Mathematical statistics in physics and astronomy

Radu Stoica
Université Lille - Laboratoire Paul Painlevé
Observatoire de Paris - Institut de Mécanique Céleste et Calcul d Ephémérides
radu.stoica@math.univ-lille1.fr

Tartu, September, 2014

About me

- associate professor at Université Lille
- working together with the Toravere Observatory since 2003
- mail : `radu.stoica@univ-lille1.fr`
- web page : `http://math.univ-lille1.fr/ stoica/`
- office : Toravere Observatory (during September)
- phone : + 33 6 20 06 29 30

Course structure

- 24 hours = 4 weeks $\times$ 6 hours = 4 weeks $\times$ 2 days *times* 3 hours
- mark : project + written/oral exam (last day of the course - you need a mark of 50% to validate it)

Bibliogaphical references

• Baddeley, A. J. (2010) Analysing spatial point patterns in 'R'. Workshop Notes, Version 4.1.

• 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.

• 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/CRC.V

• Chiu, S. N., Stoyan, D., Kendall, W. S., Mecke, J. (2013) Stochastic geometry and its applications. John Wiley and Sons.

Aknowledgements

A. J. Baddeley, G. Castellan, J.-F. Coeurjolly, X. Descombes, P. Gregori, C. Lantuéjoul, M. N. M. van Liehsout, V. Martinez, J. Mateu, J. Møller, Z. Pawlas, F. Rodriguez, E. Saar, D. Stoyan, E. Tempel, R. Waagepetersen, J. Zerubia

Perfect or exact simulation

Lesson IX

Statistical inference problems

Monte Carlo Maximum likelihood estimation

Parameter estimation based on pseudo-likelihood

Model validation : residual analysis for point processes

Statistical pattern detection

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" question :

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

"The" answer :

- generally, this "common" property 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 we are looking for, it forms a pattern that has relevant geometrical characteristics

"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 $\rightarrow$ www.spatstat.org
- C++ library : **MPPLIB** by A. G. Steenbeek, M. N. M. van Lieshout, R. S. Stoica and contributors $\rightarrow$ available at simple demand

Toravere Observatory : huge quantities of data sets and software, and the questions going together with

- there are plenty of future Nobel subjects for you :)
- contact : E. Tempel, L. J. Liivamägi, E. Saar and myself

Forestry data (1) : the points positions exhibit attraction $\rightarrow$ clustered distribution

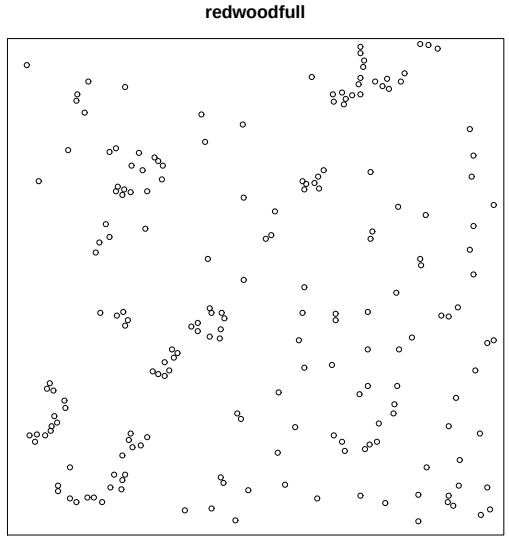


Figure: Redwoodfull data from the **spatstat** package

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

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

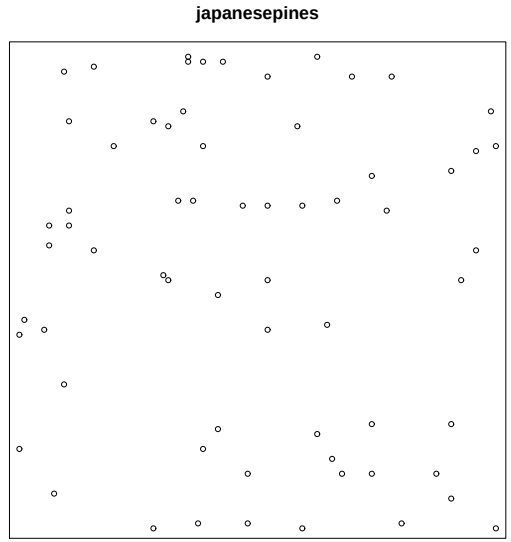


Figure: Japanese data from the **spatstat** package

In order to see all the available data sets

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

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

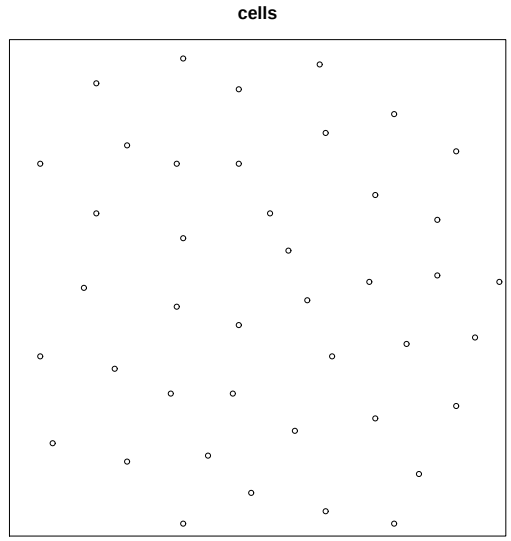


Figure: Cell data from the **spatstat** package

```
> 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

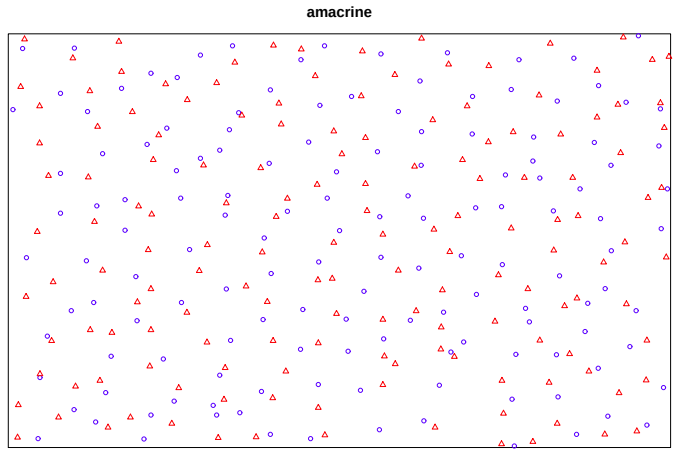


Figure: Amacrine data from the **spatstat** package

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

Animal epidemiology : sub-clinical mastitis for diary herds

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

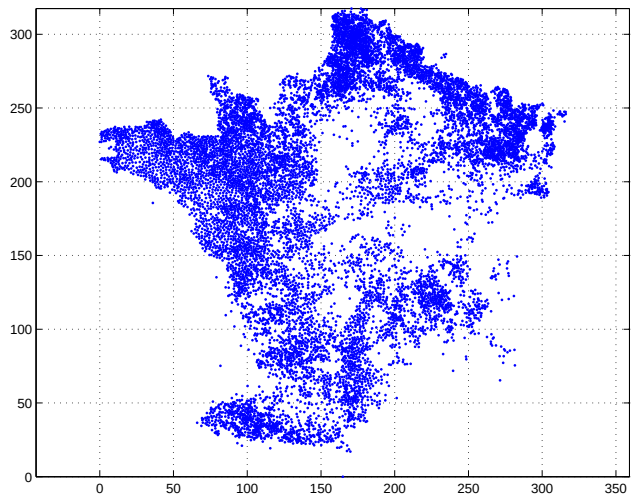


Figure: The spatial distribution of the farms outlines almost the entire French territory (INRA Avignon).

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 $\rightarrow$ 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

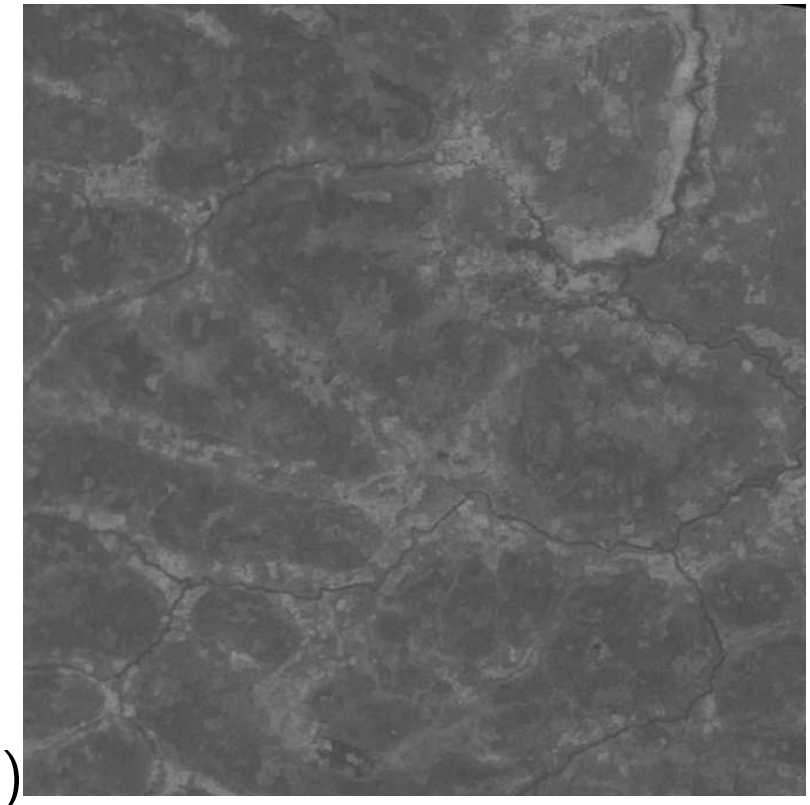
a)  b) 

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

Thin networks : some comments

- road and hydrographic networks $\rightarrow$ 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

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

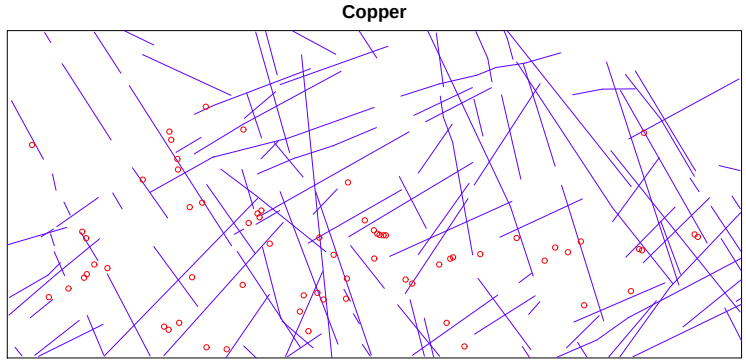


Figure: Copper data from the **spatstat** package

