www.csgb.dk

RESEARCH REPORT

2014

Michaela Prokešová, Jiří Dvořák and Eva B.V. Jensen

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

Two-step estimation procedures for inhomogeneous shot-noise Cox processes[*]

Michaela Prokešová[1], Jiří Dvořák[1] and Eva B.V. Jensen[2]

[1]Department of Probability and Mathematical Statistics, Charles University in Prague
[2]Department of Mathematics, Centre for Stochastic Geometry and Advanced Bioimaging, Aarhus University

Abstract

In the present paper we develop 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. Two-step estimation with Palm likelihood has not been considered before. 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, Palm likelihood, Asymptotic normality

1 Introduction

Cox point processes (sometimes also called doubly stochastic point processes in the literature) are the preferred point process models for analysis of clustered point patterns ([3, 4, 6, 16, 20, 22, 23, 26]). 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 fall into 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., [22]). 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 [21] and further generalized in [15] without discussing the statistical inference for the model. Note that the class of shot-noise Cox processes also includes the very popular Poisson Neyman-Scott processes like the Thomas process ([28], [16, 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 and data case, see e.g., [23, Section 7.3] for an example. Therefore, the easier to compute moment estimation methods (eventhough less efficient than the maximum likelihood estimation) are often preferred in the applications.

Several moment estimation methods applicable to the stationary shot-noise Cox processes are available in the literature: minimum contrast estimation ([5, Chapter 6]), composite likelihood ([8]), Palm likelihood ([25], [27]). According to simulation studies, as the ones presented in [7] and [8], 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 the applied point of view) a two-step estimation procedure was introduced in [30] where first the inhomogeneous 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 [10], the same two-step estimation procedure was investigated with minimum contrast based on the inhomogeneous g-function in the second step. The main assumption of this two-step estimation procedure is the second-order intensity-reweighted stationarity (SOIRS) of the inhomogeneous processes to be analyzed. SOIRS implies existence of a well defined inhomogeneous g- and K-function (see [1]) used in the second step of the estimation procedure. However, the decomposition of the second-order intensity function $\lambda^{(2)}$

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

into a product of the first-order intensity function and the inhomogenous paircorrelation function g implied by SOIRS enables a generalization of the other estimation methods from the stationary case to the SOIRS case as well. In a recent paper ([17]), 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 the 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 Section 2 while shot-noise Cox processes are introduced in Section 3.

An overview of moment estimation methods for stationary Cox processes is given in Section 4. These methods are adapted to the inhomogeneous case in Section 5. In Section 6, the focus is on two-step estimation with Palm likelihood and in Section 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 Section 8.

2   Background

In this section, we briefly introduce the basic notions relating to spatial point processes needed in the following, including first- and second-order properties. For more detailed information, see the standard references [4] and [26].

Let $\mathcal{B}(\mathbb{R}^d) = \mathcal{B}^d$ be the Borel subsets of $\mathbb{R}^d$. Let X be a point process on $\mathcal{X} \in \mathcal{B}^d$. For $A \in \mathcal{B}^d$, $|A|$ 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 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.

For any given point $u \in \mathbb{R}^d$, let $\mathrm{d}u$ be the infinitesimal region that contains the point u. Following [6] we can define the (first-order) intensity function λ of X by

$$\lambda(u) = \lim_{|\mathrm{d}u| \to 0} \left(\frac{\mathsf{E} |X \cap \mathrm{d}u|}{|\mathrm{d}u|} \right), \tag{2.1}$$

so that $\lambda(u) \, \mathrm{d}u$ is the mean number of points from X occurring in $\mathrm{d}u$. The second-order intensity function $\lambda^{(2)}(u, v)$ is defined by

$$\lambda^{(2)}(u, v) = \lim_{|\mathrm{d}u|, |\mathrm{d}v| \to 0} \left(\frac{\mathsf{E} |(X \cap \mathrm{d}u)||(X \cap \mathrm{d}v)|}{|\mathrm{d}u||\mathrm{d}v|} \right). \tag{2.2}$$

When X is simple (does not have multiple points), then $\lambda^{(2)}(u, v)|\mathrm{d}u||\mathrm{d}v|$ may for $u \neq v$ be interpreted as the probability that $\mathrm{d}u$ and $\mathrm{d}v$ each contain a point from X. Higher order intensity functions $\lambda^{(k)}$ are defined analogously. In the literature the intensity functions are also called product densities since they are in fact densities of the factorial moment measures of the process X, see [4] for details.

The point process X is stationary if its distribution is invariant with respect to the simultaneous shifts of all the points in X. Under stationarity, $\lambda(u) = \lambda$ is constant and we can write

$$\lambda^{(2)}(u, v) = \lambda^{(2)}(0, v - u) = \lambda \lambda_o(v - u). \tag{2.3}$$

Thus, the second order intensity function can be reduced to an equivalent function of only one argument and, moreover, the function λ_o is well defined by the decomposition in equation (2.3). The function λ_o is in fact equal to the (first-order) intensity function of the Palm distribution of X and is therefore sometimes called the Palm intensity. Recall that the Palm distribution is the distribution of X conditioned by the occurrence of a point from X at the origin, see [4] for details.

Two important characteristics may be defined by means of $\lambda^{(2)}$. The first one is the pair-correlation function $g(u,v) = \frac{\lambda^{(2)}(u,v)}{\lambda^2}$ which is also sometimes called simply the g-function. Because of the reducibility (2.3) of $\lambda^{(2)}$, the g-function of a stationary point process is a function of just one argument $g(u,v) = g(u-v)$, $u,v \in \mathbb{R}^d$. The second characteristic is the K-function defined by

$$K(r) = \int_{\|u\|<r} g(u)\,\mathrm{d}u = \int_{B(o,r)} g(u)\,\mathrm{d}u, \qquad r > 0. \qquad (2.4)$$

It can be shown that $\lambda K(r)$ is the mean number of further points of the point pattern in a ball $B(x,r)$ centered at a typical point x of the point process.

There are several ways to define an inhomogeneous point process, including inhomogeneity introduced by transformation [18] or local scaling [12], but in the sequel we will only deal with the most often used type of inhomogeneity – the second-order intensity-reweighted stationarity (SOIRS) – which was introduced in [1]. It is characterized by the fact that the inhomogeneous g-function $g(u,v) = \frac{\lambda^{(2)}(u,v)}{\lambda(u)\lambda(v)}$ is translation invariant and thus equal to a well-defined function of only one argument $v - u$. Under SOIRS, we can decompose $\lambda^{(2)}$ as follows:

$$\lambda^{(2)}(u,v) = \lambda(u)\lambda(v)g(v-u) = \lambda(u)\lambda_u(v) = \lambda(v)\lambda_v(u), \qquad (2.5)$$

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

The inhomogeneous K-function is defined by the relation (2.4) used in the stationary case but it does not have the simple interpretation from the stationary case anymore.

