10-27-2020

Numerical Study of Gap Distributions in Determinantal Point Process on Low Dimensional Spheres: *L*-Ensemble of *O*(*n*) Model Type for *n* = 2 and *n* = 3

Xiankui Yang
University of South Florida

Numerical Study of Gap Distributions in Determinantal Point Process on Low Dimensional Spheres:

L-ensemble of $O(n)$ model type for $n = 2$ and $n = 3$

by

Xiankui Yang

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Art
Department of Mathematics and Statistics
College of Mathematics and Statistics
University of South Florida

Major Professor: Seung-Yeop Lee, Ph.D.
Lu Lu, Ph.D.
Kandethody Ramachandran, Ph.D.

Date of Approval:
October 24, 2020

Keywords: Poisson Point Process, Circular Unitary Ensemble, Ginibre Ensemble, Hypothesis Test

Table of Contents

Abstract

Poisson point process is the most well-known point process with many applications. Unlike Poisson point process, which is the random set of non-intersecting points, determinantal point process refers to certain class of point processes where the points tend to interact with each other. The interaction often leads to more uniformly distributed points compared to those in Poisson point process.

In this article, we study the gap distribution of certain class of determinantal point process, L-ensemble of O(n) model type, and compare the distribution with the ones from the other known determinantal point process that appears in random matrices. Our numerical results suggest that our determinantal point process model converges to the known random matrices, hinting at the universality in the statistical limit.

1. Introduction

A point process is simply speaking the set of random points. For example, one-dimensional point process can represent the times of the certain event such as earthquake. The origin of modern point process theory can be traced back to Poisson in 1837 (Also See [6], 2020). The Poisson point process is applied in cosmology, geodesy, and ecology, and it is commonly used for modeling the point pattern of attraction between display points (Afshang, Saha, and Dhillon, 2017).

When one wants to understand or to measure a large system, the point process can be used to sample data points from the system. The Poisson point process may not be optimal sometimes as it may select too many points from some parts of the system while missing out some other points. For instance, an employer may want to know the operation of the company by interviewing her employees. If she uses the Poisson point process to select the employees, it might happen that all interviews are from a few departments while missing other departments. For another example, if a telecommunications company establishes the telecom base stations hypothetically after the Poisson point process, then it is quite possible to have two base stations close to each other, which is not an effective way to cover the area. To avoid those cases, we need a process that selects points more uniformly. Although one can choose points uniformly a low dimensional case such as in the above examples, such uniform choice is not possible when the dimension is very large. There the determinantal point process can be practically useful tool as a method of selection.

Determinantal point process (DPP) refers to the class of point processes with certain determinantal structure. Unlike Poisson point process, locations of the points are affected by each other in such a way that they tend to repel each other, hence spread rather uniformly without clumping together. Determinantal point process also appeared in random matrices where certain specific DPPs tend to pop up in the thermodynamic limit, regardless of the specifics of the models, hence exhibiting a universality. In statistics DPP can be used to balance the diversity and the quality of the sample (Mariet and Sra, n.d.). If there are many features in a large system, the DPP would choose a sample such that each item in the sample represents different features such that any two items do not share the similar features.

The uniform versus the clustering natures of the point process may be captured by looking at the gap between the neighboring points. For example, let $\{X_1, X_2, \dots, X_n, \dots\}$ be a point process on $\mathbb{R}$ with the order $X_1 \leq X_2 \leq \cdots \leq X_n \leq \cdots$. Then $X_2 - X_1, X_3 - X_2, \dots, X_n - X_{n-1}, \dots$ are the gaps. In $\mathbb{R}^2$, let $\{X_1, X_2, \dots, X_n, \dots\}$ be a point process. $d(X_i, X_j)$ is a Euclidean distance between X_i and X_j where $i \neq j$, and X_i and X_j are random variables. Then the gap for ith point is given by $min_{i \neq j}\{d(X_i, X_j)\}$ for $i, j \in \mathbb{N}$. Gap distribution tells us the important information about the point process. We will use gap distribution to compare the point processes. In the gap distribution of the Poisson points process, the clustering nature leads to the higher concentration near the small gaps comparing to gap distribution of determinantal point process. To explore these two point processes, we study the following problems.

1. Numerically confirming exponential gap distribution for Poisson point process.

2. Finding what kinds of gap distribution for determinantal point processes.

3. Comparing our specific determinantal point processes with those from random matrices

4. Checking uniformity of the points in a determinantal point process.

In this paper, we just talk about the one-dimensional case and the two-dimensional case as mathematical model. In section 2, we will introduce the definition of point process and the definition of Poisson point process. We also introduce the way to sample a Poisson point process. For one-dimensional case and two-dimensional case, the gap distribution of a Poisson point process is discussed. In section 3, the definition of determinantal point process is introduced, and we also introduce L-ensemble to structure the kernel of determinantal point process. An algorithm of sampling a DPP was discussed according to Kulesza's (2012) paper. In section 4, we consider the one-dimensional case and two-dimensional case for determinantal point process. The factor that affects the number of points of a DPP is discussed. The gap distribution of DPPs is explored with the numerical data. In section 5, we use the results of the previous sections to answer the above four questions, and we make a little further discussion based on the conclusion that we got for one- and two-dimensional cases.

2. Poisson Point Process

The Poisson point process was named from its natural relation to the Poisson distribution, developed by Poisson as a limiting case of the binomial distribution [Last and Penrose, 2018], which is the probability of the sum of n Bernoulli variables with probability p. In stochastic process, a Poisson process is frequently used for modeling the times at which arrivals enter a system [Gallager, 1999]. For example, for a Poisson point process $(X_1, X_2, ...,)$, when a point exists at ith time, we denote the point as $X_i = 1$ and name it as an "arrival"; on the contrary, we denote it as $X_i = 0$ and name it as an "no arrival". A point process is a random collection of at most countably many points, possibly with multiplicities [Last and Penrose, 2018].

2.1 Point process

A point process is a random and at most countable collection Z of points in some space $\mathbb{X}$ [Last and Penrose, 2018]. For example, $\mathbb{X}$ could be the d-dimensional Euclidean space $\mathbb{R}^d$. For an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the collection Z could be a mapping $\omega \mapsto Z(\omega)$ from Ω into the system of countable subsets of $\mathbb{X}$. Then we can identify Z with a family of mappings

$$\omega \mapsto \eta(\omega, B) := \mathrm{card}(Z(\omega) \cap B), B \subset \mathbb{X}$$

counting the number of points that Z has in B.

Let $(\mathbb{X}, \mathcal{X})$ be a measurable space, where $\mathcal{X}$ is the collection of subsets of $\mathbb{X}$, and let $\mathbf{N}_{<\infty}(\mathbb{X}) \equiv \mathbf{N}_{<\infty}$ denote the space of all measures μ on $\mathbb{X}$ such that $\mu(B) \in \mathbb{N}_0$ for all $B \in \mathcal{X}$, where $\mathbb{N}_0$ means $\mathbb{N} \cup \{0\}$. Then let $\mathbf{N}(\mathbb{X}) \equiv \mathbf{N}$ be the space of all measures that can be written as a countable sum of measures from $\mathbf{N}_{<\infty}$. An example of an element of $\mathbf{N}$ is the zero measure 0 that is identically zero on $\mathcal{X}$. Another example is the Dirac measure δ_x at a point $x \in \mathbb{X}$ given by $\delta_x(B) = \mathbf{1}_B(x)$. More generally, any finite or infinite sequence $(x_n)_{n=1}^k$ of elements of $\mathbb{X}$, where $k \in \overline{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$ is the number of terms in the sequence, can be used to define a measure

$$\mu = \sum_{n=1}^{k} \delta_{x_n}. \tag{2.1}$$

The $\mu \in \mathbf{N}$ and

$$\mu(B) = \sum_{n=1}^{k} \mathbf{1}_B(x_n), \quad B \in \mathcal{X}$$

We can allow for $k = 0$ in (2.1). In this case μ is the zero measure. The points $x_1, x_2, \cdots$ are not assumed to be pairwise distinct. If $x_i = x_j$ for some $i, j \leq k$ with $i \neq j$, then μ is said to have multiplicities. In fact, the multiplicity of x_i is the number card{j ≤ k: $x_i = x_j$}. Any μ of form (2.1) is interpreted as a counting measure with possible multiplicities.

Let $\mathcal{N}(\mathbb{X}) \equiv \mathcal{N}$ denote the $\sigma-$field generated by the collection of all subsets of $\mathbf{N}$ of the form

$$\{\mu \in \mathbf{N}: \mu(B) = k\}, \qquad B \in \mathcal{X}, k \in \mathbb{N}_0.$$

This means that $\mathcal{N}$ is the smallest $\sigma-$field on $\mathbf{N}$ such that $\mu \mapsto \mu(B)$ is measurable for all $B \in \mathcal{X}$.

Definition 2.1 (Last and Brandt, 2005) A point process on $\mathbb{X}$ is a random element η of $(\mathbf{N}, \mathcal{N})$, that is a measurable mapping $\eta: \Omega \to \mathbf{N}$.

If η is a point process on $\mathbb{X}$ and $B \in \mathcal{X}$, then we denote by $\eta(B)$ the mapping $\omega \mapsto \eta(\omega, B) = \eta(\omega)(B)$. By the definitions of η and the $\sigma-$field $\mathcal{N}$ these are random variables taking values in $\overline{\mathbb{N}}_0 = \overline{\mathbb{N}} \cup \{0\}$, that is, for $k \in \overline{\mathbb{N}}_0$.

$$\{\eta(B) = k\} \equiv \{\omega \in \Omega: \eta(\omega, B) = k\} \in \mathcal{F}. \tag{2.2}$$

Conversely, a mapping $\eta: \Omega \to \mathbf{N}$ is a point process if (2.2) holds. In this case we call $\eta(B)$ the number of points of η in B.

Example 2.2 (Last and Brandt, 2005) Let $\mathbb{Q}$ be a probability measure on $\mathbb{X}$ and suppose that $X_1, X_2, \cdots, X_m$ are independent random element in $\mathbb{X}$ with distribution $\mathbb{Q}$. Then

$$\eta = \delta_{x_1} + \delta_{x_2} + \cdots + \delta_{x_m}$$

is a point process on $\mathbb{X}$. Because

$$\mathbb{P}(\eta(B) = k) = \binom{m}{k} \mathbb{Q}(B)^k (1 - \mathbb{Q}(B))^{m-k}, k = 0, \cdots, m,$$

η is referred to as a binomial process with sample size m and sampling distribution $\mathbb{Q}$.

2.2 Poisson process

For the Poisson process, arrivals may occur at arbitrary positive times, and the probability of an arrival at any instant is 0, which means that we cannot directly describe a Poisson process in terms of the probability of an arrival at any given instant. Therefore, we first describe arrival processes.

Definition 2.3 (Gallager, 1999) An arrival process is a sequence of increasing random variables, $0 < S_1 < S_2 < \cdots$, where $S_i < S_{i+1}$ means that $S_{i+1} - S_i$ is a positive random variables, i.e., a random variable X such that $\mathrm{F}_X(0) = 0$, which is equal to $\mathbb{P}(X \leq 0) = 0$. The random variables $S_1, S_2, \cdots$, are called arrival

epochs and represent the times at which some repeating phenomenon occurs. Then, we let $X_1 = S_1$ and $X_i = S_i - S_{i-1}$ for $i > 1$, and we get an alternative process $X_1, X_2, \cdots$, a sequence of interarrival times.

A Poisson process is an example of an arrival process, which the interarrival times are defined to be independent and identically distributed (IID). Therefore, we can theoretically describe the Poisson process as follows.

Definition 2.4 (Gallager, 1999) A renewal process is an arrival process for which the sequence of interarrival times is a sequence of IID random variables.

Definition 2.5 (Gallager, 1999) A Poisson process is a renewal process in which the interarrival intervals have an exponential distribution function; i.e., for some real $\lambda > 0$, each X_i has the density $f_X(x) = \lambda \exp(-\lambda x)$ for $x \geq 0$.

From above definition, we assume that $\{N(t); t > 0\}$ is the Poisson counting process consisting of a nonnegative integer random variable $N(t)$ for each $t > 0$, i.e. $N(t) = max\{n | S_n < t\}$.

Theorem 2.6 (Gallager, 1999) For a Poisson process of rate λ, and for any $t > 0$, the PMF for $N(t)$, where $N(t)$ is the number of arrivals in $(0, t]$, is given by the Poisson PMF,

$$p_{N(t)}(n) = \frac{(\lambda t)^n \exp(-\lambda t)}{n!} \tag{2.3}$$

Proof: Recalling the definition 2.3, 2.4 and 2.5, we can obtain that

$$S_n = \sum_{i=1}^{n} X_n \tag{2.4}$$

For a Poisson process, S_n is the sum of n IID random variables, which have the same density function $f_X(x) = \lambda \exp(-\lambda x)$ for $x \geq 0$. Since the density of the sum of two independent random variables can be

found by convolving their densities, we can obtain that for $t \geq 0$, the density of S_n is

$$f_{S_n}(t) = \frac{\lambda^n t^{n-1} \exp(-\lambda t)}{(n-1)!} \tag{2.5}$$

For giving n and t, we have two ways of calculating $\Pr\{t < S_{n+1} \leq t + \delta\}$ for some vanishingly small δ. The first way is based on the already known density of S_{n+1} and gives

$$\Pr\{t < S_{n+1} \leq t + \delta\} = \int_t^{t+\delta} f_{S_{n+1}}(\tau)\, d\tau = f_{S_{n+1}}(t)\big(\delta + o(\delta)\big)$$

The term $o(\delta)$ is used to describe a function of δ that goes to 0 faster than δ as $\delta \to 0$. Thus $\Pr\{t < S_{n+1} \leq t + \delta\} = f_{S_{n+1}}(t)\big(\delta + o(\delta)\big)$ is simply a consequence of the fact that S_n has a continuous probability density in the interval $(t, t + \delta]$.

The second way to calculate $\Pr\{t < S_{n+1} \leq t + \delta\}$ is to first observe that the probability of more than 1 arrival in $(t, t + \delta]$ is $o(\delta)$. Ignoring this possibility, $\{t < S_{n+1} \leq t + \delta\}$ occurs if exactly n arrivals are in the interval $(0, t]$ and one arrival occurs in $(t, t + \delta]$. Because of the independent increment property, this is an event of probability $p_{N(t)}(n)(\lambda\delta + o(\delta))$. Thus

$$p_{N(t)}(n)(\lambda\delta + o(\delta)) + o(\delta) = f_{S_{n+1}}(t)\big(\delta + o(\delta)\big).$$

Dividing by δ and taking the limit $\delta \to 0$, we get

$$\lambda p_{N(t)}(n) = f_{S_{n+1}}(t)$$

From 2.5, we can obtain that

$$p_{N(t)}(n) = \frac{(\lambda t)^n exp(-\lambda t)}{n!}.$$ ■

From theorem 2.6, we can have that the distribution of the Poisson process is related to the Poisson distribution, which we can denote that $N(t) \sim Poi(\lambda t)$. Therefore, recalling the definition 2.1, we have the alternative definition of the Poisson process.

Definition 2.7 (Last and Penrose, 2018) A Poisson process with intensity measure λ is a point process η on $\mathbb{X}$ with the following two properties:

(i) For every $B \in \mathcal{X}$, the distribution of $\eta(B)$ is Poisson with parameter λ(B), that is to say $\mathbb{P}(\eta(B) = k) = \text{Poi}(\lambda(B); k)$ for all $k \in \mathbb{N}_0$.

