Integrated Nested Laplace Approximation (INLA)

What is it? A numerical method to do fast approximate bayesian inference

Why? We do not want to way for the MCMC to converge..

Where can it be applied? The (wide) class of Latent Gaussian Models

How does it work? Uses GMRF and sparse matrix computations, Laplace approximation, numerical integration

How do we use it Already implemented in the `R-INLA` library

Latent Gaussian models

A very general way of specifying the problem is by modelling the mean for the i-th unit by means of an additive linear predictor, defined on a suitable scale (e.g. logistic for binomial data)

$$\eta_i = \alpha + \sum_{l=1}^{L} f_l(u_{li}) + \sum_{k=1}^{K} \beta_k z_{ki} + \epsilon_i$$

where

- α is the intercept
- $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)$ quantify the effect of $\boldsymbol{x} = (x_1, \ldots, x_K)$ on the response
- $\boldsymbol{f} = (f_1, \ldots, f_L)$ is a set of functions defined in terms of some covariates $\boldsymbol{z} = (z_1, \ldots, z_K)$

And assume

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

Latent Gaussian Models: a Unified framework

Observations: $\boldsymbol{y}$ Assumed conditionally independent given $\boldsymbol{x}$ and $\boldsymbol{\theta}_1$

$$\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}_1 \sim \prod_i \pi(y_i|x_i, , \boldsymbol{\theta}).$$

Latent field: $\boldsymbol{x}$ Assumed to be a GMRF with sparse precision matrix $\boldsymbol{Q}(\boldsymbol{\theta}_2)$

$$\boldsymbol{x}|\boldsymbol{\theta}_1 \sim \mathcal{N}(0, \boldsymbol{Q}(\boldsymbol{\theta}_2)^{-1})$$

The latent field $\boldsymbol{x}$ can be large $(10^1 - 10^6)$

Hyperparameters: $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ Precision parameters of the Gaussian field and parameters of the likelihood

$$\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$$

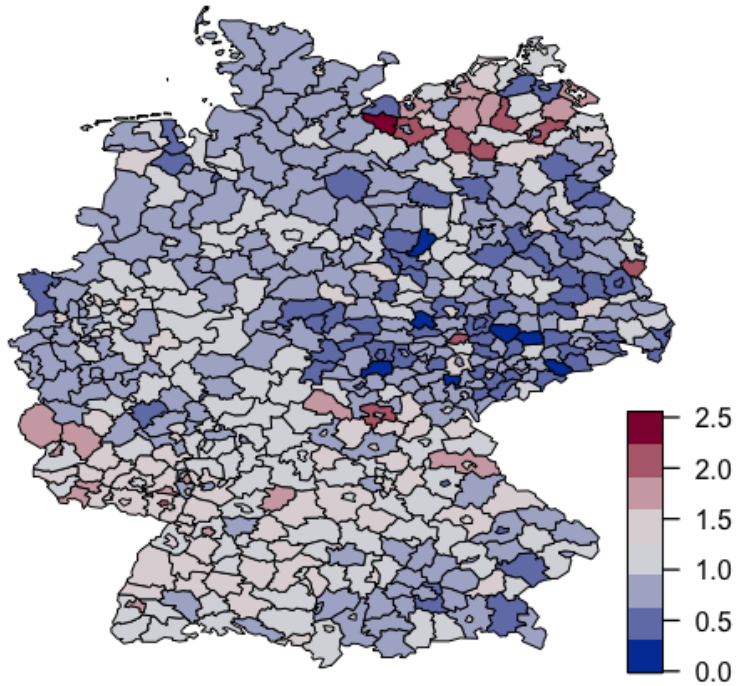
The vector $\boldsymbol{\theta}$ is usually small (1-10)

Example: Disease Mapping in Germany

We observed larynx cancer mortality counts for males in 544 district of Germany from 1986 to 1990 and want to make a model.

Information available:

- y_i The count in disctrict i
- E_i An offset, expected number of cases in district i
- c_i A covariate (level of smoking consumption in district i)
- s_i Spatial location i (district)



Example: Disease Mapping in Germany

- Poisson likelihood

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

- Laten Gaussian model

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

  The latent field is $\boldsymbol{x} = \{\mu, (f_s(\cdot)), (f(\cdot)), u_1, \ldots, u_n\}$

