UNIVERSITAT DE BARCELONA

UAB
Universitat Autònoma de Barcelona

ADVANCED MATHEMATICS
MASTER'S FINAL PROJECT

Determinantal point processes

Author:
Rubén Jiménez Lumbreras

Supervisor:
Dr. Joaquim Ortega Cerdà

Facultat de Matemàtiques i Informàtica

September 28th, 2025

Abstract

In this Master's Final Project, we develop a survey of determinantal point processes, which form a family of simple point processes whose characteristic property is that their joint intensities

$$\rho_k(x_1, \ldots, x_k) = \det(\mathbb{K}(x_i, x_j))_{1 \leq i,j \leq k} \quad \forall k \geq 1 \text{ and } x_1, \ldots, x_k \in E$$

are given by the determinants of a kernel $\mathbb{K}$ in the appropriate space E. Intuitively, these processes describe and allow us to model the behaviour of objects or points that verify certain repulsion interactions between them, determined by their respective kernels.

After introducing the necessary preliminaries and using them to define this type of processes, then we proceed to review some of their most remarkable properties. Among them, we especially focus on their existence and uniqueness results, as well as on other selected properties that describe probabilistic behaviours of these processes and justify their interest.

Finally, we delve into some specific examples of determinantal point processes, with the aim of showing that they appear naturally in various fields, such as in the study of certain random matrices, whose eigenvalues are examples and give rise to processes of the family. In addition, examples of these processes have also recently appeared in the development of machine learning solutions, something that we also explore in the last section of the project.

2020 Mathematics Subject Classification: 60G55, 60G57.

Acknowledgements

I would like to deeply thank Dr. Joaquim Ortega Cerdà for all the support, time, dedication and help he has given me throughout this project, which has made it possible.

I would also like to thank my master's classmates and professors for their help and the good moments I had the pleasure of sharing with them during this past year.

Finally, I also want to thank my family for all their support, help, and advice. Thanks to them, I have been able to get to where I am and become who I am.

Contents

1. Introduction

In few words, determinantal point processes are a family of repulsive point processes with rich mathematical properties and a preferred choice for diversity-related machine learning applications. Now, let's explain what this actually means step by step and informally, aiming to motivate the interest of these processes and to fix some basic ideas that will be useful during the rest of the project.

Indeed, as mentioned above, DPPs are a family of point processes, so if want to describe DPPs we need to start by introducing first this general notion. Point processes can be seen as random subsets of a certain space, for example, we can think about subsets of the real line $\mathbb{R}$, the plane $\mathbb{R}^2$ or the natural numbers $\mathbb{N}$ among others. That means that, in some way, point processes are still random variables that can take several values and that we can sample, the only difference is that instead of getting numbers with certain probabilities we get sets of points, that again can be more or less likely to appear.

As happens with random variables, that can follow many standard distributions like the Gaussian or the uniform ones, the same is true for point processes, which can also exhibit many different behaviours depending on their type. For instance, there are some well-studied examples of point processes that tend to select their points in tightly-packed clusters, others favour instead repulsion among their points and usually distribute them evenly across the available space, while others are known for their rotation and/or translation invariances, etc. Among them, we already said that DPPs form a family of repulsive point processes, that is, the random points they generate tend to present repulsions, being it unlikely to find two points close together.

Nevertheless, DPPs are not the only type of point process that exhibits this behaviour and, in fact, there are several alternatives that are more repulsive, for example Gibbs processes. So, what makes determinantal processes unique is the other two items that we have already mentioned in our introductory statement: their rich mathematical properties and the fact that they are a preferred choice for diversity-related machine learning applications.

On the one hand, regarding their properties, the most distinct one is that their joint intensities are given by the determinants of a kernel $\mathbb{K}$. These functions, that can be computed for certain types of point processes, describe their behaviour by measuring how likely is to find points of the process at certain locations and, in this case, these probabilities take the form of a determinant. This aspect of DPPs is what we use to define them and, actually, it is the reason why they are repulsive processes, since the likelihood of finding two points near each other will be the determinant of a matrix with two similar rows and thus a value very close to 0.

Not only that, but DPPs have many other relevant properties, some of which will be explained in detail during the project. For example, it is a well-known fact that these determinantal processes satisfy a certain version of the Central Limit Theorem, something that, in particular and under the appropriate setting, tells us that the number of points found in an increasing sequence of sets tends to follow a Gaussian distribution.

In addition, we will also see that a certain concentration inequality holds for DPPs, something that allows us to use them in Montecarlo techniques and measure estimation tasks. Finally, it is also worth mentioning that there is a strong connection between these point processes and Random Matrix Theory, since some examples of the former can be seen as sets of eigenvalues of the latter. This relationship increases the reach and versatility of the rich mathematical theory to which DPPs give rise.

On the other hand, focusing now on their applications, DPPs have traditionally been used to model a wide variety of physical phenomena, ranging from quantum particles and heavy atomic nuclei to the behaviour of plasma. More recently, these point processes have also found success in machine learning systems, thanks to their ability to model a wide variety of repulsive behaviours as well as their computational tractability. Indeed, there are many fast and efficient algorithms to sample these determinantal processes, some of which can be used in general settings while others are more specific and benefit from additional optimization.

As a consequence, DPPs are useful in many machine learning tasks that require and have a focus on output diversity. For example, during this project, we will explain how these point processes can be used to build complete and insightful text summaries, to retrieve diverse images following the instructions of users as well as to increase the variability in recommender systems ranking lists.

That being said, now that we know what DPPs are, we are ready to introduce our first and strikingly simple example of a member of this family of point processes, the Carries process, that appears explained in [10]. Indeed, to build this random set, we can start by taking a sequence of independent and uniformly distributed natural numbers between 0 and 9, as the ones that follow:

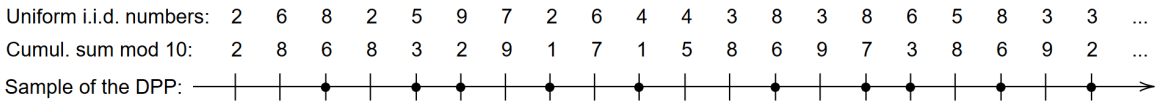


FIGURE 1. Sample from the Carries DPP

Then, the idea is to add them successively and to include the current position in our random set whenever a carry happens, that is, whenever the count becomes greater or equal than 10 and we just take the last digit, as done in the usual addition algorithm. Proceeding in this way, we could show that the resulting points form a DPP, with its joint intensities given by determinants of a rather difficult to express kernel. Although this fact may not be seen immediately, what is clear is that this process naturally favours repulsion between its points, since once the carry happens in one location it is unlikely that it will happen again in the next sum.

From here, another easy example of a determinantal process can be built as follows. We can start by considering the unit square as our space and then we can divide it into a 10×10 grid of smaller and equal squares in the natural way. In this situation, to sample our process, we just need to randomly select a point inside each of the

cells following a uniform distribution in them. Then, the obtained process is indeed determinantal, since its joint intensities are described by the determinants of a certain kernel. For further details, a more general version of this DPP will be revisited in Example 3.3, but something that can be easily seen again is that this process exhibits some kind of repulsion, because the restriction of having one point in each cell reduces the likelihood of finding two of them close together.

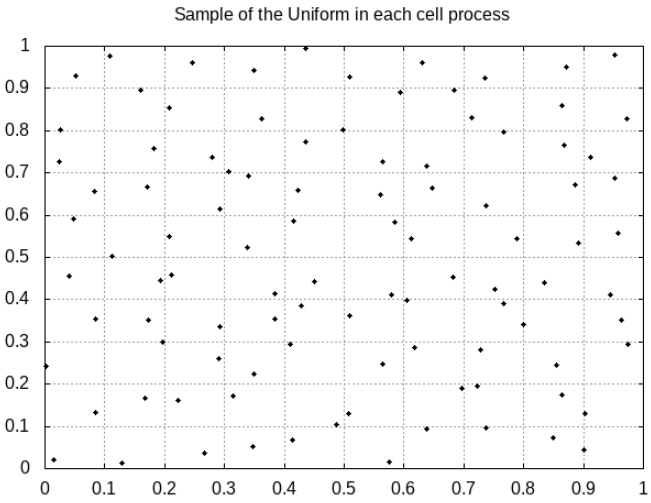


FIGURE 2. Sample from the DPP described above.

Finally, to close this introduction, we can also say a word about the history of determinantal point processes. These processes are a relatively recent mathematical object, making their first appearances in the early and mid 1960s as models for the behaviour of fermions under the appropriate physical conditions. In addition, it was during these years when several examples of DPPs were also found in the field of Random Matrix Theory, as the sets of eigenvalues of certain matrices (see Section 7). Nevertheless, it was not until the year 1975 when the formal notion of a determinantal process was introduced by Macchi in [16], aiming to create a unified theory that comprised all the isolated examples that were known to that date. Since then, as we have already explained it, these processes have been explored in detail and have been used in many remarkable theoretical and practical applications. So, today and despite their short history, DPPs are of great interest to the Probability and Mathematical Physics communities.

2. Preliminaries

2.1. **Point processes in Polish spaces.** From here, after this informal introduction, now we begin our survey of DPPs. Nevertheless, before we delve into this topic from a rigorous mathematical perspective, we need to start by setting our framework and defining some related concepts that will appear through the project. To do it, as our reference, we will mainly follow Chapters 1.2 and 4.2 of [11] unless otherwise stated.

Indeed, with that in mind, we start by focusing our attention on the notion of point processes, that we will consider in the context of locally compact Polish spaces equipped with a Radon measure. So, we begin by recalling this two concepts.

Definition 2.1 (Polish space)**.** *Let E be a topological space, we call E a Polish space if it is separable and completely metrizable by a certain metric d.*

Definition 2.2 (Radon measure)**.** *Let E be a locally compact topological space and let $\mathcal{B}(E)$ be its Borel σ-algebra, we say that μ is a Radon measure in E if it is a measure in $(E, \mathcal{B}(E))$, it is regular ($\mu(A) = \sup\{\mu(K) \,|\, K \subseteq A,\ K \text{ compact}\} = \inf\{\mu(U) \,|\, A \subseteq U,\ U \text{ open}\}$ $\forall A$ μ-measurable such that $\mu(A) < \infty$) and it is finite on compact sets.*

Therefore, from here and until the end, otherwise stated, for us E will always be a locally compact Polish space equipped with a fixed Radon measure μ. Now, with the aim of fixing ideas, as an example, note that some usual spaces verify these properties.

Example 2.3 (Open subsets of $\mathbb{R}^n$ equipped with the Lebesgue measure)**.** Indeed, let D be any open subset of $\mathbb{R}^n$ and consider μ the Lebesgue measure, it is clear that $\mathbb{R}^n$ is a Polish space, because it is complete and separable with its usual metric. Then, using the fact that every open subset of a Polish space is a Polish space (see Examples in Section 26 of [3]), we can conclude that D is a Polish space. That, combined with the well-known properties of the Lebesgue measure, implies that D equipped with μ is a locally compact Polish space with a Radon measure.

Example 2.4 (Discrete sets equipped with the counting measure)**.** Now, let Λ be a finite or countable set (for example, $\Lambda = \mathbb{N}$) and let μ be the corresponding counting measure, that assigns mass one to each of the elements of Λ. Then, injecting Λ in $\mathbb{N}$, we can easily induce a complete metric in Λ and, as all finite subsets of Λ are clearly μ-measurable, open and compact, we can conclude that Λ equipped with μ is a locally compact Polish space with a Radon measure.

We will be especially interested in these particular spaces, so it is convenient to keep them in mind. That being said, now, with the objective of introducing the notion of point processes, we start by recalling a type of Radon measures called point measures.

Definition 2.5 (Point measure)**.** *Let μ be a Radon measure in E, then we say that μ is a point measure in E if there exists $S \subseteq E$ discrete such that $\mu = \sum_{x \in S} a_x \delta_x$ for some $a_x \in \mathbb{N}$. Note that, given $x \in E$, δ_x denotes the Dirac measure of x.*

Not only that, but also denote by $M(E)$ the set of all Radon measures in E. Then, if we take any $D \in \mathcal{B}(E)$ and any open interval $I \subseteq [0, \infty)$, it is clear that we can

consider the corresponding subsets $\{\mu \in M(E) \,|\, \mu(D) \in I\}$ of Radon measures, which can be used to generate a σ-algebra in $M(E)$, that we will denote by $\mathcal{M}$. So, with this construction, $(M(E), \mathcal{M})$ is a measurable space.

Finally, combining all of the above, we are ready to present the definition of a point process.

Definition 2.6 (Point process)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, then a point process $\mathscr{X}$ on E is a measurable map $\mathscr{X} : (\Omega, \mathcal{F}) \longrightarrow (M(E), \mathcal{M})$ such that $\mathscr{X}(\omega)$ is a point measure P-a.s.*

In particular, given $D \in \mathcal{B}(E)$ and $\mathscr{X}$ a point process in E, we will denote by $\mathscr{X}(D)$ the random variable $\mathscr{X}(D) : \Omega \longrightarrow [0, \infty]$ given by $\mathscr{X}(D)(\omega) = (\mathscr{X}(\omega))(D)\ \forall \omega \in \Omega$, that is, the random variable that describes the measure that $\mathscr{X}$ assigns to D.

It is worth mentioning that, although we will mainly work with the previous definition of point processes, there are alternative ways of looking at them, which can be useful in some contexts. For example, another common possibility is to consider them as stochastic processes indexed by the bounded subsets of E, which count the number of random points that appear in these subsets of the space.

Example 2.7. Let $X_1, \ldots, X_n$ be n real-valued random variables, then, following the usual notations, we can consider $\delta_{X_i}\ \forall i \in \{1, \ldots, n\}$ the corresponding random Dirac measures. Note that, given any $i \in \{1, \ldots, n\}$, $D \in \mathcal{B}(E)$ and I an open interval of $[0, \infty)$, we can observe that

$$\{\omega \in \Omega \,|\, \delta_{X_i}(D) \in I\} = \begin{cases} \emptyset & \text{if } 0 \notin I,\ 1 \notin I \\ \{\omega \in \Omega \,|\, X_i(\omega) \in D\} & \text{if } 0 \notin I,\ 1 \in I \\ \{\omega \in \Omega \,|\, X_i(\omega) \notin D\} & \text{if } 0 \in I,\ 1 \notin I \\ \Omega & \text{if } 0 \in I,\ 1 \in I \end{cases}$$

something the shows that δ_{X_i} is measurable and, in particular, a point process. Not only that, but we can define $\mathscr{X} = \sum_{i=1}^{n} \delta_{X_i}$ and, using a similar argument as above, we can conclude that $\mathscr{X}$ is also a point process in $\mathbb{R}$, the one that assigns measure one to the points given by the random variables X_i.

From here, we will focus on a particular type of point processes in which we will be especially interested, the simple point processes, that we proceed to define.

Definition 2.8 (Simple point process)**.** *Let $\mathscr{X}$ be a point process in E, we say that $\mathscr{X}$ is simple if a.s., for all $x \in E$, we have that $\mathscr{X}(\{x\}) \in \{0, 1\}$.*

So, in other words, intuitively, a simple point process can be thought as a collection of random points in E, or, more precisely, as random discrete subset of the space, whose points are the ones that are assigned measure 1 by $\mathscr{X}$.

Once introduced these notions, recall that, in general, to describe the distribution of a point process $\mathscr{X}$, one has to be able to determine the values $P(\mathscr{X} \in A)\ \forall A \in \mathcal{M}$. However, note that the elements of $\mathcal{M}$ can be expressed as (countable) unions,

(countable) intersections and/or complements of the sets $\{\mu \in M(E) \,|\, \mu(D) \in I\}$ for $D \in \mathcal{B}(E)$ and I open interval of $[0, \infty)$. So, by the properties of the probability P, it is actually enough to focus the probabilities $P(\mathscr{X} \in (\bigcap_{i=1}^{l} A_i) \cap (\bigcap_{i=l+1}^{k} A_i^c))$, where $A_i = \{\mu \in M(E) \,|\, \mu(D_i) \in I_i\}$ $\forall i \in \{1, \ldots, k\}$ for any $D_1, \ldots, D_k \in \mathcal{B}(E)$, $I_1, \ldots, I_k$ open intervals of $[0, \infty)$ and $l \in \{1, \ldots, k\}$.

Not only that, but also, in this situation, we can observe that as point measures are integer-valued:

$$P(\mathscr{X} \in (\bigcap_{i=1}^{l} A_i) \cap (\bigcap_{i=l+1}^{k} A_i^c)) = \sum_{\substack{n_i \in \mathbb{N} \cap I_i \text{ for } i \leq l \\ n_i \in \mathbb{N} \cap (\mathbb{R} \setminus I_i) \text{ for } i > l}} P(\mathscr{X}(D_1) = n_1, \ldots, \mathscr{X}(D_k) = n_k)$$

Taking the above into account, then we conclude that to deduce the distribution of a point process $\mathscr{X}$, one has to know the probabilities $P(\mathscr{X}(D_1) = n_1, \ldots, \mathscr{X}(D_k) = n_k)$ for any $D_1, \ldots, D_k \in \mathcal{B}(E)$ and $n_1, \ldots, n_k \in \mathbb{N}$.

Nevertheless, in the particular case of simple point processes, considering the previous probabilities may not be the most useful way to understand or describe the behaviour of the process, especially if one is interested in studying the attraction or repulsion properties of its random points. In this case, it is more convenient to describe the distribution of the process using the following functions already mentioned in the introduction, known as joint intensities.

Definition 2.9 (Joint intensities)**.** *Let $\mathscr{X}$ be a simple point process in E and let $k \geq 1 \in \mathbb{N}$. Then, we define the k-th joint intensity of $\mathscr{X}$ (with respect to μ) as a locally integrable function $\rho_k : E^k \longrightarrow [0, \infty)$, if it exists, such that $\rho_k(x_1, \ldots, x_k) = 0$ if $x_i = x_j$ for some $i \neq j$ and that, for any $D_1, \ldots, D_k \in \mathcal{B}(E)$ pairwise disjoint sets, it verifies:*

$$\mathbb{E}(\prod_{i=1}^{k} \mathscr{X}(D_i)) = \int_{\prod_{i=1}^{k} D_i} \rho_k(x_1, \ldots, x_k) d\mu(x_1) \ldots d\mu(x_k)$$

Intuitively, being informal, one can see the joint intensities as an unordered version of the finite joint densities of the process, where, for $k \geq 1$, $\rho_k(x_1, \ldots, x_k)$ is proportional to the probability to find a point of the process in each of the infinitesimal disjoint regions $d\mu(x_1) \ldots d\mu(x_k)$ around $x_1, \ldots, x_k$ respectively. We will precise this claim in Theorem 2.17. Now, with the aim of further clarifying this notion, we start by providing a simple example of the joint intensities.

Example 2.10. Indeed, let X_1 and X_2 be two real random variables independent and identically distributed as $U(-1, 1)$. In this situation, assume that $f : \mathbb{R}^2 \longrightarrow [0, \infty)$ denotes their usual joint density function (so, $f(x_1, x_2) = \frac{1}{4}$ $\forall (x_1, x_2) \in (-1, 1)^2$ and 0 otherwise). Then, using similar arguments as the ones from Example 2.7, one can consider $\mathscr{X} = \delta_{X_1} + \delta_{X_2}$ and show that it is a simple point process. Now, we will study its joint intensities.

To do it, let any $D \in \mathcal{B}(\mathbb{R})$, we can observe that $\mathbb{E}(\mathscr{X}(D)) = \mathbb{E}(\delta_{X_1}(D)+\delta_{X_2}(D)) = P(X_1 \in D)+P(X_2 \in D) = \int_{D\cap(-1,1)} \frac{1}{2}dx + \int_{D\cap(-1,1)} \frac{1}{2}dx = \int_{D\cap(-1,1)} dx$, so we can take $\rho_1 : \mathbb{R} \longrightarrow [0,\infty)$ given by $\rho_1 = \mathcal{X}_{(-1,1)}$.

Not only that, but we can also consider any $D_1, D_2 \in \mathcal{B}(\mathbb{R})$ disjoint and deduce that $\mathbb{E}(\mathscr{X}(D_1)\mathscr{X}(D_2)) = \mathbb{E}(\delta_{X_1}(D_1)\delta_{X_1}(D_2) + \delta_{X_1}(D_1)\delta_{X_2}(D_2) + \delta_{X_1}(D_2)\delta_{X_2}(D_1) + \delta_{X_2}(D_1)\delta_{X_2}(D_2)) = 2P(X_1 \in D_1, X_2 \in D_2) = 2\int_{D_1\times D_2} \frac{1}{4}\mathcal{X}_{(-1,1)^2}(x_1,x_2)dx_1dx_2$, so we can take $\rho_2 : \mathbb{R}^2 \longrightarrow [0,\infty)$ given by $\rho_2(x_1,x_2) = \frac{1}{2}\mathcal{X}_{(-1,1)^2}(x_1,x_2)\ \forall x_1 \neq x_2 \in \mathbb{R}$ and 0 otherwise.

Finally, it is clear that for any $k \geq 3$ and $D_1, \dots, D_k \in \mathcal{B}(\mathbb{R})$ pairwise disjoint sets, $\mathbb{E}(\prod_{i=1}^k \mathscr{X}(D_i)) = 0$ because the process has only two points, so we deduce that the k-th joint intensities are constant equal to 0. That completes the example.

From here, looking at the above example, we can notice that, in fact, the previous observation is valid not only for 2 but for any number of random variables in our general space E, something that allows us to compute the joint intensities of these simple points processes given by random variables. We can state this result as the following proposition (that generalises Exercise 1.2.5 from [11]).

Proposition 2.11. *Let $X_1, \dots, X_n$ be a.s pairwise distinct and exchangeable random variables in our space E such that $f : E^n \longrightarrow [0,\infty)$ denotes their joint density function and let $\mathscr{X} = \sum_{i=1}^n \delta_{X_i}$ be the corresponding simple point process. Then, for all $1 \leq k \leq n$, the k-th joint intensity ρ_k of $\mathscr{X}$ exists and it is given by $\rho_k(x_1,\dots,x_k) = \frac{n!}{(n-k)!}\int_{E^{n-k}} f(x_1,\dots,x_n)d\mu(x_{k+1})\dots d\mu(x_n)\ \forall x_1,\dots,x_k$ pairwise distinct.*

Proof. Indeed, using the notations introduced by the statement, we start by considering any $D_1, \dots, D_k \in \mathcal{B}(E)$ pairwise disjoint subsets of E. Then, to deduce the desired expression for the k-th joint intensity of the process, it will be enough to study the following expectation $\mathbb{E}(\prod_{i=1}^k \mathscr{X}(D_i)) = \mathbb{E}((\sum_{i=1}^n \delta_{X_i}(D_1))\dots(\sum_{i=1}^n \delta_{X_i}(D_k))) = \mathbb{E}((\sum_{i=1}^n \mathbb{1}_{\{X_i\in D_1\}})\dots(\sum_{i=1}^n \mathbb{1}_{\{X_i\in D_k\}})) = \mathbb{E}((\sum_{1\leq i_1,\dots,i_k\leq n} \mathbb{1}_{\{X_{i_1}\in D_1,\dots,X_{i_k}\in D_k\}}))$.

From here, we observe that the Borel sets are pairwise disjoint, something that implies that we can consider only the indices i_j that are pairwise distinct. Not only that, but recall that the random variables are exchangeable, so the expected values do not depend on the particular indices i_j. That, combined with the fact that there are $\frac{n!}{(n-k)!}$ possible combinations of indices, lets us conclude $\mathbb{E}(\prod_{i=1}^k \mathscr{X}(D_i)) = \frac{n!}{(n-k)!}\mathbb{E}(\mathbb{1}_{\{X_1\in D_1,\dots,X_k\in D_k\}}) = \frac{n!}{(n-k)!}\int_{\prod_{i=1}^k D_i}\int_{E^{n-k}} f(x_1,\dots,x_n)d\mu(x_n)\dots d\mu(x_1)$.

Finally, recalling the definition of the joint intensities, it is clear that ρ_k exists and it is given by $\rho_k(x_1,\dots,x_k) = \frac{n!}{(n-k)!}\int_{E^{n-k}} f(x_1,\dots,x_n)d\mu(x_{k+1})\dots d\mu(x_n)\ \forall x_1,\dots,x_k$ pairwise distinct. □

Remark 2.12. Notice that, continuing in the situation stated by the previous proposition, it is easy to observe that if $k > n$, then the corresponding k-th joint intensity of the process exists and it is constant equal to zero, because the process has not enough points to divide them among all the pairwise disjoint subsets.

Once this result has been shown, recall that we have defined the joint intensities as functions that let us describe how the point process behaves in pairwise disjoint Borel subsets of the space E. Nevertheless, it is possible to extend this property to subsets that are not necessarily disjoint, for example allowing repetitions of the same subsets, something that we proceed to detail in the following theorem.

Theorem 2.13. *Let $\mathscr{X}$ be a simple point process in E and let ρ_k be its k-th joint intensity for any $k \geq 1$. Then, given $r \leq k$ and $k_1, \dots, k_r \geq 1$ natural numbers such that $\sum_{i=1}^r k_i = k$ and given $D_1, \dots, D_r \in \mathcal{B}(E)$ pairwise disjoint sets, we have that:*

$$\mathbb{E}(\prod_{i=1}^{r} \binom{\mathscr{X}(D_i)}{k_i} k_i!) = \int_{\prod_{i=1}^r D_i^{k_i}} \rho_k(x_1, \dots, x_k) d\mu(x_1) \dots d\mu(x_k)$$

Proof. Since this result is left to the reader in [11], to show it we have mainly used our own ideas. Indeed, continuing with the above notations, we can start by observing that $\mathbb{E}(\prod_{i=1}^r \binom{\mathscr{X}(D_i)}{k_i} k_i!) = \mathbb{E}(\prod_{i=1}^r \frac{\mathscr{X}(D_i)!}{k_i!(\mathscr{X}(D_i)-k_i)!} k_i!) = \mathbb{E}(\prod_{i=1}^r \frac{\mathscr{X}(D_i)!}{(\mathscr{X}(D_i)-k_i)!}) = \mathbb{E}(\prod_{i=1}^r \mathscr{X}(D_i)(\mathscr{X}(D_i)-1)\dots(\mathscr{X}(D_i)-k_i+1))$. From here, in order to be able to further develop this last expression, let $D \in \mathcal{B}(E)$ any Borel set, we start by defining $A_l = \{(x_1, \dots, x_l) \in D^l \,|\, x_i \neq x_j \ \forall\ 1 \leq i \neq j \leq l\}$ for any $l \geq 1 \in \mathbb{N}$ (note that with this definition $A_1 = D$). Then, we proceed to show by induction in l that:

$$\int_{A_l} d\mathscr{X}(x_1) \dots d\mathscr{X}(x_l) = \mathscr{X}(D)(\mathscr{X}(D)-1)\dots(\mathscr{X}(D)-l+1)$$

Initial case ($l = 1$): it is clear that, looking at the point process as a random Radon measure in E, as we defined them, we have that $\int_{A_1} d\mathscr{X}(x_1) = \int_D d\mathscr{X}(x) = \mathscr{X}(D)$. This is exactly what we needed to conclude.

Induction step: from here, we assume that the result is true for any $l - 1 \geq 1$ and we will show it now for $l \geq 2$. Indeed, we can observe that

$$\int_{A_l} d\mathscr{X}(x_1) \dots d\mathscr{X}(x_l) = \int_D \int_{D\setminus\{x_1\}} \cdots \int_{D\setminus\{x_1,\dots,x_{l-1}\}} d\mathscr{X}(x_l) \dots d\mathscr{X}(x_1) =$$

(computing the innermost integral) =

$$\int_D \int_{D\setminus\{x_1\}} \cdots \int_{D\setminus\{x_1,\dots,x_{l-2}\}} (\mathscr{X}(D) - \mathscr{X}(\{x_1, \dots x_{l-1}\})) d\mathscr{X}(x_{l-1}) \dots d\mathscr{X}(x_1) =$$

(note that above $\mathscr{X}(\{x_1, \dots x_{l-1}\}) = l - 1$ because the measure $\mathscr{X}^{l-1}$ is discrete)

$$= \int_D \int_{D\setminus\{x_1\}} \cdots \int_{D\setminus\{x_1,\dots,x_{l-2}\}} (\mathscr{X}(D) - (l-1)) d\mathscr{X}(x_{l-1}) \dots d\mathscr{X}(x_1) =$$

$$(\mathscr{X}(D) - l + 1) \int_{A_{l-1}} d\mathscr{X}(x_1) \dots d\mathscr{X}(x_{l-1}) =$$

(applying our induction hypothesis) =

$$\mathscr{X}(D)(\mathscr{X}(D)-1)\dots(\mathscr{X}(D)-l+1)$$

that is exactly what we wanted to deduce.

Not only that, but we can also notice that, given any $x_1, \dots, x_l \in D$ pairwise distinct points, we can find pairwise disjoint neighbourhoods for them in D, because it follows from our definitions that D is metrizable, and all metric spaces are regular. Recall that regular topological spaces are the ones where we can find disjoints open sets that

separate any point from any closed set that does not contain the point, so in our case we could build our desired neighbourhoods iteratively. In particular, that shows A_l is open in D^l.

Moreover, recall that E was separable, so D, that can be seen as a separable metric space, verifies the second countable axiom, something that implies by the definition of the product topology that $A_l = \bigcup_{i=1}^{\infty} A_l^{i,1} \times \cdots \times A_l^{i,l}$ for some open sets $A_l^{i,j} \subseteq D$. Finally, making the appropriate changes, we can refine this covering by rectangles in such a way that we obtain a covering of A_l by (not necessarily open) rectangles that are pairwise disjoint. Notice that, to do it, we will need to treat the boundary of our original and new rectangles (that follow from our intersections) as degenerated rectangles, that is, product sets, themselves. So, using these ideas, inspired in part by

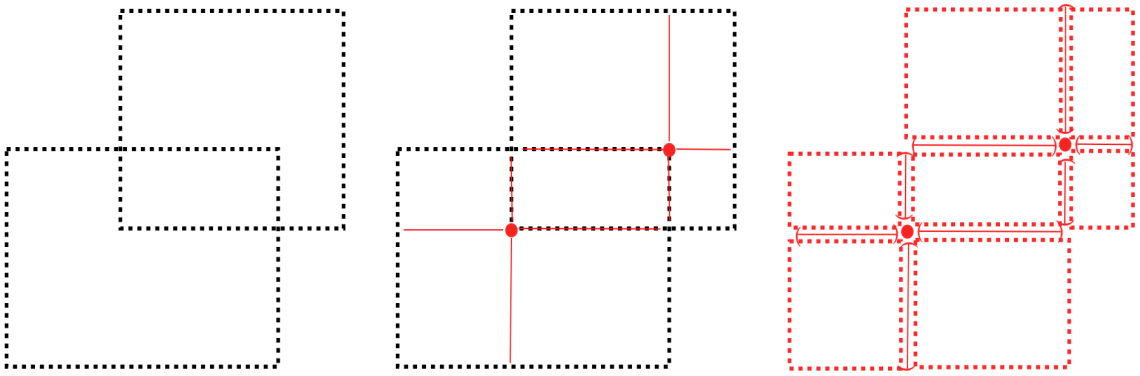


FIGURE 3. Representation of the idea used to refine the covering.

the usual covering lemmas in $\mathbb{R}^n$, as the Whitney covering lemma (originally introduced in [25]), we can deduce that $A_l = \bigcup_{i=1}^{\infty} B_l^{i,1} \times \cdots \times B_l^{i,l}$ for some Borel sets $B_l^{i,j} \subseteq D$ such that the products are pairwise disjoint between them.

From here, recovering the expected value we where trying to deduce and using the above results, we can conclude that:

$$\mathbb{E}(\prod_{i=1}^{r} \binom{\mathscr{X}(D_i)}{k_i} k_i!) = \mathbb{E}(\prod_{i=1}^{r} \int_{A_{k_i}} d\mathscr{X}(x_1) \dots d\mathscr{X}(x_{k_i})) =$$

$$\mathbb{E}(\int_{\bigcup_{i_1,\dots,i_r \geq 1} B_{k_1}^{i_1,1} \times \cdots \times B_{k_1}^{i_1,k_1} \times \cdots \times B_{k_r}^{i_r,1} \times \cdots \times B_{k_r}^{i_r,k_r}} d\mathscr{X}(x_1) \dots d\mathscr{X}(x_k)) =$$

(by the definition of the product measure) =

$$\mathbb{E}(\sum_{i_1,\dots,i_r \geq 1} \mathscr{X}(B_{k_1}^{i_1,1}) \dots \mathscr{X}(B_{k_1}^{i_1,k_1}) \dots \mathscr{X}(B_{k_r}^{i_r,1}) \dots \mathscr{X}(B_{k_r}^{i_r,k_r})) =$$

$$\sum_{i_1,\dots,i_r \geq 1} \int_{B_{k_1}^{i_1,1} \times \cdots \times B_{k_1}^{i_1,k_1} \times \cdots \times B_{k_r}^{i_r,1} \times \cdots \times B_{k_r}^{i_r,k_r}} \rho_k(x_1, \dots, x_k) d\mu(x_1) \dots d\mu(x_k) =$$

(by using known properties of the integrals) =

$$\int_{\bigcup_{i_1,\dots,i_r \geq 1} B_{k_1}^{i_1,1} \times \cdots \times B_{k_1}^{i_1,k_1} \times \cdots \times B_{k_r}^{i_r,1} \times \cdots \times B_{k_r}^{i_r,k_r}} \rho_k(x_1, \dots, x_k) d\mu(x_1) \dots d\mu(x_k) =$$

$$\int_{\prod_{i=1}^{r} A_{k_i}} \rho_k(x_1, \dots, x_k) d\mu(x_1) \dots d\mu(x_k)$$

Finally, using that $\rho_k(x_1,\dots,x_k)=0$ if $x_i=x_j$ for some $1\leq i\neq j\leq k$, we can change A_{k_i} in the integrals by $D_i^{k_i}$, something that, at the end, lets us conclude that $\mathbb{E}(\prod_{i=1}^r \binom{\mathscr{X}(D_i)}{k_i}k_i!)=\int_{\prod_{i=1}^r D_i^{k_i}}\rho_k(x_1,\dots,x_k)d\mu(x_1)\dots d\mu(x_k)$. That ends the proof. □

Notice that, in particular, using the notations introduced in the proof, we have also managed to show in between that $\mathscr{X}^l(A_l)=\binom{\mathscr{X}(D)}{l}l!$. In fact, this equality allows us to easily modify our previous proof to produce a similar result for any open set in E^k.

Corollary 2.14. *Let $\mathscr{X}$ be a simple point process in E and let ρ_k be its k-th joint intensity for any $k\geq 1$. Then, given $D\subseteq E^k$ an open subset, if $A=\{(x_1,\dots,x_k)\in D\,|\,x_i\neq x_j\ \forall 1\leq i\neq j\leq k\}$, it follows that:*

$$\mathbb{E}(\mathscr{X}^k(A))=\int_D \rho_k(x_1,\dots,x_k)d\mu(x_1)\dots d\mu(x_k)$$

Proof. Indeed, continuing with the notation introduced in the statement, we can start by observing that, since E is a Polish space, the topology in D is also given by a metric, something that shows that D is regular and lets us conclude that $A\subseteq D$ is open in D. That, combined with the fact that D is open in E^k, implies that A is also open in E^k.

Now, we can apply the same arguments as in the proof of the theorem to deduce that we can cover A in such a way that $A=\bigcup_{i=1}^\infty B_i^1\times\dots\times B_i^k$ for some Borel sets $B_i^j\subseteq E$ such that the products are pairwise disjoint between them.

Finally, if follows from the definition of the k-th joint intensity ρ_k of $\mathscr{X}$ that $\mathbb{E}(\mathscr{X}^k(A))=\mathbb{E}(\mathscr{X}^k(\bigcup_{i=1}^\infty B_i^1\times\dots\times B_i^k))=\sum_{i=1}^\infty\mathbb{E}(\mathscr{X}(B_i^1)\dots\mathscr{X}(B_i^k))=$ (definition of ρ_k) $\sum_{i=1}^\infty\int_{B_i^1\times\dots\times B_i^k}\rho_k(x_1,\dots,x_k)d\mu(x_1)\dots d\mu(x_k)=\int_A\rho_k(x_1,\dots,x_k)d\mu(x_1)\dots d\mu(x_k)$ and we use that $\rho_k(D\setminus A)=\{0\}$ to obtain the desired result. □

From here, thanks to these last results, we can now show other remarkable properties that the joint intensities of a point process verify. In particular, we will start by focusing on the expression "the joint intensities of a point process", because, indeed, they are unique.

Proposition 2.15 (Uniqueness of the joint intensities of a point process)**.** *Let $\mathscr{X}$ be a simple point process in E, assume that, for any given $k\geq 1$, $\rho_k^1:E^k\longrightarrow[0,\infty)$ and $\rho_k^2:E^k\longrightarrow[0,\infty)$ are k-th joint intensities of $\mathscr{X}$. Then, we conclude that $\rho_k^1=\rho_k^2$ a.e. in E^k.*

Proof. In this case, using the above notations and our own ideas, we can start by observing that, given $D\subseteq E^k$ any open set, one has that $\int_D\rho_k^1(x_1,\dots,x_k)d\mu(x_1)\dots d\mu(x_k)=\mathbb{E}(\mathscr{X}^k(A))=\int_D\rho_k^2(x_1,\dots,x_k)d\mu(x_1)\dots d\mu(x_k)$, where $A=\{(x_1,\dots,x_k)\in D\,|\,x_i\neq x_j\ \forall 1\leq i\neq j\leq k\}$, something that follows directly from the previous corollary.

Then, in particular, if we denote by $B_r(x)=\{y\in E^k\,|\,d^k(x,y)<r\}$ the open ball of centre $x\in E^k$ and radius $r>0$ (given by the corresponding metric d^k in E^k), we can conclude that $\rho_k^1(x)=$ (applying the Lebesgue differentiation theorem)$=\lim_{r\to 0^+}\frac{1}{\mu^k(B_r(x))}\int_{B_r(x)}\rho_k^1(y_1,\dots,y_k)d\mu(y_1)\dots d\mu(y_k)=$ (using the above observation

for $D = B_r(x)) = \lim_{r\to 0^+} \frac{1}{\mu^k(B_r(x))} \int_{B_r(x)} \rho_k^2(y_1,\dots,y_k) d\mu(y_1)\dots d\mu(y_k) = \rho_k^2(x)$ a.e. in E^k, that is what we wanted to see. As a comment, notice that, in these last equalities, we have used twice the Lebesgue differentiation theorem, which is a standard result whose proof can be found for example in page 98 of [7]. □

Now, given a simple point process $\mathscr{X}$, recall that we originally introduced its joint intensities because we wanted to have a more convenient way of describing its behaviour other than stating the probabilities $P(\mathscr{X}(D_1) = n_1, \dots, \mathscr{X}(D_k) = n_k)$ for any $D_1, \dots, D_k \in \mathcal{B}(E)$ and $n_1, \dots, n_k \in \mathbb{N}$, that can be difficult to manage, especially if we want to study and focus on the repulsion properties of the process.

The fact that these intensities are indeed able to give us this description and determine the distribution of our process $\mathscr{X}$ is not true in general, but it can be the case under the proper assumptions, as we show below.

Proposition 2.16. *Let $\mathscr{X}$ be a simple point process in E, assume that $\mathscr{X}$ has exponential tails for all compact sets in E, that is, $\forall K \subseteq E$ compact, there exists $c > 0$ and $k_0 \in \mathbb{N}$ such that $P(\mathscr{X}(K) \geq k) \leq e^{-ck}$ $\forall k \geq k_0$. Then, if the joint intensities of $\mathscr{X}$ exist, they determine its distribution.*

Proof. Indeed, to prove this result, that appears only as a brief Remark 1.2.4 in our reference [11], it suffices to show that the joint intensities of $\mathscr{X}$ determine the probabilities $p_n := P(\mathscr{X}(D_1) = n_1, \dots, \mathscr{X}(D_k) = n_k)$ for any $n = (n_1, \dots, n_k) \in \mathbb{N}^k$ and $D_1, \dots, D_k \in \mathcal{B}(E)$.

In fact, given any $D_1, \dots, D_k \in \mathcal{B}(E)$, notice that if we consider their intersections and certain decompositions similar, for example, to $D_1 = (D_1 \cap D_2) \cup (D_1 \cap D_2^c)$, we can show that there exist $A_1, \dots, A_l \in \mathcal{B}(E)$ pairwise disjoint Borel sets such that, for all $1 \leq i \leq k$, we have that $D_i = \bigcup_{j\in J} A_j$ for some $J \subseteq \{1, \dots, l\}$. That, combined with the fact that if $D_i = \bigcup_{j\in J} A_j$ then $\mathscr{X}(D_i) = \sum_{j\in J} \mathscr{X}(A_j)$, allows us to conclude that to determine the distribution of $(\mathscr{X}(D_1), \dots, \mathscr{X}(D_k))$, it is enough to determine the one of the random vector $(\mathscr{X}(A_1), \dots, \mathscr{X}(A_l))$. So, in other words, that means that we can assume that our starting Borel sets $D_1, \dots, D_k$ are pairwise disjoint without loss of generality.

In addition, given any $h \in \{1, \dots, k\}$, recall that Borel sets can be expressed as (countable) unions, (countable) intersections and/or complements of open sets. Then, as E is locally compact, we deduce by topology results that every element in E has a neighbourhood system of compact sets, something that, knowing that E is also separable, shows that all open sets in E are a countable union of compact sets. Not only that, but we can also observe that all closed sets of E are a countable union of compacts, because the intersection of a closed set and a compact one is compact (E can be written as a countable union of compacts).

That, combined with certain properties of compact sets in Polish spaces that we do not detail here, proves that all Borel sets of E are a countable union of compact sets, so we can take $D_h = \cup_{i=1}^{\infty} A_h^i$ for some $A_h^i \subseteq E$ compacts.

From here, considering $K_h^i \coloneqq \cup_{j=1}^i A_h^j$ for all $i \in \mathbb{N}$, we notice that the sets K_h^i are also compact and they define an increasing sequence such that $D_h = \cup_{i=1}^\infty K_h^i$. With that in mind, we can observe that $\forall \omega \in \Omega$ (the probability space where $\mathscr{X}$ is defined), $\mathscr{X}(w)$ is a measure except maybe in a set of measure 0 and, as $D_h = \cup_{i=1}^\infty K_h^i$ is increasing, by a known property of measures deduced from their σ-additivity, we can conclude that $\mathscr{X}(\omega)(D) = \lim_i \mathscr{X}(\omega)(K_h^i)$. So, we have that, as a limit of random variables, $\mathscr{X}(D) = \lim_i \mathscr{X}(K_h^i)$ a.s.

Moreover, taking into account that the above is true for all $h \in \{1, \dots, k\}$, we also deduce that $(\mathscr{X}(D_1), \dots, \mathscr{X}(D_k)) = \lim_i (\mathscr{X}(K_1^i), \dots, \mathscr{X}(K_k^i))$. So, to determine the distribution of $(\mathscr{X}(D_1), \dots, \mathscr{X}(D_k))$, it suffices to determine the law of the random vectors $(\mathscr{X}(K_1^i), \dots, \mathscr{X}(K_k^i))$ for all $i \geq 1 \in \mathbb{N}$.

Then, as all these sets are compact, we conclude that we can assume that the sets $D_1, \dots, D_k$ are compact without loss of generality and focus now on proving that the corresponding probabilities p_n are determined by the joint intensities of the process.

