Statistics for determinantal point process models and its usefulness

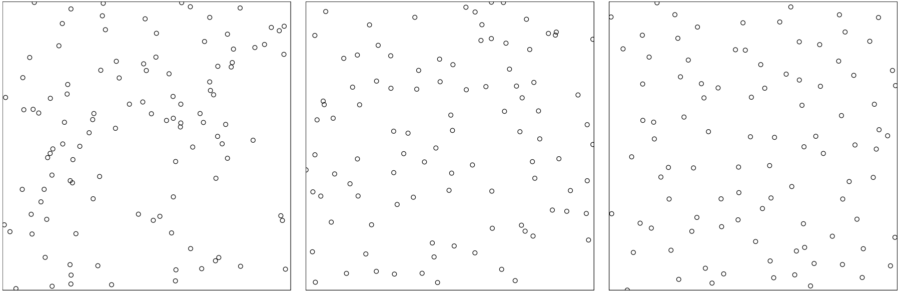
September 28, 2018

Jesper Møller
jm@math.aau.dk

Department of Mathematical Sciences
Aalborg University
Denmark

AALBORG UNIVERSITY
DENMARK

Introduction

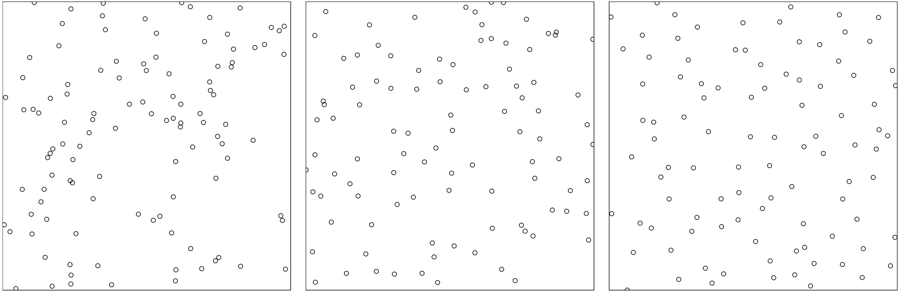


Poisson DPP DPP with stronger inhibition

- DPPs are inhibitive/regular/repulsive point processes.
- Introduced by O. Macchi in 1975 to model fermions in quantum mechanics.
- Several theoretical studies appeared in the 2000's.

Jesper Møller

Introduction



Poisson | DPP | DPP with stronger inhibition

- DPPs are inhibitive/regular/repulsive point processes.
- Introduced by O. Macchi in 1975 to model fermions in quantum mechanics.
- Several theoretical studies appeared in the 2000's.
- **Until recently statistical models and inference had been largely unexplored.**

Dept. of Mathematical Sciences
Aalborg University
Denmark

Main questions re. DPPs

How do we make statistical inference (parameter estimation; simulation based procedures used e.g. for model checking)?

Dept. of Mathematical Sciences
Aalborg University
Denmark

Main questions re. DPPs

How do we make statistical inference (parameter estimation; simulation based procedures used e.g. for model checking)?

How do we construct useful classes of parametric models?
And ensure the DPP exists?

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Main questions re. DPPs

How do we make statistical inference (parameter estimation; simulation based procedures used e.g. for model checking)?

How do we construct useful classes of parametric models?
And ensure the DPP exists?

How flexible will they be — to what extent are DPPs repulsive?

Dept. of Mathematical Sciences
Aalborg University
Denmark

Main questions re. DPPs

How do we make statistical inference (parameter estimation; simulation based procedures used e.g. for model checking)?

How do we construct useful classes of parametric models?
And ensure the DPP exists?

How flexible will they be — to what extent are DPPs repulsive?

- Focus on DPPs defined on $\mathbb{R}^d$;
- something about DPPs defined on other spaces ($\mathbb{S}^d$).

Gibbs point processes — the usual class of point processes used for modelling inhibition

Example: A *Strauss hard-core process* on a compact set $S \subset \mathbb{R}^d$ has density

$$f(\{x_1,\ldots,x_n\}) = \frac{1}{c(r,R,\beta,\gamma)} \beta^n \gamma^{\sum_{i<j} \mathbf{1}_{\{\|x_i-x_j\| \le R\}}} \prod_{i<j} \mathbf{1}_{\{\|x_i-x_j\|>r\}}, \qquad \{x_1,\ldots,x_n\} \subset S,$$

w.r.t. the unit rate Poisson process, where $n = 0, 1, \ldots$, and $0 < r < R$ and $\beta, \gamma > 0$ are parameters.

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Gibbs point processes — the usual class of point processes used for modelling inhibition

Example: A *Strauss hard-core process* on a compact set $S \subset \mathbb{R}^d$ has density

$$f(\{x_1,\ldots,x_n\}) = \frac{1}{c(r,R,\beta,\gamma)}\beta^n\gamma^{\sum_{i<j}\mathbf{1}_{\{\|x_i-x_j\|\le R\}}}\prod_{i<j}\mathbf{1}_{\{\|x_i-x_j\|>r\}}, \qquad \{x_1,\ldots,x_n\}\subset S,$$

w.r.t. the unit rate Poisson process, where $n = 0, 1, \ldots$, and $0 < r < R$ and $\beta, \gamma > 0$ are parameters.

- The normalizing constant $c(r, R, \beta, \gamma)$ is intractable.

Dept. of Mathematical Sciences
Aalborg University
Denmark

Gibbs point processes — the usual class of point processes used for modelling inhibition

Example: A *Strauss hard-core process* on a compact set $S \subset \mathbb{R}^d$ has density

$$f(\{x_1,\ldots,x_n\}) = \frac{1}{c(r,R,\beta,\gamma)}\beta^n\gamma^{\sum_{i<j}\mathbf{1}_{\{\|x_i-x_j\|\le R\}}}\prod_{i<j}\mathbf{1}_{\{\|x_i-x_j\|>r\}}, \qquad \{x_1,\ldots,x_n\}\subset S,$$

w.r.t. the unit rate Poisson process, where $n = 0,1,\ldots$, and $0 < r < R$ and $\beta,\gamma > 0$ are parameters.

- The normalizing constant $c(r,R,\beta,\gamma)$ is intractable.
- Don't know any moment properties, including the intensity.

Gibbs point processes — the usual class of point processes used for modelling inhibition

Example: A *Strauss hard-core process* on a compact set $S \subset \mathbb{R}^d$ has density

$$f(\{x_1,\ldots,x_n\}) = \frac{1}{c(r,R,\beta,\gamma)}\beta^n \gamma^{\sum_{i<j} \mathbf{1}_{\{\|x_i - x_j\| \le R\}}} \prod_{i<j} \mathbf{1}_{\{\|x_i - x_j\| > r\}}, \qquad \{x_1,\ldots,x_n\} \subset S,$$

w.r.t. the unit rate Poisson process, where $n = 0, 1, \ldots$, and $0 < r < R$ and $\beta, \gamma > 0$ are parameters.

- The normalizing constant $c(r, R, \beta, \gamma)$ is intractable.
- Don't know any moment properties, including the intensity.
- Don't know the distribution of (e.g.) the number of points.

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Gibbs point processes — the usual class of point processes used for modelling inhibition

Example: A *Strauss hard-core process* on a compact set $S \subset \mathbb{R}^d$ has density

$$f(\{x_1,\dots,x_n\}) = \frac{1}{c(r,R,\beta,\gamma)} \beta^n \gamma^{\sum_{i<j} \mathbf{1}_{\{\|x_i-x_j\|\le R\}}} \prod_{i<j} \mathbf{1}_{\{\|x_i-x_j\|>r\}}, \qquad \{x_1,\dots,x_n\} \subset S,$$

w.r.t. the unit rate Poisson process, where $n = 0, 1, \dots$, and $0 < r < R$ and $\beta, \gamma > 0$ are parameters.

- The normalizing constant $c(r,R,\beta,\gamma)$ is intractable.
- Don't know any moment properties, including the intensity.
- Don't know the distribution of (e.g.) the number of points.
- (Long) MCMC runs are needed...

Dept. of Mathematical Sciences
Aalborg University
Denmark

Gibbs point processes — the usual class of point processes used for modelling inhibition

Jesper Møller

Example: A *Strauss hard-core process* on a compact set $S \subset \mathbb{R}^d$ has density

$$f(\{x_1,\ldots,x_n\}) = \frac{1}{c(r,R,\beta,\gamma)}\beta^n\gamma^{\sum_{i<j}\mathbf{1}_{\{\|x_i-x_j\|\le R\}}}\prod_{i<j}\mathbf{1}_{\{\|x_i-x_j\|>r\}}, \qquad \{x_1,\ldots,x_n\}\subset S,$$

w.r.t. the unit rate Poisson process, where $n = 0, 1, \ldots$, and $0 < r < R$ and $\beta, \gamma > 0$ are parameters.

- The normalizing constant $c(r, R, \beta, \gamma)$ is intractable.
- Don't know any moment properties, including the intensity.
- Don't know the distribution of (e.g.) the number of points.
- (Long) MCMC runs are needed...
- Edge effects have been ignored: the restriction to an observation window $W \subsetneq S$ is not a Strauss hard-core process.

Dept. of Mathematical Sciences
Aalborg University
Denmark

Gibbs point processes — the usual class of point processes used for modelling inhibition

Example: A *Strauss hard-core process* on a compact set $S \subset \mathbb{R}^d$ has density

$$f(\{x_1,\ldots,x_n\}) = \frac{1}{c(r,R,\beta,\gamma)}\beta^n \gamma^{\sum_{i<j} \mathbf{1}_{\{\|x_i-x_j\|\le R\}}} \prod_{i<j} \mathbf{1}_{\{\|x_i-x_j\|>r\}}, \qquad \{x_1,\ldots,x_n\} \subset S,$$

w.r.t. the unit rate Poisson process, where $n = 0, 1, \ldots$, and $0 < r < R$ and $\beta, \gamma > 0$ are parameters.

