Adaptive estimating function inference for nonstationary determinantal point processes

Frédéric Lavancier\textsuperscript{1}\textsuperscript{*} | Arnaud Poinas\textsuperscript{2} | Rasmus Waagepetersen\textsuperscript{3}

\textsuperscript{1}Laboratoire de Mathématiques Jean Leray, University of Nantes
\textsuperscript{2}IRMAR, University of Rennes 1
\textsuperscript{3}Department of Mathematical Sciences, Aalborg University

Correspondence
Frédéric Lavancier, Jean Leray Mathematics Institute, University of Nantes, 2 rue de la Houssinière, 44322 Nantes Cedex 3, France.
Email: frederic.lavancier@univ-nantes.fr

Abstract
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 article, we establish asymptotic normality of estimating function estimators in a very general setting of nonstationary point processes. We then adapt this result to the case of nonstationary 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 a 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 and by an application to a real dataset.

Keywords
asymptotic normality, determinantal point processes, estimating functions, joint intensities, nonstationary, 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., Møller & Waagepetersen, 2004, 2007, 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 article, 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 (Baddeley, Rubak, & Turner, 2015; Illian, Penttinen, Stoyan, & Stoyan, 2008; Møller & Waagepetersen, 2004) as well as determinantal point processes (Lavancier, Møller, & Rubak, 2015; Macchi, 1975; Shirai & Takahashi, 2003; Soshnikov, 2000). 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.

Theoretical results have been established in a variety of special cases of first- and second-order estimating functions for Cox and cluster processes (Guan, 2006; Guan & Loh, 2007; Schoenberg, 2005; Waagepetersen, 2007; Waagepetersen & Guan, 2009) and for the closely related Palm likelihood estimators (Prokešová, Dvóřák, & Jensen, 2016; Prokešová & Jensen, 2013; Tanaka, Ogata, & Stoyan, 2008). The common general structure of the estimating functions on the other hand calls for a general theoretical setup, which is the first contribution of this article. Our setup also covers third- or higher order estimating functions and combinations of such estimating functions, providing a general unifying framework.

The literature on statistical inference for continuous determinantal point processes is quite limited. A Bayesian approach is considered in Affandi, Fox, Adams, and Taskar (2014), while likelihood and minimum contrast estimation methods are discussed in Lavancier et al. (2015). In fact, maximum likelihood is not feasible in general and only an approximated version of the likelihood is proposed in the stationary case in Lavancier et al. (2015), without theoretical guarantees. On the other hand, consistency and asymptotic normality of minimum contrast estimators based on the pair correlation function or the Ripley’s $K$ function have been established for determinantal point processes in Biscio and Lavancier (2017) but only in the stationary case. Based on the general setup, our second main contribution is to provide a detailed theoretical study of estimating function estimators for general nonstationary determinantal point processes.

Specializing to second-order estimating functions, a common approach (Guan, 2006; Tanaka et al., 2008) 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. We demonstrate its usefulness in simulation studies for determinantal point processes as well as an example of a cluster process. The practical
advantage of the adaptive choice is further illustrated by an application to a dataset of locations of Japanese pines.

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 \( X \cap B \) and \( |B| \) denote the Lebesgue measure of \( B \). 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 article, 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

\[
\mathbb{E} \sum_{u_1, \ldots, u_n \in X} \mathbb{I}_{u_1 \in B_1, \ldots, u_n \in B_n} = \int_{\mathbb{R}^d} \rho^{(n)}(u_1, \ldots, u_n) du_1 \cdots du_n, \tag{1}
\]

for Borel sets \( B_i \subseteq \mathbb{R}^d, i = 1, \ldots, n \), 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 likelihood of observing yet another point at \( u \). The Campbell formula

\[
\mathbb{E} \sum_{u_1, \ldots, u_n \in X} f(u_1, \ldots, u_n) = \int f(u_1, \ldots, u_n) \rho^{(n)}(u_1, \ldots, u_n) du_1 \cdots du_n
\]

follows immediately from the definition of \( \rho^{(n)} \) for any nonnegative function \( f : (\mathbb{R}^d)^n \to [0, \infty) \).

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 \( \Theta \) 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 bounded window \( W_n \subseteq \mathbb{R}^d \). We estimate the unknown parameter \( \theta^* \) by the solution \( \hat{\theta}_n \) of \( e_n(\theta) = 0 \) (or one of the solutions if there are many), where

\[
e_n(\theta) = \begin{bmatrix}
\sum_{u_1, \ldots, u_{q_1} \in \mathbb{R}_n \cap W_n} f_1(u_1, \ldots, u_{q_1}; \theta) - \int_{W_n} f_1(u; \theta) \rho^{(q_1)}(u; \theta) du \\
\vdots \\
\sum_{u_1, \ldots, u_{q_l} \in \mathbb{R}_n \cap W_n} f_l(u_1, \ldots, u_{q_l}; \theta) - \int_{W_n} f_l(u; \theta) \rho^{(q_l)}(u; \theta) du
\end{bmatrix}.
\]
for \( l \) given functions \( f_i : (\mathbb{R}^d)^n \times \Theta \rightarrow \mathbb{R}^k \) such that \( \sum_i k_i = p \).

A basic assumption for the following theorem 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 these conditions are detailed in Appendix A1 while the proof of the theorem is given in the Supplementary material.

**Theorem 1.** Under Assumptions (F1) through (F3) (or (F3’)), (X1) and (X2), with a probability tending to one as \( n \rightarrow \infty \), there exists a \( |W_n|^{1/2} \)-consistent sequence of roots \( \hat{\theta}_n \) of the estimating equations \( e_n(\theta) = 0 \). Precisely, for all \( \varepsilon > 0 \), there exists \( A > 0 \) such that

\[
\mathbb{P}(\exists \hat{\theta}_n : e_n(\hat{\theta}_n) = 0 \text{ and } |W_n|^{1/2} \| \hat{\theta}_n - \theta^* \| < A) > 1 - \varepsilon
\]

for a sufficiently large \( n \).