```
> 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")
```

Cosmology (1) : spatial distribution of galactic filaments

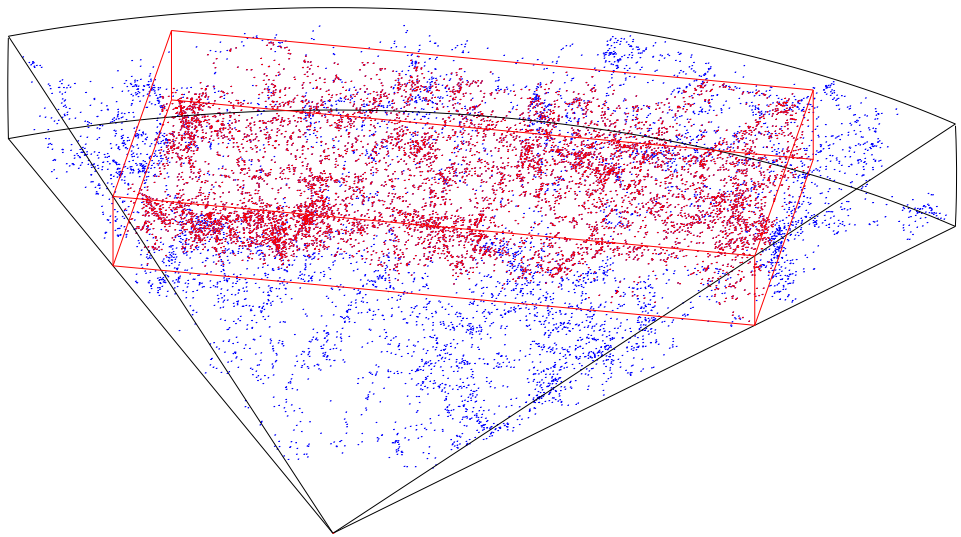
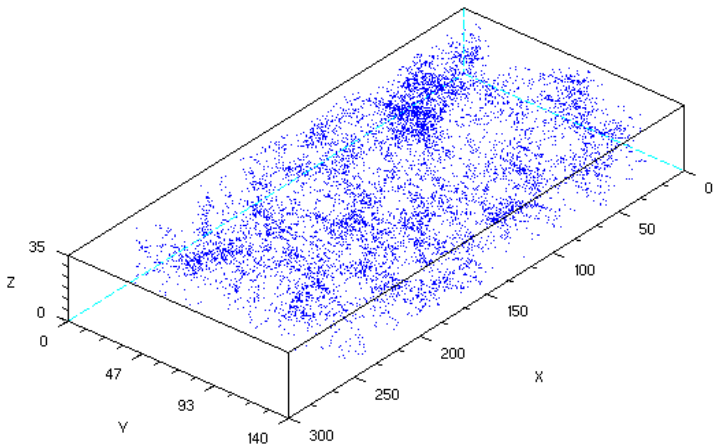
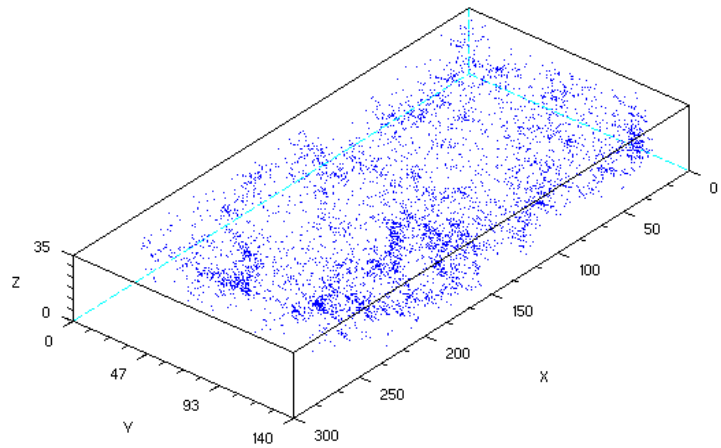


Figure: Cuboidal sample from the North Galactic Cap of the 2dF Galaxy Redshift Survey. Diameter of a galaxy $\sim 30 \times 3261.6$ light years.

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

Real data

- 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 ?

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 $(q, \cos i, \triangle z)$
- spatial data framework : location $(q, \cos i)$ and marks $\triangle z$
- $\triangle z = z_f - z_i$: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 (1)

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 temporal behaviour ?

Satellite debris :

- after explosion, the debris of two colliding artificial satellites distributes a long an orbit around Earth
- test the uniformity of the debris along the orbit
- characterize the spatio-temporal distribution of the entire set of debris
- extremely interesting set : the evolution of the objects dynamics is known from the initial state, *i.e.* no debris at all

→ video spatial debris

Spatio-temporal data (2)

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 $\rightarrow$ 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

$\rightarrow$ video roads dynamics

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 $\rightarrow$ stochastic geometry
- the pattern is described by means of a probability density $\rightarrow$ stochastic modelling
- the probability density allows the computation of average quantities and descriptors (these are integrals) related to the pattern

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 $\rightarrow$ exploratory analysis, model formulation, simulation, inference

- comets example : random fields $\longrightarrow$ another probabilistic mathematical tool
- unifying random fields and marked point processes is a mathematical challenge
- spatio - temporal example : requires new clean and appropriate mathematics, based on both 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

Perfect or exact simulation

Lesson IX

Statistical inference problems

Monte Carlo Maximum likelihood estimation

Parameter estimation based on pseudo-likelihood

Model validation : residual analysis for point processes

Statistical pattern detection

Measure and integration theory $\rightarrow$ blackboard

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

Construction of a point processes

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 Ω :

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

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

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

with $B \in \mathcal{B}$ 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 $\mathbf{w}$ in a observation window W

- a points configuration is $\mathbf{w} = \{w_1, w_2, \ldots, w_n\}$, with n the corresponding number of points
- locally finite : $n(\mathbf{w})$ is finite whenever the volume of W is finite
- simple : $w_i \neq w_j$ for $i \neq j$
- in this course, W is almost always a compact set ...
- from time to time we will consider also the case $W = \mathbb{R}^d$ but this should not worry you ... too much ...

$\rightarrow$ drawing

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 $\mathbf{x} = \{x_n = (w_n, m_n)\}$ *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 : $M = \{1, 2, \ldots, k\}$

... 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 proces X_w, that is

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

for any $w \in W$.
A point process X on W is isotropic if it has the same distribution as the rotated proces $\mathbf{r}X$, that is

$$\{w_1, \ldots, w_n\} \stackrel{\mathcal{L}}{=} \{\mathbf{r}w_1, \ldots, \mathbf{r}w_n\}$$

for any rotation matrix $\mathbf{r}$.

- motion invariant : stationary and isotropic
- marked case : in principle easy to generalize, but take care ...

Intuitive description of a point process : being able to say how many points of the process X can be found in any neighbourhood of 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 within this context 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

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

that are called *void events*.

Definition

The distribution of a point process X is determined by the finite dimensional distributions of its count function, i.e. the joint distribution of $N(B_1), \ldots, N(B_m)$ for any $B_1, \ldots, B_m \in \mathcal{B}$ and $m \in \mathbb{N}$.

Theorem

The distribution of a point process is uniquely determined by its void probabilities

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

Remarks :

- the previous definition is sometimes considered a mathematical result ...
- the proof of the theorem shows first, that the events set $\mathcal{F}$ can be built using the void set made of F_B's. Then, a measure theory argument says that two measures acting in the same way on a first original event set, they will act similarly on a second event set generated from the previous one
- this result is easy to generalize for marked point processes, but with care ...
- the generalization of the previous result can be done also with respect to W : complete and separable (Polish) space

Binomial point process

The trivial random pattern : a single random point x uniformly distributed in 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

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

for Borel subset $B_1, \ldots, B_n$ of W.

$\rightarrow$ 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

$$n = N(W) = n(\mathbf{x}_W)$$

and

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

- 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 process is simple : all points are isolated
- 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*

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

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

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

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

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

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

Stationary Poisson point process

Motivation :

- convergence binomial towards Poisson
- $\rightarrow$ 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 ρ, 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

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

- ρ : 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 $B_1, \ldots, B_k$ are disjoint Borel sets, then $N(B_1), \ldots, N(B_k)$ are independent Poisson variable with means $\rho\nu(B_1), \ldots, \rho\nu(B_k)$.Thus

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

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

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

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

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

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

Mapping theorem

- a Poisson process on W mapped into another space W' by a function ψ is still a Poisson process, provided ψ is measurable and has no atoms
- ψ has no atoms $\leftrightarrow$ ψ does not map several points from W into a single point in W'
- consequence : the projection of a Poisson process is also a Poisson process

- if ψ is a linear transformation, then we have :

Theorem

Linear transformation of a Poisson process. *Let $A : \mathbb{R}^d \to \mathbb{R}^d$ a non-singular mapping. If X is a stationary Poisson process of intensity ρ, then $AX = \{Aw : w \in X\}$ is also a stationary Poisson process and its intensity is $\rho|det(A^{-1})|$, where $det(A^{-1})$ is the determinant of the inverse of A.*

- Spherical contact distribution :

$$\begin{aligned} F(r) &= \mathbb{P}(d(w, X) < r) \\ &= 1 - \mathbb{P}(d(w, X) > r) \\ &= 1 - \mathbb{P}(b(w, r) \cap X = \emptyset) \end{aligned}$$

with $b(w, r)$ the ball centred in $w \in W$ and with radius r. For the stationary Poisson point processes on $W \subset \mathbb{R}^2$, we get

$$F(r) = 1 - \mathbb{P}(N(b(o, r)) = 0) = 1 - \exp(-\rho\pi r^2)$$

Remark : the spherical contact distribution does not completely characterize a point process, because it is obtained for a particular set B

- Conditioning and binomial point processes. Let X be a stationary Poisson point process on W a compact set in $\mathbb{R}^d$, and consider the conditioning $N(W) = n$. The resulting process is a binomial point process with n points.

  This is easily verified, by computing the void probabilities in a compact subset $B \subset W \rightarrow$ Exercice 1 and 2
- a Poisson point process on $\mathbb{R}^d$ is obtained by adding "disjoint" boxes till covering the whole domain ...

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

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

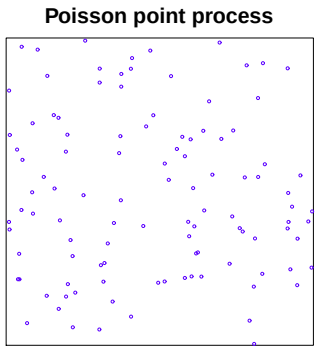
The corresponding probability measure : weighted 'counting" of objects

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

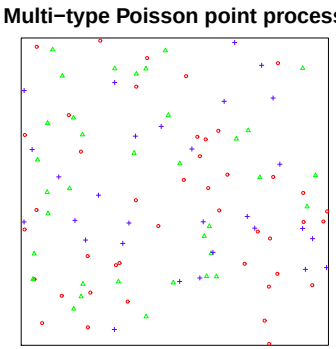
for all $F \in \mathcal{F}$.

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

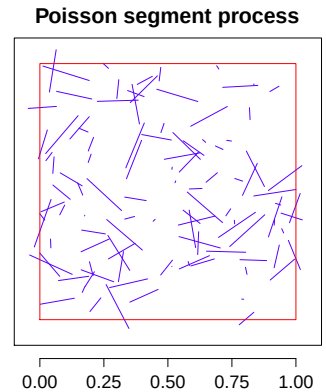
Simulations results of some Poissonian point processes : the domain is $W = [0,1] \times [0,1]$ and the intensity parameter is $\rho = 100$



a)



b)



c)

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

→ Exercice 3

Some general facts concerning the binomial and the Poisson point processes

- the law is completely known $\rightarrow$ analytical formulas,
- independence $\rightarrow$ no interaction $\rightarrow$ no structure in the data ...
- completely random patterns : null or the default hypothesis that we want to reject
- more complicate models can be built $\rightarrow$ specifying a probability density $p(\mathbf{x})$ w.r.t. the reference measure given by the unit intensity Poisson point process. This probability measure is written as

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

with μ 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 $\mathbf{1}_F\{\mathbf{y}\}$ with $p(\mathbf{y})$ ...

Perfect or exact simulation

Lesson IX

Statistical inference problems

Monte Carlo Maximum likelihood estimation

Parameter estimation based on pseudo-likelihood

