Ann Inst Stat Math (2017) 69:513–542
DOI 10.1007/s10463-016-0556-y

Two-step estimation procedures for inhomogeneous shot-noise Cox processes

Michaela Prokešová[1] · **Jiří Dvořák**[1] · **Eva B. Vedel Jensen**[2]

Received: 11 December 2014 / Revised: 27 November 2015 / Published online: 27 February 2016

Abstract In the present paper, we discuss and compare several two-step estimation procedures for inhomogeneous shot-noise Cox processes. The intensity function is parametrized by the inhomogeneity parameters while the pair-correlation function is parametrized by the interaction parameters. The suggested procedures are based on a combination of Poisson likelihood estimation of the inhomogeneity parameters in the first step and an adaptation of a method from the homogeneous case for estimation of the interaction parameters in the second step. The adapted methods, based on minimum contrast estimation, composite likelihood and Palm likelihood, are compared both theoretically and by means of a simulation study. The general conclusion from the simulation study is that the three estimation methods have similar performance. Two-step estimation with Palm likelihood has not been considered before and is motivated by the superior performance of the Palm likelihood in the stationary case for estimation of certain parameters of interest. Asymptotic normality of the two-step estimator with Palm likelihood is proved.

Keywords Shot-noise Cox processes · Inhomogeneous spatial point processes · Two-step estimation methods

✉ Michaela Prokešová
prokesov@karlin.mff.cuni.cz

[1] Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University in Prague, Sokolovská 83, Praha 8, 186 75, Czech Republic

[2] Department of Mathematics, Centre for Stochastic Geometry and Advanced Bioimaging, Aarhus University, Ny Munkegade 118, Bldg. 1530, 8000 Aarhus C, Denmark

1 Introduction

Cox point processes (sometimes also called doubly stochastic point processes) are the preferred point process models for analysis of clustered point patterns (Cox 1955; Matérn 1971; Daley and Vere-Jones 2003; Diggle 2003; Møller and Waagepetersen 2003, 2007; Daley and Vere-Jones 2008; Illian et al. 2008; Chiu et al. 2013). These processes are able to model clustering of different strength on different scales as well as inhomogeneity dependent on spatial covariates. As such they are used in a large spectrum of applications, e.g., in biology, ecology and epidemiology.

Spatial Cox point process models include two large classes—the log-Gaussian Cox processes and the shot-noise Cox processes. Since these two classes have somewhat different properties they are usually considered separately in the literature and their statistical inference is based on different methods (see e.g., Møller and Waagepetersen 2003). In the present paper, we will consider the shot-noise Cox processes and the problem of parameter estimation of inhomogeneous models coming from this class. The shot-noise Cox processes were introduced in Møller (2003) and further generalized in Hellmund et al. (2008) without discussing the statistical inference for this model class. Note that the class of shot-noise Cox processes includes the very popular Poisson Neyman–Scott processes such as the Thomas process (see e.g., Thomas 1949; Illian et al. 2008, Section 6.3.2).

Maximum likelihood estimation for these processes is computationally very intensive (even more so for inhomogeneous models) and involves the development of a special MCMC numerical algorithm for each particular model, see e.g., Møller and Waagepetersen (2007, Section 7.3) for an example. Therefore, the easier-to-compute moment estimation methods (even though less efficient than the maximum likelihood estimation) are often preferred in applications.

Several moment estimation methods applicable to the stationary shot-noise Cox processes are available in the literature: minimum contrast estimation (Diggle 1983, Chapter 6), composite likelihood (Guan 2006), Palm likelihood (Tanaka et al. 2007; Prokešová and Jensen 2013). According to simulation studies, as the ones presented in Guan (2006) and Dvořák and Prokešová (2012), the efficiency of the different estimators on middle-sized observation windows depends on the considered model and the parameter of main interest. There is no uniformly best estimator.

For the nonstationary case (which is much more interesting from an applied point of view) a two-step estimation procedure was introduced in Waagepetersen and Guan (2009) where first the first-order intensity function $\lambda(u)$ is estimated and then, conditionally on $\lambda(u)$, the inhomogeneous K-function is used for the minimum contrast estimation of the interaction parameters of the Cox process. In Guan (2009), the same two-step estimation procedure was investigated with minimum contrast based on the pair-correlation function in the second step. These two-step estimation procedures work for inhomogeneous spatial point processes which are second-order intensity-reweighted stationary (SOIRS). This model class, introduced in Baddeley et al. (2000), is characterized by a translation-invariant pair-correlation function. Accordingly, the second-order intensity function can for SOIRS processes be decomposed as

$$\lambda^{(2)}(u, v) = \lambda(u)\lambda(v)g(v - u).$$

However, this decomposition enables a generalization of the other estimation methods from the stationary case to the SOIRS case as well. In the recent paper (Jalilian et al. 2013), two-step composite likelihood was discussed.

In the present paper, we investigate the above-mentioned two-step estimation procedures for SOIRS inhomogeneous shot-noise Cox processes, including conditions for the validity of asymptotic results for these two-step estimation procedures. Further, we generalize the Palm likelihood estimation to a two-step estimation procedure for SOIRS inhomogeneous Cox processes and derive conditions for consistency and asymptotic normality of the estimators. Finally, we compare the efficiency of all the considered two-step estimation procedures on middle-sized observation windows in a simulation study.

The paper is organized as follows. Basic notions relating to spatial point processes are given in Sect. 2, while shot-noise Cox processes are introduced in Sect. 3. An overview of moment estimation methods for stationary Cox processes is given in Sect. 4. These methods are adapted to the inhomogeneous case in Sect. 5. In Sect. 6, the focus is on two-step estimation with Palm likelihood and in Sect. 7 asymptotic normality of this two-step estimator with Palm likelihood is proved. The performance of the developed two-step estimation methods is compared in a simulation study presented in Sect. 8. Conclusions and perspectives are found in Sect. 9.

2 Background

In this section, we briefly introduce the notation relating to spatial point processes used in the following. For more detailed information, see standard references such as Daley and Vere-Jones (2008) and Chiu et al. (2013).

Let $\mathscr{B}(\mathbb{R}^d) = \mathscr{B}^d$ be the Borel subsets of $\mathbb{R}^d$. Let X be a point process on $\mathscr{X} \in \mathscr{B}^d$. For $A \in \mathscr{B}^d$, $|A|$ will denote the volume of A and $|X \cap A|$ the number of points from X in A (we use the notation $|\cdot|$ for the suitable Hausdorff measure of the set). For $R > 0$, $B(o, R)$ is the ball centered at the origin o with radius R and $A \oplus R = \bigcup_{x \in A} B(x, R)$. The Euclidean norm of the vector $x \in \mathbb{R}^d$ is denoted by $\|x\|$ and I is the indicator function.

Throughout this paper, we consider point processes for which the intensity function λ and the second-order intensity function $\lambda^{(2)}$ exist. For a stationary point process X, the intensity function is constant, $\lambda(u) = \lambda$, say, and the second-order intensity function can be decomposed as

$$\lambda^{(2)}(u, v) = \lambda^{(2)}(0, v - u) = \lambda\lambda_o(v - u). \tag{1}$$

The function λ_o is the (first-order) intensity function of the Palm distribution of X, sometimes called the Palm intensity. The Palm distribution may be interpreted as the distribution of X conditioned by the occurrence of a point from X at the origin.

The interaction between points may be described by the pair-correlation function

$$g(u, v) = \frac{\lambda^{(2)}(u, v)}{\lambda(u)\lambda(v)},$$

also called the g-function. For stationary point processes, the g-function is translation invariant, $g(u, v) = g(v - u)$, say. Alternatively, the interaction between points in a stationary point process may be described by Ripley's K-function. For $A \in \mathscr{B}^d$ with $0 < |A| < \infty$, the K-function satisfies

$$\lambda K(r) = \frac{1}{\lambda|A|} \mathsf{E} \sum_{u \in X \cap A} \sum_{v \in X \setminus \{u\}} I(\|u - v\| \leq r), \quad r > 0. \tag{2}$$

Note that the right-hand side of (2) does not depend on A. Since $\lambda|A|$ is the mean number of points in A, $\lambda K(r)$ may be interpreted as the mean number of further points of the point process at distance at most r from a typical point of the point process. The g- and K- functions are related in the following way

$$K(r) = \int_{B(o,r)} g(u)\,\mathrm{d}u, \quad r > 0. \tag{3}$$

In the following, we will also consider inhomogeneous point processes that are second-order intensity-reweighted stationary (SOIRS); see Baddeley et al. (2000). These processes are characterized by a translation-invariant g-function, but they may have a nonconstant intensity function. They can be obtained by location-dependent thinning of a stationary process. Under SOIRS, we can decompose $\lambda^{(2)}$ as follows:

$$\lambda^{(2)}(u, v) = \lambda(u)\lambda(v)g(v - u) = \lambda(u)\lambda_u(v) = \lambda(v)\lambda_v(u), \tag{4}$$

where $\lambda_u(v)$ is the intensity function at v of the Palm distribution of X conditioned by the event that a point of X occurs at location u. This possibility of decomposing $\lambda^{(2)}$ in a multiplicative way will be important for the estimation procedures developed in Sects. 5 and 6.

In the analysis of SOIRS processes, the so-called inhomogeneous K-function is used. This function is defined by the relation (3) used in the stationary case.

3 Shot-noise Cox processes

The focus of this paper is on shot-noise Cox processes driven by a random field of the form

$$\Lambda(u) = \sum_{(r,v) \in \Pi_\Upsilon} r k(u, v), \quad u \in \mathscr{X}, \tag{5}$$

where Π_Υ is a Poisson process on $\mathbb{R}^+ \times \mathbb{R}^d$ with intensity measure Υ and k is a smoothing kernel, i.e., a non-negative function integrable in both coordinates; see Møller (2003) and Hellmund et al. (2008) for further details.

The shot-noise Cox process X is stationary if the kernel k is translation invariant $k(u, v) = k(v - u)$ and the measure Υ has the form $\Upsilon(\mathrm{d}(r, v)) = \mu V(\mathrm{d}r)\,\mathrm{d}v$, where $\mu > 0$ and $V(\mathrm{d}r)$ is an arbitrary measure on $\mathbb{R}^+$ satisfying the integrability assumption $\int_{\mathbb{R}^+} \min(1, r) V(\mathrm{d}r) < \infty$. A large variety of models may be obtained according to the choice of V. The popular class of Poisson cluster processes is obtained when V is

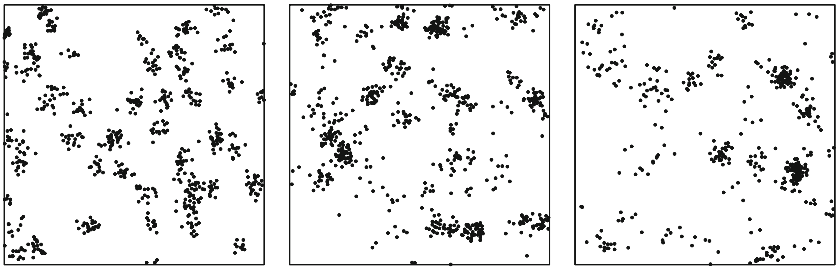

Fig. 1 Realizations of (from *left* to *right*) a stationary Thomas process, a stationary gamma shot-noise Cox process and an inhomogeneous gamma shot-noise Cox process. For details, see the *text*

equal to the Dirac measure, $V(\mathrm{d}r) = \delta_1(\mathrm{d}r)$. Figure 1, left panel, shows a realization of such a Poisson cluster process with Gaussian kernel. This process is a stationary Thomas process (Thomas 1949).

All shot-noise Cox processes can be viewed as generalized cluster processes. The measure V determines the distribution of the number of points in the clusters. By choosing an appropriate measure V, we can obtain very variable number of points in the clusters.

