Determinantal point process models and statistical inference

Frédéric Lavancier,
Laboratoire de Mathématiques Jean Leray, Nantes
Equipe Serpico, Inria Rennes

Joint works with **Christophe Biscio**, **Jesper Møller** and **Ege Rubak** (Aalborg University, Denmark).

UNIVERSITÉ DE NANTES

GDR GeoSto

In short

- DPPs are general models for random processes with **negative dependencies**.
- Their general definition has been introduced by O. Macchi in 1975 to model fermions (i.e. particles in repulsion)
- DPPs arise in many examples of Probability theory (much before 1975)
- They have been used for statistical purposes more recently (for about 5 years).

Example 1: the "descent subsequence"

Aim : sample points with negative dependence from a regular grid of N points.

Example 1: the "descent subsequence"

Aim : sample points with negative dependence from a regular grid of N points.

1. Draw $N+1$ random numbers from $\{0,\ldots,9\}$

1 1 6 3 9 1 4 0 1 4 7 0 4 3 3 8 7 9 9 4 3 9 4 1 6 8 9 0 2 1 0

$\circ$ $\circ$

Example 1: the "descent subsequence"

Aim : sample points with negative dependence from a regular grid of N points.

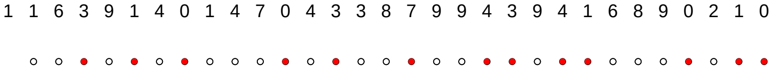
1. Draw $N + 1$ random numbers from $\{0, \ldots, 9\}$
2. Keep those points where there is a "descent"

1 1 6 3 9 1 4 0 1 4 7 0 4 3 3 8 7 9 9 4 3 9 4 1 6 8 9 0 2 1 0

Example 1: the "descent subsequence"

Aim : sample points with negative dependence from a regular grid of N points.

1. Draw $N+1$ random numbers from $\{0,\ldots,9\}$
2. Keep those points where there is a "descent"



This is a DPP!

With $N = 100$.

From the "descent subsequence"

From an independent sampling

Example 2: the Ginibre DPP

Let M be an $n \times n$ matrix with iid complex Gaussian entries.
Then the n eigenvalues of M are

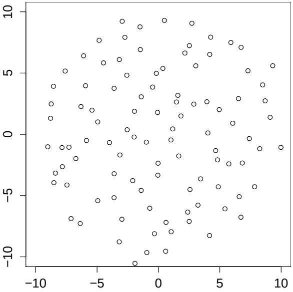
- "mainly" concentrated in the complex disk with radius $\sqrt{n}$
- distributed as a DPP on the complex plane

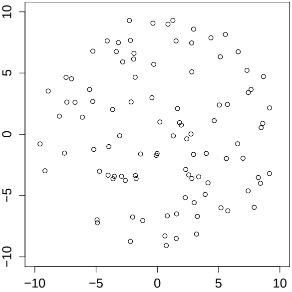
Example 2: the Ginibre DPP

Let M be an $n \times n$ matrix with iid complex Gaussian entries.
Then the n eigenvalues of M are

- "mainly" concentrated in the complex disk with radius $\sqrt{n}$
- distributed as a DPP on the complex plane

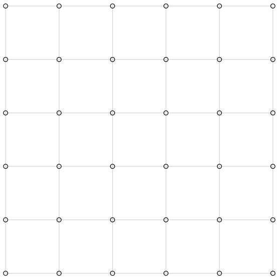
Left : Ginibre DPP ($n = 100$). Right : 100 iid points in the disk.





Example 3: spanning tree

Aim : sample edges (with negative dependence) from a graph

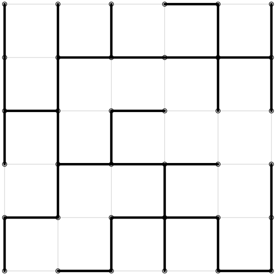


Example 3: spanning tree

Aim : sample edges (with negative dependence) from a graph

Random spanning tree

"tree" = any 2 vertices are connected by exactly one path
"spanning" = all vertices are connected
"random" = uniform over all possible spanning trees

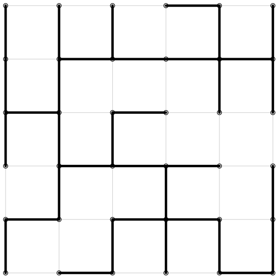


Example 3: spanning tree

Aim : sample edges (with negative dependence) from a graph

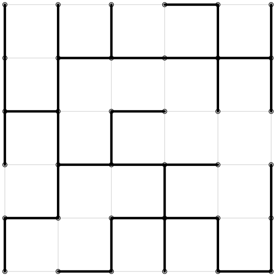
Random spanning tree

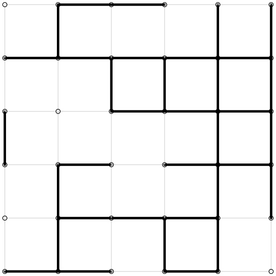
"tree" = any 2 vertices are connected by exactly one path
"spanning" = all vertices are connected
"random" = uniform over all possible spanning trees



The repartition of the edges on the graph is a DPP!

Left: spanning tree. Right: independent sampling of the edges





Why DPPs may be of interest for statisticians?

Why DPPs may be of interest for statisticians?

- Random matrices naturally arise in Statistics

Why DPPs may be of interest for statisticians?

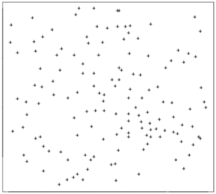
- Random matrices naturally arise in Statistics
- In Machine Learning : for instance, by imposing diversity (through a prior with negative dependencies) in the categories of an unsupervised classification problem.

Why DPPs may be of interest for statisticians?