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.

**Remark 1.** While the parameter \( \theta^* \) is generally uniquely defined (in the sense that \( \theta \mapsto \mathbb{P}_\theta \) is injective) and verifies \( E(e_n(\theta^*)) = 0 \), the solution to \( e_n(\theta) = 0 \) may not be unique. The above theorem states that there exists a consistent and asymptotically Gaussian sequence of solutions, but unicity is not guaranteed. This drawback is unfortunately common in most asymptotic results for estimating functions inference (see the references in introduction, Sørensen, 1999, the handbook by Heyde, 1997 or the discussion in Van der Vaart, 2000, sect. 5.6). Nonetheless, it can be proved that the solution is unique for \( n \) sufficiently large whenever \( \lim_{n \rightarrow \infty} e_n(\theta)/|W_n| \) admits a unique zero, see Jacod and Sørensen (2018). But simple sufficient conditions to ensure the latter condition elude us.

### 2.3 Second-order estimating functions

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

\[
e_n(\theta) = \sum_{u,v \in X_n \setminus W_n} f(u, v; \theta) - \int_{W_n^2} f(u, v; \theta) \rho^{(2)}(u, v; \theta) du dv.
\]

In Guan (2006), the author 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 Equation (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, for example, the score functions of second-order composite likelihood (Guan, 2006; Waagepetersen, 2007) and Palm likelihood functions (Prokešová et al., 2016; Prokešová & Jensen, 2013; Tanaka et al., 2008) as well as score functions of minimum contrast object functions based on nonparametric estimates of summary statistics as the Ripley’s \( K \)
or the pair correlation function. For the second-order composite likelihood defined in Equation (4) in Guan (2006),

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

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 Waagepetersen (2007). The score of the Palm likelihood as generalized to the inhomogeneous case in Prokešová et al. (2016) is obtained with

\[ f(u, v; \theta) = \frac{\nabla_\theta \rho^{(2)}(u, v; \theta)}{\rho^{(2)}(u, v; \theta)} \left( \frac{\rho^{(2)}(u, w; \theta)}{\rho(u, \theta)} \right) \, dw. \]

In Prokešová et al. (2016), the authors also regarded the second-order composite likelihood proposed in Waagepetersen (2007) as a generalization of the stationary case Palm likelihood but the interpretation as a second-order composite likelihood given in Waagepetersen (2007) is more straightforward.

Considering a class of estimating functions of the form (2), a natural question is what is the optimal choice of \( f \)? A solution to this problem is provided in Deng, Guan, Waagepetersen, and Zhang (2017), 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 Lavancier and Rochet (2016, 2017).

### 2.4 Adaptive version

Consider second-order composite likelihood using Equation (3) but only \( R \) close pairs. The resulting weight function is then of the form

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

As mentioned in the previous section, using only \( R \) close pairs may be beneficial both for statistical efficiency and computational tractability. However, the possible improvement depends strongly on the chosen \( R \). Simulation studies such as in Prokešová et al. (2016) and Deng et al. (2017) 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, for example, be the standard deviation of the distribution for dispersal of offspring around parents.
For a determinantal point process (DPP), the parameter would typically be a correlation scale parameter in the kernel of the DPP, see Section 3. In practice, these parameters are not known and among the quantities that need to be estimated. In Guan (2006), it is suggested to choose an $R$ that minimizes a goodness of fit criterion for the fitted point process model while the choice of $R$ in Tanaka et al. (2008) and Waagepetersen and Guan (2009) is done by inspection of a nonparametric estimate of the pair correlation function (a similar approach is suggested by Heagerty & Lele, 1998 and Bai, Kang, & Song, 2014 in the context of pairwise composite likelihood for random fields). 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 $|g(u, v; \theta) - 1| \leq M(u, v; \theta)$ where

$$M(u, v; \theta) = \max_{s \in [u,v]} |g(s, \theta) - 1|.$$

Note that for DPPs, $M(u, v; \theta) = 1$ (see the next section) and for stationary point processes, $M(u, v; \theta)$ does not depend on $u$ and $v$. 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 Equation (4) by the constraint

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

for a small $\epsilon$. If, for example, $\epsilon = 1\%$ 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 $M(u, v; \theta)$. Note that this choice of pairs of points is adaptive in that it depends on $\theta$.

We then modify the function $f_R$ to be

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

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 $w(r) = \epsilon^{1/(r^2 - 1)}$ for $-1 \leq r \leq 1$, while $w(r) = 0$ otherwise. The user needs to specify a value of $\epsilon$ but in contrast to the original tuning parameter $R$, $\epsilon$ has an intuitive meaning independent of the underlying point process. We choose $\epsilon = 1\%$.

We emphasize that choosing $\epsilon = 1\%$ is not necessarily optimal. An optimal $\epsilon$ might be found by maximizing the Godambe information as a function of $\epsilon$ but this is not straightforward and the computational advantages of our approach would be lost. In fact, if Godambe optimality is key, we suggest to consider the previously mentioned approach by Deng et al. (2017) to identify an optimal second-order estimating function.

### 3 ASYMPTOTIC RESULTS FOR DPPS

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

$$\rho^{(n)}(u_1, \ldots, u_n) = \det[K](u_1, \ldots, u_n),$$
for all \( \{u_1, \ldots, u_n\} \subset \mathbb{R}^d \), where \([K](u_1, \ldots, u_n)\) 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 DPP with kernel \( K \) exists, it is unique. General conditions for existence are presented in Lavancier et al. (2015). In particular, if \( K \) admits the form

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

for a function \( C : \mathbb{R}^d \to \mathbb{R} \) with \( C(0) = 1 \), then a sufficient condition for existence of a DPP with kernel \( K \) is that \( \rho \) 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 \( \rho \) is the intensity of the DPP.

