

Chapter 1. Introduction

Dra. Rosa Abellana

Departament Fonaments Clínics. Facultat de Medicina. Universitat de Barcelona
rabellana@ub.edu
<https://orcid.org/0000-0003-4487-4431>

Table of contents

1 Introduction

- Type of spatial data

2 Geostatistical data

- Definition
- Examples

3 Lattice data

- Definition
- Examples

4 Point pattern

- Definition
- Examples

5 Parametrization spatial data

6 Spatial Structure

Introduction

- **Spatial data** consists of a series of geo-referenced observations: for each individual of the sample, in addition to annotating the values of the interested variables, also takes note of the exact place where produces the data.
- **Geo-referenced data** is interesting when the position of the data on the space contains notable information for the analysis.
- The **spatial data analysis**, the space information is taken into account in the models and predictions.

Geographic coordinates (longitude, latitude) can be used:

- as predicted variables (tendency)
- to define the spatial dependence: observations nearby in space tend to be alike

Geographic LOCATIONS :

- by points (coordinates:longitude east-west and latitude north-south)

Example: Temperature registered in a meteorological station at 12:00 a.m. on the 5th July

- by area (delimited region)

Example: Number of births recorded in January 2012 in the province of Barcelona

Distribution of the points can be:

- *regular o gridded.*

Example of points: Height above the sea level measured at the intersections of meridians and parallels, separated by 1 degree.

Example of area: Areas are the pixels defined with a satellite image.

- *irregular or non-gridded.*

Example of points: Temperature registered in meteorological stations of a region.

Example of area: Number of cardiovascular deaths reported in January of 2010 in each Spanish region.

Type of spatial data

The spatial data are classified in three types:

- ① Geostatistical data
- ② Lattice data
- ③ Point process

1. GEOSTATISTICAL DATA:

- Point observations of a continuously varying quantity over a region.
- The variable can be measured wherever of the study region.
- The researcher chooses a few observation points and measures the value of the variable.
- The data is referenced by points.

Examples:

- Weekly concentrations of ozone in U.S.
Annual acid rain deposition.
- Richness of iron within an ore body.

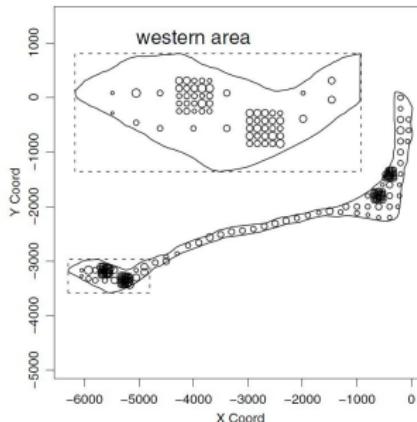
- Background: One important problem in geostatistics is to **predict** the ore grade in a mining block from observed samples (Matheron, 1963).
The main objective of geostatistical analysis is to predict the variable of study from observed samples throughout a region of study.
- Geostatistical data presents a **positive spatial correlation**.
- The spatial correlation is fitted by a function of distance among the points.

Rongelap Island (South Pacific)

Data format (x_i, y_i, t_i) :

- $i = 1, \dots, 157$
 x_i identifies a spatial location
- y_i is a photon emission count attributable to radioactive caesium
- t_i is the time (in seconds) over which y_i was accumulated
(Diggle et al. 1988)

1

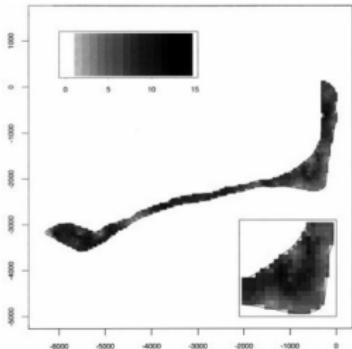


¹

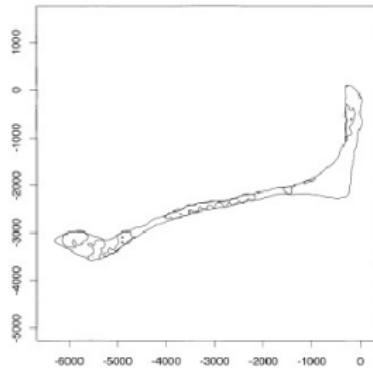
<http://www2.stat.duke.edu/~fei/samsi/Readings/DiggTawnMoye1988.pdf>

Introduction
Geostatistical data
Lattice data
Point pattern
Parametrization spatial data
Spatial Structure

Definition
Examples



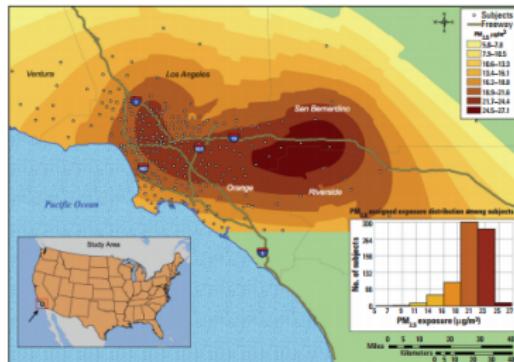
Predicted intensity (counts per units time)
of ^{137}Cs



Contour plot of the value 0.05 of the probability
of Cs intensity exceeding 15 units per unit the time

Ambient Air Pollution and Atherosclerosis in Los Angeles

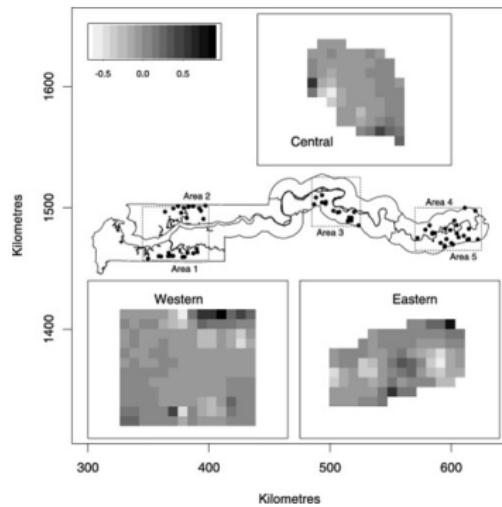
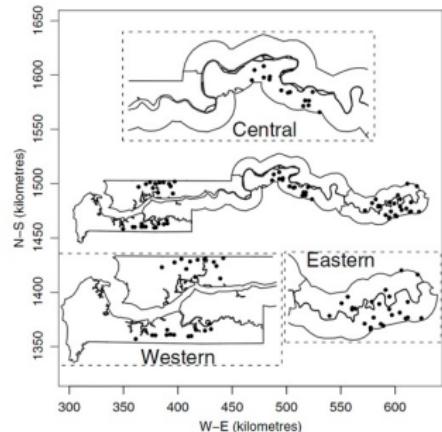
- Objective: Investigated association between residential ambient $PM_{2.5}$ and carotid artery intima-media thickness Exposition levels of pollution:
- 23 state and local district monitoring stations
- $PM_{2.5}$ data were interpolated and linked with the ZIP code centroids of each subject.



Childhood malaria in the Gambia

Data information:

- Cross-sectional (2,276 children)
1–4 years of age
- 65 villages



2. LATTICE DATA

- The locations are areas instead of coordinates for a given point of interest.
- Regions are pre-defined, normally by administrative criterion: (*secciones censales, municipios, countries, ...*).
- Counts or averages of a quantity on subregions that make up a larger region.
- Statistical models for lattice data need to express the fact that observations nearby in space tend to be alike (not predictable pattern)

Examples:

- Number of deaths due to SIDS in the counties of NC.
Incidence rate of diabetes in the *comarcas* of Catalonia.

Spatio-seasonal modelling of the incidence rate of malaria in Mozambique

- Data: Malaria incidence rate.
Region: Manhiça area (Province of Maputo, Mozambique).
- Cohort children less or equal to 10 years old.
- Following period: December 1996 - July 1999.
- 115 administrative neighbourhoods.
- Covariates: Sex, Climate Seasons, Age, Period.

Spatio-seasonal modelling of the incidence rate of malaria

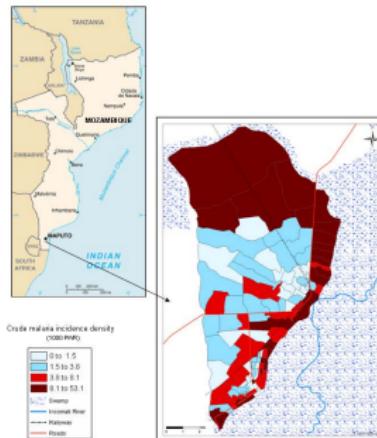
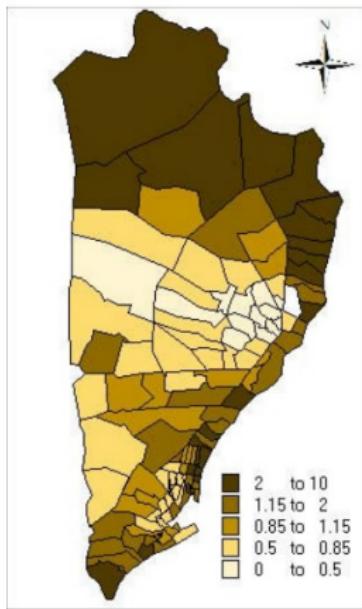


Figure: Area of study and crude malaria incidence for each neighbourhood ²

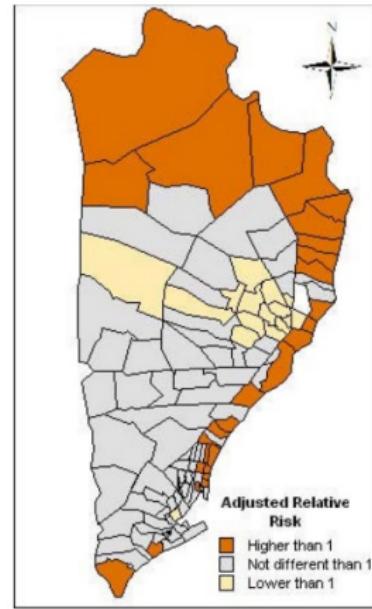
²<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2584655/>

Introduction
Geostatistical data
Lattice data
Point pattern
Parametrization spatial data
Spatial Structure

Definition
Examples



Neighbourhood-specific adjusted relative risk estimated
by the spatial model

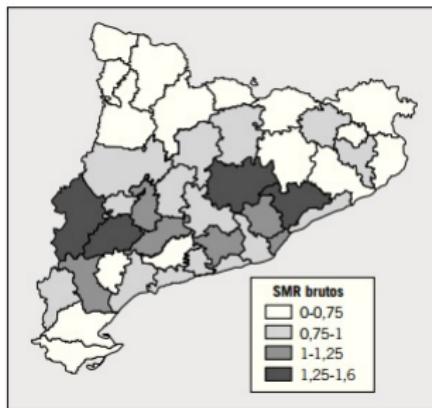


Posterior probability of the neighbourhood-specific adjusted
relative risk to be different than 1.

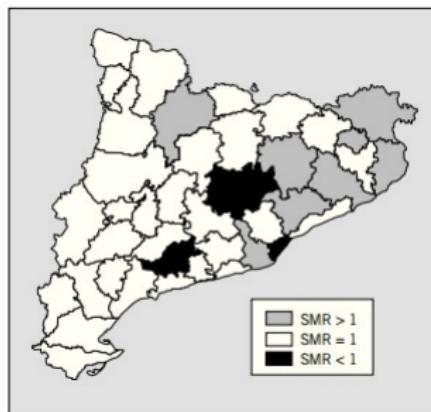
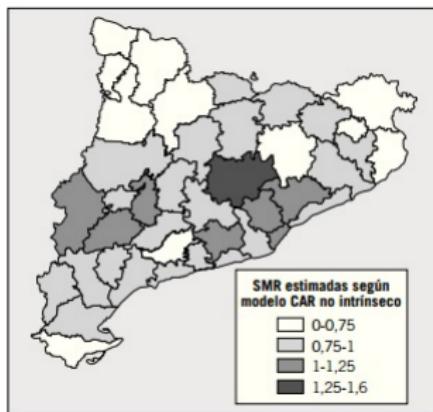
Incidence rate of diabetes type I in Catalonia

Data: Incidence of Type 1 diabetes in Catalonia. Subjects younger than 30 years. Regions: 40 comarcas. Period:1989-1998.

Figure: Incidence rate of diabetes type I in Catalonia



Incidence rate of diabetes type I in Catalonia



3. POINT PATTERN

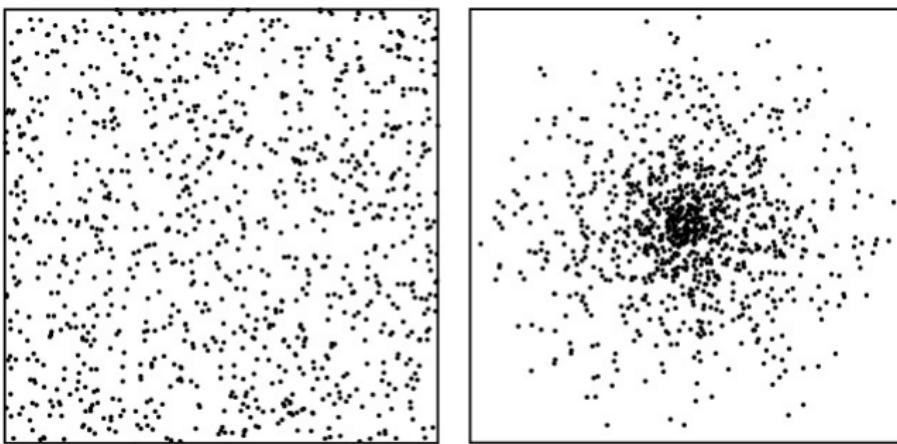
- Points on a map. (The researcher doesn't select the points)
- Point patterns arise when the important variable to be analysed is the location of events.
- The question of interest is whether the pattern is exhibiting complete spatial randomness, clustering, or regularity.

Examples

- Location of bird's nests in a suitable habitat.
Location of longleaf pines in a natural forest in NC – evidence of clustering?
- Diameter of the longleaf pines – Do larger trees cluster?

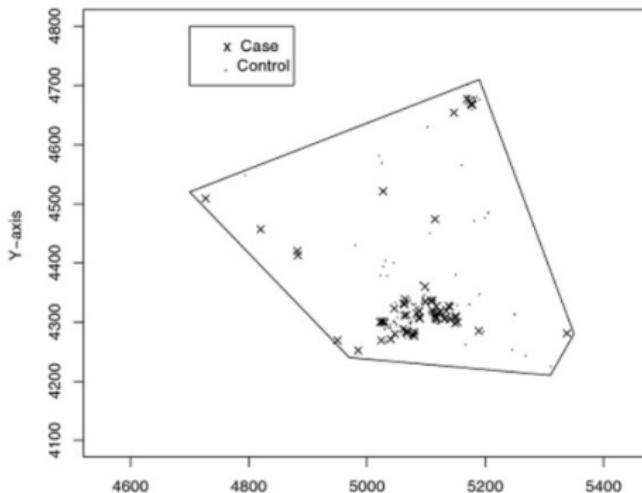
The size variable (in example 3) is usually called a mark variable, and the process is called a *marked spatial point process*.

Figure: Examples of point pattern



Leukemia in North Humberside (England)

Data Case : 62 cases of childhood leukemia and lymphoma diagnosed between 1974 and 1986. Data Control: 141 controls selected at random. Locations: Centroids of postal codes of home addresses



Potential source of pollution

Data Case :Respiratory cancer in the vicinity of a former industrial waste incinerator (58 cases). Region: In south Lancashire. Period: 1974-83

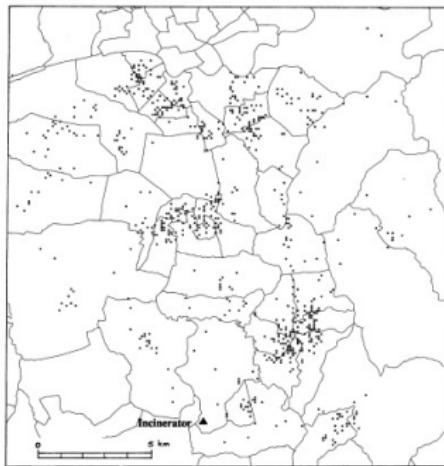


Figure 7 Locations of lung cancers, Chorley and South Ribble, Lancashire, 1974-83

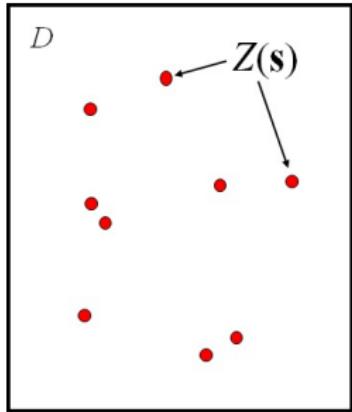
3. PARAMETRIZATION SPATIAL DATA

- Cressie (1993) propose a mathematics model to include the three types of spatial process:
Let $\mathbf{s} \in R^d$ the location of a data (d usually is 2).
Let $\mathbf{Z}(\mathbf{s})$ random variable (o random vector) that is measured on the place \mathbf{s} .
- Stochastic process:

$$\{\mathbf{Z}(\mathbf{s}) : \mathbf{s} \in D\}.$$

The subset D can be fixed or random.

One realization of the stochastic process is denoted as
 $\{\mathbf{z}(\mathbf{s}) : \mathbf{s} \in D\}.$



D: is the spatial domain or area of interest

s: contains the spatial coordinates

Z: is a value located at the spatial coordinates

Types of spatial Data:

$$\{\mathbf{Z}(\mathbf{s}) : \mathbf{s} \in D\}.$$

Geostatistics: Z random; D fixed, infinite, continuous

Lattice Models: Z random; D fixed, finite, (ir)regular grid

Point Patterns: Z = 1; D random, finite

Spatial point pattern: Z=1 Marked point pattern: Z random

SPATIAL STRUCTURE

- Large-scale structure (Global, over the entire region)

Mean function of a geostatistical process

- Mean vector of lattice data
- Intensity of spatial point process

Small-scale structure (Local, highly localized region)

- Variogram, covariance function of geostatistical process
- Neighbour weights for lattice process
- Nearest-neighbour functions for spatial point process

- Geostatistical data

- GeoR: Analysis of Geostatistical Data

<https://cran.r-project.org/web/packages/geoR/geoR.pdf>

- Gstat: Spatial and Spatio-Temporal Geostatistical Modelling, Prediction and Simulation

<https://cran.r-project.org/web/packages/gstat/gstat.pdf>

- Lattice data

- Maptools

<http://cran.r-project.org/web/packages/maptools/index.html>

- spdep: Spatial Dependence

<http://cran.r-project.org/src/contrib/Descriptions/spdep.html> or
<http://spatial.nhh.no/R/spdep/>

- R2WinBUGS: Running 'WinBUGS' and 'OpenBUGS' from R-package

<https://cran.r-project.org/web/packages/R2WinBUGS/R2WinBUGS.pdf>

- Point Process

- Spatial Point Pattern Analysis, Model-Fitting, Simulation, Tests

<https://cran.r-project.org/web/packages/spatstat/spatstat.pdf>

Bibliography

- Banerjee, S.; Carlin, BP.; Gelfrand A.E. *Hierarchical modelling and analysis for spatial data*. Chapman and Hall CRC, 2004.
- Bivand, R. S.; Pebesma, E. J.; Gómez-Rubio, V. *Applied spatial data analysis with R*. Springer, 2008.
- Cressie, N.A.C. *Statistics for spatial data*. New York: John Wiley and Sons, 1993.
- Abellana R, Ascaso C, Aponte J, et al. Spatio-seasonal modeling of the incidence rate of malaria in Mozambique. *Malaria Journal*. 2008;7:228. doi:10.1186/1475-2875-7-228.
- Abellana R, Ascaso C, Carrasco JI, Castell C, Tresserras R. Geographical variability of the incidence of Type 1 diabetes in subjects younger than 30 years in Catalonia, Spain. *Medicina Clínica*. 2009;4:454.

2.EXPLORATORY DATA ANALYSIS

Rosa Abellana

Department of Basic Clinical Practice. University of Barcelona

Table of contents

- 1 Geostatistical model
- 2 Stationary and Isotropy
- 3 Exploratory data analysis
- 4 Descriptive analysis methods. Large-scale variation

1. GEOSTATISTICAL MODEL

General model for spatial data:

$$\{\mathbf{Z}(\mathbf{s}) : \mathbf{s} \in D\},$$

and in case of geostatistical data:

- $D \subset \mathbb{R}^d$ and $d = 2$ (usually)
- spatial locations \mathbf{s} are fixed.

A very popular statistical model for many kinds of data:

$$\text{Data} = \text{mean} + \text{residual}$$

where:

- mean: non-random quantity (a number)
- residual: random variable with mean zero.

Geostatistical model:

$$Z(s) = m(s) + \varepsilon(s)$$

where:

$m(s)$: Large-scale variation, function of mean

$\varepsilon(s)$: Small-scale variation. It is a zero-mean random field, associated with a covariance function C :

$$C(s_i, s_j) = \text{cov}(Z(s_i), Z(s_j)) = \text{cov}(\varepsilon(s_i), \varepsilon(s_j))$$

To fit geostatistical process needs identify both types of variability.

The covariance function C satisfies:

- *Symmetry*, i.e. $C(s_i, s_j) = C(s_j, s_i)$ for all $s_i, s_j \in D$.
- *Non-negative definiteness*,

$$\sum \sum a_1 a_2 C(s_i, s_j) \geq 0$$

for all n , all sequences $a_i : i = 1, \dots, n$ and all sequences of spatial locations $s_i : i = 1, \dots, n$. (Non-negative definiteness).

- **Stationary.** Three types of stationarity:

- ① **Strict stationarity.** The joint probability distribution of the data depends only on the relative positions of the sites at which the data were taken (the process is invariant to translations). The joint distribution of

$$(\mathbf{Z}(s_1), \dots, \mathbf{Z}(s_k))$$

is the same as

$$(\mathbf{Z}(s_1 + v), \dots, \mathbf{Z}(s_k + v))$$

for any m spatial points s_1, \dots, s_k and any $v \in D$.

2 Weak stationary (second-order stationary).

- the mean is constant, $E(Z(s)) = \mu$ for all $s \in D$.
- the covariance at two sites depends on only the sites relative positions.

$$C(Z(s_1), Z(s_2)) = C(s_1 - s_2) \text{ for all } s_1, s_2 \in D$$

Weak stationary implies that the variance of the process does not depend on the locations:

$$\text{Var}(Z(s)) = C(0) = \sigma^2 \text{ for all } s \in D$$

3 Intrinsic stationary.

This is the more general stationary.

- the mean is constant, $E(Z(s)) = \mu$ for all $s \in D$.
- the $\text{Var}[Z(s_1) - Z(s_2)]$ depends on only the sites relative positions $s_1 - s_2$, there is a function γ such that:

$$\text{Var}[Z(s_1) - Z(s_2)] = 2\gamma(s_1 - s_2)$$

$2\gamma(s_1 - s_2) = 2\gamma(h)$ is called **variogram function**

Most common uses

$$\frac{1}{2} \text{var}[Z(s_1) - Z(s_2)] = \gamma(h)$$

the **semivariogram function**

Intrinsic stationary is less restrictive.

A second-order stationary random process with covariance C is intrinsically stationary.

$$\begin{aligned} \text{Var}[Z(s_i) - Z(s_j)] &= \text{Var}(Z(s_i)) + \text{Var}(Z(s_j)) - 2\text{Cov}(Z(s_i), Z(s_j)) = \\ &= 2C(\mathbf{0}) - 2C(s_i - s_j) \end{aligned}$$

The variogram is (depends only the distance (h) between their corresponding locations):

$$2\gamma(h) = 2C(\mathbf{0}) - 2C(h)$$

The converse is not true in general, only in case that the random process must have a bounded variogram.

$$\lim_{h \rightarrow \infty} \gamma(h) = M < +\infty$$

- **Isotropy.**

A second order stationary process is isotropic if the covariance between any two values depends only on the Euclidean distance between their corresponding locations:

$$C(Z(s_1), Z(s_2)) = C(||s_1 - s_2||)$$

for all h.

A intrinsic process is isotropic if the variogram depends on the Euclidean distance between their corresponding locations:

$$\text{Var}(Z(s + h), Z(s)) = 2\gamma(||h||)$$

for all h.

The process that don't have this property is called **anisotropic**.

3.EXPLORATORY DATA ANALYSIS (EDA) Objectives:

- Detect spatial dependence
- Detect outliers observations
- Detect large scale variation (mean of the process)
- Study the stationary of the residual process
- Study the isotropy of the residual process
- Help in the proposal of parametric models for both the mean and residual dependency

Descriptive analysis methods. Large-scale variation

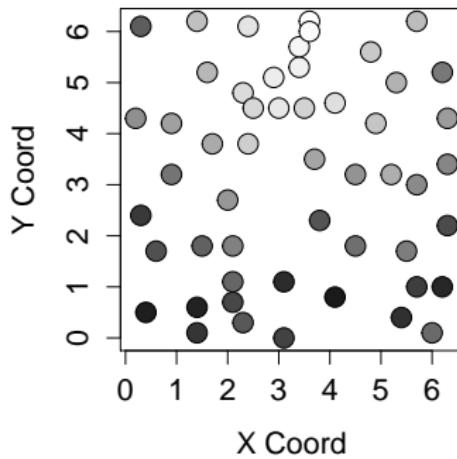
- **Traditional Numerical Summaries:** Mean, median, mode, standard deviation, range, interquartile range (IQR). They reduce the data to a few numbers, not useful for geostatistical data, because the numerical summaries ignore the location.
- **Stem-and-leaf plot:** Gives more information about the distribution of the data (if the data are iid), but location is ignored.

- **Graphs to explore the large-scale structure**

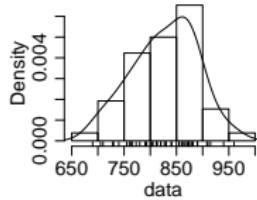
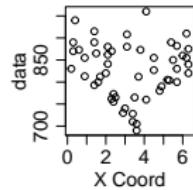
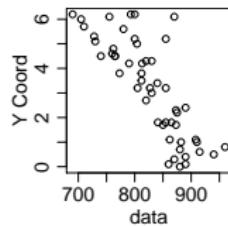
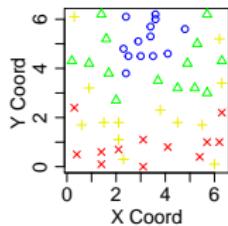
Let Z be the spatial process of interest (eg.ozone concentrations), and Z_i the value of Z at a particular location.

- 3-D Scatterplot: a plot of Z_i versus location.
- Plot of Z_i versus each marginal coordinate (latitude, longitude).
- Plot of mean or median of Z_i versus row index or column index(Also:Box-plot)
- 2-D Scatterplot of data locations with symbols indicating whether Z_i is above or below the median.
- Contour plot of Z_i . This needs some smoothing.

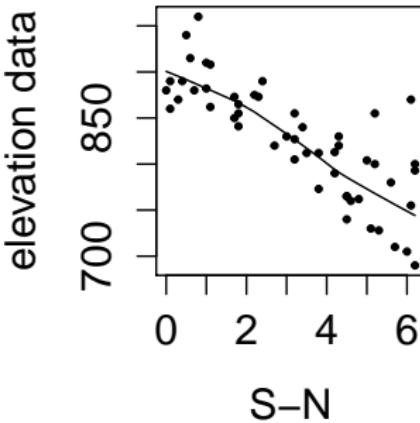
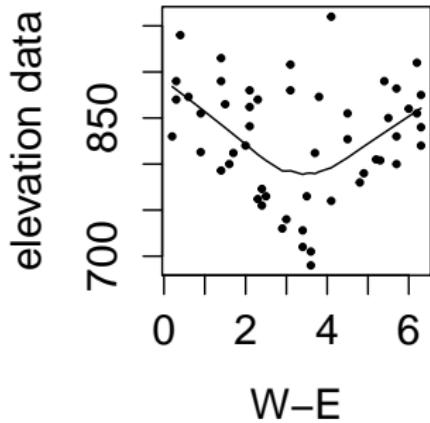
Example: The measured surface elevations Z_i at each 52 locations x_i .



Example: Descriptive analysis: Elevations data



Example: Descriptive analysis: Elevations data



Practice in R

- **Example gridded-data:** The coal seam on the Rodena Mine Property in Greence County, Pennsylvania. The data frame contains 208 coal core samples collected on a grid given by x and y planar coordinates.

Example non-gridded-data: Scallop data were collected during a 1990 survey cruise off the east coast of North America. Scallop counts were obtained using a dredge. Any scallop smaller than 70 mm was termed a prerecruit. Total catch is the sum of prerecruits and recruits.

- 1. Theoretical variogram
- 2. Methods to explore the small-scale dependence
 - 4. Covariogram and Correlogram
- 2. Monte Carlo test for spatial correlation
 - 5. Isotropy
 - 6. Types of anisotropy

3. EXPLORATORY DATA ANALYSIS (II)

Rosa Abellana

Department of Basic Clinical Practice. University of Barcelona

- 1. Theoretical variogram
- 2. Methods to explore the small-scale dependence
 - 4. Covariogram and Correlogram
- 2. Monte Carlo test for spatial correlation
 - 5. Isotropy
- 6. Types of anisotropy

Table of contents

- 1 1. Theoretical variogram
- 2 2. Methods to explore the small-scale dependence
- 3 4. Covariogram and Correlogram
- 4 2. Monte Carlo test for spatial correlation
- 5 5. Isotropy
- 6 6. Types of anisotropy

Definition of variogram

Variogram is an important tool for describing how the spatial data are related with distance. It is defined in terms of dissimilarity in data values between two locations separated by a distance h .

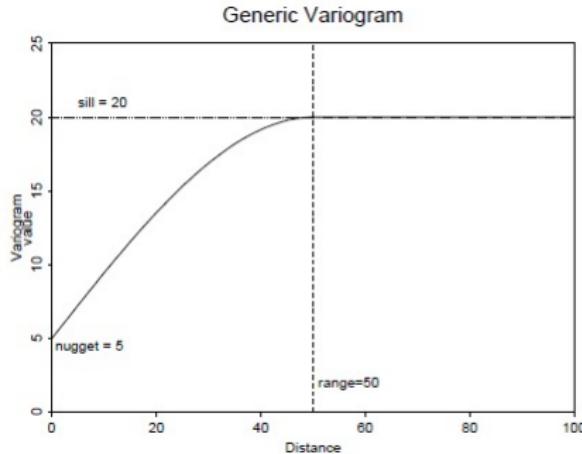
$Z(\cdot)$ intrinsic stationary process. The variogram is defined as:

$$\frac{1}{2}V(Z(s) - Z(t)) = \frac{1}{2}E((Z(s) - Z(t))^2) = \gamma(h),$$

where $h = s - t$, $s, t \in D$.

- 1. Theoretical variogram**
- 2. Methods to explore the small-scale dependence**
 - 4. Covariogram and Correlogram**
 - 2. Monte Carlo test for spatial correlation**
 - 5. Isotropy**
 - 6. Types of anisotropy**

Theoretical variogram



A typical variogram can be described using three parameters:

- Nugget effect, represents micro-scale variation or measurement error. It is estimated from the empirical variogram at $h = 0$.
- Range, is the distance at which the variogram reaches the plateau, i.e., the distance (if any) at which data are no longer correlated.
- Sill is the variance of the random field $\text{Var}(Z)$, disregarding the spatial structure. It is the plateau the variogram reaches at the range.

- The variogram can be estimated using the method of the moments estimators (Matheron, 1971):

$$\gamma(h) = \frac{1}{2N(h)} \sum_{N(h)} (Z(s_i) - Z(s_j))^2$$

where $N(h)$ are the number of data pairs s_i and s_j separated by h .

- Plot $\gamma(h)$ versus different values of h .
- Note that this implicitly assumes stationary of some kind. You can display the variogram along selected directions (e.g., N-S, NW-SE,E-W, and NE-WE) on the same 2-D graph.

- The estimator is biased when the observations themselves are used (Z_1, \dots, Z_n) if there is a non-constant mean (the mean depends on location). It is approximately unbiased when the residuals are used.

