Annals of the Institute of Statistical Mathematics manuscript No.
(will be inserted by the editor)

Infill asymptotics for logistic regression estimators for parameters of the intensity function of spatial point processes

M.N.M. van Lieshout · C. Lu

Received: date / Revised: date

Abstract This paper discusses infill asymptotics for logistic regression estimators for spatial point processes whose intensity functions are of log-linear form. First, we establish strong consistency and asymptotic normality for the parameters of a Poisson point process model. We also propose consistent estimators for the asymptotic covariance matrix. Next, we extend the results to general point process models for which replicated realizations are available and, under proper conditions, extend the central limit theorem to estimators from other unbiased estimating equations that are based on the Campbell–Mecke theorem. In a simulation study, we demonstrate the efficiency of a regular dummy point process in logistic regression estimation and pseudo-likelihood estimation. Finally, we demonstrate the approach on data on kitchen fires in the Twente region in the Netherlands.

Keywords & Phrases: Campbell–Mecke theorem, Infill asymptotics, Logistic regression estimator, Spatial point process, Unbiased estimating equation.

2010 Mathematics Subject Classification: 60G55, 62M30.

This research was funded by the Dutch Research Council (NWO) for the project 'Data riven Risk Management for Fire Services' (18004) when C. Lu was employed at the University of Twente.

M.N.M. van Lieshout
Centrum Wiskunde & Informatica
P.O. Box 94079, NL-1090 GB, Amsterdam, The Netherlands
Department of Applied Mathematics, University of Twente
P.O. Box 217, NL-7500 AE, Enschede, The Netherlands
E-mail: m.n.m.van.lieshout@cwi.nl

C. Lu
Centrum Wiskunde & Informatica
P.O. Box 94079, NL-1090 GB, Amsterdam, The Netherlands

1 Introduction

Spatial point process models have been of great interest in such diverse disciplines as forestry (Waagepetersen 2008), geo-sciences (Bray and Schoenberg 2013) and epidemiology (Dong et al. 2023). In a previous study on chimney fire occurrences in the Twente region of the Netherlands (Lu et al. 2023), we developed a heterogeneous Poisson point process model. The intensity function we proposed was log-linear in a number of explanatory covariates. Similar model structures have been widely used, for instance, in Møller and Waagepetersen (2007) and Cœurjolly and Møller (2014).

To estimate the parameters of a Poisson point process model, one may consider maximum likelihood estimation (e.g., Kutoyants 1998; Møller and Waagepetersen 2004). In practice, this approach involves an integral that needs to be approximated numerically, e.g. using quadrature points (Berman and Turner 1992). For some non-Poisson point processes, including the random stochastic adsorption model (Van Lieshout 2006), maximum likelihood estimation is also tractable. However, the likelihood – if available at all – involves an intractable normalizing constant that must be approximated by cluster expansions (e.g., Ogata and Tanemura 1981), Markov chain Monte Carlo simulations (e.g., Møller and Waagepetersen 2004) or approximate Bayesian computation methods (Sisson and Fan 2011; Stoica et al. 2017; Vihrs, Møller and Gelfand 2022), which are computationally expensive. Therefore, alternative approaches have been developed, such as minimum contrast estimation (e.g., Guyon 1995) and, particularly, M-estimators based on unbiased estimating equations (e.g., Van der Vaart 1998; Sørensen 1999). Early examples of the latter approach are Takacs and Fiksel's pioneering work in the 1980s (Fiksel 1984; Takacs 1985; Fiksel 1988) and maximum pseudo-likelihood estimation (Besag 1977; Baddeley and Turner 2000) for Gibbs and Markov point process models (e.g., Van Lieshout 2000; Georgii 2011). More recent methods include Poisson likelihood estimation (Schoenberg 2005), logistic regression estimation (Baddeley et al. 2014) and the quasi-likelihood approach (Guan, Jalilian and Waagepetersen 2015). Typically, these estimators are based on either the Campbell–Mecke theorem or the Nguyen–Zessin theorem (see, e.g., Daley and Vere-Jones 2008). In Lu et al. (2023), we preferred to use logistic regression estimation because – in contrast to the quasi-likelihood approach – it is easy to implement using standard software for generalized linear models (Baddeley, Rubak and Turner 2016), and because – other than Poisson likelihood estimation – it involves a so-called dummy point process which can be tuned to focus on salient regions and times efficiently.

From a theoretical perspective, it is important to analyze the asymptotic properties of parameter estimators in order to establish statistical inference for point process models. In the current context, two asymptotic regimes can be formulated: increasing-domain asymptotics and infill asymptotics (Ripley 1988). In the former, the observation window expands; in the latter, the window remains fixed but more and more points are observed in it. Note that this can be achieved in various ways, for example by scaling of the intensity function or by observing independent realisations of a point process distribution in the window. A hybrid framework that combines the two regimes can be considered as well (e.g., Kutoyants 1998).

For Poisson likelihood estimators, a vast body of asymptotic theory exists. Assuming that the point process of interest is Poisson, Rathbun and Cressie (1994) proved increasing-domain asymptotics and Kutoyants (1998) gave a detailed and comprehensive account in the hybrid framework. For general point processes, Schoenberg (2005) provided simple but sufficient conditions for consistency in the increasing-domain regime; Waagepetersen (2007) studied infill asymptotics for a class of clustered point processes. For modified Poisson likelihood estimators designed to handle partially observed covariates, Rathbun, Shiffman and Gwaltney (2007) developed increasing-domain asymptotics for Poisson point processes in space and time; infill asymptotics based on scaling the intensity function of Poisson and Neyman-Scott point processes using similar estimators were treated in Waagepetersen (2008). Recently, for variable selection purposes, Thurman et al. (2015) considered increasing-domain asymptotics for regularized versions of Poisson likelihood estimators, whereas Choiruddin, Cœurjolly and Letué (2023) worked in the hybrid framework. Similar results for mis-specified models were also derived by Choiruddin, Cœurjolly and Waagepetersen (2021) for model selection purposes.

In contrast, the literature for logistic regression estimators is more limited. Rathbun (2013) discussed asymptotic theory for Poisson point processes in the increasing-domain regime, whilst Waagepetersen (2008) let the intensity function increase. Baddeley et al. (2014) developed increasing-domain asymptotics for stationary spatial Gibbs point processes. Afterwards, Choiruddin, Cœurjolly and Letué (2018) proved asymptotic normality for general point process models also in the increasing-domain framework. In this paper, we study consistency and asymptotic normality for logistic regression estimators for general spatial point processes in the infill regime and extend our central limit theorem to the estimators obtained from general unbiased estimating equations that are based on the Campbell–Mecke theorem. Note that the infill framework is quite natural for Poisson point patterns observed in a fixed window; it is also suitable for general point patterns when replications are available (e.g., Pawlas 2011; Chiu et al. 2013; Van Lieshout 2021). In practice, replications can be obtained by observing point patterns within fixed domains over different time windows or by dividing a large-scale point pattern into independent, identically distributed pieces at various locations.

Specifically, our contributions are three-fold. First, we develop replication-based infill asymptotic theory for logistic regression estimators. Other than in most of the literature (e.g., Waagepetersen 2008), we give direct, explicit and self-contained proofs and provide sufficient conditions that are easy to verify in practice. Second, our asymptotic results enable us to improve the efficiency of logistic regression estimation by using inhomogeneous and regular dummy point processes. We propose consistent estimators for the asymptotic covariance matrices and demonstrate their practical use in a simulation study and on a data set of kitchen fires. Third, we extend our central limit theorem to estimators obtained from general unbiased estimating equations, paving the way for further studies on the design of test functions.

The remainder of the paper is organized as follows. Section 2 introduces the background and notation. Section 3 presents infill asymptotic results for logistic regression estimators for Poisson point process models. Section 4 discusses extensions to general point processes and to unbiased estimating equations. In Section 5, we validate the approach by a simulation study before showing it in action on kitchen fire data. Finally, the paper finishes with some conclusions and suggestions for further research.

2 Background and Notation

Let X be a spatial point process on a bounded non-empty open set $W \subset \mathbb{R}^d$ equipped with the Borel σ-algebra (see, e.g., Daley and Vere-Jones 2008). Suppose that the first-order moment measure of X, which is defined as

$$\Lambda(B) := \mathbb{E}\left\{\sum_{\boldsymbol{x}\in X} 1(\boldsymbol{x}\in B)\right\}$$

for any Borel subset $B \subset W$, exists as a σ-finite measure that is absolutely continuous with respect to Lebesgue measure with Radon–Nikodym derivative λ. Here $1(\cdot)$ denotes the indicator function. Then $\lambda : W \to [0,\infty)$ is called the intensity function of X.

Higher order moment measures are defined analogously. The second-order factorial moment measure of X is

$$\Lambda^{(2)}(B_1 \times B_2) := \mathbb{E}\left\{\sum_{\boldsymbol{x},\boldsymbol{y}\in X}^{\neq} 1(\boldsymbol{x}\in B_1, \boldsymbol{y}\in B_2)\right\}$$

for any Borel subsets $B_1, B_2 \subset W$. The superscript indicates that the sum ranges over distinct points $\boldsymbol{x} \neq \boldsymbol{y}$. Suppose that this measure exists as a σ-finite measure that is absolutely continuous with respect to Lebesgue measure with Radon–Nikodym derivative $\lambda^{(2)}$. Then $\lambda^{(2)} : W \times W \to [0,\infty)$ is called the second-order product density function of X. Upon scaling, one obtains the pair correlation function $g(\boldsymbol{x},\boldsymbol{y}) := \lambda^{(2)}(\boldsymbol{x},\boldsymbol{y})/[\lambda(\boldsymbol{x})\lambda(\boldsymbol{y})]$ of X, provided that $\lambda(\boldsymbol{x})\lambda(\boldsymbol{y}) > 0$.

2.1 Parametric intensity function

In this paper, we assume that X admits an intensity function λ that is of log-linear form offset by a measurable function b and parameterized by a vector $\boldsymbol{\theta}$ in some parameter space $\boldsymbol{\Theta} \subset \mathbb{R}^m$:

$$\lambda(\boldsymbol{u};\boldsymbol{\theta}) := b(\boldsymbol{u}) \exp\left[\boldsymbol{\theta}^\top \boldsymbol{z}(\boldsymbol{u})\right], \tag{1}$$

where $\boldsymbol{u} \in W$ denotes a location, $b : W \to [0, \infty)$ is a measurable function that serves as the baseline or reference intensity, $\boldsymbol{z} = [z_1, \ldots, z_m]^\top : W \to \mathbb{R}^m$ is an m-dimensional measurable vector of covariates and $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_m]^\top$ is the parameter vector. The gradient vector of $\lambda(\boldsymbol{u}; \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ then takes the form $\nabla\lambda(\boldsymbol{u}; \boldsymbol{\theta}) = \lambda(\boldsymbol{u}; \boldsymbol{\theta})\boldsymbol{z}(\boldsymbol{u})$. Conditions must be imposed on b and $\boldsymbol{z}$ to ensure that λ is integrable. In the sequel, it will sometimes also be necessary to assume that b, and thus λ, is strictly positive. When this is the case, we will state it explicitly.

2.2 Logistic regression estimation

Unbiased estimating equations for the parameters of a spatial point process model in general, and logistic regression estimation in particular, are based on the Campbell–Mecke theorem (see, e.g., Daley and Vere-Jones 2008). Consider a spatial point process X on W with intensity function λ. For any real-valued measurable function f defined on W such that $f\lambda$ is absolutely integrable, the Campbell–Mecke theorem reads

$$\mathbb{E}\left\{\sum_{\boldsymbol{x} \in X} f(\boldsymbol{x})\right\} = \int_W f(\boldsymbol{u})\lambda(\boldsymbol{u})\mathrm{d}\boldsymbol{u}, \tag{2}$$

where $\boldsymbol{x}$ runs through the points of X. When λ is parameterized by a vector $\boldsymbol{\theta}$ as $\lambda(\boldsymbol{u}; \boldsymbol{\theta})$, (2) provides a basis for estimating $\boldsymbol{\theta}$.

Logistic regression estimation is based on the vector function

$$\boldsymbol{f}(\boldsymbol{u}; \boldsymbol{\theta}) := \nabla \log\left[\frac{\lambda(\boldsymbol{u}; \boldsymbol{\theta})}{\lambda(\boldsymbol{u}; \boldsymbol{\theta}) + \rho(\boldsymbol{u})}\right] = \frac{\rho(\boldsymbol{u})/\lambda(\boldsymbol{u}; \boldsymbol{\theta})}{\lambda(\boldsymbol{u}; \boldsymbol{\theta}) + \rho(\boldsymbol{u})}\nabla\lambda(\boldsymbol{u}; \boldsymbol{\theta}). \tag{3}$$

Here, one assumes that $\lambda(\boldsymbol{u}; \boldsymbol{\theta})$ is a positive-valued differentiable function such that its gradient vector $\nabla\lambda(\boldsymbol{u}; \boldsymbol{\theta})$ is absolutely integrable, and that $\rho(\boldsymbol{u})$ is a positive-valued measurable function also defined on W. The idea is then to estimate both sides of (2) and solve the equations for $\boldsymbol{\theta}$. In order to approximate the right-hand side, one may use a 'dummy' point process D on W which is independent of X and has integrable intensity function ρ. Applying the Campbell–Mecke theorem to D, one finds that

$$\sum_{\boldsymbol{x} \in D} \frac{1}{\lambda(\boldsymbol{x}; \boldsymbol{\theta}) + \rho(\boldsymbol{x})}\nabla\lambda(\boldsymbol{x}; \boldsymbol{\theta})$$

is an unbiased estimator for the right-hand side of (2) with $\boldsymbol{f}$ as in (3). Hence,

$$\boldsymbol{s}(X, D; \boldsymbol{\theta}) := \sum_{\boldsymbol{x} \in X} \frac{\rho(\boldsymbol{x})/\lambda(\boldsymbol{x}; \boldsymbol{\theta})}{\lambda(\boldsymbol{x}; \boldsymbol{\theta}) + \rho(\boldsymbol{x})}\nabla\lambda(\boldsymbol{x}; \boldsymbol{\theta}) - \sum_{\boldsymbol{x} \in D} \frac{1}{\lambda(\boldsymbol{x}; \boldsymbol{\theta}) + \rho(\boldsymbol{x})}\nabla\lambda(\boldsymbol{x}; \boldsymbol{\theta}) = \boldsymbol{0} \tag{4}$$

is an unbiased estimating equation. It is interesting to observe that the middle part of (4) is exactly the gradient of a logistic log-likelihood function

$$l(X,D;\boldsymbol{\theta}) \coloneqq \sum_{\boldsymbol{x}\in X} \log\left[\frac{\lambda(\boldsymbol{x};\boldsymbol{\theta})}{\lambda(\boldsymbol{x};\boldsymbol{\theta})+\rho(\boldsymbol{x})}\right] + \sum_{\boldsymbol{x}\in D} \log\left[\frac{\rho(\boldsymbol{x})}{\lambda(\boldsymbol{x};\boldsymbol{\theta})+\rho(\boldsymbol{x})}\right]. \quad (5)$$

Thus, (4) can also be interpreted as the score function and solved using standard software for logistic regression for exponential family models such as the R-package *stats* (Venables 2002). The existence and uniqueness of a maximizer of the log-likelihood function (5) are ensured under proper conditions (Silvapulle 1981).

For simplicity, in the remainder of this paper, we write $\boldsymbol{s}(\boldsymbol{\theta})$ and $l(\boldsymbol{\theta})$ for $\boldsymbol{s}(X,D;\boldsymbol{\theta})$ and $l(X,D;\boldsymbol{\theta})$, and suppress the dependence on the point patterns X and D. Moreover, we use $\boldsymbol{\theta_0}$ and $\widehat{\boldsymbol{\theta}}$ to denote the true value and the estimator of $\boldsymbol{\theta}$, respectively.

2.3 First two moments of first-order U-statistics

Following Reitzner and Schulte (2013), we call random vectors of the form

$$\boldsymbol{H} \coloneqq \sum_{\boldsymbol{x}\in X} \left[h_1(\boldsymbol{x}),\ldots,h_m(\boldsymbol{x})\right]^\top$$

first-order U-statistics of X if every $h_l\lambda$ is absolutely integrable.

Suppose that X has second-order product density function $\lambda^{(2)}$. By the Campbell–Mecke theorem and in analogy to (2),

$$\mathbb{E}\left\{\sum_{\boldsymbol{x}_1,\boldsymbol{x}_2\in X}^{\neq} f(\boldsymbol{x}_1,\boldsymbol{x}_2)\right\} = \int\int_{W\times W} f(\boldsymbol{u},\boldsymbol{v})\lambda^{(2)}(\boldsymbol{u},\boldsymbol{v})\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v}$$

for any real-valued measurable function f defined on $W\times W$ such that $f\lambda^{(2)}$ is absolutely integrable. Then

$$\mathbb{E}\{\boldsymbol{H}\} = \left[\int_W h_l(\boldsymbol{u})\lambda(\boldsymbol{u})\mathrm{d}\boldsymbol{u}\right]_{l=1}^m$$

and

$$\begin{aligned}\mathbb{E}\left\{\boldsymbol{H}\boldsymbol{H}^\top\right\} &= \left[\int_W h_k(\boldsymbol{u})h_l(\boldsymbol{u})\lambda(\boldsymbol{u})\mathrm{d}\boldsymbol{u}\right]_{k,l=1}^m \\ &+ \left[\int\int_{W\times W} h_k(\boldsymbol{u})h_l(\boldsymbol{v})\lambda^{(2)}(\boldsymbol{u},\boldsymbol{v})\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v}\right]_{k,l=1}^m.\end{aligned}$$

If $\mathbb{E}\{\boldsymbol{H}\boldsymbol{H}^\top\}$ is finite and $\lambda > 0$, the entries in the covariance matrix of $\boldsymbol{H}$ are finite and can be expressed in terms of the pair correlation function $g(\boldsymbol{u}, \boldsymbol{v})$ as

$$\mathbf{Cov}\{\boldsymbol{H}\} = \left[\int_W h_k(\boldsymbol{u})h_l(\boldsymbol{u})\lambda(\boldsymbol{u})\mathrm{d}\boldsymbol{u}\right]_{k,l=1}^m + \left[\int\int_{W\times W} h_k(\boldsymbol{u})h_l(\boldsymbol{v})(g(\boldsymbol{u},\boldsymbol{v})-1)\lambda(\boldsymbol{u})\lambda(\boldsymbol{v})\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v}\right]_{k,l=1}^m.$$

2.4 Infill asymptotic regime

The infill asymptotic regime considered in this paper is as defined in Pawlas (2011), Chiu et al. (2013) and Van Lieshout (2021).