Example 1 (Gamma shot-noise Cox process) Let $V(\mathrm{d}r) = r^{-1}\exp(-\theta r)\,\mathrm{d}r$, where $\theta > 0$ is a parameter. Note that V is not integrable in the neighborhood of 0. As a consequence, the corresponding shot-noise Cox process X is not a cluster process in the classical sense (Illian et al. 2008, Section 6.3) since the number of "clusters" in any compact set is infinite. However, because the weights of the majority of the clusters are very small, X is still a well-defined Cox process. The name gamma shot-noise Cox process refers to the fact that V is the Lévy measure of a gamma distributed random variable (Hellmund et al. 2008, Section 4). Figure 1, middle panel, shows a realization of a stationary gamma shot-noise Cox process. The point process has the same Gaussian kernel k and intensity as the Thomas process in Fig. 1, left panel, but has clearly larger variability in the cluster sizes.

The moment properties of the shot-noise Cox processes are easily available (Hellmund et al. 2008, Section 4). In particular, for the intensity function we have

$$\lambda(u) = \mu \int_{\mathbb{R}^+} r\, V(\mathrm{d}r) \int_{\mathbb{R}^d} k(u, v)\,\mathrm{d}v,$$

and for the pair-correlation function

$$g(u, v) = 1 + \frac{\mu \int_{\mathbb{R}^+} r^2 V(\mathrm{d}r) \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} k(u, w)k(v, w)\,\mathrm{d}w}{\lambda(u)\,\lambda(v)}.$$

Note that in both equations, a product of separate integrals for V and k appears—this will be important in the estimation procedures developed in Sects. 5 and 6. Moreover, for parametric forms of V such as $V(\mathrm{d}r) = r^{-1}\exp(-\theta r)\,\mathrm{d}r$, both integrals with respect to V are simple functions of the parameters, $\int_{\mathbb{R}+} r\, V(\mathrm{d}r) = 1/\theta$ and $\int_{\mathbb{R}+} r^2 V(\mathrm{d}r) = 1/\theta^2$.

If we apply location-dependent thinning to a stationary shot-noise Cox process specified by μ, V and k, a new shot-noise Cox process is obtained with the same μ and V, but with different kernel function. The process is second-order intensity-reweighted stationary (SOIRS). An example of a realization of such a process is shown in Fig. 1, right panel. It is obtained by location-dependent thinning of the point process shown in Fig. 1, middle panel. For a more detailed description of the process, see Sect. 8.

4 Estimation in the stationary case

In this section we give an overview of the moment estimation methods for the stationary Cox process models, available in the literature. All of them are based on the second-order intensity function $\lambda^{(2)}$ or on characteristics derived from this function.

Let W denote a compact observation window on which we observe the point process X. We will assume a parametric model for X. The vector of unknown parameters will be denoted by η. Particularly, we assume that the stationary Cox point process X is characterized by its second-order intensity function $\lambda^{(2)}(\cdot;\eta)$ (or by some other equivalent characteristic like K, g or λ_o). As explained in the previous section, these characteristics are for many shot-noise Cox process models available in a reasonably tractable form as functions of the parameter η and thus the maximization of the respective estimation criteria is numerically feasible.

4.1 Minimum contrast

This estimation method was in the context of spatial statistics described as early as in Diggle (1983, Chapter 5). It can be based either on the K-function or the pair-correlation function g; see e.g., Diggle (2003, Chapter 6). In the version based on the g-function it is required that the process X is isotropic as well as stationary. Under isotropy, the g-function is a function of a scalar argument.

The vector of parameters η is estimated by minimizing the discrepancy measure

$$\int_r^R \left[\hat{K}^q(u) - K^q(u;\eta)\right]^2 \mathrm{d}u \quad \text{or} \quad \int_r^R \left[\hat{g}^q(u) - g^q(u;\eta)\right]^2 \mathrm{d}u \tag{6}$$

between the nonparametric estimate $\hat{K}$ or $\hat{g}$ and the theoretical value $K(\cdot;\eta)$ or $g(\cdot;\eta)$, respectively.

The constants q, r and R are used to control the sampling fluctuations in the estimators of K and g. Recommendations concerning the choice of tuning parameters and other practical aspects can be found in Diggle (2003, Section 6.1.1). Asymptotic properties of the minimum contrast estimator, based on the K-function, are discussed in Heinrich (1992) and Guan and Sherman (2007) for the stationary case. In Heinrich (1992) strong consistency and asymptotic normality for minimum contrast estimators, based on the K-function, was proved for stationary Poisson cluster processes.

In Guan and Sherman (2007) asymptotic normality for minimum contrast estimators, based on the K-function, was shown for stationary processes, fulfilling a strong mixing assumption.

4.2 Composite likelihood

The composite likelihood approach is a general statistical methodology (Lindsay 1988). In the context of point processes it is based on adding together individual log-likelihoods for single points or pairs of points of the process X to form a composite log-likelihood. Several versions of composite likelihood have been suggested for estimation of different types of spatial point processes (Baddeley et al. 2000; Guan 2006; Møller and Waagepetersen 2007). Composite likelihood suitable for estimation of Cox processes was introduced in Guan (2006). It uses the second-order intensity function $\lambda^{(2)}(\cdot;\eta)$ to obtain the probability density for two points of X occurring at locations x and y

$$f(x,y;\eta)=\frac{\lambda^{(2)}(y-x;\eta)}{\int_W\int_W\lambda^{(2)}(u-v;\eta)\mathrm{d}u\mathrm{d}v}. \tag{7}$$

After adding the individual log-likelihoods, the composite log-likelihood is obtained

$$\begin{aligned}\log CL(\eta)=\sum_{x,y\in X\cap W,\,0<\|y-x\|<R}\Bigg[&\log\lambda^{(2)}(y-x;\eta)\\&-\log\left(\int_W\int_W\lambda^{(2)}(u-v;\eta)I(\|u-v\|<R)\,\mathrm{d}u\,\mathrm{d}v\right)\Bigg].\end{aligned} \tag{8}$$

Here, only pairs of points with distance less than R are considered. Disregarding the pairs of points separated by distance R or larger is motivated by the fact that pairs of points far apart are often nearly independent. They do not carry much information about the parameter η, but increase the variability of the estimator. Consistency and asymptotic normality of the composite likelihood estimator in the stationary case are proved in Guan (2006) under suitable mixing assumptions.

Note that in the stationary case the squared intensity λ^2 cancels out in (7) so that

$$f(x,y;\eta)=\frac{g(y-x;\eta)}{\int_W\int_W g(u-v;\eta)\mathrm{d}u\mathrm{d}v},$$

and (8) can be used with g instead of $\lambda^{(2)}$.

4.3 Palm likelihood

The Palm likelihood estimator for isotropic stationary point processes was introduced in Tanaka et al. (2007) and uses a very "geometrical" approach. It is based on the process of differences between the points of the observed point process X. Let

$$Y(R) = \{y - x : x \neq y \in X \cap W, \|y - x\| < R\}$$

be the point process of differences of points in X observed on W with mutual distance smaller than R. Evidently, $Y(R)$ is a point process contained in $B(o, R)$. The intensity function of this point process can be derived as follows. Let A be a Borel subset of $B(o, R)$. Then,

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

where $\gamma_W(u) = |W \cap (W + u)|$ is the set covariance of the window W, see Chiu et al. (2013, p. 17) for further details. The point process $Y(R)$ has thus an intensity function concentrated on $B(o, R)$ of the form

$$\lambda_R(u) = \gamma_W(u)\lambda\lambda_o(u; \eta), \quad u \in B(o, R).$$

The Palm log-likelihood

$$\log L_P(\eta) = \sum_{\substack{x \neq y \in X \cap W \\ (y-x) \in B(o,R)}} \log\left(|X \cap W|\lambda_o(y - x; \eta)\right) - |X \cap W| \int_{B(o,R)} \lambda_o(r; \eta)\,\mathrm{d}r \tag{9}$$

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

An alternative way of arriving at the Palm likelihood goes as follows. Let

$$Y_x = \{y - x, x \neq y \in X\}, \; x \in X \cap W.$$

Each Y_x is an inhomogeneous point process with intensity function equal to the Palm intensity $\lambda_o(\cdot; \eta)$ of the original process X. Ignoring the interactions in the process Y_x, i.e., approximating Y_x by a Poisson process, the log-likelihood of $Y_x \cap B(o, R)$ is (up to a constant) the following:

$$\sum_{y \in X \cap W, 0 < ||x - y|| < R} \log \lambda_o(x - y; \eta) - \int_{\mathbb{R}^d} I(||u|| < R)\lambda_o(u; \eta)\mathrm{d}u.$$

By treating all the Y_x, $x \in X \cap W$, as independent, identically distributed replications (and ignoring the edge effects caused by a bounded observation window W), we can sum the individual log-likelihoods over $x \in X \cap W$ and get an equivalent version of the Palm log-likelihood

$$\log L_P(\eta) = \sum_{x \neq y \in X \cap W, ||x-y|| < R} \log \lambda_o(x-y;\eta) - |X \cap W| \int_{B(o,R)} \lambda_o(r;\eta)\,\mathrm{d}r. \quad (10)$$

Note that even though the Palm likelihood estimation was derived by using the process of differences, it is a second-order moment method because it is based on the second-order characteristic λ_o of the observed point process X. Strong consistency and asymptotic normality of the Palm likelihood estimator are proved for stationary Cox processes in Prokešová and Jensen (2013) under suitable mixing assumptions.

5 Estimation in the inhomogeneous case

For the inhomogeneous (nonstationary) point processes the methods reviewed in the previous section cannot be used directly. Nevertheless, under the SOIRS assumption they can be adapted to the inhomogeneous case due to the product structure (4) of $\lambda^{(2)}$.

Following these ideas, Waagepetersen and Guan (2009) introduced for SOIRS point processes a two-step estimation procedure where first the first-order intensity function $\lambda(u)$ is estimated and then, conditionally on $\lambda(u)$, the inhomogeneous K-function is used in a minimum contrast estimation of the interaction parameters of the Cox process. Alternatively, Guan (2009) used minimum contrast estimation with the pair-correlation function g in the second step.

The minimum contrast estimation based on the K-function (MCK) is definitely the most frequently used method in the stationary case, but this method is actually not necessarily the most efficient. Simulation studies in Guan (2006) and Dvořák and Prokešová (2012) show that in many cases, minimum contrast estimation with the g-function (MCg) is superior to MCK.

In some cases, composite likelihood estimation (CL) is more efficient than any of the MC methods for estimation of interaction parameters, such as the scale of the kernel function in the cluster process. This applies in particular to cases where the total number of points observed in different clusters is very variable. Examples are log-Gaussian Cox processes with exponential correlation kernel or shot-noise Cox processes with nonatomic shape measure V. On the other hand, Palm likelihood is often superior to any other method when estimating the parameter μ for a Thomas process.

Since MCK and MCg are in the stationary case in some situations inferior to CL or PL it is natural also to consider the two-step estimators for the inhomogeneous case based on the CL and PL methods. Composite likelihood was done in a recent paper (Jalilian et al. 2013). For the Palm likelihood, we will introduce the new two-step estimator in Sect. 6.

In the remaining part of this section, we review the estimation of the inhomogeneity parameters in the first step and of the interaction parameters in the second step by minimum contrast estimation or composite likelihood estimation.

Throughout the section, X will be a SOIRS Cox process with second-order intensity function of the form

$$\lambda^{(2)}(u, v) = \lambda_\beta(u)\lambda_\beta(v)g_\eta(v - u).$$

Here, $\eta \in \mathbb{R}^q$ is a vector of interaction parameters that parametrizes the pair-correlation function g and $\beta \in \mathbb{R}^t$ is the vector of inhomogeneity parameters that parametrizes the first-order intensity function $\lambda(u)$. Thus, the full model is parametrized by $\psi = (\beta, \eta) \in \Psi \subset \mathbb{R}^{t+q}$, and we assume that it is possible to separate the inhomogeneity and interaction parameters, so that the model is not overspecified. Below, we give an example of such a model.

