Bayesian Statistics with R-INLA - Part 1

Geilo, January, 2023

Sara Martino

NTNU
Norwegian University of Science and Technology

Outline

Introduction

Last slide from Geir's talk

- Assume a Hierarchical model

$$\begin{aligned} \mathbf{y} &\sim p(\mathbf{y}|\mathbf{x}, \theta) && \text{Observations} \\ \mathbf{x} &\sim p(\mathbf{x}|\theta) && \text{Latent process} \\ \theta &\sim p(\theta) && \text{parameters} \end{aligned}$$

- Main interest in θ

$$p(\theta|\mathbf{y}) = \int_{\mathbf{x}} p(\theta, \mathbf{x}|\mathbf{y}) d\mathbf{x}$$

Last slide from Geir's talk

- Assume a Hierarchical model

$$\begin{aligned} \mathbf{y} &\sim p(\mathbf{y}|\mathbf{x}, \theta) && \text{Observations} \\ \mathbf{x} &\sim p(\mathbf{x}|\theta) && \text{Latent process Gaussian!!!} \\ \theta &\sim p(\theta) && \text{parameters} \end{aligned}$$

- Main interest in θ

$$p(\theta|\mathbf{y}) = \int_{\mathbf{x}} p(\theta, \mathbf{x}|\mathbf{y}) d\mathbf{x}$$

- For some models it is possible to create an **independent proposal** $p(\theta)$ that is very close to the target $p(\theta|\mathbf{y})$!!

Last slide from Geir's talk

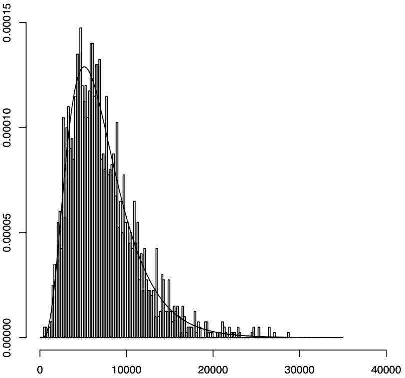


Figure 5.12 *The histogram of the posterior marginal for* κ *based on* 2000 *successive samples from the independence samples constructed from A2. The solid line is the approximation* $\widetilde{\pi}(\kappa|\boldsymbol{y})$.

- Why not just *use* the proposal then??...this was start for INLA!

What is inla?

The short answer:

> *INLA is a fast method to do approximate Bayesian inference with latent Gaussian models and* `INLA` *(and 'inlabru') are* `R`*-packages that implement this method with a flexible and simple interface.*

What is inla?

The short answer:

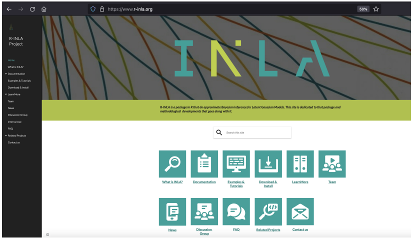
> *INLA is a fast method to do approximate Bayesian inference with latent Gaussian models and* `INLA` *(and 'inlabru') are* `R`*-packages that implement this method with a flexible and simple interface.*

The (much) longer answer:

- Rue, Martino, and Chopin (2009) "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations." *JRSSB*
- Rue, Riebler, Sørbye, Illian, Simpson, Lindgren (2017) "Bayesian Computing with INLA: A Review." *Annual Review of Statistics and Its Application*
- Martino, Riebler "Integrated Nested Laplace Approximations (INLA)" (2021) *arXiv:1907.01248*

Where?

The software, information, examples and help can be found at `http://www.r-inla.org`



- paper
- tutorials
- discussion group
- ...

So... Why should you use R-INLA?

- Why approximate inference?
- What type of problems and models can we solve?
- When can we use it?

Bayesian Inference

- **Likelihood** $\pi(\mathbf{y}|\theta)$
- **Prior** $\pi(\theta)$
- **Posterior**

$$\pi(\theta|\mathbf{y}) = \frac{\pi(\mathbf{y}, \theta)}{\pi(\mathbf{y})} \propto \pi(\mathbf{y}|\theta)\pi(\theta)$$

Background I

- A big obstacle for Bayesian modelling is the issue of "Bayesian Computing"

Background I

- A big obstacle for Bayesian modelling is the issue of "Bayesian Computing"
- MCMC is a very general solution but can be slow, can have convergence problems . . .