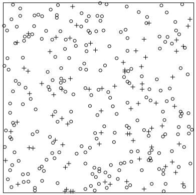
- Random matrices naturally arise in Statistics
- In Machine Learning : for instance, by imposing diversity (through a prior with negative dependencies) in the categories of an unsupervised classification problem.
- In surveys or random designs : negative dependencies in the sampling may reduce the variance of an estimate (ex : Monte Carlo approximation of the integral of a smooth function)

Why DPPs may be of interest for statisticians?

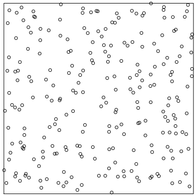
- Random matrices naturally arise in Statistics
- In Machine Learning : for instance, by imposing diversity (through a prior with negative dependencies) in the categories of an unsupervised classification problem.
- In surveys or random designs : negative dependencies in the sampling may reduce the variance of an estimate (ex : Monte Carlo approximation of the integral of a smooth function)
- Some real-data naturally involve negative dependencies:



Cellular Network



Kidney cells (hamster)



Oak and Beech trees

Gibbs point processes vs DPPs (on $\mathbb{R}^d$)

Gibbs point processes: The usual class when modelling repulsiveness on $\mathbb{R}^d$ (e.g. Strauss model).

Gibbs point processes vs DPPs (on $\mathbb{R}^d$)

Gibbs point processes: The usual class when modelling repulsiveness on $\mathbb{R}^d$ (e.g. Strauss model).

In general:

- moments are not expressible in closed form;
- likelihoods involve intractable normalizing constants;
- elaborate McMC methods are needed for simulations and approximate likelihood inference;
- for infinite Gibbs point processes defined on $\mathbb{R}^d$, 'things' become rather complicated (existence and uniqueness)

Gibbs point processes vs DPPs (on $\mathbb{R}^d$)

DPPs possess a number of appealing properties:

Gibbs point processes vs DPPs (on $\mathbb{R}^d$)

DPPs possess a number of appealing properties:

(a) simple conditions for existence;

Gibbs point processes vs DPPs (on $\mathbb{R}^d$)

DPPs possess a number of appealing properties:

(a) simple conditions for existence;

(b) all orders of moments are known;

Gibbs point processes vs DPPs (on $\mathbb{R}^d$)

DPPs possess a number of appealing properties:

(a) simple conditions for existence;

(b) all orders of moments are known;

(c) the density of the DPP restricted to any compact set is expressible on closed form;

Gibbs point processes vs DPPs (on $\mathbb{R}^d$)

DPPs possess a number of appealing properties:

(a) simple conditions for existence;

(b) all orders of moments are known;

(c) the density of the DPP restricted to any compact set is expressible on closed form;

(d) the DPP on any compact set can easily be simulated;

Gibbs point processes vs DPPs (on $\mathbb{R}^d$)

DPPs possess a number of appealing properties:

(a) simple conditions for existence;

(b) all orders of moments are known;

(c) the density of the DPP restricted to any compact set is expressible on closed form;

(d) the DPP on any compact set can easily be simulated;

(e) parametric models are available, and inference can be done by MLEs or using the moments.

Definition of a DPP on $\{1, \dots, N\}$

A point process on $\{1, \dots, N\}$ is a probability measure on the set of all subsets of $\{1, \dots, N\}$.

Definition of a DPP on $\{1,\ldots,N\}$

A point process on $\{1,\ldots,N\}$ is a probability measure on the set of all subsets of $\{1,\ldots,N\}$.

Let K be a $N\times N$ matrix.

Definition

X is a DPP on $\{1,\ldots,N\}$ with kernel matrix K if for any subset A of $\{1,\ldots,N\}$

$$P(A\subset X)=\det K_A$$

where $K_A=[K_{ij}]_{i,j\in A}$.

Definition of a DPP on $\{1, \ldots, N\}$

A point process on $\{1, \ldots, N\}$ is a probability measure on the set of all subsets of $\{1, \ldots, N\}$.

Let K be a $N \times N$ matrix.

Definition

X is a DPP on $\{1, \ldots, N\}$ with kernel matrix K if for any subset A of $\{1, \ldots, N\}$

$$P(A \subset X) = \det K_A$$

where $K_A = [K_{ij}]_{i,j \in A}$.

Interpretation: for $i, j \in \{1, \ldots, N\}$,

$K_{ii} = P(i \in X)$

$K_{ij} \approx$ measure of similarity between i and j. If K is symmetric

$$P(i, j \in X) = \begin{vmatrix} K_{ii} & K_{ij} \\ K_{ji} & K_{jj} \end{vmatrix} = P(i \in X)P(j \in X) - K_{ij}^2$$

Definition of a DPP on $\mathbb{R}^d$

In the following, we focus on DPPs on the continuous space $\mathbb{R}^d$

Definition of a DPP on $\mathbb{R}^d$

In the following, we focus on DPPs on the continuous space $\mathbb{R}^d$

Notation :

- X : spatial point process on $\mathbb{R}^d$

Definition of a DPP on $\mathbb{R}^d$

In the following, we focus on DPPs on the continuous space $\mathbb{R}^d$

Notation :

- X : spatial point process on $\mathbb{R}^d$
- For any borel set $B \subseteq \mathbb{R}^d$, $X_B = X \cap B$.

Definition of a DPP on $\mathbb{R}^d$

In the following, we focus on DPPs on the continuous space $\mathbb{R}^d$

Notation :

- X : spatial point process on $\mathbb{R}^d$
- For any borel set $B \subseteq \mathbb{R}^d$, $X_B = X \cap B$.
- For $n > 0$, $\rho^{(n)}$ is the n'th *order product density function* of X.
  Intuitively,
  $$\rho^{(n)}(x_1, \ldots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n$$
  is the probability that for each $i = 1, \ldots, n$,
  X has a point in a region around x_i of volume $\mathrm{d}x_i$.

