HAL open science

Mixing properties and central limit theorem for associated point processes

Arnaud Poinas, Bernard Delyon, Frédéric Lavancier

To cite this version:

Arnaud Poinas, Bernard Delyon, Frédéric Lavancier. Mixing properties and central limit theorem for associated point processes. 2017. hal-01519096v2

HAL Id: hal-01519096

https://hal.science/hal-01519096v2

Preprint submitted on 27 Sep 2017 (v2), last revised 19 Feb 2018 (v3)

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

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

Mixing properties and central limit theorem for associated point processes

ARNAUD POINAS[1,*] BERNARD DELYON[1,**] FRÉDÉRIC LAVANCIER[2,†]

[1] *IRMAR – Campus de Beaulieu - Bat. 22/23 – 263 avenue du Général Leclerc – 35042 Rennes – France*
E-mail: [*]arnaud.poinas@univ-rennes1.fr
E-mail: [**]bernard.delyon@univ-rennes1.fr

[2] *Laboratoire de Mathématiques Jean Leray – BP 92208 – 2, Rue de la Houssinière – F-44322 Nantes Cedex 03 – France. Inria, Centre Rennes Bretagne Atlantique, France.*
E-mail: [†]Frederic.Lavancier@univ-nantes.fr

Positively (resp. negatively) associated point processes are a class of point processes that induce attraction (resp. inhibition) between the points. As an important example, determinantal point processes (DPPs) are negatively associated. We prove α-mixing properties for associated spatial point processes by controlling their α-coefficients in terms of the first two intensity functions. A central limit theorem for functionals of associated point processes is deduced, using both the association and the α-mixing properties. We discuss in detail the case of DPPs, for which we obtain the limiting distribution of sums, over subsets of close enough points of the process, of any bounded function of the DPP. As an application, we get the asymptotic properties of the parametric two-step estimator of some inhomogeneous DPPs.

Keywords: determinantal point process, parametric estimation, strong mixing, negative association, positive association.

1. Introduction

Positive association (PA) and negative association (NA) [1, 18] are properties that quantify the dependence between random variables. They have found many applications in limit theorems for random fields [9, 39]. Even if the extension of PA to point processes have been used in analysis of functionals of random measures [10, 19], there are no general applications of PA or NA to limit theorems for point processes. We contribute in this paper to this aspect for spatial point processes on $\mathbb{R}^d$. We especially discuss in detail the case of determinantal point processes (DPPs for short), that are an important example of negatively associated point processes. DPPs are a type of repulsive point processes that were first introduced by Macchi [31] in 1975 to model systems of fermions in the context of quantum mechanics. They have been extensively studied in Probability theory with applications ranging from random matrix theory to non-intersecting random walks, random spanning trees and more (see [25]). From a statistical perspective, DPPs have applications in machine learning [28], telecommunication [14, 32, 21], biology, forestry [29] and computational statistics [2].

As a first result, we relate the association property of a point process to its α-mixing properties. First introduced in [35], α-mixing is a measure of dependence between random variables, which is actually more popular than PA or NA. It has been used extensively to prove central limit theorems for dependent random variables [6, 15, 22, 26, 35]. More details about mixing can be found in [7, 15]. We derive in Section 2 an important covariance inequality for associated point processes (Theorem 2.5), that turns out to be very similar to inequalities established in [17] for weakly dependent continuous random processes. We show that this inequality implies α-mixing and precisely allows to control the α-mixing coefficients by the first two intensity functions of the point process. This result for point processes is in contrast with the case of random fields where it is known that association does not imply α-mixing in general (see Examples 5.10-5.11 in [9]). However, this implication holds true for integer-valued random fields (see [16] or [9]). As explained in [16], this is because the σ-algebras generated by countable sets are much poorer than σ-algebras generated by continuous sets. In fact, by this aspect and some others (for instance our proofs boil down to the control of the number of points in bounded sets), point processes are very similar to discrete processes.

We then establish in Section 3 a general central limit theorem (CLT) for random fields defined as a function of an associated point process (Theorem 3.1). A standard method for proving this kind of theorems is to rely on sufficiently fast decaying α-mixing coefficients along with some moment assumptions. We use an alternative procedure that exploits both the mixing properties and the association property. This results in weaker assumptions on the underlying point process, that can have slower decaying mixing coefficients. This improvement allows in particular to include all standard DPPs, some of them being otherwise excluded with the first approach (like for instance DPPs associated to the Bessel-type kernels [4]).

Section 4 discusses in detail the case of DPPs, where we derive a tight explicit bound for their α-mixing coefficients and prove a central limit theorem for certain functionals of a DPP (Theorem 4.4). Specifically, these functionals write as a sum of a bounded function of the DPP, over subsets of close-enough points of the DPP. A particular case concerns sums over p-tuple of close enough points of the DPP, which are frequently encountered in asymptotic inference. Limit theorems in this setting have been established in [37] when $p = 1$, and in [3] for stationary DPPs and $p \geqslant 1$. We thus extend these studies to sums over any subsets and without the stationary assumption. As a statistical application, we consider the parametric estimation of second-order intensity reweighted stationary DPPs. These DPPs have an inhomogeneous first order intensity, but translation-invariant higher order (reweighted) intensities. We prove that the two-step estimator introduced in [38], designed for this kind of inhomogeneous point process models, is consistent and asymptotically normal when applied to DPPs.

2. Associated point processes and α-mixing

2.1. Notation

In this paper, we consider locally finite simple point processes on $\mathbb{R}^d$, for a fixed $d \in \mathbb{N}$. Some theoretical background on point processes can be found in [11, 33]. We denote by Ω the set of locally finite point configurations in $\mathbb{R}^d$. For $X \in \Omega$ and $A \subset \mathbb{R}^d$, we write

$$N(A) := \mathrm{card}(X \cap A)$$

for the random variable representing the number of point of X that fall in A. We also denote by $\mathscr{B}(A)$ the Borel σ-algebra of A and by $\mathcal{E}(A)$ the σ-algebra generated by $X \cap A$, defined by

$$\mathcal{E}(A) := \sigma(\{X \in \Omega : N(B) = m\}, B \in \mathscr{B}(A), m \in \mathbb{N}).$$

The notation $|.|$ will have a different meaning depending on the object it is applied. For $x \in \mathbb{R}^d$, $|x|$ stands for the euclidean norm. For a set $A \subset \mathbb{R}^d$, $|A| := \int_A \mathrm{d}x$ is the euclidean volume of A, and for a set $I \subset \mathbb{Z}^d$ we write $|I|$ for the cardinal of I. For A, B two subsets of $\mathbb{R}^d$ (resp. $\mathbb{Z}^d$) we define $\mathrm{dist}(A, B)$ as $\inf_{x\in A, y\in B} |y-x|$ and $\mathrm{diam}(A)$ as $\sup_{x,y\in A} |y-x|$ where $|.|$ is the associated norm on $\mathbb{R}^d$ (resp. $\mathbb{Z}^d$). For $i \in \mathbb{Z}^d$, $|i|_1$ denotes the ℓ_1-norm. Finally, we write $\mathcal{B}(x, r)$ for the euclidean ball centred at x with radius r and $\|.\|_p$ for the p-norm of random variables and functions where $1 \leqslant p \leqslant \infty$.

We recall that the intensity functions of a point process (when they exist), with respect to the Lebesgue measure, are defined as follows.

Definition 2.1. *Let $X \in \Omega$ and $n \geqslant 1$ be an integer. If there exists a non-negative function $\rho_n : (\mathbb{R}^d)^n \to \mathbb{R}$ such that*

$$\mathbb{E}\left[\sum_{x_1,\cdots,x_n\in X}^{\neq} f(x_1, \cdots, x_n)\right] = \int_{(\mathbb{R}^d)^n} f(x_1, \cdots, x_n)\rho_n(x_1, \cdots, x_n)\mathrm{d}x_1 \cdots \mathrm{d}x_n \tag{2.1}$$

for all locally integrable function $f : (\mathbb{R}^d)^n \to \mathbb{R}$ then ρ_n is called the nth order intensity function of X.

In particular, $\rho_n(x_1, \cdots, x_n)\mathrm{d}x_1 \cdots \mathrm{d}x_n$ can be viewed as the probability that X has a point in each of the infinitesimally small sets around $x_1, \cdots, x_n$ with volumes $\mathrm{d}x_1, \cdots, \mathrm{d}x_n$ respectively.

We further introduce the notation

$$D(x,y) := \rho_2(x,y) - \rho_1(x)\rho_1(y). \tag{2.2}$$

This quantity is involved in the following equality, deduced from the previous definition and used several times throughout the paper:

$$\mathrm{Cov}(N(A), N(B)) = \int_{A\times B} D(x,y)\mathrm{d}x\mathrm{d}y. \tag{2.3}$$

2.2. Negative and positive association

Our goal in this section is to prove a crucial covariance inequality and to deduce an α-mixing property for associated point processes. We recall that associated point processes are defined the following way (see Definitions 2.11-2.12 in [5] for example).

Definition 2.2. *A point process X is said to be negatively associated (NA for short) if, for all family of pairwise disjoint Borel sets $(A_i)_{1\leqslant i\leqslant k}$ and $(B_i)_{1\leqslant i\leqslant l}$ such that*

$$(\cup_i A_i) \cap (\cup_j B_j) = \varnothing \tag{2.4}$$

and for all coordinate-wise increasing functions $F : \mathbb{N}^k \mapsto \mathbb{R}$ and $G : \mathbb{N}^l \mapsto \mathbb{R}$ it satisfies

$$\mathbb{E}[F(N(A_1),\cdots,N(A_k))G(N(B_1),\cdots,N(B_l))] \\ \leqslant \mathbb{E}[F(N(A_1),\cdots,N(A_k))]\mathbb{E}[G(N(B_1),\cdots,N(B_l))]. \tag{2.5}$$

Similarly, a point process is said to be positively associated (PA for short) if it satisfies the reverse inequality for all family of pairwise disjoint Borel sets $(A_i)_{1\leqslant i\leqslant k}$ and $(B_i)_{1\leqslant i\leqslant l}$ (but not necessarily satisfying (2.4)). If a point process is NA or PA it is said to be associated.

The main difference between the definition of PA and NA is the restriction (2.4) that only affects NA point processes. Notice that without (2.4), $\mathbb{E}[N(A)]^2 \geqslant \mathbb{E}[N(A)^2]$ contradicts (2.5) hence the need to consider functions depending on disjoint sets for NA point processes.

These inequalities extend to the more general case of functionals of point processes. The first thing we need is a more general notion of increasing functions. We associate to Ω the partial order $X \leqslant Y$ iff $X \subset Y$ and then call a function on Ω increasing if it is increasing respective to this partial order. The association property can then be extended to these functions. A proof in a general setting can be found in [30, Lemma 3.6] but we give an alternative elementary one in Appendix A.

Theorem 2.3. *Let X be a NA point process on $\mathbb{R}^d$ and A, B be disjoint subsets of $\mathbb{R}^d$. Let $F : \Omega \mapsto \mathbb{R}$ and $G : \Omega \mapsto \mathbb{R}$ be bounded increasing functions, then*

$$\mathbb{E}[F(X\cap A)G(X\cap B)] \leqslant \mathbb{E}[F(X\cap A)]\mathbb{E}[G(X\cap B)]. \tag{2.6}$$

If X is PA then, for all $A, B \subset \mathbb{R}^d$ non necessarily disjoint,

$$\mathbb{E}[F(X\cap A)G(X\cap B)] \geqslant \mathbb{E}[F(X\cap A)]\mathbb{E}[G(X\cap B)]. \tag{2.7}$$

Association is a very strong dependence condition. As proved in the following theorem, it implies a strong covariance inequality that is only controlled by the behaviour of the first two intensity functions of X (assuming their existence). To state this result, we need to introduce the following seminorm for functionals over point processes.

Definition 2.4. *For any $A \subset \mathbb{R}^d$, $\|.\|_A$ is the seminorm on the functions $f : \Omega \mapsto \mathbb{C}$ defined by*

$$\|f\|_A := \sup_{\substack{X\in\Omega, X\subset A\\ x\in A}} |f(X) - f(X\cup\{x\})|.$$

Note that $\|.\|_A$ is a Lipschitz norm in the sense that it controls the way $f(X)$ changes when a point is added to $X \cap A$.

