ENVIRONMENTAL STATISTICS

Spatial point processes

Linda Altieri

linda.altieri@unibo.it

University of Bologna

INTRODUCTION TO THE COURSE

Course contents

Tentative program

- Introduction to point process data
- Descriptive measures for point data
- Homogeneous Poisson processes
- Tests for complete spatial randomness
- Interpoint interaction
- Inhomogeneous Poisson processes
- Point process model criticism

In order to successfully attend the course

Please install on your laptop:

- R software (www.r-project.org) - please check for the latest version
- R-studio (https://www.rstudio.com/products/rstudio/download) *optional*
- package `spatstat` (please update to, at least, version 2.0-0)

If you are not familiar with R, an introductory course can be handed out - email me

References

THEORY

- P. J. Diggle, Statistical Analysis of Spatial and Spatio-Temporal Point Patterns. Third Edition
- J. B. Illian et al., Statistical Analysis and Modelling of Spatial Point Patterns

SOFTWARE

- A. Baddeley, Analysing spatial point patterns in R

Course material online

- Course slides
- Notes taken during classes
- Course R script
- Exercise - pdf file
- Exercise - R script
- Exam example - pdf file
- Exam example - R workspace
- Exam example - R script with solution

INTRODUCTION TO POINT PROCESS DATA

Point process data

Point process data $\neq$ Areal data

- one point is one single event, not an aggregate number
- a-dimensional data
- the process is continuous, not discrete

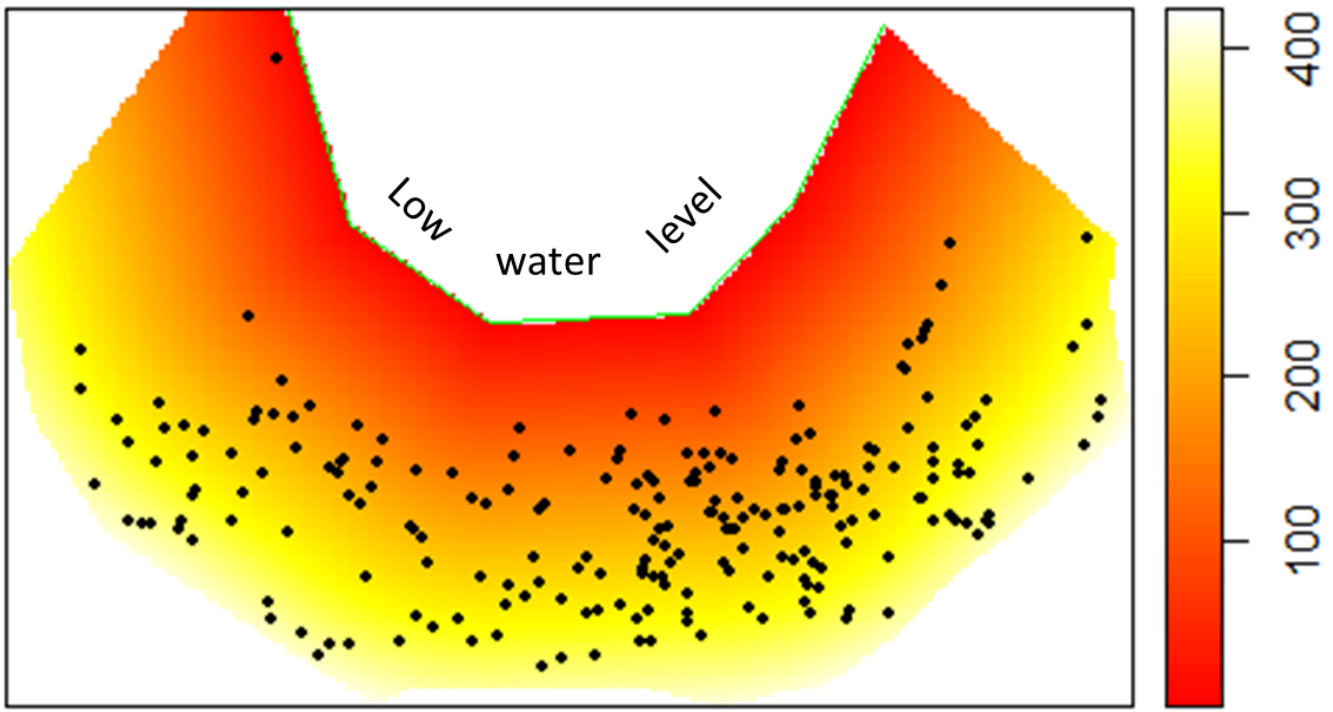
Basic question when dealing with point process data:
WHERE does the phenomenon of interest occur?

$\downarrow$

Objectives

- understand the law regulating the phenomenon
- forecast future events/events in other areas

Example



Point process data

Typical examples of point data

- Environmental data - e.g.
  - locations of a tree species of interest,
  - locations of animal stocks,
  - coordinates of earthquakes' centres,
  - locations of wildfires ....
- Health data - locations of occurrence of a disease

Data are detailed ($\neq$ areal data), each point marks a **single** event of interest

Areal data may be seen as aggregation/discretization of point data (e.g. for privacy reasons, for computational easiness...)

Point process dataset

A dataset must include the following **essential information**:

1. Couples of **coordinates** locating the events over the (2D) space
2. Observation **window**

Please note: nothing else is necessary

Potential **extra information**

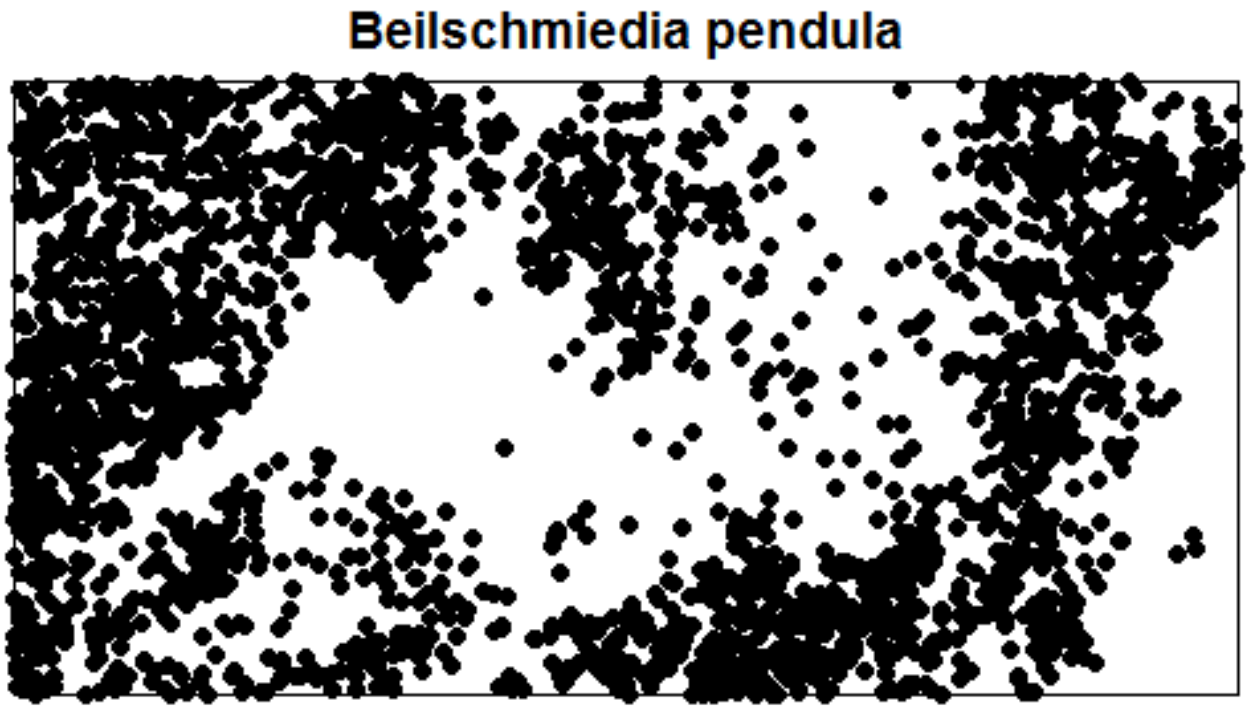
1. **Covariates**: known over the whole observation window, irrespective of the point locations
2. **Marks**: an additional variable linked to the points - e.g. tree diameter, animal gender, species...

Example - tropical rainforest trees

bei - Beilschmiedia pendula

Dataset regarding the location of 3604 trees over a rectangular observation window of 1000 by 500 metres in the tropical island of Barro Colorado, Panama.

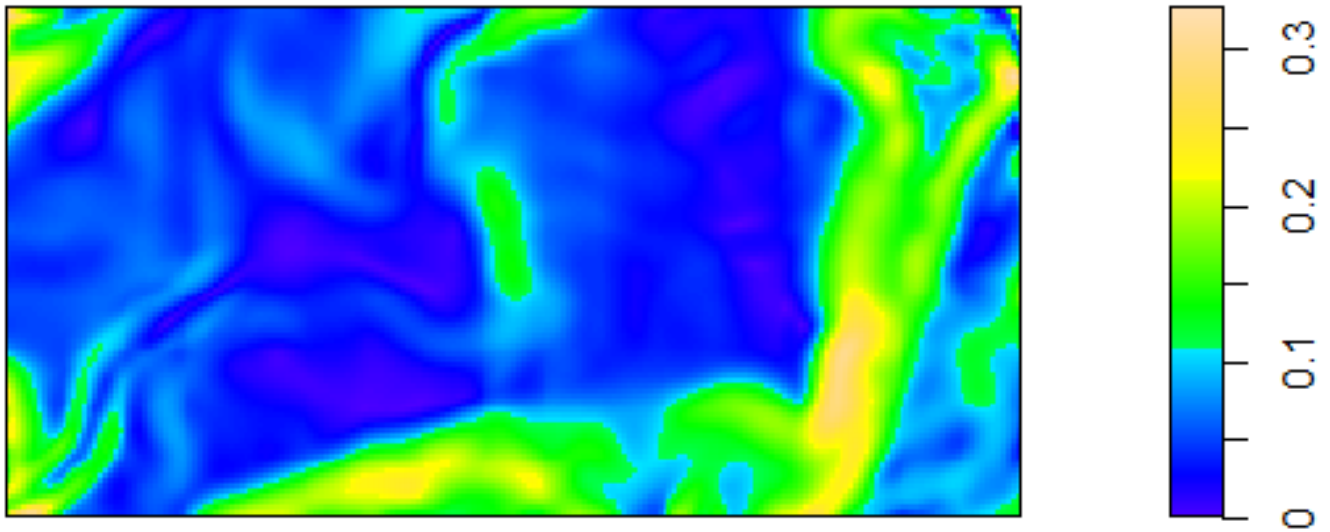
Available in `spatstat`



Example - tropical rainforest trees

Potential **questions** about the data:

- are trees randomly distributed?
- if not, do they have a clustering or repulsive behaviour?
- is their spatial distribution influenced by the available covariates?
- what is a suitable model for these data?
- ...

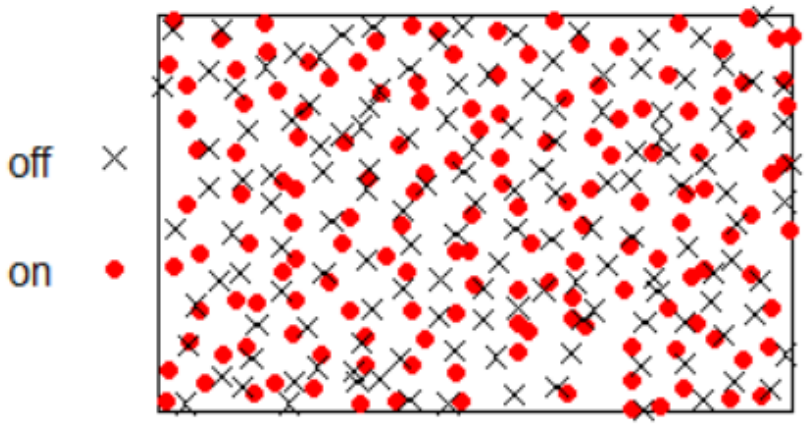


→ **R script**

Example - retina cells

amacrine

A point pattern of amacrine cells in the retina of a rabbit. There are 152 'on' cells and 142 'off' cells in a rectangular sampling frame



Possible additional **questions**:

- do the two types of cell behave the same way?
- do the two types interact?
- what type of interaction takes place?
- ...

$\rightarrow$ **`R script`**

Formal definition of a point pattern

A *point pattern* is

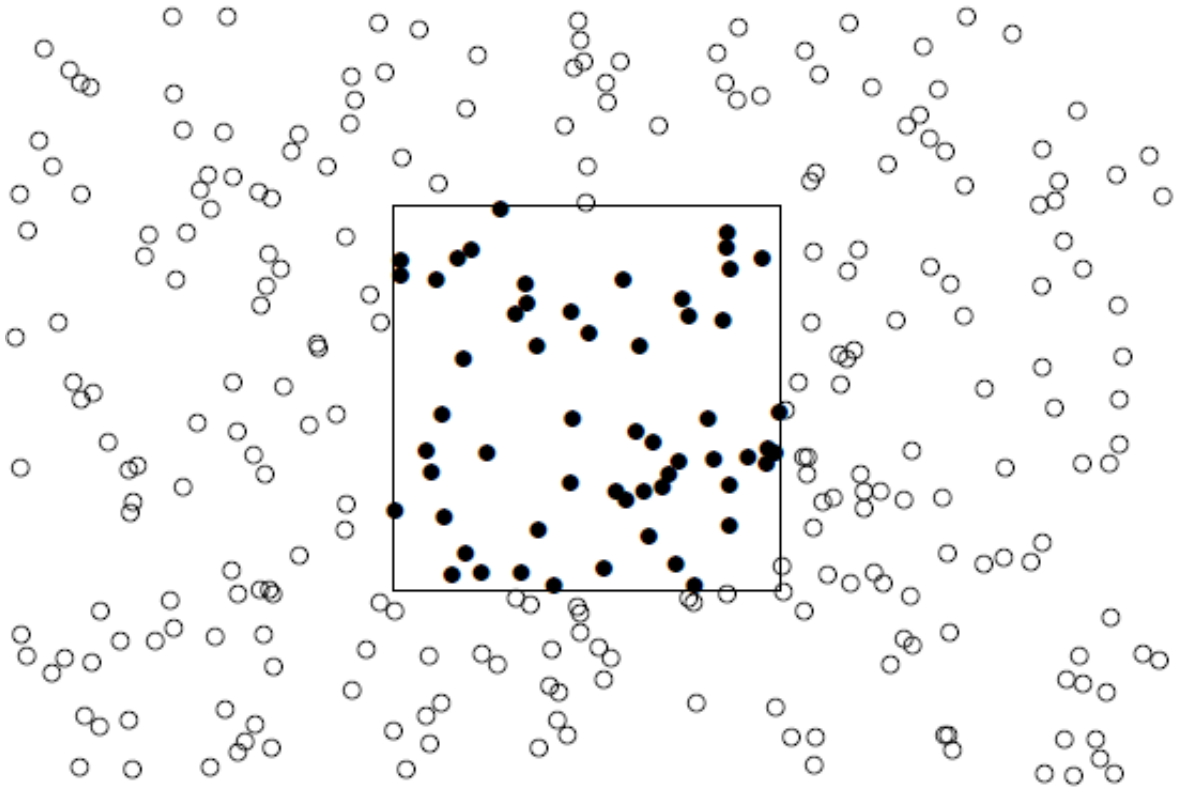
- a collection of spatial locations irregularly distributed over an observation window
- a realization of a stochastic process called *point process*

Note Points may be called *events*, in order to distinguish them from arbitrary locations in space

Note Assumption in this course: 2D space, 1 occurrence over time

Patterns and processes

The point process is *continuous* over space
(i.e. points can potentially occur anywhere)
and also exists beyond the observation window