3 Shot-noise Cox processes

A Cox point process on a set $\mathcal{X} \in \mathcal{B}^d$ is a doubly-stochastic process which, conditionally on the realization of the random driving field $\Lambda(u), u \in \mathcal{X}$, is a Poisson process with intensity function Λ.

A shot-noise Cox process X has a driving field of the form

$$\Lambda(u) = \sum_{(r,v) \in \Pi_U} rk(u,v), \quad u \in \mathcal{X}, \qquad (3.1)$$

where Π_U is a Poisson measure on $\mathbb{R}^+ \times \mathbb{R}^d$ with intensity measure U and k is a smoothing kernel, i.e., a non-negative function integrable in both coordinates. Under some basic integrability assumptions (3.1) is an almost surely locally integrable field and X is a well-defined Cox process, see [21] and [15] for details.

The shot-noise Cox process X is stationary if the kernel k is just a function of the difference of the two arguments $k(u,v) = k(v-u)$ and the measure U has the form $U(\mathrm{d}(r,v)) = \mu V(\mathrm{d}r)\mathrm{d}v$, where $\mu > 0$ and $V(\mathrm{d}r)$ may be 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 recovered when V is equal to the Dirac measure.

Example 1 (Poisson cluster process)**.** If $V(\mathrm{d}r) = \delta_1(\mathrm{d}r)$ is simply a Dirac measure concentrated in 1, then X is a Poisson cluster process with cluster centers coming from a stationary Poisson process on $\mathbb{R}^d$ with intensity μ (given by U). Let further $k(u) = c\tilde{k}(u)$ where $c > 0$ and $\tilde{k}$ is a probability density on $\mathbb{R}^d$. Then, conditionally on the positions of the cluster centers, the clusters are independent with Poisson distributed number of points with mean value $c = \int k(v)\,\mathrm{d}v$ and the points within the cluster are distributed independently, according to the probability density $\tilde{k}$ around the cluster center. Thus in this case we get the class of Poisson Neyman-Scott processes, see e.g., [16, Section 6.3.2], which also includes the well-known Thomas process with the Gaussian kernel k ([28]). Fig. 1, left panel, shows a realization of a stationary Thomas process. $\qquad\square$

In fact, all shot-noise Cox processes can be viewed as generalized cluster processes. We can rewrite the shot-noise Cox process X as $X = \bigcup_{(r,v)\in\Pi_U} X_v$ where X_v is the cluster centered around a point located at v. Conditionally on Π_U, the cluster processes X_v are independent Poisson processes with intensity function $rk(\cdot, v)$. Note that even for a compact set A the number of cluster centers in A may be infinute almost surely, see [21], [15] for details. Therefore, shot-noise Cox processes may be regarded as a generalization of the standard cluster processes ([16, Section 6.3]).

If we assume the stationary case for simplicity then, under the condition that $\int_{\mathbb{R}^+} rV(\mathrm{d}r) < \infty$, almost surely only a finite number of the clusters X_v will have nonzero number of points. If we condition by the positions of the centers only (and not by the whole Π_U) and assume for simplicity that k is a probability density, then the shifted cluster processes $(X_v - v)$ are independent identically distributed, and the number of points in a cluster has a mixed Poisson distribution with mixing distribution governed by the measure V.

Thus, the measure V determines the distribution of the number of points in the clusters. The standard cluster processes produce point patterns with pretty homogenous clusters. By choosing an appropriate measure V, we can obtain a much more variable number of points in the clusters than in Example 1.

Example 2 (Gamma shot-noise Cox process)**.** Let V be defined by $V(\mathrm{d}r) = r^{-1}\exp(-\theta r)\,\mathrm{d}r$ where $\theta > 0$ is a parameter. Note that V is not integrable in the neighbourhood of 0. As a consequence, the corresponding shot-noise Cox process X is not a cluster process in the classical sense ([16, 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 ([15, Section 4]). Fig. 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 variablity in the number of points in different clusters. $\qquad\square$
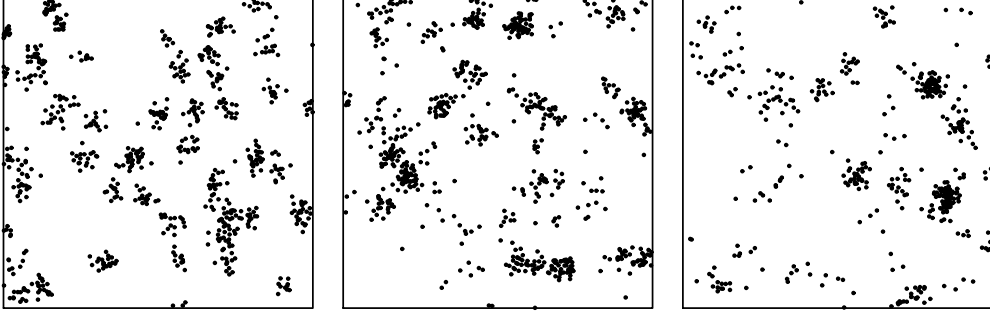
Figure 1: Realizations of (from left to right) a stationary Thomas process, a stationary gamma shot-noise Cox process and an inhomogeneous gamma shot-noise Cox process. For details, see the text.

The moment properties of the shot-noise Cox processes are easily available ([15, Section 4]), in particular for the intensity function we have

$$\lambda(u) = \mu \int_{\mathbb{R}^+} r\, V(\mathrm{d}r) \int_{\mathbb{R}^d} k(u,v)\, \mathrm{d}v, \qquad (3.2)$$

and for the pair-correlation function

$$g(u,v) = 1 + \frac{\mu \int_{\mathbb{R}^+} r^2 V(\mathrm{d}r) \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} k(u,w)k(v,w)\, \mathrm{d}w}{\lambda(u)\, \lambda(v)}. \qquad (3.3)$$

Note that in both equations we have a product of separate integrals for V and k – this will be important in the estimation procedures developed in Sections 5 and 6. Moreover, for the parametric models as the ones described in the examples above both integrals with respect to V are simple functions of the model parameters. In Examples 1 and 2, $\int_{\mathbb{R}^+} r\, V(\mathrm{d}r) = 1, 1/\theta$ and $\int_{\mathbb{R}^+} r^2 V(\mathrm{d}r) = 1, 1/\theta^2$, respectively.