We now consider a parametric family of DPPs on \( \mathbb{R}^d \) with kernels \( K_{\theta} \) where \( \theta \in \Theta \) and \( \Theta \subseteq \mathbb{R}^p \) (see Lavancier et al., 2015; Biscio & Lavancier, 2016, for examples of such families). Henceforth, we assume that \( K_{\theta} \) is symmetric, continuous, and the DPP with kernel \( K_{\theta} \) exists for all \( \theta \in \Theta \). Note that in general, it is possible that two different kernels generate the same DPP distribution. This identifiability issue especially arises in the case of a discrete state space, where the distribution of a DPP is only identified up to flips of the signs of the rows and columns of its matrix kernel (see Engel & Schneider, 1980 or Rising, Kulesza, & Taskar, 2015). However, in the continuous case, corresponding to our framework, the kernel of a DPP is uniquely determined whenever the intensity function is positive, see Proposition S.3.1 and its corollary in the Supplementary material. Assuming a positive intensity function is not restrictive for statistical applications of DPPs.

An expression for the likelihood of a DPP on a bounded window is provided in Lavancier et al. (2015), where likelihood-based inference for stationary DPPs is discussed. 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, minimum contrast estimation is considered in Biscio and Lavancier (2016), based on the pair correlation function or Ripley’s \( K \)-function, but only for stationary DPPs. In the following, we consider general nonstationary DPPs and the estimator \( \hat{\theta}_n \) obtained by solving \( e_n(\theta) = 0 \), where \( e_n \) is given by (2). Note that the distribution for any classical parametric DPP model (showcased in Lavancier et al., 2015; Biscio & Lavancier, 2016) is uniquely determined by its first two order intensity functions, in the sense that \( \theta \mapsto (\rho(\cdot; \theta), \rho^{(2)}(\cdot; \theta)) \) is injective. This justifies the use of second-order estimating functions for DPPs. We establish in Section 3.1 using Theorem 1, the asymptotic properties of the estimator \( \hat{\theta}_n \), 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 \( \beta \) only appearing in the intensity function and a parameter \( \psi \) only appearing in the pair correlation function. Following Waagepetersen and Guan (2009), it is natural to consider a two-step estimation procedure where in a first step \( \beta \) is estimated by a Poisson likelihood score estimating function, which provides a consistent estimate of the intensity (Schoenberg (2005)), and in a second step the remaining parameter \( \psi \) is estimated by a second-order estimating function as in Equation (2), where \( \beta \) 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 1.
3.1 Second-order estimating functions for DPPs

In this part and in the rest of the document, we consider the following notation. For any set $W \subset \mathbb{R}^d$ and $r > 0$, we write $W \oplus r := \cup_{x \in W} B(x, r)$ and $W \ominus r := \{ x \in W, B(x, r) \subset W \}$ for the dilation and erosion of the set $W$, where $B(x, r)$ denotes the ball centered in $x$ with radius $r$.

We assume a realization of a DPP $X$ with kernel $K_\theta^*$, $\theta^* \in \text{Int}(\Theta)$, is observed on a bounded window $W_n \subset \mathbb{R}^d$. We estimate the unknown parameter $\theta^*$ 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 setup 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 A1. The condition (F1) in this case demands that $\theta \mapsto f(u, v; \theta)$ is twice continuously differentiable in a neighbourhood of $\theta^*$ and for $\theta$ in this neighbourhood, the derivatives are bounded with respect to $(u, v)$ uniformly in $\theta$. Moreover, from (F2), there exists $R > 0$ such that for all $\theta$ in a neighbourhood of $\theta^*$,

$$f(u, v; \theta) = 0 \quad \text{if} \quad \|u - v\| > R. \tag{7}$$

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

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

where we recall that in this setting

$$\rho^{(2)}(u, v; \theta) = K_\theta(u, u)K_\theta(v, v) - K_\theta(u, v)^2. \tag{8}$$

The assumptions (F3) and (F3') are technical and needed for the consistency 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 (5), we prove in Lemmas 1 and 2 that (F3) is generally satisfied even if $X$ is not stationary.

Finally, as shown in the proof of Theorem 2 below, the assumptions (X1) through (X3) in Theorem 1 are implied by the following:

\begin{itemize}
  \item[(D1)] $\theta \mapsto K_\theta(u, v)$ is twice continuously differentiable in a neighborhood of $\theta^*$, for all $u, v \in \mathbb{R}^d$. Moreover, the first and second derivative of $K_\theta$ with respect to $\theta$ are bounded with respect to $u, v \in \mathbb{R}^d$ uniformly in $\theta$ in a neighborhood of $\theta^*$.
  \item[(D2)] The kernel $K_\theta^*$ satisfies, for some $\varepsilon > 0$,
    $$\sup_{\|u - v\| > r} K_{\theta^*}(u, v) = o(r^{-(d+\varepsilon)/2}).$$
  \item[(D3)] $\lim \inf_n \lambda_{\min}(|W_n|^{-1} \Sigma_n) > 0$, where $\Sigma_n := \text{Var}(e_n(\theta^*))$ and $\lambda_{\min}(|W_n|^{-1} \Sigma_n)$ denotes the smallest eigenvalue of $|W_n|^{-1} \Sigma_n$.
  \item[(W)] $\exists \varepsilon > 0 \text{ s.t. } \partial W_n \ominus (R + \varepsilon) = o(|W_n|)$, where $\partial$ in this context denotes the boundary of a set, $R$ is defined in (7), and $|W_n| \to \infty$, as $n \to \infty$.
\end{itemize}
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\(|u-v|>2\), \(K_\theta(u, v) = O(r^{-(d+1)/2})\), including the most repulsive stationary DPP (see Biscio \& Lavancier, 2016; Lavancier et al., 2015). 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 is not bounded in any direction. It is, for instance, fulfilled if \(W_n\) is a Cartesian product of \(d\) intervals whose lengths tend to infinity.

