Lecture 4: Models for point processes

Jean-François Coeurjolly

http://www-ljk.imag.fr/membres/Jean-Francois.Coeurjolly/

GDR GeoSto

LABORATOIRE JEAN KUNTZMANN MATHÉMATIQUES APPLIQUÉES - INFORMATIQUE

UNIVERSITÉ Grenoble Alpes

Why modelling ? Classification

Cox point processes

Gibbs point processes

Why modelling ?

The main objectives of this section are

- to present more realistic models than the too simple Poisson point process to take into account the spatial dependence between points.
- to present statistical methodologies to infer these models.

Why modelling ?

The main objectives of this section are

- to present more realistic models than the too simple Poisson point process to take into account the spatial dependence between points.
- to present statistical methodologies to infer these models.

We can distinguish several classes of models for spatial point processes

1. point processes based on the thinning of a Poisson point processes, on the superimposition of Poisson point processes. [sometimes hard to relate the stochastic process producing the realization and the physical phenomenon producing the data]
2. Cox point processes (which include Cluster point processes,. . .).
3. Gibbs point processes.
4. Determinantal point processes.

An attempt to classify these models . . .

Model	Allows to model	Are moments expressible in a closed form ?	Density w.r.t. Poisson ?
Cox	**attraction**	**yes**	**no**
Gibbs	**repulsion** but also attraction	**no**	**yes**
Determinantal	**repulsion**	**yes**	**yes**

An attempt to classify these models ...

Model	Allows to model	Are moments expressible in a closed form ?	Density w.r.t. Poisson ?
Cox	**attraction**	**yes**	**no**
Gibbs	**repulsion** but also attraction	**no**	**yes**
Determinantal	**repulsion**	**yes**	**yes**

This lecture only focuses on the first two classes of point processes, that is on Cox and Gibbs point processes.

Definition of Cox point processes

We let $S \subseteq \mathbb{R}^d$ throughout this section. B denotes any bounded domain $\subseteq S$.

Definition

Suppose that $Z = \{Z(u) : u \in S\}$ is a nonnegative random field so that with probability one, $u \to Z(u)$ is a locally integrable function. If the conditional distribution of $\mathbf{X}$ given Z is a Poisson process on S with intensity function Z, then $\mathbf{X}$ is said to be a *Cox process* driven by Z.

Remarks :

- Z is a random field means that $Z(u)$ is a random variable $\forall u \in S$.
- if $\mathrm{E}Z(u)$ exists and is locally integrable then w.p. 1, $Z(u)$ is a locally integrable function.

Basic properties of Cox point processes

Proposition

1. Provided $Z(u)$ has finite expectation and variance for any $u \in S$

$$\rho(u) = \mathrm{E} Z(u),\ \rho^{(2)}(u,v) = \mathrm{E}[Z(u)Z(v)],\ g(u,v) = \frac{\mathrm{E}[Z(u)Z(v)]}{\rho(u)\rho(v)}.$$

2. The void probabilities are given by

$$v(B) = \mathrm{E}\exp\left(-\int_B Z(u)\mathrm{d}u\right)$$

for bounded $B \subseteq S$.

Proof : direct consequence of the fact that $\mathbf{X}|Z$ is a Poisson point process with intensity function Z.

Over-dispersion of Cox processes

Proposition

Let A, B bounded sets of S, then

$$\mathrm{Cov}(N(A), N(B)) = \int_A \int_B \mathrm{Cov}(Z(u), Z(v))\mathrm{d}u\mathrm{d}v + \int_{A\cap B} \mathrm{E}Z(u)\mathrm{d}u$$

Over-dispersion of Cox processes

Proposition

Let A, B bounded sets of S, then

$$\mathrm{Cov}(N(A), N(B)) = \int_A \int_B \mathrm{Cov}(Z(u), Z(v))\mathrm{d}u\mathrm{d}v + \int_{A\cap B} \mathrm{E}Z(u)\mathrm{d}u$$

Consequence :

- In particular, $\mathrm{Var}N(A) \geq \mathrm{E}N(A)$ with equality only when $\mathbf{X}$ is a Poisson process.
- $\Rightarrow$ over-dispersion of the counting variables.

Over-dispersion of Cox processes

Proposition

Let A, B bounded sets of S, then

$$\mathrm{Cov}(N(A), N(B)) = \int_A \int_B \mathrm{Cov}(Z(u), Z(v)) \mathrm{d}u\mathrm{d}v + \int_{A\cap B} \mathrm{E}Z(u)\mathrm{d}u$$

Consequence :

- In particular, $\mathrm{Var}N(A) \geq \mathrm{E}N(A)$ with equality only when $\mathbf{X}$ is a Poisson process.
- $\Rightarrow$ over-dispersion of the counting variables.

Other remarks :

- Most of models have pcf such that $g \geq 1$ (but a few exceptions $\exists$).
- If $S = \mathbb{R}^d$ and $\mathbf{X}$ is stationary and/or isotropic then $\mathbf{X}$ is stationary and/or isotropic.
- Explicit expressions of the F, G and J functions in the stationary case are in general difficult to derive.

A first example

Definition

A *mixed Poisson process* is a Cox process where $Z(u) = Z_0$ is given by a positive random variable for any $u \in S$, i.e. $\mathbf{X}|Z_0$ follows a homogeneous Poisson process with intensity Z_0.

This has a limited interest …but the following properties are really easy to derive

1. $\mathbf{X}$ is stationary and (provided Z_0 has first two moments)

$$\rho = \mathrm{E}Z_0 \quad \text{and} \quad g(u,v) = \frac{\mathrm{E}[Z_0^2]}{\mathrm{E}[Z_0]^2} \geq 1.$$

2. The K and L functions are given by

$$K(r) = \beta\omega_d r^d \quad \text{and} \quad L(r) = \beta^{1/d} r \geq r$$

where $\omega_d = |B(0,1)|$ and $\beta = \frac{\mathrm{E}[Z_0^2]}{\mathrm{E}[Z_0]^2}$.

Neymann-Scott processes

Definition

Let C be a stationary Poisson process on $\mathbb{R}^d$ with intensity $\kappa > 0$. Conditional on C, let $\mathbf{X}_c, c \in C$ be independent Poisson processes on $\mathbb{R}^d$ where $\mathbf{X}_c$ has intensity function

$$\rho_c(u) = \alpha k(u - c)$$

where $\alpha > 0$ is a parameter and k is a kernel (i.e. for all $c \in \mathbb{R}^d$, $u \to k(u - c)$ is a density function). Then $\mathbf{X} = \cup_{c \in C} \mathbf{X}_c$ is a Neymann-Scott process with cluster centres C and clusters $\mathbf{X}_c, c \in C$.

Neymann-Scott processes

Definition

Let C be a stationary Poisson process on $\mathbb{R}^d$ with intensity $\kappa > 0$. Conditional on C, let $\mathbf{X}_c, c \in C$ be independent Poisson processes on $\mathbb{R}^d$ where $\mathbf{X}_c$ has intensity function

$$\rho_c(u) = \alpha k(u - c)$$

where $\alpha > 0$ is a parameter and k is a kernel (i.e. for all $c \in \mathbb{R}^d$, $u \to k(u - c)$ is a density function). Then $\mathbf{X} = \cup_{c \in C} \mathbf{X}_c$ is a Neymann-Scott process with cluster centres C and clusters $\mathbf{X}_c, c \in C$.

- $\mathbf{X}$ is also a Cox process on $\mathbb{R}^d$ driven by $Z(u) = \sum_{c \in C} \alpha k(u - c)$.
- Simulating a Neymann-Scott process (on W) is very simple (if k has compact support $T < \infty$)
  1. Generate $C \sim \text{Poisson}(W \oplus T, \kappa)$.
  2. For each $c \in C$, generate $\mathbf{X}_c \sim \text{Poisson}(W, \rho_c)$.
  3. Concatenate all the $\mathbf{X}_c$'s.
- If k has unbounded support, an exact simulation is still possible.