The pattern of interest is *completely*
observed over a well defined observation window

Observation window

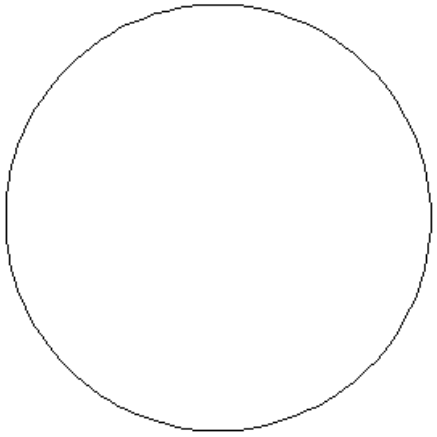
The observation window may be denoted as

$$W = \bigcup A_i$$

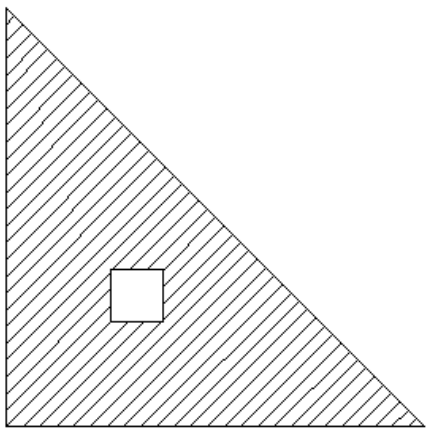
and is the overall area over which data are observed; it can be partitioned into sub-regions A_i

- it is exogenously fixed and known, both in shape and size
- it can take many shapes

Circular window



Polygonal window with holes



Window as binary mask

→ **R script**

Marks and covariates

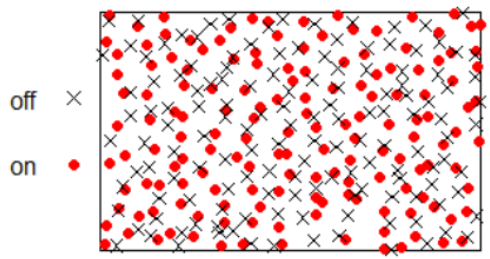
Marks

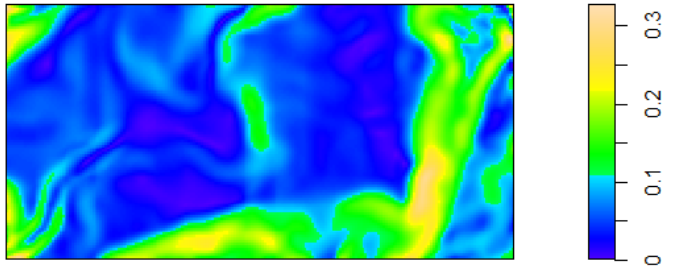
- information linked to the event of interest
- part of the response
- only observed at point locations

Covariates

- exist prior to the event of interest
- deterministic
- known over the entire observation window

Marks and covariates may be of any type: categorical, ordered, discrete, continuous, ...





Aim of point process data analysis

Understanding the point generating process, based on the available information:

Random

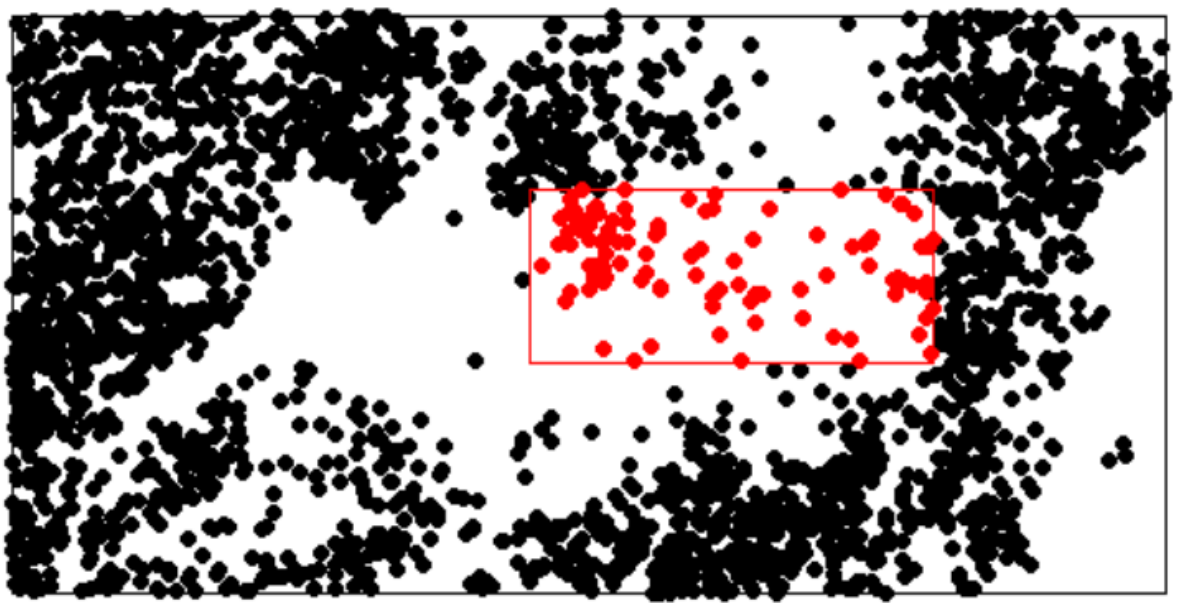
- point location
- point number
- (optional) mark realizations

Fixed

- shape and size of the observation window
- (optional) covariates

Notation

- $N(W)$: random variable, number of points over the whole observation window
- $N(A)$: random variable, number of points over region A
- $P(N(A) = n)$: distribution of the number of points
- $E(N(A)) = \Lambda(A)$: expected number of points in A, according to the distribution of $N(A)$



Point process intensity

$$\Lambda(A) = \int_A \lambda(u)\, du$$

the expected number of points in A depends on the *risk* of having an event on each spatial location u in A ($u = (x, y)$ coordinates)

$\lambda(u)$ is the **intensity function** on each spatial location u

$$\lambda(u) = \lim_{|du| \to 0} \frac{E(N(du))}{|du|}$$

the process intensity is the expected number of points in an extremely small area around u; at the limit, it represents the *risk* of having events in u

Note: $\lambda(u)$ is defined for every possible location within the observation window, not only where events occur

Intensity of a point process

The intensity function $\lambda(u)$ is always **non-negative**

Its value at each location u **may be constant or not**, and may depend on

- a baseline value due to specific phenomenon characteristics
- covariates
- interpoint interaction
- other factors (e.g. random effects)

$\lambda(u)$ **is what we model**; according to what we add to the model, we obtain different classes of point processes

To be sure the intensity is non-negative, we model $\log(\lambda(u))$

Intensity vs probability

Both intensity and probability are non-negative, and they are related, but they are not the same

Over a very small area u we have

- $p_u(i)$ the **probability** of observing i points (i can be 0, 1, 2, ...)
  $\sum_{i=0}^{\infty} p_u(i) = 1$
- $\lambda(u)$ the "**risk**" of observing points in u
  Its complement to 1 is the risk of *not observing* points $1 - \lambda(u)$
  The sum (integral) $\sum_W \lambda(u) \neq 1$
- the probability of having i points depends on λ (we will see the Poisson distribution..)

Point process analysis is interested in the *risk* of observing events over an area, rather than on the *probability* of a specific number of events

Point process classes

There are three main point process classes

1. **Poisson processes** - intensity may be constant or vary, still it is a function of known quantities. Points do not interact and there are no random effects, only an error term

2. **Cox processes** - intensity may depend on an underlying process; conditional on this process, points do not interact

3. **Gibbs processes** - the presence of events in an area (i.e. the intensity) depends, positively or negatively, on the presence of events in surrounding areas: points interact in a clustering or repulsive way

Stationarity

Intuitively:
a point process is stationary if the **expected number of points within an area A only depends on the area size**, not on its spatial location

A more formal definition:
a point process is stationary if it is **invariant to translations**:

$$X =^d X_t$$

$$N(A) =^d N(A_t)$$

where X_t is the shifted process from area A to area A_t

Stationarity $\equiv$ homogeneity

Stationarity

Examples of stationary/homogeneous processes:

- points are randomly and uniformly scattered over the observation window (homogeneous Poisson process)
- points are distributed over space according to a random intensity function, whose parameters are constant over the observation window (stationary Cox process)

Examples of non stationary/inhomogeneous processes:

- intensity depends on the spatial location e.g. via covariates. If an area is shifted, the covariates' value changes, therefore expected number of points changes
- the point configuration may depend on the spatial location in other ways, e.g. point clusters may be larger in some areas than others

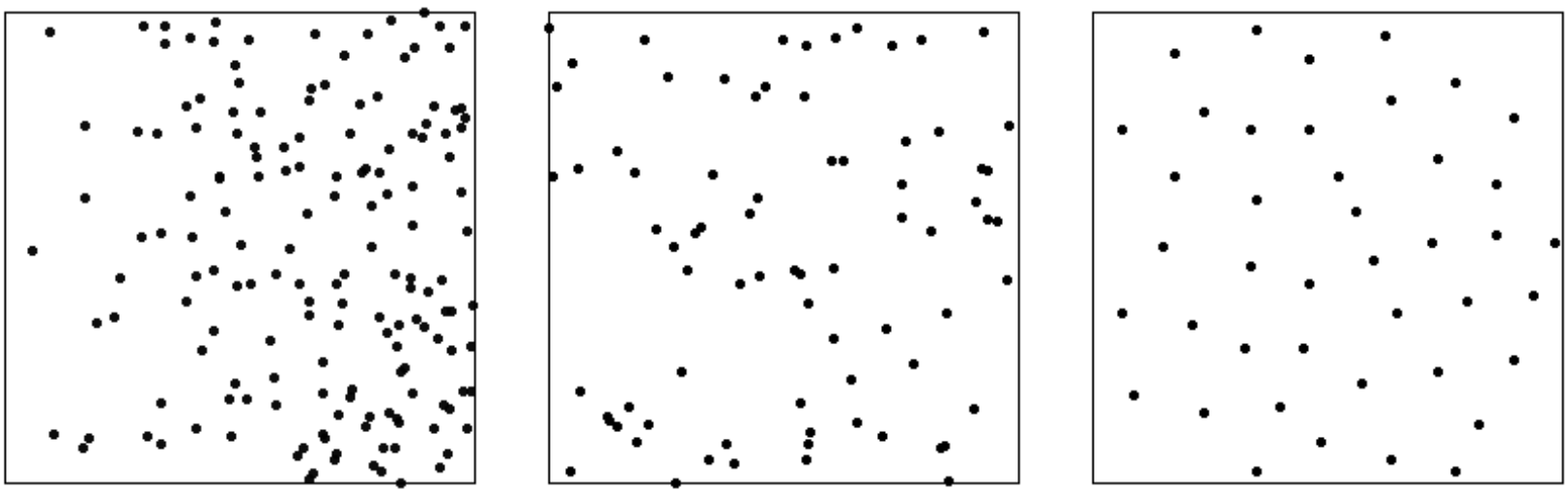
HOMOGENEOUS POISSON PROCESSES AND TESTS FOR COMPLETE SPATIAL RANDOMNESS

Homogeneous Poisson process - intuitively

A point process X is a Homogeneous Poisson process if

- the probability of finding a point is constant everywhere over the window
- points do not interact, i.e. finding a point in a location does not influence the probability of finding a point closeby

Are these homogeneous Poisson processes?



Homogeneous Poisson process - formal properties

Key element: the **intensity** $\lambda(u) = \lambda$ **is constant** everywhere.

- the number of points over the window W is Poisson distributed with parameter $\Lambda(W)$

$$N(W) \sim Poisson(\Lambda(W)) = Poisson(\lambda|W|)$$

$$P(N(W) = n) = e^{-\lambda|W|} \frac{(\lambda|W|)^n}{n!}$$

- for a partition $\{A_1, \ldots, A_m\}$ of W

$N(A_i) \sim Poisson(\Lambda(A_i)) = Poisson(\lambda|A_i|)$ for $i = 1, \ldots, m$

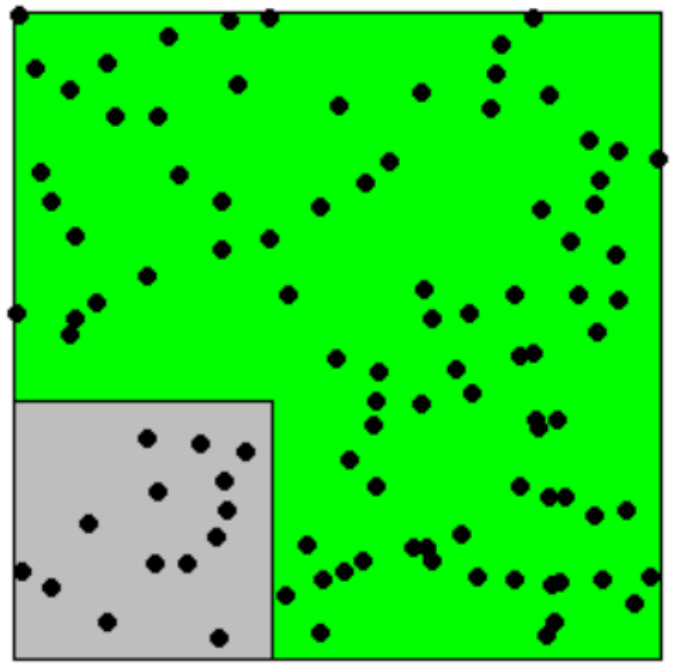
$$N(W) \sim Poisson\left(\sum_i \Lambda(A_i)\right) = Poisson\left(\lambda \sum_i |A_i|\right) = Poisson(\lambda|W|)$$

In this simple process, an **intuitive estimator** for the intensity is the *observed* number of points divided by the window area

$$\hat{\lambda} = \frac{n}{|W|}$$

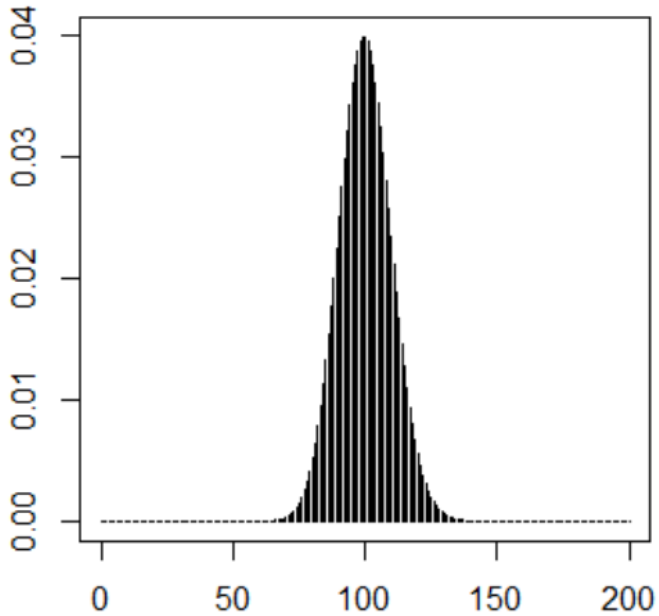
Homogeneous Poisson process

Example



Homogeneous Poisson process

Example



Note: it looks approximately normal, since $N(W)$ is large

Process likelihood

The likelihood of a homogeneous Poisson process is

$$L(X|\lambda) = \exp(|W|)\exp(-\lambda|W|)\lambda^n$$

where $\exp(|W|)$ is an offset (known intercept) determining the *exposure* of the process to the risk of having events in the window W.