Background I

- A big obstacle for Bayesian modelling is the issue of "Bayesian Computing"
- MCMC is a very general solution but can be slow, can have convergence problems ...
- There are (somewhat) generic tools based on MCMC like `JAGS/OpenBUGS/stan`...

Background II

GLM/GAM/GLMM/GAMM/++

- Perhaps the most used class of statistical models

Background II

GLM/GAM/GLMM/GAMM/++

- Perhaps the most used class of statistical models
- All these models can be seen as Bayesian hyperarchical models...

Background II

GLM/GAM/GLMM/GAMM/++

- Perhaps the most used class of statistical models
- All these models can be seen as Bayesian hyperarchical models...
- ...more specifically then can be cast into the class of "Latent Gaussian Models"

Background II

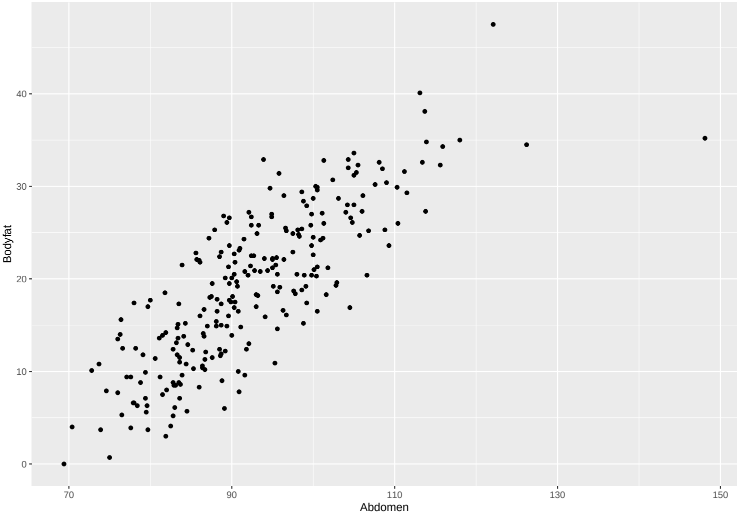
GLM/GAM/GLMM/GAMM/++

- Perhaps the most used class of statistical models
- All these models can be seen as Bayesian hyperarchical models…
- …more specifically then can be cast into the class of "Latent Gaussian Models"
- For such "restricted" class INLA beats MCMC in terms of velocity and accuracy.

Bayesian Hierarchical models

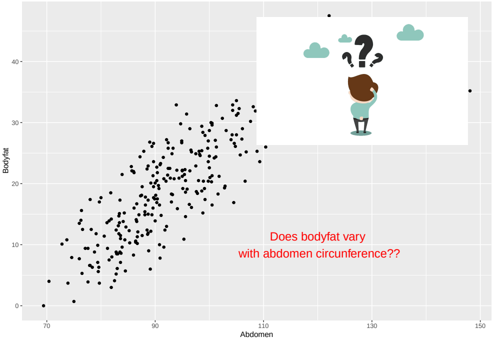
Build a Bayesian model, a simple example

1. We observed something…



Build a Bayesian model, a simple example

2. We formulate questions ..



Build a Bayesian model, a simple example

3. We formulate a model to answer the questions ..

- The observational model
- How data depend on each other/ or other quantities
- What is our "information" prior to the observation process

A Bayesian regression model

- Perc. Body fat $(y_1, \ldots, y_n)$ are Gaussian distributed
    - The mean η_i depends on the abdomen circunference
    - The precision is constant

$$y_i|\eta_i, \tau \sim \mathcal{N}(\eta_i, \tau^{-1})$$
$$\eta_i = \alpha + \beta x_i$$

- Need priors for α, β, τ
    - $(\alpha, \beta) \sim \mathcal{N}(0, \mathrm{diag}(\sigma^2_\alpha, \sigma^2_\beta))$ with $\sigma^2_\alpha, \sigma^2_\beta$ known
    - $\tau \sim \mathrm{Gamma}(a, b)$ with a, b known

A Bayesian hyerarchical model

- **Observation model**

$$\mathbf{y} \mid \underbrace{\alpha, \beta}_{\mathbf{x}}, \underbrace{\tau}_{\theta}$$

