Lecture 5: Intrinsic GMRFs

Gaussian Markov random fields

David Bolin
Chalmers University of Technology
February 2, 2015

Intrinsic GMRFs (IGMRF)

Intrinsic GMRFs are

- improper, i.e., they have precision matrices not of full rank.
- Often used as prior distributions in various applications.

Of particular importance are IGMRFs that are invariant to any trend that is a polynomial of the locations of the nodes up to a specific order.

Today we will define IGMRFs and look at some popular constructions:

- Random walks models on the line
- ICAR models on regular lattices
- Besag models on irregular graphs

Improper GMRF

Definition

Let $\mathbf{Q}$ be an $n \times n$ SPSD matrix with rank $n - k > 0$. Then $\mathbf{x} = (x_1, \ldots, x_n)^T$ is an improper GMRF of rank $n - k$ with parameters $(\boldsymbol{\mu}, \mathbf{Q})$, if its density is

$$\pi(\mathbf{x}) = (2\pi)^{\frac{-(n-k)}{2}} \left(|\mathbf{Q}|^*\right)^{1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{Q} (\mathbf{x} - \boldsymbol{\mu})\right). \quad (1)$$

Further, $\mathbf{x}$ is an improper GMRF wrt to the labelled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where

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

$|\cdot|^*$ denote the generalised determinant: product of the non-zero eigenvalues.

Comments

- We will continue denoting $(\boldsymbol{\mu}, \mathbf{Q})$ "mean" and "precision", even if these quantities formally do not exist.
- The precision matrix has at least one zero eigenvalue.
- An improper GMRF is always proper on an appropriate subspace.
- We *define* sampling from an IGMRF as sampling from the proper density on the appropriate subspace.
- The eigenvectors corresponding to the zero eigenvalues define directions that the GMRF 'does not care about'
- This is *the* important property: We can use this indifference!
- If we use an improper GMRF as a prior, the *posterior* is typically proper.

Example: GMRFs under linear constraints

Consider a GMRF $\mathbf{x}$ under the hard contraint $\mathbf{A}\mathbf{x} = \mathbf{a}$.

We can *define* the density $\pi^*(\mathbf{x}) = \pi(\mathbf{x}|\mathbf{A}\mathbf{x} = \mathbf{a})$ for any $\mathbf{x} \in \mathbb{R}^n$.

The null space of the precision of $\pi^*(\mathbf{x})$ is spanned by the columns of $\mathbf{A}^T$.

Take $\mathbf{x} = \tilde{\mathbf{x}} + \hat{\mathbf{x}}$, where $\tilde{\mathbf{x}}$ is in the null space of the precision matrix $\hat{\mathbf{x}}$ is in the orthogonal subspace. We then have

$$\pi^*(\mathbf{x}) = \pi^*(\hat{\mathbf{x}}) = \pi(\mathbf{x}|\mathbf{A}\mathbf{x} = \mathbf{a})$$

Thus, π^* is invariant to the addition of any $\tilde{\mathbf{x}}$ from the space spanned by the rows of $\mathbf{A}$.

Markov properties

The Markov properties are interpreted as those obtained from the limit of a proper density.

Let the columns of $\mathbf{A}^T$ span the null space of $\mathbf{Q}$

$$\mathbf{Q}(\gamma) = \mathbf{Q} + \gamma \mathbf{A}^T \mathbf{A}.$$

Each element in $\mathbf{Q}(\gamma)$ tends to the corresponding one in $\mathbf{Q}$ as $\gamma \to 0$.

$$\mathsf{E}(x_i \mid \mathbf{x}_{-i}) = \mu_i - \frac{1}{Q_{ii}} \sum_{j \sim i} Q_{ij}(x_j - \mu_j)$$

is interpreted as $\gamma \to 0$.

Polynomial IGMRFs

The order of the IGMRF is the rank deficiency of the precision matrix.

Polynomial IGMRFs

We use the term *polynomial IGMRF* of order k if the field is invariant to the addition of a polynomial of order $< k$.

An intrinsic GMRF of first order is an improper GMRF of rank $n-1$ *where* $\mathbf{Q1} = \mathbf{0}$.

The condition $\mathbf{Q1} = \mathbf{0}$ means that $\sum_j Q_{ij} = 0, \quad i = 1, \ldots, n$

Local behaviour

Consider a first order IGMRF. With $\boldsymbol{\mu} = \mathbf{0}$, we have

$$\mathsf{E}(x_i \mid \mathbf{x}_{-i}) = -\frac{1}{Q_{ii}} \sum_{j:j\sim i} Q_{ij} x_j$$

and

$$-\sum_{j:j\sim i} Q_{ij}/Q_{ii} = 1$$

- The conditional mean of x_i is a weighted mean of its neighbours.
- No shrinking towards an overall level.This is important!
- We can concentrate on the deviation from *any* overall level without having to specify what it is
- This is a perfect property for a prior
- Many IGMRFs are constructed such that the *deviation* from the overall level is a smooth curve in time or a smooth surface in space.

Building IGMRFs on the line

Definition (Forward difference)

Define the first-order forward difference of a function $f(\cdot)$ as

$$\Delta f(z) = f(z+1) - f(z).$$

Higher-order forward differences are defined recursively:

$$\Delta^k f(z) = \Delta \Delta^{k-1} f(z)$$

so

$$\Delta^2 f(z) = f(z+2) - 2f(z+1) + f(z) \tag{2}$$

and in general for $k = 1, 2, \ldots,$

$$\Delta^k f(z) = (-1)^k \sum_{j=0}^{k} (-1)^j \binom{k}{j} f(z+j).$$

IGMRFs of first order on the line

Location of the node i is i (think "time").

Assume *independent increments*

$$\Delta x_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \kappa^{-1}), \quad i = 1, \ldots, n-1, \qquad \text{so} \tag{3}$$

$$x_j - x_i \sim \mathcal{N}(0, (j-i)\kappa^{-1}) \text{ for } i < j. \tag{4}$$

The density for $\mathbf{x}$ is

$$\pi(\mathbf{x} \mid \kappa) \propto \kappa^{(n-1)/2} \exp\left(-\frac{\kappa}{2}\sum_{i=1}^{n-1}(\Delta x_i)^2\right) \tag{5}$$

$$= \kappa^{(n-1)/2} \exp\left(-\frac{\kappa}{2}\sum_{i=1}^{n-1}(x_{i+1} - x_i)^2\right), \tag{6}$$

or

$$\pi(\mathbf{x}) \propto \kappa^{(n-1)/2} \exp\left(-\frac{\kappa}{2}\mathbf{x}^T\mathbf{R}\mathbf{x}\right) \tag{7}$$

The structure matrix

The *structure matrix*

$$\mathbf{R} = \begin{pmatrix} 1 & -1 & & & & & \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & -1 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & -1 & 2 & -1 & \\ & & & & -1 & 2 & -1 \\ & & & & & -1 & 1 \end{pmatrix}. \tag{8}$$

- The $n-1$ independent increments ensure that the rank of $\mathbf{Q}$ is $n-1$.
- Denote this model by RW1(κ) or short RW1.