- Robust definition of semivariogram. The moment estimator is sensitive to outliers in the data. The robust estimator given by Cressie and Hawkins (1980):

$$\frac{1}{0.914 + 0.988/|N(r_u)|} \left\{ \frac{1}{|N(r_u)|} \sum_{(i,j) \in N(r_u)} |z(\mathbf{s}_i) - z(\mathbf{s}_j)|^{1/2} \right\}^4$$

- 1. Theoretical variogram
- 2. Methods to explore the small-scale dependence
 - 4. Covariogram and Correlogram
 - 2. Monte Carlo test for spatial correlation
 - 5. Isotropy
 - 6. Types of anisotropy

2. Methods to explore the small-scale dependence

- ① Variogram cloud
- ② Sampling variogram

Variogram cloud

- Plot $(Z(s_i) - Z(s_j))^2$ versus $\|s_i - s_j\|^{1/2}$ (Euclidean distance) for all pairs of observations.
- It may be advisable to bin the lags and plot a box-plot for each bin.
- The square-root differences $(Z(s_i) - Z(s_j))^{1/2}$ are more resistant to outliers.
- The variogram cloud implicitly assumes isotropy (does not differentiate any directions)

1. Theoretical variogram
2. Methods to explore the small-scale dependence
 4. Covariogram and Correlogram
 2. Monte Carlo test for spatial correlation
 5. Isotropy
 6. Types of anisotropy

Example Variogram cloud and bins

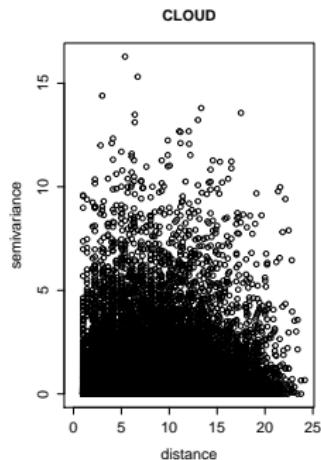


Figure: Variogram cloud

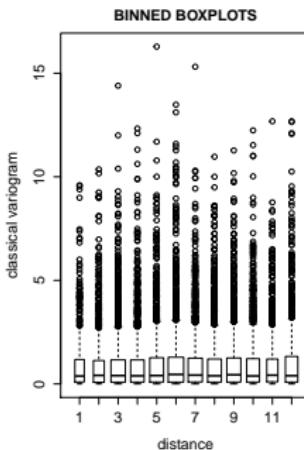


Figure: Box-plot for each bin

1. Theoretical variogram
2. Methods to explore the small-scale dependence
 4. Covariogram and Correlogram
 2. Monte Carlo test for spatial correlation
 5. Isotropy
 6. Types of anisotropy

Example Variogram cloud and bins: Modulus

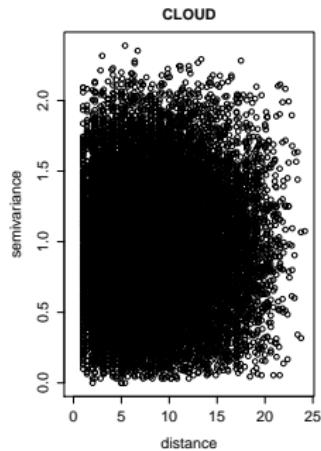


Figure: Variogram cloud

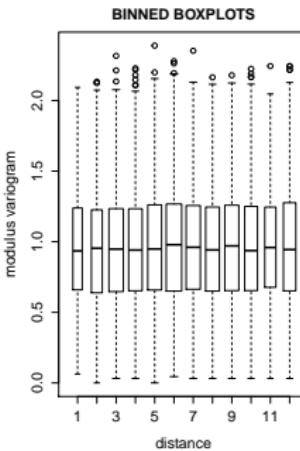
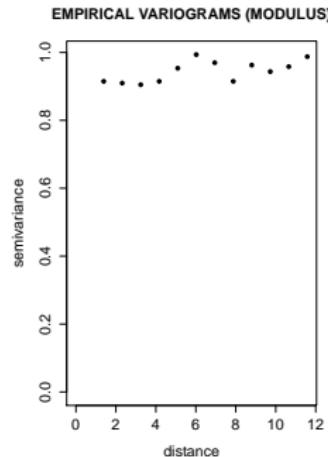
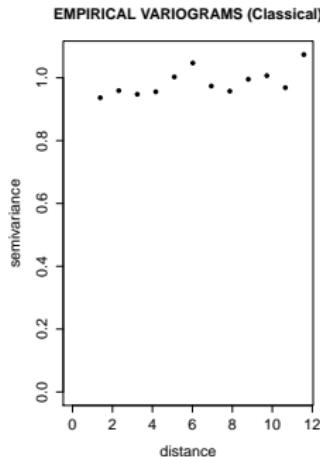


Figure: Box-plot for each bin

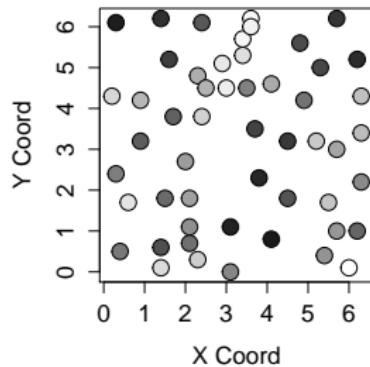
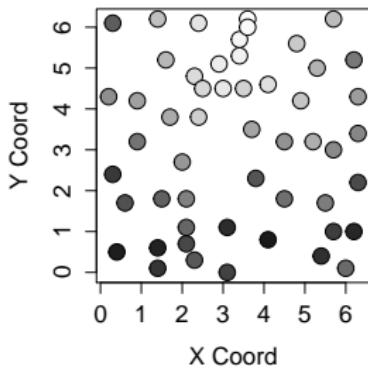
- 1. Theoretical variogram
- 2. Methods to explore the small-scale dependence
 - 4. Covariogram and Correlogram
 - 2. Monte Carlo test for spatial correlation
 - 5. Isotropy
 - 6. Types of anisotropy

Example Empirical variogram



- 1. Theoretical variogram
- 2. Methods to explore the small-scale dependence
 - 4. Covariogram and Correlogram
 - 2. Monte Carlo test for spatial correlation
 - 5. Isotropy
 - 6. Types of anisotropy

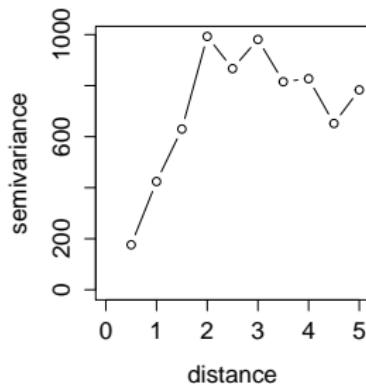
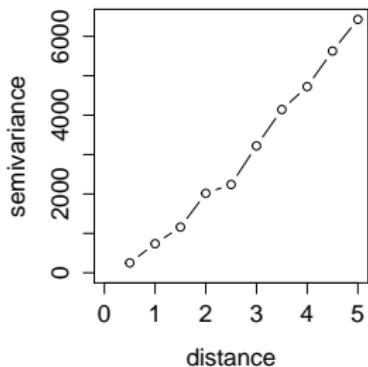
Example: The measured surface elevations Z_i at each 52 locations x_i .



Left-hand panel shows the original data. The right-hand panels show the residuals from a second-order (quadratic) polynomial trend surfaces.

- 1. Theoretical variogram
- 2. Methods to explore the small-scale dependence
 - 4. Covariogram and Correlogram
 - 2. Monte Carlo test for spatial correlation
 - 5. Isotropy
 - 6. Types of anisotropy

Example: Empirical variogram of elevations dataset



Empirical variograms for the original data (left panel) and for residuals from a quadratic trend surface (right panel).

- 1. Theoretical variogram
- 2. Methods to explore the small-scale dependence
 - 4. Covariograma and Correlogram
- 2. Monte Carlo test for spatial correlation
 - 5. Isotropy
 - 6. Types of anisotropy

4. Covariograma and Correlogram

Covariogram (analogous to covariance) and correlogram (analogous to correlation coefficient).

Useful methods for measuring spatial correlation.

They describe similarity in values between two locations.

Covariogram: $C(h) = \text{cov}[Z(s), Z(s + h)]$

Its estimator:

$$\hat{C}(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} [(z(s_i) - \bar{z})(z(s_i + h) - \bar{z})]$$

where \bar{z} is the sample mean.

- 1. Theoretical variogram
- 2. Methods to explore the small-scale dependence
 - 4. Covariogram and Correlogram
 - 2. Monte Carlo test for spatial correlation
 - 5. Isotropy
 - 6. Types of anisotropy

At $h=0$, $\hat{C}(0)$, variance of the random field. The relationship between variogram and covariogram is:

$$\gamma(h) = C(0) - C(h)$$

The correlogram is defined as:

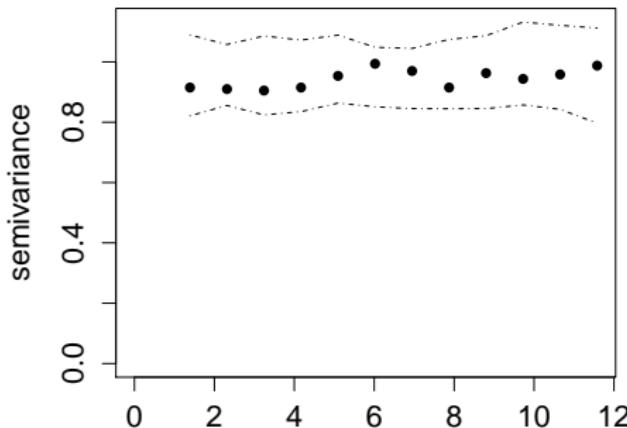
$$\rho(h) = \frac{C(h)}{C(0)} = 1 - \frac{\gamma(h)}{C(0)}$$

- 1. Theoretical variogram
- 2. Methods to explore the small-scale dependence
 - 4. Covariogram and Correlogram
- 2. Monte Carlo test for spatial correlation
 - 5. Isotropy
 - 6. Types of anisotropy

2. Monte Carlo test for spatial correlation

Varigram envelopes computed from random permutations of the original or the residual data can be used to test the hypothesis of constant variogram (no spatial correlation)

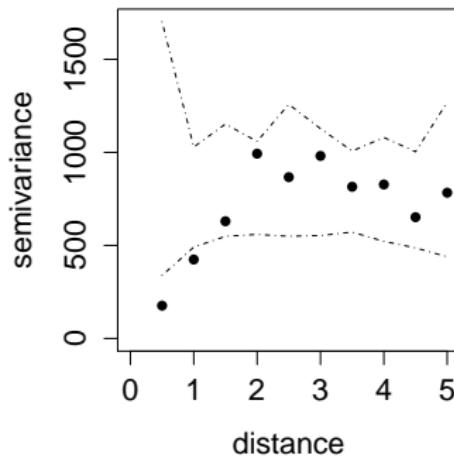
CONFIDENCE BANDS FOR INDEPENDENT |



- 1. Theoretical variogram
- 2. Methods to explore the small-scale dependence
 - 4. Covariogram and Correlogram
- 2. Monte Carlo test for spatial correlation
 - 5. Isotropy
 - 6. Types of anisotropy

Monte Carlo envelope for Elevation Data

Monte carlo envelopes independent model



- 1. Theoretical variogram
- 2. Methods to explore the small-scale dependence
 - 4. Covariogram and Correlogram
- 2. Monte Carlo test for spatial correlation
 - 5. Isotropy
 - 6. Types of anisotropy

PRACTICAL RULES, number of lags and lag increment

- 1. The distance reliability for sample variogram $h < D/2$ where D is the maximum distance over the field of data
- 2. The sample variogram should only be considered for distances h for which the number of pairs is greater than 30.

5. Isotropy

- A second order stationary process is isotropic if the covariance between any two values depends only on the Euclidean distance between their corresponding locations:

$$C(Z(s_1), Z(s_2)) = C(||s_1 - s_2||)$$

for all h.

- An intrinsic process is isotropic if the variogram depends on the Euclidean distance between their corresponding locations:

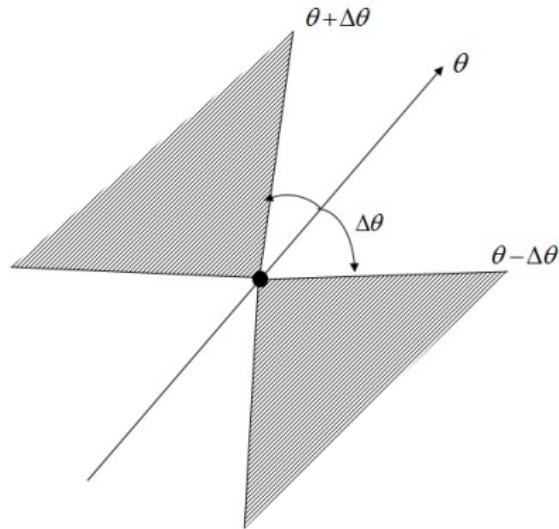
$$\text{Var}(Z(s + h), Z(s)) = 2\gamma(||h||)$$

for all h.

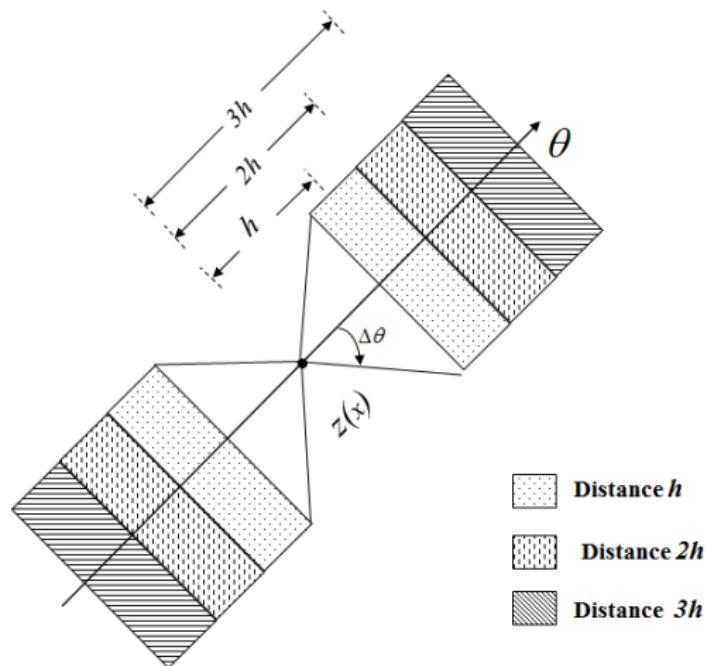
- The process that don't have this property is called **anisotropic**. To study the isotropy, uses the directional variogram.

- 1. Theoretical variogram
- 2. Methods to explore the small-scale dependence
 - 4. Covariogram and Correlogram
- 2. Monte Carlo test for spatial correlation
 - 5. Isotropy
- 6. Types of anisotropy

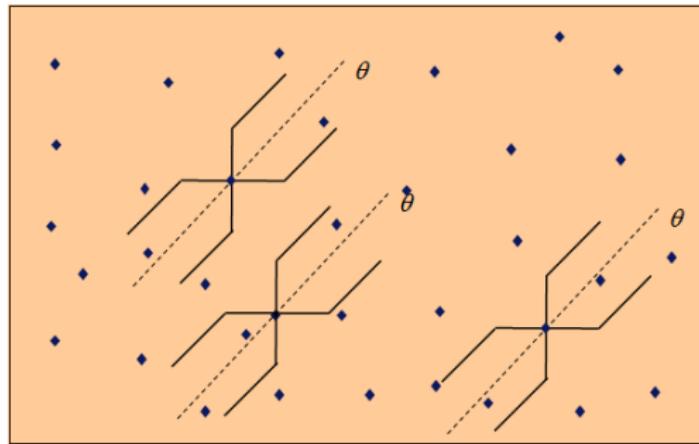
For each direction, θ is defined a tolerance $\Delta\theta$ and it is used only the points that are into directions $\theta - \Delta\theta$ and $\theta + \Delta\theta$



- 1. Theoretical variogram
- 2. Methods to explore the small-scale dependence
 - 4. Covariogram and Correlogram
- 2. Monte Carlo test for spatial correlation
 - 5. Isotropy
- 6. Types of anisotropy

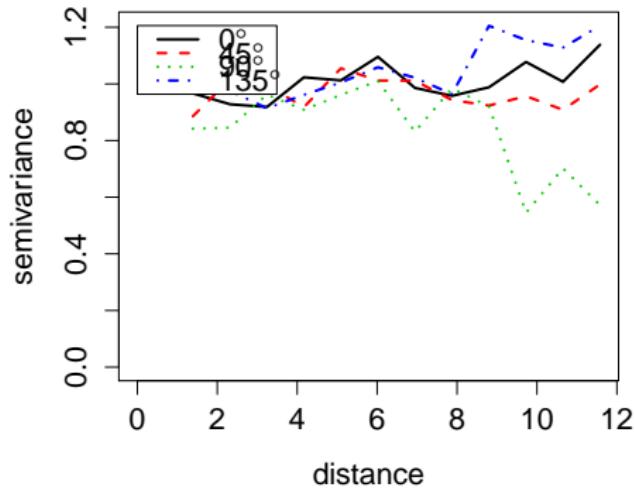


- 1. Theoretical variogram
- 2. Methods to explore the small-scale dependence
 - 4. Covariogram and Correlogram
- 2. Monte Carlo test for spatial correlation
 - 5. Isotropy
- 6. Types of anisotropy



- 1. Theoretical variogram
- 2. Methods to explore the small-scale dependence
 - 4. Covariogram and Correlogram
- 2. Monte Carlo test for spatial correlation
 - 5. Isotropy
- 6. Types of anisotropy

DIRECTIONAL EMPIRICAL VARIOGRAM

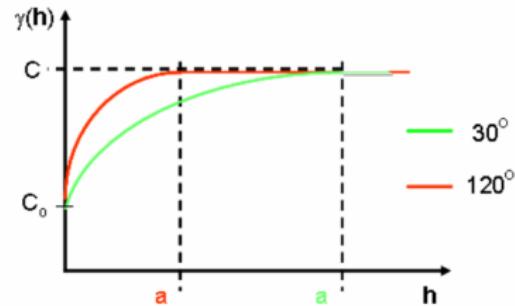
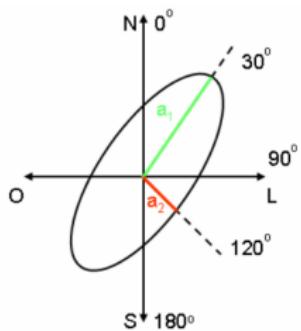


6.Types of anisotropy

- Geometry anisotropy. Directional semivariograms with same model function, same sill and different ranges.
- Zonal anisotropy. Directional semivariograms with same model function, same ranges and different sill.
- Combined anisotropy. Directional semivariograms with same model function, different ranges and sill.

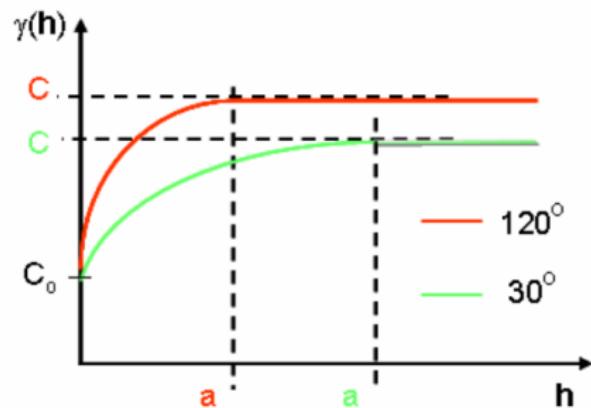
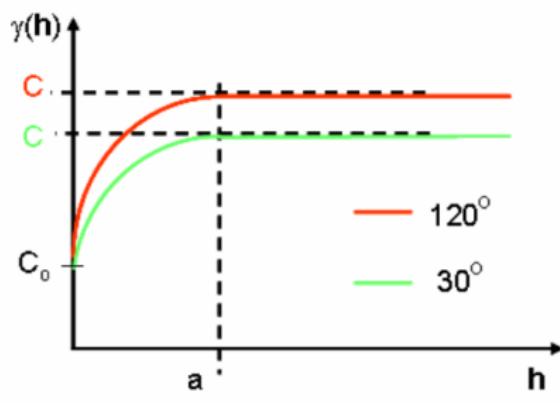
- 1. Theoretical variogram
- 2. Methods to explore the small-scale dependence
 - 4. Covariogram and Correlogram
- 2. Monte Carlo test for spatial correlation
 - 5. Isotropy
- 6. Types of anisotropy

Example: Geometry anisotropy



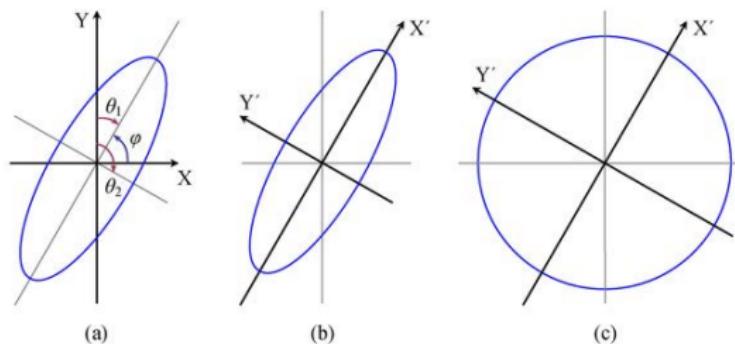
- 1. Theoretical variogram
- 2. Methods to explore the small-scale dependence
 - 4. Covariogram and Correlogram
- 2. Monte Carlo test for spatial correlation
 - 5. Isotropy
- 6. Types of anisotropy

Example: Zonal anisotropy (left panel) and combined anisotropy (right panel)



- 1. Theoretical variogram
- 2. Methods to explore the small-scale dependence
- 4. Covariogram and Correlogram
- 2. Monte Carlo test for spatial correlation
- 5. Isotropy
- 6. Types of anisotropy

Transformation to solve geometry anisotropy



Transformation of the coordinates into (x', y') in terms of φ

$$\begin{bmatrix} \cos(\varphi) & \sin(\varphi) \\ -\sin(\varphi) & \cos(\varphi) \end{bmatrix}$$

Lengthen the minor axis

$$\begin{bmatrix} 1 & 0 \\ 0 & 1/R \end{bmatrix}$$

$R =$ shorter /longer range

4.VARIOGRAM

Rosa Abellana

Departament Fonaments Clínics. Facultat de Medicina. Universitat de Barcelona
rabellana@ub.edu
<https://orcid.org/0000-0003-4487-4431>

September 27, 2017

Table of contents

1 1.Parameters of a variogram

2 2.Theoretical Variogram Models

3 3.Methods of fitting

1. Parameters of a variogram

The main goal of a variogram analysis is to construct a variogram that best estimates the autocorrelation structure of the underlying stochastic process. Most variogram are defined through a several parameters;

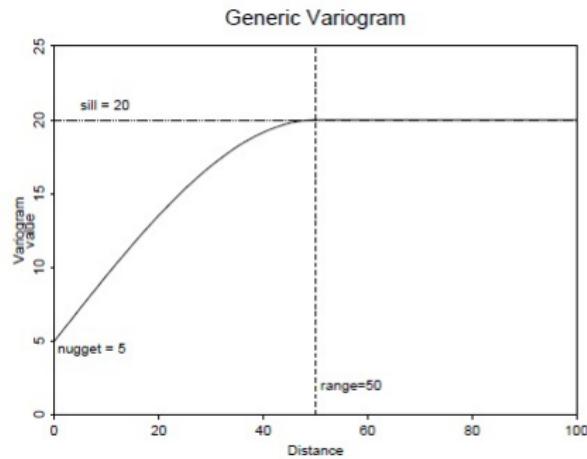
- ① *Nugget effect.* Presents micro-scale variation or measurement error. It is estimated from the empirical variogram as the value of $\lim_{h \rightarrow 0^+} \gamma(h) = \tau^2$.
- ② *Sill.* The $\lim_{h \rightarrow \infty} \gamma(h) = \tau^2 + \sigma^2$ representing the variance of the random field. And the sill minus the nugget, which is simply σ^2 is called the *partial sill*.
- ③ *Range.* The distance at which data are no longer autocorrelated.

If for some distance, \mathbf{h} , the variables $Z(x)$ and $Z(x+h)$ are uncorrelated then the variogram is constant:

$$\gamma(h) = \frac{1}{2}E[(Z(x) - Z(x + h))^2] = \sigma^2 - E[Z(x)Z(x + h)] = \sigma^2$$

Range: Distance from which there is no correlation

Sill: Variance of the variable Z



- A second order process:

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h}),$$

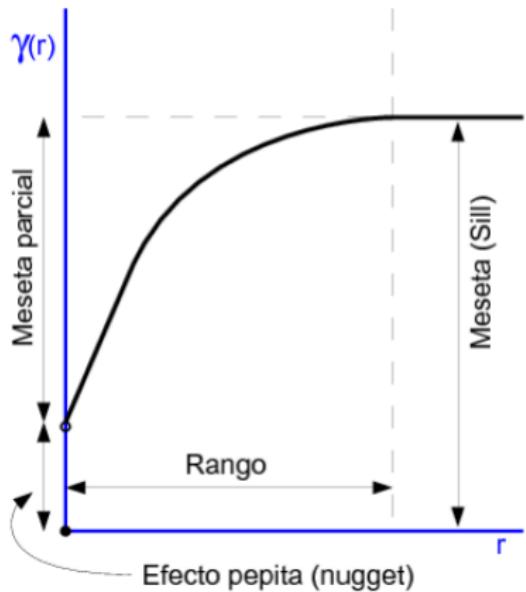
where $C(\mathbf{h}) = \text{Cov}(\mathbf{Z}(\mathbf{s}), \mathbf{Z}(\mathbf{t}))$, if $\mathbf{h} = \mathbf{s} - \mathbf{t}$.

- Function $C(\mathbf{h})$ is called autocovariance or covariogram.
- If a intrinsic stationary process verifies that:
 $\lim_{\|\mathbf{h}\| \rightarrow \infty} \gamma(\mathbf{h}) = M < \infty$ then the process is second-order stationary and

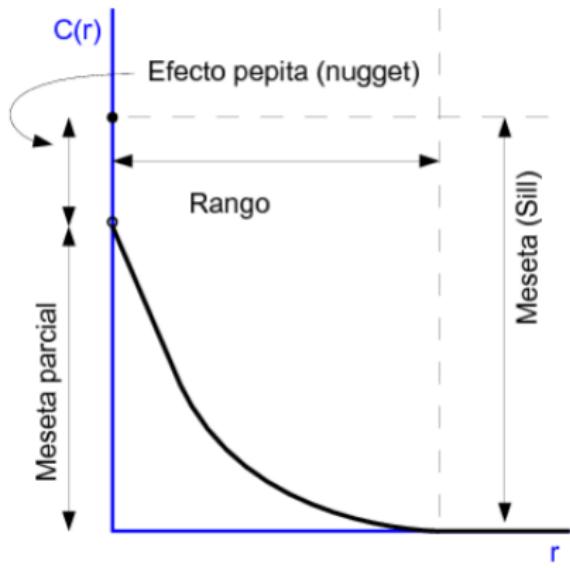
$$C(\mathbf{h}) = M - \gamma(\mathbf{h}).$$

- Usually it is assume stationary (intrinsic) and isotropic and the variogram function it was fitted by a parametric function

Variograma



Covariograma



2.Theoretical Variogram Models

1.Pure Nugget effect. The pure nugget effect thus corresponds to the total absence of auto-correlation.

$$y(h) = \begin{cases} 0 & \text{si } |h| = 0 \\ s & \text{si } |h| \neq 0 \end{cases}$$

Figure: Pure Nugget effect



2. Spherical model (without nugget effect)

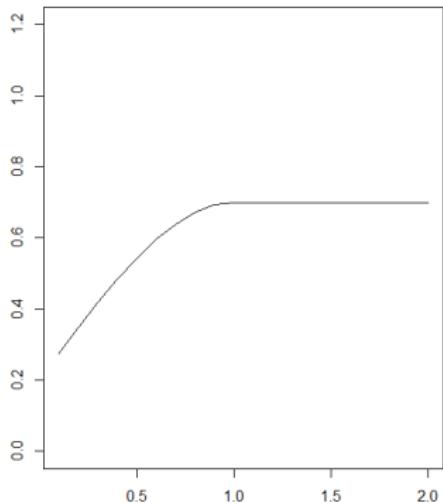
$$y(h) = \begin{cases} \sigma^2 \left(\frac{3h}{2\phi} - \frac{1}{2} \left(\frac{h}{\phi} \right)^3 \right) & \text{si } 0 \leq h \leq \phi \\ \sigma^2 & \text{si } h > \phi \end{cases}$$

Nugget=0 Sill= σ^2 Range= ϕ

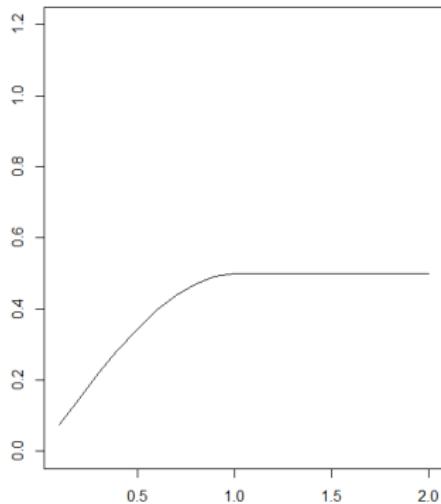
Spherical model (with nugget effect)

$$y(h) = \begin{cases} \tau^2 + \sigma^2 \left(\frac{3h\phi}{2\phi} - \frac{1}{2} \left(\frac{h}{\phi} \right)^3 \right) & \text{si } 0 \leq h \leq \phi \\ \tau^2 + \sigma^2 & \text{si } h > \phi \end{cases}$$

Nugget= τ^2 Sill= $\tau^2 + \sigma^2$ Range= ϕ



Spherical; tau2=0.2,sigma2=0.5,phi=1



Spherical; tau2=0,sigma2=0.5,phi=1

3. Exponential Model:

$$\gamma(h; \sigma, \phi) = \sigma^2(1 - e^{-h/\phi}) \quad h > 0$$

Nugget: 0. Sill : σ^2 (Asymptotically). Effective range: $3 \times \phi$ (distance to reach the 95% of the Sill).

Exponential Model. With nugget effect

$$\gamma(h; \sigma, \phi) = \tau^2 + \sigma^2(1 - e^{-h/\phi}) \quad h > 0$$

Nugget: τ^2 . Sill : $\tau^2 + \sigma^2$. Effective range: $3 \times \phi$ (distance to reach the 95% of the Sill).

4. Gaussian model

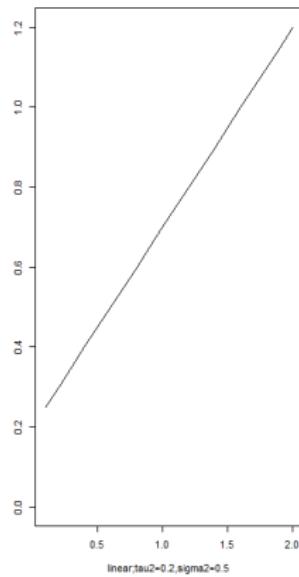
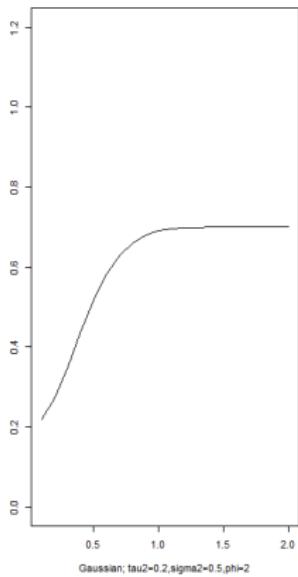
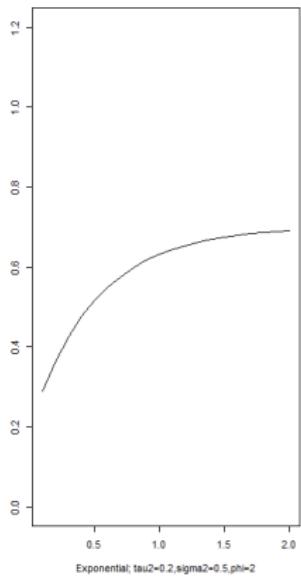
$$\gamma(h; \sigma, \phi) = \tau^2 + \sigma^2(1 - e^{-h^2/\phi^2}) \quad h > 0$$

Nugget: τ^2 . Sill : $\tau^2 + \sigma^2$. Range: $\sqrt{3} \times \phi$ (distance to reach the 95% of the Sill).

5. Linear model

$$\gamma(h; \sigma, \phi) = \tau^2 + \sigma^2 h \quad h > 0$$

Nugget: τ^2 . Sill : ∞ . Range: Don't have



6. Rational quadratic

$$\gamma(h; \sigma, \phi) = \tau^2 + \frac{\sigma^2 h^2 \text{ffl}}{(1 + h^2/\phi)} \quad h > 0$$

7. Wave

$$\gamma(h; \sigma, \phi) = \tau^2 + \sigma^2 \left(1 - \frac{\sin(h/\phi)}{h/\phi}\right) \quad h > 0$$

8.Power Law

$$\gamma(h; \sigma, \phi) = \tau^2 + \sigma^2 h^\lambda \quad h > 0$$

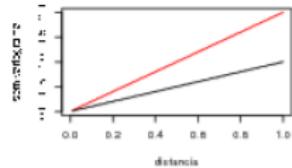
9.Matérn

$$\gamma(h; \sigma, \phi) = \tau^2 + \sigma^2 \left[1 - \frac{1}{2^{\nu-1} \Gamma(\nu)} \left(\frac{2\sqrt{\nu}h}{\phi}\right)^\nu K_\nu \left(\frac{2\sqrt{\nu}h}{\phi}\right)\right] \quad h > 0$$

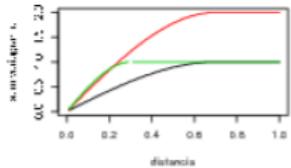
Special cases $\nu = 1/2$ exponential model and $\nu \rightarrow \infty$ Gaussian model

1.Parameters of a variogram
2.Theoretical Variogram Models
3.Methods of fitting

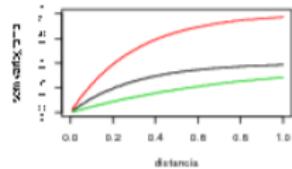
modelo lineal



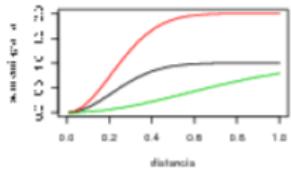
modelo esférico



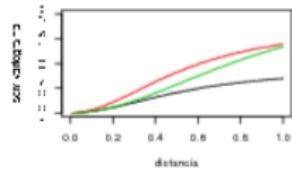
modelo exponencial



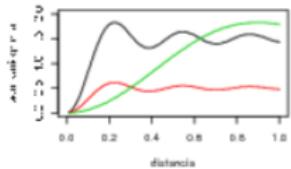
modelo gausiano



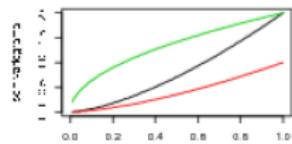
modelo racional cuadrático



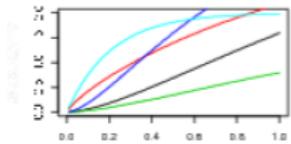
modelo ondulado



modelo potencial



modelo de Matérn



Spherical	Sill	ϕ	Range
Black	1	0.7	0.7
Red	2	0.7	0.7
Green	1	0.3	0.3
Exponential	Sill	ϕ	Range(95%)
Black	1	0.3	0.9
Red	2	0.3	0.9
Green	1	0.8	2.4
Gaussian	Sill	ϕ	Range
Black	1	0.3	0.51
Red	2	0.3	0.51
Green	1	0.8	1.38
Rational Q.	Sill	ϕ	Range
Black	3	0.3	-
Red	6	0.3	-
Green	3	0.8	-

Wave	Sill	ϕ	Range
Black	1.5	0.05	
Red	0.5	0.05	
Green	1.5	0.05	
Potential	σ^2	λ	
Black	2	1.5	
Red	1	1.5	
Green	2	0.5	
Mattern	Sill	ν	ϕ
Black	4	1	2
Red	4	0.4	2
Green	2	1	2
Dark Blue	4	1	1
Light Blue	2	0.5	0.3

3.Methods of fitting

- ① **Ordinary non-linear least squares** (not appropriate: because the lag classes have different variability and they are correlated). Minimizes:

$$\sum_{u=1}^k (\hat{\gamma}(r_u) - \gamma(r_u))^2$$

- ② **Weighted non-linear least squares** A weighted non-linear least squares estimator of $\gamma(h; \theta)$ is defined as a value θ that minimizes the weighted residual sum of squares function:

$$\sum_{u=1}^k \frac{|N(r_u)|}{\gamma(r_u; \theta)^2} (\hat{\gamma}(r_u) - \gamma(r_u))^2$$

- ③ **Generalized non-linear least squares**

$$GRSS(\theta) = (\hat{\gamma} - \gamma(\theta))' \hat{Var}(\hat{\gamma}) (\hat{\gamma} - \gamma(\theta))$$

4 Maximum likelihood (and restricted maximum likelihood) Applicable to processes with second-order stationary errors only.

- Estimates β and θ simultaneously.
- Let $V = V(\theta)$ denote the covariance matrix of $Z = (Z_1, \dots, Z_n)$ and let X denote the model matrix for the model:

$$Z = X\beta + \epsilon$$

The log-likelihood functions is

$$\log L(\beta, \theta; Z) = -\frac{1}{2} \log |V(\theta)| - \frac{1}{2} (Z - X\beta)' V(\theta)^{-1} (Z - X\beta).$$

- a MLE is a value $(\hat{\beta}, \hat{\theta})$ that maximizes $\log L(\beta, \theta)$.
- Generally a MLE must be found by numerical optimization routines (e.g. Newton-Raphson). Thus we have to be concerned with starting values, and convergence criteria.

