Bayesian Spatial Analysis

Dr. Jarad Niemi

STAT 615 - Iowa State University

November 9, 2017

Spatial modeling

Three main types of spatial data:

- Point/geo-referenced
- Areal-referenced
- Point process/pattern

Point-referenced spatial data

Features

- Some spatial domain $\mathcal{D}$ is under study

Examples

Point-referenced spatial data

Features

- Some spatial domain $\mathcal{D}$ is under study
- Measured spatial locations $s \in \mathcal{D}$ are pre-determined

Examples

Point-referenced spatial data

Features

- Some spatial domain $\mathcal{D}$ is under study
- Measured spatial locations $s \in \mathcal{D}$ are pre-determined
- Some quantity, $Y(s)$, is measured at each location $s \in \mathcal{D}$

Examples

Point-referenced spatial data

Features

- Some spatial domain $\mathcal{D}$ is under study
- Measured spatial locations $s \in \mathcal{D}$ are pre-determined
- Some quantity, $Y(s)$, is measured at each location $s \in \mathcal{D}$

Examples

- Air quality monitoring

Point-referenced spatial data

Features

- Some spatial domain $\mathcal{D}$ is under study
- Measured spatial locations $s \in \mathcal{D}$ are pre-determined
- Some quantity, $Y(s)$, is measured at each location $s \in \mathcal{D}$

Examples

- Air quality monitoring
- Coastal tide level monitoring

Point-referenced spatial data

Features

- Some spatial domain $\mathcal{D}$ is under study
- Measured spatial locations $s \in \mathcal{D}$ are pre-determined
- Some quantity, $Y(s)$, is measured at each location $s \in \mathcal{D}$

Examples

- Air quality monitoring
- Coastal tide level monitoring
- Earthquake monitoring

Point-referenced spatial data

Features

- Some spatial domain $\mathcal{D}$ is under study
- Measured spatial locations $s \in \mathcal{D}$ are pre-determined
- Some quantity, $Y(s)$, is measured at each location $s \in \mathcal{D}$

Examples

- Air quality monitoring
- Coastal tide level monitoring
- Earthquake monitoring
- Bird point counts

Areal-referenced spatial data

Features

- Some set of spatial regions $1, \ldots, S$ are pre-determined

Examples

Areal-referenced spatial data

Features

- Some set of spatial regions $1, \ldots, S$ are pre-determined
- Some quantity, Y_s, is measured as an aggregate over that region

Examples

Areal-referenced spatial data

Features

- Some set of spatial regions $1, \ldots, S$ are pre-determined
- Some quantity, Y_s, is measured as an aggregate over that region

Examples

- Disease occurrence per county

Areal-referenced spatial data

Features

- Some set of spatial regions $1, \ldots, S$ are pre-determined
- Some quantity, Y_s, is measured as an aggregate over that region

Examples

- Disease occurrence per county
- Unemployment rate per state

Areal-referenced spatial data

Features

- Some set of spatial regions $1, \ldots, S$ are pre-determined
- Some quantity, Y_s, is measured as an aggregate over that region

Examples

- Disease occurrence per county
- Unemployment rate per state
- Inflation per country

Point-process spatial data

Features

- Some spatial domain $\mathcal{D}$ is under study

Examples

Point-process spatial data

Features

- Some spatial domain $\mathcal{D}$ is under study
- Spatial locations $s \in \mathcal{S} \subset \mathcal{D}$ are random

Examples

Point-process spatial data

Features

- Some spatial domain $\mathcal{D}$ is under study
- Spatial locations $s \in \mathcal{S} \subset \mathcal{D}$ are random
- $Y(s) = 1$ indicates an occurence of the event

Examples

Point-process spatial data

Features

- Some spatial domain $\mathcal{D}$ is under study
- Spatial locations $s \in \mathcal{S} \subset \mathcal{D}$ are random
- $Y(s) = 1$ indicates an occurence of the event

Examples

- Locations of Mayan ruins

Point-process spatial data

Features

- Some spatial domain $\mathcal{D}$ is under study
- Spatial locations $s \in \mathcal{S} \subset \mathcal{D}$ are random
- $Y(s) = 1$ indicates an occurence of the event

Examples

- Locations of Mayan ruins
- Locations of invasive species

Point-process spatial data

Features

- Some spatial domain $\mathcal{D}$ is under study
- Spatial locations $s \in \mathcal{S} \subset \mathcal{D}$ are random
- $Y(s) = 1$ indicates an occurence of the event

Examples

- Locations of Mayan ruins
- Locations of invasive species
- Locations of caught Lingcod

Point-referenced spatial data

Let $Y(s)$ for $s \in \mathcal{D} \subseteq \mathbb{R}^d$ be a spatial process.

Point-referenced spatial data

Let $Y(s)$ for $s \in \mathcal{D} \subseteq \mathbb{R}^d$ be a spatial process. Let $E[Y(s)] = 0$ for all $s \in \mathcal{D}$ because we will model the mean separately.

Point-referenced spatial data

Let $Y(s)$ for $s \in \mathcal{D} \subseteq \mathbb{R}^d$ be a spatial process. Let $E[Y(s)] = 0$ for all $s \in \mathcal{D}$ because we will model the mean separately.

Assumptions:

- Stationarity

Point-referenced spatial data

Let $Y(s)$ for $s \in \mathcal{D} \subseteq \mathbb{R}^d$ be a spatial process. Let $E[Y(s)] = 0$ for all $s \in \mathcal{D}$ because we will model the mean separately.

Assumptions:

- Stationarity
  - Intrinsic stationarity

Point-referenced spatial data

Let $Y(s)$ for $s \in \mathcal{D} \subseteq \mathbb{R}^d$ be a spatial process. Let $E[Y(s)] = 0$ for all $s \in \mathcal{D}$ because we will model the mean separately.

Assumptions:

- Stationarity
    - Intrinsic stationarity
    - Weak stationarity

Point-referenced spatial data

Let $Y(s)$ for $s \in \mathcal{D} \subseteq \mathbb{R}^d$ be a spatial process. Let $E[Y(s)] = 0$ for all $s \in \mathcal{D}$ because we will model the mean separately.

Assumptions:

- Stationarity
  - Intrinsic stationarity
  - Weak stationarity
  - Strong stationarity

Point-referenced spatial data

Let $Y(s)$ for $s \in \mathcal{D} \subseteq \mathbb{R}^d$ be a spatial process. Let $E[Y(s)] = 0$ for all $s \in \mathcal{D}$ because we will model the mean separately.

Assumptions:

- Stationarity
  - Intrinsic stationarity
  - Weak stationarity
  - Strong stationarity
- Isotropy

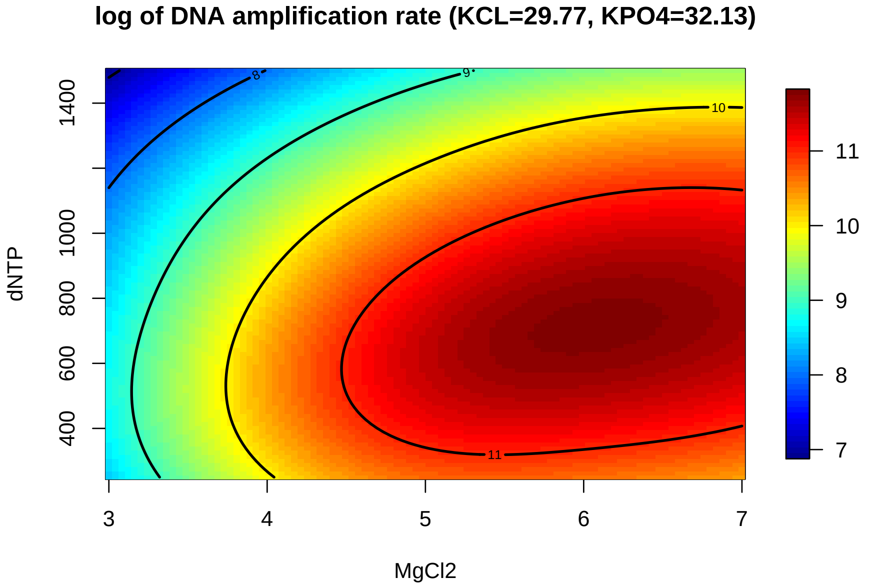
Point-referenced spatial data

Let $Y(s)$ for $s \in \mathcal{D} \subseteq \mathbb{R}^d$ be a spatial process. Let $E[Y(s)] = 0$ for all $s \in \mathcal{D}$ because we will model the mean separately.

Assumptions:

- Stationarity
  - Intrinsic stationarity
  - Weak stationarity
  - Strong stationarity
- Isotropy
- Gaussian process

Example spatial process



Intrinsic stationarity

Definition

A process $Y(s)$ is intrinsically stationary if ($E[Y(s+h) - Y(s)] = 0$ and)

$$E[(Y(s+h) - Y(s))^2] = Var[Y(s+h) - Y(s)] = 2\gamma(h)$$

when $s, s+h \in \mathcal{D}$.

Intrinsic stationarity

Definition

A process $Y(s)$ is intrinsically stationary if ($E[Y(s+h) - Y(s)] = 0$ and)

$$E[(Y(s+h) - Y(s))^2] = Var[Y(s+h) - Y(s)] = 2\gamma(h)$$

when $s, s+h \in \mathcal{D}$. We call $2\gamma(h)$ the variogram and $\gamma(h)$ the semivariogram.

Intrinsic stationarity

Definition

A process $Y(s)$ is intrinsically stationary if ($E[Y(s+h) - Y(s)] = 0$ and)

$$E[(Y(s+h) - Y(s))^2] = Var[Y(s+h) - Y(s)] = 2\gamma(h)$$

when $s, s+h \in \mathcal{D}$. We call $2\gamma(h)$ the variogram and $\gamma(h)$ the semivariogram.

Definition

A process $Y(s)$ is isotropic if the semivariogram function depends only on $||h||$, the length of the separation vector. Otherwise the process is anisotropic.

Weak stationarity

Definition

A process $Y(s)$ has weak stationarity if ($E[Y(s)] = \mu$ and) $Cov[Y(s), Y(s+h)] = C(h)$ when $s, s+h \in \mathcal{D}$.