Let $\{Y_i\}$ and $\{E_i\}$, with $i \in \mathbb{N}$, be two independent sequences of independent and identically distributed spatio-temporal point processes with intensity functions λ and ρ, respectively. Set

$$X_n := \bigcup_{i=1}^n Y_i, \quad D_n := \bigcup_{i=1}^n E_i.$$

Write λ_n for the intensity function of X_n and ρ_n for that of the 'dummy' point process D_n. Thus $\lambda_n = n\lambda$ and $\rho_n = n\rho$, that is, we assume that the intensity functions of X_n and D_n increase at the same rate. If λ is of the form (1), for all $n \in \mathbb{N}$, the function (4) based on X_n and D_n becomes

$$\boldsymbol{s}_n(\boldsymbol{\theta}) = \sum_{\boldsymbol{x}\in X_n} \frac{\rho(\boldsymbol{x})}{\lambda(\boldsymbol{x};\boldsymbol{\theta}) + \rho(\boldsymbol{x})}\boldsymbol{z}(\boldsymbol{x}) - \sum_{\boldsymbol{x}\in D_n} \frac{\lambda(\boldsymbol{x};\boldsymbol{\theta})}{\lambda(\boldsymbol{x};\boldsymbol{\theta}) + \rho(\boldsymbol{x})}\boldsymbol{z}(\boldsymbol{x}). \tag{6}$$

Note that the terms in the two sums above do not depend on n, while the subscript $\boldsymbol{x}$ runs through the points of X_n and D_n. Taking the limit as $n \to \infty$, one obtains an asymptotic regime that Ripley (1988) calls 'infill asymptotics'.

Intuitively, under this regime, the estimate for the parameter vector $\boldsymbol{\theta}$ will become more precise when the points observed in the fixed window W become more dense. Our aim in the remainder of this paper is to analyze the asymptotic behaviour of the estimator $\widehat{\boldsymbol{\theta}}_n$ defined by (6) as $n \to \infty$.

3 Infill Asymptotics for Poisson Point Process Models

In this section, we derive infill asymptotics for logistic regression estimators in the case that the point process of interest is Poisson. It is motivated by a previous study (Lu et al. 2023) where the Poisson point process of interest existed in a fixed window only.

For the sake of completeness, we recall the definition of Poisson point processes (see, e.g., Daley and Vere-Jones 2008).

Definition 1 A point process X defined as in the beginning of Section 2 is a Poisson point process if it satisfies the following properties:

(i) for any bounded Borel set $B \subset W$, the number of points that fall in B is Poisson distributed with mean $\Lambda(B)$;
(ii) for disjoint bounded Borel sets $B_1, B_2 \subset W$, the numbers of points that fall in B_1 and B_2 are independent.

The probability distribution of a Poisson point process is completely specified by its first-order moment measure. The assumption of a log-linear intensity function parameterized as (1) is natural and constitutes an exponential family with the components of $\sum_{\boldsymbol{x}\in X} \boldsymbol{z}(\boldsymbol{x})$ as sufficient statistics. Moreover, the pair correlation function is equal to one. Since Poisson point processes are infinitely divisible, the infill asymptotic regime defined in Section 2.4 is particularly appropriate.

Let $P_{\boldsymbol{\theta_0}}$ denote the distribution of (X_n, D_n) under the true parameter value $\boldsymbol{\theta_0}$. For ease of referencing, we list the following conditions that are required to derive the asymptotic results.

(C1) $\{Y_i\}$ and $\{E_i\}$ with $i \in \mathbb{N}$ are two independent sequences of independent and identically distributed spatial point processes on some bounded open set $W \subset \mathbb{R}^d$ and defined on the same underlying probability space Ω. Set $X_n := \cup_{i=1}^n Y_i$ and $D_n := \cup_{i=1}^n E_i$.
(C2) Y_i is a Poisson point process with intensity function $\lambda(\boldsymbol{u};\boldsymbol{\theta})$ given by (1), where $b > 0$ is an integrable function, $\boldsymbol{z}$ is a measurable vector of covariates and the parameter vector $\boldsymbol{\theta}$ lies in an open set $\boldsymbol{\Theta} \subset \mathbb{R}^m$; E_i has integrable intensity function $\rho(\boldsymbol{u}) > 0$.
(C3) E_i has bounded pair correlation function $g(\boldsymbol{u}, \boldsymbol{v})$: $\sup_{(\boldsymbol{u},\boldsymbol{v})\in W\times W} g(\boldsymbol{u}, \boldsymbol{v}) < \infty$.
(C4) For every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, there exist $\epsilon_1(\boldsymbol{\theta}), \epsilon_2(\boldsymbol{\theta}) > 0$ such that $\epsilon_1(\boldsymbol{\theta}) < \inf_{\boldsymbol{u}\in W} \rho(\boldsymbol{u})/\lambda(\boldsymbol{u};\boldsymbol{\theta})$ and $\sup_{\boldsymbol{u}\in W} \rho(\boldsymbol{u})/\lambda(\boldsymbol{u};\boldsymbol{\theta}) < \epsilon_2(\boldsymbol{\theta})$.
(C5) The elements of the measurable covariate vector $\boldsymbol{z}$ are all bounded: $\sup_{\boldsymbol{u}\in W} \|\boldsymbol{z}(\boldsymbol{u})\| < \infty$.
(C6) The parameter space $\boldsymbol{\Theta}$ is convex.
(C7) The parametric model for λ is identifiable: $\lambda(\boldsymbol{u};\boldsymbol{\theta}) = \lambda(\boldsymbol{u};\widetilde{\boldsymbol{\theta}})$ almost everywhere on W implies $\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}$.
(C8) The $[m\times m]$-dimensional matrix $\boldsymbol{U}(\boldsymbol{\theta_0})$, whose (k,l)-th entry reads $\int_W \lambda(\boldsymbol{u};\boldsymbol{\theta_0})\rho(\boldsymbol{u}) z_k(\boldsymbol{u}) z_l(\boldsymbol{u})/[\lambda(\boldsymbol{u};\boldsymbol{\theta_0}) + \rho(\boldsymbol{u})]\mathrm{d}\boldsymbol{u}$ is invertible.

A few remarks on some of the conditions are appropriate. In condition (C2), Y_i is assumed to be a Poisson point process, however, this assumption will be relaxed in Section 4.1. In condition (C3), E_i is not required to be a Poisson point process, as its realizations are only used to approximate integrals. Condition (C4) is reasonable, recalling the rule of thumb recommended by Baddeley et al. (2014) for selecting ρ such that $\rho \approx 4\lambda$. Together with condition (C7), it is necessary for strong consistency (cf., Theorem 1). Condition (C5) is needed to ensure that the moments of (6) are finite. Condition (C6) is a technical condition used in proofs, and, finally, condition (C8) helps ensure the existence of the estimator $\widehat{\boldsymbol{\theta}}_n$ as $n \to \infty$ (cf., Theorem 3) as well as that of its asymptotic covariance matrix (cf., Theorem 2).

3.1 Strong consistency

To establish strong consistency, we start our investigations with the asymptotic behaviour of the scaled logistic log-likelihood function.

Lemma 1 *Assume that the conditions (C1)–(C2) and (C4)–(C5) hold. Define* $l_n(\boldsymbol{\theta}) \coloneqq l(X_n, D_n; \boldsymbol{\theta})$ *by (5) with* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$*. Then, as* $n \to \infty$*,* $l_n(\boldsymbol{\theta})/n$ *converges* $P_{\boldsymbol{\theta_0}}$*-almost surely to*

$$\int_W \left\{ \lambda(\boldsymbol{u}; \boldsymbol{\theta_0}) \log \left[\frac{\lambda(\boldsymbol{u}; \boldsymbol{\theta})}{\lambda(\boldsymbol{u}; \boldsymbol{\theta}) + \rho(\boldsymbol{u})} \right] + \rho(\boldsymbol{u}) \log \left[\frac{\rho(\boldsymbol{u})}{\lambda(\boldsymbol{u}; \boldsymbol{\theta}) + \rho(\boldsymbol{u})} \right] \right\} \mathrm{d}\boldsymbol{u}.$$

Proof Under conditions (C1)–(C2) and recalling the logistic log-likelihood function (5),

$$\frac{l_n(\boldsymbol{\theta})}{n} = \frac{1}{n} \sum_{\boldsymbol{x} \in X_n} \log \left[\frac{\lambda(\boldsymbol{x}; \boldsymbol{\theta})}{\lambda(\boldsymbol{x}; \boldsymbol{\theta}) + \rho(\boldsymbol{x})} \right] + \frac{1}{n} \sum_{\boldsymbol{x} \in D_n} \log \left[\frac{\rho(\boldsymbol{x})}{\lambda(\boldsymbol{x}; \boldsymbol{\theta}) + \rho(\boldsymbol{x})} \right],$$

which consists of two first-order U-statistics defined on X_n and D_n. We derive their strong convergence separately.

Write the first-order U-statistic defined on X_n as

$$\frac{1}{n} \sum_{\boldsymbol{x} \in X_n} \log \left[\frac{\lambda(\boldsymbol{x}; \boldsymbol{\theta})}{\lambda(\boldsymbol{x}; \boldsymbol{\theta}) + \rho(\boldsymbol{x})} \right] = \frac{1}{n} \sum_{i=1}^{n} \left\{ \sum_{\boldsymbol{x} \in Y_i} \log \left[\frac{\lambda(\boldsymbol{x}; \boldsymbol{\theta})}{\lambda(\boldsymbol{x}; \boldsymbol{\theta}) + \rho(\boldsymbol{x})} \right] \right\}.$$

The sum is the average of independent and identically distributed real-valued random variables. By the Campbell–Mecke theorem,

$$\mathbb{E}_{\boldsymbol{\theta_0}} \left\{ \sum_{\boldsymbol{x} \in Y_i} \log \left[\frac{\lambda(\boldsymbol{x}; \boldsymbol{\theta})}{\lambda(\boldsymbol{x}; \boldsymbol{\theta}) + \rho(\boldsymbol{x})} \right] \right\} = \int_W \log \left[\frac{\lambda(\boldsymbol{u}; \boldsymbol{\theta})}{\lambda(\boldsymbol{u}; \boldsymbol{\theta}) + \rho(\boldsymbol{u})} \right] \lambda(\boldsymbol{u}; \boldsymbol{\theta_0}) \mathrm{d}\boldsymbol{u}. \tag{7}$$

Conditions (C2) and (C5) imply that the intensity function $\lambda(u; \boldsymbol{\theta_0})$ is integrable on W. By conditions (C2) and (C4),

$$\log \left[\frac{1}{1 + \epsilon_2(\boldsymbol{\theta})} \right] < \log \left[\frac{\lambda(\boldsymbol{u}; \boldsymbol{\theta})}{\lambda(\boldsymbol{u}; \boldsymbol{\theta}) + \rho(\boldsymbol{u})} \right] < \log \left[\frac{1}{1 + \epsilon_1(\boldsymbol{\theta})} \right].$$

Thus, the $P_{\boldsymbol{\theta_0}}$-mean in (7) is finite for all $\boldsymbol{\theta}$. Kolmogorov's strong law of large numbers implies that the first-order U-statistic defined on X_n converges $P_{\boldsymbol{\theta_0}}$-almost surely to the integral in the right-hand side of (7).

Similarly, for the first-order U-statistic defined on D_n, as $n \to \infty$,

$$\begin{aligned} \frac{1}{n} \sum_{\boldsymbol{x} \in D_n} \log \left[\frac{\rho(\boldsymbol{x})}{\lambda(\boldsymbol{x}; \boldsymbol{\theta}) + \rho(\boldsymbol{x})} \right] &\overset{P_{\boldsymbol{\theta_0}}-a.s.}{\to} \mathbb{E}_{\boldsymbol{\theta_0}} \left\{ \sum_{\boldsymbol{x} \in E_i} \log \left[\frac{\rho(\boldsymbol{x})}{\lambda(\boldsymbol{x}; \boldsymbol{\theta}) + \rho(\boldsymbol{x})} \right] \right\} \\ &= \int_W \log \left[\frac{\rho(\boldsymbol{u})}{\lambda(\boldsymbol{u}; \boldsymbol{\theta}) + \rho(\boldsymbol{u})} \right] \rho(\boldsymbol{u}) \mathrm{d}\boldsymbol{u} \end{aligned}$$

because $\rho(\boldsymbol{u})$ is integrable by condition (C2) and

$$\log\left[\frac{\epsilon_1(\boldsymbol{\theta})}{1+\epsilon_1(\boldsymbol{\theta})}\right] < \log\left[\frac{\rho(\boldsymbol{u})}{\lambda(\boldsymbol{u};\boldsymbol{\theta})+\rho(\boldsymbol{u})}\right] < \log\left[\frac{\epsilon_2(\boldsymbol{\theta})}{1+\epsilon_2(\boldsymbol{\theta})}\right]$$

by conditions (C2) and (C4).

The proof is completed by combining the two strong convergence results.

The next theorem is concerned with strong consistency of minimizers of $U_n(\boldsymbol{\theta}) := -l_n(\boldsymbol{\theta})/n$.

Theorem 1 *Assume that the conditions (C1)–(C2) and (C4)–(C7) hold. Define $l_n(\boldsymbol{\theta}) := l(X_n, D_n; \boldsymbol{\theta})$ by (5) with $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. If $\widehat{\boldsymbol{\theta}}_n := \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} l_n(\boldsymbol{\theta})$ exists for all $n \in \mathbb{N}$, then $\widehat{\boldsymbol{\theta}}_n$ converges $P_{\boldsymbol{\theta_0}}$-almost surely to $\boldsymbol{\theta_0}$ as $n \to \infty$.*

Proof Conditions (C2) and (C6) ensure that the parameter space $\boldsymbol{\Theta}$ is open and convex.

First, we prove that, for every ω in the underlying probability space Ω, the realizations of the function $\boldsymbol{\theta} \mapsto U_n(\boldsymbol{\theta})$ are convex. By condition (C2), the intensity function $\lambda(\boldsymbol{x};\boldsymbol{\theta})$ has the log-linear form (1) and is thus twice differentiable with respect to the parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Then the Hessian matrix of $-l_n(\boldsymbol{\theta})$ reads

$$\left[\sum_{\boldsymbol{x}\in X_n\cup D_n} \frac{\lambda(\boldsymbol{x};\boldsymbol{\theta})\rho(\boldsymbol{x})}{(\lambda(\boldsymbol{x};\boldsymbol{\theta})+\rho(\boldsymbol{x}))^2} z_k(\boldsymbol{x}) z_l(\boldsymbol{x})\right]_{k,l=1}^{m}.$$

By decomposition, it can be written into the product of a matrix $\boldsymbol{M}$ and its transpose as $\boldsymbol{M}^\top \boldsymbol{M}$, where $\boldsymbol{M}$ is a $[|X_n \cup D_n| \times m]$-dimensional matrix given by

$$\boldsymbol{M} := \left[\frac{(\lambda(\boldsymbol{x}_i;\boldsymbol{\theta})\rho(\boldsymbol{x}_i))^{1/2}}{\lambda(\boldsymbol{x}_i;\boldsymbol{\theta})+\rho(\boldsymbol{x}_i)} z_k(\boldsymbol{x}_i)\right]_{i=1,k=1}^{|X_n\cup D_n|,m}.$$

Here, $|X_n \cup D_n|$ denotes the number of points in $X_n \cup D_n$, the subscript i runs through all points in $X_n \cup D_n$ and k runs through m covariates. One can readily obtain that the Hessian matrix of $-l_n(\boldsymbol{\theta})$ is positive semi-definite for every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, which implies that $-l_n(\boldsymbol{\theta})$ as a function of $\boldsymbol{\theta}$, and thus $-l_n(\boldsymbol{\theta})/n$, is convex.

Second, we prove that, as $n \to \infty$, $U_n(\boldsymbol{\theta}) - U_n(\boldsymbol{\theta_0})$ converges $P_{\boldsymbol{\theta_0}}$-almost surely to a function $K(\boldsymbol{\theta}, \boldsymbol{\theta_0})$ which is non-negative and vanishes only at $\boldsymbol{\theta} = \boldsymbol{\theta_0}$. By Lemma 1, under conditions (C1)–(C2) and (C4)–(C5), $U_n(\boldsymbol{\theta}) - U_n(\boldsymbol{\theta_0})$ converges $P_{\boldsymbol{\theta_0}}$-almost surely to

$$\int_W \left\{\lambda(\boldsymbol{u};\boldsymbol{\theta_0}) \log\left[\frac{\lambda(\boldsymbol{u};\boldsymbol{\theta_0})}{\lambda(\boldsymbol{u};\boldsymbol{\theta})}\right] - (\lambda(\boldsymbol{u};\boldsymbol{\theta_0}) + \rho(\boldsymbol{u})) \log\left[\frac{\lambda(\boldsymbol{u};\boldsymbol{\theta_0}) + \rho(\boldsymbol{u})}{\lambda(\boldsymbol{u};\boldsymbol{\theta}) + \rho(\boldsymbol{u})}\right]\right\} \mathrm{d}\boldsymbol{u}.$$

Denote this limit by $K(\boldsymbol{\theta}, \boldsymbol{\theta_0})$ and the integrand by $k(\boldsymbol{u}; \boldsymbol{\theta}, \boldsymbol{\theta_0})$. Clearly, $K(\boldsymbol{\theta_0}, \boldsymbol{\theta_0}) = 0$. Furthermore,

$$k(\boldsymbol{u}; \boldsymbol{\theta}, \boldsymbol{\theta_0}) = \lambda(\boldsymbol{u}; \boldsymbol{\theta}) \left\{ \frac{\lambda(\boldsymbol{u}; \boldsymbol{\theta_0})}{\lambda(\boldsymbol{u}; \boldsymbol{\theta})} \log \left[\frac{\lambda(\boldsymbol{u}; \boldsymbol{\theta_0})}{\lambda(\boldsymbol{u}; \boldsymbol{\theta})} \right] - \left(\frac{\lambda(\boldsymbol{u}; \boldsymbol{\theta_0})}{\lambda(\boldsymbol{u}; \boldsymbol{\theta})} + \frac{\rho(\boldsymbol{u})}{\lambda(\boldsymbol{u}; \boldsymbol{\theta})} \right) \log \left[\frac{\frac{\lambda(\boldsymbol{u};\boldsymbol{\theta_0})}{\lambda(\boldsymbol{u};\boldsymbol{\theta})} + \frac{\rho(\boldsymbol{u})}{\lambda(\boldsymbol{u};\boldsymbol{\theta})}}{1 + \frac{\rho(\boldsymbol{u})}{\lambda(\boldsymbol{u};\boldsymbol{\theta})}} \right] \right\}.$$

By condition (C2), $\lambda(\boldsymbol{u}; \boldsymbol{\theta}) > 0$. Consider the function $a \mapsto a \log a - (a + b) \log[(a+b)/(1+b)]$ with $a, b > 0$. Its derivative with respect to a is $\log[a(1+b)/(a+b)]$. The function is strictly decreasing when $a \in (0, 1)$ and strictly increasing when $a \in (1, +\infty)$. For $a = 1$, $a \log a - (a+b) \log[(a+b)/(1+b)] = 0$. Thus, $k(\boldsymbol{u}; \boldsymbol{\theta}, \boldsymbol{\theta_0})$ is non-negative and is strictly positive when $\lambda(\boldsymbol{u}; \boldsymbol{\theta}) \neq \lambda(\boldsymbol{u}; \boldsymbol{\theta_0})$.