To do it, we start by noticing that, by Theorem 2.13, the factorial moments of the random vector $X \coloneqq (\mathscr{X}(D_1), \dots, \mathscr{X}(D_k))$ are $\mathbb{E}(\prod_{i=1}^k \binom{\mathscr{X}(D_i)}{m_i} m_i!)\ \forall m_1, \dots, m_k \geq 0 \in \mathbb{N}$ and can be computed from the joint intensities. Then, equivalently, knowing that the moments of a random vector can be expressed as a linear combination of its factorial moments (something that can be easily shown by induction), we deduce that the joint intensities of $\mathscr{X}$ determine the moments of the random vector.

Now, we observe that, by hypothesis, we know that $\forall 1 \leq i \leq k$, there exist $n_0^i \in \mathbb{N}$ and $c_i > 0$ such that $P(\mathscr{X}(D_i) \geq n) \leq e^{-c_i n}\ \forall n \geq n_0^i$. From here, we can deduce that, if we take $n_0 = \max\{n_0^i \mid 1 \leq i \leq k\}$ and $c = \min\{c_i \mid 1 \leq i \leq k\}$, then, for all $n = (n_1, \dots, n_k) \in \mathbb{N}^k$ verifying $\|n\|_\infty \geq n_0$, we have that $p_n \leq e^{-c\|n\|_\infty}$.

So, in particular, for all $s \in B_{\frac{c}{2\sqrt{k}}}(0) \subseteq \mathbb{R}^k$ and $m \in \mathbb{N}^k$ a multi-index, we conclude that:

$$\begin{gathered}\sum_{n\in\mathbb{N}^k} e^{-s\cdot n} n^m p_n \leq \sum_{n\in\mathbb{N}^k,\ \|n\|_\infty<n_0} e^{-s\cdot n} n^m p_n + \sum_{n\in\mathbb{N}^k,\ \|n\|_\infty\geq n_0} e^{-s\cdot n} n^m p_n \leq \\ C_1 + \sum_{n\in\mathbb{N}^k} e^{\|s\|_2\|n\|_2} n^m e^{-c\|n\|_\infty} \leq C_1 + \sum_{n\in\mathbb{N}^k} e^{\frac{c}{2}\|n\|_\infty} n^m e^{-c\|n\|_\infty} \leq \\ C_1 + \sum_{n\in\mathbb{N}^k} e^{-\frac{c}{2}\|n\|_\infty} \|n\|_\infty^{|m|} \leq C_1 + \sum_{l=0}^\infty e^{-\frac{c}{2}l} l^{|m|} (l+1)^k \leq C\end{gathered}$$

where $C_1 > 0$ and $C > 0$ are positive constants independent on s.

That shows that the previous series converges uniformly in $A \coloneqq B_{\frac{c}{2\sqrt{k}}}(0) \subseteq \mathbb{R}^k$, so we can consider $M : A \longrightarrow \mathbb{R}$ the moment generating function of the random vector X, given by $M(s) = \mathbb{E}(e^{-s\cdot X})\ \forall s \in A$. Notice that we have proved that $M \in \mathcal{C}^\infty(A)$ and it is clear that M depends only on the joint intensities of $\mathscr{X}$, because $\forall s \in A$ we have that $M(s) = \mathbb{E}(e^{-s\cdot X}) = \mathbb{E}(\sum_{l=0}^\infty \frac{(-s\cdot X)^l}{l!}) =$ (by Fubini's theorem) $= \sum_{l=0}^\infty \frac{\mathbb{E}((-s\cdot X)^l)}{l!}$, an expression that can be computed from the moments of X.

Now, we merely need to use the fact that the moment generating function of a random vector determines its law (see for example, page 285 of [4]) to finish the proof. □

Finally, to complete our study of the joint intensities, notice that we still need to clarify in which sense the joint intensities of a process can be thought as the functions that describe, for certain $x_1, \dots, x_k \in E$ pairwise distinct, the probability of finding a point of the process in each of the infinitesimal neighbourhoods $d\mu(x_1), \dots, d\mu(x_k)$. Indeed, using the previous results and recalling that E is a Polish space, so it is metrizable by a certain metric d, we can show the following theorem, that justifies why joint intensities can be viewed in this helpful and intuitive way.

Theorem 2.17. *Let $\mathscr{X}$ be a simple point process in E such that, for any $k \geq 1$, its k-th joint intensity ρ_k exists and is locally bounded. Then, if we denote by $B_r(x) = \{y \in E \,|\, d(x,y) < r\} \subseteq E$ the open ball of centre $x \in E$ and radius $r > 0$, we can conclude that:*

$$\rho_k(x_1, \dots, x_k) = \lim_{\epsilon \to 0^+} \frac{P(\mathscr{X}(B_\epsilon(x_i)) = 1 \text{ for each } 1 \leq i \leq k)}{\prod_{i=1}^k \mu(B_\epsilon(x_i))}$$

a.e. in E^k for $x_1, \dots, x_k$ pairwise distinct.

Proof. To show this result using our own ideas, let $x_1, \dots, x_k \in E$ be any pairwise distinct points, we can start by noticing that, as E can be seen as a metric space with distance d, there exists $\epsilon_0 > 0$ such that $\forall 0 < \epsilon \leq \epsilon_0$, the open balls $B_\epsilon(x_1), \dots, B_\epsilon(x_k)$ are pairwise disjoint. Then, except maybe in a set of measure 0, by the Lebesgue differentiation theorem, we can deduce that:

$$\rho_k(x_1, \dots, x_k) = \lim_{\epsilon \to 0^+} \frac{1}{\prod_{i=1}^k \mu(B_\epsilon(x_i))} \int_{\prod_{i=1}^k B_\epsilon(x_i)} \rho_k(y_1, \dots, y_k) d\mu(y_1) \dots d\mu(y_k) =$$

$$\text{(by the definition of } \rho_k \text{ and considering } \epsilon \leq \epsilon_0 \text{ in the limit)} =$$

$$\lim_{\epsilon \to 0^+} \frac{\mathbb{E}(\prod_{i=1}^k \mathscr{X}(B_\epsilon(x_i)))}{\prod_{i=1}^k \mu(B_\epsilon(x_i))} =$$

$$\lim_{\epsilon \to 0^+} \frac{P(\mathscr{X}(B_\epsilon(x_i)) = 1 \text{ for each } 1 \leq i \leq k)}{\prod_{i=1}^k \mu(B_\epsilon(x_i))} + \lim_{\epsilon \to 0^+} \frac{\mathbb{E}(\prod_{i=1}^k \mathscr{X}(B_\epsilon(x_i)) \mid A^\epsilon) P(A^\epsilon)}{\prod_{i=1}^k \mu(B_\epsilon(x_i))}$$

where we have considered $A^\epsilon := \{\omega \in \Omega \mid \mathscr{X}(\omega)(B_\epsilon(x_i)) \geq 1 \text{ for each } 1 \leq i \leq k,\ \mathscr{X}(\omega)(B_\epsilon(x_j)) \geq 2 \text{ for some } j\}$.

Not only that, but if we define, for every $j \in \{1, \dots, k\}$, the random events $A_j^\epsilon := \{\omega \in \Omega \mid \mathscr{X}(\omega)(B_\epsilon(x_i)) \geq 1 \text{ for each } 1 \leq i \neq j \leq k,\ \mathscr{X}(\omega)(B_\epsilon(x_j)) \geq 2\}$, it is clear that $A^\epsilon = \bigcup_{j=1}^k A_j^\epsilon$, so, in particular, we conclude that:

$$0 \leq \lim_{\epsilon \to 0^+} \frac{\mathbb{E}(\prod_{i=1}^k \mathscr{X}(B_\epsilon(x_i)) \mid A^\epsilon) P(A^\epsilon)}{\prod_{i=1}^k \mu(B_\epsilon(x_i))} \leq \lim_{\epsilon \to 0^+} \sum_{j=1}^k \frac{\mathbb{E}(\prod_{i=1}^k \mathscr{X}(B_\epsilon(x_i)) \mid A_j^\epsilon) P(A_j^\epsilon)}{\prod_{i=1}^k \mu(B_\epsilon(x_i))}$$

As a consequence, to end the proof, let $j \in \{1, \dots, k\}$ arbitrary and fixed, it suffices to show that:

$$\lim_{\epsilon \to 0^+} \frac{\mathbb{E}(\prod_{i=1}^k \mathscr{X}(B_\epsilon(x_i)) \mid A_j^\epsilon) P(A_j^\epsilon)}{\prod_{i=1}^k \mu(B_\epsilon(x_i))} = 0$$

Indeed, to do it, we can observe that, given $n_1, \dots, n_k \in \mathbb{N}$ such that $n_j = 2$ and $n_i = 1\ \forall 1 \leq i \neq j \leq k$, one has that:

$$2\mathbb{E}(\prod_{i=1}^{k}\binom{\mathscr{X}(B_\epsilon(x_i))}{n_i}) = \mathbb{E}((\prod_{i=1}^{k}\mathscr{X}(B_\epsilon(x_i)))\cdot(\mathscr{X}(B_\epsilon(x_j))-1)) =$$
$$\text{(conditioning the above expected values to } A_j^\epsilon \text{ and } (A_j^\epsilon)^c) =$$
$$\mathbb{E}((\prod_{i=1}^{k}\mathscr{X}(B_\epsilon(x_i)))\cdot(\mathscr{X}(B_\epsilon(x_j))-1)\mid A_j^\epsilon)P(A_j^\epsilon)+C_\epsilon =$$
$$\text{(using that } \mathscr{X}(B_\epsilon(x_j))-1\geq 1 \text{ if } A_j^\epsilon \text{ has happened and monotony properties)} =$$
$$\mathbb{E}(\prod_{i=1}^{k}\mathscr{X}(B_\epsilon(x_i))\mid A_j^\epsilon)P(A_j^\epsilon)+C_\epsilon \geq \mathbb{E}(\prod_{i=1}^{k}\mathscr{X}(B_\epsilon(x_i))\mid A_j^\epsilon)P(A_j^\epsilon)$$

where $C_\epsilon > 0 \in \mathbb{R}$ denotes the positive value corresponding to the other term of the expansion, that there is no need to specify.

Finally, combining the above inequality, previous results, and the Lebesgue differentiation theorem, we can focus on the limit and conclude that:

$$0 \leq \lim_{\epsilon\to 0^+}\frac{\mathbb{E}(\prod_{i=1}^{k}\mathscr{X}(B_\epsilon(x_i))\mid A_j^\epsilon)P(A_j^\epsilon)}{\prod_{i=1}^{k}\mu(B_\epsilon(x_i))} \leq \lim_{\epsilon\to 0^+}\frac{\mathbb{E}\left(\prod_{i=1}^{k}\binom{\mathscr{X}(B_\epsilon(x_i))}{n_i}n_i!\right)}{\prod_{i=1}^{k}\mu(B_\epsilon(x_i))} =$$
$$\text{(applying the already proven Theorem 2.13)} =$$
$$\lim_{\epsilon\to 0^+}\frac{1}{\prod_{i=1}^{k}\mu(B_\epsilon(x_i))}\int_{\prod_{i=1}^{k}B_\epsilon(x_i)^{n_i}}\rho_{k+1}(y_1,\dots,y_{k+1})d\mu(y_1)\dots d\mu(y_{k+1}) =$$
$$\lim_{\epsilon\to 0^+}\mu(B_\epsilon(x_j))\frac{1}{\prod_{i=1}^{k}\mu(B_\epsilon(x_i))^{n_i}}\int_{\prod_{i=1}^{k}B_\epsilon(x_i)^{n_i}}\rho_{k+1}(y_1,\dots,y_{k+1})d\mu(y_1)\dots d\mu(y_{k+1})$$

where we can distinguish two different cases.

If $\mu(\{x_j\})=0$, then it is clear that $\lim_{\epsilon\to 0^+}\mu(B_\epsilon(x_j))=0$, something that, combined with the fact that ρ_{k+1} is locally bounded, lets us deduce that the desired limit is 0.

If $\mu(\{x_j\})>0$, then $\mu(B_\epsilon(x_j))$ is bounded for all $\epsilon\leq 1$ and we notice that $\mu^{k+1}(D\times\{x_j\})=\mu^k(D)\mu(\{x_j\})$ for all $D\subset E^k$ measurable. As a consequence, except maybe in a set of measure 0 in E^k, using again the Lebesgue differentiation theorem, we conclude that:

$$0 \leq \lim_{\epsilon\to 0^+}\frac{\mathbb{E}(\prod_{i=1}^{k}\mathscr{X}(B_\epsilon(x_i))\mid A_j^\epsilon)P(A_j^\epsilon)}{\prod_{i=1}^{k}\mu(B_\epsilon(x_i))} \leq \mu(B_1(x_j))\rho_{k+1}(z)=0$$

where $z=(x_1,\dots,x_j,\dots,x_k,x_j)\in E^{k+1}$ and $\rho_{k+1}(z)=0$ because, by definition, we know that $\rho_{k+1}(y_1,\dots,y_{k+1})=0$ if $y_i=y_j$ for some $i\neq j$ between 1 and $k+1$.

That is exactly what we wanted to see, something that shows that the limit is 0 and completes the proof. □

As a comment, notice that the above theorem also provides an alternative proof of the fact that the joint intensities of a simple point process are unique *a.e.*, in the particular case they are locally bounded.

2.2. **Integral kernels.** From here, before introducing the concept of determinantal point processes, it is also worth shifting our attention towards the notion of integral kernels and their associated operators, that are key in the study of these point processes.

Indeed, to do it, we assume that $D \subseteq E$ denotes in this section a compact subset of our locally compact Polish space E and, then, we start by defining what an integral operator is in this context.

Definition 2.18 (Integral operator)**.** *Let $\mathbb{K} \in L^2(D^2, \mu^2)$ be a square integrable function in D^2, we define its associated integral operator $\mathscr{K} : L^2(D, \mu) \longrightarrow L^2(D, \mu)$ as the one given by:*

$$\mathscr{K} f(x) = \int_D \mathbb{K}(x, y) f(y) d\mu(y) \qquad a.e.\ x \in D \text{ and } \forall f \in L^2(D, \mu)$$

Although the above is the general notion of an integral operator in D, we are especially interested in the particular case where our integral kernel $\mathbb{K}$ is Hermitian, that is, $\mathbb{K}(x, y) = \overline{\mathbb{K}(y, x)}$ a.e. for $x, y \in D$. This is the case because, under these assumptions, one can state and show the following proposition.

Proposition 2.19. *Let $\mathbb{K} \in L^2(D^2, \mu^2)$ be an Hermitian kernel and let $\mathscr{K}$ be its associated integral operator, then $\mathscr{K}$ is a well-defined compact and self-adjoint bounded linear operator.*

Proof. Indeed, to show this result, we start by noticing that, given any $f \in L^2(D, \mu)$, we have that:

$$\|\mathscr{K} f\|_2^2 = \int_D |\mathscr{K} f(x)|^2 d\mu(x) \leq \int_D (\int_D |\mathbb{K}(x, y) f(y)| d\mu(y))^2 d\mu(x) \leq$$
$$\int_D (\int_D |\mathbb{K}(x, y)|^2 d\mu(y))(\int_D |f(y)|^2 d\mu(y)) d\mu(x) = \|\mathbb{K}\|_2^2 \|f\|_2^2$$

where we have used the Cauchy-Schwarz inequality. That, combined with the fact that $\mathscr{K}$ is linear because the Lebesgue integral is it, lets us conclude that $\mathscr{K}$ is a well-defined bounded linear operator.

From here, given any $f, g \in L^2(D, \mu)$, we can also deduce that:

$$\langle \mathscr{K} f, g \rangle = \int_D \mathscr{K} f(x) \overline{g(x)} d\mu(x) = \int_D \int_D \mathbb{K}(x, y) f(y) \overline{g(x)} d\mu(y) d\mu(x) =$$
$$\text{(applying Fubini's theorem because the expression is integrable)} =$$
$$\int_D f(y) \int_D \mathbb{K}(x, y) \overline{g(x)} d\mu(x) d\mu(y) = \int_D f(y) \overline{\int_D \mathbb{K}(y, x) g(x) d\mu(x)} d\mu(y) = \langle f, \mathscr{K} g \rangle$$

so we conclude that $\mathscr{K}$ is self-adjoint.

Finally, we observe that, given any $C \subseteq D^2$ measurable set, the function $\mathcal{X}_C$ can be approximated by characteristic functions of the form $\mathcal{X}_{A \times B}$ for some $A, B \subseteq D$ measurable, something that follows directly from the definition of the product measure μ^2 and its measurable sets.

That, combined with the fact that simple functions are dense in $L^2(D^2, \mu^2)$ and that $\mathcal{X}_{A \times B}(x, y) = \mathcal{X}_A(x) \mathcal{X}_B(y)\ \forall (x, y) \in D^2$, shows that the linear combinations of these products of characteristic functions are also dense in $L^2(D^2, \mu^2)$. In particular, $\mathbb{K}$ can be approximated these functions $\{\mathbb{K}_l\}_l$ and, consequently, $\mathscr{K}$ can also be approximated by the corresponding integral operators $\mathscr{K}_l$, that have finite rank, so we conclude that $\mathscr{K}$ is compact. □

Thanks to this result, using the fact that under these hypothesis our integral operator $\mathscr{K}$ is compact and self-adjoint, we can then apply the well-known Spectral theorem for compact self-adjoint operators in a separable Hilbert space (see for example page 365 of [17]), that lets us conclude that there exists $\{\varphi_j\}_j \subseteq L^2(D,\mu)$ a (countable) orthonormal basis of eigenfunctions of $\mathscr{K}$ with corresponding eigenvalues $\{\lambda_j\}_j \subseteq \mathbb{R}$ such that:

$$\mathscr{K} f = \sum_j \lambda_j \langle f, \varphi_j \rangle \varphi_j \quad \forall f \in L^2(D,\mu)$$

where $\langle\, , \rangle$ denotes the usual inner product in $L^2(D,\mu)$.

Consequently, for the remainder of this section, we will focus our attention mainly on this particular type of integral operators. From here, before moving on, we need to define some common properties these operators can satisfy.

Definition 2.20 (Trace class operator)**.** *Let $\mathbb{K} \in L^2(D^2,\mu^2)$ be an Hermitian kernel, $\mathscr{K}$ its associated operator and $\{\lambda_j\}_j$ its eigenvalues, then we say that $\mathscr{K}$ is of trace class if:*

$$\sum_j |\lambda_j| < \infty$$

Definition 2.21 (Non-negative definite operator)**.** *Let $\mathbb{K} \in L^2(D^2,\mu^2)$ be an Hermitian kernel and let $\mathscr{K}$ be its associated integral operator, then we say that $\mathscr{K}$ is non-negative definite (or positive semi-definite) if $\langle \mathscr{K} f, f\rangle \geq 0\ \forall f \in L^2(D,\mu)$.*

Notice that, actually, the above properties could be extended and considered for any linear bounded operator in a separable Hilbert space, but, in this project, we will only need to consider them in the described context.

Not only that, but also, as a comment, observe that with the already introduced notations we have that $\mathscr{K}\varphi_i = \lambda_i \varphi_i\ \forall i$ and also, by the continuity of the inner product, $\langle \mathscr{K} f, f\rangle = \sum_j \lambda_j \langle f, \varphi_j\rangle\langle \varphi_j, f\rangle = \sum_j \lambda_j |\langle f, \varphi_j\rangle|^2\ \forall f \in L^2(D,\mu)$, so we conclude that $\mathscr{K}$ is non-negative definite if, and only if, all its eigenvalues λ_j are non-negative.

Finally, once we have recalled these definitions, we are ready to introduce the main result of this section, that allows us to deduce many properties of our integral kernel $\mathbb{K}$ from the ones that its associated operator $\mathscr{K}$ verifies.

Theorem 2.22. *Let $\mathbb{K} \in L^2(D^2,\mu^2)$ be an Hermitian kernel and let $\mathscr{K}$ be its associated integral operator that has $\{\varphi_j\}_j$ as a basis of eigenfunctions with eigenvalues $\{\lambda_j\}_j$. Then, if $\mathscr{K}$ is of trace class and non-negative definite, we can conclude that:*

(1) The series $\sum_j \lambda_j \varphi_j(x)\overline{\varphi_j}(y)$ converges absolutely $\mu - a.e.$ for $x, y \in D$.
(2) The same series converges in $L^2(D,\mu)$ as a function of y $\mu - a.e.$ for $x \in D$ and vice versa. Not only that, but the series is convergent in $L^2(D^2,\mu^2)$.
(3) The integral kernel $\mathbb{K}$ is given by $\mathbb{K}(x,y) = \sum_j \lambda_j \varphi_j(x)\overline{\varphi_j}(y)$ $\mu^2 - a.e.$ for $(x,y) \in D^2$.
(4) If we redefine $\mathbb{K}$ with the above expression, then the function $f : D^k \longrightarrow \mathbb{C}$ given by $f(x_1,\dots,x_k) = \det(\mathbb{K}(x_i,x_j))_{i,j\leq k}$ $\mu^k - a.e$ for $(x_1,\dots,x_k) \in D^k$ is well-defined and integrable on D^k.

(5) *Finally, redefining* $\mathbb{K}$ *as before, we also have that* $f(x_1, \dots, x_k) \geq 0 \; \forall k \geq 1$ *and* $\mu^k - a.e.$ *for* $(x_1, \dots, x_k) \in D^k$.

Proof. In this situation, to prove the theorem, we start by noticing that if the orthonormal basis $\{\varphi_j\}_j$ is finite, then most of the claims are trivial and the ones that are not can be shown in a similar way to the infinite case. So, without loss of generality and reindexing the sequences, we can assume that $\{\varphi_j\}_j = \{\varphi_j\}_{j \geq 1 \in \mathbb{N}}$.

That being said, focusing our attention on the first claim, we recall that to show that the series is absolutely convergent, it suffices to deduce that $\sum_j |\lambda_j \varphi_j(x)\varphi_j(y)| < \infty \; \mu - a.e.$ for $x, y \in D$. With that in mind, we can observe that, by the Cauchy-Schwarz inequality, $(\sum_j |\lambda_j \varphi_j(x)\varphi_j(y)|)^2 \leq (\sum_j |\lambda_j||\varphi_j(x)|^2)(\sum_j |\lambda_j||\varphi_j(y)|^2) \; \forall x, y \in D$.

From here, we notice that:

$$\int_D \sum_j |\lambda_j||\varphi_j(x)|^2 d\mu(x) = (\text{by Tonelli's theorem}) = \sum_j \int_D |\lambda_j||\varphi_j(x)|^2 d\mu(x) = \sum_j |\lambda_j| \int_D |\varphi_j(x)|^2 d\mu(x) = \sum_j |\lambda_j| < \infty$$

because we are assuming that the operator is of trace class. Consequently, we can deduce that $\sum_j |\lambda_j||\varphi_j(x)|^2 < \infty \; \mu - a.e.$ for $x \in D$. This last observation lets us conclude the desired result.

Now, regarding the second claim, we can start by considering, for all $N \geq 1 \in \mathbb{N}$, the partial sums $S_N(x, y) = \sum_{j \leq N} \lambda_j \varphi_j(x)\overline{\varphi_j}(y)$ of our series and then, as functions of y and given any $1 \leq N < M$, we can observe that:

$$\|S_M(x, \cdot) - S_N(x, \cdot)\|_2^2 \leq \int_D (\sum_{N+1 \leq j \leq M} |\lambda_j \varphi_j(x)||\varphi_j(y)|)^2 d\mu(y) \leq (\text{as before}) \leq$$
$$(\sum_{N+1 \leq j \leq M} |\lambda_j||\varphi_j(x)|^2) \int_D \sum_{N+1 \leq j \leq M} |\lambda_j||\varphi_j(y)|^2 d\mu(y) =$$
$$(\sum_{N+1 \leq j \leq M} |\lambda_j||\varphi_j(x)|^2)(\sum_{N+1 \leq j \leq M} |\lambda_j|) \leq (\sum_{N+1 \leq j} |\lambda_j||\varphi_j(x)|^2)(\sum_{N+1 \leq j} |\lambda_j|)$$

that goes to 0 when $N \longrightarrow \infty \; \mu - a.e.$ for $x \in D$, because the series are convergent $\mu - a.e.$ for $x \in D$ as we have already seen. Then, for these values of x the original series is Cauchy in $L^2(D, \mu)$ as a function of y, so it is convergent.

Not only that, but it is also clear that we could have exchanged the roles of x and y above to show the analogous result. Moreover, in $L^2(D^2, \mu^2)$, we deduce that:

$$\|S_M - S_N\|_2^2 \leq \int_{D^2} (\sum_{N+1 \leq j \leq M} |\lambda_j \varphi_j(x)||\varphi_j(y)|)^2 d\mu(x) d\mu(y) \leq$$
$$\int_D \int_D (\sum_{N+1 \leq j \leq M} |\lambda_j||\varphi_j(x)|^2)(\sum_{N+1 \leq j \leq M} |\lambda_j||\varphi_j(y)|^2) d\mu(x) d\mu(y) =$$
$$(\int_D \sum_{N+1 \leq j \leq M} |\lambda_j||\varphi_j(x)|^2 d\mu(x))^2 = (\sum_{N+1 \leq j \leq M} |\lambda_j|)^2 \leq (\sum_{N+1 \leq j} |\lambda_j|)^2$$

something that proves the series $\sum_j \lambda_j \varphi_j(x)\overline{\varphi_j}(y)$ is Cauchy and also convergent in $L^2(D^2, \mu^2)$, as a function of $(x, y) \in D^2$.

Moving on, except on a set of μ^2-measure 0 in D^2, given any $f \in L^2(D, \mu)$, we can observe that:

$$\langle \mathbb{K}(x, \cdot), \overline{f} \rangle = \mathscr{K} f(x) = \sum_j \lambda_j \langle f, \varphi_j \rangle \varphi_j(x) = \sum_j \lambda_j \int_D f(y)\overline{\varphi_j}(y)\varphi_j(x) d\mu(y) =$$
$$(\text{by Fubini's theorem}) = \int_D f(y) \sum_j \lambda_j \varphi_j(x)\overline{\varphi_j}(y) d\mu(y) = \langle \sum_j \lambda_j \varphi_j(x)\overline{\varphi_j}(\cdot), \overline{f} \rangle$$

something that shows that $\mathbb{K}(x,y) = \sum_j \lambda_j \varphi_j(x)\overline{\varphi_j}(y)$, that is exactly what we wanted to conclude. Notice that, to apply Fubini's theorem and exchange the order of the integrals, we have used that, in fact, the series $\sum_j |\lambda_j \varphi_j(x)\varphi_j(y)|$ also converges in $L^2(D,\mu)$ as a function of y by the same arguments as before, so:

$$\int_D \sum_j |f(y)\lambda_j\varphi_j(x)\overline{\varphi_j}(y)|d\mu(y) = \int_D |f(y)|(\sum_j |\lambda_j\varphi_j(x)\varphi_j(y)|)d\mu(y) \leq \|f\|_2\| \sum_j |\lambda_j\varphi_j(x)\varphi_j(\cdot)|\|_2 < \infty$$

Now, redefining $\mathbb{K}$ using the corresponding series, notice $\mathbb{K}(x,y)$ is well-defined $\mu - a.e.$ for $x, y \in D$, because we have already seen that this is the case for the series. In particular, the function f given by $f(x_1,\dots,x_k) = \det(\mathbb{K}(x_i,x_j))_{i,j\leq k} =$ (applying the definition of the determinant) $= \sum_{\sigma\in S_k} \operatorname{sgn}(\sigma) \prod_{i=1}^k \mathbb{K}(x_i, x_{\sigma(i)})$ is well-defined $\mu - a.e.$ for $x_1,\dots,x_k \in D$, so it is also well-defined in D^k μ^k-almost everywhere (recall that the product of sets with measure 0 has measure 0 in product measure).

From here, using this redefinition of $\mathbb{K}$ as a series, that we already know is absolutely convergent $\mu - a.e.$ for $x, y \in D$ so we can rearrange it freely, we proceed develop the expression of the determinant function f and conclude that, for all $(x_1,\dots,x_k) \in D^k$ except maybe on a set of μ^k-measure 0:

$$f(x_1,\dots,x_k) = \det(\mathbb{K}(x_i,x_j))_{i,j\leq k} = \sum_{\sigma\in S_k} \operatorname{sgn}(\sigma) \prod_{i=1}^k \mathbb{K}(x_i, x_{\sigma(i)}) =$$
$$\sum_{\sigma\in S_k} \operatorname{sgn}(\sigma) \prod_{i=1}^k \sum_j \lambda_j\varphi_j(x_i)\overline{\varphi_j}(x_{\sigma(i)}) =$$
$$\sum_{\sigma\in S_k} \operatorname{sgn}(\sigma) \sum_{j_1,\dots,j_k} \prod_{i=1}^k \lambda_{j_i}\varphi_{j_i}(x_i)\overline{\varphi_{j_i}}(x_{\sigma(i)}) =$$
$$\sum_{j_1,\dots,j_k} \sum_{\sigma\in S_k} \operatorname{sgn}(\sigma)(\prod_{i=1}^k \lambda_{j_i}\varphi_{j_i}(x_i))(\prod_{i=1}^k \overline{\varphi_{j_i}}(x_{\sigma(i)})) =$$
$$\sum_{j_1,\dots,j_k} \prod_{i=1}^k \lambda_{j_i}\varphi_{j_i}(x_i) \sum_{\sigma\in S_k} \operatorname{sgn}(\sigma) \prod_{i=1}^k \overline{\varphi_{j_i}}(x_{\sigma(i)}) =$$
$$\sum_{j_1,\dots,j_k} \det(\overline{\varphi_{j_i}}(x_h))_{i,h\leq k} \prod_{i=1}^k \lambda_{j_i}\varphi_{j_i}(x_i) =$$
$$\sum_{j_1,\dots,j_k \text{pairwise distinct}} \det(\overline{\varphi_{j_i}}(x_h))_{i,h\leq k} \prod_{i=1}^k \lambda_{j_i}\varphi_{j_i}(x_i) =$$
$$\sum_{j_1<\dots<j_k} \sum_{\sigma\in S_k} \det(\overline{\varphi_{\sigma(j_i)}}(x_h))_{i,h\leq k} \prod_{i=1}^k \lambda_{\sigma(j_i)}\varphi_{\sigma(j_i)}(x_i) =$$
$$\sum_{j_1<\dots<j_k} (\prod_{i=1}^k \lambda_{j_i}) \det(\overline{\varphi_{j_i}}(x_h))_{i,h\leq k} \sum_{\sigma\in S_k} \operatorname{sgn}(\sigma) \prod_{i=1}^k \varphi_{\sigma(j_i)}(x_i) =$$
$$\sum_{j_1<\dots<j_k} (\prod_{i=1}^k \lambda_{j_i}) \det(\varphi_{j_h}(x_i))_{i,h\leq k} \det(\overline{\varphi_{j_i}}(x_h))_{i,h\leq k}$$

Not only that, but now, using these expressions, we can also deduce that f is integrable in D^k. Indeed, we have that:

$$\int_{D^k} |f(x_1,\dots,x_k)|d\mu(x_1)\dots d\mu(x_k) \leq$$
$$\int_{D^k} \sum_{\sigma\in S_k} \sum_{j_1,\dots,j_k} \prod_{i=1}^k \lambda_{j_i} |\varphi_{j_i}(x_i)\varphi_{j_i}(x_{\sigma(i)})|d\mu(x_1)\dots d\mu(x_k) =$$
$$\sum_{\sigma\in S_k} \sum_{j_1,\dots,j_k} (\prod_{i=1}^k \lambda_{j_i}) \int_{D^k} \prod_{i=1}^k |\varphi_{j_i}(x_i)\varphi_{j_i}(x_{\sigma(i)})|d\mu(x_1)\dots d\mu(x_k) =$$
$$\sum_{\sigma\in S_k} \sum_{j_1,\dots,j_k} (\prod_{i=1}^k \lambda_{j_i}) \prod_{i=1}^k \int_D |\varphi_{j_i}(x_i)\varphi_{j_{\sigma^{-1}(i)}}(x_i)|d\mu(x_i) \leq$$
(using the Cauchy-Schwarz inequality and the orthonormality of $\{\varphi_j\}_j$) $\leq$
$$\sum_{\sigma\in S_k} \sum_{j_1,\dots,j_k} \prod_{i=1}^k \lambda_{j_i} = \sum_{\sigma\in S_k} (\sum_{j_1} \lambda_{j_1}) \dots (\sum_{j_k} \lambda_{j_k}) = \sum_{\sigma\in S_k} (\sum_j \lambda_j)^k < \infty$$

because we know that our operator $\mathscr{K}$ is of trace class.

Finally, focusing our attention on the last claim, given any $k \geq 1$ and $(x_1, \ldots, x_k) \in D^k$, then, except maybe in a set of μ^k-measure 0, we can directly observe that:

$$f(x_1, \ldots, x_k) = \sum_{j_1 < \cdots < j_k} \left(\prod_{i=1}^k \lambda_{j_i}\right) \det(\varphi_{j_h}(x_i))_{i,h \leq k} \det(\overline{\varphi_{j_i}}(x_h))_{i,h \leq k} =$$
$$\sum_{j_1 < \cdots < j_k} \left(\prod_{i=1}^k \lambda_{j_i}\right) \det(\varphi_{j_h}(x_i))_{i,h \leq k} \overline{\det(\varphi_{j_i}(x_h))_{i,h \leq k}} =$$
$$\sum_{j_1 < \cdots < j_k} \left(\prod_{i=1}^k \lambda_{j_i}\right) |\det(\varphi_{j_h}(x_i))_{i,h \leq k}|^2 \geq 0$$

that is exactly what we wanted to show, something that completes the proof. □

Before ending these preliminaries, notice that all of the above in this subsection is valid for $D \subseteq E$ a compact subset of E. Nevertheless, in an attempt to weaken our assumptions and extend our framework to the whole space E, we can consider $\mathbb{K} \in L^2_{\text{loc}}(E^2, \mu^2)$ to be a locally square integrable and Hermitian kernel. Then, in a similar way to what we did before, we can also define the $\mathscr{K}$, an associated integral operator, given by:

$$\mathscr{K} f(x) = \int_E \mathbb{K}(x, y) f(y) d\mu(y) \quad \text{a.e. } x \in E$$

and for all $f \in L^2(E, \mu)$ with a.e. compact support.

In this way, given $D \subseteq E$ compact, we can consider $\mathscr{K}_D \in \mathcal{L}(L^2(D, \mu))$ the restriction of $\mathscr{K}$ to D, defined by:

$$\mathscr{K}_D f(x) = \int_D \mathbb{K}(x, y) f(y) d\mu(y) \quad \text{a.e. } x \in D \text{ and } \forall f \in L^2(D, \mu)$$

Then, if we assume that $\mathscr{K}$ is locally of trace class ($\mathscr{K}_D$ is trace class $\forall D \subseteq E$ compact) and non-negative definite ($\langle \mathscr{K} f, f \rangle \geq 0$ for all $f \in L^2(E, \mu)$ with a.e. compact support), the restrictions of $\mathbb{K}$ and $\mathscr{K}$ to any compact subset D of E (recall that we assume E is locally compact) will verify everything we have already seen in this subsection, something that will be enough for our study of determinantal point processes.

3. Definition of determinantal point processes

Once we have recalled the necessary concepts, we are ready to define the notion of a determinantal point process, which is central to this project.

Definition 3.1 (Determinantal point process)**.** *Let $\mathscr{X}$ be a simple point process in E and let $\mathbb{K} : E^2 \longrightarrow \mathbb{C}$ be a locally square integrable and Hermitian kernel such that $\mathscr{K}$ is locally of trace class and non-negative definite, then we say that $\mathscr{X}$ is a determinantal point process (DPP) with kernel $\mathbb{K}$ if its joint intensities w.r.t μ exist and are given by:*

$$\rho_k(x_1, \dots, x_k) = \det(\mathbb{K}(x_i, x_j))_{i,j \leq k}$$

for all $k \geq 1$ and $\mu^k - a.e.$ for $(x_1, \dots, x_k) \in E^k$.

Focusing on this definition and following again Chapter 4.2 of [11], we start by providing some remarks on it, trying to justify that, thanks to our hypothesis, it is coherent and makes sense to consider these determinants as the joint intensities of a point process.

Remarks 3.2. First of all, continuing with the notations used in the definition, notice that we already know that the expression $\det(\mathbb{K}(x_i, x_j))_{i,j \leq k}$ is well-defined for all $k \geq 1$ and $\mu^k - a.e.$ for $(x_1, \dots, x_k) \in E^k$, because we showed that this was the case in every compact set D^k with $D \subseteq E$ compact (Theorem 2.22) and E is a locally compact space.

Not only that, but we have also proved that $\det(\mathbb{K}(x_i, x_j))_{i,j \leq k}$ is locally integrable in E^k, because it is integrable on every compact set D^k with $D \subseteq E$ compact. This a property that the joint intensities must verify by definition, recalling that their integrals on certain compact sets are related, as we already know, to some expected values given by the point process.

Moreover, recall that, by definition, the joint intensities of a point process have to be non-negative, something that is coherent with the proposed notion, because we have already shown that, under these hypothesis, $\det(\mathbb{K}(x_i, x_j))_{i,j \leq k} \geq 0$.

Finally, as a last remark, notice that at this point we do not have any guarantees about the existence or uniqueness of these DPPs, something that will be discussed in detail in the following section.

That being said, it is also worth mentioning that the above is not the most general definition one can consider for determinantal point processes, since for example the hypothesis of $\mathbb{K}$ being Hermitian is actually not necessary and could be removed (as done in [11]). Nevertheless, unless otherwise stated, we will restrict ourselves to the previously defined case.

From here, once we have clarified and explained the previous definition, with the aim of fixing ideas, now we revisit one of the examples briefly mentioned in the introduction, filling in the missing details by ourselves.

Example 3.3. Indeed, let $A_1, \dots, A_n$ be any number of pairwise disjoint open subsets of E of positive and finite measure. Then, in each of them, we can take a random point

uniformly distributed, that is, for all $i \in \{1, \dots, n\}$, we consider X_i a random variable that follows an uniform distribution in the open subset A_i, such that $X_1, \dots, X_n$ are independent.

In this situation, we can also consider the corresponding simple point process $\mathscr{X} = \sum_{i=1}^{n} \delta_{X_i}$ defined in $A \coloneqq \bigcup_{i=1}^{n} A_i$ (a locally compact Polish space equipped the appropriate Radon measure), that we can identify with the set of random points we have chosen. Then, we have that $\mathscr{X}$ is a DPP with kernel $\mathbb{K} : A^2 \longrightarrow \mathbb{C}$ given by:

$$\mathbb{K}(x,y) = \sum_{i=1}^{n} \frac{\mathcal{X}_{A_i}(x)\mathcal{X}_{A_i}(y)}{\mu(A_i)} = \begin{cases} \frac{1}{\mu(A_i)} & \text{if } x, y \in A_i \text{ for some } 1 \leq i \leq n \\ 0 & \text{otherwise} \end{cases} \qquad \forall x, y \in A$$

as we will see.

To do it, we start by noticing that, as A has finite measure, $\mathbb{K}$ is square integrable in A^2 because it is bounded. Moreover, it is easy to see that $\mathbb{K}$ is Hermitian and, using its expression as a sum and the fact that the functions $\mathcal{X}_{A_i}/\sqrt{\mu(A_i)}$ clearly form an orthonormal system, one can conclude that the associated integral operator $\mathscr{K}$ is locally of trace class and non-negative definite, using similar arguments to the ones that appeared in the preliminaries. Finally, studying the corresponding determinants, given any $x_1, \dots, x_k \in A$, we can distinguish two cases:

If there exists $1 \leq i \neq j \leq k$ such that $x_i, x_j \in A_h$ for some h, then it is clear $\mathbb{K}(x_l, x_i) = \mathbb{K}(x_l, x_j)\ \forall l \in \{1, \dots, k\}$, so the matrix has two columns that are equal and $\det(\mathbb{K}(x_i, x_j))_{i,j \leq k} = 0$

If there exists an injective map $f : \{1, \dots, k\} \longrightarrow \{1, \dots, n\}$ such that $x_i \in A_{f(i)}\ \forall 1 \leq i \leq k$, then it is also easy to deduce that $\mathbb{K}(x_i, x_j) = 0\ \forall i \neq j$, because the points belong to different open subsets A_h, so the matrix is diagonal and $\det(\mathbb{K}(x_i, x_j))_{i,j \leq k} = \prod_{i=1}^{k} \mathbb{K}(x_i, x_i) = \frac{1}{\prod_{i=1}^{k} \mu(A_{f(i)})}$.

From here, focusing on the joint intensities of this point process $\mathscr{X}$ (that exist, something that can be shown with similar arguments to the ones used in the proof of Proposition 2.11), we can apply their alternative expression and recall that, for any $k \geq 1$ and any $x_1, \dots, x_k \in A$ except maybe in a set of measure 0, one has:

$$\rho_k(x_1, \dots, x_k) = \lim_{\epsilon \to 0^+} \frac{P(\mathscr{X}(B_\epsilon(x_i)) = 1 \text{ for each } 1 \leq i \leq k)}{\prod_{i=1}^{k} \mu(B_\epsilon(x_i))}$$

if the points are pairwise distinct. Then, we can distinguish two different cases, as before:

If there exists $1 \leq i \neq j \leq k$ such that $x_i, x_j \in A_h$ for some h, then it is clear that, if the points are pairwise distinct, $P(\mathscr{X}(B_\epsilon(x_i)) = 1 \text{ for each } 1 \leq i \leq k) = 0$ for $\epsilon > 0$ small enough, because there is only one random point in A_h and $B_\epsilon(x_i) \cup B_\epsilon(x_j) \subseteq A_h$, $B_\epsilon(x_i) \cap B_\epsilon(x_j) = \emptyset$ if we take $\epsilon > 0$ small enough. So we conclude that $\rho_k(x_1, \dots, x_k) = 0 = \det(\mathbb{K}(x_i, x_j))_{i,j \leq k}$, the same that happens by definition if the points are not pairwise distinct.

If there exists an injective map $f : \{1, \dots, k\} \longrightarrow \{1, \dots, n\}$ such that $x_i \in A_{f(i)}$ $\forall 1 \leq i \leq k$, then it is also easy to deduce that:

$$\rho_k(x_1, \dots, x_k) = \lim_{\epsilon \to 0^+} \frac{P(\mathscr{X}(B_\epsilon(x_i))=1 \text{ for each } 1 \leq i \leq k)}{\prod_{i=1}^k \mu(B_\epsilon(x_i))} =$$
$$\lim_{\epsilon \to 0^+} \prod_{i=1}^k \frac{P(\mathscr{X}(B_\epsilon(x_i))=1)}{\mu(B_\epsilon(x_i))} = \lim_{\epsilon \to 0^+} \prod_{i=1}^k \frac{P(X_i \in B_\epsilon(x_i))}{\mu(B_\epsilon(x_i))} =$$
$$\lim_{\epsilon \to 0^+} \prod_{i=1}^k \frac{\mu(B_\epsilon(x_i))}{\mu(B_\epsilon(x_i))\mu(A_{f(i)})} = \prod_{i=1}^k \frac{1}{\mu(A_{f(i)})} = \det(\mathbb{K}(x_i, x_j))_{i,j \leq k}$$

That completes the proof and shows $\mathscr{X}$ is indeed a DPP with kernel $\mathbb{K}$.

Now, before moving on and following [13], one can notice that in the discrete case the definition of a determinantal point process can be particularized in a nice and more understandable way, thanks to the following statement, that we can easily show.

