



**HAL**  
open science

## Spatial statistics and Bayesian modelling

Radu S. Stoica

► **To cite this version:**

| Radu S. Stoica. Spatial statistics and Bayesian modelling. Doctoral. France. 2020. hal-03049366

**HAL Id: hal-03049366**

**<https://hal.science/hal-03049366v1>**

Submitted on 9 Dec 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Spatial statistics and Bayesian modelling

Radu Stoica

Université de Lorraine - IUT Charlemagne

Institut Elie Cartan de Lorraine

`radu-stefan.stoica@univ-lorraine.fr`

Nancy, M2 IMSD - S3

# Table of contents

## Course 1. Introduction and some Mathematical background

Introduction

Some data sets and their related questions

Mathematical background

# Course 1. Introduction : course organisation, data sets and examples

## About me

- ▶ professor at Université de Lorraine
- ▶ mail : `radu-stefan.stoica@univ-lorraine.fr`
- ▶ web page :  
`https://sites.google.com/site/radustefanstoica/`
- ▶ office : 120
- ▶ phone : + 33 6 20 06 29 30

## Course structure

- ▶ 6 blocks, where 1 block = 1h30 course + 2h30 theoretical and practical exercises
- ▶ mark : exam to be defined

## Bibliographical references

- Baddeley, A. J., Rubak, E., Turner, R. (2016). Spatial point patterns. Methodology and applications with R. Chapman and Hall.
- Chiu, S. N., Stoyan, D., Kendall, W. S., Mecke, J. (2013). Stochastic geometry and its applications. John Wiley and Sons.
- Diggle, P. J. (2003). Statistical analysis of spatial point patterns. Arnold Publishers.
- van Lieshout, M. N. M. (2000). Markov point processes and their applications. Imperial College Press.
- van Lieshout, M. N. M. (2019). Theory of spatial statistics. A concise introduction. Chapman and Hall, CRC Press.

- Martinez, V. J., Saar, E. (2002). Statistics of the galaxy distribution. Chapman and Hall.
- Møller, J., Waagepetersen, R. P. (2004). Statistical inference and simulation for spatial point processes. Chapman and Hall.
- Tegmark, M. (2014). Our mathematical Universe. Alfred A. Knopf, New York.

# Acknowledgements

L. Ben Allal, A. J. Baddeley, P. Bermolen, M. Bodea, G. Castellan, G. Caumon, M. Clausel, J.-F. Coeurjolly, T. Danelian, N. Dante, Yu. Davydov, M. Deaconu, X. Descombes, N. Emelyanov, M. Fouchard, E. Gay, P. Gregori, C. Gouache, P. Heinrich, D. Hestroffer, L. Hurtado, R. Kipper, F. Kleinschroth, I. Kovalenko, A. Kretzschmar, M. Kruuse, C. Lantuéjoul, Q. Laporte-Chabasse, A. Lejay, M. N. M. van Liehsout, S. Liu, L. J. Liivamägi, R. Marchand, V. Martinez, J. Mateu, J. Møller, E. Mordecki, F. Mortier, Z. Pawlas, A. Philippe, F. Rihmi, E. Saar, D. Stoyan, E. Tempel, C. V. Tran, G. B. Valsecchi, A. Vienne, R. Waagepetersen, J. Zerubia

and the list is far from being complete ...

# What is this course about ?

**Spatial data analysis** : investigate and describe data sets whose elements have two components

- ▶ position : spatial coordinates of the elements in a certain space
- ▶ characteristic : value(s) of the measure(s) done at this specific location
- ▶ application domains : image analysis, environmental sciences, astronomy



## "The" questions :

- ▶ what is the data subset made of those elements having a certain "common" property ?
- ▶ what is the information contained by the data ?

## "The" answer :

- ▶ generally, these "common" property and information may be described by a **statistical analysis**
- ▶ the spatial coordinates of the spatial data elements add a **morphological** component to the answer
- ▶ the data subset and the information we are looking for, they form a **pattern** that has relevant **geometrical** characteristics

## Bayesian paradigm :

- ▶ use prior knowledge in order to answer the question

"The" question re-formulated :

- ▶ what is the pattern hidden in the data ?
- ▶ what are the geometrical and the statistical characteristics of this pattern ?

**Aim of the course** : provide you with some mathematical tools to allow you formulate answers to these questions

# Examples : data sets and related questions

For the purpose of this course : software and data sets are available

- ▶ R library : **spatstat** by A. Baddeley, R. Turner and contributors → [www.spatstat.org](http://www.spatstat.org)
- ▶ C++ library : **MPPLIB** by A. G. Steenbeek, M. N. M. van Lieshout, R. S. Stoica and contributors → available at simple demand

## My research partners :

- ▶ mathematics : Université de Lorraine, INRIA, Université de Lille, Universitat Jaume I Castellon, CWI Amsterdam
- ▶ astronomy and cosmology : Tartu Observatory, Observatorio Astronomico de Valencia, Observatoire de Paris IMCCE
- ▶ environmental sciences : RING and GeoRessources, INRA, CIRAD
- ▶ industry : SEDIF, St. Gobain

Forestry data (1) : the points positions exhibit attraction →  
clustered distribution

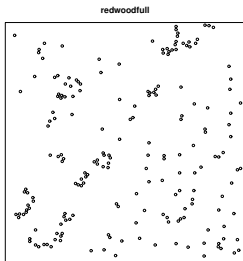


Figure: Redwoodfull data from the **spatstat** package

```
> library(spatstat)
> data(redwoodfull) ; plot(redwoodfull)
```

Forestry data (2) : the points positions exhibit neither attraction nor repulsion → completely random distribution

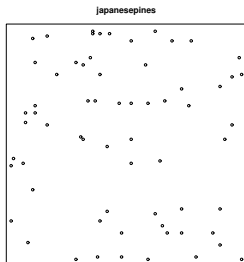


Figure: Japanese data from the **spatstat** package

In order to see all the available data sets

```
> data(package="spatstat")
```

Biological data (1) : the points positions exhibit repulsion →  
regular distribution

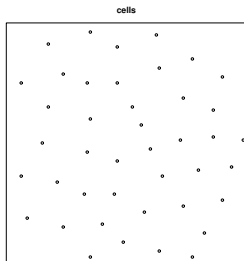


Figure: Cell data from the **spatstat** package

```
> data(cells)
> cells
planar point pattern: 42 points
window: rectangle = [0, 1] x [0, 1] units
```

Biological data (2) : two types of cells exhibiting attraction and repulsion depending on their relative positions and types

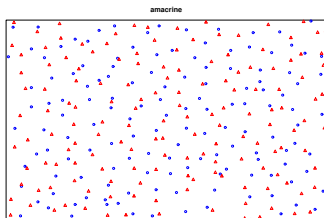


Figure: Amacrine data from the **spatstat** package

```
> data(amacrine) ;  
plot(amacrine,cols=c("blue","red"))
```



Geological data : two types of patterns, line segments and points  
→ are these patterns independent ?

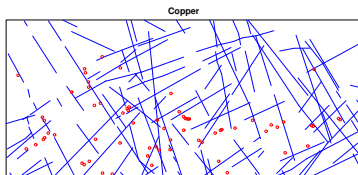
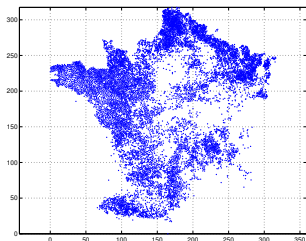


Figure: Copper data from the **spatstat** package

```
> attach(copper) ; L=rotate(Lines,pi/2) ;  
P=rotate(Points,pi/2)  
> plot(L,main="Copper",col="blue") ;  
points(P$x,P$y,col="red")
```

## Animal epidemiology : sub-clinical mastitis for diary herds

- ▶ points → farms location
- ▶ to each farm → disease score (continuous variable)
- ▶ **clusters pattern detection** : regions where there is a lack of hygiene or rigour in farm management

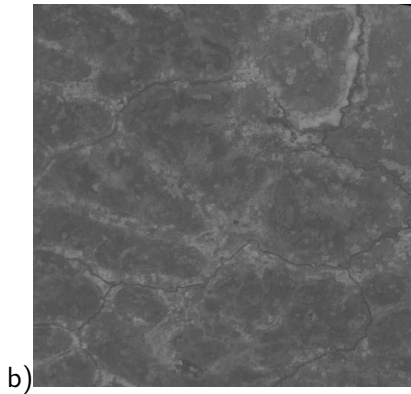
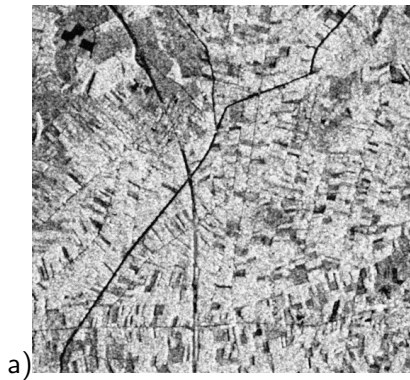


**Figure:** The spatial distribution of the farms outlines almost the entire French territory (INRA Avignon).

## Cluster pattern : some comments

- ▶ particularity of the disease : can spread from animal to animal but not from farm to farm
- ▶ cluster pattern : several groups (regions) of points that are close together and have the “same statistical properties”
- ▶ clusters regions → approximate it using interacting small regions (random disks)
- ▶ local properties of the cluster pattern : small regions where locally there are a lot of farms with a high disease score value
- ▶ problem : pre-visualisation is difficult ...

## Image analysis : road and hydrographic networks

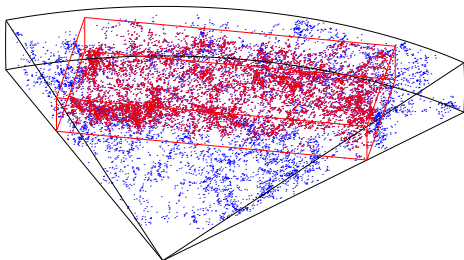


**Figure:** a) Rural region in Malaysia (<http://southport.jpl.nasa.gov>), b) Forest galleries (BRGM).

## Thin networks : some comments

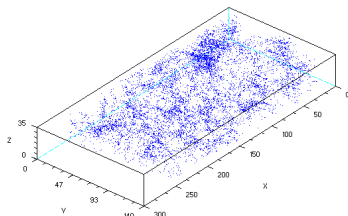
- ▶ road and hydrographic networks → approximate it by connected random segments
- ▶ topologies : roads are “straight” while rivers are “curved”
- ▶ texture : locally homogeneous, different from its right and its left with respect a local orientation
- ▶ avoid false alarms : small fields, buildings,etc.
- ▶ local properties of the network : connected segments covered by a homogeneous texture

# Cosmology (1) : spatial distribution of galactic filaments

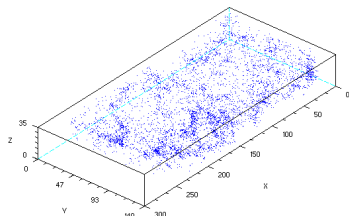


**Figure:** Cuboidal sample from the North Galactic Cap of the 2dF Galaxy Redshift Survey. Diameter of a galaxy  $\sim 30 \times 3261.6$  light years.

## Cosmology (2) : study of mock catalogs



a)



b)

**Figure:** Galaxy distribution : a) Homogeneous region from the 2dfN catalog, b) A mock catalogue within the same volume

# Cosmology (3) : questions and observations

## Few words about the 2dF GRS and SDSS catalogues

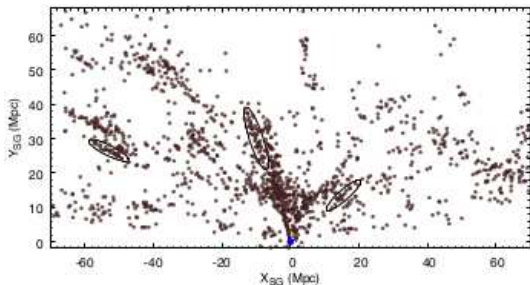
- ▶ filaments, walls and clusters : different size and contrast
- ▶ inhomogeneity effects (only the brightest galaxies are observed)
- ▶ filamentary network the most relevant feature
- ▶ local properties of the filamentary network : connecting random cylinders containing a “lot” of galaxies “along” its main axis

## Mock catalogues

- ▶ how “filamentary” they are w.r.t the real observation ?
- ▶ how the theoretical models producing the synthetic data fits the reality ?



## Cosmology (4) : cluster detection



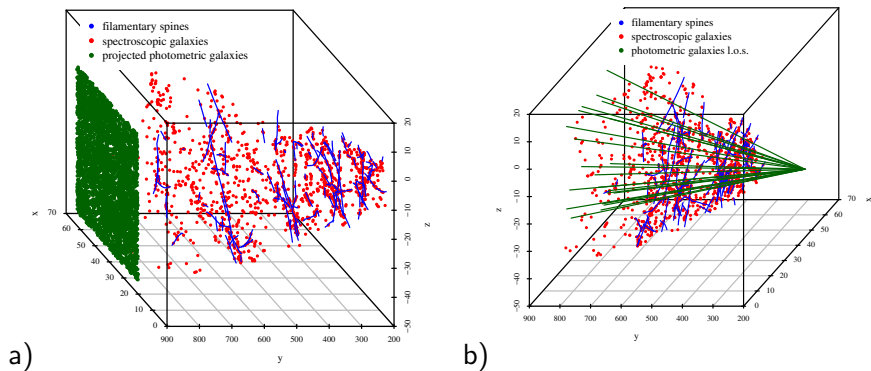
**Figure:** Distribution of galaxies in the 2MRS data set. Positions of galaxies are given in supergalactic coordinates, where observer is located at the origin of coordinates (marked as blue point on the figure). The thickness of the slice shown in the figure is 15 Mpc. Some galaxy clusters are marked with black ellipses to highlight the elongation of galaxy groups/clusters along the line of sight.

# Cosmology (5) : questions and observations

Few words about the 2MRS data set:

- ▶ more galaxies are observed
- ▶ price to pay : lack of precision for the third coordinates
- ▶ consequence : finger-of-God effect is much more important
- ▶ galaxy groups and clusters seem elongated along the line of sight
- ▶ inhomogeneity effects

## Cosmology (6) : influence of the new observations on the already detected structures



**Figure:** Photometric galaxies (green dots), spectroscopical galaxies (red dots) and filaments (blue) : a) photometrical galaxies projected on a sphere, b) photometrical galaxies lines of sight

## Few words about the SDSS Data Release 12 data set:

- ▶ much more galaxies are observed
- ▶ price to pay : bigger lack of precision for the third coordinates
- ▶ question : how this new data set is related to the already existing detected structures
- ▶ need to compare and to relate complex patterns in 2d and 3d

# Cosmology (7) : knowing the filamentary structure what is the galaxy distribution ?

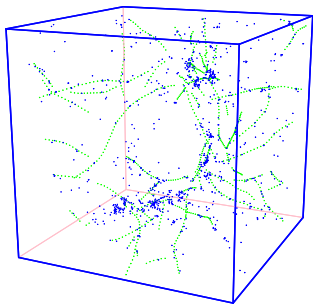
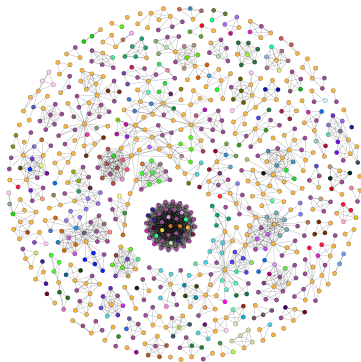


Figure: Data set : galaxy positions with spines (filaments main axes)

## Some questions related to galaxy distribution

- ▶ galaxies are spread as pearls on a necklace : (Tempel et al., 2014)
- ▶ inhomogeneity : filaments presence or interactions ?
- ▶ which of these factors controls the galaxy distribution ?
- ▶ what type of interaction : gravitational, territorial, component oriented ?
- ▶ what is the interaction range ?
- ▶ what is the size of a cluster ?

# Network science



**Figure:** Collaborations among researchers within the Loria laboratory : HAL data set (2018).

## Description of the network :

- ▶ each node represents a researcher,
- ▶ the edges are collaborative links
- ▶ nodes' color represent the affiliation to a laboratory.
- ▶ all Loria members are coloured in yellow, while the members of the other labs are differently coloured

## Related questions :

- ▶ what determines the occurrence of a collaborative link ?
- ▶ how the presence of cooperating individuals can be characterised ?
- ▶ how to quantify the cooperation behaviour of a research team ?



# Oort cloud comets (1)

## The comets dynamics :

- ▶ comets parameters  $(z, q, \cos i, \omega, \Omega)$   $\rightarrow$  inverse of the semi-major axis, perihelion distance, inclination, perihelion argument, longitude of the ascending node
- ▶ variations of the orbital parameters
- ▶ initial state : parameters before entering the planetary region of our Solar System
- ▶ final state : current state
- ▶ question : do these perturbation exhibit any spatial pattern ?

## Oort cloud comets (2)

### Study of the planetary perturbations

- ▶  $\sim 10^7$  perturbations were simulated
- ▶ data  $\rightarrow$  set of triplets  $(q, \cos i, \Delta z)$
- ▶ spatial data framework : location  $(q, \cos i)$  and marks  $\Delta z$
- ▶  $\Delta z = z_f - z_i$  : perturbations of the cometary orbital energy
- ▶ local properties of the perturbations : locations are uniformly spread in the observation domain, marks tend to be important whenever they are close to big planets orbits
- ▶ reformulated question : do these planetary observations exhibit an **observable** spatial pattern ?
- ▶ problem : pre-visualisation is **very** difficult ...

# Spatio-temporal data

Time dimension available :

- ▶ the previous example may be considered snapshots
- ▶ more recent data sets have also a temporal coordinate
- ▶ **question** : what is the pattern hidden in the data and its spatio-temporal description ?

# Binary asteroids

## Bayesian orbit determination :

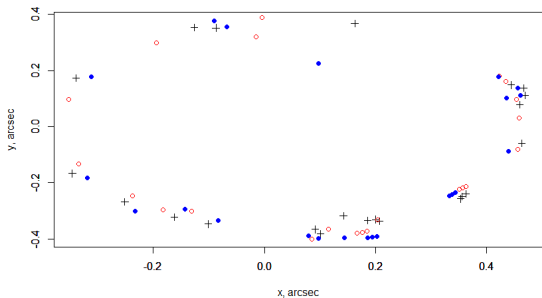
- ▶ input data :  $n$  observations at times  $t = (t_1, \dots, t_n)$
- ▶ observational equation :

$$\mathbf{w} = \psi(\mathbf{s}) + \varepsilon$$

with

- ▶  $\mathbf{w} = (x_1, y_1, \dots, x_n, y_n)$  : the rectangular coordinates of the secondary asteroid with primary asteroid in the origin
- ▶  $\mathbf{s} = (a, e, i, \Omega, \omega, T, P)$  : the apparent orbit Keplerian elements and the period of revolution  $P$
- ▶  $\varepsilon$  : noise and measure errors
- ▶ question : what are the best parameters  $\mathbf{s}$  fitting the observations  $\mathbf{w}$
- ▶ problem : the  $z_1, \dots, z_n$  coordinates are not observable
- ▶ our solution : elliptical regression + Bayesian modelling

## Illustration example :



**Figure:** Orbit detection result ( asteroid 283 Emma) : observed data (black points), our MCMC Bayesian detection method (blue points), N. V. Emeliyanov least square method (red points)

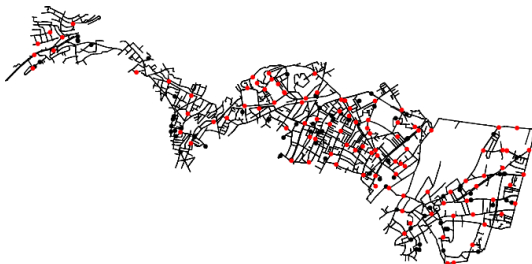
## Roads dynamics in Central Africa region :

- ▶ in forest region with rare woods, road networks appear and disappear within the territory of an exploitation concession
- ▶ there is a difference between “classical” road networks and “exploitation” networks → mining galleries
- ▶ this roads dynamics may be relevant in many aspects : health of the forest, respect of rules for the enterprises, environmental behaviour and understanding
- ▶ characterize the distribution and the dynamics of the road network

→ video roads dynamics

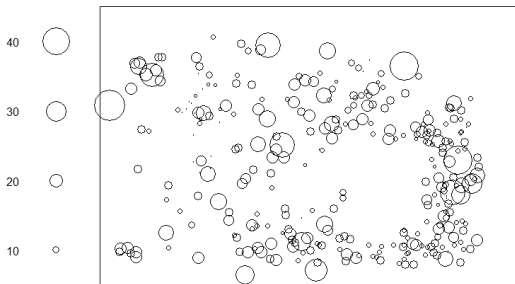
## Study the spatio-temporal spread of failures in a water distribution network : failures (black points), detector's activation (red points)

- ▶ information available : position, activation date, alert type, etc.
- ▶ SEDIF data and questions : do the detectors work ? do the failures form a particular pattern ?
- ▶ how to integrate the temporal dynamics ?



## Paleontology : Guérande salinas - fairy rings

- ▶ growing rings : territories occupied by cyano-bacterias
- ▶ the size of a ring is proportional with its age
- ▶ what determines the spatial distribution of the rings : edge effects, water arrival, interaction ?
- ▶ how to integrate the temporal dyanmics ?



**Figure:** The distribution of the morphostructures and their sizes (diameter in cm)



## Sismology : characterisation of earthquakes occurrences

- ▶ earthquakes : space-time events
- ▶ self-excitation phenomenon
- ▶ characterize and predict the seismic activity in different regions

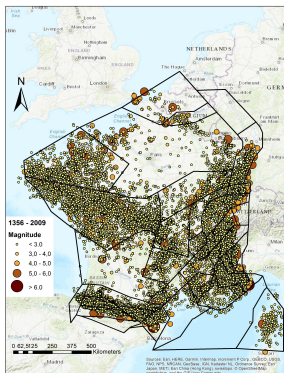


Figure: Representation of positions of earthquakes on the French territory

# Synthesis

**Hypothesis** : the pattern we are looking for can be approximated by a configuration of random objects that interact

- ▶ marked points pattern : repulsive or attractive marked points
- ▶ clusters pattern : superposing random disks
- ▶ filamentary network : connected and aligned segments

**Important remark** :

- ▶ locally : the number of objects is finite

## Marked point processes :

- ▶ probabilistic models for random points with random characteristics
- ▶ origin → **stochastic geometry**
- ▶ the pattern is described by means of a probability density → stochastic modelling
- ▶ the probability density allows the computation of average quantities and descriptors (these are integrals) related to the pattern
- ▶ conversely, whenever a pattern is observed, the probabilistic framework allows the derivation of the law of parameters conditioned on the observation : Bayesian inference
- ▶ not the only integrator ...

## Remarks :

- ▶ there exist also deterministic mathematical tools able to treat pattern recognition problems
- ▶ probability is cool : the phenomenon is not controlled, but understood
- ▶ probability thinking framework offers simultaneously **the analysis and the synthesis** abilities of the proposed method
- ▶ probabilistic approach deeply linked with physics :
  - ▶ exploratory analysis
  - ▶ model formulation
  - ▶ simulation
  - ▶ statistical inference

- ▶ comets example : random fields  $\rightarrow$  another probabilistic mathematical tool
  - ▶ unifying random fields and marked point processes is a mathematical challenge
  - ▶ determinantal point processes ... ?
- ▶ binary system example :
  - ▶ not especially a point process
  - ▶ the pattern we are looking for is a "tube" of ellipses
  - ▶ common points : Gibbsian formulation of the problem to optimize and Bayesian approach

- ▶ spatio - temporal examples :
  - ▶ spatio-temporal marked point processes
  - ▶ random sets theory
  - ▶ general idea : new data sets require new mathematics → **stochastic processes and stochastic geometry**
  - ▶ still, partial answers to these questions can be given using the tools presented in this course
- ▶ present challenge : **big data**
  - ▶ examples : particle tracking (quantum dots), cosmological data, etc.

### Important question :

- ▶ are you interested in an internship : IECL, Inria, GeoRessources, CRAN + travelling ?
- ▶ what about a phd ?

# Mathematical background

Measure and integration theory → blackboard

- ▶  $\sigma$ -algebra
- ▶ measurable space, sets, functions
- ▶ measure
- ▶ measure space, integral with respect to a measure
- ▶ probability space, probability measure

# Table of contents

## Course 2. Point processes, Binomial and Poisson point processes

Definition of a point process

Binomial point process

Poisson point process

Few words about self-exciting point processes

Exercises : binomial and Poisson point processes, simulation



## Course 2. Point processes, Binomial and Poisson point processes

### Construction of a point process : mathematical ingredients

- ▶ observation window : the measure space  $(W, \mathcal{B}, \nu)$ , with  $W \subset \mathbb{R}^d$ ,  $\mathcal{B}$  the Borel  $\sigma$ -algebra and  $0 < \nu(W) < \infty$  the Lebesgue measure
- ▶ points configuration space : probability space  $(\Omega, \mathcal{F}, \mathbb{P})$

## Configuration space construction :

- ▶ state space  $\Omega$  :

$$W_n \text{ is the set of all n-tuples } \{w_1, \dots, w_n\} \subset W$$
$$\Omega = \cup_{n=0}^{\infty} W_n, \quad n \in \mathbb{N}$$

- ▶ events space  $\mathcal{F}$  : the  $\sigma$ - algebra given by

$$\mathcal{F} = \sigma(\{w = \{w_1, \dots, w_n\} \in \Omega : n(w_B) = n(w \cap B) = m\})$$

for any bounded  $B \in \mathcal{B}$  and  $m \in \mathbb{N}$

- ▶ probability measure  $\mathbb{P}$  : **the model** answering our questions

## Definition

*A point process in  $W$  is a measurable mapping from a probability space  $(\mathcal{S}, \mathcal{A})$  in  $(\Omega, \mathcal{F})$ . Its distribution is given by*

$$\mathbb{P}(X \in F) = \mathbb{P}\{\omega \in \mathcal{S} : X(\omega) \in F\},$$

*with  $F \in \mathcal{F}$ . The realization of a point process is random set of points in  $W$ . We shall sometimes identify  $X$  and  $\mathbb{P}(X \in F)$  and call them both a point process.*

**Remarks :** point process  $\Rightarrow$  random configuration of points  $w$  in a observation window  $W$ . In the following, it is considered that :

- ▶ a points configuration is  $w = \{w_1, w_2, \dots, w_n\}$ , with  $n$  the corresponding number of points
- ▶ the process is locally finite :  $n(w \cap B)$  is finite whenever  $\nu(B)$  is finite
- ▶ the process is simple :  $w_i \neq w_j$  for  $i \neq j$

**Marked point processes** : attach characteristics to the points  $\rightarrow$   
extra-ingredient : marks probability space  $(M, \mathcal{M}, \nu_M)$

## Definition

*A marked point process is a random sequence  $x = \{x_n = (w_n, m_n)\}$  such that the points  $w_n$  are a point process in  $W$  and  $m_n$  are the marks corresponding for each  $w_n$ .*

**Examples :**

- ▶ random circles :  $M = (0, \infty)$
- ▶ random segments :  $M = (0, \infty) \times [0, \pi]$
- ▶ multi-type process :  $M = \{1, 2, \dots, k\}$

... and all the possible combinations ...  $\rightarrow$  drawing

**Stationarity and isotropy.** A point process  $X$  on  $W$  is **stationary** if it has the same distribution as the translated process  $X_w$ , that is

$$\{w_1, \dots, w_n\} \stackrel{\mathcal{L}}{=} \{w_1 + w, \dots, w_n + w\}$$

for any  $w \in W$ .

A point process  $X$  on  $W$  is **isotropic** if it has the same distribution as the rotated process  $rX$ , that is

$$\{w_1, \dots, w_n\} \stackrel{\mathcal{L}}{=} \{rw_1, \dots, rw_n\}$$

for any rotation matrix  $r$ .

- ▶ motion invariant : stationary and isotropic
- ▶ marked case : in principle easy to generalize, but take care ...
- ▶ counter example : a point process on a half plane is not stationary

Intuitive characterisation of a point process : being able to say how many points of the process  $X$  can be found in a neighbourhood in  $W$

The mathematical tools for point processes : should be able to do the following

- ▶ count the points of a point process in a **small neighbourhood of a point in  $W$** , and then extend the neighbourhood
- ▶ count the points of a point process in a **small neighbourhood of a typical point of the process  $X$** , and then extend the neighbourhood
- ▶ “counting” means using a **probability measure based counter**

Let  $X$  be a point process on  $W$ , and let us consider the counting variable

$$N(B) = n(X_B), \quad B \in \mathcal{B},$$

representing the number of points “falling” in  $B$ .

Let us consider also the sets of the form

$$F_B = \{x \in \Omega : n(x_B) = 0\},$$

that are called *void events*.