Model validation : residual analysis for point processes

Statistical pattern detection

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 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

$\rightarrow$ blackboard

More properties of the Poisson process

Stationary Poisson point processes : compute the moment measures, the factorial moment measures and the product densities

- find relations between all these for this process

→ Exercice 4

Definition

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

Proposition

If $X_i \sim Poissson(W, \rho_i)$, $i = 1, 2, \ldots$ 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 $X \sim Poisson(W, \rho)$.

$\rightarrow$ stable character of the Poisson process

Definition

Let be $q : W \to [0, 1]$ 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

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

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* $q(\xi), \quad \xi \in W$ *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* ρ_{thin} *and* $\rho - \rho_{thin}$*, respectively.*

Corollary

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

Remarks : utility of the previous results

- the Poisson process is invariant under independent thinning
- easy procedure for simulate non-stationary Poisson process
- the $n-$th product density measure of an independently thinned point process is

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

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

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

but watch out ...

$\rightarrow$ Exercice 5 : explain the envelope tests

Campbell moment measures

Present context :

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

Definition

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

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

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

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

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

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

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

from which we can get the second order moment measure

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

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 $h(x, \mathbf{x})$ measured in the location of a point process X : the function h may depend on X

Campbell - Mecke formula

Theorem

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

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

Proof.

$\to$ blackboard $\square$

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

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

and

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

Proof.

Follow the same proof scheme as previously. □

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

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

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

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

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

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

Perfect or exact simulation

Lesson IX

Statistical inference problems

Monte Carlo Maximum likelihood estimation

Parameter estimation based on pseudo-likelihood

Model validation : residual analysis for point processes

Statistical pattern detection

Interior and exterior conditioning

Present context :

- counting points measures : count the points in a small neighbourhood and then integrate using the Campbell Mecke formula $\rightarrow$ the small neighbourhood is a small region in W
- idea : the "same" counting points measures $\rightarrow$ the small neighbourhood is a small region in W "centred" in a point of the process X $\rightarrow$ interior conditionning
- 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 conditionning

A review of the Palm theory

- construction $\rightarrow$ 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

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

- for stationary point processes

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

Slivnyak - Mecke theorem

Theorem

If $X \sim Poisson(W, \rho)$, then for functions $g : W \times \Omega \to [0, \infty)$, 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 : $\to$ blackboard
- generalization : rather easy ... (the same comment as for the Campbell - Mecke theorem)

- 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 telecomunications : knowing, in this location I have a mobile phone antena, how the quality of the signal change if I add randomly more antenas ? (the group of F. Baccelli)

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

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

- 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 Υ 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

$\rightarrow$ Exercice 6 and 7

Assumption : X is a stationary point process

- the nearest neighbour distance distribution function

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

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 :

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

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

$$\begin{aligned} F(r) &= 1 - \exp[-\rho\pi r^2] \\ G(r) &= F(r) \\ J(r) &= 1 \end{aligned}$$

$\rightarrow$ Exercice 8

Reduced Palm distributions
Campbell and Slivnyak theorems
Summary statistics : the K and L functions
Exterior conditioning : conditional intensity

Lesson VI

Cox processes
Cluster processes
Boolean model
Capacity functional.Choquet theorem

Lesson VII

Probability density of a point processes
Interacting marked point processes
Markov point processes

Lesson VIII

Monte Carlo simulation
Markov chains : a little bit of theory
Metropolis-Hastings algorithm
MH algorithm for sampling marked point processes
Spatial birth-and-death processes

Perfect or exact simulation

Lesson IX

Statistical inference problems

Monte Carlo Maximum likelihood estimation

Parameter estimation based on pseudo-likelihood

Model validation : residual analysis for point processes

Statistical pattern detection

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 follows 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

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

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

- the analogue of Campbell-Mecke formula reads

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

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

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

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

- the function $P^!_\cdot(F)$ is defined uniquely up to an $\mu^{(1)}-$null set
- it is possible to find a version such that for fixed $w \in W$, $P^!_w(\cdot)$ 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

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

- for stationary point processes

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

- 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 $\rightarrow$ 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 :

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

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

- 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 σ_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

$$L(r) = r$$

- theoretical explanations $\rightarrow$ blackboard

Exterior conditioning : conditional intensity

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

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

  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 $\mathbf{x}$, $\Lambda(\cdot; \mathbf{x})$ is a locally finite Borel measure $\rightarrow$ the first order Papangelou kernel

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

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

- 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 $p(\mathbf{x})$ with respect to a Poisson process with non-atomic finite intensity measure ν. Then X has Papangelou conditional intensity

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

for $u \notin \mathbf{x} \in \Omega$.

Proof.

$\rightarrow$ blackboard

Importance of the conditional intensity :

- intuitive interpretation :

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

  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 $\mathbf{x}$ 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

$$\lambda(u; \mathbf{x}) = \lambda(u; \emptyset)$$

for all patterns $\mathbf{x}$ satisfying $\mathbf{x} \cap b(u, r) = \emptyset \rightarrow$ 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 $\rightarrow$ blackboard

$\rightarrow$ Exercice 9

Reduced Palm distributions
Campbell and Slivnyak theorems
Summary statistics : the K and L functions
Exterior conditioning : conditional intensity

Lesson VI

Cox processes
Cluster processes
Boolean model
Capacity functional.Choquet theorem

Lesson VII

Probability density of a point processes
Interacting marked point processes
Markov point processes

Lesson VIII

Monte Carlo simulation
Markov chains : a little bit of theory
Metropolis-Hastings algorithm
MH algorithm for sampling marked point processes
Spatial birth-and-death processes

Perfect or exact simulation

Lesson IX

Statistical inference problems

Monte Carlo Maximum likelihood estimation

Parameter estimation based on pseudo-likelihood

Model validation : residual analysis for point processes

Statistical pattern detection

Some main ideas to be retained till now

- counting measures $\rightarrow$ summary statistics for point pattern characterization
- two categories : interpoint distances (F, G and J) and second order characteristics (ρ, K and L)
- possible extension of the summary statistics : marks, non-stationary processes, different observation spaces W case and spatio-temporal $\rightarrow$ some of these topics are already solved, others are rather hot research topics
- non-parametrical estimation of the summary statistics : kernel estimation and management of the border effects + numerical sensitivity $\rightarrow$ not presented in this course, but Enn and his team are experts of this domain ...
- central limit available : statistical tests

- summary statistics for parameter estimation of the undergoing model $\Rightarrow$ 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)
- good exploring tool : **spatstat** provides also some 3d estimators ...