Weak stationarity

Definition

A process $Y(s)$ has weak stationarity if ($E[Y(s)] = \mu$ and) $Cov[Y(s), Y(s+h)] = C(h)$ when $s, s+h \in \mathcal{D}$. We call $C(h)$ the covariance function or covariogram.

Weak stationarity

Definition

A process $Y(s)$ has weak stationarity if ($E[Y(s)] = \mu$ and) $Cov[Y(s), Y(s+h)] = C(h)$ when $s, s+h \in \mathcal{D}$. We call $C(h)$ the covariance function or covariogram.

Since $\gamma(h) = C(0) - C(h)$, a weakly stationary process is also intrinsicly stationary.

Weak stationarity

Definition

A process $Y(s)$ has weak stationarity if ($E[Y(s)] = \mu$ and) $Cov[Y(s), Y(s+h)] = C(h)$ when $s, s+h \in \mathcal{D}$. We call $C(h)$ the covariance function or covariogram.

Since $\gamma(h) = C(0) - C(h)$, a weakly stationary process is also intrinsicly stationary.

If the spatial process is ergodic, then $C(h) \to 0$ as $||h|| \to \infty$

Weak stationarity

Definition

A process $Y(s)$ has weak stationarity if ($E[Y(s)] = \mu$ and) $Cov[Y(s), Y(s+h)] = C(h)$ when $s, s+h \in \mathcal{D}$. We call $C(h)$ the covariance function or covariogram.

Since $\gamma(h) = C(0) - C(h)$, a weakly stationary process is also intrinsicly stationary.

If the spatial process is ergodic, then $C(h) \to 0$ as $||h|| \to \infty$ and $\lim_{||h|| \to \infty} \gamma(h) = C(0)$.

Weak stationarity

Definition

A process $Y(s)$ has weak stationarity if ($E[Y(s)] = \mu$ and) $Cov[Y(s), Y(s+h)] = C(h)$ when $s, s+h \in \mathcal{D}$. We call $C(h)$ the covariance function or covariogram.

Since $\gamma(h) = C(0) - C(h)$, a weakly stationary process is also intrinsicly stationary.

If the spatial process is ergodic, then $C(h) \to 0$ as $||h|| \to \infty$ and $\lim_{||h|| \to \infty} \gamma(h) = C(0)$. Thus

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

Weak stationarity

Definition

A process $Y(s)$ has weak stationarity if ($E[Y(s)] = \mu$ and) $Cov[Y(s), Y(s+h)] = C(h)$ when $s, s+h \in \mathcal{D}$. We call $C(h)$ the covariance function or covariogram.

Since $\gamma(h) = C(0) - C(h)$, a weakly stationary process is also intrinsicly stationary.

If the spatial process is ergodic, then $C(h) \to 0$ as $||h|| \to \infty$ and $\lim_{||h|| \to \infty} \gamma(h) = C(0)$. Thus

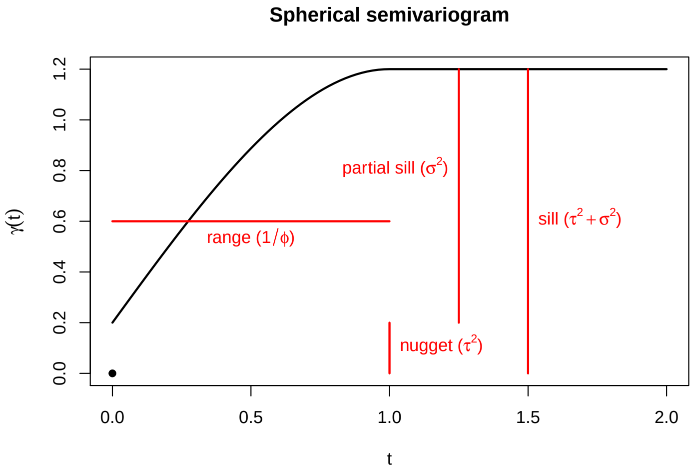
$$C(h) = C(0) - \gamma(h) = \lim_{||u|| \to \infty} \gamma(u) - \gamma(h).$$

Thus, if the process is ergodic, an intrinsicly stationary process is also weakly stationary.

Covariance functions for isotropic models

Model	Covariance function, $C(t)$	Semivariogram, $\gamma(t)$
Linear	$C(t)$ does not exist	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 t & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$
Spherical	$C(t) = \begin{cases} 0 \\ \sigma^2 \left[1 - \frac{3}{2}\phi t + \frac{1}{2}(\phi t)^3\right] \\ \tau^2 + \sigma^2 \end{cases}$	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 & \text{if} \\ \tau^2 + \sigma^2 \left[\frac{3}{2}\phi t - \frac{1}{2}(\phi t)^3\right] & 0 \\ 0 & \text{o} \end{cases}$
Exponential	$C(t) = \begin{cases} \sigma^2 \exp(-\phi t) \\ \tau^2 + \sigma^2 \end{cases}$	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 [1 - \exp(-\phi t)] & t > \\ 0 & \text{oth} \end{cases}$
Powered exponential	$C(t) = \begin{cases} \sigma^2 \exp(-\lvert\phi t\rvert^p) \\ \tau^2 + \sigma^2 \end{cases}$	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 [1 - \exp(-\lvert\phi t\rvert^p)] \\ 0 \end{cases}$
Gaussian	$C(t) = \begin{cases} \sigma^2 \exp(-\phi^2 t^2) \\ \tau^2 + \sigma^2 \end{cases}$	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 \left[1 - \exp(-\phi^2 t^2)\right] \\ 0 \end{cases}$
Rational quadratic	$C(t) = \begin{cases} \sigma^2 \left(1 - \frac{t^2}{(1+\phi^2)}\right) \\ \tau^2 + \sigma^2 \end{cases}$	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 \frac{t^2}{(1+\phi^2)} & t > 0 \\ 0 & \text{otherwise} \end{cases}$
Wave	$C(t) = \begin{cases} \sigma^2 \frac{\sin(\phi t)}{\phi t} \\ \tau^2 + \sigma^2 \end{cases}$	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 \left[1 - \frac{\sin(\phi t)}{\phi t}\right] & t > \\ 0 & \text{other} \end{cases}$
Power law	$C(t)$ does not exist	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 t^\lambda & t > 0 \\ 0 & \text{otherwise} \end{cases}$
Matérn	$C(t) = \begin{cases} \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (2\sqrt{v} t\phi)^\nu K_\nu(2\sqrt{\nu} t\phi) \\ \tau^2 + \sigma^2 \end{cases}$	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 \left[1 - \frac{(2\sqrt{\nu} t\phi)^\nu}{2^{\nu-1}\Gamma(\nu)} (2\sqrt{v} \right. \\ 0 \end{cases}$
Matérn ($\nu = 3/2$)	$C(t) = \begin{cases} \sigma^2 (1 + \phi t) \exp(-\phi t) \\ \tau^2 + \sigma^2 \end{cases}$	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 [1 - (1 + \phi t) \exp(-\phi \\ 0 \end{cases}$

Spherical semivariogram



Matérn

Perhaps the most important isotropic process is the Matérn process with covariance

$$C(t) = \begin{cases} \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(2\sqrt{\nu}t\phi)^\nu K_\nu(2\sqrt{\nu}t\phi) & t > 0 \\ \tau^2 + \sigma^2 & t = 0 \end{cases}$$

and variogram

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 \left[1 - \frac{(2\sqrt{\nu}t\phi)^\nu}{2^{\nu-1}\Gamma(\nu)} K_\nu(2\sqrt{\nu}t\phi)\right] & t > 0 \\ 0 & \text{otherwise} \end{cases}$$

Matérn

Perhaps the most important isotropic process is the Matérn process with covariance

$$C(t) = \begin{cases} \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(2\sqrt{\nu}t\phi)^\nu K_\nu(2\sqrt{\nu}t\phi) & t > 0 \\ \tau^2 + \sigma^2 & t = 0 \end{cases}$$

and variogram

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 \left[1 - \frac{(2\sqrt{\nu}t\phi)^\nu}{2^{\nu-1}\Gamma(\nu)} K_\nu(2\sqrt{\nu}t\phi)\right] & t > 0 \\ 0 & \text{otherwise} \end{cases}$$

where

- ν controls the smoothness of the spatial process ($\lfloor\nu\rfloor$ number of times process realizations are mean square differentiable) while
- ϕ is a spatial scale parameter.

Matérn

Perhaps the most important isotropic process is the Matérn process with covariance

$$C(t) = \begin{cases} \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(2\sqrt{\nu}t\phi)^\nu K_\nu(2\sqrt{\nu}t\phi) & t > 0 \\ \tau^2 + \sigma^2 & t = 0 \end{cases}$$

and variogram

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 \left[1 - \frac{(2\sqrt{\nu}t\phi)^\nu}{2^{\nu-1}\Gamma(\nu)} K_\nu(2\sqrt{\nu}t\phi)\right] & t > 0 \\ 0 & \text{otherwise} \end{cases}$$

where

- ν controls the smoothness of the spatial process ($\lfloor \nu \rfloor$ number of times process realizations are mean square differentiable) while
- ϕ is a spatial scale parameter.

Special cases are the exponential ($\nu = 1/2$) and Gaussian ($\nu \to \infty$).

Strong stationarity

Definition

A process $Y(s)$ is strongly (or strictly) stationary if, for any set of $n \geq 1$ sites $\{s_1, \ldots, s_n\}$ and any $h \in \mathbb{R}^d$,

$$(Y(s_1), \ldots, Y(s_n))^\top \stackrel{d}{=} (Y(s_1 + h), \ldots, Y(s_n + h))^\top$$

where $\stackrel{d}{=}$ means equal in distribution.

Strong stationarity

Definition

A process $Y(s)$ is strongly (or strictly) stationary if, for any set of $n \geq 1$ sites $\{s_1, \ldots, s_n\}$ and any $h \in \mathbb{R}^d$,

$$(Y(s_1), \ldots, Y(s_n))^\top \stackrel{d}{=} (Y(s_1 + h), \ldots, Y(s_n + h))^\top$$

where $\stackrel{d}{=}$ means equal in distribution.

If we assume all variances exist, then strong stationarity implies weak stationarity.

