



HAL
open science

Asymptotic inference of stationary and non-stationary determinantal point processes

Arnaud Poinas

► **To cite this version:**

Arnaud Poinas. Asymptotic inference of stationary and non-stationary determinantal point processes. Probability [math.PR]. Université de Rennes, 2019. English. NNT : 2019REN1S024 . tel-02286143

HAL Id: tel-02286143

<https://theses.hal.science/tel-02286143>

Submitted on 13 Sep 2019

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.

THÈSE DE DOCTORAT DE

L'UNIVERSITE DE RENNES 1
COMUE UNIVERSITE BRETAGNE LOIRE

Ecole Doctorale N°601
*Mathématique et Sciences et Technologies
de l'Information et de la Communication*
Spécialité : *Mathématiques et leurs interactions*

Par

Arnaud POINAS

Statistiques asymptotiques des processus ponctuels déterminan- taux stationnaires et non stationnaires

Thèse présentée et soutenue à RENNES, le 4 Juillet 2019
Unité de recherche : IRMAR

Rapporteurs avant soutenance :

Bartłomiej BLASZCZYSZYN, Directeur de recherche INRIA, ENS Paris
Marie-Colette VAN LIESHOUT, Professeur, University of Twente

Composition du jury :

Examineurs : Rémi BARDENET, Chargé de recherche CNRS, Université de Lille
Agnès DESOLNEUX, Directeur de recherche CNRS, ENS Paris-Saclay
Bartłomiej BLASZCZYSZYN, Directeur de recherche INRIA, ENS Paris
Marie-Colette VAN LIESHOUT, Professeur, University of Twente

Dir. de thèse : Bernard DELYON, Professeur, Université de Rennes 1

Co-dir. de thèse : Frédéric LAVANCIER, Maître de conférences, Université de Nantes

Remerciements

Tout d'abord, je tiens à remercier mes deux directeurs : Bernard Delyon et Frédéric Lavancier, pour tout leur investissement ces trois dernières années. Réaliser ma thèse avec vous deux a été un immense plaisir ; vous m'avez apporté tout ce que je pouvais attendre de mes directeurs de thèse.

Bernard, dès que je coinçais sur un problème ou que j'avais une question, je pouvais tout de suite aller te voir dans le bureau voisin et tu étais toujours prêt à me consacrer autant de temps qu'il le fallait pour m'aider. Et une bonne partie des résultats des chapitres 2 et 3 ont été obtenus grâce à ton aide immense.

Quant à toi, Frédéric, tu as aussi été tout le temps disponible pour m'aider. Toujours rapide à répondre à mes mails, quelle que soit l'heure à laquelle je pouvais les envoyer. Je ne compte pas le nombre de fois que tu as relu tout ce que j'ai pu écrire. Tu m'as appris comment rédiger des papiers et comment faire des talks, moi qui au début de la thèse rédigeais des phrases telles que "In this documents, we will study the comportement of [...]". Merci à tous les deux pour tout ce que vous avez pu m'apprendre ces trois dernières années.

Un grand merci également à Bartek Błaszczyszyn et Marie-Colette Van Lieshout d'avoir accepté de rapporter ma thèse. Je vous remercie de l'intérêt que vous avez porté à mon travail et d'avoir pris le temps de relire attentivement tout ce que j'ai rédigé. Enfin, je vous remercie aussi, ainsi que Rémi Bardenet et Agnès Désolneux, pour avoir accepté d'être présent dans mon jury de thèse. En particulier, je remercie Rémi de m'avoir aussi donné la possibilité de continuer à travailler sur les DPPs avec lui et Adrien Hardy, après que ma thèse soit validée.

J'en profite aussi pour remercier toutes les personnes que j'ai pu rencontrer lors des rencontres Franco-Danoises de statistiques spatiales, personnes avec lesquelles j'ai passé de très bons moments. En particulier, je remercie Rasmus Waagepetersen pour m'avoir permis de partir 3 semaines au Danemark pour aller travailler avec lui, ce qui fut une expérience très enrichissante. Merci aussi à Christophe Biscio pour m'avoir logé durant ces trois semaines bien que je ne sois pas le plus agréable des colocataires et merci aussi pour m'avoir aidé à m'intégrer à l'équipe de statistiques spatiales et pour les bons moments qu'on a pu partager à Aalborg.

Je remercie aussi tous les chercheurs de l'IRMAR et de l'ENS Rennes que j'ai pu avoir en cours de la L3 au M2 et qui ont forgé mes connaissances les plus avancées en Mathématiques. Merci aussi au personnel administratif de l'IRMAR qui a toujours été d'une très grande aide pour toute démarche administrative durant ma thèse.

Je souhaite également remercier toutes les personnes qui m'ont guidé d'une façon ou d'une autre vers la recherche en Mathématiques. Pour commencer, merci à Nicolas Basbois et Boris Bellini mes professeurs de Mathématiques et Physique-Chimie de MPSI à Stanislas Cannes. Alors que je ne savais pas vraiment ce que je voulais faire plus tard, ce sont eux qui m'ont orienté vers une MP* au lycée Massena à Nice, suivie de l'ENS, en plus d'avoir été de très bons professeurs dont je me souviens encore très bien 9 ans

plus tard.

Merci à Jean-Daniel Boissonnat pour m'avoir pris en stage en L3. Ce fut ma toute première expérience avec le milieu de la recherche qui m'était alors inconnu. C'est à l'issue de ce stage que je fus pour la première fois convaincu que je voulais faire Mathématicien-Chercheur (et que ce ne serait pas en géométrie algorithmique). De même, merci à Philippe Thieullen pour m'avoir pris en stage en M1. C'est à l'issue de ce stage que j'ai décidé de m'orienter vers les Probabilités/Statistiques alors que j'étais complètement indécis sur mon orientation future auparavant.

Merci à tous les amis que je me suis fait en cours de route et avec qui j'ai pu partager tout un tas de bons moments au cours de mes études. De Brice, Pauline, Elodie et Anaïs au lycée, à Benjamin, Zoïs et Rubén à l'ENS, en passant par (presque) toute ma classe de MPSI pour ne donner que quelques noms. Je remercie également mes camarades de bureau Thi, Ninon, Tristan et Thibault avec qui j'ai passé 3 années de thèse très agréables.

Enfin, je remercie du fond du cœur mes parents. Si je devais dédier cette thèse à quelqu'un, je le ferai à eux sans hésiter. Merci à vous de m'avoir constamment soutenu jusqu'au bout. Malgré la distance, malgré les problèmes que j'ai pu avoir et malgré les problèmes que vous avez pu avoir, vous avez toujours été à mes côtés et je vous en suis éternellement reconnaissant. Je ne serais pas là aujourd'hui si vous ne m'aviez pas soutenu jusqu'au bout. Je ne pouvais pas espérer avoir de meilleurs parents et je suis fier d'avoir la chance d'être votre fils. Merci pour tout. Merci aussi à mes grands-parents pour le soutien qu'ils m'ont apporté. Avoir pu faire mes études avec l'encouragement de toute ma famille a toujours été une énorme source de motivation.

En conclusion, je remercie toutes les personnes qui, d'une façon ou d'une autre, ont eu un impact positif sur ma vie et sans qui je ne serais peut-être pas là aujourd'hui à terminer la rédaction de ma thèse. Merci à tous.

Table of contents

1	Introduction	5
1.1	Random point patterns and repulsiveness	5
1.2	Preliminaries on point processes	8
1.2.1	Definition of a point process	8
1.2.2	Determinantal point processes	9
1.2.3	Parametric families of determinantal point processes	10
1.3	Outline of the PHD	12
1.4	Quantifying dependency in a point process	12
1.4.1	α -mixing and β -mixing	12
1.4.2	Positive and negative association	13
1.5	Presentation of the PHD's results	14
1.5.1	Chapter 2	14
1.5.2	Chapter 3	15
1.5.3	Chapter 4	16
1.5.4	Chapter 5	17
1.5.5	Appendix	18
2	Mixing properties and central limit theorem for associated point processes	21
2.1	Introduction	21
2.2	Associated point processes and α -mixing	22
2.2.1	Notation	22
2.2.2	Negative and positive association	23
2.2.3	Application to α -mixing	26
2.3	Central limit theorem for associated point processes	26
2.4	Application to determinantal point processes	30
2.4.1	Negative association and α -mixing for DPPs	30
2.4.2	Central limit theorem for functionals of DPPs	32
2.4.3	Application to the two-step estimation of an inhomogeneous DPP	35
2.A	Proof of Theorem 2.2.3	38
2.B	Auxiliary results	39
3	A bound of the β-mixing coefficient for point processes in terms of their intensity functions	47
3.1	Introduction	47
3.2	Preliminaries	48
3.2.1	Intensities of point processes	48
3.2.2	Mixing	48
3.2.3	Lower sum transform	49
3.3	β -mixing of point processes with known intensity functions	50

3.4	Application to determinantal point processes	52
4	Adaptive estimating function inference for non-stationary determinantal point processes	55
4.1	Introduction	55
4.2	Estimating functions based on joint intensities	56
4.2.1	Joint intensity functions and Campbell formulae	56
4.2.2	A general asymptotic result for estimating functions	57
4.2.3	Second order estimating functions	57
4.2.4	Adaptive version	58
4.3	Asymptotic results for determinantal point processes	59
4.3.1	Second order estimating functions for DPPs	60
4.3.2	Two-step estimation for a separable parameter	64
4.4	Simulation study	65
4.4.1	Performance of adaptive estimating function	66
4.4.2	Two-step versus simultaneous	67
4.5	Discussion	70
4.A	Assumptions and proof of Theorem 4.2.1	70
4.B	Proof of Lemma 4.A.1	77
4.C	Identifiability of the kernel of a continuous DPP	78
4.D	Examples of DPP families satisfying (F3) or (F3')	79
4.E	Supplementary tables for Section 4.4.1	80
5	Consistency and approximation of the likelihood of stationary continuous determinantal point processes	87
5.1	Introduction	87
5.2	Likelihood of DPPs	88
5.2.1	Likelihood expression	88
5.2.2	Approximation of the likelihood for stationary DPPs	89
5.2.3	Consistency of the MLE	90
5.2.4	Correcting edge effects	92
5.2.5	Estimation of the intensity by MLE	93
5.3	Simulation study	94
5.3.1	Performance of the approximate MLE for Gaussian-type DPPs	94
5.4	Proofs of Section 5.2	97
5.4.1	Proof of Proposition 5.2.2	98
5.4.2	Proof of Proposition 5.2.3	98
5.4.3	Proof of Theorem 5.2.4	101
5.A	Technical Lemmas	106
A	Supplementary Results	113
A.1	Negative association of finite DPPs	113
A.2	Generalized Cramér-Wold device	118

Chapter 1

Introduction

1.1 Random point patterns and repulsiveness

One of the main topics of spatial statistics is the study of datasets consisting of a finite set of points falling into an observation window that is a compact subset of \mathbb{R}^d (where, usually, $d = 1, 2$ or 3). The earliest and probably most known example is the study in 1854 of the physician John Snow [95] of the spatial distribution of cases of Cholera during an outbreak in London which allowed him to trace back its origin to a contaminated water pump. In practice, a lot of spatial point patterns are observed in a wide variety of scientific disciplines. Some examples are given in Figure 1. The mathematical models that serve to study these kind of spatial data are point processes.

The simplest way to describe a point process on \mathbb{R}^d is as a probability measure on the σ -algebra of locally finite subset of \mathbb{R}^d , where a subset X of \mathbb{R}^d is said to be locally finite if for all bounded subsets A of \mathbb{R}^d the set $X \cap A$ is finite. The most commonly studied point process is the Poisson point process (or PPP). A PPP X is defined by a locally integrable function ρ called its intensity such that the number $N(A)$ of points of the PPP that falls into a subset A of \mathbb{R}^d follows a Poisson distribution with mean $\int_A \rho(x) dx$ and, for any disjoint sets $A_1, \dots, A_n \subset \mathbb{R}^d$, the random variables $N(A_1), \dots, N(A_n)$ are mutually independent. As a consequence, conditionally to $N(A)$, all points of $X \cap A$ are independent and identically distributed for any set $A \subset \mathbb{R}^d$. This is why PPPs are used to model spatial data consisting of points that can be considered to be independent from each other.

The main limitation of PPPs is obviously the assumption that all points are independently distributed and therefore have no interaction between each others whereas a lot of point patterns in practice feature some positive dependency (attractiveness) or negative dependency (repulsiveness) between points. A simulation of a PPP is shown in the top left of Figure 2. In comparison, some examples of point patterns with a repulsive behaviour are given in Figure 1. This is why more complex models of point processes have been developed in order to model attractive or repulsive point patterns. In this thesis, we focus mainly on the study of repulsive point patterns. The main models for these kind of data are Matérn's hardcore point processes, perturbed lattice point processes, Gibbs point processes and, the main focus of this thesis, determinantal point processes.

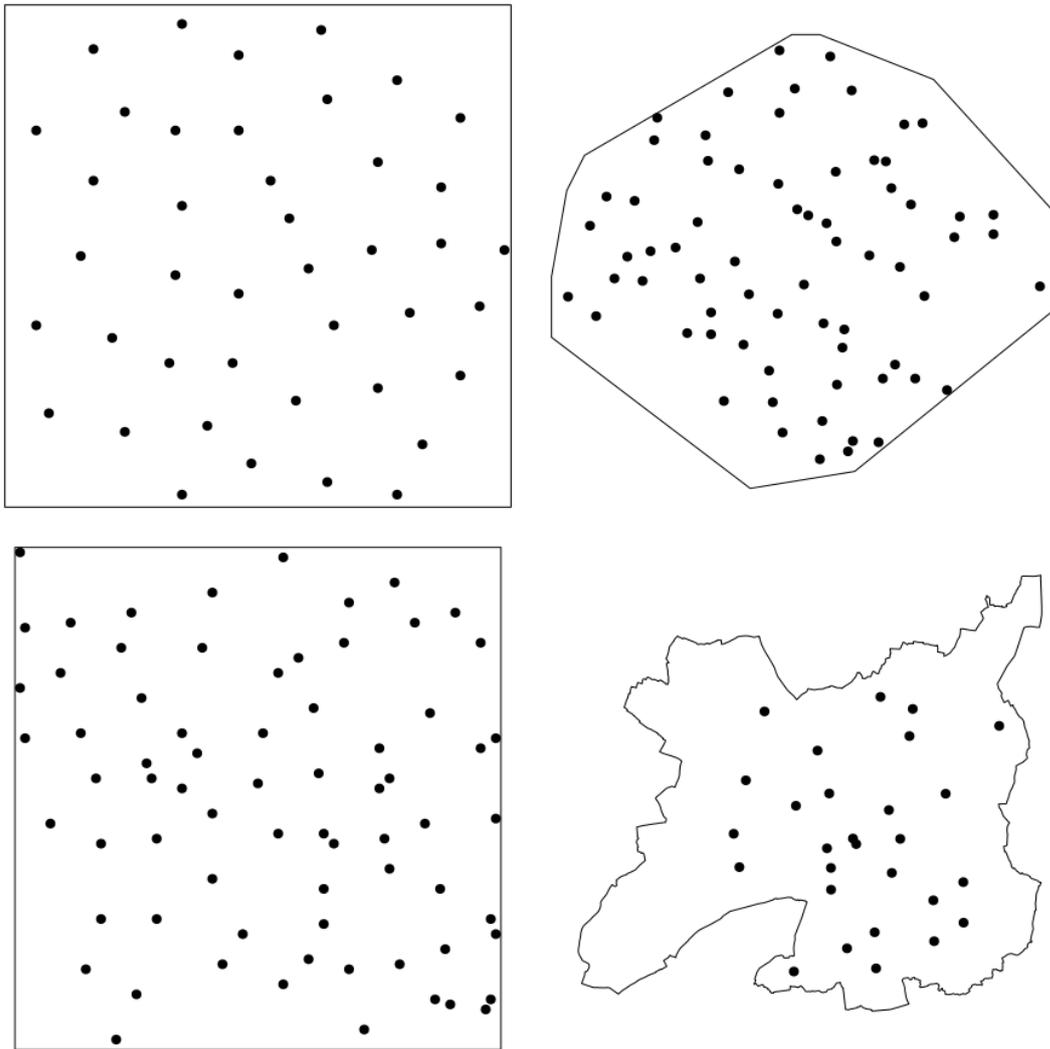


Figure 1: Examples of spatial point patterns. The top left figure shows the location of the centres of 42 biological cells in an histological section. The top right figure shows the location of 68 Messor ants nests at a site in Greece. The bottom left picture shows the location of 71 pine saplings in a Swedish forest. The bottom right picture shows the location of the 28 polling places in the French commune of Rennes. The first three datasets are provided in the *spatstat.data* package [5] of R [87]. The last dataset is provided by the website <https://data.rennesmetropole.fr>.

Matérn’s hardcore point processes are defined by taking a PPP and removing points so that no pair of points are closer than a certain distance R (called the hardcore radius) from each other while perturbed lattice point processes are defined by randomly shifting a lattice by i.i.d. random variables. Both models are easy to simulate and convenient to use but are too restrictive for a lot of statistical applications. On the opposite, Gibbs point processes are a more general class of point processes that offer a large variety of interactions between points and have been extensively used for modeling repulsiveness but they are generally difficult to work with. For example, they can only be simulated by MCMC methods, their density involves an intractable constant making maximum likelihood inference difficult and no closed form is known for their moments. More details can be found in [77, 100]

DPPs on the other hand are a more restrictive family than Gibbs point processes

[44] but are generally easier to work with while still being very flexible and allowing a large variety of interactions between points. Compared to Gibbs point processes, they can be exactly simulated and they have a closed form expression for their moments and Janossy densities. This allows the inference of DPPs by maximum likelihood estimation or various methods of moments. This is why DPPs present themselves as a nice alternative to model repulsive point patterns. Examples of simulations of DPPs, Gibbs point processes and Matérn's hardcore point processes are shown in Figure 2.

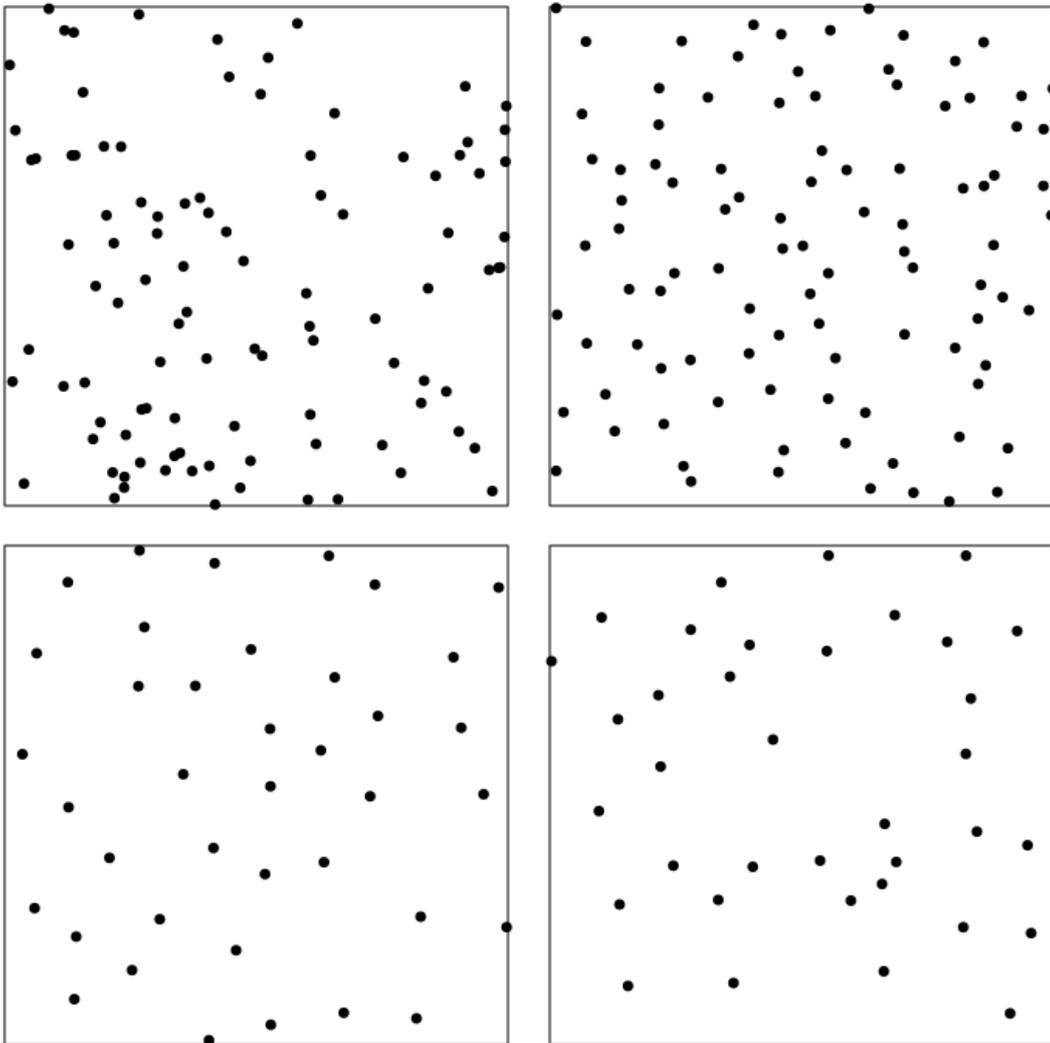


Figure 2: Examples of simulations of four different types of point processes on the window $W = [0, 1]^2$. The top left figure is a realization of an homogeneous PPP with intensity $\rho = 100$. The top right figure is a realization of a DPP with Bessel-type kernel (4.4.1) with parameters $\rho = 100$ and $\alpha = 0.05$. The bottom left figure is a realization of a type II Matérn's point process with hardcore radius $R = 0.1$. The bottom right figure is a realization of a Strauss point process (a special type of Gibbs point process defined in [98]) with parameters $\beta = 100$, $\gamma = 0.2$ and $R = 0.1$. All simulations have been done in R [87] using the *spatstat* [7] package.

1.2 Preliminaries on point processes

1.2.1 Definition of a point process

We consider (simple) point processes on the measured space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu)$ where $d \in \mathbb{N}$ is a constant, $\mathcal{B}(\mathbb{R}^d)$ is the set of *Borel* subsets of \mathbb{R}^d and μ is a Radon measure. Let Ω be the space of locally finite subsets of \mathbb{R}^d :

$$\Omega := \{X \subset \mathbb{R}^d : \forall S \in \mathcal{B}_0, \text{card}(X \cap S) < +\infty\}$$

where \mathcal{B}_0 denotes the space of bounded Borel sets of \mathbb{R}^d . We then consider \mathcal{F} , the σ -algebra generated by the sets

$$\{X \in \Omega, \text{card}(X \cap S) = m\}$$

for all $S \in \mathcal{B}_0$ and $m \in \mathbb{N}$. A (simple) point process X is then defined as a measurable application from a measured space into (Ω, \mathcal{F}) . We write $X = \emptyset$ when X corresponds to the empty point pattern and we denote by $N(S)$ the random variables $N(S) := \text{card}(X \cap S)$. General properties about point processes can be found in [31, 100] for example.

Several ways of defining a point process include using joint intensity functions, Janossy densities, void probabilities or the Laplace functional. Joint intensity functions of a point process are defined the following way:

Definition 1.2.1. *Let $X \in \Omega$ be a point process and $n \geq 1$ be an integer. If there exists a non-negative function $\rho_n : (\mathbb{R}^d)^n \rightarrow \mathbb{R}$ such that*

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

for all locally integrable functions $f : (\mathbb{R}^d)^n \rightarrow \mathbb{R}$, where the symbol \neq means that the sum is done for distinct x_i , then ρ_n is called the n -th order joint intensity function of X .

Thus, ρ_n is defined as the density of the factorial moment measures (see [31]) of X with respect to μ . Therefore, $\rho_n(x_1, \dots, x_n) d\mu(x_1) \cdots d\mu(x_n)$ can be simply interpreted as the probability of X having a point in each of the infinitesimal balls centered at x_i with volume $d\mu(x_i)$. Janossy densities on the other hand are defined the following way:

Definition 1.2.2. *Let $X \in \Omega$ be a point process and $S \in \mathcal{B}_0(\mathbb{R}^d)$. If there exists non-negative functions $j_n^S : (\mathbb{R}^d)^n \rightarrow \mathbb{R}$ for all integers $n \geq 1$ such that*

$$\mathbb{E}[f(X \cap S)] = \sum_{n \geq 0} \frac{1}{n!} \int_{S^n} f(x) j_n^S(x) d\mu^n(x)$$

for all locally integrable functions $f : \bigcup_{n \geq 0} S^n \rightarrow \mathbb{R}$, then the functions j_n^S are called the Janossy densities of X .

For Janossy densities of a point process X , $j_n^S(x_1, \dots, x_n) d\mu(x_1) \cdots d\mu(x_n)$ can be interpreted as the probability of X having exactly n points in S , one in each of the infinitesimal balls centered at x_i with volume $d\mu(x_i)$. Void probabilities are defined as the probabilities $\mathbb{P}(X \cap S = \emptyset)$ for all $S \in \mathcal{B}_0$. Since the set of events $\{\{X \cap S = \emptyset\}, S \in \mathcal{B}_0\}$ generates the σ -algebra \mathcal{F} then the knowledge of void probabilities is enough to

define a point process. Finally, the Laplace functional of a point process X is defined as the functional

$$L : f \mapsto \mathbb{E} \left[\exp \left(- \sum_{x \in X} f(x) \right) \right]$$

for all measurable non-negative functionals f with bounded support. This is the point process equivalent of the Laplace transform for real random variables.

1.2.2 Determinantal point processes

The introduction of DPPs under their current form comes from Macchi [75] to model fermion systems. Even if instances of DPPs can be found in earlier papers [67, 46], Macchi is the first one to have introduced them under the general form that is mostly used today.

A DPP is defined by a function $K : (\mathbb{R}^d)^2 \rightarrow \mathbb{C}$ called the kernel of the DPP. The most common way of defining a DPP is by its joint intensity functions:

Definition 1.2.3. *Let $K : (\mathbb{R}^d)^2 \rightarrow \mathbb{C}$. X is said to be a determinantal point process on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu)$ with kernel K if its joint intensity functions exist and satisfy*

$$\rho_n(x_1, \dots, x_n) = \det(K[x]) \quad (1.2.1)$$

for all integer n and for all $x_1, \dots, x_n \in \mathbb{R}^d$ where $K[x]$ is defined as the matrix $(K(x_i, x_j))_{1 \leq i, j \leq n}$.

A first consequence of Definition 1.2.3 is that the first order intensity of a DPP writes $\rho_1(x) = K(x, x)$ which shows that the local density of the point process is given by the diagonal values of its kernel. A second consequence is that $\rho_2(x, y) - \rho(x)\rho(y) = -|K(x, y)|^2 < 0$ showing that any pair of points has less chance to appear than both points individually causing points to repel each other. In particular, the higher $|K(x, y)|$ is, the stronger the repulsion is, so the off-diagonal values of the kernel of a DPP show the pairwise repulsion between points. More generally, for any configuration of points (x_1, \dots, x_n) , the closer two points x_i and x_j are from each other, the closer the i -th and j -th rows and columns of the matrix $K[x]$ will be, and the lower the values of the determinant and ρ_n are, which shows that DPPs favor configurations of points that are distant from each other. This explains the repulsive nature of determinantal point process.

The main issue with the general definition 1.2.3 is that it is not clear for which kernel K a DPP is well-defined. The usual assumptions considered on K for the DPP to be well-defined are the following:

Condition \mathcal{H} : The function $K : (\mathbb{R}^d)^2 \rightarrow \mathbb{C}$ is a continuous locally square integrable hermitian measurable function such that its associated integral operator on $L^2(\mathbb{R}^d, \mu)$

$$\mathcal{K} : f \mapsto \left(\mathcal{K}f : x \mapsto \int_{\mathbb{R}^d} K(x, y)f(y)d\mu(y) \right)$$

is locally of trace class with eigenvalues in $[0, 1]$.

Here, $L^2(\mathbb{R}^d, \mu)$ denotes the space of square integrable functions on \mathbb{R}^d . Under Condition \mathcal{H} , it is also possible to define DPPs through their void probabilities or their Laplace transform:

Proposition 1.2.4 ([94]). *Let X be a DPP with kernel K satisfying condition \mathcal{H} , then for all $S \in \mathcal{B}_0$,*

$$\mathbb{P}(X \cap S = \emptyset) = \det(Id - \mathcal{K}_S)$$

where \det refers to the Fredholm determinant, Id is the identity operator on $L^2(\mathbb{R}^d, \mu)$ and \mathcal{K}_S is the projection of \mathcal{K} on $L^2(S, \mu)$. Moreover, the Laplace transform of X writes

$$L(f) = \det(Id - \mathcal{K}[f])$$

where $\mathcal{K}[f]$ is the integral operator with kernel

$$K_{[f]}(x, y) := \sqrt{1 - e^{-f(x)}} K(x, y) \sqrt{1 - e^{-f(y)}}.$$

In order to get the Janossy densities of a DPP, it is worth noting that since K is assumed to be a non-negative hermitian function under Condition \mathcal{H} then, by Mercer's theorem, for all compact set $S \subset \mathbb{R}^d$ there exists a sequence of eigenvalues $\lambda_i^S \in [0, 1]$ and orthonormal eigenfunctions ϕ_i^S of $L^2(S, \mu)$ such that

$$K(x, y) = \sum_i \lambda_i^S \phi_i^S(x) \overline{\phi_i^S(y)}$$

for almost all $(x, y) \in S^2$. When all $\lambda_i^S \in [0, 1[$, we define the kernel

$$L_S(x, y) := \sum_i \frac{\lambda_i^S}{1 - \lambda_i^S} \phi_i^S(x) \overline{\phi_i^S(y)}$$

and we can express the Janossy densities of X the following way:

Proposition 1.2.5 ([75, 97]). *Let X be a DPP with kernel K satisfying condition \mathcal{H} and let $S \subset \mathbb{R}^d$ be a compact set. If the eigenvalues of \mathcal{K}_S are in $[0, 1[$ then, for all integers n ,*

$$j_n^S(x_1, \dots, x_n) = \det(Id - \mathcal{K}_S) \det(L_S[x]).$$

Finally, DPPs are mostly studied under two different settings. When μ is a counting measure over a finite space then the DPP is said to be discrete. Discrete DPPs have found applications for example in machine learning [68] to remove redundancies and get a diverse subset of a dataset, in sampling theory [72] to get a fixed size sampling design with fixed first order inclusion probability and some control over the second order inclusion probability or in graph theory as the distribution of uniform spanning trees [28]. When μ is absolutely continuous with respect to the Lebesgue measure then the DPP is said to be continuous. Continuous DPPs also appear in various probability fields as the eigenvalue's distribution of some families of random matrices [46], as a combinatoric tool to study the asymptotic of the longest increasing subsequence of random perturbations [21], position of non-intersecting Brownian motions with given starting point and end point [67], zeros of some families of random analytic functions [82] and also as the distribution of Fermion systems in quantum physics [75] to give a few examples. From a statistical perspective, continuous DPPs have found applications in telecommunication [36, 76, 47] to model the spatial configuration of wireless networks, in biology [1] to model the spatial distribution of epidermal nerve fibers, in forestry [69] to model the spatial distribution of trees in a forest and in computational statistics as a tool to get fast Monte Carlo methods [11].

1.2.3 Parametric families of determinantal point processes

There exists several ways to create parametric families of DPPs. One of the most common one is the following result from [69] :

Proposition 1.2.6. *Let $C : \mathbb{R}^d \mapsto \mathbb{R}^d$ be a square integrable symmetric function on \mathbb{R}^d such that its Fourier transform*

$$\widehat{C} : x \mapsto \int_{\mathbb{R}^d} C(t) \exp(-2i\pi t \cdot x) dt$$

takes values in $[0, 1]$. Then, the function $K(x, y) = C(y - x)$ is a DPP kernel on (\mathbb{R}^d, μ) where μ is the Lebesgue measure.

Since symmetric functions C with non negative Fourier transform are covariance functions, this proposition implies that we can consider as many parametric families of DPPs as there are parametric families of covariance functions. The assumption that $\widehat{C} \leq 1$ simply adds a bound on the parameters of the family. A well-known example is the Gaussian-type DPP kernel family that writes

$$\left\{ K^{\rho, \alpha}(x, y) = \rho \exp\left(-\frac{\|y - x\|^2}{\alpha^2}\right), 0 \leq \rho(\sqrt{\pi}\alpha)^d \leq 1 \right\}$$

where $\|\cdot\|$ is the euclidean norm of \mathbb{R}^d . In this family, the parameter ρ controls the intensity of the point process while the parameter α controls its range of interaction. The constraint $\rho(\sqrt{\pi}\alpha)^d \leq 1$ is a consequence of the bound $\widehat{C} \leq 1$ and means that a trade-off is needed between the repulsiveness of the DPP and the number of points it can generate. More examples of parametric families of stationary DPP kernels can be found in [15, 69]

Furthermore, we can also construct non-stationary DPP families by considering location dependent thinning of stationary DPPs. The following result is an immediate consequence from the definition of DPPs:

Proposition 1.2.7. *Let X be a DPP with kernel K and $\rho : \mathbb{R}^d \rightarrow \mathbb{R}^d$ be a function taking values in $[0, 1]$. Let X' be the point process obtained by retaining all $x \in X$ independently and with probability $\rho(x)$, then X' is a DPP with kernel*

$$K'(x, y) := \sqrt{\rho(x)}K(x, y)\sqrt{\rho(y)}.$$

We can then work with parametric families of non-stationary DPPs with kernels of the form

$$\{K^\theta(x, y) := \sqrt{\rho_\theta(x)}C_\theta(y - x)\sqrt{\rho_\theta(y)}, \|\widehat{C}_\theta\|_\infty \|\rho_\theta\|_\infty < 1\}.$$

They are called second-order intensity reweighted stationary [6, 103] since their second order properties are translation-invariants even if their first order intensity is inhomogeneous.

Finally, a last general method to obtain parametric families of DPPs is to start from one DPP and consider all its thinned and scaled versions. This is, for example, how is defined the family of scaled β -Ginibre DPPs [36].

Proposition 1.2.8. *Let X be a DPP with kernel K , $\rho \in [0, 1]$ and $\alpha \in \mathbb{R}_+^*$. Let X' be the point process obtained from retaining, independently and with probability ρ , each point of X and then applying the homothety of ratio α to the remaining points. Then X' is a DPP with kernel*

$$K^{\rho, \alpha} = \frac{\rho}{\alpha^d} K\left(\frac{x}{\alpha}, \frac{y}{\alpha}\right).$$

1.3 Outline of the PHD

In this thesis, we focus on the asymptotic inference of continuous DPPs in a specific framework called increasing domain asymptotic. Let $(W_n)_{n \geq 0}$ be a sequence of increasing bounded subsets of \mathbb{R}^d , meaning that $W_n \subset W_{n+1}$ for all integers n . Unlike the classical iid framework, when working in the increasing domain asymptotic framework we consider a unique observation of a point process within a given window W_n . Therefore, the asymptotic is considered on the size of the window and, as a consequence, on the number of point observed. This is why we also usually consider that either $\bigcup_{n \geq 0} W_n = \mathbb{R}^d$ or, more simply, that the d -dimensional volume of W_n goes to ∞ . In most applications, windows are rectangular but in order to get results as general as possible, we will usually consider windows of any shape provided that their boundaries are not too distorted as to limit edge effects.

The main problem in the increasing domain asymptotic setting for DPPs is that, since all points repel each other, we need to control their spatial dependency. To this goal, we study mixing properties of DPPs. We use these results to deduce a general central limit theorem (CLT) for a wide class of statistics on both stationary and non-stationary DPPs. We then apply this CLT to get asymptotic results on moment-based estimating equations. Finally, we consider the problem of proving the consistency and approximating the maximum likelihood estimator of stationary DPPs. All results of the PHD are detailed in Section 1.5.

1.4 Quantifying dependency in a point process

1.4.1 α -mixing and β -mixing

There exists a lot of different ways to measure dependency, some of the most common ones can be found in [23]. We study in Chapters 2 and 3 α -mixing and β -mixing for DPPs, the two most basic notions of dependency of a stochastic process. Introduced by Rosenblatt in [91] for the α -mixing and by Rozanov and Volkonskii in [92] for the β -mixing, they have been extensively used in the literature to get asymptotic properties of random processes, random fields and point processes. Some examples can be found in [30, 86, 103] for α -mixing and [9, 101] for β -mixing.

The α -mixing and β -mixing coefficients of two σ -algebras are defined the following way:

Definition 1.4.1. *Consider a probability space $(E, \mathcal{T}, \mathbb{P})$ and \mathcal{A}, \mathcal{B} two sub σ -algebras of \mathcal{T} . Let $\mathbb{P}_{\mathcal{A}}$ and $\mathbb{P}_{\mathcal{B}}$ be the respective restrictions of \mathbb{P} to \mathcal{A} and \mathcal{B} and define the probability $\mathbb{P}_{\mathcal{A} \otimes \mathcal{B}}$ on the product σ -algebra by $\mathbb{P}_{\mathcal{A} \otimes \mathcal{B}}(A \times B) = \mathbb{P}(A \cap B)$ for all $A \in \mathcal{A}$ and $B \in \mathcal{B}$. The α -mixing coefficient and β -mixing coefficient are defined as the following measures of dependence between \mathcal{A} and \mathcal{B} :*

$$\begin{aligned}\alpha(\mathcal{A}, \mathcal{B}) &:= \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{A}, B \in \mathcal{B}\} \\ \beta(\mathcal{A}, \mathcal{B}) &:= \|\mathbb{P}_{\mathcal{A} \otimes \mathcal{B}} - \mathbb{P}_{\mathcal{A}} \otimes \mathbb{P}_{\mathcal{B}}\|_{TV}\end{aligned}$$

where $\|\cdot\|_{TV}$ is the total variation metric.

An immediate consequence of their definition is that $\alpha(\mathcal{A}, \mathcal{B}) = 0$ iff $\beta(\mathcal{A}, \mathcal{B}) = 0$ iff \mathcal{A} and \mathcal{B} are independent. So, the smaller the α -mixing and β -mixing are, the closer to independence the σ -algebras can be considered to be. Moreover, the mixing coefficients satisfy $2\alpha(\mathcal{A}, \mathcal{B}) \leq \beta(\mathcal{A}, \mathcal{B})$. α -mixing is more often used but β -mixing

implies stronger covariance inequalities [89] as well as a coupling theorem known as Berbee's Lemma [13].

The definitions of the mixing coefficients are adapted to point processes the following way. We write $\sigma(X)$ for the σ -algebra generated by a random variable X and $\text{dist}(A, B) := \inf_{x \in A, y \in B} \|y - x\|$ for the distance between two sets A and B .

Definition 1.4.2. *Let X be a point process on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu)$. The α -mixing and β -mixing coefficients of X associated to the sizes p and q and to the distance r are defined by*

$$\begin{aligned}\alpha_{p,q}(r) &:= \sup\{\alpha(\sigma(X \cap A), \sigma(X \cap B)) : \mu(A) \leq p, \mu(B) \leq q, \text{dist}(A, B) > r\}, \\ \beta_{p,q}(r) &:= \sup\{\beta(\sigma(X \cap A), \sigma(X \cap B)) : \mu(A) \leq p, \mu(B) \leq q, \text{dist}(A, B) > r\},\end{aligned}$$

for all $p, q, r \in \mathbb{R}_+$. We also define

$$\begin{aligned}\alpha_{p,\infty}(r) &:= \sup_q \alpha_{p,q}(r), \quad \beta_{p,\infty}(r) := \sup_q \beta_{p,q}(r), \\ \alpha_{\infty,\infty}(r) &:= \sup_{p,q} \alpha_{p,q}(r), \quad \beta_{\infty,\infty}(r) := \sup_{p,q} \beta_{p,q}(r),\end{aligned}$$

for all $p \in \mathbb{R}_+$.

In most applications, the coefficients $\alpha_{\infty,\infty}(r)$ and $\beta_{\infty,\infty}(r)$ tend to be impractical to use. This is why the parameters p and q have been introduced to control the size of the sets considered in the definition of the mixing coefficients. However, the main information carried by these coefficients is their rate of decay with respect to r . It shows how fast events that happen further and further away from each others become independent. As an example of application, one of the most well-known result on α -mixing random variables is the central limit theorem by Bolthausen [20] that can be adapted to point processes under the main assumption that

$$\forall p \in \mathbb{R}_+, \quad \sum_{r \in \mathbb{N}} r^{d-1} \alpha_{p,\infty}(r) < +\infty.$$

1.4.2 Positive and negative association

A different kind of dependency property satisfied by DPPs is negative association. It is defined the following way:

Definition 1.4.3. *A point process X on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu)$ is called negatively associated if it satisfies for all disjoint sets $A, B \in \mathcal{B}_0(\mathbb{R}^d)$ and functions $f, g : \Omega \rightarrow \mathbb{R}$ increasing with respect to the inclusion (i.e. $X \subset Y \Rightarrow f(X) \leq f(Y)$),*

$$\text{Cov}(f(X \cap A), g(X \cap B)) \leq 0.$$

A direct consequence of the negative association property is that the random variables $N(A)$ and $N(B)$ are negatively correlated for any pair of bounded disjoint sets A, B . Since the covariance between $N(A)$ and $N(B)$ can be written as

$$\text{Cov}(N(A), N(B)) = \int_{A \times B} (\rho_2(x, y) - \rho_1(x)\rho_1(y)) d\mu(x) d\mu(y)$$

therefore $\rho_2(x, y) - \rho_1(x)\rho_1(y)$ is negative almost everywhere as a consequence of negative association which shows that any pair of points will repel each other. Thus, negative association can be interpreted as a stronger property than pairwise repulsiveness. The negative association of discrete DPPs has been proved in [73] using algebraic

tools. A more probabilistic proof is provided in Appendix A.1. As for continuous DPPs, one can deduce from the previous references and [45, 74] that they remain negatively associated. Just like α -mixing and β -mixing, the negative association property has been used to prove central limit theorems but only in the case of random fields [27]. Instead, it has been used in [18, 19] to deduce various properties of random geometric models driven by DPPs. We will show in Chapter 2 that a CLT for point processes can also be obtained from negative association.

Also of note is the analogous property for attractive point processes which is called positive association and defined the following way:

Definition 1.4.4. *A point process X on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu)$ is called positively associated if it satisfies for all sets $A, B \in \mathcal{B}_0(\mathbb{R}^d)$ and increasing functions $f, g : \Omega \rightarrow \mathbb{R}$*

$$\text{Cov}(f(X \cap A), g(X \cap B)) \geq 0.$$

Note that the sets A and B are not restricted to be disjoint in the definition of positive association compared to negative association. This is because $\text{Cov}(f(X \cap A), f(X \cap A))$ is always non-negative which is why we only consider disjoint sets for negative association but not necessarily for positive association. Another difference is that positive association can be deduced from the FKG inequality for discrete point processes and its analogous equivalent [29] for continuous point processes while no such result exists for negative association. This is why positive association tends to be easier to prove than negative association in general.

1.5 Presentation of the PHD's results

In Chapter 2 we show a general central limit theorem for functionals of stationary and non-stationary DPPs that write as a sum of a test function over all subsets of the point process. In Chapter 3 we prove a β -mixing inequality for point processes that only depends on their joint intensity functions and apply it to DPPs. In Chapter 4 we prove the asymptotic normality of estimators of stationary and non-stationary DPPs built from a wide class of estimating functions. Finally, we study the asymptotic properties of the maximum likelihood estimator for stationary DPPs in Chapter 5 and give a tractable asymptotic approximation of the DPPs likelihood as well as a proof of its consistency.

1.5.1 Chapter 2

We start by showing that association (positive or negative) of a point process implies a general covariance inequality that only depends on the covariance between the number of points of the point process between different sets:

Theorem 1.5.1. *Let X be an associated point process and $A, B \subset \mathbb{R}^d$ be two disjoint bounded subsets. Let $f : \Omega \rightarrow \mathbb{R}$ and $g : \Omega \rightarrow \mathbb{R}$ be two functions such that $f(X \cap A)$ and $g(X \cap B)$ are bounded, then*

$$|\text{Cov}(f(X \cap A), g(X \cap B))| \leq \|f\|_A \|g\|_B |\text{Cov}(N(A), N(B))|$$

where $\|\cdot\|_A$ is the Lipschitz semi-norm defined by

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

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

As a consequence, we get that association implies an α -mixing bound that only depends on the first two joint intensity functions of the point process:

Proposition 1.5.2. *Let X be an associated point process on \mathbb{R}^d whose first two intensity functions are well-defined, then for all $p, q > 0$,*

$$\begin{cases} \alpha_{p,q}(r) \leq pq \sup_{\|x-y\| \geq r} |\rho_2(x,y) - \rho(x)\rho(y)|, \\ \alpha_{p,\infty}(r) \leq ps_d \int_r^\infty t^{d-1} \sup_{\|x-y\|=t} |\rho_2(x,y) - \rho(x)\rho(y)| dt. \end{cases}$$

where s_d is the area of the d -dimensional sphere.

Therefore, the rate of decay of the α -mixing coefficients in each cases only depends on the pairwise repulsion of the point process. In the case of DPPs, since $|\rho_2(x,y) - \rho(x)\rho(y)| = |K(x,y)|^2$ then we deduce that the $\alpha_{p,q}(r)$ coefficients decay at least at the same rate than the kernel squared does when $\|y-x\|$ goes to infinity. We even show that this rate of decay is optimal for a wide class of DPPs by giving a lower bound on the α -mixing coefficients with the same rate of decay. Unfortunately, the rates of decay of the $\alpha_{p,\infty}(r)$ coefficients are not as good and applying the CLT by Bolthausen and Guyon [20, 50] for α -mixing random variables would force us to exclude the most repulsive DPPs that have the slowest decaying kernel. In order to solve this issue, we prove a more general CLT, based not on α -mixing but on the covariance inequality in Theorem 1.5.1, which can be applied to all the parametric families of DPP kernels commonly used in spatial statistics.

Theorem 1.5.3. *Let X be a DPP with bounded kernel K satisfying \mathcal{H} . Let $\tau > 0$ and $f : \Omega \rightarrow \mathbb{R}$ be a function of the form*

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

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

$$(H1) \quad |\partial W_n \oplus (\tau + \nu)| = o(|W_n|);$$

$$(H2) \quad \sup_{\|y-x\| > r} |K(x,y)| = o(r^{-(d+\varepsilon)/2});$$

$$(H3) \quad \liminf_n |W_n|^{-1} \sigma_n^2 > 0.$$

Then,

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

Here, $\partial W_n \oplus (\tau + \nu)$ denotes the set of points at distance less than $\tau + \nu$ from the boundary of W_n .