Definition of a DPP on $\mathbb{R}^d$

In the following, we focus on DPPs on the continuous space $\mathbb{R}^d$

Notation :

- X : spatial point process on $\mathbb{R}^d$
- For any borel set $B \subseteq \mathbb{R}^d$, $X_B = X \cap B$.
- For $n > 0$, $\rho^{(n)}$ is the n'th *order product density function* of X. Intuitively,

$$\rho^{(n)}(x_1, \ldots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n$$

  is the probability that for each $i = 1, \ldots, n$,
  X has a point in a region around x_i of volume $\mathrm{d}x_i$.

  In particular $\rho = \rho^{(1)}$ is the *intensity function*.

Definition of a DPP on $\mathbb{R}^d$

Let C be a function from $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$.
Denote $[C](x_1, \ldots, x_n)$ the $n \times n$ matrix with entries $C(x_i, x_j)$.

$$\text{Ex :} \quad [C](x_1) = C(x_1, x_1) \quad [C](x_1, x_2) = \begin{pmatrix} C(x_1, x_1) & C(x_1, x_2) \\ C(x_2, x_1) & C(x_2, x_2) \end{pmatrix}.$$

Definition of a DPP on $\mathbb{R}^d$

Let C be a function from $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$.
Denote $[C](x_1, \dots, x_n)$ the $n \times n$ matrix with entries $C(x_i, x_j)$.

$$\text{Ex :} \quad [C](x_1) = C(x_1, x_1) \quad [C](x_1, x_2) = \begin{pmatrix} C(x_1, x_1) & C(x_1, x_2) \\ C(x_2, x_1) & C(x_2, x_2) \end{pmatrix}.$$

Definition

X is a *determinantal point process* with *kernel* C, denoted $X \sim \mathrm{DPP}(C)$, if its product density functions satisfy

$$\rho^{(n)}(x_1, \dots, x_n) = \det[C](x_1, \dots, x_n), \quad n = 1, 2, \dots$$

Definition of a DPP on $\mathbb{R}^d$

Let C be a function from $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$.
Denote $[C](x_1, \dots, x_n)$ the $n \times n$ matrix with entries $C(x_i, x_j)$.

$$\text{Ex :} \quad [C](x_1) = C(x_1, x_1) \quad [C](x_1, x_2) = \begin{pmatrix} C(x_1, x_1) & C(x_1, x_2) \\ C(x_2, x_1) & C(x_2, x_2) \end{pmatrix}.$$

Definition

X is a *determinantal point process* with *kernel* C, denoted $X \sim \text{DPP}(C)$, if its product density functions satisfy

$$\rho^{(n)}(x_1, \dots, x_n) = \det[C](x_1, \dots, x_n), \quad n = 1, 2, \dots$$

Basic example : The Poisson point process with intensity $\rho(x)$ is the special case where $C(x, x) = \rho(x)$ and $C(x, y) = 0$ if $x \neq y$.

First properties (if $X \sim DPP(C)$ exists)

First properties (if $X \sim DPP(C)$ exists)

- The intensity of X is $\rho(x) = C(x, x)$.

First properties (if $X \sim DPP(C)$ exists)

- The intensity of X is $\rho(x) = C(x,x)$.
- The pair correlation function is

$$g(x,y) := \frac{\rho^{(2)}(x,y)}{\rho(x)\rho(y)} = \frac{\det[C](x,y)}{C(x,x)C(y,y)} = 1 - \frac{C(x,y)C(y,x)}{C(x,x)C(y,y)}$$

First properties (if $X \sim DPP(C)$ exists)

- The intensity of X is $\rho(x) = C(x, x)$.
- The pair correlation function is

$$g(x, y) := \frac{\rho^{(2)}(x, y)}{\rho(x)\rho(y)} = \frac{\det[C](x, y)}{C(x, x)C(y, y)} = 1 - \frac{C(x, y)C(y, x)}{C(x, x)C(y, y)}$$

- Thus $g \leq 1$ (i.e. repulsiveness) if C is Hermitian.

First properties (if $X \sim DPP(C)$ exists)

- The intensity of X is $\rho(x) = C(x,x)$.
- The pair correlation function is

$$g(x,y) := \frac{\rho^{(2)}(x,y)}{\rho(x)\rho(y)} = \frac{\det[C](x,y)}{C(x,x)C(y,y)} = 1 - \frac{C(x,y)C(y,x)}{C(x,x)C(y,y)}$$

- Thus $g \leq 1$ (i.e. repulsiveness) if C is Hermitian.
- If $X \sim \text{DPP}(C)$, then $X_B \sim \text{DPP}(C_B)$

First properties (if $X \sim DPP(C)$ exists)

- The intensity of X is $\rho(x) = C(x,x)$.
- The pair correlation function is

$$g(x,y) := \frac{\rho^{(2)}(x,y)}{\rho(x)\rho(y)} = \frac{\det[C](x,y)}{C(x,x)C(y,y)} = 1 - \frac{C(x,y)C(y,x)}{C(x,x)C(y,y)}$$

- Thus $g \leq 1$ (i.e. repulsiveness) if C is Hermitian.
- If $X \sim \text{DPP}(C)$, then $X_B \sim \text{DPP}(C_B)$
- Any smooth transformation or independent thinning of a DPP is still a DPP with explicit given kernel.

First properties (if $X \sim DPP(C)$ exists)

- The intensity of X is $\rho(x) = C(x,x)$.
- The pair correlation function is

$$g(x,y) := \frac{\rho^{(2)}(x,y)}{\rho(x)\rho(y)} = \frac{\det[C](x,y)}{C(x,x)C(y,y)} = 1 - \frac{C(x,y)C(y,x)}{C(x,x)C(y,y)}$$

- Thus $g \leq 1$ (i.e. repulsiveness) if C is Hermitian.
- If $X \sim \text{DPP}(C)$, then $X_B \sim \text{DPP}(C_B)$
- Any smooth transformation or independent thinning of a DPP is still a DPP with explicit given kernel.
- Given a kernel C, there exists at most one $\text{DPP}(C)$.

