



Adaptive estimating function inference for non-stationary determinantal point processes

Frédéric Lavancier, Arnaud Poinas, Rasmus Waagepetersen

► To cite this version:

Frédéric Lavancier, Arnaud Poinas, Rasmus Waagepetersen. Adaptive estimating function inference for non-stationary determinantal point processes. *Scandinavian Journal of Statistics*, 2021, 48 (1), pp.87-107. 10.1111/sjos.12440 . hal-01816528

HAL Id: hal-01816528

<https://hal.science/hal-01816528>

Submitted on 15 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Adaptive estimating function inference for non-stationary determinantal point processes

FRÉDÉRIC LAVANCIER^{1,*} ARNAUD POINAS^{2,**} RASMUS WAAGEPETERSEN^{3,†}

¹*Laboratoire de Mathématiques Jean Leray – BP 92208 – 2, Rue de la Houssinière – F-44322 Nantes Cedex 03 – France. Inria, Centre Rennes Bretagne Atlantique, France.*

E-mail: *frederic.lavancier@univ-nantes.fr

²*IRMAR – Campus de Beaulieu - Bat. 22/23 – 263 avenue du Général Leclerc – 35042 Rennes – France*

E-mail: **arnaud.poinas@univ-rennes1.fr

³*Department of Mathematical Sciences – Aalborg University - Fredrik Bajersvej 7G – DK-9220 Aalborg – Denmark*

E-mail: †rw@math.aau.dk

Estimating function inference is indispensable for many common point process models where the joint intensities are tractable while the likelihood function is not. In this paper we establish asymptotic normality of estimating function estimators in a very general setting of non-stationary point processes. We then adapt this result to the case of non-stationary determinantal point processes which are an important class of models for repulsive point patterns. In practice often first and second order estimating functions are used. For the latter it is common practice to omit contributions for pairs of points separated by a distance larger than some truncation distance which is usually specified in an ad hoc manner. We suggest instead a data-driven approach where the truncation distance is adapted automatically to the point process being fitted and where the approach integrates seamlessly with our asymptotic framework. The good performance of the adaptive approach is illustrated via simulation studies for non-stationary determinantal point processes.

Keywords: asymptotic normality, determinantal point processes, estimating functions, joint intensities, non-stationary, repulsive.

1. Introduction

A common feature of spatial point process models (except for the Poisson process case) is that the likelihood function is not available in a simple form. Numerical approximations of the likelihood function are available [see e.g. 12, 13, for reviews] but the approaches are often computationally demanding and the distributional properties of the approximate maximum likelihood estimates may be difficult to assess. Therefore much work has focused on establishing computationally simple estimation methods that do not require knowledge of the likelihood function.

In this paper we focus on estimation methods for point processes which have known joint intensity functions. This includes many cases of Cox and cluster point process models [12, 7, 1] as well as determinantal point processes [11, 17, 16, 8]. These classes of models are quite different since realizations of Cox and cluster point processes are aggregated while determinantal point processes produce regular point pattern realizations.

Knowledge of an n th order joint intensity enables the use of the so-called Campbell formulae for computing expectations of statistics given by random sums indexed by n -tuples of distinct points in a point process. Unbiased estimating functions can then be constructed from such statistics by subtracting their expectations. So far mainly the cases of first and second order joint intensities have been considered where the first order joint intensity is simply the intensity function. However, consideration of higher order estimating functions may be worthwhile to obtain more precise estimators or to identify parameters in complex point process models.

Theoretical results have been established in a variety of special cases of first and second order estimating functions for Cox and cluster processes [15, 5, 22, 6, 23] and for the closely related Palm likelihood estimators [20, 18, 19]. The common general structure of the estimating functions on the other hand calls for a general theoretical set-up which is the first contribution of this paper. Our set-up also covers third or higher order estimating functions and combinations of such estimating functions.

The literature on statistical inference for determinantal point processes is quite limited with theoretical results so far only available in case of minimum contrast estimation for stationary determinantal point processes [3]. Based on the general set-up our second main contribution is to provide a detailed theoretical study of estimating function estimators for general non-stationary determinantal point processes.

Specializing to second-order estimating functions, a common approach [5, 20] is to restrict the random sum to pairs of R -close points for some user-specified $R > 0$. This may lead to faster computation and improved statistical efficiency. The properties of the resulting estimators depend strongly on R but only ad hoc guidance is available for the choice of R . Moreover, it is difficult to account for ad hoc choices of R when establishing theoretical results. Our third contribution is a simple intuitively appealing adaptive choice of R which leads to a theoretically tractable estimation procedure and we demonstrate its usefulness in simulation studies for determinantal point processes as well as an example of a cluster process.

2. Estimating functions based on joint intensities

A point process X on $\mathbb{R}^d$, $d \geq 1$, is a locally finite random subset of $\mathbb{R}^d$. For $B \subseteq \mathbb{R}^d$, we let $N(B)$ denote the random number of points in $X \cap B$. That X is locally finite means that $N(B)$ is finite almost surely whenever B is bounded. The so-called joint intensities of a point process are described in Section 2.1. In this paper we mainly focus on determinantal point processes, detailed in Section 3. A prominent feature of determinantal point processes is that they have known joint intensity functions of any order.

2.1. Joint intensity functions and Campbell formulae

For integer $n \geq 1$, the joint intensity $\rho^{(n)}$ of n th order is defined by

$$\mathbb{E} \sum_{u_1, \dots, u_n \in X}^{\neq} \mathbb{1}_{u_1 \in B_1, \dots, u_n \in B_n} = \int_{\times_{i=1}^n B_i} \rho^{(n)}(u_1, \dots, u_n) du_1 \cdots du_n \quad (1)$$

for Borel sets $B_i \subseteq \mathbb{R}^d$, $i = 1, \dots, n$, assuming that the left hand side is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^d$. The $\neq$ over the summation sign means that the sum is over pairwise distinct points in X . Of special interest are the cases $n = 1$ and $n = 2$ where the intensity function $\rho = \rho^{(1)}$ and the second order joint intensity $\rho^{(2)}$ determine the first and second order moments of the count variables $N(B)$, $B \subseteq \mathbb{R}^d$. The pair correlation function $g(u, v)$ is defined as

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

whenever $\rho(u)\rho(v) > 0$ (otherwise we define $g(u, v) = 0$). The product $\rho(u)g(u, v)$ can be interpreted as the intensity of X at u given that $v \in X$. Hence $g(u, v) > 1$ (< 1) means that presence of a point at v increases (decreases) the likeliness of observing yet another point at u . The Campbell formula

$$\mathbb{E} \sum_{u_1, \dots, u_n \in X}^{\neq} f(u_1, \dots, u_n) = \int f(u_1, \dots, u_n) \rho^{(n)}(u_1, \dots, u_n) du_1 \cdots du_n$$

follows immediately from the definition of $\rho^{(n)}$ for any non-negative function $f : (\mathbb{R}^d)^n \rightarrow [0, \infty[$.

2.2. A general asymptotic result for estimating functions

Consider a parametric family of distributions $\{\mathbb{P}_\theta : \theta \in \Theta\}$ of point processes on $\mathbb{R}^d$, where Θ is a subset of $\mathbb{R}^p$. We assume a realization of the point process X with distribution $\mathbb{P}_{\theta^*}$, $\theta^* \in \text{Int}(\Theta)$, is observed on a window $W_n \subset \mathbb{R}^d$. We estimate the unknown parameter θ^* by the solution $\hat{\theta}_n$ of $e_n(\theta) = 0$ where

$$e_n(\theta) = \begin{pmatrix} \sum_{u_1, \dots, u_{q_1} \in X \cap W_n}^{\neq} f_1(u_1, \dots, u_{q_1}; \theta) - \int_{W_n^{q_1}} f_1(u; \theta) \rho^{(q_1)}(u; \theta) du \\ \vdots \\ \sum_{u_1, \dots, u_{q_l} \in X \cap W_n}^{\neq} f_l(u_1, \dots, u_{q_l}; \theta) - \int_{W_n^{q_l}} f_l(u; \theta) \rho^{(q_l)}(u; \theta) du \end{pmatrix}$$

for l given functions $f_i : (\mathbb{R}^d)^{q_i} \times \Theta \rightarrow \mathbb{R}^{k_i}$ such that $\sum_i k_i = p$.

A basic assumption for the following theorem (verified in Appendix A) is that a central limit theorem is available for $e_n(\theta^*)$ (assumption (X3)). In addition to this, a number of

technical assumptions (F1) through (F3) (or (F3')), (X1) and (X2) regarding existence and differentiability of joint intensities as well as differentiability of the f_i are needed. All the conditions are listed in Appendix A.

Theorem 2.1. *Under Assumptions (F1) through (F3) (or (F3')), (X1) and (X2), with a probability tending to one as $n \rightarrow \infty$, there exists a sequence of roots $\hat{\theta}_n$ of the estimating equations $e_n(\theta) = 0$ for which*

$$\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta^*.$$

Moreover, if (X3) holds true, then

$$|W_n| \Sigma_n^{-1/2} H_n(\theta^*)(\hat{\theta}_n - \theta^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I_p).$$

where $\Sigma_n = \text{Var}(e_n(\theta^*))$, $H_n(\theta^*)$ is defined in (F3), and I_p is the $p \times p$ identity matrix.

2.3. Second order estimating functions

Referring to the previous section, much attention has been devoted to instances of the case $l = 1$, $q_1 = 2$ and $k_1 = p$. In this case we obtain a second-order estimating function of the form

$$e_n(\theta) = \sum_{u,v \in X \cap W_n}^{\neq} f(u, v; \theta) - \int_{W_n^2} f(u, v; \theta) \rho^{(2)}(u, v; \theta) du dv. \quad (2)$$

[5] noted that for computational and statistical efficiency it may be advantageous to use only close pairs of points rather than all pairs of points. Thus in (2) it is common practice to introduce an indicator $\mathbb{1}_{\|u-v\| \leq R}$ for some constant $0 < R$ or choose f so that $f(u, v) = 0$ whenever $\|u - v\| > R$. We discuss a method for choosing R in Section 2.4.

The general form (2) includes e.g. the score functions of second-order composite likelihood [5, 22] and Palm likelihood functions [20, 18, 19] as well as score functions of minimum contrast object functions based on non-parametric estimates of summary statistics as the K or the pair correlation function. For the second-order composite likelihood of [5],

$$f(u, v; \theta) = \frac{\nabla_{\theta} \rho^{(2)}(u, v; \theta)}{\rho^{(2)}(u, v; \theta)} - \frac{\int_{W^2} \nabla_{\theta} \rho^{(2)}(u, v; \theta) du dv}{\int_{W^2} \rho^{(2)}(u, v; \theta) du dv}$$

while

$$f(u, v; \theta) = \frac{\nabla_{\theta} \rho^{(2)}(u, v; \theta)}{\rho^{(2)}(u, v; \theta)}$$

for the second-order composite likelihood proposed in [22]. The score of the Palm likelihood as generalized to the inhomogeneous case in [19] is obtained with

$$f(u, v; \theta) = \frac{\nabla_{\theta} \frac{\rho^{(2)}(u, v; \theta)}{\rho(u; \theta)}}{\rho^{(2)}(u, v; \theta) / \rho(u; \theta)} - \frac{1}{N(W) - 1} \int_W \nabla_{\theta} \left(\frac{\rho^{(2)}(u, w; \theta)}{\rho(u; \theta)} \right) dw.$$

[19] also regarded the second-order composite likelihood proposed in [22] as a generalization of the stationary case Palm likelihood but the interpretation as a second-order composite likelihood given in [22] is more straightforward.

Considering a class of estimating functions of the form (2) a natural question is what is the optimal choice of f ? [4] provides a solution to this problem where an approximation of the optimal f is obtained by solving numerically a certain integral equation. This yields a statistically optimal estimation procedure but is computationally demanding and requires specification of third and fourth order joint intensities. When computational speed and ease of use is an issue, there is still scope for simpler methods. Moreover, given several (simple) estimation methods, it is possible to combine them adaptively in order to build a final estimator that achieves better properties than each initial estimator, see [9, 10].

2.4. Adaptive version

Consider second-order composite likelihood using only R close pairs. The weight function f is then of the form

$$f_R(u, v; \theta) = \mathbb{1}_{\|u-v\| \leq R} \frac{\nabla_{\theta} \rho^{(2)}(u, v; \theta)}{\rho^{(2)}(u, v; \theta)}. \quad (3)$$

The performance of the parameter estimates can depend strongly on the chosen R . Simulation studies such as in [19] and [4] usually compare results for several values of R corresponding to different multiples of some parameter associated with ‘range of correlation’. For a cluster process this parameter could e.g. be the standard deviation of the distribution for dispersal of offspring around parents. For a determinantal point process the parameter would typically be a correlation scale parameter in the kernel of the determinantal point process, see Section 3. In practice these parameters are not known and among the quantities that need to be estimated. [5] suggested to choose an R that minimizes a goodness of fit criterion for the fitted point process model while [23] suggested to choose R by inspection of a non-parametric estimate of the pair correlation function. Both approaches imply extra work and ad hoc decisions by the user and it becomes very complex to determine the statistical properties of the resulting parameter estimates.

A typical behaviour of many pair correlation functions is that $g(u, v; \theta)$ converges to a limiting value of 1 when $\|u - v\|$ increases and $|g(u, v; \theta) - 1| \leq |g(u, u; \theta) - 1|$ where the upper bound does not depend on u . If $g(u, v; \theta) = 1$ for $\|u - v\| > r_0$ then counts of points are uncorrelated when they are observed in regions separated by a distance of r_0 . Following the idea that R should depend on some range property of the point process we therefore suggest to replace the constraint $\|u - v\| < R$ in (3) by the constraint

$$\frac{|g(u, v; \theta) - 1|}{|g(u, u; \theta) - 1|} > \varepsilon,$$

for a small ε . If e.g. $\varepsilon = 1\%$ this means that we only consider pairs of points (u, v) so that the difference between $g(u, v; \theta)$ and the limiting value 1 is within 1% of the maximal

value $|g(u, u; \theta) - 1|$. Note that this choice of pairs of points is adaptive in that it depends on θ .