1.5.2 Chapter 3

In this chapter, we show the following general inequality on β -mixing coefficients of point processes depending uniquely on their n -th order intensity functions:

Theorem 1.5.4. *Let X be a point process on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu)$ such that $\mathbb{E}[4^{|X \cap A|}] < +\infty$ for all bounded sets $A \subset \mathbb{R}^d$. Then, for all $p, q, r \in \mathbb{R}_+$,*

$$\beta_{p,q}(r) \leq \sup_{\substack{\mu(A) < p, \mu(B) < q \\ \text{dist}(A,B) > r}} \left(\sum_{m,n=0}^{+\infty} \frac{2^{n+m-1}}{m!n!} \int_{A^m \times B^n} |\rho_m(x)\rho_n(y) - \rho_{m+n}(x,y)| d\mu^m(x)d\mu^n(y) \right).$$

Since the expressions $(\rho_m(x)\rho_n(y) - \rho_{m+n}(x,y))/\rho_m(x)\rho_n(y)$ can be interpreted as the correlation between X having points at the locations x_1, \dots, x_m and X having points at the locations y_1, \dots, y_n then Theorem 1.5.4 shows that the faster these correlations decrease when the distance between the points goes to infinity the faster the decay of the β -mixing coefficients is. When applied to determinantal point processes, we get the following β -mixing inequality:

Theorem 1.5.5. *Let X be a DPP with bounded kernel K satisfying \mathcal{H} , then*

$$\beta_{p,q}(r) \leq 2pq(1 + 2p\|K\|_\infty)(1 + 2q\|K\|_\infty)e^{2\|K\|_\infty(p+q)} \sup_{\|y-x\| \geq r} |K(x,y)|^2.$$

Note that the rate of decay of the β -mixing coefficients in Theorem 1.5.5 is the same as the rate of decay of the α -mixing coefficients in Proposition 1.5.2. The main drawback of this result, compared to the α -mixing inequality in Proposition 1.5.2, is the exponential dependency of the bound on the $\beta_{p,q}(r)$ with respect to p, q making it impractical to use in most applications.

1.5.3 Chapter 4

Chapter 4 focuses on asymptotic properties of estimators built from estimating functions. Given a family of continuous point processes $\{\mathbb{P}^\theta, \theta \in \Theta\}$, where $\Theta \subset \mathbb{R}^p$ for some $p \in \mathbb{N}$, and an increasing sequence of observation windows $(W_n)_{n \geq 0}$, we say that a function e_n is an estimating function when θ is estimated by a solution (or one of the solutions) $\hat{\theta}_n$ of the equation $e(\theta_n) = 0$. This is a very general statistical framework that encompasses other methods like minimum contrast estimation (when the contrast is differentiable) or maximum likelihood estimation (when the likelihood is differentiable). In particular, we say that e_n is a k -th order estimating function if it is of the form

$$e_n(X) = \sum_{x_1, \dots, x_k \in X \cap W_n}^{\neq} f_0(x_1, \dots, x_k; \theta) - \int_{W_n^k} f_0(t; \theta) \rho_k(t; \theta) d\mu(t),$$

where $f_0 : (\mathbb{R}^d)^k \times \Theta \rightarrow \mathbb{R}$ is a test function. In this chapter, we consider the more general framework of combinations of estimating functions of various orders:

$$e_n(\theta) = \begin{pmatrix} \sum_{x_1, \dots, x_{q_1} \in X \cap W_n}^{\neq} f_1(x_1, \dots, x_{q_1}; \theta) - \int_{W_n^{q_1}} f_1(x; \theta) \rho^{(q_1)}(x; \theta) d\mu(x) \\ \vdots \\ \sum_{x_1, \dots, x_{q_l} \in X \cap W_n}^{\neq} f_l(x_1, \dots, x_{q_l}; \theta) - \int_{W_n^{q_l}} f_l(x; \theta) \rho^{(q_l)}(x; \theta) d\mu(x) \end{pmatrix}.$$

where each $f_i : (\mathbb{R}^d)^{q_i} \times \Theta \rightarrow \mathbb{R}^{k_i}$ is a test function and $\sum_i k_i = p$. We show that under some regularity assumptions on the observation windows, the test functions and the point process then there exists a $\mu(W_n)$ -consistent sequence of roots $\hat{\theta}_n$ of the equation $e_n(\theta) = 0$ that is also asymptotically normally distributed. More precisely, for all $\varepsilon > 0$, there exists $A > 0$ such that

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

for a sufficiently large n and

$$\mu(W_n)\text{Var}(e_n(\theta^*))^{-1/2}H_n(\theta^*)(\hat{\theta}_n - \theta^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I_p),$$

where I_p is the $p \times p$ identity matrix and

$$H_n(\theta^*) := \frac{1}{\mu(W_n)} \begin{pmatrix} \int_{W_n^{q_1}} f_1(x; \theta^*) \nabla_{\theta} \rho^{(q_1)}(x; \theta^*)^T d\mu(x) \\ \vdots \\ \int_{W_n^{q_l}} f_l(x; \theta^*) \nabla_{\theta} \rho^{(q_l)}(x; \theta^*)^T d\mu(x) \end{pmatrix}.$$

We then focus in particular on the case of second-order estimating functions ($l = 1$ and $q_1 = 2$) and two-step estimation ($l = 2$, $q_1 = 1$ and $q_2 = 2$) of determinantal point processes and show how the various assumptions on X and the test functions f_i are simplified in both cases.

For second-order estimating functions, one of the main assumptions considered on the test function f is that $f(u, v) = 0$ if $\|v - u\|$ is greater than some constant R . In practical applications, the choice of this R is very important. If R is chosen too small then the estimating function will ignore a lot of pairs of points but taking R to high will make the estimating functions consider pair of points that are so far from each other they are nearly uncorrelated and will contribute mostly as noise. If we define $g(u, v; \theta) := \rho_2(u, v; \theta) / \rho_1(u; \theta)\rho_1(v; \theta)$ as the pair correlation function between u and v then we know that $g(u, v; \theta)$ converges towards 1 when the distance between u and v goes to infinity. Following the idea that we only want to take into account pair of points that are correlated enough, we want to keep pairs of point (u, v) such that $g(u, v; \theta) - 1$ is higher than some fraction of $g(u, u; \theta) - 1$ and $g(v, v; \theta) - 1$. This means that we want to choose R such that

$$\|v - u\| > R \Leftrightarrow \left| \frac{g(u, v; \theta^*) - 1}{g(u, u; \theta^*) - 1} \right| < \varepsilon \text{ and } \left| \frac{g(u, v; \theta^*) - 1}{g(v, v; \theta^*) - 1} \right| < \varepsilon$$

for a small ε (for example $\varepsilon = 1\%$). A common approach to choose such an R is to inspect a non-parametric estimation of the pair correlation function as in [8, 51, 103]. The issue with this approach is that, not only it requires extra work from the user, this complex method of choosing R makes it difficult to get any theoretical result on the estimating equation. In order to solve this issue, we propose instead to modify the test function to be of the form

$$f_{\text{adap}}(u, v; \theta) = w \left(\varepsilon \frac{\max(g(u, u; \theta) - 1, g(v, v; \theta) - 1)}{g(u, v; \theta) - 1} \right) f_0(u, v; \theta)$$

for some test function f_0 and weight function w of bounded support $[-1, 1]$. Then, we show that not only f_{adap} fits our consistency result and central limit theorem but we show in some simulation studies that it performs as well as test functions of the form

$$f(u, v; \theta) = \mathbb{1}_{\|v-u\| \leq R} f_0(u, v; \theta)$$

with an optimal choice of R but without the need for the user to find this optimal R .

1.5.4 Chapter 5

In the final chapter of this manuscript, we consider the problem of maximum likelihood estimation (MLE for short) of stationary determinantal point processes. Given a family $\{\mathbb{P}^\theta, \theta \in \Theta\}$ of continuous DPPs on an increasing sequence of compact observation

windows $(W_n)_{n \geq 1}$, as a consequence of Proposition 1.2.5 the (normalized) log-likelihood of this DPP family can be expressed as

$$l_n(\theta|X) = 1 + \frac{1}{\mu(W_n)} (\log \det(Id - \mathcal{K}_{W_n}^\theta) + \log \det(L_{W_n}^\theta[X \cap W_n])).$$

Both the expression of $L_{W_n}^\theta$ and the Fredholm determinant of $Id - \mathcal{K}_{W_n}^\theta$ depend on the spectral decomposition of the projection of \mathcal{K} on $L^2(W_n, \mu)$. Computing this spectral decomposition is not an issue for discrete DPPs where \mathcal{K} is a finite matrix but cannot be done for continuous DPPs. We propose in this chapter an asymptotic approximation of the likelihood of continuous stationary DPPs based on the fact that stationary DPPs have kernels of the form $K(x, y) = K_0(y - x)$ and when K_0 is square integrable, it admits a Fourier transform taking values in $[0, 1]$ by Proposition 1.2.6. We use the following asymptotic approximations:

$$L_{W_n}(x, y) \approx \int_{\mathbb{R}^d} \frac{\hat{K}_0(t)}{1 - \hat{K}_0(t)} \exp(2i\pi t \cdot (y - x)) d\mu(t)$$

and

$$\frac{1}{\mu(W_n)} \log \det(Id - \mathcal{K}_{W_n}^\theta) \approx \int_{\mathbb{R}^d} \log(1 - \hat{K}_0(x)) d\mu(x),$$

to get an approximation $\tilde{l}_n(\theta|X)$ of $l_n(\theta|X)$ that does not depend on the spectral decomposition of the kernels but only on their Fourier transform which are explicitly known for all classical parametric families of stationary DPP kernels used in spatial statistics.

The main issue with an asymptotic approximation of the likelihood is that it ignores any kind of edge effect of the DPP, worsening the inference of the most repulsive DPPs that have the stronger edge effects. As a fix to this issue in the case of rectangular window, we consider the idea of replacing the window with a flat torus, bringing points on the edge closer to each other in order to mitigate the edge effect. We show in a simulation study that the approximated likelihood $\tilde{l}_n(\theta|X)$ with the edge effect correction gives better result than DPP inference with more common moment methods like minimum contrast estimation based on the pair correlation function.

Another issue with the MLE is the difficulty to get any theoretical result about its consistency. One of the main problem is that $\det(L_{W_n}^\theta[X])$ vanishes when two points of X get arbitrarily close to each other. Unfortunately, no relationship between how close some points of X are from each other and the value of the determinant is known, making the likelihood difficult to control unless we can force all points of the DPP to be at a distance at least ε from each other. This motivates the idea to work with approximations of DPPs on $(\varepsilon\mathbb{Z})^d$, $\varepsilon > 0$, an arbitrary small regular grid of \mathbb{R}^d , with kernels $\varepsilon^d K^\theta[(\varepsilon\mathbb{Z})^d]$. We show that for any small enough $\varepsilon > 0$ and under weak assumptions on the DPP family $\{\mathbb{P}^\theta, \theta \in \Theta\}$ the MLE and our approximated MLE are equivalent and converges to the true parameter θ .

1.5.5 Appendix

In this section, we give two complementary results to the PHD. In [74], Lyons proved the negative association of general DPPs as a consequence of the negative association of discrete DPPs. The proof of the negative association of discrete DPPs is done using various exterior algebra properties making the proof difficult to understand for those who are not familiar with this algebraic formalism. Our first result is a modified version of Lyon's proof that only uses probabilistic tools.

Our second result is a generalization of the Cramér-Wold device. The Cramér-Wold device is used, among others, as a way to generalize unidimensional CLTs to multidimensional CLTs. Unfortunately, it only works when the variance in the CLT is convergent which is not assumed in Theorems 2.3.1 and 2.4.4 where we need it. Therefore, we show that the Cramér-Wold device can be extended to the case of non-convergent variance.

Chapter 2

Mixing properties and central limit theorem for associated point processes

This chapter is an article [84] accepted for publication in *Bernoulli* journal so some considerations and definitions are redundant with the introduction. For the same reason, some notations are specific to this chapter.

2.1 Introduction

Positive association (PA) and negative association (NA) [3, 41] are properties that quantify the dependence between random variables. They have found many applications in limit theorems for random fields [26, 104]. Even if the extension of PA to point processes have been used in analysis of functionals of random measures [29, 42], there are no general applications of PA or NA to limit theorems for point processes. We contribute in this chapter 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 [75] 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 [61]). From a statistical perspective, DPPs have applications in machine learning [68], telecommunication [36, 76, 47], biology, forestry [69] and computational statistics [11].

As a first result, we relate the association property of a point process to its α -mixing properties. First introduced in [91], 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 [20, 37, 50, 62, 91]. More details about mixing can be found in [23, 37]. We derive in Section 2.2 an important covariance inequality for associated point processes (Theorem 2.2.5), that turns out to be very similar to inequalities established in [39] for weakly dependent continuous random processes. We show that this inequality implies α -mixing and precisely allows one 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 [26]). However, this implication holds true for integer-valued random fields (see [38] or [26]). As explained in [38], 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 2.3 a general central limit theorem (CLT) for random fields defined as a function of an associated point process (Theorem 2.3.1). A standard method for proving this kind of theorem 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 one 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 [15]).

Section 2.4 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 2.4.4). 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 [97] when $p = 1$, and in [14] for stationary DPPs and $p \geq 1$. 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 [103], designed for this kind of inhomogeneous point process models, is consistent and asymptotically normal when applied to DPPs.

2.2 Associated point processes and α -mixing

2.2.1 Notation

In this chapter, 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 [31, 77]. 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

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

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

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

The notation $|\cdot|$ 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 $A \subset \mathbb{R}^d$, $|A| := \int_A dx$ 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 $\text{dist}(A, B)$ as $\inf_{x \in A, y \in B} |y - x|$ and $\text{diam}(A)$ as $\sup_{x, y \in A} |y - x|$ where $|\cdot|$ 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 centered at x with radius r and $\|\cdot\|_p$ for the p -norm of random variables and functions where $1 \leq p \leq \infty$.

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

Definition 2.2.1. Let $X \in \Omega$ and $n \geq 1$ be an integer. If there exists a non-negative function $\rho_n : (\mathbb{R}^d)^n \rightarrow \mathbb{R}$ such that

$$\mathbb{E} \left[\sum_{x_1, \dots, x_n \in X}^{\neq} f(x_1, \dots, x_n) \right] = \int_{(\mathbb{R}^d)^n} f(x_1, \dots, x_n) \rho_n(x_1, \dots, x_n) dx_1 \cdots dx_n$$

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

In particular, $\rho_n(x_1, \dots, x_n) dx_1 \cdots dx_n$ can be viewed as the probability that X has a point in each of the infinitesimally small sets around x_1, \dots, x_n with volumes dx_1, \dots, dx_n respectively.

We further introduce the notation

$$D(x, y) := \rho_2(x, y) - \rho_1(x)\rho_1(y). \quad (2.2.1)$$

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

$$\text{Cov}(N(A), N(B)) = \int_{A \times B} D(x, y) dx dy. \quad (2.2.2)$$

2.2.2 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 [19] for example).

Definition 2.2.2. A point process X is said to be negatively associated (NA for short) if, for all families of pairwise disjoint Borel sets $(A_i)_{1 \leq i \leq k}$ and $(B_i)_{1 \leq i \leq l}$ such that

$$(\cup_i A_i) \cap (\cup_j B_j) = \emptyset \quad (2.2.3)$$

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

$$\begin{aligned} \mathbb{E}[F(N(A_1), \dots, N(A_k))G(N(B_1), \dots, N(B_l))] \\ \leq \mathbb{E}[F(N(A_1), \dots, N(A_k))]\mathbb{E}[G(N(B_1), \dots, N(B_l))]. \end{aligned} \quad (2.2.4)$$

Similarly, a point process is said to be positively associated (PA for short) if it satisfies the reverse inequality for all families of pairwise disjoint Borel sets $(A_i)_{1 \leq i \leq k}$ and $(B_i)_{1 \leq i \leq l}$ (but not necessarily satisfying (2.2.3)).

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 (2.2.3) that only affects NA point processes. Notice that without (2.2.3), $\mathbb{E}[N(A)]^2 \leq \mathbb{E}[N(A)^2]$ contradicts (2.2.4) 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 $X \leq Y$ 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 [74, Lemma 3.6] but we give an alternative elementary one in Appendix 2.A.

Theorem 2.2.3. *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*

$$\mathbb{E}[F(X \cap A)G(X \cap B)] \leq \mathbb{E}[F(X \cap A)]\mathbb{E}[G(X \cap B)]. \quad (2.2.5)$$

If X is PA then, for all $A, B \subset \mathbb{R}^d$ not necessarily disjoint,

$$\mathbb{E}[F(X \cap A)G(X \cap B)] \geq \mathbb{E}[F(X \cap A)]\mathbb{E}[G(X \cap B)]. \quad (2.2.6)$$

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 2.2.4. *For any $A \subset \mathbb{R}^d$, $\|\cdot\|_A$ 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 $\|\cdot\|_A$ 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 2.2.5. *Let X be an associated point process and $A, B \subset \mathbb{R}^d$ be two disjoint bounded subsets. Let $f : \Omega \rightarrow \mathbb{R}$ and $g : \Omega \rightarrow \mathbb{R}$ be two functions such that $f(X \cap A)$ and $g(X \cap B)$ are bounded, then*

$$|\text{Cov}(f(X \cap A), g(X \cap B))| \leq \|f\|_A \|g\|_B |\text{Cov}(N(A), N(B))|. \quad (2.2.7)$$

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 [25] 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 $f_+, f_- : \Omega \rightarrow \mathbb{R}$, $\mathcal{E}(A)$ -measurable, and $g_+, g_- : \Omega \rightarrow \mathbb{R}$, $\mathcal{E}(B)$ -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 $x \in A \setminus X$, $f_+(X \cup \{x\}) - f_+(X) = f(X \cup \{x\} \cap A) - f(X \cap A) + \|f\|_A$ 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 , (2.2.5) 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 (2.2.5)

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

Since these expressions are equal to

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

adding these two expressions together yields the upper bound in (2.2.7):

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

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

A similar inequality as in Theorem 2.2.5 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 $\Im(f)$ refer to the real and imaginary part of f respectively.

Corollary 2.2.6. *Let X be an associated point process and $A, B \subset \mathbb{R}^d$ be two disjoint bounded subsets. Let $f : \Omega \rightarrow \mathbb{C}$ and $g : \Omega \rightarrow \mathbb{C}$ be two functions such that $f(X \cap A)$ and $g(X \cap B)$ are bounded, then*

$$|\text{Cov}(f(X \cap A), g(X \cap B))| \leq 4\|f\|_A \|g\|_B |\text{Cov}(N(A), N(B))|.$$

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 (2.2.1) is well-defined. As a consequence of Theorem 2.2.5 and from (2.2.2), 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 $\text{dist}(A, B)$ tends to infinity, as specified by the following corollary.

Corollary 2.2.7. *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 $\text{dist}(A, B) \geq r$. Then, for all functions $f : \Omega \rightarrow \mathbb{R}$ and $g : \Omega \rightarrow \mathbb{R}$ such that $f(X \cap A)$ and $g(X \cap B)$ are bounded,*

$$|\text{Cov}(f(X \cap A), g(X \cap B))| \leq s_d |A| \|f\|_A \|g\|_B \int_r^\infty t^{d-1} \sup_{|x-y|=t} |D(x, y)| dt, \quad (2.2.8)$$

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 $\text{dist}(A, B) \geq r$ then, from (2.2.2),

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

The final result is then a consequence of Theorem 2.2.5 and Corollary 2.2.6. \square

2.2.3 Application to α -mixing

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

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

In particular, \mathcal{A} and \mathcal{B} are independent iff $\alpha(\mathcal{A}, \mathcal{B}) = 0$. This definition leads to the essential covariance inequality due to Davydov [32] and later generalized by Rio [88]: For all random variables X, Y measurable with respect to \mathcal{A} and \mathcal{B} respectively,

$$|\text{Cov}(X, Y)| \leq 8\alpha^{1/r}(\mathcal{A}, \mathcal{B})\|X\|_p\|Y\|_q, \quad \text{where } p, q, r \in [1, \infty] \text{ and } \frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1. \quad (2.2.9)$$

This definition is adapted to random fields the following way (see [37] or [50]). Let $Y = (Y_i)_{i \in \mathbb{Z}^d}$ be a random fields on \mathbb{Z}^d and define

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

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

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

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

As a consequence of Corollary 2.2.7, the α -mixing coefficients of an associated point process tend to 0 when $D(x, y)$ vanishes fast enough as $|y - x|$ goes to infinity. More precisely, we have the following inequalities.

Proposition 2.2.8. *Let X be an associated point process on \mathbb{R}^d whose first two intensity functions are well-defined, then for all $p, q > 0$,*

$$\begin{cases} \alpha_{p,q}(r) \leq pq \sup_{|x-y| \geq r} |D(x, y)|, \\ \alpha_{p,\infty}(r) \leq ps_d \int_r^\infty t^{d-1} \sup_{|x-y|=t} |D(x, y)| dt. \end{cases} \quad (2.2.10)$$

Proof. We can write

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

so Proposition 2.2.8 is a direct consequence of Theorem 2.2.5 and Corollary 2.2.7 applied to indicator functions. \square

2.3 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 center x_i and side length s , where

$s \geq R$ 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 centered random field $(Y_i)_{i \in \mathbb{Z}^d}$ defined by

$$Y_i := f_i(X \cap C_i) - \mathbb{E}[f_i(X \cap C_i)], \quad i \in \mathbb{Z}^d, \quad (2.3.1)$$

and we are interested in this section in 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 2.2.8, we could directly use one of the different CLT for α -mixing random fields that already exist in the literature [20, 37, 50] 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 (2.2.8). 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 2.4.2 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 [66, 54] and other papers. As an example, DPPs have been proved to be Brillinger mixing in [14, 53]. However, this condition applies to stationary point processes only.

Theorem 2.3.1. *Consider the random field Y given by (2.3.1), 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 $\sigma_n^2 := \text{Var}(S_n)$. 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) $\sup_{|x-y| \geq r} |D(x, y)| = \underset{r \rightarrow \infty}{o} (r^{-(d+\varepsilon)\frac{2+\delta}{\delta}})$ where D is given by (2.2.1);

(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 (2.3.1). Moreover, we have $\text{dist}(C_i, C_j) \geq \frac{1}{\sqrt{d}}(|i - j|_1 R - sd)$ as a consequence of Lemma 2.B.2, and since $|\bigcup_{i \in I} C_i| \leq s^d |I|$, this gives us the inequality

$$\forall p, q > 0, \quad \forall r > \frac{sd}{R}, \quad \alpha_{p,q}^Y(r) \leq \alpha_{ps^d, qs^d}^X \left(\frac{1}{\sqrt{d}}(rR - sd) \right),$$

where we denote by α^X, α^Y the α -mixing coefficients of X and Y respectively. In particular, conditions (C1), (C3) and identity (2.2.10) yields

$$\forall p, q > 0, \quad \alpha_{p,q}^Y(r) = \underset{r \rightarrow \infty}{o} (r^{-(d+\varepsilon)\frac{2+\delta}{\delta}}). \quad (2.3.2)$$

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 [20] and Guyon [50], while the second step exploits elements from Ibragimov and Linnik [62]. 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 $\sup_i \|Y_i\|_\infty := \sup_i \|f_i(\cdot \cap C_i)\|_\infty = M < \infty$ instead of Assumption (C2). Since $\alpha_{p,q}^Y(r)$ is non increasing in r and is a $o(r^{-d})$ by (2.3.2), we can choose a sequence $(r_n)_{n \in \mathbb{N}}$ such that

$$\alpha_{p,q}^Y(r_n) \sqrt{|I_n|} \rightarrow 0 \quad \text{and} \quad r_n^{-d} \sqrt{|I_n|} \rightarrow \infty.$$

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

$$S_{i,n} = \sum_{\substack{j \in I_n \\ |i-j|_1 \leq r_n}} Y_j, \quad S_{i,n}^* = S_n - S_{i,n}, \quad a_n = \sum_{i \in I_n} \mathbb{E}[Y_i S_{i,n}],$$

$$\bar{S}_n = \frac{1}{\sqrt{a_n}} S_n \quad \text{and} \quad \bar{S}_{i,n} = \frac{1}{\sqrt{a_n}} S_{i,n}.$$

We have $\sigma_n^2 = \text{Var}(S_n) = a_n + \sum_{i \in I_n} \mathbb{E}[Y_i S_{i,n}^*]$ and, as a consequence of the typical covariance inequality (2.2.9) for α -mixing random variables, we get

$$\begin{aligned} \left| \sum_{i \in I_n} \mathbb{E}[Y_i S_{i,n}^*] \right| &\leq \sum_{\substack{i,j \in I_n \\ |i-j|_1 > r_n}} |\text{Cov}(Y_i, Y_j)| \leq 8M^2 \sum_{\substack{i,j \in I_n \\ |i-j|_1 > r_n}} \alpha_{1,1}^Y(|i-j|_1) \\ &\leq 8M^2 |I_n| \sum_{r > r_n} |\{k \in \mathbb{Z}^d : |k|_1 = r\}| \alpha_{1,1}^Y(r). \end{aligned}$$

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 its values in $\{-r, \dots, r\}$ and the last coordinate is fixed by the other ones, up to the sign, since $|k|_1 = r$. Therefore,

$$\left| \sum_{i \in I_n} \mathbb{E}[Y_i S_{i,n}^*] \right| \leq 16M^2 |I_n| \sum_{r > r_n} (2r+1)^{d-1} \alpha_{1,1}^Y(r).$$

By Assumption (2.3.2), 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 this will be a consequence of the following condition (see [17, 20])

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

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

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

It was proved by Bolthausen [20] that $\mathbb{E}[A_1^2]$ and $\mathbb{E}[|A_2|]$ vanish when n goes to infinity if $\sum r^{d-1} \alpha_{p,q}^Y(r) < \infty$ for $p+q \leq 4$ which is the case here. We show that $\mathbb{E}[A_3]$ vanishes at infinity using (2.2.8). Notice that we have

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

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 $\text{dist}(C_j, B_j) \geq (Rr_n - sd)/\sqrt{d}$ (see Lemma 2.B.2). We have $\|f_j\|_{C_j} \leq 2M$ 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 contain x then

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

Lemma 2.B.2 gives us the bound $|J_x| \leq (2sd/R + 1)^d$ and thus $\|g_j\|_{B_j} \leq 2\lambda M(2sd/R + 1)^d/\sqrt{a_n}$. Finally, using Corollary 2.2.7 we get

$$\begin{aligned} |\mathbb{E}[A_3]| &\leq \frac{4|I_n|sd}{\sqrt{a_n}} |C_j| \|f_j\|_{C_j} \|g_j\|_{B_j} \int_{\text{dist}(B_j, C_j)}^{\infty} t^{d-1} \sup_{|x-y| \geq t} |D(x, y)| dt \\ &\leq 16s_d M^2 \left(\frac{2s^2d}{R} + s \right)^d \lambda \frac{|I_n|}{a_n} \int_{\frac{1}{\sqrt{d}}(Rr_n - sd)}^{\infty} t^{d-1} \sup_{|x-y| \geq t} |D(x, y)| dt. \end{aligned} \quad (2.3.3)$$

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

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

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

Let $\sigma_n^2(N) := \text{Var}(S_{1,n})$, 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 $N \rightarrow \infty$ and by assumption (C4) we have that $|I_n| \leq c\sigma_n^2$ for a sufficiently large n , where c is a positive constant. By (2.2.9),

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

By assumption (C3) and the choice of γ we have $\sum (2r+1)^{d-1} \alpha_{1,1}^Y(r)^{\frac{\delta\gamma}{2+\delta\gamma}} < \infty$ 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

$$\frac{1}{\sigma_n^2} |\text{Cov}(S_{1,n}, S_{2,n})| \leq 16 c M C_N \sum_{r=0}^{\infty} (2r+1)^{d-1} \alpha_{1,1}^Y(r)^{\frac{\delta\gamma}{2+\delta\gamma}},$$

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

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

concluding the proof. \square

2.4 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 2.4.3, we get the consistency and the asymptotic normality of the two-step estimation method of [103] for a parametric inhomogeneous DPP.

2.4.1 Negative association and α -mixing for DPPs

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

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

The function $K : (\mathbb{R}^d)^2 \rightarrow \mathbb{C}$ 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 $K : (\mathbb{R}^d)^2 \rightarrow \mathbb{C}$ 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 [61, 94, 74]. In particular, from [45, Theorem 1.4] and [74, Theorem 3.7], we know that DPPs are NA.

Theorem 2.4.1 ([45, 74]). *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 (2.2.1). Hence, using the last theorem and Proposition 2.2.8 we get the following strong mixing coefficients of a DPP, where we define

$$\omega(r) := \sup_{|x-y| \geq r} |K(x, y)|. \quad (2.4.1)$$

Corollary 2.4.2. *Let X be a DPP with kernel K satisfying \mathcal{H} . Then, for all $p, q > 0$,*

$$\begin{cases} \alpha_{p,q}(r) \leq \sup_{\substack{|A| < p, |B| < q \\ \text{dist}(A,B) > r}} \int_{A \times B} |K(x, y)|^2 dx dy \leq pq\omega(r)^2, \\ \alpha_{p,\infty}(r) \leq ps_d \int_r^\infty \omega^2(t) t^{d-1} dt. \end{cases} \quad (2.4.2)$$

It is worth noticing that this result, and so the covariance inequality (2.2.7), 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 $\sup_{\substack{|A| < p, |B| < q \\ \text{dist}(A,B) > r}} |\text{Cov}(N(A), N(B))|$ when r goes to infinity, as stated in the following proposition.

Proposition 2.4.3. *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 $\|\cdot\|$ is the operator norm. Then, for all $p, q, r > 0$,*

$$\begin{aligned} (1 - \|\mathcal{K}\|)^{\frac{(p+q)\|\mathcal{K}\|_\infty}{\|\mathcal{K}\|}} \sup_{\substack{|A| < p, |B| < q \\ \text{dist}(A,B) > r}} \int_{A \times B} |K(x, y)|^2 dx dy &\leq \alpha_{p,q}(r) \\ &\leq \sup_{\substack{|A| < p, |B| < q \\ \text{dist}(A,B) > r}} \int_{A \times B} |K(x, y)|^2 dx dy. \end{aligned} \quad (2.4.3)$$

Proof. The upper bound for $\alpha_{p,q}(r)$ is just the one in (2.4.2). 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 $\text{dist}(A, B) > r$. By definition, for any such sets A and B , $\alpha_{p,q}(r) \geq |\mathbb{P}(N(A) = 0)\mathbb{P}(N(B) = 0) - \mathbb{P}(N(A \cup B) = 0)|$. The void probabilities of DPPs are known (see [94]) and equal to

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

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

$$\begin{aligned} &\mathbb{P}(N(A) = 0)\mathbb{P}(N(B) = 0) - \mathbb{P}(N(A \cup B) = 0) \\ &= \exp\left(-\sum_{n \geq 1} \frac{\text{Tr}(\mathcal{K}_{A \cup B}^n)}{n}\right) \left(\exp\left(\sum_{n \geq 1} \frac{\text{Tr}(\mathcal{K}_{A \cup B}^n) - \text{Tr}(\mathcal{K}_A^n) - \text{Tr}(\mathcal{K}_B^n)}{n}\right) - 1\right) \\ &\geq \exp\left(-\sum_{n \geq 1} \frac{\text{Tr}(\mathcal{K}_{A \cup B}^n)}{n}\right) \sum_{n \geq 1} \frac{\text{Tr}(\mathcal{K}_{A \cup B}^n) - \text{Tr}(\mathcal{K}_A^n) - \text{Tr}(\mathcal{K}_B^n)}{n}. \end{aligned} \quad (2.4.4)$$

Using the classical trace inequality we get

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

thus

$$\exp\left(-\sum_{n \geq 1} \frac{\text{Tr}(\mathcal{K}_{A \cup B}^n)}{n}\right) \geq (1 - \|\mathcal{K}\|)^{\frac{(p+q)\|\mathcal{K}\|_\infty}{\|\mathcal{K}\|}}. \quad (2.4.5)$$

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

$$\begin{aligned} \text{Tr}(\mathcal{K}_{A \cup B}^n) - \text{Tr}(\mathcal{K}_A^n) - \text{Tr}(\mathcal{K}_B^n) &= \int_{(A \cup B)^n} K(x_1, x_2) \cdots K(x_{n-1}, x_n) K(x_n, x_1) dx_1 \cdots dx_n \\ &\quad - \int_{A^n \cup B^n} K(x_1, x_2) \cdots K(x_{n-1}, x_n) K(x_n, x_1) dx_1 \cdots dx_n, \end{aligned} \quad (2.4.6)$$

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 $n \geq 3$ since K is assumed to be non-negative. Finally, by combining (2.4.4), (2.4.5) and (2.4.6) we get the lower bound in (2.4.3). \square

2.4.2 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

$$f(X) := \sum_{S \subset X} f_0(S), \quad (2.4.7)$$

where f_0 is a bounded function vanishing when $\text{diam}(S) > \tau$ 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 (2.4.7) often takes the form

$$f(X) = \frac{1}{p!} \sum_{x_1, \dots, x_p \in X}^{\neq} f_0(x_1, \dots, x_p), \quad (2.4.8)$$

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 (2.4.8) has been investigated in [97] when $p = 1$ and in [14] for general p and stationary DPPs.

In the next theorem, we extend these settings to functionals like (2.4.7) applied to general non stationary DPPs. Some discussion and comments are provided after its proof. We use Minkowski's notation and write $A \oplus r$ for the set $\bigcup_{x \in A} \mathcal{B}(x, r)$.

Theorem 2.4.4. *Let X be a DPP associated to a kernel K that satisfies \mathcal{H} and that is further bounded. Let $\tau > 0$ and $f : \Omega \rightarrow \mathbb{R}$ be a function of the form*

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

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

$$(H1) \quad |\partial W_n \oplus (\tau + \nu)| = o(|W_n|);$$

$$(H2) \quad \omega(r) = o(r^{-(d+\varepsilon)/2});$$

$$(H3) \quad \liminf_n |W_n|^{-1} \sigma_n^2 > 0.$$

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 2.3.1, 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 2.3.1 to an auxiliary function, close to f , as follows. Define S^0 as the barycenter of the set S . We write

$$f_W(X) = \sum_{S \subset X} f_0(S) \mathbb{1}_W(S^0) \quad (2.4.9)$$

for the sum over the subsets of points of X with barycenter 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 $0 < s \leq \nu/\sqrt{d}$. 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 $\tilde{W}_n = \bigcup_{i \in I_n} C_i$. An illustration of these definitions is provided in Figure 3. Since $f_{\tilde{W}_n}(X) = \sum_{i \in I_n} f_{C_i}(X)$ and each f_{C_i} are $\mathcal{E}(C_i \oplus \tau)$ -measurable then $f_{\tilde{W}_n}$ is the ideal candidate to use Theorem 2.3.1 on. Thus, we first prove that the difference between $f_{\tilde{W}_n}$ and $f(X \cap W_n)$ is asymptotically negligible and then that $f_{\tilde{W}_n}$ satisfies the conditions of Theorem 2.3.1.

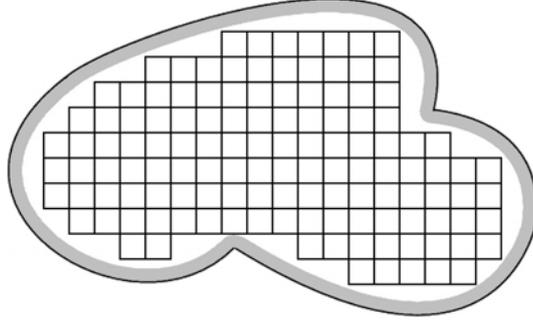


Figure 3: Example of illustration of the definition of \tilde{W}_n . Here, the black border is ∂W_n , the grey area corresponds to $(\partial W_n \oplus \tau) \cap W_n$ and the square lattice corresponds to \tilde{W}_n .

First of all, notice that $\text{dist}(C_i, \partial W_n) \geq \tau$ 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 \tilde{W}_n and we get

$$|W_n \setminus \tilde{W}_n| \leq |\partial W_n \oplus (\tau + s\sqrt{d})|.$$

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

$$f(X \cap W_n) - f_{\tilde{W}_n}(X) = \sum_{S \subset X \cap W_n} f_0(S) \mathbb{1}_{W_n \setminus \tilde{W}_n}(S^0). \quad (2.4.10)$$

Since f_0 vanishes when two points of S are at distance further than τ , then the sum in (2.4.10) only concerns the subsets S of $X \cap ((W_n \setminus \tilde{W}_n) \oplus \tau) \cap W_n$. By Lemma 2.B.6, the variance of $f(X \cap W_n) - f_{\tilde{W}_n}(X)$ is then a $O(|(W_n \setminus \tilde{W}_n) \oplus \tau|)$, 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 $\sigma_n^{-1}(f_{\tilde{W}_n}(X) - \mathbb{E}[f_{\tilde{W}_n}(X)])$. Moreover, we have

$$\begin{aligned} & |\text{Cov}(f_{\tilde{W}_n}(X), f(X \cap W_n) - f_{\tilde{W}_n}(X))| \\ & \leq \sigma_n \sqrt{\text{Var}(f(X \cap W_n) - f_{\tilde{W}_n}(X))} = \sigma_n o\left(\sqrt{|W_n|}\right) = o(\sigma_n^2) \end{aligned}$$

by Assumptions (H1), (H3) and Lemma 2.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 $\text{Var}(f_{\tilde{W}_n}(X))^{-1/2}(f_{\tilde{W}_n}(X) - \mathbb{E}[f_{\tilde{W}_n}(X)])$.

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 2.3.1. A rough bound on f gives us $|f_{C_i}(X)| \leq \|f_0\|_\infty 2^{N(C_i \oplus \tau)}$ so, by Lemma 2.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 $|I_n| = s^{-d} |\tilde{W}_n| = O(|W_n|)$ and $\text{Var}(f_{\tilde{W}_n}(X)) \sim \sigma_n^2$, we have

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

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

We highlight some extensions of this result.

- i) Since the superposition of independent PA (respectively NA) point processes remains a PA (respectively NA) point process, then Theorem 2.4.4 holds true for α -determinantal point processes where $\alpha \in \{-1/m : m \in \mathbb{N}^*\}$, see [94] for more information about α -DPPs.
- ii) Theorem 2.4.4 also extends to \mathbb{R}^q -valued functions f where $q \geq 2$. Let $\Sigma_n := \text{Var}(f(X \cap W_n))$. 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 2.4.4 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. Since Σ_n does not necessary converge, this result is not a direct application of the Cramér-Wold device. Instead, a detailed proof is given in Lemma A.2.1.

- iii) In (2.4.7), f_0 only depends on finite subsets of \mathbb{R}^d and not on the order of the points in each subset. Nonetheless, we can easily extend (2.4.7) to functions of the form

$$f(X) = \sum_{n \geq 0} \frac{1}{n!} \sum_{x_1, \dots, x_n \in X}^{\neq} f_0(x_1, \dots, x_n)$$

where f_0 is a bounded function on $\bigcup_{n \geq 0} (\mathbb{R}^d)^n$ that vanishes when two of its coordinates are at a distance greater than τ . Then f still satisfy Theorem 2.4.4. This is because we can write

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

where f_0^{sym} is the symmetrization of f_0 defined by

$$f_0^{sym}(\{x_1, \dots, x_n\}) := \frac{1}{n!} \sum_{\sigma \in S_n} f_0(x_{\sigma(1)}, \dots, x_{\sigma(n)})$$

where S_n is the symmetric group on $\{1, \dots, n\}$. Since f_0^{sym} is also bounded and vanishes when $\text{diam}(S) > \tau$ then it satisfies the required assumptions for Theorem 2.4.4.

Let us comment the assumptions of Theorem 2.4.4.

- Condition (H1) makes clear the idea that W_n must grow to \mathbb{R}^d as $n \rightarrow \infty$, 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 $|\Delta_n^{(k)}| \rightarrow \infty$ 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 [61]) have an exponential decay. Moreover, all translation-invariant kernels used in spatial statistics (see [69] and [15]) 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 [69, 15]), 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 2.3.1. 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 2.4.4, 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 [97] 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, [97] 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 2.B.7 that

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

implies (H3) and is much easier to verify.

2.4.3 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

$$K_{\beta, \psi}(x, y) = \sqrt{\rho_{\beta}(x)} C_{\psi}(y - x) \sqrt{\rho_{\beta}(y)}, \quad \forall x, y \in \mathbb{R}^2, \quad (2.4.11)$$

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. $\rho^{(n)}(x_1, \dots, x_n)/(\rho_\beta(x_1) \dots \rho_\beta(x_n))$, are translation-invariant for $n \geq 2$. 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 ρ_{\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 [69]. For later use, we call \mathcal{H}' this latter assumptions on $K_{\beta,\psi}$.

Consider the observation of a DPP X with kernel K_{β^*,ψ^*} , along with the covariates z , within a window $W_n := [an, bn] \times [cn, dn]$ where $b > a$ and $d > c$. Waagepetersen and Guan [103] 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

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

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

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

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

$$\mathcal{K}_\psi(t) := \int_{\|u\| \leq t} g_\psi(u) du$$

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 [103], 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 (2.4.11). This setting allows us to apply Theorem 2.4.4 and get rid of some restrictive mixing assumptions needed in [103].

The asymptotic covariance matrix of $(\hat{\beta}_n, \hat{\psi}_n)$ depends on two matrices defined in [103, Section 3.1], where they are denoted by Σ_n and I_n . We do not reproduce their expression, which is hardly tractable. An assumption in [103] 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 [103], 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 2.4.4.

Theorem 2.4.5. *Let X be a DPP with kernel K_{β^*, ψ^*} given by (2.4.11) 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 $r_l \geq 0$,

(W2) ρ_β and \mathcal{K}_ψ are twice continuously differentiable as functions of β and ψ ,

(W3) $\sup_{\|x\| \geq r} C_{\psi^*}(x) = O(r^{-1-\varepsilon})$,

(W4) Condition N3 in [103] (concerning the matrices I_n and $\tilde{\Sigma}_n$) is satisfied.

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

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

Proof. Let ρ_k be the k th intensity function of the DPP with kernel $(x, y) \mapsto C_{\psi^*}(y - x)$. In order to apply Theorem 1 in [103] we need to show that

(i) ρ_2, ρ_3 are bounded and there is a constant M such that for all $u_1, u_2 \in \mathbb{R}^2$,

$$\int |\rho_3(0, v, v + u_1) - \rho_1(0)\rho_2(0, u_1)|dv < M$$

and

$$\int |\rho_4(0, u_1, v, v + u_2) - \rho_2(0, u_1)\rho_2(0, u_2)|dv < M,$$

(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

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

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

$$\int_{\mathbb{R}^2} C_{\psi^*}(v)^2 dv \leq 2\pi \int_0^\infty r \sup_{\|x\|=r} |C_{\psi^*}(x)|^2 dr$$

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 2.4.2 decreases slower than what we desire. But, the only place this assumption is used in [103] is to get a CLT for some functionals of the same form as in our Theorem 2.4.4 leading to the asymptotic normality of $(\hat{\beta}_n, \hat{\psi}_n)$ in their Lemma 5. The same result can then also be derived as a consequence of our Theorem 2.4.4 with Assumption (W3). A formal proof of a more general result, namely Theorem 4.3.1, is given in Chapter 4. \square

2.A Proof of Theorem 2.2.3

We use the following variant of the monotone class theorem (see [34, Theorem 22.1]).

Theorem 2.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 $\tilde{1}$. Then, \mathcal{S} contains all bounded functions measurable over $\sigma(\mathcal{C})$.*

Now, let A, B_1, \dots, 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 configurations in A and we define \mathcal{S} as the set of functions $f : \Omega_A \mapsto \mathbb{R}$ such that

$$\mathbb{E}[\hat{f}(X \cap A)g(N(B_1), \dots, N(B_k))] \leq \mathbb{E}[\hat{f}(X \cap A)]\mathbb{E}[g(N(B_1), \dots, N(B_k))], \quad (2.A.1)$$

where $\hat{f}(X) := \sup_{Y \leq X} f(Y)$. 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 (2.2.4), we know that \mathcal{S} contains the set \mathcal{C} of functions of the form $f(N(A_1), \dots, N(A_k))$ where the A_i are pairwise disjoint Borel subsets of A . In particular, since point processes over A are generated by the set of random vectors $\{(N(A_1), \dots, N(A_k)) : A_i \subset A \text{ disjoint}, k \in \mathbb{N}\}$, then we only need to verify that \mathcal{S} and \mathcal{C} satisfy the hypothesis of Theorem 2.A.1 to conclude.

- Stability of \mathcal{S} by bounded monotonic convergence: Since (2.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 , $h \leq k \Rightarrow \hat{h} \leq \hat{k}$ and $h \geq k \Rightarrow \hat{h} \geq \hat{k}$. 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 $Y \subset X$, $f_n(Y) \leq f(Y)$. Taking the supremum then the limit gives us $g(X) \leq \hat{f}(X)$. Moreover, for all $Y \subset X$, $g(X) \geq \hat{f}_n(X) \geq f_n(Y)$. Taking the limit gives us that $g(X) \geq f(Y)$ for all $Y \subset X$ so $g(X) \geq \hat{f}(X)$ which proves that $g = \hat{f}$. Using the monotone convergence theorem we conclude that

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

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 2.B.1, \hat{f}_n also converges uniformly (and therefore in L^1) to \hat{f} . As a consequence, (2.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 $\tilde{1}$ so we only need to prove that \mathcal{C} is stable by multiplication. Let A_1, \dots, A_r and A'_1, \dots, A'_s be two sequences of pairwise distinct Borel subsets of A . Let $f \in \mathcal{C}$ of the form $f(N(A_1), \dots, N(A_r))$ and $h \in \mathcal{C}$ of the form $h(N(A'_1), \dots, N(A'_s))$. We can write

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

so $f \cdot h$ can be expressed as a function of the number of points in the subsets $A_i \setminus \cup_j A'_j$, $A'_i \setminus \cup_j A_j$ 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 satisfying $\mathbb{E}[f(X \cap A)g(X \cap B)] \leq \mathbb{E}[f(X \cap A)]\mathbb{E}[g(X \cap B)]$ for a fixed f we obtain the same result which concludes the proof.