Existence

C must be non-negative definite. Henceforth assume

(C1) C is a continuous complex covariance function.

Existence

C must be non-negative definite. Henceforth assume

(C1) C is a continuous complex covariance function.

By Mercer's theorem, for any compact set $S \subset \mathbb{R}^d$, C restricted to $S \times S$, denoted C_S, has a spectral representation,

$$C_S(x, y) = \sum_{k=1}^{\infty} \lambda_k^S \phi_k^S(x) \overline{\phi_k^S(y)}, \quad (x, y) \in S \times S,$$

where $\lambda_k^S \geq 0$ and $\int_S \phi_k^S(x) \overline{\phi_l^S(x)} \, \mathrm{d}x = \mathbf{1}_{\{k=l\}}$.

Existence

C must be non-negative definite. Henceforth assume

(C1) C is a continuous complex covariance function.

By Mercer's theorem, for any compact set $S \subset \mathbb{R}^d$, C restricted to $S \times S$, denoted C_S, has a spectral representation,

$$C_S(x,y) = \sum_{k=1}^{\infty} \lambda_k^S \phi_k^S(x) \overline{\phi_k^S(y)}, \quad (x,y) \in S \times S,$$

where $\lambda_k^S \geq 0$ and $\int_S \phi_k^S(x) \overline{\phi_l^S(x)} \, \mathrm{d}x = \mathbf{1}_{\{k=l\}}$.

Theorem (Macchi, 1975)

Under (C1), existence of DPP(C) is equivalent to :

$$\lambda_k^S \leq 1 \textit{ for all compact } S \subset \mathbb{R}^d \textit{ and all } k.$$

Density on a compact set S

Let $X \sim \mathrm{DPP}(C)$ and $S \subset \mathbb{R}^d$ be any compact set.

Recall that $C_S(x,y) = \sum_{k=1}^{\infty} \lambda_k^S \phi_k^S(x)\overline{\phi_k^S(y)}$.

Theorem (Macchi (1975))

If $\lambda_k^S < 1 \ \forall k$, then $X_S \ll Poisson(S,1)$, with density

$$f(\{x_1,\dots,x_n\}) = \exp(|S| - D)\det[\tilde{C}](x_1,\dots,x_n),$$

where $D = -\sum_{k=1}^{\infty} \log(1-\lambda_k^S)$ and $\tilde{C} : S \times S \to \mathbb{C}$ is given by

$$\tilde{C}(x,y) = \sum_{k=1}^{\infty} \frac{\lambda_k^S}{1-\lambda_k^S} \phi_k^S(x)\overline{\phi_k^S(y)}$$

Simulation

Let $X \sim \mathrm{DPP}(C)$.

We want to simulate X_S for $S \subset \mathbb{R}^d$ compact.

Recall that $X_S \sim \mathrm{DPP}(C_S)$ with $C_S(x,y) = \sum_{k=1}^{\infty} \lambda_k^S \phi_k^S(x) \overline{\phi_k^S(y)}$.

Simulation

Let $X \sim \mathrm{DPP}(C)$.

We want to simulate X_S for $S \subset \mathbb{R}^d$ compact.

Recall that $X_S \sim \mathrm{DPP}(C_S)$ with $C_S(x,y) = \sum_{k=1}^{\infty} \lambda_k^S \phi_k^S(x) \overline{\phi_k^S(y)}$.

Theorem (Hough et al. (2006))

For $k \in \mathbb{N}$, let B_k be independent Bernoulli r.v. with mean λ_k^S. Define

$$K(x,y) = \sum_{k=1}^{\infty} B_k \phi_k^S(x) \overline{\phi_k^S(y)}, \quad (x,y) \in S \times S.$$

Then $DPP(C_S) \stackrel{d}{=} DPP(K)$.

So simulating X_S is equivalent to simulate $\mathrm{DPP}(K)$ with

$$K(x,y) = \sum_{k=1}^{\infty} B_k \phi_k^S(x) \overline{\phi_k^S(y)}, \quad (x,y) \in S \times S.$$

So simulating X_S is equivalent to simulate $\mathrm{DPP}(K)$ with

$$K(x,y) = \sum_{k=1}^{\infty} B_k \phi_k^S(x)\overline{\phi_k^S(y)}, \quad (x,y) \in S \times S.$$

1. Simulate $M := \sup\{k \geq 0 : B_k \neq 0\}$ (by the inversion method).

So simulating X_S is equivalent to simulate DPP(K) with

$$K(x,y) = \sum_{k=1}^{\infty} B_k \phi_k^S(x)\overline{\phi_k^S(y)}, \quad (x,y) \in S \times S.$$

1. Simulate $M := \sup\{k \geq 0 : B_k \neq 0\}$ (by the inversion method).
2. Given $M = m$, generate $B_1, \ldots, B_{m-1}$

So simulating X_S is equivalent to simulate $\mathrm{DPP}(K)$ with

$$K(x,y) = \sum_{k=1}^{\infty} B_k \phi_k^S(x)\overline{\phi_k^S(y)}, \quad (x,y) \in S \times S.$$

1. Simulate $M := \sup\{k \geq 0 : B_k \neq 0\}$ (by the inversion method).
2. Given $M = m$, generate $B_1, \dots, B_{m-1}$
3. simulate $\mathrm{DPP}(K)$ given $B_1, \dots, B_M$ and $M = m$. The kernel K becomes w.l.o.g.

$$K(x,y) = \sum_{k=1}^{n} \phi_k^S(x)\overline{\phi_k^S(y)}$$

and $\mathrm{DPP}(K)$ is a *determinantal projection process*.