- The normalizing constant $c(r, R, \beta, \gamma)$ is intractable.
- Don't know any moment properties, including the intensity.
- Don't know the distribution of (e.g.) the number of points.
- (Long) MCMC runs are needed...
- Edge effects have been ignored: the restriction to an observation window $W \subsetneq S$ is not a Strauss hard-core process.
- On $\mathbb{R}^d$ a 'local specification' is needed and the issue of phase transition has to be clarified.

Dept. of Mathematical Sciences
Aalborg University
Denmark

Notation

- X : simple point process on $\mathbb{R}^d$, with Lebesgue measure as reference measure (or any locally compact Polish space equipped with a Radon measure).

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Notation

- X : simple point process on $\mathbb{R}^d$, with Lebesgue measure as reference measure (or any locally compact Polish space equipped with a Radon measure).
- For any Borel set $B \subseteq \mathbb{R}^d$, $X_B = X \cap B$, $X(B) = \#X_B$.

Dept. of Mathematical Sciences
Aalborg University
Denmark

Notation

- X : simple point process on $\mathbb{R}^d$, with Lebesgue measure as reference measure (or any locally compact Polish space equipped with a Radon measure).
- For any Borel set $B \subseteq \mathbb{R}^d$, $X_B = X \cap B$, $X(B) = \#X_B$.
- For any integer $n > 0$, denote $\rho^{(n)}$ the n'th *order joint intensity* of X:

$$\mathrm{E}\left[X(B_1)\cdots X(B_n)\right] = \int_{B_1}\cdots\int_{B_n}\rho^{(n)}(x_1,\ldots,x_n)\,\mathrm{d}x_1\cdots\mathrm{d}x_n$$

  for any disjoint Borel sets $B_1,\ldots,B_n \subseteq \mathbb{R}^d$.

  - 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 $\mathrm{d}x_i$.

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Notation

- X : simple point process on $\mathbb{R}^d$, with Lebesgue measure as reference measure (or any locally compact Polish space equipped with a Radon measure).
- For any Borel set $B \subseteq \mathbb{R}^d$, $X_B = X \cap B$, $X(B) = \#X_B$.
- For any integer $n > 0$, denote $\rho^{(n)}$ the n'th *order joint intensity* of X:

$$\mathrm{E}\left[X(B_1)\cdots X(B_n)\right] = \int_{B_1}\cdots\int_{B_n} \rho^{(n)}(x_1,\ldots,x_n)\,\mathrm{d}x_1\cdots\mathrm{d}x_n$$

for any disjoint Borel sets $B_1,\ldots,B_n \subseteq \mathbb{R}^d$.

  - 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*.

Notation

- X : simple point process on $\mathbb{R}^d$, with Lebesgue measure as reference measure (or any locally compact Polish space equipped with a Radon measure).
- For any Borel set $B \subseteq \mathbb{R}^d$, $X_B = X \cap B$, $X(B) = \#X_B$.
- For any integer $n > 0$, denote $\rho^{(n)}$ the n'th *order joint intensity* of X:

$$\mathrm{E}\left[X(B_1)\cdots X(B_n)\right] = \int_{B_1}\cdots\int_{B_n} \rho^{(n)}(x_1,\ldots,x_n)\,\mathrm{d}x_1\cdots\mathrm{d}x_n$$

for any disjoint Borel sets $B_1,\ldots,B_n \subseteq \mathbb{R}^d$.

  - 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*.
  - $\rho^{(n)}(x_1,\ldots,x_n) = 0$ if $x_i = x_j$ for some $i \neq j$.

Definition of a DPP

Jesper Møller

Definition

Let C be a function $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$, called the *kernel*.
$X \sim \mathrm{DPP}(C)$ if

$$\rho^{(n)}(x_1, \ldots, x_n) = \det\{C(x_i, x_j)\}_{i,j=1,\ldots,n}, \quad n = 1, 2, \ldots$$

Dept. of Mathematical Sciences
Aalborg University
Denmark

Definition of a DPP

Definition

Let C be a function $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$, called the *kernel*.
$X \sim \mathrm{DPP}(C)$ if

$$\rho^{(n)}(x_1, \ldots, x_n) = \det\{C(x_i, x_j)\}_{i,j=1,\ldots,n}, \quad n = 1, 2, \ldots$$

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

Definition of a DPP

Definition

Let C be a function $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$, called the *kernel*.
$X \sim \mathrm{DPP}(C)$ if

$$\rho^{(n)}(x_1, \ldots, x_n) = \det\{C(x_i, x_j)\}_{i,j=1,\ldots,n}\,, \quad n = 1, 2, \ldots$$

- The Poisson process with intensity $\rho(x)$ is the special case where $C(x, x) = \rho(x)$ and $C(x, y) = 0$ if $x \neq y$.
- For existence, conditions on the kernel C are mandatory. E.g. $\det\{C(x_i, x_j)\}_{i,j=1,\ldots,n} \geq 0$.

Definition of a DPP

Definition

Let C be a function $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$, called the *kernel*.
$X \sim \mathrm{DPP}(C)$ if

$$\rho^{(n)}(x_1, \ldots, x_n) = \det\{C(x_i, x_j)\}_{i,j=1,\ldots,n}, \quad n = 1, 2, \ldots$$

- The Poisson process with intensity $\rho(x)$ is the special case where $C(x, x) = \rho(x)$ and $C(x, y) = 0$ if $x \neq y$.
- For existence, conditions on the kernel C are mandatory. E.g. $\det\{C(x_i, x_j)\}_{i,j=1,\ldots,n} \geq 0$.

For ease of exposition, assume

(C1) $\quad C$ is a continuous (complex) covariance function.

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

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

- *Intensity*: $\rho(x) = C(x, x)$.

Dept. of Mathematical Sciences
Aalborg University
Denmark

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

- *Intensity*: $\rho(x) = C(x,x)$.
- *Inhibition (negative correlation)*:

$$\rho^{(n)}(x_1,\dots,x_n) \leq \rho(x_1)\cdots\rho(x_n)$$

with equality iff X is a Poisson process with intensity function ρ.

Dept. of Mathematical Sciences
Aalborg University
Denmark

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

- *Intensity*: $\rho(x) = C(x,x)$.
- *Inhibition (negative correlation)*:

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

  with equality iff X is a Poisson process with intensity function ρ.
- *Pair correlation function*:

$$g(x,y) := \frac{\rho^{(2)}(x,y)}{\rho(x)\rho(y)} = 1 - \frac{C(x,y)C(y,x)}{C(x,x)C(y,y)} = 1 - |R(x,y)|^2 \leq 1$$

  where R is the correlation function corresponding to C.

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

- *Intensity*: $\rho(x) = C(x,x)$.
- *Inhibition (negative correlation)*:

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

  with equality iff X is a Poisson process with intensity function ρ.
- *Pair correlation function*:

$$g(x,y) := \frac{\rho^{(2)}(x,y)}{\rho(x)\rho(y)} = 1 - \frac{C(x,y)C(y,x)}{C(x,x)C(y,y)} = 1 - |R(x,y)|^2 \leq 1$$

  where R is the correlation function corresponding to C.
- Any *smooth transformation* or *independent thinning* of X is still a DPP with an explicitly given kernel.

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

- *Intensity*: $\rho(x) = C(x,x)$.
- *Inhibition (negative correlation)*:

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

  with equality iff X is a Poisson process with intensity function ρ.
- *Pair correlation function*:

$$g(x,y) := \frac{\rho^{(2)}(x,y)}{\rho(x)\rho(y)} = 1 - \frac{C(x,y)C(y,x)}{C(x,x)C(y,y)} = 1 - |R(x,y)|^2 \leq 1$$

  where R is the correlation function corresponding to C.
- Any *smooth transformation* or *independent thinning* of X is still a DPP with an explicitly given kernel.
- For any Borel set $B \subset \mathbb{R}^d$, $X_B \sim \mathrm{DPP}(C_B)$ where $C_B(x,y) = C(x,y)$ if $x,y \in B$ and $C_B(x,y) = 0$ else.

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

- *Intensity*: $\rho(x) = C(x,x)$.
- *Inhibition (negative correlation)*:

$$\rho^{(n)}(x_1,\dots,x_n) \le \rho(x_1)\cdots\rho(x_n)$$

  with equality iff X is a Poisson process with intensity function ρ.
- *Pair correlation function*:

$$g(x,y) := \frac{\rho^{(2)}(x,y)}{\rho(x)\rho(y)} = 1 - \frac{C(x,y)C(y,x)}{C(x,x)C(y,y)} = 1 - |R(x,y)|^2 \le 1$$

  where R is the correlation function corresponding to C.
- Any *smooth transformation* or *independent thinning* of X is still a DPP with an explicitly given kernel.
- For any Borel set $B \subset \mathbb{R}^d$, $X_B \sim \mathrm{DPP}(C_B)$ where $C_B(x,y) = C(x,y)$ if $x,y \in B$ and $C_B(x,y) = 0$ else.
- Given C, there *exists at most one* $\mathrm{DPP}(C)$.