- A restricted MLE (REML) is defined as a value θ that maximizes the log-likelihood function associated with $n - \text{rank}(X)$ linearly independent error contrasts. It is less biased than MLE

Model Selection Procedures

- Visual inspection of semivariogram plot
- Minimized weighted (or generalized) residual sum of squares function
- Maximized log-likelihood (or restricted log-likelihood) function, $\log L(\hat{\beta}, \hat{\theta})$.
- Penalized likelihood criteria, e.g. Akaike's Information Criterion

$$AIC = -2\log L(\hat{\beta}, \hat{\theta}) + \text{number of estimated parameters}$$

5.Kriging

Rosa Abellana

Departament Fonaments Clínics. Facultat de Medicina. Universitat de Barcelona
rabellana@ub.edu
<https://orcid.org/0000-0003-4487-4431>

Table of contents

- 1 1.Kriging
- 2 2.Model
- 3 3.Kriging Types
- 4 4.Properties of kriging predictor

Background

- The origin of the word kriging is from D.G. Krige, a South African mining engineer who in the 1950's developed empirical methods for predicting ore grades at unsampled locations using the known grades of ore sampled at nearby sites.
- Krige's original method is what is now called ordinary kriging. There have been several modifications and extensions (e.g., universal kriging, indicator kriging, disjunctive kriging, and others) but they are all based on very similar ideas.

Model

Spatial process: $Z(s_i); s \in D \subset R^d$

Observed values $Z(s_1), \dots Z(s_n)$ n locations inside the region of study.

We consider the geostatistic process:

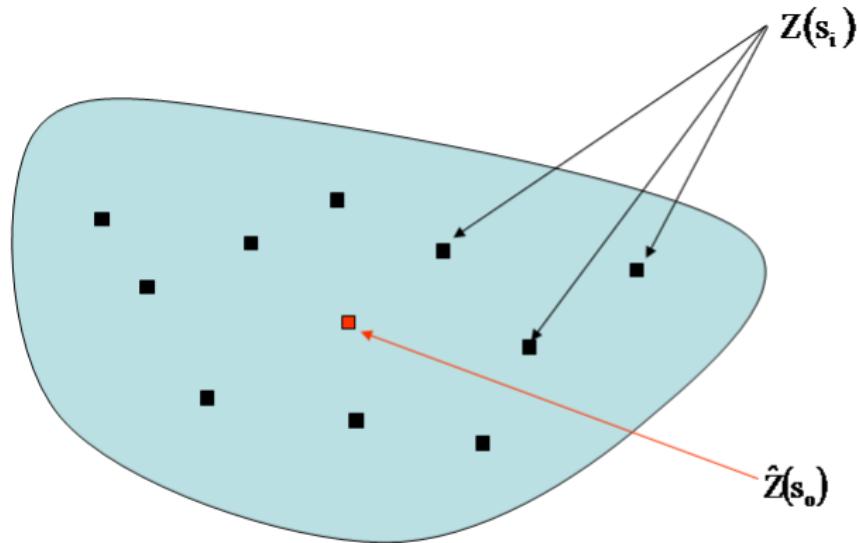
$$Z(s) = m(s) + \varepsilon(s)$$

where: •

- $m(s)$ is the mean of the process or large-scale variation.
- $\varepsilon(s) = Z(s) - m(s)$ spatial error or small-scale variation

The $\varepsilon(s)$ are intrinsic stationarity and you have the variogram $\gamma(h)$
The goal of the **kriging**: to predict the value of $Z(s_0)$ at s_0 .

The basic idea is to estimate the attribute value, $\hat{Z}(s_0)$, at a location where we do not know the true value



Kriging Types

- **Simple kriging.** Assumes there is no trend, $m(s)$ is known and constant and in the majority of cases this is an unrealistic assumption.
- **Ordinary kriging.** Assumes there is no trend, $m(s)$ is unknown and constant and focuses on the spatially correlated component.
- **Universal kriging.** Assumes z values change because of a drift (trend) in addition to autocorrelation. The $m(s)$ is not constant. Trend component expressed as a 1st order (plane) or 2nd order polynomials (quadratic surface). Kriging is performed on residual after the trend is removed.

- **External Drift Kriging.** Particular case of Universal kriging, when an explicative variable different than latitude and longitude it is used.
- **Indicator kriging.** Z is a binary variable.
- **Cokriging.** Adds second variable which is correlated with the primary function.
 - It assumes the correlation between the variables can be used to improve the prediction of the primary variable

Properties of kriging predictor:

- ① It is a *linear* combination of the data values,

$$\hat{Z}(s_0) = \sum_{i=1}^n \lambda_i Z(s_i)$$

- ② It is *unbiased*, it satisfies

$$E[\hat{Z}(s_0)] = E[Z(s_0)]$$

- ③ Among all function of the data that satisfy the first 2 properties, it is the *best* in the sense that minimizes the variance of prediction error, $\text{var}[\hat{Z}(s_0) - Z(s_0)]$

Kriging is also known as the **Best Linear Unbiased Predictor (BLUP)**.

Simple kriging

Assumes that:

$$Z(s) = m + \epsilon(s) \implies \epsilon(s) = Z(s) - m$$

m is constant and known. Then the predictor of the random error is defined as:

$$\hat{\epsilon}(s_0) = \sum_{i=1}^n \lambda_i \epsilon(s_i)$$

Then the predictor of the $Z(s_0)$ is:

$$\hat{Z}(s_0) = m + \sum_{i=1}^n \lambda_i \epsilon(s_i)$$

The unbiased predictor:

$$E[\hat{Z}(s_0)] - E[Z(s_0)] = m + \sum_{i=1}^n \lambda_i E[\epsilon(s_i)] - m = 0$$

Estimation of λ_i that minimizes the variance of prediction error

$$\begin{aligned} \text{Var}(\hat{\epsilon}(s_0) - \epsilon(s_0)) &= E[(\hat{\epsilon}(s_0) - \epsilon(s_0))^2] = E\left[\left(\sum_{i=1}^n \lambda_i \epsilon(s_i) - \epsilon(s_0)\right)^2\right] = \\ &E\left[\left(\sum_{i=1}^n \lambda_i \epsilon(s_i)\right)^2\right] - 2E\left[\sum_{i=1}^n \lambda_i \epsilon(s_i) \epsilon(s_0)\right] + E[\epsilon(s_0)]^2 = \\ &\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j E[\epsilon(s_i) \epsilon(s_j)] - 2\lambda_i \sum_{i=1}^n E[\epsilon(s_i) \epsilon(s_0)] + \sigma^2 = \\ &\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(s_i - s_j) - 2 \sum_{i=1}^n \lambda_i C(s_i - s_0) + \sigma^2 \end{aligned}$$

$$\frac{\partial \text{Var}(\bullet)}{\partial \lambda_1} = 2 \sum_{j=1}^n \lambda_j C(s_1 - s_j) - 2C(s_1 - s_0) = 0$$

$$\frac{\partial \text{Var}(\bullet)}{\partial \lambda_2} = 2 \sum_{j=1}^n \lambda_j C(s_2 - s_j) - 2C(s_2 - s_0) = 0$$

In general:

$$\sum_{j=1}^n \lambda_j C(s_i - s_j) = C(s_i - s_0)$$

Matrix form:

$$\begin{pmatrix} C_{11} & C_{12} & \dots & C_{1n} \\ C_{21} & C_{22} & \dots & C_{2n} \\ \vdots & & & \vdots \\ C_{n1} & C_{n2} & \dots & C_{nn} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix} = \begin{pmatrix} C(s_1 - s_0) \\ C(s_2 - s_0) \\ \vdots \\ C(s_n - s_0) \end{pmatrix}$$

Variance of Simple Kriging prediction. The minimized variance, called the kriging variance, is

$$\begin{aligned}\sigma_{SK}(s_0) &= \text{Var}(\hat{\epsilon}(s_0) - \epsilon(s_0)) = \\&= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(s_i - s_j) - 2 \sum_{i=1}^n \lambda_i C(s_i - s_0) + \sigma^2 = \\&= \sum_{i=1}^n \lambda_i C(s_i - s_0) - 2 \sum_{i=1}^n \lambda_i C(s_i - s_0) + \sigma^2 = \\&= \sigma^2 - \sum_{i=1}^n \lambda_i C(s_i - s_0)\end{aligned}$$

Prediction. Effect of mean, m .

$$\begin{aligned}\hat{Z}(s) &= m + \sum_{i=1}^n \lambda_i \epsilon(s_i) = m + \sum_{i=1}^n \lambda_i (Z(s_i) - m) = \\ &= m + \sum_{i=1}^n \lambda_i Z(s_i) - m \sum_{i=1}^n \lambda_i = \\ &= m(1 - \sum_{i=1}^n \lambda_i) + \sum_{i=1}^n \lambda_i Z(s_i)\end{aligned}$$

What happened if the observed values $Z(s_i)$ are not spatial correlated?

$$\begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix} = \begin{pmatrix} C(s_1 - s_0) \\ C(s_2 - s_0) \\ \vdots \\ C(s_n - s_0) \end{pmatrix}$$

$$\sigma^2 \lambda_i = C(s_i - s_0)$$

$$\lambda_i = \frac{C(s_i - s_0)}{\sigma^2} = \rho(s_i - s_0)$$

What happened if the observed values $Z(s_i)$ are not spatial correlated with the point to predicted?

$$\begin{pmatrix} C_{11} & C_{12} & \dots & C_{1n} \\ C_{21} & C_{22} & \dots & C_{2n} \\ \vdots & & & \vdots \\ C_{n1} & C_{n2} & \dots & C_{nn} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Solution of the system $\lambda_i = 0$ The prediction:

$$\hat{Z}(s_0) = m + \sum_{i=1}^n \lambda_i (Z(s_i) - m) = m$$

Prediction intervals

A $100(1 - \alpha)\%$ prediction interval for $Z(s_0)$ is as follows:

$$\hat{Z}(s_0) \pm z_{\alpha/2} \sigma_{OK}(s_0),$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ percentage point of the standard normal distribution.

Ordinary Kriging The theory of ordinary kriging is based on the same geostatistical model we have been using all along, with two important restrictions:

- The mean $m(s)$ is assumed to be constant (unknown).
- The semivariogram $\gamma(h)$ is assumed to be known.

$$Z(s) = m(s) + \varepsilon(s)$$

$$Z(s) = m + \varepsilon(s)$$

$$\hat{Z}(s_0) = \sum_{i=1}^n \lambda_i Z(s_i)$$

It is unbiased, if it satisfies that:

$$E[Z(s_0)] - E[\hat{Z}(s_0)] = 0$$

$$\begin{aligned} E[Z(s_0)] - E[\hat{Z}(s_0)] &= E[Z(s_0)] - E\left[\sum_{i=1}^n \lambda_i Z(s_i)\right] = \\ &= m - \sum_{i=1}^n \lambda_i m = m \left(1 - \sum_{i=1}^n \lambda_i\right) = 0 \end{aligned}$$

To satisfy this condition

$$\sum_{i=1}^n \lambda_i = 1$$

The properties we have imposed on our predictor lead us to minimize

$$\text{var}[\hat{Z}(s_0) - Z(s_0)]$$

subject to the restriction

$$\sum_{i=1}^n \lambda_i = 1$$

(Use The method of Lagrange multipliers)

Minimize:

$$\text{Var}(\hat{Z}(s_0) - Z(s_0)) = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(s_i - s_j) - 2 \sum_{i=1}^n \lambda_i C(s_i - s_0) + \sigma^2$$

subject to the restriction:

$$\sum_{i=1}^n \lambda_i = 1 \implies m \left(\sum_{i=1}^n \lambda_i - 1 \right)$$

Solution obtained from the equations:

$$\sum_{j=1}^n \lambda_j C(s_i - s_j) - m = C(s_i - s_0)$$

Noticed that $C(s_i - s_j) = C(0) - \gamma(s_i - s_j)$

Also it's possible to estimated λ_i from the variogram function.

Variance of Ordinary Kriging prediction

$$\sigma_{OK}(s_0) = \sigma^2 - \sum_{i=1}^n \lambda_i C(s_i - s_0) + m$$

How can we choose the most appropriate method for interpolation? Leave-one-out Cross validation:

- ① Remove a known point from data set
- ② Use remaining points to estimate the value at the point removed
- ③ Compare the estimated to known value
- ④ Repeat for all points

Calculated:

$$VC1 = \frac{1}{n} \sum_{i=1}^n \frac{\hat{Z}_{-1}(s_i) - Z(s_i)}{\hat{\sigma}_{-1}(s_i)}$$

$$VC2 = \frac{1}{n} \sum_{i=1}^n \left(\frac{\hat{Z}_{-1}(s_i) - Z(s_i)}{\hat{\sigma}_{-1}(s_i)} \right)^2$$

$$VC3 = \frac{1}{n} \sum_{i=1}^n \left(\hat{Z}_{-1}(s_i) - Z(s_i) \right)^2$$

Goodness of fit

$VC1 \approx 0$, $VC2 \approx 1$, $VC3$ low values

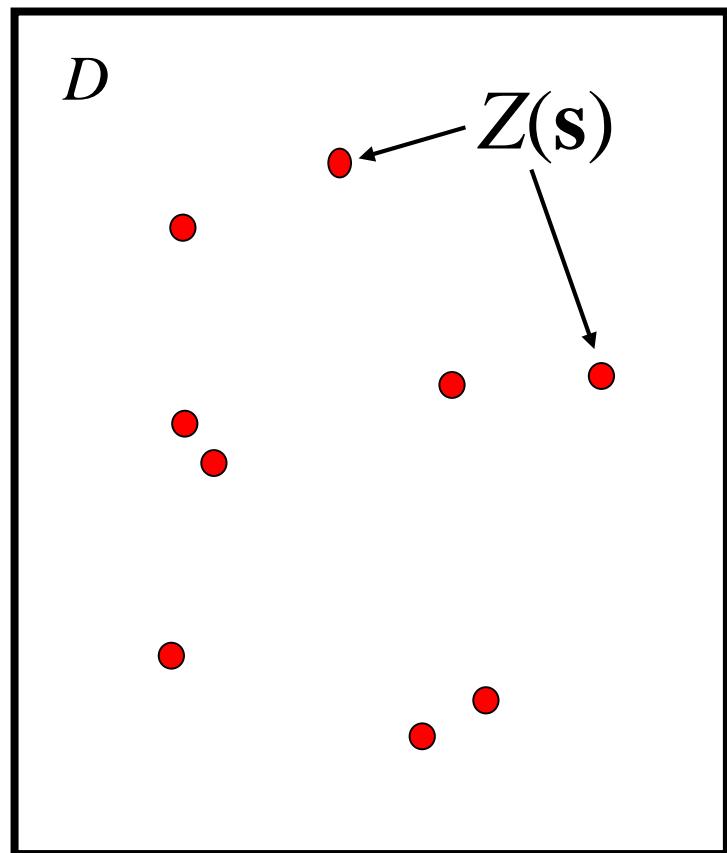
Lattice Data

- The neighborhood matrix
- Exploratory data analysis

Dra. ROSA ABELLANA

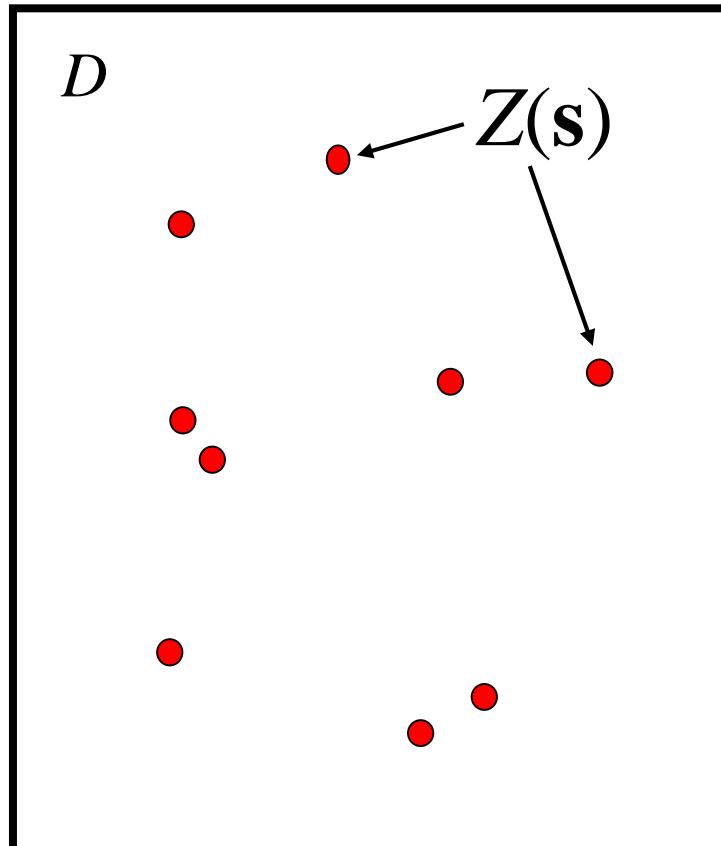
DPT. Fonaments Clínics(UB)

Notation



- D : Spatial domain or interest area
- \mathbf{s} : spatial coordinates
- Z : *study variable*

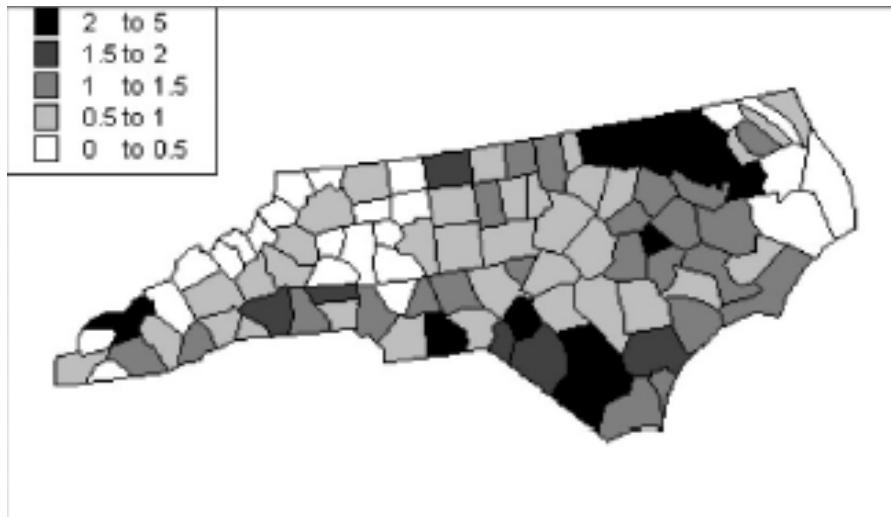
Type of spatial data



$$\{ Z(s) : s \in D \}$$

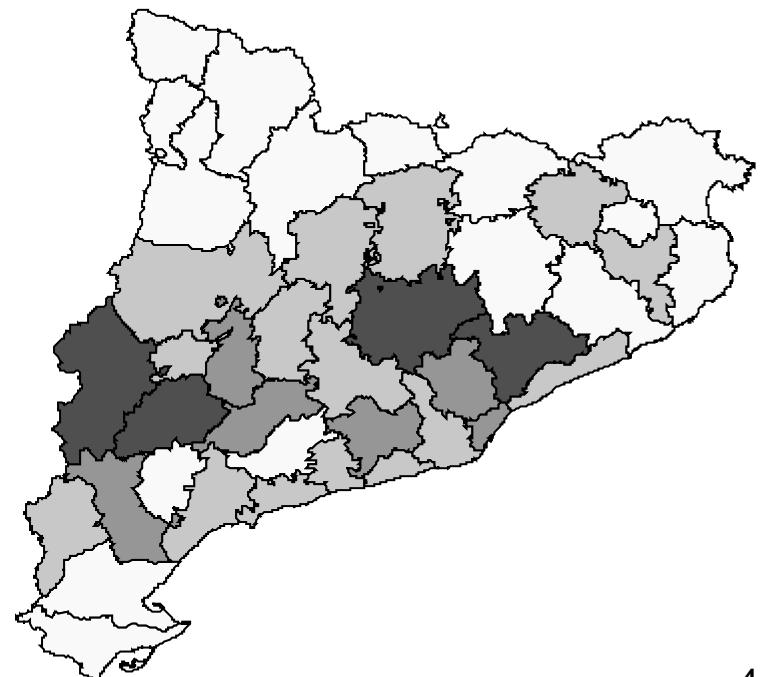
- **Geo statistics:** Z random; D fixed, infinity,
- **Lattice Models:** Z random; D fixed, finite, (ir)regular grilled
- **Point Patterns:** D random, finite
 - **Spatial point pattern:** $Z=1$
 - **Marked point pattern:** Z random

Examples Lattice Data



North Carolina: Sudden
infant death standardized
mortality

Catalonia: SMR
Diabetes type I



Examples (II)

Lattice data

- Number of patients by region in a county.
- Number of fruits for each tree in a field
- Number of accidents for road section.
- Number of fishes by river section.

Neighborhood

Modelling of data at an areal level, it is necessary to define

Adjacency matrix

that characterizes the neighbourhood structure of the data being analysed

Adjacency matrix

1. CHOOSE A NEIGHBORHOOD CRITERION

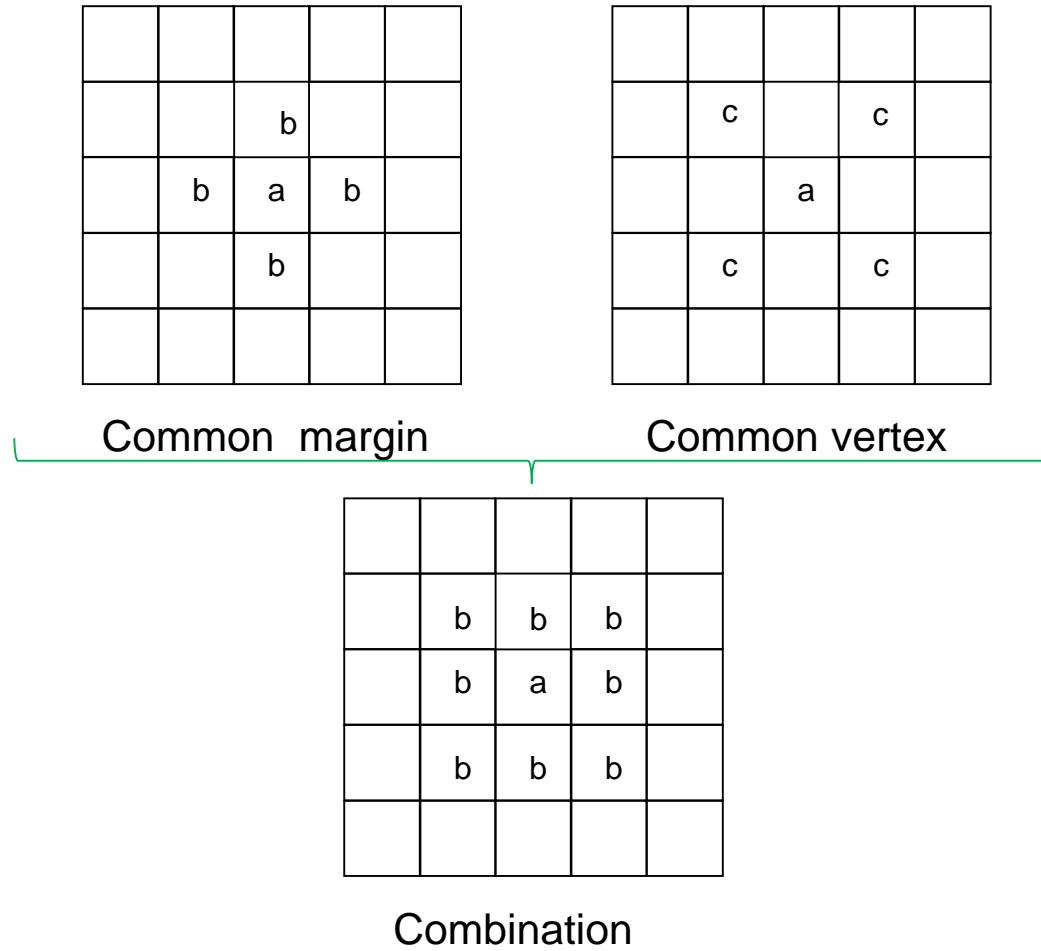
- a. Contiguity based neighbors
- b. Distance based neighbors k nearest neighbors
- c. Distance based neighbors

2. DEFINE SPATIAL WEIGHTS MATRICES

- a. Binary weights (0/1)
- b. Row-standardized weights
- c. According the distances between regions
other criteria

a. Contiguity based neighbors

Regular grid

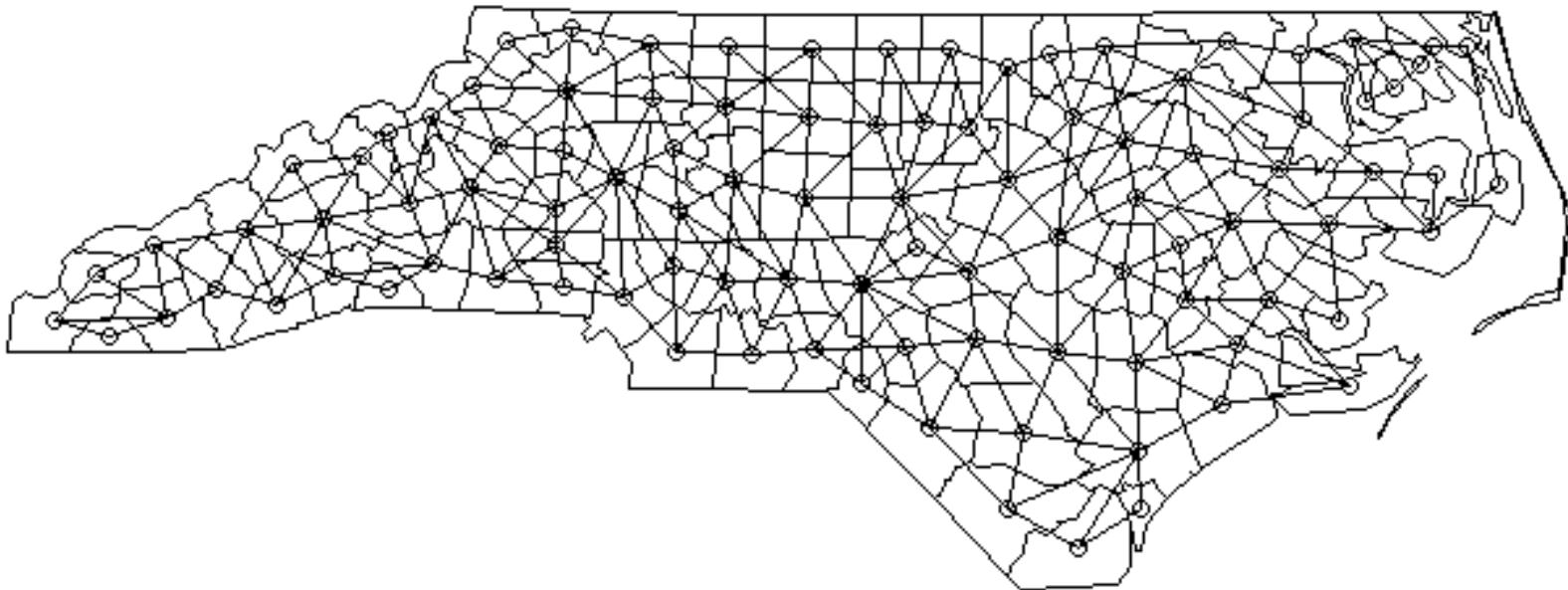


a. Contiguity based neighbors

Irregular grid

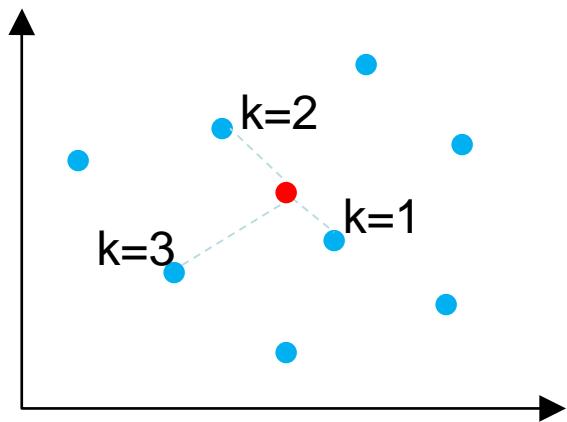


Queen =TRUE

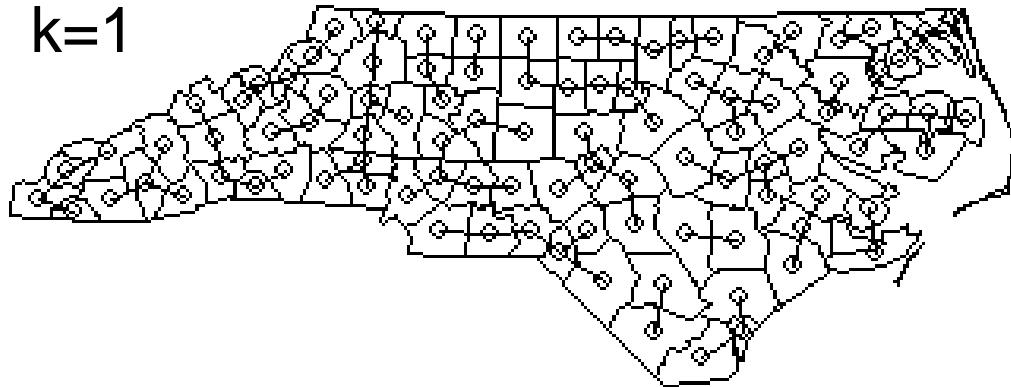


A single shared boundary point meets the contiguity condition

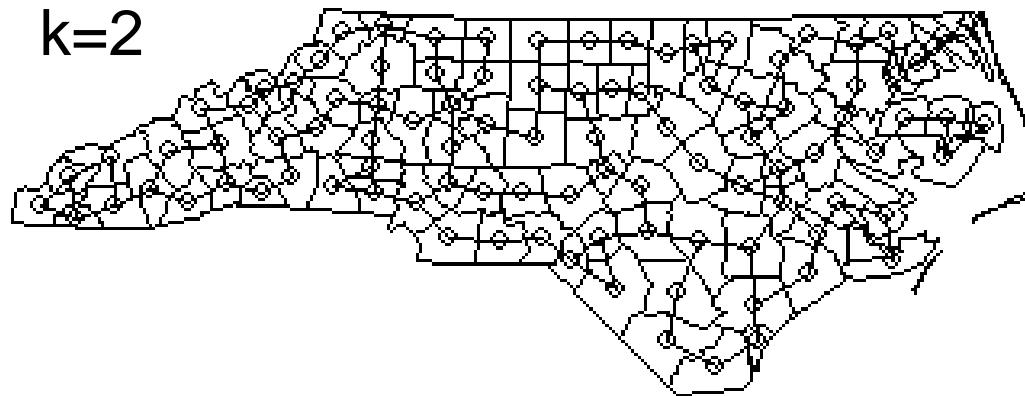
b. Distance based neighbors k nearest neighbors



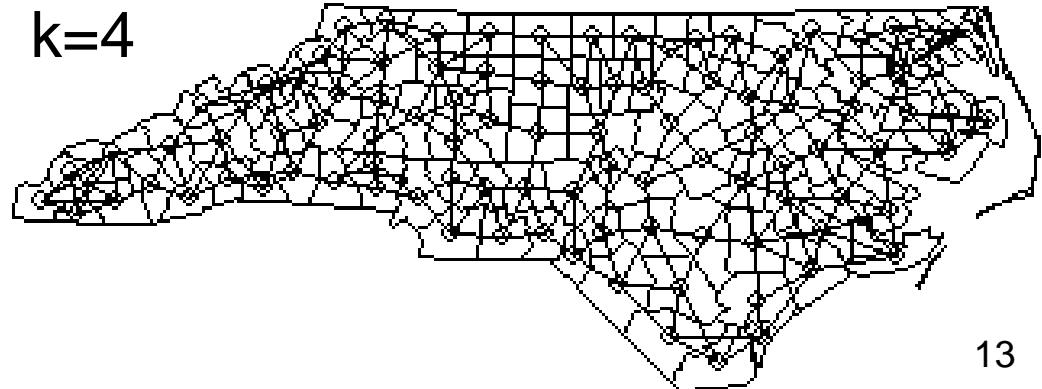
$k=1$



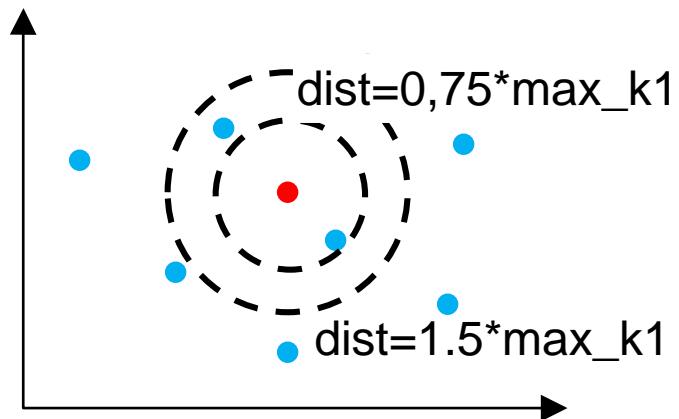
$k=2$



$k=4$

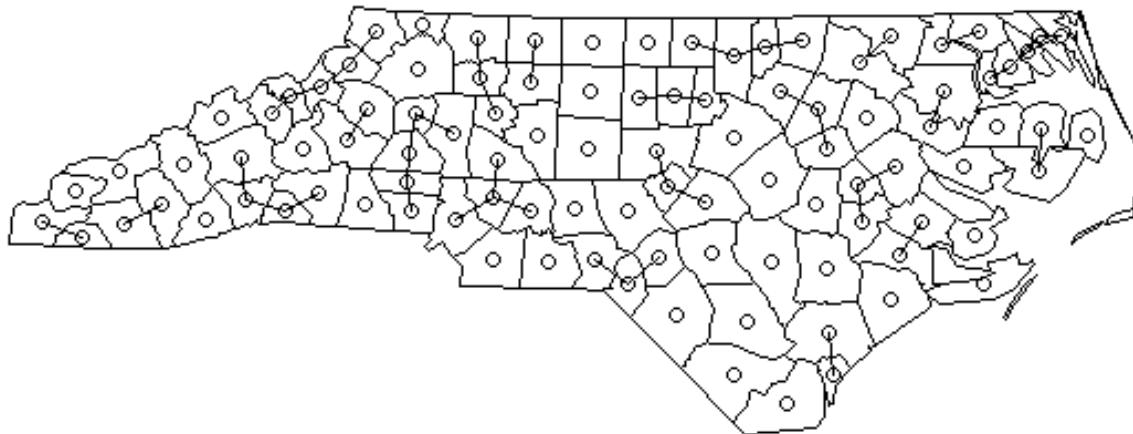


c. Distance based neighbors

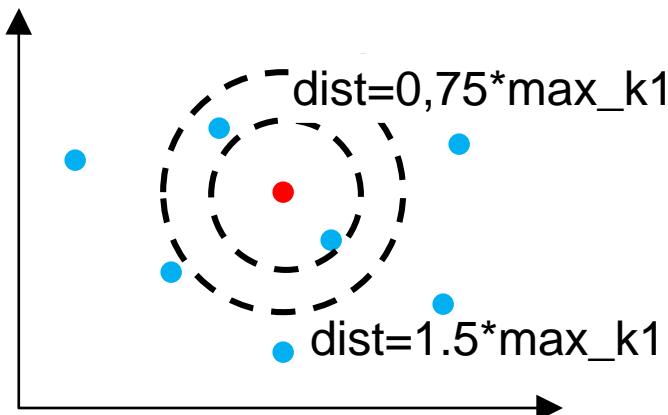
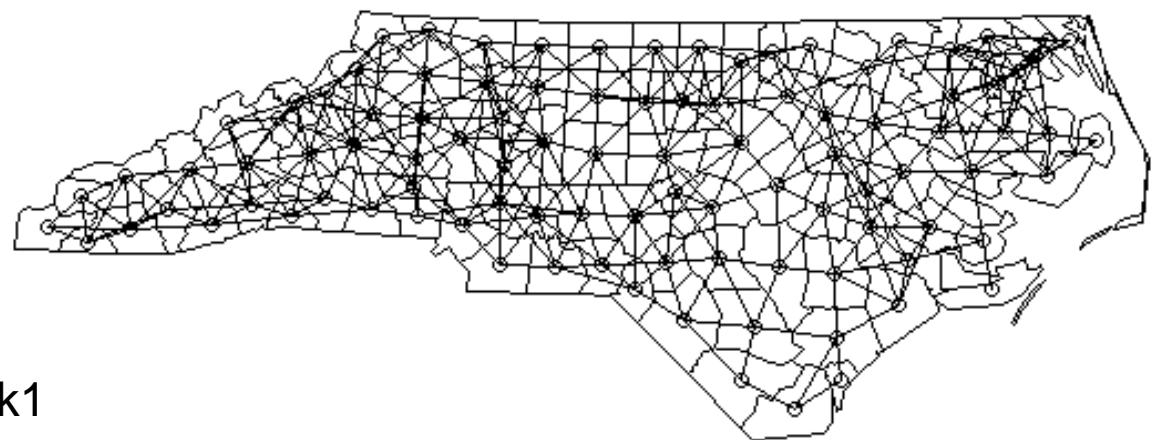