(ii) For every $m \in \mathbb{N}$ and all pairwise disjoint sets $B_1, B_2, \cdots, B_m \in \mathcal{X}$, the random variables $\eta(B_1), \eta(B_2), \cdots, \eta(B_m)$ are independent.

Property (i) of Definition 2.7 is responsible for the name of the Poisson process. A point process with property (ii) is said to be completely independent. One also says that η has independent increments or is completely random.

2.3 The one-dimensional Poisson process

Poisson points process requires that every point needs to be sampled independently. In this section, we explore the Poisson points process in one-dimensional case, especially on a unit circle. From section 2.2, we know that the interarrival times in a Poisson process are defined to be independent and identical distribution. Let $N(t) = max\{n|S_n < t\}$ be the number of arrivals in $(0, t]$. According to theorem 2.6 we have

$$\mathbb{P}(N(t) = 0) = e^{-\lambda t} \text{ and } \mathbb{P}(N(t) = 1) = \lambda t e^{-\lambda t} \tag{2.6}$$

Now we let $S_1, S_2, \ldots, S_n, \ldots$ be a Poisson points process on a real line $\mathbb{R}$ and $X_i = S_i - S_{i-1}$ be the difference between two consecutive points in a Poisson points process. Without loss of generality, we assume that the location of X_i is s on a real line. Then

$$\mathbb{P}(X_i > t \mid S_i = s) = \mathbb{P}(\text{no points on}(s, s+t] | S_i = s)$$

$$= \mathbb{P}(\text{no points on}(s, s+t])$$

$$= \mathbb{P}(N(t) = 0) = e^{-\lambda t} \tag{2.7}$$

A sequence $S_1, S_2, \ldots, S_n, \ldots$ of random variables, which have the property that the interarrival times $S_1, S_2 - S_1, S_3 - S_2, \cdots$ are independent random variables with an exponential(λ) distribution, is the one-dimensional Poisson process with intensity λ.

From (2.3), the distribution of the number of points only depends on the length of the interval rather than on its location. Therefore, without loss of generality, we can assume an interval starting at 0. Then we consider a Poisson point process on the interval $[0, a]$. If there is one point on $[0, a]$. Then, for $0 < s < a$,

$$\mathbb{P}(S_1 \le s \mid N([0,a]) = 1) = \frac{\mathbb{P}(S_1 \le s, N([0,a]) = 1)}{\mathbb{P}(N([0,a]) = 1)}$$

$$= \frac{P\big(N([0,s]) = 1, N\big((s,a]\big) = 0\big)}{P(N([0,a]) = 1)} \tag{2.8}$$

Since the interarrival times are independent, and from formula 2.6,

$$\mathbb{P}(S_1 \le s \mid N([0,a]) = 1) = \frac{\lambda s e^{-\lambda s} e^{-\lambda(a-s)}}{\lambda a e^{-\lambda a}} = \frac{s}{a} \tag{2.9}$$

Therefore, for one point of a Poisson process on the interval $[0, a]$, the random variable S_1 is uniformly distributed over the interval $[0, a]$.

If there are two points on $[0, a]$. Then, for $0 < s < t < a$,

$$\mathbb{P}(S_1 \le s, S_2 \le t \mid N([0,a]) = 2)$$

$$= \frac{\mathbb{P}\big(N([0,t]) = 2, N\big((t,a]\big) = 0\big)}{\mathbb{P}(N([0,a]) = 2)} - \frac{\mathbb{P}\big(N([0,s]) = 0, N\big((s,t]\big) = 2, N\big((t,a]\big) = 0\big)}{\mathbb{P}(N([0,a]) = 2)}$$

$$= \frac{\frac{(\lambda t)^2}{2} e^{-\lambda t} e^{-\lambda(a-t)}}{\frac{(\lambda a)^2}{2} e^{-\lambda a}} - \frac{e^{-\lambda s} \frac{(\lambda(t-s))^2}{2} e^{-\lambda(t-s)} e^{-\lambda(a-t)}}{\frac{(\lambda a)^2}{2} e^{-\lambda a}} = \frac{t^2 - (t-s)^2}{a^2} \tag{2.10}$$

Lemma 2.8 If S_1 and S_2 are two independent random variables, both uniformly sampling over $[0, a]$, and let $V = min\{S_1, S_2\}$ and $U = max\{S_1, S_2\}$; then, the joint distribution function of V and U is given by

$$\mathbb{P}(V \leq s, U \leq t\ ) = \frac{t^2 - (t-s)^2}{a^2} \text{ for } 0 \leq s < t \leq a \tag{2.11}$$

Proof: Since S_1 and S_2 are independent and V and U just order S_1 and S_2, we can obtain that

$$\mathbb{P}(V \leq s, U \leq t\ ) = \mathbb{P}(V \leq t, U \leq t\ ) - \mathbb{P}(s < V \leq t) \tag{2.12}$$

Since S_1 and S_2 are sampled from the uniformly distribution over $[0, a]$, V and U also follow the uniformly distribution and $V < U$. Therefore,

$$\mathbb{P}(V \leq t, U \leq t\ ) = \frac{t^2}{a^2} \; and \; \mathbb{P}(s < V \leq t) = \mathbb{P}(s < V \leq t, s < U \leq t) = \frac{(t-s)^2}{a^2}.$$

Then, we can get the (2.11).

From above lemma 2.8, we can find that if we ignore the order of S_1 and S_2 in the Poisson process, from (2.10), S_1 and S_2 are independent and uniformly distributed over $[0, a]$. With more step checking, we have the following inference.

Corollary 2.9 There is a Poisson process with n points over the interval $[a, b]$; then, the locations of these points are independent and uniformly distributed over $[a, b]$.

2.3.1 Poisson process on a unit circle

A unit circle is a fitting model for doing the Poisson process because every point on the circle could be the initial point in a process. From corollary 2.9, we can uniformly sample points on the unit circle, and

then we can check the gap distribution of two neighboring points to judge if the process is the Poisson process with definition 2.5.

(1) Gap distribution with a fixed point

Let n be the n points sampling at each processing and m be the m samples. For each circulation we sample $x_1, x_2, \dots, x_n$ from $[0,2\pi]$, and let $X_i = \{x^{(1)}, x^{(2)}, \dots, x^{(n)}\}$ be the ordered points at ith circulation. Then let Y_i be the gap of a points being closest to point (1,0) with counterclockwise direction.

Algorithm 2.10 Sampling m times Poisson process on a unit circle

Input n, m
For $i = 1,2, \dots, m$
  $X_i =$ sampling n order points from $unif[0,2\pi]$
  $Y_i =$ the gap of a point closest to point (1,0)
End
Output $Y = \{Y_1, Y_2, \dots, Y_m\}$

We fixed $m = 100000$ and separately set $n = 1000, 2000, 3000, 4000$. Then we obtain the histograms whose bin width follows Freedman-Diaconis rule (Bin is $2 \cdot \text{IQR} \cdot n^{-1/3}$, IQR is interquartile range), i.e. figure 2.2. Here we let our x-axis as

$$x - axis: gaps \times \lambda = gaps \times \frac{n}{2\pi}$$

The lines in the histograms are the theoretical gap distribution exponential(1).

Table 2.1 The information of gaps distribution for different n on a unit circle

n	m	$mean\ (\mu)$	$\lambda = 1/\mu$	$\lambda = n/2\pi$
1000	100000	0.0062933	158.9002	159.1549
2000	100000	0.0031313	319.3574	318.3099
3000	100000	0.0020899	478.4841	477.4648
4000	100000	0.0015665	638.3852	636.6198

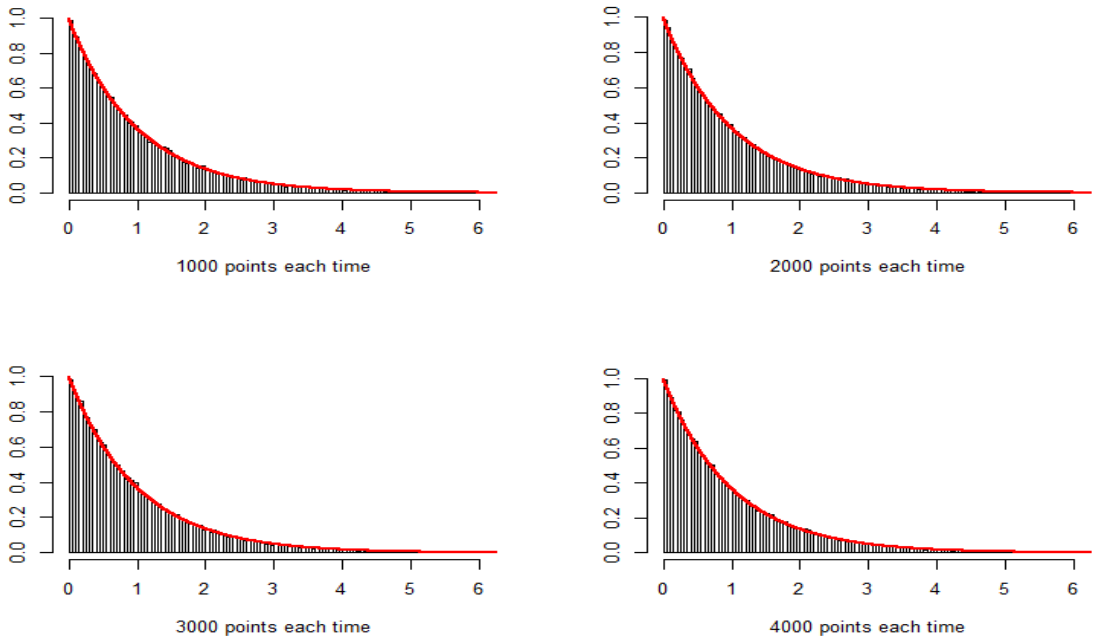

Figure 2.2 Histograms of gaps between (1,0) and the closest point with counterclockwise direction, in the condition of $n = 1000, 2000, 3000, 4000$ and $m = 100000$.

From table 2.1, the numerical λ is almost equal to the theoretical value of λ. Through doing the Z statistic test, the z values for $n = 1000, 2000, 3000, 4000$ are 0.507, -1.039, -0.674 and -0.880. The p-values for $n = 1000, 2000, 3000, 4000$ are 0.612, 0.299, 0.500 and 0.378. Therefore, we cannot reject the null hypothesis that the simulation λ is equal to the theoretical λ at a significant level 0.95 for $n = 1000, 2000, 3000, 4000$.

(2) Gap distribution with all random points

Let n be the n points sampling at each processing and m be the m samples. For each circulation we sample $x_1, x_2, \dots, x_n$ from $[0,2\pi]$, and let $X_i = \{x^{(1)}, x^{(2)}, \dots, x^{(n)}\}$ be the ordered points at ith circulation. Then let Y_i be the gaps of two neighboring points with counterclockwise direction, i.e. $Y_i = \{y_1, y_2, \dots y_n\}$, where $y_j = x^{(j+1)} - x^{(j)}$ for $j = 1,2, \dots, n-1$ and $y_n = 2\pi - (x^{(n)} - x^{(1)})$.

Algorithm 2.11 Sampling m times Poisson process on a unit circle

Input n, m

For $i = 1,2, \dots, m$

$X_i =$ sampling n order points from $unif[0,2\pi]$

$Y_i =$ the gaps of two neighboring points

End

Output $Y = \{Y_1, Y_2, \dots, Y_m\}$

We fix $m = 100$ and separately set $n = 1000, 2000, 3000, 4000$. Then we get the histograms with breaks of Freedman-Diaconis rule (Bin is $2 \cdot \text{IQR} \cdot n^{-1/3}$), i.e. figure 2.3.

$$x - axis: gaps \times \lambda = gaps \times \frac{n}{2\pi}$$

If these processes are Poisson processes, we have $y_1, y_2, \cdots y_n$ are independent and identical distribution. Then from (2.6), we have

$$\mathbb{P}(N(y_1) = 0, N(y_2) = 0, \dots, N(y_n) = 0) = \prod_{j=1}^{n} \mathbb{P}(N(y_j) = 0) = \prod_{j=1}^{n} e^{-\lambda y_j} = e^{-n\lambda y_1} \qquad (2.13)$$

Since this process is on a unit circle, every point could be the initial point of a process. From (2.13), we have the theoretical λ for this exponential distribution, which is $n/2\pi$. In figure 2.3, the histograms show the numerical data density, and the red lines show the theoretical exponential density function. In table 2.4(a), we can find that the simulate λ is the same as the theoretical λ. In table 2.4(b) and (2.13), we can find that n has effects on the distribution. If we use n to time the gaps, we can get the same gap distribution on the unit circle, i.e. figure 2.5.

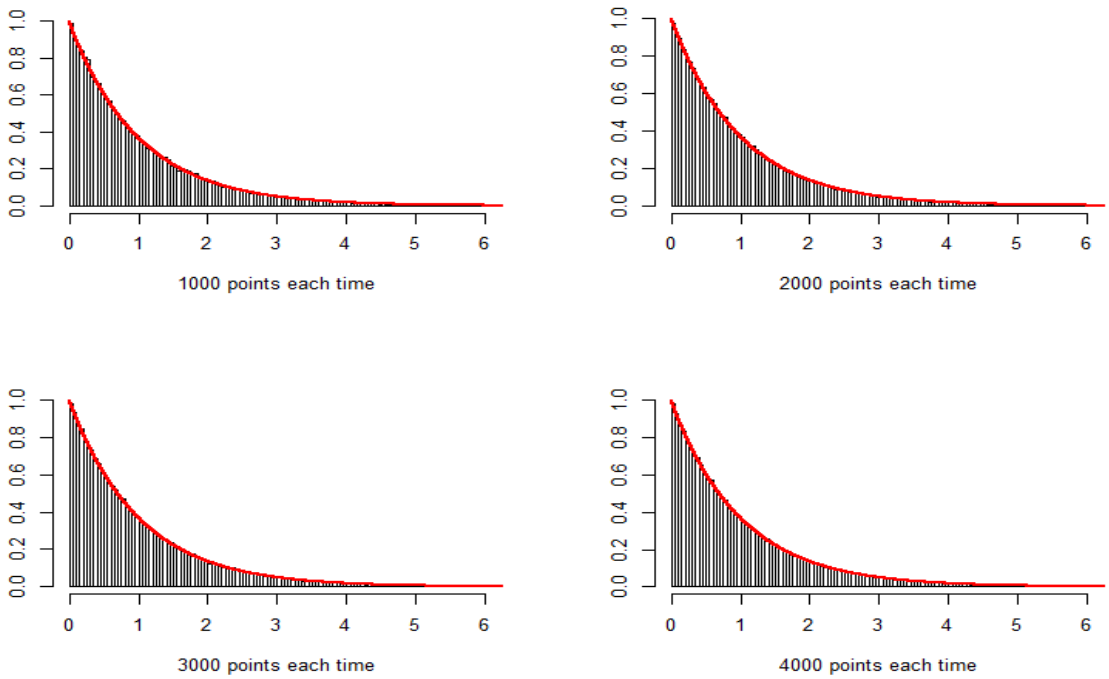

Figure 2.3 Histograms of gaps between neighboring points with $n = 1000, 2000, 3000, 4000$ and $m = 100$.

