Standard and robust intensity parameter estimation for stationary determinantal point processes

Christophe A.N. Biscio[1] and Jean-François Coeurjolly[2]

[1]Department of Mathematical Sciences, Aalborg University, Denmark, christophe@math.aau.dk.
[2]Univ. Grenoble Alpes, France, Jean-Francois.Coeurjolly@upmf-grenoble.fr.

Abstract

This work is concerned with the estimation of the intensity parameter of a stationary determinantal point process. We consider the standard estimator, corresponding to the number of observed points per unit volume and a recently introduced median-based estimator more robust to outliers. The consistency and asymptotic normality of estimators are obtained under mild assumptions on the determinantal point process. We illustrate the efficiency of the procedures in a simulation study.

Keywords: Repulsive point processes; Statistical inference; Robust statistics; Sample quantiles; Brillinger mixing.

1 Introduction

Spatial point patterns are datasets containing the random locations of some event of interest which arise in many scientific fields such as biology, epidemiology, seismology and hydrology. Spatial point processes are the stochastic models generating such data. We refer to Stoyan et al. (1995), Illian et al. (2008) or Møller and Waagepetersen (2004) for an overview on spatial point processes. The Poisson point process is the reference process to model random locations of points without interaction. Many alternative models such as Cox point processes (including Neymann-Scott processes, shot noise Cox processes, log-Gaussian Cox processes) or Gibbs point processes allow us to introduce clustering effects or to produce regular patterns (see again e.g. Møller and Waagepetersen (2004) or Illian et al. (2008)). First introduced by Macchi (1975), the interesting class of determinantal point processes has been revisited recently by Lavancier et al. (2015) in a statistical context. Such processes are in particular designed to model repulsive point patterns.

In this paper, we focus on stationary point processes, that is on point processes with distribution invariant by translation, and on first order characteristics for such processes, that is on the intensity parameter denoted by λ. The nonnegative real

parameter λ measures the mean number of points per unit volume and is needed for the estimation of second-order characteristics of point processes such as the pair correlation function or the Ripley's K-function, see for instance Møller and Waagepetersen (2004). Thus, the estimation of λ has been the subject of a large literature (see e.g. Illian et al. (2008)). Asymptotic properties for estimators of λ are non trivial and may be particularly challenging to obtain for some models such as the class of Gibbs point processes.

In this paper, we investigate the theoretical and practical properties of two different estimators of λ for the class of stationary determinantal point processes. The first estimator is the standard one, corresponding to the number of observed points divided by the volume of the observation domain. The second one is a median-based estimator recently proposed by Coeurjolly (2016) to handle outliers such as extra points or missing points. The form of these two estimators is not novel and follows the aforementioned references. Asymptotic properties for these two estimators have been established under general conditions on the underlying point process. However, these conditions have been checked mainly for Cox processes. We propose two contributions. First, we provide conditions on the kernel C, defining a determinantal point process (see Section 2 for details), which ensure that the standard and the median-based estimators are consistent and satisfy a central limit theorem. Second, we investigate the finite-sample size properties of the proposed procedures through a simulation study, where, in particular, we evaluate the ability of the estimators to be robust to outliers.

The rest of the paper is organized as follows. A short background on stationary determinantal point processes is presented in Section 2. Section 3.1 focuses on the standard estimator and details asymptotic properties for this estimator as well as an estimator of its asymptotic variance. Section 3.2 deals with the median-based estimator. Finally, we conduct a simulation study in Section 4 to compare these estimators in different scenarios. Proofs of the main results are postponed to Appendix.

2 Stationary Determinantal point processes

2.1 Background and definition

For $d \geq 1$, let $\mathbf{X}$ be a spatial point process defined on $\mathbb{R}^d$, which we see as a random locally finite subset of $\mathbb{R}^d$. Let $\mathcal{B}(\mathbb{R}^d)$ denote the class of bounded Borel sets in $\mathbb{R}^d$. For $u = (u^1, \ldots, u^d)^\top \in \mathbb{R}^d$ and for $A, B \in \mathcal{B}(\mathbb{R}^d)$, we denote by $|u| = \max_{i=1,\ldots,d} |u^i|$ and by $d(A, B)$ the minimal distance between A and B.

For any $W \in \mathcal{B}(\mathbb{R}^d)$, we denote by $|W|$ its Lebesgue measure, by $N(\mathbf{X} \cap W)$ the number of points in $\mathbf{X} \cap W$ and a realization of $\mathbf{X} \cap W$ is of the form $\mathbf{x} = \{x_1, \ldots, x_m\} \subset W$ for some nonnegative finite integer m. We consider simple point processes which means that two points of the process never coincide almost surely. For further details about point processes, we refer to Daley and Vere-Jones (2003, 2008) and Møller and Waagepetersen (2004).

The factorial moment measures are quantities of special interest for point processes. For any integer $l \geq 1$, $\mathbf{X}$ is said to have an l-th order factorial moment measure $\alpha^{(l)}$

if for all non-negative measurable functions h defined on $\mathbb{R}^{dl}$,

$$\mathrm{E}\sum_{u_1,\ldots,u_l\in\mathbf{X}}^{\neq} h(u_1,\ldots,u_l) = \int_{\mathbb{R}^{dl}} h(u_1,\ldots,u_l)\,\alpha^{(l)}(\mathrm{d}u_1\times\cdots\times\mathrm{d}u_l) \tag{1}$$

where the sign $\neq$ over the summation means that $u_1,\ldots,u_l$ are pairwise distinct. If $\alpha^{(l)}$ admits a density with respect to the Lebesgue measure on $\mathbb{R}^{dl}$, this density is called the l-th order product density of $\mathbf{X}$ and is denoted by ρ_l. Note that $\rho_1=\lambda$ and that for the homogeneous Poisson point process $\rho_l(u_1,\ldots,u_l)=\lambda^l$. We assume from now on, that λ is a positive real number.

The rest of this section is devoted to stationary determinantal point processes on $\mathbb{R}^d$. We refer to Hough et al. (2009) or Lavancier et al. (2015) for a review on non-stationary determinantal point processes on $\mathbb{C}^d$.

Definition 1. *Let $C:\mathbb{R}^d\to\mathbb{R}$ be a function. A point process $\mathbf{X}$ on $\mathbb{R}^d$ is a stationary determinantal point process (DPP) with kernel C and we denote for short $\mathbf{X}\sim DPP(C)$, if for all $l\geq 1$ its l-th order product density satisfies the relation*

$$\rho_l(x_1,\ldots x_l)=\det[C](x_1,\ldots,x_l)$$

for every $(x_1,\ldots,x_l)\in\mathbb{R}^{dl}$, where $[C](x_1,\ldots,x_l)$ denotes the $l\times l$ matrix with entries $C(x_i-x_j)$, $1\leq i,j\leq l$.

Conditions on the kernel C are required to ensure the existence of $DPP(C)$. To introduce the result, let S be a compact set of $\mathbb{R}^d$ and consider the function from $S\times S$, $(x,y)\mapsto C(x-y)$. By the Mercer theorem (see Riesz and Nagy (1990)), if we assume that C is continuous, the following series expansion holds

$$C(x-y)=\sum_{k=1}^{\infty}\beta_k^S\phi_k(x)\overline{\phi_k(y)} \tag{2}$$

where $\{\phi_k\}_{k\in\mathbb{N}}$ is an orthonormal basis of $L^2(S)$ and where β_k^S, $k\geq 1$, are real numbers. Finally, let $\mathcal{F}(h)$ denote the Fourier transform for a function $h\in L^1(\mathbb{R}^d)$ defined for all $t\in\mathbb{R}^d$ by

$$\mathcal{F}(h)(t)=\int_{\mathbb{R}^d}h(x)e^{-2i\pi x\cdot t}\mathrm{d}x,$$

a definition which can be extended to $L^2(\mathbb{R}^d)$ by Plancherel's theorem (see Stein and Weiss (1971)). The following result gives a sufficient condition to ensure the existence.

Proposition 2 (Lavancier et al. (2015, Proposition 1), Hough et al. (2009))**.** *Assume C is a symmetric continuous real-valued function in $L^2(\mathbb{R}^d)$. Then $DPP(C)$ exists if and only if one of the two statements is satisfied:*
(i) For all compact $S\subset\mathbb{R}^d$ and $k\geq 1$, $0\leq\beta_k^S\leq 1$.
(ii) $0\leq\mathcal{F}(C)\leq 1$.

For most of the kernels, Proposition 2 (ii) provides a more useful way of characterizing existence. To rephrase this condition, any real-valued continuous covariance function C in $L^2(\mathbb{R}^d)$ with $\mathcal{F}(C) \leq 1$ defines a stationary DPP. Some of the results presented hereafter (Propositions 4 and 5) require a slightly more restrictive condition, namely $\mathcal{F}(C) < 1$. To sum up, we consider the following assumption denoted by $\mathcal{A}_C$.

