UNIVERSITA' DEGLI STUDI DI MILANO BICOCCA

BSM6 - Bayesian spatial modelling

Lecturer: Riccardo Corradin

Introduction

Several modern approaches in various fields, such as climatology, ecology, environmental health, real estate marketing, etc., face the task of analyzing structured data that are

- highly multivariate, with covariates and response variables;
- geographically referenced;
- temporally correlated.

Among these, spatial data analysis deals with observations that depend on a specific continuous domain, which describe their dispersion over a coordinate set.

Those coordinates are providing informations and insights that can be helpful to better understand and interpret what we are studying.

Introduction

Spatial data are usually classified into 3 types, depending on the specific structure of the data.

- **Point-referenced data** (geostatistical data): $Y(\boldsymbol{s})$ random vector at location $\boldsymbol{s} \in \mathbb{R}^r$. The coordinate $\boldsymbol{s}$ varies continuously over D, a subset of $\mathbb{R}^r$ that contains a r-dim rectangle of positive volume.
- **Areal data**: $Y(\boldsymbol{s})$, $\boldsymbol{s} \in D$, and D is partitioned into a finite number of areal units with well-defined boundaries.
- **Point pattern data**: D is random, and the index set of D gives the locations of random events that are the spatial point pattern. For example, $Y(\boldsymbol{s}) = 1$ for all $\boldsymbol{s} \in D$.

As statistician, we want to investigate if the spatial domain has an impact on the data structure. Specifically, we want to study if there is any **spatial pattern** in data $Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2), \ldots, Y(\boldsymbol{s}_n)$.

→ **Spatial pattern** suggests that measurements near to each other will tend to take more similar values than those for units far from each other.

→ **Independent measurements** for the units no pattern.

Point-referenced data

Point-referenced data

In this framework, data $Y(\boldsymbol{s})$ are given at specific locations $\boldsymbol{s} \in D \subseteq \mathbb{R}^r$. For example, $Y(\boldsymbol{s})$ are level of a pollutant at site $\boldsymbol{s}$.

While we can assume the existence of a pollutant level at all possible sites, in practice the data are a partial realization of a **spatial process** at specific locations $\{\boldsymbol{s}_1, \dots, \boldsymbol{s}_n\}$.

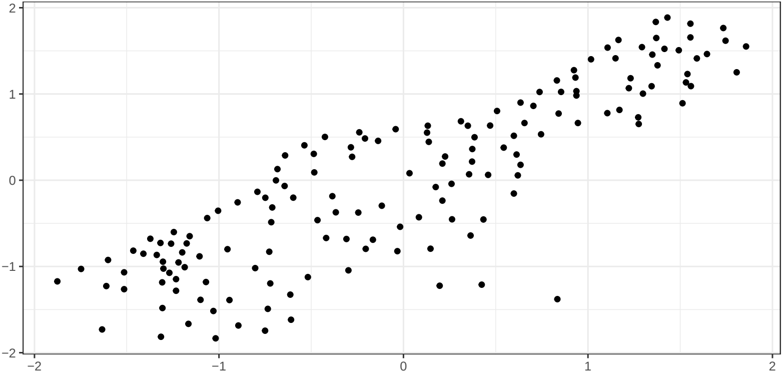
A fundamental object to deal with this type of data is the **underlying stochastic process** $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D\}$.

We observe the process at fixed locations. Hence, the data we observe are

$$y_{1:n} = (y(\boldsymbol{s}_1), \dots, y(\boldsymbol{s}_n))^\intercal.$$

- The process is centered in $\mu(\boldsymbol{s}) = \mathrm{E}[Y(\boldsymbol{s})]$, which is the mean parameter.
- We also assume that the variance of $Y(\boldsymbol{s})$ exists at each $\boldsymbol{s} \in D$.
- Usually, we also assume that $(Y(\boldsymbol{s}_1), \dots, Y(\boldsymbol{s}_n))^\intercal$ is distributed according to a multivariate Gaussian distribution.

Point-referenced data



Locations (we also have topsoil heavy metal concentrations along with a number of soil and landscape variables at the observation locations) collected in a flood plain of the river Meuse, near the village of Stein (NL).