Table 2.4(a) The information of gaps distribution for different n on a unit circle

n	m	$mean\ (\mu)$	$\lambda = 1/\mu$	$\lambda = n/2\pi$
1000	100	0.0062832	159.1549	159.1549
2000	100	0.0031416	318.3099	318.3099
3000	100	0.0020944	477.4648	477.4648
4000	100	0.0015708	636.6198	636.6198

Table 2.4(b) The information of gaps distribution for different n on a unit circle

simulate $\lambda = 1/\mu$	$a = \lambda/n$	$b = 1/a$
159.1549	0.1591549	6.283185
318.3099	0.1591549	6.283185
477.4648	0.1591549	6.283185
636.6198	0.1591549	6.283185

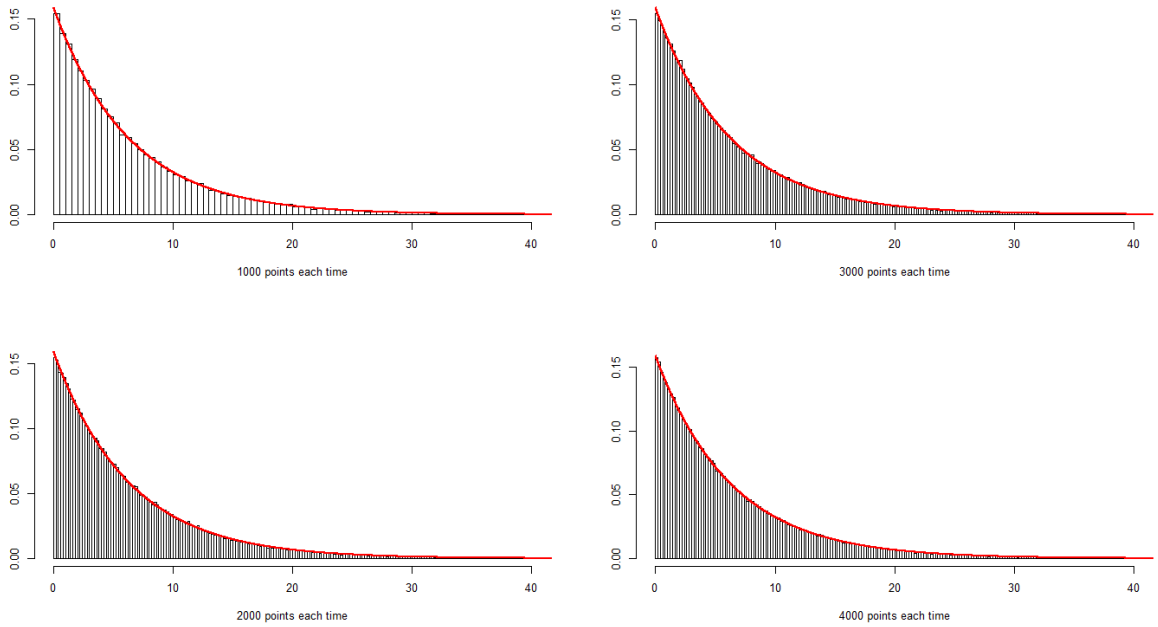

Figure 2.5 Histograms of gaps* n between neighboring points with $n = 1000, 2000, 3000, 4000$ and $m = 100$.

Through the results of (1) and (2) cases, we can find that for the same n, the gap distribution with a fixed point is equal to the gap distribution with all random points. The reason is that every random point on the circle is independent with other points. For example, a Poisson process on a circle $(x_1, x_2, \dots, x_n)$, the x_1 could be considered as a fixed initial point as the point $(1,0)$ and $(x_2, \dots, x_n)$ could be considered as a new Poisson process. Then we cycle through this perspective. Hence the gap distribution for this Poisson process is similar with the gap distribution that we do n times Poisson process with a fixed initial point $(1,0)$.

In one dimensional case, when we sample a Poisson point process, we can sample a series of points with a uniform distribution. Besides, these points should be sampled independently which could ensure the gaps between near points are independents satisfying the definition 2.5.

2.4 The two-dimensional Poisson process

In the above section, we use a unit circle to discuss the one-dimensional Poisson process. because every point on the circle can be considered as the initial point in the Poisson process. Therefore, in this section, we could use a similar model to explore the two-dimensional Poisson process, i.e. we can sample points on the unit sphere.

In the one-dimensional case, we can generate points with exponentially distributed gaps. We can have a similar procedure for the two-dimensional case. In a real plane, without loss of generality, we assume the initial point X_1 is the reference point. Let $s > 0$ be the linear distance between X_1 and X_2. Let

$$M_s = N(C_s) \tag{2.14}$$

where C_s is the circular region of radius s, i.e. C_s has area πs^2, and $N(C_s)$ is number of particles in region C_s. Therefore,

$$\mathbb{P}(M_s = n) = \frac{(\lambda\pi s^2)^n exp(-\lambda\pi s^2)}{n!} \tag{2.15}$$

Let R_i be the linear distance between X_1 and X_{i+1}. We want $R_i \leq s$ if and only if $M_s \geq i$. Let $i = 1$; then,

$$\mathbb{P}(R_1^2 \leq s^2) = \mathbb{P}(R_1 \leq s) = \mathbb{P}(M_s \geq 1) = 1 - e^{-\lambda\pi s^2} \tag{2.16}$$

In other words, R_1^2 is exponential$(\lambda\pi)$ distribution. For general i, we can similarly write

$$\mathbb{P}\left(R_i^2 \leq s^2\right) = \mathbb{P}(R_i \leq s) = \mathbb{P}(M_s \geq i) = 1 - \sum_{j=0}^{i-1} \frac{(\lambda\pi s^2)^j exp(-\lambda\pi s^2)}{j!} \tag{2.17}$$

Hence R_i^2 has a $Gamma(i, \lambda\pi)$ distribution.

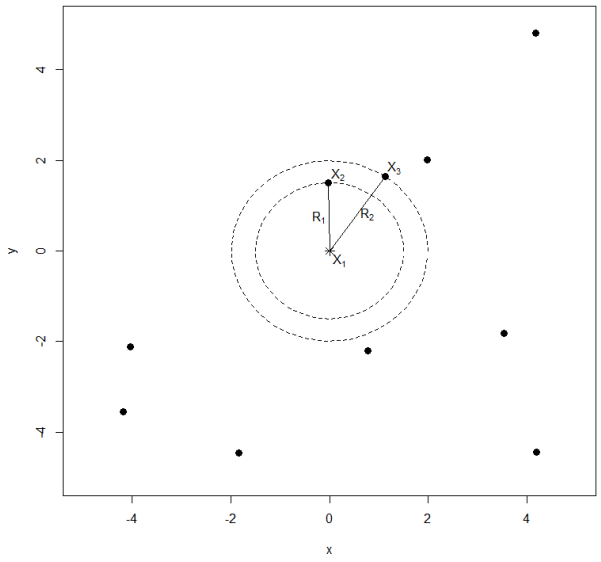

Figure 2.6 The Poisson process in the plane with two gaps of the two points closest to the origin.

2.4.1 Poisson process on a unit sphere

For a Poisson process, similar to the case of the unit circle, we sample n points on a unit sphere. Then for each point A, we find the point $B \neq A$ that is closest to the point A. Next, we collect the gap distance between A and B. Finally, we repeat this processing m times to get the data of gap distribution of the Poisson process on the sphere.

Algorithm 2.12 Sampling m times Poisson process on a unit sphere

Input n, m
For $i = 1,2, \dots, m$
$X_i =$ a set of n points from a sphere
$D_i =$ a set of the gaps of two closest points
End
Output D= $\{D_1, D_2, \dots, D_m\}$

We fix $m = 100$ and separately set $n = 1000, 2000, 3000, 4000$. Then we get the histograms with breaks of Freedman-Diaconis rule (Bin is $2 \cdot \text{IQR} \cdot n^{-1/3}$), i.e. figure 2.7.

$$x - axis:\ gaps^2 \times \lambda = gaps^2 \times \frac{n}{4}$$

On the sphere, every point could be the reference point. From (2.16), the squares of the gaps between two closest points follow the exponential$(\lambda\pi)$ distribution. For a circle centered at point A with radius AB on the sphere, there does not exist any points in this circle except the point A. For a Poisson process, λ is given by the number of points on the unit sphere by (2.15).

$$\mathbb{E}\left(R_j^2\pi\right) = \frac{4\pi}{n} \tag{2.18}$$

$$\mathbb{E}\left(R_j^2\right) = \frac{4}{n} = \frac{1}{\lambda} \tag{2.19}$$

Therefore, the theoretical λ is $n/4$. From figure 2.7, we can directly see that the gaps of closest points on the sphere follow exponential distribution. Then the simulate λs are approximately equal to the theoretical λ.

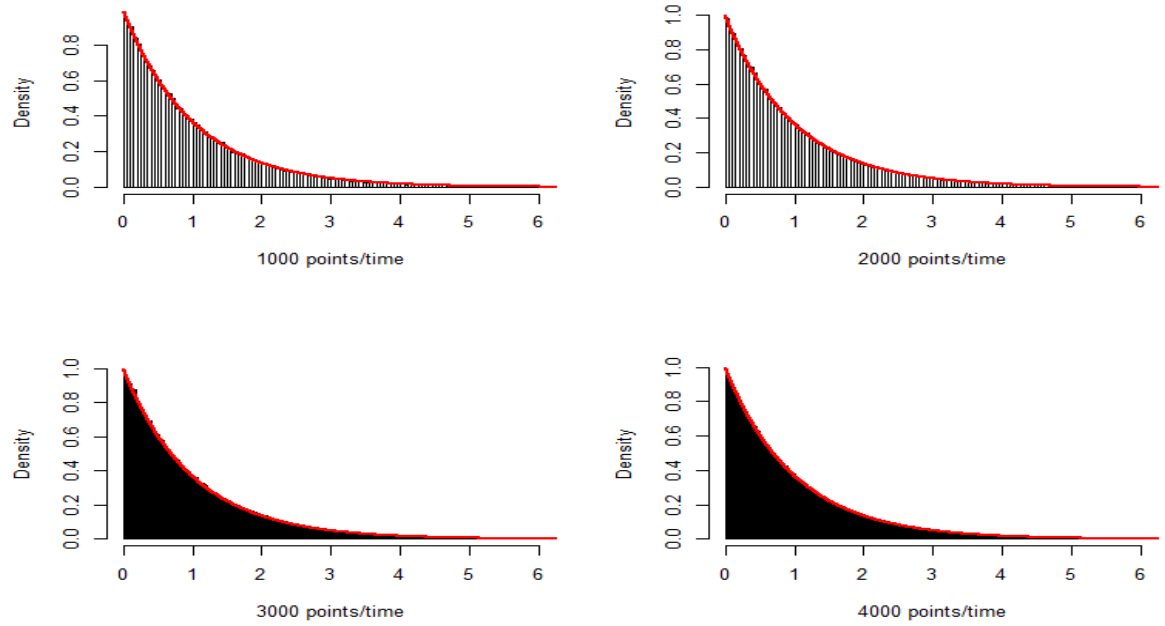

Figure 2.7 Histograms of gaps between closest points with $n = 1000, 2000, 3000, 4000$ and $m = 200$.

Table 2.8 The information of square gaps distribution for different n on a unit sphere

n	m	$mean\ (\mu)$	$\lambda = 1/\mu$	$\lambda = n/4$
1000	200	0.004009029	249.437	250
2000	200	0.002000778	499.8055	500
3000	200	0.001336522	748.2104	750
4000	200	0.0009996887	1000.311	1000

In both one-dimensional case and in two-dimensional case, the gaps between neighboring points follow the exponential distribution. If we uniformly sample points from a space, the process is a Poisson process. For one-dimensional case with region $[a, b]$, the gaps between neighboring points follow exponential$\big(n/(b-a)\big)$ distribution. For two-dimensional case with region $[a, b] \times [c, d]$, the square gaps between closest points follow exponential$(n\pi/[(b-a)(c-d)])$ distribution.

3. Determinantal Point Process

The exponential distribution shows that gaps between two neighbor points could be smaller with a higher probability, which means that it is easy to sample a series of points gathering in one location. If we want these points could cover the full space, we need to sample many points. With these points, if we want to deal with the information of these points, it is difficult, and it would take a long time. Therefore, we need to find another process such that the points that we sample can be evenly distributed in the space.

Determinantal point processes (DPPs) were first called "fermion process", which give the distribution of fermion systems at thermal equilibrium [Macchi, 1975, Kulesza and Taskar, 2012]. Although DPPs are initially applied in physics, the specific DPPs appeared in major results in random matrix theory and continue working in random matrix as an important role [Diaconis, 2003, Johansson, 2002, Kulesza and Taskar, 2012]. The term "determinantal" was first used by Borodin and Olshanski [2000] and has become a standard concept. The aspects of DPPS are described including definition, structures, and algorithms.

3.1 Determinantal point process

Let $\mathcal{P}$ be a point process on a ground set $\mathcal{Y}$, which is the probability measure of the finite subsets of $\mathcal{Y}$. For the remainder of this paper, we assume that $\mathcal{P}$ is discrete finite point processes. Then without loss of generality, we can also assume that $\mathcal{Y} = \{1,2,\cdots,N\}$; in this setting we sometimes refer to elements of $\mathcal{Y}$ as items.

In the discrete case, a point process in simply a probability measure on $2^{\mathcal{Y}}$ (power set), the set of all subsets of $\mathcal{Y}$. A sample from $\mathcal{P}$ might be the empty set, the entirety of $\mathcal{Y}$, or anything in between.

Definition 3.1 (Kulesza and Taskar, 2012) $\mathcal{P}$ is called a determinantal point process if, when $\boldsymbol{Y}$ is a random subset drawn according to $\mathcal{P}$, we have, for every $A \subseteq \mathcal{Y}$,

$$\mathcal{P}(\mathrm{A} \subseteq \boldsymbol{Y}) = \det(K_A) \tag{3.1}$$

for some real, symmetric $N \times N$ matrix K indexed by the elements of $\mathcal{Y}$. Here, $K_A \equiv [K_{i,j}]_{i,j \in A}$ denotes the restriction of K to the entries indexed by elements of A, and we adopt $\det(K_\emptyset) = 1$.

For instance, let $\mathcal{Y} = \{1,2,3,4\}$ and $K = \begin{bmatrix} k_{11} k_{12} k_{13} k_{14} \\ k_{21} k_{22} k_{23} k_{24} \\ k_{31} k_{32} k_{33} k_{34} \\ k_{41} k_{42} k_{43} k_{44} \end{bmatrix}$. If we assume $A = \{1,3\}$, then we have the probability measure $\mathcal{P}(\mathrm{A} \subseteq \boldsymbol{Y}) = \mathcal{P}\{1,*,3,*\} = \det(K_A) = \begin{vmatrix} k_{11} & k_{13} \\ k_{31} & k_{33} \end{vmatrix}$.

Since $\mathcal{P}$ is a probability measure, all principal minors $\det(K_A)$ of K must be nonnegative. Therefore, K itself must be positive semidefinite. In other words, for every $A \subseteq \mathcal{Y}$, $\mathcal{P}(\mathrm{A} \subseteq \boldsymbol{Y}) = \det(K_A) \in [0,1]$, and thus the eigenvalues and minors of K must satisfy $[0,1]$. Therefore, for any K, $0 \preccurlyeq K \preccurlyeq I$.

From the above example, we can see that $\mathcal{P}(A \subseteq \boldsymbol{Y})$ is to compute the probability of any subset A being included in $\boldsymbol{Y}$. In other words, $\mathcal{P}(A \subseteq \boldsymbol{Y})$ compute the marginal probability of A belongs to $\boldsymbol{Y}$. Since K contains all the information of computing marginal probabilities, K is called marginal kernel. If $A = \{i\}$ is a singleton, then we have

$$\mathcal{P}(\{i\} \subseteq \boldsymbol{Y}) = K_{ii} \tag{3.2}$$

That is, the diagonal of K gives the marginal probabilities of $\boldsymbol{Y}$ containing the individual elements of $\mathcal{Y}$.