2.B Auxiliary results

Lemma 2.B.1. *Let E be a set and $f, g : E \rightarrow \mathbb{R}$ be two functions, then*

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

Proof. The proposition becomes trivial once we write

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

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

Lemma 2.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 center $x_i = R \cdot i$ and $x_j = R \cdot j$. Then,*

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

Moreover, each cube intersects at most $(2sd/R)^d$ other cubes with centers on $R \cdot \mathbb{Z}^d$ and side length s .

Proof. Since each point of a d -dimensional square with side length s is at distance at most $s\sqrt{d}/2$ from its center, we get

$$\text{dist}(C_i, C_j) \geq \sqrt{(Ri_1 - Rj_1)^2 + \dots + (Ri_d - Rj_d)^2} - s\sqrt{d}$$

which takes its minimum when $|i_l - j_l| = r/d$ for all $1 \leq l \leq d$ hence $\text{dist}(C_i, C_j) \geq rR/\sqrt{d} - s\sqrt{d}$.

In particular, if $|i - j|_1 > sd/R$ then $C_i \cap C_j = \emptyset$, hence for all $i \in \mathbb{Z}^d$

$$|\{j \in \mathbb{Z}^d : C_i \cap C_j \neq \emptyset, i \neq j\}| \leq |\{j : 0 < |i - j|_1 \leq sd/R\}| \leq \left(\frac{2sd}{R} + 1\right)^d$$

\square

Lemma 2.B.3. *Let M and N be two $n \times n$ semi-positive definite matrices such that $0 \leq M \leq N^{-1}$ where \leq denotes the Loewner order. Then,*

$$\det(Id - MN) \geq 1 - \text{Tr}(MN)$$

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

$$\begin{aligned}
\det(Id - M) &= \prod_{\lambda \in \text{Sp}(M)} \exp(\log(1 - \lambda)) \\
&= \prod_{\lambda \in \text{Sp}(M)} \exp\left(-\sum_{n=0}^{\infty} \frac{\lambda^n}{n}\right) \\
&= \exp\left(-\sum_{n=0}^{\infty} \frac{\text{Tr}(M^n)}{n}\right) \\
&\geq \exp\left(-\sum_{n=0}^{\infty} \frac{\text{Tr}(M)^n}{n}\right) \\
&= \exp(\log(1 - \text{Tr}(M))) \\
&= 1 - \text{Tr}(M).
\end{aligned} \tag{2.B.1}$$

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 $0 \leq M \leq N^{-1}$ then $0 \leq SMS^T \leq Id$ and by applying (2.B.1) this concludes the proof:

$$\det(Id - MN) = \det(Id - SMS^T) \geq 1 - \text{Tr}(SMS^T) = 1 - \text{Tr}(MN).$$

□

Proposition 2.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 $\|A\|_{\infty} := \sup |a_{i,j}|$ for any matrix A . Then,*

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

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

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

while the inequality $N^T M_2^{-1} N \leq M_1$ gives us (see Lemma 2.B.3)

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

Therefore,

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

where $\Delta_i(M_1)$ means the (i, i) minor of the matrix M_1 and $\text{adj}(M_1)$ 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 $\Delta_i(M_1) \leq \prod_{j \neq i} M_1(j, j) \leq \|M\|_\infty^{k-1}$. 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 2.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

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

□

Lemma 2.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 $W \subset \mathbb{R}^d$ and all bounded functions $g : \bigcup_{p>0} (\mathbb{R}^d)^p \rightarrow \mathbb{R}$ such that $g(S)$ vanishes when $\text{diam}(S) > \tau$ for a given constant $\tau > 0$,*

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

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

$$\sum_{S \subset X \cap W} g(S) = \sum_{p \geq 0} \sum_{\substack{S \subset X \cap W \\ |S|=p}} g(S) \quad \text{a.s.}$$

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

$$\begin{aligned}
& \sum_{k=0}^p \mathbb{E} \left[\sum_{\substack{S, T \subset X \cap W \\ |S|=|T|=p, |S \cap T|=k}} g(S)g(T) \right] \\
&= \sum_{k=0}^p \mathbb{E} \left[\sum_{\substack{U \subset X \cap W \\ |U|=2p-k}} \sum_{\substack{S' \subset S \subset U \\ |S'|=k, |S|=p}} g(S)g(S' \cup (U \setminus S)) \right] \\
&= \sum_{k=0}^p \frac{1}{(2p-k)!} \int_{W^{2p-k}} \sum_{\substack{S' \subset S \subset \{x_1, \dots, x_{2p-k}\} \\ |S'|=k, |S|=p}} g(S)g(S' \cup (U \setminus S)) \rho_{2p-k}(x_1, \dots, x_{2p-k}) dx_1 \cdots dx_{2p-k} \\
&= \sum_{k=0}^p \frac{\binom{p}{k} \binom{2p-k}{p}}{(2p-k)!} \int_{W^{2p-k}} g(x_1, \dots, x_p) g(x_1, \dots, x_k, x_{p+1}, \dots, x_{2p-k}) \rho_{2p-k}(x) dx. \quad (2.B.3)
\end{aligned}$$

Since the determinant of a positive semi-definite matrix is smaller than the product of its diagonal terms, we have $|\rho_{2p-k}(x)| \leq \|K\|_\infty^{2p-k}$. Moreover, as a consequence of our assumptions on g , each term for $k \geq 1$ in (2.B.3) is bounded by

$$\begin{aligned}
& \frac{1}{p!(p-k)!} \binom{p}{k} \int_{W^{2p-k}} \|g\|_\infty^2 \|K\|_\infty^{2p-k} \mathbb{1}_{\{0 \leq |x_i - x_1| \leq \tau, \forall i\}} dx \\
&\leq \frac{|W|}{p!} \binom{p}{k} \|g\|_\infty^2 \|K\|_\infty^{2p-k} |\mathcal{B}(0, \tau)|^{2p-k-1} \\
&\leq \frac{|W|}{p!} \binom{p}{k} \|g\|_\infty^2 (1 + \|K\|_\infty)^{2p} (1 + |\mathcal{B}(0, \tau)|)^{2p}.
\end{aligned}$$

Hence,

$$\sum_{k=1}^p \mathbb{E} \left[\sum_{\substack{S, T \subset X \cap W \\ |S|=|T|=p, |S \cap T|=k}} g(S)g(T) \right] \leq |W| \|g\|_\infty^2 \frac{C_1^p}{p!} \quad (2.B.4)$$

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 $k \geq 1$ in (2.B.3) are $O(|W|)$, 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

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

Using Proposition 2.B.4, we get

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

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

$$\begin{aligned} \int_{W^p} \mathbb{1}_{\{0 < |x_k - x_j| \leq \tau, \forall j, k\}} |K(x_i, y)|^2 dx &\leq |\mathcal{B}(0, \tau)|^{p-1} \int_W |K(x_i, y)|^2 dx_i \\ &\leq |\mathcal{B}(0, \tau)|^{p-1} s_d \int_{\mathbb{R}^d} r^{d-1} \omega(r)^2 dr \end{aligned}$$

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

$$\begin{aligned} \int_{W^{2p}} g(x)g(y)|K(x_i, y_j)|^2 dx dy &\leq \|g\|_\infty |\mathcal{B}(0, \tau)|^{p-1} \int_{W^{p+1}} g(x)|K(x_i, y_1)|^2 dx dy_1 \\ &\leq |W| \|g\|_\infty^2 |\mathcal{B}(0, \tau)|^{2p-2} s_d \int_{\mathbb{R}^d} r^{d-1} \omega(r)^2 dr. \quad (2.B.5) \end{aligned}$$

By combining (2.B.4) and (2.B.5), we get the bound

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

where

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

is a constant independent from p and W . Finally,

$$\sum_{p \geq 0} \text{Var} \left(\sum_{\substack{S \subset X \cap W \\ |S|=p}} g(S) \right) = O(|W|)$$

and

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

which is $O(|W|)$ concluding the proof. \square

Proposition 2.B.7. *Let $p \in \mathbb{N}$, $f : \mathbb{R}^p \rightarrow \mathbb{R}_+$ 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 $\|K\| < 1$ where $\|K\|$ is the operator norm of the integral operator associated with K . If, for a given increasing sequence of compact sets $W_n \subset \mathbb{R}^d$,

$$\liminf_n \frac{1}{|W_n|} \int_{W_n^p} f(x) \det(K[x]) dx > 0, \quad (2.B.6)$$

then

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

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

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

We showed in Lemma 2.B.6 that $|W|^{-1}\text{Var}(N(W))$ is bounded by a constant $C > 0$ so we are only interested in the behaviour of $\text{Cov}(F(X \cap W), N(W))$. We start by developing $\mathbb{E}[F(X \cap W)N(W)]$ as

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

We also have

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

hence

$$\begin{aligned} & \text{Cov}(F(X \cap W), N(W)) \\ &= \frac{1}{p!} \int_{W^p} f(x) \det(K[x]) \left(p - \int_W (K(a, a) - \det(K[x, a]) \det(K[x])^{-1}) da \right) dx. \end{aligned} \tag{2.B.7}$$

Using Schur's complement, we get

$$K(a, a) - \det(K[x, a]) \det(K[x])^{-1} = K_{ax} K[x]^{-1} K_{ax}^T \tag{2.B.8}$$

where we define K_{ax} as the vector $(K(a, x_1), \dots, K(a, x_p))$. Moreover, since we look at our point process in a compact window W , a well-known property of DPPs (see [61]) 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, $\forall x, y \in W$,

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

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

$$\int_W K_{ax} K[x]^{-1} K_{ax}^T da = \text{Tr} \left(K[x]^{-1} \int_W K_{ax}^T K_{ax} da \right) = \text{Tr}(K[x]^{-1} L[x]) \leq p \|\mathcal{K}\|. \tag{2.B.9}$$

Finally, since f is non negative, by combining (2.B.7), (2.B.8) and (2.B.9) we get the lower bound

$$\begin{aligned} \text{Var}(F(X \cap W)) &\geq \frac{\text{Cov}(F(X \cap W), N(W))^2}{\text{Var}(N(W))} \\ &\geq \frac{(1 - \|\mathcal{K}\|)^2}{C(p-1)!^2 |W|} \left(\int_{W^p} f(x) \det(K[x]) dx \right)^2 \end{aligned}$$

which proves the proposition. □

Chapter 3

A bound of the β -mixing coefficient for point processes in terms of their intensity functions

This chapter is an article [83] published in *Statistics & Probability Letters* so some considerations and definitions are redundant with the introduction. For the same reason, some notations are specific to this chapter.

3.1 Introduction

In asymptotic inference for dependent random variables, it is necessary to quantify the dependence between σ -algebras. Some of the first measures of dependence that have been introduced are the alpha-mixing coefficients [91] and the beta-mixing coefficients [92]. They have been used to establish moment inequalities, exponential inequalities and central limit theorems for stochastic processes (see [23, 80, 89] for more details about mixing) with various applications in statistics, see for instance [22, 33]. In this chapter, we focus on spatial point processes. As detailed below, for these models, alpha-mixing has been widely studied and exploited in the literature, but not beta-mixing in spite of its stronger properties. In a lesser extent, some alternative measures of dependence have also been used for spatial point processes, namely Brillinger mixing [14, 58] (which only applies to stationary point processes but has been established in [58] under suitable conditions on the β -mixing coefficients) and association as in Chapter 2.

The main models used in spatial point processes are Gibbs point processes, Cox processes and determinantal point processes, see [79] for a recent review. An α -mixing inequality is established for Gibbs point processes in the Dobrushin uniqueness region in [43]. It has been used to show asymptotic normality of maximum likelihood and pseudo-likelihood estimates [65]. Similarly, some inhomogeneous Cox processes like the Neyman-Scott process have also been showed to satisfy α -mixing inequalities in [103]. These inequalities are at the core of asymptotic inference results in [30, 86, 103]. Finally, an α -mixing inequality has also been showed for determinantal point processes in (2.4.2) and used to get the asymptotic normality of a wide class of estimators of these models.

On the other hand, β -mixing is a stronger property than α -mixing. It implies stronger covariance inequalities [89] as well as a coupling theorem known as Berbee's Lemma [13] used in various limit theorems (for example in [9, 101]). Nevertheless, it rarely appears in the literature in comparison to α -mixing. This is especially true

for point processes where there has been no β -mixing property established for any of the above examples. Nevertheless, β -mixing coefficients have still been used several times in random geometry and point process statistics [52, 56, 55, 57]. In particular, it is argued in [55] that the β -mixing coefficient cannot be replaced by the α -mixing coefficient when used to obtain bounds for point process characteristics related with the Palm distribution. Our goal is to establish a general inequality for the β -mixing coefficients of a point process in terms of its intensity functions.

We begin in Section 3.2 by recalling the basic definitions and properties of the α -mixing and β -mixing coefficients and we introduce the lower sum transform which is the main technical tool that we use throughout the chapter. Then, a general inequality for the β -mixing coefficients of a point process that depends only on its n -th order intensity functions is proved in Section 3.3. As an example, we deduce a β -mixing inequality in the special case of determinantal point processes (DPPs) in Section 3.4 whose rate of decay is optimal for a wide class of DPPs.

3.2 Preliminaries

3.2.1 Intensities of point processes

In this chapter, we consider simple point processes on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu)$ equipped with the euclidean norm $\|\cdot\|$ where d is a fixed integer, $\mathcal{B}(\mathbb{R}^d)$ the Borel- σ -algebra and μ the Lebesgue measure (more information on spatial point processes can be found in [31, 77]). We denote by Ω (resp. Ω_F) the set of locally finite (resp. finite) point configurations in \mathbb{R}^d . For all functions $f : \Omega_F \rightarrow \mathbb{R}$, $n \in \mathbb{N}$ and $x = (x_1, \dots, x_n) \in (\mathbb{R}^d)^n$, we write $f(x)$ for $f(\{x_1, \dots, x_n\})$ by an abuse of notation. Finally, we write $|A|$ for the cardinal of a finite set A and $\|f\|_\infty$ for the uniform norm of a function f .

We begin by recalling that the n -th order intensity functions (also called n -th order product density) are defined the following way (see [77]).

Definition 3.2.1. *Let X be a simple point process on \mathbb{R}^d and $n \geq 1$ be an integer. If there exists a non negative function $\rho_n : (\mathbb{R}^d)^n \rightarrow \mathbb{R}$ such that*

$$\mathbb{E} \left[\sum_{x_1, \dots, x_n \in X}^{\neq} f(x_1, \dots, x_n) \right] = \int_{(\mathbb{R}^d)^n} f(x) \rho_n(x) d\mu^n(x). \quad (3.2.1)$$

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

In the rest of the chapter, all point processes will be considered to admit bounded n -th order intensity function for all $n \geq 1$.

3.2.2 Mixing

Consider a probability space $(\mathcal{X}, \mathcal{F}, \mathbb{P})$ and \mathcal{A}, \mathcal{B} two sub σ -algebras of \mathcal{F} . Let $\mathbb{P}_{\mathcal{A}}$ and $\mathbb{P}_{\mathcal{B}}$ be the respective restrictions of \mathbb{P} to \mathcal{A} and \mathcal{B} and define the probability $\mathbb{P}_{\mathcal{A} \otimes \mathcal{B}}$ on the product σ -algebra by $\mathbb{P}_{\mathcal{A} \otimes \mathcal{B}}(A \times B) = \mathbb{P}(A \cap B)$ for all $A \in \mathcal{A}$ and $B \in \mathcal{B}$. The α -mixing and β -mixing coefficients (also called strong-mixing and absolute regularity coefficients) are defined as the following measures of dependence between \mathcal{A} and \mathcal{B} [80, 89]:

$$\alpha(\mathcal{A}, \mathcal{B}) := \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{A}, B \in \mathcal{B}\}, \quad (3.2.2)$$

$$\beta(\mathcal{A}, \mathcal{B}) := \|\mathbb{P}_{\mathcal{A} \otimes \mathcal{B}} - \mathbb{P}_{\mathcal{A}} \otimes \mathbb{P}_{\mathcal{B}}\|_{TV}, \quad (3.2.3)$$

where $\|\cdot\|_{TV}$ is the total variation of a signed norm.

For a given point process X and a bounded set $A \subset \mathbb{R}^d$, we denote by $\mu(A) := \int_A d\mu(x)$ the volume of A and $\mathcal{E}(A)$ the σ -algebra generated by $X \cap A$. Finally, for all $A, B \subset \mathbb{R}^d$, we write $\text{dist}(A, B)$ for the infimum of $\|y - x\|$ where $(x, y) \in A \times B$. The β -mixing coefficients of the point process X are then defined by

$$\beta_{p,q}(r) := \sup\{\beta(\mathcal{E}(A), \mathcal{E}(B)) : \mu(A) \leq p, \mu(B) \leq q, \text{dist}(A, B) > r\},$$

and we say that the point process X is beta-mixing if $\beta_{p,q}(r)$ vanishes when $r \rightarrow +\infty$ for all $p, q > 0$. The α -mixing coefficients can be defined in a similar way.

Our goal is to prove that under appropriate assumptions over the intensity functions ρ_n of X we have a β -mixing property.

3.2.3 Lower sum transform

The main tool we use throughout this chapter is the so-called lower sum operator (see [2]). Notice that when f is a symmetric function the term in the expectation in (3.2.1) can be written as $n! \sum_{Z \subset X} f(Z) \mathbb{1}_{|Z|=n}$. This motivates the following definition:

Definition 3.2.2. *Let f be a real function defined over Ω_F . The lower sum of f is the linear operator defined by*

$$\hat{f} : X \mapsto \sum_{Z \subset X} f(Z). \quad (3.2.4)$$

As shown in Example 4.19 in [2], this operator admits the following inverse transform.

Proposition 3.2.3 ([2, Theorem 4.18]). *The operator (3.2.4) admits an inverse transform \check{f} , called the lower difference of f , defined by*

$$\check{f} : X \mapsto \sum_{Z \subset X} (-1)^{|X \setminus Z|} f(Z). \quad (3.2.5)$$

These definitions extend to functions over Ω_F^2 by defining

$$\hat{f} : (X_1, X_2) \mapsto \sum_{Z_1 \subset X_1, Z_2 \subset X_2} f(Z_1, Z_2)$$

and

$$\check{f} : (X_1, X_2) \mapsto \sum_{Z_1 \subset X_1, Z_2 \subset X_2} (-1)^{|X_1 \setminus Z_1| + |X_2 \setminus Z_2|} f(Z_1, Z_2).$$

In a similar way, we could also extend these definitions to Ω_F^n for any n but we will only need the case $n \leq 2$ for the remaining of the chapter. These operators allow us to give an explicit expression for the expectation of a functional of a point process with respect to its intensity functions.

Proposition 3.2.4. *If X is an almost surely finite point process such that $\mathbb{E}[4^{|X|}] < +\infty$, then*

$$\mathbb{E}[f(X)] = \sum_{n=0}^{+\infty} \frac{1}{n!} \int_{(\mathbb{R}^d)^n} \check{f}(x) \rho_n(x) d\mu^n(x) \quad (3.2.6)$$

for all bounded functions $f : \Omega_F \rightarrow \mathbb{R}$. Moreover, if X' is a point process independent from X satisfying the same assumptions than X and with n -th order intensity functions ρ'_n , then

$$\mathbb{E}[f(X, X')] = \sum_{m,n=0}^{+\infty} \frac{1}{m!n!} \int_{(\mathbb{R}^d)^{m+n}} \check{f}(x, y) \rho_m(x) \rho'_n(y) d\mu^m(x) d\mu^n(y) \quad (3.2.7)$$

for all bounded functions $f : \Omega_F^2 \rightarrow \mathbb{R}$.

Proof. Using the bound $|\check{f}(x)| \leq \|f\|_\infty \text{card}\{Z, Z \subset X\} = \|f\|_\infty 2^{|x|}$ we get

$$\sum_{n \geq 0} \mathbb{E} \left[\left| \sum_{\substack{Z \subset X \\ |Z|=n}} \check{f}(Z) \right| \right] \leq \sum_{n \geq 0} \mathbb{E} \left[2^{|X|} \binom{|X|}{n} \right] \|f\|_\infty = \|f\|_\infty \mathbb{E} [4^{|X|}] < +\infty. \quad (3.2.8)$$

Since we can write

$$f(X) = \hat{f}(X) = \sum_{Z \subset X} \check{f}(Z) = \sum_{n \geq 0} \sum_{\substack{Z \subset X \\ |Z|=n}} \check{f}(Z) \quad \text{a.s.},$$

then

$$\mathbb{E}[f(X)] = \sum_{n \geq 0} \mathbb{E} \left[\sum_{\substack{Z \subset X \\ |Z|=n}} \check{f}(Z) \right] = \sum_{n=0}^{+\infty} \frac{1}{n!} \int_{(\mathbb{R}^d)^n} \check{f}(x) \rho_n(x) d\mu^n(x)$$

where the inversion of the first sum and the expectation is a consequence of (3.2.8). Similarly, for all functions $f : \Omega_F^2 \rightarrow \mathbb{R}$ we have

$$\begin{aligned} \mathbb{E}[f(X, X')] &= \mathbb{E}[\mathbb{E}[f(X, X') | X']] \\ &= \mathbb{E} \left[\sum_{m=0}^{+\infty} \frac{1}{m!} \int_{(\mathbb{R}^d)^m} \left(\sum_{z \subset x} (-1)^{m-|z|} f(z, X') \right) \rho_m(x) d\mu^m(x) \right] \\ &= \sum_{m,n=0}^{+\infty} \frac{1}{m!n!} \int_{(\mathbb{R}^d)^{m+n}} \check{f}(x, y) \rho_m(x) \rho'_n(y) d\mu^m(x) d\mu^n(y), \end{aligned}$$

where all inversions of expectation with sum and integrals can be justified in a similar way than (3.2.8). \square

3.3 β -mixing of point processes with known intensity functions

Our main result is the following inequality showing that if $\rho_m(x)\rho_n(y) - \rho_{m+n}(x, y)$ vanish fast enough when $\|y - x\| \rightarrow +\infty$ for all $m, n \in \mathbb{N}$, then the underlying point process is β -mixing.

Theorem 3.3.1. *Let X be a simple point process on (\mathbb{R}^d, μ) such that $\mathbb{E}[4^{|X \cap A|}] < +\infty$ for all bounded subsets $A \subset \mathbb{R}^d$. Then, for all $p, q, r \in \mathbb{R}_+$,*

$$\beta_{p,q}(r) \leq \sup_{\substack{\mu(A) < p, \mu(B) < q \\ \text{dist}(A,B) > r}} \left(\sum_{m,n=0}^{+\infty} \frac{2^{n+m-1}}{m!n!} \int_{A^m \times B^n} |\rho_m(x)\rho_n(y) - \rho_{m+n}(x, y)| d\mu^m(x) d\mu^n(y) \right). \quad (3.3.1)$$

Before giving the proof of Theorem 3.3.1, we need the following lemmas showing the behaviour of $f(X \cap A, X \cap B)$ and $f(X \cap A, X' \cap B)$ under the lower difference operator.

Lemma 3.3.2. *Let $A \subset \mathbb{R}^d$, $f : \Omega_F \rightarrow \mathbb{R}$ and define $f_A : X \mapsto f(X \cap A)$. Then,*

$$\widetilde{f}_A(X) = \check{f}(X) \mathbb{1}_{X \subset A}.$$

Proof. If $X \subset A$ then the result is trivial. Otherwise, there exists $x \in X \setminus A$ and we can write

$$\begin{aligned} \widetilde{f}_A(X) &= \sum_{Z \subset X, Z \ni x} (-1)^{|X \setminus Z|} f(Z \cap A) + \sum_{Z \subset X, Z \not\ni x} (-1)^{|X \setminus Z|} f(Z \cap A) \\ &= \sum_{Z \subset X, Z \not\ni x} (-1)^{|X \setminus Z| - 1} f((Z \cup \{x\}) \cap A) + \sum_{Z \subset X, Z \not\ni x} (-1)^{|X \setminus Z|} f(Z \cap A) \\ &= \sum_{Z \subset X, Z \not\ni x} (-1)^{|X \setminus Z| - 1} f(Z \cap A) + \sum_{Z \subset X, Z \not\ni x} (-1)^{|X \setminus Z|} f(Z \cap A) = 0. \end{aligned}$$

□

This result can be extended to multivariate functions: The lower difference of $(X_1, X_2) \rightarrow f(X_1 \cap A_1, X_2 \cap A_2)$ is $\check{f}(X_1, X_2) \mathbb{1}_{\{X_1 \subset A_1\}} \mathbb{1}_{\{X_2 \subset A_2\}}$.

Lemma 3.3.3. *For all $f : \Omega_F^2 \rightarrow \mathbb{R}$ and A, B disjoint subsets of \mathbb{R}^d , let us define the function $g : X \mapsto f(X \cap A, X \cap B)$. The lower difference of g satisfies*

$$\check{g}(X) = \check{f}(X \cap A, X \cap B) \mathbb{1}_{\{X \subset A \cup B\}}.$$

Proof. Using Lemma 3.3.2 we get that $\check{g}(X) = 0$ whenever X is not a subset of $A \cup B$. Otherwise, since A and B are disjoint sets,

$$\check{g}(X) = \sum_{Z \subset X} (-1)^{|X \setminus Z|} f(Z \cap A, Z \cap B) = \sum_{\substack{U \subset X \cap A \\ V \subset X \cap B}} (-1)^{|(X \cap A) \setminus U| + |(X \cap B) \setminus V|} f(U, V)$$

which, by definition, is equal to $\check{f}(X \cap A, X \cap B)$. □

We now have the necessary tools required for the proof of Theorem 3.3.1.

Proof of Theorem 3.3.1. Let $p, q > 0$ and A, B be two disjoint subsets of \mathbb{R}^d such that $\mu(A) \leq p$ and $\mu(B) \leq q$. Using one of the characterizations of the total variation distance, the β -mixing coefficient between $\mathcal{E}(A)$ and $\mathcal{E}(B)$ can be expressed as

$$\beta(\mathcal{E}(A), \mathcal{E}(B)) = \frac{1}{2} \sup_{\|f\|_\infty=1} |E[f(X \cap A, X \cap B)] - E[f(X \cap A, X' \cap B)]|$$

where X' is an independent copy of X . Since $X \cap A$, $X' \cap B$ and $X \cap B$ are finite a.s. we can apply (3.2.7) which, combined with Lemma 3.3.2, gives us

$$\mathbb{E}[f(X \cap A, X' \cap B)] = \sum_{m,n=0}^{+\infty} \frac{1}{m!n!} \int_{A^m \times B^n} \check{f}(x, y) \rho_m(x) \rho_n(y) d\mu^m(x) d\mu^n(y). \quad (3.3.2)$$

On the other hand, by combining (3.2.6) with Lemma 3.3.3, we get

$$\mathbb{E}[f(X \cap A, X \cap B)] = \sum_{n=0}^{+\infty} \frac{1}{n!} \int_{(A \cup B)^n} \check{f}(x \cap A, x \cap B) \rho_n(x) d\mu^n(x).$$

Since A and B are disjoint sets and by symmetry of $\check{f}(x \cap A, x \cap B) \rho_n(x)$, we can simplify the above expression into

$$\begin{aligned} \mathbb{E}[f(X \cap A, X \cap B)] &= \sum_{n=0}^{+\infty} \sum_{m=0}^n \frac{1}{n!} \binom{n}{m} \int_{A^m \times B^{n-m}} \check{f}(x, y) \rho_n(x, y) d\mu^m(x) d\mu^{n-m}(y) \\ &= \sum_{m,n=0}^{+\infty} \frac{1}{m!n!} \int_{A^m \times B^n} \check{f}(x, y) \rho_{m+n}(x, y) d\mu^m(x) d\mu^n(y). \quad (3.3.3) \end{aligned}$$

Combining (3.3.2) and (3.3.3) yields that $|\mathbb{E}[f(X \cap A, X \cap B)] - \mathbb{E}[f(X \cap A, X' \cap B)]|$ is equal to

$$\left| \sum_{m,n=0}^{+\infty} \frac{1}{m!n!} \int_{A^m \times B^n} \check{f}(x,y)(\rho_m(x)\rho_n(y) - \rho_{m+n}(x,y))d\mu^m(x)d\mu^n(y) \right|$$

which is bounded by

$$\sum_{m,n=0}^{+\infty} \frac{2^{n+m}}{m!n!} \int_{A^m \times B^n} |\rho_m(x)\rho_n(y) - \rho_{m+n}(x,y)|d\mu^m(x)d\mu^n(y)$$

when $\|f\|_\infty = 1$ and where we used the bound $|\check{f}(x,y)| \leq 2^{|x|+|y|}$. \square

3.4 Application to determinantal point processes

We can directly apply Theorem 3.3.1 to determinantal point processes. First introduced in [75] under its current form to model fermion systems, DPPs are a broad class of repulsive point processes. We recall that a DPP X with kernel $K : (\mathbb{R}^d)^2 \rightarrow \mathbb{R}$ is defined by its intensity functions

$$\rho_n(x_1, \dots, x_n) = \det(K[x]) \quad \forall x \in (\mathbb{R}^d)^n, \forall n \in \mathbb{N}$$

where we denote by $K[x]$ the matrix $(K(x_i, x_j))_{1 \leq i, j \leq n}$. Existence and uniqueness conditions as well as general information on DPPs can be found in [61]. The application of Theorem 3.3.1 to DPPs gives us the following β -mixing condition:

Theorem 3.4.1. *Let X be a DPP with kernel K and define*

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

If K is bounded and $\omega(r) \xrightarrow{r \rightarrow +\infty} 0$ then X is β -mixing. In particular,

$$\beta_{p,q}(r) \leq 2pq(1 + 2p\|K\|_\infty)(1 + 2q\|K\|_\infty)e^{2\|K\|_\infty(p+q)}\omega(r)^2.$$

Unfortunately, this result does not give a bound for $\beta_{p,\infty}(r)$ which yet is necessary in almost all limit theorems based on beta-mixing.

Proof. Since $\mathbb{E}[4^{|X \cap A|}] < +\infty$ for all bounded sets A (see Lemma 2.B.5) then the β -mixing coefficients of X satisfy (3.3.1) by Theorem 3.3.1. Let $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_m)$, we need to control $|\det(K[x])\det(K[y]) - \det(K[x,y])|$ where $\|x - y\| \geq r$. By Lemma 2.B.4, we get the bound

$$0 \leq \det(K[x])\det(K[y]) - \det(K[x,y]) \leq nm\|K\|_\infty^{n+m-2} \sum_{i=1}^n \sum_{j=1}^m K(x_i, y_j)^2.$$

Injecting this bound into (3.3.1) gives us

$$\begin{aligned} \beta_{p,q}(r) &\leq \sum_{n,m=0}^{+\infty} \frac{n^2 m^2 2^{n+m-1} p^{n-1} q^{m-1} \|K\|_\infty^{n+m-2}}{n!m!} \sup_{\substack{|A| < p, |B| < q \\ \text{dist}(A,B) > r}} \int_{A \times B} |K(x,y)|^2 d\mu(x)d\mu(y) \\ &\leq \sum_{n,m=0}^{+\infty} \frac{n^2 m^2 2^{n+m-1} p^n q^m \|K\|_\infty^{n+m-2}}{n!m!} \omega(r)^2 \\ &= 2pq(1 + 2p\|K\|_\infty)(1 + 2q\|K\|_\infty)e^{2(p+q)\|K\|_\infty} \omega(r)^2. \end{aligned} \tag{3.4.1}$$

In particular, if $\omega(r)$ vanishes when $r \rightarrow +\infty$ then X is β -mixing. \square

In conclusion, the β -mixing coefficients of DPPs decay at the same rate as $|K(x, y)|^2$ does when x and y deviate from each other. For example, kernels of the Ginibre ensemble or the Gaussian unitary ensemble have an exponential decay (see [61]). Moreover, among translation-invariant kernels used in spatial statistics (see [15, 69]), all kernels of the Laguerre-Gaussian family also have an exponential decay while kernels of the Whittle-Matérn and Cauchy family satisfy $\omega(r) = o(r^{-d})$ and kernels of the Bessel family satisfy $\omega(r) = o(r^{-(d+1)/2})$.

It is also worth noticing that Theorem 3.3.1 is optimal in the sense that for a wide class of DPPs, the β -mixing coefficients $\beta_{p,q}(r)$ do not decay faster, when r goes to infinity, than the supremum of $\int_{A \times B} |K(x, y)|^2 d\mu(x) d\mu(y)$ for all A, B such that $\mu(A) \leq p$, $\mu(B) \leq q$ and $\text{dist}(A, B) \geq r$ as stated in the following proposition.

Proposition 3.4.2. *Let X be a DPP with a non-negative bounded kernel K such that the eigenvalues of its associated integral operator are all in $[0, M]$ where $M < 1$. Then, for all $p, q, r > 0$,*

$$\begin{aligned} & 2(1 - M)^{\frac{(p+q)\|K\|_\infty}{M}} \sup_{\substack{\mu(A) < p, \mu(B) < q \\ \text{dist}(A, B) > r}} \int_{A \times B} |K(x, y)|^2 d\mu(x) d\mu(y) \leq \beta_{p,q}(r) \\ & \leq 2(1 + 2p\|K\|_\infty)(1 + 2q\|K\|_\infty) e^{2(p+q)\|K\|_\infty} \sup_{\substack{\mu(A) < p, \mu(B) < q \\ \text{dist}(A, B) > r}} \int_{A \times B} |K(x, y)|^2 d\mu(x) d\mu(y). \end{aligned}$$

Proof. The first inequality is a consequence of the fact that $\beta_{p,q}(r) \geq 2\alpha_{p,q}(r)$ and Proposition 2.4.3. The second inequality is equivalent to (3.4.1) once the sum has been developed. \square

Chapter 4

Adaptive estimating function inference for non-stationary determinantal point processes

This chapter is a work that has been submitted to a mathematics journal so some considerations and definitions are redundant with the introduction. For the same reason, some notations are specific to this chapter.

4.1 Introduction

A common feature of spatial point process models (except for the Poisson process case) is that the likelihood function is not available in a simple form. Numerical approximations of the likelihood function are available (see e.g. [77, 78, for reviews]) but the approaches are often computationally demanding and the distributional properties of the approximate maximum likelihood estimates may be difficult to assess. Therefore much work has focused on establishing computationally simple estimation methods that do not require knowledge of the likelihood function.

In this chapter we focus on estimation methods for point processes which have known joint intensity functions. This includes many cases of Cox and cluster point process models [77, 63, 7] as well as determinantal point processes [75, 97, 94, 69]. These classes of models are quite different since realizations of Cox and cluster point processes are aggregated while determinantal point processes produce regular point pattern realizations.

Knowledge of an n th order joint intensity enables the use of the so-called Campbell formulae for computing expectations of statistics given by random sums indexed by n -tuples of distinct points in a point process. Unbiased estimating functions can then be constructed from such statistics by subtracting their expectations. So far mainly the cases of first and second order joint intensities have been considered where the first order joint intensity is simply the intensity function.

Theoretical results have been established in a variety of special cases of first and second order estimating functions for Cox and cluster processes [93, 48, 102, 49, 103] and for the closely related Palm likelihood estimators [99, 86, 85]. The common general structure of the estimating functions on the other hand calls for a general theoretical set-up which is the first contribution of this chapter. Our set-up also covers third or higher order estimating functions and combinations of such estimating functions, providing a general unifying framework.

The literature on statistical inference for continuous determinantal point processes is quite limited. A Bayesian approach is considered in [1], while likelihood and minimum contrast estimation methods are discussed in [69]. Consistency and asymptotic normality of the associated estimator are so far only available in case of minimum contrast estimation for stationary determinantal point processes [16]. Based on the general set-up our second main contribution is to provide a consistency and asymptotic normality result of estimating function estimators for general non-stationary determinantal point processes.

Specializing to second-order estimating functions, a common approach [48, 99] is to restrict the random sum to pairs of R -close points for some user-specified $R > 0$. This may lead to faster computation and improved statistical efficiency. The properties of the resulting estimators depend strongly on R but only ad hoc guidance is available for the choice of R . Moreover, it is difficult to account for ad hoc choices of R when establishing theoretical results. Our third contribution is a simple intuitively appealing adaptive choice of R which leads to a theoretically tractable estimation procedure and we demonstrate its usefulness in simulation studies for determinantal point processes as well as an example of a cluster process.

4.2 Estimating functions based on joint intensities

A point process X on \mathbb{R}^d , $d \geq 1$, is a locally finite random subset of \mathbb{R}^d . For $B \subseteq \mathbb{R}^d$, we let $N(B)$ denote the random number of points in $X \cap B$ and $|B|$ the Lebesgue measure of B . That X is locally finite means that $N(B)$ is finite almost surely whenever B is bounded. The so-called joint intensities of a point process are described in Section 4.2.1. In this chapter we mainly focus on determinantal point processes, detailed in Section 4.3. A prominent feature of determinantal point processes is that they have known joint intensity functions of any order.

4.2.1 Joint intensity functions and Campbell formulae

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

$$\mathbb{E} \sum_{u_1, \dots, u_n \in X}^{\neq} \mathbb{1}_{u_1 \in B_1, \dots, u_n \in B_n} = \int_{\times_{i=1}^n B_i} \rho^{(n)}(u_1, \dots, u_n) du_1 \cdots du_n \quad (4.2.1)$$

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

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

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

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

follows immediately from the definition of $\rho^{(n)}$ for any non-negative function f .

4.2.2 A general asymptotic result for estimating functions

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

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

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

A basic assumption for the following theorem is that a central limit theorem is available for $e_n(\theta^*)$ (assumption (X3)). In addition to this, a number of technical assumptions (F1) through (F3) (or (F3')), (X1) and (X2) regarding existence and differentiability of joint intensities as well as differentiability of the f_i are needed. All these conditions as well as the proof of the following theorem are given in Section 4.A of the appendix.

Theorem 4.2.1. *Under Assumptions (F1) through (F3) (or (F3')), (X1) and (X2), with a probability tending to one as $n \rightarrow \infty$, there exists a $|W_n|$ -consistent sequence of roots $\hat{\theta}_n$ of the estimating equations $e_n(\theta) = 0$. Precisely, for all $\varepsilon > 0$, there exists $A > 0$ such that*

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

for a sufficiently large n .

Moreover, if (X3) holds true, then

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

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

Remark 4.2.2. *While the parameter θ^* is generally uniquely defined (in the sense that $\theta \mapsto \mathbb{P}_\theta$ is injective) and verifies $\mathbb{E}(e_n(\theta^*)) = 0$, the solution to $e_n(\theta) = 0$ may not be unique. The above theorem states that there exists a consistent and asymptotically Gaussian sequence of solutions, but uniqueness is not guaranteed. This drawback is unfortunately common in most asymptotic results for estimating functions inference, see the references in introduction, [96], or the handbook by C. Heyde [59]. Nonetheless it can be proved that the solution is unique for n sufficiently large whenever $\lim_{n \rightarrow \infty} e_n(\theta)/|W_n|$ admits a unique zero, see [64]. But to ensure the latter condition, we would need some very strong additional properties on the parametric form of the test functions and of the joint intensity functions.*

4.2.3 Second order estimating functions

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

$$e_n(\theta) = \sum_{u, v \in X \cap W_n}^{\neq} f(u, v; \theta) - \int_{W_n^2} f(u, v; \theta) \rho^{(2)}(u, v; \theta) dudv. \quad (4.2.2)$$

In [48], Y. Guan noted that for computational and statistical efficiency it may be advantageous to use only close pairs of points rather than all pairs of points. Thus in (4.2.2) it is common practice to introduce an indicator $\mathbb{1}_{\|u-v\|\leq R}$ for some constant $0 < R$ or choose f so that $f(u, v) = 0$ whenever $\|u - v\| > R$. We discuss a method for choosing R in Section 4.2.4.

The general form (4.2.2) includes e.g. the score functions of second-order composite likelihood [48, 102] and Palm likelihood functions [99, 86, 85] as well as score functions of minimum contrast object functions based on non-parametric estimates of summary statistics as the K or the pair correlation function. For the second-order composite likelihood in [48],

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

while

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

for the second-order composite likelihood proposed in [102]. The score of the Palm likelihood as generalized to the inhomogeneous case in [85] is obtained with

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

In [85], the authors also regarded the second-order composite likelihood proposed in [102] as a generalization of the stationary case Palm likelihood but the interpretation as a second-order composite likelihood given in [102] is more straightforward.

Considering a class of estimating functions of the form (4.2.2) a natural question is what is the optimal choice of f ? A solution to this problem is provided in [35] where an approximation of the optimal f is obtained by solving numerically a certain integral equation. This yields a statistically optimal estimation procedure but is computationally demanding and requires specification of third and fourth order joint intensities. When computational speed and ease of use is an issue, there is still scope for simpler methods. Moreover, given several (simple) estimation methods, it is possible to combine them adaptively in order to build a final estimator that achieves better properties than each initial estimator, see [70, 71].

4.2.4 Adaptive version

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

$$f_R(u, v; \theta) = \mathbb{1}_{\|u-v\|\leq R} \frac{\nabla_{\theta} \rho^{(2)}(u, v; \theta)}{\rho^{(2)}(u, v; \theta)}. \quad (4.2.3)$$

As mentioned in the previous section, using only R close pairs may be beneficial both for statistical efficiency and computational tractability. However, the performances depends strongly on the chosen R . Simulation studies such as in [85] and [35] usually compare results for several values of R corresponding to different multiples of some parameter associated with ‘range of correlation’. For a cluster process this parameter could e.g. be the standard deviation of the distribution for dispersal of offspring around parents. For a determinantal point process the parameter would typically be a correlation scale parameter in the kernel of the determinantal point process, see Section 4.3. In

practice these parameters are not known and among the quantities that need to be estimated. In [48] it is suggested to choose an R that minimizes a goodness of fit criterion for the fitted point process model while the choice of R in [103] is done by inspection of a non-parametric estimate of the pair correlation function (a similar approach is suggested by [51] and [8] in the context of pairwise composite likelihood for random fields). Both approaches imply extra work and ad hoc decisions by the user and it becomes very complex to determine the statistical properties of the resulting parameter estimates. A typical behaviour of many pair correlation functions is that $g(u, v; \theta)$ converges to the limiting value of 1 when $\|u - v\|$ increases and $|g(u, v; \theta) - 1| \leq M(u, v; \theta)$ where

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

Note that for DPPs, $M(u, v; \theta) = 1$ (see the next section) and for stationary point processes, $M(u, v; \theta)$ does not depend on u and v . If $g(u, v; \theta) = 1$ for $\|u - v\| > r_0$ then counts of points are uncorrelated when they are observed in regions separated by a distance of r_0 .

Following the idea that R should depend on some range property of the point process we therefore suggest to replace the constraint $\|u - v\| < R$ in (4.2.3) by the constraint

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

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

We then modify the function f_R to be

$$f_{\text{adap}}(u, v; \theta) = w \left(\varepsilon \frac{M(u, v; \theta)}{g(u, v; \theta) - 1} \right) \frac{\nabla_{\theta} \rho^{(2)}(u, v; \theta)}{\rho^{(2)}(u, v; \theta)} \quad (4.2.4)$$

where w is some weight function of bounded support $[-1, 1]$. Later on, when establishing asymptotic results, we will also assume that w is differentiable. A common example of admissible weight function is $w(r) = e^{1/(r^2-1)}$ for $-1 \leq r \leq 1$, while $w(r) = 0$ otherwise. The user needs to specify a value of ε but in contrast to the original tuning parameter R , ε has an intuitive meaning independent of the underlying point process. We choose $\varepsilon = 1\%$. In the simulation study in Section 4.4.1 we also consider $\varepsilon = 5\%$ in order to investigate the sensitivity to the choice of ε .

We emphasize that choosing $\varepsilon = 1\%$ or 5% is not necessarily optimal. An optimal ε might be found by maximizing the Godambe information as a function of ε but this is not straightforward and the computational advantages of our approach would be lost. In fact, if Godambe optimality is key, we suggest to consider the previously mentioned approach by [35] to identify an optimal second order estimating function.

4.3 Asymptotic results for determinantal point processes

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

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

for all $\{u_1, \dots, u_n\} \subset \mathbb{R}^d$, where $[K](u_1, \dots, u_n)$ is the matrix with entries $K(u_i, u_j)$. The intensity function is thus $\rho(u) = K(u, u)$, $u \in \mathbb{R}^d$. If a determinantal point process

with kernel K exists it is unique. General conditions for existence are presented in [69]. In particular, if K admits the form

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

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

We now consider a parametric family of DPPs on \mathbb{R}^d with kernels K_θ where $\theta \in \Theta$ and $\Theta \subseteq \mathbb{R}^p$ (see [69, 15, for examples of such families]). Henceforth, we assume that K_θ is symmetric, continuous and the DPP with kernel K_θ exists for all $\theta \in \Theta$. Note that in general, it is possible that two different kernels generate the same DPP distribution. This identifiability issue especially arises in the case of a discrete state space, where the distribution of a DPP is only identified up to flips of the signs of the rows and columns of its matrix kernel (see [40] or [90]). However, in the continuous case, corresponding to our framework, the kernel of a DPP is uniquely determined whenever the intensity function is positive, see Proposition 4.C.1 and its corollary in the appendix. Assuming a positive intensity function is not restrictive for statistical applications of DPPs.

An expression for the likelihood of a DPP on a bounded window is provided in [69], where likelihood based inference for stationary DPPs is discussed. However, the expression depends on a spectral representation of K which is rarely known in practice and must be approximated numerically. Letting n denote the number of observed points, the likelihood further requires the computation of an $n \times n$ dense matrix which can be time consuming for large n . As an alternative, minimum contrast estimation is considered in [15], based on the pair correlation function or Ripley's K -function, but only for stationary DPPs. In the following, we consider general non-stationary DPPs and the estimator $\hat{\theta}_n$ obtained by solving $e_n(\theta) = 0$ where e_n is given by (4.2.2). Note that the distribution for any classical parametric DPP model showcased in [69, 15] is uniquely determined by its first two order intensity functions, in the sense that $\theta \mapsto (\rho(\cdot; \theta), \rho^{(2)}(\cdot, \cdot; \theta))$ is injective. This justifies the use of second order estimating functions for DPPs.

We establish in Section 4.3.1 using Theorem 4.2.1 the asymptotic properties of the estimate $\hat{\theta}_n$ where e_n is given by (4.2.2) for a wide class of test functions f . In Section 4.3.2, we focus on a particular case of the DPP model, where the parameter $\theta = (\beta, \psi)$ can be separated into a parameter β only appearing in the intensity function and a parameter ψ only appearing in the pair correlation function. Following [103], it is natural to consider a two-step estimation procedure where in a first step β is estimated by a Poisson likelihood score estimating function, and in a second step the remaining parameter ψ is estimated by a second order estimating function as in (4.2.2), where β is replaced by $\hat{\beta}_n$ obtained in the first step. The asymptotic properties of this two-step procedure again follow as a special case of Theorem 4.2.1.