Theorem 2.5. *Let X be an associated point process and $A, B \subset \mathbb{R}^d$ be two disjoint bounded subsets. Let $f : \Omega \to \mathbb{R}$ and $g : \Omega \to \mathbb{R}$ be two functions such that $f(X \cap A)$ and $g(X \cap B)$ are bounded, then*

$$|Cov(f(X \cap A), g(X \cap B))| \leqslant \|f\|_A \|g\|_B |Cov(N(A), N(B))|. \tag{2.8}$$

Moreover, if X is PA then it also satisfies the same inequality for all $A, B \subset \mathbb{R}^d$ not necessarily disjoint.

Proof. The proof mimics the one from [8] for associated random fields. We only consider the case of NA point processes but the PA case can be treated in the same way.
Consider the functions $f_+, f_- : \Omega \to \mathbb{R}$, $\mathcal{E}(A)$-measurable, and $g_+, g_- : \Omega \to \mathbb{R}$, $\mathcal{E}(B)$-measurable, defined by

$$\begin{cases} f_\pm(X) = f(X \cap A) \pm \|f\|_A N(A), \\ g_\pm(X) = g(X \cap B) \pm \|g\|_B N(B). \end{cases}$$

For all $x \in A \backslash X$, $f_+(X \cup \{x\}) - f_+(X) = f(X \cup \{x\} \cap A) - f(X \cap A) + \|f\|_A$ which is positive by definition of $\|f\|_A$. f_+ is thus an increasing function. With the same reasoning, g_+ is also increasing and f_-, g_- are decreasing. f_+ is not bounded but it is non-negative and almost surely finite so it can be seen as an increasing limit of the sequence of functions $\min(f_+, k)$ when k goes to infinity. These functions are non-negative, increasing and bounded so for any k and any bounded increasing function g, (2.6) applies where f is replaced by $\min(f_+, k)$. By a limiting argument, the same inequality holds true for $f = f_+$. We can also treat the other functions the same way and we get from (2.6)

$$\mathrm{Cov}(f_+(X), g_+(X)) \leqslant 0 \quad \text{and} \quad \mathrm{Cov}(f_-(X), g_-(X)) \leqslant 0.$$

Since these expressions are equal to

$$\begin{aligned} \mathrm{Cov}(f_\pm(X), g_\pm(X)) = \mathrm{Cov}(f(X \cap A), g(X \cap B)) + \|f\|_A \|g\|_B \mathrm{Cov}(N(A), N(B)) \\ \pm (\|g\|_B \mathrm{Cov}(f(X \cap A), N(B)) + \|f\|_A \mathrm{Cov}(N(A), g(X \cap B))), \end{aligned}$$

adding these two expressions together yields the upper bound in (2.8):

$$\mathrm{Cov}(f(X \cap A), g(X \cap B)) \leqslant -\|f\|_A \|g\|_B \mathrm{Cov}(N(A), N(B)).$$

The lower bound is obtained by replacing f by $-f$ in the previous expression. □

A similar inequality as in Theorem 2.5 can also be obtained for complex-valued functions since $\|\Re(f)\|_A$ and $\|\Im(f)\|_A$ are bounded by $\|f\|_A$, where $\Re(f)$ and $\Im(f)$ refers to the real and imaginary part of f respectively.

Corollary 2.6. *Let X be an associated point process and $A, B \subset \mathbb{R}^d$ be two disjoint bounded subsets. Let $f : \Omega \to \mathbb{C}$ and $g : \Omega \to \mathbb{C}$ be two functions such that $f(X \cap A)$ and $g(X \cap B)$ is bounded, then*

$$|Cov(f(X \cap A), g(X \cap B))| \leqslant 4\|f\|_A \|g\|_B |Cov(N(A), N(B))|.$$

Moreover, if X is PA then it also satisfies the same inequality for all $A, B \subset \mathbb{R}^d$ not necessarily disjoint.

If the first two intensity functions of X are well-defined then D in (2.2) is well-defined. As a consequence of Theorem 2.5 and from (2.3), if $|D(x, y)|$ vanishes fast enough when $|y - x|$ goes to infinity then any two events respectively in $\mathcal{E}(A)$ and $\mathcal{E}(B)$ will get closer to independence as $\mathrm{dist}(A, B)$ tends to infinity, as specified by the following corollary.

Corollary 2.7. *Let X be an associated point process on $\mathbb{R}^d$ whose first two intensity functions are well-defined. Let A, B be two bounded disjoint sets of $\mathbb{R}^d$ such that $dist(A, B) \geqslant r$. Then, for all functions $f : \Omega \to \mathbb{R}$ and $g : \Omega \to \mathbb{R}$ such that $f(X \cap A)$ and $g(X \cap B)$ is bounded,*

$$|Cov(f(X \cap A), g(X \cap B))| \leqslant s_d |A| \, \|f\|_A \|g\|_B \int_r^\infty t^{d-1} \sup_{|x-y|=t} |D(x, y)| \mathrm{d}t. \tag{2.9}$$

where s_d is the $(d-1)$-dimensional area measure of the unit sphere in $\mathbb{R}^d$. Moreover, if f and/or g are complex-valued functions, the same inequality holds true with an extra factor 4 on the right hand side.

Proof. Consider A, B to be two bounded disjoint sets of $\mathbb{R}^d$ such that $\mathrm{dist}(A, B) \geqslant r$ then, from (2.3),

$$\begin{aligned}
|\mathrm{Cov}(N(A), N(B)| &= \int_{A \times B} |D(x, y)| \mathrm{d}x \mathrm{d}y \\
&\leqslant |A| \sup_{x \in A} \int_B |D(x, y)| \mathrm{d}y \\
&\leqslant |A| \sup_{x \in A} \int_{\mathcal{B}(x,r)^c} |D(x, y)| \mathrm{d}y \\
&\leqslant |A| \sup_{x \in A} \int_{\mathcal{B}(x,r)^c} \sup_{\substack{u \in \mathbb{R}^d \\ |u-x|=|y-x|}} |D(x, u)| \mathrm{d}y \\
&\leqslant |A| s_d \int_r^\infty t^{d-1} \sup_{|u-v|=t} |D(u, v)| \mathrm{d}t.
\end{aligned}$$

The final result is then a consequence of Theorem 2.5 and Corollary 2.6. □

2.3. Application to α-mixing

Let us first recall some generalities about mixing. Consider a probability space $(\mathcal{X}, \mathcal{F}, \mathbb{P})$ and $\mathscr{A}, \mathscr{B}$ two sub σ-algebra of $\mathcal{F}$. The α-mixing coefficient is defined as the following measure of dependence between $\mathscr{A}$ and $\mathscr{B}$:

$$\alpha(\mathscr{A}, \mathscr{B}) := \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathscr{A}, B \in \mathscr{B}\}.$$

In particular, $\mathscr{A}$ and $\mathscr{B}$ are independent iff $\alpha(\mathscr{A}, \mathscr{B}) = 0$. This definition leads to the essential covariance inequality due to Davydov [12] and later generalised by Rio [34]: For all random variables X, Y measurable with respect to $\mathscr{A}$ and $\mathscr{B}$ respectively,

$$|\mathrm{Cov}(X, Y)| \leqslant 8\alpha^{1/r}(\mathscr{A}, \mathscr{B}) \|X\|_p \|Y\|_q, \quad \text{where } p, q, r \in [1, \infty] \text{ and } \frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1. \tag{2.10}$$

This definition is adapted to random fields the following way (see [15] or [22]). Let $Y = (Y_i)_{i \in \mathbb{Z}^d}$ be a random fields on $\mathbb{Z}^d$ and define

$$\alpha_{p,q}(r) := \sup\{\alpha(\sigma(\{Y_i, i \in A\}), \sigma(\{Y_i, i \in B\})) : |A| \leqslant p, |B| \leqslant q, \mathrm{dist}(A, B) > r\}$$

with the convention $\alpha_{p,\infty}(r) = \sup_q \alpha_{p,q}(r)$. The coefficients $\alpha_{p,q}(r)$ describe how close two events happening far enough from each other are from being independent. The parameters p and q play an important role since, in general, we cannot get any information directly on the behaviour of $\alpha_{\infty,\infty}(r)$.

We can adapt this definition to point processes the following way. For a point process X on $\mathbb{R}^d$, define

$$\alpha_{p,q}(r) := \sup\{\alpha(\mathcal{E}(A), \mathcal{E}(B)) : |A| \leqslant p, |B| \leqslant q, \mathrm{dist}(A, B) > r\}$$

with the convention $\alpha_{p,\infty}(r) = \sup_q \alpha_{p,q}(r)$.

As a consequence of Corollary 2.7, the α-mixing coefficients of an associated point process tends to 0 when $D(x, y)$ vanishes fast enough as $|y - x|$ goes to infinity. More precisely, we have the following inequalities.

Proposition 2.8. *Let X be an associated point process on $\mathbb{R}^d$ whose first two intensity functions are well-defined, then for all $p, q > 0$,*

$$\begin{cases} \alpha_{p,q}(r) \leqslant pq \sup\limits_{|x-y|\geqslant r} |D(x,y)|, \\ \alpha_{p,\infty}(r) \leqslant ps_d \displaystyle\int_r^\infty t^{d-1} \sup\limits_{|x-y|=t} |D(x,y)| \mathrm{d}t. \end{cases} \tag{2.11}$$

Proof. We can write

$$\alpha(\mathcal{E}(A), \mathcal{E}(B)) = \sup_{\substack{\mathscr{A}\in\mathcal{E}(A) \\ \mathscr{B}\in\mathcal{E}(B)}} \mathrm{Cov}(\mathbb{1}_{\mathscr{A}}(X\cap A), \mathbb{1}_{\mathscr{B}}(X\cap B))$$

so Proposition 2.8 is a direct consequence of Theorem 2.5 and Corollary 2.7 applied to indicator functions. □

3. Central limit theorem for associated point processes

Consider the lattice $(x_i)_{i\in\mathbb{Z}^d}$ defined by $x_i = R\cdot i$, where $R > 0$ is a fixed constant. We denote by C_i, $i \in \mathbb{Z}^d$, the d-dimensional cube with centre x_i and side length s, where $s \geqslant R$ is another fixed constant. Note that the union of these cubes forms a covering of $\mathbb{R}^d$. Let X be an associated point process and $(f_i)_{i\in\mathbb{Z}^d}$ be a family of real-valued measurable functions defined on Ω. We consider the centred random field $(Y_i)_{i\in\mathbb{Z}^d}$ defined by

$$Y_i := f_i(X\cap C_i) - \mathbb{E}[f_i(X\cap C_i)], \quad i\in\mathbb{Z}^d, \tag{3.1}$$

and we are interested in this section by the asymptotic behavior of $S_n := \sum_{i\in I_n} Y_i$, where $(I_n)_{n\in\mathbb{N}}$ is a sequence of strictly increasing finite domains of $\mathbb{Z}^d$.

As a consequence of Proposition 2.8, we could directly use one of the different CLT for α-mixing random fields that already exist in the literature [6, 15, 22] to get the asymptotic distribution of S_n. But, the coefficients $\alpha_{p,\infty}$ decreasing much slower than the coefficients $\alpha_{p,q}$, this would imply an unnecessary strong assumption on D. Precisely, this would require $D(x,y)$ to decay at a rate at least $o(|y-x|^{-2(d+\varepsilon)\frac{2+\delta}{\delta}})$, where $\varepsilon > 0$ and δ is a positive constant depending on the behaviour of the moments of X. In the next theorem, we bypass this issue by exploiting both the behaviour of the mixing coefficients $\alpha_{p,q}$ when $p < \infty$ and $q < \infty$, and the association property through inequality (2.9). We show that we can still get a CLT when $D(x,y)$ decays at a rate $o(|y-x|^{-(d+\varepsilon)\frac{2+\delta}{\delta}})$. This improvement is important to include DPPs with a slow decaying kernel, thus inducing more repulsiveness, such as Bessel-type kernels, see the applications to DPPs in Section 4.2 and especially the discussion at the end of the section. Let us also remark that another technique, based on the convergence of moments, is sometimes used to establish a CLT for point processes. This has been exploited especially for Brillinger mixing point processes in [27, 24] and other papers. As an example, DPPs have been proved to be Brillinger mixing in [3, 23]. However, this condition applies to stationary point processes only.