Example 2 Let X be the stationary gamma shot-noise Cox process in $\mathbb{R}^2$ with parameters $\mu, \theta > 0$ and smoothing kernel density k equal to the bivariate Gaussian density

$$k_{\sigma^2}(u) = \frac{1}{2\pi\sigma^2} \exp\left(\frac{-\|u\|^2}{2\sigma^2}\right), \quad u \in \mathbb{R}^2.$$

Suppose we observe X in a compact window W. Furthermore, let $h_{\beta^*}(u)$ be a nonconstant function, parametrized by the vector parameter $\beta^* = (\beta_1, \ldots \beta_{t-1})$, and let each point x of the process X be independently thinned with the probability $\frac{h_{\beta^*}(x)}{\max_{v \in W} h_{\beta^*}(v)}$. If we let $\beta_0 = \log(\frac{\mu}{\theta} / \max_{v \in W} h_{\beta^*}(v))$ and $\beta = (\beta_0, \beta^*)$, then the intensity function of the resulting inhomogeneous shot-noise Cox process Y is parametrized by β and takes the form

$$\lambda_\beta(u) = \exp(\beta_0)h_{\beta^*}(u).$$

The pair-correlation function is unchanged by the thinning

$$g_{(\sigma,\mu)}(v - u) = 1 + \frac{1}{4\pi\sigma^2\mu} \exp\left(\frac{-\|v - u\|^2}{4\sigma^2}\right)$$

and parametrized by the interaction parameter $\eta = (\sigma, \mu)$. In applications, a log-linear form of the intensity is often used

$$\lambda_\beta(u) = \exp(z(u)\beta^T), \quad u \in W,$$

where $z(u)$ is a vector of covariates observed at the location u.

The two-step estimation procedure in Waagepetersen and Guan (2009) can be described as follows. At first, the inhomogeneity parameter β is estimated by disregarding the interactions in the model, using the Poisson log-likelihood

$$\log L_1(\beta) = \sum_{x \in X \cap W} \log \lambda_\beta(x) - \int_W \lambda_\beta(u)\,\mathrm{d}u \tag{11}$$

only. The value $\hat{\beta}$ at which L_1 attains its maximal value is then taken to be the estimate of β.

In the second step, the interaction parameters η are estimated with the intensity function $\hat{\lambda} = \lambda_{\hat{\beta}}$ taken as fixed. The inhomogeneous K-function can be estimated by

$$\hat{K}(r) = \sum_{x,y \in W \cap X} \frac{I(0 < \|x-y\| < r)}{\hat{\lambda}(x)\hat{\lambda}(y)} w_{x,y}, \qquad r > 0,$$

where $w_{x,y}$ is an edge correction weight (see Baddeley et al. 2000). Analogously, it is possible to estimate the pair-correlation function by kernel smoothing of the differences between the observed points from X, reweighted by the reciprocal of $\hat{\lambda}(x)\hat{\lambda}(y)$; see Guan (2009) for the exact formula. Of course, the precision of the estimators of K and g depends heavily on the precision of $\hat{\lambda}$. Under an appropriate parametric model $\lambda = \lambda_\beta$, the estimates of K and g will be more stable than in the case where a nonparametric estimate of λ, obtained by kernel smoothing, is used.

Now the minimum contrast (6) can be employed for the estimation of the interaction parameters η in the same way as for the homogeneous case.

In Waagepetersen (2007), it was shown that the estimate of the inhomogeneity parameter β obtained by the Poisson likelihood L_1 differs negligibly from the estimate obtained by a more complicated and computationally much more demanding second-order estimation equation, which corresponds to the score equation of the full composite likelihood (8) in the inhomogeneous case. This finding supports the use of the first-order intensity function in L_1 for the estimation of β and it appears reasonable to estimate the interaction parameter η conditionally on $\hat{\beta}$ being fixed.

The two-step composite likelihood estimation was suggested in Jalilian et al. (2013). Here, formula (8) is rewritten as

$$\begin{aligned} \log CL(\eta) = \sum_{x,y \in X \cap W,\, 0<\|x-y\|<R} \Bigg[& \log(\hat{\lambda}(x)\hat{\lambda}(y) g_\eta(y-x)) \\ & - \log\left(\int_W \int_W \hat{\lambda}(u)\hat{\lambda}(v) g_\eta(u-v) I(\|u-v\| < R)\, \mathrm{d}u\, \mathrm{d}v \right) \Bigg], \end{aligned} \tag{12}$$

and maximized with respect to the interaction parameter η for fixed $\hat{\lambda}$. As in the homogeneous case, $R > 0$ is a tuning parameter. This two-step maximization is computationally much less demanding than maximization of the full composite likelihood (8) with respect to the complete parameter ψ.

6 Two-step estimation with Palm likelihood

In this section, we generalize the Palm likelihood estimator from the stationary case to a two-step estimation procedure for SOIRS inhomogeneous shot-noise Cox processes. The first step is the same as in the previous section so the inhomogeneity parameter β is still estimated using the Poisson likelihood (11). However, in order to estimate the interaction parameter, we need to generalize the Palm likelihood (10) to the inhomogeneous case and this is not a straightforward problem. There are, in fact, several possibilities.

The first option is to mimic formula (10) closely and just plug-in instead of $\lambda_o(y-x)$ the inhomogeneous version of the Palm intensity $\lambda_x(y) = \lambda(y)g(y-x)$ which now depends on both locations x and y. As a consequence, the quantity $|X \cap W|$ must be replaced by a sum over $x \in X \cap W$ and we get

$$\log L_{P1}(\eta) = \sum_{\substack{x,y \in X \cap W \\ 0<\|x-y\|<R}} \log(\hat{\lambda}(y) g_\eta(y-x)) - \sum_{x \in X \cap W} \int_{B(x,R)} \hat{\lambda}(u) g_\eta(u-x)\,\mathrm{d}u. \tag{13}$$

Note that $\log L_{P1}$ can also be rewritten as

$$\log L_{P1}(\eta) = \sum_{x \in X \cap W} \left(\sum_{\substack{z \in ((X \cap W)-x) \\ 0<\|z\|<R}} \log(\hat{\lambda}(x+z) g_\eta(z)) - \int_{B(x,R)} \hat{\lambda}(u) g_\eta(u-x)\,\mathrm{d}u \right).$$

Thus, L_{P1} is actually equal to the composite log-likelihood composed from the Poisson likelihoods of the difference processes $Y_x = \{y - x : y \in X \cap W,\ 0 < \|y - x\| < R\}$ with intensity functions (apart from edge effects) equal to $\lambda_x(u)$. This corresponds to the second method of derivation of the homogeneous Palm likelihood.

The second option is to use the whole process of differences $Y = \{x - y : x \neq y \in X \cap W\} \cap B(o, R)$ viewed for the purpose of approximate inference as a superposition of independent Poisson processes Y_x, $x \in X \cap W$. The intensity of the difference process Y is (again apart from edge effects) equal to $\sum_{x \in X \cap W} \lambda(x+u)g(u)$. Thus, the Palm likelihood L_{P2} defined as the Poisson likelihood of the process Y can be expressed as

$$\begin{aligned} \log L_{P2}(\eta) = & \sum_{\substack{z=w-y: w,y \in X \cap W \\ 0<\|z\|<R}} \log \left(\sum_{x \in X \cap W} \hat{\lambda}(x+z) g_\eta(z) \right) \\ & - \int_{B(o,R)} \sum_{x \in X \cap W} \hat{\lambda}(x+u) g_\eta(u)\,\mathrm{d}u. \end{aligned} \tag{14}$$

However, note that the second term in (13) and (14) is actually the same and since $\hat{\lambda}$ does not depend on η, both (13) and (14) may be written as

$$C + \sum_{\substack{z=w-y: w,y \in X \cap W \\ 0<\|z\|<R}} \log g_\eta(z) - \sum_{x \in X \cap W} \int_{B(x,R)} \hat{\lambda}(u) g_\eta(u-x)\,\mathrm{d}u,$$

as a function of η (for a suitably chosen constant C). Thus, the two derivations lead to the same Palm likelihood estimation which we will denote L_{P1} in the sequel.

The third option for generalization of the Palm likelihood is based on the following observation for the homogeneous case: the normalized number of points $|X \cap W|/|W|$ is an unbiased estimator of the constant intensity λ of a stationary process X. Thus, the complete version of the homogeneous Palm likelihood (9) can be expressed as

$$\begin{aligned}\log L_P(\eta) &= \sum_{\substack{x \neq y \in X \cap W \\ \|y-x\| < R}} \log\left(|X \cap W| \lambda_o(y-x;\eta)\right) \\ &\quad - |X \cap W| \int_{\mathbb{R}^d} I(\|u\| < R)\lambda_o(u;\eta)\,\mathrm{d}u \\ &= \sum_{\substack{x \neq y \in X \cap W \\ \|y-x\| < R}} \log\left(\hat{\lambda}|W|\lambda_o(y-x;\eta)\right) - \int_{\mathbb{R}^d} \hat{\lambda}|W| I(\|u\| < R)\lambda_o(u;\eta)\,\mathrm{d}u.\end{aligned}$$

Since $|W|$ in the first term does not change the maximum of L_P, it can be omitted and we get

$$\log L_P(\eta) = \sum_{\substack{x \neq y \in X \cap W \\ \|y-x\| < R}} \log\left(\hat{\lambda}\lambda_o(y-x;\eta)\right) - \int_W \hat{\lambda} \int_{B(v,R)} \lambda_o(u-v;\eta)\,\mathrm{d}u\,\mathrm{d}v.$$

If we now in the inhomogeneous case use $\hat{\lambda}(x)$ instead of $\hat{\lambda}$, decompose the Palm intensity $\lambda_x(u) = \lambda(u)g(u-x)$ and change the order of integration in the second term, we get a third version of the inhomogeneous Palm likelihood

$$\begin{aligned}\log L_{P3}(\eta) = &\sum_{\substack{x \neq y \in Y \cap W \\ \|x-y\| < R}} \log\left(\hat{\lambda}(x)\hat{\lambda}(y)g_\eta(y-x)\right) \\ &- \int_{B(o,R)} \int_{W \cap (W-u)} \hat{\lambda}(v)\hat{\lambda}(v+u)g_\eta(u)\,\mathrm{d}v\,\mathrm{d}u. \end{aligned} \tag{15}$$

Finding the estimate $(\hat{\beta}, \hat{\eta})$ by the two-step estimation corresponds to solving the score equation

$$U(\beta,\eta) = (U_1(\beta), U_2(\beta,\eta)) = 0, \tag{16}$$

where

$$U_1(\beta) = \sum_{x \in X \cap W} \frac{\lambda'_\beta(x)}{\lambda_\beta(x)} - \int_W \lambda'_\beta(u)\,\mathrm{d}u$$

is the score function for the Poisson log-likelihood (11),

$$U_2(\beta,\eta)=\frac{\mathrm{d}\log L_{P1}(\eta)}{\mathrm{d}\eta}=\sum_{\substack{x\neq y\in X\cap W\\ \|y-x\|<R}}\frac{g'_\eta(y-x)}{g_\eta(y-x)}-\sum_{x\in X\cap W}\int_{B(x,R)}\lambda_\beta(u)g'_\eta(u-x)\,\mathrm{d}u$$

is the score function for $\log L_{P1}$ and

$$\begin{aligned}U_2(\beta,\eta)=\frac{\mathrm{d}\log L_{P3}(\eta)}{\mathrm{d}\eta}&=\sum_{x\neq y\in X\cap W,\|y-x\|<R}\frac{g'_\eta(y-x)}{g_\eta(y-x)}\\&\quad-\int_{B(o,R)}\int_{W\cap(W-u)}\lambda_\beta(v)\lambda_\beta(v+u)g'_\eta(u)\,\mathrm{d}v\,\mathrm{d}u\end{aligned}$$

is the score function for $\log L_{P3}$. Here, λ'_β and g'_η denote the derivatives of the intensity function and the pair-correlation function with respect to β and η, respectively.