Properties

Full conditionals

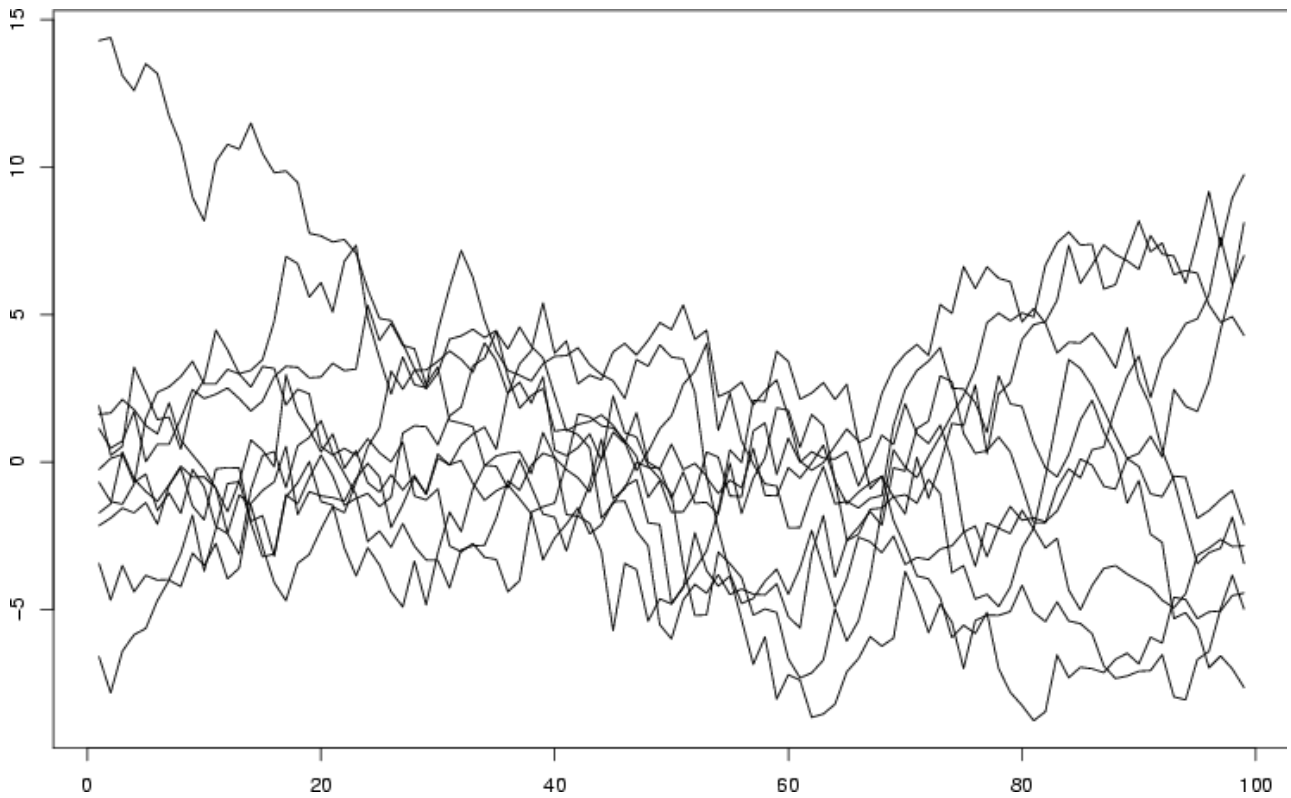
$$x_i \mid \mathbf{x}_{-i}, \kappa \sim \mathcal{N}(\frac{1}{2}(x_{i-1} + x_{i+1}),\ 1/(2\kappa)), \quad 1 < i < n, \tag{9}$$

If the intersection between $\{i, \ldots, j\}$ and $\{k, \ldots, l\}$ is empty for $i < j$ and $k < l$, then

$$\mathrm{Cov}(x_j - x_i, x_l - x_k) = 0. \tag{10}$$

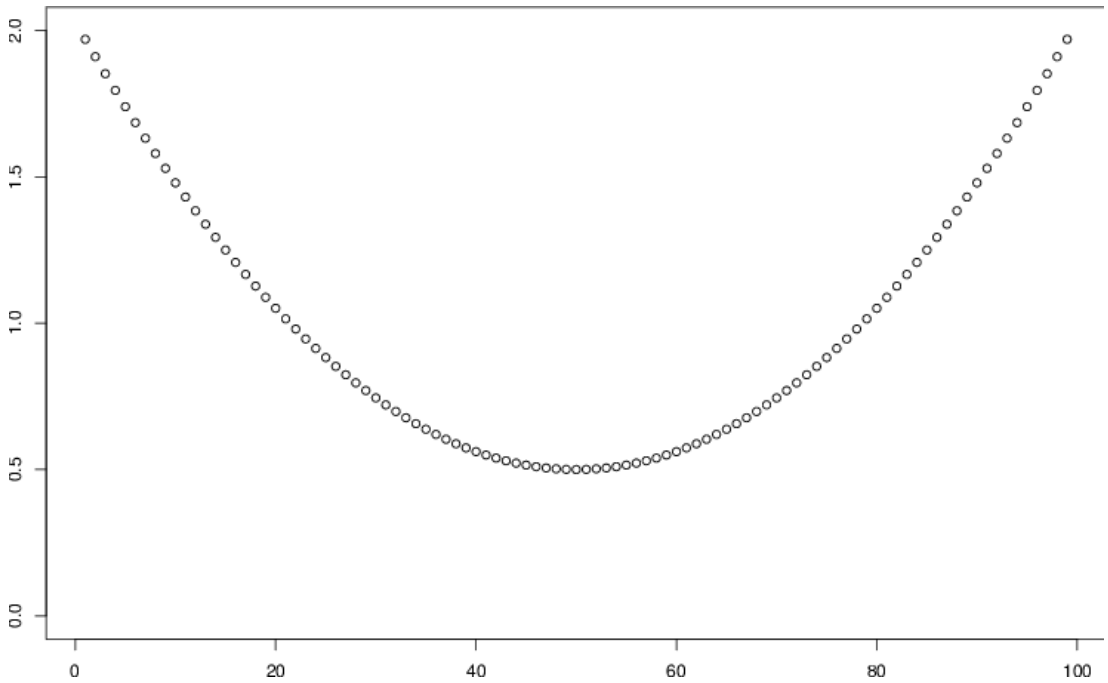
Properties coincide with those of a *Wiener process*.

Example



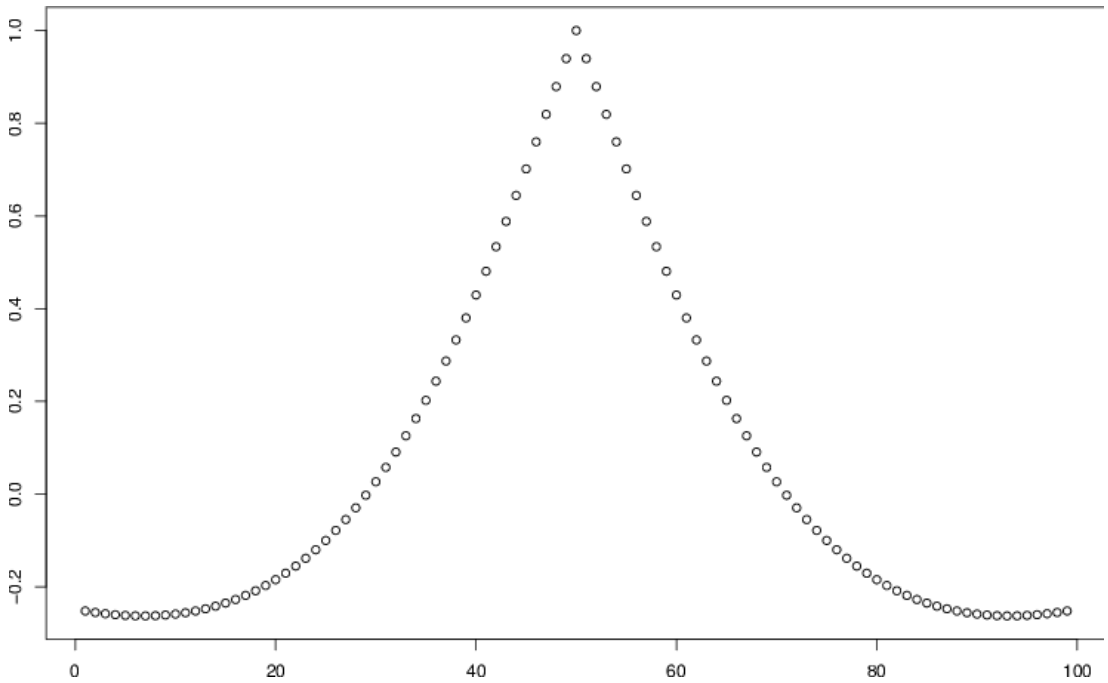
Samples with $n = 99$ and $\kappa = 1$ by conditioning on the constraint $\sum x_i = 0$.

Example



Marginal variances $\mathrm{Var}(x_i)$ for $i = 1, \ldots, n$.

Example



$\text{Corr}(x_{n/2}, x_i)$ for $i = 1, \ldots, n$

An alternative first order IGMRF

Let

$$\pi(\mathbf{x}) \propto \kappa^{(n-1)/2} \exp\left(-\frac{\kappa}{2}\sum_{i=1}^{n}(x_i - \overline{\mathbf{x}})^2\right) \tag{11}$$

where $\overline{\mathbf{x}}$ is the empirical mean of $\mathbf{x}$.
Assume n is even and

$$x_i = \begin{cases} 0, & 1 \leq i \leq n/2 \\ 1, & n/2 < i \leq n \end{cases}. \tag{12}$$

Thus $\mathbf{x}$ is locally constant with two levels.

Evaluate the density at this configuration under the RW1 model and the alternative (11), we obtain

$$\kappa^{(n-1)/2} \exp\left(-\frac{\kappa}{2}\right) \quad \text{and} \quad \kappa^{(n-1)/2} \exp\left(-n\frac{\kappa}{8}\right) \tag{13}$$

The log ratio of the densities is then of order $\mathcal{O}(n)$.