The log-likelihood is

$$l(X|\lambda) = |W| - \lambda|W| + n\log(\lambda).$$

By setting its derivative with respect to λ equal to zero, we obtain

$$\frac{\partial}{\partial\lambda}\left[-\lambda|W| + n\log(\lambda)\right] = -|W| + \frac{n}{\lambda} = 0$$

$$\hat{\lambda} = \frac{n}{|W|}$$

Estimation of a homogeneous Poisson process

The intuitive estimator

$$\hat{\lambda} = \frac{n}{|W|}$$

is also the ML estimator, and it is known to enjoy the following properties:

- it is unbiased

$$E[\hat{\lambda}] = \lambda$$

- its variance is

$$V[\hat{\lambda}] = \frac{\lambda}{|W|}$$

Note: this estimator is also unbiased for any stationary/homogeneous process, even for non-Poisson ones

Confidence intervals

(Approximate) confidence intervals for the estimator $\hat{\lambda}$:

given

$$\hat{\lambda}|W| = n \sim Poisson(\lambda|W|)$$

the confidence interval at a $(1-\alpha)$ level is given by

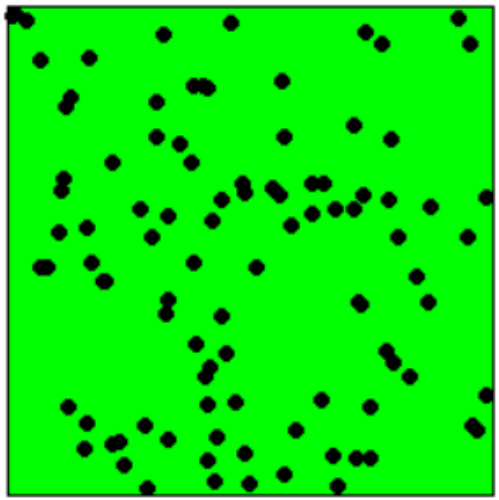
$$\left(\frac{z_{\alpha/2}}{2} - \sqrt{n}\right)^2 \leq \lambda|W| \leq \left(\frac{z_{\alpha/2}}{2} + \sqrt{n+1}\right)^2$$

Note:

1. this implies a normal approximation of the Poisson distribution, under the hypothesis that $\lambda|W|$ is large
2. the interval is adjusted with a *continuity correction* or *Yates' correction* (the term $1/2$)
   See, e.g., Armitage et al (2001) Statistical methods in medical research. Blackwell Science, chapters 4 and 5

Confidence intervals

Example



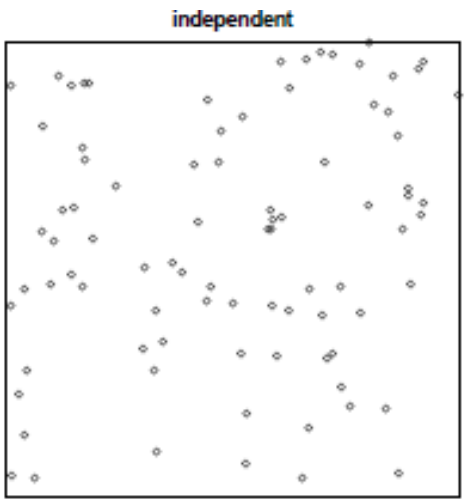
→ **R script**

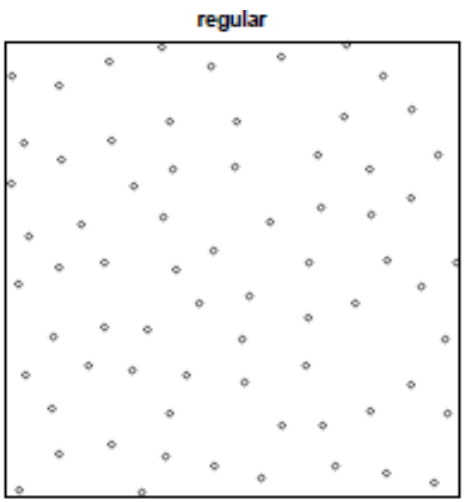
Complete spatial randomness

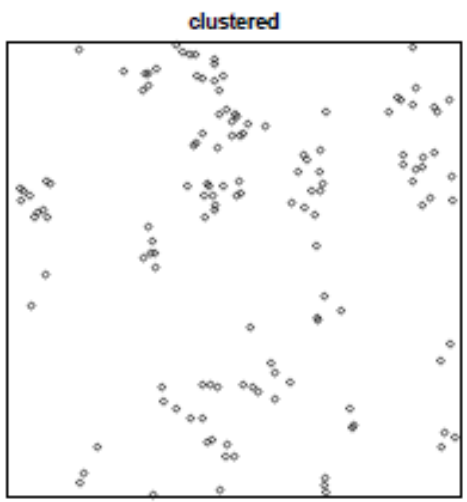
The distribution of points which follow a homogeneous Poisson process is also known as **Complete spatial randomness** (CSR)

The homogeneous Poisson process is almost impossible to find in real situations

However, it is a useful benchmark since all its features are known; it can be used as a reference process for comparison.
Tests for CSR are run on a dataset in order to understand whether the data distribution is random, clustered or repulsive, and to get hints about the most suitable model to fit







Tests fot Complete Spatial Randomness

CHI-SQUARE TEST

The window W is partitioned into P equivalent (=of equal size) sub-areas; then, a standard Chi-square test is performed

$$\chi = \sum_{p=1}^{P} r_p^2 = \sum_{p=1}^{P} \frac{(o_p - e_p)^2}{e_p}$$

where

- o_p is the observed number of points in area p
- e_p is the expected number of points in area p

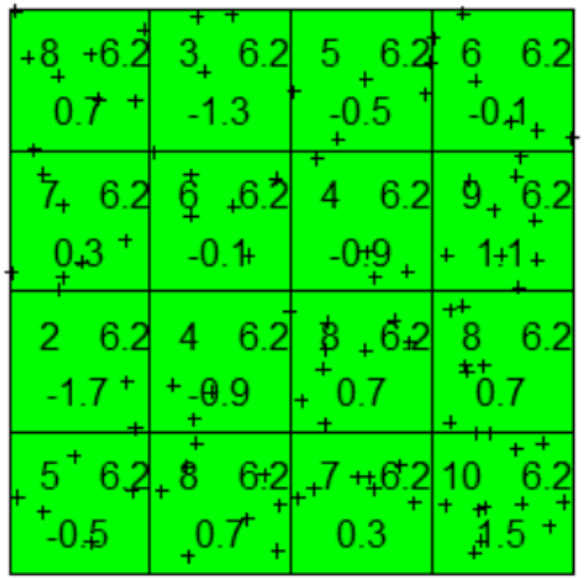
If the hypothesis of CSR is correct,

$$\chi | H_0 \sim \chi^2_{(P-1)}$$

otherwise data are clustered or repulsive

Tests for Complete Spatial Randomness

CHI-SQUARE TEST



- top-left numbers: $o_p = N(A_p)$
- top-right numbers: $e_p = \hat{\lambda}|A_p| = \hat{\lambda}|W|/P = 6.2$
- bottom numbers: $r_p = (o_p - e_p)/\sqrt{e_p}$

$\chi = 12.32$ and the p-value for χ^2_{15} is 0.69, thus H_0=CSR is not rejected

$\rightarrow$ **`R script`**

Tests for Complete Spatial Randomness

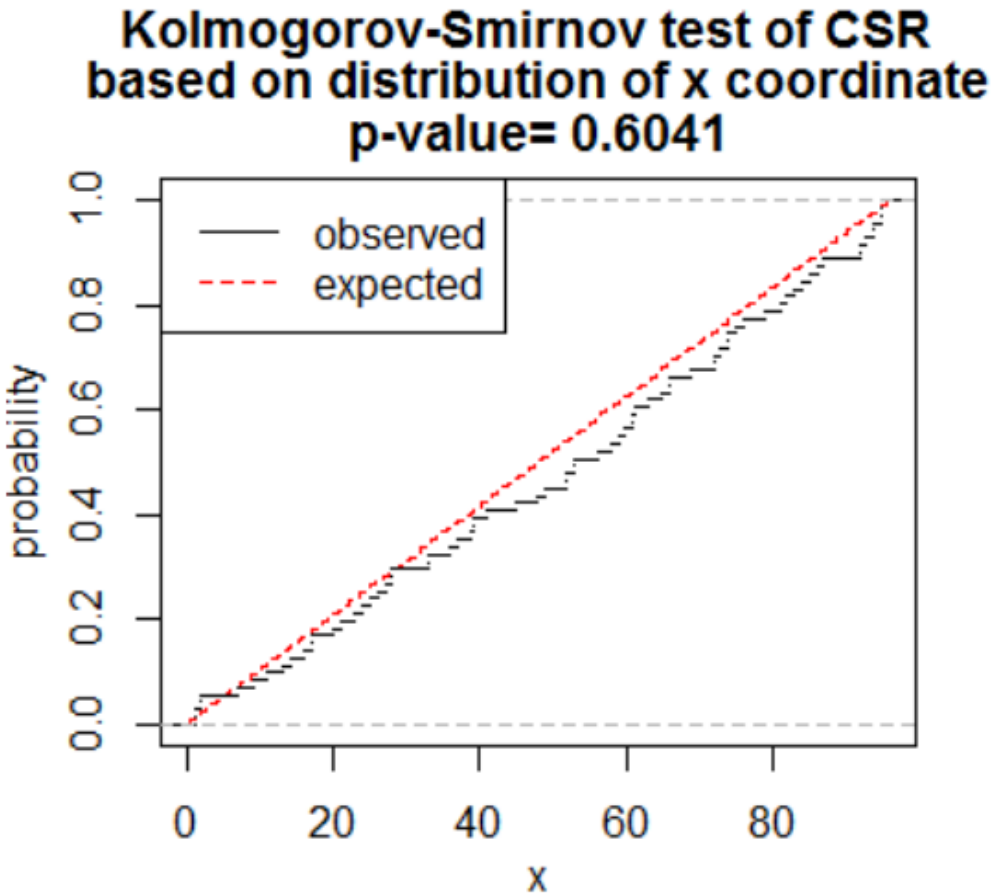
KOLMOGOROV-SMIRNOV TEST

1. a spatial function is chosen, which is defined for all spatial locations over the window
   - a covariate
   - coordinates x and y
2. the function is evaluated at all events locations
3. its empirical cumulative distribution function is computed
4. the empirical cdf is compared to the cdf under CSR

Tests for Complete Spatial Randomness

KOLMOGOROV-SMIRNOV TEST

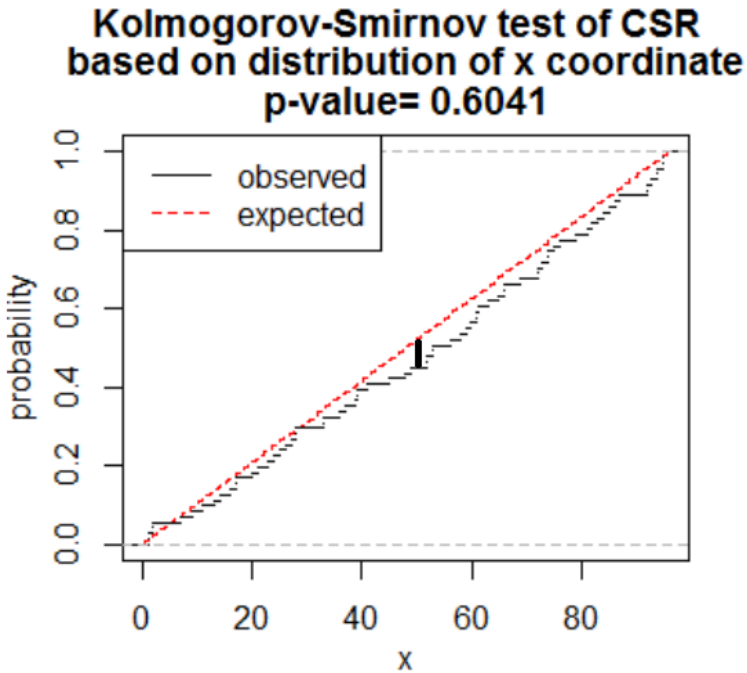
Example: x coordinate



Tests for Complete Spatial Randomness

KOLMOGOROV-SMIRNOV TEST

Example: x coordinate



Test statistic for n observations: $D_n = \sup |F_n(x) - F(x)|$

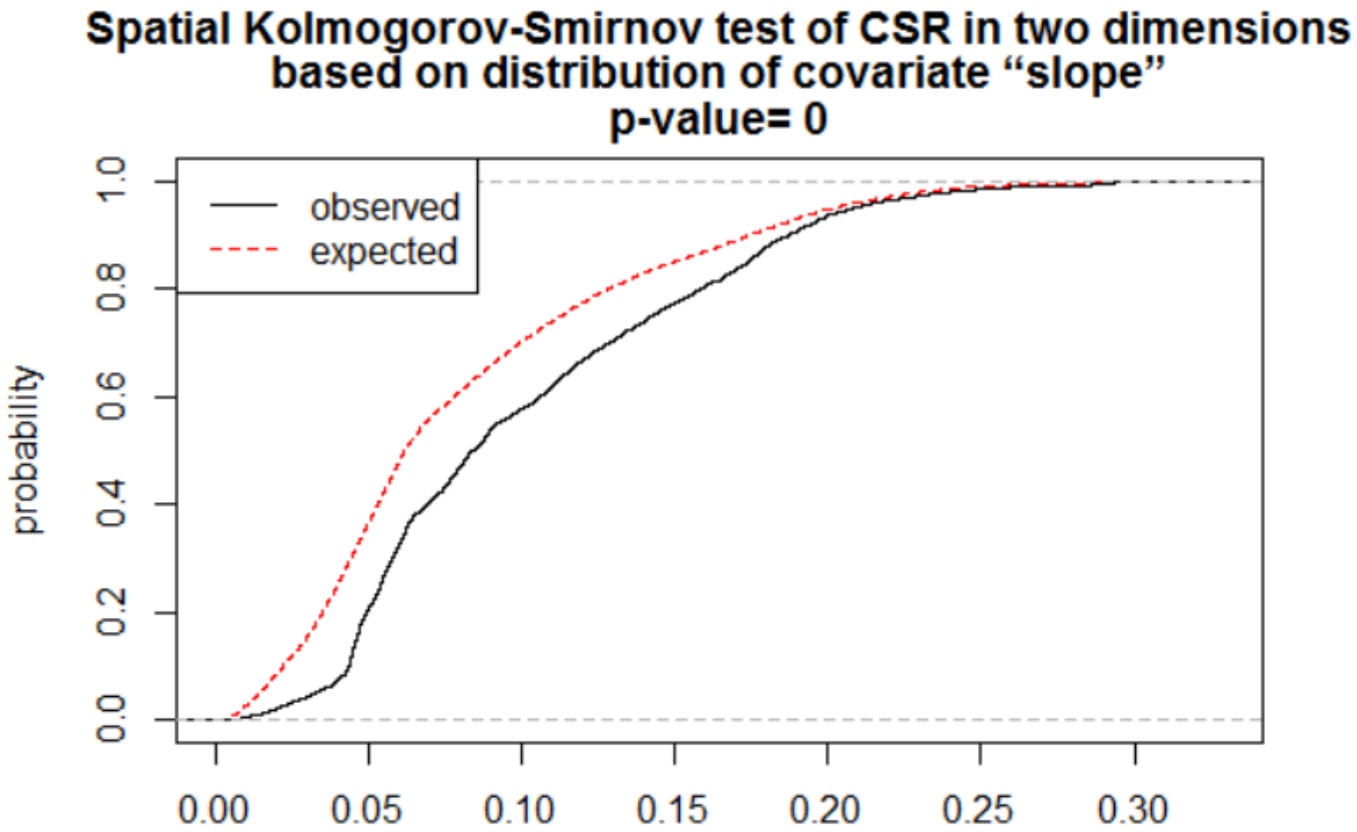
H_0: data come from a homogeneous Poisson process $\Rightarrow D_n \rightarrow_{n\rightarrow\infty} 0$

The distribution of $\sqrt{n}D_n$ is known under H_0; it is used for comparison of the test statistic in order to decide CSR rejection