- outline important aspects of a point pattern : clustering, repulsion, completely randomness

$\rightarrow$ real need for models able to reproduce these characteristics

$\rightarrow$ counting is not always obvious ...

Cox processes

Definition

Let Υ be a random locally finite diffuse measure on $(W, \mathcal{B})$. If the conditional distribution of X given Υ is a Poisson process on W with intensity measure Υ, X is said to be a Cox point process with driving measure Υ. 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 $Z = \mathbf{z}$ is a distribution of the Poisson process with intensity function $\mathbf{z} \Rightarrow$

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

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

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

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

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

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

- if $\rho^{(2)}(u,v) = \mathbb{E}Z(u)Z(v)$ exists, then it is the product density
- proof : use the results from Exercice 4 and the total expectation law

- the pair correlation function is

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

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

- the variance $VarN(B)$ is obtained using the total variance law, and it is

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

$\Rightarrow$ over - dispersion of the Cox process counting variables

- the void probabilities of Cox processes are

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

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 Π
- $\Rightarrow X_{thin}$ is a Cox process driven by $Z_{thin}(w) = \Pi(w)Z(w)$

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 $X_c, \quad c \in C$ are independent. Then X is called a Poisson cluster point process.

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 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 α, then X is a Neyman-Scott Poisson process.

$\rightarrow$ drawing + Exercice 10

Theorem

Let X be a Neyman-Scott Poisson process such that C is a stationary Poisson process with intensity κ. 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 which have density k.

Proof.

→ blackboard

□

Matérn cluster process (Matérn 1960,1986)

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

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

$\rightarrow$ drawing + show on your computer Exercice 11 + data sets (redwoodfull, japanesepines, celss)

Remarks :

- usually in applications Z is unobserved
- we cannot distinguish a Cox process X from its corresponding Poisson process $X|Z$ when only one 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

Boolean model

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

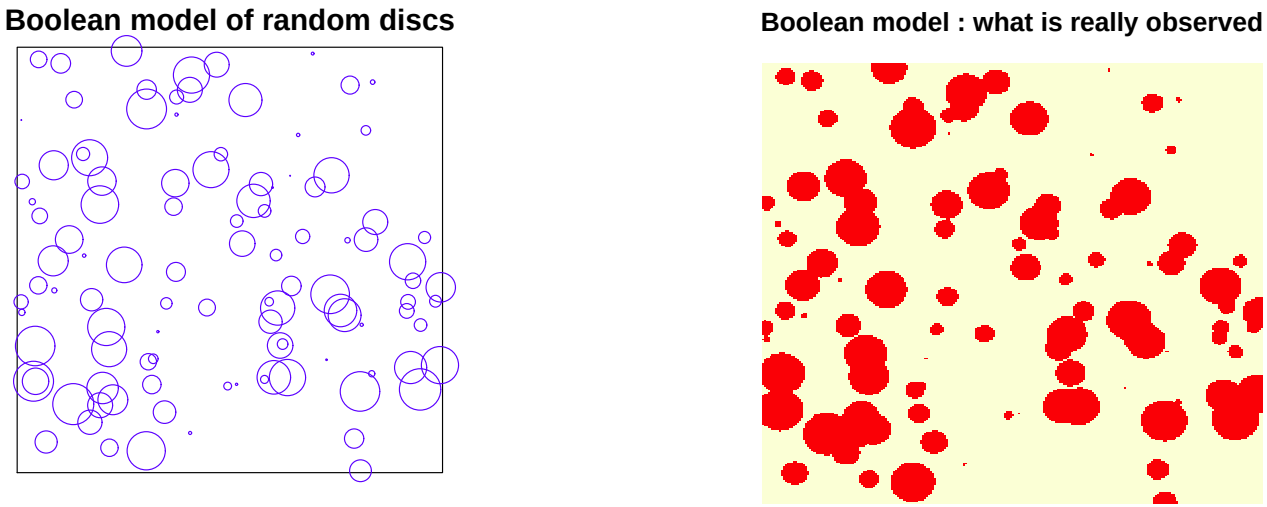
- germs : a stationary Poisson point process X of intensity ρ on $\mathbb{R}^d$
- grains : a sequence of i.i.d. random compact sets $\Gamma_1, \Gamma_2, \ldots$ and independent of X

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

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

The random set Γ_0 is said to be the typical grain. The set Γ 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



a) b)

Figure: Boolean model of random discs : complete and incomplete views.

Remarks :

- classical references : Matheron (1975), Molchanov (1997), Lantuéjoul (2002), Chiu et al. (2013)
- important practical applications $\rightarrow$ one of the first models of complex patttern
- no structure $\leftrightarrow$ no objects interactions
- Neyman-Scott processes may be seen as Boolean models as well ...

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 Γ *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 Γ *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 Γ *is*

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

- the reflection of the typical grain :

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

- the Minkowski addition :

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

Proof.

$\rightarrow$ blackboard

Basic characteristics of the Boolean model

- the volume fraction
- covariance
- contact distribution

$\rightarrow$ blackboard

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.

$\rightarrow$ blackboard $\square$

Perfect or exact simulation

Lesson IX

Statistical inference problems

Monte Carlo Maximum likelihood estimation

Parameter estimation based on pseudo-likelihood

Model validation : residual analysis for point processes

Statistical pattern detection

Probability density of a point processes

Present context :

- the independence property of the Poisson based processes does not allow to introduce point interactions
- interactions can be introduced by means of a probability measure w.r.t a Poissonian reference measure μ
- the distribution of such a point process writes as

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

- let μ be the standard unit intensity Poisson process
- the point process distribution w.r.t μ writes as

$$\mathbb{P}(X \in F) = \\ = \sum_{n=0}^{\infty} \frac{\exp[-\nu(W)]}{n!} \int_W \cdots \int_W \mathbf{1}(\{w_1, \ldots, w_n\} \in F) \times \\ p(\{w_1, \ldots, w_n\}) d\nu(w_1) \ldots d\nu(w_n),$$

  whenever $n > 0$. If $n = 0$, we take $\exp[-\nu(W)]\mathbf{1}(\emptyset \in F)p(\emptyset)$. 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 ν_M

- usually the probability density is known up to a constant : $p \propto h$
- the normalizing constant or the partition function is given by

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

that becomes

$$\alpha = \sum_{n=0}^{\infty} \frac{\exp[-\nu(W)]}{n!} \int_W \cdots \int_W h(\{w_1, \ldots, w_n\}) d\nu(w_1) \ldots d\nu(w_n) \tag{2}$$

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

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

$$p(\mathbf{w}) \propto \prod_{w_i \in \mathbf{w}} \rho(w_i).$$

Let $\Upsilon(B) = \int_B \rho(w) d\nu(w)$ be the associated intensity measure.
We asssume $0 < \Upsilon(B) < \infty$ for any $B \subseteq W$.
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

$$p(\mathbf{w}) = \exp[\nu(W) - \Upsilon(W)] \prod_{w_i \in \mathbf{w}} \rho(w_i)$$

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

$$p(\mathbf{w}) = \exp[(1-\rho)\nu(W)]\rho^n$$

Interacting marked point processes

Construction of the probability density :

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

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

  for any $k-$tuplet of objects

- the density is the product of all these functions

$$p(\mathbf{x}) = \alpha \prod_{x_i \in \mathbf{x}} \phi(x_i)^{(1)} \cdots \prod_{\{x_{i_1}, \ldots, x_{i_k}\} \in \mathbf{x}} \phi(x_{i_1}, \ldots, x_{i_k})^{(k)} \qquad (3)$$

- α the normalizing constant is not known

- the probability densities (3) are suitable for modelling provided they are integrable on Ω ; that is

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

- 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 with probability density p w.r.t the reference measure μ. The process X is stable in the sense of Ruelle, if it exists $\Lambda > 0$ such that

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

Proposition

The probability density of a stable point process is integrable.

Proof.

The integrability of $p(\mathbf{x})$ follows directly from the preceding condition :

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

□

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*

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

Proposition

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

Proof.

It is easy to show by induction that

$$p(\mathbf{x}) = p(\emptyset)\Lambda^{n(\mathbf{x})}, \quad \forall \mathbf{x} \in \Omega.$$

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

- the local stability implies the hereditary condition

$$p(\mathbf{x}) = 0 \Rightarrow p(\mathbf{y}) = 0, \quad \text{if} \quad \mathbf{x} \subseteq \mathbf{y}.$$

- this condition allows the definition of the conditional intensity as

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

  assuming $0/0 = 0$
- the conditional intensity is also known in the literature as the Papangelou intensity condition (we have already meet it)

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 $\mathbf{x} \subseteq \mathbf{y}$ implies

$$\lambda(\eta; \mathbf{x}) \leq \lambda(\eta; \mathbf{y}),$$

and repulsive otherwise

$$\lambda(\eta; \mathbf{x}) \geq \lambda(\eta; \mathbf{y}),$$

- 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

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

- 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 $\mathbf{x} \in \Omega$ *such that* $\eta \sim \zeta$ *for all* $\eta, \zeta \in \mathbf{x}$. *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 μ. The process X is Markov if for all $\mathbf{x} \in \Omega$ such that $p(\mathbf{x}) > 0$, the following conditions are simultaneously fulfilled :

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

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

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

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

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

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

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

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

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

Proof.

→ blackboard

□

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

Theorem

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

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

for all $\mathbf{x} \in \Omega$.

Proof.

$\to$ blackboard

□

Remarks :

- the previous result simplifies the writing of the probability density of an interacting point process
- taking $\phi_c(\mathbf{y}) = 1$ whenever $\mathbf{y}$ 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

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

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

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

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)

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

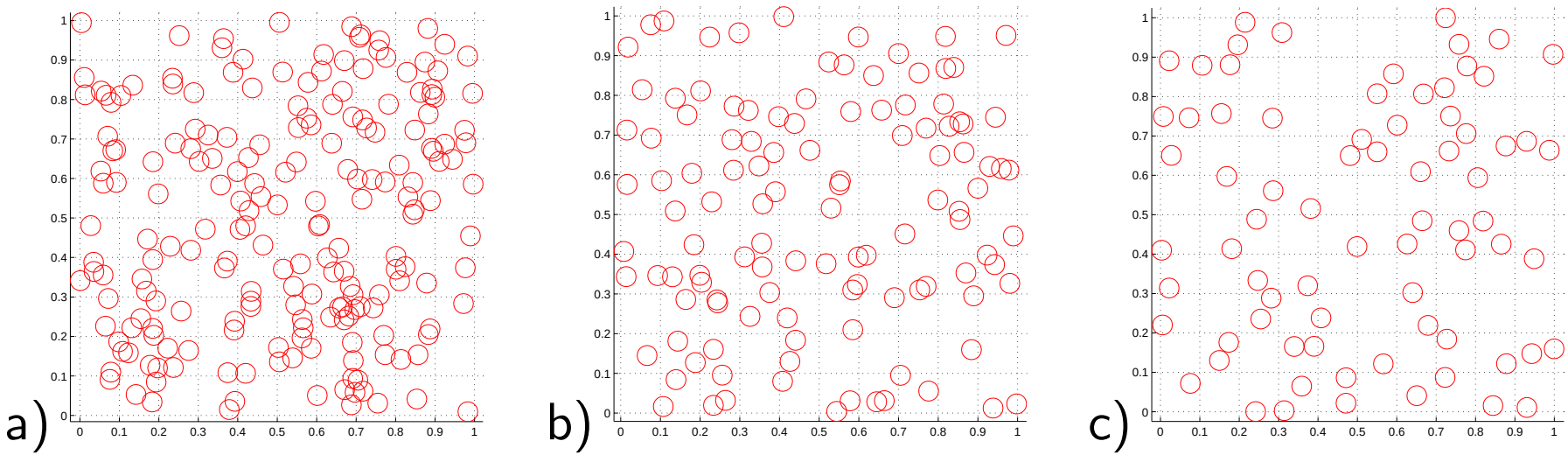


Figure: Strauss model realisations for different parameter values : a) $\gamma = 1.0$, b) $\gamma = 0.5$ and c) $\gamma = 0.0$.

