ESTIMATION OF THE DENSITY OF A DETERMINANTAL PROCESS

YANNICK BARAUD

ABSTRACT. We consider the problem of estimating the density Π of a determinantal process N from the observation of n independent copies of it. We use an aggregation procedure based on robust testing to build our estimator. We establish non-asymptotic risk bounds with respect to the Hellinger loss and deduce, when n goes to infinity, uniform rates of convergence over classes of densities Π of interest.

1. INTRODUCTION

The starting point of this work goes back to 2007 when Persi Diaconis visited our Laboratory Jean-Alexandre Dieudonné in Nice. At that time, he explained that determinantal processes were emerging in many areas and that there was no statistical procedure to estimate their distributions. Almost five years later, it still seems to be the case. The aim of this paper is therefore to contribute to the study of these processes. Our aim is not only to focus on statistical estimation but also to discuss some related problems. For example, how the class $\mathcal{D}$ of all determinantal densities can be parametrized? Is there an identifiable way of doing it? Another natural question, at least for a Statistician, is to understand how the elements of $\mathcal{D}$ can be approximated. Are there some specific parametric sets that should be used to approximate the densities lying in $\mathcal{D}$? If so, what can be said about the approximation properties of these sets? Finally, given n independent copies of a determinantal process N, we propose an estimator of the density Π of N. We establish non-asymptotic risk bounds for our estimator and deduce uniform rates of convergence over classes of Π of interest. It turns out that our estimation strategy is robust with respect to the assumption that N is a determinantal process. This means that the risk bounds we get are not only valid when Π belongs to the class $\mathcal{D}$ but also when Π is close enough to it (in the Hellinger distance). Our approach is based on T-estimation as introduced by Birgé (2006). More precisely, we start with a suitable family of models, which typically consists of compact sets of densities, and the role of which is to provide a good approximation of the elements of $\mathcal{D}$. Then, we discretize these models.

This results in a family of points $(\Pi_{\mathfrak{m}})_{\mathfrak{m}\in\mathfrak{M}}$ of $\mathcal{D}$ and we finally use the data in order to select a suitable point among the $\Pi_{\mathfrak{m}}$. The way we select this point, which provides our estimator of Π, is based on robust testing and aims at finding an element among the $\Pi_{\mathfrak{m}}$ which is as close as possible to the target density Π. We establish non-asymptotic risk bounds for our estimator and show how they depend on the approximation properties of the models we started from. Under *a posteriori* assumptions on Π and for a suitable choice of the models, we specify this bounds and derive rates of convergence.

For an introduction to determinantal processes, we refer the interested reader to Lyons (2003), Hough *et al* (2006) and the books by Anderson *et al* (2010) and Hough *et al* (2009) as well as the references therein. Part of the popularity of determinantal processes comes from the fact that they naturally arise in the study of the eigenvalues of large random matrices. Recently, Borodin *et al* (2010) showed that these processes are also involved in the process of "caries" when adding a column of numbers.

The paper is organized as follows. In Section 2 we settle the probabilistic background as well as our main notations and conventions. We introduce determinantal processes in Section 3 and tackle the problem of estimating of their densities in Section 4. Finally, Section 5 is devoted to the proofs.

2. The background

2.1. **Notations and conventions.** Throughout this paper we use the conventions $\sum_{\varnothing} = 0$ and $\prod_{\varnothing} = 1$ and set $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$ and $\mathbb{R}_+^* = (0, +\infty)$. Given a finite set A, $|A|$ denotes the cardinality of A and for $z \in \mathbb{C}$, $\Re(z)$, $\overline{z}$ and $|z|$ denote the real part, conjugate and modulus of z respectively. We denote by $\mathcal{P}$ the class of all finite subset J of $\mathbb{N}^*$ and set $\mathcal{P}^* = \mathcal{P} \setminus \{\varnothing\}$. Moreover, we set

$$\Lambda = \{\lambda \in [0,1]^{\mathbb{N}^*}, \ \ |\lambda|^2 = \sum_{j \geq 1} \lambda_j^2 < +\infty\}.$$

All along, we consider a metric space $(\mathcal{X}, d)$ which we endow with its Borel σ-field $\mathcal{B}(\mathcal{X})$ and a σ-finite measure μ. Roughly speaking, a point process on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ will correspond to a random choice of a family of distinct points among $\mathcal{X}$. One should typically think of $\mathcal{X}$ as $\{1, \ldots, p\}$, $\mathbb{N}$, $\mathbb{R}$ or $\mathbb{R}^p$ for some positive integer p. When $\mathcal{X}$ is not finite, we denote by $\mathbb{H}$ the Hilbert space of measurable and complex-valued functions ϕ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ satisfying

$$\|\phi\|^2 = \int_{\mathcal{X}} |\phi|^2 \, d\mu < +\infty.$$

We endow $\mathbb{H}$ with the Hermitian inner product defined for $\phi, \psi \in \mathbb{H}$ by

$$\langle \phi, \psi \rangle = \int_{\mathcal{X}} \overline{\phi} \psi d\mu.$$

For conveniency, we adopt the convention that $\langle ., . \rangle$ is linear with respect to the second argument and not the first one, as usually the case. In order to keep our

notation as simple as possible, when $\mathcal{X}$ is finite, say $\mathcal{X} = \{1, \ldots, p\}$, we embed $\mathcal{X}$ into $\mathbb{N}^*$ and use

$$\mathbb{H} = \ell_2(\mathbb{N}^*) = \{\phi \in \mathbb{C}^{\mathbb{N}^*}, \; \sum_{i \geq 1} |\phi(i)|^2 < +\infty\}.$$

More precisely, a mapping ϕ on $\mathcal{X} = \{1, \ldots, p\}$ with values in $\mathbb{C}$ will be viewed as a sequence $(\phi(i))_{i \geq 1} \in \ell_2(\mathbb{N}^*)$ with $\phi(i) = 0$ for all $i > p$. Since $\mathbb{H}$ is an infinite dimensional Hilbert space (whatever $\mathcal{X}$), we may define $\mathbb{U}$ as the set of all orthonormal sequences $\Phi = (\phi_j)_{j \geq 1}$ in $\mathbb{H}$. For $J \in \mathcal{P}^*$ and $\Phi \in \mathbb{U}$, we set $\Phi_J = (\phi_j)_{j \in J}$ and given an (ordered) finite subset α of $\mathcal{X}$, denote by $\Phi_{\alpha,J}$ the $|\alpha| \times |J|$-matrix

$$\Phi_{\alpha,J} = (\phi_j(x))_{x \in \alpha, j \in J}.$$

We extend this notation for rectangle matrices A with entries in $\mathbb{C}$: $A_{\alpha,J} = (A_{i,j})_{i \in \alpha, j \in J}$. Moreover, A^* denotes the transpose of the conjugate of A, that is, if $A = (A_{i,j})_{i=1,\ldots,k, j=1,\ldots,k'}$, $A^* = (\overline{A}_{j,i})_{j=1,\ldots,k', i=1,\ldots,k}$.

Finally, we recall that the Hellinger distance h between two densities p, q on a measured space $(E, \mathcal{E}, \nu)$ is defined by the formula

$$h^2(p, q) = \frac{1}{2} \int_E (\sqrt{p} - \sqrt{q})^2 \, d\nu.$$

For the sake of simplicity, we shall keep the same notation h throughout this paper even though the measured space $(E, \mathcal{E}, \mu)$ may vary.

2.2. **The probabilistic background.** In this section, our aim is to introduce the probabilistic background we shall use throughout this paper. We denote by $\mathbb{X}$ the class of all finite subsets of $\mathcal{X}$ and for $k \in \mathbb{N}$, denote by $\mathbb{X}_k$ the class of those subsets with cardinality k. By convention, $\mathbb{X}_0 = \{\varnothing\}$. We identify $\mathbb{X}$ with the set of finite measures of the form

$$\overline{\alpha} = \sum_{x \in \alpha} \delta_x \text{ with } \alpha \in \mathbb{X}$$

and denote the same way α and $\overline{\alpha}$ so that for all $B \in \mathcal{B}$, $\alpha(B)$ means $|\alpha \cap B|$. We equip $\mathbb{X}$ with the smallest σ-field $\mathcal{B}(\mathbb{X})$ for which the mappings

$$\mathbb{X} \rightarrow \mathbb{N}$$
$$M_B : \alpha \mapsto \alpha(B)$$

are measurable for all $B \in \mathcal{B}(\mathcal{X})$. In particular, the subsets $\mathbb{X}_k = M_{\mathcal{X}}^{-1}(k)$ are measurable for all $k \in \mathbb{N}$. We endow $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ with the measure L defined for all measurable functions f from $\mathbb{X}$ into $\mathbb{R}_+$ by

$$\int_{\mathbb{X}} f(\alpha) dL(\alpha) = f(\varnothing) + \sum_{k \geq 1} \frac{1}{k!} \int_{\mathcal{X}^{\vee k}} f(\{x_1, \ldots, x_k\}) d\mu(x_1) \ldots d\mu(x_k)$$

where $\mathcal{X}^{\vee k}$ is the set of all k-uplets $(x_1, \ldots, x_k) \in \mathcal{X}^k$ with distinct coordinates. If $\mathcal{X}$ is finite, say $\mathcal{X} = \{1, \ldots, p\}$, and if μ is the counting measure on $\mathcal{X}$ then L is merely the counting measure on $\mathbb{X}$.

Throughout this paper, a point process N on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is a random variable defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in $(\mathbb{X}, \mathcal{B}(\mathbb{X}), L)$.

3. INTRODUCTION TO DETERMINANTAL PROCESSES

In this section, our aim is to define a determinantal process on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. To do so, we adopt the point of view developed in Hough *et al* (2006). In particular, we start with the simpler case of *determinantal projection processes*.

3.1. **Determinantal projection processes.**

Definition 1. *Given $J \in \mathcal{P}^*$ and $\Phi \in \mathbb{U}$, a determinantal projection process N of rank $|J|$ with parameter $\Phi_J = (\phi_j)_{j \in J}$ is a point process with density (with respect to L) given by*

(3.1) $$\Pi_J^\Phi(\alpha) = |\det[\Phi_{\alpha, J}]|^2 \, \mathbb{1}_{\mathbb{X}_{|J|}}(\alpha) \ \text{ for all } \alpha \in \mathbb{X}.$$

When $J = \varnothing$, by convention $\Pi_\varnothing^\Phi = \delta_\varnothing$.

If the matrix $\Phi_{\alpha, J} = (\phi_j(x))_{x \in \alpha, j \in J}$ depends on an ordering on the set α, $|\det[\Phi_{\alpha, .}]|$ does not and we shall therefore omit to specify one. It follows from the definition of Π_J^Φ that with probability 1, $|N(\mathcal{X})| = |J|$. Hence, a determinantal projection process N of rank $|J|$ consists of $|J|$ distinct points of $\mathcal{X}$. The location of these points depends on the geometry of the ϕ_j for $j \in J$. If the ϕ_j are real-valued, a configuration $\alpha = \{x_1, \ldots, x_k\}$ of points is all the more likely that the volume of the parallelepiped based on the $|J|$ vectors $((\phi_j(x_1), \ldots, \phi_j(x_k))_{j \in J}$ is large.

