Master's Thesis

Autoregressive spatial models: an application to linguistic surveys

Adrián Pérez Bote

Master in Statistical Techniques

Academic year 2015-2016

Master's Thesis Proposal

Título en galego: Modelos espaciais autorregresivos: unha aplicación a enquisas lingüísticas

Título en español: Modelos espaciales autorregresivos: una aplicación a encuestas lingüísticas

English title: Autoregressive spatial models: an application to linguistic surveys

Modalidade: Modalidade A

Autor/a: Adrián Pérez Bote, Universidade de Santiago de Compostela

Director/a: Rosa M. Crujeiras Casais, Dpto. de Estatística e Investigación Operativa, Universidade de Santiago de Compostela

Breve resumo do traballo:

Na análise de datos discretos que presentan dependencia espacial, son de especial interese os modelos baseados na teoría de Campos Gaussianos de Markov (en inglés, Gaussian Markov Random Fields, GMRF). Dous casos particulares son os modelos CAR (modelos condicionalmente autoregresivos) e SAR (modelos simultáneamente autoregresivos). Neste traballo estes modelos preséntanse primeiro desde un punto de vista teórico, incluíndo a estimación por Máxima Verosimilitude dos parámetros, e depois empréganse para modelizar datos reais. A variábel que se pretende explicar é a porcentaxe de falantes de lingua galega en Galicia e de euskera no País Vasco, estudando estes territorios divididos por concellos.

Dona Rosa M. Crujeiras Casais, profesora contratada doutora do Dpto. de Estatística e Investigación Operativa, Universidade de Santiago de Compostela, informa que o Traballo Fin de Máster titulado

Autoregressive spatial models: an application to linguistic surveys

foi realizado baixo a súa dirección por don Adrián Pérez Bote para o Máster en Técnicas Estatísticas. Estimando que o traballo está rematado, da a súa conformidade para a súa presentación e defensa ante un tribunal.

En Vigo, a 2 de xaneiro de 2016.

A directora: O autor:

Dona Rosa M. Crujeiras Casais Don Adrián Pérez Bote

Contents

Abstract

Resumo en galego

Este traballo de fin de máster ten como obxectivo o estudo dos modelos espaciais autorregresivos tanto desde un punto de vista teórico como aplicándoos para modelizar datos reais.

En primeiro lugar estudaranse os Campos Gaussianos de Markov que nos serán útiles para introducir os modelos que empregaremos: modelos condicionalmente autorregresivos e modelos simultaneamente autorregresivos. Despois da introdución destes dous modelos, abordarase un estudo da técnica utilizada para estimar os parámetros (Máxima Verosimilitude) e efectuaranse simulacións para ilustrar as propiedades dos estimadores dos parámetros de cada modelo. Posteriormente os modelos serán empregados co obxectivo de modelizar os datos dos que se dispón, porcentaxe de falantes de lingua galega en Galicia e de lingua vasca en Euskadi, dividindo estes territorios en municipios e realizando antes unha análise exploratoria e unha aproximación mediante regresión linear.

Finalmente introduciranse brevemente modelos espazo-temporais desde un punto de vista soamente teórico, xa que non se dispón de datos para aplicalos á materia de estudo.

English abstract

This MSc Thesis has as main goal the study of spatial autoregressive models from both a theoretical point of view and applying them to model real data.

First of all Gaussian Markov Random Fields will be studied, because they will be useful to introduce the models that will be applied: conditional autoregressive models and simultaneous autoregressive models. After the introduction of these two models, the study of the technique used to estimate the parameters (Maximum Likelihood) will be approached, and simulations will be carried out in order to illustrate the properties of the estimators of the parameters in each model. Afterwards, the models will be applied with the objective of modeling the data at disposal, the percentage of speakers of Galician language in Galicia and Basque language in the Basque Country, dividing these territories into municipalities and doing before some exploratory analysis and an approach by linear regression.

Finally, spatiotemporal models will be briefly introduced only in a theoretical way, as there is no data in the study field at disposal to apply them.

Introduction

The objective of this MSc Thesis is to present some techniques and models techniques and models that allow to model and to understand the spatial distribution of the percentage of speakers of a given language on a territory where two languages have a relevant amount of speakers.

The data chosen are the percentage of population that speaks always Galician language in each municipality of Galicia and the percentage that considers Basque its main language in each municipality of the Basque country. A reasonable assumption is made on this data: the spatial-related territories present some correlation on the percentage of speakers of the given language. Other assumptions are made, as that there are variables that explain part of the variability of the response (such as percentage of Galician-speakers or Basque-speakers).

Chapter 1 presents an approach to Gaussian Markov Random Fields. For a proper characterization, it is necessary to introduce the concepts of conditional independence and undirected graphs and their elements. After defining Gaussian Markov Random Fields some properties will be enunciated. All this is very useful to approach further issues.

In Chapter 2, conditional autoregressive and simultaneous autoregressive models are introduced theoretically, without losing insight on the relation between them and the general Markov Random Fields. When applying models to data, the estimators of the parameters are obtained by Maximum Likelihood method. The details of this procedure are presented in this chapter, as well as simulations to illustrate their properties such as if they are unbiased or consistent. This chapter ends with an approach to exploratory analysis to determine whether some data should be studied using these models or not. The Moran's I will be used for this purpose.

Chapter 3 is the main part of the thesis. The linguistic data are taken, studied and modeled with the techniques that have been introduced. The procedure is analogous for both cases, each of one in a different section. First of all some exploratory analysis and linear regression is developed, that in the two sections suggests formally the use of autoregressive models, because it is clear that the residuals of the linear regression are spatially correlated. Afterwards, two different spatial structures are created to use in the autoregressive models, one of them based on common borders and the other on distances between the centroids of two municipalities. The models are fitted and after comparing them some conclusions are enunciated.

Three appendixes are added in the end. The first of them is a brief introduction to spatiotemporal autoregression. As the percentage of speakers of a language in a territory evolves over time and it is influenced by close territories, it would be interesting to apply spatiotemporal models. However, the lack of data makes it impossible, so the only approach that can be done is theoretical. In the second appendix the programming techniques used to fit the models are explained. In the third one a scheme of the notation used along the thesis is presented.

Bibliographical notes: the main contents of this manuscript have been obtained from references focused on Gaussian Markov Random Fields, such as [6], and on SAR and CAR, such as [1], [3] and [4]. The data analysis is based on some code described in [2] while it is developed using the language for statistical computing R [5].

Chapter 1

Gaussian Markov Random Fields

Gaussian Markov Random Fields (GMRFs) are introduced in this chapter. The basic concept of conditional independence is defined in the first section and in the second one there is an approach to undirected graphs and a little recall of the multivariate normal distribution. In the third section GMRFs are defined and some of their properties are presented. For further information, check [6] (pp 1-30), a monograph on GMRF. The link between GMRFs and autoregressive models will soon arise.

Along this chapter, random vectors taking values on points or regions on a spatial domain will be considered. Note that, in this setting, there is no natural ordering for the indexes (or elements) of the vector.

1.1 Conditional independence

An approach to *conditional independence* concept is going to be presented in order to understand GMRFs. Recall that two random variables X and Y are *independent* ($X \perp Y$) iff

$$p(x, y) = p_X(x)p_Y(y).$$

Two variables X and Y are called *conditionally independent* given a third variable Z ($X \perp Y|Z$) iff

$$p(x, y|z) = p_{X|Z}(x|z)p_{Y|Z}(y|z).$$

It is easy to realize that independence implies conditional independence, while the reciprocal is not true. This fact is illustrated with the example in Figure 1.1. In this case, equiprobable squares show the probabilities of each of the three variables X, Y and Z and the dependencies between them. Note that $p_X(x) = 5/8$, $p_Y(y) = 1/4$, $p_Z(z) = 1/4$. The random variables X and Y are clearly not independent, as $p(x, y) = 1/8 \neq p_X(x)p_Y(y) = 5/32$. However, they are conditionally independent given Z because

$$p(x, y|z) = \frac{1}{4} = p_{X|Z}(x|z)p_{Y|Z}(y|z).$$

The following theorem is useful to verify conditional independence.

Theorem 1.1. *Let X, Y and Z be random variables.*

$$X \perp Y|Z \Leftrightarrow p(x, y, z) = f(x, z)g(y, z)$$

for some functions f and g and for all z with $p_Z(z) > 0$.

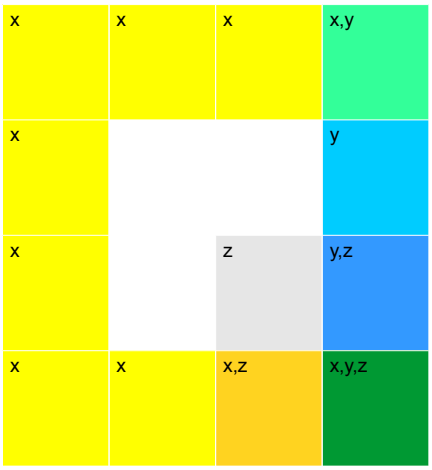


Figure 1.1: Graphical representation of three random variables: X, Y and Z. Each labelled and colored square is equiprobable, and its label indicates whether the probability of each of the variables is 1 (the corresponding letter appears on it) or 0 (it does not appear).

It is easy to extend the concept of conditional independence to the multivariate case, the one that is useful in our context. Given $\mathbf{X} = (X_1, \cdots, X_{n_1})^T$, $\mathbf{Y} = (Y_1, \cdots, Y_{n_2})^T$ and $\mathbf{Z} = (Z_1, \cdots, Z_{n_3})^T$, $\mathbf{X}$ and $\mathbf{Y}$ are called *conditionally independent* given $\mathbf{Z}$ ($\mathbf{X} \perp \mathbf{Y} | \mathbf{Z}$) iff

$$p(\mathbf{x},\mathbf{y}|\mathbf{z}) = p_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}|\mathbf{z}) p_{\mathbf{Y}|\mathbf{Z}}(\mathbf{y}|\mathbf{z}).$$

Theorem 1.1 holds also with multivariate variables.

1.2 Elements of a GMRF

1.2.1 Undirected graphs

Undirected graphs are going to be introduced now, as well as some other basic concepts such as *neighbors* of a node or of a set, *paths*, *separated nodes* or *sets* and *subgraphs*.

Undirected graph: tuple $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is the set of nodes in the graph and $\mathcal{E}$ is the set of edges $\{i, j\}$, where i and j are distinct nodes.

If $\{i, j\} \in \mathcal{E}$, there is an undirected edge from node i to node j. From now on, it is assumed that $\mathcal{V} = \{1, \cdots, n\}$, so the graph is said to be *labelled*.

Neighbors of node i: all nodes in $\mathcal{G}$ having an edge to node i,

$$ne(i) = \{j \in \mathcal{V} : \{i, j\} \in \mathcal{E}\}.$$

Neighbors of a set $\mathcal{A} \subset \mathcal{V}$: all nodes not in $\mathcal{A}$ having an edge to any node in $\mathcal{A}$,

$$ne(\mathcal{A}) = \bigcup_{i \in \mathcal{A}} ne(i) \setminus \mathcal{A}.$$

It is written $i \overset{\mathcal{G}}{\sim} j$ if node i and j are neighbors. If the graph is implicit it can be written just $i \sim j$.

Path from i_1 to i_m: sequence of distinct nodes in $\mathcal{V}$, $i_1, i_2, \cdots, i_m$ for which $(i_j, i_{j+1}) \in \mathcal{E}$ for $j = 1, \cdots, m-1$.

Separated nodes or sets: given a subset $\mathcal{C} \subset \mathcal{V}$, it is said that it *separates* two nodes $i, j \notin \mathcal{C}$ if every path from i to j contains at least one node from $\mathcal{C}$. This can be extended to disjoint sets $\mathcal{A}, \mathcal{B} \subset \mathcal{V} \backslash \mathcal{C}$, that are called *separated* by $\mathcal{C}$ if all $i \in \mathcal{A}$ and $j \in \mathcal{B}$ are separated by $\mathcal{C}$.

Subgraph: let $\mathcal{A}$ be a subset of $\mathcal{V}$. The graph restricted to $\mathcal{A}$ is denoted by $\mathcal{G}^{\mathcal{A}}$. This graph is obtained after removing all nodes not in $\mathcal{A}$ and all edges where at least one node does not belong to $\mathcal{A}$.

1.2.2 Normal distribution

The normal distribution is widely used in many areas of statistics. In particular, it is essential to understand and employ GMRFs because they are based on it. Recall now briefly this distribution and give some of the basic properties of the sum and splitting of multivariate normal distributions.

Take a normal random vector $\mathbf{X} = (X_1, \ldots, X_n)^T$, $n < \infty$, with vector of means $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ (always semipositive definite). Its density function is

$$p_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{n/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right), \ \mathbf{x} \in \mathbb{R}^n.$$

The distribution of $\mathbf{X}$ is denoted by $\mathbf{X} \in \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The following statement is a basic and important property of normal random vectors.

1. If $\mathbf{X} \in \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{X}' \in \mathcal{N}_n(\boldsymbol{\mu}', \boldsymbol{\Sigma}')$ are independent, then

$$\mathbf{X} + \mathbf{X}' \in \mathcal{N}_n(\boldsymbol{\mu} + \boldsymbol{\mu}', \boldsymbol{\Sigma} + \boldsymbol{\Sigma}').$$

Consider a normal random vector divided into two parts, $\mathbf{X} = (\mathbf{X}_A^T, \mathbf{X}_B^T)^T$. In this case $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ have to be split accordingly:

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{AA} & \boldsymbol{\Sigma}_{AB} \\ \boldsymbol{\Sigma}_{BA} & \boldsymbol{\Sigma}_{BB} \end{pmatrix}.$$

Basic properties of split normal random vectors are enunciated now.

2. $\mathbf{X}_A \in \mathcal{N}(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_{AA})$.
3. $\boldsymbol{\Sigma}_{AB} = \mathbf{0}$ iff $\mathbf{X}_A$ and $\mathbf{X}_B$ are independent.
4. $p(\mathbf{X}_A | \mathbf{X}_B = \mathbf{x}_B) \in \mathcal{N}(\boldsymbol{\mu}_{A|B}, \boldsymbol{\Sigma}_{A|B})$, where

$$\boldsymbol{\mu}_{A|B} = \boldsymbol{\mu}_A + \boldsymbol{\Sigma}_{AB} \boldsymbol{\Sigma}_{BB}^{-1} (\mathbf{x}_B - \boldsymbol{\mu}_B) \quad \text{and}$$

$$\boldsymbol{\Sigma}_{A|B} = \boldsymbol{\Sigma}_{AA} - \boldsymbol{\Sigma}_{AB} \boldsymbol{\Sigma}_{BB}^{-1} \boldsymbol{\Sigma}_{BA}.$$

1.3 Definition and properties of GMRFs

Gaussian Markov Random Fields, which are very useful for further study of spatial autoregressive models, will be presented. In this section, GMRFs wrt to a given labelled graph will be defined. Some of their properties, as well as an example and the specification through full conditionals, will be introduced. It will be shown that an intuitive approach based on graphs can be very useful to get information about precisions and conditional correlations between elements.

1.3.1 Definition and example

Let $\mathbf{X} = (X_1, \cdots, X_n)^T$ follow a normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph were $\mathcal{V} = \{1, \cdots, n\}$ and $\mathcal{E}$ the set of edges such that there is no edge between nodes i and j iff $X_i \perp X_j | \mathbf{X}_{-ij}$, where $\mathbf{X}_{-ij}$ denotes vector $\mathbf{X}$ except ith and jth components. Then, it is said that $\mathbf{X}$ is a GMRF wrt $\mathcal{G}$.

There is a direct connection between the graph $\mathcal{G}$ and the parameters of the normal distribution. It is known that the mean $\boldsymbol{\mu}$ does not have any influence on the pairwise conditional independence properties of $\mathbf{X}$, so this information can be deduced from the covariance matrix $\boldsymbol{\Sigma}$. The *precision matrix* $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$ is used for this, as can be noticed in the following theorem.

Theorem 1.2. *Lex $\boldsymbol{X} = (X_1, \cdots, X_n)^T$ be normal distributed with mean vector $\boldsymbol{\mu}$ and precision matrix $\boldsymbol{Q} > 0$. Then for X_i, X_j with $i \neq j$,*

$$X_i \perp X_j | \boldsymbol{X}_{-ij} \Leftrightarrow Q_{ij} = 0.$$

The formal definition of a GMRF is presented below.