Simulation of a determinantal projection process:

$$K(x,y)=\sum_{k=1}^{n}\phi_k^S(x)\overline{\phi_k^S(y)}=\boldsymbol{v}(y)^*\boldsymbol{v}(x),\quad \boldsymbol{v}(x)=(\phi_1^S(x),\dots,\phi_n^S(x))^T.$$

DPP(K) has a.s. n points $(X_1,\dots,X_n)$ that can be generated by the following Gram-Schmidt procedure

---

sample X_n from the density $p_n(x)=\|\boldsymbol{v}(x)\|^2/n$;
set $\boldsymbol{e}_1=\boldsymbol{v}(X_n)/\|\boldsymbol{v}(X_n)\|$;
for $i=(n-1)$ to 1 **do**
  sample X_i from the density (given $X_{i+1},\dots,X_n$) :

$$p_i(x)=\frac{1}{i}\left[\|\boldsymbol{v}(x)\|^2-\sum_{j=1}^{n-i}|\boldsymbol{e}_j^*\boldsymbol{v}(x)|^2\right],\quad x\in S$$

  set $\boldsymbol{w}_i=\boldsymbol{v}(X_i)-\sum_{j=1}^{n-i}\left(\boldsymbol{e}_j^*\boldsymbol{v}(X_i)\right)\boldsymbol{e}_j$, $\boldsymbol{e}_{n-i+1}=\boldsymbol{w}_i/\|\boldsymbol{w}_i\|$

Summary

Therefore, given a kernel C :

- condition for existence of DPP(C) are known*
- all moments of DPP(C) are explicitly known
- the density of DPP(C) on any compact set is known*
- DPP(C) can be easily and quickly simulated on any compact set*

* if the spectral representation of C_S is known on any S (see later for an approximation).

Stationary models

Consider a stationary kernel : $C(x, y) = C_0(x - y), \quad x, y \in \mathbb{R}^d.$

Stationary models

Consider a stationary kernel : $C(x,y) = C_0(x-y), \quad x, y \in \mathbb{R}^d.$

Recall (C1): C_0 is a continuous covariance function.

Moreover, if $C_0 \in L^2(\mathbb{R}^d)$ we can define its Fourier transform

$$\varphi(x) = \int C_0(t) \mathrm{e}^{-2\pi \mathrm{i} x \cdot t} \, \mathrm{d}t, \quad x \in \mathbb{R}^d.$$

Stationary models

Consider a stationary kernel : $C(x, y) = C_0(x - y), \quad x, y \in \mathbb{R}^d$.

Recall (C1): C_0 is a continuous covariance function.

Moreover, if $C_0 \in L^2(\mathbb{R}^d)$ we can define its Fourier transform

$$\varphi(x) = \int C_0(t) \mathrm{e}^{-2\pi \mathrm{i} x \cdot t} \, \mathrm{d}t, \quad x \in \mathbb{R}^d.$$

Theorem

Under (C1), if $C_0 \in L^2(\mathbb{R}^d)$, then existence of DPP$(C_0)$ is equivalent to

$$\varphi \leq 1.$$

Stationary models

Consider a stationary kernel : $C(x, y) = C_0(x - y), \quad x, y \in \mathbb{R}^d$.

Recall (C1): C_0 is a continuous covariance function.

Moreover, if $C_0 \in L^2(\mathbb{R}^d)$ we can define its Fourier transform

$$\varphi(x) = \int C_0(t) \mathrm{e}^{-2\pi \mathrm{i} x \cdot t} \, \mathrm{d}t, \quad x \in \mathbb{R}^d.$$

Theorem

Under (C1), if $C_0 \in L^2(\mathbb{R}^d)$, then existence of DPP$(C_0)$ is equivalent to

$$\varphi \leq 1.$$

To construct parametric families of DPP :
Consider parametric families of C_0 and rescale so that $\varphi \leq 1$.

$\rightarrow$ *This will induce restriction on the parameter space.*

Several parametric families of covariance functions are available, with closed form expressions for their Fourier transform.

Several parametric families of covariance functions are available, with closed form expressions for their Fourier transform.

- For $d = 2$, the circular covariance function with range α is given by

$$C_0(x) = \rho \frac{2}{\pi} \left(\arccos(\|x\|/\alpha) - \|x\|/\alpha \sqrt{1 - (\|x\|/\alpha)^2} \right) \mathbf{1}_{\|x\|<\alpha}.$$

  DPP(C_0) exists iff $\varphi \leq 1 \Leftrightarrow \rho\alpha^2 \leq 4/\pi$.

  $\Rightarrow$ Tradeoff between the intensity ρ and the range of repulsion α.

Several parametric families of covariance functions are available, with closed form expressions for their Fourier transform.

- For $d = 2$, the circular covariance function with range α is given by

$$C_0(x) = \rho\frac{2}{\pi}\left(\arccos(\|x\|/\alpha) - \|x\|/\alpha\sqrt{1-(\|x\|/\alpha)^2}\right)\mathbf{1}_{\|x\|<\alpha}.$$

  DPP(C_0) exists iff $\varphi \leq 1 \Leftrightarrow \rho\alpha^2 \leq 4/\pi$.

  $\Rightarrow$ Tradeoff between the intensity ρ and the range of repulsion α.

- Whittle-Matérn (includes Exponential and Gaussian) :

$$C_0(x) = \rho\,\frac{2^{1-\nu}}{\Gamma(\nu)}\|x/\alpha\|^{\nu}K_{\nu}(\|x/\alpha\|), \quad x \in \mathbb{R}^d,$$

  DPP(C_0) exists iff $\rho \leq \frac{\Gamma(\nu)}{\Gamma(\nu+d/2)(2\sqrt{\pi}\alpha)^d}$

Several parametric families of covariance functions are available, with closed form expressions for their Fourier transform.

- For $d = 2$, the circular covariance function with range α is given by

$$C_0(x) = \rho \frac{2}{\pi} \left(\arccos(\|x\|/\alpha) - \|x\|/\alpha \sqrt{1 - (\|x\|/\alpha)^2} \right) \mathbf{1}_{\|x\|<\alpha}.$$

  DPP(C_0) exists iff $\varphi \leq 1 \Leftrightarrow \rho\alpha^2 \leq 4/\pi$.

