon/2,
\end{aligned}\tag{26}$$

for all n larger than some $N>0$ from the continuity assumptions and from (10), where C_ϵ is the constant from (18) in proof of Theorem 1. Now combining (18) in the proof of Theorem 1 with (26), we get that (18) holds also for $\tilde{U}_n(\theta)$ for $\tilde{C}_\epsilon = C_\epsilon/2$.

The validity of (19) follows in exactly the same way as the validity of (19) in the proof in Theorem 1. □

Proof (Proof of Lemma 1) Let us consider two sets A and B with $d(A,B)\ge k$ and $|A|,\ |B|\le p$. Let us rewrite the cluster process X as $X=\bigcup_{c\in C}X_c$, where X_c is the cluster centered around a mother point located at c and C is the stationary Poisson process of mothers.

We denote $X_1=\bigcup_{c\in A\oplus\frac{k}{2}}X_c$ and $X_2=X\backslash X_1$. Then, X_1 and X_2 are independent processes and $X=X_1\cup X_2$. Let G_1, G_2 be measurable subsets of the locally finite subsets of $\mathbf{R}^d$ and $C_1=\{X\cap A\in G_1\},\ C_2=\{X\cap B\in G_2\}$ be arbitrary fixed events from $\mathcal{F}^X(A), \mathcal{F}^X(B)$, respectively. Moreover, we let $D_1=\{X_1\cap B=\emptyset\}$, $D_2=\{X_2\cap A=\emptyset\}$ and $D=D_1\cap D_2$.

Then, $P(C_1\cap C_2)=P(C_1\cap C_2\cap D)+P(C_1\cap C_2\cap D^C)$ and $P(C_1\cap C_2\cap D)=P(X_1\cap A\in G_1, X_1\cap B=\emptyset)P(X_2\cap B\in G_2, X_2\cap A=\emptyset)$. Similarly,

$$\begin{aligned}
P(C_1)P(C_2) &= P(X_1\cap A\in G_1, X_1\cap B=\emptyset)P(X_2\cap B\in G_2, X_2\cap A=\emptyset)P(D)\\
&\quad+P(C_1\cap D)P(C_2\cap D^C)+P(C_1\cap D^C)P(C_2\cap D)\\
&\quad+P(C_1\cap D^C)P(C_2\cap D^C).
\end{aligned}$$

Thus,

$$|P(C_1 \cap C_2) - P(C_1)P(C_2)| \leq 4P(D^C) \leq 4P(D_1^C) + 4P(D_2^C),$$

and

$$P(D_1^C) \leq E|X_1 \cap B| = \mu\nu \int_{A \oplus \frac{k}{2}} \int_B k(c,u)\mathrm{d}u\mathrm{d}c \leq \mu\nu|B| \int_{\mathbf{R}^d \setminus B(o,\frac{k}{2})} h(v)\mathrm{d}v,$$

since the distance of $c \in A \oplus \frac{k}{2}$ and $x \in B$ is always larger or equal to $\frac{k}{2}$. Similarly,

$$P(D_2^C) \leq \mu\nu|A| \int_{\mathbf{R}^d \setminus B(o,\frac{k}{2})} h(v)\mathrm{d}v.$$

Thus,

$$\frac{\alpha(p;k)}{\max(p,1)} \leq \mu\nu \left(\frac{p}{p} + \frac{p}{p}\right) \int_{\mathbf{R}^d \setminus B(o,\frac{k}{2})} h(v)\mathrm{d}v = \mathcal{O}\left(\int_{\frac{k}{2}}^{\infty} v^{d-1-d-\epsilon}\mathrm{d}v\right) = \mathcal{O}(k^{-\epsilon}),$$

where we at the second equality changed into polar coordinates. This concludes the proof. □