Diagonal entries close to 1 correspond to elements of $\mathcal{Y}$ that are almost always selected by the DPP. If $A = \{i, j\}$ is a two-element set, then

$$\mathcal{P}(i, j \subseteq \boldsymbol{Y}) = \begin{vmatrix} K_{ii} & K_{ij} \\ K_{ji} & K_{jj} \end{vmatrix} = K_{ii}K_{jj} - K_{ij}K_{ji} = \mathcal{P}(i \subseteq \boldsymbol{Y})\mathcal{P}(j \subseteq \boldsymbol{Y}) - K_{ij}^2 \tag{3.3}$$

Thus, the pairs of off-diagonal elements determine the negative correlations. That is, large values of K_{ij} imply that ith term and jth term tend not to be chosen by a DPP.

From (3.3), if we set the marginal kernel as measurements of similarity between pairs of elements in $\mathcal{Y}$, then the highly similar elements are rarely to appear together in DPPs. If $K_{ij} = \sqrt{K_{ii}K_{jj}}$, then ith term and jth term do not appear together almost surely in a DPP. On the contrary, if K is diagonal, which means that there are no correlations, the elements appear independently in DPPs. Overall, correlations, i.e. the degree of similarity between two terms, are always negative in DPPs.

Example 3.2 Let K measure the distance between points on the plane, which we divide the $[0,1]$ of x-axis and y-axis both to 40 terms, i.e. $x = \frac{1}{40}, \frac{2}{40}, \dots, \frac{40}{40}$ and $y = \frac{1}{40}, \frac{2}{40}, \dots, \frac{40}{40}$. Then combining the x coordinates and y coordinated to get the 1600 points, i.e. $(x_1, y_1), \dots, (x_{1600}, y_{1600})$. Then we use the DPP algorithm to get the series of points in figure 3.1 left, which L-ensemble is a 1600×1600 matrix with

$$L_{ij} = exp\{-((x_i - x_j)^2 + (y_i - y_j)^2)/0.05\}$$

Finally, we use the Poisson processing to independently get the series of points in figure 3.1 right. From these two graphs, we can see that the DPP points seem to uniformly fall in the plane. On the contrary, for the Poisson process, some points assemble in a place.

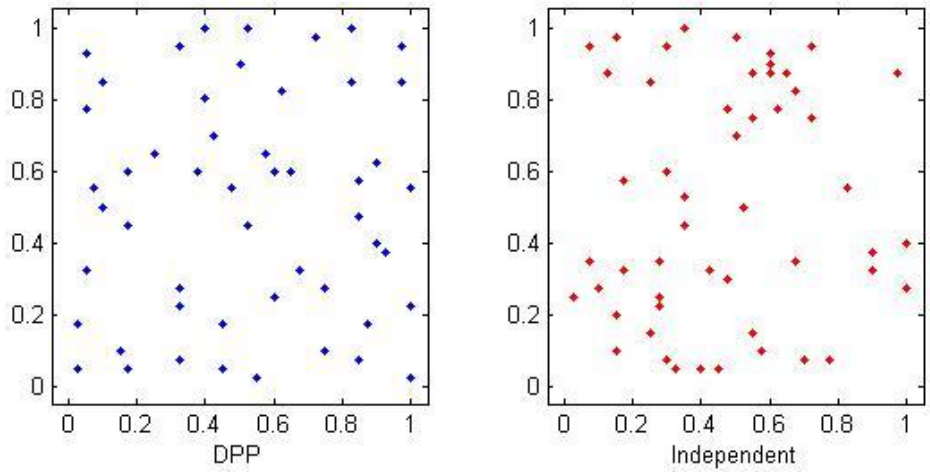

Figure 3.1 A set of points in the plane drawn from a DPP (left), and the same number of points sampled independently.

3.2 L-ensemble

For modeling real data, it is difficult to structure the marginal kernel K; therefore, L-ensemble is useful to define a DPP, where L is a real and symmetric matrix indexed by the elements of $\mathcal{Y}$:

$$\mathcal{P}_L(\boldsymbol{Y} = Y) \propto \det(L_Y) \tag{3.4}$$

As for K, it obviously shows that L must be positive semidefinite. However, since the equation 3.4 is only a statement of proportionality, the eigenvalues of L need not be less than one. In conclusion, any positive semidefinite L defines an L-ensemble.

From (3.4), we can explore the relationship between $\mathcal{P}_L(\boldsymbol{Y} = Y)\ and\ \det(L_Y)$. First, we assume $N = 2$, then we can obtain

$$\begin{aligned}\det(L+I) &= \begin{vmatrix} L_{11}+1 & L_{12} \\ L_{21} & L_{22}+1 \end{vmatrix} = \begin{vmatrix} L_{11} & L_{12} \\ L_{21} & L_{22}+1 \end{vmatrix} + \begin{vmatrix} 1 & 0 \\ L_{21} & L_{22}+1 \end{vmatrix} \\ &= \begin{vmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{vmatrix} + \begin{vmatrix} L_{11} & L_{12} \\ 0 & 1 \end{vmatrix} + \begin{vmatrix} 1 & 0 \\ L_{21} & L_{22} \end{vmatrix} + \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} \\ &= \det\left(L_{1,2}\right) + \det(L_1) + \det(L_2) + \det(L_\emptyset)\end{aligned} \tag{3.5}$$

Then, we continue assuming $N = 3,4,5,\cdots$. We can final obtain that

$$\det(L + I) = \sum_{Y \subseteq \mathcal{Y}} \det(L_Y) \tag{3.6}$$

From (3.6), we can also obtain that

$$\mathcal{P}_L(\boldsymbol{Y} = Y) = \frac{\det(L_Y)}{\sum_{Y \subseteq \mathcal{Y}} \det(L_Y)} = \frac{\det(L_Y)}{\det(L + I)} \tag{3.7}$$

Theorem 3.4 (Kulesza and Taskar, 2012) For any $A \subseteq \mathcal{Y}$,

$$\sum_{A \subseteq Y \subseteq \mathcal{Y}} \det(L_Y) = \det(L + I_{\bar{A}}) \tag{3.8}$$

Where $I_{\bar{A}}$ is the diagonal matrix with ones in the diagonal positions corresponding to elements of $\bar{A} = \mathcal{Y} - A$, and zeros everywhere else.

Proof: If $A = \mathcal{Y}$, (3.8) holds trivially. Now suppose that when we give A, $|\bar{A}| = k > 0$, then we let i be an element of $\mathcal{Y}$ where $i \in \bar{A}$. Then we can write:

$$L + I_{\bar{A}} = \begin{pmatrix} L_{ii} + 1 & L_{i\bar{\imath}} \\ L_{\bar{\imath}i} & L_{\mathcal{Y}-\{i\}} + I_{\mathcal{Y}-\{i\}-A} \end{pmatrix} \tag{3.9}$$

where $L_{\bar{\imath}i}$ is the sub-column of the ith column of L whose rows correspond to $\bar{\imath}$, and similarly for $L_{i\bar{\imath}}$.

Using the same way of equation 3.5, then,

$$\begin{aligned} \det(L + I_{\bar{A}}) &= \begin{vmatrix} L_{ii} & L_{i\bar{\imath}} \\ L_{\bar{\imath}i} & L_{\mathcal{Y}-\{i\}} + I_{\mathcal{Y}-\{i\}-A} \end{vmatrix} + \begin{vmatrix} 1 & 0 \\ L_{\bar{\imath}i} & L_{\mathcal{Y}-\{i\}} + I_{\mathcal{Y}-\{i\}-A} \end{vmatrix} \\ &= \det\left(L + I_{\overline{A \cup \{i\}}}\right) + \det\left(L_{\mathcal{Y}-\{i\}} + I_{\mathcal{Y}-\{i\}-A}\right) \end{aligned} \tag{3.10}$$

We can now apply the inductive hypothesis separately to each term, giving

$$\begin{aligned} \det(L + I_{\bar{A}}) &= \sum_{A \cup \{i\} \subseteq Y \subseteq \mathcal{Y}} \det(L_Y) + \sum_{A \subseteq Y \subseteq \mathcal{Y}-\{i\}} \det(L_Y) \\ &= \sum_{A \subseteq Y \subseteq \mathcal{Y}} \det(L_Y) \end{aligned} \tag{3.11}$$

Where we observe that every Y either contains i and is included only in the first sum or does not contain i and is included only in the second sum.

From (3.1) and (3.8), we have

$$\mathcal{P}(A \subseteq \boldsymbol{Y}) = \sum_{A \subseteq Y \subseteq \mathcal{Y}} \mathcal{P}_L(\boldsymbol{Y} = Y) = \det(K_A) \tag{3.12}$$

$$= \sum_{A \subseteq Y \subseteq \mathcal{Y}} \frac{\det(L_Y)}{\det(L+I)} = \frac{1}{\det(L+I)} \sum_{A \subseteq Y \subseteq \mathcal{Y}} \det(L_Y) \tag{3.13}$$

$$= \frac{\det(L + I_{\bar{A}})}{\det(L+I)} = \det\{(L + I_{\bar{A}})(L+I)^{-1}\} \tag{3.14}$$

When A is equal to $\mathcal{Y}$, we have

$$\det(K) = \det\{(L+0)(L+I)^{-1}\} \tag{3.15}$$

Therefore, for a DPP, we can directly let:

$$K = L(L+I)^{-1} = (L+I-I)(L+I)^{-1} = I - (L+I)^{-1} \tag{3.16}$$

$$L = K(I-K)^{-1} \tag{3.17}$$

Then we can obtain that:

$$\begin{aligned}
\mathcal{P}(A \subseteq \boldsymbol{Y}) &= \det\{(L + I_{\bar{A}})(L+I)^{-1}\} \\
&= \det\{I_{\bar{A}}(L+I)^{-1} + I - (L+I)^{-1}\} \\
&= \det\{I - I_A(L+I)^{-1}\} \\
&= \det\{I_{\bar{A}} + I_A K\} \\
&= \begin{vmatrix} I_{|\overline{A}| \times |\overline{A}|} & 0 \\ K_{A\overline{A}} & K_A \end{vmatrix} \\
&= \det(K_A)
\end{aligned}$$

The inverse part of (3.16) always exists due to the positive coefficient on the identity matrix. However, the inverse part of equation 3.17 may not exists if any eigenvalue of K achieve the upper bound of 1. For this circumstance, the DPP is not a L-ensemble. When we do the eigendecomposition of L to get $L = \sum_{n=1}^{N} \lambda_n \boldsymbol{v}_n \boldsymbol{v}_n^T$, then we can also obtain the eigendecomposition of $K = \sum_{n=1}^{N} \frac{\lambda_n}{\lambda_n + 1} \boldsymbol{v}_n \boldsymbol{v}_n^T$.

3.3 Example of L-ensemble

L-ensemble of a DPP gives that L is a positive semidefinite matrix. Then we can find a B being a $D \times N$ matrix such that $L = B^{\mathrm{T}}B$, where $D \leq N$. Denote the columns of B by B_i for $i = 1,2,\cdots,N$. Then

$$\mathcal{P}_L(\boldsymbol{Y} = Y) \propto \det(L_Y) = \det\left(B_i^{\mathrm{T}}B_j\right)_{i,j\in Y} \tag{3.18}$$

Therefore, any real matrix B can structure a L-ensemble for a DPP.

In many real situations, we hope the diversity is balanced against some underlying preferences for different items in $\mathcal{Y}$. Therefore, we want to do the decomposition of the DPP, which is called quality-diversity decomposition. From 3.18, we have $L = B^{\mathrm{T}}B$, where B is a $D \times N$ matrix. Then we can continue to decompose B_i for $i = 1,2,\cdots,N$. Then

$$B_i = q_i\phi_i \tag{3.19}$$

where $q_i \in \mathbb{R}^+$ is a quality term, and $\phi_i \in \mathbb{R}^D$ is a diversity feature vector and $\|\phi_i\|^2 = 1$. For this decomposition, the quality term represents the "goodness" of an item i. Then the kernel L is changed to

$$L_{ij} = B_i^{\mathrm{T}}B_j = q_i\phi_i^{\mathrm{T}}\phi_j q_j \tag{3.20}$$

Since $\|\phi_i\|^2 = 1$, $\phi_i^{\mathrm{T}}\phi_j \in [-1,1]$. We denote $S_{ij} = \phi_i^{\mathrm{T}}\phi_j$, where $S \in \mathbb{R}^{N\times N}$. Then

$$S_{ii} = \phi_i^{\mathrm{T}}\phi_i = 1 \tag{3.21}$$

$$L_{ii} = q_i\phi_i^{\mathrm{T}}\phi_i q_i = q_i^2 \tag{3.22}$$

$$S_{ij} = \phi_i^{\mathrm{T}}\phi_j = \frac{q_i\phi_i^{\mathrm{T}}\phi_j q_j}{q_i q_j} = \frac{L_{ij}}{\sqrt{L_{ii}L_{jj}}} \tag{3.23}$$

Therefore, we can use S and q_i to rewrite the equation 3.18:

$$\mathcal{P}_L(\boldsymbol{Y} = Y) \propto \det(L_Y) = \det\left(q_i \phi_i^{\mathrm{T}} \phi_j q_j\right)_{i,j \in Y} = \left(\prod_{i \in Y} q_i^2\right) \det(S_Y) \tag{3.24}$$

where the first term represents the quality part, and the second term represents the diversity of the selected items.

From above analysis, we can find that it is convenient and efficient to structure a L-ensemble of a DPP rather than directly finding the kernel K. Therefore, in this research, we focus on using the kernel L for sampling Y from a DPP.

3.4 Algorithm – Sample DPP

Due to Hough et al. [2009, Kulesza and Taskar, 2012], we have an efficient algorithm to sample a configuration Y from a DPP. The input to the algorithm is an eigendecomposition of the DPP kernel L, and $\boldsymbol{e}_i$ is the ith standard basis N-vector, which is all zeros except for a one in the ith position.

Algorithm 3.1 (Kulesza and Taskar, 2012) Sampling from a DPP

Input: eigendecomposition $\{(\boldsymbol{v}_n, \lambda_n)\}_{n=1}^N$ of L
$J \longleftarrow \emptyset$
For $n = 1,2,\cdots,N$
  $J \longleftarrow J \cup \{n\}$ with probability $\frac{\lambda_n}{\lambda_n + 1}$
End
$V \longleftarrow \{\boldsymbol{v}_n\}_{n \in J}$
$Y \longleftarrow \emptyset$
While $|V| > 0$
  Select i from $\mathcal{Y}$ with probability $\frac{1}{|V|} \sum_{\boldsymbol{v} \in V} \left(\boldsymbol{v}^{\mathrm{T}} \boldsymbol{e}_i\right)^2$
  $Y \longleftarrow Y \cup i$
  $V \longleftarrow V_\perp$, an orthonormal basis for the subspace of V orthogonal to $\boldsymbol{e}_i$
End
Output: Y

In this algorithm, we have two main loops. The first loop is to help us ensure how many points that we need to sample from a DPP. The second loop is that based on a set of orthonormal vectors V, we can sample Y from $\mathcal{Y}$. To ensure the result of this algorithm sampled from $\mathcal{P}_L$. We separate this algorithm to two parts.