Strong stationarity

Definition

A process $Y(s)$ is strongly (or strictly) stationary if, for any set of $n \geq 1$ sites $\{s_1, \ldots, s_n\}$ and any $h \in \mathbb{R}^d$,

$$(Y(s_1), \ldots, Y(s_n))^\top \stackrel{d}{=} (Y(s_1 + h), \ldots, Y(s_n + h))^\top$$

where $\stackrel{d}{=}$ means equal in distribution.

If we assume all variances exist, then strong stationarity implies weak stationarity.

The reverse is not necessarily true.

Gaussian process

Definition

$Y(s)$ is a Gaussian process if, for any $n \geq 1$ and any set of sites $\{s_1, \ldots, s_n\}$, $Y = (Y(s_1), \ldots, Y(s_n))^\top$ has a multivariate normal distribution.

Gaussian process

Definition

$Y(s)$ is a Gaussian process if, for any $n \geq 1$ and any set of sites $\{s_1, \ldots, s_n\}$, $Y = (Y(s_1), \ldots, Y(s_n))^\top$ has a multivariate normal distribution.

For a Gaussian process, weak stationarity and strong stationarity are equivalent.

Bayesian estimation of Gaussian process parameters

Suppose we observe data at some locations $s_1, \ldots, s_n$.

Bayesian estimation of Gaussian process parameters

Suppose we observe data at some locations $s_1, \ldots, s_n$. Collectively, we have $y = (y(s_1), \ldots, y(s_n))$.

Bayesian estimation of Gaussian process parameters

Suppose we observe data at some locations $s_1, \ldots, s_n$. Collectively, we have $y = (y(s_1), \ldots, y(s_n))$. Let's assume the data arise from a Gaussian process

Bayesian estimation of Gaussian process parameters

Suppose we observe data at some locations $s_1, \ldots, s_n$. Collectively, we have $y = (y(s_1), \ldots, y(s_n))$. Let's assume the data arise from a Gaussian process and according to a particular covariance function

Bayesian estimation of Gaussian process parameters

Suppose we observe data at some locations $s_1, \ldots, s_n$. Collectively, we have $y = (y(s_1), \ldots, y(s_n))$. Let's assume the data arise from a Gaussian process and according to a particular covariance function. Collectively refer to the parameters as θ, then our objective is

$$p(\theta|y) \propto p(y|\theta)p(\theta).$$

Bayesian estimation of Gaussian process parameters

Suppose we observe data at some locations $s_1, \ldots, s_n$. Collectively, we have $y = (y(s_1), \ldots, y(s_n))$. Let's assume the data arise from a Gaussian process and according to a particular covariance function. Collectively refer to the parameters as θ, then our objective is

$$p(\theta|y) \propto p(y|\theta)p(\theta).$$

Suppose we assume the Matérn covariance function and a common mean μ so that $\theta = (\mu, \nu, \phi, \tau^2, \sigma^2)$.

Bayesian estimation of Gaussian process parameters

Suppose we observe data at some locations $s_1, \ldots, s_n$. Collectively, we have $y = (y(s_1), \ldots, y(s_n))$. Let's assume the data arise from a Gaussian process and according to a particular covariance function. Collectively refer to the parameters as θ, then our objective is

$$p(\theta|y) \propto p(y|\theta)p(\theta).$$

Suppose we assume the Matérn covariance function and a common mean μ so that $\theta = (\mu, \nu, \phi, \tau^2, \sigma^2)$. Then we have

$$p(\mu, \nu, \phi, \tau^2, \sigma^2|y) \propto N(y; \mu, \Sigma)p(\mu, \nu, \phi, \tau^2, \sigma^2)$$

Bayesian estimation of Gaussian process parameters

Suppose we observe data at some locations $s_1, \ldots, s_n$. Collectively, we have $y = (y(s_1), \ldots, y(s_n))$. Let's assume the data arise from a Gaussian process and according to a particular covariance function. Collectively refer to the parameters as θ, then our objective is

$$p(\theta|y) \propto p(y|\theta)p(\theta).$$

Suppose we assume the Matérn covariance function and a common mean μ so that $\theta = (\mu, \nu, \phi, \tau^2, \sigma^2)$. Then we have

$$p(\mu, \nu, \phi, \tau^2, \sigma^2|y) \propto N(y; \mu, \Sigma)p(\mu, \nu, \phi, \tau^2, \sigma^2)$$

where Σ is constructed from the parameters ν, ϕ, τ^2, and σ^2 and the distances between locations, e.g. $||s_1 - s_2||$.

Consider point-referenced data at spatial locations $s_1, \ldots, s_n$, model this data as

$$Y(s) = \mu(s) + w(s) + \epsilon(s)$$

If we constrain ourselves to isotropic models, the Matérn class is suggested as a general tool (Banerjee pg. 37).

Consider point-referenced data at spatial locations $s_1, \ldots, s_n$, model this data as

$$Y(s) = \mu(s) + w(s) + \epsilon(s)$$

If we constrain ourselves to isotropic models, the Matérn class is suggested as a general tool (Banerjee pg. 37). If $w = (w(s_1), \ldots, w(s_n))^\top$ and $\epsilon = (\epsilon(s_1), \ldots, \epsilon(s_n))^\top$, then a general model is

Consider point-referenced data at spatial locations $s_1, \ldots, s_n$, model this data as

$$Y(s) = \mu(s) + w(s) + \epsilon(s)$$

If we constrain ourselves to isotropic models, the Matérn class is suggested as a general tool (Banerjee pg. 37). If $w = (w(s_1), \ldots, w(s_n))^\top$ and $\epsilon = (\epsilon(s_1), \ldots, \epsilon(s_n))^\top$, then a general model is

$$Var[w] = \sigma^2 H(\phi) \qquad Var[\epsilon] = \tau^2 \mathrm{I}$$

Consider point-referenced data at spatial locations $s_1, \ldots, s_n$, model this data as

$$Y(s) = \mu(s) + w(s) + \epsilon(s)$$

If we constrain ourselves to isotropic models, the Matérn class is suggested as a general tool (Banerjee pg. 37). If $w = (w(s_1), \ldots, w(s_n))^\top$ and $\epsilon = (\epsilon(s_1), \ldots, \epsilon(s_n))^\top$, then a general model is

$$Var[w] = \sigma^2 H(\phi) \qquad Var[\epsilon] = \tau^2 \mathrm{I}$$

where H is a correlation matrix with $H_{ij} = \rho(s_i - s_j; \phi)$ and ρ is a valid isotropic correlation function on $\mathbb{R}^r$, i.e. Matérn:

$$\rho(u; \nu, \phi) = \frac{(u/\phi)^\nu K_\nu(u/\phi)}{2^{\nu-1}\Gamma(\nu)}$$

as defined in `geoR:matern`.

Consider point-referenced data at spatial locations $s_1, \ldots, s_n$, model this data as

$$Y(s) = \mu(s) + w(s) + \epsilon(s)$$

If we constrain ourselves to isotropic models, the Matérn class is suggested as a general tool (Banerjee pg. 37). If $w = (w(s_1), \ldots, w(s_n))^\top$ and $\epsilon = (\epsilon(s_1), \ldots, \epsilon(s_n))^\top$, then a general model is

$$Var[w] = \sigma^2 H(\phi) \qquad Var[\epsilon] = \tau^2 \mathrm{I}$$

where H is a correlation matrix with $H_{ij} = \rho(s_i - s_j; \phi)$ and ρ is a valid isotropic correlation function on $\mathbb{R}^r$, i.e. Matérn:

$$\rho(u; \nu, \phi) = \frac{(u/\phi)^\nu K_\nu(u/\phi)}{2^{\nu-1}\Gamma(\nu)}$$

as defined in `geoR:matern`. The overall mean is modeled separately and uses covariates $x(s)$ via

$$\mu(s) = x(s)^\top \beta.$$

Bayesian estimation for spatial random effects

Let $\theta = (\beta, \sigma^2, \tau^2, \phi)$, then parameter estimates may be obtained from the posterior distribution:

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

Bayesian estimation for spatial random effects

Let $\theta = (\beta, \sigma^2, \tau^2, \phi)$, then parameter estimates may be obtained from the posterior distribution:

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

where

$$Y|\theta \sim N(X\beta, \sigma^2 H(\phi) + \tau^2 \mathrm{I}).$$

Bayesian estimation for spatial random effects

Let $\theta = (\beta, \sigma^2, \tau^2, \phi)$, then parameter estimates may be obtained from the posterior distribution:

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

where

$$Y|\theta \sim N(X\beta, \sigma^2 H(\phi) + \tau^2 \mathrm{I}).$$

Typically, independent priors are chosen so that

$$p(\theta) = p(\beta)p(\sigma^2)p(\tau^2)p(\phi).$$

As a general rule, non-informative priors can be chosen for β, e.g. $p(\beta) \propto 1$.

