Jiří Dvořák; Michaela Prokešová

Asymptotic properties of the minimum contrast estimators for projections of inhomogeneous space-time shot-noise Cox processes

Applications of Mathematics, Vol. 61 (2016), No. 4, 387–411

Persistent URL: `http://dml.cz/dmlcz/145792`

61 (2016) APPLICATIONS OF MATHEMATICS No. 4, 387–411

ASYMPTOTIC PROPERTIES OF THE MINIMUM CONTRAST ESTIMATORS FOR PROJECTIONS OF INHOMOGENEOUS SPACE-TIME SHOT-NOISE COX PROCESSES

Jiří Dvořák, Michaela Prokešová, Praha

(Received September 20, 2015)

Abstract. We consider a flexible class of space-time point process models—inhomogeneous shot-noise Cox point processes. They are suitable for modelling clustering phenomena, e.g. in epidemiology, seismology, etc. The particular structure of the model enables the use of projections to the spatial and temporal domain. They are used to formulate a stepwise estimation method to estimate different parts of the model separately. In the first step, the Poisson likelihood approach is used to estimate the inhomogeneity parameters. In the second and third steps, the minimum contrast estimation based on K-functions of the projected processes is used to estimate the interaction parameters. We study the asymptotic properties of the resulting estimators and formulate a set of conditions sufficient for establishing consistency and asymptotic normality of the estimators under the increasing domain asymptotics.

Keywords: space-time point process; shot-noise Cox process; minimum contrast estimation; projection process; increasing domain asymptotics

MSC 2010: 60G55, 62F12

1. Introduction

Statistical inference for spatial and temporal point processes has seen a significant development over the past few decades and it is now a well-established field [18], [19], [15], [2]. It might seem tempting to simply extend the methods from the purely spatial setting to the space-time setting but a lot of caution should be taken in such a case. A space-time point process in $\mathbb{R}^d \times \mathbb{R}$ should not be considered just a point process in $\mathbb{R}^{d+1}$. The temporal coordinate plays a distinct role and hence devoted space-time methods should be used for statistical inference. Developing

The research has been supported by the Czech Science Foundation, project no. 16-03708S.

the methods for the space-time setting is an ongoing process and many interesting problems remain open.

In the present paper we focus on the problem of parameter estimation for the so-called shot-noise Cox processes (doubly stochastic point processes with a particular structure of the driving random field, see [13]). They constitute a flexible class of parametric models suitable for modelling clustered point patterns. As is the case for Cox point processes in general, obtaining the maximum likelihood estimates is often computationally too demanding for such models. The likelihood contains expectation of a complicated integral term with respect to the distribution of the driving random field. Hence, moment-based estimation methods are preferable in this case due to the lower computational demands for both the stationary and non-stationary processes.

For the space-time point processes, estimation of the moment properties—particularly of the second and higher orders—often turns out to be problematic due to the high variability of the estimates. This may in turn cause low stability of the derived moment-based estimators of the model parameters. In an attempt to remedy this problem, the paper [16] proposed a step-wise estimation method for inhomogeneous space-time shot-noise Cox processes with a particular model structure. It is based on the projections of the space-time process to the spatial and temporal domain. The idea of using the projection processes is due to [14]. The estimation method proposed in [16] proved to be applicable in realistic scenarios but asymptotic properties of the resulting estimators were not studied. In this paper, we fill in this gap—we provide a set of sufficient conditions under which the consistency and asymptotic normality of the estimators under the increasing domain asymptotics can be proved.

After presenting the necessary background material in Sections 2 and 3, the model parametrization is given in Section 4. Then, the estimation procedure is described in Section 5. The main results of the paper are formulated in Section 6 together with the discussion of the assumptions. The proofs are deferred to Appendices.

2. Background

We start with introducing the necessary notation and concepts relating to the space-time point processes. For more detailed information on space-time point processes see [3].

We denote by $\mathbb{R}$ the set of real numbers and by $\mathcal{B}(\mathbb{R})$ the corresponding Borel σ-algebra. For simplicity we restrict our attention to the space-time point processes in $\mathbb{R}^2 \times \mathbb{R}$, i.e. the spatial domain is $\mathbb{R}^2$ and the temporal domain is $\mathbb{R}$. Furthermore, $\mathcal{B}(\mathbb{R}^2)$ and $\mathcal{B}(\mathbb{R}^2 \times \mathbb{R})$ denote the Borel σ-algebras on the appropriate spaces. In the latter case we emphasize the role of the temporal coordinate by using the notation $\mathbb{R}^2 \times \mathbb{R}$.

Throughout this paper, we regard a simple space-time point process X as a random countable locally finite subset of $\mathbb{R}^2 \times \mathbb{R}$. A point $(u,t) \in X$ is thus interpreted as an event of the process which occurs at the location $u \in \mathbb{R}^2$ at the time $t \in \mathbb{R}$.

The volume of a Borel set B, i.e. its Lebesgue measure of appropriate dimension, is denoted by $|B|$, $\|x\|$ is the Euclidean norm of a vector x and I is the indicator function.

The kth-order factorial moment measure α_k of X is defined as

$$\alpha_k(A) = \mathbb{E}\Bigg(\sum^{\neq}_{(u_1,t_1),\dots,(u_k,t_k)\in X} I[((u_1,t_1),\dots,(u_k,t_k)) \in A] \Bigg), \quad A \in (\mathcal{B}(\mathbb{R}^2 \times \mathbb{R}))^{\otimes k},$$

where $\neq$ denotes that the summation is over k-tuples of distinct points of X.

If the density of the measure α_k w.r.t. the Lebesgue measure of dimension $3k$ exists for some $k \in \mathbb{N}$, we call it the kth-order intensity function of X and denote it by ϱ_k. We assume in the following that the first- and second-order intensity functions ϱ_1 and ϱ_2 of X exist. For simplicity, we use the notation $\varrho_1 = \varrho$ and we call ϱ the intensity function (or simply the intensity if it is constant).

Just as in [16], we consider a particular type of inhomogeneity of the process X, the second-order intensity reweighted stationarity (SOIRS), see [1]. It means that the inhomogeneous pair-correlation function of X defined as

$$g((u,t),(v,s)) = \frac{\varrho_2((u,t),(v,s))}{\varrho(u,t)\varrho(v,s)}$$

depends only on the differences $(v-u, s-t)$.

Consider a stationary process X_0 (its distribution is invariant w.r.t. translations of $\mathbb{R}^2 \times \mathbb{R}$). Now we form a thinned process X by randomly deleting or retaining each point, independently of the others. The retention probabilities are given by a function $f\colon \mathbb{R}^2 \times \mathbb{R} \to [0,1]$. We call f the inhomogeneity function. The thinned process X is an example of a SOIRS process.

If the kth-order intensity function $\varrho_{0,k}$ of the stationary process X_0 exists, so does the kth-order intensity function ϱ_k of the thinned process X and it has the form

$$\varrho_k((u_1,t_1),\dots,(u_k,t_k)) = \varrho_{0,k}((u_1,t_1),\dots,(u_k,t_k)) \prod_{i=1}^{k} f((u_i,t_i)),$$
$$(u_1,t_1),\dots,(u_k,t_k) \in \mathbb{R}^2 \times \mathbb{R}.$$

This in turn means that the pair-correlation functions of X and X_0 are the same.

Following [14], we define the space-time K-function of a SOIRS process X as

$$K(r,t)=\int_{\mathbb{R}}\int_{\mathbb{R}^2} I[\|s\|\leqslant r,|\tau|\leqslant t]g(s,\tau)\,\mathrm{d}s\,\mathrm{d}\tau,\quad r\geqslant 0,\ t\geqslant 0.$$

Here $\mathrm{d}\cdot$ denotes integration with respect to the Lebesgue measure over the appropriate space.

We consider the space-time observation window in the product form $W\times T$, $W\in\mathcal{B}(\mathbb{R}^2)$, $T\in\mathcal{B}(\mathbb{R})$. We assume that it has a finite positive Lebesgue measure.

The assumed existence of the second-order intensity function of X implies that for any pair of distinct points $(u,t)\neq(v,s)$ from X (with both $u,v\in W$ or both $t,s\in T$) we have $u\neq v$ and $t\neq s$ almost surely. Hence, we may define the spatial and temporal projection process as

$$\begin{aligned}X_s&=\{u\in\mathbb{R}^2\colon\ \exists\, t\in T \text{ such that } (u,t)\in X\},\\ X_t&=\{t\in\mathbb{R}:\exists\, u\in W \text{ such that } (u,t)\in X\}.\end{aligned}$$

The moment characteristics of the projection processes X_s, X_t are fully determined by the characteristics of the space-time process X. Let the kth-order intensity function ϱ_k of the space-time process X exist. Then the kth-order intensity function $\varrho_{s,k}$ of X_s exists and takes the form

$$\varrho_{s,k}(u_1,\ldots,u_k)=\int_T\ldots\int_T \varrho_k((u_1,t_1),\ldots,(u_k,t_k))\,\mathrm{d}t_1\ldots\mathrm{d}t_k,\ u_1,\ldots,u_k\in\mathbb{R}^2,$$

and similarly for $\varrho_{t,k}$ of X_t.

3. Space-time shot-noise Cox processes

A space-time shot-noise Cox process is a Cox process with the driving field Λ of the form

$$\Lambda(u,t)=\sum_{(r,v,s)\in\Phi} r\,k((u,t),(v,s)),\ (u,t)\in\mathbb{R}^2\times\mathbb{R}, \tag{3.1}$$

where Φ is a Poisson process on $(0,\infty)\times\mathbb{R}^2\times\mathbb{R}$ with intensity measure U, and k is a probability density function on $\mathbb{R}^2\times\mathbb{R}$. In the following, k will be called the smoothing kernel. Some basic integrability assumptions must be fulfilled in order to secure the existence of the shot-noise Cox process with given U and k, see [13].

The space-time process X driven by the random field (3.1) is stationary if the kernel k depends only on the difference of its arguments, $k((u,t),(v,s)) = k(v-u, s-t)$, and the intensity measure U has the product form

$$U(\mathrm{d}r,\, \mathrm{d}v,\, \mathrm{d}s) = \mu\, V(\mathrm{d}r)\, \mathrm{d}v\, \mathrm{d}s. \tag{3.2}$$

Here $\mu > 0$ and $V(\mathrm{d}r)$ is an arbitrary measure on $(0,\infty)$ satisfying the integrability condition $\int_0^\infty \min(1,r)V(\mathrm{d}r) < \infty$.

A shot-noise Cox point process can be viewed as a generalized cluster process. The measure V determines the distribution of the number of points in a cluster. By choosing an appropriate measure V, we can obtain a much more variable number of points in individual clusters than for the classical Poisson-Neyman-Scott process [15], Sec. 5.3.