Encodes information about observed data

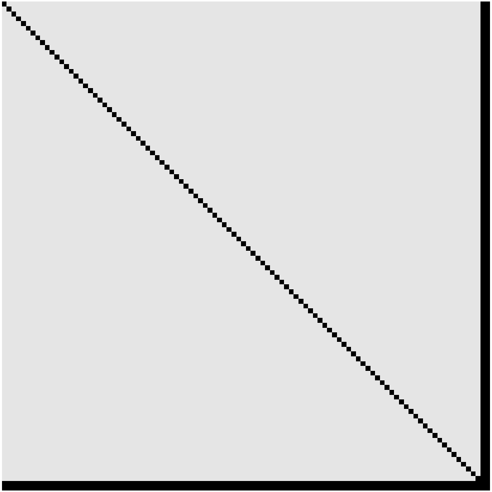
- **Latent model**

$$\mathbf{x} = (\alpha, \beta, \eta) \sim \mathcal{N}(0, \mathbf{Q}^{-1}(\theta))$$

The unobserved process

- **Hyperparameters** $\theta = \tau$

Precision Matrix $\mathbf{Q}(\theta)$



Bayesian Computations

From this we can compute the posterior distribution

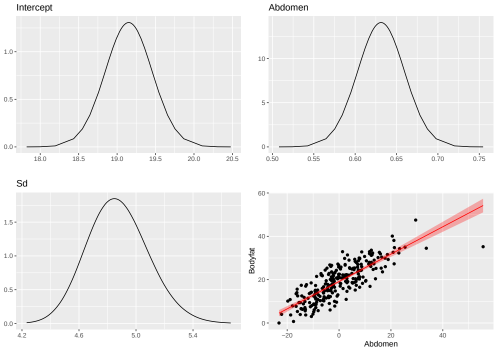
$$\pi(\mathbf{x}, \theta|\mathbf{y}) \propto \pi(\mathbf{y}|\mathbf{x}, \theta)\pi(\mathbf{x})\pi(\theta)$$

and then the corresponding posterior marginal distributions:

$$\begin{aligned} \pi(x_j|\mathbf{y}) &\qquad j = 1, 2 \\ \pi(\tau|\mathbf{y}) & \\ \pi(\eta_i|\mathbf{y}) &\qquad i = 1, \ldots, n \end{aligned}$$

Results

- Assign priors to α, β, τ
- Use Bayes theorem to compute posterior distributions



On the prior choice. . . .

- Priors are an important part of the model
- There are several "schools" about priors
- "non-informative" priors are not always a good choice
- for complex models priors can have a large influence on the results.

Real-world datasets are usually much more complicated!

Using a Bayesian framework:

- Build (hierarchical) models to account for potentially complicated dependency structures in the data.
- Attribute uncertainty to model parameters and latent variables using priors.

Two main challenges:

1. Need computationally efficient methods to calculate posteriors.
2. Select priors in a sensible way

Bayesian hierarchical models

INLA can be used with Bayesian hierarchical models where we model in different stages or levels:

- **Stage 1:** What is the distribution of the responses?
- **Stage 2:** What is the distribution of the underlying unobserved (latent) components?
- **Stage 3:** What are our prior beliefs about the parameters controlling the components in the model?

Stage 1: The data generating process

How is our data ($\boldsymbol{y}$) generated from the underlying components ($\boldsymbol{x}$) and hyperparameters ($\boldsymbol{\theta}$) in the model:

- Gaussian response? (temperature, rainfall, fish weight ...)

Stage 1: The data generating process

How is our data ($\boldsymbol{y}$) generated from the underlying components ($\boldsymbol{x}$) and hyperparameters ($\boldsymbol{\theta}$) in the model:

- Gaussian response? (temperature, rainfall, fish weight ...)
- Count data? (people infected with a disease in each area)

Stage 1: The data generating process

How is our data ($\boldsymbol{y}$) generated from the underlying components ($\boldsymbol{x}$) and hyperparameters ($\boldsymbol{\theta}$) in the model:

- Gaussian response? (temperature, rainfall, fish weight ...)
- Count data? (people infected with a disease in each area)
- Point pattern? (locations of trees in a forest)

Stage 1: The data generating process

How is our data ($\boldsymbol{y}$) generated from the underlying components ($\boldsymbol{x}$) and hyperparameters ($\boldsymbol{\theta}$) in the model:

- Gaussian response? (temperature, rainfall, fish weight ...)
- Count data? (people infected with a disease in each area)
- Point pattern? (locations of trees in a forest)
- Binary data? (yes/no response, binary image)