**Theorem 2.** Under Assumptions (D1) and (D2), if assumptions (F1) through (F3) (or (F3')) are satisfied for \(f = f_t\), with a probability tending to one as \(n \to \infty\), there exists a \(|W_n|^{-1/2}\)-consistent sequence of roots \(\hat{\theta}_n\) of the estimating equations \(e_n(\theta) = 0\). If moreover (W) and (D3) holds true, then

\[ |W_n|^{-1/2} H_n(\theta^*)(\hat{\theta}_n - \theta^*) \overset{\mathcal{D}}{\to} \mathcal{N}(0, I_p). \]

**Proof.** We deduce from (8) that (D1) implies (X1). Moreover, it was shown in Poinas, Delyon, and Lavancier (2017) 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 (1) in the case \(l = 1\) and \(q_1 = 2\).

In the case of a stationary \(X\) and \(f\) given by (5), the following lemma shows that (F3) is satisfied under mild assumptions that are violated only in degenerate cases. For instance, if \(p = 1\), the last assumption boils down to \(\nabla_\theta \rho^{(2)}(0, t; \theta^*) \neq 0\) for some \(t \neq 0\) such that \(|K_\theta(t)| > \sqrt{\epsilon K_\theta(0)}\). In particular, it is not difficult to verify these assumptions for the stationary parametric kernels considered in our simulation study of Section 4, namely the Bessel-type and the Gaussian kernels, see the Supplementary material.

**Lemma 1.** Assume (W) and (D2), suppose \(X\) is stationary and let \(f\) be as in (5). Assume that \(w\) is positive on \([0, 1]\), vanishes on \([1, \infty]\) and is differentiable on \(\mathbb{R}_+\). If \(t \mapsto f(0, t; \theta^*)\) is integrable and \(\text{span}\{\nabla_\theta \rho^{(2)}(0, t; \theta^*) : |K_\theta(t)| > \sqrt{\epsilon K_\theta(0)}\} = \mathbb{R}^p\), then (F3) is satisfied.

**Proof.** By definition of \(w\) and (D2), there exists \(R > 0\) such that \(f(0, t; \theta^*) = 0\) when \(|t| \geq R\). By Lemma 4, since \(t \mapsto f(0, t; \theta^*)\) is integrable then \(H_n(\theta^*) = \int_{||t||<R} h(t)dt\) where the function \(h : \mathbb{R}^d \to \mathbb{R}^{p \times p}\) is defined by

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

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 \(|K_\theta(t)| > \sqrt{\epsilon K_\theta(0)}\), \(\phi \in \mathbb{R}^p\) and note that since \(w(\epsilon K_\theta(0)^2/K_\theta(t)^2) > 0\) for \(t \in A\) and \(h(t)\) is continuous and positive semidefinite,

\[ \phi^T H(\theta^*) \phi = 0 \iff \forall t \in A, \phi^T h(t) \phi = 0 \iff \forall t \in A, \nabla_\theta \rho^{(2)}(0, t; \theta^*)^T \phi = 0 \iff \phi \in \left(\text{span}\{\nabla_\theta \rho^{(2)}(0, t; \theta^*) : t \in A\}\right)'. \]

By assumption \(\text{span}\{\nabla_\theta \rho^{(2)}(0, t; \theta^*) : t \in A\} = \mathbb{R}^p\) whereby \(\phi = 0\), which concludes the proof.
Similarly, we can show that even in the nonstationary case, condition (F3) is satisfied for the function in (5) 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. In particular, a straightforward calculus carried out in the Supplementary material shows that the nonstationary Bessel-type kernel used in our simulation study satisfies these assumptions.

**Lemma 2.** Assume (W), (D2) and that $K_{\theta^*}$ is bounded. Let $f$ be as in (5) and define $h : (\mathbb{R}^d)^2 \to \mathbb{R}^{p \times p}$ by

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

Assume that $w$ is positive on $[0, 1)$, vanishes on $[1, \infty)$, and is differentiable on $\mathbb{R}_+$. If $\sup_{u \in \mathbb{R}^d} |\mathbb{E} h(u, v)dv| < +\infty$ and if there exists $\mu > 1$ and $\delta > 0$ such that for all $u \in \mathbb{R}^d$ and for all unit vectors $\phi$ of $\mathbb{R}^p$ there exists a subset $A$ of $\{v : K_{\theta^*}(u, v)^2 > \mu K_{\theta^*}(u, u)K_{\theta^*}(v, v)\}$ 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 $K_{\theta^*}$ is bounded, there exists $R > 0$ such that $h(u, v) = 0$ when $|v - u| \geq R$. The integral in (F3) writes

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

where

$$\epsilon_n = \frac{1}{|W_n|} \int_{W_n \ominus (W_n \ominus \mathcal{R})} \int_{W_n} h(u, v) \mathbb{I}_{|u - v| \leq R} dv du.$$