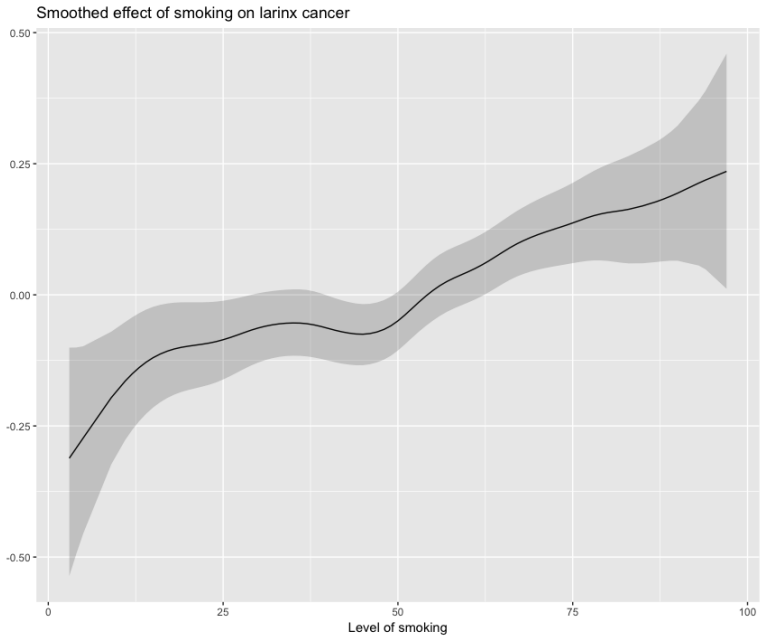
- Hyperparameters: τ_c , τ_f , τ_η : The precisions (inverse variances) of the covariate effect, spatial effect and unstructured effect, respectively.

Example: Disease Mapping in Germany

Posterior of interest

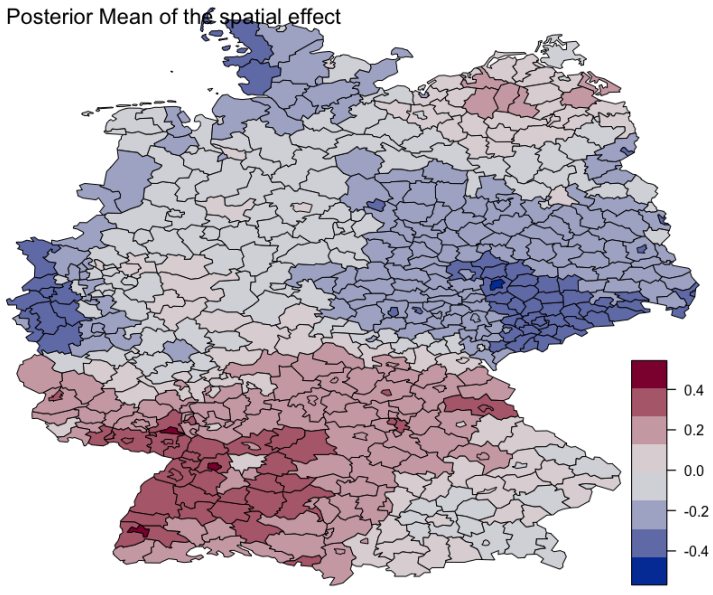
Effect of the covariate:

$$\pi(f(c_i)|\boldsymbol{y})$$



Structured spatial effect:

$$\pi(f_s(s_i)|\boldsymbol{y})$$



INLA computing scheme

From the posterior $\pi(\boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y})$ we are mostly interested in

$$\pi(\theta_j|\boldsymbol{y}) \text{ and } \pi(x_i|\boldsymbol{y})$$

- Approximate $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ using Laplace approximation
  - Use numerical integration to approximate

$$\pi(\theta_j|\boldsymbol{y}) = \int \pi(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta}_{-j}$$

  - This integral is not difficult to solve (dimension of $\boldsymbol{\theta}$ is small)
- Approximate $\pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})$ using Laplace approximation
  - Use numerical integration to approximate

$$\pi(x_i|\boldsymbol{y}) = \int \pi(x_i|\boldsymbol{\theta}, \boldsymbol{y}) d\boldsymbol{\theta}$$

  - This integral is not difficult to solve (dimension of $\boldsymbol{\theta}$ is small)

Smoothing noisy observations

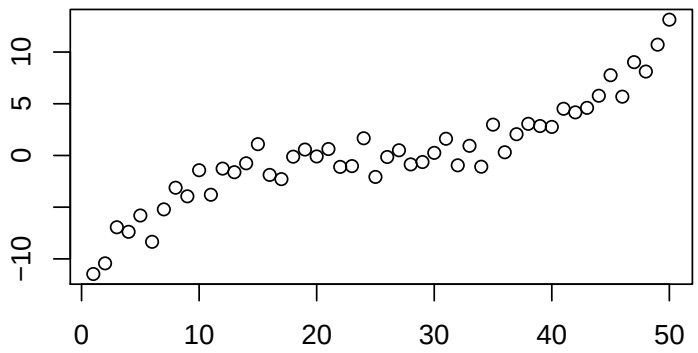
Assume

$$y_i = f(i) + \epsilon_i$$

where

$$\epsilon_i \sim \mathcal{N}(0, 1)$$

$f(i)$ smooth function of i



We have noisy observation, we want to recover the f function

Hierarchical Model

Data Gaussian Observations with known precision

$$y_i|x_i \sim \mathcal{N}(x_i, 1)$$

Latent Model : A Gaussian model for the smooth function (RW2 model)