Proposition 3.4. *Let Λ be a countable set equipped with the counting measure μ and let $\mathscr{X}$ be a determinantal point process in Λ with kernel $\mathbb{K}$, then, if we consider $\mathscr{X}$ as a random subset of Λ, we have that:*

$$P(\{x_1, \dots, x_k\} \subseteq \mathscr{X}) = \det(\mathbb{K}(x_i, x_j))_{i,j \leq k}$$

for all $k \geq 1$ and $x_1, \dots, x_k \in \Lambda$ pairwise distinct.

Proof. Indeed, continuing with the notations introduced above, we can directly apply the definitions and the results we have already seen in the preliminaries section to conclude that:

$$P(\{x_1, \dots, x_k\} \subseteq \mathscr{X}) = P(x_1 \in \mathscr{X}, \dots, x_k \in \mathscr{X}) =$$
$$P(\mathscr{X}(\{x_i\}) = 1 \text{ for each } 1 \leq i \leq k) = \lim_{\epsilon \to 0^+} \frac{P(\mathscr{X}(B_\epsilon(x_i))=1 \text{ for each } 1 \leq i \leq k)}{\prod_{i=1}^k \mu(B_\epsilon(x_i))} =$$
$$\rho_k(x_1, \dots x_k) = \det(\mathbb{K}(x_i, x_j))_{i,j \leq k}$$

that is what we wanted to see. □

Notice that in fact, in this last proof, we have actually managed to observe that $P(\{x_1, \dots, x_k\} \subseteq \mathscr{X}) = \rho_k(x_1, \dots, x_k)$, which implies that in the discrete case we could use these equalities $P(\{x_1, \dots, x_k\} \subseteq \mathscr{X}) = \det(\mathbb{K}(x_i, x_j))_{i,j \leq k}$ for all $k \geq 1$ and $x_1, \dots, x_k \in \Lambda$ pairwise distinct directly to define what a determinantal point process is, without the need to introduce first the notion of the joint intensities of a point process (as done for example in [13]).

4. Existence and uniqueness of DPPs

As we have explained it in the previous section, it remains to be studied in which cases the determinantal point processes exist. In other words, we still need to check for which kernels $\mathbb{K}$ verifying the hypothesis of the definition, a determinantal point process with such a kernel actually exists. Not only that, but if they do exist, we would also need to analyse whether they are unique, that is, whether being a determinantal point process with a particular kernel determines the distribution of the point process. To do it, during this section, we will mainly follow Chapters 4.2 and 4.5 from our reference [11] unless otherwise stated.

4.1. **Existence of DPPs.**

With that in mind, we will first try to answer the question regarding the existence of these processes. To do it, we start by stating and showing a useful property about the restriction of determinantal point processes that we will need for this section.

Lemma 4.1 (Restriction of DPPs)**.** *Let $\mathscr{X}$ be a determinantal point process in E with kernel $\mathbb{K} : E^2 \longrightarrow \mathbb{C}$ and let $D \in \mathcal{B}(E)$ be any Borel set. Then, we have that the restriction of $\mathscr{X}$ to D, $\mathscr{X} \cap D$, is a determinantal point process with kernel $\mathbb{K}_{|D^2}$.*

Proof. Indeed, continuing with the above notations and filling in the details by ourselves, it is clear that $\mathscr{X} \cap D$ exists and it is a simple point process, because it is the process defined by taking the (random) points of $\mathscr{X}$ that are in D.

Moreover, it is also easy to see that $\mathbb{K}_{|D^2}$ is locally square integrable and Hermitian, since the same is true for $\mathbb{K}$. Not only that, but we also have that the integral operator defined by $\mathbb{K}_{|D^2}$ can be seen as the restriction of $\mathscr{K}$ to the functions of $L^2(D, \mu)$ with a.e. compact support, so it is also locally of trace class and non-negative definite. So, the above shows that, in our context, it makes sense to check if $\mathscr{X} \cap D$ is a determinantal point process with kernel $\mathbb{K}_{|D^2}$.

Now, given any $k \geq 1$ and $D_1, \ldots, D_k \in \mathcal{B}(D)$ pairwise disjoint sets, it is clear that for all $1 \leq i \leq k$ we have that $D_i \in \mathcal{B}(E)$ and then, we can also see that:

$$\mathbb{E}(\textstyle\prod_{i=1}^{k}(\mathscr{X} \cap D)(D_i)) =$$
$$\text{(using that } \mathscr{X} \text{ and } \mathscr{X} \cap D \text{ coincide in } D_i \text{ for all } 1 \leq i \leq k) =$$
$$\mathbb{E}(\textstyle\prod_{i=1}^{k}\mathscr{X}(D_i)) =$$
$$\text{(by the definition of the joint intensities and the fact that } \mathscr{X} \text{ is a DPP}) =$$
$$\textstyle\int_{\prod_{i=1}^{k} D_i} \det(\mathbb{K}(x_j, x_j))_{i,j \leq k} d\mu(x_1) \ldots d\mu(x_k) =$$
$$\textstyle\int_{\prod_{i=1}^{k} D_i} \det(\mathbb{K}_{|D^2}(x_j, x_j))_{i,j \leq k} d\mu(x_1) \ldots d\mu(x_k)$$

That shows, by definition, that the joint intensities of $\mathscr{X} \cap D$ exist and are given by $\rho_k(x_1, \ldots, x_k) = \det(\mathbb{K}_{|D^2}(x_j, x_j))_{i,j \leq k}$ $\forall k \geq 1,\ x_1, \ldots, x_k \in D$. So, that lets us conclude that $\mathscr{X} \cap D$ is a determinantal point process with kernel $\mathbb{K}_{|D^2}$, that is what we wanted to deduce. That completes the proof. □

From here, to study the existence of determinantal point processes, we begin by focusing our attention on a particular type of functions, for which it is relatively easy to show the result, that is, prove that there exist DPPs with them as kernels.

Proposition 4.2. *Let $\{\varphi_k\}_{k=1}^n$ be a finite orthonormal system of $L^2(E,\mu)$ and consider the function $\mathbb{K}: E^2 \longrightarrow \mathbb{C}$ given by $\mathbb{K}(x,y) = \sum_{k=1}^n \varphi_k(x)\overline{\varphi_k}(y)\ \forall (x,y) \in E^2$, then we have that there exists $\mathscr{X}$ a determinantal point process with kernel $\mathbb{K}$ that has n points almost surely.*

Proof. To do it, continuing with the above notations and changing slightly the proof given in [11], we can start by observing that $\mathbb{K}$ is square integrable in E^2, because the products $\varphi_k(x)\overline{\varphi_k}(y)$ that define it are also square integrable, as functions of $(x,y) \in E^2$. Not only that, but also, for all $x, y \in E$, we deduce that $\mathbb{K}(x,y) = \sum_{k=1}^n \varphi_k(x)\overline{\varphi_k}(y) = \overline{\sum_{k=1}^n \overline{\varphi_k}(x)\varphi_k(y)} = \overline{\mathbb{K}(y,x)}$, so $\mathbb{K}$ is an Hermitian kernel. That, combined with the fact that its associated integral operator $\mathscr{K}: L^2(E,\mu) \longrightarrow L^2(E,\mu)$ is given by:

$$\mathscr{K}f(x) = \int_E \mathbb{K}(x,y)f(y)d\mu(y) = \int_E \sum_{k=1}^n \varphi_k(x)\overline{\varphi_k}(y)f(y)d\mu(y) = \sum_{k=1}^n \varphi_k(x)\int_E \overline{\varphi_k}(y)f(y)d\mu(y) = \sum_{k=1}^n \langle f, \varphi_k\rangle \varphi_k(x)$$

a.e. $x \in E$ and $\forall f \in L^2(E,\mu)$, shows that the kernel $\mathbb{K}$ is in the hypothesis of our definition of determinantal point processes. Indeed, extending $\{\varphi_k\}_{k=1}^n$ to an orthonormal basis of $L^2(E,\mu)$ and having in mind the Spectral theorem, we notice that $\mathscr{K}$ is locally of trace class and positive semi-definite, diagonalizing in this basis and having 1 and 0 as its only eigenvalues.

From here, let $x_1, \ldots, x_n \in E$ be any points in E, then we have already seen that $\det(\mathbb{K}(x_i,x_j))_{i,j\leq n} \geq 0$ and, in fact, $(\mathbb{K}(x_i,x_j))_{i,j\leq n} = (\varphi_j(x_i))_{i,j\leq n}(\overline{\varphi_i}(x_j))_{i,j\leq n}$ by the definition of the matrix product. Then, using this last observation and the properties of the determinant, we can also deduce that:

$$\begin{gathered}
\int_{E^n} \det(\mathbb{K}(x_i,x_j))_{i,j\leq n} d\mu(x_1)\ldots d\mu(x_n) = \\
\int_{E^n} \det(\varphi_j(x_i))_{i,j\leq n} \det(\overline{\varphi_j}(x_i))_{i,j\leq n} d\mu(x_1)\ldots d\mu(x_n) = \\
\int_{E^n} \Big(\sum_{\sigma\in S_n} \prod_{i=1}^n \varphi_{\sigma(i)}(x_i)\Big)\Big(\sum_{\tau\in S_n} \prod_{i=1}^n \overline{\varphi_{\tau(i)}}(x_i)\Big) d\mu(x_1)\ldots d\mu(x_n) = \\
\sum_{\sigma\in S_n}\sum_{\tau\in S_n} \int_{E^n} \prod_{i=1}^n \varphi_{\sigma(i)}(x_i)\overline{\varphi_{\tau(i)}}(x_i) d\mu(x_1)\ldots d\mu(x_n) = \\
\sum_{\sigma\in S_n}\sum_{\tau\in S_n} \prod_{i=1}^n \int_E \varphi_{\sigma(i)}(x_i)\overline{\varphi_{\tau(i)}}(x_i) d\mu(x_i) = \\
\text{(the product of integrals is 0 if } \sigma \neq \tau \text{ because of the orthogonality)} = \\
\sum_{\sigma\in S_n} \prod_{i=1}^n \int_E \varphi_{\sigma(i)}(x_i)\overline{\varphi_{\sigma(i)}}(x_i) d\mu(x_i) = \\
\text{(using again that } \{\varphi_k\}_{k=1}^n \text{ is an orthonormal system)} = \\
\sum_{\sigma\in S_n} \prod_{i=1}^n 1 = \sum_{\sigma\in S_n} 1 = n!
\end{gathered}$$

so we conclude that the previous determinant of our kernel $\mathbb{K}$ divided by $n!$ is a probability density.

In detail, that means that there exists $(X_1, \ldots, X_n): \Omega \longrightarrow E^n$ a random vector defined in some probability space Ω such that the function $f: E^n \longrightarrow [0,\infty)$, given

by $f(x_1,\dots,x_n)=\frac{1}{n!}\det(\mathbb{K}(x_i,x_j))_{i,j\leq n}\ \forall (x_1,\dots,x_n)\in E^n$, is its probability density function.

Not only that, but we can also deduce, given any $1\leq a<b\leq n$, that $P(X_a=X_b)=0$. To do it, notice that this probability can be calculated as an integral of $f(x_1,\dots,x_{b-1},x_a,x_{b+1},\dots,x_n)$ as a function of $x_1,\dots,x_{b-1},x_{b+1},\dots,x_n\in E$, that is constant equal to 0 because it is based on the determinant of a matrix that has two equal columns, so the integral is also 0. That means that the random variables $X_1,\dots,X_n$ are pairwise distinct almost surely.

Moreover, given any $x_1,\dots,x_n\in E$ points in E and $\sigma\in S_n$ a permutation, we can also notice that:

$$f(x_{\sigma(1)},\dots,x_{\sigma(n)})=\frac{1}{n!}\det(\mathbb{K}(x_{\sigma(i)},x_{\sigma(j)}))_{i,j\leq n}=$$
$$\frac{1}{n!}\det(P_\sigma)\det(\mathbb{K}(x_i,x_j))_{i,j\leq n}\det(P_\sigma^{-1})=\frac{1}{n!}\det(\mathbb{K}(x_i,x_j))_{i,j\leq n}$$

where P_σ denotes the permutation matrix associated to σ, so the random variables $X_1,\dots,X_n$ are exchangeable.

Thanks to these last observations, we can now apply Proposition 2.11 and conclude that the point process $\mathscr{X}=\sum_{i=1}^n\delta_{X_i}$ exists and has its joint intensities given by $\rho_k(x_1,\dots,x_k)=\frac{1}{(n-k)!}\int_{E^{n-k}}\det(\mathbb{K}(x_i,x_j))_{i,j\leq n}d\mu(x_{k+1})\dots d\mu(x_n)\ \forall x_1,\dots,x_k\in E^k$ if $1\leq k\leq n$ and $\rho_k(x_1,\dots,x_k)=0\ \forall x_1,\dots,x_k\in E^k$ otherwise. So, to complete the proof, it suffices to see that actually $\rho_k(x_1,\dots,x_k)=\det(\mathbb{K}(x_i,x_j))_{i,j\leq k}\ \forall x_1,\dots,x_k\in E^k$ for all $k\geq 1$, which implies that $\mathscr{X}$ is a determinantal point process with kernel $\mathbb{K}$ (and it is easy to see that it has n random points almost surely).

Indeed, to do it, we start by noticing that, if $k=n$, then the result is clear and already proven. So, we can now distinguish two cases, depending on the value of k we focus on.

If $k>n$, we can deduce that, similar to what we saw before, $(\mathbb{K}(x_i,x_j))_{i,j\leq k}=(\varphi_j(x_i))_{i\leq k,j\leq n}(\overline{\varphi_i}(x_j))_{i\leq n,j\leq k}$, so it is a product of two matrices of rank smaller or equal than n, something that implies that the rank of $(\mathbb{K}(x_i,x_j))_{i,j\leq k}$ is also smaller or equal than n, so $\det(\mathbb{K}(x_i,x_j))_{i,j\leq k}=0$, that is exactly what we had to conclude in this case.

If $k<n$, then we can apply the expressions of the determinant we deduced in our preliminaries (proof of Theorem 2.22) to observe that, given any $x_1,\dots,x_k\in E$:

$$\rho_k(x_1,\dots,x_k)=\frac{1}{(n-k)!}\int_{E^{n-k}}\det(\mathbb{K}(x_i,x_j))_{i,j\leq n}d\mu(x_{k+1})\dots d\mu(x_n)=$$
$$\frac{1}{(n-k)!}\int_{E^{n-k}}\sum_{\sigma\in S_n}\operatorname{sgn}(\sigma)\sum_{j_1,\dots,j_n\leq n}\prod_{i=1}^n\varphi_{j_i}(x_i)\overline{\varphi_{j_i}}(x_{\sigma(i)})d\mu(x_{k+1})\dots d\mu(x_n)=$$
$$\frac{1}{(n-k)!}\sum_{\sigma\in S_n}\operatorname{sgn}(\sigma)\sum_{j_1,\dots,j_n\leq n}\int_{E^{n-k}}\prod_{i=1}^n\varphi_{j_i}(x_i)\overline{\varphi_{j_i}}(x_{\sigma(i)})d\mu(x_{k+1})\dots d\mu(x_n)=$$
$$\frac{1}{(n-k)!}\sum_{\sigma\in S_n}\operatorname{sgn}(\sigma)\sum_{j_1,\dots,j_n\leq n}(\prod_{i=1}^k\varphi_{j_i}(x_i)\overline{\varphi_{j_{\sigma^{-1}(i)}}}(x_i))\cdot$$
$$\prod_{i=k+1}^n\int_E\varphi_{j_i}(x_i)\overline{\varphi_{j_{\sigma^{-1}(i)}}}(x_i)d\mu(x_i)=$$
(using that if $j_a=j_b$ for $a\neq b$, then the expression is 0, as already seen) $=$

$$\frac{1}{(n-k)!}\sum_{\sigma\in S_n,\sigma|_{\{k+1,\dots,n\}}=id}\operatorname{sgn}(\sigma)\sum_{j_1<\cdots<j_n\leq n}\prod_{i=1}^{k}\varphi_{j_i}(x_i)\overline{\varphi_{j_{\sigma^{-1}(i)}}}(x_i)=$$
$$\frac{1}{(n-k)!}\sum_{j_1<\cdots<j_n\leq n}\sum_{\sigma\in S_k}\operatorname{sgn}(\sigma)\prod_{i=1}^{k}\varphi_{j_i}(x_i)\overline{\varphi_{j_i}}(x_{\sigma(i)})=$$
$$\frac{1}{(n-k)!}\sum_{j_{k+1}<\cdots<j_n\leq n}\sum_{j_1,\dots,j_k\leq n}\sum_{\sigma\in S_k}\operatorname{sgn}(\sigma)\prod_{i=1}^{k}\varphi_{j_i}(x_i)\overline{\varphi_{j_i}}(x_{\sigma(i)})=$$
$$\frac{1}{(n-k)!}\sum_{j_{k+1}<\cdots<j_n\leq n}\det(\mathbb{K}(x_i,x_j))_{i,j\leq k}=$$
$$\frac{1}{(n-k)!}(n-k)!\det(\mathbb{K}(x_i,x_j))_{i,j\leq k}=\det(\mathbb{K}(x_i,x_j))_{i,j\leq k}$$

something that shows the desired equality and completes the proof. That is, we have managed to see that $\mathscr{X}$ is a determinantal point process with kernel $\mathbb{K}$ and has n random points almost surely. □

It is worth mentioning that this type of kernels are called (finite dimensional) projection kernels, because their corresponding integral operators are actually the projection operators in the subspace generated by the system $\{\varphi_k\}_{k=1}^n$, something that follows easily from the fact that:

$$\mathscr{K}f=\sum_{k=1}^{n}\langle f,\varphi_k\rangle\varphi_k\quad\forall f\in L^2(E,\mu)$$

that we have already seen.

Moving on, we can observe that, in fact, finite dimensional projection kernels are not the only ones that define determinantal point processes, and the following theorem goes in this direction, as we now see.

Theorem 4.3. *Let $\mathbb{K}:E^2\longrightarrow\mathbb{C}$ be a function given by $\mathbb{K}(x,y)=\sum_{k=1}^{n}\lambda_k\varphi_k(x)\overline{\varphi_k}(y)$ $\forall x,y\in E$, where $n\in\mathbb{N}\cup\{\infty\}$, $\{\varphi_k\}_{k=1}^n$ is an orthonormal system and $\lambda_k\in[0,1]$ $\forall 1\leq k\leq n$, assuming $\mathscr{K}$ is of trace class. Then, consider $I_1,\dots,I_n$ independent random variables such that $I_k\sim B(\lambda_k)$ $\forall 1\leq k\leq n$ and $\mathbb{K}_I:E^2\longrightarrow\mathbb{C}$ given by $\mathbb{K}_I(x,y)=\sum_{k=1}^{n}I_k\varphi_k(x)\overline{\varphi_k}(y)$ $\forall x,y\in E$.*

In this case, if we denote by $\mathscr{X}_I$ the (random) determinantal point process defined by $\mathbb{K}_I$ after sampling the Bernoulli random variables, we have that $\mathscr{X}_I$ exists, is well-defined and is a determinantal point process with kernel $\mathbb{K}$.

Proof. Indeed, continuing with the above notations and changing again some details from the proof found in [11], we start by noticing that without loss of generality we can assume $n=\infty$. That is the case because, if it was finite, we could extend the orthonormal system $\{\varphi_k\}_{k=1}^n$ to an infinite one and take $\lambda_k=0$ for the added functions, in such a way that, at the end, our new kernels $\mathbb{K}$ and $\mathbb{K}_I$ would be the same that in the finite case (almost surely for the random one) but we could apply the result for the infinite case.

Now, we can also observe that $\mathbb{K}$ is a square integrable and Hermitian kernel such that $\mathscr{K}$ is of trace class and non-negative definite, so it makes sense (with our definition of DPP) to talk about a determinantal point process with kernel $\mathbb{K}$. To do it, we can easily see that $\mathbb{K}$ is well-defined and square integrable in E^2, using the same techniques

that in the proof of Theorem 2.22. Not only that, but also, for all $x, y \in E$, we have that $\mathbb{K}(x,y) = \sum_{k=1}^{\infty} \lambda_k \varphi_k(x)\overline{\varphi_k}(y) = \overline{\sum_{k=1}^{\infty} \lambda_k \overline{\varphi_k}(x)\varphi_k(y)} = \overline{\mathbb{K}(y,x)}$, so $\mathbb{K}$ is an Hermitian kernel. Finally, its associated integral operator $\mathscr{K} : L^2(E,\mu) \longrightarrow L^2(E,\mu)$ is given by:

$$\mathscr{K} f(x) = \sum_{k=1}^{\infty} \lambda_k \langle f, \varphi_k \rangle \varphi_k(x)$$

a.e. $x \in E$ and $\forall f \in L^2(E,\mu)$, something that, again, follows directly from the proof of Theorem 2.22. That clearly shows it is non-negative definite, so $\mathbb{K}$ is in the hypothesis of our definition of determinantal point processes.

That being said, from here, to show that $\mathscr{X}_I$ exists and is actually well-defined, we can observe that $\mathbb{E}(\sum_{k=1}^{\infty} I_k) = \sum_{k=1}^{\infty} \mathbb{E}(I_k) = \sum_{k=1}^{\infty} \lambda_k < \infty$ because we know by hypothesis that $\mathscr{K}$ is of trace class and $\{\lambda_k\}_{k=1}^{\infty}$ are its eigenvalues. In particular, we can conclude that $\sum_{k=1}^{\infty} I_k < \infty$ a.s., something that implies that the number of Bernoulli random variables taking the value 1 is finite almost surely.

In other words, almost surely, $\mathbb{K}_I$ is a finite dimensional projection kernel and, in these cases, we can apply Proposition 4.2 to conclude that it defines a determinantal point process. So, that shows that the point process $\mathscr{X}_I$ exists, is well-defined and is a (random) determinantal point process with (random) kernel $\mathbb{K}_I$ almost surely.

From here, we proceed to study the joint intensities of this point process, that, even though it is a determinantal point process with a certain (random) kernel in each realization, in principle, it is not clear that it is a determinantal point process (with a fixed deterministic kernel). Indeed, to do it, given $k \geq 1$ and given any $x_1, \ldots, x_k \in E$ pairwise distinct, we can observe that (except maybe in a subset of E^k of measure 0):

$$\rho_k(x_1, \ldots, x_k) = \lim_{\epsilon \to 0^+} \frac{P(\mathscr{X}_I(B_\epsilon(x_i))=1 \text{ for each } 1 \leq i \leq k)}{\prod_{i=1}^{k} \mu(B_\epsilon(x_i))} =$$

$$\text{(considering the random variable } N := \sum_{i=1}^{\infty} I_i) =$$

$$\lim_{\epsilon \to 0^+} \sum_{m=1}^{\infty} \frac{P(\mathscr{X}_I(B_\epsilon(x_i))=1 \text{ for each } 1 \leq i \leq k \mid N=m)}{\prod_{i=1}^{k} \mu(B_\epsilon(x_i))} P(N=m) =$$

$$\lim_{\epsilon \to 0^+} \sum_{m=1}^{\infty} \sum_{i_1 < \cdots < i_m} \frac{P(\mathscr{X}_I(B_\epsilon(x_i))=1 \text{ for each } 1 \leq i \leq k \mid N=m, I_{i_1}=1, \ldots, I_{i_m}=1)}{\prod_{i=1}^{k} \mu(B_\epsilon(x_i))} \cdot$$

$$P(I_{i_1} = 1, \ldots, I_{i_m} = 1 | N = m) P(N = m) =$$

$$\text{(denoting by } \mathbb{K}_J \text{ the corresponding realization of the random kernel)} =$$

$$\sum_{m=1}^{\infty} \sum_{i_1 < \cdots < i_m} \det(\mathbb{K}_J(x_i, x_j))_{i,j \leq k} P(I_{i_1} = 1, \ldots, I_{i_m} = 1 | N = m) P(N = m) =$$

$$\mathbb{E}(\det(\mathbb{K}_I(x_i, x_j))_{i,j \leq k})$$

where in this last equality we have applied known properties of the expected value. That shows, in particular, that $\rho_k(x_1, \ldots, x_k) = \mathbb{E}(\det(\mathbb{K}_I(x_i, x_j))_{i,j \leq k})$ μ^k-a.e. for $(x_1, \ldots, x_k) \in E^k$.

So, in conclusion and to complete the proof, it suffices to deduce that given μ^k-almost any $(x_1, \ldots, x_k) \in E^k$ we have that $\mathbb{E}(\det(\mathbb{K}_I(x_i, x_j))_{i,j \leq k}) = \det(\mathbb{K}(x_i, x_j))_{i,j \leq k}$, because in this case $\mathscr{X}_I$ would be a determinantal point process with kernel $\mathbb{K}$ by definition. To do it, applying again results seen in the preliminaries, we can start by noticing that:

$$\det(\mathbb{K}(x_i, x_j))_{i,j \leq k} = \sum_{j_1 < \cdots < j_k} (\prod_{i=1}^{k} \lambda_{j_i}) |\det(\varphi_{j_h}(x_i))_{i,h \leq k}|^2 =$$

$$\sum_{j_1<\cdots<j_k}(\prod_{i=1}^{k}\mathbb{E}(I_{j_i}))|\det(\varphi_{j_h}(x_i))_{i,h\leq k}|^2 =$$
$$\text{(using that the Bernoulli random variables are independent)} =$$
$$\sum_{j_1<\cdots<j_k}\mathbb{E}(\prod_{i=1}^{k}I_{j_i})|\det(\varphi_{j_h}(x_i))_{i,h\leq k}|^2 =$$
$$\sum_{j_1<\cdots<j_k}\mathbb{E}((\prod_{i=1}^{k}I_{j_i})|\det(\varphi_{j_h}(x_i))_{i,h\leq k}|^2) =$$
$$\text{(by the monotonous convergence theorem)} =$$
$$\mathbb{E}(\sum_{j_1<\cdots<j_k}(\prod_{i=1}^{k}I_{j_i})|\det(\varphi_{j_h}(x_i))_{i,h\leq k}|^2)$$

Finally, given any $w\in\Omega$, the probability space where the Bernoulli random variables are defined, we denote $\mathbb{K}_{I(\omega)}(x,y)=\sum_{k=1}^{n}I_k(\omega)\varphi_k(x)\overline{\varphi_k}(y)\ \forall x,y\in E$ and, except maybe in a subset of measure 0 of Ω, it is a finite dimensional projection kernel, so we can apply the results seen in the preliminaries and conclude that:

$$\det(\mathbb{K}_{I(\omega)}(x_i,x_j))_{i,j\leq k}=\sum_{j_1<\cdots<j_k}(\prod_{i=1}^{k}I_{j_i}(\omega))|\det(\varphi_{j_h}(x_i))_{i,h\leq k}|^2$$

In particular, combining all of the above and taking expectations, then we have that:

$$\mathbb{E}(\sum_{j_1<\cdots<j_k}(\prod_{i=1}^{k}I_{j_i})|\det(\varphi_{j_h}(x_i))_{i,h\leq k}|^2)=\mathbb{E}(\det(\mathbb{K}_I(x_i,x_j))_{i,j\leq k})$$

so we conclude that $\det(\mathbb{K}(x_i,x_j))_{i,j\leq k}=\mathbb{E}(\det(\mathbb{K}_I(x_i,x_j))_{i,j\leq k})$, something that completes the proof. □

Finally, thanks to all of the above, we proceed to characterise the kernels that actually give rise to determinantal point processes in terms of the norm of their associated integral operator, if it can be extended to an operator on $L^2(E)$.

Theorem 4.4. *Let $\mathbb{K}:E^2\longrightarrow\mathbb{C}$ be a locally square integrable and Hermitian kernel such that $\mathscr{K}$ is locally of trace class, non-negative definite and extends to a bounded operator $\mathscr{K}'$ on $L^2(E)$. Then, we have that $\mathbb{K}$ defines a determinantal point process in E if, and only if, the norm of $\mathscr{K}'$ is smaller of equal than 1, that is, $\|\mathscr{K}'\|\leq 1$.*

Proof. Since this long proof is only very briefly outlined in [11], we proceed to fill in the details with our own ideas, partially inspired by the proof of the Kolmogorov Extension Theorem provided in [23]. Indeed, continuing with the above notations, we can start by observing that the extended bounded operator $\mathscr{K}'$ is self-adjoint, because $\mathscr{K}$ is it. To do it, given any $f,g\in L^2(E,\mu)$ functions, we use the fact that the continuous functions with compact support are dense in $L^2(E,\mu)$ and consider sequences $\{f_n\}_{n\geq 1},\ \{g_n\}_{n\geq 1}\subseteq L^2(E)$ of continuous functions with compact support such that $\lim_n f_n=f$ and $\lim_n g_n=g$.

Then, given $\epsilon>0$, there exists $n_0\in\mathbb{N}$ such that, $\forall n\geq n_0$, we have that $\|f-f_n\|\leq\frac{\epsilon}{2(\|\mathscr{K}'\|\|g\|+1)}$, $\|f_n\|\leq\|f\|+1$ and then also $\|g-g_n\|\leq\frac{\epsilon}{2(\|\mathscr{K}'\|(\|f\|+1)+1)}$. So, in particular, for any $n\geq n_0$, we can observe that:

$$|\langle\mathscr{K}'f,g\rangle-\langle\mathscr{K}'f_n,g_n\rangle|=|\langle\mathscr{K}'f,g\rangle-\langle\mathscr{K}'f_n,g\rangle+\langle\mathscr{K}'f_n,g\rangle-\langle\mathscr{K}'f_n,g_n\rangle|\leq$$
$$|\langle\mathscr{K}'f,g\rangle-\langle\mathscr{K}'f_n,g\rangle|+|\langle\mathscr{K}'f_n,g\rangle-\langle\mathscr{K}'f_n,g_n\rangle|=$$
$$|\langle\mathscr{K}'f-\mathscr{K}'f_n,g\rangle|+|\langle\mathscr{K}'f_n,g-g_n\rangle|\leq\|\mathscr{K}'(f-f_n)\|\|g\|+\|\mathscr{K}'f_n\|\|g-g_n\|\leq$$

$$\|\mathscr{K}'\|\|f-f_n\|\|g\|+\|\mathscr{K}'\|\|f_n\|\|g-g_n\|\leq$$
$$\|\mathscr{K}'\|\frac{\epsilon}{2(\|\mathscr{K}'\|\|g\|+1)}\|g\|+\|\mathscr{K}'\|(\|f\|+1)\frac{\epsilon}{2(\|\mathscr{K}'\|(\|f\|+1)+1)}\leq\epsilon$$

This shows that $\langle\mathscr{K}'f,g\rangle=\lim_n\langle\mathscr{K}'f_n,g_n\rangle$ and, by the properties of the inner product, exchanging the roles of f and g, we also deduce that $\langle f,\mathscr{K}'g\rangle=\lim_n\langle f_n,\mathscr{K}'g_n\rangle$.

Finally, using the above and the already studied properties of the integral operator $\mathscr{K}$, we can conclude that:

$$\langle\mathscr{K}'f,g\rangle=\lim_n\langle\mathscr{K}'f_n,g_n\rangle=\lim_n\langle\mathscr{K}f_n,g_n\rangle=$$
$$\text{(considering } f_n \text{ and } g_n \text{ as functions defined in the union of their supports}$$
$$\text{and restricting } \mathscr{K} \text{ to this compact set)} =$$
$$\lim_n\langle f_n,\mathscr{K}g_n\rangle=\lim_n\langle f_n,\mathscr{K}'g_n\rangle=\langle f,\mathscr{K}'g\rangle$$

that is what we wanted to see. That shows that $\mathscr{K}'$ is self-adjoint.

From here, recall that it is a well-known fact that the norm of self-adjoint operators can be described by the inner product of the space, in our case, we have that $\|\mathscr{K}'\|=\sup\{|\langle\mathscr{K}'f,f\rangle|\mid f\in L^2(E,\mu),\ \|f\|=1\}$. Not only that, but also, given any $K\subseteq E$ a compact set, we already know that our hypothesis imply that the restriction $\mathscr{K}_K$ is a compact and self-adjoint operator in $L^2(K,\mu)$, so we deduce that $\|\mathscr{K}_K\|=\sup\{|\langle\mathscr{K}_Kf,f\rangle|\mid f\in L^2(K,\mu),\ \|f\|=1\}$.

Now, using this observation, as an auxiliary result, we will show that $\|\mathscr{K}'\|\leq 1$ if, and only if, for all $K\subseteq E$ compact, we have that $\|\mathscr{K}_K\|\leq 1$. Indeed, to do it, on the one hand, we start by assuming that $\|\mathscr{K}'\|\leq 1$ and then it is clear that, given any $K\subseteq E$ a compact set, we have that:

$$\|\mathscr{K}_K\|=\sup\{|\langle\mathscr{K}_Kf,f\rangle|\mid f\in L^2(K,\mu),\ \|f\|=1\}=$$
$$\text{(relating } L^2(K,\mu) \text{ with the set of functions in } L^2(E,\mu) \text{ with support in } K) =$$
$$\sup\{|\langle\mathscr{K}'f,f\rangle|\mid f\in L^2(E,\mu),\ f \text{ has support in } K,\ \|f\|=1\}\leq$$
$$\sup\{|\langle\mathscr{K}'f,f\rangle|\mid f\in L^2(E,\mu),\ \|f\|=1\}=\|\mathscr{K}'\|\leq 1$$

something that proves the first implication.

On the other hand, assuming now that for all $K\subseteq E$ we have that $\|\mathscr{K}_K\|\leq 1$, then, given any $f\in L^2(E,\mu)$ such that $\|f\|=1$, it suffices to see that $|\langle\mathscr{K}'f,f\rangle|\leq 1$. To do it, as before, we can consider $\{f_n\}_{n\geq 1}\subseteq L^2(E,\mu)$ a sequence of continuous functions with compact support such that $\lim_n f_n=f$ and then, we already know that $|\langle\mathscr{K}'f,f\rangle|=|\lim_n\langle\mathscr{K}'f_n,f_n\rangle|=\lim_n|\langle\mathscr{K}'f_n,f_n\rangle|$.

So, again, fixing any $n\geq 1\in\mathbb{N}$, it will be enough to see that $|\langle\mathscr{K}'f_n,f_n\rangle|\leq 1$. But it is clear that $|\langle\mathscr{K}'f_n,f_n\rangle|=|\langle\mathscr{K}_{\text{supp }f_n}f_n,f_n\rangle|\leq 1$, something that completes this part of the proof.

That being said, given any $K\subseteq E$ compact set, we can also deduce from our hypothesis that the integral operator $\mathscr{K}_K$ is non-negative definite. Indeed, we merely need to observe that, given $f\in L^2(K,\mu)$, we have that $\langle\mathscr{K}_Kf,f\rangle=$ (thinking f as a

function in $L^2(E,\mu)$ with support in K) $= \langle \mathscr{K} f, f\rangle \geq 0$ because we know that $\mathscr{K}$ is non-negative definite, something that proves the claim.

From here, taking into account all of the above, we are ready to show the equivalence stated by the theorem, by proving both implications.

$\Longrightarrow$ Indeed, in this case, by reduction to absurdity, we assume that $\mathbb{K}$ defines a determinantal point process in E but also that $\|\mathscr{K}^{'}\| > 1$. Then, using the previous observations, we can deduce that it exists $K \subseteq E$ a compact set such that $\|\mathscr{K}_K\| > 1$ and consider the restriction of the determinantal point process with kernel $\mathbb{K}$ to K, that we denote by $\mathscr{X}$. Notice that, by the Lemma 4.1, we already know that $\mathscr{X}$ is also a determinantal point process with kernel $\mathbb{K}_{|K^2}$ and it is clear that $\mathscr{K}_K$ can be seen as the integral operator defined by $\mathbb{K}_{|K^2}$, that is compact, self-adjoint, non-negative definite and of trace class.

With that in mind, we can apply the results we described in the preliminaries to conclude that there exists $n \in \mathbb{N}\cup\{\infty\}$, $\{\varphi_k\}_{k=1}^n$ an orthonormal basis of $L^2(K,\mu)$ and $\{\lambda_k\}_{k=1}^n$ a set of real and non-negative eigenvalues that we can consider in decreasing order such that $\mathbb{K}_{|K^2}(x,y) = \sum_{k=1}^n \lambda_k \varphi_k(x)\overline{\varphi_k}(y)$ μ-a.e. for $x, y \in K$. Not only that, but also, for compact and self-adjoint operators as $\mathscr{K}_K$, it is known that its norm coincides with the maximum absolute value of its eigenvalues, so $\lambda_1 > 1$.

Now, we can consider $\mathscr{Y}$ the point process given by sampling $\mathscr{X}$ and then independently removing each of the obtained points with probability $1 - \frac{1}{\lambda_1} \in (0,1)$. It is clear that $\mathscr{Y}$ exists and it is a well-defined point process in K and, from here, we proceed to see that it is actually a determinantal point process with kernel $\frac{1}{\lambda_1}\mathbb{K}_{|K^2}$. Indeed, to do it, we can compute its joint intensities, observing that, given any $k \geq 1 \in \mathbb{N}$ and $D_1, \ldots, D_k \in \mathcal{B}(K)$ pairwise disjoint Borel sets, we have that:

$$\mathbb{E}(\textstyle\prod_{i=1}^k \mathscr{Y}(D_i)) =$$

$$\textstyle\sum_{n_1,\ldots,n_k \geq 0} \mathbb{E}(\prod_{i=1}^k \mathscr{Y}(D_i) \mid \mathscr{X}(D_i) = n_i \text{ for all } i \leq k) P(\mathscr{X}(D_i) = n_i \text{ for all } i \leq k)$$

(using the fact that the points of $\mathscr{X}$ are removed independently to define $\mathscr{Y}$) =

$$\textstyle\sum_{n_1,\ldots,n_k \geq 0} (\prod_{i=1}^k \mathbb{E}(\mathscr{Y}(D_i) \mid \mathscr{X}(D_i) = n_i)) P(\mathscr{X}(D_i) = n_i \text{ for all } i \leq k) =$$

(if $\mathscr{X}(D_i) = n_i$, then $\mathscr{Y}(D_i)$ behaves like a binomial random variable) =

$$\textstyle\sum_{n_1,\ldots,n_k \geq 0} (\prod_{i=1}^k \frac{n_i}{\lambda_1}) P(\mathscr{X}(D_i) = n_i \text{ for all } i \leq k) =$$

$$\textstyle\frac{1}{\lambda_1^k} \sum_{n_1,\ldots,n_k \geq 0} (\prod_{i=1}^k n_i) P(\mathscr{X}(D_i) = n_i \text{ for all } i \leq k) = \frac{1}{\lambda_1^k}\mathbb{E}(\prod_{i=1}^k \mathscr{X}(D_i)) =$$

$$\textstyle\frac{1}{\lambda_1^k} \int_{\prod_{i=1}^k D_i} \det(\mathbb{K}_{|K^2}(x_i,x_j))_{i,j\leq k} d\mu(x_1)\ldots d\mu(x_k) =$$

$$\textstyle\int_{\prod_{i=1}^k D_i} \det(\frac{1}{\lambda_1}\mathbb{K}_{|K^2}(x_i,x_j))_{i,j\leq k} d\mu(x_1)\ldots d\mu(x_k)$$

that is exactly what we wanted to deduce. That shows by definition that we have that $\rho_k(x_1,\ldots,x_k) = \det(\frac{1}{\lambda_1}\mathbb{K}_{|K^2}(x_i,x_j))_{i,j\leq k}$ μ^k-a.e. for $(x_1,\ldots,x_k) \in K^k$. In particular, we conclude that $\mathscr{Y}$ is a determinantal point process in K with kernel $\frac{1}{\lambda_1}\mathbb{K}_{|K^2}$ (notice that it is clear that $\frac{1}{\lambda_1}\mathbb{K}_{|K^2}$ is still a square integrable and Hermitian kernel in K^2 with a trace class and non-negative definite associated integral operator).

Finally, using that $\mathscr{K}_K$ is of trace class by hypothesis and applying the results already shown in the preliminaries, we deduce that $\det(\mathbb{K}_{|K^2}(x_i, x_j))_{i,j\leq k}$, as function of $x_1, \ldots, x_k \in K$, is integrable in K^k, so in particular $\mathbb{E}(\mathscr{X}(K)) < \infty$ and the number of points of $\mathscr{X}$ in K is finite almost surely. That, combined with the fact that $1-\frac{1}{\lambda_1} > 0$, shows that it is possible to remove all the original points of $\mathscr{X}$ when sampling $\mathscr{Y}$, which implies that $P(\mathscr{Y}(K) = 0) > 0$.

Nevertheless, we can also observe that $\frac{1}{\lambda_1}\mathbb{K}_{|K^2}(x, y) = \sum_{k=1}^{n} \frac{\lambda_k}{\lambda_1}\varphi_k(x)\overline{\varphi_k}(y)$ μ-a.e. for $x, y \in K$ with $\frac{\lambda_k}{\lambda_1} \in [0, 1]$ and $\frac{\lambda_1}{\lambda_1} = 1$, something that, by Proposition 4.2 and observations made during that proof of Theorem 4.3, shows that $\mathscr{Y}$ must have at least one point almost surely, that is, $P(\mathscr{Y}(K) = 0) = 0$.

That leads us to a clear contradiction, so we have managed to see that, if $\|\mathscr{K}'\| > 1$, then there cannot exist a determinantal point process in E with kernel $\mathbb{K}$. That completes the proof of this first implication.

$\Longleftarrow$ On the other hand, we assume that $\|\mathscr{K}'\| \leq 1$ and then, we have to conclude that there exists a determinantal point process with kernel $\mathbb{K}$ defined in E. To do it, we can start by observing that, given any $K \subseteq E$ compact set, we can consider the restriction $\mathbb{K}_{|K^2}$, that is an Hermitian and square integrable kernel such that its integral operator $\mathscr{K}_K$ is compact, self-adjoint, non-negative definite and of trace class. Then, as before, we know that there exists $n \in \mathbb{N} \cup \{\infty\}$, $\{\varphi_k\}_{k=1}^{n}$ an orthonormal basis of $L^2(K, \mu)$ and $\{\lambda_k\}_{k=1}^{n}$ a set of real and non-negative eigenvalues such that $\mathbb{K}_{|K^2}(x, y) = \sum_{k=1}^{n} \lambda_k\varphi_k(x)\overline{\varphi_k}(y)$ μ-a.e. for $x, y \in K$. In fact, in this case, using that now $\|\mathscr{K}_K\| \leq 1$, we can deduce that $\lambda_k \in [0, 1]$ for all $1 \leq k \leq n$.

So, we find ourselves in the hypothesis of Theorem 4.3 and, in particular, we conclude that there exists a determinantal point process in K with kernel $\mathbb{K}_{|K^2}$.

From here, recall that $M(E)$ denoted the set of all Radon measures in E and that we considered $\mathcal{M}$ the σ-algebra in $M(E)$ generated by the corresponding subsets $\{\mu \in M(E) \mid \mu(D) \in I\}$, where $D \in \mathcal{B}(E)$ and $I \subseteq [0, \infty)$ denotes an open interval. In fact, similarly, we can also define $M'(E)$ the set of all Radon measures in E that are point measures, that is, given $\mu \in M(E)$, $\mu \in M'(E)$ if an only if $\mu = \sum_{x\in S} a_x\delta_x$ for some $a_x \in \mathbb{N}$ and $S \subseteq E$ countable. Denote also by $\mathcal{M}'$ the analogous σ-algebra generated the sets $\{\mu \in M'(E) \mid \mu(D) \in I\}$, where, again, $D \in \mathcal{B}(E)$ and $I \subseteq [0, \infty)$ denotes an open interval.