  $\Rightarrow$ Tradeoff between the intensity ρ and the range of repulsion α.

- Whittle-Matérn (includes Exponential and Gaussian) :

$$C_0(x) = \rho \, \frac{2^{1-\nu}}{\Gamma(\nu)} \|x/\alpha\|^{\nu} K_{\nu}(\|x/\alpha\|), \quad x \in \mathbb{R}^d,$$

  DPP(C_0) exists iff $\rho \leq \frac{\Gamma(\nu)}{\Gamma(\nu+d/2)(2\sqrt{\pi}\alpha)^d}$

- Generalized Cauchy

$$C_0(x) = \frac{\rho}{(1 + \|x/\alpha\|^2)^{\nu+d/2}}, \quad x \in \mathbb{R}^d.$$

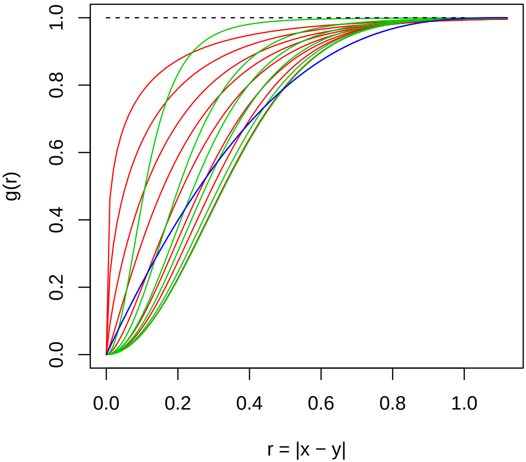
  DPP(C_0) exists iff $\rho \leq \frac{\Gamma(\nu+d/2)}{\Gamma(\nu)(\sqrt{\pi}\alpha)^d}$

Pair correlation functions of DPP(C_0) for previous models when $\rho = 1$ and $\alpha = \alpha_{max}(\nu)$:

In blue : C_0 is the circular covariance function.

In red : C_0 is Whittle-Matérn, for different values of ν

In green : C_0 is generalized Cauchy, for different values of ν



- Are the previous parametric families "rich enough"?

- How repulsive a stationary DPP can be?

  - Less repulsive DPP = Poisson Point Process.

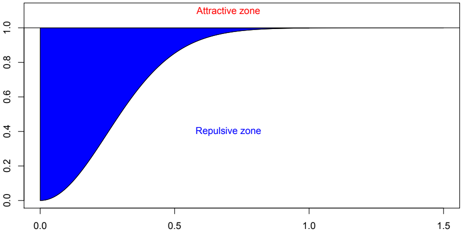
  - What is the most repulsive DPP?

- We introduce criteria of repulsiveness based on the pair correlation function g.

How repulsive a stationary DPP can be?

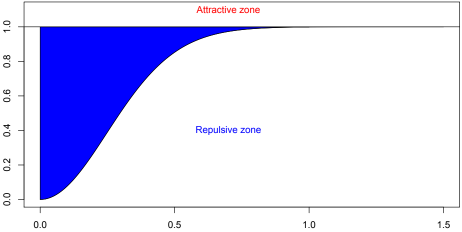
How repulsive a stationary DPP can be?

Criteria of repulsiveness based on the pair correlation function.



How repulsive a stationary DPP can be?

Criteria of repulsiveness based on the pair correlation function.



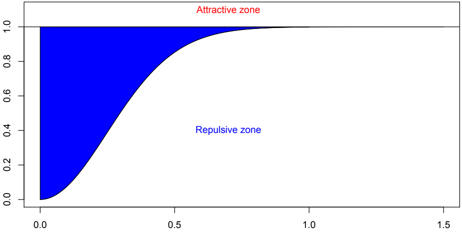
Let X and Y be two DPPs with the same intensity ρ and pcf g_X, g_Y.

Definition

- X is *globally more repulsive* than Y if "g_X has a larger blue zone", i.e. $\int(1 - g_X) \geq \int(1 - g_Y)$.

How repulsive a stationary DPP can be?

Criteria of repulsiveness based on the pair correlation function.



Let X and Y be two DPPs with the same intensity ρ and pcf g_X, g_Y.

Definition

- X is *globally more repulsive* than Y if "g_X has a larger blue zone", i.e. $\int(1-g_X) \geq \int(1-g_Y)$.

- X is *locally more repulsive* than Y if "g_X is more flat near 0", i.e. $g_X(0) = g_Y(0) = 0$, $\nabla g_X(0) = \nabla g_Y(0) = 0$ and $\Delta g_X(0) \leq \Delta g_Y(0)$.

Note that for a hardcore point process: $g(0) = \nabla g(0) = \Delta g(0) = 0$.

Let J_ν be the Bessel function of the first kind with order ν.

Proposition

There exists a unique DPP that is both the most globally and the most locally repulsive DPP with intensity ρ. Its kernel in dimension $d = 2$ is given by:

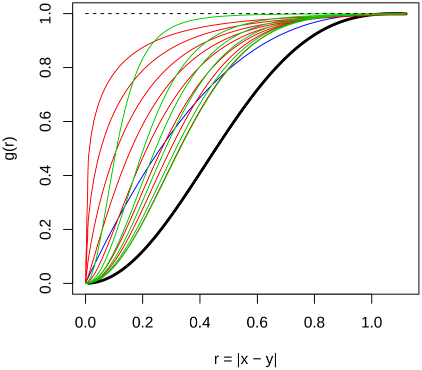
$$C_0(x) = \sqrt{\rho}\frac{J_1(2\sqrt{\pi\rho}||x||)}{\sqrt{\pi}||x||} \quad \text{and} \quad \varphi(x) = \mathbf{1}_{||x||^2 \leq \rho\pi^{-1}}$$

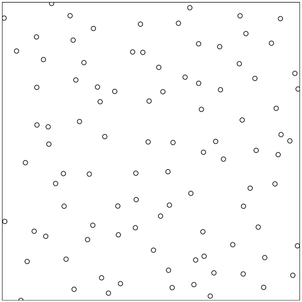
Let J_ν be the Bessel function of the first kind with order ν.