Proof (Proof of Theorem 3) To show the existence of

$$\lim_{n\to\infty} |W_n \ominus R| \mathrm{Var}_{\theta_0} U_n(\theta),$$

let us write in detail that

$$\begin{aligned}
&|W_n \ominus R| \mathrm{Var}_{\theta_0} U_n(\theta) \\
&\quad = \frac{1}{|W_n \ominus R|} \mathrm{Var}_{\theta_0} \left(\sum_{x,y\in X}^{\neq} I_{W_n \ominus R}(x) I_{B(o,R)}(x-y) \frac{\mathrm{d}\lambda_0(y-x;\theta)}{\mathrm{d}\theta} \frac{1}{\lambda_0(y-x;\theta)} \right) \\
&\qquad - \frac{(2\int_{B(o,R)} \frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta}\mathrm{d}u)^T}{|W_n \ominus R|} \\
&\qquad \times \mathrm{Cov}_{\theta_0} \left(\sum_{x,y\in X}^{\neq} I_{W_n \ominus R}(x) I_{B(o,R)}(x-y) \frac{\mathrm{d}\lambda_0(y-x;\theta)}{\mathrm{d}\theta} \frac{1}{\lambda_0(y-x;\theta)}, \sum_{z\in X} I_{W_n \ominus R}(z) \right) \\
&\qquad + \frac{\mathrm{Var}_{\theta_0}(X \cap |W_n \ominus R|)}{|W_n \ominus R|} \left(\int_{B(o,R)} \frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta}\mathrm{d}u \right)^T \left(\int_{B(o,R)} \frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta}\mathrm{d}u \right).
\end{aligned}$$

The last term converges to

$$\lambda(1 + \gamma_{\mathrm{red}}^{(2)}(\mathbf{R}^d)) \left(\int_{B(o,R)} \frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta}\mathrm{d}u \right)^T \left(\int_{B(o,R)} \frac{\mathrm{d}\lambda_0(u;\theta)}{\mathrm{d}\theta}\mathrm{d}u \right).$$

The variance and covariance in the first and second term can be expressed as integrals with respect to the factorial moment measures. Using the relations between the moment and cumulant measures and disintegration of $\gamma^{(k)}$ for stationary processes, the terms can be further reexpressed as a combination of mixed integrals with respect to the reduced factorial cumulant measures up to the fourth order. Examining them one by one, it can be shown that under the assumptions of boundedness and continuity of $\frac{\mathrm{d}\lambda_0(x;\theta)}{\mathrm{d}\theta}\frac{1}{\lambda_0(x;\theta)}$ and the finiteness of the total variation of $\gamma_{\text{red}}^{(k)}$, $k = 2, 3, 4$, the normalization by $1/|W_n \ominus R|$ is the correct one to make them all converge.

To prove the asymptotic normality of the MPL estimator, denote by $U_n'(\theta)$ the derivative $\frac{\mathrm{d}U_n(\theta)}{\mathrm{d}\theta}$. The mean value theorem yields

$$U_n(\hat{\theta}_n) = U_n(\theta_0) + U_n'(\theta_n^*)(\hat{\theta}_n - \theta_0) = 0,$$

for some $\theta_n^* = \theta_0 + Q(\hat{\theta}_n - \theta_0)$, where Q is a diagonal matrix with diagonal elements between 0 and 1. Thus for proving the asymptotic normality of $\hat{\theta}_n$, it is enough to prove by Slutsky's lemma that:

$$U_n'(\theta_n^*) \to M \text{ in probability,} \tag{27}$$

$$\sqrt{|W_n \ominus R|}\, U_n(\theta_0) \text{ converges in distribution to } N(0, \Sigma(\theta_0)). \tag{28}$$

To show (27), we observe that

$$|U_n'(\theta_n^*) - M| \le |U_n'(\theta_n^*) - U_n'(\theta_0)| + |U_n'(\theta_0) - M|.$$

Now, $|U_n'(\theta_0) - M|$ converges to 0 almost surely from the ergodicity of X since $M = E_{\theta_0} U_n'(\theta_0)$.

To show that $|U_n'(\theta_n^*) - U_n'(\theta_0)|$ converges to 0 in probability, let us denote by $m^f(\delta)$ the supremum from the formula (14) taken over $|\theta_1 - \theta_2| < \delta$ and by $m^h(\delta)$ the supremum from the formula (15) taken again over $|\theta_1 - \theta_2| < \delta$. Since both the $m^f(\delta)$ and $m^h(\delta)$ converge to 0 as $\delta \to 0$, we can find for a given $\epsilon > 0$ a $\Delta > 0$ so that

$$m^f(\delta) < \frac{\epsilon}{4}\frac{1}{\lambda^2 K(R;\theta_0)} \qquad \text{and} \qquad m^h(\delta) < \frac{\epsilon}{4}\frac{1}{\lambda}$$