We then modify the function f_R to be

$$f_{\text{adap}}(u, v; \theta) = w \left(\varepsilon \frac{g(u, u; \theta) - 1}{g(u, v; \theta) - 1} \right) \frac{\nabla_{\theta} \rho^{(2)}(u, v; \theta)}{\rho^{(2)}(u, v; \theta)} \quad (4)$$

where w is some weight function of bounded support $[-1, 1]$. Later on, when establishing asymptotic results, we will also assume that w is differentiable. A common example of admissible weight function is $w(r) = e^{1/(r^2-1)}$ for $-1 \leq r \leq 1$, while $w(r) = 0$ otherwise. The user needs to specify a value of ε but in contrast to the original tuning parameter R , ε has an intuitive meaning independent of the underlying point process. We choose $\varepsilon = 1\%$. In the simulation study in Section 4.1 we also consider $\varepsilon = 5\%$ in order to investigate the sensitivity to the choice of ε .

3. Asymptotic results for determinantal point processes

A point process X is a determinantal point process (DPP for short) with kernel $K : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ if for all $n \geq 1$, the joint intensity $\rho^{(n)}$ exists and is of the form

$$\rho^{(n)}(u_1, \dots, u_n) = \det[K](u_1, \dots, u_n)$$

for all $\{u_1, \dots, u_n\} \subset \mathbb{R}^d$, where $[K](u_1, \dots, u_n)$ is the matrix with entries $K(u_i, u_j)$. The intensity function is thus $\rho(u) = K(u, u)$, $u \in \mathbb{R}^d$. If a determinantal point process with kernel K exists it is unique. General conditions for existence are presented in [8]. In particular, if K admits the form

$$K(u, v) = \sqrt{\rho(u)\rho(v)}C(u - v)$$

for a function $C : \mathbb{R}^d \rightarrow \mathbb{R}$ with $C(0) = 1$, then a sufficient condition for existence of a DPP with kernel K is that ρ is bounded and that C is a square integrable continuous covariance function with spectral density bounded by $1/\|\rho\|_{\infty}$. The normalization $C(0) = 1$ ensures that ρ is the intensity of the DPP.

We now consider a parametric family of DPPs on $\mathbb{R}^d$ with kernels K_{θ} where $\theta \in \Theta$ and $\Theta \subseteq \mathbb{R}^p$ [see 8, 2, for examples of such families]. Henceforth, we assume that K_{θ} is symmetric and the DPP with kernel K_{θ} exists for all $\theta \in \Theta$.

[8] provide an expression for the likelihood of a DPP on a bounded window and discuss likelihood based inference for stationary DPPs. However, the expression depends on a spectral representation of K which is rarely known in practice and must be approximated numerically. Letting n denote the number of observed points, the likelihood further requires the computation of an $n \times n$ dense matrix which can be time consuming for large n . As an alternative, [2] consider minimum contrast estimation based on the pair correlation function or Ripley's K -function, but only for stationary DPPs. In the

following, we consider general non-stationary DPPs and the estimator $\hat{\theta}_n$ obtained by solving $e_n(\theta) = 0$ where e_n is given by (2).

We establish in Section 3.1 using Theorem 2.1 the asymptotic properties of the estimate $\hat{\theta}_n$ where e_n is given by (2) for a wide class of test functions f . In Section 3.2, we focus on a particular case of the DPP model, where the parameter $\theta = (\beta, \psi)$ can be separated into a parameter β only appearing in the intensity function and a parameter ψ only appearing in the pair correlation function. Following [23], it is natural to consider a two-step estimation procedure where in a first step β is estimated by a Poisson likelihood score estimating function, and in a second step the remaining parameter ψ is estimated by a second order estimating function as in (2), where β is replaced by $\hat{\beta}_n$ obtained in the first step. The asymptotic properties of this two-step procedure again follow as a special case of Theorem 2.1.

3.1. Second order estimating functions for DPPs

We assume a realization of a DPP X with kernel K_{θ^*} , $\theta^* \in \text{Int}(\Theta)$, is observed on a window $W_n \subset \mathbb{R}^d$. We estimate the unknown parameter θ^* by the solution $\hat{\theta}_n$ of $e_n(\theta) = 0$ where $e_n(\theta)$ is given by (2) for a given $\mathbb{R}^p$ -valued function f . Therefore, we are in a special case of the set-up in Section 2.2 with $l = 1$, $q_1 = 2$, $k_1 = p$ and we assume that $f_1 = f$ satisfies the assumptions (F1) through (F3) (or (F3')) listed in Appendix A. The condition (F1) in this case demands that $\theta \mapsto f(u, v; \theta)$ is twice continuously differentiable in a neighbourhood of θ^* and for θ in this neighbourhood, the derivatives are bounded with respect to (u, v) uniformly in θ . Moreover, from (F2), there exists $R > 0$ such that for all θ in a neighbourhood of θ^* ,

$$f(u, v; \theta) = 0 \quad \text{if} \quad \|u - v\| > R. \quad (5)$$

Concerning (F3) (or (F3')), this condition controls the asymptotic behaviour of the matrix $H_n(\theta)$ given by

$$H_n(\theta) = \frac{1}{|W_n|} \int_{W_n^2} f(u, v; \theta) \nabla_{\theta} \rho^{(2)}(u, v; \theta)^T du dv,$$

where we recall that in this setting

$$\rho^{(2)}(u, v; \theta) = K_{\theta}(u, u)K_{\theta}(v, v) - K_{\theta}(u, v)^2. \quad (6)$$

The assumptions (F3) and (F3') are technical and needed for the identifiability of the estimation procedure. When H_n is a symmetric matrix, assumption (F3) seems simpler to verify than (F3'). As an important example, when f is defined as in (4), we prove in Lemmas 3.2 and 3.3 that (F3) is generally satisfied even if X is not stationary.

Finally, as shown in the proof of Theorem 3.1 below, the assumptions (X1) through (X3) in Theorem 2.1 become:

- (D1) $\theta \mapsto K_\theta(u, v)$ is twice continuously differentiable in a neighborhood of θ^* , for all $u, v \in \mathbb{R}^d$. Moreover, the first and second derivative of K_θ with respect to θ are bounded with respect to $u, v \in \mathbb{R}^d$ uniformly in θ in a neighborhood of θ^* .
- (D2) The kernel K_{θ^*} satisfies, for some $\varepsilon > 0$,

$$\sup_{\|u-v\|>r} K_{\theta^*}(u, v) = o(r^{-(d+\varepsilon)/2}).$$

- (D3) $\liminf_n \lambda_{\min}(|W_n|^{-1}\Sigma_n) > 0$ where $\Sigma_n := \text{Var}(e_n(\theta^*))$.

- (W) $\exists \varepsilon > 0$ s.t. $|\partial W_n \oplus (R + \varepsilon)| = o(|W_n|)$, where ∂ in this context denotes the boundary of a set, R is defined in (5), and $|W_n| \rightarrow \infty$, as $n \rightarrow \infty$.

Let us briefly comment on these assumptions. (D1) is a standard regularity assumption. Condition (D2) is not restrictive since all standard parametric kernel families satisfy $\sup_{\|u-v\|>r} K_\theta(u, v) = O(r^{-(d+1)/2})$, including the most repulsive stationary DPP [see 8, 2]. Condition (D3) ensures that the asymptotic variance in the central limit theorem below is not degenerated. Finally, Assumption (W) makes specific the fact that W_n is not too irregularly shaped and tends to infinity in all directions. It is for instance fulfilled if W_n is a Cartesian product of d intervals whose lengths tends to infinity.

Theorem 3.1. *Under Assumptions (D1) and (D2), if assumptions (F1) through (F3) (or (F3')) are satisfied for $f_1 = f$, with a probability tending to one as $n \rightarrow \infty$, there exists a sequence of roots $\hat{\theta}_n$ of the estimating equations $e_n(\theta) = 0$ for which*

$$\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta^*.$$

If moreover (W) and (D3) holds true, then

$$|W_n| \Sigma_n^{-1/2} H_n(\theta^*)(\hat{\theta}_n - \theta^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I_p).$$

Proof. We deduce from (6) that (D1) implies (X1). Moreover, it was shown in [14] that (X2) is a consequence of (D2) and that (X3) is a consequence of (D2), (D3) and (W). Thus, we can conclude by applying Theorem (2.1) in the case $l = 1$ and $q_1 = 2$. $\square$

In the case of a stationary X and f given by (4), the following lemma shows that (F3) is satisfied under mild assumptions that are violated only in degenerate cases. For instance, if $p = 1$, the main assumption boils down to $\nabla_\theta \rho^{(2)}(0, t; \theta^*) \neq 0$ for some $t \neq 0$ such that $|K_{\theta^*}(t)| > \sqrt{\varepsilon} K_{\theta^*}(0)$.

Lemma 3.2. *Assume (W) and (D2), suppose X is stationary and let f be as in (4). Let $h : \mathbb{R}^d \rightarrow \mathcal{M}_p(\mathbb{R})$ be defined by*

$$h(t) = w \left(\frac{\varepsilon K_{\theta^*}(0)^2}{K_{\theta^*}(t)^2} \right) \frac{\nabla_\theta \rho^{(2)}(0, t; \theta^*) \nabla_\theta \rho^{(2)}(0, t; \theta^*)^T}{\rho^{(2)}(0, t; \theta^*)}.$$

Assume that w is positive on $[0, 1[$. If h is integrable at the origin and if $\text{span}\{\nabla_\theta \rho^{(2)}(0, t; \theta^) : |K_{\theta^*}(t)| > \sqrt{\varepsilon} K_{\theta^*}(0)\} = \mathbb{R}^p$, then (F3) is satisfied.*

Proof. By definition of w and (D2), there exists $R > 0$ such that $h(t) = 0$ when $\|t\| \geq R$. By Lemma A.1, $H_n(\theta^*)$ converges towards the positive semi-definite matrix $H(\theta^*) = \int_{\|t\| < R} h(t) dt$. In this case, proving (F3) is equivalent to showing that $\phi^T H(\theta^*) \phi = 0$ only if $\phi = 0$. For this, let A be the set of t such that $|K_{\theta^*}(t)| > \sqrt{\varepsilon} K_{\theta^*}(0)$, $\phi \in \mathbb{R}^p$ and note that since $w(\varepsilon K_{\theta^*}(0)^2 / K_{\theta^*}(t)^2) > 0$ for $t \in A$ and $h(t)$ is continuous and positive semi-definite,

$$\begin{aligned} \phi^T H(\theta^*) \phi = 0 &\Leftrightarrow \forall t \in A, \phi^T h(t) \phi = 0 \\ &\Leftrightarrow \forall t \in A, \nabla_{\theta} \rho^{(2)}(0, t; \theta^*)^T \phi = 0 \\ &\Leftrightarrow \phi \in \left(\text{span}\{\nabla_{\theta} \rho^{(2)}(0, t; \theta^*) : t \in A\} \right)^{\perp}. \end{aligned}$$

By assumption $\text{span}\{\nabla_{\theta} \rho^{(2)}(0, t; \theta^*) : t \in A\} = \mathbb{R}^p$ whereby $\phi = 0$, which concludes the proof. $\square$

Similarly, we can show that even in the non-stationary case, condition (F3) is satisfied for the function in (4) but under slightly stronger assumptions on $\nabla_{\theta} \rho^{(2)}(u, v; \theta^*)$. Namely, we demand that all functions $v \mapsto \nabla_{\theta} \rho^{(2)}(u, v; \theta^*)$ are not contained in a single hyperplane of $\mathbb{R}^p$ nor confined around 0. This is similar in essence to what we have assumed in the previous corollary but with the need of a uniform condition with respect to u . Functions that do not satisfy these requirements are arguably degenerate.

Lemma 3.3. Assume (W), (D2) and that K_{θ^*} is bounded. Let f be as in (4) and define $h : \mathbb{R}^d \rightarrow \mathcal{M}_p(\mathbb{R})$ by

$$h(u, v) = w \left(\frac{\varepsilon K_{\theta^*}(u, u) K_{\theta^*}(v, v)}{K_{\theta^*}(u, v)^2} \right) \frac{\nabla_{\theta} \rho^{(2)}(u, v; \theta^*) \nabla_{\theta} \rho^{(2)}(u, v; \theta^*)^T}{\rho^{(2)}(u, v; \theta^*)}.$$

Assume that w is positive on $[0, 1[$. If $\sup_u \|h(u, \cdot)\|$ is integrable and if there exists $\mu > 1$ and $\delta > 0$ such that for all $u \in \mathbb{R}^d$ and for all unit vectors ϕ of $\mathbb{R}^p$ there exists a subset A of $\{v : K_{\theta^*}(u, v)^2 > \mu \varepsilon K_{\theta^*}(u, u) K_{\theta^*}(v, v)\}$ of positive Lebesgue measure $|A| > 0$ and satisfying

$$\forall v \in A, |\phi^T \nabla_{\theta} \rho^{(2)}(u, v; \theta^*)| > \delta$$

then (F3) is satisfied.