## Theorem

*The distribution of a point process  $X$  on a complete, separable metric space  $(W, d)$  is determined by the finite dimensional distributions of its count function, i.e. the joint distribution of  $N(B_1), \dots, N(B_m)$  for any bounded  $B_1, \dots, B_m \in \mathcal{B}$  and  $m \in \mathbb{N}$ .*

## Theorem

*The distribution of a simple point process on a complete, separable metric space  $(W, d)$  is uniquely determined by its void probabilities*

$$v(B) = \mathbb{P}(N(B) = 0), \quad B \in \mathcal{B}.$$

# Choquet's theorem

## Theorem

*The distribution of a random closed set  $X$  is entirely determined by the functional*

$$T_X(K) = \mathbb{P}(X \cap K \neq \emptyset)$$

*for every compact  $K$  in  $W$ .*

- ▶ random closed sets : more general object than a point process  
→ this object cannot be counted, but it can be observed through a compact window ...
- ▶ practical applications → Boolean model

# Binomial point process

The **trivial random pattern** : a single random point  $x$  uniformly distributed in a compact  $W$  such that

$$\mathbb{P}(x \in B) = \frac{\nu(B)}{\nu(W)}$$

for all  $B \in \mathcal{F}$ .

**More interesting point pattern** :  $n$  independent points distributed uniformly such that

$$\begin{aligned}\mathbb{P}(x_1 \in B_1, \dots, x_n \in B_n) &= \\ &= \mathbb{P}(x_1 \in B_1) \cdot \dots \cdot \mathbb{P}(x_n \in B_n) \\ &= \frac{\nu(B_1) \cdot \dots \cdot \nu(B_n)}{\nu(W)^n}\end{aligned}$$

for Borel subsets  $B_1, \dots, B_n$  of the compact  $W$ .

→ drawing

## Properties

- ▶ this process earns its name from a distributional probability
- ▶ the r.v.  $N(B)$  with  $B \subseteq W$  follows a binomial distribution with parameters

$$n = N(W) = n(x_W)$$

and

$$\rho = \frac{\nu(B)}{\nu(W)}$$

- ▶ the **intensity** of the binomial point process, or the mean number of points per unit volume

$$\rho = \frac{n}{\nu(W)}$$

- ▶ the mean number of points in the set  $B$

$$\mathbb{E}(N(B)) = np = \rho\nu(B)$$

- ▶ the binomial point process is simple
- ▶ number of points in different subsets of  $W$  are not independent even if the subsets are disjoint

$$N(B) = m \Rightarrow N(W \setminus B) = n - m$$

- ▶ the distribution of the point process is characterized by the *finite dimensional distributions*

$$\mathbb{P}(N(B_1) = n_1, \dots, N(B_k) = n_k) \quad \text{for } k = 1, 2, \dots$$

such that  $n_1 + n_2 + \dots + n_k \leq n$

- ▶ if the  $B_k$  are disjoint Borel sets with  $B_1 \cup \dots \cup B_k = W$  and  $n_1 + \dots + n_k = n$ , the finite-dimensional distributions are given by the multinomial probabilities

$$\begin{aligned} & \mathbb{P}(N(B_1) = n_1, \dots, N(B_k) = n_k) \\ &= \frac{n!}{n_1! \dots n_k!} \frac{\nu(B_1)^{n_1} \dots \nu(B_k)^{n_k}}{\nu(W)^n} \end{aligned}$$

- ▶ the void probabilities for the binomial point process are given by

$$\nu(B) = \mathbb{P}(N(B) = 0) = \frac{(\nu(W) - \nu(B))^n}{\nu(W)^n}$$

# Stationary Poisson point process

**Motivation** : what happens if extend  $W$  towards  $\mathbb{R}^d$  ?

- ▶ convergence binomial towards Poisson
- ▶ → drawing + blackboard

**Definition** : a stationary (homogeneous) Poisson point process  $X$  is characterized by the following fundamental properties

- ▶ **Poisson distribution of points counts.** The random number of points of  $X$  in a bounded Borel set  $B$  has a Poisson distribution with mean  $\rho\nu(B)$  for some constant  $\rho$ , that is

$$\mathbb{P}(N(B) = m) = \frac{(\rho\nu(B))^m}{m!} \exp(-\rho\nu(B))$$

- ▶ **Independent scattering.** The number of points of  $X$  in  $k$  disjoint Borel sets form  $k$  independent random variables, for arbitrary  $k$

## Properties

- ▶ **simplicity** : no duplicate points
- ▶ the mean number of points in a Borel set  $B$  is

$$\mathbb{E}(N(B)) = \rho\nu(B)$$

- ▶  $\rho$  : the **intensity or density** of the Poisson process, and it represents the mean number of points in a set of unit volume
- ▶  $0 < \rho < \infty$ , since for  $\rho = 0 \Rightarrow$  the process contains no points, while for  $\rho = \infty$  we get a pathological case



- ▶ if  $B_1, \dots, B_k$  are disjoint Borel sets, then  $N(B_1), \dots, N(B_k)$  are independent Poisson variable with means  $\rho\nu(B_1), \dots, \rho\nu(B_k)$ . Thus

$$\begin{aligned} & \mathbb{P}(N(B_1) = n_1, \dots, N(B_k) = n_k) \\ &= \frac{\rho^{n_1 + \dots + n_k} \nu(B_1)^{n_1} \dots \nu(B_k)^{n_k}}{n_1! \cdot \dots \cdot n_k!} \exp\left(-\sum_{i=1}^k \rho\nu(B_i)\right), \end{aligned}$$

- ▶ this formula can be used to compute joint probabilities for overlapping sets
- ▶ the void probabilities for the Poisson point process are given by

$$\nu(B) = \mathbb{P}(N(B) = 0) = \exp(-\rho\nu(B))$$

- ▶ the Poisson point process with  $\rho = \text{ct.}$  is stationary and isotropic
- ▶ if the intensity is a function  $\rho : W \rightarrow \mathbb{R}^+$  such that

$$\int_B \rho(w) d\nu(w) < \infty$$

for bounded subsets  $B \subseteq W$ , then we have a **inhomogeneous Poisson process** with mean

$$\mathbb{E}(N(B)) = \int_B \rho(w) d\nu(w) = \Upsilon(B)$$

- ▶  $\Upsilon$  is called the **intensity measure**
- ▶ we have already seen that for the stationary Poisson process :  
 $\Upsilon(B) = \rho\nu(B)$

## Theorem

*Conditionning a Poisson point process.* Let  $X$  be a stationary Poisson point process on  $\mathbb{R}^d$  with intensity  $\rho > 0$  and  $W$  a bounded Borel set with  $\nu(W) > 0$ . Then, conditional on the event  $\{N(W) = n\}$ ,  $X$  restricted to  $W$  is a binomial point process of  $n$  points.

- ▶ proof of the theorem : → [Exercice 1](#)
- ▶ bonus : → [Exercice 2, 3 and 4](#)

# Some properties of the Poisson point process

## Theorem

*Interval theorem.* Let  $X$  be a stationary point process on  $(0, \infty)$  with intensity  $\rho$  and let the points of  $X$  be written in ascending order :

$$0 < X_1 < X_2 < \dots < X_n < \dots$$

The the random variables :

$$Y_1 = X_1, Y_n = X_n - X_{n-1}$$

are independently, identically distributed according to  $g(y) = \rho \exp(-\rho y)$  for  $y > 0$ .

- ▶ bus paradox : if to the initial process a symmetric independent copy on  $(-\infty, 0)$  is added, then the interval between two consecutive points of the process containing 0 is longer than any other interval between two consecutive points
- ▶ no extension of this result for  $d \geq 2$

Maybe most important marked Poisson point process : the unit intensity Poisson point process with i.i.d. marks on a compact  $W$

- ▶ number of objects  $\sim \text{Poisson}(\nu(W))$
- ▶ locations and marks i.i.d. :  $w_i \sim \frac{1}{\nu(W)}$  and  $m_i \sim \nu_M$

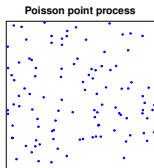
The corresponding probability measure : weighted ‘counting’ of objects

$$\mathbb{P}(X \in F) = \sum_{n=0}^{\infty} \frac{e^{-\nu(W)}}{n!} \int_{W \times M} \cdots \int_{W \times M} \mathbf{1}_F\{(w_1, m_1), \dots, (w_n, m_n)\} \\ \times d\nu(w_1) d\nu_M(m_1) \dots d\nu(w_n) d\nu_M(m_n)$$

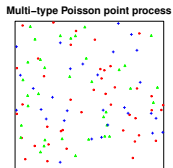
for all  $F \in \mathcal{F}$ .

**Remark** : the simulation of this process is straightforward, while the knowledge of the probability distribution allows analytical computations of the interest quantities

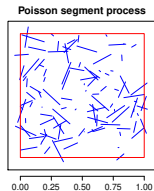
Simulations results of some Poissonian point processes : the domain is  $W = [0, 1] \times [0, 1]$  and the intensity parameter is  $\rho = 100$



a)



b)



c)

Figure: Poisson based models realizations : a) unmarked, b) multi-type and c) Poisson process of segments.

## Definition

A disjoint union  $\cup_{i=1}^{\infty} X_i$  of point processes  $X_1, X_2, \dots$  is called *superposition*.

## Proposition

If  $X_i \sim \text{Poisson}(W, \rho_i)$ ,  $i = 1, 2, \dots$  are mutually independent and if  $\rho = \sum \rho_i$  is locally integrable, then with probability one,  $X = \cup_{i=1}^{\infty} X_i$  is a disjoint union and est  $X \sim \text{Poisson}(W, \rho)$ .

→ *stable character of the Poisson process*

## Definition

Let be  $q : W \rightarrow [0, 1]$  a function and  $X$  a point process on  $W$ . The point process  $X_{thin} \subset X$  obtained by including the  $\xi \in X$  in  $X_{thin}$  with probability  $q(\xi)$ , where points are included/excluded independently of each other, is said to be an *independent thinning* of  $X$  with retention probabilities  $q(\xi)$ .

Formally, we can set

$$X_{thin} = \{\xi \in X : R(\xi) \leq q(\xi)\},$$

with the random variables  $R(\xi) \sim \mathcal{U}[0, 1]$ ,  $\xi \in W$ , mutually independent and independent of  $X$ .



## Proposition

Suppose that  $X \sim \text{Poisson}(W, \rho)$  is subject to independent thinning with retention probabilities  $q(\xi)$ ,  $\xi \in W$  and let

$$\rho_{thin} = q(\xi)\rho(\xi), \quad \xi \in W.$$

Then  $X_{thin}$  and  $X \setminus X_{thin}$  are independent Poisson processes with intensity functions  $\rho_{thin}$  and  $\rho - \rho_{thin}$ , respectively.

## Corollary

Suppose that  $X \sim \text{Poisson}(W, \rho)$  with  $\rho$  bounded by a positive constant  $C$ . Then  $X$  is distributed as independent thinning of a  $\text{Poisson}(W, C)$  with retention probabilities  $q(\xi) = \rho(\xi)/C$ .

## Some general facts concerning the Poisson point processes

- ▶ the Poisson point process is as important for spatial statistics as the Gaussian process in classical probability theory
- ▶ the law is completely known  $\rightarrow$  analytical formulas
- ▶ the Poisson process is invariant under independent thinning
- ▶ easy procedure for simulate non-stationary Poisson process
- ▶ completely random patterns : null or the default hypothesis that we want to reject
- ▶ independence  $\rightarrow$  no interaction  $\rightarrow$  no structure in the data

- ▶ two stationary Poisson point processes, they are not absolutely continuous with respect to each other, except if one process has unit intensity or if they have the same intensity
- ▶ two inhomogeneous Poisson point processes with strictly positive intensities, they are absolutely continuous with respect to each other
- ▶ more complicated models can be built  $\rightarrow$  specifying a probability density  $p(x)$  w.r.t. the reference measure given by the unit intensity Poisson point process. This probability measure is written as

$$\mathbb{P}(X \in F) = \int_F p(x) \mu(dx)$$

with  $\mu$  the reference measure.

**Remark :** in this case the normalizing constant is not available from an analytical point of view. To check this replace in the expression of  $\mu(\cdot)$  the indicator function  $1_F\{y\}$  with  $p(y)$  ...

$\rightarrow$  **Exercise 5**

## Few words about self-exciting point processes

**Hawkes processes** : a point process defined by its intensity of events conditional on the past  $\lambda^*(t) \rightarrow$  the intensity of the process evolves with the time depending on the points arrived in the configuration : **no more independence**

$$\lambda^*(t) = \lambda + \sum_{i=1}^n \mu(t - t_i)$$

such as

- ▶  $(t_i)_{1 \leq i \leq n}$  the sequence of arrival times of events that have occurred up to  $t$ .
- ▶  $\lambda$  background intensity.
- ▶  $\mu : [0, +\infty[ \rightarrow [0, +\infty[$  excitation function.

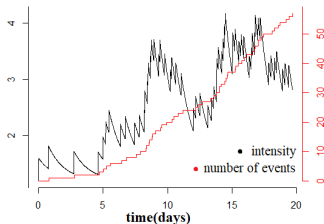
## Remarks :

- ▶ if  $\mu = 0 \Rightarrow$  classical Poisson process
- ▶ modelling : several models available for the excitation functions
- ▶ simulation : thinning method
- ▶ inference : the conditional intensity allows the construction of a likelihood function
- ▶ application : sismology and epidemiology (extend the definition of the conditional intensity)

**Exponential model** : an example of excitation function for Hawkes processes

$$\mu(t) = \alpha \exp(-\beta t)$$

with  $\alpha < \beta$ . The parameter  $\alpha$  gives the instantaneous influence of events and  $\beta$  the rate at which it decreases.



**Figure:** Number of events and conditional intensity of a Hawkes process, with exponential excitation function with parameters  $\alpha = 0.6$ ,  $\beta = 0.8$  et  $\lambda = 1.2$ .

## Exercises<sup>1</sup> : binomial and Poisson point processes, simulation

*Exercise 1.* Prove the Theorem related to the conditioning of a stationary Poisson point process.

*Hint :* Compute the void probabilities in subsets  $B \subset W$ .

*Exercise 2.* The spherical contact distribution for a point process is given by

$$F(r) = \mathbb{P}(d(w, X) \leq r)$$

where  $d(w, X)$  is the minimum distance from a given point  $w \in W$  to the point process  $X$ . Compute the expression of  $F(r)$  for a stationary Poisson point process with intensity  $\rho > 0$ .

---

<sup>1</sup>Part of the theoretical and practical exercises is due to the generous help of Zbyněk Pawlas from Charles University in Prague and Marie-Colette van Lieshout from CWI Amsterdam.

*Exercise 3.* Let  $X$  be a stationary Poisson point process in  $\mathbb{R}^2$ . Denote by  $D_X$  the distance from the origin to the nearest point in  $X$ . Calculate the mean and the variance of the random variable  $D_X$ .

*Exercise 4.* Let  $X_1, X_2, \dots$  be independent and exponentially distributed with parameter  $\rho$  and define a point process on  $\mathbb{R}^+$  by

$$X = \{X_1, X_1 + X_2, X_1 + X_2 + X_3, \dots\}$$

Calculate  $\mathbb{P}(N((0, t]) = 0)$  for  $t \in \mathbb{R}^+$ .



*Exercise 5.* Install the R package **spatstat**. Download the following documents :

- ▶ *Package spatstat*
  - ▶ *Analysing spatial point patterns in 'R'* by A. J. Baddeley
- 
- a) Simulate and print 10 realizations of a Binomial point process with  $n = 10$  points in the square  $W = [0, 1]^2$ .
  - b) Simulate and print 10 realization of a homogeneous Poisson point processes with intensity parameter  $\rho = 10$  in the square  $W = [0, 1]^2$ . Compare the realizations of the previous two processes. What do you observe ?

- c) Simulate an inhomogeneous Poisson point process given by the intensity
- ▶  $\rho(x_1, x_2) = 100(x^2 + y)$  in the domain  $W = [0, 2] \times [0, 1]$ .
  - ▶  $\rho(x_1, x_2) = 100 \exp[-(x^2 + y^2)]$  in the domain given by the ellipse  $\frac{x^2}{4} + y^2 - 1 = 0$
- d) Simulate and print a realization of a multi-type marked Point process with the following parameters : the locations process is a stationary Poisson point process in  $[0, 1]^2$ , while the marks probability law is the uniform distribution over three point types.
- e) Simulate and print a realization of a Poisson process of random discs. The centres locations process is the previous Poisson process, while the disk radius follows an uniform distribution on the interval  $[0, 0.05]$ .

- f) Simulate and print a realization of a Poisson process of random segments. The centres locations process is the previous Poisson process, while the orientation and lengths parameters are independently uniformly distributed in the intervals  $[0, \pi]$  and  $[0, 0.2]$ .
- g) Propose statistics that may characterize the previous processes ? Try to approximate them by simulation. How can this be used as a statistical test ?
- h) Test if the following spatstat datasets may be considered as a stationary Poisson point process : cells, swedishpines, japanesepines, redwood ?

*Hint* : Use the help. Some of the commands you may be interested in are ppp, psp, rpoispp, rmpoispp, owin, runifpoint.

# Table of contents

## Course 3. Moment and factorial moment measures

Moment and factorial moment measures, product densities

Moment measures for the Poisson process

Campbell measures

Campbell - Mecke formula

Exercises : moment and factorial moment measures

# Cours 3. Moment and factorial moment measures, product densities

Present context :

- ▶ mathematical background
- ▶ definition of a marked point process
- ▶ Binomial and Poisson point process
- ▶ important result : the point process law is determined by counts of points

Let  $X$  be a point process on  $W$ . The counts of points in bounded Borel regions of  $B \subset W$ ,  $N(B)$  characterize the point process and they are well defined random variables

- ▶ it is difficult to average the pattern  $X$
- ▶ it is possible to compute moments of the  $N(B)$ 's

The appropriate mathematical tools are :

- ▶ the moment measures
- ▶ the factorial moment measures
- ▶ the product densities

→ blackboard

# Moment measures of the Poisson process

- ▶ stationary Poisson point processes → Exercise 6
- ▶ the  $n$ -th product density measure of an independently thinned point process is

$$\rho_{\text{thin}}^{(n)}(w_1, \dots, w_n) = \rho^{(n)}(w_1, \dots, w_n) \prod_{i=1}^n q(w_i)$$

this gives the invariance under independent thinning of the  $n$ -th point correlation function (van Lieshout, 2011)

$$\frac{\rho_{\text{thin}}^{(n)}(w_1, \dots, w_n)}{\rho_{\text{thin}}(w_1) \cdot \dots \cdot \rho_{\text{thin}}(w_n)} = \frac{\rho^{(n)}(w_1, \dots, w_n)}{\rho(w_1) \cdot \dots \cdot \rho(w_n)}$$

- ▶ → Exercise 9 : explain the envelope tests

# Campbell measures

## Present context :

- ▶ counting points *i.e.* computing moment and factorial moment measures → very interesting tool for analysing point patterns : allow the computation of average quantities
- ▶ still, compute an average pattern → difficult and challenging problem
- ▶ **idea** : counting points that have some specific properties → **Campbell measures**



## Definition

Let  $X$  be a point process on  $W$ . The Campbell measure is

$$C(B \times F) = \mathbb{E}[N(B)1\{X \in F\}],$$

for any bounded  $B \in \mathcal{B}$  and  $F \in \mathcal{F}$ .

The first order moment measure can be expressed as a Campbell measure :

$$C(B \times \Omega) = \mathbb{E}[N(B)] = \mu^{(1)}(B).$$

- ▶ the moment and Campbell measures are not necessarily finite : their respective extension to an unique  $\sigma$ -finite measure can be shown (see van Lieshout, 2000) → Exercise 7

Higher order Campbell measures are constructed in a similar manner. For instance, the second order Campbell measure is

$$C^{(2)}(B_1 \times B_2 \times F) = \mathbb{E} [N(B_1)N(B_2)1\{X \in F\}],$$

from which we can get the second order moment measure

$$C^{(2)}(B_1 \times B_2 \times \Omega) = \mathbb{E} [N(B_1)N(B_2)] = \mu^{(2)}(B_1 \times B_2)$$

Remark :

- ▶ the moment measures allow to average functions  $h(x)$  measured in the location of a point process  $X$  : the function  $h$  **does not depend** on  $X$
- ▶ the Campbell measures allow to average functions  $h(x, x)$  measured in the location of a point process  $X$  : the function  $h$  **may depend** on  $X$

# Campbell - Mecke formula

## Theorem

Let  $h : W \times \Omega \rightarrow [0, \infty)$  a measurable function that is either non-negative either integrable with respect to the Campbell measure. Then

$$\mathbb{E} \left[ \sum_{w \in X} h(w, X) \right] = \int_W \int_{\Omega} h(w, x) dC(w, x).$$

Proof.

→ Exercise 8

□

# A more general Campbell-Mecke formulas

## Theorem

For a point process  $X$  and arbitrary nonnegative measurable function  $h$  that does not depend on  $X$  we have

$$\mathbb{E} \sum_{w_1, \dots, w_n \in X} h(w_1, \dots, w_n) = \int_W \cdots \int_W h(w_1, \dots, w_n) d\mu^{(n)}(w_1, \dots, w_n)$$

and

$$\mathbb{E} \sum_{\substack{\neq \\ w_1, \dots, w_n \in X}} h(w_1, \dots, w_n) = \int_W \cdots \int_W h(w_1, \dots, w_n) d\alpha^{(n)}(w_1, \dots, w_n)$$

## Proof.

Follow the same proof scheme as previously. □

## Remarks :

- ▶ If the function  $h$  does not depend on the point process  $X$ , the Campbell - Mecke becomes

$$\mathbb{E} \left[ \sum_{w \in X} h(w) \right] = \int_{\mathcal{W}} h(w) d\mu^{(1)}(w).$$

- ▶ point process of intensity function  $\rho(w)$

$$\mathbb{E} \left[ \sum_{w \in X} h(w) \right] = \int_{\mathcal{W}} h(w) \rho(w) d\nu(w).$$

- ▶ the preceding formula becomes for a stationary Poisson point process of intensity  $\rho > 0$

$$\mathbb{E} \left[ \sum_{w \in X} h(w) \right] = \rho \int_W h(w) d\nu(w),$$

this new formula is true for any stationary point process but is difficult to relate its intensity with its distribution ...

- ▶ point process of second order intensity function  $\rho^{(2)}(u, v)$

$$\mathbb{E} \left[ \sum_{\substack{\neq \\ u, v \in X}} h(u, v) \right] = \int_W \int_W h(u, v) \rho^{(2)}(u, v) d\nu(u) d\nu(v).$$



## Exercises : moment and factorial moment measures

*Exercise 6.* Let  $X$  be a stationary Poisson point process in  $\mathbb{R}^2$  with intensity parameter  $\rho$ . Compute the first and second order moment and factorial measures. Compute the corresponding product densities and the pair correlation function.

*Exercise 7.* Let  $X$  be a finite point process on a compact subset  $W \subset \mathbb{R}^d$  with the number of points given by  $p_n = \frac{1}{n(n-1)}$  for  $n \geq 2$  and zero otherwise

- Show that  $p_n$  is a probability function.
- Show that the first order moment measure on  $W$  is not finite.

*Exercise 8.* Prove the Campbell-Mecke Theorem.

*Hint :* start by considering indicator functions

$h(w, x) = 1\{(w, x) \in A \times F\}$  for some bounded Borel set  $A \in \mathcal{B}$  and some  $F \in \mathcal{F}$ .



### Exercise 9.

- a) Simulate and print a realization of a Poisson point processes with intensity parameter  $\rho = 100$  on the square  $W = [0, 1] \times [0, 1]$ . Use the **spatstat** package to compute and print estimates of the empty space function and of the pair correlation function for a pattern of points obtained by the simulation of the previous process. Compare the obtained values with their corresponding theoretical values.
- b) Build envelope tests based on these characteristics to compare the observed pattern with the realization of a Poisson process.
- c) Analyse the data sets : `redwoodfull`, `japanesepines` and `cells`. How, the empty space function and the pair correlation function can be used to diagnose clustering, repulsion or completely randomness of a pattern ? Try to propose a model for one of these data sets and test it.

*Hint* : Use the help. Commands you may be interested in are `pcf`, `Fest`, `envelope`, `density`.

*Exercise 10.  $\chi^2$  test of homogeneity.* Let  $x$  be an observed point pattern of a finite point process in a compact  $W \subset \mathbb{R}^2$ . Let us consider the hypothesis test given by :

- ▶  $H_0$  : the point process is stationary Poisson,
- ▶  $H_1$  : the point process is not stationary Poisson.

Divide the window  $W$  into quadrats  $B_1, \dots, B_m$  and count the numbers of points  $n_1, \dots, n_m$  in each quadrat. Under the null hypothesis, the  $n_j$ s are realisations of independent Poisson random variables with expected values  $\mu_j = \rho a_j$  where  $\rho$  is the unknown intensity and  $a_j = \nu(B_j)$ .

Given the total number of points  $n = \sum_j n_j$ , and the total window area  $a = \sum_j a_j$ , the estimated intensity is  $\hat{\rho} = n/a$ , and the expected count in quadrat  $B_j$  is  $e_j = \hat{\rho}a_j = na_j/a$ . The test statistic is

$$T = \sum_{j=1}^m \frac{(n_j - e_j)^2}{e_j}$$

and under  $H_0$  it follows a  $\chi^2$  distribution with  $m - 1$  degrees of freedom.

- a) Simulate and print a realization of a Poisson point processes with intensity parameter  $\rho = 200$  on the square  $W = [0, 1] \times [0, 1]$ .
- b) For a realisation of such a process visualise the decomposition in equal surface quadrats. What do you observe if you increase the number of quadrats ?
- c) For the same realisation visualise the decomposition in quadrats given by a Vornoi tessellation. What decomposition may you prefer and why ?

- d) Implement the  $\chi^2$  test of homogeneity based on equal surface quadrats.
- e) Implement the test using a Voronoi decomposition. How do you explain the obtained result ?
- f) Repeat the previous points for the realisation of an inhomogeneous Poisson point proces given by  $\rho(x, y) = 100(x^2 + y)$  in the domain  $W = [0, 2] \times [0, 1]$ .
- g) Analyse the data sets : redwoodfull, japanesepines and cells. Based on the  $\chi^2$  test, check if these patterns tend to be rather *regular* or *clustered*.
- h) What are the advantages and the drawbacks of using this test ?

*Hint* : Use the help. Commands you may be interested in are `quadratcount`, `intensity`, `dirichlet`, `quadrat.test`.

### Exercise 11.

- a) Simulate and print a realisation of an inhomogeneous Poisson point processes with intensity function given by  $\rho(x, y) = (x^2 + y^2)$  in the square  $W = [0, 5]^2$ .
- b) Intensity estimator proposed by Diggle is

$$\hat{\rho}(w) = \sum_{i=1}^n \frac{1}{e(w)} \kappa(w - x_i)$$

with the probability density kernel  $\kappa(w) \geq 0$  such that  $\int_{\mathbb{R}^2} \kappa(w) d\nu(w) = 1$  and the correction for bias due to edge effects

$$e(w) = \int_W \kappa(w - u) d\nu(u)$$

Outside the window  $W$ , the estimated intensity is zero. Use the **R** function `density` to estimate the intensity function for the previous realisation.

- c) Plot several density estimates using different kernel's bandwidths. What do you notice? Use the **R** command `bw.diggle` in order to find an optimal estimate of the kernel's bandwidth.
- d) Estimate the intensity of some real data sets: `redwoodfull`, `cells`, `japanesepines`. Are these patterns stationary? Motivate your answer.

- e) The **R spatstat** data set `bei` contains a pattern of points and covariates. Use the `help` to learn about the data structure. Type the following code to visualise the data set:

```
x11()  
map.data=persp(bei.extra$elev, theta=-45, phi=18,  
  expand=7,border=NA, apron=TRUE, shade=0.3,  
  box=FALSE, visible=TRUE,  
  colmap=terrain.colors(128),  
  main="Tropical rain forest - data")  
perspPoints(bei, Z=bei.extra$elev, M=map.data, pch=16,  
  col="blue",cex=0.5)
```

- f) Estimate the intensity of the `bei` data set point pattern ? Is the pattern stationary ? Motivate your answer.
- g) Is the intensity depending of one the available covariates ? A possible manner is to use the  $\chi^2$  test based on quadrats. First of all use the function `quanttess` to make a tessellation of the domain depending on the quantiles of the elevation covariate. Next, use the `quadrat.test` to reject a Poissonian assumption. What model you test in this way ? What are the alternatives you have considered ?