The moment properties of shot-noise Cox processes are easily available [10], Sec. 4. For a space-time processes satisfying (3.2) we have

$$\varrho(u,t) = \mu \int_0^\infty rV(\mathrm{d}r) \int_{\mathbb{R}^2\times\mathbb{R}} k((u,t),(v,s))\, \mathrm{d}v\, \mathrm{d}s,\ \ (u,t) \in \mathbb{R}^2 \times \mathbb{R},$$

$$g((u,t),(v,s)) = 1 + \frac{\mu \int_0^\infty r^2 V(\mathrm{d}r) \int_{\mathbb{R}^2\times\mathbb{R}} k((u,t),(w,\tau))k((v,s),(w,\tau))\, \mathrm{d}w\, \mathrm{d}\tau}{\varrho(u,t)\varrho(v,s)},$$
$$(u,t),(v,s) \in \mathbb{R}^2 \times \mathbb{R}.$$

In both the equations we have a product of separate integrals for V and k. This property will be important for the estimation procedure discussed below. In order to simplify the notation we write in the following $V_1 = \int_0^\infty rV(\mathrm{d}r)$ and $V_2 = \int_0^\infty r^2 V(\mathrm{d}r)$.

Location-dependent thinning using an inhomogeneity function f applied to a stationary shot-noise Cox process specified by μ, V and k, yields a SOIRS shot-noise Cox process with the same μ and V but with a new kernel $\widetilde{k}((u,t),(v,s)) = f(u,t) \times k(v-u, s-t)$. In the following, however, we prefer the parametrization using the function f and the kernel $k(y-x)$ as opposed to the inhomogeneous kernel function $\widetilde{k}$.

4. Model specification and parametrization

Let X_0 be a stationary space-time shot-noise Cox process on $\mathbb{R}^2 \times \mathbb{R}$ specified by the constant $\mu > 0$, the measure V on $\mathbb{R}^+$ (parametrized by a scalar parameter θ) and the homogeneous kernel function $k(u,t)$. We denote by $\varrho_{0,k}$, $\varrho_{0,s,k}$ and $\varrho_{0,t,k}$ the kth-order intensity functions of X_0 and of the projection processes $X_{0,s}$ and $X_{0,t}$, respectively. Throughout this section, we assume that $\varrho_{0,2}$ exists and is bounded so that the pair-correlation function of X_0 is properly defined.

Let X be the SOIRS process obtained by location-dependent thinning from X_0 using the inhomogeneity function f. As before, we denote by ϱ_k, $\varrho_{s,k}$, and $\varrho_{t,k}$ the kth-order intensity functions of X and of the projection processes X_s and X_t, respectively.

Following [7] and [14], we adopt a pragmatic assumption that the inhomogeneity function f has a space-time product structure. In particular, we assume the parametric form of f

$$f(u,t;\beta_s,\beta_t) = f_1(z_1(u)\beta_s^{\mathrm{T}})f_2(z_2(t)\beta_t^{\mathrm{T}}), \tag{4.1}$$

where $z_1(u)$ and $z_2(t)$ are vectors of spatial and temporal covariates, respectively, and f_1, f_2 are positive, strictly increasing functions on $\mathbb{R}$. The vectors β_s, β_t denote the unknown inhomogeneity parameters. For a more concise notation we write in the following way: $f_1(z_1(u)\beta_s^{\mathrm{T}}) = f_1(u;\beta_s)$, and similarly for f_2. Moreover, we assume

$$\max_{u\in\mathbb{R}^2} f_1(u;\beta_s) = 1 = \max_{t\in\mathbb{R}} f_2(t;\beta_t).$$

This assumption prevents overparametrization of the model. We further set $\beta_0 = \log(\mu V_1)$ and hence the intensity function ϱ of X is parametrized by the vector $\beta = (\beta_0,\beta_s,\beta_t)$.

Further, following an example of a structured space-time Poisson cluster process in [14], Sec. 5, we assume a product structure of the kernel function k, i.e.

$$k(u,t) = k_1(u;\widetilde{\psi})k_2(t;\widetilde{\xi}), \quad u\in\mathbb{R}^2,\ t\in\mathbb{R}, \tag{4.2}$$

where $k_1(\cdot;\widetilde{\psi})$ and $k_2(\cdot;\widetilde{\xi})$ are probability density functions on $\mathbb{R}^2$ and $\mathbb{R}$, parametrized by the (vector) parameters $\widetilde{\psi}$ and $\widetilde{\xi}$, respectively.

The assumptions (4.1) and (4.2) allow us to introduce a tractable estimation procedure (the second-order moment characteristics of the projected processes have a tractable form) but do not imply spatio-temporal separability of the process X.

In the following we take advantage of the notation

$$K_1(v-u;\widetilde{\psi}) = \int_{\mathbb{R}^2} k_1(u-w;\widetilde{\psi})k_1(v-w;\widetilde{\psi})\,\mathrm{d}w, \quad u,v\in\mathbb{R}^2,$$
$$K_2(s-t;\widetilde{\xi}) = \int_{\mathbb{R}} k_2(t-\tau;\widetilde{\xi})k_2(s-\tau;\widetilde{\xi})\,\mathrm{d}\tau, \quad s,t\in\mathbb{R}.$$

The moment characteristics of X are derived easily from the model assumptions and the formulae in Section 3. The intensity function and the pair-correlation function are

$$\varrho(u,t;\beta) = \mu V_1 f_1(u;\beta_s)f_2(t;\beta_t), \quad (u,t)\in\mathbb{R}^2\times\mathbb{R},$$
$$g(v,s;\mu,\theta,\widetilde{\psi},\widetilde{\xi}) = 1 + \frac{V_2}{\mu(V_1)^2}K_1(v;\widetilde{\psi})K_2(s;\widetilde{\xi}), \quad v\in\mathbb{R}^2,\ s\in\mathbb{R}.$$

As in [16], we get for the intensity functions and the pair-correlation functions of X_t and X_s that

$$\begin{aligned}
\varrho_t(t;\beta) &= \mu V_1 f_2(t;\beta_t)\int_W f_1(w;\beta_s)\,\mathrm{d}w, \quad t\in\mathbb{R},\\
\varrho_s(u;\beta) &= \mu V_1 f_1(u;\beta_s)\int_T f_2(\tau;\beta_t)\,\mathrm{d}\tau, \quad u\in\mathbb{R}^2,\\
g_t(s;\xi) &= 1 + C_t\,\frac{V_2}{\mu\,(V_1)^2}K_2(s;\widetilde{\xi}) = 1+\xi_0 K_2(s;\widetilde{\xi}), \quad t,s\in\mathbb{R},\\
g_s(u;\psi) &= 1 + C_s\,\frac{V_2}{\mu\,(V_1)^2}K_1(u;\widetilde{\psi}) = 1+\psi_0 K_1(u;\widetilde{\psi}), \quad u,v\in\mathbb{R}^2,
\end{aligned}$$

where the constants C_t, C_s are given by

$$C_t = \frac{1}{(\int_W f_1(w;\beta_s)\,\mathrm{d}w)^2}\int_W\int_W f_1(u;\beta_s)f_1(v;\beta_s)K_1(v-u;\widetilde{\psi})\,\mathrm{d}u\,\mathrm{d}v, \tag{4.3}$$

$$C_s = \frac{1}{(\int_T f_2(\tau;\beta_t)\,\mathrm{d}\tau)^2}\int_T\int_T f_2(s;\beta_t)f_2(t;\beta_t)K_2(s-t;\widetilde{\xi})\,\mathrm{d}s\,\mathrm{d}t. \tag{4.4}$$

We also use the notation $\psi_0 = C_s\,V_2/[\mu\,(V_1)^2]$, $\xi_0 = C_t\,V_2/[\mu\,(V_1)^2]$ and $\psi=(\psi_0,\widetilde{\psi})$, $\xi=(\xi_0,\widetilde{\xi})$.

Clearly, the pair-correlation function g_t of X_t depends on the "spatial" part of the model (f_1 and k_1) only through the constant C_t. Analogously, the pair-correlation function g_s of X_s depends on f_2 and k_2 only through C_s.

In the following we use mainly the K-functions of the projection processes:

$$\begin{aligned}
K_s(r;\psi) &= \int_{\|u\|\leqslant r} g_s(u;\psi)\,\mathrm{d}u = \pi r^2 + \psi_0\int_{\|u\|<r}K_1(u;\widetilde{\psi})\,\mathrm{d}u, \quad r\geqslant 0,\\
K_t(t;\xi) &= \int_{-t}^{t} g_t(s;\xi)\,\mathrm{d}s = 2t + \xi_0\int_{|s|<t}K_2(s;\widetilde{\xi})\,\mathrm{d}s, \quad t\geqslant 0.
\end{aligned}$$

Below we also use the following notation: $\varrho^{(1)}$ and $\varrho^{(2)}$ are the first- and second-order derivatives of the intensity function of X w.r.t. the parameter β; $K_t^{(1)}(t;\xi), K_t^{(2)}(t;\xi)$ are the first- and second-order derivatives of $K_t(t;\xi)$ w.r.t. ξ; $K_s^{(1)}(r;\psi), K_s^{(2)}(r;\psi)$ are the first- and second-order derivatives of $K_s(r;\psi)$ w.r.t. ψ, assuming that the appropriate derivatives exist.

Finally, we denote by β^*, ψ^*, ξ^*, μ^*, θ^* the true parameter values governing the distribution of the process X.

5. Estimation method using projection processes

In this section we briefly describe the estimation method proposed in the paper [16], based on the notion of the spatial and temporal projection processes introduced in [14]. It is a step-wise estimation procedure analogous to the estimation of spatial SOIRS Cox point processes, in [21]. Even though the estimation method in [16] can be based either on the pair-correlation functions or the K-functions of the projection processes, we focus in the following only on the latter case—it enables us to formulate the asymptotic properties of the estimators. For more details on the estimation method see [16].

With the aim of discussing the asymptotic properties of the estimators, we focus here on estimation of β, ξ and ψ. For the parameters μ and θ, the calculations depend on the particular form of the measure V (more precisely, on the form of V_1 and V_2) and it is complicated to discuss the asymptotics in the general setting. Also, the inhomogeneity parameter β and the clustering parameters ξ and ψ are likely to be of main interest in practical applications.

In the step-wise estimation procedure, we take advantage of the special spatio-temporal structure of the process X. In the first step, the parameter β of the intensity function ϱ is estimated. In the second and third step, conditionally on the knowledge of ϱ, the K-functions of the projection processes are estimated from the data and used for minimum contrast estimation in order to estimate the parameters ξ and ψ, respectively. Finally, we also show how the parameters μ and θ of the underlying Poisson measure can be estimated from the previous estimates and the total number of points observed in $W \times T$.

5.1. First step. Ignoring for the moment the inter-point interactions, the inhomogeneity parameter β may be estimated by means of the Poisson log-likelihood score function given by

$$U_1(\beta) = \sum_{(u,t)\in X\cap(W\times T)} \frac{\varrho^{(1)}(u,t;\beta)}{\varrho(u,t;\beta)} - \int_{W\times T} \varrho^{(1)}(v,s;\beta)\,\mathrm{d}v\,\mathrm{d}s.$$

The estimate $\widehat{\beta}$ is obtained as a solution of the (vector) equation $U_1(\beta) = 0$.