holds for all $\delta < \Delta$. Then similarly as at the end of the proof of Theorem 1, we have that

$$\sup_{|\theta_1-\theta_0|<\Delta} |U_n'(\theta_1) - U_n'(\theta_0)| < \frac{\epsilon}{4}\frac{1}{\lambda^2 K(R;\theta_0)} A_n + \frac{\epsilon}{4}\frac{1}{\lambda}\frac{|X \cap (W_n \ominus R)|}{|W_n \ominus R|},$$

and, thus from the ergodicity of X for a given $\eta > 0$, we can find an N_1 large enough so that

$$P_{\theta_0}\left(|U_n'(\theta_1) - U_n'(\theta_0)| > \epsilon\right) < \frac{\eta}{2}, \tag{29}$$

for all $n > N_1$ and θ_1 closer than Δ to θ_0.

Now, we have to remember that since $\hat{\theta}_n$ converges to θ_0 almost surely and thus also in probability from Theorem 1, so does θ_n^*. Thus for a given $\eta > 0$, we can find an N_2 such that

$$P_{\theta_0}(|\theta_n^* - \theta_0| > \Delta) < \frac{\eta}{2} \qquad \text{for all}\, n > N_2,$$

which together with (29) gives that for a given $\epsilon > 0$ and $\eta > 0$

$$P_{\theta_0}\left(|U_n'(\theta_n^*) - U_n'(\theta_0)| > \epsilon\right) < \eta \qquad \text{for all } \ n > \max(N_1, N_2).$$

Thus $|U_n'(\theta_n^*) - U_n'(\theta_0)|$ converges in probability to 0, which completes the proof of (27).

To prove (28), we will use a blocking method similar to the one used in Guan et al. (2007).

Let α and η be positive constants such that $2d/(d + \epsilon) < \eta < \alpha < 1$, and let $\rho_n = \rho(W_n)$, $l_n = \rho_n^\alpha$ and $m_n = \rho_n^\alpha - \rho_n^\eta$.

For a fixed n, let us cover $\mathbf{R}^d$ by the union of disjoint d-dimensional cubes $\{K_n^j\}$ of sidelength l_n. Let $C_n^j \subset K_n^j$ be the closed cube with the same center as K_n^j, but with sidelength m_n. In the sequel, we will consider the collection $\{C_n^j, \, j \in J_n\}$ of all cubes contained in W_n. Note that the distance between any two distinct cubes C_n^j, $C_n^{j'}$ is at least ρ_n^η, which goes to infinity as n increases. Thus by the strong mixing, $X \cap C_n^j$ and $X \cap C_n^{j'}$ become asymptotically independent and furthermore the volume (and thus the observed information available in these sets) of W_n and $\bigcup_{j \in J_n} C_n^j$ are of the same order. This is the main idea of the blocking method. For the formal development of the argument, we need some more notation.

Let $U_n(\theta_0)_i$ denote the i-th component of the score function $U_n(\theta_0)$ and let

$$\begin{aligned}
S_n &= \sqrt{|W_n \ominus R|}\, U_n(\theta_0)_i, \\
s_n^j &= \sqrt{|C_n^j \ominus R|}\, U(\theta_0; C_n^j)_i, \\
s_n &= \left(\sum_{j \in J_n} s_n^j\right) \Big/ \sqrt{k_n}, \\
s_n' &= \left(\sum_{j \in J_n} s_n'^j\right) \Big/ \sqrt{k_n}
\end{aligned}$$

where $U(\theta_0; C_n^j)_i$ is the i-th component of the score function $U_n(\theta_0)$ with W_n replaced by C_n^j, k_n is the number of elements in J_n and $\{s_n'^j, \, j \in J_n\}$ is a collection of independent identically distributed random variables with the same distribution as s_n^j.

To prove that $\sqrt{|W_n \ominus R|}\, U_n(\theta_0)_i$ converges in distribution to $N(0, \sigma_i(\theta_0))$, where $\sigma_i(\theta_0) = (\Sigma(\theta_0))_{i,i}$, it is enough to show the following three facts:

$$(S_n - s_n) \to 0 \text{ in probability}, \tag{30}$$
$$(\phi_n(t) - \phi_n'(t)) \to 0 \text{ for all } t \in \mathbf{R}, \tag{31}$$
$$s_n' \to N(0, \sigma_i(\theta_0)) \text{ in distribution}, \tag{32}$$

where ϕ_n denotes the characteristic function of s_n and ϕ_n' of s_n', respectively.