- h) Another possible strategy to analyse the dependence of the intensity on a covariate is to compare the observed distribution of the values of the covariate at the data points with the values of that covariate at all spatial locations in the observation window. Do you have an explanation for it ?
- i) Use the function `cdf.test.ppp` to test whether the intensity of the point pattern in `bei` data set depends on the elevation covariate. Plot the different results.

# Table of contents

## Course 4. Palm theory

Interior and exterior conditioning

A review of the Palm theory

Slivnyak - Mecke theorem

Applications : summary statistics

## Cours 4. Palm theory

### Interior and exterior conditioning : present context

- ▶ counting points measures : count points in a small region in  $W \rightarrow$  integrate using the Campbell Mecke formula
- ▶ **idea** : count points in a small region in  $W$  that is “centred” in a point of the process  $X \rightarrow$  **interior conditioning**
- ▶ **question** : how the measures applied to a process  $X$  change, if we add or if we delete a point from the current configuration  
 $\rightarrow$  **exterior conditioning**

## Palm distributions

- ▶ present construction → **blackboard**
- ▶ the Palm distributions of  $X$  at  $w \in W$  can be interpreted as

$$P_w(F) = \mathbb{P}(X \in F | N(\{w\}) > 0)$$

- ▶ the Campbell - Mecke formula can be expressed as

$$\mathbb{E} \left[ \sum_{w \in X} h(w, X) \right] = \int_W \int_{\Omega} h(w, x) dP_w(x) d\mu^{(1)}(w)$$

- ▶ for stationary point processes

$$\begin{aligned} \mathbb{E} \left[ \sum_{w \in X} h(w, X) \right] &= \rho \int_W \int_{\Omega} h(w, x) dP_w(x) d\nu(w) \\ &= \rho \int_W \int_{\Omega} h(w, x + w) dP_o(x) d\nu(w) \end{aligned}$$

# Slivnyak - Mecke theorem

## Theorem

If  $X \sim \text{Poisson}(W, \rho)$ , then for functions  $h : W \times \Omega \rightarrow [0, \infty)$ , we have

$$\mathbb{E} \sum_{w \in X} h(w, X \setminus \{w\}) = \int_W \mathbb{E} h(w, X) \rho(w) d\nu(w),$$

(where the left hand side is finite if and only if the right hand side is finite).

► proof : → Exercise 12

# General Slivnyak - Mecke theorem

## Theorem

If  $X \sim \text{Poisson}(W, \rho)$ , then for any  $n \in \mathbb{N}$  and any functions  $h : W^n \times \Omega \rightarrow [0, \infty)$ , we have

$$\begin{aligned} \mathbb{E} \sum_{w_1, \dots, w_n \in X}^{\neq} h(w_1, \dots, w_n, X \setminus \{w_1, \dots, w_n\}) \\ = \int_W \cdots \int_W \mathbb{E} h(w_1, \dots, w_n, X) \prod_{i=1}^n \rho(w_i) d\nu(w_i) \end{aligned}$$

where the  $\neq$  over the summation sign means that the  $n$  points  $w_1, \dots, w_n$  are pairwise distinct.

- ▶ proof : similar to the previous one + induction. For  $n \geq 2$  consider

$$\tilde{h}(w, x) = \sum_{w_2, \dots, w_n \in x}^{\neq} h(w, w_2, \dots, w_n, x \setminus \{w_2, \dots, w_n\})$$

- ▶ this theorem is a very strong result, since it allows computing averages of a Poisson point process knowing that one or several points belong to the process ...
- ▶ application in telecommunications : knowing, in this location I have a mobile phone antenna, how the quality of the signal change if I add randomly more antennas ? (the group of F. Baccelli)

- ▶ combining the Campbell-Mecke and the Slivnyak-Mecke theorem, we obtain for a Poisson process

$$\int_{\Omega} h(x) dP_w(x) = \int_{\Omega} h(x \cup \{w\}) dP(x)$$

- ▶ in words : the Palm distribution of a Poisson process with respect to  $w$  is simply the Poisson distribution plus an added point at  $w$
- ▶ a more mathematical formulation : the Palm distribution  $P_w^{\Upsilon}(\cdot)$  of a Poisson process of intensity measure  $\Upsilon$  and distribution  $\mathbb{P}^{\Upsilon}$  is the convolution  $\mathbb{P}^{\Upsilon} \star \delta_w$  of  $\mathbb{P}^{\Upsilon}$  with an additional deterministic point at  $w$
- ▶ explanation : **blackboard**

→ Exercise 13 and 14



Assumption :  $X$  is a stationary point process

- ▶ the nearest neighbour distance distribution function

$$G(r) = P_w(d(w, X \setminus \{w\}) \leq r) \quad (1)$$

with  $P_w$  the Palm distribution. The translation invariance of the distribution of  $X \rightarrow$  inherited by the Palm distribution  $\rightarrow$   $G(r)$  is well-defined and does not depend on the choice of  $w$ .

- ▶ replacing the Palm distribution in (1) by the distribution of  $X \rightarrow$  the spherical contact distribution or the empty space function

$$F(r) = \mathbb{P}(d(w, X) \leq r)$$

with  $\mathbb{P}$  the distribution of  $X$ .

- ▶ the J function : compares nearest neighbour to empty distances

$$J(r) = \frac{1 - G(r)}{1 - F(r)}$$

defined for all  $r > 0$  such that  $F(r) < 1$

The J function describes the morphology of a point pattern with respect to a Poisson process :

$$J(r) \text{ is } \begin{cases} = 1 & \text{Poisson : complete random} \\ \leq 1 & \text{clustering : attraction} \\ \geq 1 & \text{regular : repulsion} \end{cases}$$

For the stationary Poisson process of intensity parameter  $\rho$ , on  $W \subset \mathbb{R}^2$ , these statistics have exact formulas :

$$F(r) = 1 - \exp[-\rho\pi r^2]$$

$$G(r) = F(r)$$

$$J(r) = 1$$

→ proof the formulas at the blackboard + Exercice 17

# Table of contents

## Cours 5. Palm theory (continued). Papangelou conditional intensity.

Reduced Palm distributions

Campbell and Slivnyak theorems

Summary statistics : the  $K$  and  $L$  functions

Few words about summary statistics estimation

Applications : summary statistics

Exterior conditioning : conditional intensity

Exercises : Palm theory - interior and exterior conditioning

## Cours 5. Palm theory (continued). Papangelou conditional intensity.

### Reduced Palm distributions : present context

- ▶ Palm distributions : count the points in a neighbourhood centred on a point of the process  $\rightarrow$  this point is counted as well
- ▶ idea : in some applications (telecommunications) we may wish to measure the effect of a point process in a location being a point of the process, while this particular point has no effect on the entire process  $\rightarrow$  **reduced Palm distributions**
- ▶ the following mathematical development is rather easy to follow since it is similar to what we have already seen during until now

# Reduced Campbell measure

## Definition

Let  $X$  be a simple point process on the complete, separable metric space  $(W, d)$ . The reduced Campbell measure is

$$C^!(B \times F) = \mathbb{E} \left[ \sum_{w \in X \cap B} 1_{\{X \setminus \{w\} \in F\}} \right],$$

for any bounded  $B \in \mathcal{B}$  and  $F \in \mathcal{F}$ .

- ▶ the analogue of Campbell-Mecke formula reads

$$\mathbb{E} \left[ \sum_{w \in X} h(w, X \setminus \{w\}) \right] = \int_W \int_{\Omega} h(w, x) dC^!(w, x).$$

- ▶ assuming the first order moment measure  $\mu^{(1)}$  of  $X$  exists and it is  $\sigma$ -finite, we can apply Radon-Nikodym theory to write

$$C^1(B \times F) = \int_B P_w^1(F) d\mu^{(1)}(w),$$

for any bounded  $B \in \mathcal{B}$  and  $F \in \mathcal{F}$

- ▶ the function  $P_w^1(F)$  is defined uniquely up to an  $\mu^{(1)}$ -null set
- ▶ it is possible to find a version such that for fixed  $w \in W$ ,  $P_w^1(\cdot)$  is a probability distribution  $\rightarrow$  **the reduced Palm distribution**

# Campbell and Slivnyak theorems

- ▶ the reduced Palm distribution can be interpreted as the conditional distribution

$$P_w^!(F) = \mathbb{P}(X \setminus \{w\} \in F | N(\{w\}) > 0)$$

- ▶ the Campbell-Mecke formula equivalent

$$\mathbb{E} \left[ \sum_{w \in X} h(w, X \setminus \{w\}) \right] = \int_W \int_{\Omega} h(w, x) dP_w^!(x) d\mu^{(1)}(w).$$



- ▶ for stationary point processes

$$\begin{aligned} & \mathbb{E} \left[ \sum_{w \in X} h(w, X \setminus \{w\}) \right] \\ &= \rho \int_W \int_{\Omega} h(w, x) dP_w^!(x) d\nu(w) \\ &= \rho \int_W \int_{\Omega} h(w, x + w) dP_o^!(x) d\nu(w) \end{aligned}$$

- ▶ the Slivnyak-Mecke theorem : for a Poisson process on  $W$  with distribution  $\mathbb{P}$ , we have

$$P_w^!(\cdot) = \mathbb{P}(\cdot)$$

- ▶ there is a general result linking the reduced Palm distribution and the distribution of a Gibbs process → a little bit later in this course ...

## Example

*The nearest neighbour distribution  $G(r)$  of stationary process can be expressed in terms of the Palm distributions*

$$G(r) = 1 - P_o(X \in \Omega : N(b(o, r)) = 1),$$

*and the reduced Palm distributions*

$$G(r) = 1 - P_o^!(X \in \Omega : N(b(o, r)) = 0),$$

*where  $N(b(o, r))$  is the number of points inside the ball centred at the origin  $o$  of radius  $r$ .*

# Summary statistics : the $K$ and $L$ functions

The  $K$  function : theoretical explanations → blackboard

- ▶ maybe one of the most used summary statistic
- ▶ for a stationary process, its definition depending on the reduced Palm distribution is

$$\rho K(r) = \mathbb{E}_o^! [N(b(o, r))]$$

- ▶ the  $L$  function is

$$L(r) = \left[ \frac{K(r)}{\omega_d} \right]^{1/d}$$

with  $\omega_d = \nu(b(0, 1))$  the volume of the  $d$ -dimensional unit ball

- ▶ for stationary point processes, the pair correlation function is

$$g(r) = \frac{K'(r)}{\sigma_d r^{d-1}}$$

with  $\sigma_d$  the surface area of the unit sphere in  $\mathbb{R}^d$

- ▶ for the stationary Poisson process we have

$$K(r) = \omega_d r^d, \quad g(r) = 1$$

and

$$L(r) = r$$

- ▶ → **Exercise 15 and 16**

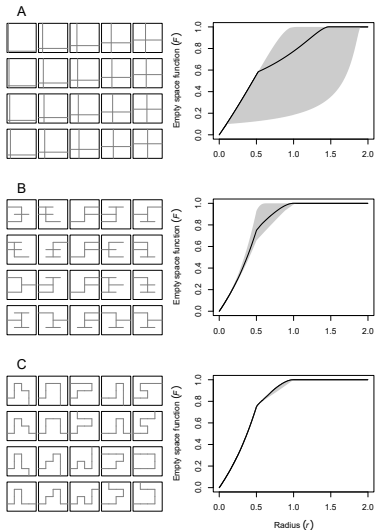
# Few words about summary statistics estimation

- ▶ 'Robbins' theorem
- ▶ spatial sampling bias
- ▶ unbiased sampling rules
- ▶ Horvitz-Thompson and spatial Horvitz-Thompson estimators
- ▶ sampling bias for point processes
- ▶ application : Fritz Kleinschroth's work on road network dynamics and Maarja's phd results ...  
→ blackboard

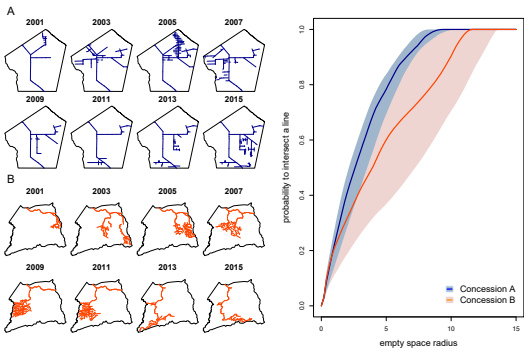
# Applications : summary statistics

## Road network dynamics :

- ▶ **data**: spatio-temporal evolution of a road network due to logging activity → you have already seen the video ...
- ▶ **questions**:
  - ▶ do the exploitation companies respect the regulations for preserving the forest ?
  - ▶ does the economical activity affect the resilience capacity of the forest ?
- ▶ **exploratory analysis**: empty space function → road-less space measure
- ▶ **challenge**: build a stochastic model
- ▶ **people**: F. Kleinschroth, J. R. Healey, S. Gourlet-Fleury, F. Mortier, M. N. M. van Lieshout
- ▶ **paper**: (Kleinschroth et al., 2017)



**Figure:** Toy model for explaining the behaviour of the empty space function: the simulated roads have the length, so the same density of roads per unit of surface.



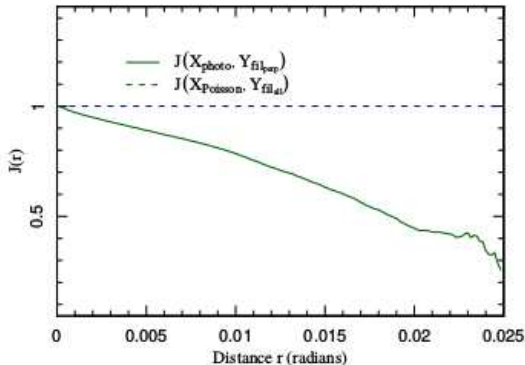
**Figure:** Empty space function for characterization the road network evolution in two logging companies.



## Cosmology : influence of the new observations on the already detected structure

- ▶ **data**: SDSS Data Release 12 - photometrical galaxies
- ▶ **question**:
  - ▶ how this new data set is related to the already detected structure ?
- ▶ **exploratory analysis**: adapt the  $F$ ,  $G$  and  $J$  function to establish possible dependence of different types of patterns
- ▶ **challenge**: the position of the photometrical galaxies is not entirely known
- ▶ **people**: M. Kruuse, E. Tempel, R. Kipper
- ▶ **paper**: in preparation

## Preliminary result :



**Figure:** Estimation of the bivariate  $J$  function. The considered sets were the photometric galaxies and the projection on the sphere of the filamentary spines that are rather perpendicular on the line of sight. This result indicates positive association of these two patterns.

## Exterior conditioning : conditional intensity

- ▶ assume that for any fixed bounded Borel set  $B \in \mathcal{B}$ , the reduced Campbell measure  $C^!(B \times \cdot)$  is absolutely continuous with respect to the distribution  $\mathbb{P}(\cdot)$  of  $X$

- ▶ then

$$C^!(B \times F) = \int_F \Lambda(B; x) d\mathbb{P}(x)$$

for some measurable function  $\Lambda(B; \cdot)$  specified uniquely up to a  $\mathbb{P}$ -null set

- ▶ moreover, one can find a version such that for fixed  $x$ ,  $\Lambda(\cdot; x)$  is a locally finite Borel measure  $\rightarrow$  the first order Papangelou kernel

- ▶ if  $\Lambda(\cdot; x)$  admits a density  $\lambda(\cdot; x)$  with respect to the Lebesgue measure  $\nu(\cdot)$  on  $W$ , the Campbell-Mecke theorem becomes

$$\begin{aligned}
 & \mathbb{E} \left[ \sum_{w \in X} h(w, X \setminus \{w\}) \right] \\
 &= \int_W \int_{\Omega} h(w, x) dC^1(w, x) \\
 &= \mathbb{E} \left[ \int_W h(w, X) \lambda(w; X) d\nu(w) \right]
 \end{aligned}$$

- ▶ the function  $\lambda(\cdot; \cdot)$  is called **the Papangelou conditional intensity**
- ▶ the previous result is known as **the Georgii-Nguyen-Zessin formula**

- ▶ the case where the distribution of  $X$  is dominated by a Poisson process is especially important

### Theorem

*Let  $X$  be a finite point process specified by a density  $p(x)$  with respect to a Poisson process with non-atomic finite intensity measure  $\nu$ . Then  $X$  has Papangelou conditional intensity*

$$\lambda(u; x) = \frac{p(x \cup \{u\})}{p(x)}$$

for  $u \notin x \in \Omega$ .

Proof.

→ Exercise 18



## Importance of the conditional intensity :

- ▶ intuitive interpretation :

$$\lambda(u; x)d\nu(u) = \mathbb{P}(N(du) = 1 | X \cap (d\nu(u))^c = x \cap (d\nu(u))^c),$$

the infinitesimal probability of finding a point in a region  $d\nu(u)$  around  $u \in W$  given that the point process agrees with the configuration  $x$  outside of  $d\nu(u)$

- ▶ the “conditional reverse” of the Palm distributions

- ▶ describe the local interactions in a point pattern  $\rightarrow$  Markov point processes
- ▶ if

$$\lambda(u; x) = \lambda(u; \emptyset)$$

for all patterns  $x$  satisfying  $x \cap b(u, r) = \emptyset \rightarrow$  the process has 'interactions of range  $r$  at  $u$ '

- ▶ in other words, points further than  $r$  away from  $u$  do not contribute to the conditional intensity at  $u$

- ▶ integrability of the model
- ▶ convergence of the Monte Carlo dynamics able to simulate the model
- ▶ differential characterization of Gibbs point processes → blackboard
  - ▶ the Slivnyak-Mecke theorem links Palm distributions with the distribution of a Poisson point process
  - ▶ this characterization links the Palm distributions with the distributions of a finite point process
  - ▶ this characterization is an essential element for extending finite point process to  $\mathbb{R}^d$



## Exercises : Palm theory - interior and exterior conditioning

*Exercise 12.* Prove the Slivnyak-Mecke Theorem.

*Hint :* start by considering the distribution function of Poisson point process in  $W$  and compute the desired expectation.

*Exercise 13.* Compute the average number of pairs of points in a stationary Poisson process of intensity  $\rho$  on the planar unit square separated by a distance that does not exceed some fixed  $r < \sqrt{2}$ .

- Do this computation conditioning on the event  $N(W) = n$ .
- Do this computation using the Campbell - Mecke formula.

*Hint :*  $\mathbb{E} [\mathbb{E}[X|Y]] = \mathbb{E}[X]$ .

*Exercise 14.* Let  $U_1$  and  $U_2$  be two independent random variables with uniform distribution on the interval  $[0, r]$ ,  $r > 0$ . Define a point process  $X$  in  $\mathbb{R}^2$  as

$$X = \bigcup_{m,n \in \mathbb{Z}} (U_1 + mr, U_2 + nr), \quad m, n \in \mathbb{Z}$$

where  $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$ . Determine the intensity measure and the Palm distributions of  $X$ .

*Exercise 15.* Let  $X$  be a stationary Poisson point process on  $\mathbb{R}^d$  with intensity parameter  $\rho$ . Prove that :

$$K(r) = \omega_d r^d \quad \text{and} \quad L(r) = r,$$

where  $\omega_d$  is the volume of the unit sphere in  $\mathbb{R}^d$ .

*Exercise 16.* This exercise studies alternative definitions for the Palm distribution and the  $G$ -function. Let  $X$  be a stationary point process in  $\mathbb{R}^d$  with intensity  $\rho$ .

a) Show that

$$P_v(F) = \frac{1}{\rho \nu(A)} \mathbb{E} \sum_{u \in X} \mathbf{1}\{u \in A, X+v-u \in F\}, \quad v \in \mathbb{R}^d, F \in \mathcal{F}$$

for an arbitrary set  $A \subset \mathbb{R}^d$  with  $0 < \nu(A) < \infty$ .

b) Show that

$$G(r) = \frac{1}{\rho\nu(A)} \mathbb{E} \sum_{u \in X} 1\{u \in A, (X \setminus \{u\}) \cap b(u, r) \neq \emptyset\}, \quad r > 0,$$

for an arbitrary set  $A \subset \mathbb{R}^d$  with  $0 < \nu(A) < \infty$ .

*Hint* : use the Campbell-Mecke theorem

*Exercise 17.*

- a) Simulate and print a realization of a Poisson point processes with intensity parameter  $\rho = 100$  on the square  $W = [0, 1] \times [0, 1]$ . Use the **spatstat** package to compute and print estimates of the  $G$ ,  $K$  and of the  $J$  function for a pattern of points obtained by the simulation of the previous process. Compare the obtained values with their corresponding theoretical values.
- b) Build envelope tests based on these characteristics to compare the observed pattern with the realization of a Poisson process.

- c) Analyse the data sets : `redwoodfull`, `japanesepines` and `swedishpines`.
- d) Analyse the behaviour of the different tree species in the data set : `lansing`.

*Hint* : `help(lansing)`. For the last two points use all the summary statistics you know.

*Exercise 18.* Prove the Papangelou conditional intensity Theorem.  
*Hint* : use the Campbell-Mecke and the Georgii-Nguyen-Zessin formulas.

# Table of contents

## Cours 6. Poisson based point processes : Cox and Boolean processes

Cox processes

Cluster processes

Boolean model

Capacity functional. Choquet theorem

Exercises : Poisson based point processes

## Cours 6. Poisson based point processes.

### Direct application of counting measures and Palm distributions for point process analysis

- ▶ counting measures  $\rightarrow$  summary statistics for point pattern characterization
- ▶ two categories : interpoint distances ( $F$ ,  $G$  and  $J$ ) and second order characteristics ( $\rho$ ,  $K$  and  $L$ )
- ▶ possible extension of the summary statistics : marks, non-stationary processes, different observation spaces  $W$  case and spatio-temporal



- ▶ non-parametrical estimation of the summary statistics : kernel estimation and management of the border effects + numerical sensitivity and computational cost
- ▶ central limit available : statistical tests
- ▶ envelope simulation based tests : reject a particular model
- ▶ summary statistics for parameter estimation of a given model :
  - ▶ these statistics are an "equivalent" of the moments in probability theory, hence they do not entirely determine the model to be estimated
  - ▶ (Baddeley and Silverman, 1984)
  - ▶ (Bedford and Berg, 1997)

- ▶ good exploring tool : **spatstat**
- ▶ outline important characteristics of a point pattern : clustering, repulsion, completely randomness
- ▶ it is **difficult to differentiate between interaction and inhomogeneity** if only one realisation is available
- ▶ **need for models** able to reproduce these characteristics
- ▶ counting or choosing a typical point is not always obvious : **consider random measure theory**

# Cox processes

## Definition

Let  $\Upsilon$  be a random locally finite diffuse measure on  $(W, \mathcal{B})$ . If the conditional distribution of  $X$  given  $\Upsilon$  is a Poisson process on  $W$  with intensity measure  $\Upsilon$ ,  $X$  is said to be a **Cox point process with driving measure  $\Upsilon$** . Sometimes  $X$  is also called **doubly stochastic Poisson process**.

## Remarks :

- ▶ if there exists a random field  $Z = \{Z(w), w \in W\}$  such that

$$\Upsilon(B) = \int_B Z(w) d\nu(w)$$

then  $X$  is a **Cox process with driving function  $Z$**

- ▶ the conditional distribution of  $X$  given  $Z = z$  is a distribution of the Poisson process with intensity function  $z \Rightarrow$

$$\mathbb{E}[N(B)|Z = z] = \int_B z(w) d\nu(w)$$

- ▶ the first order factorial moment measure is obtained using the law of the total expectation

$$\begin{aligned} \mu^{(1)}(B) &= \alpha^{(1)}(B) = \mathbb{E}[N(B)] \\ &= \mathbb{E}[\mathbb{E}[N(B)|Z = z]] = \mathbb{E}\left[\int_B Z(w) d\nu(w)\right] \\ &= \mathbb{E}[\Upsilon(B)] = \int_B \mathbb{E}Z(w) d\nu(w) \end{aligned}$$

- ▶ if  $\rho(w) = \mathbb{E}Z(w)$  exists then it is the intensity function

- ▶ similarly, it can be shown that the second order factorial moment measure is

$$\begin{aligned}\alpha^{(2)}(B_1 \times B_2) &= \mathbb{E}[\Upsilon(B_1)\Upsilon(B_2)] \\ &= \mathbb{E}\left[\int_{B_1} Z(u)d\nu(u) \int_{B_2} Z(v)d\nu(v)\right] \\ &= \mathbb{E}\left[\int_{B_1} \int_{B_2} Z(u)Z(v)d\nu(u)d\nu(v)\right] \\ &= \int_{B_1} \int_{B_2} \mathbb{E}[Z(u)Z(v)] d\nu(u)d\nu(v)\end{aligned}$$

- ▶ if  $\rho^{(2)}(u, v) = \mathbb{E}Z(u)Z(v)$  exists, then it is the product density

- ▶ the pair correlation function is

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

- ▶ depending on  $Z$  it is possible to obtain analytic formulas for the second order characteristics ( $g, K$  and  $L$ ) and the interpoint distance characteristic ( $F, G$  and  $J$ )

- ▶ the variance  $\text{Var}N(B)$  is obtained using the total variance law, and it is

$$\text{Var}N(B) = \mathbb{E}N(B) + \text{Var} \left[ \int_B Z(w) d\nu(w) \right] \geq \mathbb{E}N(B)$$

⇒ over - dispersion of the Cox process counting variables

- ▶ the void probabilities of Cox processes are

$$\begin{aligned} \mathbb{P}(N(B) = 0) &= \mathbb{E}1\{N(B) = 0\} \\ &= \mathbb{E} [\mathbb{E}1\{N(B) = 0\} | Z = z] = \mathbb{E} [\mathbb{P}(N(B) = 0 | Z = z)] \\ &= \mathbb{E} \left[ \exp \left( - \int_B Z(w) d\nu(w) \right) \right] \\ &= \mathbb{E} [\exp(-\Upsilon(B))] \end{aligned}$$

## Trivial Cox process : mixed Poisson processes

- ▶  $Z(w) = Z_0$  a common positive random variable for all locations  $w \in W$
- ▶  $X|Z_0$  follows a homogeneous Poisson process with intensity  $Z_0$
- ▶ the driving measure is  $\Upsilon(B) = Z_0\nu(B)$

## Thinning of Cox processes :

- ▶  $X$  is a Cox process driven by  $Z$
- ▶  $\Pi = \{\Pi(w) : w \in W\} \subseteq [0, 1]$  is a random field which is independent of  $(X, Z)$
- ▶  $X_{thin}|\Pi \rightarrow$  the point process obtained by independent thinning of the points in  $X$  with retention probabilities  $\Pi$
- ▶  $\Rightarrow X_{thin}$  is a Cox process driven by  $Z_{thin}(w) = \Pi(w)Z(w)$



## Log Gaussian Cox processes

- ▶ introduced independently by astronomers (Coles and Jones, 1991) and statisticians (Møller et. al., 1998)
- ▶ consider  $Y = \log Z$  is a Gaussian field
  - ▶ for any integer  $n > 0$ , locations  $\xi_1, \dots, \xi_n \in \mathbb{R}^d$  and numbers  $a_1, \dots, a_n \in \mathbb{R}$ ,  $\sum_{i=1}^n a_i Y(\xi_i)$  follows a normal distribution
- ▶ the Cox process  $X$  driven by  $Z = \exp(Y)$  is a *log Gaussian Cox Process* (LGCP)

- ▶ the distribution of  $(X, Y)$  is entirely determined by the mean and the covariance function

$$m(\xi) = \mathbb{E}Y(\xi) \quad \text{and} \quad c(\xi, \eta) = \text{Cov}(Y(\xi)Y(\eta))$$

- ▶ covariance function :
  - ▶ for simplicity it may be considered translation invariant

$$c(\xi, \eta) = c(\xi - \eta)$$

of the form

$$c(\xi) = \sigma^2 r(\xi/\alpha)$$

- ▶ the function  $r : \mathbb{R}^d \rightarrow [-1, 1]$  is a correlation function for a Gaussian field iff  $r$  is positive definite

$$\sum_{i=1}^n a_i a_j r(\xi_i, \xi_j) > 0 \quad \text{for all} \quad \xi_1, \dots, \xi_n \in \mathbb{R}^d, a_1, \dots, a_n \in \mathbb{R}$$

- ▶ weak conditions are required on  $m$  and  $r$  in order to get

$$\Upsilon(B) = \int_B Z(\xi) d\nu(\xi)$$

for bounded  $B \subset \mathbb{R}^d$ . For instance, we may require  $\xi \rightarrow Y(\xi)$  continuous almost surely

- ▶ as example, this is satisfied by continuous  $m$  and  $r$  such that

$$r(\xi) = \exp(-\|\xi\|^\delta), \quad 0 \leq \delta \leq 2$$

with  $\delta$  controlling the smoothness of the realizations of the Gaussian field

- ▶  $\delta = 1$  : exponential correlation function
  - ▶  $\delta = 1/2$  : stable correlation function
  - ▶  $\delta = 2$  : Gaussian correlation function
- ▶ there is a one-to-one correspondence between  $(m, c)$  and  $(g, \rho)$  *Rightarrow* the distribution of  $(X, Y)$  is uniquely determined by  $(\rho, g)$

→ Exercise 19

# Cluster processes

## Definition

Let  $C$  be a point process (parent process), and for each  $c \in C$  let  $X_c$  be a finite point process (daughter process). Then

$$X = \bigcup_{c \in C} X_c$$

is called a *cluster point process*.