Theorem 3.1. *Consider the random field Y given by (3.1), a sequence $(I_n)_{n\in\mathbb{N}}$ of strictly increasing finite domains of $\mathbb{Z}^d$ and $S_n = \sum_{i\in I_n} Y_i$. Let $\sigma_n^2 := Var(S_n)$. Assume that for some $\varepsilon, \delta > 0$ the following conditions are satisfied:*

(C1) X is an associated point process on $\mathbb{R}^d$ whose first two intensity functions are well-defined;
(C2) $\sup_{i\in\mathbb{Z}^d} \|Y_i\|_{2+\delta} = M < \infty$;
(C3) $\sup_{|x-y|\geqslant r} |D(x,y)| = \underset{r\to\infty}{o}(r^{-(d+\varepsilon)\frac{2+\delta}{\delta}})$ where D is given by (2.2);
(C4) $\liminf_n |I_n|^{-1}\sigma_n^2 > 0$.

Then

$$\frac{1}{\sigma_n} S_n \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

Proof. First, we notice that Y inherits its strong mixing coefficients from X. This is due to the fact that we have $\sigma(\{Y_i : i \in I\}) \subset \mathcal{E}(\bigcup_{i\in I} C_i)$ for all $I \subset \mathbb{Z}^d$ as a consequence of (3.1). Moreover, we have $\mathrm{dist}(C_i, C_j) \geqslant \frac{1}{\sqrt{d}}(|i-j|_1 R - sd)$ as a consequence of Lemma B.2, and since $|\bigcup_{i\in I} C_i| \leqslant s^d|I|$, this gives us the inequality

$$\forall p, q > 0,\ \forall r > \frac{sd}{R},\quad \alpha^Y_{p,q}(r) \leqslant \alpha^X_{ps^d, qs^d}\left(\frac{1}{\sqrt{d}}(rR - sd)\right),$$

where we denote by α^X, α^Y the α-mixing coefficients of X and Y respectively. In particular, conditions (C1), (C3) and identity (2.11) yields

$$\forall p, q > 0,\ \alpha^Y_{p,q}(r) = \underset{r\to\infty}{o}(r^{-(d+\varepsilon)\frac{2+\delta}{\delta}}). \tag{3.2}$$

We deal with the proof in two steps: first, we consider the case of bounded variables and then we extend the result to the more general case.

The first step of the proof follows the approach used by Bolthausen [6] and Guyon [22], while the second step exploits elements from [26]. The main difference lies in the way we deal with the term A_3 that appears later on in the proof.

First step: Bounded variables. Without loss of generality, we consider that $\mathbb{E}[f_i(X \cap C_i)] = 0$ for all $i \in \mathbb{Z}^d$. Suppose that we have $\sup_i \|Y_i\|_\infty := \sup_i \|f_i(. \cap C_i)\|_\infty = M < \infty$ instead of Assumption (C2). Since $\alpha^Y_{p,q}(r)$ is non increasing in r and is a $o(r^{-d})$ by (3.2), we can choose a sequence $(r_n)_{n\in\mathbb{N}}$ such that

$$\alpha^Y_{p,q}(r_n)\sqrt{|I_n|} \to 0 \ \text{ and } \ r_n^{-d}\sqrt{|I_n|} \to \infty.$$

For $i \in \mathbb{Z}^d$, define

$$S_{i,n} = \sum_{\substack{j\in I_n \\ |i-j|_1 \leqslant r_n}} Y_j, \quad S^*_{i,n} = S_n - S_{i,n}, \quad a_n = \sum_{i\in I_n} \mathbb{E}[Y_i S_{i,n}], \quad \bar{S}_n = \frac{1}{\sqrt{a_n}} S_n, \quad \bar{S}_{i,n} = \frac{1}{\sqrt{a_n}} S_{i,n}.$$

We have $\sigma_n^2 = \mathrm{Var}(S_n) = a_n + \sum_{i\in I_n} \mathbb{E}[Y_i S^*_{i,n}]$ and, as a consequence of the typical covariance inequality (2.10) for α-mixing random variables, we get

$$\left|\sum_{i\in I_n} \mathbb{E}[Y_i S^*_{i,n}]\right| \leqslant \sum_{\substack{i,j\in I_n \\ |i-j|_1 > r_n}} |\mathrm{Cov}(Y_i, Y_j)| \leqslant 8M^2 \sum_{\substack{i,j\in I_n \\ |i-j|_1 > r_n}} \alpha^Y_{1,1}(|i-j|_1) \leqslant 8M^2|I_n| \sum_{r > r_n} |\{k \in \mathbb{Z}^d : |k|_1 = r\}|\alpha^Y_{1,1}(r).$$

The number of $k \in \mathbb{Z}^d$ satisfying $|k|_1 = r$ is bounded by $2(2r+1)^{d-1}$. This is because each of the $d-1$ first coordinates of k takes their values in $\{-r, \cdots, r\}$ and the last coordinate is fixed by the other ones, up to the sign, since $|k|_1 = r$. Therefore,

$$\left|\sum_{i\in I_n} \mathbb{E}[Y_i S^*_{i,n}]\right| \leqslant 16M^2|I_n| \sum_{r > r_n} (2r+1)^{d-1}\alpha^Y_{1,1}(r).$$

By Assumption (3.2), this quantity is $o(|I_n|)$ and thus $\sigma_n^2 \sim a_n$ as a consequence of Assumption (C4). We then only need to prove the asymptotic normality of $\bar{S}_n$. Moreover, since $\sup_n \mathbb{E}[\bar{S}_n^2] < \infty$ then, by Stein's Lemma, this will be a consequence of the condition

$$\lim_{n\to\infty} \mathbb{E}[(i\lambda - \bar{S}_n)\exp(i\lambda\bar{S}_n)] = 0, \quad \forall \lambda \in \mathbb{R}.$$

We can split this expression into $(i\lambda - \bar{S}_n)\exp(i\lambda\bar{S}_n) = A_1 - A_2 - A_3$ where

$$\begin{cases} A_1 = i\lambda\exp(i\lambda\bar{S}_n)\left(1 - \dfrac{1}{a_n}\displaystyle\sum_{j\in I_n} Y_j S_{j,n}\right), \\ A_2 = \dfrac{1}{\sqrt{a_n}}\exp(i\lambda\bar{S}_n)\displaystyle\sum_{j\in I_n} Y_j\left(1 - i\lambda\bar{S}_{j,n} - \exp(-i\lambda\bar{S}_{j,n})\right), \\ A_3 = \dfrac{1}{\sqrt{a_n}}\displaystyle\sum_{j\in I_n} Y_j \exp\left(i\lambda(\bar{S}_n - \bar{S}_{j,n})\right). \end{cases}$$

It was proved by Bolthausen [6] that $\mathbb{E}[A_1^2]$ and $\mathbb{E}[|A_2|]$ vanish when n goes to infinity if $\sum r^{d-1}\alpha^Y_{p,q}(r) < \infty$ for $p+q \leqslant 4$ which is the case here. We show that $\mathbb{E}[A_3]$ vanishes at infinity using (2.9). Notice that we have

$$|\mathbb{E}[A_3]| \leqslant \frac{|I_n|}{\sqrt{a_n}} \sup_{j\in I_n} \left| \mathrm{Cov}\left(f_j(X\cap C_j), \exp\left(\frac{i\lambda}{\sqrt{a_n}} \sum_{\substack{k\in I_n \\ |k-j|_1 > r_n}} f_k(X\cap C_k) \right)\right)\right|.$$

Define the function

$$g_j : X \mapsto \exp\left(\frac{i\lambda}{\sqrt{a_n}} \sum_{\substack{k\in I_n \\ |k-j|_1 > r_n}} f_k(X\cap C_k) \right).$$

This function is bounded by 1 and $\mathcal{E}(B_j)$-measurable where $B_j := \bigcup_{k\in I_n,\, |k-j|_1>r_n} C_k$ is a bounded set and $\mathrm{dist}(C_j, B_j) \geqslant (Rr_n - sd)/\sqrt{d}$ (see Lemma B.2). We have $\|f_j\|_{C_j} \leqslant 2M$ and for all $X \in \Omega$, for all $x \in B_j$, if we denote by $J_x = \{k : x \in C_k\}$ the set of cubes that contains x then

$$|g_j(X\cup\{x\}) - g_j(X)| = \left| 1 - \exp\left(\frac{i\lambda}{\sqrt{a_n}} \sum_{k\in J_x} (f_k(X\cap C_k\cup\{x\}) - f_k(X\cap C_k)) \right)\right| \leqslant \frac{2\lambda M |J_x|}{\sqrt{a_n}}.$$

Lemma B.2 gives us the bound $|J_x| \leqslant (2sd/R)^d$ and thus $\|g_j\|_{B_j} \leqslant \frac{2\lambda M(2sd)^d}{R^d\sqrt{a_n}}$. Finally, using Corollary 2.7 we get

$$\begin{aligned} |\mathbb{E}[A_3]| &\leqslant \frac{4|I_n|s_d}{\sqrt{a_n}} |C_j| \|f_j\|_{C_j} \|g_j\|_{B_j} \int_{\mathrm{dist}(B_j,C_j)}^{\infty} t^{d-1} \sup_{|x-y|\geqslant t} |D(x,y)| \mathrm{d}t \\ &\leqslant 16 s_d \lambda M^2 \left(\frac{2s^2 d}{R}\right)^d \frac{|I_n|}{a_n} \int_{\frac{1}{\sqrt{d}}(Rr_n - sd)}^{\infty} t^{d-1} \sup_{|x-y|\geqslant t} |D(x,y)| \mathrm{d}t. \end{aligned} \tag{3.3}$$

By assumption (C3) we have that $t^{d-1}\sup_{|x-y|\geqslant t}|D(x,y)|$ is integrable and by assumption (C4) we have $|I_n| = O(a_n)$ which shows that $\lim_{n\to\infty}\mathbb{E}[A_3] = 0$ concluding the proof of the theorem for bounded variables.

Second step: General Case. For $N > 0$, we define

$$\begin{cases} S_{1,n} := \sum_{i\in I_n} (F_N(Y_i) - \mathbb{E}[F_N(Y_i)]) \quad \text{where} \quad F_N : x \mapsto x\mathbb{1}_{|x|\leqslant N}, \\ S_{2,n} := \sum_{i\in I_n} (\widetilde{F}_N(Y_i) - \mathbb{E}[\widetilde{F}_N(Y_i)]) \quad \text{where} \quad \widetilde{F}_N : x \mapsto x\mathbb{1}_{|x|>N}. \end{cases}$$

Let $\sigma_n^2(N) := \mathrm{Var}(S_{1,n})$, from the first step of the proof we have $\sigma_n(N)^{-1}S_{1,n} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$. Let $1 > \gamma > (1 + \frac{\varepsilon}{d}(1+\frac{\delta}{2}))^{-1}$ and define $C_N := \sup_i \|Y_i\mathbb{1}_{|Y_i|>N}\|_{2+\delta\gamma}$. By assumption (C2) we have that C_N vanishes when $N \to \infty$ and by assumption (C4) we have that $|I_n| \leqslant c\sigma_n^2$ for a sufficiently large n, where c is a positive constant. By (2.10),

$$\begin{aligned} \frac{1}{\sigma_n^2}\mathrm{Var}(S_{2,n}) &= \frac{1}{\sigma_n^2} \sum_{i,j\in I_n} \mathrm{Cov}(\widetilde{F}_N(Y_i), \widetilde{F}_N(Y_j)) \\ &\leqslant \frac{|I_n|}{\sigma_n^2} C_N^2 \sup_{i\in I_n} \sum_{j\in I_n} 8\alpha^Y_{1,1}(|i-j|_1)^{\frac{\delta\gamma}{2+\delta\gamma}} \\ &\leqslant 16\, c\, C_N^2 \sum_{r=0}^{\infty} (2r+1)^{d-1} \alpha^Y_{1,1}(r)^{\frac{\delta\gamma}{2+\delta\gamma}}. \end{aligned}$$

By assumption (C3) and the choice of γ we have $\sum(2r+1)^{d-1}\alpha^Y_{1,1}(r)^{\frac{\delta\gamma}{2+\delta\gamma}} < \infty$ so $\sigma_n^{-1}S_{2,n}$ converges in mean square to 0 when N goes to infinity, uniformly in n. With the same reasoning, we also get the inequality

$$\frac{1}{\sigma_n^2}|\mathrm{Cov}(S_{1,n}, S_{2,n})| \leqslant 16\, c\, M C_N \sum_{r=0}^{\infty} (2r+1)^{d-1} \alpha^Y_{1,1}(r)^{\frac{\delta\gamma}{2+\delta\gamma}},$$

where the right hand side tends to 0 when N goes to infinity, uniformly in n. Hence $\sigma_n^2(N)$ tends to σ_n^2 uniformly in n as N goes to infinity.