Max_k1=distance
maximum between
centroids

dist=0.75*max_k1



dist=1.5*max_k1



2. DEFINE SPATIAL WEIGHTS MATRICES

a. Binary weights (0/1)

Z_1	Z_2
Z_3	Z_4

$$W = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

Properties:

Matrix of 0's and 1's, where:

- 1 indicates that region j^{th} is neighbor of the region i^{th}
- 0 indicates that region j^{th} is not neighbor of the region i^{th}
- $w_{ii}=0$
- $w_{ij}=w_{ji}$ (Symmetry matrix)

Grid of 3x3 = 9 regions

1	2	3
4	5	6
7	8	9

$$W = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

b. Row-standardized weights

Spatial weights style of row standardization

$$\frac{w_{ij}}{\sum_{j \in J} w_{ij}}$$

$$W = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ \dots & & & & & & & & \end{bmatrix}$$

c. According the distances between regions or other criteria

Degree of relationship between two regions can depend on:

- distance among their geographic centroids
- length of the share boundaries

Definitions

$$\text{Anselin (1993)} \longrightarrow w_{ij} = d_{ij}^{-2}$$

Where d_{ij} Euclidean distance between region i^{th} and j^{th} i

$$\text{Cliff i Odr (1973)} \longrightarrow w_{ij} = d_{ij}^{-a} [\beta_{i(j)}]^b$$

where β proportion of the length of the share boundaries between region i^{th} and j^{th} ($a,b>0$)

A few papers justify the selection of weighted matrix. The most used it is the binary definition (share boundaries)

Eranest Arul (2007). Evaluated the influence for different weighted matrix. Data number of births with problems of malformations in New South Wales (Austràlia)

Exploratory data analysis

Exploratory analyses of space data consists in a group of technical that allows to describe the spatial distribution, identify spatial outliers and discover spatial clusters



Definition Spatial autocorrelation

Types of tests

Global
Local

GLOBAL TEST

1. Moran's Test (Moran 1948)

$$I = \frac{N \sum_i \sum_j w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\left(\sum_{i \neq j} w_{ij} \right) \sum_{i \neq j} (z_i - \bar{z})^2}$$

where w_{ij} are the spatial weights between region i th and j th

Interpretation

$$I \begin{cases} < E(I) \text{ negative spatial correlation} \\ = E(I) \text{ spatial independence} \\ > E(I) \text{ positive spatial correlation} \end{cases}$$

H_0 : No spatial correlation

H_a : Positive spatial correlation

1. Normal approximation

Under the null hypothesis, the I is asymptotically normal with

$$E(I) = -\frac{1}{n-1}$$

$$\text{Var}(I) = \frac{n^2(n-1)S_1 - n(n-1)S_2 - 2S_0^2}{(n+1)(n-1)^2 S_0^2}$$

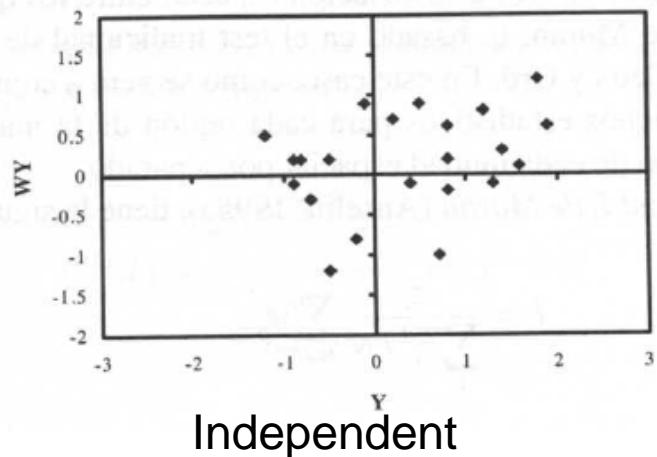
on $S_0 = \sum_{i \neq j} w_{ij}$ $S_1 = \frac{1}{2} \sum_{i \neq j} (w_{ij} + w_{ji})^2$ $S_2 = \sum_k \left(\sum_j w_{kj} + \sum_i w_{ik} \right)^2$

2. –Permutation test for Moran's I statistic

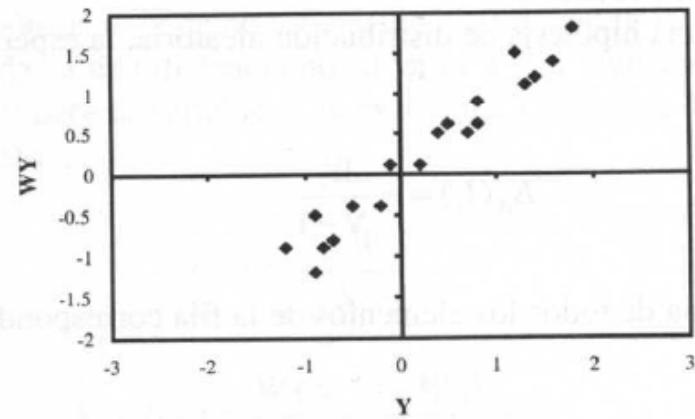
- The values of the variable of interest are randomly assigned to entities
- Establish position occupies the observed I Moran test in relation to the values of I Moran simulated under the hypotheses of spatial independence
- The P-value then estimates the proportion of C-values as extreme or more extreme than the observed C.

MORAN'S SCATTERPLOT

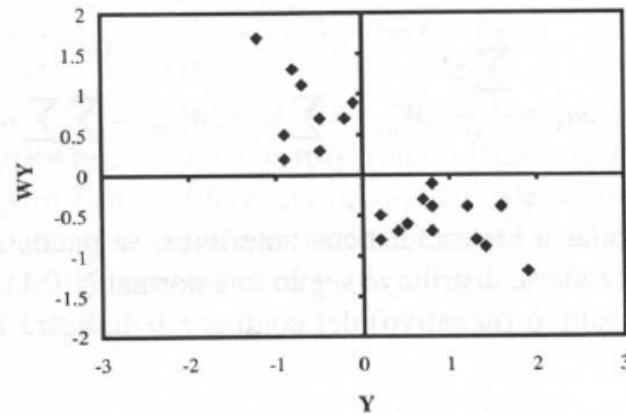
Variable of interest on the x-axis and the spatially weighted sum of values of neighbors- the spatially lagged values- on the y-axes.



Independent



Positive spatial dependence



Negative spatial dependence

2. Geary's Test (Geary 1954)

$$c = \frac{(n-1) \sum_{i=1}^n \sum_j w_{ij} (z_i - z_j)^2}{\left(\sum_{i \neq j} w_{ij} \right) \sum_i (z_i - \bar{z})^2}$$

where w_{ij} are the spatial weights between region i th and j th

- C too small \Rightarrow positive spatial autocorrelation.
- C too large \Rightarrow negative spatial autocorrelation.

Normal assumptions

$$E(c) = 1$$

$$\text{Var}(c) = \frac{(2S_1 + S_2)(n-1)S_2 - 4S_0}{2(n+1)S_0^2}$$

Same interpretation than Moran's test

$$c' = 1 - \frac{\sum_{i=1}^N \sum_j w_{ij} (z_i - z_j)^2}{\left(\sum_{i \neq j} w_{ij} \right) \sum_i (z_i - \bar{z})^2}$$

Moran's I is a measure of global spatial autocorrelation, while Geary's C is more sensitive to local spatial autocorrelation.

LOCALS INDICADORS

Local Moran

Break global measures down into their components and by extension construct localizes tests to detect clusters and hotspots.

Local de Moran I_i

$$I_i = \frac{(z_i - \bar{z}) \sum_j^n w_{ij} (z_j - \bar{z})}{\sum_{i=1}^n (z_i - \bar{z})^2}$$

I_i (positive) detect clusters:
observations with very similar
neighbors

I_i (negative) detect hotspots:
observations with very different
neighbors.

Maptools Package (for use with R - WIN, MacOS, Linux) - Freeware

<http://cran.r-project.org/web/packages/maptools/index.html>

Set of tools for manipulating and reading geographic data, in particular ESRI shapefiles; C code used from shapelib. It includes binary access to GSHHS shoreline files.

The package also provides interface wrappers for exchanging spatial objects with packages such as PBSmapping, spatstat, maps, RArclInfo, Stata tmap, WinBUGS, Mondrian, and others.

sf

.Support for simple features, a standardized way to encode spatial vector data. Binds to 'GDAL' for reading and writing data, to 'GEOS' for geometrical operations, and to 'PROJ' for projection conversions and datum transformations.

spdep: Spatial Dependence (for use with R - WIN, MacOS, Linux) - Freeware
<http://cran.r-project.org/src/contrib/Descriptions/spdep.html> or
<http://spatial.nhh.no/R/spdep/>

A collection of functions that run in the R language to:

- Create spatial weights matrix objects from polygon contiguities, from point patterns by distance and tessellations, for summarising these objects, and for permitting their use in spatial data analysis.
- Tests for spatial autocorrelation, including global Moran's I, Geary's c, Hubert/Mantel general cross product statistic, Empirical Bayes estimates and Assun,^cao/Reis Index, Getis/Ord G and multicoloured join count statistics, local Moran's I and Getis/Ord G, saddlepoint approximations for global and local Moran's I.
- Estimate spatial simultaneous autoregressive (SAR) models.

Bibliografia:

Banerjee S Carlin BP, Gelfrand A.E. (2004) Hierarchical Modelling and Analysis for Spatial Data. Chapman & Hall /CRC

Bibliografia addicional:

Cliff A and Odr J (1973). Spatial autocorrelation, London:Piom

Lattice Data

Spatial autoregressive models
SAR and CAR models

Dra. ROSA ABELLANA

DEPARTAMENT SALUT PÚBLICA. UNIVERSITAT DE
BARCELONA

Introduction- Spatial Statistics Approaches

- It is common to assume that the observations are independent and identically distributed.
- When working with spatial data the observations are not independent.
- Spatial dependence can be modelled in different ways using statistical models.

Classical Model

- The vector of the response variable Z is multivariate normal and

$$Z = \mu + e$$

- Supposed than the mean of Z depends on some covariables X

$$Z = X^T \beta + e$$

- Assume that the random errors are independent identically distributed according to a Normal distribution :

- $e \sim N(0, \Lambda)$

where Λ is a diagonal matrix with elements σ^2

- To account the spatial correlation
a specific form of the variance matrix Λ
- Two approaches:
 - Simultaneous Autoregressive (SAR)
 - Conditionally Autoregressive (CAR)

Simultaneous autoregressive model(SAR)

The SAR specification:

$$e_i = \sum_{i=1}^n b_{ij} e_i + \varepsilon_i \quad \begin{aligned} \varepsilon_i &\sim N(0, \sigma^2 I) \\ b_{ij} &: \text{elements of adjacency matrix} \end{aligned}$$

Matrix form

$$e = Be + \varepsilon$$

Then the regression model

$$\begin{aligned} Z &= X^T \beta + e \\ Z &= X^T \beta + Be + \varepsilon \end{aligned}$$

Under this model Z is distributed according a multivariate normal with

$$E(Z) = X^t \beta$$

$$Var(Z) = \sigma^2 (I - B)^{-1} I (I - B)^{-T}$$

Re-parametrization of the model

$(I - B)$ Must be non-singular

Re-parametrization of this model can be writing $B = \rho W$

$$Z = X^T \beta + \rho W e + \varepsilon \quad \varepsilon_i \sim N(0, \sigma^2 I)$$

b_{ij} : elements of adjacency matrix

$$E(Z) = X^t \beta$$

$$Var(Z) = \sigma^2 (1 - \rho W)^{-1} I (1 - \rho W)^{-T}$$

Conditions to variance and covariance matrix will be non-singular:

W: neighborhood matrix binary (B)

$(I - \rho B)$ is non-singular, we need to impose that ρ ,

$\rho \in (1/\lambda_1, 1/\lambda_n)$ where λ are eigenvalues of B

W : neighborhood standardized by rows (W)

$(I - \rho W)$ is non-singular, we need impose that $\rho \in (-1, 1)$

Interpretation of the SAR error model

SAR model for the random error generate a global spatial dependence

Global: covariance matrix involved all the locations of the system

$$Var(Z) = \sigma^2 (1 - \rho W)^{-1} I (1 - \rho W)^{-T}$$



$$u = [I - \rho W]^{-1}$$

$$[I - \rho W]^{-1} = I + \rho W + \rho^2 W^2 + \dots \quad \text{Leontieff expansion}$$

$$Var(Z) = I + \rho W + \rho W' + \rho^2 (W^2 + WW' + W'^2) \dots$$

Potency of W involves a contiguity of higher order.

Potency of ρ involves progressive loss of the value of the covariance

Conditionally autoregressive models (CAR)

Introduced by Besag 1974, is called **conditionally autoregressive models (CAR)**.

$$Z = X^T \beta + e$$

Definition: Conditional distribution of the spatial error

$$f(e_i | e_{-i})$$

Normal distribution of the residual error

$$E[e_i | e_{-i}] = \sum_{j \neq i} c_{ij} (Z_j - X^t \beta)$$

$$Var[e_i | e_{-i}] = \sigma_i^2 \quad i = 1, \dots, n$$

Joint distribution

Multivariate distribution of Z

$$Z \sim NMV(X\beta, Q^-) \quad Q = D^{-1}(I - C)$$

Where $C = c_{ij}$ and D is diagonal matrix with elements $d_i = \sigma_i^2$

Restrictions for Q:

Ensure symmetric matrix $Q = D^{-1}(I - C)$

$$\frac{c_{ij}}{\tau_i^2} = \frac{c_{ji}}{\tau_j^2} \quad \text{For all } i \text{ and } j$$

1. Binary definition for C (non-standardized by row)

No problems of symmetry.

2. Row-standardized weights

If C is **standardized by rows** of the binary matrix

$w_{ij}=1$ if i and j are neighbors

$w_{ij}=0$ other case

$$c_{ij} = \frac{w_{ij}}{w_{i+}} \quad \tau_i^2 = \frac{\tau^2}{w_{i+}}$$

Conditional distribution of Z_i is Normal with expectation and variance

$$E[e_i|e_{-i}] = \sum_{j \neq i} \frac{w_{ij}}{w_{i+}} (Z_j - X^t \beta)$$

$$Var[e_i|e_{-i}] = \frac{\sigma^2}{w_{i+}}$$

Joint distribution of Z

$$Z \sim NMV(X\beta, \sigma^2 Q^-) \quad \text{where}$$

$$Q = (I_w - W)$$

$$I_w = \text{diag}(w_{i+})$$

Handicap

$(I_w - W)\mathbf{1} = 0$ then Q is singular



Joint distribution might be improper

The joint distribution is improper but the conditionals are proper
Model called as **intrinsically autoregressive model (IAR)**

IAR it is used to define a random spatial effects

The impropriety can be remedied. Redefine:

$$Q = (I_w - \rho W) \quad \text{Standardized by row}$$

Choice ρ to make Q^{-1} nonsingular

It is guarantee if

$$\rho \in (1/\lambda_1, 1/\lambda_n)$$

where $\lambda_1 < \dots < \lambda_n$ are the ordered eigenvalues of

$$I_w^{-1/2} \rho W I_w^{-1/2}$$

$$\lambda_1 < 0, \text{ and } \lambda_n < 1$$

Binary definition for W (non-standardized by row)

$$Q = D^{-1}(I - \rho W) \text{ is nonsingular if } \rho \in (1/\lambda_1, 1/\lambda_n)$$

values $\lambda_1 < \dots < \lambda_n$ eigenvalues of W

COMPARATION BETWEEN SAR AND CAR

Grid 3x3 (definition of neighbors tower criterion). MATRIX W

	1	2	3	4	5	6	7	8	9
1		1		1					
2	1		1		1				
3		1				1			
4	1				1		1		
5		1		1		1		1	
6			1		1				1
7				1				1	
8					1		1		1
9						1		1	

SAR

$$\Sigma = \sigma^2 (I - \rho W)^{-1} (I - \rho W)^{-T}$$

$\rho=0.25$

	1	2	3	4	5	6	7	8	9
1	1,86	1,33	0,63	1,33	1,25	0,67	0,63	0,67	0,39
2	1,33	2,48	1,33	1,25	2,00	1,25	0,67	1,02	0,67
3	0,63	1,33	1,86	0,67	1,25	1,33	0,39	0,67	0,63
4	1,33	1,25	0,67	2,48	2,00	1,02	1,33	1,25	0,67
5	1,25	2,00	1,25	2,00	3,50	2,00	1,25	2,00	1,25
6	0,67	1,25	1,33	1,02	2,00	2,48	0,67	1,25	1,33
7	0,63	0,67	0,39	1,33	1,25	0,67	1,86	1,33	0,63
8	0,67	1,02	0,67	1,25	2,00	1,25	1,33	2,48	1,33
9	0,39	0,67	0,63	0,67	1,25	1,33	0,63	1,33	1,86

CAR

$$\Sigma = \sigma^2 (I - \rho W)^{-1}$$

	1	2	3	4	5	6	7	8	9
1	1,20	0,39	0,13	0,39	0,25	0,11	0,13	0,11	0,05
2	0,39	1,32	0,39	0,25	0,50	0,25	0,11	0,18	0,11
3	0,13	0,39	1,20	0,11	0,25	0,39	0,05	0,11	0,13
4	0,39	0,25	0,11	1,32	0,50	0,18	0,39	0,25	0,11
5	0,25	0,50	0,25	0,50	1,50	0,50	0,25	0,50	0,25
6	0,11	0,25	0,39	0,18	0,50	1,32	0,11	0,25	0,39
7	0,13	0,11	0,05	0,39	0,25	0,11	1,20	0,39	0,13
8	0,11	0,18	0,11	0,25	0,50	0,25	0,39	1,32	0,39
9	0,05	0,11	0,13	0,11	0,25	0,39	0,13	0,39	1,20

1. Different values but the properties are the same.
2. The interior sites, and those with more neighbors have a larger variance. The variance are not stationary

SAR Binary weights

$$\Sigma = \sigma^2 (I - \rho W)^{-1} (I - \rho W)^{-T}$$

$$\rho=0.25$$

	1	2	3	4	5	6	7	8	9
1	1,86	1,33	0,63	1,33	1,25	0,67	0,63	0,67	0,39
2	1,33	2,48	1,33	1,25	2,00	1,25	0,67	1,02	0,67
3	0,63	1,33	1,86	0,67	1,25	1,33	0,39	0,67	0,63
4	1,33	1,25	0,67	2,48	2,00	1,02	1,33	1,25	0,67
5	1,25	2,00	1,25	2,00	3,50	2,00	1,25	2,00	1,25
6	0,67	1,25	1,33	1,02	2,00	2,48	0,67	1,25	1,33
7	0,63	0,67	0,39	1,33	1,25	0,67	1,86	1,33	0,63
8	0,67	1,02	0,67	1,25	2,00	1,25	1,33	2,48	1,33
9	0,39	0,67	0,63	0,67	1,25	1,33	0,63	1,33	1,86

CAR Binary weights

$$\Sigma = \sigma^2 (I - \rho W)^{-1}$$

	1	2	3	4	5	6	7	8	9
1	1,20	0,39	0,13	0,39	0,25	0,11	0,13	0,11	0,05
2	0,39	1,32	0,39	0,25	0,50	0,25	0,11	0,18	0,11
3	0,13	0,39	1,20	0,11	0,25	0,39	0,05	0,11	0,13
4	0,39	0,25	0,11	1,32	0,50	0,18	0,39	0,25	0,11
5	0,25	0,50	0,25	0,50	1,50	0,50	0,25	0,50	0,25
6	0,11	0,25	0,39	0,18	0,50	1,32	0,11	0,25	0,39
7	0,13	0,11	0,05	0,39	0,25	0,11	1,20	0,39	0,13
8	0,11	0,18	0,11	0,25	0,50	0,25	0,39	1,32	0,39
9	0,05	0,11	0,13	0,11	0,25	0,39	0,13	0,39	1,20

1. Different values but the properties are the same.
2. The interior sites, and those with more neighbors have a larger variance. The variance are not stationary

3. Covariance decreases with the distance

SAR Binary weights

	Z_2	Z_5	Z_3	Z_6	Z_9
Z_1	1.33	1.25	0.63	0.67	0.39

CAR Binary weights

	Z_2	Z_5	Z_3	Z_6	Z_9
Z_1	0.39	0.25	0.13	0.11	0.05

The correlation is not constant(effect of variance is not constant).

	Z_2	Z_5	Z_3	Z_6	Z_9
Z_1	0.62	0.49	0.34	0.31	0.21

	Z_2	Z_5	Z_3	Z_6	Z_9
Z_1	0,31	0,19	0,10	0,09	0,04

4. The ρ parameter controls the amount of autocorrelation.

If $\rho = 0.10$ $\text{cor}(Z_1, Z_2) = 0.21$

If $\rho = -0.25$ $\text{cor}(Z_1, Z_2) = -0.62$

If $\rho = 0.10$ $\text{cor}(Z_1, Z_2) = 0.10$

If $\rho = -0.25$ $\text{cor}(Z_1, Z_2) = -0.31$

5. The covariances are not invariant to translation:

$$\text{cov}(Z_1, Z_3) = 0.63$$

$$\text{cov}(Z_4, Z_6) = 1.25$$

$$\text{cov}(Z_1, Z_3) = 0.13$$

$$\text{cov}(Z_4, Z_6) = 0.18$$

6. The models are anisotropic:

$$\text{cov}(Z_2, Z_3) = 1.33 \text{ but } \text{cov}(Z_2, Z_5) = 2.00$$

$$\text{cor}(Z_2, Z_3) = 0.62 \text{ but } \text{cor}(Z_2, Z_5) = 0.69$$

$$\text{cov}(Z_2, Z_3) = 0.39 \text{ but } \text{cov}(Z_2, Z_5) = 0.5$$

$$\text{cor}(Z_2, Z_3) = 0.31 \text{ but } \text{cor}(Z_2, Z_5) = 0.36$$

Grid of 3x3 (definition of neighbors move tower). **MATRIX W* Row Standardized weights**

	1	2	3	4	5	6	7	8	9
1		1/2		½					
2	1/3		1/3		1/3				
3		1/2				1/2			
4	1/3				1/3		1/3		
5		1/4		1/4		1/4		1/4	
6			1/3		1/3				1/3
7				1/2				1/2	
8					1/3		1/3		1/3
9						1/2		1/2	

Example matrix Q^- with $\rho=0.9$ $Q = (I_w - \rho W)$

	1	2	3	4	5	6	7	8	9
1	0.95	0.50	0.36	0.50	0.36	0.29	0.36	0.29	0.26
2	0.50	0.75	0.50	0.36	0.39	0.36	0.29	0.29	0.29
3	0.36	0.5	0.95	0.29	0.36	0.50	0.26	0.29	0.36
4	0.50	0.36	0.29	0.75	0.39	0.29	0.50	0.36	0.29
5	0.36	0.39	0.36	0.39	0.61	0.39	0.36	0.39	0.36
6	0.29	0.36	0.50	0.29	0.39	0.75	0.29	0.36	0.50
7	0.36	0.29	0.26	0.50	0.36	0.29	0.95	0.50	0.36
8	0.29	0.29	0.29	0.36	0.39	0.36	0.50	0.75	0.5
9	0.26	0.29	0.36	0.29	0.36	0.50	0.36	0.5	0.95

1. The interior sites, those with more neighbors have smaller variance
2. The variances are not stationary
3. Covariance decreases with “distance”.

Example matrix Q^- with $\rho=0.9$ $Q = (I_w - \rho W)$

	1	2	3	4	5	6	7	8	9
1	0.95	0.50	0.36	0.50	0.36	0.29	0.36	0.29	0.26
2	0.50	0.75	0.50	0.36	0.39	0.36	0.29	0.29	0.29
3	0.36	0.5	0.95	0.29	0.36	0.50	0.26	0.29	0.36
4	0.50	0.36	0.29	0.75	0.39	0.29	0.50	0.36	0.29
5	0.36	0.39	0.36	0.39	0.61	0.39	0.36	0.39	0.36
6	0.29	0.36	0.50	0.29	0.39	0.75	0.29	0.36	0.50
7	0.36	0.29	0.26	0.50	0.36	0.29	0.95	0.50	0.36
8	0.29	0.29	0.29	0.36	0.39	0.36	0.50	0.75	0.5
9	0.26	0.29	0.36	0.29	0.36	0.50	0.36	0.5	0.95

1. The interior sites, those with more neighbors have smaller variance
2. The variances are not stationary

3. The covariance decrease with the distances

	Z_2	Z_5	Z_3	Z_6	Z_9
Z_1	0.5	0.36	0.36	0.29	0.26

And the correlations also decrease with the distances

	Z_2	Z_5	Z_3	Z_6	Z_9
Z_1	0.59	0.47	0.37	0.35	0.28

4. For $\rho=0.92$ $\text{cor}(Z_1, Z_2)=0.6171$

	Z_2	Z_5	Z_3	Z_6	Z_9
Z_1	0.62	0.50	0.41	0.38	0.31

5 . The covariances are not translation invariant

$$\text{cov}(Z_1, Z_3) = 0.5 \quad \text{cov}(Z_4, Z_6) = 0.29$$

6. The model is not isotropic.

$$\begin{aligned} \text{cov}(Z_2, Z_3) &= 0.5 \quad \text{but} \quad \text{cov}(Z_2, Z_5) = 0.39 \\ \text{cor}(Z_2, Z_3) &= 0.42 \quad \text{but} \quad \text{cov}(Z_2, Z_5) = 0.58 \end{aligned}$$

Bibliografia:

BOOKS

Cressie N.A. (1993). *Statistics for Spatial Data*. Wiley Series in Probability and Mathematical Statistics. Canada. **Chapter 6**

Banerjee S Carlin BP, Gelfrand A.E. (2004) Hierarchical Modelling and Analysis for Spatial Data. Chapman & Hall /CRC. **Chapter 3**

Clayton D.and Bernardinelli L. (1992). *Bayesian methods for mapping disease risks*. In small Area Studies in Geographical and Environmental Epidemiology (P.Elliott, J. Cuzich, et al) Oxford University Press. 205-20.

Moreno R i Vayà E. (2000). Técnicas econométricas para el tratamiento de datos espaciales: la econometría espacial. Ediciones UB. 44 manuals.

PAPERS

Besag J. (1974). *Spatial interaction and the statistical analysis of lattice systems*. Journal of the Royal Statistical Society, Series B. **36**:192-236.

Besag J., York J. and Mollié A.(1991). Bayesian image restoration, with applications in spatial statistics. Annals of the Institute of Statistical Mathematics. **43**: 1-59.

Breslow N.E. and Clayton D.G.(1993). *Approximate Inference in Generalized Linear Mixed Models*, Journal of the American Statistical Association. **88**: 9-25.

Clayton D. and Kaldor J. (1987). *Empirical Bayes Estimates of Age-standardized Relative Risks for Use in Disease Mapping*. Biometrics, **43**: 671-681

Lattice Data

- Disease Mapping
- Regression models

Dra. ROSA ABELLANA

DPT. Fonaments Clínics (UB)

Disease Mapping

- Disease mapping aims to determine the underlying disease risk from scattered epidemiological data and to represent it on a smoothed colored map.
- The response variable Y is a count (number of incident or deaths) per region
- The probability distribution of is no longer a Normal model. It is more suitable to use a Poisson model.

- The classical linear model will not be adequate because:
 - ✓ the assumption of normality of errors will be violated
 - ✓ the relationship between the response variable and the explanatory variables will not be lineal

Generalized linear models

- Model proposed for Nelder and Wedderburn (1972).
- Generalized linear model (GLM) is a generalization of ordinary linear regression.
- Allowing:
 - ✓ The distribution of Y , therefore the errors, is any member of an exponential family: such as the Gaussian (normal), Binomial, Poisson, Gamma families of distributions.

- ✓ The relationship between the expectation of the response variable and the covariate is nonlinear.
- ✓ The link function $g(\cdot)$, related the $\mu_i = E(Yi)$ and the covariates:

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik}$$

- ✓ The link function must be invertible, we can also write:

$$\mu_i = g^{-1} (\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik})$$

Some common link function $g(\cdot)$

LINK	$g(\mu_i)$
Identity	μ_i
Log	$\log_e(\mu_i)$
Inverse	μ_i^{-1}
Inverse-square	μ_i^{-2}
Square-root	$\mu_i^{-1/2}$
Logit	$\log\left(\frac{\mu_i}{1 - \mu_i}\right)$
Probit	$\Phi^{-1}(\mu_i)$
Log-log	$-\log_e[-\log_e(\mu_i)]$
Complementary log-log	$\log_e[-\log_e(1 - \mu_i)]$

Count Data Models

- Variable coming from account be studied using Poisson regression
- Number of cases , Y_i , is assume to be distributed as a Poisson distribution

$$Y_i \sim Poisson(\mu_i)$$

- Exemples:
 - Number of cancer cases in a region
 - Cells with chromosomal anomalies
 - Traffic accidents

GLM: Regression Poisson

$$y_i \sim Poisson(\mu_i)$$

$$\log(\mu) = \beta_0 + \beta_1 \cdot X$$

$$E(Y) = \mu = \exp(\beta_0 + \beta_1 \cdot X) \quad \text{Var}(Y) = E(Y)$$

Exemple:

- Number of epileptic episodes in a month
- Explanatory variable: Age, sex, treatment

R-package:

```
glm(num_epileptic ~ Age + sex + treatmet), family = poisson , data = Dta.epi
```

Interpretation of regression coefficients

$$E(Y) = \mu = \exp(\beta_0 + \beta_1 X) = \exp(\beta_0) \cdot \exp(\beta_1 X)$$

$\exp(\beta_1)$ = with every unit increase in X

The X variable has multiplicative effect of $\exp(\beta_1)$ on the mean of Y, that is μ :

- If $\beta_1 = 0$, then $\exp(\beta_1) = 1$, and the expected count is $\exp(\beta_0)$ and, Y and X are not related.
- If $\beta_1 > 0$, then $\exp(\beta_1) > 1$, and the expected count is $\exp(\beta_1)$ times larger than when $X = 0$
- If $\beta_1 < 0$, then $\exp(\beta_1) < 1$, and the expected count is $\exp(\beta_1)$ times smaller than when $X = 0$

Model for incidence or mortality rates

- Count data is expressed as a rate.
- The observation units where the count has been carried out are not comparable because they correspond to:
 - ✓ Different population size
 - ✓ Different duration of study period

$$Rate = \frac{\text{number of cases}}{\text{People time at risk}} = \frac{Y}{N \cdot t}$$

- The expected value of the rate is $E\left(\frac{Y}{N \cdot t}\right) = \frac{E(Y)}{N \cdot t} = \frac{\mu}{N \cdot t}$
- The Poisson log-linear regression model for the expected rate:

$$\log\left(\frac{\mu}{N \cdot t}\right) = \beta_0 + \beta_1 \cdot X$$

$$\log(\mu) = \log(N \cdot t) + \beta_0 + \beta_1 \cdot X$$

- $\log(N \cdot t)$ is in the linear predictor but don't have regression coefficient.

$$\mu = (N \cdot t) \cdot \exp(\beta_0 + \beta_1 \cdot X)$$

- Number of expected cases is proportional to $N \cdot t$. Glm models must be define as an offset

R-package:

`glm(y~Age+sex offset=lpob, family =poisson , data = BBDD)`

Interpretation of the coefficients:

$\exp(\beta)$ is a quotient of rates, Risk Relative (RR)

$$RR = \frac{\text{Rate incidence in exposed}}{\text{Rate incidence in not exposed}}$$

$\beta_1 > 0$, then RR>1. Risk factor: the probability to have the disease is superior if you have the risk factor

$\beta_1 < 0$, then RR<1. Protective factor: the probability to have the disease is lower if you have the risk factor

$\beta_1 = 0$, then RR=1. Disease and factor risk are not associated. The probability to have disease is the same if you have or not the risk factor.

R-package:

`glm(y~Age+year, offset=lpob, family =poisson , data = BBDD)`

Y =number of breast cancer in Tarragona by age group and year

Age=age groups (<=50, >50 years)  Reference category <=50

Year=Year of diagnostic (2015,2016,2017)

Pob=population per year and age groups

- $\beta_{age>50} = 0,22 \longrightarrow RR = \exp(0,22) = 1,25$

Woman older than 50 years has 25% more risk of cancer diagnostic

- $\beta_{year} = 0,024 \longrightarrow RR_{year} = \exp(0,024) = 1,024$

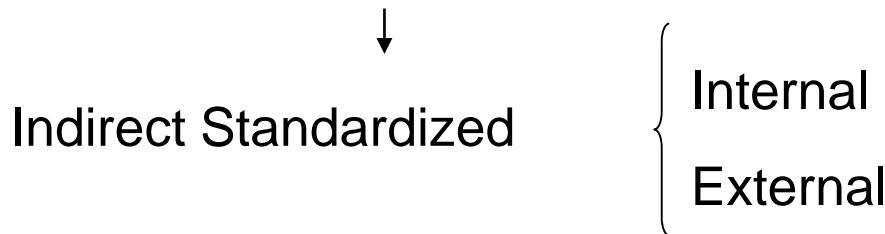
Percentage annual increase 2,4%

Application: Disease Mapping

Lattice data in the field of health sciences study the risk of suffering or dying of a particular area.

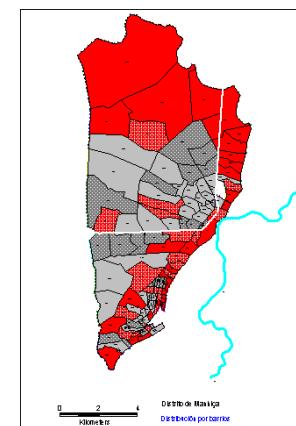
Main Objective: Analyze the geographical variability of rates of illness.