Assumption $\mathcal{A}_C$: C is a symmetric and continuous function, $C \in L^2(\mathbb{R}^d)$, $C(0) = \lambda$ and $0 \leq \mathcal{F}(C) < 1$.

2.2 Mixing-type properties

We continue this section by discussing Brillinger mixing and R-dependence type properties for stationary DPPs. To introduce the first one, we assume that the factorial moment measure exists until a certain $l \geq 1$. Then, the l-th order factorial cumulant moment measure $\gamma_{[l]}$ (see Daley and Vere-Jones (2003, 2008)) is defined for any $A_1, \ldots, A_l$ in $\mathcal{B}(\mathbb{R}^d)$ by

$$\gamma_{[l]}\left(\prod_{i=1}^{l} A_i\right) = \sum_{j=1}^{l} (-1)^{j-1}(j-1)! \sum_{B_1,\ldots,B_j \in \mathcal{P}_j^l} \prod_{i=1}^{j} \alpha^{(|B_i|)}\left(\prod_{l_i \in B_i} A_{l_i}\right),$$

where for all $j \leq l$, $\mathcal{P}_j^l$ denotes the set of all partitions of $\{1, \ldots, l\}$ into j non empty sets $B_1, \ldots, B_j$. For stationary point processes, we define, for $l \geq 2$, the so-called reduced version of the factorial cumulant moment measure $\gamma_{[l]}^{red}$ by

$$\gamma_{[l]}\left(\prod_{i=1}^{l} A_i\right) = \int_{A_l} \gamma_{[l]}^{red}\left(\prod_{i=1}^{l-1}(A_i - x)\right) \mathrm{d}x$$

where $A_1, \ldots, A_l \in \mathcal{B}(\mathbb{R}^d)$ and for $i = 1, \ldots, l-1$, $A_i - x$ stands for the translation of A_i by x. By Hahn-Jordan decomposition (see Dudley (2002, Theorem 5.6.1)), we may write $\gamma_{[l]}^{red} = \gamma_{[l]}^{+red} - \gamma_{[l]}^{-red}$ where $\gamma_{[l]}^{+red}$ and $\gamma_{[l]}^{-red}$ are two measures. The total variation measure of $\gamma_{[l]}^{red}$ is then defined as $|\gamma_{[l]}^{red}| = \gamma_{[l]}^{+red} + \gamma_{[l]}^{-red}$. Then, a point process is said to be Brillinger mixing, if for $l \geq 2$,

$$\left|\gamma_{[l]}^{red}\right|\left(\mathbb{R}^{d(l-1)}\right) < +\infty.$$

Brillinger mixing is adapted to DPPs as shown by the following result.

Theorem 3 (Biscio and Lavancier (2016b))**.** *Any DPP with kernel C verifying $\mathcal{A}_C$ is Brillinger mixing.*

For $R > 0$, R-dependence is simpler to define. A point process $\mathbf{X}$ is R-dependent if for all $A, B \in \mathcal{B}(\mathbb{R}^d)$ verifying $d(A, B) > R$, $\mathbf{X} \cap A$ and $\mathbf{X} \cap B$ are independent. This criterion is satisfied by the large subclass of DPPs with compactly supported kernel C.

Proposition 4. *Let $\mathbf{X} \sim DPP(C)$ be a DPP with kernel C verifying $\mathcal{A}_C$ and such that $C(x) = 0$ for $|x| > R$ for some $R > 0$. Then, $\mathbf{X}$ is R-dependent.*

Proof. Let A and B be two compact subsets in $\mathbb{R}^d$ such that $d(A,B) > R$. We first need to remind briefly the definition of the density for a DPP. More details can be found in Macchi (1975) and Lavancier et al. (2015), particularly in its supplementary materials. By assumption $\mathcal{A}_C$, $\mathbf{X} \cap S \sim DPP(C) \cap S$ is absolutely continuous with respect to the homogeneous Poisson process on S with unit intensity and has density

$$f_S(\{x_1,\dots,x_n\}) = e^{|S|} P(N(\mathbf{X}\cap S)=0)\ \det[\tilde{C}](x_1,\dots,x_n), \quad x_1,\dots,x_n \in S \quad (3)$$

where $\tilde{C}$ is defined by

$$\tilde{C}(x-y) = \sum_{k=1}^{\infty} C_k^S(x-y) \quad (4)$$

with $C_1^S(x-y) = C(x-y)$ and $C_k^S(x-y) = \int_S C_{k-1}^S(x-z)C(z-y)\mathrm{d}z$ for $k>1$, see Appendix G in the supplementary materials of Lavancier et al. (2015). By induction, if C is compactly supported, so is $\tilde{C}$. Now, let $x_1,\dots,x_p \in A$ and $y_1,\dots,y_q \in B$, for $p,q \geq 1$. Since $\tilde{C}(x) = 0$ for $|x| > R$, $[\tilde{C}](x_1,\dots,x_p,y_1,\dots,y_q)$ is a block diagonal matrix. Then, by applying (3) with $S = A \cup B$, it is straightforwardly seen that

$$f_S(\{x_1,\dots,x_p,y_1,\dots,y_q\}) \propto f_A(\{x_1,\dots,x_p\}) f_B(\{y_1,\dots,y_q\})$$

where the normalizing constant is determined by the condition $\int f_S = 1$, whereby we deduce the result. □

For some statistical applications, another type of mixing coefficient often used is the α-mixing coefficient defined as follows for spatial point processes (see e.g. Politis et al. (1998)): let $j,k \geq 1$ and $m > 0$

$$\alpha_{j,k}(m) = \sup\{|P(A\cap B) - P(A)P(B)| :\ A \in \mathcal{F}(\Lambda_1),\ B \in \mathcal{F}(\Lambda_2),\\ \Lambda_1 \in \mathcal{B}(\mathbb{R}^d),\ \Lambda_2 \in \mathcal{B}(\mathbb{R}^d),\ |\Lambda_1| \leq j,\ |\Lambda_2| \leq k,\ d(\Lambda_1,\Lambda_2) \geq m\} \quad (5)$$

where $\mathcal{F}(\Lambda_i)$ is the σ-algebra generated by $\mathbf{X}\cap\Lambda_i$, $i=1,2$. It is still an open question to know whether there are general conditions on the kernel C of a DPP providing a control of α-mixing coefficients. However, an obvious consequence of Proposition 4 is that for DPPs with compactly supported kernel, $\alpha_{j,k}(m) = 0$ for any $m > R$ and any $j,k \geq 1$.

2.3 On the distribution of the number of points

For general point processes, it is sometimes not easy to see what the distribution of the number of points in a compact set S is. For DPPs, we can actually show that this distribution is, for large compact set S, quite close to the probability distribution of a Poisson variable. This interesting behaviour will be exploited in Section 3.2. We let $\Pi(\theta)$ denote a Poisson random variable with parameter θ. The following proposition is based on results obtained by Zacharovas and Hwang (2010).

Proposition 5. *Let* $\mathbf{X}$ *be a DPP with kernel* C *verifying* $\mathcal{A}_C$. *Define for any* $m \geq 0$

$$d_0(S,m) = \mathrm{P}(N(\mathbf{X}\cap S) = m) - \mathrm{P}(\Pi(\lambda|S|) = m)$$
$$d_1(S,m) = \mathrm{P}(N(\mathbf{X}\cap S) = m) - \mathrm{P}(\Pi(\lambda|S|) = m)\left(1 - |S|\omega(m,\lambda|S|)\check{C}_0/2\right)$$

where $\check{C}_0 = \int_{\mathbb{R}^d} C^2(x)\mathrm{d}x$ *and* $\omega(m,\ell) = ((m-\ell)^2 - m)/\ell^2$ *for any* $m \geq 0$ *and* $\ell > 0$. *Then, there exists three constants* κ_0, κ_1 *and* κ_1', *independent of* m, *such that for all compact* $S \subset \mathbb{R}^d$ *we have*

$$|d_0(S,m)| \leq \frac{\kappa_0}{\sqrt{|S|}} \qquad \textit{and} \qquad |d_1(S,m)| \leq \frac{\kappa_1}{\sqrt{|S|}} + \frac{\kappa_1'}{|S|}. \tag{6}$$

In particular,

$$\kappa_0 = \sqrt{3}(\sqrt{e}-1)\frac{\check{C}_0\sqrt{\lambda}}{(\lambda - \check{C}_0)^2} \qquad \textit{and} \qquad \kappa_1 = \frac{\sqrt{15}(\sqrt{e}-1)}{2}\frac{\check{C}_0^2\sqrt{\lambda}}{(\lambda-\check{C}_0)^3}. \tag{7}$$

3 Estimators of λ

We are interested in the estimation of the intensity parameter λ based on a single realization of a DPP, $\mathbf{X}$, observed on an increasing sequence of bounded domains $W_n \subset \mathbb{R}^d$. The standard estimator is considered in Section 3.1 while the median-based estimator is studied in Section 3.2.