Finally, for all constant $\nu > 0$ arbitrary small, we can choose N such that $\mathbb{E}[\sigma_n^{-1}|S_{2,n}|] \leqslant \nu$ and $|1 - \sigma_n(N)/\sigma_n| \leqslant \nu$ for all n sufficiently large. By looking at the characteristic function of $\sigma_n^{-1}S_n$ we get

$$\begin{aligned}\left|\mathbb{E}\left(e^{\frac{ixS_n}{\sigma_n}}\right) - e^{-\frac{1}{2}x^2}\right| &\leqslant \left|\mathbb{E}\left(e^{\frac{ixS_{1,n}}{\sigma_n}}\right) - \mathbb{E}\left(e^{\frac{ixS_{1,n}}{\sigma_n(N)}}\right)\right| + \left|\mathbb{E}\left(e^{\frac{ixS_{1,n}}{\sigma_n(N)}}\right) - e^{-\frac{1}{2}x^2}\right| + \left|\mathbb{E}\left(e^{\frac{ixS_{2,n}}{\sigma_n}} - 1\right)\right| \\ &\leqslant x\mathbb{E}\left(\left|\frac{S_{1,n}}{\sigma_n(N)}\right|\right)\left|1 - \frac{\sigma_n(N)}{\sigma_n}\right| + o(1) + x\nu \\ &\leqslant 2x\nu + o(1)\end{aligned}$$

concluding the proof. □

4. Application to determinantal point processes

In this section, we give a CLT for a wide class of functionals of DPPs. This result is a key tool for the asymptotic inference of DPPs. As an application treated in Section 4.3, we get the consistency and the asymptotic normality of the two-step estimation method of [38] for a parametric inhomogeneous DPP.

4.1. Negative association and α-mixing for DPPs

We recall that a DPP X on $\mathbb{R}^d$ is defined trough its intensity functions with respect to the Lebesgue measure that must satisfy

$$\forall n \in \mathbb{N},\ \forall x \in (\mathbb{R}^d)^n,\ \rho_n(x_1, \cdots, x_n) = \det(K[x]) \ \text{ with } K[x] := (K(x_i, x_j))_{i,j\in\{1,\cdots,n\}}.$$

The function $K : (\mathbb{R}^d)^2 \to \mathbb{C}$ is called the kernel of X and is assumed to satisfy the following standard general condition ensuring the existence of X.

Condition $\mathcal{H}$: The function $K : (\mathbb{R}^d)^2 \to \mathbb{C}$ is a locally square integrable hermitian measurable function such that its associated integral operator $\mathcal{K}$ is locally of trace class with eigenvalues in $[0, 1]$.

This condition is not necessary for existence, in particular there are examples of DPPs having a non-hermitian kernel. It is nonetheless very general and is assumed in most studies on DPPs. Basic properties of DPPs can be found in [25, 36, 30]. In particular, from [20, Theorem 1.4] and [30, Theorem 3.7], we know that DPPs are NA.

Theorem 4.1 ([20, 30]). *Let K satisfy Condition $\mathcal{H}$, then a DPP with kernel K is NA.*

By definition, for a DPP with kernel K we have $D(x, y) = -|K(x, y)|^2$ where D is introduced in (2.2). Hence, using the last theorem and Proposition 2.8 we get the following strong mixing coefficients of a DPP, where we define

$$\omega(r) := \sup_{|x-y|\geqslant r} |K(x, y)|. \tag{4.1}$$

Corollary 4.2. *Let X be a DPP with kernel K satisfying $\mathcal{H}$. Then, for all $p, q > 0$,*

$$\begin{cases} \alpha_{p,q}(r) \leqslant \sup\limits_{\substack{|A|<p,|B|<q \\ dist(A,B)>r}} \int_{A\times B} |K(x, y)|^2 \leqslant pq\omega(r)^2, \\ \alpha_{p,\infty}(r) \leqslant ps_d \int_r^\infty \omega^2(t)t^{d-1}\mathrm{d}t. \end{cases} \tag{4.2}$$

It is worth noticing that this result, and so the covariance inequality (2.8), is optimal in the sense that for a wide class of DPPs, the α-mixing coefficient $\alpha_{p,q}(r)$ do not decay faster than $\sup_{\substack{|A|<p,|B|<q \\ dist(A,B)>r}} |\mathrm{Cov}(N(A),N(B))|$ when r goes to infinity, as stated in the following proposition.

Proposition 4.3. *Let X be a DPP with kernel K satisfying $\mathcal{H}$. We further assume that K is bounded, takes its values in $\mathbb{R}_+$ and is such that $\|\mathcal{K}\| < 1$ where $\|.\|$ is the operator norm. Then, for all $p, q, r > 0$,*

$$(1-\|\mathcal{K}\|)^{\frac{(p+q)\|K\|_\infty}{\|\mathcal{K}\|}} \sup_{\substack{|A|<p,|B|<q \\ dist(A,B)>r}} \int_{A\times B} |K(x,y)|^2 \leqslant \alpha_{p,q}(r) \leqslant \sup_{\substack{|A|<p,|B|<q \\ dist(A,B)>r}} \int_{A\times B} |K(x,y)|^2. \tag{4.3}$$

Proof. The upper bound for $\alpha_{p,q}(r)$ is just the one in (4.2). The lower bound is obtained through void probabilities. Let $p, q, r > 0$ and $A, B \subset \mathbb{R}^d$ such that $|A| < p$, $|B| < q$ and $\mathrm{dist}(A, B) > r$. By definition, for any such sets A and B, $\alpha_{p,q}(r) \geqslant |\mathbb{P}(N(A) = 0)\mathbb{P}(N(B) = 0) - \mathbb{P}(N(A \cup B) = 0)|$. The void probabilities of DPPs are known (see [36]) and equal to

$$\mathbb{P}(N(A) = 0) = \exp\left(-\sum_{n\geqslant 1} \frac{\mathrm{Tr}(\mathcal{K}_A^n)}{n}\right)$$

where $\mathcal{K}_A$ is the projection of $\mathcal{K}$ on the set of square integrable functions $f : A \to \mathbb{R}$. Moreover, $\mathbb{P}(N(A) = 0)\mathbb{P}(N(B) = 0) - \mathbb{P}(N(A \cup B) = 0) \geqslant 0$ by negative association, and we have

$$\begin{aligned}
&\mathbb{P}(N(A) = 0)\mathbb{P}(N(B) = 0) - \mathbb{P}(N(A \cup B) = 0) \\
&= \exp\left(-\sum_{n\geqslant 1} \frac{\mathrm{Tr}(\mathcal{K}_{A\cup B}^n)}{n}\right)\left(\exp\left(\sum_{n\geqslant 1} \frac{\mathrm{Tr}(\mathcal{K}_{A\cup B}^n) - \mathrm{Tr}(\mathcal{K}_A^n) - \mathrm{Tr}(\mathcal{K}_B^n)}{n}\right) - 1\right) \\
&\geqslant \exp\left(-\sum_{n\geqslant 1} \frac{\mathrm{Tr}(\mathcal{K}_{A\cup B}^n)}{n}\right) \sum_{n\geqslant 1} \frac{\mathrm{Tr}(\mathcal{K}_{A\cup B}^n) - \mathrm{Tr}(\mathcal{K}_A^n) - \mathrm{Tr}(\mathcal{K}_B^n)}{n}.
\end{aligned} \tag{4.4}$$

Using the classical trace inequality we get

$$\mathrm{Tr}(\mathcal{K}_{A\cup B}^n) \leqslant \|\mathcal{K}_{A\cup B}\|^{n-1}\mathrm{Tr}(\mathcal{K}_{A\cup B}) \leqslant \|\mathcal{K}\|^{n-1}\int_{A\cup B} K(x,x)\mathrm{d}x \leqslant \|\mathcal{K}\|^{n-1}(p+q)\|K\|_\infty,$$

thus

$$\exp\left(-\sum_{n\geqslant 1} \frac{\mathrm{Tr}(\mathcal{K}_{A\cup B}^n)}{n}\right) \geqslant (1-\|\mathcal{K}\|)^{\frac{(p+q)\|K\|_\infty}{\|\mathcal{K}\|}}. \tag{4.5}$$

Moreover, since A and B are disjoint sets, we can write

$$\begin{aligned}
\mathrm{Tr}(\mathcal{K}_{A\cup B}^n) - \mathrm{Tr}(\mathcal{K}_A^n) - \mathrm{Tr}(\mathcal{K}_B^n) = &\int_{(A\cup B)^n} K(x_1,x_2)\cdots K(x_{n-1},x_n)K(x_n,x_1)\mathrm{d}x_1\cdots\mathrm{d}x_n \\
&- \int_{A^n\cup B^n} K(x_1,x_2)\cdots K(x_{n-1},x_n)K(x_n,x_1)\mathrm{d}x_1\cdots\mathrm{d}x_n,
\end{aligned} \tag{4.6}$$

which vanishes when $n = 1$, is equal to $2\int_{A\times B} |K(x,y)|^2$ when $n = 2$ and is non-negative for $n \geqslant 3$ since K is assumed to be non-negative. Finally, by combining (4.4), (4.5) and (4.6) we get the lower bound in (4.3). □

4.2. Central limit theorem for functionals of DPPs

We investigate the asymptotic distribution of functions that can be written as a sum over subsets of close enough points of X, namely

$$f(X) := \sum_{S\subset X} f_0(S), \tag{4.7}$$

where f_0 is a bounded function vanishing when $\mathrm{diam}(S) > \tau$ for a certain fixed constant $\tau > 0$. The typical example, encountered in asymptotic inference, concerns functions f_0 that are supported on sets S having exactly p elements, in which case (4.7) often takes the form

$$f(X) = \frac{1}{p!} \sum_{x_1,\cdots,x_p \in X}^{\neq} f_0(x_1, \cdots, x_p), \tag{4.8}$$

where the sum is done over ordered p-tuples of X and the symbol $\neq$ means that we consider distinct points. The asymptotic distribution of (4.8) has been investigated in [37] when $p = 1$ and in [3] for general p and stationary DPPs.

In the next theorem, we extend these settings to functionals like (4.7) applied to general non stationary DPPs. Some discussion and comments are provided after its proof. We use Minkwoski's notation and write $A \oplus r$ for the set $\bigcup_{x \in A} \mathcal{B}(x, r)$.

Theorem 4.4. *Let X be a DPP associated to a kernel K that satisfies $\mathcal{H}$ and that is further bounded. Let $\tau > 0$ and $f : \Omega \to \mathbb{R}$ be a function of the form*

$$f(X) := \sum_{S \subset X} f_0(S)$$

where f_0 is a bounded function vanishing when $diam(S) > \tau$. Let $(W_n)_{n \in \mathbb{N}}$ be a sequence of increasing subsets of $\mathbb{R}^d$ such that $|W_n| \to \infty$ and let $\sigma_n^2 := Var(f(X \cap W_n))$. Assume that there exists $\varepsilon > 0$ and $\nu > 0$ such that the following conditions are satisfied:

(H1) $|\partial W_n \oplus (\tau + \nu)| = o(|W_n|)$;
(H2) $\omega(r) = o(r^{-(d+\varepsilon)/2})$;
(H3) $\liminf_n |W_n|^{-1} \sigma_n^2 > 0$.