When we apply a location dependent thinning with inhomogeneity function $f(u)$ to a stationary shot-noise Cox process specified by μ, V and k, a new SOIRS shot-noise Cox process is obtained with the same μ and V, but with a new kernel function $\tilde{k}(u,v) = f(u)k(v - u)$. In the rest of the paper we will consider such SOIRS shot-noise Cox processes. It turns out to be useful to use the parametrization by the homogeneous kernel function $k(v - u)$ and the inhomogeneity function f, instead of using the inhomogeneous kernel function $\tilde{k}$. One of the reasons is that the inhomogeneity function f does not enter into the formula for the g-function, since we have

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

For an example of a realization of an inhomogeneous SOIRS gamma shot-noise Cox process, see Fig. 1, right panel. The point process is a location dependent thinning of the point process described in Ex. 2 and shown in Fig. 1, middle panel. The inhomogeneity in the direction of the x-axis is clearly visible as well as the effect of the thinning procedure on the clusters in the low intensity areas of the observation window. For a more detailed description of the model, see Section 8.

6

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 it.

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 [5, Chapter 5]. It can be based either on the K-function or the pair correlation function g, see e.g., [6, 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

$$\int_r^R \left[\hat{K}^q(u) - K^q(u; \eta) \right]^2 \mathrm{d}u \qquad \text{or} \qquad \int_r^R \left[\hat{g}^q(u) - g^q(u; \eta) \right]^2 \mathrm{d}u \qquad (4.1)$$

between the nonparametric estimate $\hat{K}$ or $\hat{g}$ and its 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 estimates of K and g. Recommendations concerning the choice of tuning parameters and other practical aspects can be found in [6, Section 6.1.1]. Asymptotic properties of the minimum contrast estimator, based on the K-function, are discussed in [9] and [14] for stationary case. In [14] strong consistency and asymptotic normality for minimum contrast estimators, based on the K-function, was proved for stationary Poisson cluster processes. In [9] 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 ([19]). 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 ([1], [8], [23]). Composite likelihood suitable for estimation of Cox processes was introduced in [8]. 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:

$$f(x, y; \eta) = \frac{\lambda^{(2)}(y - x; \eta)}{\int_W \int_W \lambda^{(2)}(u - v; \eta) \mathrm{d}u\mathrm{d}v}. \tag{4.2}$$

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

$$\begin{aligned}
&\log CL(\eta) \\
&= \sum_{x,y \in X \cap W, 0 < \|y-x\| < R} \Big[\log \lambda^{(2)}(y - x; \eta) \\
&\qquad - \log\Big(\int_W \int_W \lambda^{(2)}(u - v; \eta) I(\|u - v\| < R) \,\mathrm{d}u\,\mathrm{d}v \Big) \Big],
\end{aligned} \tag{4.3}$$

Here, only pairs of points with distance less than R are considered. Disregarding the pairs of points separated by distance larger than R 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 [8] under suitable mixing assumptions.

Note that in the stationary case the squared intensity λ^2 cancels in (4.2) so that

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

and (4.3) 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 [27] and uses a very "geometrical" approach. It is based on the process of differences among 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 [26, p. 126] for further details). The point process $Y(R)$ has thus an intensity function concentrated on $B(o, R)$ of the form

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

The Palm log-likelihood

$$\log L_P(\eta) = \sum_{x \neq y \in X \cap W, (y-x) \in B(o,R)} \log\left(|X \cap W|\lambda_o(y-x;\eta)\right)$$
$$- |X \cap W| \int_{B(o,R)} \lambda_o(r;\eta)\,\mathrm{d}r, \quad (4.4)$$

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

$$\log L_P(\eta) = \sum_{x \neq y \in X \cap W, ||x-y|| < R} \log \lambda_o(x-y;\eta) - |X \cap W| \int_{B(o,R)} \lambda_o(r;\eta)\,\mathrm{d}r. \quad (4.5)$$

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. An extension of the Palm likelihood estimation to non-isotropic stationary point processes was introduced in [25]. Moreover, strong consistency and asymptotic normality of the Palm likelihood estimator are proved for stationary Cox processes in [25] 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 (2.5) of $\lambda^{(2)}$, implying the existence of a well-defined inhomogeneous g-function and

K-function, that can be estimated from the data once we know the intensity function $\lambda(u)$.

Following these ideas, [30] introduced a two-step estimation procedure where first the inhomogeneous 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, [10] used minimum contrast estimation with the inhomogeneous g-function in the second step.

The minimum contrast estimation based on the K-function (MCK) is definitively the most frequently used method in the stationary case, but this method is actually not necessarily the most efficient. Simulation studies in [8] and [7] 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 when the total number of points observed in different clusters vary a lot. 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.

Therefore, there is a need for deriving two-step estimators for the inhomogeneous case, based on the other methods from Section 4. For composite likelihood it was done in a recent paper [17]. For the Palm likelihood we will introduce the new two-step estimator in Section 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 product density 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 we do not have overspecification in the model. Below, we show an example of such a separation.

Example 2 (continued). 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 $\frac{h_\beta(x)}{\max_{v \in W} h_\beta(v)}$.

Then, the resulting inhomogeneous shot-noise Cox process Y has first-order intensity function

$$\lambda_\beta(u) = \frac{\mu}{\theta} \frac{h_\beta(u)}{\max_{v \in W} h_\beta(v)}$$

and inhomogeneous pair-correlation function

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

In applications, the intensity function has often a log-linear form

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

where $z(u)$ is a vector of covariates observed at the location u. When we reparametrize the intensity function λ_β as

$$\lambda_\beta(u) = \exp(\beta_0)h_\beta(u), \tag{5.1}$$

where $\beta_0 = \log(\frac{\mu}{\theta}/\max_{v \in W} h_\beta(v))$, then λ is parametrized by the inhomogeneity parameter $(\beta_0, \beta_1, \ldots \beta_{t-1}) \in \mathbb{R}^t$ and the (inhomogeneous) pair-correlation function is parametrized by the interaction parameter $\eta = (\sigma, \mu)$. $\qquad\square$

The two-step estimation procedure in [30] can be described as follows. At first, the inhomogeneity parameter β is estimated by disregarding the interaction in the model, using the Poisson log-likelihood

$$\log L_1(\beta) = \sum_{x \in X \cap W} \log \lambda_\beta(x) - \int_W \lambda_\beta(u)\,\mathrm{d}u \tag{5.2}$$

only. The value $\widehat{\beta}$ 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 K-function is well defined for the inhomogeneous case and, using the estimate $\hat{\lambda}$ of the intensity function, it is possible to estimate the K-function of the observed process X by

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

where $w_{x,y}$ is an edge correction weight (see [1]). Analogously, it is possible to estimate the inhomogeneous g-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 [10] for the exact formula. Of course, the precision of the estimates of K and g depends heavily on the precision of $\widehat{\lambda}$. 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 (4.1) can be employed for the estimation of the interaction parameters η in the same way as for the homogeneous case.

In [29], 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 (4.3) 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 rest of the interaction parameters η conditionally on $\widehat{\beta}$ fixed.