The fact that Π_J^Φ is a density on $\mathbb{X}$ might not be clear at first sight. In fact, when $\mathcal{X} = \{1, \ldots, p\}$ this comes the the Cauchy-Binet formula: if A, B are $k \times p$ and $p \times k$ matrices respectively with $p \geq k$, the Cauchy-Binet formula asserts that

(3.2) $$\det[AB] = \sum_{\alpha \in \mathbb{X}_k} \det A_{\{1, \ldots, k\}, \alpha} \det B_{\alpha, \{1, \ldots, k\}}.$$

Again, note that this formula is independent of the choice of an ordering on α. By using the Cauchy-Binet formula with $B = (\phi_j(x))_{x \in \mathcal{X}, j \in J} = \Phi_{\mathcal{X}, J}$, $A = B^*$ and by using the fact the family $(\phi_j)_{j \in J}$ is orthonormal we get

$$\int_{\mathbb{X}_k} \Pi_J^\Phi(\alpha) dL(\alpha) = \sum_{\alpha \in \mathbb{X}_k} \overline{\det[\Phi_{\alpha, J}]} \det[\Phi_{\alpha, J}] = \sum_{\alpha \in \mathbb{X}_k} \det[\Phi_{J, \alpha}^*] \det[\Phi_{\alpha, J}]$$

$$= \det[\Phi_{J, \mathcal{X}}^* \Phi_{\mathcal{X}, J}] = \det[\langle \phi_i, \phi_j \rangle_{i, j \in J}] = 1.$$

When $\mathcal{X}$ is no longer finite, the Cauchy-Binet formula can be extended by using the identity below from which we can deduce in a similar way as above that Π_J^Φ is a density.

Proposition 1. *Let $\Phi = (\phi_1, \ldots, \phi_k)$ and $\Psi = (\psi_1, \ldots, \psi_k)$ be two elements of $\mathbb{H}^k$. We have that*

$$\det \left[(\langle \phi_i, \psi_j \rangle)_{i,j=1,\ldots,k} \right] = \int_{\mathbb{X}_k} \det \left[\Phi^*_{\{1,\ldots,k\},\alpha} \right] \det \left[\Psi_{\alpha,\{1,\ldots,k\}} \right] dL(\alpha).$$

This identity is known for a long time, especially when $\mathcal{X}$ is a compact interval of $\mathbb{R}$ and μ the Lebesgue measure on $\mathcal{X}$ (see de Bruijn (1955)). The general form of this identity can be found in Baik and Rains (2001).

3.2. **The general case.** As proved in Hough *et al* (2006), the distribution of a (finite) determinantal process N can be viewed as a mixture of densities of some determinantal projection processes. More precisely, a determinantal process can be defined as follows.

Definition 2. *Let $\Phi \in \mathbb{U}$ and $\lambda \in \Lambda$. A determinantal process N with parameters (Φ, λ) is a point process with density*

(3.3) $$\Pi^{\Phi,\lambda} = \sum_{J \in \mathcal{P}} p_J^\lambda \Pi_J^\Phi \quad \text{where} \quad p_J^\lambda = \prod_{j \in J} \lambda_j^2 \prod_{j \notin J} (1 - \lambda_j^2) \text{ for all } J \in \mathcal{P}.$$

We use the convention $\Pi_\varnothing^\Phi = \delta_\varnothing$.

Since $\lambda_j \in [0,1]$ for all $j \geq 1$ and

(3.4) $$\sum_{j \geq 1} \lambda_j^2 < +\infty,$$

the numbers p_J^λ are nonnegative and well defined (the infinite product $\prod_{j \notin J}(1 - \lambda_j^2)$ converges for all $J \in \mathcal{P}$). Besides,

$$\sum_{J \in \mathcal{P}} p_J^\lambda = \prod_{j \geq 1}(\lambda_j^2 + (1 - \lambda_j^2)) = 1.$$

Consequently, $\Pi^{\Phi,\lambda}$ is indeed an (at most countable) mixture of densities. Given $J \in \mathcal{P}^*$, it is not difficult to see that for the particular choice $\lambda = \lambda_J = (\mathbb{1}_{j \in J})_{j \geq 1}$, the density $\Pi^{\Phi,\lambda}$ is that of a determinantal projection process with parameter Φ_J: indeed, for $J' = J$, $p_{J'}^\lambda = 1$ and for $J' \neq J$, $p_{J'}^\lambda = 0$.

As explained in Hough *et al* (2006), another way of defining a determinantal process is as follows. First simulate a sequence $(Z_j)_{j \geq 1}$ of independent Bernoulli random variables with respective parameters $(\lambda_j^2)_{j \geq 1}$. Consider the subset $\widehat{J}$ of those indices $j \geq 1$ such that $Z_j = 1$. Finally choose N according to a determinantal projection process of rank $|\widehat{J}|$ with parameter $\Phi_{\widehat{J}}$. With such a description, Condition (3.4) is easy to understand: together with the Borel-Cantelli lemma, it ensures that $\widehat{J}$ is finite almost surely. It is also clear that the distribution of a determinantal process remains unchanged if we change the labelling of the pairs $((\lambda_j, \phi_j))_{j \geq 1}$. That is, for all bijection σ on $\mathbb{N}^*$, the parameters $((\phi_j)_{j \geq 1}, (\lambda_j)_{j \geq 1})$ and $((\phi_{\sigma(j)})_{j \geq 1}, (\lambda_{\sigma(j)})_{j \geq 1})$ lead to the same determinantal

distribution. In particular, with no loss of generality, we may assume that the sequence $\lambda = (\lambda_j)_{j \geq 1}$ is non-increasing with respect to j.

In the literature, one usually associates to a determinantal process N a square integrable kernel K on $\mathcal{X}^2$ which defines a self-adjoint compact operator on $\mathbb{H}$ by the formula

$$(3.5) \qquad T_K : \phi \quad \begin{array}{c} \mathbb{H} \to \mathbb{H} \\ \to \left[x \mapsto \int_{\mathcal{X}} K(x,y)\phi(y)d\mu(y) \right]. \end{array}$$

The sequences $(\lambda_j^2)_{j \geq 1}$ and $\Phi = (\phi_j)_{j \geq 1}$ mentioned above correspond then to the eigenvalues and associated eigenvectors of T_K. Conversely, given the sequences $\lambda = (\lambda_j)_{j \geq 1}$ and $\Phi = (\phi_j)_{j \geq 1}$ and provided that $\mu(\mathcal{X}) < +\infty$, the kernel K can be obtained by the fomula (Mercer's Theorem)

$$(3.6) \qquad K(x,y) = \sum_{j \geq 1} \lambda_j^2 \phi_j(x) \overline{\phi}_j(y)$$

where the series converge absolutely for almost every $(x,y) \in \mathcal{X}^2$. When $\mathcal{X} = \{1,\dots,p\}$, K is merely (any) $p \times p$ Hermitian matrix with eigenvalues in $[0,1]$. Interestingly, the kernel K can be related to the distribution of N by the following formula which holds for all measurable functions f from $\mathbb{X}$ into $\mathbb{R}_+$

$$\mathbb{E} \left[\sum_{\alpha \subset N} f(\alpha) \right] = \int_{\mathbb{X}} f(\alpha) \det[K_{\alpha,\alpha}] dL(\alpha) \quad \text{where} \quad K_{\alpha,\alpha} = (K(x,y))_{x \in \alpha, y \in \alpha}.$$

The mapping $\alpha \mapsto \det[K_{\alpha,\alpha}]$ determines the distribution of N and is called the *correlation function*. When $\mathcal{X} = \{1,\dots,p\}$, this formula simply says that for all $\alpha \subset \mathcal{X}$

$$\mathbb{P}\left[\alpha \subset N\right] = \det[K_{\alpha,\alpha}].$$

3.3. **Hellinger distance and determinantal process.** In the previous section, we have seen that the distribution of a determinantal process can be parametrized by a pair (Φ, λ) in $\mathbb{U} \times \Lambda$ and that, conversely, any choice of such a pair allows to define a determinantal process. The aim of this section is to relate the Hellinger distance between the distributions of two determinantal processes associated to two distinct pairs (Φ, λ) and (Ψ, γ) to some distance between these pairs. Again, we start with the simpler case of a determinantal projection process.

3.3.1. *Case of a determinantal projection process.* Let $\Phi = (\phi_j)_{j \geq 1}$ and $\Psi = (\psi_j)_{j \geq 1}$ be two elements of $\mathbb{U}$ and J, J' two elements of $\mathcal{P}$. If $|J| \neq |J'|$, the supports of Π_J^Φ and $\Pi_{J'}^\Psi$ are disjoint (the densities are supported by $\mathbb{X}_{|J|}$ and $\mathbb{X}_{|J'|}$ respectively) and hence $h^2(\Pi_J^\Phi, \Pi_{J'}^\Psi) = 1$. If $J = J' = \varnothing$, $\Pi_J^\Phi = \Pi_{J'}^\Psi$ and therefore $h^2(\Pi_J^\Phi, \Pi_{J'}^\Psi) = 0$. Consequently, the only case we need to consider is the one where $|J| = |J'| \geq 1$. In fact, as already mentioned, we may re-index one of the two sequences, say Ψ, in order to have $J = J'$ without changing the distribution Π^Ψ. By doing so, the following result holds.

Proposition 2. *Let $J \in \mathcal{P}^*$ and $\Phi, \Psi \in \mathbb{U}$. We have,*

$$(3.7) \qquad h^2(\Pi_J^\Phi, \Pi_J^\Psi) \;=\; 1 - \int_{\mathbb{X}_k} \left|\det\left[\Phi_{\alpha,J}\right]\right| \left|\det\left[\Psi_{\alpha,J}\right]\right| dL(\alpha)$$

$$\leq\; 1 - \left|\det\left[\left(\langle \phi_i, \psi_j \rangle\right)_{i,j \in J}\right]\right|.$$

Moreover,

$$h^2(\Pi_J^\Phi, \Pi_J^\Psi) \leq \frac{5}{2} \sum_{j \in J} \|\phi_j - \psi_j\|^2.$$

Up to constants, the last inequality is sharp. For example, if $J = \{1\}$, $\Pi_{\{1\}}^\Phi$ and $\Pi_{\{1\}}^\Psi$ correspond to the two densities on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ given by

$$(3.8) \qquad \Pi_{\{1\}}^\Phi(x) = |\phi_1(x)|^2 \quad \text{and} \quad \Pi_{\{1\}}^\Psi(x) = |\psi_1(x)|^2 \quad \text{for all} \quad x \in \mathcal{X}$$

and hence, if ϕ_1, ψ_1 are two nonnegative real-valued functions on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$,

$$h^2\left(\Pi_{\{1\}}^\Phi, \Pi_{\{1\}}^\Psi\right) = \frac{1}{2} \int_{\mathcal{X}} \left(\sqrt{\Pi_{\{1\}}^\Phi} - \sqrt{\Pi_{\{1\}}^\Psi}\right)^2 d\mu = \frac{1}{2} \|\phi_1 - \psi_1\|^2.$$

Clearly, this equality is no longer true when the nonnegativity assumption on ϕ_1 and ψ_1 is violated. Nevertheless, Proposition 2 says that the inequality remains true (up to a constant). The proof of this proposition is postponed to Section 5.1.