$$\pi(\mathbf{x}|\theta) \propto \theta^{(n-2)/n} \exp\left\{-\frac{\theta}{2}\sum_{i=2}^{n}(x_i - 2x_{i-1} + x_{i-2})^2\right\}$$

Hyperparameter The precision of the smooth function θ. We assign a Gamma prior

$$\pi(\theta) \propto \theta^{a-1} \exp(-b\theta)$$

Posterior marginal for hyperparameter

We have that

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

so

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

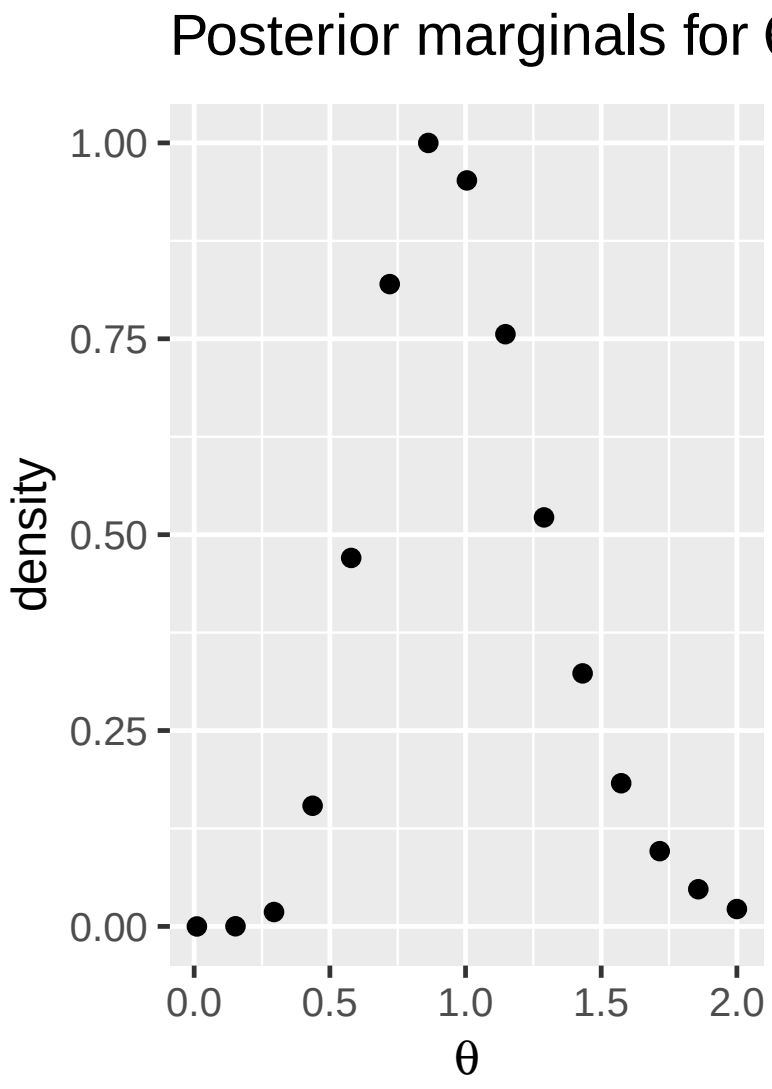
Since the likelihood is Gaussian, then $\pi(\mathbf{y}, \mathbf{x}|\theta)$ is also Gaussian. We have then:

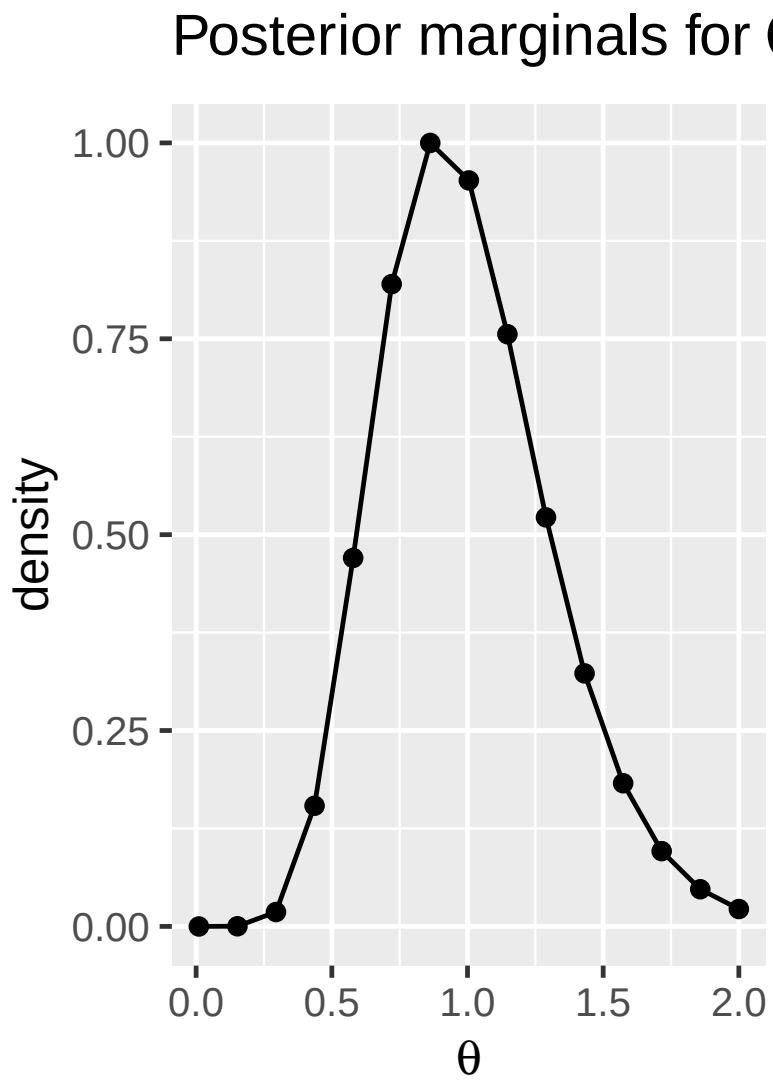
$$\pi(\theta|\mathbf{y}) \propto \frac{\overbrace{\pi(\mathbf{y}, \mathbf{x}|\theta)}^{\text{Gaussian}}\ \pi(\theta)}{\underbrace{\pi(\mathbf{x}|\theta, \mathbf{y})}_{\text{Gaussian}}}$$

This is valid for any $\mathbf{x}$

Posterior marginal for the hyperparameter

Select a grid of points to represent the density $\pi(\theta|\mathbf{x})$





Posterior marginals for latent field

Again we have that

$$\mathbf{x}, \mathbf{y}|\theta \sim \mathbf{N}(\cdot, \cdot)$$

so also $\pi(x_i|\theta, \mathbf{y})$ is Gaussian!!