Definition 1.3. *A random vector $\boldsymbol{X} = (X_1, \cdots, X_n)^T$ with support in $\mathbb{R}^n$ is called a GMRF wrt a labelled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with mean $\boldsymbol{\mu}$ and SPD precision matrix $\boldsymbol{Q}$ iff its density has the form*

$$p_{\boldsymbol{X}}(\boldsymbol{x}) = (2\pi)^{n/2} |\boldsymbol{Q}|^{1/2} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{Q} (\boldsymbol{x} - \boldsymbol{\mu})\right)$$

and

$$Q_{ij} \neq 0 \Leftrightarrow \{i, j\} \in \mathcal{E} \text{ for all } i \neq j.$$

If $\boldsymbol{Q}$ is a completely dense matrix then $\mathcal{G}$ is fully connected. This implies that any normal distribution with SPD covariance matrix is a GMRF and vice versa. Note that $\boldsymbol{\Sigma}$ is always SPD, so it admits a diagonalization with non negative eigenvalues, and therefore $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$ has also non negative eigenvalues and it is SPD. However, the nice properties of GMRFs are really useful when the precision matrices are sparse (that is, a matrix in which most of the elements are zero), the ones that are going to be used in this text.

The elements of $\boldsymbol{Q}$ provide much information about $\mathbf{X}$ as it can be seen in the next result. The proof can be seen in [6] (pp 23, 24).

Theorem 1.4. *Let $\boldsymbol{X}$ be a GMRF wrt $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with mean $\boldsymbol{\mu}$ and precision matrix $\boldsymbol{Q} > 0$, then*

- $\mathbb{E}(X_i | \boldsymbol{X}_{-i}) = \mu_i - \dfrac{1}{Q_{ii}} \sum_{j=j\sim i} Q_{ij}(X_j - \mu_j),\ i = 1, \cdots, n,$
- $\text{Prec}(X_i | \boldsymbol{X}_{-i}) = Q_{ii},\ i = 1, \cdots, n,$
- $\text{Corr}(X_i, X_j | \boldsymbol{X}_{-ij}) = -\dfrac{Q_{ij}}{\sqrt{Q_{ii} Q_{jj}}},\ i = 1, \cdots, n, i \neq j.$

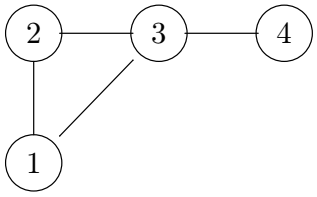


Figure 1.2: Example of a simple graph corresponding to a GMRF.

The interpretation of coefficients of $\boldsymbol{\Sigma}$ and $\boldsymbol{Q}$ are quite different. The covariance matrix gives information about the marginal variance of each component of $\mathbf{X}$ (diagonal) and the covariance between every two components of $\mathbf{X}$ (off−diagonal), while the precision matrix provides the conditional precisions of each component (diagonal) and conditional correlation between two components (off−diagonal).

A simple example of GMRF is going to be introduced to complete this interpretation of the coefficients. Observe the graph represented in Figure 1.2. A possible precision matrix for a GMRF wrt this graph is the following:

$$\boldsymbol{Q} = \begin{pmatrix} 1 & -1/2 & -1/2 & 0 \\ -1/2 & 1 & -1/2 & 0 \\ -1/2 & -1/2 & 1 & -1/2 \\ 0 & 0 & -1/2 & 1 \end{pmatrix}.$$

Applying Theorem 1.4 it is easy to deduce conditional expectancies, precisions and correlations of the components of the GMRFs. For example, taking $\boldsymbol{\mu} = \mathbf{0}_4$,

- $\mathbb{E}(X_2|\mathbf{X}_{-2}) = -\frac{1}{Q_{22}} \sum_{j=j\sim 2} Q_{2j}(X_j) = -\left(\frac{-X_1}{2} + \frac{-X_3}{2}\right) = \frac{X_1 + X_3}{2}$,
- $\text{Prec}(X_2|\mathbf{X}_{-2}) = 1$,
- $\text{Corr}(X_2, X_3|\mathbf{X}_{-23}) = -\frac{Q_{23}}{\sqrt{Q_{22}Q_{33}}} = \frac{1}{2}$.

Note that, in this example, conditional correlations between every pair of non-equal elements is either $\frac{1}{2}$, if they are neighbors, or 0, if they are not.

1.3.2 Properties of GMRFs

So far, it has been seen that if two nodes of $\mathcal{G}$ are neighbors, the corresponding off−diagonal entry of $\boldsymbol{Q}$ is not zero, and vice versa. It turns out that more information about conditional independence can be extracted from $\mathcal{G}$.

Theorem 1.5. *Let $\boldsymbol{X} = (X_1, \cdots, X_n)^T$ be a GMRF wrt $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Then the following statements are equivalent.*

1. *The pairwise Markov property:*

$$X_i \perp X_j | \boldsymbol{X}_{ij} \quad \text{if } \{i,j\} \notin \mathcal{E} \text{ and } i \neq j.$$

2. *The local Markov property:*

$$X_i \perp \boldsymbol{X}_{-\{i,ne(i)\}} | \boldsymbol{X}_{ne(i)} \quad \text{for every} \quad i \in \mathcal{V}.$$

3. *The global Markov property:*

$$\boldsymbol{X}_{\mathcal{A}} \perp \boldsymbol{X}_{\mathcal{B}} | \boldsymbol{X}_{\mathcal{C}}$$

for all disjoint sets $\mathcal{A}, \mathcal{B}$ and $\mathcal{C}$ where $\mathcal{C}$ separates $\mathcal{A}$ and $\mathcal{B}$, and $\mathcal{A}$ and $\mathcal{B}$ are non-empty.

Let $\mathcal{A}$ be a subset of $\mathcal{V}$. The conditional distribution for $\mathbf{X}_{\mathcal{A}} \subset \mathbf{X}$ given the rest $\mathbf{X}_{\overline{\mathcal{A}}}$ will be studied now. From now on, $\mathcal{B}$ is denoted by the complementary of $\mathcal{A}$, so

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_{\mathcal{A}} \\ \\ \mathbf{x}_{\mathcal{B}} \end{pmatrix}, \ \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{\mathcal{A}} \\ \\ \boldsymbol{\mu}_{\mathcal{B}} \end{pmatrix}, \ \boldsymbol{Q} = \begin{pmatrix} \boldsymbol{Q}_{\mathcal{A}\mathcal{A}} & \boldsymbol{Q}_{\mathcal{A}\mathcal{B}} \\ \\ \boldsymbol{Q}_{\mathcal{B}\mathcal{A}} & \boldsymbol{Q}_{\mathcal{B}\mathcal{B}} \end{pmatrix}.$$

The following is a powerful generalization of Theorem 1.5.

Theorem 1.6. *Let $\boldsymbol{X}$ be a GMRF wrt $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with mean $\boldsymbol{\mu}$ and precision matrix $\boldsymbol{Q} > 0$. Let $\mathcal{A} \subset \mathcal{V}$ and $\mathcal{B} = \mathcal{V} \backslash \mathcal{A}$, where $\mathcal{A}, \mathcal{B} \neq \emptyset$. The conditional distribution of $\boldsymbol{X}_{\mathcal{A}} | \boldsymbol{X}_{\mathcal{B}}$ is then a GMRF wrt the subgraph $\mathcal{G}^{\mathcal{A}}$ with mean $\boldsymbol{\mu}_{\mathcal{A}|\mathcal{B}}$ and precision matrix $\boldsymbol{Q}_{\mathcal{A}|\mathcal{B}} > 0$, where*

$$\boldsymbol{\mu}_{\mathcal{A}|\mathcal{B}} = \boldsymbol{\mu}_{\mathcal{A}} - \boldsymbol{Q}_{\mathcal{A}\mathcal{A}}^{-1} \boldsymbol{Q}_{\mathcal{A}\mathcal{B}} (\boldsymbol{X}_{\mathcal{B}} - \boldsymbol{\mu}_{\mathcal{B}})$$

and

$$\boldsymbol{Q}_{\mathcal{A}|\mathcal{B}} = \boldsymbol{Q}_{\mathcal{A}\mathcal{A}}.$$

This result implies that the conditional mean only depends on values of $\boldsymbol{\mu}$ and $\boldsymbol{Q}$ in $\mathcal{A} \cup ne(\mathcal{A})$, since Q_{ij} is zero unless $j \in ne(i)$. Its proof can be checked in [6] (pp 27).

Introduce now the *canonical parametrization* for a GMRF, different from the usual parametrization of a normal distribution.

Definition 1.7. *A GMRF $\boldsymbol{X}$ wrt $\mathcal{G}$ with canonical parameters $\boldsymbol{b}$ and $\boldsymbol{Q} > 0$ has density*

$$p(\boldsymbol{X}) \propto \exp\left(-\frac{1}{2} \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} + \boldsymbol{b}^T \boldsymbol{x} \right),$$

where $\propto$ denotes proportionality. The precision matrix is $\boldsymbol{Q}$ and the mean is $\boldsymbol{Q}^{-1}\boldsymbol{b}$. The canonical parametrization is denoted as

$$\boldsymbol{X} \sim \mathcal{N}_n^C(\boldsymbol{b}, \boldsymbol{Q}).$$

Therefore, the distribution $\mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{Q}^{-1})$ (usual parametrization) is equal to $\mathcal{N}_n^C(\boldsymbol{Q\mu}, \boldsymbol{Q})$. The link between this density and the one in Definition 1.3 is presented now.

On the one hand,

$$\exp\left(-\frac{1}{2}\mathbf{x}^T\boldsymbol{Q}\mathbf{x} + \mathbf{b}^T\mathbf{x}\right) = \exp\left(-\frac{1}{2}\mathbf{x}^T\boldsymbol{Q}\mathbf{x} + \boldsymbol{Q\mu}\mathbf{x}\right) = \exp\left(-\frac{1}{2}\mathbf{x}^T\boldsymbol{Q}\mathbf{x}\right)\exp\left(\boldsymbol{Q\mu}\mathbf{x}\right).$$

On the other hand,

$$\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{Q}(\mathbf{x}-\boldsymbol{\mu})\right) = \exp\left(-\frac{1}{2}\mathbf{x}^T\boldsymbol{Q}\mathbf{x}\right)\exp\left(\frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{Q}\mathbf{x}\right)\exp\left(\frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q\mu}\right)\exp\left(-\frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{Q\mu}\right).$$

However,

$$\exp\left(\frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{Q}\mathbf{x}\right)\exp\left(\frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q\mu}\right) = \exp\left(\boldsymbol{Q\mu}\mathbf{x}\right)$$

and $\exp\left(-\frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{Q\mu}\right)$ is a constant wrt $\mathbf{x}$.

1.3.3 Specification through full conditionals

Aside from specifying a GMRF by its mean and precision matrix, it can be specified as well through full conditionals $\{p(X_i|\mathbf{X}_{-i})\}_{i=1}^n$. However, some further assumptions on the full conditionals are required, so that it is sure that they correspond to a valid GMRF.

Theorem 1.8. *Let $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\boldsymbol{k} = (k_1, \ldots, k_n) \in (\mathbb{R}^+)^n$. Given n normal full conditionals $\{p(X_i|\boldsymbol{X}_{-i})\}_{i=1}^n$ with conditional mean*

$$\mathbb{E}(X_i|\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i}) = \mu - \sum_{j:j\sim i}\beta_{ij}(x_j - \mu_j),$$

where $j \sim i$ denotes that i and j are neighbors, and precision

$$Prec(X_i|\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i}) = k_i,$$

then $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ is a GMRF wrt labelled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with mean $\boldsymbol{\mu}$ and precision matrix $\boldsymbol{Q}$ where

$$Q_{ij} = \begin{cases} k_i\beta_{ij} & i \neq j \\ k_i & i = j \end{cases}$$

provided $k_i\beta_{ij} = k_j\beta_{ji},\ i \neq j$ and $\boldsymbol{Q} > 0$.

Chapter 2

Spatial Autoregressive Models

Recall that given a set of data, a linear regression model relates the n observations on the dependent random variable Y to k explanatory random variables $X_1, \ldots, X_k$ whose importance is given by k fixed effects and with error captured by the n disturbances $\boldsymbol{\epsilon} = \epsilon_1, \cdots, \epsilon_n$, that are supposed to be independent and equally distributed belonging to a normal with null mean.

Now suppose that the set of data is collected over space (maybe two dimensional space, easier to handle than three dimensional one). In many cases, it seems reasonable to think that the place of the individual from which the data is taken influences the dependent variable, but not as if it was another explanatory variable (imagine for example average temperature, that descends as going north) but because of proximity among individuals. Hence, the observations are no longer independent. For example, consider the language spoken by each small community on a multilingual region. Of course explanatory variables as population density where each community lives may be very significant. However, it seems obvious that spatial proximity between communities is also relevant.

In order to model this influence, conditional autoregressive (CAR) and simultaneous autoregressive (SAR) models are introduced in this chapter. Note that SAR is used as the acronym of Spatial Autoregression in many publications so this may lead to confusion. In this context, individuals refer to sample units. In the previous example, the individuals are the small communities of speakers spatially related to each other when close in space.

Different theoretical approaches to conditional autoregressive and simultaneous autoregressive models, as well as more further information about them and other alternative models, can be found in [1] (pp 69-87), [4] (pp 203-253) and [3] (pp 245-260).

2.1 Conditional Autoregressive Model

2.1.1 Introduction to the model

Conditional autoregressive model (CAR) can be seen as an analogous of multivariate linear regression modifying the distribution of errors.

Recall that, given n realizations of one response variable Y and k explanatory variables $X_1, \ldots, X_k$, a linear regression model for Y over $X_1, \ldots, X_k$, under the usual assumptions of independence, homocedasticity and errors normality, can be summarized in distribution terms as follows,

$$\mathbf{Y} \in \mathcal{N}_n(\boldsymbol{\mathcal{X}}\boldsymbol{\gamma}, \sigma^2 \mathbf{I}_n), \tag{2.1}$$

where the design matrix is given by

$$\boldsymbol{\mathcal{X}} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix},$$

being $x_{im}, i = 1, \ldots, n$ a realization of the random variable $X_m, m = 1, \ldots, k$. Note that the column of ones is only included if an intercept is needed into de model.

Therefore, in linear regression models the residuals follow the normal distribution

$$\boldsymbol{\epsilon} \in \mathcal{N}_n(\mathbf{0}_n, \sigma^2 \mathbf{I}_n).$$

However, in CAR, residuals are no longer independent and they follow the distribution below,

$$\boldsymbol{\epsilon} \in \mathcal{N}_n(\mathbf{0}_n, \sigma^2 \boldsymbol{\Omega}),$$

where $\boldsymbol{\Omega} = (\mathbf{I}_n - \phi \mathbf{W})^{-1}$. The $n \times n$ matrix $\mathbf{W}$ is called *spatial weight matrix*. This matrix contains all the information about the spatial dependence structure. The CAR spatial weight matrix is as follows. If individuals i and j are spatially related, $W_{ij} = W_{ji} > 0$. Otherwise $W_{ij} = W_{ji} = 0$, so $\mathbf{W}$ is always symmetric. Diagonal weights are set to zero. The scalar parameter $\phi \in \mathbb{R}^+$ provides a measure of the intensity of the influence that the spatial relation has into the model.

Therefore, given a set of n spatial points, the considered response is a n-dimensional random vector $\mathbf{Y} = (Y_1, \ldots, Y_n)$ and it is modeled by CAR as

$$\mathbf{Y} \in \mathcal{N}_n(\boldsymbol{\mathcal{X}}\boldsymbol{\gamma}, \sigma^2 (\mathbf{I}_n - \phi \mathbf{W})^{-1}). \tag{2.2}$$

This is in fact a particular case of a GMRF, that have been introduced in Chapter 1. The mean vector would be

$$\boldsymbol{\mu} = \boldsymbol{\mathcal{X}}\boldsymbol{\gamma}$$

and the precision matrix

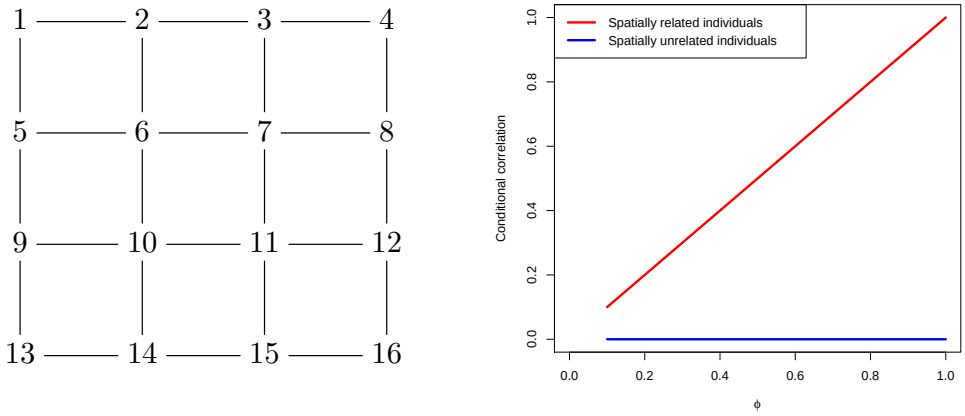
$$\mathbf{Q} = \frac{\mathbf{I}_n - \phi \mathbf{W}}{\sigma^2}.$$

Hence, applying Theorem 1.8, it can be seen that $k_i = k$ for all $i = 1, \ldots, n$, $\sigma^2 = 1/k$ and $\beta_{ij} = -\phi W_{ij}$ when $i \neq, j,\ i, j = 1, \ldots, n$. Therefore, the precision matrix in a CAR model is

$$\mathbf{Q} = k \cdot \begin{pmatrix} 1 & \beta_{12} & \cdots & \beta_{1n} \\ \beta_{12} & 1 & \cdots & \beta_{2n} \\ \vdots & \vdots & & \vdots \\ \beta_{1n} & \beta_{2n} & \cdots & 1 \end{pmatrix},$$

where $k = 1/\sigma^2$ and $\beta_{ij} = -\phi W_{ij}$.