Point-referenced data

A fundamental concept is the **stationarity** of the underlying process. Under stationarity, the characteristic of such a processes such as mean, variance and covariance do not change upon shifting the support.

- The process is said to be **strictly (strong) stationary** if, for any $n \geq 1$, any set of sites $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$, and $\boldsymbol{h} \in \mathbb{R}^r$, we have

$$(Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n)) \overset{d}{=} (Y(\boldsymbol{s}_1 + \boldsymbol{h}), \ldots, Y(\boldsymbol{s}_n + \boldsymbol{h})), \qquad \text{with } \boldsymbol{s}_j + \boldsymbol{h} \in D.$$

- The process is said to be **weakly stationary** if
  - → $\mu(\boldsymbol{s}) = \mu$, i.e. constant mean over the spatial domain.
  - → For the covariance term, we have

$$\text{cov}(Y(\boldsymbol{s}), Y(\boldsymbol{s} + \boldsymbol{h})) = C(\boldsymbol{h}), \qquad \text{with } \boldsymbol{s} + \boldsymbol{h} \in D.$$

  In practice, the covariance can be summarized in a covariance function.

- The process is said to be **intrinsic stationary** if
  - → $\mathrm{E}[Y(\boldsymbol{s} + \boldsymbol{h}) - Y(\boldsymbol{s})] = 0$.
  - → Also,

$$\mathrm{E}[(Y(\boldsymbol{s} + \boldsymbol{h}) - Y(\boldsymbol{s}))^2] = \text{var}(Y(\boldsymbol{s} + \boldsymbol{h}) - Y(\boldsymbol{s})) = 2\gamma(\boldsymbol{h}),$$

  depending solely on $\boldsymbol{h}$, where $2\gamma(\boldsymbol{h})$ is called variogram and $\gamma(\boldsymbol{h})$ semivariogram.

Note that: strong stationarity ⇒ weak stationarity ⇒ intrinsic stationarity.

Point-referenced data

We recall the following relation between the semivariogram and the covariance function.

$$\gamma(\boldsymbol{h}) = C(0) - C(\boldsymbol{h}) \quad \Leftrightarrow \quad C(\boldsymbol{h}) = C(0) - \gamma(\boldsymbol{h}) = \lim_{||\boldsymbol{u}|| \to \infty} \gamma(\boldsymbol{u}) - \gamma(\boldsymbol{h}).$$

Finally, the model is said to be isotropic if $\gamma(\boldsymbol{h}) = \gamma(||\boldsymbol{h}||)$. A model that is both isotropic and stationary is called homogeneous.

Example: exponential semivariogram/covariance.

$$\gamma(d) = \gamma(||\boldsymbol{h}||) = \begin{cases} \tau^2 + \sigma^2(1 - \mathrm{e}^{-\phi d}) & \text{if } d > 0 \\ 0 & \text{if } d = 0 \end{cases}$$

$$C(d) = C(||\boldsymbol{h}||) = \begin{cases} \sigma^2 \mathrm{e}^{-\phi d} & \text{if } d > 0 \\ \tau^2 + \sigma^2 & \text{if } d = 0 \end{cases}$$

Nuggets: $\tau^2 = \lim_{d \to 0^+} \gamma(d)$, represent the non-spatial variability

Range: $R = 1/\phi$, where ϕ is the decay parameter

Sill: $\tau^2 + \sigma^2 = \lim_{d \to +\infty} \gamma(d)$

Note that, for $d > 0$, we use the notation $C(d) = \sigma^2 \rho(d, \phi)$.

Point-referenced data

Example: powered exponential semivariogram/covariance.

$$\gamma(d) = \gamma(||\boldsymbol{h}||) = \begin{cases} \tau^2 + \sigma^2(1 - \mathrm{e}^{-|\phi d|^p}) & \text{if } d > 0 \\ 0 & \text{if } d = 0 \end{cases}$$

$$C(d) = C(||\boldsymbol{h}||) = \begin{cases} \sigma^2 \mathrm{e}^{-|\phi d|^p} & \text{if } d > 0 \\ \tau^2 + \sigma^2 & \text{if } d = 0 \end{cases}$$

with $0 < p \leq 2$ and $\rho(d, \phi) = e^{-|\phi d|^p}$.