3.3.2. *The general case.* Since $\Pi^{\Phi,\lambda}$ and $\Pi^{\Psi,\gamma}$ are mixtures, the problem of bounding the Hellinger distance between these two densities amounts to understanding how, more generally, the Hellinger distance behaves with respect to mixtures of densities. More precisely, let p, q be two densities on the measured space $(T, \mathcal{T}, m)$ and $(P_t)_{t \in T}$ and $(Q_t)_{t \in T}$ two families of densities on a measured space $(E, \mathcal{A}, \nu)$. What can we say about the Hellinger distance between the two mixtures

$$P = \int_T P_t p(t) dm(t) \quad \text{and} \quad Q = \int_T Q_t q(t) dm(t)$$

when we known how far p is from q and the P_t from the Q_t? The following result gives an answer.

Proposition 3. *If m and ν are both σ-finite,*

$$h^2(P, Q) \leq 2h^2(p, q) + 2 \int_T h^2(P_t, Q_t) q(t) dm(t).$$

The proof of this result is postponed to Section 5.2.

We may apply Proposition 3 with the choices $T = \mathcal{P}$ (m being the counting measure on $\mathcal{P}$), $(E, \mathcal{A}, \nu) = (\mathbb{X}, \mathcal{B}(\mathbb{X}), L)$, $p = p^\lambda$ the density defined on $\mathcal{P}$ by $p^\lambda(J) = p_J^\lambda$ for all $J \in \mathcal{P}$, $q = p^\gamma$ defined analogously and for $t = J \in \mathcal{P}$, $P_t = \Pi_J^\Phi$ and $Q_t = \Pi_J^\Psi$. We obtain the following result the proof of which is detailed in Section 5.3.

Proposition 4. *Let* $\Phi, \Psi \in \mathbb{U}$ *and* $\lambda, \gamma \in \Lambda$ *and set*

$$\check{\lambda} = \left(1 - \sqrt{1 - \lambda_j^2}\right)_{j \geq 1} \quad \text{and} \quad \check{\gamma} = \left(1 - \sqrt{1 - \gamma_j^2}\right)_{j \geq 1}.$$

The following inequalities hold

(3.9)
$$h^2(p^\lambda, p^\gamma) \ \leq \ |\lambda - \gamma|^2 + \left|\check{\lambda} - \check{\gamma}\right|^2$$

(3.10)
$$\sum_{J \in \mathcal{P}} p_J^\gamma h^2(\Pi_J^\Phi, \Pi_J^\Psi) \ \leq \ \frac{5}{2} \sum_{j \geq 1} \gamma_j^2 \, \|\phi_j - \psi_j\|^2 .$$

In particular,

(3.11)
$$h^2(\Pi^{\Phi,\lambda}, \Pi^{\Psi,\gamma}) \leq 2 \left[|\lambda - \gamma|^2 + \left|\check{\lambda} - \check{\gamma}\right|^2\right] + 5 \sum_{j \geq 1} \gamma_j^2 \, \|\phi_j - \psi_j\|^2 .$$

4. Statistical estimation

Throughout this section, we consider a point process N on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ with density Π with respect to L. Given n independent copies $N_1, \ldots, N_n$ of N, our aim is to estimate Π. One may naturally think of N as being a determinantal process which means that Π belongs to the set $\mathcal{D}$ of all determinantal distributions. Nevertheless, our result is robust with respect to such an assumption in the sense that Π may not belong to $\mathcal{D}$. In this case, one may rather consider $\mathcal{D}$ as an approximation set for Π. Before turning to the estimation of Π, we shall first discuss some identifiability issues which are of independent interest and may therefore be skipped.

4.1. **Identifiability and exterior algebra.** When N is a determinantal process, we may write $\Pi = \Pi^{\Phi,\lambda}$ for some pair $(\Phi, \lambda) \in \mathbb{U} \times \Lambda$ or, alternatively, define Π from some kernel K on $\mathcal{X}^2$ as in (3.6). If these two approaches provide a parametrization of $\mathcal{D}$, none is identifiable. More precisely, two distinct pairs in $\mathbb{U} \times \Lambda$ or two distinct kernels may parametrize the same determinantal distribution. This lack of identifiability is already true if one restricts to the simpler class of determinantal projection processes. A simple counter-example can be obtained from (3.8) with $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mu) = ([0,1], \mathcal{B}([0,1], dx)$ by taking $\phi_1(x) = e^{ix}$ and $\psi_1(x) = e^{2ix}$. In this case, the corresponding kernels $K_1(x,y) = e^{i(x-y)}$ and $K_2(x,y) = e^{2i(x-y)}$ are distinct but both parametrize the uniform distribution on $[0,1]$. It is also clear from this counter-example that there is no hope to estimate ϕ_1, which is not identifiable either.

Consequently, a question arises. How can we define a one-to-one parametrization of $\mathcal{D}$? As we shall see, this problem is rather difficult. In fact, we shall partially answer this question by restricting ourself to the case where $\mathcal{X} = \{1, \ldots, p\}$ and by focusing on the class $\mathcal{D}_{p,k}$ of all determinantal projection distributions of rank k with $k \in \{1, \ldots, p-1\}$ and $p \geq 2$. It follows from Definition 1 that for

each element $\Pi \in \mathcal{D}_{p,k}$, there exists an orthonormal family $\phi_1, \ldots, \phi_k$ (which is certainly not unique) such that for all $\alpha \in \mathbb{X}_k$

(4.1) $$\Pi(\alpha) = \left| \det \left[\Phi_{\alpha, \{1, \ldots, k\}} \right] \right|^2 .$$

Let us now consider the exterior algebra $E = \bigwedge^k \mathbb{C}^p$ consisting of the sums of k-blades $\phi_1 \wedge \ldots \wedge \phi_k$ with $\phi_1, \ldots, \phi_k \in \mathbb{C}^p$. Since we shall only use the algebraic properties of these objects and more specifically their connections with determinants, we shall not define them and rather refer the interested reader to Mac Lane and Birkhoff (1988) (Chapter XVI, Section 7). Denoting by $e_1, \ldots, e_p$ the canonical basis of $\mathbb{C}^p$, this exterior algebra E can be viewed as a $\mathbb{C}$-linear space, a basis of which being given by the k-blades of the form

$$e_\alpha = e_{i_1} \wedge \ldots \wedge e_{i_k}$$

where $\alpha = (i_1, \ldots, i_k)$ (with $1 \le i_1 < i_2 \ldots < i_k \le p$) varies among $\mathbb{X}_k$. This linear space can be equipped with an Hermitian inner product $[.,.]$ for which the elements $(e_\alpha)_{\alpha \in \mathbb{X}_k}$ provide an orthonormal family of E. Besides, for a k-blade $\phi_1 \wedge \ldots \wedge \phi_k$

(4.2) $$[e_\alpha, \phi_1 \wedge \ldots \wedge \phi_k] = \det \left[\Phi_{\alpha, \{1, \ldots, k\}} \right] \quad \text{for all} \quad \alpha \in \mathbb{X}_k.$$

Let $\mathcal{S}_E$ be the unit sphere of $(E, [.,.])$, G the subset of $\mathcal{S}_E$ gathering the elements of the form $\phi_1 \wedge \ldots \wedge \phi_k$ for $\phi_1, \ldots, \phi_k$ being an orthonormal family of $\mathbb{C}^p$ and G_+ be the subset of $\mathcal{S}_E$ defined by

$$G_+ = \{ g_+ = \sum_{\alpha \in \mathbb{X}_k} |[e_\alpha, g]| \, e_\alpha \mid g \in G \}.$$

It follows from (4.1) and (4.2) that the mapping

$$
\begin{aligned}
G_+ &\to \mathcal{D}_{p,k} \\
g_+ &\mapsto \Pi_{g+} : \alpha \mapsto |[e_\alpha, g]|^2 = |[e_\alpha, g_+]|^2
\end{aligned}
$$

is surjective. It is also clearly one-to-one and provides thus an identifiable parametrization of the elements of $\mathcal{D}_{p,k}$ by those of G_+. In fact, if Δ denotes the Hermitian distance on E defined for $g, g' \in E$ by $\Delta^2(g, g') = [g - g', g - g']$, (G_+, Δ) and $(\mathcal{D}_{p,k}, \sqrt{2}h)$ are isometric: by (3.7), for all $g_+, g'_+ \in G_+$,

$$
\begin{aligned}
\Delta^2(g_+, g'_+) &= \sum_{\alpha \in \mathbb{X}_k} \left| [g_+, e_\alpha] - [g'_+, e_\alpha] \right|^2 = 2 \left[1 - \sum_{\alpha \in \mathbb{X}_k} |[g, e_\alpha]| \, |[g', e_\alpha]| \right] \\
&= 2h^2(\Pi_{g+}, \Pi_{g'+}).
\end{aligned}
$$

The metric dimension (in the sense given in Birgé (2006)) of a set of densities is usually closely related to the minimax rate of estimation over this set. Roughly speaking, if the metric dimension of the set is D, one can expect that the minimax rate be of order D/n. The above isometry shows that the metric dimension $D_{p,k}$ of $(\mathcal{D}_{p,k}, h)$ is the same as that of the subset $G_+ \subset E$ for the Hermitian distance. In particular $D_{p,k}$ is not larger than the dimension of E (in the usual sense, viewed as a linear space on $\mathbb{R}$), that is $D_{p,k} \le 2\binom{p}{k}$. This upper

bound is unfortunately very crude and we shall see that the minimax rates can be much faster. We believe that the metric dimension of G_+ is actually of order kp.