Under condition (C7), strong consistency then follows from an appeal to the Proposition below Guyon (1995, Theorem 3.4.4).

Remark 1 Note that the proofs of Lemma 1 and Theorem 1 do not depend on the assumption that Y_i, and thus X_n, is a Poisson point process. However, strong consistency relies on the concavity of $l_n(\boldsymbol{\theta})$.

3.2 Asymptotic normality

To establish asymptotic normality, we start our investigations with the Taylor series of the function (6).

For every component of (6), denoted by $\boldsymbol{s}_{n,i}(\boldsymbol{\theta})$ with $1 \leq i \leq m$, the second-order Taylor expansion of $\boldsymbol{s}_{n,i}(\widehat{\boldsymbol{\theta}}_n)$ with respect to $\boldsymbol{\theta}$ at $\boldsymbol{\theta_0}$ reads

$$\begin{aligned} 0 &= \boldsymbol{s}_{n,i}(\widehat{\boldsymbol{\theta}}_n) \\ &= \boldsymbol{s}_{n,i}(\boldsymbol{\theta_0}) + \nabla \boldsymbol{s}_{n,i}(\boldsymbol{\theta_0})(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}) + \frac{1}{2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0})^\top \nabla^2 \boldsymbol{s}_{n,i}(\boldsymbol{\theta'^{,i}})(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}), \end{aligned} \tag{8}$$

where $\nabla \boldsymbol{s}_{n,i}(\boldsymbol{\theta})$ is the $[1 \times m]$-dimensional vector containing the first-order partial derivatives of $\boldsymbol{s}_{n,i}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ and $\nabla^2 \boldsymbol{s}_{n,i}(\boldsymbol{\theta})$ is the $[m \times m]$-dimensional matrix containing the second-order partial derivatives of $\boldsymbol{s}_{n,i}(\boldsymbol{\theta})$. Moreover, $\boldsymbol{\theta'^{,i}}$ is a convex combination of $\widehat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta_0}$ which, by condition (C6), lies in $\boldsymbol{\Theta}$ as well.

Write $\nabla \boldsymbol{s}_n(\boldsymbol{\theta_0})$ for the matrix whose i-th row is $\nabla \boldsymbol{s}_{n,i}(\boldsymbol{\theta_0})$ and assume its inverse is well-defined. Heuristically, the idea is to ignore the error term which is the quadratic form in (8) and rearrange the remaining terms to obtain

$$n^{1/2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}) \approx \left[-\frac{\nabla \boldsymbol{s}_n(\boldsymbol{\theta_0})}{n} \right]^{-1} \frac{\boldsymbol{s}_n(\boldsymbol{\theta_0})}{n^{1/2}}. \tag{9}$$

In the next two lemmas, we study the asymptotic behaviours of the terms $-\nabla \boldsymbol{s}_n(\boldsymbol{\theta_0})/n$ and $\boldsymbol{s}_n(\boldsymbol{\theta_0})/n^{1/2}$, respectively.

Lemma 2 *Assume that the conditions (C1)–(C2) and (C5) hold. Define $\boldsymbol{s}_n(\boldsymbol{\theta}) := \boldsymbol{s}(X_n, D_n; \boldsymbol{\theta})$ by (6) with $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Then, as $n \to \infty$, $-\nabla \boldsymbol{s}_n(\boldsymbol{\theta_0})/n$ converges $P_{\boldsymbol{\theta_0}}$-almost surely to*

$$\boldsymbol{U} := \left[\int_W \frac{\lambda(\boldsymbol{u};\boldsymbol{\theta_0})\rho(\boldsymbol{u})}{\lambda(\boldsymbol{u};\boldsymbol{\theta_0})+\rho(\boldsymbol{u})} z_k(\boldsymbol{u}) z_l(\boldsymbol{u}) \mathrm{d}\boldsymbol{u} \right]_{k,l=1}^{m}.$$

Proof Under conditions (C1)–(C2) and recalling (6),

$$-\frac{\nabla \boldsymbol{s}_n(\boldsymbol{\theta_0})}{n} = \left[\sum_{\boldsymbol{x} \in X_n \cup D_n} \frac{\lambda(\boldsymbol{x};\boldsymbol{\theta_0})\rho(\boldsymbol{x})}{n\left(\lambda(\boldsymbol{x};\boldsymbol{\theta_0})+\rho(\boldsymbol{x})\right)^2} z_k(\boldsymbol{x}) z_l(\boldsymbol{x}) \right]_{k,l=1}^{m}.$$

To prove component-wise strong convergence, consider the (k,l)-th entry of the matrix above which, recalling condition (C1), is given by

$$\frac{1}{n} \sum_{i=1}^{n} \left\{ \sum_{\boldsymbol{x} \in Y_i \cup E_i} \frac{\lambda(\boldsymbol{x};\boldsymbol{\theta_0})\rho(\boldsymbol{x})}{(\lambda(\boldsymbol{x};\boldsymbol{\theta_0})+\rho(\boldsymbol{x}))^2} z_k(\boldsymbol{x}) z_l(\boldsymbol{x}) \right\}.$$

The sum is the average of independent and identically distributed real-valued random variables. By the Campbell–Mecke theorem,

$$\mathbb{E}_{\boldsymbol{\theta_0}} \left\{ \sum_{\boldsymbol{x} \in Y_i \cup E_i} \frac{\lambda(\boldsymbol{x};\boldsymbol{\theta_0})\rho(\boldsymbol{x}) z_k(\boldsymbol{x}) z_l(\boldsymbol{x})}{(\lambda(\boldsymbol{x};\boldsymbol{\theta_0})+\rho(\boldsymbol{x}))^2} \right\} = \int_W \frac{\lambda(\boldsymbol{u};\boldsymbol{\theta_0})\rho(\boldsymbol{u}) z_k(\boldsymbol{u}) z_l(\boldsymbol{u})}{\lambda(\boldsymbol{u};\boldsymbol{\theta_0})+\rho(\boldsymbol{u})} \mathrm{d}\boldsymbol{u}.$$

By condition (C5), the covariate terms in the integrand are bounded. By condition (C2), $0 < \lambda(\boldsymbol{u};\boldsymbol{\theta_0})\rho(\boldsymbol{u})/(\lambda(\boldsymbol{u};\boldsymbol{\theta_0})+\rho(\boldsymbol{u})) < \rho(\boldsymbol{u})$ and $\rho(\boldsymbol{u})$ is integrable. Kolmogorov's strong law of large numbers implies the claimed $P_{\boldsymbol{\theta_0}}$-almost sure convergence.

Lemma 3 *Assume that the conditions (C1)–(C3) and (C5) hold. Define $\boldsymbol{s}_n(\boldsymbol{\theta}) := \boldsymbol{s}(X_n, D_n; \boldsymbol{\theta})$ by (6) with $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Then, as $n \to \infty$, $\boldsymbol{s}_n(\boldsymbol{\theta_0})/n^{1/2}$ converges under $P_{\boldsymbol{\theta_0}}$ in distribution to an m-dimensional normally distributed random vector with mean zero and covariance matrix*

$$\begin{aligned}\boldsymbol{V} := & \left[\int_W \frac{\lambda(\boldsymbol{u};\boldsymbol{\theta_0})\rho(\boldsymbol{u})}{\lambda(\boldsymbol{u};\boldsymbol{\theta_0})+\rho(\boldsymbol{u})} z_k(\boldsymbol{u}) z_l(\boldsymbol{u}) \mathrm{d}\boldsymbol{u} \right]_{k,l=1}^{m} \\ & + \left[\int\int_{W \times W} \frac{\lambda(\boldsymbol{u};\boldsymbol{\theta_0})\lambda(\boldsymbol{v};\boldsymbol{\theta_0})\rho(\boldsymbol{u})\rho(\boldsymbol{v}) z_k(\boldsymbol{u}) z_l(\boldsymbol{v})}{(\lambda(\boldsymbol{u};\boldsymbol{\theta_0})+\rho(\boldsymbol{u}))(\lambda(\boldsymbol{v};\boldsymbol{\theta_0})+\rho(\boldsymbol{v}))} (g(\boldsymbol{u},\boldsymbol{v})-1) \mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v} \right]_{k,l=1}^{m}.\end{aligned}$$

Proof Under conditions (C1)–(C2) and recalling (6),

$$\frac{\boldsymbol{s}_n(\boldsymbol{\theta_0})}{n^{1/2}} = \sum_{\boldsymbol{x} \in X_n} \frac{\rho(\boldsymbol{x}) \boldsymbol{z}(\boldsymbol{x})}{n^{1/2}(\lambda(\boldsymbol{x};\boldsymbol{\theta_0})+\rho(\boldsymbol{x}))} - \sum_{\boldsymbol{x} \in D_n} \frac{\lambda(\boldsymbol{x};\boldsymbol{\theta_0}) \boldsymbol{z}(\boldsymbol{x})}{n^{1/2}(\lambda(\boldsymbol{x};\boldsymbol{\theta_0})+\rho(\boldsymbol{x}))}.$$

It consists of two first-order U-statistics defined on X_n and D_n:

$$\begin{aligned}\frac{\boldsymbol{s}_n(\boldsymbol{\theta_0})}{n^{1/2}} &= \frac{\boldsymbol{s}_n(X_n;\boldsymbol{\theta_0})}{n^{1/2}} - \frac{\boldsymbol{s}_n(D_n;\boldsymbol{\theta_0})}{n^{1/2}} \\ &= n^{1/2}\frac{1}{n}\sum_{i=1}^{n}\left\{\sum_{\boldsymbol{x}\in Y_i}\frac{\rho(\boldsymbol{x})\boldsymbol{z}(\boldsymbol{x})}{\lambda(\boldsymbol{x};\boldsymbol{\theta_0})+\rho(\boldsymbol{x})}\right\} \\ &\quad - n^{1/2}\frac{1}{n}\sum_{i=1}^{n}\left\{\sum_{\boldsymbol{x}\in E_i}\frac{\lambda(\boldsymbol{x};\boldsymbol{\theta_0})\boldsymbol{z}(\boldsymbol{x})}{\lambda(\boldsymbol{x};\boldsymbol{\theta_0})+\rho(\boldsymbol{x})}\right\}.\end{aligned}$$

We discuss the two first-order U-statistics separately.

Note that the term in the first curly brackets is the average of independent and identically distributed real-valued random vectors. By the Campbell–Mecke theorem and recalling Section 2.3,

$$\mathrm{M}_{\boldsymbol{\theta_0}} := \mathbb{E}_{\boldsymbol{\theta_0}}\left\{\sum_{\boldsymbol{x}\in Y_i}\frac{\rho(\boldsymbol{x})\boldsymbol{z}(\boldsymbol{x})}{\lambda(\boldsymbol{x};\boldsymbol{\theta_0})+\rho(\boldsymbol{x})}\right\} = \left[\int_W \frac{\rho(\boldsymbol{u})\lambda(\boldsymbol{u};\boldsymbol{\theta_0})z_l(\boldsymbol{u})}{\lambda(\boldsymbol{u};\boldsymbol{\theta_0})+\rho(\boldsymbol{u})}\mathrm{d}\boldsymbol{u}\right]_{l=1}^{m}$$

and, because Y_i is a Poisson point process under condition (C2),

$$\mathbf{Cov}_{\boldsymbol{\theta_0}}\left\{\sum_{\boldsymbol{x}\in Y_i}\frac{\rho(\boldsymbol{x})\boldsymbol{z}(\boldsymbol{x})}{\lambda(\boldsymbol{x};\boldsymbol{\theta_0})+\rho(\boldsymbol{x})}\right\} = \left[\int_W \frac{\lambda(\boldsymbol{u};\boldsymbol{\theta_0})\rho^2(\boldsymbol{u})z_k(\boldsymbol{u})z_l(\boldsymbol{u})}{(\lambda(\boldsymbol{u};\boldsymbol{\theta_0})+\rho(\boldsymbol{u}))^2}\mathrm{d}\boldsymbol{u}\right]_{k,l=1}^{m}. \tag{10}$$

By condition (C5), the covariate terms in the integrand are bounded. By condition (C2), $0 < \lambda(\boldsymbol{u};\boldsymbol{\theta_0})\rho^2(\boldsymbol{u})/(\lambda(\boldsymbol{u};\boldsymbol{\theta_0})+\rho(\boldsymbol{u}))^2 < \lambda(\boldsymbol{u};\boldsymbol{\theta_0})$ and $\lambda(\boldsymbol{u};\boldsymbol{\theta_0})$ is integrable. The multi-variate Lindeberg–Lévy central limit theorem implies that $n^{-1/2}(\boldsymbol{s}_n(X_n;\boldsymbol{\theta_0}) - n\mathrm{M}_{\boldsymbol{\theta_0}})$ converges under $P_{\boldsymbol{\theta_0}}$ in distribution to an m-dimensional normally distributed random vector with mean zero and covariance matrix given by (10).

Analogously, by condition (C3), $n^{-1/2}(\boldsymbol{s}_n(D_n,\boldsymbol{\theta_0})-n\mathrm{M}_{\boldsymbol{\theta_0}})$ converges under $P_{\boldsymbol{\theta_0}}$ in distribution to an m-dimensional normally distributed random vector with mean zero and covariance matrix

$$\begin{aligned}\mathbf{Cov}_{\boldsymbol{\theta_0}}\left\{\sum_{\boldsymbol{x}\in E_i}\frac{\lambda(\boldsymbol{x};\boldsymbol{\theta_0})\boldsymbol{z}(\boldsymbol{x})}{\lambda(\boldsymbol{x};\boldsymbol{\theta_0})+\rho(\boldsymbol{x})}\right\} &= \left[\int_W \frac{\lambda^2(\boldsymbol{u};\boldsymbol{\theta_0})\rho(\boldsymbol{u})z_k(\boldsymbol{u})z_l(\boldsymbol{u})}{(\lambda(\boldsymbol{u};\boldsymbol{\theta_0})+\rho(\boldsymbol{u}))^2}\mathrm{d}\boldsymbol{u}\right]_{k,l=1}^{m} \\ + \left[\int\int_{W\times W}\frac{\lambda(\boldsymbol{u};\boldsymbol{\theta_0})\lambda(\boldsymbol{v};\boldsymbol{\theta_0})\rho(\boldsymbol{u})\rho(\boldsymbol{v})z_k(\boldsymbol{u})z_l(\boldsymbol{v})}{(\lambda(\boldsymbol{u};\boldsymbol{\theta_0})+\rho(\boldsymbol{u}))(\lambda(\boldsymbol{v};\boldsymbol{\theta_0})+\rho(\boldsymbol{v}))}\right. & \left.(g(\boldsymbol{u},\boldsymbol{v})-1)\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v}\right]_{k,l=1}^{m}.\end{aligned}$$

By condition (C2),

$$0 < \frac{\lambda(\boldsymbol{u};\boldsymbol{\theta_0})\lambda(\boldsymbol{v};\boldsymbol{\theta_0})\rho(\boldsymbol{u})\rho(\boldsymbol{v})}{(\lambda(\boldsymbol{u};\boldsymbol{\theta_0})+\rho(\boldsymbol{u}))(\lambda(\boldsymbol{v};\boldsymbol{\theta_0})+\rho(\boldsymbol{v}))} < \lambda(\boldsymbol{u};\boldsymbol{\theta_0})\lambda(\boldsymbol{v};\boldsymbol{\theta_0})$$

and both $\lambda(\boldsymbol{u};\boldsymbol{\theta_0})$ and $\lambda(\boldsymbol{v};\boldsymbol{\theta_0})$ are integrable.

Applying Lévy's continuity theorem and using the independence of X_n and D_n under condition (C1), the weak limits for the two first-order U-statistics can be combined, which completes the proof.

Remark 2 The pair correlation function $g(\boldsymbol{u}, \boldsymbol{v})$ of E_i can be tuned to control the covariance matrix $\boldsymbol{V}$. When E_i, and thus D_n, is a Poisson point process, $g(\boldsymbol{u}, \boldsymbol{v}) \equiv 1$ and the second term in $\boldsymbol{V}$ vanishes. Moreover, considering regular point processes E_i, whose $g(\boldsymbol{u}, \boldsymbol{v}) < 1$, may be useful to reduce the variance.

Recalling (9), we are now in a position to conjecture a central limit theorem for logistic regression estimators. For a formal proof, though, we need to analyze the error term in the Taylor series (8). The next theorem is concerned with asymptotic normality.

Theorem 2 *Assume that the conditions (C1)–(C3), (C5)–(C6) and (C8) hold. Define* $\boldsymbol{s}_n(\boldsymbol{\theta}) := \boldsymbol{s}(X_n, D_n; \boldsymbol{\theta})$ *by (6) with* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ *and let* $\widehat{\boldsymbol{\theta}}_n$ *be a consistent estimator sequence such that* $\boldsymbol{s}_n(\widehat{\boldsymbol{\theta}}_n) = \mathbf{0}$. *Then, as* $n \to \infty$, $n^{1/2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0})$ *converges under* $P_{\boldsymbol{\theta_0}}$ *in distribution to an* m*-dimensional normally distributed random vector with mean zero and covariance matrix* $\boldsymbol{U}^{-1}\boldsymbol{V}(\boldsymbol{U}^{-1})^\top$, *where* $\boldsymbol{U}, \boldsymbol{V}$ *are as defined in Lemma 2 and Lemma 3.*

Proof Consider the error term in the Taylor series (8) which, under conditions (C1) and (C2) and recalling (6), reads

$$\frac{1}{2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0})^\top \nabla^2 \boldsymbol{s}_{n,i}(\boldsymbol{\theta'^{,i}})(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}) = \frac{1}{2}\sum_{k=1}^{m}\sum_{l=1}^{m}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0})_k(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0})_l \times$$
$$\left\{ \sum_{\boldsymbol{x} \in X_n \cup D_n} \frac{\left(\lambda(\boldsymbol{x}; \boldsymbol{\theta'^{,i}}) - \rho(\boldsymbol{x})\right) \lambda(\boldsymbol{x}; \boldsymbol{\theta'^{,i}})\rho(\boldsymbol{x}) z_i(\boldsymbol{x}) z_k(\boldsymbol{x}) z_l(\boldsymbol{x})}{\left(\lambda(\boldsymbol{x}; \boldsymbol{\theta'^{,i}}) + \rho(\boldsymbol{x})\right)^3} \right\}$$

for all choices of $\boldsymbol{\theta'^{,i}}$ among the convex combinations of $\widehat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta_0}$. By condition (C6), $\boldsymbol{\theta'^{,i}}$ lies in $\boldsymbol{\Theta}$.