The two-step composite likelihood estimation was suggested in [17]. Here, formula (4.3) is rewritten as

$$
\log CL(\eta) = \sum_{\substack{x,y \in X \cap W \\ 0 < \|x-y\| < R}} \Big[\log(\hat{\lambda}(x)\hat{\lambda}(y)g_\eta(y-x))
$$
$$
- \log\Big(\int_W \int_W \hat{\lambda}(u)\hat{\lambda}(v)g_\eta(u-v)I(\|u-v\| < R)\,\mathrm{d}u\,\mathrm{d}v \Big) \Big], \tag{5.3}
$$

and (5.3) with a fixed value $\hat{\lambda}$ of the intensity function from the first step is then maximized with respect to the interaction parameter η. 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 (4.3) 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 (5.2). However, in order to estimate the interaction parameters, we need to generalize the Palm likelihood (4.5) to the inhomogeneous case and this is not a straightforward problem. There are, in fact, several possibilities.

The first option is to mimic formula (4.5) 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 $x \in X \cap W$ thus arriving at

$$
\log L_{P1}(\eta) = \sum_{\substack{x,y \in X \cap W \\ 0 < \|x-y\| < R}} \log(\hat{\lambda}(y)g_\eta(y-x)) - \sum_{x \in X \cap W} \int_{B(x,R)} \hat{\lambda}(u)g_\eta(u-x)\,\mathrm{d}u. \tag{6.1}
$$

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

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

Thus, L_{P1} is actually equal to the composite loglikelihood 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 $\lambda_x(u)$ and it 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

$$
\begin{aligned}
\log L_{P2}(\eta) = &\sum_{\substack{z = w - y : w, y \in X \cap W \\ 0 < \|z\| < R}} \log\Big(\sum_{x \in X \cap W} \hat{\lambda}(x + z)g_\eta(z) \Big) \\
&- \int_{B(o,R)} \sum_{x \in X \cap W} \hat{\lambda}(x + u)g_\eta(u)\,\mathrm{d}u.
\end{aligned}
\tag{6.2}
$$

However note that the second term in (6.1) and (6.2) is actually the same and since $\hat{\lambda}$ does not depend on η, both (6.1) and (6.2) may be written as

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

as a function of η. 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 the stationary process X. Thus, the complete version of the homogeneous Palm likelihood (4.4) can be expressed as

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

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

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

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

$$\log L_{P3}(\eta) = \sum_{\substack{x \neq y \in Y \cap W \\ \|x-y\| < R}} \log\big(\hat{\lambda}(x)\hat{\lambda}(y)g_\eta(y-x)\big)$$
$$- \int_{B(o,R)} \int_{W \cap (W-u)} \hat{\lambda}(v)\hat{\lambda}(v+u)g_\eta(u)\,\mathrm{d}v\,\mathrm{d}u. \tag{6.3}$$

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

$$U(\beta,\eta) = (U_1(\beta), U_2(\beta,\eta)) = 0, \tag{6.4}$$

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 (5.2),

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

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

$$U_2(\beta,\eta) = \frac{\mathrm{d}\log L_{P3}(\eta)}{\mathrm{d}\eta}$$
$$= \sum_{\substack{x \neq y \in X \cap W \\ \|y-x\| < R}} \frac{g'_\eta(y-x)}{g_\eta(y-x)} - \int_{B(o,R)} \int_{W \cap (W-u)} \lambda_\beta(v)\lambda_\beta(v+u)g'_\eta(u)\,\mathrm{d}v\,\mathrm{d}u, \tag{6.5}$$

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 (6.4) 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 (6.1), obtaining the following unbiased version

$$\log L_{P1}(\eta) = \sum_{x,y \in X \cap W, 0 < \|x-y\| < R} \log(\hat{\lambda}(y)g_\eta(y-x))$$
$$- \sum_{x \in X \cap W} \int_{B(x,R) \cap W} \hat{\lambda}(u)g_\eta(u-x)\,\mathrm{d}u. \tag{6.6}$$

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 R too large may lead to numerical instability of the maximization procedure, see Section 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 Section 8.

7 Asymptotic properties

In [30], 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 $\int_{\mathbb{R}^+} r^k V(\mathrm{d}r) < \infty$, $k \in \mathbb{N}$. Then, X has well-defined moment measures up to the k-th order and all reduced factorial cumulant measures up to the k-th order have finite total variation.*

Proof. The first statement follows from Theorem 3 and Proposition 2 in [15]. It is well-known for cluster processes (see e.g., [13]) 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 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. 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 k-th 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. $\square$

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

$$\alpha_{p_1,p_2}(m) = \sup\{\alpha(\mathcal{F}^X(A), \mathcal{F}^X(B)) : d(A, B) \geq m, |A| \leq p_1, |B| \leq p_2\}, \quad (7.1)$$

where $\mathcal{F}^X(A)$ 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 measurable sets A, B in $\mathcal{B}^d$ and

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

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 $h(v) = \mathcal{O}(|v|^{-(2d+\epsilon)})$, as $|v| \to \infty$, then $\frac{\alpha_{p,p}(m)}{\max(p,1)} \leq \mathcal{O}(m^{-d-\epsilon})$.

Proof. Let us rewrite X as $\bigcup_{(r,v)\in\Pi_U} X_v$, where X_v is the cluster centered around a point located at v with intensity function $rk(\cdot, v)$. Denote $X_1 = \bigcup_{(r,v)\in\Pi_U, v\in A} X_v$. Then, using the fact that $\mathsf{E}(X_1 \cap B) = \mu \int_{\mathbb{R}^+} r V(\mathrm{d}r) \int_A \int_B k(v, u)\,\mathrm{d}u\,\mathrm{d}v$ for any $A, B \in \mathcal{B}^d$, the proof is exactly the same as the proof of Lemma 1 in [25]. $\square$

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 [30,

Theorem 1]. In particular, Theorem 3 below covers all the point process models considered in [30]. 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} (formula (6.6)) or L_{P3} (formula (6.3)) 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 $\Sigma_n = |W_n|^{-1} \mathsf{Var}(U_n(\beta_0, \eta_0))$ 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

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

and

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

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

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

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

Theorem 3. *Let X be a SOIRS Cox process in $\mathbb{R}^2$ whose kth-order intensity functions $\lambda_\beta^{(k)}$ satisfy*

$$\lambda_\beta^{(k)}(u_1, \ldots, u_k) = \lambda^{(k)}(u_1, \ldots u_k) \prod_{i=1}^k \lambda_\beta(u_i), \tag{7.2}$$