4.2. **The main result.** Let us now turn to the statistical part of this paper. As already mentioned, our aim is to estimate the density Π of a point process N from the observation of n independent copies of it. Our estimation strategy is based on T-estimation. More precisely, we start with an at most countable family $\{\Pi_\mathfrak{m},\ \mathfrak{m} \in \mathfrak{M}\}$ of candidate determinantal densities, the choice of which will be explained below, and we use a test possessing robustness properties in view of selecting the closest element to Π among the $\Pi_\mathfrak{m}$. We shall not detail the statistical procedure here and rather refer the reader to Birgé (2006) (Theorem 9) or Baraud (2011). Nevertheless, in order to give an brief account of the estimation strategies described there, let us merely say that they allow to endow $\{\Pi_\mathfrak{m},\ \mathfrak{m} \in \mathfrak{M}\}$ with a (random) binary relation $\propto$ by means of a statistical test based on the observations. Given a pair $(\Pi_\mathfrak{m}, \Pi_{\mathfrak{m}'})$ of distinct candidate densities, we either have $\Pi_\mathfrak{m} \propto \Pi_{\mathfrak{m}'}$ or $\Pi_{\mathfrak{m}'} \propto \Pi_\mathfrak{m}$, the test being build in such way that the former (respectively the latter) relation is likely to occur when $h(\Pi, \Pi_\mathfrak{m})$ is small compared to $h(\Pi, \Pi_{\mathfrak{m}'})$ and vice-versa. If the relation $\propto$ were a total order on $\{\Pi_\mathfrak{m},\ \mathfrak{m} \in \mathfrak{M}\}$, a natural idea would be to define the estimator of Π as the minimal element of $(\{\Pi_\mathfrak{m},\ \mathfrak{m} \in \mathfrak{M}\}, \propto)$. Unfortunately, this is not the case since $\propto$ fails to be transitive in general. A nice idea, which is actually due to Birgé, is to define the estimator as the element $\widehat{\Pi}$ of $\{\Pi_\mathfrak{m},\ \mathfrak{m} \in \mathfrak{M}\}$ minimizing the quantity $\Pi_\mathfrak{m} \mapsto \mathrm{crit}(\Pi_\mathfrak{m}) = \sup\{h(\Pi_\mathfrak{m}, \Pi_{\mathfrak{m}'}), \Pi_{\mathfrak{m}'} \propto \Pi_\mathfrak{m}\}$. The property of the test ensures that the value of criterion at $\Pi_\mathfrak{m}$ is likely to be large when $\Pi_\mathfrak{m}$ is far from Π (provided that there exist some elements $\Pi_{\mathfrak{m}'}$ which are closer to Π) while it is likely to be small as soon as $\Pi_\mathfrak{m}$ lies in a small enough neighborhood of Π. With such an estimation strategy, we can design an estimator possessing the following property.

Proposition 5. *Let* $\mathbf{\Pi} = \{\Pi_\mathfrak{m},\ \mathfrak{m} \in \mathfrak{M}\}$ *be an at most countable family of densities on* $(\mathbb{X}, \mathcal{B}(\mathbb{X}), L)$ *and* π *a sub-probability on* $\mathfrak{M}$*, that is*

$$\sum_{\mathfrak{m} \in \mathfrak{M}} \pi(\mathfrak{m}) \leq 1 \ \text{ and } \ \pi(\mathfrak{m}) \geq 0 \ \text{ for all } \ m \in \mathfrak{M}.$$

There exist a universal constant $C > 0$ *and an estimator* $\widehat{\Pi} = \widehat{\Pi}(\mathbf{\Pi}, \pi)$ *solely based on* $N_1, \ldots, N_n$ *such that whatever the density* Π*,*

$$C\mathbb{E}\left[h^2(\Pi, \widehat{\Pi})\right] \leq \inf_{\mathfrak{m} \in \mathfrak{M}} \left[h^2(\Pi, \Pi_\mathfrak{m}) + \frac{\log(1/\pi(\mathfrak{m}))}{n}\right].$$

Proof. Proposition 3 page 363 of Baraud (2011) (Example 1, Density Estimation) ensures that the random measure $n^{-1} \sum_{i=1}^n \delta_{N_i}$ with intensity Π satisfies the assumption of Corollary 5 page 373 of Baraud (2011). The result follows by applying this corollary with $s = \Pi$, $\Lambda = \mathfrak{M}$, $s_\lambda = \Pi_\mathfrak{m}$ and $\overline{\Delta}(\Pi_\mathfrak{m}) = \log(1/\pi(\mathfrak{m}))$

for all $\mathfrak{m} \in \mathfrak{M}$, the summation condition (4) of Baraud (2011) being satisfied under the assumption that π is a sub-probability. $\square$

Before turning to the choice of the family $\{\Pi_{\mathfrak{m}}, \; \mathfrak{m} \in \mathfrak{M}\}$, let us comment on the role of π in our result. When π is a probability, it can be interpreted as a prior on the family $\{\Pi_{\mathfrak{m}}, \; \mathfrak{m} \in \mathfrak{M}\}$ and gives thus a bayesian flavor to our approach. Intuitively, our procedure tends to advantage densities $\Pi_{\mathfrak{m}}$ associated to values of $\pi(\mathfrak{m})$ which are not too small.

We design our family $\{\Pi_{\mathfrak{m}}, \; \mathfrak{m} \in \mathfrak{M}\}$ in view of possessing good approximation properties with respect to the elements of the class $\mathcal{D}$. Inequality (3.11) tells us that one can approximate a determinantal density $\Pi^{\Phi,\lambda}$ (with respect to the Hellinger distance) by suitably approximating the sequence λ and the functions ϕ_j of Φ corresponding to those indices j for which λ_j is large enough. To do so, we introduce compacts subsets of Λ and $\mathbb{H}$ respectively defined as follows. Concerning $\lambda = (\lambda_j)_{j \geq 1}$, with no loss of generality, we may assume the sequence is non-increasing with respect to j and it is therefore natural to introduce compact sets of the form

$$\Lambda_j = \{\gamma \in \Lambda, \gamma_{j'} = 0 \text{ for all } j' > j\}$$

for different values of $j \geq 1$. This amounts to approximating λ by the truncated sequence keeping the j first entries of λ, the others being turned to 0. In order to approximate the ϕ_j, we introduce an at most countable family $\mathcal{H} = (H_m)_{m \in \mathcal{M}}$ of compact subsets of $(\mathbb{H}, \| \; \|)$. Examples of such compacts sets will be given in Section 4.3 for the purpose of providing rates of convergence. Given a compact subset H of $(\mathbb{H}, \| \; \|)$ and some positive number η, we denote by $H[\eta]$ a maximal η-separated subset of H, that is, any subset $H' \subset H$ of maximal cardinality satisfying the property: for all $\phi, \phi' \in H'$ with $\phi \neq \phi'$, $\|\phi - \phi'\| > \eta$. The maximality of $H[\eta]$ implies that for all $\phi \in H$, there exists $\phi' \in H[\eta]$ such that $\|\phi - \phi'\| \leq \eta$. This means that $H[\eta]$ is an η-net for the compact set H. By applying Proposition 5 to a suitable discretization of the compact sets Λ_j and H_m, we deduce the result below. Its proof is detailed in Section 5.4.

Theorem 1. *Let $\mathcal{H} = (H_m)_{m \in \mathcal{M}}$ be an at most countable families of compact subsets H_m of $(\mathbb{H}, \| \; \|)$ and let π be a sub-probability on $\mathcal{M}$. There exists a density estimator $\widehat{\Pi}$ such that whatever the density Π on $(\mathbb{X}, \mathcal{B}(\mathbb{X}), L)$,*

$$C\mathbb{E}\left[h^2(\Pi, \widehat{\Pi})\right] \leq \inf_{\Phi \in \mathbb{U}, \lambda \in \Lambda} \left[h^2(\Pi, \Pi^{\Phi,\lambda}) + \inf_{j \geq 1}\left(\sum_{j'=1}^{j} O(\mathcal{H}, \pi, \phi_{j'}) + \sum_{j' > j} \lambda_{j'}^2\right)\right]$$

where for all $j \geq 1$,

$$O(\mathcal{H}, \pi, \phi_j) = \inf_{m \in \mathcal{M}}\left[\inf_{\psi \in H_m} \|\phi_j - \psi\|^2 + \frac{1}{n}\log\left(\frac{|H_m[1/\sqrt{n}]|\, n}{\pi(m)}\right)\right]$$

and C is a positive universal constant.

Let us now comment on this risk bound. The term $h^2(\Pi, \Pi^{\Phi,\lambda})$ corresponds to the approximation of Π by an element of $\mathcal{D}$. It expresses the fact that our estimation procedure is robust with respect to the assumption that Π belongs to $\mathcal{D}$. The quantity $\sum_{j'>j} \lambda_{j'}^2$ is the bias term that we get for approximating λ by the elements of Λ_j. Given $j \geq 1$ and $m \in \mathcal{M}$, $\inf_{\psi \in H_m} \|\phi_j - \psi\|^2$ corresponds to the best approximation of ϕ_j by some element of the compact set H_m. Enlarging H_m (for the inclusion) makes this term smaller but may increase the quantity $n^{-1} \log(|H_m[1/\sqrt{n}]| \, n/\pi(m))$ which measures in some sense the massiveness of H_m. The quantity $O(\mathcal{H}, \pi, \phi_j)$ corresponds to the best trade-off that can achieved between these two terms among the family $\mathcal{H}$. It is typically the bound we would get for estimating the function ϕ_j alone by a model selection procedure among $\mathcal{H}$ (up to possible extra logarithmic factors). The sum $\sum_{j'=1}^{j} O(\mathcal{H}, \pi, \phi_{j'})$ is therefore the risk bound we get for estimating the j first elements $\phi_1, \ldots, \phi_j$ of $\Phi = (\phi_{j'})_{j' \geq 1}$.

In order to specify these quantities, let us turn to the following typical situation. Let $(S_m)_{m \in \mathcal{M}}$ be a family of finite-dimensional subspaces of $\mathbb{H}$ with respective dimension $D_m \geq 1$ (viewed as a linear space on $\mathbb{R}$) and for $m \in \mathcal{M}$, let us take $H_m = \mathcal{S} \cap S_m$ where $\mathcal{S}$ denotes the unit sphere of $\mathbb{H}$. The following results hold.

Proposition 6. *For all* $n \geq 1$,

$$(4.3) \qquad \log\left|H_m[1/\sqrt{n}]\right| \leq D_m \log(2\sqrt{n} + 1).$$

Besides, for all $\phi \in \mathcal{S}$

$$(4.4) \qquad \inf_{\psi \in H_m} \|\phi - \psi\| \leq 4 \inf_{\psi \in S_m} \|\phi - \psi\|.$$

The first inequality gives a control of the maximal size of a $1/\sqrt{n}$-separated subset of H_m. The second one shows that H_m and S_m share similar approximation properties with respect to the elements of $\mathcal{S}$. The proof of the proposition is delayed to Section 5.5. With such a result, we deduce from Theorem 1 the following corollary.