With that in mind, notice that we have already shown, during the proof of Proposition 2.16, that any Borel set of E can be written as a countable union of compact sets so, in particular, we can express $E = \cup_{i=1}^{\infty} A_i$ with A_i compact for all $i \in \mathbb{N}$ and, if we define the sets $K_n \coloneqq \bigcup_{i=1}^{n} A_i$ for all $n \in \mathbb{N}$, then we have that $E = \bigcup_{n=1}^{\infty} K_n$ is a countable union of increasing compact sets. In addition, without loss of generality, it is clear that we can assume that $\mu(K_n) > 0$ for all $n \in \mathbb{N}$.

Then, given any $n \geq 1$, we can claim, thanks to our previous results, that there exists a determinantal point process in K_n with kernel $\mathbb{K}_{|K_n^2}$, that we can denote by $\mathscr{X}_n$. Now, applying the definition of these point processes, we know that $\mathscr{X}_n$ is a measurable map

$\mathscr{X}_n : (\Omega_n, \mathcal{F}_n) \longrightarrow (M^{'}(K_n), \mathcal{M}_n^{'})$ such that $\mathscr{X}_n(\omega)$ is a simple point measure a.s. $\omega \in \Omega_n$, where $(\Omega_n, \mathcal{F}_n, \tau_n)$ is a certain probability space and $(M^{'}(K_n), \mathcal{M}_n^{'})$ denotes the corresponding measurable space of point measures defined in K_n. In this situation, we can define in this last measurable space $(M^{'}(K_n), \mathcal{M}_n^{'})$ the map $P_n : \mathcal{M}_n^{'} \longrightarrow [0,1]$ given by $P_n(A) = \tau_n(\mathscr{X}_n^{-1}(A)) \ \forall A \in \mathcal{M}_n^{'}$ and we proceed to see that it is a well-defined probability measure.

Lemma 4.5. *The map $P_n : \mathcal{M}_n^{'} \longrightarrow [0,1]$ given by $P_n(A) = \tau_n(\mathscr{X}_n^{-1}(A)) \ \forall A \in \mathcal{M}_n^{'}$ is a well-defined probability measure in $(M^{'}(K_n), \mathcal{M}_n^{'})$.*

Proof. Indeed, notice that P_n is clearly well-defined. If $A \in \mathcal{M}_n^{'}$, then we can deduce that $\mathscr{X}_n^{-1}(A) \in \mathcal{F}_n$ because $\mathscr{X}_n$ is measurable, so it makes sense to consider $\tau_n(\mathscr{X}_n^{-1}(A))$. Moreover, by definition, it is clear that $P_n(A) \in [0,1]$ for all $A \in \mathcal{M}_n^{'}$ and we can observe that $P_n(\emptyset) = \tau_n(\mathscr{X}_n^{-1}(\emptyset)) = \tau_n(\emptyset) = 0$ and also $P_n(M^{'}(K_n)) = \tau_n(\mathscr{X}_n^{-1}(M^{'}(K_n))) = \tau_n(\Omega_n) = 1$.

Finally, given any sequence $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{M}_n^{'}$ of pairwise disjoint sets, we have that $P_n(\bigcup_{i=1}^{\infty} A_i) = \tau_n(\mathscr{X}_n^{-1}(\bigcup_{i=1}^{\infty} A_i)) =$ (by general properties that any function between two sets verifies) $= \tau_n(\bigcup_{i=1}^{\infty} \mathscr{X}_n^{-1}(A_i)) =$ (again, it is clear and a general fact that preimages preserve the disjoint property) $= \sum_{i=1}^{\infty} \tau_n(\mathscr{X}_n^{-1}(A_i)) = \sum_{i=1}^{\infty} P_n(A_i)$. That shows that P_n is a probability measure in $(M^{'}(K_n), \mathcal{M}_n^{'})$. □

From here, we can also consider the restriction map $r_n : M^{'}(E) \longrightarrow M^{'}(K_n)$ given by $r_n(\mu) = \mu_{|K_n}$ for all $\mu \in M^{'}(E)$, where $\mu_{|K_n}$ denotes the original measure restricted to K_n, and one can deduce that r_n is a measurable map. Indeed, given any $D \in \mathcal{B}(K_n)$ and $I \subseteq [0, \infty)$ open interval, we can consider $\{\mu \in M^{'}(K_n) \mid \mu(D) \in I\}$ and it is clear that $r_n^{-1}(\{\mu \in M^{'}(K_n) \mid \mu(D) \in I\}) = \{\mu \in M^{'}(E) \mid \mu(D) \in I\}$, something that shows that r_n is measurable.

Now, given any $1 \leq n \leq m \in \mathbb{N}$, then it is clear by definition that $K_n \subseteq K_m$. So, similarly and as a particular case, we can also define $r_{m,n} : M^{'}(K_m) \longrightarrow M^{'}(K_n)$ the corresponding restriction map given by $r_{m,n}(\mu) = \mu_{|K_n} \ \forall \mu \in M^{'}(K_m)$ and it has the same properties, that is, it is measurable. In fact, not only that, but we can recall that the restriction of $\mathscr{X}_m$ to K_n is a determinantal point process in K_n with kernel $\mathbb{K}_{|K_n^2}$ (it has the same law that $\mathscr{X}_n$) and, in this context, that translates into $r_{m,n} \circ \mathscr{X}_m$ having the same law that $\mathscr{X}_n$, which means that $P_n(A) = P_m(r_{m,n}^{-1}(A))$ for all $A \in \mathcal{M}_n^{'}$.

That being said, given now any $D \in \mathcal{B}(E)$ Borel set and $I \subseteq [0, \infty)$ open interval, we can clearly express $D = \bigcup_{n=1}^{\infty}(D \cap K_n)$ and denote $D_n = D \cap K_n \in \mathcal{B}(K_n)$ for all $n \geq 1$. Not only that, but given any $\mu \in M^{'}(E)$ point measure, we can also observe that $\mu(D) \in I$ if, and only if, there exists $n_0 \geq 1$ such that $\mu(D_n) \in I$ for all $n \geq n_0$, something that shows that $\{\mu \in M^{'}(E) \mid \mu(D) \in I\} = \bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} \{\mu \in M^{'}(E) \mid \mu(D_i) \in I\}$. That, combined with the definition of $\mathcal{M}^{'}$, shows that $\mathcal{M}^{'}$ is the σ-algebra generated by the sets $\{\mu \in M^{'}(E) \mid \mu(V) \in J\}$ for any $n \geq 1$, $V \in \mathcal{B}(K_n)$ and $J \subseteq [0, \infty)$ open interval. In other words, $\mathcal{M}^{'}$ is the σ-algebra generated by the sets $r_n^{-1}(A_n)$ for any $n \geq 1$ and $A_n \in \mathcal{M}_n^{'}$. Then, if we consider

$\mathcal{A} = \{r_n^{-1}(A_n) \mid n \geq 1,\ A_n \in \mathcal{M}_n^{'}\}$ the set formed by all of these preimages, we can also observe that $\mathcal{A}$ is an algebra.

Lemma 4.6. *The family of sets $\mathcal{A} = \{r_n^{-1}(A_n) \mid n \geq 1,\ A_n \in \mathcal{M}_n^{'}\}$ is an algebra of $M^{'}(E)$.*

Proof. Indeed, as $\emptyset \in \mathcal{M}_1^{'}$, we have that $\emptyset = r_1^{-1}(\emptyset)$ and thus $\emptyset \in \mathcal{A}$. Moreover, given any $1 \leq n \leq m$ and $A_n \in \mathcal{M}_n^{'}$, $A_m \in \mathcal{M}_m^{'}$, we can observe that:

$$r_n^{-1}(A_n) \cup r_m^{-1}(A_m) = (r_{m,n} \circ r_m)^{-1}(A_n) \cup r_m^{-1}(A_m) = r_m^{-1}(r_{m,n}^{-1}(A_n)) \cup r_m^{-1}(A_m) = r_m^{-1}(r_{m,n}^{-1}(A_n) \cup A_m)$$

so we conclude that $r_n^{-1}(A_n) \cup r_m^{-1}(A_m) \in \mathcal{A}$.

Finally, by properties of the preimages, we also have that $M'(E) \setminus r_n^{-1}(A_n) = r_n^{-1}(M'(K_n)) \setminus r_n^{-1}(A_n) = r_n^{-1}(M'(K_n) \setminus A_n)$, so $M'(E) \setminus r_n^{-1}(A_n) \in \mathcal{A}$. That shows that $\mathcal{A}$ is an algebra of $M'(E)$. □

From here, we can define in $\mathcal{A}$ the map $P^{'} : \mathcal{A} \longrightarrow [0,1]$ given by $P^{'}(A) = P_n(A_n)$ for all $A = r_n^{-1}(A_n) \in \mathcal{A}$ and now our objective is to show that it is a well-defined pre-measure in $(M'(E), \mathcal{A})$. To do it, given any $1 \leq n \leq m$ and $A_n \in \mathcal{M}_n^{'}$, $A_m \in \mathcal{M}_m^{'}$ such that $r_n^{-1}(A_n) = r_m^{-1}(A_m)$, we can deduce that $r_m^{-1}(r_{m,n}^{-1}(A_n)) = r_m^{-1}(A_m)$ and thus $r_m(r_m^{-1}(r_{m,n}^{-1}(A_n))) = r_m(r_m^{-1}(A_m))$, something that combined with the fact that r_m is exhaustive implies that $r_{m,n}^{-1}(A_n) = A_m$. So, we conclude that $P_n(A_n) = P_m(r_{m,n}^{-1}(A_n)) = P_m(A_m)$, something that shows that $P^{'}$ is well-defined.

Not only that, but given any $1 \leq n \leq m$ and $A_n \in \mathcal{M}_n^{'}$, $A_m \in \mathcal{M}_m^{'}$ such that $r_n^{-1}(A_n) \cap r_m^{-1}(A_m) = \emptyset$, we can also observe that:

$$r_n^{-1}(A_n) \cap r_m^{-1}(A_m) = \emptyset \iff r_m^{-1}(r_{m,n}^{-1}(A_n)) \cap r_m^{-1}(A_m) = \emptyset \iff r_m^{-1}(r_{m,n}^{-1}(A_n) \cap A_m) = \emptyset \iff r_{m,n}^{-1}(A_n) \cap A_m = \emptyset$$

since these restriction maps are, again, clearly exhaustive. As a consequence, we can deduce that:

$$P^{'}(r_n^{-1}(A_n) \cup r_m^{-1}(A_m)) = P_m(r_{m,n}^{-1}(A_n) \cup A_m) =$$

$$P_m(r_{m,n}^{-1}(A_n)) + P_m(A_m) = P_n(A_n) + P^{'}(r_m^{-1}(A_m)) = P^{'}(r_n^{-1}(A_n)) + P^{'}(r_m^{-1}(A_m))$$

That allows us to conclude that $P^{'}$ is additive. Moreover, it is also clear that $P'(\emptyset) = P_1(\emptyset) = 0$ and $P'(M^{'}(E)) = P_1(M^{'}(K_1)) = 1$.

Finally, regarding the σ-additivity of $P^{'}$, as we have already proven that $P^{'}$ is additive and finite ($P^{'}(M^{'}(E)) = 1$), it is a very well-known fact that its σ-additivity is equivalent to $P^{'}$ being sequentially continuous.

Lemma 4.7. *The map $P^{'}$ considered above is sequentially continuous.*

Proof. Indeed, given any family $\{F_i\}_{i=1}^{\infty} \subseteq \mathcal{A}$ such that $F_{i+1} \subseteq F_i\ \forall i \geq 1 \in \mathbb{N}$ and $\bigcap_{i=1}^{\infty} F_i = \emptyset$, it suffices to show that $\lim_i P^{'}(F_i) = 0$. To do it, we proceed by reduction to absurdity, assuming that $\lim_i P^{'}(F_i) \neq 0$ and thus, as it is a decreasing sequence

that is bounded by below, we can deduce that there must exist $\epsilon > 0 \in \mathbb{R}$ such that $\lim_i P^{'}(F_i) = \epsilon$ and, in particular, $P^{'}(F_i) \geq \epsilon$ for all $i \geq 1$.

With that in mind, given any $i \geq 1$, we can also observe that $F_i \in \mathcal{A}$, so by definition there exists $n_i \geq 1$ and $V_i \in \mathcal{M}_{n_i}^{'}$ such that $F_i = r_{n_i}^{-1}(V_i)$. Not only that, but given any $n \geq 1$, we can notice that:

$$\epsilon \leq \lim_i P^{'}(F_i) \leq \lim_i P^{'}(r_n^{-1}(r_n(F_i))) = \lim_i P_n(r_n(F_i))$$

something that combined with the fact that P_n is a probability measure shows that $P_n(\bigcap_{i=1}^{\infty} r_n(F_i)) \geq \epsilon$ and thus $P^{'}(r_n^{-1}(\bigcap_{i=1}^{\infty} r_n(F_i))) = P^{'}(\bigcap_{i=1}^{\infty} r_n^{-1}(r_n(F_i))) \geq \epsilon$.

Then, we can define the set $W_n = \bigcap_{i=1}^{\infty} r_n^{-1}(r_n(F_i))$ and it is clear that we have that $P^{'}(W_n) \geq \epsilon$ and $\bigcap_{n=1}^{\infty} W_n = \bigcap_{n=1}^{\infty} \bigcap_{i=1}^{\infty} r_n^{-1}(r_n(F_i)) = \bigcap_{i=1}^{\infty} \bigcap_{n=1}^{\infty} r_n^{-1}(r_n(F_i)) = \bigcap_{i=1}^{\infty} F_i = \emptyset$. Moreover, we can also observe that for any $1 \leq n \leq m$ we have that:

$$W_m = \bigcap_{i=1}^{\infty} r_m^{-1}(r_m(F_i)) \subseteq \bigcap_{i=1}^{\infty} r_m^{-1}(r_{m,n}^{-1}(r_{m,n}(r_m(F_i)))) = \bigcap_{i=1}^{\infty} r_n^{-1}(r_n(F_i)) = W_n$$

So, at the end, without loss of generality, we can consider that we have a family $\{F_i\}_{i=1}^{\infty} \subseteq \mathcal{A}$ such that $\bigcap_{i=1}^{\infty} F_i = \emptyset$ and that for all $i \geq 1$ verifies that $F_{i+1} \subseteq F_i$, $P^{'}(F_i) \geq \epsilon$ and $F_i = r_i^{-1}(G_i)$ for some $G_i \in \mathcal{M}_i^{'}$.

That being said, given any $n \geq 1 \in \mathbb{N}$, recall that $M(K_n)$ is the space of all Radon measures in K_n. In addition, recall that $K_n \subseteq E$ is a Polish space because E is Polish and closed subspaces of Polish spaces are Polish, as stated in an example in Section 26 of [3]. That, combined with the fact that every finite Borel measure in K_n is a Radon measure (as deduced from Theorem 7.1.7 in [5]), allows us to conclude that, in fact, $M(K_n)$ is also the space of finite Borel measures in K_n. Then, in this space, we can consider the corresponding weak-star topology that, in this context, is the topology that has as a subbasis the sets:

$$\{\mu \in M(K_n) \mid \mu(C) < l\} \text{ for any } C \subseteq K_n \text{ closed and } l \in \mathbb{R}$$
$$\{\mu \in M(K_n) \mid \mu(K_n) > l\} \text{ for any } l \in \mathbb{R}$$

In this situation, using again the fact that K_n is a Polish space and applying Theorem 7.11 of [24], we can conclude that $M(K_n)$ is also a Polish space with this topology. In fact, not only that, but at the same time we can deduce that the Borel σ-algebra given by this weak-star topology coincides with our original σ-algebra $\mathcal{M}_n$, applying analogous arguments to the ones that are stated in Theorem 2.3 of [8].

Finally, considering the corresponding subspace $M^{'}(K_n)$, we now know that it is also a Polish space. Indeed, to show it, it suffices to see that $M^{'}(K_n)$ is a closed subspace. To do it, given any $\{\mu_n\}_{n\in\mathbb{N}} \subseteq M^{'}(K_n)$ sequence of point measures and $\mu \in M(K_n)$ such that $\lim_n \mu_n = \mu$, we can apply Proposition 4.9 of [24] to observe that this convergence is equivalent to the fact that $\lim_n \mu_n(D) = \mu(D)$ for any $D \in \mathcal{B}(K_n)$ such that $\mu(\partial D) = 0$. Then, in this context, it is clear that μ must be integer-valued and thus also a point measure, so $\mu \in M^{'}(K_n)$ and we conclude that $M^{'}(K_n)$ is closed.

As a consequence, applying again Theorem 7.1.7 in [5], we can deduce that the probability measure P_n is inner-regular in $(M^{'}(K_n), \mathcal{M}_n^{'})$, as desired.

From here, given any $1 \leq j \leq i$, we can start by observing that $F_i \subseteq F_j$ and thus $r_i^{-1}(G_i) \subseteq r_j^{-1}(G_j)$, something that allows us to conclude that $r_i^{-1}(G_i) \subseteq r_i^{-1}(r_{i,j}^{-1}(G_j))$ and $G_i \subseteq r_{i,j}^{-1}(G_j)$.

Moreover, given again any $1 \leq j \leq i$, we can also observe that the restriction map $r_{i,j} : M^{'}(K_i) \longrightarrow M^{'}(K_j)$ is continuous with the weak-star topologies we are considering and now we proceed to prove it. Indeed, given any $\mu \in M^{'}(K_i)$ and $\{\mu_n\}_{n\geq 1} \subseteq M^{'}(K_i)$ such that $\lim_n \mu_n = \mu$, to conclude the result suffices to see that $\lim_n r_{i,j}(\mu_n) = r_{i,j}(\mu)$. To do it, as we know that $\lim_n \mu_n = \mu$, applying Proposition 4.9 of [24], we can deduce that $\lim_n \mu_n(D) = \mu(D)$ for any $D \in \mathcal{B}(K_i)$ such that $\mu(\partial D) = 0$ and, in particular, $\lim_n \mu_n(D) = \mu(D)$ for any $D \in \mathcal{B}(K_j)$ such that $\mu(\partial D) = 0$, something that implies that $\lim_n(r_{i,j}(\mu_n))(D) = (r_{i,j}(\mu))(D)$ for any $D \in \mathcal{B}(K_j)$ such that $\mu(\partial D) = 0$. Finally, an application of the same proposition allows us to conclude the desired result and shows that the restriction map $r_{i,j}$ is continuous.

With that in mind, using that we have already seen that the probability measure P_i is inner-regular, we can conclude that there exists a compact subset $C_i \subseteq G_i$ such that $P_i(C_i) \geq P_i(G_i) - \frac{\epsilon}{2^{i+1}}$. Then, using these compact sets, for any $i \geq 1$, we can define the set $C_i^{'} = \bigcap_{j=1}^{i} r_{i,j}^{-1}(C_j)$ and now we proceed to study their properties.

Indeed, given any $i \geq 1$, we can start by seeing that $C_i^{'} = \bigcap_{j=1}^{i} r_{i,j}^{-1}(C_j) \subseteq C_i \subseteq G_i$. Moreover, as the maps $r_{i,j}$ are continuous, we can observe that $C_i^{'}$ is an intersection of closed sets and thus it is closed, something that combined with the fact that $C_i^{'} \subseteq C_i$ and C_i is compact allows us to conclude that $C_i^{'}$ is also compact. In addition, we can also deduce that:

$$P_i(C_i^{'}) \geq P_i(G_i) - \sum_{j=1}^{i} \frac{\epsilon}{2^{j+1}} \geq \epsilon - \sum_{j=1}^{i} \frac{\epsilon}{2^{j+1}} \geq \epsilon/2 > 0$$

where we have used that $P_i(r_{i,j}^{-1}(C_j)) = P_j(C_j)$. So, in particular, $C_i^{'}$ is not empty.

Not only that, but given any $1 \leq j \leq i$, notice that we also have that:

$$C_i^{'} = \bigcap_{k=1}^{i} r_{i,k}^{-1}(C_k) \subseteq \bigcap_{k=1}^{j} r_{i,k}^{-1}(C_k) = \bigcap_{k=1}^{j} (r_{j,k} \circ r_{i,j})^{-1}(C_k) = \bigcap_{k=1}^{j} r_{i,j}^{-1}(r_{j,k}^{-1}(C_k)) = r_{i,j}^{-1}(\bigcap_{k=1}^{j} r_{j,k}^{-1}(C_k)) = r_{i,j}^{-1}(C_j^{'})$$

so we can conclude that $C_i^{'} \subseteq r_{i,j}^{-1}(C_j^{'})$.

From here, for any $i \geq 1$, we can consider the set $H_i = r_i^{-1}(C_i^{'})$ and then we can start by observing that $H_i \neq \emptyset$ because $C_i^{'} \neq \emptyset$ and r_i exhaustive. Moreover, it is also easy to see that these sets are decreasing, as $H_{i+1} = r_{i+1}^{-1}(C_{i+1}^{'}) \subseteq r_{i+1}^{-1}(r_{i+1,i}^{-1}(C_i^{'})) = r_i^{-1}(C_i^{'}) = H_i$. In addition, we can also deduce that $H_i = r_i^{-1}(C_i^{'}) \subseteq r_i^{-1}(G_i) = F_i$ and thus $\bigcap_{i=1}^{\infty} H_i = \emptyset$, because we know that $\bigcap_{i=1}^{\infty} F_i = \emptyset$.

Now, in this setting, we will try to arrive to a contradiction. To do it, using the axiom of choice, we can start by selecting a measure $\mu_i \in H_i$ for each $i \geq 1$ and considering the corresponding sequence $\{\mu_i\}_{i\geq 1} \subseteq M^{'}(E)$. Notice that for any $i_0 \geq 1$ and $i \geq i_0$, we have that:

$$r_{i_0}(H_i) = r_{i_0}(r_i^{-1}(C_i^{'})) = r_{i,i_0}(r_i(r_i^{-1}(C_i^{'}))) = r_{i,i_0}(C_i^{'}) \subseteq r_{i,i_0}(r_{i,i_0}^{-1}(C_{i_0}^{'})) = C_{i_0}^{'}$$

so, in particular, we can deduce that $r_{i_0}(\mu_i) \in r_{i_0}(H_i) \subseteq C_{i_0}^{'}$.

Now, recalling the properties of compact sets and that $C_1^{'}$ is compact, we can start by finding a subsequence $\{\mu_{i,1}\}_{i\geq 1}$ of $\{\mu_i\}_{i\geq 1}$ such that $\{r_1(\mu_{i,1})\}_{i\geq 1}$ converges. Then, using now that $C_2^{'}$ is compact, we can also find a subsequence of this last one, $\{\mu_{i,2}\}_{i\geq 1}$, such that $\{r_2(\mu_{i,2})\}_{i\geq 1}$ converges. Iteratively, applying the same idea, we can obtain nested subsequences $\{\mu_{i,n}\}_{i\geq 1}$ for all $n \geq 1$ such that $\{r_n(\mu_{i,n})\}_{i\geq 1}$ converges.

From here, using these convergent subsequences and applying the usual diagonalization argument, we can consider now the sequence $\{\mu_{i,i}\}_{i\geq 1}$ obtained from the previous ones. In this situation, thanks to the fact that the subsequences of a convergent sequence are convergent, we can see that $\{r_n(\mu_{i,i})\}_{i\geq 1}$ converges when $i \to \infty$ for each $n \geq 1$, and it must do so to a point measure that we can denote by $\eta_n \in C_n^{'}$.

Focusing our attention on these limit measures, we can now observe that they are compatible. Indeed, given any $1 \leq n \leq m$, we can deduce that $r_n(\mu_{i,i}) = r_{m,n}(r_m(\mu_{i,i}))$ and then, taking the limit $i \to \infty$ and using the continuity of $r_{m,n}$, we can conclude that $\eta_n = r_{m,n}(\eta_m)$, that is what we wanted to see.

Finally, using the previous restricted measures obtained as limits, we can now define, for any $n \geq 1$, the measure $\overline{\eta}_n \in M^{'}(E)$ as the one that coincides with η_n in K_n and that verifies that $\overline{\eta}_n(E \setminus K_n) = 0$. Notice that it is clear that $\overline{\eta}_n$ is a well-defined measure because η_n is it. Then, with these measures, we can also introduce the measure $\mu : \mathcal{B}(E) \longrightarrow [0, \infty]$ given by $\mu(D) = \lim_n \overline{\eta}_n(D)$ for all $D \in \mathcal{B}(E)$ and now we proceed to show that it is indeed a measure.

To do it, we can observe that the limit of the definition exists because the sequence is increasing, that $\mu(\emptyset) = \lim_n \overline{\eta}_n(\emptyset) = \lim_n 0 = 0$ and that $\mu(D) = \lim_n \overline{\eta}_n(D) \geq 0$ for all $D \in \mathcal{B}(E)$. Moreover, given any family $\{D_i\}_{i\geq 1} \subseteq \mathcal{B}(E)$ of pairwise disjoint sets, we can observe that:

$$\mu(\bigcup_{i=1}^{\infty} D_i) = \lim_n \overline{\eta}_n(\bigcup_{i=1}^{\infty} D_i) = \lim_n \sum_{i=1}^{\infty} \overline{\eta}_n(D_i) = \sum_{i=1}^{\infty} \lim_n \overline{\eta}_n(D_i) = \sum_{i=1}^{\infty} \mu(D_i)$$

where we have applied the monotonous convergence theorem. That shows that μ is a measure and it is also clear that it must be a point measure by definition, so $\mu \in M^{'}(E)$.

With that in mind, given any $n \geq 1$ and any $D \in \mathcal{B}(K_n)$, now we can notice that $\mu(D) = \lim_n \overline{\eta}_n(D) = \eta_n(D)$. That shows that, for every $n \geq 1$, we have that $r_n(\mu) = \eta_n \in C_n^{'}$ and thus, by definition, $\mu \in H_n$, something that implies that $\mu \in \bigcap_{i=1}^{\infty} H_i = \emptyset$. This is a clear contradiction, so we have managed to show that $P^{'}$ is σ-additive. $\square$

From here, once that we know that $P^{'}$ is σ-additive, we can conclude that $P^{'}$ is indeed a pre-measure in $\mathcal{A}$. As a consequence, using the fact that $\mathcal{M}^{'}$ is the σ-algebra generated by $\mathcal{A}$ and applying the Hahn-Kolmogorov theorem [23], we can claim that there exists P a probability measure in $(M^{'}(E), \mathcal{M}^{'})$ such that it extends $P^{'}$.

With that in mind, now it will be easy to show that there exists a determinantal point process with kernel $\mathbb{K}$ defined in E. To do it, we can start by considering the above probability space $(M^{'}(E), \mathcal{M}^{'}, P)$ and the corresponding inclusion map defined

by it $\mathscr{X} : (M^{'}(E), \mathcal{M}^{'}) \longrightarrow (M(E), \mathcal{M})$. Then, we will see that $\mathscr{X}$ is actually the desired determinantal point process.

Lemma 4.8. *The map $\mathscr{X}$ defined above is a determinantal point process with kernel $\mathbb{K}$ defined in E.*

Proof. To begin, it is clear that $\mathscr{X}$ is a measurable map, because we can recall that the σ-algebra $\mathcal{M}^{'}$ is the one induced in $M^{'}(E) \subseteq M(E)$ from the original σ-algebra $\mathcal{M}$. That shows, by definition, that $\mathscr{X}$ is a point process in E. Not only that, but given any $x \in E$, we can also observe that there exists $n \geq 1$ such that $x \in K_n$, since $E = \bigcup_{n=1}^{\infty} K_n$. As a consequence, we can apply the definition of $\mathscr{X}$ to conclude that:

$$P(\mathscr{X}(\{x\}) \in \{0,1\}) = P(\{\mu \in M^{'}(E) \mid \mu(\{x\}) \in [0, \tfrac{3}{2})\}) =$$
$$P_n(\{\mu \in M^{'}(K_n) \mid \mu(\{x\}) \in [0, \tfrac{3}{2})\}) = \tau_n(\mathscr{X}_n(\{x\}) \in \{0,1\}) = 1$$

because we know that $\mathscr{X}_n$ is a simple point process. That shows, by definition, that $\mathscr{X}$ is in fact a simple point process in E.

Moreover, given any $k \geq 1$ and $D_1, \ldots, D_k \in \mathcal{B}(E)$ pairwise disjoint Borel sets, we can observe that:

$$\mathbb{E}(\textstyle\prod_{i=1}^{k} \mathscr{X}(D_i)) =$$
(by sequential continuity of measures and the fact that $D_i = \bigcup_{n=1}^{\infty} D_i \cap K_n \ \forall i$) =
$$\mathbb{E}(\textstyle\prod_{i=1}^{k} \lim_n \mathscr{X}(D_i \cap K_n)) = \mathbb{E}(\lim_n \prod_{i=1}^{k} \mathscr{X}(D_i \cap K_n)) =$$
(by the monotonous convergence theorem) =
$$\lim_n \mathbb{E}(\textstyle\prod_{i=1}^{k} \mathscr{X}(D_i \cap K_n)) = \lim_n \mathbb{E}(\prod_{i=1}^{k} \mathscr{X}_n(D_i \cap K_n)) =$$
$$\lim_n \textstyle\int_{\prod_{i=1}^{k} D_i \cap K_n} \det(\mathbb{K}_{|K_n^2}(x_i, x_j))_{i,j \leq k} d\mu(x_1) \ldots d\mu(x_k) =$$
$$\lim_n \textstyle\int_{\prod_{i=1}^{k} D_i \cap K_n} \det(\mathbb{K}(x_i, x_j))_{i,j \leq k} d\mu(x_1) \ldots d\mu(x_k) =$$
$$\lim_n \textstyle\int_{\prod_{i=1}^{k} D_i} \mathcal{X}_{K_n^k} \det(\mathbb{K}(x_i, x_j))_{i,j \leq k} d\mu(x_1) \ldots d\mu(x_k) =$$
(again, by the monotonous convergence theorem) =
$$\textstyle\int_{\prod_{i=1}^{k} D_i} \det(\mathbb{K}(x_i, x_j))_{i,j \leq k} d\mu(x_1) \ldots d\mu(x_k)$$

where we have used the fact that $\mathbb{E}(\prod_{i=1}^{k} \mathscr{X}(D_i \cap K_n)) = \mathbb{E}(\prod_{i=1}^{k} \mathscr{X}_n(D_i \cap K_n))$. Indeed, it is easy to see that:

$$P(\mathscr{X}(D_i \cap K_n) = h) = P(\{\mu \in M^{'}(E) \mid \mu(D_i \cap K_n) = h\}) =$$
(by the definition of the probability measures P and $P^{'}$) =
$$P_n(\{\mu \in M^{'}(K_n) \mid \mu(D_i \cap K_n) = h\}) = \tau_n(\mathscr{X}_n(D_i \cap K_n) = h)$$

for all $1 \leq i \leq k$, $n \geq 1$ and $h \geq 0 \in \mathbb{N}$, something that implies the desired equality of expected values.

The above calculations show that the k-th joint intensity of $\mathscr{X}$ w.r.t μ exists and it is given by $\rho_k(x_1, \ldots, x_k) = \det(\mathbb{K}(x_i, x_j))_{i,j \leq k}$ μ^k-a.e. for $(x_1, \ldots, x_k) \in E^k$. So, at the end, we have managed to show that $\mathscr{X}$ is a determinantal point process with kernel $\mathbb{K}$ defined in E. $\square$

So, in particular, we now know that there exists a determinantal point process with kernel $\mathbb{K}$ in E, that is what we wanted to conclude. That completes the proof of the original existence theorem. $\square$

4.2. **Uniqueness of DPPs.** Finally, once we have studied the existence of determinantal point processes, now we can focus our attention on their uniqueness. Indeed, to do it, we start by recalling that we already know thanks to our preliminaries (Proposition 2.16) that if a point process has exponential tails, then its joints intensities determine its distribution.

So, to show that determinantal point processes are actually unique, having their distributions determined by their corresponding kernel, it will suffice to prove that these point processes have exponential tails.

Proposition 4.9. *Let $\mathscr{X}$ be a determinantal point process in E with kernel $\mathbb{K}$, then we have that $\mathscr{X}$ has exponential tails for all compact sets in E, that is, $\forall K \subseteq E$ compact, there exists $c > 0$ and $k_0 \in \mathbb{N}$ such that $P(\mathscr{X}(K) \geq k) \leq e^{-ck}$ $\forall k \geq k_0$.*

Proof. Indeed, continuing with the notations introduced above, we can start by recalling Hadamard's inequality, that claims that given $A = (a_{i,j})_{i,j} \in \mathbb{C}^{n\times n}$ any Hermitian and positive semi-definite matrix, then we have that $\det(A) \leq \prod_{i=1}^n a_{i,i}$. To prove it, we merely need to apply the Cholesky decomposition of the original matrix $A = CC^*$ with $C = (c_{i,j})_{i,j}$ and conclude that:

$$\det(A) = \det(CC^*) = \det(C)\det(C^*) = |\det(C)|^2 = \prod_{i=1}^n |c_{i,i}|^2 \leq \prod_{i=1}^n \|c_i\|^2 = \prod_{i=1}^n \langle c_i, c_i\rangle = \prod_{i=1}^n (\sum_{j=1}^n c_{i,j}\overline{c_{i,j}}) = \prod_{i=1}^n a_{i,i}$$

where c_i denotes the i-th row of C and $\langle\ ,\ \rangle$ the usual inner product in $\mathbb{C}^n$.

With that in mind, let $n \geq 1 \in \mathbb{N}$ be any natural number and let $(x_1, \dots, x_n) \in K^n$, then we already know that there exist $\{\lambda_k\}_k$ non-negative and $\{\varphi_k\}_k$ an orthonormal basis of $L^2(K,\mu)$ such that we have that $\mathbb{K}(x,y) = \sum_k \lambda_k\varphi_k(x)\overline{\varphi_k}(y)$ $\mu - a.e.$ for $x, y \in K$, which implies that:

$$\mathbb{K}(x_i, x_j) = \overline{\mathbb{K}(x_j, x_i)} \quad \forall 1 \leq i, j \leq n$$

$$z^*(\mathbb{K}(x_i,x_j))_{i,j\leq n}z = z^*(\sqrt{\lambda_k}\varphi_k(x_i))_{i,k}((\sqrt{\lambda_k}\varphi_k(x_i))_{i,k})^*z =$$
$$= (((\sqrt{\lambda_k}\varphi_k(x_i))_{i,k})^*z)^*((\sqrt{\lambda_k}\varphi_k(x_i))_{i,k})^*z = \|((\sqrt{\lambda_k}\varphi_k(x_i))_{i,k})^*z\|_2^2 \geq 0 \quad \forall z \in \mathbb{C}^n$$

so have that the matrix $(\mathbb{K}(x_i,x_j))_{i,j\leq n}$ is Hermitian and positive semi-definite μ^k-a.e.

Using this fact, and combining it with the properties we saw in the preliminaries, we can also observe that:

$$\mathbb{E}(\tbinom{\mathscr{X}(K)}{n}n!) = \int_{K^n} \det(\mathbb{K}(x_i,x_j))_{i,j\leq n} d\mu(x_1)\dots d\mu(x_n) \leq$$
$$\text{(using Hadamard's inequality)} \leq \int_{K^n} \prod_{i=1}^n \mathbb{K}(x_i,x_i) d\mu(x_1)\dots d\mu(x_n) =$$
$$(\int_K \mathbb{K}(x_1,x_1)d\mu(x_1))\dots(\int_K \mathbb{K}(x_n,x_n)d\mu(x_n)) = (\int_K \mathbb{K}(x,x)d\mu(x))^n$$

Now, thanks to the above deductions, we are ready to show that $\mathscr{X}$ has exponential tails. Indeed, we set $\kappa(K) = \int_K \mathbb{K}(x,x)d\mu(x)$ and for any fixed $s > 0 \in \mathbb{R}$ one has:

$$P(\mathscr{X}(K) \geq k) = P((s+1)^{\mathscr{X}(K)} \geq (s+1)^k) \leq \text{ (using Markov's inequality) } \leq$$
$$\frac{\mathbb{E}((s+1)^{\mathscr{X}(K)})}{(s+1)^k} = (s+1)^{-k}\mathbb{E}((s+1)^{\mathscr{X}(K)}) = \text{(by Newton's binomial formula)} =$$
$$e^{\log((s+1)^{-k})}\mathbb{E}(\sum_{i=0}^{\mathscr{X}(K)} \binom{\mathscr{X}(K)}{i} s^i) = \text{(if } \mathscr{X}(K) < i\text{, then } \binom{\mathscr{X}(K)}{i} = 0) =$$
$$e^{-k\log(s+1)}\mathbb{E}(\sum_{i=0}^{\infty} \binom{\mathscr{X}(K)}{i} s^i) = \text{(by the monotonous convergence theorem)} =$$
$$e^{-k\log(s+1)} \sum_{i=0}^{\infty} \mathbb{E}(\binom{\mathscr{X}(K)}{i}) s^i = e^{-k\log(s+1)} \sum_{i=0}^{\infty} \mathbb{E}(\binom{\mathscr{X}(K)}{i} i!) \frac{s^i}{i!} \leq$$
$$e^{-k\log(s+1)} \sum_{i=0}^{\infty} \frac{\kappa(K)^i s^i}{i!} = e^{-k\log(s+1)} e^{\kappa(K)s} = e^{\kappa(K)s - k\log(s+1)} =$$
$$e^{-k(\log(s+1) - \frac{\kappa(K)s}{k})} \leq e^{-kc}$$

for $k \geq \lfloor \frac{2\kappa(K)s}{\log(s+1)} \rfloor + 1$ and $c = \frac{\log(s+1)}{2}$.

So we merely need to take $k_0 = \lfloor \frac{2\kappa(K)s}{\log(s+1)} \rfloor + 1$, $c = \frac{\log(s+1)}{2}$ and they verify the desired property. That completes the proof. □

In conclusion, as already explained, using this last result one can deduce that determinantal point processes, if they exist, are unique. That means that, if one considers two determinantal point processes $\mathscr{X}_1$ and $\mathscr{X}_2$ with kernel $\mathbb{K}$, then $\mathscr{X}_1$ and $\mathscr{X}_2$ actually follow the same distribution and behave in the same way.

5. Algorithms to sample DPPs

From here, although in the previous sections we have already given the definition of a determinantal point process and shown their existence under the proper assumptions, the descriptions that we have seen until now of these point processes do not give us an immediate and easy way to sample them, something necessary in practical applications, that will be discussed in Section 8.

With that in mind, to solve this issue and mainly following Chapter 4.4 of [11] (filling in some small details by ourselves), now we will provide a general algorithm to sample a determinantal point process of a given kernel $\mathbb{K}$. Indeed, to do it, let $\mathscr{X}$ be a determinantal point process in E with kernel $\mathbb{K} : E^2 \longrightarrow \mathbb{C}$, we can start by observing, applying our definition, that $\mathbb{K}$ must be a locally square integrable and Hermitian kernel such that $\mathscr{K}$ is locally of trace class and non-negative definite.

In fact, we further assume that $\mathbb{K}$ is square integrable and $\mathscr{K}$ is of trace class. That ensures that the number of points of our point process is finite almost surely, something can be easily deduced from our preliminaries and the results and proofs of Proposition 4.2 and Theorem 4.3. Then, thanks to this extra assumption, we can safely try to sample our process in practical applications, something that clearly would not be possible if the number of points to sample was infinite. Notice that this is a reasonable assumption that is satisfied, for example, when we consider our original determinantal point process restricted to a compact set.

Now, that being said, we can apply the results we have seen in the preliminaries to conclude that there exists $m \in \mathbb{N} \cup \{\infty\}$ (for practical applications we could consider that m is finite), $\{\varphi_k\}_{k=1}^m$ an orthonormal basis of $L^2(E,\mu)$ and $\{\lambda_k\}_{k=1}^m \subseteq \mathbb{R}$ a set of non-negative eigenvalues such that $\mathbb{K}(x,y) = \sum_{k=1}^m \lambda_k \varphi_k(x)\overline{\varphi_k}(y)$ μ-a.e. for $x, y \in E$. Not only that, but applying the existence Theorem 4.4, we can deduce that these eigenvalues of $\mathscr{K}$ must be in the interval $[0,1]$, that is, $\lambda_i \in [0,1]$ for all $1 \leq k \leq m$.

That, combined with the result stated in Theorem 4.3, tells us that to sample our determinantal point process it will be enough to sample first m independent Bernoulli random variables of parameter λ_k respectively and then sample the corresponding projection determinantal point process. So, without loss of generality, we can focus on providing an algorithm to sample a DPP whose kernel $\mathbb{K}$ is given by $\mathbb{K}(x,y) = \sum_{k=1}^n \varphi_k(x)\overline{\varphi_k}(y)$ for all $x, y \in E$, where n is finite and $\{\varphi_k\}_{k=1}^n$ now denotes an orthonormal system of $L^2(E,\mu)$. Here, it is important to notice that, usually, the functions φ_k are actually equivalence classes and their images are not well-defined. Nevertheless, since we can assume that in this sampling context $\mathbb{K}$ is completely determined, we can fix adequate representatives of these functions to ensure that, indeed, $\mathbb{K}(x,y) = \sum_{k=1}^n \varphi_k(x)\overline{\varphi_k}(y)$ $\forall x, y \in E$. Thanks to this, we can evaluate the expressions that appear in what remains of the section without problems.

That being said, focusing on this type of kernels, recall that we have already explained that they are called projection kernels because their associated integral operator is the projection operator on the subspace generated by its eigenvectors of eigenvalue 1, that is, the projection operator on the subspace $H = \mathrm{Span}(\{\varphi_k\}_{k=1}^n)$.

In this situation, as a preliminary before stating our sampling algorithm, we can start by seeing that $\mathbb{K}$ is the reproducing kernel of H and, in particular, H is a reproducing kernel Hilbert space. Indeed, to do it, given any $f \in H$ and $x \in E$, we merely need to observe that $\langle f, \mathbb{K}(\cdot, x)\rangle = (\mathscr{K} f)(x) = f(x)$. Not only that, but given any $V \subseteq H$ subspace, it is clear that V must also be a reproducing kernel Hilbert space. So, we can denote by $\mathbb{K}_V$ the reproducing kernel of V and thus $\mathbb{K}_H = \mathbb{K}$.

As a consequence, given any $x \in E$, it is convenient to consider the function $\mathbb{K}(\cdot, x)$ defined E and we can denote it by $\mathscr{K}\delta_x$. As a comment, recalling that δ_x corresponds to the Dirac measure of x, notice that this notation is coherent, because heuristically:

$$\mathscr{K}\delta_x = \int_E \mathbb{K}(\cdot, y)\delta_x(y)d\mu(y) = \mathbb{K}(\cdot, x)$$

something that especially works when E is discrete and thus δ_x is given by $\delta_x(\{y\}) = 1$ if $y = x$ and $\delta_x(\{y\}) = 0$ otherwise. Similarly, given any subspace $V \subseteq H$, we can also denote by $\mathscr{K}_V\delta_x$ its own $\mathbb{K}_V(\cdot, x)$ function based on the reproducing kernel of V.