3.1 Standard estimator of λ

In this section, we assume that $\{W_n\}_{n\geq 1}$ is a sequence of bounded convex subsets of $\mathbb{R}^d$ such that for all $n \geq 1$, $W_n \subset W_{n+1}$ and there exists an Euclidean ball included in W_n with radius denoted by $r(W_n)$ tending to infinity as n tends to infinity. To shorten, such a sequence $\{W_n\}_{n\in\mathbb{N}}$ is said to be regular.

For a point process $\mathbf{X}$, the standard estimator of λ is given by

$$\widehat{\lambda}_n^{\mathrm{std}} = \frac{N(\mathbf{X}\cap W_n)}{|W_n|}. \tag{8}$$

Further, it is well-known that if the stationary point process is ergodic, a property established by Soshnikov (2000) for stationary DPPs, then this unbiased estimator is strongly consistent as $n \to \infty$. Using the Brillinger mixing property (Theorem 3), we obtain the following result.

Proposition 6 (Biscio and Lavancier (2016b); Soshnikov (2002))**.** *Let* $\mathbf{X}$ *be a stationary DPP with pair correlation function* g, *kernel* C *verifying* $\mathcal{A}_C$ *and let* $\{W_n\}_{n\in\mathbb{N}}$ *be a regular sequence of subsets of* $\mathbb{R}^d$. *Then, as* $n \to \infty$

$$\sqrt{|W_n|}\left(\widehat{\lambda}_n^{\mathrm{std}} - \lambda\right) \to \mathcal{N}(0,\sigma^2)$$

in distribution, where $\sigma^2 = \lambda + \lambda^2\int_{\mathbb{R}^d}(g(w)-1)dw$. *In particular, for* $\mathbf{X} \sim DPP(C)$, *we have* $\sigma^2 = \lambda - \check{C}_0 = \lambda - \int_{\mathbb{R}^d} C(x)^2\mathrm{d}x$.

The last result is not restricted to DPPs and is valid for a lot of spatial point processes (including some Cox processes, Gibbs point processes,...), up to the form of σ^2 that is in general known only in terms of the pair correlation function. The estimation of σ^2 has therefore been an important topic. We refer the reader to

Heinrich and Prokešová (2010) for a discussion of this challenging topic. In the latter paper, the following estimator is proposed.

$$\widehat{\sigma}_n^2 = \widehat{\lambda}_n^{\mathrm{std}} + \sum_{x,y\in\mathbf{X}\cap W_n}^{\neq} \frac{k\left(\frac{y-x}{|W_n|^{1/d}b_n}\right)}{|(W_n-x)\cap(W_n-y)|} - |W_n|b_n^d\widehat{\lambda}_n^{\mathrm{std}}(\widehat{\lambda}_n^{\mathrm{std}} - |W_n|^{-1})\int_{W_n} k(x)\,\mathrm{d}x$$

where $k:\mathbb{R}^d\to[0,\infty)$ plays the role of a kernel and $\{b_n\}_{n\geq 1}$ is a sequence of real numbers playing the role of a bandwidth. Heinrich and Prokešová (2010) obtained in particular the following result which can be directly applied to DPPs with kernel C satisfying $\mathcal{A}_C$.

Proposition 7 (Heinrich and Prokešová (2010))**.** *Let $k:\mathbb{R}^d\to[0,\infty)$ be a symmetric, bounded and continuous function at the origin of $\mathbb{R}^d$ such that $k(0)=1$. Let $\{W_n\}_{n\geq 1}$ be a sequence of regular subsets of $\mathbb{R}^d$ and $\{b_n\}_{n\geq 1}$ be a sequence of real numbers such that as $n\to\infty$*

$$b_n\to 0,\quad b_n^2|W_n|^{1/d}\to 0\quad \textit{and}\quad b_n|W_n|^{1/d}r(W_n)^{-1}\to 0$$

where $r(W_n)$ stands for the inball radius of W_n. If $\mathbf{X}$ is a Brillinger mixing point process, then as $n\to\infty$, $\widehat{\sigma}_n^2\to\sigma^2$ in L^2.

3.2 Median-based estimator of λ

For any real-valued random variable Y, we denote by $F_Y(\cdot)$ its cdf, by $F_Y^{-1}(p)$ its quantile of order $p\in(0,1)$ and by $\mathrm{Me}_Y = F_Y^{-1}(1/2)$ its theoretical median. Based on a sample $\mathbf{Y}=(Y_1,\ldots,Y_n)$ of n identically distributed random variables, we denote by $\widehat{F}(\cdot;\mathbf{Y})$ the empirical cdf, and by $\widehat{F}^{-1}(p;\mathbf{Y})$ the sample quantile of order p given by

$$\widehat{F}^{-1}(p;\mathbf{Y}) = \inf\{x\in\mathbb{R} : p\leq \widehat{F}(x;\mathbf{Y})\}. \tag{9}$$

The sample median is simply denoted by $\widehat{\mathrm{Me}}(\mathbf{Y}) = \widehat{F}^{-1}(1/2;\mathbf{Y})$.

In this section, we assume the following assumption, denoted by $\mathcal{A}_{W_n}$ for the sequence of bounded domain $\{W_n\}_{n\geq 1}$.

Assumption $\mathcal{A}_{W_n}$: The domain of observation W_n can be decomposed as $W_n = \cup_{k\in\mathcal{K}_n} C_{n,k}$ where the cells $C_{n,k}$ are non-overlapping and equally sized with volume $c_n = |C_{n,k}|$ and where $\mathcal{K}_n$ is a subset of $\mathbb{Z}^d$ with cardinality $k_n = |\mathcal{K}_n|$. As $n\to\infty$, $k_n\to\infty$, $c_n\to\infty$.

The standard estimator of λ is given by (8). To define a more robust one, we can note that

$$\widehat{\lambda}_n^{\mathrm{std}} = \frac{1}{k_n}\sum_{k\in\mathcal{K}_n}\frac{N(\mathbf{X}\cap C_{n,k})}{c_n} \tag{10}$$

since $|W_n| = k_n c_n$, i.e. $\widehat{\lambda}_n^{\mathrm{std}}$ is nothing else than the sample mean of intensity estimators computed in cells $C_{n,k}$. The strategy adopted by Coeurjolly (2016) was to replace the sample mean by the sample median, which is known to be more robust to outliers. Quantile estimators based on count data or more generally on discrete data can cause some troubles in the asymptotic theory (see e.g. David and Nagaraja

(2003)). To bypass the discontinuity problem of the count variables $N(\mathbf{X} \cap C_{n,k})$, we follow a well-known technique (e.g. Machado and Santos Silva (2005)) which introduces smoothness. Let $(U_k, k \in \mathcal{K}_n)$ be a collection of independent and identically distributed random variables, distributed as $U \sim \mathcal{U}([0,1])$. Then, for any $k \in \mathcal{K}_n$, we define

$$Z_{n,k} = N(\mathbf{X} \cap C_{n,k}) + U_k \quad \text{and} \quad \mathbf{Z} = (Z_{n,k},\ k \in \mathcal{K}_n). \tag{11}$$

By **Assumption** $\mathcal{A}_{W_n}$ and the stationarity of $\mathbf{X}$, the variables $Z_{n,k}$ are identically distributed and we let $Z \sim Z_{n,k}$. The jittering effect shows up right away: the cdf of Z is given for any $t \geq 0$ by

$$F_Z(t) = P(N(\mathbf{X} \cap C_{n,0}) \leq \lfloor t \rfloor - 1) + P(N(\mathbf{X} \cap C_{n,0}) = \lfloor t \rfloor)\,(t - \lfloor t \rfloor),$$

and is continuously differentiable whereby we deduce that Z admits a density f_Z at t given by $f_Z(t) = P(N(\mathbf{X} \cap C_{n,0}) = \lfloor t \rfloor)$. We define the jittered median-based estimator of λ by

$$\widehat{\lambda}_n^{\text{med}} = \frac{\widehat{\text{Me}}(\mathbf{Z})}{c_n} \tag{12}$$

where the sample median is defined by (9). To derive asymptotic properties for $\widehat{\lambda}_n^{\text{med}}$, we need to consider the subclass of compactly supported DPPs, summarized by the following assumption.

Assumption $\mathcal{A}_{C_R}$. The kernel C of the stationary DPP satisfies $\mathcal{A}_C$. There exists $R > 0$ such that $C(x) = 0$ for any $|x| > R$.

Finally a technical condition, ensuring the asymptotic positivity of the density at the median is required.

Assumption $\mathcal{A}_{\text{med}}$. $\liminf_{n\to\infty} s_n > 0$, where $s_n = \sqrt{c_n}\text{P}(N(\mathbf{X} \cap C_{n,0}) = \lfloor \text{Me}_Z \rfloor)$.