The local structure of the RW1 model

- The RW1 model only penalises the *local* deviation from a constant level.
- The alternative penalises the *global* deviation from a constant level.

This *local* behaviour is advantageous in applications if the mean level of $\mathbf{x}$ is approximately or locally constant.

A similar argument also apply for polynomial IGMRFs of higher order constructed using forward differences of order k as independent Gaussian increments.

The RW2 model for regular locations

Let $s_i = i$ for $i = 1, \ldots, n$, with a constant distance between consecutive nodes.

Use the second order increments

$$\Delta^2 x_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \kappa^{-1}) \tag{14}$$

for $i = 1, \ldots, n-2$, to define the joint density of $\mathbf{x}$

$$\begin{aligned} \pi(\mathbf{x}) &\propto \kappa^{(n-2)/2} \exp\left(-\frac{\kappa}{2}\sum_{i=1}^{n-2}(x_i - 2x_{i+1} + x_{i+2})^2\right) && (15) \\ &= \kappa^{(n-2)/2} \exp\left(-\frac{\kappa}{2}\mathbf{x}^T\mathbf{R}\mathbf{x}\right) && (16) \end{aligned}$$

The structure matrix

The structure matrix is now

$$\mathbf{R} = \begin{pmatrix} 1 & -2 & 1 & & & & & & \\ -2 & 5 & -4 & 1 & & & & & \\ 1 & -4 & 6 & -4 & 1 & & & & \\ & 1 & -4 & 6 & -4 & 1 & & & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & & \\ & & & 1 & -4 & 6 & -4 & 1 & \\ & & & & 1 & -4 & 6 & -4 & 1 \\ & & & & & 1 & -4 & 5 & -2 \\ & & & & & & 1 & -2 & 1 \end{pmatrix}.$$

Remarks

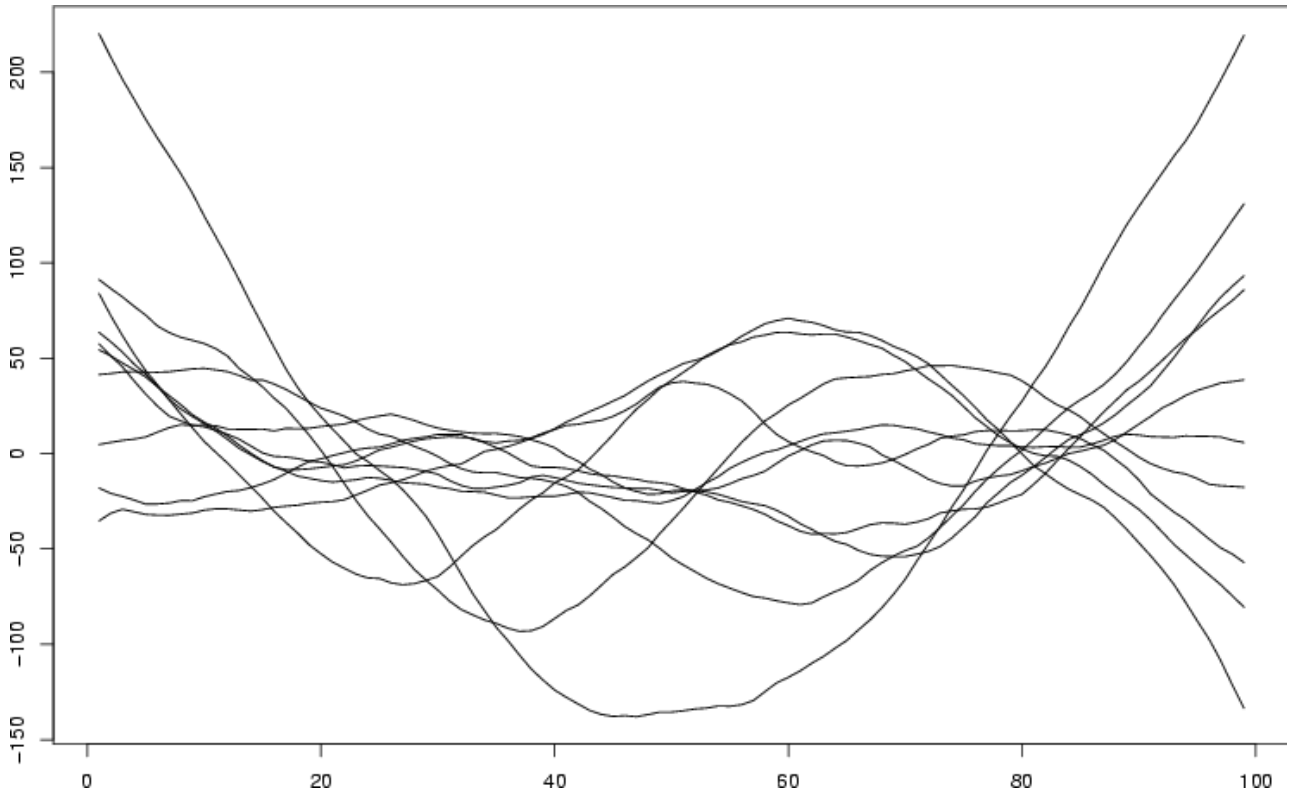
The conditional mean and precision is

$$\begin{aligned} \mathsf{E}(x_i \mid \mathbf{x}_{-i}, \kappa) &= \frac{4}{6}(x_{i+1} + x_{i-1}) - \frac{1}{6}(x_{i+2} + x_{i-2}), \\ \mathrm{Prec}(x_i \mid \mathbf{x}_{-i}, \kappa) &= 6\kappa, \end{aligned}$$

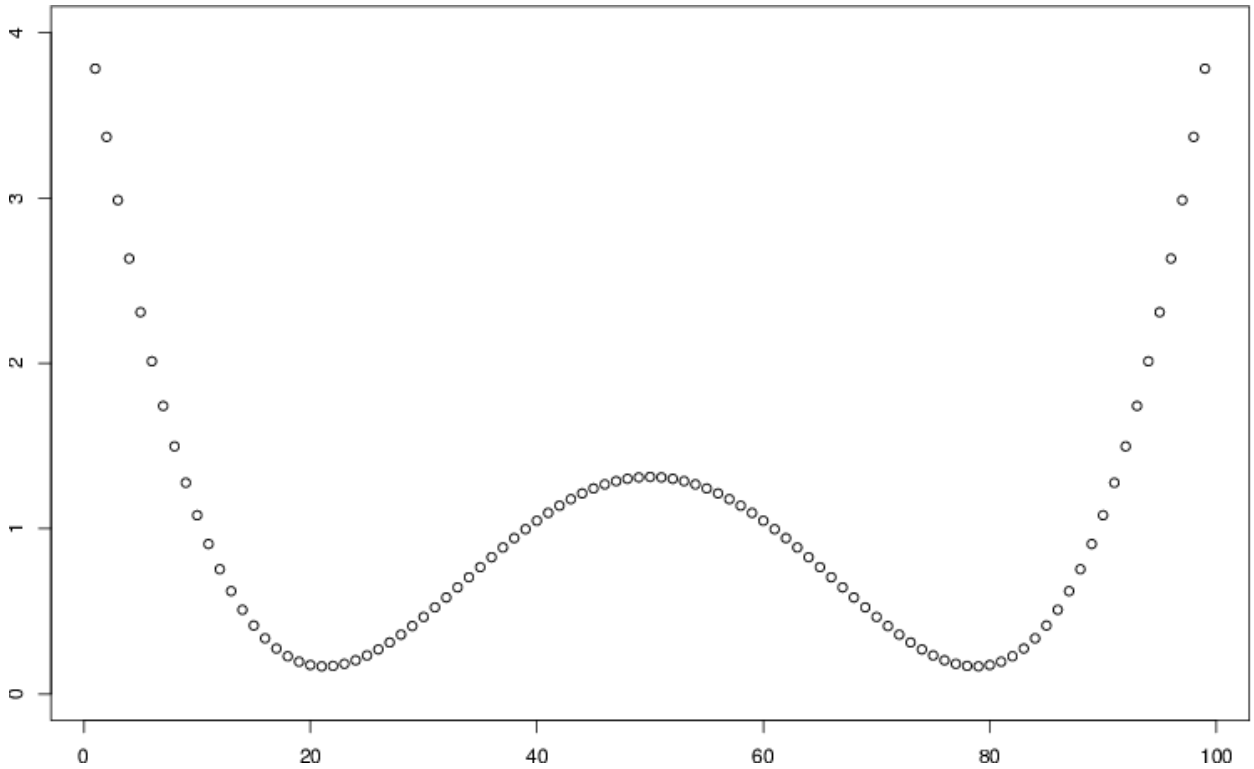
respectively for $2 < i < n - 2$.