First of all, note that

$$\sum_{\boldsymbol{x} \in X_n \cup D_n} \frac{\left(\lambda(\boldsymbol{x}; \boldsymbol{\theta'^{,i}}) - \rho(\boldsymbol{x})\right) \lambda(\boldsymbol{x}; \boldsymbol{\theta'^{,i}})\rho(\boldsymbol{x}) z_i(\boldsymbol{x}) z_k(\boldsymbol{x}) z_l(\boldsymbol{x})}{n\left(\lambda(\boldsymbol{x}; \boldsymbol{\theta'^{,i}}) + \rho(\boldsymbol{x})\right)^3} \tag{11}$$

is bounded in absolute value by

$$G_{n,ikl} := \sum_{\boldsymbol{x} \in X_n \cup D_n} \frac{|z_i(\boldsymbol{x}) z_k(\boldsymbol{x}) z_l(\boldsymbol{x})|}{n}.$$

This bound does not depend on $\boldsymbol{\theta'^{,i}}$ and thus does not depend on $\widehat{\boldsymbol{\theta}}_n$. By the Campbell–Mecke theorem,

$$\mathbb{E}_{\boldsymbol{\theta_0}}\left\{G_{n,ikl}\right\} = \int_W \left(\lambda(\boldsymbol{u}; \boldsymbol{\theta_0}) + \rho(\boldsymbol{u})\right) |z_i(\boldsymbol{u}) z_k(\boldsymbol{u}) z_l(\boldsymbol{u})| \, \mathrm{d}\boldsymbol{u}.$$

Under conditions (C2) and (C5), $\mathbb{E}_{\boldsymbol{\theta_0}}\{G_{n,ikl}\}$ is a non-negative constant. If $\mathbb{E}_{\boldsymbol{\theta_0}}\{G_{n,ikl}\}$ is strictly positive, by Markov's inequality, for any $\delta > 0$, there always exists a finite $H_{ikl}(\delta) := \mathbb{E}_{\boldsymbol{\theta_0}}\{G_{n,ikl}\}/\delta$ such that

$$\mathbb{P}_{\boldsymbol{\theta_0}}\left\{|G_{n,ikl}| \geq H_{ikl}(\delta)\right\} \leq \delta.$$

If $\mathbb{E}_{\boldsymbol{\theta_0}}\{G_{n,ikl}\} = 0$, the product of the three covariate terms $z_i z_k z_l$ is zero almost everywhere on W and therefore $|G_{n,ikl}| = 0$ almost surely. Consequently $\mathbb{P}_{\boldsymbol{\theta_0}}\{|G_{n,ikl}| \geq H_{ikl}(\delta)\} = 0$ for any $H_{ikl}(\delta) > 0$. Since $H_{ikl}(\delta)$ depends only on δ but not on n, $G_{n,ikl}$, and thus (11), is bounded in probability under $P_{\boldsymbol{\theta_0}}$. This result applies to every component of (11) with $1 \leq i, k, l \leq m$.

Move back to the Taylor expansion (8). By Lemma 2, $-\nabla \boldsymbol{s}_n(\boldsymbol{\theta_0})/n$ converges almost surely, and thus in probability, under $P_{\boldsymbol{\theta_0}}$ to $\boldsymbol{U}$. Collecting all i with $1 \leq i \leq m$ and recalling the results obtained above, (8) can be rewritten as

$$\left[\boldsymbol{U} + o_P(1) - \frac{1}{2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0})^\top O_P(1)\right](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}) = \frac{\boldsymbol{s}_n(\boldsymbol{\theta_0})}{n}.$$

Because of the assumed consistency of the estimating sequence, $(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0})^\top$ times $O_P(1)$ converges under $P_{\boldsymbol{\theta_0}}$ in probability to zero. Furthermore, by condition (C8), the matrix $\boldsymbol{U}$ is invertible. Thus, $\boldsymbol{U} + o_P(1)$ is also invertible with a probability tending to one as $n \to \infty$. Multiplication by this inverse yields that

$$n^{1/2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}) = [\boldsymbol{U} + o_P(1)]^{-1}\,\frac{\boldsymbol{s}_n(\boldsymbol{\theta_0})}{n^{1/2}}.$$

The remainder of the proof is a straightforward application of Slutsky's theorem. Obviously, $[\boldsymbol{U} + o_P(1)]^{-1}$ converges under $P_{\boldsymbol{\theta_0}}$ in probability to $\boldsymbol{U}^{-1}$. By Lemma 3, $\boldsymbol{s}_n(\boldsymbol{\theta_0})/n^{1/2}$ converges under $P_{\boldsymbol{\theta_0}}$ in distribution to an m-dimensional normally distributed random vector with mean zero and covariance matrix $\boldsymbol{V}$. Thus, as $n \to \infty$, $n^{1/2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0})$ converges under $P_{\boldsymbol{\theta_0}}$ in distribution to an m-dimensional normally distributed random vector with mean zero and covariance matrix $\boldsymbol{U}^{-1}\boldsymbol{V}(\boldsymbol{U}^{-1})^\top$.

Remark 3 In Theorem 1, we have proved that $\widehat{\boldsymbol{\theta}}_n$ is a strongly consistent estimator. Here, observing that $\boldsymbol{U}$ and $\boldsymbol{V}$ are constant matrices, we thus provide the convergence rate: $(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0})_l = O_P(n^{-1/2})$ for every $1 \leq l \leq m$.

3.3 Existence of the estimator

Note that both strong consistency and asymptotic normality of the logistic regression estimator rely on the attainment of $\widehat{\boldsymbol{\theta}}_n$. For the sake of completeness, the next theorem deals with existence of the estimator.

Theorem 3 *Assume that the conditions (C1)–(C5) and (C8) hold. Define* $\boldsymbol{s}_n(\boldsymbol{\theta}) \coloneqq \boldsymbol{s}(X_n, D_n; \boldsymbol{\theta})$ *by (6) with* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. *Then, for every* n, *an estimator* $\widehat{\boldsymbol{\theta}}_n$ *exists that solves* $\boldsymbol{s}_n(\widehat{\boldsymbol{\theta}}_n) = \boldsymbol{0}$ *with a probability tending to one as* $n \to \infty$. *Moreover,* $\widehat{\boldsymbol{\theta}}_n$ *can be chosen to be consistent.*

Proof The proof follows from an appeal to Sørensen (1999, Corollary 2.6). By condition (C2), $\boldsymbol{s}_n(\boldsymbol{\theta})$ is continuously differentiable with respect to $\boldsymbol{\theta}$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Thus, we only need to verify Sørensen (1999, Condition 2.5). For elegance of writing, we omit the proof here but provide a proof in more general cases in Theorem 9.

3.4 Construction of asymptotic confidence intervals

Central limit theorems such as Theorem 2 can be used to construct approximate confidence intervals, a prime tool for uncertainty quantification in applications, for instance, in Lu et al. (2023). Note that the entries in the covariance matrix $n^{-1}\boldsymbol{U}^{-1}\boldsymbol{V}(\boldsymbol{U}^{-1})^\top$ of $\widehat{\boldsymbol{\theta}}_n$ decrease as n increases. However, as they depend on the unknown true intensity function $\lambda(\boldsymbol{u};\boldsymbol{\theta_0})$, they must be estimated. For the Poisson point process model considered in Lu et al. (2023), we used a simple plug-in estimator of $\lambda(\boldsymbol{u};\boldsymbol{\theta_0})$ based on the dummy point process D_n. Here, we investigate consistent estimators based on replications.

Consider two sequences of independent and identically distributed point processes $A_{1,n} := \cup_{i=1}^n B_{1,i}$ and $A_{2,n} := \cup_{i=1}^n B_{2,i}$ defined on W that are mutually independent and also independent from X_n and D_n. Assume that $B_{1,i}$ and $B_{2,i}$ have the same intensity function $\rho(\boldsymbol{u})$ as E_i. Furthermore, let $B_{1,i}$ be a Poisson point process and $B_{2,i}$ distributed as E_i having pair correlation function $g(\boldsymbol{u},\boldsymbol{v})$. We propose estimators for $\boldsymbol{U}$ and $\boldsymbol{V}$ defined in Theorem 2 as follows:

$$n\widehat{\boldsymbol{U}} := \sum_{i=1}^n \left[\sum_{\boldsymbol{x}\in B_{1,i}} \frac{\lambda(\boldsymbol{x};\widehat{\boldsymbol{\theta}}_n) z_k(\boldsymbol{x}) z_l(\boldsymbol{x})}{\lambda(\boldsymbol{x};\widehat{\boldsymbol{\theta}}_n)+\rho(\boldsymbol{x})} \right]_{k,l=1}^m$$

and

$$\begin{aligned} n\widehat{\boldsymbol{V}} := n\widehat{\boldsymbol{U}} &+ \sum_{i=1}^n \left[\sum_{\boldsymbol{x},\boldsymbol{y}\in B_{2,i}}^{\neq} \frac{\lambda(\boldsymbol{x};\widehat{\boldsymbol{\theta}}_n) z_k(\boldsymbol{x})}{\lambda(\boldsymbol{x};\widehat{\boldsymbol{\theta}}_n)+\rho(\boldsymbol{x})} \frac{\lambda(\boldsymbol{y};\widehat{\boldsymbol{\theta}}_n) z_l(\boldsymbol{y})}{\lambda(\boldsymbol{y};\widehat{\boldsymbol{\theta}}_n)+\rho(\boldsymbol{y})} \right]_{k,l=1}^m \\ &- \sum_{i=1}^n \left[\sum_{\boldsymbol{x},\boldsymbol{y}\in B_{1,i}}^{\neq} \frac{\lambda(\boldsymbol{x};\widehat{\boldsymbol{\theta}}_n) z_k(\boldsymbol{x})}{\lambda(\boldsymbol{x};\widehat{\boldsymbol{\theta}}_n)+\rho(\boldsymbol{x})} \frac{\lambda(\boldsymbol{y};\widehat{\boldsymbol{\theta}}_n) z_l(\boldsymbol{y})}{\lambda(\boldsymbol{y};\widehat{\boldsymbol{\theta}}_n)+\rho(\boldsymbol{y})} \right]_{k,l=1}^m . \end{aligned}$$

Recall that $\rho(\boldsymbol{u})$ and $g(\boldsymbol{u},\boldsymbol{v})$ are known explicitly by assumption.

Theorem 4 *Assume that the conditions (C1)–(C3) and (C5)–(C6) hold. Define* $\boldsymbol{s}_n(\boldsymbol{\theta}) := \boldsymbol{s}(X_n, D_n; \boldsymbol{\theta})$ *by (6) with* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ *and let* $\widehat{\boldsymbol{\theta}}_n$ *be a consistent estimator sequence such that* $\boldsymbol{s}_n(\widehat{\boldsymbol{\theta}}_n) = \boldsymbol{0}$. *Then* $\widehat{\boldsymbol{U}}$ *and* $\widehat{\boldsymbol{V}}$ *are consistent estimators of the matrices* $\boldsymbol{U}$ *and* $\boldsymbol{V}$ *defined in Lemma 2 and Lemma 3.*

Proof To show the consistency of approximating $n^{-1}\boldsymbol{U}^{-1}\boldsymbol{V}(\boldsymbol{U}^{-1})^\top$ by $(n\widehat{\boldsymbol{U}})^{-1}$ $n\widehat{\boldsymbol{V}}$ $((n\widehat{\boldsymbol{U}})^{-1})^\top$ using estimators $n\widehat{\boldsymbol{U}}$ and $n\widehat{\boldsymbol{V}}$, we need to prove that $\widehat{\boldsymbol{U}}$ and $\widehat{\boldsymbol{V}}$ converge in $P_{\boldsymbol{\theta_0}}$-probability to $\boldsymbol{U}$ and $\boldsymbol{V}$, respectively.

For

$$\widehat{\boldsymbol{U}} = \frac{1}{n}\sum_{i=1}^n \left[\sum_{\boldsymbol{x}\in B_{1,i}} \frac{\lambda(\boldsymbol{x};\widehat{\boldsymbol{\theta}}_n) z_k(\boldsymbol{x}) z_l(\boldsymbol{x})}{\lambda(\boldsymbol{x};\widehat{\boldsymbol{\theta}}_n)+\rho(\boldsymbol{x})} \right]_{k,l=1}^m ,$$

we prove component-wise convergence. First, we show that, for every $1 \leq k, l \leq m$,

$$\left| \frac{1}{n} \sum_{i=1}^{n} \sum_{\boldsymbol{x} \in B_{1,i}} \frac{\lambda(\boldsymbol{x}; \widehat{\boldsymbol{\theta}}_n) z_k(\boldsymbol{x}) z_l(\boldsymbol{x})}{\lambda(\boldsymbol{x}; \widehat{\boldsymbol{\theta}}_n) + \rho(\boldsymbol{x})} - \frac{1}{n} \sum_{i=1}^{n} \sum_{x \in B_{1,i}} \frac{\lambda(\boldsymbol{x}; \boldsymbol{\theta_0}) z_k(\boldsymbol{x}) z_l(\boldsymbol{x})}{\lambda(\boldsymbol{x}; \boldsymbol{\theta_0}) + \rho(\boldsymbol{x})} \right| = o_P(1).$$

Since $\lambda(\boldsymbol{x}; \boldsymbol{\theta})$ is differentiable with respect to $\boldsymbol{\theta}$, one can derive the first-order Taylor series for the terms between the absolute brackets and bound the left-hand side above by

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{\boldsymbol{x} \in B_{1,i}} \left| \frac{\lambda(\boldsymbol{x}; \bar{\boldsymbol{\theta}}) \rho(\boldsymbol{x}) z_k(\boldsymbol{x}) z_l(\boldsymbol{x})}{\left(\lambda(\boldsymbol{x}; \bar{\boldsymbol{\theta}}) + \rho(\boldsymbol{x})\right)^2} \sum_{j=1}^{m} \left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}\right)_j z_j(\boldsymbol{x}) \right|,$$

where $\bar{\boldsymbol{\theta}}$ is a convex combination of $\widehat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta_0}$ lying in Θ by condition (C6). Considering that $0 < \lambda(\boldsymbol{x}; \bar{\boldsymbol{\theta}})\rho(\boldsymbol{x}) / \, (\lambda(\boldsymbol{x}; \bar{\boldsymbol{\theta}}) + \rho(\boldsymbol{x}))^2 < 1$, by condition (C5), one can find a constant vector $\boldsymbol{c} = [c_1, \ldots, c_m]^\top$ such that the term above is further bounded by

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{\boldsymbol{x} \in B_{1,i}} |z_k(\boldsymbol{x}) z_l(\boldsymbol{x})| \left| \sum_{j=1}^{m} \left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}\right)_j c_j \right|. \tag{12}$$

Since $\widehat{\boldsymbol{\theta}}_n$ depends on X_n and D_n while $A_{1,n} = \cup_{i=1}^{n} B_{1,i}$ is independent of X_n, D_n, one may analyze the two terms $n^{-1} \sum_{i=1}^{n} \sum_{\boldsymbol{x} \in B_{1,i}} |z_k(\boldsymbol{x}) \; z_l(\boldsymbol{x})|$ and $|\sum_{j=1}^{m} (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0})_j c_j|$ separately. Note that the former is the average of independent and identically distributed real-valued random variables. By the Campbell–Mecke theorem,

$$\mathbb{E}_{\boldsymbol{\theta_0}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{\boldsymbol{x} \in B_{1,i}} |z_k(\boldsymbol{x}) z_l(\boldsymbol{x})| \right\} = \int_W |z_k(\boldsymbol{u}) z_l(\boldsymbol{u})| \, \rho(\boldsymbol{u}) \mathrm{d}\boldsymbol{u}.$$

Condition (C5) implies that the covariate terms in the integrand are bounded and condition (C2) that the intensity function ρ is integrable on W. By Kolmogorov's strong law of large numbers, $n^{-1} \sum_{i=1}^{n} \sum_{\boldsymbol{x} \in B_{1,i}} |z_k(\boldsymbol{x}) z_l(\boldsymbol{x})|$ converges $P_{\boldsymbol{\theta_0}}$-almost surely, and thus in $P_{\boldsymbol{\theta_0}}$-probability, to the integral above which is $O(1)$. Since, by assumption, $|\sum_{j=1}^{m} (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0})_j c_j|$ converges in probability to zero, (12) converges in $P_{\boldsymbol{\theta_0}}$-probability to zero.

On the other hand, Kolmogorov's strong law of large numbers also implies that

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{\boldsymbol{x} \in B_{1,i}} \frac{\lambda(\boldsymbol{x}; \boldsymbol{\theta_0}) z_k(\boldsymbol{x}) z_l(\boldsymbol{x})}{\lambda(\boldsymbol{x}; \boldsymbol{\theta_0}) + \rho(\boldsymbol{x})} \overset{P_{\boldsymbol{\theta_0}}-a.s.}{\rightarrow} \int_W \frac{\lambda(\boldsymbol{u}; \boldsymbol{\theta_0}) \rho(\boldsymbol{u})}{\lambda(\boldsymbol{u}; \boldsymbol{\theta_0}) + \rho(\boldsymbol{u})} z_k(\boldsymbol{u}) z_l(\boldsymbol{u}) \mathrm{d}\boldsymbol{u},$$

from which convergence in $P_{\boldsymbol{\theta_0}}$-probability follows. By the triangle inequality, $\widehat{\boldsymbol{U}}$ converges in $P_{\boldsymbol{\theta_0}}$-probability to $\boldsymbol{U}$.