The interaction function $\gamma : W \times W \to [0, 1]$ 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 η to $\mathbf{x} \setminus \{\eta\}$ is

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

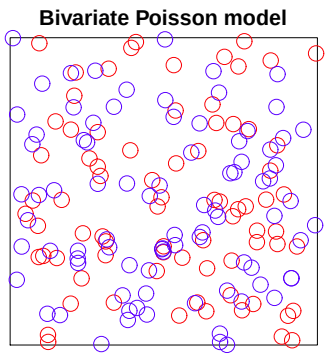
where $\partial(u) = \{v \in \mathbf{x} : d(u, v) \leq r\}$

The Strauss model is a locally stable model with $\Lambda = \beta$ and Markov with interaction range r.
The interaction functions applied to cliques are

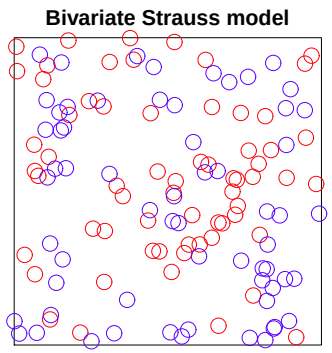
$$\begin{aligned} \phi_c(\emptyset) &= \alpha \\ \phi_c(\{u\}) &= \beta \\ \phi_c(\{u, v\}) &= \gamma(u, v) \end{aligned}$$

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

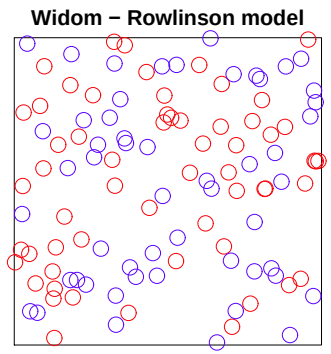
Multi-type pairwise interaction processes



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 0.25.

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

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

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 $(w, 1) \notin \mathbf{x}$ to the configuration $\mathbf{x}$ is

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

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, $\lambda((w,m);\mathbf{x}') \geq \lambda((w,m);\mathbf{x})$ for all $\mathbf{x}' \subseteq \mathbf{x}$ and $(w,m) \in W \times M$.
The interaction functions are

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

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

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

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

- the parameters $\beta_m > 0$, $m \in M$ control the intensity of the points of type m.
- the measurable functions $\gamma_{ij} : [0, \infty) \to [0, 1]$ 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 $(w, m) \notin \mathbf{x}$, the conditional intensity is

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

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

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

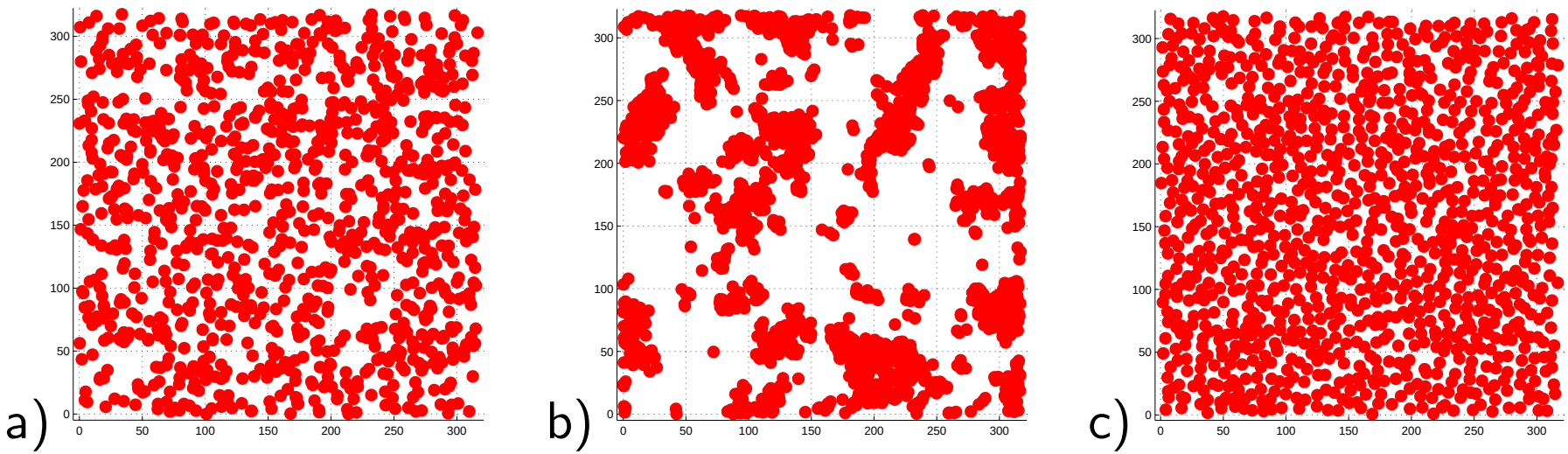


Figure: Area interaction model realisations for different parameter values : a) $\gamma = 1.0$, b) $\gamma > 1.0$ and c) $\gamma < 1.0$.

Remarks :

- the first probability density based point process producing clusters $\rightarrow$ 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.

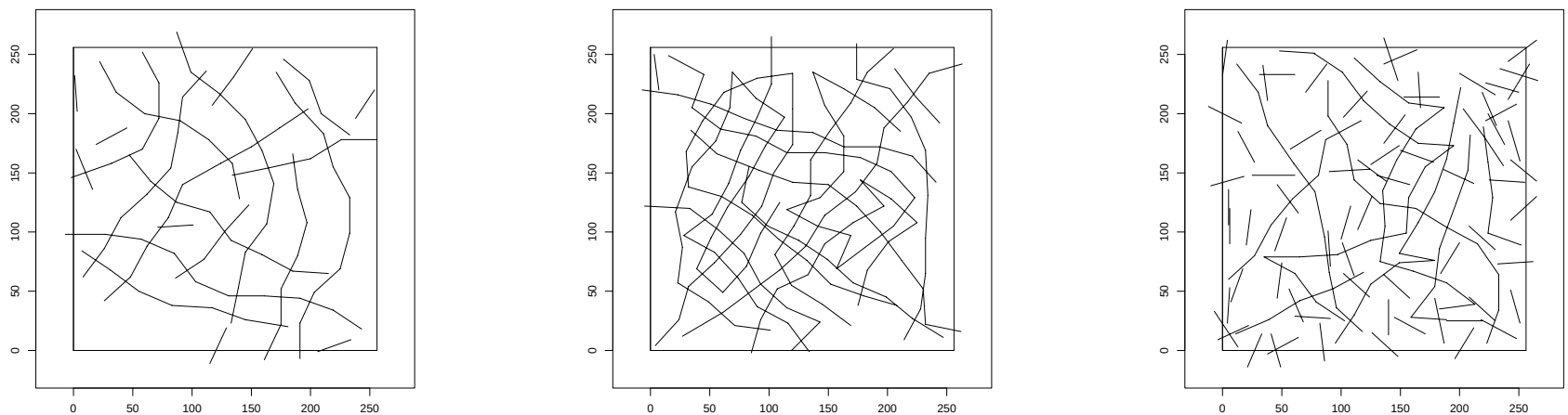
$\rightarrow$ blackboard $\square$

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)

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

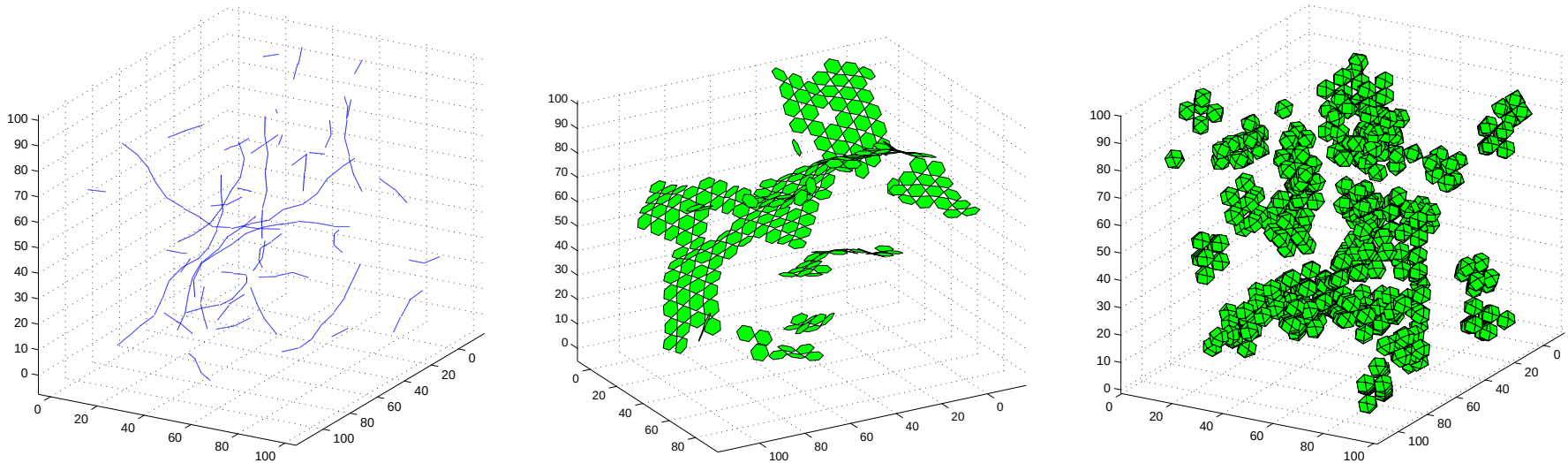


Figure: Random shapes generated with Bisous model.

Remarks :

- Candy and Bisous are based on compound interactions $\rightarrow$ 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 $\rightarrow$ kissing number
- $\rightarrow$ 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 current undergoing work with M. N. M. van Lieshout and classical literature in mathematical morphology

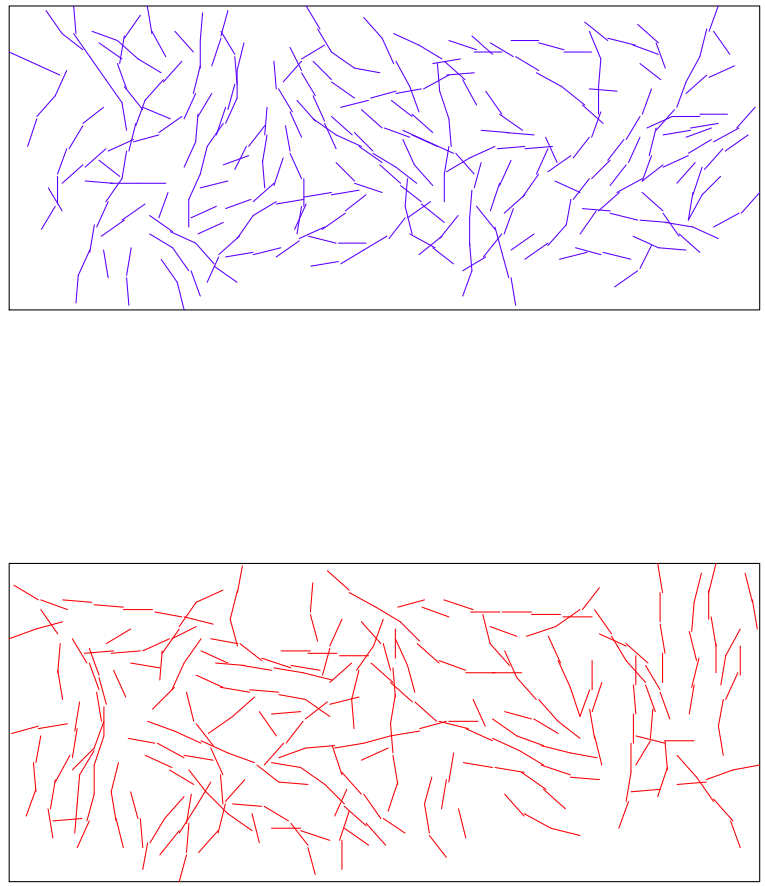


Figure: Realizations of the Candy model obtained with different samplers.

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

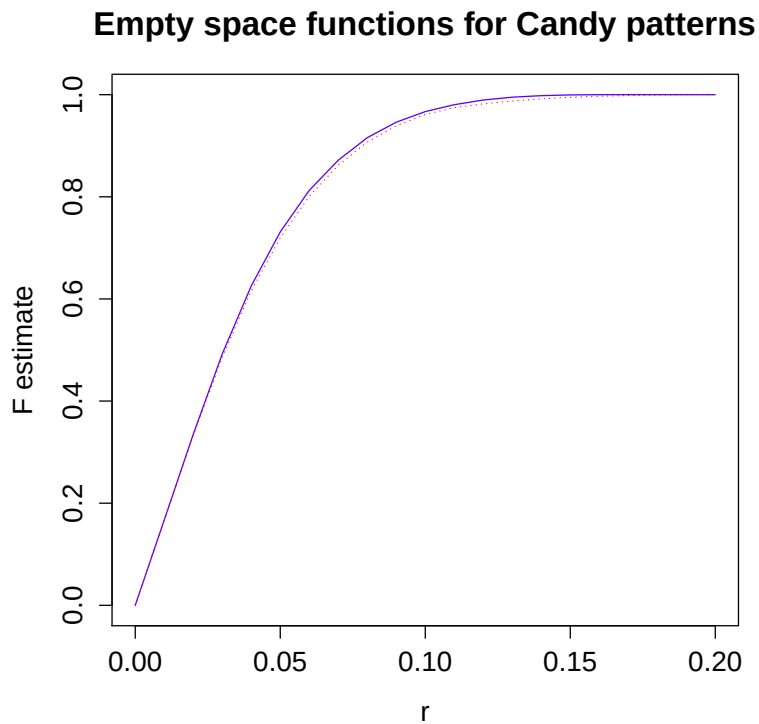


Figure: Estimation of the empty space function for the previous Candy realizations

J function for multi-type segment pattern : undergoing work with Marie-Colette van Lieshout

- undergoing work with Marie-Colette van Lieshout (CWI Amsterdam) and a group of people from CIRAD Montpellier
- show : R script testJFuncData.R

Open questions :

- may the F function be used to characterize and to compare different filamentary patterns ?
- may a general J function be used to characterise the interactions of filaments, clusters and walls ?
- temporal behaviour ? if yes, for what type of model ?
- planar and cluster patterns ?

Perfect or exact simulation

Lesson IX

Statistical inference problems

Monte Carlo Maximum likelihood estimation

Parameter estimation based on pseudo-likelihood

Model validation : residual analysis for point processes

Statistical pattern detection

Markov chain Monte Carlo algorithms

Problem : sampling or simulation probability distributions

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

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

Remarks :

- almost all the point process models and Markov random fields
- exceptions : Poisson processes and also permanental and determinental point processes
- Markov chain theory and simulation is whole domain in probability and statistics $\rightarrow$ very general working framework

Basic MCMC algorithm

Algorithm

$\mathbf{x} = \textit{My first MCMC sampler}\,(T)$

1. *choose an initial condition* $\mathbf{x}_0$
2. *for* $i = 1$ *to* T, *do*
$$\{ \quad \mathbf{x}_i = \textit{Update}(\mathbf{x}_{i-1}) \quad \}$$
3. *return* $\mathbf{x}_T$.

Principles of the MCMC algorithm :

- simulates a Markov chain
- the *Update* function reproduces the transition kernel of the Markov chain
- the output $\mathbf{x}_T$ is asymptotically distributed according to π whenever $T \to \infty$
- if the simulated Markov chain has good properties $\to$ 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 :

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

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

Transition kernel : it is the "engine" of the Markov chain, that is a mapping $P : \Omega \times \mathcal{F} \to [0, 1]$ 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 $x \in \Omega$

Invariant distribution : the probability distribution π that satisfies

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

- interpretation : the transition kernel changes of the current state, but the new state is distributed according to π

Reversible chain : the probability of going from A to B equals the probability of going from B to A

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

The reversibility implies invariance. Indeed, since $P(x,\Omega) = 1$ and considering $B = \Omega$ in the reversibility equations, we get

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

Equilibrium distribution : the invariant distribution is an equilibrium distribution if and only if :

$$\lim_{n \to \infty} P^n(x, A) = \pi(A)$$

for all measurable sets $A \in \mathcal{F}$ and any $x \in \Omega$; P^n is the $n-$th application of the transition kernel

Markov chains : convergence properties

Important : key elements whenever building a transition kernels

- *aperiodicity :* no deterministic loops
- *irreducibility :* the chain can go from any state to any other state
- *recurrence :* the chain can go from any state to any other state "often enough" $\rightarrow$ independence of the initial conditions
- *ergodicity :* the chains distribution converge towards its equilibrium distribution "fast enough" (an ergodic chain is recurrent ..)

Proposition

The invariant distribution of an aperiodic and irreducible chain is unique and it is also the equilibrium distribution of the chain.

Remark : the previous result is verified, for all starting points $\mathbf{x} \in \Omega'$ such that $\pi(\Omega \setminus \Omega') = 0$. Hence, the convergence depend on the initial conditions. The recurrence property banishes these null-sets.

Proposition

The ergodic chain reaches the equilibrium regime, fast enough, from any initial state. The large numbers law and the central limit theorem can be used whenever sampling with an ergodic chain.

Remark : the speed of convergence of the chain may be the same for all the initial conditions, the chain is uniformly ergodic. If the speed of convergence depends on the starting a point, we may have a geometrically ergodic chain

- drift condition : technical mathematical tool for establishing convergence properties of a Markov chain

Metropolis-Hastings algorithm

Principle :

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

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

  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

$\rightarrow$ blackboard + Exercise 12 + comment Elmo

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

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

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

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

- acceptance probability

$$\alpha(\mathbf{x} \to \mathbf{x} \cup \{\zeta\}) = \min\left\{1, \frac{p_d p(\mathbf{x} \cup \{\zeta\})}{p_b p(\mathbf{x})} \times \frac{\nu(W)}{n(\mathbf{x}) + 1}\right\} \quad (10)$$

Death : remove an object

- the inverse movement of birth
- acceptance probability

$$\alpha(\mathbf{x} \to \mathbf{x} \setminus \{\zeta\}) = \min\left\{1, \frac{p_b p(\mathbf{x} \setminus \{\zeta\})}{p_d p(\mathbf{x})} \times \frac{n(\mathbf{x})}{\nu(K)}\right\} \quad (11)$$

Remark : note the appearance of the Papangelou intensity in the acceptance probability $\Rightarrow$ local stability property guarantees good convergence properties of the Markov chain

A transition kernel doing these transformations is

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

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 $b(\mathbf{x}, \eta) = \frac{1}{\nu(W)}$ and the death rate is $d(\mathbf{x}, \eta) = \frac{1}{n(\mathbf{x})}$

Algorithm

$\mathbf{y} = \textit{Update}(\mathbf{x})$

1. *Choose "birth" or "death" with probabilities* p_b *and* p_d*, respectively.*
2. *If "birth" was chosen, then generate a new object following* $b(\mathbf{x}, \eta)$*. Accept the new configuration,* $\mathbf{y} = \mathbf{x} \cup \{\eta\}$ *with the probability* $\alpha(\mathbf{x}, \mathbf{y})$ *given by* (10)*.*
3. *If "death" was chosen, then select the object to be removed using* $d(\mathbf{x}, \eta)$*. Accept the new configuration,* $\mathbf{y} = \mathbf{x} \setminus \{\eta\}$ *with the probability* $\alpha(\mathbf{x}, \mathbf{y})$ *given by* (11)*.*
4. *Return the present configuration.*

Theorem

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

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

Fix $p_b, p_d \in (0, 1)$ with $p_b + p_d \leq 1$ and let $p(\mathbf{x})$ be the probability density of a marked point process on $W \times M$. The point process is locally stable and $p(\mathbf{x})$ is built w.r.t the standard Poisson process μ. 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.

Remark :

- the same result holds if change moves are introduced with care ... $\rightarrow$ 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

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