Definition 3.5 (Kulesza and Taskar, 2012) A DPP is called elementary if every eigenvalue of its marginal kernel is in $\{0,1\}$. We write $\mathcal{P}^V$, where V is a set of orthonormal vectors, to denote an elementary DPP with marginal kernel $K^V = \sum_{\boldsymbol{v}\in V} \boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}$

From this definition, now we assume that there is a DPP kernel L. When we do orthonormal eigendecomposition, we can get $L = \sum_{n=1}^{N} \lambda_n \boldsymbol{v}_n \boldsymbol{v}_n^T$. Then $K = \sum_{n=1}^{N} \frac{\lambda_n}{\lambda_n+1} \boldsymbol{v}_n \boldsymbol{v}_n^T$. For each n, we get a number a from $[0,1]$. If the a is less than $\frac{\lambda_n}{\lambda_n+1}$, we denote $\lambda'_n = 1$; otherwise, $\lambda'_n = 0$. Then the new kernel $K' = \sum_{n=1}^{N} \lambda'_n \boldsymbol{v}_n \boldsymbol{v}_n^T$, which is the elementary DPP kernel.

Lemma 3.6 (Kulesza and Taskar, 2012) Let W_n for $n = 1,2,\cdots,N$ be an arbitrary sequence of $k \times k$ rank-one matrices, and let $(W_n)_i$ denote the ith column of W_n. Let $W_J = \sum_{n\in J} W_n$. Then

$$\det(W_J) = \sum_{n_1,n_2,\cdots,n_k\in J,distinct} det\left(\left[(W_{n_1})_1 (W_{n_2})_2 \cdots (W_{n_k})_k\right]\right) \tag{3.25}$$

Proof: Since $W_J = \sum_{\boldsymbol{n}\in J} W_n$, we can expand the first column of W_J as

$$\det(W_J) = \sum_{n\in J} det\left(\left[(W_n)_1 (W_J)_2 \cdots (W_J)_k\right]\right) \tag{3.26}$$

Then we continue to expand the 2nd, 3rd, $\cdots$, kth column of W_J. Then

$$\det(W_J) = \sum_{n_1,n_2,\cdots,n_k\in J} det\left(\left[(W_{n_1})_1 (W_{n_2})_2 \cdots (W_{n_k})_k\right]\right) \tag{3.27}$$

Since W_n is the rank-one matrix, the determinant of the mixture matrix is zero if it contains two or more columns of the same W_n. Therefore, $n_1, n_2, \cdots, n_k$ should be distinct.

Lemma 3.7 (Kulesza and Taskar, 2012) A DPP with kernel $L = \sum_{n=1}^{N} \lambda_n \boldsymbol{v}_n \boldsymbol{v}_n^{\mathrm{T}}$ is a mixture of elementary DPPs:

$$\mathcal{P}_L = \frac{1}{\det(L+I)} \sum_{J \subseteq 1,2,\cdots,N} \mathcal{P}^{V_J} \prod_{n \in J} \lambda_n \tag{3.28}$$

Where V_J denotes the set $\{\boldsymbol{v}_n\}_{n \in J}$.

Proof: Let $A \in \mathcal{Y}$ be an arbitrary set with $|A| = k$. Denote $W_n = [\boldsymbol{v}_n \boldsymbol{v}_n^{\mathrm{T}}]_A$ for $n = 1,2,\cdots,N$. Here all of the W_n have rank one. From the definition of the elementary DPP, the kernel $K^{V_J} = \sum_{\boldsymbol{n} \in J} W_n$. Therefore, the right-hand side of 3.28 should be

$$\begin{aligned}
&\frac{1}{\det(L+I)} \sum_{J \subseteq 1,2,\cdots,N} det\left(\sum\nolimits_{\boldsymbol{n} \in J} W_n\right) \prod_{n \in J} \lambda_n \\
&\quad = \frac{1}{\det(L+I)} \sum_{J \subseteq 1,2,\cdots,N} \sum_{n_1,n_2,\cdots,n_k \in J, distinct} det\left(\left[\left(W_{n_1}\right)_1 \left(W_{n_2}\right)_2 \cdots \left(W_{n_k}\right)_k\right]\right) \prod_{n \in J} \lambda_n \\
&\quad = \frac{1}{\det(L+I)} \sum_{n_1,n_2,\cdots,n_k=1, distinct}^{N} \det\left(\left[\left(W_{n_1}\right)_1 \left(W_{n_2}\right)_2 \cdots \left(W_{n_k}\right)_k\right]\right) \sum_{\{n_1,n_2,\cdots,n_k\} \in J} \prod_{n \in J} \lambda_n \\
&\quad = \frac{1}{\det(L+I)} \sum_{n_1,n_2,\cdots,n_k=1, distinct}^{N} \det\left(\left[\left(W_{n_1}\right)_1 \cdots \left(W_{n_k}\right)_k\right]\right) \frac{\lambda_{n_1}}{\lambda_{n_1}+1} \cdots \frac{\lambda_{n_k}}{\lambda_{n_k}+1} \prod_{n=1}^{N} (\lambda_n \\
&\quad + 1) \\
&= \sum_{n_1,n_2,\cdots,n_k=1, distinct}^{N} \det\left(\left[\frac{\lambda_{n_1}}{\lambda_{n_1}+1}\left(W_{n_1}\right)_1 \cdots \frac{\lambda_{n_k}}{\lambda_{n_k}+1}\left(W_{n_k}\right)_k\right]\right)
\end{aligned} \tag{3.29}$$

Here, we use the fact $\det(L+I) = \prod_{n=1}^{N} (\lambda_n + 1)$. By the definition of marginal kernel K in terms of the eigendecomposition of L, we have

$$det\left(\sum_{n=1}^{N} \frac{\lambda_n}{\lambda_n+1} W_n\right) = det(K_A) \tag{3.30}$$

Since A is an arbitrary set, (3.28) holds.

As we proof of lemma 3.7, we can see that it is equal to sample Y from an elementary DPP comparing with sampled from an original DPP.

Lemma 3.8 (Kulesza and Taskar, 2012) If $\boldsymbol{Y}$ is drawn according to an elementary DPP $\mathcal{P}^V$, then $|\boldsymbol{Y}| = |V|$ with probability one.

Proof: By the definition of DPP kernel, K^V has rank $|V|$. Therefore, $\mathcal{P}^V(Y \subseteq \boldsymbol{Y}) = 0$ whenever $|\boldsymbol{Y}| > |V|$. Hence, $|\boldsymbol{Y}| \leq |V|$. Then

$$\mathbb{E}(|\boldsymbol{Y}|) = \sum_{n=1}^{N} \mathbb{I}(n \in \boldsymbol{Y}) = \sum_{n=1}^{N} K_{nn}^V = \mathrm{tr}(K^V) = \sum_{n \in V} \|\boldsymbol{v}_n\|^2 = |V| \tag{3.31}$$

Therefore, $\mathbb{P}(|\boldsymbol{Y}| = |V|) = 1$, i.e., $|\boldsymbol{Y}| = |V|$ almost surely.

Based on the lemma 3.7 and 3.8, we can see that the first loop of algorithm helps us ensure the elementary DPP $\mathcal{P}^V$ with the orthonormal eigendecomposition of L. Since the elementary DPP has the fixed cardinality, the first loop also assures the numbers of items that we choose from the DPP. The second loop of algorithm is sampling Y from the elementary DPP $\mathcal{P}^V$.

Let us give an example how the algorithm works on the circle as figure 3.2. We first uniformly get 100 points on a unit circle and use the distance between points as the feature to structure the kernel L. For example, there are 100 points $(x_1, x_2, \dots, x_{100})$ on a unit circle. Then we can define the L-ensemble as

$$L_{ij} = exp\{\langle x_i, x_j \rangle / 0.1\}$$

where $\langle x_i, x_j \rangle$ is the dot product of coordinates of points x_i and x_j. The figure 3.2 shows the probability of catching a point in next loop of algorithm. The x-axis is the location for each point on the circle. The range is $[0,2\pi]$. From this picture, we can see that after we sample a point from the elementary DPP, the probability around this point has decreased. Here we use distance between points as our feature; therefore,

this result directly shows that if the features of two items are more similar, it is less likely to sample them in a DPP.

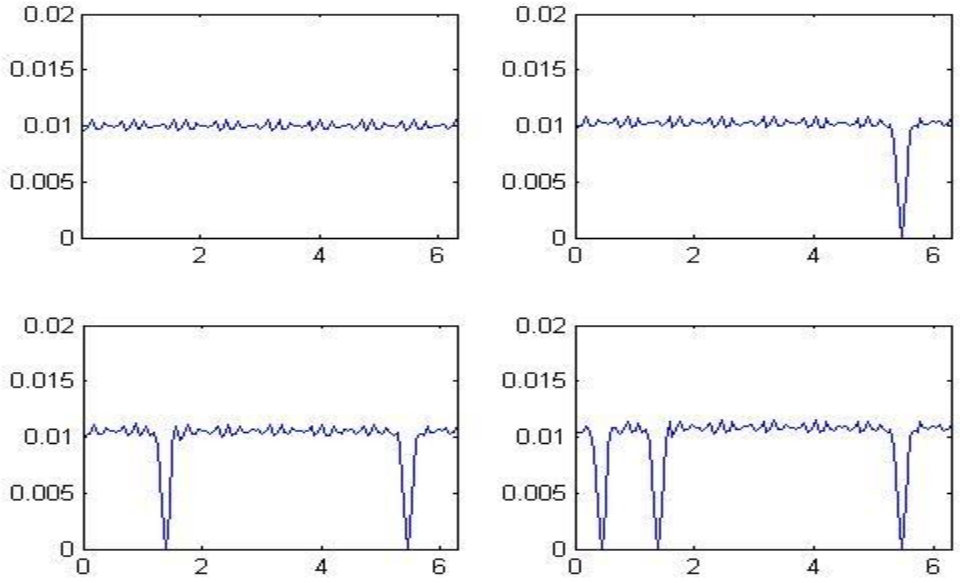

Figure 3.2 A DPP on a unit circle. The upper left is probability of sampling a point at the beginning. The upper right is the probability of sampling a point after 1 time.

4. Gap Distribution in Determinantal Point Process

Gap distribution is roughly speaking the distribution of the distance between two neighboring points. Gap distribution is an important aspect of the point process. For example, one-dimensional Poisson point process can be characterized in terms of the gap distribution done (definition 2.5). Therefore, we will study the gap distribution of determinantal point process (DPP) and compare it with the gap distribution from other DPP.

In the Poisson point process, the gap distribution shows that the sample points may gather in one location. For example, in figure 4.1, we sample a DPP and a Poisson point process with 42 points on the unit circle. Compared to the Poisson point process through the graphs, we can obviously see that the gaps of determinantal point process are almost uniform size. Therefore, we can roughly state that the DPP gives the balance gap distribution comparing with the Poisson point process.

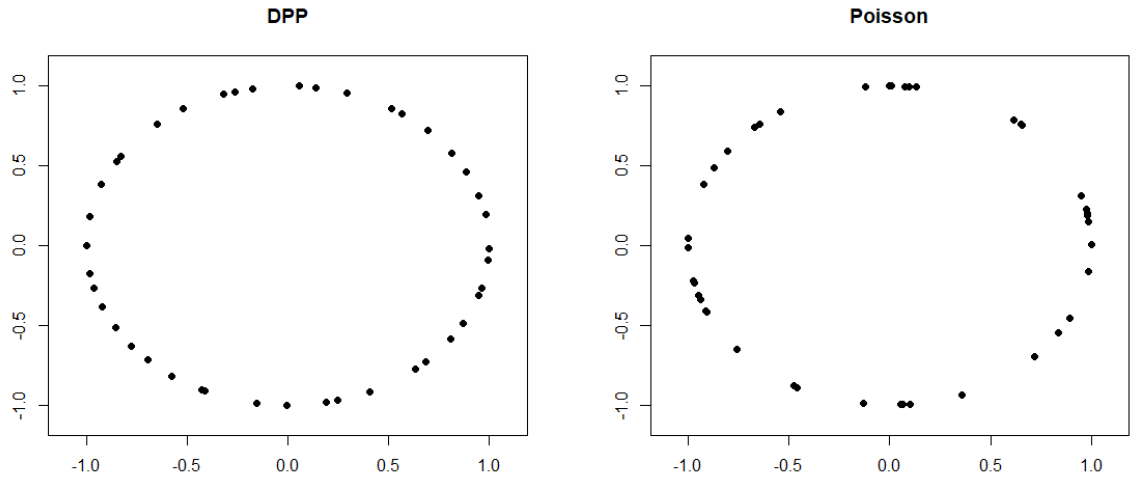

Figure 4.1 A DPP and a Poisson point processing form the same population with 10,000 points on a unit circle.

In this section we will explore the one-dimensional case and the two-dimensional case of the DPP. Same as what we assumed in the Poisson point process, we will sample the points from the population on the unit circle and on the unit sphere. For each case, we want to find three aspects of DPP. First, how many points will be sampled with different population sizes or scales? Second, for each case what distribution does the gap distribution of a DPP satisfy? Third, we will test if the gap distribution equals the distribution that we find in the second question.

In the Poisson point process, we simulate the points on a unit circle and a unit sphere because any point on a unit circle or a unit sphere could be considered as the initial/reference point. It is convenience for us to measure a point process without considering the bound of the population space. Therefore, a unit circle and a unit sphere can also be considered as the environment for exploring the gap distribution in the determinantal point process.

4.1 Gap distribution in one-dimensional DPP

We first explore the one-dimensional case. We randomly sample n points on the unit circle and use these n points to structure the L-ensemble. We then perform eigen-decomposition on the L-ensemble. Finally, we use the algorithm 3.1 to sample a determinantal point process. Since the n points are on the unit circle, we can use θ's with the range $[0,2\pi]$ to express their locations. For each point X_i it has a coordinate vector $X_i = \big(cos(\theta_i), sin(\theta_i)\big)^T$. For our case we want to sample the points almost uniformly on the circle; therefore, we can use the distance between points as the feature in L-ensemble. Given two points X_i, X_j from $X_1, \dots, X_n$, we write the inner product of these two points:

$$\langle X_i, X_j \rangle = |X_i||X_j| cos(\theta_i - \theta_j) = cos(\theta_i - \theta_j) = cos(\theta_j - \theta_i). \quad (4.1)$$

Since they are on the unit circle, $|X_i| = |X_j| = 1$ and $|\theta_i - \theta_j|$ is the arclength distance between X_i and X_j. We structure our L-ensemble as follows:

$$L_{ij} = exp\left(\frac{\langle X_i, X_j \rangle}{\sigma}\right) \quad (4.2)$$

where σ is a positive parameter. Then we can sample a DPP on a unit circle with the algorithm 4.1.

Algorithm 4.1 Sampling a DPP on a unit circle

Input: N, σ

Sample N points $X_1, \dots, X_N$ from $[0, 2\pi]$

Structure the L-ensemble $L_{ij} = exp\left(\frac{cos(X_i - X_j)}{\sigma}\right)$

Eigendecomposition $\{(\boldsymbol{v}_n, \lambda_n)\}_{n=1}^N$ of L

Using algorithm 3.1 with $\{(\boldsymbol{v}_n, \lambda_n)\}_{n=1}^N$ to get Y

Output: X_Y

From algorithm 3.1, we can find that the number of points in a DPP depends on the number of elements in J. For an matrix L we have that $\mathbb{E}(J) = \sum_{i=1}^N \frac{\lambda_i}{\lambda_i + 1}$. Here, we assume that $N = 1000, 2000, 3000, 4000$ and $\sigma = 0.1$. Then we operate algorithm 4.1, repeat the step "using algorithm 3.1" for a fixed matrix L and get table 4.2 (a). We continue operating algorithm 4.1 with the same repeating times and get table 4.2(b).