## Definition

Let  $X$  be a cluster point process such that  $C$  is a Poisson point process and conditional on  $C$ , the processes  $X_c$ ,  $c \in C$  are independent. Then  $X$  is called a *Poisson cluster point process*.

# Neyman-Scott processes

## Definition

*Let  $X$  be a Poisson cluster point process such that centred daughter processes  $X_c - c$  are independent of  $C$ . Given  $C$ , let the points of  $X_c - c$  be i.i.d. with probability density function  $k$  on  $\mathbb{R}^d$  and  $N(X_c)$  be i.i.d. random variables. Then  $X$  is called a Neyman-Scott process. If moreover  $N(X_c)$  given  $C$  has a Poisson distribution with intensity  $\alpha$ , then  $X$  is a Neyman-Scott Poisson process.*

→ drawing + Exercice 20

## Theorem

Let  $X$  be a Neyman-Scott Poisson process such that  $C$  is a stationary Poisson process with intensity  $\kappa$ . Then  $X$  is stationary process with intensity  $\rho = \alpha\kappa$  and pair correlation function

$$g(u) = 1 + \frac{h(u)}{\kappa},$$

where

$$h(u) = \int k(v)k(u+v)d\nu(v)$$

is the density for the difference between two independent points distributed according to  $k$ .

Proof.

→ Exercise 21



## Other very known cluster point processes

Matérn cluster process (Matérn 1960,1986)

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

is the uniform density on the ball  $b(o, r)$

Thomas process (Thomas 1949)

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

is the density for  $\mathcal{N}_d(0, \omega^2 I_d)$ , i.e. for  $d$  independent normally distributed variables with mean 0 and variance  $\omega^2 > 0$

- ▶ both kernels are isotropic
- ▶ the Thomas process pair correlation function is

$$g(u) = 1 + \frac{1}{\kappa(4\pi\omega^2)^{d/2}} \exp \left[ -\frac{\|u\|^2}{4\omega^2} \right]$$

and its  $K$ -function for  $d = 2$  is

$$K(r) = \pi r^2 + \frac{1 - \exp[-r^2/(4\omega^2)]}{\kappa}$$

- ▶ other summary statistics can be also computed
- ▶ the expressions of the summary statistics are more complicated for the Matérn process

→ drawing the processes ...



## Remarks :

- ▶ usually in applications  $Z$  is unobserved
- ▶ one cannot distinguish a Cox process  $X$  from its corresponding Poisson process  $X|Z$  whenever a single realisation of  $X$  is available
- ▶ open question : which of the two models might be most appropriate, i.e. whether  $Z$  should be random or “systematic” /deterministic

- ▶ prior knowledge of the observed phenomenon
- ▶ Bayesian setting of the intensity function  $\Rightarrow$  Cox processes
- ▶ if we want to investigate the dependence of certain covariates associated to  $Z$ , these may be treated as systematic terms, while unobserved effects may be treated as random terms
- ▶ Cox process: more flexible models for clustered patterns than inhomogeneous Poisson point processes
- ▶ parameter estimation methods: minimum contrast, Palm distributions, composite likelihood
  - ▶ based on the  $K$  function and the Palm distributions
  - ▶ spatstat: the `kppm` function

→ Exercise 22

# Boolean model

Random objects “centred” around Poissonian points  $\rightarrow$  germs and grains

- ▶ germs : a stationary Poisson point process  $X$  of intensity  $\rho$  on  $\mathbb{R}^d$
- ▶ grains : a sequence of i.i.d. random compact sets  $\Gamma_1, \Gamma_2, \dots$  and independent of  $X$

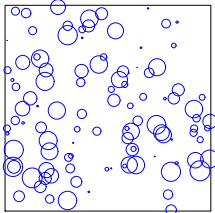
The **Boolean model is the random set** obtained by the replacement of the germs by the appropriately shifted corresponding set, and taking the set union as it follows

$$\Gamma = \bigcup_{n=1}^{\infty} (\Gamma_n + w_n) = (\Gamma_1 + w_1) \cup (\Gamma_2 + w_2) \cup \dots$$

The random set  $\Gamma_0$  is said to be the **typical grain**. The set  $\Gamma$  is also called the **Poisson germ-grain** model.

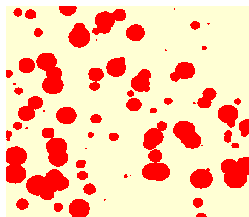
The Boolean model observation is an **incomplete** observation of a marked point process, since the locations points is **not available**

Boolean model of random discs



a)

Boolean model : what is really observed



b)

**Figure:** Boolean model of random discs : marked point process (complete) and random sets perspectives.

## Remarks :

- ▶ classical references : Matheron (1975), Molchanov (1997), Lantuéjoul (2002), Chiu et al. (2013)
- ▶ important practical applications → one of the first models of complex pattern
- ▶ Neyman-Scott processes are Boolean models as well
- ▶ very convenient models since they allow the analytical computation of quantities describing them → this is due mainly to the independence properties allowed in the construction of the process
- ▶ independence  $\leftrightarrow$  no objects interactions  $\leftrightarrow$  no structure

# Capacity functional. Choquet theorem

- ▶ in general, for random sets it is rather difficult to use moment and factorial measures  $\leftrightarrow$  it is not possible to “count” points
- ▶ un-marked and marked point processes are particular random sets

## Definition

*The capacity functional the random closed set  $\Gamma$  is*

$$T_{\Gamma}(K) = \mathbb{P}(\Gamma \cap K \neq \emptyset)$$

*for  $K$  an element of the family  $\mathcal{K}$  of compact sets in  $\mathbb{R}^d$ .*

## Theorem

*(Choquet theorem). The distribution of a random closed set  $\Gamma$  is completely determined by the capacity functionals  $T_{\Gamma}(K)$  for all  $K \in \mathcal{K}$ .*

# Capacity functional of the Boolean model

## Proposition

The capacity functional of the Boolean model  $\Gamma$  is

$$T_{\Gamma}(K) = 1 - \exp \left[ -\rho \mathbb{E}(\nu(\check{\Gamma}_0 \oplus K)) \right].$$

- ▶ the reflection of the typical grain :

$$\check{\Gamma}_0 = -\Gamma_0 = \{-w : w \in A\}, \quad \text{for } A \subset \mathbb{R}^d$$

- ▶ the Minkowski addition :

$$A \oplus B = \{u + v : u \in A, v \in B\}, \quad \text{for } A, B \subset \mathbb{R}^d$$

Proof.

→ blackboard: Exercise 23



## Basic characteristics of the Boolean model

- ▶ the volume fraction : the mean fraction of volume occupied by  $\Gamma$  in a region of unit volume

$$p = \mathbb{P}(o \in \Gamma) = 1 - \exp[-\rho \mathbb{E}(\nu(\Gamma_o))]$$

- ▶ the non-centred covariance

$$C(r) = \mathbb{P}(o \in \Gamma \text{ and } r \in \Gamma) = 2p - 1 + (1 - p)^2 \exp[\rho \eta_{\Gamma_o}]$$

with  $\eta_{\Gamma_o} = \mathbb{E}[\nu(\Gamma_o \cap (\Gamma_o - r))]$

- ▶ contact distribution

$$H_B(r) = 1 - \frac{1 - T_{\Gamma}(rB)}{1 - p}$$

→ Exercise 24 + some more explanations for the contact distribution



# Stability of the Boolean model

## Proposition

*The following properties are satisfied :*

*i) the union of two independent Boolean models is a Boolean model*

*ii) a Boolean model dilated by a non-empty compact subset of  $\mathbb{R}^d$  is a Boolean model*

*iii) the intersection between a Boolean model and a compact subset of  $\mathbb{R}^d$  is a Boolean model*

*iv) the cross-section of a Boolean model by an  $i$ -flat is a Boolean model*

**Proof.**

→ blackboard: Exercise 25



## Exercises : Poisson based point processes

*Exercise 19.* The intensity and pair correlation function of an LGCP are given by

$$\rho(\xi) = \exp(m(\xi) + c(\xi, \xi)/2), \quad g(\xi, \eta) = \exp(c(\xi, \eta)).$$

*Hint :* the moment generating function of the normal distribution with mean  $m$  and variance  $\sigma^2$  is  $\exp(mt + \sigma^2 t^2/2)$ .

*Exercise 20.* In the case of a Neyman-Scott Poisson process as defined during the course, show that  $X_c$  given  $C$  are independent Poisson processes with intensity function  $\rho(w) = \alpha k(w - c)$ .

*Hint :* compute the void probabilities.

*Exercise 21.* Prove the stationarity of a Neyman-Scott Poisson process and the computation of its pair-correlation function.

*Hint :* compute the void probabilities

*Exercise 22.* Use the **spatstat** package to compute and print estimates of the known summary statistics (second order and interpoint distances) for

- a) Thomas process with parameters  $\alpha = 10$ ,  $\kappa = 10$  and  $\omega^2 = 0.01$  in a window  $W = [0, 1] \times [0, 1]$ .
- b) Matérn cluster process with parameters  $\alpha = 10$ ,  $\kappa = 10$  and  $r = 0.1$  in a window  $W = [0, 1] \times [0, 1]$ .
- c) LGCP with exponential correlation function and parameters  $m = 4$ ,  $\sigma^2 = 0.2$ ,  $\alpha = 0.1$

- d) What is the theoretical intensity of these processes ? Do you see any differences between two realizations of these two processes, respectively ? How, can you use these observations in order to chose an appropriate model for a given data set ?
- e) Estimate the model parameters of the three precedent models on the data set redwood.
- f) Compare the three results using an envelope test based on the  $L$  function.

*Hint* : Spatstat commands : `rThomas`, `rMatClust`, `rLGCP`, `kppm`, `envelope`. Install the package `RandomFields`. Play with model parameters in order to obtain different configuration topologies.

*Exercise 23.* Prove the capacity functional formula of the Boolean model.

*Exercise 24.* Let  $\Gamma$  be a Boolean process on  $\mathbb{R}^d$  with intensity parameter  $\rho$ . Prove the formulas presented in the course, for the volume fraction, the covariance and the contact distribution function.

*Exercise 25.* Prove the stability properties of the Boolean model.

# Table of contents

## Cours 7. Gibbs models.

Gibbs models

Construction of the probability density

Markov point processes

Exercises : Gibbs models

# Gibbs models

## Construction of the probability density of a point process :

- ▶ the independence property of the Poisson based processes does not allow to introduce point interactions
- ▶ **Gibbs models** are flexible point processes that allow the specification of point interactions via a a probability density
- ▶ in the following, let us consider  $(W, d)$  a complete, separable metric space such that  $W \subset \mathbb{R}^d$  and  $0 < \nu(W) < \infty$
- ▶ let  $\mu$  be the unit intensity Poisson point process on  $W$
- ▶ the condition that  $W$  has to be finite is required in order to obtain a well defined probability density for the considered Gibbs model

- ▶ the probability density of a Gibbs model is a Radon-Nykodim derivative w.r.t the Poissonian reference measure  $\mu$
- ▶ within a statistical physics context, the probability density of such a point process has the form

$$p(x) = \frac{\exp[-U(x|\theta)]}{\alpha(\theta)}$$

with  $U(x|\theta)$  the energy function,  $\theta$  the model parameters and

$$\alpha(\theta) = \int_{\Omega} \exp[-U(x|\theta)] d\mu(x)$$

the normalizing constant or the partition function



- ▶ under these circumstances, the probability distribution of a Gibbs model writes as

$$\mathbb{P}(X \in F) = \int_F p(x) d\mu(x)$$

and by introducing the expression of  $\mu$ , it is further expressed as

$$\begin{aligned} \mathbb{P}(X \in F) &= \\ &= \sum_{n=0}^{\infty} \frac{\exp[-\nu(W)]}{n!} \int_W \cdots \int_W 1(\{x_1, \dots, x_n\} \in F) \times \\ &\quad p(\{x_1, \dots, x_n\}) d\nu(x_1) \dots d\nu(x_n), \end{aligned}$$

whenever  $n > 0$ . If  $n = 0$ , we take  $\exp[-\nu(W)]1(\emptyset \in F)p(\emptyset)$ . If  $\nu(W) = 0$ , then  $P(X = \emptyset) = 1$ . For applications, we always assume that  $\nu(W) > 0$ .

- ▶ the marked case writes in a similar way by introducing also the marks distribution  $\nu_M$

- ▶ usually the probability density is known only up to a constant :  $p \propto h = \exp(-U)$
- ▶ the normalizing constant or the partition function is given by

$$\alpha = \int_{\Omega} h(x) d\mu(x)$$

that becomes

$$\alpha = \sum_{n=0}^{\infty} \frac{\exp[-\nu(W)]}{n!} \int_W \cdots \int_W h(\{x_1, \dots, x_n\}) d\nu(x_1) \cdots d\nu(x_n) \quad (2)$$

- ▶ the previous quantity is **not always available under analytical closed form**
- ▶ this is the main difficulty to be solved while using this approach ...

**Normalizing constant for the Poisson process** : Let  $\rho$  be the intensity function of a Poisson point process on  $W$ . Its probability density up to a normalizing constant is

$$\rho(x) \propto \prod_{x_i \in x} \rho(x_i).$$

Let  $\Upsilon(B) = \int_B \rho(w) d\nu(w)$  be the associated intensity measure. By using (2), we get

$$\alpha = \exp[-\nu(W)] \sum_{n=0}^{\infty} \frac{\Upsilon(W)^n}{n!} = \exp[\Upsilon(W) - \nu(W)],$$

that gives for the complete probability density

$$\rho(x) = \exp[\nu(W) - \Upsilon(W)] \prod_{x_i \in W} \rho(x_i)$$

If the process is stationary  $\rho(x) = \rho = \text{ct.}$ , then the probability density is

$$\rho(x) = \exp[(1 - \rho)\nu(W)]\rho^n$$

## Remarks :

- ▶ the probability density is specified only for finite point processes
- ▶ the extension to  $\mathbb{R}^d$  of a finite point process specified by a probability density is possible under some conditions (see (Møller and Waagpetersen, 2004, section 6.4))
- ▶ two such conditions that are equivalent :
  - ▶ local specification for Markov point processes
  - ▶ differential characterisation of Gibbs point process : link Palm distribution and conditional intensity
- ▶ phase transition - if such an extension is possible, does it surely leads to an unique probability measure ?

# Construction of the probability density

- ▶ specify the interaction functions  $\phi^{(k)} : \Omega \rightarrow \mathbb{R}^+$

$$\phi(x_{i_1}, \dots, x_{i_k})^{(k)}$$

for any  $k$ -tuple of objects

- ▶ the un-normalized probability density is the product of all these functions

$$h(x) = \prod_{x_i \in \mathcal{X}} \phi(x_i)^{(1)} \dots \prod_{\{x_{i_1}, \dots, x_{i_k}\} \in \mathcal{X}} \phi(x_{i_1}, \dots, x_{i_k})^{(k)} \quad (3)$$

- ▶ clearly, the energy function is obtained by taking  $U(x) = -\log h(x)$
- ▶  $\alpha$  the normalizing constant is difficult to be determined : untractable mathematical formula

- ▶ the un-normalized probability densities (3) are suitable for modelling provided they are integrable on  $\Omega$  ; that is

$$\alpha = \int_{\Omega} h(x) d\mu(x) < \infty.$$

- ▶ the following results ensure the integrability of the probability density of a marked point process  $\rightarrow$  the Ruelle stability conditions

## Definition

Let  $X$  be a marked point process given by the un-normalized probability density  $h$  w.r.t the reference measure  $\mu$ . The process  $X$  is stable in the sense of Ruelle, if it exists  $\Lambda > 0$  such that

$$h(x) \leq \Lambda^{n(x)}, \quad \forall x \in \Omega. \quad (4)$$

## Proposition

The un-normalized probability density of a stable point process is integrable.



### Proof.

The integrability of  $h(x)$  follows directly from the preceding condition :

$$\begin{aligned} \int_{\Omega} h(x) \mu(dx) &\leq \int_{\Omega} \Lambda^{n(x)} \mu(dx) \\ &= \sum_{n=0}^{\infty} \frac{\exp[-\nu(W)] [\Lambda \nu(W)]^n}{n!} = \exp[\nu(W)(\Lambda - 1)]. \end{aligned}$$



## Definition

*Under the same hypotheses as in Prop. 5, a marked point process is said to be locally stable if it exists  $\Lambda > 0$  such that*

$$h(x \cup \{\eta\}) \leq \Lambda h(x), \quad \forall x \in \Omega, \eta \in W \times M \setminus x \quad (5)$$

## Proposition

*A locally stable point process is stable in the sense of Ruelle.*

Proof.

It is easy to show by induction that

$$h(x) \leq h(\emptyset)\Lambda^{n(x)}, \quad \forall x \in \Omega.$$

The local stability of a point process (5) implies its integrability (4). □

- ▶ the conditional intensity for a point process  $X$  with probability density  $p$  is

$$\lambda(\eta; x) = \frac{p(x \cup \{\eta\})}{p(x)} = \frac{h(x \cup \{\eta\})}{h(x)}, \quad x \in \Omega, \eta \in W \times M \setminus x,$$

taking  $a/0 = 0$  for  $a \geq 0$

- ▶ the conditional intensity is also known in the literature as the Papangelou intensity condition (we have already meet it)
- ▶ we shall often consider functions  $h : \Omega \rightarrow [0, \infty[$  which are hereditary

$$h(x) > 0 \Rightarrow h(y) > 0, \quad \text{for } y \subset x.$$

- ▶ if  $p$  is hereditary, then there is a one-to-one correspondence between  $p$  and  $\lambda$

Importance of the conditional intensity : key element in modelling

- ▶ plays a similar role as the conditional probabilities for Markov random fields
- ▶ integrability
- ▶ convergence properties of the MCMC algorithms used to sample from  $p$
- ▶ the process  $X$  is attractive if  $x \subseteq y$  implies

$$\lambda(\eta; x) \leq \lambda(\eta; y),$$

and repulsive otherwise

$$\lambda(\eta; x) \geq \lambda(\eta; y),$$

- ▶ attractive processes tend to cluster the points, while the repulsive ones tend to distance the points
- ▶ these conditions are important also for **exact MCMC algorithms**
- ▶ there exist processes that are neither attractive nor repulsive
- ▶ there are processes that are integrable but not locally stable : Lennard - Jones (statistical physics)

# Markov point processes

The conditional intensity of an interacting point process is given by

$$\lambda(\eta; \mathbf{x}) = \phi(\eta)^{(1)} \prod_{x_i \in \mathbf{x}} \phi(x_i, \eta)^{(2)} \dots \prod_{\{x_{i_1}, \dots, x_{i_k}\} \in \mathbf{x}} \phi(x_{i_1}, \dots, x_{i_k}, \eta)^{(k+1)}$$

- ▶ difficult to manipulate
- ▶ possible simplifications : limit the order of interactions  $\rightarrow$  only pairs of points for instance
- ▶ limit the range of the interaction : a point interact only with its closest neighbours

Let  $\sim$  be a symmetrical and reflexive relation between points belonging to  $W \times M$ . This may be a typical neighbourhood relation based on a metric (Euclidean, Hausdorff) or on sets intersection.

### Definition

*A clique is a configuration  $x \in \Omega$  such that  $\eta \sim \zeta$  for all  $\eta, \zeta \in x$ .  
The empty set is a clique.*



## Definition

Let  $X$  be a marked point process on  $W \times M$  with probability density  $p$  w.r.t the reference measure  $\mu$ . The process  $X$  is Markov if for all  $x \in \Omega$  such that  $p(x) > 0$ , the following conditions are simultaneously fulfilled :

(i)  $p(y) > 0$  for all  $y \subseteq x$  (hereditary)

(ii)  $\frac{p(x \cup \{\zeta\})}{p(x)}$  depends only on  $\zeta$  and  $\partial(\zeta) \cap x = \{\eta \in x : \eta \sim \zeta\}$ .

This process is known in the literature as the **Ripley-Kelly Markov process**.

**Example :** The probability density w.r.t to  $\mu$  of a marked Poisson process on  $W \times M$  with constant intensity function  $(\rho(\eta) = \beta > 0)$  is

$$p(x) = \beta^{n(x)} \exp[(1 - \beta)\nu(W)].$$

Clearly  $p(x) > 0$  for all configurations  $x$ . Its Papangelou conditional intensity is

$$\lambda(\eta; x) = \beta 1\{\eta \notin x\}.$$

Hence, the Poisson process is Markov, independently of the interaction functions  $\phi^{(k)}$ . This agrees with the choice of the **Poisson process** for modelling a **completely random structure**.

The following result is known as the **spatial Markov property**.

→ drawing

### Theorem

*Let  $X$  be a Markov point process with density  $p(\cdot)$  on  $W$  and consider a Borel set  $A \subseteq W$ . Then the conditional distribution of  $X \cap A$  given  $X \cap A^c$  depends only on  $X$  restricted to the neighbourhood*

$$\partial(A) \cap A^c = \{u \in W \setminus A : u \sim a \text{ for some } a \in A\}.$$

Proof.

→ Exercise 26



The following result is known as the **Hammersley-Clifford theorem**.

### Theorem

A marked point process density  $p : \Omega \rightarrow \mathbb{R}^+$  is Markov with respect to the interaction relation  $\sim$  if and only if there is a measurable function  $\phi_c : \Omega \rightarrow \mathbb{R}^+$  such that

$$p(x) = \prod_{\substack{\text{cliques} \\ y \subseteq x}} \phi_c(y), \quad \alpha = \phi(\emptyset) \quad (6)$$

for all  $x \in \Omega$ .

Proof.

→ Exercise 27



## Remarks :

- ▶ the previous result simplifies the writing of the probability density of an interacting point process
- ▶ taking  $\phi_c(y) = 1$  whenever  $y$  is not a clique leads us to the equivalence of (3) and (6)
- ▶ Markov point processes are known in physics community as **Gibbs point processes**

$$p(x) = \frac{1}{Z} \exp[-U(x)] = \frac{1}{Z} \exp \left[ - \sum_{\substack{\text{cliques} \\ z \subseteq x}} U_c(z) \right],$$

with  $Z$  the partition function,  $U$  the system energy and  $U_c = \log \phi_c$  the clique potential

- ▶ all the Markov processes are Gibbs
- ▶ the reciprocal is not true

Poisson process as a Markov process : the probability density of a Poisson point process is

$$p(x) = e^{(1-\beta)\nu(W)} \prod_{x \in x} \beta.$$

Hence, the interactions functions applied to cliques are

$$\begin{aligned}\phi_c(\emptyset) &= e^{(1-\beta)\nu(W)} \\ \phi_c(\{u\}) &= \beta\end{aligned}$$

with  $\phi_c \equiv 1$  for the cliques made of more than one object. The potential of the cliques made of a single object is

$$U_c(u) = -\log \beta,$$

while  $U_c = 0$  otherwise. This confirms the lack of interaction in the Poisson process. It validates also, the choice of this process to model patterns exhibiting no particular morphological structure.

Distance interaction model - Strauss model : (Strauss, 1975),  
(Kelly and Ripley, 1976)

$$p(x) = \alpha\beta^{n(x)}\gamma^{sr(x)}, \quad \alpha, \beta > 0, \gamma \in [0, 1]$$

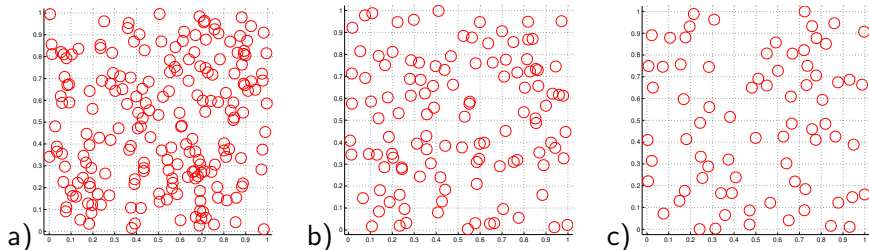


Figure: Strauss model realisations for different parameter values : a)  $\gamma = 1.0$ , b)  $\gamma = 0.5$  and c)  $\gamma = 0.0$ .

The interaction function  $\gamma : W \times W \rightarrow [0, 1]$  is

$$\gamma(u, v) = \begin{cases} \gamma & \text{if } d(u, v) \leq r \\ 1 & \text{otherwise} \end{cases}$$

The conditional intensity of adding a point  $\eta$  to  $x \setminus \{\eta\}$  is

$$\lambda(u; x) = \beta \gamma^{\text{card} \partial(u)}$$

where  $\partial(u) = \{v \in x : d(u, v) \leq r\}$



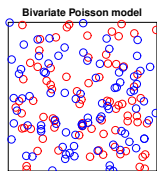
The Strauss model is a locally stable model with  $\Lambda = \beta$  and Markov with interaction range  $r$ .

The interaction functions applied to cliques are

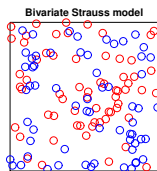
$$\begin{aligned}\phi_c(\emptyset) &= \alpha \\ \phi_c(\{u\}) &= \beta \\ \phi_c(\{u, v\}) &= \gamma(u, v)\end{aligned}$$

and  $\phi_c \equiv 1$  if the cliques have three or more objects. The interaction potentials are obtained taking  $U_c = -\log \phi_c$ .

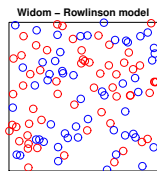
# Multi-type pairwise interaction processes



a)



b)



c)

**Figure:** Bivariate pairwise interaction processes with  $r = 0.05$  and : a)  $\gamma_{1,2} = \gamma_{2,1} = 1.0$ , b)  $\gamma_{1,2} = \gamma_{2,1} = 0.75$  and c)  $\gamma_{1,2} = \gamma_{2,1} = 0$ . Circles around the points have a radius of 0.025.

Widom-Rowlinson or penetrable spheres model : this model is described by the mark space  $M = \{1, 2\}$  and the density

$$p(x) = \alpha \prod_{(w,m) \in x} \beta_m \prod_{(u,1), (v,2) \in x} 1\{\|u - v\| > r\} \quad (7)$$

w.r.t the standard Poisson point process on  $W \times M$  with  $\nu_M(1) = \nu_M(2)$ .

The parameters  $\beta_1 > 0$  and  $\beta_2 > 0$  control the number of particles of type 1 and 2, respectively.

The conditional intensity for adding  $(w, 1) \notin x$  to the configuration  $x$  is

$$\lambda((w, 1); x) = \beta_1 1\{d(u, w) > r \text{ for all the } (u, 2) \in x\}.$$

A similar expression is available for adding an object of type 2.

The Widom-Rowlinson is hereditary and locally stable with

$$\Lambda = \max\{\beta_1, \beta_2\}.$$

Furthermore,  $\lambda((w, m); x') \geq \lambda((w, m); x)$  for all  $x' \subseteq x$  and  $(w, m) \in W \times M$ .

The interaction functions are

$$\begin{aligned}\phi_c(\emptyset) &= \alpha \\ \phi_c(\{(w, m)\}) &= \beta_m \\ \phi_c(\{(u, 1), (v, 2)\}) &= 1\{d(u, v) > r\}\end{aligned}$$

and  $\phi_c \equiv 1$  if the cliques have two or more objects of the same type.

**Multi-type pairwise interaction process** : consider  $M = \{1, \dots, I\}$  with  $I \in \mathbb{N}$  and  $\nu_M$  the uniform distribution on  $M$ . The probability density w.r.t the standard multi-type process is

$$p(x) = \alpha \prod_{(w,m) \in x} \beta_m \prod_{(u,i) \neq (v,j) \in x} \gamma_{ij}(d(u,v)). \quad (8)$$

- ▶ the parameters  $\beta_m > 0$ ,  $m \in M$  control the intensity of the points of type  $m$ .
- ▶ the measurable functions  $\gamma_{ij} : [0, \infty) \rightarrow [0, 1]$  describe the interaction between each type pair of objects  $i, j \in M$
- ▶ symmetric functions :  $\gamma_{ij} \equiv \gamma_{ji}$  for all  $i, j \in M$

For  $(w, m) \notin x$ , the conditional intensity is

$$\lambda((w, m); x) = \beta_m \prod_{(u, i) \in x} \gamma_{im}(d(u, w)).$$

This process is locally stable with  $\Lambda = \max_{m \in M} \beta_m$ , anti-monotonic and Markov under smooth assumptions on the functions  $\gamma_{ij}$ .