Stage 1: The data generating process

How is our data ($\boldsymbol{y}$) generated from the underlying components ($\boldsymbol{x}$) and hyperparameters ($\boldsymbol{\theta}$) in the model:

- Gaussian response? (temperature, rainfall, fish weight ...)
- Count data? (people infected with a disease in each area)
- Point pattern? (locations of trees in a forest)
- Binary data? (yes/no response, binary image)
- Survival data? (recovery time, time to death)

Stage 1: The data generating process

How is our data ($\boldsymbol{y}$) generated from the underlying components ($\boldsymbol{x}$) and hyperparameters ($\boldsymbol{\theta}$) in the model:

- Gaussian response? (temperature, rainfall, fish weight ...)
- Count data? (people infected with a disease in each area)
- Point pattern? (locations of trees in a forest)
- Binary data? (yes/no response, binary image)
- Survival data? (recovery time, time to death)

Stage 1: The data generating process

How is our data ($\boldsymbol{y}$) generated from the underlying components ($\boldsymbol{x}$) and hyperparameters ($\boldsymbol{\theta}$) in the model:

- Gaussian response? (temperature, rainfall, fish weight . . .)
- Count data? (people infected with a disease in each area)
- Point pattern? (locations of trees in a forest)
- Binary data? (yes/no response, binary image)
- Survival data? (recovery time, time to death)

This information is placed into our likelihood $\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})$

Stage 1: The data generating process

We assume that *given* the underlying components ($\boldsymbol{x}$) and hyperparameters ($\boldsymbol{\theta}$) the data are independent on each other

$$\pi(\mathbf{y}|\mathbf{x}, \theta) = \prod_{i \in \mathcal{I}} \pi(y_i|x_i, \theta)$$

Stage 1: The data generating process

We assume that *given* the underlying components ($\boldsymbol{x}$) and hyperparameters ($\boldsymbol{\theta}$) the data are independent on each other

$$\pi(\mathbf{y}|\mathbf{x},\theta) = \prod_{i\in\mathcal{I}} \pi(y_i|x_i,\theta)$$

This implies that all the dependence structure in the data is explained in Stage II !!

Stage 1: The data generating process

We assume that *given* the underlying components ($\boldsymbol{x}$) and hyperparameters ($\boldsymbol{\theta}$) the data are independent on each other

$$\pi(\mathbf{y}|\mathbf{x}, \theta) = \prod_{i \in \mathcal{I}} \pi(y_i|x_i, \theta)$$

This implies that all the dependence structure in the data is explained in Stage II !!

Can you think of a model that does not respect this condition?

Stage 2: The dependence structure

The underlying unobserved components $\boldsymbol{x}$ are called **latent components** and can be:

- Fixed effects for covariates
- Unstructured random effects (individual effects, group effects)
- Structured random effects (AR(1), regional effects, ...)

These are linked to the responses in the likelihood through linear predictors.

Stage 3: The hyperparameters

The likelihood and the latent model typically have hyperparameters that control their behavior.

The hyperparameters $\boldsymbol{\theta}$ can include:

Stage 3: The hyperparameters

The likelihood and the latent model typically have hyperparameters that control their behavior.

The hyperparameters $\boldsymbol{\theta}$ can include:

Examples likelihood:

- Variance of observation noise
- Dispersion parameter in the negative binomial model
- Probability of a zero (zero-inflated models)

Stage 3: The hyperparameters

The likelihood and the latent model typically have hyperparameters that control their behavior.

The hyperparameters $\boldsymbol{\theta}$ can include:

Examples likelihood:

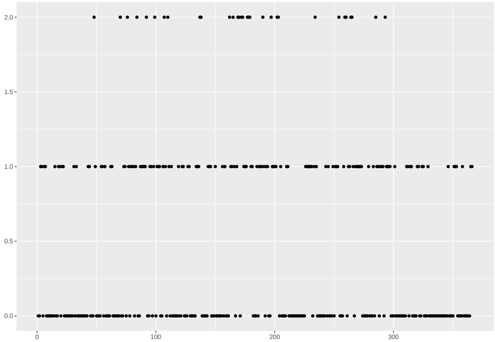
- Variance of observation noise
- Dispersion parameter in the negative binomial model
- Probability of a zero (zero-inflated models)

Examples latent model:

- Variance of unstructured effects
- Correlation of multivariate effects
- Range and variance of spatial effects
- Autocorrelation parameter