Bayesian estimation for spatial random effects

Let $\theta = (\beta, \sigma^2, \tau^2, \phi)$, then parameter estimates may be obtained from the posterior distribution:

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

where

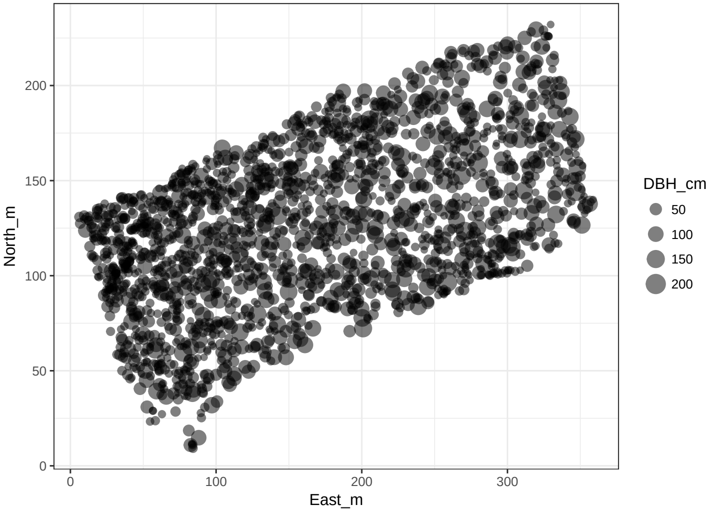
$$Y|\theta \sim N(X\beta, \sigma^2 H(\phi) + \tau^2 \mathrm{I}).$$

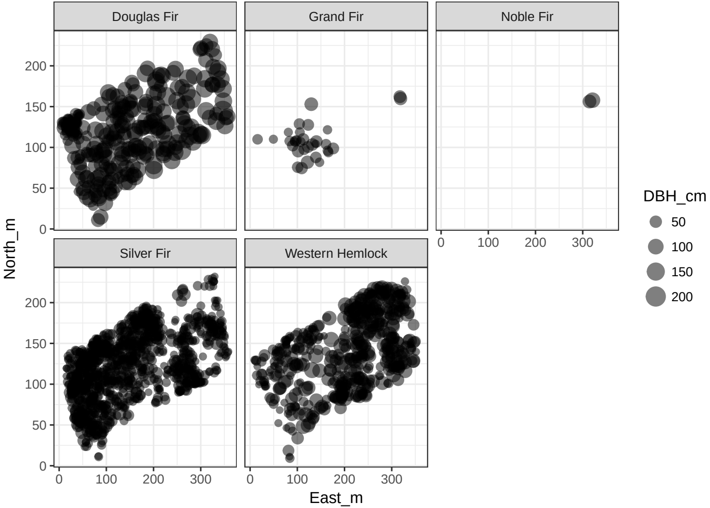
Typically, independent priors are chosen so that

$$p(\theta) = p(\beta)p(\sigma^2)p(\tau^2)p(\phi).$$

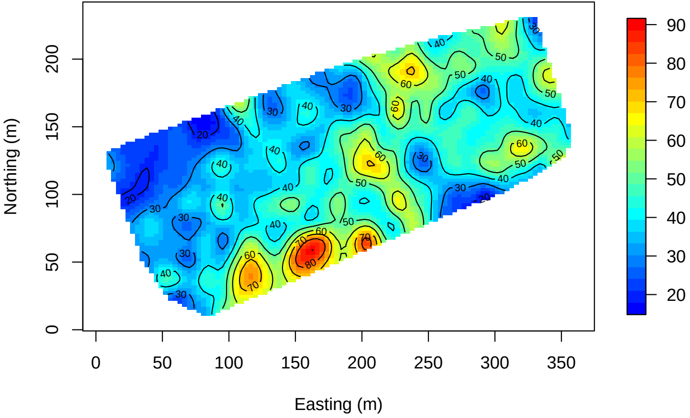
As a general rule, non-informative priors can be chosen for β, e.g. $p(\beta) \propto 1$. However, improper (or vague proper) priors for the variance parameters can lead to improper (or computationally improper) posteriors.

Diameter at breast height (DBH) for an experimental forest




Douglas Fir
Grand Fir
Noble Fir
Silver Fir
Western Hemlock
North_m
East_m
0
50
100
150
200
0
100
200
300
DBH_cm
50
100
150
200

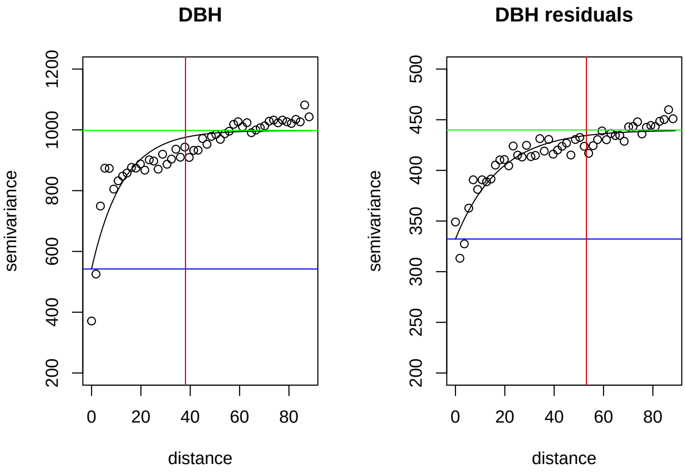
Interpolation of mean DBH (ignoring species)



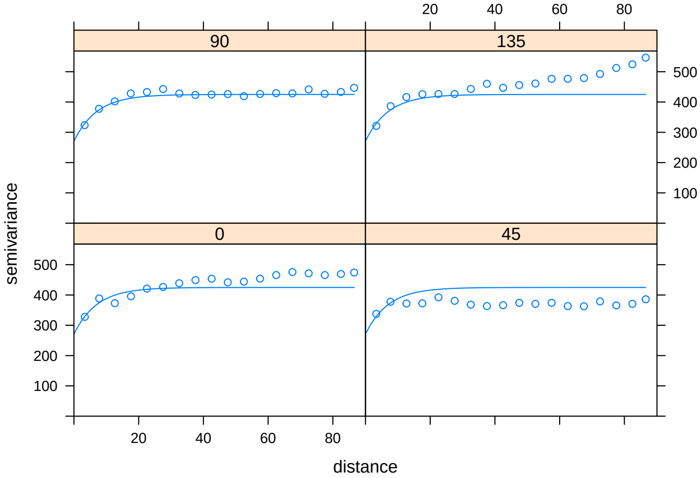
Regression

```
## variog: computing omnidirectional variogram
## variofit: covariance model used is exponential
## variofit: weights used: equal
## variofit: minimisation function used: nls
##
## Call:
## lm(formula = DBH_cm ~ Species, data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -78.423  -9.969  -3.561  10.924 118.277
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)              89.423      1.303  68.629   <2e-16 ***
## SpeciesGrand Fir        -51.598      4.133 -12.483   <2e-16 ***
## SpeciesNoble Fir         -5.873     15.744  -0.373    0.709
## SpeciesSilver Fir       -68.347      1.461 -46.784   <2e-16 ***
## SpeciesWestern Hemlock  -48.062      1.636 -29.377   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.19 on 1950 degrees of freedom
## Multiple R-squared:  0.5332,Adjusted R-squared:  0.5323
## F-statistic: 556.9 on 4 and 1950 DF,  p-value: < 2.2e-16
```

Variogram (exponential model)



Isotropy?



90
135
0
45
20
40
60
80
100
200
300
400
500
semivariance
distance

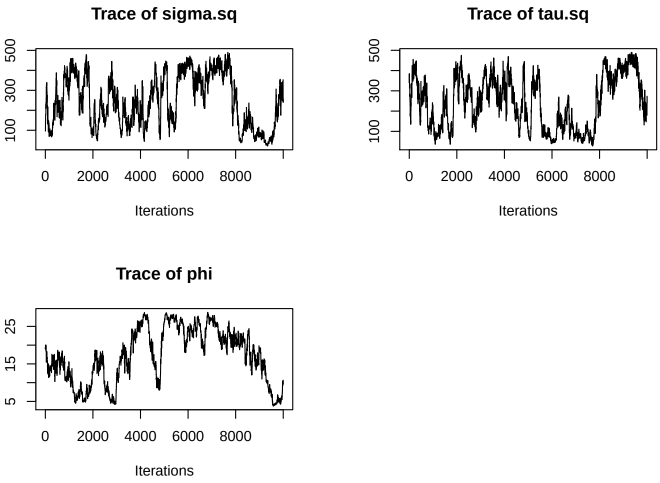
spBayes

```
p = nlevels(d$Species)
r = spLM(DBH_cm ~ Species,
        data = d,
        coords = as.matrix(d[c('East_m','North_m')]),
        knots = c(6,6,.1), # for spatial prediction
        cov.model = 'exponential',

        starting = list(tau.sq   = fit.DBH.resid$nugget,
                        sigma.sq = fit.DBH.resid$cov.pars[1],
                        phi      = fit.DBH.resid$cov.pars[2]),

        tuning = list(tau.sq   = 0.015,
                      sigma.sq = 0.015,
                      phi      = 0.015),

        priors = list(beta.Norm   = list(rep(0,p), diag(1000,p)),
                      phi.Unif    = c(3/1,3/0.1),
                      sigma.sq.IG = c(2,200),
                      tau.sq.IG   = c(3,300)),

        n.samples = 10000,
        n.report = 200,
        verbose=TRUE)
```

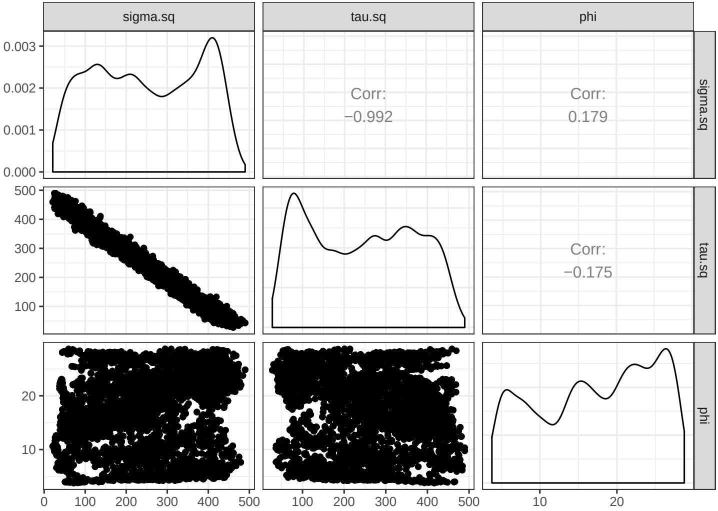
```
## ----------------------------------------
##  General model description
## ----------------------------------------
## Model fit with 1955 observations.
##
## Number of covariates 5 (including intercept if specified).
##
## Using the exponential spatial correlation model.
##
## Using modified predictive process with 36 knots.
##
## Number of MCMC samples 10000.
##
## Priors and hyperpriors:
##  beta normal:
##  mu: 0.000 0.000 0.000 0.000 0.000
##  cov:
##  1000.000 0.000 0.000 0.000 0.000
##  0.000 1000.000 0.000 0.000 0.000
##  0.000 0.000 1000.000 0.000 0.000
##  0.000 0.000 0.000 1000.000 0.000
##  0.000 0.000 0.000 0.000 1000.000
##
##  sigma.sq IG hyperpriors shape=2.00000 and scale=200.00000
##  tau.sq IG hyperpriors shape=3.00000 and scale=300.00000
##  phi Unif hyperpriors a=3.00000 and b=30.00000
## -------------------------------------------------
##  Sampling
## -------------------------------------------------
## Sampled: 200 of 10000, 2.00%
## Report interval Metrop. Acceptance rate: 36.50%
## Overall Metrop. Acceptance rate: 36.50%
## -------------------------------------------------
## Sampled: 400 of 10000, 4.00%
## Report interval Metrop. Acceptance rate: 36.50%
```