Proof. By definition of w , (D2) and the fact that K_{θ^*} is bounded, there exists $R > 0$ such that $h(u, v) = 0$ when $\|v - u\| \geq R$. The integral in (F3) writes

$$H_n(\theta^*) = \int_{W_n^2} h(u, v) \mathbb{1}_{\|u-v\| \leq R} dv du = \int_{W_n^{\ominus R}} \int_{W_n} h(u, v) \mathbb{1}_{\|u-v\| \leq R} dv du + \varepsilon_n$$

where

$$\varepsilon_n = \int_{W_n \setminus W_n^{\ominus R}} \int_{W_n} h(u, v) \mathbb{1}_{\|u-v\| \leq R} dv du.$$

By (W), we have

$$\begin{aligned} \frac{\|\varepsilon_n\|}{|W_n|} &\leq \frac{|W_n \setminus W_n^{\ominus R}|}{|W_n|} \int_{\|v-u\| < R} \sup_{u \in \mathbb{R}^d} \|h(u, v)\| dv \\ &\leq \frac{|\partial W_n \oplus R|}{|W_n|} \int_{\|v-u\| < R} \sup_{u \in \mathbb{R}^d} \|h(u, v)\| dv \rightarrow 0, \end{aligned}$$

and for all ϕ ,

$$\phi^T \left(\int_{W_n^{\ominus R}} \int_{W_n} h(u, v) \mathbb{1}_{\|u-v\| \leq R} du dv \right) \phi = \int_{W_n^{\ominus R}} \left(\int_{\|u-v\| \leq R} \phi^T h(u, v) \phi dv \right) du.$$

By our assumption on $\nabla_\theta \rho^{(2)}$, there exists a set A of positive Lebesgue measure such that

$$\begin{aligned} \forall v \in A, \nabla_\theta \rho^{(2)}(u, v; \theta^*) &\in \text{span}\{\phi\} \cap \mathcal{B}(0, \delta)^C \\ \text{and } w \left(\frac{\varepsilon K_{\theta^*}(u, u) K_{\theta^*}(v, v)}{K_{\theta^*}(u, v)^2} \right) &> \inf_{x \in [0, 1/\mu]} w(x). \end{aligned}$$

Hence for $\|\phi\| = 1$,

$$\begin{aligned} &\frac{1}{|W_n|} \phi^T \left(\int_{W_n^{\ominus R}} \int_{W_n} h(u, v) \mathbb{1}_{\|u-v\| \leq R} du dv \right) \phi \\ &\geq \frac{\inf_{x \in [0, 1/\mu]} w(x)}{|W_n| \|\rho^{(2)}(\cdot, \cdot; \theta^*)\|_\infty} \int_{W_n^{\ominus R}} \left(\int_A |\phi^T \nabla_\theta \rho^{(2)}(u, v; \theta^*)|^2 dv \right) du \\ &\geq \frac{|W_n^{\ominus R}| |A| \delta^2 \inf_{x \in [0, 1/\mu]} w(x)}{|W_n| \|\rho^{(2)}(\cdot, \cdot; \theta^*)\|_\infty} \\ &\rightarrow \frac{|A| \delta^2 \inf_{x \in [0, 1/\mu]} w(x)}{\|\rho^{(2)}(\cdot, \cdot; \theta^*)\|_\infty} > 0 \end{aligned}$$

and since the limit does not depend on ϕ , (F3) is satisfied. $\square$

3.2. Two-step estimation for a separable parameter

We consider a family of kernels

$$K_\theta(u, v) = \sqrt{\rho(u; \beta)} C(u, v; \psi) \sqrt{\rho(v; \beta)},$$

where $\theta := (\beta^T, \psi^T)^T \in \Theta \subset \mathbb{R}^{p+q}$ with $\beta \in \mathbb{R}^p$ and $\psi \in \mathbb{R}^q$, $\rho(\cdot; \beta)$ are non-negative functions, and $C(\cdot, \cdot; \psi)$ are correlation functions, in particular $C(u, u; \psi) = 1$ for any ψ . Note that in this case the DPP with kernel K_θ has intensity $\rho(\cdot; \beta)$ and its pair correlation function is $g(u, v; \psi) = 1 - C^2(u, v; \psi)$.

As in the preceding section, we assume a DPP X with kernel K_{θ^*} , $\theta^* \in \text{Int}(\Theta)$, is observed on a window $W_n \subset \mathbb{R}^d$. In the spirit of [23], we estimate θ^* in two steps. First, β^* is estimated as the solution $\hat{\beta}_n$ of $s_n(\beta) = 0$ where

$$s_n(\beta) = \sum_{u \in X \cap W_n} \frac{\nabla_{\beta} \rho(u; \beta)}{\rho(u; \beta)} - \int_{W_n} \nabla_{\beta} \rho(u; \beta) du$$

is the score function for a Poisson point process. Then, we estimate ψ^* by the solution $\hat{\psi}_n$ of $u_n((\hat{\beta}_n, \psi) = 0$ where

$$u_n(\theta) = \sum_{u, v \in X \cap W_n}^{\neq} f(u, v; \theta) - \int_{W_n^2} f(u, v; \theta) \rho^{(2)}(u, v; \theta) du dv$$

for a given $\mathbb{R}^q$ -valued function f and where $\rho^{(2)}(u, v; \theta) = \rho^{(2)}(u, v; \beta, \psi) = \rho(u; \beta) \rho(v; \beta) (1 - C^2(u, v; \psi))$ in this case. Here and in the following for convenience of notation e.g. identify $u_n(\beta, \psi)$ with $u_n(\theta)$ when $\theta = (\beta^T, \psi^T)^T$.

This two-step procedure is a particular estimating equation procedure, since $\hat{\theta}_n := (\hat{\beta}_n^T, \hat{\psi}_n^T)^T$ is obtained as the solution of $e_n(\theta) = 0$ where $e_n(\theta) = (s_n(\beta)^T, u_n(\beta, \psi)^T)^T$. Thus, this is a particular case of the setting in Section 2.2 where $l = 2$, $q_1 = 1$, $q_2 = 2$, $f_1 = \nabla_{\beta} \rho(u; \beta) / \rho(u; \beta)$ and $f_2 = f$.

We assume in the following theorem the same conditions on the DPP X as in the previous section. Similarly, we assume that (F1) through (F3) (or (F3')) are satisfied for f_1 and f_2 . In this particular case, the matrix H_n involved in (F3) simply writes

$$H_n(\beta, \psi) = \begin{pmatrix} H_n^{1,1}(\beta, \psi) & 0 \\ H_n^{2,1}(\beta, \psi) & H_n^{2,2}(\beta, \psi) \end{pmatrix}$$

where

$$\begin{aligned} H_n^{1,1}(\beta) &= \frac{1}{|W_n|} \int_{W_n} \frac{\nabla_{\beta} \rho(u; \beta) \nabla_{\beta} \rho(u; \beta)^T}{\rho(u; \beta)} du, \\ H_n^{2,1}(\beta, \psi) &= \frac{1}{|W_n|} \int_{W_n^2} f(u, v; \beta, \psi) \nabla_{\beta} \rho^{(2)}(u, v; \beta, \psi)^T du dv, \\ H_n^{2,2}(\beta, \psi) &= \frac{1}{|W_n|} \int_{W_n^2} f(u, v; \beta, \psi) \nabla_{\psi} \rho^{(2)}(u, v; \beta, \psi)^T du dv. \end{aligned}$$

Since it is a non symmetric matrix, condition (F3') is more applicable than (F3). Mild conditions ensuring (F3') in the stationary case are provided in Lemma 3.5.

Theorem 3.4. *Under Assumptions (D1) and (D2), if assumptions (F1) through (F3) (or (F3')) are satisfied for $f_1 = \nabla_{\beta} \rho(u; \beta) / \rho(u; \beta)$ and $f_2 = f$, then with a probability tending to one as $n \rightarrow \infty$, there exists a sequence of solutions $\hat{\theta}_n = (\hat{\beta}_n^T, \hat{\psi}_n^T)^T$ to the estimating equation $(s_n(\beta)^T, u_n(\beta, \psi)^T)^T = 0$ for which*

$$\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta^*.$$

If moreover (W) and (D3) hold true, then

$$|W_n| \Sigma_n^{-1/2} H_n(\theta^*)(\hat{\theta}_n - \theta^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I_{p+q}).$$

Proof. The proof follows the same lines as the proof of Theorem 3.1. $\square$

The next lemma is similar to Lemma 3.2. The main technical condition is not restrictive. When $q = 1$ it boils down to $\nabla_\psi(1 - C^2(0, t; \psi^*)) \neq 0$ for some t such that $C(0, t; \psi^*) \geq \sqrt{\varepsilon} C(0, 0; \psi^*)$.

Lemma 3.5. Assume that for all θ , $K_\theta(u, v)$ only depends on $u - v$, in which case $\rho(u; \beta) = \beta$ with $\beta > 0$ and $C(u, v; \psi) = C(0, v - u; \psi)$ with $\psi \in \mathbb{R}^q$. Then the output of the first step is $\hat{\beta}_n = N(X \cap W_n)/|W_n|$. In the second step, assume

$$\begin{aligned} f(u, v; \beta, \psi) &= w \left(\varepsilon \frac{g(u, u; \psi) - 1}{g(u, v; \psi) - 1} \right) \frac{\nabla_\psi \rho^{(2)}(u, v; \beta, \psi)}{\rho^{(2)}(u, v; \beta, \psi)} \\ &= w \left(\varepsilon \frac{C(0, 0; \psi)^2}{C(0, v - u; \psi)^2} \right) \frac{\nabla_\psi(1 - C^2(0, v - u; \psi))}{1 - C^2(0, v - u; \psi)} \end{aligned}$$

and let

$$h(t) = w \left(\varepsilon \frac{C(0, 0; \psi)^2}{C(0, t; \psi)^2} \right) \frac{\nabla_\psi(1 - C^2(0, t; \psi^*)) \nabla_\psi(1 - C^2(0, t; \psi^*))^T}{1 - C^2(0, t; \psi^*)}.$$

If h is integrable at the origin and if $\text{span}\{\nabla_\psi(1 - C^2(0, t; \psi^*)) : C(0, t; \psi^*) \geq \sqrt{\varepsilon} C(0, 0; \psi^*)\} = \mathbb{R}^q$, then (F3') is satisfied under (W), (D1) and (D2).

Proof. By definition of w and (D2), there exists $R > 0$ such that $h(t) = 0$ when $\|t\| \geq R$. Since $K_\theta(u, v)$ and f are invariant by translation then $H_n(\theta)$ converges by Lemma A.1. In particular, we have

$$\begin{aligned} H_n^{1,1}(\beta) &\rightarrow \frac{1}{\beta}, \\ H_n^{2,2}(\beta, \psi) &\rightarrow \beta^2 \int_{\|t\| \leq R} h(t; \psi) dt, \\ H_n^{2,1}(\beta, \psi) &\rightarrow 2\beta \int_{\|t\| \leq R} w \left(\varepsilon \frac{C(0, 0; \psi)^2}{C(0, t; \psi)^2} \right) \nabla_\psi(1 - C^2(0, t; \psi)) dt. \end{aligned}$$

The limit of $H_n(\theta)$ is continuous by (D1). In this case, proving (F3') is equivalent to showing that the limit of $H_n(\theta^*)$ is invertible. Since this matrix is block triangular and $\beta > 0$ then it is invertible if and only if the limit of $H_n^{2,1}(\theta^*)$ is invertible. This is done the same way as in Lemma 3.2. $\square$

4. Simulation study

In this section we use simulation studies to investigate the performance of our adaptive estimating function and to compare two-step estimation with simultaneous estimation.

4.1. Performance of adaptive estimating function

In order to assess the adaptive test function (4) against the truncated test function (3) with a prescribed R , we consider a DPP model in $\mathbb{R}^2$ with a Bessel-type kernel

$$K(u, v) = \sqrt{\rho(u)\rho(v)} \frac{J_1(2\|u - v\|/\alpha)}{\|u - v\|/\alpha},$$

where J_1 denotes the Bessel function of the first kind, ρ is the intensity and α controls the range of interaction of the DPP. For existence, ρ and α must satisfy

$$\alpha^2 \|\rho\|_\infty \leq \frac{1}{\pi}. \quad (7)$$

This relation shows the tradeoff between the expected number of points and the strength of repulsiveness that we can obtain. This model is a particular instance of the Bessel-type DPP introduced in [2]. It covers a large range of repulsiveness, from the Poisson point process (when α is close to 0) to the most repulsive DPP (when $\alpha = 1/\sqrt{\pi\|\rho\|_\infty}$).

For this model, we consider three constant values of ρ , $\rho \in \{50, 100, 1000\}$, corresponding to homogeneous DPPs, and an inhomogeneous situation where $\rho(u) = \rho(x, y) = 20 \exp(4x)$ when $u \in [0, 1]^2$. The latter case corresponds to a log-linear intensity function involving two parameters. For each ρ , three values of α are considered: a small one, a medium one, and a last one close to the maximal possible value satisfying (7). Examples of point patterns simulated on $[0, 1]^2$ are displayed in Figure 1. All simulations are carried out using R [21], in particular the library `spatstat` [1].

We estimate ρ and α by a two-step procedure as studied in Section 3.2 from realisations of the DPP on $W = [0, 1]^2$. The alternative global approach of Section 3.1 is discussed in the next section. In the first step, the parameters arising in ρ are estimated by the score function for a Poisson point process. This gives $\hat{\rho} = N(X \cap W)/|W|$ in the homogeneous cases. In the second step, we consider the estimating equation based on (3) where θ is α in this setting and when $R \in \{0.05, 0.1, 0.25\}$, and based on the adaptive test function (4) with $\varepsilon = 0.01$ and the weight function w given at the end of Section 2.4. This yields four different estimators of α . The root mean square errors (RMSEs) of these estimators and the mean computation time estimated from 1000 replications are summarised in Table 1. Boxplots are displayed in Figure S1 in the supplementary material. Note that the codes have not been optimised, but the same computational strategy has been used for all methods, making the comparison of the mean computation time meaningful.