Note that (16) is an unbiased estimating equation for L_{P3}. To get an unbiased estimating equation also for L_{P1}, we would need to include an edge correction into the integrals in the second term of (13), obtaining the following unbiased version

$$\log L_{P1}(\eta)=\sum_{\substack{x,y\in X\cap W\\ 0<\|x-y\|<R}}\log(\hat{\lambda}(y)g_\eta(y-x))-\sum_{x\in X\cap W}\int_{B(x,R)\cap W}\hat{\lambda}(u)g_\eta(u-x)\,\mathrm{d}u. \tag{17}$$

As in the stationary case, R is a user-specified tuning constant that may influence the efficiency of the estimator. Obviously, if ρ is the (practical) interaction range of the process, we have $g(u)=1$ (or $g(u)\approx 1$) for $\|u\|>\rho$. Thus, by using $R>\rho$, we only introduce additional variance into the estimation of the interaction parameter η. Moreover, using too large a R may lead to numerical instability of the maximization procedure; see Sect. 8 for details. Thus, we recommend to use R somewhat smaller than the likely interaction range of the analyzed point pattern. For a more detailed discussion of the influence of the choice of R on the estimation for a selection of shot-noise Cox process models, see Sect. 8.

7 Asymptotic properties

In Waagepetersen and Guan (2009), asymptotic normality of the estimators from the two-step estimation procedure with the minimum contrast based on the K-function is proved under certain moment and mixing conditions. Fulfillment of these conditions is discussed for Poisson Neyman–Scott processes and log-Gaussian Cox processes. These conditions are also satisfied for shot-noise Cox processes as we show in the two following lemmas.

Lemma 1 *Let X be a stationary shot-noise Cox process satisfying $\int_{\mathbb{R}+} r^k V(\mathrm{d}r)<\infty$ for some $k\in\mathbb{N}$. Then, X has well-defined moment measures up to the kth order and all reduced factorial cumulant measures up to the kth order have finite total variation.*

Proof The first statement follows from Theorem 3 and Proposition 2 in Hellmund et al. (2008). It is well-known for cluster processes (see e.g., Heinrich 1988) that if the parent process has reduced factorial cumulant measures of finite total variation up to order k and the distribution of the number of points in the clusters has finite moments up to order k, then also all reduced factorial cumulant measures of the cluster process up to order k have finite total variation. For any shot-noise Cox process X, it is possible to define an approximating shot-noise Cox process with only a finite number of clusters in a bounded region, i.e., with $\int_{\mathbb{R}+} V(\mathrm{d}r) < \infty$, and with the same moment measures up to the order k. The construction is based on Proposition 3 in Hellmund et al. (2008). This approximating process is then just a standard cluster process with stationary Poisson distribution of parents and as such with reduced factorial cumulant measures up to the kth order of finite total variation. Since these reduced factorial cumulant measures are identical to those of the original shot-noise Cox process X, the second statement follows. □

Lemma 2 *Let X be a stationary shot-noise Cox process in $\mathbb{R}^d$ with $\int_{\mathbb{R}+} rV(\mathrm{d}r) < \infty$ so that the first-order moment measure is well-defined. Let*

$$\alpha_{p_1,p_2}(m) = \sup\{\alpha(\mathscr{F}^X(A), \mathscr{F}^X(B)) : d(A, B) \geq m, |A| \leq p_1, |B| \leq p_2\}, \quad (18)$$

where $p_1, p_2 > 0$, $\mathscr{F}^X(A)$ denotes the σ-algebra generated by $X \cap A$, $d(A, B)$ denotes the Hausdorff distance between A and B, the supremum is taken over all sets A, B in $\mathscr{B}^d$ and

$$\alpha(\mathscr{F}_1, \mathscr{F}_2) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \mathscr{F}_1, B \in \mathscr{F}_2\}$$

denotes the standard strong mixing coefficient.

If there exists a function h such that $k(c, v) = h(v - c)$ and an $\epsilon > 0$ such that $h(v) = \mathscr{O}(|v|^{-(2d+\epsilon)})$, as $|v| \to \infty$, then $\frac{\alpha_{p,p}(m)}{\max(p,1)} \leq \mathscr{O}(m^{-d-\epsilon})$.

Proof Let us rewrite X as $\bigcup_{(r,v)\in\Pi_\Upsilon} X_v$, where X_v is the cluster centered around a point located at v with intensity function $rk(\cdot, v)$. Denote $X_1 = \bigcup_{(r,v)\in\Pi_\Upsilon, v\in A} X_v$. Then, using the fact that $\mathsf{E}(X_1 \cap B) = \mu \int_{\mathbb{R}+} rV(\mathrm{d}r) \int_A \int_B k(v, u)\mathrm{d}u\,\mathrm{d}v$ for any $A, B \in \mathscr{B}^d$, the proof is exactly the same as the proof of Lemma 1 in Prokešová and Jensen (2013).

Asymptotic normality of the estimators obtained by the two-step estimation procedure with the composite likelihood in the second step, based on the formula (12), is briefly discussed in Section 6.3 of Jalilian et al. (2013).

For the two-step estimation procedure with Palm likelihood in the second step, consistency and asymptotic normality can be shown along the same lines as in Waagepetersen and Guan (2009, Theorem 1). In particular, Theorem 1 below covers all the point process models considered in Waagepetersen and Guan (2009). For simplicity we restrict ourselves to the case of $\mathbb{R}^d = \mathbb{R}^2$.

We will consider an expanding window asymptotics such that X is observed on a sequence of windows $\{W_n\}$ expanding to $\mathbb{R}^2$. The estimators obtained from $X \cap$

W_n by the two-step estimation with either L_{P1} (17) or L_{P3} (15) are denoted $\hat{\beta}_n$ and $\hat{\eta}_n$. The corresponding score functions obtained for $X \cap W_n$ are $U_n(\beta, \eta) = (U_{n,1}(\beta), U_{n,2}(\beta, \eta))$. Further, we denote by β_0 and η_0 the true values of the parameters to be estimated.

Let $\Sigma_n = |W_n|^{-1} \operatorname{Var}(U_n(\beta_0, \eta_0))$ be the information matrix for the considered score function and let us define

$$I_n = \begin{pmatrix} I_{n,11} & I_{n,12} \\ 0 & I_{n,22} \end{pmatrix} = \frac{1}{|W_n|} \left(- \mathsf{E} \left. \frac{\mathrm{d}U_n(\beta, \eta)}{\mathrm{d}(\beta, \eta)^T} \right|_{(\beta,\eta)=(\beta_0,\eta_0)} \right),$$

where

$$I_{n,11} = \frac{1}{|W_n|} \int_{W_n} \frac{(\lambda'_{\beta_0}(u))^T \lambda'_{\beta_0}(v)}{\lambda_{\beta_0}(u)} \, \mathrm{d}u \tag{19}$$

and

$$I_{n,22} = \frac{1}{|W_n|} \int_{W_n} \int_{B(v,R)\cap W_n} \frac{(g'_{\eta_0}(u-v))^T g'_{\eta_0}(u-v)}{g_{\eta_0}(u-v)} \lambda_{\beta_0}(u) \lambda_{\beta_0}(v) \, \mathrm{d}u \, \mathrm{d}v \tag{20}$$

are the same for L_{P1} and L_{P3}, while

$$I_{n,12} = \frac{1}{|W_n|} \int_{W_n} \lambda_{\beta_0}(v) \int_{B(v,R)\cap W_n} (\lambda'_{\beta_0}(u))^T g'_{\eta_0}(u-v) \, \mathrm{d}u \, \mathrm{d}v \tag{21}$$

for L_{P1} and a double of this matrix for L_{P3}.

Theorem 1 *Let X be a SOIRS Cox process in $\mathbb{R}^2$ whose kth-order intensity functions $\lambda_\beta^{(k)}$ satisfy*

$$\lambda_\beta^{(k)}(u_1, \ldots, u_k) = \lambda^{(k)}(u_1, \ldots u_k) \prod_{i=1}^{k} \lambda_\beta(u_i), \qquad k \in \mathbb{N}, \tag{22}$$

where λ_β is the first-order intensity function of X and $\lambda^{(k)}$ are kth-order intensity functions of a stationary Cox process. Let $\{W_n\}_{n=1}^{\infty}$ be a sequence of observation windows $W_n = [an, bn] \times [cn, dn]$, where $(b-a) > 0$, $(d-c) > 0$ and $0 \in Int(W_n)$. For $s > 0$ let $A_{i,j} = [is, (i+1)s) \times [js, (j+1)s) \oplus R$, $i, j \in \mathbb{Z}^2$, and

$$\alpha^F_{p_1,p_2}(m) = \sup \Bigg\{ \alpha(\mathscr{F}^X(B_1), \mathscr{F}^X(B_2)) : B_1 = \bigcup_{M_1} A_{i,j}, B_2 = \bigcup_{M_2} A_{i,j},$$
$$|M_1| \le p_1, |M_2| \le p_2, d(M_1, M_2) \ge m, M_1, M_2 \subset \mathbb{Z}^2 \Bigg\},$$

where $d(M_1, M_2)$ denotes the minimal distance between M_1 and M_2 in the grid $\mathbb{Z}^2$ and $\alpha(\mathscr{F}_1, \mathscr{F}_2)$ is the standard strong mixing coefficient.

Assume

(A0) $\lambda_\beta(u) = f(z(u)\beta^T)$ *for some strictly increasing positive differentiable function* f *and* $\|z(u)\| < K_1, u \in \mathbb{R}^2$, *for some* $K_1 > 0$ *(bounded covariates);*
(A1) $\lambda^{(2)}$ *and* $\lambda^{(3)}$ *are bounded and there exists* K_2 *so that* $\int |\lambda^{(3)}(0, v, v + u_1) - \lambda^{(1)}(0)\lambda^{(2)}(0, u_1)| \mathrm{d}v < K_2$ *and* $\int |\lambda^{(4)}(0, u_1, v, v+u_2) - \lambda^{(2)}(0, u_1)\lambda^{(2)}(0, u_2)| \mathrm{d}v < K_2$ *for all* $u_1, u_2 \in \mathbb{R}^2$;
(A2) $\lambda_\beta(u)$ *and* $g_\eta(u)$ *have well-defined first and second derivatives with respect to* β *and* η, *and these are continuous functions of* (u, β) *and* (u, η), *respectively;*
(A3) $\liminf_{n\to\infty}(\lambda_{n,ii}) > 0$, $i = 1, 2$, *where* $\lambda_{n,11}$ *and* $\lambda_{n,22}$ *are the smallest eigenvalues of* $I_{n,11}$ *and* $I_{n,22}$, *respectively. The information matrices* Σ_n *converge to a positive definite matrix* Σ *as* $n \to \infty$;
(A4) $\lambda^{(4+2\nu)}(u_1, \ldots, u_{4+2\nu}) < \infty$ *for some* $\nu \in \mathbb{N}$;
(A5) *There exists an* $s > 0$ *such that* $\alpha^F_{2,\infty}(m) = \mathscr{O}(m^{-\delta})$ *for some* $\delta > 2(2 + \nu)/\nu$.