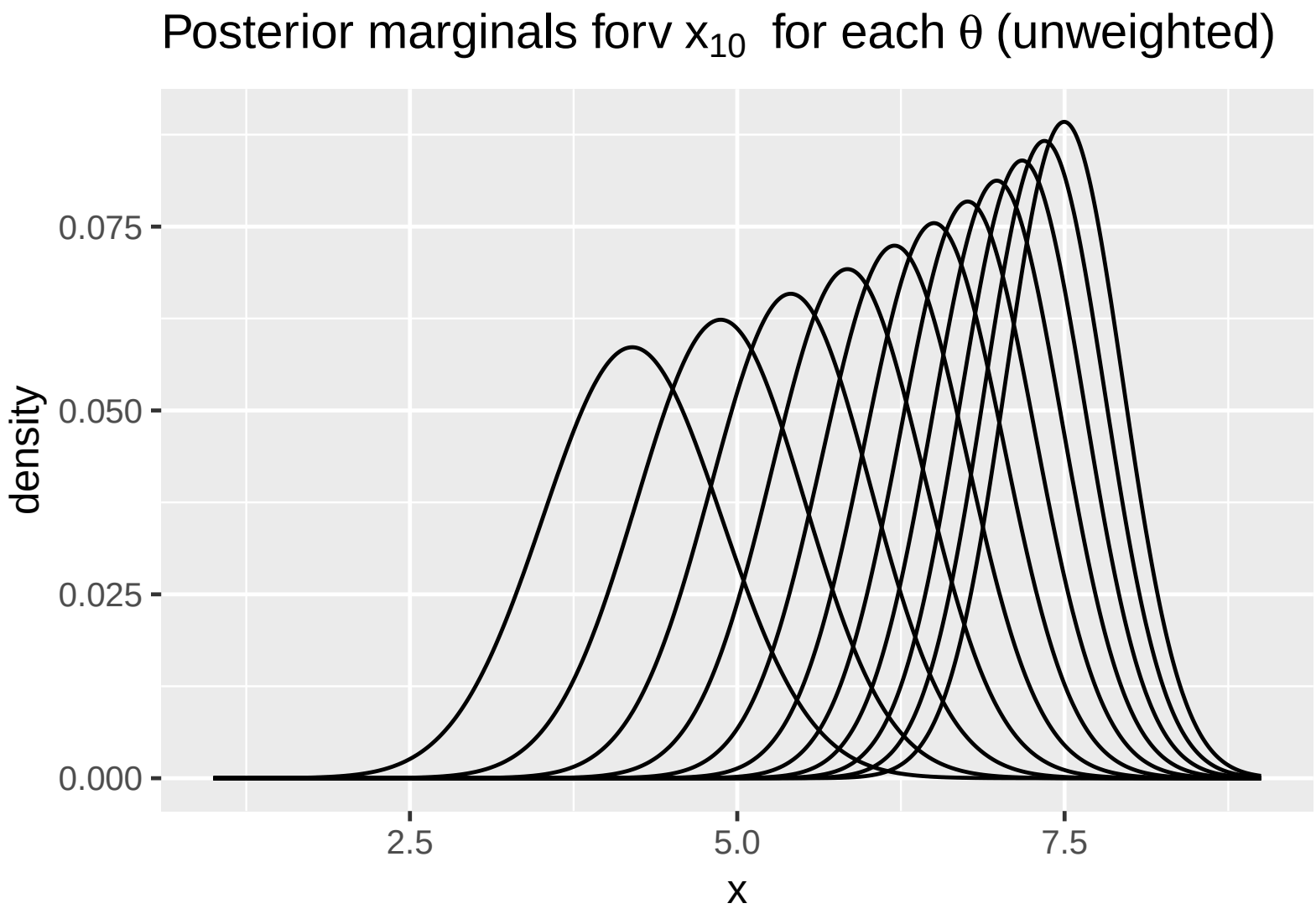
We compute

$$\begin{aligned} \pi(x_i|\mathbf{y}) &= \int \pi(x_i|\theta, \mathbf{y})\pi(\theta|\mathbf{y})d\theta \\ &\approx \sum_k \pi(x_i|\theta_k, \mathbf{y})\pi(\theta_k|\mathbf{y})\Delta_k \end{aligned}$$

where $\theta_k, k = 1, \ldots, K$ are the representative points of $\pi(\theta|\mathbf{y})$ and Δ_k are the corresponding weights