Two classical Neymann-Scott processes

We obtain specific models by choosing specific kernel densities.

1. the *Matérn cluster process* where

$$k(u) = \mathbf{1}(\|u\| \leq R)\frac{1}{\omega_d R^d}$$

   is the uniform density on the $B(0,R)$.

2. the *Thomas process* where

$$k(u) = \left(\frac{1}{2\pi\sigma^2}\right)^{d/2} \exp\left(-\frac{\|u\|^2}{2\sigma^2}\right)$$

   is the density of $\mathcal{N}(0, \sigma^2 I_d)$.

When R is small or when σ is small, then point pattern exhibit strong attraction.

Basic properties of Neymann-Scott processes

- κ is the mean number of cluster centres per unit square, α is the mean number of daughters points per cluster.
- $\mathbf{X}$ is stationary (since Z is stationary) and is isotropic if $k(u) = k(\|u\|)$.
- Intensity of $\mathbf{X}$: $\rho(u) = \alpha\kappa$.
- The (stationary) pair correlation function is given by

$$g(u, v) = 1 + \frac{k * k(v-u)}{\kappa} \geq 1 \quad \text{where} \quad k*k(u) = \int k(c)k(u-c)\mathrm{d}c.$$

- The F, G and J functions are also expressible in terms of k. In particular

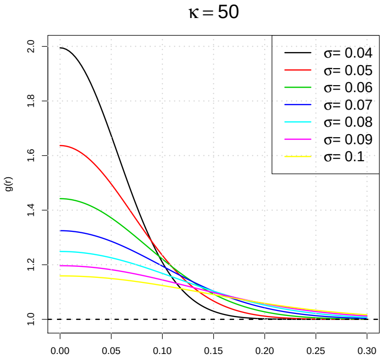
$$J(r) = \int k(u) \exp\left(-\alpha \int_{\|v\|\leq r} k(u+v)\mathrm{d}v\right)\mathrm{d}u$$

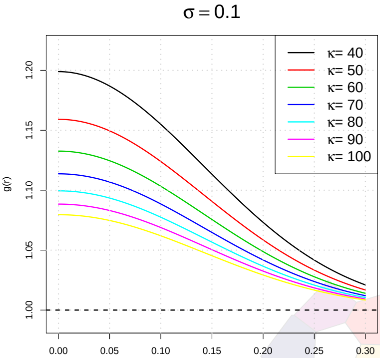
whereby we deduce that $\exp(-\alpha) \leq J(r) \leq 1$.

Back to the Thomas process

Recall that k is the density of a $\mathcal{N}(0, \sigma^2 I_d)$. Applying the previous results, we get (for the pcf)

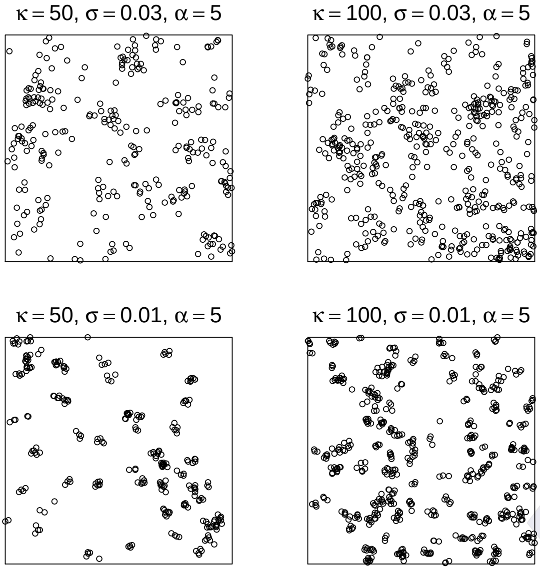
$$g(r) = 1 + \frac{1}{(4\pi\sigma^2)^{d/2}} \exp\left(-r^2/(4\sigma^2)\right)/\kappa$$



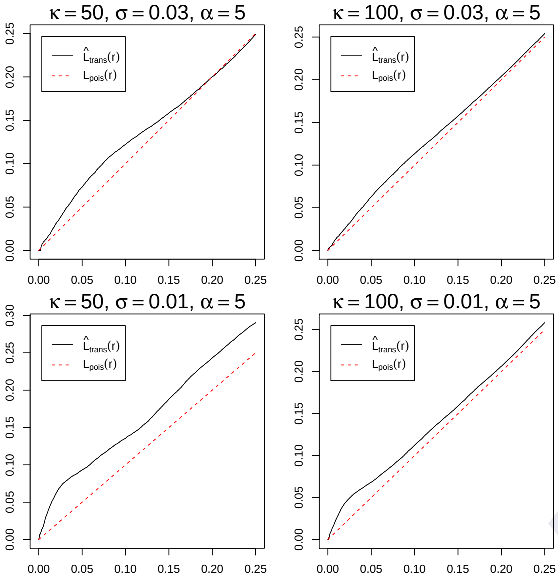


(similar developments can be done for the K, L, J functions and with more work for the Matérn process).

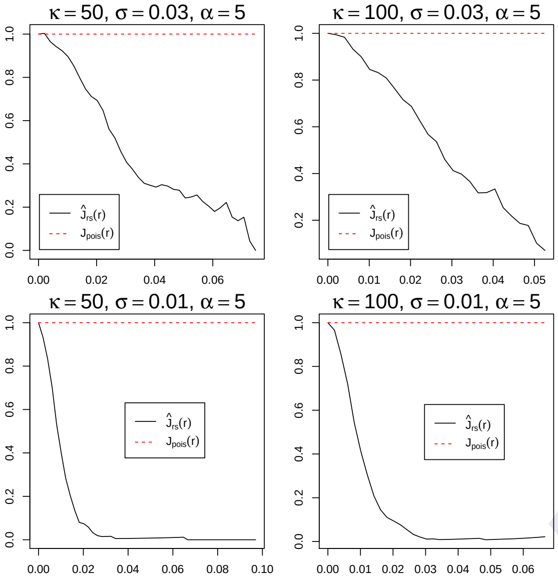
Four realizations of Thomas point processes



Correponding L estimates



Correponding J estimates



Complements

- Inhomogeneous Neymann-Scott processes can be obtained by replacing the intensity parameter κ by a spatial function $\kappa(u)$.
- The natural extension of NS processes is given by shot-noise Cox processes which is a Cox process driven by

$$Z(u) = \sum_{(c,\gamma)\in\Phi} \gamma k(c,u)$$

where $k(\cdot,\cdot)$ is a kernel and Φ is a Poisson point process on $\mathbb{R}^d \times (0,\infty)$ with a locally integrable intensity function ζ. (see e.g. Møller and Waagepetersen 2004 for complements).

Log-Gaussian Cox processes

Definition

Let $\mathbf{X}$ be a Cox process on $\mathbb{R}^d$ driven by $Z = \exp Y$ where Y is a Gaussian random field. Then, $\mathbf{X}$ is said to be a *log Gaussian Cox process* (LGCP).

Log-Gaussian Cox processes

Definition

Let $\mathbf{X}$ be a Cox process on $\mathbb{R}^d$ driven by $Z = \exp Y$ where Y is a Gaussian random field. Then, $\mathbf{X}$ is said to be a *log Gaussian Cox process* (LGCP).

Remarks :

- we could consider $Z = h(Y)$ for some non-negative function h, but the exp leads to tractable calculations.
- another possibility : using a χ^2 field, i.e. $Z(u) = Y_1(u)^2 + \ldots + Y_m(u)^2$ are the Y_i's are independent Gaussian fields with zero mean.
- LGCP are easy to simulate since the problem is transfered to generate a Gaussian field (which can be handled by several methods).
- The mean and covariance function of Y determine the distribution of $\mathbf{X}$.