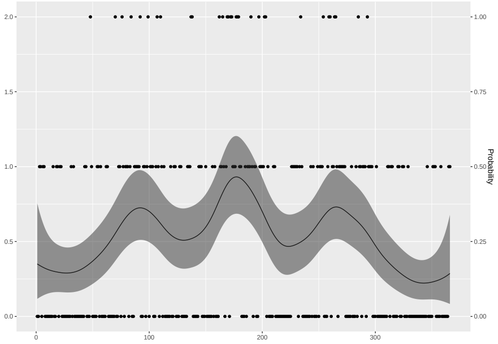
Example 1: Tokyo rainfall data

Rainfall over 1 mm in the Tokyo area for each calendar day during two years (1983-84) are registered.



Tokyo rainfall data

Rainfall over 1 mm in the Tokyo area for each calendar day during two years (1983-84) are registered.



Stage 1: The data

$$y_i \mid p_i \sim \text{Binomial}(n_i, p_i),$$

for $i = 1, 2, ..., 366$

$$n_i = \begin{cases} 1, & \text{for 29 February} \\ 2, & \text{other days} \end{cases}$$

$$y_i = \begin{cases} \{0,1\}, & \text{for 29 February} \\ \{0,1,2\}, & \text{other days} \end{cases}$$

Linear predictor

$$logit(p_i) = \eta_i \quad \Leftrightarrow \quad p_i = \frac{1}{1+\exp(-\eta_i)}$$

- probability of rain on day i depends on η_i
- the likelihood has no hyperparameters θ

Stage 2: The latent model

$$\eta_i = \alpha + u_i + v_i$$

where

- α is a global mean (Gaussian)

Stage 2: The latent model

$$\eta_i = \alpha + u_i + v_i$$

where

- α is a global mean (Gaussian)
- $\mathbf{u}$ is a AR-model like

$$u_i = \phi u_{i-1} + \epsilon_i, \qquad \epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$$

  with parameters (ϕ, σ_ϵ)

Stage 2: The latent model

$$\eta_i = \alpha + u_i + v_i$$

where

- α is a global mean (Gaussian)
- **u** is a AR-model like

$$u_i = \phi u_{i-1} + \epsilon_i, \qquad \epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$$

  with parameters (ϕ, σ_ϵ)

- **v** can be a unstructured term/"random effect"/slowly varying trend. Typically Gaussian with precision parameter τ_v controlling the smoothness.

Stage 2: The latent model

$$\eta_i = \alpha + u_i + v_i$$

where

- α is a global mean (Gaussian)
- $\mathbf{u}$ is a AR-model like

$$u_i = \phi u_{i-1} + \epsilon_i, \qquad \epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$$

  with parameters (ϕ, σ_ϵ)
- $\mathbf{v}$ can be a unstructured term/“random effect”/slowly varying trend. Typically Gaussian with precision parameter τ_v controlling the smoothness.

Stage 2: The latent model

$$\eta_i = \alpha + u_i + v_i$$

where

- α is a global mean (Gaussian)
- $\mathbf{u}$ is a AR-model like

$$u_i = \phi u_{i-1} + \epsilon_i, \qquad \epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$$

  with parameters (ϕ, σ_ϵ)
- $\mathbf{v}$ can be a unstructured term/"random effect"/slowly varying trend. Typically Gaussian with precision parameter τ_v controlling the smoothness.

This gives the latent model $\mathbf{x} = (\alpha, \mathbf{u}, \mathbf{v}, \eta) \sim \mathcal{N}(0, \mathbf{Q}^{-1}(\theta))$.

Stage 3: Hyperparameters

Hyperparameters control the smoothness of the effects in the latent model

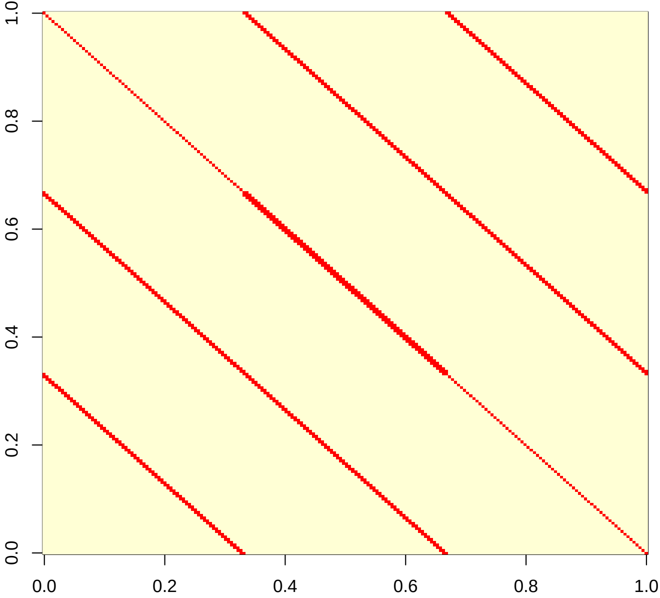
$$\theta = (\phi, \sigma_\epsilon, \sigma_v)$$