Then, there exists a sequence $\{(\hat\beta_n, \hat\eta_n)\}_{n\geq 1}$ *for which* $U_n(\hat\beta_n, \hat\eta_n) = 0$ *with probability converging to 1 and*

$$|W_n|^{1/2}\{(\hat\beta_n, \hat\eta_n) - (\beta_0, \eta_0)\} I_n \Sigma_n^{-1/2} \xrightarrow{D} N(0, \mathbf{1}),$$

where $N(0, \mathbf{1})$ *is the standard normal* $(t + q)$*-dimensional distribution.*

Proof The proof is analogous to the proof of Waagepetersen and Guan (2009, Theorem 1) for the two-step estimation with minimum contrast for the K-function. However, we have used a different mixing assumption (A5) formulated directly for the mixing coefficient of a random field. Our assumption is weaker than the one in Waagepetersen and Guan (2009) and it suffices for the application of the central limit theorem 3.3.1 in Guyon (1991) for random fields, which is needed in the proof. □

Remark If the kernel k of a stationary shot-noise Cox process is bounded and the assumption of Lemma 1 is satisfied, then it follows from the formulae for $\lambda^{(k)}$ in Hellmund et al. (2008, Section 4) that these are bounded and continuous. So are the densities of the reduced factorial cumulant measures up to order k. Moreover, since the kth order reduced factorial cumulant measures have finite total variation, it follows that the integrals of the densities of the reduced factorial cumulant measures up to order k are bounded. Thus, Lemma 1 for $k = 4$ implies assumption (A1).

Remark In Waagepetersen and Guan (2009, Theorem 1) a stronger mixing assumption is used

(Av) there exist constants $a > 8R^2$ and $\delta > 2(2 + \nu)/\nu$ such that $\alpha_{a,\infty}(m) = \mathscr{O}(m^{-\delta})$.

This assumption is formulated for the mixing coefficient of the point process X and as such it implies our assumption (A5). However, it is unnecessarily strong and no simple conditions are available for Poisson Neyman–Scott processes or shot-noise Cox processes which would ensure fulfillment of (Av). The assumption

$$\sup_{w\in[-m/2,m/2]^2} \left\{ \int_{\mathbb{R}^2\setminus[-m,m]^2} k(v - w)\,\mathrm{d}v \right\} = \mathscr{O}(m^{-\delta-2}) \tag{23}$$

presented in Waagepetersen and Guan (2009, Appendix E) is not sufficient for (Av). Nevertheless it is sufficient for assumption (A5), as the following lemma shows.

Lemma 3 *Let X be a stationary shot-noise Cox process in $\mathbb{R}^2$ with well-defined first-order moment measure and kernel function k, satisfying* (23)*. Then X satisfies condition (A5).*

Proof For a given s, let $n = ms - \frac{s}{2} - R > 0$, and consider the sets $E_1 = A_{0,0} - (s/2, s/2)$, $E_2 = \mathbb{R}^2 \backslash [-n, n]^2$ and $E_3 = [-n/2, n/2]^2$. Further, using the cluster representation of X, let $X_1 = \bigcup_{(r,v)\in\Pi_\Upsilon, v\in E_3} X_v$, $X_2 = X\backslash X_1$. Then X_1, X_2 are independent cluster processes and by standard arguments, like those in Waagepetersen and Guan (2009, Appendix E), we get

$$\begin{aligned}
\alpha(\mathscr{F}^X(E_1), \mathscr{F}^X(E_2)) \le\ & 5(\mathsf{E}\,|X_1 \cap E_2| + \mathsf{E}\,|X_2 \cap E_1|) \\
\le\ & 5\mu \int_{\mathbb{R}^+} rV(\mathrm{d}r)\Bigg(\int_{[-\frac{n}{2},\frac{n}{2}]^2} \int_{\mathbb{R}^2\backslash[-n,n]^2} k(u-v)\,\mathrm{d}u\,\mathrm{d}v \\
& + \int_{\mathbb{R}^2\backslash[-\frac{n}{2},\frac{n}{2}]^2} \int_{E_1} k(u-v)\,\mathrm{d}u\,\mathrm{d}v\Bigg) \\
\le\ & const\Bigg(|E_3| \sup_{v\in[-\frac{n}{2},\frac{n}{2}]^2} \int_{\mathbb{R}^2\backslash[-n,n]^2} k(u-v)\,\mathrm{d}u\,\mathrm{d}v \\
& + |E_1| \sup_{v\in E_1} \int_{\mathbb{R}^2\backslash[-\frac{n}{2},\frac{n}{2}]^2} k(u-v)\,\mathrm{d}u\,\mathrm{d}v\Bigg).
\end{aligned}$$

If m is sufficiently large such that $E_1 \subset [-n/4, n/4]^2$ we get from (23) that both terms on the right-hand side are $O(m^{-\delta})$. This implies (A5) for $\alpha^F_{1,\infty}(m)$.
For $\alpha^F_{2,\infty}(m)$ we just need to consider for some $(i, j) \in \mathbb{Z}^2$:

$$\begin{aligned}
E_1 &= (A_{0,0} \cup A_{i,j}) - (s/2, s/2) \\
E_2 &= (\mathbb{R}^2\backslash[-n, n]^2)\backslash([-n, n]^2 + (is, js)) \\
E_3 &= [-n/2, n/2]^2 \cup ([-n/2, n/2]^2 + (is, js)).
\end{aligned}$$

We get by similar arguments as the ones given above

$$\begin{aligned}
\alpha(\mathscr{F}^X(E_1), \mathscr{F}^X(E_2)) \le const\Bigg(&|E_3| \sup_{v\in[-\frac{n}{2},\frac{n}{2}]^2} \int_{\mathbb{R}^2\backslash[-n,n]^2} k(u-v)\,\mathrm{d}u\,\mathrm{d}v \\
&+ |E_1| \sup_{v\in(A_{0,0}-(s/2,s/2))} \int_{\mathbb{R}^2\backslash[-\frac{n}{2},\frac{n}{2}]^2} k(u-v)\,\mathrm{d}u\,\mathrm{d}v\Bigg),
\end{aligned}$$

where we have used the stationarity of X. Thus, again if $\frac{s}{2} + R < \frac{n}{4}$ holds, we get from (23) that both terms on the right-hand side are $O(m^{-\delta})$. This implies (A5) for $\alpha^F_{2,\infty}(m)$. $\square$

The inhomogeneous shot-noise Cox process, as defined at the end of Sect. 3, inherits the mixing properties of the unthinned homogeneous process, since the inhomogeneous process was constructed by location-dependent thinning. Therefore, condition (23) for the homogeneous kernel k ensures that (A5) is fulfilled also for the inhomogeneous shot-noise Cox process X.

Remark The incomplete argument in Waagepetersen and Guan (2009, Appendix E) stems from the fact that a set $E_1 = [-h, h]^2$ was considered for some $h > 0$ and it was assumed that whatever Borel set A with fixed volume a will fit into such E_1. However, for $\alpha_{a,\infty}(m)$ to be of order $\mathscr{O}(m^{-\delta})$, a universal set E_1 would be needed, which could cover all Borel sets of volume $\le a$. Unfortunately, this is not possible, since the set A may be arbitrarily "thin" and so there will always exist some set A which is not a subset of any fixed square E_1. Therefore, the tail condition (23) can only assure (A5) for the mixing coefficient of the random field and not (Av) for the mixing coefficient of the point process X.

It is possible to use Theorem 1 to derive approximate confidence intervals for the parameter estimates, if we are able to compute the information matrix Σ_n. Below, we give the formulae for the submatrices of the block representation, corresponding to the decomposition into the following two parts of the score function

$$\Sigma_n = |W_n|^{-1}\,\mathsf{Var}(U_{n,1}(\beta_0), U_{n,2}(\beta_0,\eta_0)) = \begin{pmatrix} \Sigma_{n,11} & \Sigma_{n,12} \\ \Sigma_{n,12}^T & \Sigma_{n,22} \end{pmatrix}.$$

For both L_{P1} and L_{P3}, we obtain the same expression

$$\Sigma_{n,11} = I_{n,11} + \frac{1}{|W_n|}\int_{W_n}\int_{W_n} (\lambda'_{\beta_0}(u))^T \lambda'_{\beta_0}(v)\big(g'_{\eta_0}(u-v) - 1\big)\,\mathrm{d}u\,\mathrm{d}v.$$

For L_{P1} we get

$$\begin{aligned}\Sigma_{n,12} = \frac{1}{|W_n|}\Bigg[&\int_{W_n^3} \frac{(\lambda'_{\beta_0}(w))^T}{\lambda_{\beta_0}(w)} g'_{\eta_0}(u-v) I(\|u-v\| < R) \\ &\times\left(\frac{\lambda^{(3)}_{\beta_0}(w,u,v)}{g_{\eta_0}(u-v)} - \lambda_{\beta_0}(u)\lambda^{(2)}_{\beta_0}(w,v)\right)\mathrm{d}w\,\mathrm{d}u\,\mathrm{d}v \\ &+ \int_{W_n^2} (\lambda'_{\beta_0}(u))^T g'_{\eta_0}(u-v) I(\|u-v\| < R)\lambda_{\beta_0}(v)\,\mathrm{d}u\,\mathrm{d}v\Bigg]\end{aligned}$$

and for L_{P3}

$$\Sigma_{n,12} = \frac{1}{|W_n|}\Bigg[\int_{W_n^3} \frac{(\lambda'_{\beta_0}(w))^T}{\lambda_{\beta_0}(w)} \frac{g'_{\eta_0}(u-v)}{g_{\eta_0}(u-v)} I(\|u-v\| < R)$$

$$\times \left(\lambda^{(3)}_{\beta_0}(w,u,v) - \lambda_{\beta_0}(w)\lambda^{(2)}_{\beta_0}(u,v) \right) \mathrm{d}w\,\mathrm{d}u\,\mathrm{d}v$$
$$+ 2\int_{W_n^2} (\lambda'_{\beta_0}(u))^T g'_{\eta_0}(u-v) I(\|u-v\| < R)\lambda_{\beta_0}(v)\,\mathrm{d}u\,\mathrm{d}v \Bigg].$$

Finally,

$$\Sigma_{n,22,LP3} = \frac{1}{|W_n|}\Bigg[2\int_{W_n^2} \frac{(g'_{\eta_0}(u-v))^T g'_{\eta_0}(u-v) I(\|u-v\| < R)}{g_{\eta_0}(u-v)}$$
$$\lambda_{\beta_0}(u)\lambda_{\beta_0}(v)\,\mathrm{d}u\,\mathrm{d}v$$
$$+ 4\int_{W_n^3} \frac{(g'_{\eta_0}(u-v))^T g'_{\eta_0}(v-w)}{g_{\eta_0}(u-v)g_{\eta_0}(v-w)} I(\|u-v\|, \|v-w\| < R)$$
$$\lambda_{\beta_0}(u)\lambda_{\beta_0}(v)\lambda_{\beta_0}(w)\lambda^{(3)}(u,v,w)\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w$$
$$+ \int_{W_n^4} \frac{(g'_{\eta_0}(u-v))^T g'_{\eta_0}(w-z)}{g_{\eta_0}(u-v)g_{\eta_0}(w-z)} I(\|u-v\|, \|w-z\| < R)$$
$$\left(\lambda^{(4)}_{\beta_0}(u,v,w,z) - \lambda^{(2)}_{\beta_0}(u,v)\lambda^{(2)}_{\beta_0}(w,z) \right) \mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w\,\mathrm{d}z \Bigg]$$

and for L_{P1}

$$\Sigma_{n,22,LP1} = \Sigma_{n,22,LP3} + \frac{1}{|W_n|}$$
$$\Bigg[-3\int_{W_n^3} (g'_{\eta_0}(u-v))^T g'_{\eta_0}(v-w) I(\|u-v\|, \|v-w\| < R)$$
$$\lambda_{\beta_0}(u)\lambda_{\beta_0}(v)\lambda_{\beta_0}(w)\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w$$
$$+ \int_{W_n^4} (g'_{\eta_0}(u-v))^T g'_{\eta_0}(w-z) I(\|u-v\|, \|w-z\| < R)$$
$$\left(\lambda^{(2)}_{\beta_0}(v,z)\lambda_{\beta_0}(u)\lambda_{\beta_0}(w) - 2\frac{\lambda^{(3)}(v,u,z)}{g_{\eta_0}(u-v)}\lambda_{\beta_0}(w) \right) \mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w\,\mathrm{d}z \Bigg].$$

See Sect. 8.3 for more details about the computation of the approximate confidence intervals in practice.