The remainder of the proof mostly proceeds along similar lines for the terms in $\widehat{\boldsymbol{V}}$. For instance, we need to show that, for every $1 \leq k, l \leq m$,

$$\frac{1}{n}\sum_{i=1}^{n}\sum\nolimits^{\neq}_{\boldsymbol{x},\boldsymbol{y}\in B_{2,i}} \frac{\lambda(\boldsymbol{x};\widehat{\boldsymbol{\theta}}_n)z_k(\boldsymbol{x})}{\lambda(\boldsymbol{x};\widehat{\boldsymbol{\theta}}_n)+\rho(\boldsymbol{x})}\frac{\lambda(\boldsymbol{y};\widehat{\boldsymbol{\theta}}_n)z_l(\boldsymbol{y})}{\lambda(\boldsymbol{y};\widehat{\boldsymbol{\theta}}_n)+\rho(\boldsymbol{y})}$$
$$\stackrel{P_{\boldsymbol{\theta_0}}}{\to} \int\int_{W\times W} \frac{\lambda(\boldsymbol{u};\boldsymbol{\theta_0})\lambda(\boldsymbol{v};\boldsymbol{\theta_0})\rho(\boldsymbol{u})\rho(\boldsymbol{v})z_k(\boldsymbol{u})z_l(\boldsymbol{v})}{(\lambda(\boldsymbol{u};\boldsymbol{\theta_0})+\rho(\boldsymbol{u}))(\lambda(\boldsymbol{v};\boldsymbol{\theta_0})+\rho(\boldsymbol{v}))} g(\boldsymbol{u},\boldsymbol{v})\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v}.$$

To do so, we first note that the absolute value of

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\sum\nolimits^{\neq}_{\boldsymbol{x},\boldsymbol{y}\in B_{2,i}} \frac{\lambda(\boldsymbol{x};\widehat{\boldsymbol{\theta}}_n)z_k(\boldsymbol{x})}{\lambda(\boldsymbol{x};\widehat{\boldsymbol{\theta}}_n)+\rho(\boldsymbol{x})}\frac{\lambda(\boldsymbol{y};\widehat{\boldsymbol{\theta}}_n)z_l(\boldsymbol{y})}{\lambda(\boldsymbol{y};\widehat{\boldsymbol{\theta}}_n)+\rho(\boldsymbol{y})}\right\}$$
$$-\frac{1}{n}\sum_{i=1}^{n}\left\{\sum\nolimits^{\neq}_{\boldsymbol{x},\boldsymbol{y}\in B_{2,i}} \frac{\lambda(\boldsymbol{x};\boldsymbol{\theta_0})z_k(\boldsymbol{x})}{\lambda(\boldsymbol{x};\boldsymbol{\theta_0})+\rho(\boldsymbol{x})}\frac{\lambda(\boldsymbol{y};\boldsymbol{\theta_0})z_l(\boldsymbol{y})}{\lambda(\boldsymbol{y};\boldsymbol{\theta_0})+\rho(\boldsymbol{y})}\right\}$$

is $o_P(1)$.This can be readily seen from arguments similar to those in the proof above. In a second step, we show that

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\sum\nolimits^{\neq}_{\boldsymbol{x},\boldsymbol{y}\in B_{2,i}} \frac{\lambda(\boldsymbol{x};\boldsymbol{\theta_0})z_k(\boldsymbol{x})}{\lambda(\boldsymbol{x};\boldsymbol{\theta_0})+\rho(\boldsymbol{x})}\frac{\lambda(\boldsymbol{y};\boldsymbol{\theta_0})z_l(\boldsymbol{y})}{\lambda(\boldsymbol{y};\boldsymbol{\theta_0})+\rho(\boldsymbol{y})}\right\}$$

converges in $P_{\boldsymbol{\theta_0}}$-probability to

$$\int\int_{W\times W} \frac{\lambda(\boldsymbol{u};\boldsymbol{\theta_0})\lambda(\boldsymbol{v};\boldsymbol{\theta_0})\rho(\boldsymbol{u})\rho(\boldsymbol{v})z_k(\boldsymbol{u})z_l(\boldsymbol{v})}{(\lambda(\boldsymbol{u};\boldsymbol{\theta_0})+\rho(\boldsymbol{u}))(\lambda(\boldsymbol{v};\boldsymbol{\theta_0})+\rho(\boldsymbol{v}))} g(\boldsymbol{u},\boldsymbol{v})\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v}$$

by Kolmogorov's strong laws of large numbers, noting that $g(\boldsymbol{u},\boldsymbol{v})$ is bounded by condition (C3) so that the limit is finite.

Remark 4 Under condition (C8), the asymptotic covariance matrix $n^{-1}\boldsymbol{U}^{-1}$ $\boldsymbol{V}(\boldsymbol{U}^{-1})^{\top}$ can be estimated consistently by $(n\widehat{\boldsymbol{U}})^{-1}\; n\widehat{\boldsymbol{V}}\; ((n\widehat{\boldsymbol{U}})^{-1})^{\top}$, using the estimators $n\widehat{\boldsymbol{U}}$ and $n\widehat{\boldsymbol{V}}$ in Theorem 4.

4 Extensions of the Asymptotic Results

The theory developed in Section 3 can be extended in two directions. First, the assumption that X_n is a Poisson point process may be relaxed. In Section 4.1, we give sufficient conditions for consistency and asymptotic normality and discuss the applicability for some specific families of spatial point processes. Second, the Campbell–Mecke theorem provides the theoretical foundation for a wide range of unbiased estimating equations by choosing vector functions $\boldsymbol{f}$ other than (3). In Section 4.2, we extend the central limit theorem (cf., Theorem 2) to the general case, introduce the necessary background and provide complete proofs.

4.1 Infill asymptotics for general point process models

For general point processes, we shall need the following modified conditions.

(C9) Y_i is a point process with intensity function $\lambda(\boldsymbol{u};\boldsymbol{\theta})$ given by (1), where $b > 0$ is an integrable function, $\boldsymbol{z}$ is a measurable vector of covariates and the parameter vector $\boldsymbol{\theta}$ lies in an open set $\boldsymbol{\Theta} \subset \mathbb{R}^m$; E_i has integrable intensity function $\rho(\boldsymbol{u}) > 0$.

(C10) Both E_i and Y_i have bounded pair correlation functions $g(\boldsymbol{u},\boldsymbol{v})$ and $h(\boldsymbol{u},\boldsymbol{v})$: $\sup_{(\boldsymbol{u},\boldsymbol{v})\in W\times W} g(\boldsymbol{u},\boldsymbol{v}) < \infty$ and $\sup_{(\boldsymbol{u},\boldsymbol{v})\in W\times W} h(\boldsymbol{u},\boldsymbol{v}) < \infty$.

The next theorems are concerned with strong consistency, asymptotic normality and existence of logistic regression estimators for general point processes in analogy to Theorems 1, 2 and 3.

Theorem 5 *Assume that the conditions (C1), (C4)–(C7) and (C9) hold. Define $l_n(\boldsymbol{\theta}) := l(X_n, D_n;\boldsymbol{\theta})$ by (5) with $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. If $\widehat{\boldsymbol{\theta}}_n := \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} l_n(\boldsymbol{\theta})$ exists for all $n \in \mathbb{N}$, then $\widehat{\boldsymbol{\theta}}_n$ converges $P_{\boldsymbol{\theta_0}}$-almost surely to $\boldsymbol{\theta_0}$ as $n \to \infty$.*

Proof The same proofs as of Lemma 1 and Theorem 1 can be applied straightforwardly.

Theorem 6 *Assume that the conditions (C1), (C5)–(C6) and (C8)–(C10) hold. Define $\boldsymbol{s}_n(\boldsymbol{\theta}) := \boldsymbol{s}(X_n, D_n;\, \boldsymbol{\theta})$ by (6) with $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and let $\widehat{\boldsymbol{\theta}}_n$ be a consistent estimator sequence such that $\boldsymbol{s}_n(\widehat{\boldsymbol{\theta}}_n) = \boldsymbol{0}$. Then, as $n \to \infty$, $n^{1/2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0})$ converges under $P_{\boldsymbol{\theta_0}}$ in distribution to an m-dimensional normally distributed random vector with mean zero and covariance matrix $\boldsymbol{U}^{-1}\boldsymbol{V}(\boldsymbol{U}^{-1})^\top$, where $\boldsymbol{U}$ is as defined in Lemma 2 and $\boldsymbol{V}$ is now given by*

$$\begin{aligned}\boldsymbol{V} :=& \left[\int_W \frac{\lambda(\boldsymbol{u};\boldsymbol{\theta_0})\rho(\boldsymbol{u})}{\lambda(\boldsymbol{u};\boldsymbol{\theta_0})+\rho(\boldsymbol{u})} z_k(\boldsymbol{u})z_l(\boldsymbol{u})\mathrm{d}\boldsymbol{u}\right]_{k,l=1}^m \\ &+ \left[\int\int_{W\times W} \frac{\lambda(\boldsymbol{u};\boldsymbol{\theta_0})\lambda(\boldsymbol{v};\boldsymbol{\theta_0})\rho(\boldsymbol{u})\rho(\boldsymbol{v})z_k(\boldsymbol{u})z_l(\boldsymbol{v})}{(\lambda(\boldsymbol{u};\boldsymbol{\theta_0})+\rho(\boldsymbol{u}))(\lambda(\boldsymbol{v};\boldsymbol{\theta_0})+\rho(\boldsymbol{v}))}(g(\boldsymbol{u},\boldsymbol{v})-1)\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v}\right]_{k,l=1}^m \\ &+ \left[\int\int_{W\times W} \frac{\lambda(\boldsymbol{u};\boldsymbol{\theta_0})\lambda(\boldsymbol{v};\boldsymbol{\theta_0})\rho(\boldsymbol{u})\rho(\boldsymbol{v})z_k(\boldsymbol{u})z_l(\boldsymbol{v})}{(\lambda(\boldsymbol{u};\boldsymbol{\theta_0})+\rho(\boldsymbol{u}))(\lambda(\boldsymbol{v};\boldsymbol{\theta_0})+\rho(\boldsymbol{v}))}(h(\boldsymbol{u},\boldsymbol{v})-1)\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v}\right]_{k,l=1}^m .\end{aligned}$$

Proof Similar proofs as of Lemma 2, Lemma 3 and Theorem 2 can be used. Note that, by condition (C10), the pair correlation function of Y_i, $h(\boldsymbol{u},\boldsymbol{v})$, is bounded, which ensures that the entries of $\boldsymbol{V}$ are finite.

Theorem 7 *Assume that the conditions (C1), (C4)–(C5) and (C8)–(C10) hold. Define $\boldsymbol{s}_n(\boldsymbol{\theta}) := \boldsymbol{s}(X_n, D_n;\boldsymbol{\theta})$ by (6) with $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Then, for every n, an estimator $\widehat{\boldsymbol{\theta}}_n$ exists that solves $\boldsymbol{s}_n(\widehat{\boldsymbol{\theta}}_n) = \boldsymbol{0}$ with a probability tending to one as $n \to \infty$. Moreover, $\widehat{\boldsymbol{\theta}}_n$ can be chosen to be consistent.*

Proof A similar proof as of Theorem 3 can also be applied.

As an example, consider an inhomogeneous Neyman-Scott point process (Waagepetersen 2007) with a stationary Poisson parent point process of intensity $\kappa > 0$ and clusters that are independent Poisson point processes with intensity functions $\mu k(\boldsymbol{u}-\boldsymbol{v})\exp[\boldsymbol{\theta}^\top \boldsymbol{z}(\boldsymbol{u})]$ centred around the parent points $\boldsymbol{v}$. Here, μ is the baseline offspring intensity and $k(\cdot)$ is some probability density function determining the spread of offspring points in clusters. The intensity function of the combined offspring process then reads $\kappa\mu\exp[\boldsymbol{\theta}^\top \boldsymbol{z}(\boldsymbol{u})]$, which is of the log linear form (1) upon adding a component with the entry one to $\boldsymbol{z}$ and combining κ and μ.

Also the strength of interactions may enter into the intensity function. For instance, consider a log-Gaussian Cox process (Coles and Jones 1991; Møller, Syversveen and Waagepetersen 1998) driven by a Gaussian random field with mean function $(\boldsymbol{\theta}^\top \boldsymbol{z}(\boldsymbol{u}))\log b(\boldsymbol{u})$ and covariance function $\sigma^2 c(\boldsymbol{u},\boldsymbol{v})$ where $c(\boldsymbol{u},\boldsymbol{v})$ is some correlation function and $\sigma^2 > 0$ denotes the variance at distance zero. Its intensity function reads $b(\boldsymbol{u})\exp[\boldsymbol{\theta}^\top \boldsymbol{z}(\boldsymbol{u}) + \sigma^2/2]$, which still has the log-linear form (1) upon adding a component with the entry one to $\boldsymbol{z}$. Should $c(\boldsymbol{u},\boldsymbol{v})$ depend on further parameters, e.g. the decay rate of interactions, additional estimating equations are required.

To estimate the asymptotic covariance matrix (cf., Theorem 6), as in Section 3.4, consider two sequences of independent and identically distributed point processes $A_{1,n} := \cup_{i=1}^n B_{1,i}$ and $A_{2,n} := \cup_{i=1}^n B_{2,i}$ defined on W that are mutually independent and also independent from X_n and D_n. Again, assume that $B_{1,i}$ and $B_{2,i}$ have the same intensity function $\rho(\boldsymbol{u})$ as E_i. Furthermore, let $B_{1,i}$ be a Poisson point process and $B_{2,i}$ distributed as E_i having pair correlation function $g(\boldsymbol{u},\boldsymbol{v})$. Additionally, let $B_{3,i}$ be another sequence of independent replications from the same distribution as the data. Thus, their intensity function and pair correlation function are given by $\lambda(\boldsymbol{u};\boldsymbol{\theta_0})$ and $h(\boldsymbol{u},\boldsymbol{v})$. Write $A_{3,n} = \cup_{i=1}^n B_{3,i}$ for the union and assume that $A_{3,n}$ is independent from $X_n, D_n, A_{1,n}$ and $A_{2,n}$. With $\widehat{\boldsymbol{U}}$ as in Theorem 4, set

$$\begin{aligned} n\widehat{\boldsymbol{V}} := n\widehat{\boldsymbol{U}} &+ \sum_{i=1}^n \left[\sum_{\boldsymbol{x},\boldsymbol{y}\in B_{2,i}}^{\neq} \frac{\lambda(\boldsymbol{x};\widehat{\boldsymbol{\theta}}_n) z_k(\boldsymbol{x})}{\lambda(\boldsymbol{x};\widehat{\boldsymbol{\theta}}_n)+\rho(\boldsymbol{x})} \frac{\lambda(\boldsymbol{y};\widehat{\boldsymbol{\theta}}_n) z_l(\boldsymbol{y})}{\lambda(\boldsymbol{y};\widehat{\boldsymbol{\theta}}_n)+\rho(\boldsymbol{y})} \right]_{k,l=1}^m \\ &+ \sum_{i=1}^n \left[\sum_{\boldsymbol{x},\boldsymbol{y}\in B_{3,i}}^{\neq} \frac{\rho(\boldsymbol{x}) z_k(\boldsymbol{x})}{\lambda(\boldsymbol{x};\widehat{\boldsymbol{\theta}}_n)+\rho(\boldsymbol{x})} \frac{\rho(\boldsymbol{y}) z_l(\boldsymbol{y})}{\lambda(\boldsymbol{y};\widehat{\boldsymbol{\theta}}_n)+\rho(\boldsymbol{y})} \right]_{k,l=1}^m \\ &- 2\sum_{i=1}^n \left[\sum_{\boldsymbol{x},\boldsymbol{y}\in B_{1,i}}^{\neq} \frac{\lambda(\boldsymbol{x};\widehat{\boldsymbol{\theta}}_n) z_k(\boldsymbol{x})}{\lambda(\boldsymbol{x};\widehat{\boldsymbol{\theta}}_n)+\rho(\boldsymbol{x})} \frac{\lambda(\boldsymbol{y};\widehat{\boldsymbol{\theta}}_n) z_l(\boldsymbol{y})}{\lambda(\boldsymbol{y};\widehat{\boldsymbol{\theta}}_n)+\rho(\boldsymbol{y})} \right]_{k,l=1}^m . \end{aligned}$$

The following theorem establishes consistency.

Theorem 8 *Assume that the conditions (C1), (C5)–(C6) and (C9)–(C10) hold. Define* $\boldsymbol{s}_n(\boldsymbol{\theta}) := \boldsymbol{s}(X_n, D_n; \boldsymbol{\theta})$ *by (6) with* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ *and let* $\widehat{\boldsymbol{\theta}}_n$ *be a consistent estimator sequence such that* $\boldsymbol{s}_n(\widehat{\boldsymbol{\theta}}_n) = \boldsymbol{0}$. *Then* $\widehat{\boldsymbol{U}}$ *and* $\widehat{\boldsymbol{V}}$ *are consistent estimators of the matrices* $\boldsymbol{U}$ *and* $\boldsymbol{V}$ *defined in Lemma 2 and Theorem 6.*

Proof The same proof as of Theorem 4 can be applied. Note that, by condition (C10), $h(\boldsymbol{u},\boldsymbol{v})$ is bounded, which also ensures that

$$\int\int_{W\times W} \frac{\lambda(\boldsymbol{u};\boldsymbol{\theta_0})\lambda(\boldsymbol{v};\boldsymbol{\theta_0})\rho(\boldsymbol{u})\rho(\boldsymbol{v})z_k(\boldsymbol{u})z_l(\boldsymbol{v})}{(\lambda(\boldsymbol{u};\boldsymbol{\theta_0})+\rho(\boldsymbol{u}))(\lambda(\boldsymbol{v};\boldsymbol{\theta_0})+\rho(\boldsymbol{v}))} h(\boldsymbol{u},\boldsymbol{v})\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v}$$

is finite.

In the literature, bootstrapping and kernel techniques have been proposed to approximate the asymptotic covariance matrix. In contrast to the former (Guan and Loh 2007), our estimators are consistent and computationally more efficient. An example of the latter is the estimator proposed by Cœurjolly and Guan (2014), but in the increasing-domain regime.

4.2 General unbiased estimating equations

The Campbell–Mecke theorem gives rise to a great variety of unbiased estimating equations. Although strong consistency of these estimators in general may not hold due to the loss of the likelihood interpretation of logistic regression estimators, asymptotic normality and existence of the estimator hold under appropriate conditions.

Recalling (2), an unbiased estimating equation based on the Campbell–Mecke theorem (e.g., Daley and Vere-Jones 2008) can take the form

$$\boldsymbol{s}(X,D;\boldsymbol{\theta}) := \sum_{\boldsymbol{x}\in X} \boldsymbol{f}(\boldsymbol{x};\boldsymbol{\theta}) - \sum_{\boldsymbol{x}\in D} \boldsymbol{f}(\boldsymbol{x};\boldsymbol{\theta})\frac{\lambda(\boldsymbol{x};\boldsymbol{\theta})}{\rho(\boldsymbol{x})} = \boldsymbol{0}, \tag{13}$$

for some test vector function $\boldsymbol{f} : W \to \mathbb{R}^m$ such that every component of $\boldsymbol{f}$, denoted by f_i with $1 \le i \le m$, has the property that $f_i\lambda$ is absolutely integrable. For instance, the test function $\boldsymbol{f}$ may contain some non-negative quadrature weight functions (e.g., Waagepetersen 2008; Guan and Shen 2010) depending on the first- and second-order characteristics of X to reduce the bias caused by numerical approximation or ignoring the interaction structure. In the quasi-likelihood approach (Guan, Jalilian and Waagepetersen 2015) and the increasing-domain framework, the optimal $\boldsymbol{f}$ is the solution to a Fredholm integral equation.

To extend our central limit theorem to such unbiased estimating equations in the infill asymptotic regime, we shall need the following additional conditions.