Tests for Complete Spatial Randomness

KOLMOGOROV-SMIRNOV TEST

Example: covariate 'slope'



→ **R script**

Interpoint interaction

DISTANCE-BASED METHODS
It is another class of methods to understand whether a dataset is uniformly distributed (CSR)

These methods are based on interpoint distances, which are compared to distances under CSR, and use simulation techniques

Three main functions measure the interpoint interaction:

- **empty space distance**: between an arbitrary point u and the closest event x
- **pairwise distance**: between all pairs of events
- **nearest neighbour distance**: between each event and the nearest one

Interpoint interaction

NEAREST NEIGHBOUR DISTANCE

It is the distance between each event x_i in the window and the nearest event x_j

$$d(x_i) = \min_j ||x_i - x_j|| \text{ for } x_i, x_j \in X \text{ , } j \neq i$$

This distance is only defined for the data points, therefore if we have n events, we have n distances of type $d(x_i)$

The function of interest is the cdf of these distances

$$G(r) = P(d(x_i) \leq r) \text{ for } r > 0$$

r is the distance, and the function tells how likely it is that, starting from any event, we find other events at a distance $\leq r$

The function is known for a homogeneous Poisson process:

$$G_{Poi}(r) = 1 - \exp\left(-\lambda\pi r^2\right)$$

where $\exp\left(-\lambda\pi r^2\right)$ is the probability of having 0 points within a disk of radius r

Interpoint interaction

NEAREST NEIGHBOUR DISTANCE

This is how we estimate the function:

1. for each event x_i, we compute the distance $d(x_i)$ to the closest event
2. at each distance r, we count how many $d(x_i) \leq r$, i.e. we compute $\sum_{i=1}^{n} \mathbf{1}\left(d(x_i) \leq r\right)$
3. we compute the empirical cdf

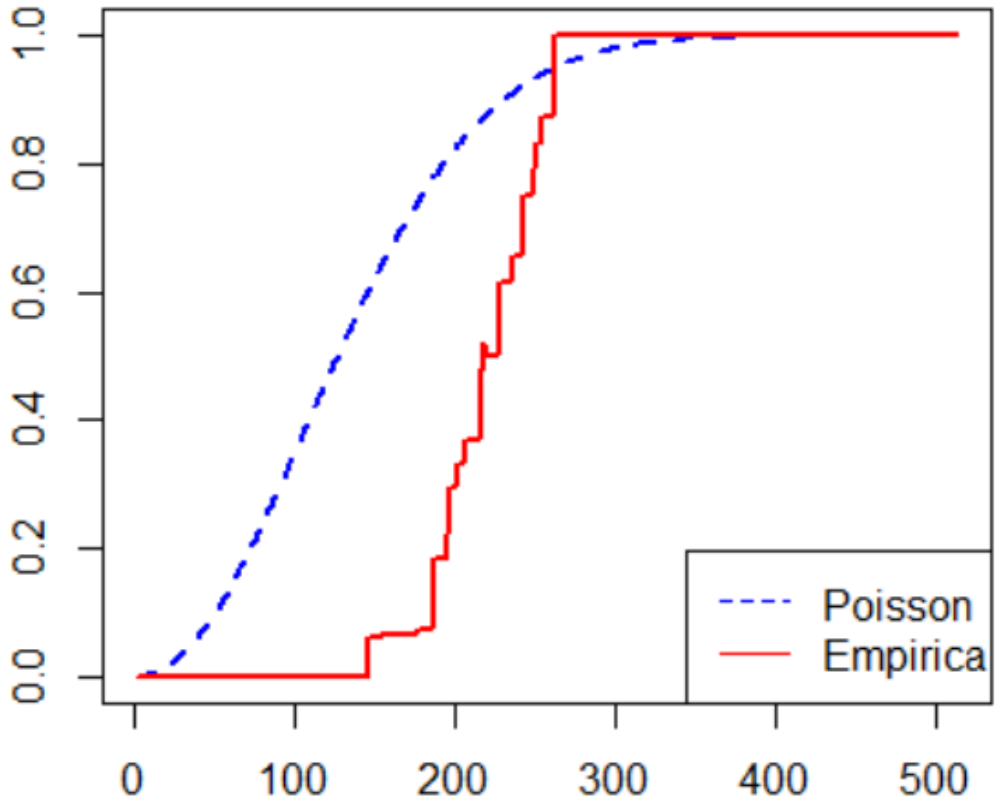
$$\hat{G}(r) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\left(d(x_i) \leq r\right)$$

4. we compare the ecdf to the function under CSR: given $\hat{\lambda} = n/|W|$

$$\hat{G}_{Poi}(r) = 1 - \exp\left(-\hat{\lambda}\pi r^2\right)$$

Interpoint interaction

NEAREST NEIGHBOUR DISTANCE



Smaller values wrt CSR: **regular/repulsive** pattern

distances between events are larger than for a Poisson process

Greater values wrt CSR: **clustered** pattern

distances between events are shorter than for a Poisson process

Interpoint interaction

NEAREST NEIGHBOUR DISTANCE

MC test to assess the departure from CSR

1. we choose a number S of simulations, e.g. $S = 99$
2. we generate S homogeneous Poisson processes with exactly n points
3. we compute the ecdf $\hat{G}_s$, $s = 1, \ldots, S$ for all S datasets
4. at each distance r, we take the minimum and maximum value among all S ecdf $\hat{G}_s$

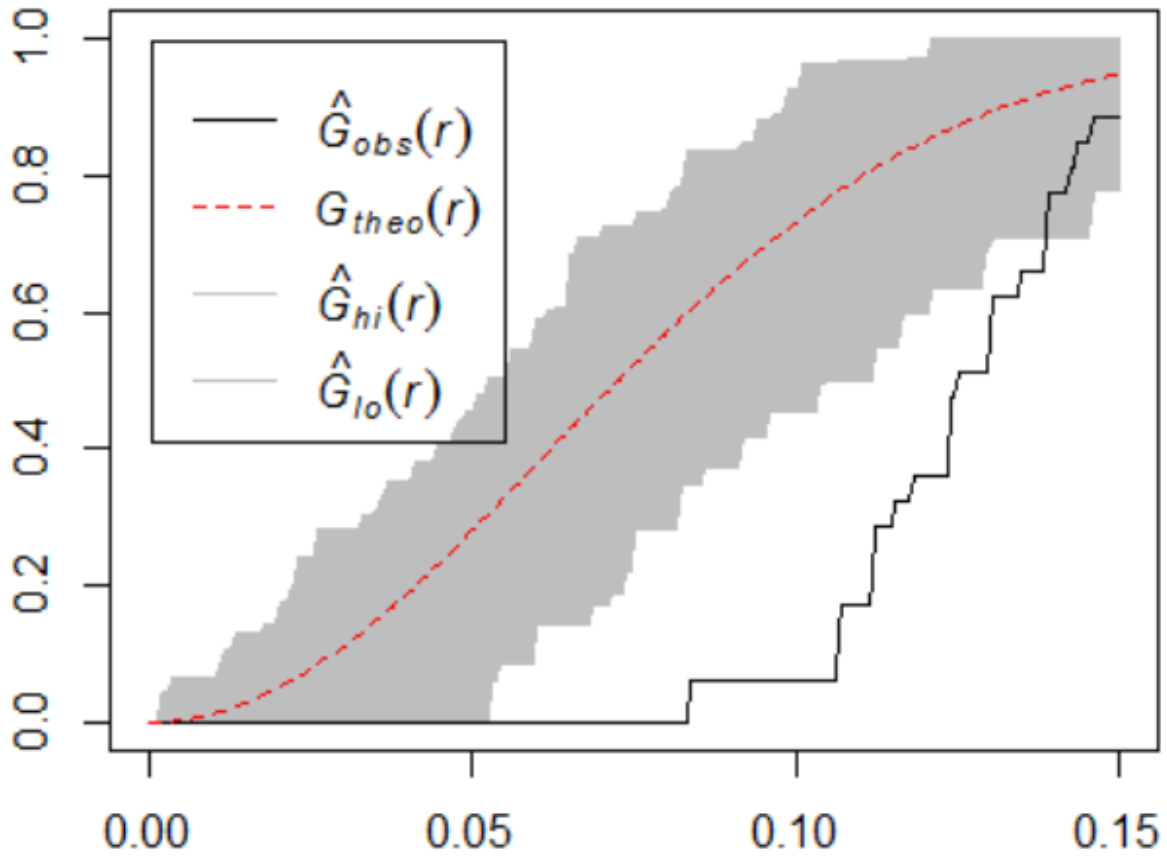
$$L_G(r) = \min_s(\hat{G}_s(r))$$

$$U_G(r) = \max_s(\hat{G}_s(r))$$

5. the two resulting curves $L_G(r)$ and $U_G(r)$ determine, at each r, the critical value of the MC test to reject H_0 =CSR
6. at each r, the test rejecting H_0 has a significance level $\alpha = \frac{2}{S+1}$ (if, e.g. $S = 99$, $\alpha = 0.02$)

Interpoint interaction

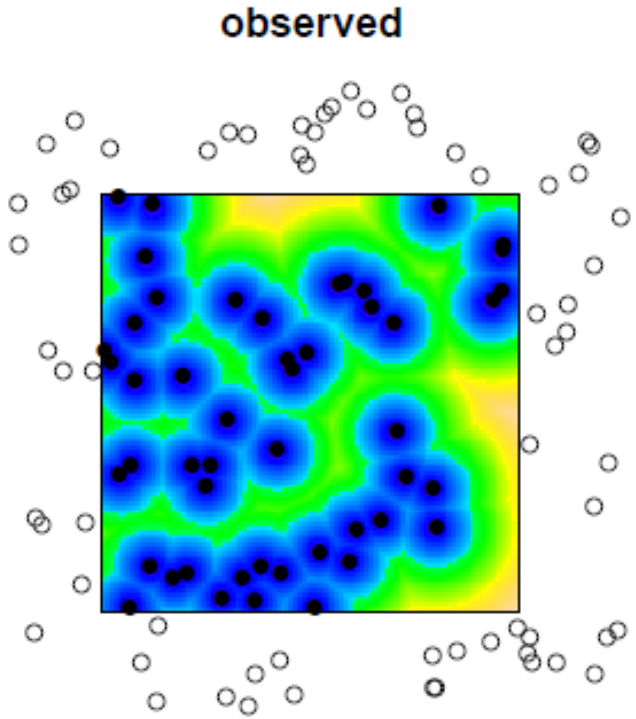
NEAREST NEIGHBOUR DISTANCE

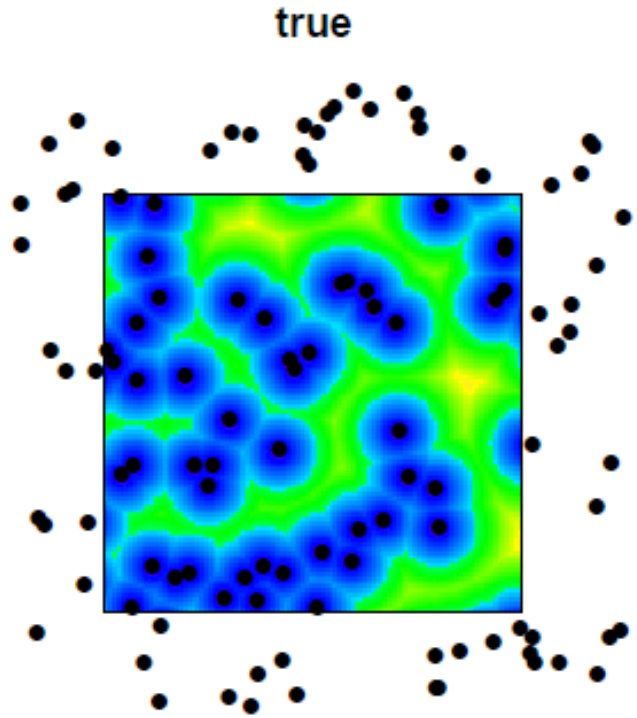


Grey area: area between $L_G(r)$ and $U_G(r)$ for 99 simulations with n points

Interpoint interaction

Issue:
we do not observe events outside the window W, still they might be very close to the events inside the window and influence them, but we do not take this into account





Interpoint interaction

Drawback:
the estimator for the G function (as for the other distance functions) is **negatively biased**, they generally overestimate distances

$\downarrow$

Solution:
apply corrections for the so-called 'edge effects'

The corrections are multiplicative factors plugged into the estimates so that estimators are unbiased

Several proposals in the literature - we will see some in the lab

$\rightarrow$ **`R script`**

Model choice

Based on explorative analysis

- if CSR is not rejected
  ↓
  1. $\lambda(u) = \lambda$ constant
  2. unbiased estimator $\hat{\lambda}$
  3. the expected number of points for this model is $\hat{\lambda}|W|$

  Model:

$$N(W) \sim Poisson(\lambda|W|)$$
$$\log(\lambda) = \beta_0$$

- if CSR is rejected
  ↓
  1. choose one of the other available point process model classes

INHOMOGENEOUS POISSON PROCESSES

Inhomogeneous Poisson process

Key element: **intensity** $\lambda(u)$ **is NOT constant** over the window.

Assumption: variations in the intensity are **only due to fixed effects** (e.g. covariates), while we do not account for

- interpoint interaction
- random effects (e.g. spatial effect)

The number of points over the window W is

$$N(W) \sim Poisson(\Lambda(W)) = Poisson\left(\int_W \lambda(u)du\right)$$

Inhomogeneous Poisson process

The intensity is modelled with a log-linear model, and can be written as

$$\log(\lambda(u)) = z(u)^{'}\beta.$$

where $z(u)$ is a covariate vector (intercept included) which is defined at each spatial location u.