The model

We can write the model as

$$\begin{aligned}
\theta &\sim \pi(\theta) \\
\mathbf{x}|\theta &\sim \pi(\mathbf{x}|\theta) \\
\mathbf{y}|\mathbf{x},\theta &\sim \prod_i \pi(y_i|\eta_i,\theta)
\end{aligned}$$

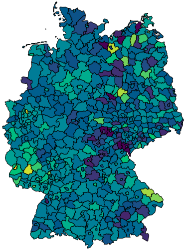
Precision matrix



Example: disease mapping

We observed larynx cancer mortality counts for males in 544 district of Germany from 1986 to 1990 and want to understand the spatial distribution and the inpact of covariates.

- y_i: The count at location i.
- E_i: An offset; expected number of cases in district i.
- c_i: A covariate (level of smoking consumption) at i
- $\boldsymbol{s}_i$: spatial location i .



Bayesian disease mapping

- **Stage 1:** We choose a Poisson distribution for the responses, so that

$$y_i \mid \eta_i \sim \text{Poisson}(E_i \exp(\eta_i)))$$

Bayesian disease mapping

- **Stage 1:** We choose a Poisson distribution for the responses, so that

$$y_i \mid \eta_i \sim \text{Poisson}(E_i \exp(\eta_i)))$$

- **Stage 2:** η_i is a linear function of the latent components: a covariate c_i, a spatially structured effect $\mathbf{u}$, an unstructured effect $\mathbf{v}$ likelihood by

$$\eta_i = \mu + \beta\, c_i + u_i + v_i$$

Bayesian disease mapping

- **Stage 1:** We choose a Poisson distribution for the responses, so that

$$y_i \mid \eta_i \sim \text{Poisson}(E_i \exp(\eta_i)))$$

- **Stage 2:** η_i is a linear function of the latent components: a covariate c_i, a spatially structured effect $\mathbf{u}$, an unstructured effect $\mathbf{v}$ likelihood by

$$\eta_i = \mu + \beta\ c_i + u_i + v_i$$

- **Stage 3:**

Bayesian disease mapping

- **Stage 1:** We choose a Poisson distribution for the responses, so that

$$y_i \mid \eta_i \sim \text{Poisson}(E_i \exp(\eta_i)))$$

- **Stage 2:** η_i is a linear function of the latent components: a covariate c_i, a spatially structured effect $\mathbf{u}$, an unstructured effect $\mathbf{v}$ likelihood by

$$\eta_i = \mu + \beta\, c_i + u_i + v_i$$

- **Stage 3:**
  - τ_u: Precision parameter for the structured effect

Bayesian disease mapping

- **Stage 1:** We choose a Poisson distribution for the responses, so that

$$y_i \mid \eta_i \sim \text{Poisson}(E_i \exp(\eta_i)))$$

- **Stage 2:** η_i is a linear function of the latent components: a covariate c_i, a spatially structured effect $\mathbf{u}$, an unstructured effect $\mathbf{v}$ likelihood by

$$\eta_i = \mu + \beta\, c_i + u_i + v_i$$

- **Stage 3:**
  - τ_u: Precision parameter for the structured effect
  - τ_v: Precision parameter for the unstructured effect

Bayesian disease mapping

- **Stage 1:** We choose a Poisson distribution for the responses, so that

$$y_i \mid \eta_i \sim \text{Poisson}(E_i \exp(\eta_i)))$$

- **Stage 2:** η_i is a linear function of the latent components: a covariate c_i, a spatially structured effect $\mathbf{u}$, an unstructured effect $\mathbf{v}$ likelihood by

$$\eta_i = \mu + \beta\ c_i + u_i + v_i$$

- **Stage 3:**
  - τ_u: Precision parameter for the structured effect
  - τ_v: Precision parameter for the unstructured effect