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

$$\alpha_{p_1,p_2}^F(m) = \sup \left\{ \alpha(\mathcal{F}^X(B_1), \mathcal{F}^X(B_2)) : B_1 = \bigcup_{M_1} A_{i,j}, B_2 = \bigcup_{M_2} A_{i,j}, \right.$$

$$\left. |M_1| \le p_1, |M_2| \le p_2, d(M_1, M_2) \ge m, M_1, M_2 \subset \mathbb{Z}^2 \right\},$$

where $d(M_1, M_2)$ denotes the minimal distance between M_1 and M_2 in the grid $\mathbb{Z}^2$ and $\alpha(\mathcal{F}_1, \mathcal{F}_2)$ 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, for all* $u_1, u_2 \in \mathbb{R}^2$, $\int |\lambda^{(3)}(0, v, v + u_1) - \lambda^{(1)}(0)\lambda^{(2)}(0, u_1)| \, dv < K_2$ *and* $\int |\lambda^{(4)}(0, u_1, v, v + u_2) - \lambda^{(2)}(0, u_1)\lambda^{(2)}(0, u_2)| \, dv < K_2$;

(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 it holds* $\alpha_{2,\infty}^F(m) = \mathcal{O}(m^{-\delta})$ *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 turning 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 [30, Theorem 1] for the two-step estimation with minimum contrast for the K-function. But 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 [30] and it suffices for the application of the central limit theorem 3.3.1 in [11] for random fields, which is needed in the proof. $\square$

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 formulas for $\lambda^{(k)}$ in [15, 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 k-th 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 [30, Theorem 1] a stronger mixing assumption is used

(Av) there exists a constant $a > 8R^2$ such that $\alpha_{a,\infty}(m) = \mathcal{O}(m^{-\delta})$ for some $\delta > 2(2 + \nu)/\nu$.

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

$$\sup_{w \in [-m/2, m/2]^2} \left\{ \int_{\mathbb{R}^2 \setminus [-m,m]^2} k(v - w) \, dv \right\} = \mathcal{O}(m^{-\delta-2}), \tag{7.3}$$

presented in [30, Appendix E] is not sufficient for (Av). Nevertheless it is sufficient for assumption (A5), as the following lemma shows.

Lemma 4. *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 (7.3). 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 $X_1 = \bigcup_{(r,v) \in \Pi_U, v \in E_3} X_v$, $X_2 = X \backslash X_1$. Then X_1, X_2 are independent cluster processes and by standard arguments (like those in [30, Appendix E]), we get

$$\alpha(\mathcal{F}^X(E_1), \mathcal{F}^X(E_2)) \leq 5(\mathsf{E}\,|X_1 \cap E_2| + \mathsf{E}\,|X_2 \cap E_1|)$$

$$\leq 5\mu \int_{\mathbb{R}^+} rV(\mathrm{d}r) \Big(\int_{[-\frac{n}{2}, \frac{n}{2}]^2} \int_{\mathbb{R}^2 \backslash [-n,n]^2} k(u-v)\,\mathrm{d}u\,\mathrm{d}v$$

$$+ \int_{\mathbb{R}^2 \backslash [-\frac{n}{2}, \frac{n}{2}]^2} \int_{E_1} k(u-v)\,\mathrm{d}u\,\mathrm{d}v \Big)$$

$$\leq const \Big(|E_3| \sup_{v \in [-\frac{n}{2}, \frac{n}{2}]^2} \int_{\mathbb{R}^2 \backslash [-n,n]^2} k(u-v)\,\mathrm{d}u\,\mathrm{d}v$$

$$+ |E_1| \sup_{v \in E_1} \int_{\mathbb{R}^2 \backslash [-\frac{n}{2}, \frac{n}{2}]^2} k(u-v)\,\mathrm{d}u\,\mathrm{d}v \Big).$$

If m is sufficiently large such that $E_1 \subset [-n/4, n/4]^2$ we get from (7.3) that both terms on the right hand side are $O(m^{-\delta})$. This implies (A5) for $\alpha_{1,\infty}^F(m)$.

For $\alpha_{2,\infty}^F(m)$ we just need to consider $E_1 = (A_{0,0} \cup A_{i,j}) - (s/2, s/2)$ for some $(i,j) \in \mathbb{Z}^2$, $E_2 = (\mathbb{R}^2 \backslash [-n,n]^2) \backslash ([-n,n]^2 + (is, js))$ and $E_3 = [-n/2, n/2]^2 \cup ([-n/2, n/2]^2 + (is, js))$. We get by similar arguments as the ones given above

$$\alpha(\mathcal{F}^X(E_1), \mathcal{F}^X(E_2)) \leq const \Big(|E_3| \sup_{v \in [-\frac{n}{2}, \frac{n}{2}]^2} \int_{\mathbb{R}^2 \backslash [-n,n]^2} k(u-v)\,\mathrm{d}u\,\mathrm{d}$$

$$+ |E_1| \sup_{v \in (A_{0,0} - (s/2, s/2))} \int_{\mathbb{R}^2 \backslash [-\frac{n}{2}, \frac{n}{2}]^2} k(u-v)\,\mathrm{d}u\,\mathrm{d}v \Big),$$

where we have used the stationarity of X. Thus, again if $\frac{s}{2} + R < \frac{n}{4}$ holds, we get from (7.3) that both terms on the right hand side are $O(m^{-\delta})$. This implies (A5) for $\alpha_{2,\infty}^F(m)$. $\qquad \square$

The inhomogeneous shot-noise Cox process, as defined at the end of Section 3, inherits the mixing properties of the unthinned homogeneous process, since the inhomogeneous process was derived by location dependent thinning. Therefore, condition (7.3) for the homogeneous kernel k ensures that (A5) is fulfilled also for the inhomogeneous shot-noise Cox process X.

Remark. The incomplete argument in [30, 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 $\mathcal{O}(m^{-\delta})$, a universal set E_1 would be needed, which could cover all Borel sets of volume $\leq a$. 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 (7.3) 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 3 to derive approximate confidence intervals for the parameter estimates, if we are able to compute the information matrix Σ_n. Below, we give the formulas for the submatrices of the block representation, corresponding to the decomposition into the following two parts of the score function

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

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

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

For L_{P3} we get

$$\Sigma_{n,12} = \frac{1}{|W_n|} \left[\int_{W_n^3} \frac{(\lambda'_{\beta_0}(w))^T}{\lambda_{\beta_0}(w)} \frac{g'_{\eta_0}(u-v)}{g_{\eta_0}(u-v)} \right.$$
$$\times I(\|u-v\| < R)(\lambda_{\beta_0}^{(3)}(w,u,v) - \lambda_{\beta_0}(w)\lambda_{\beta_0}^{(2)}(u,v)) \, dw \, du \, dv$$
$$\left. + 2 \int_{W_n^2} (\lambda'_{\beta_0}(u))^T g'_{\eta_0}(u-v) I(\|u-v\| < R) \lambda_{\beta_0}(v) \, du \, dv \right],$$

and for L_{P1}

$$\Sigma_{n,12} = \frac{1}{|W_n|} \left[\int_{W_n^3} \frac{(\lambda'_{\beta_0}(w))^T}{\lambda_{\beta_0}(w)} g'_{\eta_0}(u-v) \right.$$
$$\times I(\|u-v\| < R) \left(\frac{\lambda_{\beta_0}^{(3)}(w,u,v)}{g_{\eta_0}(u-v)} - \lambda_{\beta_0}(u)\lambda_{\beta_0}^{(2)}(w,v) \right) dw \, du \, dv$$
$$\left. + \int_{W_n^2} (\lambda'_{\beta_0}(u))^T g'_{\eta_0}(u-v) I(\|u-v\| < R) \lambda_{\beta_0}(v) \, du \, dv \right].$$