Proposition 8. *Assume that the sequence of domains satisfies $\mathcal{A}_{W_n}$, that the DPP with kernel C, $\mathbf{X}$, satisfies $\mathcal{A}_{C_R}$ and that $\mathcal{A}_{\text{med}}$ holds. Then, as $n \to \infty$,*
(a)

$$\sqrt{|W_n|}s_n \left(\frac{\widehat{\text{Me}}(\mathbf{Z})}{c_n} - \frac{\text{Me}_Z}{c_n} \right) \to \mathcal{N}(0, 1/4) \tag{13}$$

in distribution.
(b) $\text{Me}_Z - \lambda c_n = o(\sqrt{c_n})$.
(c) If in addition, $\sqrt{k_n}(\text{Me}_Z - \lambda c_n)/\sqrt{c_n} \to 0$ *as* $n \to \infty$, *then*

$$2\sqrt{|W_n|}s_n \left(\widehat{\lambda}_n^{\text{med}} - \lambda \right) \to \mathcal{N}(0, 1) \tag{14}$$

in distribution.

It is worth comparing (14) with Coeurjolly (2016, Corollary 6) devoted to Cox processes. For Cox processes, the author is able to provide conditions on the Cox process for which s_n admits a limit equal to $(2\pi\sigma^2)^{-1/2}$, where $\sigma^2 = \lambda + \int_{\mathbb{R}^d}(g(x)-1)\text{d}x$. To prove this, Coeurjolly (2016) uses explicitly the connection between Poisson and Cox point processes. We were unable to prove any result of that type for stationary

DPP (and actually conjecture that there is no limit). An interesting fact can however be noticed. Since DPP are purely repulsive models, the asymptotic variance of $\widehat{\lambda}^{\mathrm{std}}$, that is σ^2 is always bounded by λ, that is, by the corresponding asymptotic variance under the Poisson model. Transferring this to the median-based estimator, we conjecture that $|W_n|\operatorname{Var}(\widehat{\lambda}^{\mathrm{med}})$ is asymptotically bounded by the corresponding variance under the Poisson case, which is precisely $\pi\lambda/2$. By replacing λ by its estimate, we have the basis to propose an asymptotic conservative confidence interval for λ.

We end this section by stating a sufficient condition which ensures $\mathcal{A}_{\mathrm{med}}$.

Proposition 9. *If the DPP with kernel C, $\mathbf{X}$, satisfies $\mathcal{A}_C$, then $\mathcal{A}_{\mathrm{med}}$ holds if*

$$\max\left((2\pi\lambda)^{-1/2}-\kappa_0,(2\pi\lambda)^{-1/2}\left(1+\frac{\check{C}_0}{2\lambda}\right)-\kappa_1\right)>0 \tag{15}$$

where κ_0 and κ_1 are given by (7) *and $\check{C}_0=\int_{\mathbb{R}^d}C^2(x)\mathrm{d}x$.*

Condition (15) is a theoretical condition, which allows us to understand what kind of kernels C can satisfy $\mathcal{A}_{\mathrm{med}}$. From a practical point of view, if we assume that the data can be modelled by a stationary DPP, the condition (15) can be tested by plugging an estimate of λ and $\check{C}_0$. Note that $\check{C}_0=\lambda-\sigma^2$ can be estimated by $\widehat{\lambda}_n^{\mathrm{med}}-\widehat{\sigma}_n^2$ where $\widehat{\sigma}_n^2$ is the estimate detailed in Section 3.1 into which we can also replace $\widehat{\lambda}_n^{\mathrm{std}}$ by $\widehat{\lambda}_n^{\mathrm{med}}$ if the presence of outliers is suspected.

Proof. Using Proposition 5 (and the notation therein), we have

$$s_n=\sqrt{c_n}\mathrm{P}(N(\mathbf{X}\cap C_{n,0})=\lfloor\mathrm{Me}_Z\rfloor)\geq\max\left(L_{0,n},L_{1,n}\right)$$

where

$$\begin{aligned}L_{0,n}=&\sqrt{c_n}\mathrm{P}(\Pi(\lambda c_n)=\lfloor\mathrm{Me}_Z\rfloor)-\kappa_0\\ L_{1,n}=&\sqrt{c_n}\mathrm{P}(\Pi(\lambda c_n)=\lfloor\mathrm{Me}_Z\rfloor)\left(1-c_n\omega(\lfloor\mathrm{Me}_Z\rfloor,\lambda c_n)\check{C}_0/2\right)-\kappa_1.\end{aligned}$$

As $n\to\infty$, Proposition 8 (a) and (23) yield that

$$\mathrm{P}(\Pi(\lambda c_n)=\lfloor\mathrm{Me}_Z\rfloor)\to(2\pi\lambda)^{-1/2}\quad\text{ and }\quad c_n\omega(\lfloor\mathrm{Me}_Z\rfloor,\lambda c_n)\to-1$$

whereby we deduce the result. □

4 Simulation study

In this section, we investigate the performances of (8) and (12) for planar stationary DPPs. The study follows the one done in Coeurjolly (2016) which was mainly designed for Cox processes.

We consider the following DPP model introduced by Biscio and Lavancier (2016a). For $\nu>0$, let j_ν be the first positive zeros of the Bessel function of the first kind J_ν and define the constant M by $M\lambda^{1/d}=\left(2^{d-2}j^2_{\frac{d-2}{2}}\Gamma\left(\frac{d}{2}\right)\right)^{1/d}/\pi^{1/2}$.When $d=2$,

$M\lambda^{1/2} = j_0/\pi^{1/2} \approx 1.357$. Now, let $R \in (0, M]$. We define the kernel $C_R = u_R * u_R$ where

$$u_R(x) = \kappa \, \frac{J_{\frac{d-2}{2}}\left(2j_{\frac{d-2}{2}}\frac{|x|}{R}\right)}{|x|^{\frac{d-2}{2}}} \, \mathbf{1}_{\left\{|x|<\frac{R}{2}\right\}},$$

and $\kappa^2 = \frac{4\Gamma(d/2)}{\lambda\pi^{d/2}R^2}\left(J'_{\frac{d-2}{2}}(j_{\frac{d-2}{2}})\right)^{-2}$. Note that there exist many other kernels of DPPs which are compactly supported, see for instance Biscio and Lavancier (2016a, Proposition 4.1). The advantage of the kernel C_R is that its Fourier transform is explicit and thus $\mathcal{A}_{C_R}$ can be investigated. In particular, for any, $x \in \mathbb{R}^d$, it can be shown that, $\mathcal{F}(C_R)(x) \leq \frac{R^d}{M^d}$. Thus, the kernel C_R satisfies $\mathcal{A}_{C_R}$ for all $R < M$. Two different versions of this model, denoted DPP1 and DPP2, obtained by setting R to the values $R = M/4$ and $R = 3M/4$ respectively, are considered in the simulation study. Figure 1 depicts the pair correlation functions g for these models as well as a realization of each of these processes. It is to be noted that the models DPP1 and DPP2 satisfy $\mathcal{A}_{C_R}$ and $\mathcal{A}_{\text{med}}$. Specifically, the constants involved in (15) are numerically evaluated to 0.057 and 0.021 respectively.

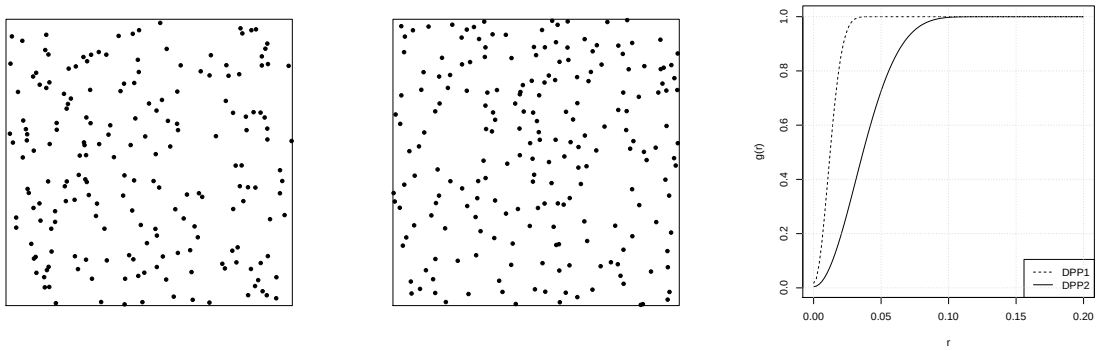


Figure 1: Left (resp. middle): Realization of the model DPP1 (resp. DPP2) on the domain $[-1, 1]^2$. The intensity equals $\lambda = 50$; Right: Pair correlation function g for the models DPP1,DPP2.

The models DPP1, DPP2 are generated on $W_n = [-n, n]^2$ for $n = 1, 2$ and we consider the three following settings: let $\mathbf{y}$ be a realization from one of the two models described above, generated on W_n and with m points. The observed point pattern is denoted by $\mathbf{x}$ and is obtained as follows.

(A) Pure case: no modification is considered, $\mathbf{x} = \mathbf{y}$.

(B) A few points are added: in a sub-square Δ_n with side-length $|\Delta_n|^{1/2} = n/5$ included in W_n and randomly chosen, we generate a point process $\mathbf{y}^{\text{add}}$ of $n^{\text{add}} = \rho\, m$ uniform points in Δ_n. We choose $\rho = 0.05$ or 0.1. Then, we define $\mathbf{x} = \mathbf{y} \cup \mathbf{y}^{\text{add}}$.