Then,

$$\frac{1}{\sigma_n}(f(X \cap W_n) - \mathbb{E}[f(X \cap W_n)]) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Proof. In order to apply Theorem 3.1, we would like to rewrite f as a sum over cubes of a lattice. Unfortunately, for disjoint sets $A, B \subset \mathbb{R}^d$, $f(X \cap A) + f(X \cap B) \neq f(X \cap (A \cup B))$ in general. Instead, we apply Theorem 3.1 to an auxiliary function, close to f, as follows. Define S^0 as the barycentre of the set S. We write

$$f_W(X) = \sum_{S \subset X} f_0(S) \mathbb{1}_W(S^0) \tag{4.9}$$

for the sum over the subsets of points of X with barycentre in $W \subset \mathbb{R}^d$. Now, we split $\mathbb{R}^d$ into little cubes the following way. Let C_0 be a given d-dimensional cube with a given side-length $0 < s \leqslant \nu/\sqrt{d}$. For all $i \in \mathbb{Z}^d$, let C_i be the translation of C_0 by the vector $s \cdot i$. Let $I_n := \{i : C_i \oplus \tau \subset W_n\}$ and $\widetilde{W}_n = \bigcup_{i \in I_n} C_i$. An illustration of these definitions is provided in Figure 1. Since $f_{\widetilde{W}_n}(X) = \sum_{i \in I_n} f_{C_i}(X)$ and each f_{C_i} are $\mathcal{E}(C_i \oplus \tau)$-measurable then $f_{\widetilde{W}_n}$ is the ideal candidate to use Theorem 3.1 on. Thus, we first prove that the difference between $f_{\widetilde{W}_n}$ and $f(X \cap W_n)$ is asymptotically negligible and then that $f_{\widetilde{W}_n}$ satisfies the conditions of Theorem 3.1.

First of all, notice that $dist(C_i, \partial W_n) \geqslant \tau$ for all $i \in I_n$. Therefore, for any point in W_n at a distance greater than $\tau + s\sqrt{d}$ from ∂W_n, the cube C_i of side-length s containing it is at a distance at least τ from ∂W_n, hence it is one of the C_i in $\widetilde{W}_n$ and we get

$$|W_n \backslash \widetilde{W}_n| \leqslant |\partial W_n \oplus (\tau + s\sqrt{d})|.$$

Hence, by Assumption (H1), $|W_n| \sim |\widetilde{W}_n|$. Now,

$$f(X \cap W_n) - f_{\widetilde{W}_n}(X) = \sum_{S \subset X \cap W_n} f_0(S) \mathbb{1}_{W_n \backslash \widetilde{W}_n}(S^0). \tag{4.10}$$

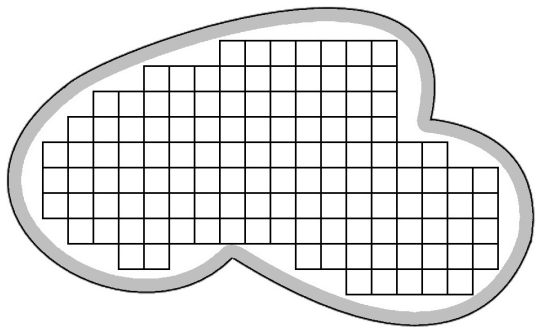

Figure 1. Example of illustration of the definition of $\widetilde{W}_n$. Here, the black border is ∂W_n, the grey area corresponds to $(\partial W_n \oplus \tau) \cap W_n$ and the square lattice corresponds to $\widetilde{W}_n$.

Since f_0 vanishes when two points of S are at distance further than τ, then the sum in (4.10) only concerns the subsets S of $X \cap ((W_n \backslash \widetilde{W}_n) \oplus \tau) \cap W_n)$. By Lemma B.6, the variance of $f(X \cap W_n) - f_{\widetilde{W}_n}(X)$ is then a $O(|(W_n \backslash \widetilde{W}_n) \oplus \tau|)$, whence a $o(|W_n|)$ and finally a $o(\sigma_n^2)$ by Assumption (H3). Therefore, $\sigma_n^{-1}(f(X \cap W_n) - \mathbb{E}[f(X \cap W_n)])$ has the same limiting distribution as $\sigma_n^{-1}(f_{\widetilde{W}_n}(X) - \mathbb{E}[f_{\widetilde{W}_n}(X)])$. Moreover, we have

$$|\mathrm{Cov}(f_{\widetilde{W}_n}(X), f(X \cap W_n) - f_{\widetilde{W}_n}(X))| \leqslant \sigma_n \sqrt{\mathrm{Var}(f(X \cap W_n) - f_{\widetilde{W}_n}(X))} = \sigma_n o(\sqrt{|W_n|}) = o(\sigma_n^2)$$

by Assumptions (H1), (H3) and Lemma B.6 proving that $\sigma_n^{-1}(f(X \cap W_n) - \mathbb{E}[f(X \cap W_n)])$ has the same limiting distribution as $\mathrm{Var}(f_{\widetilde{W}_n}(X))^{-1/2}(f_{\widetilde{W}_n}(X) - \mathbb{E}[f_{\widetilde{W}_n}(X)])$.

We conclude by showing that the random variables $Y_i = f_{C_i}(X) - \mathbb{E}[f_{C_i}(X)]$ satisfy the assumptions of Theorem 3.1. A rough bound on f gives us $|f_{C_i}(X)| \leqslant \|f_0\|_\infty 2^{N(C_i \oplus \tau)}$ so, by Lemma B.5,

$$\forall n \in \mathbb{N}, \quad \sup_{i \in \mathbb{Z}^d} \mathbb{E}[|Y_i|^n] < \infty.$$

This means that the Y_i's satisfy Assumption (C2) for all $\delta > 0$ and thus (C3) as a consequence of (H2). Finally, since $|I_n| = s^{-d}|\widetilde{W}_n| = O(|W_n|)$ and $\mathrm{Var}(f_{\widetilde{W}_n}(X)) \sim \sigma_n^2$, we have

$$\liminf_n |I_n|^{-1} \mathrm{Var}(f_{\widetilde{W}_n}(X)) > 0$$

by Assumption (H3), which concludes the proof of the theorem. □

We highlight two direct extensions of this result.

i) Since the superposition of independent PA (respectively NA) point processes remains a PA (respectively NA) point process, then Theorem 4.4 holds true for α-determinantal point processes where $\alpha \in \{-1/m : m \in \mathbb{N}^*\}$, see [36] for more information about α-DPPs.

ii) Theorem 4.4 also extends to $\mathbb{R}^q$-valued functions f where $q \geqslant 2$. Let $\Sigma_n := \mathrm{Var}(f(X \cap W_n))$. If we replace (H3) by

$$\liminf_n |W_n|^{-1} \lambda_{\min}(\Sigma_n) > 0$$

where $\lambda_{\min}(\Sigma_n)$ denotes the smallest eigenvalue of Σ_n, then Theorem 4.4 holds true with the conclusion

$$(\Sigma_n)^{-1/2}(f(X \cap W_n) - \mathbb{E}[f(X \cap W_n)]) \xrightarrow{\mathcal{L}} \mathcal{N}(0, Id_q)$$

where Id_q is the $q \times q$ identity matrix.

Let us comment the assumptions of Theorem 4.4.

- Condition (H1) makes clear the idea that W_n must grow to $\mathbb{R}^d$ as $n \to \infty$, without being a too irregular set. In the simple case where W_n is the Cartesian product of intervals, i.e. $W_n = \Delta_n^{(1)} \times \cdots \times \Delta_n^{(d)}$, then (H1) is equivalent to $|\Delta_n^{(k)}| \to \infty$ for all k.
- Condition (H2) is not really restrictive and is satisfied by all classical kernel families. For example, the kernels of the Ginibre ensemble and of the Gaussian unitary ensemble (see [25]) have an exponential decay. Moreover, all translation-invariant kernels used in spatial statistics (see [29] and [4]) satisfy $\omega(r) = O(r^{-(d+1)/2})$: the Gaussian and the Laguerre-Gaussian covariance functions have an exponential decay; the Whittle-Matérn and the Cauchy covariance functions satisfy $\omega(r) = o(r^{-d})$; and in the case of the most repulsive DPP in dimension d (as determined in [29, 4]), which is the slowest decaying Bessel-type kernel, its kernel is given by

$$K(x,y) = \frac{\sqrt{\rho\Gamma(\frac{d}{2}+1)}}{\pi^{\frac{d}{4}}} \frac{J_{\frac{d}{2}}(2\sqrt{\pi}\Gamma(\frac{d}{2}+1)^{\frac{1}{d}}\rho^{\frac{1}{d}}||y-x||)}{||y-x||^{\frac{d}{2}}} \Rightarrow \omega(r) = O\left(r^{-\frac{d+1}{2}}\right),$$

  where $\rho > 0$ is a constant. While this DPP satisfies Condition (H2), we point out that its α-mixing coefficients decay too slowly to be able to derive a CLT only from them, see the discussion before Theorem 3.1. This justifies the importance of Condition (C3) in this theorem, obtained by the NA property, and which leads to Condition (H2).
- Condition (H3) is harder to control in the broad setting of Theorem 4.4, but we can get sufficient conditions in some particular cases. For example, if $f_0(S) = \mathbb{1}_{|S|=1}$ and K is a translation-invariant continuous kernel then it was shown in [37] that Condition (H3) holds when K is not the Fourier transform of an indicator function. In the peculiar case where K is the Fourier transform of an indicator function, [37] proved that the limiting distribution is still Gaussian but the rate of convergence is different. As another example extending the previous one, assume that f_0 is a non-negative function supported on the set $\{S \subset X : |S| = p\}$ for a given integer $p > 0$ and assume that the highest eigenvalue of the integral operator $\mathcal{K}$ associated to K is less than 1. Then, we show in Proposition B.7 that

$$\liminf_n \frac{1}{|W_n|} \int_{W_n^p} f_0(x) \det(K[x]) \mathrm{d}x > 0$$

  implies (H3) and is much easier to verify.

4.3. Application to the two-step estimation of an inhomogeneous DPP

In this section, we consider DPPs on $\mathbb{R}^2$ with kernel of the form

$$K_{\beta,\psi}(x,y) = \sqrt{\rho_\beta(x)} C_\psi(y-x) \sqrt{\rho_\beta(y)}, \quad \forall x, y \in \mathbb{R}^2, \tag{4.11}$$

where $\beta \in \mathbb{R}^p$ and $\psi \in \mathbb{R}^q$ are two parameters, C_ψ is a correlation function and ρ_β is of the form $\rho_\beta(x) = \rho(z(x)\beta^T)$ where ρ is a known positive strictly increasing function and z is a p-variate bounded function called covariates. This form implies that the first order intensity, corresponding to $\rho_\beta(x)$, is inhomogeneous and depends on the covariates $z(x)$ through the parameter β. But all higher order intensity functions once normalized, i.e. $\rho^{(n)}(x_1, \ldots, x_n)/(\rho_\beta(x_1) \ldots \rho_\beta(x_n))$, are translation-invariant for $n \geqslant 2$. In particular, the pair correlation (the case $n = 2$) is invariant by translation. This kind of inhomogeneity is sometimes named second-order intensity reweighted stationarity and is frequently assumed in the spatial point process community.

Existence of DPPs with a kernel like above is for instance ensured if $\rho_\beta(x)$ is bounded by $\rho_{\max}$ and C_ψ is a continuous, square-integrable correlation function on $\mathbb{R}^d$ whose Fourier transform is less than $1/\rho_{\max}$, see [29]. For later use, we call $\mathcal{H}'$ the previous assumptions on $K_{\beta,\psi}$.

Consider the observation of a DPP X with kernel $K_{\beta*,\psi*}$, along with the covariates z, within a window $W_n := [an, bn] \times [cn, dn]$ where $b > a$ and $d > c$. Waagepetersen and Guan [38] have proposed the following

two-step estimation procedure of (β^*, ψ^*) for second-order intensity reweighted stationary models. First, $\hat{\beta}_n$ is obtained by solving

$$u_{n,1}(\beta) := \sum_{u \in X \cap W_n} \frac{\nabla \rho_\beta(u)}{\rho_\beta(u)} - \int_{W_n} \nabla \rho_\beta(u) \mathrm{d}u = 0.$$

where $\nabla \rho_\beta$ denotes the gradient with respect to β. In the second step, $\hat{\psi}_n$ is obtained by minimizing $m_{n,\hat{\beta}_n}$ where

$$m_{n,\beta}(\psi) := \int_{r_l}^{r} \left(\left(\sum_{u,v \in X \cap W_n} \frac{\mathbb{1}_{\{0<|u-v|\leqslant t\}}}{\rho_\beta(u)\rho_\beta(v)|W_n \cap W_{n,u-v}|} \right)^c - K_\psi(t)^c \right)^2 \mathrm{d}t.$$

Here r_l, r and c are user-specified non-negative constants, $W_{n,u-v}$ is W_n translated by $u-v$ and K_ψ is the Ripley K-function defined by

$$K_\psi(t) := \int_{\|u\|\leqslant t} g_\psi(u) \mathrm{d}u$$

where $g_\psi(u) := 1 - C_\psi(u)^2/C_\psi(0)^2$ is the pair correlation function of X. If we define

$$u_{n,2}(\beta, \psi) := -|W_n| \frac{\partial m_{n,\beta}(\psi)}{\partial \psi},$$

then the two-step procedure amounts to solve

$$u_n(\beta, \psi) := (u_{n,1}(\beta), u_{n,2}(\beta, \psi)) = 0.$$

The asymptotic properties of this two-step procedure are established in [38], under various moments and mixing assumptions, with a view to inference for Cox processes. We state hereafter the asymptotic normality of $(\hat{\beta}_n, \hat{\psi}_n)$ in the case of DPPs with kernel of the form (4.11). This setting allows us to apply Theorem 4.4 and get rid of some restrictive mixing assumptions needed in [38].