(C11) The component functions $f_i(\boldsymbol{u},\boldsymbol{\theta})$ are twice continuously differentiable with respect to $\boldsymbol{\theta}$. Its components and the first- and second-order partial derivatives, denoted by $\frac{\partial}{\partial\theta_l}f_i(\boldsymbol{u};\boldsymbol{\theta})$ and $\frac{\partial^2}{\partial\theta_k\partial\theta_l}f_i(\boldsymbol{u};\boldsymbol{\theta})$, in a neighbourhood of $\boldsymbol{\theta_0}$ are bounded in absolute value by some functions $d_i(\boldsymbol{u}), d_{il}(\boldsymbol{u})$ and $d_{ikl}(\boldsymbol{u})$, respectively, that are absolutely integrable with respect to $\lambda(\boldsymbol{u};\boldsymbol{\theta_0})$.

(C12) The second-order partial derivatives satisfy an adapted Hölder condition in a neighbourhood of $\boldsymbol{\theta_0}$: there exist some $\alpha \in (0, 1]$ and some functions $e_{ikl}(\boldsymbol{u})$ which are absolutely integrable with respect to $\lambda(u; \boldsymbol{\theta_0})$ such that

$$\left| \frac{\partial^2}{\partial\theta_k \partial\theta_l} f_i(\boldsymbol{u}; \boldsymbol{\theta}) - \frac{\partial^2}{\partial\theta_k \partial\theta_l} f_i(\boldsymbol{u}; \boldsymbol{\theta_0}) \right| \leq e_{ikl}(\boldsymbol{u}) \|\boldsymbol{\theta} - \boldsymbol{\theta_0}\|^{\alpha}.$$

(C13) The $[m \times m]$-dimensional matrix $\widetilde{\boldsymbol{U}}$, whose (k, l)-th entry reads $\int_W f_k(\boldsymbol{u}; \boldsymbol{\theta_0})\lambda(\boldsymbol{u}; \boldsymbol{\theta_0}) z_l(\boldsymbol{u}) \mathrm{d}\boldsymbol{u}$ with $\boldsymbol{\theta_0} \in \boldsymbol{\Theta}$, is invertible.

(C14) The product of any two component functions, $f_k(\boldsymbol{u}; \boldsymbol{\theta_0}) f_l(\boldsymbol{u}; \boldsymbol{\theta_0})$, is absolutely integrable with respect to $\lambda(\boldsymbol{u}; \boldsymbol{\theta_0})$.

A few remarks on the conditions are appropriate. Conditions (C11) and (C12) restrict the smoothness of $\boldsymbol{f}$ and its first-order and second-order derivatives. Condition (C13) is the counterpart of condition (C8) for general unbiased estimating equations; the technical constraints in condition (C14) ensure that the entries in the covariance matrices of the corresponding estimators are finite.

For elegance of writing, we start our investigations with the following lemma.

Lemma 4 *Assume that the conditions (C1), (C5), (C9) and (C11) hold. Define $\boldsymbol{s}_n(\boldsymbol{\theta}) := \boldsymbol{s}(X_n, D_n; \boldsymbol{\theta})$ by (13) with $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Then, as $n \to \infty$, for all $\beta > 0$,*

$$\sup_{\boldsymbol{\theta'_k} \in C_{n,\beta}} \left\| -\frac{1}{n} \nabla \boldsymbol{s}_n(\boldsymbol{\theta'_1}, \ldots, \boldsymbol{\theta'_m}) - \widetilde{\boldsymbol{U}} \right\|$$

converges under $P_{\boldsymbol{\theta_0}}$ in probability to zero, where $C_{n,\beta} := \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta_0}\| \leq \beta/n^{1/2}\}$, $\nabla \boldsymbol{s}_n(\boldsymbol{\theta})$ is the $[m \times m]$-dimensional matrix containing the first-order partial derivatives of $\boldsymbol{s}_n$ with $\boldsymbol{s}_n(\boldsymbol{\theta'_1}, \ldots, \boldsymbol{\theta'_m})$ obtained upon replacing $\boldsymbol{\theta}$ by $\boldsymbol{\theta'_k}$ throughout the k-th row and

$$\widetilde{\boldsymbol{U}} := \left[\int_W f_k(\boldsymbol{u}; \boldsymbol{\theta_0}) \lambda(\boldsymbol{u}; \boldsymbol{\theta_0}) z_l(\boldsymbol{u}) \mathrm{d}\boldsymbol{u} \right]_{k,l=1}^m .$$

Proof Under conditions (C1), (C9) and (C11), recalling the unbiased estimating equation (13), $-\nabla \boldsymbol{s}_n(\boldsymbol{\theta'_1}, \ldots, \boldsymbol{\theta'_m})/n$ reads

$$\frac{1}{n} \left[\sum_{\boldsymbol{x} \in X_n} -\frac{\partial f_k(\boldsymbol{x}; \boldsymbol{\theta'_k})}{\partial \theta_l} + \sum_{\boldsymbol{x} \in D_n} \frac{\lambda(\boldsymbol{x}; \boldsymbol{\theta'_k})}{\rho(\boldsymbol{x})} \left(\frac{\partial f_k(\boldsymbol{x}; \boldsymbol{\theta'_k})}{\partial \theta_l} + f_k(\boldsymbol{x}; \boldsymbol{\theta'_k}) z_l(\boldsymbol{x}) \right) \right]_{k,l=1}^m .$$

Fix $\beta > 0$. We prove the convergence of the supremum component-wise.

First, by condition (C11), for large enough n and every $1 \leq i, k, l \leq m$,

$$|f_i(\boldsymbol{u}; \boldsymbol{\theta})| \leq d_i(\boldsymbol{u}), \quad \left| \frac{\partial f_i(\boldsymbol{u}; \boldsymbol{\theta})}{\partial \theta_l} \right| \leq d_{il}(\boldsymbol{u}), \quad \left| \frac{\partial^2 f_i(\boldsymbol{u}; \boldsymbol{\theta})}{\partial \theta_k \partial \theta_l} \right| \leq d_{ikl}(\boldsymbol{u}), \qquad (14)$$

for all $\boldsymbol{u} \in W$ and $\boldsymbol{\theta} \in C_{n,\beta}$. Note that, taking n large enough, $C_{n,\beta}$ lies entirely within $\boldsymbol{\Theta}$, as the parameter space $\boldsymbol{\Theta}$ is open by condition (C9). Furthermore, under condition (C9), the first-order Taylor expansion of $\lambda(u;\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ at $\boldsymbol{\theta_0}$ reads

$$\lambda(\boldsymbol{u};\boldsymbol{\theta}) - \lambda(\boldsymbol{u};\boldsymbol{\theta_0}) = b(\boldsymbol{u}) \exp\left[\underline{\boldsymbol{\theta}}^\top \boldsymbol{z}(\boldsymbol{u})\right] \sum_{i=1}^{m} (\boldsymbol{\theta} - \boldsymbol{\theta_0})_i z_i(\boldsymbol{u}),$$

where $\underline{\boldsymbol{\theta}}$ is a convex combination of $\boldsymbol{\theta}$ and $\boldsymbol{\theta_0}$. Note that if $\boldsymbol{\theta}$ lies in $C_{n,\beta}$, then so does $\underline{\boldsymbol{\theta}}$. By condition (C5), the covariate terms in the inner product $\underline{\boldsymbol{\theta}}^\top \boldsymbol{z}(\boldsymbol{u})$ and the sum above are bounded. Thus, recalling the definition of $C_{n,\beta}$, when $n \to \infty$, one obtains that $\lambda(\boldsymbol{u};\boldsymbol{\theta}) = \lambda(\boldsymbol{u};\boldsymbol{\theta_0}) + b(\boldsymbol{u})o(1)$ for all $\boldsymbol{u} \in W$ and $\boldsymbol{\theta} \in C_{n,\beta}$.

Next, consider $-\nabla \boldsymbol{s}_n(\boldsymbol{\theta'_1}, \ldots, \boldsymbol{\theta'_m})/n$. In the remainder of this proof, denote its (k,l)-th entry by $[-\nabla \boldsymbol{s}_{n,k}(\boldsymbol{\theta'_k})/n]_l$. The difference between $[-\nabla \boldsymbol{s}_{n,k}(\boldsymbol{\theta'_k})/n]_l$ and $[-\nabla \boldsymbol{s}_n(\boldsymbol{\theta_0})/n]_{k,l}$ reads

$$\begin{aligned} &-\frac{1}{n}\sum_{\boldsymbol{x}\in X_n} \frac{\partial f_k(\boldsymbol{x};\boldsymbol{\theta'_k})}{\partial \theta_l} + \frac{1}{n}\sum_{\boldsymbol{x}\in X_n} \frac{\partial f_k(\boldsymbol{x};\boldsymbol{\theta_0})}{\partial \theta_l} \\ &+ \frac{1}{n}\sum_{\boldsymbol{x}\in D_n} \frac{\lambda(\boldsymbol{x};\boldsymbol{\theta'_k})}{\rho(\boldsymbol{x})}\frac{\partial f_k(\boldsymbol{x};\boldsymbol{\theta'_k})}{\partial \theta_l} - \frac{1}{n}\sum_{\boldsymbol{x}\in D_n} \frac{\lambda(\boldsymbol{x};\boldsymbol{\theta_0})}{\rho(\boldsymbol{x})}\frac{\partial f_k(\boldsymbol{x};\boldsymbol{\theta_0})}{\partial \theta_l} \\ &+ \frac{1}{n}\sum_{\boldsymbol{x}\in D_n} \frac{\lambda(\boldsymbol{x};\boldsymbol{\theta'_k})}{\rho(\boldsymbol{x})} f_k(\boldsymbol{x};\boldsymbol{\theta'_k}) z_l(\boldsymbol{x}) - \frac{1}{n}\sum_{\boldsymbol{x}\in D_n} \frac{\lambda(\boldsymbol{x};\boldsymbol{\theta_0})}{\rho(\boldsymbol{x})} f_k(\boldsymbol{x};\boldsymbol{\theta_0}) z_l(\boldsymbol{x}). \end{aligned} \tag{15}$$

Below, we analyze the asymptotic behavior of the second line in (15). The analysis of the other two lines proceeds along similar lines.

Under condition (C11), the first-order Taylor expansion of $\frac{\partial}{\partial \theta_l} f_k(\boldsymbol{x};\boldsymbol{\theta'_k})$ with respect to $\boldsymbol{\theta}$ at $\boldsymbol{\theta_0}$ reads

$$\frac{\partial f_k(\boldsymbol{x};\boldsymbol{\theta'_k})}{\partial \theta_l} - \frac{\partial f_k(\boldsymbol{x};\boldsymbol{\theta_0})}{\partial \theta_l} = \sum_{i=1}^{m} (\boldsymbol{\theta'_k} - \boldsymbol{\theta_0})_i \frac{\partial^2 f_k(\boldsymbol{x};\boldsymbol{\theta''_k})}{\partial \theta_i \partial \theta_l},$$

where $\boldsymbol{\theta''_k}$ is a convex combination of $\boldsymbol{\theta'_k}$ and $\boldsymbol{\theta_0}$, and thus $\boldsymbol{\theta''_k} \in C_{n,\beta}$. Recalling that $\lambda(\boldsymbol{x};\boldsymbol{\theta'_k}) = \lambda(\boldsymbol{x};\boldsymbol{\theta_0}) + b(\boldsymbol{x})o(1)$, the second line in (15) is bounded in absolute value by

$$\sum_{\boldsymbol{x}\in D_n} \left\{ \sum_{i=1}^{m} \frac{\lambda(\boldsymbol{x};\boldsymbol{\theta_0})}{n\rho(\boldsymbol{x})} \left|(\boldsymbol{\theta'_k} - \boldsymbol{\theta_0})_i\right| \left| \frac{\partial^2 f_k(\boldsymbol{x};\boldsymbol{\theta''_k})}{\partial \theta_i \partial \theta_l} \right| + \left| \frac{b(\boldsymbol{x})o(1)}{n\rho(\boldsymbol{x})} \right| \left| \frac{\partial f_k(\boldsymbol{x};\boldsymbol{\theta'_k})}{\partial \theta_l} \right| \right\},$$

which, recalling the definition of $C_{n,\beta}$ and the bound functions in (14), is further bounded by

$$\sum_{i=1}^{m} \left\{ \sum_{\boldsymbol{x}\in D_n} \frac{\beta}{n^{3/2}} \frac{\lambda(\boldsymbol{x};\boldsymbol{\theta_0})}{\rho(\boldsymbol{x})} d_{kil}(\boldsymbol{x}) \right\} + \sum_{\boldsymbol{x}\in D_n} \frac{1}{n}\frac{b(\boldsymbol{x})}{\rho(\boldsymbol{x})} |o(1)| d_{kl}(\boldsymbol{x}) \tag{16}$$

for large enough n. Note that this bound depends only on n and not on $\boldsymbol{\theta'_k}$. Furthermore, by condition (C11), one can apply the Campbell-Mecke theorem and Markov's inequality as in the proof of Theorem 2 to obtain that the inner sum in the first term of (16) converges under $P_{\boldsymbol{\theta_0}}$ in probability to zero. Since this result applies to every component of the inner sum with $1 \leq i \leq m$ and also to the second term, (16) converges under $P_{\boldsymbol{\theta_0}}$ in probability to zero. Hence,

$$\sup_{\boldsymbol{\theta'_k} \in C_{n,\beta}} \left| \frac{1}{n} \sum_{\boldsymbol{x} \in D_n} \frac{\lambda(\boldsymbol{x}; \boldsymbol{\theta'_k})}{\rho(\boldsymbol{x})} \frac{\partial f_k(\boldsymbol{x}; \boldsymbol{\theta'_k})}{\partial \theta_l} - \frac{1}{n} \sum_{x \in D_n} \frac{\lambda(\boldsymbol{x}; \boldsymbol{\theta_0})}{\rho(\boldsymbol{x})} \frac{\partial f_k(\boldsymbol{x}; \boldsymbol{\theta_0})}{\partial \theta_l} \right| = o_P(1).$$

Applying similar proofs to the other two lines in (15), recalling condition (C5), and combining the obtained results, by the triangle inequality, one obtains that

$$\sup_{\boldsymbol{\theta'_k} \in C_{n,\beta}} \left| \left[-\frac{1}{n} \nabla \boldsymbol{s}_{n,k}(\boldsymbol{\theta'_k}) \right]_l - \left[-\frac{1}{n} \nabla \boldsymbol{s}_n(\boldsymbol{\theta_0}) \right]_{k,l} \right| = o_P(1). \tag{17}$$

Finally, consider the analogue of Lemma 2 for the unbiased estimating equation (13). Under conditions (C1), (C9) and (C11), $-\nabla \boldsymbol{s}_n(\boldsymbol{\theta_0})/n$ reads

$$\frac{1}{n} \sum_{i=1}^{n} \left[\sum_{\boldsymbol{x} \in Y_i} -\frac{\partial f_k(\boldsymbol{x}; \boldsymbol{\theta_0})}{\partial \theta_l} + \sum_{\boldsymbol{x} \in E_i} \frac{\lambda(\boldsymbol{x}; \boldsymbol{\theta_0})}{\rho(\boldsymbol{x})} \left(\frac{\partial f_k(\boldsymbol{x}; \boldsymbol{\theta_0})}{\partial \theta_l} + f_k(\boldsymbol{x}; \boldsymbol{\theta_0}) z_l(\boldsymbol{x}) \right) \right]_{k,l=1}^{m}.$$

Again, by conditions (C5) and (C11), one can apply the Campbell–Mecke theorem and Kolmogorov's strong law of large numbers to obtain that $-\nabla \boldsymbol{s}_n(\boldsymbol{\theta_0})/n$ converges $P_{\boldsymbol{\theta_0}}$-almost surely, and thus in $P_{\boldsymbol{\theta_0}}$-probability, to $\widetilde{\boldsymbol{U}}$. Accordingly,

$$\left| \left[-\frac{1}{n} \nabla \boldsymbol{s}_n(\boldsymbol{\theta_0}) \right]_{k,l} - \widetilde{\boldsymbol{U}}_{k,l} \right| = o_P(1). \tag{18}$$

The proof is completed by applying the triangle inequality to (17) and (18).

The next theorem is concerned with the extended central limit theorem.

Theorem 9 *Assume that the conditions (C1), (C4)–(C5) and (C9)–(C14) hold. Define* $\boldsymbol{s}_n(\boldsymbol{\theta}) := \boldsymbol{s}(X_n, D_n; \boldsymbol{\theta})$ *by (13) with* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. *Then, as* $n \to \infty$, *for every* n, *an estimator* $\widehat{\boldsymbol{\theta}}_n$ *exists that solves* $\boldsymbol{s}_n(\boldsymbol{\theta}) = \boldsymbol{0}$ *with a probability tending to one. Moreover,* $\widehat{\boldsymbol{\theta}}_n$ *converges under* $P_{\boldsymbol{\theta_0}}$ *in probability to* $\boldsymbol{\theta_0}$ *and* $n^{1/2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0})$ *converges under* $P_{\boldsymbol{\theta_0}}$ *in distribution to an* m*-dimensional normally distributed random vector with mean zero and covariance matrix* $\widetilde{\boldsymbol{U}}^{-1}\widetilde{\boldsymbol{V}}(\widetilde{\boldsymbol{U}}^{-1})^{\top}$, *where* $\widetilde{\boldsymbol{U}}$ *is as defined in Lemma 4 and* $\widetilde{\boldsymbol{V}}$ *is given by*

$$\begin{aligned}
\widetilde{\boldsymbol{V}} := & \left[\int_W f_k(\boldsymbol{u}; \boldsymbol{\theta_0}) f_l(\boldsymbol{u}; \boldsymbol{\theta_0}) \lambda(\boldsymbol{u}; \boldsymbol{\theta_0}) \left(1 + \frac{\lambda(\boldsymbol{u}; \boldsymbol{\theta_0})}{\rho(\boldsymbol{u})} \right) \mathrm{d}\boldsymbol{u} \right]_{k,l=1}^{m} \\
& + \left[\int\int_{W \times W} f_k(\boldsymbol{u}; \boldsymbol{\theta_0}) \lambda(\boldsymbol{u}; \boldsymbol{\theta_0}) f_l(\boldsymbol{v}; \boldsymbol{\theta_0}) \lambda(\boldsymbol{v}; \boldsymbol{\theta_0}) (g(\boldsymbol{u}, \boldsymbol{v}) - 1) \mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v} \right]_{k,l=1}^{m} \\
& + \left[\int\int_{W \times W} f_k(\boldsymbol{u}; \boldsymbol{\theta_0}) \lambda(\boldsymbol{u}; \boldsymbol{\theta_0}) f_l(\boldsymbol{v}; \boldsymbol{\theta_0}) \lambda(\boldsymbol{v}; \boldsymbol{\theta_0}) (h(\boldsymbol{u}, \boldsymbol{v}) - 1) \mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v} \right]_{k,l=1}^{m}.
\end{aligned}$$

Proof First of all, by condition (C11), $\boldsymbol{s}_n(\boldsymbol{\theta})$ is twice continuously differentiable with respect to $\boldsymbol{\theta}$. Then we verify the analogues of Lemmas 2 and 3 for the unbiased estimating equation (13).