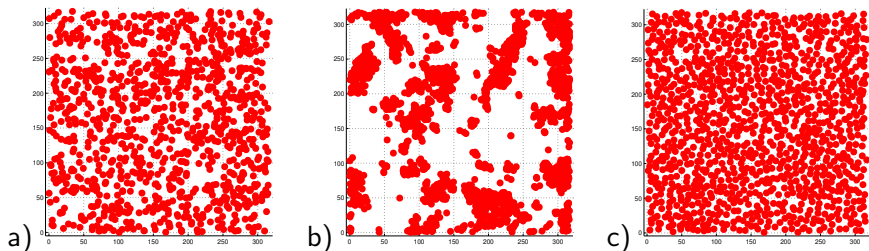
The interaction functions are

$$\begin{aligned}\phi_c(\emptyset) &= \alpha \\ \phi_c(\{(w, m)\}) &= \beta_m \\ \phi_c(\{(u, i), (v, j)\}) &= \gamma_{ij}(d(u, v))\end{aligned}$$

with  $\phi_c \equiv 1$  for cliques of three objects and more.

Area interaction model :  
(Baddeley and van Lieshout, 1995)

$$p(x) \propto \beta^{n(x)} \gamma^{-\nu[\Gamma(x)]}, \quad \beta, \gamma > 0 \quad (9)$$



**Figure:** Area interaction model realisations for different parameter values : a)  $\gamma = 1.0$ , b)  $\gamma > 1.0$  and c)  $\gamma < 1.0$ .

## Remarks :

- ▶ the first probability density based point process producing clusters → alternative to the Strauss process ...
- ▶ the model should be re-parametrized in order to be identifiable

## Proposition

*The area interaction process given by (9) is a Markov point process.*

## Proof.

→ Exercice 28 + Exercice 29



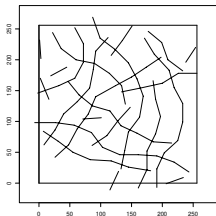


## Candy model :

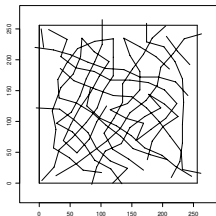
(van Lieshout and Stoica, 2003), (Stoica, Descombes and Zerubia, 2004)

$$p(\mathbf{x}) \propto \gamma_f^{n_f(\mathbf{x})} \gamma_s^{n_s(\mathbf{x})} \gamma_d^{n_d(\mathbf{x})} \gamma_o^{n_o(\mathbf{x})} \gamma_r^{n_r(\mathbf{x})},$$

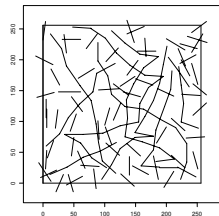
with  $\gamma_f, \gamma_s, \gamma_d > 0$  and  $\gamma_o, \gamma_r \in [0, 1]$



a)



b)



c)

Figure: Candy model realisations.

## Bisous model :

(Stoica, Gregori and Mateu, 2005)

$$p(x) \propto \left[ \prod_{s=0}^q \gamma_s^{n_s(x)} \right] \prod_{\kappa \in \Gamma \subset \mathcal{R}} \gamma_{\kappa}^{n_{\kappa}(x)} \quad \gamma_s > 0, \gamma_{\kappa} \in [0, 1]$$

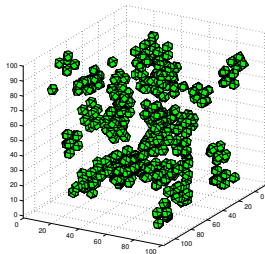
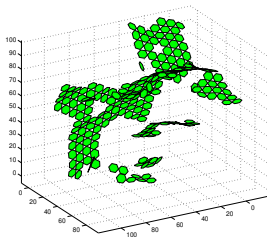
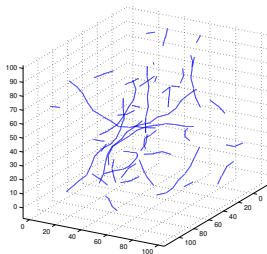


Figure: Random shapes generated with Bisous model.

## Remarks :

- ▶ Candy and Bisous are based on compound interactions → **drawing + explanations**
- ▶ connections are produced by giving different weights for the repulsive interactions
- ▶ the conditional intensity is bounded

$$\lambda(\zeta; \mathbf{x}) \leq \prod_{s=0}^q \max\{\gamma_s, \gamma_s^{-1}\}^{12} = \Lambda.$$

this gives the name of the model → **kissing number**

- ▶ → **blackboard - Candy**
- ▶ Markov range :  $4r_h + 2r_a$
- ▶ the models are locally stable but the exact simulation is sometimes difficult ...

Compare two random sets : idea inspired by work with M. N. M. van Lieshout and classical literature in mathematical morphology

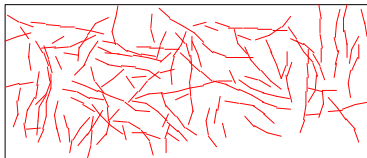
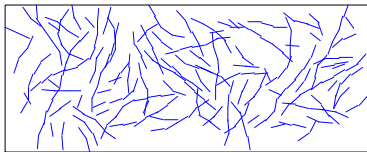


Figure: Realizations of the Candy model obtained with different samplers.

Empty space function : these probability distributions should be similar  $\Rightarrow$  Kolmogorov-Smirnov  $p$ -value is higher than 0.8

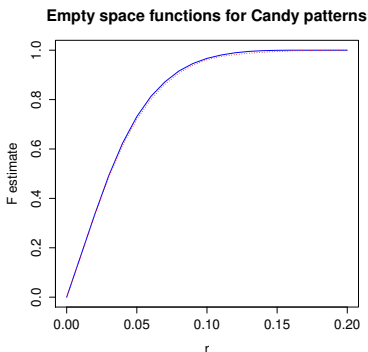


Figure: Estimation of the empty space function for the previous Candy realizations

## Exercises : Gibbs models

*Exercise 26.* Prove the Spatial Markov property Theorem.

*Exercise 27.* Prove the Hammersley - Clifford Theorem.

*Exercise 28.* Let us consider the area-interaction point process as defined in the course.

- a) Write its conditional intensity.
- b) Prove that the model is locally stable and for  $\beta, \gamma > 0$ .
- c) What kind of pattern is generated if  $\gamma < 1$ ,  $\gamma = 1$  and  $\gamma > 1$  ?
- d) Prove the model is Markov and exhibit its interaction functions. What order are these interactions ?

*Exercise 29.*

- a) Simulate attractive and repulsive area-interaction processes in  $W = [0, 1]^2$ . Use the R spatat function : `rmh`.
- b) To an attractive and repulsive pattern obtained previously fit the area-interaction process. What do you notice ? Are the differences between the theoretical and estimated parameters important ? Why ? Use the R spatat function : `ppm`.
- c) Consider now the data set `redwood`. After an exploratory analysis of the data set, fit an area-interaction process to it. Test the model using envelope tests. How do you interpret the results ? How do you compare these results with the ones obtained at the Exercice 22 ?

# Table of contents

## Cours 8. Markov chains. Definitions, properties, algorithms.

Monte Carlo simulation

Markov chains : a little bit of theory

Metropolis-Hastings algorithm

MH algorithm for sampling marked point processes

Perfect or exact simulation

Spatial birth-and-death processes

Exercises : Markov chains. Definitions, properties, algorithms.



# Markov chain Monte Carlo algorithms

**Problem** : sampling or simulation probability distributions

$$\pi(A) = \int_A p(x) d\mu(x)$$

that are not available in closed form  $\leftrightarrow$  normalising constant  
analytically intractable

# Basic MCMC algorithm

## Algorithm

$x = \text{My first MCMC sampler } (T)$

1. choose an initial condition  $x_0$

2. for  $i = 1$  to  $T$ , do

{  
     $x_i = \text{Update}(x_{i-1})$   
}

3. return  $x_T$ .

## Principles of the MCMC algorithm :

- ▶ simulates a Markov chain
- ▶ the *Update* function reproduces the transition kernel of the Markov chain
- ▶ the output  $x_T$  is asymptotically distributed according to  $\pi$  whenever  $T \rightarrow \infty$
- ▶ if the simulated Markov chain has good properties  $\rightarrow$  statistical inference is possible
- ▶ several solutions : Gibbs sampler, Metropolis-Hastings, birth and death processes, stochastic adsorption, RJMCMC, exact simulation (CFTP, clan of ancestors, etc.)

# Markov chains : a little bit of theory

Let  $(\Omega, \mathcal{F}, \mu)$  a probability space.

**Markov chain** : a sequence of random variables  $\{X_n\}$  such that :

$$\mathbb{P}(X_{n+1}|X_0, \dots, X_n) = \mathbb{P}(X_{n+1}|X_n)$$

The chain is **homogeneous** if the probabilities from going from one state to another do not change in time.

**Transition kernel** : a mapping  $P : \Omega \times \mathcal{F} \rightarrow [0, 1]$  such that

- ▶  $P(\cdot, A)$  is measurable for any  $A \in \mathcal{F}$
- ▶  $P(x, \cdot)$  is a probability measure on  $(\Omega, \mathcal{F})$  for any fixed  $x \in \Omega$

As on discrete spaces the  $n$ -step transition probability kernel is defined iteratively. Let  $P^0(x, A) = \delta_x(A)$  the Dirac measure defined by

$$\delta_x(A) = \begin{cases} 1 & x \in A \\ 0 & x \notin A, \end{cases}$$

and, for  $n \geq 1$  the  $n$ -step transition probability kernel is defined inductively

$$P^n(x, A) = \int_{\Omega} P(x, dy) P^{n-1}(y, A) \quad x \in \Omega, A \in \mathcal{F}.$$

**Theorem: Chapman-Kolmogorov equations.** For any  $m$  with  $0 \leq m \leq n$ ,

$$P^n(x, A) = \int_{\Omega} P^m(x, dy) P^{n-m}(y, A), \quad x \in \Omega, A \in \mathcal{F}.$$

**Proof:** See (Meyn and Tweedie 2009, Thm. 3.4.2 pp.61). It is easy to see that

$$\begin{aligned} P^n(x, A) &= \int_{\Omega} P(x, dy) P^{n-1}(y, A) \\ &= \int_{\Omega} P(x, dy) \int_{\Omega} P(y, dz) P^{n-2}(z, A) \\ &= \int_{\Omega} P^2(x, dz) P^{n-2}(z, A) \\ &\vdots \\ &= \int_{\Omega} P^m(x, dy) P^{n-m}(y, A). \end{aligned}$$

- ▶ in a very intuitive way, this result states that to get from  $x$  to  $A$  in  $n + m$  steps, the chain must pass through some  $y$  on the  $n$ th step
- ▶ in the discrete case the Chapman-Kolmogorov equations may be interpreted as a matrix product
- ▶ in the general case the kernel  $P^n$  is an operator acting on both bounded measurable functions and  $\sigma$ -finite measures  $\mu$  on  $\mathcal{F}$  via

$$P^n f(x) = \int_{\Omega} P^n(x, dy) f(y), \quad \mu P^n(A) = \int_{\Omega} \mu(dx) P^n(x, A).$$

**Definition: invariant measure.** A  $\sigma$ -finite measure on  $(\Omega, \mathcal{F})$  is invariant, if it verifies

$$\pi(A) = \int_{\Omega} \pi(dx)P(x, A), \quad A \in \mathcal{F}.$$

- ▶ if an initial condition is sampled according to the invariant distribution  $\pi$ , then the action of the transition kernel will produce new states always distributed according to  $\pi$
- ▶ simple verification:

$$\begin{aligned} \pi(A) &= \int_{\Omega} \pi(dx)P(x, A) = \int_{\Omega} \left[ \int_{\Omega} \pi(dy)P(y, dx) \right] P(x, A) \\ &= \int_{\Omega} \pi(dy) \left[ \int_{\Omega} P(y, dx)P(x, A) \right] = \int_{\Omega} \pi(dy)P^2(y, A) \\ &\quad \vdots \\ &= \int_{\Omega} \pi(dy)P^n(y, A), \end{aligned}$$

for any  $n$  and all  $A \in \mathcal{F}$ .



- ▶ now, consider any starting distribution  $\mu$
- ▶ if a limiting measure  $\gamma_\mu$  exists in a suitable topology on the space of probability measures, such as

$$\gamma_\mu(A) = \lim_{n \rightarrow \infty} \int_{\Omega} \mu(dx) P^{(n)}(x, A)$$

for sets  $A \in \mathcal{F}$

- ▶ then

$$\begin{aligned} \gamma_\mu(A) &= \lim_{n \rightarrow \infty} \int_{\Omega} \mu(dx) \int_{\Omega} P^{n-1}(x, dy) P(y, A) \\ &= \int_{\Omega} \gamma_\mu(dy) P(y, A), \end{aligned}$$

hence  $\gamma_\mu$  is an invariant probability measure.

- ▶ if the chain has an unique invariant measure  $\pi$
- ▶ then
  - ▶ the limit  $\gamma_\mu$  equals  $\pi$
  - ▶ it is independent of any initial condition  $\mu$
- ▶ in this case, the invariant measure is the **equilibrium distribution** of the chain

**Definition: reversibility.** A transition kernel  $P(x, A)$  is said to be reversible with respect to a measure  $\pi$ , if the integral

$$\int_A \int_B 1_A(x)1_B(y)\pi(dx)P(x, dy) \quad (10)$$

is symmetric under the interchange of  $A$  and  $B$ .

- ▶ applying the previous definition to the formula (10) leads directly to

$$\int_A \pi(dx)P(x, B) = \int_B \pi(dx)P(x, A)$$

- ▶ hence, the probability of going from  $A$  to  $B$  equals the probability of going backwards from  $B$  to  $A$ .

- ▶ if  $P(x, \Omega) = 1$  and  $B = \Omega$ , it follows immediately

$$\pi(A) = \int_{\Omega} \pi(dx) P(x, A),$$

hence, the reversibility of the transition kernel guarantees the invariance of the considered measure

- ▶ practical point of view: the `Update()` mechanism should reproduce a reversible transition kernel
- ▶ the reversibility condition alone  $\rightarrow$  only the construction of a sampling algorithm that needs the initial conditions to be chosen according to the distribution of interest
- ▶ problem solved: if the Markov chain is irreducible, aperiodic and recurrent

# Irreducibility

- ▶ crucial property in the setup of MCMC algorithms
- ▶ intuitive description: guarantees that from any initial point  $x \in \Omega$ , the chain may reach any region  $A \in \mathcal{F}$  of the configuration space
- ▶ first measure of the sensitivity of the Markov chain to the initial conditions
- ▶ it leads to convergence conditions of the simulated chain towards the desired equilibrium distribution.
- ▶ for presenting it, we need first the definition of the stopping time of the chain in a set  $A$ .

**Definition: stopping time at  $A$ .** For any set  $A \in \mathcal{F}$ , the quantity

$$\tau_A = \min\{n \geq 1 : X_n \in A\}$$

is called the stopping time at  $A$ . If  $X_n \notin A$  for every  $n$ , then by convention  $\tau_A = +\infty$ .

- ▶ discrete case: irreducibility means, that all the chain states communicate, that is

$$\mathbb{P}_x(\tau_y < \infty) > 0 \quad \forall x, y \in \Omega.$$

- ▶ continuous state space:  $\mathbb{P}_x(\tau_y < \infty) = 0$
- ▶ in order to correctly define irreducibility for general state spaces, an auxiliary measure  $\phi$  is needed

**Definition:  $\phi$ -irreducibility.** The Markov chain  $(X_n)$  is  $\phi$ -irreducible if there exists a measure  $\phi$  on  $\mathcal{F}$  such that, whenever  $\phi(A) > 0$ , we have  $\mathbb{P}_x(\tau_A < \infty) > 0$  for all  $x \in \Omega$ .

An equivalent formulation of the definition of the  $\phi$ -irreducibility is that for all  $x \in \Omega$ , whenever  $\phi(A) > 0$ , there exists  $n > 0$  such that  $P^n(x, A) > 0$ .

- ▶ a  $\phi$ -irreducible Markov chain is able to reach any set  $A$  which is “big enough”, independently of the initial condition
- ▶ in the following we will give the tools needed for establishing the irreducibility of a transition kernel

**Definition: small sets.** A set  $C$  is a small set if there exists  $n \in \mathbb{N}^*$  and a nonzero measure  $\nu_n$  such that

$$P^n(x, A) \geq \nu_n(A) \quad (11)$$

for all  $x \in C$  and  $A \in \mathcal{F}$ . Whenever the equation (11) is verified, the set  $C$  is called  $\nu_n$ -small.

**Theorem.** Let  $(X_n)$  be a  $\phi$ -irreducible chain. For every set  $A \in \mathcal{F}$  such that  $\phi(A) > 0$ , there exists  $n \in \mathbb{N}^*$  and a small set  $C \subset A$  such that  $\nu_n(C) > 0$ . Moreover,  $\Omega$  can be decomposed in a countable partition of small sets.

- ▶ proof: (Meyn and Tweedie, 2009) Prop. 5.2.4 pp.105-106
- ▶ the small sets are tool for "discretizing" a continuous space state



- ▶ communicating states (*i.e.* irreducibility): desirable property whenever building Markov chain for sampling probability distributions
- ▶ cyclic behaviour: naturally considered as undesirable

**Definition: aperiodic chain.** Let  $P(\cdot, \cdot)$  be a  $\phi$ -irreducible transition kernel. Suppose that there is a set  $A \in \mathcal{F}$ , a probability measure  $\nu$  with  $\nu(A) = 1$ , a constant  $\epsilon > 0$  and an integer  $n_0 \geq 1$  such that

$$P^{n_0}(x, \cdot) \geq \epsilon \nu(\cdot) \quad \forall x \in A.$$

The induced Markov chain is aperiodic if

$$\gcd\{m : \exists \epsilon_m > 0 \text{ such that } P^m(x, \cdot) \geq \epsilon_m \nu(\cdot) \quad \forall x \in A\} = 1.$$

- ▶ practical consequence: aperiodic Markov chains may be constructed by simply allowing the Update procedure to remain in the initial state with a positive probability

**Theorem.** Suppose the chain  $(X_n)$  has invariant probability measure  $\pi$ . Assume that the chain is  $\pi$ -irreducible and aperiodic. Then there is a set  $\Omega' \subseteq \Omega$  such that  $\pi(\Omega') = 1$  and

$$\sup_{A \in \mathcal{F}} |P^n(x, A) - \pi(A)| \rightarrow 0 \quad (12)$$

for each  $x \in \Omega'$ .

- ▶ proof: (Athreya et al. 1996), Thm. 1 pp. 72
- ▶ the results allows to build Update() procedures to sample from  $\pi$
- ▶ the result does not prevent the Markov chain of a set of configurations  $\Omega''$  with  $\pi(\Omega'') = 0$  for which the limit in (12) differs from zero

# Recurrence

- ▶ the recurrence is the Markov chain property that guarantees the non-existence of null-sets, such as  $\Omega''$ .
- ▶ it guarantees the independence of a MCMC sampling algorithm with respect to the initial conditions

**Definition: number of passages.** For any set  $A \in \mathcal{F}$ , the quantity

$$\eta_A = \sum_{n=1}^{\infty} 1_A(X_n)$$

is the number of passages of  $(X_n)$  in  $A$ .

**Definition: Harris recurrence.** A set  $A$  is Harris recurrent if  $\mathbb{P}_x(\eta_A = \infty) = 1$  for all  $x \in A$ . The chain  $(X_n)$  is Harris recurrent if it is  $\phi$ -irreducible and if any set  $A \in \mathcal{F}$  such that  $\phi(A) > 0$  is Harris recurrent.

- ▶ Harris recurrence is a “better” property than  $\phi$ -irreducibility
- ▶ in practice, the Update mechanisms that are  $\phi$ -irreducible, they are also Harris recurrent
- ▶ the following result presents: mathematical tool for proving Harris recurrence, based on the so-called drift condition

**Theorem.** Let  $(X_n)$  be a  $\phi$ -irreducible Markov chain and suppose there exist a small set  $C \in \mathcal{F}$  and function  $V : \Omega \rightarrow (0, \infty)$  such that the level sets

$$C_V(\alpha) = \{x \in \Omega : V(x) \leq \alpha\}$$

are small. The chain is Harris recurrent if the drift  $\Delta V(x) = PV(x) - V(x)$  is negative for any  $x \notin C$ , that is :

$$PV(x) = \int_{\Omega} P(x, dy)V(y) \leq V(x), \quad x \notin C.$$

- ▶ proof: (Meyn and Tweedie, 2009), Thm. 9.1.8 pp. 206.

# Ergodicity

- ▶ whenever a Harris recurrent Markov chain is simulated through an Update procedure, LLN and CLT can be used with the obtained samples, only after the chain reaches the equilibrium
- ▶ for the non-perfect MCMC algorithms, *i.e.* this is impossible
- ▶ the only thing we know: the chain should reach the equilibrium, but we do not know exactly when this will happen
- ▶ possible solution to this problem → build an ergodic chain
- ▶ no need to wait till equilibrium
- ▶ the only thing needed: enough samples in order to be able to apply the LLN or the CLT

**Definition: ergodicity.** The Markov chain  $(X_n)$  is ergodic if it is both Harris recurrent and aperiodic.

- ▶ for ergodic chains, (12) holds independently of the initial conditions
- ▶ the speed of convergence of the chain may be the same for all the initial conditions  $\rightarrow$  the chain is **uniformly ergodic**
- ▶ if the speed of convergence depends on the starting a point, we may have a **geometrically ergodic** chain

**Definition: total variation norm.** The total variation norm of a bounded signed measure  $\nu$  on  $(\Omega, \mathcal{F})$  is defined as

$$\|\nu\| = \sup_{A \in \mathcal{F}} \nu(A) - \inf_{A \in \mathcal{F}} \nu(A).$$

The total variation distance between two such measures  $\nu_1$  and  $\nu_2$  is  $\|\nu_1 - \nu_2\|$ .

**Theorem.** Let  $(X_n)$  be a Markov chain  $\phi$ -irreducible and aperiodic. The chain is geometrically ergodic if there exists a function  $V : \Omega \rightarrow [1, \infty)$ , constants  $b < \infty$  and  $a < 1$ , and a small set  $C \in \mathcal{F}$  such that

$$PV(x) \leq aV(x) + b1_C(x), \quad \forall x \in \Omega. \quad (13)$$

- ▶ proof: (Meyn and Tweedie, 2009), Thm. 15.0.1 pp. 363
- ▶ geometric ergodicity means that the iterations of the transition kernel approach the equilibrium distribution at a geometric speed
- ▶ it can be shown, that the sets  $C_V(\alpha)$  are small for any  $\alpha > 0$   
→ the geometric ergodicity drift condition implies the drift condition for Harris recurrence.



# Metropolis-Hastings algorithm

## Principle :

- ▶ consider the chain in the state  $x_i = x$
- ▶ propose a new state  $x_f = y$  using the proposal density  $q(x_i \rightarrow x_f)$
- ▶ accept this new state with probability

$$\alpha(x, y) = \min \left\{ 1, \frac{p(y)q(y \rightarrow x)}{p(x)q(x \rightarrow y)} \right\}$$

if not remain in the previous state

- ▶ iterate as many times as we need (... in theory till infinity ...)

## Properties

- ▶  $\alpha(\cdot, \cdot)$  is a solution of the detailed balance equation  $\rightarrow$  reversibility is preserved
- ▶ very few conditions are required for  $q(\cdot \rightarrow \cdot)$  so that the chain has all the convergence properties
- ▶  $q(\cdot \rightarrow \cdot)$  should be simple to calculate and to simulate
- ▶ the knowledge of the normalizing constant of  $p(\cdot)$  is not needed

$\rightarrow$  Exercise 30

# MH algorithm for sampling marked point processes

**Idea** : the transition kernel propose to add an object to the configuration with probability  $p_b$  or propose to delete an object from the configuration with the probability  $p_d$

**Birth** : add an object

- ▶ initial state :  $x_i = x$  an object configuration
- ▶ final state :  $x_f = x \cup \{\zeta\}$
- ▶ proposal density to add an object : choose uniformly its location in  $W$  and its mark independently according to  $\nu_M$

$$q(x_i \rightarrow x_f) = q(x \rightarrow x \cup \{\zeta\}) = p_b \frac{1\{\zeta_w \in W\}}{\nu(W)}$$

- ▶ proposal density to remove an object : choose uniformly an object from  $x \cup \{\zeta\}$

$$q(x_f \rightarrow x_i) = q(x \cup \{\zeta\} \rightarrow x) = p_d \frac{1_{\{\zeta \in x \cup \{\zeta\}\}}}{n(x) + 1}$$

- ▶ acceptance probability

$$\alpha(x \rightarrow x \cup \{\zeta\}) = \min \left\{ 1, \frac{p_d p(x \cup \{\zeta\})}{p_b p(x)} \times \frac{\nu(W)}{n(x) + 1} \right\} \quad (14)$$

Death : remove an object

- ▶ the inverse movement of birth
- ▶ acceptance probability

$$\alpha(x \rightarrow x \setminus \{\zeta\}) = \min \left\{ 1, \frac{p_b p(x \setminus \{\zeta\})}{p_d p(x)} \times \frac{n(x)}{\nu(K)} \right\} \quad (15)$$

A transition kernel doing these transformations is

$$\begin{aligned}
 P(x, A) = & p_b \int_K b(x, \eta) \alpha(x, y := x \cup \{\eta\}) \mathbf{1}\{y \in A\} d\sigma(\eta) \\
 & + p_d \sum_{\eta \in x} d(x, \eta) \alpha(x, y := x \setminus \{\eta\}) \mathbf{1}\{y \in A\} \\
 & + \mathbf{1}\{x \in A\} \left[ 1 - p_b \int_K b(x, \eta) \alpha(x, x \cup \{\eta\}) d\sigma(\eta) \right. \\
 & \left. - p_d \sum_{\eta \in x} d(x, \eta) \alpha(x, x \setminus \{\eta\}) \right],
 \end{aligned}$$

where  $K = W \times M$ ,  $d\sigma(\eta) = d\sigma((w, m)) = d\nu(w) \times d\nu_M(m)$  et  $0 < p_b + p_d \leq 1$ . The birth rate is  $b(x, \eta) = \frac{1}{\nu(W)}$  and the death rate is  $d(x, \eta) = \frac{1}{n(x)}$

## Remarks :

- ▶ the Papangelou intensity appears in the acceptance probability
- ▶ local stability property guarantees good convergence properties of the Markov chain
- ▶ → blackboard : discuss reversibility

## Algorithm

$y = \text{Update}(x)$

1. Choose “birth” or “death” with probabilities  $p_b$  and  $p_d$ , respectively.
2. If “birth” was chosen, then generate a new object following  $b(x, \eta)$ . Accept the new configuration,  $y = x \cup \{\eta\}$  with the probability  $\alpha(x, y)$  given by (14).
3. If “death” was chosen, then select the object to be removed using  $d(x, \eta)$ . Accept the new configuration,  $y = x \setminus \{\eta\}$  with the probability  $\alpha(x, y)$  given by (15).
4. Return the present configuration.



**Theorem.** Let be  $b, d$  and  $q$  as described previously. Assume that  $b(x, \eta)$  and  $d(x, \eta)$  are strictly positive on their corresponding definition domain, respectively, and

$$\lim_{n \rightarrow \infty} u_n = \lim_{n \rightarrow \infty} \left[ \sup_{\eta \in W \times M, x \in \Xi_n} \frac{d(x \cup \{\eta\}, \eta)}{b(x, \eta)} \right] \rightarrow 0.$$

Fix  $p_b, p_d \in (0, 1)$  with  $p_b + p_d \leq 1$  and let  $p(x)$  be the probability density of a marked point process on  $W \times M$ . The point process is locally stable and  $p(x)$  is built w.r.t the standard Poisson process  $\mu$ . The MH sampler defined previously simulates a Markov chain with invariant measure  $\pi = \int p d\mu$  who is  $\phi$ -irreducible, Harris recurrent and geometric ergodic.

→ proof: Exercise 31

Remark :

- ▶ the same result holds if change moves are introduced with care ... → explain ...