The number of points is therefore distributed as

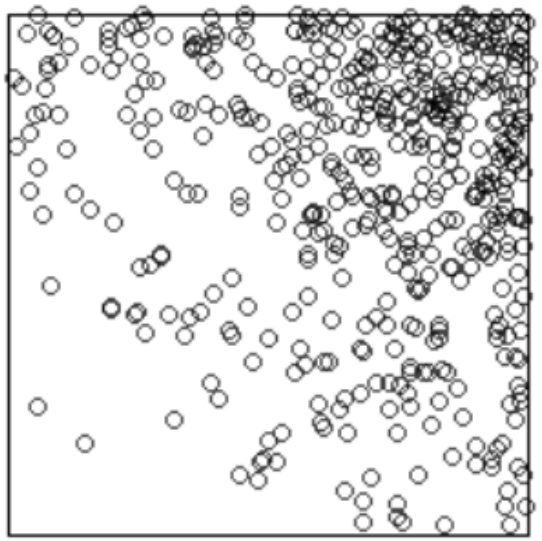
$$N(W) \sim Poisson\left(\int_W \exp\left(z(u)^{'}\beta\right) du\right)$$

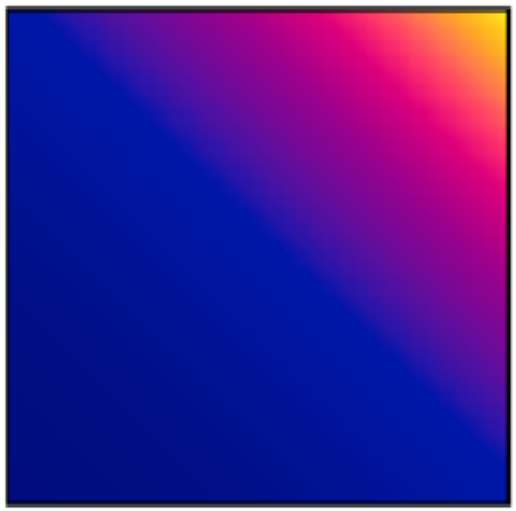
In the special case $z(u)^{'}\beta = \beta_0$, i.e. the model only includes an intercept, we get the homogeneous Poisson process with $\lambda = \exp(\beta_0)$

$$N(W) \sim Poisson\left(\int_W \exp\left(\beta_0\right) du\right) = Poisson\left(\exp\left(\beta_0\right)|W|\right) =$$
$$\sim Poisson\left(\lambda|W|\right)$$

Inhomogeneous Poisson process

Example





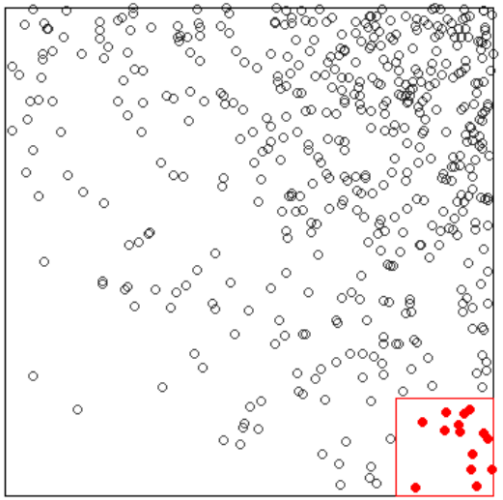
Data generating process:

$$\log(\lambda(u)) = \frac{x}{2} + \frac{y}{2}$$

$W = 5 \times 5$ square

Inhomogeneous Poisson process

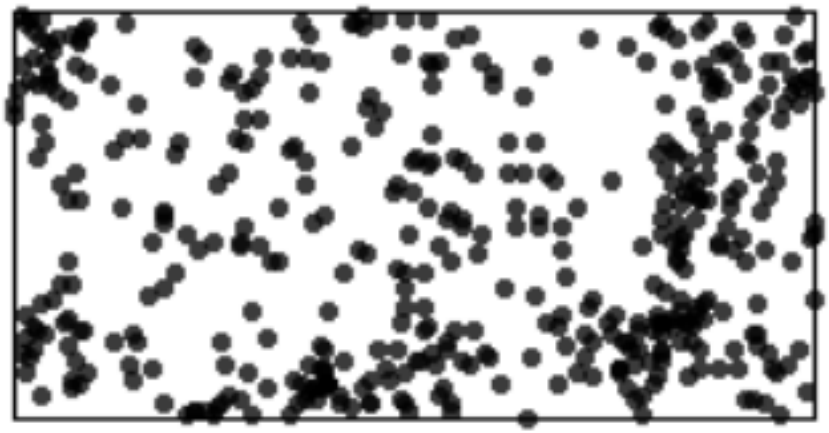
Example

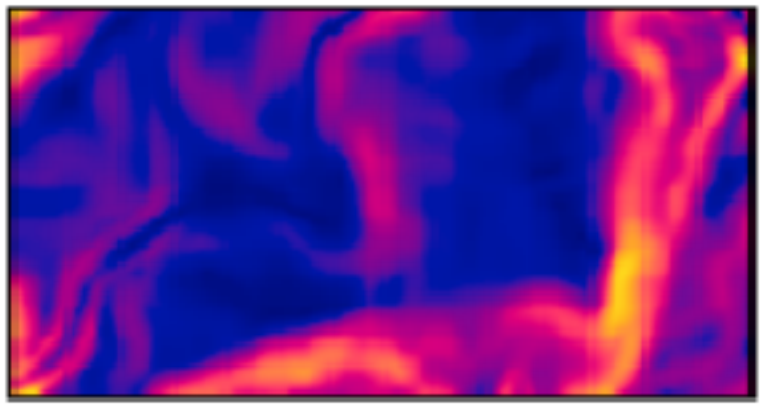


$|A| = 1$

Inhomogeneous Poisson process

Example





Data generating process:

$$\log(\lambda(u)) = s(u)$$

Likelihood

Let's recall the homogeneous Poisson process likelihood

$$L(X|\lambda) = \exp(|W|)\exp(-\lambda|W|)\lambda^n$$

and let's generalize it to a inhomogeneous process

$$\begin{aligned} L(X|\Lambda) &= \exp(|W|)\exp\left(-\int_W \lambda(u)du\right)\prod_{i=1}^{n}\lambda(x_i) \\ &= \exp(|W|)\exp\left(-\int_W \exp(z(u)'\beta)du\right)\prod_{i=1}^{n}\exp(z(x_i)'\beta) \end{aligned}$$

where the last term only depends on the intensity at the event locations

The log-likelihood is

$$l(X|\Lambda) = |W| - \int_W \lambda(u)du + \sum_{i=1}^{n}\log(\lambda(x_i)).$$

Estimating the intensity for Poisson processes

Issue

apart from the very special homogeneous case, estimating Poisson processes is hard as the integral in the likelihood is *intractable* (not analytically computable)

↓ Non parametric methods

↓ Approximate methods

↓ **Gridding**

Non parametric methods

These methods can potentially be applied to any dataset, since they do not assume a model for the data.

They are only based on the information available from the data; the use the information 'around' a location to estimate the location intensity

Example:

$$\hat{\lambda}(u) = \frac{N(du, h)}{|du, h|}$$

the intensity estimator at a location u is based on

- $N(du, h)$, the number of points within a circle of radius h centered at u
- $|du, h|$ the area of the considered circle

it is the same idea as the estimator for a homogeneous Poisson process, except that it is applied to many small areas within the window

Non parametric methods

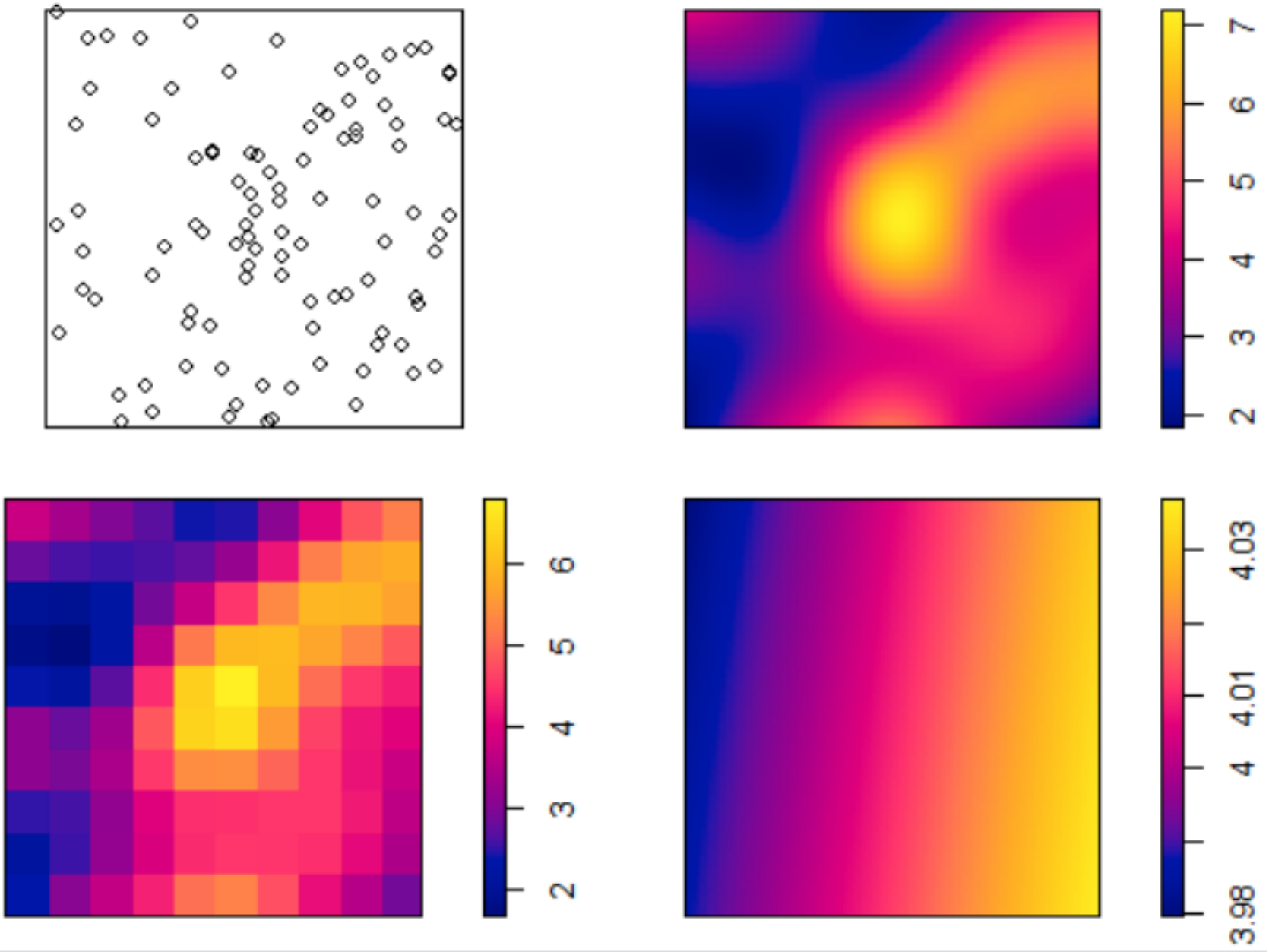
The parameter h is exogenous and it is called *bandwidth*

h defines how wide the area influencing the estimate around u is. The greater h, the more similar the estimates for different $\lambda(u)$ (the intensity function is *smooth*)

Many non parametric estimators exist, all based on the idea of a *bandwidth*

When it comes to practice, the intensity function is estimated for a fixed list of points over the window (e.g. the centroids of a very fine grid)

Non parametric methods - example



Non parametric methods - comment

Advantages of these methods

- easyness
- lack of assumptions

Limits of these methods

- everything depends on the choice of h
- they only describe the observed data, and cannot be used for prediction

$\rightarrow$ **`R script`**

Gridding

Idea
W partitioned into equivalent cells (pixels)
Y new variable = count of the number of points within each pixel

The parameter λ takes one value for each cell, i.e. it takes a finite number of values over the window

Continuous space is transformed into *discrete* space
Integrals are replaced by *sums*

Results will not be precise, but computations are much easier
The precision depends on the number of pixels

Gridding

$C_1, \ldots, C_S$ partition of W in S pixels, of constant area $|C|$
The number of points for each pixel, $N(C_s)$ is called Y_s

Let's use the properties of Poisson processes

$$Y_s \sim Poisson(\Lambda(C_s)) \text{ for } s = 1, \ldots, S$$

$$\Lambda(C_s) = \int_{C_s} \lambda(u)du$$

$$\Lambda(W) = \sum_{s=1}^{S} \Lambda(C_s) = \int_W \lambda(u)du$$

$$N(W) = \sum_{s=1}^{S} Y_s \sim Poisson(\Lambda(W)).$$

Response variable: vector $Y = (Y_1, \ldots, Y_S)'$

Gridding

The model **approximation** consists on assuming that the intensity is constant inside each pixel

$$\Lambda(C_s) \approxeq \lambda_s |C| \text{ for } s = 1, \ldots, S$$

$$\Lambda(W) \approxeq |C| \sum_{s=1}^{S} \lambda_s.$$

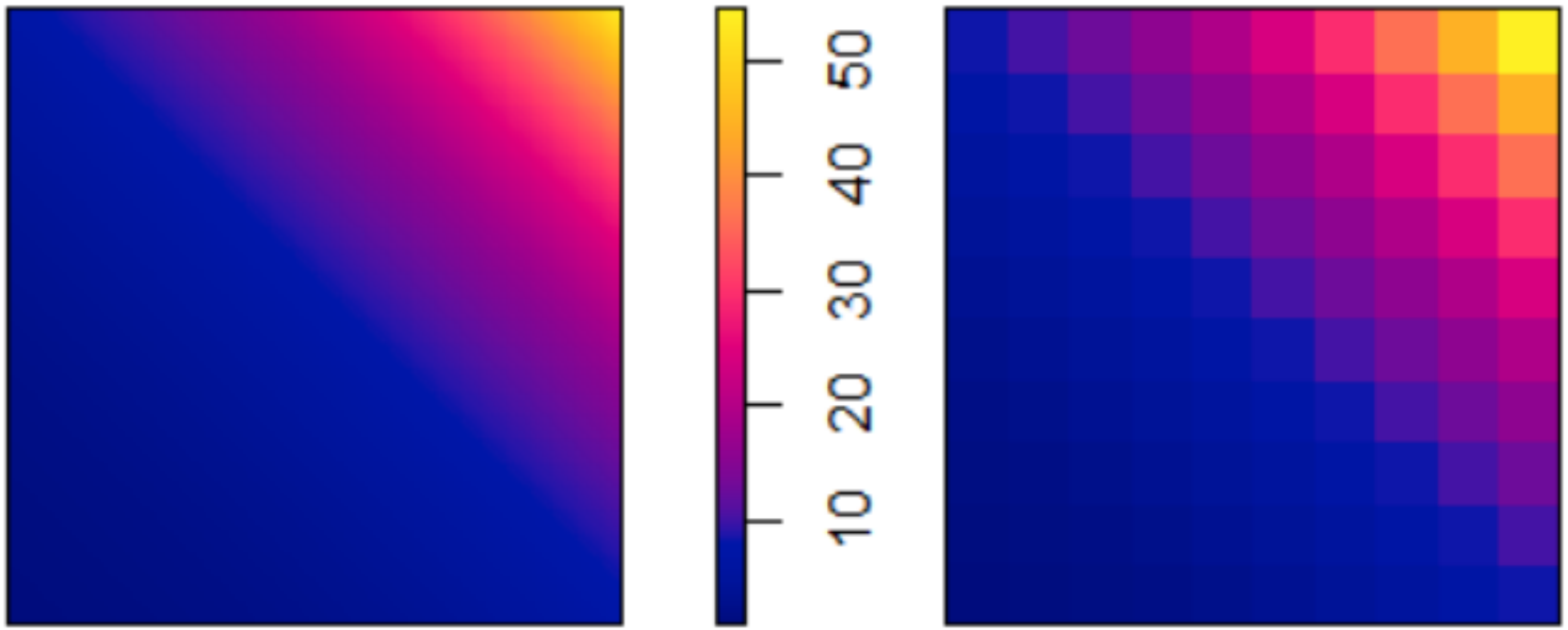
where λ_s is a representative value within cell s
Example: $\lambda_s = \exp(\bar{z}_s \beta)$ the representative value depends on the covariate average value within the cell

This allows the intractable integral to be approximated:

$$\int_W \lambda(u) du \approxeq |C| \sum_{s=1}^{S} \lambda_s$$

Gridding - example

The intensity function, which is actually continuous over space, is transformed into a piecewise linear function



Intuitively, the finer the grid, the better the approximation

Gridding - approximate model

This approximation implies we have a **homogeneous Poisson process within each pixel**, with an expected number of points equal to $\Lambda(C_s) = \lambda_s|C|$

$$Y_s|\lambda_s \sim Poisson(\lambda_s|C|)$$

The model becomes

$$\log(\Lambda(C_s)) = \log(|C|) + z_s'\beta$$

where $\log(|C|)$ is the cell offset

Gridding - approximate model

The approximate model likelihood becomes

$$L(X) \approxeq \exp(|W|) \exp\left(-|C| \sum_{s=1}^{S} \lambda_s\right) \prod_{s=1}^{S} \lambda_s^{Y_s}$$

- the integral is turned into a (tractable!)sum
- in the product term, the intensity for each cell is counted as many times as the number of points in the cell; this way, it remains a product of n terms