Finally, we can also consider the measure μ_H in E as the one that is absolutely continuous with respect to μ (the original Radon measure in E) with density function $f : E \longrightarrow [0, \infty)$ given by $f(x) = \|\mathscr{K}_H\delta_x\|^2 = \|\mathbb{K}(\cdot, x)\|^2$ for all $x \in E$. Then, we can deduce that, since $\dim(H) = n$, the normalized measure $\frac{1}{n}\mu_H$ is a probability measure in E. Indeed, to show it, we merely need to observe that:

$$\frac{\mu_H(E)}{n} = \int_E \frac{1}{n} f(x)d\mu(x) = \int_E \frac{1}{n}\|\mathbb{K}(\cdot, x)\|^2 d\mu(x) = \int_E \frac{1}{n}\|\sum_{k=1}^n \varphi_k(x)\overline{\varphi_k}\|^2 d\mu(x) =$$
$$\text{(using that the functions } \{\varphi_k\}_{k=1}^n \text{ form an orthonormal system)} =$$
$$\int_E \frac{1}{n}\sum_{k=1}^n |\varphi_k(x)|^2 d\mu(x) = \sum_{k=1}^n \frac{1}{n}\int_E |\varphi_k(x)|^2 d\mu(x)$$
$$\text{(using again the same property of the functions } \{\varphi_k\}_{k=1}^n) =$$
$$\sum_{k=1}^n \frac{1}{n} = 1$$

so it is a probability measure. Moreover, given any subspace $V \subseteq H$, we can also consider its corresponding induced measure μ_V in E as the one absolutely continuous with respect to μ that has as its density $f : E \longrightarrow [0, \infty)$ given by $f(x) = \|\mathscr{K}_V\delta_x\|^2 = \|\mathbb{K}_V(\cdot, x)\|^2$ for all $x \in E$. Then, using exactly the same arguments as before with an orthonormal basis of V, we can conclude that $\frac{1}{\dim(V)}\mu_V$ is a probability measure in E.

With all of the above in mind, now we are ready to provide an algorithm to sample these finite dimensional projection determinantal point processes.

Algorithm 1: Sampling of a finite projection DPP algorithm

Input : The kernel $\mathbb{K}$, the space $H_n = H$ and $n \geq 1 \in \mathbb{N}$.
Output: The set of sampled points $\{x_1, \ldots, x_n\} \subseteq E$.
while $n \geq 1$ **do**
(1) Sample a point x_n in E according to the probability measure $\frac{1}{n}\mu_{H_n}$.
(2) Consider $H_{n-1} \subset H_n$ the orthogonal complement of the function $\mathscr{K}_{H_n}\delta_{x_n}$ in H_n. Notice that, in fact, $H_{n-1} = \{f \in H_n \mid \langle f, \mathscr{K}_{H_n}\delta_{x_n}\rangle = 0\} = \{f \in H_n \mid \langle f, \mathbb{K}_{H_n}(\cdot, x_n)\rangle = 0\} = \{f \in H_n \mid f(x_n) = 0\}$.
(3) Decrease n by 1.

Now, we proceed to show that the points generated by the above algorithm do in fact follow the distribution of the points of a determinantal point process with kernel $\mathbb{K}$, as stated by the following theorem.

Theorem 5.1. *Let $\{\varphi_k\}_{k=1}^n$ be a finite orthonormal system of $L^2(E,\mu)$ and let $\mathbb{K}: E^2 \longrightarrow \mathbb{C}$ be the function given by $\mathbb{K}(x,y) = \sum_{k=1}^n \varphi_k(x)\overline{\varphi_k}(y)\ \forall (x,y) \in E^2$, then consider any random vector $(X_1,\ldots,X_n)$ taking values in E^n that can be sampled (from X_n to X_1) following the previous algorithm. In this situation, we have that the point process $\mathscr{X}$ that these random variables define, that is, $\mathscr{X} = \sum_{i=1}^n \delta_{X_i}$, is a determinantal point process in E with kernel $\mathbb{K}$.*

Proof. Indeed, using the notations introduced by the statement, we will start by showing that our random vector $(X_1,\ldots,X_n)$ is absolutely continuous with respect to our Radon measure μ^n in E^n and that its joint density function $f: E^n \longrightarrow [0,\infty)$ is given by:

$$f(x_1,\ldots,x_n) = \prod_{i=1}^n \frac{\|\mathscr{K}_{H_i}\delta_{x_i}\|^2}{i} \quad \forall (x_1,\ldots,x_n) \in E^n$$

To do it, we proceed by induction from n to 1. Focusing first on our initial case, following our algorithm, we know that X_n is sampled from the probability measure $\frac{1}{n}\mu_H$, something that allows us to conclude that it is an absolutely continuous random variable with density $f_n: E \longrightarrow [0,\infty)$ given by $f_n(x) = \frac{1}{n}\|\mathscr{K}_H\delta_x\|^2$ for all $x \in E$. That completes the initial case.

From here, for $1 \leq k < n$, we now assume that the random vector $(X_{k+1},\ldots,X_n)$ is absolutely continuous with density function $f_{k+1}: E^{n-k} \longrightarrow [0,\infty)$ given by the corresponding expression and we have to show that the same is true for $(X_k,\ldots,X_n)$. Indeed, we can start by considering the function $f_k: E^{n-k+1} \longrightarrow [0,\infty)$ given by $f_k(x_k,\ldots,x_n) = \prod_{i=k}^n (\|\mathscr{K}_{H_i}\delta_{x_i}\|^2/i)\ \forall (x_k,\ldots,x_n) \in E^{n-k+1}$ and then, given any $A \in \mathcal{B}(E)$ and $B \in \mathcal{B}(E^{n-k})$ Borel sets of the corresponding spaces, we can observe that:

$$P((X_k,\ldots,X_n) \in A \times B) = P((X_k,(X_{k+1},\ldots,X_n)) \in A \times B) =$$
$$P(X_k \in A, (X_{k+1},\ldots,X_n) \in B) =$$
(by the continuous version of the law of total probabilities, since E is Polish) =
$$\int_{E^{n-k}} P(X_k \in A, (X_{k+1},\ldots,X_n) \in B \mid (X_{k+1},\ldots,X_n) = x) f_{k+1}(x) d\mu^{n-k}(x) =$$
$$\int_B P(X_k \in A \mid (X_{k+1},\ldots,X_n) = x) f_{k+1}(x) d\mu^{n-k}(x) =$$
(fixed the values of $X_{k+1},\ldots,X_n$, then X_k is sampled with a density) =
$$\int_B (\int_A \tfrac{1}{k}\|\mathscr{K}_{H_k}\delta_y\| d\mu(y)) f_{k+1}(x) d\mu^{n-k}(x) = \text{(by Tonelli's theorem)} =$$
$$\int_{A\times B} \tfrac{1}{k}\|\mathscr{K}_{H_k}\delta_{x_k}\| f_{k+1}(x) d\mu^{n-k+1}(x_k,\ldots,x_n) = \int_{A\times B} f_k(x) d\mu^{n-k+1}(x)$$

so, in particular, we have that $P((X_k,\ldots,X_n) \in A \times B) = \int_{A\times B} f_k(x) d\mu^{n-k+1}(x)$.

From here, given $D \subseteq E^{n-k+1}$ any open set in E^{n-k+1}, we can use the same arguments than in the proof of Theorem 2.13 and, using the corresponding covering lemma by rectangles, we can conclude that $D = \bigcup_{i=1}^\infty A_i \times B_i$ for some $A_i \in \mathcal{B}(E)$ and

$B_i \in \mathcal{B}(E^{n-k})$ Borel sets such that their products are pairwise disjoint between them. With that in mind, we can now use this expression to deduce that:

$$P((X_k,\dots,X_n)\in D) = P((X_k,(X_{k+1},\dots,X_n))\in \bigcup_{i=1}^{\infty} A_i\times B_i) =$$
$$P(\bigcup_{i=1}^{\infty}\{(X_k,(X_{k+1},\dots,X_n))\in A_i\times B_i\}) =$$
$$\sum_{i=1}^{\infty} P((X_k,(X_{k+1},\dots,X_n))\in A_i\times B_i) =$$
$$\sum_{i=1}^{\infty}\int_{A_i\times B_i} f_k(x)d\mu^{n-k+1}(x) = \int_D f_k(x)d\mu^{n-k+1}(x)$$

where we have used the monotonous convergence theorem in this last equality.

Finally, since the probability law of the random vector $(X_k,\dots,X_n)$ is a probability measure in E^{n-k+1} and E^{n-k+1} is a Polish space, we can apply Theorem 7.1.7 in [5] to conclude that it is a Radon measure. That, combined with the fact that Radon measures are fully described by how they act on open sets by definition, allows us to conclude that the law of $(X_k,\dots,X_n)$ coincides with the measure in E^{n-k+1} that is absolutely continuous with respect to μ^{n-k+1} and has density f_k.

So, at the end, we have proved that $P((X_k,\dots,X_n)\in C) = \int_C f_k(x)d\mu^{n-k+1}(x)$ for every Borel set $C\in\mathcal{B}(E^{n-k+1})$, something that shows that the random vector $(X_k,\dots,X_n)$ is absolutely continuous and has f_k as its density function.

As a consequence, we have managed to complete the induction step and thus we can conclude that $(X_1,\dots,X_n)$ is an absolutely continuous random vector with the desired density function.

Now, given any $(x_1,\dots,x_n)\in E^n$, we can define the functions $\psi_i := \mathscr{K}\delta_{x_i}$ for all $1\leq i\leq n$ and we proceed to see that $\mathscr{K}_{H_i}\psi_i = \mathscr{K}_{H_i}\delta_{x_i}$. To do it, given any $1\leq i\leq n$, we can start by noticing that $\mathbb{K}_{H_i}$ is Hermitian, because given any $x,y\in E$, we can observe that $\mathbb{K}_{H_i}(x,y) = \langle \mathbb{K}_{H_i}(\cdot,y),\mathbb{K}_{H_i}(\cdot,x)\rangle = \overline{\langle \mathbb{K}_{H_i}(\cdot,x),\mathbb{K}_{H_i}(\cdot,y)\rangle} = \overline{\mathbb{K}_{H_i}(y,x)}$. Then, using this fact we can conclude that for any $x\in E$ we have that:

$$\mathscr{K}_{H_i}\psi_i(x) = \int_E \mathbb{K}_{H_i}(x,y)\psi_i(y)d\mu(y) = \langle\psi_i,\mathbb{K}_{H_i}(\cdot,x)\rangle = \langle\mathscr{K}_H\delta_{x_i},\mathbb{K}_{H_i}(\cdot,x)\rangle =$$
$$\langle\mathbb{K}(\cdot,x_i),\mathbb{K}_{H_i}(\cdot,x)\rangle = \overline{\langle\mathbb{K}_{H_i}(\cdot,x),\mathbb{K}(\cdot,x_i)\rangle} = \overline{\mathbb{K}_{H_i}(x_i,x)} = \mathbb{K}_{H_i}(x,x_i) = \mathscr{K}_{H_i}\delta_{x_i}(x)$$

something that shows the desired equality of functions $\mathscr{K}_{H_i}\psi_i = \mathscr{K}_{H_i}\delta_{x_i}$.

So, combining all of the above, we can now conclude that the joint density of our random vector $(X_1,\dots,X_n)$, that we denote by $f: E^n\longrightarrow[0,\infty)$, is given by:

$$f(x_1,\dots,x_n) = \prod_{i=1}^{n}\frac{\|\mathscr{K}_{H_i}\psi_i\|^2}{i} = \frac{1}{n!}\prod_{i=1}^{n}\|\mathscr{K}_{H_i}\psi_i\|^2 \quad \forall (x_1,\dots,x_n)\in E^n$$

That being said, given any $(x_1,\dots,x_n)\in E^n$, recall that by definition, for all $i<n$, we have that $H_i = H_{i+1}\cap \mathrm{Span}(\{\mathscr{K}_{H_{i+1}}\delta_{x_{i+1}}\})^{\perp} = H_{i+1}\cap\mathrm{Span}(\{\mathscr{K}_{H_{i+1}}\psi_{i+1}\})^{\perp}$, something that combined with the fact that $\langle g,\mathscr{K}_{H_{i+1}}\psi_{i+1}\rangle = \langle g,\psi_{i+1}\rangle$ for all $g\in H_{i+1}$ (a very well-known property of projection maps), allows us to conclude that $H_i = H_{i+1}\cap\mathrm{Span}(\{\psi_{i+1}\})^{\perp}$. As a consequence, applying iteratively this expression and using the properties of the orthogonal subspace (the intersection of orthogonal subspaces is the orthogonal subspace of the sum), we can conclude that $H_i = H\cap\mathrm{Span}(\{\psi_j\}_{j=i+1}^{n})^{\perp}$.

In addition, given any $1 \leq i, j \leq n$, we can use the fact that $\{\varphi_k\}_{k=1}^n$ is an orthonormal basis of H to observe that:

$$\langle \psi_i, \psi_j \rangle = \langle \sum_{k=1}^n \langle \psi_i, \varphi_k \rangle \varphi_k, \psi_j \rangle = \sum_{k=1}^n \langle \psi_i, \varphi_k \rangle \langle \varphi_k, \psi_j \rangle = \sum_{k=1}^n \langle \psi_i, \varphi_k \rangle \overline{\langle \psi_j, \varphi_k \rangle}$$

So, if we denote by $G = (g_{i,j})_{i,j \leq n}$ the matrix that has as entries these inner products, that is, $g_{i,j} = \langle \psi_i, \psi_j \rangle$ for all $1 \leq i, j \leq n$, we have managed to see that $G = AA^*$, where $A = (\langle \psi_i, \varphi_j \rangle)_{i,j \leq n}$ is the matrix that has by rows the coordinates of ψ_i in $\{\varphi_k\}_{k=1}^n$.

With all of the above in mind, following the definition and results for the volume (with and without sign) of a set of vectors stated in Section 3 of [18], we can notice that:

$$\prod_{i=1}^n \|\mathscr{K}_{H_i} \psi_i\|^2 = (\prod_{i=1}^n \|\mathscr{K}_{H_i} \psi_i\|)^2 = \mathrm{Vol}^+(\psi_1, \ldots, \psi_n)^2 = |\mathrm{Vol}(\psi_1, \ldots, \psi_n)|^2 =$$
$$|\det(A)|^2 = \det(A)\overline{\det(A)} = \det(A)\det(A^*) = \det(AA^*) = \det(G)$$

that is what we wanted to see. That, combined with the fact that for all $1 \leq i, j \leq n$ we have $\langle \psi_i, \psi_j \rangle = \langle \mathscr{K}_H \delta_{x_i}, \mathscr{K}_H \delta_{x_j} \rangle = \langle \mathbb{K}(\cdot, x_i), \mathbb{K}(\cdot, x_j) \rangle = \mathbb{K}(x_j, x_i)$ and thus $\det(G) = \det(\mathbb{K}(x_i, x_j))_{i,j \leq n}$, allows us to conclude that:

$$f(x_1, \ldots, x_n) = \frac{1}{n!} \det(\mathbb{K}(x_i, x_j))_{i,j \leq n} \quad \forall (x_1, \ldots, x_n) \in E^n$$

Finally, once that we know that the density function of $(X_1, \ldots, X_n)$ is given by $f(x_1, \ldots, x_n) = \frac{1}{n!} \det(\mathbb{K}(x_i, x_j))_{i,j \leq n}$ for all $(x_1, \ldots, x_n) \in E^n$, we can apply exactly the same arguments we used in the proof of Proposition 4.2 to conclude that $\mathscr{X} = \sum_{i=1}^n \delta_{X_i}$ is a determinantal point process in E with kernel $\mathbb{K}$, that is what we had to show. That completes the proof. □

As a comment, without entering into detail, it is worth mentioning that the above algorithm is one of the most efficient known if our aim is to sample general determinantal point processes. Nevertheless, for several individual determinantal point processes, as the ones that come from uniform spanning trees or from the eigenvalues of certain random matrices, there are specific algorithms that are faster and more efficient than the previous one.

Not only that, but also in practical applications, for example in the field of machine learning, it is usually enough to restrict ourselves to the finite case, where we assume that E is a finite set of cardinal $N \geq 1 \in \mathbb{N}$ considered with the counting measure and the usual topology. In this situation, one can propose other algorithms for sampling determinantal point processes in E, that for example, may not require to explicitly compute the eigenvalues $\{\lambda_k\}_k$ and eigenvectors $\{\varphi_k\}_k$ of $\mathscr{K}$. These algorithms are also generally faster than the one introduced above, mainly because they are based in usual matrix factorization algorithms and thus they can benefit from the many optimization and high speed techniques originally developed for the corresponding factorizations.

6. Selected properties of DPPs

From here, once that we have introduced the notion of determinantal point processes and studied the basic questions that arise from their definition, now we will focus on analysing some selected and impactful properties that the point processes of this family verify.

6.1. **Repulsive behaviour.**

As mentioned earlier, one of the most distinctive features of determinantal point processes is their repulsive nature, that lowers the probability of finding two points of the process close together. With the aim of studying this property and comparing it to the repulsion that a usual Poisson process exhibits, we will start by introducing the notion of the pair correlation function of a point process. It is worth mentioning that to develop this Subsection we have followed no particular reference, although this repulsive behaviour of DPPs is a very well-studied property that can be found in many sources.

Definition 6.1 (Pair correlation function)**.** *Let $\mathscr{X}$ be a simple point process in E whose first and second order joint intensities ρ_1 and ρ_2 exist. Then, we define its pair correlation function $g : E^2 \longrightarrow \mathbb{R}$ as the one given by:*

$$g(x,y) = \begin{cases} \frac{\rho_2(x,y)}{\rho_1(x)\rho_1(y)} & \text{if } \rho_1(x) > 0 \text{ and } \rho_1(y) > 0 \\ 0 & \text{otherwise} \end{cases}$$

for all $x, y \in E^2$.

In other words and intuitively, recall that the joint intensities of a point process can be understood as a measure of "how likely it is that certain points belong to the process" and, in our case, this would mean that $\rho_1(x)$ is the likelihood of x belonging to $\mathscr{X}$ and $\rho_2(x,y)$ is the likelihood of x and y belonging simultaneously to it. So, in this sense, the objective of this function is to quantify how the likelihood of finding the points together changes when considering their interaction. As a consequence, if $g(x,y) < 1$, that tells us that introducing one of the points reduces the likelihood of finding the other and thus the process is repulsive.

From here, using now the expression of the joint intensities of a determinantal point process, notice that it is immediate to express the above correlation function in terms of the kernel $\mathbb{K}$ of the process and to justify its repulsiveness.

Proposition 6.2. *Let $\mathscr{X}$ be a determinantal point process in E with kernel $\mathbb{K}$. Then, we have that its pair correlation function verifies that:*

$$g(x,y) = g(y,x) = 1 - \frac{|\mathbb{K}(x,y)|^2}{\mathbb{K}(x,x)\mathbb{K}(y,y)} \leq 1$$

μ^2-a.e for $(x,y) \in E^2$ such that $\mathbb{K}(x,x) > 0$ and $\mathbb{K}(y,y) > 0$. In addition, if we assume that $\mathbb{K}$ is continuous and redefine g in a set of measure 0, we have that $\lim_{x \to y} g(x,y) = 0$ for every $y \in E$.

Proof. Indeed, to show this result and continuing with the above notations, we only need to substitute the joint intensities of our determinantal point process in the definition of the pair correlation function. To do it, given almost any $(x, y) \in E^2$ such that $\mathbb{K}(x, x) > 0$ and $\mathbb{K}(y, y) > 0$, we can observe that:

$$g(x,y) = \frac{\rho_2(x,y)}{\rho_1(x)\rho_1(y)} = \frac{\mathbb{K}(x,x)\mathbb{K}(y,y) - |\mathbb{K}(x,y)|^2}{\mathbb{K}(x,x)\mathbb{K}(y,y)} = 1 - \frac{|\mathbb{K}(x,y)|^2}{\mathbb{K}(x,x)\mathbb{K}(y,y)} = g(y,x)$$

that is what we wanted to see, where we have used that the kernel $\mathbb{K}$ is Hermitian.

Not only that, but assuming also now that $\mathbb{K}$ is a continuous function and redefining g to make the above equalities hold for all $(x, y) \in E^2$ such that $\mathbb{K}(x, x) > 0$ and $\mathbb{K}(y, y) > 0$, it is clear that given any $y \in E$ we have that $\lim_{x\to y} \mathbb{K}(x, y) = \mathbb{K}(y, y)$ and $\lim_{x\to y} \mathbb{K}(x, x) = \mathbb{K}(y, y)$. As a consequence, we can conclude that:

$$\lim_{x\to y} g(x,y) = \lim_{x\to y}(1 - \frac{|\mathbb{K}(x,y)|^2}{\mathbb{K}(x,x)\mathbb{K}(y,y)}) = 1 - \frac{\mathbb{K}(y,y)^2}{\mathbb{K}(y,y)^2} = 0$$

where we have also considered that $g(x, y) = 0$ if $\mathbb{K}(x, x) = 0$ or $\mathbb{K}(y, y) = 0$. □

On the other hand, we can also compare this behaviour to what happens in a Poisson process, where the points are independently placed in the space. Indeed, here, given almost any $x, y \in E$ such that $\rho_1(x) > 0$ and $\rho_1(y) > 0$, it is easy to see that:

$$g(x,y) = \frac{\rho_2(x,y)}{\rho_1(x)\rho_1(y)} = \frac{\rho_1(x)\rho_1(y)}{\rho_1(x)\rho_1(y)} = 1$$

where we have used the definition of the joint intensities and the independence property of the Poisson process. That reinforces the idea that DPPs can be used to describe repulsive behaviours.

As a comment, it is worth highlighting that this repulsiveness is very useful in applications and it explains why DPPs are one of the main choices to model repulsive particles or to select diverse outputs. We will further develop these ideas in Section 8.

6.2. **Determinantal point processes of order n.** That being said, now we will develop another remarkable property of DPPs with our own ideas, that describes how their composition behaves. Indeed, under certain conditions, it is possible to observe that the composition of two DPPs is, in general, not a DPP. Nevertheless, it still has an interesting structure as a point process, with its joint intensities being given by determinants of products of kernels. We precise this fact in the following statement.

Theorem 6.3. *Let $\mathscr{X}$ be a determinantal point process in E with a locally bounded kernel $\mathbb{K}$ and let $\mathbb{H} : E^2 \longrightarrow \mathbb{C}$ be a continuous kernel such that its restrictions give rise to determinantal point processes in all countable subsets of E equipped with the counting measure. Then, consider $\mathscr{Y}$ the point process obtained by sampling first the process $\mathscr{X}$ and then sampling the DPP induced by the restriction of $\mathbb{H}$ in the set of points of the first. In this situation, we have that $\mathscr{Y}$ is a point process in E with joint intensities given by:*

$$\rho_k(x_1, \dots, x_k) = \det((\mathbb{H}(x_i, x_j))_{i,j\leq k}(\mathbb{K}(x_i, x_j))_{i,j\leq k})$$

for all $k \geq 1$ and μ^k-a.e. for $(x_1, \dots, x_k) \in E^k$.

Proof. Indeed, to show this result, continuing with the notations introduced above, first of all it is clear that $\mathscr{Y}$ is a well-defined point process thanks to our hypothesis. So with that in mind, now we proceed to compute its joint intensities. To do it, given any $x_1, \dots, x_k \in E$ pairwise distinct, then except maybe in a set of μ^k-measure zero in E^k, we have that:

$$\rho_k(x_1, \dots, x_k) = \lim_{\epsilon \to 0^+} \frac{P(\mathscr{Y}(B_\epsilon(x_i))=1 \text{ for each } 1 \leq i \leq k)}{\prod_{i=1}^k \mu(B_\epsilon(x_i))} =$$
$$\text{(applying the Law of total probability)} =$$
$$\lim_{\epsilon \to 0^+} \frac{P(\mathscr{Y}(B_\epsilon(x_i))=1 \text{ for each } 1 \leq i \leq k \mid \mathscr{X}(B_\epsilon(x_i))=1 \text{ for each } 1 \leq i \leq k) P(\mathscr{X}(B_\epsilon(x_i))=1 \text{ for each } 1 \leq i \leq k)}{\prod_{i=1}^k \mu(B_\epsilon(x_i))}$$
$$+ \lim_{\epsilon \to 0^+} \frac{P(\mathscr{Y}(B_\epsilon(x_i))=1 \text{ for each } 1 \leq i \leq k \mid A^\epsilon) P(A^\epsilon)}{\prod_{i=1}^k \mu(B_\epsilon(x_i))}$$

where $A^\epsilon := \{\omega \in \Omega \mid \mathscr{X}(\omega)(B_\epsilon(x_i)) \geq 1 \text{ for each } 1 \leq i \leq k,\ \mathscr{X}(\omega)(B_\epsilon(x_j)) \geq 2 \text{ for some } 1 \leq j \leq k\}$. From here, defining the events $A_j^\epsilon := \{\omega \in \Omega \mid \mathscr{X}(\omega)(B_\epsilon(x_i)) \geq 1 \text{ for each } 1 \leq i \neq j \leq k,\ \mathscr{X}(\omega)(B_\epsilon(x_j)) \geq 2\}$ for every $1 \leq j \leq k$, we can observe that $A^\epsilon = \bigcup_{j=1}^k A_j^\epsilon$ and thus we conclude that:

$$C := \lim_{\epsilon \to 0^+} \frac{P(\mathscr{Y}(B_\epsilon(x_i))=1 \text{ for each } 1 \leq i \leq k \mid A^\epsilon) P(A^\epsilon)}{\prod_{i=1}^k \mu(B_\epsilon(x_i))} \leq$$
$$\lim_{\epsilon \to 0^+} \sum_{j=1}^k \frac{P(\mathscr{Y}(B_\epsilon(x_i))=1 \text{ for each } 1 \leq i \leq k \mid A_j^\epsilon) P(A_j^\epsilon)}{\prod_{i=1}^k \mu(B_\epsilon(x_i))} =$$
$$\sum_{j=1}^k \lim_{\epsilon \to 0^+} \frac{P(\mathscr{Y}(B_\epsilon(x_i))=1 \text{ for each } 1 \leq i \leq k \mid A_j^\epsilon) P(A_j^\epsilon)}{\prod_{i=1}^k \mu(B_\epsilon(x_i))}$$

From here, we aim to see that all these limits are 0 (and in particular $C = 0$). Indeed, given any $j \in \{1, \dots, k\}$, we can observe that:

$$0 \leq \lim_{\epsilon \to 0^+} \frac{P(\mathscr{Y}(B_\epsilon(x_i))=1 \text{ for each } 1 \leq i \leq k \mid A_j^\epsilon) P(A_j^\epsilon)}{\prod_{i=1}^k \mu(B_\epsilon(x_i))} \leq \lim_{\epsilon \to 0^+} \frac{P(A_j^\epsilon)}{\prod_{i=1}^k \mu(B_\epsilon(x_i))} \leq$$
$$\lim_{\epsilon \to 0^+} \frac{\mathbb{E}(\prod_{i=1}^k \mathscr{X}(B_\epsilon(x_i)) \mid A_j^\epsilon) P(A_j^\epsilon)}{\prod_{i=1}^k \mu(B_\epsilon(x_i))} \leq 0$$

noticing that this last inequality has already been shown during the proof of Theorem 2.17 (here we have used that $\mathbb{K}$ is locally bounded). So, now we can conclude that:

$$\rho_k(x_1, \dots, x_k) =$$
$$\lim_{\epsilon \to 0^+} \frac{P(\mathscr{Y}(B_\epsilon(x_i))=1 \text{ for each } 1 \leq i \leq k \mid \mathscr{X}(B_\epsilon(x_i))=1 \text{ for each } 1 \leq i \leq k) P(\mathscr{X}(B_\epsilon(x_i))=1 \text{ for each } 1 \leq i \leq k)}{\prod_{i=1}^k \mu(B_\epsilon(x_i))}$$
$$= \det(\mathbb{K}(x_i, x_j))_{i,j \leq k} \times$$
$$\lim_{\epsilon \to 0^+} P(\mathscr{Y}(B_\epsilon(x_i)) = 1 \text{ for each } 1 \leq i \leq k \mid \mathscr{X}(B_\epsilon(x_i)) = 1 \text{ for each } 1 \leq i \leq k)$$

and we focus on this last limit. Indeed, conditioning to the fact that $\mathscr{X}(B_\epsilon(x_i)) = 1$ for each $1 \leq i \leq k$, we can denote by $p_1^\epsilon, \dots, p_k^\epsilon$ these unique points of $\mathscr{X}$ that belong to $B_\epsilon(x_1), \dots, B_\epsilon(x_k)$ respectively. Then, recalling the definition of $\mathscr{Y}$, we know that this point process is based on the determinantal point process induced by the kernel $\mathbb{H}$ in a certain countable set Λ such that $\{p_1^\epsilon, \dots, p_k^\epsilon\} \subseteq \Lambda$. That implies that, by the properties of DPPs in discrete settings, we have that:

$$\lim_{\epsilon\to 0^+} P(\mathscr{Y}(B_\epsilon(x_i)) = 1 \text{ for each } 1 \leq i \leq k \mid \mathscr{X}(B_\epsilon(x_i)) = 1 \text{ for each } 1 \leq i \leq k)$$
$$= \lim_{\epsilon\to 0^+} \det(\mathbb{H}(p_i^\epsilon, p_j^\epsilon))_{i,j\leq k} = \det(\mathbb{H}(x_i, x_j))_{i,j\leq k}$$

where we have used the continuity of the kernel $\mathbb{H}$ in this last step. Then, combining all of the above, we have managed to see that:

$$\rho_k(x_1, \ldots, x_k) = \det(\mathbb{H}(x_i, x_j))_{i,j\leq k} \det(\mathbb{K}(x_i, x_j))_{i,j\leq k} =$$
$$\det((\mathbb{H}(x_i, x_j))_{i,j\leq k}(\mathbb{K}(x_i, x_j))_{i,j\leq k})$$

that is what we wanted to deduce, something that completes the proof. □

From here, with the above result in mind and aiming to make sense of this composition of DPPs, we can introduce the notion of the determinantal point process of order n, for any $n \geq 1$. This generalisation of the usual DPP can be defined as follows.

Definition 6.4. *Let $n \geq 1$, let $\mathscr{X}$ be a simple point process in E and let $\mathbb{K}_1, \ldots, \mathbb{K}_n$ be complex-valued functions in E^2. Then we say that $\mathscr{X}$ is a determinantal point process of order n with kernels $\mathbb{K}_1, \ldots, \mathbb{K}_n$ if its joint intensities w.r.t μ exist and are given by:*

$$\rho_k(x_1, \ldots, x_k) = \det((\mathbb{K}_1(x_i, x_j)_{i,j\leq k}) \cdots (\mathbb{K}_n(x_i, x_j)_{i,j\leq k}))$$

for all $k \geq 1$ and μ^k-a.e. for $(x_1, \ldots, x_k) \in E^k$.

Notice that, using similar arguments than the ones exposed in Theorem 6.3 and under certain hypothesis, it may be the case that the composition of a DPP of order n with a usual DPP yields a DPP of order $n + 1$ (and, in particular, the composition of two DPPs produces a DPP of order 2, as shown in the theorem).

In addition, regarding the hypothesis of the theorem, observe that they may seem limiting, especially the condition on $\mathbb{H}$ that imposes that its restrictions should give rise to determinantal point processes in all countable subsets of E equipped with the counting measure. Nevertheless, a particular case where these hypothesis are easily satisfied is in the discrete one, as shown in the following proposition.

Proposition 6.5. *Let Λ be a countable set equipped with the counting measure μ and let $\mathbb{K} : \Lambda^2 \longrightarrow \mathbb{C}$ be a kernel that gives rise to a determinantal point process in Λ. Then we have that its restrictions give rise to a determinantal point process in all countable subsets of Λ equipped, again, with the counting measure.*

Proof. To show this result, we only need to notice that it follows directly from Lemma 4.1. Indeed, given any $D \subseteq \Lambda$, we can start by observing that it is a Borel set $D \in \mathcal{B}(\Lambda)$, as mentioned in Example 2.4. Then, applying the corresponding lemma with $E = \Lambda$, we can conclude that the restriction of $\mathscr{X}$ to D, $\mathscr{X} \cap D$, is a determinantal point process with kernel $\mathbb{K}_{|D^2}$. In particular, the restriction of $\mathbb{K}$ to D^2 gives rise to a determinantal point process in D, that is exactly what we wanted to see. □

From here, in the context of the above Theorem 6.3, we can also observe that the following result about determinantal point processes holds, which essentially tells us that, in the discrete case, the complementary of one of these processes is also determinantal.

Proposition 6.6. *Let Λ be a countable set equipped with the counting measure μ and let $\mathscr{X}$ be a determinantal point process in Λ with kernel $\mathbb{K}$. Then, if we consider $\mathscr{Y}$ the point process defined by taking the points that do not belong to $\mathscr{X}$, we have that $\mathscr{Y}$ is also a determinantal point process with kernel $\delta - \mathbb{K}$, where $\delta : \Lambda^2 \longrightarrow \mathbb{C}$ is given by $\delta(x,y) = 1$ if $x = y$ and $\delta(x,y) = 0$ otherwise, for all $x, y \in \Lambda$.*

Proof. Indeed, continuing with the notation introduced above, we can start by observing that, since $\mathbb{K}$ is locally square integrable and Hermitian and $\mathscr{K}$ is locally of trace class, it is also immediate to see that the same is true for $\delta - \mathbb{K}$. In addition, denoting by $\mathscr{K}'$ the integral operator associated to $\delta - \mathbb{K}$, regarding the non-negativity, we can conclude that given any function $f : \Lambda \longrightarrow \mathbb{C}$ with compact support we have:

$$\langle \mathscr{K}' f, f\rangle = \langle \sum_{y\in\Lambda}(\delta(\cdot,y) - \mathbb{K}(\cdot,y))f(y), f\rangle = \sum_{x,y\in\Lambda}(\delta(x,y) - \mathbb{K}(x,y))f(y)f(x)$$
$$= \sum_{x,y\in\Lambda}\delta(x,y)f(y)f(x) - \sum_{x,y\in\Lambda}\mathbb{K}(x,y)f(y)f(x) = \sum_{x\in\Lambda} f(x)^2 - \langle \mathscr{K} f, f\rangle \geq$$
$$\|f\|^2 - \|\mathscr{K} f\|\|f\| \geq \|f\|^2 - \|\mathscr{K}\|\|f\|^2 \geq 0$$

because $\|\mathscr{K}\| \leq 1$ by Theorem 4.4 and where the above sums are finite thanks to the compactness of the support of f. That shows that it makes sense to consider a determinantal point process in Λ with kernel $\delta - \mathbb{K}$, according to our definition.

From here, to show that $\mathscr{Y}$ is actually determinantal with kernel $\delta - \mathbb{K}$, we proceed to compute its joint intensities. Given any $x_1, \ldots, x_n \in \Lambda$ pairwise distinct, using the description of the joint intensities seen in Proposition 3.4, we can deduce that:

$$\rho_n(x_1, \ldots, x_n) = P(\{x_1, \ldots, x_n\} \subseteq \mathscr{Y}) = P(x_1 \notin \mathscr{X}, \ldots, x_n \notin \mathscr{X}) =$$
$$1 - P(\cup_{k=1}^n \{x_i \in \mathscr{X}\}) =$$
$$\text{(applying the usual formula for the probability of the union of events)} =$$
$$1 - \sum_{k=1}^n (-1)^{k-1} \sum_{1\leq i_1<\cdots<i_k\leq n} P(\{x_{i_1}, \ldots, x_{i_k}\} \subseteq \mathscr{X}) =$$
$$1 + \sum_{k=1}^n (-1)^{k} \sum_{1\leq i_1<\cdots<i_k\leq n} \det(\mathbb{K}(x_{i_a}, x_{i_b}))_{a,b\leq k}$$
$$= \sum_{k=0}^n \sum_{1\leq i_1<\cdots<i_k\leq n} \det(-\mathbb{K}(x_{i_a}, x_{i_b}))_{a,b\leq k} = \det(\delta(x_i, x_j) - \mathbb{K}(x_i, x_j))_{i,j\leq n}$$

where in this last equality we have applied a purely combinatorial result, that can be found stated in Corollary 6.164 of [9]. That ends the proof. □

In our case, assuming the hypothesis and notations of Theorem 6.3, recall that we know that the restrictions of $\mathbb{H}$ define DPPs in all countable subsets of E. In particular and taking the set of points of $\mathscr{X}$ as our countable set $\Lambda \subseteq E$, the above proposition tells us that the complementary of this DPP induced by $\mathbb{H}$ in Λ is also a DPP. So, in conclusion, we find ourselves in the following situation. We have a set of random points in E that form a DPP and, in addition, these points are divided into two subsets that, by themselves, are also DPPs of order 2.

These DPPs could be especially relevant in practical applications related to machine learning, a context where the hypothesis of Theorem 6.3 are very frequently satisfied. For example, they could be used to model two groups of particles that coexist, with

repulsion occurring between all particles, but also exhibiting different behaviours within each group.

As a comment, notice that it is possible that these last constructions could be also iterated further, involving DPPs of higher order.

6.3. **Central Limit Theorem.** From here, another interesting property that motivates the consideration of determinantal point processes is that, for them, a certain version of the Central Limit Theorem holds. So, given a family of determinantal point processes, under the proper assumptions, we can observe that any function of them normalized converges in distribution to a standard normal random variable.

Theorem 6.7 (Central Limit Theorem for linear statistics of DPPs)**.** *Let $\{\mathscr{X}_l\}_{l\in\mathbb{N}}$ be a family of determinantal point process in E with respective kernels $\{\mathbb{K}_l\}_{l\in\mathbb{N}}$ and let $\{f_l\}_{l\in\mathbb{N}}$ be also a family of real-valued, bounded and measurable functions defined in E whose support has a compact closure. Then, for each value of $l \geq 0$, we define the random variable $S_{f_l} \coloneqq \sum_{x\in\mathscr{X}_l} f_l(x)$ understanding $\mathscr{X}_l$ as a random discrete subset of E and assume that $\lim_l \mathrm{Var}(S_{f_l}) = \infty$, $\sup_{x\in E}|f_l(x)| = o(\mathrm{Var}(\mathrm{S_{f_l}})^{\epsilon})$ and $\mathbb{E}(S_{|f_l|}) = O((\mathrm{Var}(S_{f_l}))^{\delta})$ for all $\epsilon > 0$ and for some $\delta > 0$. In this situation, we can conclude that:*

$$\frac{S_{f_l} - \mathbb{E}(S_{f_l})}{\sqrt{\mathrm{Var}(S_{f_l})}} \xrightarrow{\mathcal{L}} N(0,1) \quad \textit{when } l \longrightarrow \infty$$

Proof. Indeed, to show this result that can be found proven as Theorem 1 of [22], we can start by noticing that given any $l \geq 0$, it makes sense to consider the random variable S_{f_l} because it is finite almost surely. To do it, observe that we know that f_l is bounded and that $\mathrm{supp} f_l$ is inside a compact set K. Then, we can consider the restriction of the corresponding determinantal point process $\mathscr{X}_l$ to K and, as $\mathscr{H}_l$ is locally of trace class by definition, we can deduce that the restricted kernel $\mathbb{K}_{|K^2}$ is under the hypothesis of Theorem 4.3, something that allows us to conclude (as we explained in the proof of the theorem) that the number of points of $\mathscr{X}_l \cap K$ is finite μ-a.s. That, combined with the fact that f_l is bounded, clearly implies that $|S_{f_l}| \leq \sum_{x\in\mathscr{X}_l} |f_l(x)| < \infty$ is finite almost surely and that all of its moments exist.

That being said, given any $l \geq 0$, now we proceed to study the cumulants of the random variable S_{f_l}, that we will denote by S_f without making explicit the index l to ease the notation. In fact, what we aim to show is the following lemma.

Lemma 6.8. *Let S_f be the random variable considered above and denote by $C_n(S_f)$ its n-th cumulant for every $n \geq 1$. Then, for all $n \geq 1$ we have that:*

$$C_n(S_f) = \sum_{m=1}^{n} \sum_{\substack{n_1,\dots,n_m\geq 1 \\ n_1+\cdots+n_m=n}} \frac{(-1)^{m-1}}{m} \frac{n!}{n_1!\dots n_m!} \times$$

$$\int_{E^m} f(x_1)^{n_1}\mathbb{K}(x_1,x_2)\dots f(x_m)^{n_m}\mathbb{K}(x_m,x_1)d\mu(x_1)\dots d\mu(x_m)$$

Proof. Following the proof of Lemma 1 from [21] and filling in the details, first of all, recall that the cumulants of the random variable S_f are defined through the Taylor coefficients of the logarithm of the characteristic function of S_f. In detail, that means that they are given by the equality $\log(\mathbb{E}(e^{itS_f})) = \sum_{n=1}^{\infty} C_n(S_f)\frac{(it)^n}{n!}$ that is satisfied for all $t \in \mathbb{R}$. With that in mind, to show the desired expression for them, we will start by studying the moments of S_f.

Indeed, let any $N \geq 1 \in \mathbb{N}$, we can consider the N-th moment of S_f, $\mathbb{E}((S_f)^N)$, and then, if we denote by $x_1, \dots, x_h$ the set of (random) points where $\mathscr{X}$ has mass that are inside the closure of supp$f$ (notice that h is also a random variable finite almost surely), we can deduce that:

$$\mathbb{E}((S_f)^N) = \mathbb{E}(S_f \cdots S_f) = \mathbb{E}((\sum_{n_1=1}^{h} f(x_{n_1})) \cdot (\sum_{n_N=1}^{h} f(x_{n_N}))) =$$
$$\mathbb{E}(\prod_{i=1}^{N}(\sum_{n_i=1}^{h} f(x_{n_i}))) = (\text{exchanging the order of the product and the sum }) =$$
$$\mathbb{E}(\sum_{n_1,\dots,n_N} \prod_{i=1}^{N} f(x_{n_i}))$$

From here, fixed a term of the previous sum, then we can group together the indices n_i that coincide in this term of the sum. This produces a partition $\mathcal{M} = \{M_1, \dots, M_r\}$ of the set $\{1, \dots, N\}$ in the following way: if $i, j \in M_k$ for any $1 \leq k \leq r$, then $n_i = n_j$ in the fixed term of the sum. Denote now $s_k = |M_k|$ for every $1 \leq k \leq r$ and then we can observe that $\mathcal{M}$ is a partition of $\{1, \dots, N\}$, a set that is independent of the fixed term of the sum. That allows us to express the previous sum in terms of the partitions, first adding over the partitions of $\{1, \dots, N\}$ and then over the possible indices that are representatives of the elements of the partition (as a small subtlety here and later on, notice given a partition $\mathcal{M}$ we need a consistent way of labelling its elements $\mathcal{M} = \{M_1, \dots, M_r\}$, maybe by enforcing $\min M_i \leq \min M_j$ if $i \leq j$). So, at the end, let Part_N be the set of partitions of $\{1, \dots, N\}$, we conclude that:

$$\mathbb{E}((S_f)^N) = \mathbb{E}(\sum_{\mathcal{M}\in \text{Part}_N} \sum_{n_1 \neq \dots \neq n_r} \prod_{i=1}^{r} f(x_{n_i})^{s_i}) =$$
$$\sum_{\mathcal{M}\in \text{Part}_N} \mathbb{E}(\sum_{n_1 \neq \dots \neq n_r} \prod_{i=1}^{r} f(x_{n_i})^{s_i})$$

Now, fixed the partition $\mathcal{M} \in \text{Part}_N$, we can consider the corresponding term of the sum and then show the following property.