Traceplots

```
plot(r$p.theta.samples, density=FALSE)
```

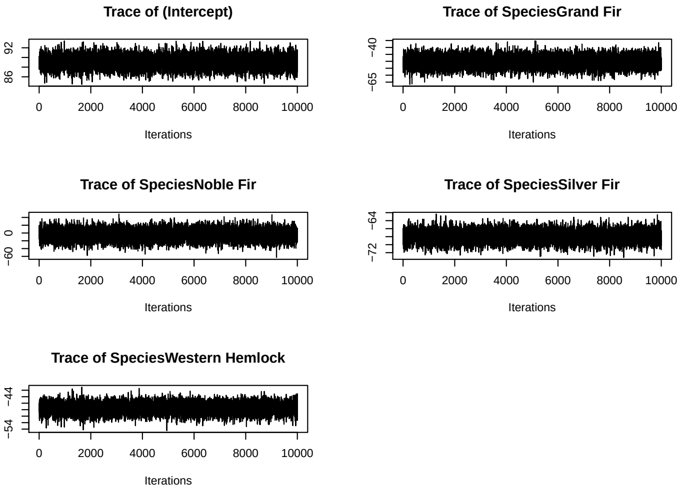


```
r$p.theta.samples[burnin:nreps,] %>%
  as.data.frame %>%
  GGally::ggpairs() +
  theme_bw()
```



Traceplot 2s

```
plot(r$p.beta.samples, density=FALSE)
```



Summary statistics

```r
summary(r$p.theta.samples)
```

```
##
## Iterations = 1:10000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 10000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##            Mean      SD Naive SE Time-series SE
## sigma.sq 246.59 127.324  1.27324         31.224
## tau.sq   244.94 126.725  1.26725         30.490
## phi       17.24   7.204  0.07204          2.743
##
## 2. Quantiles for each variable:
##
##            2.5%    25%    50%    75% 97.5%
## sigma.sq 43.200 133.07 238.13 366.61 446.7
## tau.sq   51.208 124.40 252.81 355.24 449.0
## phi       4.593  11.46  17.55  23.58  27.9
```

Summary statistics 2

```
summary(r$p.beta.samples)


##
## Iterations = 1:10000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 10000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##                            Mean     SD Naive SE Time-series SE
## (Intercept)              89.011  1.291  0.01291        0.01291
## SpeciesGrand Fir        -50.361  4.125  0.04125        0.04125
## SpeciesNoble Fir         -4.406 14.034  0.14034        0.14034
## SpeciesSilver Fir       -67.923  1.458  0.01458        0.01458
## SpeciesWestern Hemlock -47.605  1.631  0.01631        0.01631
##
## 2. Quantiles for each variable:
##
##                           2.5%    25%     50%     75%  97.5%
## (Intercept)              86.48  88.13  89.006  89.873  91.53
## SpeciesGrand Fir        -58.49 -53.11 -50.306 -47.617 -42.21
## SpeciesNoble Fir        -32.44 -13.91  -4.382   5.187  22.83
## SpeciesSilver Fir       -70.79 -68.90 -67.917 -66.944 -65.09
## SpeciesWestern Hemlock -50.79 -48.69 -47.615 -46.506 -44.44
```

Spatial surface

If interest resides in w, draws can be obtained using the following relationship

$$p(w|y) = \int p(w|\sigma^2, \phi, y) p(\sigma^2, \phi|y) d\sigma^2 d\phi$$

which suggests the following strategy:

Spatial surface

If interest resides in w, draws can be obtained using the following relationship

$$p(w|y) = \int p(w|\sigma^2, \phi, y) p(\sigma^2, \phi|y) d\sigma^2 d\phi$$

which suggests the following strategy:

1. Run the MCMC sampler to obtain draws $(\sigma^2, \phi)^{(g)} \sim p(\sigma^2, \phi|y)$
2. After burn-in and for $g = 1, \ldots, G$, sample $w^{(g)} \sim p(w|(\sigma^2, \phi)^{(g)}, y)$.

Prediction

For prediction at points $s_{01}, \ldots, s_{0m}$ and denoting $Y_0 = (Y(s_{01}), \ldots, Y(s_{0m}))^\top$ and design matrix X_0 having rows $x(s_{0j})^\top$,

Prediction

For prediction at points $s_{01}, \ldots, s_{0m}$ and denoting $Y_0 = (Y(s_{01}), \ldots, Y(s_{0m}))^\top$ and design matrix X_0 having rows $x(s_{0j})^\top$, we have the following relationship

$$p(y_0|y, X, X_0) = \int p(y_0|y, \theta, X_0)p(\theta|y, X)d\theta$$

Prediction

For prediction at points $s_{01}, \ldots, s_{0m}$ and denoting $Y_0 = (Y(s_{01}), \ldots, Y(s_{0m}))^\top$ and design matrix X_0 having rows $x(s_{0j})^\top$, we have the following relationship

$$p(y_0|y, X, X_0) = \int p(y_0|y, \theta, X_0)p(\theta|y, X)d\theta \approx \frac{1}{G}\sum_{g=1}^{G} p(y_0|y, \theta^{(g)}, X_0).$$

Prediction

For prediction at points $s_{01}, \ldots, s_{0m}$ and denoting $Y_0 = (Y(s_{01}), \ldots, Y(s_{0m}))^\top$ and design matrix X_0 having rows $x(s_{0j})^\top$, we have the following relationship

$$p(y_0|y, X, X_0) = \int p(y_0|y, \theta, X_0)p(\theta|y, X)d\theta \approx \frac{1}{G}\sum_{g=1}^{G} p(y_0|y, \theta^{(g)}, X_0).$$

It is more common to take draws $y_0^{(g)} \sim p(y_0|y, \theta^{(g)}, X_0)$ and estimate the predictive distribution using

$$p(y_0|y, X, X_0) \approx \frac{1}{G}\sum_{g=1}^{G} \delta_{y_0^{(g)}}$$

Prediction

For prediction at points $s_{01}, \ldots, s_{0m}$ and denoting $Y_0 = (Y(s_{01}), \ldots, Y(s_{0m}))^\top$ and design matrix X_0 having rows $x(s_{0j})^\top$, we have the following relationship

$$p(y_0|y, X, X_0) = \int p(y_0|y, \theta, X_0)p(\theta|y, X)d\theta \approx \frac{1}{G}\sum_{g=1}^{G} p(y_0|y, \theta^{(g)}, X_0).$$

It is more common to take draws $y_0^{(g)} \sim p(y_0|y, \theta^{(g)}, X_0)$ and estimate the predictive distribution using

$$p(y_0|y, X, X_0) \approx \frac{1}{G}\sum_{g=1}^{G} \delta_{y_0^{(g)}}$$

where $p(y_0|y, \theta, X_0)$ has a conditional normal distribution.

Predictions are not conditionally independent

Consider the joint distribution for y and $y_0 = y(s_0)$ (a scalar for simplicity),

Predictions are not conditionally independent

Consider the joint distribution for y and $y_0 = y(s_0)$ (a scalar for simplicity), then

$$\left(\begin{array}{c} y \\ y_0 \end{array}\right) \sim N\left(\left[\begin{array}{c} X\beta \\ X_0\beta \end{array}\right], \left[\begin{array}{cc} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{array}\right]\right)$$

Predictions are not conditionally independent

Consider the joint distribution for y and $y_0 = y(s_0)$ (a scalar for simplicity), then

$$\left(\begin{array}{c} y \\ y_0 \end{array} \right) \sim N\left(\left[\begin{array}{c} X\beta \\ X_0\beta \end{array} \right], \left[\begin{array}{cc} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{array} \right] \right)$$

where

$$\begin{array}{ll} \Omega_{11} & = \sigma^2 H(\phi) + \tau^2 \mathrm{I} \\ \Omega_{22} & = \sigma^2 + \tau^2 \\ \Omega_{12}^\top & = \sigma^2 [\rho(d_{01};\phi), \ldots, \rho(d_{0n};\phi)] \end{array}$$

and $d_{ij} = ||s_i - s_j||$.

Predictions are not conditionally independent

Consider the joint distribution for y and $y_0 = y(s_0)$ (a scalar for simplicity), then

$$\left(\begin{array}{c} y \\ y_0 \end{array} \right) \sim N\left(\left[\begin{array}{c} X\beta \\ X_0\beta \end{array} \right], \left[\begin{array}{cc} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{array} \right] \right)$$

where

$$\begin{array}{ll} \Omega_{11} & = \sigma^2 H(\phi) + \tau^2 \mathrm{I} \\ \Omega_{22} & = \sigma^2 + \tau^2 \\ \Omega_{12}^\top & = \sigma^2[\rho(d_{01};\phi),\ldots,\rho(d_{0n};\phi)] \end{array}$$

and $d_{ij} = ||s_i - s_j||$.

Thus $y_0|y,\theta,X,X_0$ is normal with

$$\begin{array}{rl} E[Y(s_0)|y,\theta,X,X_0] & = x_0^\top\beta + \Omega_{12}^\top\Omega_{22}^{-1}(y - X\beta) \\ Var[Y(s_0)|y,\theta,X,X_0] & = \sigma^2 + \tau^2 - \Omega_{12}^\top\Omega_{22}^{-1}\Omega_{12} \end{array}$$

Generalized linear spatial modeling

Let $Y(s)$ be the response of interest with

$$E[Y(s)] = g^{-1}(x(s)^\top \beta + w(s))$$

where $w(s)$ is our spatial random effect.

Generalized linear spatial modeling

Let $Y(s)$ be the response of interest with

$$E[Y(s)] = g^{-1}(x(s)^\top \beta + w(s))$$

where $w(s)$ is our spatial random effect.

For example, Poisson regression

$$Y(s) \sim Po(e^{x(s)^\top \beta + w(s)}).$$

Generalized linear spatial modeling

Let $Y(s)$ be the response of interest with

$$E[Y(s)] = g^{-1}(x(s)^\top \beta + w(s))$$

where $w(s)$ is our spatial random effect.