Example: Gaussian semivariogram/covariance.

$$C(d) = C(||\boldsymbol{h}||) = \begin{cases} \sigma^2 e^{-\phi^2 d^2} & \text{if } d > 0 \\ \tau^2 + \sigma^2 & \text{if } d = 0 \end{cases}$$

and $\rho(d, \phi) = e^{-\phi^2 d^2}$.

Point-referenced data

We introduce a first **Bayesian spatial regression model** for point-referenced data. We consider a model of the form

$$Y(\boldsymbol{s}) = \boldsymbol{x}^\intercal(\boldsymbol{s})\boldsymbol{\beta} + \omega(\boldsymbol{s}) + \epsilon(\boldsymbol{s}),$$

where the residual term of the model is partitioned in two parts:

- $\omega(\boldsymbol{s})$ is the spatial residual term, where $\{\omega(\boldsymbol{s})\}$ is a spatial Gaussian process, capturing the residual spatial association. Its distribution is indexed by the dispersion parameters σ^2 and ϕ.
- $\{\epsilon(\boldsymbol{s})\}$ is a sequence of uncorrelated pure error terms, variability at distances smaller than the smallest interlocation distance, with a distribution indexed by τ^2.

Let X be a $n \times p$ matrix with $\boldsymbol{x}^\intercal(\boldsymbol{s}_i)$ being its ith row, and $\boldsymbol{\omega} = (\omega(\boldsymbol{s}_1), \dots, \omega(\boldsymbol{s}_n))^\intercal$. The model specification is completed by setting

$$\begin{aligned}
\boldsymbol{Y} \mid \boldsymbol{\omega}, \boldsymbol{\beta}, \tau^2 &\sim N(\mathrm{X}\boldsymbol{\beta} + \boldsymbol{\omega}, \tau^2 \mathrm{I}_n), \\
\boldsymbol{\omega} \mid \boldsymbol{\theta} &\sim N(\boldsymbol{0}, \Sigma(\boldsymbol{\theta})), \qquad \text{where } [\Sigma(\boldsymbol{\theta})]_{ij} = \sigma^2 \rho(||\boldsymbol{s}_i - \boldsymbol{s}_j||, \boldsymbol{\theta}), \\
\boldsymbol{\beta} &\sim N(\boldsymbol{b}_0, \Lambda_0), \\
\boldsymbol{\theta} = (\sigma^2, \phi, \tau^2) &\sim \pi(\sigma^2, \phi, \tau^2).
\end{aligned}$$

where $\sim \pi(\sigma^2, \phi, \tau^2)$ usually is assumed to factorize in independent components.

Point-referenced data

Alternatively, we can integrate out the random effect $\boldsymbol{\omega}$, obtaining as marginal model

$$\begin{aligned} \boldsymbol{Y} \mid \boldsymbol{\beta}, \tau^2 &\sim N(\mathrm{X}\boldsymbol{\beta}, \sigma^2 H(\phi) + \tau^2 \mathrm{I}_n), \qquad \text{where } [H(\phi)]_{ij} = \rho(||\boldsymbol{s}_i - \boldsymbol{s}_j||, \phi), \\ \boldsymbol{\beta} &\sim N(\boldsymbol{b}_0, \Lambda_0), \\ \boldsymbol{\theta} = (\sigma^2, \phi, \tau^2) &\sim \pi(\sigma^2, \phi, \tau^2). \end{aligned}$$

Finally, considering the support of the parameters in $\boldsymbol{\theta}$, we can consider as priors

$$\begin{aligned} \sigma^2 &\sim IG(a_\sigma, b_\sigma), \\ \tau^2 &\sim IG(a_\tau, b_\tau), \\ \phi &\sim IG(a_\phi, b_\phi). \end{aligned}$$

The model can be implemented in STAN, by suitably constructing the correlation/covariance matrices needed in the spatial residual term (first specification) or in the marginal distribution of the data (second specification).

Point-referenced data

As usual in spatial analysis, one of our main scopes is to perform **kriging**. In a Bayesian framework, kriging is nothing but **Bayesian prediction**.