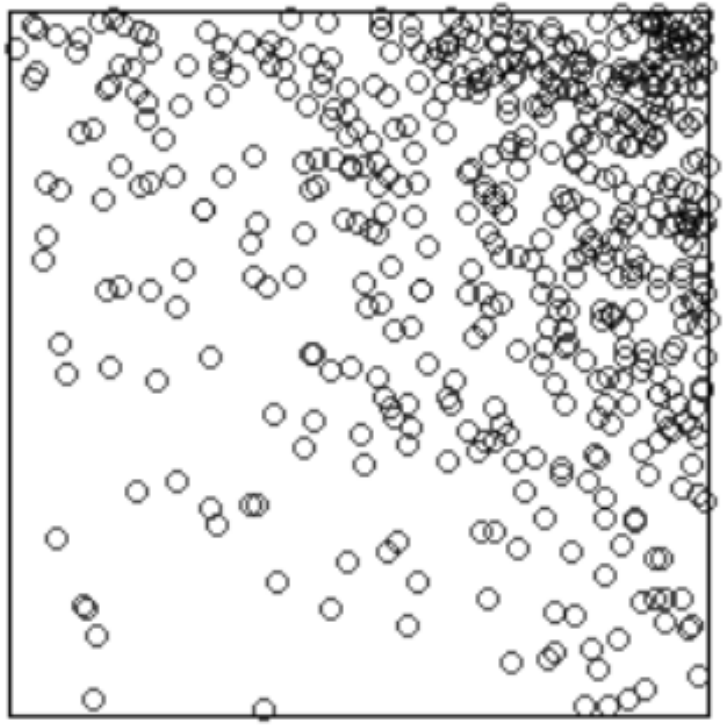
Consequently, the log-likelihood is

$$l(X|\lambda) \approxeq |W| - |C| \sum_{s=1}^{S} \lambda_s + \sum_{s=1}^{S} Y_s \log(\lambda_s)$$

Example

Gridding

Original data



Observed counts

8	12	11	10	15	28	26	43
3	7	2	12	12	12	17	23
3	2	4	4	6	13	17	26
1	4	3	3	8	7	17	21
1	2	0	5	9	9	8	18
0	2	4	0	1	5	9	7
3	0	0	3	4	2	5	8
2	0	1	0	0	1	7	4

Note: we lose detail and precision, but this is NOT an approximation
Let's assume the model $\log(\lambda(u)) = \frac{x}{2} + \frac{y}{2}$

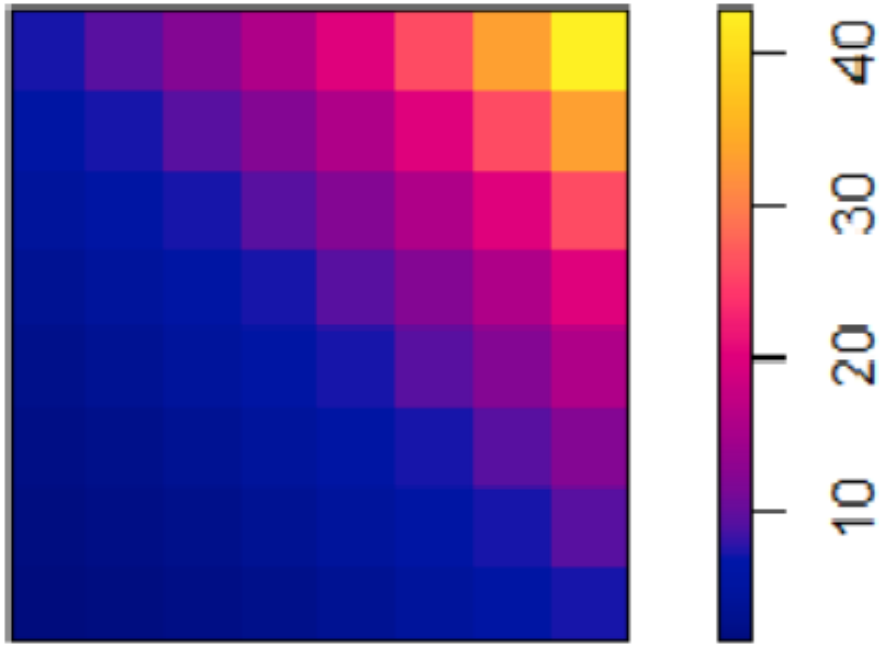
Example

Gridding

Observed counts

8	12	11	10	15	28	26	43
3	7	2	12	12	12	17	23
3	2	4	4	6	13	17	26
1	4	3	3	8	7	17	21
1	2	0	5	9	9	8	18
0	2	4	0	1	5	9	7
3	0	0	3	4	2	5	8
2	0	1	0	0	1	7	4

Expected counts

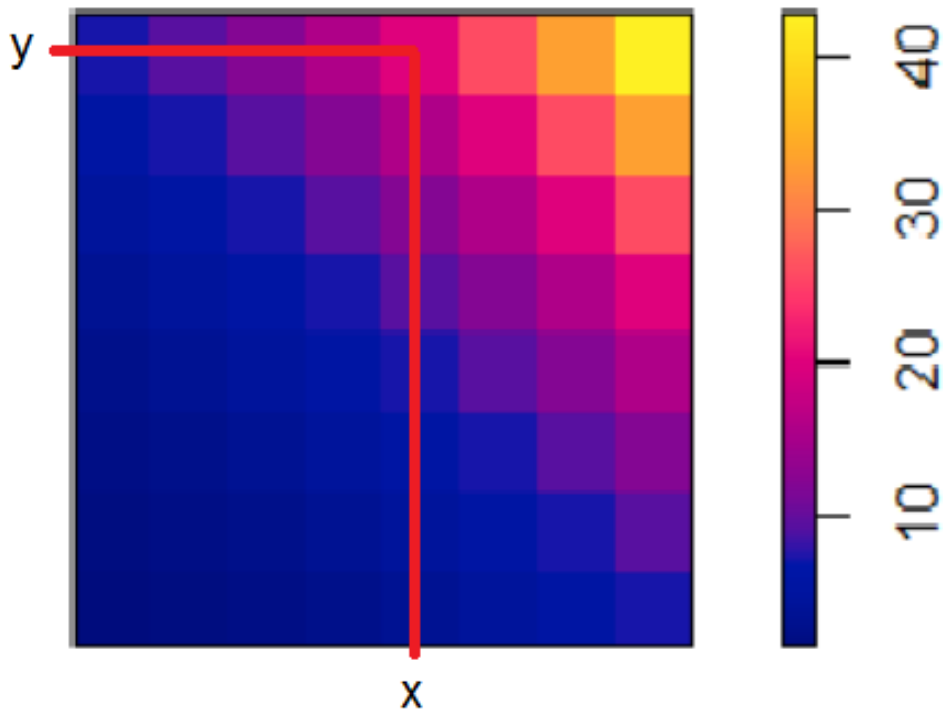


Note: here we approximate, by assuming constant expectation within each cell

Example

Gridding

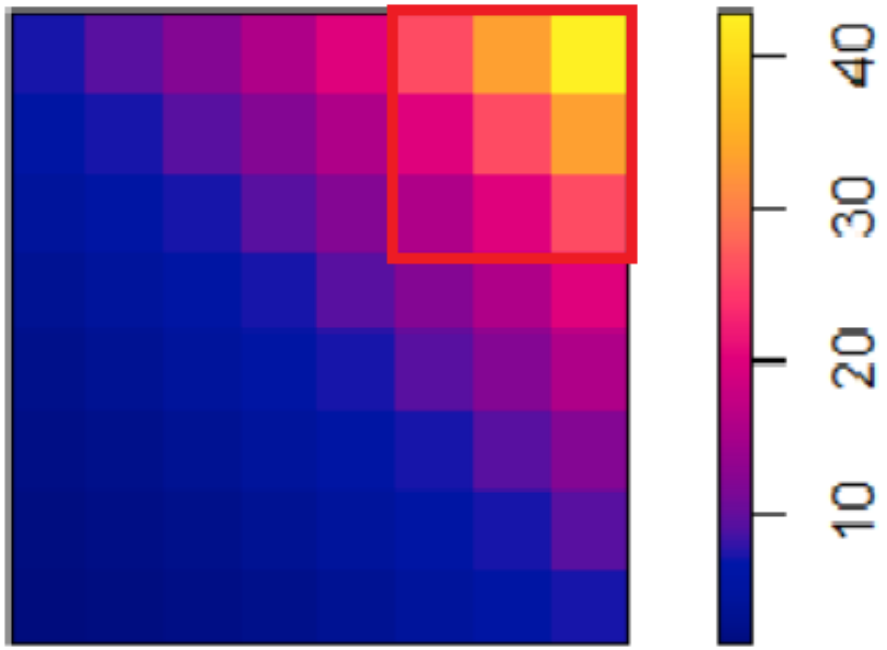
8	12	11	10	15	28	26	43
3	7	2	12	12	12	17	23
3	2	4	4	6	13	17	26
1	4	3	3	8	7	17	21
1	2	0	5	9	9	8	18
0	2	4	0	1	5	9	7
3	0	0	3	4	2	5	8
2	0	1	0	0	1	7	4



Example

Gridding

8	12	11	10	15	28	26	43
3	7	2	12	12	12	17	23
3	2	4	4	6	13	17	26
1	4	3	3	8	7	17	21
1	2	0	5	9	9	8	18
0	2	4	0	1	5	9	7
3	0	0	3	4	2	5	8
2	0	1	0	0	1	7	4



Estimation of inhomogeneous Poisson processes

Up to now, we have only approximated a process, without worrying about parameter estimation

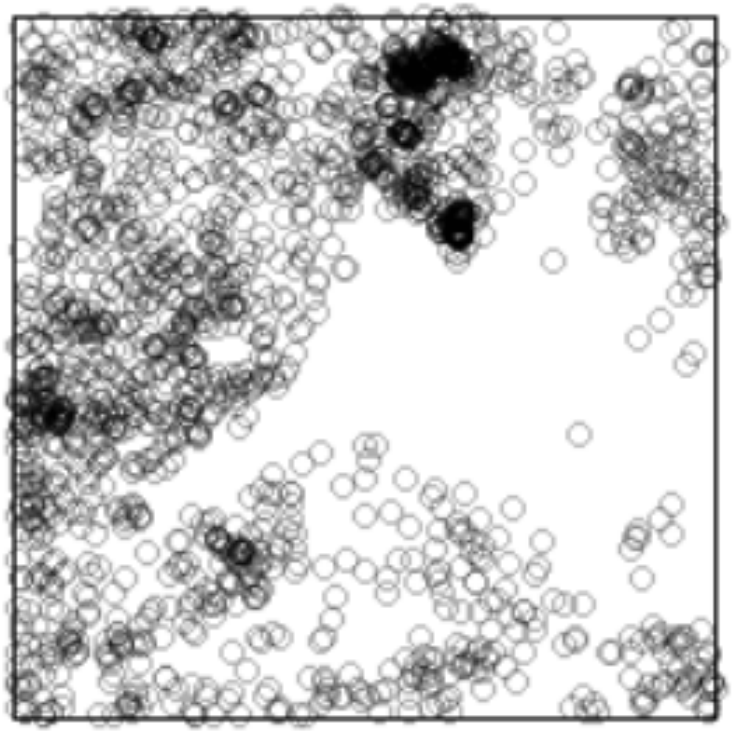
Estimate methods for inhomogeneous processes are very complex. They use the idea of gridding:

1. a model is chosen for the data
2. a grid of cell centroids is fixed
3. all covariates are evaluated
   - at the event locations - to compute the last term in the likelihood
   - at the grid centroids - to approximate the integral in the likelihood
4. advanced numerical integration techniques are used in order to obtain the β coefficients

In this course, we "trust" the estimates produced by specific functions in R

Examples of inhomogeneous Poisson processes

We will fit some simple Poisson models on this dataset

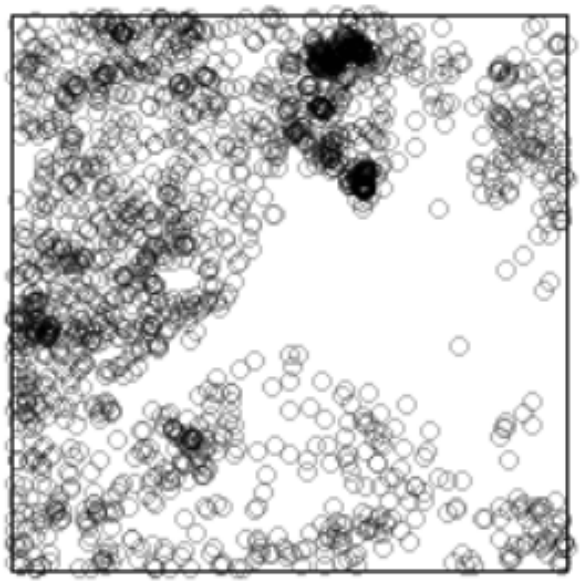


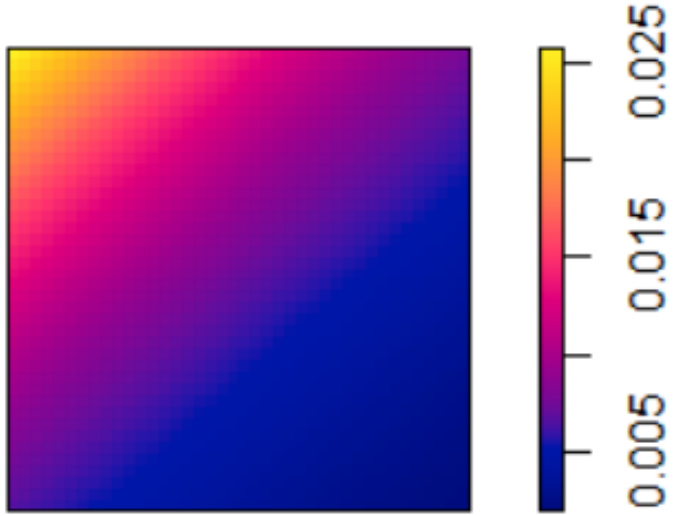
Window area $|W| = 500 \times 500 = 250{\cdot}000$ square metres
Number of points $n = 2052$
Homogeneous intensity $\hat{\lambda} = 0.008$ events for square metre

Examples of inhomogeneous Poisson processes

1. log-linear trend on the coordinates

$$\log(\lambda(u)) = \beta_0 + \beta_1 x + \beta_2 y$$





$$\beta = (\beta_0, \beta_1, \beta_2) = c(-5.025, -0.002, 0.003)$$

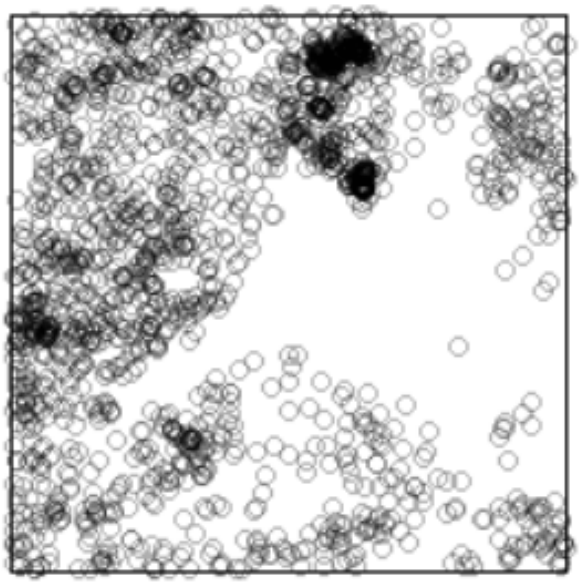
$$\lambda(u) = \exp(-5.025 - 0.002x + 0.003y)$$

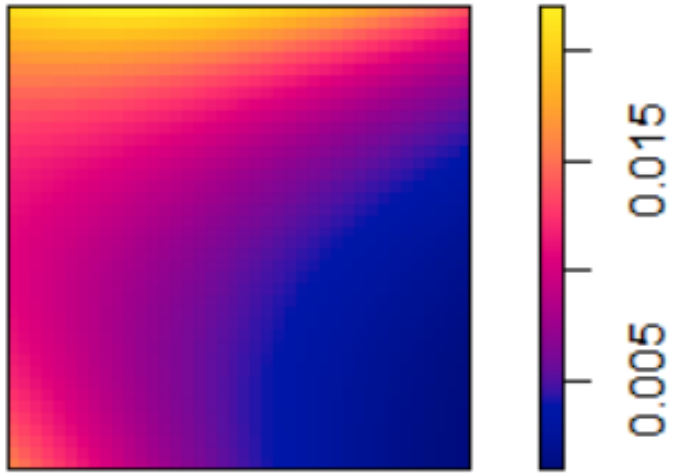
$\rightarrow$ **R script**

Examples of inhomogeneous Poisson processes

2 log-quadratic trend on the coordinates

$$\log(\lambda(u)) = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 y^2 + \beta_5 xy$$