Particular cases

- In the following we let

$$m(u) = \mathrm{E}\, Y(u) \quad \text{and } c(u,v) = \mathrm{Cov}(Y(u), Y(v))$$

  and we focus on the case where $c(u,v)$ depends only on $\|v-u\|$ (covariance function invariant by translation and by rotation).

- Conditions on c are needed to get a covariance function. Among functions satisfying these properties we find :
  - the *power exponential family* satisfies these conditions

$$c(u,v) = \sigma^2 r(\|v-u\|/\alpha) \text{ with } r(t) = \exp\left(-t^\delta\right), t \geq 0$$

    with $\alpha, \sigma > 0$. $\delta = 1$ is the exponential correlation function ; $\delta = 1/2$ is the stable correlation function ; $\delta = 2$ is the Gaussian correlation function.
  - the *cardinal sine correlation* :

$$c(u,v) = \sigma^2 r(\|v-u\|/\alpha) \text{ with } r(t) = \frac{\sin(t)}{t}, t \geq 0$$

Summary statistics for the LGCP

Proposition

Let $\mathbf{X}$ be a LGCP then under the previous notation

1. the intension function of $\mathbf{X}$ is

$$\rho(u) = \exp\left(m(u) + c(u,u)/2\right).$$

2. The pair correlation function g of $\mathbf{X}$ is

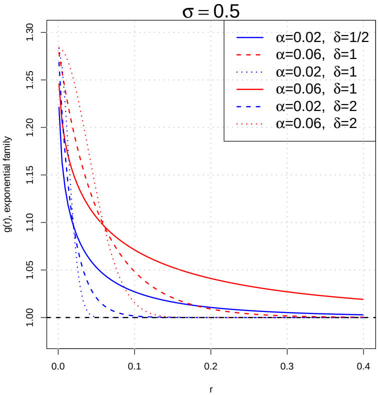
$$g(u,v) = \exp(c(u,v)).$$

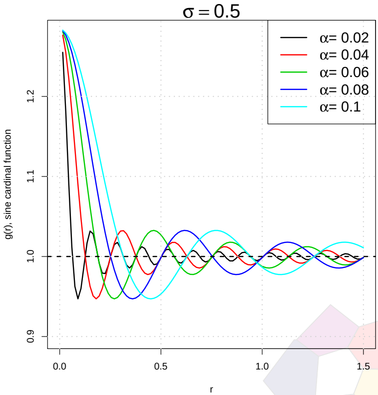
<u>Proof</u> : based on the fact that for $U \sim \mathcal{N}(\zeta, \sigma^2)$, the Laplace transform of U is $\mathrm{E}\exp(tU) = \exp(\zeta + \sigma^2 t/2)$.

- one to one correspondendce between (m, c) and (ρ, g).
- If c is translation invariant then $\mathbf{X}$ is second order reweighted stationary (stationary if m is constant, and isotropic if in addition $c(u,v)$ depends only on $\|v-u\|$).

A few plots of pair correlation function

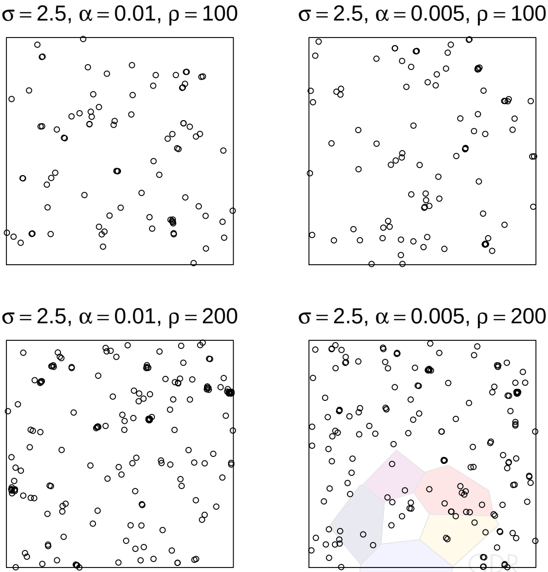
- pcf for the *power exponential family* : $\log g(r) = \sigma^2 \exp\left(-\left(\frac{r}{\alpha}\right)^\delta\right), \quad \alpha, \sigma, \delta > 0$
- pcf for the *cardinal sine correlation* : $\log g(r) = \sigma^2 \frac{\sin(r/\alpha)}{r/\alpha}, \quad \alpha, \sigma > 0$



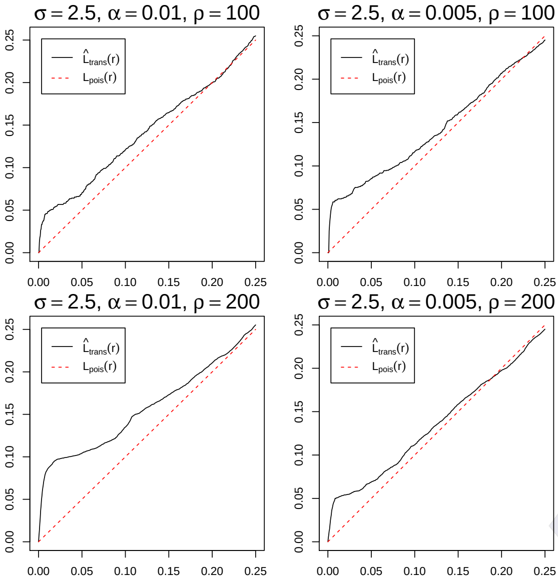


Four realizations of (stationary) LGCP point processes

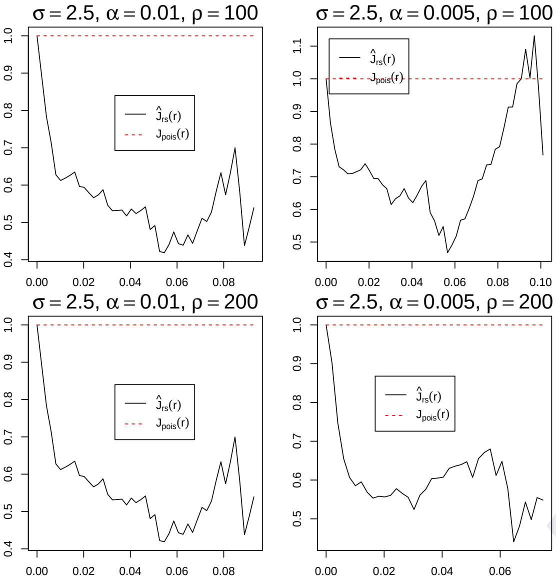
- with exponential correlation function ($\delta = 1$).
- The mean m of the Gaussian process is such that $\rho = \exp(m + \sigma^2/2)$.



Correponding L estimates



Correponding J estimates



Is likelihood available ?

- Assume (only here) that S is a bounded domain, then the density of $\mathbf{X}_S$ w.r.t a Poisson processes with unit rate is given by

$$f(\mathbf{x}) = \mathrm{E}\left[\exp\left(|S| - \int_S Z(u)\mathrm{d}u\right)\prod_{u\in\mathbf{x}} Z(u)\right]$$

for finite point configurations $\mathbf{x} \subset S$. Explicit expression of the expectation is usually unknown and the integral may be difficult to calculate.
$\Rightarrow$ MLE is usually impossible to calculate (approximations or Bayesian should be used)

Is likelihood available ?