5.2. Second step. In the next step we use $\widehat{\beta}$ to calculate the semi-parametric estimate $\widehat{K}_t(t;\widehat{\beta})$ of $K_t(t;\xi)$,

$$\widehat{K}_t(t;\widehat{\beta}) = \frac{1}{|T|} \sum_{s,\tau\in X_t}^{\neq} \frac{I(|s-\tau| \leqslant t)}{|T\cap T_{s-\tau}|\varrho_t(s;\widehat{\beta}_t)\varrho_t(\tau;\widehat{\beta}_t)},$$

where $T_{s-\tau}$ denotes the set T shifted by $s-\tau$. We use the translation edge-correction factor $|T \cap T_{s-\tau}|$, see [6], because it is convenient for our discussion on the asymptotic properties of the estimators. We then minimize the discrepancy

$$m_{2,\widehat{\beta}}(\xi) = \int_{t_0}^{t_1} (\widehat{K}_t(t;\widehat{\beta})^{c_2} - K_t(t;\xi)^{c_2})^2 \,\mathrm{d}t,$$

where $c_2 > 0$ is the variance-stabilizing exponent, usually taking on values $c_2 = 1/2$ or $1/4$, and $0 \leqslant t_0 < t_1$ are fixed constants. Assuming differentiability of $m_{2,\widehat{\beta}}(\cdot)$, this corresponds to solving the estimating equation

$$\begin{aligned}(5.1)\qquad U_2(\widehat{\beta},\xi) &= -|T|\frac{\partial m_{2,\widehat{\beta}}(\xi)}{\partial \xi} \\ &= 2c_2|T|\int_{t_0}^{t_1} [\widehat{K}_t(t;\widehat{\beta})^{c_2} - K_t(t;\xi)^{c_2}]K_t(t;\xi)^{c_2-1}K_t^{(1)}(t;\xi)\,\mathrm{d}t = 0.\end{aligned}$$

5.3. Third step. The estimation of ψ is analogous to the estimation of ξ in the second step. We use

$$\widehat{K}_s(r;\widehat{\beta}) = \sum_{x,y\in X_s}^{\neq} \frac{I(\|x-y\| \leqslant r)}{|W \cap W_{x-y}|\varrho_s(x;\widehat{\beta}_s)\varrho_s(y;\widehat{\beta}_s)}$$

to construct the discrepancy criterion to be minimized:

$$(5.2)\qquad m_{3,\widehat{\beta}}(\psi) = \int_{r_0}^{r_1} (\widehat{K}_s(u;\widehat{\beta})^{c_3} - K_s(u;\psi)^{c_3})^2 \,\mathrm{d}u.$$

Here W_{x-y} denotes the set W shifted by the vector $x-y$, $c_3 > 0$ is the variance-stabilizing exponent (the usual choices are $c_3 = 1/2$ or $1/4$) and $0 \leqslant r_0 < r_1$ are fixed constants.

Assuming differentiability of $m_{3,\widehat{\beta}}(\cdot)$, minimizing (5.2) corresponds to solving the estimating equation

$$\begin{aligned}U_3(\widehat{\beta},\psi) &= -|W|\frac{\partial m_{3,\widehat{\beta}}(\psi)}{\partial \psi} \\ &= 2c_3|W|\int_{r_0}^{r_1} [\widehat{K}_s(r;\widehat{\beta})^{c_3} - K_s(r;\psi)^{c_3}]K_s(r;\psi)^{c_3-1}K_s^{(1)}(r;\psi)\,\mathrm{d}r = 0.\end{aligned}$$

Altogether, the described estimation procedure can be formulated as solving the vector estimating equation $U(\beta,\xi,\psi) = (U_1(\beta), U_2(\beta,\xi), U_3(\beta,\psi)) = 0$ to obtain the parameter estimates $\widehat{\beta}$, $\widehat{\xi}$ and $\widehat{\psi}$.

5.4. Final step. Finally, we briefly comment on how the parameters μ and θ of the underlying Poisson measure can be estimated. Once the estimates of ψ and ξ have been computed, we can plug them into formulas (4.3), (4.4), and from $\widehat{\psi}_0$ or $\widehat{\xi}_0$ obtain the estimate of $\alpha = V_2/[\mu(V_1)^2]$. Finally, we calculate $\widehat{\theta}$ and $\widehat{\mu}$ from $\widehat{\alpha}$ and the equation

$$X(W \times T) = \widehat{\mu V_1} \int_W f_1(u; \widehat{\beta}_s)\,\mathrm{d}u \int_T f_2(t; \widehat{\beta}_t)\,\mathrm{d}t,$$

where $X(W \times T)$ plays the role of the estimate of $\mathbb{E}\,X(W \times T) = \int_{W\times T} \varrho(u, t; \beta)\,\mathrm{d}u\,\mathrm{d}t$. The actual form of the calculations depends on the precise form of V_1 and V_2 depending on θ.

6. Asymptotic properties

In this section we discuss the asymptotic properties of the estimators under the so-called increasing domain asymptotics, i.e. when the data is observed on an increasing sequence of compact observation windows. The proofs are deferred to Appendices. They are inspired by paper [21], which discusses the minimum contrast estimation in the purely spatial case, but contain important changes.

We consider the following asymptotic regime: $W_n \times T_n$, $n \geqslant 1$, is an increasing sequence of compact observation windows such that $W_n \times T_n \nearrow \mathbb{R}^2 \times \mathbb{R}$ and that for any $h \in \mathbb{R}^2$ and $k \in \mathbb{R}$,

$$\lim_{n\to\infty} \frac{|W_n \times T_n|}{|(W_n \cap W_{n,h}) \times (T_n \cap T_{n,k})|} = 1,$$

where $W_{n,h}$ denotes the set W_n shifted by $h \in \mathbb{R}^2$, and similarly for $T_{n,k}$. Moreover, it is required that $|\partial(W_n \times T_n)|/|W_n \times T_n| \to 0$, $n \to \infty$, where $|\partial(W_n \times T_n)|$ is the Hausdorff measure of the boundary of $W_n \times T_n$.

Let $(\widehat{\beta}_n, \widehat{\xi}_n, \widehat{\psi}_n)$ be the estimated parameter values calculated from $W_n \times T_n$, i.e. the solution of the equation $U_n(\beta, \xi, \psi) = (U_{n,1}(\beta), U_{n,2}(\beta, \xi), U_{n,3}(\beta, \psi)) = 0$, where

$$U_{n,1}(\beta) = \sum_{(u,t)\in X\cap(W_n\times T_n)} \frac{\varrho^{(1)}(u, t; \beta)}{\varrho(u, t; \beta)} - \int_{W_n\times T_n} \varrho^{(1)}(v, s; \beta)\,\mathrm{d}v\,\mathrm{d}s,$$

$$U_{n,2}(\beta, \xi) = 2c_2|T_n| \int_{t_0}^{t_1} [\widehat{K}_{t,n}(t; \beta)^{c_2} - K_t(t; \xi)^{c_2}] K_t(t; \xi)^{c_2-1} K_t^{(1)}(t; \xi)\,\mathrm{d}t,$$

$$U_{n,3}(\beta, \psi) = 2c_3|W_n| \int_{r_0}^{r_1} [\widehat{K}_{s,n}(r; \beta)^{c_3} - K_s(r; \psi)^{c_3}] K_s(r; \psi)^{c_3-1} K_s^{(1)}(r; \psi)\,\mathrm{d}r,$$

and $\widehat{K}_{t,n}$ and $\widehat{K}_{s,n}$ are the semi-parametric estimates of K_t and K_s, calculated using $X_t \cap T_n$ and $X_s \cap W_n$, respectively.

Note that for all $n \geqslant 1$, we use the same temporal projection process X_t (projected from the fixed spatial region W) to define $U_{n,2}$ and the same spatial projection process X_s (projected from the fixed time interval T) to define $U_{n,3}$. If, e.g. T_n were used to define the projection processes $X_s^{(n)}$, the resulting asymptotic regime for $U_{n,3}$ would be a combination of the increasing domain asymptotics and the so-called infill asymptotics. In particular, the intensity function of $X_s^{(n)}$ (at any location) would be an increasing, unbounded function of n. Moreover, the second-order moment characteristics of $X_s^{(n)}$ converge to those of a Poisson process and thus in the limit they provide no information about the clustering parameters.

Following the approach of [21], we approximate the functions $U_{n,2}(\beta^*, \xi^*)$ and $U_{n,3}(\beta^*, \psi^*)$ by

$$\widetilde{U}_{n,2}(\beta^*, \xi^*) = 2c_2^2 |T_n| \int_{t_0}^{t_1} [\widehat{K}_{t,n}(t; \beta^*) - K_t(t; \xi^*)] K_t(t; \xi^*)^{2c_2-2} K_t^{(1)}(t; \xi^*)\, \mathrm{d}t,$$

$$\widetilde{U}_{n,3}(\beta^*, \psi^*) = 2c_3^2 |W_n| \int_{r_0}^{r_1} [\widehat{K}_{s,n}(r; \beta^*) - K_s(r; \psi^*)] K_s(r; \psi^*)^{2c_3-2} K_s^{(1)}(r; \psi^*)\, \mathrm{d}r.$$