For L_{P3} we get

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

and for L_{P1}

$$\Sigma_{n,22,LP1} = \Sigma_{n,22,LP3}$$

$$+ \frac{1}{|W_n|} \Bigg[-3 \int_{W_n^3} (g'_{\eta_0}(u-v))^T g'_{\eta_0}(v-w) I(\|u-v\|, \|v-w\| < R)$$

$$\times \lambda_{\beta_0}(u) \lambda_{\beta_0}(v) \lambda_{\beta_0}(w) \, \mathrm{d}u \, \mathrm{d}v \, \mathrm{d}w$$

$$+ \int_{W_n^4} (g'_{\eta_0}(u-v))^T g'_{\eta_0}(w-z) I(\|u-v\|, \|w-z\| < R)$$

$$\times \Big(\lambda_{\beta_0}^{(2)}(v,z) \lambda_{\beta_0}(u) \lambda_{\beta_0}(w) - 2 \frac{\lambda^{(3)}(v,u,z)}{g_{\eta_0}(u-v)} \lambda_{\beta_0}(w) \Big) \, \mathrm{d}u \, \mathrm{d}v \, \mathrm{d}w \, \mathrm{d}z \Bigg].$$

8 Simulation study

8.1 Design of the simulation study

To compare the performance of the developed two-step estimation methods we applied them 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

$$f(x) = \exp(\beta_1 x_1 - \max(\beta_1, 0)), \quad x = (x_1, x_2) \in W. \tag{8.1}$$

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

$$\lambda_\beta(x) = \exp(\beta_0 + \beta_1 x_1), \quad x \in W, \tag{8.2}$$

where $\beta_0 = \log \mu - \log \theta - \max(\beta_1, 0)$ and the interaction parameter is $\eta = (\mu, \sigma)$. The total intensity of X on W is thus

$$\mathsf{E} \, |X \cap W| = \int_W \lambda_\beta(x) \, \mathrm{d}x = \frac{\mu}{\theta \cdot |\beta_1|} \left(1 - \exp(-|\beta_1|) \right). \tag{8.3}$$

In the first estimation step, we used the Poisson log-likelihood score function (5.2) to estimate the parameter $\beta = (\beta_0, \beta_1)$. The estimation was performed by means of the function `ppm` from the R package `Spatstat` ([2]). The vector of interaction parameters $\eta = (\mu, \sigma)$ was estimated in the second step by the methods described in Sections 5 and 6.

The minimum contrast estimation, using the K-function (MCK) and the pair-correlation function (MCg), was performed by a `Spatstat` routine. The value of

the tuning parameter r (see equation (4.1)) 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 of 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 [7] and [24].

The composite likelihood (CL) and Palm likelihood estimates (PL) 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 parameter R, the composite and Palm likelihood estimates were computed using three different values of $R = 0.1, 0.2$ and 0.3.

Finally, the remaining parameter θ was identified from the equation (8.3) 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 a mean number of points in $X \cap W$ ranging from approx. 310 to 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. Fig. 2 shows realizations of the point processes for the combination of parameters considered.

8.2   Results of the simulation study

Tables 1–3 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 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 $R = 0.01$) and PL3 (with $R = 0.01$) is at most $2\,\%$ for all the considered models, thus all these four estimators produce very good estimates, see Table 1. Both minimum contrast methods have very similar
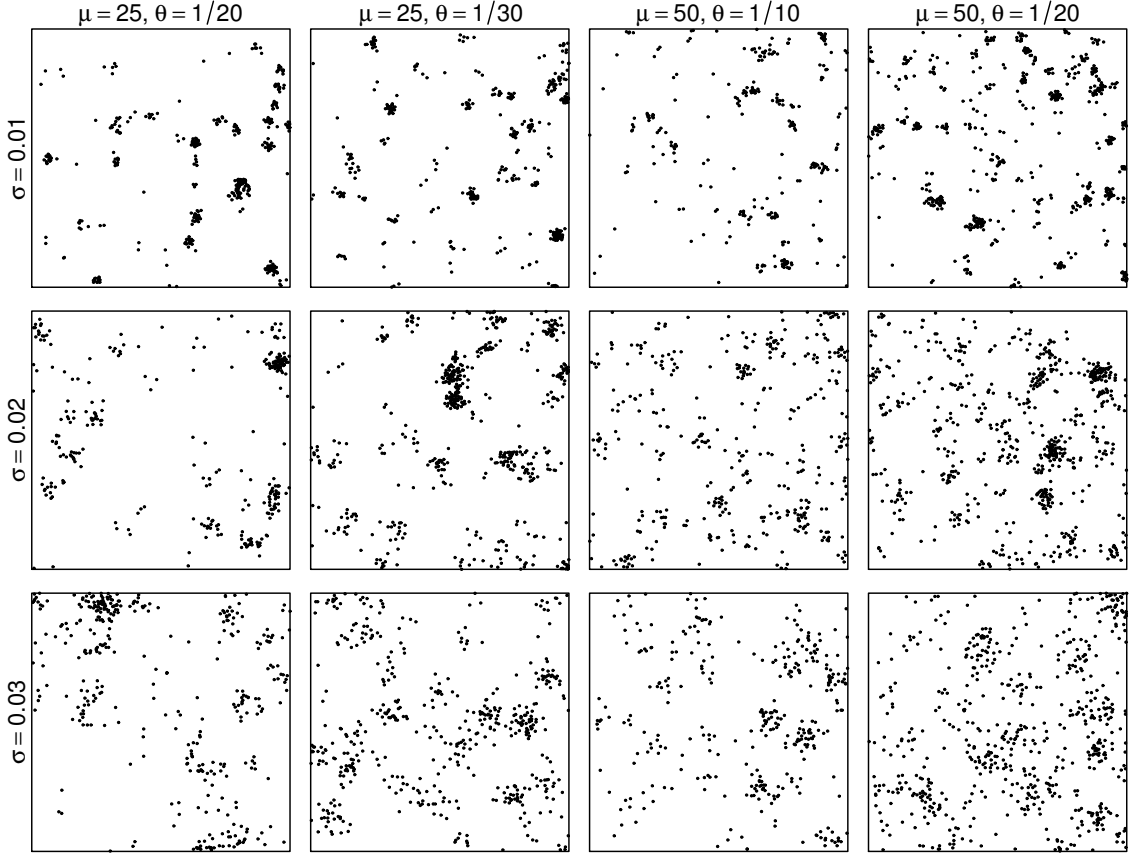
Figure 2: Realizations of the point processes used in the simulation study. For details, see Section 8.1.