By (W), we have

$$\|\epsilon_n\| \leq \frac{|W_n \ominus W_n \ominus \mathcal{R}|}{|W_n|} \sup_{u \in \mathbb{R}^d} \int_{\mathbb{R}^d} \|h(u, v)\| dv \leq \frac{|\partial W_n \ominus \mathcal{R}|}{|W_n|} \sup_{u \in \mathbb{R}^d} \int_{\mathbb{R}^d} \|h(u, v)\| dv \to 0,$$

and for all $\phi$,

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

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

$$\forall v \in A, |\phi^T \nabla_\theta \rho^{(2)}(u, v; \theta^*)| > \delta \quad \text{and} \quad w \left( \frac{\varepsilon K_{\theta^*}(u, u)K_{\theta^*}(v, v)}{K_{\theta^*}(u, v)^2} \right) > \inf_{x \in [0, 1/\mu]} w(x).$$
Hence for $\|\phi\| = 1$,
\[
\frac{1}{|W_n|} \phi^T \left( \int_{W_n \Theta R} \int_{W_n} h(u, v) 1_{\| u - v \| \leq K} \, dudv \right) \phi \\
\geq \inf_{x \in [0, 1/\mu]} w(x) \frac{|W_n|}{\|\rho^{(2)}(\cdot; \theta^{*})\|_\infty} \int_{W_n \Theta R} \left( \int_A |\phi^T \nabla \rho^{(2)}(u, v; \theta^{*})|^2 \, dv \right) \, du \\
\geq |W_n \ominus R| |A| \delta^2 \inf_{x \in [0, 1/\mu]} w(x) \frac{|W_n|}{\|\rho^{(2)}(\cdot; \theta^{*})\|_\infty} \\
= \left( \frac{|W_n| - |W_n \cap (\partial W_n \ominus R)|}{|W_n|} \right) \frac{|A| \delta^2 \inf_{x \in [0, 1/\mu]} w(x)}{\|\rho^{(2)}(\cdot; \theta^{*})\|_\infty} \\
\to \frac{|A| \delta^2 \inf_{x \in [0, 1/\mu]} w(x)}{\|\rho^{(2)}(\cdot; \theta^{*})\|_\infty} > 0,
\]
where the limit is a consequence of (W). Since the limit does not depend on $\phi$, then (F3) is satisfied.

\[\blacksquare\]

### 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 $\theta := (\beta^T, \psi^T) \in \Theta \subset \mathbb{R}^{p+q}$ with $\beta \in \mathbb{R}^p$ and $\psi \in \mathbb{R}^q$, $\rho(\cdot; \beta)$ are nonnegative functions, and $C(\cdot, \cdot; \psi)$ are correlation functions, in particular $C(u, u; \psi) = 1$ for any $\psi$. Note that in this case, the DPP with kernel $K_\theta$ 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 $K_\theta^*$, $\theta^* \in \text{Int}(\Theta)$, is observed on a bounded window $W_n \subset \mathbb{R}^d$. In the spirit of Waagepetersen and Guan (2009), we estimate $\theta^*$ in two steps. First, $\beta^*$ is estimated as the solution $\hat{\beta}_n$ of $s_n(\beta) = 0$ where
\[
s_n(\beta) = \sum_{u \in \mathcal{X} \cap W_n} \frac{\nabla \rho(u; \beta)}{\rho(u; \beta)} - \int_{W_n} \frac{\nabla \rho(u; \beta)}{\rho(u; \beta)} \, du,
\]
is the score function for a Poisson point process. Then, we estimate $\psi^*$ by the solution $\hat{\psi}_n$ of $u_n(\hat{\beta}_n, \psi) = 0$ where
\[
u_n(\theta) = \sum_{u, v \in \mathcal{X} \cap W_n} f(u, v; \theta) - \int_{W_n^2} f(u, v; \theta) \rho^{(2)}(u, v; \theta) \, dudv,
\]
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, we 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 \( \hat{\theta}_n = (\hat{\beta}_n, \hat{\psi}_n)^T \) is obtained as the solution of \( e_n(\theta) = 0 \), where \( e_n(\theta) = (s_n(\beta), u_n(\beta, \psi))^T \). 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

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

where

\[
H_n^{1,1}(\beta) = \frac{1}{|W_n|} \int_{W_n} \frac{\nabla_\beta \rho(u; \beta) \nabla_\beta \rho(u; \beta)^T}{\rho(u; \beta)} du,
\]

\[
H_n^{2,1}(\beta, \psi) = \frac{1}{|W_n|} \int_{W_n^2} f(u, v; \beta, \psi) \nabla_\beta \rho^{(2)}(u, v; \beta, \psi)^T dudv,
\]

\[
H_n^{2,2}(\beta, \psi) = \frac{1}{|W_n|} \int_{W_n^2} f(u, v; \beta, \psi) \nabla_\psi \rho^{(2)}(u, v; \beta, \psi)^T dudv.
\]

Since it is a nonsymmetric matrix, condition (F3') is more applicable than (F3). Mild conditions ensuring (F3') in the stationary case are provided in Lemma 3.

**Theorem 3.** 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 \( n \to \infty \), there exists a \( |W_n|^{1/2} \)-consistent sequence of roots \( \hat{\theta}_n \) of the estimating equations \( e_n(\theta) = 0 \). 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 2. □

The next lemma is similar to Lemma 1. When \( q = 1 \), the last technical condition boils down to \( \nabla_\psi (1 - C^2(0, t; \psi^*)) \neq 0 \) for some \( t \) such that \( C(0, t; \psi^*) \geq \sqrt{\varepsilon} C(0, 0; \psi^*) \). In particular, the stationary kernels in Section 4 satisfy the required assumptions, see the Supplementary material.