Proposition

There exists a unique DPP that is both the most globally and the most locally repulsive DPP with intensity ρ. Its kernel in dimension $d = 2$ is given by:

$$C_0(x) = \sqrt{\rho}\frac{J_1(2\sqrt{\pi\rho}||x||)}{\sqrt{\pi}||x||} \quad \text{and} \quad \varphi(x) = \mathbf{1}_{||x||^2 \leq \rho\pi^{-1}}$$





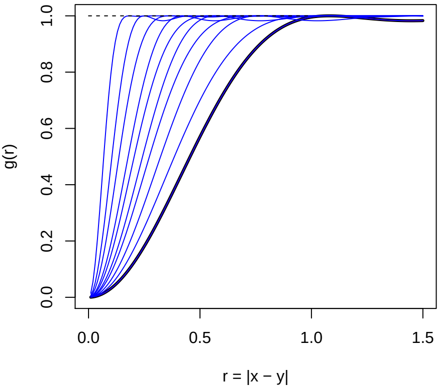
New parametric families of kernels

New parametric families of kernels

- Laguerre-Gauss family, for $n \in \mathbb{N}$ and $0 < \alpha \leq \alpha_{max}(d, \rho, n)$

$$C_0(x) \propto L_{n-1}^{\frac{d}{2}}\left(\frac{1}{n}\left\|\frac{x}{\alpha}\right\|^2\right) e^{-\frac{1}{n}\|\frac{x}{\alpha}\|^2} \quad \varphi(x) \propto e^{-n(\pi\alpha\|x\|)^2} \sum_{k=0}^{n-1} \frac{(\pi\sqrt{n}\|\alpha x\|)^{2k}}{k!}$$

$\longrightarrow$ covers all possible degrees of repulsiveness.

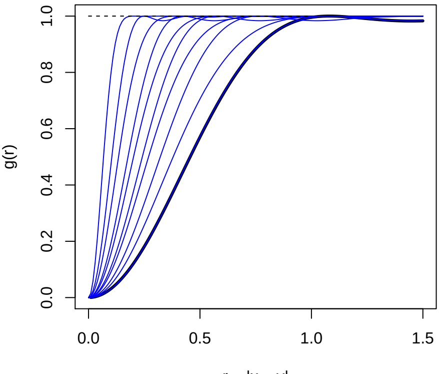


- Generalized-Sinc family, for $\sigma \geq 0$, $0 < \alpha \leq \alpha_{max}(d, \rho, \sigma)$

$$C_0(x) \propto \frac{J_{\frac{\sigma+d}{2}}\left(2\|\frac{x}{\alpha}\|\sqrt{\frac{\sigma+d}{2}}\right)}{\left(2\|\frac{x}{\alpha}\|\sqrt{\frac{\sigma+d}{2}}\right)^{\frac{\sigma+d}{2}}} \qquad \varphi(x) \propto \left(1 - \frac{2(\pi\|\alpha x\|)^2}{\sigma + d}\right)_+^{\frac{\sigma}{2}}$$

$\longrightarrow$ covers all possible degrees of repulsiveness.

$\longrightarrow$ φ is compactly supported.



Approximation of stationary DPP's models

Consider a stationary kernel C_0 and $X \sim \mathrm{DPP}(C_0)$.

- The simulation and the density of X_S requires the expansion

$$C_S(x,y) = C_0(y-x) = \sum_{k=1}^{\infty} \lambda_k^S \phi_k^S(x) \overline{\phi_k^S(y)}, \quad (x,y) \in S \times S,$$

but in general λ_k^S and ϕ_k^S are not expressible on closed form.

Approximation of stationary DPP's models

Consider a stationary kernel C_0 and $X \sim \mathrm{DPP}(C_0)$.

- The simulation and the density of X_S requires the expansion

$$C_S(x,y) = C_0(y-x) = \sum_{k=1}^{\infty} \lambda_k^S \phi_k^S(x)\overline{\phi_k^S(y)}, \quad (x,y) \in S \times S,$$

but in general λ_k^S and ϕ_k^S are not expressible on closed form.

- Consider w.l.g. the unit box $S = [-\frac{1}{2}, \frac{1}{2}]^d$ and the Fourier expansion

$$C_0(y-x) = \sum_{k \in \mathbb{Z}^d} c_k \mathrm{e}^{2\pi \mathrm{i} k\cdot(y-x)}, \quad y - x \in S.$$

Approximation of stationary DPP's models

Consider a stationary kernel C_0 and $X \sim \mathrm{DPP}(C_0)$.

- The simulation and the density of X_S requires the expansion

$$C_S(x,y) = C_0(y-x) = \sum_{k=1}^{\infty} \lambda_k^S \phi_k^S(x)\overline{\phi_k^S(y)}, \quad (x,y) \in S \times S,$$

but in general λ_k^S and ϕ_k^S are not expressible on closed form.

- Consider w.l.g. the unit box $S = [-\frac{1}{2}, \frac{1}{2}]^d$ and the Fourier expansion

$$C_0(y-x) = \sum_{k \in \mathbb{Z}^d} c_k \mathrm{e}^{2\pi \mathrm{i} k\cdot(y-x)}, \quad y-x \in S.$$

The Fourier coefficients are

$$c_k = \int_S C_0(u)\mathrm{e}^{-2\pi \mathrm{i} k\cdot u}\,\mathrm{d}u$$

Approximation of stationary DPP's models

Consider a stationary kernel C_0 and $X \sim \mathrm{DPP}(C_0)$.