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 ($\sigma = 0.02, 0.03$) 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)

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. 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.

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	.004	.006	.007	.020	.013	.009	.009	.009	.011	.011	.011
			.003	−.047	.005	.013	.009	−.019	−.019	−.019	−.001	−.001	−.001
25	1/20	0.02	.007	.009	.006	.020	.041	.015	.016	.016	.015	.023	.023
			−.010	−.045	.002	.015	.024	−.072	−.072	−.072	−.016	−.009	−.009
25	1/20	0.03	.017	.017	.020	.022	.041	.029	.034	.034	.019	.057	.124
			−.021	−.048	.021	.016	.031	−.124	−.147	−.147	−.027	−.006	.013
25	1/30	0.01	.003	.005	.009	.038	.043	.018	.019	.019	.023	.025	.025
			−.004	−.048	.006	.024	.026	−.011	−.011	−.011	.008	.009	.009
25	1/30	0.02	.006	.008	.005	.020	.036	.016	.017	.017	.016	.038	.056
			−.009	−.039	.001	.020	.030	−.069	−.069	−.069	−.012	.001	.003
25	1/30	0.03	.011	.013	.013	.016	.034	.027	.033	.033	.014	.040	.108
			−.034	−.057	.012	.005	.013	−.131	−.154	−.154	−.035	−.023	−.008
50	1/10	0.01	.007	.007	.010	.013	.012	.008	.008	.008	.010	.010	.010
			.003	−.045	.012	.014	.012	−.020	−.020	−.020	.003	.003	.003
50	1/10	0.02	.012	.013	.010	.021	.054	.020	.022	.022	.018	.051	.069
			−.017	−.052	−.001	.006	.023	−.098	−.098	−.098	−.019	−.003	−.001
50	1/10	0.03	.020	.023	.040	.023	.049	.046	.052	.052	.021	.055	NA
			−.038	−.066	.044	.006	.017	−.018	−.020	−.020	−.041	−.022	NA
50	1/20	0.01	.003	.005	.005	.011	.011	.008	.008	.008	.009	.009	.009
			−.002	−.046	.002	.004	.004	−.026	−.026	−.026	−.006	−.006	−.006
50	1/20	0.02	.006	.008	.005	.016	.028	.015	.016	.016	.012	.043	.074
			−.007	−.038	.006	.013	.018	−.090	−.090	−.090	−.010	.007	.011
50	1/20	0.03	.012	.013	.014	.020	.041	.038	.045	.045	.016	.037	.057
			−.021	−.045	.018	.018	.038	−.166	−.190	−.190	−.021	−.008	.001

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 equation (8.3), 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

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. 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.

μ	θ	σ	MCK	MCg	CL 0.1	0.2	0.3	PL1 0.1	0.2	0.3	PL3 0.1	0.2	0.3
25	1/20	0.01	.098	.111	.249	.139	.126	.154	.154	.154	.125	.125	.125
			.156	.177	.089	.093	.104	.229	.229	.229	.183	.183	.183
25	1/20	0.02	.159	.173	.263	.221	.201	.325	.325	.325	.197	.198	.198
			.183	.200	.091	.102	.109	.363	.363	.363	.221	.218	.218
25	1/20	0.03	.277	.300	.446	.261	.272	.710	.766	.766	.299	.326	.330
			.270	.289	.096	.114	.120	.596	.632	.632	.292	.290	.287
25	1/30	0.01	.097	.102	.263	.155	.133	.145	.145	.145	.122	.122	.122
			.141	.148	.089	.086	.087	.207	.207	.207	.162	.162	.162
25	1/30	0.02	.136	.146	.233	.195	.178	.344	.345	.345	.208	.210	.211
			.183	.194	.101	.091	.096	.376	.376	.376	.231	.227	.227
25	1/30	0.03	.223	.230	.314	.293	.307	.679	.733	.733	.278	.293	.295
			.251	.260	.072	.118	.150	.585	.620	.620	.284	.284	.281
50	1/10	0.01	.068	.082	.111	.086	.086	.101	.101	.101	.081	.081	.081
			.113	.144	.061	.068	.079	.175	.175	.175	.125	.125	.125
50	1/10	0.02	.122	.137	.180	.187	.187	.314	.315	.315	.166	.169	.169
			.148	.172	.063	.091	.093	.350	.351	.351	.169	.164	.164
50	1/10	0.03	.255	.272	.440	.276	.317	1.07	1.12	1.12	.342	.362	NA
			.247	.268	.040	.123	.154	.742	.781	.781	.288	.287	NA
50	1/20	0.01	.065	.070	.120	.088	.086	.104	.104	.104	.083	.083	.083
			.100	.110	.055	.063	.069	.169	.169	.169	.122	.122	.122
50	1/20	0.02	.088	.095	.125	.137	.132	.243	.243	.243	.119	.122	.123
			.135	.147	.064	.087	.095	.331	.332	.332	.153	.148	.148
50	1/20	0.03	.173	.179	.240	.220	.259	.810	.864	.864	.238	.255	.257
			.191	.202	.077	.100	.101	.651	.692	.692	.220	.220	.219

point patterns with smaller number of points and loose clusters. For all the estimators, the precision of the estimates 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 ((μ, θ, σ) = $(50, 1/10, 0.01), (50, 1/20, 0.01), (50, 1/20, 0.02)$), in all the other models CL with an appropriate value of R produces the best estimates of θ. In most cases, PL3 shows similar behaviour 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

Eventhough 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 (6.1) 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 (6.1) since all the difference processes Y_x have the same weight.

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. 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 last column shows the relative mean squared errors (upper row) and relative mean biases (lower row) of the estimated inhomogeneity parameter $\hat{\beta}_1$.