The Bessel-type kernel and the aforementioned test functions used in the two-step estimation procedure fulfill the assumptions of Theorem 3.4 and Lemma 3.5 (for the homogeneous case), ensuring nice asymptotic properties of the estimators considered

in this section. This is confirmed by the estimated RMSE's reported in Table 1, that decrease when the intensity ρ increases [which mimics the effect of an increasing window since rescaling the window by a factor $1/k$ is equivalent to change ρ into $k^2\rho$ and α into α/k , see (2.4) in 8]. Moreover, these RMSE's show that the best choice of R in the test function (3) clearly depends on the range of interaction of the underlying process. This emphasizes the importance of a data-driven approach to choosing R since the range is unknown in practice. Fortunately, the performance of the adaptive method is, except for the case $\rho = 100, \alpha = 0.01$, always better than the worst choice of R and very close to the best R . For the exceptional case, the small differences in performance can be explained by Monte Carlo error. Further, use of the adaptive method implies only little or no extra computational effort. In presence of many points, the adaptive version is in fact much faster to compute than the estimator based on (3) with the choice of a too large R , see for instance the results for $\rho = 1000$ and $R = 0.25$.

Table S2 in the supplementary material shows the root mean square errors of the adaptive estimator using $\varepsilon = 0.05$. The RMSEs obtained with $\varepsilon = 0.05$ are bigger than those obtained with $\varepsilon = 0.01$. Nevertheless, the adaptive method with $\varepsilon = 0.05$ still performs well in the sense that it usually performs better than the worst R and usually almost as good as the best R . Because the above estimation methods sometimes fail to converge, we also report in Table S1 in the supplementary material the percentages of times each method has converged in our simulation study. These percentages are similar for all methods. Note that the results in Table 1 and in Figure S1 are based on 1000 simulations where all four methods have converged.

4.2. Two-step versus simultaneous

Most models used in spatial statistics involve a separable parameter $\theta = (\beta, \psi)$ where β only appears in the intensity function and ψ only appears in the pair correlation function. This makes the two-step procedure described in Section 3.2 available, as exploited in the previous simulation study. However a simultaneous second order estimating equation approach might be a better alternative. It is not easy to compare the respective performance of the two approaches through the asymptotic variances obtained in Section 3.1 and Section 3.2. In this section, we show through an example why the two-step procedure seems preferable.

We consider a stationary model with parameter $\theta = (\rho, \psi)$, where ρ is the intensity and the pair correlation function writes $g(u, v; \theta) = g(r; \psi)$ with $r = \|u - v\|$. In this case the two-step procedure, based on the observation $X \cap W$ and using the adaptive test function (4), provides $\hat{\rho} = N(X \cap W)/|W|$ and $\hat{\psi}$ is the root of

$$e_2(\psi) = \sum_{r_{ij}} w \left(\varepsilon \frac{g(0; \psi) - 1}{g(r_{ij}; \psi) - 1} \right) \frac{\nabla_{\psi} g(r_{ij}; \psi)}{g(r_{ij}; \psi)} - N(X \cap W)^2 \int_W w \left(\varepsilon \frac{g(0; \psi) - 1}{g(r; \psi) - 1} \right) \nabla_{\psi} g(r; \psi) dF(r). \quad (8)$$

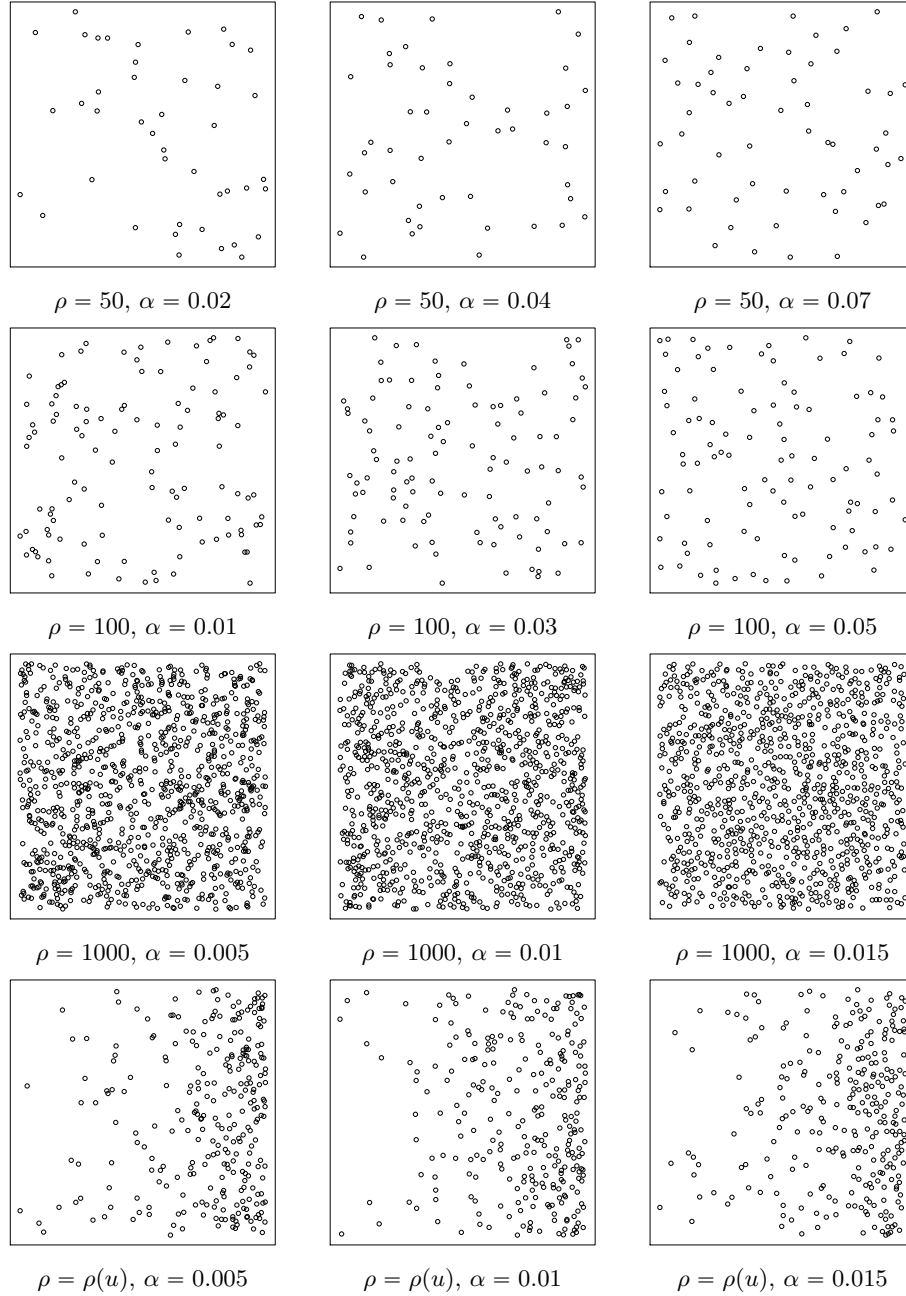


Figure 1. Examples of point patterns simulated from a Bessel-type DPP on $[0, 1]^2$ for different values of ρ and α . For the last row, $\rho(x, y) = 20 \exp(4x)$.

ρ	α		$R = 0.05$	$R = 0.1$	$R = 0.25$	Adaptive	$\tilde{R}$
50	0.02	RMSE:	5.84	5.83	6.29	5.97	0.047
			(0.15)	(0.17)	(0.19)	(0.18)	(0.020)
	0.04	TIME:	0.43	0.48	0.68	0.64	
		RMSE:	15.60	9.18	9.19	9.25	0.106
	0.07		(0.44)	(0.20)	(0.22)	(0.21)	(0.037)
		TIME:	0.48	0.50	0.68	0.73	
100	0.01	RMSE:	13.32	8.25	8.22	8.15	0.147
			(0.33)	(0.23)	(0.24)	(0.24)	(0.050)
	0.03	TIME:	0.50	0.45	0.59	0.72	
		RMSE:	2.44	2.45	2.58	2.63	0.024
	0.05		(0.08)	(0.08)	(0.09)	(0.09)	(0.009)
		TIME:	0.44	0.57	1.22	0.70	
1000	0.005	RMSE:	5.34	5.12	5.28	5.27	0.064
			(0.13)	(0.13)	(0.14)	(0.13)	(0.019)
	0.01	TIME:	0.40	0.47	0.98	0.70	
		RMSE:	5.78	4.43	4.50	4.53	0.139
	0.015		(0.12)	(0.12)	(0.10)	(0.12)	(0.022)
		TIME:	0.52	0.56	1.16	0.95	
Inhom	0.005	RMSE:	0.67	0.88	0.83	0.72	0.015
			(0.02)	(0.02)	(0.02)	(0.02)	(0.003)
	0.01	TIME:	3.83	19.04	110.07	9.38	
		RMSE:	0.57	0.59	0.61	0.56	0.028
	0.015		(0.01)	(0.02)	(0.01)	(0.01)	(0.005)
		TIME:	2.68	10.40	60.79	6.84	
	0.005	RMSE:	0.47	0.46	0.52	0.47	0.026
			(0.01)	(0.01)	(0.01)	(0.01)	(0.002)
	0.01	TIME:	2.53	9.81	55.78	7.75	
		RMSE:	1.58	1.65	1.66	1.61	0.014
	0.015		(0.04)	(0.04)	(0.04)	(0.04)	(0.005)
		TIME:	0.89	2.50	10.30	1.19	
	0.005	RMSE:	1.34	1.36	1.36	1.32	0.025
			(0.03)	(0.03)	(0.03)	(0.03)	(0.008)
	0.01	TIME:	0.76	1.86	7.66	1.22	
		RMSE:	1.43	1.47	1.48	1.40	0.030
	0.015		(0.03)	(0.03)	(0.03)	(0.03)	(0.006)
		TIME:	0.86	1.90	7.46	1.40	

Table 1. Estimated root mean square errors ($\times 10^3$) and mean computation time (in seconds) of $\hat{\alpha}$ for a Bessel-type DPP on $[0, 1]^2$, for different values of ρ and α . The 3 first estimators use the test function (3) with $R = 0.05$, $R = 0.1$ and $R = 0.25$ respectively, while the last estimator is the adaptive version based on (4). The standard errors of the RMSE estimations are given in parenthesis. The last column gives the averages of "practical ranges" (i.e. maximal solution to $|g(r) - 1| = 0.01$) used for the adaptive estimator, along with their standard deviations in parenthesis. For each value of ρ and α , these quantities are computed from 1000 simulations where all four estimation methods have converged.

Here F denotes the cumulative distribution function of $R = \|U - V\|$ where U and V are independent variables uniformly distributed on W and $\{r_{ij}\}$ is the set of all pairwise distances of $X \cap W$. On the other hand, by a simultaneous procedure using the same test function, we get that $\hat{\psi}$ is the root of

$$e(\psi) = \sum_{r_{ij}} w \left(\varepsilon \frac{g(0; \psi) - 1}{g(r_{ij}; \psi) - 1} \right) \frac{\nabla_{\psi} g(r_{ij}; \psi)}{g(r_{ij}; \psi)} - \frac{\sum_{r_{ij}} w \left(\varepsilon \frac{g(0; \psi) - 1}{g(r_{ij}; \psi) - 1} \right)}{\int w \left(\varepsilon \frac{g(0; \psi) - 1}{g(r; \psi) - 1} \right) g(r; \psi) dF(r)} \int w \left(\varepsilon \frac{g(0; \psi) - 1}{g(r_{ij}; \psi) - 1} \right) \nabla_{\psi} g(r; \psi) dF(r), \quad (9)$$

while $\hat{\rho}$ is given by

$$\hat{\rho}^2 = \frac{1}{|W|^2} \frac{\sum_{r_{ij}} w \left(\varepsilon \frac{g(0; \hat{\psi}) - 1}{g(r_{ij}; \hat{\psi}) - 1} \right)}{\int w \left(\varepsilon \frac{g(0; \hat{\psi}) - 1}{g(r; \hat{\psi}) - 1} \right) g(r; \hat{\psi}) dF(r)}. \quad (10)$$

The more complicated expression of (9) in comparison with (8) implies that $e(\psi)$ can be highly irregular in ψ . Figure S2 in the supplementary material shows an example for one realisation of a DPP with a Gaussian kernel with range ψ . For this example $e(\psi)$ exhibits many different roots, although the dataset contains a fairly large number of points (about 1000). The consequence is an extreme sensitivity to the initial parameter when we try to solve $e(\psi) = 0$. In contrast $e_2(\psi) = 0$ has one clear solution. This advocates the use of the two-step approach.

Due to the aforementioned very strong sensitivity to the initial value of ψ , conclusions from comparison of the simultaneous estimate of ψ with the two-step estimate of ψ can be quite arbitrary. However, we report in Figure S3 in the supplementary material the distribution of estimates of ρ from 1000 simulations of a Bessel-type DPP with $\rho = 1000$ and $\psi = \alpha = 0.01$, using either (10) from the simultaneous approach or $\hat{\rho} = N(X \cap W)/|W|$ from the two-step approach. For the simultaneous method we either chose the true value $\alpha = 0.01$ as the starting point for the numerical solution of $e(\alpha) = 0$ to get $\hat{\alpha}$, or fixed $\hat{\alpha}$ at the true value, i.e. $\hat{\alpha} = 0.01$, in (10). The estimate $\hat{\rho} = N(X \cap W)/|W|$ is unequivocally better than (10) in terms of root mean square error, even when the true value of α is used for $\hat{\alpha}$ in (10). This confirms our recommendation.

The simultaneous estimation approach in this example is covered by our theoretical results in Sections 3.1 and 3.2. It shows that while our consistency result guarantees the existence of a consistent sequence of parameter estimates (roots) there could also exist other non-consistent sequences.

5. Discussion

In this paper we provide a very general asymptotic framework for estimating function inference for spatial point processes with known joint intensities. Specific asymptotic results are obtained for determinantal point processes.