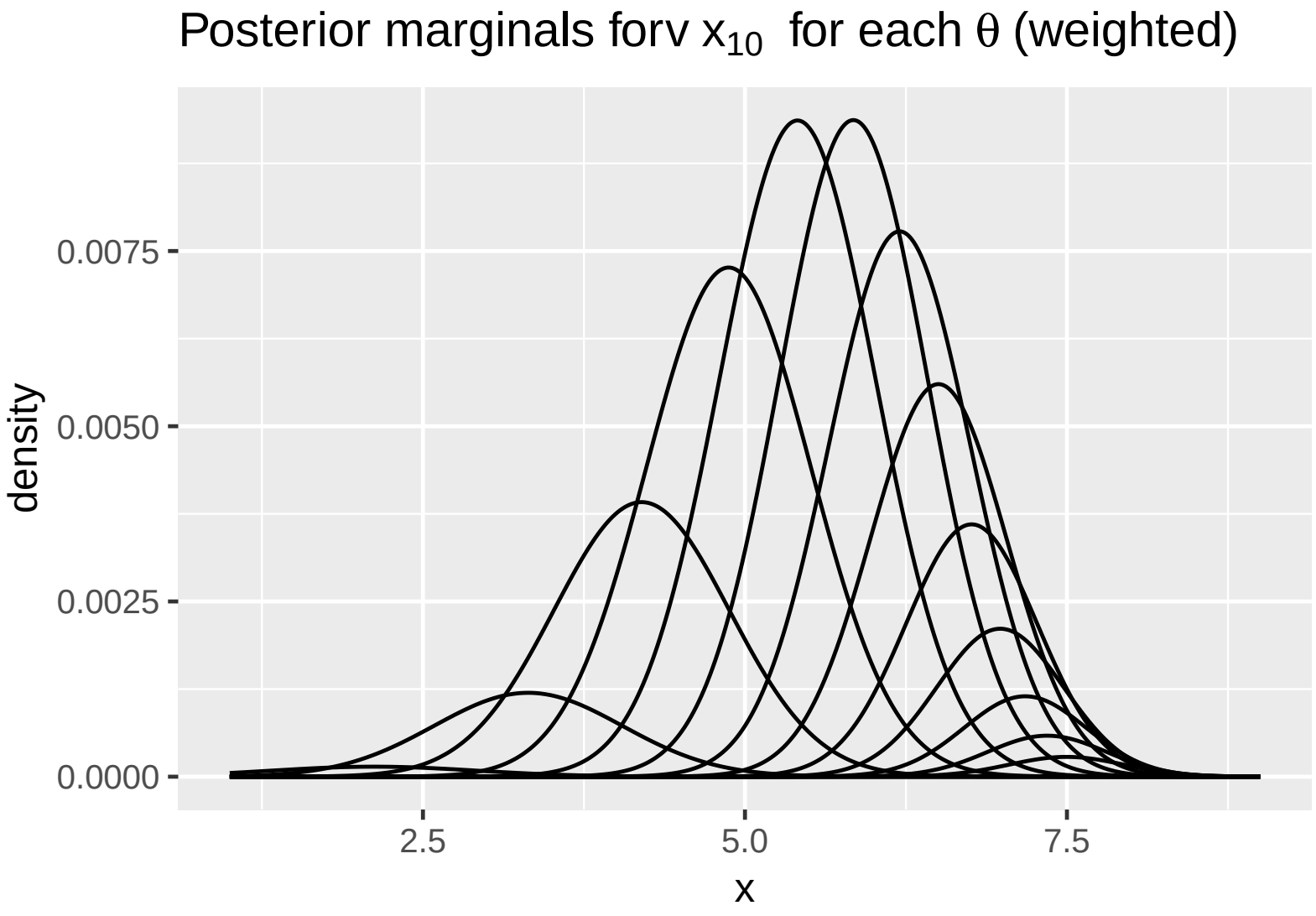
Posterior marginals for latent field

Compute the conditional posterior marginal for x_i given each θ_k



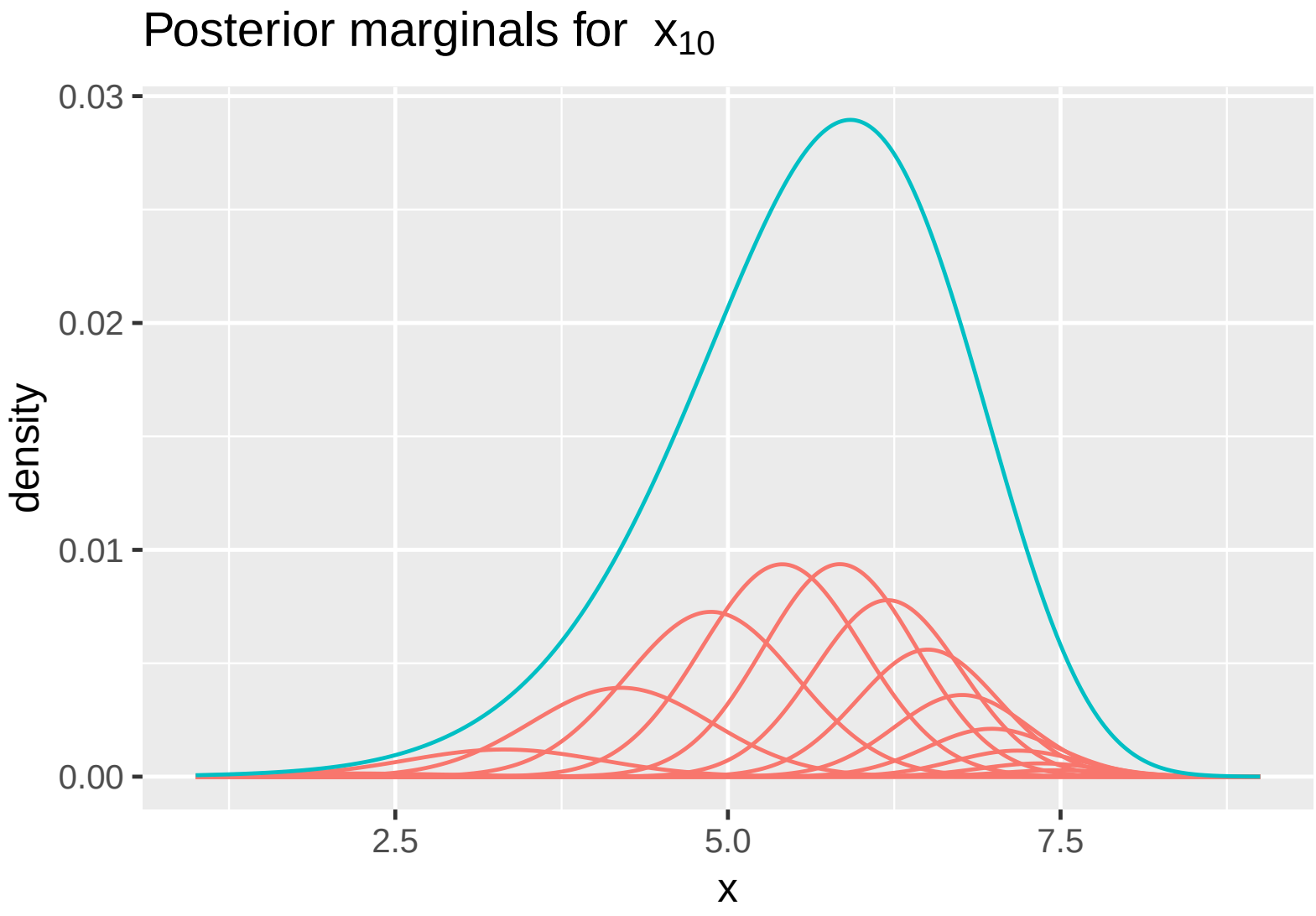
Posterior marginals for latent field

Weighted the conditional posterior marginal for $\pi(x_i|\theta_k, \mathbf{y})$ by $\pi(\theta_k|\mathbf{y})$´$_\mathbf{k}$