For $\widetilde{U}_{n,2}$, this approximation is based on the Taylor series expansion of the function x^{c_2}, applied on $\widehat{K}_{t,n}(t;\beta)^{c_2} - K_t(t;\xi)^{c_2}$, and similarly for $\widetilde{U}_{n,3}$. We further define

$$\begin{aligned}
\Sigma_{n,11} &= |W_n \times T_n|^{-1} \operatorname{Var}(U_{n,1}(\beta^*)),\\
\widetilde{\Sigma}_{n,22} &= |T_n|^{-1} \operatorname{Var}(\widetilde{U}_{n,2}(\beta^*, \xi^*)),\\
\widetilde{\Sigma}_{n,33} &= |W_n|^{-1} \operatorname{Var}(\widetilde{U}_{n,3}(\beta^*, \psi^*)),\\
J_n(\beta, \xi, \psi) &= -\frac{\partial}{\partial(\beta, \xi, \psi)^{\mathrm{T}}} U_n(\beta, \xi, \psi)\\
&= -\begin{pmatrix} \frac{\partial}{\partial \beta^{\mathrm{T}}} U_{n,1}(\beta) & \frac{\partial}{\partial \beta^{\mathrm{T}}} U_{n,2}(\beta, \xi) & \frac{\partial}{\partial \beta^{\mathrm{T}}} U_{n,3}(\beta, \psi) \\ 0 & \frac{\partial}{\partial \xi^{\mathrm{T}}} U_{n,2}(\beta, \xi) & 0 \\ 0 & 0 & \frac{\partial}{\partial \psi^{\mathrm{T}}} U_{n,3}(\beta, \psi) \end{pmatrix}\\
&= \begin{pmatrix} J_{n,11}(\beta) & J_{n,12}(\beta, \xi) & J_{n,13}(\beta, \psi) \\ 0 & J_{n,22}(\beta, \xi) & 0 \\ 0 & 0 & J_{n,33}(\beta, \psi) \end{pmatrix},\\
I_{n,11} &= \frac{1}{|W_n \times T_n|} \int_{W_n \times T_n} \frac{\varrho^{(1)}(v, s; \beta^*)^{\mathrm{T}} \varrho^{(1)}(v, s; \beta^*)}{\varrho(v, s; \beta^*)}\, \mathrm{d}v\, \mathrm{d}s,\\
I_{n,12} &= -2c_2^2 \int_{t_0}^{t_1} H_{n,2}(t; \beta^*) K_t(t; \xi^*)^{2c_2-2} K_t^{(1)}(t; \xi^*)\, \mathrm{d}t,
\end{aligned}$$

$$
\begin{aligned}
I_{n,13} &= -2c_3^2 \int_{r_0}^{r_1} H_{n,3}(r;\beta^*) K_s(r;\psi^*)^{2c_3-2} K_s^{(1)}(r;\psi^*)\,\mathrm{d}r, \\
I_{22} &= 2c_2^2 \int_{t_0}^{t_1} K_t(t;\xi^*)^{2c_2-2} K_t^{(1)}(t;\xi^*)^{\mathrm{T}} K_t^{(1)}(t;\xi^*)\,\mathrm{d}t, \\
I_{33} &= 2c_3^2 \int_{r_0}^{r_1} K_s(r;\psi^*)^{2c_3-2} K_s^{(1)}(r;\psi^*)^{\mathrm{T}} K_s^{(1)}(r;\psi^*)\,\mathrm{d}r,
\end{aligned}
$$

where

$$
\begin{aligned}
H_{n,2}(t;\beta^*) &= \mathbb{E}\, \frac{\partial}{\partial \beta^{\mathrm{T}}} \widehat{K}_{t,n}(t;\beta)|_{\beta=\beta^*} \\
&= -2 \int_{T_n}\int_{T_n} \frac{I\{|s-\tau|<t\}}{|T_n \cap T_{n,s-\tau}|} \frac{\varrho_t^{(1)}(s;\beta^*)}{\varrho_t(s;\beta^*)} g_t(s-\tau;\xi^*)\,\mathrm{d}s\,\mathrm{d}\tau, \\
H_{n,3}(t;\beta^*) &= \mathbb{E}\, \frac{\partial}{\partial \beta^{\mathrm{T}}} \widehat{K}_{s,n}(r;\beta)|_{\beta=\beta^*} \\
&= -2 \int_{W_n}\int_{W_n} \frac{I\{\|u-v\|<r\}}{|W_n \cap W_{n,u-v}|} \frac{\varrho_s^{(1)}(u;\beta^*)}{\varrho_s(u;\beta^*)} g_s(u-v;\psi^*)\,\mathrm{d}u\,\mathrm{d}v.
\end{aligned}
$$

Now we can formulate the consistency theorem, inspired by [21].

Theorem 1. *Apart from the model assumptions formulated above, let the following conditions be met:*

- (A1) ϱ *is twice continuously differentiable as a function of* β,
- (A2) $\exists\, C_1 < \infty$ *such that* $\|z_1(u)\| < C_1, \|z_2(t)\| < C_1$, $u \in \mathbb{R}^2$, $t \in \mathbb{R}$,
- (A3) I_{22} *and* I_{33} *are positive definite matrices and* $\liminf \omega_{n,11} > 0$, *where* $\omega_{n,11}$ *is the smallest eigenvalue of* $I_{n,11}$,
- (A4) $\widetilde{\Sigma}_{n,22}$ *and* $\widetilde{\Sigma}_{n,33}$ *converge to positive definite matrices* $\widetilde{\Sigma}_{22}$ *and* $\widetilde{\Sigma}_{33}$, *respectively,*
- (A5) $K_s(r;\psi), K_s^{(1)}(r;\psi), K_s^{(2)}(r;\psi)$ *exist and are continuous functions of* (r,ψ),
- (A6) $K_t(t;\xi), K_t^{(1)}(t;\xi), K_t^{(2)}(t;\xi)$ *exist and are continuous functions of* (t,ξ),
- (A7) $t_0 \geqslant 0$ *for* $c_2 \geqslant 2$, *otherwise* $t_0 > 0$; *similarly,* $r_0 \geqslant 0$ *for* $c_3 \geqslant 2$, *otherwise* $r_0 > 0$,
- (A8) $\varrho_{0,2}$ *and* $\varrho_{0,3}$ *exist and are bounded, and the second-order reduced factorial cumulant measure of* X_0 *has finite total variation,*
- (A9) $\exists\, C_2 < \infty$ *such that for all* $u_1, u_2 \in \mathbb{R}^2$:

$$
\int_{\mathbb{R}^2} |\varrho_{0,s,4}(0,u_1,v,u_2+v) - \varrho_{0,s,2}(0,u_1)\,\varrho_{0,s,2}(0,u_2)|\,\mathrm{d}v < C_2,
$$

- (A10) $\exists\, C_3 < \infty$ *such that for all* $s_1, s_2 \in \mathbb{R}$:

$$
\int_{\mathbb{R}} |\varrho_{0,t,4}(0,s_1,\tau,s_2+\tau) - \varrho_{0,t,2}(0,s_1)\,\varrho_{0,t,2}(0,s_2)|\,\mathrm{d}\tau < C_3.
$$

Then there is a sequence $\{(\widehat{\beta}_n, \widehat{\xi}_n, \widehat{\psi}_n)\}_{n\geqslant 1}$ *for which* $U_n(\widehat{\beta}_n, \widehat{\xi}_n, \widehat{\psi}_n) = 0$ *with probability tending to* 1 *and the sequence*

$$M_n = (|W_n \times T_n|^{1/2}(\widehat{\beta}_n - \beta^*),\ |T_n|^{1/2}(\widehat{\xi}_n - \xi^*),\ |W_n|^{1/2}(\widehat{\psi}_n - \psi^*))$$

is bounded in probability, i.e. $\forall\, \varepsilon > 0\ \exists\, \delta > 0\colon\ \mathbb{P}(\|M_n\| > \delta) \leqslant \varepsilon$ *for* n *sufficiently large.*

Concerning the assumptions of Theorem 1, condition (A1) is not restrictive. It covers, among others, the log-linear model for ϱ, which is the most popular in applications. Assumption (A2) of bounded covariates can be easily justified in practice.

Regarding (A3), it is sufficient for I_{22} to be a positive definite matrix that there are distinct values $t_0 < \tau_1 < \tau_2 < \ldots < \tau_q < t_1$ (q is the number of elements of the vector ξ^*) such that the matrix with rows $K_t^{(1)}(\tau_i; \xi^*)$ has full rank. For a detailed example see [21], Sec. 3.3. The smallest eigenvalue of $I_{n,11}$ depends on all the values of the covariates in $\mathbb{R}^2 \times \mathbb{R}$, making it difficult to discuss the condition on $\omega_{n,11}$ in the general setting. The same applies to condition (A4) on the limiting behaviour of the matrices $\widetilde{\Sigma}_{n,22}$ and $\widetilde{\Sigma}_{n,33}$.

Conditions (A5) and (A6) in fact impose restrictions on the kernels k_1 and k_2, respectively. For example, Gaussian kernels satisfy these conditions. For the uniform (Matérn-type) kernels the dimension comes into play. Namely, a short calculation shows that for a uniform circular kernel k_1 in $\mathbb{R}^2$, (A5) is fulfilled. On the other hand, for a uniform kernel k_2 in $\mathbb{R}$, (A6) does not hold.

Assumption (A7) is only technical. In applications one can use very small positive values of t_0 and r_0 without hesitation.

Condition (A8) relates to the stationary (unthinned) version X_0 of the process. It follows from [10], Sec. 4, that if the kernel k is bounded and $\int_0^\infty r^m V(\mathrm{d}r) < \infty$ for some $m \in \mathbb{N}$, then $\varrho_{0,m}$ is bounded and all reduced factorial cumulant measures up to order m have finite total variation. Thus, (A8) holds if these conditions are fulfilled for $m = 3$.

A sufficient condition for (A9) can be formulated in terms of the second- and fourth-order intensity functions of the space-time process X_0. It suffices if there is a constant $\widetilde{C}_2$ such that for all $u_1, u_2 \in \mathbb{R}^2$ and $s_1, s_2, s_3 \in \mathbb{R}$ we have:

$$\int_{\mathbb{R}^2} |\varrho_{0,4}((0,0),(u_1,s_1),(v,s_2),(u_2+v,s_3)) \\ - \varrho_{0,2}((0,0),(u_1,s_1))\,\varrho_{0,2}((0,0),(u_2,s_3-s_2))|\,\mathrm{d}v < \widetilde{C}_2.$$

For (A10) a similar sufficient condition can be formulated.

We now proceed to the formulation of the asymptotic normality results for the estimators considered above. First, we discuss the properties of the estimator $\widehat{\beta}_n$

based on the space-time process X. For a Borel set $A \in \mathcal{B}(\mathbb{R}^2 \times \mathbb{R})$, denote by $\mathcal{F}^X(A)$ the σ-algebra generated by $X \cap A$.

For general σ-algebras $\mathcal{F}_1$ and $\mathcal{F}_2$, let $\alpha(\mathcal{F}_1, \mathcal{F}_2) = \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|, A \in \mathcal{F}_1, B \in \mathcal{F}_2\}$ denote the standard strong mixing coefficient [4]. For $h > 0$ let $A_{ijk} = [ih, (i+1)h) \times [jh, (j+1)h) \times [kh, (k+1)h)$, $(i,j,k) \in \mathbb{Z}^3$, and

$$\alpha^F_{p_1,p_2}(m) = \sup\Big\{\alpha(\mathcal{F}^X(S_1), \mathcal{F}^X(S_2))\colon\ S_1 = \bigcup_{M_1} A_{ijk},\ S_2 = \bigcup_{M_2} A_{ijk},\\ |M_1| \leqslant p_1,\ |M_2| \leqslant p_2,\ d(M_1, M_2) \geqslant m,\ M_1, M_2 \subset \mathbb{Z}^3\Big\},$$

where $|M|$ is the cardinality of the set $M \subseteq \mathbb{Z}^3$ and $d(M_1, M_2)$ denotes the minimal distance between M_1 and M_2 in the grid $\mathbb{Z}^3$.

Theorem 2. *Apart from the model assumptions formulated above and* (A1)–(A10)*, suppose there exist* $\delta > 0$ *and* $\nu \in \mathbb{N}$, $0 < \delta < \nu$, *such that*

(B1) $\varrho_{0,2+\nu}((u_1, t_1), \ldots, (u_{2+\nu}, t_{2+\nu})) < \infty$,
(B2) *there exist* $h > 0$ *and* $d > 3(2+\delta)/\delta$ *such that* $\alpha^F_{2,\infty}(m) = \mathcal{O}(m^{-d})$,
(B3) *the matrix* $\Sigma_{n,11}$ *converges to a positive definite matrix* Σ_{11}.