(C) A few points are deleted: let Δ_n be a randomly chosen sub-square included in W_n. The volume of Δ_n is chosen such that $\mathrm{E}(N(\mathbf{X} \cap \Delta_n)) = \rho\ \mathrm{E}(N(\mathbf{X} \cap W_n)) = \rho\lambda|W_n|$, with $\mathbf{X} \sim$ DPP1 or DPP2, and we choose either $\rho = 0.05$ or 0.1. Then, we define $\mathbf{x} = \mathbf{y} \setminus \Delta_n$, i.e. $\mathbf{x}$ is the initial configuration thinned by 5% or 10% of its points (on average) located in the sub-square Δ_n.

An illustration of settings (B) and (C) is proposed in Figure 2. We conduct a Monte Carlo simulation and generate 500 replications of the models DPP1, DPP2 with intensity parameter $\lambda = 50$ and for the three different settings (A)-(C). For each replication, we evaluate $\widehat{\lambda}_n^{\mathrm{std}}$ and $\widehat{\lambda}_n^{\mathrm{med}}$ for different number of non-overlapping and equally sized cells k_n. More precisely, we choose $k_n = 9, 16, 25, 36, 49$. The empirical results can be sometimes quite influenced by the choice of the number of blocks k_n. In a separate analysis not reported, we have noticed that, depending on the situation, the estimates could be far from λ for some k_n but, also, that there are consecutive values of k_n producing close values. Following this empirical finding, we propose the data-driven estimator for λ, denoted by $\widetilde{\lambda}_n^{\mathrm{med}}$ and defined as the median of the median-based estimators, that is

$$\widetilde{\lambda}_n^{\mathrm{med}} = \widehat{\mathrm{Me}}\left(\left\{\widehat{\lambda}_n^{\mathrm{med}},\ k_n = 9, 16, 25, 36, 49\right\}\right). \tag{16}$$

The estimator $\widetilde{\lambda}_n^{\mathrm{med}}$ is very simple and quick to evaluate. It is a reasonable procedure as it follows standard ideas from aggregated estimators (see e.g. Lavancier and Rochet (2016)). Let us add that it requires only to tune a grid of reasonable values for the number of block cells. To set this grid, we suggest to start with a small number of blocks, 9 or 16, and increase it until the estimate $\widehat{\lambda}_n^{\mathrm{med}}$ significantly deviates from the other ones.

Tables 1-3 summarize the results. We report empirical results for $\widehat{\lambda}_n^{\mathrm{std}}$, for the median-based estimator $\widehat{\lambda}_n^{\mathrm{med}}$ for $k_n = 9, 25$ and 49 and for the data-driven estimator $\widetilde{\lambda}_n^{\mathrm{med}}$. Table 1 reports empirical means and standard deviations for the pure case (A). Tables 2 and 3 are respectively related to the settings (B) and (C). The two latter can affect significantly the bias of the estimator. In both tables, we report the bias of the different estimators and the gain (in percent) in terms of mean squared error of $\widehat{\lambda} = \widehat{\lambda}_n^{\mathrm{med}}$ or $\widehat{\lambda} = \widetilde{\lambda}_n^{\mathrm{med}}$ with respect to $\widehat{\lambda}_n^{\mathrm{std}}$, i.e. for each model and each value of ρ, n, k_n, we compute

$$\widehat{\mathrm{Gain}}(\widehat{\lambda}) = \left(\frac{\widehat{\mathrm{MSE}}(\widehat{\lambda}_n^{\mathrm{std}}) - \widehat{\mathrm{MSE}}(\widehat{\lambda})}{\widehat{\mathrm{MSE}}(\widehat{\lambda}_n^{\mathrm{std}})}\right) \times 100\% \tag{17}$$

where $\widehat{\mathrm{MSE}}$ is the empirical mean squared error based on the 500 replications. Thus a positive (resp. negative) empirical gain means that the median-based estimator is more efficient (resp. less efficient) than the standard procedure.

Table 1 shows that the standard and the median-based estimators are consistent when n increases. The parameter k_n looks crucial when $n = 1$. In particular the bias seems to increase with k_n. When $n = 2$, its influence is much less important. It is also interesting to note that the choice of k_n does not change that much the empirical standard deviations. With absence of outliers, the standard estimator

obviously outperforms the median-based estimators, but it is interesting to note that the loss of efficiency is not too important as n increases. Our data-driven median-based estimator, surprisingly, exhibits very nice properties. The bias seems to be averaged over the k_n when $n = 1$ and the procedure is even able to reduce the standard deviation. Following the remark after Proposition 8 the standard deviation of the median-based estimator is difficult to estimate but it can be bounded by the corresponding standard deviation under the Poisson case. According to the simulation setting, this upper-bound is equal to $\sqrt{\pi\lambda/2}/(2n)$, which is equal to 4.4 when $n = 1$ and 2.2 when $n = 2$. This indeed bounds the empirical standard deviation of $\widetilde{\lambda}_n^{\text{med}}$ and $\widehat{\lambda}_n^{\text{med}}$ for any k_n.

Regarding Tables 2 and 3, we can observe that $\widehat{\lambda}_n^{\text{std}}$ gets biased. As expected, this bias is less important for $\widehat{\lambda}_n^{\text{med}}$. When $n = 1$ and $\rho = 0.05$, the standard estimator remains much better than the median-based estimator: the gain is negative and can reach very low values. This behaviour also holds in the case (B) when $\rho = 0.1$. In the other situations, the median-based estimator outperforms the standard procedure with a positive gain for almost all the values of k_n. The fluctuation of the gain with k_n is not very satisfactory and justifies again the introduction of a data-driven procedure. In the setting (B), like $\widehat{\lambda}_n^{\text{med}}$, $\widetilde{\lambda}_n^{\text{med}}$ outperforms the standard estimator when $n = 2$ and behaves similarly to the standard estimator when $n = 1$ and $\rho = 0.1$. Like in the setting (A), the standard deviation of $\widetilde{\lambda}_n^{\text{med}}$ is shown to be smaller than the ones of $\widehat{\lambda}_n^{\text{med}}$ for all values of k_n, which explains why we observe a higher gain. The conclusion for the setting (C) is unambiguous: the performances of $\widetilde{\lambda}_n^{\text{med}}$ are very good even for small observation window or when only 5% of points on average are deleted. Again, the gain of $\widetilde{\lambda}_n^{\text{med}}$ is larger than the gains obtained from $\widehat{\lambda}_n^{\text{med}}$ for all the values of k_n, except for the DPP2 model when $n = 2$ an $k_n = 49$ for which the observed empirical gain is slightly larger. As a general comment for Tables 2 and 3, we observe the more repulsive the point pattern, the higher the performances of the robust estimates.

	Empirical mean (Standard Deviation)				
	$\widehat{\lambda}_n^{\text{std}}$	$\widehat{\lambda}_n^{\text{med}}$			$\widetilde{\lambda}_n^{\text{med}}$
		$k_n = 9$	25	49	
DPP1					
$n = 1$	49.7 (3.5)	50.6 (4.3)	52.1 (4.3)	54.1 (4.6)	52.1 (3.9)
$n = 2$	49.5 (1.6)	49.8 (2.1)	50.1 (2.1)	50.7 (2.1)	50.1 (1.8)
DPP2					
$n = 1$	50.0 (3.0)	51.1 (3.7)	52.8 (3.9)	55.3 (3.7)	52.9 (3.4)
$n = 2$	50.0 (1.5)	50.3 (1.8)	50.6 (1.9)	51.2 (1.9)	50.6 (1.6)

Table 1: Empirical means and standard deviations between brackets of estimates of the intensity $\lambda = 50$ for different models of determinantal point processes (DPP1, DPP2). The empirical results are based on 500 replications simulated on $[-n, n]^2$ for $n = 1, 2$. The first column corresponds to the standard estimator given by (8) while the following ones correspond to the median-based estimators given by (12) for different number of cells k_n and by (16) for the data-driven procedure.