Take a simple example in order to illustrate the effect of ϕ. Figure 2.1a represents 16 individuals (each one labelled with a natural number) and the spatial relation between two of them with a segment



(a) (b)

Figure 2.1: Plot (a) represents 16 individuals with a spatial structure. A segment joining two of them indicates that they are spatially related. Plot (b) shows the effect of ϕ on conditional correlation.

connecting them. Let $\mathbf{Y} = (Y_1, \ldots, Y_{16})$ be a random 16-dimensional vector generated by a CAR model, that is to say,

$$\mathbf{Y} \in \mathcal{N}_{16}(\boldsymbol{\mathcal{X}}\boldsymbol{\gamma}, \sigma^2(\mathbf{I}_{16} - \phi\mathbf{W})^{-1}).$$

The mean of the random vector, $\boldsymbol{\mathcal{X}}\boldsymbol{\gamma}$, is irrelevant to study the importance of ϕ, so it is necessary to focus on the covariance matrix, $\sigma^2(\mathbf{I}_{16} - \phi\mathbf{W})^{-1}$. The spatial weight matrix, $\mathbf{W}$, is built in the simplest way: $W_{ij} = W_{ji} = 1$ if individuals i and j are spatially related and $W_{ij} = W_{ji} = 0$ otherwise, where $i,\ j \in \{1, \ldots, n\}$.

The precision matrix is calculated easily:

$$\mathbf{Q} = \frac{1}{\sigma^2}\ \mathbf{Q}_0,$$

where

$$Q_{0ij} = \begin{cases} 1 & i = j \\ -\phi & i \neq j,\ i \sim j \\ 0 & \text{otherwise.} \end{cases}$$

Recall that

$$\text{Corr}(X_i, X_j | \mathbf{X}_{-ij}) = -\frac{Q_{ij}}{\sqrt{Q_{ii}Q_{jj}}},\ i = 1, \cdots, n, i \neq j,$$

so, taking $\sigma^2 = 1$, $\text{Corr}(X_i, X_j|\mathbf{X}_{-ij}) = \phi$ if i and j are spatially related, and the conditional correlation is 0 otherwise. As ϕ varies, conditional correlations change. This effect is shown in Figure 2.1b.

2.1.2 Estimation by Maximum Likelihood

In this subsection, the Maximum Likelihood (ML) estimation method for CAR parameters will be introduced. A detailed approach can be found in [4] (pp 203-253).

Given a sample realization $\mathbf{y}$ of a n-dimensional vector $\mathbf{Y}$, $\mathbf{W}$ a weight matrix and $\boldsymbol{\mathcal{X}}$ a design matrix, the log-likelihood of the model (2.2) is

$$l(\boldsymbol{\gamma}, \sigma^2, \phi) = C - \frac{1}{2}\log|\mathbf{I}_n - \phi\mathbf{W}| - \frac{n}{2}\log(\sigma^2) - \frac{(\mathbf{y} - \boldsymbol{\mathcal{X}}\boldsymbol{\gamma})^T(\mathbf{I}_n - \phi\mathbf{W})(\mathbf{y} - \boldsymbol{\mathcal{X}}\boldsymbol{\gamma})}{2\sigma^2},$$

being C a constant. It can be directly obtained from de density function of the multivariate normal.

The estimators of the parameters of the model are found by maximizing $l(\boldsymbol{\gamma}, \sigma^2, \phi)$. The estimations for $\boldsymbol{\gamma}, \sigma^2, \phi$ will be denoted by $\widehat{\boldsymbol{\gamma}}, \widehat{\sigma^2}, \widehat{\phi}$ respectively.

The objective function that has to be maximized is 3-dimensional, so specific tools for multi-dimensional optimization are needed. The problem can be approached from two alternatives: first it can be solved directly (for instance, using steepest gradient descent methods or quasi-newton methods such as BFGS), but it can be computationally expensive, so another option is to consider Generalized Least Squares (GLS), that allows to accelerate the calculation of the estimators.

The GLS method is a variation of Ordinary Least Squares (OLS). It introduces the covariance matrix in the formulation of the estimators, and then it applies a transformation on the data using this matrix, allowing the implementation of an OLS procedure. There is still a optimization process that will be approached using a Newton modified method. As the involved model is linear, the estimators obtained with GLS and ML are the same.

Recall that, under some regularity conditions, the ML estimators are consistent, asymptotically unbiased and asymptotically normal. The following conditions are needed to ensure consistency: the model has to be identifiable, the parameter space has to be compact and the likelihood function needs to exhibit upper semi-continuity and dominance. Normality requires that the model does not suffer from the following issues: estimation on boundary, data boundary parameter-dependence, increasing number of parameters or increasing information. In CAR model all the conditions can be assumed to be right, so the introduced estimators exhibit consistency and asymptotic normality.

2.1.3 Simulations

The study of CAR model finishes with three simulations of 100 data vectors obtained from a CAR model with $\phi = 0.3$ and $\sigma^2 = 1$ in the first case and $\phi = 0.2$ and $\sigma^2 = 1$ in the second and third case. The number of individuals will be 16 in the first simulation, exactly with the same spatial structure that was taken in the example of Subsection 2.1.1, while the second and the third simulation will be performed in 10×10 and 20×20 dimensional regular lattices respectively. The weight matrix are analogous to the $\mathbf{W}$ that was introduced in that example. These three simulations will be performed taking two different mean vectors: a null vector in the first case and a vector that allocates as the mean in each point the sum of its two coordinates in the second case. The explanatory variables considered in this last one are two: the coordinate of the considered point in the horizontal axis and in the vertical axis. Therefore, the intercept is null, so

$$\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2)^T = (0, 1, 1)^T.$$

A CAR model is estimated with each data vector. This means that given the data and the spatial structure, 100 estimations for the two parameters ϕ and σ are found by ML when taking null mean vector. When this vector is not null, 100 estimations are computed for three more parameters: γ_0, γ_1 and γ_2. This method is implemented in the R function spautolm, from the package spdep, as it is described in Appendix B.

Variance, bias and mean squared error (MSE) of the estimators of the parameters will be estimated for each of the simulated scenarios, and at the end it will be checked if the estimators are asymptotically unbiased and consistent, i.e, if when the number of points gets bigger the expectation of the estimator converges to the real parameter and its variance converges to zero.

Bias is estimated by averaging the difference between the real value of the parameter and each of its estimations, while the estimation of MSE is obtained by averaging these differences squared. Recall

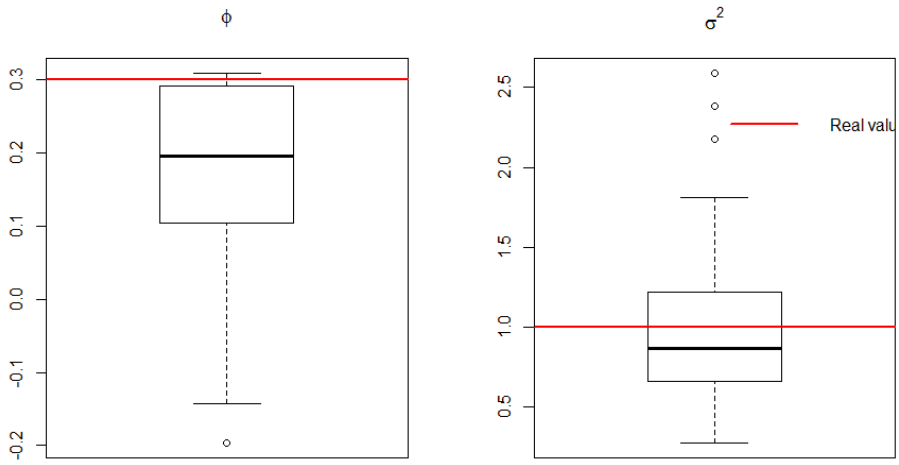


Figure 2.2: Boxplots of the Maximum Likelihood estimations for ϕ and σ and with their real values obtained from 100 simulated data vectors in the 16 points lattice.

that, being θ a parameter and $\widehat{\theta}$ an estimator, $MSE(\widehat{\theta}) = Var(\widehat{\theta}) + (Bias(\widehat{\theta}))^2$, so the variance is estimated from this equation.

Null mean vector

A boxplot of the estimations of each of the two parameters for the 16 points lattice can be observed in Figure 2.2. Note the apparent large bias of $\widehat{\phi}$, as almost all of the estimated values are below the real value of ϕ. In Figure 2.3, histograms of the same estimations are shown. It seems to be clear that $\widehat{\phi}$ does not exhibit normality at all in this simulated scenario, while $\widehat{\sigma^2}$ maybe does.

The complete results are shown in Table 2.1. It seems that $\widehat{\phi}$ is a biased estimator for small lattices, but as the lattice gets larger the bias converges to zero, so it is asymptotically unbiased, as

	$\widehat{\phi}$			$\widehat{\sigma^2}$		
n	$\widehat{Bias}$	$\widehat{Var}$	$\widehat{MSE}$	$\widehat{Bias}$	$\widehat{Var}$	$\widehat{MSE}$
16	-0.1209	0.0288	0.0435	-0.0257	0.1829	0.1835
100	-0.0214	0.0029	0.0034	0.0163	0.0212	0.0215
400	-0.0086	0.0006	0.0007	0.0103	0.0063	0.0064

Table 2.1: Estimated bias, variance and mean squared error of the estimators of the CAR model parameters ϕ and σ^2 obtained for each lattice by 100 simulations when taking a null mean vector.

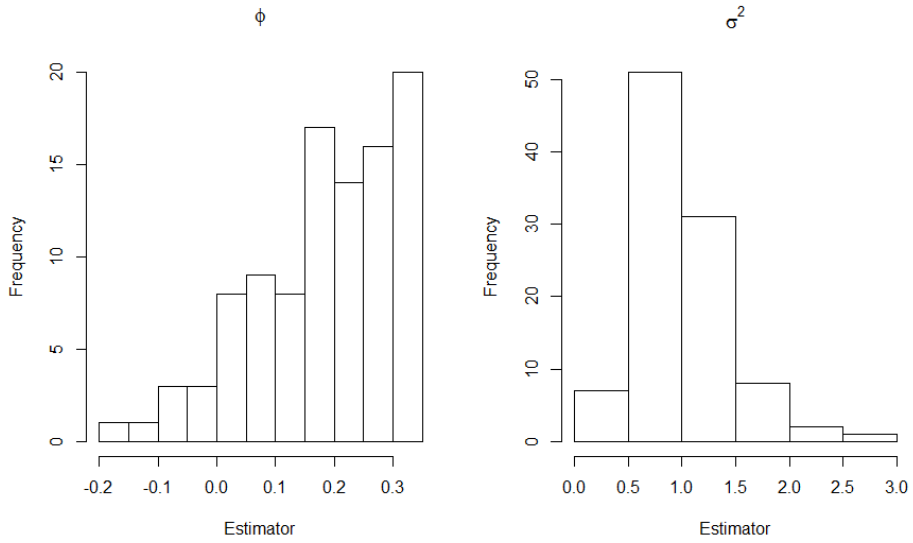


Figure 2.3: Histograms of the ML estimations for ϕ and σ obtained from 100 data vectors in the 16 points lattice.

it was expected since it is a ML estimator. Its variance converges to zero too, so the consistency of $\widehat{\phi}$ appears to be confirmed too. Finally, the study of the normality of $\widehat{\phi}$ is going to be approached. The p-values obtained from de Lilliefors normality test (a variation of the basic Kolmogorov-Smirnov test) are shown in the left part of Table 2.4. It seems that as the lattice gets larger, the estimations became normal, something expected since ML estimators are asymptotically normal. These properties of ML estimators were introduced in Subsection 2.1.2.

With respect to $\widehat{\sigma^2}$, it seems that both unbiasedness and consistency are plausible, as well as asymptotic normality.

Not null mean vector

Results for this case are shown in Table 2.2 and in Table 2.3. It can be observed that the estimators of the regression parameters exhibit good properties: they seem to be consistent and not only asymptotically unbiased but unbiased for the designs considered.

The estimators of ϕ and σ^2 appear to behave similarly as in the null mean vector case. Both of them appear to be consistent, while $\widehat{\sigma^2}$ seems unbiased and $\widehat{\phi}$ only asymptotically unbiased.

Studying normality by the p-values shown in the right part of Table 2.4, it can be concluded that all estimators exhibit asymptotic normality as in the null mean vector case. Moreover, the estimators $\widehat{\gamma_0}$, $\widehat{\gamma_1}$ and $\widehat{\gamma_2}$ are accepted to be normal even in small lattices.

2.2 Simultaneous Autoregressive Model

2.2.1 Introduction to the model

Simultaneous autoregressive model (SAR) employs a different specification for error covariance with respect to CAR. Given a n-dimensional dependent random vector $\mathbf{Y}$ and a design matrix $\boldsymbol{\mathcal{X}}$ consisting

	$\widehat{\gamma_0}$			$\widehat{\gamma_1}$			$\widehat{\gamma 2}$		
n	$\widehat{Bias}$	$\widehat{Var}$	$\widehat{MSE}$	$\widehat{Bias}$	$\widehat{Var}$	$\widehat{MSE}$	$\widehat{Bias}$	$\widehat{Var}$	$\widehat{MSE}$
16	-0.0304	1.7443	1.7452	-0.0060	0.1816	0.1816	0.0572	0.1282	0.1315
100	-0.0111	0.1881	0.1882	-0.0019	0.0026	0.0026	-0.0013	0.0039	0.0039
400	-0.0248	0.0537	0.0543	0.0008	0.0002	0.0002	0.0008	0.0003	0.0003

Table 2.2: Estimated bias, variance and mean squared error of the estimators of the CAR model parameters γ_0, γ_1 and γ_2 obtained for each lattice by 100 simulations when taking a null mean vector.

	$\widehat{\phi}$			$\widehat{\sigma^2}$		
n	$\widehat{Bias}$	$\widehat{Var}$	$\widehat{MSE}$	$\widehat{Bias}$	$\widehat{Var}$	$\widehat{MSE}$
16	-0.2642	0.1039	0.1737	0.2049	0.1571	0.1991
100	-0.0401	0.0043	0.0059	-0.0075	0.0201	0.0201
400	-0.0133	0.0006	0.0008	0.0081	0.0064	0.0064

Table 2.3: Estimated bias, variance and mean squared error of the estimators of the CAR model parameters obtained for each lattice by 100 simulations when taking a non null mean vector.

	Null mean vector		Not null mean vector				
n	$\widehat{\phi}$	$\widehat{\sigma^2}$	$\widehat{\gamma_0}$	$\widehat{\gamma_1}$	$\widehat{\gamma_2}$	$\widehat{\phi}$	$\widehat{\sigma^2}$
16	< 0.001	0.001	0.7560	0.2830	0.2868	0.0101	0.0420
100	0.3015	0.0736	0.5759	0.5928	0.9016	< 0.001	0.5742
400	0.2031	0.4370	0.1951	0.7096	0.5153	0.4079	0.4457

Table 2.4: P-values of the Lilliefors (Kolmogorov-Smirnov) normality test ran over the estimators of the CAR model parameters for each lattice by 100 simulations.