Then $|W_n \times T_n|^{1/2}(\widehat{\beta}_n - \beta^*) I_{n,11} \Sigma^{-1/2}_{n,11} \xrightarrow{d} N(0, \mathbf{1})$, *where* $\xrightarrow{d}$ *denotes convergence in distribution and* $\mathbf{1}$ *is the identity matrix of appropriate dimension.*

Now we focus on the properties of the estimator $\widehat{\xi}_n$ based on the temporal projection process X_t. For a Borel set $B \in \mathcal{B}(\mathbb{R})$, denote by $\mathcal{F}^{X_t}(B)$ the σ-algebra generated by $X_t \cap B$. For $h > 0$ let $B_i = [ih, (i+1)h)$, $i \in \mathbb{Z}$, and

$$\alpha^F_{t,p_1,p_2}(m) = \sup\Big\{\alpha(\mathcal{F}^{X_t}(S_1 \oplus t_1), \mathcal{F}^{X_t}(S_2 \oplus t_1))\colon\ S_1 = \bigcup_{M_1} B_i,\ S_2 = \bigcup_{M_2} B_i,\\ |M_1| \leqslant p_1,\ |M_2| \leqslant p_2,\ d(M_1, M_2) \geqslant m,\ M_1, M_2 \subset \mathbb{Z}\Big\},$$

where $|M|$ is the cardinality of the set $M \subseteq \mathbb{Z}$ and $d(M_1, M_2)$ denotes the minimal distance between M_1 and M_2 in the grid $\mathbb{Z}$. Also, $S_i \oplus t_1$ denotes the set S_i dilated by the distance t_1, where t_1 is the upper limit used in the minimum contrast criterion (5.1).

Theorem 3. *Apart from the model assumptions formulated above and* (A1)–(A10)*, suppose there exist* $\delta > 0$ *and* $\nu \in \mathbb{N}$, $0 < \delta < \nu$, *such that*

(C1) $\varrho_{0,4+2\nu}((u_1, t_1), \ldots, (u_{4+2\nu}, t_{4+2\nu})) < \infty$,
(C2) *there exist* $h > 0$ *and* $d > (2+\delta)/\delta$ *such that* $\alpha^F_{t,2,\infty}(m) = \mathcal{O}(m^{-d})$.

Then $|T_n|^{1/2}(\widehat{\xi}_n - \xi^*) I_{22} \widetilde{\Sigma}^{-1/2}_{n,22} \xrightarrow{d} N(0, \mathbf{1})$.

Finally, we discuss the properties of the estimator $\widehat{\psi}_n$ based on the spatial projection process X_s. For a Borel set $C \in \mathcal{B}(\mathbb{R}^2)$, denote by $\mathcal{F}^{X_s}(C)$ the σ-algebra generated by $X_s \cap C$. For $h > 0$ let $C_{ij} = [ih, (i+1)h) \times [jh, (j+1)h)$, $(i, j) \in \mathbb{Z}^2$, and

$$\alpha^F_{s,p_1,p_2}(m) = \sup\Big\{\alpha(\mathcal{F}^{X_s}(S_1 \oplus r_1),\ \mathcal{F}^{X_s}(S_2 \oplus r_1))\colon\ S_1 = \bigcup_{M_1} C_{ij},\ S_2 = \bigcup_{M_2} C_{ij},$$
$$|M_1| \leqslant p_1,\ |M_2| \leqslant p_2,\ d(M_1, M_2) \geqslant m,\ M_1, M_2 \subset \mathbb{Z}^2\Big\},$$

where $|M|$ is the cardinality of the set $M \subseteq \mathbb{Z}^2$ and $d(M_1, M_2)$ denotes the minimal distance between M_1 and M_2 in the grid $\mathbb{Z}^2$. Also, $S_i \oplus r_1$ denotes the set S_i dilated by the distance r_1, where r_1 is the upper limit used in the minimum contrast criterion (5.2).

Theorem 4. *Apart from the model assumptions formulated above and* (A1)–(A10), *suppose there exist* $\delta > 0$ *and* $\nu \in \mathbb{N}$, $0 < \delta < \nu$, *such that*

(C1) $\varrho_{0,4+2\nu}((u_1, t_1), \dots, (u_{4+2\nu}, t_{4+2\nu})) < \infty$,
(D2) *there exist* $h > 0$ *and* $d > 2(2+\delta)/\delta$ *such that* $\alpha^F_{s,2,\infty}(m) = \mathcal{O}(m^{-d})$.

Then $|W_n|^{1/2}(\widehat{\psi}_n - \psi^*)I_{33}\widetilde{\Sigma}^{-1/2}_{n,33} \xrightarrow{d} N(0, \mathbf{1})$.

It is possible to formulate sufficient conditions for the mixing assumptions above in terms of conditions on the tail behaviour of the kernels k_1 and k_2. They are generally easier to verify than the original mixing conditions (B2), (C2), and (D2). The proofs follow the one given in Appendix C of the recent paper [5].

Lemma 1. *Let* X_0 *be a stationary shot-noise Cox process in* $\mathbb{R}^2 \times \mathbb{R}$ *with well-defined first-order moment measure and smoothing kernel* $k(u, t) = k_1(u)k_2(t)$, $u \in \mathbb{R}^2$, $t \in \mathbb{R}$, *satisfying*

$$\sup_{(u,t)\in[-m/2,m/2]^3}\Big\{\int_{\mathbb{R}^3\setminus[-m,m]^3} k_1(v-u)k_2(s-t)\,\mathrm{d}(v,s)\Big\} = \mathcal{O}(m^{-d-3}). \tag{6.1}$$

Then X_0 *satisfies condition* (B2).

Furthermore, if k_2 *satisfies* $\sup_{s\in[-m/2,m/2]}\big\{\int_{\mathbb{R}\setminus[-m,m]} k_2(s-\tau)\,\mathrm{d}\tau\big\} = \mathcal{O}(m^{-d-1})$, *then* $X_{0,t}$ *satisfies condition* (C2). *Also, condition* (D2) *holds for* $X_{0,s}$ *if* k_1 *satisfies* $\sup_{u\in[-m/2,m/2]^2}\big\{\int_{\mathbb{R}^2\setminus[-m,m]^2} k_1(v-u)\,\mathrm{d}v\big\} = \mathcal{O}(m^{-d-2})$.

The inhomogeneous process X, formed by location-dependent thinning of the stationary process X_0, inherits the mixing properties of X_0. Thus, (6.1) ensures that

(B2) holds also for X, and similarly for the other conditions and the assumptions (C2) and (D2), respectively.

7. Discussion

The estimation method for space-time shot-noise Cox processes, proposed in [16] and based on projections of the process to the spatial and temporal domain, proved to be practically applicable but lacking theoretical foundations. In the present paper we remedy this flaw by providing a set of sufficient conditions under which the estimates of the parameters governing the inhomogeneity and clustering can be proved to be consistent and asymptotically normally distributed.

The reason for presenting the asymptotic normality results in Section 6 as three separate theorems is that due to the different normalization required for each estimation step, we cannot prove joint asymptotic normality for the vector $(\widehat{\beta}_n, \widehat{\xi}_n, \widehat{\psi}_n)$ using the current methodology. The crucial point is that the variances of $U_{n,1}(\beta^*)$, $U_{n,2}(\beta^*, \xi^*)$, and $U_{n,3}(\beta^*, \psi^*)$ grow with different orders. In particular, they grow as $|W_n \times T_n|$, $|T_n|$, and $|W_n|$, respectively. In an attempt to prove joint asymptotic normality, depending on the normalization factors used, we either encounter singular limiting variance matrices or fail to check the uniform integrability condition (see condition (b) at the end of Appendix B). This is one of the main differences from the purely spatial case discussed in [21].

The other important difference is that we use different mixing coefficients which are more suitable in this context and enable us to establish in Lemma 1 sufficient conditions for the required mixing properties in terms of the tail behaviour of the kernels k_1 and k_2. For more detailed discussion on the different choice of the mixing coefficient, see [17].

Finally, we remark that the practical application of the theoretical results presented in Section 6 lies e.g. in the construction of confidence regions for the parameters β, ξ, and ψ. This can be done by using a plug-in approach, as discussed in Sec. 3.2 and App. B of [21]. The construction is based on computing the estimates of $\Sigma_{n,11}$, $\widetilde{\Sigma}_{n,22}$, and $\widetilde{\Sigma}_{n,33}$, together with $I_{n,11}$, I_{22}, and I_{33}, using the estimated values $\widehat{\beta}_n$, $\widehat{\xi}_n$, and $\widehat{\psi}_n$.

Appendix A. Proof of Theorem 1

The proof is based on the general asymptotic result given in [21], App. C. We formulate here only the statement of the result, for the proof see [21], App. C.

Consider a parametrized family of probability measures $\mathbb{P}_\theta, \theta \in \mathbb{R}^p$, and a sequence of estimating functions $u_n \colon \mathbb{R}^p \to \mathbb{R}^p$, $n \geqslant 1$. The distribution of $\{u_n(\theta)\}_{n \geqslant 1}$ is governed by $\mathbb{P} = \mathbb{P}_{\theta^*}$, where θ^* denotes the true parameter value. For a matrix $A = (a_{ij})$, let $\|A\|_M = \max_{ij} |a_{ij}|$ and let $J_n(\theta) = -\partial u_n(\theta)/\partial\theta^{\mathrm{T}}$.

Lemma 2. *Assume that there is a sequence of invertible symmetric matrices V_n such that*

(a) $\|V_n^{-1}\|_M \to 0$,

(b) *there exists $l > 0$ such that* $\mathbb{P}\big(\inf_{\|\varphi\|=1}\{\varphi V_n^{-1} J_n(\theta^*) V_n^{-1}\varphi^{\mathrm{T}}\} < l\big) \to 0$,

(c) *for any $d > 0$,* $\gamma_{nd} = \sup_{\|(\theta-\theta^*)V_n\|\leqslant d} [\|V_n^{-1}\{J_n(\theta) - J_n(\theta^*)\}V_n^{-1}\|_M] \to 0$ *in probability under $\mathbb{P} = \mathbb{P}_{\theta^*}$,*

(d) *the sequence $u_n(\theta^*)V_n^{-1}$ is bounded in probability, i.e., for each $\varepsilon > 0$ there exists d such that $\mathbb{P}(\|u_n(\theta^*)V_n^{-1}\| > d) \leqslant \varepsilon$ for n sufficiently large.*

Then for each $\varepsilon > 0$ there exists $d > 0$ such that

$$\mathbb{P}\big(\exists\, \widetilde{\theta}_n\colon\ u_n(\widetilde{\theta}_n) = 0 \text{ and } \|(\widetilde{\theta}_n - \theta^*)V_n\| < d\big) > 1 - \varepsilon$$

whenever n is sufficiently large.