Corollary 1. *Let* $\mathbb{S} = (S_m)_{m \in \mathcal{M}}$ *be an at most countable family of finite dimensional subspaces of* $(\mathbb{H}, \| \ \|)$ *with dimensions* $D_m \geq 1$ *and let* π *be a sub-probability on* $\mathcal{M}$. *There exists a density estimator* $\widehat{\widehat{\Pi}} = \widehat{\Pi}(\mathbb{S}, \pi)$ *such that whatever the density* Π *on* $(\mathbb{X}, \mathcal{B}(\mathbb{X}), L)$,

$$C\mathbb{E}\left[h^2(\Pi, \widehat{\Pi})\right] \leq \inf_{\Phi \in \mathbb{U}, \lambda \in \Lambda} \left[h^2(\Pi, \Pi^{\Phi,\lambda}) + \inf_{j \geq 1}\left(\sum_{j'=1}^{j} O(\mathbb{S}, \pi, \phi_{j'}) + \sum_{j'>j} \lambda_{j'}^2\right)\right]$$

where for all $j \geq 1$,

$$O(\mathbb{S}, \pi, \phi_j) = \inf_{m \in \mathcal{M}}\left[\inf_{\psi \in S_m} \|\phi_j - \psi\|^2 + \frac{D_m \log n + \log(1/\pi(m))}{n}\right]$$

and C *is a positive universal constant.*

For illustration, let us consider the elementary situation where $\mathcal{X} = \{1, \ldots, p\}$ and assume that $\Pi \in \mathcal{D}_{p,k}$ is a determinantal projection process on $\mathcal{X}$ of rank k as in Section 4.1. In this case, one can choose $\mathbb{S} = \{S_0\}$ where S_0 is the linear subspace of dimension p of $\mathbb{H} = \ell_2(\mathbb{N}^*)$ gathering the elements of the form $(u_1, \ldots, u_p, 0, \ldots)$ with $(u_1, \ldots, u_p) \in \mathbb{C}^p$ and π the Dirac mass at 0. Whatever $\Pi \in \mathcal{D}_{p,k}$ there exists $\Phi \in \mathbb{U}$ with $\phi_1, \ldots, \phi_k \in S_0 \cap \mathcal{S}$ such that $\Pi = \Pi^{\Phi, \lambda}$ with $\lambda_1 = \ldots = \lambda_k = 1$ and $\lambda_j = 0$ for $j > k$. Since $O(\{S_0\}, \pi, \phi_{j'}) \leq p \log n / n$ for all $j' \in \{1, \ldots, k\}$ and $\sum_{j' > k} \lambda_{j'}^2 = 0$, by applying Corollary 1 we derive the risk bound

$$C\mathbb{E}\left[h^2(\Pi, \widehat{\Pi})\right] \leq \frac{kp \log n}{n} \quad \text{for all} \quad \Pi \in \mathcal{D}_{p,k}.$$

This inequality shows that the minimax rate of estimation over $\mathcal{D}_{p,k}$ is not larger than $kp \log n / n$. Since we expect that the metric dimension of $\mathcal{D}_{p,k}$ is of order kp, we believe that the logarithmic factor could probably be dropped.

4.3. **Rates of convergence.** In this section, we assume that $\mathcal{X} = [0,1]^k$ for some integer $k \geq 1$. Our aim is to deduce from Corollary 1 some rates of convergence towards Π when it is of the form $\Pi^{\Phi, \lambda}$ for some parameter $(\Phi, \lambda) \in \mathbb{U} \times \Lambda$. To do so, we make some *a posteriori* smoothness assumptions on $\Phi = (\phi_j)_{j \geq 1}$. More precisely, we assume that the ϕ_j are real-valued and belong to classes $\mathcal{B}_{p,p}^\beta([0,1]^k)$ of (possibly) anisotropic real-valued Besov functions indexed by a number $p \in (0, +\infty]$ and a smoothness parameter $\beta = (\beta_i)_{i=1,\ldots,k} \in (0, +\infty)^k$. When $p = +\infty$, $\mathcal{B}_{\infty,\infty}^\beta([0,1]^k)$ is merely the class of anisotropic β-Hölderian functions on $[0,1]^k$, which means that a function in $\mathcal{B}_{\infty,\infty}^\beta([0,1]^k)$ is β_i-Holderian on $[0,1]$ when we keep all the coordinates fixed expect the i-th. For a more precise definition of these smoothness classes we refer to Hochmuth (2002), at least when $k = 2$. The definition there can easily be generalized to larger values of k. Denoting by $|\phi|_{\beta,p,p}$ the Besov semi-norm of a function ϕ in $\mathcal{B}_{p,p}^\beta([0,1]^k)$, we set for any $R > 0$

$$\mathbb{U}_{p,p}^\beta(R) = \left\{(\phi_j)_{j \geq 1} \in \mathbb{U} \mid \phi_j \in \mathcal{B}_{p,p}^\beta([0,1]^k), \ |\phi_j|_{\beta,p,p} \leq R, \ \forall j \geq 1\right\}$$

and for $j_0 \in \mathbb{N}^*$,

$$\mathbb{U}_{p,p}^\beta(R, j_0) = \left\{(\phi_j)_{j \geq 1} \in \mathbb{U} \mid \phi_j \in \mathcal{B}_{p,p}^\beta([0,1]^k), \ |\phi_j|_{\beta,p,p} \leq R, \ \forall j = 1, \ldots, j_0\right\}.$$

In order to approximate the elements of such class, we use the following result of Akakpo (2009).

Proposition 7. *Let $p > 0$, $k \in \mathbb{N}^*$ and $r \in \mathbb{N}$. There exist a collection of linear spaces $(S_m)_{m \in \mathcal{M}_{k,r}}$ with $\mathcal{M}_{k,r} = \bigcup_{D \geq 1} \mathcal{M}_{k,r}(D)$ and a positive number $C_{k,r}$ such that for all positive integer D,*

$$(4.5) \qquad |\mathcal{M}_{k,r}(D)| \leq e^{C_{r,k}D}, \quad \sup_{m \in \mathcal{M}_{k,r}(D)} \dim(S_m) \leq C_{r,k}D$$

and

(4.6)
$$\inf_{m \in \mathcal{M}_{k,r}(D)} \inf_{\psi \in S_m} \|\phi - \psi\| \leq C(k, r, p) |\phi|_{\beta, p, p} D^{-\overline{\beta}/k}$$

for all $\phi \in \mathcal{B}_{p,p}^{\beta}([0,1]^k)$ and β satisfying

(4.7)
$$\sup_{1 \leq i \leq k} \beta_i < r + 1 \quad and \quad \overline{\beta} = \left(\frac{1}{k} \sum_{i=1}^{k} \frac{1}{\beta_i} \right)^{-1} > k \left[\left(p^{-1} - 2^{-1} \right) \vee 0 \right].$$

Hereafter, k being fixed, we consider the family of linear spaces $\mathbb{S} = (S_m)_{m \in \mathcal{M}}$ indexed with $\mathcal{M} = \bigcup_{r \geq 0} \mathcal{M}_{k,r}$ (omitting thus the dependency with respect to k) and endow $\mathcal{M}$ with the sub-probability π defined by

$$\pi(m) = \inf \left\{ e^{-(1+C_{r,k})D-r} \mid (r, D) \in \mathbb{N} \times \mathbb{N}^*, m \in \mathcal{M}_{k,r}(D) \right\}.$$

By using the first part of (4.5),

$$\sum_{m \in \mathcal{M}} \pi(m) \leq \sum_{r \geq 0} \sum_{D \geq 1} \sum_{m \in \mathcal{M}_{k,r}(D)} e^{-(1+C_{r,k})D-r} \leq \sum_{r \geq 0} e^{-r} \sum_{D \geq 1} e^{-D} \leq 1$$

and hence π is a sub-probability on $\mathcal{M}$.

By applying our Corollary 1 with the family $(S_m)_{m \in \mathcal{M}}$ and this sub-probability π, we deduce the following uniform rates of convergence over the classes of densities of determinantal projection processes of rank $j_0 \geq 1$ and parameter Φ belonging to $\mathbb{U}_{p,p}^{\beta}(R, j_0)$. The proof of the result is delayed to Section 5.6.

Proposition 8. *There exists an estimator $\widehat{\Pi}$ such that for all $j_0 \geq 1$, $R > 0$, $\beta \in (0, +\infty)^k$ and $p \in (0, +\infty]$ such that $\overline{\beta} > k \left[\left(p^{-1} - 2^{-1} \right) \vee 0 \right]$, we have*

$$\sup_{\Phi \in \mathbb{U}_{p,p}^{\beta}(R, j_0)} \mathbb{E} \left[h^2(\Pi_{\{1,\dots,j_0\}}^{\Phi}, \widehat{\Pi}) \right] \leq C j_0 \left(\frac{\log n}{n} \right)^{\frac{2\overline{\beta}}{2\overline{\beta}+k}},$$

where C denotes some positive number depending on k, R, p and β only.

When $j_0 = 1$, $\Pi_{\{1\}}^{\Phi}$ is merely a density on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ of the form $|\phi_1|^2$ for some function ϕ_1 of unit norm belonging to $\mathcal{B}_{p,p}^{\beta}([0,1]^k)$. Note that $\Pi_{\{1\}}^{\Phi} = |\phi_1|^2$ also belongs to $\mathcal{B}_{p,p}^{\beta}([0,1]^k)$ and up to the logarithmic factor, the rate we get is the usual one for estimating a density in $\mathcal{B}_{p,p}^{\beta}([0,1]^k)$.

Let us now establish uniform rates of convergence towards more general classes of determinantal densities. To do so, we also need to make *a posteriori* assumptions

on λ. More precisely, we assume that it belongs to classes of the form

$$\Lambda_\alpha^{(a)}(A) \;=\; \left\{ \lambda \in \Lambda \mid \sum_{j'>j} \lambda_{j'}^2 \leq Aj^{-\alpha}, \;\; \forall j \geq 1 \right\}$$

$$\text{or}$$

$$\Lambda_\alpha^{(g)}(A) \;=\; \left\{ \lambda \in \Lambda \mid \sum_{j'>j} \lambda_{j'}^2 \leq Ae^{-\alpha j}, \;\; \forall j \geq 1 \right\}$$

for some $A, \alpha > 0$. These sets contain sequences of $(\lambda_j)_{j \geq 1}$ which are decreasing polynomially and exponentially fast respectively. We get the following result the proof of which is delayed to Section 5.6.