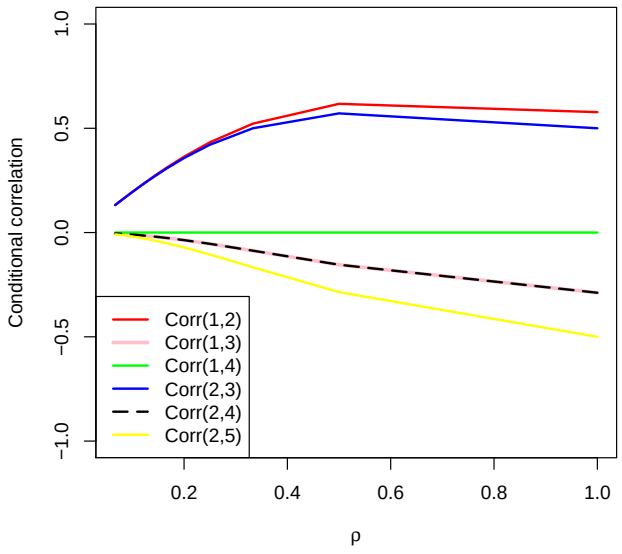


Figure 2.4: Effect of ρ on conditional correlation between some pairs of individuals.

on the observations of k explanatory random variables $X_m, m = 1, \ldots, k$ (adding a first column of ones), SAR model considers

$$\mathbf{Y} \in \mathcal{N}_n(\boldsymbol{\mathcal{X}}\gamma, \sigma^2\boldsymbol{\Omega}), \tag{2.3}$$

where in this case, the matrix $\boldsymbol{\Omega}$ is given by

$$\boldsymbol{\Omega} = [(\mathbf{I}_n - \rho\mathbf{W})^T(\mathbf{I}_n - \rho\mathbf{W})]^{-1},$$

where ρ is a scalar parameter and $\mathbf{W}$ is the SAR spatial weight matrix. SAR $\mathbf{W}$ is not necessarily symmetric whereas CAR spatial weight matrix satisfy this condition. This allows to use non symmetric influences between two individuals.

Therefore, the precision matrix for SAR is

$$\boldsymbol{Q} = \frac{(\mathbf{I}_n - \rho\mathbf{W})^T(\mathbf{I}_n - \rho\mathbf{W})}{\sigma^2}.$$

Its explicit form cannot be given in a simple way in terms of $\mathbf{W}$ as it can be done for CAR. Take the same naive example that was studied in the previous subsection to guess how ρ influences on the conditional dependence structure and to find $\boldsymbol{Q}$ for this specific case. The effect of the parameter ρ on conditional correlations between six pairs of individuals is shown in Figure 2.4. The weight matrix chosen for this example is the same taken for CAR, so it is a symmetric one.

Taking, for example, $\rho = 0.3$, a general insight of the precision matrix is given by Figure 2.5. Observe that this matrix has many null elements, which will be useful in posterior high-dimensional models.

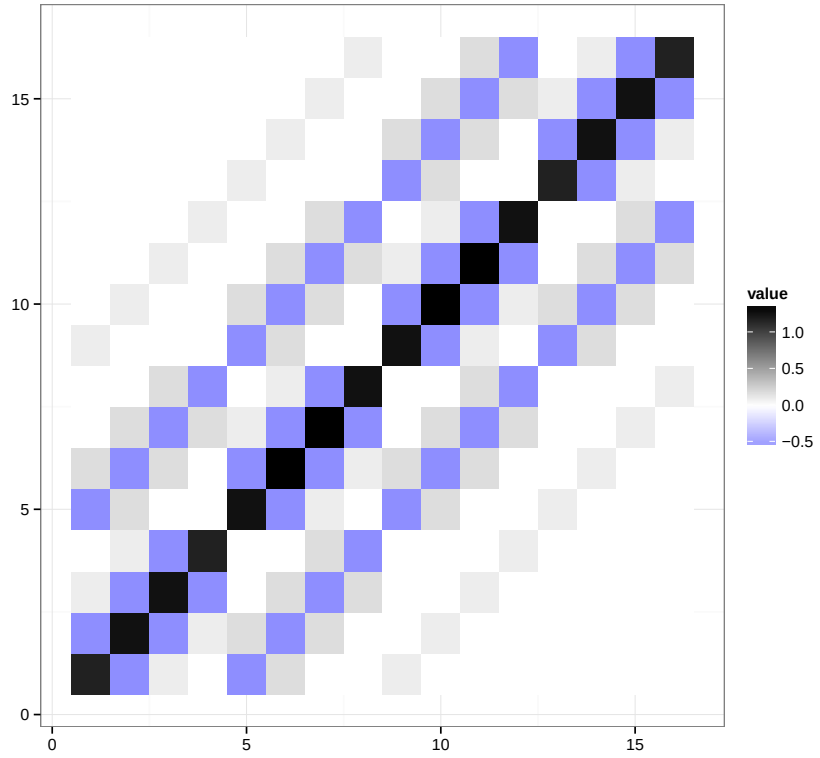


Figure 2.5: Representation of the precision matrix for the given example. White blocks correspond to null elements. Note that the elements of the matrix are placed in a non-canonical way.

2.2.2 Estimation by Maximum Likelihood

The ML estimation method for SAR will be introduced in this subsection. The log-likelihood is analogous to the log-likelihood of CAR shown in Subsection 2.1.2.

Given a sample realization $\mathbf{y}$ of a n-dimensional vector $\mathbf{Y}$ and $\mathbf{W}$, $\boldsymbol{\mathcal{X}}$ (the weight and the design matrix, respectively), the log-likelihood becomes

$$l(\boldsymbol{\gamma}, \sigma^2, \rho) = C - \frac{1}{2}\log|(\mathbf{I}_n - \rho\mathbf{W})^T(\mathbf{I}_n - \rho\mathbf{W})| - \frac{n}{2}\log(\sigma^2) - \frac{(\mathbf{y} - \boldsymbol{\mathcal{X}}\boldsymbol{\gamma})^T(\mathbf{I}_n - \rho\mathbf{W})^T(\mathbf{I}_n - \rho\mathbf{W})(\mathbf{y} - \boldsymbol{\mathcal{X}}\boldsymbol{\gamma})}{2\sigma^2},$$

where C is a constant. It is obtained from the density function of the SAR model, the following multivariate normal:

$$\mathcal{N}_n(\boldsymbol{\mathcal{X}}\boldsymbol{\gamma}, \sigma^2[(\mathbf{I}_n - \rho\mathbf{W})^T(\mathbf{I}_n - \rho\mathbf{W})^{-1}]).$$

The estimators of the parameters of the model are found by maximizing $l(\boldsymbol{\gamma}, \sigma^2, \rho)$. The estimations for $\boldsymbol{\gamma}, \sigma^2, \rho$ will be denoted by $\widehat{\boldsymbol{\gamma}}, \widehat{\sigma^2}, \widehat{\rho}$ respectively.

The optimization problem is solved analogously to how it was solved in Subsection 2.1.2. As it was explained in that subsection, the estimators $\widehat{\boldsymbol{\gamma}}, \widehat{\sigma^2}, \widehat{\rho}$ are consistent, asymptotically unbiased and asymptotically normal.

2.2.3 Simulations

As in the previous subsection, this description of SAR models ends with three simulations, each of them of one hundred data vectors, in lattices of dimensions 4×4, 10×10 and 20×20 taking parameters $\rho = 0.3$, $\sigma^2 = 1$ in the first case and $\rho = 0.2$, $\sigma^2 = 1$ in the second and third cases. The spatial structure of the smallest lattice is given by the same weight matrix $\mathbf{W}$ that was used in the example of Subsection 2.1.1, and for the bigger lattices $\mathbf{W}$ are analogous. This three simulations will be performed taking the two different mean vectors that were used in Section 2.1.3. Therefore, in the non null mean vector case, the regression parameters are

$$\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2)^T = (0, 1, 1)^T.$$

For each one of the simulated data vectors the estimations of both the parameters are found using ML method. The R function used is again spautolm, in which SAR or CAR can be chosen as arguments, as it is explained in Appendix B. Recall that the estimators are consistent, asymptotically normal and asymptotically unbiased, since they are obtained by ML method under appropiate conditions.

Null mean vector

The complete results obtained are shown in Table 2.5 and in the left part of Table 2.8. Observing the case of the 16 points lattice, the bias of the estimations of ρ seems to be relevantly lower, in absolute value, than the one of the estimator of the scalar parameter ϕ in CAR (check Subsection 2.1.3). However, the bias of the estimations of σ^2 are slightly higher, also taking the absolute value. Figure 2.6 shows the boxplots of the 100 parameter estimators $\widehat{\rho}$ and $\widehat{\sigma^2}$. The Lilliefors normality test provides a really small p-value for the estimator of ρ, so normality is rejected, as it was for $\widehat{\phi}$ in CAR. On the other hand, there are no evidences against normality of σ^2.

Observing now the results for the three lattices, the variance of the estimations of ρ appears to be also much lower than the ones of ϕ in CAR, while de variances of the estimations of σ^2 may be similar in both models. Both estimators appear to be asymptotically normal.

In conclusion, it seems that $\widehat{\rho}$ is a biased estimator, even though with a much smaller bias than $\widehat{\phi}$ in CAR. It is also asymptotically normal, asymptotically unbiased and consistent, as it was expected by the fact that it is a ML estimator.

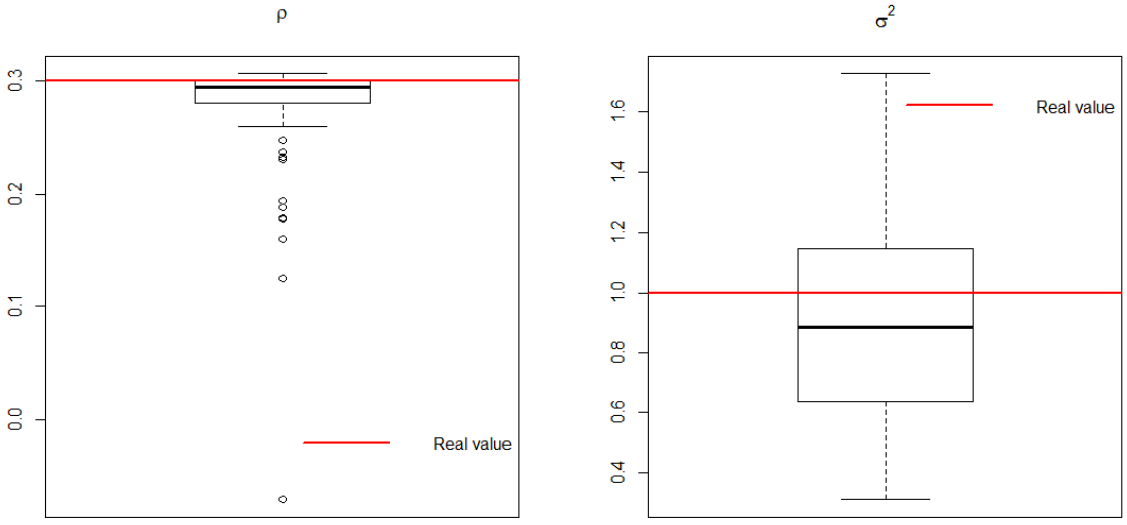


Figure 2.6: Boxplots of the Maximum Likelihood estimations for ρ and σ and with their expected values obtained from 100 simulated data vectors in the 16 point lattice.

	$\widehat{\rho}$			$\widehat{\sigma^2}$		
n	$\widehat{Bias}$	$\widehat{Var}$	$\widehat{MSE}$	$\widehat{Bias}$	$\widehat{Var}$	$\widehat{MSE}$
16	-0.0204	0.0027	0.0032	-0.0856	0.1227	0.1300
100	-0.0100	0.0007	0.0008	-0.0293	0.0162	0.0171
400	-0.0007	0.0001	0.0001	-0.0086	0.0049	0.0049

Table 2.5: Estimated bias, variance and mean squared error of the estimators of the SAR model parameters ρ and σ^2 obtained for each lattice by 100 simulations when taking a null mean vector.

	$\widehat{\gamma_0}$			$\widehat{\gamma_1}$			$\widehat{\gamma 2}$		
n	$\widehat{Bias}$	$\widehat{Var}$	$\widehat{MSE}$	$\widehat{Bias}$	$\widehat{Var}$	$\widehat{MSE}$	$\widehat{Bias}$	$\widehat{Var}$	$\widehat{MSE}$
16	-0.1792	2.9433	2.9754	0.0230	0.3125	0.3130	0.0285	0.3078	0.3087
100	-0.0199	0.5146	0.5150	-0.0001	0.0078	0.0078	-0.0076	0.0084	0.0084
400	0.0083	0.1711	0.1711	0.0017	0.0009	0.0009	0.0011	0.0008	0.0008

Table 2.6: Estimated bias, variance and mean squared error of the estimators of the SAR model parameters γ_0, γ_1 and γ_2 obtained for each lattice by 100 simulations when taking a not null mean vector.

	$\widehat{\rho}$			$\widehat{\sigma^2}$		
n	$\widehat{Bias}$	$\widehat{Var}$	$\widehat{MSE}$	$\widehat{Bias}$	$\widehat{Var}$	$\widehat{MSE}$
16	-0.0447	0.0115	0.0135	-0.1637	0.1465	0.1733
100	-0.0187	0.0012	0.0015	-0.0067	0.0211	0.0211
400	-0.0048	0.0001	0.0001	0.0067	0.0064	0.0064

Table 2.7: Estimated bias, variance and mean squared error of the estimators of the SAR model parameters ρ and σ^2 obtained for each lattice by 100 simulations when taking a not null mean vector.

It seems that the estimator of the σ^2 is consistent, normal and unbiased, so it has even better properties than its analogous in CAR.

Not null mean vector

The results are shown in Table 2.6, in Table 2.7 and in the right part of 2.8.

The estimators of the regression parameters appear to exhibit good properties: they seem to be normal, unbiased and consistent.

The estimators of ρ and σ^2 appear to behave similarly as in the null mean vector case. Consistency can be accepted for both, while only $\widehat{\sigma^2}$ is unbiased and $\widehat{\rho}$ has a small bias. However, it seems to be asymptotically unbiased. Therefore, these estimators exhibit the properties ensured by ML method, and perhaps some of them stronger ones: $\widehat{\gamma_0}, \widehat{\gamma_1}, \widehat{\gamma_2}$ and σ^2 appear to be normal.

2.3 Exploratory analysis for autoregressive models

When working with data it appears to be really helpful to begin the analysis by developing some exploratory analysis. If the data are spatially related, as the examples that will be analyzed in this

	Null mean vector		Not null mean vector				
n	$\widehat{\rho}$	$\widehat{\sigma^2}$	$\widehat{\gamma_0}$	$\widehat{\gamma_1}$	$\widehat{\gamma_2}$	$\widehat{\rho}$	$\widehat{\sigma^2}$
16	<0.001	0.0988	0.4965	0.6127	0.8950	<0.001	0.0984
100	0.1170	0.7438	0.5490	0.2771	0.0370	0.0192	0.6481
400	0.2541	0.8262	0.3666	0.7381	0.5118	0.9745	0.4056

Table 2.8: P-values of the Lilliefors (Kolmogorov-Smirnov) normality test ran over the estimators of the SAR model parameters for each lattice by 100 simulations.

manuscript, some previous steps should be considered.

First of all, plots are a fast way to get some insight of the data. A simple representation can be obtained by dividing the spatial area into individuals and colouring each one depending on the value of the corresponding response variable. If a plain regression model is fitted before the autoregressive modeling, an analogous plot with the residuals of this model can be helpful to catch the possible spatial dependence.

After that, a spatial weight matrix has to be built to capture the spatial structure of the data. There are different ways to face this. A wider approach to the construction of this matrix can be checked in [1] (pp 69-87) and in [3] (pp 245-260). The chosen spatial structure for data analysis in Chapter 3 is the simplest one, that has already been used in Subsections 2.1.1 and 2.2.1: $W_{ij} = W_{ji} = 1$ if individuals i and j are spatially related and $W_{ij} = W_{ji} = 0$ otherwise.

However, in this case there are several ways to decide whether two individuals are spatial related or not. Two approaches will be used in this thesis: as the individuals are spatial regions, they can share borders (and one may consider them as neighbors) or the neighboring structure may be constructed taking distances from regions centroids. In the second case a concrete distance has to be specified in order to decide if two individuals are close or not. From now on the first of this structures will be referred as "border spatial structure" while the other will be called "distance spatial structure".

Finally, a measure of spatial dependence is going to be introduced: Moran's I. This is the spatial analogue of the statistic for measuring association in time series: the lagged autocorrelation coefficient. The reader is referred to [1] (pp 71,72), although a brief introduction will be included here. The Moran's I statistic is defined as:

$$I = \frac{n \sum_i \sum_j W_{ij}(Y_i - \mathbb{E}(\mathbf{Y})(Y_j - \mathbb{E}(\mathbf{Y}))}{(\sum_{i \neq j} W_{ij}) \sum_i (Y_i - \mathbb{E}(\mathbf{Y}))^2},$$

where $\mathbf{Y} = (Y_1, \ldots, Y_n)$ and $\mathbf{W} = (W_{ij})$ were introduced in Subsection 2.1.1. Summation over i and j is done over all the observations.

Under the null model where the $Y_i,\ i = 1, \ldots, n$ are i.i.d., I is asymptotically normally distributed with mean $-1/(n-1)$ and variance of the form

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

where $S_0 = \sum_{i \neq j} W_{ij},\ S_1 = (1/2) \sum_{i \neq j} (W_{ij} + W_{ji})^2$ and $S_2 = \sum_k (\sum_j W_{kj} + \sum_i W_{ik})^2$.