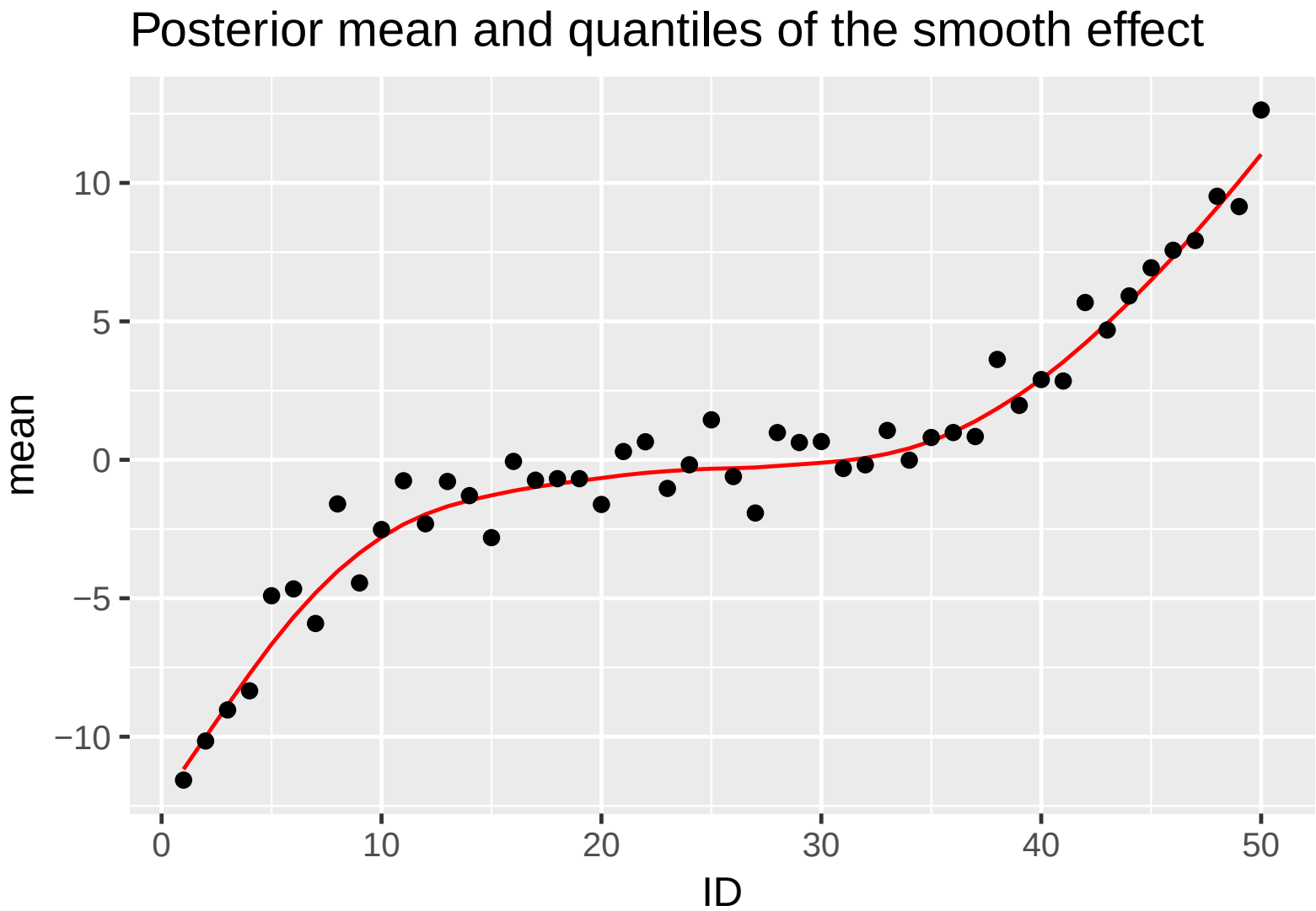
Posterior marginals for latent field

Sum to get the posterior marginal for $x_i|\mathbf{y}$



Fitted Spline

The posterior marginals are used to calculate summary statistics, like means, variances and credible intervals:



Extending the method

This is the basic idea behind INLA. It is quite simple. However, we need to extend this basic idea so we can deal with

- More than one hyperparameter
- Non-Gaussian observations

Non-Gaussian Observations: Approximating $\pi(\boldsymbol{x}|)\theta\mathbf{y}$

Let $\boldsymbol{x}$ denote a GMRF with precision matrix $\boldsymbol{Q}$ and mean $\boldsymbol{\mu}$. Approximate

$$\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y}) \propto \exp\left(-\frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \sum_{i=1}^{n} \log \pi(y_i|x_i)\right)$$

by using a second-order Taylor expansion of $\log \pi(y_i|x_i)$ around $\boldsymbol{\mu}_0$, say.

Recall

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2 = a + bx - \frac{1}{2}cx^2$$

with $b = f'(x_0) - f''(x_0)x_0$ and $c = -f''(x_0)$.

The GMRF approximation (II)

Thus,

$$\tilde{\pi}(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y}) \propto \exp\left(-\frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \sum_{i=1}^{n}(a_i + b_i x_i - 0.5 c_i x_i^2)\right)$$
$$\propto \exp\left(-\frac{1}{2}\boldsymbol{x}^T(\boldsymbol{Q} + \text{diag}(\boldsymbol{c}))\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x}\right)$$

to get a Gaussian approximation with precision matrix $\boldsymbol{Q} + \text{diag}(\boldsymbol{c})$ and mean given by the solution of $(\boldsymbol{Q} + \text{diag}(\boldsymbol{c}))\boldsymbol{\mu} = \boldsymbol{b}$. The canonical parameterization is

$$\mathcal{N}_C(\mathbf{b}, \boldsymbol{Q} + \text{diag}(\boldsymbol{c}))$$

which corresponds to

$$\mathcal{N}((\boldsymbol{Q} + \text{diag}(\boldsymbol{c}))^{-1}\mathbf{b}, (\boldsymbol{Q} + \text{diag}(\boldsymbol{c}))^{-1}).$$

The GMFR approximation - One dimensional example

Assume

$$y|\lambda \sim \text{Poisson}(\lambda) \text{ Likelihood}$$
$$\lambda = \exp(x) \text{ Likelihood}$$
$$x \sim \mathcal{N}(0, 1) \text{ Latent Model}$$

we have that

$$\pi(x|y) \propto \pi(y|x)\pi(x) \propto \exp\{-\frac{1}{2}x^2 + \underbrace{xy - exp(x)}_{\text{non-gaussian part}}\}$$

(Show R-code `Taylor_expansion.R`)

Non-Gaussian Observations

In many cases $\pi(\boldsymbol{x}|\boldsymbol{y}, \theta)$ is very close to a Gaussian distribution, and can be replaced with a Laplace approximation:

- This means that all the really hard, high-dimensional integrals with respect to the latent field are easy, and only the integrals with respect to the hyperparameters remain
- If the number of hyperparameters is low, these integrals can be done efficiently numerically

Limitations

- The dimension of the latent field x can be large ($10^2 - 10^6$)
- The dimension of the hyperparameters θ must be small (≤ 9)

In other words, each random effect can be big, but there cannot be too many random effects unless they share parameters.

Gaussian Markov Random Fields

A GMRF $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ is a random vector following a multivariate Gaussian distribution

$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}^{-1}) \text{ where } \boldsymbol{Q}^{-1} = \boldsymbol{\Sigma}$$

and that is endowed with some Markov properties like

$$x_j \perp x_i | \boldsymbol{x}_{-ij}$$

where $\boldsymbol{x}_{-ij}$ indicates "all elements of $\boldsymbol{x}$ other than i and j"

The easiest example is a AR(1) model

Gaussian Markov Random Fields

If $\boldsymbol{\Sigma}$ is the covariance matrix of a Gaussian vector and $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$ is the precision matrix, we have that

$$x_i \perp x_j \iff \Sigma_{ij} = 0$$

and

$$x_i \perp x_j \iff Q_{ij} = 0$$

GMRF have sparse precision matrices....this means it is "easy" to compute determinant and invert $\boldsymbol{Q}$