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

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

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

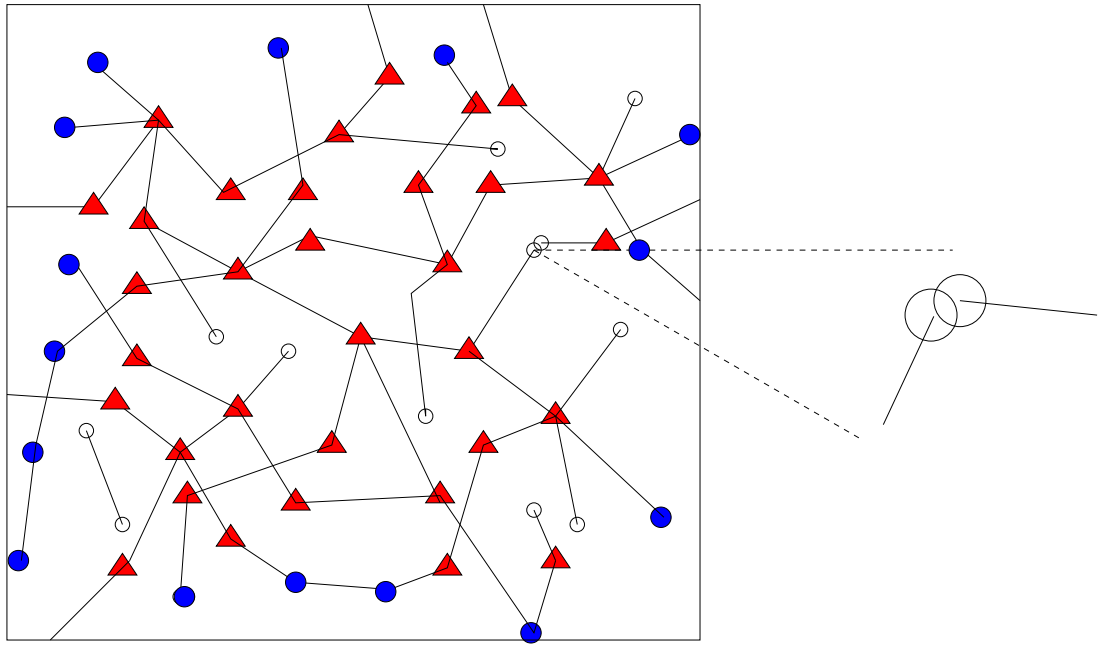
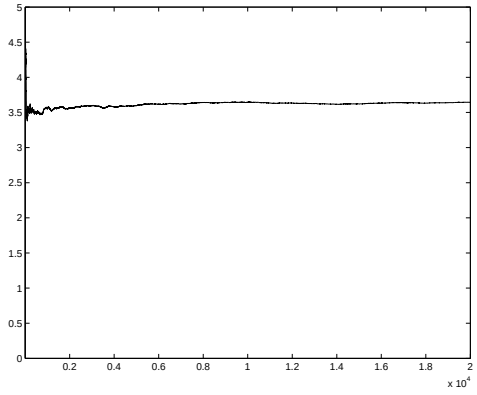
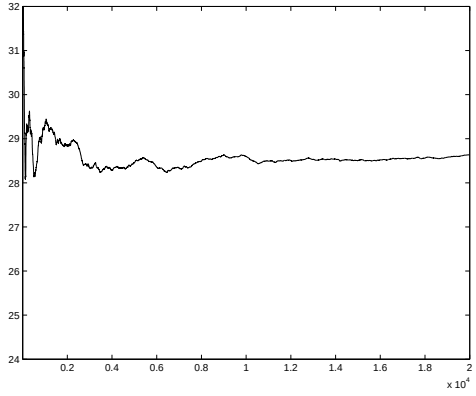


Figure: Extremities marked by triangles are connected and further than $\frac{1}{2}l_{\max} + r_c$ to the boundary, those labeled by a black disk are closer than $\frac{1}{2}l_{\max} + r_c$ to the boundary of K.

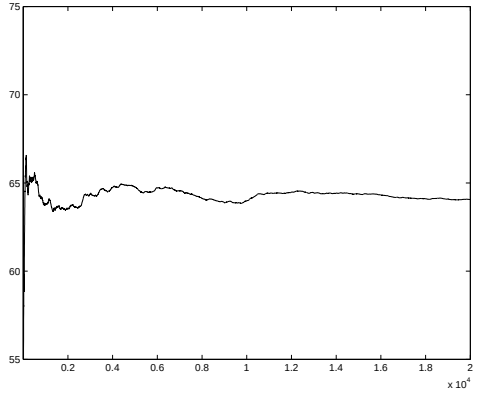
MH algorithm for sampling the Candy model : dynamics behaviour through the sufficient statistics analysis



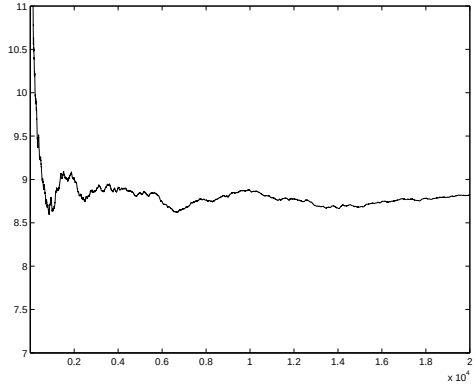
$\bar{n}_f = 3.64$



$\bar{n}_s = 28.63$



$\bar{n}_d = 64.07$



$\bar{n}_o = 8.82$

- great adaptability and theoretical convergence
- easy to use
- but appropriate solutions need to be found for each new problem or situation ...
- the dynamics has to be built depending on the model
- the general framework, even if having good theoretical properties, it is not always the most efficient from a numerical point of view ...
- but if you do not have good theoretical properties, your results will be bad anyway ...

Spatial birth-and-death processes

Theoretical background : continuous time Markov chains $\rightarrow$ 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 very 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 Λ.

1. *Simulate a Poisson process on $W \times M$ with intensity function $\rho = \Lambda$. Let $\mathbf{x}_\Lambda$ be the obtained configuration.*
2. *Set $\mathbf{x} = \emptyset$ the realisation of p, and continue*
3. *For every point η in $\mathbf{x}_\Lambda$ do*
   - *the point is added to $\mathbf{x}$: $\mathbf{x} = \mathbf{x} \cup \{\eta\}$ with probability*

$$\frac{\lambda(\eta; \mathbf{x}_\Lambda)}{\Lambda}$$

   - *otherwise the point is deleted from $\mathbf{x}_\Lambda$: $\mathbf{x}_\Lambda = \mathbf{x}_\Lambda \setminus \{\eta\}$*
4. *Return $\mathbf{x}$*

Remarks :

- this algorithm has all the desired convergence properties
- comparing with MH algorithm, this algorithm may easily produce independent samples ...
- nevertheless, for strong interactions Λ may be very high
- Candy and Bisous models with strong interactions cannot be simulated with this algorithm

Perfect or exact simulation

- classical MCMC methods $\rightarrow$ 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 π defined on the discrete state space $\Omega = \{\omega_1, \omega_2, \ldots, \omega_m\}$. The induced Markov chain (X_n) is represented by its transition functiones $\phi(\cdot, \cdot)$ such that

$$X_{n+1} = \phi(X_n, V_n), \tag{12}$$

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_ns
- if at a certain moment $n \in -T, \ldots, 0$ all the chains are in the same state or they coalesced, that is

$$X_n \equiv \mathbf{x}$$

  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 π

$\rightarrow$ blackboard drawing

Extraordinary smart idea :

- launching m parallel chains is not always feasible
- if Ω can be ordered

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

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).

Perfect algorithms for sampling marked point processes

Previous ideas hold for perfect sampling of marked point processes

CFTP based algorithms :

- spatial birth-and-death and clan of ancestors : thinning principles
- Metropolis-Hastings
- locally stable point processes
- monotonic and anti-monotonic point processes
- dominating process : stationary Poisson $\rightarrow$ (CFTP and clan of ancestors), Λ parameter is important
- the transition function ingredients : conditional intensity, Λ and the V_ns

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)

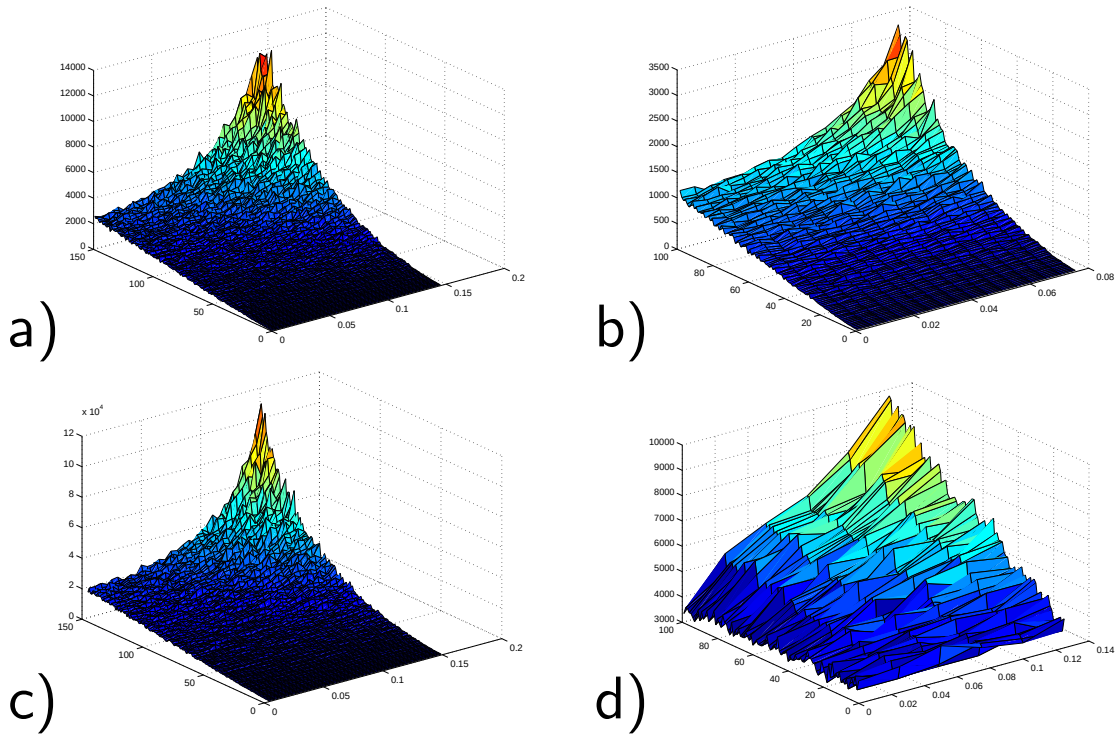


Figure: Exact simulation algorithms applied to Strauss model : a) CFTP, b) clan of ancestors, c) Metropolis-Hastings and d) Gibbs.

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

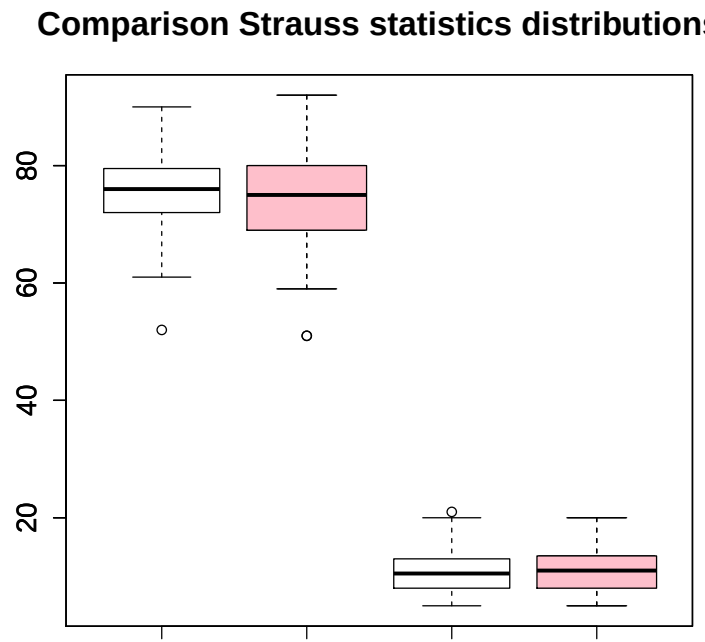


Figure: Boxplots comparison for the n and s_r statistics distributions : white - the distributions obtained using the exact algorithm, pink (dark couloured) - the distributions obtained using the Metropolis - Hastings algorithm.

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

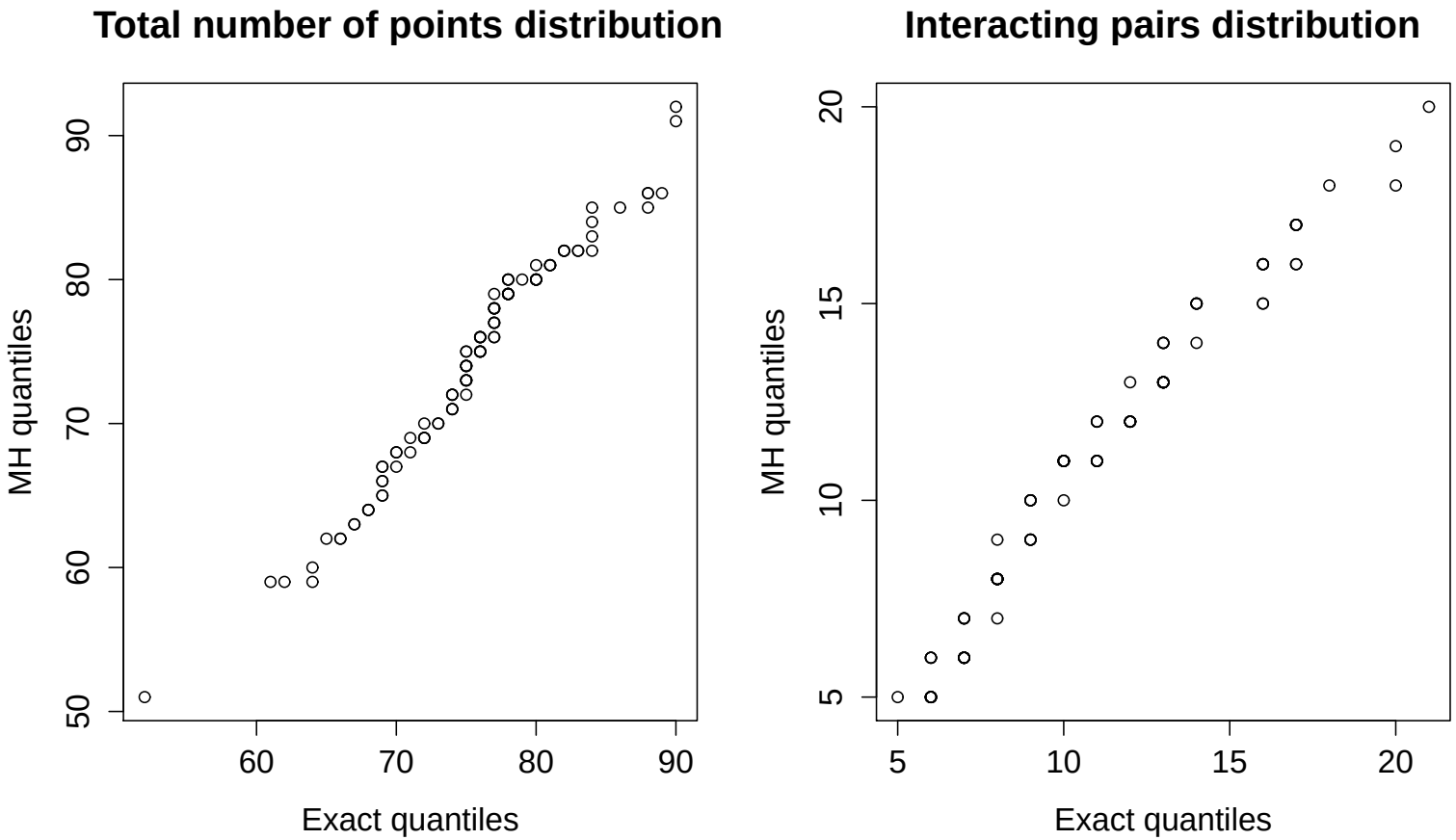


Figure: Qqplot comparison for the n and s_r statistics distributions.

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 $\rightarrow$ exact algorithms able to be tailored to the model

Lesson IX

- Statistical inference problems
- Monte Carlo Maximum likelihood estimation
- Parameter estimation based on pseudo-likelihood
- Model validation : residual analysis for point processes
- Statistical pattern detection

Statistical inference problems

Problem I : parameter estimation

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

Problem II : pattern detection

- observe the data $\mathbf{d}$ and find $\mathbf{x}$ "hidden"
- the model parameters are : hidden, modelled, unknown
- open problem : the detected pattern does it really exist ...?

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

- shape modelling $\rightarrow$ "crystal ball" ?
- observe a phenomenon and propose a model doing the "same" ...
- needs the time dimension
- open problem : time,
- what it is the time "quanta" $\rightarrow$ generating element and interactions as for marked point processes ?

Monte Carlo Maximum likelihood estimation

Exponential family models :

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

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

where $t(\mathbf{x})$ and θ represent the sufficient statistics vector and the model parameters vector, respectively. The normalising constant $Z(\theta)$ is unknown.

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

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

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

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

since we have

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

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 $\mathbb{X}_1, \mathbb{X}_2, \ldots, \mathbb{X}_n$ are samples obtained from $p(\mathbf{y}|\psi)$.
Hence, the Monte-Carlo counterpart of the log-likelihood is :

$$l_n(\theta) = \langle t(\mathbf{x}), \theta - \psi \rangle - \log \left(\frac{1}{n} \sum_{i=1}^{n} \exp\langle t(\mathbb{X}_i), \theta - \psi \rangle \right).$$

Theorem

The log-likelihood of an exponential family model is a convex function.

- proof : see (Monfort 1997, Thm.3, pp. 61)
- $l_n(\theta) \to l(\theta)$ almost sureley
- all these suggest that local optimisation procedures applied to $l_n(\theta)$ may give interesting results

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 :

$$-\nabla^2 l_n(\theta) = \mathbb{V}ar_{n,\theta,\psi}[t(\mathbb{X})]$$

where

$$\mathbb{V}ar_{n,\theta,\psi}[t(\mathbb{X})] = \mathbb{E}_{n,\theta,\psi}[t(\mathbb{X})t(\mathbb{X})^t] - \mathbb{E}_{n,\theta,\psi}[t(\mathbb{X})]\mathbb{E}_{n,\theta,\psi}[t(\mathbb{X})^t].$$

Newton-Raphson method :

$$\theta_{k+1} = \theta_k - [\nabla^2 l_n(\theta_k)]^{-1} \nabla l_n(\theta_k) \tag{13}$$

for $k = 1, 2, \ldots,$

- $l_n(\cdot)$ is computed using n samples from $p(\mathbf{x}|\psi)$
- the computation of the gradient and Hessian inverse is numerically unstable
- useful only if the initial value is close enough from the solution

Stochastic gradient :

$$\theta_{k+1} = \theta_k + \epsilon_k[t(\mathbf{x}) - t(\mathbb{X}_k)]$$

where $\epsilon_k > 0$ is a decreasing sequence while $\mathbb{X}_k$ is a sample of $p(\mathbf{x}|\theta_k)$

- 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
- joint pattern detection and parameter estimation

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 $\| \theta_k - \psi \| >$ threshold
- obtain a reference value θ_0 close enough to the maximum likelihood estimator

Asymptotic results

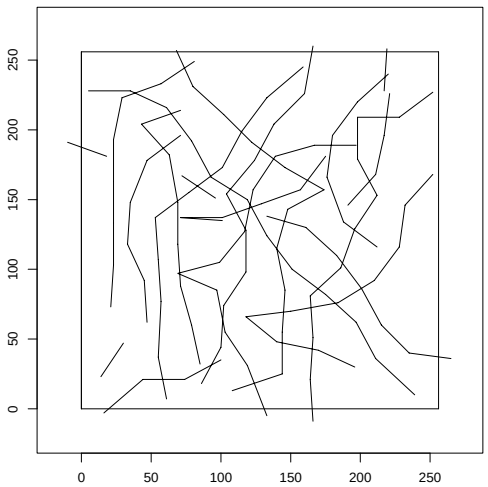
The random variable $\sqrt{n}(\widehat{\theta}_n - \widehat{\theta})$ whenever $n \to \infty$, it converges in distribution towards a normal random variable of zero mean and variance $I(\widehat{\theta})^{-1}\Gamma I(\widehat{\theta})^{-1}$:

$$\sqrt{n}(\widehat{\theta}_n - \widehat{\theta}) \to \mathcal{N}(0, I(\widehat{\theta})^{-1}\Gamma I(\widehat{\theta})^{-1}).$$

- the matrix
$$I(\widehat{\theta}) = \mathbb{V}ar_{\widehat{\theta}}[t(\mathbb{X})] = -\nabla^2 l(\widehat{\theta})$$
is the Fisher information of $\widehat{\theta}$
- the matrix Γ is the matrix of asymptotic covariance of the normalised Monte Carlo gradient $\sqrt{n}\nabla l_n(\widehat{\theta})$
- an approximation of these matrices may be found in (van Lieshout and Stoica, 2003)
- under some assumptions, the error $\widehat{\theta} - \theta_0$ can be estimated by the diagonal of the inverse of $-\nabla^2 l_n(\widehat{\theta}_n)$

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 :

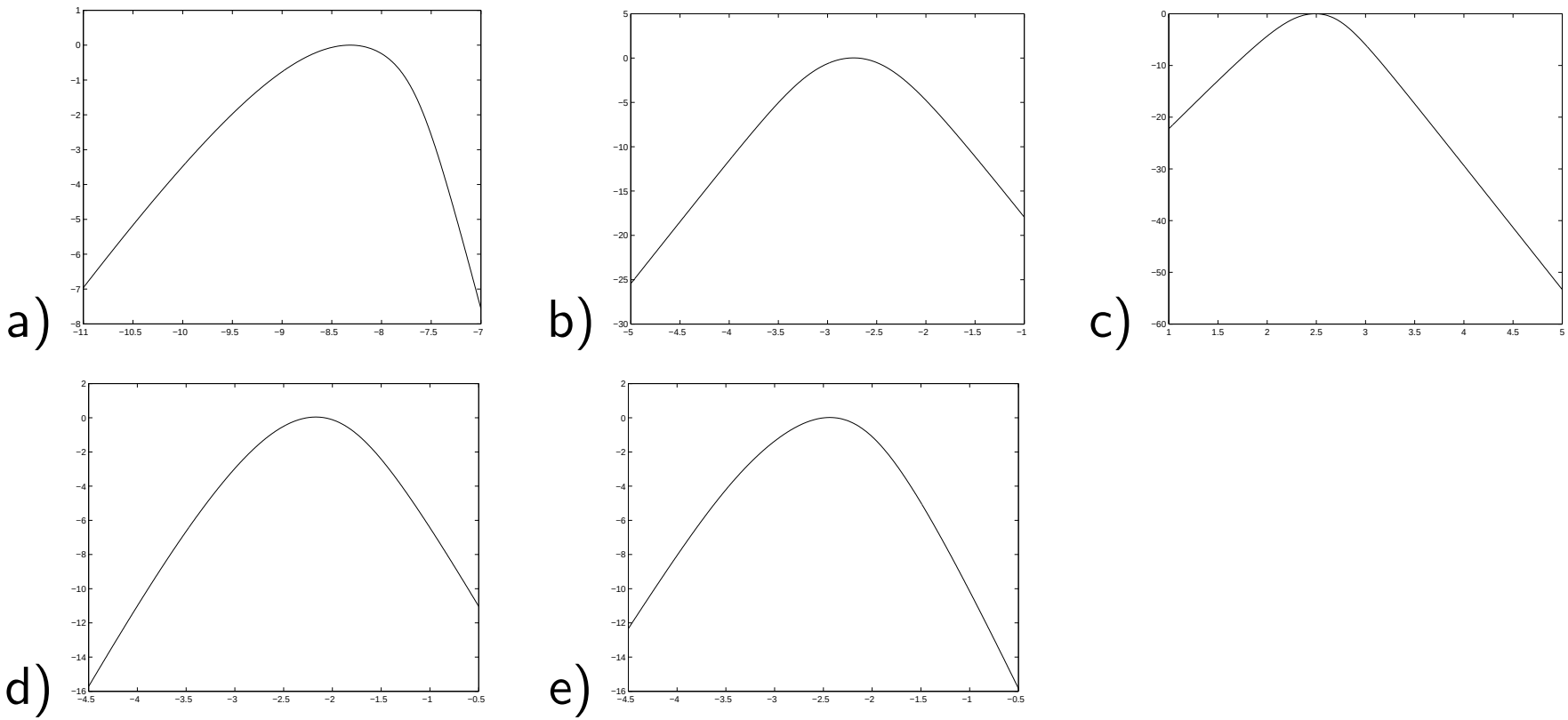


Figure: Monte Carlo approximation of the log likelihood function. The X axis represents the variation of a single component. The Y axis represents the values of the Monte Carlo log likelihood with all other components of $\hat{\theta}^0$ fixed : a - $\theta_f \in [-11, -7]$, b - $\theta_s \in [-5, -1]$, c - $\theta_d \in [1, 5]$, d - $\theta_o \in [-4.5, -0.5]$, e - $\theta_r \in [-4.5, -0.5]$.

Parameter estimation 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

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

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

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

Properties :

- 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 based on the "whole" window (Jensen and Møller, 1991)
- lacks of statistical significance (except for the Poisson process) : there is no real link with the true model behind the pattern
- but easy to be implemented : this was the motivation to introduce it in the middle of 70s (Besag, 1975)

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

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

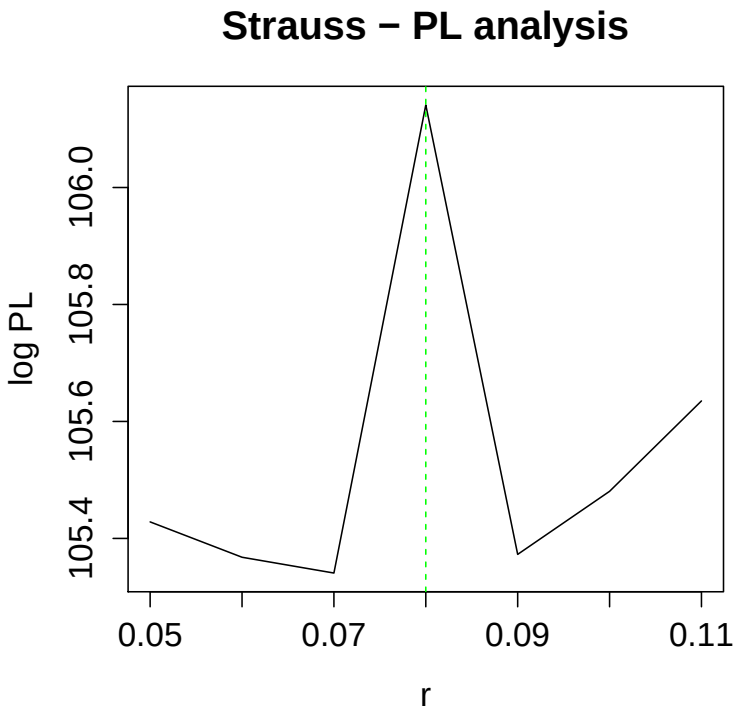
  with the 'first order term' $S(u)$ that describes spatial inhomogeneity and/or covariates effects and the 'higher order' term that describes interobject interaction

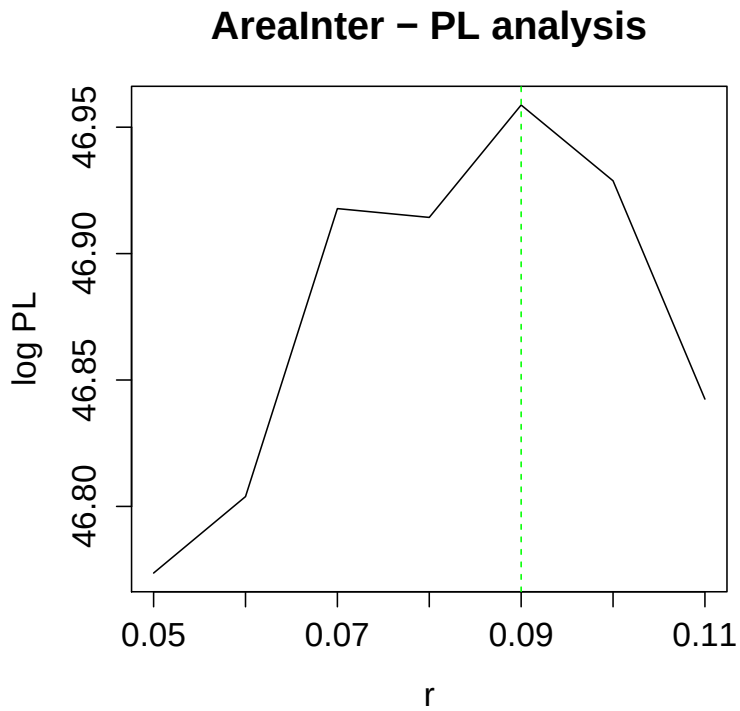
- refer to the **spatstat** documentation

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
```