To eliminate possible differences among regions caused by variables, such as gender or age



Standardized mobility or mortality ratio

$$\hat{r}_i = \text{SMR}_i = \frac{Y_i}{E_i}$$



Example of indirect standardization of rates: Mortality rates

REGION A

Age groups	Nº deaths	Population	Rate
20-29	5	100	0,05
30-39	5	100	0,05
40-49	20	200	0,10
50-59	75	500	0,15
60-69	240	600	0,40
Total	345	1500	0,23

REGION B

Age groups	Nº deaths	Population	Rate
20-29	20	500	0,04
30-39	24	400	0,06
40-49	20	200	0,10
50-59	30	300	0,10
60-69	32	100	0,32
Total	126	1500	0,084

Mortality rate for all the country
(Reference population A+B)

Age groups	Nº deaths	Population	Rates
20-29	25	600	0,04
30-39	29	500	0,06
40-49	40	400	0,10
50-59	105	800	0,13
60-69	272	700	0,39

Indirect Standardization (INTERNAL)

Compute using the specific rate of the Standard population

REGION A

Age groups	Nº deaths	Population	Expected
20-29	5	100	4
30-39	5	100	6
40-49	20	200	20
50-59	75	500	65
60-69	240	600	234
Total	345	1500	329

REGION B

Nº visits	Population	Expected
20	500	20
24	400	24
20	200	20
30	300	39
32	100	39
126	1500	142

$$SMR_i = \frac{Observed_i}{Expected_i}$$

Region A: $SMR_A = \frac{345}{329} = 1,05$
 IC 95% : (0.94; 1.16) pvalue=0.39

Region B: $SMR_B = \frac{126}{142} = 0,89$
 IC 95% : (0.73; 1.04) pvalue=0.18

Interpretation of SMR

$SMR_i > 1$ indicates that there is a higher risk in the region of study than in the population of reference

$SMR_i < 1$ less risk

$SMR_i = 1$ equal risk to the population of reference

Model for SMR

$$Y_i \sim Poisson(\mu_i)$$

$$\log\left(\frac{E(Y)}{E}\right) = \alpha + \beta \cdot X$$

$$\log(Y) = \text{log}(E) + \alpha + \beta \cdot X$$



$$\mu = E \times \exp(\alpha + \beta X)$$

Problems with Poisson

- A characteristic of the Poisson distribution is that its mean is equal to its variance.
- An issue of importance is when **empirical variance** in the data exceeds the **nominal variance** under presumed model.
- This is known as **overdispersion**.
- Overdispersion leads to underestimated standard errors and overestimated significance of regression parameters

Causes of overdispersion

- correlation between individual responses
 - e.g. litters of rats
- cluster sampling
 - e.g. areas; schools; classes; children
- omitted unobserved variables
- excess zero counts (structural and sampling zeros)
- presence of spatial correlation

Detection Over dispersion in Poisson regression

- If the Poisson model fits the data reasonably, we would expect the residual **deviance** to be roughly equal to the **degrees of freedom (n-p)**.

The **residual deviance** is the difference between the **deviance** of the current **model** and the maximum **deviance** of the **ideal model** where the predicted values are identical to the observed.

- That the **residual deviance is so large** suggests that the variation of the data **exceeds** the variation of a Poisson-distributed variable, for which the variance equals the mean.

Solutions of this problem

Source of overdispersion:

- Unknown
 - ✓ Fit a Poisson quasi likelihood.
 - ✓ Fit Negative Binomial GLM.
- Arises from excess of zeroes
 - ✓ Zero-inflated Poisson regression
- Arises from presence of spatial correlation in data:
 - ✓ Add random effects with a spatial correlation matrix.

1. Fit a Poisson quasi likelihood

To introduce a dispersion parameter into the Poisson model, so that the variance of the response is

$$V(Y) = \phi\mu$$

If $\phi > 1$, therefore, the variance of Y increases more rapidly than its mean.

We use a method-of-moments estimator for the dispersion parameter.

$$\Phi = \frac{1}{n - k - 1} \sum \frac{(Y_i - \widehat{\mu}_i)^2}{\widehat{\mu}_i}$$

2. Fit a Negative Binomial

A simple way to allow for a higher variance is to use **Negative Binomial distribution instead of the Poisson**



Under the Poisson the mean, μ_i , is assumed to be constant within classes.

However, by defining a specific distribution for mean, heterogeneity within class is allowed.

- Assume μ to be Gamma :

$$E(\mu) = \mu \quad \text{and} \quad \text{Var}(\mu) = \mu^2/\alpha$$

- The distribution of $Y| \mu$ to be the Poisson distribution with conditional mean $E(Y | \mu) = \mu$

The marginal distribution of Y_i , $\underset{\mu}{\sim}$ is a **binomial negative**

$$P(Y_i) = \int_{\mu} P(Y_i | \mu_i) \cdot f(\mu_i) d\mu_i$$

With:

$$E(Y) = E[E[Y | \mu]] = E[\mu] = \mu$$

$$\text{Var}(Y) = E[\text{Var}(Y | \mu)] + \text{Var}(E[Y | \mu]) = E[\mu] + \text{Var}(\mu) = \mu + \mu^2/\alpha$$

If $\alpha \rightarrow \infty$, there would be **no over-dispersion**.

4. Overdispersion when there are more zeroes

- A particular kind of over dispersion obtains when there are more zeroes in the data than is consistent with a Poisson distribution
- Theory suggests that the excess of zeros are generated by a separate process from the count values.
- The excess zeros can be modeled independently.

- Model proposed for count data with an excess of zeroes, the zero-inflated Poisson regression (or ZIP) model, due to Lambert (1992).
- ZIP employs two components that correspond two zero generating processes:
 1. Binary distribution that generate extra zeros (or logit models commonly used)
 2. Poisson distribution that generated count (counting zeros) some of which may be zero.

It assumes that with probability π_i the only possible observation is 0, and with probability $1 - \pi_i$, a Poisson(μ) random variable is observed

The probability of observing a 0 count is:

$$P(Y_i = 0) = \pi_i + (1 - \pi_i) \cdot e^{-\mu_i}$$

The probability of observing any particular nonzero count y_i is:

$$P(Y_i = y_i) = (1 - \pi_i) \cdot \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} \quad \text{where } y_i = 1, 2, \dots$$

$$E(Y_i) = (1 - \pi_i) \cdot \mu_i$$

$$\text{Var}(Y_i) = (1 - \pi_i) \cdot \mu_i \cdot (1 + \pi_i \mu_i)$$

$\pi_i = 0$ Variance=Expectation

$\pi_i > 0$ Variance>Expectation

5. Overdispersion: Presence of spatial correlation data:

- Generalized linear mixed model (GLMM) is an extension to the generalized lineal model.
- The linear predictor contains random effects in addition to the usual fixed effects.
- Mixed model related to conditional mean to random effects.

$$E(y | b) = \mu^b$$

$$g(\mu^b) = \eta^b = X\beta + Zb$$

where X is the design matrix for the fixed effects

β Is the vector of the fixed parameters

Z is the design matrix for the random effects

b is the random effects vector (dim= N)

Assume that random effects is distributed as normal with mean zero and variance $\Sigma(\delta)$

$$\mathbf{b} \sim \text{NMV}(0, \Sigma(\delta))$$

Different structured for random effects

1. Spatial structured model

Introduced in this context by Clayton and Kaldor (1987)

Intrinsic CAR model $\rho=1$ $\mathbf{b} \sim N\left(0, \tau_s^2 (\mathbf{D}_w - \mathbf{W})^{-1}\right)$

The conditional distribution of the random effects is a Normal

$$b_i | b_{j \in \delta_i} \sim N\left(\bar{b}_i, \sqrt{\frac{\sigma_s^2}{n_i}}\right)$$

2. Heterogeneity model $b \sim NMV(0, \tau^2 I)$

Without spatial structure, assumes not spatial correlation

3. Spatial+ Heterogeneity model

- **Non-intrinsic CAR** Proposed by Besag, York and Mollié, 1991
- Two variability components: over dispersion non-structured and another over dispersion spatial structured.
- The distribution for random effects:

In this case

$$b \sim N\left(0, \tau_H^2 I_N + \tau_S^2 (D_w - \rho W)^{-1}\right)$$

$$E[b_i | b_j \ i \neq j] = \bar{b}_i = \frac{\sum_{j \in \partial_i} b_j}{n_i}$$

$$\text{Var}(b_i | b_j \ i \neq j) = \sigma_H^2 + \frac{\sigma_S^2}{n_i}$$

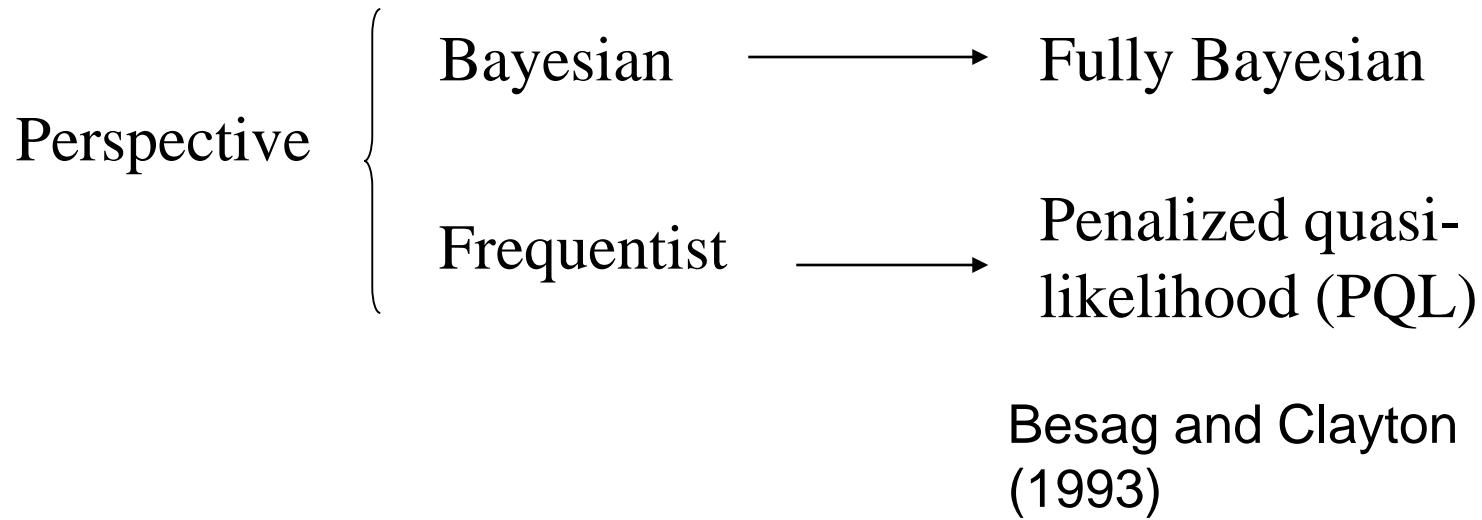
Estimation of the parameters

Maximum likelihood of the GLMM parameters

$$\begin{aligned} L(Y; \theta, \phi, \Sigma(\varpi)) &= \int \prod_{i=1}^N f_Y(y_i; \varpi, \phi) \cdot f(b, \Sigma(\theta)) d(b) = \\ &\int \prod_{i=1}^N \exp \left\{ \frac{(y_i \cdot \theta_i - b(\theta_i))}{a(\phi)} + c(y, \phi) \right\} \cdot f(b, \Sigma(\varpi)) d(b) \\ &= \int \prod_{i=1}^N \exp \left\{ \frac{(y_i \cdot \theta_i - b(\theta_i))}{a(\phi)} + c(y, \phi) \right\} \cdot \frac{1}{\sqrt{(2\pi)^N \Sigma(\varpi)}} \exp \left\{ -\frac{1}{2} b \cdot \Sigma^{-1}(\varpi) \cdot b^T \right\} db \end{aligned}$$

Since the random effects are not independent, the maximum likelihood function is not able to assess in a closed form.

Estimation of the parameters:

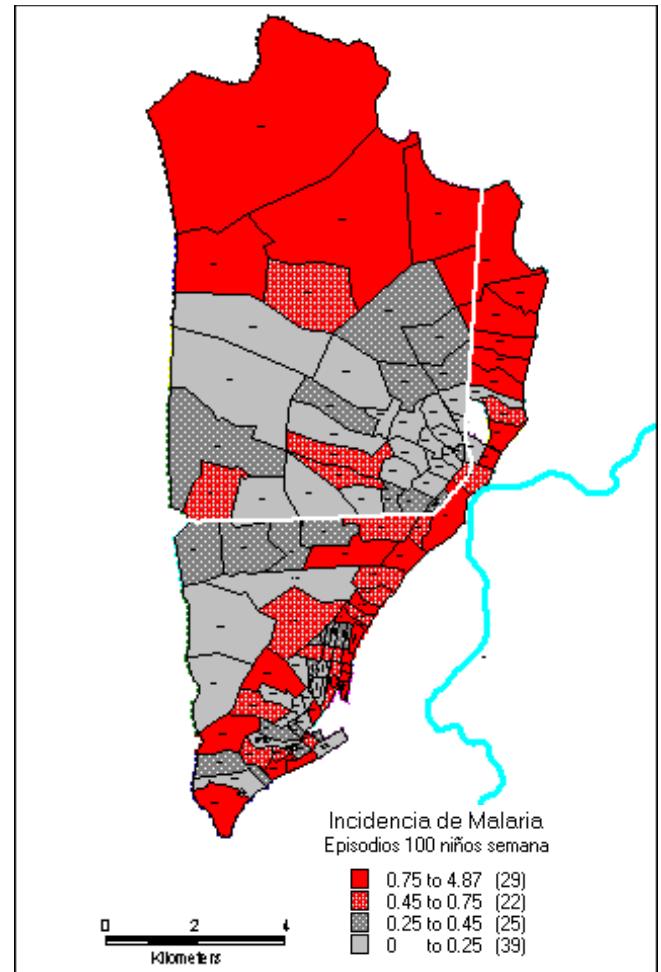


Example

Data: Following of 2006 children between 1 and 10 years old

Cohort (2 years) visited once a week, considering a week at risk

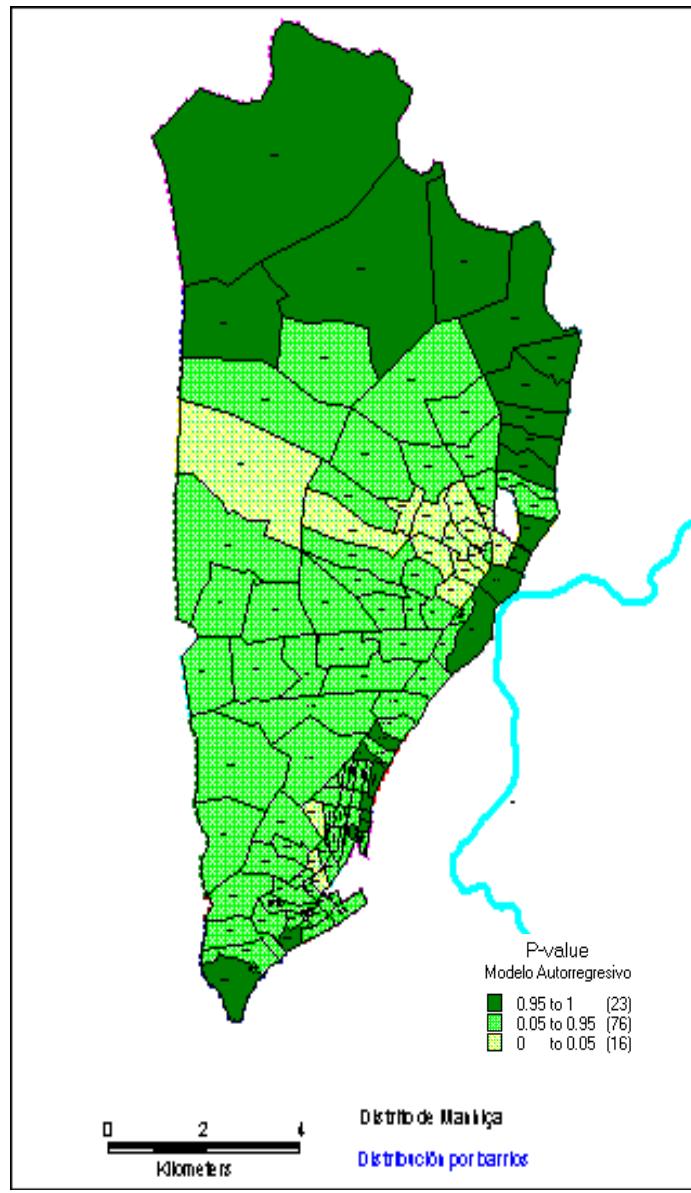
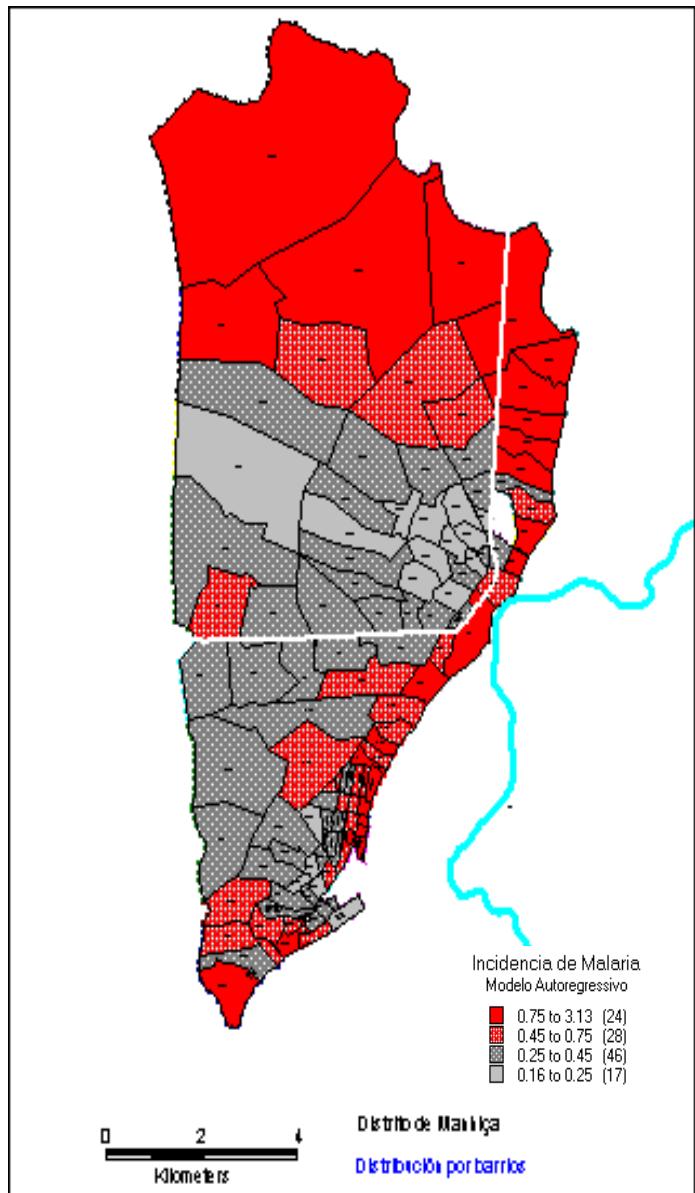
Incidence is expressed by number of episodes for 100 children week at risk



RESULTS

Model	Mean ⁽¹⁾	Desv. Het	Desv. CAR	DIC
Heterogeneity	0.43	0.748		194.97
CAR	0.43		0.731	176.11
Heterogeneity + CAR	0.42	0.327	0.664	178.65

100 children week at risk



Bibliografia:

BOOKS

Cressie N.A. (1993). *Statistics for Spatial Data*. Wiley Series in Probability and Mathematical Statistics. Canada. **Chapter 6**

Banerjee S Carlin BP, Gelfrand A.E. (2004) Hierarchical Modelling and Analysis for Spatial Data. Chapman & Hall /CRC. **Chapter 3**

Clayton D.and Bernardinelli L. (1992). *Bayesian methods for mapping disease risks*. In small Area Studies in Geographical and Environmental Epidemiology (P.Elliott, J. Cuzich, et al) Oxford University Press. 205-20.

Moreno R i Vayà E. (2000). Técnicas econométricas para el tratamiento de datos espaciales: la econometría espacial. Ediciones UB. 44 manuals.

PAPERS

Besag J. (1974). *Spatial interaction and the statistical analysis of lattice systems*. Journal of the Royal Statistical Society, Series B. **36**:192-236.

Besag J., York J. and Mollié A.(1991). Bayesian image restoration, with applications in spatial statistics. Annals of the Institute of Statistical Mathematics. **43**: 1-59.

Breslow N.E. and Clayton D.G.(1993). *Approximate Inference in Generalized Linear Mixed Models*, Journal of the American Statistical Association. **88**: 9-25.

Clayton D. and Kaldor J. (1987). *Empirical Bayes Estimates of Age-standardized Relative Risks for Use in Disease Mapping*. Biometrics, **43**: 671-681

Lattice Date

Bayesian Estimation
Hierarchical Models

Dra. ROSA ABELLANA

DEPARTMENT PUBLIC HEALTH. UNIVERSITY OF
BARCELONA

BAYESIAN INFERENCE

Classical framework unknown parameter, θ , is **fixed, non-random quantities**. The estimate of this parameter is performed by a function of the sample, the estimator is a random variable.

Bayesian framework the unknown parameter, θ , it is **a random variable** and it has a distribution of probability.

Combine two distributions:

- **Initial distribution or prior distribution** of the parameter, $p(\theta)$: expresses our uncertainty about θ before seeing the data.
- The **likelihood function** $p(x|\theta)$, the information of the parameter, θ , that we obtain from the data

Bayesian inference. Use the theorem of Bayes:

$$p(\theta | x) = \frac{p(x | \theta)p(\theta)}{p(x)} \alpha$$

A posterior distribution of the parameter

Prior distribution

Likelihood function

- $p(\theta|x)$ the **posterior distribution** is the final result of the Bayesian analysis and all the information on the parameter, it is obtained from this distribution. Posterior expresses our uncertainty about θ after seeing the data.

Point estimation

Given the posterior distribution $p(\theta|y)$, the Bayesian estimate of θ can be

Posterior mean $\hat{\theta} = E(\theta | y)$

Posterior median $\hat{\theta} : \int_{-\infty}^{\hat{\theta}} p(\theta | y) d\theta = 0.5$

Posterior mode $\hat{\theta} : p(\hat{\theta} | y) = \sup_{\theta} p(\theta | y)$

Interval estimation $(1 - \alpha)\%$

The same that the median if we want to calculate the percentile $\alpha/2$ and the $1 - \alpha/2$ of $p(\theta|y)$, that is to say the points q_l and q_u

$$\int_{-\infty}^{q_L} p(\theta | y) d\theta = \alpha/2$$

$$\int_{q_u}^{\infty} p(\theta | y) d\theta = 1 - \alpha/2$$

Therefore $P(q_l < \theta < q_u) = 1 - \alpha$.

Choosing prior distribution

When the prior and posterior come from the same family of distributions the prior is said to be **conjugate** to the likelihood

Likelihood	Parameter	Prior	Posterior
Normal	Mean	Normal	Normal
Normal	Accuracy	Gamma	Gamma
Binomial	Prob Event	Beta	Beta
Poisson	Rate or Mean	Gamma	Gamma

Conjugate prior distributions are mathematically convenient

Computations for non-conjugate priors are harder, but possible using Monte Carlo Markow Chain

Example 1: Bayesian inference using a Normal distribution

Known variance, unknown mean, θ

We suppose have a sample of data of a Normal $x_i \sim N(\theta, \sigma^2)$ ($i=1,..n$).

Assuming that σ^2 is known

Prior distribution of θ is $N(\mu_p, \sigma^2/n_0)$

The posterior distribution:

$$\begin{aligned} p(\theta|x) &\propto \prod_i p(x_i|\theta) p(\theta) \\ &\propto \exp\left[-\frac{\sum_i (x_i - \theta)^2}{2\sigma^2}\right] \times \exp\left[-\frac{(\theta - \mu)^2 n_0}{2\sigma^2}\right] \end{aligned}$$

Therefore, deleting all the terms that do not depend of θ

$$p(\theta|x) = N\left(\frac{n_0\mu + n\bar{x}}{n_0 + n}, \frac{\sigma^2}{n_0 + n}\right)$$

Concentration of (THMs) in drinking water

We suppose that we want to estimate the average concentration of THM in a zone.

We have 2 independent measures x_1 and x_2 . The mean is 130 $\mu\text{g/l}$.

We suppose that the standard deviation is $\sigma = 5 \mu\text{g/l}$

Which is the estimate of the average concentration, θ ?

Standard analyses

The estimation of θ would be $\bar{x} = 130 \mu\text{g/l}$

The estimation of the standard error would be $\sigma_e / \sqrt{n} = 5 / \sqrt{2} = 3.5 \mu\text{g/l}$

A 95% confidence interval θ $\bar{x} \pm 1.96 \times \sigma_e / \sqrt{n}$ 132.1 a 136.9 $\mu\text{g/l}$

We suppose that you have historical data on concentrations levels of THM. The average of the concentration is 120 $\mu\text{g/l}$ and the standard deviation is 10 $\mu\text{g/l}$

Prior distribution for $\theta \sim N(120, 10^2)$

If we express the prior standard deviation prior as $\sigma_e / \sqrt{n_0}$

Then $n_o = (\sigma_e / 10)^2 = 0.25$

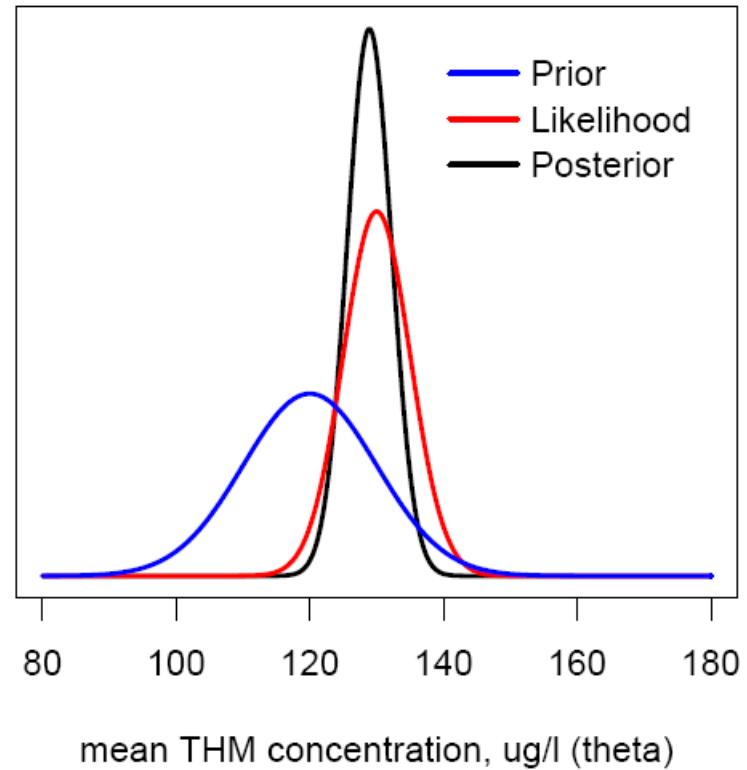
So our prior can be written as $\theta \sim N(120, \sigma_e^2 / 0.25)$

$$\theta \sim N(120, 5^2 / 0.25)$$

And the posterior distribution of θ

$$\begin{aligned} p(\theta_{zq} | \mathbf{x}) &= \text{Normal} \left(\frac{0.25 \times 120 + 2 \times 130}{0.25 + 2}, \frac{5^2}{0.25 + 2} \right) \\ &= \text{Normal}(128.9, 3.33^2) \end{aligned}$$

Obtaining an interval for θ of 122.4 to 135.4 $\mu\text{g/l}$ – 95%- Credibility interval-



Example 2: Estimation for disease risk in a single area

We are interested in estimating the rate or relative risk for some Poisson data

We suppose that we observe $x=5$ cases of leukemia in a region and that the number of cases expected standardized by age-gender is $E=2.8$.

We assume that the likelihood for x is a Poisson (λE) , where λ is the unknown relative risk

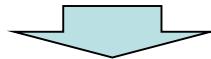
$$p(x|\lambda, E) = \frac{(\lambda E)^x e^{-\lambda E}}{x!}$$

We assume that the distribution a priori for λ is a Gamma(a, b)

$$p(\lambda) = \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda} \quad \begin{aligned} E(\lambda) &= a/b \\ \text{Var}(\lambda) &= a/b^2 \end{aligned}$$

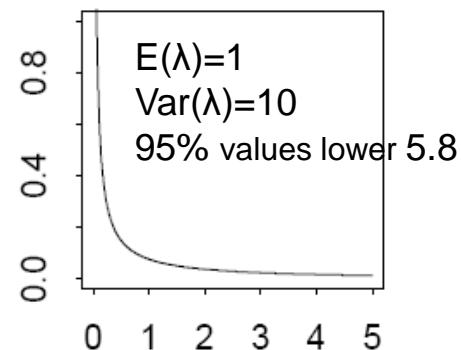
The posterior distribution for λ will be

$$\begin{aligned} p(\lambda|x, E) &\propto \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda} \frac{(\lambda E)^x e^{-\lambda E}}{x!} \\ &\propto \lambda^{a+x-1} e^{-(b+E)\lambda} \propto \text{Gamma}(a+x, b+E) \end{aligned}$$

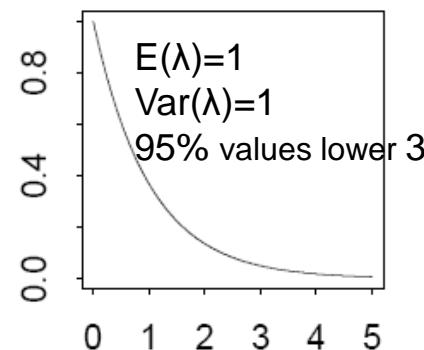


$$E(p(\lambda|x, E)) = \frac{a+x}{b+E}$$

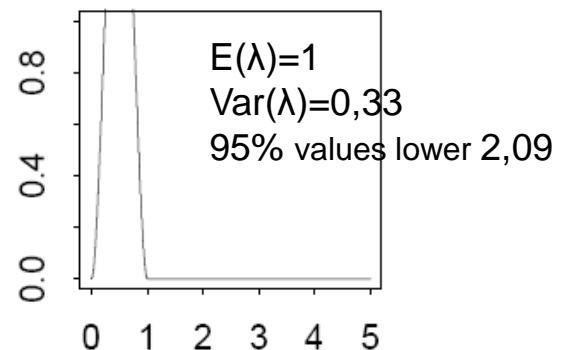
Gamma(0.1,0.1)



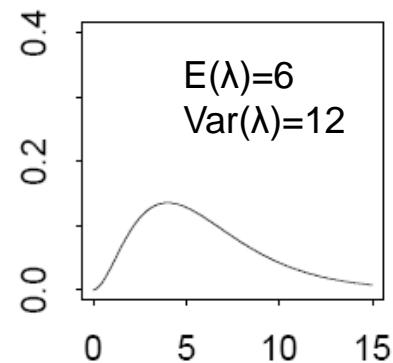
Gamma(1,1)



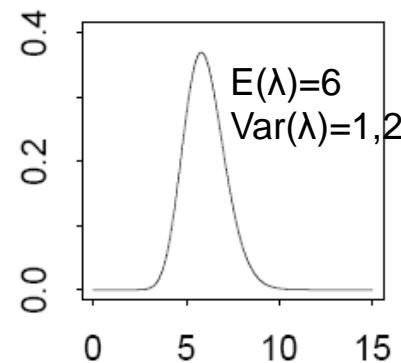
Gamma(3,3)



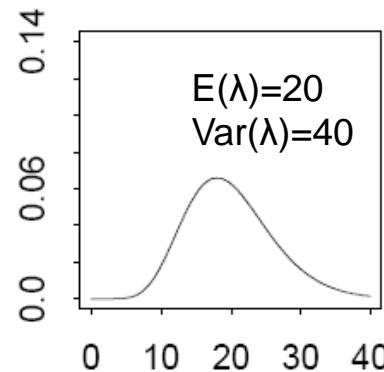
Gamma(3,0.5)



Gamma(30,5)



Gamma(10,0.5)



X~Gamma(α,β)

$$\left\{ \begin{array}{l} E(x)=\alpha/\beta \\ Var(X)=\alpha/\beta^2 \end{array} \right.$$

The posterior distribution for λ will be

$$\begin{aligned} p(\lambda|x, E) &\propto \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda} \frac{(\lambda E)^x e^{-\lambda E}}{x!} \\ &\propto \lambda^{a+x-1} e^{-(b+E)\lambda} \propto \text{Gamma}(a+x, b+E) \end{aligned}$$

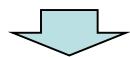
Vague Prior Information about λ

$$\lambda \sim \text{Gamma}(0.1, 0.1)$$

$$E(\lambda) = 0.1/0.1 = 1$$

$$\text{Var}(\lambda) = 0.1/0.1^2 = 10$$

$$95^{\text{th}} \text{ percentile} = 5.8$$



Posterior distribution Gamma(5.1, 2.9)

Posterior mean for λ

$$E(\lambda|x) = 5.1/2.9 = 1.76$$

$$\text{CI}(95\%): [0.58, 3.58]$$

$$(\text{SMR} = 5/2.8 = 1.78)$$

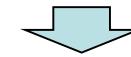
Strong Prior Information about λ

$$\lambda \sim \text{Gamma}(48, 40)$$

$$E(\lambda) = 48/40 = 1.2$$

$$\text{Var}(\lambda) = 48/40^2 = 0.03$$

$$95^{\text{th}} \text{ percentile} = 1.5$$



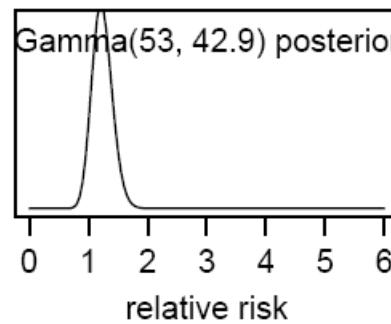
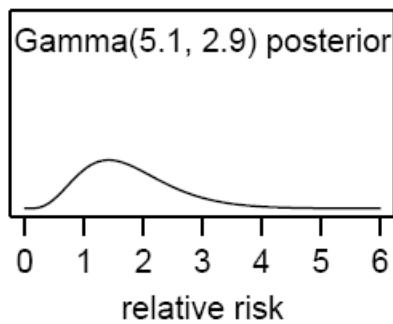
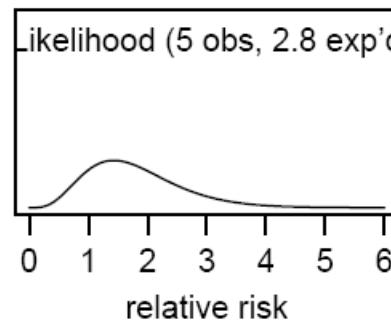
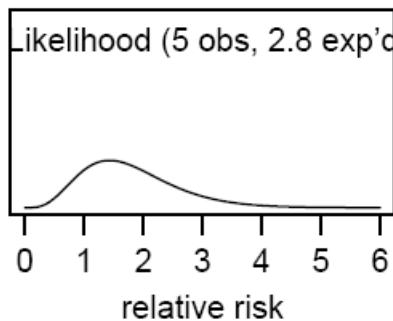
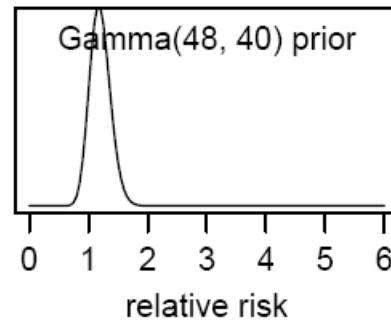
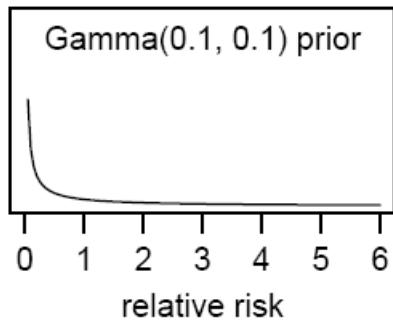
Posterior distribution Gamma(53, 42.9)

Posterior mean for λ

$$E(\lambda|x) = 53/42.9 = 1.24$$

$$\text{CI}(95\%): [0.92, 1.56]$$

$$(\text{SMR} = 5/2.8 = 1.78)$$



Where does the prior come from?