The asymptotic covariance matrix of $(\hat{\beta}_n, \hat{\psi}_n)$ depends on two matrices defined in [38, Section 3.1], where they are denoted by $\widetilde{\Sigma}_n$ and I_n. We do not reproduce their expression, which is hardly tractable. An assumption in [38] ensures the asymptotic non-degeneracy of this covariance matrix and we also need this assumption in our case, see (W4) below. Unfortunately, as discussed in [38], it is hard to check this assumption for a given model, particularly because it depends on the covariates z. We are confronted by the same limitation in our setting. On the other hand, the other assumptions of the following theorem are not restrictive. In particular almost all standard kernels satisfy (W3) below, see the discussion after Theorem 4.4.

Theorem 4.5. *Let X be a DPP with kernel K_{β^*,ψ^*} given by (4.11) and satisfying $\mathcal{H}'$. Let $(\hat{\beta}_n, \hat{\psi}_n)$ the two-step estimator defined above. We assume the following.*

(W1) $r_l > 0$ if $c < 1$; otherwise $r_l \geqslant 0$,
(W2) ρ_β and K_ψ are twice continuously differentiable as functions of β and ψ,
(W3) $\sup_{\|x\|\geqslant r} C_{\psi^*}(x) = O(r^{-1-\varepsilon})$,
(W4) Condition N3 in [38] (concerning the matrices I_n and $\widetilde{\Sigma}_n$) is satisfied.

Then, there exists a sequence $\{(\hat{\beta}_n, \hat{\psi}_n) : n \geqslant 1\}$ for which $u_n(\hat{\beta}_n, \hat{\psi}_n) = 0$ with a probability tending to one and

$$|W_n|^{1/2}[(\hat{\beta}_n, \hat{\psi}_n) - (\beta^*, \psi^*)] I_n \widetilde{\Sigma}_n^{-1/2} \xrightarrow{\mathcal{L}} \mathcal{N}(0, Id).$$

Proof. Let ρ_k be the kth intensity function of the DPP with kernel $(x, y) \mapsto C_{\psi^*}(y - x)$. In order to apply Theorem 1 in [38] we need to show that

(i) ρ_2, ρ_3 are bounded and there is a constant M such that for all $u_1, u_2 \in \mathbb{R}^2$, $\int |\rho_3(0, v, v + u_1) - \rho_1(0)\rho_2(0, u_1)| \mathrm{d}v < M$ and $\int |\rho_4(0, u_1, v, v + u_2) - \rho_2(0, u_1)\rho_2(0, u_2)| \mathrm{d}v < M$,
(ii) $\|\rho_{4+2\delta}\|_\infty < \infty$ for some $\delta > 0$,
(iii) $\alpha_{a,\infty}(r) = O(r^{-d})$ for some $a > 8r^2$ and $d > 2(2 + \delta)/\delta$.

The first property (i) is a consequence of (W3). This is because we can write

$$|\rho_3(0,v,v+u_1)-\rho_1(0)\rho_2(0,u_1)| = |2C_{\psi*}(v)C_{\psi*}(u_1)C_{\psi*}(v+u_1) - C_{\psi*}(0)(C_{\psi*}(v+u_1)^2 + C_{\psi*}(v)^2)|$$

which is bounded by $2|C_{\psi*}(0)|(C_{\psi*}(v+u_1)^2 + C_{\psi*}(v)^2)$ and

$$\int_{\mathbb{R}^2} C_{\psi*}(v)^2 \mathrm{d}v \leqslant 2\pi \int_0^\infty r \sup_{\|x\|=r} |C_{\psi*}(x)|^2 \mathrm{d}r$$

which is finite by Assumption (W3). The term $\rho_4(0,u_1,v,v+u_2) - \rho_2(0,u_1)\rho_2(0,u_2)$ can be treated the same way. For a DPP, (ii) is satisfied for any $\delta > 0$. Finally, (iii) is the one that causes an issue since, as stated before, the α-mixing coefficient we get in Corollary 4.2 decreases slower than what we desire. But, the only place this assumption is used in [38] is to prove the asymptotic normality of their estimator in their Lemma 5, which can also be derived as a consequence of our Theorem 4.4 with Assumption (W3). □

Appendix A: Proof of Theorem 2.3

We use the following variant of the monotone class theorem (see [13, Theorem 22.1]).

Theorem A.1. *Let $\mathcal{S}$ be a set of bounded functions stable by bounded monotone convergence and uniform convergence. Let $\mathcal{C}$ be a subspace of $\mathcal{S}$ such that $\mathcal{C}$ is an algebra containing the constant function $\widetilde{1}$. Then, $\mathcal{S}$ contains all bounded functions measurable over $\sigma(\mathcal{C})$.*

Now, let $A, B_1, \cdots, B_k$ be pairwise distinct Borel subsets of $\mathbb{R}^d$ and $g : \mathbb{N}^k \mapsto \mathbb{R}$ be a coordinate-wise increasing function. We denote by Ω_A the set of locally finite point configuration in A and we define $\mathcal{S}$ as the set of functions $f : \Omega_A \mapsto \mathbb{R}$ such that

$$\mathbb{E}[\hat{f}(X\cap A)g(N(B_1),\cdots,N(B_k))] \leqslant \mathbb{E}[\hat{f}(X\cap A)]\mathbb{E}[g(N(B_1),\cdots,N(B_k)], \tag{A.1}$$

where $\hat{f}(X) := \sup_{Y\leqslant X} f(Y)$. Note that $\hat{f}$ is an increasing function and that f is increasing iff $\hat{f} = f$. Our goal is to prove that $\mathcal{S}$ contains all bounded functions supported over A. Because of the definition of NA point processes (2.5), we know that $\mathcal{S}$ contains the set $\mathcal{C}$ of functions of the form $f(N(A_1),\cdots,N(A_k))$ where the A_i are pairwise disjoints Borel subsets of A. In particular, since point processes over A are generated by the set of random vectors $\{(N(A_1),\cdots,N(A_k)) : A_i \subset A \text{ disjoints}, k\in\mathbb{N}\}$, then we only need to verify that $\mathcal{S}$ and $\mathcal{C}$ satisfy the hypothesis of Theorem A.1 to conclude.

- Stability of $\mathcal{S}$ by bounded monotonic convergence: Since (A.1) is invariant if we add a constant to f and f is bounded then we can consider f to be positive. Now, notice that for all functions h and k, $h \leqslant k \Rightarrow \hat{h} \leqslant \hat{k}$ and $h \geqslant k \Rightarrow \hat{h} \geqslant \hat{k}$. So, if we take a positive bounded monotonic sequence $f_n \in \mathcal{S}$ that converges to a bounded function f, then $\hat{f}_n$ is also a positive bounded monotonic sequence that consequently converges to a function g. Suppose that $(f_n)_n$ is an increasing sequence (the decreasing case can be treated similarly) and let us show that $g = \hat{f}$. Let $X \in \Omega_A$, for all $Y \subset X$, $f_n(Y) \leqslant f(Y)$. Taking the suppremum then the limit gives us $g(X) \leqslant \hat{f}(X)$. Moreover, for all $Y \subset X$, $g(X) \geqslant \hat{f}_n(X) \geqslant f_n(Y)$. Taking the limit gives us that $g(X) \geqslant f(Y)$ for all $Y \subset X$ so $g(X) \geqslant \hat{f}(X)$ which proves that $g = \hat{f}$. Using the monotone convergence theorem we conclude that

$$\begin{aligned} &\mathbb{E}[\hat{f}_n(X\cap A)] \to \mathbb{E}[\hat{f}(X\cap A)] \\ \text{and}\quad &\mathbb{E}[\hat{f}_n(X\cap A)g(N(B_1),\cdots,N(B_k))] \to \mathbb{E}[\hat{f}(X\cap A)g(N(B_1),\cdots,N(B_k))], \end{aligned} \tag{A.2}$$

  which proves that $f \in \mathcal{S}$
- Stability of $\mathcal{S}$ by uniform convergence: Let f_n be a sequence over $\mathcal{S}$ converging uniformly to a function f then, by Lemma B.1, $\hat{f}_n$ also converges uniformly (and therefore in L^1) to $\hat{f}$. As a consequence, (A.2) is also satisfied in this case so $f \in \mathcal{S}$.

- $\mathcal{C}$ is an algebra: It is easily shown that $\mathcal{C}$ is a linear space containing $\widetilde{1}$ so we only need to prove that $\mathcal{C}$ is stable by multiplication. Let $A_1, \cdots, A_r$ and $A'_1 \cdots, A'_s$ be two sequences of pairwise distinct Borel subsets of A. Let $f = f(N(A_1), \cdots, N(A_r)) \in \mathcal{C}$ and $h = h(N(A'_1), \cdots, N(A'_r)) \in \mathcal{C}$. We can write

$$N(A_i) = N(A_i \backslash \cup_j A'_j) + \sum_j N(A_i \cap A'_j) \quad \text{and} \quad N(A'_i) = N(A'_i \backslash \cup_j A_j) + \sum_j N(A'_i \cap A_j),$$

so $f \cdot h$ can be expressed as a function of the number of points in the subsets $A_i \backslash \cup_j A'_j$, $A'_i \backslash \cup_j A_j$ and $A_i \cap A'_j$ that are all pairwise distinct Borel subsets of A, proving that $\mathcal{C}$ is stable by multiplication.

This concludes the proof that $\mathcal{S}$ contains all bounded functions supported over Ω_A. By doing the same exact reasoning on the set of bounded functions g which satisfy $\mathbb{E}[f(X \cap A)g(X \cap B)] \leqslant \mathbb{E}[f(X \cap A)]\mathbb{E}[g(X \cap B)]$ for a fixed f we obtain the same result which concludes the proof.

Appendix B: Auxiliary results

Lemma B.1. *Let E be a set and $f, g : E \to \mathbb{R}$ be two functions, then*

$$\left| \sup_{x \in E} f(x) - \sup_{y \in E} g(y) \right| \leqslant \|f - g\|_\infty$$

Proof. The proposition becomes trivial once we write

$$f(x) \leqslant g(x) + \|f - g\|_\infty \leqslant \sup_{y \in E} g(y) + \|f - g\|_\infty$$

Taking the supremum yields the first inequality. Moreover, by symmetry of f and g the second one follows similarly. □

Lemma B.2. *Let $i, j \in \mathbb{Z}^d$ such that $|i - j|_1 := \sum_{l=1}^d |i_l - j_l| = r$. Let $s, R > 0$ and C_i, C_j be the d-dimensional cubes with side length s and respective centre $x_i = R \cdot i$ and $x_j = R \cdot j$. Then,*

$$dist(C_i, C_j) \geqslant \frac{1}{\sqrt{d}}(rR - sd).$$

Moreover, each cube intersects at most $(2sd/R)^d$ other cubes.