- Verify directly that $\mathbf{Q}\mathbf{S}_1 = \mathbf{0}$ and that the rank of $\mathbf{Q}$ is $n - 2$.
- IGMRF of second order: invariant to the adding line to $\mathbf{x}$.
- Known as the second order random walk model, denoted by RW2(κ) or simply RW2 model.

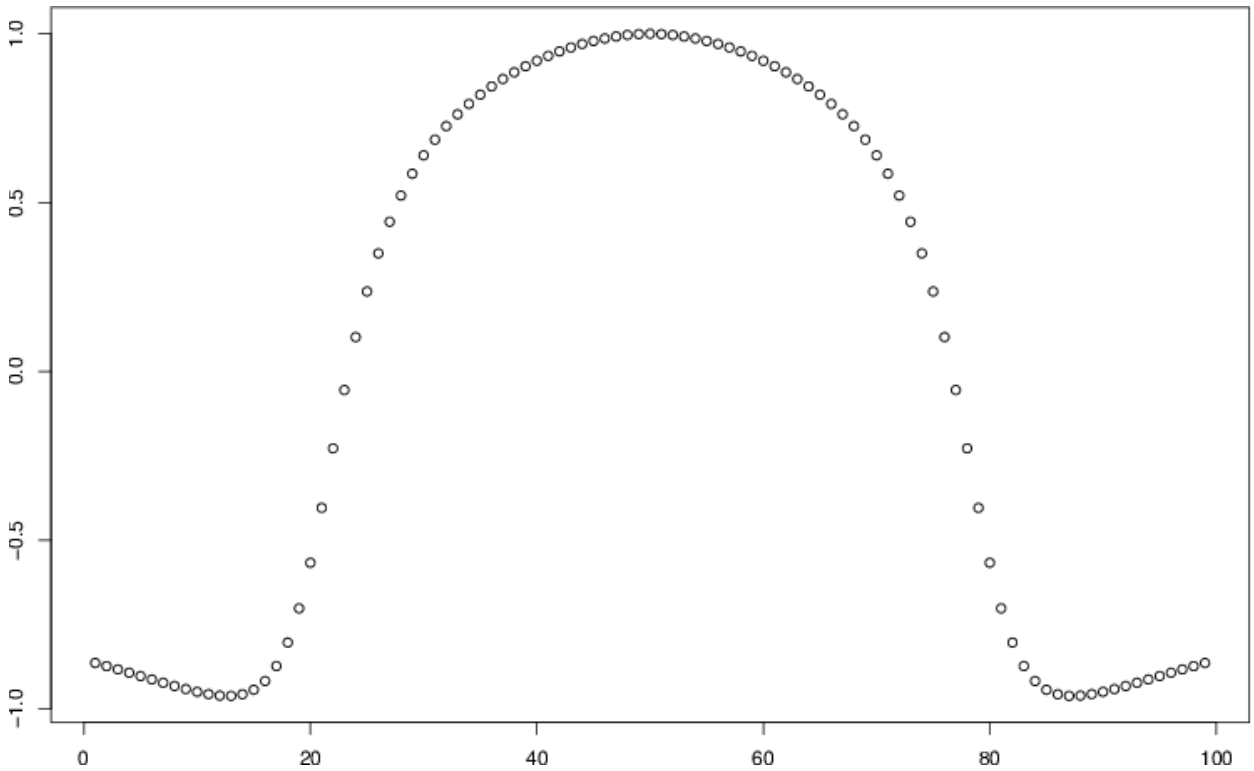
Example



Samples with $n = 99$ and $\kappa = 1$ by conditioning on the constraints.



Marginal variances $\mathrm{Var}(x_i)$ for $i = 1, \ldots, n$.



$\mathrm{Corr}(x_{n/2}, x_i)$ for $i = 1, \ldots, n$

Smoothness and limiting properties

The forward difference of kth order as an approximation to the kth derivative of $f(z)$,

$$f'(z) = \lim_{h \to 0} \frac{f(z+h) - f(z)}{h}$$

We can think of smoothness of the in the limit as the distance between the points goes to zero relative to the domain size

In the limit we are asking that the kth derivative isn't too large.

We can show (later), that this ends up meaning that we have k weak derivatives, or (almost) $k - 1/2$ continuous derivatives

Thus, the RW1 model contraints the field to have 1/2 derivative, and the RW2 model has 3/2 derivatives (Think Hölder continuity)

What about irregular locations?

We can define these models at irregular locations.

Let the location of x_i is s_i but $s_{i+1} - s_i$ is not constant.

Assume $s_1 < s_2 < \cdots < s_n$ and let $\delta_i \stackrel{\text{def}}{=} s_{i+1} - s_i$.

There has been a surprising number of false starts in how to define the models in this case!

It turns out that we just need to make sure that the differences over the irregular locations still converges to the derivatives.

Everything else works the same!

The first order RW for irregular locations

The RW1 model is straightforward to extend to irregular locations

Consider x_i as the realisation of a continuous time Wiener process $W(t)$, at time s_i.

Definition (Wiener process)

- A Wiener process with precision κ is a continuous-time stochastic process $W(t)$ for $t \geq 0$ with $W(0) = 0$ and such that the increments $W(t) - W(s)$ are Gaussian with mean 0 and variance $(t - s)/\kappa$ for any $0 \leq s < t$.
- Increments for non overlapping time intervals are independent.
- For $\kappa = 1$, this process is called a standard Wiener process.

Brownian bridge

The full conditional is in this case known as the the *Brownian bridge*:

$$\mathsf{E}(x_i \mid \mathbf{x}_{-i}, \kappa) = \frac{\delta_i}{\delta_{i-1} + \delta_i} x_{i-1} + \frac{\delta_{i-1}}{\delta_{i-1} + \delta_i} x_{i+1} \tag{17}$$

and

$$\mathrm{Prec}(x_i \mid \mathbf{x}_{-i}, \kappa) = \kappa \left(\frac{1}{\delta_{i-1}} + \frac{1}{\delta_i} \right). \tag{18}$$

κ is a precision parameter.

The joint density

The precision matrix is now

$$Q_{ij} = \kappa \begin{cases} \frac{1}{\delta_{i-1}} + \frac{1}{\delta_i} & j = i \\ -\frac{1}{\delta_i} & j = i+1 \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

for $1 < i < n$.

A proper correction at the boundary gives the remaining diagonal terms $Q_{11} = \kappa/\delta_1$, $Q_{nn} = \kappa/\delta_{n-1}$.

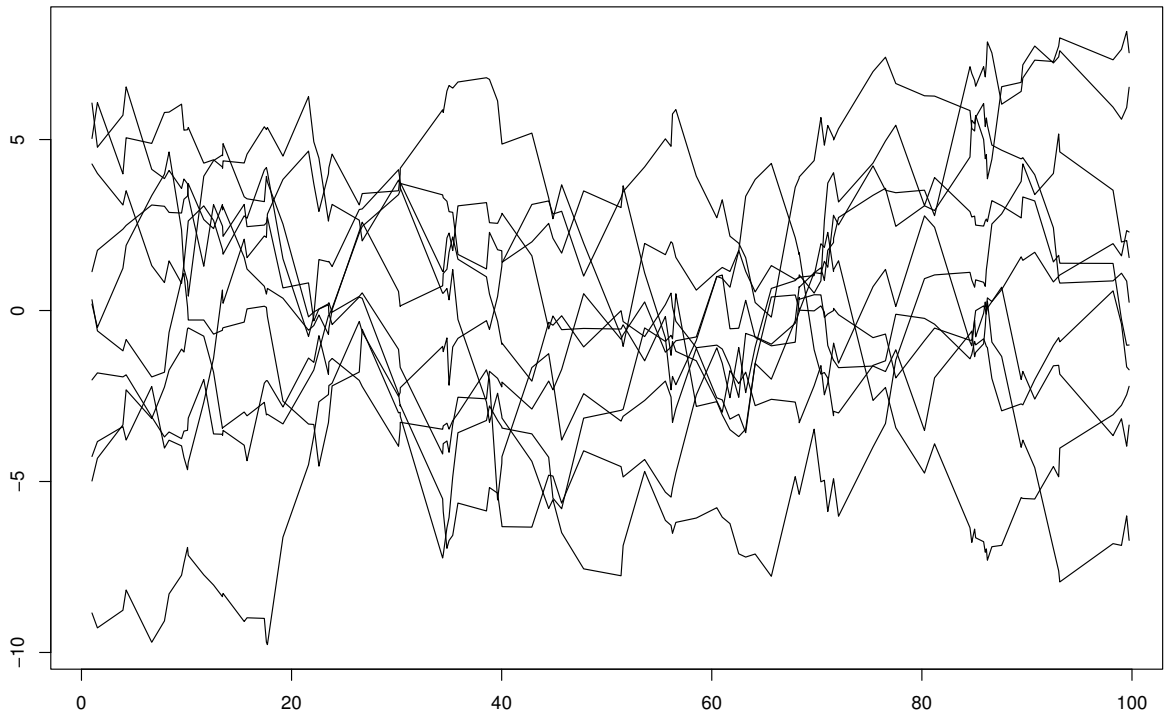
The joint density is

$$\pi(\mathbf{x} \mid \kappa) \propto \kappa^{(n-1)/2} \exp\left(-\frac{\kappa}{2} \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2/\delta_i \right), \tag{20}$$

and is invariant to the addition of a constant.