Proposition 9. *There exists an estimator* $\widehat{\Pi}$ *such that for all* $A, \alpha, R > 0$, $\beta \in (0, +\infty)^k$ *and* $p \in (0, +\infty]$ *such that* $\overline{\beta} > k \left[\left(p^{-1} - 2^{-1} \right) \vee 0 \right]$, *we have*

$$(4.8) \qquad \sup_{(\Phi,\lambda) \in \mathbb{U}_{p,p}^\beta(R) \times \Lambda_\alpha^{(a)}(A)} \mathbb{E}\left[h^2(\Pi^{\Phi,\lambda}, \widehat{\Pi}) \right] \;\leq\; C \left(\frac{\log n}{n} \right)^{\frac{2\alpha\overline{\beta}}{(2\overline{\beta}+k)(1+\alpha)}}$$

$$(4.9) \qquad \sup_{(\Phi,\lambda) \in \mathbb{U}_{p,p}^\beta(R) \times \Lambda_\alpha^{(g)}(A)} \mathbb{E}\left[h^2(\Pi^{\Phi,\lambda}, \widehat{\Pi}) \right] \;\leq\; C \left(\frac{(\log n)^{2+k/(2\overline{\beta})}}{n} \right)^{\frac{2\overline{\beta}}{2\overline{\beta}+k}}$$

where C *denotes some positive number depending on* k, A, R, p, α *and* β *only.*

5. Proofs

5.1. **Proof of Proposition 2.** The first equality is clear since by (3.1) the Hellinger affinity between Π_J^Φ and Π_J^Ψ equals

$$\rho(\Pi_J^\Phi, \Pi_J^\Psi) \;=\; \int_{\mathbb{X}_k} \sqrt{\Pi_J^\Phi(\alpha) \Pi_J^\Psi(\alpha)} dL(\alpha)$$

$$=\; \int_{\mathbb{X}_k} |\det [\Phi_{\alpha,J}]| \, |\det [\Psi_{\alpha,J}]| \, dL(\alpha).$$

For the second part we use Proposition 1 and get

$$1 - \int_{\mathbb{X}_k} |\det [\Phi_{\alpha,J}]| \, |\det [\Psi_{\alpha,J}]| \, dL(\alpha) \;\leq\; 1 - \left| \int_{\mathbb{X}_k} \det \left[\Phi_{J,\alpha}^* \right] \det \left[\Psi_{\alpha,J} \right] dL(\alpha) \right|$$

$$=\; 1 - \left| \det \left[\left(\langle \phi_i, \psi_j \rangle \right)_{i,j \in J} \right] \right|.$$

Let us now prove the last inequality and set $a = \sum_{j \in J} \|\phi_j - \psi_j\|^2$. If $a > 2/5$ then the result is true since the Hellinger distance is bounded by 1. We may therefore assume that $a \leq 2/5$. In the remaining part of the proof we consider

the linear space $\mathcal{M}_{J \times J}(\mathbb{C})$ of $|J| \times |J|$ matrices indexed by J with entries in $\mathbb{C}$. We endow $\mathcal{M}_{J \times J}(\mathbb{C})$ with the Hilbert-Schmidt norm defined by

$$\|A\| = \left(\sum_{i \in J} \sum_{j \in J} |A_{i,j}|^2 \right)^{1/2}.$$

It is well-known that this norm is sub-multiplicative in the sense that for all $A, B \in \mathcal{M}_{J \times J}(\mathbb{C})$, $\|AB\| \leq \|A\| \, \|B\|$ and it also satisfies

(5.1) $$|\mathrm{Tr}(AB)| \leq \|A\| \, \|B\|.$$

One can decompose the matrix $A = (\langle \phi_i, \psi_j \rangle)_{i,j \in J}$ as $A = D + B$ where D is diagonal with entries $D_{i,i} = \langle \phi_i, \psi_i \rangle$ and $B = A - D$. Since $\|\phi_i\| = \|\psi_i\| = 1$ for all i

(5.2) $$|D_{i,i}| \geq \Re(D_{i,i}) = 1 - \frac{\|\phi_i - \psi_i\|^2}{2} \geq 1 - \frac{a}{2} > 0 \quad \forall i = 1, \ldots, k$$

and hence, D is non-singular. We may therefore write

$$\det A = \det D \det(I + M) \quad \text{with} \quad M = D^{-1}B$$

and since for all i, $\sum_{j \in J} |\langle \phi_i, \psi_j \rangle|^2 \leq \|\phi_i\|^2 = 1$,

$$\|M\|^2 = \sum_{i \in J} \frac{1}{|D_{i,i}|^2} \sum_{j \neq i} |\langle \phi_i, \psi_j \rangle|^2 \leq \sum_{i \in J} \frac{1 - |D_{i,i}|^2}{|D_{i,i}|^2} = \Delta_J^2$$

and by using (5.2) with the fact that $u \mapsto u^{-2}(1 - u^2)$ is decreasing on $(0, +\infty)$,

(5.3) $$\|M\|^2 \leq \Delta_J^2 \leq \frac{1 - a/4}{(1 - a/2)^2} \sum_{j \in J} \|\phi_j - \psi_j\|^2 \leq \frac{9}{16} < 1.$$

The matrix $I + M$ is therefore non-singular and we may write $I + M = e^L$ for some matrix $L \in \mathcal{M}_{J \times J}(\mathbb{C})$. In fact,

$$L = M - \frac{M^2}{2} + \sum_{p \geq 3} (-1)^{p-1} \frac{M^p}{p}$$

where the series converge normally in $(\mathcal{M}_{J \times J}(\mathbb{C}), \|\,\|)$. Moreover,

$$\det(I + M) = e^{\mathrm{Tr}(L)}.$$

Since the mapping $L \mapsto \mathrm{Tr}(L)$ is linear and continuous on $(\mathcal{M}_{J \times J}(\mathbb{C}), \|\,\|)$ and since $\mathrm{Tr}(M) = 0$,

$$\mathrm{Tr}(L) = -\frac{\mathrm{Tr}(M^2)}{2} + \sum_{p \geq 3} (-1)^{p-1} \frac{\mathrm{Tr}(M^p)}{p}.$$

By using (5.1) and the sub-multiplicative property of the Hilbert-Schmidt norm, we get

$$\left| \mathrm{Tr}(L) + \frac{\mathrm{Tr}(M^2)}{2} \right| \leq \sum_{p \geq 3} \frac{\|M\|^p}{p} \leq \frac{\|M\|^3}{3(1 - \|M\|)}$$

and thus, by using that $\|M\| \leq \Delta_J$,

$$
\begin{aligned}
\Re(\operatorname{Tr}(L)) \;&\geq\; -\Re\left(\frac{\operatorname{Tr}(M^2)}{2}\right) - \frac{\|M\|^3}{3(1-\|M\|)} \\
&\geq\; -\frac{\|M\|^2}{2}\left(1 + \frac{2\|M\|}{3(1-\|M\|)}\right) \\
&\geq\; -\frac{\Delta_J^2}{2}\left(1 + \frac{2\Delta_J}{3(1-\Delta_J)}\right).
\end{aligned}
$$

This inequality together with the fact that $\log u \geq -(1-u)/u$ for all $u > 0$ leads to

$$
\begin{aligned}
|\det A| \;&=\; |\det D|\, e^{\Re(\operatorname{Tr}(L))} \\
&\geq\; \exp\left[\frac{1}{2}\sum_{i\in J}\log\left(|D_{i,i}|^2\right) - \frac{\Delta_J^2}{2}\left(1 + \frac{2\Delta_J}{3(1-\Delta_J)}\right)\right] \\
&\geq\; \exp\left[-\frac{\Delta_J^2}{2} - \frac{\Delta_J^2}{2}\left(1 + \frac{2\Delta_J}{3(1-\Delta_J)}\right)\right] \\
&=\; \exp\left[-c(\Delta_J)\Delta_J^2\right]
\end{aligned}
$$

with $c(u) = 1 + u/[3(1-u)]$ for $u \geq 0$. By using the facts that $a \leq 2/5$, c is increasing and (5.3), we get

$$
\begin{aligned}
h^2(\Pi_J^\Phi, \Pi_J^\Psi) \;&\leq\; 1 - |\det A| \\
&\leq\; 1 - \exp\left[-c\left(\frac{a(1-a/4)}{(1-a/2)^2}\right)\Delta_J^2\right] \\
&\leq\; c\left(\frac{a(1-a/4)}{(1-a/2)^2}\right)\frac{(1-a/4)}{(1-a/2)^2}\sum_{j\in J}\|\phi_j - \psi_j\|^2 \\
&\leq\; \frac{5}{2}\sum_{j\in J}\|\phi_j - \psi_j\|^2.
\end{aligned}
$$

5.2. **Proof of Proposition 3.** Let us set

$$
R = \int_T P_t q(t)\, dm(t).
$$

Since $h^2(P,Q) \leq 2h^2(P,R) + 2h^2(R,Q)$, it remains to bound each of those terms from above. The measures ν and m being σ-finite, we may apply Fubini-Tonnelli theorem. By using the Cauchy-Schwarz inequality, we bound the first

term as follows.

$$
\begin{aligned}
2h^2(P,R) &= \int_E \left(\sqrt{P} - \sqrt{R}\right)^2 d\nu = \int_E \frac{(P-R)^2}{(\sqrt{P}+\sqrt{R})^2} d\nu \\
&\leq \int_E \frac{\left(\int_T P_t(p(t)-q(t))dm(t)\right)^2}{P+R} d\nu \\
&\leq \int_E \frac{\left(\int_T \sqrt{P_t}(\sqrt{p(t)}-\sqrt{q(t)})\sqrt{P_t}(\sqrt{p(t)}+\sqrt{q(t)})dm(t)\right)^2}{P+R} d\nu \\
&\leq \int_E \left[\int_T P_t\left(\sqrt{p(t)}-\sqrt{q(t)}\right)^2 dm(t) \times \frac{\int_T P_t(\sqrt{p(t)}+\sqrt{q(t)})^2 dm(t)}{P+R}\right] d\nu \\
&\leq 2\int_T \left[\int_E P_t(x)\left(\sqrt{p(t)}-\sqrt{q(t)}\right)^2 d\nu(x)\right] dm(t) = 4h^2(p,q).
\end{aligned}
$$

Let us now turn to the second term. By using similar arguments,

$$
\begin{aligned}
2h^2(R,Q) &\leq \int_E \frac{\left(\int_T (P_t-Q_t)q(t)dm(t)\right)^2}{R+Q} d\nu \\
&= \int_E \frac{\left(\int_T (\sqrt{P_t}-\sqrt{Q_t})\sqrt{q(t)} \times (\sqrt{P_t}+\sqrt{Q_t})\sqrt{q(t)}dm(t)\right)^2}{R+Q} d\nu \\
&\leq \int_E \left[\int_T \left(\sqrt{P_t}-\sqrt{Q_t}\right)^2 q(t)dm(t) \times \frac{\int_T (\sqrt{P_t}+\sqrt{Q_t})^2 q(t)dm(t)}{R+Q}\right] d\nu \\
&\leq 4\int_T h^2(P_t,Q_t)q(t)dm(t).
\end{aligned}
$$

We conclude by adding these two upper bounds.

5.3. **Proof of Proposition 4.** Inequality (3.11) derives from (3.9), (3.10) and Proposition 3. Hence, it remains to prove (3.9) and (3.10).