Bayesian disease mapping

- **Stage 1:** We choose a Poisson distribution for the responses, so that

$$y_i \mid \eta_i \sim \text{Poisson}(E_i \exp(\eta_i)))$$

- **Stage 2:** η_i is a linear function of the latent components: a covariate c_i, a spatially structured effect $\mathbf{u}$, an unstructured effect $\mathbf{v}$ likelihood by

$$\eta_i = \mu + \beta\ c_i + u_i + v_i$$

- **Stage 3:**
  - τ_u: Precision parameter for the structured effect
  - τ_v: Precision parameter for the unstructured effect

The latent field is $\boldsymbol{x} = (\mu, \beta, \mathbf{u}, \mathbf{v})$, the hyperparameters are $\boldsymbol{\theta} = (\tau_u, \tau_v)$, and must be given a prior.

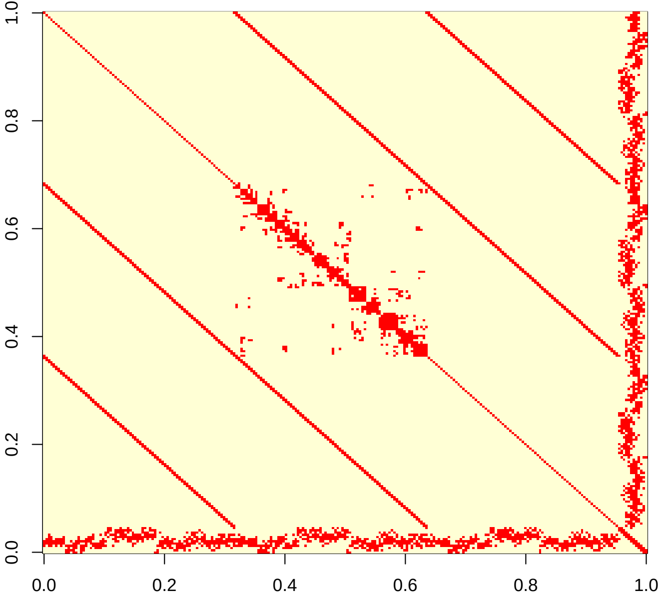
The model

We can write the model as

$$\theta \sim \pi(\theta)$$
$$\mathbf{x}|\theta \sim \pi(\mathbf{x}|\theta)$$
$$\mathbf{y}|\mathbf{x},\theta \sim \prod_i \pi(y_i|\eta_i,\theta)$$

Identical as the one before!!!!

Precision Matrix



Latent Gaussian models

What have we learned so far

Models of the kind:

$$
\begin{aligned}
\theta &\sim \pi(\theta) \\
\mathbf{x}|\theta &\sim \pi(\mathbf{x}|\theta) = \mathcal{N}(0, \mathbf{Q}^{-1}(\theta)) \\
\mathbf{y}|\mathbf{x}, \theta &\sim \prod_i \pi(y_i|\eta_i, \theta)
\end{aligned}
$$

occurs in many, seemingly unrelated, statistical models.

We call this **Latent Gaussian models**.

Other example of LGM

- Generalised linear (mixed) models
- Stochastic volatility
- Generalised additive (mixed) models
- Measurement error models
- Spline smoothing
- Semiparametric regression
- Space-varying (semiparametric) regression models
- Disease mapping
- Log-Gaussian Cox-processes
- Model-based geostatistics (*)
- Spatio-temporal models
- Survival analysis
- +++

Characteristics of LGM

- The latent part of the hierarchical model is Gaussian:

$$\boldsymbol{x}|\boldsymbol{\theta} \sim N(0, Q^{-1}(\theta))$$

- The expected value is $\mathbf{0}$
- The *precision* matrix (inverse covariance matrix) is $Q(\theta)$

The general set-up

The mean of the observation i, μ_i, is connected to the linear predictor, η_i, through a link function g,

$$\eta_i = g(\mu_i) = \mu + \boldsymbol{z}_i^\top \boldsymbol{\beta} + \sum_\gamma w_{\gamma,i} f_\gamma(c_{\gamma,i}) + v_i, \quad i = 1, 2, \ldots, n$$

where

$$\begin{aligned}
\mu &: \text{Intercept} \\
\boldsymbol{\beta} &: \text{Fixed effects of covariates } \boldsymbol{z} \\
\{f_\gamma(\cdot)\} &: \text{Non-linear/smooth effects of covariates } \boldsymbol{c} \\
\{w_{\gamma,i}\} &: \text{Known weights defined for each observed data point} \\
\boldsymbol{v} &: \text{Unstructured error terms}
\end{aligned}$$

Specification of the latent field

- Collect all parameters (random variables) in the latent field $\mathbf{x} = \{\mu, \beta, \{f_\gamma(\cdot)\}, \eta\}$.