Since $E_{\theta_0} S_n = E_{\theta_0} s_n = 0$, it is enough to show that $\mathrm{Var}_{\theta_0}(S_n - s_n) \to 0$ to prove (30). Obviously

$$|W_n \ominus R| \geq k_n |C_n^j \ominus R|,$$

and $W_n \ominus R \subseteq W_n \subseteq (\partial W_n \oplus \sqrt{d} l_n) \cup (\cup_{j \in J_n} K_n^j)$, thus

$$|W_n \ominus R| \leq |\partial W_n \oplus (\sqrt{d}\, l_n)| + k_n |K_n^j|.$$

Since $|\partial W_n \oplus (\sqrt{d}\, l_n))| = \mathcal{O}((\rho_n)^{(d-1)+\alpha})$ and $k_n \leq |W_n \oplus (\sqrt{d}\, l_n))|/l_n^d = \mathcal{O}(\rho_n^{d(1-\alpha)})$, we have

$$\lim_{n\to\infty} \frac{|\partial W_n \oplus (\sqrt{d}\, l_n))|}{k_n |C_n^j \ominus R|} \leq \lim_{n\to\infty} \frac{\mathcal{O}(\rho_n^{d-1+\alpha})}{\mathcal{O}(\rho_n^{d(1-\alpha)} m_n^d)} = \lim_{n\to\infty} \mathcal{O}(\rho_n^{\alpha-1}) = 0,$$

and

$$\lim_{n\to\infty} \frac{k_n |K_n^j|}{k_n |C_n^j \ominus R|} = \lim_{n\to\infty} \frac{l_n^d}{m_n^d} = 1.$$

Thus,

$$\lim_{n\to\infty} \frac{|W_n \ominus R|}{k_n |C_n^j \ominus R|} = 1. \tag{33}$$

Thus, in order to show that $\mathrm{Var}_{\theta_0}(S_n - s_n) \to 0$, it is enough to show that

$$\mathrm{Var}_{\theta_0}\left(\sum_{\substack{x \in X \cap (W_n \ominus R) \\ y \in X, (y-x) \in B(o,R)}}^{\neq} \frac{\left(\frac{\mathrm{d}\lambda_0(y-x;\theta)}{\mathrm{d}\theta} \frac{1}{\lambda_0(y-x;\theta)} \Big|_{\theta=\theta_0} \right)_i}{\sqrt{|W_n \ominus R|}} \right.$$
$$\left. - \sum_{j=1}^{k_n} \sum_{\substack{x \in X \cap (C_n^j \ominus R) \\ y \in X, (y-x) \in B(o,R)}}^{\neq} \frac{\left(\frac{\mathrm{d}\lambda_0(y-x;\theta)}{\mathrm{d}\theta} \frac{1}{\lambda_0(y-x;\theta)} \Big|_{\theta=\theta_0} \right)_i}{\sqrt{|W_n \ominus R|}} \right) \to 0, \tag{34}$$

and

$$\mathrm{Var}_{\theta_0}\left(\frac{|X\cap(W_n\ominus R)|-\sum_{j=1}^{k_n}|X\cap(C_n^j\ominus R)|}{\sqrt{|W_n\ominus R|}}\left(\frac{\mathrm{d}\int_{B(o,R)}\lambda_0(r;\theta)\mathrm{d}r}{\mathrm{d}\theta}\bigg|_{\theta=\theta_0}\right)_i\right)\to 0. \tag{35}$$

Let us denote by V_n the set $(W_n\ominus R)\setminus(\bigcup_{j=1}^{k_n}(C_n^j\ominus R))$. Then from the boundedness assumptions of Theorem 1, the variance from (35) is bounded from above by

$$\mathrm{const}\frac{\lambda|V_n|}{|W_n\ominus R|}(1+\gamma_{\mathrm{red}}^{(2)}(V_n))\le \mathrm{const}'\frac{\lambda|V_n|}{|W_n\ominus R|},$$

when $\gamma_{\mathrm{red}}^{(2)}$ has finite total variation. Similarly, but with a substantialy larger amount of algebra, it is possible to derive an upper bound for the variance from (34) of the same form const$\frac{\lambda|A_n|}{|W_n\ominus R|}$ where the constant is a combination of the total variations of $\gamma_{\mathrm{red}}^{(2)}$, $\gamma_{\mathrm{red}}^{(3)}$ and $\gamma_{\mathrm{red}}^{(4)}$.