The performance of second order estimating functions depends strongly on a tuning parameter that controls which pairs of points are used in the estimation. Our adaptive choice of this tuning parameter is intuitively appealing, easy to implement and performs well in the simulation studies considered. It moreover seamlessly integrates with the asymptotic results where the use of the adaptive method poses no extra theoretical difficulties. Though we focus in this paper on determinantal point processes, the adaptive method is applicable for any spatial point process with known pair correlation function. As an example we provide in Section 3 of the supplementary material a simulation study in case of a cluster process.

Appendix A: Assumptions and proof of Theorem 2.1

Our general Theorem 2.1 depends on a number of assumptions. The setting is the same as in Section 2.2. We moreover define $\text{diam}(x)$ as the largest distance between two coordinates of x . The assumptions (F1) through (F3) are mainly related to the test functions f_i , while for X we assume (X1) through (X3).

- (F1) For all $i = 1, \dots, l$ and for all $x \in (\mathbb{R}^d)^{q_i}$, $\theta \mapsto f_i(x; \theta)$ is twice continuously differentiable in a neighbourhood of θ^* . Moreover, the first and second derivative of f_i with respect to θ are bounded with respect to $x \in (\mathbb{R}^d)^{q_i}$ uniformly in θ belonging to this neighbourhood.
- (F2) There exists a constant $R > 0$ such that for all θ in a neighbourhood of θ^* , all functions $x \mapsto f_i(x; \theta)$ vanish when $\text{diam}(x) > R$.

Define the matrices $H_n(\theta)$ by

$$H_n(\theta) = \begin{pmatrix} H_n^1(\theta) \\ \vdots \\ H_n^l(\theta) \end{pmatrix},$$

where for all i

$$H_n^i(\theta) := \frac{1}{|W_n|} \int_{W_n^{q_i}} f_i(x; \theta) \nabla_{\theta} \rho^{(q_i)}(x; \theta)^T dx.$$

- (F3) The matrices $H_n(\theta^*)$ satisfy

$$\liminf_{n \rightarrow \infty} \left(\inf_{\|\phi\|=1} \phi^T H_n(\theta^*) \phi \right) > 0.$$

- (F3') There exists a neighbourhood of θ^* such that for all n high enough and all θ in this neighbourhood, $H_n(\theta)$ is invertible and $\|H_n(\theta)^{-1}\|$ is uniformly bounded with respect to n and θ , where $\|\cdot\|$ stands for any matrix norm.
- (X1) For all θ in a neighbourhood of θ^* and all q_i , $i = 1, \dots, l$, the intensity functions $x \mapsto \rho^{(q_i)}(x; \theta)$ are well-defined and bounded. Moreover, $\theta \mapsto \rho^{(q_i)}(x; \theta)$ is twice

continuously differentiable in a neighbourhood of θ^* , for all $x \in (\mathbb{R}^d)^{q_i}$. Finally, the first and second derivative of $\rho^{(q_i)}$ with respect to θ are bounded with respect to $x \in (\mathbb{R}^d)^{q_i}$ uniformly in θ belonging to this neighbourhood.

- (X2) For all q_i , $i = 1, \dots, l$, the intensity functions $\rho^{(q_i)}(\cdot; \theta^*), \dots, \rho^{(2q_i)}(\cdot; \theta^*)$ of X are well-defined. Moreover, the intensity functions $\rho^{(q_i)}(\cdot; \theta^*), \dots, \rho^{(2q_i-1)}(\cdot; \theta^*)$ are bounded and for all bounded sets $W \subset \mathbb{R}^d$ there exists a constant $C_0 > 0$, so that $\int_W \varphi_i(x_1) dx_1 < C_0$, $i = 1, \dots, l$ where φ_i is the function

$$\begin{aligned} \varphi_i : x_1 \mapsto & \sup_{\text{diam}(x) < R} \sup_{\text{diam}(y) < R} \sup_{y_1 \in W} \rho^{(2q_i)}(x_1, x_2, \dots, x_{q_i}, y_1, \dots, y_{q_i}; \theta^*) \\ & - \rho^{(q_i)}(x_1, x_2, \dots, x_{q_i}; \theta^*) \rho^{(q_i)}(y_1, \dots, y_{q_i}; \theta^*) \end{aligned}$$

with R coming from (F2).

- (X3) X satisfies the central limit theorem

$$\Sigma_n^{-1/2} e_n(\theta^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I_p),$$

where e_n is defined in Section 2.2 and $\Sigma_n = \text{Var}(e_n(\theta^*))$.

Assumptions (F1) and (F2) are basic regularity conditions on the f_i 's. Similarly (X1) and (X2) ensure that the intensity functions of X exist and are sufficiently regular. The technical assumptions are in fact (F3) (or (F3')) and (X3). While the latter strongly depends on the underlying point process (see [23] for Cox processes and [14] for DPPs), the former can be simplified in some cases. For example, if $H_n(\theta^*)$ are symmetrical matrices for all n then (F3) writes $\liminf_n \lambda_{\min}(H_n(\theta^*)) > 0$ where $\lambda_{\min}(H_n(\theta^*))$ denotes the smallest eigenvalue of $H_n(\theta^*)$. If the matrices $H_n(\theta^*)$ are not symmetrical, Assumption (F3') will be preferred since (F3) does not translate well for non-symmetrical matrices. Furthermore, if X is stationary and all f_i 's are invariant by translation, then $H_n(\theta)$ converges towards a matrix $H(\theta)$ explicitly given in Lemma A.1 below, thus Assumption (F3) simply becomes $\inf_{\|\phi\|=1} \phi^T H(\theta^*) \phi > 0$ and (F3') is satisfied whenever $H(\theta^*)$ is invertible by continuity of $H(\theta)$.

Lemma A.1. *Assume (W), (X1), (F2) and let $\theta \in \mathbb{R}^p$. Suppose that all $\rho^{(q_i)}(\cdot; \theta)$'s and $f_i(\cdot; \theta)$'s are invariant by translation, i.e. $f_i(u_1, u; \theta) = f_i(0, u - u_1; \theta)$ where we denote by u the vector $(u_2, \dots, u_{q_i})$. If $u \mapsto f_i(0, u; \theta)$ is integrable for all i such that $q_i \geq 2$, then $H_n(\theta)$ converges to a matrix $H(\theta)$. In particular, for all i we have*

$$\lim_{n \rightarrow \infty} H_n^i(\theta) = \int_{\|t\| \leq R} f_i(0, t; \theta) \nabla_{\theta} \rho^{(q_i)}(0, t; \theta)^T dt.$$

This lemma is verified in Section B. We now turn to the proof of the theorem. To prove the consistency of $\hat{\theta}_n$ and get its rate of convergence we apply the following result, where $\|\cdot\|$ stands for any matrix norm.

Theorem A.2 ([23]). *Suppose that $e_n(\theta)$ is continuously differentiable with respect to θ and define*

$$J_n(\theta) := -\frac{d}{d\theta^T} e_n(\theta) := -\left(\frac{\partial}{\partial \theta_j} e_n(\theta)_i \right)_{1 \leq i, j \leq p}.$$

Suppose that for all $\alpha > 0$

$$\sup_{\theta \in M_n^\alpha(\theta^*)} \left\| \frac{1}{|W_n|} (J_n(\theta) - J_n(\theta^*)) \right\| \xrightarrow{\mathbb{P}} 0, \quad (11)$$

where

$$M_n^\alpha(\theta^*) := \left\{ \theta \in \Theta : \|\theta - \theta^*\| \leq \frac{\alpha}{\sqrt{|W_n|}} \right\},$$

and suppose that there exists $l > 0$ such that

$$\mathbb{P} \left(\frac{1}{|W_n|} \inf_{\|\phi\|=1} \phi^T J_n(\theta^*) \phi < l \right) \rightarrow 0. \quad (12)$$

Assume, moreover, that the class of random vectors

$$\left\{ \frac{1}{\sqrt{|W_n|}} e_n(\theta^*) : n \in \mathbb{N} \right\}$$

is stochastically bounded. Then, for all $\varepsilon > 0$, there exists $d > 0$ such that

$$\mathbb{P}(\exists \hat{\theta}_n : e_n(\hat{\theta}_n) = 0 \text{ and } |W_n| \|\hat{\theta}_n - \theta^*\| < d) > 1 - \varepsilon \quad (13)$$

for a sufficiently large n .

We now verify the assumptions of Theorem A.2. There is no loss in generality by assuming that all f_i are symmetric functions. Otherwise we can just replace $f_i(x)$ by its symmetrized version $(q_i!)^{-1} \sum_{u \in \pi(x)} f_i(u)$ where $\pi(x)$ denotes the set of all vectors obtained by permuting the components of x . This does not change the value of $e_n(\theta)$ and each symmetrized function still satisfies Assumptions (F1) through (F3). We will use at several places the following result.

Lemma A.3. *Let X be a point process satisfying Assumption (X2). Consider any $i \in \{1, \dots, l\}$, any bounded set $W \subset \mathbb{R}^d$, and any symmetric bounded function $g : (\mathbb{R}^d)^{q_i} \rightarrow \mathbb{R}^{k_i}$ vanishing when two of its components are at a distance greater than R for a given constant $R > 0$. Then*

$$\left\| \text{Var} \left(\sum_{x_1, \dots, x_{q_i} \in X \cap W}^{\neq} g(x_1, \dots, x_{q_i}) \right) \right\| = O(|W|).$$

Proof. Since g is a symmetric function, then $g(x_1, \dots, x_{q_i})$ does not depend on the order of the x_i . Thus, for any set of q_i points $S = \{x_1, \dots, x_{q_i}\}$, we can write $g(S)$ for the value of g at an arbitrary order of the points in S and we write

$$\sum_{\substack{\neq \\ x_1, \dots, x_{q_i} \in X \cap W}} g(x_1, \dots, x_{q_i}) = q_i! \sum_{S \subset X \cap W} g(S) \mathbb{1}_{|S|=q_i}.$$

We start by expanding $\mathbb{E}[(\sum_{S \subset X \cap W} g(S) \mathbb{1}_{|S|=q_i}) (\sum_{S \subset X \cap W} g^T(S) \mathbb{1}_{|S|=q_i})]$ as

$$\begin{aligned} & \sum_{k=0}^{q_i} \mathbb{E} \left[\sum_{\substack{S, T \subset X \cap W \\ |S|=|T|=q_i, |S \cap T|=k}} g(S) g^T(T) \right] \\ &= \sum_{k=0}^{q_i} \mathbb{E} \left[\sum_{\substack{U \subset X \cap W \\ |U|=2q_i-k}} \sum_{\substack{S' \subset S \subset U \\ |S'|=k, |S|=q_i}} g(S) g^T(S' \cup (U \setminus S)) \right] \\ &= \sum_{k=0}^{q_i} \frac{1}{(2q_i - k)!} \int_{W^{2q_i-k}} \sum_{\substack{S' \subset S \subset \{x_1, \dots, x_{2q_i-k}\} \\ |S'|=k, |S|=q_i}} g(S) g^T(S' \cup (\{x_1, \dots, x_{2q_i-k}\} \setminus S)) \rho^{(2q_i-k)}(x; \theta^*) dx \\ &= \sum_{k=0}^{q_i} \frac{\binom{q_i}{k} \binom{2q_i-k}{q_i}}{(2q_i - k)!} \int_{W^{2q_i-k}} g(x_1, \dots, x_{q_i}) g^T(x_1, \dots, x_k, x_{q_i+1}, \dots, x_{2q_i-k}) \rho^{(2q_i-k)}(x; \theta^*) dx. \end{aligned} \tag{14}$$

By Assumption (X2), the functions $\rho^{(q_i)}, \dots, \rho^{(2q_i-1)}$ are all bounded. Moreover, as a consequence of our assumptions on g , each component of each term for $k \geq 1$ in (14) is bounded by

$$\frac{1}{q_i!(q_i - k)!} \binom{q_i}{k} \int_{W^{2q_i-k}} \|g\|_\infty^2 \|\rho^{(2q_i-k)}\|_\infty \mathbb{1}_{\{0 \leq |x_i - x_1| \leq R, \forall i\}} dx$$

which is $O(|W|)$. However, for $k = 0$, the term is $O(|W|^2)$. Instead of controlling this term alone, we consider its difference with the remaining term in the variance we are looking at, that is

$$\begin{aligned} & \frac{1}{(q_i!)^2} \int_{W^{2q_i}} g(x) g^T(y) \rho^{(2q_i)}(x, y; \theta^*) dx dy \\ & \quad - \mathbb{E} \left[\sum_{S \subset X \cap W} g(S) \mathbb{1}_{|S|=q_i} \right] \mathbb{E} \left[\sum_{S \subset X \cap W} g(S) \mathbb{1}_{|S|=q_i} \right]^T \\ &= \frac{1}{(q_i!)^2} \int_{W^{2q_i}} g(x) g^T(y) (\rho^{(2q_i)}(x, y; \theta^*) - \rho^{(q_i)}(x; \theta^*) \rho^{(q_i)}(y; \theta^*)) dx dy. \end{aligned}$$

All of its components are bounded by

$$\frac{\|g\|_\infty^2 |W| |\mathcal{B}(O, R)|^{2q_i-2}}{(q_i!)^2} \int_W \sup_{\text{diam}(x) < R} \sup_{\text{diam}(y) < R} \sup_{y_1 \in W} \rho^{(2q_i)}(x_1, x_2, \dots, x_{q_i}, y_1, \dots, y_{q_i}; \theta^*) \\ - \rho^{(q_i)}(x_1, x_2, \dots, x_{q_i}; \theta^*) \rho^{(q_i)}(y_1, \dots, y_{q_i}; \theta^*) dx_1$$

which is $O(|W|)$ by Assumption (X2). $\square$

The regularity conditions on $e_n(\theta)$ in Theorem A.2 are consequences of (F1), (X1). The stochastic behavior of $e_n(\theta^*)$ is easily deduced from the previous lemma.