The analogue of Lemma 2 has been established in Lemma 4. For Lemma 3, under conditions (C1) and (C9), $\boldsymbol{s}_n(\boldsymbol{\theta})/n^{1/2}$ now reads

$$\frac{\boldsymbol{s}_n(\boldsymbol{\theta_0})}{n^{1/2}} = n^{1/2}\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{\boldsymbol{x}\in Y_i}\boldsymbol{f}(\boldsymbol{x};\boldsymbol{\theta_0}) - \frac{1}{n}\sum_{i=1}^{n}\sum_{\boldsymbol{x}\in E_i}\boldsymbol{f}(\boldsymbol{x};\boldsymbol{\theta_0})\frac{\lambda(\boldsymbol{x};\boldsymbol{\theta_0})}{\rho(\boldsymbol{x})}\right\}.$$

The two terms between the curly bracket above are both averages of independent and identically distributed real-valued random vectors. By conditions (C9)–(C11), their means are finite and identical. Under the additional conditions (C4) and (C14), their variances are also finite and given by the entries of $\widetilde{\boldsymbol{V}}$. Then, similar to the proof of Lemma 3, one can apply the Campbell–Mecke theorem and multi-variate Lindeberg–Lévy central limit theorem to obtain that $\boldsymbol{s}_n(\boldsymbol{\theta_0})/n^{1/2}$ converges under $P_{\boldsymbol{\theta_0}}$ in distribution to an m-dimensional normally distributed random vector with mean zero and covariance matrix $\widetilde{\boldsymbol{V}}$.

From Sørensen (1999, Condition 2.7, Corollary 2.8 and Theorem 2.9), by condition (C13), we thus only need to prove that, for all $\beta > 0$,

$$\sup_{\boldsymbol{\theta}\in C_{n,\beta}}\left\|-\frac{1}{n}\boldsymbol{s}_n(\boldsymbol{\theta})\right\|, \quad \sup_{\boldsymbol{\theta'_k}\in C_{n,\beta}}\left\|-\frac{1}{n}\nabla\boldsymbol{s}_n(\boldsymbol{\theta'_1},\ldots,\boldsymbol{\theta'_m}) - \widetilde{\boldsymbol{U}}\right\|,$$

and

$$\sup_{\boldsymbol{\theta'_k}\in C_{n,\beta}}\left\|-\frac{1}{n}\nabla^2\boldsymbol{s}_{n,i}(\boldsymbol{\theta'_1},\ldots,\boldsymbol{\theta'_m}) - \widetilde{\boldsymbol{Q}}_i\right\|$$

all converge under $P_{\boldsymbol{\theta_0}}$ in probability to zero, where $\nabla^2\boldsymbol{s}_{n,i}(\boldsymbol{\theta})$ is the $m\times m$ matrix containing the second-order partial derivatives of $\boldsymbol{s}_{n,i}$ and $\boldsymbol{s}_{n,i}(\boldsymbol{\theta'_1},\ldots,\boldsymbol{\theta'_m})$ is obtained upon replacing $\boldsymbol{\theta}$ by $\boldsymbol{\theta'_k}$ throughout the k-th row. In order words, $\boldsymbol{s}_{n,i}(\boldsymbol{\theta'_1},\ldots,\boldsymbol{\theta'_m})_{k,l} \coloneqq \frac{\partial^2}{\partial\theta_k\partial\theta_l}\boldsymbol{s}_{n,i}(\boldsymbol{\theta'_k})$. Furthermore, $\widetilde{\boldsymbol{Q}}_i$ is given by

$$\begin{aligned}\widetilde{\boldsymbol{Q}}_i \coloneqq &\left[\int_W f_i(\boldsymbol{u},\boldsymbol{\theta_0})\lambda(\boldsymbol{u},\boldsymbol{\theta_0})z_k(\boldsymbol{u})z_l(\boldsymbol{u})\mathrm{d}\boldsymbol{u}\right]_{k,l=1}^{m} \\ &+\left[\int_W \left(\frac{\partial f_i(\boldsymbol{u},\boldsymbol{\theta_0})}{\partial\theta_k}z_l(\boldsymbol{u}) + \frac{\partial f_i(\boldsymbol{u},\boldsymbol{\theta_0})}{\partial\theta_l}z_k(\boldsymbol{u})\right)\lambda(\boldsymbol{u},\boldsymbol{\theta_0})d\boldsymbol{u}\right]_{k,l=1}^{m}.\end{aligned}$$

In Lemma 4, we already proved that

$$\sup_{\boldsymbol{\theta'_k}\in C_{n,\beta}}\left\|-\frac{1}{n}\nabla\boldsymbol{s}_n(\boldsymbol{\theta'_1},\ldots,\boldsymbol{\theta'_m}) - \widetilde{\boldsymbol{U}}\right\| \stackrel{P_{\boldsymbol{\theta_0}}}{\to} \boldsymbol{0}.$$

The remainder of the proof mostly proceeds along similar lines for the other two cases.

Condition (C12) is required in proving the convergence of some terms in the entries of $-\nabla^2 \boldsymbol{s}_{n,i}(\boldsymbol{\theta'_1}, \ldots, \boldsymbol{\theta'_m})/n$. For instance, we need to prove that

$$\sup_{\boldsymbol{\theta'_k} \in C_{n,\beta}} \left| \frac{1}{n} \sum_{\boldsymbol{x} \in D_n} \frac{\lambda(\boldsymbol{x}; \boldsymbol{\theta'_k})}{\rho(\boldsymbol{x})} \frac{\partial^2 f_i(\boldsymbol{x}; \boldsymbol{\theta'_k})}{\partial \theta_k \partial \theta_l} - \frac{1}{n} \sum_{\boldsymbol{x} \in D_n} \frac{\lambda(\boldsymbol{x}; \boldsymbol{\theta_0})}{\rho(\boldsymbol{x})} \frac{\partial^2 f_i(\boldsymbol{x}; \boldsymbol{\theta_0})}{\partial \theta_k \partial \theta_l} \right| = o_P(1).$$

Since $f_i(\boldsymbol{u}, \boldsymbol{\theta})$ is only twice continuously differentiable with respect to $\boldsymbol{\theta}$, the proof in Lemma 4 does not apply here. However, recalling from the proof of Lemma 4 that $\lambda(\boldsymbol{x}; \boldsymbol{\theta'}) = \lambda(\boldsymbol{x}; \boldsymbol{\theta_0}) + b(\boldsymbol{x}) o(1)$, the absolute value term above is bounded by

$$\frac{1}{n} \sum_{\boldsymbol{x} \in D_n} \left\{ \frac{\lambda(\boldsymbol{x}; \boldsymbol{\theta_0})}{\rho(\boldsymbol{x})} \left| \frac{\partial^2 f_i(\boldsymbol{x}; \boldsymbol{\theta'_k})}{\partial \theta_k \partial \theta_l} - \frac{\partial^2 f_i(\boldsymbol{x}; \boldsymbol{\theta_0})}{\partial \theta_k \partial \theta_l} \right| + \left| \frac{b(\boldsymbol{x}) o(1)}{\rho(\boldsymbol{x})} \right| \left| \frac{\partial^2 f_i(\boldsymbol{x}; \boldsymbol{\theta'_k})}{\partial \theta_k \partial \theta_l} \right| \right\},$$

which, under the adapted Hölder condition in condition (C12) and using condition (C11), is further bounded by

$$\sum_{\boldsymbol{x} \in D_n} \left\{ \frac{\beta^\alpha}{n^{\alpha/2+1}} \frac{\lambda(\boldsymbol{x}; \boldsymbol{\theta_0})}{\rho(\boldsymbol{x})} e_{ikl}(\boldsymbol{x}) + \frac{1}{n} \frac{b(\boldsymbol{x})}{\rho(\boldsymbol{x})} \left| o(1) \right| d_{ikl}(\boldsymbol{x}) \right\}$$

for large enough n. This bound depends only on n and not on $\boldsymbol{\theta'}$. By conditions (C11) and (C12), one can again apply the Campbell-Mecke theorem and Markov's inequality as in the proof of Theorem 2 to obtain that the bound converges under $P_{\boldsymbol{\theta_0}}$ in probability to zero.

Remark 5 Looking back on logistic regression estimators given by (3), its component functions are smooth with respect to $\boldsymbol{\theta}$ and the third-order partial derivatives are bounded. Thus, the second-order derivatives satisfy the adapted Hölder condition. Moreover, its component functions and the first- and second-order derivatives are bounded under condition (C5).

5 The asymptotics in practice

In this section, we apply our asymptotic results to establish confidence intervals for intensity estimation in practice. First, we conduct two simulation studies: the first investigates the use of regular dummy point processes with pair correlation functions smaller than one in reducing the variance of logistic regression estimators, as noted in Remark 2. The second demonstrates the utility of the extended central limit theorem in Theorem 9 through an example based on a pseudo-likelihood estimating equation. Finally, we present a case study estimating kitchen fire risk in an eastern region of the Netherlands.

5.1 Simulation on regular dummy point processes

We consider two examples of spatial point process models for $X_n = \cup_{i=1}^n Y_i$. In the first example, Y_i is an inhomogeneous Poisson point process with intensity function $\lambda(\boldsymbol{u}) = \exp[\theta_1 + \theta_2 u_x + \theta_3 u_y]$. Note that u_x and u_y are spatial coordinates at location $\boldsymbol{u}$. In the second example, Y_i is a log-Gaussian Cox point process driven by a Gaussian random field with mean function $m(\boldsymbol{u}) = \theta_5 u_x + \theta_6 u_y$ and covariance function $c(\boldsymbol{u}, \boldsymbol{v}) = \sigma^2 \exp[-\|\boldsymbol{u} - \boldsymbol{v}\|/\gamma]$. Thus, its intensity function reads $\lambda(\boldsymbol{u}) = \exp[\theta_5 u_x + \theta_6 u_y + \sigma^2/2]$. For elegance, we write $\theta_4 = \sigma^2/2$.

We also consider two options for dummy point processes $D_n = \cup_{i=1}^n E_i$. In the first option, E_i is a homogeneous Poisson point process with intensity function $\rho(\boldsymbol{u}) = \rho_1$ and pair correlation function $g(\boldsymbol{u}, \boldsymbol{v}) \equiv 1$. In the second option, E_i is a Gaussian determinantal point process also with intensity function $\rho(\boldsymbol{u}) = \rho_1$ but with pair correlation function $g(\boldsymbol{u}, \boldsymbol{v}) < 1$. Note that determinantal point processes have also been used for similar efficiency improvement purposes in Monte Carlo integration (Gautier, Bardenet and Valko 2019). Moreover, their realizations are easy to obtain (Lavancier, Møller and Rubak 2015), and the pair correlation functions of Gaussian determinantal point processes are known explicitly, $g(\boldsymbol{u}, \boldsymbol{v}) = 1 - \exp[-2(\|\boldsymbol{u} - \boldsymbol{v}\|/\eta)^2]$.

Specifically, we set $\theta_1 = 0$, $\theta_2 = 1$, $\theta_3 = 2$, $\theta_4 = 0.5$, $\theta_5 = 1$, $\theta_6 = 2$, $\gamma = 0.1$, $\rho_1 = 30$ and $\eta = 0.1$ We conduct two experiments, one in which Y_i is Poisson and one in which Y_i is a log-Gaussian Cox point process. In each experiment, we generate n replications $Y_1, \dots Y_n$ and form their union X_n. We also generate n replications $E_1, \dots, E_n$ and take the union D_n from either of the two dummy point process options. For each experiment, we carry out logistic regression estimation and report the parameter estimates as well as their approximate variances and 95% confidence intervals using the consistent estimators proposed in Section 3.4. The results for $n = 200$, $n = 600$ and $n = 1000$ are shown in Tables 1 and 2. We observe that for both point process models, the estimators using the regular dummy point process option achieve a smaller variance than those using the Poisson point process option. Additionally, as expected, such difference in variance decreases as n increases (cf., Theorems 2 and 6).

5.2 Simulation on a pseudo-likelihood type estimating equation

We consider the following estimating equation based on pseudo-likelihood, as proposed by Besag (1977) and discussed in Rathbun, Shiffman and Gwaltney (2007) and Waagepetersen (2008),

$$\boldsymbol{s}(X, D; \boldsymbol{\theta}) := \sum_{\boldsymbol{x} \in X} \boldsymbol{z}(\boldsymbol{x}) - \sum_{\boldsymbol{x} \in D} \boldsymbol{z}(\boldsymbol{x}) \frac{\lambda(\boldsymbol{x}; \boldsymbol{\theta})}{\rho(\boldsymbol{x})} = \mathbf{0}.$$

Table 1: Simulation results for the Poisson point process model using Poisson or regular dummy point process options in logistic regression estimation.

n	**Dummy**	**Parameter**	**True**	**Estimate**	**Variance**	95% **CI**
200	Poisson	θ_1	0	0.042	0.012	[-0.175, 0.259]
		θ_2	1	0.965	0.015	[0.727, 1.203]
		θ_3	2	1.897	0.016	[1.650, 2.144]
	Regular	θ_1	0	0.054	0.011	[-0.155, 0.263]
		θ_2	1	0.968	0.014	[0.736, 1.200]
		θ_3	2	1.887	0.015	[1.647, 2.127]
600	Poisson	θ_1	0	-0.015	0.0040	[-0.140, 0.110]
		θ_2	1	1.001	0.0047	[0.867, 1.135]
		θ_3	2	2.045	0.0052	[1.903, 2.187]
	Regular	θ_1	0	0.015	0.0039	[-0.108, 0.138]
		θ_2	1	0.949	0.0044	[0.819, 1.079]
		θ_3	2	2.039	0.0049	[1.901, 2.177]
1000	Poisson	θ_1	0	-0.003	0.0025	[-0.101, 0.094]
		θ_2	1	0.962	0.0028	[0.858, 1.066]
		θ_3	2	2.030	0.0032	[1.919, 2.141]
	Regular	θ_1	0	-0.007	0.0023	[-0.102, 0.088]
		θ_2	1	1.009	0.0026	[0.908, 1.110]
		θ_3	2	2.013	0.0030	[1.906, 2.120]

Table 2: Simulation results for the log-Gaussian Cox process model using Poisson or regular dummy point process options in logistic regression estimation.

n	**Dummy**	**Parameter**	**True**	**Estimate**	**Variance**	95% **CI**
200	Poisson	θ_4	0.5	0.440	0.012	[0.225, 0.655]
		θ_5	1	1.112	0.012	[0.897, 1.328]
		θ_6	2	2.020	0.014	[1.788, 2.252]
	Regular	θ_4	0.5	0.472	0.011	[0.266, 0.678]
		θ_5	1	1.125	0.011	[0.919, 1.331]
		θ_6	2	1.942	0.012	[1.719, 2.165]
600	Poisson	θ_4	0.5	0.493	0.0037	[0.374, 0.612]
		θ_5	1	1.044	0.0041	[0.918, 1.170]
		θ_6	2	1.987	0.0043	[1.858, 2.116]
	Regular	θ_4	0.5	0.491	0.0037	[0.372, 0.610]
		θ_5	1	1.080	0.0038	[0.959, 1.201]
		θ_6	2	1.954	0.0040	[1.829, 2.079]
1000	Poisson	θ_4	0.5	0.494	0.0021	[0.403, 0.585]
		θ_5	1	1.013	0.0026	[0.914, 1.112]
		θ_6	2	1.976	0.0028	[1.873, 2.079]
	Regular	θ_4	0.5	0.455	0.0021	[0.364, 0.546]
		θ_5	1	1.041	0.0024	[0.944, 1.138]
		θ_6	2	2.008	0.0026	[1.908, 2.108]

By Lemma 4 and Theorem 9, the components of the asymptotic covariance matrix are given by

$$\widetilde{\boldsymbol{U}} = \left[\int_W \lambda(\boldsymbol{u};\boldsymbol{\theta_0}) z_k(\boldsymbol{u}) z_l(\boldsymbol{u}) \mathrm{d}\boldsymbol{u} \right]_{k,l=1}^{m}$$

and

$$\widetilde{\boldsymbol{V}} = \left[\int_W \lambda(\boldsymbol{u};\boldsymbol{\theta_0})\left(1+\frac{\lambda(\boldsymbol{u};\boldsymbol{\theta_0})}{\rho(\boldsymbol{u})}\right) z_k(\boldsymbol{u})z_l(\boldsymbol{u})\mathrm{d}\boldsymbol{u}\right]_{k,l=1}^m + \left[\int\int_{W\times W} \lambda(\boldsymbol{u};\boldsymbol{\theta_0})\lambda(\boldsymbol{v};\boldsymbol{\theta_0}) z_k(\boldsymbol{u})z_l(\boldsymbol{v})(g(\boldsymbol{u},\boldsymbol{v})+h(\boldsymbol{u},\boldsymbol{v})-2)\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v}\right]_{k,l=1}^m .$$

To estimate the asymptotic covariance matrix (cf., Theorem 9), we consider three sequences of independent and identically distributed point processes, in analogy to Section 4.1. Specifically, we define $A_{1,n} := \cup_{i=1}^n B_{1,i}$, where $B_{1,i}$ is a Poisson point process with the same intensity function $\rho(\boldsymbol{u})$ as E_i. Also set $A_{2,n} := \cup_{i=1}^n B_{2,i}$ where $B_{2,i}$ shares the intensity function $\rho(\boldsymbol{u})$ and pair correlation function $g(\boldsymbol{u},\boldsymbol{v})$ of E_i. Moreover, set $A_{3,n} := \cup_{i=1}^n B_{3,i}$ where $B_{3,i}$ is an independent replication from the same distribution as Y_i. Then, we approximate

$$n\widehat{\widetilde{\boldsymbol{U}}} := \sum_{i=1}^n \left[\sum_{\boldsymbol{x}\in B_{1,i}} \frac{\lambda(\boldsymbol{x};\widehat{\boldsymbol{\theta}}_n)}{\rho(\boldsymbol{x})} z_k(\boldsymbol{x})z_l(\boldsymbol{x})\right]_{k,l=1}^m$$

and

$$\begin{aligned} n\widehat{\widetilde{\boldsymbol{V}}} := &\sum_{i=1}^n \left[\sum_{\boldsymbol{x}\in B_{1,i}} \frac{\lambda(\boldsymbol{x};\widehat{\boldsymbol{\theta}}_n)}{\rho(\boldsymbol{x})}\left(1+\frac{\lambda(\boldsymbol{x};\widehat{\boldsymbol{\theta}}_n)}{\rho(\boldsymbol{x})}\right) z_k(\boldsymbol{x})z_l(\boldsymbol{x})\right]_{k,l=1}^m \\ &+ \sum_{i=1}^n \left[\sum_{\boldsymbol{x},\boldsymbol{y}\in B_{2,i}}^{\neq} \frac{\lambda(\boldsymbol{x};\widehat{\boldsymbol{\theta}}_n)\lambda(\boldsymbol{y};\widehat{\boldsymbol{\theta}}_n)}{\rho(\boldsymbol{x})\rho(\boldsymbol{y})} z_k(\boldsymbol{x})z_l(\boldsymbol{y})\right]_{k,l=1}^m \\ &+ \sum_{i=1}^n \left[\sum_{\boldsymbol{x},\boldsymbol{y}\in B_{3,i}}^{\neq} z_k(\boldsymbol{x})z_l(\boldsymbol{y})\right]_{k,l=1}^m \\ &- 2\sum_{i=1}^n \left[\sum_{\boldsymbol{x},\boldsymbol{y}\in B_{1,i}}^{\neq} \frac{\lambda(\boldsymbol{x};\widehat{\boldsymbol{\theta}}_n)\lambda(\boldsymbol{y};\widehat{\boldsymbol{\theta}}_n)}{\rho(\boldsymbol{x})\rho(\boldsymbol{y})} z_k(\boldsymbol{x})z_l(\boldsymbol{y})\right]_{k,l=1}^m . \end{aligned}$$

In the experiments, we consider the spatial point process models as in Section 5.1, along with the same two dummy point process options. We then conduct the simulation using the pseudo-likelihood type estimating equation defined above and report the parameter estimates, approximate variances and 95% confidence intervals. The results for n = 200, n = 600 and n = 1000 are presented in Tables 3 and 4. We observe that for both point process models, the estimators based on pseudo-likelihood exhibit larger variance than those from logistic regression estimation. Additionally, the regular dummy point process option consistently yields a smaller variance than the Poisson point process option. However, both reductions in variance decreases as n increases.