The proof of (30) is complete by observing that

$$\frac{|V_n|}{|W_n\ominus R|}=\frac{|W_n\ominus R|-k_n|C_n^j\ominus R|}{|W_n\ominus R|}\ \to\ 0,\quad \text{as}\, n\to\infty,$$

according to (33).

To show (31), we will use the mixing assumptions. Let us define

$$V_j=\exp\left(\iota t\frac{s_n^j}{\sqrt{k_n}}\right),$$

where ι denotes the imaginary unit. Then,

$$\phi_n(t)=E\left(\prod_{j\in J_n}V_j\right),\qquad \phi_n'(t)=\prod_{j\in J_n}EV_j,$$

and

$$|\phi_n(t)-\phi_n'(t)|\le\sum_{j=1}^{k_n-1}\left|E\left(\prod_{s=1}^{j+1}V_s\right)-E\left(\prod_{s=1}^{j}V_s\right)EV_{j+1}\right|. \tag{36}$$

If we denote $Z_j=\prod_{s=1}^{j}V_s$ and $Y_j=V_j$, then obviously $Z_j\in\mathcal{F}^X(\cup_{s=1}^{j}C_n^s)$, $Y_j\in\mathcal{F}^X(C_n^{j+1})$, $|\cup_{s=1}^{j}C_n^s|=j(m_n)^d$, $|C_n^{j+1}|=(m_n)^d$. Recall that $d(\cup_{s=1}^{j}C_n^s,C_n^{j+1})\ge(\rho_n)^\eta$. Since both random variables Z_j and Y_j are bounded in absolute value by 1, we obtain the following bound on their covariance by means of the strong mixing coefficient (see e.g. Lemma 1.2.1 in Zhengyan and Chuanrong 1996)

$$\mathrm{Cov}(Z_j, Y_j) \leq 4\alpha(j(m_n)^d, (\rho_n)^\eta) \leq \mathcal{O}(j(\rho_n^\alpha - \rho_n^\eta)^d(\rho_n)^{-\eta\epsilon}) = \mathcal{O}(j(\rho_n)^{\alpha d-\eta\epsilon}).$$

Finally from the obvious observation $k_n \leq |W_n|/(\rho_n)^{\alpha d} = \mathcal{O}((\rho_n)^{d-\alpha d})$ and from (36), we find

$$|\phi_n(t) - \phi_n'(t)| \leq k_n \mathcal{O}(k_n(\rho_n)^{\alpha d-\eta\epsilon}) \leq \mathcal{O}((\rho_n)^{2d-\alpha d-\eta\epsilon}),$$

which under the assumptions we made about α and η goes to 0 as $\rho_n \to \infty$ and (31) is proved.

(32) is just an application of the Lyapunov central limit theorem.

Finally, since the convergence of $\sqrt{|W_n \ominus R|}\, c \cdot U_n(\theta_0) \to N(0, c\,\Sigma(\theta_0)c^T)$ in distribution for any $c \in \mathbf{R}^q$ follows the same type of derivations as above, the proof of the asymptotic normality of the vector $\sqrt{|W_n \ominus R|}\, U_n(\theta_0)$ follows directly from the Cramér–Wold device. $\square$

Acknowledgments This work was supported by projects GAČR 201/08/P100 and 201/10/0472 from the Czech Science Foundation and by Centre for Stochastic Geometry and Advanced Bioimaging, funded by the Villum Foundation.

References

Baddeley, A. J., Turner, R. (2000). Practical maximum pseudolikelihood for spatial point processes. *Australian & New Zealand Journal of Statistics, 42*, 283–322.

Bolthausen, E. (1982). On the central limit theorem for stationary mixing random fields. *Annals of Probability, 10*, 1047–1050.

Crowder, M. J. (1986). On consistency and inconsistency of estimating equations. *Econometric Theory, 2*, 305–330.

Daley, D. J., Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes* (2nd ed.). New York: Springer.

Davidson, J. (1994). *Stochastic Limit Theory*. New York: Oxford University Press.

Diggle, P. J. (2003). *Statistical Analysis of Spatial Point Patterns*. New York: Oxford University Press.