Proof. Since each point of a d-dimensional square with side length s is at distance at most $s\sqrt{d}/2$ from it's centre, we get $\text{dist}(C_i, C_j) \geqslant \sqrt{(Ri_1 - Rj_1)^2 + \cdots + (Ri_d - Rj_d)^2} - s\sqrt{d}$ which takes it's minimum when $|i_l - j_l| = r/d$ for all $1 \leqslant l \leqslant d$ hence $\text{dist}(C_i, C_j) \geqslant rR/\sqrt{d} - s\sqrt{d}$.
In particular, if $|i - j|_1 > sd/R$ then $C_i \cap C_j = \varnothing$, hence for all $i \in \mathbb{Z}^d$

$$|\{j \in \mathbb{Z}^d : C_i \cap C_j \neq \varnothing, i \neq j\}| \leqslant |\{j : 0 < |i - j|_1 \leqslant sd/R\}| \leqslant \left(\frac{2sd}{R}\right)^d$$

□

Lemma B.3. *Let M and N be two $n \times n$ semi-positive definite matrices such that $0 \leqslant M \leqslant N^{-1}$ where $\leqslant$ denotes the Loewner order. Then,*

$$\det(Id - MN) \geqslant 1 - Tr(MN)$$

Proof. First, let's consider the case where $N = Id$. If $\mathrm{Tr}(M) \geqslant 1$ then $\det(Id - M) \geqslant 0 \geqslant 1 - \mathrm{Tr}(M)$. Otherwise, we denote by $\mathrm{Sp}(M)$ the spectrum of M and since all eigenvalues are in [0,1[, we can write

$$\begin{aligned}
\det(Id - M) &= \prod_{\lambda \in \mathrm{Sp}(M)} \exp(\ln(1-\lambda)) \\
&= \prod_{\lambda \in \mathrm{Sp}(M)} \exp\left(-\sum_{n=0}^{\infty} \frac{\lambda^n}{n}\right) \\
&= \exp\left(-\sum_{n=0}^{\infty} \frac{\mathrm{Tr}(M^n)}{n}\right) \\
&\geqslant \exp\left(-\sum_{n=0}^{\infty} \frac{\mathrm{Tr}(M)^n}{n}\right) \\
&= \exp(\log(1 - \mathrm{Tr}(M))) \\
&= 1 - \mathrm{Tr}(M). \qquad \text{(B.1)}
\end{aligned}$$

Getting back to the general case, we can write N as $S^T S$ and by Sylvester's determinant identity we get that $\det(Id - MN) = \det(Id - SMS^T)$. Since we assumed that $0 \leqslant M \leqslant N^{-1}$ then $0 \leqslant SMS^T \leqslant Id$ and by applying (B.1) this concludes the proof:

$$\det(Id - MN) = \det(Id - SMS^T) \geqslant 1 - \mathrm{Tr}(SMS^T) = 1 - \mathrm{Tr}(MN)$$

□

Proposition B.4. *Let M be a $n \times n$ semi-definite positive matrix of the form $M = \begin{pmatrix} M_1 & N \\ N^T & M_2 \end{pmatrix}$ where M_1 is a $k \times k$ semi-definite positive matrix, M_2 is a $(n-k) \times (n-k)$ semi-definite positive matrix and N is a $k \times (n-k)$ matrix. We define $||A||_\infty = \sup |a_{i,j}|$. Then,*

$$0 \leqslant \det(M_1)\det(M_2) - \det(M) \leqslant k(n-k)\, Tr(N^T N)||M||_\infty^{n-2}.$$

Proof. First, we assume that M_1 and M_2 are invertible. Using Schur's complement, we can write

$$\det(M) = \det(M_1)\det(M_2)\det(Id - M_1^{-1} N^T M_2^{-1} N)$$

where $0 \leqslant N^T M_2^{-1} N \leqslant M_1$ with $\leqslant$ being the Loewner order. $N^T M_2^{-1} N$ being semi-definite positive implies

$$\det(M) \leqslant \det(M_1)\det(M_2),$$

while the inequality $N^T M_2^{-1} N \leqslant M_1$ gives us (see Lemma B.3)

$$\det(M) \geqslant \det(M_1)\det(M_2)(1 - \mathrm{Tr}(M_1^{-1} N^T M_2^{-1} N)).$$

Therefore,

$$\begin{aligned}
0 \leqslant \det(M_1)\det(M_2) - \det(M) &\leqslant \mathrm{Tr}(\mathrm{adj}(M_1) N^T \mathrm{adj}(M_2) N) \\
&\leqslant \mathrm{Tr}(\mathrm{adj}(M_1))\mathrm{Tr}(\mathrm{adj}(M_2))\mathrm{Tr}(N^T N) = \sum_{i=1}^{k} \Delta_i(M_1) \sum_{j=1}^{n-k} \Delta_j(M_2)\mathrm{Tr}(N^T N),
\end{aligned}$$

where $\Delta_i(M_1)$ means the (i,i) minor of the matrix M_1 and $\mathrm{adj}(M_1)$ is the transpose of the matrix of cofactor of M_1. But, since all principal sub-matrices of M_1 and M_2 are positive definite matrices then their determinant is lower than the product of their diagonal entries, meaning that $\Delta_i(M_1) \leqslant \prod_{j \neq i} M_1(j,j) \leqslant ||M||_\infty^{k-1}$. Doing the same thing for the terms $\Delta_j(M_2)$ gives us the desired result.

If M_1 or M_2 is not invertible, a limit argument using the continuity of the determinant leads to the same conclusion.

□

Lemma B.5. *Let X be a DPP with bounded kernel K satisfying $\mathcal{H}$, $s > 0$ and $n > 0$, then*

$$\sup_{A\subset\mathbb{R}^d, |A|=s} \mathbb{E}[2^{nN(A)}] < \infty$$

Proof. Let $n \in \mathbb{N}$ and $A \subset \mathbb{R}^d$ such that $|A| = s$. Since the determinant of a positive semi-definite matrix is always smaller than the product of its diagonal coefficients we get

$$\begin{aligned}\mathbb{E}[2^{nN(A)}] &= \mathbb{E}\left[\sum_{k=0}^{\infty}\binom{N(A)}{k}(2^n-1)^k\right]\\ &= \sum_{k=0}^{\infty}\frac{(2^n-1)^k}{k!}\int_{A^k}\det(K[x])\mathrm{d}x\\ &\leqslant e^{(2^n-1)\|K\|_\infty|A|} < \infty.\end{aligned}$$

□

Lemma B.6. *Let X be a DPP on $\mathbb{R}^d$ with bounded kernel K satisfying $\mathcal{H}$ such that $\omega(r) = O(r^{-\frac{d+\varepsilon}{2}})$ for a certain $\varepsilon > 0$. Then, for all bounded Borel sets $W \subset R^d$ and all bounded functions $g : \bigcup_{p>0}(\mathbb{R}^d)^p \to \mathbb{R}$ such that $g(S)$ vanishes when $diam(S) > \tau$ for a given constant $\tau > 0$,*

$$Var\left(\sum_{S\subset X\cap W} g(S)\right) = O(|W|). \tag{B.2}$$

Proof. Since W is bounded then $N(W)$ is almost surely finite and we can write

$$\sum_{S\subset X\cap W} g(S) = \sum_{p\geqslant 0}\sum_{\substack{S\subset X\cap W\\ |S|=p}} g(S).$$

Looking at the variance of each term individually, we start by developing $\mathbb{E}\left[\left(\sum_{S\subset X\cap W} g(S)\mathbb{1}_{|S|=p}\right)^2\right]$ as

$$\sum_{k=0}^{p}\mathbb{E}\left[\sum_{\substack{S,T\subset X\cap W\\ |S|=|T|=p, |S\cap T|=k}} g(S)g(T)\right] \tag{B.3}$$

$$\begin{aligned}&= \sum_{k=0}^{p}\mathbb{E}\left[\sum_{\substack{U\subset X\cap W\\ |U|=2p-k}}\sum_{\substack{S'\subset S\subset U\\ |S'|=k, |S|=p}} g(S)g(S'\cup(U\backslash S))\right]\\ &= \sum_{k=0}^{p}\frac{1}{(2p-k)!}\int_{W^{2p-k}}\sum_{\substack{S'\subset S\subset\{x_1,\cdots,x_{2p-k}\}\\ |S'|=k, |S|=p}} g(S)g(S'\cup(U\backslash S))\rho_{2p-k}(x_1,\cdots,x_{2p-k})\mathrm{d}x_1\cdots\mathrm{d}x_{2p-k}\\ &= \sum_{k=0}^{p}\frac{1}{(2p-k)!}\binom{p}{k}\binom{2p-k}{p}\int_{W^{2p-k}} g(x_1,\cdots,x_p)g(x_1,\cdots,x_k,x_{p+1},\cdots,x_{2p-k})\rho_{2p-k}(x)\mathrm{d}x. \end{aligned} \tag{B.4}$$

Since the determinant of a positive semi-definite matrix is smaller than the product of its diagonal terms, we have $|\rho_{2p-k}(x)| \leqslant \|K\|_\infty^{2p-k}$. Moreover, as a consequence of our assumptions on g, each term for $k \geqslant 1$

in (B.4) is bounded by

$$\frac{1}{p!(p-k)!}\binom{p}{k}\int\limits_{W^{2p-k}}\|g\|_\infty^2\|K\|_\infty^{2p-k}\mathbb{1}_{\{0\leqslant|x_i-x_1|\leqslant\tau,\ \forall i\}}\mathrm{d}x\leqslant\frac{|W|}{p!}\binom{p}{k}\|g\|_\infty^2\|K\|_\infty^{2p-k}|\mathcal{B}(0,\tau)|^{2p-k-1}$$
$$\leqslant\frac{|W|}{p!}\binom{p}{k}\|g\|_\infty^2(1+\|K\|_\infty)^{2p}(1+|\mathcal{B}(0,\tau)|)^{2p}.$$

Hence,

$$\sum_{k=1}^{p}\mathbb{E}\left[\sum_{\substack{S,T\subset X\cap W\\|S|=|T|=p,|S\cap T|=k}}g(S)g(T)\right]\leqslant|W|\|g\|_\infty^2\frac{C_1^p}{p!}\tag{B.5}$$

where $C_1=2(1+\|K\|_\infty)^2(1+\mathcal{B}(0,\tau))^2$ is a constant independent from p and W. However, even if all terms for $k\geqslant1$ in (B.4) are a $O(|W|)$, this is not the case of the term for $k=0$ which is a $O(|W|^2)$. Instead of controlling this term alone, we consider its difference with the remaining term in the variance we are looking at, that is

$$\frac{1}{(p!)^2}\int_{W^{2p}}g(x)g(y)\rho_{2p}(x,y)\mathrm{d}x\mathrm{d}y-\mathbb{E}\left[\left(\sum_{S\subset X\cap W}g(S)\mathbb{1}_{|S|=p}\right)\right]^2$$
$$=\frac{1}{(p!)^2}\int_{W^{2p}}g(x)g(y)(\rho_{2p}(x,y)-\rho_p(x)\rho_p(y))\mathrm{d}x\mathrm{d}y.$$

Using Proposition B.4, we get

$$|\rho_{2p}(x,y)-\rho_p(x)\rho_p(y)|\leqslant p^2\|K\|_\infty^{2p-2}\sum_{1\leqslant i,j\leqslant p}K(x_i,y_j)^2.$$

Now, notice that for all $y\in\mathbb{R}^d$ and $1\leqslant i\leqslant p$,

$$\int_{W^p}\mathbb{1}_{\{0<|x_k-x_j|\leqslant\tau,\ \forall j,k\}}|K(x_i,y)|^2\mathrm{d}x\leqslant|\mathcal{B}(0,\tau)|^{p-1}\int_W|K(x_i,y)|^2\mathrm{d}x_i\leqslant|\mathcal{B}(0,\tau)|^{p-1}s_d\int_{\mathbb{R}^d}r^{d-1}\omega(r)^2\mathrm{d}r$$

which is finite because of our assumption on $\omega(r)$. Thus, we obtain the inequality

$$\int_{W^{2p}}g(x)g(y)|K(x_i,y_j)|^2\mathrm{d}x\mathrm{d}y\leqslant\|g\|_\infty|\mathcal{B}(0,\tau)|^{p-1}\int_{W^{p+1}}g(x)|K(x_i,y_1)|^2\mathrm{d}x\mathrm{d}y_1$$
$$\leqslant|W|\|g\|_\infty^2|\mathcal{B}(0,\tau)|^{2p-2}s_d\int_{\mathbb{R}^d}r^{d-1}\omega(r)^2\mathrm{d}r.\tag{B.6}$$