8 Simulation study

8.1 Design of the simulation study

To compare the performance of the developed two-step estimation methods and to assess the influence of the choice of the tuning constant R for the PL and CL methods we applied the MCK, MCg, CL and PL estimation procedures to realizations from the inhomogeneous gamma shot-noise Cox process (see Example 2) with parameters μ

and θ, observed on the unit square $W = [0, 1]^2$. We chose the smoothing kernel $k(u)$ to be the Gaussian kernel function with standard deviation σ (density of a zero-mean bivariate radially symmetric normal distribution).

First, we have generated realizations of a homogeneous version of the process (with the intensity $\frac{\mu}{\theta}$) and then applied the location-dependent thinning, using the inhomogeneity function

$$f(x) = \exp(\beta_1 x_1 - \max(\beta_1, 0)), \quad x = (x_1, x_2) \in W.$$

Note that f is properly scaled to fulfill the condition $\max_W f = 1$. The intensity function of the thinned process is therefore $\frac{\mu}{\theta} f(x), \;\; x \in W$. We can express the intensity function as

$$\lambda_\beta(x) = \exp(\beta_0 + \beta_1 x_1), \quad x \in W,$$

where $\beta_0 = \log\mu - \log\theta - \max(\beta_1, 0)$ and the interaction parameter is $\eta = (\mu, \sigma)$. The mean number of points in W is thus

$$\mathsf{E}\,|X \cap W| = \int_W \lambda_\beta(x)\,\mathrm{d}x = \frac{\mu}{\theta \cdot |\beta_1|}\,(1 - \exp(-|\beta_1|)). \tag{24}$$

In the first estimation step, we used the Poisson log-likelihood function (11) to estimate the parameter $\beta = (\beta_0, \beta_1)$. The estimation was performed by means of the function `ppm` from the `R` package `Spatstat` (Baddeley and Turner 2005; Baddeley et al. 2015). The vector of interaction parameters $\eta = (\mu, \sigma)$ was estimated in the second step by the methods described in Sects. 5 and 6.

The minimum contrast estimation, using the inhomogeneous K-function (MCK) and the pair-correlation function (MCg), was performed by a `Spatstat` routine. The value of the tuning parameter r, see Eq. (6), was chosen as the minimal observed interpoint distance in the given point pattern (which is a standard choice in similar situations in the literature) while the value of the tuning parameter R was 4σ. The value of 4σ corresponds to the practical range of interaction in the considered point process. Using larger values of R would result in no further gain of information, only in larger variability of the estimates. The variance stabilizing exponent q was chosen to be $1/4$ for MCK and $1/2$ for MCg, based on our previous studies Dvořák and Prokešová (2012) and Prokešová and Dvořák (2014). This established choice of tuning constants can be considered an advantage of the minimum contrast methods in this simulation study while for the composite likelihood (CL) and Palm likelihood (PL) methods the influence of the tuning parameter R has not been studied before and no recommendation is available.

The CL and PL estimates were obtained by a grid search for σ combined with numerical maximization in μ (combination of golden section search and successive parabolic interpolation performed by the `R` function `optimize`). Simultaneous maximization for the complete vector (μ, σ) by various optimization algorithms turned out to be numerically unstable. In order to investigate the influence of the tuning

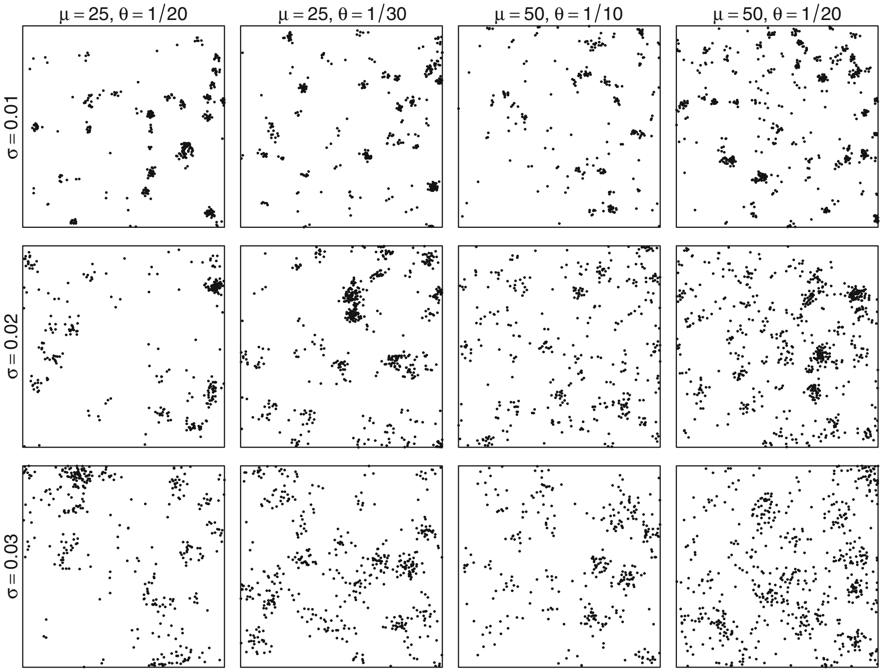

Fig. 2 Realizations of the point processes used in the simulation study. For details, see Sect. 8.1

parameter R for the CL and PL methods, the estimates were computed using three different values of $R = 0.1$, 0.2 and 0.3.

Finally, the remaining parameter θ was estimated from (24) where $\mathsf{E}\,|X \cap W|$ was replaced by the actual number of observed points in W and μ and β_1 were similarly replaced by their respective estimates.

To study properties of the estimators under different cluster size distributions we chose the values of μ and θ to be 25 or 50 and 1/10, 1/20 or 1/30, respectively. Different degree of clustering was obtained by taking the values of σ to be 0.01, 0.02 or 0.03. For the inhomogeneity function we use the parameter value $\beta_1 = 1$.

We disregarded the two extreme combination of parameters ($\mu = 25, \theta = 1/10$ and $\mu = 50, \theta = 1/30$). The remaining combinations of parameter values result in the mean number of points in $X \cap W$ ranging from approx. 310–630. For each combination of parameters we generated 500 independent realizations from our model and re-estimated the parameters. All the estimation procedures were applied to the same set of simulated patterns. Figure 2 shows realizations of the point processes for the considered combinations of parameters.

8.2 Results of the simulation study

Tables 1, 2 and 3 show relative mean squared errors (MSEs) of the estimators and relative mean biases. Relative quantities are for MSEs obtained by dividing by the square of the true value of the estimated parameter while in case of biases we have divided

Table 1 Relative mean squared errors (upper row) and relative mean biases (lower row) of the estimators of σ, determined by simulation of the point process models with the specified combinations of the parameters μ, θ and σ, shown in the left column

μ	θ	σ	MCK	MCg	CL			PL1			PL3		
					0.1	0.2	0.3	0.1	0.2	0.3	0.1	0.2	0.3
25	1/20	0.01	**0.004**	0.006	0.007	0.020	0.013	0.009	0.009	0.009	0.011	0.011	0.011
			0.003	−0.047	0.005	0.013	0.009	−0.019	−0.019	−0.019	**−0.001**	**−0.001**	**−0.001**
25	1/20	0.02	0.007	0.009	**0.006**	0.020	0.041	0.015	0.016	0.016	0.015	0.023	0.023
			−0.010	−0.045	**0.002**	0.015	0.024	−0.072	−0.072	−0.072	−0.016	−0.009	−0.009
25	1/20	0.03	**0.017**	**0.017**	0.020	0.022	0.041	0.029	0.034	0.034	0.019	0.057	0.124
			−0.021	−0.048	0.021	0.016	0.031	−0.124	−0.147	−0.147	−0.027	**−0.006**	0.013
25	1/30	0.01	**0.003**	0.005	0.009	0.038	0.043	0.018	0.019	0.019	0.023	0.025	0.025
			−0.004	−0.048	0.006	0.024	0.026	−0.011	−0.011	−0.011	0.008	0.009	0.009
25	1/30	0.02	0.006	0.008	**0.005**	0.020	0.036	0.016	0.017	0.017	0.016	0.038	0.056
			−0.009	−0.039	**0.001**	0.020	0.030	−0.069	−0.069	−0.069	−0.012	0.001	0.003
25	1/30	0.03	**0.011**	0.013	0.013	0.016	0.034	0.027	0.033	0.033	0.014	0.040	0.108
			−0.034	−0.057	0.012	**0.005**	0.013	−0.131	−0.154	−0.154	−0.035	−0.023	−0.008
50	1/10	0.01	**0.007**	**0.007**	0.010	0.013	0.012	0.008	0.008	0.008	0.010	0.010	0.010
			0.003	−0.045	0.012	0.014	0.012	−0.020	−0.020	−0.020	**0.003**	**0.003**	**0.003**
50	1/10	0.02	0.012	0.013	**0.010**	0.021	0.054	0.020	0.022	0.022	0.018	0.051	0.069
			−0.017	−0.052	**−0.001**	0.006	0.023	−0.098	−0.098	−0.098	−0.019	−0.003	**−0.001**
50	1/10	0.03	**0.020**	0.023	0.040	0.023	0.049	0.046	0.052	0.052	0.021	0.055	NA
			−0.038	−0.066	0.044	**0.006**	0.017	−0.018	−0.020	−0.020	−0.041	−0.022	NA
50	1/20	0.01	**0.003**	0.005	0.005	0.011	0.011	0.008	0.008	0.008	0.009	0.009	0.009
			−0.002	−0.046	**0.002**	0.004	0.004	−0.026	−0.026	−0.026	−0.006	−0.006	−0.006
50	1/20	0.02	0.006	0.008	**0.005**	0.016	0.028	0.015	0.016	0.016	0.012	0.043	0.074
			−0.007	−0.038	**0.006**	0.013	0.018	−0.090	−0.090	−0.090	−0.010	0.007	0.011
50	1/20	0.03	**0.012**	0.013	0.014	0.020	0.041	0.038	0.045	0.045	0.016	0.037	0.057
			−0.021	−0.045	0.018	0.018	0.038	−0.166	−0.190	−0.190	−0.021	−0.008	**0.001**

The estimation methods considered are MCK, MCg, CL, PL1 and PL3. For the three latter methods with tuning parameter $R = 0.1, 0.2$ and 0.3, respectively. The value closest to 0 in each row is indicated in boldface

by the true parameter value. The overall conclusion is that there is no uniformly best estimator. The performance of the different estimators depends both on the particular parameter which is to be estimated and on the tuning parameter R. However, the performance (according to the MSE) of the four estimators MCK, MCg, CL (with properly chosen R) and PL3 (with properly chosen R) is quite similar. Let us discuss the results for each of the parameters in more detail.