Specification of the latent field

- Collect all parameters (random variables) in the latent field $\mathbf{x} = \{\mu, \beta, \{f_\gamma(\cdot)\}, \eta\}$.
- A latent Gaussian model is obtained by assigning Gaussian priors to all elements of $\mathbf{x}$.

Specification of the latent field

- Collect all parameters (random variables) in the latent field $\mathbf{x} = \{\mu, \beta, \{f_\gamma(\cdot)\}, \eta\}$.
- A latent Gaussian model is obtained by assigning Gaussian priors to all elements of $\mathbf{x}$.
- Very flexible due to many different forms of the unknown functions $\{f_\gamma(\cdot)\}$:

Specification of the latent field

- Collect all parameters (random variables) in the latent field $\mathbf{x} = \{\mu, \beta, \{f_\gamma(\cdot)\}, \eta\}$.
- A latent Gaussian model is obtained by assigning Gaussian priors to all elements of $\mathbf{x}$.
- Very flexible due to many different forms of the unknown functions $\{f_\gamma(\cdot)\}$:
- Hyperparameters account for variability and length/strength of dependence

Flexibility through f-functions

The functions $\{f_\gamma\}$ in the linear predictor make it possible to capture very different types of random effects in the same framework:

- $f(\texttt{time})$:, For example, an AR(1) process, RW1 or RW2
- $f(\texttt{spatial location})$:, For example, a Matern field
- $f(\texttt{covariate})$:, For example, a RW1 or RW2 on the covariate values
- $f(\texttt{time}, \texttt{spatial location})$ can be a spatio-temporal effect
- And much more

Additivity

- One of the most useful features of the framework is the additivity.
- Effects can easily be removed and added without difficulty.
- Each component might add a new latent part and might add new hyperparameters, but the modelling framework and computations stay the same.

Additivity

- One of the most useful features of the framework is the additivity.
- Effects can easily be removed and added without difficulty.
- Each component might add a new latent part and might add new hyperparameters, but the modelling framework and computations stay the same.

OBS: The *linear* predictor needs to stay linear!! So effects can be added but not multiplied
Why??

A small point to think about

From a Bayesian point of view fixed effects and random effects are all the same.

- Fixed effects are also random
- They only differ in the prior we put on them

So. . . which model fit the INLA framework??

1. Latent **Gaussian** model
2. The latent field has a sparse precision matrix (Markov properties)
3. The data are conditionally independent given the latent field
4. The predictor is linear

Assume that, given $\eta = (\eta_1, \ldots, \eta_n)$ the observations $y = (y_1, \ldots, y_n)$ are independent and Poisson distributed with parameter $\lambda_i = \exp(\eta_i)$ i.e.

$$y_i|\eta_i = \text{Poisson}(\lambda_i); i = 1, \ldots, n$$

1. $\eta_i = \alpha + \beta x_i + U_i$ where
$$\begin{aligned} \alpha, \beta &\sim \mathcal{N}(0,1) \\ U_i &\sim \mathcal{N}(0,1) \text{ for } i = 1, \ldots, n \end{aligned}$$
2. $\eta_i = \alpha + \beta x_i + V_i$ where
$$\begin{aligned} \alpha, \beta &\sim \mathcal{N}(0,1) \\ V_i &\sim \text{Bernoulli}(0.4) \text{ for } i = 1, \ldots, n \end{aligned}$$
3. $\eta_i = \alpha + \beta x_i$ where
$$\alpha, \beta \sim \mathcal{N}(0,1)$$
4. $\eta_i = \alpha + \beta x_i + U_i V_i$ where
$$\begin{aligned} \alpha, \beta &\sim \mathcal{N}(0,1) \\ U_i &\sim \mathcal{N}(0,1) \text{ for } i = 1, \ldots, n \\ V_i &\sim \mathcal{N}(0,1) \text{ for } i = 1, \ldots, n \end{aligned}$$

Why precision matrix