For example, Poisson regression

$$Y(s) \sim Po(e^{x(s)^\top \beta + w(s)}).$$

For GLMs (other than linear models), $w(s)$ cannot be integrated out

Generalized linear spatial modeling

Let $Y(s)$ be the response of interest with

$$E[Y(s)] = g^{-1}(x(s)^\top \beta + w(s))$$

where $w(s)$ is our spatial random effect.

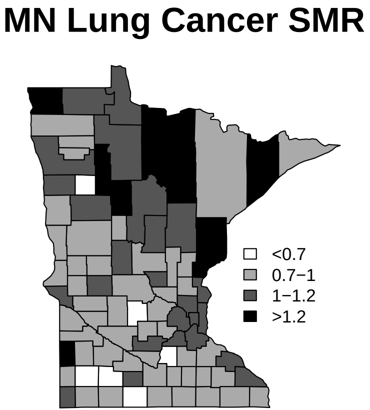
For example, Poisson regression

$$Y(s) \sim Po(e^{x(s)^\top \beta + w(s)}).$$

For GLMs (other than linear models), $w(s)$ cannot be integrated out and therefore a common MCMC strategy is

1. Sample $\beta| \ldots$.
2. Sample $w| \ldots$.
3. Sample $\theta| \ldots$ (the spatial parameters [no nugget]).

Choropleth



Modeling areal units

Let Y_i represent the SMR for lung cancer in MN county i.

Modeling areal units

Let Y_i represent the SMR for lung cancer in MN county i. Consider the model defined by conditional distributions:

$$Y_i | y_{-i} \sim N\left(\sum_{j \in n_i} y_j / m_i, \tau^2 / m_i\right)$$

where

- n_i indicates the neighbors of i

Modeling areal units

Let Y_i represent the SMR for lung cancer in MN county i. Consider the model defined by conditional distributions:

$$Y_i|y_{-i} \sim N\left(\sum_{j\in n_i} y_j/m_i, \tau^2/m_i\right)$$

where

- n_i indicates the neighbors of i
- m_i indicates the number of neighbors for i

Modeling areal units

Let Y_i represent the SMR for lung cancer in MN county i. Consider the model defined by conditional distributions:

$$Y_i | y_{-i} \sim N\left(\sum_{j \in n_i} y_j / m_i, \tau^2 / m_i \right)$$

where

- n_i indicates the neighbors of i
- m_i indicates the number of neighbors for i

Modeling areal units

Let Y_i represent the SMR for lung cancer in MN county i. Consider the model defined by conditional distributions:

$$Y_i | y_{-i} \sim N\left(\sum_{j \in n_i} y_j / m_i, \tau^2 / m_i\right)$$

where

- n_i indicates the neighbors of i
- m_i indicates the number of neighbors for i

This defines a *Markov Random Field*.

Brook's Lemma

It is clear that given $p(y_1, \ldots, y_n)$, the *full conditionals*, i.e. $p(y_i|y_{-i})$, are determined.

Brook's Lemma

It is clear that given $p(y_1,\ldots,y_n)$, the *full conditionals*, i.e. $p(y_i|y_{-i})$, are determined.

Definition

Brook's Lemma states that

$$\frac{p(y_1,\ldots,y_n)}{p(y_1',\ldots,y_n')} = \frac{p(y_1|y_2,\ldots,y_n)}{p(y_1'|y_2,\ldots,y_n)} \cdot \frac{p(y_2|y_1',y_3,\ldots,y_n)}{p(y_2'|y_1',y_3,\ldots,y_n)} \cdots \frac{p(y_n|y_1',\ldots,y_{n-1}')}{p(y_n'|y_1',\ldots,y_{n-1}')}$$

for all $(y_1',\ldots,y_n')$.

Brook's Lemma

It is clear that given $p(y_1,\ldots,y_n)$, the *full conditionals*, i.e. $p(y_i|y_{-i})$, are determined.

Definition

Brook's Lemma states that

$$\frac{p(y_1,\ldots,y_n)}{p(y_1',\ldots,y_n')} = \frac{p(y_1|y_2,\ldots,y_n)}{p(y_1'|y_2,\ldots,y_n)} \cdot \frac{p(y_2|y_1',y_3,\ldots,y_n)}{p(y_2'|y_1',y_3,\ldots,y_n)} \cdots \frac{p(y_n|y_1',\ldots,y_{n-1}')}{p(y_n'|y_1',\ldots,y_{n-1}')}$$

for all $(y_1',\ldots,y_n')$.

If

$$p(y_1',\ldots,y_n') = \int \frac{p(y_1|y_2,\ldots,y_n)}{p(y_1'|y_2,\ldots,y_n)} \cdot \frac{p(y_2|y_1',y_3,\ldots,y_n)}{p(y_2'|y_1',y_3,\ldots,y_n)} \cdots \frac{p(y_n|y_1',\ldots,y_{n-1}')}{p(y_n'|y_1',\ldots,y_{n-1}')} dy_1,\ldots,dy_n < \infty$$

then $p(y_1,\ldots,y_n)$ is a proper joint distribution.

Conditionally autoregressive models

More generally, we can consider

$$Y_i|y_{-i} \sim N\left(\sum_{j\neq i} b_{ij}y_j, \tau_i^2\right)$$

Through Brook's Lemma, we have

$$p(y_1,\ldots,y_n) \propto \exp\left(-\frac{1}{2}y^\top D^{-1}[\mathrm{I}-B]y\right)$$

where

- B has elements b_{ij}
- D is diagonal with elements τ_i^2

Conditionally autoregressive models

More generally, we can consider

$$Y_i|y_{-i} \sim N\left(\sum_{j\neq i} b_{ij}y_j, \tau_i^2\right)$$

Through Brook's Lemma, we have

$$p(y_1,\ldots,y_n) \propto \exp\left(-\frac{1}{2}y^\top D^{-1}[\mathrm{I}-B]y\right)$$

where

- B has elements b_{ij}
- D is diagonal with elements τ_i^2

In order for $D^{-1}[\mathrm{I}-B]$ to be symmetric, we need $\frac{b_{ij}}{\tau_i^2} = \frac{b_{ji}}{\tau_j^2}$ for all i,j.

Proximity matrix

Definition

A proximity matrix is a an $n \times n$ matrix, W, with elements

- $w_{ii} = 0$ and

Proximity matrix

Definition

A proximity matrix is a an $n \times n$ matrix, W, with elements

- $w_{ii} = 0$ and
- w_{ij} representing the "distance" between unit i and unit j

Proximity matrix

Definition

A proximity matrix is a an $n \times n$ matrix, W, with elements

- $w_{ii} = 0$ and
- w_{ij} representing the “distance” between unit i and unit j

Proximity matrix

Definition

A proximity matrix is a an $n \times n$ matrix, W, with elements

- $w_{ii} = 0$ and
- w_{ij} representing the "distance" between unit i and unit j

Common choices for w_{ij} are

- 1 if i is a neighbor of j and 0 otherwise

Proximity matrix

Definition

A proximity matrix is a an $n \times n$ matrix, W, with elements

- $w_{ii} = 0$ and
- w_{ij} representing the "distance" between unit i and unit j

Common choices for w_{ij} are

- 1 if i is a neighbor of j and 0 otherwise
    - neighbors defined by those who share an edge

Proximity matrix

Definition

A proximity matrix is a an $n \times n$ matrix, W, with elements

- $w_{ii} = 0$ and
- w_{ij} representing the “distance” between unit i and unit j

Common choices for w_{ij} are

- 1 if i is a neighbor of j and 0 otherwise
  - neighbors defined by those who share an edge
  - neighbors defined by those who share a point

Proximity matrix

Definition

A proximity matrix is a an $n \times n$ matrix, W, with elements

- $w_{ii} = 0$ and
- w_{ij} representing the "distance" between unit i and unit j

Common choices for w_{ij} are

- 1 if i is a neighbor of j and 0 otherwise
  - neighbors defined by those who share an edge
  - neighbors defined by those who share a point
  - neighbors defined by those who are within distance δ

Proximity matrix

Definition

A proximity matrix is a an $n \times n$ matrix, W, with elements

- $w_{ii} = 0$ and
- w_{ij} representing the "distance" between unit i and unit j

Common choices for w_{ij} are

- 1 if i is a neighbor of j and 0 otherwise
  - neighbors defined by those who share an edge
  - neighbors defined by those who share a point
  - neighbors defined by those who are within distance δ
  - K-nearest neighbors

Proximity matrix

Definition

A proximity matrix is a an $n \times n$ matrix, W, with elements

- $w_{ii} = 0$ and
- w_{ij} representing the "distance" between unit i and unit j

Common choices for w_{ij} are

- 1 if i is a neighbor of j and 0 otherwise
  - neighbors defined by those who share an edge
  - neighbors defined by those who share a point
  - neighbors defined by those who are within distance δ
  - K-nearest neighbors
- "distance"

Proximity matrix

Definition

A proximity matrix is a an $n \times n$ matrix, W, with elements

- $w_{ii} = 0$ and
- w_{ij} representing the "distance" between unit i and unit j

Common choices for w_{ij} are

- 1 if i is a neighbor of j and 0 otherwise
  - neighbors defined by those who share an edge
  - neighbors defined by those who share a point
  - neighbors defined by those who are within distance δ
  - K-nearest neighbors
- "distance"
  - inverse intercentroidal distance

Proximity matrix

Definition

A proximity matrix is a an $n \times n$ matrix, W, with elements

- $w_{ii} = 0$ and
- w_{ij} representing the "distance" between unit i and unit j

Common choices for w_{ij} are

- 1 if i is a neighbor of j and 0 otherwise
  - neighbors defined by those who share an edge
  - neighbors defined by those who share a point
  - neighbors defined by those who are within distance δ
  - K-nearest neighbors
- "distance"
  - inverse intercentroidal distance
  - inverse minimum distance plus c

Intrinsicially autoregressive model

Recall

$$p(y_1, \ldots, y_n) \propto \exp\left(-\frac{1}{2} y^\top D^{-1}[\mathrm{I} - B]y\right)$$