	Bias (Gain of MSE %)				
	$\widehat{\lambda}_n^{\text{std}}$	$\widehat{\lambda}_n^{\text{med}}$			$\widetilde{\lambda}_n^{\text{med}}$
		$k_n = 9$	25	49	
$\rho = 0.05$					
DPP1, $n = 1$	1.9 (0)	1.7 (-48)	3.0 (-73)	5.2 (-203)	3.2 (-62)
$n = 2$	1.9 (0)	0.4 (18)	0.8 (12)	1.4 (-1)	0.8 (36)
DPP2, $n = 1$	2.3 (0)	2.0 (-39)	3.5 (-86)	5.9 (-253)	3.6 (-72)
$n = 2$	2.3 (0)	0.8 (36)	1.1 (30)	1.8 (10)	1.2 (42)
$\rho = 0.1$					
DPP1, $n = 1$	4.7 (0)	2.8 (13)	4.0 (-8)	6.0 (-61)	4.2 (1)
$n = 2$	4.8 (0)	1.1 (71)	1.3 (70)	1.9 (66)	1.4 (77)
DPP2, $n = 1$	5.0 (0)	3.2 (19)	4.4 (-4)	6.8 (-87)	4.7 (1)
$n = 2$	4.8 (0)	1.1 (76)	1.3 (77)	2.0 (69)	1.4 (80)

Table 2: Bias and empirical gains in percent between brackets, see (17), for the standard and median based estimators for different values of k_n. The empirical results are based on 500 replications generated on $[-n, n]^2$ for $n = 1, 2$ for the models DPP1, DPP2 where 5% or 10% of points are added to each configuration. This corresponds to the case (B) described in details above.

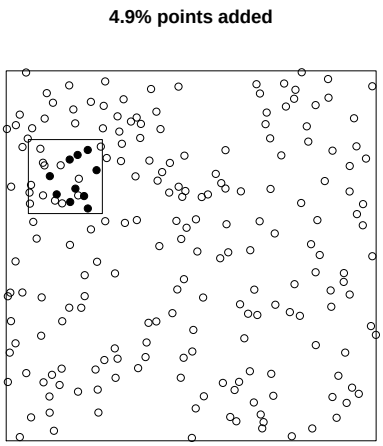
	Bias (Gain of MSE %)				
	$\widehat{\lambda}_n^{\text{std}}$	$\widehat{\lambda}_n^{\text{med}}$			$\widetilde{\lambda}_n^{\text{med}}$
		$k_n = 9$	25	49	
$\rho = 0.05$					
DPP1, $n = 1$	-2.8 (0)	-2.0 (0)	-0.7 (3)	1.6 (-10)	-0.6 (30)
$n = 2$	-2.9 (0)	-2.8 (-9)	-1.8 (25)	-0.6 (52)	-1.7 (40)
DPP2, $n = 1$	-2.6 (0)	-1.5 (-9)	0.2 (4)	2.7 (-51)	0.2 (28)
$n = 2$	-2.6 (0)	-2.3 (-4)	-1.1 (41)	-0.1 (56)	-1.1 (53)
$\rho = 0.1$					
DPP1, $n = 1$	-5.4 (0)	-4.7 (-3)	-2.3 (35)	0.0 (46)	-2.3 (48)
$n = 2$	-5.4 (0)	-5.0 (-1)	-1.9 (69)	-1.7 (74)	-2.3 (69)
DPP2, $n = 1$	-4.9 (0)	-3.9 (5)	-0.8 (41)	2.0 (36)	-0.9 (56)
$n = 2$	-5.2 (0)	-4.4 (12)	-1.3 (79)	-1.1 (83)	-1.6 (80)

Table 3: Bias and empirical gains in percent between brackets, see (17), for the standard and median based estimators for different values of k_n. The empirical results are based on 500 replications generated on $[-n, n]^2$ for $n = 1, 2$ for the models DPP1, DPP2 where 5% or 10% of points are deleted to each configuration. This corresponds to the case (C) described in details above.

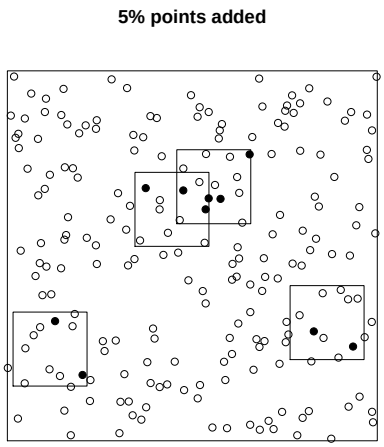
The differences of performances of the median-based estimators between the settings (B) or (C) were not expected. To investigate this more, we extend the simulation study. For the case (B), we investigate a different number of randomly chosen sub-squares (specifically 1,2 and 4 sub-squares) and with different side-length (specifically $n/10, n/5$ and $2n/5$) into which points are added. For the setting (C), we also investigate the possibility to delete on average $\rho = 5\%$ of the initial points

in 1, 2 or 4 randomly chosen sub-squares. We consider only the DPP2 model and the estimator $\widetilde{\lambda}_n^{\text{med}}$. Table 4 reports empirical results based on 500 replications. Like Tables 2 and 3, we report the empirical bias and gain. For the contamination (B), like Table 2, we observe that when $n = 1$, the results are not in favor of $\widetilde{\lambda}_n^{\text{med}}$. When $n = 2$, we remark, as expected, that the larger the sub-squares Δ_n, the lower the gain. The differences are quite similar when the number of sub-squares increases: the larger the number of sub-squares, the lower the gain. As a conclusion, when data exhibit repulsion with a suspicion of areas with extra points, we recommend to use the estimator $\widetilde{\lambda}_n^{\text{med}}$ if those areas are not too large. Regarding the contamination (C), the conclusion is different. Even if we observe that the performances of the estimator decrease with the number of sub-squares, the gain is still very significant. In other words, the estimator $\widetilde{\lambda}_n^{\text{med}}$ is shown to be very robust to missing information for repulsive point patterns.

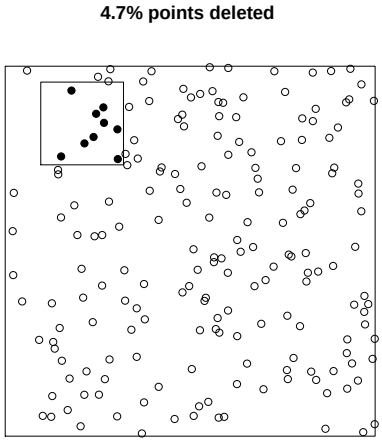
To go further, as suggested by one reviewer, we investigate another type of outliers which is the addition (resp. deletion) of points uniformly on W (resp. on $\mathbf{X}$). When 5% of points are added on average, we observe empirical biases and gains of 7.7. and -105% when $n = 1$ and 5.5 and -26% when $n = 2$. This clearly shows the limitation of the median-based estimator for repulsive patterns. It should not be used at all if we think that extra data are uniform on the observation domain. When, we delete 5% of points uniformly, we observe -2.5 and 53% for the empirical bias and gain, when $n = 1$ and -3.4 and 23% when $n = 2$. Surprisingly, the median-based estimator tends to be quite efficient compared to the standard estimator. It is somehow difficult to explain why the empirical bias increases with n. Overall, we think that a median-based estimator, is not tailored-made to take into account for outliers which are not "enough" isolated. We believe another approach should be considered for such a problem, like the one proposed by Redenbach et al. (2015). In the mentioned paper, the authors construct an MCMC algorithm which estimates the parameters of the superposition of a Strauss point process and a Poisson point process.



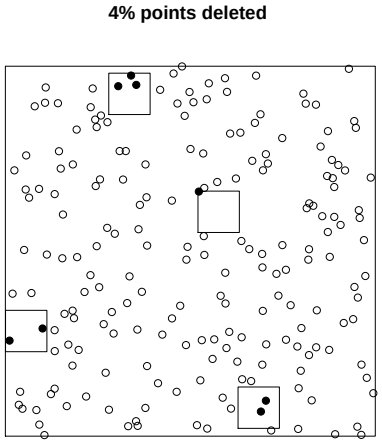
(a) Contamination (B), one sub-square



(b) Contamination (B), four sub-squares



(c) Contamination (C), one sub-square



(d) Contamination (C), four sub-squares

Figure 2: Examples of realizations after a contamination of type (B) or (C). The initial patterns are realizations of the DPP2 model with intensity $\lambda = 50$, on the domain $[-1, 1]^2$. For the setting (B) (Figures (a) and (b)), the filled circles represent extra points added to the initial pattern. For the setting (C), the filled circles represent deleted points from the initial pattern. On average, 5% of points are added or deleted.

	Number of sub-squares		
	1	2	4
Contamination (B)			
$\lvert\Delta_n\rvert^{1/2} = n/10, n = 1$	4.3 (8)	5.3 (-25)	7.0 (-80)
$n = 2$	1.3 (81)	2.2 (68)	3.6 (39)
$\lvert\Delta_n\rvert^{1/2} = n/5, n = 1$	5.1 (-11)	6.1 (-55)	7.2 (-96)
$n = 2$	2.0 (72)	3.2 (48)	4.6 (7)
$\lvert\Delta_n\rvert^{1/2} = 2n/5, n = 1$	6.6 (-66)	7.3 (-93)	7.5 (-109)
$n = 2$	3.7 (35)	4.9 (1)	5.1 (-16)
Contamination (C)			
$n = 1$	-0.9 (63)	-1.4 (59)	-2 (55)
$n = 2$	-1.0 (78)	-1.3 (71)	-2 (55)

Table 4: Bias and empirical gains in percent between brackets, see (17), for the estimator $\widetilde{\lambda}_n^{\text{med}}$ given by (16). The empirical results are based on 500 replications generated on $[-n, n]^2$ for $n = 1, 2$ for the model DPP2. To each point pattern, we add (contamination (B)) or delete (contamination (C)) on average 5% of points in 1,2 or 4 randomly chosen sub-squares Δ_n. In the setting (B), we investigate different values for the side-length of Δ_n.