4.3.1 Second order estimating functions for DPPs

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

We assume a realization of a DPP X with kernel K_{θ^*} , $\theta^* \in \text{Int}(\Theta)$, is observed on a bounded window $W_n \subset \mathbb{R}^d$. We estimate the unknown parameter θ^* by the solution $\hat{\theta}_n$

of $e_n(\theta) = 0$ where $e_n(\theta)$ is given by (4.2.2) for a given \mathbb{R}^p -valued function f . Therefore, we are in a special case of the set-up in Section 4.2.2 with $l = 1$, $q_1 = 2$, $k_1 = p$ and we assume that $f_1 = f$ satisfies the assumptions (F1) through (F3) (or (F3')) listed in Appendix 4.A. The condition (F1) in this case demands that $\theta \mapsto f(u, v; \theta)$ is twice continuously differentiable in a neighbourhood of θ^* and for θ in this neighbourhood, the derivatives are bounded with respect to (u, v) uniformly in θ . Moreover, from (F2), there exists $R > 0$ such that for all θ in a neighbourhood of θ^* ,

$$f(u, v; \theta) = 0 \quad \text{if} \quad \|u - v\| > R. \quad (4.3.1)$$

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

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

where we recall that in this setting

$$\rho^{(2)}(u, v; \theta) = K_{\theta}(u, u)K_{\theta}(v, v) - K_{\theta}(u, v)^2. \quad (4.3.2)$$

The assumptions (F3) and (F3') are technical and needed for the consistency of the estimation procedure. When H_n is a symmetric matrix, assumption (F3) seems simpler to verify than (F3'). As an important example, when f is defined as in (4.2.4), we prove in Lemmas 4.3.2 and 4.3.3 that (F3) is generally satisfied even if X is not stationary.

Finally, as shown in the proof of Theorem 4.3.1 below, the assumptions (X1) through (X3) in Theorem 4.2.1 are implied by the following:

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

(D2) The kernel K_{θ^*} satisfies, for some $\varepsilon > 0$,

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

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

(W) $\exists \varepsilon > 0$ s.t. $|\partial W_n \oplus (R + \varepsilon)| = o(|W_n|)$, where ∂ in this context denotes the boundary of a set, R is defined in (4.3.1), and $|W_n| \rightarrow \infty$, as $n \rightarrow \infty$.

Let us briefly comment on these assumptions. (D1) is a standard regularity assumption. Condition (D2) is not restrictive since all standard parametric kernel families satisfy $\sup_{\|u-v\|>r} K_{\theta}(u, v) = O(r^{-(d+1)/2})$, including the most repulsive stationary DPP (see [69, 15]). Condition (D3) ensures that the asymptotic variance in the central limit theorem below is not degenerated. Finally, Assumption (W) makes specific the fact that W_n is not too irregularly shaped and is not bounded in any direction. It is for instance fulfilled if W_n is a Cartesian product of d intervals whose lengths tends to infinity.

Theorem 4.3.1. *Under Assumptions (D1) and (D2), if assumptions (F1) through (F3) (or (F3')) are satisfied for $f_1 = f$, with a probability tending to one as $n \rightarrow \infty$, there exists a $|W_n|$ -consistent sequence of roots $\hat{\theta}_n$ of the estimating equations $e_n(\theta) = 0$. If moreover (W) and (D3) holds true, then*

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

Proof. We deduce from (4.3.2) that (D1) implies (X1). Moreover, it was shown in the proof of Lemma 2.B.6 that (X2) is a consequence of (D2) and that (X3) is a consequence of (D2), (D3) and (W). Thus, we can conclude by applying Theorem 4.2.1 in the case $l = 1$ and $q_1 = 2$. \square

In the case of a stationary X and f given by (4.2.4), the following lemma shows that (F3) is satisfied under mild assumptions that are violated only in degenerate cases. For instance, if $p = 1$, the last assumption boils down to $\nabla_{\theta} \rho^{(2)}(0, t; \theta^*) \neq 0$ for some $t \neq 0$ such that $|K_{\theta^*}(t)| > \sqrt{\varepsilon} K_{\theta^*}(0)$. In particular it is not difficult to verify these assumptions for the stationary parametric kernels considered in our simulation study of Section 4.4, namely the Bessel-type and the Gaussian kernels, see Section 4.D of the appendix.

Lemma 4.3.2. *Assume (W) and (D2), suppose X is stationary and let f be as in (4.2.4). Assume that w is positive on $[0, 1]$, $t \mapsto f(0, t; \theta^*)$ is integrable and*

$$\text{span}\{\nabla_{\theta} \rho^{(2)}(0, t; \theta^*) : |K_{\theta^*}(t)| > \sqrt{\varepsilon} K_{\theta^*}(0)\} = \mathbb{R}^p,$$

then (F3) is satisfied.

Proof. By definition of w and (D2), there exists $R > 0$ such that $f(0, t; \theta^*) = 0$ when $\|t\| \geq R$. By Lemma 4.A.1, since $t \mapsto f(0, t; \theta^*)$ is integrable then $H_n(\theta^*)$ converges towards the positive semi-definite matrix $H(\theta^*) = \int_{\|t\| < R} h(t) dt$ where the function $h : \mathbb{R}^d \rightarrow \mathbb{R}^{p \times p}$ is defined by

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

In this case, proving (F3) is equivalent to showing that $\phi^T H(\theta^*) \phi = 0$ only if $\phi = 0$. For this, let A be the set of t such that $|K_{\theta^*}(t)| > \sqrt{\varepsilon} K_{\theta^*}(0)$, $\phi \in \mathbb{R}^p$ and note that since $w(\varepsilon K_{\theta^*}(0)^2 / K_{\theta^*}(t)^2) > 0$ for $t \in A$ and $h(t)$ is continuous and positive semi-definite,

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

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

Similarly, we can show that even in the non-stationary case, condition (F3) is satisfied for the function in (4.2.4) but under some slightly stronger assumptions on $\nabla_{\theta} \rho^{(2)}(u, v; \theta^*)$. Namely, we demand that all functions $v \mapsto \nabla_{\theta} \rho^{(2)}(u, v; \theta^*)$ are not contained in a single hyperplane of \mathbb{R}^p nor confined around 0. This is similar in essence to what we have assumed in the previous corollary but with the need of a uniform condition with respect to u . Functions that do not satisfy these requirements are arguably degenerate. In particular, a straightforward calculus carried out in the appendix shows that the non-stationary Bessel-type kernel used in our simulation study satisfies these assumptions.

Lemma 4.3.3. *Assume (W), (D2) and that K_{θ^*} is bounded. Let f be as in (4.2.4) and define $h : (\mathbb{R}^d)^2 \rightarrow \mathbb{R}^{p \times p}$ by*

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

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

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

then (F3) is satisfied.

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

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

where

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

By (W), we have

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

and for all ϕ ,

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

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

$$\forall v \in A, |\phi^T \nabla_{\theta} \rho^{(2)}(u, v; \theta^*)| > \delta \quad \text{and} \quad w \left(\frac{\varepsilon K_{\theta^*}(u, u) K_{\theta^*}(v, v)}{K_{\theta^*}(u, v)^2} \right) > \inf_{x \in [0, 1/\mu]} w(x)$$

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

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

where the limit is a consequence of (W). Since the limit does not depend on ϕ , then (F3) is satisfied. \square

4.3.2 Two-step estimation for a separable parameter

We consider a family of kernels

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

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

As in the preceding section, we assume a DPP X with kernel K_{θ^*} , $\theta^* \in \text{Int}(\Theta)$, is observed on a bounded window $W_n \subset \mathbb{R}^d$. In the spirit of [103], we estimate θ^* in two steps. First, β^* is estimated as the solution $\hat{\beta}_n$ of $s_n(\beta) = 0$ where

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

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

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

for a given \mathbb{R}^q -valued function f and where

$$\rho^{(2)}(u, v; \theta) = \rho^{(2)}(u, v; \beta, \psi) = \rho(u; \beta)\rho(v; \beta)(1 - C^2(u, v; \psi))$$

in this case. Here and in the following, for convenience of notation, we identify $u_n(\beta, \psi)$ with $u_n(\theta)$ when $\theta = (\beta^T, \psi^T)^T$.

This two-step procedure is a particular estimating equation procedure, since $\hat{\theta}_n := (\hat{\beta}_n^T, \hat{\psi}_n^T)^T$ is obtained as the solution of $e_n(\theta) = 0$ where $e_n(\theta) = (s_n(\beta)^T, u_n(\beta, \psi)^T)^T$. Thus, this is a particular case of the setting in Section 4.2.2 where $l = 2$, $q_1 = 1$, $q_2 = 2$, $f_1 = \nabla_\beta \rho(u; \beta)/\rho(u; \beta)$ and $f_2 = f$.

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

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

where

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

Since it is a non symmetric matrix, condition (F3') is more applicable than (F3). Mild conditions ensuring (F3') in the stationary case are provided in Lemma 4.3.5.

Theorem 4.3.4. *Under Assumptions (D1) and (D2), if assumptions (F1) through (F3) (or (F3')) are satisfied for $f_1 = \nabla_\beta \rho(u; \beta) / \rho(u; \beta)$ and $f_2 = f$, then with a probability tending to one as $n \rightarrow \infty$, there exists a $|W_n|$ -consistent sequence of roots $\hat{\theta}_n$ of the estimating equations $e_n(\theta) = 0$. If moreover (W) and (D3) hold true, then*

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

Proof. The proof follows the same lines as the proof of Theorem 4.3.1. \square

The next lemma is similar to Lemma 4.3.2. When $q = 1$ the last technical condition boils down to $\nabla_\psi(1 - C^2(0, t; \psi^*)) \neq 0$ for some t such that $C(0, t; \psi^*) \geq \sqrt{\varepsilon} C(0, 0; \psi^*)$. In particular, the stationary kernels in Section 4.4 satisfy the required assumptions, see Section 4.D.

Lemma 4.3.5. *Assume that for all θ , $K_\theta(u, v)$ only depends on $u - v$, in which case $\rho(u; \beta) = \beta$ with $\beta > 0$ and $C(u, v; \psi) = C(0, v - u; \psi)$ with $\psi \in \mathbb{R}^q$. Then the output of the first step is $\hat{\beta}_n = N(X \cap W_n) / |W_n|$. In the second step, assume*

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

Assume that w is positive on $[0, 1]$, $t \mapsto f(0, t; \theta^)$ is integrable and that $\text{span}\{\nabla_\psi(1 - C^2(0, t; \psi^*)) : C(0, t; \psi^*) > \sqrt{\varepsilon}\} = \mathbb{R}^q$, then (F3') is satisfied under (W), (D1) and (D2).*

Proof. By definition of w and (D2), there exists $R > 0$ such that $f(0, t; \theta^*) = 0$ when $\|t\| \geq R$. Since K_θ and f are invariant by translation and $t \mapsto f(0, t; \theta^*)$ is integrable then $H_n(\theta)$ converges by Lemma 4.A.1. In particular, we have

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

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

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

The limit of $H_n(\theta)$ is continuous by (D1). In this case, proving (F3') is equivalent to showing that the limit of $H_n(\theta^*)$ is invertible. Since this matrix is block triangular and $\beta > 0$ then it is invertible if and only if the limit of $H_n^{2,1}(\theta^*)$ is invertible. This is done the same way as in Lemma 4.3.2. \square

4.4 Simulation study

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

4.4.1 Performance of adaptive estimating function

In order to assess the adaptive test function (4.2.4) against the truncated test function (4.2.3) with a prescribed R , we consider a DPP model in \mathbb{R}^2 with a Bessel-type kernel

$$K(u, v) = \sqrt{\rho(u)\rho(v)} \frac{J_1(2\|u - v\|/\alpha)}{\|u - v\|/\alpha}, \quad (4.4.1)$$

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

$$\alpha^2 \|\rho\|_\infty \leq \frac{1}{\pi}. \quad (4.4.2)$$

This relation shows the tradeoff between the expected number of points and the strength of repulsiveness that we can obtain. This model is a particular instance of the Bessel-type DPP introduced in [15]. It covers a large range of repulsiveness, from the Poisson point process (when α is close to 0) to the most repulsive DPP (when $\alpha = 1/\sqrt{\pi\|\rho\|_\infty}$).

For this model, we consider three constant values of ρ , $\rho \in \{50, 100, 1000\}$, corresponding to homogeneous DPPs, and an inhomogeneous situation where $\rho(u) = \rho(x, y) = 20 \exp(4x)$ when $u \in [0, 1]^2$. The latter case corresponds to a log-linear intensity function involving two parameters. For each ρ , three values of α are considered: a small one, a medium one, and a last one close to the maximal possible value satisfying (4.4.2). Examples of point patterns simulated on $[0, 1]^2$ are displayed in Figure 4. All simulations are carried out using R [87], in particular the library `spatstat` [7].

We estimate ρ and α by a two-step procedure as studied in Section 4.3.2 from realizations of the DPP on $W = [0, 1]^2$. The alternative global approach of Section 4.3.1 is discussed in the next section. In the first step, the parameters arising in ρ are estimated by the score function for a Poisson point process. This gives $\hat{\rho} = N(X \cap W)/|W|$ in the homogeneous cases. In the second step, we consider the estimating equation based on (4.2.3) where θ is α in this setting and when $R \in \{0.05, 0.1, 0.25\}$, and based on the adaptive test function (4.2.4) with $\varepsilon = 0.01$ and the weight function w given at the end of Section 4.2.4. This yields four different estimators of α . The root mean square errors (RMSEs) of these estimators and the mean computation time estimated from 1000 replications are summarised in Table 1. Boxplots are displayed in Figure 5 in the appendix. Note that the codes have not been optimised, but the same computational strategy has been used for all methods, making the comparison of the mean computation time meaningful.

The Bessel-type kernel and the aforementioned test functions used in the two-step estimation procedure fulfill the assumptions of Theorem 4.3.4 and Lemma 4.3.5 (for the homogeneous case), ensuring nice asymptotic properties of the estimators considered in this section. This is confirmed by the estimated RMSE's reported in Table 1, that decrease when the intensity ρ increases which mimics the effect of an increasing window since rescaling the window by a factor $1/k$ is equivalent to change ρ into $k^2\rho$ and α into α/k , see (2.4) in [69]. Moreover, these RMSE's show that the best choice of R in the test function (4.2.3) clearly depends on the range of interaction of the underlying process. This emphasizes the importance of a data-driven approach to choosing R since the range is unknown in practice. Fortunately, the performance of the adaptive method is always better than the worst choice of R and very close to the best R except for the case $\rho = 100, \alpha = 0.01$. For the exceptional case, the small differences in performance can be explained by Monte Carlo error. Further, use of the adaptive method implies only little or no extra computational effort. In presence of many points, the adaptive

version is in fact much faster to compute than the estimator based on (4.2.3) with the choice of a too large R , see for instance the results for $\rho = 1000$ and $R = 0.25$.

Table 3 in the appendix shows the root mean square errors of the adaptive estimator using $\varepsilon = 0.05$. The RMSEs obtained with $\varepsilon = 0.05$ are bigger than those obtained with $\varepsilon = 0.01$. Nevertheless, the adaptive method with $\varepsilon = 0.05$ still performs well in the sense that it usually performs better than the worst R and usually almost as good as the best R . Because the above estimation methods sometimes fail to converge, we also report in Table 2 in the appendix the percentages of times each method has converged in our simulation study. These percentages are similar for all methods. Note that the results in Table 1 and in Figure 5 are based on 1000 simulations where all four methods have converged.

4.4.2 Two-step versus simultaneous

Most models used in spatial statistics involve a separable parameter $\theta = (\beta, \psi)$ where β only appears in the intensity function and ψ only appears in the pair correlation function. This makes the two-step procedure described in Section 4.3.2 available, as exploited in the previous simulation study. However a simultaneous second order estimating equation approach might be a better alternative. It is not easy to compare the respective performance of the two approaches through the asymptotic variances obtained in Section 4.3.1 and Section 4.3.2. In this section, we show through an example why the two-step procedure seems preferable.

We consider a stationary model with parameter $\theta = (\rho, \psi)$, where ρ is the intensity and the pair correlation function writes $g(u, v; \theta) = g(r; \psi)$ with $r = \|u - v\|$. In this case the two-step procedure, based on the observation $X \cap W$ and using the adaptive test function (4.2.4), provides $\hat{\rho} = N(X \cap W)/|W|$ and $\hat{\psi}$ is the root of

$$e_2(\psi) = \sum_{r_{ij}} w \left(\varepsilon \frac{g(0; \psi) - 1}{g(r_{ij}; \psi) - 1} \right) \frac{\nabla_{\psi} g(r_{ij}; \psi)}{g(r_{ij}; \psi)} - N(X \cap W)^2 \int_W w \left(\varepsilon \frac{g(0; \psi) - 1}{g(r; \psi) - 1} \right) \nabla_{\psi} g(r; \psi) dF(r). \quad (4.4.3)$$

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

$$e(\psi) = \sum_{r_{ij}} w \left(\varepsilon \frac{g(0; \psi) - 1}{g(r_{ij}; \psi) - 1} \right) \frac{\nabla_{\psi} g(r_{ij}; \psi)}{g(r_{ij}; \psi)} - \frac{\sum_{r_{ij}} w \left(\varepsilon \frac{g(0; \psi) - 1}{g(r_{ij}; \psi) - 1} \right)}{\int w \left(\varepsilon \frac{g(0; \psi) - 1}{g(r; \psi) - 1} \right) g(r; \psi) dF(r)} \int w \left(\varepsilon \frac{g(0; \psi) - 1}{g(r; \psi) - 1} \right) \nabla_{\psi} g(r; \psi) dF(r), \quad (4.4.4)$$

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

$$\hat{\rho}^2 = \frac{1}{|W|^2} \frac{\sum_{r_{ij}} w \left(\varepsilon \frac{g(0; \hat{\psi}) - 1}{g(r_{ij}; \hat{\psi}) - 1} \right)}{\int w \left(\varepsilon \frac{g(0; \hat{\psi}) - 1}{g(r; \hat{\psi}) - 1} \right) g(r; \hat{\psi}) dF(r)}. \quad (4.4.5)$$

The more complicated expression of (4.4.4) in comparison with (4.4.3) implies that $e(\psi)$ can be highly irregular in ψ . Figure 6 in the appendix shows an example for

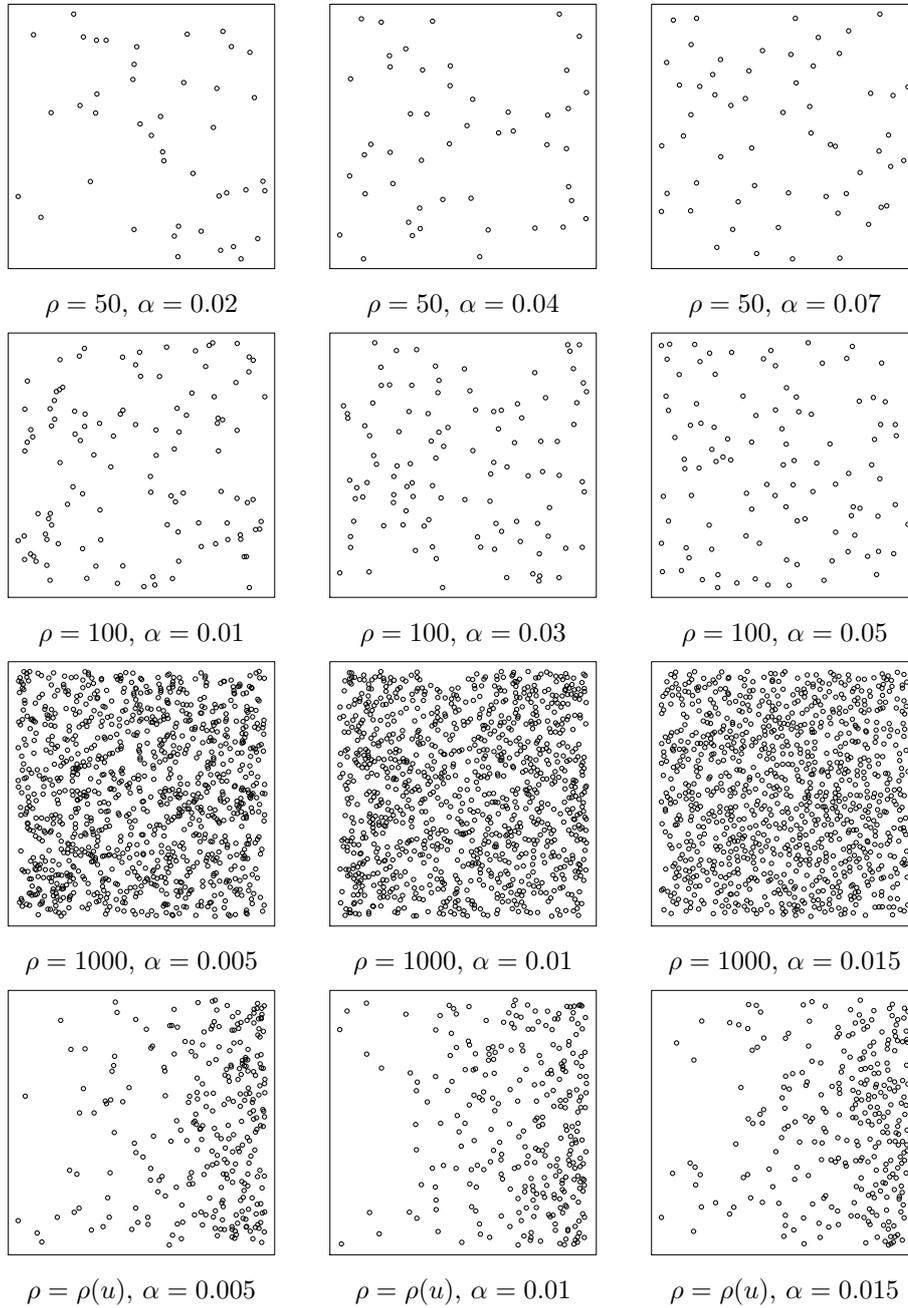


Figure 4: Examples of point patterns simulated from a Bessel-type DPP on $[0, 1]^2$ for different values of ρ and α . For the last row, $\rho(x, y) = 20 \exp(4x)$.

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

Table 1: Estimated root mean square errors ($\times 10^3$) and mean computation time (in seconds) of $\hat{\alpha}$ for a Bessel-type DPP on $[0, 1]^2$, for different values of ρ and α . The 3 first estimators use the test function (4.2.3) with $R = 0.05$, $R = 0.1$ and $R = 0.25$ respectively, while the last estimator is the adaptive version based on (4.2.4). The standard errors of the RMSE estimations are given in parenthesis. The last column gives the averages of "practical ranges" (i.e. maximal solution to $|g(r) - 1| = 0.01$) used for the adaptive estimator, along with their standard deviations in parenthesis. For each value of ρ and α , these quantities are computed from 1000 simulations where all four estimation methods have converged.

one realization of a DPP with a Gaussian kernel with range ψ . For this example $e(\psi)$ exhibits many different roots, although the dataset contains a fairly large number of points (about 1000). The consequence is an extreme sensitivity to the initial parameter when we try to solve $e(\psi) = 0$. In contrast $e_2(\psi) = 0$ has one clear solution. This advocates the use of the two-step approach.

Due to the aforementioned very strong sensitivity to the initial value of ψ , conclusions from comparison of the simultaneous estimate of ψ with the two-step estimate of ψ can be quite arbitrary. However, we report in Figure 7 in the appendix the distribution of estimates of ρ from 1000 simulations of a Bessel-type DPP with $\rho = 1000$ and $\psi = \alpha = 0.01$, using either (4.4.5) from the simultaneous approach or $\hat{\rho} = N(X \cap W)/|W|$ from the two-step approach. For the simultaneous method we either chose the true value $\alpha = 0.01$ as the starting point for the numerical solution of $e(\alpha) = 0$ to get $\hat{\alpha}$, or fix $\hat{\alpha}$ at the true value, i.e. $\hat{\alpha} = 0.01$, in (4.4.5). The estimate $\hat{\rho} = N(X \cap W)/|W|$ is unequivocally better than (4.4.5) in terms of root mean square error, even when the true value of α is used for $\hat{\alpha}$ in (4.4.5). This confirms our recommendation.

The simultaneous estimation approach in this example is covered by our theoretical results in Sections 4.3.1 and 4.3.2. It shows that while our consistency result guarantees the existence of a consistent sequence of parameter estimates (roots) there could also exist other non-consistent sequences.

4.5 Discussion

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

The performance of second order estimating functions depends strongly on a tuning parameter R that controls which pairs of points are used in the estimation. Although not statistically optimal, our adaptive procedure for selecting this tuning parameter is intuitively appealing and easy to implement. The method depends on a new tuning parameter ε for which it is easier to identify reasonable values than for the original tuning parameter R . The resulting estimation procedure is computationally tractable and performs well in terms of mean squared error in the simulation studies considered. It moreover seamlessly integrates with the asymptotic results where the use of the adaptive method poses no extra theoretical difficulties.

Though we focus in this chapter on determinantal point processes, the adaptive method is applicable for any spatial point process with known pair correlation function. As an example we provide in Section 4.E of the appendix a simulation study in case of a cluster process.

4.A Assumptions and proof of Theorem 4.2.1

Our general Theorem 4.2.1 depends on a number of assumptions. The setting is the same as in Section 4.2.2. We moreover define $\text{diam}(x)$ as the largest distance between two coordinates of x . The assumptions (F1) through (F3) are mainly related to the test functions f_i , while for X we assume (X1) through (X3).

- (F1) For all $i = 1, \dots, l$ and for all $x \in (\mathbb{R}^d)^{q_i}$, $\theta \mapsto f_i(x; \theta)$ is twice continuously differentiable in a neighbourhood of θ^* . Moreover, the first and second derivative

of f_i with respect to θ are bounded with respect to $x \in (\mathbb{R}^d)^{q_i}$ uniformly in θ belonging to this neighbourhood.

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

Define the matrices $H_n(\theta)$ by

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

where for all i

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

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

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

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

- (X1) For all θ in a neighbourhood of θ^* and all q_i , $i = 1, \dots, l$, the intensity functions $x \mapsto \rho^{(q_i)}(x; \theta)$ are well-defined and bounded. Moreover, $\theta \mapsto \rho^{(q_i)}(x; \theta)$ is twice continuously differentiable in a neighbourhood of θ^* , for all $x \in (\mathbb{R}^d)^{q_i}$. Finally, the first and second derivative of $\rho^{(q_i)}$ with respect to θ are bounded with respect to $x \in (\mathbb{R}^d)^{q_i}$ uniformly in θ belonging to this neighbourhood.

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

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

with R coming from (F2).

- (X3) X satisfies the central limit theorem

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

where e_n is defined in Section 4.2.2 and $\Sigma_n = \text{Var}(e_n(\theta^*))$.

Assumptions (F1) and (F2) are basic regularity conditions on the f_i 's. Similarly (X1) and (X2) ensure that the intensity functions of X exist and are sufficiently regular. The technical assumptions are in fact (F3) (or (F3')) and (X3). While the latter strongly depends on the underlying point process (see [103] for Cox processes and Theorem 2.4.4 for DPPs), the former can be simplified in some cases. For example, if

$H_n(\theta^*)$ are symmetrical matrices for all n then (F3) writes $\liminf_n \lambda_{\min}(H_n(\theta^*)) > 0$ where $\lambda_{\min}(H_n(\theta^*))$ denotes the smallest eigenvalue of $H_n(\theta^*)$. If the matrices $H_n(\theta^*)$ are not symmetrical, Assumption (F3') will be preferred since (F3) does not translate well for non-symmetrical matrices. Furthermore, if X is stationary, all f_i 's are invariant by translation, and the sequence of windows $\{W_n\}_{n \geq 1}$ satisfies (W) in Section 4.3.1, then $H_n(\theta)$ converges, towards a matrix $H(\theta)$ explicitly given in Lemma 4.A.1 below. Assumption (F3) thus simply becomes $\inf_{\|\phi\|=1} \phi^T H(\theta^*) \phi > 0$ and (F3') is satisfied whenever $H(\theta^*)$ is invertible by continuity of $H(\theta)$. In specific applications of Theorem 4.2.1, further conditions on the sequence of observation windows $\{W_n\}_{n \geq 1}$ may be required, see e.g. (W) in Section 4.3.1.

Lemma 4.A.1. *Assume (W), (X1), (F2) and let $\theta \in \mathbb{R}^p$. Suppose that all $\rho^{(q_i)}(\cdot; \theta)$'s and $f_i(\cdot; \theta)$'s are invariant by translation, i.e. $f_i(u_1, u; \theta) = f_i(0, u - u_1; \theta)$ where we denote by u the vector (u_2, \dots, u_{q_i}) . If $u \mapsto f_i(0, u; \theta)$ is integrable for all i such that $q_i \geq 2$, then $H_n(\theta)$ converges to a matrix $H(\theta)$. In particular, for all i we have*

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

This lemma is verified in Section 4.B. We now turn to the proof of the theorem. To prove the consistency of $\hat{\theta}_n$ and get its rate of convergence we apply the following result, where $\|\cdot\|$ stands for any matrix norm.

Theorem 4.A.2 ([103]). *Suppose that $e_n(\theta)$ is continuously differentiable with respect to θ and define*

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

Suppose that for all $\alpha > 0$

$$\sup_{\theta \in M_n^\alpha(\theta^*)} \left\| \frac{1}{|W_n|} (J_n(\theta) - J_n(\theta^*)) \right\| \xrightarrow{\mathbb{P}} 0, \quad (4.A.1)$$

where

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

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

$$\mathbb{P} \left(\frac{1}{|W_n|} \inf_{\|\phi\|=1} \phi^T J_n(\theta^*) \phi < l \right) \rightarrow 0. \quad (4.A.2)$$

Assume, moreover, that the class of random vectors

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

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

$$\mathbb{P}(\exists \hat{\theta}_n : e_n(\hat{\theta}_n) = 0 \text{ and } |W_n| \|\hat{\theta}_n - \theta^*\| < A) > 1 - \varepsilon \quad (4.A.3)$$

for a sufficiently large n .

We now verify the assumptions of Theorem 4.A.2. There is no loss in generality by assuming that all f_i are symmetric functions. Otherwise we can just replace $f_i(x)$ by its symmetrized version $(q_i!)^{-1} \sum_{u \in \pi(x)} f_i(u)$ where $\pi(x)$ denotes the set of all vectors obtained by permuting the components of x . This does not change the value of $e_n(\theta)$ and each symmetrized function still satisfies Assumptions (F1) through (F3). We will use at several places the following result.

Lemma 4.A.3. *Let X be a point process satisfying Assumption (X2). Consider any $i \in \{1, \dots, l\}$, any bounded set $W \subset \mathbb{R}^d$, and any symmetric bounded function $g : (\mathbb{R}^d)^{q_i} \rightarrow \mathbb{R}^{k_i}$ vanishing when two of its components are at a distance greater than R for a given constant $R > 0$. Then*

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

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

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

where $|S|$ denotes the cardinality of S .

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

$$\begin{aligned} & \sum_{k=0}^{q_i} \mathbb{E} \left[\sum_{\substack{S, T \subset X \cap W \\ |S|=|T|=q_i, |S \cap T|=k}} g(S) g^T(T) \right] \\ &= \sum_{k=0}^{q_i} \mathbb{E} \left[\sum_{\substack{U \subset X \cap W \\ |U|=2q_i-k}} \sum_{\substack{S' \subset S \subset U \\ |S'|=k, |S|=q_i}} g(S) g^T(S' \cup (U \setminus S)) \right] \\ &= \sum_{k=0}^{q_i} \frac{1}{(2q_i - k)!} \int_{W^{2q_i-k}} \sum_{\substack{S' \subset S \subset \{x_1, \dots, x_{2q_i-k}\} \\ |S'|=k, |S|=q_i}} g(S) g^T(S' \cup (\{x_1, \dots, x_{2q_i-k}\} \setminus S)) \rho^{(2q_i-k)}(x; \theta^*) dx \\ &= \sum_{k=0}^{q_i} \frac{\binom{q_i}{k} \binom{2q_i-k}{q_i}}{(2q_i - k)!} \int_{W^{2q_i-k}} g(x_1, \dots, x_{q_i}) g^T(x_1, \dots, x_k, x_{q_i+1}, \dots, x_{2q_i-k}) \rho^{(2q_i-k)}(x; \theta^*) dx. \end{aligned} \tag{4.A.4}$$

By Assumption (X2), the functions $\rho^{(q_i)}, \dots, \rho^{(2q_i-1)}$ are all bounded. Moreover, as a consequence of our assumptions on g , each entry of each term for $k \geq 1$ in (4.A.4) is bounded by

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

which is $O(|W|)$. However, we are unable to get a similar control for $k = 0$. Therefore, instead of controlling this term alone, we consider its difference with the remaining

term in the variance we are looking at, that is

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

All of its components are bounded by

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

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

The regularity conditions on $e_n(\theta)$ in Theorem 4.A.2 are consequences of (F1), (X1). The stochastic behavior of $e_n(\theta^*)$ is easily deduced from the previous lemma.

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

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

is stochastically bounded.

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

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

we know that $\text{Var}(e_n^i(\theta^*))$ is $O(|W_n|)$ by Lemma 4.A.3 under Assumptions (X2) and (F2). \square

To apply Theorem 4.A.2 under Assumptions (F1) through (F3), it remains to show the following lemma.

Lemma 4.A.5. *Under Assumptions (F1) through (F3), (X1) and (X2) we have for all $\alpha > 0$,*

$$\sup_{\theta \in M_n^\alpha(\theta^*)} \left\| \frac{1}{|W_n|} (J_n(\theta) - J_n(\theta^*)) \right\| \xrightarrow{\mathbb{P}} 0. \quad (4.A.5)$$

where $M_n^\alpha(\theta^*)$ is defined as in Theorem 4.A.2 and there exists $l > 0$ such that

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

Proof. Let $\alpha > 0$ and an integer n big enough such that $M_n^\alpha(\theta^*)$ is a neighbourhood of θ^* where Assumptions (F1), (F2) and (X1) are satisfied. For all $\theta \in M_n^\alpha(\theta^*)$, we write

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

For all i , the derivative of $x \mapsto f_i(x; \theta) \rho^{(q_i)}(x; \theta)$ is bounded on $W_n^{q_i}$. So, we can write

$$\begin{aligned} J_n^i(\theta) = & -q_i! \sum_{\substack{x \subset X \cap W_n \\ |x|=q_i}} \frac{d}{d\theta^T} f_i(x; \theta) + \int_{W_n^{q_i}} \frac{d}{d\theta^T} f_i(x; \theta) \rho^{(q_i)}(x; \theta) dx \\ & + \int_{W_n^{q_i}} f_i(x; \theta) \nabla_\theta \rho^{(q_i)}(x; \theta)^T dx. \end{aligned} \quad (4.A.6)$$

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

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

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

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

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

It remains to prove that there exists $l > 0$ such that (4.A.5) holds. By Assumption (F3) choose $\varepsilon > 0$ so that $\liminf_{n \rightarrow \infty} \phi^T H_n(\theta^*) \phi > \varepsilon$ and let $l = \varepsilon/2$. In the case where $\theta = \theta^*$, the second term in (4.A.6) is just the expectation of the first one and the third term is equal to $|W_n| H_n^i(\theta^*)$ which is deterministic. Thus when $\theta = \theta^*$, the L^2 norm of the first two terms in (4.A.6) is equal to

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

which is $O(\sqrt{|W_n|})$ by Lemma 4.A.3. Hence it vanishes in probability when divided by $|W_n|$. Denote by a_n the first two terms in $\phi^T J_n(\theta^*) \phi / |W_n|$ and by b_n the last term which is $\phi^T H_n(\theta^*) \phi$. Then

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

which concludes the proof. \square

To apply Theorem 4.A.2 under the alternative Assumption (F3') instead of (F3), we proceed as follows. Consider n to be large enough and θ to be in a neighbourhood of θ^* such that $H_n(\theta)$ is invertible and $H_n(\theta)^{-1}$ is uniformly bounded with respect to n and θ . Let $\tilde{e}_n(\theta) = H_n(\theta)^{-1}e_n(\theta)$ and let us show that we can apply Theorem 4.A.2 to \tilde{e}_n . Obviously, \tilde{e}_n has the same roots as e_n , is continuously differentiable since $\theta \mapsto H_n(\theta)$ and $\theta \mapsto H_n(\theta)^{-1}$ are continuously differentiable from Assumptions (F1) and (F2), and the family $\{\tilde{e}_n(\theta^*)/\sqrt{|W_n|} : n \in \mathbb{N}\}$ is stochastically bounded. Let $\tilde{J}_n(\theta) = -\frac{d}{d\theta^T}\tilde{e}_n(\theta)$. It remains to show the following lemma.

Lemma 4.A.6. *Under Assumptions (F1), (F2), (F3'), (X1) and (X2) we have*

$$\sup_{\theta \in M_n^\alpha(\theta^*)} \left\| \frac{1}{|W_n|} (\tilde{J}_n(\theta) - \tilde{J}_n(\theta^*)) \right\| \xrightarrow{\mathbb{P}} 0. \quad (4.A.7)$$

and

$$\mathbb{P} \left(\lim_{n \rightarrow \infty} \frac{1}{|W_n|} \inf_{\|\phi\|=1} \phi^T \tilde{J}_n(\theta^*) \phi < 1/2 \right) = 0. \quad (4.A.8)$$

Proof. We have

$$\tilde{J}_n(\theta) = H_n(\theta)^{-1}J_n(\theta) - T_n(\theta) \quad \text{where} \quad T_n(\theta)_{i,j} = \sum_{k=1}^p \frac{\partial}{\partial \theta_j} H_n(\theta)_{i,k}^{-1} e_n(\theta)_k. \quad (4.A.9)$$

For any $\theta \in M_n^\alpha(\theta^*)$, since all terms in (4.A.6) are bounded by Assumptions (F1) and (X1), using Assumption (F2) we get

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

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

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

Moreover,

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

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

From Lemma 4.A.5 we know that

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

Finally, we observe that

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

where the boundedness of $\partial H_n(\theta)^{-1}/\partial\theta_j$ for each j was noted above and the boundedness of $\mathbb{E}\left(\sup_{\theta \in M_n^\alpha(\theta^*)} \|J_n(\theta)\|\right)$ follows by considerations in the last part of the proof of Lemma 4.A.5 as a consequence of the regularity assumptions imposed on $H_n(\theta)$ by Assumption (F3'). This finishes proving (4.A.7). The result (4.A.8) is then a consequence of the fact that $H_n(\theta^*)^{-1}J_n(\theta^*)$ converges towards I_p when n goes to infinity. \square

Finally, by Lemmas 4.A.4, 4.A.5 and 4.A.6, we can apply Theorem 4.A.2 and the first part in the statements of Theorem 4.2.1 is deduced.

Now, for each $n \in \mathbb{N}$, we define $\hat{\theta}_n$ as the closest root of e_n to θ^* , if e_n has any, otherwise let $\hat{\theta}_n = 0$. Theorem 4.A.2 tells us that $\mathbb{P}(e_n(\hat{\theta}_n) = 0) \rightarrow 1$ and $\sqrt{|W_n|}(\hat{\theta}_n - \theta^*)$ is bounded in probability.

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

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

We know that $e_n(\hat{\theta}_n)/\sqrt{|W_n|}$ converges in distribution towards 0 and by Theorem 4.A.2 we also know that

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

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

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

4.B Proof of Lemma 4.A.1

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

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

where

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

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

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

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

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

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

4.C Identifiability of the kernel of a continuous DPP

In this section, we show that most continuous DPP kernels are uniquely characterized by their associated DPP distribution. In particular, DPPs with positive first order intensity function are characterized by a unique continuous kernel. This is for instance the case for any stationary DPP with intensity $\rho > 0$.

Proposition 4.C.1. *Let K and \widetilde{K} be two continuous DPP kernels generating the same DPP distribution. Then, there exists a closed set A such that*

$$\widetilde{K}(x, y) = \begin{cases} K(x, y) & \text{if } (x, y) \in A^2 \cup (A^c)^2 \\ -K(x, y) & \text{if } (x, y) \in A \times A^c \cup A^c \times A. \end{cases}$$

In particular, $K(x, y) = \widetilde{K}(x, y) = 0$ for all $(x, y) \in \partial A \times \mathbb{R}^d \cup \mathbb{R}^d \times \partial A$ and $K(x, x) = \widetilde{K}(x, x) = 0$ for all $x \in \partial A$ since K and \widetilde{K} are continuous.

Proof. Since K and \widetilde{K} are continuous then their n -th order intensity functions are also continuous and are thus equal everywhere. In particular, for any finite set D , the matrices $(K(x, y))_{x, y \in D}$ and $(\widetilde{K}(x, y))_{x, y \in D}$ are symmetrical and have the same principal minors. As shown in [90] using a result of [40], we get that there exists a function $s(x)$ taking values in $\{\pm 1\}$ such that $\widetilde{K}(x, y) = s(x)K(x, y)s(y)$ for all $x, y \in D$. Since this is true for any finite set D , this result can directly be extended to any countable dense subset D of \mathbb{R}^d . Now, let A be the closure of $s^{-1}(\{1\})$, then, by continuity of K and \widetilde{K} ,

$$\widetilde{K}(x, y) = \begin{cases} K(x, y) & \text{if } (x, y) \in A^2 \cup (A^c)^2 \\ -K(x, y) & \text{if } (x, y) \in A \times A^c \cup A^c \times A. \end{cases}$$

□

Corollary 4.C.2. *Any DPP distribution on \mathbb{R}^d with positive first order intensity function admits at most one continuous kernel generating it.*

Proof. By proposition 4.C.1, for any pair of continuous kernel K, \widetilde{K} generating the same DPP distribution there exists a closed set A such that $K(x, y) = \widetilde{K}(x, y)$ if x and y are both in A or its complementary and $K(x, x) = \widetilde{K}(x, x) = 0$ on the boundary of A . But, since $K(x, x) > 0$ everywhere then $\partial A = \emptyset$ and $K = \widetilde{K}$ everywhere. □

4.D Examples of DPP families satisfying (F3) or (F3')

We illustrate the use of Lemmas 4.3.2, 4.3.3 and 4.3.5 through two examples which correspond to the kernel families used in our simulation study.

- (i) As a first example, we consider the stationary Gaussian kernel on \mathbb{R}^d

$$K(u, v) = \rho \exp(-\|v - u\|^2/\alpha^2)$$

with parameters $\theta = (\rho, \alpha)$ where $0 < \rho \leq (\sqrt{\pi}\alpha)^{-d}$ in which case the test function considered in Lemma 4.3.2 (which is also the one used in our simulation study) writes

$$\begin{aligned} f(u, v; \theta) &= \omega \left(\frac{\varepsilon g(0, 0; \theta)}{g(u, v; \theta)} \right) \frac{\nabla_{\theta} \rho^{(2)}(u, v; \theta)}{\rho^{(2)}(u, v; \theta)} \\ &= \omega \left(\varepsilon \exp(2\|v - u\|^2/\alpha^2) \right) \left(\frac{-4\|v - u\|^2 \exp(-2\|v - u\|^2/\alpha^2)}{\alpha^3(1 - \exp(-2\|v - u\|^2/\alpha^2))} \right). \end{aligned}$$

Since $u \mapsto f(0, u; \theta)$ has bounded support and $\lim_{\|v - u\| \rightarrow 0} f(u, v; \theta) = \begin{pmatrix} 2/\rho \\ -2/\alpha \end{pmatrix}$

then $u \mapsto f(0, u; \theta)$ is integrable for any θ .

Moreover, since one entry of the vectors $\nabla_{\theta} \rho^{(2)}(u, v; \theta)/\rho^{(2)}(u, v; \theta)$ is constant and the other one is not, then

$$\text{span}\{\nabla_{\theta} \rho^{(2)}(0, u; \theta), \|u\|^2 \leq -\alpha^2 \log(\varepsilon)/2\} = \mathbb{R}^2$$

for all θ . This proves that the assumptions of Lemma 3.2 are satisfied, which in turn implies (F3). Similarly, if we consider the two-step estimation with the same kernels and the same f , then the same reasoning shows that we can apply Lemma 3.5 and get that (F3') is satisfied in that case. All classical families of stationary DPP kernels can be treated in a similar way.

- (ii) As a second example, we consider the non stationary DPP on \mathbb{R}^2 used in our simulation study, with kernel

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

where we assume that ρ is bounded and satisfies $\rho(u) \geq C$ for some constant $C > 0$ and all $u \in \mathbb{R}^d$ and where the parameter is α where $0 < \alpha \leq (\pi\|\rho\|_{\infty})^{-1/2}$. Using a second order estimating function with the same f as before, we get that the function h in Lemma 3.3 writes

$$\begin{aligned} h(u, v) &= w \left(\varepsilon \left(\frac{\|v - u\|/\alpha^*}{J_1(2\|v - u\|/\alpha^*)} \right)^2 \right) \\ &\quad \times \sqrt{\rho(u)\rho(v)} \frac{16}{(\alpha^*)^2} \frac{\left(\frac{J_2(2\|v - u\|/\alpha^*)J_1(2\|v - u\|/\alpha^*)}{\|v - u\|/\alpha^*} \right)^2}{1 - \left(\frac{J_1(2\|v - u\|/\alpha^*)}{\|v - u\|/\alpha^*} \right)^2} \end{aligned}$$

whereby

$$\left| \int_{\mathbb{R}^d} h(u, v) dv \right| \leq \frac{16\|\rho\|_{\infty}}{(\alpha^*)^2} \int_{\mathbb{R}^d} \frac{\left(\frac{J_2(2\|t\|/\alpha^*)J_1(2\|t\|/\alpha^*)}{\|t\|/\alpha^*} \right)^2}{1 - \left(\frac{J_1(2\|t\|/\alpha^*)}{\|t\|/\alpha^*} \right)^2} \mathbb{1}_{\{\sqrt{\varepsilon}\|t\|/\alpha^* \leq J_1(2\|t\|/\alpha^*)\}} dt.$$

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

Table 2: Percentage of times the estimation methods have converged for the models and estimators considered in Section 4.4.