- Assume (only here) that S is a bounded domain, then the density of $\mathbf{X}_S$ w.r.t a Poisson processes with unit rate is given by

$$f(\mathbf{x}) = \mathrm{E}\left[\exp\left(|S| - \int_S Z(u)\mathrm{d}u\right)\prod_{u\in\mathbf{x}} Z(u)\right]$$

for finite point configurations $\mathbf{x} \subset S$. Explicit expression of the expectation is usually unknown and the integral may be difficult to calculate.
$\Rightarrow$ MLE is usually impossible to calculate (approximations or Bayesian should be used)

- In most of applications, we only observe the realization of $\mathbf{X}$.
$\Rightarrow$ Z should be considered as a latent process generating the point process, which is not observed.

Inference by minimum contrast estimation

- Assume we observe the realization of a stationary Cox point process which belongs to a parametric family with parameter θ (ex : $\theta = (\alpha, \kappa, \sigma^2)$ for the Thomas process, $\theta = (\mu, \alpha, \sigma^2)$ for a LGCP with exponential correlation function).
- For most of Cox point processes, $\rho = \rho_\theta$, $K = K_\theta$ or $g = g_\theta$ functions are expressible in a closed form, for instance :
  - for a planar ($d = 2$) **Thomas process** (NS process with Gaussian kernel) : $\rho = \alpha\kappa$ and

$$g_\theta(r) = 1 + \frac{1}{\sqrt{4\pi\sigma^2}} \exp\left(-r^2/(4\sigma^2)\right)/\kappa \quad \text{and} \quad K_\theta(r) = \pi r^2 + \left(1 - \exp\left(-r^2/(4\sigma^2)\right)\right)/\kappa$$

  - for a **LGCP with exponential correlation function**

$$\rho = \exp(m + \sigma^2/2) \quad \text{and} \quad \log g_\theta(r) = \sigma^2 \exp(-r/alpha).$$

Inference by minimum contrast estimation (2)

- Then the idea is to estimate θ using a **minimum contrast approach** : i.e. define $\hat{\theta}$ as the minimizer of

$$\int_{r_1}^{r_2} \left| \widehat{K}(r)^q - K_\theta(r)^q \right|^2 \mathrm{d}r \quad \text{or} \quad \int_{r_1}^{r_2} \left| \widehat{g}(r)^q - g_\theta(r)^q \right|^2 \mathrm{d}r$$

  where
  - $\widehat{K}(r)$ and $\widehat{g}(r)$ are the nonparametric estimates of $K(r)$ and $g(r)$.
  - where $[r_1, r_2]$ is a set of r fixed values.
  - q is a power parameter (adviced in the literature to be set to $q = 1/4$ or $1/2$).
- If $\mathbf{X}$ is not stationary but second-order reweighted stationary, the estimation can be done into two steps
  - estimate the intensity parametrically (e.g. Poisson likelihood estimator for instance).
  - use the inhomogeneous K or g function in the minimum contrast estimation method.

A short simulation

- we generated 200 replications of a Thomas process on W with parameters $\kappa = 100$, $\sigma^2 = 10^{-4}$ and $\alpha = 5$
- we estimated the parameters σ^2 and κ using the minimimum contrast estimat based on the K function.
- Then α is estimated using $\widehat{\alpha} = \widehat{\rho}/\widehat{\kappa}$

	Parameter κ	
	$W = [0,1]^2$	$W = [0,2]^2$
Emp. mean	98.9	102.4
Emp. var.	251.9	78.1

	Parameter α	
	$W = [0,1]^2$	$W = [0,2]^2$
Emp. mean	4.9	4.9
Emp. var.	40.1	6.1

	Parameter σ^2	
	$W = [0,1]^2$	$W = [0,2]^2$
Emp. mean	1.01×10^{-4}	9.7×10^{-5}
Emp. var.	1.5×10^{-5}	8.2×10^{-6}

Exercise 7 : `R` and Cox pp - Simulation study in `R`

- Reproduce the previous simlulation study with other parameters. Hint : use the `R` functions `intensity, rThomas, kppm`.
- Same question with a LGCP with exponential covariance function. Hint : use `rLGCP` to generate Log-Gaussian Cox processes.

Exercise 8 : `R` and Cox pp - Dataset `finpines`

1. Plot the K or L function, and construct envelopes by simulation and justify the next question. After justifying it, model the data by a stationary Thomas process and estimate its parameters.
2. Simulate m realizations (m as large as possible) of Thomas process with estimated parameters. Construct 95% confidence bands based on the m estimated K functions, plot the original one. Is the Thomas process well-suited ?

Exercise 9 : `R` and Cox pp - Dataset `bei`

Same exercise for the `beilschmedia` dataset, except that the intensity is modeled as an inhomogeneous function of elevation and gradient of elevation.

Introduction to Gibbs point processes

- the objective of this section is to introduce a new class of point processes : the class of Gibbs point processes.
- **Gibbs point process** :
  - are mainly used to model **repulsion** between point (but a few models allows also to produce **aggregated models**). That's why this kind of models are widely used in statistical physics to model particles systems.
  - are defined (in a bounded domain) by a **density** w.r.t. a Poisson point process
    $\Rightarrow$ **very easy** to interpret the model and the parameters.
  - their main drawback : **moments are not expressible** in a closed form and density known up to a scalar
    $\Rightarrow$ **specific inference methods** are required.

Important restriction of this section

- Throughout this chapter : we assume that the point process $\mathbf{X}$ is defined in a bounded domain $S \subset \mathbb{R}^d$ ($|S| < \infty$).
- Gibbs point processes defined on $\mathbb{R}^d$ are of particular interest :
  - in statistical physics because they can model **phase transition** .
  - in asymptotic statistics : if for instance we want to prove the convergence of an estimator as the window expands to $\mathbb{R}^d$

  However, the formalism is more complicated and technical and this is not considered here.

$\Rightarrow$ from now, $\mathbf{X}$ is a **finite point process in** S **(bounded)** taking values in N_f (space of finite configurations of points)

$$N_f = \{\mathbf{x} \subset S : n(\mathbf{x}) < \infty\}.$$

Most of the results presented here have an extension to $S = \mathbb{R}^d$*.*

Definition of Gibbs point processes

Definition

A finite point process $\mathbf{X}$ on a bounded domain S $(0 < |S| < \infty)$ is said to be a Gibbs point process if it admits a density f w.r.t. a Poisson point process with unit rate, i.e. for any $F \subseteq N_f$

$$P(\mathbf{X} \in F) = \sum_{n \geq 0} \frac{\exp(-|S|)}{n!} \times \int_S \dots \int_S \mathbf{1}(\{x_1, \dots, x_n\} \in F) f(\{x_1, \dots, x_n\}) \mathrm{d}x_1 \dots \mathrm{d}x_n$$

where the term $n = 0$ is read as $\exp(-|S|)\mathbf{1}(\emptyset \in F) f(\emptyset)$.

- Gpp can be viewed as a perturbation of a Poisson point process.
- f is easily interpretable since it is in some sense a weight w.r.t. a Poisson process.

The simplest example ...

is the inhomogeneous Poisson point process. Indeed for $\mathbf{X} \sim$ Poisson(S,ρ) (such that $\mu(S) < \infty$), we recall that $\mathbf{X}$ admits a density w.r.t. to a Poisson point process with unit rate given for any $\mathbf{x} \in N_f$ by

$$f(\mathbf{x}) = \exp(|S| - \mu(S)) \prod_{u \in \mathbf{x}} \rho(u).$$

The simplest example ...

is the inhomogeneous Poisson point process. Indeed for $\mathbf{X} \sim \text{Poisson}(S, \rho)$ (such that $\mu(S) < \infty$), we recall that $\mathbf{X}$ admits a density w.r.t. to a Poisson point process with unit rate given for any $\mathbf{x} \in N_f$ by

$$f(\mathbf{x}) = \exp(|S| - \mu(S)) \prod_{u \in \mathbf{x}} \rho(u).$$

In most of cases, f is specified up to a proportionality $f = c^{-1} h$ where $h : N_f \to \mathbb{R}^+$ is a known function.
$\Rightarrow$ c is given by

$$c = \sum_{n \geq 0} \frac{\exp(-|S|)}{n!} \int_S \dots \int_S h(\{x_1, \dots, x_n\}) \mathrm{d}x_1 \dots \mathrm{d}x_n = \mathrm{E}[h(Y)]$$

where $Y \sim \text{Poisson}(S, 1)$.