## Optimality of the MH dynamics

- ▶ theoretical convergence properties
- ▶ local computation
- ▶ no need of the normalising constant
- ▶ highly correlated samples : only one element changed per accepted transition
- ▶ allows improvements : transition kernels that “help” the model

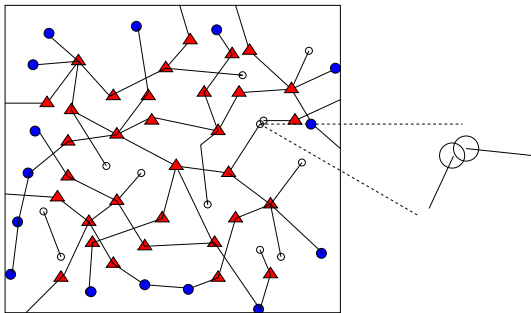
## Tailored to the model proposal distribution

$$b(x, \eta) = \frac{p_1}{\nu(K)} + p_2 b_a(x, \eta),$$

with  $p_1 + p_2 = 1$  and  $b_a(x, \eta)$  a probability density given by

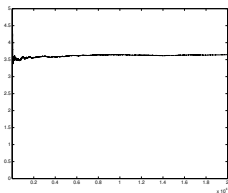
$$b_a(x, \eta) = \frac{1}{n(A(x))} \sum_{x \in A(x)} \tilde{b}(x, \eta).$$

- ▶ the role of  $b_a(x, \eta)$  : propose the birth of a new point in those regions where the interactions between the new born and the other configuration members is favoured or not penalised by the model
- ▶  $A(x)$  : the set of marked points in a configuration that are not exhibiting yet “good” interactions

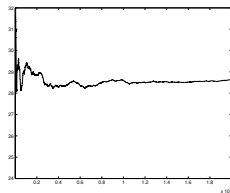


**Figure:** Extremities marked by triangles are connected and further than  $\frac{1}{2}l_{\max} + r_c$  to the boundary, those labeled by a black disk are closer than  $\frac{1}{2}l_{\max} + r_c$  to the boundary of  $K$ .

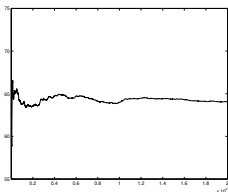
# MH algorithm for sampling the Candy model : dynamics behaviour through the sufficient statistics analysis



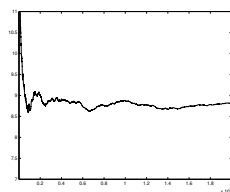
$$\bar{n}_f = 3.64$$



$$\bar{n}_s = 28.63$$



$$\bar{n}_d = 64.07$$



$$\bar{n}_o = 8.82$$

- ▶ great adaptability and theoretical convergence
- ▶ easy to use
- ▶ appropriate solutions need to be found for each new model
- ▶ the general framework, even if it has good theoretical properties, it is not always the most efficient from a numerical point of view
- ▶ still, if no good theoretical properties are available, the results will be always bad

→ Exercise 32

# Perfect or exact simulation

+ la schioapa sunt ...

- ▶ classical MCMC methods → need a theoretical infinite time till convergence
- ▶ dependence on the initial conditions
- ▶ perfect or exact MCMC methods indicate by themselves whenever the convergence is attained
- ▶ these methods are perfect within the limits of the random number generators of the computers
- ▶ historical paper : (Propp and Willson, 1996)

## Principle : discrete case

Let us build a MCMC sampler for  $\pi$  defined on the discrete state space  $\Omega = \{\omega_1, \omega_2, \dots, \omega_m\}$ . The induced Markov chain  $(X_n)$  is represented by its transition functions  $\phi(\cdot, \cdot)$  such that

$$X_{n+1} = \phi(X_n, V_n), \quad (16)$$

where  $V_n$  are i.i.d random variables.



## Key idea :

- ▶ consider  $m$ ,  $(X_n(\omega_i))$  all initialised with a different state, that evolve from  $-T < 0$  till 0
- ▶ the chains are coupled : they use the same  $V_n$ s
- ▶ if at a certain moment  $n \in -T, \dots, 0$  all the chains are in the same state or they **coalesced**, that is

$$X_n \equiv x$$

then they will all remain in the same state, till the time 0

- ▶ the influence of the initial conditions just ... **vanished**
- ▶ if the chains are started before the time  $-T$ , at infinite, the chains will be all in the same state, at the same moment
- ▶ it comes out that  $X_0$  is a perfect sampler from the equilibrium distribution  $\pi$

→ blackboard : drawing

Extraordinary smart idea :

- ▶ launching  $m$  parallel chains is not always feasible
- ▶ if  $\Omega$  can be ordered

$$\omega_{\min} = \omega_{(1)} < \omega_{(2)} < \dots < \omega_{(m)} = \omega_{\max}$$

and if the transition kernel respect this order relation

$$\omega \leq \omega' \quad \Rightarrow \quad \phi(\omega) \leq \phi(\omega')$$

then only the states  $\omega_{\min}$  and  $\omega_{\max}$  are needed

- ▶ the behaviour of the other chains is bounded by the extremal chains  $X_n(\omega_{\min})$  and  $X_n(\omega_{\max})$

This idea is known under the name **Coupling From The Past (CFTP)**.

# Spatial birth-and-death processes

**Theoretical background** : continuous time Markov chains → the very nice book of S. Resnick (2005)

- ▶ history : the first MCMC sampler for marked point processes
- ▶ the simulation of locally stable marked point process is rather simple : thinning procedure
- ▶ the simulated pattern is “hidden” in a dominating Poisson process

## Algorithm

Let  $p$  be the probability density w.r.t. the standard Poisson process, of a locally stable marked point process on  $W \times M$ . Its corresponding Papangelou conditional intensity bound is  $\Lambda$ .

Set  $X^{(0)} = x_0$  for some configuration  $x_0 \in \Omega$  with  $p(x_0) > 0$ . For  $n = 0, 1, \dots$ , if  $X^{(n)} = x$ , do :

- ▶ the sojourn time  $T^{(n)}$  in the state  $X^{(n)}$  is exponentially distributed with mean  $\frac{1}{n(x) + \Lambda\nu(W)}$
- ▶ the next transition is a death with probability  $\frac{n(x)}{n(x) + \Lambda\nu(W)}$ , obtained by selecting one point  $\eta$  from the current configuration with probability  $\frac{1}{n(x)}$ , and then deleting it, that is  $x = x \setminus \{\eta\}$
- ▶ with probability  $\frac{\Lambda\nu(W)}{n(x) + \Lambda\nu(W)}$  the next transition is a birth, obtained by generating a new point  $\eta$  according to the probability density  $\frac{1}{\nu(W)}$  and accepting it with probability  $\frac{\lambda(\eta; x)}{\Lambda}$ , hence  $x = x \cup \{\eta\}$ ,

## Remarks :

- ▶ this algorithm has equivalent convergence properties compared to MH algorithm (van Lieshout 2000, Thm. 3.3 pp. 86; Møller and Waagepetersen 2004, Prop. G7 pp. 276)
- ▶ nevertheless, for strong interactions  $\Lambda$  may be very high
- ▶ Candy and Bisous models with strong interactions cannot be simulated with this algorithm
- ▶ but ... **extremely important tool** in building perfect simulation algorithms

# Perfect algorithms for sampling marked point processes

## Coupling From The Past algorithms :

- ▶ dominating process :
  - ▶ stationary Poisson  $\rightarrow$  (CFTP and clan of ancestors),  $\Lambda$  parameter is important
  - ▶ it runs into the past, and then into the future till time 0 : doubling time scheme till coalescence
- ▶ the coupling and transition function ingredients : the  $V_{ns}$ , conditional intensity and  $\Lambda$
- ▶ spatial birth-and-death based algorithms : CFTP and clan of ancestors
- ▶ Metropolis-Hastings based algorithms : different dominant process + control of the birth and death proposal probabilities
- ▶ apply to locally stable point processes
- ▶ monotonic and anti-monotonic point processes

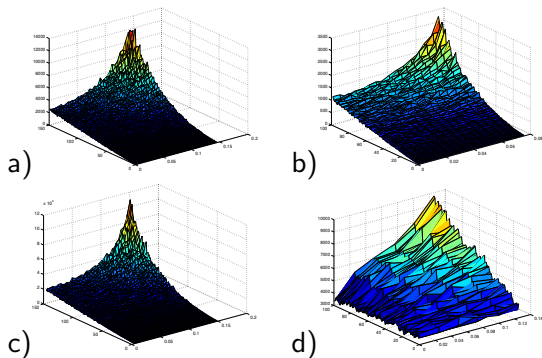
## Gibbs sampler algorithm :

- ▶ discrete probability densities  $\rightarrow$  approximation of the continuous marked point processes probability density
- ▶ does not require : order relation, dominating process, monotonic or anti-monotonic relation
- ▶ still the algorithm is more efficient if these properties are exhibited by the considered model
- ▶ Potts like models

## Remarks :

- ▶ CFTP algorithm is implemented within the **spatstat** package
- ▶ all these algorithms are implemented within the **MPPLIB** C++ package

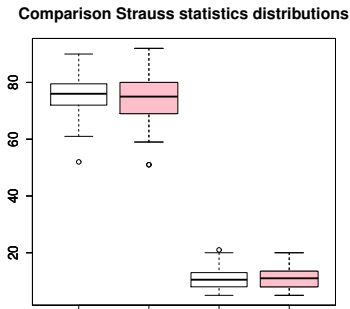
## Strauss model : convergence speed for exact sampling methods (van Lieshout and Stoica, 2006)



**Figure:** Exact simulation algorithms applied to Strauss model : a) CFTP, b) clan of ancestors, c) Metropolis-Hastings and d) Gibbs.



# Comparison exact simulation and MH algorithm for the Strauss model (1)



**Figure:** Boxplots comparison for the  $n$  and  $s_r$  statistics distributions : white - the distributions obtained using the exact algorithm, pink (dark coloured) - the distributions obtained using the Metropolis - Hastings algorithm.

# Comparison exact simulation and MH algorithm for the Strauss model (2)

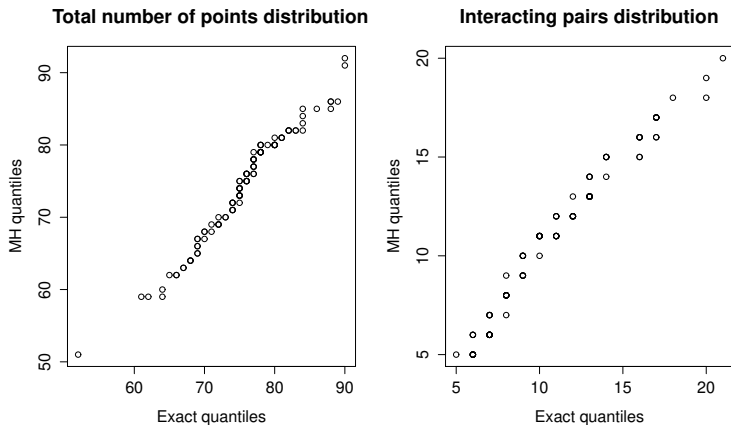


Figure: Qqplot comparison for the  $n$  and  $s_r$  statistics distributions.

→ Exercise 33

# Open questions MCMC methods :

## Classical algorithms - MH based dynamics

- ▶ good convergence properties but convergence at infinity
- ▶ burning-in methods + de-correlation techniques
- ▶ great adaptability : tailored to the model moves
- ▶ manipulate several objects during one move : work of X. Descombes
- ▶ link with RJMCMC : great adaptability, but difficult to state convergence proofs, hence difficult to use ...

## Perfect simulation algorithms

- ▶ the simulated chain indicates by itself whenever convergence is reached
- ▶ parameter dependence : can be applied in practice only to a restricted range of parameters
- ▶ neither change moves, nor tailored moves
- ▶ study existing algorithm : Fill algorithm, forward simulation and simulated tempering
- ▶ challenging perspective : synthesis of both families of algorithms → exact algorithms able to be tailored to the model

## Exercises : Markov chain Monte Carlo algorithms

*Exercise 30.* Look at the following code lines and explain the role of each function :

```
algo.mh = function(x0,n)
{x=x0;
  for(i in 1:n)
    { y=q.prop(x);
      a.ratio=(p.density(y)*q.density(y,x))/
              (p.density(x)*q.density(x,y));
      u=runif(1,0,1);
      if(u<=a.ratio){ x=y; }}
  x;
}
```

```
q.prop = function (x)
{
  delta=0.1;
  lim=0.5*delta;
  res=runif(1,x-lim,x+lim);
}
```

```
q.density = function (x,y)
{
  delta=0.1;
  lim=0.5*delta;
  res=dunif(y,x-lim,x+lim);
}
```

```
p.density = function (y)
{
  d1=100;
  d2=100;
  if(y>=0)
    { res=df(y,d1,d2); }
  else
    { res=0; }
}
```

```
x0=0.5;
m=10;
n=1000;
x=1:n;
for (i in 1:n)
{
  x[i]=algo.mh(x0,m);
  x0=x[i];
}
```

Using the previous code, answer the following questions :

- a) Simulate  $n = 1000$  random variables distributed according to a Fisher distribution  $F(\nu_1, \nu_2)$  of parameters  $\nu_1 = \nu_2 = 100$ .
- b) Plot the histogram of the obtained values. On the same plot, add the theoretical density. Plot the empirical cross-correlation function.
- c) For reducing the correlation of the samples obtained using the Metropolis-Hasting algorithm, one common techniques is to separate the samples; Re-do this exercice, by taking the samples every  $m = \{1, 5, 10, 50, 100\}$  iterations. Interpretation of the obtained results.



- d) Repeat the exercise for  $\delta = 0.001, 1.0, 100$ . If needed the very first samples given by the algorithm can be dropped of. This is called *burn-in* time. Try different values for it.
- e) Simulate  $n = 1000$  random variables following a Fisher distribution with parameters  $\nu_1 = \nu_2 = 100$ . In this case, we should obtain  $E[X] = 1.02$ . What is the value of the variance we should obtain ? Build confidence intervals for the samples mean. Make a statistical test to verify that the simulated variables have the desired theoretical mean.
- f) Plot the evolution of the mean, the confidence intervalls, the  $p$ -values and the empirical levels for the confidence intervals and the tests, while the number of samples increases.
- g) Suppose that we do not know what is the sampled distribution, build a test to verify the values obtained from the empirical mean. As before, build and plot confidence intervals, statistical tests and empirical levels for the samples mean.

*Exercise 31.* Prove the convergence of the MH sampler for marked point processes.

*Exercise 32.* Look at the following code lines and explain the role of each function :

```
mo=list(cif="strauss",par=list(beta=100,gamma=0.1,r=0.1),
        w=square(1))
X=rmh(model=mo,start=list(n.start=100),
      control=list(nrep=10000,nsave=100,nburn=100,
                  track=TRUE))
plot(X,cols="blue",main="Strauss model")
```

- a) The previous setting of the MH dynamics saves all the intermediate patterns. To have access to these patterns use :

```
patternX=attr(X,"saved")
```

Make a plot of the evolution of the sufficient statistics of the model.

- b) If  $\gamma = 1$  in a Strauss model a Poisson point process is simulated. Use the sufficient statistics plots in order to find a parameter setting of the MH dynamics that may indicate convergence of the algorithm.
- c) Simulate different realisations of the Strauss model for different model parameters. For  $\beta = 100$  and  $\gamma = 1$  build confidence intervals and statistical tests in order to verify the obtained values for the mean of the sufficient statistics.
- d) Repeat the previous question but for  $\beta = 50$ ,  $\gamma = 0.5$  and  $r = 0.1$ .

*Exercise 33.* Consider a Strauss process in  $W = [0, 1]^2$  with the following parameters  $\beta = 100, \gamma = 0.25$  and  $r = 0.005$

- a) Simulate 1000 samples using a Metropolis-Hastings dynamics and plot the time series evolution and diagnostics corresponding to the sufficient statistics.
- b) Simulate 1000 samples using the CFTP dynamics and plot the time series evolution and diagnostics corresponding to the sufficient statistics.
- c) Compare the sufficient statistics samples using exploratory analysis tools (boxplots, qqplots, histograms) and also statistical tests (Kolmogorov-Smirnoff).
- d) How many samples do we need in order to obtain Metropolis-Hastings based patterns that may be considered “good enough” ?

# Table of contents

## Cours 9. Parameter estimation. Model validation.

Statistical inference problems

Parameter estimation based on pseudo-likelihood

Monte Carlo Maximum likelihood estimation

Parameter estimation using the ABC Shadow algorithm

Model validation : residual analysis for point processes

Exercises : Parameter estimation. Model validation.

# Parameter estimation and model validation : the statistical inference problems

## Problem I : parameter estimation

- ▶ observe the pattern  $x$  and find the model parameters  $\theta$  able to statistically reproduce it
- ▶ complete and incomplete data : pseudo-likelihood, Monte Carlo maximum likelihood, EM, ABC Shadow
- ▶ open problem : sampling  $p(\theta|x)$

## Problem II : pattern detection

- ▶ observe the data  $d$  and find  $x$  “hidden”
- ▶ the model parameters are : hidden, modelled, unknown
- ▶ open problem : the does the detected pattern really exist ?

**Problem III** : sample the joint law of the pattern and the parameters  $p(x, \theta)$

- ▶ shape modelling → “crystal ball” ?
- ▶ look at a phenomenon and propose a model providing statistical prediction and validation of the observed pattern
- ▶ needs also the time dimension
- ▶ open problem : how to introduce time
- ▶ what it is the time “quanta” → generating element and interactions as for marked point processes → a twisted spaghetti - see pasta geometry of Max Tegmark ?

## Parameter estimation based on pseudo-likelihood

The pseudo-likelihood of a marked point process  $\mathbb{X}$  with conditional intensity  $\lambda_\theta(\zeta; \mathbf{x})$  observed on the bounded set  $W$  is expressed as

$$\begin{aligned} PL_W(\theta; \mathbf{x}) &= \\ &= \left[ \prod_{x_i \in \mathbf{x}} \lambda_\theta(x_i; \mathbf{x}) \right] \exp \left[ - \int_{W \times M} \lambda_\theta((w, m); \mathbf{x}) \nu(dw) \nu_M(dm) \right]. \end{aligned}$$

→ blackboard : construction of the pseudo-likelihood

The pseudo-likelihood estimator is given by the solution of the equation :

$$\frac{\partial PL_W(\theta; \mathbf{x})}{\partial \theta} = 0$$



## Remarks :

- ▶ the  $PL$  is concave for exponential models
- ▶ no normalising constant needed ...
- ▶ it "amplifies" the interaction weights : check the formula - for a Strauss process the interactions are counted twice ...
- ▶ consistency and asymptotic normality of the estimator : if we observe the model in a finite window, then it converges towards the parameters estimated using the pseudo-vraisemblance based on the observation of the "whole" window (Jensen and Møller, 1991)

- ▶ it can be used to have a “good” initial condition for other more elaborate methods
  - ▶ “good” results for mild interactions : (Mateu and Montes, 2001)
- ▶ **lacks of statistical significance** : there is no real link with the true model behind the pattern
  - ▶ **except for the Poisson process : in this case the pseudo-likelihood is the true likelihood**
- ▶ easy to be implemented :
  - ▶ this was the motivation to introduce it in the middle of 70s (Besag, 1975)
  - ▶ see (Baddeley, Rubak and Turner, 2016) for implementation details within the **spatstat** package

→ Exercise 34 + Exercise 35

## Implementation within R spatstat package

- ▶ stationary Strauss process :

$$\log \lambda_{\theta}(\zeta; \mathbf{x}) = \log \beta + (\log \gamma)t(\zeta, \mathbf{x})$$

with  $t(\zeta, \mathbf{x})$  the number of pairs of objects closer than the distance  $r$  in the configuration  $\mathbf{y}$

- ▶ general structure of the conditional intensity

$$\log \lambda_{\theta}(\zeta; \mathbf{x}) = \eta S(\zeta) + \phi V(\zeta, \mathbf{x}),$$

with the 'first order term'  $S(u)$  that describes spatial inhomogeneity and/or covariates effects and the 'higher order' term that describes interobject interaction

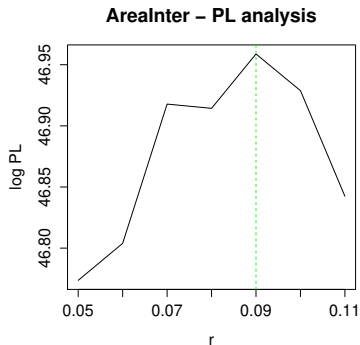
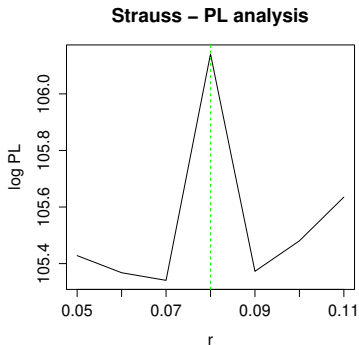
- ▶ refer to the **spatstat** documentation

→ caution : definition of the model ...

# Applications

**Pseudo-likelihood profile analysis** : the range parameters

```
>radius = data.frame(r=seq(0.05,0.11, by=0.01))  
>pradius = profilepl(radius, Strauss, japanesepines)  
>plot(pradius,main="Strauss : PL analysis")
```



Fitting the model to the pattern :

```
> ppm(japanesepines, 1, Strauss(r=0.08), rbord=0.08)
```

```
Stationary Strauss process
```

```
First order term:  beta 77.93567
```

```
Interaction:  Strauss process interaction distance:  
0.08
```

```
Fitted interaction parameter gamma:  0.7953
```

# Monte Carlo Maximum likelihood estimation

Exponential family models :

- ▶ very general framework
- ▶ the point processes models that were presented are given by

$$p(\mathbf{x}|\theta) = \frac{h(\mathbf{x}|\theta)}{Z(\theta)} = \frac{\exp\langle t(\mathbf{x}), \theta \rangle}{Z(\theta)}$$

where  $h(\mathbf{x}|\theta)$ ,  $t(\mathbf{x})$  and  $\theta$  represent the un-normalized probability density w.r.t. the standard Poisson process, the sufficient statistics vector and the model parameters vector, respectively. The normalising constant  $Z(\theta)$  is unknown.

→ blackboard : general notions about the exponential family models

The configuration  $x$  is **entirely** observed, hence the log-likelihood with respect a known parameter  $\psi$  can be written as follows :

$$l(\theta) = \langle t(x), \theta - \psi \rangle - \log \frac{Z(\theta)}{Z(\psi)}$$

It is easy to check, that the normalizing constants ratio is

$$\frac{Z(\theta)}{Z(\psi)} = \mathbb{E} [\exp \langle t(\mathbb{X}), \theta - \psi \rangle],$$

since we have

$$\begin{aligned}\frac{Z(\theta)}{Z(\psi)} &= \frac{1}{Z(\psi)} \int_{\Omega} h(\mathbf{x}|\theta) d\mu(\mathbf{x}) \\ &= \frac{1}{Z(\psi)} \int_{\Omega} h(\mathbf{x}|\theta) \frac{h(\mathbf{x}|\psi)}{h(\mathbf{x}|\psi)} d\mu(\mathbf{x}) \\ &= \int_{\Omega} \frac{h(\mathbf{x}|\theta)}{h(\mathbf{x}|\psi)} \frac{h(\mathbf{x}|\psi)}{Z(\psi)} d\mu(\mathbf{x}) \\ &= \mathbb{E} \left[ \frac{h(\mathbb{X}|\theta)}{h(\mathbb{X}|\psi)} \right]\end{aligned}$$



The Monte Carlo approximation of the normalizing constants ratio gives :

$$\frac{Z(\theta)}{Z(\psi)} \approx \frac{1}{n} \sum_{i=1}^n \exp\langle t(\mathbb{X}_i), \theta - \psi \rangle,$$

where  $\mathbb{X}_1, \mathbb{X}_2, \dots, \mathbb{X}_n$  are samples obtained from  $p(y|\psi)$ .

Hence, the Monte-Carlo counterpart of the log-likelihood is :

$$l_n(\theta) = \langle t(x), \theta - \psi \rangle - \log \left( \frac{1}{n} \sum_{i=1}^n \exp\langle t(\mathbb{X}_i), \theta - \psi \rangle \right).$$

## Theorem

*The log-likelihood of an exponential family model is a convex function.*

- ▶ proof : see (Monfort 1997, Thm.3, pp. 61)
- ▶  $l_n(\theta) \rightarrow l(\theta)$  almost surely
- ▶ all these suggest that local optimisation procedures applied to  $l_n(\theta)$  may give interesting results

→ Exercise 36

# MCMC local optimisation procedures

The gradient of the MCMC log-likelihood is

$$\nabla l_n(\theta) = t(\mathbf{x}) - \mathbb{E}_{n,\theta,\psi}[t(\mathbb{X})]$$

where

$$\mathbb{E}_{n,\theta,\psi}[t(\mathbb{X})] = \frac{\sum_{i=1}^n t(\mathbb{X}_i) \exp\langle t(\mathbb{X}_i), \theta - \psi \rangle}{\sum_{i=1}^n \exp\langle t(\mathbb{X}_i), \theta - \psi \rangle}$$

that is the **Monte Carlo importance sampling** approximation of  $\mathbb{E}_\theta t(\mathbb{X})$ .

Similarly, the Hessian can be computed too :

$$-\nabla^2 l_n(\theta) = \text{Var}_{n,\theta,\psi}[t(\mathbb{X})]$$

where

$$\text{Var}_{n,\theta,\psi}[t(\mathbb{X})] = \mathbb{E}_{n,\theta,\psi}[t(\mathbb{X})t(\mathbb{X})^t] - \mathbb{E}_{n,\theta,\psi}[t(\mathbb{X})]\mathbb{E}_{n,\theta,\psi}[t(\mathbb{X})^t].$$

Newton-Raphson method :

$$\theta_{k+1} = \theta_k - [\nabla^2 l_n(\theta_k)]^{-1} \nabla l_n(\theta_k) \quad (17)$$

for  $k = 1, 2, \dots,$

- ▶  $l_n(\cdot)$  is computed using  $n$  samples from  $p(x|\psi)$
- ▶ the computation of the gradient and Hessian inverse is numerically unstable
- ▶ useful only if the initial value is close enough from the solution

Iterative gradient method :

$$\begin{cases} l_n(\theta_k + \rho(\theta_k)\nabla l_n(\theta_k)) = \max_{\rho \in \mathbb{R}} l_n(\theta_k + \rho\nabla l_n(\theta_k)) \\ \theta_{k+1} = \theta_k + \rho(\theta_k)\nabla l_n(\theta_k) \end{cases}$$

where  $\rho(\theta_k)$  is the optimal step (Descombes et al. '99, Stoica '01).

- ▶ re-sampling if  $\| \theta_k - \psi \| > \text{threshold}$
- ▶ obtain a reference value  $\theta_0$  close enough to the maximum likelihood estimator

## Stochastic gradient :

$$\theta_{k+1} = \theta_k + \epsilon_k [t(x) - t(\mathbb{X}_k)]$$

where  $\epsilon_k > 0$  is a decreasing sequence while  $\mathbb{X}_k$  is a sample of  $p(x|\theta_k)$

- ▶ very simple, but finding an optimal sequence  $\{\epsilon_k\}$  is an open problem
- ▶ L. Younes, G. Winkler : Markov random fields
- ▶ R. Moyeed and A. Baddeley : point processes

## Asymptotic results MCMCML estimation

The random variable  $\sqrt{n}(\hat{\theta}_n - \hat{\theta})$  whenever  $n \rightarrow \infty$ , it converges in distribution towards a normal random variable of zero mean and variance  $I(\hat{\theta})^{-1}\Gamma I(\hat{\theta})^{-1}$  :

$$\sqrt{n}(\hat{\theta}_n - \hat{\theta}) \rightarrow \mathcal{N}(0, I(\hat{\theta})^{-1}\Gamma I(\hat{\theta})^{-1}).$$

- ▶ the matrix

$$I(\hat{\theta}) = \text{Var}_{\hat{\theta}}[t(\mathbb{X})] = -\nabla^2 I(\hat{\theta})$$

is the Fisher information of  $\hat{\theta}$

- ▶ the matrix  $\Gamma$  is the matrix of the asymptotic covariance of the normalised Monte Carlo gradient  $\sqrt{n}\nabla I_n(\hat{\theta})$

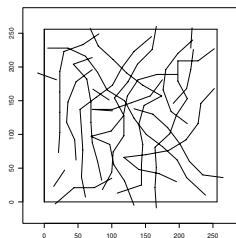


- ▶ the variance of the components of  $\hat{\theta} - \theta_0$  can be estimated by taking the diagonal elements of the inverse of  $-\nabla^2 l_n(\hat{\theta}_n)$ 
  - ▶ it represents the error between the maximum likelihood estimate and the true model parameters
- ▶ the variance of the components of  $\sqrt{n}(\hat{\theta}_n - \hat{\theta})$  can be estimated by taking the diagonal elements of  $I(\hat{\theta})^{-1}\Gamma I(\hat{\theta})^{-1}$ 
  - ▶ it represents the error between the maximum likelihood estimate and its Monte Carlo counterpart
- ▶ refer to (Monfort, 1997), (Geyer, 1999) and (van Lieshout and Stoica, 2003) for the computation of these matrices

→ show the next example + blackboard : incomplete data observation - EM algorithm

# MCML example

Candy model : (van Lieshout and Stoica, 2003)



Model parameters
$\theta_f = -8.5$
$\theta_s = -3.0$
$\theta_d = 2.5$
$\theta_o = -2.5$
$\theta_r = -2.5$

Sufficient statistics
$n_f = 4$
$n_s = 34$
$n_d = 63$
$n_o = 12$
$n_r = 9$

**Figure:** Realization (left) of the reference model given by the parameters in the middle table. The observed values of the sufficient statistics are listed at right.