Intrinsicially autoregressive model

Recall

$$p(y_1, \ldots, y_n) \propto \exp\left(-\frac{1}{2}y^\top D^{-1}[\mathrm{I} - B]y\right)$$

if we set $w_{i+} = \sum_{j=1}^{n} w_{ij}$, $b_{ij} = w_{ij}/w_{i+}$, and $\tau_i^2 = \tau^2/w_{i+}$,

Intrinsicially autoregressive model

Recall

$$p(y_1, \ldots, y_n) \propto \exp\left(-\frac{1}{2} y^\top D^{-1} [\mathrm{I} - B] y\right)$$

if we set $w_{i+} = \sum_{j=1}^n w_{ij}$, $b_{ij} = w_{ij}/w_{i+}$, and $\tau_i^2 = \tau^2/w_{i+}$, we have

$$p(y_1, \ldots, y_n) \propto \exp\left(-\frac{1}{2\tau^2} y^\top [D_w - W] y\right)$$

Intrinsicially autoregressive model

Recall

$$p(y_1, \ldots, y_n) \propto \exp\left(-\frac{1}{2} y^\top D^{-1}[\mathrm{I} - B]y\right)$$

if we set $w_{i+} = \sum_{j=1}^n w_{ij}$, $b_{ij} = w_{ij}/w_{i+}$, and $\tau_i^2 = \tau^2/w_{i+}$, we have

$$p(y_1, \ldots, y_n) \propto \exp\left(-\frac{1}{2\tau^2} y^\top [D_w - W]y\right)$$

where

- W is our proximity matrix and

Intrinsicially autoregressive model

Recall

$$p(y_1,\ldots,y_n) \propto \exp\left(-\frac{1}{2}y^\top D^{-1}[\mathrm{I}-B]y\right)$$

if we set $w_{i+} = \sum_{j=1}^n w_{ij}$, $b_{ij} = w_{ij}/w_{i+}$, and $\tau_i^2 = \tau^2/w_{i+}$, we have

$$p(y_1,\ldots,y_n) \propto \exp\left(-\frac{1}{2\tau^2}y^\top [D_w - W]y\right)$$

where

- W is our proximity matrix and
- D_w has diagonal elements w_{i+}

Intrinsicially autoregressive model

Recall

$$p(y_1,\ldots,y_n) \propto \exp\left(-\frac{1}{2}y^\top D^{-1}[\mathrm{I}-B]y\right)$$

if we set $w_{i+}=\sum_{j=1}^n w_{ij}$, $b_{ij}=w_{ij}/w_{i+}$, and $\tau_i^2=\tau^2/w_{i+}$, we have

$$p(y_1,\ldots,y_n) \propto \exp\left(-\frac{1}{2\tau^2}y^\top [D_w-W]y\right)$$

where

- W is our proximity matrix and
- D_w has diagonal elements w_{i+}

Intrinsicially autoregressive model

Recall

$$p(y_1,\ldots,y_n) \propto \exp\left(-\frac{1}{2}y^\top D^{-1}[\mathrm{I}-B]y\right)$$

if we set $w_{i+} = \sum_{j=1}^n w_{ij}$, $b_{ij} = w_{ij}/w_{i+}$, and $\tau_i^2 = \tau^2/w_{i+}$, we have

$$p(y_1,\ldots,y_n) \propto \exp\left(-\frac{1}{2\tau^2}y^\top [D_w - W]y\right)$$

where

- W is our proximity matrix and
- D_w has diagonal elements w_{i+}

This can be rewritten as

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

Intrinsicially autoregressive model

Recall

$$p(y_1,\ldots,y_n) \propto \exp\left(-\frac{1}{2}y^\top D^{-1}[\mathrm{I}-B]y\right)$$

if we set $w_{i+} = \sum_{j=1}^n w_{ij}$, $b_{ij} = w_{ij}/w_{i+}$, and $\tau_i^2 = \tau^2/w_{i+}$, we have

$$p(y_1,\ldots,y_n) \propto \exp\left(-\frac{1}{2\tau^2}y^\top [D_w - W]y\right)$$

where

- W is our proximity matrix and
- D_w has diagonal elements w_{i+}

This can be rewritten as

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

This is called the *intrinsically autoregressive* model.

Proper CAR models

To make this proper,

$$p(y_1, \ldots, y_n) \propto \exp\left(-\frac{1}{2\tau^2} y^\top [D_w - \rho W] y\right)$$

with

- $\rho \in (1/\lambda_{(n)}, 1/\lambda_{(1)})$ where

Proper CAR models

To make this proper,

$$p(y_1, \ldots, y_n) \propto \exp\left(-\frac{1}{2\tau^2} y^\top [D_w - \rho W] y\right)$$

with

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

Proper CAR models

To make this proper,

$$p(y_1, \ldots, y_n) \propto \exp\left(-\frac{1}{2\tau^2} y^\top [D_w - \rho W] y\right)$$

with

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

Proper CAR models

To make this proper,

$$p(y_1, \ldots, y_n) \propto \exp\left(-\frac{1}{2\tau^2} y^\top [D_w - \rho W] y\right)$$

with

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

The full conditionals are

$$Y_i | y_{-i} \sim N\left(\rho \sum_{j \neq i} w_{ij} y_j / w_{i+}, \tau^2 / w_{i+}\right)$$

Proper CAR models

To make this proper,

$$p(y_1, \ldots, y_n) \propto \exp\left(-\frac{1}{2\tau^2} y^\top [D_w - \rho W] y\right)$$

with

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

The full conditionals are

$$Y_i | y_{-i} \sim N\left(\rho \sum_{j \neq i} w_{ij} y_j / w_{i+}, \tau^2 / w_{i+}\right)$$

a prior for ρ that induces a reasonable amount of spatial association should put most of its mass near 1.

Issues with the proper CAR

The full condition for the proper CAR, i.e.

$$Y_i|y_{-i} \sim N\left(\rho \sum_{j\neq i} w_{ij}y_j/w_{i+}, \tau^2/w_{i+}\right)$$

indicates some issues with this model:

Issues with the proper CAR

The full condition for the proper CAR, i.e.

$$Y_i|y_{-i} \sim N\left(\rho \sum_{j\neq i} w_{ij}y_j/w_{i+}, \tau^2/w_{i+}\right)$$

indicates some issues with this model:

- τ^2 does not play a role in spatial association.

Issues with the proper CAR

The full condition for the proper CAR, i.e.

$$Y_i|y_{-i} \sim N\left(\rho \sum_{j\neq i} w_{ij}y_j/w_{i+}, \tau^2/w_{i+}\right)$$

indicates some issues with this model:

- τ^2 does not play a role in spatial association.
- ρ is a proportional reaction to the weighted average of its neighbors.

Issues with the proper CAR

The full condition for the proper CAR, i.e.

$$Y_i|y_{-i} \sim N\left(\rho \sum_{j\neq i} w_{ij} y_j / w_{i+}, \tau^2 / w_{i+}\right)$$

indicates some issues with this model:

- τ^2 does not play a role in spatial association.
- ρ is a proportional reaction to the weighted average of its neighbors.
- If $\rho < 1$, then the expected value of the current location is less than the weighted average of its neighbors.

Issues with the proper CAR

The full condition for the proper CAR, i.e.

$$Y_i|y_{-i} \sim N\left(\rho \sum_{j\neq i} w_{ij}y_j/w_{i+}, \tau^2/w_{i+}\right)$$

indicates some issues with this model:

- τ^2 does not play a role in spatial association.
- ρ is a proportional reaction to the weighted average of its neighbors.
- If $\rho < 1$, then the expected value of the current location is less than the weighted average of its neighbors.
- If $\rho = 0$, then we have conditional independence.

Issues with the proper CAR

The full condition for the proper CAR, i.e.

$$Y_i|y_{-i} \sim N\left(\rho \sum_{j\neq i} w_{ij}y_j/w_{i+}, \tau^2/w_{i+}\right)$$

indicates some issues with this model:

- τ^2 does not play a role in spatial association.
- ρ is a proportional reaction to the weighted average of its neighbors.
- If $\rho < 1$, then the expected value of the current location is less than the weighted average of its neighbors.
- If $\rho = 0$, then we have conditional independence. But the variance decreases with the number of neighbors which is perplexing.

Issues with the proper CAR

The full condition for the proper CAR, i.e.

$$Y_i|y_{-i} \sim N\left(\rho \sum_{j \neq i} w_{ij} y_j / w_{i+}, \tau^2 / w_{i+}\right)$$

indicates some issues with this model:

- τ^2 does not play a role in spatial association.
- ρ is a proportional reaction to the weighted average of its neighbors.
- If $\rho < 1$, then the expected value of the current location is less than the weighted average of its neighbors.
- If $\rho = 0$, then we have conditional independence. But the variance decreases with the number of neighbors which is perplexing.
- ρ needs to be very close to 1 to obtain a consequential amount of spatial association.

Dealing with ρ in the proper CAR

- Choose ρ so the CAR model is proper

Dealing with ρ in the proper CAR

- Choose ρ so the CAR model is proper
- Choose $\rho = 1$ (improper IAR model) and constrain $\sum_{i=1}^{n} Y_i = 0$

Dealing with ρ in the proper CAR

- Choose ρ so the CAR model is proper
- Choose $\rho = 1$ (improper IAR model) and constrain $\sum_{i=1}^{n} Y_i = 0$
- Choose $\rho = 1$ and estimate a mean (remove mean from the fixed effect)

Dealing with ρ in the proper CAR

- Choose ρ so the CAR model is proper
- Choose $\rho = 1$ (improper IAR model) and constrain $\sum_{i=1}^{n} Y_i = 0$
- Choose $\rho = 1$ and estimate a mean (remove mean from the fixed effect)
- Let $\rho \sim Be(18, 2)$ (Banerjee pg 164) and estimate it.

Leroux CAR

The Leroux et al. (1999) CAR tries to ameliorate these issues.