Table 3: Simulation results for the Poisson point process model using Poisson or regular dummy point process options in pseudo-likelihood type estimation.

n	**Dummy**	**Parameter**	**True**	**Estimate**	**Variance**	95% **CI**
200	Poisson	θ_1	0	0.029	0.012	[-0.190, 0.248]
		θ_2	1	0.975	0.015	[0.737, 1.213]
		θ_3	2	1.909	0.016	[1.661, 2.157]
	Regular	θ_1	0	0.032	0.011	[-0.018, 0.242]
		θ_2	1	0.979	0.014	[0.747, 1.211]
		θ_3	2	1.914	0.015	[1.673, 2.155]
600	Poisson	θ_1	0	-0.015	0.0041	[-0.140, 0.110]
		θ_2	1	1.002	0.0047	[0.867, 1.137]
		θ_3	2	2.044	0.0053	[1.902, 2.186]
	Regular	θ_1	0	0.020	0.0040	[-0.103, 0.143]
		θ_2	1	0.947	0.0044	[0.817, 1.077]
		θ_3	2	2.032	0.0050	[1.894, 2.170]
1000	Poisson	θ_1	0	-0.007	0.0025	[-0.105, 0.091]
		θ_2	1	0.960	0.0029	[0.855, 1.065]
		θ_3	2	2.038	0.0032	[1.927, 2.149]
	Regular	θ_1	0	-0.012	0.0023	[-0.107, 0.083]
		θ_2	1	1.006	0.0027	[0.905, 1.107]
		θ_3	2	2.024	0.0030	[1.917, 2.131]

Table 4: Simulation results for the log-Gaussian Cox process model using Poisson or regular dummy point process options in pseudo-likelihood type estimation.

n	**Dummy**	**Parameter**	**True**	**Estimate**	**Variance**	95% **CI**
200	Poisson	θ_4	0.5	0.435	0.012	[0.220, 0.650]
		θ_5	1	1.115	0.012	[0.900, 1.330]
		θ_6	2	2.025	0.014	[1.793, 2.257]
	Regular	θ_4	0.5	0.472	0.012	[0.257, 0.687]
		θ_5	1	1.132	0.010	[0.936, 1.328]
		θ_6	2	1.935	0.012	[1.720, 2.150]
600	Poisson	θ_4	0.5	0.476	0.0040	[0.353, 0.599]
		θ_5	1	1.057	0.0043	[0.939, 1.195]
		θ_6	2	2.004	0.0045	[1.873, 2.013]
	Regular	θ_4	0.5	0.484	0.0039	[0.362, 0.606]
		θ_5	1	1.083	0.0039	[0.960, 1.206]
		θ_6	2	1.961	0.0042	[1.834, 2.088]
1000	Poisson	θ_4	0.5	0.493	0.0022	[0.401, 0.585]
		θ_5	1	1.016	0.0027	[0.915, 1.117]
		θ_6	2	1.975	0.0028	[1.871, 2.079]
	Regular	θ_4	0.5	0.446	0.0022	[0.354, 0.538]
		θ_5	1	1.048	0.0025	[0.950, 1.146]
		θ_6	2	2.017	0.0026	[1.917, 2.117]

5.3 Application to kitchen fire data

To demonstrate the utility of our asymptotic results in practical applications, we present a case study on fire risk prediction. The data set consists of 403 kitchen fire incidents occurring between 2011 and 2020 in Twente, an eastern

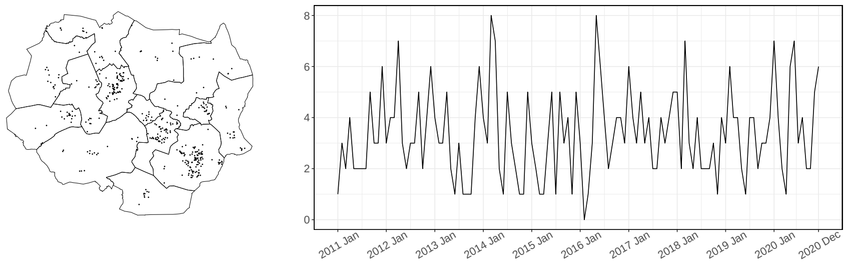

Fig. 1: Spatial and temporal projections of kitchen fire incidents in Twente.

region of the Netherlands. Spatial and temporal projections of point patterns of fires are plotted in Figure 1. It is obvious that kitchen fires exhibit clear spatial heterogeneity, with more incidents in urban areas. For example, the cities Enschede, Hengelo and Almelo can be clearly identified along the diagonal). The temporal patterns are more diffuse. According to firefighters, the risk of kitchen fires is closely linked to the density of buildings in a given area. However, specific building information for individual fire incidents is not recorded. Additionally, the population size in a neighbourhood also influences the risk of kitchen fires, as kitchen fires often result from improper use of kitchen facilities in densely populated areas. To estimate the fire risk, the following intensity function model is proposed:

$$\lambda(\boldsymbol{u};\boldsymbol{\theta}) = B(\boldsymbol{u})\exp[\theta_1 + \theta_2 P(\boldsymbol{u})],$$

where $B(\boldsymbol{u})$ denotes the general density of buildings at a location $\boldsymbol{u}$ and $P(\boldsymbol{u})$ denote the population size. Note that $B(\boldsymbol{u})$ is available in precise locations, while $P(\boldsymbol{u})$ is counted by Statistics Netherlands in the grid of predefined area boxes. To enable point process modelling, we assume $P(\boldsymbol{u})$ to be constant in an area box. The plots for these two variables are displayed in Figure 2.

To align with the asymptotic regime defined in Section 2.4, we treat the kitchen fires occurring on individual days as independent and identically distributed spatial point patterns given the absence of temporal dependence. Spatial dependence may exist; for instance, neighboring households might install kitchens with facilities of similar quality. Over a ten-year period, the kitchen fires lead to 3,650 independent and identically distributed point processes, excluding leap days. We fit the intensity function above using the logistic regression estimator with stationary dummy point processes. To further demonstrate the utility of regular dummy point processes in variance reduction, as discussed in Remark 2 and showed in Section 5.1, we again conduct two experiments for parameter estimation, one using a Poisson point process as the dummy point process and the other using a determinantal point process. We approximate the asymptotic covariance matrix using the consistent estimator proposed in Section 3.4. The parameter estimates, computed vari-

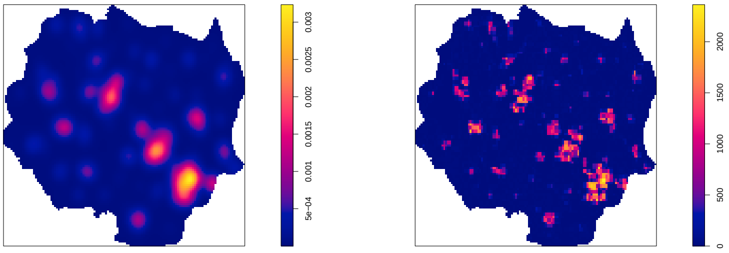

Fig. 2: Influencing variables for kitchen fire occurrences suggested by firefighters: building density $B(\boldsymbol{u})$ (left) and population size $P(\boldsymbol{u})$ (right).

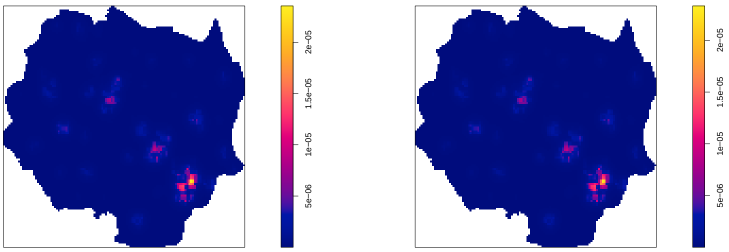

Fig. 3: Estimated kitchen fire risk for an individual day using Poisson (left) and determinantal (right) dummy point processes in logistic regression estimation.

Table 5: Estimation results for the kitchen fire model using Poisson or regular dummy point process options in logistic regression estimation.

Dummy	**Parameter**	**Estimate**	**Variance**	95% **CI**
Poisson	θ_1	-1.568e1	7.353e-3	[-1.585e1, -1.552e1]
	θ_2	1.090e-3	5.383e-9	[9.464e-4, 1.234e-3]
Regular	θ_1	-1.568e1	7.315e-3	[-1.585e1, -1.551e1]
	θ_2	1.084e-3	5.378e-9	[9.400e-4, 1.228e-3]

ances and 95% confidence intervals of θ_1, θ_2 are summarized in Table 5. Note that, since $\hat{\theta}_2$ is positive, regions with higher population size are prone to more kitchen fires. In addition, we visualize the estimated kitchen fire risk for both experiments in Figure 3. The resulting plots are similar, as both methods produce unbiased parameter estimates. However, using determinantal dummy point processes results in a smaller estimation variance compared to Poisson point processes, highlighting their efficiency.

6 Conclusion

In this paper, under conditions that are easy to verify in practice, we established strong consistency and asymptotic normality for logistic regression estimators for the parameters in the intensity function of a spatial point process model under an infill framework. Such a regime is quite natural for Poisson point processes observed in a fixed window, as is the case in our previous study on chimney fire prediction (Lu et al. 2023) which motivated this work. It is also suitable for general point processes when replications are available. We also extended our central limit theorem to estimators obtained from solving general unbiased estimating equations based on the Campbell–Mecke theorem.

From a practical perspective, our simulation studies illustrate how these results can be used to obtain approximate confidence intervals based on consistent estimators for the asymptotic covariance matrices. Logistic regression estimators are computationally fast, as they can be implemented using standard software. In addition, we showed the advantage of using regular dummy point processes to improve estimation precision. Furthermore, we also presented an example on a different pseudo-likelihood type estimating equation. Last but not least, as a real-world application, we demonstrate how our asymptotic results enable uncertainty quantification for parameter estimates of kitchen fire intensity in point pattern data.

For future work, note that additional higher order estimating equations are usually required to estimate interaction parameters for general point processes. Thus, the first direction would be to study infill asymptotics there. Moreover, we assumed that the intensity function is log-linear in covariates, which simplified the computations. We believe, however, that similar asymptotic results can be proved for a much wider class of intensity functions. Finally, it would also be interesting to study infill asymptotic results for Gibbs and Markov point processes under similar appropriate conditions on Papangelou conditional intensity functions.

Acknowledgements We thank Maurits de Graaf and Clément Lezane for valuable discussions and comments.

References

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

Baddeley, A., Rubak, E. and Turner, R. (2016). *Spatial point patterns: Methodology and applications with R.* Boca Raton: CRC Press.

Baddeley, A., Cœurjolly, J.-F., Rubak, E. and Waagepetersen, R. (2014). Logistic regression for spatial Gibbs point processes. *Biometrika*, 101, 377–392.

Berman, M. and Turner, R. (1992). Approximating point process likelihoods with GLIM. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 41, 31–38.

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

Bray, A. and Schoenberg, F. (2013). Assessment of point process models for earthquake forecasting. *Statistical Science*, 28, 510–520.

Chiu, S. N., Stoyan, D., Kendall, W. S. and Mecke, J. (2013). *Stochastic geometry and its applications.* Chichester: Wiley.

Choiruddin, A., Cœurjolly, J.-F. and Letué, F. (2018). Convex and non-convex regularization methods for spatial point processes intensity estimation. *Electronic Journal of Statistics*, 12, 1210–1255.

Choiruddin, A., Cœurjolly, J.-F. and Letué, F. (2023). Adaptive lasso and Dantzig selector for spatial point processes intensity estimation. *Bernoulli*, 29, 1849–1876.

Choiruddin, A., Cœurjolly, J.-F. and Waagepetersen, R. (2021). Information criteria for inhomogeneous spatial point processes. *Australian & New Zealand Journal of Statistics*, 63, 119–143.

Cœurjolly, J.-F. and Guan, Y. (2014). Covariance of empirical functionals for inhomogeneous spatial point processes when the intensity has a parametric form. *Journal of Statistical Planning and Inference*, 155, 79–92.

Cœurjolly, J.-F. and Møller, J. (2013). Variational approach for spatial point process intensity estimation. *Bernoulli*, 20, 1097–1125.

Coles, P. and Jones, B. (1991). A lognormal model for the cosmological mass distribution. *Monthly Notices of the Royal Astronomical Society*, 248, 1–13.

Daley, D. J. and Vere-Jones, D. (2008). *An introduction to the theory of point processes. Volume II: General Theory and Structure.* New York: Springer.

Dong, Z., Zhu S., Xie Y., Mateu J. and Rodríguez-Cortés F. J. (2023). Non-stationary spatio-temporal point process modeling for high-resolution COVID-19 data. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 7, 368–386.

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

Fiksel, T. (1988). Estimation of interaction potentials of Gibbsian point processes. *Statistics*, 19, 77–86.

Gautier, G., Bardenet, R. and Valko, M. (2019). On two ways to use determinantal point processes for Monte Carlo integration. In *Conference on Neural Information Processing Systems*, 32. Vancouver.

Georgii, H. (1995). *Gibbs measures and phase transitions.* Berlin: De Gruyter.

Guan, Y., Jalilian, A. and Waagepetersen, R. (2015). Quasi-likelihood for spatial point processes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 77, 677–697.

Guan, Y. and Loh, J. M. (2007). A thinned block bootstrap variance estimation procedure for inhomogeneous spatial point patterns. *Journal of the American Statistical Association*, 102, 1377–1386.

Guan, Y. and Shen, Y. (2010). A weighted estimating equation approach for inhomogeneous spatial point processes. *Biometrika*, 97, 867–880.

Guyon, X. (1995). *Random fields on a network: Modeling, statistics and applications.* New York: Springer.

Kutoyants, Y. A. (1998). *Statistical inference for spatial Poisson processes.* New York: Springer.

Lavancier, F., Møller, J. and Rubak, E. (2015). Determinantal point process models and statistical inference. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 77, 853–877.

Lieshout, M. N. M. van (2000). *Markov point processes and their applications.* London: Imperial College Press.

Lieshout, M. N. M. van (2006). Maximum likelihood estimation for random sequential adsorption. *Advances in Applied Probability (SGSA)*, 38, 889–898.

Lieshout, M. N. M. van (2021). Infill asymptotics for adaptive kernel estimators of spatial intensity. *Australian & New Zealand Journal of Statistics*, 63, 159–181.

Lu, C., Lieshout, M. N. M. van, Graaf, M. de and Visscher, P. (2023). Data-driven chimney fire risk prediction using machine learning and point process tools. *Annals of Applied Statistics*, 17, 3088–3111.

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

Møller, J. and Waagepetersen, R. (2004). *Statistical inference and simulation for spatial point processes.* Boca Raton: CRC Press.

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

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

Pawlas, Z. (2011). Estimation of summary characteristics from replicated spatial point processes. *Kybernetika*, 47, 880–892.

Rathbun, S. L. (2013). Optimal estimation of Poisson intensity with partially observed covariates. *Biometrika*, 100, 277–281.

Rathbun, S. L. and Cressie, N. (1994). Asymptotic properties of estimators for the parameters of spatial inhomogeneous Poisson point processes. *Advances in Applied Probability*, 26, 122–154.

Rathbun, S. L., Shiffman, S. and Gwaltney, C. J. (2007). Modelling the effects of partially observed covariates on Poisson process intensity. *Biometrika*, 94, 153–165.

Reitzner, M. and Schulte, M. (2013). Central limit theorems for U-statistics of Poisson point processes. *Annals of Probability*, 41, 3879–3909.

Ripley, B. D. (1988). *Statistical inference for spatial processes.* Cambridge: Cambridge University Press.

Schoenberg, F. (2005). Consistent parametric estimation of the intensity of a spatial-temporal point process. *Journal of Statistical Planning and Inference*, 128, 79–93.

Silvapulle, M. J. (1981). On the existence of maximum likelihood estimators for the binomial response models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 43, 310–313.

Sisson S. A. and Fan Y. (2011). Likelihood-free MCMC. In: *Handbook of Markov Chain Monte Carlo*, 12. New York: CRC Press.

Stoica, R. S., Philippe, A., Gregori, P. and Mateu, J. (2017). ABC shadow algorithm: A tool for statistical analysis of spatial patterns. *Statistics and Computing*, 27, 1225–1238.

Sørensen, M. (1999). On asymptotics of estimating functions. *Brazilian Journal of Probability and Statistics*, 13, 111–136.

Takacs, R. (1985). Estimator for the pair-potential of a Gibbsian point process. *Statistics*, 17, 429–433.

Thurman, A. L., Fu, R., Guan, Y. and Zhu, J. (2015). Regularized estimating equations for model selection of clustered spatial point processes. *Statistica Sinica*, 25, 173–188.

Vaart, A. W. van der (1998). *Asymptotic statistics.* Cambridge: Cambridge University Press.

Venables, W. N. and Ripley B. D. (2002). *Modern applied statistics with S.* New York: Springer.

Vihrs N., Møller J. and Gelfand A. E. (2022). Approximate Bayesian inference for a spatial point process model exhibiting regularity and random aggregation. *Scandinavian Journal of Statistics*, 49, 185–210.

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

Waagepetersen, R. (2008). Estimating functions for inhomogeneous spatial point processes with incomplete covariate data. *Biometrika*, 95, 351–363.