To prove Theorem 1 we apply Lemma 2 succesively as follows. In the first step we apply it to $u_n = U_{n,1}$ with $V_n = |W_n \times T_n|^{1/2} \cdot \mathbf{1}$, where $\mathbf{1}$ is the identity matrix of the appropriate dimension. It follows that there is a sequence $\{\widehat{\beta}_n\}_{n\geqslant 1}$ such that $|W_n \times T_n|^{1/2}\|\widehat{\beta}_n - \beta^*\|$ is bounded in probability and $U_{n,1}(\widehat{\beta}_n) = 0$ with probability tending to 1. This also implies $\widehat{\beta}_n \to \beta^*$ in probability as $n \to \infty$.

In the second step we use Lemma 2 on $u_n(\cdot) = U_{n,2}(\widehat{\beta}_n, \cdot)$ with $V_n = (|T_n|\widetilde{\Sigma}_{n,22})^{1/2}$ to show that there is a sequence $\{\widehat{\xi}_n\}_{n\geqslant 1}$ such that $U_{n,2}(\widehat{\beta}_n, \widehat{\xi}_n) = 0$ with probability tending to 1 and $|T_n|^{1/2}\|\widehat{\xi}_n - \xi^*\|$ is bounded in probability. The difficult part in using this lemma is to show boundedness in probability of $|T_n|^{-1/2}U_{n,2}(\widehat{\beta}_n, \xi^*)\widetilde{\Sigma}_{n,22}^{-1/2}$ in condition (d). To do this, we use a Taylor series expansion

$$\text{(A.1)}\qquad \begin{aligned} |T_n|^{-1/2}U_{n,2}(\widehat{\beta}_n, \xi^*)\widetilde{\Sigma}_{n,22}^{-1/2} &= |T_n|^{-1/2}U_{n,2}(\beta^*, \xi^*)\widetilde{\Sigma}_{n,22}^{-1/2} \\ &\quad - |T_n|^{-1/2}(\widehat{\beta}_n - \beta^*)J_{n,12}(\widetilde{\beta}, \xi^*)\widetilde{\Sigma}_{n,22}^{-1/2}, \end{aligned}$$

where $\|\widetilde{\beta} - \beta^*\| \leqslant \|\widehat{\beta}_n - \beta^*\|$. In this way we can show boundedness in probability of the two terms on the right-hand side which are easier to handle than $|T_n|^{-1/2}U_{n,2}(\widehat{\beta}_n, \xi^*)\widetilde{\Sigma}_{n,22}^{-1/2}$.

Finally, in the third step we use Lemma 2 in a similar way on $u_n(\cdot) = U_{n,3}(\widehat{\beta}_n, \cdot)$ with $V_n = (|W_n|\widetilde{\Sigma}_{n,33})^{1/2}$ to show that there is a sequence $\{\widehat{\psi}_n\}_{n\geqslant 1}$ such that $|W_n|^{1/2}\|\widehat{\psi}_n - \psi^*\|$ is bounded in probability and $U_{n,3}(\widehat{\beta}_n, \widehat{\psi}_n) = 0$ with probability tending to 1. To verify condition (d) of the lemma we use a similar Taylor series expansion as above.

Thus, Theorem 1 will be proved, once we check the conditions of Lemma 2 for the three cases described above.

First step. Condition (a) of Lemma 2 follows from the fact that $|W_n \times T_n| \to \infty$ as $n \to \infty$.

To verify condition (b), first note that $|W_n \times T_n|^{-1} J_{n,11}(\beta^*)$ is a real symmetric matrix and $\inf_{\|\varphi\|=1} \{\varphi J_{n,11}(\beta^*)|W_n \times T_n|^{-1}\varphi^{\mathrm{T}}\}$ is equal to the smallest eigenvalue of $|W_n \times T_n|^{-1} J_{n,11}(\beta^*)$ [9]. Since the eigenvalues of a matrix are in fact the roots of a certain polynomial, [12], Sec. 2 gives uniform continuity of the mapping g: $A \mapsto \inf_{\|\varphi\|=1} \{\varphi A \varphi^{\mathrm{T}}\}$.

Lemma 5 below shows that $|W_n \times T_n|^{-1} J_{n,11}(\beta^*) - I_{n,11}$ converges to 0 in probability as $n \to \infty$. Taking advantage of the uniform continuity, it is easy to show that also $g(|W_n \times T_n|^{-1} J_{n,11}(\beta^*)) - g(I_{n,11})$ converges to 0 in probability as $n \to \infty$. This and assumption (A3) verify condition (b), where we take $l = \frac{1}{2} \liminf \omega_{n,11} > 0$.

Regarding condition (c), it is sufficient to show that

$$\gamma_{nd}^{ij} = \sup_{\|(\theta-\theta^*)|W_n \times T_n|^{1/2}\| \leqslant d} \frac{|J_{n,11}^{ij}(\theta) - J_{n,11}^{ij}(\theta^*)|}{|W_n \times T_n|}$$

converges in probability to 0 as $n \to \infty$, where $J_{n,11}^{ij}$ is the (i,j)-th element of the matrix $J_{n,11}$. To do this, we need to control the differences of the type $|\varrho^{(2)}(u,t;\beta) - \varrho^{(2)}(u,t;\beta^*)|$, where the arguments (u,t) are the same in both terms. Such differences can be uniformly bounded from above by assumptions (A1) and (A2). The required convergence follows.

Using the Campbell theorem and assumptions (A2) and (A8), it is easy to show for each element of the vector $|W_n \times T_n|^{-1/2} U_{n,1}(\beta^*)$ that its mean is 0 and its variance is bounded from above by the same constant for all n. This implies that $|W_n \times T_n|^{-1/2} U_{n,1}(\beta^*)$ is bounded in probability and hence condition (d) is verified.

Second step. Condition (a) of Lemma 2 follows easily from the fact that $|T_n| \to \infty$, $n \to \infty$, and $\widetilde{\Sigma}_{n,22} \to \widetilde{\Sigma}_{22}$ as $n \to \infty$ by assumption (A4). It also secures invertibility of $\widetilde{\Sigma}_{n,22}$, at least for n large enough.

Regarding condition (b), we show in Lemma 5 that $|T_n|^{-1} J_{n,22}(\widehat{\beta}_n, \xi^*) - I_{22} \to 0$ in probability as $n \to \infty$, and thus

$$|T_n|^{-1} \widetilde{\Sigma}_{n,22}^{-1/2} J_{n,22}(\widehat{\beta}_n, \xi^*) \widetilde{\Sigma}_{n,22}^{-1/2} - \widetilde{\Sigma}_{n,22}^{-1/2} I_{22} \widetilde{\Sigma}_{n,22}^{-1/2} \to 0$$

in probability as $n \to \infty$.

By assumption (A3), I_{22} is a positive definite matrix and hence all its eigenvalues are positive. This implies that lim inf of the smallest eigenvalue of the matrix $\widetilde{\Sigma}_{n,22}^{-1/2} I_{22} \widetilde{\Sigma}_{n,22}^{-1/2}$ is positive. Now, one can use the same argument as in the first step.

To verify condition (c), we need to check that for any $d > 0$, $\gamma_{nd} \to 0$ in probability as $n \to \infty$, where

$$\gamma_{nd} = \sup_{\|(\xi-\xi^*)|T_n|^{1/2}\widetilde{\Sigma}_{n,22}^{1/2}\| \leqslant d} \left\| \frac{\widetilde{\Sigma}_{n,22}^{-1/2}(J_{n,22}(\widehat{\beta}_n, \xi) - J_{n,22}(\widehat{\beta}_n, \xi^*))\widetilde{\Sigma}_{n,22}^{-1/2}}{|T_n|} \right\|_M.$$

By assumption (A4), $\widetilde{\Sigma}_{n,22}$ converges to a deterministic positive definite matrix and thus it is sufficient to verify the convergence in probability for

$$\widetilde{\gamma}_{nd} = \sup_{\|(\xi-\xi^*)|T_n|^{1/2}\| \leqslant d} \left\| \frac{J_{n,22}(\widehat{\beta}_n, \xi) - J_{n,22}(\widehat{\beta}_n, \xi^*)}{|T_n|} \right\|_M.$$

As in the first step, we verify the convergence for each element of $\widetilde{\gamma}_{nd}$ separately. Using the continuity assumptions in (A6), the difference in $\widetilde{\gamma}_{nd}$ can be made arbitrarily small by choosing n large enough and thus the required convergence is obtained and condition (c) is verified.

In view of equation (A.1), it is sufficient for condition (d) to be satisfied that the following quantities are bounded in probability (note that $|T_n|^{1/2}(\widehat{\beta}_n - \beta^*)$ is bounded in probability from the first step):

$$|T_n|^{-1/2}U_{n,2}(\beta^*, \xi^*)\widetilde{\Sigma}_{n,22}^{-1/2}, \quad |T_n|^{-1}J_{n,12}(\widetilde{\beta}, \xi^*)\widetilde{\Sigma}_{n,22}^{-1/2},$$

where $\|\widetilde{\beta} - \beta^*\| \leqslant \|\widehat{\beta}_n - \beta^*\|$. Regarding the first term, consider the approximation

$$|T_n|^{-1/2}U_{n,2}(\beta^*, \xi^*) = |T_n|^{-1/2}\widetilde{U}_{n,2}(\beta^*, \xi^*) + |T_n|^{-1/2}V_{n,2}(\beta^*, \xi^*),$$

where $\widetilde{U}_{n,2}$ is defined in Section 6. We argue below that $|T_n|^{-1/2}V_{n,2}(\beta^*, \xi^*)$ converges to 0 in probability. By the Campbell theorem, $|T_n|^{-1/2}\widetilde{U}_{n,2}(\beta^*, \xi^*)\widetilde{\Sigma}_{n,22}^{-1/2}$ has mean 0 and its variance is the identity matrix. It follows that $|T_n|^{-1/2}\widetilde{U}_{n,2}(\beta^*, \xi^*)$ and hence also $|T_n|^{-1/2}U_{n,2}(\beta^*, \xi^*)$ is bounded in probability.

Regarding the second term, i.e. $|T_n|^{-1}J_{n,12}(\widetilde{\beta}, \xi^*)\widetilde{\Sigma}_{n,22}^{-1/2}$, it can be checked that it is bounded in probability by using Lemmas 3 and 4, assumptions (A1), (A2), (A4), (A6), and (A7), and the Cauchy-Schwarz inequality and the Fubini theorem in appropriate places.

It remains to verify that $|T_n|^{-1/2}V_{n,2}(\beta^*, \xi^*)$ converges to 0 in probability but this can be checked using the same methods and assumptions. Combining these results with (A4), we get condition (d) for the second step.