- It is in principle subjective.
- It might be obtained by experts.
- It might be more convincing to be based on historical data.
- If you don't have any information about the parameters. To be objective, select a **non informative prior**.

Recommendations

- Location parameters (e.g. mean, coefficients of regression)

$$\theta \sim \text{Unif}(-100, 100)$$

$$\theta \sim \text{Normal}(0, 100000)$$

- Parameters scale $\tau = \sigma^{-2}$ (inverse variance, precision)

$$\tau \sim \text{Gamma}(\epsilon, \epsilon) \quad \epsilon=0.001$$

Why is computation important?

More than one parameter of interest $\theta = (\theta_1, \theta_2, \dots, \theta_k)$

First find the **joint posterior** distribution of all the parameters,

$$p(\theta_1, \theta_2, \dots, \theta_n | x) \sim p(x | \theta_1, \theta_2, \dots, \theta_n) \cdot p(\theta_1, \theta_2, \dots, \theta_n)$$

Next obtain the **marginal distribution marginal** of each parameters

$$p(\theta_1 | x) = \int \int \dots \int p((\theta_1, \theta_2, \dots, \theta_n | x) d\theta_2 \dots d\theta_n$$

$P(x | \theta)$ and $p(\theta)$ will be available in closed form

But $p(\theta_1, \theta_2, \dots, \theta_n | x)$ is usually high dimensional and not analytically tractable.

- To avoid to resolve analytically these integrals you can use a Markov chain Monte Carlo (MCMC) algorithm.
- This algorithm allows obtaining a sequence of observations which are approximated from a specified multivariate probability distribution, when direct sampling is difficult.
- This sequence can be used to approximate the posterior distribution of the parameters
- There are several MCMC algorithm, we will use the Gibbs Sampling.

Gibbs sampling, works from the full conditioned distributions

Sampling of Gibbs

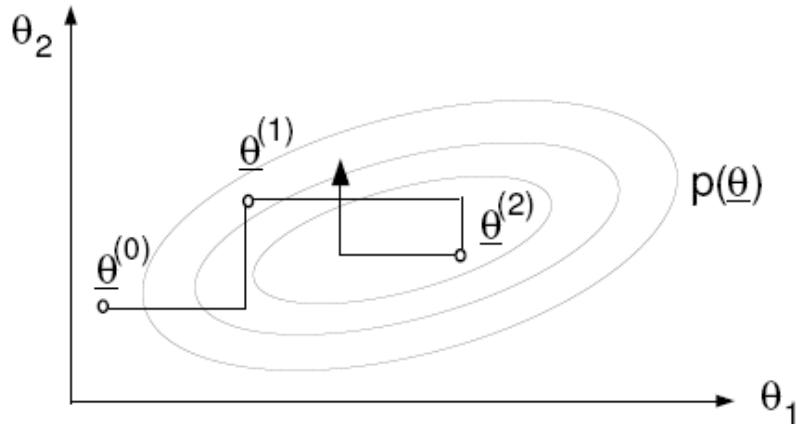
We suppose that have n parameters with distributions conditioned

$$p(\theta_1 | \theta_2, \dots, \theta_n) \quad p(\theta_2 | \theta_1, \dots, \theta_n) \quad p(\theta_n | \theta_2, \dots, \theta_{n-1})$$

- 1) Choose initial values. $[\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_n^{(0)}]$
- 2) Sample $\theta_1^{(1)}$ from $\theta_1^{(1)} \sim p(\theta_1 | \theta_2^{(0)}, \dots, \theta_n^{(0)}, x)$
Sample $\theta_2^{(1)}$ from $\theta_2^{(1)} \sim p(\theta_2 | \theta_1^{(1)}, \dots, \theta_n^{(0)}, x)$
....
Sample $\theta_k^{(1)}$ from $\theta_k^{(1)} \sim p(\theta_k | \theta_1^{(1)}, \dots, \theta_{k-1}^{(1)}, x)$
- 3) Repeat step 2 many 1000 s(or n) of times

The conditional distributions are called **full conditionals** because they are conditioned to all the other parameters

Example with k=2

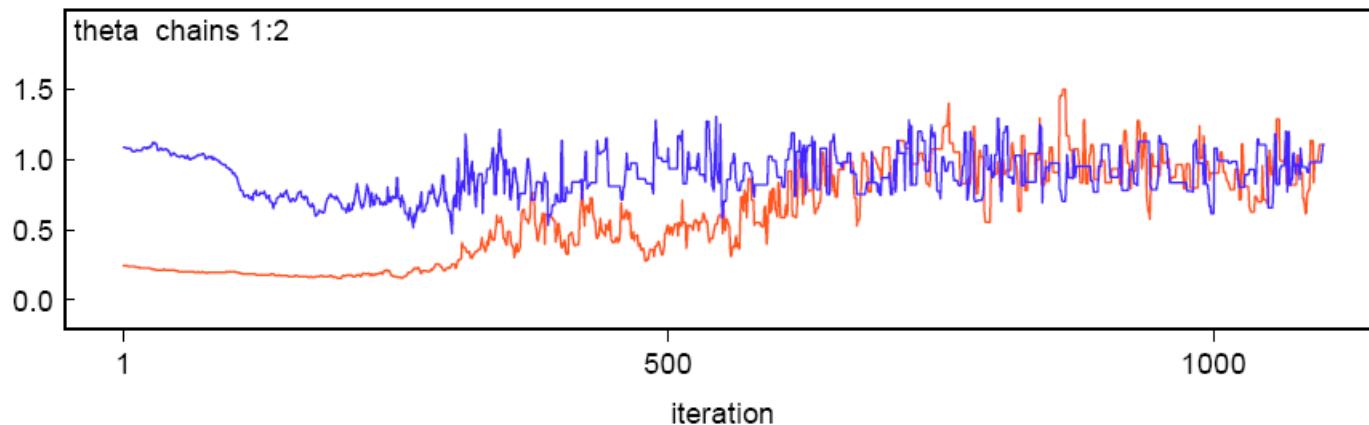


- | | |
|---------------------------------|---|
| 2) Sample $\theta_1^{(1)}$ From | $\theta_1^{(1)} \sim p(\theta_1 \theta_2^{(0)}, x)$ |
| Sample $\theta_2^{(1)}$ From | $\theta_2^{(1)} \sim p(\theta_2 \theta_1^{(1)}, x)$ |
| Sample $\theta_1^{(2)}$ From | $\theta_1^{(2)} \sim p(\theta_1 \theta_2^{(1)}, x)$ |

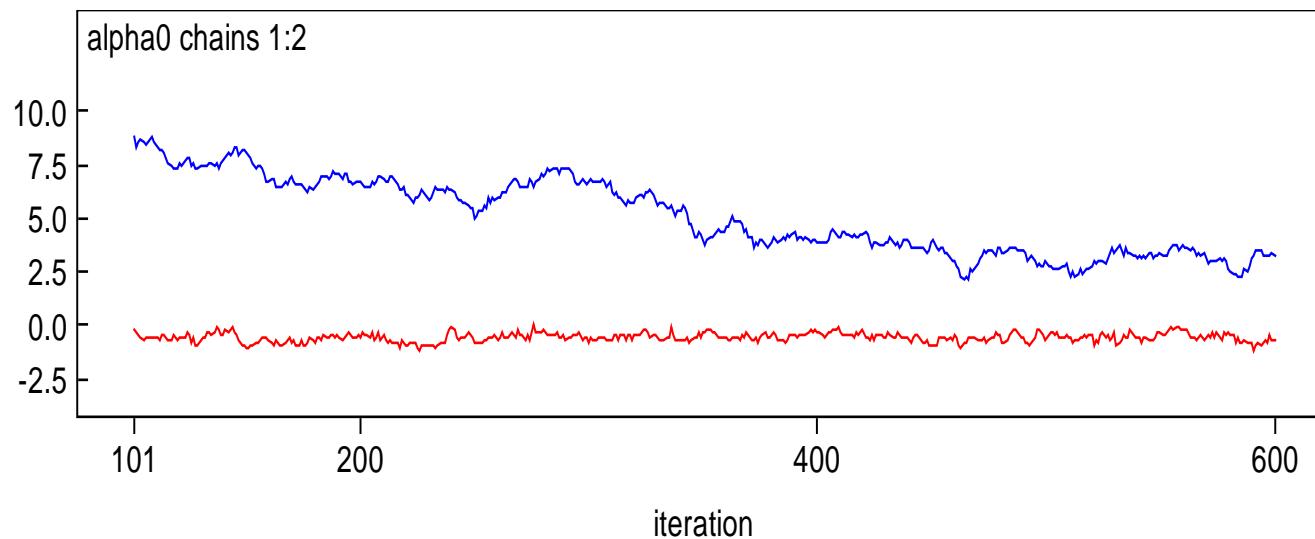
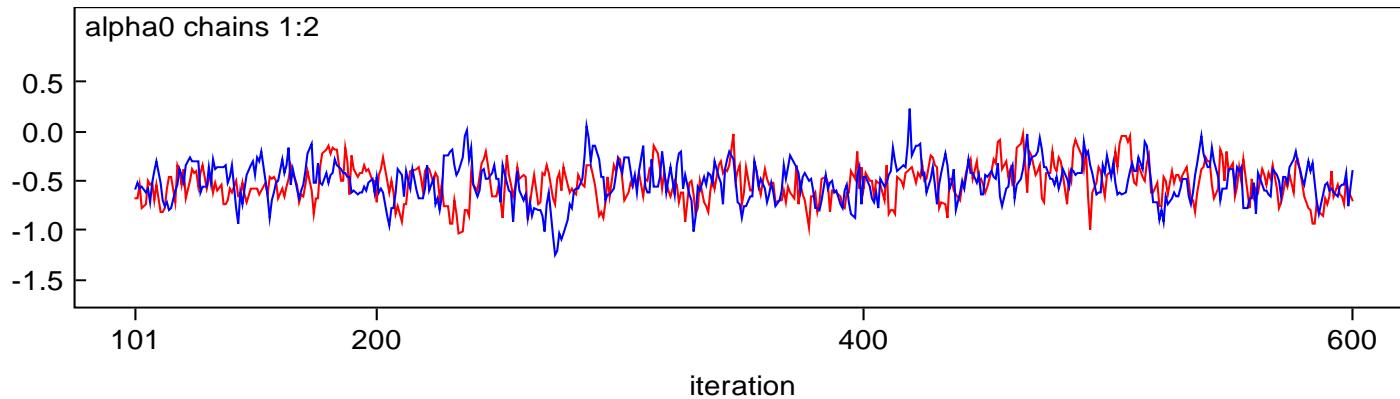
Checking convergence

- Using MCMC algorithm, there is a one main issue to consider:
 - To check the convergence of the algorithm
- Practice, is to run multiple chains for the same model but with different initial points, and visually inspect “trace” plot for convergence
- Convergence is to a distribution not to a single value.
- Once convergence reached, samples should look like a random scatter about a stable mean value.

Trace Plot



Example convergence and no convergence

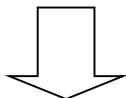


Sampling of Gibbs

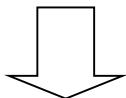


We stroll for the sample space

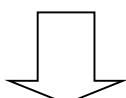
First samples are discarded



Burning process



Check if the chains came from posterior
distribution



Convergence to posterior distribution.

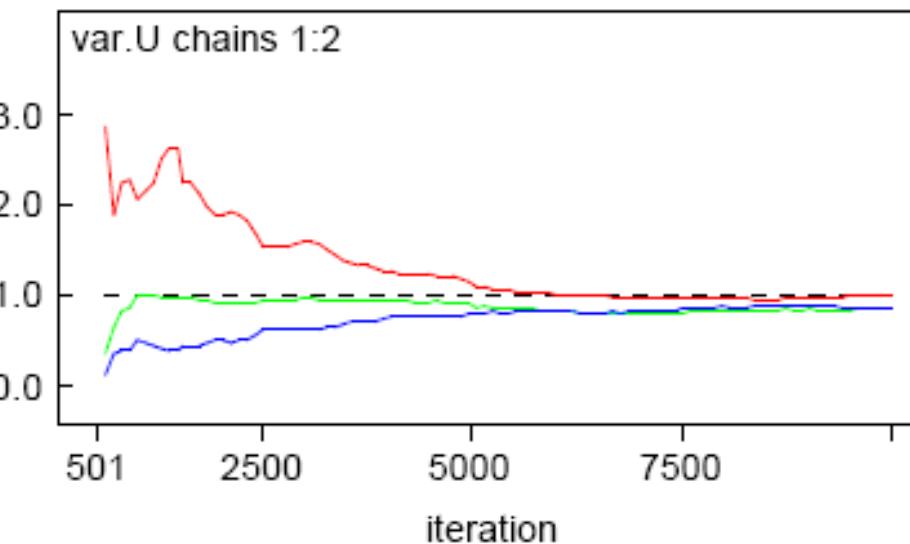
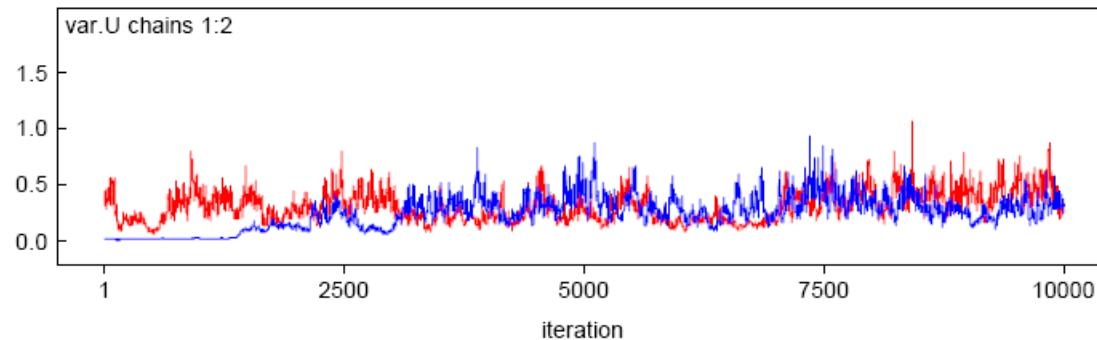
Statistician of Gelman-Rubin

Formally, can use the Gelman-Rubin statistician based in the ratio of the *between* and *within* chains variance (ANOVA)

Intuitively, the behavior of all the chains would have to be the same.

Or, the variance within the chain would have to be the same that the variance along the chain

Example of Gelman and Rubin with the Winbugs



B_{pooled}

$W_{\text{average of width}}$

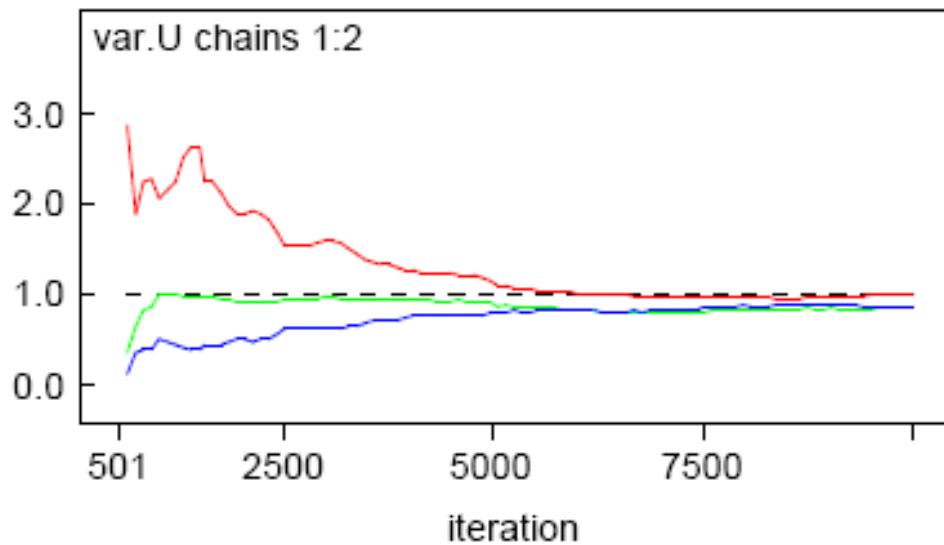
Gelman-Rubin statistic

- Generate multiple chains starting at *over-dispersed* initial values.
- Denote the number of chains generated by M and the length of each chain by $2T$.
- Take as a measure of posterior variability the width of the $100(1-a)\%$ credible interval for the parameter of interest (in WinBUGS, $a = 0.2$).
- From the final T iterations we calculate the empirical credible interval for each chain. We then calculate the average width of the intervals across the M chains and denote this by $\textcolor{blue}{W}$
- Calculate the width of the empirical credible interval based on all MT samples pooled together, $\textcolor{green}{B}$.

The ratio of pooled to average interval widths $R = B / W$ should be greater than 1 if the starting values are suitably overdispersed;

It will also tend to 1 as convergence is approached.

and so we might assume convergence for practical purposes if $R < 1.05$, say.



Inference using posterior samples from MCMC runs

- Once you have run enough updates to obtain an appropriate samples from the posterior distribution, you may summarize these samples.
- Typically report either mean or median of the posterior samples for each parameter of interest as a point estimate
- 2.5% and 97.5% percentiles of the posterior samples for each parameter give a 95 % posterior credible interval

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
alpha0	-0.5553	0.1904	0.003374	-0.9365	-0.5563	-0.1763	2001	20000
alpha1	0.08693	0.3123	0.006873	-0.5502	0.09157	0.6964	2001	20000
alpha12	-0.8358	0.4388	0.01105	-1.731	-0.8253	-0.002591	2001	20000
alpha2	1.359	0.2744	0.005812	0.8204	1.354	1.923	2001	20000
sigma	0.2855	0.146	0.005461	0.04489	0.2752	0.6132	2001	20000

Generalized linear mixed models mixed via Fully Bayesian

Generalized linear mixed model, the mean conditioned to the random effects, μ^b , of the outcome variable, Y, is related to the explanatory variables by the link function , $g(\cdot)$, :

$$g(\mu^b) = \eta^b = X\beta + Zb$$

where X is design matrix of the fixed effects,
 β is the parameters of the fixed effects,
Z is design the matrix of the random effects
and b is a vector of dimension n that contains the random effects

$$b \sim NMV(0, \Sigma)$$

This joint distribution of the random effects is expresses as

$$f[b | \delta] \longrightarrow \text{It allowed to define spatial correlation}$$

And the conditional distribution of each random effect denotes

$$f[b_i | b_j, j \neq i, \delta]$$

In this model the unknown parameters are β , b , and δ , to which will assign them a priori distribution,

$$f[\beta], f[b | \delta], f[\delta]$$

Based to the beliefs of the researcher.

The **joint posterior distribution** of the unknown parameters is obtained by theorem of Bayes :

$$f[b, \delta, \beta | Y] \propto f[Y | b, \delta, \beta] \cdot f[b | \delta] \cdot f[\delta] \cdot f[\beta]$$

And the **marginal posterior distributions** of the parameters are defined

$$f[\beta | Y] = \int_{\beta} \int_{b_1} \int_{b_2} \dots \int_{b_N} \int_{\delta} f[b, \delta, \beta | Y] d\beta db_1 db_2 \dots db_N d\delta$$

$$f[b_i | Y] = \int_{\beta} \int_{b_1} \dots \int_{b_{i-1}} \int_{b_{i+1}} \dots \int_{b_N} \int_{\delta} f[b, \delta, \beta | Y] d\beta db_1 db_{i-1} db_{i+1} \dots db_N d\delta$$

$$f[\delta | Y] = \int_{\beta} \int_{b_1} \int_{b_2} \dots \int_{b_N} f[b, \delta, \beta | Y] d\beta db_1 db_2 \dots db_N$$

To avoid to resolve analytically these integrals, simulation by Chains of Markov Monte Carlo, (MCMC) (Gilks, Richardson and Spiegelhalter, 1996).

Conditional distributions are:

$$f[\beta | b, \delta, Y] \propto f[\beta] \cdot \prod_{i=1}^N l[Y_i; \beta, \delta]$$

$$f[b_i | b_j, j \neq i, \delta, \beta, Y] \propto f[b_i | b_j, j \neq i] \cdot l[Y_i; \beta, \delta]$$

$$f[\delta | b, \beta, Y] \propto f[\delta] \cdot f[b | \delta]$$

Gibbs sampling process:

1) Choose initial values for the parameters of interest:

$$\delta^{(0)}, b_1^{(0)}, \dots, b_N^{(0)}, \beta^{(0)}$$

2) Sample a new value of $\delta^{(1)}$ from his conditioned distribution. The conditioned parameters are substituted by the initial values:

$$f[\delta | b^{(0)}, \beta^{(0)}, Y] \propto f[\delta] \cdot f[b^{(0)} | \delta]$$

3) Sample a new value of the first random effect, $b_1^{(1)}$

$$f[b_1 | b_j^{(0)} \ j \neq 1, \delta^{(0)}, \beta^{(0)}, Y] \propto f[b_1 | b_j^{(0)} \ j \neq 1] \cdot l[Y_i; b^{(0)}, \beta, \delta^{(0)}]$$

Until the random effect n-th $b_N^{(1)}$

$$f[b_N | b_j^{(0)} \ j \neq N, \delta^{(0)}, \beta^{(0)}, Y] \propto f[b_N | b_j^{(0)} \ j \neq N] \cdot l[Y_i; b^{(0)}, \beta, \delta^{(0)}]$$

4) Finally, sample a new value of the fixed effect β from:

$$f[\beta | b^{(0)}, \delta^{(0)}, Y] \propto f[\beta] \cdot \prod_{i=1}^N l[Y_i; \beta, b^{(0)}, \delta^{(0)}]$$

Bibliography

Books

Banerjee S Carlin BP, Gelfrand To.E. (2004) Hierarchical Modelling and Analysis for Spatial Dates. Chapman & Hall /CRC. **Chapter 5**

Mollie To. (1999). Bayesian and Empirical Bayes Approaches Tone disease mapping in Disease Mapping and Risk Assessment for Public Health edited by Lawson

Example: Disease mapping

BUGS code

```
model {  
  for(i in 1 : I) {  
    O[i] ~ dpois(mu[i])  
    log(mu[i]) <- log(E[i]) + alpha + theta[i]  
    theta[i] ~ dnorm(0, tau) # area-specific random effects  
  
    R[i] <- exp(alpha + theta[i]) # area-specific relative risk  
  
    SMR[i] <- O[i] / E[i] # Observed SMRs - this is just a function of  
                           # the data, not of stochastic parameters, but  
                           # is included in the model code so that values  
                           # of SMR can be inspected and mapped  
  }  
  # Priors:  
  alpha ~ dflat() # uniform prior on overall intercept  
  
  tau ~ dgamma(0.001, 0.001) # prior on precision of area random effects  
  sigma <- 1/sqrt(tau) # between-area sd of random effects  
}
```

DADES EN RETICULES

Introduction to
WinBUGS and GeoBUGS

Dra. ROSA ABELLANA

DEPARTAMENT SALUT PÚBLICA. UNIVERSITAT DE
BARCELONA

WinBUGS software Bayesian inference using Gibbs Sampling

- Language for specifying complex Bayesian models
- Constructs object-oriented internal representation of the model
- Simulation from full conditionals using Gibbs sampling
- Current version (WinBUGS 1.4) runs in Windows
- R2WinBUGS :A Package for Running WinBUGS from R
- WinBUGS is freely available from:

<https://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/>

Installing WinBUGS 1.4 in Windows

One way to do this is as follows:

1. Copy WinBUGS14.exe to your computer
2. Go into Explore and double click on WinBUGS14.exe
3. Follow the instructions in the dialog box
4. You should have a new directory called WinBUGS14 within Program Files
5. Inside the WinBUGS14 directory is a program called WinBUGS14.exe
6. Right-click on the pretty WinBUGS icon, select 'create shortcut', then drag this shortcut to the desktop.
7. Double click on WinBUGS14.exe to run WinBUGS14..

Another way to do this is as follows:

1. Unzip the folder winbugs143_unrestricted.zip in a new folder in your PC.
2. Go to the step 5.

Obtaining the key for unrestricted use

Download the key for unrestricted use. In the same web page as Winbugs installation.

After following the instructions given in the key, check that the **keys.ocf** file in ..\WinBUGS14\Bugs\Code\ has been updated.

WinBUGS formats of data

Type of formats :

1. Rectangular

n[] r[]

47 0

148 18

...

360 24

END

2. R package:

```
list(N=12,n = c(47,148,119,810,211,196,  
148,215,207,97,256,360),  
r = c(0,18,8,46,8,13,9,31,14,8,29,24))
```

Function BUGS

- $p <- \text{equals}(x,.7) = 1$ si $x = 0.7$, 0 otherwise.
 - $\tau <- 1/\text{pow}(s,2)$ compute= $1/s^2$.
 - $s <- 1/\sqrt{\tau}$ compute $s = 1/\tau^{1/2}$.
-
- $p <- \text{step}(x-0.7) = 1 \quad \text{if } x \geq 0.7; \quad 0 \text{ otherwise.}$
- E.g: Average the samples of p , obtain the probability that $x \geq 0.7$
-
- Look 'Model Specification/Logical nodes' of the winBUGS manual for all the functions.

Most common distributions

Expression	Distribution	Us
dbin	binomial	$r \sim \text{dbin}(p,n)$
dnorm	normal	$x \sim \text{dnorm}(\mu,\tau)$
dpois	Poisson	$r \sim \text{dpois}(\lambda)$
dunif	uniform	$x \sim \text{dunif}(a,b)$
dgamma	gamma	$x \sim \text{dgamma}(a,b)$

Normal distribution is parameterized by the mean and the precision (inverse of variance) = $1/\text{sd}^2$.

Look 'Model Specification/The BUGS language: stochastic nodes/Distributions'

Some aspects of language BUGS

- \sim represents stochastic dependence, e.g. $r \sim \text{dunif}(a,b)$

- Can use arrays and loops

```
for (i in 1:n){  
  r[i] ~ dbin(p[i],n[i])  
  p[i] ~ dunif(0,1)  
}
```

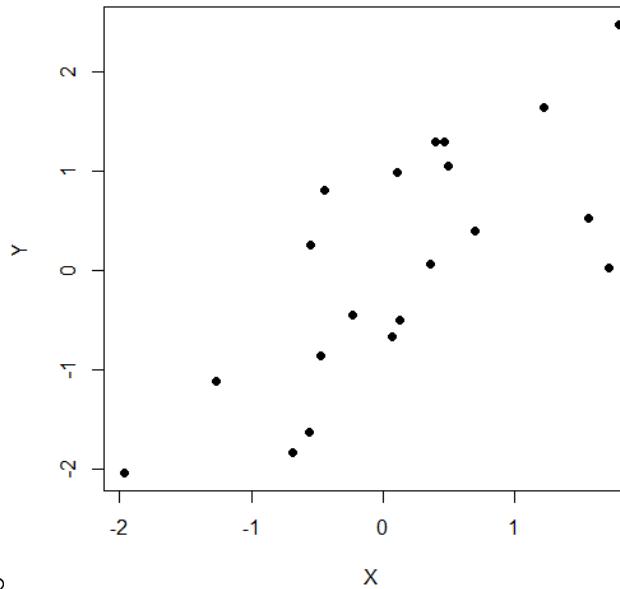
- Some functions can appear on left-hand-side of an expression, e.g.

```
logit(p[i])<- a + b*x[i]  
log(m[i]) <- c + d*y[i]
```

- $\text{mean}(p[])$ to take mean of whole array, $\text{mean}(p[m:n])$ to take mean of elements m to n. Also for $\text{sum}(p[])$.

Simple linear regression in WinBUGS

We observed the data:



Call:

```
lm(formula = Y ~ X)
```

Residuals:

Min	1Q	Median	3Q	
-1.5123	-0.6019	-0.1857	0.8309	1.2

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.04017	0.19197	-0.209	0.836588
X	0.92174	0.20027	4.603	0.000221 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8491 on 18 degrees of freedom

Multiple R-squared: 0.5406, Adjusted R-squared: 0.5151

F-statistic: 21.18 on 1 and 18 DF, p-value: 0.000221

WinBUGS:

#The model

```
model{
  for(i in 1:n){
    y[i]~dnorm(mu[i],tau)  #likelihood distribution
    mu[i] <- alpha + beta*x[i]
  }
  #prior distributions
  tau~dgamma(0.01,0.01)
  alpha~dnorm(0,0.01)
  beta~dnorm(0,0.01)
}
```

#The data

```
list(n=20,
X=c(-0.5605,-0.2302,1.5587,0.0705,0.1293,1.7151,0.4609,-1.2651,-0.6869,-
0.4457,1.2241,0.3598,0.4008,0.1107,-0.5558,1.7869,0.4979,-1.9666,0.7014,-
0.4728),
Y=c( -1.6283,-0.4482,0.5327,-0.6584,-0.4958,0.0284,1.2987,-1.1117,-
1.825,0.8082,1.6505,0.0647,1.2959,0.9888,0.2657,2.4756,1.0518,-
2.0285,0.3954,-0.8533))
```

#The initial values

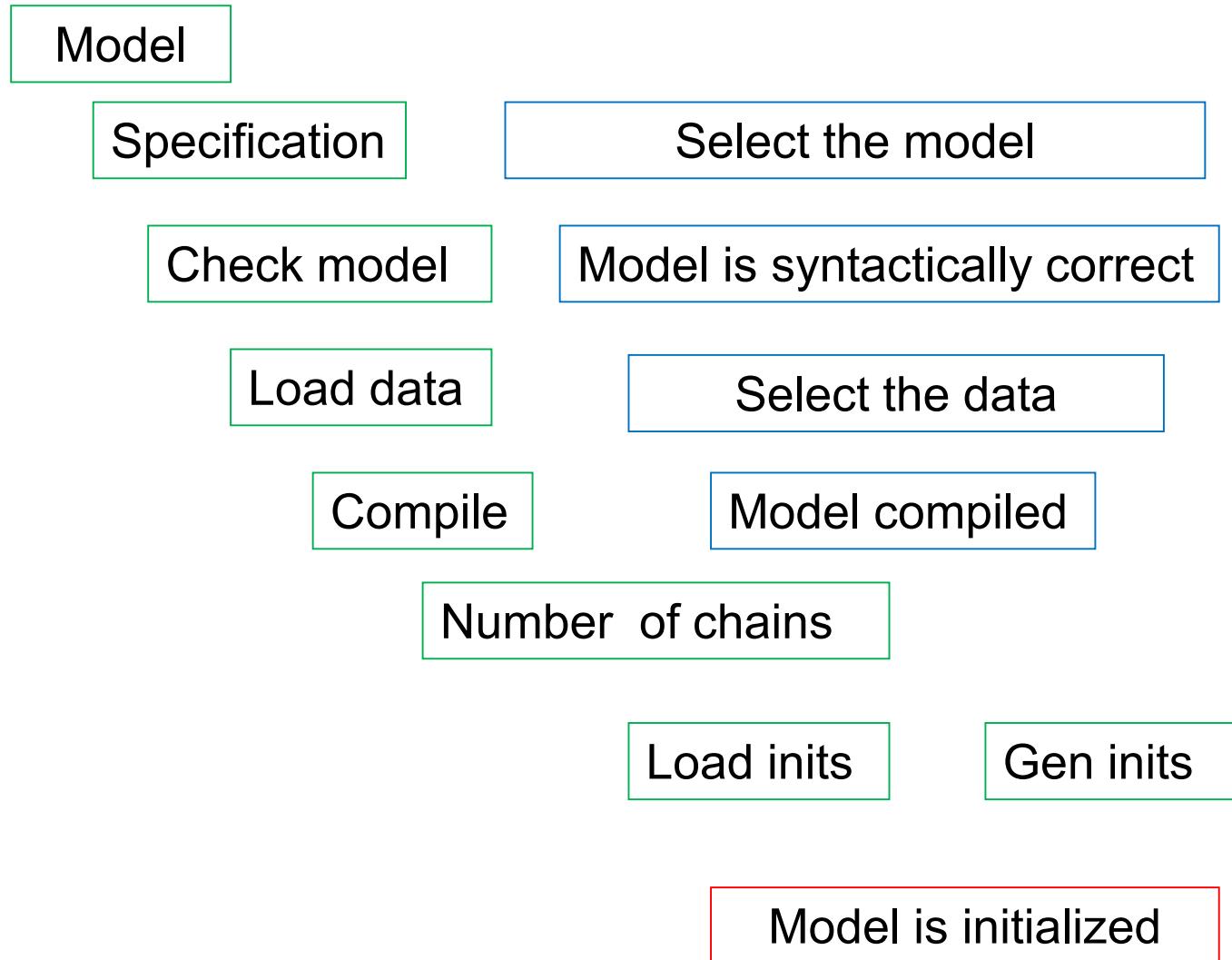
Chain 1:

```
list(alpha=1, beta=0, tau=0.01)
```

Chain 2:

```
list(alpha=-1,beta=-2,tau=0.8)
```

1. Running WinBUGS:



2. For each variable you'd like to analyze, go to
“inference/sample”

Enter the name of the variable in the “nodes” box,
and click “set”. Samples of variables you don’t
request to record are automatically discarded.