Table 4.2(a) The information of number of points in the DPP with an L

N	σ	m	$mean(length(Y))$	$var(length(Y))$	$\mathbb{E}(J)$
1000	0.1	100	36.8	1.1717	36.9669
2000	0.1	100	37.86	1.3135	37.9892
3000	0.1	100	38.46	1.1196	38.5799
4000	0.1	100	38.89	1.7151	38.9847

Table 4.2(b) The information of $\mathbb{E}(J)$ in the DPP with random L.

N	σ	m	$mean(\mathbb{E}(J))$	$var(\mathbb{E}(J))$
1000	0.1	100	36.9797	0.00013
2000	0.1	100	37.9979	0.00002
3000	0.1	100	38.5775	0.00001
4000	0.1	100	38.983	0.000006

From table 4.2(a), we can have a null hypothesis that the average number of points in the DPP with an L is equal to the expectation of J, i.e. $mean(length(Y)) = \mathbb{E}(J)$. Since for each DPP with the same L, the number of points in DPPs are independent and identically distributed. Hence, we can use the statistics method called $z - test$.

$$Z = \frac{mean\big(length(Y)\big) - \mathbb{E}(J)}{\sqrt{var\big(length(Y)\big)/m}} \sim N(0,1) \tag{4.3}$$

where $var\big(length(Y)\big)$ is the variance of number of points in the DPP with an L, and m is the number of the DPPs. Then, for $N = 1000, 2000, 3000, 4000$ and $\sigma = 0.1$, $|Z_N| = 1.54, 1.13, 1.13, 0.723$. Since $z = 1.96$ at the significant level 0.05, $|Z_N| < z$ shows that the average number of points in DPP with an L is equal to the expectation J at the significant level of 0.05. Then from table 4.2(b), we can have another null hypothesis that the average expectation of J with random L is equal to the average number of points in the DPP with an L, i.e. $mean(length(Y)) = mean(\mathbb{E}(J))$. Since N points are sampled randomly, L's are independent and identically distributed. $\mathbb{E}(J)$'s are also independent and identically distributed. Hence, we can also use the statistics z-test

$$Z = \frac{mean\big(length(Y)\big) - mean\big(\mathbb{E}(J)\big)}{\sqrt{var\big(length(Y)\big)/m + var\big(\mathbb{E}(J)\big)/m}} \sim N(0,1) \tag{4.4}$$

Where $var\big(\mathbb{E}(J)\big)$ is the variance of expectation of J. Then, for $N = 1000, 2000, 3000, 4000$ and $\sigma = 0.1$, $|Z_N| = 1.66, 1.20, 1.11, 0.71,$. Since $z = 1.96$ at the significant level 0.05, $|Z_N| < z$ shows that the average expectation J with random L is equal to the average number of points in the DPP with an L at the significant level of 0.05. Combine the above two hypotheses, we can have that the average number of points is equal for random L's.

From the above tests, we have that when we sample a DPP, the number of points in the DPP with the same N and σ do not depend on the random L. Therefore, we can let the following DPPs be sampled from the random L's.

For $N = 1000, 2000, 3000$ and $\sigma = 0.05, 0.1, 0.15, 0.2$, we separately sample $m = 200$ times DPP for each case. Then we can get the gap distribution with counterclockwise direction as table 4.3 and figure 4.4. Like the way that we work on the Poisson process, for each situation we use $mean(\mathbb{E}(J))$ to time the gaps and over the circumference.

$$x - axis: gaps \times \frac{n}{2\pi} \tag{4.5}$$

From table 4.3, the histograms show that for different N and σ the gap distributions give the almost same means and similar standard deviations, and the skewness and kurtosis are tiny different. Furthermore, from figure 4.4, we can see that the histograms for different N and σ show the similarity of curves. Then we can use the Gaussian kernel to estimate the density function for each gap distribution, and we can plot the density function and the cumulative distribution function for each gap distribution as figure 4.5. Then we can directly see that whatever density function or cumulative distribution function the gap distributions seem to be the same for different N and σ. Therefore, we can state a null hypothesis that for different N and σ the gap distribution gives the same shape.

According to Kolmogorov-Smirnov two-sample tests in Pratt's and Gibbons' (1981) book "Concepts of nonparametric theory", we can test two histograms have the same shape. Then we can get the P-values for different combinations of N and σ as table 4.6. From the results, we can find that there are no P-values less than 0.1 even though no P-values less than 0.05. Therefore, we cannot reject the null hypothesis.

Table 4.3 The statistic information of gap distribution for $N = 1000, 2000, 3000$ and $\sigma = 0.05, 0.1, 0.15, 0.2$

N	σ	m	mean	standard deviation	skewness	kurtosis
1000	0.05	200	1	0.4219	0.5289	3.2375
1000	0.1	200	1	0.4330	0.5787	3.2675
1000	0.15	200	1	0.4334	0.5973	3.3295
1000	0.2	200	1	0.4213	0.5904	3.3322
2000	0.05	200	1	0.4279	0.5082	3.1369
2000	0.1	200	1	0.4272	0.4985	3.1179
2000	0.15	200	1	0.4334	0.5426	3.1878
2000	0.2	200	1	0.4290	0.5593	3.2697
3000	0.05	200	1	0.4282	0.5238	3.1587
3000	0.1	200	1	0.4270	0.4979	3.1033
3000	0.15	200	1	0.4247	0.5125	3.2221
3000	0.2	200	1	0.4311	0.5789	3.2590

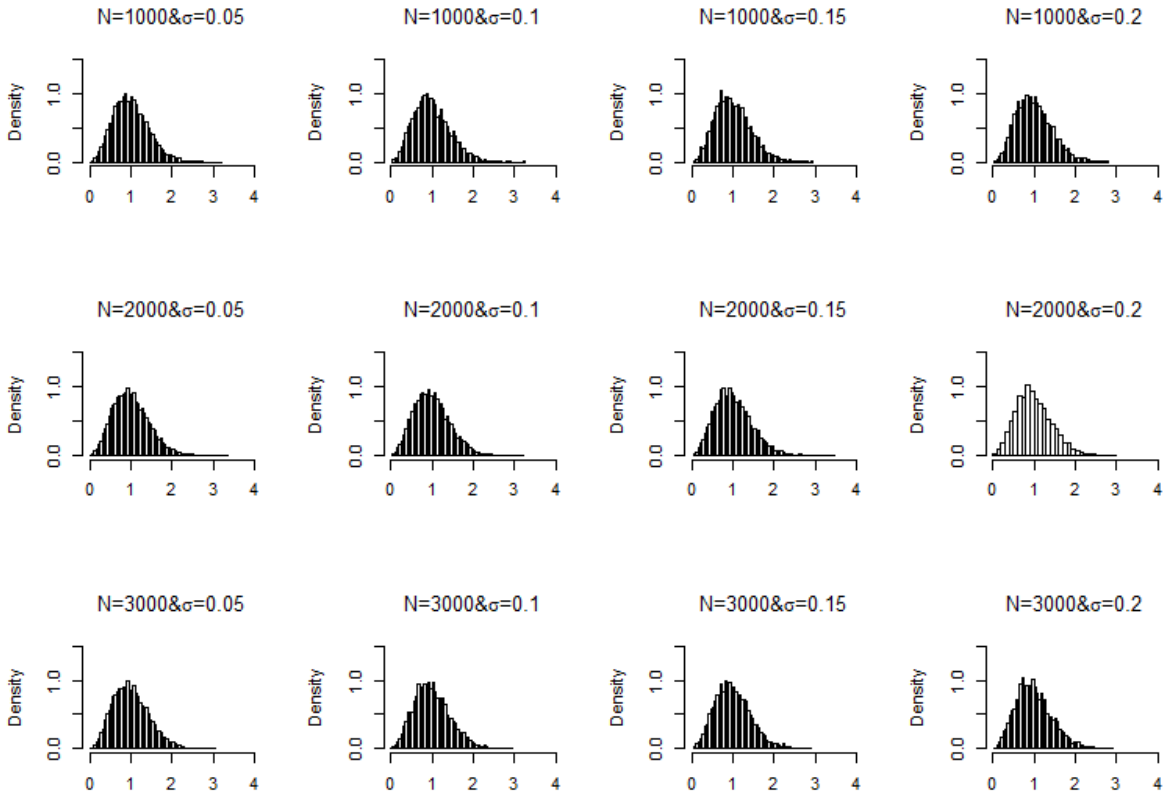

Figure 4.4 The histograms of gaps between neighboring points with counterclockwise direction for $N = 1000, 2000, 3000$ and $\sigma = 0.05, 0.1, 0.15, 0.2$.

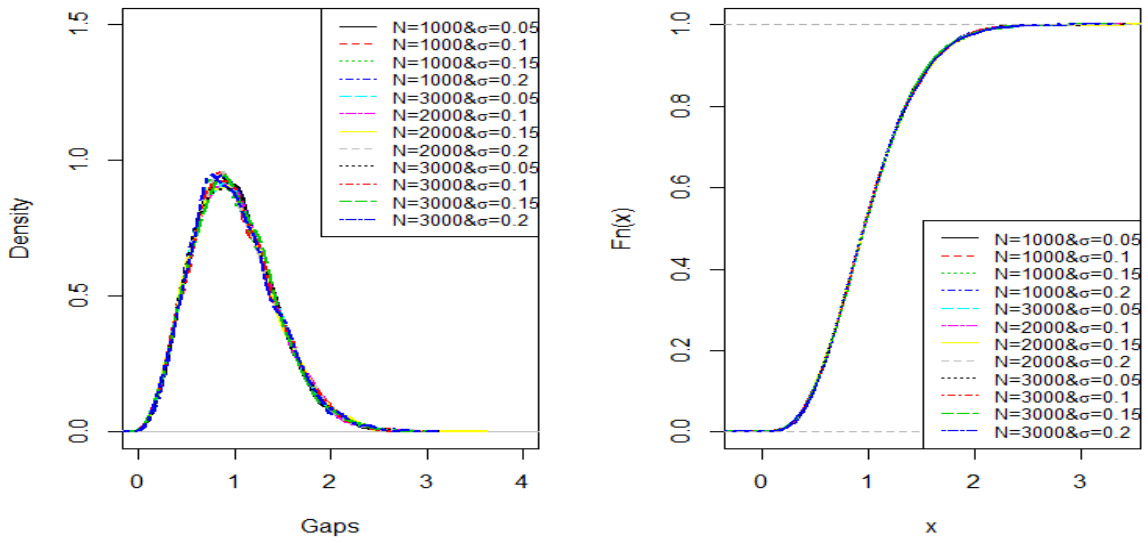

Figure 4.5 The plots of density function and cumulative distribution function (CDF) for $N = 1000, 2000, 3000$ and $\sigma = 0.05, 0.1, 0.15, 0.2$.

Table 4.6 The P-value of Kolmogorov-Smirnov two-sample tests for different combination of $N_{1,2,3} = 1000, 2000, 3000$ and $\sigma_{1,2,3,4} = 0.05, 0.1, 0.15, 0.2$.

$N\&\sigma$	$N_1\&\sigma_1$	$N_1\&\sigma_2$	$N_1\&\sigma_3$	$N_1\&\sigma_4$	$N_2\&\sigma_1$	$N_2\&\sigma_2$	$N_2\&\sigma_3$	$N_2\&\sigma_4$	$N_3\&\sigma_1$	$N_3\&\sigma_2$	$N_3\&\sigma_3$	$N_3\&\sigma_4$
$N_1\&\sigma_1$	1											
$N_1\&\sigma_2$	0.3869	1										
$N_1\&\sigma_3$	0.2353	0.9019	1									
$N_1\&\sigma_4$	0.963	0.668	0.7461	1								
$N_2\&\sigma_1$	0.3523	0.3731	0.3852	0.7154	1							
$N_2\&\sigma_2$	0.5196	0.2809	0.4376	0.7808	1.0000	1						
$N_2\&\sigma_3$	0.3542	0.7243	0.8225	0.6067	0.7837	0.8671	1					
$N_2\&\sigma_4$	0.7154	0.9601	0.6793	0.8957	0.7525	0.7647	0.7818	1				
$N_3\&\sigma_1$	0.2839	0.1755	0.3939	0.7217	0.9831	0.996	0.9032	0.6849	1			
$N_3\&\sigma_2$	0.8351	0.7342	0.4698	0.5317	0.8165	0.8386	0.7415	0.9047	0.6019	1		
$N_3\&\sigma_3$	0.8026	0.4422	0.3063	0.8311	0.836	0.8633	0.3481	0.6352	0.5876	0.8924	1	
$N_3\&\sigma_4$	0.3276	0.8835	0.8817	0.7536	0.579	0.6232	0.8046	0.6519	0.6588	0.7398	0.3049	1

Following our algorithm for sampling DPPs, we first randomly choose N points to structure the matrix L. Therefore, each DPP is sampled from the random matrix L. Circular unitary ensemble is a type of random complex matrix whose eigenvalues are uniformly on the unit circle. We can guess that the gap distribution of points that are sampled from the circular unitary ensemble (CUE) may give the same shape

as the gap distribution of a DPP. Then using the wolfram language and system, we can sample 100 points from the circular unitary ensemble (CUE) 100 times. Then same as what we do for the gaps of DPP, we use 100 to time the gaps of points sampled from the CUE. We can get the histogram of gap distribution, the graph of a density function, and the graph of cumulative distribution function as in figure 4.7. Comparing with the gap distribution of DPP, we can state a null hypothesis that the gap distribution of DPP has the same shape with the gap distribution of CUE. We can do Kolmogorov-Smirnov two-sample tests for this hypothesis. The P-values of comparing CUE with DPP for different $N = 1000, 2000, 3000$ and $\sigma = 0.05, 0.1, 0.15, 0.2$ are 0.8116, 0.3293, 0.5848, 0.7963, 0.6616, 0.7789, 0.3778, 0.7956, 0.5334, 0.7357, 0.9739, 0.6089. Since each P-value is not less than 0.05, we cannot reject our null hypothesis at a significant level of 0.05. Therefore, we can obtain that the gap distributions of DPP and CUE have the same shape at the significant level 0.05.

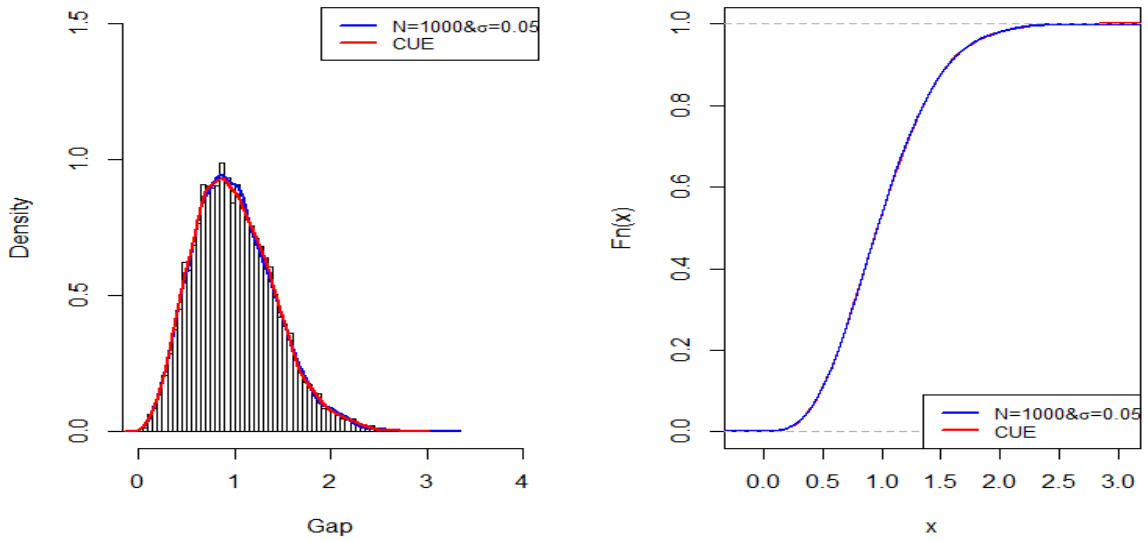

Figure 4.7 The histogram of gap distribution for CUE, graph of density function and graph of cumulative distribution function for CUE and DPP with $N = 1000$ & $\sigma = 0.05$.