Third step. For verifying conditions (a)–(d) of Lemma 2, the same arguments can be used as in the second step. The only difference is that $|W_n|$ appears instead of $|T_n|$ and K_s instead of K_t. Thus, the details are omitted here.

We conclude the proof with three technical lemmas. The first is a version of Lemma 1 in [21], App. C.

Lemma 3. *Under assumptions* (A2), (A8), *and* (A9), *the variance of*

$$\sum_{u,v\in X_s\cap W_n}^{\neq} \frac{I(\|u-v\|\leqslant r)h(u,v)}{|W_n|\lambda_s(u;\beta^*)\lambda_s(v;\beta^*)}$$

is $\mathcal{O}(|W_n|^{-1})$ *for any bounded function* $h(u,v)$ *symmetric in its arguments and for any* $r \geqslant 0$.

Under assumptions (A2), (A8), *and* (A10), *the variance of*

$$\sum_{t,s\in X_t\cap T_n}^{\neq} \frac{I(|s-t|\leqslant \tau)h(t,s)}{|T_n|\lambda_t(t;\beta^*)\lambda_t(s;\beta^*)}$$

is $\mathcal{O}(|T_n|^{-1})$ *for any bounded function* $h(t,s)$ *symmetric in its arguments and for any* $\tau \geqslant 0$.

P r o o f. The two parts of this lemma can be proved using the same arguments. We focus here on the first part (spatial projection process).

Let $\varphi(u,v) = I(\|u-v\|\leqslant r)h(u,v)/\{\lambda_s(u;\beta^*)\lambda_s(v;\beta^*)\}$. Assumption (A2) ensures that λ_s is bounded from below by a positive constant and hence φ is bounded from above. Also, φ is a symmetric function. By the Campbell theorem, the variance of the sum above is equal to

$$\begin{aligned}|W_n|^{-2}&\int_{W_n^4}\varphi(u,v)\varphi(w,z)[\lambda_{s,4}(u,v,w,z)-\lambda_{s,2}(u,v)\lambda_{s,2}(w,z)]\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w\,\mathrm{d}z\\ &+4|W_n|^{-2}\int_{W_n^3}\varphi(u,v)\varphi(v,w)\lambda_{s,3}(u,v,w)\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w\\ &+2|W_n|^{-2}\int_{W_n^2}\varphi(u,v)^2\lambda_{s,2}(u,v)\,\mathrm{d}u\,\mathrm{d}v.\end{aligned}$$

It then follows by direct calculation that each of the three terms is $\mathcal{O}(|W_n|^{-1})$. □

The second lemma is a generalized version of Lemma 2 in [21], App. C.

Lemma 4. *Consider a sequence* $\{\breve{\beta}_n\}_{n\geqslant 1}$ *such that* $\breve{\beta}_n \to \beta^*$ *in probability as* $n\to\infty$. *Under the assumptions of Theorem* 1, $\sup_{t\in[t_0,t_1]} |\widehat{K}_{t,n}(t;\breve{\beta}_n)^c - K_t(t;\xi^*)^c|$ *is* $o_P(1)$ *for any* $0 < t_0 < t_1 < \infty$ *and for any* $c\in\mathbb{R}$. *If* $c \geqslant 0$, *we may take* $t_0 = 0$. *A similar statement holds also for* $\sup_{r\in[r_0,r_1]} |\widehat{K}_{s,n}(r;\breve{\beta}_n)^c - K_s(r;\psi^*)^c|$.

P r o o f. Let us first remark that the condition $t_0 > 0$ if $c < 0$ is needed to avoid division by 0, since $K(0;\xi^*) = 0$.

Due to (A1) and (A2), the intensity function λ_t is bounded and continuous as a function of β. Thus, it is possible to show the convergence $\widehat{K}_{t,n}(t;\breve{\beta}_n) - \widehat{K}_{t,n}(t;\beta^*) \to 0$ in probability as $n \to \infty$ for any $t \geqslant 0$. By Lemma 3 we get $\widehat{K}_{t,n}(t;\beta^*) \to K_t(t;\xi^*)$ in probability for any $t \geqslant 0$ and hence also $\widehat{K}_{t,n}(t;\breve{\beta}_n) \to K_t(t;\xi^*)$ in probability as $n \to \infty$ for any $t \geqslant 0$. Using monotonicity of $\widehat{K}_{t,n}(t;\breve{\beta}_n)^c$ and $K_t(t;\xi^*)^c$ as functions of t, the result follows by arguments similar to those in the proof of the Glivenko-Cantelli theorem, see, e.g. [20], p. 266.

The same type of argument can be used to prove the second part of the lemma. □

Lemma 5. *Under the conditions of Theorem* 1, *the following assertions hold:*

(a) $|W_n \times T_n|^{-1} J_{n,11}(\beta^*) - I_{n,11} \to 0$ *in probability as* $n \to \infty$,

(b) $|T_n|^{-1} J_{n,22}(\widehat{\beta}_n, \xi^*) - I_{22} \to 0$ *in probability as* $n \to \infty$,

(c) $|W_n|^{-1} J_{n,33}(\widehat{\beta}_n, \psi^*) - I_{33} \to 0$ *in probability as* $n \to \infty$.

Proof. (a) By the Campbell theorem, $|W_n \times T_n|^{-1} J_{n,11}(\beta^*) - I_{n,11}$ has mean 0. Using assumptions (A2) and (A8), one can show that the variance of each element of the matrix in question is $\mathcal{O}(|W_n \times T_n|^{-1})$. This implies the required convergence in probability.

(b) Note that $|T_n|^{-1} J_{n,22}(\widehat{\beta}_n, \xi^*) = I_{22} - V_n$, where

$$V_n = 2c_2 \int_{t_0}^{t_1} [\widehat{K}_{t,n}(t;\widehat{\beta}_n)^{c_2} - K_t(t;\xi^*)^{c_2}] \\ \times [(c_2-1)K_t(t;\xi^*)^{c_2-2} K_t^{(1)}(t;\xi^*)^{\mathrm{T}} K_t^{(1)}(t;\xi^*) + K_t(t;\xi^*)^{c_2-1} K_t^{(2)}(t;\xi^*)]\,\mathrm{d}t.$$

It is now sufficient to show that $V_n \to 0$ in probability. Denote

$$s_n(c) = \sup_{t\in[t_0,t_1]} |\widehat{K}_{t,n}(t;\widehat{\beta}_n)^c - K_t(t;\xi^*)^c|.$$

Then we can write (using any matrix norm)

$$\|V_n\| \leqslant 2c_2 s_n(c_2) \int_{t_0}^{t_1} \|(c_2-1)K_t(t;\xi^*)^{c_2-2} K_t^{(1)}(t;\xi^*)^{\mathrm{T}} K_t^{(1)}(t;\xi^*) \\ + K_t(t;\xi^*)^{c_2-1} K_t^{(2)}(t;\xi^*)\|\,\mathrm{d}t \leqslant \text{const.}\ (t_1 - t_0) s_n(c_2),$$

since the integrand can be bounded from above by assumptions (A2), (A6), and (A7). By Lemma 4, $s_n(c_2) \to 0$ in probability and thus $\|V_n\| \to 0$ in probability as $n \to \infty$. This concludes the proof.

(c) The proof follows the arguments in (b), starting with $|W_n|^{-1} J_{n,33}(\widehat{\beta}_n, \psi^*) = I_{33} - \widetilde{V}_n$, where $\widetilde{V}_n$ is the remainder term, and finishing with the use of Lemma 4. □

Appendix B. Proof of Theorem 4

The proofs of Theorems 2–4 follow the same lines of reasoning. Hence, we give the details of the proof only for Theorem 4—it is more technically complicated than the proof of Theorem 2 and essentially equivalent to the proof of Theorem 3. At the same time it illustrates the necessity of using the central limit theorem for random fields.

Consider the Taylor series expansion

$$\begin{aligned}(U_{n,1}(\beta^*), U_{n,3}(\beta^*,\psi^*)) &= (U_{n,1}(\widehat{\beta}_n), U_{n,3}(\widehat{\beta}_n,\widehat{\psi}_n)) \\ &\quad + \{(\widehat{\beta}_n,\widehat{\psi}_n) - (\beta^*,\psi^*)\}\begin{pmatrix} J_{n,11}(\widetilde{\beta}_n) & J_{n,13}(\widetilde{\beta}_n,\widetilde{\psi}_n) \\ 0 & J_{n,33}(\widetilde{\beta}_n,\widetilde{\psi}_n)\end{pmatrix},\end{aligned}$$

where $(\widetilde{\beta}_n,\widetilde{\psi}_n)$ is between $(\widehat{\beta}_n,\widehat{\psi}_n)$ and (β^*,ψ^*). We focus on the second part of the vector equation:

$$U_{n,3}(\beta^*,\psi^*) = U_{n,3}(\widehat{\beta}_n,\widehat{\psi}_n) + (\widehat{\beta}_n - \beta^*)J_{n,13}(\widetilde{\beta}_n,\widetilde{\psi}_n) + (\widehat{\psi}_n - \psi^*)J_{n,33}(\widetilde{\beta}_n,\widetilde{\psi}_n).$$

We multiply both sides of the equation from the right by $|W_n|^{-1/2}\widetilde{\Sigma}_{n,33}^{-1/2}$ and discuss each term separately.

On the left-hand side, the term $|W_n|^{-1/2}U_{n,3}(\beta^*,\psi^*)\widetilde{\Sigma}_{n,33}^{-1/2}$ depends only on the true parameter values and we will show below that it has asymptotically a standard normal distribution. Moving on to the right-hand side, $|W_n|^{-1/2}U_{n,3}(\widehat{\beta}_n,\widehat{\psi}_n)\widetilde{\Sigma}_{n,33}^{-1/2}$ equals 0 with probability tending to 1 by Theorem 1.

We rewrite the remaining terms as follows:

$$\begin{aligned}|W_n|^{-1/2}(\widehat{\beta}_n - \beta^*)J_{n,13}(\widetilde{\beta}_n,\widetilde{\psi}_n)\widetilde{\Sigma}_{n,33}^{-1/2} &= |W_n|^{1/2}(\widehat{\beta}_n - \beta^*)\frac{J_{n,13}(\widetilde{\beta}_n,\widetilde{\psi}_n)}{|W_n|}\widetilde{\Sigma}_{n,33}^{-1/2}, \\ |W_n|^{-1/2}(\widehat{\psi}_n - \psi^*)J_{n,33}(\widetilde{\beta}_n,\widetilde{\psi}_n)\widetilde{\Sigma}_{n,33}^{-1/2} &= |W_n|^{1/2}(\widehat{\psi}_n - \psi^*)\frac{J_{n,33}(\widetilde{\beta}_n,\widetilde{\psi}_n)}{|W_n|}\widetilde{\Sigma}_{n,33}^{-1/2}.\end{aligned}$$

Under the assumptions of the theorem, one can show that

- ▷ $|W_n|^{1/2}(\widehat{\beta}_n - \beta^*)$ converges to 0 in probability, since $|W_n \times T_n|^{1/2}(\widehat{\beta}_n - \beta^*)$ is bounded in probability by Theorem 1;
- ▷ $|W_n|^{-1}J_{n,13}(\widetilde{\beta}_n,\widetilde{\psi}_n) - I_{n,13}$ converges to 0 in probability, by using similar continuity arguments as in the proof of Theorem 1;
- ▷ the elements of the matrix $I_{n,13}$ are bounded.