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Existence

For any compact set $S \subset \mathbb{R}^d$, there is 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 $\{\phi_k^S\}$ is a set of orthonormal basis functions for $L^2(S)$, i.e.,

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

For any compact set $S \subset \mathbb{R}^d$, there is 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 $\{\phi_k^S\}$ is a set of orthonormal basis functions for $L^2(S)$, i.e.,

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

Theorem (Macchi, 1975)

Under (C1), existence of $\mathrm{DPP}(C)$ *is equivalent to*

$$\textit{spectrum}(C) \leq 1$$

meaning that

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

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Density on a compact set S

Theorem (Macchi, 1975)

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

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Density on a compact set S

Theorem (Macchi, 1975)

Let $X \sim DPP(C)$ and $S \subset \mathbb{R}^d$ be compact. If $\lambda_k^S < 1\ \forall k$, then $X_S \ll Poisson(S, 1)$, with density

$$f_S(\{x_1, \ldots, x_n\}) = \exp(|S| - D) \det\{\tilde{C}(x_i, x_j)\}_{i,j=1,\ldots,n},$$

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

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

Density on a compact set S

Theorem (Macchi, 1975)

Let $X \sim DPP(C)$ and $S \subset \mathbb{R}^d$ be compact. If $\lambda_k^S < 1 \; \forall k$, then $X_S \ll Poisson(S, 1)$, with density

$$f_S(\{x_1, \ldots, x_n\}) = \exp(|S| - D) \det\{\tilde{C}(x_i, x_j)\}_{i,j=1,\ldots,n} \, ,$$

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

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

- Thus to calculate the density/likelihood we need the spectral representation.

Dept. of Mathematical Sciences
Aalborg University
Denmark

Density on a compact set S

Theorem (Macchi, 1975)

Let $X \sim DPP(C)$ and $S \subset \mathbb{R}^d$ be compact. If $\lambda_k^S < 1 \ \forall k$, then $X_S \ll Poisson(S, 1)$, with density

$$f_S(\{x_1, \dots, x_n\}) = \exp(|S| - D) \det\{\tilde{C}(x_i, x_j)\}_{i,j=1,\dots,n} \,,$$

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

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

- Thus to calculate the density/likelihood we need the spectral representation.
- Conversely, using f_S, existence of X_S is ensured by that

$$\lambda_k^S = \frac{\tilde{\lambda}_k^S}{1 + \tilde{\lambda}_k^S} < 1.$$

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Simulation

Let $X \sim \mathrm{DPP}(C)$. Suppose we want to simulate X_S for $S \subset \mathbb{R}^d$ compact.

Simulation

Let $X \sim \mathrm{DPP}(C)$. Suppose we want to simulate X_S for $S \subset \mathbb{R}^d$ compact.

Theorem (Hough, Krishnapur, Peres & Viràg (2006))

Let $B_1, B_2, \ldots$ be independent Bernoulli variables with means $\lambda_1^S, \lambda_2^S, \ldots$, and

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

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Simulation

Let $X \sim \mathrm{DPP}(C)$. Suppose we want to simulate X_S for $S \subset \mathbb{R}^d$ compact.

Theorem (Hough, Krishnapur, Peres & Viràg (2006))

Let $B_1, B_2, \ldots$ be independent Bernoulli variables with means $\lambda_1^S, \lambda_2^S, \ldots$, and

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

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

Dept. of Mathematical Sciences
Aalborg University
Denmark

Simulation

Let $X \sim \mathrm{DPP}(C)$. Suppose we want to simulate X_S for $S \subset \mathbb{R}^d$ compact.

Theorem (Hough, Krishnapur, Peres & Viràg (2006))

Let $B_1, B_2, \ldots$ be independent Bernoulli variables with means $\lambda_1^S, \lambda_2^S, \ldots$, and

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

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

We start by producing n points:

$$n \sim \sum_{k=1}^{\infty} B_k, \quad \mathrm{E}[n] = \sum_{k=1}^{\infty} \lambda_k^S, \quad \mathrm{Var}[n] = \sum_{k=1}^{\infty} \lambda_k^S (1 - \lambda_k^S),$$

where since C is continuous,

$$\sum_{k=1}^{\infty} \lambda_k^S = \int_S C(x,x)\,\mathrm{d}x < \infty.$$

Dept. of Mathematical Sciences
Aalborg University
Denmark

Simulation (cont'd)

Next, if e.g. $B_1 = \ldots = B_n = 1$ and $B_k = 0$ for $k > n$, we simulate the DPP with finite rank kernel

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

Simulation (cont'd)

Next, if e.g. $B_1 = \ldots = B_n = 1$ and $B_k = 0$ for $k > n$, we simulate the DPP with finite rank kernel

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

This is a *projection kernel*, and $\mathrm{DPP}(K)$ can be simulated:

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Simulation (cont'd)

Next, if e.g. $B_1 = \ldots = B_n = 1$ and $B_k = 0$ for $k > n$, we simulate the DPP with finite rank kernel

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

This is a *projection kernel*, and $\mathrm{DPP}(K)$ can be simulated: we have n points $X_1, \ldots, X_n$, where $X_k \,|\, X_1 = x_1, \ldots, X_{k-1} = x_{k-1}$ follows the density

$$p(x_k \,|\, x_1, \ldots, x_{k-1}) = \frac{\rho(x_k \,|\, x_1, \ldots, x_{k-1})}{n-k+1}, \qquad x_k \in S,$$

where

$$\rho(x_k \,|\, x_1, \ldots, x_{k-1}) = \frac{\det\{K(x_i, x_j)\}_{i,j=1,\ldots,k}}{\det\{K(x_i, x_j)\}_{i,j=1,\ldots,k-1}}$$

is the intensity function for a projection DPP with $n-k+1$ points in S.

Simulation (cont'd)

Next, if e.g. $B_1 = \ldots = B_n = 1$ and $B_k = 0$ for $k > n$, we simulate the DPP with finite rank kernel

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

This is a *projection kernel*, and $\mathrm{DPP}(K)$ can be simulated: we have n points $X_1, \ldots, X_n$, where $X_k \mid X_1 = x_1, \ldots, X_{k-1} = x_{k-1}$ follows the density

$$p(x_k \mid x_1, \ldots, x_{k-1}) = \frac{\rho(x_k \mid x_1, \ldots, x_{k-1})}{n-k+1}, \qquad x_k \in S,$$

where

$$\rho(x_k \mid x_1, \ldots, x_{k-1}) = \frac{\det\{K(x_i, x_j)\}_{i,j=1,\ldots,k}}{\det\{K(x_i, x_j)\}_{i,j=1,\ldots,k-1}}$$

is the intensity function for a projection DPP with $n-k+1$ points in S.

At each step we use rejection sampling...

Status

- We let a DPP on $\mathbb{R}^d$ be specified through a continuous (complex) covariance function $C : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$.

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Status

- We let a DPP on $\mathbb{R}^d$ be specified through a continuous (complex) covariance function $C : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$.
- C determines the moment properties of the DPP.

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Status

- We let a DPP on $\mathbb{R}^d$ be specified through a continuous (complex) covariance function $C : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$.
- C determines the moment properties of the DPP.
- Given the spectral representation of C on a compact set S:
  - there is an existence condition;
  - the distribution of the number of points falling in S is known;
  - can simulate the process on S;
  - can calculate the density/likelihood;
  - have no problem with boundary effects.

Status

- We let a DPP on $\mathbb{R}^d$ be specified through a continuous (complex) covariance function $C : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$.
- C determines the moment properties of the DPP.
- Given the spectral representation of C on a compact set S:
  - there is an existence condition;
  - the distribution of the number of points falling in S is known;
  - can simulate the process on S;
  - can calculate the density/likelihood;
  - have no problem with boundary effects.

But we need to know the spectral representation — typically we don't!

Stationary kernels

Consider a stationary kernel,

$$C(x, y) = C_0(x - y), \qquad x, y \in \mathbb{R}^d,$$

with Fourier transform (spectral density) $\varphi := \mathcal{F}(C_0)$:

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

Dept. of Mathematical Sciences
Aalborg University
Denmark

Stationary kernels

Consider a stationary kernel,

$$C(x,y) = C_0(x-y), \qquad x,y \in \mathbb{R}^d,$$

with Fourier transform (spectral density) $\varphi := \mathcal{F}(C_0)$:

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

Theorem (Lavancier, Møller & Rubak, 2015)

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

$$DPP(C_0) \text{ exists} \quad \Leftrightarrow \quad \varphi \le 1.$$

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Stationary kernels

Consider a stationary kernel,

$$C(x,y) = C_0(x-y), \qquad x,y \in \mathbb{R}^d,$$

with Fourier transform (spectral density) $\varphi := \mathcal{F}(C_0)$:

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

Theorem (Lavancier, Møller & Rubak, 2015)

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

$$DPP(C_0) \text{ exists} \quad \Leftrightarrow \quad \varphi \leq 1.$$

$\rightarrow$ *This induces a restriction*:

There will be a trade-off between intensity and the degree of repulsiveness:

If the intensity is large, the "effective support" of $C_0 = \mathcal{F}^{-1}(\varphi)$ must be small (i.e. small range of correlation/interaction, as detailed later).

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Approximation

For simplicity and specificity, consider $S = [-1/2, 1/2]^d$.

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Approximation

For simplicity and specificity, consider $S = [-1/2, 1/2]^d$.
Approximate X_S by $X^{\text{app}} \sim DPP_S(C_{\mathrm{app}})$ where

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

Turns out to work well in practice!

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Approximation

For simplicity and specificity, consider $S = [-1/2, 1/2]^d$.
Approximate X_S by $X^{\mathrm{app}} \sim DPP_S(C_{\mathrm{app}})$ where

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

Turns out to work well in practice!

For $x - y \in S$, it is effectively the Fourier expansion

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

since for "interesting models" (for any reasonable expected number of points) the existence condition $\varphi \leq 1$ implies $C_0(t) \approx 0$ for $t \notin S$ and

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

Dept. of Mathematical Sciences
Aalborg University
Denmark

Modelling based on spectral densities

Idea: Start by modelling $0 \leq \varphi \leq 1$ s.t. $C_0 := \mathcal{F}^{-1}(\varphi) \in L^2(\mathbb{R}^d)$.