For the Poisson point process, we have the gap distribution as an exponential distribution. For exponential distribution, the random gaps fall near the 0 with a higher probability. This situation makes many random points of the Poisson point process fall in the same place. However, if we do the determinantal point process, we can transfer the exponential distribution to the other distribution that gaps have a higher probability of falling around the mean of gaps like figure 4.8. Besides, according to table 4.3 and table 2.4(b), the means of gap distributions of these two point processes are the same. Therefore, the DPP improve the Poisson point process.

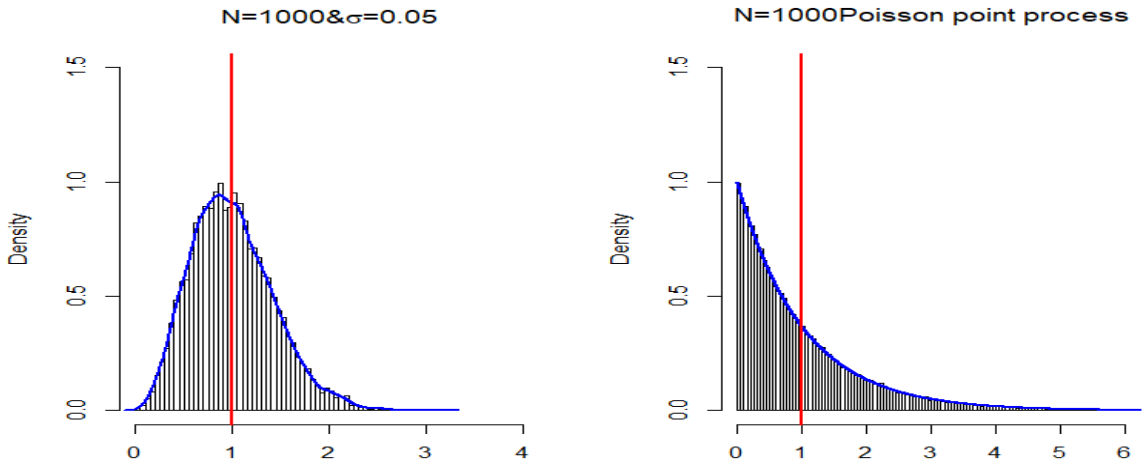

Figure 4.8 The gaps distributions for Poisson point process and DPP with mean.

In the one-dimensional case, we can find that the number of points sampled from a DPP is related to the expectation J, $\mathbb{E}(J) = \sum_{i=1}^{N} \frac{\lambda_i}{\lambda_i+1}$, where λ_i is the eigenvalue of L. Through Kolmogorov-Smirnov two-sample tests, we find that for different N and σ, the gap distribution of DPPs gives the same distribution as the circular unitary ensemble distribution. Comparing with the Poisson point process, the DPP gives the better gap distribution that gaps fall around the mean of gaps, which means that the points that we sampled nearly uniformly fall in the space.

4.2 Gap distribution in two-dimensional DPP

For the one-dimensional case, we find that the DPP optimizes the Poisson point process. Then for a higher-dimensional case such as the two-dimensional case, we do not ensure the statement also being true. Therefore, we need also to check the three aspects of DPPs mentioned at the beginning of this section.

For the two-dimensional case, we can sample a DPP from a unit sphere because every point on the unit sphere could be considered as the reference point such that it is convenience for us to measure the gaps between the closest points as what we do in the Poisson point process. Since we want the DPP to optimize the Poisson point process, we first sample n random points on the unit sphere. Then we use these points to structure the L. We continue doing eigen-decomposition on the L. Finally, using algorithm 3.1 to sample a determinantal point process on a unit sphere. Since the points are on the unit sphere, we can uniformly sample θ from $[0,2\pi]$ and h from $[-1,1]$. The coordinate vector of each point can be created by $(\sqrt{1-h^2}cos(\theta), \sqrt{1-h^2}sin(\theta), h)$; then we can use the inner product of coordinate vectors of points as the elements of L, same as (4.1) and (4.2). Therefore, for a unit sphere, we can sample a DPP with algorithm 4.2.

Algorithm 4.2 Sampling a DPP on a unit sphere

Input: N, σ

Sample N points $X_1, \dots, X_N$ with $\theta \in [0,2\pi]$ and $h \in [-1,1]$

Structure the L-ensemble $L_{ij} = exp\left(\frac{\sqrt{1-h_i^2}\sqrt{1-h_j^2}cos(\theta_i-\theta_j)+h_ih_j}{\sigma}\right)$

Eigendecomposition $\{(\boldsymbol{v}_n, \lambda_n)\}_{n=1}^N$ of L

Using algorithm 3.1 with $\{(\boldsymbol{v}_n, \lambda_n)\}_{n=1}^N$ to get Y

Output: X_Y

Since we also use algorithm 3.1 in algorithm 4.2, we can obtain that the number of DPP depends on the expectation J, i.e. $\mathbb{E}(J) = \sum_{i=1}^N \frac{\lambda_i}{\lambda_i+1}$ where λ_i is the eigenvalues of L. We assume that $N =$

$1000, 2000, 3000, 4000$ and $\sigma = 0.1$ and then do the same operation as the one-dimensional case to get table 4.9(a) and 4.9(b).

Table 4.9(a) The information of number of points in the DPP with a L

N	σ	m	$mean(length(Y))$	$var(length(Y))$	$\mathbb{E}(J)$
1000	0.1	100	309.86	25.4125	309.8348
2000	0.1	100	333.65	31.3207	333.4854
3000	0.1	100	346.09	27.7999	345.7298
4000	0.1	100	355.35	21.1995	354.4752

Table 4.9(b) The information of $\mathbb{E}(J)$ in the DPP with random L.

N	σ	m	$mean(\mathbb{E}(J))$	$var(\mathbb{E}(J))$
1000	0.1	100	309.0103	0.3635
2000	0.1	100	333.3693	0.0621
3000	0.1	100	345.8952	0.0244
4000	0.1	100	354.4919	0.0121

Since for each DPP with the same L, DPPs are independent and identically distributed. From table 4.9(a), we can have the null hypothesis that the average number of points in the DPP with an L is equal to the expectation J, i.e. $mean(length(Y)) = \mathbb{E}(J)$. Then by (4.3), for $N = 1000, 2000, 3000, 4000$ and $\sigma = 0.1$, $|Z_N| =$ 0.05, 0.2941, 0.6832,1.90. Since $z = 1.96$ at the significant level 0.05, $|Z_N| < z$ shows that for the two dimensional case, the average number of points in DPP with an L is equal to the expectation J at the significant level of 0.05. Then we sample N points randomly, the matrix L's are independent and identically distributed. By table 4.9(b), we can state a null hypothesis that the average number of points in the DPP with an L is equal to the average of expectation J for different L's. According to (4.4), for $N = 1000, 2000, 3000, 4000$ and $\sigma = 0.1$, $|Z_N| = 1.6736, 0.5011, 0.3693, 1.8632$. Since $|Z_N| \leq 1.96$, we cannot reject the null hypothesis at the significant level 0.05. We can state that the average number of points in the DPP with an L and the average

of expectation J for different L's are equal. Therefore, we can sample DPPs with random L for the following checking.

For $N = 1000, 2000, 3000$ and $\sigma = 0.1, 0.15, 0.2$, we separately sample 100 times DPP for each case. Then we can get the gap distribution of closest points as table 4.10 and figure 4.11. Here the x-axis is defined as

$$x - axis: gaps \times \frac{n}{4} \tag{4.6}$$

According to the Poisson point process in the two-dimensional case, we measure the square gaps timing the average number of points in DPPs over the area of the unit sphere as our gap distribution. From table 4.10, we can see that skewness for each case is positive which shows that the distribution is right-skewed. Each case's kurtosis is higher than 3 which means it has a higher perk than the normal distribution. Therefore, the more points fall near the mean. The mean, standard deviation, skewness, and kurtosis for each case does not exist big differences, and by figure 4.12, the histograms of all cases look like similar. We can make a guess that the gap distribution for different N and σ can be the same. Then we plot graph of density function and cumulative distribution function as figure 4.11 shows. The graphs for different cases in figure 4.11 are overlapped. Therefore, we can make a null hypothesis that the gap distribution for different N and σ has the same shape.

Table 4.10 The statistic information of gap distribution for $N = 1000, 2000, 3000$ and $\sigma = 0.1, 0.15, 0.2$

N	σ	m	mean	standard deviation	skewness	kurtosis
1000	0.1	100	1.412	0.2644	0.9096	4.0127
1000	0.15	100	1.442	0.2669	0.8599	3.8440
1000	0.2	100	1.423	0.2656	0.9282	3.9630
2000	0.1	100	1.437	0.2645	0.8639	3.8378
2000	0.15	100	1.438	0.2650	0.9022	4.0715
2000	0.2	100	1.431	0.2638	0.8723	3.7995
3000	0.1	100	1.445	0.2641	0.8617	3.9111
3000	0.15	100	1.437	0.2631	0.8492	3.7781
3000	0.2	100	1.429	0.2627	0.8925	3.9164

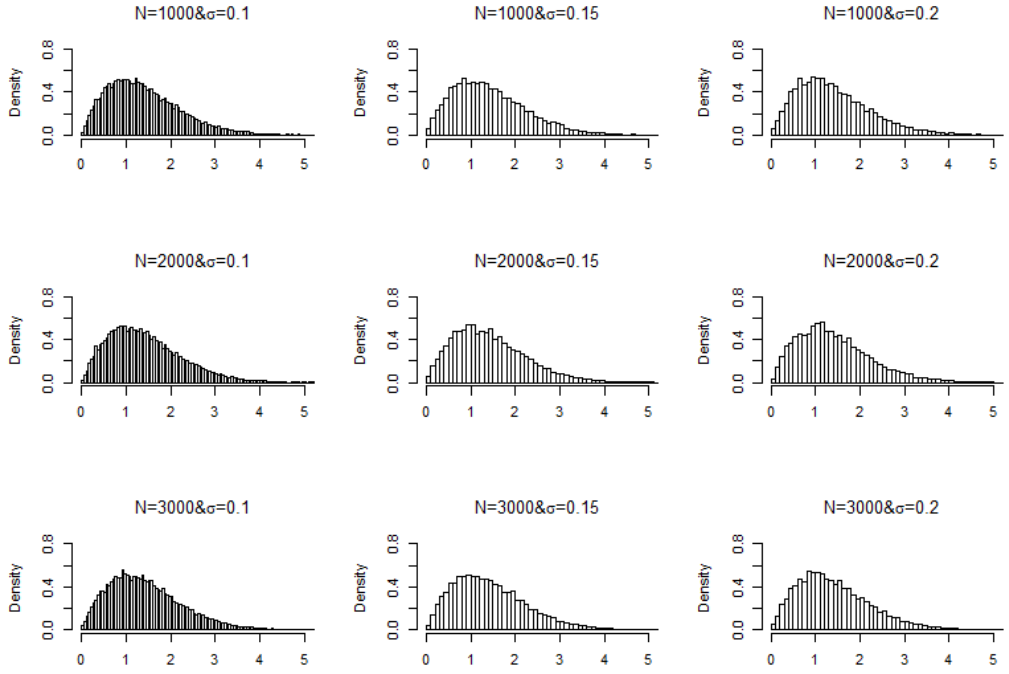

Figure 4.12 The histograms of gaps between closest points with counterclockwise direction for $N = 1000, 2000, 3000$ and $\sigma = 0.1, 0.15, 0.2$.

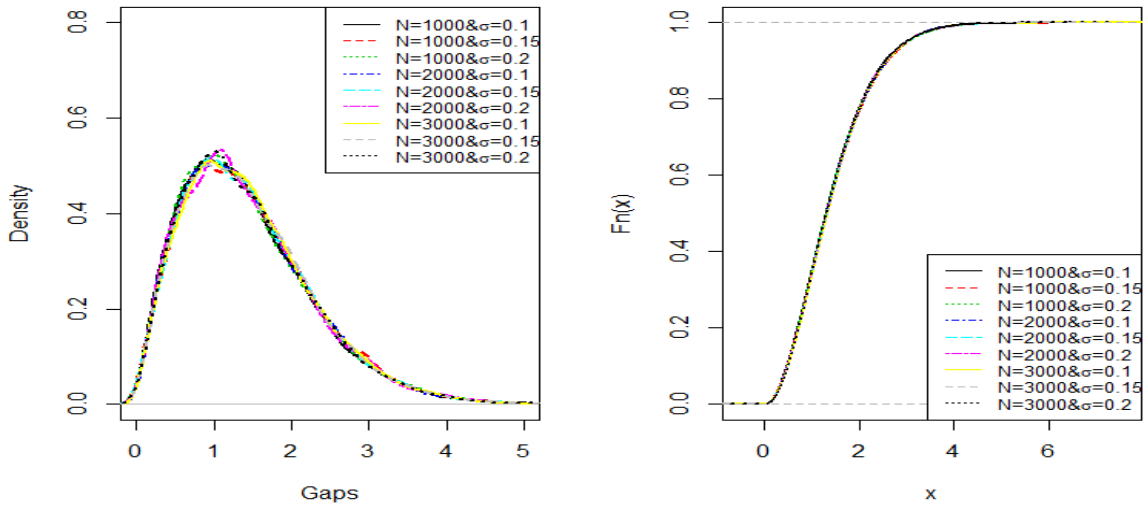

Figure 4.11 The plots of density function and cumulative distribution function (CDF) for $N = 1000, 2000, 3000$ and $\sigma = 0.1, 0.15, 0.2$.

According to Kolmogorov-Smirnov two-sample tests in Pratt's and Gibbons' (1981) book "Concepts of nonparametric theory", the two histograms with the same shape can be test. Then we combine different cases to get the table 4.13 about the P-values of Kolmogorov-Smirnov test. Then we can obtain that for any combination of N and σ, the gap distribution gives the same shape at the significant level of 0.05 because all P-values of Kolmogorov-Smirnov test are more than 0.05.

Table 4.13 The P-value of Kolmogorov-Smirnov two-sample tests for different combination of $N_{1,2,3} = 1000, 2000, 3000$ and $\sigma_{1,2,3} = 0.1, 0.15, 0.2$.