Lemma 6.9. *In the situation described above, we have that:*

$$\mathbb{E}(\sum_{n_1 \neq \dots \neq n_r} \prod_{i=1}^{r} f(x_{n_i})^{s_i}) = \int_{E^r} f^{s_1}(y_1) \dots f^{s_r}(y_r) \det(\mathbb{K}(y_i, y_j))_{i,j\leq r} d\mu(y_1) \dots d\mu(y_r)$$

Proof. Indeed, to show this result, we can start by considering the function g defined in E^r given by $g(y_1, \dots, y_r) = \prod_{i=1}^{r} f(y_i)^{s_i}$ for all $(y_1, \dots, y_r) \in E^r$ and it is clear that it has support inside a compact set and that it is a measurable and bounded function because the same is true for f. With that in mind, then it suffices to show that:

$$\mathbb{E}(\sum_{n_1 \neq \dots \neq n_r} g(x_{n_1}, \dots, x_{n_r})) = \int_{E^r} g(y_1, \dots, y_n) \det(\mathbb{K}(y_i, y_j))_{i,j\leq r} d\mu(y_1) \dots d\mu(y_r)$$

To do it, let $D \subseteq E^r$ be any open set of E^r such that the corresponding characteristic function $\mathcal{X}_D$ has support inside a compact set. Then, if we consider the previous equality for this function we have that:

$$\mathbb{E}(\sum_{n_1 \neq \ldots \neq n_r} \mathcal{X}_D(x_{n_1}, \ldots, x_{n_r})) =$$
$$\int_{E^r} \mathcal{X}_D(y_1, \ldots, y_n) \det(\mathbb{K}(y_i, y_j))_{i,j \leq r} d\mu(y_1) \ldots d\mu(y_r) \iff$$
$$\mathbb{E}(\sum_{n_1 \neq \ldots \neq n_r} \mathcal{X}_D(x_{n_1}, \ldots, x_{n_r})) = \int_D \det(\mathbb{K}(y_i, y_j))_{i,j \leq r} d\mu(y_1) \ldots d\mu(y_r)$$

and this last equality we already know is true by applying Corollary 2.14.

From here, we can define the measures $\mu_1 : \mathcal{B}(E^r) \longrightarrow [0, \infty]$ and $\mu_2 : \mathcal{B}(E^r) \longrightarrow [0, \infty]$ given by:

$$\mu_1(B) = \int_{K^r} \mathcal{X}_B(y_1, \ldots, y_n) \det(\mathbb{K}(y_i, y_j))_{i,j \leq r} d\mu(y_1) \ldots d\mu(y_r)$$
$$\mu_2(B) = \mathbb{E}(\sum_{n_1 \neq \ldots \neq n_r} \mathcal{X}_B(x_{n_1}, \ldots, x_{n_r}))$$

respectively for all $B \subseteq E^r$ Borel set and where K denotes the compact inside which f has its support. Then, it is clear that both of them are actually measures (because they are σ- additive) and that they are finite, because for μ_2 we already know that the number of points of the DPP is finite in the compact K μ-a.s. On the other hand, by our preliminaries, we also know that the determinant of the kernel $\mathbb{K}$ restricted to a compact as K is also integrable.

Now, we can use the fact that E^r is a Polish space and applying Theorem 7.1.7 in [5] we can deduce that both μ_1 and μ_2 are Radon measures. That, combined with Corollary 2.14 and the fact that Radon measures are fully described by how they behave on open sets, allows us to conclude that $\mu_1 = \mu_2$. In other words, we have managed to see that for all Borel sets $B \subseteq K^r$, we have that:

$$\mathbb{E}(\sum_{n_1 \neq \ldots \neq n_r} \mathcal{X}_B(x_{n_1}, \ldots, x_{n_r})) = \int_{E^r} \mathcal{X}_B(y_1, \ldots, y_n) \det(\mathbb{K}(y_i, y_j))_{i,j \leq r} d\mu(y_1) \ldots d\mu(y_r)$$

Not only that, but it is also clear that if the previous equality is true for some set of bounded and measurable functions with compact support, then it is also true for any linear combination of them, because the sums, the integral and the expected value are linear. That, combined with all of the above, allows us to conclude that the equality is true for simple functions with support inside K^r.

Finally, knowing that any non-negative bounded and measurable function can be approximated by an increasing sequence of simple functions, allows us to apply the monotonous convergence theorem to conclude that the equality is true for any of these functions with support inside K^r. From here, we can use the usual arguments to extend the result to any measurable bounded and complex-valued function with support in K^r and thus we have managed to show the result:

$$\mathbb{E}(\sum_{n_1 \neq \ldots \neq n_r} g(x_{n_1}, \ldots, x_{n_r})) = \int_{E^r} g(y_1, \ldots, y_n) \det(\mathbb{K}(y_i, y_j))_{i,j \leq r} d\mu(y_1) \ldots d\mu(y_r)$$

that is what we wanted to see. That completes the proof of this lemma. □

That being done, substituting this expression in the sum we were working with, now we have that:

$$\mathbb{E}((S_f)^N) = \sum_{\mathcal{M}\in \mathrm{Part}_N} \int_{E^r} f^{s_1}(y_1)\dots f^{s_r}(y_r)\det(\mathbb{K}(y_i,y_j))_{i,j\leq r} d\mu(y_1)\dots d\mu(y_r)$$

and again we proceed to develop one of the terms of this sum. To do it, focusing on the determinant, it is a well-known fact that:

$$\det(\mathbb{K}(y_i,y_j))_{i,j\leq r} = \sum_{\sigma\in S_r} \mathrm{sgn}(\sigma)\prod_{i=1}^{r}\mathbb{K}(y_i,y_{\sigma(i)})$$

and, from here, given any permutation $\sigma\in S_r$, notice that we can write it as a composition of disjoint cycles $\sigma = \tau_1\circ\dots\circ\tau_q$. Not only that, but then by the properties of the sign of a permutation, we have that $\mathrm{sgn}(\sigma) = \mathrm{sgn}(\tau_1\circ\dots\circ\tau_q) = \prod_{\alpha=1}^{q}(-1)^{p_\alpha - 1}$ where p_α denotes the length of τ_α for all $\alpha\in\{1,\dots,q\}$. Using this fact to develop the expression of the determinant, we can observe that every permutation is uniquely determined (in a one-to-one sense) by these disjoint cycles and the partition of $\{1,\dots,r\}$ that they induce. So, applying this to rewrite the sum of the determinant, we have that:

$$\begin{gathered}
\det(\mathbb{K}(y_i,y_j))_{i,j\leq r} = \textstyle\sum_{\sigma\in S_r}\mathrm{sgn}(\sigma)\prod_{i=1}^{r}\mathbb{K}(y_i,y_{\sigma(i)}) = \\
\textstyle\sum_{\sigma\in S_r}(\prod_{\alpha=1}^{q}(-1)^{p_\alpha-1})(\prod_{\alpha=1}^{q}\prod_{j=1}^{p_\alpha}\mathbb{K}(y_{t_j^\alpha},y_{\tau_\alpha(t_j^\alpha)})) = \\
\textstyle\sum_{\sigma\in S_r}\prod_{\alpha=1}^{q}((-1)^{p_\alpha-1}\prod_{j=1}^{p_\alpha}\mathbb{K}(y_{t_j^\alpha},y_{\tau_\alpha(t_j^\alpha)})) = \\
\textstyle\sum_{\mathcal{K}\in\mathrm{Part}_r}\sum_{(\tau_1,\dots,\tau_q)\text{ cyclic permutations}}\prod_{\alpha=1}^{q}((-1)^{p_\alpha-1}\prod_{j=1}^{p_\alpha}\mathbb{K}(y_{t_j^\alpha},y_{\tau_\alpha(t_j^\alpha)})) = \\
\textstyle\sum_{\mathcal{K}\in\mathrm{Part}_r}\sum_{\tau_1\text{ cy. per. }K_1}\cdots\sum_{\tau_q\text{ cy. per. }K_q}\prod_{\alpha=1}^{q}((-1)^{p_\alpha-1}\prod_{j=1}^{p_\alpha}\mathbb{K}(y_{t_j^\alpha},y_{\tau_\alpha(t_j^\alpha)})) = \\
\textstyle\sum_{\mathcal{K}\in\mathrm{Part}_r}\prod_{\alpha=1}^{q}((-1)^{p_\alpha-1}\sum_{\sigma\text{ cy. per. }K_\alpha}\prod_{j=1}^{p_\alpha}\mathbb{K}(y_{t_j^\alpha},y_{\sigma(t_j^\alpha)}))
\end{gathered}$$

where $\mathcal{K} = \{K_1,\dots,K_q\}$ and $K_\alpha = \{t_1^\alpha,\dots,t_{p_\alpha}^\alpha\}$ for all $\alpha\in\{1,\dots,q\}$. Then, combining this with the previous sum we finally obtain that:

$$\begin{gathered}
\mathbb{E}((S_f)^N) = \textstyle\sum_{\mathcal{M}\in\mathrm{Part}_N}\int_{E^r} f^{s_1}(y_1)\dots f^{s_r}(y_r)\det(\mathbb{K}(y_i,y_j))_{i,j\leq r}d\mu(y_1)\dots d\mu(y_r) = \\
\textstyle\sum_{\mathcal{M}\in\mathrm{Part}_N}\sum_{\mathcal{K}\in\mathrm{Part}_r}\int_{E^r} f^{s_1}(y_1)\dots f^{s_r}(y_r)\times \\
\textstyle\prod_{\alpha=1}^{q}((-1)^{p_\alpha-1}\sum_{\sigma\text{ cy. per. }K_\alpha}\prod_{j=1}^{p_\alpha}\mathbb{K}(y_{t_j^\alpha},y_{\sigma(t_j^\alpha)}))d\mu(y_1)\dots d\mu(y_r) = \\
\textstyle\sum_{\mathcal{M}\in\mathrm{Part}_N}\sum_{\mathcal{K}\in\mathrm{Part}_r}\int_{E^{p_1}}\cdots\int_{E^{p_q}} f^{s_1}(y_1)\dots f^{s_r}(y_r)\times \\
\textstyle\prod_{\alpha=1}^{q}((-1)^{p_\alpha-1}\sum_{\sigma\text{ cy. per. }K_\alpha}\prod_{j=1}^{p_\alpha}\mathbb{K}(y_{t_j^\alpha},y_{\sigma(t_j^\alpha)}))d\mu(y_1)\dots d\mu(y_r) = \\
\textstyle\sum_{\mathcal{M}\in\mathrm{Part}_N}\sum_{\mathcal{K}\in\mathrm{Part}_r}\prod_{\alpha=1}^{q}((-1)^{p_\alpha-1}\int_{E^{p_\alpha}} f^{s_{t_1^\alpha}}(y_{t_1^\alpha})\dots f^{s_{t_{p_\alpha}^\alpha}}(y_{t_{p_\alpha}^\alpha})\times \\
\textstyle\sum_{\sigma\text{ cy. per. }K_\alpha}\prod_{j=1}^{p_\alpha}\mathbb{K}(y_{t_j^\alpha},y_{\sigma(t_j^\alpha)}))d\mu(y_{t_1^\alpha})\dots d\mu(y_{t_{p_\alpha}^\alpha})
\end{gathered}$$

where we have exchanged the order of the integrals and the product as we did before with the sum. From here, to continue developing the previous expression, now we aim to exchange the order of the first two sums. To do it, given $\mathcal{M} = \{M_1,\dots,M_r\}\in\mathrm{Part}_N$ and then a partition $\mathcal{K} = \{K_1,\dots,K_q\}\in\mathrm{Part}_r$, we can construct a new partition $\mathcal{P} = \{P_1,\dots,P_q\}\in\mathrm{Part}_N$ such that P_i is obtained by joining $P_i = \cup_{j\in K_i}M_j$ for all

$1 \leq i \leq q$. With that in mind, observe that this operation is actually a bijection, because given $\mathcal{P}$ and the partitions of its elements P_i we can recover the two original partitions $\mathcal{M}$ and $\mathcal{K}$ (here we use the consistent way of labelling the elements of $\mathcal{M}$). That allows us to re-index the sum in terms of these new partitions and conclude that:

$$\mathbb{E}((S_f)^N) = \sum_{\mathcal{P}\in \mathrm{Part}_N} \sum_{(\mathcal{P}_1,\dots,\mathcal{P}_q) \text{ part. of } (P_1,\dots,P_q)} \prod_{\alpha=1}^{q} ((-1)^{p_\alpha - 1} \int_{E^{p_\alpha}} f^{s_{t_1^\alpha}}(y_{t_1^\alpha}) \cdots \times$$
$$f^{s_{t_{p_\alpha}^\alpha}}(y_{t_{p_\alpha}^\alpha}) \sum_{\sigma \text{ cy. per. } K_\alpha} \prod_{j=1}^{p_\alpha} \mathbb{K}(y_{t_j^\alpha}, y_{\sigma(t_j^\alpha)})) d\mu(y_{t_1^\alpha}) \dots d\mu(y_{t_{p_\alpha}^\alpha}) =$$
$$\sum_{\mathcal{P}\in \mathrm{Part}_N} \prod_{i=1}^{q} (\sum_{\mathcal{P}_i = \{P_{i,1},\dots,P_{i,t_i}\} \text{ part. of } P_i} (-1)^{t_i - 1} \int_{E^{t_i}} f^{|P_{i,1}|}(y_1) \cdots \times$$
$$f^{|P_{i,t_i}|}(y_{t_i}) \sum_{\sigma \in S_{t_i} \text{ cyclic}} \prod_{j=1}^{t_i} \mathbb{K}(y_j, y_{\sigma(j)}) d\mu(y_1) \dots d\mu(y_{t_i}))$$

where we have used one more time the same technique to exchange the order of the sums and products and also changed the name of the variables in the integrals to make easier the notation.

That being done, recall that our actual objective was to show that the cumulants of S_f verify a certain expression. Indeed, to do it, we can use the following well-known result that relates the cumulants of a random variable and its moments and that can be found, for example, in page 227 of [6]:

$$\mathbb{E}((S_f)^N) = \sum_{\mathcal{P}=\{P_1,\dots,P_k\}\in \mathrm{Part}_N} C_{|P_1|}(S_f) \dots C_{|P_k|}(S_f)$$

With that in mind, comparing it to our own development of the moments of S_f and given any $n \geq 1$, we can take the factor in the previous expression corresponding to the cumulant and, adapting the indexes (notice that the previous $|P_i|$ now is n, the previous t_i is m and that the i subindexes are no longer required) and the names of the elements in the partitions (from P_i to R_i to avoid confusion), we conclude that:

$$C_n(S_f) = \sum_{\mathcal{P}=\{R_1,\dots,R_m\}\in \mathrm{Part}_n} (-1)^{m-1} \int_{E^m} f^{|R_1|}(y_1) \cdots \times$$
$$f^{|R_m|}(y_m) \sum_{\sigma \in S_m \text{ cyclic}} \prod_{j=1}^{m} \mathbb{K}(y_j, y_{\sigma(j)}) d\mu(y_1) \dots d\mu(y_m)$$

From here, notice that the specific partition $\mathcal{P} = \{R_1, \dots, R_m\} \in \mathrm{Part}_n$ in the previous formula does not matter, that is, what follows only depends on the sizes $|R_1|, \dots, |R_m|$. So, knowing that there exist $\frac{n!}{|R_1|!\dots|R_m|!} \frac{1}{m!}$ partitions $\mathcal{P} = \{R_1, \dots, R_m\} \in \mathrm{Part}_n$ with the same sizes (to deduce this number, consider all possible orderings of the n values, that are $n!$, and we need to divide for re-orderings inside the R_i, because they do not change the partition, and divide by $m!$ because the order of the $R_1, \dots, R_m$ also does not change $\mathcal{P}$), we can re-index the above sum:

$$C_n(S_f) = \sum_{m=1}^{n} \sum_{\substack{n_1,\dots,n_m \geq 1 \\ n_1+\dots+n_m=n}} (-1)^{m-1} \frac{n!}{n_1!\dots n_m!} \frac{1}{m!} \int_{E^m} f^{n_1}(y_1) \cdots \times$$
$$f^{n_m}(y_m) \sum_{\sigma \in S_m \text{ cyclic}} \prod_{j=1}^{m} \mathbb{K}(y_j, y_{\sigma(j)}) d\mu(y_1) \dots d\mu(y_m) =$$
$$\sum_{m=1}^{n} \sum_{\sigma \in S_m \text{ cyclic}} \sum_{\substack{n_1,\dots,n_m \geq 1 \\ n_1+\dots+n_m=n}} (-1)^{m-1} \frac{n!}{n_1!\dots n_m!} \frac{1}{m!} \int_{E^m} f^{n_1}(y_1) \cdots \times$$
$$f^{n_m}(y_m) \prod_{j=1}^{m} \mathbb{K}(y_j, y_{\sigma(j)}) d\mu(y_1) \dots d\mu(y_m)$$

Finally, applying a change of variable in the integral that reorders $y_1, \dots, y_n$ in an appropriate way, reordering also the factors and reindexing the inner sum if necessary,

we can observe that the particular cyclic permutation σ does not matter and, as there are $(m-1)!$ of them, we have that:

$$C_n(S_f) = \sum_{m=1}^{n}(m-1)!\sum_{\substack{n_1,\dots,n_m\geq 1\\ n_1+\cdots+n_m=n}}(-1)^{m-1}\frac{n!}{n_1!\dots n_m!}\frac{1}{m!}\int_{E^m} f^{n_1}(y_1)\cdots\times$$
$$f^{n_m}(y_m)\prod_{j=1}^{m}\mathbb{K}(y_j,y_{j+1})d\mu(y_1)\dots d\mu(y_m) =$$
$$\sum_{m=1}^{n}\sum_{\substack{n_1,\dots,n_m\geq 1\\ n_1+\cdots+n_m=n}}\frac{(-1)^{m-1}}{m}\frac{n!}{n_1!\dots n_m!}\times$$
$$\int_{E^m} f(x_1)^{n_1}\mathbb{K}(x_1,x_2)\dots f(x_m)^{n_m}\mathbb{K}(x_m,x_1)d\mu(x_1)\dots d\mu(x_m)$$

where we have considered $y_{m+1}=y_1$ in the first expression. That is the formula that we wanted to show for the cumulants, so the lemma is proven. □

That being done, once we have managed to prove this expression for the cumulants of the random variables S_{f_l} for every $l\geq 0$, we will use it to study their asymptotic behaviour, as stated in the following lemma.

Lemma 6.10. *Let $n\geq 1$ be any natural number and let $\delta>0$ be the value that appears in the statement of the theorem, then we have that the n-th cumulants of the above random variables verify that $C_n(S_{f_l}) = O((\mathrm{Var}(S_{f_l}))^{\delta+\epsilon})$ for any $\epsilon>0$.*

Proof. Indeed, using our own ideas (inspired by Lemma 2 from [22]), let $n\geq 1$ be any natural number, we can start by observing that, thanks to the above development for the n-th cumulant of S_{f_l}, we have that $C_n(S_{f_l})$ is a linear combination of expressions:

$$\int_{E^m} f_l(x_1)^{n_1}\mathbb{K}_l(x_1,x_2)\dots f_l(x_m)^{n_m}\mathbb{K}_l(x_m,x_1)d\mu(x_1)\dots d\mu(x_m)$$

for some $n_i\geq 1$ and $m\geq 1$ and now we proceed to see that each of them is $O((\mathrm{Var}(S_{f_l}))^{\delta+\epsilon})$, distinguishing two cases.

If $m=1$, we can observe that $|\int_E f_l^n(x)\mathbb{K}_l(x,x)d\mu(x)|\leq\int_E|f_l^n(x)|\mathbb{K}_l(x,x)d\mu(x)\leq$ (recalling that f_l is bounded) $\leq\|f_l^{n-1}\|_\infty\int_E|f_l(x)|\mathbb{K}_l(x,x)d\mu(x)=\|f_l^{n-1}\|_\infty\mathbb{E}(S_{|f_l|})=O((\mathrm{Var}(S_{f_l}))^{\delta+\epsilon})$ for any $\epsilon>0$, where we have used Lemma 6.9 and the fact that we know that $\|f_l\|_\infty=o((\mathrm{Var}(S_{f_l}))^{\epsilon'})$ for any $\epsilon'>0$ and $\mathbb{E}(S_{|f_l|})=O((\mathrm{Var}(S_{f_l}))^{\delta})$ by hypothesis. That shows this case.

Otherwise, if $m>1$, we can start by noticing that in the integral

$$|\int_{E^m} f_l(x_1)^{n_1}\mathbb{K}_l(x_1,x_2)\dots f_l(x_m)^{n_m}\mathbb{K}_l(x_m,x_1)d\mu(x_1)\dots d\mu(x_m)|$$

we can write $f_l=\max(f_l,0)-\max(0,-f_l)=f_{+,l}-f_{-,l}$ and thus, developing the obtained expression, we can observe that the above integral is bounded by a linear combination of similar integrals where we substitute f_l by its positive or negative part. Let now be

$$|\int_{E^m} f_{\pm,l}(x_1)^{n_1}\mathbb{K}_l(x_1,x_2)\dots f_{\pm,l}(x_m)^{n_m}\mathbb{K}_l(x_m,x_1)d\mu(x_1)\dots d\mu(x_m)|$$

one of the terms of this linear combination, that we fix. Then, with the aim of showing that it is $O((\mathrm{Var}(S_{f_l}))^{\delta+\epsilon})$ for any $\epsilon>0$, we consider again two different cases.

On the one hand, if $m = 2$, before developing the previous integral we can start by considering the variance $\mathrm{Var}(S_{f_{\pm,l}^{n_1}+f_{\pm,l}^{n_2}})$ and then, recalling our definitions, we notice that:

$$\mathrm{Var}(S_{f_{\pm,l}^{n_1}+f_{\pm,l}^{n_2}}) = \mathbb{E}((S_{f_{\pm,l}^{n_1}+f_{\pm,l}^{n_2}})^2) - \mathbb{E}((S_{f_{\pm,l}^{n_1}+f_{\pm,l}^{n_2}}))^2 =$$
$$\mathbb{E}((\sum_{x\in\mathscr{X}_l}(f_{\pm,l}(x)^{n_1} + f_{\pm,l}(x)^{n_2}))^2) - \mathbb{E}((S_{f_{\pm,l}^{n_1}+f_{\pm,l}^{n_2}}))^2 =$$
$$\mathbb{E}(\sum_{x\neq y\in\mathscr{X}_l}(f_{\pm,l}(x)^{n_1} + f_{\pm,l}(x)^{n_2})(f_{\pm,l}(y)^{n_1} + f_{\pm,l}(y)^{n_2})) +$$
$$\mathbb{E}(\sum_{x\in\mathscr{X}_l}(f_{\pm,l}(x)^{n_1} + f_{\pm,l}(x)^{n_2})^2) - \mathbb{E}(\sum_{x\in\mathscr{X}_l}(f_{\pm,l}(x)^{n_1} + f_{\pm,l}(x)^{n_2}))^2 =$$
$$\text{(applying Lemma 6.9 to rewrite these expected values)} =$$
$$\int_{E^2}(f_{\pm,l}(x)^{n_1} + f_{\pm,l}(x)^{n_2})(f_{\pm,l}(y)^{n_1} + f_{\pm,l}(y)^{n_2}) \times$$
$$(\mathbb{K}_l(x,x)\mathbb{K}_l(y,y) - |\mathbb{K}_l(x,y)|^2)d\mu(x)d\mu(y) +$$
$$\int_E(f_{\pm,l}(x)^{n_1} + f_{\pm,l}(x)^{n_2})^2\mathbb{K}_l(x,x)d\mu(x) -$$
$$(\int_E(f_{\pm,l}(x)^{n_1} + f_{\pm,l}(x)^{n_2})\mathbb{K}_l(x,x)d\mu(x))^2 =$$
$$\int_{E^2}(f_{\pm,l}(x)^{n_1} + f_{\pm,l}(x)^{n_2})(f_{\pm,l}(y)^{n_1} + f_{\pm,l}(y)^{n_2})\mathbb{K}_l(x,x)\mathbb{K}_l(y,y)d\mu(x)d\mu(y)$$
$$-\int_{E^2}(f_{\pm,l}(x)^{n_1} + f_{\pm,l}(x)^{n_2})(f_{\pm,l}(y)^{n_1} + f_{\pm,l}(y)^{n_2})|\mathbb{K}_l(x,y)|^2d\mu(x)d\mu(y) +$$
$$\int_E(f_{\pm,l}(x)^{n_1} + f_{\pm,l}(x)^{n_2})^2\mathbb{K}_l(x,x)d\mu(x) -$$
$$(\int_E(f_{\pm,l}(x)^{n_1} + f_{\pm,l}(x)^{n_2})\mathbb{K}_l(x,x)d\mu(x))^2 =$$
$$-\int_{E^2}(f_{\pm,l}(x)^{n_1} + f_{\pm,l}(x)^{n_2})(f_{\pm,l}(y)^{n_1} + f_{\pm,l}(y)^{n_2})|\mathbb{K}_l(x,y)|^2d\mu(x)d\mu(y)$$
$$+\int_E(f_{\pm,l}(x)^{n_1} + f_{\pm,l}(x)^{n_2})^2\mathbb{K}_l(x,x)d\mu(x)$$

So, in other words, we have managed to show that:

$$\mathrm{Var}(S_{f_{\pm,l}^{n_1}+f_{\pm,l}^{n_2}}) = -\int_{E^2}(f_{\pm,l}(x)^{n_1} + f_{\pm,l}(x)^{n_2})(f_{\pm,l}(y)^{n_1}$$
$$+f_{\pm,l}(y)^{n_2})|\mathbb{K}_l(x,y)|^2d\mu(x)d\mu(y) + \int_E(f_{\pm,l}(x)^{n_1} + f_{\pm,l}(x)^{n_2})^2\mathbb{K}_l(x,x)d\mu(x)$$

something that, combined with the fact that $\mathrm{Var}(S_{f_{\pm,l}^{n_1}+f_{\pm,l}^{n_2}}) \geq 0$, tells us that:

$$\int_{E^2}(f_{\pm,l}(x)^{n_1} + f_{\pm,l}(x)^{n_2})(f_{\pm,l}(y)^{n_1} + f_{\pm,l}(y)^{n_2})|\mathbb{K}_l(x,y)|^2d\mu(x)d\mu(y) \leq$$
$$\int_E(f_{\pm,l}(x)^{n_1} + f_{\pm,l}(x)^{n_2})^2\mathbb{K}_l(x,x)d\mu(x)$$

With that in mind, focusing on the desired integral, we are ready to conclude that:

$$\int_{E^2} f_{\pm,l}(x_1)^{n_1}|\mathbb{K}_l(x_1,x_2)|^2 f_{\pm,l}(x_2)^{n_2}d\mu(x_1)d\mu(x_2) \leq$$
$$\int_{E^2}(f_{\pm,l}(x_1)^{n_1} + f_{\pm,l}(x_1)^{n_2})|\mathbb{K}_l(x_1,x_2)|^2(f_{\pm,l}(x_2)^{n_1} + f_{\pm,l}(x_2)^{n_2})d\mu(x_1)d\mu(x_2) \leq$$
$$\text{(thanks to the inequality we have just deduced)} \leq$$
$$\int_E(f_{\pm,l}(x)^{n_1} + f_{\pm,l}(x)^{n_2})^2\mathbb{K}_l(x,x)d\mu(x) \leq$$
$$\int_E f_{\pm,l}(x)^{2n_1}\mathbb{K}_l(x,x)d\mu(x) + \int_E f_{\pm,l}(x)^{2n_2}\mathbb{K}_l(x,x)d\mu(x) +$$
$$2\int_E f_{\pm,l}(x)^{n_1}f_{\pm,l}(x)^{n_2}\mathbb{K}_l(x,x)d\mu(x) \leq$$
$$(\|f_l\|_\infty^{2n_1-1} + 2\|f_l\|_\infty^{n_1+n_2-1} + \|f_l\|_\infty^{2n_2-1})\int_E|f_l(x)|\mathbb{K}_l(x,x)d\mu(x) =$$
$$(\|f_l\|_\infty^{2n_1-1} + 2\|f_l\|_\infty^{n_1+n_2-1} + \|f_l\|_\infty^{2n_2-1})\mathbb{E}(S_{|f_l|}) = O((\mathrm{Var}(S_{f_l}))^{\delta+\epsilon})$$

for any $\epsilon > 0$, where we have used Lemma 6.9 and the fact the we know that $\|f_l\|_\infty = o((\mathrm{Var}(S_{f_l}))^{\epsilon'})$ for any $\epsilon' > 0$ and $\mathbb{E}(S_{|f_l|}) = O((\mathrm{Var}(S_{f_l}))^{\delta})$ by hypothesis.

On the other hand, if $m > 2$, denote by $F_{\pm,l}^{n_i} : L^2(E,\mu) \longrightarrow L^2(E,\mu)$ the operator multiplication by $f_{\pm,l}^{n_i}$ for all $i \in \{1,\dots,m\}$ and then, using it to develop the expression of the corresponding integral, we can observe that:

$$
\begin{gathered}
|\int_{E^m} f_{\pm,l}(x_1)^{n_1}\mathbb{K}_l(x_1,x_2)\dots f_{\pm,l}(x_m)^{n_m}\mathbb{K}_l(x_m,x_1)d\mu(x_1)\dots d\mu(x_m)| = \\
|\int_{E^2} f_{\pm,l}(x_1)^{\frac{n_1}{2}}\mathbb{K}_l(x_1,x_2)f_{\pm,l}(x_2)^{\frac{n_2}{2}}(\int_{E^{m-2}} f_{\pm,l}(x_2)^{\frac{n_2}{2}}\mathbb{K}_l(x_2,x_3)\cdots \times \\
f_{\pm,l}(x_m)^{n_m}\mathbb{K}_l(x_m,x_1)f_{\pm,l}(x_1)^{\frac{n_1}{2}}d\mu(x_3)\dots d\mu(x_m))d\mu(x_1)d\mu(x_2)| \leq \\
\int_{E^2} |f_{\pm,l}(x_1)^{\frac{n_1}{2}}\mathbb{K}_l(x_1,x_2)f_{\pm,l}(x_2)^{\frac{n_2}{2}}| \cdot |\int_{E^{m-2}} f_{\pm,l}(x_2)^{\frac{n_2}{2}}\mathbb{K}_l(x_2,x_3)\cdots \times \\
f_{\pm,l}(x_m)^{n_m}\mathbb{K}_l(x_m,x_1)f_{\pm,l}(x_1)^{\frac{n_1}{2}}d\mu(x_3)\dots d\mu(x_m)|d\mu(x_1)d\mu(x_2) \leq \\
\text{(by the Cauchy-Schwarz inequality)} \leq \\
(\int_{E^2} f_{\pm,l}(x_1)^{n_1}|\mathbb{K}_l(x_1,x_2)|^2 f_{\pm,l}(x_2)^{n_2}d\mu(x_1)d\mu(x_2))^{\frac{1}{2}} \times \\
\|\int_{E^{m-2}} f_{\pm,l}(x_2)^{\frac{n_2}{2}}\mathbb{K}_l(x_2,x_3)\dots f_{\pm,l}(x_m)^{n_m}\mathbb{K}_l(x_m,x_1)f_{\pm,l}(x_1)^{\frac{n_1}{2}}d\mu(x_3)\dots d\mu(x_m)\|_2
\end{gathered}
$$

where in this last norm we consider the integral as a function of x_1 and x_2. Now, focusing on the second factor in the expression and denoting by $g : E^2 \longrightarrow \mathbb{C}$ the function given by $g(x_m,x_1) = f_{\pm,l}(x_m)^{\frac{n_m}{2}}\mathbb{K}_l(x_m,x_1)f_{\pm,l}(x_1)^{\frac{n_1}{2}}$ for all $(x_m,x_1) \in E^2$, we can conclude that:

$$
\begin{gathered}
\|\int_{E^{m-2}} f_{\pm,l}(x_2)^{\frac{n_2}{2}}\mathbb{K}_l(x_2,x_3)\dots f_{\pm,l}(x_m)^{n_m}\mathbb{K}_l(x_m,x_1)f_{\pm,l}(x_1)^{\frac{n_1}{2}}d\mu(x_3)\dots d\mu(x_m)\|_2 \\
= \|\int_{E^{m-2}} f_{\pm,l}(x_2)^{\frac{n_2}{2}}\mathbb{K}_l(x_2,x_3)\dots f_{\pm,l}(x_m)^{\frac{n_m}{2}}g(x_m,x_1)d\mu(x_3)\dots d\mu(x_m)\|_2 = \\
\|f_{\pm,l}(x_2)^{\frac{n_2}{2}}\int_E \mathbb{K}_l(x_2,x_3)f_{\pm,l}(x_3)^{n_3}\int_E \mathbb{K}_l(x_3,x_4)\cdots \times \\
\int_E \mathbb{K}_l(x_{m-1},x_m)f_{\pm,l}(x_m)^{\frac{n_m}{2}}g(x_m,x_1)d\mu(x_m)\dots d\mu(x_3)\|_2 = \\
\|\|F_{\pm,l}^{\frac{n_2}{2}}\mathscr{K}_l F_{\pm,l}^{n_3}\dots F_{\pm,l}^{n_{m-1}}\mathscr{K}_l F_{\pm,l}^{\frac{n_m}{2}}(g(\cdot,x_1))\|_{2,\text{ in } x_2}\|_{2,\text{ in } x_1} \leq \\
\text{(by the properties of the operator norm)} \leq \\
\|\|F_{\pm,l}^{\frac{n_2}{2}}\|\|\mathscr{K}_l\|\|F_{\pm,l}^{n_3}\|\dots\|F_{\pm,l}^{n_{m-1}}\|\|\mathscr{K}_l\|\|F_{\pm,l}^{\frac{n_m}{2}}\|\|g(x_m,x_1)\|_{2,\text{ in } x_m}\|_{2,\text{ in } x_1} \leq \\
\|f_l\|_\infty^c\|\mathscr{K}_l\|^{m-1}\|g\|_2 \leq \|f_l\|_\infty^c\|g\|_2
\end{gathered}
$$

for a certain constant $c > 0$ independent of l, where we have used that $\|\mathscr{K}_l\| \leq 1$ because we are considering a determinantal point process and that for any function $h \in L^2(E^2,\mu^2)$, its norm can be written as:

$$
\begin{gathered}
\|h\|_2 = (\int_{E^2}|h(x_1,x_2)|^2 d\mu(x_1)d\mu(x_2))^{\frac{1}{2}} = (\int_E(\int_E |h(x_1,x_2)|^2 d\mu(x_2))d\mu(x_1))^{\frac{1}{2}} = \\
(\int_E((\int_E |h(x_1,x_2)|^2 d\mu(x_2))^{\frac{1}{2}})^2 d\mu(x_1))^{\frac{1}{2}} = (\int_E(\|h(x_1,x_2)\|_{2,\text{ in } x_2})^2 d\mu(x_1))^{\frac{1}{2}} = \\
\|\|h(x_1,x_2)\|_{2,\text{ in } x_2}\|_{2,\text{ in } x_1}
\end{gathered}
$$

So in conclusion, combining all of the above and using the result for the case $m = 2$, we can finally deduce that:

$$|\int_{E^m} f_{\pm,l}(x_1)^{n_1}\mathbb{K}_l(x_1,x_2)\dots f_{\pm,l}(x_m)^{n_m}\mathbb{K}_l(x_m,x_1)d\mu(x_1)\dots d\mu(x_m)| \leq$$
$$\|f_l\|_\infty^c(\int_{E^2} f_{\pm,l}(x_1)^{n_1}|\mathbb{K}_l(x_1,x_2)|^2 f_{\pm,l}(x_2)^{n_2}d\mu(x_1)d\mu(x_2))^{\frac{1}{2}} \times$$
$$(\int_{E^2} f_{\pm,l}(x_m)^{n_m}|\mathbb{K}_l(x_m,x_1)|^2 f_{\pm,l}(x_1)^{n_1}d\mu(x_m)d\mu(x_1))^{\frac{1}{2}} \leq$$
$$O((\mathrm{Var}(S_{f_l}))^{\frac{\delta+\epsilon}{2}})O((\mathrm{Var}(S_{f_l}))^{\frac{\delta+\epsilon}{2}}) = O((\mathrm{Var}(S_{f_l}))^{\delta+\epsilon})$$

where we have used again that $\|f_l\|_\infty = o((\mathrm{Var}(S_{f_l}))^{\epsilon'})$ for any $\epsilon' > 0$ and thus this factor can be easily absorbed into the rest with the right choice of ϵ'.

Finally, notice that we have managed to see that $C_n(S_{f_l})$ is a linear combination of factors that are $O((\mathrm{Var}(S_{f_l}))^{\delta+\epsilon})$, something that allows us to conclude that $C_n(S_{f_l}) = O((\mathrm{Var}(S_{f_l}))^{\delta+\epsilon})$ for any $\epsilon > 0$. That completes the proof. □

From here, with all of the above in mind, we can now recall that it is a well-known result from probability that convergence of higher order cumulants of a sequence of random variables to 0 implies convergence in distribution to a Gaussian law (see Theorem 1 of [12]). So, that tells us that to complete the proof of this version of the Central Limit Theorem for DPPs, it suffices to see that the cumulants of our sequence of random variables

$$\left\{\frac{S_{f_l} - \mathbb{E}(S_{f_l})}{\sqrt{\mathrm{Var}(S_{f_l})}}\right\}_{l\geq 0}$$

fulfil the following result.

Lemma 6.11. *There exists $N \geq 3$ such that the cumulants of the above sequence of random variables verify that:*

$$C_1(\tfrac{S_{f_l}-\mathbb{E}(S_{f_l})}{\sqrt{\mathrm{Var}(S_{f_l})}}) = 0 \quad \forall l \geq 0$$
$$C_2(\tfrac{S_{f_l}-\mathbb{E}(S_{f_l})}{\sqrt{\mathrm{Var}(S_{f_l})}}) = 1 \quad \forall l \geq 0$$
$$\lim_{l\to\infty} C_n(\tfrac{S_{f_l}-\mathbb{E}(S_{f_l})}{\sqrt{\mathrm{Var}(S_{f_l})}}) = 0 \quad \forall n > N$$

Proof. Indeed, to show this result, given any $l \geq 0$, we can start by computing the first two cumulants of the random variable S_{f_l}. To do it, substituting $n = 1$ and $n = 2$ in the combinatorial formula already proven for them in Lemma 6.8, we can easily observe that:

$$C_1(S_{f_l}) = \int_E f(x)\mathbb{K}(x,x)d\mu(x) = \mathbb{E}(S_{f_l})$$
$$C_2(S_{f_l}) = \int_E f(x)^2\mathbb{K}(x,x)d\mu(x) - \int_{E^2} f(x_1)f(x_2)|\mathbb{K}(x_1,x_2)|^2 d\mu(x_1)d\mu(x_2) = \mathrm{Var}(S_{f_l})$$

where we have used previous developments of the variance in this last equality.

Then, with that in mind, focusing now on the normalized version of the random variables and applying the definition of the cumulants, we can observe that for any

$t \in \mathbb{R}$ we have that:

$$\log\left(\mathbb{E}\left(e^{it\frac{S_{f_l}-\mathbb{E}(S_{f_l})}{\sqrt{\mathrm{Var}(S_{f_l})}}}\right)\right) = \log\left(\mathbb{E}\left(e^{i\frac{t}{\sqrt{\mathrm{Var}(S_{f_l})}}S_{f_l}}\right)\right) - it\frac{\mathbb{E}(S_{f_l})}{\sqrt{\mathrm{Var}(S_{f_l})}} =$$

$$\sum_{n=1}^{\infty}\frac{C_n(S_f)}{(\mathrm{Var}(S_{f_l}))^{\frac{n}{2}}}\frac{(it)^n}{n!} - it\frac{\mathbb{E}(S_{f_l})}{\sqrt{\mathrm{Var}(S_{f_l})}} = \frac{C_1(S_f)}{\sqrt{\mathrm{Var}(S_{f_l}))}}it - it\frac{\mathbb{E}(S_{f_l})}{\sqrt{\mathrm{Var}(S_{f_l})}} + \sum_{n=2}^{\infty}\frac{C_n(S_f)}{(\mathrm{Var}(S_{f_l}))^{\frac{n}{2}}}\frac{(it)^n}{n!}$$

something that allows us to conclude that for all $n \geq 3$, we have that:

$$C_1(\frac{S_{f_l}-\mathbb{E}(S_{f_l})}{\sqrt{\mathrm{Var}(S_{f_l})}}) = \frac{C_1(S_f)}{\sqrt{\mathrm{Var}(S_{f_l}))}} - \frac{\mathbb{E}(S_{f_l})}{\sqrt{\mathrm{Var}(S_{f_l})}} = 0$$

$$C_2(\frac{S_{f_l}-\mathbb{E}(S_{f_l})}{\sqrt{\mathrm{Var}(S_{f_l})}}) = \frac{C_2(S_f)}{\mathrm{Var}(S_{f_l})} = 1$$

$$C_n(\frac{S_{f_l}-\mathbb{E}(S_{f_l})}{\sqrt{\mathrm{Var}(S_{f_l})}}) = \frac{C_n(S_f)}{(\mathrm{Var}(S_{f_l}))^{\frac{n}{2}}}$$

In addition, now we can take $N = 2\delta + 2$ and then, given any $n > N$, we can use the big O properties of our cumulants deduced in Lemma 6.10 to observe that:

$$\lim_{l\to\infty} C_n(\frac{S_{f_l}-\mathbb{E}(S_{f_l})}{\sqrt{\mathrm{Var}(S_{f_l})}}) = \lim_{l\to\infty}\frac{C_n(S_f)}{(\mathrm{Var}(S_{f_l}))^{\frac{n}{2}}} = \lim_{l\to\infty}\frac{O((\mathrm{Var}(S_{f_l}))^{\delta+1})}{(\mathrm{Var}(S_{f_l}))^{\frac{n}{2}}} = 0$$

that is exactly what remained to be seen. That completes the proof of this lemma. □

So, combining all of the above and applying Theorem 1 from [12], we can conclude at the end that:

$$\frac{S_{f_l}-\mathbb{E}(S_{f_l})}{\sqrt{\mathrm{Var}(S_{f_l})}} \xrightarrow{\mathcal{L}} N(0,1) \quad \text{when } l \longrightarrow \infty$$

that is exactly what we wanted to see, something that completes the proof of this CLT for DPPs. □

6.4. **Concentration inequality.** Moving on and following the reference [19] (filling in some minor details by ourselves as usual), now we will focus on another interesting and useful property of projection determinantal point processes, that allows us to describe how close are functions of the process to their expected value, under not very restrictive conditions.

To do it, before stating and showing the result, we need to recall some notions. Indeed, on the one hand, given any $n \geq 1 \in \mathbb{N}$, denote by $M_n^*(E)$ the set of all simple point measures in E with total measure n and, in it, we can consider the total variation distance $d : M_n^*(E) \times M_n^*(E) \longrightarrow [0,\infty)$ given by $d(\mu,\tau) = 2\sup_{D\in\mathcal{B}(E)}|\mu(D)-\tau(D)|$ for all $\mu,\tau \in M_n^*(E)$. Then, it is clear that d is a well-defined distance in $M_n^*(E)$ and thus $(M_n^*(E), d)$ is a metric space.

On the other hand, to prove the result, we also need to introduce a technical property that projection DPPs satisfy, called the stochastic covering property.

To do it, let $\mathscr{X}$ be a projection determinantal point process in E with kernel $\mathbb{K} : E^2 \longrightarrow \mathbb{C}$ given by $\mathbb{K}(x,y) = \sum_{k=1}^{n}\varphi_k(x)\overline{\varphi_k}(y)$ μ-a.e. for $x, y \in E$, where

$\{\varphi_k\}_{k=1}^n$ denotes a finite orthonormal system of $L^2(E,\mu)$. Then, given a set of points $x_n,\dots,x_1\in E$, recall that in Section 5 we introduced the subspaces $H_n,\dots,H_1$ associated to the DPP and the sequence of points, defined as follows.