Lemma A.4. *The class of random vectors*

$$\left\{ \frac{1}{\sqrt{|W_n|}} e_n(\theta^*) : n \in \mathbb{N} \right\}$$

is stochastically bounded.

Proof. The result follows by showing that each component $e_n^i(\theta^*)$ of $e_n(\theta^*)$ is stochastically bounded. By Chebyshev's inequality, we just need to bound $|W_n|^{-1} \text{Var}(e_n^i(\theta^*))$. Letting

$$e_n^i(\theta) := \sum_{x_1, \dots, x_{q_i} \in X \cap W_n}^{\neq} f_i(x_1, \dots, x_{q_i}; \theta) - \int_{W_n^{q_i}} f_i(x; \theta) \rho^{(q_i)}(x; \theta) dx,$$

we know that $\text{Var}(e_n^i(\theta^*))$ is $O(|W_n|)$ by Lemma A.3 under Assumptions (X2) and (F2). $\square$

To apply Theorem A.2 under Assumptions (F1) through (F3), it remains to show the following lemma.

Lemma A.5. *Under Assumptions (F1) through (F3), (X1) and (X2) we have for all $\alpha > 0$,*

$$\sup_{\theta \in M_n^\alpha(\theta^*)} \left\| \frac{1}{|W_n|} (J_n(\theta) - J_n(\theta^*)) \right\| \xrightarrow{\mathbb{P}} 0. \quad (15)$$

where $M_n^\alpha(\theta^*)$ is defined as in Theorem A.2 and there exists $l > 0$ such that

$$\mathbb{P} \left(\frac{1}{|W_n|} \inf_{\|\phi\|=1} \phi^T J_n(\theta^*) \phi < l \right) \rightarrow 0.$$

Proof. We write

$$J_n(\theta) = \begin{pmatrix} J_n^1(\theta) \\ \vdots \\ J_n^l(\theta) \end{pmatrix} := - \begin{pmatrix} \frac{d}{d\theta^T} e_n^1(x; \theta) \\ \vdots \\ \frac{d}{d\theta^T} e_n^l(x; \theta) \end{pmatrix}$$

By definition, for all i

$$J_n^i(\theta) = -q_i! \sum_{\substack{x \in X \cap W_n \\ |x|=q_i}} \frac{d}{d\theta^T} f_i(x; \theta) + \int_{W_n^{q_i}} \frac{d}{d\theta^T} f_i(x; \theta) \rho^{(q_i)}(x; \theta) dx \\ + \int_{W_n^{q_i}} f_i(x; \theta) \nabla_{\theta} \rho^{(q_i)}(x; \theta)^T dx. \quad (16)$$

Now, recall that f_i , $\frac{d}{d\theta^T} f_i$, $\rho^{(q_i)}$ and $\nabla_{\theta} \rho^{(q_i)}$ are all continuously differentiable with respect to θ by Assumption (F1) and (X1). Moreover, the first and second derivatives of f_i and $\rho^{(q_i)}$ with respect to θ are bounded with respect to x and θ by the same assumptions. Therefore, since $M_n^{\alpha}(\theta^*)$ is a decreasing sequence of compact sets, there exist constants $C_1, C_2 > 0$ not depending on n , x and θ such that by a Taylor expansion,

$$\sup_{\theta \in M_n^{\alpha}(\theta^*)} \|J_n^i(\theta) - J_n^i(\theta^*)\| \leq \frac{\alpha}{\sqrt{|W_n|}} \left(C_1 \sum_{\substack{x \in X \cap W_n \\ |x|=q_i}} \mathbb{1}_{\text{diam}(x) \leq R} + C_2 \int_{W_n^{q_i}} \mathbb{1}_{\text{diam}(x) \leq R} dudv \right)$$

where the indicator functions arise as a consequence of Assumption (F2). Moreover,

$$\mathbb{E} \left[\sum_{\substack{x \in X \cap W_n \\ |x|=q_i}} \mathbb{1}_{\text{diam}(x) \leq R} \right] = \int_{W_n^{q_i}} \rho^{(q_i)}(x; \theta^*) \mathbb{1}_{\text{diam}(x) \leq R} dx = O(|W_n|)$$

since $\rho^{(q_i)}$ is bounded by Assumption (X2). This shows that $\mathbb{E}[\sup_{\theta \in M_n^{\alpha}(\theta^*)} \|J_n^i(\theta) - J_n^i(\theta^*)\|]$ is $O(\sqrt{|W_n|})$.

It remains to prove that there exists $l > 0$ such that (A.5) holds. By Assumption (F3) choose $\varepsilon > 0$ so that $\liminf_{n \rightarrow \infty} \phi^T H_n(\theta^*) \phi > \varepsilon$ and let $l = \varepsilon/2$. In the case where $\theta = \theta^*$, the second term in (16) is just the expectation of the first one and the third term is equal to $|W_n| H_n^i(\theta^*)$ which is deterministic. Thus when $\theta = \theta^*$, the L^2 norm of the first two terms in (16) is equal to

$$\sqrt{\text{Var} \left(\sum_{\substack{x \in X \cap W_n \\ |x|=q_i}} \frac{d}{d\theta^T} f_i(x; \theta^*) \right)}$$

which is $O(\sqrt{|W_n|})$ by Lemma A.3. Hence it vanishes in probability when divided by $|W_n|$. Denote by a_n the first two terms in $\phi^T J_n(\theta^*) \phi / |W_n|$ and by b_n the last term which is $\phi^T H_n(\theta^*) \phi$. Then

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{1}{|W_n|} \inf_{\|\phi\|=1} \phi^T J_n(\theta^*) \phi < l \right) \leq \lim_{n \rightarrow \infty} \mathbb{P} \left(\inf_{\|\phi\|=1} b_n - \varepsilon/2 < l, |a_n| < \varepsilon/2 \right) \\ \leq \lim_{n \rightarrow \infty} \mathbb{P} \left(\inf_{\|\phi\|=1} b_n < \varepsilon \right) = 0$$

which concludes the proof. $\square$

To apply Theorem A.2 under the alternative Assumption (F3') instead of (F3), we proceed as follows. Consider n to be large enough and θ to be in a neighbourhood of θ^* such that $H_n(\theta)$ is invertible and $H_n(\theta)^{-1}$ is uniformly bounded with respect to n and θ . Let $\tilde{e}_n(\theta) = H_n(\theta)^{-1}e_n(\theta)$ and let us show that we can apply Theorem A.2 to $\tilde{e}_n$. Obviously, $\tilde{e}_n$ has the same roots as e_n , is continuously differentiable since $\theta \mapsto H_n(\theta)$ and $\theta \mapsto H_n(\theta)^{-1}$ are continuously differentiable from Assumptions (F1) and (F2), and the family $\{\tilde{e}_n(\theta^*)/\sqrt{|W_n|} : n \in \mathbb{N}\}$ is stochastically bounded. Let $\tilde{J}_n(\theta) = -\frac{d}{d\theta^T}\tilde{e}_n(\theta)$. It remains to show the following lemma.

Lemma A.6. *Under Assumptions (F1), (F2), (F3'), (X1) and (X2) we have*

$$\sup_{\theta \in M_n^\alpha(\theta^*)} \left\| \frac{1}{|W_n|} (\tilde{J}_n(\theta) - \tilde{J}_n(\theta^*)) \right\| \xrightarrow{\mathbb{P}} 0. \quad (17)$$

and

$$\mathbb{P} \left(\lim_{n \rightarrow \infty} \frac{1}{|W_n|} \inf_{\|\phi\|=1} \phi^T \tilde{J}_n(\theta^*) \phi < 1/2 \right) = 0. \quad (18)$$

Proof. We have

$$\tilde{J}_n(\theta) = H_n(\theta)^{-1} J_n(\theta) - T_n(\theta) \quad \text{where} \quad T_n(\theta)_{i,j} = \sum_{k=1}^p \frac{\partial}{\partial \theta_j} H_n(\theta)_{i,k}^{-1} e_n(\theta)_k. \quad (19)$$

For any $\theta \in M_n^\alpha(\theta^*)$, since all terms in (16) are bounded by Assumptions (F1) and (X1), using Assumption (F2) we get

$$\mathbb{E}[|e_n(\theta) - e_n(\theta^*)|] \leq \frac{\alpha}{\sqrt{|W_n|}} \mathbb{E} \left(\sup_{\theta \in M_n^\alpha(\theta^*)} \|J_n(\theta)\| \right) = O(\sqrt{|W_n|}).$$

By (F1), (X1) and (F3'), $\frac{\partial}{\partial \theta_j} H_n(\theta)^{-1} = (\frac{\partial}{\partial \theta_j} H_n(\theta)) H_n(\theta)^{-2}$ is bounded on $M_n^\alpha(\theta^*)$ for a large enough n . It follows for all i, j ,

$$\frac{1}{|W_n|} \sup_{\theta \in M_n^\alpha(\theta^*)} \left\| \sum_{k=1}^p \frac{\partial}{\partial \theta_j} H_n(\theta)_{i,k}^{-1} e_n(\theta^*)_k - T_n(\theta)_{i,j} \right\| \xrightarrow{\mathbb{P}} 0.$$

Moreover,

$$\sup_{\theta \in M_n^\alpha(\theta^*)} \sum_{k=1}^p \frac{\partial}{\partial \theta_j} H_n(\theta)_{i,k}^{-1} \frac{e_n(\theta^*)_k}{|W_n|} \xrightarrow{\mathbb{P}} 0$$

as a consequence of Lemma A.4. Hence $|W_n|^{-1} \sup_{\theta \in M_n^\alpha(\theta^*)} \|T_n(\theta)\| \xrightarrow{\mathbb{P}} 0$ and thus $|W_n|^{-1} \|T_n(\theta^*)\| \xrightarrow{\mathbb{P}} 0$. Therefore, we only need to look at the behaviour of $H_n(\theta)^{-1} J_n(\theta)$.

From Lemma A.5 we know that

$$\sup_{\theta \in M_n^\alpha(\theta^*)} \left\| \frac{1}{|W_n|} H_n(\theta^*)^{-1} (J_n(\theta) - J_n(\theta^*)) \right\| \xrightarrow{\mathbb{P}} 0.$$

Finally, we observe that

$$\begin{aligned} & \mathbb{E} \left(\sup_{\theta \in M_n^\alpha(\theta^*)} \|(H_n(\theta)^{-1} - H_n(\theta^*)^{-1})J_n(\theta)\| \right) \\ & \leq \frac{\alpha}{\sqrt{|W_n|}} \mathbb{E} \left(\sup_{\theta \in M_n^\alpha(\theta^*)} \|J_n(\theta)\| \right) \sup_{\theta \in M_n^\alpha(\theta^*)} \sup_{1 \leq j \leq p} \left\| \frac{\partial}{\partial \theta_j} H_n(\theta)^{-1} \right\| = O(\sqrt{|W_n|}) \end{aligned}$$

where the boundedness of $\partial H_n(\theta)^{-1}/\partial \theta_j$ for each j was noted above and the boundedness of $\mathbb{E} \left(\sup_{\theta \in M_n^\alpha(\theta^*)} \|J_n(\theta)\| \right)$ follows by considerations in the last part of the proof of Lemma A.5 as a consequence of the regularity assumptions imposed on $H_n(\theta)$ by Assumption (F3'). This finishes proving (17). The result (18) is then a consequence of the fact that $H_n(\theta^*)^{-1}J_n(\theta^*)$ converges towards I_p when n goes to infinity. $\square$

Finally, by Lemmas A.4, A.5 and A.6, we can apply Theorem A.2 and the first part in the statements of Theorem 2.1 is deduced.

Now, for each $n \in \mathbb{N}$, we define $\hat{\theta}_n$ as the closest root of e_n to θ^* , if e_n has any, otherwise let $\hat{\theta}_n = 0$. Theorem A.2 tells us that $\mathbb{P}(e_n(\hat{\theta}_n) = 0) \rightarrow 1$ and $\sqrt{|W_n|}(\hat{\theta}_n - \theta^*)$ is bounded in probability.