Let us prove (3.9). To do so, we set $a^{-1} = e/[2(e-1)] < 1$ and prove the stronger inequality

$$
h^2(p^\lambda, p^\gamma) \leq a^{-1}\left[|\lambda-\gamma|^2 + |\check{\lambda}-\check{\gamma}|^2\right].
$$

If there exists $j \geq 1$ such that $|\lambda_j - \gamma_j|^2 + |\check{\lambda}_j - \check{\gamma}_j|^2 > a$ then the result is clear since $h^2(p^\lambda, p^\gamma) \leq 1$. Otherwise,

$$
\begin{aligned}
h^2(p^\lambda, p^\gamma) &= 1 - \sum_{J \in \mathcal{P}} \prod_{j \in J} \lambda_j \gamma_j \prod_{j \notin J} \check{\lambda}_j \check{\gamma}_j = 1 - \prod_{j \geq 1} \left(\lambda_j \gamma_j + \check{\lambda}_j \check{\gamma}_j \right) \\
&= 1 - \prod_{j \geq 1} \left[1 - \frac{1}{2} \left(|\lambda_j - \gamma_j|^2 + |\check{\lambda}_j - \check{\gamma}_j|^2 \right) \right] \\
&= 1 - \exp \left[\sum_{j \geq 1} \log \left(1 - \frac{1}{2} \left(|\lambda_j - \gamma_j|^2 + |\check{\lambda}_j - \check{\gamma}_j|^2 \right) \right) \right]
\end{aligned}
$$

and by using that $\log(1 - u) \geq [2a^{-1} \log(1 - a/2)] u$ for all $u \in [0, a/2]$, we get

$$
\begin{aligned}
h^2(p^\lambda, p^\gamma) &\leq \frac{-\log(1 - a/2)}{a} \sum_{j \geq 1} \left[|\lambda_j - \gamma_j|^2 + |\check{\lambda}_j - \check{\gamma}_j|^2 \right] \\
&= \frac{1}{a} \left[|\lambda - \gamma|^2 + |\check{\lambda} - \check{\gamma}|^2 \right]
\end{aligned}
$$

with our choice of a.

Let us now prove (3.10). By using Proposition 2 we have that for all $J \in \mathcal{P}$, $J \neq \varnothing$

(5.4) $$ h^2(\Pi_J^\Phi, \Pi_J^\Psi) \leq \frac{5}{2} \sum_{j \in J} \|\phi_j - \psi_j\|^2 \,. $$

With the convention $\sum_\varnothing = 0$, this inequality remains true when $J = \varnothing$ since in this case $\Pi_J^\Phi = \Pi_J^\Psi = \delta_\varnothing$ and thus $h^2(\Pi_J^\Phi, \Pi_J^\Psi) = 0$. We may therefore write

$$
\begin{aligned}
\sum_{J \in \mathcal{P}} p_J^\gamma h^2(\Pi_J^\Phi, \Pi_J^\Psi) &\leq \frac{5}{2} \sum_{J \in \mathcal{P}} p_J^\gamma \sum_{j \in J} \|\phi_j - \psi_j\|^2 \\
&\leq \frac{5}{2} \sum_{j \geq 1} \|\phi_j - \psi_j\|^2 \sum_{J \in \mathcal{P}, j \in J} p_J^\gamma \\
&\leq \frac{5}{2} \sum_{j \geq 1} \gamma_j^2 \|\phi_j - \psi_j\|^2 \sum_{J \in \mathcal{P}, j \in J} \prod_{j' \in J, j' \neq j} \gamma_{j'}^2 \prod_{j' \notin J} (1 - \gamma_{j'}^2) \\
&= \frac{5}{2} \sum_{j \geq 1} \gamma_j^2 \|\phi_j - \psi_j\|^2 \prod_{j' \geq 1, j' \neq j} (\gamma_{j'}^2 + (1 - \gamma_{j'}^2)) \\
&= \frac{5}{2} \sum_{j \geq 1} \gamma_j^2 \|\phi_j - \psi_j\|^2
\end{aligned}
$$

as claimed.

5.4. **Proof of Theorem 1.** The proof is based on Proposition 5 with suitable choices of $(\Pi_m)_{m \in \mathfrak{M}}$ and sub-probability π' on $\mathfrak{M}$.

Let j be some positive integer. We set

$$\Lambda_j[1/n] = \{\gamma \in \Lambda | \ \gamma_\ell \in \{i/n, i = 1, \ldots n\} \text{ if } \ell \leq j, \ \gamma_\ell = 0 \ \text{ otherwise}\},$$

and

$$\mathbb{U}_j = \{\Phi_{\{1,\ldots,j\}} | \ \Phi \in \mathbb{U}\} \subset \mathbb{H}^j.$$

We endow $\mathbb{H}^j$ with the distance $d(\cdot, \cdot)$ defined for $(\phi_j)_{\ell=1,\ldots,j}$ and $(\psi_j)_{\ell=1,\ldots,j}$ in $\mathbb{H}^j$ by

$$d^2((\phi_j)_{\ell=1,\ldots,j}, (\psi_j)_{\ell=1,\ldots,j}) = \sum_{\ell=1}^{j} \|\phi_\ell - \psi_\ell\|^2 .$$

For all $m_1, \ldots, m_j \in \mathcal{M}$ and $\Phi \in \mathbb{U}$, there exist $\widetilde{\phi}_1 \in H_{m_1}[1/\sqrt{n}], \ldots, \widetilde{\phi}_j \in H_{m_j}[1/\sqrt{n}]$ such that for all $\ell \in \{1, \ldots, j\}$,

$$(5.5) \qquad \left\|\phi_\ell - \widetilde{\phi}_\ell\right\| \leq \inf_{\psi \in H_{m_\ell}} \|\phi_\ell - \psi\| + 1/\sqrt{n}$$

since $H_{m_\ell}[1/\sqrt{n}]$ is a $1/\sqrt{n}$-net for H_{m_ℓ}. For such an element $(\widetilde{\phi}_j)_{\ell=1,\ldots,j} \in \mathbb{H}^j$, there exists $\overline{\Phi} \in \mathbb{U}$ such that

$$d((\widetilde{\phi}_j)_{\ell=1,\ldots,j}, \overline{\Phi}_{\{1,\ldots,j\}}) \leq \inf_{\Psi \in \mathbb{U}} d((\widetilde{\phi}_j)_{\ell=1,\ldots,j}, \Psi_{\{1,\ldots,j\}}) + \frac{1}{\sqrt{n}}.$$

Since $\overline{\Phi}$ only depends on $\widetilde{\phi}_1, \ldots, \widetilde{\phi}_j$, the cardinality of the set $\mathbb{U}(j, m_1, \ldots, m_j)$ gathering such $\overline{\Phi}$ when $\widetilde{\phi}_1, \ldots, \widetilde{\phi}_j$ vary among $H_{m_1}[1/\sqrt{n}], \ldots, H_{m_j}[1/\sqrt{n}]$ respectively is not larger than $\prod_{\ell=1}^{j} |H_{m_\ell}[1/\sqrt{n}]|$. Besides, by using that $\Phi \in \mathbb{U}$

$$
\begin{aligned}
d(\Phi_{\{1,\ldots,j\}}, \overline{\Phi}_{\{1,\ldots,j\}}) \ & \leq \ d(\Phi_{\{1,\ldots,j\}}, (\widetilde{\phi}_j)_{\ell=1,\ldots,j}) + \inf_{\Psi \in \mathbb{U}} d((\widetilde{\phi}_j)_{\ell=1,\ldots,j}, \Psi_{\{1,\ldots,j\}}) + \frac{1}{\sqrt{n}} \\
& \leq \ 2d(\Phi_{\{1,\ldots,j\}}, (\widetilde{\phi}_j)_{\ell=1,\ldots,j}) + \frac{1}{\sqrt{n}}
\end{aligned}
$$

and hence, by using (5.5)

$$
\begin{aligned}
d^2(\Phi_{\{1,\ldots,j\}}, \overline{\Phi}_{\{1,\ldots,j\}}) \ & \leq \ 8\sum_{\ell=1}^{j} \left\|\phi_j - \widetilde{\phi}_j\right\|^2 + \frac{2}{n} \\
(5.6) \qquad\qquad & \leq \ 16\sum_{\ell=1}^{j} \inf_{\psi \in H_{m_\ell}} \|\phi_\ell - \psi\|^2 + \frac{16j + 2}{n}.
\end{aligned}
$$

For all $\lambda \in \Lambda$, let us set $\widetilde{\lambda} = (\lambda_\ell \mathbb{1}_{\ell \leq j})_{\ell \geq 1} \in \Lambda$. Since for all $\ell \geq 1$, $\lambda_\ell \in [0, 1]$, we have

$$
\begin{aligned}
\left|\lambda - \widetilde{\lambda}\right|^2 + \left|\check{\lambda} - \check{\widetilde{\lambda}}\right|^2 &= \sum_{\ell \geq 1} \left(\left|\lambda_\ell - \widetilde{\lambda}_\ell\right|^2 + \left|\sqrt{1 - \lambda_\ell^2} - \sqrt{1 - \widetilde{\lambda}_\ell^2}\right|^2 \right) \\
&\leq \sum_{\ell > j} \left(\lambda_\ell^2 + \left|1 - \sqrt{1 - \lambda_\ell^2}\right|^2 \right) \\
&\leq \sum_{\ell > j} \left(\lambda_\ell^2 + \frac{\lambda_\ell^4}{\left(1 + \sqrt{1 - \lambda_\ell^2}\right)^2} \right) \\
&\leq 2 \sum_{\ell > j} \lambda_\ell^2.
\end{aligned}
$$
(5.7)

By construction of $\Lambda_j[1/n]$, there exists $\overline{\lambda} \in \Lambda_j[1/n]$ such that $\left|\widetilde{\lambda}_\ell - \overline{\lambda}_\ell\right| \leq 1/n$ for all $\ell \geq 1$ and by using that for all $u, v \in [0, 1]$, $\left|\sqrt{1 - u^2} - \sqrt{1 - v^2}\right| \leq \sqrt{2|u - v|}$, for such an element of $\Lambda_j[1/n]$,

$$
\begin{aligned}
\left|\overline{\lambda} - \widetilde{\lambda}\right|^2 + \left|\check{\overline{\lambda}} - \check{\widetilde{\lambda}}\right|^2 &\leq \sum_{\ell=1}^{j} \left(\left|\overline{\lambda}_\ell - \widetilde{\lambda}_\ell\right|^2 + \left|\sqrt{1 - \overline{\lambda}_\ell^2} - \sqrt{1 - \widetilde{\lambda}_\ell^2}\right|^2 \right) \\
&\leq 3 \sum_{\ell=1}^{j} \left|\overline{\lambda}_\ell - \widetilde{\lambda}_\ell\right| \leq \frac{3j}{n}.
\end{aligned}
$$
(5.8)

By using (5.7) and (5.8) with the triangular inequality, we obtain that

$$
\left|\lambda - \overline{\lambda}\right|^2 + \left|\check{\lambda} - \check{\overline{\lambda}}\right|^2 \leq 4 \sum_{\ell > j} \lambda_\ell^2 + \frac{6j}{n}.
$$
(5.9)