First of all, consider the space $H=\mathrm{Span}(\{\varphi_k\}_{k=1}^n)$ and, as discussed in Section 5, it is a reproducing kernel Hilbert space with kernel $\mathbb{K}$ (recall that here we need to select a representative of the $L^2(E,\mu)$ functions φ_k and redefine $\mathbb{K}$ in a set of measure 0 accordingly). Then, if we take $V\subseteq H$ any subspace, it is clear that V must also be a reproducing kernel Hilbert space. So, in this context, if we denote by $\mathbb{K}_V$ the reproducing kernel of V and also $\mathscr{K}_V\delta_x\coloneqq\mathbb{K}_V(\cdot,x)$ for all $x\in E$, we can define, recursively, $H_n\coloneqq H$ and then assuming that H_{i+1} as already been introduced for $1\leq i\leq n-1$ we have that:

$$H_i\coloneqq\{f\in H_{i+1}\mid\langle f,\mathscr{K}_{H_{i+1}}\delta_{x_{i+1}}\rangle=0\}=\{f\in H_{i+1}\mid\langle f,\mathbb{K}_{H_{i+1}}(\cdot,x_{i+1})\rangle=0\}=\{f\in H_{i+1}\mid f(x_{i+1})=0\}$$

Then, with that in mind, we have the following property involving these spaces.

Proposition 6.12 (Stochastic covering property)**.** *Let $\mathscr{X}$ be a determinantal process in E such that $\mathscr{X}(E)=n$ almost surely, then we have that for all $k\in\{1,\dots,n\}$ and for all $x_k,\dots,x_n\in E$ there exists $(\Omega_k,\mathcal{F}_k,P_k)$ a probability space and a random vector defined in it $((Z_1,\dots,Z_{k-1}),(Y_1,\dots,Y_k))\in E^{k-1}\times E^k$ such that:*

$$|\{Y_1,\dots,Y_k\}\setminus\{Z_1,\dots,Z_{k-1}\}|=1\quad P_k\text{ -almost surely}$$

and its (marginal) probability density functions $f_Y:E^k\longrightarrow\mathbb{R}$ of $(Y_1,\dots,Y_k)$ and $f_Z:E^{k-1}\longrightarrow\mathbb{R}$ of $(Z_1,\dots,Z_{k-1})$ (if $k>1$) are given by:

$$f_Y(p_1,\dots,p_k)=\prod_{i=1}^k\frac{\|\mathscr{K}_{H_i}\delta_{p_i}\|^2}{i}\qquad f_Z(p_1,\dots,p_{k-1})=\prod_{i=1}^{k-1}\frac{\|\mathscr{K}_{\widehat{H}_i}\delta_{p_i}\|^2}{i}$$

for all $p_1,\dots,p_k\in E$ respectively. Here, the subspaces $H_n,\dots,H_1$ are the ones given by the sequence of points $x_n,\dots,x_{k+1},p_k,\dots,p_1$ and, on the contrary, the subspaces $\widehat{H}_n,\dots,\widehat{H}_1$ are the ones given by the sequence $x_n,\dots,x_k,p_{k-1},\dots,p_1$, defined as in the previous comment.

In fact, this property can also be considered for general homogeneous and simple point processes with the appropriate changes. Notice that we will not give the proof of this proposition, because it is very long and complex and thus out of the scope of this project. More details can be found in our reference [19].

That being said, with all of the above in mind, now we are ready to introduce the result, that is a particular case of Theorem 3.4 from [19].

Theorem 6.13. *Let $\mathscr{X}$ be a projection determinantal process in E with $n\geq 1\in\mathbb{N}$ points almost surely, we consider $f:M_n^*(E)\longrightarrow\mathbb{R}$ a Lipschitz-1 function (Lipschitz function with Lipschitz constant equal to 1) with respect to the total variation distance and then, as a consequence, we can conclude that:*

$$P(f(\mathscr{X})-\mathbb{E}(f(\mathscr{X}))\geq a)\leq\exp\left(\frac{-a^2}{8n}\right)$$

$$P(|f(\mathscr{X}) - \mathbb{E}(f(\mathscr{X}))| \geq a) \leq 2\exp\left(\frac{-a^2}{8n}\right)$$

for all $a \geq 0 \in \mathbb{R}$.

Proof. Indeed, to show this result, let $\mathbb{K} : E^2 \longrightarrow \mathbb{C}$ denote the kernel of the determinantal point process $\mathscr{X}$, using the properties we already know about the joint intensities of the process we can start by observing that:

$$0 \leq \mathbb{E}(\tbinom{\mathscr{X}(E)}{2}2!) = \int_{E^2}(\mathbb{K}(x,x)\mathbb{K}(y,y) - |\mathbb{K}(x,y)|^2)d\mu(x)d\mu(y) =$$
$$(\int_E \mathbb{K}(x,x)d\mu(x))^2 - \int_{E^2}|\mathbb{K}(x,y)|^2 d\mu(x)d\mu(y) =$$
$$n^2 - \int_{E^2}|\mathbb{K}(x,y)|^2 d\mu(x)d\mu(y)$$

something that shows that $\int_{E^2}|\mathbb{K}(x,y)|^2 d\mu(x)d\mu(y) < \infty$. As a consequence, by usual arguments from the preliminaries, the kernel $\mathbb{K}$ is square integrable in E^2 and its associated integral operator $\mathscr{K}$ is bounded in $L^2(E,\mu)$, compact and self-adjoint.

From here, applying the Spectral theorem to $\mathscr{K}$ and noticing that $\mathscr{X}$ is a projection determinantal point process by hypothesis, we can conclude that there exists $\{\varphi_k\}_{k=1}^m \subseteq L^2(E,\mu)$ an orthonormal system such that $\mathbb{K}(x,y) = \sum_{k=1}^m \varphi_k(x)\overline{\varphi_k(y)}$ for all $(x,y) \in E^2$ (selecting representatives of φ_k and redefining $\mathbb{K}$ appropriately) and, in fact, notice that:

$$n = \int_E \mathbb{K}(x,x)d\mu(x) = \sum_{k=1}^m \int_E |\varphi_k(x)|^2 d\mu(x) = \sum_{k=1}^m 1 = m$$

With that in mind, observe that now we are in the hypothesis of Theorem 5.1 and thus we can consider a random vector $(X_1,\ldots,X_n)$ taking values in E^n that can be sampled following Algorithm 1 (from X_n to X_1) and assume without loss of generality that our point process $\mathscr{X}$ is given by $\mathscr{X} = \sum_{k=1}^n \delta_{X_k}$. Not only that, but in this case we can also redefine f as a function $f : E^n \longrightarrow \mathbb{R}$ in such a way that $f(\mathscr{X})$ is now $f(X_1,\ldots,X_n)$ and f is symmetric (that is, $f(x_1,\ldots,x_n) = f(x_{\sigma(1)},\ldots,x_{\sigma(n)})$ for all $x_1,\ldots,x_n \in E$ and $\sigma \in S_n$).

Now, using these expressions for our determinantal point process and function, given any $1 \leq k \leq n$, we can define the random variable $M_{n+1} = 0$ and also $M_k = \mathbb{E}(f(X_1,\ldots,X_n) \mid X_k,\ldots,X_n) - \mathbb{E}(f(X_1,\ldots,X_n))$ and then we start by showing the following lemma.

Lemma 6.14. *Given any* $1 \leq k \leq n+1$*, the random variable* M_k *defined above satisfies that* $M_k + \mathbb{E}(f(X_1,\ldots,X_n)) =$

$$= \int_{E^{k-1}} f(x_1,\ldots,x_{k-1},X_k,\ldots,X_n)\prod_{i=1}^{k-1}(\frac{\|\mathscr{K}_{H_i}\delta_{x_i}\|^2}{i})d\mu(x_1)\ldots d\mu(x_{k-1})$$

where we have used the same notations that in Theorem 5.1.

Proof. To show this result, given first any $1 \leq k \leq n$ and $A_k,\ldots,A_n \in \mathcal{B}(E)$ Borel sets, notice that the above integral is clearly measurable with respect to $X_k,\ldots,X_n$, so

we can apply the definition of conditional expectation and then it suffices to see that:

$$\mathbb{E}(\mathbb{1}_{\{X_k\in A_k,\ldots,X_n\in A_n\}}f(X_1,\ldots,X_n)) =$$
$$\mathbb{E}(\mathbb{1}_{\{X_k\in A_k,\ldots,X_n\in A_n\}}\int_{E^{k-1}} f(x_1,\ldots,x_{k-1},X_k,\ldots,X_n)\times$$
$$\prod_{i=1}^{k-1}(\frac{\|\mathscr{K}_{H_i}\delta_{x_i}\|^2}{i})d\mu(x_1)\ldots d\mu(x_{k-1}))$$

To do it, on the one hand, recalling the expression for the joint density of the random variables $X_1,\ldots,X_n$ (seen in Theorem 5.1), we can observe that:

$$\mathbb{E}(\mathbb{1}_{\{X_k\in A_k,\ldots,X_n\in A_n\}}f(X_1,\ldots,X_n)) =$$
$$\int_{E^n}\mathbb{1}_{\{x_k\in A_k,\ldots,x_n\in A_n\}}f(x_1,\ldots,x_n)\prod_{i=1}^{n}(\frac{\|\mathscr{K}_{H_i}\delta_{x_i}\|^2}{i})d\mu(x_1)\ldots d\mu(x_n)$$

On the other hand, using in fact the same technique and ideas, we can also conclude that:

$$\mathbb{E}(\mathbb{1}_{\{X_k\in A_k,\ldots,X_n\in A_n\}}\int_{E^{k-1}} f(x_1,\ldots,x_{k-1},X_k,\ldots,X_n)\times$$
$$\prod_{i=1}^{k-1}(\frac{\|\mathscr{K}_{H_i}\delta_{x_i}\|^2}{i})d\mu(x_1)\ldots d\mu(x_{k-1})) =$$
$$\int_{E^{n-k+1}}\mathbb{1}_{\{x_k\in A_k,\ldots,x_n\in A_n\}}\int_{E^{k-1}} f(x_1,\ldots,x_{k-1},x_k,\ldots,x_n)\times$$
$$\prod_{i=1}^{k-1}(\frac{\|\mathscr{K}_{H_i}\delta_{x_i}\|^2}{i})d\mu(x_1)\ldots d\mu(x_{k-1})\prod_{i=k}^{n}(\frac{\|\mathscr{K}_{H_i}\delta_{x_i}\|^2}{i})d\mu(x_k)\ldots d\mu(x_n) =$$
$$\int_{E^n}\mathbb{1}_{\{x_k\in A_k,\ldots,x_n\in A_n\}}f(x_1,\ldots,x_n)\prod_{i=1}^{n}(\frac{\|\mathscr{K}_{H_i}\delta_{x_i}\|^2}{i})d\mu(x_1)\ldots d\mu(x_n)$$

This shows that both expectations coincide.

Finally, regarding the case $k=n+1$, notice that by the definition of M_{n+1} we have that $M_{n+1}+\mathbb{E}(f(X_1,\ldots,X_n)) = \mathbb{E}(f(X_1,\ldots,X_n))$ and it is also clear that:

$$\int_{E^n} f(x_1,\ldots,x_{k-1},x_k,\ldots,x_n)\prod_{i=1}^{n}(\frac{\|\mathscr{K}_{H_i}\delta_{x_i}\|^2}{i})d\mu(x_1)\ldots d\mu(x_n) = \mathbb{E}(f(X_1,\ldots,X_n))$$

that is what we wanted to see. That completes the proof of this lemma. □

From here, now we aim to show as another lemma that the increments between these random variables M_k are bounded by 2, that is, given any $1\leq k\leq n$, we have that $|M_k - M_{k+1}|\leq 2$.

Lemma 6.15. *With the same notations as in Lemma 6.14, we have that* $|M_k - M_{k+1}|\leq 2$ *for all* $1\leq k\leq n$.

Proof. To prove this result, again in the context of Theorem 5.1 and given any $1\leq k\leq n$, we can start by observing that, applying the stochastic covering property for projection determinantal point processes (Proposition 6.12) and considering the (random) points $X_k,\ldots,X_n$, we have that there exists $(\Omega_k,\mathcal{F}_k,P_k)$ a probability space and $((Z_1,\ldots,Z_{k-1}),(Y_1,\ldots,Y_k))\in E^{k-1}\times E^k$ a random vector defined in it such that:

$$|\{Y_1,\ldots,Y_k\}\setminus\{Z_1,\ldots,Z_{k-1}\}| = 1 \quad P_k\text{ -almost surely}$$

and also that the (marginal) probability density functions $f_Y : E^k \longrightarrow \mathbb{R}$ of $(Y_1, \ldots, Y_k)$ and $f_Z : E^{k-1} \longrightarrow \mathbb{R}$ of $(Z_1, \ldots, Z_{k-1})$ (if $k > 1$) are given by:

$$f_Y(x_1, \ldots, x_k) = \prod_{i=1}^{k} \frac{\|\mathscr{K}_{H_i}\delta_{x_i}\|^2}{i} \qquad f_Z(x_1, \ldots, x_{k-1}) = \prod_{i=1}^{k-1} \frac{\|\mathscr{K}_{\widehat{H}_i}\delta_{x_i}\|^2}{i}$$

for all $x_1, \ldots, x_k \in E$ respectively. Here, notice that the subspaces $H_n, \ldots, H_1$ are the ones given by the sequence of (possibly random) points $X_n, \ldots, X_{k+1}, x_k, \ldots, x_1$ and, on the other hand, the subspaces $\widehat{H}_n, \ldots, \widehat{H}_1$ are the ones given by the sequence of (possibly random) points $X_n, \ldots, X_k, x_{k-1}, \ldots, x_1$, something that implies that in general $H_i \neq \widehat{H}_i$ for all $i \leq k-1$.

Related to this fact, observe that in this case the probability space $(\Omega_k, \mathcal{F}_k, P_k)$ is random and depends on the values of $X_k, \ldots, X_n$, but we do not reflect this fact in the notation for simplicity purposes.

With that in mind, we can then apply the properties of f, the previous expression shown for these random variables $M_k,\ M_{k+1}$ and the notions and ideas of the proof of Theorem 5.1 to conclude that:

$$\begin{gathered}
|M_k - M_{k+1}| = |\int_{E^{k-1}} f(x_1, \ldots, x_{k-1}, X_k, \ldots, X_n) \times \\
\prod_{i=1}^{k-1}(\tfrac{\|\mathscr{K}_{\widehat{H}_i}\delta_{x_i}\|^2}{i}) d\mu(x_1) \ldots d\mu(x_{k-1}) \\
- \int_{E^k} f(x_1, \ldots, x_k, X_{k+1}, \ldots, X_n) \prod_{i=1}^{k}(\tfrac{\|\mathscr{K}_{H_i}\delta_{x_i}\|^2}{i}) d\mu(x_1) \ldots d\mu(x_k)| = \\
\text{(by the law of the unconscious statistician)} = \\
|\int_{\Omega_k} f((Z_1, \ldots, Z_{k-1})(\omega), X_k, \ldots, X_n) dP_k(\omega) - \\
\int_{\Omega_k} f((Y_1, \ldots, Y_k)(\omega), X_{k+1}, \ldots, X_n) dP_k(\omega)| \leq \\
\int_{\Omega_k} |f((Z_1, \ldots, Z_{k-1})(\omega), X_k, \ldots, X_n) - f((Y_1, \ldots, Y_k)(\omega), X_{k+1}, \ldots, X_n)| dP_k(\omega) \\
\leq \int_{\Omega_k} d((\textstyle\sum_{i=1}^{k-1} \delta_{Z_i})(\omega) + \sum_{j=k}^{n} \delta_{X_j}, (\sum_{i=1}^{k} \delta_{Y_i})(\omega) + \sum_{j=k+1}^{n} \delta_{X_j}) dP_k(\omega) = \\
2 \int_{\Omega_k} \sup_{A \in \mathcal{B}(E)} |((\textstyle\sum_{i=1}^{k} \delta_{Y_i} - \sum_{i=1}^{k-1} \delta_{Z_i})(\omega))(A) - \delta_{X_k}(A)| dP_k(\omega) \leq \\
2 \int_{\Omega_k} dP_k(\omega) = 2
\end{gathered}$$

where we have used that $\{Y_1, \ldots, Y_k\} = \{Z_1, \ldots, Z_k\} \cup \{x\}$ for some $x \in E$ P_k -almost surely and that f is a Lipschitz-1 function with respect to the total variation distance. That completes the proof of the lemma. $\square$

Now, taking into account all of the above, given any $k \in \{1, \ldots, n\}$, we can consider the corresponding increment $M_k - M_{k+1}$ and then we have that:

$$\mathbb{E}(e^{\frac{\lambda}{2}(M_k - M_{k+1})} \mid X_{k+1}, \ldots, X_n) \leq e^{\frac{\lambda^2}{2}}$$

for all $\lambda \geq 0$. Indeed, given any $\lambda \geq 0$, we can start by considering the function $h : [-1, 1] \longrightarrow \mathbb{R}$ given by $h(x) = \frac{e^{-\lambda}}{2}(1 - x) + \frac{e^{\lambda}}{2}(x + 1)$ for all $x \in [-1, 1]$ and it is clear that it is an affine function. In fact, since $h(-1) = e^{-\lambda}$ and $h(1) = e^{\lambda}$, notice that h is the chord that connects these two points over the curve $e^{\lambda x}$, something that

combined with the fact that $e^{\lambda x}$ is a convex function in x allows us to conclude that $e^{\lambda x} \leq h(x)$ for all $x \in [-1, 1]$.

From here, using that we already know that $\frac{1}{2}(M_k - M_{k+1}) \in [-1, 1]$ and applying the linearity of the expected value, we can observe that:

$$\mathbb{E}(e^{\frac{\lambda}{2}(M_k - M_{k+1})} \mid X_{k+1}, \ldots, X_n) \leq \mathbb{E}(h(\tfrac{1}{2}(M_k - M_{k+1})) \mid X_{k+1}, \ldots, X_n) =$$
$$h(\tfrac{1}{2}\mathbb{E}(M_k - M_{k+1} \mid X_{k+1}, \ldots, X_n)) = h(0) = \tfrac{1}{2}(e^{\lambda} + e^{-\lambda})$$

where we have used that $\mathbb{E}(M_k - M_{k+1} \mid X_{k+1}, \ldots, X_n) = \mathbb{E}(M_k \mid X_{k+1}, \ldots, X_n) - M_{k+1} = M_{k+1} - M_{k+1} = 0$ by construction (martingale property). So, now it suffices to see that:

$$e^{\lambda} + e^{-\lambda} \leq 2e^{\frac{\lambda^2}{2}}$$

To do it, we can develop the Taylor expansions of these exponentials to deduce that:

$$e^{\lambda} + e^{-\lambda} \leq 2e^{\frac{\lambda^2}{2}} \iff \sum_{m=0}^{\infty} \frac{\lambda^m + (-\lambda)^m}{m!} \leq \sum_{m=0}^{\infty} 2\frac{(\frac{\lambda^2}{2})^m}{m!} \iff$$
$$\sum_{m=0}^{\infty} \frac{2\lambda^{2m}}{(2m)!} \leq \sum_{m=0}^{\infty} \frac{2\lambda^{2m}}{2^m m!} \impliedby 2^m m! \leq (2m)! \ \ \forall m \geq 0$$

and this final claim is true for $m = 0$, $m = 1$ and for $m \geq 2$, since in this last case both sides are a product of $2m$ factors with larger ones in the right. Consequently, combining all of the above, we have managed to see that $\mathbb{E}(e^{\frac{\lambda}{2}(M_k - M_{k+1})} \mid X_{k+1}, \ldots, X_n) \leq e^{\frac{\lambda^2}{2}}$.

From here, using this inequality and applying it for all the increments in an iterative way (as done in the proof of Theorem 7.2.1 in [1]), we can also deduce that:

$$\mathbb{E}(e^{\frac{\lambda}{2}M_1}) = \mathbb{E}(e^{\frac{\lambda}{2}\sum_{k=1}^{n}(M_k - M_{k+1})}) = \mathbb{E}(\textstyle\prod_{k=1}^{n} e^{\frac{\lambda}{2}(M_k - M_{k+1})}) =$$
$$\mathbb{E}(\mathbb{E}(\textstyle\prod_{k=1}^{n} e^{\frac{\lambda}{2}(M_k - M_{k+1})} \mid X_2, \ldots, X_n)) =$$
$$\mathbb{E}(\textstyle\prod_{k=2}^{n} e^{\frac{\lambda}{2}(M_k - M_{k+1})}\mathbb{E}(e^{\frac{\lambda}{2}(M_1 - M_2)} \mid X_2, \ldots, X_n)) \leq \mathbb{E}(\textstyle\prod_{k=2}^{n} e^{\frac{\lambda}{2}(M_k - M_{k+1})})e^{\frac{\lambda^2}{2}} \leq$$
$$\text{(applying iteratively the same argument)} \leq e^{\frac{\lambda^2 n}{2}}$$

Finally, now we are ready to show the desired inequality. To do it, we merely need to observe that $f(\mathscr{X}) - \mathbb{E}(f(\mathscr{X})) = M_1$ and then, as a consequence, choosing $\lambda = a/2n$ and applying the Markov inequality we can conclude that:

$$P(f(\mathscr{X}) - \mathbb{E}(f(\mathscr{X})) \geq a) = P(M_1 \geq a) = P(e^{\frac{\lambda M_1}{2}} \geq e^{\frac{\lambda a}{2}}) \leq \frac{\mathbb{E}(\exp(\frac{\lambda M_1}{2}))}{\exp(\frac{\lambda a}{2})} \leq$$
$$\frac{\exp(\frac{\lambda^2 n}{2})}{\exp(\frac{\lambda a}{2})} = \exp(\tfrac{1}{2}(\lambda^2 n - \lambda a)) = \exp(\tfrac{1}{2}(\tfrac{a^2}{4n^2}n - \tfrac{a}{2n}a)) = \exp(\tfrac{a^2}{8n} - \tfrac{a^2}{4n}) =$$
$$\exp(\tfrac{a^2}{8n} - \tfrac{2a^2}{8n}) = \exp(\tfrac{-a^2}{8n})$$

that is what we wanted to see.

To end the proof, regarding the two-sided inequality, we can easily obtain it as a corollary of the previous one. Indeed, given any ω such that $|f(\mathscr{X}(w)) - \mathbb{E}(f(\mathscr{X}))| \geq a$, then by definition of the absolute value we have that $f(\mathscr{X}(w)) - \mathbb{E}(f(\mathscr{X})) \geq a$ or

$f(\mathscr{X}(w)) - \mathbb{E}(f(\mathscr{X})) \leq -a$, something that allows us to conclude that:

$$P(|f(\mathscr{X}) - \mathbb{E}(f(\mathscr{X}))| \geq a) \leq$$
$$P(f(\mathscr{X}) - \mathbb{E}(f(\mathscr{X})) \geq a) + P((-f)(\mathscr{X}) - \mathbb{E}((-f)(\mathscr{X})) \geq a) \leq$$
$$2\exp(\tfrac{-a^2}{8n})$$

where we have applied the previous result for both f and $-f$. That is what remained to be seen, something that completes the proof. □

In fact, although here we have stated and shown the result for projection determinantal point processes, there is a similar version of this theorem for general determinantal point processes, that can be found at [19].

As a comment, before moving on, notice that this type of concentration properties can be useful, for example, when one tries to perform Montecarlo integration or measure estimation using determinantal point processes.

For example, just as a toy application of the inequality we have shown, we can start by considering $K \subseteq [0,1]^m$ a compact set and then take $\mathscr{X}$ a projection determinantal point process in $[0,1]^m$ with kernel $\mathbb{K}$ such that $\mathbb{E}(\mathscr{X}([0,1]^m)) = n$ and also $\mathbb{K}(x,x)$ is constant for almost all $x \in [0,1]^m$, with the usual Lebesgue measure.

From here, we can define the estimator $f : M_n^*([0,1]^m) \longrightarrow \mathbb{R}$ given by $f(\mu) = \mu(K)$ for all measures $\mu \in M_n^*([0,1]^m)$ and it is easy to see that it is Lipschitz-1. Indeed, given any $\mu, \tau \in M_n^*([0,1]^m)$, it is immediate to observe that:

$$|f(\mu) - f(\tau)| = |\mu(K) - \tau(K)| \leq 2 \sup_{D \in \mathcal{B}([0,1]^m)} |\mu(D) - \tau(D)| = d(\mu, \tau)$$

In addition, using the fact that $\mathbb{K}(x,x)$ is constant for almost all $x \in [0,1]^m$, we can also deduce that:

$$\mathbb{E}(f(\mathscr{X})) = \mathbb{E}(\mathscr{X}(K)) = \textstyle\int_K \mathbb{K}(x,x)dx = C\int_K dx = C|K| = C(\int_{[0,1]^m} dx)|K| =$$
$$(\textstyle\int_{[0,1]^m} \mathbb{K}(x,x)dx)|K| = \mathbb{E}(\mathscr{X}([0,1]^m))|K| = n|K|$$

With that in mind, now we can finally use our result to conclude that:

$$P\left(\left|\frac{\mathscr{X}(K)}{n} - |K|\right| \geq \epsilon\right) = P\left(|f(\mathscr{X}) - \mathbb{E}(f(\mathscr{X}))| \geq n\epsilon\right) \leq 2\exp\left(-\frac{n\epsilon^2}{8}\right)$$

In other words, this tells us that when we increase the number of points n, the error of the approximation of the measure of K goes to 0 exponentially, something that motivates the use of this type of determinantal point processes to estimate the measure of the set.

7. Examples of DPPs

Although we have already developed the general notion of a determinantal point process and studied some of its more relevant properties, notice that we have not yet seen any natural example from this family of processes in detail. But, in fact, there are many of these classical examples that are determinantal, some of which appear in Chapters 4.3 and 6 from [11], the reference that we will mainly follow in this section. Actually, as explained in the introduction, DPPs were considered in abstract, by Macchi in 1975 (see [16]), to try to obtain a global and unified understanding of many of these interesting examples that were already known to have joint intensities that could be expressed as a determinant.

7.1. **Ginibre ensemble.** For example, one of these natural point processes that is determinantal (something that was known before the notion of DPP even existed) is the Ginibre ensemble, that is formed by the eigenvalues of a random matrix with independent and identically distributed standard complex Gaussian entries. Now, we proceed to state this result and to show the determinantal nature of this process.

Theorem 7.1. *Let $M \in \mathbb{C}^{n\times n}$ be a random matrix of size $n \geq 1$ with independent and identically distributed standard complex Gaussian entries, we consider $\mathscr{X} = \sum_{i=1}^{n} \delta_{Z_i}$ the point process in $\mathbb{C}$ formed by $Z_1, \ldots, Z_n$ the eigenvalues of M. Then, we have that $\mathscr{X}$ is a determinantal point process in $\mathbb{C}$ with kernel $\mathbb{K} : \mathbb{C}^2 \longrightarrow \mathbb{C}$ given by:*

$$\mathbb{K}(z,w) = \sum_{k=0}^{n-1} \frac{(z\overline{w})^k}{k!} \quad \text{for all } z, w \in \mathbb{C}$$

with respect to the background measure μ with $d\mu(z) = \frac{1}{\pi} e^{-|z|^2} dm(z)$, where m denotes the usual Lebesgue measure in $\mathbb{C}$.

Proof. To show this result, as a preliminary, we can start by considering the space $L^2(\mathbb{C}, \mu)$ (it is clear that $\mathbb{C}$ is a Polish space and μ a Radon measure) and, in it, we can define the functions $\{\varphi_k\}_{k=0}^{n-1} \subseteq L^2(\mathbb{C}, \mu)$ given by $\varphi_k(z) = \frac{z^k}{\sqrt{k!}}$ for all $z \in \mathbb{C}$. Then, we can observe that these functions form an orthonormal system in $L^2(\mathbb{C}, \mu)$.

Indeed, to prove it, given any $\alpha, \beta \in \{0, \ldots, n-1\}$, we can compute the inner product between φ_α and φ_β to deduce that:

$$\langle \varphi_\alpha, \varphi_\beta \rangle = \textstyle\int_{\mathbb{C}} \varphi_\alpha(z)\overline{\varphi_\beta}(z) d\mu(z) = \frac{1}{\pi\sqrt{\alpha!\beta!}} \int_{\mathbb{C}} z^\alpha \overline{z}^\beta e^{-|z|^2} dz =$$

$$\text{(applying a change of variable to polar coordinates)} =$$

$$\frac{1}{\pi\sqrt{\alpha!\beta!}} \int_0^\infty \int_0^{2\pi} (re^{\theta i})^\alpha (re^{-\theta i})^\beta e^{-r^2} r dr d\theta = \frac{1}{\pi\sqrt{\alpha!\beta!}} \left(\int_0^\infty r^{\alpha+\beta+1} e^{-r^2} dr\right)\left(\int_0^{2\pi} e^{(\alpha-\beta)\theta i} d\theta\right)$$

and now we distinguish two different cases.

On the one hand, if $\alpha \neq \beta$, we can observe that:

$$\int_0^{2\pi} e^{(\alpha-\beta)\theta i} d\theta = \left(\frac{e^{(\alpha-\beta)\theta i}}{(\alpha-\beta)i}\right)\Big|_{\theta=0}^{\theta=2\pi} = \frac{1}{(\alpha-\beta)i}(e^{2\pi(\alpha-\beta)i} - 1) = 0$$

where we have used that $\alpha - \beta \in \mathbb{Z}$, something that shows that $\langle \varphi_\alpha, \varphi_\beta \rangle = 0$ and the system is orthogonal.

On the other hand, focusing now of the case where $\alpha = \beta$, we can then conclude that:

$$\|\varphi_\alpha\|^2 = \tfrac{1}{\pi \alpha!}(\int_0^\infty r^{2\alpha+1} e^{-r^2} dr)(\int_0^{2\pi} 1 d\theta) = \tfrac{2}{\alpha!} \int_0^\infty r^{2\alpha+1} e^{-r^2} dr = \\ \text{(applying the change of variable } s = r^2) = \\ \tfrac{1}{\alpha!} \int_0^\infty s^\alpha e^{-s} ds = \tfrac{\Gamma(\alpha+1)}{\alpha!} = \tfrac{\alpha!}{\alpha!} = 1$$

where we have used the definition and the known properties of the Gamma function. That shows, by definition, that the set $\{\varphi_k\}_{k=0}^{n-1} \subseteq L^2(\mathbb{C}, \mu)$ is an orthonormal system of $L^2(\mathbb{C}, \mu)$.

With that in mind, notice that through these functions we can express our kernel $\mathbb{K}$ as $\mathbb{K}(z, w) = \sum_{k=0}^{n-1} \varphi_k(z)\overline{\varphi_k}(w)$ for all $z, w \in \mathbb{C}$. Then, using this expression and applying Proposition 4.2, we can conclude that there exists a (projection) determinantal point process with kernel $\mathbb{K}$ defined in $(\mathbb{C}, \mu)$ that has n points almost surely. So, thanks to this proposition, we have managed to see that the result we are now trying to check makes sense. In fact, not only that, but analysing the proof of the same proposition, we can easily conclude that to show this theorem it suffices to see that the joint density of $(Z_1, \ldots, Z_n)$ is given by $\frac{1}{n!} \det \mathbb{K}(z_i, z_j)_{i,j \leq n}$ for all $z_1, \ldots, z_n \in \mathbb{C}$.

As a consequence, in what follows, we will focus on computing this joint density function. To do it, first of all, we can start by recalling that if X is a standard complex random variable, then its density function $f : \mathbb{C} \longrightarrow \mathbb{R}$ is given by:

$$f(z) dx \wedge dy = \frac{1}{\pi} e^{-|z|^2} dx \wedge dy$$

for all $z \in \mathbb{C}$. Notice that in the above expression, we have included the differential forms that accompany it, because they will play a key role in what follows. And, in relation with them, using the previous notations, a first observation that we can do is that:

$$dz \wedge d\overline{z} = d(x + iy) \wedge d(x - iy) = \\ dx \wedge dx - i dx \wedge dy + i dy \wedge dx + dy \wedge dy = -2i dx \wedge dy = \tfrac{2}{i} dx \wedge dy$$

where we have used the known properties of the wedge product and the differential forms. So, that means that we can also express f as:

$$f(z) dx \wedge dy = \frac{i}{2\pi} e^{-|z|^2} dz \wedge d\overline{z}$$

for all $z \in \mathbb{C}$.

From here, recall the entries of our random matrix $M = (M_{i,j})_{i,j}$ are independent and identically distributed standard complex random variables, something that, thanks to the previous comments, allows us to conclude that their joint density function is

given by:

$$\frac{i^{n^2}}{(2\pi)^{n^2}} e^{-\sum_{i,j}|m_{i,j}|^2} \bigwedge_{i,j} dm_{i,j} \wedge d\overline{m}_{i,j} = \frac{i^{n^2}}{(2\pi)^{n^2}} e^{-\sum_{i=1}^{n}\langle m_{i,\cdot}, m_{i,\cdot}\rangle} \bigwedge_{i,j} dm_{i,j} \wedge d\overline{m}_{i,j} =$$
$$\frac{i^{n^2}}{(2\pi)^{n^2}} e^{-\mathrm{tr}(M^*M)} \bigwedge_{i,j} dm_{i,j} \wedge d\overline{m}_{i,j}$$

Now, in this situation, to obtain the joint density of the eigenvalues $(Z_1, \dots, Z_n)$ of M we would like to apply a change of variable that goes from M to $(Z_1, \dots, Z_n)$.

To do it, we can consider the Schur form of the matrix M, that allow us to express it as $M = V(Z+T)V^*$ for some matrices V, Z and T such that V is unitary, T is strictly upper triangular and $Z = \mathrm{diag}(Z_1, \dots, Z_n)$. This expression is not unique in general, but if we assume that the eigenvalues are pairwise different (something that happens with probability 1) and order them, it is a well-known property of the Schur decomposition that we can impose $V_{j,j} \geq 0$ to make it unique. Notice that this is indeed something that we can impose, because if we take any matrix $\Theta = \mathrm{diag}(e^{i\theta_1}, \dots, e^{i\theta_n})$, clearly unitary, then:

$$V\Theta(Z + \Theta^* T \Theta)(V\Theta)^* = V\Theta(Z + \Theta^* T \Theta)\Theta^* V^* = V(\Theta Z \Theta^* + T)V^* = V(Z+T)V^*$$

and we can achieve $(V\Theta)_{j,j} \geq 0$ by choosing $e^{i\theta_j} = \frac{\overline{V_{j,j}}}{|V_{j,j}|}$ for all $j \in \{1, \dots, n\}$.

So, setting ourselves in this situation, we can consider M as a function of V, Z, T and then, applying the change of variable formula for densities, we can conclude that if $T = (T_{i,j})_{i,j}$ and $V = (V_{i,j})_{i,j}$, we have that the joint density of the entries of Z, T and V (that are random variables) is given by:

$$\frac{i^{n^2}}{(2\pi)^{n^2}} e^{-\mathrm{tr}(M^*M)} \bigwedge_{i,j} dm_{i,j} \wedge d\overline{m}_{i,j} = \frac{i^{n^2}}{(2\pi)^{n^2}} e^{-\mathrm{tr}(V(Z+T)^*(Z+T)V^*)} \bigwedge_{i,j} dm_{i,j} \wedge d\overline{m}_{i,j} =$$
$$\frac{i^{n^2}}{(2\pi)^{n^2}} e^{-\mathrm{tr}((Z+T)^*(Z+T))} \bigwedge_{i,j} dm_{i,j} \wedge d\overline{m}_{i,j} = \frac{i^{n^2}}{(2\pi)^{n^2}} e^{-\mathrm{tr}(Z^*Z)} e^{-\mathrm{tr}(T^*T)} \bigwedge_{i,j} dm_{i,j} \wedge d\overline{m}_{i,j}$$

where we have applied the properties of the trace, the fact that V is unitary and that Z, T are diagonal and strictly upper triangular matrices respectively.

With that in mind, to finish the development of the above expression, we will focus ourselves on the factor $\bigwedge_{i,j} dm_{i,j} \wedge d\overline{m}_{i,j}$, trying to write it as a function of the differential forms of V, T and Z. Indeed, to analyse it, considering the differential of M as a function of V, T, Z and applying the usual rules for differentiation, we can start by observing that:

$$\begin{aligned} dM &= d(V(Z+T)V^*) = (dV)(Z+T)V^* + Vd((Z+T)V^*) \\ &= (dV)(Z+T)V^* + V(dZ+dT)V^* + V(Z+T)dV^* \end{aligned}$$

Now, notice that V is a unitary matrix, that is, we have that $V^*V = I_n$ and, differentiating both sides of this equality, we can conclude that $d(V^*)V + V^*dV = 0 \iff d(V^*)V = -V^*dV \iff d(V^*) = -V^*(dV)V^*$. Thanks to this last expression for the differential $d(V^*)$, we can substitute it above to deduce that:

$$\begin{aligned} dM &= (dV)(Z+T)V^* + V(dZ+dT)V^* - V(Z+T)V^*(dV)V^* \\ &= V(V^*(dV)(Z+T) + dZ + dT - (Z+T)V^*dV)V^* \end{aligned}$$

that is a convenient expression for the differential of M. Then, introducing the notations $\Lambda = V^*(dM)V = (\lambda_{i,j})_{i,j}$, $\Omega = V^*dV = (\omega_{i,j})_{i,j}$ and $S = T + Z = (s_{i,j})_{i,j}$ (here we consider them as functions, not already evaluated in the corresponding random variables), we have managed to conclude that:

$$\Lambda = \Omega S - S\Omega + dS$$

As a comment, notice that we have just seen that $d(V^*)V + V^*dV = 0$, something that, combined with the above notation, implies that $d(V^*)V + V^*dV = 0 \iff (V^*dV)^* + V^*dV \iff \Omega^* + \Omega = 0 \iff \Omega = -\Omega^*$. That is, we have that $\omega_{i,j} = -\overline{\omega}_{j,i}$ for all $1 \leq i, j \leq n$.

Now, recall that our immediate objective was to compute the differential form $\bigwedge_{i,j} dm_{i,j} \wedge d\overline{m}_{i,j}$, that is given by the entries of dM seen as a function of V, Z, T. In addition, notice that V^* is a unitary matrix and that we also have that $\Lambda = V^*(dM)V$ and thus $dM = V\Lambda V^*$. So, to calculate it, with this last equality in mind, we can start introducing the new notation $K = \Lambda V^* = (k_{i,j})_{i,j}$ and we will proceed it in two steps.

Firstly, we can observe that $dM = VK = (Vk_{\cdot,1}|\dots|Vk_{\cdot,n})$ and then, for any $1 \leq j \leq n$ fixed, it is immediately clear that the map that sends the j-th column $k_{\cdot,j}$ to $Vk_{\cdot,j}$ is linear. That, combined with the fact that $\det(V)\overline{\det(V)} = \det(V)\det(V^*) = \det(VV^*) = \det(I_n) = 1$ because V is unitary, allows us to conclude that except maybe for a change of sign related to the order of the one-forms:

$$\begin{gathered}\textstyle\bigwedge_i dm_{i,j} = dm_{1,j} \wedge \cdots \wedge dm_{n,j} = (\sum_{l=1}^n v_{1,l}k_{l,j}) \wedge \cdots \wedge (\sum_{l=1}^n v_{n,l}k_{l,j}) = \\ \textstyle\sum_{\sigma \in S_n} v_{1,\sigma(1)}k_{\sigma(1),j} \wedge \cdots \wedge v_{n,\sigma(n)}k_{\sigma(n),j} = \sum_{\sigma \in S_n}(\prod_{i=1}^n v_{i,\sigma(i)})k_{\sigma(1),j} \wedge \cdots \wedge k_{\sigma(n),j} \\ \textstyle= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma)(\prod_{i=1}^n v_{i,\sigma(i)})k_{1,j} \wedge \cdots \wedge k_{n,j} = \det(V) \bigwedge_i k_{i,j}\end{gathered}$$

$$\begin{gathered}\textstyle\bigwedge_i dm_{i,j} \wedge d\overline{m}_{i,j} = (\bigwedge_i dm_{i,j}) \wedge (\bigwedge_i d\overline{m}_{i,j}) = \\ \textstyle\det(V)\overline{\det(V)}(\bigwedge_i k_{i,j}) \wedge (\bigwedge_i \overline{k}_{i,j}) = \bigwedge_i k_{i,j} \wedge \overline{k}_{i,j}\end{gathered}$$

From here, regarding the matrix K, recall that $K = \Lambda V^*$ and, as a consequence, it is clear that $K^* = V\Lambda^*$. Then, focusing now on the transformation from K to $K^* = (k^*_{i,j})_{i,j}$, observe that here we have only conjugated and transposed the matrix, something that allows us to conclude that, except maybe for change of sign related to the orientation that we can again disregard (because we are just trying to compute the corresponding joint density):

$$\bigwedge_{i,j} k_{i,j} \wedge \overline{k}_{i,j} = \bigwedge_{i,j} \overline{k}_{j,i} \wedge k_{j,i} = \bigwedge_{i,j} k^*_{i,j} \wedge \overline{k^*}_{i,j}$$

In fact, not only that, but notice that the same argument also applies for the transformation from Λ to Λ^*.

Finally, using again that $K^* = V\Lambda^* = (V\lambda^*_{\cdot,1}|\dots|V\lambda^*_{\cdot,n})$ is a linear transformation in each of the columns and that $|\det(V)| = 1$, we can conclude as before that by the

change of variable formula and for any $1 \leq j \leq n$ we have that:

$$\bigwedge_i k^*_{i,j} \wedge \overline{k^*}_{i,j} = \bigwedge_i \lambda^*_{i,j} \wedge \overline{\lambda^*}_{i,j}$$

That being said, combining all of the above, at the end we can deduce that except maybe for a change of sign related to the orientation that we will ignore for good from now on, because it is not relevant for the computation of the joint density, we have that:

$$\begin{gathered}\textstyle\bigwedge_{i,j} dm_{i,j} \wedge d\overline{m}_{i,j} = (\bigwedge_i dm_{i,1} \wedge d\overline{m}_{i,1}) \wedge \cdots \wedge (\bigwedge_i dm_{i,n} \wedge d\overline{m}_{i,n}) = \\ (\textstyle\bigwedge_i k_{i,1} \wedge \overline{k}_{i,1}) \wedge \cdots \wedge (\bigwedge_i k_{i,n} \wedge \overline{k}_{i,n}) = \\ \textstyle\bigwedge_{i,j} k_{i,j} \wedge \overline{k}_{i,j} = \bigwedge_{i,j} k^*_{i,j} \wedge \overline{k^*}_{i,j} = \\ (\textstyle\bigwedge_i k^*_{i,1} \wedge \overline{k^*}_{i,1}) \wedge \cdots \wedge (\bigwedge_i k^*_{i,n} \wedge \overline{k^*}_{i,n}) = \\ (\textstyle\bigwedge_i \lambda^*_{i,1} \wedge \overline{\lambda^*}_{i,1}) \wedge \cdots \wedge (\bigwedge_i \lambda^*_{i,n} \wedge \overline{\lambda^*}_{i,n}) = \\ \textstyle\bigwedge_{i,j} \lambda^*_{i,j} \wedge \overline{\lambda^*}_{i,j} = \bigwedge_{i,j} \lambda_{i,j} \wedge \overline{\lambda}_{i,j}\end{gathered}$$

that is exactly what we wanted to see.