Papangelou conditional intensity

Definition

The Papangelou conditional intensity for a point process $\mathbf{X}$ with density f is defined by

$$\lambda(u, \mathbf{x}) = \frac{f(\mathbf{x} \cup u)}{f(\mathbf{x})}$$

for any $\mathbf{x} \in N_f$ and $u \in S$ $(u \notin \mathbf{x})$, taking $a/0 = 0$ for $a \geq 0$.

- λ does not depend on c.
- for Poisson(S, ρ), $\lambda(u, \mathbf{x}) = \rho(u)$ does not depend on $\mathbf{x}$!
- $\lambda(u, \mathbf{x})\mathrm{d}u$ can be interpreted as the conditional probability of observing a point in an infinitesimal region containing u of size $\mathrm{d}u$ given the rest of $\mathbf{X}$ is $\mathbf{x}$.

Attraction, repulsion, heredity

Definition

We often say that $\mathbf{X}$ (or f) is

- **attractive** if

$$\lambda(u,\mathbf{x}) \le \lambda(u,\mathbf{y}) \text{ whenever } \mathbf{x} \subset \mathbf{y}.$$

- **repulsive** if

$$\lambda(u,\mathbf{x}) \ge \lambda(u,\mathbf{y}) \text{ whenever } \mathbf{x} \subset \mathbf{y}.$$

- **hereditary** if

$$f(\mathbf{x}) > 0 \Rightarrow f(\mathbf{y}) > 0 \text{ for any } \mathbf{y} \subset \mathbf{x}.$$

- if f is hereditary, then $f \Leftrightarrow \lambda$ (one-to-one correspondence).

Existence of a Gpp in S ($|S| < \infty$)

Proposition

Let $\phi^\star : S \to \mathbb{R}^+$ be a function so that $c^\star = \int_S \phi^\star(u)\mathrm{d}u < \infty$. Let $h = cf$, we say that $\mathbf{X}$ (or f) satisfies the

- local stability property if for any $\mathbf{x} \in N_f$, $u \in S$

$$h(\mathbf{x} \cup u) \leq \phi^\star(u)h(\mathbf{x}) \Leftrightarrow \lambda(u, \mathbf{x}) \leq \phi^\star(u).$$

- the Ruelle stability property if for any $\mathbf{x} \in N_f$ and for $\alpha > 0$

$$h(\mathbf{x}) \leq \alpha \prod_{u \in \mathbf{x}} \phi^\star(u).$$

local stability condition $\Rightarrow$ Ruelle stability condition (and that f is hereditary) $\Rightarrow$ existence of point process in S.

Proof : the first implication is obvious ; for the last one it consists in checking that $c < \infty$.

Pairwise interaction point processes

For simplicity, we focus on the isotropic case.

Definition

A istotropic parwise interaction point process (PIPP) has a density of the form (for any $\mathbf{x} \in N_f$)

$$f(\mathbf{x}) \propto \prod_{u \in \mathbf{x}} \phi(u) \prod_{\{u,v\} \subseteq \mathbf{x}} \phi_2(\|v-u\|)$$

where $\phi : S \to \mathbb{R}^+$ and $\phi_2 : \mathbb{R}^+_* \to \mathbb{R}+$.

- If ϕ is constant (equal to β) then the Gpp is said to be homogeneous (note that $\prod_{u \in \mathbf{x}} \phi(u) = \beta^{n(\mathbf{x})}$).
- ϕ_2 is called the interaction function.
- this class of models is hereditary
- f is repulsive if $\phi_2 \leq 1$, in which case the process is locally stable if $\int_S \phi(u) \mathrm{d}u$.

Strauss point process

Among the class of PIPP, the main example is the Strauss point process defined by

$$f(\mathbf{x}) \propto \beta^{n(\mathbf{x})}\gamma^{s_R(\mathbf{x})} \qquad \lambda(u,\mathbf{x}) = \beta\gamma^{t_R(u,\mathbf{x})}$$

where $\beta > 0, R < \infty$, where $s_R(\mathbf{x})$ is the number of R-close pairs of points in $\mathbf{x}$ and $t_R(u,\mathbf{x}) = s_R(\mathbf{x} \cup u) - s_R(\mathbf{x})$ is the number of R-close neighbours of u in $\mathbf{x}$

$$s_R(\mathbf{x}) = \sum_{\{u,v\}\in\mathbf{x}} \mathbf{1}(\|v-u\| \le R) \text{ and } t_R(u,\mathbf{x}) = \sum_{v\in\mathbf{x}} \mathbf{1}(\|v-u\| \le R).$$

Strauss point process

Among the class of PIPP, the main example is the Strauss point process defined by

$$f(\mathbf{x}) \propto \beta^{n(\mathbf{x})}\gamma^{s_R(\mathbf{x})} \qquad \lambda(u,\mathbf{x}) = \beta\gamma^{t_R(u,\mathbf{x})}$$

where $\beta > 0, R < \infty$, where $s_R(\mathbf{x})$ is the number of R-close pairs of points in $\mathbf{x}$ and $t_R(u,\mathbf{x}) = s_R(\mathbf{x} \cup u) - s_R(\mathbf{x})$ is the number of R-close neighbours of u in $\mathbf{x}$

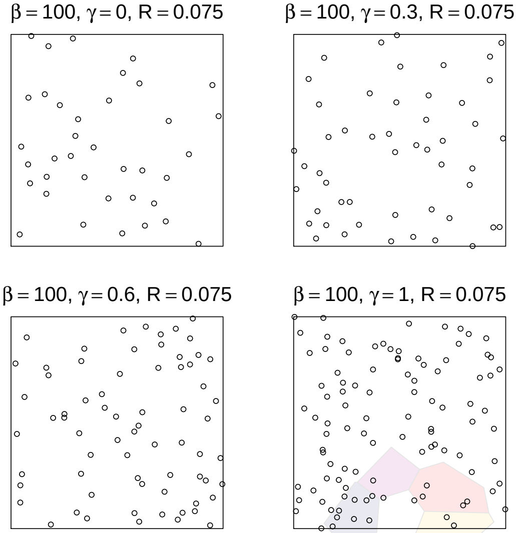
$$s_R(\mathbf{x}) = \sum_{\{u,v\}\in\mathbf{x}} \mathbf{1}(\|v-u\| \le R) \text{ and } t_R(u,\mathbf{x}) = \sum_{v\in\mathbf{x}} \mathbf{1}(\|v-u\| \le R).$$

The parameter γ is called the **interaction parameter** :

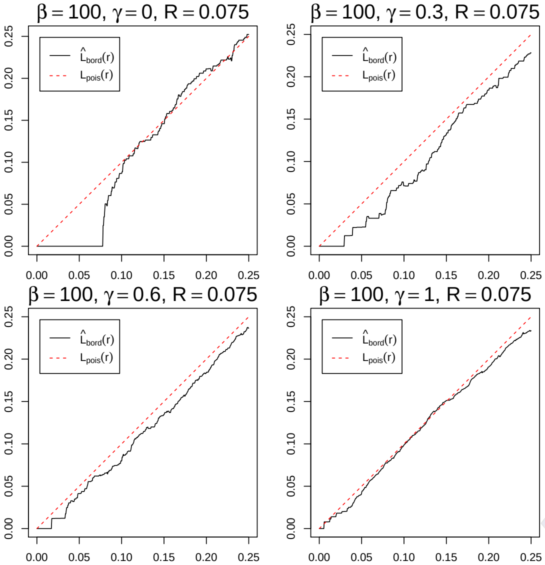
- $\gamma = 1$: homogeneous Poisson point process with intensity β.
- $0 < \gamma < 1$: repulsive point process.
- $\gamma = 0$: hard-core process with hard-core R ; the points are prohibited from being closer han R.
- $\gamma > 1$: the model is not well-defined (if there exists a set $A \subset S$ with $|A| > 0$ and $diam(A) \le R$, then $c > \sum_{n\ge 0} \frac{(\beta|A|)^n}{n!}\gamma^{n(n-1)/2} = \infty$).