5 Conclusion

In this paper, we focus on the class of stationary determinantal point processes and present two estimators of the intensity parameter for which we prove asymptotic properties. Among the two estimators, one of them, namely the median-based estimator is tailored to be robust to outliers. The median-based estimator depends on a tuning estimator, the number of blocks into which the original window is divided. The empirical findings show that the results are quite sensitive to this parameter. To correct that sensitivity we propose a combined approach and define the estimator $\widetilde{\lambda}_n^{\text{med}}$ as the median of median-based estimators computed for different number of blocks. The estimator $\widetilde{\lambda}_n^{\text{med}}$ is very simple and quick to perform.

As a general conclusion of the simulation study, this combined estimator turns out to be robust to outliers. When at least 5% out of 200 or more points, on average, lying in possibly different areas of the observation domain are assumed to be not observed, we recommend the use of the combined estimator. If at least 10% of points are added to a regular pattern with at least 200 points, on average, and when these extra points located in 1,2 or 4 areas, we also recommend to use the combined estimator. This also holds, if 5% of points are added to a regular pattern with an average of at least 800 points, or with patterns with at least an average of 200 points for which the extra points are localized in one small area (say 1/25th of the area of the observation domain). In the other situations, we recommend to use the standard estimator.

In this work, we did not aim at detecting outliers or detecting areas where problems are suspected (abundance or lack of points). If the assumption of stationarity seems valid, an inspection of the scan statistics (see e.g. Baddeley et al. (2015)) or a large difference between the median-based estimator and the standard estimator of the intensity parameter might allow the user to reconsider the observation window in a

second step. This has not been considered in this papper.

Appendix

Proof of Proposition 5

It can be shown from the spectral decomposition (2) (see e.g. Lavancier et al. (2015)), that the number of points in S satisfies $N(\mathbf{X} \cap S) = \sum_{k=1}^{\infty} \mathcal{B}(\beta_k^S)$ in distribution, where $\mathcal{B}(\beta_k^S)$ are independent Bernoulli random variables with parameters β_k^S.

For $n \geq 1$, we define $N_n(\mathbf{X} \cap S) = \sum_{k=1}^{n} \mathcal{B}(\beta_k^S)$. Under the assumptions of Proposition 2, $N_n(\mathbf{X} \cap S)$ converges in distribution to $N(\mathbf{X} \cap S)$ as n tends to infinity. The random variable $N_n(\mathbf{X} \cap S)$ is nothing else than a Poisson-Binomial distribution.

We first prove the first part of (6). By Zacharovas and Hwang (2010, Theorem 3.4), we have for all $m \geq 0$,

$$|\mathrm{P}(N_n(\mathbf{X} \cap S)) = m) - \mathrm{P}(\Pi(\theta_{1,n}) = m)| \leq \sqrt{3}(\sqrt{e} - 1)\frac{\theta_{2,n}\sqrt{\theta_{1,n}}}{(\theta_{1,n} - \theta_{2,n})^2} \tag{18}$$

where $\theta_{i,n} = \sum_{k=1}^{n} (\beta_k^S)^i$ for $i = 1, 2$. As $n \to \infty$, we have

$$\theta_{1,n} \to \sum_{k \geq 1} \beta_k^S = \mathrm{E}(N(\mathbf{X} \cap S)) = \lambda |S| \tag{19}$$

$$\theta_{2,n} \to \sum_{k \geq 1} (\beta_k^S)^2 = \mathrm{E}(N(\mathbf{X} \cap S)) - \mathrm{Var}(N(\mathbf{X} \cap S)) = \int_{S^2} C^2(x - y)\mathrm{d}x\mathrm{d}y. \tag{20}$$

We note first, that, a change of variables yields $\int_{S^2} C^2(x-y)\mathrm{d}x\mathrm{d}y < |S|\,\check{C}_0$ and that, second, by Parseval's identity and since $0 \leq \mathcal{F}(C) < 1$,

$$\lambda - \check{C}_0 = \lambda - \int_{\mathbb{R}^d} \mathcal{F}(C)^2(x)\mathrm{d}x > \lambda - \int_{\mathbb{R}^d} \mathcal{F}(C)(x)\mathrm{d}x = 0$$

whereby we deduce that

$$\lim_{n \to \infty} (\theta_{1,n} - \theta_{2,n}) \geq |S| \left(\lambda - \check{C}_0\right) > 0.$$

Hence, the first part of (6) is obtained by letting n tend to infinity in (18). For the second part of (6), we use Zacharovas and Hwang (2010, Theorem 4.2) which states that for some constant $\tilde{c}$ independent of m, n and S

$$|\Delta_n(S, m)| \leq \frac{\sqrt{15}(\sqrt{e} - 1)}{2} \frac{\theta_{2,n}^2 \sqrt{\theta_{1,n}}}{(\theta_{1,n} - \theta_{2,n})^3} + \tilde{c} \frac{\theta_{3,n}\sqrt{\theta_{1,n}}}{(\theta_{1,n} - \theta_{2,n})^{5/2}} \tag{21}$$

where $\Delta_n(S, m) = \mathrm{P}(N_n(\mathbf{X} \cap S) = m) - \mathrm{P}(\Pi(\theta_{1,n}) = m)\,(1 - \omega(\theta_{1,n}, m)\theta_{2,n}/2)$ and $\theta_{3,n} = \sum_{k=1}^{n} (\beta_k^S)^3$. Since $\beta_k^S \leq 1$, then using (19) and (20), we can show that

$$\lim_{n \to \infty} \frac{\sqrt{15}(\sqrt{e} - 1)}{2} \frac{\theta_{2,n}^2 \sqrt{\theta_{1,n}}}{(\theta_{1,n} - \theta_{2,n})^3} \to \kappa_1 \quad \text{and} \quad \sup_{n \to \infty} \frac{\theta_{3,n}\sqrt{\theta_{1,n}}}{(\theta_{1,n} - \theta_{2,n})^{5/2}} \leq \frac{1}{|S|} \frac{\lambda^{3/2}}{(\lambda - \check{C}_0)^{5/2}}.$$

Using these results, we deduce the second part of (6) by letting $n \to \infty$ in (21).

Proof of Proposition 8

We start with two ingredients used in (a)-(c). First, from Proposition 5

$$\sup_{m\geq 0} |\mathrm{P}(N(\mathbf{X}\cap C_{n,0})=m) - \mathrm{P}(\Pi(\lambda c_n)=m)| = \mathcal{O}(c_n^{-1/2}). \tag{22}$$

Second using Stirling's formula we have for any $\omega\in\mathbb{R}$ as $n\to\infty$

$$\mathrm{P}(\Pi(\lambda c_n)=\lfloor v_n\rfloor) \sim \begin{cases} (2\pi\lambda c_n)^{-1/2} & \text{if } v_n=\lambda c_n + o(c_n^{1/2}) \\ (2\pi\lambda c_n)^{-1/2}e^{-\omega^2/2} & \text{if } v_n=\lambda c_n+\omega c_n^{1/2}. \end{cases} \tag{23}$$

(a) Under the assumptions (i)-(v) described below, Coeurjolly (2016)[Theorem 4.3] proved that

$$\sqrt{|W_n|}s_n\left(\frac{\widehat{\mathrm{Me}}(\mathbf{Z})}{c_n} - \frac{\mathrm{Me}_Z}{c_n}\right) \to \mathcal{N}(0,1/4) \tag{24}$$

in distribution.
(i) As $n\to\infty$, $k_n\to\infty, c_n\to\infty$ and $k_n/c_n^{\eta'/2}\to 0$ where $0<\eta'<\eta$ where η is given by (iv).
(ii) $\forall t_n=\lambda c_n+\mathcal{O}(\sqrt{c_n/k_n})$, $\mathrm{P}(N(\mathbf{X}\cap C_{n,0})=\lfloor t_n\rfloor)/\mathrm{P}(N(\mathbf{X}\cap C_{n,0})=\lfloor\lambda c_n\rfloor)\to 1$.
(iii) $\liminf_{n\to\infty} s_n>0$ and $\limsup_{n\to\infty} s_n<\infty$.
(iv) $\mathbf{X}$ has a pair correlation function g satisfying $\int_{\mathbb{R}^d}|g(w)-1|\mathrm{d}w<\infty$.
(v) There exists $\eta>0$ such that

$$\alpha(m)=\sup_{p\geq 1}\frac{\alpha_{p,p}(m)}{p}=\mathcal{O}(m^{-d(1+\eta)}) \quad \text{and} \quad \alpha_{2,\infty}(m)=\mathcal{O}(m^{-d(1+\eta)})$$

where $\alpha_{j,k}(m)$ for $j,k\in\mathbb{N}\cup\{\infty\}$ is defined by (5).