Synthesis parameter estimation

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

Pseudo-likelihood :

- 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

Open questions :

- range parameters
- parameters of the mark distribution
- posterior sampling
- incomplete data : EM algorithms converges towards the first local maximum $\rightarrow$ a lot of references available ...
- ABC methods : empirical methods for parameter estimation $\rightarrow$ 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
- $\triangle I(x_i) = h(x_i, \mathbb{X} \setminus \eta)$: 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 $A \subseteq W \times M$

$$R(A, \widehat{h}, \widehat{\theta}) = I(A, \widehat{h}, \widehat{\lambda})$$
$$= \sum_{x_i \in \mathbf{x}_A} \widehat{h}(x_i, \mathbf{x} \setminus x_i) - \int_A \widehat{\lambda}(\eta; \mathbf{x}) \widehat{h}(\eta, \mathbf{x}) (\nu \times \nu_M)(d\eta)$$

since the function h may depend on the model, $\widehat{h}$ 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 $h(\eta, \mathbf{x}) = 1$

$$R(A, 1, \widehat{\theta}) = n(\mathbf{x} \cap A) - \int_A \widehat{\lambda}(\eta; \mathbf{x})(\nu \times \nu_M)(d\eta)$$

- inverse residuals $h(\eta, \mathbf{x}) = 1/\lambda(\eta; \mathbf{x})$ (equivalent with the Stoyan-Grabarnik diagnostic)

$$R(A, \frac{1}{\widehat{\lambda}}, \widehat{\theta}) = \sum_{x_i \in \mathbf{x}_A} \frac{1}{\widehat{\lambda}(x_i; \mathbf{x}_A)} - \int_A \mathbf{1}\{\widehat{\lambda}(\eta; \mathbf{x}) > 0\}(\nu \times \nu_M)(d\eta)$$

- Pearson residuals $h(\eta, \mathbf{x}) = 1/\sqrt{\lambda(\eta; \mathbf{x})}$ (analogy with Poisson log-linear regression)

$$R(A, \frac{1}{\sqrt{\widehat{\lambda}}}, \widehat{\theta}) = \sum_{x_i \in \mathbf{x}_A} \frac{1}{\sqrt{\widehat{\lambda}(x_i; \mathbf{x}_A)}} - \int_A \sqrt{\widehat{\lambda}(\eta; \mathbf{x})}(\nu \times \nu_M)(d\eta)$$

Remark

The inverse and Pearson residuals we need $\lambda_{\theta(\mathbf{x})}(x_i;\mathbf{x}) > 0$ *for all* $x_i \in \mathbf{x}$ *for any pattern yy, while* $\lambda_{\theta(\mathbf{x})}(\eta;\mathbf{x}) = 0$ *is allowed for* $\eta \notin \mathbf{x}$

Properties

- expectation

$$\mathbb{E}\left[R(A,\widehat{h},\widehat{\theta})\right] = \int_A \mathbb{E}\left[h_{\widehat{\theta}(\mathbb{X}\cup\{\eta\})}(\eta,\mathbb{X})\lambda(\eta,\mathbb{X}) - h_{\widehat{\theta}(\mathbb{X})}(\eta,\mathbb{X})\lambda_{\widehat{\theta}(\mathbb{X})}(\eta,\mathbb{X})\right]$$

- 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)

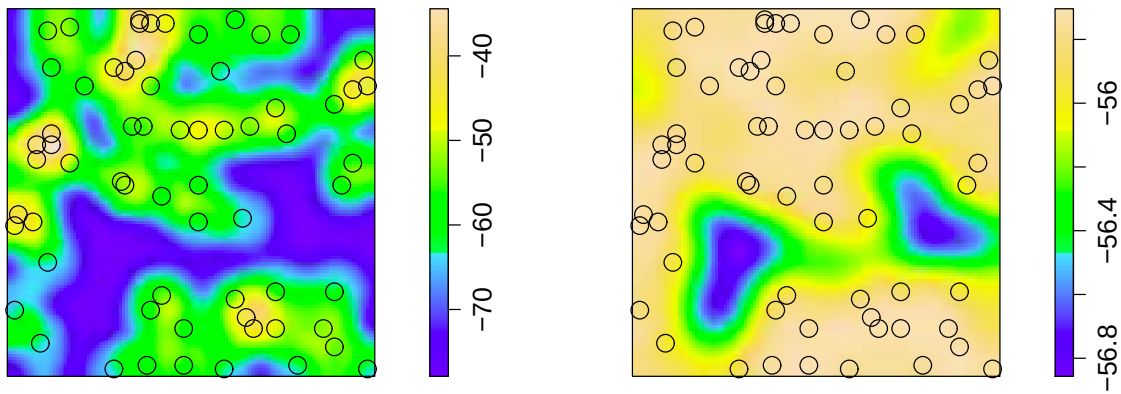


Figure: Raw residual analysis, from left to right : Strauss(r=0.08) and AreaInt(r=0.09)

- 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")`

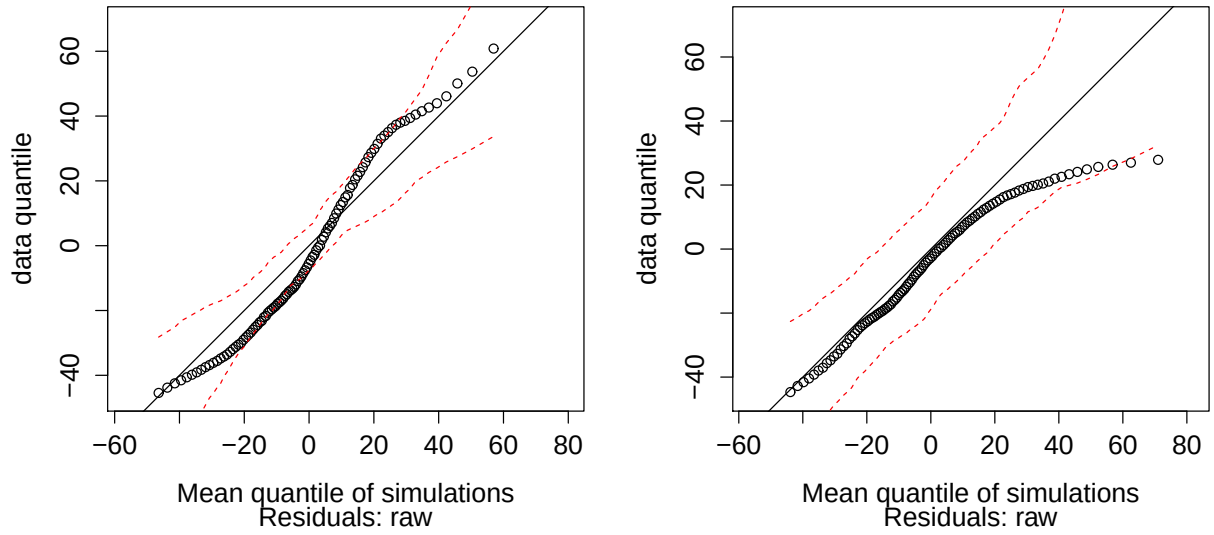


Figure: Q-Q plot analysis, from left to right : Strauss(r=0.08) and AreaInt(r=0.09)

- Strauss(r=0.08) : over-estimates repulsion, but under-estimates close attraction
- AreaInter(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 $\rightarrow$ link with the central limit theorems for computing confidence intervals
- open question : validating pattern detection result ... ?

Statistical pattern detection

Build the pattern model : probability density construction conditionally on the data observation

$$p(\mathbf{x}, \theta|\mathbf{d}) \propto \exp\left[-\frac{U_{\mathbf{d}}(\mathbf{x}|\theta) + U_i(\mathbf{x}|\theta)}{Z(\theta)} + \log p(\theta)\right]$$

- interaction energy $U_i(\mathbf{x}|\theta) \rightarrow$ objects interactions (geometrical shape of the structure)
- data energy $U_{\mathbf{d}}(\mathbf{x}|\theta)$ induced by the data field $\mathbf{d} \rightarrow$ object locations
- if the interaction parameters are unknown $\rightarrow$ prior model $p(\theta)$

Pattern estimator : the object configuration that maximises the probability density

$$(\widehat{\mathbf{x}}, \widehat{\theta}) = \arg\min_{\Omega\times\Psi}\left\{\frac{U_{\mathbf{d}}(\mathbf{x}|\theta) + U_i(\mathbf{x}|\theta)}{Z(\theta)} - \log p(\theta)\right\}$$

with Ψ the model parameters space

Simulated annealing : global optimisation technique

- sampling from $p(\mathbf{x}, \theta)^{1/T}$ while slowly $T \to 0$
- convergence towards the uniform distribution on the configuration subspace minimizing $U(\mathbf{x}, \theta)$ (Stoica, Gregori and Mateu, 2005)
- inhomogeneous Markov chain ...

Algorithm

$\mathbf{x} = \textit{Simulated Annealing}\ (T_0, \delta, T)$

1. *choose an initial condition* $\mathbf{x}_0$
2. *for* $i = 1$ *to* T *do*

$$\{$$
$$\mathbf{x}_i = \textit{Update}\ (\mathbf{x}_{i-1}, T_{i-1}, \delta)$$
$$T_i = T_0/[\log(i) + 1]$$
$$\}$$

3. *return* $\mathbf{x}_T$.

Level sets estimators :

- visit maps for compact regions in K (Heinrich, Stoica and Tran, 2012) :

$$\{T(x) > \alpha\} \Rightarrow \{T_n(x) > \alpha\}$$

- two challenges : discretisation and Monte Carlo approximations
- average behaviour of the pattern (fixed temperature)

Build the machine ...

Galaxies catalogs :

- interaction energy : Bisous model (random cylinders)
- data energy : local tests (density and spread of galaxies inside a cylinder)

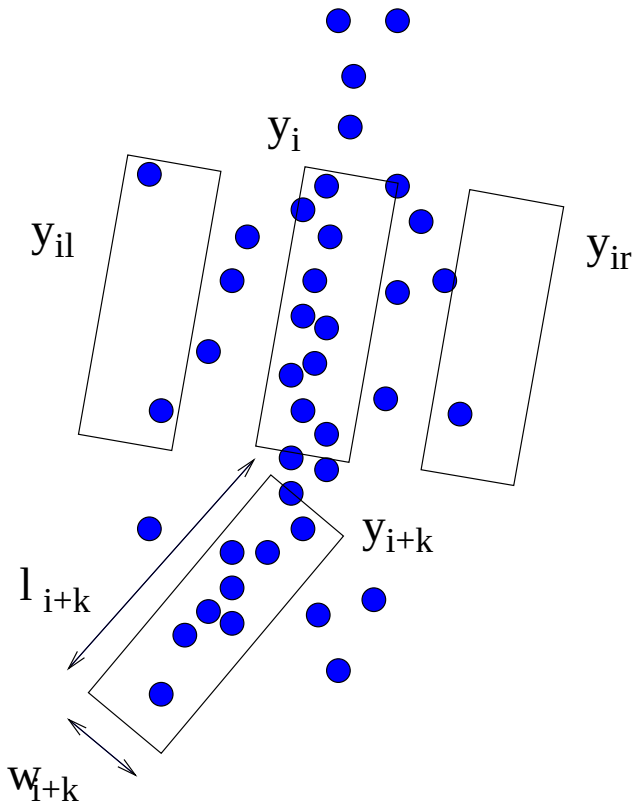


Figure: Locating interacting cylinders in a field of points.

Epidemiological data :

- interaction energy : Strauss and Area-interaction model (random disks)
- data energy : local statistical test (the average score of the farms covered by a disk)

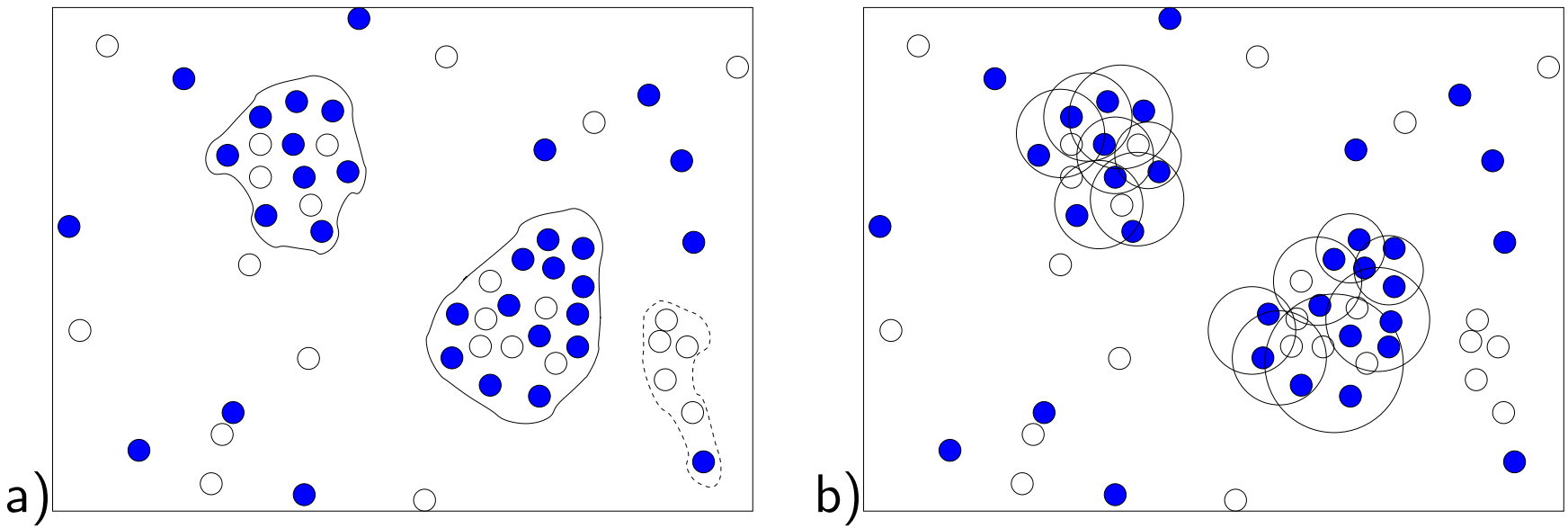


Figure: Data→ field of marked points : a) observed clusters, b) clusters approximated by random disks.

Road network extraction

(Stoica, Descombes, van Lieshout and Zerubia, 2002)

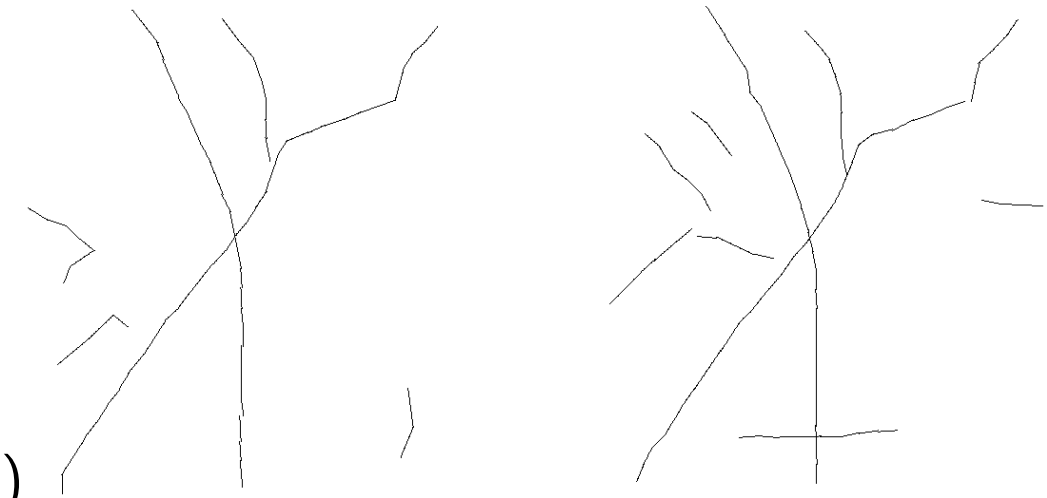
a)  b) 

Figure: Rural region in Malaysia : a) original image; b) obtained results.