By combining (3.11) with (5.6) and (5.9), for all $\Phi \in \mathbb{U}$ and $\lambda \in \Lambda$, there exists $(\overline{\Phi}, \overline{\lambda}) \in \mathbb{U}(j, m_1, \ldots, m_j) \times \Lambda_j[1/n]$ such that

$$
\begin{aligned}
h^2(\Pi^{\Phi, \lambda}, \Pi^{\overline{\Phi}, \overline{\lambda}}) &\leq 2 \left[\left|\lambda - \overline{\lambda}\right|^2 + \left|\check{\lambda} - \check{\overline{\lambda}}\right|^2 \right] + 5 \sum_{j \geq 1} \overline{\lambda}_j^2 \left\| \phi_j - \overline{\phi}_j \right\|^2 \\
&\leq 8 \sum_{\ell > j} \lambda_\ell^2 + \frac{12j}{n} + 5 d^2(\Phi_{\{1, \ldots, j\}}, \overline{\Phi}_{\{1, \ldots, j\}}) \\
&\leq 8 \sum_{\ell > j} \lambda_\ell^2 + 80 \sum_{\ell=1}^{j} \inf_{\psi \in H_{m_\ell}} \left\| \phi_\ell - \psi \right\|^2 + \frac{92j + 10}{n}.
\end{aligned}
$$
(5.10)

Let us now set

$$\mathfrak{M} = \bigcup_{j \geq 1} \bigcup_{m_1 \in \mathcal{M}, \ldots, m_j \in \mathcal{M}} \{j\} \times \bigotimes_{\ell=1}^{j} \{m_j\} \times \mathbb{U}(j, m_1, \ldots, m_j) \times \Lambda_j[1/n]$$

and for $\mathfrak{m} = (j, m_1, \ldots, m_j, \Psi, \gamma) \in \mathfrak{M}$, $\Pi_\mathfrak{m} = \Pi^{\Psi, \gamma}$ and

$$\pi'(\mathfrak{m}) = \frac{1}{(2n)^j} \prod_{\ell=1}^{j} \frac{\pi(m_\ell)}{|H_{m_\ell}[1/\sqrt{n}]|}.$$

Since for all $j \geq 1$ and $m_1, \ldots, m_j \in \mathcal{M}$, $|\mathbb{U}(j, m_1, \ldots, m_j)| \leq \prod_{\ell=1}^{j} |H_{m_\ell}[1/\sqrt{n}]|$ and $|\Lambda_j[1/n]| \leq n^j$,

$$
\begin{aligned}
\sum_{\mathfrak{m} \in \mathfrak{M}} \pi'(\mathfrak{m}) &\leq \sum_{j \geq 1} \sum_{m_1, \ldots, m_j \in \mathcal{M}} \sum_{\Psi \in \mathbb{U}(j, m_1, \ldots, m_j)} \sum_{\gamma \in \Lambda_j[1/n]} \frac{1}{(2n)^j} \prod_{\ell=1}^{j} \frac{\pi(m_\ell)}{|H_{m_\ell}[1/\sqrt{n}]|} \\
&\leq \sum_{j \geq 1} \frac{1}{2^j} \left(\sum_{m \in \mathcal{M}} \pi(m) \right)^j \leq 1
\end{aligned}
$$

and hence, π' is a sub-probability on $\mathfrak{M}$. By using Proposition 5 with the family of densities $(\Pi_\mathfrak{m})_{\mathfrak{m} \in \mathfrak{M}}$ and the sub-probability π' on $\mathfrak{M}$, we obtain an estimator $\widehat{\Pi}$ for which $\mathbb{E}\left[h^2(\Pi, \widehat{\Pi}) \right]$ is, up to a universal constant $C > 0$, not larger than

$$h^2(\Pi, \Pi_\mathfrak{m}) + \frac{\log(1/\pi'(\mathfrak{m}))}{n} \leq 2h^2(\Pi, \Pi^{\Phi, \lambda}) + 2h^2(\Pi^{\Phi, \lambda}, \Pi_\mathfrak{m}) + \frac{\log(1/\pi'(\mathfrak{m}))}{n}$$

whatever $(\Phi, \lambda) \in \mathbb{U} \times \Lambda$ and $\mathfrak{m} \in \mathfrak{M}$. In particular by using (5.10), for any choices of $j \geq 1$ and $m_1, \ldots, m_j \in \mathcal{M}$, there exists some $\mathfrak{m} = (j, m_1, \ldots, m_j, \overline{\Phi}, \overline{\lambda}) \in \mathfrak{M}$ such that

$$
h^2(\Pi^{\Phi, \lambda}, \Pi_\mathfrak{m}) + \frac{\log(1/\pi'(\mathfrak{m}))}{n}
$$

$$
= h^2(\Pi^{\Phi, \lambda}, \Pi^{\overline{\Phi}, \overline{\lambda}}) + \frac{1}{n} \sum_{\ell=1}^{j} \log \left(\frac{2 |H_{m_\ell}[1/\sqrt{n}]| n}{\pi(m_\ell)} \right)
$$

$$
\leq 102 \left[\sum_{\ell > j} \lambda_\ell^2 + \sum_{\ell=1}^{j} \left(\inf_{\psi \in H_{m_\ell}} \|\phi_\ell - \psi\|^2 + \frac{1}{n} \log \left(\frac{2 |H_{m_\ell}[1/\sqrt{n}]| n}{\pi(m_\ell)} \right) + \frac{1}{n} \right) \right].
$$

Finally, we get the result from the fact that $\Phi, \lambda, j, m_1, \ldots, m_j$ are arbitrary.

5.5. **Proof of Proposition 6.** For all $m \in \mathcal{M}$, $H_m[1/\sqrt{n}]$ is a $1/\sqrt{n}$- separated subset of the unit ball of a finite-dimensional linear space S_m on $\mathbb{R}$ of dimension D_m. Consequently, for all $m \in \mathcal{M}$

$$\log \left| H_m[1/\sqrt{n}] \right| \leq D_m \log(2\sqrt{n} + 1).$$

Let $a < 1$. If $\inf_{\psi \in S_m} \|\phi - \psi\| \geq a$, then by the triangular inequality, for all ψ' in H_m

$$\|\phi - \psi'\| \leq 2 \leq \frac{2}{a} \inf_{\psi \in S_m} \|\phi - \psi\|.$$

Otherwise, there exists $\psi \in S_m$ such that $\|\phi - \psi\| < a$. Hence, $\|\psi\| \geq \|\phi\| - \|\phi - \psi\| \geq 1 - a > 0$. In particular, $\psi \neq 0$ and we may set $\psi' = \psi / \|\psi\| \in H_m$. Since $\|\phi\| = 1$, we have

$$
\begin{aligned}
\|\phi - \psi'\| &= \left\| \phi - \frac{\psi}{\|\psi\|} \right\| = \left\| \phi - \frac{\phi}{\|\psi\|} + \frac{\phi}{\|\psi\|} - \frac{\psi}{\|\psi\|} \right\| \\
&\leq \left| \frac{\|\psi\| - \|\phi\|}{\|\psi\|} \right| + \frac{1}{\|\psi\|} \|\phi - \psi\| \\
&\leq \frac{2 \|\phi - \psi\|}{\|\psi\|} \leq \frac{2 \|\phi - \psi\|}{1 - a}.
\end{aligned}
$$

We get the result by choosing $a = 1/2$.

5.6. **Proofs of Propositions 8 and 9.** By using the collections of linear spaces $\mathbb{S}$ and our choice of π, and by using some classical optimization with respect to $m \in \mathcal{M}$, we get that for all $j \geq 1$

$$O(\mathbb{S}, \pi, \phi_j) \leq C (\log n / n)^{2\overline{\beta}/(2\overline{\beta}+k)}$$

where C is a positive constant depending on R, k, β and p. Up to the logarithmic factor, this bounds correspond to the usual estimation rate over $\mathcal{B}_{p,p}^{\beta}([0,1]^k)$. When $\lambda = (\mathbb{1}_{j \leq j_0})_{j \geq 1}$, $\Pi^{\Phi, \lambda} = \Pi_{\{1, \ldots, j_0\}}^{\Phi}$ and Corollary 1 leads to Proposition 8. For all $\lambda \in \Lambda_\alpha^{(a)}(A)$, we get from Corollary 1 that

$$C' \mathbb{E}\left[h^2(\Pi, \widehat{\Pi}) \right] \leq \inf_{j \geq 1} \left[j (\log n / n)^{2\overline{\beta}/(2\overline{\beta}+k)} + A j^{-\alpha} \right].$$

The minimum is achieved for j of order $(n / \log n)^{2\overline{\beta}/[(2\overline{\beta}+k)(1+\alpha)]}$, which leads to the rate $(\log n / n)^{2\alpha\overline{\beta}/[(2\overline{\beta}+k)(1+\alpha)]}$ as claimed. The other rate is obtained by arguing similarly and by choosing j of order

$$\frac{2\overline{\beta}}{\alpha(2\overline{\beta}+k)} \log n.$$

References

Akakpo, N. (2009). *Estimation adaptative par selection de partitions en rectangles dyadiques*. PhD thesis, University Paris XI.

Anderson, G. W., Guionnet, A., and Zeitouni, O. (2010). *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge.

Baik, J. and Rains, E. M. (2001). Algebraic aspects of increasing subsequences. *Duke Math. J.*, 109(1):1–65.

Baraud, Y. (2011). Estimator selection with respect to Hellinger-type risks. *Probab. Theory Relat. Fields*, 151(1-2):353–401.

Birgé, L. (2006). Model selection via testing: an alternative to (penalized) maximum likelihood estimators. *Ann. Inst. H. Poincaré Probab. Statist.*, 42(3):273–325.

Borodin, A., Diaconis, P., and Fulman, J. (2010). On adding a list of numbers (and other one-dependent determinantal processes). *Bull. Amer. Math. Soc. (N.S.)*, 47(4):639–670.

de Bruijn, N. G. (1955). On some multiple integrals involving determinants. *J. Indian Math. Soc. (N.S.)*, 19:133–151 (1956).

Hochmuth, R. (2002). Wavelet characterizations for anisotropic Besov spaces. *Appl. Comput. Harmon. Anal.*, 12(2):179–208.

Hough, J. B., Krishnapur, M., Peres, Y., and Virág, B. (2006). Determinantal processes and independence. *Probab. Surv.*, 3:206–229 (electronic).

Hough, J. B., Krishnapur, M., Peres, Y., and Virág, B. (2009). *Zeros of Gaussian analytic functions and determinantal point processes*, volume 51 of *University Lecture Series*. American Mathematical Society, Providence, RI.

Lyons, R. (2003). Determinantal probability measures. *Publ. Math. Inst. Hautes Études Sci.*, (98):167–212.

Mac Lane, S. and Birkhoff, G. (1988). *Algebra*. Chelsea Publishing Co., New York, third edition.

Université de Nice Sophia-Antipolis, Laboratoire J.A. Dieudonné, UMR CNRS 7351, Parc Valrose, 06108 Nice cedex 02

E-mail address: baraud@unice.fr