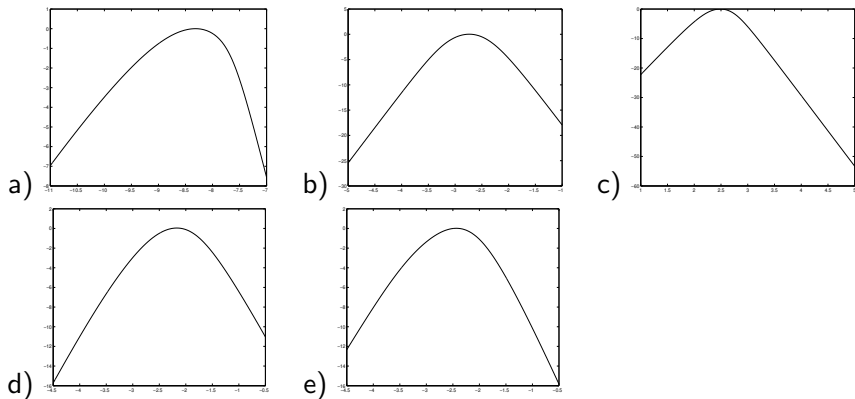
**Results** : estimation of the parameters from the reference configuration given by the Candy model

Initial parameters	Iterative method	Monte Carlo MLE
$\theta_f^i = -9.5$	$\hat{\theta}_f^0 = -8.37$	$\hat{\theta}_f^n = -8.32$
$\theta_s^i = -4.0$	$\hat{\theta}_s^0 = -2.74$	$\hat{\theta}_s^n = -2.73$
$\theta_d^i = 1.5$	$\hat{\theta}_d^0 = 2.46$	$\hat{\theta}_d^n = 2.47$
$\theta_o^i = -3.5$	$\hat{\theta}_o^0 = -2.13$	$\hat{\theta}_o^n = -2.17$
$\theta_r^i = -3.5$	$\hat{\theta}_r^0 = -2.42$	$\hat{\theta}_r^n = -2.42$

**Asymptotics** : estimation errors (central limit theorems available)

Asymptotic standard deviation of MLE	MCSE
0.51	0.002
0.23	0.003
0.17	0.001
0.30	0.002
0.31	0.005

## Log-likelihood ratio approximation :



**Figure:** Monte Carlo approximation of the log likelihood function. The X axis represents the variation of a single component. The Y axis represents the values of the Monte Carlo log likelihood with all other components of  $\hat{\theta}^0$  fixed : a -  $\theta_f \in [-11, -7]$ , b -  $\theta_s \in [-5, -1]$ , c -  $\theta_d \in [1, 5]$ , d -  $\theta_o \in [-4.5, -0.5]$ , e -  $\theta_r \in [-4.5, -0.5]$ .

# Parameter estimation using the ABC Shadow algorithm

Parameter estimation : the pattern detection dual problem

- ▶ let  $p(\theta|y)$  be the conditional distribution of the model parameters given the pattern

$$p(\theta|y) = \frac{\exp[-U(y|\theta)]p(\theta)}{Z(y)c(\theta)}$$

with  $p(\theta)$  the prior density for the model parameters and  $Z(y)$  the normalization constant

- ▶ the posterior law is defined on  $\Theta$  a compact region in  $\mathbb{R}^m$  ; the parameter space  $\Theta$  is endowed with the Borel algebra  $\mathcal{T}_\Theta$

- ▶ the parameter estimator is

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \{p(\theta|y)\}$$

- ▶ optimisation procedure : requires sampling  $p(\theta|y)$
- ▶ sampling the posterior law is not straightforward  $\rightarrow$  requires the evaluation of the ratio  $c(\theta)/c(\psi)$
- ▶ if  $p(\theta)$  is the uniform distribution over  $\Theta$  then  $\hat{\theta}$  is the maximum likelihood estimator

**Remark :** the model parameters taken into account by the posterior distribution are the “interaction” parameters. And not the “range” parameters ...

# Sampling posterior laws : theoretical solution

**Auxiliary variable method** : given by (Møller, Pettitt, Reeves and Berthelsen, 2006)

- ▶ principle : use an auxiliary variable  $\mathbb{X}$  defined by the probability density  $a(x|\theta, y) \rightarrow$  sample the joint distribution

$$\pi(\theta, x|y) = a(x|\theta, y)p(\theta|y) = a(x|\theta, y) \frac{\exp[-U(y|\theta)]p(\theta)}{Z(y)c(\theta)}$$

- ▶ present context : the auxiliary variable  $\rightarrow$  auxiliary pattern ...
- ▶ **key idea** : the final mathematical construction leads to the simplification of the normalizing constants ratio in the MH algorithm sampling  $\pi(\theta, x|y)$

→ blackboard: explain the algorithm

Remarks :

- ▶ very elegant solution
- ▶ exact sampling of the auxiliary variable
- ▶ choice of the auxiliary variable density : behaviour of the simulated chain ...



# Sampling posterior laws : approximate solution

## Approximate Bayesian Computation

- ▶ generic name for numerical simulation methods allowing model selection based on an approximate sampling from the posterior distribution  $p(\theta|y)$
- ▶ idea originated in Montpellier → statistics and environmental sciences communities
  - ▶ (Grelaud, Robert, Marin, Rodolphe and Taly, 2009)
  - ▶ (Marin, Pudlo, Robert and Ryder, 2012)

**Algorithm ABC** : assume the observed pattern is  $y$ , fix a tolerance threshold  $\epsilon$  and an integer value  $n$ .

1. For  $i = 1$  to  $n$  do

- ▶ Generate  $\theta_i$  according to  $p(\theta)$ .
- ▶ Generate  $x_i$  according to the probability density

$$p(x|\theta_i) = \frac{\exp[-U(x|\theta_i)]}{c(\theta_i)}$$

2. Return all the  $\theta_i$ 's such that the distance between the statistics of the observation and those of the simulated pattern is small, that is

$$d(t(y), t(x_i)) \leq \epsilon$$

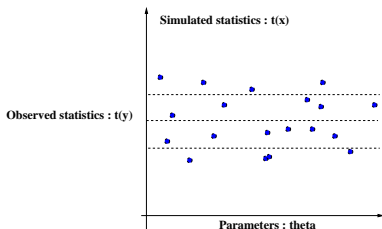


Figure: Graphical representation of the outputs of an ABC algorithm.

## Theoretical result

- ▶ (Blum, 2009) : gives the bias and the variance of the posterior distribution estimate
- ▶ (G. Biau, F. Cérous and A. Guyader , 2015) : give asymptotic features of the outputs of a slightly different algorithm
- ▶ ideas : kernel and  $k$ -nearest neighbour estimation

## Remarks :

- ▶ exact sampling from  $p(x|\theta)$  is needed
- ▶ choice of the statistics vector
  - ▶ exponential family models  $\rightarrow$  the sufficient statistics
- ▶ appropriate setting :
  - ▶ distance  $d$
  - ▶ precision parameter  $\epsilon$
  - ▶ number of neighbours  $k_n$  and bandwidth parameter  $h_n$

## Synthesis :

- ▶ ABC algorithms are useful if enough samples  $x_i$  are "close" to the observed pattern  $y$

# ABC Shadow algorithm

## Key points

- ▶ need : an algorithm with outputs "close" enough to the posterior distribution
- ▶ tool : build a Markov chain evolving "close" to an equilibrium regime given by  $p(\theta|y)$
- ▶ plan : use the auxiliary variable method ideas

## Ideal MCMC sampling of the posterior : general MH algorithm

- ▶ assume the system is in the state  $\theta$
- ▶ choose a new value  $\psi$  according to a proposal density  $q(\theta \rightarrow \psi)$
- ▶ the value  $\psi$  is accepted with probability

$$\alpha_i(\theta \rightarrow \psi) = \min \left\{ 1, \frac{p(\psi|y)p(\psi)}{p(\theta|y)p(\theta)} \frac{q(\psi \rightarrow \theta)}{q(\theta \rightarrow \psi)} \right\}$$

- ▶ consider the proposal density

$$q(\theta \rightarrow \psi) = q_{\Delta}(\theta \rightarrow \psi | x) = \frac{f(x|\psi)/c(\psi)}{I(\theta, \Delta, x)} \mathbf{1}_{b(\theta, \Delta/2)}\{\psi\}$$

with

- ▶  $x$  : outcome of a marked point process driven by the probability density  $p(x|v)$  where  $v$  is any value in  $\Theta$ .
  - ▶ pattern detection context :  $f(x|\psi) = \exp[-U(x|\psi)]$
  - ▶  $\Delta > 0$  : control parameter
  - ▶  $\mathbf{1}_{b(\theta, \Delta/2)}\{\cdot\}$  is the indicator function over  $b(\theta, \Delta/2)$ , which is the ball of centre  $\theta$  and radius  $\Delta/2$
  - ▶  $I(\theta, \Delta, x) = \int_{b(\theta, \Delta/2)} f(x|\phi)/c(\phi) d\phi$ .
  - ▶ this choice guarantees the ideal chain to be uniformly ergodic and avoids the evaluation of the ratios  $c(\theta)/c(\psi)$
  - ▶ but, it requires the computation of integrals  $I(\theta, \Delta, x) \dots$
- blackboard : drawing

# Shadow chain : approximation of the ideal chain

**Theorem** : if  $p(x|\theta)$  is a continuously differentiable function in  $\theta$

- ▶ For any fixed  $\theta \in \Theta$  and  $A \in \mathcal{T}_\Theta$ , we have

$$\lim_{\Delta \rightarrow 0^+} \int_A |q_\Delta(\theta \rightarrow \psi) - U_\Delta(\theta \rightarrow \psi)| d\psi = 0$$

- ▶ For any fixed  $\theta \in \Theta$ , we have

$$\lim_{\Delta \rightarrow 0^+} \sup_{\psi \in \Theta} \left| \frac{q_\Delta(\theta \rightarrow \psi|x)}{q_\Delta(\psi \rightarrow \theta|x)} - \frac{\frac{f(x|\psi)}{c(\psi)} \mathbf{1}_{b(\theta, \Delta/2)}(\psi)}{\frac{f(x|\theta)}{c(\theta)} \mathbf{1}_{b(\psi, \Delta/2)}(\theta)} \right| = 0$$

uniformly in  $\theta \in \Theta$ , with

- ▶  $V_\Delta$  : the volume of the ball  $b(\theta, \Delta/2)$
- ▶  $U_\Delta = \frac{1}{V_\Delta} \mathbf{1}_{b(\theta, \Delta/2)}\{\psi\}$  : uniform probability density

**Application** : simulate the shadow chain that approximate the ideal chain

- ▶ first part : use  $U_{\Delta}(\theta \rightarrow \psi)$  instead of  $q_{\Delta}(\theta \rightarrow \psi)$  for proposing new values
- ▶ second part : approximates the computation of the proposal density ratio while simplifying the normalizing constant ratio
- ▶ the shadow Markov chain accepts new states with the probability :

$$\alpha_s(\theta \rightarrow \psi) = \min \left\{ 1, \frac{p(\psi|y)p(\psi)}{p(\theta|y)p(\theta)} \times \frac{f(x; \theta)c(\psi)\mathbf{1}_{b(\psi, \Delta/2)}\{\theta\}}{f(x; \psi)c(\theta)\mathbf{1}_{b(\theta, \Delta/2)}\{\psi\}} \right\}$$



**Corollary** : the acceptance probabilities of the ideal and shadow chains are uniformly as closed as desired whenever  $\Delta \rightarrow 0_+$

**Proposition** : Let  $P_i$  and  $P_s$  be the transition kernels for the ideal and the shadow Markov chains using a general  $\Delta > 0$  and a configuration  $x \in \Omega$ . Then, for every  $\epsilon > 0$  and every  $n \in \mathbb{N}$ , there exists  $\Delta_0 = \Delta_0(\epsilon, n) > 0$  such that for every  $\Delta \leq \Delta_0$

$$|P_i^{(n)}(\theta, A) - P_s^{(n)}(\theta, A)| < \epsilon$$

uniformly in  $\theta \in \Theta$  and  $A \in \mathcal{T}_\Theta$ .

**Algorithm ABC Shadow** : assume the observed pattern is  $y$  and fix values for  $\Delta$  and  $n$  and the current state  $\theta_0$

1. Generate  $x$  according to  $p(x|\theta_0)$
2. For  $k = 1$  to  $n$  do
  - ▶ Generate a new candidate  $\psi$  following  $U_{\Delta}(\theta_{k-1} \rightarrow \psi)$ .
  - ▶ The new state  $\theta_k = \psi$  is accepted with probability  $\alpha_s(\theta_{k-1} \rightarrow \psi)$ , otherwise  $\theta_k = \theta_{k-1}$
3. Return  $\theta_n$

## Remarks :

- ▶ if several samples are needed, re-start the procedure for the same  $\Delta$  and  $n$ , with  $\theta_0 = \theta_n$ .
- ▶ depending on  $\Delta$ , the algorithm approaches the equilibrium regime of the ideal chain :

$$\|P_s^{(n)}(\theta, A) - \pi(A)\| \leq M(x, \Delta)\rho^n + \epsilon.$$

with  $\pi(A) = \int_A p(\theta|y)d\theta$  ;  $M$  and  $\rho$  : ergodicity parameters of the ideal chain

- ▶ **caution** : this is not convergence

## Application : sampling the posterior of a Gaussian model

The posterior of a Normal model with mean  $\theta_1$  and variance  $\theta_2$  is

$$p(\theta_1, \theta_2 | y = \mathbb{Y}(\omega)) \propto \frac{\exp\left(\frac{\theta_1}{\theta_2} \mathbb{Y}(\omega) - \frac{\mathbb{Y}^2(\omega)}{2\theta_2}\right)}{c(\theta_1, \theta_2)} p(\theta_1, \theta_2)$$

with

- ▶  $y = \mathbb{Y}(\omega)$  : observation issued from the supposed model
- ▶  $t(y) = (\mathbb{Y}(\omega), \mathbb{Y}^2(\omega))$  : the sufficient statistics vector
- ▶ if the sample size is  $m$  then :  
 $t(y) = (\sum_{i=1}^m \mathbb{Y}_i(\omega), \sum_{i=1}^m \mathbb{Y}_i^2(\omega))$

## Experiment :

- ▶ simulate 1000 i.i.d. Normal r. v.'s with parameters  $\theta = (\mu, \sigma^2) = (2, 9)$
- ▶  $t(y) = (1765.45, 12145.83)$
- ▶  $p(\theta_1, \theta_2)$  the uniform distribution over  $[-100, 100] \times [0, 200]$
- ▶ compare 1000 samples of the MH and ABC Shadow algorithms
- ▶  $\Delta = (0.005, 0.025)$  and  $n = 500$

Summary statistics for Normal posterior sampling						
Algorithm	$Q_5$	$Q_{25}$	$Q_{50}$	$\bar{\theta}$	$Q_{75}$	$Q_{95}$
MH $\theta_1$	1.60	1.69	1.75	1.76	1.82	1.92
ABC $\theta_1$	1.60	1.70	1.76	1.76	1.82	1.91
MH $\theta_2$	8.45	8.80	9.07	9.08	9.33	9.76
ABC $\theta_2$	8.35	8.78	9.03	9.06	9.33	9.83

Table: Empirical quantiles and mean for the posterior of the Normal model.

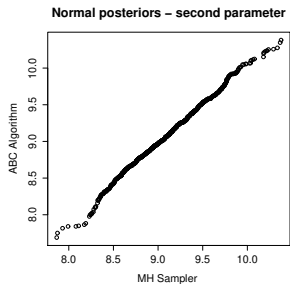
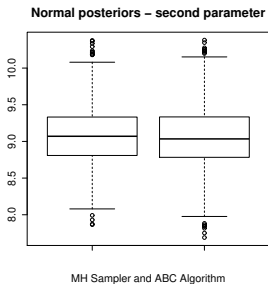
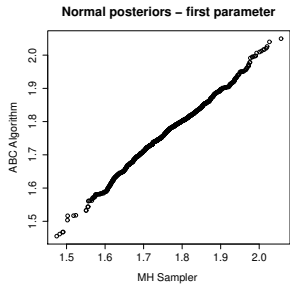
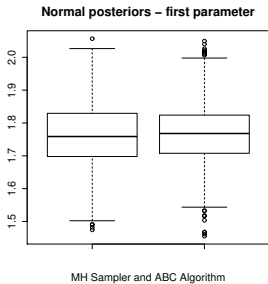
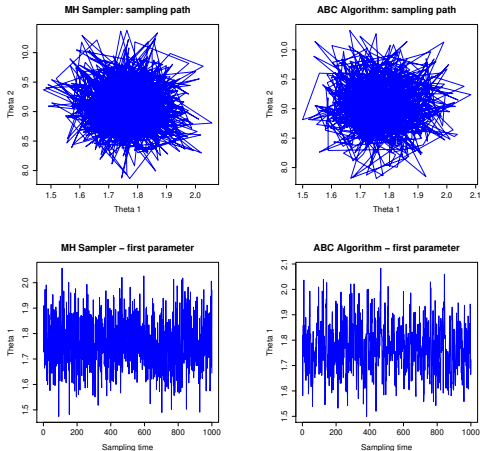


Figure: Boxplots and qqplots of the MH and ABC Shadow outputs.



**Figure:** Sample path for the Normal posterior. Left column: the MH algorithm results - from top to bottom the joint parameter path and the  $\theta_1$  time series. Right column: the ABC Shadow procedure - from top to bottom the joint parameter path and the  $\theta_1$  time series.

# Parameter estimation for marked point processes

Strauss model :

$$p(y|\theta) \propto \beta^{n(y)} \gamma^{s_r(y)}, \quad (18)$$

with

- ▶  $t(y) = (n(y), s_r(y))$  : sufficient statistics
- ▶  $\theta = (\log \beta, \log \gamma)$  : model parameters



## Experiment :

- ▶ domain  $W = [0, 1]^2$  and range parameter  $r = 0.1$
- ▶ simulate 1000 realisations of the model  $\theta = (4.60, -1.60)$  using the CFTP algorithm
- ▶ the empirical means of the sufficient statistics  $\bar{t}(y) = (\bar{n}(y), \bar{s}_r(y)) = (34.33, 5.31)$
- ▶ ABC Shadow “sampling” of  $p(\theta|\bar{t}(y)) \rightarrow 1000$  samples
- ▶  $\Delta = (0.01, 0.01)$  and  $n = 200$
- ▶  $p(\theta)$  the uniform distribution over  $[3.5, 5.5] \times [-5, 0]$

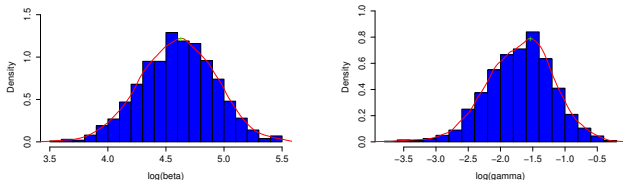


Figure: Strauss posterior sampling using the ABC procedure : histograms and kernel density estimates of the marginals. Maximum of the kernel estimated density is  $\hat{\theta} = (4.63, -1.53)$ .

Summary statistics for Strauss posterior sampling		
Algorithm	$Q_{50}$	$\bar{\theta}$
ABC $\log \beta$	4.606	4.603
ABC $\log \gamma$	-1.669	-1.700

Table: Empirical median and mean for the posterior of the Strauss model.

Candy model :

$$p(y|\theta) \propto \exp\langle \theta_d n_d(y) + \theta_s n_s(y) + \theta_f n_f(y) + \theta_r n_r(y) \rangle \quad (19)$$

with

- ▶  $t(y) = (n_d(y), n_s(y), n_f(y), n_r(y))$  : sufficient statistics
- ▶  $\theta = (\theta_d, \theta_s, \theta_f, \theta_r)$  : model parameters

## Experiment :

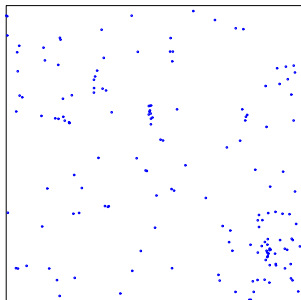
- ▶ domain  $W = [0, 3] \times [0, 1]$ , segment length  $l = 0.12$ , connection range  $r_c = 0.01$ , curvature parameters  $\tau_c = \tau_r = 0.5$  radians
- ▶ simulate 1000 realisations of the model  $\theta = (10, 7, 3, -1)$  using an Adapted MH algorithm
- ▶ the empirical means of the sufficient statistics  $\bar{t}(y) = (51.10, 101.06, 19.97, 72.89)$
- ▶ ABC Shadow “sampling” of  $p(\theta|\bar{t}(y)) \rightarrow 1000$  samples
- ▶  $\Delta = (0.01, 0.01, 0.01, 0.01)$  and  $n = 500$
- ▶  $p(\theta)$  the uniform distribution over  $[2, 12]^3 \times [-7, 0]$

Summary statistics for Candy posterior sampling		
Algorithm	$Q_{50}$	$\bar{\theta}$
ABC $\log \theta_d$	9.995	9.998
ABC $\log \theta_f$	2.977	2.975
ABC $\log \theta_s$	7.005	7.008
ABC $\log \theta_r$	-1.014	- 1.016

**Table:** Empirical median and mean for the posterior of the Candy model. Maximum of the kernel estimated density is  $\hat{\theta} = (9.96, 3.02, 6.99, -1.00)$ .

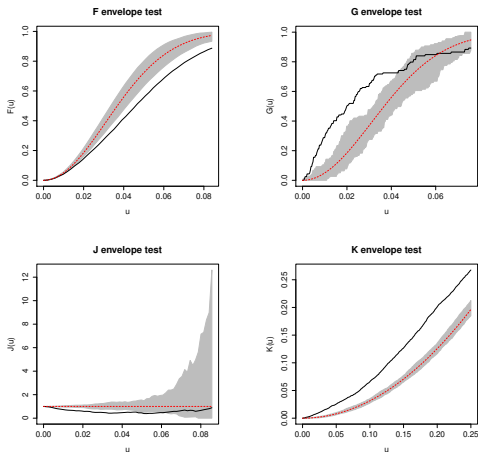
## Real data application : cosmology

- ▶ short range correlations for the galaxy distribution



**Figure:** Sample of a cosmological data set. The points represent galaxy positions in a region of our Universe.

► exploratory analysis : summary statistics  $\rightarrow$  cluster behaviour



**Figure:** Summary statistics envelope test (100 simulations) for the galaxy pattern: shaded (gray) region - the Monte Carlo envelopes, dotted line (red) - the theoretical statistics, continuous (black) line - the observed statistics.

- ▶ sampling the posterior of an area-interaction model conditionally on different range parameters
- ▶ the posterior sampling  $\rightarrow$  the probability that the model parameters indicate clustering

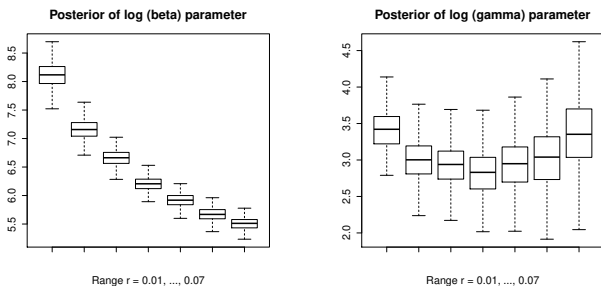
Data for the Galaxy pattern							
$r$	0.01	0.02	0.03	0.04	0.05	0.06	0.07
$n(y) = 163$							
$-a_r(y)$	135.91	114.05	96.44	82.23	69.85	59.05	49.73

**Table:** The observed sufficient statistics computed for the galaxy pattern, depending on the range parameter  $r$ . For all these parameters  $n(y)$  remains constant, while  $a_r(y)$  depends on  $r$ .



- the area interaction model :  $p(y|\theta) \propto \beta^{n(y)}\gamma^{ar(y)}$  with

$$a_r(y) = -\frac{\nu [A_r(y))]}{\pi r^2} = -\frac{\nu [\cup_{i=1}^n b(w_i, r)]}{\pi r^2}$$



**Figure:** Box-plots of the posterior distributions for the parameters of the area-interaction process estimated from the galaxy pattern, given different values for the interaction radius.

Estimation errors							
$r$	0.01	0.02	0.03	0.04	0.05	0.06	0.07
Asymptotic standard deviation							
$\hat{\sigma}_{\log \beta}$	0.20	0.17	0.13	0.11	0.10	0.10	0.08
$\hat{\sigma}_{\log \gamma}$	0.26	0.28	0.27	0.30	0.34	0.40	0.52
Monte Carlo standard deviation							
$\hat{\sigma}_{\log \beta}^{MC}$	0.001	0.002	0.001	0.002	0.002	0.002	0.003
$\hat{\sigma}_{\log \gamma}^{MC}$	0.002	0.003	0.004	0.006	0.008	0.012	0.024

**Table:** Estimation errors computed for the MAP estimates obtained from the galaxy pattern.

# Synthesis parameter estimation

## Pseudo-likelihood estimation :

- ▶ easy to compute
- ▶ good alternative whenever nothing else can be done
- ▶ consistency and central limit theorems : difficult to interpret
- ▶ no properties concerning the sufficient statistics of the model using the PL estimates of the parameters
- ▶ work of J. Mateu and P. Montes : comparison with maximum likelihood

## Monte Carlo maximum likelihood :

- ▶ general statistical framework
- ▶ numerically unstable  $\rightarrow$  but re-sampling is guaranteed to convergence, since the log-likelihood is convex
- ▶ the asymptotics are related to the true model
- ▶ **property** : the expectation of the sufficient statistics computed by the model with the ML parameters equals the observed sufficient statistics

## ABC Shadow parameter estimation :

- ▶ approximate algorithm that samples "close" to the posterior
- ▶ interesting numerical results : comparable with classical MCMCML inference
- ▶ relatively low computational cost : MCMCML needs re-sampling
- ▶ complementary tool → applications :
  - ▶ initialisation point for more rigorous methods
  - ▶ statistical testing
  - ▶ model validation
- ▶ open problem : range parameters

## Open questions :

- ▶ range parameters
- ▶ parameters of the mark distribution
- ▶ posterior sampling
- ▶ incomplete data : EM algorithms converges towards the first local maximum → a lot of references available ...
- ▶ ABC methods : empirical methods for parameter estimation  
→ control the sufficient statistics

## Model validation : residual analysis for point processes

Let  $\mathbb{X}$  be a locally stable marked point process on  $W \times M$ .

**$h$ -Innovations** : for nonnegative functions  $h$  and  $A \subseteq W \times M$

$$I(A, h, \lambda) = \sum_{x_i \in \mathbb{Y}_A} h(x_i, \mathbb{X} \setminus x_i) - \int_A \lambda(\eta; \mathbb{X}) h(\eta, \mathbb{X}) (\nu \times \nu_M)(d\eta)$$

- ▶ assuming the sum and the integral in the definition have finite expectations, the Georgii-Nguyen-Zessin formula gives

$$\mathbb{E}I(A, h, \lambda) = 0$$

- ▶  $I$  is a signed measure
- ▶  $\Delta I(x_i) = h(x_i, \mathbb{X} \setminus \eta)$  : the innovation increment ('error') attached to a point  $\eta \in \mathbb{X}$
- ▶  $dI(\eta) = -\lambda(\eta; \mathbb{X}) h(\eta, \mathbb{X})$  : the innovation increment attached to a background location  $\eta \in W \times M$

$h$ -Residuals : for  $h \geq 0$  functions and  $A \subseteq W \times M$

$$\begin{aligned} R(A, \hat{h}, \hat{\theta}) &= I(A, \hat{h}, \hat{\lambda}) \\ &= \sum_{x_i \in \mathbb{X}_A} \hat{h}(x_i, \mathbb{X} \setminus x_i) - \int_A \hat{\lambda}(\eta; \mathbb{X}) \hat{h}(\eta, \mathbb{X}) (\nu \times \nu_M)(d\eta) \end{aligned}$$

since the function  $h$  may depend on the model,  $\hat{h}$  denotes an estimate.