Also, by assumption (A4), $\widetilde{\Sigma}_{n,33}$ converges to a positive definite matrix $\widetilde{\Sigma}_{33}$. We conclude that the whole term $|W_n|^{1/2}(\widehat{\beta}_n - \beta^*)J_{n,13}(\widetilde{\beta}_n,\widetilde{\psi}_n)|W_n|^{-1}\widetilde{\Sigma}_{n,33}^{-1/2}$ converges to 0 in probability as $n \to \infty$. Hence, this term does not affect the limiting distribution of $|W_n|^{1/2}(\widehat{\psi}_n - \psi^*)$.

The next step is to show that $|W_n|^{-1} J_{n,33}(\widetilde{\beta}_n, \widetilde{\psi}_n) - I_{33}$ converges to 0 in probability. This can be done by showing that $|W_n|^{-1}(J_{n,33}(\widetilde{\beta}_n, \widetilde{\psi}_n) - J_{n,33}(\beta^*, \psi^*)) \to 0$ in probability by using continuity arguments similar to those used above for checking condition (c) of Lemma 2, and then showing that $|W_n|^{-1} J_{n,33}(\beta^*, \psi^*) - I_{33} \to 0$ in probability as $n \to \infty$, similarly to part (c) of Lemma 5.

It remains to show asymptotic normality for $|W_n|^{-1/2} U_{n,3}(\beta^*, \psi^*)\widetilde{\Sigma}_{n,33}^{-1/2}$. Consider the approximation

$$|W_n|^{-1/2} U_{n,3}(\beta^*, \psi^*) = |W_n|^{-1/2} \widetilde{U}_{n,3}(\beta^*, \psi^*) + |W_n|^{-1/2} V_{n,3}(\beta^*, \psi^*),$$

where $\widetilde{U}_{n,3}$ is defined at the beginning of Section 6. Under the assumptions of the theorem, it can be shown that $|W_n|^{-1/2} V_{n,3}(\beta^*, \psi^*)$ converges to 0 in probability and hence the limiting distribution of $|W_n|^{-1/2} U_{n,3}(\beta^*, \psi^*)\widetilde{\Sigma}_{n,33}^{-1/2}$ is the same as the limiting distribution of $|W_n|^{-1/2} \widetilde{U}_{n,3}(\beta^*, \psi^*)\widetilde{\Sigma}_{n,33}^{-1/2}$. We will focus on the latter.

Let $h > 0$ be as in assumption (D2) and define $A_{ij} = [ih, (i+1)h) \times [jh, (j+1)h)$, $(i, j) \in \mathbb{Z}^2$. In $\widetilde{U}_{n,3}(\beta^*, \psi^*)$ we may replace $\widehat{K}_{s,n}(r; \beta^*)$ by

$$\frac{1}{|W_n|} \sum_{u \in X_s \cap W_n} \sum_{v \in X_s} \frac{I(0 < \|u - v\| \leqslant r)}{\lambda_s(u; \beta^*)\lambda_s(v; \beta^*)}$$

and denote

$$\begin{aligned} Y_{ij} = 2c_3^2 &\sum_{u \in X_s \cap A_{ij}} \int_{r_0}^{r_1} \sum_{v \in X_s} \frac{I(0 < \|u - v\| \leqslant r)}{\lambda_s(u; \beta^*)\lambda_s(v; \beta^*)} \\ &\times K_s(r; \psi^*)^{2c_3 - 2} K_s^{(1)}(r; \psi^*)\, \mathrm{d}r - 2c_3^2 h^2 \int_{r_0}^{r_1} K_s(r; \psi^*)^{2c_3 - 1} K_s^{(1)}(r; \psi^*)\, \mathrm{d}r. \end{aligned}$$

Then it can be shown that

$$|W_n|^{-1/2} \widetilde{U}_{n,3}(\beta^*, \psi^*) = |W_n|^{-1/2} \sum_{(i,j) \in \mathbb{Z}^2 :\ A_{ij} \subseteq W_n} Y_{ij} + o_P(1).$$

The remainder term corresponds to those A_{ij} which hit the boundary of W_n. By our assumption, the size of the boundary grows at a slower rate than the volume of W_n. This is the key ingredient for showing that the remainder term is in fact $o_P(1)$, i.e. converges to 0 in probability.

We aim at using the Cramér-Wold theorem. To do this, we take an arbitrary non-zero vector y of the appropriate dimension and set

$$Z_{ij} = Y_{ij} y^{\mathrm{T}}, \qquad \sigma_n^2 = |W_n|^{-1} \operatorname{Var}\Bigg(\sum_{(i,j) \in \mathbb{Z}^2 :\ A_{ij} \subseteq W_n} Z_{ij} \Bigg) = y \widetilde{\Sigma}_{n,33} y^{\mathrm{T}} + o(1).$$

In this way we construct a random field $\{Z_{ij}\}$ defined on the integer lattice $\mathbb{Z}^2$. We will show below that $(\sigma_n^2|W_n|)^{-1/2} \sum_{(i,j)\in\mathbb{Z}^2:\, A_{ij}\subseteq W_n} Z_{ij}$ is asymptotically standard normal. Together with (A4) this implies that $|W_n|^{-1/2} \sum_{(i,j)\in\mathbb{Z}^2:\, A_{ij}\subseteq W_n} Z_{ij}$ converges in distribution to a normally distributed random variable with mean 0 and variance $y\widetilde{\Sigma}_{33}y^{\mathrm{T}}$. Assumption (A4) and the Cramér-Wold theorem then imply that $|W_n|^{-1/2}\widetilde{U}_{n,3}(\beta^*,\psi^*)\widetilde{\Sigma}_{n,33}^{-1/2}$ is asymptotically standard normal.

To show that $(\sigma_n^2|W_n|)^{-1/2} \sum_{(i,j)\in\mathbb{Z}^2:\, A_{ij}\subseteq W_n} Z_{ij}$ is asymptotically standard normal we use the classical central limit theorem for random fields on a lattice [8], Thm. 3.3.1, with the additional assumption of uniform integrability, see also the discussion in [11]. Namely, the following conditions must be satisfied for some $\delta > 0$:

(a) $\liminf \sigma_n^2 > 0$,

(b) $|Z_{ij}|^{2+\delta}$ are uniformly integrable,

(c) $\sum_{m=1}^{\infty} m\alpha^F_{s,2,\infty}(m)^{\delta/(2+\delta)} < \infty$.

Condition (a) is fulfilled by assumption (A4) and condition (b) follows from the moment assumption (C1). Finally, the mixing condition (c) is implied by (D2). This completes the proof.

References

[1] *A. J. Baddeley, J. Møller, R. Waagepetersen*: Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. Stat. Neerl. *54* (2000), 329–350.

[2] *D. J. Daley, D. Vere-Jones*: An Introduction to the Theory of Point Processes. Vol. II: General Theory and Structure. Probability and Its Applications, Springer, New York, 2008.

[3] *P. J. Diggle*: Spatio-temporal point processes: methods and applications. Statistical Methods for Spatio-temporal Systems (B. Finkenstädt, et al., eds.). Selected Invited Papers Based on the Presentations at the 6th Séminaire Européen de Statistique SemStat Held as a Summer School of the European Mathematical Society, Bernried, 2004, Chapman and Hall/CRC, Boca Raton, 2007, Monographs on Statistics and Applied Probability *107* (2007), 1–45.

[4] *P. Doukhan*: Mixing: Properties and Examples. Lecture Notes in Statistics 85, Springer, New York, 1994.

[5] *J. Dvořák, M. Prokešová*: Parameter estimation for inhomogeneous space-time shot-noise Cox point processes. To appear in Scand. J. Stat.

[6] *E. Gabriel*: Estimating second-order characteristics of inhomogeneous spatio-temporal point processes. Methodol. Comput. Appl. Probab. *16* (2014), 411–431.

[7] *E. Gabriel, P. J. Diggle*: Second-order analysis of inhomogeneous spatio-temporal point process data. Stat. Neerl. *63* (2009), 43–51.

[8] *X. Guyon*: Random Fields on a Network. Modeling, Statistics, and Applications. Probability and Its Applications, Springer, New York, 1995.

[9] *W. W. Hager*: Minimizing a quadratic over a sphere. SIAM J. Optim. *12* (2001), 188–208.

[10] *G. Hellmund, M. Prokešová, E. B. V. Jensen*: Lévy-based Cox point processes. Adv. Appl. Probab. *40* (2008), 603–629.

[11] *Z. Karácsony*: A central limit theorem for mixing random fields. Miskolc Math. Notes *7* (2006), 147–160.
[12] *Th. Motzkin*: From among n conjugate algebraic integers, $n-1$ can be approximately given. Bull. Am. Math. Soc. *53* (1947), 156–162.
[13] *J. Møller*: Shot noise Cox processes. Adv. Appl. Probab. *35* (2003), 614–640.
[14] *J. Møller, M. Ghorbani*: Aspects of second-order analysis of structured inhomogeneous spatio-temporal point processes. Stat. Neerl. *66* (2012), 472–491.
[15] *J. Møller, R. P. Waagepetersen*: Statistical Inference and Simulation for Spatial Point Processes. Monographs on Statistics and Applied Probability 100, Chapman & Hall/CRC, Boca Raton, 2004.
[16] *M. Prokešová, J. Dvořák*: Statistics for inhomogeneous space-time shot-noise Cox processes. Methodol. Comput. Appl. Probab. *16* (2014), 433–449.
[17] *M. Prokešová, J. Dvořák, E. B. V. Jensen*: Two-step estimation procedures for inhomogeneous shot-noise Cox processes. To appear in Ann. Inst. Stat. Math.
[18] *B. D. Ripley*: Statistical Inference for Spatial Processes. Cambridge University Press, Cambridge, 1988.
[19] *D. Stoyan, W. S. Kendall, J. Mecke*: Stochastic Geometry and Its Applications. Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Chichester, 1995.
[20] *A. W. van der Vaart*: Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics 3, Cambridge University Press, Cambridge, 1998.
[21] *R. Waagepetersen, Y. Guan*: Two-step estimation for inhomogeneous spatial point processes. J. R. Stat. Soc., Ser. B, Stat. Methodol. *71* (2009), 685–702.

Authors' address: *Jiří Dvořák*, *Michaela Prokešová*, Faculty of Mathematics and Physics, Charles University in Prague, Sokolovská 83, CZ-186 75 Praha 8, Czech Republic, e-mail: `dvorak@karlin.mff.cuni.cz`, `prokesov@karlin.mff.cuni.cz`.