**Lemma 3.** Assume that for all \( \theta \), \( K_\theta(u, v) \) 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 \( \hat{\theta}_n = N(X \cap W_n)/|W_n| \). In the second step, assume

\[
f(u, v; \beta, \psi) = w \left( \frac{\varepsilon}{1 - g(u, v; \psi)} \right) \frac{\nabla_\psi \rho^{(2)}(u, v; \beta, \psi)}{\rho^{(2)}(u, v; \beta, \psi)}
\]

\[
= w \left( \frac{\varepsilon}{C(0, v-u; \psi)^2} \right) \frac{\nabla_\psi (1 - C^2(0, v-u; \psi))}{1 - C^2(0, v-u; \psi)}.
\]

Assume that \( w \) is positive on \([0, 1]\), vanishes on \([1, \infty)\) and is differentiable on \( \mathbb{R}_+ \). If \( t \mapsto f(0, t; \theta^*) \) is integrable and span \( \{ \nabla_\psi (1 - C^2(0, t; \psi^*)) : C(0, t; \psi^*) > \sqrt{\varepsilon} \} = \mathbb{R}^q \), then (F3') is satisfied under (W), (D1), and (D2).
Proof. By definition of \( w \) and (D2), there exists \( R > 0 \) such that \( f(0, t; \theta^*) = 0 \) when \( \|t\| \geq R \). Since \( K_\theta \) and \( f \) are invariant by translation and \( t \mapsto f(0, t; \theta^*) \) is integrable then \( H_n(\theta) \) converges by Lemma 4. In particular, we have

\[
H_{n, 1}^{1, 1}(\beta) \to \frac{1}{\beta^d}, \\
H_{n, 2}^{2, 2}(\beta, \psi) \to \beta^2 \int_{\|\cdot\| \leq R} h(t; \psi) dt, \\
H_{n, 2}^{2, 1}(\beta, \psi) \to 2\beta \int_{\|\cdot\| \leq R} w \left( \frac{\varepsilon}{C(0, t; \psi)^2} \right) \nabla \psi(1 - C^2(0, t; \psi)) dt,
\]

where the function \( h(\cdot; \psi) : \mathbb{R}^d \to \mathbb{R}^{p \times p} \) is defined by

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

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 \( H_{n, 2}^{2, 1}(\theta^*) \) is invertible. This is done the same way as in Lemma 1. □

4 | SIMULATION STUDY

In this section, we use simulation studies to investigate the performance of our adaptive estimating function. In Section 5 of the Supplementary material, we additionally compare two-step estimation, when it is feasible with simultaneous estimation. Our recommendation is to use the two-step approach.

In order to assess the adaptive test function (5) against the truncated test function (4) 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, \( \rho \) is the intensity, and \( \alpha \) controls the range of interaction of the DPP. For existence, \( \rho \) and \( \alpha \) must satisfy

\[
\alpha^2 \|\rho\|_{\infty} \leq \frac{1}{\pi}.
\]  

(9)

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 Biscio and Lavancier (2016). It covers a large range of repulsiveness, from the Poisson point process (when \( \alpha \) 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 \( \rho \), \( \rho \in \{50, 100, 1000\} \), 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 \( \rho \), three values of \( \alpha \) are considered: a small one, a medium one, and a last one close to the maximal possible value satisfying (9). Examples of point patterns simulated on \([0, 1]^2\) are
displayed in Figure 1. All simulations are carried out using R (R Core Team, 2017), in particular the library spatstat (Baddeley et al., 2015).

We estimate \( \rho \) and \( \alpha \) by a two-step procedure as studied in Section 3.2 from realizations of the DPP on \( W = [0,1]^2 \). In the first step, the parameters arising in \( \rho \) 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 (4) where \( \theta = \alpha \) in this setting and when \( R \in \{0.05, 0.1, 0.25\} \), and based on the adaptive test function (5) with \( \epsilon = 0.01 \) and the weight function \( w \) given at the end of Section 2.4. This yields four different estimators of \( \alpha \). 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 S3 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 and Lemma 3 (for the homogeneous case), ensuring nice asymptotic properties of the estimators considered in this section. This is confirmed by the estimated RMSEs reported in Table 1, that decrease when the intensity \( \rho \) increases (which mimics the effect of an increasing window since rescaling the window by a factor \( 1/k \) is equivalent to change \( \rho \) into \( k^2 \rho \) and \( \alpha \) into \( \alpha/k \), see (2.4) in Lavancier et al., 2015). Moreover, these RMSEs show that the best choice of \( R \) in the test function (4) clearly depends on the range of interaction of the underlying process. This emphasizes the importance of a data-driven approach to choosing

\[
\begin{align*}
\text{Figure 1} & \quad \text{Examples of point patterns simulated from a Bessel-type determinantal point process on } [0,1]^2 \\
& \text{for different values of } \rho \text{ and } \alpha. \text{ For the last row, } \rho(x,y) = 20 \exp(4x)
\end{align*}
\]
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 $\rho$ and $\alpha$

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

Note: The three first estimators use the test function (4) with $R = 0.05$, $R = 0.1$, and $R = 0.25$, respectively, while the last estimator is the adaptive version based on (5). 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 $\rho$ and $\alpha$, these quantities are computed from 1000 simulations where all four estimation methods have converged.

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. Furthermore, 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 (4) 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 RMSEs 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 S3 are based on 1000 simulations where all four methods have converged.

5 | APPLICATION

To illustrate the practical importance of our adaptive estimating function and our asymptotic results, we consider the problem of fitting a DPP model to the point pattern data in the left plot of Figure 2. This dataset collected by Numata (1964) records the locations of 204 seedlings and saplings of Japanese black pines in an observation window of dimension 10 m by 10 m. It has previously been analysed in Ogata and Tanemura (1986) using an inhomogeneous Gibbs model and later in Lavancier et al. (2015) using an inhomogeneous DPP with kernel of the form (6) with a cubic polynomial in the Cartesian coordinates for the log-intensity and $C(u) = \exp(-||u||^2/\alpha^2)$. The estimation in Lavancier et al. (2015) was carried out using a two-step procedure where the intensity parameters were fitted in the first step by the Poisson likelihood method and in the second step $\alpha$ was estimated by minimisation of a contrast function based on the pair correlation function. This gave $\hat{\alpha} = 0.226$ and the fit was judged to be satisfying based on several goodness of fit envelope tests. However, this second step relies on the arbitrary choice of several tuning parameters similar to $R$ and no confidence intervals were provided. We also fit the same inhomogeneous DPP model by the two-step approach (detailed in Section 3.2) but using in the second step the test function (4) (for the nonadaptive approach) or (5) (for our adaptive version).