Dept. of Mathematical Sciences
Aalborg University
Denmark

Modelling based on spectral densities

Idea: Start by modelling $0 \leq \varphi \leq 1$ s.t. $C_0 := \mathcal{F}^{-1}(\varphi) \in L^2(\mathbb{R}^d)$.

Then $DPP(C_0)$ is well-defined and using the approximation based on the Fourier basis, X_S can easily be simulated and the density/likelihood can be evaluated exactly (up to series truncation).

Modelling based on spectral densities

<u>Idea</u>: Start by modelling $0 \leq \varphi \leq 1$ s.t. $C_0 := \mathcal{F}^{-1}(\varphi) \in L^2(\mathbb{R}^d)$.

Then $DPP(C_0)$ is well-defined and using the approximation based on the Fourier basis, X_S can easily be simulated and the density/likelihood can be evaluated exactly (up to series truncation).

Main drawback:

- In general the parameters in $C_0 = \mathcal{F}^{-1}(\varphi)$ (and thus in $\rho^{(n)}$) may be hard to understand/interpret.

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Intermezzo

This concludes the first part of the talk focusing on

- the probabilistic background,
- approximations for simulation and density expression.

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Intermezzo

This concludes the first part of the talk focusing on

- the probabilistic background,
- approximations for simulation and density expression.

Now we start doing statistics, so if you got lost or fell asleep you get a fresh start!

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Examples of parametric models

Focus on the following parametric models
($\rho > 0$ intensity; $\alpha > 0$ scale/range; $\nu > 0$ shape):

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Examples of parametric models

Focus on the following parametric models
($\rho > 0$ intensity; $\alpha > 0$ scale/range; $\nu > 0$ shape):

- *Whittle-Matérn model*, which includes the exponential model ($\nu = 1/2$) and the Gaussian model ($\nu = \infty$):

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

  Parameter restriction:

$$\rho \leq \frac{\Gamma(\nu)}{\Gamma(\nu + d/2)(2\sqrt{\pi}\alpha)^d}.$$

Examples of parametric models

Focus on the following parametric models
($\rho > 0$ intensity; $\alpha > 0$ scale/range; $\nu > 0$ shape):

- *Whittle-Matérn model*, which includes the exponential model ($\nu = 1/2$) and the Gaussian model ($\nu = \infty$):

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

  Parameter restriction:

$$\rho \leq \frac{\Gamma(\nu)}{\Gamma(\nu + d/2)(2\sqrt{\pi}\alpha)^d}.$$

- *Power exponential spectral model*:

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

  Parameter restriction:

$$\rho \leq \frac{\pi^{d/2}\Gamma(d/\nu+1)}{\Gamma(d/2)\alpha^d}.$$

Dept. of Mathematical Sciences
Aalborg University
Denmark

Parametric models in R (available via the spatstat package)

Parametric models are specified in R via the determinantal family functions (of class `detfamily`): `detGauss`, `detMatern`, `detPowerExp`.
E.g:

- `model <- detGauss(rho=100, alpha=0.05, d=2)`
- `model <- detMatern(rho=100, alpha=0.03, nu=0.5, d=2)`
- `model <- detPowerExp(rho=100, alpha=0.17, nu=2, d=2)`

Dept. of Mathematical Sciences
Aalborg University
Denmark

Parametric models in R (available via the spatstat package)

Parametric models are specified in R via the determinantal family functions (of class `detfamily`): `detGauss`, `detMatern`, `detPowerExp`.
E.g:

- `model <- detGauss(rho=100, alpha=0.05, d=2)`
- `model <- detMatern(rho=100, alpha=0.03, nu=0.5, d=2)`