μ	θ	σ	MCK	MCg	CL 0.1	0.2	0.3	PL1 0.1	0.2	0.3	PL3 0.1	0.2	0.3	$\hat{\beta}_1$
25	1/20	0.01	.677	.730	.792	.648	.637	.799	.799	.799	.719	.719	.719	.498
			.387	.414	.293	.305	.321	.460	.460	.460	.409	.409	.409	−.065
25	1/20	0.02	.690	.728	.790	.705	.668	1.01	1.01	1.01	.739	.737	.737	.492
			.363	.384	.249	.269	.277	.552	.553	.553	.400	.397	.397	−.008
25	1/20	0.03	1.03	1.07	1.15	.778	.800	1.93	2.05	2.05	1.08	1.14	1.14	.507
			.484	.506	.274	.298	.310	.843	.884	.884	.508	.509	.505	−.007
25	1/30	0.01	.621	.642	.800	.623	.569	.670	.670	.670	.611	.611	.611	.541
			.314	.323	.254	.250	.249	.367	.367	.367	.329	.329	.329	−.007
25	1/30	0.02	.635	.661	.674	.644	.627	.929	.931	.931	.692	.690	.690	.441
			.318	.331	.227	.217	.223	.510	.510	.510	.360	.354	.354	.044
25	1/30	0.03	.872	.889	.897	.817	.882	1.883	1.993	1.993	1.074	1.096	1.096	.511
			.436	.446	.237	.285	.325	.806	.847	.847	.480	.486	.486	.002
50	1/10	0.01	.229	.262	.280	.230	.235	.277	.277	.277	.241	.241	.241	.272
			.190	.224	.139	.143	.154	.252	.252	.252	.201	.201	.201	.028
50	1/10	0.02	.320	.352	.302	.363	.375	.557	.559	.559	.357	.361	.361	.257
			.210	.235	.109	.150	.153	.412	.413	.413	.229	.224	.224	.026
50	1/10	0.03	.631	.658	.776	.630	.660	1.82	1.89	1.89	.808	.828	NA	.266
			.328	.349	.096	.195	.227	.839	.879	.879	.376	.375	NA	.030
50	1/20	0.01	.198	.209	.231	.198	.201	.254	.254	.254	.220	.220	.220	.263
			.180	.191	.129	.137	.145	.249	.249	.249	.200	.200	.200	−.024
50	1/20	0.02	.291	.304	.304	.318	.323	.508	.509	.509	.316	.318	.318	.245
			.208	.221	.134	.156	.165	.406	.407	.407	.222	.216	.216	−.017
50	1/20	0.03	.381	.386	.366	.380	.455	1.17	1.24	1.24	.457	.479	.480	.245
			.244	.252	.121	.149	.156	.711	.753	.753	.275	.275	.273	.025

Consequently, since Y_x from the high intensity areas has a smaller weight than the correct one, we obtain an extra positive bias for $\widehat{\mu}$ ("mean number of clusters") to compensate the discrepancy between (6.1) and the data. Formula (6.3) for L_{P3} includes the approximate "probabilities" $\check{\lambda}(x)$ 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 Section 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 estimates. The best performance of the PL1 and PL3 estimates 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 extend the problem with outlier estimates and numerical instability also applies to the CL estimates with larger R (in particular $R = 0.3$).

Concerning the overall numerical complexity of the compared estimation methods, the fastest are the MCK and MCg estimates 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 $X \cap W$ needs to be incorporated.

We have also studied the correlation between the estimators. In all cases we get negative correlation between $\widehat{\sigma}$ and $\widehat{\mu}$. The absolute value of the correlation ranges between $20\,\%$ to $30\,\%$ for the tight clusters case with $\sigma = 0.01$, around $50\,\%$ for $\sigma = 0.02$ and grows up to $60\,\%$ to $70\,\%$ 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 $\widehat{\sigma}$ of the clusters, the smaller the estimated number $\widehat{\mu}$ of the clusters. The smallest correlation (in absolute value) is always obtained by the MCK and MCg estimators, the CL and PL estimates usually have $10\,\%$ larger correlation.

Since $\widehat{\theta}$ is derived from $\widehat{\mu}$, the correlation between $\widehat{\sigma}$ and $\widehat{\theta}$ follows the same pattern as the correlation between $\widehat{\sigma}$ and $\widehat{\mu}$. 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 of $X \cap W$ (a quantity not used for estimation of the other two parameters).

References

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

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

[3] Cox, D. R. (1955) Some statistical models related with series of events. *J. Roy. Stat. Soc. Ser. B*,17:129–164.

[4] Daley, D. J. and Vere-Jones, D. (1988) *An Introduction to the Theory of Point Processes*. Springer Verlag, New York.

[5] Diggle, P. J. (1983) *Statistical Analysis of Spatial Point Patterns*. Academic Press, London.

[6] Diggle, P. J. (2003) *Statistical Analysis of Spatial Point Patterns*, 2nd edition. Oxford University Press, New York.

[7] Dvořák, J. and Prokešová, M. (2012) Moment estimation methods for stationary spatial Cox processes – a comparison. *Kybernetika*, 48:1007–1026.

[8] Guan, Y. (2006) A composite likelihood approach in fitting spatial point process models. *J. Am. Stat. Assoc.*, 101:1502–1512.

[9] Guan, Y. and Sherman, M. (2007) On least squares fitting for stationary spatial point processes. *J. Roy. Statist. Soc. Ser. B*, 69:31–49.

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

[11] Guyon, X. (1991) *Random Fields on a Network.* New York:Springer.

[12] Hahn, U., Jensen, E.B.V., van Lieshout, M.N.M., Nielsen, L.S. (2003) Inhomogeneous spatial point processes by location-dependent scaling. *Adv. Appl. Prob.*, 35:319–336.

[13] 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.

[14] 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.

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

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

[17] Jalilian, A., Guan, Y. and Waagepetersen, R. P. (2013). Decomposition of variance for spatial Cox processes. *Scand. J. Statist.*, 40:119–137.

[18] Jensen, E. B. V. and Nielsen, L.S. (2000) Inhomogeneous Markov point processes by transformation. *Bernoulli*, 6:761–782.

[19] Lindsay, B.G. (1998) Composite likelihood methods. *Contemp. Math.* 80:221–239.

[20] Matérn, B. (1971) Doubly stochastic Poisson processes in the plane. In *Statistical ecology.* Volume 1. Pennsylvania State University Press.

[21] Møller, J. (2003) Shot noise Cox processes. *Adv. Appl. Prob.*, 35:614–640.

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

[23] Møller, J. and Waagepetersen, R. P. (2007) Modern statistics for spatial point processes, *Scand. J. Statist.*, 34:643–684.

[24] Prokešová, M. and Dvořák, J. (2014) Statistics for inhomogeneous space-time shot-noise Cox processes, *Methodol. Comput. Appl. Probab.*, 16, to appear.

[25] Prokešová, M. and Jensen, E. B. V. (2013) Asymptotic Palm likelihood theory for stationary point processes, *Ann. Inst. Statist. Math.*, 65:387–412.

[26] Stoyan, D., Kendall, W. S. and Mecke, J. (1995) *Stochastic Geometry and its Applications.* Second edition. Wiley, Chichester.

[27] Tanaka, U., Ogata, Y., Stoyan, D. (2007) Parameter estimation and model selection for Neymann-Scott point processes. *Biometrical Journal* 49:1–15.

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

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

[30] Waagepetersen, R. P. and Guan, Y. (2009) Two-step estimation for inhomogeneous spatial point processes, *J. R. Stat. Soc. B*, 71:685–702.