$N\&\sigma$	$N_1\&\sigma_1$	$N_1\&\sigma_2$	$N_1\&\sigma_3$	$N_2\&\sigma_1$	$N_2\&\sigma_2$	$N_2\&\sigma_3$	$N_3\&\sigma_1$	$N_3\&\sigma_2$	$N_3\&\sigma_3$
$N_1\&\sigma_1$	1								
$N_1\&\sigma_2$	0.5951	1							
$N_1\&\sigma_3$	0.2050	0.1241	1						
$N_2\&\sigma_1$	0.8601	0.5666	0.3242	1					
$N_2\&\sigma_2$	0.8397	0.7362	0.1203	0.8555	1				
$N_2\&\sigma_3$	0.8047	0.2360	0.1289	0.3645	0.3930	1			
$N_3\&\sigma_1$	0.7342	0.2822	0.2421	0.2645	0.2795	0.2139	1		
$N_3\&\sigma_2$	0.5601	0.6235	0.4132	0.7133	0.6067	0.2880	0.3179	1	
$N_3\&\sigma_3$	0.7244	0.7062	0.2815	0.1919	0.1572	0.5582	0.2043	0.07092	1

From the algorithm 4.2, we know that each DPP has a random L which is a matrix. Therefore, for the two-dimensional case, the type of random matrix is the Ginibre ensemble. The eigenvalues of a matrix from Ginibre ensemble are complex numbers, which fall in a unit circle on the complex plane. There is a hypothesis that the gap distribution of DPP in the two-dimensional case has the same shape as the gap distribution of Ginibre ensemble. According to Lozeve (2019), we can sample a series of points in a unit circle with Ginibre ensemble. Since there exists the bounds, we would delete the points on the bound to avoid the errors. The figure 4.14 shows that the density function and the cumulative distribution function of gap distribution of Ginibre ensemble compare with gap distributions of the DPPs. Then for different $N_{1,2,3} = 1000, 2000, 3000$ and $\sigma_{1,2,3} = 0.1, 0.15, 0.2$, we do the Kolmogorov-Smirnov test and get the P-value as 0.4725, 0.1508, 0.1063, 0.8333, 0.5555, 0.4717, 0.5265, 0.1823, 0.2969. Since all P-values are more than 0.05, we can accept the hypothesis that the gap distribution of DPP in the two-dimensional case and the gap distribution of Ginibre ensemble have the same shape.

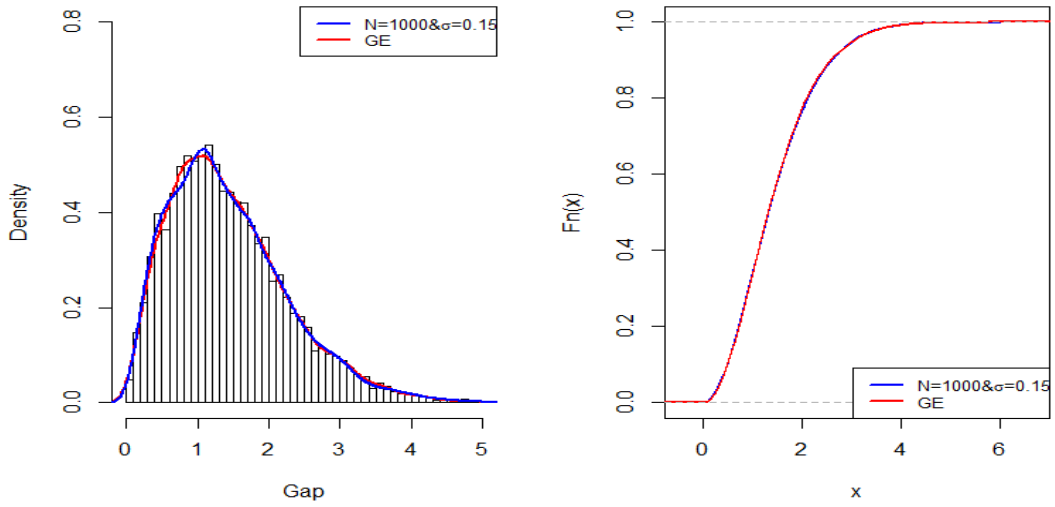

Figure 4.14 The histogram of gap distribution for GE, graph of density function and graph of cumulative distribution function for GE and DPP with $N = 1000$ & $\sigma = 0.05$.

For the two-dimensional Poisson points process, the gap distribution of exponential distribution which could be shown as the top of figure 4.15. We can also directly see the graph of gap distribution with same number of points between lines from the bottom of figure 4.15. Comparing these two graphs, the DPP make distances of points fall around the mean rather than near the 0. Although the mean has changed after we do DPP in the two-dimensional case, it does not affect showing that DPP improve the Poisson point process because the points in a DPP give more higher distance than the points in the Poisson point process.

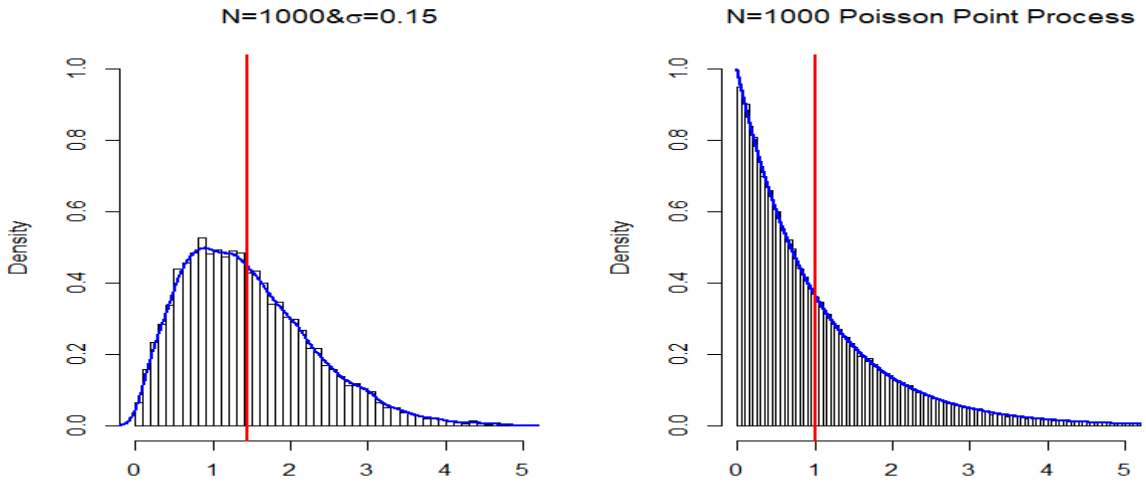

Figure 4.15 The gaps distributions for Poisson point process and DPP with mean.

In the two-dimensional case, we can find that the number of points sampled from a DPP is also related to the expectation J, $\mathbb{E}(J) = \sum_{i=1}^{N} \frac{\lambda_i}{\lambda_i+1}$, where λ_i is the eigenvalue of L. Through Kolmogorov-Smirnov two-sample tests, we find that for different N and σ, the gap distribution of DPPs gives the same distribution as the Ginibre ensemble distribution. Comparing with the Poisson point process, the DPP gives the better gap distribution that gaps fall around the mean of gaps, which means that the points that we sampled nearly uniformly fall in the space.

5. Conclusion

For the Poisson point process, we theoretically proved that the gap distribution is exponentially distributed. By the property of exponential distribution, the density probability decreases when the random variable increases. This implies that the points of a Poisson point process can form some clusters.

From section 3 and 4, we could see that the number of points in a DPP depends on the kernel matrix. The kernel matrix is difficult to be structured because the kernel matrix K should satisfy $0 \preccurlyeq K \preccurlyeq I$, which means that for any subset $A \subseteq \mathcal{Y}$, $\det(K_A) \in [0,1]$. According to Kulesza and Taskar (2012), we can find the indirect way L-ensemble instead of the kernel K, and we have that $K = L(L+I)^{-1}$.

We perform the point process on the unit circle for the one-dimensional case and on the unit sphere for the two-dimensional case. Such choice has benefits when calculating the gap since these spaces have no boundary. For the Poisson point process we confirmed that the gap distributions are the exponential distribution. Hence, when we do the determinantal point process on the unit circle, the gap distribution converges to one of the circular unitary ensemble; and when we do the determinantal point process on the unit sphere, the gap distribution converges to one of Ginibre ensemble. To support this conclusion, we use the Kolmogorov-Smirnov two-sample tests in section 4. Through figure 4.7 and 4.14, we can directly see that the determinantal point process gives more uniform point distribution the Poisson point process. The skewness and kurtosis of the gap distribution of the DPP are approximately 0.5, 3.2 for one-dimensional case and 0.9, 3.9 for two-dimensional case. However, the skewness and kurtosis of the gap distribution of

the Poisson point process are approximately 2, 9 for one- and two-dimensional case. According to the skewness and kurtosis, we can also obtain that the DPP could give more uniformly distributed points than those in the Poisson point process.

In this paper, we just talked about the one- and two- dimensional cases. Therefore, the above conclusions just suit for the one- and two- dimensional cases. However, for n dimensional case, we can assume to do the point process on the unit n-sphere. For example, we can do the Poisson point process on three-dimensional cases. In the real space, we assume that $X_1, \dots, X_n, \dots$ is a Poisson point process, and we assume that they have orders by the distance from X_1. We set X_1 as the reference point and let $d > 0$ be the linear distance between X_1 and X_2. Let

$$M_d = N(S_d) \tag{5.1}$$

where S_d is the sphere volume of radius d, i.e. S_d has the volume $\frac{4}{3}\pi d^3$. Therefore

$$\mathbb{P}(M_d = n) = \frac{\left(\lambda \frac{4}{3}\pi d^3\right)^n exp\left(-\lambda \frac{4}{3}\pi d^3\right)}{n!} \tag{5.2}$$

Let R_i be the linear distance between X_1 and X_{i+1}. We want $R_i \leq d$ if and only if $M_d \geq i$. Let $i = 1$; then,

$$\mathbb{P}(R_1^3 \leq d^3) = \mathbb{P}(R_1 \leq d) = \mathbb{P}(M_d \geq 1) = 1 - e^{-\lambda \frac{4}{3}\pi d^3} \tag{5.3}$$

In other words, R_1^3 is exponential$(\frac{4}{3}\lambda\pi)$ distribution. Then we can use a Poisson point process on the 4-sphere to check the gap distribution, which should be exponential distribution. Then we guess that the determinantal point process in higher dimension can also give more uniformly distributed points than those in the Poisson point process.

References

[1] Afshang, M., Saha, C., & Dhillon, H. S. (2017). *Nearest-Neighbor and Contact Distance Distributions for Matérn Cluster Process.* IEEE Communications Letters, *21*(12), 2686-2689. doi:10.1109/lcomm.2017.2747510

[2] Borodin, A., & Sinclair, C. D. (2009). *The Ginibre Ensemble of Real Random Matrices and its Scaling Limits*. Communications in Mathematical Physics, 291(1), 177-224. doi:10.1007/s00220-009-0874-5

[3] Diaconis, P. (2003). *Patterns in eigenvalues: The 70th Josiah Willard Gibbs lecture.* Bulletin of the American Mathematical Society, 40(02), 155-179. doi:10.1090/s0273-0979-03-00975-3

[4] Gallager, R. G. (1999). *Discrete stochastic processes*. Boston: Kluwer Acad. Publ.

[5] Hough, J., Krishnapur, M., Peres, Y., & Virág, B. (2009). *Determinantal point processes*. University Lecture Series Zeros of Gaussian Analytic Functions and Determinantal Point Processes, 47-81. doi:10.1090/ulect/051/04

[6] *Introduction to spatial point processes.* (n.d.). Retrieved September 9, 2020, from http://www.nisox.org/files/PDF/JohnsonSpatialPointProc.pdf

[7] Johansson, K. (2002). *Non-intersecting paths, random tilings and random matrices.* Probability Theory and Related Fields, 123(2), 225-280. doi:10.1007/s004400100187

[8] Kulesza, A. (2012). *Determinantal point processes for machine learning*. Place of publication not identified: World Scientific.

[9] Last, G., & Brandt, A. (2005). *Marked point processes on the real line: The dynamic approach.* New York: Springer.

[10] Last, G., & Penrose, M. (2018). *Lectures on the Poisson process.* Cambridge: Cambridge University Press.

[11] Lozeve, D. (2019, March 20). *Random matrices from the Ginibre ensemble.* Retrieved September 10, 2020, from https://www.lozeve.com/posts/ginibre-ensemble

[12] Macchi, O. (1975). *The coincidence approach to stochastic point processes*. Advances in Applied Probability, 7(01), 83-122. doi:10.1017/s0001867800040313

[13] Mariet, Z., & Sra, S. (n.d.). *Kronecker Determinantal Point Processes.* Retrieved August 9, 2020, from https://papers.nips.cc/paper/6296-kronecker-determinantal-point-processes.pdf

[14] Pratt, J. W., & Gibbons, J. D. (1981). *Concepts of nonparametric theory*. New York: Springer-Verlag.

[15] *The Poisson process.* (1984). Milton Keynes: Open University Press.

[16] (n.d.). Retrieved September 09, 2020, from https://reference.wolfram.com/language/ref/CircularUnitaryMatrixDistribution.html.en?source=foot

Appendix A: Codes

A1. Codes for Sample a Poisson point process on a unit circle (in R)

```
Pois = 0
N = 1000                    # Input N
a = runif(N,0,2*pi)         # Sample N points from [0,2π]
a= sort(a)
a1 = rep(0,N)
# Calculate the gaps
for (j in 1:(N-1)) {
   a1[j]= a[j+1]-a[j]
}
a1[N]= 2*pi-a[N]+a[1]
a1 = sort(a1)
a1 = N*a1
Pois = c(Pois,a1)
Pois = Pois[-1]
```

A2. Codes for Sample a Poisson point process on a unit sphere (in R)

```
N = 1000                    # Input N
# Sample N points
z = runif(N,-1,1)
x = rep(0,N)
y = rep(0,N)
for (i in 1:N) {
   a = runif(1,0,2*pi)
   x[i] = sqrt(1- z[i]^2)*cos(a)
   y[i]= sqrt(1- z[i]^2)*sin(a)
}
# Calculate the gaps
Pois = rep(0, N)
for (i in 1:N) {
   c = rep(0,N)
   for (j in 1:N) {
      if(i==j) {
         c[j]=0
      }else{
```

```
        a = x[i]*x[j]+y[i]*y[j]+z[i]*z[j]
        c[j] = acos(a)
      }
    }
    c1 = sort(c)
    Pois[i] = c1[2]
}
```

A3. Codes for Doing the eigendecomposition of L (In R & In MATLAB)

(1) In R

```
DecomposeKernel = function(A) {
    L1 = A
    B = eigen(A)
    L2 = B$values
    L3 = B$vectors
    L = list(L1,L2,L3)
    names(L) = c("L1","L2","L3")
    return(L)
}
```

(2) In MATLAB (Kulesza and Taskar, 2012)

```
function L = decompose_kernel(M)
    L.M = M;
    [V,D] = eig(M);
    L.V = real(V);
    L.D = real(diag(D));
```

A4. Codes for sample a DPP (In MATLAB)

```
# Codes are from Kulesza and Taskar (2012)
function Y = sample_dpp(L)
% sample a set Y from a dpp.    L is a decomposed kernel¡£
% choose eigenvectors randomly
D = L.D ./ (1+L.D);
v = find(rand(length(D),1) <= D);
k = length(v);
V = L.V(:,v);
% iterate
Y = zeros(k,1);
for i = k:-1:1
    % compute probabilities for each item
    P = sum(V.^2,2);
```

```
    P = P / sum(P);
    % choose a new item to include
    Y(i) = find(rand <= cumsum(P),1);
    % choose a vector to eliminate
    j = find(V(Y(i),:),1);
    Vj = V(:,j);
    V = V(:,[1:j-1 j+1:end]);
    % update V
    V = V - bsxfun(@times,Vj,V(Y(i),:)/Vj(Y(i)));
    % orthogonalize
    for a = 1:i-1
        for b = 1:a-1
            V(:,a) = V(:,a) - V(:,a)'*V(:,b)*V(:,b);
        end
        V(:,a) = V(:,a) / norm(V(:,a));
    end
end
Y = sort(Y);
```