- `model <- detPowerExp(rho=100, alpha=0.17, nu=2, d=2)`

Extract the kernel, spectral density, pair correlation function, K-function:

- `detkernel(model)`
- `detspecden(model)`
- `pcfmodel(model)`
- `Kmodel(model)`

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Simulation in R

Simply use the generic function `simulate` (then R automatically calls the function `simulate.detmodel`):

- ```
  model <- detGauss(rho=100, alpha=0.05, d=2)
  X <- simulate(model)
  ```


Dept. of Mathematical Sciences
Aalborg University
Denmark
  ```

Simulation in R

Simply use the generic function `simulate` (then R automatically calls the function `simulate.detmodel`):

- ```
  model <- detGauss(rho=100, alpha=0.05, d=2)
  X <- simulate(model)
  ```
- Change the window (default is the unit square):
  ```
  W <- owin(poly=list(x=c(-1,0,1),y=c(0,1,0)))
  X <- simulate(model, W=W)
  ```
  ```

Simulation in R

Simply use the generic function `simulate` (then R automatically calls the function `simulate.detmodel`):

- ```
  model <- detGauss(rho=100, alpha=0.05, d=2)
  X <- simulate(model)
  ```
- Change the window (default is the unit square):
  ```
  W <- owin(poly=list(x=c(-1,0,1),y=c(0,1,0)))
  X <- simulate(model, W=W)
  ```
- Several realizations:
  ```
  X <- simulate(model, nsim=4)
  ```
  ```

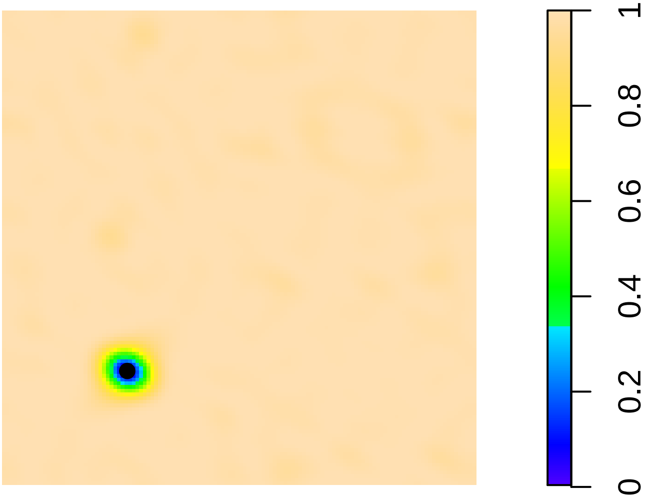
Illustration of simulation algorithm

Step 1. The first point is sampled uniformly on S (stationary case).

Dept. of Mathematical Sciences
Aalborg University
Denmark

Illustration of simulation algorithm

Step 2. The second point is sampled w.r.t. the following density:

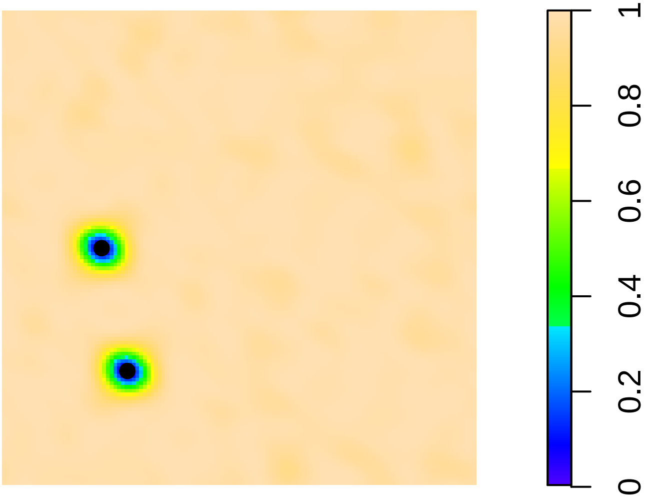


Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

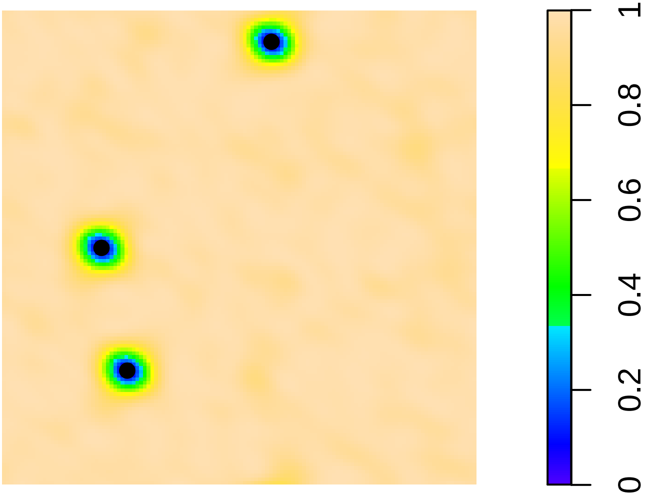
Illustration of simulation algorithm

Step 3. The third point is sampled w.r.t. the following density:



Dept. of Mathematical Sciences
Aalborg University
Denmark

Step 4. The fourth point is sampled w.r.t. the following density:

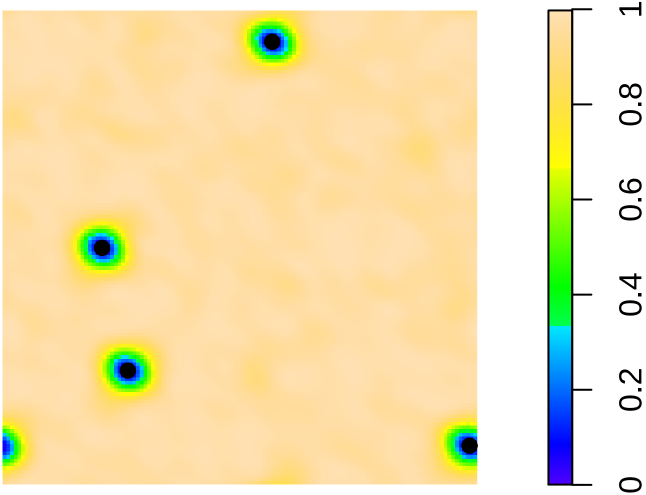


Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Illustration of simulation algorithm

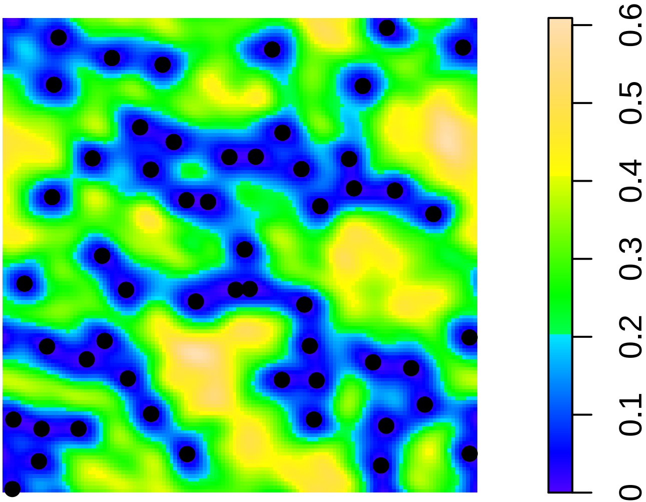
Step 5. The sixth point is sampled w.r.t. the following density:



Dept. of Mathematical Sciences
Aalborg University
Denmark

Illustration of simulation algorithm

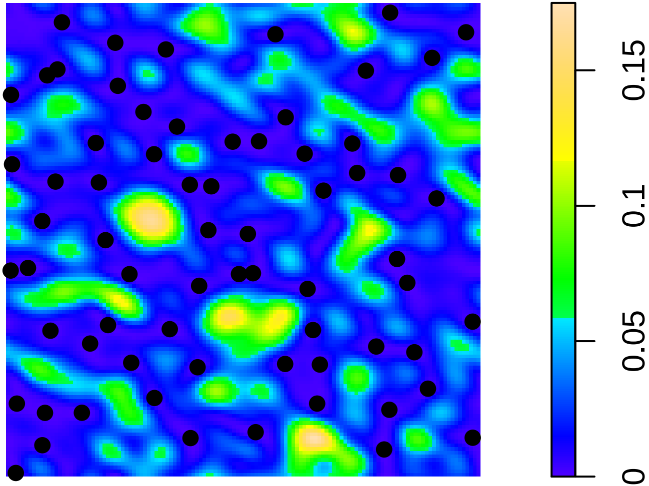
...somewhere in the middle...



Dept. of Mathematical Sciences
Aalborg University
Denmark

Illustration of simulation algorithm

Final point is sampled w.r.t. the following density:

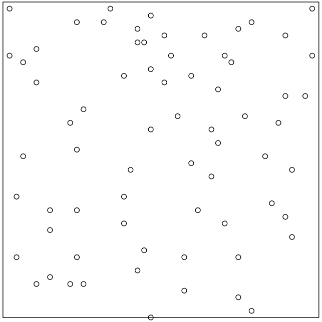


Dept. of Mathematical Sciences
Aalborg University
Denmark

Spanish towns dataset

Ripley (1988): Strauss hard-core model with 4 parameters:

r=hard-core, R=range of interaction, β=abundance, γ=interaction.



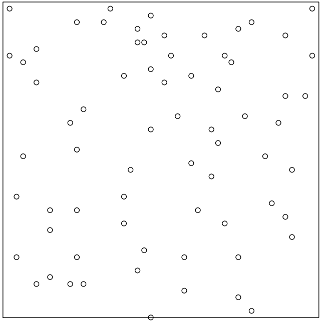
Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Spanish towns dataset

Ripley (1988): Strauss hard-core model with 4 parameters:

r=hard-core, R=range of interaction, β=abundance, γ=interaction.



Following Illian, Penttinen, Stoyan & Stoyan (2008): $\hat{r} = 0.83$, $\hat{R} = 3.5$.
Approximate likelihood method (Huang and Ogata, 1999): $\hat{\beta} = 0.12$ and $\hat{\gamma} = 0.76$.

Alternative DPP models

Gaussian, Whittle-Matérn, and power exponential spectral models fitted using the function `dppm`:

- Default estimation method is "partial likelihood" where we use $\hat{\rho} = n/|W| = 0.043$ and MLEs for the rest:

```
fit <- dppm(X, detGauss())
```

- Full likelihood:

```
fit <- dppm(X, detGauss(), method="likelihood")
```

Alternative DPP models

Gaussian, Whittle-Matérn, and power exponential spectral models fitted using the function `dppm`:

- Default estimation method is "partial likelihood" where we use $\hat{\rho} = n/|W| = 0.043$ and MLEs for the rest:
  ```
  fit <- dppm(X, detGauss())
  ```
- Full likelihood:
  ```
  fit <- dppm(X, detGauss(), method="likelihood")
  ```

Highest likelihood: fitted Whittle-Matérn model.

Dept. of Mathematical Sciences
Aalborg University
Denmark

Alternative DPP models

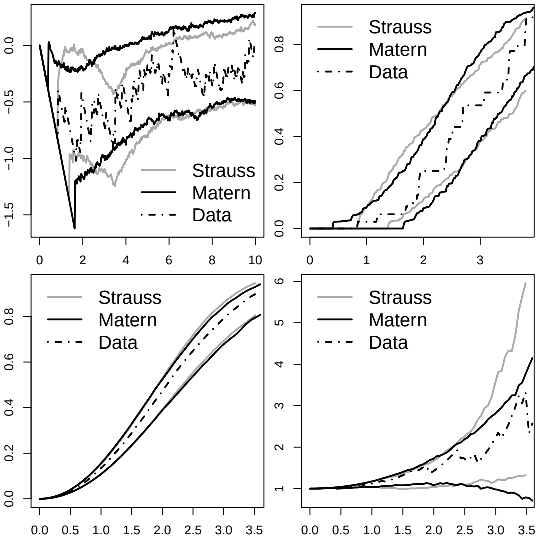
Gaussian, Whittle-Matérn, and power exponential spectral models fitted using the function `dppm`:

- Default estimation method is "partial likelihood" where we use $\hat{\rho} = n/|W| = 0.043$ and MLEs for the rest:
  ```
  fit <- dppm(X, detGauss())
  ```
- Full likelihood:
  ```
  fit <- dppm(X, detGauss(), method="likelihood")
  ```

Highest likelihood: fitted Whittle-Matérn model.

Simulation based likelihood-ratio test for the simpler Gaussian model vs the Whittle-Matérn model: $p = 3\%$.

Clockwise from top left: Non-parametric estimate of $L(r) - r$, $G(r)$, $J(r)$, $F(r)$, and simulation based 2.5% and 97.5% pointwise quantiles (based on 400 realizations).



Conclusion of data analysis

Whittle-Matérn model:

- has **less parameters**
- (arguably) provides a better fit
- has a canonical way of estimating parameters (**likelihood**)
- direct **access to the moments** (intensity, pair correlation function, ...)

Dept. of Mathematical Sciences
Aalborg University
Denmark

Conclusion of data analysis

Whittle-Matérn model:

- has **less parameters**
- (arguably) provides a better fit
- has a canonical way of estimating parameters (**likelihood**)
- direct **access to the moments** (intensity, pair correlation function, ...)

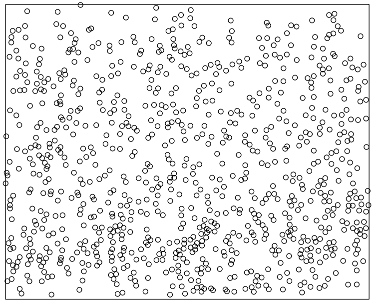
For the Strauss hard-core model

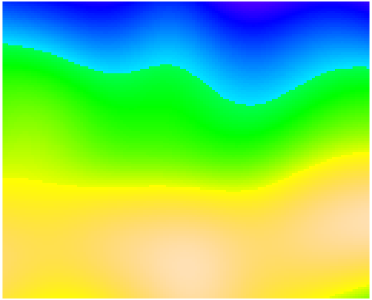
- parameter estimation relies to a certain extend on “ad-hoc” methods
- the density and moments can only be obtained by MCMC simulation.

Dept. of Mathematical Sciences
Aalborg University
Denmark

Mucous membrane dataset

Consists of the most abundant type of cell in a bivariate point pattern analysed in Møller and Waagepetersen (2004).



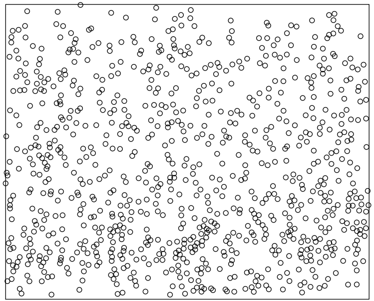


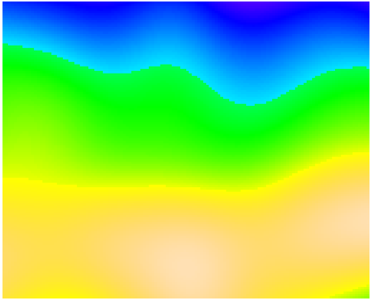
Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Mucous membrane dataset

Consists of the most abundant type of cell in a bivariate point pattern analysed in Møller and Waagepetersen (2004).





We use this to illustrate how an **inhomogenous DPP** can be fitted to a real dataset.

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Modelling inhomogeneity

Assume the correlation function is translation invariant:

$$R(x,y) = \frac{C(x,y)}{\sqrt{C(x,x)C(y,y)}} = \frac{C(x,y)}{\sqrt{\rho(x)\rho(y)}} = R_0(x-y).$$

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Modelling inhomogeneity

Assume the correlation function is translation invariant:

$$R(x,y) = \frac{C(x,y)}{\sqrt{C(x,x)C(y,y)}} = \frac{C(x,y)}{\sqrt{\rho(x)\rho(y)}} = R_0(x-y).$$

This implies

$$g(x,y) = g_0(x-y) = 1 - |R_0(x-y)|^2$$

(*second-order intensity-reweighted stationarity*; Baddeley, Møller & Waagepetersen, 2000).

Dept. of Mathematical Sciences
Aalborg University
Denmark

Modelling inhomogeneity

Assume the correlation function is translation invariant:

$$R(x,y) = \frac{C(x,y)}{\sqrt{C(x,x)C(y,y)}} = \frac{C(x,y)}{\sqrt{\rho(x)\rho(y)}} = R_0(x-y).$$

This implies

$$g(x,y) = g_0(x-y) = 1 - |R_0(x-y)|^2$$

(*second-order intensity-reweighted stationarity*; Baddeley, Møller & Waagepetersen, 2000).

- 1. Fit a parametric model to ρ depending on relevant covariates (second coordinate axis in our case).

Dept. of Mathematical Sciences
Aalborg University
Denmark

Modelling inhomogeneity

Assume the correlation function is translation invariant:

$$R(x,y) = \frac{C(x,y)}{\sqrt{C(x,x)C(y,y)}} = \frac{C(x,y)}{\sqrt{\rho(x)\rho(y)}} = R_0(x-y).$$

This implies

$$g(x,y) = g_0(x-y) = 1 - |R_0(x-y)|^2$$

(*second-order intensity-reweighted stationarity*; Baddeley, Møller & Waagepetersen, 2000).

- 1. Fit a parametric model to ρ depending on relevant covariates (second coordinate axis in our case).
- 2. Use this when estimating g_0 (by kernel methods).

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Modelling inhomogeneity

Assume the correlation function is translation invariant:

$$R(x, y) = \frac{C(x, y)}{\sqrt{C(x, x)C(y, y)}} = \frac{C(x, y)}{\sqrt{\rho(x)\rho(y)}} = R_0(x - y).$$

This implies

$$g(x, y) = g_0(x - y) = 1 - |R_0(x - y)|^2$$

(*second-order intensity-reweighted stationarity*; Baddeley, Møller & Waagepetersen, 2000).

- 1. Fit a parametric model to ρ depending on relevant covariates (second coordinate axis in our case).
- 2. Use this when estimating g_0 (by kernel methods).
- 3. Fit a parametric model for g_0 via a minimum contrast method.

Modelling inhomogeneity

Assume the correlation function is translation invariant:

$$R(x,y) = \frac{C(x,y)}{\sqrt{C(x,x)C(y,y)}} = \frac{C(x,y)}{\sqrt{\rho(x)\rho(y)}} = R_0(x-y).$$

This implies

$$g(x,y) = g_0(x-y) = 1 - |R_0(x-y)|^2$$

(*second-order intensity-reweighted stationarity*; Baddeley, Møller & Waagepetersen, 2000).

- 1. Fit a parametric model to ρ depending on relevant covariates (second coordinate axis in our case).
- 2. Use this when estimating g_0 (by kernel methods).
- 3. Fit a parametric model for g_0 via a minimum contrast method.
- 4. The estimated kernel of the DPP is

$$\hat{C}(x,y) = \sqrt{\hat{\rho}(x)}\hat{R}_0(x-y)\sqrt{\hat{\rho}(y)}.$$

Dept. of Mathematical Sciences
Aalborg University
Denmark

Simulation of the inhomogeneous DPP

NB: If a DPP $\tilde{X}$ with kernel $\tilde{C}(x, y)$ is independently thinned with retention probability $\pi(x)$ for $x \in \mathbb{R}^d$, the resulting process is a DPP with kernel

$$C(x, y) = \sqrt{\pi(x)}\tilde{C}(x, y)\sqrt{\pi(y)}.$$

Simulation of the inhomogeneous DPP

NB: If a DPP $\tilde{X}$ with kernel $\tilde{C}(x, y)$ is independently thinned with retention probability $\pi(x)$ for $x \in \mathbb{R}^d$, the resulting process is a DPP with kernel

$$C(x, y) = \sqrt{\pi(x)}\tilde{C}(x, y)\sqrt{\pi(y)}.$$

Now, let $\hat{\rho}_{\max} = \sup_{x \in S}\{\hat{\rho}(x)\}$ and define a stationary DPP $\tilde{X}$ with kernel

$$\tilde{C}(x, y) := \tilde{C}_0(x - y) := \hat{\rho}_{\max}\hat{R}_0(x - y).$$

Dept. of Mathematical Sciences
Aalborg University
Denmark

Simulation of the inhomogeneous DPP

NB: If a DPP $\tilde{X}$ with kernel $\tilde{C}(x,y)$ is independently thinned with retention probability $\pi(x)$ for $x \in \mathbb{R}^d$, the resulting process is a DPP with kernel

$$C(x,y) = \sqrt{\pi(x)}\tilde{C}(x,y)\sqrt{\pi(y)}.$$

Now, let $\hat{\rho}_{\max} = \sup_{x\in S}\{\hat{\rho}(x)\}$ and define a stationary DPP $\tilde{X}$ with kernel

$$\tilde{C}(x,y) := \tilde{C}_0(x-y) := \hat{\rho}_{\max}\hat{R}_0(x-y).$$

Then our fitted model is simulated by thinning $\tilde{X}$ with retention probability $\pi(x) = \hat{\rho}(x)/\hat{\rho}_{\max}$, since

$$C(x,y) = \sqrt{\frac{\hat{\rho}(x)}{\hat{\rho}_{\max}}}\tilde{C}(x,y)\sqrt{\frac{\hat{\rho}(y)}{\hat{\rho}_{\max}}} = \sqrt{\hat{\rho}(x)}\hat{R}_0(x-y)\sqrt{\hat{\rho}(y)} = \hat{C}(x,y).$$

Plots (omitted) of empirical summaries compared to simulations show a satisfactory fit.

Dept. of Mathematical Sciences
Aalborg University
Denmark

- On the sphere the spherical harmonics constitute a set of basis functions (given in terms of associated Legendre polynomials).

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

DPPs on the sphere

- On the sphere the spherical harmonics constitute a set of basis functions (given in terms of associated Legendre polynomials).
- Thus, to obtain a parametric model for a DPP on the sphere, we only have to make a parametric model for the eigenvalues λ_k.

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

DPPs on the sphere

- On the sphere the spherical harmonics constitute a set of basis functions (given in terms of associated Legendre polynomials).
- Thus, to obtain a parametric model for a DPP on the sphere, we only have to make a parametric model for the eigenvalues λ_k.
- Møller et al. (2018): Using classical results by Schönberg and others $\rightarrow$ characterized all *isotropic covariance functions* on the sphere and related this to DPPs.

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

DPPs on the sphere

- On the sphere the spherical harmonics constitute a set of basis functions (given in terms of associated Legendre polynomials).
- Thus, to obtain a parametric model for a DPP on the sphere, we only have to make a parametric model for the eigenvalues λ_k.
- Møller et al. (2018): Using classical results by Schönberg and others $\rightarrow$ characterized all *isotropic covariance functions* on the sphere and related this to DPPs.
- One isotropic covariance function model with known eigenvalues is the Inverse MultiQuadric, which we have implemented in R:

```
model <- detIMQ(rho=500,delta=0.998)
```

DPPs on the sphere

- On the sphere the spherical harmonics constitute a set of basis functions (given in terms of associated Legendre polynomials).
- Thus, to obtain a parametric model for a DPP on the sphere, we only have to make a parametric model for the eigenvalues λ_k.
- Møller et al. (2018): Using classical results by Schönberg and others $\rightarrow$ characterized all *isotropic covariance functions* on the sphere and related this to DPPs.
- One isotropic covariance function model with known eigenvalues is the Inverse MultiQuadric, which we have implemented in R:
  ```
  model <- detIMQ(rho=500,delta=0.998)
  ```
- Simulations rely on the previously developed code (with some modifications):
  ```
  X <- simulate(model)
  ```

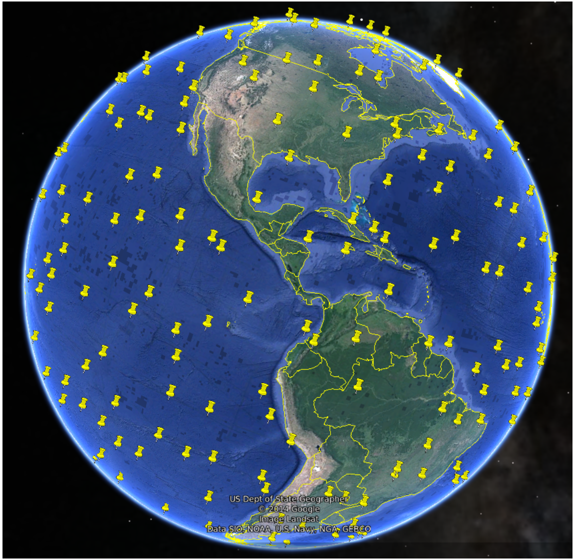
Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

DPPs on the sphere

- On the sphere the spherical harmonics constitute a set of basis functions (given in terms of associated Legendre polynomials).
- Thus, to obtain a parametric model for a DPP on the sphere, we only have to make a parametric model for the eigenvalues λ_k.
- Møller et al. (2018): Using classical results by Schönberg and others $\rightarrow$ characterized all *isotropic covariance functions* on the sphere and related this to DPPs.
- One isotropic covariance function model with known eigenvalues is the Inverse MultiQuadric, which we have implemented in R:
```
model <- detIMQ(rho=500,delta=0.998)
```
- Simulations rely on the previously developed code (with some modifications):
```
X <- simulate(model)
```
- Møller & Rubak (2016) studied Palm distributions and functional summaries for DPPs (and general point processes) on the sphere....

A simulated DPP consisting of 441 points on planet Earth



A new coupling result

Consider $X \sim \text{DPP}(C)$, defined on a locally compact Polish space Ω (e.g. $\mathbb{R}^d$ or $\mathbb{S}^d$), and satisfying certain weak conditions (C is a complex covariance function of locally trace class and with spectrum ≤ 1).

A new coupling result

Consider $X \sim \text{DPP}(C)$, defined on a locally compact Polish space Ω (e.g. $\mathbb{R}^d$ or $\mathbb{S}^d$), and satisfying certain weak conditions (C is a complex covariance function of locally trace class and with spectrum ≤ 1).

Let $u \in \Omega$ such that $C(u,u) > 0$, and let X^u follow the *reduced Palm distribution at* u: Intuitively, $X^u \sim X \setminus \{u\} | u \in X$.

A new coupling result

Consider $X \sim \text{DPP}(C)$, defined on a locally compact Polish space Ω (e.g. $\mathbb{R}^d$ or $\mathbb{S}^d$), and satisfying certain weak conditions (C is a complex covariance function of locally trace class and with spectrum ≤ 1).

Let $u \in \Omega$ such that $C(u,u) > 0$, and let X^u follow the *reduced Palm distribution at* u: Intuitively, $X^u \sim X \setminus \{u\} | u \in X$. In fact

$$X^u \sim \text{DPP}(C^u), \qquad C^u(x,y) = C(x,y) - \frac{C(x,u)C(u,y)}{C(u,u)}.$$

Dept. of Mathematical Sciences
Aalborg University
Denmark

A new coupling result

Consider $X \sim \text{DPP}(C)$, defined on a locally compact Polish space Ω (e.g. $\mathbb{R}^d$ or $\mathbb{S}^d$), and satisfying certain weak conditions (C is a complex covariance function of locally trace class and with spectrum ≤ 1).

Let $u \in \Omega$ such that $C(u,u) > 0$, and let X^u follow the *reduced Palm distribution at* u: Intuitively, $X^u \sim X \setminus \{u\} | u \in X$. In fact

$$X^u \sim \text{DPP}(C^u), \qquad C^u(x,y) = C(x,y) - \frac{C(x,u)C(u,y)}{C(u,u)}.$$

Theorem (Møller and O'Reilly, 2018 – Extension of Goldman, 2010)

There exists a coupling of X and X^u such that almost surely

$$X^u \subseteq X \qquad \text{and} \qquad \kappa^u := X \setminus X^u \text{ consists of at most one point}$$

and κ^u has intensity function

$$\rho_{\kappa^u}(x) = \rho_X(x) - \rho_{X^u}(x) = C(x,x) - C^u(x,x) = \frac{|C(x,u)|^2}{C(u,u)}.$$

Dept. of Mathematical Sciences
Aalborg University
Denmark

Quantifying repulsiveness

A *global measure of repulsiveness*:

$$p_u := \mathrm{P}\left(\kappa^u \neq \emptyset\right) = \int \frac{|C(x,u)|^2}{C(u,u)}\,\mathrm{d}x.$$

- In line with Lavancier, Møller & Rubak (2014, 2015).

Quantifying repulsiveness

A *global measure of repulsiveness*:

$$p_u := \mathrm{P}\left(\kappa^u \neq \emptyset\right) = \int \frac{|C(x,u)|^2}{C(u,u)}\,\mathrm{d}x.$$

- In line with Lavancier, Møller & Rubak (2014, 2015).
- For a Poisson process, $p_u = 0$ (for all u with $K(u,u) > 0$).

Quantifying repulsiveness

A *global measure of repulsiveness*:

$$p_u := \mathrm{P}\left(\kappa^u \neq \emptyset\right) = \int \frac{|C(x,u)|^2}{C(u,u)}\,\mathrm{d}x.$$

- In line with Lavancier, Møller & Rubak (2014, 2015).
- For a Poisson process, $p_u = 0$ (for all u with $K(u,u) > 0$).
- For a *most repulsive DPP*, $p_u = 1$ (for all u with $K(u,u) > 0$).

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Quantifying repulsiveness

A *global measure of repulsiveness*:

$$p_u := \mathrm{P}\left(\kappa^u \neq \emptyset\right) = \int \frac{|C(x,u)|^2}{C(u,u)}\,\mathrm{d}x.$$

- In line with Lavancier, Møller & Rubak (2014, 2015).
- For a Poisson process, $p_u = 0$ (for all u with $K(u,u) > 0$).
- For a *most repulsive DPP*, $p_u = 1$ (for all u with $K(u,u) > 0$).
- Trade-off between intensity and pcf:

$$p_u = \int \rho(v)(1 - g(u,v))\,\mathrm{d}v.$$

Dept. of Mathematical Sciences
Aalborg University
Denmark

Quantifying repulsiveness

A *global measure of repulsiveness*:

$$p_u := \mathrm{P}\left(\kappa^u \neq \emptyset\right) = \int \frac{|C(x,u)|^2}{C(u,u)}\,\mathrm{d}x.$$

- In line with Lavancier, Møller & Rubak (2014, 2015).
- For a Poisson process, $p_u = 0$ (for all u with $K(u,u) > 0$).
- For a *most repulsive DPP*, $p_u = 1$ (for all u with $K(u,u) > 0$).
- Trade-off between intensity and pcf:

$$p_u = \int \rho(v)(1 - g(u,v))\,\mathrm{d}v.$$

Conditioned on $\kappa^u \neq \emptyset$, the point in κ^u has density

$$f_u(x) = \frac{|C(x,u)|^2}{\|C(\cdot,u)\|_2^2}, \qquad x \in \Omega.$$

Dept. of Mathematical Sciences
Aalborg University
Denmark

Example 1: Ginibre DPP

The standard Ginibre point process is a stationary and isotropic DPP on $\mathbb{C} \equiv \mathbb{R}^2$ with kernel

$$C(x, y) = \frac{1}{\pi} \exp\left(x\overline{y} - \frac{|x|^2 + |y|^2}{2}\right), \qquad x, y \in \mathbb{C}.$$

NB: This kernel is not stationary!

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Example 1: Ginibre DPP

The standard Ginibre point process is a stationary and isotropic DPP on $\mathbb{C} \equiv \mathbb{R}^2$ with kernel

$$C(x, y) = \frac{1}{\pi} \exp\left(x\overline{y} - \frac{|x|^2 + |y|^2}{2} \right), \qquad x, y \in \mathbb{C}.$$

NB: This kernel is not stationary!

We have

$$\rho_{\kappa^u}(x) = \frac{|C(x, u)|^2}{C(u, u)} = \frac{1}{\pi} \exp(-|x - u|^2) \sim \mathrm{N}_{\mathbb{C}}(u, 1),$$

so $p_u = 1$ (i.e. most repulsive) and if Z_u is the point in κ^u, then

$$Z_u \sim \mathrm{N}_{\mathbb{C}}(u, 1).$$

Dept. of Mathematical Sciences
Aalborg University
Denmark

Example 2: Most repulsive DPPs with a stationary kernel

Fact: a DPP with a stationary kernel satisfying (C1)-(C2) is *most repulsive* iff φ is an indicator function with $\int \varphi = \rho$ (Lavancier, Møller & Rubak, 2014, 2015).

Example 2: Most repulsive DPPs with a stationary kernel

Fact: a DPP with a stationary kernel satisfying (C1)-(C2) is *most repulsive* iff φ is an indicator function with $\int \varphi = \rho$ (Lavancier, Møller & Rubak, 2014, 2015).

Natural candidate:

$$\varphi(x) \propto \mathbf{1}_{\{\|x\| \leq r\}}.$$

Then C_0 is a sinc function if $d = 1$ and a jinc-like function

$$C_0(x) = \frac{J_1(2\|x\|)}{\pi\|x\|} \qquad \text{if } d = 2.$$

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

K-functions of Ginibre and jinc-like DPPs ($\rho = 1/\pi$)

For a stationary point process: $K(r) = \int_{\|u\| \le r} g(u) \, \mathrm{d}u$, $\rho K(r) = \mathrm{E}\left[X^0(b(0,r))\right]$.

Dept. of Mathematical Sciences
Aalborg University
Denmark

K-functions of Ginibre and jinc-like DPPs ($\rho = 1/\pi$)

For a stationary point process: $K(r) = \int_{\|u\| \leq r} g(u)\,\mathrm{d}u$, $\rho K(r) = \mathrm{E}\left[X^0(b(0, r))\right]$.

Consider plot of $L(r) - r$ vs $r > 0$, where $L(r) = \sqrt{K(r)/\pi}$,