Realizations of Strauss point processes

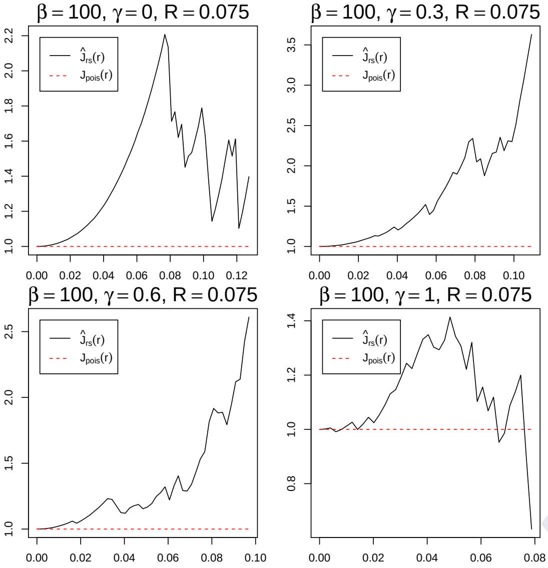
(simulation of spatial Gibbs point processes can be done using spatial birth-and-death process or using MCMC with reversible jumps, see Møller and Waagepetersen for details)



Corresponding L estimates



Corresponding J estimates



Finite range property (spatial Markov property)

Definition

A Gibbs point process $\mathbf{X}$ has a finite range R if the Papangelou conditional intensity satisfies

$$\lambda(u, \mathbf{x}) = \lambda(u, \mathbf{x} \cap B(u, R)).$$

- the probability to insert a point u into $\mathbf{x}$ depends only on some neighborhood of u.
- this definition is actually more general and leads to the definition of Markov point process (omitted here to save time).
- interesting property when we want to deal with edge effects.
- Finite range of the Strauss point process $= R$.

Other pairwise interaction point processes

- **Strauss** point process : $\phi_2(r) = \gamma^{\mathbf{1}(r \le R)}$.
- **Piecewise Strauss** point process :

$$\phi_2(r) = \gamma_1^{\mathbf{1}(r \le R_1)} \gamma_2^{\mathbf{1}(R_1 < r \le R_2)} \dots \gamma_p^{\mathbf{1}(R_{p-1} < r \le R)},$$

with $\gamma_i \in [0,1]$ and $0 \le R_1 < \dots < R_p = R < \infty$ (finite range R) .

- **Overlap area** process :

$$\phi_2(r) = \gamma^{|B(u,R/2) \cap B(v,R/2)|},$$

with $r = \|v - u\|$ with $\gamma \in [0,1]$ (finite range R) .

- **Lennard-Jones** process :

$$\phi_2(r) = \exp(\alpha_1 (\sigma/r)^6 - \alpha_2 (\sigma/r)^{12}),$$

with $\alpha \ge 0$, $\alpha_2 > 0$, $\sigma > 0$ (well-known example used in statistical physics, not locally stable but Ruelle stable) (infinite range) .

Non pairwise interaction point processes

- **Geyer's triplet** point process :

$$f(\mathbf{x}) \propto \beta^{n(\mathbf{x})} \gamma^{s_R(\mathbf{x})} \delta^{u_R(\mathbf{x})}$$

$\beta > 0$, $s_R(\mathbf{x})$ is defined as in the Strauss case and

$$u_R(\mathbf{x}) = \sum_{\{u,v,w\}} \mathbf{1}(\|v-u\| \leq R, \|w-v\| \leq R, \|w-u\| \leq R)$$

  - (i) $\gamma \in [0,1]$ and $\delta \in [0,1]$: locally stable, repulsive, finite range R.
  - (ii) $\gamma > 1$ and $\delta \in (0,1)$: locally stable, neither attractive nor repulsive, finite range R.

Non pairwise interaction point processes (2)

- **Area-interaction** point process :

$$f(\mathbf{x}) \propto \beta^{n(\mathbf{x})} \gamma^{-|U_{\mathbf{x},R}|}$$

where $U_{\mathbf{x},R} = \cup_{u \in \mathbf{x}} B(u, R)$, $\beta > 0$ and $\gamma > 0$. It is attractive for $\gamma \geq 1$ and repulsive for $0 < \gamma \leq 1$. In both cases, it is locally stable since

$$\lambda(u, \mathbf{x}) = \beta \gamma^{-|B(u,R) \setminus \cup_{v \in \mathbf{x} : \|v-u\| \leq 2R} B(v,R)|}$$

satisfies $\lambda(u, \mathbf{x}) \leq \beta$ when $\gamma \geq 1$ and $\lambda(u, \mathbf{x}) \leq \beta \gamma^{-\omega_d R^d}$ in the other case. (finite range $2R$)

GNZ formula

The following result is also a characterization of a Gibbs point process.

Georgii-Nguyen-Zeissin Formula

Let X be a finite and hereditary Gibbs point process defined on S. Then, for any function $h : S \times N_f \to \mathbb{R}^+$, we have

$$\mathrm{E}\Big[\sum_{u\in\mathbf{X}} h(u,\mathbf{X}\setminus u)\Big] = \int_S \mathrm{E}[h(u,\mathbf{X})\lambda(u,\mathbf{X})]\mathrm{d}u.$$

<u>Proof</u> : we know that $\mathrm{E}g(\mathbf{X}) = \mathrm{E}[g(\mathbf{Y})f(\mathbf{Y})]$ where f is the density of a Poisson point process with unit rate Y. Apply this to the function $g(\mathbf{X}) = \sum_{u\in\mathbf{X}} h(u,\mathbf{X}\setminus u)$

$$\begin{aligned}\mathrm{E}[g(\mathbf{X})] &= \mathrm{E}[\sum_{u\in\mathbf{Y}} h(u, Y\setminus u)f(\mathbf{Y})]\\ &= \int_S \mathrm{E}[h(u,\mathbf{Y})f(Y\cup u)]\mathrm{d}u \quad \text{from the Slivnyak-Mecke Theorem}\\ &= \int_S \mathrm{E}[h(u,\mathbf{Y})f(\mathbf{Y})\lambda(u,Y)]\mathrm{d}u \quad \text{since } \mathbf{X} \text{ is hereditary}\\ &= \int_S \mathrm{E}[h(u,\mathbf{X})\lambda(u,\mathbf{X})]\mathrm{d}u.\end{aligned}$$

First and second order intensities

Proposition

1. The intensity function is given by

$$\rho(u) = \mathrm{E}[\lambda(u, \mathbf{X})].$$

2. The second order intensity function is given by

$$\rho^{(2)}(u, v) = \mathrm{E}[\lambda(u, \mathbf{X})\lambda(v, \mathbf{X})]$$

- can be deduced from the GNZ formula.
- Except for the Poissonian case, moments are not expressible in a closed form, e.g.

$$\rho(u) = \frac{1}{c} \sum_{n \geq 0} \frac{\exp(-|S|)}{n!} \int_S \dots \int_S \lambda(u, \{x_1, \dots, x_n\}) h(\{x_1, \dots, x_n\}) \mathrm{d}x_1 \dots \mathrm{d}x_n.$$

- Approximations can be obtained using a Monte-Carlo approach or using a saddle-point approximation (very recent).