Therefore, the proof of (a) consists in verifying that $\mathcal{A}_{W_n}$, $\mathcal{A}_{C_R}$ and $\mathcal{A}_{\mathrm{med}}$ imply (i)-(v). By $\mathcal{A}_{C_R}$, $\mathbf{X}$ is R-dependent and thus for any $m>R$ and $j,k\geq 1$, $\alpha_{j,k}(m)=0$. (v) is thus obviously satisfied and $k_n/c_n^{\eta'/2}\to 0$ can always be fulfilled. From (22)

$$\frac{\mathrm{P}(N(\mathbf{X}\cap C_{n,0})=\lfloor t_n\rfloor)}{\mathrm{P}(N(\mathbf{X}\cap C_{n,0})=\lfloor\lambda c_n\rfloor)} \sim \frac{\mathrm{P}(\Pi(\lambda c_n)=\lfloor t_n\rfloor)}{\mathrm{P}(\Pi(\lambda c_n)=\lfloor\lambda c_n\rfloor)}$$

as $n\to\infty$. By $\mathcal{A}_{W_n}$, $t_n=\lambda c_n+o(c_n^{1/2})$ whereby (ii) is deduced from (23). The first part of (iii) corresponds to $\mathcal{A}_{\mathrm{med}}$ while the second part is deduced from (ii). (iv) is also clearly satisfied since $\mathbf{X}$ is Brillinger mixing.

(b) First of all, by the Brillinger mixing property, Coeurjolly (2016)[Proposition 3.1] can be applied to derive $\mathrm{Me}_Z-\lambda c_n=\mathcal{O}(\sqrt{c_n})$. Now, under the assumptions (i), (iv) and (v), we follow the proof of Coeurjolly (2016)[Theorem 4.2 Step 1] and state that: $|W_n|^{-1/2}(Z_{n,0}-\lambda c_n)\to\mathcal{N}(0,\tau^2)$ in distribution for some $\tau>0$, whereby we deduce that $\mathrm{P}(Z_{n,0}\leq\lambda c_n)-1/2\to 0$ as $n\to\infty$. Since $Z_{n,0}$ is a continuous random variable, the latter can also be rewritten as

$$\mathrm{P}(Z_{n,0}\leq\lambda c_n)-\mathrm{P}(Z_{n,0}\leq\mathrm{Me}_Z)=o(1).$$

Since F_Z is differentiable with derivative $f_Z(t)=\mathrm{P}(N(\mathbf{X}\cap C_{n,0})=\lfloor t\rfloor)$, there exists $\tilde{M}\in[\mathrm{Me}_Z\wedge\lambda c_n,\mathrm{Me}_Z\vee\lambda c_n]$ such that

$$(\mathrm{Me}_Z-\lambda c_n)\mathrm{P}(N(\mathbf{X}\cap C_{n,0})=\lfloor\tilde{M}\rfloor)=o(1) \tag{25}$$

Since $\mathrm{Me}_Z = \lambda c_n + \mathcal{O}(\sqrt{c_n})$, there exists $\omega > 0$ such that for n sufficiently large, $|\tilde{M} - \lambda c_n| \leq \omega\sqrt{c_n}$. Using (23) and the fact that the mode of a Poisson distribution is close to its intensity parameter, we have for n sufficiently large

$$\begin{aligned}\mathrm{P}(\Pi(\lambda c_n) = \lfloor \tilde{M} \rfloor) &\geq \inf_{v_n, |v_n - \lambda c_n| \leq \omega\sqrt{c_n}} \mathrm{P}(\Pi(\lambda c_n) = \lfloor v_n \rfloor) \\ &\geq \frac{1}{2}(2\pi\lambda c_n)^{-1/2} e^{-\omega^2/2} \\ &\geq \frac{1}{4} e^{-\omega^2/2} \mathrm{P}(\Pi(\lambda c_n) = \lfloor \lambda c_n \rfloor).\end{aligned}$$

Hence, denoting by $\tilde{\omega} = e^{-\omega^2/2}/4 > 0$, we deduce using (22) that for n sufficiently large

$$\begin{aligned}\sqrt{c_n}\mathrm{P}(N(\mathbf{X} \cap C_{n,0}) = \lfloor \tilde{M} \rfloor) &\geq \frac{1}{2}\sqrt{c_n}\mathrm{P}(N(\mathbf{X} \cap C_{n,0}) = \lfloor \lambda c_n \rfloor) \frac{\mathrm{P}(\Pi(\lambda c_n) = \lfloor \tilde{M} \rfloor)}{\mathrm{P}(\Pi(\lambda c_n) = \lfloor \lambda c_n \rfloor)} \\ &\geq \frac{\tilde{\omega}}{2} \liminf_{n\to\infty} s_n =: \breve{\omega} > 0\end{aligned}$$

by Assumption $\mathcal{A}_{\mathrm{med}}$. Hence, from (25), $\breve{\omega}(\mathrm{Me}_Z - \lambda c_n)/\sqrt{c_n} = o(1)$, which yields the result.

(c) From (a)-(b), taking $k_n = o(\sqrt{c_n}/(\mathrm{Me}_Z - \lambda c_n))$ is in agreement with $\mathcal{A}_{W_n}$. Equation (14) is thus a simple application of Slutsky's lemma.

Acknowledgements

The authors would like to thank Denis Allard and Alfred Stein for giving them the opportunity to contribute to this special issue. The authors are also grateful to the referees and the editor for their valuable comments and suggestions and to Frédéric Lavancier for fruitful discussions. The research of J.-F. Coeurjolly is funded by ANR-11-LABX-0025 PERSYVAL-Lab (2011, project OculoNimbus). The research of Christophe A.N. Biscio are supported by the Danish Council for Independent Research | Natural Sciences, grant 12-124675, "Mathematical and Statistical Analysis of Spatial Data", and by the "Centre for Stochastic Geometry and Advanced Bioimaging", funded by grant 8721 from the Villum Foundation.

References

A. Baddeley, E. Rubak, and R. Turner. *Spatial Point Patterns: Methodology and Applications with R*. CRC Press, 2015.

C.A.N. Biscio and F. Lavancier. Quantifying repulsiveness of determinantal point processes. *to appear in Bernoulli*, 2016a. http://arxiv.org/abs/1406.2796.

C.A.N. Biscio and F. Lavancier. Brillinger mixing of determinantal point processes and statistical applications. *Electronic Journal of Statistics*, 10(1):582–607, 2016b.

J.-F. Coeurjolly. Median-based estimation of the intensity of a spatial point process. *to appear in Annals of the Institute of Statistical Mathematics*, 2016. http://arxiv.org/abs/1410.1448.

D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods*. Springer-Verlag, New York, second edition, 2003.

D.J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes II: General theory and structure*. Springer Verlag, New York, 2008.

H.A. David and H.N. Nagaraja. *Order statistics (3rd ed.)*. Hoboken: Wiley, New Jersey, 2003.

R.M. Dudley. *Real Analysis and Probability*, volume 74. Cambridge University Press, 2002.

L. Heinrich and M. Prokešová. On estimating the asymptotic variance of stationary point processes. *Methodology and Computing in Applied Probability*, 12(3):451–471, 2010.

J.B. Hough, M. Krishnapur, Y. Peres, and B. Virág. *Zeros of Gaussian Analytic Functions and Determinantal Point Processes*, volume 51 of *University Lecture Series*. American Mathematical Society, Providence, RI, 2009.

J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan. *Statistical Analysis and Modelling of Spatial Point Patterns*. Statistics in Practice. Wiley, Chichester, 2008.

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

F. Lavancier, J. Møller, and E. Rubak. Determinantal point process models and statistical inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77 (4):853–877, 2015.

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

J.A.F. Machado and J.M.C. Santos Silva. Quantiles for counts. *Journal of the American Statistical Association*, 100(472):1226–1237, 2005.

J. Møller and R. P. Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton, 2004.

D.N. Politis, E. Paparoditis, and J.P. Romano. Large sample inference for irregularly spaced dependent observations based on subsampling. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 274–292, 1998.

C. Redenbach, A. Särkkä, and M. Sormani. Classification of points in superpositions of strauss and poisson processes. *Spatial Statistics*, 12:81–95, 2015.

F. Riesz and B.S. Nagy. *Functional Analysis*. Dover Books on Mathematics Series. Dover Publications, 1990.

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

A. Soshnikov. Gaussian limit for determinantal random point fields. *The Annals of Probability*, 30(1):171–187, 2002.

E. Stein and G. Weiss. *Introduction to Fourier Analysis on Euclidean Spaces (PMS-32)*, volume 1. Princeton university press, 1971.

D. Stoyan, W. S. Kendall, and J. Mecke. *Stochastic Geometry and Its Applications*. Wiley, Chichester, second edition, 1995.

V. Zacharovas and H.-K. Hwang. A Charlier–Parseval approach to Poisson approximation and its applications. *Lithuanian Mathematical Journal*, 50(1):88–119, 2010.