Knowing the asymptotic distribution of the Moran's I under independence it is easy to run a test on the significance of the spatial autocorrelation which will be applied to real data.

To finish this section, this test is applied to data simulated similarly to the simulations in Subsection 2.1.3 and Subsection 2.2.3, taking only the case of null mean vector. As in those simulations, three

	CAR			SAR		
$n\backslash$ significance	**0.05**	**0.01**	**0.001**	**0.05**	**0.01**	**0.001**
16	0.53	0.29	0.03	0.95	0.82	0.11
100	0.90	0.76	0.52	1.00	1.00	1.00
400	1.00	1.00	1.00	1.00	1.00	1.00

Table 2.9: Proportion of times that the null hypothesis, non-existence of spatial autocorrelation, is rejected by the Moran's test with different levels of significance, taking 100 simulations of data generated by CAR and SAR models.

lattices are considered consisting on 16, 100 and 400 points. The data generation process uses the same spatial structure as in that subsections and both models SAR and CAR. The proportion of times that the null hypothesis of the test is rejected in each case is shown in Table 2.9. Note that, clearly, when the amount of points in the lattice is large, the test is more reliable. Moreover, it seems that SAR model generation process provides much clearer spatial autocorrelation.

The Moran's test was ran using the function lm.morantest of the R package spdep. For this purpose, a spatial structure has to be built. The details of how to build it are specified in Appendix B.

Chapter 3

Application to linguistic data

In this chapter specific spatial data will be studied with the models that have been introduced in Chapter 2. This data are obtained from polls about the use of one of the two main languages spoken on a territory. The spatial regions considered are two autonomous communities in Spain, Galicia and the Basque County, each of them divided in municipalities. The percentage of speakers of galician language or basque language will be taken as the dependent variable, and the possible influence of some explanatory variables as well as the spatial dependence structure will be evaluated in CAR and SAR models. Using the notation of Chapter 2, $\mathbf{Y} = (Y_1, \ldots, Y_n)$ is a random vector that contains the percentage of speakers of the considered language in each of the n municipalities, while $\mathbf{X}_k = (X_{1k}, \ldots, X_{nk}),\ k = 1, \ldots, m$ are the m random vectors that contain the explanatory variables. The realizations of the dependent variables constitute the design matrix $\boldsymbol{\mathcal{X}}$, that was introduced in Section 2.1.1 and appears in the modeling of CAR and SAR in equations (2.2) and (2.3) respectively.

3.1 Galician language

The data for the dependent variable are taken from the Galician Statistics Institute (IGE)[1]. This data consist on the percentage of population in each municipality that speak always in galician. They were obtained in a survey performed around 1 November 2011. The purpose of this survey was to get information about knowledge and use of galician language by the whole galician population. This kind of surveys are conducted from time to time because the galician public authorities plan their institutional and educational policies basing on this data. Spatial autoregressive models may be useful to explain the obtained results of the survey in each area because it is known that language uses are "smooth" over the territory, that is to say, close places exhibit similar behaviors. First of all, a exploratory analysis on the data is performed.

3.1.1 Exploratory analysis and regression model

A representation of this data is shown in Figure 3.1 (top). Observe that the percentages are very variable over the territory. In fact, the minimum is 12.20%, corresponding to Ferrol (A Coruña), and the maximum is 97.24%, corresponding to San Xoán de Río (Ourense), so there is a very wide range of values. Observe also that this variable seems not to be equally distributed over the galician geography, but forming some clusters of many close to each other municipalities. This fact motivates the further use of autoregressive models.

[1] Data were extracted from the web of the IGE (http://www.ige.eu/web/mostrar_actividade_estatistica.jsp?idioma=gl&codigo=0206002), in particular from the information relative to the population and housing census of year 2011, in the section of information related to the use and knowledge of galician language.

A reasonable assumption is that the more countrified a community is, the higher the use of galician language on it. Those were historically isolated areas where little Spanish arrived for a long time. The complete characterization of rural areas is very complex, but a good approach is to consider the population density of the municipalities. This variable is obviously positive and also very asymmetric, as it is shown in Figure 3.2 (left). Therefore, a logarithmic transformation is performed on this variable obtaining a quite symmetric variable better for modeling the data, whose histogram is represented in Figure 3.2 (right).

Comparing the two maps in Figure 3.1 suggests that the logarithm of population density may be a significant explanatory variable for the galician-speakers data. In this subsection, the significance of this variable will be checked taking a simple linear regression, as a preliminary step before performing more sophisticated models.

The range of the dependent variable is clearly not $\mathbb{R}$, since it is a percentage. However, it is wide enough to consider a linear model as a good simplification of its behavior. The estimated regression parameters are

$$\widehat{\boldsymbol{\gamma}} = (\widehat{\gamma_0}, \widehat{\gamma_1})^T = (117.3406, -12.0966)^T,$$

where $\widehat{\boldsymbol{\gamma}}$ is the estimation of $\boldsymbol{\gamma}$, parameter vector introduced in equation (2.1).

They are both very significant with p-values much lower than 0.001, and so is the model.

The Akaike information criterion coefficient (AIC) is going to be presented for the different models in this section because it is useful to compare the goodness of fit of the different proposals, even if the models are of a different nature, as long as they are fitted with the same dataset. The lower the AIC value, the better the corresponding model. For this model AIC is 2547.4.

However, this model may be incorrect if there is spatial dependence, as guessed before, because one of the main assumptions of linear models is uncorrelation of the residuals. An exploratory approach to check this is provided by the spatial plot of the residuals of the spatial model, that is shown in Figure 3.3. It seems clearly that relevant spatial structure exists, because there are groups of spatial regions close to each other where most of them have either positive residuals or negative ones. Moreover, note that two of the "blue clusters", the one on the north and the one on the southwest, surround three of the municipalities with largest populations densities and lower percentage of galician-speakers: Ferrol, A Coruña and Vigo. This fact suggests that they apply an important influence on their neighbors.

The Moran's I is a statistic defined in order to check the spatial correlation. It was introduced in Section 2.3, where the R function to compute it is named. Here it will be applied to the residuals of the linear model. But first, it is needed to build the spatial structure. The simpler option is that all municipalities that share some border with a municipality are taken as neighbors of it. In Figure 3.4, each individual, i.e. municipality, is represented by its centroid and neighborhoods are represented by segments joining the centroids. In Section 2.3 this structure was named border spatial structure.

The R functions and procedure used to build this neighborhood structure is detailed in Appendix B. The kind of dependence structure chosen is the simplest one, that assigns ones to connected individuals and zero to the rest.

Computing the Moran's I with this structure the value obtained is

$$I = 0.4544,$$

and the p-value of the corresponding test for spatial correlation (null hypothesis is that there is no spatial correlation between observations) is lower than 0.001, so the further study of autoregressive models is justified.

Another possible spatial structure is given by taking individuals as spatial related when the distance between them is lower than a determined value. In Section 2.3 this structure was named distance spatial structure. The distance between municipalities is considered as the distance between its centroids. A

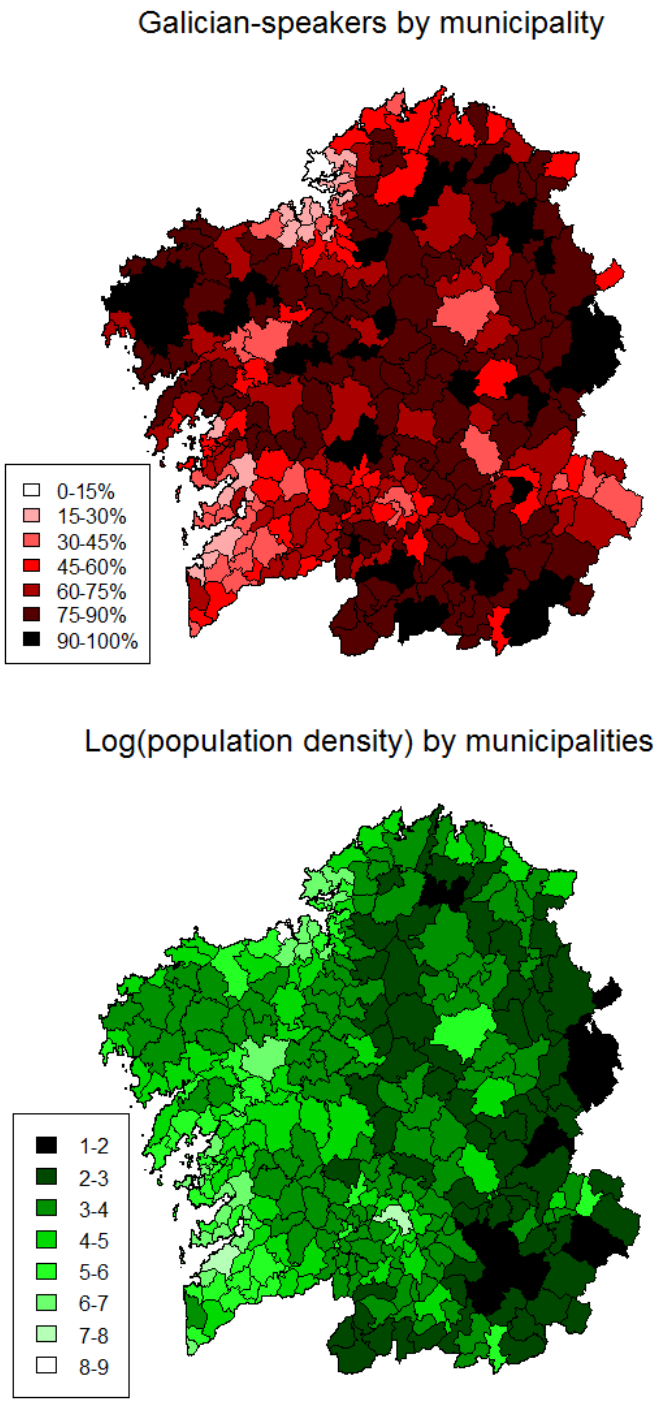


Figure 3.1: Color representation of the percentage of galician-speakers in each municipality (top) and the logarithmic transformation of the population density of each municipality (bottom).

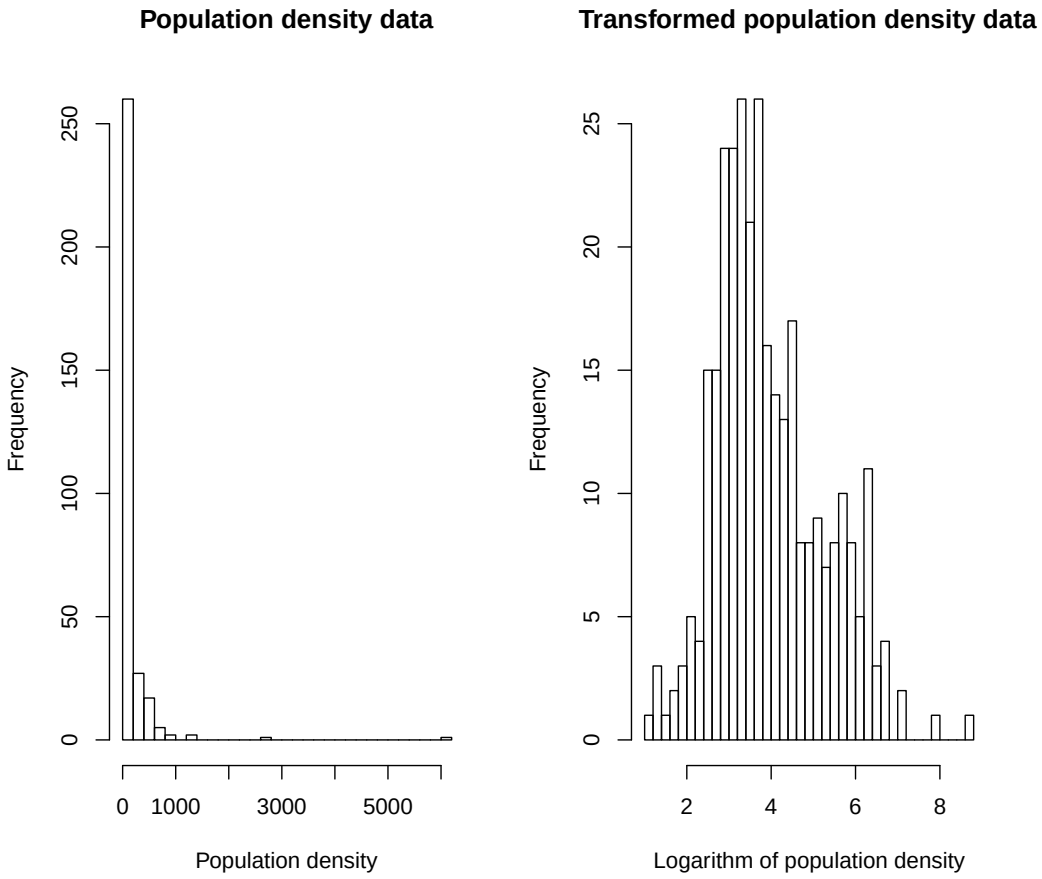


Figure 3.2: Histogram of the population data before the logarithmic transformation (left) and after the transformation (right).

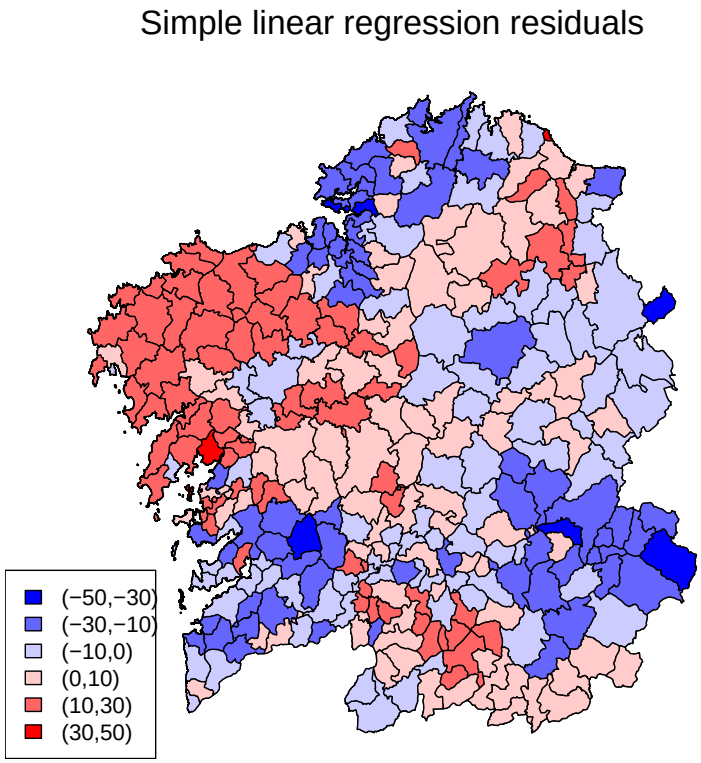


Figure 3.3: Residuals obtained from the linear regression plotted by municipalities.

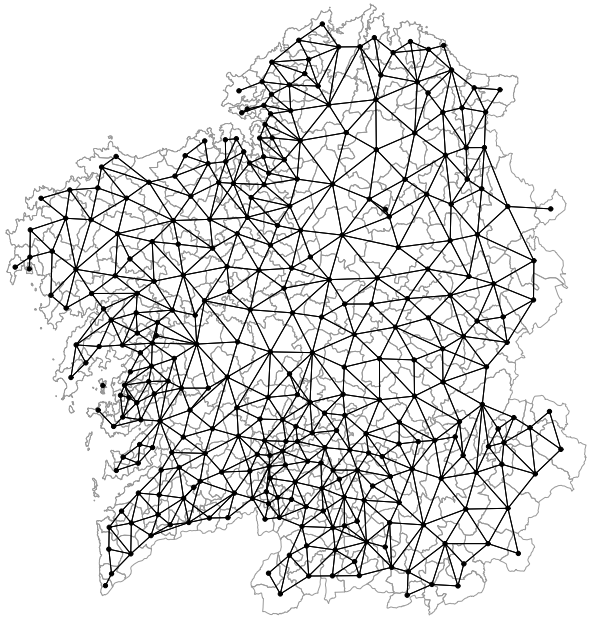


Figure 3.4: Spatial structure represented by segments joining centroids of municipalities that are neighbors. Two individuals are spatially related if they share at least a part of their borders.