We want to predict the **response** Y_0 at a new location $\boldsymbol{s}_0$, given a vector of predictors $\boldsymbol{x}_0 = \boldsymbol{x}(\boldsymbol{s}_0)$, by computing the predictive distribution

$$\begin{aligned} f(y_0 \mid \boldsymbol{x}_0, \boldsymbol{y}, \mathrm{X}) &= \int_{\mathbb{R}^p \times \mathbb{R}_+^3} \mathcal{L}(y_0, \boldsymbol{\beta}, \boldsymbol{\theta} \mid \boldsymbol{x}_0, \boldsymbol{y}, \mathrm{X}) \mathrm{d}\boldsymbol{\beta} \mathrm{d}\boldsymbol{\theta} \\ &= \int_{\mathbb{R}^p \times \mathbb{R}_+^3} \mathcal{L}(y_0 \mid \boldsymbol{x}_0, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{y}, \mathrm{X}) \pi(\boldsymbol{\beta}, \boldsymbol{\theta} \mid \boldsymbol{y}, \mathrm{X}) \mathrm{d}\boldsymbol{\beta} \mathrm{d}\boldsymbol{\theta} \end{aligned}$$

where $\pi(\boldsymbol{\beta}, \boldsymbol{\theta} \mid \boldsymbol{y}, \mathrm{X})$ denotes the posterior distribution of interest.

- The previous integral can be solved numerically, starting from an MCMC output.
- In practice, we can compute directly the quantity we need in STAN.
- Under the Gaussian model, we can write explicitly $\mathcal{L}(y_0 \mid \boldsymbol{x}_0, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{y}, \mathrm{X})$.

Point-referenced data

Recall that, from standard multivariate Gaussian proprieties, if

$$\begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} \right)$$

with $\Omega_{12} = \Omega_{21}^\intercal$. Then, the conditional distribution of $\boldsymbol{Y}_1 \mid \boldsymbol{Y}_2$ is still a Gaussian distribution, with mean and covariance matrix

$$\begin{aligned} \mathrm{E}[\boldsymbol{Y}_1 \mid \boldsymbol{Y}_2] &= \boldsymbol{\mu}_1 + \Omega_{12}\Omega_{22}^{-1}(\boldsymbol{Y}_2 - \boldsymbol{\mu}_2) \\ \mathsf{var}(\boldsymbol{Y}_1 \mid \boldsymbol{Y}_2) &= \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21} \end{aligned}$$

In our framework, we have $\boldsymbol{Y}_1 = Y(\boldsymbol{s}_0)$, $\boldsymbol{Y}_2 = \boldsymbol{y}$. Then, we have

$$\Omega_{11} = \sigma^2 + \tau^2, \quad \Omega_{12} = \boldsymbol{\gamma}^\intercal, \quad \Omega_{22} = \sigma^2 H(\phi) + \tau^2 \mathrm{I}_n,$$

where $\boldsymbol{\gamma}^\intercal = (\sigma^2\rho(d_{01}, \phi), \dots, \sigma^2\rho(d_{0n}, \phi))$. Hence,

$$\begin{aligned} \mathrm{E}[Y_0 \mid \boldsymbol{y}] &= \boldsymbol{x}_0^\intercal\boldsymbol{\beta} + \boldsymbol{\gamma}^\intercal(\sigma^2 H(\phi) + \tau^2\mathrm{I}_n)^{-1}(\boldsymbol{y} - \mathrm{X}\boldsymbol{\beta}) \\ \mathsf{var}(\boldsymbol{Y}_1 \mid \boldsymbol{Y}_2) &= \sigma^2 + \tau^2 - \boldsymbol{\gamma}^\intercal(\sigma^2 H(\phi) + \tau^2\mathrm{I}_n)^{-1}\boldsymbol{\gamma} \end{aligned}$$

Areal data

Areal data

The second type of spatial data we are considering consists of **areal data**. We recall that areal data $Y(\boldsymbol{s})$, $\boldsymbol{s} \in D$, consists of a spatial data where the domain D is partitioned into a finite number of areal units.

Hence, our realizations are $\boldsymbol{Y} = (Y_1, \dots, Y_n)$, continuous, binary, count, etc., associated to n distinct areal units $\boldsymbol{S} = \{S_1, \dots, S_n\}$.