3. Check the convergence of the chains

Menu Inference/Trace plot

Menu Inference/bgr diag

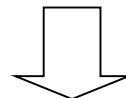
Menu inference/Autocorrelation function

4. Goodness of fit, when the samples had already converge

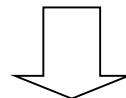
Menu Inference/DIC

Sampling of Gibbs → We stroll for the sample space

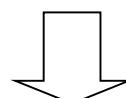
First samples are discarded



Burning process



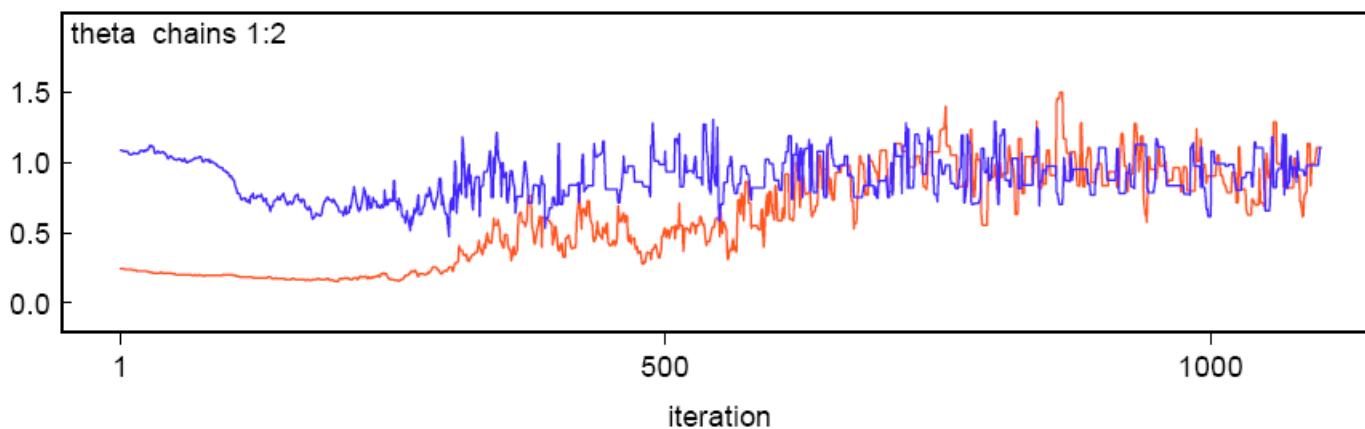
Check if the chains came from posterior distribution



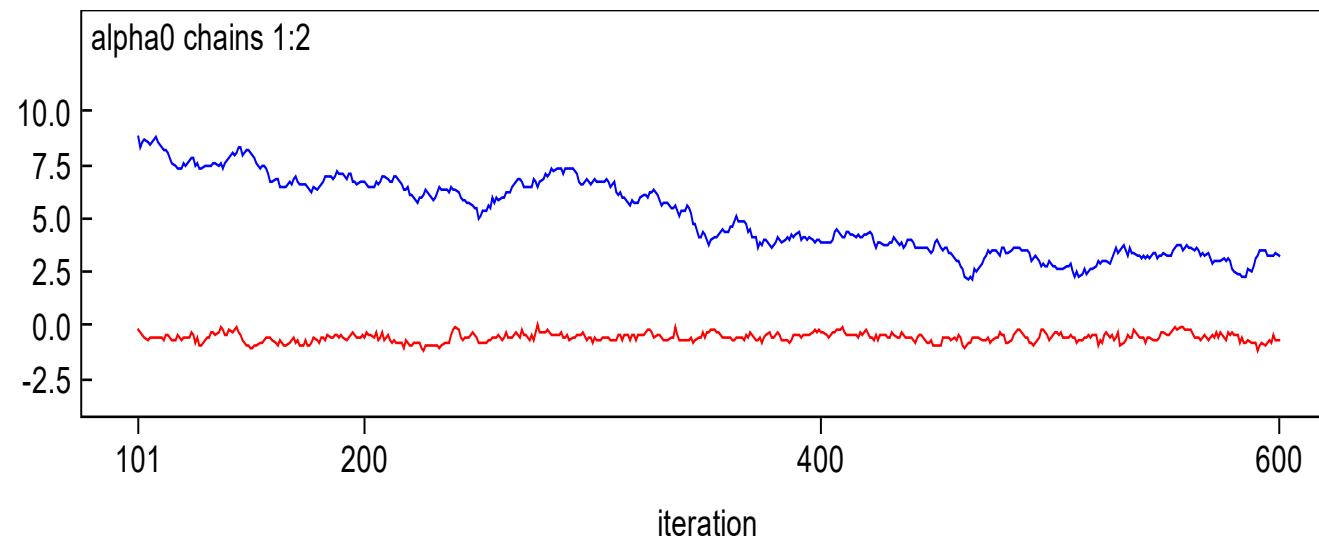
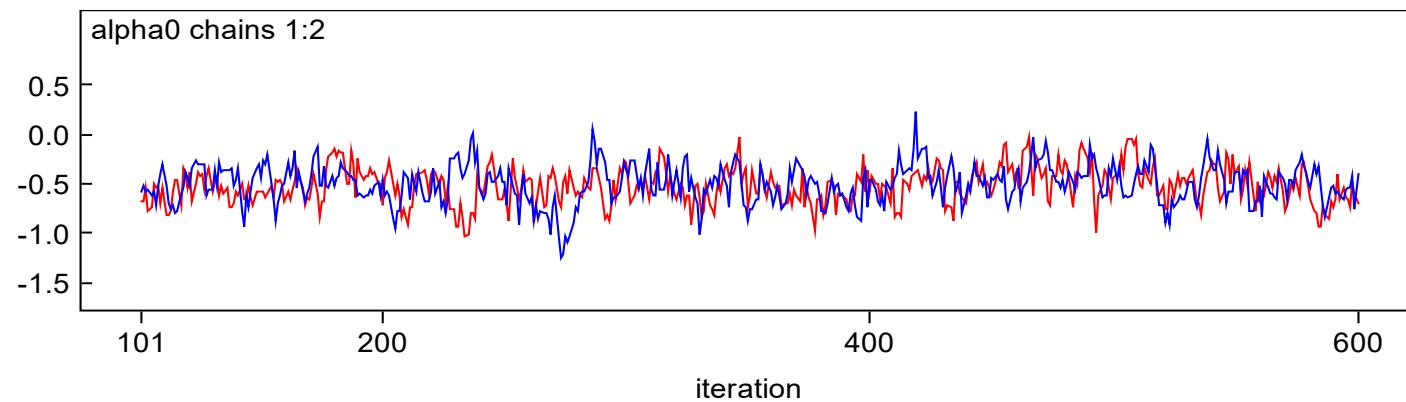
Convergence to posterior distribution.

Checking convergence

- Convergence is to a distribution not to a single value.
- Once convergence reached, samples should look like a random scatter about a stable mean value.
- Practice, is to *run* multiple chains for the same model but with different initial points, and visually inspect “trace” plot for convergence



Example convergence and no convergence



Statistician of Gelman-Rubin

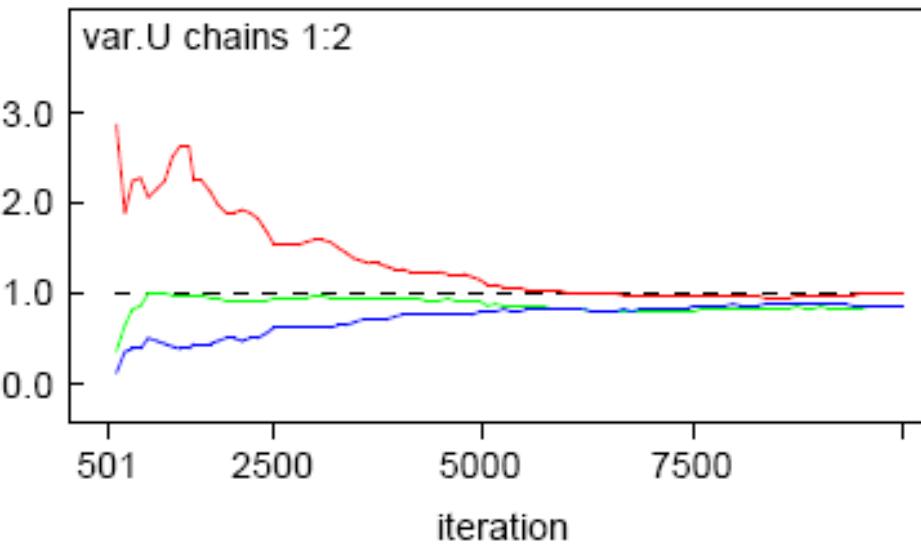
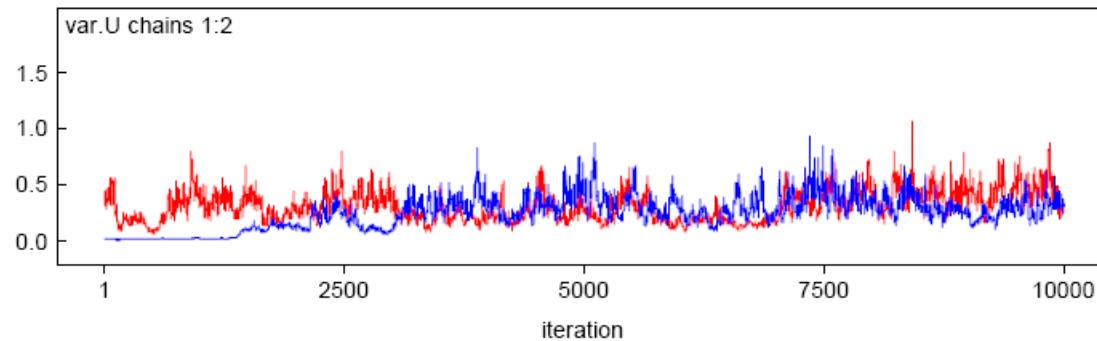
Formally, can use the Gelman-Rubin statistician based in the ratio of the *between* and *within* chains variance (ANOVA)

Gelman: the best to identify the no convergence, starting from different initial values,

Intuitively, the behavior of all the chains would have to be the same.

Or, the variance within the chain would have to **be the same** that the variance along the chain

Example of Gelman and Rubin with the Winbugs



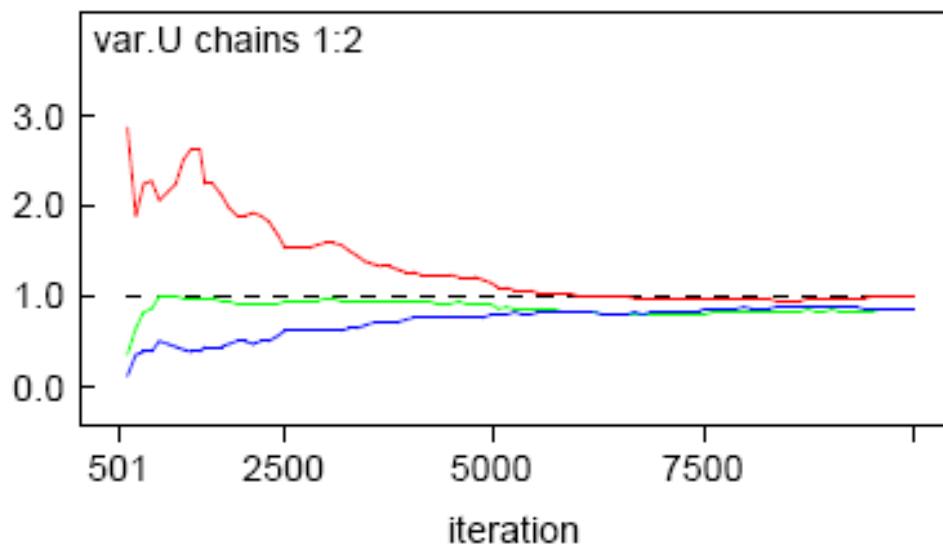
B pooled

W average of width

Gelman-Rubin statistic

- Generate multiple chains starting at *over-dispersed* initial values.
- Denote the number of chains generated by M and the length of each chain by $2T$.
- Take as a measure of posterior variability the width of the $100(1 - \alpha)\%$ credible interval for the parameter of interest (in WinBUGS, $\alpha = 0.2$).
- From the final T iterations we calculate the empirical credible interval for each chain. We then calculate the average width of the intervals across the M chains and denote this by W
- Calculate the width B of the empirical credible interval based on all MT samples pooled together.

The ratio $R = B / W$ of pooled to average interval widths should be greater than 1 if the starting values are suitably overdispersed; it will also tend to 1 as convergence is approached, and so we might assume convergence for practical purposes if $R < 1.05$, say.



GeoBUGS

GeoBUGS is an add-on module to WinBUGS which provides an interface for:

- producing maps of the output from disease mapping and other spatial models
- * creating and manipulating adjacency matrices that are required as input for the conditional autoregressive models available in WinBUGS 1.4 for carrying out spatial smoothing.

Version 1.2 of GeoBUGS contains map files for

- * Districts in Scotland (called ***Scotland***)
- * Wards in a London Health Authority (called ***London_HA***)
- * Counties in Great Britain (called ***GB_Counties***)
- * Departements in France (called ***France***)
- * Nomoi in Greece (called ***Greecenomoi***)
- * Districts in Belgium (called ***Belgium***)
- * Communes in Sardinia (called ***Sardinia***)
- * Subquarters in Munich (called ***Munich***)
- * A 15 x 15 regular grid (called ***Elevation***)
- * Wards in West Yorkshire (UK) (called ***WestYorkshire***)
- * A 4 x 4 regular grid (called ***Forest***)
- * A grid of 750 m² grid cells covering the town of Huddersfield and surroundings in northern England (called ***Huddersfield_750m_grid***)

A list of the area IDs for each map and the order in which the areas are stored in the map file can be obtained using the **export** Splus command.

GeoBUGS 1.2 also has facilities for importing user-defined maps reading polygon formats from Splus, ArclInfo and Epimap, plus a link to a program written by Yue Cui for importing ArcView shape files.

<http://www.biostat.umn.edu/~yuecui/>

Export S-plus map to Winbugs

sp2WB (library maptools)

The function exports an sp SpatialPolygons object into a S-Plus map format to be import by WinBUGS.

In Winbugs

Open the polygon file as a separate text file in WinBUGS 1. and select the appropriate **import** option from the ***Map menu***.

File with Code Rstudio: ExportImportSpatialPolygonsWinBUGS.R

DADES EN RETICULES

Generalized linear models and
Generalized mixed linear models

Dra. ROSA ABELLANA

DEPARTAMENT SALUT PÚBLICA. UNIVERSITAT DE
BARCELONA

The rates of lip cancer in 56 counties in Scotland have been analysed by Clayton and Kaldor (1987) and Breslow and Clayton (1993). The form of the data includes the observed and expected cases (expected numbers based on the population and its age and sex distribution in the county), a covariate measuring the percentage of the population engaged in agriculture, fishing, or forestry, and the "position" of each county expressed as a list of adjacent counties.

County	Observed cases O_i	Expected cases E_i	Percentage in agric. x_i	SMR	Adjacent counties
1	9	1.4	16	652.2	5,9,11,19
2	39	8.7	16	450.3	7,10
...
56	0	1.8	10	0.0	18,24,30,33,45,55

Poisson regression using R

```
> summary(results.lips)
```

Call:

```
glm(formula = O ~ X, family = poisson, data = data.lips, offset =  
log(E))
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.54227	0.06952	-7.80	6.21e-15	***
X	0.73732	0.05956	12.38	< 2e-16	***

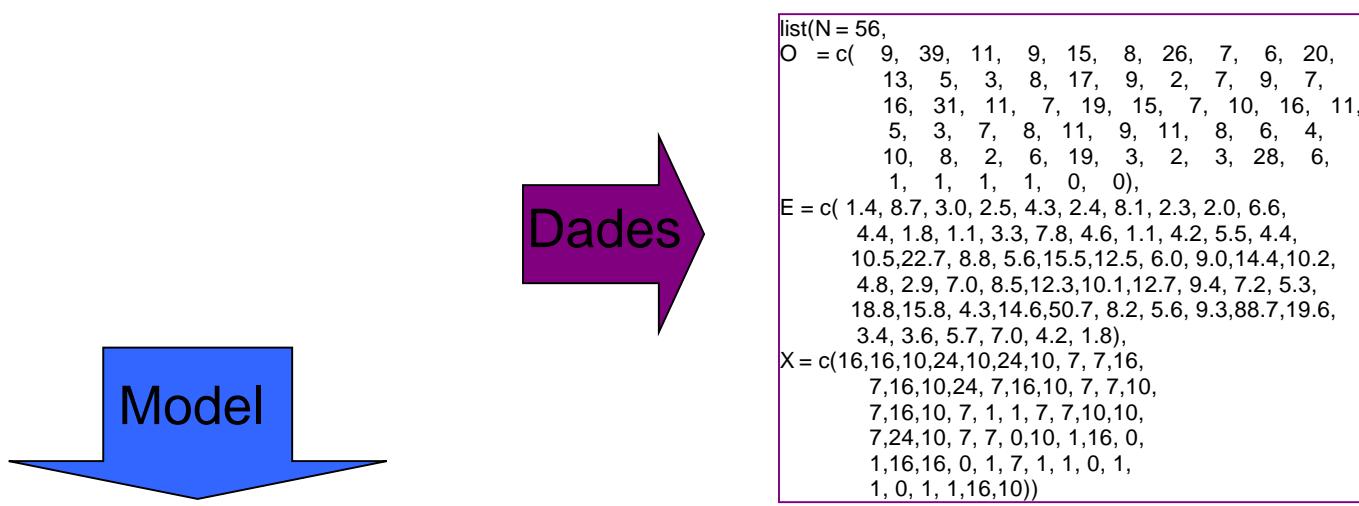
Signif. codes: 0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 380.73 on 55 degrees of freedom
Residual deviance: 238.62 on 54 degrees of freedom
AIC: 450.6

Number of Fisher Scoring iterations: 5

```
> deviance(results.lips) /54  
[1] 4.418903
```



```

model {

# Likelihood
for (i in 1 : N) {
  O[i] ~ dpois(mu[i])
  log(mu[i]) <- log(E[i]) + alpha0 + alpha1 * X[i]/10 + b[i]
  RR[i] <- exp(alpha0 + alpha1 * X[i]/10 + b[i])      }
# CAR prior distribution for random effects:
b[1:N] ~ car.normal(adj[], weights[], num[], tau)
for(k in 1:sumNumNeigh) {
  weights[k] <- 1
}

# Other priors:
alpha0 ~ dflat()
alpha1 ~ dnorm(0.0, 1.0E-5)
tau ~ dgamma(0.5, 0.0005) # prior on precision

sigma <- sqrt(1 / tau) # standard deviation
}

```



```

list(tau = 1, alpha0 = 0, alpha1 = 0,
b=c(0,0,0,0,NA,0,NA,0,0,
NA,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0))

```

The inits can be generate automatically

```
> scotland.moran.mc
```

Monte-Carlo simulation of Moran I

```
data: nc_scotland$residuals  
weights: w.scotland  
number of simulations + 1: 101
```

```
statistic = 0.33094, observed rank = 101, p-value = 0.009901  
alternative hypothesis: greater
```

Model Lips data

Generalized linear model

$$Y_i \sim Poisson(\mu_i)$$

$$\log(\mu_i) = \log(E) + \alpha_0 + \alpha_1 \cdot X$$

Code Winbugs Poisson model

```
model{
# Likelihood
  for (i in 1 : N) {
    O[i] ~ dpois(mu[i])
    log(mu[i]) <- log(E[i]) + alpha0 + alpha1 * X[i]/10
    RR[i] <- exp(alpha0 + alpha1 * X[i]/10 )      # Area-specific relative risk (for maps)
    SMR[i]<-O[i]/E[i]                            #Observed SMR
    pp.RR[i]<-step(RR[i]-1)
  }
# Priors:
  alpha0 ~ dflat()
  alpha1 ~ dnorm(0.0, 1.0E-5)
```

Files Required:

escocia-Poisson-model.odc
escocia-data.odc
escocia-poisson-inits1_2_3.odc

Examples of initial values for the parameters of the Poisson model

```
list(alpha0 = 0, alpha1 = 0)  
list(alpha0 = -0.5, alpha1 = -2)  
list(t alpha0 = 1, alpha1 = 0.5)
```

How many iterations after convergence?

- After convergence, further iterations are needed to obtain samples for posterior distribution
- More iterations== more accurate posterior estimates
- Efficiency of sample mean of parameters as estimate of theoretical posterior expectation usually assessed by calculating Monte Carlo Standard error (MC error)
- MCerror==standard error of posterior mean
- MC error depends on:
 - True variance of posterior distribution
 - Posterior sample size (number of MCMC iterations)
 - Autocorrelation in MCMC sample
- Rule of thumb: **want MC error <1% of posterior SD**

Model Lips data

Generalized linear model

$$Y_i \sim Poisson(\mu_i)$$

$$\log(\mu_i) = \log(E) + \alpha_0 + \alpha_1 \cdot X$$

Generalized linear mixed model

$$g(\mu^b) = \eta^b = X\beta + Zb$$

$$\log(\mu|b) = \log(E) + \alpha_0 + \alpha_1 \cdot X + b \cdot Z$$

$$b \sim NMV(0, \Sigma(\delta))$$

Heterogeneity model

Without spatial structure, assumes no spatial correlation

$$b \sim NMV(0, \sigma_{het} I)$$

Code Winbugs for heterogeneity model

```
model {  
for (i in 1 : l) {  
O[i] ~ dpois(mu[i])  
log(mu[i]) <- log(E[i]) + alpha0 + alpha1 * X[i]/10 + h[i]  
RR[i] <- exp(alpha0 + alpha1 * X[i]/10 + h[i]) # Area-specific adjusted relative risk  
  
h[i] ~ dnorm(0, tau.h) # unstructured random effects  
resid[i]<-exp(h[i])  
pp.resid[i]<-step(resid[i]-1)  
}  
  
# Other priors:  
alpha0 ~ dflat() # flat uniform prior on overall intercept  
alpha1 ~ dnorm(0.0, 1.0E-5)  
tau.h~gamma(0.001, 0.001)  
sigma.h<-1/sqrt(tau.h)  
}
```

Files Required:

- escocia.heterogeneity-model.odc
- escocia-data.odc
- escocia-heterogeneity-init1.odc
- escocia-heterogeneity-init2.odc
- escocia-heterogeneity-init3.odc

Examples of initial values for the parameters of the heterogeneity model

```
list(tau.h = 1, alpha0 = 0, alpha1 = 0,  
h=c(0,0,0,0,0,0,0,0,0,  
    0,0,0,0,0,0,0,0,0,  
    0,0,0,0,0,0,0,0,0,  
    0,0,0,0,0,0,0,0,0,  
    0,0,0,0,0,0,0,0,0,  
    0,0,0,0,0,0,0,0,0))
```

```
list(tau.h = 0.005, alpha0 = -0.5, alpha1 = -2,  
h=c(0,0,0,0,0,0,0,0,0,  
    0,0,0,0,0,0,0,0,0,  
    0,0,0,0,0,0,0,0,0,  
    0,0,0,0,0,0,0,0,0,  
    0,0,0,0,0,0,0,0,0,  
    0,0,0,0,0,0,0,0,0))
```

Generalized linear mixed model

$$g(\mu^b) = \eta^b = X\beta + Zb$$

$$\log(\mu|b) = \log(E) + \alpha_0 + \alpha_1 \cdot X + b \cdot Z$$

$$b \sim NMV(0, \Sigma(\delta))$$

Spatial structured model

$$Intrinsic\ CAR\ model\ \rho=1\quad b \sim N(0, \tau_s^2(D_w - W)^{-1})$$

The conditional distribution of the random effects is a Normal

$$b_i | b_{j \in \delta_i} \sim N\left(\bar{b}_i, \sqrt{\frac{\sigma_s^2}{n_i}}\right)$$

Gaussian Conditional Autoregression (CAR) models

Between-area covariance matrix:

$$v\mathbf{B} = v(\mathbf{I} - \gamma \mathbf{C})^{-1} \mathbf{M}$$

where

\mathbf{I} = N x N identity matrix

\mathbf{M} = N x N diagonal matrix, with elements M_{ii} proportional to the conditional variance of $S_i | S_j$

\mathbf{C} = N x N weight matrix, with elements C_{ij} reflecting spatial association between areas i and j

γ = controls overall strength of spatial dependence

Intrinsic CAR model in WinBUGS

Intrinsic CAR model corresponds to choosing:

- $C_{ij} = w_i/w_{i+}$ if areas i and j are adjacent and $C_{ij} = 0$ otherwise
- $M_{ii} = 1/w_{i+}$,
- $\gamma = \gamma_{\max}$ which turns out to always be 1

$$w_{ij} = \begin{cases} 1 & \text{if region } i \text{ and } j \text{ are neighbors} \\ 0 & \text{otherwise} \end{cases}$$

In WinBUGS is defined by:

```
car.normal(adj[],weights[],num[],tau)
```

Arguments ***car.normal()*** distribució

adj[]: A vector listing the ID numbers of the adjacent areas for each area

```
adj = c(  
 19, 9, 5,  
 10, 7,  
 12,  
 28, 20, 18,  
 19, 12, 1,  
  
 17, 16, 13, 10, 2,  
 .....
```

weights[]: A vector the same length as **adj[]** giving *unnormalised* weights associated with each pair of areas.

num[]: A vector of length N (the total number of areas) giving the number of neighbours n_i for each area.

```
num = c(3, 2, 1, 3, 3, 0, 5, 0, 5, 4,  
0, 2, 3, 3, 2, 6, 6, 6, 5, 3,.....
```

tau: A scalar argument representing the precision (inverse variance) parameter of the Gaussian CAR prior

num i adj can be created with the option **adj matrix** of GeoBUGS **Adjacency Tool**.

Weights:

For (j in 1:**sumNum Nneigh**){weights[]<-1}

On **sumNumNneigh**: Total number of neighbours

Code WinBUGS for spatial model

```
model
{
  for (i in 1 : N) {
    O[i] ~ dpois(mu[i])
    log(mu[i]) <- log(E[i]) + alpha0 + alpha1 * X[i]/10 + b[i]
    RR[i] <- exp(alpha0 + alpha1 * X[i]/10 + b[i])
      # Area-specific relative risk (for maps)
  }
}
```

```
# CAR prior distribution for random effects:
b[1:N] ~ car.normal(adj[], weights[], num[], tau.b)
for(k in 1:sumNumNeigh) {
  weights[k] <- 1
}
```

```
# Other priors:
alpha0 ~ dflat()
alpha1 ~ dnorm(0.0, 1.0E-5)
tau.b ~ dgamma(0.5, 0.0005)
sigma.b <- sqrt(1 / tau.b)
}
```

Files Required:

escocia.spatial-model.odc
escocia-data.odc
escocia-data-adjunts.odc
escocia-spatial-inits1.odc
escocia-spatial-inits2.odc
escocia-spatial-inits3.odc

car.proper

The proper Gaussian CAR prior distribution is specified using the distribution ***car.proper*** for the *vector* of random variables $\mathbf{S} = (S_1, \dots, S_N)$. The syntax for this distributions is as follows:

```
S[1:N] ~ car.proper(mu[], C[], adj[], num[], M[], tau, gamma)
```

where:

mu[]: A vector giving the mean for each area (this can either be entered as data, assigned a prior distribution, or specified deterministically within the model code).

C[]: A vector the same length as ***adj[]*** giving *normalised* weights associated with each pair of areas

adj[]: A vector listing the ID numbers of the adjacent areas for each area

num[] : A vector of length N (the total number of areas) giving the number of neighbours n_i for each area.

M[] : A vector of length N giving the diagonal elements M_{ii} of the conditional variance matrix

tau : A scalar parameter representing the overall precision (inverse variance) parameter.

gamma : A scalar parameter representing the overall degree of spatial dependence. This parameter is constrained to lie between bounds given by the inverse of the minimum and maximum eigenvalues of the matrix $\mathbf{M}^{-1/2} \mathbf{C} \mathbf{M}^{-1/2}$ (see appendix). *GeoBUGS 1.1Beta* contains two deterministic functions for calculating these bounds:

```
min.bound(C[], adj[], num[], M[])
max.bound(C[], adj[], num[], M[])
```

where the arguments are the same as the corresponding vectors used as arguments to the **car.proper** distribution.

Convolution model

To allow greater flexibility, Besag, York and Mollie (1991) recommend combining the intrinsic CAR prior and an exchangeable normal prior:

$$\begin{aligned}\theta_i &= S_i + H_i \\ H_i &\sim \text{Normal}(0, v^2) \\ S_i | S_{-i} &\sim \text{Normal}(m_i, s_i^2) \\ \text{where } m_i &= \frac{\sum_j w_{ij} S_j}{w_i+}; \quad s_i^2 = \frac{s^2}{w_i+}\end{aligned}$$

- Total variation in reflects a combination of spatial dependence and unstructured heterogeneity
 - The data determine the relative contribution of each component

Inference about the contribution of the spatial and heterogeneity effect in the convolution model

υ^2 (unstructured variance) i s^2 (spatial variance) are not directly comparable :

- υ^2 reflects the marginal variability of the unstructured random effect
- s^2/n_i reflects the conditional variance of the region random effect and conditioned on the values of the random effects of its neighbors

Handicap: There is no closed expression of marginal variance between regions

Solution: Estimate the marginal variance of the random effects empirically

$$s^2_{\text{marginal}} = \sum_i (S_i - \bar{S})^2 / (I - 1)$$

Relative contribution of spatial with respect to nonstructure

$$\text{frac}_{\text{spatial}} = s_{\text{marginal}}^2 / (s_{\text{marginal}}^2 + v^2)$$

$\text{frac}_{\text{spatial}} \rightarrow 1$ spatial dominates heterogeneity

$\text{frac}_{\text{spatial}} \rightarrow 0$ unstructured heterogeneity dominates the spatial

Codi WinBUGS for convolution model