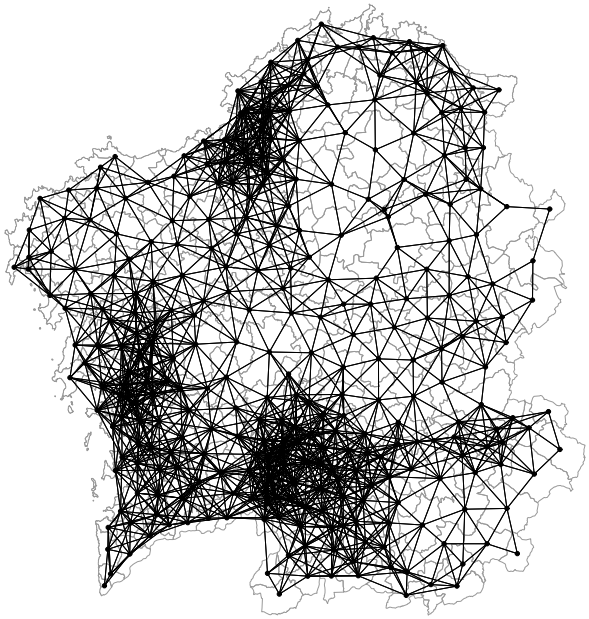


Figure 3.5: Spatial structure represented by segments joining centroids of municipalities that are neighbors. Two individuals are spatially related if the distance between their centroids is lower than 20 Km.

proper value is needed for this, that is chosen by comparing the spatial structures constructed from taking different distances. Values that generate too many or too few links are excluded. The ones that seem reasonable by this approach are compared computing their Moran's I. This leads to the choice of 20 Km. The details of this procedure in R are developed in appendix B. The resulting spatial structure is shown in Figure 3.5.

The Moran's I with this new structure takes the value

$$I = 0.43641974.$$

The associated p-value is also lower than 0.001, indicating that the residuals of the regression model also show spatial correlation taking this structure.

3.1.2 Conditional Autoregressive Model

In R package spdep there is an implemented function useful to fit CAR and SAR models, spautolm. It takes family (CAR or SAR) and weights ($\mathbf{W}$) arguments and estimates the model parameters by Maximum Likelihood (check Appendix B).

The estimated parameters of CAR (2.2) for the border spatial structure obtained are

$$\widehat{\boldsymbol{\gamma}} = (\widehat{\gamma_0}, \widehat{\gamma_1})^T = (119.2464, -12.6077)^T, \quad \widehat{\phi} = 0.1578, \quad \widehat{\sigma^2} = 115.87.$$

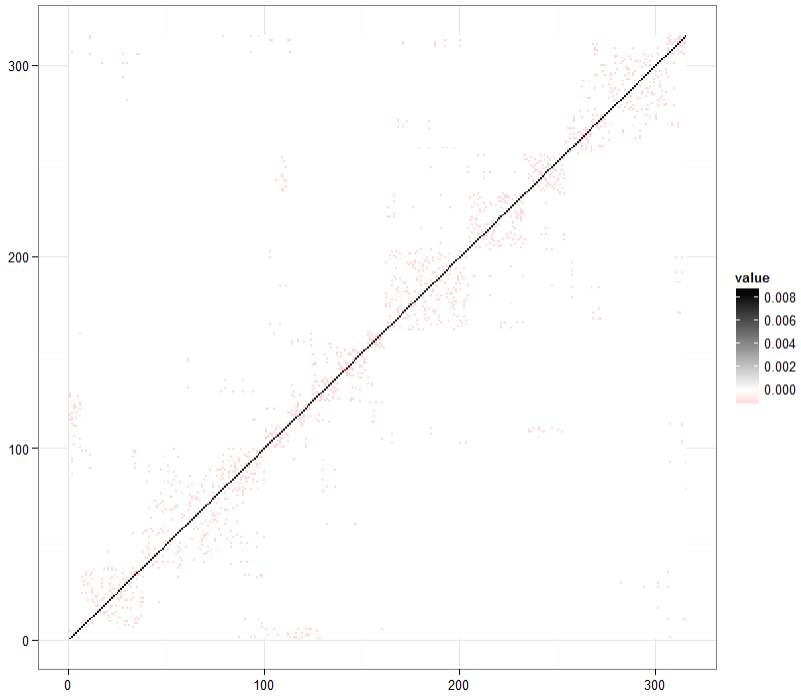


Figure 3.6: Representation of the estimated precision matrix for CAR with the border spatial structure.

The estimators $\widehat{\gamma_0}, \widehat{\gamma_1}, \widehat{\phi}$ have p-values lower than 0.001, so they are significant. From this it is easy to compute the estimated precision matrix $\widehat{\mathbf{Q}}$ for CAR. It is a high-dimensional matrix with dimension 315×315, so the color plot shown on Figure 3.6 offers an insight of it. Note that it is clearly sparse.

Compare now the goodness of fit of the linear and the conditional autoregressive model. For CAR,

$$AIC = 2436.9, \tag{3.1}$$

much lower than the one in the linear model, 2547.4, fitted in Subsection 3.1.1. This fact, as well as the result of Moran's test, confirms that CAR fits the data better than simple linear regression.

Taking the distance spatial structure and proceeding analogously, the obtained estimators are

$$\widehat{\boldsymbol{\gamma}} = (\widehat{\gamma_0}, \widehat{\gamma_1})^T = (116.0821, -11.7427)^T, \quad \widehat{\phi} = 0.0510, \quad \widehat{\sigma^2} = 137.85.$$

In this case, the estimators $\widehat{\gamma_0}, \widehat{\gamma_1}, \widehat{\phi}$ have also p-values lower than 0.001, so they are strongly significant.

The Akaike information criterion is

$$AIC = 2462.8,$$

higher than the AIC taking the border spatial structure (3.1).

3.1.3 Simultaneous Autoregressive Model

The function spautolm from R package spdep was used in the previous subsection to fit CAR model. As it was said, it includes the option of SAR too. The procedure is explained in Appendix B. The ML estimations of the parameters of SAR (2.3) obtained for the border spatial structure are

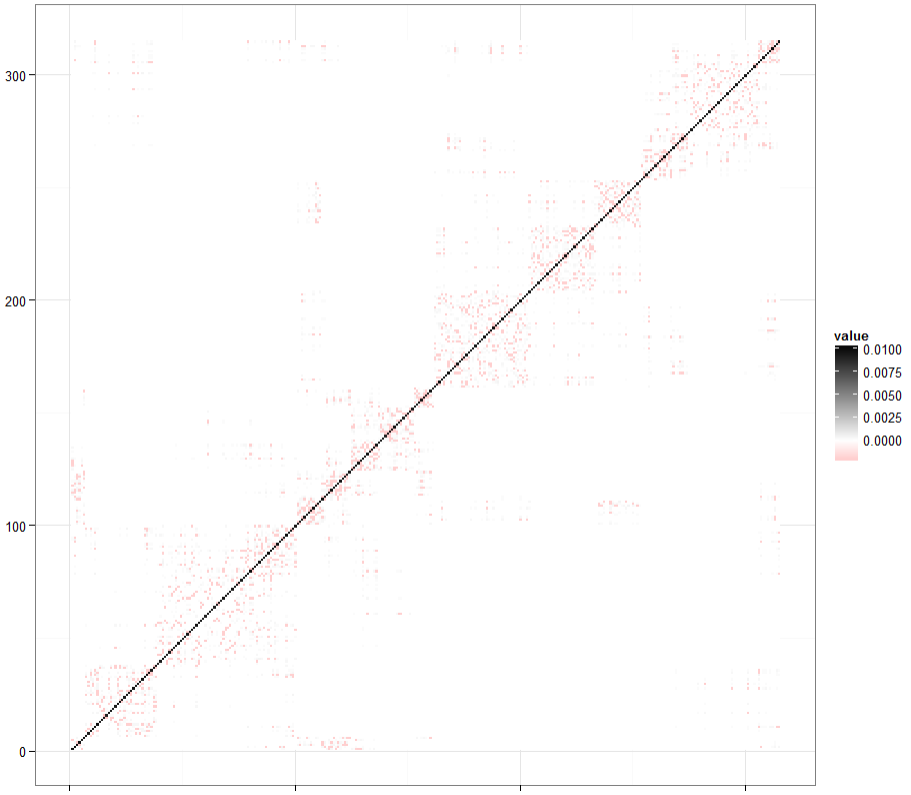


Figure 3.7: Representation of the estimated precision matrix for SAR with the border spatial structure.

$$\widehat{\boldsymbol{\gamma}} = (\widehat{\gamma_0}, \widehat{\gamma_1})^T = (119.9481, -12.5284)^T, \quad \widehat{\rho} = 0.1263, \quad \widehat{\sigma^2} = 112.74.$$

The estimators $\widehat{\gamma_0}, \widehat{\gamma_1}, \widehat{\rho}$ have p-values lower than 0.001, so they are clearly significant.

The estimated precision matrix is represented with colors in Figure 3.7. Note that it is similar to the estimated precision matrix of CAR (Figure 3.6), being both of them sparse.

The AIC of this model is

$$AIC = 2427.2, \tag{3.2}$$

a bit lower than the CAR AIC for the border spatial structure (3.1) and much lower than the coefficient of the linear model.

Taking now the distance spatial structure, the ML estimations of the parameters are

$$\widehat{\boldsymbol{\gamma}} = (\widehat{\gamma_0}, \widehat{\gamma_1})^T = (115.21030, -11.45249)^T, \quad \widehat{\rho} = 0.072459, \quad \widehat{\sigma^2} = 122.25.$$

As in the previous case, the estimators $\widehat{\gamma_0}, \widehat{\gamma_1}, \widehat{\rho}$ are strongly significant because their p-values are lower than 0.001.

The Akaike information criterion obtained is

$$AIC = 2428.8.$$

It is higher than AIC taking the border spatial structure (3.2) as it occurred in CAR model. However, in this case the difference is lower.

3.1.4 Conclusions

The autoregressive models have demonstrated to be very useful to model this data. Both of them, CAR and SAR, improve relevantly the determination coefficient compared to the linear regression model.

The best spatial structure for the percentage of galician-speakers seems to be the one that considers common borders because it provides a better AIC than the distance spatial structure. Taking a look at the residuals of the four autoregressive models in Figure 3.8, they also suggest that the border spatial structure is the best, as it shows that the residuals of CAR and SAR models have lower spatial correlation than the residuals of the analogous models changing the spatial structure.

The Moran's I is computed to check if the initial spatial correlation has been corrected with this four models. The results obtained are

$$I_a = -0.0988,\ I_b = 0.0630,\ I_c = -0.0116,\ I_d = 0.0682,$$

while the corresponding test gives the following associated p-values,

$$p_a = 0.9969,\ p_b = 0.0006,\ p_c = 0.5778,\ p_d = 0.0003,$$

where a denotes the CAR model with border spatial structure, b the CAR model with distance spatial structure, c the SAR model with border spatial structure and d the SAR model with distance spatial structure.

Observing the p-values it is clear that the adequate spatial structure to choose is the border one.

The interpretation of the estimators $\widehat{\gamma_0}, \widehat{\gamma_1}$ is quite clear. The first of them estimates the percentage of galician speakers that would be expected in a municipality whose logarithm of the population was 0. This is obviously an artifice of the model, since such individual could not exist and the percentage would be higher than the 100% (around 120% in all the cases). The second estimator is always negative (around -12), so it indicates that as the logarithm of the population density gets higher, the percentage of speakers gets lower. Equivalently, the larger the population density, the shortest the percentage of galician speakers. This confirms what was guessed at the beginning of the study.

With respect to the estimators of ϕ and ρ, the measures of spatial autocorrelation, they are harder to interpret, since there are not similar parameters with which they can be compared. The last estimator, $\widehat{\sigma^2}$, is in all cases around 120. This means that the residuals of the model exhibit a standard deviation close to 11 in a model that tries to explain percentages.

Other explanatory variables may be considered, for example, a variable that captured if a territory is mainly urban or not. However, this possible new variable would be highly correlated with population density and would make the model much more complex, because there is not a simple way to classify the regions by it.

In any case, the research until now has been successful, so another case of the same nature is going to be studied in order to get a better insight about which are the common characteristics of both situations and which others are particular of one of them.

3.2 Basque language

The data of the dependent variable, the percentage of basque-speakers, have been downloaded from the web page of Basque Statistics Office (Eustat)[2]. They were obtained in a survey performed around 1 November 2011. This survey had as a purpose to get information about different characteristics of the basque population, not only about socio-linguistic uses. The census included all the people whose residence was fixed in the Basque Country. It is reasonable to assume that spatial autoregressive models will be useful to model this data, considering the same arguments as in Section 3.1. In this

[2]Data were extracted from the web of the Eustat (http://www.eustat.eus/estadisticas/tema_460/opt_0/temas.html#axzz3vcIcdseG), specifically from the population and housing census of 2011.

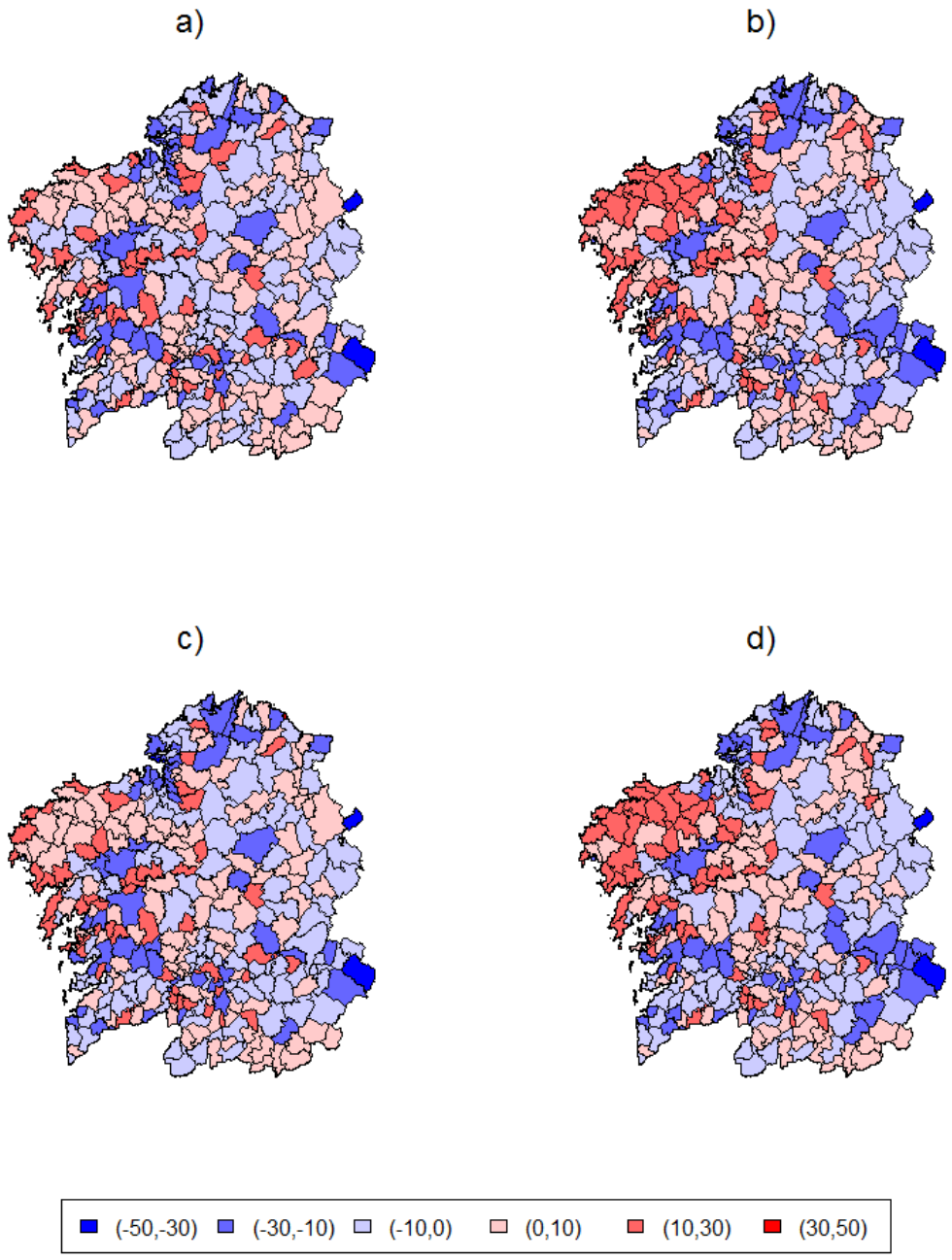


Figure 3.8: Color plot of the residuals of CAR model with border spatial structure, a), CAR model with distance spatial structure, b), SAR model with border spatial structure, c), and SAR model with distance spatial structure, d).

case, the web page presents the absolute number of basque-speakers in each municipality. Percentages of basque-speakers are needed to build the models, so they have to be computed from the absolute number and the population of each region.

The study starts with exploratory analysis, to get an overall knowledge of the data.