8.2.1 Estimation of σ

The scale parameter σ of the kernel k is the easiest one to estimate. The relative MSE of the estimators MCK, MCg, CL (with $R = 0.1$) and PL3 (with $R = 0.1$) is at most 2 % for all the considered models, thus all these four estimators produce

Table 2 Relative mean squared errors (upper row) and relative mean biases (lower row) of the estimators of μ, determined by simulation of the point process models with the specified combinations of the parameters μ, θ and σ, shown in the left column

μ	θ	σ	MCK	MCg	CL			PL1			PL3		
					0.1	0.2	0.3	0.1	0.2	0.3	0.1	0.2	0.3
25	1/20	0.01	**0.098**	0.111	0.249	0.139	0.126	0.154	0.154	0.154	0.125	0.125	0.125
			0.156	0.177	**0.089**	0.093	0.104	0.229	0.229	0.229	0.183	0.183	0.183
25	1/20	0.02	**0.159**	0.173	0.263	0.221	0.201	0.325	0.325	0.325	0.197	0.198	0.198
			0.183	0.200	**0.091**	0.102	0.109	0.363	0.363	0.363	0.221	0.218	0.218
25	1/20	0.03	0.277	0.300	0.446	**0.261**	0.272	0.710	0.766	0.766	0.299	0.326	0.330
			0.270	0.289	**0.096**	0.114	0.120	0.596	0.632	0.632	0.292	0.290	0.287
25	1/30	0.01	**0.097**	0.102	0.263	0.155	0.133	0.145	0.145	0.145	0.122	0.122	0.122
			0.141	0.148	0.089	**0.086**	0.087	0.207	0.207	0.207	0.162	0.162	0.162
25	1/30	0.02	**0.136**	0.146	0.233	0.195	0.178	0.344	0.345	0.345	0.208	0.210	0.211
			0.183	0.194	0.101	**0.091**	0.096	0.376	0.376	0.376	0.231	0.227	0.227
25	1/30	0.03	**0.223**	0.230	0.314	0.293	0.307	0.679	0.733	0.733	0.278	0.293	0.295
			0.251	0.260	**0.072**	0.118	0.150	0.585	0.620	0.620	0.284	0.284	0.281
50	1/10	0.01	**0.068**	0.082	0.111	0.086	0.086	0.101	0.101	0.101	0.081	0.081	0.081
			0.113	0.144	**0.061**	0.068	0.079	0.175	0.175	0.175	0.125	0.125	0.125
50	1/10	0.02	**0.122**	0.137	0.180	0.187	0.187	0.314	0.315	0.315	0.166	0.169	0.169
			0.148	0.172	**0.063**	0.091	0.093	0.350	0.351	0.351	0.169	0.164	0.164
50	1/10	0.03	**0.255**	0.272	0.440	0.276	0.317	1.07	1.12	1.12	0.342	0.362	NA
			0.247	0.268	**0.040**	0.123	0.154	0.742	0.781	0.781	0.288	0.287	NA
50	1/20	0.01	**0.065**	0.070	0.120	0.088	0.086	0.104	0.104	0.104	0.083	0.083	0.083
			0.100	0.110	**0.055**	0.063	0.069	0.169	0.169	0.169	0.122	0.122	0.122
50	1/20	0.02	**0.088**	0.095	0.125	0.137	0.132	0.243	0.243	0.243	0.119	0.122	0.123
			0.135	0.147	**0.064**	0.087	0.095	0.331	0.332	0.332	0.153	0.148	0.148
50	1/20	0.03	**0.173**	0.179	0.240	0.220	0.259	0.810	0.864	0.864	0.238	0.255	0.257
			0.191	0.202	**0.077**	0.100	0.101	0.651	0.692	0.692	0.220	0.220	0.219

The estimation methods considered are MCK, MCg, CL, PL1 and PL3. For the three latter methods with tuning parameter $R = 0.1, 0.2$ and 0.3, respectively. The value closest to 0 in each row is indicated in boldface

very good estimates; see Table 1. Both minimum contrast methods have very similar performance, but MCK is always slightly better than MCg. When estimating the kernel scale parameter σ with CL, it is important to choose a reasonably small value of the tuning parameter R compared to the cluster size; compare with Fig. 2. Thus, CL with $R = 0.1$ performs better than CL with larger values of R. CL with $R = 0.1$ is also practically unbiased, the small positive bias is in the majority of cases the smallest among the biases of all the considered estimators. In contrast, PL1 does not depend very much on the value of R. For models with looser clusters ($\sigma = 0.02, 0.03$), PL1 has the worst performance of all the estimators. It always has a large negative bias. For $\sigma = 0.02, 0.03$, the bias is always substantially larger than for any other estimator. As for CL, the performance of PL3 depends on R, primarily for loose clusters ($\sigma =$

Table 3 Relative mean squared errors (upper row) and relative mean biases (lower row) of the estimators of θ, determined by simulation of the point process models with the specified combinations of the parameters μ, θ and σ, shown in the left column

μ	θ	σ	MCK	MCg	CL			PL1			PL3			$\hat{\beta}_1$
					0.1	0.2	0.3	0.1	0.2	0.3	0.1	0.2	0.3	
25	1/20	0.01	0.677	0.730	0.792	0.648	**0.637**	0.799	0.799	0.799	0.719	0.719	0.719	0.498
			0.387	0.414	**0.293**	0.305	0.321	0.460	0.460	0.460	0.409	0.409	0.409	−0.065
25	1/20	0.02	0.690	0.728	0.790	0.705	**0.668**	1.01	1.01	1.01	0.739	0.737	0.737	0.492
			0.363	0.384	**0.249**	0.269	0.277	0.552	0.553	0.553	0.400	0.397	0.397	−0.008
25	1/20	0.03	1.03	1.07	1.15	**0.778**	0.800	1.93	2.05	2.05	1.08	1.14	1.14	0.507
			0.484	0.506	**0.274**	0.298	0.310	0.843	0.884	0.884	0.508	0.509	0.505	−0.007
25	1/30	0.01	0.621	0.642	0.800	0.623	**0.569**	0.670	0.670	0.670	0.611	0.611	0.611	0.541
			0.314	0.323	0.254	0.250	**0.249**	0.367	0.367	0.367	0.329	0.329	0.329	−0.007
25	1/30	0.02	0.635	0.661	0.674	0.644	**0.627**	0.929	0.931	0.931	0.692	0.690	0.690	0.441
			0.318	0.331	0.227	**0.217**	0.223	0.510	0.510	0.510	0.360	0.354	0.354	0.044
25	1/30	0.03	0.872	0.889	0.897	**0.817**	0.882	1.88	1.99	1.99	1.07	1.10	1.10	0.511
			0.436	0.446	**0.237**	0.285	0.325	0.806	0.847	0.847	0.480	0.486	0.486	0.002
50	1/10	0.01	**0.229**	0.262	0.280	0.230	0.235	0.277	0.277	0.277	0.241	0.241	0.241	0.272
			0.190	0.224	**0.139**	0.143	0.154	0.252	0.252	0.252	0.201	0.201	0.201	0.028
50	1/10	0.02	0.320	0.352	**0.302**	0.363	0.375	0.557	0.559	0.559	0.357	0.361	0.361	0.257
			0.210	0.235	**0.109**	0.150	0.153	0.412	0.413	0.413	0.229	0.224	0.224	0.026
50	1/10	0.03	0.631	0.658	0.776	**0.630**	0.660	1.82	1.89	1.89	0.808	0.828	NA	0.266
			0.328	0.349	**0.096**	0.195	0.227	0.839	0.879	0.879	0.376	0.375	NA	0.030
50	1/20	0.01	**0.198**	0.209	0.231	**0.198**	0.201	0.254	0.254	0.254	0.220	0.220	0.220	0.263
			0.180	0.191	**0.129**	0.137	0.145	0.249	0.249	0.249	0.200	0.200	0.200	−0.024
50	1/20	0.02	**0.291**	0.304	0.304	0.318	0.323	0.508	0.509	0.509	0.316	0.318	0.318	0.245
			0.208	0.221	**0.134**	0.156	0.165	0.406	0.407	0.407	0.222	0.216	0.216	−0.017
50	1/20	0.03	0.381	0.386	**0.366**	0.380	0.455	1.17	1.24	1.24	0.457	0.479	0.480	0.245
			0.244	0.252	**0.121**	0.149	0.156	0.711	0.753	0.753	0.275	0.275	0.273	0.025

The estimation methods considered are MCK, MCg, CL, PL1 and PL3. For the three latter methods with tuning parameter $R = 0.1, 0.2$ and 0.3, respectively. The value closest to 0 in each row is indicated in boldface. The last column shows the relative mean squared errors (upper row) and relative mean biases (lower row) of the estimated inhomogeneity parameter $\hat{\beta}_1$

0.02, 0.03) where it is important not to choose R too large. In one case, the estimate of σ cannot be determined for the large value of $R = 0.3$ due to numerical instability of the estimation procedure. MCK had the best overall performance (according to MSE).

8.2.2 *Estimation of μ*

The parameter μ is harder to estimate than σ and the performance of all the estimators shows the same trends in the dependence of the model parameter values; see Table 2.

The MSEs of the estimators increase with looser clusters (growing σ) and smaller number of observed points (growing θ or smaller μ). The minimum contrast methods perform also for μ very similarly, but MCK is always slightly better than MCg. In particular, MCK is less biased than MCg. CL has again the smallest bias among all the methods. The performance of CL depends on the value of the tuning parameter R and, generally, a higher precision of the estimates of μ is obtained for the larger values of $R = 0.2, 0.3$ than for estimation of σ. PL1 does not perform well. In particular, PL1 has a very large bias which grows with the model parameter σ. The performance of PL3 is comparable to that of CL and always better than that of PL1. Its performance depends only slightly on the tuning parameter R. The overall best performance (according to MSE) is again showed by MCK. All the estimators overestimate μ but the bias of MCK, MCg and PL3 is comparable (smaller than the bias of PL1 and larger than the bias of CL).

8.2.3 Estimation of θ

The parameter θ governs the distribution of the number of points in the observed clusters (or the weight of the clusters) and is the parameter hardest to estimate. A large number of observed points is necessary to estimate it well. For all estimation methods, θ is computed from Eq. (24), using $\hat{\beta}_1$ and $\hat{\mu}$. The quality of $\hat{\theta}$ depends on the quality of $\hat{\mu}$ and $\hat{\beta}_1$. Table 3 shows in the last column the MSE and bias of $\hat{\beta}_1$. Note that the MSE of $\hat{\beta}_1$ is quite large, especially for point patterns with smaller number of points and loose clusters. For all the estimators, the precision decreases with looser clusters (growing σ) and smaller number of observed points (growing θ or smaller μ). Between the MC methods, MCK is always slightly better than MCg. The best estimates of θ are obtained by MCK in three models considered in the simulation study $[(\mu, \theta, \sigma) = (50, 1/10, 0.01), (50, 1/20, 0.01), (50, 1/20, 0.02)]$, in all the other models CL with an appropriate value of R produces the best estimates of θ. In most cases, PL3 shows similar behavior as CL and is superior to PL1. All the methods overestimate the value of θ, CL has the smallest bias.

8.2.4 Further observations

Even though both L_{P1} and L_{P3} lead to unbiased estimating equations, the estimates of the parameters μ and θ governing the mean number and the distribution of the weights of the clusters had systematically larger bias for L_{P1} than for L_{P3}. This fact can be explained as follows. Formula (13) for L_{P1} does not acknowledge the "probability of observing" the difference process Y_x around the observed point $x \in X$. This "probability of observing" Y_x is the same as the probability of observing a point of the process X at location x which is proportional to $\lambda(x)$. We have a higher probability of encountering a Y_x for x from high-intensity subareas of W. This is not acknowledged in (13) since all the difference processes Y_x have the same weight. Consequently, since Y_x from the high-intensity areas has a smaller weight than the correct one, we obtain an extra positive bias for $\hat{\mu}$ ("mean number of clusters") to compensate the discrepancy between (13) and the data. Formula (15) for L_{P3} includes the approximate "probabilities" $\hat{\lambda}(x)$ of observing Y_x. Therefore, we prefer L_{P3} to L_{P1}, particularly for obviously inhomogeneous point process data. Of course, this issue of reweighting by $\lambda(x)$ is not encountered in the stationary case described in Sect. 4.3.