```
model {  
# Likelihood  
for (i in 1 : N) {  
    O[i] ~ dpois(mu[i])  
    log(mu[i]) <- log(E[i]) + alpha0 + alpha1 * X[i]/10 + s[i]+ h[i]  
    RR[i] <- exp(alpha0 + alpha1 * X[i]/10 + s[i] + h[i]) # Area-specific relative risk (for maps)  
    SMR[i]<-O[i]/E[i] #Observed SMR  
    resid[i]<-exp(s[i]+h[i])  
    pp.resid[i]<-step(resid[i]-1)  
# Exchangeable prior on unstructured random effects  
    h[i] ~ dnorm(0, tau.h)  
}  
# CAR prior distribution for random effects:  
s[1:N] ~ car.normal(adj[], weights[], num[], tau.s)  
for(k in 1:sumNumNeigh) {  
    weights[k] <- 1  
}  
s2.marginal<-sd(s[])*sd(s[]) #empirical marginal var. of spatial effects  
frac.spatial <- s2.marginal / (s2.marginal + sigma2.h)  
}
```

Other priors:

alpha0 ~ dflat()

alpha1 ~ dnorm(0.0, 1.0E-5)

tau.s ~ dgamma(0.5, 0.0005)

prior on precision

sigma.s <- sqrt(1 / tau.s)

standard deviation

tau.h ~ dgamma(0.001, 0.001)

prior on precision

sigma.h <- sqrt(1 / tau.h)

standard deviation

sigma2.h<-1/tau.h

Files Required:

escocia.convolution-model.odc

escocia-data.odc

escocia-data-adjunts.odc

escocia-convolution-inits1.odc

escocia-convolution-inits2.odc

escocia-convolution-inits3.odc

Bayesian models: Deviance information criterion (DIC)

Spielgelhalter et al (2002), proposed to compare models:

Deviance Information Criterion DIC= “Goodness of fit” +complexity

Measure goodness of fit using deviance:

$$D(\theta) = -2\log L(Y | \theta)$$

Measure complexity of the model using the effective number of parameters

$$\begin{aligned} p_D &= E_{\theta|Y}[D] - D(E_{\theta|Y}[\theta]) \\ &= \bar{D} - D(\bar{\theta}) \end{aligned}$$

DIC is defined as AIC :

$$DIC = \bar{D} + p_D$$

Lower DIC is better

Spatial Point Patterns

Introduction

Josep Lluís Carrasco

Departament de Fonaments Clínics
Universitat de Barcelona

Spatial Data Analysis

- Approaches depends on
 - How spatial data is collected
 - Underlying spatial process that generates data
- Based on that it is possible to identify three types of analysis:
 - Geostatistics
 - Lattice data
 - Spatial point patterns

Geostatistics

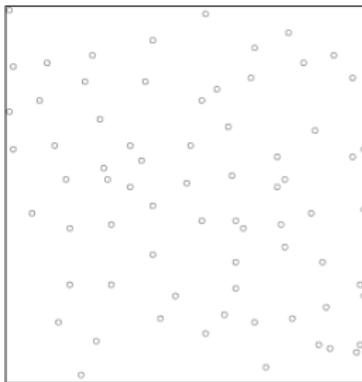
- Measurements at a limited number of observation locations (sample)
- Values in the non-sampled locations are interpolated
- In epidemiology geostatistics are usually applied to model exposure factors
- An example: Ambient Air Pollution and Atherosclerosis in Los Angeles <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1277865/pdf/ehp0113-000201.pdf>
- Outcome: Carotid artery intima-media thickness
- Exposure factor: Concentrations of $PM_{2.5}$ from 23 monitoring stations.
- $PM_{2.5}$ were extrapolated to the whole area by universal kriging

Lattice data

- Data is observed on polygon entities with defined boundaries
- Observed data are frequently aggregations within the boundaries, such as population counts.
- An example: <http://www.ij-healthgeographics.com/content/pdf/1476-072X-10-14.pdf>
- The area is subdivided in 115 neighborhoods
- Outcome: SMR of child mortality at each neighborhood
- Data is modeled by generalized linear mixed models (Poisson regression) accounting for spatial dependency using a CAR structure

Spatial Point Patterns

- Finite set of points in a given region A
- Stochastic process: the locations of the events observed within a bounded region A.
- Some examples: trees, nests, earthquake epicentres



Examples in Epidemiology

- 1854 Broad street cholera outbreak

[http://en.wikipedia.org/wiki/John_Snow_\(physician\)](http://en.wikipedia.org/wiki/John_Snow_(physician))

- It is considered the first application of spatial points pattern analysis
- Aggregation of cases was used to identify the focus outbreak.

- Urinary Schistosomiasis Infection

<https://www.ajtmh.org/view/journals/tpmd/70/4/article-p443.xml>

- Points: households
- Outcome: urine concentration levels in households
- Covariate: distance to potential focus of the infection
- Aim: to evaluate human infection clustering by household

Examples in Epidemiology

- Primary Biliary Cirrhosis (PBC)

<http://onlinelibrary.wiley.com/doi/10.1053/jhep.2001.29760/pdf>

- Aim: to describe the spatial distribution of cases of PBC.
- Spatial distribution of cases if compared to that of controls in order to detect higher risk clusters.
- Points: subject's home postcode

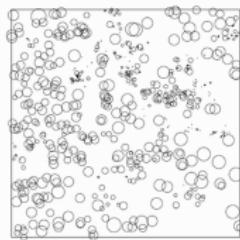
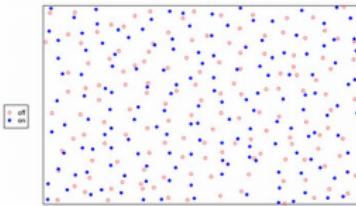
- Legionnaire's disease outbreak

http://link.springer.com/chapter/10.1007/0-387-31144-0_9

- Aim: to describe the spatial distribution of cases of legionnaire's disease.
- Spatial distribution of cases if compared to that of controls in order to detect potential sources of infection.
- Points: subject's home location

Types of data

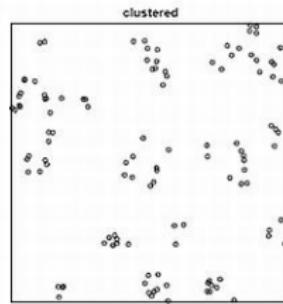
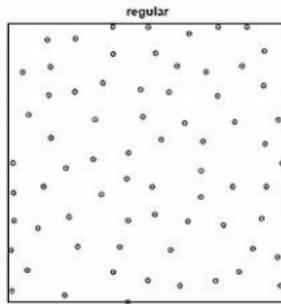
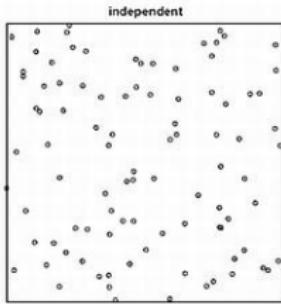
- Points. Locations of the events.
- Marks. Extra information attached to points.



- Covariates. Collected thorough out the study area (observed points and other locations).

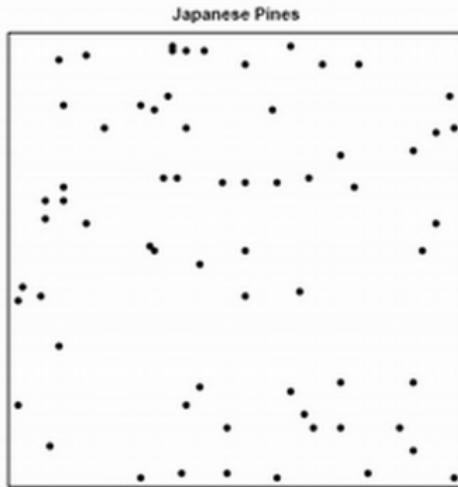
Questions of interest

- Estimation of **Intensity**. Average density of points (expected number of points per unit area)
- **Interaction**. Dependence between the points in a point pattern



Questions of interest

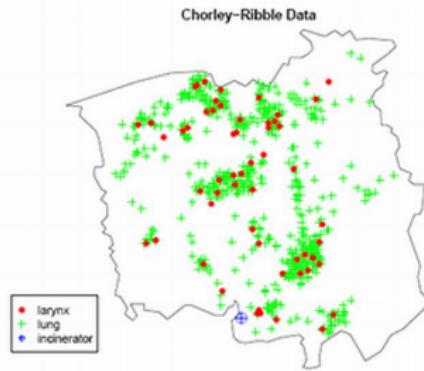
- **Example.** Locations of 65 saplings of Japanese pine in a 5.7×5.7 metre square sampling region in a natural stand.
- **Main question.** Is the spacing between saplings greater than would be expected for a random pattern? (reflecting competition for resources)



Questions of interest

Covariate effects

- Does intensity depend on covariates?
- Do covariates explain interaction between points?
- **Example.** Apparent cluster of cases of cancer of the larynx occurred near a disused industrial incinerator.
- **Main question.** Is there evidence of raised incidence of laryngeal cancer near the incinerator?

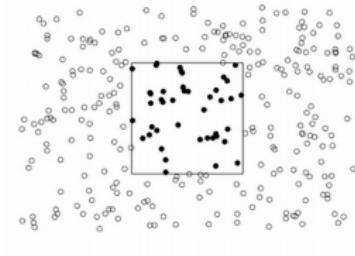


The study region

- The point process is only observed in the study region W .

$$\mathbf{x} = \{x_1, \dots, x_n\}, \mathbf{x} \in W, n \geq 0$$

- The window W must be fixed and known. Not guess!
- Covariates should also be measured at non-data locations.
- Implicitly assumed that all \mathbf{x} points within W have been mapped (observed).
- Edge effect \rightarrow biased estimates \rightarrow adjustment to eliminate the bias



Statistical Methods

- Intensity
 - Is the point process completely random? Is the intensity a flat surface in the area?
 - Does the intensity depends on covariates?
- Interaction
 - Beyond intensity is homogeneous or not, is the point process random?
 - Does the point process show a cluster or regular pattern?
 - What is the best model to account for a non random interaction?
- Case/control studies
 - Is the risk spatial variation different in cases than that of controls?
 - Is the interaction pattern different in cases than that of controls?

Spatial Point Patterns

Intensity and Randomness

Josep Lluís Carrasco

Departament de Fonaments Clínics
Universitat de Barcelona

Complete Spatial Randomness

- Homogeneous Poisson Point Process or Uniform Poisson Process with intensity λ .
- Properties
 - ➊ The number $N(A)$ of points falling in any region A is a Poisson random variable.
 - ➋ The expected number of points falling in A is $\lambda \cdot \text{area}(A)$.
 - ➌ If A_1 and A_2 are disjoint sets then $N(A_1)$ and $N(A_2)$ are independent random variables.
 - ➍ Given that $N(A) = n$, the n points are independent and uniformly distributed in A.
- Points are independent of each other and have the same propensity to be found at any location
- The Uniform Poisson Process is often the *null model*

Inhomogeneous Poisson Point Process

- The number $N(A)$ of points falling in any region A is a Poisson random variable.
- The expected number of points falling in A is

$$E [N (A)] = \int_A \lambda (u) du$$

- If A_1 and A_2 are disjoint sets then $N(A_1)$ and $N(A_2)$ are independent random variables.
- Given that $N(A) = n$, the n points are i.i.d. in A with density function

$$f (u) = \frac{\lambda (u)}{I}$$

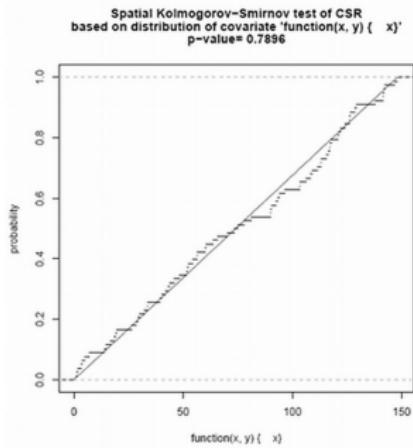
$$I = \int_A \lambda (u) du$$

Testing for CSR

- Classical approach: Chi-square test based on quadrat counts
- The region is divided in m subregions with equal area.
- Under CSR, the counts $n_j, j = 1, \dots, m$, are i.i.d. Poisson random variables with the same expected value.
- Handycaps:
 - There are different ways of departure from H_0 .
 - The power of the test depends on the size of the quadrats.

Testing for CSR

- **Kolmogorov-Smirnov test** based on a real-value function $T(x, y)$ defined at all locations (x, y) .
- The empirical distribution of $T(x, y)$ is compared with the theoretical distribution of $T(x, y)$ under CSR.
- For example $T(x, y) = x$. The theoretical distribution of x-coordinate under CSR is uniform.



Testing for CSR

- The Chi-square and KS tests can be used as exploratory analysis to investigate the dependence of the intensity on a covariate.
 - ➊ Chi-square test. Dividing the region into *regular subregions* according to the covariate.
 - ➋ KS test. Using the covariate to define $T(x,y)$.
- KS is more powerful and preferable in case of continuous covariates.
- KS test can be ineffective with factor or discrete variables because tied values.

Estimation of the intensity

- Intensity: the average density of points per unit area.

Let $N(A)$ be the random process *number of points in area A*

- Uniform or homogeneous.**

Constant throughout the area under study.

$$E[N(A)] = \lambda \cdot \text{area}(A)$$

$$\lambda = \frac{n}{\text{area}(A)}$$

- Inhomogeneous.**

Intensity depends on the location.

$$E[N(A)] = \int_A \lambda(u) du$$

Estimation can be done **parametrically** or **non-parametrically**.

Non-parametric estimation of the intensity

- Method: kernel density estimation.

$$\lambda(u) = \frac{1}{b^2} \sum_{i=1}^n \kappa\left(\frac{\|u - u_i\|}{b}\right) \frac{1}{q(\|u\|)}$$

- $\kappa(\cdot)$ is a bivariate kernel function.
- $q(\|u\|)$ is the edge correction. The kernel mass inside the window.

$$q(\|u\|) = \int_{v \in A} \kappa\left(\frac{\|u - v\|}{b}\right) dv$$

- b is the bandwidth: controls the level of smoothing.

Non-parametric estimation of the intensity

Kernel	$\kappa(u^*)$
Uniform	$\frac{1}{2} I(u^* \leq 1)$
Triangle	$\frac{1}{(2-b)} (1 - u^*) I(u^* \leq 1)$
Quartic (biweight)	$\frac{15}{16} \left[1 - (u^*)^2\right]^2 I(u^* \leq 1)$
Gaussian	$\frac{1}{\sqrt{2\pi b^{-1}}} \exp\left[-\frac{1}{2} (u^*)^2\right]$

$$u^* = \frac{u - u_i}{b}$$

Non-parametric estimation of the intensity

- The form of the kernel weakly influences intensity estimates. However, the bandwidth can have a great impact.
- There are several approaches for determining bandwidth¹
- Mean Squared Error²

$$MSE(b) = E \left[(\hat{\lambda}(x, b) - \lambda(x))^2 \right]$$

- Gaussian kernels³

$$b = \sigma N^{-1/(dim+4)}$$

- Convenient: use several values and choose a plausible value.

¹Silverman, B.W. (1986) *Density Estimation for statistics and data analysis*

²Diggle, P. (1985) A Kernel Method for Smoothing Point Process Data, *JRSS Series C*

³Scott, DW. (1992). *Multivariate density estimation: theory, practice, and visualization*

Parametric estimation of the intensity

- Log-likelihood for an HPP:

$$L(\lambda) = n \log(\lambda) - \lambda \text{area}(A)$$

- Log-likelihood for an IPP:

$$L(\lambda) = \sum_{i=1}^n \log \lambda(x_i) - \int_A \lambda(x) dx$$

- LRT for CSR

$$LRT = 2 \log \frac{L_I}{L_H}$$

- The IPP allows including spatial covariates

$$\log \lambda(x) = \sum_{j=1}^p \beta_j z_j(x)$$

Model selection and validation

Model selection

- Nested models: Analysis of Deviance (LRT)
- No nested models: AIC

Model validation

- Goodness-of-fit
 - Chi-square test on quadrat counts
 - KS test using the fitted model as null model
- Residuals
 - Raw residuals
 - Smoothed residual
 - Cumulative sum of residuals (Lurking Plot)

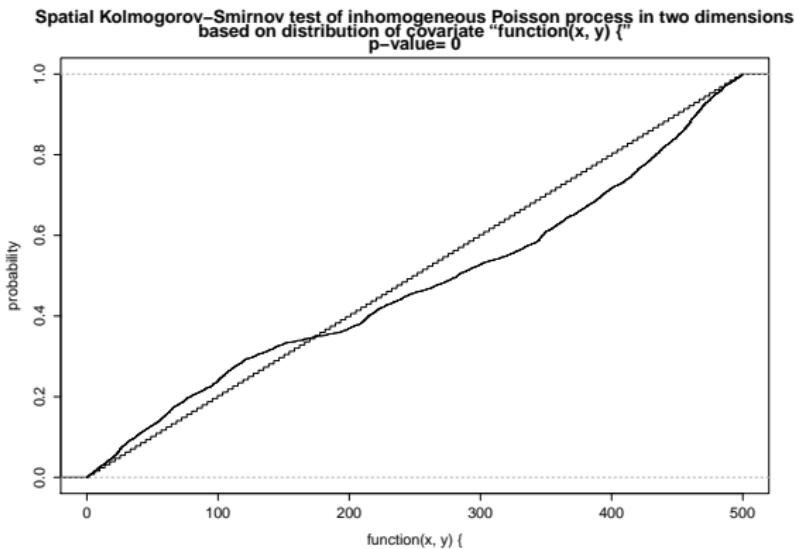
Chi-square test on quadrant counts

- Null hypothesis: the model is true.

bei							
666	601	677	478.5	130	402.8	481	319.7
2.7		9.1		-14		9	
544	601.6	165	477.9	643	402.3	298	320.2
-2.3		-14		12		-1.2	

KS test

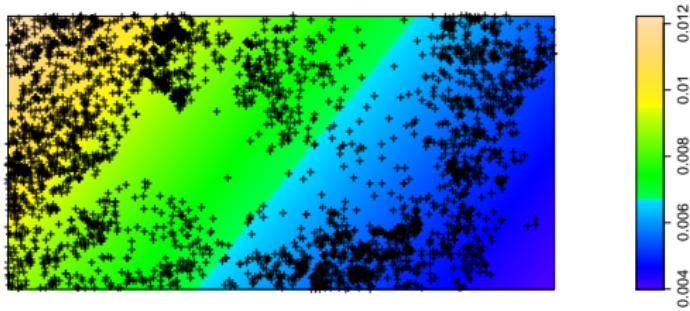
- The empirical distribution of $T(x, y)$ is compared with the theoretical distribution of $T(x, y)$ under the fitted model.



Raw Residuals

$$n(x) - \int_A \lambda(u) du$$

`predict(fit)`

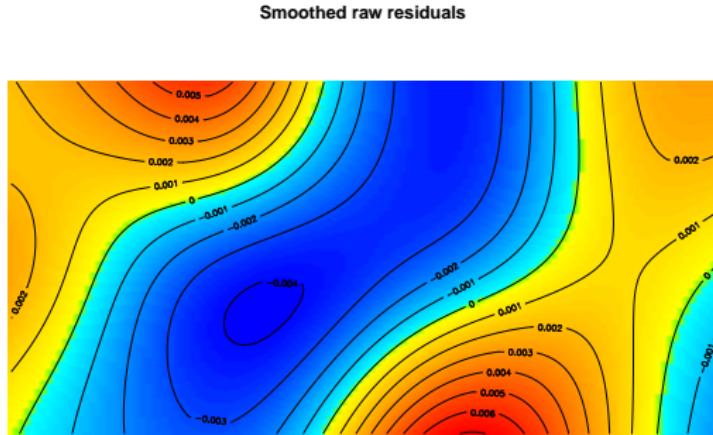


Smoothed residuals

$$s(u) = \lambda^s(u) - \lambda^f(u)$$

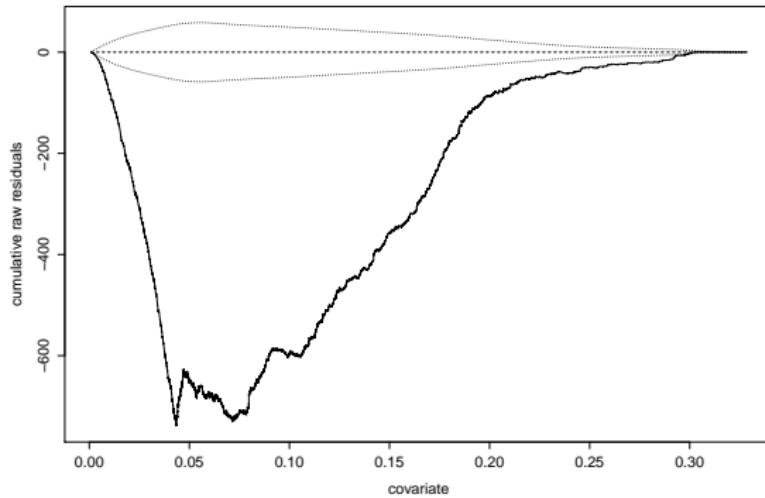
$\lambda^s(u)$ is the nonparametric kernel estimate of the intensity

$\lambda^f(u)$ is the smoothed version of the parametric estimate of the intensity according to the fitted model



Lurking plot

Useful to analyze the dependence between the residuals and covariates.



Spatial Point Patterns

Interaction

Josep Lluís Carrasco

Departament de Fonaments Clínics
Universitat de Barcelona

Outline

1 Distance Methods

2 Models for non-Poisson patterns

Outline

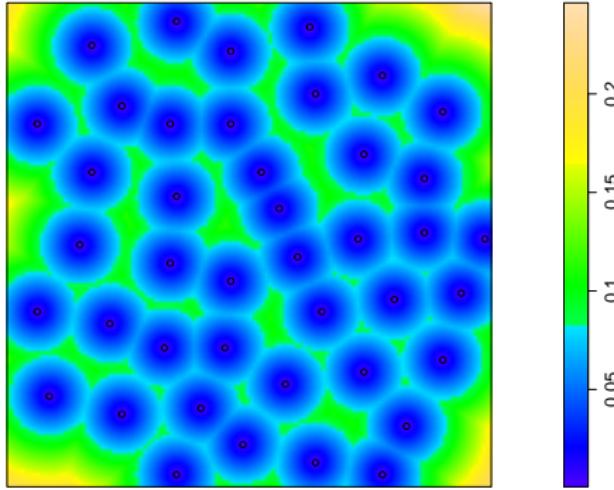
1 Distance Methods

2 Models for non-Poisson patterns

Empty space distances

- The distance from a fixed reference location u in the window to the nearest data point.

Empty space distances



Empty space distances

$$d(u, \mathbf{x}) = \min \{ \|u - x_i\| : x_i \in \mathbf{x}\}$$

- Cumulative distribution function

$$F(r) = P[d(u, \mathbf{X}) \leq r]$$

- Empirical distribution function

$$F^*(r) = \frac{1}{m} \sum_j \mathbf{1}\{d(u_j, x) \leq r\}$$

- Empirical *edge corrected* distribution function

$$\hat{F}(r) = \sum_j e(u_j, r) \mathbf{1}\{d(u_j, x) \leq r\}$$

where $e(u_j, r)$ is a weight to correct the edge effect.

Edge correction

- Simplest approach: **border method**.
- Only use those points at a distance r to the area boundary.
- For each r the function is estimated as

$$\hat{F}^b(r) = \frac{\hat{F}(r) \cap A_{(-r)}}{|A_{(-r)}|}$$

- Other option: edge effect as a censoring issue (Baddeley and Gill, 1997)

Empty space distances

- Homogeneous Poisson Point Process

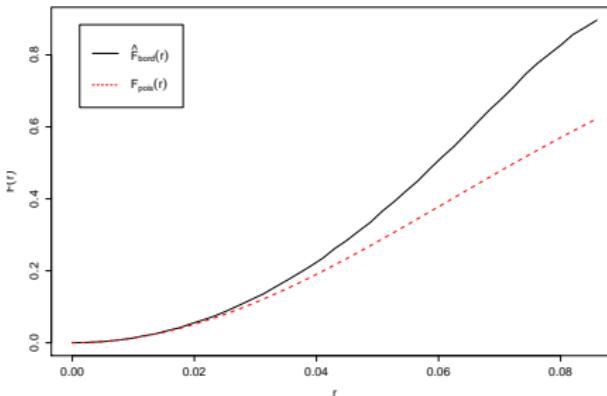
$$F_{pois}(r) = 1 - \exp(-\lambda\pi r^2)$$

- Regular pattern

$$\hat{F}(r) > F_{pois}(r)$$

- Clustered pattern

$$\hat{F}(r) < F_{pois}(r)$$



Nearest neighbour distances

$$t_i = \min_{j \neq i} \|x_i - x_j\|$$

- Cumulative distribution function

$$G(r) = P[d(u, \mathbf{X}\{u\}) \leq r | u \in \mathbf{X}]$$

- Empirical distribution function

$$G^*(r) = \frac{1}{n} \sum_i \mathbf{1}\{t_i \leq r\}$$

- Empirical *edge corrected* distribution function

$$\hat{G}(r) = \sum_i e(x_i, r) \mathbf{1}\{t_i \leq r\}$$

Nearest neighbour distances

- Homogeneous Poisson point process

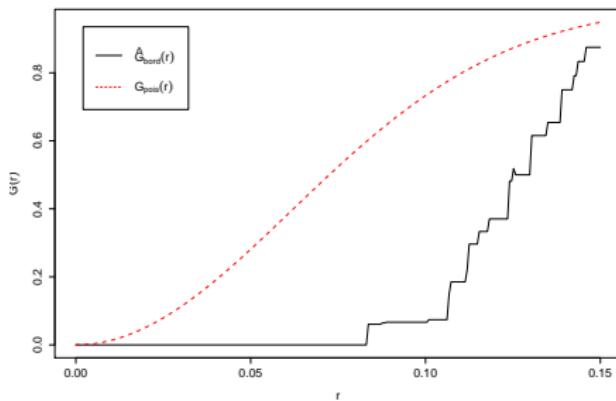
$$G_{Pois}(r) = 1 - \exp(-\lambda\pi r^2)$$

- Regular pattern

$$\hat{G}(r) < G_{Pois}(r)$$

- Clustered pattern

$$\hat{G}(r) > G_{Pois}(r)$$



Pairwise distances

$$s_{ij} = \|x_i - x_j\|$$

- Expected number of other points of the process within a distance r

$$E [n (\mathbf{X}, r)] = \lambda \cdot K (r)$$

where $K (r)$ is the Ripley's K-function.

- Estimator of $K (r)$

$$\hat{E} [n (\mathbf{X}, r)] = \frac{1}{N} \sum_i \sum_{j \neq i} \mathbf{1} \{ \|x_i - x_j\| \leq r \}$$

$$\hat{\lambda} = \frac{N}{area (A)}$$

$$\hat{K} (r) = \frac{area (A)}{N^2} \sum_i \sum_{j \neq i} \mathbf{1} \{ \|x_i - x_j\| \leq r \} e (x_i, x_j, r)$$

- **Isotropic method.** $e (x_i, x_j, r)$ is the proportion of a disc centered at point x_i and passing through point x_j that lies within the study area.

Pairwise distances

- Homogeneous Poisson point process

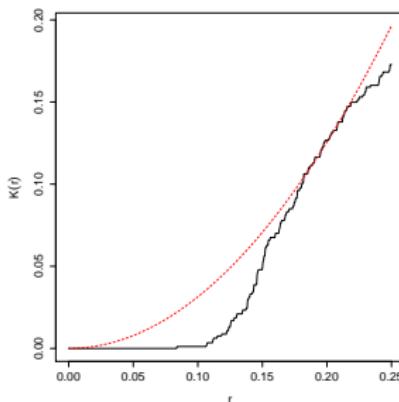
$$K_{Pois}(r) = \pi r^2$$

- Regular pattern

$$\hat{K}(r) < K_{Pois}(r)$$

- Clustered pattern

$$\hat{K}(r) > K_{Pois}(r)$$



Additional related functions

- L-function

$$L(r) = \sqrt{\frac{K(r)}{\pi}} \quad L_{Pois}(r) = r$$

- *Pair correlation function.* Probability of observing a pair of points separated by a distance r , divided by the corresponding probability for a Poisson process.

$$g(r) = \frac{K'(r)}{2\pi r} \quad g_{Pois}(r) = 1$$

Clustering $g(r) > 1$; Regularity $g(r) < 1$

- J function

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

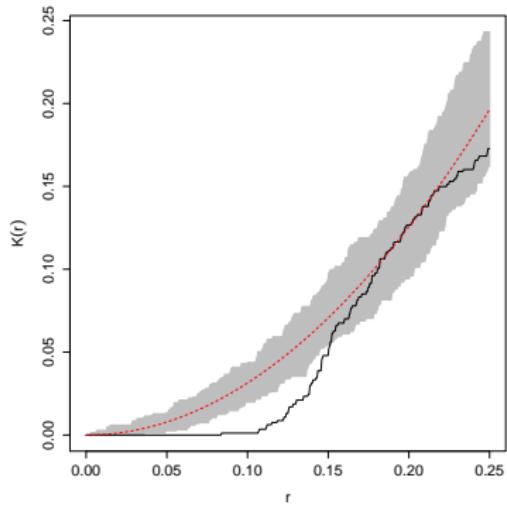
Envelopes and Montecarlo tests

- ➊ Suppose the reference curve is the theoretical K function for CSR.
- ➋ Generate M independent simulations of CSR inside the study region A.
- ➌ Compute the estimated K functions for each of these realizations.
- ➍ Obtain the pointwise upper and lower envelopes of these simulated curves.

$$L(r) = \min_j \hat{K}^{(j)}(r)$$

$$U(r) = \max_j \hat{K}^{(j)}(r)$$

Envelopes and Montecarlo tests



Envelopes and Montecarlo tests

- For any fixed value of r the significance level is $\alpha = 2/(M+1)$.
 - r must be fixed in advance
 - If the test was used as *Does $K(r)$ lie outside for any r ?* the significance level would be higher.
 - Simultaneous Monte Carlo test → simultaneous critical bands with a global significance level $1/(M+1)$.
- ① For each estimate $\hat{K}_i(r)$, $i = 1, \dots, M$ compute its maximum deviation from Poisson K function

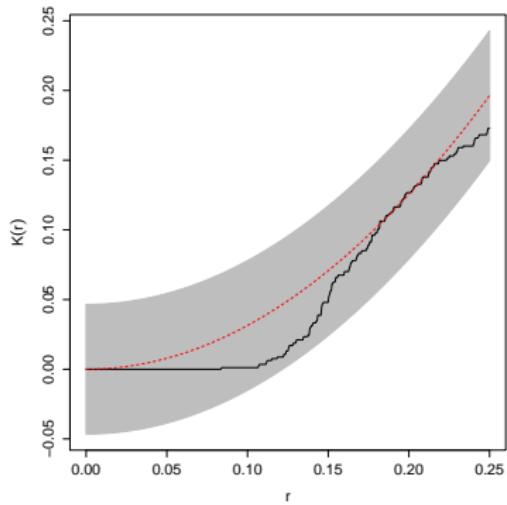
$$D_i = \max |\hat{K}_i(r) - K_{Pois}(r)|$$

② $D_{max} = \max \{D_i\}$

$$L(r) = \pi r^2 - D_{max}$$

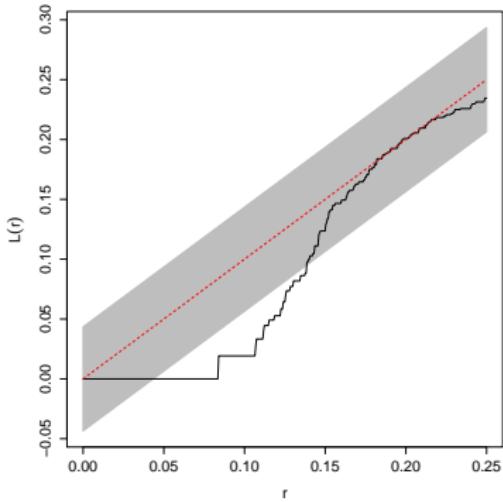
$$U(r) = \pi r^2 + D_{max}$$

Envelopes and Montecarlo tests



Envelopes and Montecarlo tests

A more powerful test is obtained if the variance is stabilized using $\sqrt{K(r)}$.



Outline

1 Distance Methods

2 Models for non-Poisson patterns

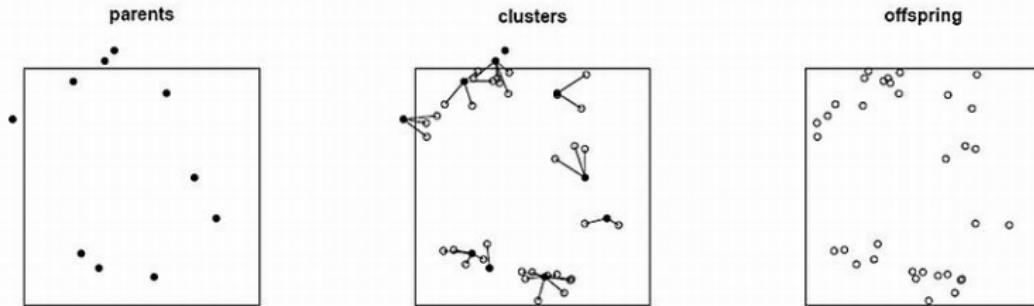
Outline

1 Distance Methods

2 Models for non-Poisson patterns

Poisson cluster processes

- Each event belongs to a particular cluster.
- The process consists of a set of *parent locations* each of which generates a set of *child locations*.
- Only *child locations* are observed.



Neyman-Scott processes

- ➊ Parent locations follow a Poisson process
- ➋ Numbers of children per parent are i.i.d..
- ➌ Locations of children around their respective parent are i.i.d.
- Matérn process.
 - Parent points come from a Poisson process with intensity $\lambda(x)$.
 - Each parent has a *Poisson* (μ) number of offspring.
 - Children has an independently and uniformly distributed displacement from its parent in a disc of radius r centred around the parent.
- Thomas process.
 - Parent points come from a Poisson process with intensity $\lambda(x)$.
 - Each parent has a *Poisson* (μ) number of offspring.
 - Children has an independently and isotropic Gaussian $N(0, \sigma^2 \mathbf{I})$ distributed displacement from its parent in a disc of radius r centred around the parent.

Cox processes

- Poisson process with a random intensity function.
- Most used: log-Gaussian Cox Process.
- Intensities $\lambda(\mathbf{u}) \sim N(\exp(\beta\mathbf{X}), C(r))$
- Covariance between locations takes the form:

$$C(r) = \sigma^2 f(r, \alpha)$$

- σ^2 and α controls the scale and the strength of the autocorrelation.
- $f(r, \alpha)$ is a decay function. Two cases are:
 - Exponential. $f(r, \alpha) = \exp(-r/\alpha)$
 - Gaussian $f(r, \alpha) = \exp(-(r/\alpha)^2)$

Model fitting: estimation method

- **Minimum contrast** (Diggle, 1983)

$$D(\theta) = \int_a^b |\hat{K}(r)^q - K_\theta(r)^q|^p dr$$

- a, b and q are used to control the sampling fluctuations in the estimates of K .
 - a, b : limits of the range of distances.
 - $q > 0$ is used to transform the K function to more powerful options.
 - If $q = 0.5$ the contrast uses the L function.
 - In spatial processes commonly $p = 2$.
- It gives consistent estimates.
- It can be very computationally intensive.

Diggle, PJ (1983). Statistical Analysis of Spatial Point Patterns. Chapter 5. Academic Press, London 1983

Model fitting: estimation method

- **Composite likelihood** (Guan, 2006).
- It maximises a composite likelihood based on knowledge of the second moment of intensity (covariance between locations).
- **Palm likelihood** (Tanaka et al, 2008).
- It assumes a Palm distribution for the point process.
- Asymptotic normality of estimates (if the likelihood is correct).

Guan, Y. (2006). Journal of the American Statistical Association 101, 1502-1512

Tanaka, U. and Ogata, Y. and Stoyan, D. (2008). Biometrical Journal 50, 43-57

Gibbs point processes

- Spatial point process models that are constructed by writing down their probability densities.

$$f(\mathbf{X}) = \alpha \left[\prod_{i=1}^n b(x_i) \right] \left[\prod_{i < j} c(x_i, x_j) \right]$$

- α is a normalising constant.
- $b(x_i)$ are the *first order* terms.
- $c(x_i, x_j)$ are the *second order* terms.
- Intensity. $\lambda(u) = b(u) [\prod_{i=1}^n c(u, x_i)]$
- Gibbs point processes can model regular patterns.
- Estimation approach: Maximum pseudolikelihood

Gibbs point processes for regular patterns

Hard core process

- Intensity. $b(u) = \lambda$
- Pairwise interaction.

$$c(u, v) = \begin{cases} 1 & \text{if } \|u - v\| > r \\ 0 & \text{if } \|u - v\| \leq r \end{cases}$$

Strauss process

- Pairwise interaction.

$$c(u, v) = \begin{cases} 1 & \text{if } \|u - v\| > r \\ \gamma & \text{if } \|u - v\| \leq r \end{cases}$$

$$0 \leq \gamma \leq 1$$

- Intensity. $b(u) = \lambda \cdot \gamma^{t(u)}$

where $t(u)$ is the number of points at a distance lower than r .

Gibbs point processes for cluster patterns

Geyer process

- Pairwise interaction.

$$c(u, v) = \begin{cases} 1 & \text{if } \|u - v\| > r \\ \gamma & \text{if } \|u - v\| \leq r \end{cases}$$

- Intensity. $b(u) = \lambda \cdot \gamma^{\min(s, t(u))}$

where

- $t(u)$ is the number of points at a distance lower than r .
- s is the saturation parameter.

Gibbs point processes for cluster patterns

Area Interaction process

- Two points interact if their discs of radius r overlaps.
- Pairwise interaction.

$$c(u, v) = \begin{cases} 1 & \text{if } \|u - v\| > 2r \\ \gamma & \text{if } \|u - v\| \leq 2r \end{cases}$$

- Intensity. $b(u) = \lambda \cdot \gamma^{-A}$

where A is the area of the region formed by the union of discs of radius r centred at the observed points.

Spatial Point Patterns

Case-control studies

Josep Lluís Carrasco

Departament de Fonaments Clínics
Universitat de Barcelona

Introduction

- The distribution of the cases of a disease can be regarded as a point process.
 - The cases point process also reflects the distribution of the population.
 - It is necessary to have an estimate of the spatial distribution of the population → sample of controls.
 - The spatial point process of the cases is compared to that of the controls.
- ① Spatial variation of the relative risk: analyze the ratio between the intensities of cases and controls.
 - ② Binary regression: probability of being a case.
 - ③ Difference of K-function of cases and controls.

Spatial variation of the relative risk

- Sample of n_1 cases and n_0 controls.
- \mathbf{x}_1 and \mathbf{x}_0 are realizations of two IPP with intensities $\lambda_1(x)$ and $\lambda_0(x)$.
- If the distribution of cases and controls is the same:

$$\lambda_1(x) = \frac{n_1}{n_0} \lambda_0(x)$$

- Relative ratio of intensities

$$\rho(x) = \frac{\lambda_1(x)}{\lambda_0(x)}$$

- Under the null hypothesis of equal spatial distribution.

$$\rho_0 = n_1/n_0$$

Spatial variation of the relative risk

- If the logarithm of the ratio of the densities is used:

$$r(x) = \log(f(x)/g(x))$$

- Null hypothesis of equal spatial distribution $r_0(x) = 0$.
- The risk ratio is non-parametrically computed at each location.
- To assess departure from the null hypothesis:

$$T = \int_A (\rho(x) - \rho_0)^2 dx$$

Spatial variation of the relative risk

- The integral can be computed on a regular grid of points $\{s_i, i = 1, \dots, p\}$

$$\hat{T} = |c| \sum_{i=1}^p (\hat{\rho}(s_i) - \hat{\rho}_0)^2$$

where $|c|$ is the area of the cells of the grid.

- Significance is computed by Monte Carlo test (permutation test).
- A probability map is calculated by comparing the observed kernel density ratios to the permuted ratios.
- Permutations approach is very computational demanding.
- Hazelton and Davies (2009) proposed an alternative based on the asymptotic properties of the kernel density estimator.

Binary regression

- Probability of being a case at a given location.

$$P(Y_i = 1 | X_i = x_i) = p(x_i) = \frac{\lambda_1(x_i)}{\lambda_0(x_i) + \lambda_1(x_i)}$$

- Relation with relative risk

$$\text{logit}(p(x)) = \log\left(\frac{p(x)}{1-p(x)}\right) = \log\left(\frac{\lambda_1(x)}{\lambda_0(x)}\right)$$

- Binary regression using GAM

$$\text{logit}(p(x)) = Z(x)' \beta + s(x)$$

Point source pollution

- Fit an IPP to the cases where intensity accounts for the distance to the pollution sources.

$$\lambda_1(x) = \rho \lambda_0(x) f(x - x_0; \theta)$$

- ρ stands for the baseline odds.
- $f(x - x_0; \theta)$ function of the distance from point x to location of source x_0 .

$$f(x - x_0; \alpha, \beta) = 1 + \alpha \exp(-\beta \|x - x_0\|^2)$$

- Alternatively we can model

$$p(x) = \frac{\lambda_1(x)}{\lambda_0(x) + \lambda_1(x)} = \frac{\rho f(x - x_0; \theta)}{1 + \rho f(x - x_0; \theta)}$$

- Estimation is carried out by maximum likelihood.

Assessment of general spatial clustering

- Spatial variation of the risk.
 - the risk is not homogeneous in the study region.
 - Cases are independent.
- Clustering.
 - Occurrence of cases is not at random.
 - The presence of a case increases the probability of other cases appearing nearby.
- Aim: to test if cases and controls are point processes that have **same intensities** and produce the **same K-functions**.

$$D(s) = K_1(s) - K_0(s)$$

- Statistical test:

$$D = \int_A \frac{D(s)}{\text{var}[D(s)]^{1/2}}$$

- Significance obtained by permutation.