Application idea :

- ▶ consider a parametric model for a marked point process  $\mathbb{X}$  observed within  $A$
- ▶ estimate the model parameters (maximum likelihood, pseudo-likelihood)
- ▶ expect the residuals  $R(A)$  to be close to 0 if the model is appropriate



Building residuals : several possible choices for  $h$

- ▶ raw residuals  $h(\eta, \mathbf{x}) = 1$

$$R(A, 1, \hat{\theta}) = n(\mathbf{x} \cap A) - \int_A \hat{\lambda}(\eta; \mathbf{x})(\nu \times \nu_M)(d\eta)$$

- ▶ inverse residuals  $h(\eta, \mathbf{x}) = 1/\lambda(\eta; \mathbf{x})$  (equivalent with the Stoyan-Grabarnik diagnostic)

$$R(A, \frac{1}{\hat{\lambda}}, \hat{\theta}) = \sum_{x_i \in X_A} \frac{1}{\hat{\lambda}(x_i; X_A)} - \int_A 1_{\{\hat{\lambda}(\eta; \mathbf{x}) > 0\}}(\nu \times \nu_M)(d\eta)$$

- ▶ Pearson residuals  $h(\eta, \mathbf{x}) = 1/\sqrt{\lambda(\eta; \mathbf{x})}$  (analogy with Poisson log-linear regression)

$$R(A, \frac{1}{\sqrt{\hat{\lambda}}}, \hat{\theta}) = \sum_{x_i \in x_A} \frac{1}{\sqrt{\hat{\lambda}(x_i; x_A)}} - \int_A \sqrt{\hat{\lambda}(\eta; \mathbf{x})} (\nu \times \nu_M)(d\eta)$$

**Remark :** the inverse and Pearson residuals we need  $\lambda_{\theta(x)}(x_i; \mathbf{x}) > 0$  for all  $x_i \in \mathbf{x}$  for any pattern  $\mathbf{x}$ , while  $\lambda_{\theta(x)}(\eta; \mathbf{x}) = 0$  is allowed for  $\eta \notin \mathbf{x}$

## Properties :

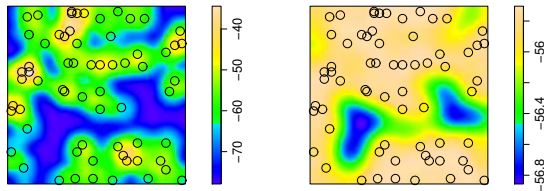
- ▶ expectation

$$\begin{aligned} & \mathbb{E} \left[ R(A, \hat{h}, \hat{\theta}) \right] \\ &= \int_A \mathbb{E} \left[ h_{\hat{\theta}(\mathbb{X} \cup \{\eta\})}(\eta, \mathbb{X}) \lambda(\eta, \mathbb{X}) - h_{\hat{\theta}(\mathbb{X})}(\eta, \mathbb{X}) \lambda_{\hat{\theta}(\mathbb{X})}(\eta, \mathbb{X}) \right] \end{aligned}$$

- ▶ variance : more complicate structures but very nice formulas for Poisson processes (Baddeley, Moller and Pakes 2008)
- ▶ these residuals do not have independent increments  $\rightarrow$  the raw innovations for Markov point processes are conditionnaly independent and uncorrelated (Baddeley, 2005)
- ▶ consistency and asymptotic normality for the residuals of stationary Gibbs point processes (Coeurjolly and Lavancier, 2013)

**Application** : smoothed residuals to test several models for japanesepines datasets

- ▶ Strauss process : only repulsion
- ▶ area-interaction process : repulsion or attraction (competition for ressources)



**Figure:** Raw residual analysis, from left to right : Strauss( $r=0.08$ ) and AreaInt( $r=0.09$ )

- ▶ R code:

```
>mjp=ppm(japanesepines, 1,Strauss(r=0.08),rbord=0.08)  
>rjp=residuals(mjp,type="raw") >plot(rjp)
```

QQ plots: comparison of empirical quantiles of the smoothed residuals with the expected quantiles under the estimated model

- ▶ interpretation in the spirit of  $K$  and  $F$  functions
- ▶ if the data pattern is more clustered than the model: heavier tails especially in the left-hand tail
- ▶ if the data pattern is more inhibited than the model: lighter tails especially in the right-hand tail
- ▶ R code : `qqplot.ppm(rjp, type="raw")`

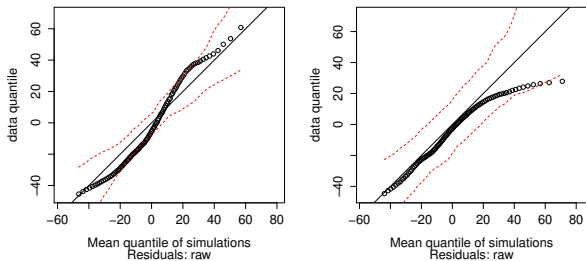


Figure: Q-Q plot analysis, from left to right : Strauss( $r=0.08$ ) and ArealInt( $r=0.09$ )

- ▶ Strauss( $r=0.08$ ): over-estimates repulsion, but under-estimates close attraction
- ▶ ArealInt( $r=0.09$ ): very well for the close attraction, underestimate the repulsion
- ▶ the best model for the entire data set : polynomial inhomogeneity and soft-core interaction

## Remarks :

- ▶ the theory is wonderful
- ▶ but the numerical results are obtained using the PL estimators
- ...
- ▶ see the remark of J. Besag
- ▶ visualisation of residuals difficult for higher dimensions
- ▶ the qq plots very informative → link with the central limit theorems for computing confidence intervals
- ▶ open question : validating pattern detection result ... ?

→ Exercise 37

## Exercises : Parameter Estimation.

*Exercise 34.* Let  $X$  be a homogeneous Poisson point process on a compact  $W \subset \mathbb{R}^d$  with unknown intensity parameter  $\theta$ .

- a) Write its probability density w.r.t. the reference given by the unit intensity Poisson process. Specify its normalising constant.
- b) Write the log-likelihood function and give the maximum likelihood estimate of  $\theta$ .



*Exercise 35.* Let  $X$  be an inhomogeneous Poisson point process on a compact  $W \subset \mathbb{R}^d$  with the intensity function  $\lambda$  depending on the parameter  $\theta$ .

- a) Write its probability density w.r.t. the reference given by the unit intensity Poisson process. Specify its normalising constant.
- b) Write the log-likelihood function and compare it with the log pseudo-likelihood of a point process.
- c) In this case, what is the difference between the intensity of the process and its conditional intensity ?

*Exercice 36.* Let  $X$  be a Strauss point process on a compact  $W \subset \mathbb{R}^d$  given by the probability density w.r.t. the unit rate Poisson point process :

$$p(x; \theta) = \alpha(\theta) \beta^{n(x)} \gamma^{s(x)}$$

where  $n(x)$  denotes the cardinality of  $x$ ,  $s(x)$  the number of  $r$ -close pairs in  $x$  and  $\theta = (\beta, \gamma)$  with  $\beta > 0$  and  $\gamma \in (0, 1)$  is the parameter vector of interest

- a) Specify the normalising constant and write the log-likelihood function.
- b) Write the system of equations for obtaining  $\theta$ .
- c) Show that the maximum likelihood estimator solves :

$$(n(x), s(x)) = (\mathbb{E}_\theta n(X), \mathbb{E}_\theta s(X))$$

*Exercice 37.* The data set `waterstriders` within the package **spatstat** contains three point patterns representing the positions of water striders in a pond. Try to answer the following question : what is the more important factor in the relative position of these insects - the distance w.r.t to its neighbour or the occupied territory ? In order to answer the question, do the following steps :

- a) Plot the data. Do a summary statistics analysis in order to propose at least two candidates for modelling ? What type of interaction is exhibited by all these patterns ?
- b) Use the profile of the pseudo-likelihood and the previous analysis in order to propose a good candidate for the interaction radius.
- c) Fit the models using `ppm`.
- d) Verify the models using envelope tests.
- e) Choose a model using the residual analysis : `residuals`, `qqplot`

# Table of contents

## Cours 10. Pattern detection and characterisation.

Statistical pattern detection

Does the detected pattern really exist ?

Exercises : Pattern detection and characterisation.

# Pattern detection and characterisation

The pattern detection problem :

- ▶ a spatial data set is observed  $\rightarrow$  pattern hidden in the data ?
- ▶ hypothesis : the pattern is the realization of a random process (MRFs, marked point processes, etc.)
- ▶ the Gibbsian modelling framework  $\rightarrow$  write a probability density model

$$p(y|\theta) = \frac{\exp[-U(x|\theta)]}{c(\theta)}$$

with  $U : \Omega \rightarrow \mathbb{R}^+$  the energy function,  $\theta$  the model parameters and  $c(\theta)$  the normalizing constant

- ▶ parameters knowledge  $\rightarrow$  prior density  $p(\theta)$
- ▶ the pattern estimator is

$$(\hat{x}, \hat{\theta}) = \arg \max_{(x, \theta) \in \Omega \times \Theta} \{p(x|\theta)p(\theta)\} = \arg \max_{(x, \theta) \in \Omega \times \Theta} \{p(x, \theta)\}.$$

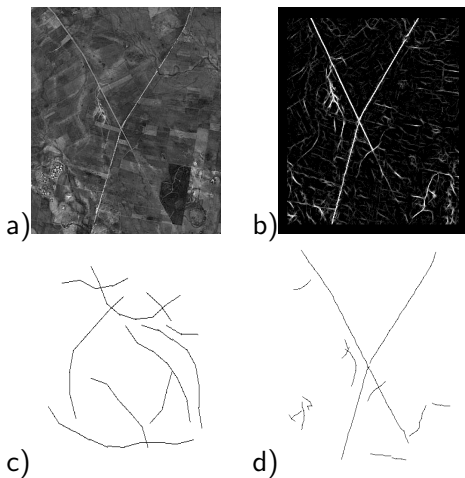
# Statistical pattern detection

Build the pattern model : probability density construction conditionally on the data observation

$$p(x, \theta | d) \propto \exp \left[ -\frac{U_d(x|\theta) + U_i(x|\theta)}{Z(\theta)} + \log p(\theta) \right]$$

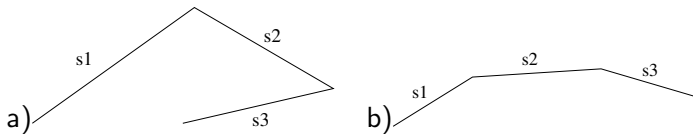
- ▶ **interaction energy**  $U_i(x|\theta)$  → objects interactions (geometrical shape of the structure)
- ▶ **data energy**  $U_d(x|\theta)$  induced by the data field  $d$  → object locations
- ▶ if the interaction parameters are unknown → **prior model**  $p(\theta)$

- ▶ role of the interaction and data energies :



**Figure:** Influence of the energy components of the model : a) original image SPOT ; and results obtained using only : the data term (b), the interaction term (c), the complete model (d)

▶ setting the model parameters :



**Figure:** Two segments configurations : a) the connectivity is favored over alignment, b) connectivity and alignment have equivalent potentials.



**Pattern estimator** : the object configuration that maximises the probability density

$$(\hat{x}, \hat{\theta}) = \arg \min_{\Omega \times \Psi} \left\{ \frac{U_d(x|\theta) + U_i(x|\theta)}{Z(\theta)} - \log p(\theta) \right\}$$

with  $\Psi$  the model parameters space.

**Simulated annealing** : global optimisation technique

- ▶ sampling from  $p(x, \theta)^{1/T}$  while slowly  $T \rightarrow 0$
- ▶ convergence towards the uniform distribution on the configuration subspace minimizing  $U(x, \theta)$  (Stoica, Gregori and Mateu, 2005)
- ▶ inhomogeneous Markov chain

Algorithm SA :  $x = \text{Simulated Annealing} (T_0, \delta, T)$

1. choose an initial condition  $x_0$
2. for  $i = 1$  to  $T$  do
  - {
  - $x_i = \text{Update} (x_{i-1}, T_{i-1}, \delta)$
  - $T_i = T_0 / [\log(i) + 1]$
  - }
3. return  $x_T$ .

▶ slow algorithm  $\rightarrow$  an alternative cooling schedule :

$$T_{n+1} = cT_n \quad \text{with} \quad c \in [0.95, 1[$$

▶ simulated tempering : improving mixing properties

Level sets estimators :

- ▶ visit maps for compact regions in  $W$  :

$$\{T(w) > \alpha\} \Rightarrow \{T_n(w) > \alpha\}$$

with  $T(w) = \mathbb{P}(w \in X)$  the probability that the structure hits a point in  $W$

- ▶ link with the capacity functional and volumic fraction

- ▶ two challenges : discretisation and Monte Carlo approximations
- ▶ Vorob'ev expectation : the level set with volume equal to the mean volume of the random set
  - ▶ demands the knowledge of the behaviour of an unknown random set, but still manageable in practice ...
- ▶ **average behaviour** of the pattern (fixed temperature)
- ▶ (Heinrich, Stoica and Tran, 2012) prove the convergence  $\mathbb{L}^1$  of these estimators

→ Exercise 38

# Build the machine ...

Filaments detection in galaxies catalogues :

- ▶ interaction energy : Bisous model (random cylinders)
- ▶ data energy : local tests (density and spread of galaxies inside a cylinder)

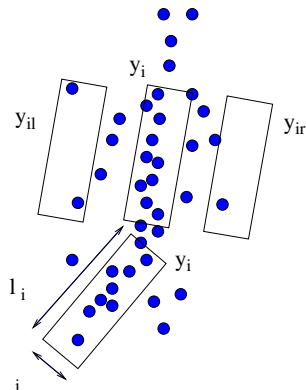
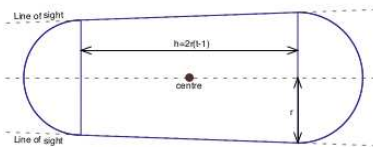


Figure: Locating interacting cylinders in a field of points.

## Cluster detection in galaxies catalogues :

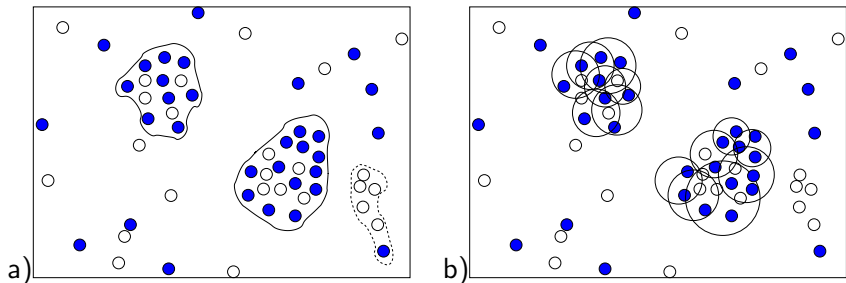
- ▶ interaction energy : Strauss and Area-interaction models (random object : trunk of a cone + two half-spheres)
- ▶ data energy : local minimum number of galaxies inside the considered random object tests



**Figure:** Cross-section of the considered random object : two half-spheres connected with a truncated cone. The object is fully determined by its centre position, radius  $r$  and shape parameter  $t \geq 1$ . Shape parameter  $t$  gives the aspect ratio of the object along and perpendicular to the line of sight; for  $t = 1$  the object is a ball. For a given  $r$  and  $t$  the height of the truncated cone is defined as  $h = 2r(t - 1)$ . The shape of the truncated cone is defined by the lines of sights, which are indicated by dashed lines on the figure. The observer is located at far left from the object.

## Cluster detection in epidemiological data :

- ▶ interaction energy : Strauss and Area-interaction models (random disks)
- ▶ data energy : local statistical test (the average score of the farms covered by a disk)



**Figure:** Data  $\rightarrow$  field of marked points : a) observed clusters, b) clusters approximated by random disks.

# Orbit determination for binary systems (1)

- ▶ interaction energy :

$$U_I(\theta) = \log p(\theta) = \sum_{i=1}^7 \log p(\theta_i),$$

where  $\theta = (a, e, i, \Omega, \omega, \tau, P)$  is the vector of orbital parameters

- ▶ Jeffreys' principle - non-informative independent priors
- ▶ our choice : uniform distributions over bounded intervals
- ▶ the intervals were chosen taking into account the *a priori knowledge* of the objects to be detected
- ▶ perspective : introduce dependence of the parameters



## Orbit determination for binary systems (2)

- ▶ data energy : sum of the distances between the observed positions and the computed positions ; these last ones are computed using the given model parameters

$$U_d(\varphi|\theta) = \sum_{i=1}^n \left[ - \left( |x_i^o - x_i^c|^l + |y_i^o - y_i^c|^l \right)^{k/l} \right] =$$

where

- ▶  $d = \{(x_i^o, y_i^o)\}, (i = 1, 2, \dots, n)$  : the  $n$  observed positions of the secondary asteroid with respect to the primary
- ▶  $\{(x_i^c, y_i^c)\}$  : the computed positions at the same time  $i$  as the corresponding observations, given the current  $\theta$
- ▶  $k, l$  : pre-fixed model parameters
  - ▶  $k = l = 2$  : Gaussian character of the data model
  - ▶  $k = l = 1$  : Laplacian character of the data model
- ▶ perspective : model choice

# Road network extraction in satellite and aerial images

(Stoica, Descombes, van Lieshout and Zerubia, 2002)

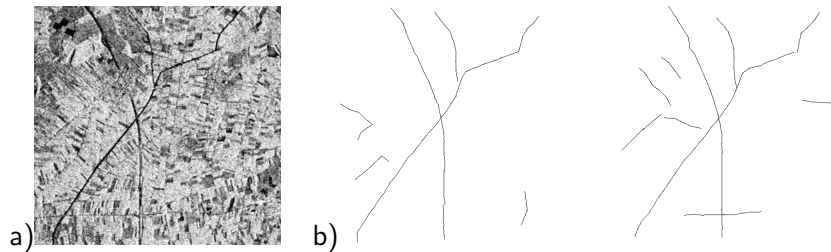
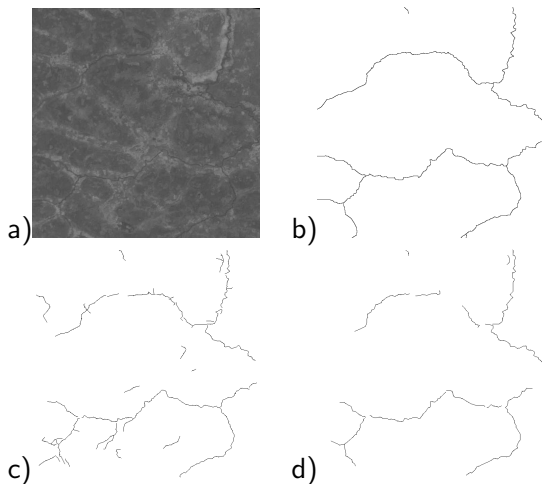


Figure: Rural region in Malaysia : a) original image; b) obtained results.

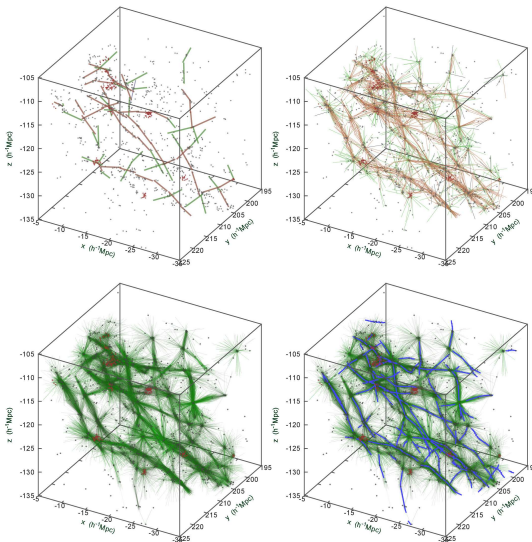
# Forest galleries : verifying the results

(Stoica, Descombes and Zerubia, 2004)



**Figure:** Forest galleries extraction : a) original image ; b) ground truth ; c)-d) obtained results. Data provided by BRGM.

## Filaments detection (1) : (Tempel, Stoica et. al., 2014)



**Figure:** Detected filamentary pattern (cylinder axes) in a small sample volume within a pattern of galaxies (points).

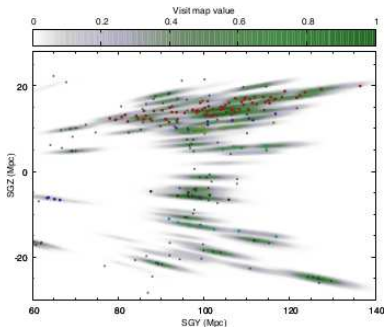
## Filaments detection (2)

(Tempel, Stoica et. al., 2014)

The movie, showing the MCMC in action is available at

[:http://www.aai.ee/elmo/sdss-filaments/](http://www.aai.ee/elmo/sdss-filaments/)

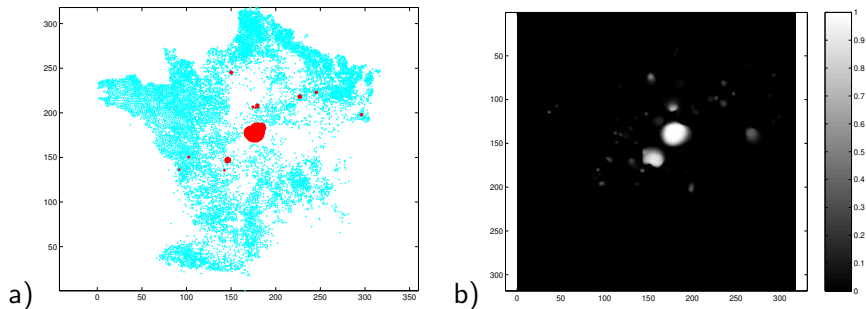
## Cluster detection in cosmology : (Tempel, Stoica et. al., 2018)



**Figure:** The distribution of galaxies in supergalactic coordinates (points) and the visit map obtained using the cluster detection process. The thickness of the slice is 4 Mpc around  $SGX = 0$ . Red points show galaxies in the Coma cluster while the other coloured points show galaxies in some clusters with at least five members. The grey points show all remaining galaxies.

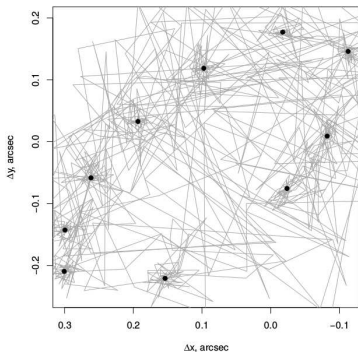
# Cluster detection in epidemiology : sub-clinical mastitis data

(Stoica, Gay and Kretzschmar, 2007)



**Figure:** Disease data scores and coordinates for the year 1996 : a) disk configuration obtained using the simulated annealing algorithm ; b) cover probabilities.

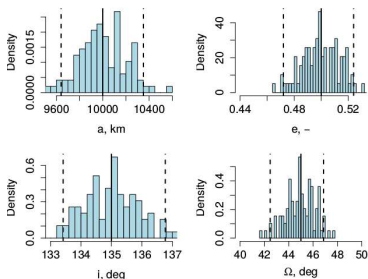
## Orbit determination (1) : (Kovalenko, Stoica and Emelyanov, 2017)



**Figure:** Simulated observations (black points):  $\Delta x$  and  $\Delta y$  correspond to relative positions of the secondary with respect to the primary. Lines show a search for the optimal solution during SA algorithm.



## Orbit determination (2) :



**Figure:** Resulting distributions of semi-major axis  $a$ , eccentricity  $e$ , inclination  $i$  and longitude of the ascending node  $\Omega$  (referenced to J2000 equatorial frame) obtained for simulated observations with the likelihood model. Solid line represents the true parameter value. The dotted lines represent the 2.5% and 97.5% quantiles of the resulting sample.

# Does the detected pattern really exist ?

**Idea** : the sufficient statistics of the model  $\rightarrow$  morphological descriptors of the shape hidden by the data

- ▶ turn the machine at constant temperature  $T = 1$
- ▶ compute the average of the sufficient statistics
- ▶ compare with the maximum value obtained for the permuted data

**Sufficient statistics** :

- ▶ Bisous model (pattern of connected cylinders) : free cylinders, cylinders with one extremity connected, cylinders with both extremities connected

# Test for the filaments existence in galaxy catalogs

Permuted data : keeping the same number of galaxies while spreading them uniformly (binomial point process)

Sufficient statistics	Data		
	NGP150	NGP200	NGP250
$\bar{n}_2$	4.13	5.83	9.88
$\bar{n}_0$	15.88	21.19	35.82
$\bar{n}_1$	21.35	35.58	46.49

Sufficient statistics	Simulated data (100 binomial catalogs)		
	NGP150	NGP200	NGP250
$\max \bar{n}_2$	0.015	0.05	0.015
$\max \bar{n}_0$	0.54	0.27	0.45
$\max \bar{n}_1$	0.39	0.24	0.33

# Test for the cluster existence epidemiological data

**Permuted data** : keeping the same farm locations while exchanging the score disease

**Results** :

- ▶ sufficient statistics for the data from the year 1996 :

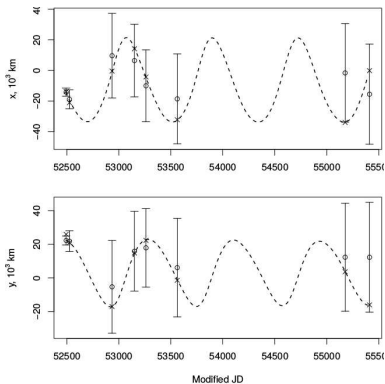
$$\bar{n}(y) = 74.10, \quad \bar{\nu}[Z(y)] = 312.46, \quad \bar{n}_o = 555.08$$

- ▶ maximum values of the sufficient statistics for 100 simulated data fields

$$\bar{n}(y) = 2.36, \quad \bar{\nu}[Z(y)] = 13.83, \quad \bar{n}_o = 2.62$$

**Interpretation** : this test does not say if the pattern is well detected, but it says that there is something to be detected ...

## Orbit determination validation : position prediction for the obtained parameter values



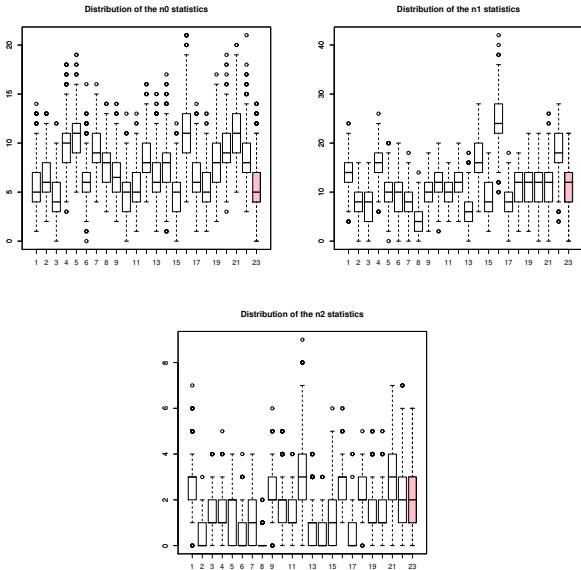
**Figure:** The calculated positions (black circles) are compared with given observed positions (crosses) by the  $x$  and  $y$  coordinates on sky-plane. Black bars denote the 2.5%-97.5% quantiles interval. Dotted line corresponds to the calculated positions for the orbit, obtained with the entire set of observations.

# How similar are two data sets ?

**Cosmology** : compare the sufficient statistics for 22 mock catalogues with the ones for the observation (Stoica, Martinez and Saar, 2010)

## Discussion

- ▶ mock catalogues exhibit filaments
- ▶ mock filaments are generally shorter, more fragmented and more dense
- ▶ Bisous model : good for testing the filamentary structure



**Figure:** Comparison of the sufficient statistics distributions for the real data (dark box plot) and the mock catalogues.

# Some conclusion and perspectives

## Spatial models and random geometry :

- ▶ Markov marked point processes allow statistical and morphological description of the pattern
- ▶ good synthesis properties
- ▶ limitations : models remain just models ...

## Perspectives :

- ▶ stochastic processes and random geometry (marked point processes, random fields) → modelling, simulation, statistical inference
- ▶ temporal dimension ...
- ▶ applications : astronomy and environmental sciences

**Acknowledgements** : this work was done together with wonderful co-authors and also with the precious help of some very generous persons ...



## Exercises : Pattern detection and characterisation.

*Exercise 38.*

- a) Build an Metropolis-Hastings algorithm for sampling the probability distribution :

$$f(x; k, \theta) \propto \Gamma(k, \theta) + 0.75\Gamma(3k, 2\theta)$$

with  $k = 6$  and  $\theta = 0.5$ . Please explain the choice of the algorithm parameters.

- b) Find the maximum of this distribution using the *Simulated Annealing* algorithm.

- c) Compare the algorithm performances depending of the following cooling schedules

$$T_{k+1} = cT_k, \quad c \in \{0.99, 0.999, 0.9999, 0.99999\}$$

and

$$T_{k+1} = \frac{T_0}{\log(k+1) + 1}, \quad k \in \mathbb{N}.$$

- d) Test the algorithm dependence on the set of initial conditions given by  $x_0 = \{0.5, 8, 25\}$ .