Position of the problem

- we observe a realization of $\mathbf{X}$ on $W = S$ ($|S| < \infty$; edge effects occur when $W \subset S$) of a parametric Gibbs point process with density which belongs to a parametric family of densities $(f_\theta = h_\theta / c_\theta)_{\theta \in \Theta}$ for $\Theta \subset \mathbb{R}^p$.
- Problem : estimate the parameter θ based on a single realization.

Position of the problem

- we observe a realization of $\mathbf{X}$ on $W = S$ ($|S| < \infty$; edge effects occur when $W \subset S$) of a parametric Gibbs point process with density which belongs to a parametric family of densities $(f_\theta = h_\theta / c_\theta)_{\theta \in \Theta}$ for $\Theta \subset \mathbb{R}^p$.
- Problem : estimate the parameter θ based on a single realization.
- *MLE approach* : the log-likelihood is $\boxed{\ell_W(\mathbf{x};\theta) = \log h_\theta - \log c_\theta}$.
  Pbm : Given a model h_θ can be computed but c_θ cannot be evaluated even for a single value of θ ; asymptotic properties are only partial.

Position of the problem

- we observe a realization of $\mathbf{X}$ on $W = S$ ($|S| < \infty$; edge effects occur when $W \subset S$) of a parametric Gibbs point process with density which belongs to a parametric family of densities $(f_\theta = h_\theta / c_\theta)_{\theta \in \Theta}$ for $\Theta \subset \mathbb{R}^p$.
- Problem : estimate the parameter θ based on a single realization.
- *MLE approach* : the log-likelihood is $\boxed{\ell_W(\mathbf{x};\theta) = \log h_\theta - \log c_\theta}$.
  Pbm : Given a model h_θ can be computed but c_θ cannot be evaluated even for a single value of θ ; asymptotic properties are only partial.
  $\Rightarrow$ several solutions exist
  1. Approximate c_θ using a Monte-Carlo approach.
  2. Bayesian approach, importance sampling method (to estimate a ratio of normalizing constants).
  3. Combine the MLE with the Ogata-Tanemura approximation.
  4. Find another method which does not involve c_θ.

Pseudo-likelihood

- To avoid the computation of the normalizing constant, the idea is to compute a likelihood based on conditional densities

$$PL_W(\mathbf{x};\theta) = \exp(-|W|) \lim_{i\to\infty} \prod_{j=1}^{m_i} f(\mathbf{x}_{A_{ij}} | \mathbf{x}_{W\setminus A_{ij}};\theta)$$

where $\{A_{ij} : j = 1, \ldots, m_i\}$ $i = 1, 2, \ldots$ are nested subdivisions of W.

- By letting $m_i \to \infty$ and $m_i \max |A_{ij}|^2 \to 0$ as $i \to \infty$ and taking the log, Jensen and Møller (91) obtained

$$LPL_W(\mathbf{x};\theta) = \sum_{u\in\mathbf{x}_W} \lambda(u, \mathbf{x} \setminus u;\theta) - \int_W \lambda(u, \mathbf{x};\theta)\mathrm{d}u$$

Comments on the Pseudo-likelihood

The MPLE is the estimate $\widehat{\theta}$ maximizing

$$LPL_W(\mathbf{x};\theta) = \sum_{u \in \mathbf{x}_W} \log \lambda(u, \mathbf{x} \setminus u; \theta) - \int_W \lambda(u, \mathbf{x}; \theta) \mathrm{d}u$$

1. **Independent on c_θ,** so the LPL is up to an integral discretization and up to edge effects very to compute.

Comments on the Pseudo-likelihood

The MPLE is the estimate $\widehat{\theta}$ maximizing

$$LPL_W(\mathbf{x};\theta) = \sum_{u\in\mathbf{x}_W} \log\lambda(u,\mathbf{x}\setminus u;\theta) - \int_W \lambda(u,\mathbf{x};\theta)\mathrm{d}u$$

1. **Independent on c_θ,** so the LPL is up to an integral discretization and up to edge effects very to compute.
2. If $\mathbf{X}$ has a finite range R, then since $\mathbf{x}$ is observed in W, we can replace W by $W_{\ominus R}$ so that for instance $\lambda(u,\mathbf{x};\theta)$ can always be computed for any $u \in W_{\ominus R}$ **(border correction)**.

Comments on the Pseudo-likelihood

The MPLE is the estimate $\widehat{\theta}$ maximizing

$$LPL_W(\mathbf{x};\theta) = \sum_{u\in\mathbf{x}_W} \log\lambda(u,\mathbf{x}\setminus u;\theta) - \int_W \lambda(u,\mathbf{x};\theta)\mathrm{d}u$$

1. **Independent on c_θ,** so the LPL is up to an integral discretization and up to edge effects very to compute.
2. If $\mathbf{X}$ has a finite range R, then since $\mathbf{x}$ is observed in W, we can replace W by $W_{\ominus R}$ so that for instance $\lambda(u,\mathbf{x};\theta)$ can always be computed for any $u \in W_{\ominus R}$ **(border correction)**.
3. If $\log\lambda(u,\mathbf{x};\theta) = \theta^\top v(u,\mathbf{x})$ (exponential family - class of all examples presented before), then LPL is a **concave** function of θ.

Comments on the Pseudo-likelihood

The MPLE is the estimate $\widehat{\theta}$ maximizing

$$LPL_W(\mathbf{x};\theta) = \sum_{u\in \mathbf{x}_W} \log \lambda(u,\mathbf{x}\setminus u;\theta) - \int_W \lambda(u,\mathbf{x};\theta)\mathrm{d}u$$

1. **Independent on c_θ,** so the LPL is up to an integral discretization and up to edge effects very to compute.
2. If $\mathbf{X}$ has a finite range R, then since $\mathbf{x}$ is observed in W, we can replace W by $W_{\ominus R}$ so that for instance $\lambda(u,\mathbf{x};\theta)$ can always be computed for any $u \in W_{\ominus R}$ **(border correction)**.
3. If $\log \lambda(u,\mathbf{x};\theta) = \theta^\top v(u,\mathbf{x})$ (exponential family - class of all examples presented before), then LPL is a **concave** function of θ.
4. under suitable conditions $\widehat{\theta}$ is a **consistent** estimate and satisfies a **CLT** (and a fast covariance estimate is available) as the window W expands to $\mathbb{R}^d$. [Jensen and Künsch'94, Billiot Coeurjolly and Drouilhet'08-'10, Coeurjolly and Rubak'12].