The expression inside the integral is a $O(\|t\|^2)$ when $\|t\| \rightarrow 0$ (it can be shown by a Taylor expansion) and equal to 0 outside a compact set, hence

$$\sup_u \left| \int_{\mathbb{R}^d} h(u, v) dv \right| < \infty.$$

Moreover,

$$\nabla_{\alpha} \rho^{(2)}(u, v; \alpha^*) \leq -\frac{4C}{\alpha^*} \frac{J_2(2\|v - u\|/\alpha^*) J_1(2\|v - u\|/\alpha^*)}{\|v - u\|/\alpha^*}$$

which is negative when $\|v - u\|$ is greater than 0 but lower than $\alpha^* j_1/2$ where j_1 denotes the first zero of J_1 . Therefore, we can always choose constants $\varepsilon_1, \varepsilon_2, \delta$ such that $|\nabla_{\alpha} \rho^{(2)}(u, v; \alpha^*)| > \delta$ and $\varepsilon \|v - u\|/\alpha^* \leq J_1(2\|v - u\|/\alpha^*)$ for all u, v such that $\varepsilon_1 \geq \|v - u\| \geq \varepsilon_2$. This shows that the assumptions of Lemma 3.3 are satisfied which implies (F3).

4.E Supplementary tables for Section 4.4.1

For the simulation study carried out in Section 4.4, considering estimation for a DPP model with a Bessel-type kernel, we report in Figure 5 the boxplots representing the distribution of the estimators and in Table 2 the percentages of times each estimation method has converged in our simulation study. These percentages are similar for all estimation methods. Table 3 displays the root mean square errors of the estimators considered in Section 4.1 where, for comparison, we also include results for the adaptive estimator using $\varepsilon = 0.05$. Conclusions based on these tables are given in Section 4.4.

Two-step versus simultaneous

Referring to Section 4.4.2, Figure 6 shows how irregular the contrast function $e(\psi)$ for the simultaneous approach can be in comparison with the contrast function $e_2(\psi)$ for the two-step approach. The underlying point pattern is displayed on the left. This is a realisation on the unit square of a homogeneous DPP with a Gaussian kernel, with intensity $\rho = 1000$ and range $\alpha = 0.01$.

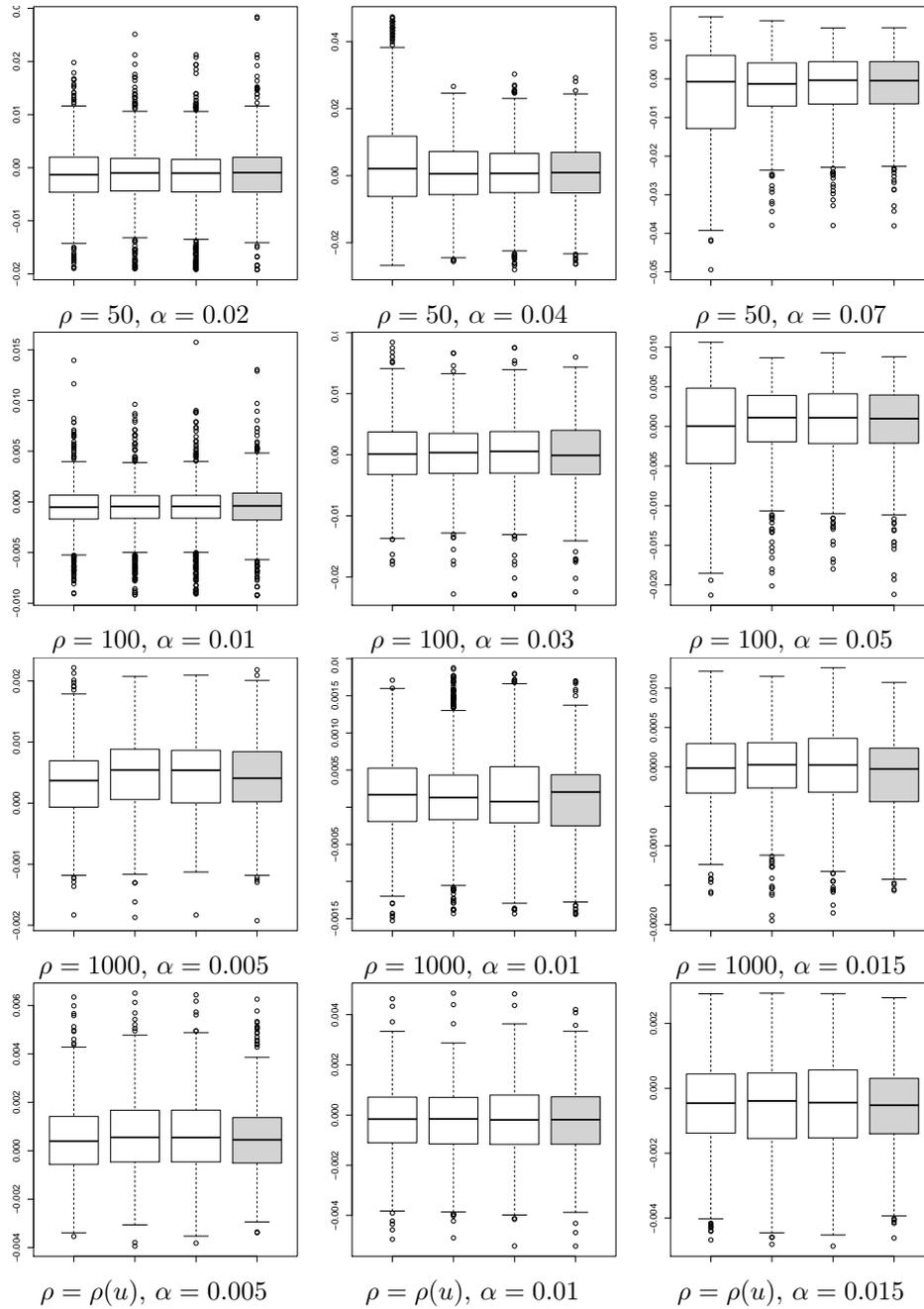


Figure 5: Distribution of $\hat{\alpha} - \alpha$ for a Bessel-type DPP on $[0, 1]^2$ for different values of ρ and α . In each subfigure, the 3 first estimators on the left use the test function (4.2.3) with $R = 0.05$, $R = 0.1$ and $R = 0.25$ respectively, while the last estimator is the adaptive version based on (4.2.4).

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

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

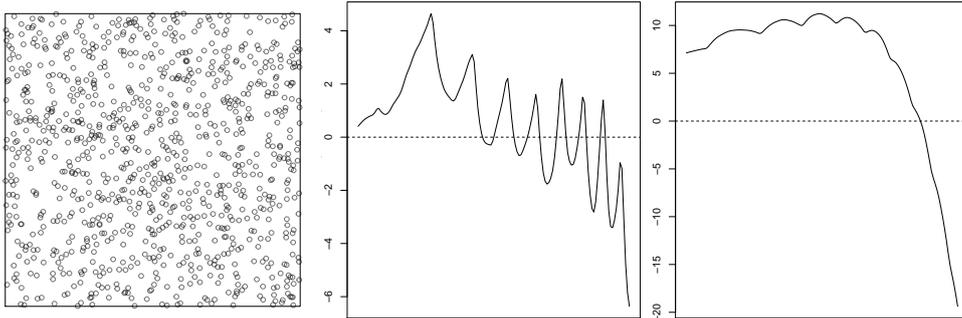


Figure 6: Realization of a DPP (left) and plots of the contrast function $e(\psi)$ (see (4.4.4)) of the simultaneous procedure (middle) and of the contrast function $e_2(\psi)$ (see (4.4.3)) of the two-step procedure (right) obtained for the data in the left plot.

Figure 7 reports the distributions of estimates of ρ over 1000 realisations on the unit square of a DPP with a Bessel-type kernel with $\rho = 1000$ and $\alpha = 0.01$. The two first estimators come from the simultaneous approach, see equation (4.4.5) where $\hat{\psi} = \hat{\alpha}$ in this setting. For the first one, the numerical solution of $e(\alpha) = 0$ to get $\hat{\alpha}$ was initialized at the true value 0.01 of α . For the second one, $\hat{\alpha}$ was fixed to the true value, i.e. $\hat{\alpha} = 0.01$. The last estimator on the right of Figure 7 is simply $\hat{\rho} = N(X \cap W)/|W|$, corresponding to the first step of the two-step procedure. The respective root mean square errors are 33.6, 31.4 and 26.

Some simulations for the Thomas model

The adaptive estimating function is also useful for clustered point processes. Here we consider a Thomas model on $[0, 1]^2$, with parent intensity $\kappa = 100$, offspring intensity $\mu = 10$ and various values of the dispersal kernel standard deviation σ . The same three estimation methods as in Section 4.4 have been evaluated, where for the adaptive

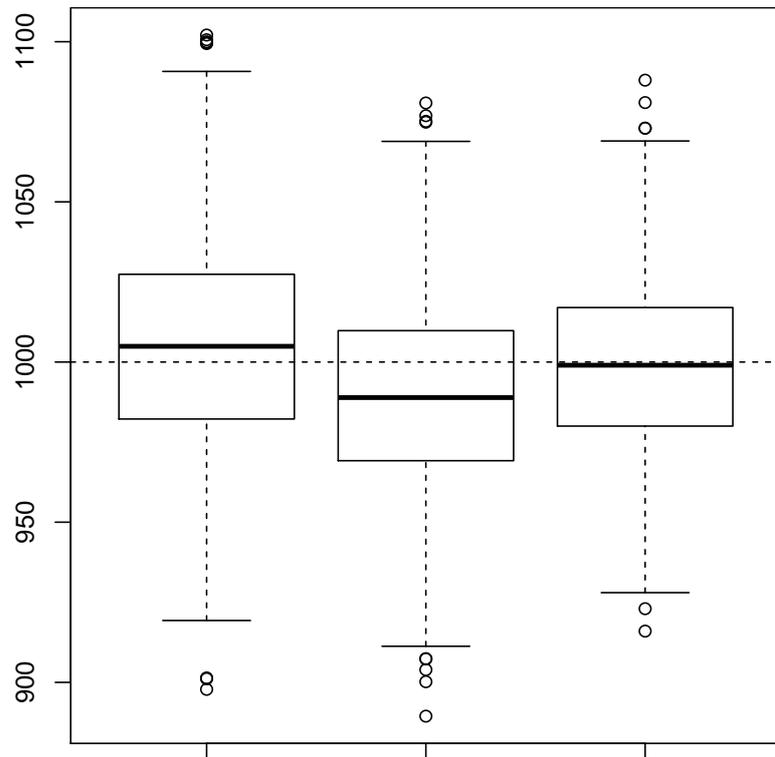


Figure 7: Distribution of estimates of ρ obtained from 1000 realisations of the Bessel-type DPP on $W = [0, 1]^2$ with $\rho = 1000$ and $\alpha = 0.01$. Left: the simultaneous estimator as given in equation (4.4.5) with initial value for the numerical solution given by the true value 0.01 of α . Middle: as left but using the true value of α instead of $\hat{\alpha}$. Right: $\hat{\rho} = N(X \cap W)/|W|$ corresponding to the first step of a two-step procedure.

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

Table 4: For the Thomas model, estimated root mean square errors of various estimators of κ and σ ($\times 10^3$). The 3 first estimators use the test function (4.2.3) with $R = 0.05$, $R = 0.1$ and $R = 0.25$ respectively; the fourth and fifth estimators are the adaptive version based on (4.2.4) where $\varepsilon = 0.01$ and $\varepsilon = 0.05$; the three last estimators are from the library `spatstat`: based on K , on Guan’s composite likelihood (clik) and on Palm likelihood - all with default settings. The standard errors of the MSE estimations are given in parenthesis.

version both $\varepsilon = 0.01$ and $\varepsilon = 0.05$ have been considered. A point pattern sample and the distribution of the estimators of κ and σ based on 1000 replications are shown in Figure 4.E for $\sigma = 0.02$, $\sigma = 0.035$ and $\sigma = 0.05$ respectively. Estimators of the library `spatstat` [7] of R [87] with default settings have also been added. These are: minimum contrast estimation based on the K -function, Guan’s composite likelihood, and Palm likelihood, see also Section 4.2.3. Table 4 summaries the estimated root mean square errors for each estimation method.

Also for the Thomas process, the adaptive method, both with $\varepsilon = 0.01$ and $\varepsilon = 0.05$, performs well compared with the three fixed R estimators. In fact for $\sigma = 0.05$, the adaptive versions are better than any of the fixed R estimators. The adaptive method also has good stable performance compared with the three `spatstat` methods. In particular, the adaptive method performs much better than Guan’s composite likelihood with default settings.

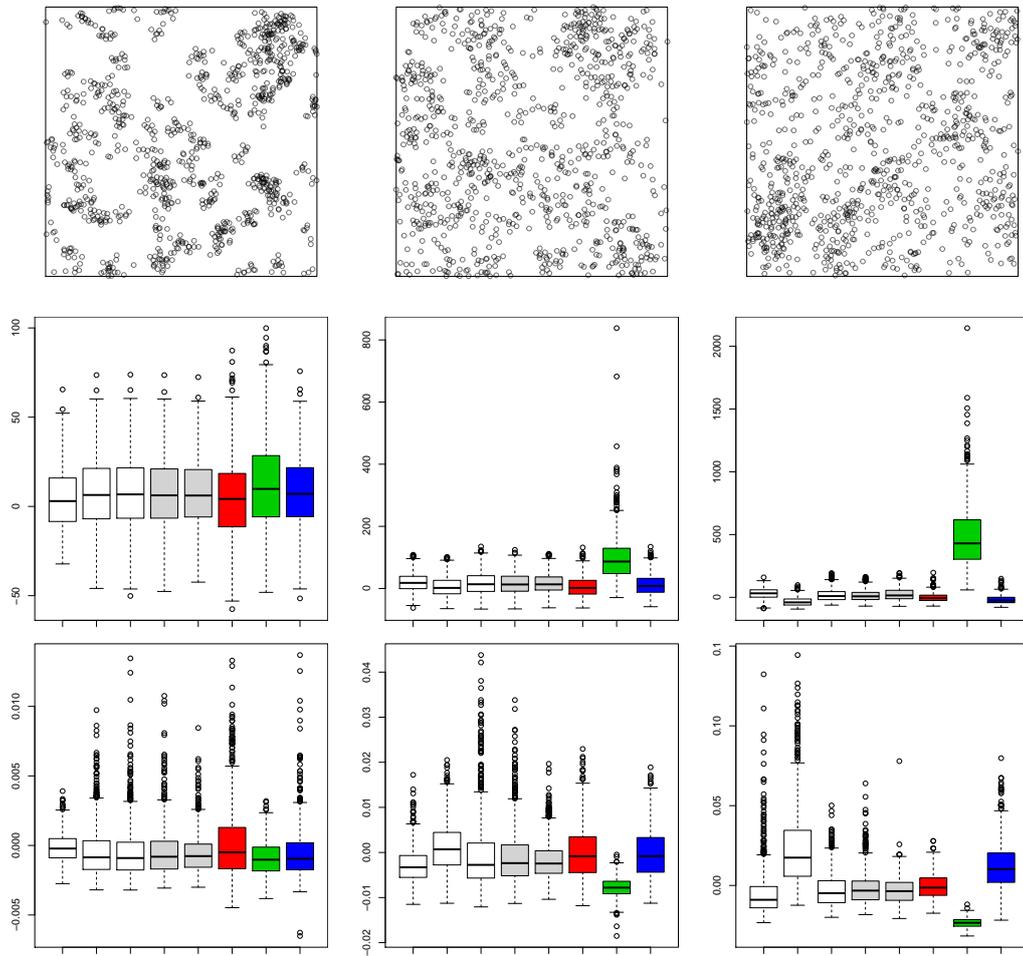


Figure 8: First row: Examples of point patterns simulated from a Thomas model on $[0, 1]^2$ for $\kappa = 100$, $\mu = 10$ and from left to right $\sigma = 0.02, 0.035, 0.05$. Second row: Distribution of estimates of κ based on 1000 replications. In each plot, the 3 first boxplots are for estimates obtained with the test function (4.2.3) with $R = 0.05$, $R = 0.1$ and $R = 0.25$ respectively; the fourth and fifth boxplots (in grey) are for the adaptive version based on (4.2.4) where $\varepsilon = 0.01$ (left) and $\varepsilon = 0.05$ (right); the three last boxplots are for methods from `spatstat`: based on K (red), on Guan's composite likelihood (green) and on Palm likelihood (blue) - all with default settings. Third row: Distribution of estimates of σ based on 1000 replications, using the same estimation methods.

Chapter 5

Consistency and approximation of the likelihood of stationary continuous determinantal point processes

5.1 Introduction

Expressions of the densities of determinantal point processes are known since they were first described as a model of fermion systems in [75]. For DPPs defined on finite spaces, the expression of its likelihood can be easily computed and its asymptotic properties in an iid setting have been well studied [24]. On the opposite, the likelihood expression of continuous DPPs is hardly tractable. Computing the likelihood requires the knowledge of a function that can only be obtained by solving an integral equation or by knowing the spectral representation of any projection of the locally trace class operator defined from the DPP kernel. Both methods are not feasible in practice. Moreover, part of the likelihood writes as the log-determinant of a random kernel matrix whose behavior is difficult to control. However, some results on the maximum likelihood estimator (MLE) of continuous DPPs are known in the iid setting [10], but because of the aforementioned difficulties no asymptotic result is known in the case of increasing domain asymptotic inference of continuous DPPs.

In the stationary case, one alternative solution to compute the likelihood proposed in [69] is to consider a Fourier series approximation of the DPP kernel on rectangular observation windows. Since the spectral decomposition of a Fourier series is explicit, the associated likelihood can then be computed. Another alternative solution given in [12] is to use a Markov chain Monte Carlo inference method based on bounds on the likelihood that does not depend on the spectral decomposition of its kernel. Neither of these two contributions studies the consistency of their approach.

In this chapter, we give an asymptotic approximation of the DPP likelihood as well as a way to correct the edge effects arising as a consequence of this approximation and we show that it improves upon [69]. We also prove that the MLE is consistent for any DPP defined on a small enough regular grid of \mathbb{R}^d as an approximation of the actual DPP.

The remainder of the chapter is as follows. In Section 5.2, we recall the form of the likelihood of DPPs and we suggest an asymptotic approximation of this likelihood as well as an edge correction method. In Section 5.3, we use various simulations of DPPs

to compare our approximate MLE to common minimum contrast inference method. Section 5.4 contains the proof of the MLE consistency. Finally, the appendix 5.A contains some technical results needed in Section 5.4.

5.2 Likelihood of DPPs

5.2.1 Likelihood expression

Let X be a DPP on (\mathbb{R}^d, μ) with kernel K^{θ^*} and associated integral operator \mathcal{K}^{θ^*} . Here, μ will either be the Lebesgue measure on \mathbb{R}^d or the counting measure on a regular grid of \mathbb{R}^d and K^{θ^*} belongs to a parametric family $\{K^\theta, \theta \in \Theta\}$ where Θ is a compact subset of \mathbb{R}^p for some integer $p \geq 1$. We consider an increasing domain asymptotic framework, meaning that we want to estimate θ^* from a unique observation of $X \cap W_n$ where W_n is an increasing sequence of subsets of \mathbb{R}^d and we write $N(W_n)$ for the number of points of $X \cap W_n$. Moreover, for all finite sets $X \subset \mathbb{R}^d$ and functions $F : (\mathbb{R}^d)^2 \rightarrow \mathbb{R}$, we write $F[X]$ for the matrix $(F(x, y))_{x, y \in X}$. If $F(x, y)$ only depends on $y - x$ then we write $F(x, y) = F_0(y - x)$. Similarly, if $F(x, y)$ only depends on $\|y - x\|$ then we also write $F(x, y) = F_0(\|y - x\|)$ by an abuse of notation. When the kernel of a DPP satisfies the former property then the DPP is stationary and when it satisfies the latter, the DPP is isotropic. We also denote by \hat{F}_0 the Fourier transform of F_0 on (\mathbb{R}^d, μ) defined by

$$\hat{F}_0(x) := \int_{\mathbb{R}^d} F_0(t) \exp(-2i\pi t \cdot x) d\mu(t).$$

Finally, for all finite hermitian matrices M we write $\lambda_{\min}(M)$ (resp. $\lambda_{\max}(M)$) for the lowest (resp. highest) eigenvalue of M . If M has a countable number of rows and columns then by an abuse of notation we denote by $\lambda_{\min}(M)$ (resp. $\lambda_{\max}(M)$) the infimum (resp. supremum) of the eigenvalues of M .

We recall that for all compact sets W , the projection \mathcal{K}_W of a DPP kernel \mathcal{K} on $L^2(W, \mu)$ is a compact operator whose kernel can be written by Mercer's theorem as

$$K_W(x, y) = \sum_i \lambda_i^W \phi_i^W(x) \bar{\phi}_i^W(y)$$

where the λ_i^W are the eigenvalues of \mathcal{K}_W and the ϕ_i^W are the corresponding family of orthonormal eigenfunctions (see [61] for more details). When $\|\mathcal{K}\| < 1$, we define the operator $\mathcal{L} = \mathcal{K}(\mathcal{I} - \mathcal{K})^{-1}$ with kernel L and for all compact sets W we define the operator $\mathcal{L}_W = \mathcal{K}_W(\mathcal{I}_W - \mathcal{K}_W)^{-1}$ with kernel

$$L_W(x, y) := \sum_i \frac{\lambda_i^W}{1 - \lambda_i^W} \phi_i^W(x) \bar{\phi}_i^W(y). \quad (5.2.1)$$

Note that, unlike with \mathcal{K} , the kernel \mathcal{L}_W is different from the restriction of \mathcal{L} to $L^2(W, \mu)$. Nevertheless, if we consider a sequence W_n of increasing subsets of \mathbb{R}^d such that $\bigcup_n W_n = \mathbb{R}^d$ then $L_{W_n}(x, y)$ converges to $L(x, y)$ for all $x, y \in \mathbb{R}^d$. Therefore, we later show that we can consider $\mathcal{L}(x, y)$ as an approximation of $\mathcal{L}_W(x, y)$ when $x, y \in W$. Another difference between \mathcal{L}_W and \mathcal{L} is that, when X is a stationary (resp. isotropic) DPP, $L(x, y)$ only depends on $y - x$ (resp. $\|y - x\|$) but this is not true for L_W . Finally, we write \mathcal{I} for the identity operator on $L^2(\mathbb{R}^d, \mu)$ and \mathcal{I}_W for its restriction on $L^2(W, \mu)$ and we recall that the expression of the likelihood of DPPs results from the following property:

Theorem 5.2.1 ([75, 97]). *Let X be a DPP on a compact set W with eigenvalues in $[0, 1[$, then X is absolutely continuous with respect to the homogeneous Poisson point process with intensity 1, and has density*

$$f(x) = \exp(\mu(W)) \det(\mathcal{I}_W - \mathcal{K}_W) \det(L_W[x])$$

for all $x \in \cup_n(\mathbb{R}^d)^n$.

In the above expression, the first determinant corresponds to the Fredholm determinant of the operator $\mathcal{I}_W - \mathcal{K}_W$ which is equal to $\prod_i(1 - \lambda_i^W)$ while the second determinant is the matrix determinant of $(L_W(x_i, x_j))_{i,j}$. The (normalized) log-likelihood of $X \cap W_n$ is therefore:

$$l_n(\theta|X) = 1 + \frac{1}{\mu(W_n)} (\log \det(\mathcal{I}_{W_n} - \mathcal{K}_{W_n}^\theta) + \log \det(L_{W_n}^\theta[X \cap W_n])) \quad (5.2.2)$$

and a maximum likelihood estimate is

$$\hat{\theta}_n \in \left\{ \arg \max_{\theta \in \Theta} l_n(X|\theta) \right\}.$$

5.2.2 Approximation of the likelihood for stationary DPPs

As can be seen in the expression of (5.2.2), computing the log-likelihood of a family of DPPs requires knowing the spectral decomposition of $\mathcal{K}_{W_n}^\theta$ for all n, θ to get the expression of $L_{W_n}^\theta$ by (5.2.1). This is possible in the case of discrete DPPs whose kernels are finite matrices, but the spectral decomposition is usually not known for continuous DPPs. This motivates the idea in [69] to approximate stationary kernels by a truncated Fourier series on rectangular windows. For example, if $W = [-1/2, 1/2]^d$ then the approximation writes

$$K^\theta(x, y) \approx \sum_{\substack{k \in \mathbb{Z}^d \\ \|k\| < N}} c_k e^{i2\pi k \cdot (y-x)}, \quad \text{where } c_k = \int_W K_0^\theta(t) e^{-i2\pi k \cdot t} d\mu(t)$$

for some truncation constant N . Since the eigenvalues and eigenvectors of this kernel approximation are explicitly known, then the log-likelihood can be computed.

Our approximation is based on a different expression of (5.2.2) in terms of the convolution products of the function $(x, y) \mapsto \mathbb{1}_{W_n}(x) K^\theta(x, y) \mathbb{1}_{W_n}(y)$ through the following two identities (see [94] for example):

$$\log \det(\mathcal{I}_{W_n} - \mathcal{K}_{W_n}^\theta) = - \sum_{k=1}^{\infty} \frac{1}{k} \int_{W_n^k} K^\theta(x_1, x_2) \cdots K^\theta(x_{k-1}, x_k) K^\theta(x_k, x_1) d\mu^k(x) \quad (5.2.3)$$

and for all $x, y \in W_n$

$$L_{W_n}^\theta(x, y) = K^\theta(x, y) + \sum_{k=1}^{\infty} \int_{W_n^k} K^\theta(x, z_1) K^\theta(z_1, z_2) \cdots K^\theta(z_{k-1}, z_k) K^\theta(z_k, y) d\mu^k(z). \quad (5.2.4)$$

These convolution products are not known either in the general case, but for stationary DPPs satisfying $\|\mathcal{K}^\theta\| < 1$ then $\hat{L}_0^\theta = \hat{K}_0^\theta / (1 - \hat{K}_0^\theta)$ as a consequence of (5.2.4) and an asymptotic approximation (see Proposition 5.2.2) gives

$$L_{W_n}^\theta(x, y) \approx L_0^\theta(y - x) = \int_{\mathbb{R}^d} \frac{\hat{K}_0^\theta(t)}{1 - \hat{K}_0^\theta(t)} \exp(2i\pi t \cdot (y - x)) d\mu(t),$$

$$\frac{1}{\mu(W_n)} \log \det(\mathcal{I}_{W_n} - \mathcal{K}_{W_n}^\theta) \approx \int_{\mathbb{R}^d} \log(1 - \hat{K}_0^\theta(x)) d\mu(x).$$

This motivates the use of the following approximation for the log-likelihood:

$$\tilde{l}_n(\theta|X) := 1 + \int_{\mathbb{R}^d} \log(1 - \hat{K}_0^\theta(x)) d\mu(x) + \frac{1}{\mu(W_n)} \log \det(L_0^\theta[X \cap W_n]), \quad (5.2.5)$$

and this approximation can be computed from \hat{K}_0^θ which is known for all classical family of stationary DPPs built from covariance functions [69]. Moreover, we introduce in Section 5.2.4 a correction to $\tilde{l}_n(\theta|X)$ that fixes the edge effects introduced by the asymptotic approximation.

In some cases, an analytic expression for L_0^θ is also known making its computation easier. For example, Gaussian-type kernels of the form

$$K(x, y) = \rho \exp\left(-\frac{\|y - x\|^2}{\alpha^2}\right), \quad 0 < \rho \leq (\sqrt{\pi}\alpha)^{-d}$$

satisfy

$$L(x, y) = \sum_{n \geq 1} \rho^n \frac{(\sqrt{\pi}\alpha)^{d(n-1)}}{n^{d/2}} \exp\left(-\frac{\|y - x\|^2}{n\alpha^2}\right).$$

Moreover, the most repulsive (see [15]) Bessel-type kernels have the form

$$K(x, y) = \rho 2^{d/2} \Gamma(d/2 + 1) \frac{J_{d/2}(\sqrt{2d}\|y - x\|/\alpha)}{(\sqrt{2d}\|y - x\|/\alpha)^{d/2}}, \quad 0 < \rho \leq \frac{d^{d/2}}{(2\pi)^{d/2} \alpha^d \Gamma(d/2 + 1)},$$

where J denotes the Bessel function of the first kind, and satisfy

$$L(x, y) = \frac{\rho 2^{d/2} \Gamma(d/2 + 1)}{1 - \rho \frac{(2\pi)^{d/2} \alpha^d \Gamma(d/2 + 1)}{d^{d/2}}} \frac{J_{d/2}(\sqrt{2d}\|y - x\|/\alpha)}{(\sqrt{2d}\|y - x\|/\alpha)^{d/2}}.$$

Finally, when $d = 1$, Whittle-Matérn-type kernels of the form

$$K(x, y) = \rho \exp\left(-\frac{|y - x|}{\alpha}\right), \quad 0 < \rho \leq \frac{1}{2\alpha}$$

satisfy

$$L(x, y) = \frac{\rho}{\sqrt{1 - 2\rho\alpha}} \exp\left(-\frac{\sqrt{1 - 2\rho\alpha}}{\alpha} |y - x|\right).$$

All these expression are easily obtained from the fact that $\hat{L}_0 = \hat{K}_0 / (1 - \hat{K}_0) = \sum_{n \geq 1} (\hat{K}_0)^n$.

5.2.3 Consistency of the MLE

The deterministic part of the log-likelihood (5.2.2) has a straightforward asymptotic:

Proposition 5.2.2. *Let X be a stationary DPP on (\mathbb{R}^d, μ) with kernel K with eigenvalues in $[0, 1[$ and let W_n be an increasing sequence of compact subsets of \mathbb{R}^d such that $\bigcup_{n \geq 0} W_n = \mathbb{R}^d$, then*

$$\frac{1}{\mu(W_n)} \log \det(\mathcal{I}_{W_n} - \mathcal{K}_{W_n}) \xrightarrow{n \rightarrow +\infty} \int_{\mathbb{R}^d} \log(1 - \hat{K}_0(x)) d\mu(x).$$

The stochastic part of the log-likelihood (5.2.2) is a random kernel matrix whose behavior is more difficult to control. The main issue is that $\det(L_{W_n}^\theta[X])$ vanishes when two points of X gets arbitrary close to each other but no relationship between how close some points of X are from each other and the value of the determinant is known, making the likelihood difficult to control. To the author's knowledge, the only related result is that, in most cases, the lowest eigenvalue of $L_{W_n}^\theta[X]$ is non zero iff $\inf_{x,y \in X} \|y-x\| > 0$ (see [4]). This motivates the idea to only consider approximations of DPPs on $(\varepsilon\mathbb{Z})^d$, $\varepsilon > 0$, an arbitrary small regular grid of \mathbb{R}^d , with kernels $\varepsilon^d K[(\varepsilon\mathbb{Z})^d]$. These DPPs approximate continuous DPPs and since all points are at a distance at least ε from each other then we can bound the logarithm of both the highest and lowest eigenvalue of the stochastic part of the likelihood for all point configurations X of $(\varepsilon\mathbb{Z})^d$, a property that is essential for most of our results. This approximation on a small regular grid is the main limitation of our contribution. Other weaker assumptions on the DPP kernels that we consider are:

Condition \mathcal{H} : $\{K^\theta, \theta \in \Theta\}$ is a compact family of non-negative stationary kernels on (\mathbb{R}^d, μ) such that $\hat{K}_0^\theta(x)$ exists and is positive for all $x \in \mathbb{R}^d$ and $\theta \in \Theta$ and satisfy

$$\sup_{\theta \in \Theta} \|\hat{K}_0^\theta\|_\infty < 1.$$

Moreover, there exists constants $C, C', \alpha, \alpha' > 0$ such that for all $x \in \mathbb{R}^d$ and $\theta \in \Theta$

$$|K_0^\theta(x)| \leq \frac{C}{1 + \|x\|^{\alpha+d}}, \quad |\nabla_x K_0^\theta(x)| \leq \frac{C'}{1 + \|x\|^{\alpha'+d}}. \quad (5.2.6)$$

Finally, for all $x \in \mathbb{R}^d$, $\theta \mapsto K_0^\theta(x)$ and $\theta \mapsto L_0^\theta(x)$ are continuous on Θ .

Most of these assumptions are not too restrictive. The continuity assumptions for $\theta \mapsto K_0^\theta$ and $\theta \mapsto L_0^\theta$ and positiveness of \hat{K}_0^θ are satisfied by most translation-invariant kernel families defined from covariance function that are used in spatial statistics (see [15, 69] for more details) such as Gaussian, Laguerre-Gaussian, Whittle-Matérn, Cauchy or Bessel-type kernels, the exception being the most repulsive [15] Bessel-type kernel whose Fourier transform is an indicator function. The rate of decay (5.2.6) needed for K_0 and $\nabla_x K_0$ is a little bit more restrictive. All previously mentioned kernel families satisfy this condition to the exception of Bessel-type kernels whose associated covariance function is not integrable, excluding some of the most repulsive kernels. The assumption of non-negativity seems unnecessarily restrictive. It is only used in Lemma 5.A.5 to transfer properties of the continuous DPPs to their discrete approximation and not in the proof of consistency of the MLE itself. This is why we think it is not a necessary assumption. This point is still under investigation.

Similarly, some weak assumptions are also needed on the sequence of observation windows $(W_n)_{n \in \mathbb{N}}$:

Condition \mathcal{W} : W_n is an increasing sequence of compact subsets of \mathbb{R}^d such that $\bigcup_{n \geq 0} W_n = \mathbb{R}^d$ and there exists an increasing sequence $r_n \rightarrow +\infty$ such that

$$\mu((\partial W_n \oplus r_n) \cap W_n) = o(\mu(W_n)). \quad (5.2.7)$$

Moreover,

$$\sum_{n \geq 0} \exp(-\delta \mu(W_n)) < +\infty \quad (5.2.8)$$

for all $\delta > 0$.

Note that condition (5.2.7) is not really restrictive and is satisfied when the boundary of W_n is not too irregular. For example, if μ is the Lebesgue measure on \mathbb{R}^d , $(W_n)_{n \geq 0}$ is a sequence of increasing convex sets with inradius $R_n \rightarrow +\infty$, then condition (5.2.7) is always satisfied by taking $r_n = \sqrt{R_n}$. Similarly, condition (5.2.8) is also unrestrictive and will be satisfied, for example, if there exists $C, \alpha > 0$ such that $\mu(W_n) \geq Cn^\alpha$.

Under these assumptions, we first have the following approximation of the stochastic part of the log-likelihood.

Proposition 5.2.3. *Let $\{K^\theta, \theta \in \Theta\}$ be a family of DPP kernels satisfying condition (\mathcal{H}) , $(W_n)_{n \in \mathbb{N}}$ be an increasing sequence of subsets of \mathbb{R}^d satisfying condition (\mathcal{W}) and let X be the realization of a DPP on $(\varepsilon\mathbb{Z})^d$ with kernel $K_{\varepsilon, W_n}^{\theta^*} := \varepsilon^d K_{W_n}^{\theta^*}$. Then, for all $\theta \in \Theta$ and small enough $\varepsilon > 0$,*

$$\mathbb{E}_{\theta^*} \left[\frac{1}{\mu(W_n)} |\log \det(L_\varepsilon^\theta[X \cap W_n]) - \log \det(L_{\varepsilon, W_n}^\theta[X \cap W_n])| \right] \xrightarrow{n \rightarrow +\infty} 0$$

Furthermore, we also get that the MLE of any DDP restricted to a small enough regular grid of \mathbb{R}^d is consistent

Theorem 5.2.4. *Let $\{K^\theta, \theta \in \Theta\}$ be a family of DPP kernels satisfying condition (\mathcal{H}) and $(W_n)_{n \in \mathbb{N}}$ be an increasing sequence of subsets of \mathbb{R}^d satisfying condition (\mathcal{W}) . Let $\hat{\theta}_{\varepsilon, n}$ be the MLE associated to an observation of a DPP X_ε on $(\varepsilon\mathbb{Z})^d$ with kernel $\varepsilon^d K_{W_n}^{\theta^*}$ where $\theta^* \in \Theta$, then the function*

$$l_\varepsilon : \theta \mapsto \lim_{n \rightarrow \infty} l_n(\theta | X_\varepsilon \cap W_n)$$

is well-defined on Θ almost surely for any small enough ε . Assuming that this function admits a unique maximum then

$$\hat{\theta}_{\varepsilon, n} \xrightarrow{a.s.} \theta^*$$

for any small enough ε .

Note that the assumption that the function l_ε admits a unique maximum is likely superfluous but we were unable to prove it.

5.2.4 Correcting edge effects

The main issue of the asymptotic approximation of the likelihood is that it ignores edge effects. As shown in Section 5.3, this does not affect DPPs with low repulsion since their edge effects are minimal but has a more prominent impact in the inference of more repulsive DPPs. One way to correct this issue for isotropic DPPs (i.e. $K(x, y) = K_0(\|y - x\|)$) and rectangular windows (i.e. $W = [a_1, b_1] \times \cdots \times [a_d, b_d]$) is to consider the observation window as the flat torus $\mathbb{T}_W := \mathbb{R} \setminus (b_1 - a_1)\mathbb{Z} \times \cdots \times \mathbb{R} \setminus (b_d - a_d)\mathbb{Z}$. This way, points close to the border of the window are brought close to each other in order to compensate edge effects.

More precisely, we replace all instances of $L_0^\theta(\|y - x\|)$ in the determinant in the stochastic part of (5.2.5) by $L_0^\theta(\|y - x\|_{\mathbb{T}_W})$ where $\|\cdot\|_{\mathbb{T}_W}$ is the flat torus metric on \mathbb{T}_W . This is equivalent to replacing L_0^θ by a periodic version of itself on the observation window. The approximate likelihood then writes

$$\tilde{l}_n^\mathbb{T}(\theta | X) = 1 + \int_{\mathbb{R}^d} \log(1 - \hat{K}_0(x)) d\mu(x) + \frac{1}{\mu(W_n)} \log \det((L_0^\theta(\|y - x\|_{\mathbb{T}_W}))_{x, y \in W_n}) \quad (5.2.9)$$

Note that, since we consider a periodic version of L_0^θ on the observation window then it can be approximated by its Fourier series which was the idea first described in [69]. This is why both (5.2.9) and the approximate likelihood used in [69] are nearly equal (see figure 9) but approximating L_{W_n} directly by L_0 instead of a truncation of its Fourier series leads to a smoother likelihood and overall slightly better results, as well as a more computationally efficient method when an analytic expression of L^θ is known.

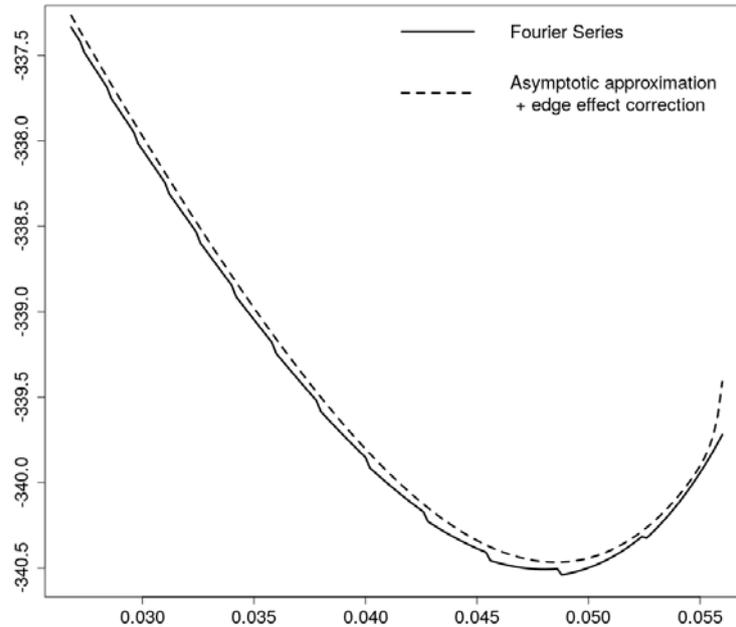


Figure 9: Comparison between two different approximations of $\alpha \mapsto l(\rho^*, \alpha|X)$ where X is simulated from a DPP with Gaussian-type kernel (5.3.1) with true parameters $\rho^* = 100$ and $\alpha^* = 0.05$ in $[0, 1]^2$. The solid line is obtained when L_{W_n} is approximated by its Fourier series with the truncation described in [69, Section 5.1] while the dashed line is obtained when L_{W_n} is approximated by L_0 with the correction described in Section 5.2.4.

5.2.5 Estimation of the intensity by MLE

In this section, we assume that we estimate parameters of homogeneous DPP kernels of the form

$$K^{\rho, \theta}(x, y) = \rho \widetilde{K}^\theta(x, y) \quad (5.2.10)$$

such that $\widetilde{K}^\theta(x, x) = 1$ for all x . The parameter ρ corresponds here to the intensity of the DPP. When jointly estimating (ρ, θ) by the approximate MLE, simulations usually show that the estimate of ρ appears to be very close to $N(W_n)/\mu(W_n)$. One explanation given in [69] is that, by doing a first order convolution approximation in (5.2.3) and (5.2.4), we get

$$l_n(X \cap W_n) \approx 1 - n\rho + \log(\rho) \frac{N(W_n)}{\mu(W_n)} + \log \det(\widetilde{K}_{W_n}^\theta[X \cap W_n])$$

and the maximum point of this approximation is $\hat{\rho} = N(X \cap W_n)/\mu(W_n)$. We even show in Proposition 5.A.6 that, in the case of Bessel type DPP kernels with parameters

(ρ, α) , $\hat{\rho} = N(X \cap W_n)/\mu(W_n)$ is always the maximum point of $\rho \mapsto \tilde{l}_n(X|\rho, \alpha)$ for any α . This result suggests that, instead of jointly estimating ρ and θ , it is more computationally efficient to directly estimate ρ by $\hat{\rho} = N(X \cap W_n)/\mu(W_n)$ and then θ by an argument of the maximum of $\theta \mapsto \tilde{l}_n(X|\hat{\rho}, \theta)$.

5.3 Simulation study

In this section we perform a simulation study to investigate the performance of our approximate MLE, with and without edge effect correction, and compare it to minimum contrast estimators based on Ripley's K function or on the pair correlation function.

5.3.1 Performance of the approximate MLE for Gaussian-type DPPs

We consider a Gaussian family of DPP kernels in \mathbb{R}^2

$$K^{\rho, \alpha}(x, y) = \rho \exp\left(-\frac{\|y - x\|^2}{\alpha^2}\right) \quad (5.3.1)$$

where $\pi\rho\alpha^2 \leq 1$. Here, the parameter ρ controls the number of points of the point process while α controls its repulsiveness. The bound on $\pi\rho\alpha^2$ is a consequence of the eigenvalues of \mathcal{K} being in $[0, 1]$, it can be interpreted as a trade-off between the repulsiveness of the DPP and how dense it is. It is easy to see that the exponential decay of $K_0^{\rho, \alpha}$ makes this family satisfy Condition (\mathcal{H}) when $\pi\rho\alpha^2$ is bounded by a constant strictly lower than 1. Moreover, since this family satisfy (5.2.10) then, as explained in Section 5.2.5, we will not jointly estimate (ρ, α) directly by MLE but we estimate ρ by $\hat{\rho} = N(X \cap W_n)/\mu(W_n)$ and α by the argument of the maximum of $\alpha \mapsto \tilde{l}_n(X|\hat{\rho}, \alpha)$.

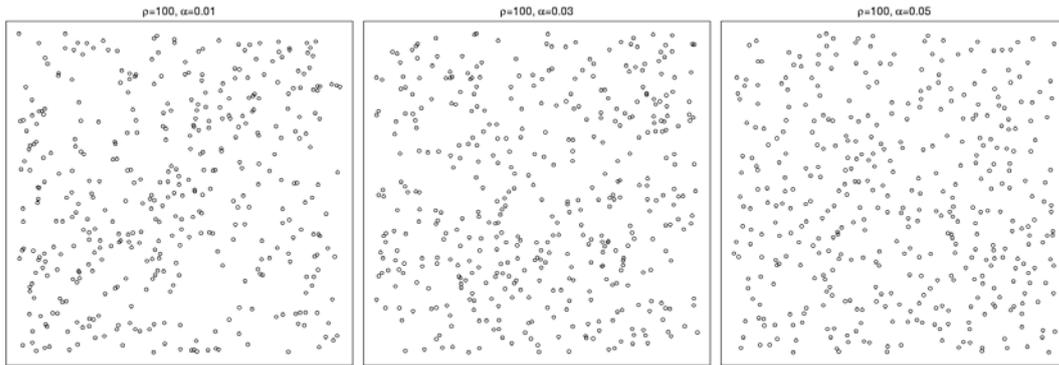


Figure 10: Examples of realization of a Gaussian-type DPP on $[0, 1]^2$ with parameters $\rho^* = 100$ and $\alpha^* \in \{0.01, 0.03, 0.05\}$

We consider realizations of a DPP with Gaussian kernel where $\rho^* = 100$, $\alpha^* \in \{0.01, 0.03, 0.05\}$ and the observation window is either $[0, 1]^2$, $[0, 2]^2$ or $[0, 3]^2$. When $\rho^* = 100$, α takes values in $]0, (10\sqrt{\pi})^{-1} \approx 0.056[$. Therefore, $\alpha^* = 0.01$ corresponds to a weakly repulsive Gaussian DPP, close to a Poisson point process, while $\alpha^* = 0.03$ corresponds to a mildly repulsive DPP and $\alpha^* = 0.05$ corresponds to a strongly repulsive DPP. Examples of such realizations are shown in Figure 10. We estimate α^* by the approximate MLE defined in (5.2.5) and compare it to its edge-corrected version defined in (5.2.9) as well as minimum contrast estimators (MCE) based on the pair correlation function (pcf) or Ripley's K function (see [16]), both being common

second-order moment estimators used in spatial statistics. All simulations of the DPP have been done in R [87] using the *spatstat* [7] package and both moment estimators were computed by the function *dppm* of the same package.

Boxplots of the difference between the estimators $\hat{\alpha}$ and the true value α^* for 500 simulations in the nine different cases are displayed in Figure 11 and corresponding mean square errors are displayed in Table 5.