To prove the asymptotic normality, we use the Taylor expansion $e_n(\hat{\theta}_n) = e_n(\theta^*) + J_n(\theta_n^0)(\hat{\theta}_n - \theta^*)$ where $\|\theta_n^0 - \theta^*\| \leq \|\hat{\theta}_n - \theta^*\|$, which implies

$$\begin{aligned} \frac{e_n(\hat{\theta}_n)}{\sqrt{|W_n|}} &= \frac{e_n(\theta^*)}{\sqrt{|W_n|}} + \frac{J_n(\theta^*)}{|W_n|} \sqrt{|W_n|}(\hat{\theta}_n - \theta^*) \\ &\quad + \frac{1}{|W_n|} (J_n(\theta_n^0) - J_n(\theta^*)) \sqrt{|W_n|}(\hat{\theta}_n - \theta^*). \end{aligned}$$

We know that $e_n(\hat{\theta}_n)/\sqrt{|W_n|}$ converges in distribution towards 0 and by Theorem A.2 we also know that

$$\left\| \frac{1}{|W_n|} (J_n(\theta_n^0) - J_n(\theta^*)) \right\| \xrightarrow{\mathbb{P}} 0$$

because θ_n^0 is closer to θ^* than $\hat{\theta}_n$ with probability tending to 1. Moreover, we saw at the end of the proof of Theorem A.2 that the variance of the first two terms of $|W_n|^{-1}J_n(\theta^*)$ vanishes when $n \rightarrow \infty$ and the last term is equal to $H_n(\theta^*)$. Finally, by Assumption (X3) and since $|W_n|^{-1}\text{Var}(e_n(\theta^*))$ is stochastically bounded (Lemma A.4), it follows by Slutsky's lemma that

$$|W_n| \Sigma_n^{-1/2} H_n(\theta^*)(\hat{\theta}_n - \theta^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I_p).$$

Appendix B: Proof of Lemma A.1

For all i , if $q_i = 1$ then $H_n^i(\theta)$ is constant. Otherwise, since f_i and X are stationary, the integral in (F3) writes

$$\begin{aligned} H_n^i(\theta) &= \int_{W_n} \int_{W_n^{q_i-1}} f_i(u_1, u; \theta) \nabla_{\theta} \rho^{(q_i)}(u_1, u; \theta)^T du du_1 \\ &= \int_{W_n^{\ominus R}} \int_{W_n^{q_i-1}} f_i(0, u - u_1; \theta) \nabla_{\theta} \rho^{(q_i)}(0, u - u_1; \theta)^T du du_1 + \varepsilon_n \end{aligned}$$

where

$$\varepsilon_n = \int_{W_n \setminus W_n^{\ominus R}} \int_{W_n^{q_i-1}} f_i(0, u - u_1; \theta) \nabla_{\theta} \rho^{(q_i)}(0, u - u_1; \theta)^T du du_1.$$

By integrability of f_i , (W), (X1) and (F2), we have

$$\begin{aligned} \frac{|\varepsilon_{n,kl}|}{|W_n|} &\leq \frac{1}{|W_n|} \int_{W_n \setminus W_n^{\ominus R}} \int_{(\mathbb{R}^d)^{q_i-1}} |f_i(0, u - u_1; \theta)_k| \|\nabla_{\theta} \rho^{(q_i)}(\cdot; \theta)\|_{\infty} du du_1 \\ &= \frac{|W_n \setminus W_n^{\ominus R}|}{|W_n|} \|\nabla_{\theta} \rho^{(q_i)}(\cdot; \theta)\|_{\infty} \int_{\|t\| \leq R} |f_i(0, t; \theta)_k| dt \\ &\leq \frac{|\partial W_n \oplus R|}{|W_n|} \|\nabla_{\theta} \rho^{(q_i)}(\cdot; \theta)\|_{\infty} \int_{\|t\| \leq R} |f_i(0, t; \theta)_k| dt \rightarrow 0, \end{aligned}$$

where $\varepsilon_{n,kl}$ denotes the kl th entry of the matrix ε_n and $f_i(\cdot)_k$ the k th component of the vector $f_i(\cdot)$. Moreover,

$$\begin{aligned} &\frac{1}{|W_n|} \int_{W_n^{\ominus R}} \int_{W_n} f_i(0, u - u_1; \theta) \nabla_{\theta} \rho^{(q_i)}(0, u - u_1; \theta)^T du du_1 \\ &= \frac{|W_n^{\ominus R}|}{|W_n|} \int_{\|t\| \leq R} f_i(0, t; \theta) \nabla_{\theta} \rho^{(q_i)}(0, t; \theta)^T dt \\ &\rightarrow \int_{\|t\| \leq R} f_i(0, t; \theta) \nabla_{\theta} \rho^{(q_i)}(0, t; \theta)^T dt, \end{aligned}$$

as $n \rightarrow \infty$. This proves the convergence of $H_n(\theta)$.

Acknowledgements

Rasmus Waagepetersen is supported by The Danish Council for Independent Research — Natural Sciences, grant DFF - 7014-00074 "Statistics for point processes in space and beyond", and by the "Centre for Stochastic Geometry and Advanced Bioimaging", funded by grant 8721 from the Villum Foundation.

References

- [1] BADDELEY, A. J., RUBAK, E. and TURNER, R. (2015). *Spatial Point Patterns: Methodology and Applications with R. Interdisciplinary Statistics*. Chapman & Hall/CRC, Boca Raton, Florida.
- [2] BISCIO, C. A. N. and LAVANCIER, F. (2016). Quantifying repulsiveness of determinantal point processes. *Bernoulli* **22** 2001-2028.
- [3] BISCIO, C. A. N. and LAVANCIER, F. (2017). Contrast estimation for parametric stationary determinantal point processes. *Scandinavian Journal of Statistics* **44** 204-229.
- [4] DENG, C., GUAN, Y., WAAGEPETERSEN, R. and ZHANG, J. (2017). Second-order quasi-likelihood for spatial point processes. *Biometrics* **73** 1311-1320.
- [5] GUAN, Y. (2006). A composite likelihood approach in fitting spatial point process models. *Journal of the American Statistical Association* **101** 1502-1512.
- [6] GUAN, Y. and LOH, J. M. (2007). A thinned block bootstrap procedure for modeling inhomogeneous spatial point patterns. *Journal of the American Statistical Association* **102** 1377-1386.
- [7] ILLIAN, J., PENTTINEN, A., STOYAN, H. and STOYAN, D. (2008). *Statistical analysis and modelling of spatial point patterns. Statistics in Practice*. Wiley, New York.
- [8] LAVANCIER, F., MØLLER, J. and RUBAK, E. (2015). Determinantal point process models and statistical inference. *Journal of Royal Statistical Society: Series B (Statistical Methodology)* **77** 853-877.
- [9] LAVANCIER, F. and ROCHET, P. (2016). A general procedure to combine estimators. *Computational Statistics & Data Analysis* **94** 175-192.
- [10] LAVANCIER, F. and ROCHET, P. (2017). A tutorial on estimator averaging in spatial point process models. *Journal de la SFDS* **158** 106-123.
- [11] MACCHI, O. (1975). The coincidence approach to stochastic point processes. *Advances in Applied Probability* **7** 83-122.
- [12] MØLLER, J. and WAAGEPETERSEN, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton.
- [13] MØLLER, J. and WAAGEPETERSEN, R. P. (2007). Modern statistics for spatial point processes. *Scandinavian Journal of Statistics* **34** 643-684.
- [14] POINAS, F., DELYON, B. and LAVANCIER, F. (2017). Mixing properties and central limit theorem for associated point processes. *to appear in Bernoulli (available at arXiv:1705.02276)*.
- [15] SCHOENBERG, F. P. (2005). Consistent parametric estimation of the intensity of a spatial-temporal point process. *Journal of Statistical Planning and Inference* **128** 79-93.
- [16] SHIRAI, T. and TAKAHASHI, Y. (2003). Random point fields associated with certain Fredholm determinants II: Fermion shifts and their ergodic and Gibbs properties. *Ann. Probab.* **31** 1533-1564.
- [17] SOSHIKOV, A. (2000). Determinantal random point fields. *Russian Mathematical Surveys* **55** 923-975.
- [18] PROKEŠOVÁ, M. and JENSEN, E. B. V. (2013). Asymptotic Palm likelihood theory

- for stationary point processes. *Annals of the Institute of Statistical Mathematics* **65** 387-412.
- [19] PROKEŠOVÁ, M., DVOŘÁK, J. and JENSEN, E. B. V. (2016). Two-step estimation procedures for inhomogeneous shot-noise Cox processes. *Annals of the Institute of Statistical Mathematics*.
 - [20] TANAKA, U., OGATA, Y. and STOYAN, D. (2008). Parameter estimation and model selection for Neyman-Scott point processes. *Biometrical Journal* **50** 43-57.
 - [21] R CORE TEAM (2017). R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria.
 - [22] WAAGEPETERSEN, R. (2007). An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics* **63** 252-258.
 - [23] WAAGEPETERSEN, R. and GUAN, Y. (2009). Two-step estimation for inhomogeneous spatial point processes. *Journal of the Royal Statistical Society, Series B* **71** 685-702.

Supplementary material for “Adaptive estimating function inference for non-stationary determinantal point processes”

FRÉDÉRIC LAVANCIER^{1,*} ARNAUD POINAS^{2,**} RASMUS WAAGEPETERSEN^{3,†}

¹*Laboratoire de Mathématiques Jean Leray – BP 92208 – 2, Rue de la Houssinière – F-44322 Nantes Cedex 03 – France. Inria, Centre Rennes Bretagne Atlantique, France.*

E-mail: [*frederic.lavancier@univ-nantes.fr](mailto:frederic.lavancier@univ-nantes.fr)

²*IRMAR – Campus de Beaulieu - Bat. 22/23 – 263 avenue du Général Leclerc – 35042 Rennes – France*

E-mail: **arnaud.poinas@univ-rennes1.fr

³*Department of Mathematical Sciences – Aalborg University - Fredrik Bajersvej 7G – DK-9220 Aalborg – Denmark*

E-mail: †rw@math.aau.dk

1. Supplementary tables for Section 4.1 of the main manuscript

For the simulation study carried out in Section 4.1 of the main manuscript, considering estimation for a DPP model with a Bessel-type kernel, we report in Figure S1 the boxplots representing the distribution of the estimators and in Table S1 the percentages of times each estimation method has converged in our simulation study. These percentages are similar for all estimation methods. Table S2 displays the root mean square errors of the estimators considered in Section 4.1 where, for comparison, we also include results for the adaptive estimator using $\varepsilon = 0.05$. Conclusions based on these tables are given in the main paper.

2. Two-step versus simultaneous

Referring to Section 4.2, Figure S2 shows how irregular the contrast function $e(\psi)$ for the simultaneous approach can be in comparison with the contrast function $e_2(\psi)$ for the two-step approach. The underlying point pattern is displayed on the left. This is

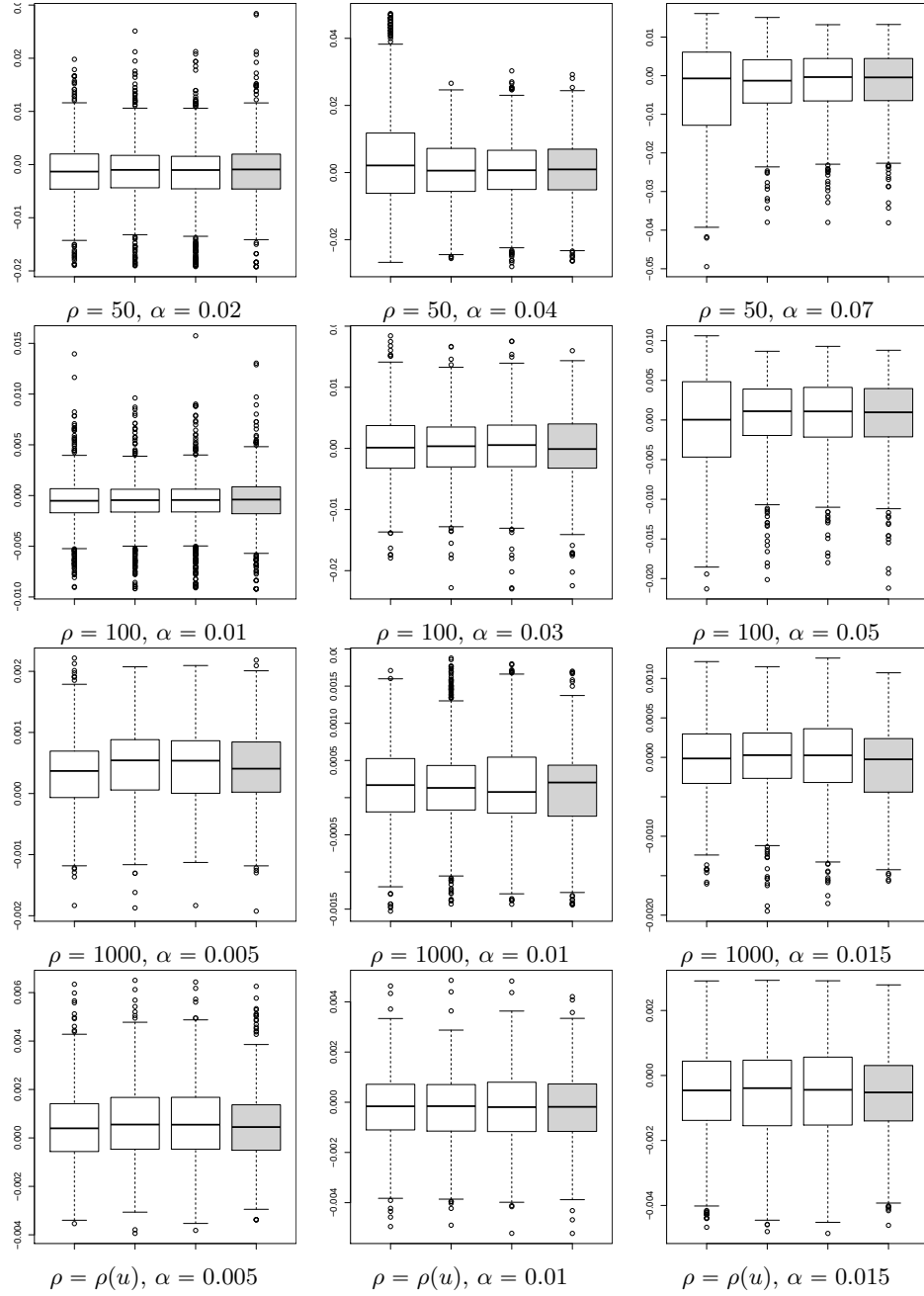


Figure S1. Distribution of $\hat{\alpha} - \alpha$ for a Bessel-type DPP on $[0, 1]^2$ for different values of ρ and α . In each subfigure, the 3 first estimators on the left use the test function (3) of the main manuscript with $R = 0.05, R = 0.1$ and $R = 0.25$ respectively, while the last estimator is the adaptive version based on (4).

ρ	α	$R = 0.05$	$R = 0.1$	$R = 0.25$	Adaptive
50	0.02	0.75	0.72	0.72	0.72
	0.04	0.97	0.85	0.80	0.85
	0.07	0.92	1.00	0.99	0.98
100	0.01	0.66	0.69	0.72	0.64
	0.03	0.99	0.96	0.89	0.93
	0.05	1.00	1.00	1.00	1.00
1000	0.005	1.00	0.95	0.95	0.95
	0.01	1.00	1.00	1.00	1.00
	0.015	1.00	1.00	1.00	1.00
Inhom	0.005	0.93	0.92	0.95	0.93
	0.01	0.98	0.98	0.98	0.96
	0.015	1.00	1.00	1.00	1.00

Table S1. Percentage of times the estimation methods have converged for the models and estimators considered in Section 4.1 of the main manuscript.