We also have a $n \times n$ matrix, here called W, that describes how different areas are, in some way, connected. Typically, we set $w_{ii} = 0$, $i = 1, \dots, n$, i.e. an observation is not connected with itself. Further, we have

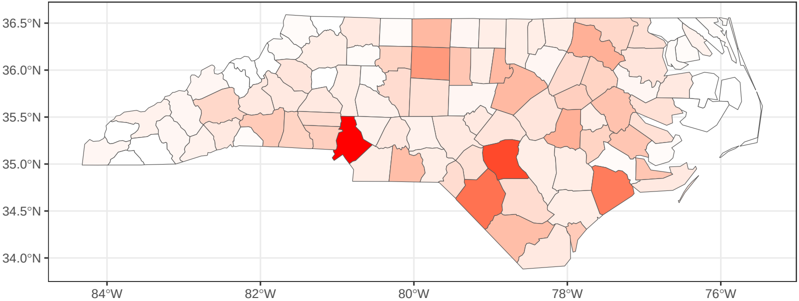
- → $w_{ij} = 1$ if the ith and the jth area share at least a common boundary.
- → w_{ij} could reflect the distance among units, e.g. a decreasing function of intercentroidal distance.

W is usually a symmetric matrix, and it can be marginally standardized by defining

$$\tilde{w}_{ij} = \frac{w_{ij}}{w_{i+}}, \qquad w_{i+} = \sum_{j=1}^{n} w_{ij}$$

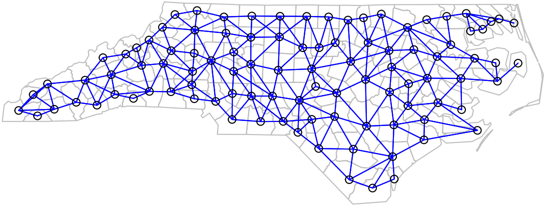
Is not symmetric anymore, but is a stochastic matrix, with row summing up to 1.

Point-referenced data



Sudden infant deaths in North Carolina for 1974-78. We also have access to other information for each specific areas, such as number of births and number of non-white birth.

Point-referenced data



The corresponding graph connecting different areas, and defining the adjacency matrix.

Areal data

Typical quantities to measure the strength of spatial association among different areal units are the following.

- **Moran's I**, which is defined as

  $$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{(\sum_{i \neq j} w_{ij}) \sum_{i=n}^{n} (y_j - \bar{y})^2}$$

  which is the analogue of lagged autocorrelation for time series. By construction, is not constricted in $[-1, 1]$.

- **Geary's C**, which is defined as

  $$C = \frac{(n-1) \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(y_i - \bar{y})^2}{2(\sum_{i \neq j} w_{ij}) \sum_{i=n}^{n} (y_j - \bar{y})^2}$$

  The C index is never negative. Small values (between 0 and 1) indicate positive spatial association.

Areal data

A relevant issue here is how we can specify a **joint distribution**, playing the role of the likelihood term, for $\boldsymbol{Y} = (Y_1, \dots, Y_n)^\intercal$, that incorporates also the spatial dependence we across areas.

A possible strategy that we can explore is to build up the **joint distribution** starting from the **full conditional distributions** of each observation, giving all the others, i.e. $\mathcal{L}(Y_i \mid \boldsymbol{Y}_{-i})$, where $\boldsymbol{Y}_{-i}$ denotes all the observations discarding the ith term.

Problem: the joint distribution can be determined by the product of the full conditionals, but the joint distribution **can be improper**.

Instead of considering the whole support, we denote by ∂_i a generic neighborhood of i. Suppose we specify the full conditionals in a local fashion, by considering

$$\mathcal{L}(Y_i \mid \boldsymbol{Y}_{-i}) = \mathcal{L}(Y_i \mid Y_j \in \partial_i), \quad i = 1, \dots, n.$$

By specify the full conditionals in the previous way, we identify a unique joint distribution, while we are inside the Markov random field domain.

Conditionally autoregressive (CAR) models are an example of Marov random fields, where the joint distribution is a Gibbs distribution, i.e. it exists but it can be improper.