So, actually, we have managed to deduce that $\bigwedge_{i,j} dm_{i,j} \wedge d\overline{m}_{i,j} = \bigwedge_{i,j} \lambda_{i,j} \wedge \overline{\lambda}_{i,j}$, something that allows us to focus on these forms $\lambda_{i,j}$. Then, given any $i, j \in \{1, \ldots, n\}$, we can use the equality $\Lambda = \Omega S - S\Omega + dS$ and, combining it with the definition of the product of matrices and the fact that S and thus dS are upper-triangular matrices, we can observe that:

$$\lambda_{i,j} = \sum_{k=1}^{j} s_{k,j}\omega_{i,k} - \sum_{k=i}^{n} s_{i,k}\omega_{k,j} + ds_{i,j} =$$

$$\begin{cases} (s_{j,j} - s_{i,i})\omega_{i,j} + \left[\sum_{k=1}^{j-1} s_{k,j}\omega_{i,k} - \sum_{k=i+1}^{n} s_{i,k}\omega_{k,j}\right] & \text{if } i > j \\ ds_{i,j} + s_{i,j}(\omega_{i,i} - \omega_{j,j}) + \left[\sum_{k=1,\, k\neq i}^{j} s_{k,j}\omega_{i,k} - \sum_{k=i,\, k\neq j}^{n} s_{i,k}\omega_{k,j}\right] & \text{if } i \leq j \end{cases}$$

That being said, notice that our objective is to calculate the product $\bigwedge_{i,j} \lambda_{i,j} \wedge \overline{\lambda}_{i,j}$. To do it, we can start by considering the forms $\lambda_{i,j}$ in the following order:

$$\lambda_{n,1}, \overline{\lambda}_{n,1}, \ldots, \lambda_{n,n}, \overline{\lambda}_{n,n}, \lambda_{n-1,1}, \overline{\lambda}_{n-1,1}, \ldots, \lambda_{n-1,n}, \overline{\lambda}_{n-1,n}, \ldots, \lambda_{1,1}, \overline{\lambda}_{1,1}, \ldots, \lambda_{1,n}, \overline{\lambda}_{1,n}$$

where one can notice that, if we fix $\lambda_{i,j}$ for some $1 \leq i, j \leq n$, the following terms have already appeared before: $\lambda_{a,b}$ and $\overline{\lambda}_{a,b}$ with $b < a$ and $i < a$ or, alternatively, with $b < a$, $i = a$ and $b < j$.

Now, with this order in mind, we can observe that the one-forms that are present inside the factors in brackets (the corresponding $\omega_{\cdot,\cdot}$ and $\overline{\omega}_{\cdot,\cdot}$) have already appeared before outside previous brackets (in the expressions $(s_{j,j} - s_{i,i})\omega_{i,j}$ or their conjugates), something that we can see distinguishing two cases. Indeed, given any $1 \leq i, j \leq n$, we have that:

Case 1: On the one hand, if $i \geq j$, in the expression inside the brackets we can find the one-forms $\omega_{i,1}, \ldots, \omega_{i,j-1}$ and $\omega_{i+1,j}, \ldots, \omega_{n,j}$ (observe that this also happens when

$i = j$). Nevertheless, notice that all of these forms are given by $\omega_{a,b}$ with $a = i$, $b < j$ and $b < a$ in the first ones, or by $\omega_{a,b}$ with $i < a$ and $b < a$ in the second ones. So, we can conclude that all of them have already appeared before in the other factor $(s_{b,b}-s_{a,a})\omega_{a,b}$ for a previous $\lambda_{a,b}$. Not only that, but using the same argument and conjugating all the previous expressions, we can observe that the one-forms $\overline{\omega}_{i,1}, \dots, \overline{\omega}_{i,j-1}$ and $\overline{\omega}_{i+1,j}, \dots, \overline{\omega}_{n,j}$ have also appeared before. That completes this case.

Case 2: On the other hand, if $i < j$, in the expression inside the brackets we can find the one-forms $\omega_{i,1}, \dots, \omega_{i,i-1}, \omega_{i,i+1}, \dots, \omega_{i,j}$ and $\omega_{i,j}, \dots, \omega_{j-1,j}, \omega_{j+1,j}, \dots, \omega_{n,j}$ and now we distinguish four more cases:

(1) Regarding the one-forms $\omega_{i,1}, \dots, \omega_{i,i-1}$, here it is clear that all of them are given by $\omega_{a,b}$ with $a = i$, $b < i$ and $b < a$, so they have already appeared in the factor $(s_{b,b} - s_{a,a})\omega_{a,b}$ for a previous $\lambda_{a,b}$.
(2) Regarding the one-forms $\omega_{i,i+1}, \dots, \omega_{i,j}$, using that we already know that $\Omega = -\Omega^*$, we can observe that they are (multiples of) $\overline{\omega}_{i+1,i}, \dots, \overline{\omega}_{j,i}$. Then, notice that all of these last forms are given by $\overline{\omega}_{a,b}$ with $i < a$ and $b < a$ and, as a consequence, they have already appeared in the factor $(\overline{s}_{b,b} - \overline{s}_{a,a})\overline{\omega}_{a,b}$ for a previous $\overline{\lambda}_{a,b}$.
(3) Regarding the one-forms $\omega_{i,j}, \dots, \omega_{j-1,j}$, using the same argument as above, we can deduce that they are (multiples of) $\overline{\omega}_{j,i}, \dots, \overline{\omega}_{j,j-1}$. That, combined with the fact that these last forms are given by $\overline{\omega}_{a,b}$ with $i < a$ and $b < a$, allows us to conclude that they have already appeared in the factor $(\overline{s}_{b,b} - \overline{s}_{a,a})\overline{\omega}_{a,b}$ for a previous $\overline{\lambda}_{a,b}$.
(4) Regarding the one-forms $\omega_{j+1,j}, \dots, \omega_{n,j}$, we can easily observe that all of them are given by $\omega_{a,b}$ with $i < a$ and $b < a$, so they have already appeared in the factor $(s_{b,b} - s_{a,a})\omega_{a,b}$ for a previous $\lambda_{a,b}$.

That is what we wanted to see. Then, applying a completely analogous argument for their conjugates, we can also conclude that the forms $\overline{\omega}_{i,1}, \dots, \overline{\omega}_{i,i-1}, \overline{\omega}_{i,i+1}, \dots, \overline{\omega}_{i,j}$ and $\overline{\omega}_{i,j}, \dots, \overline{\omega}_{j-1,j}, \overline{\omega}_{j+1,j}, \dots, \overline{\omega}_{n,j}$ have already appeared before, something that completes this last case.

So, at the end and thanks to these observations, when computing the wedge product $\bigwedge_{i,j} \lambda_{i,j} \wedge \overline{\lambda}_{i,j}$ using its properties (notice that if $a, b_1, \dots, b_h$ are one-forms and $u, d_0, \dots, d_h \in \mathbb{C}$, then $ua \wedge (d_0 a + \sum_{i=1}^{h} d_i b_i) = ud_0(a \wedge a) + ua \wedge \sum_{i=1}^{h} d_i b_i = ua \wedge \sum_{i=1}^{h} d_i b_i$), we can disregard these factors inside the brackets, something that allows us to conclude that:

$$\bigwedge_{i,j} \lambda_{i,j} \wedge \overline{\lambda}_{i,j} = \left(\prod_{i>j} |z_i - z_j|^2\right) \bigwedge_i (dz_i \wedge d\overline{z}_i) \bigwedge_{i>j} (\omega_{i,j} \wedge \overline{\omega}_{i,j}) \times \bigwedge_{i<j} (dt_{i,j} + t_{i,j}(\omega_{i,i} - \omega_{j,j}) \wedge \overline{dt_{i,j} + t_{i,j}(\omega_{i,i} - \omega_{j,j})})$$

Finally, combining all of the above, note that we have actually managed to conclude that the joint density of the entries of V (expressed through Ω), T and $Z =$

$\mathrm{diag}(Z_1,\dots,Z_n)$ is given by:

$$\frac{i^{n^2}}{(2\pi)^{n^2}} e^{-\mathrm{tr}(Z^*Z)} (\prod_{i>j} |z_i - z_j|^2) \bigwedge_i (dz_i \wedge d\overline{z}_i) \times$$
$$e^{-\mathrm{tr}(T^*T)} \bigwedge_{i>j} (\omega_{i,j} \wedge \overline{\omega}_{i,j}) \bigwedge_{i<j} (dt_{i,j} + t_{i,j}(\omega_{i,i} - \omega_{j,j}) \wedge \overline{dt_{i,j} + t_{i,j}(\omega_{i,i} - \omega_{j,j})})$$

Nevertheless, as we are only interested in the eigenvalues of our original matrix M, $(Z_1,\dots,Z_n)$, now we focus only the corresponding marginal of the previous joint density. Indeed, using that the above joint density is a product of a factor depending on $z_1,\dots,z_n$ and another that does not depend on $z_1,\dots,z_n$, then it is immediate to compute this marginal and conclude that the joint density of the eigenvalues $(Z_1,\dots,Z_n)$ is proportional to:

$$e^{-\mathrm{tr}(Z^*Z)} (\prod_{i>j} |z_i - z_j|^2) \bigwedge_i (dz_i \wedge d\overline{z}_i) = e^{-\sum_{k=1}^n |z_k|^2} (\prod_{i>j} |z_i - z_j|^2) \bigwedge_i (dz_i \wedge d\overline{z}_i)$$

So, now, to complete the proof of this theorem, taking into account that we are considering our space $(\mathbb{C},\mu)$ with respect to the measure $d\mu(z) = \frac{1}{\pi} e^{-|z|^2} dm(z)$, it suffices to see that:

$$\det \mathbb{K}(z_i, z_j)_{i,j\leq n} \propto \prod_{i>j} |z_i - z_j|^2$$

and, to do it, we will compute this determinant.

Indeed, applying the well-known properties of the Vandermonde matrices and their determinant, we can easily see that for any $z_1,\dots,z_n \in \mathbb{C}$, we have that:

$$\prod_{i>j} |z_i - z_j|^2 = \prod_{i>j} (z_i - z_j) \prod_{i>j} (\overline{z_i} - \overline{z_j}) =$$
$$\begin{vmatrix} 1 & z_1 & \dots & z_1^{n-1} \\ 1 & z_2 & \dots & z_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & z_n & \dots & z_n^{n-1} \end{vmatrix} \begin{vmatrix} 1 & \overline{z}_1 & \dots & \overline{z}_1^{n-1} \\ 1 & \overline{z}_2 & \dots & \overline{z}_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \overline{z}_n & \dots & \overline{z}_n^{n-1} \end{vmatrix} = \begin{vmatrix} 1 & z_1 & \dots & z_1^{n-1} \\ 1 & z_2 & \dots & z_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & z_n & \dots & z_n^{n-1} \end{vmatrix} \begin{vmatrix} 1 & 1 & \dots & 1 \\ \overline{z}_1 & \overline{z}_2 & \dots & \overline{z}_n \\ \vdots & \vdots & \ddots & \vdots \\ \overline{z}_1^{n-1} & \overline{z}_2^{n-1} & \dots & \overline{z}_n^{n-1} \end{vmatrix}$$

$\propto$ (using that the product of determinants is the determinant of the product) $\propto$

$$\begin{vmatrix} \mathbb{K}(z_1,z_1) & \mathbb{K}(z_1,z_2) & \dots & \mathbb{K}(z_1,z_n) \\ \mathbb{K}(z_2,z_1) & \mathbb{K}(z_2,z_2) & \dots & \mathbb{K}(z_2,z_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{K}(z_n,z_1) & \mathbb{K}(z_n,z_2) & \dots & \mathbb{K}(z_n,z_n) \end{vmatrix} = \det \mathbb{K}(z_i,z_j)_{i,j\leq n}$$

That is what we wanted to see, something that combined with all of the above, allows us to conclude that the eigenvalues of M, that is, $Z_1,\dots,Z_n$, form a determinantal point process $\mathscr{X} = \sum_{i=1}^n \delta_{Z_i}$ in $(\mathbb{C},\mu)$ with kernel $\mathbb{K}$. That completes the proof. □

7.2. **Gaussian unitary ensemble.** Nevertheless, the Ginibre ensemble is not the only determinantal point process that naturally comes from the field of Random Matrix Theory. Another very well-known example of a DPP is the Gaussian unitary ensemble, that can be obtained by considering the eigenvalues of certain Hermitian random matrices.

These point processes were introduced and studied by Wigner before the general notion of DPPs appeared, with the aim of understanding and providing a statistical framework for the behaviour of the energy levels of heavy atomic nuclei, like the ones of the uranium. In what follows, we precise in which sense these sets of eigenvalues can be seen as a determinantal point process.

Theorem 7.2. *Let $M \in \mathbb{C}^{n\times n}$ be a random matrix of size $n \geq 1$ with independent and identically distributed standard complex Gaussian entries, we consider the matrix $A = \frac{M+M^*}{\sqrt{2}}$ and the point process $\mathscr{X} = \sum_{i=1}^{n} \delta_{X_i}$ formed by $X_1, \ldots, X_n$ the eigenvalues of A. Then, we have that $\mathscr{X}$ is a determinantal point process in $\mathbb{R}$ with kernel $\mathbb{K} : \mathbb{R}^2 \longrightarrow \mathbb{R}$ given by:*

$$\mathbb{K}(x,y) = \sum_{k=0}^{n-1} H_k(x)H_k(y) \quad \text{for all } x, y \in \mathbb{R}$$

with respect to the background measure μ with $d\mu(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$ on $\mathbb{R}$, where the functions H_k are the Hermite polynomials, that can be obtained by applying the Gram-Schmidt orthogonalization scheme to $\{1, x, \ldots, x^{n-1}\}$ in $L^2(\mathbb{R}, \mu)$.

The proof of this result, as was the case with the Ginibre ensemble, is long, relatively complex and can be found in our reference [11]. So, since we already have the example of one of such proofs, to avoid extending this project too much, we have decided not to include more proofs that certain specific examples are indeed determinantal point processes.

7.3. **Spherical ensemble.** As yet another example of a determinantal point process coming from random matrices, we can present the Spherical ensemble, that can be obtained and sampled by taking the eigenvalues of certain products of random matrices. These processes, although originally introduced in the context of DPPs and random matrices by Krishnapur, were already known in the field of physics, since they appear when one considers certain models for the behaviour of plasma in a sphere. Notice that these processes, as well as previous examples of DPPs, show that there is a strong connection between physics and these family of point process. This is one of the reasons that explains the interest and the study of DPPs.

Theorem 7.3. *Let $A, B \in \mathbb{C}^{n\times n}$ be two independent random matrices of size $n \geq 1$ with independent and identically distributed standard complex Gaussian entries, we consider the matrix $M = A^{-1}B$ and the point process $\mathscr{X} = \sum_{i=1}^{n} \delta_{Z_i}$ formed by $Z_1, \ldots, Z_n$ the eigenvalues of M. Then, we have that $\mathscr{X}$ is a determinantal point process in $\mathbb{C}$ with kernel $\mathbb{K} : \mathbb{C}^2 \longrightarrow \mathbb{C}$ given by:*

$$\mathbb{K}(z,w) = (1 + z\overline{w})^{n-1} \quad \text{for all } z, w \in \mathbb{C}$$

with respect to the background measure μ with $d\mu(z) = \frac{n}{\pi(1+|z|^2)^{n+1}} dm(z)$, where m denotes the usual Lebesgue measure in $\mathbb{C}$.

7.4. **Sine kernel process.** On the other hand, it is also worth mentioning that not all relevant DPPs are naturally described by taking the eigenvalues of certain random matrices. Indeed, for example, we can focus our attention now on the Sine kernel process, which is a translation invariant process in $\mathbb{R}$ very related to the usual Fourier transform.

Theorem 7.4. *Let $\mathbb{K} : \mathbb{R}^2 \longrightarrow \mathbb{R}$ be the kernel given by:*

$$\mathbb{K}(x,y) = \frac{\sin(\pi(x-y))}{\pi(x-y)}$$

for all $x, y \in \mathbb{R}$ (if $x = y$, we define $\mathbb{K}(x,x) = 1$). Then, there exists a determinantal point process in $\mathbb{R}$ with kernel $\mathbb{K}$, that we call the Sine kernel process.

Its relation with the Fourier theory comes from the fact that $\mathscr{K}$, that is, the associated integral operator given by $\mathbb{K}$, can be seen as the projection operator in the space of functions of $L^2(\mathbb{R})$ whose Fourier transform is supported in $[-\pi, \pi]$. To show this connection, we merely need to recall basic Fourier results and then, given any $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ and almost any $x \in \mathbb{R}$, we can observe that:

$$\mathscr{K}f(x) = \int_{\mathbb{R}} \mathbb{K}(x,y)f(y)dy = \int_{\mathbb{R}} \tfrac{\sin(\pi(x-y))}{\pi(x-y)} f(y)dy = \int_{\mathbb{R}} \left(\tfrac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(x-y)u} du \right) f(y)dy$$

$$= \tfrac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} e^{ixu} \left(\tfrac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(y) e^{-iyu} dy \right) du = \tfrac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} e^{ixu} \hat{f}(u) du =$$

$$\tfrac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathbb{1}_{[-\pi,\pi]}(u) \hat{f}(u) e^{ixu} du =$$

$$\left(\mathbb{1}_{[-\pi,\pi]} \hat{f}\right)^{\check{}}(x)$$

where we have used Fubini's theorem. Then, this behaviour can be extended to all functions in $L^2(\mathbb{R})$ thanks to the density of $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$.

Notice that, in particular, this tells us that $\mathscr{K}$ is a projection operator. So thanks to Theorem 4.4, our definition of a DPP and the fact that $\mathbb{K}$ is clearly a locally square integrable and Hermitian kernel (because it is continuous and real-valued), we can conclude that, indeed, the above kernel $\mathbb{K}$ gives rise to a DPP.

7.5. **Zero set of a hyperbolic GAF.** Finally, as another natural example of a DPP that is not closely related to Random Matrix Theory, we can consider the zeros of a certain (hyperbolic) GAF. Recall that, without entering into detail, a Gaussian analytic function (GAF) f on a region $E \subseteq \mathbb{C}$ is a random variable taking values in the space of analytic functions on E such that, for all $n \geq 1$ and $z_1, \ldots, z_n \in E$, we have that $(f(z_1), \ldots, f(z_n))$ has a complex Gaussian distribution with mean zero.

With that in mind, given a GAF, notice that its set of zeros forms a point process in E. These processes coming from the zeros of GAF constitute another interesting and well-studied family of point processes, different from DPPs, but that also exhibit some repulsive properties and that can be used to model certain physical phenomena.

Nevertheless, although in general these sets of random zeros are not DPPs, there is a particular process where this is exactly the case, that can be introduced as follows.

Theorem 7.5. *Let $\{a_n\}_{n=0}^{\infty}$ be independent and identically distributed standard complex Gaussian variables and let $f : \mathbb{D} \longrightarrow \mathbb{C}$ be the Gaussian analytic function given by $f(z) = \sum_{n=0}^{\infty} a_n z^n$ for all $z \in \mathbb{D}$, we consider the point process $\mathscr{X} = \sum_{i=1}^{n} \delta_{Z_i}$ formed by $Z_1, \ldots, Z_n$ the zeros of f. Then, we have that $\mathscr{X}$ is a determinantal point process in the complex disk $\mathbb{D}$ with kernel $\mathbb{K} : \mathbb{D}^2 \longrightarrow \mathbb{C}$ given by:*

$$\mathbb{K}(z, w) = \frac{1}{\pi(1 - z\overline{w})^2} \qquad \textit{for all } z, w \in \mathbb{D}$$

with respect to the usual Lebesgue measure in $\mathbb{D}$.

As a comment, it is worth mentioning that the above kernel $\mathbb{K}$ is the Bergman kernel of $\mathbb{D}$, that is, its associated integral operator $\mathscr{K}$ is the projection operator from the (measure-normalized) $L^2(\mathbb{D}, \pi^{-1}dm)$ space onto the subspace of holomorphic functions.

8. Applications of DPPs in modelling and machine learning

Finally, it is also worth mentioning that, apart from their theoretical interest that we have already explored, determinantal point processes are also being used in practical applications, especially in the fields of modelling and machine learning, where their repulsion properties are useful to describe certain interactions between particles and to ensure diversity in the predictions.

8.1. **Applications in machine learning.** Indeed, focusing first on how determinantal point processes are being employed in the field of machine learning, their role is to ensure that there is diversity among the outputs of the predictive models. So, by using one of these processes, our objective is to encourage the system not to show only the most relevant prediction, but to also provide a variety of results among the ones available that cover multiple aspects of the user's query.

It is worth mentioning that DPPs are not the only type of point processes that can fill this role in the machine learning pipeline, since there other well-known families of processes that are also able to describe and produce a wide variety of repelling behaviours. For example, one of these alternatives is to use Gibbs point processes instead, that can in fact model stronger repulsion interactions than the ones covered by determinantal processes. Nevertheless, one of the reasons that explains the popularity of DPPs in these tasks is their computational tractability and efficiency, since there are many fast algorithms that have been developed for them. Indeed, DPPs are remarkably easy to fit and simulate, something that combined with their nice descriptions and mathematical properties, makes them a popular choice to achieve this diversity.

That being said, aiming to build our intuition on this topic and without entering into too much detail, now we will cover a few machine learning tasks where DPPs have been used, that can be found explained in Sections 5.3 and 6.6.4 from [13] and in [26].

Text summarization: On the one hand, as a first example, we can consider an extractive multi-document summarization task. In this problem, given a corpus of several related text documents, the objective is to select a small number of sentences from them that, together, provide a complete summary of the topics and contents discussed in the corpus. To solve this challenge, there are multiple suitable machine learning techniques that could be used but, when building these summaries, a common issue that most of them face is that the predicted summary sentences lack diversity, since all of them end up revolving around the same specific idea that the system found to be most relevant.

To avoid this situation, we can fit and sample a DPP instead, whose points tend to exhibit repulsive interactions. In our case, that translates into our highlighted sentences being different and thus covering different topics, aspects or ideas found in the corpus, something that is desirable if we aim to build these limited but insightful summaries.

To do it, introducing some notation, we consider X a given corpus of text and then let $E(X)$ be the set formed by all its sentences. This is the space, depending on X, where we will define our DPP and thus, a sample of the DPP could be seen as a possible

summary of the corpus X, made by some of its sentences. As a comment, notice that clearly $E(X)$ is a finite set, so we can equip it with the usual topology and counting measure to have a Polish space. From here, before introducing the DPP model and with the aim of training it later, we also need training data, $\{X^{(t)}, Y^{(t)}\}_{t=1}^{T}$, formed by many of these corpuses and their corresponding human-made extractive summaries Y, that will serve as our labels.

Then, we can consider our DPP model, whose kernel can be represented by a certain symmetric and positive semi-definite matrix L, as we are in a finite setting. In fact, as expected, these processes depend on the particular corpus X we are trying to summarise and thus, the entries of this matrix are a function of the corpus as well as of a certain set of parameters θ, in such a way that we consider that:

$$P(Y \mid X) \propto \det(L_Y(X, \theta))$$

where $L_Y(X, \theta)$ denotes the corresponding principal minor of $L(X, \theta)$, determined by the possible summary Y. From here, using our data to find the optimal value of θ (for example, its maximum likelihood estimator), we can finally obtain our DPP model ready to be used in inference.

Image diversity: On the other hand, as an example of another machine learning system where the properties of DPPs are desirable, we have the Image search engines. In these automated systems, given a query from a user, the objective is to provide them with a set of images that describe and represent said query. Nevertheless, in these situations, the queries are usually too broad and general, in such a way that it is difficult to know what the user really wanted to see. So, to avoid this problem, we try to offer a set of images that are not only relevant to the query but also diverse, aiming to maximize the chances that at least one of the proposed images corresponds to what the user was looking for.

With that in mind, to achieve this diversity among the produced images, again, a good option is to fit and sample a DPP to select them. Here, the idea is that the repulsion behaviour that points of these processes tend to exhibit will make our model more inclined to choose diverse sets of images.

That being said, to actually train and test our model with the aim of showing how suited DPPs are for this problem, one can consider the following task. Given a certain user query and a set of images X relevant to said query, we select a subset $Y_t \subseteq X$ and two additional images $i_t^+, i_t^- \in X$ obtained at random. From here, we can employ humans to decide which of the two extra images, i_t^+, is the one that produces a more diverse set together with Y_t. Then, we would like our DPP to assign higher probability to the set $Y_t \cup \{i_t^+\}$ than to $Y_t \cup \{i_t^-\}$. Repeating this process many times, we can build our dataset $\{(Y_t, i_t^+, i_t^-)\}_{t=1}^{T}$ necessary to fit our algorithm.

From here, with this data, now we are ready to train our DPP model. In this case, we define the DPP as a point process in X, which can be seen as a Polish space with the usual topology and counting measure, in such a way that its samples are sets of relevant images. In addition, as before, we describe our DPP by using a symmetric and positive semi-definite matrix L, depending on features computed from the images

in X and certain parameters θ, so that, if $|Y_t| = k$ for all $1 \leq t \leq T$, then we have that:

$$P_L^{k+1}(Y_t \cup \{i\}) \coloneqq P(Y_t \cup \{i\} \subseteq \mathscr{X} \mid |\mathscr{X}| = k+1) = \frac{\det(L_Y(X,\theta))}{\sum_{|Y'|=k+1} \det(L_{Y'}(X,\theta))}$$

for any $i \in \{i_t^+, i_t^-\}$. Notice that in this task we are interested in obtaining a DPP that produces sets of a fixed size $k+1$. As already explained through this project, this could be achieved by using a projection DPP but, in practice, it has been shown to be more flexible and convenient to just use a general DPP and condition it to sampling a set of a prefixed cardinality, as done above. Finally, with our dataset and working with these probabilities, we can find the optimal value of the parameters and end up with our fitted DPP model ready to be used.

Recommender systems: Finally, it is also worth explaining the role that DPPs have in recommender systems. In this problem, the objective is to provide the user with an ordered set of item proposals with which we hope the user will interact. To achieve this goal, for each item i in our catalogue, traditional methods focus on computing a measure of how relevant is the item to the user, q_i, and thus, as a side effect, similar items tend to get similar quality metrics. Consequently, a common issue that these systems have is that the top proposals given by the algorithm are too similar, something that may not be appealing to the user. This happens because these quality metrics q_i are computed independently for each product.

With that in mind, to solve this problem, it is possible to use heuristic and rule-based schemes that try to increase the diversity of the recommended items, but these practices usually decrease the overall quality of the proposals, since they are too rigid and it is difficult to find a balance between quality and diversity. Achieving this balance is exactly the role of DPPs, which can assign higher probabilities to highly relevant items while also naturally producing diverse sets thanks to their repulsion properties.

So, the idea is to attach the DPP to the last layer of the recommending system, using the already computed quality metrics q_i to build its kernel. In addition, the kernel of the DPP also uses some given similarity scores between pairs of items $D_{i,j}$, which are usually computed by another machine learning model. From here, given a pre-selected set of items prepared for a user, X, we can use the set of items that the user actually interacted with in the past, $Y \subseteq X$, as our labels, building our training data with these pairs $\{X^{(t)}, Y^{(t)}\}_{t=1}^T$.

That being said, we are ready to introduce our DPP, which is taken as a point process in the finite and thus Polish space X, with its samples being sets of relevant items. Then, the DPP model is described, again, by a symmetric and positive semi-definite matrix L, depending on the corresponding q_i, $D_{i,j}$ and a set of parameters θ (so we have $L \coloneqq L(X,\theta)$), in such a way that:

$$P(Y \mid X) \propto \det(L_Y(X,\theta))$$

In this case, during training, we try to maximize the above probabilities for the pairs $(X^{(t)}, Y^{(t)})$ to find the optimal values of the parameters and then, during inference, we can build the ordered set of recommendations recursively as follows. If Y is the current

incomplete list, the next element v_{next} added to the list is the one that maximizes:

$$\max_{v \in X \setminus Y} \det(L_{Y \cup \{v\}}(X, \theta^*))$$

8.2. **Applications in modelling.** On the other hand, regarding the applications of DPPs in modelling (analysed for example in [15]), here their task is to describe and represent spatial point patterns that, even though are random, they also exhibit repulsion behaviours between their points. These models, based on DPPs, can be very useful when trying to analyse these patterns, aiming for example to understand the underlying cause for the repulsion and allowing us to extract conclusions and to compute relevant metrics. Furthermore, thanks to these models, we can also generate synthetic data for further experiments as well as to predict likely positions for future particles.

In addition, as already mentioned in the machine learning section, recall DPPs can be efficiently fitted and simulated and they also benefit from very attractive mathematical properties. So, even though one can find alternative models to describe these patterns, these properties make DPPs a popular choice for this type of tasks compared to other repulsive point processes, which are more difficult to analyse, both theoretically and computationally.

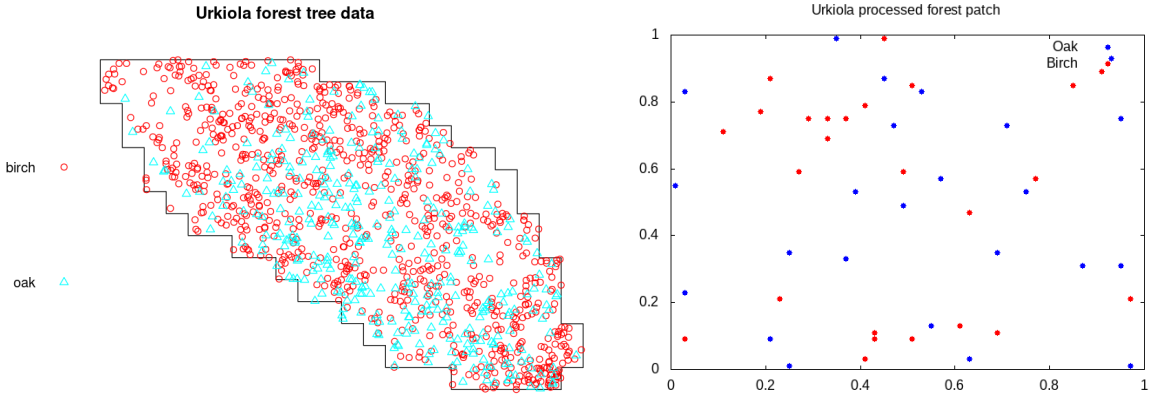
There are many phenomena, natural or not, that can be analysed and studied from this point of view, as random points or locations that are still subject to a certain kind repulsive structure. For example, in the literature, one can find instances of DPPs being used to model the behaviour of certain quantum particles, the disposition of cities and towns or the locations of base stations in communication networks. Moreover, in the field of ecology, DPPs have also been used to describe the positions of trees and of animal nests.

In relation to this last point and with the aim of exemplifying the above, now we will develop our own small toy application (inspired by the ones shown in [15]) to the task of modelling the position of trees in a forest. Indeed, to do it, we have started by considering data coming from the Urkiola Natural Park forest, which is located in the Basque Country (Spain). This data, that can be found in [14], describes the spatial positions of two different species of trees that live in the forest, oak and birch trees.

Notice that, as already mentioned, it makes sense to use a DPP model to describe the spatial locations of trees, as the ones that can be found in the Urkiola forest. This is the case because, even though the positions of trees have a random component, tied for example to how the wind dispersed their seeds, notice that the light, water and space resources among others are limited in the forest. As a consequence, this competition for resources limits the amount of trees that can grow close together at any given location, which, from our point of view, makes the tree locations exhibit a repulsive behaviour. This repulsion is exactly the one that motivates the use of DPPs in this context and explains why they can be a good modelling choice.

With that in mind, to develop our application, first of all we have started by preprocessing our dataset, since it contains too many tree locations in a region that is also too irregular, something that makes it difficult to model using our point processes. So,

to solve these problems, we have decided to focus only on a smaller square patch found inside the forest, that has sides of 25 meters and contains a total of 49 trees. After selecting the patch, we have normalized it to have all our points in the unit square $[0,1]^2$. This step does not change the spatial patterns and behaviours exhibited by the trees, so we have just done it for convenience purposes. Finally, to make the task of fitting our determinantal models easier, we have also decided discretize the unit square, considering it as a 50×50 grid and relocating the trees to the closest valid position in it. Thanks to this last transformation, we will be able to work in a finite setting, something that allows us to use faster algorithms to fit and sample our DPPs and puts us under the hypothesis of Proposition 6.5.



(A) Urkiola Natural Park forest dataset. (B) Extracted and discretized forest patch.

FIGURE 4. Observed tree location data.

From here, once our data was ready, we have used it to fit and analyse three different point process models, that aim to replicate the spatial behaviour of the locations of the trees. In what follows, we precise which are the three processes we have compared.

Gaussian determinantal point process: On the one hand, as expected, we have fitted a DPP to the observed tree location pattern and simulated it. In particular, we have chosen a Gaussian DPP, whose kernel $\mathbb{K} : [0,1]^2 \longrightarrow \mathbb{R}$ is given by the expression:

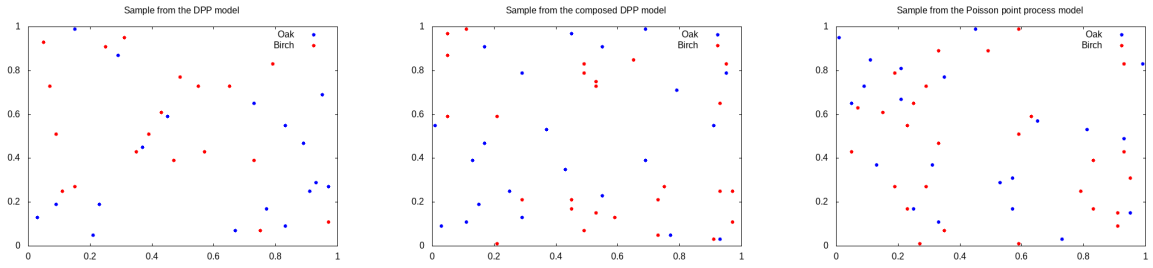
$$\mathbb{K}(x,y) = ae^{-\frac{\|x-y\|_2^2}{2\sigma^2}} \qquad \forall (x,y) \in [0,1]^2 \text{ that are valid points of the grid}$$

for suitable values of the parameters $0 \leq a \leq 1$ and $\sigma \geq 0$. This is a simple kernel that usually appears when working with DPPs in practical applications, because it has only two parameters, something that avoids overfitting the data, but it still allows us to use a to control the number of points observed and σ to modulate the strength of the repulsive interactions. To fit this process, we have tried to find the maximum likelihood estimators (MLEs) of the parameters, that is, the pair of parameters that makes more likely sampling the observed data. To do it, we have applied two approximate optimization algorithms (Montecarlo and Particle Swarm optimization), choosing at the end the most likely pair of parameters. From here, once we had fitted our model, it is worth mentioning that to sample it we have not used the general Algorithm 1

explained in this project, because in the finite setting there are simpler alternatives. So, in practice, we have ended up applying a variant of the Gaussian elimination algorithm, that is, the one that is used to compute the LU factorization of a matrix. Finally, to decide the species of each tree we sampled, we just assigned them randomly with equal probability.

Composition of Gaussian determinantal point processes: On the other hand, as an example of the theory developed in Subsection 6.2, we have decided to also fit a combined DPP to the provided data. This point process is defined by first sampling a Gaussian DPP (in fact, the same one introduced above) and then considering another Gaussian DPP defined in the points we have just sampled, that we use to decide the species of the trees. The intuition behind this combination is that, maybe, the repulsion exhibited by one of the tree species is different from the other, something that this richer model could better describe. Indeed, at a first glance, this difference in repulsions seems to be present in our data shown in Figure 4, where the birch trees appear to be more clustered. That being said and from a technical perspective, to fit and sample this combined process, we have essentially used the same techniques as in the previous case with the appropriate adaptations.

Poisson point process: Finally, and with the aim of putting our previous models into perspective, we have also decided to use a homogenous Poisson point process to describe our tree location data. This is a very simple and well-studied process that does not naturally favour repulsion interactions, something that makes it the perfect choice for deciding if modelling repulsion is actually useful in this task. In this case, recall that this type of process only depends on one parameter λ, that controls the random amount of points Poiss(λ) that the process has in the patch. Since it is known that the MLE of a Poisson distribution is the sample mean (in this case it is just the number of trees observed), it has been very easy to fit the model, since we have not had the need to use optimization algorithms. From here, once fitted, to simulate this process, we have just sampled first a Poisson random variable of parameter λ to decide the number of trees and then we have placed them uniformly and without replacement in the $50 \times 50 = 2500$ positions available in the grid, choosing the species at random.



(A) Gaussian DPP. (B) Composition of DPPs. (C) Poisson point process.

FIGURE 5. Examples sampled from our three tree location models.

That being said, once we have fitted each of our three models, we have sampled them 99 times, something that can be used to measure how well these models describe

the tree locations. Before moving on, to build our intuition, notice that one of these simulations is shown above, in Figure 5.

From here, to analyse our samples from a statistical point of view, we have considered the K statistic, given by:

$$K(r) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} e_{i,j} \mathbb{1}_{\|x_i - x_j\| \leq r} \qquad \forall r > 0 \text{ suitable}$$

where n denotes the number of points $x_1, \ldots, x_n \in [0,1]^2$ sampled and $e_{i,j}$ are certain weights that compensate for the fact that some points are close to the edges of the square. The task of this function is to summarize the spatial patterns shown in one sample, something that is a necessary step if we want to compare them. Finally, to perform this comparison, we have computed the mean absolute deviation between the K statistic of our observed data and the ones of our model samples. In addition, with the values of these statistics, we have also found a p-value for our models, that is, a value that gives us an idea of what would have been the probability of observing our actual data if the point process model was true. In simple terms and for comparison purposes, this tells us that the lower the MAD of our model is and, especially, the higher the p-value of our model is, the better it describes our data.

With that in mind, we have proceeded to perform the above statistic computations. Nevertheless, since we are interested in comparing the goodness of the two DPP models, notice that we have not only worked with the whole dataset, without distinguishing the tree species, but we have also repeated the same calculations for the oak and birch trees separately. In the following table, we summarize the achieved results.

Model	**mad total**	p **total**	**mad oak**	p **oak**	**mad birch**	p **birch**
DPP	0.0093	0.85	0.0715	0.05	0.0595	0.06
Comb. DPP	0.0114	0.77	0.0754	0.08	0.0547	0.13
PPP	0.0101	0.75	0.0729	0.05	0.0573	0.05

Looking at the obtained results and focusing first on the whole dataset without distinctions, we can conclude that there is probably a certain degree of repulsion between the tree locations, something that can be seen by the fact that the determinantal models have outperformed the Poisson one. These observations, together with what we have already explained in this section of the project, shows us that there are certain problems in which considering repulsion properties is relevant and can be effectively done by working with DPPs, something that adds to their usefulness.

On the other hand, regarding now the case in which we individually consider the tree species, here we can observe that the combined DPP has outperformed the vanilla one by a small margin, at least when comparing their p-values, something that could motivate further study into these compositions we introduced in Subsection 6.2, that are not as common and well-studied as the simple DPPs.

Nevertheless, as a disclaimer, it is worth mentioning that we cannot consider the above results as a strong evidence of the superior modelling capabilities of the two DPP models we tested, because the observed differences may not be significant enough and heavily depend on the statistic that is used to summarize the point patterns. In addition, the algorithm that we used to sample the DPPs (see Algorithm 1 in [20]) is faster than the general ones, but sometimes suffers from numerical difficulties (inherited from the ones of the LU factorization). As a consequence, the objective of this toy application was merely to build our intuition, to exemplify the use of DPPs to model spatial patterns and to motivate further study into them, something that from our point of view has indeed been achieved.

As a final comment, all the code and simulations that we have developed to perform this toy application to tree location modelling can be found in the Google Drive folder whose link appears below. The code has been written using both C and R languages, calling functions from the Spatstat R package (see [2]) in the latter cases.

`https://drive.google.com/drive/folders/1ofcbG_pMCz2erRSClQiA8o67NIYM5kRo?usp=sharing`

References

[1] Alon, N.; Spencer, J. H.: *The probabilistic method*, 3rd. ed., Hoboken: John Wiley and Sons, New Jersey, 2008.

[2] Baddeley, A.; Turner, R.: Spatstat: an R package for analyzing spatial point patterns, *J. Stat. Softw.*, 12(6):1-42, 2005.

[3] Bauer, H.: *Measure and Integration Theory*, De Gruyter, Germany, 2011.

[4] Billingsley, P.: *Probability and measure*, 3rd. ed., Wiley, New York, 1995.

[5] Bogachev, V. I.: *Measure theory*, 1st. ed., Vol. 2, Springer Berlin / Heidelberg, Berlin, 2007.

[6] Doukhan, P.: *Stochastic models for time series*, Springer Cham, Cham, 2018.

[7] Folland, G. B.: *Real analysis: modern techniques and their application*, 2nd. ed., Wiley, New York, 1999.

[8] Gaudard, M.; Hadwin, D.: Sigma-Algebras on Spaces of Probability Measures, *Scand. J. Stat.*, 16(2):169-175, 1989.

[9] Grinberg, D.: Notes on the combinatorial fundamentals of algebra, `arXiv:2008.09862v3 [math.CO]`, September 2022.

[10] Hardy, A.; Maïda, M.: Determinantal Point Processes, *Newsl. Eur. Math. Soc.*, 112:8-15, June 2019.

[11] Hough, J. B.: *Zeros of Gaussian Analytic Functions and Determinantal Point Processes*, American Mathematical Society, Providence, R.I, 2009.

[12] Janson, S.: Normal Convergence by Higher Semiinvariants with Applications to Sums of Dependent Random Variables and Random Graphs, *Ann. Probab.*, 16(1):305-12, 1988.

[13] Kulesza, A.; Taskar, B.: Determinantal Point Processes for Machine Learning, *Found. Trends Mach. Learn.*, 5(2-3):123-286, 2012.

[14] Laskurain, N. A.: *Dinámica espacio-temporal de un bosque secundario en el Parque Natural de Urkiola (Bizkaia)*, PhD thesis, University of the Basque Country, 2008 (in Spanish).

[15] Lavancier, F.; Møller, J.; Rubak, E.: Determinantal point process models and statistical inference, *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 77(4):853-77, 2015.

[16] Macchi, O.: The Coincidence Approach to Stochastic Point Processes, *Adv. in Appl. Probab.*, 7(1):83-122, 1975.

[17] Markin, M. V.: *Elementary Operator Theory*, De Gruyter, Germany, 2020.

[18] Naranjo, J. C.: *Geometria Lineal: Grau de Matemàtiques, UB*, University of Barcelona, 2014 (in Catalan).

[19] Pemantle, R.; Peres, Y.: Concentration of Lipschitz Functionals of Determinantal and Other Strong Rayleigh Measures, *Combin. Probab. Comput.*, 23(1):140-60, 2014.

[20] Poulson, J.: High-performance sampling of generic determinantal point processes, *Philos. Trans. Roy. Soc. A*, 378(2166):1-17, 2020.

[21] Soshnikov, A.: The Central Limit Theorem for Local Linear Statistics in Classical Compact Groups and Related Combinatorial Identities, *Ann. Probab.*, 28(3):1353-70, 2000.

[22] Soshnikov, A.: Gaussian Limit for Determinantal Random Point Fields, *Ann. Probab.*, 30(1):171-87, 2002.

[23] Tao, T.: *An introduction to measure theory*, 1st. ed., American Mathematical Society, Providence, R.I, 2011.

[24] Tausk, D. V.: *Weak* topology for the space of finite measures on a topological space*, Institute of Mathematics and Statistics of the University of São Paulo, 2024.

[25] Whitney, H.: Analytic Extensions of Differentiable Functions Defined in Closed Sets, *Trans. Amer. Math. Soc.*, 36(1):63-89, 1934.

[26] Wilhelm, M.; Ramanathan, A.; Bonomo, A.; Jain, S.; Chi, E. H.; Gillenwater, J.: Practical Diversified Recommendations on YouTube with Determinantal Point Processes, *Proc. ACM Int. Conf. Inf. Knowl. Manag.*, 2165-2173, 2018.