The natural limiting interpretation

- The interpretation of a RW1 model as a discretely observed Wiener-process, justifies the corrections needed.
- The underlying model *is the same*, it is only observed differently.



The RW2 model for irregular locations (I)

There are many *ad hoc* ways in which we can extend the RW2 model to irregular locations (see the book)

A (much) better alternative is to consider the RW2 model as a discretely observed continuous time process, $x(t)$, as we did for the RW1 model.

Let $x(t)$ be the solution of the stochastic differential equation

$$\Delta x(t) = \frac{dW(t)}{dt}$$

where $\Delta = d^2/dt^2$ and $W(t)$ is the standard Wiener process

This approach can be motivated by the connection between smoothing splines and integrated Wiener proesses (See Grace Wahba's paper on smoothing splines in JRSSB from 1978)

The RW2 model for irregular locations (II)

A Galerkin approximation of the diffusion, using piecewise linear test functions gives us a precision matrix with elements

$$Q_{i,i} = \frac{2}{\delta_{i-1}^2(\delta_{i-2}+\delta_{i-1})} + \frac{2}{\delta_{i-1}\delta_i}\left(\frac{1}{\delta_{i-1}}+\frac{1}{\delta_i}\right) + \frac{2}{\delta_i^2(\delta_i+\delta_{i+1})}$$

$$Q_{i,i-1} = -\frac{2}{\delta_{i-1}^2}\left(\frac{1}{\delta_{i-2}}+\frac{1}{\delta_i}\right)$$

$$Q_{i,i-2} = \frac{2}{\delta_{i-2}\delta_{i-1}(\delta_{i-2}+\delta_{i-1})}$$

and by symmetry: $Q_{i,i+1} = Q_{i+1,i}$, $Q_{i,i+2} = Q_{i+2,i}$

Details on the procedure will be given later during the SPDE lectures! If you cannot wait for that, see reference on homepage for details on this specific construction.

First order IGMRFs on regular lattices

For a lattice $\mathcal{I}_{\mathbf{N}}$ with $n = n_1 n_2$ nodes, let (i, j) denote the node in the ith row and jth column.

Use the nearest four sites of (i, j) as its neighbours

$$(i+1, j),\ (i-1, j),\ (i, j+1),\ (i, j-1).$$

Specify the precision matrix using the stencil

$$\begin{bmatrix} \cdot & -1 & \cdot \\ -1 & 4 & -1 \\ \cdot & -1 & \cdot \end{bmatrix} = 4 \begin{smallmatrix} \circ & \circ & \circ \\ \circ & \bullet & \circ \\ \circ & \circ & \circ \end{smallmatrix} - 1 \begin{smallmatrix} \circ & \bullet & \circ \\ \bullet & \circ & \bullet \\ \circ & \bullet & \circ \end{smallmatrix}$$

which means that the full conditionals are

$$x_{(i,j)} \mid \mathbf{x}_{-(i,j)}, \kappa \sim \mathcal{N}\left(\frac{x_{(i+1,j)} + x_{(i-1,j)} + x_{(i,j+1)} + x_{(i,j-1)}}{4},\ \frac{1}{4\kappa} \right)$$

The precision matrix

The precision matrix $\mathbf{Q}$ has elements

$$Q_{i,j} = \begin{cases} 4 & \text{if } i = j \\ -1 & \text{if } i \sim j \\ 0 & \text{otherwise} \end{cases}$$

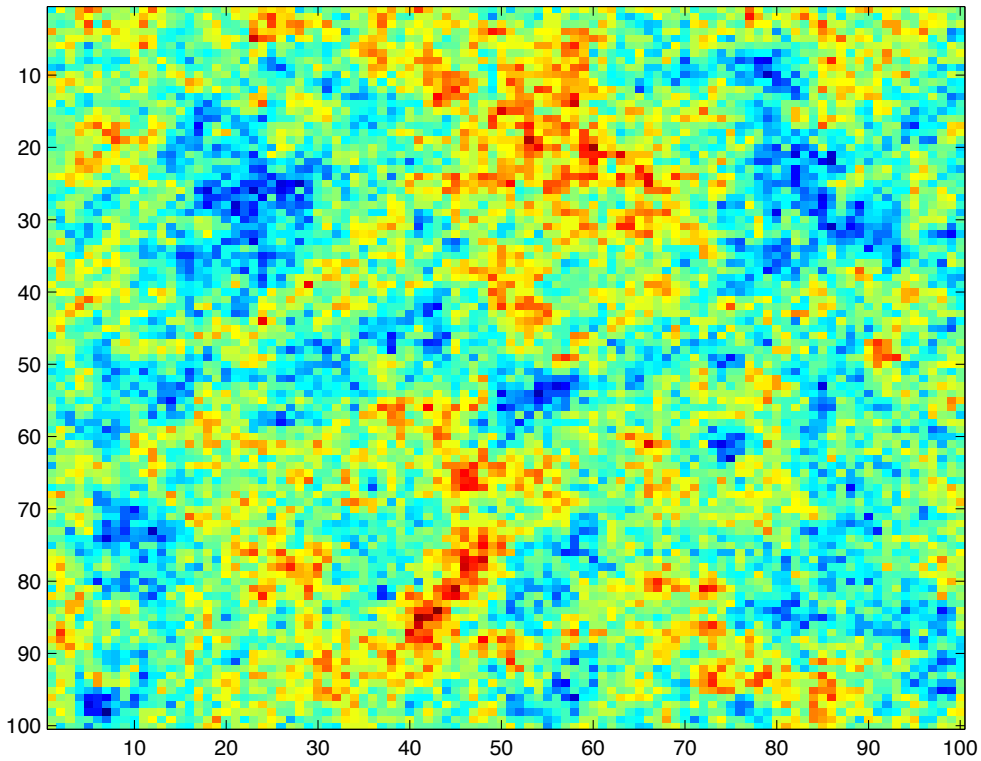
We have

$$\mathbf{Q} = \mathbf{R}_{n_1} \otimes \mathbf{I}_{n_2} + \mathbf{I}_{n_1} \otimes \mathbf{R}_{n_2}$$

where $\mathbf{R}_n$ is the structure matrix of a RW1 model of dimension $n \times n$ and $\mathbf{I}_n$ is the $n \times n$ identity matrix.

(this is useful when implementing the model)

The ICAR(1) model



Limiting behaviour

This model have an important property

> *The process converge to the de Wijs-process: a Gaussian process with variogram*

$$\log(\textit{distance})$$

This is important

- there is a limiting process
- The de Wijs process has dense precision matrix whereas the IGMRF is very sparse!

Another important connection, which we will come back to, is that the de Wijs process can be viewed as a solution to a stochastic PDE

Why do we are about the limiting distribution?

A de Wijs process is an *intrinsic, generalised Gaussian random field*.

- It has generalised covariance function $c(h) \propto log(h)$.
- The field does not make sense point-wise.
- Only "Differences of Integrals" are defined!

Interpret first order CAR models in 2D with great care!
It is not at all obvious from the definition of the model!