$L(r) - r = 0$ if Poisson, $L(r) - r \leq 1 \rightarrow$ repulsiveness.

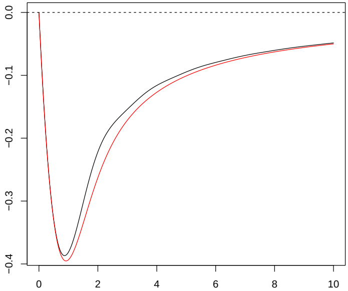
Dept. of Mathematical Sciences
Aalborg University
Denmark

K-functions of Ginibre and jinc-like DPPs ($\rho = 1/\pi$)

For a stationary point process: $K(r) = \int_{\|u\| \le r} g(u)\,\mathrm{d}u$, $\rho K(r) = \mathrm{E}\left[X^0(b(0,r))\right]$.

Consider plot of $L(r) - r$ vs $r > 0$, where $L(r) = \sqrt{K(r)/\pi}$,
$L(r) - r = 0$ if Poisson, $L(r) - r \le 1 \to$ repulsiveness.

Red: $L(r) - r$ for Ginibre DPP. Black: $L(r) - r$ for jinc-like DPP.

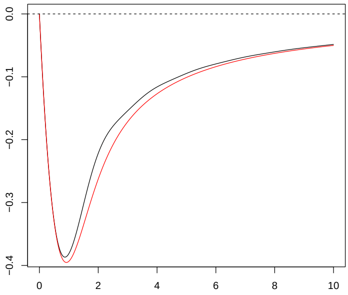


K-functions of Ginibre and jinc-like DPPs ($\rho = 1/\pi$)

For a stationary point process: $K(r) = \int_{\|u\| \le r} g(u)\,\mathrm{d}u$, $\rho K(r) = \mathrm{E}\left[X^0(b(0,r))\right]$.

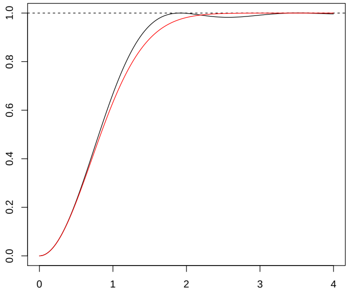
Consider plot of $L(r) - r$ vs $r > 0$, where $L(r) = \sqrt{K(r)/\pi}$,
$L(r) - r = 0$ if Poisson, $L(r) - r \le 1 \rightarrow$ repulsiveness.

Red: $L(r) - r$ for Ginibre DPP. Black: $L(r) - r$ for jinc-like DPP.



So is Ginibre DPP more repulsive?

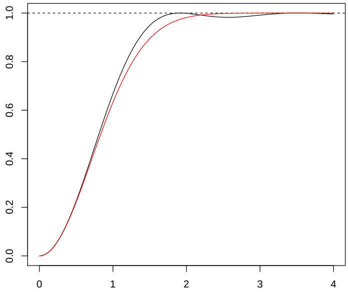
Pcfs $g_0(r) = g(u, v)$ $(r = \|u - v\|)$ of Ginibre and jinc-like DPPs



No: the two DPPs are equally repulsive according to the global repulsiveness criteria (that averages at all scales through the integral over the whole space).

Dept. of Mathematical Sciences
Aalborg University
Denmark

Pcfs $g_0(r) = g(u, v)$ $(r = \|u - v\|)$ of Ginibre and jinc-like DPPs

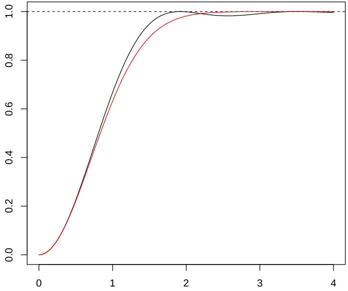


No: the two DPPs are equally repulsive according to the global repulsiveness criteria (that averages at all scales through the integral over the whole space).

• At small scales: similar behaviour (same curvature at 0 $\rightarrow$ equally locally repulsive, cf. Biscio & Lavancier, 2016).

Dept. of Mathematical Sciences
Aalborg University
Denmark

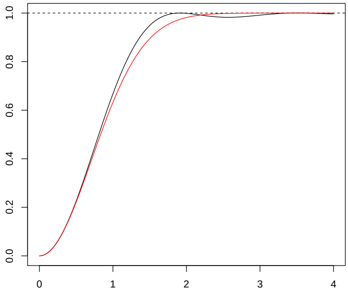
Pcfs $g_0(r) = g(u, v)$ ($r = \|u - v\|$) of Ginibre and jinc-like DPPs



No: the two DPPs are equally repulsive according to the global repulsiveness criteria (that averages at all scales through the integral over the whole space).

- At small scales: similar behaviour (same curvature at 0 $\rightarrow$ equally locally repulsive, cf. Biscio & Lavancier, 2016).
- At medium scales: Ginibre DPP more repulsive.

Pcfs $g_0(r) = g(u, v)$ $(r = \|u - v\|)$ of Ginibre and jinc-like DPPs



No: the two DPPs are equally repulsive according to the global repulsiveness criteria (that averages at all scales through the integral over the whole space).

• At small scales: similar behaviour (same curvature at 0 $\rightarrow$ equally locally repulsive, cf. Biscio & Lavancier, 2016).

• At medium scales: Ginibre DPP more repulsive.

• At large scales: jinc-like DPP more repulsive (the distribution of Z_0 is heavy tailed, cf. Møller and O'Reilly, 2018).

Dept. of Mathematical Sciences
Aalborg University
Denmark

Conclusion

DPP's possess **appealing properties**:

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

DPP's possess **appealing properties**:

- They provide flexible parametric models of repulsive point processes ("soft-core" cases and cases with more repulsion).

Conclusion

DPP's possess **appealing properties**:

- They provide flexible parametric models of repulsive point processes ("soft-core" cases and cases with more repulsion).
- Easily and quickly simulated.

Dept. of Mathematical Sciences
Aalborg University
Denmark

Conclusion

DPP’s possess **appealing properties**:

- They provide flexible parametric models of repulsive point processes (“soft-core” cases and cases with more repulsion).
- Easily and quickly simulated.
- Closed form expressions for all orders of moments.

Conclusion

DPP's possess **appealing properties**:

- They provide flexible parametric models of repulsive point processes ("soft-core" cases and cases with more repulsion).
- Easily and quickly simulated.
- Closed form expressions for all orders of moments.
- Closed form expression for the density of a DPP on any compact set.

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Conclusion

DPP's possess **appealing properties**:

- They provide flexible parametric models of repulsive point processes ("soft-core" cases and cases with more repulsion).
- Easily and quickly simulated.
- Closed form expressions for all orders of moments.
- Closed form expression for the density of a DPP on any compact set.
- Inference is feasible, including likelihood inference. Freely available software!

Dept. of Mathematical Sciences
Aalborg University
Denmark

Conclusion

DPP's possess **appealing properties**:

- They provide flexible parametric models of repulsive point processes ("soft-core" cases and cases with more repulsion).
- Easily and quickly simulated.
- Closed form expressions for all orders of moments.
- Closed form expression for the density of a DPP on any compact set.
- Inference is feasible, including likelihood inference. Freely available software!
- Can quantify repulsiveness.

Jesper Møller

Dept. of Mathematical Sciences
Aalborg University
Denmark

Conclusion

DPP's possess **appealing properties**:

- They provide flexible parametric models of repulsive point processes ("soft-core" cases and cases with more repulsion).
- Easily and quickly simulated.
- Closed form expressions for all orders of moments.
- Closed form expression for the density of a DPP on any compact set.
- Inference is feasible, including likelihood inference. Freely available software!
- Can quantify repulsiveness.

$\Rightarrow$ **Promising alternative to Gibbs point processes**.

Dept. of Mathematical Sciences
Aalborg University
Denmark

Joint work on DPPs

F. Lavancier, J. Møller and E. Rubak (2015). Determinantal point process models and statistical inference. *Journal of Royal Statistical Society: Series B (Statistical Methodology)*, 77, 853-877.

F. Lavancier, J. Møller and E. Rubak (2014). Determinantal point process models and statistical inference: Extended version (61 pages). Available at arXiv:1205.4818.

Joint work on DPPs

F. Lavancier, J. Møller and E. Rubak (2015). Determinantal point process models and statistical inference. *Journal of Royal Statistical Society: Series B (Statistical Methodology)*, 77, 853-877.

F. Lavancier, J. Møller and E. Rubak (2014). Determinantal point process models and statistical inference: Extended version (61 pages). Available at arXiv:1205.4818.

J. Møller, M. Nielsen, E. Porcu and E. Rubak (2018). Determinantal point processes on $\mathbb{S}^d$. *Bernoulli*, 24, 1171-1201.

J. Møller and E. Rubak (2016). Functional summary statistics on the sphere with an application to determinantal point processes. *Spatial Statistics*, 18, 4-23.

Dept. of Mathematical Sciences
Aalborg University
Denmark

Joint work on DPPs

F. Lavancier, J. Møller and E. Rubak (2015). Determinantal point process models and statistical inference. *Journal of Royal Statistical Society: Series B (Statistical Methodology)*, 77, 853-877.

F. Lavancier, J. Møller and E. Rubak (2014). Determinantal point process models and statistical inference: Extended version (61 pages). Available at arXiv:1205.4818.

J. Møller, M. Nielsen, E. Porcu and E. Rubak (2018). Determinantal point processes on $\mathbb{S}^d$. *Bernoulli*, 24, 1171-1201.

J. Møller and E. Rubak (2016). Functional summary statistics on the sphere with an application to determinantal point processes. *Spatial Statistics*, 18, 4-23.

J. Møller and E. O'Reilly (2018). Couplings for determinantal point processes and their reduced Palm distributions with a view to quantifying repulsiveness. Available at arXiv:1806.07347.

Joint work on permanental point processes

P. McCullagh & J. Møller (2006). The permanental process. *Advances in Applied Probability*, 38, 873-888.

J. Møller & Ege Rubak (2010). A model for positively correlated count variables. *International Statistical Review*, 78, 65-80.

E. Rubak, J. Møller and P. McCullagh (2010). Statistical inference for a class of multivariate negative binomial distributions. Research Report R-2010-10, Department of Mathematical Sciences, Aalborg University. Submitted for journal publication.

Thank you for your attention!

AALBORG UNIVERSITY
DENMARK