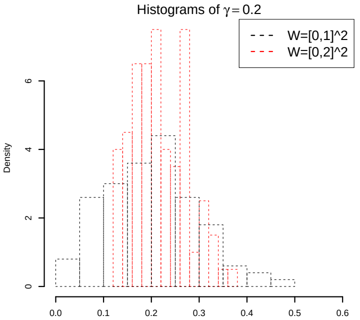
Simulation example

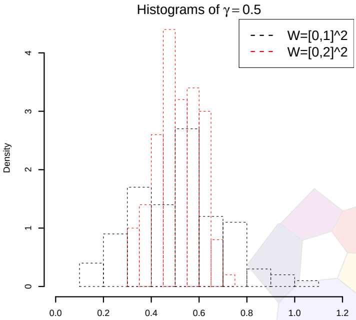
We generated 100 replications of Strauss point processes (a border correction was applied) :

1. mod1 : $\beta = 100,\ \gamma = 0.2,\ R = .05.$
2. mod2 : $\beta = 100,\ \gamma = 0.5,\ R = .05.$

	Estimates of β			
	$W = [0,1]^2$		$W = [0,2]^2$	
mod1	99.52	(17.84)	97.98	(9.24)
mod2	99.28	(20.48)	98.21	(8.53)

	Estimates of γ			
	$W = [0,1]^2$		$W = [0,2]^2$	
mod1	0.20	(0.09)	0.21	(0.06)
mod2	0.52	(0.19)	0.51	(0.09)





Takacs-Fiksel method

- Denote for any function h (eventually depending on θ)

$$L_W(\mathbf{X}, h; \theta) = \sum_{u \in \mathbf{X}_W} h(u, \mathbf{X} \backslash u; \theta) \text{ and } R_W(\mathbf{X}, h; \theta) = \int_W h(u, \mathbf{X}; \theta) \lambda(u, \mathbf{X}; \theta) \mathrm{d}u$$

- The GNZ formula states : $\mathrm{E}[L_W(\mathbf{X}, h; \theta)] = \mathrm{E}[R_W(\mathbf{X}, h; \theta)]$.
- **Idea** : if θ is a p-dimensional vector,
  1. choose p test function h_i and define the contrast

$$U_W(\mathbf{X}, \theta) = \sum_{i=1}^{p} \left(L_W(\mathbf{X}, h; \theta) - R_W(\mathbf{X}, h; \theta) \right)^2 .$$

  2. Define $\widehat{\theta}^{TF} = \mathrm{argmin}_\theta U_W(\mathbf{X}, \theta)$.

Takacs-Fiksel (2)

General comments :

- **like the MPLE** :
  - independent of c_θ, border correction possible in case of $\mathbf{X}$ has a finite range
  - consistent and asymptotically Gaussian estimate (Coeurjolly et al.'12).

Takacs-Fiksel (2)

General comments :

- **like the MPLE** :
  - independent of c_θ, border correction possible in case of $\mathbf{X}$ has a finite range
  - consistent and asymptotically Gaussian estimate (Coeurjolly et al.'12).
- **Another advantage** : interesting choices of test functions cal least to a decreasing of computation time.
  Ex : $h_i(u, \mathbf{X}) = n(B(u, r_i))\lambda^{-1}(u, \mathbf{X}; \theta) \Rightarrow R_W$ independent of θ.

Takacs-Fiksel (2)

General comments :

- **like the MPLE** :
  - independent of c_θ, border correction possible in case of $\mathbf{X}$ has a finite range
  - consistent and asymptotically Gaussian estimate (Coeurjolly et al.'12).
- **Another advantage** : interesting choices of test functions cal least to a decreasing of computation time.
  Ex : $h_i(u,\mathbf{X}) = n(B(u,r_i))\lambda^{-1}(u,\mathbf{X};\theta) \Rightarrow R_W$ independent of θ.
- Actually : **MPLE = TFE** with $h = (h_1,\ldots,h_p)^\top = \lambda^{(1)}(\cdot,\cdot;\theta)$.
  Indeed (assume $\log\lambda(u,\mathbf{X};\theta) = \theta^\top v(u,\mathbf{X})$ (for simplicity)

$$\nabla LPL_W(\mathbf{X};\theta) = \sum_{u\in\mathbf{X}_W} v(u,\mathbf{X}\setminus u) - \int_W v(u,\mathbf{X})\lambda(u,\mathbf{X};\theta)\mathrm{d}u.$$

Exercise 10 (Takacs-Fiksel and Strauss model)

1. We focus on the Strauss point process. Consider the test functions

$$h_1(u,\mathbf{X}) = \mathbf{1}(n(B(u,R) = 0)) \quad \text{and} \quad h_2(u,\mathbf{X}) = \mathbf{1}(n(B(u,R) = 1))$$

   and show that $L_W(\mathbf{X},h_i) = L_i$ $(i=1,2)$, $R_W(\mathbf{X},h_1) = \beta I_1$ and $R_W(\mathbf{X},h_2) = \beta\gamma I_2$ where L_i and I_i are independent of the parameters.

2. Write the contrast function $U_W(\mathbf{X};\theta) = (L_1 - \beta I_1)^2 + (L_2 - \beta\gamma I_2)^2$ and show that the minimization of this contrast leads to an explicit estimator of the parameters β and γ.

3. Extend the methodology for the Gibbs point process with Papangelou conditional intensity

$$\lambda(u,\mathbf{X}) = \beta\gamma_1^{t_{R_1}(u,\mathbf{X})}\gamma_2^{t_{R_2}(u,\mathbf{X})} \text{ with } t_{R_i}(u,\mathbf{X}) = \sum_{v\in\mathbf{X}} \mathbf{1}(R_{i-1} \le \|v-u\| \le R_i)$$

   for $i=1,2$ where $R_0 = 0 < R_1 < R_2$.

Exercise 11 : R and Gibbs pp - Simulation Study

- Generate m Strauss point processes with parameters $\beta = 200$, $\gamma = 0.5$ and $R = 0.05$ on $W = [0,1]^2, [0,2]^2$. Estimate the parameters β and γ (assume R is known) by pseudo-likelihood for each replication. Hint : use the `R` functions `rStrauss, ppm`
- Evaluate the empirical mean and standard deviation and comment the results. Same questions with $\gamma = 0.2$ and $\gamma = 0.8$.

Execercise 12 : R and Gibbs pp - swedishpines dataset

- Plot the K or L function and construct 95% confidence bands under CSR by simulation.
- After justifying it, model the data by a Strauss point process and estimate the parameters (with $R = .5$).
- Generate m replications of the fitted model, construct 95% confidence bands under the fitted model of the K function and plot the K function estimated on the data. Is the model well-suited ?
- Propose a 95% confidence interval for the parameters β, γ.

Complements

Other parametric approaches :

- Variational approach : (Baddeley and Dereudre'12).
- Method based on a logistic regression likelihood (Baddeley, Coeurjolly, Rubak, Waagepetersen'13).

Complements

Other parametric approaches :

- Variational approach : (Baddeley and Dereudre'12).
- Method based on a logistic regression likelihood (Baddeley, Coeurjolly, Rubak, Waagepetersen'13).

Model fitting :

- Monte-Carlo approach : we can compare a summary statistic e.g. L with $L_{\widehat{\theta}}$.
  Pbm : L_θ not expressible in a closed form and must be approximated.
- We can still use the GNZ formula : given a test function h, we can construct

$$L_W(\mathbf{X}, h; \widehat{\theta}) - R_W(\mathbf{X}, h; \widehat{\theta}) =: \text{Residuals}(\mathbf{X}, \text{h}).$$

If the model is correct, then Residuals($\mathbf{X}$, h) should be close to zero. (Baddeley et al.'05,08', Coeurjolly and Lavancier'12).

General Conclusion

The anaysis of **spatial point pattern**

- very large domain of research including probability, mathematical statistics, applied statistics
- own specific models, methodologies and software(s) to deal with.
- is involved in more and more applied fields : economy, biology, physics, hydrology, environmetrics,. . .

General Conclusion

The anaysis of **spatial point pattern**

- very large domain of research including probability, mathematical statistics, applied statistics
- own specific models, methodologies and software(s) to deal with.
- is involved in more and more applied fields : economy, biology, physics, hydrology, environmetrics,. . .

Still a lot of **challenges**

- Modelling : the "true model", problems of existence, phase transition.
- Many classical statistical methodologies need to be adapted (and proved) to s.p.p. : robust methods, resampling techniques, multiple hypothesis testing.
- High-dimensional problems : $S = \mathbb{R}^d$ with d large, selection of variables, regularization methods,. . .
- Space-time point processes.