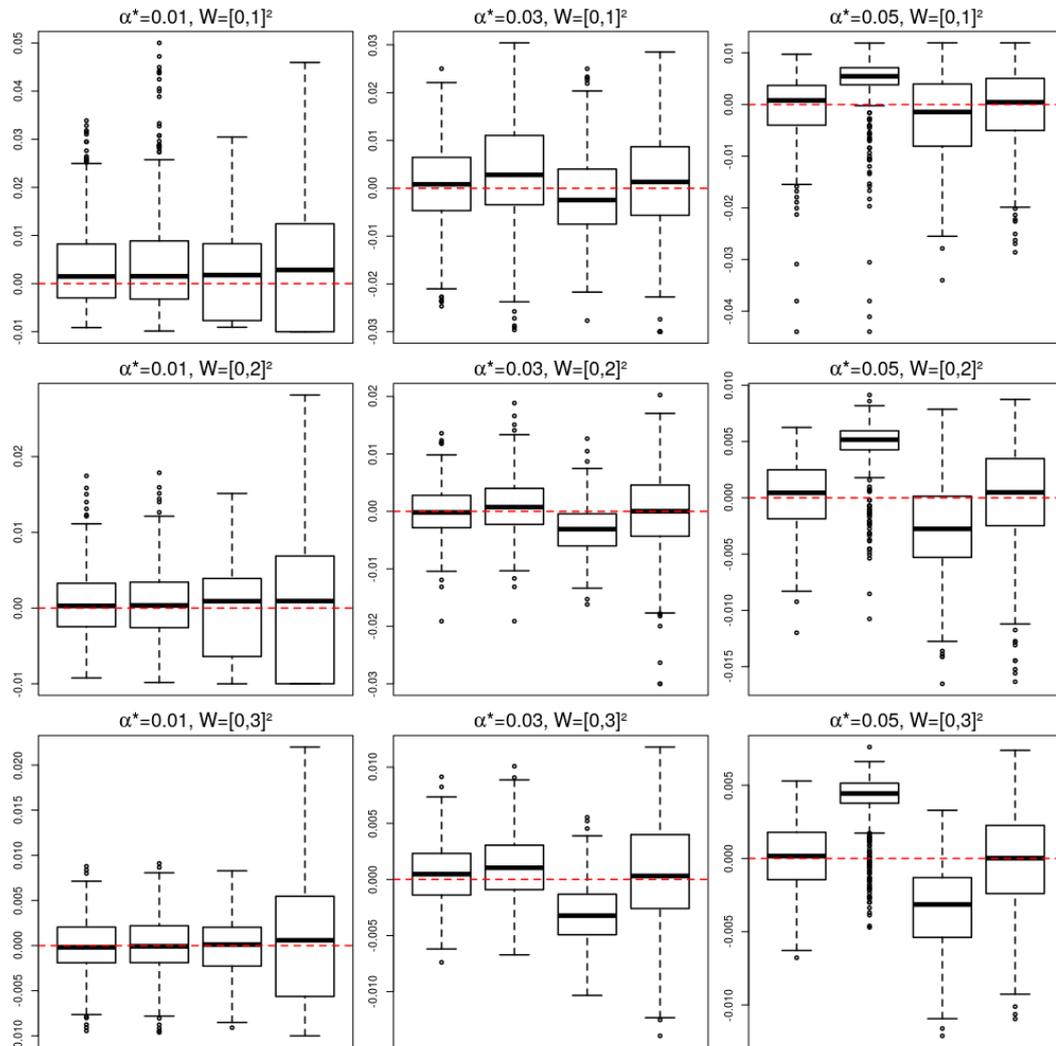


Figure 11: Boxplots of $\hat{\alpha} - \alpha^*$ generated from 500 simulations of Gaussian-type DPPs (5.3.1) on windows $W = [0, 1]^2$, $[0, 2]^2$ or $[0, 3]^2$ with true parameters $\rho^* = 100$ and $\alpha^* = 0.01, 0.03$ or 0.05 . The first two estimators are the approximate MLE from $\tilde{l}_n^{\mathbb{T}}(X|\hat{\rho}, \alpha)$ and $\tilde{l}_n(X|\hat{\rho}, \alpha)$, while the last two are MCE based on the pair correlation function and Ripley's K function.

Note that, when $\alpha^* = 0.01$ and $\alpha^* = 0.03$, inference based on the approximate likelihood $\tilde{l}_n(X|\hat{\rho}, \alpha)$ outperforms moment based inference for windows bigger than $[0, 2]^2$. These results are expected from maximum likelihood based inference and shows that hundreds of points are enough for $\tilde{l}_n(X|\hat{\rho}, \alpha)$ to be a good enough approximation of the true likelihood when the underlying DPP is not too repulsive. The issue lies within the case $\alpha^* = 0.05$ where the estimation is heavily biased due to strong edge effects making $\tilde{l}_n(X|\hat{\rho}, \alpha)$ not a good approximation of the true likelihood for low values of n . As can be seen in Figure 12, $\tilde{l}_n(X|\hat{\rho}, \alpha)$ will usually be decreasing with respect to

α	Window	MLE based on \tilde{l}^\top	MLE based on \tilde{l}	MCE (pcf)	MCE (K)
0.01	$[0, 1]^2$	0.83	1.25	0.86	1.81
	$[0, 2]^2$	0.21	0.24	0.31	0.74
	$[0, 3]^2$	0.090	0.095	0.17	0.48
0.03	$[0, 1]^2$	0.81	1.75	0.77	1.17
	$[0, 2]^2$	0.18	0.23	0.27	0.46
	$[0, 3]^2$	0.079	0.10	0.17	0.23
0.05	$[0, 1]^2$	0.41	0.54	0.74	0.51
	$[0, 2]^2$	0.088	0.28	0.23	0.21
	$[0, 3]^2$	0.051	0.20	0.19	0.12

Table 5: Estimated mean square errors ($\times 10^4$) of $\hat{\alpha}$ for Gaussian-type DPPs on different windows and with different values of α , each computed from 500 simulations. The first two columns corresponds to estimators based on the approximate MLEs (5.2.5) and (5.2.9). The last two columns corresponds to MCE estimators based on the pcf and Ripley's K function.

α when the underlying DPP is very repulsive and therefore it often estimates α by the highest value the parameter can take, which is $1/\sqrt{\pi\hat{\rho}}$, explaining its behaviour.

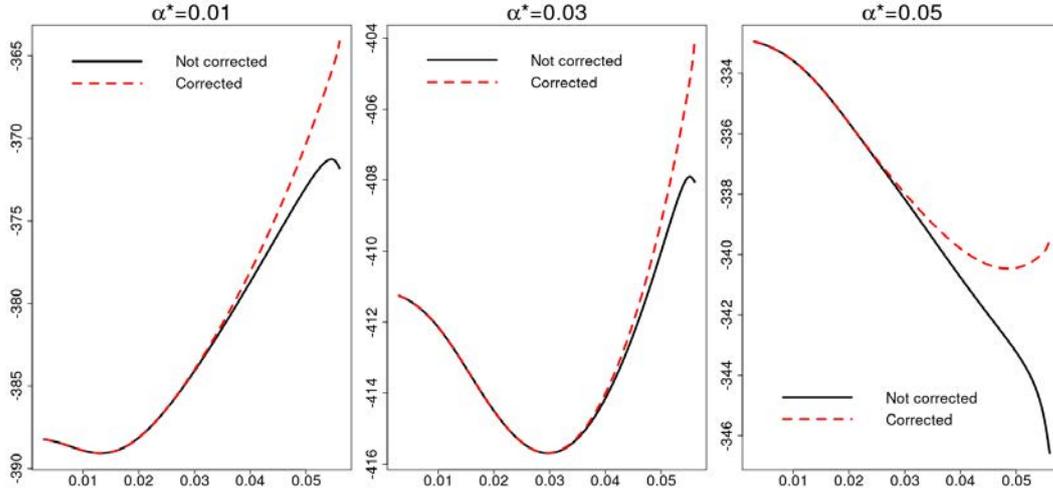


Figure 12: Comparison between $-\tilde{l}_n(X|100, \alpha)$ (solid lines) and $-\tilde{l}_n^\top(X|100, \alpha)$ (dashed lines) with respect to α where X has been simulated from a DPP on $[0, 1]^2$ with a Gaussian-type kernel (5.3.1) with true parameters $\rho^* = 100$ and, from left to right, $\alpha^* = 0.01, 0.03$ or 0.05 .

On the other hand, the correction introduced in (5.2.9) gives more accurate values of the likelihood for high values of α as can be seen in Figure 12. This explains why this estimator outperforms the others in nearly every cases and especially the most repulsive ones. The main issue of this correction is that it is limited to rectangular windows but these results suggest that, for a window with a different shape, using a similar idea of replacing the euclidean distance in the expression of $l_n(X|\theta)$ with a distance that brings points on the edge closer to each other should give similar results.

Finally, the main drawback with the MLE is its heavy computation time required due to the need to optimize a function defined as the log-determinant of an $n \times n$ matrix, where n is the number of points observed. For comparison, each MCE took less than 1 second on a regular computer to compute one estimator in each case considered in Figure 11 but each computation of the approximate MLE took about a second when $W = [0, 1]^2$, about 20 seconds when $W = [0, 2]^2$ and about 100 seconds when $W = [0, 3]^2$.

5.4 Proofs of Section 5.2

In this section, we give the proofs of the propositions stated in Section 5.2. We will use several times these two expressions of the expectation of a functional of X:

$$\mathbb{E}_{\theta^*}[f(X \cap W)] = \det(\text{Id} - \mathcal{K}_W^\theta) \sum_{k \geq 0} \frac{1}{k!} \int_{W^k} f(x) \det(L_W^{\theta^*}[x]) d\mu(x) \quad (5.4.1)$$

$$\mathbb{E}_{\theta^*} \left[\sum_{x_1, \dots, x_k \in X \cap W}^{\neq} f(x_1, \dots, x_k) \right] = \int_{W^k} f(x) \det(K^{\theta^*}[x]) d\mu(x) \quad (5.4.2)$$

for all W compact, $f : \bigcup_{n \geq 0} W^n \rightarrow \mathbb{R}$ symmetrical and integrable and where the symbol \neq in the second expression means that the sum is over p -tuples of distinct points of $X \cap W$. Here, (5.4.1) is the consequence of Theorem 5.2.1 while (5.4.2) is the definition of the k -th order intensity function of a determinantal point process. In the proofs of Propositions 5.2.3 and 5.2.4, as explained in Section 5.2, we work with DPPs on $((\varepsilon\mathbb{Z})^d, \mu)$, where μ is the counting measure on $\varepsilon\mathbb{Z}^d$ and with kernels $K_\varepsilon^\theta(x, y) := \varepsilon^d K^\theta(x, y)$. In this case, the associated integral operator \mathcal{K}_ε is simply the infinite matrix $K_\varepsilon[(\varepsilon\mathbb{Z})^d]$ and \mathcal{L}_ε can be simply defined as the infinite matrix $K_\varepsilon[(\varepsilon\mathbb{Z})^d](\text{Id} - K_\varepsilon[(\varepsilon\mathbb{Z})^d])^{-1}$ when $\|\mathcal{K}_\varepsilon\| < 1$. Finally, by Lemma 5.A.5, ε is chosen small enough such that the DPPs are well-defined and satisfy the following assumptions

Condition \mathcal{H}_ε : $\{K_\varepsilon^\theta, \theta \in \Theta\}$ is a compact family of stationary kernels on $((\varepsilon\mathbb{Z})^d, \mu)$, where μ is the counting measure on $\varepsilon\mathbb{Z}^d$, satisfying

$$\inf_{\theta \in \Theta} \lambda_{\min}(\mathcal{K}_\varepsilon^\theta) > 0 \quad \text{and} \quad \sup_{\theta \in \Theta} \lambda_{\max}(\mathcal{K}_\varepsilon^\theta) < 1.$$

Moreover, there exists constants $C, C', \alpha, \alpha' > 0$ such that for all $x \in (\varepsilon\mathbb{Z})^d$ and $\theta \in \Theta$

$$|K_{0,\varepsilon}^\theta(x)| \leq \frac{C}{1 + \|x\|^{\alpha+d}} \quad \text{and} \quad |L_{0,\varepsilon}^\theta(x)| \leq \frac{C'}{1 + \|x\|^{\alpha'+d}}. \quad (5.4.3)$$

In the proofs of Propositions 5.2.3 and 5.2.4, we mostly avoid using ε in indices as to avoid a clutter of notations. And finally, we will consider the following constants:

$$\lambda_m^\theta := \lambda_{\min}(\mathcal{K}_\varepsilon^\theta), \quad \lambda_M^\theta := (1 - \lambda_{\max}(\mathcal{K}_\varepsilon^\theta))^{-1} \quad \text{and} \quad A_\Theta := \sup_{\theta \in \Theta} \max(|\log(\lambda_m^\theta)|, |\log(\lambda_M^\theta)|).$$

In particular, $A_\Theta < +\infty$ under condition \mathcal{H}_ε .

5.4.1 Proof of Proposition 5.2.2

In this section, we give a proof of a stronger Proposition than 5.2.2 that will be needed later for the proof of the consistency of $\hat{\theta}$:

Proposition 5.4.1. *Let X be a DPP on (\mathbb{R}^d, μ) whose kernel K^θ belong to a family satisfying Condition (\mathcal{H}) and let $(W_n)_{n \geq 0}$ be an increasing sequence of compact subsets of \mathbb{R}^d such that $\bigcup_{n \geq 0} W_n = \mathbb{R}^d$, then*

$$\frac{1}{\mu(W_n)} \log \det(\mathcal{I}_{W_n} - \mathcal{K}_{W_n}^\theta) \xrightarrow{n \rightarrow +\infty} \int_{\mathbb{R}^d} \log(1 - \hat{K}_0^\theta(x)) d\mu(x)$$

uniformly in θ and the limit is continuous with respect to θ .

Proof. Let $\theta \in \Theta$, since all eigenvalues of $\mathcal{K}_{W_n}^\theta$ are in $[0, 1[$ then we can expand the logarithm of the Fredholm determinant into

$$\log \det(\mathcal{I}_{W_n} - \mathcal{K}_{W_n}^\theta) = - \sum_{k \geq 1} \frac{\text{Tr}((\mathcal{K}_{W_n}^\theta)^k)}{k} = - \sum_{k \geq 1} \frac{1}{k} \int_{W_n^k} K^\theta(x_1, x_2) \cdots K^\theta(x_k, x_1) d\mu^k(x).$$

Now, we know that for any $x_1 \in \mathbb{R}^d$ the function

$$(x_2, \dots, x_k) \mapsto K^\theta(x_1, x_2) \cdots K^\theta(x_k, x_1)$$

is integrable and its integral is equal to $(K_0^\theta)^{*k}(0)$ where $(K_0^\theta)^{*k}$ is the k -th time self-convolution of K_0^θ . Moreover, since these functions are all dominated by (5.4.3) then

$$\begin{aligned} & \sup_{\theta \in \Theta} \left| \frac{1}{\mu(W_n)} \int_{W_n^k} K^\theta(x_1, x_2) \cdots K^\theta(x_k, x_1) d\mu^k(x) - (K_0^\theta)^{*k}(0) \right| \\ & \leq \frac{1}{\mu(W_n)} \int_{W_n \times (\mathbb{R}^d)^{k-1} \setminus (W_n)^{k-1}} \frac{C^k}{(1 + \|x_2 - x_1\|^{d+\alpha}) \cdots (1 + \|x_1 - x_k\|^{d+\alpha})} d\mu(x_1) \cdots d\mu(x_k) \end{aligned}$$

which vanishes when $n \rightarrow +\infty$ by Lemma 5.A.4. Finally, since

$$\frac{1}{\mu(W_n)} |\text{Tr}((\mathcal{K}_{W_n}^\theta)^k)| \leq \frac{1}{\mu(W_n)} \|\mathcal{K}^\theta\|^{k-1} \text{Tr}(\mathcal{K}_{W_n}^\theta) \leq C (\sup_{\theta \in \Theta} \|\mathcal{K}^\theta\|)^{k-1}$$

which is summable with respect to k and doesn't depend on n and θ , therefore

$$\frac{1}{\mu(W_n)} \log \det(\mathcal{I}_{W_n} - \mathcal{K}_{W_n}^\theta) \xrightarrow{n \rightarrow +\infty} - \sum_{k \geq 1} \frac{(K_0^\theta)^{*k}(0)}{k} = \int_{\mathbb{R}^d} \log(1 - \hat{K}_0^\theta(x)) d\mu(x)$$

uniformly in θ by the dominated convergence theorem. Moreover, since $\theta \mapsto K_0^\theta(x)$ is continuous for all $x \in \mathbb{R}^d$ and $\text{Tr}((\mathcal{K}_{W_n}^\theta)^k)/k$ is bounded by $\mu(W_n) C (\sup_{\Theta} \|\mathcal{K}^\theta\|)^{k-1}/k$ which is summable, then the dominated convergence theorem also gives us the continuity of $\theta \mapsto \log \det(\mathcal{I}_{W_n} - \mathcal{K}_{W_n}^\theta)$ for all $n \in \mathbb{N}$ hence $\theta \mapsto \int_{\mathbb{R}^d} \log(1 - \hat{K}_0^\theta(x)) d\mu(x)$ is continuous as a uniform limit of continuous functions. \square

5.4.2 Proof of Proposition 5.2.3

By Lemma 5.A.5, we consider a DPP X on $(\varepsilon\mathbb{Z})^d$ with kernel K_ε from a parametric family satisfying Condition $(\mathcal{H}_\varepsilon)$ and a sequence of observation windows $(W_n)_{n \geq 0}$ satisfying Condition (\mathcal{W}) . In order to prove Proposition 5.2.3, we start by showing that the stochastic part of the likelihood gets close to its expectation when $n \rightarrow +\infty$ with the help of the concentration inequality [81] from Pemantle and Peres. One of the results we will need is the following bound on the eigenvalues of the random matrices involved in the likelihood:

Lemma 5.4.2. *Let X be a DPP on $(\varepsilon\mathbb{Z})^d$ with kernel K and $W \subset (\varepsilon\mathbb{Z})^d$, then*

$$\lambda_{\min}(L[X \cap W]) \geq \lambda_{\min}(\mathcal{K}), \quad \lambda_{\min}(L_W[X \cap W]) \geq \lambda_{\min}(\mathcal{K})$$

and

$$\lambda_{\max}(L[X \cap W]) \leq (1 - \lambda_{\max}(\mathcal{K}))^{-1}, \quad \lambda_{\max}(L_W[X \cap W]) \leq (1 - \lambda_{\max}(\mathcal{K}))^{-1}$$

Proof. Since $L[X \cap W]$ is a sub-matrix of $L[(\varepsilon\mathbb{Z})^d] := \mathcal{L} = \mathcal{K}(Id - \mathcal{K})^{-1}$ then by [60, Theorem 4.3.28], we know that $\lambda_{\max}(L[X \cap W]) \leq \lambda_{\max}(\mathcal{L}) \leq (1 - \lambda_{\max}(\mathcal{K}))^{-1}$ and $\lambda_{\min}(L[X \cap W]) \geq \lambda_{\min}(\mathcal{L}) \geq \lambda_{\min}(\mathcal{K})$. We get the second set of inequalities with the same reasoning applied to L_W and by using the fact that $\lambda_{\min}(\mathcal{K}_W) \geq \lambda_{\min}(\mathcal{K})$ and $\lambda_{\max}(\mathcal{K}_W) \leq \lambda_{\max}(\mathcal{K})$ since \mathcal{K}_W is a sub-matrix of \mathcal{K} . \square

Lemma 5.4.3. *Under the same assumptions as Proposition 5.2.3, for all $\theta \in \Theta$,*

$$\frac{1}{\mu(W_n)} (\log \det(L_{W_n}^\theta[X \cap W_n]) - \mathbb{E}_{\theta^*}[\log \det(L_{W_n}^\theta[X \cap W_n])]) \xrightarrow{a.s.} 0$$

Proof. Let $X \subset W_n$ and $a \in W_n \setminus X$, then for all $\theta \in \Theta$,

$$\log \det(L_{W_n}^\theta[X \cup \{a\} \cap W_n]) - \log \det(L_{W_n}^\theta[X \cap W_n]) = \log \left(\frac{\det(L_{W_n}^\theta[X \cup \{a\} \cap W_n])}{\det(L_{W_n}^\theta[X \cap W_n])} \right).$$

By Lemma 5.4.2 and [60, Theorem 4.3.28], we get

$$\begin{aligned} \lambda_m^\theta &\leq \lambda_{\min}(L_{W_n}^\theta[X \cup \{a\} \cap W_n]) \leq \frac{\det(L_{W_n}^\theta[X \cup \{a\} \cap W_n])}{\det(L_{W_n}^\theta[X \cap W_n])} \\ &\leq \lambda_{\max}(L_{W_n}^\theta[X \cup \{a\} \cap W_n]) \leq \lambda_M^\theta. \end{aligned}$$

This shows that $\mu(W_n)^{-1} \log \det(L_{W_n}^\theta[\cdot])$ is a Lipschitz function on the finite point configurations in W_n with constant $A_\Theta/\mu(W_n)$. By Theorem 3.5 of [81], we get for all $a \in \mathbb{R}_+$

$$\begin{aligned} \mathbb{P}_{\theta^*} \left(\frac{1}{\mu(W_n)} |\log \det(L_{W_n}^\theta[X \cap W_n]) - \mathbb{E}_{\theta^*}[\log \det(L_{W_n}^\theta[X \cap W_n])]| > a \right) \\ \leq 5 \exp \left(- \frac{a^2 \mu(W_n)^2 / A_\Theta^2}{16(a\mu(W_n)/A_\Theta + 2\mathbb{E}_{\theta^*}[N(W_n)])} \right). \end{aligned} \quad (5.4.4)$$

Since $\mathbb{E}_{\theta^*}[N(W_n)] = K_0^{\theta^*}(0)\mu(W_n)$ then the right term in (5.4.4) is $O(\exp(-a^2\mu(W_n)))$ and vanishes when n goes to infinity uniformly in θ giving us the convergence in probability. Additionally, since W_n satisfies (5.2.8) then by the Borel-Cantelli Lemma the convergence holds almost surely. \square

Now, it only remains to show that

$$\frac{1}{\mu(W_n)} (\log \det(L^\theta[X \cap W_n]) - \log \det(L_{W_n}^\theta[X \cap W_n])) \quad (5.4.5)$$

converges in mean towards 0 to conclude the proof of Proposition 5.2.3. This is done by showing the existence of an operator $\mathcal{N}_{W_n}^\theta$ whose trace is $o(\mu(W_n))$ and such that

$$0 \leq \mathcal{L}^\theta|_{W_n} - \mathcal{L}_{W_n}^\theta \leq \mathcal{N}_{W_n}^\theta,$$

where $\mathcal{L}^\theta|_{W_n}$ is the projection of \mathcal{L}^θ on $L^2(W_n)$.

Lemma 5.4.4. *Let X be a DPP on $((\varepsilon\mathbb{Z})^d, \mu)$ with kernel K such that $\|\mathcal{K}\| < 1$ and let W be a compact of \mathbb{R}^d . Define the operator $\mathcal{M} := (\mathcal{L} + Id)^{-1} = Id - \mathcal{K}$ and the operator \mathcal{N}_W with kernel*

$$N_W(x, y) = \int_{W^C} L(x, z)L(z, y)d\mu(z) \quad \forall x, y \in W.$$

Finally, if we write $\mathcal{L}|_W$ the projection of \mathcal{L} on $L^2(W)$ then

$$0 \leq \mathcal{L}|_W - \mathcal{L}_W \leq \mathcal{N}_W.$$

Proof. We denote by \mathcal{P} the projection on $L^2(W)$, \mathcal{P}^\perp the projection on $L^2(W^C)$, Id the identity operator on $L^2(\mathbb{R}^d)$, Id_W the identity operator on $L^2(W)$ and Id_{W^C} the identity operator on $L^2(W^C)$. Since \mathcal{K}_W is by definition equal to the projection of \mathcal{K} on $L^2(W)$, with these notations we can write

$$\mathcal{M} = \begin{pmatrix} Id_W - \mathcal{K}_W & -\mathcal{P}\mathcal{K}\mathcal{P}^\perp \\ -\mathcal{P}^\perp\mathcal{K}\mathcal{P} & Id_{W^C} - \mathcal{K}_{W^C} \end{pmatrix} = \begin{pmatrix} (\mathcal{L}_W + Id_W)^{-1} & -\mathcal{P}\mathcal{K}\mathcal{P}^\perp \\ -\mathcal{P}^\perp\mathcal{K}\mathcal{P} & (\mathcal{L}_{W^C} + Id_{W^C})^{-1} \end{pmatrix}$$

and

$$\mathcal{M}^{-1} = \begin{pmatrix} \mathcal{L}|_W + Id_W & \mathcal{P}\mathcal{L}\mathcal{P}^\perp \\ \mathcal{P}^\perp\mathcal{L}\mathcal{P} & \mathcal{L}|_{W^C} + Id_{W^C} \end{pmatrix}.$$

Then, using the Schur complement we get

$$0 \leq \mathcal{L}|_W - \mathcal{L}_W = \mathcal{P}\mathcal{L}\mathcal{P}^\perp(\mathcal{L}|_{W^C} + Id_{W^C})^{-1}\mathcal{P}^\perp\mathcal{L}\mathcal{P}.$$

Finally, since \mathcal{N}_W is defined as $(\mathcal{P}\mathcal{L}\mathcal{P}^\perp)(\mathcal{P}^\perp\mathcal{L}\mathcal{P})$ and $(\mathcal{L}|_{W^C} + Id_{W^C})^{-1} \leq Id_{W^C}$ this concludes the first part of the lemma. \square

Now, we rewrite (5.4.5) as

$$\mathbb{E}_{\theta^*} \left[\frac{1}{\mu(W_n)} |\log \det (Id + (L^\theta[X \cap W_n] - L_{W_n}^\theta[X \cap W_n])L_{W_n}^\theta[X \cap W_n]^{-1})| \right].$$

By Lemma 5.4.4 we know that

$$0 \leq L^\theta[X \cap W_n] - L_{W_n}^\theta[X \cap W_n] \leq N_{W_n}^\theta[X \cap W_n]$$

where N^θ is defined as in Lemma 5.4.4. Therefore, using Lemma 5.A.3 we obtain the expression

$$0 \leq \log \det(L^\theta[X \cap W_n]) - \log \det(L_{W_n}^\theta[X \cap W_n]) \leq \text{Tr}(N_{W_n}^\theta[X \cap W_n]L_{W_n}^\theta[X \cap W_n]^{-1}).$$

But, since $\lambda_{\min}(L_{W_n}^\theta[X \cap W_n]) \geq \lambda_m^\theta$ by Lemma 5.4.2 then

$$\begin{aligned} \mathbb{E}_{\theta^*} [\text{Tr}(N_{W_n}^\theta[X \cap W_n]L_{W_n}^\theta[X \cap W_n]^{-1})] &\leq (\lambda_m^\theta)^{-1} \mathbb{E}_{\theta^*} [\text{Tr}(N_{W_n}^\theta[X \cap W_n])] \\ &= (\lambda_m^\theta)^{-1} \mathbb{E}_{\theta^*} \left[\sum_{x \in X \cap W_n} N_{W_n}^\theta(x, x) \right] \\ &= (\lambda_m^\theta)^{-1} \int_{W_n} N_{W_n}^\theta(x, x) K_0^{\theta^*}(0) d\mu(x) \\ &= (\lambda_m^\theta)^{-1} \|K^{\theta^*}\|_\infty \text{Tr}(\mathcal{N}_{W_n}^\theta). \end{aligned}$$

Finally, we write

$$\frac{1}{\mu(W_n)} \text{Tr}(\mathcal{N}_{W_n}^\theta) = \frac{1}{\mu(W_n)} \int_{W_n \times W_n^c} L_0^\theta(y-x)^2 d\mu(x) d\mu(y).$$

But, by Condition $(\mathcal{H}_\varepsilon)$, L_0^θ is a square integrable function on $(\varepsilon\mathbb{Z})^d$ for all θ, ε . Therefore, by Lemma 5.A.4,

$$\frac{1}{\mu(W_n)} \left(\int_{W_n \times \mathbb{R}^d} L_0^\theta(y-x)^2 d\mu(x) d\mu(y) - \int_{W_n \times W_n} L_0^\theta(y-x)^2 d\mu(x) d\mu(y) \right) \xrightarrow{n \rightarrow +\infty} 0$$

hence $\mu(W_n)^{-1} \text{Tr}(\mathcal{N}_{W_n}^\theta)$ vanishes when $n \rightarrow +\infty$ concluding the proof.

5.4.3 Proof of Theorem 5.2.4

In order to prove Theorem 5.2.4, we first show that the likelihood converges to a continuous limit $\theta \mapsto l(\theta)$ admitting a unique maximum in θ^* . We then conclude by showing that the convergence is uniform in θ . Once again, X is considered to be a DPP on $(\varepsilon\mathbb{Z})^d$ from a parametric family satisfying Condition $(\mathcal{H}_\varepsilon)$ and the the sequence of observation windows $(W_n)_{n \geq 0}$ satisfy Condition (\mathcal{W}) .

Lemma 5.4.5. *Under the same assumptions as Theorem 5.2.4, $\mu(W_n)^{-1} \mathbb{E}_{\theta^*}[l_n(X|\theta)]$ converges to a limit $l(\theta)$ for all $\theta \in \Theta$. Moreover, the function l is continuous on Θ .*

Proof. We saw in section 5.4.1 that the deterministic part of $l_n(X|\theta)$ converges uniformly to a continuous function. Moreover, we showed in Section 5.4.2 that

$$\lim_{n \rightarrow +\infty} \frac{1}{\mu(W_n)} \left(\mathbb{E}_{\theta^*}[\log \det(L_{W_n}^\theta[X \cap W_n])] - \mathbb{E}_{\theta^*}[\log \det(L^\theta[X \cap W_n])] \right) = 0$$

so we only need to show the convergence of $\mu(W_n)^{-1} \mathbb{E}_{\theta^*}[\log \det(L^\theta[X \cap W_n])]$ to a continuous function to prove the theorem. Using the power series expression of the matrix logarithm of $Id - L^\theta[X \cap W_n]/e^{A_\Theta}$ (which has all its eigenvalues in $[0, 1]$), we get

$$\begin{aligned} \log \det(L^\theta[X \cap W_n]) &= \text{Tr}(\log(L^\theta[X \cap W_n])) \\ &= N(W_n)A_\Theta + \text{Tr}(\log(Id - (Id - L^\theta[X \cap W_n])/e^{A_\Theta})) \\ &= N(W_n)A_\Theta - \sum_{k \geq 1} \frac{1}{k} \text{Tr}((Id - L^\theta[X \cap W_n])/e^{A_\Theta})^k. \end{aligned}$$

We know that $\mu(W_n)^{-1} \mathbb{E}_{\theta^*}[N(W_n)A_\Theta] = K_0^{\theta^*}(0)A_\Theta$. Moreover, by Lemma 5.4.2 we get the bound

$$\begin{aligned} &\frac{1}{k\mu(W_n)} \left| \mathbb{E}_{\theta^*} \left[\text{Tr}((Id - L^\theta[X \cap W_n])/e^{A_\Theta})^k \right] \right| & (5.4.6) \\ &\leq \frac{\mathbb{E}_{\theta^*}[N(W_n)]}{k\mu(W_n)} \lambda_{\max}(Id - L^\theta[X \cap W_n]/e^{A_\Theta})^k \\ &\leq \frac{K_0^{\theta^*}(0)}{k} \left(1 - \frac{\lambda_m^\theta}{e^{A_\Theta}} \right)^k \\ &\leq \frac{K_0^{\theta^*}(0)}{k} \left(1 - e^{-2A_\Theta} \right)^k. & (5.4.7) \end{aligned}$$

Since (5.4.7) does not depend on n nor θ then by dominated convergence we only need to show that $\mu(W_n)^{-1} \mathbb{E}_{\theta^*}[\text{Tr}((Id - L^\theta[X \cap W_n])/e^{A_\Theta})^k]$, and thus $\mu(W_n)^{-1} \mathbb{E}_{\theta^*}[\text{Tr}(L^\theta[X \cap$

$W_n]^k]$), converges for all k to conclude. Moreover, the dominated convergence theorem also tells us that after proving the convergence, we will only need to show that $\lim_{n \rightarrow +\infty} \mu(W_n)^{-1} \mathbb{E}_{\theta^*} [\text{Tr}(L^\theta[X \cap W_n]^k)]$ is continuous for all k to prove the continuity of the limit function.

We can develop the trace the following way:

$$\text{Tr}(L^\theta[X \cap W_n]^k) = \sum_{x_1, \dots, x_k \in (X \cap W_n)^k} f^\theta(x_1, \dots, x_k)$$

where we define the function $f^\theta : (\mathbb{R}^d)^k \rightarrow \mathbb{R}$ by

$$f^\theta : x \rightarrow L_0^\theta(x_2 - x_1) L_0^\theta(x_3 - x_2) \cdots L_0^\theta(x_n - x_{n-1}) L_0^\theta(x_1 - x_k).$$

Since we only know an expression of the expectation when the sum is over distinct points of $X \cap W_n$ we need to split the sum into all the different cases possible. Let \mathcal{V}_k be the set of partitions of $\{1, \dots, k\}$. For all $V = \{V_1, \dots, V_l\} \in \mathcal{V}_k$, we write v_i^j for the elements of V_i where $1 \leq j \leq |V_i|$. We also define the function $f_V^\theta : (\mathbb{R}^d)^l \rightarrow \mathbb{R}$ where we made identical all variables in $f^\theta(x)$ whose indices are in a same subset of V (an example is given in 5.4.6). We also define the function h_V the same way, where h is the function dominating all L_0^θ in Condition $(\mathcal{H}_\varepsilon)$. Finally, we write $\delta_V(x)$ for the function equal to 1 if, for all $1 \leq i, j \leq k$, $x_i = x_j$ when i and j are in the same subset of V and $x_i \neq x_j$ when i and j are in different subsets, and 0 otherwise. With these notations, we can use (5.4.2) to express the expectation of $\text{Tr}(L^\theta[X \cap W_n]^k)$ as

$$\begin{aligned} \mathbb{E}_{\theta^*} [\text{Tr}(L^\theta[X \cap W_n]^k)] &= \mathbb{E}_{\theta^*} \left[\sum_{x_1, \dots, x_k \in (X \cap W_n)^k} \sum_{V \in \mathcal{V}_k} f^\theta(x) \delta_V(x) \right] \\ &= \sum_{V \in \mathcal{V}_k} \int_{W_n^{|V|}} f_V^\theta(y) \det(K^{\theta^*}[y]) d\mu(y). \end{aligned}$$

We want to show that all f_V^θ are integrable with respect to their $|V| - 1$ last coordinates in order to apply Lemma 5.A.4. We say that two sets V_i and V_j of the partition V are connected if there exists $a \in V_i$ and $b \in V_j$ such that $|a - b| = 1 \pmod{k}$. This way, we define a graph $G = (V, E)$ where $E = \{(V_i, V_j), V_i \text{ and } V_j \text{ are connected}\}$ (an example is given in 5.4.6). It is easy to show that this graph is connected. Thus, we can write

$$|f_V^\theta(y)| \leq \|L^\theta\|_\infty^{k-|E|} \prod_{(V_a, V_b) \in E} |L_0^\theta(y_b - y_a)|.$$

For all $i \in \{1, \dots, |V|\}$, we define $G[i] = (V[i], E[i])$ as the graph induced by cutting the vertex V_i . Since we assumed that $L^\theta \in L^1((\varepsilon\mathbb{Z})^d, \mu)$ and we know that V_i is not an isolated vertex, we get

$$\begin{aligned} \int_{\mathbb{R}^d} |f_V^\theta(y)| d\mu(y_i) &\leq \|L_0^\theta\|_\infty^{k-|E|} \prod_{(V_a, V_b) \in E[i]} |L_0^\theta(y_b - y_a)| \int_{\mathbb{R}^d} \prod_{\{j: (V_i, V_j) \in E\}} |L_0^\theta(y_j - y_i)| d\mu(y_i) \\ &\leq \|L^\theta\|_1 \|L^\theta\|_\infty^{k-|E[i]|-1} \prod_{(V_a, V_b) \in E[i]} |L_0^\theta(y_b - y_a)|. \end{aligned}$$

So, we can keep integrating with respect to other variables y_j as long as V_j is not an isolated vertex of $G[i]$. Thus, proving that f_V^θ is integrable with respect to its $|V| - 1$ last coordinates is equivalent to showing that there exists a sequence of removal of all but one vertices from G that doesn't make the graph unconnected at any step. This

can be done by only removing non articulation points. Since any connected graph with more than 2 points has at least 2 non articulation points this can always be done. By induction, this shows that

$$\int_{\mathbb{R}^d} |f_V^\theta(y)| dy_2 \cdots dy_l \leq \|L^\theta\|_1^{|V|-1} \|L^\theta\|_\infty^{k-(|V|-1)} \quad (5.4.8)$$

and thus all f_V^θ are integrable with respect to their $|V| - 1$ last coordinates. Moreover, since f_V^θ and K^{θ^*} are translation invariant, then

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{1}{\mu(W_n)} \int_{W_n \times (\mathbb{R}^d)^{l-1}} f_V^\theta(y) \det(K^{\theta^*}[y]) d\mu(y) \\ = \int_{(\mathbb{R}^d)^{l-1}} f_V^\theta(0, z_2, \dots, z_l) \det(K^{\theta^*}[0, z_2, \dots, z_l]) d\mu(z_2) \cdots d\mu(z_l) \end{aligned}$$

by Lemma 5.A.4 which shows the convergence. Finally, the function

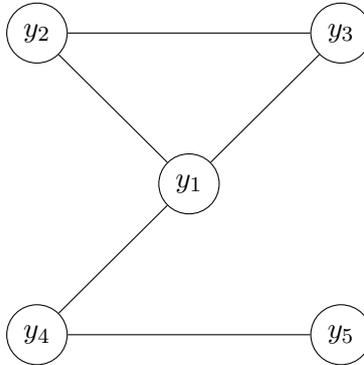
$$\theta \mapsto f_V^\theta(0, z_2, \dots, z_l) \det(K^{\theta^*}[0, z_2, \dots, z_l])$$

is continuous for all z_i and is bounded by $\|K^{\theta^*}\|_\infty^l h_V(0, z_2, \dots, z_l)$ which is integrable and doesn't depends on θ , proving that the limit is continuous.

Example 5.4.6. If $n = 8$ and $V = \{\{1, 4\}, \{2\}, \{3\}, \{5, 7, 8\}, \{6\}\}$ then f_V is written

$$\begin{aligned} f_V^\theta(y_1, y_2, y_3, y_4, y_5) \\ = f^\theta(y_2 - y_1) f^\theta(y_3 - y_2) f^\theta(y_1 - y_3) f^\theta(y_4 - y_1) f^\theta(y_5 - y_4) f^\theta(y_4 - y_5) f^\theta(0) f^\theta(y_1 - y_4) \end{aligned}$$

and the associated graph is



□

This result, combined with Lemma 5.4.4, shows that the function

$$l : \theta \mapsto \lim_{n \rightarrow \infty} l_n(\theta | X_\varepsilon \cap W_n)$$

is well-defined almost surely for all $\theta \in \Theta$ and continuous. We assumed that this function has a unique maximum and since we know that

$$\mathbb{E}_{\theta^*} [l_n(\theta^* | X) - l_n(\theta | X)] = \frac{1}{\mu(W_n)} D_{KL}(\mathbb{P}_{W_n}^{\theta^*} \| \mathbb{P}_{W_n}^\theta) \geq 0$$

where D_{KL} is the Kulback-Leibler divergence and $\mathbb{P}_{W_n}^\theta$ is the probability measure associated with the DPP with kernel $K_{W_n}^\theta$, taking the limit yields that

$$l(\theta^*) \geq l(\theta)$$

for all $\theta \in \Theta$. Therefore, θ^* is the unique maximum of l .

Finally, in order to conclude, we need to show that the almost sure convergence of the likelihood towards l is uniform in θ . We showed that l is continuous and we also showed in Section 5.4.1 that the deterministic part of the likelihood uniformly converges with respect to θ . Thus, it remains to prove the uniform convergence of the stochastic part of the likelihood. Consider the function

$$\phi_{\theta_0}(\eta) = \limsup_{n \rightarrow +\infty} \sup_{|\theta - \theta_0| \leq \eta} \frac{1}{\mu(W_n)} |\log \det(L_{W_n}^\theta[X \cap W_n]) - \log \det(L_{W_n}^{\theta_0}[X \cap W_n])|.$$

Let Θ_d be a countable dense subset of Θ and $\theta_1 \in \Theta_d$ such that $|\theta_1 - \theta_0| \leq \eta$. Then,

$$\begin{aligned} \phi_{\theta_0}(\eta) &\leq 2 \limsup_{n \rightarrow +\infty} \sup_{|\theta - \theta_1| \leq 2\eta} \frac{1}{\mu(W_n)} |\log \det(L_{W_n}^\theta[X \cap W_n]) - \log \det(L_{W_n}^{\theta_1}[X \cap W_n])| \\ &\leq 2 \limsup_{n \rightarrow +\infty} \sup_{|\theta - \theta_1| \leq 2\eta} \frac{1}{\mu(W_n)} |\log \det(L^\theta[X \cap W_n]) - \log \det(L^{\theta_1}[X \cap W_n])| \\ &\quad + 2 \limsup_{n \rightarrow +\infty} \sup_{\theta \in \Theta} \frac{1}{\mu(W_n)} |\log \det(L^\theta[X \cap W_n]) - \log \det(L_{W_n}^\theta[X \cap W_n])|. \end{aligned}$$

By Lemma 5.A.2, we get

$$\begin{aligned} &\frac{1}{\mu(W_n)} |\log \det(L^\theta[X \cap W_n]) - \log \det(L^{\theta_1}[X \cap W_n])| \\ &\leq \frac{e^{A_\Theta} \sqrt{N(W_n)}}{\mu(W_n)} \text{Tr}((L^\theta[X \cap W_n] - L^{\theta_1}[X \cap W_n])^2) \\ &= e^{A_\Theta} \sqrt{\frac{N(W_n)}{\mu(W_n)}} \left(\frac{N(W_n)}{\mu(W_n)} |L_0^\theta(0) - L_0^{\theta_1}(0)|^2 \right. \\ &\quad \left. + \frac{1}{\mu(W_n)} \sum_{x, y \in X \cap W_n}^{\neq} |L_0^\theta(y-x) - L_0^{\theta_1}(y-x)|^2 \right)^{1/2}. \end{aligned}$$

Then, by the strong law of large number in Lemma 5.A.1, we obtain that

$$\limsup_{n \rightarrow +\infty} \sup_{|\theta - \theta_1| \leq 2\eta} |\log \det(L^\theta[X \cap W_n]) - \log \det(L^{\theta_1}[X \cap W_n])|$$

is bounded by

$$\begin{aligned} &e^{A_\Theta} \sqrt{C} \left(C \sup_{|\theta - \theta_1| \leq 2\eta} |L_0^\theta(0) - L_0^{\theta_1}(0)|^2 \right. \\ &\quad \left. + \int_{\mathbb{R}^d} \sup_{|\theta - \theta_1| \leq 2\eta} |L_0^\theta(z) - L_0^{\theta_1}(z)|^2 (K_0^{\theta^*}(0)^2 - K_0^{\theta^*}(z)^2) d\mu(z) \right)^{1/2} \end{aligned}$$

where the bound is true for all event on a set with probability 1 since we applied the law of large number twice for each $\theta_1 \in \Theta_d$ which are countable. Finally, $|L^\theta(0) - L^{\theta_1}(0)|^2$ vanishes when $|\theta - \theta_1| \rightarrow 0$ by continuity of $\theta \mapsto L^\theta(x)$ for all x and the integral also vanishes by dominated convergence since all L^θ are dominated by a square integrable function.

Using the same reasoning as in the proof of Theorem 5.2.3, we also get

$$\begin{aligned}
& \frac{1}{\mu(W_n)} \sup_{\Theta} |\log \det(L^\theta[X \cap W_n]) - \log \det(L_{W_n}^\theta[X \cap W_n])| \\
& \leq \sup_{\Theta} \frac{(\lambda_n^\theta)^{-1}}{\mu(W_n)} \text{Tr}(N_{W_n}^\theta[X \cap W_n]) \\
& \leq \frac{e^{A_\Theta}}{\mu(W_n)} \sum_{x \in X \cap W_n} \int_{W_n^c} \sup_{\theta \in \Theta} L^\theta(x, y)^2 d\mu(y). \tag{5.4.9}
\end{aligned}$$

Since all L^θ are uniformly bounded in Condition $(\mathcal{H}_\varepsilon)$ by a square integrable function h then (5.4.9) is a Lipschitz function of X with coefficient $\mu(W_n)^{-1} e^{A_\Theta} \|h\|_{L^2}$. Moreover, the expectation of (5.4.9) also vanishes when $n \rightarrow +\infty$ by Lemma 5.A.4 hence, by using Pemantle and Peres result [81] as in the proof of Theorem 5.4.3, (5.4.9) vanish when n goes to infinity almost surely.

We then conclude that for all $\theta_0 \in \Theta$, $\phi_{\theta_0}(\eta) \rightarrow 0$ when $\eta \rightarrow 0$ almost surely. Moreover, since the convergence doesn't depend on θ_0 then for all countable subset Θ_d of Θ , $\sup_{\Theta_d} \phi_\theta(\eta) \rightarrow 0$ when $\eta \rightarrow 0$ almost surely.

Now, let $g(\theta) := \lim_{n \rightarrow +\infty} \mu(W_n)^{-1} \log \det(L_{W_n}^\theta[X \cap W_n])$. Since g is continuous on Θ by Theorem 5.4.5 and Θ is compact then g is equicontinuous. Then, for all $\theta \in \Theta$ and all integer $k \geq 1$, let $\theta_1^k, \dots, \theta_{q_k}^k$ such that $\Theta \subset \bigcup_i \mathcal{B}(\theta_i^k, 1/k)$ and let $\Theta_d = \bigcup_{k \geq 1} \{\theta_1^k, \dots, \theta_{q_k}^k\}$. Up to relabeling the θ_i^j we can consider that $|\theta - \theta_1^k| \leq 1/k$ for all k therefore

$$\begin{aligned}
|\log \det(L_{W_n}^\theta[X \cap W_n]) - g(\theta)| & \leq |\log \det(L_{W_n}^\theta[X \cap W_n]) - \log \det(L_{W_n}^{\theta_1^k}[X \cap W_n])| \\
& \quad + |\log \det(L_{W_n}^{\theta_1^k}[X \cap W_n]) - g(\theta_1^k)| + |g(\theta_1^k) - g(\theta)|
\end{aligned}$$

and then

$$\begin{aligned}
\sup_{\theta \in \Theta} |\log \det(L_{W_n}^\theta[X \cap W_n]) - g(\theta)| & \leq \sup_{\Theta_d} \phi_\theta(1/k) + \max_{1 \leq i \leq q_k} |\log \det(L_{W_n}^{\theta_i^k}[X \cap W_n]) - g(\theta_i^k)| \\
& \quad + \sup_{|\theta - \theta'| < 1/k} |g(\theta') - g(\theta)|.
\end{aligned}$$

Finally, we can choose k such that $\sup_{\theta \in \Theta_d} \phi_\theta(1/k)$ and $\sup_{|\theta - \theta'| < 1/k} |g(\theta') - g(\theta)|$ are arbitrary small with probability 1 and since $\max_{1 \leq i \leq q_k} |\log \det(L_{W_n}^{\theta_i^k}[X \cap W_n]) - g(\theta_i^k)|$ vanishes for all k also with probability 1 then

$$\sup_{\theta \in \Theta} |\log \det(L_{W_n}^\theta[X \cap W_n]) - g(\theta)| \xrightarrow{a.s.} 0.$$

Finally, we can conclude about the consistency of $\hat{\theta}_n$ by first assuming that $\hat{\theta}_n$ doesn't converge almost surely towards θ^* . In that case, for all event ω in a set of probability > 0 there exist a subsequence $\phi(n)$ and $\varepsilon > 0$ such that $|\hat{\theta}_{\phi(n)}(\omega) - \theta^*| > \varepsilon$. But, since Θ is compact, then there also exists a subsequence $\hat{\theta}_{\psi(n)}(\omega)$ of $\hat{\theta}_{\phi(n)}(\omega)$ that converges towards a certain $\theta_0(\omega) \in \Theta$. We know that for all $\theta \in \Theta$,

$$l_{\psi(n)}(\hat{\theta}_{\psi(n)}(\omega)|X) > l_{\psi(n)}(\theta|X).$$

The right term converges almost surely to $l(\theta)$ and the left term converges almost surely to $l(\theta_0(\omega))$ by uniform convergence and continuity of l . Thus, $\theta_0(\omega) = \theta^*$ almost surely by uniqueness of the maximum which contradict our previous assumption. Therefore, $\hat{\theta}_n$ converges almost surely towards θ^* .