- The simulation and the density of X_S requires the expansion

$$C_S(x,y) = C_0(y-x) = \sum_{k=1}^{\infty} \lambda_k^S \phi_k^S(x) \overline{\phi_k^S(y)}, \quad (x,y) \in S \times S,$$

but in general λ_k^S and ϕ_k^S are not expressible on closed form.

- Consider w.l.g. the unit box $S = [-\frac{1}{2}, \frac{1}{2}]^d$ and the Fourier expansion

$$C_0(y-x) = \sum_{k \in \mathbb{Z}^d} c_k \mathrm{e}^{2\pi \mathrm{i} k \cdot (y-x)}, \quad y - x \in S.$$

The Fourier coefficients are

$$c_k = \int_S C_0(u) \mathrm{e}^{-2\pi \mathrm{i} k \cdot u} \, \mathrm{d}u \approx \int_{\mathbb{R}^d} C_0(u) \mathrm{e}^{-2\pi \mathrm{i} k \cdot u} \, \mathrm{d}u = \varphi(k)$$

which is a good approximation if $C_0(u) \approx 0$ for $|u| > \frac{1}{2}$.

Approximation of stationary DPP's models

Consider a stationary kernel C_0 and $X \sim \mathrm{DPP}(C_0)$.

- The simulation and the density of X_S requires the expansion

$$C_S(x,y) = C_0(y-x) = \sum_{k=1}^{\infty} \lambda_k^S \phi_k^S(x)\overline{\phi_k^S(y)}, \quad (x,y) \in S \times S,$$

but in general λ_k^S and ϕ_k^S are not expressible on closed form.

- Consider w.l.g. the unit box $S = [-\frac{1}{2}, \frac{1}{2}]^d$ and the Fourier expansion

$$C_0(y-x) = \sum_{k \in \mathbb{Z}^d} c_k \mathrm{e}^{2\pi \mathrm{i} k \cdot (y-x)}, \quad y - x \in S.$$

The Fourier coefficients are

$$c_k = \int_S C_0(u) \mathrm{e}^{-2\pi \mathrm{i} k \cdot u} \, \mathrm{d}u \approx \int_{\mathbb{R}^d} C_0(u) \mathrm{e}^{-2\pi \mathrm{i} k \cdot u} \, \mathrm{d}u = \varphi(k)$$

which is a good approximation if $C_0(u) \approx 0$ for $|u| > \frac{1}{2}$.

- Example: For the circular covariance, this is true whenever $\rho|S| > 5$.

Approximation of stationary models

So, DPP(C_0) on S can be approximated by DPP($C_{\mathrm{app},0}$) with

$$C_{\mathrm{app},0}(y-x) = \sum_{k\in\mathbb{Z}^d} \varphi(k)\mathrm{e}^{2\pi\mathrm{i}(y-x)\cdot k}, \quad x, y \in S,$$

where φ is the Fourier transform of C_0.

This **kernel approximation** allows us

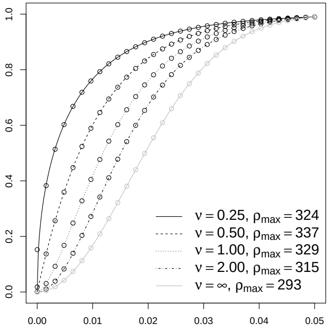
- to simulate DPP(C_0) on S, by simulating DPP($C_{\mathrm{app},0}$)
- to compute the (approximated) density of DPP(C_0) on S.

Turns out to be a very good approximation in most cases.

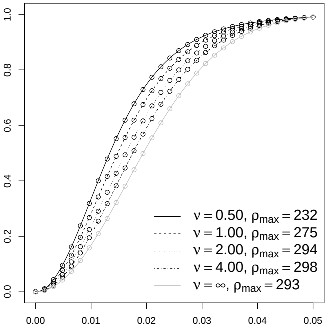
Exemples of approximations

– Solid lines : theoretical pair correlation function

∘ Circles : pair correlation from the approximated kernel



Whittle-Matérn



Generalized Cauchy

Consider a stationary and isotropic parametric DPP(C_0), with

$$C_0(x-y) = \rho R_\theta(\|x-y\|),$$

where $R_\theta(0) = 1$ and θ is some parameter.

Consider a stationary and isotropic parametric DPP(C_0), with

$$C_0(x - y) = \rho R_\theta(\|x - y\|),$$

where $R_\theta(0) = 1$ and θ is some parameter.

For this model:

- $\rho =$ intensity.
- Pair correlation function:

$$g(x, y) = 1 - R_\theta(\|x - y\|)^2.$$

Parameter estimation

In

$$C_0(x-y) = \rho R_\theta(\|x-y\|)$$

we can estimate (ρ, θ)

- by MLE using the kernel approximation for the likelihood
- or using the moments, for instance :

$$\hat{\rho} = \#\{\text{obs. points}\}/[\text{area of obs. window}]$$

$$\hat{\theta} = \text{argmin}_\theta \int_{r_{\min}}^{r_{\max}} \left| \sqrt{\widehat{g}(r)} - \sqrt{g_\theta(r)} \right|^2 \mathrm{d}r$$

Conclusion

• DPPs provide **flexible** (parametric) models of repulsive point processes and possess **appealing properties**:

- Easily and rather quickly simulated
- Closed form expressions for all moments.
- Closed form expression for the density on any bounded set.
- Parametric models are available
- Inference is feasible, including likelihood inference.
- More: Laplace transform, mixing properties, concentration inequalities, asymptotic inference,...

• Parametric models, inference methods and so on are available in R in the package `spatstat`.

Conclusion

Some open questions :

- Can we simulate DPP(C) without the spectral representation of C?
- Theoretical properties for the MLE?
- Theory for inference in the non-homogeneous case
- Semi or non-parametric inference
- ...

Conclusion

Some open questions :

- Can we simulate DPP(C) without the spectral representation of C?
- Theoretical properties for the MLE?
- Theory for inference in the non-homogeneous case
- Semi or non-parametric inference
- ...

THANK YOU