As stated in the discussion for the particular parameters, a good choice of the tuning parameter R is crucial for the performance of PL estimators. The best performance of the PL1 and PL3 estimators is always obtained with $R = 0.1$. For larger $R = 0.2, 0.3$, the maximization of the Palm likelihood gets numerically less stable. We have observed a certain number of very large outlier estimates $\hat{\sigma}$ of σ. In some cases the procedure can even diverge. This happened for one-point pattern with the true parameter values $\mu = 50, \theta = 1/10, \sigma = 0.03$ and PL3 with $R = 0.3$. Therefore for this case there is NA in the tables. To a smaller extent the problem with outlier estimates and numerical instability also applies to the CL estimates with larger R (in particular $R = 0.3$).

It is also worth noting that when estimating μ and θ, the PL3 version is rather robust w.r.t. the choice of R. When these parameters are of particular interest, one may opt for the PL3 method with confidence that a possible unlucky choice of R would not compromise the resulting estimates.

Concerning the overall numerical complexity of the compared estimation methods, the fastest are the MCK and MCg methods as implemented in `Spatstat`. CL and PL estimates are somewhat slower to compute because of the grid search for σ. They have comparable computation time that increases with increasing value of the tuning parameter R, since more data from $X \cap W$ need to be incorporated.

We have also studied the correlation between the estimators. In all cases we get negative correlation between $\hat{\sigma}$ and $\hat{\mu}$. The absolute value of the correlation ranges between 0.2 and 0.3 for the tight clusters case with $\sigma = 0.01$, around 0.5 for $\sigma = 0.02$ and grows up to $0.6-0.7$ for the loose clusters with $\sigma = 0.03$. This is nicely explainable by the fact that with larger σ we observe "looser" and therefore also less distinguishable clusters in the point pattern. Thus the larger the estimated size $\hat{\sigma}$ of the clusters, the smaller the estimated number $\hat{\mu}$ of the clusters. The smallest correlation (in absolute value) is always obtained by the MCK and MCg estimators, the CL and PL estimators usually have 10 % larger correlation.

Since $\hat{\theta}$ is derived from $\hat{\mu}$, the correlation between $\hat{\sigma}$ and $\hat{\theta}$ follows the same pattern as the correlation between $\hat{\sigma}$ and $\hat{\mu}$. The only difference is that it is uniformly approximately 10 % smaller in absolute value in all the cases. This loss in the dependence is explainable by the transformation and the use of the total number of observed points in $X \cap W$ (a quantity not used for estimation of the other two parameters).

8.3 Asymptotic standard deviations for Palm likelihood estimates

In Theorem 1 we have derived the asymptotic normality for the Palm likelihood estimators PL1 and PL3 when the observation window W_n expands towards $\mathbb{R}^2$. To check the applicability of the asymptotics for finite sample sizes we have computed the asymptotic standard deviations and the confidence intervals for σ and μ based on Theorem 1, using the data from the simulation study. We show the results for the PL3 method, since according to the simulation results in the preceding subsection PL3 is to be preferred.

Table 4 Approximate asymptotic standard deviations, empirical standard deviations and coverage (cvrg) of the nominal 95 % approximate confidence interval for the estimates $\hat{\sigma}$ and $\hat{\mu}$ computed by the PL3 method

μ	θ	σ	as. sd ($\hat{\sigma}$)	emp. sd ($\hat{\sigma}$)	cvrg (σ)	as. sd ($\hat{\mu}$)	emp. sd ($\hat{\mu}$)	cvrg (μ)
25	1/20	0.01	0.0008	0.0010	0.95	9.8	7.6	0.97
		0.02	0.0033	0.0024	0.98	12	9.6	0.95
		0.03	0.0042	0.0040	0.96	11	12	0.90
25	1/30	0.01	0.0010	0.0015	0.95	9.4	7.7	0.97
		0.02	0.0026	0.0025	0.96	9.2	9.8	0.90
		0.03	0.0041	0.0034	0.98	11	11	0.90
50	1/10	0.01	0.0009	0.0010	0.93	13	13	0.92
		0.02	0.0028	0.0027	0.96	17	19	0.89
		0.03	0.0045	0.0041	0.97	20	25	0.88
50	1/20	0.01	0.0007	0.0010	0.95	14	13	0.95
		0.02	0.0025	0.0022	0.96	16	19	0.93
		0.03	0.0044	0.0038	0.98	21	22	0.92

For details see Sect. 8.3

When applying Theorem 1 the information matrix I_n and the covariance matrix Σ_n must be computed. The matrix I_n can be computed by numerical integration from the formulae (19)–(21). In principle, the covariance matrix Σ_n can be computed by numerical integration from the formulae at the end of Sect. 7. However, these formulae include intensity functions of the 3rd and 4th order which are quite complicated or even unavailable in closed form for a particular model. An alternative is to use the empirical covariance matrix of the score function based on the simulations. By this method we obtained the asymptotic standard deviations of $\hat{\sigma}$ and $\hat{\mu}$.

In Table 4, the asymptotic standard deviations are compared with the standard deviations of $\hat{\sigma}$ and $\hat{\mu}$ computed from the simulations. Further we have determined the fraction of the estimates which fall into the approximate 95 % confidence interval, centered around the theoretical value of the estimated parameter and with length determined by the asymptotic standard deviation. This fraction is reported as coverage in Table 4. Note that the parameter θ is not included in Table 4 because Theorem 1 is not directly applicable to this parameter which is estimated using $\hat{\beta}_1$ and $\hat{\mu}$ in (24).

The simulation results in Table 4 show good agreement between the empirical standard deviations and the asymptotic standard deviations from the simulated parameter estimates. The coverage of the confidence interval are also close to the nominal 95 %. Note that the results are somewhat better for the easier-to-estimate parameter σ than for μ.

9 Conclusions and perspectives

The performance of the different two-step estimators has been compared in a simulation study of several inhomogeneous gamma shot-noise Cox processes. The conclusion is that the performance of the four methods MCK, MCg, CL and PL3 is nearly equivalent. All of them are able to estimate the interaction parameter σ (the scale of the kernel k) very precisely. The parameter μ is also estimated well by all the methods. The worst performance was observed when estimating the distribution of the weights of the clusters (i.e., parameter θ).

Further we have investigated how the asymptotic theory derived for the Palm likelihood estimator in Sect. 7 applies in the situation of the simulation study. We have computed the approximate asymptotic standard deviations and confidence intervals based on the asymptotic normality derived in Theorem 1. They show good agreement with their empirical counterparts computed directly from the simulations.

The inhomogeneous spatial point process model studied in the present paper is an example from the most well-known class of inhomogeneous point processes: the second-order intensity-reweighted stationary processes, originally proposed in Baddeley et al. (2000). It is, of course, important to be able to test the fit of such a model. Permutation and bootstrap tests have recently been developed for this purpose; see Hahn and Jensen (2015). It is also worth stressing, that is not necessary to use a parametric form of the inhomogeneity function if the user feels uncomfortable with that. Instead, the inhomogeneity function may be estimated directly from data using a nonparametric estimate of the intensity function. This alternative procedure has been tried out in Hahn and Jensen (2015).

Acknowledgements This project has been supported by the Czech Science Foundation, Project No. P201/10/0472, Charles University Grant Agency, Project No. 664313, and by Centre for Stochastic Geometry and Advanced Bioimaging, funded by a Grant from The Villum Foundation. We thank the referees for constructive comments and suggestions both concerning the content and exposition of the paper.

References

Baddeley, A. J., Turner, R. (2005). Spatstat: An R package for analyzing spatial point patterns. *Journal of Statistical Software*, *12*, 1–42.

Baddeley, A. J., Møller, J., Waagepetersen, R. (2000). Non- and semiparametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, *54*, 329–350.

Baddeley, A. J., Rubak, E., Turner, R. (2015). *Spatial point patterns: Methodology and applications with R*. London: Chapman and Hall/CRC Press.

Chiu, S. N., Stoyan, D., Kendall, W. S., Mecke, J. (2013). *Stochastic geometry and its applications* (3rd ed.). Chichester: Wiley.

Cox, D. R. (1955). Some statistical models related with series of events. *Journal of the Royal Statistical Society: Series B*, *17*, 129–164.

Daley, D. J., Vere-Jones, D. (2003). *An introduction to the theory of point processes. Volume 1: Elementary theory and methods*. New York: Springer.

Daley, D. J., Vere-Jones, D. (2008). *An introduction to the theory of point processes. Volume 2: General theory and structure*. New York: Springer.

Diggle, P. J. (1983). *Statistical analysis of spatial point patterns*. London: Academic Press.

Diggle, P. J. (2003). *Statistical analysis of spatial point patterns* (2nd ed.). New York: Oxford University Press.

Dvořák, J., Prokešová, M. (2012). Moment estimation methods for stationary spatial Cox processes—A comparison. *Kybernetika*, *48*, 1007–1026.

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

Guan, Y. (2009). A minimum contrast estimation procedure for estimating the second-order parameters of inhomogeneous spatial point processes. *Statistics and its Interface*, *2*, 91–99.

Guan, Y., Sherman, M. (2007). On least squares fitting for stationary spatial point processes. *Journal of the Royal Statistical Society: Series B*, *69*, 31–49.
Guyon, X. (1991). *Random fields on a network*. New York: Springer.
Hahn, U., Jensen, E. B. V. (2015). Hidden second-order stationary spatial point processes. *Scandinavian Journal of Statistics*,. doi:10.1111/sjos.12185.
Heinrich, L. (1988). Asymptotic Gaussianity of some estimators for reduced factorial moment measures and product densities of stationary Poisson cluster processes. *Statistics*, *19*, 87–106.
Heinrich, L. (1992). Minimum contrast estimates for parameters of spatial ergodic point processes. In: Transactions of the 11th Prague Conference on Random Processes, Information Theory and Statistical Decision Functions. Prague: Academic Publishing House.
Hellmund, G., Prokešová, M., Jensen, E. B. V. (2008). Lévy-based Cox point processes. *Advances in Applied Probability*, *40*, 603–629.
Illian, J., Penttinen, A., Stoyan, H., Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*. Chichester: Wiley.
Jalilian, A., Guan, Y., Waagepetersen, R. P. (2013). Decomposition of variance for spatial Cox processes. *Scandinavian Journal of Statistics*, *40*, 119–137.
Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary Mathematics*, *80*, 221–239.
Matérn, B. (1971) Doubly stochastic Poisson processes in the plane. In: Statistical ecology (volume 1). University Park: Pennsylvania State University Press.
Møller, J. (2003). Shot noise Cox processes. *Advances in Applied Probability*, *35*, 614–640.
Møller, J., Waagepetersen, R. P. (2003). *Statistical inference and simulation for spatial point processes*. Boca Raton: Chapman & Hall/CRC.
Møller, J., Waagepetersen, R. P. (2007). Modern statistics for spatial point processes. *Scandinavian Journal of Statistics*, *34*, 643–684.
Prokešová, M., Dvořák, J. (2014). Statistics for inhomogeneous space-time shot-noise Cox processes. *Methodology and Computing in Applied Probability*, *16*, 433–449.
Prokešová, M., Jensen, E. B. V. (2013). Asymptotic Palm likelihood theory for stationary point processes. *Annals of the Institute of Statistical Mathematics*, *65*, 387–412.
Tanaka, U., Ogata, Y., Stoyan, D. (2007). Parameter estimation and model selection for Neyman–Scott point processes. *Biometrical Journal*, *49*, 1–15.
Thomas, M. (1949). A generalization of Poisson's binomial limit for use in ecology. *Biometrika*, *36*, 18–25.
Waagepetersen, R. P. (2007). An estimating function approach to inference for inhomogeneous Neyman–Scott processes. *Biometrics*, *63*, 252–258.
Waagepetersen, R. P., Guan, Y. (2009). Two-step estimation for inhomogeneous spatial point processes. *Journal of the Royal Statistical Society: Series B*, *71*, 685–702.