Leroux CAR

The Leroux et al. (1999) CAR tries to ameliorate these issues. The joint distribution is

$$Y \sim N(0, \tau^2[\rho(D_w - W) + (1 - rho)\mathrm{I}]^{-1})$$

Leroux CAR

The Leroux et al. (1999) CAR tries to ameliorate these issues. The joint distribution is

$$Y \sim N(0, \tau^2[\rho(D_w - W) + (1 - rho)\mathrm{I}]^{-1})$$

and the conditional distributions are

$$Y_i|y_{-i} \sim N\left(\frac{\rho\sum_{j\neq i} w_{ij}y_j}{rho\sum_{j\neq i} w_{ij}y_j + 1 - \rho}, \frac{\tau^2}{rho\sum_{j\neq i} w_{ij}y_j + 1 - \rho}\right).$$

Leroux CAR

The Leroux et al. (1999) CAR tries to ameliorate these issues. The joint distribution is

$$Y \sim N(0, \tau^2[\rho(D_w - W) + (1 - rho)\mathrm{I}]^{-1})$$

and the conditional distributions are

$$Y_i|y_{-i} \sim N\left(\frac{\rho \sum_{j\neq i} w_{ij}y_j}{rho \sum_{j\neq i} w_{ij}y_j + 1 - \rho}, \frac{\tau^2}{rho \sum_{j\neq i} w_{ij}y_j + 1 - \rho}\right).$$

This distribution is proper so long as $0 \le \rho < 1$.

Leroux CAR

The Leroux et al. (1999) CAR tries to ameliorate these issues. The joint distribution is

$$Y \sim N(0, \tau^2[\rho(D_w - W) + (1 - rho)\mathrm{I}]^{-1})$$

and the conditional distributions are

$$Y_i|y_{-i} \sim N\left(\frac{\rho\sum_{j\neq i} w_{ij}y_j}{rho\sum_{j\neq i} w_{ij}y_j + 1 - \rho}, \frac{\tau^2}{rho\sum_{j\neq i} w_{ij}y_j + 1 - \rho}\right).$$

This distribution is proper so long as $0 \le \rho < 1$. Lee (2011) argued that this CAR should be preferred for a variety of reasons.

CAR as a model for random effects

Let

- Y_i represent the (continuous) response for observation i

CAR as a model for random effects

Let

- Y_i represent the (continuous) response for observation i
- X_i represent explanatory variables for observation i

CAR as a model for random effects

Let

- Y_i represent the (continuous) response for observation i
- X_i represent explanatory variables for observation i
- $s[i]$ represent the areal unit for observation i

CAR as a model for random effects

Let

- Y_i represent the (continuous) response for observation i
- X_i represent explanatory variables for observation i
- $s[i]$ represent the areal unit for observation i

CAR as a model for random effects

Let

- Y_i represent the (continuous) response for observation i
- X_i represent explanatory variables for observation i
- $s[i]$ represent the areal unit for observation i

then a possible model is

$$Y_i = X_i^\top \beta + \omega_{s[i]} + \epsilon_i$$

CAR as a model for random effects

Let

- Y_i represent the (continuous) response for observation i
- X_i represent explanatory variables for observation i
- $s[i]$ represent the areal unit for observation i

then a possible model is

$$Y_i = X_i^\top \beta + \omega_{s[i]} + \epsilon_i$$

where

- $\epsilon_i \stackrel{ind}{\sim} N(0, \sigma^2)$ is noise

CAR as a model for random effects

Let

- Y_i represent the (continuous) response for observation i
- X_i represent explanatory variables for observation i
- $s[i]$ represent the areal unit for observation i

then a possible model is

$$Y_i = X_i^\top \beta + \omega_{s[i]} + \epsilon_i$$

where

- $\epsilon_i \stackrel{ind}{\sim} N(0, \sigma^2)$ is noise
- ω_s is the spatial random effect associated with areal unit s, e.g.

CAR as a model for random effects

Let

- Y_i represent the (continuous) response for observation i
- X_i represent explanatory variables for observation i
- $s[i]$ represent the areal unit for observation i

then a possible model is

$$Y_i = X_i^\top \beta + \omega_{s[i]} + \epsilon_i$$

where

- $\epsilon_i \stackrel{ind}{\sim} N(0, \sigma^2)$ is noise
- ω_s is the spatial random effect associated with areal unit s, e.g.

CAR as a model for random effects

Let

- Y_i represent the (continuous) response for observation i
- X_i represent explanatory variables for observation i
- $s[i]$ represent the areal unit for observation i

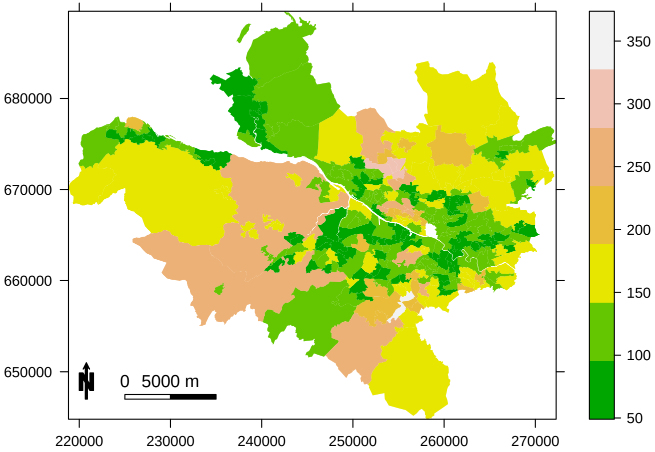
then a possible model is

$$Y_i = X_i^\top \beta + \omega_{s[i]} + \epsilon_i$$

where

- $\epsilon_i \stackrel{ind}{\sim} N(0, \sigma^2)$ is noise
- ω_s is the spatial random effect associated with areal unit s, e.g.

$$p(\omega_1, \ldots, \omega_S) \propto \exp\left(-\frac{1}{2\tau^2}\omega^\top [D_w - \rho W]\omega\right)$$


680000
670000
660000
650000
220000
230000
240000
250000
260000
270000
0 5000 m
N
350
300
250
200
150
100
50

Housing price model

Let

- Y_i be the logarithm of the median home price in each Intermediate Geography (IG) to the north of the river Clude in the Greater Glasgow and Clyde health board,

Housing price model

Let

- Y_i be the logarithm of the median home price in each Intermediate Geography (IG) to the north of the river Clude in the Greater Glasgow and Clyde health board,
- use explanatory variables
  - crime: crime rate (number of crimes per 10,000 people) in each IG (logged),
  - rooms: median number of rooms in a property in each IG,
  - type: predominant property type in each IG with levels: detached, flat, semi, terrace,
  - sales: percentage of properties that sold in each IG in a year, and
  - driveshop: average time taken to drive to a shopping centre in minutes (logged).

Housing price model

Assume

$$Y_i \stackrel{ind}{\sim} N(X_i\beta + \omega_i, \nu^2)$$

Housing price model

Assume

$$Y_i \stackrel{ind}{\sim} N(X_i\beta + \omega_i, \nu^2)$$

or, alternatively,

$$Y_i = X_\beta + \omega_i + \epsilon_i, \quad \epsilon_i \stackrel{ind}{\sim} N(0, \nu^2)$$

Housing price model

Assume

$$Y_i \stackrel{ind}{\sim} N(X_i\beta + \omega_i, \nu^2)$$

or, alternatively,

$$Y_i = X_\beta + \omega_i + \epsilon_i, \quad \epsilon_i \stackrel{ind}{\sim} N(0, \nu^2)$$

where

- β are the regression parameters and

Housing price model

Assume

$$Y_i \stackrel{ind}{\sim} N(X_i\beta + \omega_i, \nu^2)$$

or, alternatively,

$$Y_i = X_\beta + \omega_i + \epsilon_i, \quad \epsilon_i \stackrel{ind}{\sim} N(0, \nu^2)$$

where

- β are the regression parameters and
- ω_i are assumed to come from an intrinsic CAR model with proximity matrix indicating those regions that share a border

```
propertydata.spatial@data$logprice <- log(propertydata.spatial@data$price)
propertydata.spatial@data$logdriveshop <- log(propertydata.spatial@data$driveshop)


###################################################
### code chunk number 9: CARBayes.Rnw:495-498
###################################################
library(splines)
form <- logprice~ns(crime,3)+rooms+sales+factor(type) + logdriveshop
model <- lm(formula=form, data=propertydata.spatial@data)


###################################################
### code chunk number 10: CARBayes.Rnw:505-510
###################################################
library(spdep)
W.nb <- poly2nb(propertydata.spatial, row.names = rownames(propertydata.spatial@data))
W.list <- nb2listw(W.nb, style="B")
resid.model <- residuals(model)
moran.mc(x=resid.model, listw=W.list, nsim=1000)


##
##  Monte-Carlo simulation of Moran I
##
## data:  resid.model
## weights: W.list
## number of simulations + 1: 1001
##
## statistic = 0.2733, observed rank = 1001, p-value = 0.000999
## alternative hypothesis: greater
```

```
##
## Call:
## lm(formula = form, data = propertydata.spatial@data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.91319 -0.15992  0.00136  0.15647  0.81675
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          4.436135   0.157971  28.082  < 2e-16 ***
## ns(crime, 3)1       -0.358967   0.089006  -4.033 7.24e-05 ***
## ns(crime, 3)2       -0.617084   0.165152  -3.736 0.000229 ***
## ns(crime, 3)3       -0.299454   0.126516  -2.367 0.018670 *
## rooms                0.193827   0.029268   6.623 2.02e-10 ***
## sales                0.002034   0.000362   5.619 4.93e-08 ***
## factor(type)flat    -0.215967   0.066412  -3.252 0.001298 **
## factor(type)semi    -0.153610   0.057750  -2.660 0.008301 **
## factor(type)terrace -0.280023   0.072634  -3.855 0.000146 ***
## logdriveshop        -0.089084   0.025588  -3.482 0.000585 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2243 on 260 degrees of freedom
## Multiple R-squared:  0.6206,Adjusted R-squared:  0.6075
## F-statistic: 47.26 on 9 and 260 DF,  p-value: < 2.2e-16
```

Bayesian analysis using Leroux CAR

```
W <- nb2mat(W.nb, style = "B")
model.spatial <- S.CARleroux(formula = form,
                             data = propertydata.spatial@data,
                             family = "gaussian",
                             W = W,
                             burnin = 20000,
                             n.sample = 120000, thin = 10)

## Setting up the model

## Error:  the formula inputted contains an error, e.g the variables may be different lengths.
```