This is nothing like the behaviour in 1D

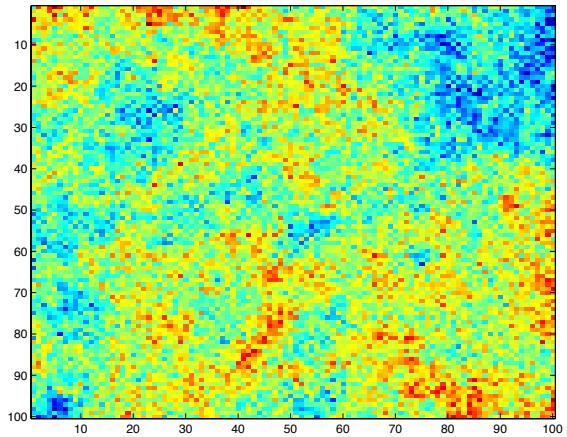
Don't bring your 1D intuition to 2D - it will let you down!

This is different to working with (common) isotropic covariance functions: they don't care about dimension!

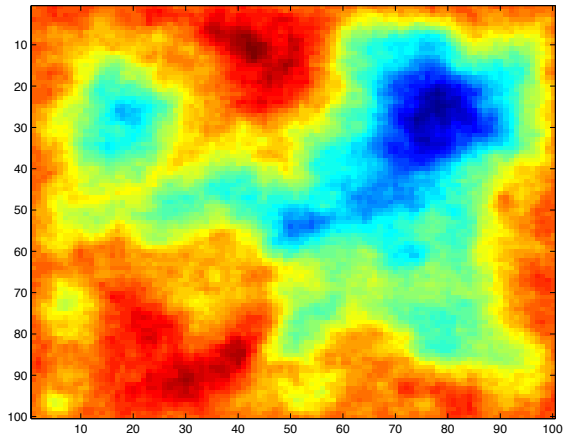
IGMRFs of higher order on regular lattices

In the ICAR(1) model the conditional mean is the average of the neighbouring values. If we want something smoother we can try:

- Make the neighbourhood bigger by averaging second neighbours.
  - Sensible! (And common)
  - Doesn't work!



- Replace $\mathbf{Q}$ with $\mathbf{Q}^2$.
  - Not as obious.
  - The "correct" thing to do!



A second order IGMRF in two dimensions

Consider a regular lattice $\mathcal{I}_{\mathbf{N}}$ in $d = 2$ dimensions Choose the "independent" increments

$$\left(x_{(i+1,j)} + x_{(i-1,j)} + x_{(i,j+1)} + x_{(i,j-1)}\right) - 4x_{(i,j)} \sim \mathsf{N}(0, \kappa^{-1})$$

The motivation for this choice is that the construction says

$$\left(\Delta_i^2 + \Delta_j^2\right) x_{i,j} \sim \mathsf{N}(0, \kappa^{-1})$$

Invariant to adding a first order polynomial

$$p_{1,2}(i, j) = \beta_{00} + \beta_{10} i + \beta_{01} j,$$

Interlude: Why "independent"?

There are hidden constraints in the increments due to the more complicated geometry on a lattice than on the line.

Example

Let $n = 3$ where all nodes are neighbours. Then $x_1 - x_2 = \epsilon_1$, $x_2 - x_3 = \epsilon_2$, and $x_3 - x_1 = \epsilon_3$, where ϵ_1, ϵ_2 and ϵ_3 are the increments.

This implies that

$$\epsilon_1 + \epsilon_2 + \epsilon_3 = 0$$

which is the 'hidden' linear constraint.

The thin plate spline connection

The precision matrix (apart from boundary effects) should have non-zero elements

$$-\left(\Delta_i^2+\Delta_j^2\right)^2=-\left(\Delta_i^4+2\Delta_i^2\Delta_j^2+\Delta_j^4\right)$$

which is a negative difference approximation to the *biharmonic* differential operator

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)^2=\frac{\partial^4}{\partial x^4}+2\frac{\partial^4}{\partial x^2\partial y^2}+\frac{\partial^4}{\partial y^4}.$$

The fundamental solution of the biharmonic equation

$$\left(\frac{\partial^4}{\partial x^4}+2\frac{\partial^4}{\partial x^2\partial y^2}+\frac{\partial^4}{\partial y^4}\right)\phi(x,y)=0$$

is the *thin plate spline*!

Comments

Full conditionals in the interior

$$\mathrm{E}(x_i \mid \mathbf{x}_{-i}) = \frac{1}{20}\left(8 \; \begin{array}{ccccc} \circ & \circ & \circ & \circ & \circ \\ \circ & \circ & \bullet & \circ & \circ \\ \circ & \bullet & \circ & \bullet & \circ \\ \circ & \circ & \bullet & \circ & \circ \\ \circ & \circ & \circ & \circ & \circ \end{array} \; - 2 \; \begin{array}{ccccc} \circ & \circ & \circ & \circ & \circ \\ \circ & \bullet & \circ & \bullet & \circ \\ \circ & \circ & \circ & \circ & \circ \\ \circ & \bullet & \circ & \bullet & \circ \\ \circ & \circ & \circ & \circ & \circ \end{array} \; - 1 \; \begin{array}{ccccc} \circ & \circ & \bullet & \circ & \circ \\ \circ & \circ & \circ & \circ & \circ \\ \bullet & \circ & \circ & \circ & \bullet \\ \circ & \circ & \circ & \circ & \circ \\ \circ & \circ & \bullet & \circ & \circ \end{array} \right) \tag{21}$$

and

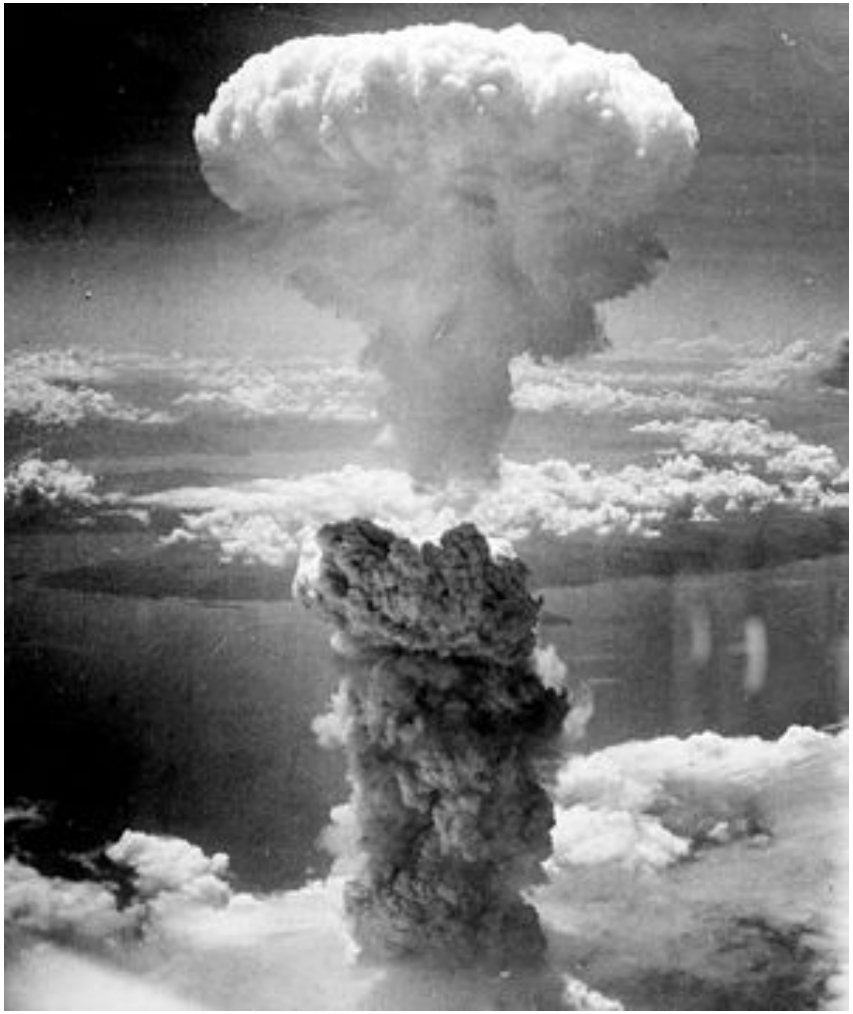
$$\mathrm{Prec}(x_i \mid \mathbf{x}_{-i}) = 20\kappa. \tag{22}$$

This model is sometimes called the ICAR(2) or SAR(1) model

In the limit, it has (almost) 1 continuous derivate. This is again less than in 1D! The limiting properties of a GMRF depends strongly on dimension.