3.2.1 Exploratory analysis and multivariate regression model

First of all, observe and compare the maps shown in Figure 3.9. The relation between population density and percentage of basque-speakers is much less clear than in the previous section. However, two new variables arise from a quick sight of the top map: latitude and longitude.

This fact has a historical explanation. Centuries ago, the territory where basque language was spoken was wider. Spanish become more and more important over the years and it increased its territory from the south and from the west. Therefore, nowadays there are many southern and eastern municipalities in the Basque Country where the percentage of speakers is very close to zero.

Due to all this, the modeling of the data will start with a multiple linear regression, where the explanatory variables will be the logarithm of the population density, the latitude and the longitude.

The same consideration as in the previous section should be done on the dependent variable, as it is a percentage, so its range is bounded. However, this model is very useful as a first approach, as it will be shown in the following.

The estimated regression parameters, obtained with the function lm in R, are

$$\widehat{\boldsymbol{\gamma}} = (\widehat{\gamma_0}, \widehat{\gamma_1}, \widehat{\gamma_2}, \widehat{\gamma_3})^T = (61.7907, -6.1979, 13.4627, 14.7497)^T$$

where $\widehat{\boldsymbol{\gamma}}$ estimates de parameter vector $\boldsymbol{\gamma}$ (2.1). The parameters γ_1, γ_2 and γ_3 are associated with the logarithm of the population density, latitude and longitude, respectively. All the p-values are lower than 0.001, so the explanatory variables are all significant. AIC for this model is

$$AIC = 2128.3. \tag{3.3}$$

This model may be incorrect if the residuals have spatial correlation, because independence of the residuals is a basic hypothesis of the linear model. The map in Figure 3.10 shows a color plot of the residuals on the regression, and it seems that there is spatial correlation. The are at least two relevant clusters of negative residuals, in the north-west and in the south-west, and some smaller clusters of positive residuals in the north.

In order to check this guesses, the Moran's I (introduced in Section 2.3) will be computed for a given spatial structure. The same two structures explained in Section 2.3 and used for Galician language data in Subsection 3.1.1 will be considered. In that case, the border spatial structure provided more accurate results than the distance one. This fact will be checked in this new case.

The Moran's I obtained from the border spatial structure is

$$I = 0.4411,$$

and the p-value of the corresponding test is lower than 0.001. Taking the distance spatial structure, the value of the Moran's I is 0.2453.

The test provides a p-value lower than 0.001. This justifies the modeling of the data with spatial autoregressive models, as it was guessed before, taking both of the spatial structures than have just been introduced. In this section the autoregressive model applied will be SAR.

3.2.2 Simultaneous Autoregressive Model

The ML estimation of the SAR parameters (2.3) obtained from the function spautolm, taking the border spatial structure, are

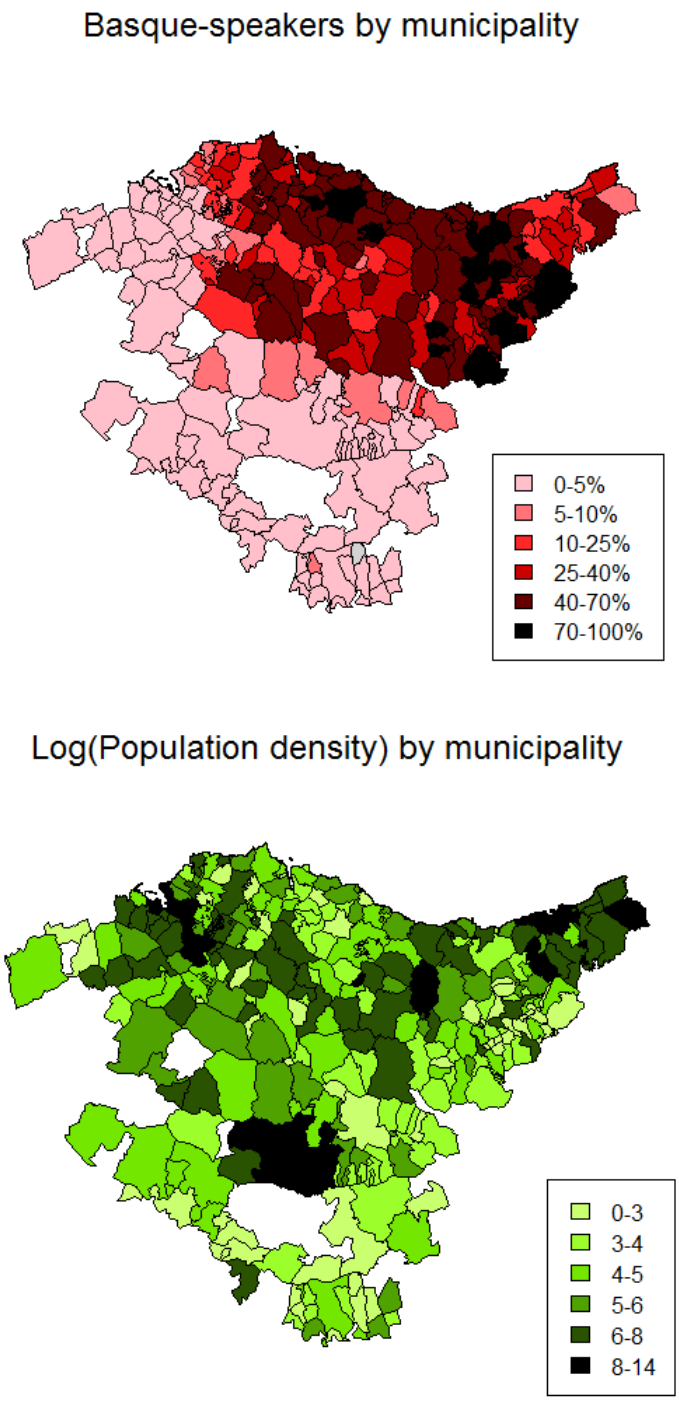


Figure 3.9: Color representation of the percentage of basque-speakers in each municipality (top) and the logarithmic transformation of the population density of each municipality (bottom). Municipalities with lack of data are colored in gray.

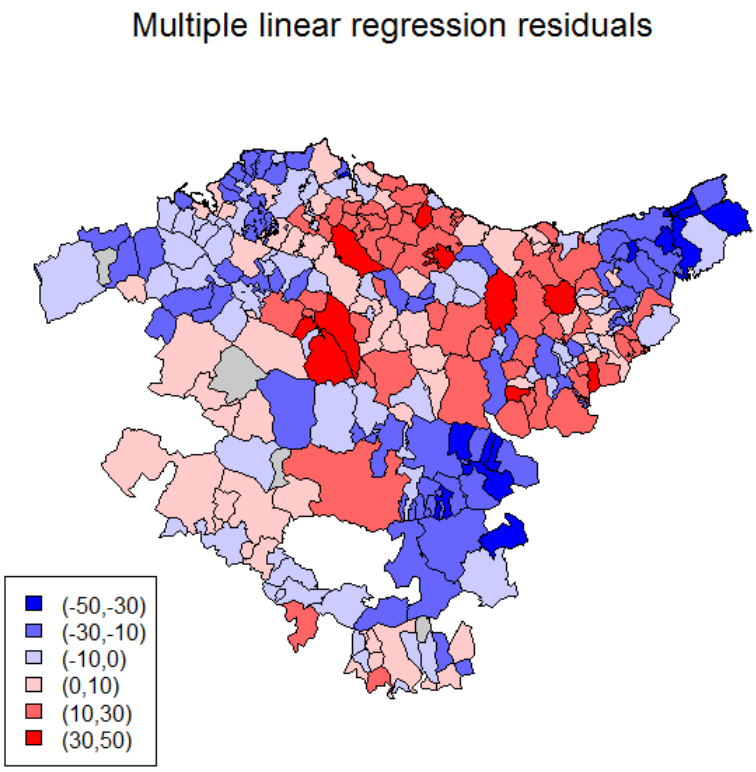


Figure 3.10: Color representation of the percentage residuals of the lienar model.

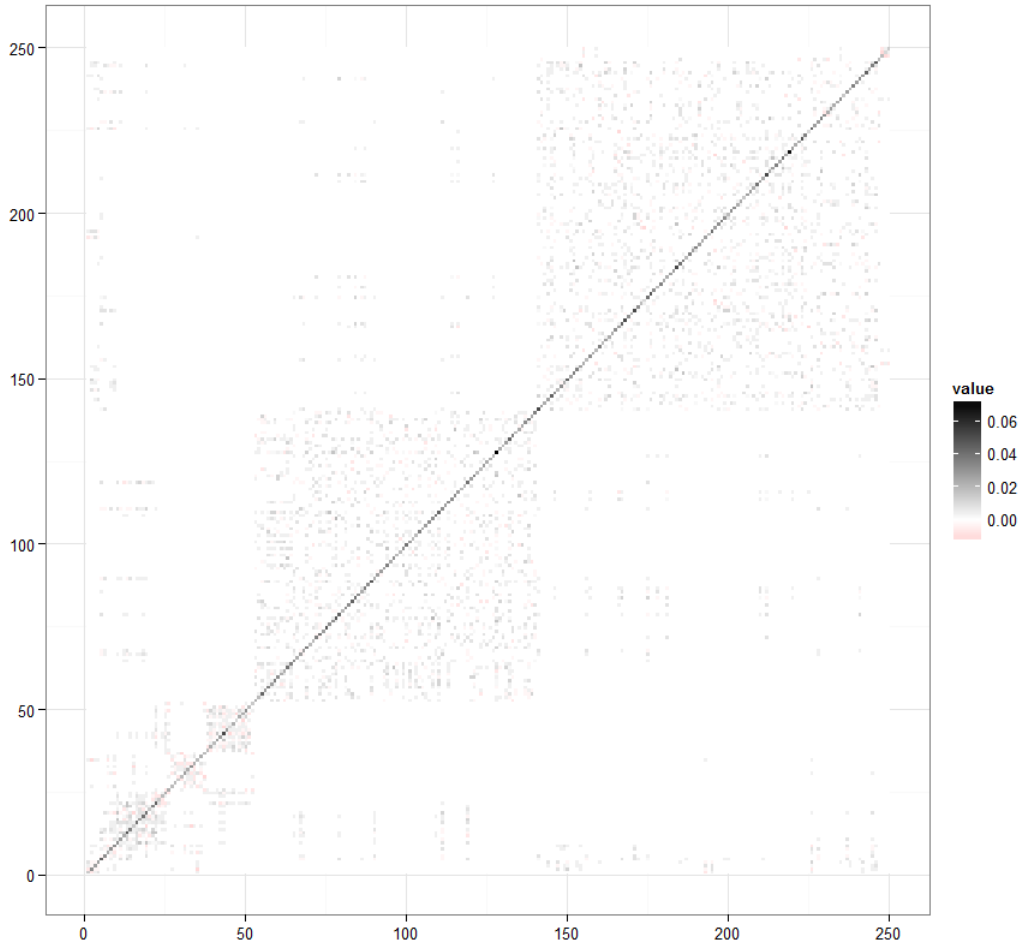


Figure 3.11: Representation of the estimated precision matrix for SAR with the border spatial structure.

$$\widehat{\boldsymbol{\gamma}} = (\widehat{\gamma_0}, \widehat{\gamma_1}, \widehat{\gamma_2}, \widehat{\gamma_3})^T = (45.8174, -2.7297, 9.5441, 15.6168)^T, \ \widehat{\rho} = 0.7705, \ \widehat{\sigma^2} = 142.66.$$

The estimators $\widehat{\gamma_0}, \widehat{\gamma_1}, \widehat{\gamma_3}$ and $\widehat{\rho}$ have associated p-values lower than 0.001, while $\widehat{\gamma_2}$, with a p-value of 0.001383. Therefore, they are all very significant.

The estimated precision matrix of the SAR model is represented in Figure 3.11. It is a sparse matrix, most of its elements are null or very close to zero.

The Akaike information criterion is computed so that this model can be compared with the linear regression.

$$AIC = 2001.6.$$

Note that this coefficient is relevantly lower than the determination coefficient of the previous linear model (3.3).

Take now the distance spatial structure. The estimations of the parameters are the following,

$$\widehat{\boldsymbol{\gamma}} = (\widehat{\gamma_0}, \widehat{\gamma_1}, \widehat{\gamma_2}, \widehat{\gamma_3})^T = (44.7261, -4.1543, 5.0089, 10.8782)^T, \ \widehat{\rho} = 0.9382, \ \widehat{\sigma^2} = 198.68.$$

The p-values associated with $\widehat{\gamma_0}, \widehat{\gamma_1}$ and $\widehat{\rho}$ are lower than 0.001, while $\widehat{\gamma_2}$ and $\widehat{\gamma_3}$ have p-values of 0.23905 and 0.02136 respectively. Therefore, taking a significance level of 5%, the parameter γ_2 is not significantly different from zero.

The SAR model without the variable "latitude" is considered. The new estimations of the parameters obtained are

$$\widehat{\boldsymbol{\gamma}} = (\widehat{\gamma_0}, \widehat{\gamma_1}, \widehat{\gamma_2})^T = (41.2969, -4.0814, 11.0870)^T,\ \widehat{\rho} = 0.93823,\ \widehat{\sigma^2} = 188.44.$$

In this case γ_2 is the parameter associated with longitude. The estimators $\widehat{\gamma_1}$ and $\widehat{\rho}$ have p-values lower than 0.001. The p-value of $\widehat{\gamma_0}$ is 0.003911 and the one of $\widehat{\gamma_2}$ is 0.028507, so the four estimators are significant with a level of 5%.

The AIC of these two models are 2046.5 for the first one and 2045.7 for the second one. For all this, taking the distance spatial structure the best SAR model is obtained eliminating "latitude" as a explanatory variable. However, the best SAR model is provided by the border spatial structure.

3.2.3 Conclusions

The second data modeled provide similar conclusions than the first one. In both galician and basque cases, spatial autoregressive models have been very useful, improving AIC compared to linear models. However, the study developed in this section has been more complicated. New explanatory variables were needed and taking one of the spatial structures, the distance one, implicated the suppression of one of the explanatory variables. Again, the border spatial structure produced the best results. This is also suggested by taking a look to Figure 3.12 as it shows that spatial correlation of residuals is higher with the distance spatial structure (right), albeit residuals are less correlated in both of them than with multiple linear regression (Figure 3.10).

Analogously as in Subsection 3.1.4, the Moran's I is computed to check if the original spatial correlation has been corrected with this two SAR models. Moran's I was introduced in Section 2.3, where its asymptotic distribution under spatial uncorrelation was shown. The results obtained are

$$I_a = -0.0486,\ I_b = 0.0146,$$

where a refers to the SAR model with the border spatial structure and b, with the distance spatial structure.

The corresponding test, obtained using the asymptotic distribution of Moran's I taking as the null hypothesis that the values of the response variable are i.i.d, gives the following associated p-values,

$$p_a = 0.8656,\ p_b = 0.1124,$$

so, again, the border spatial structure provides better results. However, in this case, the two models behave properly.

It is easy to interpret the obtained estimators $\widehat{\gamma_0}, \widehat{\gamma_1}, \widehat{\gamma_2}, \widehat{\gamma_3}$ in the different models that have been fitted. The first of them estimates the percentage of basque speakers that would be expected in a municipality whose logarithm of the population was 0 and whose latitude and longitude were also 0. This is clearly an artifice of the model, since such individual does not exist. In fact, it is not even a territory that could be placed in the Basque Country!. The second estimator is always negative (between -2.7 and -4.2 in SAR models, while its value is more negative in the linear regression model), so it indicates that as the logarithm of the population density gets higher, the percentage of speakers gets lower. Equivalently, the larger the population density, the shortest the percentage of basque speakers. Observing now the third and fourth estimators, that correspond to latitude and longitude, it appears that they are always positive, except for the last model fitted, where latitude was eliminated. This confirms the previous assumptions, that is to say, the percentage of basque speakers is lower when moving west or south, although the last one is not significant when considering SAR with the distance spatial structure.