Doukhan, P. (1994). *Mixing: Properties and Examples*. New York: Springer.

Guan, Y. (2006). A composite likelihood approach in fitting spatial point process models. *Journal of the American Statistical Association, 101*, 1502–1512.

Guan, Y., Sherman, M. (2007). On least squares fitting for stationary spatial point processes. *Journal of the Royal Statistical Society: Series B, 69*, 31–49.

Guan, Y., Sherman, M., Calvin, J. A. (2007). On asymptotic properties of the mark variogram estimator of a marked point process. *Journal of Statistical Planning and Inference, 137*, 148–161.

Guyon, X. (1995). *Random Fields on a Network*. New York: Springer.

Heagerty, P. J., Lumley, T. (2000). Window subsampling of estimating functions with application to regression models. *Journal of the American Statistical Association, 95*, 197–211.

Heinrich, L. (1988). Asymptotic Gaussianity of some estimators for reduced factorial moment measures and product densities of stationary poisson cluster processes. *Statistics, 19*, 87–106.

Heinrich, L. (1992). Minimum contrast estimates for parameters of spatial ergodic point processes. In *Transactions of the 11th Prague Conference on Random Processes, Information Theory and Statistical Decision Functions*. Prague: Academic Publishing House.

Heinrich, L. (2012). Asymptotic methods in statistics of random point processes. In E. Spodarev (Ed.), *Lecture notes on Stochastic Geometry, Spatial Statistics and Random Fields*. New York: Springer (to appear).

Ibragimov, I. A., Linnik, Y. V. (1971). *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Noordhoff.

Illian, J., Penttinen, A., Stoyan, H., Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. Chichester: Wiley.

Jolivet, E. (1981). Central limit theorem and convergence of empirical processes of stationary point processes. In P. Bartfai, J. Tomko (Eds.), *Point Processes and Queueing Problems*. Amsterdam: North-Holland.

Jonsdottir, K. Y., Rønn-Nielsen, A., Mouridsen, K., Jensen, E. B. V. (2011). Lévy based modelling in brain imaging. In C. S. G. B. Research (Ed.), *Report* (pp. 11–2). Aarhus: Department of Mathematics, Aarhus University (Submitted).

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary Mathematics, 80*, 221–239.

Møller, J., Waagepetersen, R. P. (2003). *Statistical Inference and Simulation for Spatial Point Processes*. Boca Raton: Chapman & Hall/CRC.

Møller, J., Waagepetersen, R. P. (2007). Modern statistics for spatial point processes. *Scandinavian Journal of Statistics, 34*, 643–684.

Møller, J., Syversveen, A. R., Waagepetersen, R. P. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics, 25*, 451–482.

Nelder, J. A., Mead, R. (1965). A simplex method for function minimization. *Computer Journal, 7*, 308.

Ogata, Y., Katsura, K. (1991). Maximum likelihood estimates of fractal dimension for random spatial patterns. *Biometrika, 78*, 463–474.

Politis, D. N., Sherman, M. (2001). Moment estimation for statistics from marked point processes. *Journal of the Royal Statistical Society: Series B, 63*, 261–275.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences, 42*, 43–47.

Rue, H., Martino, S., Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society: Series B, 71*, 319–392.

Stein, M. L. (1999). *Interpolation of Spatial Data*. New York: Springer.

Stoyan, D., Kendall, W. S., Mecke, J. (1995). *Stochastic Geometry and its Applications* (2nd ed.). Chichester: Wiley.

Tanaka, U., Ogata, Y., Stoyan, D. (2008). Parameter Estimation and Model Selection for Neyman–Scott Point Processes. *Biometrical Journal, 50*, 43–57.

Thomas, M. (1949). A generalization of Poisson's binomial limit for use in ecology. *Biometrika, 36*, 18–25.

Waagepetersen, R. P. (2007). An estimating function approach to inference for inhomogeneous Neyman–Scott processes. *Biometrics, 63*, 252–258.

Wills, J. M. (1970). Zum Verhältnis von Volumen zu Oberfläche bei Convexen Körpern. *Archiv der Mathematik, 21*, 557–560.

Zhengyan, L., Chuanrong, L. (1996). *Limit Theory for Mixing Dependent Random Variables*. Dordrecht: Kluwer Academic Publishers.