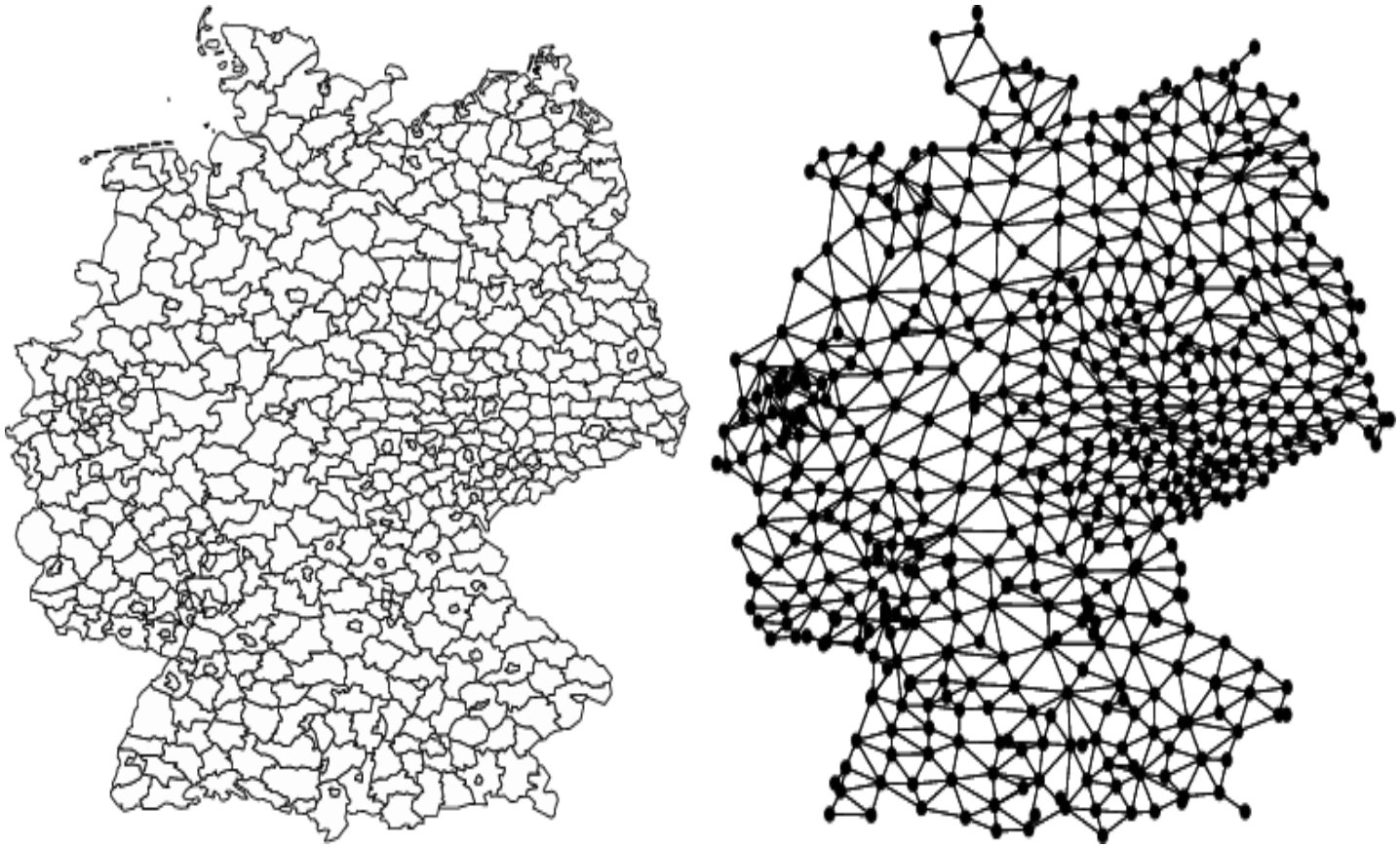
We can generalise to higher orders, but we need fractional stochastic PDEs to do it properly: We will come back to this later!

What happens when we're off the lattice?



But things don't tend to happen on regular lattices...

Consider the map of the 544 regions in Germany.
Two regions are *neighbours* if they share a common border.



The Besag construction

Between neighbouring regions i and j, say, we define a "independent" Gaussian increment

$$x_i - x_j \sim \mathcal{N}(0, \kappa^{-1}) \tag{23}$$

Assume independent increments yields

$$\pi(\mathbf{x}) \propto \kappa^{(n-1)/2} \exp\left(-\frac{\kappa}{2} \sum_{i \sim j} (x_i - x_j)^2\right). \tag{24}$$

"$i \sim j$" denotes the set of all *unordered* pairs of neighbours.

Number of increments $|i \sim j|$ is larger than n, but the rank of the corresponding precision matrix is still $n - 1$.

The precision matrix

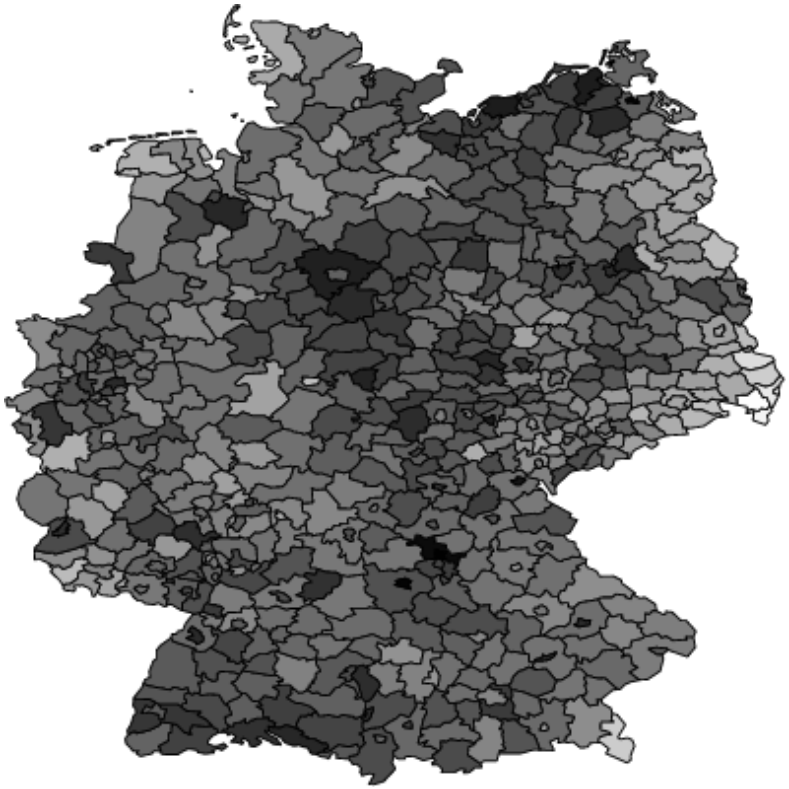
Let n_i denote the number of neighbours of region i.
The precision matrix $\mathbf{Q}$ is

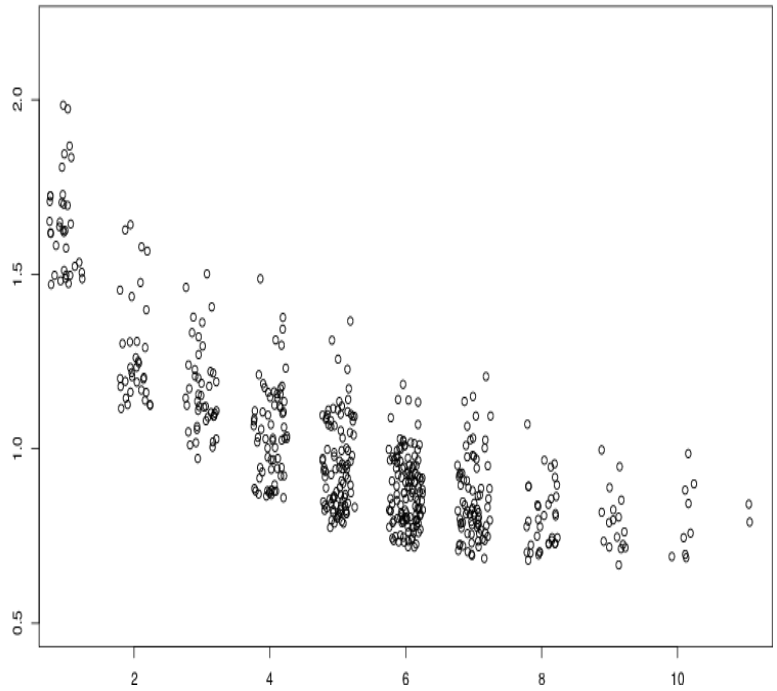
$$Q_{ij} = \kappa \begin{cases} n_i & i = j, \\ -1 & i \sim j, \\ 0 & \text{otherwise}, \end{cases} \tag{25}$$

The full conditionals are

$$x_i \mid \mathbf{x}_{-i}, \kappa \ \sim \ \mathcal{N}(\frac{1}{n_i} \sum_{j:j\sim i} x_j, \ \frac{1}{n_i \kappa}). \tag{26}$$

Example





- Left: A sample from the model
- Right: The marginal variance of the nodes as a function of the number of neighbors

Weighted variants

Incorporate symmetric weights w_{ij} for each pair of adjacent nodes i and j.