**Figure 2** Japanese pines dataset and kernel estimate of its pair correlation function [Colour figure can be viewed at wileyonlinelibrary.com]
For the nonadaptive approach, the default choice of $R$ provided by spatstat (Baddeley et al., 2015) is one quarter of the smallest side length of the observation window, which in the current case gives $R = 2.5$. This choice of $R$ is subject to criticism since it does not take into account the correlation properties of the data-generating point process. Alternatively one could, following the references mentioned in Section 2.4, choose an $R$ based on inspection of the kernel estimate of the pair correlation function shown in the right plot in Figure 2. This suggests using a value of $R$ around 0.4. However, this last approach completely eludes a theoretical underpinning. First, the asymptotic properties of the kernel estimator itself are complicated and second it is not possible to handle mathematically the visual assessment of $R$. We use instead our adaptive procedure and obtain $\hat{\alpha} = 0.241$ for the range parameter and the adaptively chosen value of $\hat{R} = 0.37$, which is in agreement with the visual inspection of the pair correlation function.

Based on our Theorem 3 in Section 3.2, we further in a standard way obtain a 95% approximate confidence interval for $\alpha$ (estimate $\pm 1.96$ times asymptotic standard deviation). For ease of implementation, we use a parametric bootstrap to estimate the asymptotic standard deviation (alternatively one could use numerical integration to compute the asymptotic covariance matrix). More precisely, we generate 1000 realisations of the fitted model and refit the model for each simulation. The empirical standard deviation of the resulting simulated estimates is then an estimate of the standard deviation of $\hat{\alpha}$. Note that our consistency result in Theorem 3 is a requirement for the validity of this approach (see for instance Beran, 1997). We obtain the specific estimate 0.026 and hence the approximate 95% confidence interval [0.19; 0.29] for $\alpha$. This result confirms that there is significant inhibition between the Japanese pines and, in particular, provides strong evidence against the inhomogeneous Poisson process model ($\alpha = 0$).

Note that a similar bootstrap approach is not possible in the nonadaptive case, where $R$ is chosen by visual inspection, since the variability of this choice cannot be included in an automatic procedure.

6 | DISCUSSION

In this article, 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 $R$ that controls which pairs of points are used in the estimation. Although not statistically optimal, our adaptive procedure for selecting this tuning parameter is intuitively appealing and easy to implement. The method depends on a new tuning parameter $\epsilon$ for which it is easier to identify reasonable values than for the original tuning parameter $R$. The resulting estimation procedure is computationally tractable and performs well in terms of mean squared error 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 article 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 7 of the Supplementary material, a simulation study in case of a cluster process.

ACKNOWLEDGEMENTS

Rasmus Waagepetersen was 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.

**ORCID**

Frédéric Lavancier https://orcid.org/0000-0001-7138-6658

**REFERENCES**

Affandi, R. H., Fox, E., Adams, R., & Taskar, B. (2014). *Learning the parameters of determinantal point process kernels*. Paper presented at the International Conference on Machine Learning, 1224–1232.

Baddeley, A. J., Rubak, E., & Turner, R. (2015). *Spatial point patterns: Methodology and applications with R*. Interdisciplinary statistics. Boca Raton, FL: Chapman & Hall.

Bai, Y., Kang, J., & Song, P. X.-K. (2014). Efficient pairwise composite likelihood estimation for spatial-clustered data. *Biometrics*, 70(3), 661–670.

Beran, R. (1997). Diagnosing bootstrap success. *Annals of the Institute of Statistical Mathematics, 49*(1), 1–24.

Biscio, C., & Lavancier, F. (2016). Quantifying repulsiveness of determinantal point processes. *Bernoulli*, 22, 2001–2028.

Biscio, C., & Lavancier, F. (2017). Contrast estimation for parametric stationary determinantal point processes. *Scandinavian Journal of Statistics, 44*, 204–229.

Deng, C., Guan, Y., Waagepetersen, R., & Zhang, J. (2017). Second-order quasi-likelihood for spatial point processes. *Biometrics, 73*(4), 1311–1320.

Engel, G. M., & Schneider, H. (1980). Matrices diagonally similar to a symmetric matrix. *Linear Algebra and its Applications, 29*, 131–138.

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

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

Heagerty, P. J., & Lele, S. R. (1998). A composite likelihood approach to binary spatial data. *Journal of the American Statistical Association, 93*(443), 1099–1111.

Heyde, C. C. (1997). *Quasi-likelihood and its application: A general approach to optimal parameter estimation*. Springer-Verlag, New-York: Springer Science & Business Media.

Illian, J., Penttinen, A., Stoyan, H., & Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*. New York, NY: Wiley.

Jacod, J., & Sørensen, M. (2018). A review of asymptotic theory of estimating functions. *Statistical Inference for Stochastic Processes, 21*, 415–434.

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

Lavancier, F., & Rochet, P. (2016). A general procedure to combine estimators. *Computational Statistics & Data Analysis, 94*, 175–192.

Lavancier, F., & Rochet, P. (2017). A tutorial on estimator averaging in spatial point process models. *Journal de la SFDS, 158*(3), 106–123.

Macchi, O. (1975). The coincidence approach to stochastic point processes. *Advances in Applied Probability, 7*, 83–122.