5.A Technical Lemmas

When W_n is a convex averaging sequence, the following Lemma is a well-known result as a consequence of DPPs being ergodic. Since our setting is slightly different, we provide a different proof.

Lemma 5.A.1. *Let X be a stationary DPP on (\mathbb{R}^d, μ) with kernel $K_0(\cdot) \in L^2(\mathbb{R}^d, \mu)$ and assume that W_n satisfies Condition (W). Then,*

$$\frac{N(W_n)}{\mu(W_n)} \xrightarrow{a.s.} K_0(0).$$

and for all bounded function $f \in L^1(\mathbb{R}^d, \mu)$,

$$\frac{1}{\mu(W_n)} \sum_{x,y \in X \cap W_n}^{\neq} f(y-x) \xrightarrow{a.s.} \int_{\mathbb{R}^d} f(x)(K_0(0)^2 - K_0(x)^2) d\mu(x)$$

Proof. We have $\mathbb{E}[\mu(W_n)^{-1}N(W_n)] = K_0(0)$ and since $N(W_n)$ is a 1-Lipschitz function then by Pemantle and Peres [81] inequality, we get

$$\mathbb{P}_{\theta^*} \left(\left| \frac{N(W_n)}{\mu(W_n)} - K_0(0) \right| > \frac{1}{k} \right) \leq 5 \exp \left(- \frac{\mu(W_n)^2}{16(k\mu(W_n) + 2k^2K_0(0)\mu(W_n))} \right)$$

which is summable w.r.t. k by our assumption on W_n hence $\mu(W_n)^{-1}N(W_n) \xrightarrow{a.s.} K_0(0)$ by Borel-Cantelli's lemma.

For the second expression, we define

$$S_W := \sum_{x,y \in X \cap W}^{\neq} f(y-x).$$

and we first assume that f is non-negative. We can get the limit of the expectation of $\mu(W_n)^{-1}S_n$ by Lemma 5.A.4 since W_n satisfies (5.2.7):

$$\begin{aligned} \mathbb{E} \left[\frac{S_{W_n}}{\mu(W_n)} \right] &= \frac{1}{\mu(W_n)} \int_{W_n^2} f(y-x) \det(K[0, y-x]) d\mu(x) d\mu(y) \\ &\xrightarrow{n \rightarrow +\infty} \int_{\mathbb{R}^d} f(x) \det(K_0(0)^2 - K_0(x)^2) d\mu(x). \end{aligned}$$

We can also develop the variance into

$$\begin{aligned} \text{Var} \left(\frac{S_{W_n}}{\mu(W_n)} \right) &= \frac{1}{\mu(W_n)^2} \left(2 \int_{W_n^2} f(y-x)^2 \det(K[0, y-x]) d\mu(x) d\mu(y) \right. \\ &\quad \left. + 4 \int_{W_n^3} f(y-x) f(z-x) \det(K[0, y-x, z-x]) d\mu^3(x, y, z) \right. \\ &\quad \left. + \int_{W_n^4} f(y-x) f(t-z) (\det(K[x, y, z, t]) - \det(K[x, y]) \det(K[z, t])) d\mu^4(x, y, z, t) \right). \end{aligned} \tag{5.A.1}$$

The first term in (5.A.1) is bounded by

$$2\mu(W_n) \|K\|_{\infty}^2 \int_{\mathbb{R}^d} f(z)^2 d\mu(z) = O(\mu(W_n)).$$

Similarly, the second term in (5.A.1) is bounded by

$$4\mu(W_n)\|K\|_\infty^3 \int_{\mathbb{R}^d} f^{*2}(z)d\mu(z) = O(\mu(W_n)).$$

In order to bound correctly the last term, we remark that

$$\det(K[x, y, z, t]) - \det(K[x, y]) \det(K[z, t])$$

can be expanded into 20 terms, each bounded by $\|K\|_\infty^3|K(y-t)|$, $\|K\|_\infty^3|K(y-z)|$, $\|K\|_\infty^3|K(x-t)|$ or $\|K\|_\infty^3|K(x-z)|$. Moreover, we have

$$\begin{aligned} & \int_{W_n^4} f(y-x)f(t-z)|K_0(y-t)|d\mu(x)d\mu(y)d\mu(z)d\mu(t) \\ & \leq \mu(W_n) \int_{\mathbb{R}^d} (f * |K_0| * f)(x)d\mu(x) = O(\mu(W_n)) \end{aligned} \quad (5.A.2)$$

and all other terms can be dealt the same way. Finally, the variance (5.A.1) is $O(\mu(W_n)^{-1})$ which is summable if $\mu(W_n) \sim n^2$. In that case,

$$\frac{1}{\mu(W_n)} \sum_{x,y \in X \cap W_n}^{\neq} f(y-x) \xrightarrow{a.s.} \int_{\mathbb{R}^d} f(x) \det(K_0(0)^2 - K_0(x)^2) d\mu(x)$$

using Bienaymé-Tchebychev's inequality and Borel-Cantelli's lemma. In the more general case, since $\mu(W_n) \rightarrow +\infty$ then we can choose a sequence $(V_n)_{n \in \mathbb{N}}$ such that $\mu(V_n) \sim n^2$ and for all n there exists a unique m such that $W_m \subset V_n \subset W_{m+1}$. Now, for all k , there exists n such that $V_n \subset W_k \subset V_{n+1}$ and since S_W is increasing then

$$\frac{S_{V_n}}{\mu(V_n)} \frac{\mu(V_n)}{\mu(V_{n+1})} \leq \frac{S_{W_k}}{\mu(W_k)} \leq \frac{S_{V_{n+1}}}{\mu(V_{n+1})} \frac{\mu(V_{n+1})}{\mu(V_n)}.$$

The right and left term of the inequality converges almost surely towards their expectation since $\mu(V_n) \sim n^2$ which conclude the proof when f is non-negative. The general case is obtained by decomposing f into its positive and negative part and applying the previous result to each part. \square

Lemma 5.A.2. *For all $n \times n$ symmetrical positive definite matrix A, B ,*

$$|\log \det(A) - \log \det(B)| \leq \sqrt{n} \|A - B\|_F \max(\lambda_{\min}(A)^{-1}, \lambda_{\min}(B)^{-1})$$

where $\|\cdot\|_F$ is the Froebenius norm.

Proof. Let $f : t \mapsto \log \det(tA + (1-t)B)$. It's derivative is equal to

$$f'(t) = \text{Tr}((A - B)(tA + (1-t)B)^{-1}).$$

Moreover, $\lambda_{\min}(tA + (1-t)B) \geq t\lambda_{\min}(A) + (1-t)\lambda_{\min}(B) \geq \min(\lambda_{\min}(A), \lambda_{\min}(B))$ therefore

$$\sup_{[0,1]} |f'(t)| \leq \|A - B\|_F \|(tA + (1-t)B)^{-1}\|_F \leq \sqrt{n} \|A - B\|_F \max(\lambda_{\min}(A)^{-1}, \lambda_{\min}(B)^{-1}).$$

Since $f(0) = \log \det(B)$ and $f(1) = \log \det(A)$ we get the desired bound. \square

Lemma 5.A.3. *Let $n \in \mathbb{N}$ and A, B be two $n \times n$ positive semi-definite matrices. Then,*

$$0 \leq \log \det(I + AB) \leq \text{Tr}(AB)$$

Proof. We first assume that B is the identity matrix. Let $\lambda_1, \dots, \lambda_n$ be the eigenvalues (with multiplicity) of A . Then,

$$0 \leq \log \det(I + M) = \sum_{i=1}^n \log(1 + \lambda_i) \leq \sum_{i=1}^n \lambda_i = \text{Tr}(A).$$

In the general case, Sylvester's determinant identity gives us

$$0 \leq \log \det(I + AB) = \log \det(I + A^{1/2}BA^{1/2}) \leq \text{Tr}(A^{1/2}BA^{1/2}) = \text{Tr}(AB).$$

□

Lemma 5.A.4. *Let $f : (\mathbb{R}^d)^k \mapsto \mathbb{R}$ be a translation invariant function integrable with respect to its $k-1$ last coordinates and the background measure μ . Let W_n be a sequence of increasing compact subsets of \mathbb{R}^d such that*

$$\mu((\partial W_n \oplus r_n) \cap W_n) = o(\mu(W_n)),$$

then

$$\frac{1}{\mu(W_n)} \int_{W_n^k} f(x) d\mu(x) \xrightarrow{n \rightarrow +\infty} \int_{(\mathbb{R}^d)^{k-1}} f(0, x_2, \dots, x_k) d\mu(x_2) \cdots d\mu(x_k) \quad (5.A.3)$$

Proof. We write $W_n \ominus r_n$ for the set $W_n \setminus (\partial W_n \oplus r_n)$ of points at distance at least r_n from the boundary of W_n . Since f is translation invariant then the right term in (5.A.3) is equal to

$$\frac{1}{\mu(W_n)} \int_{W_n \times (\mathbb{R}^d)^{k-1}} f(x) d\mu(x).$$

Then,

$$\begin{aligned} & \left| \int_{(\mathbb{R}^d)^{k-1}} f(0, x_2, \dots, x_k) d\mu(x_2) \cdots d\mu(x_k) - \frac{1}{\mu(W_n)} \int_{W_n^k} f(x) d\mu(x) \right| \\ &= \frac{1}{\mu(W_n)} \int_{W_n \times \mathbb{R}^{d(k-1)} \setminus (W_n^{k-1})} f(x) d\mu(x) \\ &= \frac{1}{\mu(W_n)} \int_{W_n \ominus r_n} \int_{\mathbb{R}^{d(k-1)} \setminus (W_n^{k-1})} f(x) d\mu(x) \\ & \quad + \frac{1}{\mu(W_n)} \int_{(\partial W_n \oplus r_n) \cap W_n} \int_{\mathbb{R}^{d(k-1)} \setminus (W_n^{k-1})} f(x) d\mu(x) \\ &\leq \frac{1}{\mu(W_n)} \int_{W_n \ominus r_n} \left(\int_{(\mathbb{R}^d)^{k-1}} f(0, y) \mathbb{1}_{\{\forall i, \|y_i\| > r_n\}} d\mu(y) \right) d\mu(x) \\ & \quad + \frac{1}{\mu(W_n)} \int_{(\partial W_n \oplus r_n) \cap W_n} \left(\int_{(\mathbb{R}^d)^{k-1}} f(0, y) d\mu(y) \right) d\mu(x) \\ &\leq \int_{\mathcal{B}(0, r_n)^{k-1}} f(0, y) d\mu(y) + \frac{\mu((\partial W_n \oplus r_n) \cap W_n)}{\mu(W_n)} \|f(0, \cdot)\|_{L^1} \end{aligned}$$

which converges to 0 because f is integrable with respect to its last $k-1$ variables. □

Lemma 5.A.5. *Let $\{K^\theta, \theta \in \Theta\}$ be a compact family of DPP kernels on \mathbb{R}^d (with the Lebesgue measure) satisfying Condition (\mathcal{H}) and, for all ε , define $K_\varepsilon^\theta(x, y) := \varepsilon^d K^\theta(x, y)$. Then, for small enough ε , $\{K_\varepsilon^\theta, \theta \in \Theta\}$ is a compact family of DPP kernels on $(\varepsilon\mathbb{Z})^d$ (with the counting measure μ) satisfying condition $(\mathcal{H}_\varepsilon)$.*

Proof. First, we need to show that $\sup_{\Theta} K_{\varepsilon}^{\theta}[(\varepsilon\mathbb{Z})^d]$ is an hermitian matrix with eigenvalues in $[0, 1]$ for a small enough ε . This property is part of Condition $(\mathcal{H}_{\varepsilon})$ but also implies that the associated DPP is well defined. By Gershgorin circle theorem and the positiveness of K^{θ} , the eigenvalues of $\mathcal{K}_{\varepsilon} := K_{\varepsilon}^{\theta}[(\varepsilon\mathbb{Z})^d]$ are bounded by

$$\sum_{x \in (\varepsilon\mathbb{Z})^d} \varepsilon^d K_0^{\theta}(x).$$

Moreover, we can write

$$\begin{aligned} \sup_{\Theta} \left| \sum_{x \in (\varepsilon\mathbb{Z})^d} \varepsilon^d K_0^{\theta}(x) - \int_{\mathbb{R}^d} K_0^{\theta}(x) dx \right| &\leq \frac{\varepsilon\sqrt{d}}{2} \varepsilon^d \sum_{x \in (\varepsilon\mathbb{Z})^d} \sup_{\Theta} \sup_{x \in C_{x,\varepsilon}} |\nabla_x K_0^{\theta}(x)| \\ &\leq \frac{\varepsilon\sqrt{d}}{2} \varepsilon^d \sum_{x \in (\varepsilon\mathbb{Z})^d} \sup_{t \in C_{x,\varepsilon}} \frac{C'}{1 + \|t\|^{\alpha'+d}} \end{aligned}$$

where $C_{x,\varepsilon}$ is the d -dimensional hypercube centered in x with side length ε and the last inequality is obtained as a consequence of Condition (\mathcal{H}) . If $\|x\| \geq 3\varepsilon\sqrt{d}/2$ then we can write

$$\begin{aligned} \sup_{t \in C_{x,\varepsilon}} \frac{C'}{1 + \|t\|^{\alpha'+d}} &\leq \frac{C'}{1 + (\|x\| - \varepsilon\sqrt{d}/2)^{\alpha'+d}} \\ &\leq \inf_{t \in C_{x,\varepsilon}} \frac{C'}{1 + (\|t\| - \varepsilon\sqrt{d})^{\alpha'+d}} \leq \varepsilon^{-d} \int_{C_{x,\varepsilon}} \frac{C'}{1 + (\|x\| - \varepsilon\sqrt{d})^{\alpha'+d}} dx \end{aligned}$$

hence

$$\begin{aligned} \sup_{\Theta} \left| \sum_{x \in (\varepsilon\mathbb{Z})^d} \varepsilon^d K_0^{\theta}(x) - \int_{\mathbb{R}^d} K_0^{\theta}(x) d\mu(x) \right| &\leq \frac{\varepsilon\sqrt{d}}{2} \left(\varepsilon^d \sum_{\substack{x \in (\varepsilon\mathbb{Z})^d \\ \|x\| \leq 3\varepsilon\sqrt{d}/2}} \sup_{t \in C_{x,\varepsilon}} \frac{C'}{1 + \|t\|^{\alpha'+d}} + \int_{\mathbb{R}^d} \frac{C'}{1 + \|x\|^{\alpha'+d}} dx \right) \end{aligned}$$

which vanishes when $\varepsilon \rightarrow 0$. Therefore, the largest eigenvalue of $\mathcal{K}_{\varepsilon}$ uniformly converges to $\hat{K}_0^{\theta}(0) < 1$ hence, we can find a small enough ε such that $\sup_{\Theta} \|K_{\varepsilon}^{\theta}[(\varepsilon\mathbb{Z})^d]\| < 1$ and the associated DPPs are all well-defined. Moreover, it is proved in [4] that, under the assumption that

$$|K_0^{\theta}(x)| \leq \frac{C}{1 + \|x\|^{\alpha'+d}} \quad \text{and} \quad \forall x \in \mathbb{R}^d, \quad \inf_{\Theta} \hat{K}_0^{\theta}(x) > 0,$$

both being satisfied as a consequence of Condition (\mathcal{H}) , then $\inf_{\Theta} \lambda_{\min}(K_{\varepsilon}^{\theta}[(\varepsilon\mathbb{Z})^d]) > 0$. Finally, in order to show the bound (5.4.3) on L_0 , recall that we can write

$$L_{0,\varepsilon}(x) = \sum_{n \geq 1} K_{0,\varepsilon}(x)^{*n}$$

where $*n$ denotes the n -th self discrete convolution on \mathbb{Z}^d . Therefore,

$$\sup_x |L_{0,\varepsilon}^{\theta}(x)| \leq C \sum_{n \geq 0} \left(\varepsilon^d \sum_{x \in (\varepsilon\mathbb{Z})^d} |K_0^{\theta}(x)| \right)^n.$$

But, since ε has been chosen such that

$$\sup_{\Theta} \varepsilon^d \sum_{x \in (\varepsilon\mathbb{Z})^d} K_0^\theta(x) < 1$$

and K_0^θ is positive, we get a uniform bound for $L_{0,\varepsilon}^\theta$. Now,

$$K_{0,\varepsilon}^\theta(x)^{*n} = \varepsilon^{nd} \sum_{y_1, \dots, y_n \in (\varepsilon\mathbb{Z})^d} K^\theta(y_1) K^\theta(y_2 - y_1) \cdots K^\theta(x - y_n)$$

and since at least one element of $\{\|y_1\|, \|y_2 - y_1\|, \dots, \|x - y_n\|\}$ is greater than $\|x\|/n$, then

$$|K_{0,\varepsilon}^\theta(x)^{*n}| \leq n \frac{Cn^{\alpha+d}}{\|x\|^{\alpha+d}} \left(\varepsilon^d \sum_{x \in (\varepsilon\mathbb{Z}^d)} |K_0^\theta(x)| \right)^n.$$

But, since ε has been chosen such that

$$\sup_{\Theta} \varepsilon^d \sum_{x \in (\varepsilon\mathbb{Z})^d} K_0^\theta(x) < 1$$

then

$$\sup_{\Theta} \sum_{n \geq 0} Cn^{\alpha+d+1} \left(\varepsilon^d \sum_{x \in (\varepsilon\mathbb{Z}^d)} |K_0^\theta(x)| \right)^n < +\infty$$

giving us the desired bound on $L_{0,\varepsilon}^\theta$. □

Proposition 5.A.6. *Let X be a DPP with Bessel kernel*

$$K^{\rho,\alpha}(x, y) = \rho 2^{d/2} \Gamma(d/2 + 1) \frac{J_{d/2}(\sqrt{2d}\|y - x\|/\alpha)}{(\sqrt{2d}\|y - x\|/\alpha)^{d/2}},$$

then, for all $\alpha > 0$,

$$\frac{N(W_n)}{\mu(W_n)} \in \arg \max_{0 < \rho < v_d^\alpha} \tilde{l}_n(X|\rho, \alpha) \quad (5.A.4)$$

where

$$v_d^\alpha = \frac{d^{d/2}}{(2\pi\alpha^2)^{d/2} \Gamma(d/2 + 1)}$$

is the upper bound for ρ for which X is well-defined.

Proof. The Fourier transform of the Bessel kernel is known:

$$\hat{K}_0(x) = \frac{\rho}{v_d^\alpha} \mathbb{1}_{\|x\| \leq \sqrt{\frac{d}{2\pi^2\alpha^2}}}$$

By noticing that v_d^α is the volume of the d -dimensional ball with radius $\sqrt{\frac{d}{2\pi^2\alpha^2}}$, we get that

$$\int_{\mathbb{R}^d} \log(1 - \hat{K}_0^{\rho,\alpha}(x)) d\mu(x) = v_d^\alpha \log(1 - \rho/v_d^\alpha).$$

Moreover, $L^{\rho,\alpha}$ can be written as $\frac{\rho}{1-\rho/v_d^\alpha} F^\alpha$, where F^α is a function not depending on ρ . Therefore, $\log \det(L^{\rho,\alpha}[X \cap W_n])$ can be expressed as the sum of

$$N(W_n) \log \left(\frac{\rho}{1 - \rho/v_d^\alpha} \right)$$

and an expression not depending on ρ . Finally, $\tilde{l}_n(X|\rho, \alpha)$ is twice differentiable with respect to ρ with derivative

$$\frac{-1}{1 - \rho/v_d^\alpha} + \frac{N(W_n)}{\mu(W_n)\rho(1 - \rho/v_d^\alpha)}.$$

It is easy to see that this expression vanishes only when $\rho = N(W_n)/\mu(W_n)$ with the second derivative being negative at this point, concluding the proof. \square

Appendix A

Supplementary Results

A.1 Negative association of finite DPPs

In this section, we denote by X a discrete determinantal process on $[n] := \{1, \dots, n\}$, $|X|$ its cardinality, K its matrix kernel and X_i the Bernoulli random variables equal to 1 when $i \in X$. Our goal is to prove that X is negatively associated:

Theorem A.1.1. *Let X be a DPP on $[n]$ and $F, G : \mathcal{P}([n]) \rightarrow \mathbb{R}$ be increasing function (with respect to the inclusion). For all A, B disjoint subsets of $[n]$,*

$$\mathbb{E}[F(X \cap A)G(X \cap B)] \leq \mathbb{E}[F(X \cap A)]\mathbb{E}[G(X \cap B)]. \quad (\text{A.1.1})$$

Note that the functions F and G can be expanded the following way.

Lemma A.1.2. *Let $F : \mathcal{P}([n]) \rightarrow \mathbb{R}$ be an increasing function and let $A \subset [n]$. We call an event \mathcal{A} increasing if its indicator function is increasing. Then, $F(X \cap A)$ can be written as*

$$F(X \cap A) = c + \sum_{i=1}^k a_i \mathbb{1}_{\mathcal{A}_i}(X)$$

where c is a constant, $a_i > 0$ for all i and the events \mathcal{A}_i are increasing and in the σ -algebra generated by $X \cap A$.

Proof. Let $c = \min(F)$ and $\tilde{F}(X) = F(X \cap A) - c$. \tilde{F} is non-negative and we call $0 = b_0 < b_1 < \dots < b_k$ the different values taken by \tilde{F} . Consider the events $\mathcal{A}_i = \{\tilde{F} \geq a_i\}$. The \mathcal{A}_i are increasing events of $\sigma(X \cap A)$ and we can write

$$\tilde{F} = \sum_{i=1}^k (b_i - b_{i-1}) \mathbb{1}_{\mathcal{A}_i},$$

proving the lemma. □

A proof of Theorem A.1.1 is given in [73] using exterior algebra properties. Our goal is to give a detailed proof that only uses probabilistic tools. We recall a few properties of discrete DPPs.

Proposition A.1.3. *Let X be a DPP on $[n]$ with kernel K . Its complementary $\bar{X} := [n] \setminus X$ is also a DPP on $[n]$ with kernel $I - K$.*

Proof. Let $A \subset [n]$, by using the inclusion–exclusion principle we get

$$\begin{aligned}
\mathbb{P}(A \subset \bar{X}) &= \mathbb{P}(A \cap X = \emptyset) \\
&= 1 - \mathbb{P}(\{i\} \subset X \text{ for some } i \in A) \\
&= 1 - \sum_{k=1}^{|A|} (-1)^{k-1} \sum_{i_1 < \dots < i_k \in A} \mathbb{P}(\{i_1, \dots, i_k\} \subset X) \\
&= \sum_{S \subset A} (-1)^{|S|} \det(K_S) \\
&= \det((I - K)_A).
\end{aligned}$$

□

Proposition A.1.4. *Let X be a DPP on $[n]$ with kernel K , where K is a projection matrix with rank k . Then, $|X| = k$ almost surely. Such a DPP is called a projection DPP.*

Proof. Let $A \subset [n]$ such that $|A| > k$. Then, $\mathbb{P}(A \subset X) = \det(K_A) = 0$ since K is of rank k . So, $|X| \leq k$ almost surely. Moreover,

$$\mathbb{E}[|X|] = \sum_{i=1}^n \mathbb{P}(i \in X) = \text{Tr}(K) = k$$

concluding the proof. □

Finally, an immediate consequence of the identity

$$\det \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} = \det(C) \det(A - BC^{-1}B^T)$$

is that a DPP X stays determinantal when conditioned by the absence or presence of points in X .

Proposition A.1.5. *Let X be a DPP on $[n]$ with kernel K . Let $A \subset [n]$ such that $\mathbb{P}(A \subset X) > 0$ then, conditionally to $\{A \subset X\}$, $X \setminus A$ is a determinantal process on $[n] \setminus A$ with kernel*

$$K^A := K_{A^c} - K_{A^c A} K_A^{-1} K_{A^c A}^T \text{ where } K_{A^c A} = (K_{i,j})_{i \in A^c, j \in A}.$$

Let $A \subset [n]$ such that $\mathbb{P}(A \cap X = \emptyset) > 0$ then, conditionally to $\{A \cap X = \emptyset\}$, X is a determinantal process on $[n] \setminus A$ with kernel $I - (I - K)^A$. Moreover, projection DPPs under both conditioning stays projection DPPs.

We now give a probabilistic proof of Theorem A.1.1.

Proof of Theorem A.1.1. Let $F, G : \mathcal{P}([n]) \rightarrow \mathbb{R}$ be increasing functions and A, B be two disjoint subsets of $[n]$. Using Proposition A.1.2 we can write $F(X \cap A)$ and $G(X \cap B)$ as the sum of a constant and a positive linear combination of indicator functions of increasing events of, respectively, $\sigma(X \cap A)$ and $\sigma(X \cap B)$. Considering that the constant doesn't affect the inequality (A.1.1), negative association only has to be proved for functions $F = \mathbb{1}_{\mathcal{A}}$ and $G = \mathbb{1}_{\mathcal{B}}$ where \mathcal{A} and \mathcal{B} are increasing events in, respectively, $\sigma(X \cap A)$ and $\sigma(X \cap B)$. This is done in three steps. In the first step, we consider the case where X is a projection DPP and where $F = \mathbb{1}_{\mathcal{A}}$ and $G = X_i$. In the second step, we extend the result to the case where $F = \mathbb{1}_{\mathcal{A}}$ and $G = \mathbb{1}_{\mathcal{B}}$. Finally, we

conclude in the last step. We also say that an event \mathcal{A} ignores i (or does not depends on i) if $\mathcal{A} \in \sigma(X \setminus \{i\})$. This is equivalent to saying that for all $X \in \mathcal{A}$, $X \cup \{i\} \in \mathcal{A}$ and $X \setminus \{i\} \in \mathcal{A}$.

Step 1: In the first step of the proof, we consider that X is a projection DPP (see Proposition A.1.4) and that $F = X_i$ and $G = \mathbb{1}_{\mathcal{A}}$ where \mathcal{A} is an increasing event ignoring i . In this case, the negative association property (A.1.1) can be written as

$$\mathbb{E}[X_i \mathbb{1}_{\mathcal{A}}] \leq \mathbb{E}[X_i] \mathbb{P}(\mathcal{A}). \quad (\text{A.1.2})$$

If $\mathbb{E}[X_i] = 0$ or 1 the result is trivial so we consider that $\mathbb{E}[X_i] \in]0, 1[$. This inequality is then equivalent to $\mathbb{P}(\mathcal{A} | i \in X) \leq \mathbb{P}(\mathcal{A})$ as well as $\mathbb{P}(\mathcal{A}) \leq \mathbb{P}(\mathcal{A} | i \notin X)$. By combining these two inequalities we get that (A.1.2) is also equivalent to $\mathbb{P}(\mathcal{A} | i \in X) \leq \mathbb{P}(\mathcal{A} | i \notin X)$. In the end, we obtained three inequalities equivalent to the one we want to prove that we use throughout the proof.

$$\begin{aligned} (\text{A.1.2}) &\Leftrightarrow \mathbb{P}(\mathcal{A} | i \in X) \leq \mathbb{P}(\mathcal{A}) \\ &\Leftrightarrow \mathbb{P}(\mathcal{A}) \leq \mathbb{P}(\mathcal{A} | i \notin X) \\ &\Leftrightarrow \mathbb{P}(\mathcal{A} | i \in X) \leq \mathbb{P}(\mathcal{A} | i \notin X). \end{aligned} \quad (\text{A.1.3})$$

We show that X satisfies these inequalities by induction on the size of $[n]$. The key property making the induction work is Proposition A.1.5 stating that the measures $\mathbb{P}(\cdot | i \notin X)$ and $\mathbb{P}(\cdot \cup \{i\} | i \in X)$ are still projection determinantal measures on $[n] \setminus \{i\}$ for all $i \in [n]$.

$n=1$: In this case, $X_i = X_1$ so $\mathcal{A} = \emptyset$ or $\mathcal{P}([1])$. If $\mathcal{A} = \emptyset$ then $X_i \mathbb{1}_{\mathcal{A}} = 0$ and otherwise $\mathbb{P}(\mathcal{A}) = 1$ so the negative association is satisfied in both cases.

$n-1 \rightarrow n$: We consider that all projection DPPs on $[l]$, where $l \leq n-1$, satisfies (A.1.2). If we consider the case where $\mathbb{P}(i, j \in X) = 0$ for all $j \neq i$, we get

$$\mathbb{P}(\mathcal{A} | i \in X) = \mathbb{P}(\mathcal{A} | X = \{i\}) = \begin{cases} 1 & \text{if } \{i\} \in \mathcal{A} \\ 0 & \text{otherwise.} \end{cases}$$

If $\mathbb{P}(\mathcal{A} | i \in X) = 0$ then (A.1.2) is satisfied. Otherwise, $\{i\} \in \mathcal{A}$ and since \mathcal{A} ignore i then $\emptyset \in \mathcal{A}$ and thus, because \mathcal{A} is increasing, we get $\mathcal{A} = \mathcal{P}([n])$ which implies (A.1.2). We then consider the case where the set $C := \{j \in [n], \mathbb{P}(i, j \in X) = 0\}$ contains an element distinct from i . The fact that $\sum_j X_j = k = \sum_j \mathbb{P}(j \in X)$ gives us

$$\sum_{\substack{j \in [n] \\ j \neq i}} \mathbb{E}[X_j \mathbb{1}_{\mathcal{A}} | i \in X] = (k-1) \mathbb{P}(\mathcal{A} | i \in X) = \left(\sum_{\substack{j \in [n] \\ j \neq i}} \mathbb{P}(j \in X | i \in X) \right) \mathbb{P}(\mathcal{A} | i \in X).$$

But, if $j \in C$ then, conditionally to $i \in X$, $j \notin X$ a.s. so the previous equality stays true if we consider sums on $[n] \setminus C$ instead. In particular, we can choose a $j \notin C$ distinct from i such that

$$\mathbb{E}[X_j \mathbb{1}_{\mathcal{A}} | i \in X] \geq \mathbb{E}[X_j | i \in X] \mathbb{P}(\mathcal{A} | i \in X).$$

This inequality correspond to the reverse of (A.1.2) conditionally to $i \in X$. We can especially deduce that we have the same equivalence as (A.1.3) but with opposite signs, hence j satisfies

$$\mathbb{P}(\mathcal{A} | i, j \in X) \geq \mathbb{P}(\mathcal{A} | i \in X, j \notin X). \quad (\text{A.1.4})$$

Since we took j such that $\mathbb{P}(i, j \in X) \neq 0$, when conditioning $\mathbb{P}(\mathcal{A} | i \in X)$ by j we get

$$\mathbb{P}(\mathcal{A} | i \in X) = \mathbb{P}(\mathcal{A} | i, j \in X) \mathbb{P}(j \in X | i \in X) + \mathbb{P}(\mathcal{A} | i \in X, j \notin X) \mathbb{P}(j \notin X | i \in X) \quad (\text{A.1.5})$$

where we consider the right term as being zero if $\mathbb{P}(i \in X, j \notin X) = 0$. Since X satisfies

$$\mathbb{P}(j \in X | i \in X) = \frac{K(i, i)K(j, j) - K(i, j)^2}{K(i, i)} \leq K(j, j) = \mathbb{P}(j \in X)$$

then (A.1.3) is satisfied for the event $\mathcal{A} = \{j \in X\}$ therefore

$$\mathbb{P}(j \in X | i \in X) \leq \mathbb{P}(j \in X | i \notin X).$$

Then, by combining this result with (A.1.4), we can bound (A.1.5) by

$$\mathbb{P}(\mathcal{A} | i, j \in X) \mathbb{P}(j \in X | i \notin X) + \mathbb{P}(\mathcal{A} | i \in X, j \notin X) \mathbb{P}(j \notin X | i \notin X) \quad (\text{A.1.6})$$

because the difference between (A.1.6) et (A.1.5) is then equal to

$$(\mathbb{P}(\mathcal{A} | i, j \in X) - \mathbb{P}(\mathcal{A} | i \in X, j \notin X))(\mathbb{P}(j \in X | i \notin X) - \mathbb{P}(j \in X | i \in X)) \geq 0$$

The only thing left to do is applying the induction hypothesis on the measures $\mathbb{P}(\cdot | j \notin X)$ and $\mathbb{P}(\cdot | j \in X)$ on $\sigma(X \setminus \{j\})$. In order to do that, we use the following lemma.

Lemma A.1.6. *Let \mathcal{A} be an increasing event of $\mathcal{P}([n])$ ignoring i and $j \neq i \in [n]$. We define the events $\mathcal{A}/j = \{X \in \mathcal{A} : j \in X\}$ and $\mathcal{A} \setminus j = \{X \in \mathcal{A} : j \notin X\}$ in $\sigma(X \setminus \{j\})$. These events are increasing and ignore i .*

Proof. Let's start by showing that they are increasing. Let $X \in \mathcal{A}/j$ and $Y \subset [n]$ containing X . Since \mathcal{A} is increasing and $j \in X$ then $Y \in \mathcal{A}$ and $j \in Y$ so $Y \in \mathcal{A}/j$ which shows that \mathcal{A}/j is increasing. Similarly, let $X \in \mathcal{A} \setminus j$ and $X \subset Y \subset [n] \setminus \{j\}$. Then $j \notin Y$ and since \mathcal{A} is increasing then $Y \in \mathcal{A}$ so $Y \in \mathcal{A} \setminus j$ and we can conclude that $\mathcal{A} \setminus j$ is increasing.

Now, let's show that they ignore i . Let $X \in \mathcal{A}/j$ and thus $X \in \mathcal{A}$ and $j \in X$. Since \mathcal{A} ignores i then $X \cup \{i\}$ and $X \setminus \{i\}$ are also elements of \mathcal{A} that contains j so $X \cup \{i\} \in \mathcal{A}/j$ and $X \setminus \{i\} \in \mathcal{A}/j$ hence \mathcal{A}/j ignores i . Similarly, let $X \in \mathcal{A} \setminus j$ and thus $j \notin X$. Since \mathcal{A} ignore i then $X \cup \{i\}$ and $X \setminus \{i\}$ are also element of \mathcal{A} that doesn't contain j so $X \cup \{i\} \in \mathcal{A} \setminus j$ and $X \setminus \{i\} \in \mathcal{A} \setminus j$ hence $\mathcal{A} \setminus j$ ignores i . \square

This lemma allows us to apply the induction hypothesis to $\mathbb{P}(\mathcal{A}/j | j \in X)$ and $\mathbb{P}(\mathcal{A} \setminus j | j \notin X)$. As a consequence, we get

$$\mathbb{P}(\mathcal{A} | i, j \in X) = \mathbb{P}(\mathcal{A}/j | i, j \in X) \leq \mathbb{P}(\mathcal{A}/j | i \notin X, j \in X) = \mathbb{P}(\mathcal{A} | i \notin X, j \in X)$$

as well as

$$\mathbb{P}(\mathcal{A} | i \in X, j \notin X) = \mathbb{P}(\mathcal{A} \setminus j | i \in X, j \notin X) \leq \mathbb{P}(\mathcal{A} \setminus j | i, j \notin X) = \mathbb{P}(\mathcal{A} | i, j \notin X).$$

By applying these two inequalities to (A.1.6), we get the result we were looking for:

$$\begin{aligned} \mathbb{P}(\mathcal{A} | i \in X) &\leq \mathbb{P}(\mathcal{A} | i \notin X, j \in X) \mathbb{P}(j \in X | i \notin X) \\ &\quad + \mathbb{P}(\mathcal{A} | i \notin X, j \notin X) \mathbb{P}(j \notin X | i \notin X) = \mathbb{P}(\mathcal{A} | i \notin X). \end{aligned}$$

Step 2: In the second step of the proof, we consider that $F = \mathbb{1}_{\mathcal{A}}$ and $G = \mathbb{1}_{\mathcal{B}}$ where \mathcal{A} and \mathcal{B} are increasing events in, respectively, $\sigma(X \cap A)$ and $\sigma(X \cap B)$. If $\mathbb{P}(\mathcal{B}) = 0$ the result is trivial so we consider the case where $\mathbb{P}(\mathcal{B}) > 0$. The negative association property (A.1.1) can be written in this case as

$$\mathbb{P}(\mathcal{A})\mathbb{P}(\mathcal{B}) \geq \mathbb{P}(\mathcal{A} \cap \mathcal{B}) \Leftrightarrow \mathbb{P}(\mathcal{A}) \geq \mathbb{P}(\mathcal{A} | \mathcal{B}).$$

This inequality is proved by induction on the size of $[n]$.

$n=1$: By symmetry, we suppose that \mathcal{A} depends on X_1 and \mathcal{B} ignores 1. The only increasing event of $\mathcal{P}(\{1\})$ depending on 1 is $\{1 \in X\}$ so $\mathbb{1}_{\mathcal{A}} = X_1$ which was done in step 1.

$n-1 \rightarrow n$: In this case, by using the fact that $\sum_j X_j = k = \sum_j \mathbb{P}(j \in X)$ we get that

$$\sum_{j \in [n]} \mathbb{P}(j \in X | \mathcal{B}) = k = \sum_{j \in [n]} \mathbb{P}(j \in X). \quad (\text{A.1.7})$$

We showed in the first step that if i is ignored by \mathcal{B} then $\mathbb{P}(i \in X | \mathcal{B}) \leq \mathbb{P}(i \in X)$ meaning that we can choose an $i \in [n]$ that is not ignored by \mathcal{B} such that

$$\mathbb{P}(i \in X | \mathcal{B}) \geq \mathbb{P}(i \in X) > 0,$$

because, if $\mathbb{P}(i \in X) = 0$ then $\mathbb{P}(i \in X | \mathcal{B}) = 0$ too and so both quantities doesn't change the sum (A.1.7). By conditioning on i , we obtain

$$\mathbb{P}(\mathcal{A} | \mathcal{B}) = \mathbb{P}(\mathcal{A} | \mathcal{B}, i \in X) \mathbb{P}(i \in X | \mathcal{B}) + \mathbb{P}(\mathcal{A} | \mathcal{B}, i \notin X) \mathbb{P}(i \notin X | \mathcal{B})$$

where we consider the second term as zero is $\mathbb{P}(i \notin X) = 0$. By using the induction hypothesis on $\mathbb{P}(\cdot | i \notin X)$ and $\mathbb{P}(\cdot \cup \{i\} | i \in X)$ combined with the lemma A.1.6 we get

$$\mathbb{P}(\mathcal{A} | \mathcal{B}, i \notin X) \leq \mathbb{P}(\mathcal{A} | i \notin X) \quad \text{and} \quad \mathbb{P}(\mathcal{A} | \mathcal{B}, i \in X) \leq \mathbb{P}(\mathcal{A} | i \in X)$$

so $\mathbb{P}(\mathcal{A} | \mathcal{B})$ is bounded by

$$\mathbb{P}(\mathcal{A} | i \in X) \mathbb{P}(i \in X | \mathcal{B}) + \mathbb{P}(\mathcal{A} | i \notin X) \mathbb{P}(i \notin X | \mathcal{B}). \quad (\text{A.1.8})$$

In the first step we saw that

$$\mathbb{P}(\mathcal{A} | i \in X) \leq \mathbb{P}(\mathcal{A} | i \notin X)$$

because i is ignored by \mathcal{A} . Combined with the choice of i , we deduce that $\mathbb{P}(\mathcal{A} | \mathcal{B})$ is bounded by

$$\mathbb{P}(\mathcal{A} | i \in X) \mathbb{P}(i \in X) + \mathbb{P}(\mathcal{A} | i \notin X) \mathbb{P}(i \notin X) = \mathbb{P}(\mathcal{A}) \quad (\text{A.1.9})$$

because the difference between (A.1.9) and (A.1.8) is equal to

$$(\mathbb{P}(\mathcal{A} | i \notin X) - \mathbb{P}(\mathcal{A} | i \in X)) (\mathbb{P}(i \in X | \mathcal{B}) - \mathbb{P}(i \in X)) \geq 0,$$

which conclude our theorem when X is a projection DPP.

Step 3: Let's get back to the more general case where X is a determinantal process whose matrix kernel K is symmetric with eigenvalues in $[0, 1]$. We can build R and R' as the positive semi-definite square root of K and $I - K$. We define Z as the projection DPP with matrix kernel

$$\begin{pmatrix} K & RR' \\ R'R & I - K \end{pmatrix}$$

symmetrical and idempotent, defined on $[2n]$ by $Z = X \cup \bar{X}$ where \bar{X} is a copy of the complementary of X independent from X . This is a consequence of Propositions A.1.4 and A.1.3. The distribution of Z is then

$$\mathbb{P}(Z = A) = \begin{cases} \mathbb{P}(X = A \cap [n]) & \text{if } |A| = n \\ 0 & \text{otherwise.} \end{cases}$$

Let \mathbb{P}_z be the associated measure of Z then

$$\mathbb{P}(\mathcal{A}) = \mathbb{P}_z(\bar{\mathcal{A}}) \quad \text{where} \quad \bar{\mathcal{A}} = \{A \cup B, A \in \mathcal{A} \text{ and } |A| + |B| \geq n\}.$$

Moreover, we can easily check that \mathcal{A} and $\bar{\mathcal{A}}$ depends on the same set of variables and that \mathcal{A} increasing implies $\bar{\mathcal{A}}$ increasing. Since Z satisfies the negative association property then it shows that X satisfies it too for all functions $F = \mathbb{1}_{\mathcal{A}}$ and $G = \mathbb{1}_{\mathcal{B}}$ where \mathcal{A} and \mathcal{B} are increasing events depending on a distinct set of variables and thus for all increasing functions by Proposition A.1.2 concluding the proof. \square

A.2 Generalized Cramér-Wold device

This result is part of an article [17] published in Statistics and Probability Letters.

The Cramér-Wold device is a useful statistical tool to extend one dimensional CLT to a multi-dimensional setting. Let X_n be a sequence of p -dimensional random variables. The Cramér-Wold device states that X_n converges in distribution to a random variable X iff for all $a \in \mathbb{R}^p$, $a^T X_n$ converges in distribution to $a^T X$. In statistical applications we often want to show that $\text{Var}(X_n)^{-1/2} X_n$ converges in distribution to a standard normal distribution for some sequence of statistics X_n . When $\text{Var}(X_n)$ converges to a fixed positive definite matrix, a direct consequence of the Cramér-Wold device is that

$$\text{Var}(X_n)^{-1/2} X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, I_p) \Leftrightarrow \forall a \in \mathbb{R}^p, \quad \text{Var}(a^T X_n)^{-1/2} a^T X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1) \quad (\text{A.2.1})$$

where I_p is the $p \times p$ identity matrix. Therefore, the multidimensional CLT on X_n can be deduced by the unidimensional CLT on all its projections. When $\text{Var}(X_n)$ does not converges, as in Theorems 2.3.1 and 2.4.4, this result does not hold. Thus, we give a proof of (A.2.1) under the assumptions that the eigenvalues of $\text{Var}(X_n)$ are bounded.

Lemma A.2.1. *Let $\{X_n\}_{n \in \mathbb{N}}$ be a sequence of random variables in \mathbb{R}^p such that*

$$0 < \liminf_{n \rightarrow \infty} \lambda_{\min}(\text{Var}(X_n)) < \limsup_{n \rightarrow \infty} \lambda_{\max}(\text{Var}(X_n)) < \infty,$$

where for a symmetrical matrix M , $\lambda_{\min}(M)$ and $\lambda_{\max}(M)$ denote the minimal and maximal eigenvalues of M .

Then, $\text{Var}(X_n)^{-1/2} X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, I_p)$ if for all $a \in \mathbb{R}^p$,

$$\left(a^T \text{Var}(X_n) a\right)^{-\frac{1}{2}} a^T X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

Assume by contradiction that $\text{Var}(X_n)^{-1/2} X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} Y \sim \mathcal{N}(0, I_p)$ does not hold. Then there exists a bounded and continuous function f such that $\mathbb{E}f(\text{Var}(X_n)^{-1/2} X_n) - \mathbb{E}f(Y)$ does not converge toward 0. Thus, there exists $\varepsilon > 0$ and a strictly increasing function $b : \mathbb{N} \rightarrow \mathbb{N}$ such that for all $n \in \mathbb{N}$,

$$\left|\mathbb{E}f(\text{Var}(X_{b(n)})^{-1/2} X_{b(n)}) - \mathbb{E}f(Y)\right| > \varepsilon. \quad (\text{A.2.2})$$

Further, since for all $a \in \mathbb{R}^p$, $\left(a^T \text{Var}(X_n) a\right)^{-\frac{1}{2}} a^T X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1)$, we also have

$$\left(a^T \text{Var}(X_{b(n)}) a\right)^{-\frac{1}{2}} a^T X_{b(n)} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

for any $a \in \mathbb{R}^p$. Since the sequence of eigenvalues of the matrices $\text{Var}(X_n)$ is bounded, there exists by the Bolzano-Weierstrass theorem a strictly increasing function $c : \mathbb{N} \rightarrow \mathbb{N}$ such that $\{c(n)\}_{n \in \mathbb{N}} \subset \{b(n)\}_{n \in \mathbb{N}}$, and a matrix Σ such that

$$\text{Var}(X_{c(n)}) \xrightarrow[n \rightarrow \infty]{} \Sigma.$$

Since

$$\left(a^T \text{Var}(X_{c(n)}) a\right)^{-\frac{1}{2}} a^T X_{c(n)} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

we obtain by multiplying with $a^T \text{Var}(X_{c(n)}) a$ and using Slutsky lemma, that

$$a^T X_{c(n)} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} a^T \mathcal{N}(0, \Sigma).$$

Hence, by the Cramér-Wold device,

$$X_{c(n)} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \Sigma)$$

Therefore, by Slutsky lemma,

$$\text{Var}(X_{c(n)})^{-\frac{1}{2}} X_{c(n)} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, I_p).$$

From this we can conclude

$$\left| \mathbb{E}f(\text{Var}(X_{c(n)})^{-1/2} X_{c(n)}) - \mathbb{E}f(Y) \right| \leq \varepsilon$$

for n large enough which contradicts (A.2.2).