Forest galleries : verifying the results

(Stoica, Descombes and Zerubia, 2004)

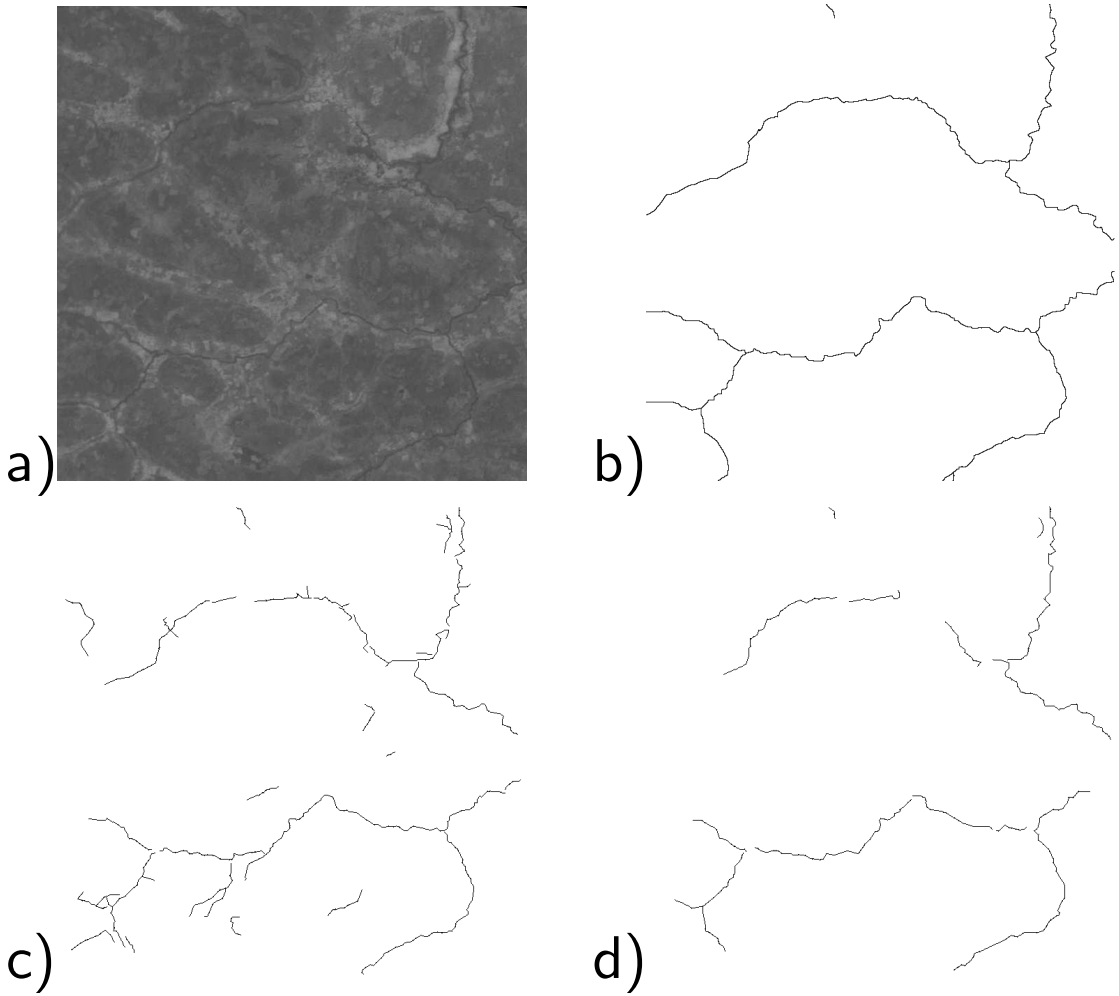


Figure: Forest galleries extraction : a) original image ; b) ground truth ; c)-d) obtained results. Data provided by BRGM.

Galaxy catalogue (1) : (Tempel, Stoica et. al., 2014)

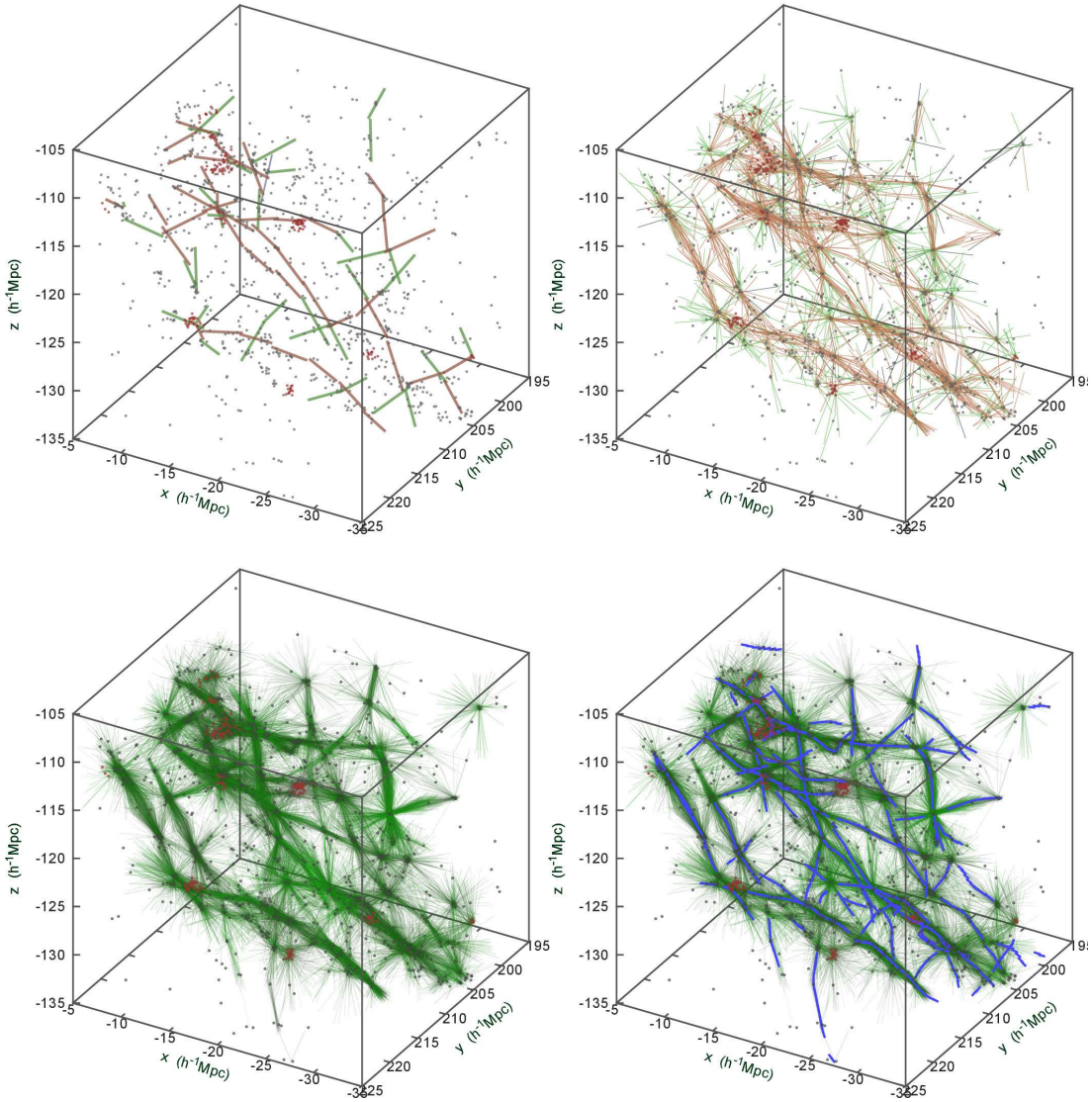


Figure: Detected filamentary pattern (cylinder axes) in a small sample volume within a pattern of galaxies (points).

Galaxy catalogue (2)

(Tempel, Stoica et. al., 2014)
The movie, showing the MCMC in action is available at
:`http://www.aai.ee/ elmo/sdss-filaments/`

Epidemiology : sub-clinical mastitis data

(Stoica, Gay and Kretzschmar, 2007)

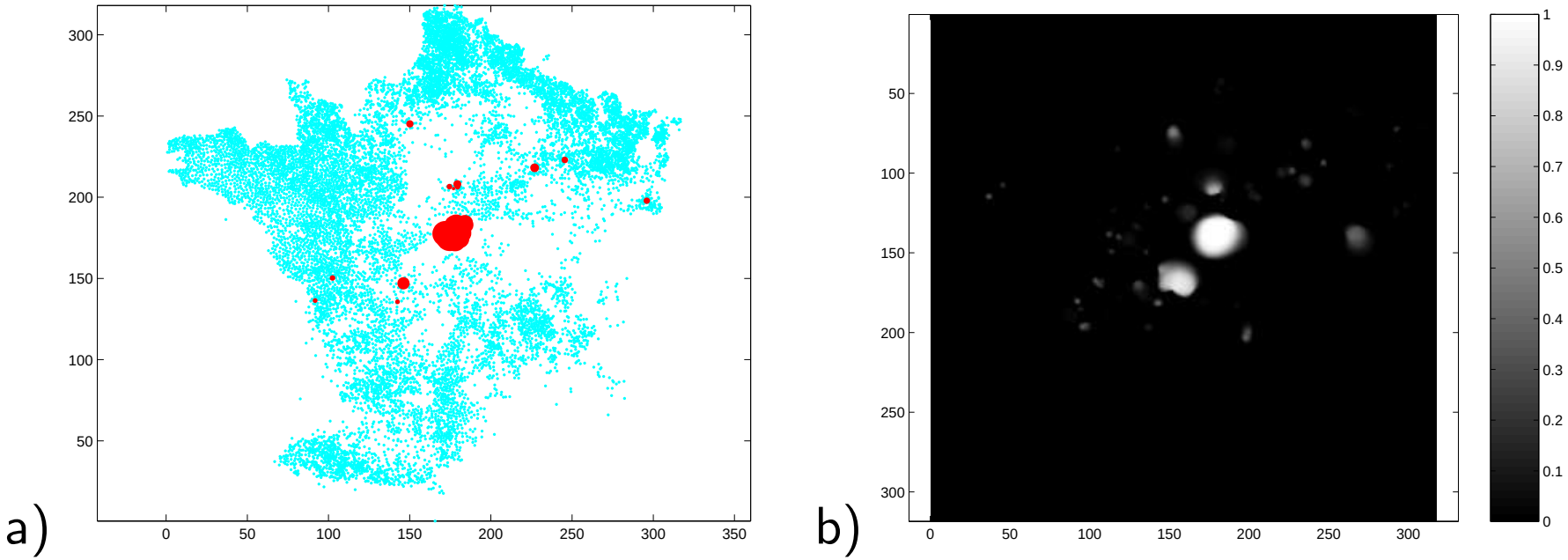


Figure: Disease data scores and coordinates for the year 1996 : a) disk configuration obtained using the simulated annealing algorithm ; b) cover probabilities.

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 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 epidemiological data

Permuted data : keeping the same farm locations while exchanging the score disease

Results :

- sufficient statistics for the data from the year 1996 :

$$\bar{n}(\mathbf{y}) = 74.10, \quad \bar{\nu}[Z(\mathbf{y})] = 312.46, \quad \bar{n}_o = 555.08$$

- maximum values of the sufficient statistics for 100 simulated data fields

$$\bar{n}(\mathbf{y}) = 2.36, \quad \bar{\nu}[Z(\mathbf{y})] = 13.83, \quad \bar{n}_o = 2.62$$

Interpretation : this test does not say if the pattern is well detected, but it says that there is something to be detected ...

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

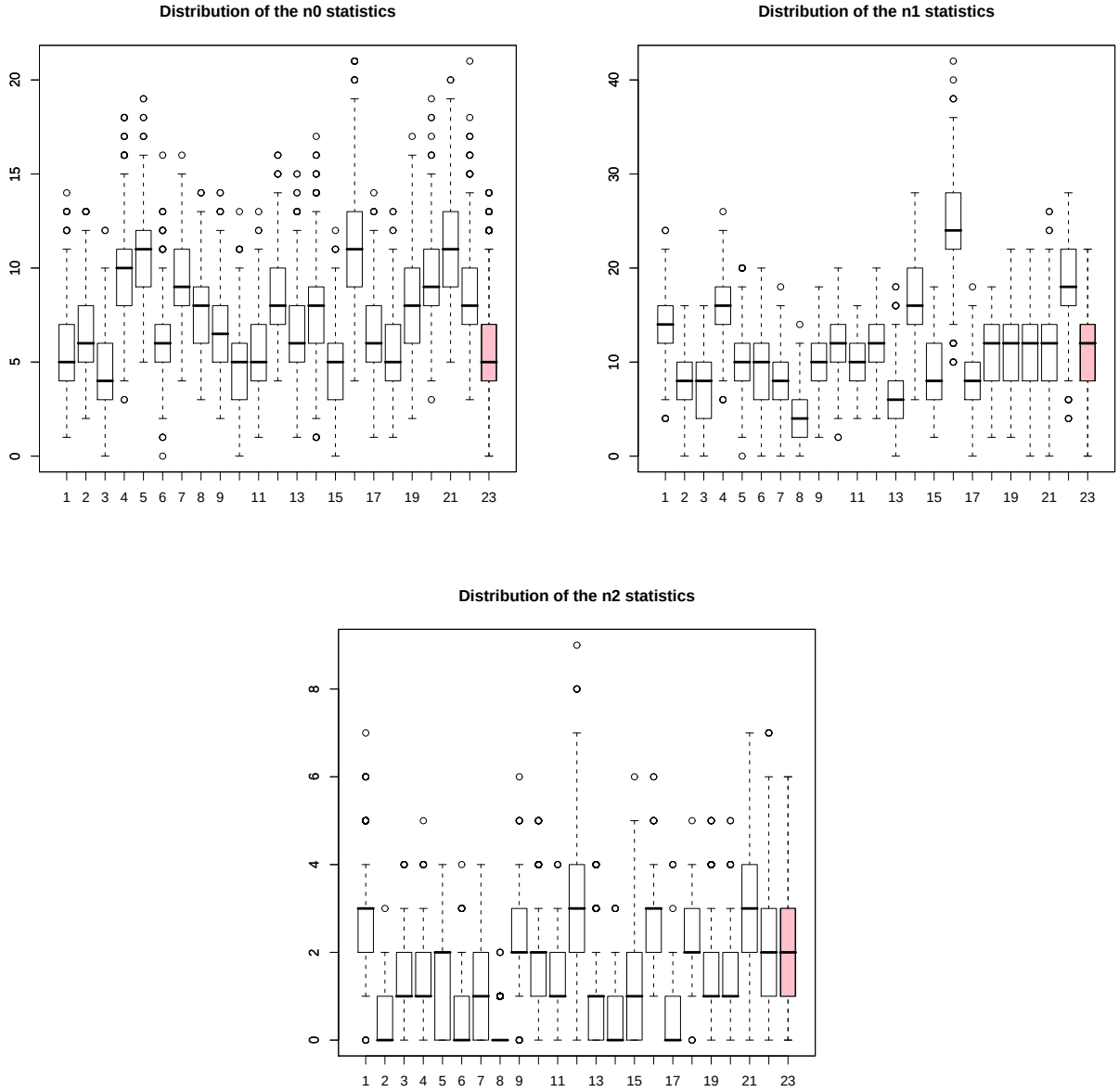


Figure: Comparison of the sufficient statistics distributions for the real data (dark box plot) and the mock catalogues.

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 :

- random geometry (marked point processes, random fields) $\rightarrow$ 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 ... Some of them are today with us :) ...