For example, $w_{ij} = 1/d(i,j)$

Assuming independent increments

$$x_i - x_j \sim \mathcal{N}(0, 1/(w_{ij}\kappa)), \tag{27}$$

then

$$\pi(\mathbf{x}) \propto \kappa^{(n-1)/2} \exp\left(-\frac{\kappa}{2}\sum_{i\sim j} w_{ij}(x_i - x_j)^2\right). \tag{28}$$

The precision matrix

The precision matrix is

$$Q_{ij} = \kappa \begin{cases} \sum_{k:k\sim i} w_{ik} & i = j, \\ -1/w_{ij} & i \sim j, \\ 0 & \text{otherwise}, \end{cases} \tag{29}$$

and with mean and precision

$$\frac{\sum_{j:j\sim i} x_j w_{ij}}{\sum_{j:j\sim i} w_{ij}} \quad \text{and} \quad \kappa \sum_{j:j\sim i} w_{ij}, \tag{30}$$

respectively.

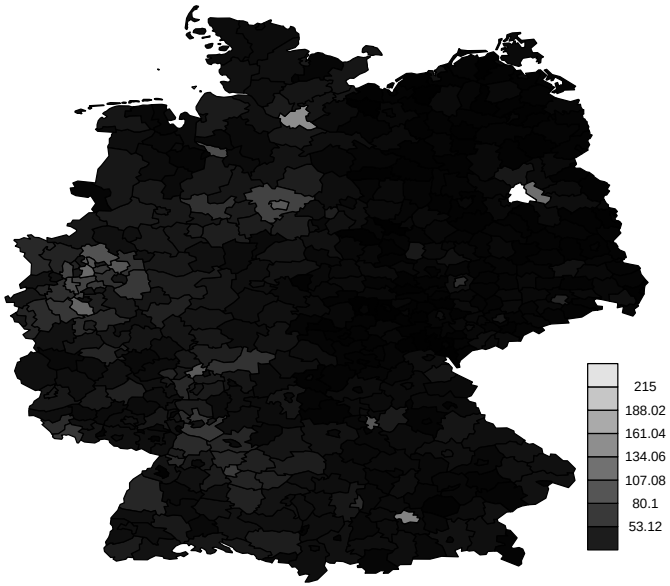
This seems too easy...

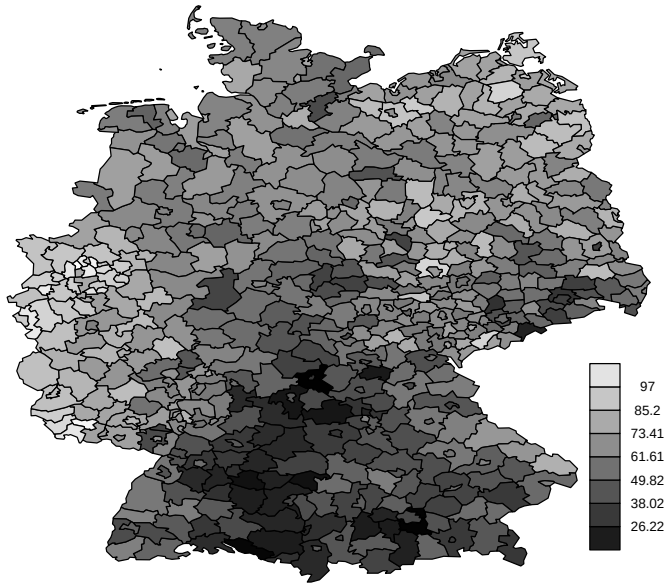
What's the catch?

- In general there is no longer an underlying continuous stochastic process that we can relate to this density.
- If we change the spatial resolution or split a region into two new ones, we change the model.

This is bad!

Example: Larnyx cancer mortality





- Larnyx cancer mortality rates are observed in the 544 districts of Germany from 1986 to 1990.
- We also know the level of smoking consumption, the number of people in each district, and their age distribution.

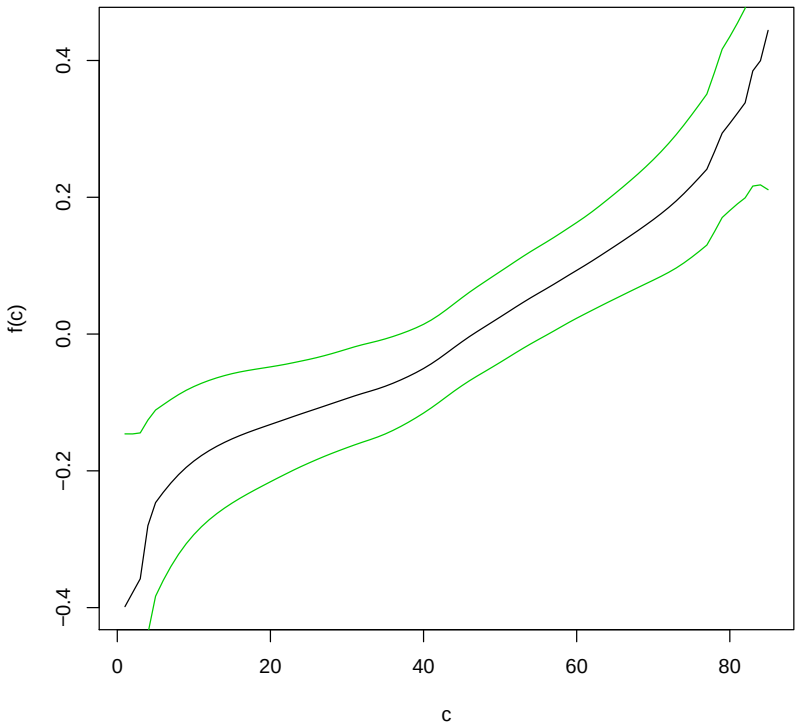
Larnyx cancer: model

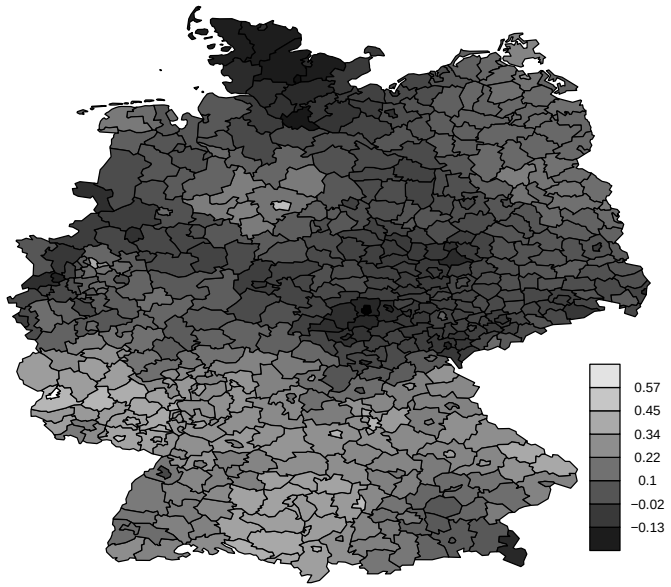
- We assume the data to be conditionally independent Poisson random variables with mean $E_i \exp(\eta_i)$.
- E_i is fixed and accounts for demographic variation and η_i is the log-relative risk.
- The model for η_i takes the following form

$$\eta_i = \mu + f(s_i) + f(c_i) + u_i$$

- μ is an intercept and $\mathbf{u}$ is an unstructured random effect.
- $f(c)$ is a smooth effect of the smoking consumption c, which we model using a RW1 model.
- $f(s)$ is a spatial effect, which we model as a Besag model.

Larnyx cancer results





- Left: The estimated effect of the smoking consumption c, together with marginal confidence bands.
- Right: The spatial effect.

Tomorrow: Project 1

There will be two projects in the course

- You will work on Project 1 during the first two computer labs
- and Project 2 during the last two computer labs.

You can work alone or in pairs, and I strongly recommend that you work in pairs.

- These projects may take some time to complete

Tomorrow, I will start with some basics for how to use sparse matrices in R and Matlab

I will put up the project on the course homepage this afternoon

- The project description is quite brief. You will probably need more details from me.

Also, if you need more help outside of the computer labs, you are welcome to come by my office.