Bibliography

- [1] R.H. Affandi, E. Fox, R. Adams, and B. Taskar. Learning the parameters of determinantal point process kernels. In *International Conference on Machine Learning*, pages 1224–1232, 2014.
- [2] M. Aigner. *Combinatorial Theory*. Springer-Verlag, 1979.
- [3] K. Alam and K.M.L. Saxena. Positive dependence in multivariate distributions. *Comm. Statist. Theory Methods A*, 10:1183–1196, 1981.
- [4] F. Bachoc and R. Furrer. On the smallest eigenvalues of covariance matrices of multivariate spatial processes. *Stat*, 5:102–107, 2016.
- [5] A. Baddeley, R. Turner, and E. Rubak. *spatstat.data: Datasets for 'spatstat'*, 2018. R package version 1.4-0.
- [6] A. J. Baddeley, J. Møller, and R. Waagepetersen. Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54(3):329–350, 2000.
- [7] A. J. Baddeley, E. Rubak, and R. Turner. *Spatial Point Patterns: Methodology and Applications with R*. Interdisciplinary Statistics. Chapman & Hall/CRC, Boca Raton, Florida, 2015.
- [8] Y. Bai, J. Kang, and P.X.-K. Song. Efficient pairwise composite likelihood estimation for spatial-clustered data. *Biometrics*, 70(3):661–670, 2014.
- [9] M. Banna, F. Merlevède, and P. Youssef. Bernstein-type inequality for a class of dependent random matrices. *Random Matrices: Theory Appl.*, 5, 2016. 28pp.
- [10] Y. Baraud. Estimation of the density of a determinantal process. *Confluentes Math.*, 5(1):3–21, 2013.
- [11] R. Bardenet and A. Hardy. Monte carlo with determinantal point processes. *arXiv:1605.00361*, 2016. preprint.
- [12] R. Bardenet and M. Titsias RC AUEB. Inference for determinantal point processes without spectral knowledge. In *Advances in Neural Information Processing Systems 28*, pages 3393–3401. Curran Associates, Inc., 2015.
- [13] H.C.P. Berbee. *Random walks with stationary increments and renewal theory*. Mathematisch Centrum, 1979.
- [14] C.A.N. Biscio and F. Lavancier. Brillinger mixing of determinantal point processes and statistical applications. *Electron. J. Statist.*, 10:582–607, 2016.

- [15] C.A.N. Biscio and F. Lavancier. Quantifying repulsiveness of determinantal point processes. *Bernoulli*, 22:2001–2028, 2016.
- [16] C.A.N. Biscio and F. Lavancier. Contrast estimation for parametric stationary determinantal point processes. *Scandinavian Journal of Statistics*, 44:204–229, 2017.
- [17] C.A.N. Biscio, A. Poinas, and R. Waagepetersen. A note on gaps in proofs of central limit theorems. *Statistics and Probability Letters*, 135:7–10, 2018.
- [18] B. Błaszczyszyn and D. Yogeshwaran. Clustering and percolation of point processes. *Electron. J. Probab.*, 18:20 pp., 2013.
- [19] 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.
- [20] E. Bolthausen. On the central limit theorem for stationary mixing random fields. *The Annals of Probability*, 10:1047–1050, 1982.
- [21] A. Borodin, A. Okounkov, and G. Olshanski. Asymptotics of plancherel measures for symmetric groups. *J Amer Math Soc*, 13, 06 1999.
- [22] D. Bosq. *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*. Springer Science & Business Media, 2012.
- [23] R.C. Bradley. Basic properties of strong mixing conditions. a survey and some open questions. *Probability Surveys*, 2:107–144, 2005.
- [24] V.-E. Brunel, A. Moitra, P. Rigollet, and J. Urschel. Maximum likelihood estimation of determinantal point processes. *arXiv:1701.06501*, 2017. preprint.
- [25] A. Bulinski and E. Shabanovich. Asymptotic behaviour of some functionals of positively and negatively dependent random fields. *Fundam. Prikl. Mat.*, 4:479–492, 1998.
- [26] A. Bulinski and A. Shashkin. *Limit Theorems for Associated Random Fields and Related Systems*. Springer-Verlag, 2007.
- [27] A. Bulinski and E. Spodarev. *Central Limit Theorems for Weakly Dependent Random Fields*, pages 337–383. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
- [28] R. Burton and R. Pemantle. Local characteristics, entropy and limit theorems for spanning trees and domino tilings via transfer-impedances. *Ann. Probab.*, 21, 9 1999.
- [29] R. Burton and E. Waymire. Scaling limits for associated random measures. *Ann. Appl. Probab.*, 13:1267–1278, 1985.
- [30] J. Coeurjolly and J. Møller. Variational approach for spatial point process intensity estimation. *Bernoulli*, 20:1097–1125, 2014.
- [31] 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.

- [32] Y. Davydov. On convergence of distributions which are generated by stationary random processes. *Teor. Veroyatnost. i Primenen.*, 13:730–737, 1968.
- [33] J. Dedecker. *Weak Dependence: With Examples and Applications*. Springer Science & Business Media, 2007.
- [34] C. Dellacherie and P. Meyer. *Probabilities and Potential*. Elsevier Science Ltd, 1978.
- [35] C. Deng, Y. Guan, R. Waagepetersen, and J. Zhang. Second-order quasi-likelihood for spatial point processes. *Biometrics*, 73(4):1311–1320, 2017.
- [36] 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.
- [37] P. Doukhan. *Mixing: Properties and Examples*. Springer-Verlag, 1994.
- [38] 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.
- [39] P. Doukhan and S. Louhichi. A new weak dependence condition and applications to moment inequalities. *Stochastic Process. Appl.*, 84:313–342, 1999.
- [40] G. M. Engel and H. Schneider. Matrices diagonally similar to a symmetric matrix. *Linear Algebra and its Applications*, 29:131–138, 1980.
- [41] J.D. Esary, F. Proschan, and D.W. Walkup. Association of random variables, with applications. *Ann. Math. Statist.*, 38:1466–1474, 1967.
- [42] S.N. Evans. Association and random measures. *Probability Theory and Related Fields*, 86:1–19, 1990.
- [43] H. Föllmer. A covariance estimate for gibbs measures. *Journal of Functional Analysis*, 46:387–395, 1982.
- [44] H. Georgii and H. Yoo. Conditional intensity and gibbsianness of determinantal point processes. *Journal of Statistical Physics*, 118(1):55–84, Jan 2005.
- [45] S. Ghosh. Determinantal processes and completeness of random exponentials: the critical case. *Probability Theory and Related Fields*, 163(3):643–665, 2015.
- [46] J. Ginibre. Statistical ensembles of complex, quaternion, and real matrices. *J. Mathematical Phys.*, 4:440–449, 1965.
- [47] 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.
- [48] Y. Guan. A composite likelihood approach in fitting spatial point process models. *Journal of the American Statistical Association*, 101:1502–1512, 2006.
- [49] Y. Guan and J. M. Loh. A thinned block bootstrap procedure for modeling inhomogeneous spatial point patterns. *Journal of the American Statistical Association*, 102:1377–1386, 2007.
- [50] X. Guyon. *Random Fields on a Network*. Springer-Verlag, 1994.

- [51] P.J. Heagerty and S.R. Lele. A composite likelihood approach to binary spatial data. *Journal of the American Statistical Association*, 93(443):1099–1111, 1998.
- [52] L. Heinrich. Normal approximation for some mean-value estimates of absolutely regular tessellations. *Math. Methods of Statistics*, 3:1–24, 1994.
- [53] L. Heinrich. On the strong Brillinger-mixing property of α -determinantal point processes and some applications. *Appl. Math.*, 61(4):443–461, 2016.
- [54] 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.
- [55] L. Heinrich, S. Lück, and V. Schmidt. Asymptotic goodness-of-fit tests for the palm mark distribution of stationary point processes with correlated marks. *Bernoulli*, 20:1673–1697, 2014.
- [56] L. Heinrich and E. Liescher. Strong convergence of kernel estimators for product densities of absolutely regular point processes. *J. Nonparametric Statistics*, 8:65–96, 1997.
- [57] L. Heinrich and I. Molchanov. Central limit theorem for a class of random measures associated with germ-grain models. *Adv. Appl. Prob.*, 31:283–314, 1999.
- [58] L. Heinrich and Z. Pawlas. Absolute regularity and brillinger-mixing of stationary point processes. *Lithuanian Math. J.*, 53:293–310, 2013.
- [59] C.C. Heyde. *Quasi-likelihood and its application: a general approach to optimal parameter estimation*. Springer Science & Business Media, 1997.
- [60] Johnson C.R. Horn R.A. *Matrix Analysis*. CUP, 2ed. edition, 2012.
- [61] J.B. Hough, M. Krishnapur, Y. Peres, and B. Virag. *Zeros of Gaussian Analytic Functions and Determinantal Point Processes*. American Mathematical Society, 2009.
- [62] I.A. Ibragimov and Y.V. Linnik. *Independant and stationnary sequences of random variables*. Wolters-Noordhoff, 1971.
- [63] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan. *Statistical analysis and modelling of spatial point patterns*. Statistics in Practice. Wiley, New York, 2008.
- [64] J. Jacod and M. Sørensen. A review of asymptotic theory of estimating functions. *Statistical Inference for Stochastic Processes*, pages 1–20, 2017.
- [65] J.L. Jensen. Asymptotic normality of estimates in spatial point processes. *Scandinavian Journal of Statistics*, 20:97–109, 1993.
- [66] E. Jolivet. Central limit theorem and convergence of empirical processes for stationary point processes. In *Point processes and queuing problems (Colloquium, Keszthely, 1978)*, volume 24 of *Colloq. Math. Soc. János Bolyai*, pages 117–161. North-Holland, Amsterdam-New York, 1981.
- [67] S. Karlin and J. McGregor. Coincidence probabilities. *Pacific J. Math.*, 9(4):1141–1164, 1959.

- [68] A. Kulesza and B. Taskar. Determinantal point processes for machine learning. *Foundations and Trends in Machine Learning*, 5(2-3):123–286, 2012.
- [69] 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.
- [70] F. Lavancier and P. Rochet. A general procedure to combine estimators. *Computational Statistics & Data Analysis*, 94:175–192, 2016.
- [71] F. Lavancier and P. Rochet. A tutorial on estimator averaging in spatial point process models. *Journal de la SFDS*, 158(3):106–123, 2017.
- [72] V. Loonis and X. Mary. Determinantal sampling designs. *J. Statist. Plann. Inference*, 199:60–88, 2019.
- [73] R. Lyons. Determinantal probability measures. *Publications Mathématiques de l’IHÉS*, 98:167–212, 2003.
- [74] R. Lyons. Determinantal probability: Basic properties and conjectures. *Proceedings of the International Congress of Mathematicians, Seoul, Korea, IV*:137–161, 2014.
- [75] O. Macchi. The coincidence approach to stochastic point processes. *Advances in Applied Probability*, 7:83–122, 1975.
- [76] N. Miyoshi and T. Shirai. A cellular network model with ginibre configured base stations. *Advances in Applied Probability*, 46:832–845, 2014.
- [77] J. Møller and R.P. Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, 2004.
- [78] J. Møller and R.P. Waagepetersen. Modern statistics for spatial point processes. *Scandinavian Journal of Statistics*, 34:643–684, 2007.
- [79] J. Møller and R.P. Waagepetersen. Some recent developments in statistics for spatial point patterns. *Annual Review of Statistics and Its Application*, 4(1):317–342, 2017.
- [80] B. Nahapetian. *Limit Theorems and Some Applications in Statistical Physics*, volume 123 of *Teubner-Texte zur Mathematik*. Teubner Verlag, 1991.
- [81] R. Pemantle and Y. Peres. Concentration of lipschitz functionals of determinantal and other strong rayleigh measures. *Combin. Probab. Comput.*, 23:140–160, 2014.
- [82] Y. Peres and B. Virág. Zeros of the i.i.d. gaussian power series: a conformally invariant determinantal process. *Acta Math.*, 194(1):1–35, 2005.
- [83] A. Poinas. A β -mixing inequality for point processes induced by their intensity functions. *Statistics and Probability Letters*, 148:88–93, 2019.
- [84] A. Poinas, B. Delyon, and F. Lavancier. Mixing properties and central limit theorem for associated point processes. *Bernoulli*, 25(3):1724–1754, 2019.
- [85] M. Prokešová, J. Dvořák, and E.B.V. Jensen. Two-step estimation procedures for inhomogeneous shot-noise Cox processes. *Annals of the Institute of Statistical Mathematics*, 2016.

- [86] M. Prokešová and E.B.V Jensen. Asymptotic palm likelihood theory for stationary point processes. *Ann Inst Stat Math*, 65:387–412, 2013.
- [87] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017.
- [88] E. Rio. Covariance inequalities for strongly mixing processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 29(4):587–597, 1993.
- [89] E. Rio. *Asymptotic Theory of Weakly Dependent Random Processes*. Springer Berlin Heidelberg, 2017.
- [90] J. Rising, A. Kulesza, and B. Taskar. An efficient algorithm for the symmetric principal minor assignment problem. *Linear Algebra and its Applications*, 473:126–144, 2015.
- [91] M. Rosenblatt. A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U.S.A.*, 42:43–47, 1956.
- [92] Y.A. Rozanov and V.A. Volkonskii. Some limit theorems for random functions I. *Theory Probab. Appl.*, 4:178–197, 1959.
- [93] F. P. Schoenberg. Consistent parametric estimation of the intensity of a spatial-temporal point process. *Journal of Statistical Planning and Inference*, 128:79–93, 2005.
- [94] 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.
- [95] J. Snow. *On the Mode of Communication of Cholera*. John Churchill, 1855.
- [96] M. Sørensen. On asymptotics of estimating functions. *Brazilian Journal of Probability and Statistics*, 13(2):111–136, 1999.
- [97] A. Soshnikov. Determinantal random point fields. *Russian Math. Surveys*, 55:923–975, 2000.
- [98] D.J. Strauss. A model for clustering. *Biometrika*, 62:467–475, 1975.
- [99] U. Tanaka, Y. Ogata, and D. Stoyan. Parameter estimation and model selection for Neyman-Scott point processes. *Biometrical Journal*, 50(1):43–57, 2008.
- [100] M. Van Lieshout. *Markov Point Processes And Their Applications*. World Scientific Publishing Company, 2000.
- [101] G. Viennet. Inequalities for absolutely regular sequences: application to density estimation. *Probab. Theory Relat. Fields*, 107:467–492, 1997.
- [102] R. Waagepetersen. An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics*, 63:252–258, 2007.
- [103] R. Waagepetersen and Y. Guan. Two-step estimation for inhomogeneous spatial point processes. *Journal of the Royal Statistical Society*, 71:685–702, 2009.
- [104] 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.

Titre : Statistiques asymptotiques des processus ponctuels déterminantaux stationnaires et non stationnaires

Mot clés : Processus ponctuel déterminantal, Théorème limite central, Statistiques spatiales

Resumé : Ce manuscrit est dédié à l'étude de l'estimation paramétrique d'une famille de processus ponctuels appelée processus déterminantaux. Ces processus sont utilisés afin de générer et modéliser des configurations de points possédant de la dépendance négative, dans le sens où les points ont tendance à se repousser entre eux. Dans une première partie, nous montrons un théorème limite central pour une classe générale de statistiques sur les processus déterminantaux. Dans une seconde partie, nous montrons une inégalité de bêta-mélange générale pour les processus ponctuels. Dans une troisième partie, nous appliquons le théorème limite central obtenu à la première partie à une classe générale de fonctions estimantes basées sur des méthodes de moments. Finalement, dans la dernière partie, nous donnons une approximation asymptotique de la log-vraisemblance des processus déterminantaux qui est calculable numériquement et nous étudions la consistance de son maximum.

Title : Asymptotic inference of stationary and non-stationary determinantal point processes

Keywords : Determinantal point process, Central limit theorem, Spatial statistics

Abstract : This manuscript is devoted to the study of parametric estimation of a point process family called determinantal point processes. These point processes are used to generate and model point patterns with negative dependency, meaning that the points tend to repel each other. In the first chapter, we prove a central limit theorem for a wide class of statistics on determinantal point processes. In the second chapter, we show a general beta-mixing inequality for point processes. In the third chapter, we apply the central limit theorem showed in the first chapter to a wide class of moment-based estimating functions. Finally, in the last chapter, we give an asymptotic approximation of the log-likelihood that is computationally tractable and we study the consistency of its maximum.

Résumé étendu de la thèse

Introduction

Un des sujets d'étude principaux des statistiques spatiales est l'étude de jeux de données constitués d'un ensemble fini de points dans une fenêtre d'observation qui est un compact de \mathbb{R}^d (en général, $d = 1, 2$ ou 3). L'exemple le plus ancien et probablement le plus connu est l'étude par le médecin John Snow [8] de la répartition spatiale de cas de choléra lors d'une épidémie à Londres ce qui lui a permis de remonter à une pompe à eau contaminée qui en était la source. En pratique, de nombreux jeux de données ponctuel spatiaux sont observés dans toutes les disciplines scientifiques. Quelques exemples sont donnés dans la figure 1. Le modèle mathématique utilisé pour modéliser ce type de donnée est appelé un processus ponctuel.

La façon la plus simple de décrire un processus ponctuel sur \mathbb{R}^d est en tant que mesure de probabilité sur l'ensemble Ω des ensembles localement finis de \mathbb{R}^d , où un sous-ensemble X de \mathbb{R}^d est dit localement fini si pour tout sous-ensemble borné A de \mathbb{R}^d l'ensemble $X \cap A$ est fini. Le processus ponctuel le plus connu et le plus étudié est le processus ponctuel de Poisson (ou PPP). Un PPP X est défini par une fonction ρ localement intégrable appelée "intensité" telle que le nombre $N(A)$ de points du PPP observés dans un sous-ensemble A de \mathbb{R}^d suit une loi de poisson de paramètre $\int_A \rho(x) dx$ et, pour tout ensembles disjoints $A_1, \dots, A_n \subset \mathbb{R}^d$, les variables aléatoires $N(A_1), \dots, N(A_n)$ sont mutuellement indépendantes. En conséquence, conditionnellement à $N(A)$, les points de $X \cap A$ sont mutuellement indépendant et identiquement distribués pour tout $A \subset \mathbb{R}^d$. C'est pourquoi les PPPs sont utilisés pour modéliser tout jeu de données spatiales constitué de points qui peuvent être considérés comme étant indépendant les uns des autres.

Évidemment, les PPPs sont limités par le fait que tous les points n'ont aucune interaction les uns avec les autres alors qu'en pratique de nombreux jeux de données présentent de la dépendance positive (attraction) ou négative (répulsion) entre leurs points. Une simulation d'un PPP est présentée en haut à droite de la Figure 2. En comparaison, divers exemples de données ponctuels exhibant un comportement répulsif sont présentés dans la Figure 1. C'est pourquoi des modèles de processus ponctuels plus complexes ont été développés afin de pouvoir étudier ce genre de donnée. Dans cette thèse, nous nous concentrons sur l'étude de processus ponctuels répulsifs. Les principaux modèles existants sont les processus ponctuels de Matérn (aussi appelé "hardcore"), les grilles perturbées, les processus ponctuels de Gibbs et, le sujet principal de cette thèse, les processus déterminantaux (où DPPs).

Les processus ponctuels de Matérn sont défini à partir des PPPs en retirant des points de sorte qu'aucune paire de point ne soit à distance plus petite qu'une constante R (appelé le rayon hardcore) alors que les grilles perturbées sont définies en déplaçant de façon aléatoire les points d'une grille régulière par des variables aléatoires iid. Ces deux modèles sont faciles à simuler et assez simple à manipuler mais trop restrictifs pour la plupart des applications statistiques. Au contraire, les processus ponctuels de Gibbs forment une famille très générale de processus ponctuels qui offrent une importante gamme possible d'interactions entre les points et ont été beaucoup utilisés pour modéliser de la répulsion mais sont en général difficiles à manipuler. Par exemple, ils ne peuvent être simulés qu'à partir de méthodes MCMC, leur densité fait intervenir une constante impossible à calculer facilement ce qui rend difficile une estimation par maximum de vraisemblance et aucune

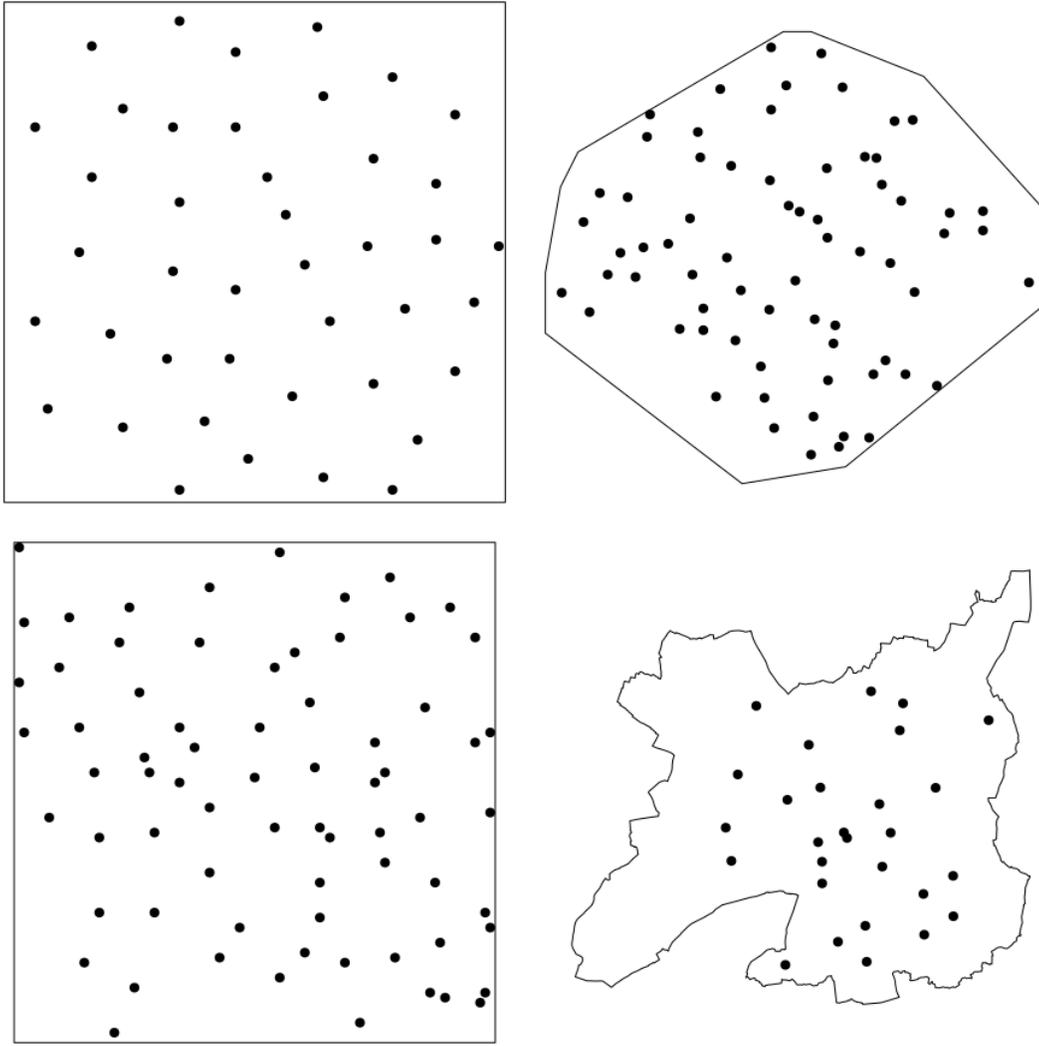


Figure 1: Exemples de données ponctuelles spatiales. La figure en haut à gauche correspond à l'emplacement du centre de 42 cellules biologiques dans une coupe histologique. La figure en haut à droite correspond à l'emplacement de 68 nids de fourmis moissonneuses sur un site en Grèce. La figure en bas à gauche correspond à l'emplacement de 71 plants de pin dans une forêt suédoise. La figure en bas à droite correspond à l'emplacement des 28 bureaux de vote dans la commune de Rennes. Les trois premiers jeux de données font partie du package *spatstat.data* [1] du logiciel R [7]. Le dernier jeu de données est fourni par le site internet <https://data.rennesmetropole.fr>.

expression de forme fermée n'est connue pour leurs moments. Pour plus de détails, voir [6, 10].

D'un autre côté, les DPPs sont une famille plus restrictive que les processus ponctuels de Gibbs [4] mais sont plus faciles à manipuler tout en étant une famille très flexible qui offre une large gamme d'interactions possibles entre les points. Comparé aux processus de Gibbs, ils peuvent être simulés de façon exacte et possèdent une expression de forme fermée pour leurs moments et leurs densités de Janossy. Cela permet d'utiliser des méthodes de

moment ainsi que le maximum de vraisemblance pour leur estimation. C'est pourquoi les DPPs se présentent comme une bonne alternative pour modéliser les jeux de données ponctuelles possédant de la répulsion. On trouvera un exemple de simulation des DPPs, des processus ponctuels de Gibbs et des processus ponctuels de Matérn dans la figure 2.

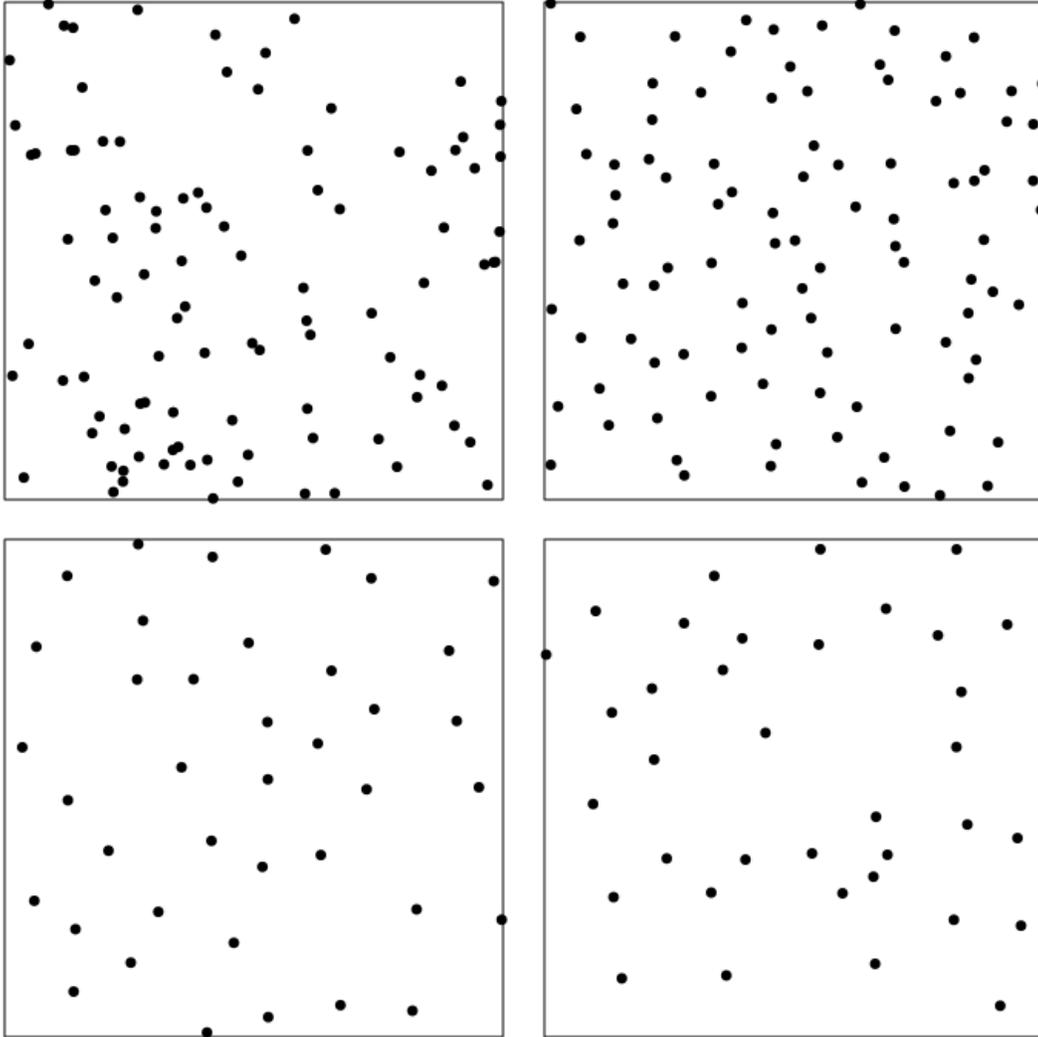


Figure 2: Exemples de simulation de 4 différents types de processus ponctuels sur la fenêtre $W = [0, 1]^2$. En haut à gauche se trouve une simulation d'un PPP homogène d'intensité $\rho = 100$. En haut à droite se trouve une simulation d'un DPP de noyau de type Bessel de paramètres $\rho = 100$ et $\alpha = 0.05$. En bas à gauche se trouve une simulation d'un processus ponctuel de Matérn de rayon hardcore $R = 0.1$. En bas à droite se trouve une simulation d'un processus ponctuel de Strauss (un cas particulier d'un processus ponctuel de Gibbs défini dans [9]) de paramètres $\beta = 100$, $\gamma = 0.2$ et $R = 0.1$. Toutes les simulations ont été réalisées sur R [7] avec le package *spatstat* [2].

Nous rappelons qu'un DPP continu est défini par ses fonctions d'intensité jointe que l'on note ρ_n . Si on considère une fonction $K : (\mathbb{R}^d)^2 \rightarrow \mathbb{C}$ satisfaisant l'hypothèse

\mathcal{H} : K est une fonction continue, localement de carré intégrable et hermitienne telle que son opérateur intégral associé sur $L^2(\mathbb{R}^d, \mu)$

$$\mathcal{K} : f \mapsto \left(\mathcal{K}f : x \mapsto \int_{\mathbb{R}^d} K(x, y) f(y) d\mu(y) \right)$$

est localement de trace classe à valeurs propres dans $[0, 1]$.

alors un processus ponctuel X est dit DPP de noyau K s'il vérifie

$$\rho_n(x_1, \dots, x_n) = \det ((K(x_i, x_j))_{1 \leq i, j \leq n})$$

pour tout entier n .

Dans cette thèse, nous nous concentrons sur l'étude des statistiques asymptotiques de DPPs continus dans un cadre spécifique appelé "increasing domain asymptotic". Considérons $(W_n)_{n \geq 0}$ une suite croissante de sous-ensemble de \mathbb{R}^d , c'est-à-dire que $W_n \subset W_{n+1}$ pour tout entier n . À la différence du cadre iid, nous considérons l'observation d'une unique réalisation du processus ponctuel dans la fenêtre W_n . Dans ce cas-là, les propriétés asymptotiques dépendent de la taille de la fenêtre et donc indirectement du nombre de points observés. C'est pour cela que l'on suppose en général que $\bigcup_{n \geq 0} W_n = \mathbb{R}^d$ où, plus simplement, que le volume d -dimensionnel de W_n tend vers l'infini. Dans la plupart des applications, les fenêtres sont considérées comme étant rectangulaires mais, afin d'obtenir des résultats aussi généraux que possibles, nous considérons que les fenêtres peuvent avoir n'importe quelle forme du tant que leur frontière ne soit pas trop distordue afin de limiter les effets de bord.

La difficulté principale de ce cadre d'étude pour les DPPs est que comme tous les points générés par un DPP se repoussent les uns les autres, alors il y a besoin de contrôler leur dépendance spatiale pour obtenir des résultats asymptotiques. Dans ce but, nous étudions les propriétés de mélange des DPPs. Nous utilisons ces résultats pour établir un théorème limite central (TLC) pour une large famille de statistiques sur les DPPs stationnaires et non-stationnaires. Nous appliquons ensuite ce TLC pour obtenir des résultats asymptotiques sur les équations estimantes basées sur des méthodes de moment. Enfin, nous considérons le problème de prouver la consistance et de trouver une bonne approximation du maximum de vraisemblance d'un DPP stationnaire. Nous allons maintenant détailler les divers résultats de la thèse.

Chapitre 2

Nous commençons par montrer que la propriété d'association (positive ou négative) d'un processus ponctuel implique une inégalité de covariance générale qui ne dépend que de la covariance entre le nombre de points du processus ponctuel qui se trouvent dans des ensembles disjoints:

Theorem 1. *Soit X un processus ponctuel associé et $A, B \subset \mathbb{R}^d$ deux ensembles bornés. Soit $f : \Omega \rightarrow \mathbb{R}$ et $g : \Omega \rightarrow \mathbb{R}$ deux fonctions telles que $X \mapsto f(X \cap A)$ et $X \mapsto g(X \cap B)$ soient bornés, alors*

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

où $\|\cdot\|_A$ est la semi-norme de Lipschitz définie par

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

De plus, si X est positivement associé alors il vérifie la même inégalité avec $A, B \subset \mathbb{R}^d$ pas forcément disjoint.

En conséquence, on en déduit que l'association implique une borne sur les coefficients d' α -mélange qui ne dépend que des deux premières fonctions d'intensité jointe du processus ponctuel.

Proposition 2. *Soit X un processus ponctuel associé sur \mathbb{R}^d dont les deux premières fonctions d'intensité jointe soient bien définies, alors pour tout $p, q > 0$,*

$$\begin{cases} \alpha_{p,q}(r) \leq pq \sup_{\|x-y\| \geq r} |\rho_2(x,y) - \rho(x)\rho(y)|, \\ \alpha_{p,\infty}(r) \leq ps_d \int_r^\infty t^{d-1} \sup_{\|x-y\|=t} |\rho_2(x,y) - \rho(x)\rho(y)| dt. \end{cases}$$

où s_d est l'aire de la sphère d -dimensionnelle.

En conséquence, la vitesse de décroissance des coefficients de α -mélange dans chaque cas ne dépend que de la répulsion des paires de points du processus ponctuel. Dans le cas des DPPs, puisque $|\rho_2(x,y) - \rho(x)\rho(y)| = |K(x,y)|^2$ alors on en déduit que les coefficients $\alpha_{p,q}(r)$ décroissent au moins à la même vitesse que le noyau du DPP au carré lorsque $\|y-x\|$ tend vers l'infini. Nous prouvons également que cette vitesse de décroissance est optimale pour une large classe de DPPs en exhibant une borne inférieure pour les coefficients d' α -mélange avec la même vitesse de décroissance. Malheureusement, les vitesses de décroissance des coefficients $\alpha_{p,\infty}(r)$ ne sont pas aussi bonnes et si on devait appliquer le TLC de Bolthausen et Guyon [3, 5] pour les variables aléatoires α -mélangeante, cela nous forcerait d'exclure les DPPs les plus répulsifs qui sont ceux avec les noyaux qui décroissent le moins vite. Afin de parer à ce problème, nous prouvons un TLC plus général basé non pas sur les coefficients de α -mélange mais sur l'inégalité de covariance du Théorème 1. Le TLC obtenu peut s'appliquer à toutes les familles paramétriques classiques de noyaux de DPPs communément utilisés en statistiques spatiales.

Theorem 3. *Soit X un DPP de noyau borné K satisfaisant \mathcal{H} . Soit $\tau > 0$ et $f : \Omega \rightarrow \mathbb{R}$ une fonction de la forme*

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

où f_0 est une fonction bornée qui s'annule sur l'ensemble $\sup_{x,y \in S} \|y-x\| > \tau$. Soit $(W_n)_{n \in \mathbb{N}}$ une suite croissante de sous-ensembles de \mathbb{R}^d telle que $|W_n| \rightarrow \infty$ et soit $\sigma_n^2 := \text{Var}(f(X \cap W_n))$. Supposons qu'il existe $\varepsilon > 0$ et $\nu > 0$ tels que les conditions suivantes soient satisfaites:

$$(H1) \quad |\partial W_n \oplus (\tau + \nu)| = o(|W_n|);$$

$$(H2) \quad \sup_{\|y-x\| > r} |K(x,y)| = o(r^{-(d+\varepsilon)/2});$$

(H3) $\liminf_n |W_n|^{-1} \sigma_n^2 > 0$.

Alors,

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

Ici, $\partial W_n \oplus (\tau + \nu)$ correspond à l'ensemble des points à distance inférieure à $\tau + \nu$ de la frontière de W_n .

Chapitre 3

Dans ce chapitre, nous prouvons une inégalité sur les coefficients de β -mélange d'un processus ponctuel qui ne dépend que de ses fonctions d'intensité jointe.

Theorem 4. Soit X un processus ponctuel sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu)$ tel que $\mathbb{E}[4^{|X \cap A|}] < +\infty$ pour toute ensemble borné $A \subset \mathbb{R}^d$. Alors, pour tout $p, q, r \in \mathbb{R}_+$,

$$\beta_{p,q}(r) \leq \sup_{\substack{\mu(A) < p, \mu(B) < q \\ \text{dist}(A,B) > r}} \left(\sum_{m,n=0}^{+\infty} \frac{2^{n+m-1}}{m!n!} \int_{A^m \times B^n} |\rho_m(x) \rho_n(y) - \rho_{m+n}(x,y)| d\mu^m(x) d\mu^n(y) \right).$$

Comme les identités $(\rho_m(x) \rho_n(y) - \rho_{m+n}(x,y)) / \rho_m(x) \rho_n(y)$ peuvent s'interpréter comme la corrélation entre X ayant un point en x_1, \dots, x_m et X ayant des points en y_1, \dots, y_n alors Théorème 4 prouve que plus ces corrélations décroissent vite lorsque la distance entre les x_i et y_j tend vers l'infini, plus les coefficients de β -mélange décroissent rapidement. Lorsque ce résultat est appliqué aux DPPs, on obtient l'inégalité de β -mélange suivante:

Theorem 5. Soit X un DPP de noyau borné K satisfaisant \mathcal{H} , alors

$$\beta_{p,q}(r) \leq 2pq(1 + 2p\|K\|_\infty)(1 + 2q\|K\|_\infty)e^{2\|K\|_\infty(p+q)} \sup_{\|y-x\| \geq r} |K(x,y)|^2.$$

Il est à noter que la vitesse de décroissance des coefficients de β -mélange dans le Théorème 5 est la même que celle des coefficients de α -mélange dans la Proposition 2. Le gros défaut de ce résultat, comparé à l'inégalité sur les coefficients de α -mélange obtenue dans la Proposition 2, est la dépendance exponentielle de la borne sur les $\beta_{p,q}(r)$ par rapport à p et q ce qui la rend inutilisable dans la plupart des applications.

Chapitre 4

Dans le chapitre 4, nous nous intéressons aux propriétés asymptotiques des estimateurs construits à partir de fonctions estimantes. Considérons une famille de processus ponctuels continus $\{\mathbb{P}^\theta, \theta \in \Theta\}$, où $\Theta \subset \mathbb{R}^p$ pour un certain $p \in \mathbb{N}$, ainsi qu'une suite croissante de fenêtres d'observation $(W_n)_{n \geq 0}$, on dit que la fonction e_n est une fonction estimante lorsque θ est estimé par la solution (où l'une des solutions) $\hat{\theta}_n$ de l'équation $e_n(\theta_n) = 0$. C'est un cadre statistique très général qui englobe d'autres méthodes telles que l'estimation par minimum de contraste (lorsque le contraste est dérivable) où l'estimation par maximum de vraisemblance (lorsque la vraisemblance est dérivable). En particulier, on dit que e_n est une fonction estimante d'ordre k si elle s'écrit sous la forme

$$e_n(X) = \sum_{x_1, \dots, x_k \in X \cap W_n}^{\neq} f_0(x_1, \dots, x_k; \theta) - \int_{W_n^k} f_0(t; \theta) \rho_k(t; \theta) d\mu(t),$$

où $f_0 : (\mathbb{R}^d)^k \times \Theta \rightarrow \mathbb{R}^p$ est une fonction test. Dans ce chapitre, on s'intéresse à un cadre plus général où l'on combine des fonctions estimantes d'ordre différents.

$$e_n(\theta) = \begin{pmatrix} \sum_{x_1, \dots, x_{q_1} \in X \cap W_n}^{\neq} f_1(x_1, \dots, x_{q_1}; \theta) - \int_{W_n^{q_1}} f_1(x; \theta) \rho^{(q_1)}(x; \theta) d\mu(x) \\ \vdots \\ \sum_{x_1, \dots, x_{q_l} \in X \cap W_n}^{\neq} f_l(x_1, \dots, x_{q_l}; \theta) - \int_{W_n^{q_l}} f_l(x; \theta) \rho^{(q_l)}(x; \theta) d\mu(x) \end{pmatrix}.$$

où chaque $f_i : (\mathbb{R}^d)^{q_i} \times \Theta \rightarrow \mathbb{R}^{k_i}$ est une fonction test et $\sum_i k_i = p$. Nous montrons que sous certaines hypothèses de régularité sur les fenêtres d'observation, les fonctions tests et le processus ponctuel alors il existe une suite $\mu(W_n)$ -consistante de racines $\hat{\theta}_n$ de l'équation $e_n(\theta) = 0$ qui sont asymptotiquement normaux. Plus précisément, pour tout $\varepsilon > 0$, il existe $A > 0$ tel que

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

pour un n assez grand, et

$$\mu(W_n) \text{Var}(e_n(\theta^*))^{-1/2} H_n(\theta^*)(\hat{\theta}_n - \theta^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I_p),$$

où I_p est la matrice identité de taille $p \times p$ et

$$H_n(\theta^*) := \frac{1}{\mu(W_n)} \begin{pmatrix} \int_{W_n^{q_1}} f_1(x; \theta^*) \nabla_{\theta} \rho^{(q_1)}(x; \theta^*)^T d\mu(x) \\ \vdots \\ \int_{W_n^{q_l}} f_l(x; \theta^*) \nabla_{\theta} \rho^{(q_l)}(x; \theta^*)^T d\mu(x) \end{pmatrix}.$$

Chapitre 5

Dans le dernier chapitre de cette thèse, nous nous concentrons sur le problème de l'estimation par maximum de vraisemblance de DPPs stationnaires. Si on considère une famille $\{\mathbb{P}^\theta, \theta \in \Theta\}$ de DPPs continus observée sur une suite croissante de fenêtres d'observation compactes $(W_n)_{n \geq 1}$, alors la log-vraisemblance (renormalisée) de la famille de DPPs s'écrit

$$l_n(\theta|X) = 1 + \frac{1}{\mu(W_n)} (\log \det(Id - \mathcal{K}_{W_n}^\theta) + \log \det(L_{W_n}^\theta[X \cap W_n])).$$

où $L_{W_n}^\theta$ et le déterminant de Fredholm de $Id - \mathcal{K}_{W_n}^\theta$ sont deux éléments dont le calcul fait intervenir la décomposition spectrale de la projection de \mathcal{K} sur $L^2(W_n, \mu)$. Ce n'est pas possible numériquement pour les DPPs continus car cette décomposition n'est jamais explicitement connue. Nous proposons dans ce chapitre une approximation asymptotique de la vraisemblance de DPPs continus stationnaires basé sur le fait que ces DPPs ont un noyau de la forme $K(x, y) = K_0(y - x)$ et lorsque K_0 est de carré intégrable, il admet une transformée de Fourier \hat{K}_0 prenant à valeurs dans $[0, 1]$. Nous montrons alors les approximations suivantes:

$$L_{W_n}(x, y) \approx \int_{\mathbb{R}^d} \frac{\hat{K}_0(t)}{1 - \hat{K}_0(t)} \exp(2i\pi t \cdot (y - x)) d\mu(t)$$

et

$$\frac{1}{\mu(W_n)} \log \det (Id - \mathcal{K}_{W_n}^\theta) \approx \int_{\mathbb{R}^d} \log(1 - \hat{K}_0(x)) d\mu(x),$$

afin d'obtenir une approximation $\tilde{l}_n(\theta|X)$ de $l_n(\theta|X)$ qui ne dépend pas de la décomposition spectrale des noyaux mais seulement de leur transformée de Fourier qui est connue explicitement pour toutes les familles classiques de DPPs stationnaires utilisées en statistiques spatiales.

Le défaut principal d'une approximation asymptotique de la vraisemblance est qu'elle ignore toute sorte d'effet de bord du DPP ce qui est problématique pour les DPPs les plus répulsifs qui subissent les effets de bord les plus forts. Afin de remédier à ce problème, nous nous intéressons à l'idée de remplacer cette fenêtre par un tore plat afin de rapprocher les points qui se trouvent sur des bords opposés de la fenêtre dans le but de mitiger ces effets de bord. Nous montrons par diverses simulations que cette approximation du maximum de vraisemblance avec la correction des effets de bord permet d'obtenir de meilleurs résultats que les méthodes d'inférence plus classiques telles que l'estimation par minimum de contraste basée sur le fonction K de Ripley ou la fonction de corrélation par paires.

References

- [1] A. Baddeley, R. Turner, and E. Rubak. *spatstat.data: Datasets for 'spatstat'*, 2018. R package version 1.4-0.
- [2] A. J. Baddeley, E. Rubak, and R. Turner. *Spatial Point Patterns: Methodology and Applications with R*. Interdisciplinary Statistics. Chapman & Hall/CRC, Boca Raton, Florida, 2015.
- [3] E. Bolthausen. On the central limit theorem for stationary mixing random fields. *The Annals of Probability*, 10:1047–1050, 1982.
- [4] H. Georgii and H. Yoo. Conditional intensity and gibbsianness of determinantal point processes. *Journal of Statistical Physics*, 118(1):55–84, Jan 2005.
- [5] X. Guyon. *Random Fields on a Network*. Springer-Verlag, 1994.
- [6] J. Møller and R.P. Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, 2004.
- [7] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017.
- [8] J. Snow. *On the Mode of Communication of Cholera*. John Churchill, 1855.
- [9] D.J. Strauss. A model for clustering. *Biometrika*, 62:467–475, 1975.
- [10] M. Van Lieshout. *Markov Point Processes And Their Applications*. World Scientific Publishing Company, 2000.