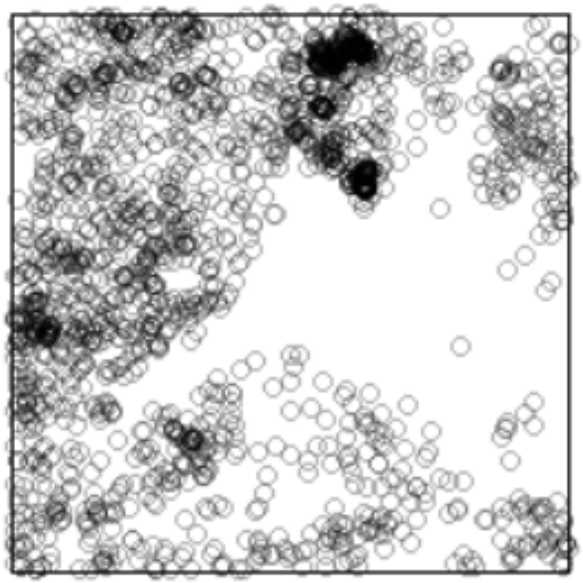
$\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4.\beta_5) =$
$(-4.117, -0.004, -0.004, 0.001, 0.001, 0.001)$

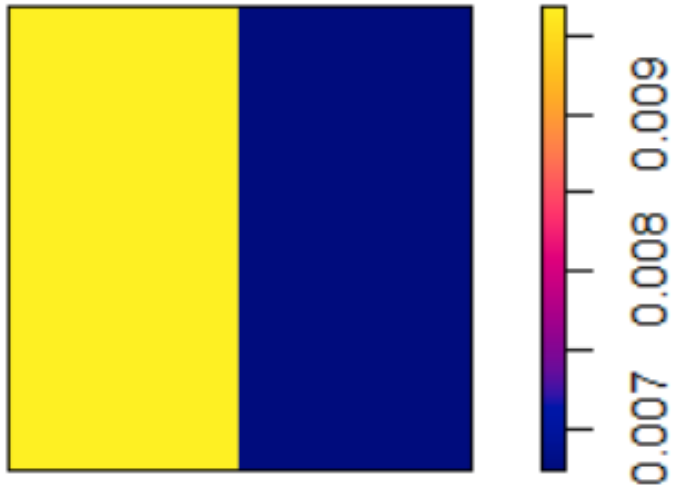
→ **R script**

Examples of inhomogeneous Poisson processes

3 two values for $\lambda(u)$: before and after a fixed coordinate x^*
let's build the dummy variable $d_{x^*} = 1$ if $x \geq x^*$

$$\log(\lambda(u)) = \beta_0 + \beta_1 d_{x^*}$$





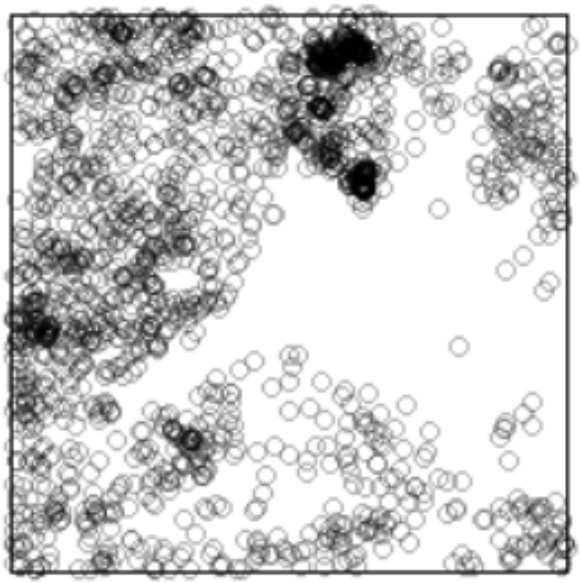
$\beta = (\beta_0, \beta_1) = (-4.638, -0.363)$

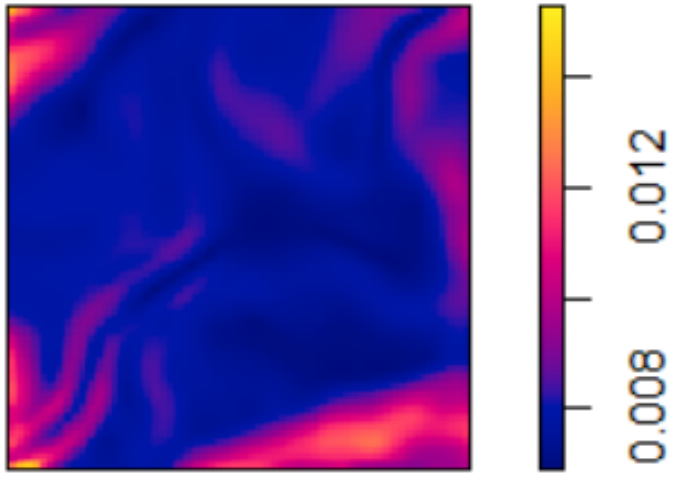
→ **R script**

Examples of inhomogeneous Poisson processes

4 log-linear dependency on a covariate

$$\log(\lambda(u)) = \beta_0 + \beta_1 z(u)$$





$\beta = (\beta_0, \beta_1) = (-4.976, 2.414)$

→ **R script**

MODEL CRITICISM

Assessment of Poisson processes

The assessment of point process models is extremely complex

We will only deal with simple tools for Poisson processes, which are similar to methods you (probably) already know

For Poisson processes, the focus is on evaluating the significance/need of fixed effects

MODEL CRITICISM
Model Diagnostics

Significance of coefficients

The significance of each coefficient for the fixed effects may be evaluated with the usual tests

t-test

Tests a simple hypothesis for a fixed effect:
H0: $\beta = 0$
H1: $\beta \neq 0$

Test statistic: $\frac{\hat{\beta}}{se(\hat{\beta})} \sim t_{df}$

where $df \approxeq n - p$ are obtained with approximate methods

Special case: homogeneous Poisson
the confidence interval for the $\hat{\beta}_0$ coefficient can be obtained starting from the interval for $\hat{\lambda}$ (see previous slides)

$\rightarrow$ **`R script`**

Goodness of fit

Goodness of fit tests rely on the null hypothesis that the model is good for the data

CHI SQUARE TEST

H_0: $N(A_p) \sim Poisson(\hat{\Lambda}_p)$, $p = 1, \ldots, P$

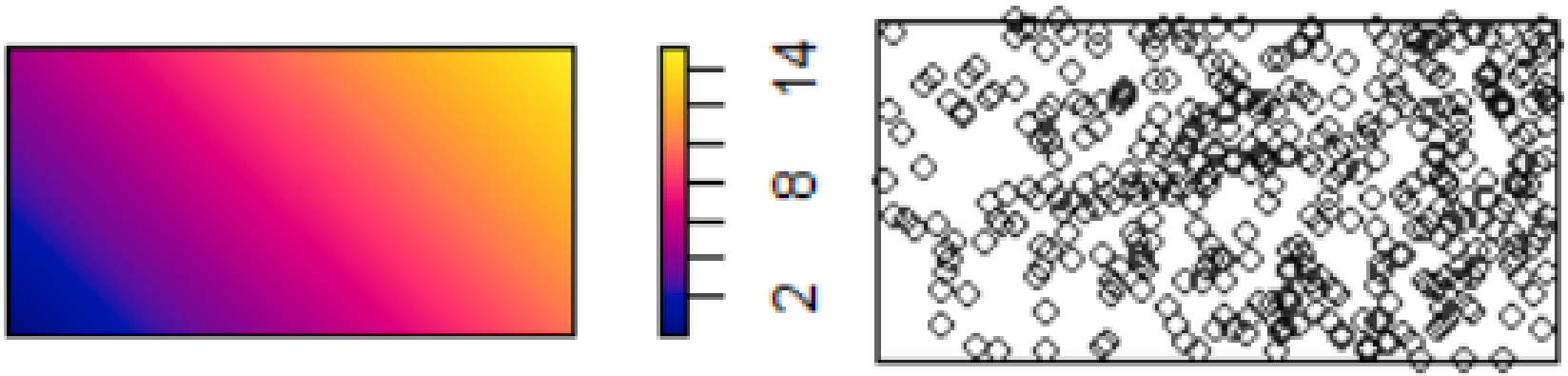
the expected counts $\hat{e}_p$ over the chosen partition in P areas depend on the model parameters

$$\chi^2 = \sum_{p=1}^{P} \frac{(o_p - \hat{e}_p)^2}{\hat{e}_p} | H_0 \sim \chi^2_{(P-b)}$$

where

- b is the number of parameters of the model
- $\hat{e}_p = \hat{\Lambda}_p$ is the expected number of points over area p under the model
- $\hat{\Lambda}_p = |C| \sum_{C_s \in A_p} \hat{\lambda}_s$

Goodness of fit - example



Data generating process:

$$\log(\lambda(u)) = x + y$$

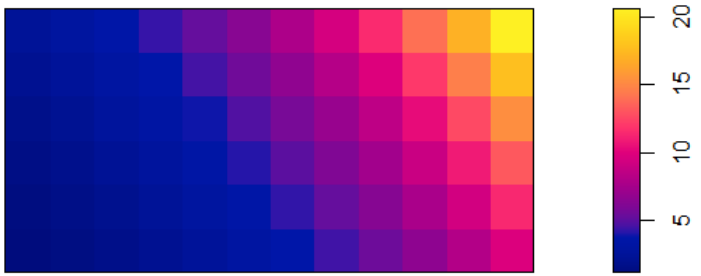
Window 10×5 units

Goodness of fit - example

Model:

$$\log(\lambda(u)) = \beta_1 x + \beta_2 y$$

Let's estimate the model using $S = 12 \times 6 = 72$ cells



For each cell C_s we have

$$\log(\hat{\lambda}_s) = \hat{\beta}_1 \bar{x}_s + \hat{\beta}_2 \bar{y}_s$$

Under the model, we expect to observe a number of points equal to

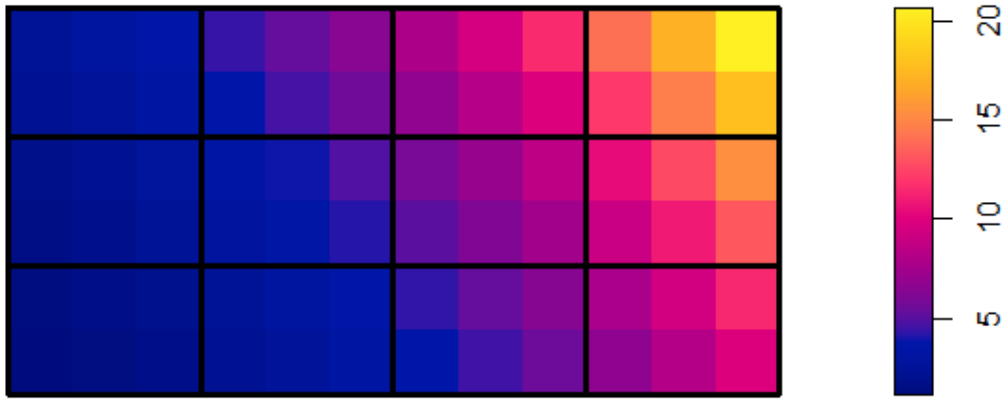
$$\hat{\lambda}_s |C_s|$$

for each cell, and a total number of points equal to

$$\sum_s \hat{\lambda}_s |C_s|$$

Goodness of fit - example

Let's choose $P = 12$ areas for the χ^2 test



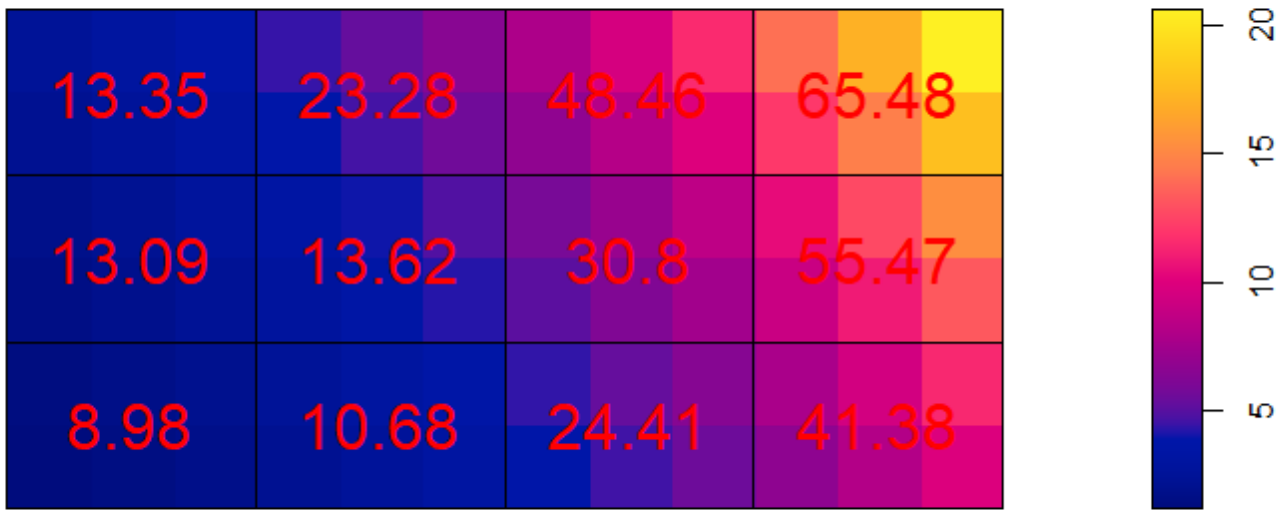
For each area A_p we have

$$N(A_p) \sim Poisson(\hat{\Lambda}(A_p))$$

$$\hat{\Lambda}(A_p) = \sum_{C_s \in A_p} \hat{\lambda}_s |C|$$

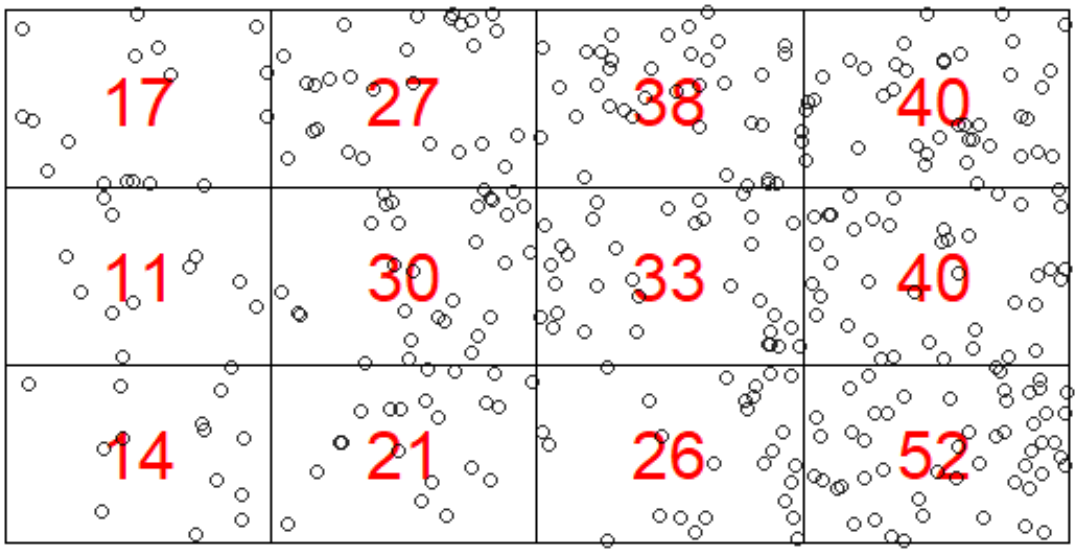
Goodness of fit - example

Expected number of points for each area $N(A_p)$



Goodness of fit - example

Observed number of points for each area n_p



Goodness of fit - example

H_0: $N(A_p) \sim Poisson(\hat{\Lambda}_p), p = 1, \ldots, P$

$$\chi^2 = \sum_{p=1}^{P} \frac{(n_p - \hat{\Lambda}_p)^2}{\hat{\Lambda}_p} | H_0 \sim \chi^2_{10}$$

p-value < 0.001

The model does not fit the data well

Goodness of fit - example

Model for data 'bei':