Areal data

We introduce the **CAR** case with continuous Y_is, Gaussian distributed. The same framework can be extended more in general to **exponential family models**. We then set, for each single observation, a model of the form

$$Y_i \mid \boldsymbol{y}_{-i} \sim N\left(\sum_{j\neq i} b_{ij} y_j, \tau_i^2\right), \quad i = 1, \ldots, n.$$

These **full conditional are compatible**, and we obtain as joint distribution

$$f(y_1, \ldots, y_n) \propto \exp\left\{-\frac{1}{2}\boldsymbol{y}^\intercal D^{-1}(\mathrm{I}_n - B)\boldsymbol{y}\right\},$$

where $[B]_{ij} = b_{ij}$ and $D = \mathrm{diag}(\tau_1^2, \ldots, \tau_n^2)$.

Ideally, the previous expression suggests us a joint multivariate normal distribution for $\boldsymbol{Y}$, with $\mathbf{0}$ mean and covariance matrix $\Sigma = (\mathrm{I}_n - B)^{-1} D$.

But we should be careful, as Σ^{-1} and hence $D^{-1}(\mathrm{I}_n - B)$ must be symmetric and nonsingular.

Areal data

First, to **enforce symmetry** in $D^{-1}(\mathrm{I}_n - B)$, we should satisfy the following conditions

$$\frac{b_{ij}}{\tau_i^2} = \frac{b_{ji}}{\tau_j^2}, \qquad i, j = 1, \ldots, n.$$

From the previous, is apparent that B does not need to be symmetric.

From the definition of the Gaussian CAR model, b_{ij} relates observation i and j. Hence, we can return back to our proximity matrix W, which is assumed to be symmetric.

Suppose now that we set $b_{ij} = w_{ij}/w_{i+}$ and $\tau_i^2 = \tau^2/w_{i+}$. Then the previous is satisfied since W is symmetric and

$$\frac{w_{ij} w_{i+}}{\tau_2 w_{i+}} = \frac{w_{ji} w_{j+}}{\tau_2 w_{j+}}, \qquad i, j = 1, \ldots, n,$$

leading to full conditionals of the form

$$Y_i \mid \boldsymbol{y}_{-i} \sim N\left(\sum_{j=1}^{n} \frac{w_{ij}}{w_{i+}} y_j, \frac{\tau^2}{w_{i+}} \right).$$

Areal data

From the previous full conditionals, the **joint distribution** we obtain takes form

$$f(y_1, \ldots, y_n) \propto \exp\left(-\frac{1}{2\tau^2}\boldsymbol{y}^\intercal(D_w - W)\boldsymbol{y}\right),$$

where D_w is a diagonal matrix with $[D_w]_{ii} = w_{i+}$.

Secondly, we can note a second aspect. Unfortunately, within the previous construction we have

$$(D_w - W)\mathbf{1} = \mathbf{0},$$

hence $(D_w - W) = \Sigma^{-1}$ is singular, so that Σ does not exist.

Note that, while for Σ singular we do not have a density function, but a distribution that lives in a lower dimensional space, when Σ^{-1} is singular we do have a density function, but not integrable, hence improper.

With some algebra, the previous density function can be rewritten as

$$f(y_1, \ldots, y_n) \propto \exp\left(-\frac{1}{2\tau^2}\sum_{i \neq j} w_{ij}(y_i - y_j)^2\right),$$

The impropery is still apparent from the previous, we can add any constant to all the Y_is, and the previous is unaffected. However, it can still be used as improper model. The previous is usually referred to as intrinsically autoregressive (IAR) model.

Areal data

A slight variation of the previous model gives us a proper distribution. We redefine $\Sigma^{-1} = (D_w - \rho W)$, by suitably choosing ρ such that Σ^{-1} is nonsingular.

The nonsingularity is guaranteed by setting $\rho \in (1/\lambda_{(1)}, 1/\lambda_{(n)})$, where $\lambda_{(1)} < \lambda_{(2)} \cdots < \lambda_{(n)}$ are the ordered eigenvalues of $D_w^{1/2} W D_w^{-1/2}$.

Moreover, since $\mathrm{tr}(D_w^{1/2} W D_w^{-1/2}) = 0 = \sum_{i=1}^{n} \lambda_{(i)}$, then we have $\lambda_{(1)} < 0$, $\lambda_{(n)} > 0$ and 0 being in the set of interest $(1/\lambda_{(1)}, 1/\lambda_{(n)})$.