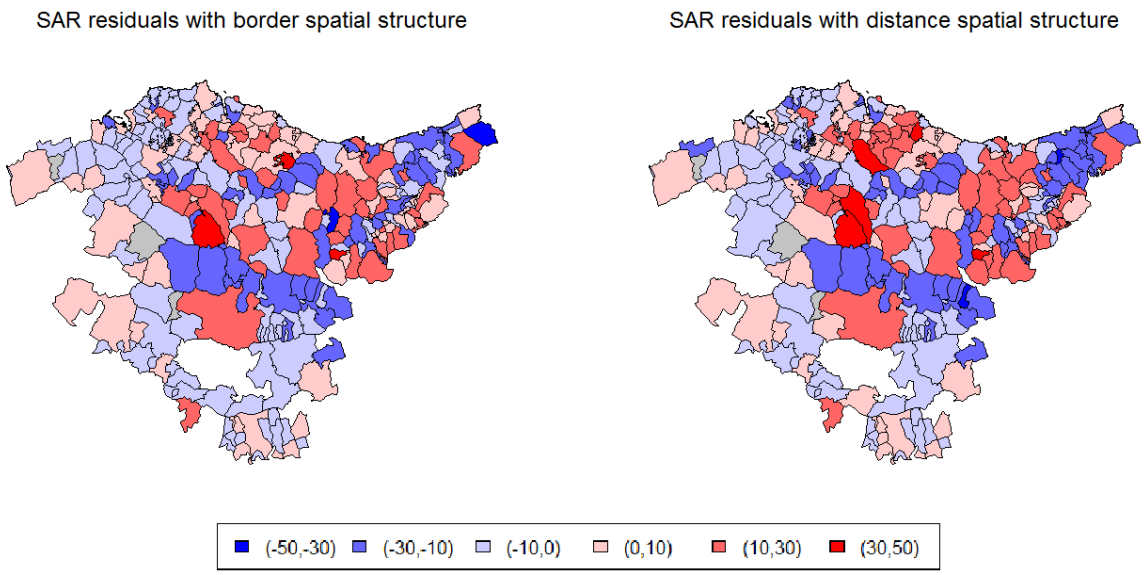


Figure 3.12: Color plot of the residuals of SAR model with border spatial structure (left) and with distance spatial structure (right).

With respect to the estimators of ρ, the measure of spatial autocorrelation, they are more difficult to interpret. Comparing them to the values of $\widehat{\rho}$ in Subsection 3.1.3, it seems that the influences between close municipalities are similar in Galicia and the Basque Country. The last estimator, $\widehat{\sigma^2}$, is between 140 and 200 in all the SAR models fitted. This means that the residuals of the model exhibit higher standard deviation than the models fitted in Section 3.1.

The possibility of adding a new variable that captures if a municipality is urban or not deserves the same consideration as in Subsection 3.1.4. Moreover, in this case the inclusion of a new variable would be even more problematic because the model itself is already more complex.

Appendix A

An extension to Spatiotemporal Autoregressive Models

Time is a qualitatively different variable from space, so when data are collected over space and also over time new models have to be developed. In this appendix, one of this models is going to be introduced: spatiotemporal autoregressive model (STAR).

An example where this model would be useful is the one that was studied in this text. It seemed reasonable that close communities would influence each other linguistic habits. But a more complex approach appears naturally when thinking about this process over years. As the changes in the use of a language are usually slow over time, the influence of a community into others is produced over years, or even generations. Unfortunately, there is a lack os data in the percentage of galician and basque-speakers over time, so STAR will only be considered theoretically. A wider approach to STAR and how it is related to spatial autoregressive models can be found in [3] (pp 251-253).

Take a time discrete variable t moving between 1 and T. Let $\mathbf{Y}_t$ be a random n-dimensional vector that takes different values at different times t, $t = 1, \dots, T$. Consider first a model that uses only past data and the assumption that explanatory variables do not change over time. Therefore, the matrix $\boldsymbol{\mathcal{X}}$ taken is analogue to the one used in CAR and SAR models. Then,

$$\mathbf{Y}_t = \tau \mathbf{Y}_{t-1} + \rho \mathbf{W} \mathbf{Y}_{t-1} + \boldsymbol{\mathcal{X}} \boldsymbol{\gamma} + \boldsymbol{\epsilon}_t, \tag{A.1}$$

where $\mathbf{W}$ is a spatial weight matrix.

This is equivalent to

$$\mathbf{Y}_t = \mathbf{G} \mathbf{Y}_{t-1} + \boldsymbol{\mathcal{X}} \gamma + \boldsymbol{\epsilon}_t, \tag{A.2}$$

with $\mathbf{G} = \tau \mathbf{I}_n + \rho \mathbf{W}$. The scalar parameter τ governs dependence between each point at time t and $t-1$. The larger τ, the stronger the dependence. The scalar parameter ρ provides the strength of the spatial dependence between each point at time t and its neighboring points at time $(t-1)$. As the spatial dependence is already given, errors are independent, so

$$\boldsymbol{\epsilon}_t \in \mathcal{N}_n(\mathbf{0}_n, \sigma^2 \mathbf{I}_n),$$

where $\boldsymbol{\epsilon}_t$ denotes the errors of the model at time t.

Make another assumption related to $\mathbf{G}$. Suppose that $\lim_{t \to \infty} \mathbf{G}^t = \mathbf{0}_n$. This restricts also the scalar parameters. Assuming that $\tau, \rho \geq 0$ and that $\mathbf{W}$ has a principal eigenvalue of 1, then $\tau + \rho < 1$.

As the explanatory variable at time t depends on its values at time $(t-1)$, it is obvious that this last one depends on its values at time $t-2$, so recursively from the equation (A.2),

$$\mathbf{Y}_t = \mathbf{G}^t \mathbf{Y}_0 + (\mathbf{I}_n + \mathbf{G} + \mathbf{G}^2 + \cdots + \mathbf{G}^{t-1}) \boldsymbol{\mathcal{X}} \gamma + \mathbf{u},$$

being

$$\mathbf{u} = \mathbf{G}^{t-1}\boldsymbol{\epsilon}_1 + \cdots + \mathbf{G}\boldsymbol{\epsilon}_{t-1} + \boldsymbol{\epsilon}_t.$$

A more complete form of STAR is introduced now. It considers not only past data but also present data. Therefore the form is analogue to the last one but adding the term $\lambda\mathbf{W}\mathbf{Y}_t$. The obtained equation is

$$\mathbf{Y}_t = \lambda\mathbf{W}\mathbf{Y}_t + \tau\mathbf{Y}_{t-1} + \rho\mathbf{W}\mathbf{Y}_{t-1} + \boldsymbol{\mathcal{X}}\gamma + \boldsymbol{\epsilon}_t, \tag{A.3}$$

equivalent to

$$\mathbf{Y}_t = \lambda\mathbf{W}\mathbf{Y}_t + \mathbf{G}\mathbf{Y}_{t-1} + \boldsymbol{\mathcal{X}}\gamma + \boldsymbol{\epsilon}_t.$$

Given the matrix $\mathbf{A} = (\mathbf{I}_n - \lambda\mathbf{W})^{-1}$, the equation can be rewritten as

$$\mathbf{Y}_t = \mathbf{A}\mathbf{G}\mathbf{Y}_{t-1} + \mathbf{A}\boldsymbol{\mathcal{X}}\gamma + \mathbf{A}\boldsymbol{\epsilon}_t.$$

Analogously as is was done before, the dependent variable at time t can depend only on its first values,

$$\mathbf{Y}_t = (\mathbf{A}\mathbf{G})^t\mathbf{Y}_0 + (\mathbf{I}_n + \mathbf{A}\mathbf{G} + (\mathbf{A}\mathbf{G})^2 + \cdots + (\mathbf{A}\mathbf{G})^{t-1})\mathbf{A}\boldsymbol{\mathcal{X}}\gamma + \mathbf{A}\boldsymbol{\epsilon}_t,$$

where $\mathbf{u} = (\mathbf{A}\mathbf{G})^{t-1}\boldsymbol{\epsilon}_1 + \cdots + \mathbf{A}\mathbf{G}\boldsymbol{\epsilon}_{t-1} + \boldsymbol{\epsilon}_t$.

Consider now that the explanatory variables evolve over time. In this case, the model has to include the realization of the explanatory variables at time t and at time $(t-1)$, $\boldsymbol{\mathcal{X}}_t, \boldsymbol{\mathcal{X}}_{t-1}$, as well as the influence of the realizations of the explanatory variables near each point at time t and $(t-1)$, $\mathbf{W}\boldsymbol{\mathcal{X}}_t, \mathbf{W}\boldsymbol{\mathcal{X}}_{t-1}$. This leads us to the following equation.

$$\mathbf{Y}_t = \lambda\mathbf{W}\mathbf{Y}_t + \tau\mathbf{Y}_{t-1} + \rho\mathbf{W}\mathbf{Y}_{t-1} + \boldsymbol{\mathcal{X}}_t\gamma_1 + \mathbf{W}\boldsymbol{\mathcal{X}}_t\gamma_2 + \boldsymbol{\mathcal{X}}_{t-1}\gamma_3 + \mathbf{W}\boldsymbol{\mathcal{X}}_{t-1}\gamma_4 + \mathbf{I}_n k + \boldsymbol{\epsilon}. \tag{A.4}$$

A brief illustration about how socio-linguistic date would be approached with spatiotemporal autoregressive models is going to be introduced. Imagine that the percentage of speakers of a given language in different areas of a territory was available over time, for example, every 5 years, and take it as the response variable, $\mathbf{Y}_t$, where t is an index that indicates the year when the data were taken. Clearly, some useful explanatory variables can be considered, as they were in Chapter 3. Realizations of these variables shape the design matrix $\boldsymbol{\mathcal{X}}$. However, it seems reasonable that the influences between spatial-related areas make percentages evolve over time. Think about a group of several people that arrive to a given town from a place with very different socio-linguistic uses. At first, this town would have a different value on the response variable than the towns around it, maybe an outlier, but as years go on, this situation would be reverted because of the influences of the close towns. This is one of the events that autoregressive models try to capture.

Consider the simplest case of STAR introduced (A.1). In this one, the realizations of the explanatory variables are considered to be fixed over time, so in the case that there is available data for different years, a reasonable solution would be to use the average. A weight matrix would be needed too. The simple ones used in Chapter 3 are appropriate here. Once the model is fitted, the interpretation of the estimators of the coefficients $\boldsymbol{\gamma}$ would be analogue as the one performed to CAR or SAR, while the estimation of τ would provide information about the strength of the influence of the percentage of speakers in an area at time t and at time $t+1$. The estimation of ρ would give insight into the implications of the socio-linguistic uses on the neighbors of an area at time $t-1$ and the uses on the given area at time t.

Another approach based on equation A.3 could be performed. It is similar to the previous one, but another parameter has to be estimated, λ. The estimation gives information analogue to ϕ in CAR (Subsection 2.1.1) or ρ in SAR (Subsection 2.2.1).

Finally, if data of the explanatory variables is available over time, the most complete model would be the one described in equation A.4. This one adds parameters with respect to A.3. The interpretation of the estimations of $\boldsymbol{\gamma}_1$, $\boldsymbol{\gamma}_2$, $\boldsymbol{\gamma}_3$ and $\boldsymbol{\gamma}_4$ is quite intuitive: the influence on the response variable at time t of each of the explanatory variables in the same area at time t, in neighbor areas at time t, in the same area at time $t-1$ and in neighbor areas at time $t-1$, respectively.

Appendix B

CAR and SAR fitting in R

The R package spdep was used to fit CAR and SAR models along this thesis. The elements needed and steps to fit CAR and SAR are summarized in this appendix. They are widely explained in [2] (pp 237-287), as well as many other approaches related to spatial data developed with R.

Elements needed:

- A object of class sp (SpatialPoints or SpatialPolygons) that contains the spatial structure of the data.
- Vectors that contain the realization of each explanatory variable and of the response variable.
- Functions dnearneigh, poly2nb, nb2listw and spautolm of R package spdep.

Steps:

1. Create an object of the class nb that contains the neighborhood structure of the data. This can be done with a wide variety of functions. The ones used in this text are the following:

   - Function dnearneigh, when the object of class sp contains the point coordinates and the spatial structure used is the distance one, as it was in Chapter 2 and in Section 3.1.1 and 3.2.1. The usage of this function is:

     ```
     dnearneigh(x, d1, d2 ...)
     ```

     where x is a matrix of point coordinates or a SpatialPoints object, d1 the lower distance bound and d2 the upper distance bound.
   - Function poly2nb, when the object of class sp contains a list of polygons and the spatial structure used is the border one, as it was in Sections 3.1.1 and 3.2.1. The usage of this function is

     ```
     poly2nb(pl...)
     ```

     where pl is a list of polygons of class extending SpatialPolygons.

2. Supplement the neighbors list of the object of class nb with spatial weights for the chosen coding scheme. This is done with the function nb2listw. Its usage is the following.

   ```
   nb2listw(neighbors, style="W", zero.policy=NULL ...)
   ```

where neighbors is the object of class nb that has just been created, style allows to choose the type of spatial weight matrix and the value of zero policy indicates if the computation must stop for any empty neighbor sets or not.

All over this text, in Chapter 2 and in Section 3.1.1 and 3.2.1, the type of spatial weight matrix is "B", the simplest binary type. In Section 3.1.1, the zero.policy was allocated with TRUE, because there is a municipality with no neighbors when using the border spatial structure (an island, Illa de Arousa).

This function returns an object of class listw.

3. Fit the CAR or SAR model with the function spautolm. The usage of this function is

```
spautolm(formula, data=list(), listw, family = "SAR", zero.policy = NULL ...)
```

A symbolic description of the model to be fit is included in formula, while a data frame containing the realization of the explanatory and response variables of the model is included in data. The object of class listw that has been created in the last step has to be specified in listw. The parameter family admits "SAR" or "CAR", depending of the model that is being fitted, and zero.policy indicates if the computation must stop for any empty neighbor sets or not.

In Section 3.1.1 the zero.policy had to be allocated with TRUE, as it was explained in the second step.

Appendix C

Notation

X: random variable.

x: realization of X.

$\mathbf{X} = (X_1, \cdots, X_n)^T$: random vector whose realizations are n-dimensional.

$\mathbf{x} = (x_1, \cdots, x_n)^T$: sample realization of $\mathbf{X}$.

$\mathbf{X}_{-i}$: random vector $\mathbf{X}$ except the component i, $i = 1, \ldots, n$.

$p_X(x)$: probability that the random variable X takes the value x.

$p_{\mathbf{X}}(\mathbf{x})$: probability that the random vector $\mathbf{X}$ takes the value $\mathbf{x}$.

$p_{X|z}(x|z)$: probability that the random variable X takes the value x given that the random variable Z takes the value z.

$p_{\mathbf{X}|\mathbf{z}}(\mathbf{x}|\mathbf{z})$: probability that the random vector $\mathbf{X}$ takes the value $\mathbf{x}$ given that the random vector $\mathbf{Z}$ takes the value $\mathbf{z}$.

$\mathbb{E}(X) = \mu$: expected value of X.

$\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}$: n-dimensional vector of expected values of $\mathbf{X}$.

$\mathbb{E}(X_i|\mathbf{X}_{-i} = \mathbf{x}_{-i})$: expected value of X_i conditioned to that $\mathbf{X}_{-i} = \mathbf{x}_{-i}$, $i = 1, \ldots, n$.

$\boldsymbol{\Sigma}$: covariance matrix of $\mathbf{X}$ where each element Σ_{ij} consists on covariance between X_i and X_j, $i, j = 1, \ldots, n$.

$\text{Prec}(X)$: reciprocal of the variance of X.

$\text{Prec}(\mathbf{X})$: vector whose elements are the reciprocals of the variances of $\mathbf{X}$.

$\text{Prec}(X_i|\mathbf{X}_{-i} = \mathbf{x}_{-i})$: reciprocal of the variance of X_i conditioned to that $\mathbf{X}_{-i} = \mathbf{x}_{-i}$, $i = 1, \ldots, n$.

$\text{Corr}(X_i, X_j)$: correlation between the random variables X_i, X_j, $i, j = 1, \ldots, n$.

$\text{Corr}(X_i, X_j|\mathbf{X}_{-ij} = \mathbf{x}_{-ij})$: correlation between X_i and X_j conditioned to that $\mathbf{X}_{-ij} = \mathbf{x}_{-ij}$, $i, j = 1, \ldots, n$.

$\boldsymbol{I}_n$: Identity matrix of dimension $n \times n$.

$\mathbf{0}_n$: Null squared matrix of dimension $n \times n$.

$\boldsymbol{A} > 0$, being $\boldsymbol{A}$ a matrix: matrix $\boldsymbol{A}$ is semipositive definite.

$\boldsymbol{\mathcal{X}}$: matrix consisting on the observations of k explanatory random variables $X_m, m = 1, \ldots, k$, adding a first column of ones (only if an intercept is needed into the model).

wrt: with respect to

iff: if and only if

SPD: semipositive definite

Bibliography

[1] Banerjee S, Carlin BP, Gelfand AE (2004) *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC, New York.

[2] Bivand RS, Pebesma EJ, Gómez-Rubio V (2008) *Applied Spatial Data Analysis with R. Springer Science+ Business Media*, New York.

[3] Gelfand AE, Diggle PJ, Fuentes M, Guttorp P (2010) *Handbook of Spatial Statistics*. Chapman & Hall/CRC, New York.

[4] Mardia KV (1990) *Maximum Likelihood Estimation for Spatial Models*. Department of Statistics, University of Leeds, Leeds.

[5] R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

[6] Rue H, Held L (2005) *Gaussian Markov Random Fields. Theory and Applications*. Chapman & Hall/CRC, New York.