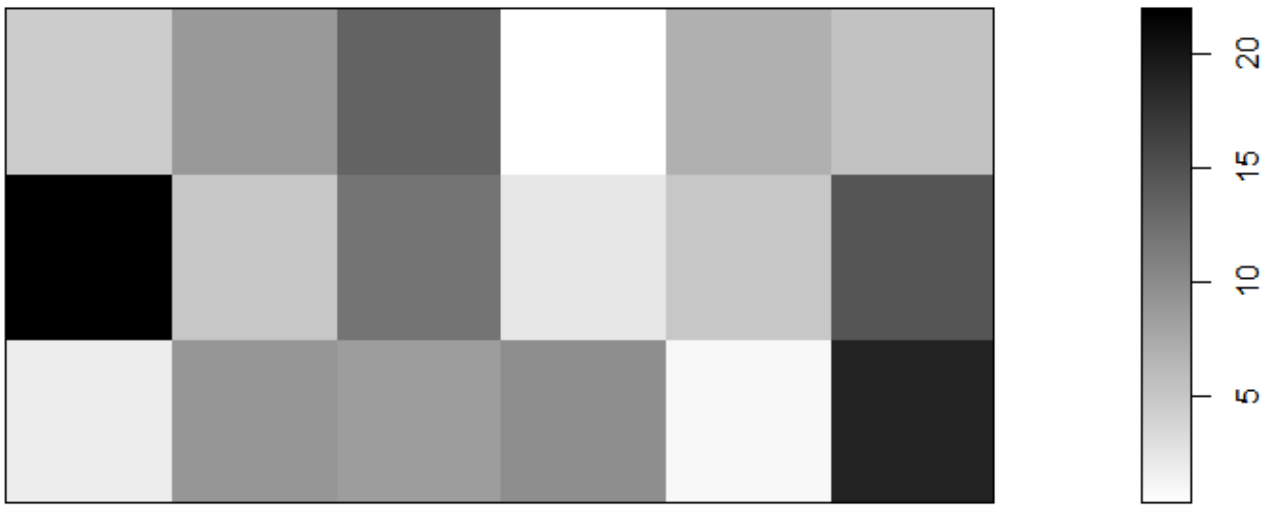
$$\log(\lambda(u)) = \beta_0 + \beta_1 x + \beta_2 y$$

380 356.3 1.3	608 303.2 18	178 258.2 -5	56 220 -11	331 187.3 11
451 325.9 6.9	21 277.8 -15	53 236.7 -12	52 201.4 -11	237 171.6 5
272 286.4 -0.85	125 243.9 -7.6	259 207.6 3.6	473 176.9 22	108 150.7 -3.5

$\chi^2 = 1708 | H_0 \sim \chi^2_{12}$
p-value< 0.001

Goodness of fit - example

Pearson's residuals (in absolute value) may be used to evaluate where the model performs the worst



→ **R script**

Goodness of fit

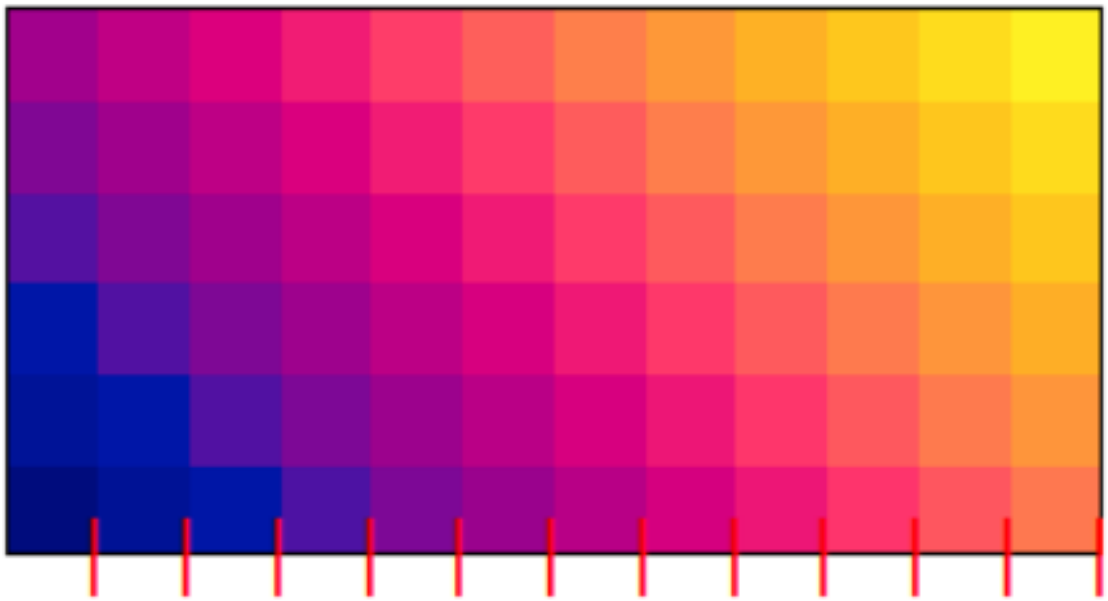
KOLMOGOROV SMIRNOV TEST

H_0: the empirical distribution of the spatial covariate follows the expected distribution under the model

Example: x covariate

- we choose a sequence of values x_p in W
- expected distribution under the model: based on the estimated intensity, for each x_p we compute the expected proportion of points having coordinate $x \leq x_p$
- empirical distribution: based on data, for each x_p we compute the observed proportion of points having coordinate $x \leq x_p$

Goodness of fit - example

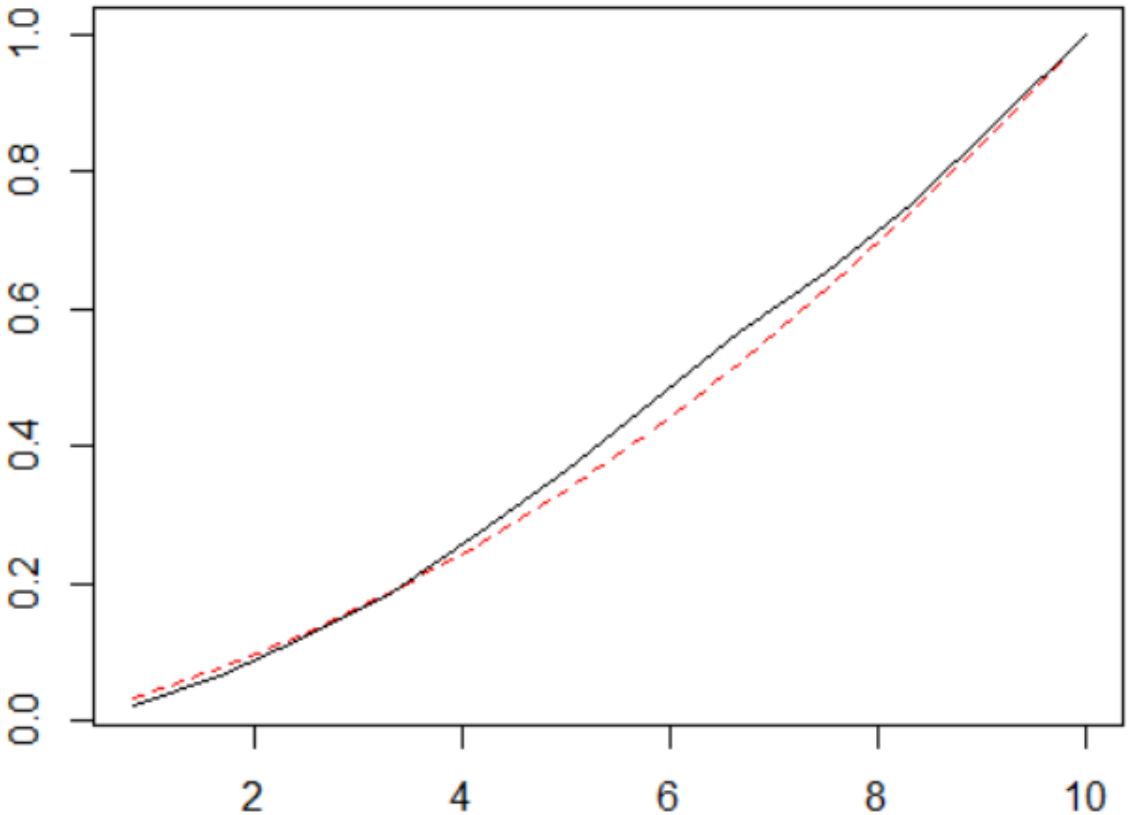


- r is chosen using the bottom-right x coordinates of each pixel

- we compute the cdf under the model (i.e. using $\hat{\Lambda}_s$)

- we compare it to the empirical cdf (by counting the number of points left to each value of r)

Goodness of fit - example



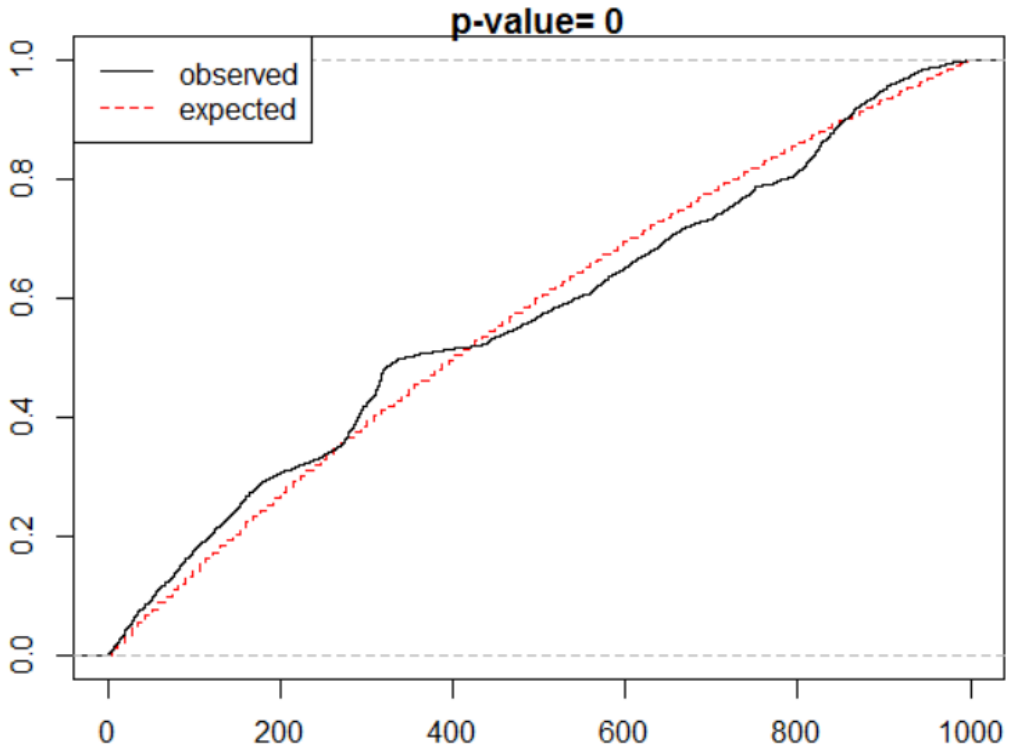
The Kolmogorov Smirnov rejects the null hypothesis of having a good model (p-value < 0.001)

Goodness of fit - example

Model for data 'bei':

$$\log(\lambda(u)) = \beta_0 + \beta_1 x + \beta_2 y$$

KS test with covariate x



$D_n = 0.07$, p-value< 0.001

$\rightarrow$ **R script**

Residual diagnostics

Generally, it is of interest to evaluate residual correlation

Many traditional diagnostic tools are useless for point processes

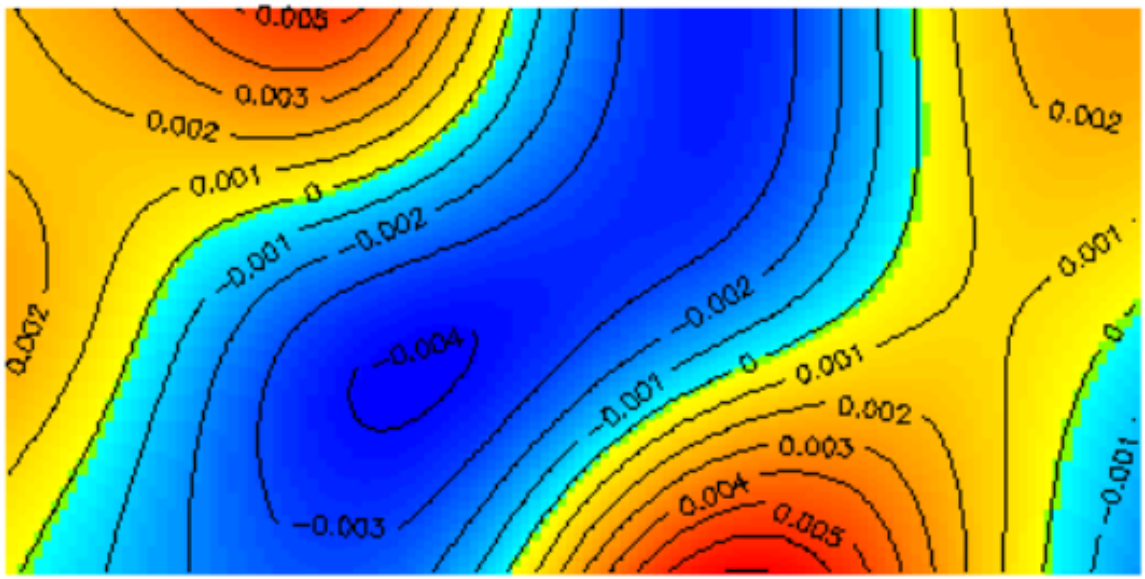
We use a plot of 'smoothed' residuals, using a nonparametric technique similar to the ones employed for intensity estimation.
If a trend can be seen, residuals are correlated and the model is not good for the data

Note: it is a visual/informal test

Residual diagnostics - example

Model for data 'bei':

$$\log(\lambda(u)) = \beta_0 + \beta_1 x + \beta_2 y$$



→ **R script**

MODEL CRITICISM
Model Comparison

Information criteria

Methods for evaluating a model based on

- parameter estimates
- penalty for complexity

This applies to any model, provided they are fit on the same set of data

AIC (Akaike Information Criterion)

$$AIC = -2l(\hat{\beta}, \hat{\theta}) + 2p$$

$AIC \in [-\infty; \infty]$ and smaller values are preferred

AIC are automatically available in R for all Poisson processes

$\rightarrow$ **`R script`**

Model comparison

Likelihood ratio test

It is a test for nested models.

It requires fitting both models on the same set of data

$$LRT = -2 \ln \left(\frac{L_0}{L_1} \right) = 2(l_1 - l_0)$$

where p_1 and p_0 are the numbers of parameters of the two models

The two likelihoods are computed by plugging in the fixed effects estimates for M_0 and M_1.

H_0: M_0 is sufficient for explaining the data behaviour wrt M_1

$$LRT | H_0 \sim \chi^2_{(p_1 - p_0)}$$

Large value = evidence in favour of H_1

$\rightarrow$ **`R script`**