Møller, J., & Waagepetersen, R. P. (2004). *Statistical inference and simulation for spatial point processes*. Boca Raton, FL: Chapman and Hall.

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

Numata, M. (1964). Forest vegetation, particularly pine stands in the vicinity of Choshi-flora and vegetation in Choshi, Chiba prefecture VI. *Bulletin of the Choshi Marine Laboratory, 6*, 27–37.

Ogata, Y., & Tanemura, M. (1986). *Likelihood estimation of interaction potentials and external fields of inhomogeneous spatial point patterns*. In I. Francis, B. Manly, & F. Lam (Eds.), *Pacific statistical congress* (pp. 150–154). Amsterdam, The Netherlands: Elsevier.
Poinas, F., Delyon, B., & Lavancier, F. (2017). Mixing properties and central limit theorem for associated point processes. *Bernoulli* arXiv:1705.02276. 25(3), 1724–1754.

Prokešová, M., Dvořák, J., & Jensen, E. (2016). Two-step estimation procedures for inhomogeneous shot-noise Cox processes. *Annals of the Institute of Statistical Mathematics*. 69(3), 513–542.

Prokešová, M., & Jensen, E. B. V. (2013). Asymptotic palm likelihood theory for stationary point processes. *Annals of the Institute of Statistical Mathematics*, 65(2), 387–412.

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rising, J., Kulesza, A., & Taskar, B. (2015). An efficient algorithm for the symmetric principal minor assignment problem. *Linear Algebra and its Applications*, 473, 126–144.

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.

Shirai, T., & Takahashi, Y. (2003). Random point fields associated with certain Fredholm determinants II: Fermion shifts and their ergodic and Gibbs properties. *Annals of Probability*, 31(3), 1533–1564.

Sørensen, M. (1999). On asymptotics of estimating functions. *Brazilian Journal of Probability and Statistics*, 13(2), 111–136.

Soshnikov, A. (2000). Determinantal random point fields. *Russian Mathematical Surveys*, 55, 923–975.

Tanaka, U., Ogata, Y., & Stoyan, D. (2008). Parameter estimation and model selection for Neyman-Scott point processes. *Biometrical Journal*, 50(1), 43–57.

van der Vaart, A. W. (2000). *Asymptotic statistics* (Vol. 3). Cambridge, England: Cambridge University Press.

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

Waagepetersen, R., & Guan, Y. (2009). Two-step estimation for inhomogeneous spatial point processes. *Journal of the Royal Statistical Society, Series B*, 71, 685–702.

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section at the end of this article.

---

**How to cite this article:** Lavancier F, Poinas A, Waagepetersen R. Adaptive estimating function inference for nonstationary determinantal point processes. *Scand J Statist*. 2021;48:87–107. https://doi.org/10.1111/sjos.12440

**APPENDIX A**

Our general Theorem 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 $i = 1, \ldots, l$ and for all $x \in (\mathbb{R}^d)^{\theta_1}, \theta \mapsto f_i(x; \theta)$ is twice continuously differentiable in a neighbourhood of $\theta^*$. Moreover, the first and second derivative of $f_i$ with respect to $\theta$ are bounded with respect to $x \in (\mathbb{R}^d)^{\theta_1}$ uniformly in $\theta$ belonging to this neighbourhood.

(F2) There exists a constant $R > 0$ such that for all $\theta$ in a neighbourhood of $\theta^*$, all functions $x \mapsto f_i(x; \theta)$ vanish when $\text{diam}(x) > R$. 
Define the matrices $H_n(\theta)$ by

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

where for all $i$

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

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

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

(F3') There exists a neighbourhood of $\theta^*$ such that for all $n$ high enough and all $\theta$ in this neighbourhood, $H_n(\theta)$ is invertible and $\|H_n(\theta)^{-1}\|$ is uniformly bounded with respect to $n$ and $\theta$, where $\| \cdot \|$ stands for any matrix norm.

(X1) For all $\theta$ in a neighbourhood of $\theta^*$ and all $q_i$, $i = 1, \ldots, l$, 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 $\theta^*$, for all $x \in (\mathbb{R}^d)^q$. Finally, the first and second derivative of $\rho^{(q_i)}$ with respect to $\theta$ are bounded with respect to $x \in (\mathbb{R}^d)^q$ uniformly in $\theta$ belonging to this neighbourhood.

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

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

with $R$ coming from (F2).

(X3) $X$ satisfies the central limit theorem

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

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
LAVANCIER et al.

process (see Waagepetersen & Guan, 2009 for Cox processes and Poinas et al., 2017 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 $\lim \inf_n \lambda_{\min}(H_n(\theta^*)) > 0$, 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 nonsymmetrical matrices. Furthermore, if $X$ is stationary, all $f_i$s are invariant by translation, and the sequence of windows $\{W_n\}_{n \geq 1}$ satisfies (W) in Section 3.1, then $H_n(\theta)$ converges towards a matrix $H(\theta)$ explicitly given in Lemma 4 below. Assumption (F3), thus simply, becomes $\inf_{\|\phi\|=1} \phi^T H(\theta^*) \phi > 0$ and (F3’) is satisfied whenever $H(\theta^*)$ is invertible by continuity of $H(\theta)$. In specific applications of Theorem 1, further conditions on the sequence of observation windows $\{W_n\}_{n \geq 1}$ may be required, see, for example, (W) in Section 3.1.

**Lemma 4.** 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, that is, $f_i(u_1, u; \theta) = f_i(0, u - u_1; \theta)$ where $u$ is the vector $(u_2, \cdots, u_{q_i})$ and $u - u_1 = (u_2 - u_1, \cdots, u_{q_i} - u_1)$. 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

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

The proof of this lemma is available in the Supplementary material.