ρ	α	$R = 0.05$	$R = 0.1$	$R = 0.25$	$\varepsilon = 0.01$	$\varepsilon = 0.05$
50	0.02	5.49	5.45	5.95	5.53	7.13
	0.04	14.92	8.81	8.79	8.87	8.71
	0.07	13.08	8.10	8.07	8.04	8.82
100	0.01	2.30	2.27	2.45	2.49	2.77
	0.03	5.05	4.99	5.16	5.10	5.27
	0.05	5.75	4.40	4.47	4.50	5.10
1000	0.005	0.68	0.87	0.83	0.73	0.73
	0.01	0.57	0.59	0.61	0.56	0.59
	0.015	0.47	0.46	0.52	0.47	0.51
Inhom	0.005	1.58	1.65	1.66	1.61	1.57
	0.01	1.34	1.36	1.36	1.32	1.37
	0.015	1.43	1.47	1.48	1.40	1.46

Table S2. RMSE ($\times 10^3$) for the same simulations as in Table 1 of the main manuscript, with the addition of the adaptive estimator using $\varepsilon = 0.05$. These quantities are computed from 1000 simulations where all five estimation methods have converged (explaining the differences with Table 1 of the main manuscript).

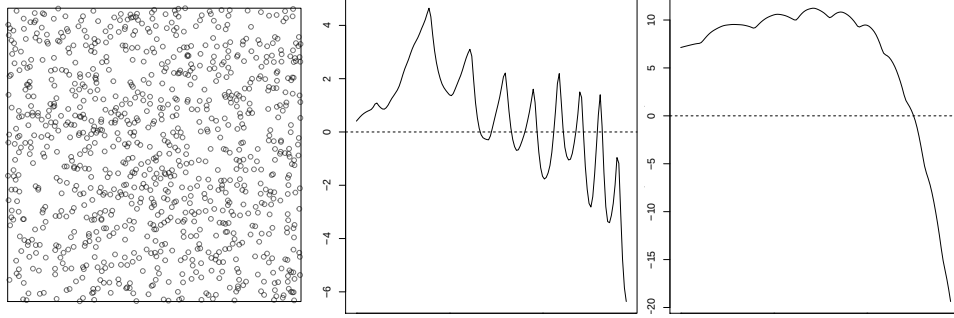


Figure S2. Realization of a DPP (left) and plots of the contrast function $e(\psi)$ (see (9) in the main manuscript) of the simultaneous procedure (middle) and of the contrast function $e_2(\psi)$ (see (8)) of the two-step procedure (right) obtained for the data in the left plot.

a realisation on the unit square of a homogeneous DPP with a Gaussian kernel, with intensity $\rho = 1000$ and range $\alpha = 0.01$.

Figure S3 reports the distributions of estimates of ρ over 1000 realisations on the unit square of a DPP with a Bessel-type kernel with $\rho = 1000$ and $\alpha = 0.01$. The two first estimators come from the simultaneous approach, see equation (10) of the main manuscript where $\hat{\psi} = \hat{\alpha}$ in this setting. For the first one, the numerical solution of $e(\alpha) = 0$ to get $\hat{\alpha}$ was initialized at the true value 0.01 of α . For the second one, $\hat{\alpha}$ was fixed to the true value, i.e. $\hat{\alpha} = 0.01$. The last estimator on the right of Figure S3 is simply $\hat{\rho} = N(X \cap W)/|W|$, corresponding to the first step of the two-step procedure. The respective root mean square errors are 33.6, 31.4 and 26. See the main manuscript for further discussion.

3. Some simulations for the Thomas model

The adaptive estimating function is also useful for clustered point processes. Here we consider a Thomas model on $[0, 1]^2$, with parent intensity $\kappa = 100$, offspring intensity $\mu = 10$ and various values of the dispersal kernel standard deviation σ . The same three estimation methods as in Section 4.1 of the main manuscript have been evaluated, where for the adaptive version both $\varepsilon = 0.01$ and $\varepsilon = 0.05$ have been considered. A point pattern sample and the distribution of the estimators of κ and σ based on 1000 replications are shown in Figure 3 for $\sigma = 0.02$, $\sigma = 0.035$ and $\sigma = 0.05$ respectively. Estimators of the library `spatstat` [1] of R [2] with default settings have also been added. These are: minimum contrast estimation based on the K -function, Guan's composite likelihood, and Palm likelihood, see also Section 2.3 in the main manuscript. Table S3 summarises the estimated root mean square errors for each estimation method.

Also for the Thomas process, the adaptive method, both with $\varepsilon = 0.01$ and $\varepsilon = 0.05$, performs well compared with the three fixed R estimators. In fact for $\sigma = 0.05$, the

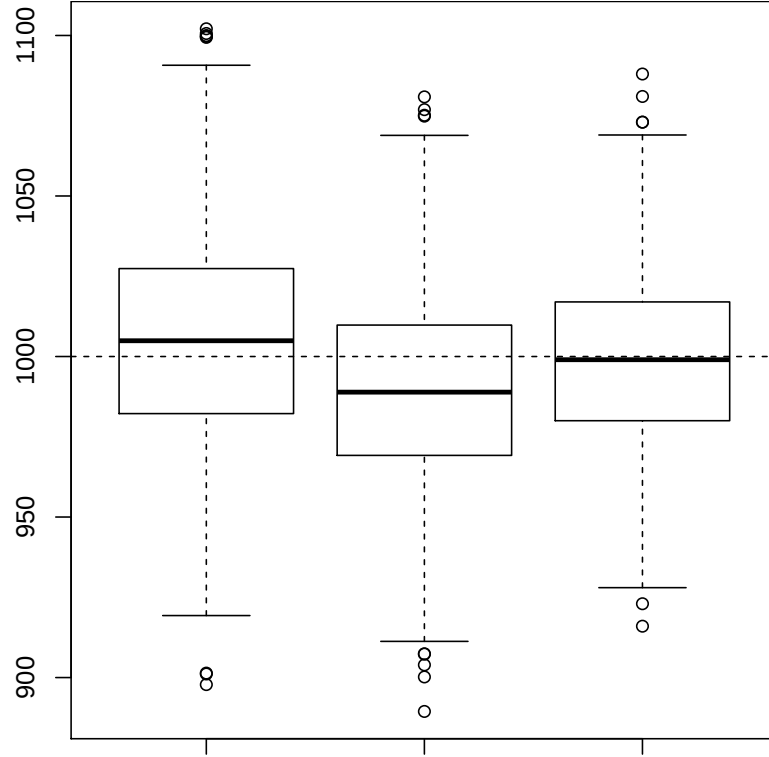


Figure S3. Distribution of estimates of ρ obtained from 1000 realisations of the Bessel-type DPP on $W = [0, 1]^2$ with $\rho = 1000$ and $\alpha = 0.01$. Left: the simultaneous estimator as given in equation (10) with initial value for the numerical solution given by the true value 0.01 of α . Middle: as left but using the true value of α instead of $\hat{\alpha}$. Right: $\hat{\rho} = N(X \cap W)/|W|$ corresponding to the first step of a two-step procedure.

σ		$R = 0.05$	$R = 0.1$	$R = 0.25$	$\varepsilon = 0.01$	$\varepsilon = 0.05$	K	clik	Palm
0.02	$\hat{\kappa}$	17 (0.40)	21 (0.48)	21 (0.47)	21 (0.48)	20 (0.47)	23 (0.54)	28 (0.70)	21 (0.49)
	$\hat{\sigma}$	1.04 (0.02)	1.84 (0.06)	1.94 (0.08)	1.79 (0.07)	1.51 (0.04)	2.60 (0.09)	1.54 (0.03)	1.92 (0.09)
0.035	$\hat{\kappa}$	35 (0.79)	31 (0.75)	40 (0.95)	38 (0.89)	35 (0.85)	33 (0.81)	121 (4.20)	35 (0.90)
	$\hat{\sigma}$	4.80 (0.09)	5.54 (0.14)	7.92 (0.32)	6.24 (0.21)	4.59 (0.10)	5.76 (0.14)	8.04 (0.07)	5.50 (0.12)
0.05	$\hat{\kappa}$	54 (1.24)	49 (0.92)	53 (2.05)	47 (1.74)	53 (1.89)	35 (1.55)	554 (13.64)	39 (1.02)
	$\hat{\sigma}$	18.30 (1.28)	36.47 (1.47)	12.17 (0.41)	11.22 (0.53)	8.94 (0.61)	8.12 (0.25)	23.69 (0.13)	19.47 (0.74)

Table S3. For the Thomas model, estimated root mean square errors of various estimators of κ and σ ($\times 10^3$). The 3 first estimators use the test function (3) of the main manuscript with $R = 0.05$, $R = 0.1$ and $R = 0.25$ respectively; the fourth and fifth estimators are the adaptive version based on (4) where $\varepsilon = 0.01$ and $\varepsilon = 0.05$; the three last estimators are from the library **spatstat**: based on K , on Guan’s composite likelihood (clik) and on Palm likelihood - all with default settings. The standard errors of the MSE estimations are given in parenthesis.

adaptive versions are better than any of the fixed R estimators. The adaptive method also has good stable performance compared with the three **spatstat** methods. In particular, the adaptive method performs much better than Guan’s composite likelihood with default settings.

References

- [1] BADDELEY, A. J., RUBAK, E. and TURNER, R. (2015). *Spatial Point Patterns: Methodology and Applications with R. Interdisciplinary Statistics*. Chapman & Hall/CRC, Boca Raton, Florida.
- [2] R CORE TEAM (2017). R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria.

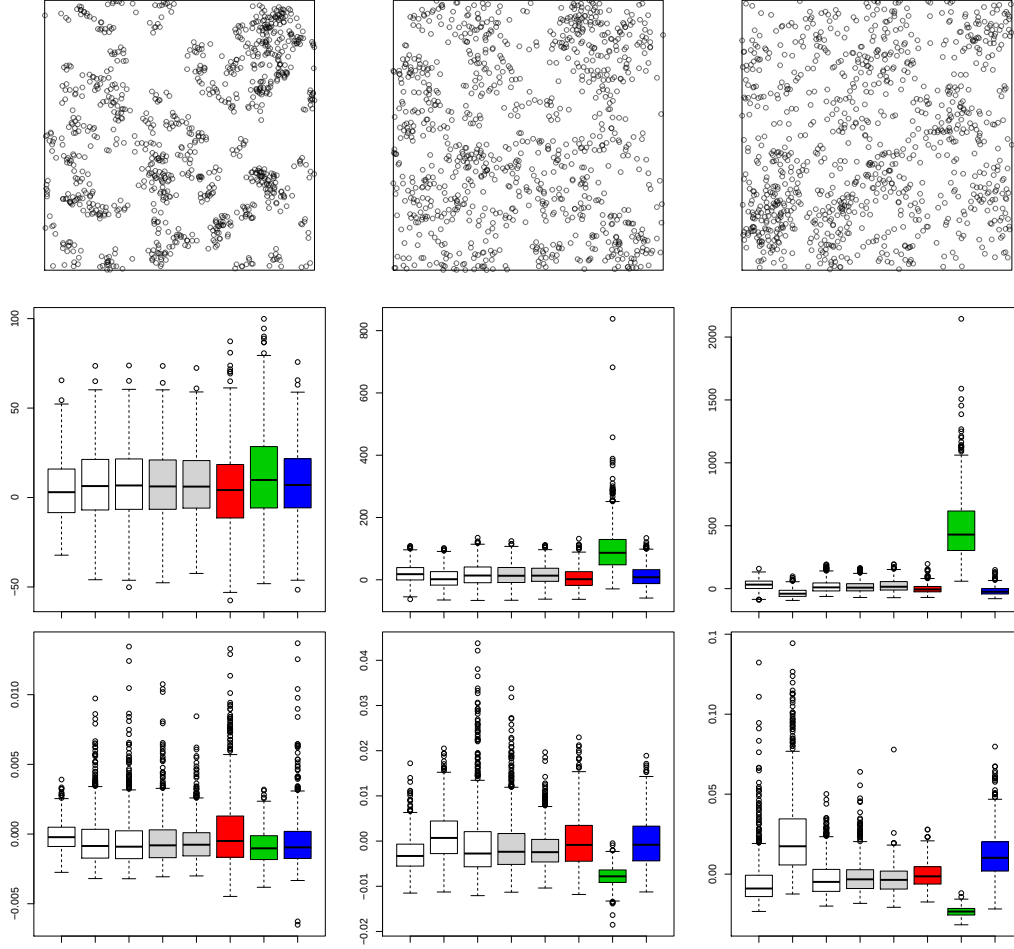


Figure S4. First row: Examples of point patterns simulated from a Thomas model on $[0, 1]^2$ for $\kappa = 100$, $\mu = 10$ and from left to right $\sigma = 0.02, 0.035, 0.05$. Second row: Distribution of estimates of κ based on 1000 replications. In each plot, the 3 first boxplots are for estimates obtained with the test function (3) of the main manuscript with $R = 0.05$, $R = 0.1$ and $R = 0.25$ respectively; the fourth and fifth boxplots (in grey) are for the adaptive version based on (4) where $\varepsilon = 0.01$ (left) and $\varepsilon = 0.05$ (right); the three last boxplots are for methods from `spatstat`: based on K (red), on Guan's composite likelihood (green) and on Palm likelihood (blue) - all with default settings. Third row: Distribution of estimates of σ based on 1000 replications, using the same estimation methods.