By combining (B.5) and (B.6), we get the bound

$$\mathrm{Var}\left(\sum_{\substack{S\subset X\cap W\\|S|=p}}g(S)\right)\leqslant|W|\|g\|_\infty^2\left(\frac{C_1^p}{p!}+\frac{C_2}{p!}\right)$$

where

$$C_2:=\left(\sup_{p\geqslant0}\frac{p^4\|K\|_\infty^{2p-2}|\mathcal{B}(0,\tau)|^{2p-2}}{p!}\right)s_d\int_{\mathbb{R}^d}r^{d-1}\omega(r)^2\mathrm{d}r$$

is a constant independent from p and W. Finally,

$$\sum_{p\geqslant0}\mathrm{Var}\left(\sum_{\substack{S\subset X\cap W\\|S|=p}}g(S)\right)=O(|W|)$$

and

$$\sum_{p>q\geqslant 0} \mathrm{Cov}\left(\sum_{\substack{S\subset X\cap W\\|S|=p}} g(S), \sum_{\substack{S\subset X\cap W\\|S|=q}} g(S)\right) \leqslant |W|\|g\|_\infty^2 \sum_{p,q\geqslant 0} \sqrt{\left(\frac{C_1^p}{p!}+\frac{C_2}{p!}\right)\left(\frac{C_1^q}{q!}+\frac{C_2}{q!}\right)} = O(|W|)$$

concluding the proof. □

Proposition B.7. *Let $p\in\mathbb{N}$, $f:\mathbb{R}^p\to\mathbb{R}_+$ be a symmetrical measurable function and define*

$$F(X) = \sum_{\substack{S\subset X\\|S|=p}} f(S).$$

Let X be a DPP with kernel K satisfying Condition $\mathcal{H}$ such that $\|\mathcal{K}\|<1$ where $\|\mathcal{K}\|$ is the operator norm of the integral operator associated with K. If, for a given increasing sequence of compact sets $W_n\subset\mathbb{R}^d$,

$$\liminf_n \frac{1}{|W_n|}\int_{W_n^p} f(x)\det(K[x])\mathrm{d}x > 0, \tag{B.7}$$

then

$$\liminf_n \frac{1}{|W_n|} Var(F(X\cap W_n)) > 0.$$

Proof. Let W be a compact subset of $\mathbb{R}^d$. The Cauchy-Schwartz inequality gives us

$$\mathrm{Cov}(F(X\cap W), N(W))^2 \leqslant \mathrm{Var}(F(X\cap W))\mathrm{Var}(N(W)).$$

We showed in Lemma B.6 that $|W|^{-1}\mathrm{Var}(N(W))$ is bounded by a constant $C>0$ so we are only interested in the behaviour of $\mathrm{Cov}(F(X\cap W), N(W))$. We start by developing $\mathbb{E}[F(X\cap W)N(W)]$:

$$\begin{aligned}
\mathbb{E}[F(X\cap W)N(W)] &= \mathbb{E}\left[\sum_{\substack{S\subset X\cap W\\|S|=p}} f(S)\sum_{x\in X\cap W} 1\right]\\
&= \mathbb{E}\left[\sum_{\substack{S\subset X\cap W\\|S|=p+1}}\sum_{x\in S} f(S\backslash\{x\}) + p\sum_{\substack{S\subset X\cap W\\|S|=p}} f(S)\right]\\
&= \frac{1}{(p+1)!}\int_{W^{p+1}}\sum_{i=1}^{p+1} f(z\backslash\{z_i\})\det(K[z])\mathrm{d}z + \frac{1}{p!}\int_{W^p} pf(x)\det(K[x])\mathrm{d}x\\
&= \frac{1}{p!}\left(\int_{W^p} f(x)\left(p\det(K[x]) + \int_W \det(K[x,a])\mathrm{d}a\right)\mathrm{d}x\right).
\end{aligned}$$

We also have

$$\mathbb{E}[F(X\cap W)]\mathbb{E}[N(W)] = \frac{1}{p!}\int_{W^p} f(x)\det(K[x])\mathrm{d}x\int_W K(a,a)\mathrm{d}a,$$

hence

$$\mathrm{Cov}(F(X\cap W), N(W)) = \frac{1}{p!}\int_{W^p} f(x)\det(K[x])\left(p - \int_W K(a,a) - \det(K[x,a])\det(K[x])^{-1}\mathrm{d}a\right)\mathrm{d}x. \tag{B.8}$$

Using Schur's complement, we get

$$K(a,a) - \det(K[x,a])\det(K[x])^{-1} = K_{ax}K[x]^{-1}K_{ax}^T \tag{B.9}$$

where we define K_{ax} as the vector $(K(a,x_1),\cdots,K(a,x_p))$. Moreover, since we look at our point process in a compact window W, a well-known property of DPPs (see [25]) is that there exists a sequence of eigenvalues λ_i in $[0,\|\mathcal{K}\|]$ and an orthonormal basis of $L^2(W)$ of eigenfunctions ϕ_i such that

$$K(x,y)=\sum_i \lambda_i\phi_i(x)\bar{\phi}_i(y) \quad \forall x,y\in W.$$

As a consequence, $\forall x,y\in W$,

$$\int_W K(x,a)K(a,y)\mathrm{d}a=\sum_i \lambda_i^2\phi_i(x)\bar{\phi}_i(y)$$

which we define as $L(x,y)$. Therefore, for all $x\in W^p$, $L[x]\leqslant\|\mathcal{K}\|K[x]$ where $\leqslant$ is the Loewner order for positive definite symmetric matrices and we get

$$\int_W K_{ax}K[x]^{-1}K_{ax}^T\mathrm{d}a=\mathrm{Tr}\left(K[x]^{-1}\int_W K_{ax}^TK_{ax}\mathrm{d}a\right)=\mathrm{Tr}(K[x]^{-1}L[x])\leqslant p\|\mathcal{K}\|. \tag{B.10}$$

Finally, since f is non negative, by combining (B.8), (B.9) and (B.10) we get the lower bound

$$\mathrm{Var}(F(X\cap W))\geqslant\frac{\mathrm{Cov}(F(X\cap W),N(W))^2}{\mathrm{Var}(N(W))}\geqslant\frac{(1-\|\mathcal{K}\|)^2}{C(p-1)!^2|W|}\left(\int_{W^p}f(x)\det(K[x])\mathrm{d}x\right)^2$$

which proves the proposition. □

References

[1] K. Alam and K.M.L. Saxena. Positive dependence in multivariate distributions. *Comm. Statist. Theory Methods A*, 10:1183–1196, 1981.

[2] R. Bardenet and A. Hardy. Monte Carlo with determinantal point processes. *arXiv preprint arXiv:1605.00361*, 2016.

[3] C.A.N. Biscio and F. Lavancier. Brillinger mixing of determinantal point processes and statistical applications. *Electron. J. Statist.*, 10:582–607, 2016.

[4] C.A.N. Biscio and F. Lavancier. Quantifying repulsiveness of determinantal point processes. *Bernoulli*, 22:2001–2028, 2016.

[5] B. Błaszczyszyn and D. Yogeshwaran. Clustering comparison of point processes with applications to random geometric models. In V. Schmidt, editor, *Stochastic Geometry, Spatial Statistics and Random Fields: Models and Algorithms*, volume 2120 of *Lecture Notes in Mathematics*, chapter 2, pages 31–71. Springer, 2014.

[6] E. Bolthausen. On the central limit theorem for stationnary mixing random fields. *The Annals of Probability*, 10:1047–1050, 1982.

[7] R.C. Bradley. Basic properties of strong mixing conditions. a survey and some open questions. *Probability Surveys*, 2:107–144, 2005.

[8] A. Bulinski and E. Shabanovich. Asymptotic behaviour of some functionals of positively and negatively dependent random fields. *Fundam. Prikl. Mat.*, 4:479–492, 1998.

[9] A. Bulinski and A. Shashkin. *Limit Theorems for Associated Random Fields and Related Systems.* Springer-Verlag, 2007.

[10] R. Burton and E. Waymire. Scaling limits for associated random measures. *Ann. Appl. Probab.*, 13:1267–1278, 1985.

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

[12] Yu. A. Davydov. The convergence of distributions which are generated by stationary random processes. *Teor. Verojatnost. i Primenen.*, 13:730–737, 1968.

[13] C. Dellacherie and P. Meyer. *Probabilities and Potential.* Elsevier Science Ltd, 1978.

[14] N. Deng, W. Zhou, and M. Haenggi. The ginibre point process as a model for wireless networks with repulsion. *IEEE Transactions on Wireless Communications*, 1:479–492, 2015.
[15] P. Doukhan. *Mixing: Properties and Examples.* Springer-Verlag, 1994.
[16] P. Doukhan, K. Fokianos, and X. Li. On weak dependence conditions: the case of discrete valued processes. *Statist. Probab. Lett.*, 82(11):1941–1948, 2012.
[17] P. Doukhan and S. Louhichi. A new weak dependence condition and applications to moment inequalities. *Stochastic Process. Appl.*, 84:313–342, 1999.
[18] J.D. Esary, F. Proschan, and D.W. Walkup. Association of random variables, with applications. *Ann. Math. Statist.*, 38:1466–1474, 1967.
[19] S.N. Evans. Association and random measures. *Probability Theory and Related Fields*, 86:1–19, 1990.
[20] S. Ghosh. Determinantal processes and completeness of random exponentials: the critical case. *Probability Theory and Related Fields*, 163(3):643–665, 2015.
[21] J. S. Gomez, A. Vasseur, A. Vergne, P. Martins, L. Decreusefond, and W. Chen. A Case Study on Regularity in Cellular Network Deployment. *IEEE Wireless Communications Letters*, 4(4):421–424, 2015.
[22] X. Guyon. *Random Fields on a Network.* Springer-Verlag, 1994.
[23] L. Heinrich. On the strong Brillinger-mixing property of α-determinantal point processes and some applications. *Appl. Math.*, 61(4):443–461, 2016.
[24] L. Heinrich and S. Klein. Central limit theorems for empirical product densities of stationary point processes. *Statistical Inference for Stochastic Processes. An International Journal Devoted to Time Series Analysis and the Statistics of Continuous Time Processes and Dynamical Systems*, 17(2):121–138, 2014.
[25] J.B. Hough, M. Krishnapur, Y. Peres, and B. Virag. *Zeros of Gaussian Analytic Functions and Determinantal Point Processes.* American Mathematical Society, 2009.
[26] I.A. Ibragimov and Y.V. Linnik. *Independant and stationnary sequences of random variables.* Wolters-Noordhoff, 1971.
[27] E. Jolivet. Central limit theorem and convergence of empirical processes for stationary point processes. In *Point processes and queuing problems (Colloqium, Keszthely, 1978)*, volume 24 of *Colloq. Math. Soc. János Bolyai*, pages 117–161. North-Holland, Amsterdam-New York, 1981.
[28] A. Kulesza and B. Taskar. Determinantal point process models for machine learning. *Foundations and Trends in Machine Learning*, 5:123–286, 2012.
[29] F. Lavancier, J. Møller, and E. Rubak. Determinantal point process models and statistical inference. *Journal of Royal Statistical Society: Series B (Statistical Methodology)*, 77:853–877, 2015.
[30] R. Lyons. Determinantal probability: Basic properties and conjectures. *Proceedings of the International Congress of Mathematicians, Seoul, Korea*, IV:137–161, 2014.
[31] O. Macchi. The coincidence approach to stochastic point processes. *Advances in Applied Probability*, 7:83–122, 1975.
[32] N. Miyoshi and T. Shirai. A cellular network model with ginibre configured base stations. *Advances in Applied Probability*, 46:832–845, 2014.
[33] J. Møller and R.P. Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes.* Chapman and Hall/CRC, Boca Raton, 2004.
[34] E. Rio. Covariance inequalities for strongly mixing processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 29(4):587–597, 1993.
[35] M. Rosenblatt. A central limit theorem and a strong mixing condition. *Probc. Nat. Acad. Sci. U.S.A.*, 42:43–47, 1956.
[36] T. Shirai and Y. Takahashi. Random point fields associated with certain fredholm determinants i: fermion, poisson and boson point processes. *Journal of Functional Analysis*, 205:414–463, 2003.
[37] A. Soshnikov. Determinantal random point fields. *Russian Math. Surveys*, 55:923–975, 2000.
[38] R. Waagepetersen and Y. Guan. Two-step estimation for inhomogeneous spatial point processes. *Journal of the Royal Statistical Society*, 71:685–702, 2009.
[39] M. Yuan, C. Su, and T. Hu. A central limit theorem for random fields of negatively associated processes. *Journal of Theoretical Probability*, 16:309–323, 2003.