Simpler bounds can be alternatively obtained by looking at the scaled matrix we defined above, $\tilde{W} = \mathrm{diag}(1/w_{1+}, \ldots, 1/w_{n+})W$. Such a matrix is not symmetric, but is row stochastic (i.e. all of its rows sum to 1).

Then, Σ^{-1} can be written as $M^{-1}(\mathrm{I}_n - \alpha \tilde{W})$, where M is diagonal. Further, if $|\alpha| < 1$, then $(\mathrm{I}_n - \alpha \tilde{W})$ is nonsingular.

Areal data

Hence, under the first constrain, with $\Sigma^{-1} = (D_w - \rho W)$, with W symmetric matrix, we have

$$Y_i \mid \boldsymbol{y}_{-i} \sim N\left(\rho \sum_{j=1}^{n} \frac{w_{ij}}{w_{i+}} y_j, \frac{\tau^2}{w_{i+}}\right), \qquad i = 1, \ldots, n.$$

Typically, we set $\rho \in (0, 1)$. For the boundary values, we have the following.

- If $\rho = 0$, then
$$Y_i \overset{iid}{\sim} N(0, \tau^2/w_{i+}), \qquad i = 1, \ldots, n.$$
- If $\rho = 1$, then we are back to the improper intrinsic CAR model.

Usually we set a prior on ρ.

→ A prior with mass near to 1 encourages spatial association among different areas.

The model can be used directly as distribution for the data, or can be combined in more complex models where the spatial association is dictated by a latent parameter level.

Areal data

In a general setting, we consider

- $\boldsymbol{S} = \{S_1, \ldots, S_n\}$ areal units.
- $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\intercal$ response variables.
- $\boldsymbol{O} = \{O_1, \ldots, O_n\}$ offsets, additional information that we have on the areal units.

The spatial pattern in the response is modelled by

- X a matrix of covarites, where $\boldsymbol{x}_i^\intercal = (x_{i1}, \ldots, x_{ip})$ is the covariate vector associated to the ith area.
- A set of random effects $\boldsymbol{\phi} = \{\phi_1, \ldots, \phi_n\}$.

We consider a generic case arising from GLMM. The model specification is given by

$$\begin{aligned}
Y_i \mid \mu_i &\overset{ind}{\sim} f(y_i \mid \mu_i, \nu^2), \qquad i = 1, \ldots, n, \\
g(\mu_i) &= \boldsymbol{x}_i^\intercal \boldsymbol{\beta} + \phi_i + O_i, \qquad i = 1, \ldots, n, \\
\boldsymbol{\beta} &\sim N(\boldsymbol{b}_0, \Lambda_0), \\
\nu^2 &\sim IG(a_\nu, b_\nu), \\
(\phi_1, \ldots, \phi_n) &\sim CAR(W, \rho, \tau^2), \\
\rho &\sim Beta(a_\rho, b_\rho), \\
\tau^2 &\sim IG(a_\tau, b_\tau).
\end{aligned}$$

Areal data

Regarding specific distributional assumption for $f(y_i \mid \mu_i, \nu^2)$, we can assume the usual ones.

- Gaussian, $Y_i \sim N(\mu_i, \nu^2)$ and $\mu_i = \boldsymbol{x}_i^\intercal \boldsymbol{\beta} + \phi_i + O_i$.
- Bernoulli, $Y_i \sim Be(\theta_i)$ and
$$\mu_i = \log(\theta_i/(1-\theta_i)) = \boldsymbol{x}_i^\intercal \boldsymbol{\beta} + \phi_i + O_i.$$
- Poisson, $Y_I \sim Poi(|mu_i)$ and $\log(\mu_i) = \boldsymbol{x}_i^\intercal \boldsymbol{\beta} + \phi_i + O_i$.

The offset is needed in case we want to adjust what we observe on some feature of the area, such as the dimension. Suppose for example that we are interested into model rates, but we observe counts. For instance, suppose we observe a count Y in an area with surface O. Our model is a Poisson with $\mathrm{E}[Y \mid x] = \mu_x$, but we are interested in

$$\log \frac{\mu_x}{O} = \beta_0 + \beta_1 x \qquad \Longrightarrow \qquad \log \mu_x = \log(O) + \beta_0 + \beta_1 x.$$

Then $\log(O)$ is the corresponding offset.