Bayesian inference using the integrated nested Laplace approximation (INLA)

Master's Degree in Biostatistics

Joaquín Martínez-Minaya, 2021-11-10
Universitat Politècnica de València
https://smeg-bayes.org/
http://vabar.es/
jmarmin@upv.es

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Outline

1. Why INLA?
2. Elements to understand how INLA works
3. Putting all the pieces together: INLA
4. `R-INLA`
5. Model Selection
6. Examples
7. References

1. Why INLA?

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

INLA as an alternative to MCMC

- MCMC is an asymptotically exact method whereas INLA is an **approximation**. Their error are frequently very similar, as has been shown in many simulation studies.

- INLA is a **fast alternative** to MCMC for the general class of latent Gaussian models (LGMs). Many familiar models can be re-cast to look like LGMs:

  - **generalized linear models**, **generalized additive models**, smoothing spline models,
  - state space models, semi-parametric regression, **random walk (first and second order)** models, longitudinal data models,
  - **spatial and spatiotemporal** models, log-Gaussian Cox processes and geostatistical and geoadditive models., etc.
- To understand INLA, we need to be familiar with:
  - Latent Gaussian models
  - Gaussian Markov Random Fields (GMRFs)
  - Laplace approximations

2. Elements to understand how INLA works

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Latent Gaussian model

Level 1 : likelihood

The first stage is formed by the **conditionally independent likelihood** function of data coming from a certain exponential family distribution:

$$p(\boldsymbol{y} \mid \boldsymbol{\theta}, \boldsymbol{\psi}_1) = \prod_{i=1}^{n} p(y_i \mid \eta_i(\boldsymbol{\theta}), \boldsymbol{\psi}_1)$$

- $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ is the response vector, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^T$ is the **latent field**,
- $\boldsymbol{\psi}_1$ is the hyperparameter vector of the exponential family distribution and
- $\eta_i(\boldsymbol{\theta})$ is the $\boldsymbol{i}$-th linear predictor that connects the data to the latent field.

Indeed each $\boldsymbol{\eta}_i$ can take a more general additive form:

$$\eta_i = \beta_0 + \sum_{j=1}^{J} \beta_k x_{ij} + \sum_{k=1}^{K} f^{(k)}(z_{ik})$$

Latent Gaussian model

Level 2: latent Gaussian field

- The second stage is formed by the **latent Gaussian field**, where we attribute a Gaussian distribution with mean $\boldsymbol{\mu}$ and precision matrix $Q(\boldsymbol{\psi}_2)$ to the latent field $\boldsymbol{\theta}$ conditioned on the hyperparameters $\boldsymbol{\psi}_2$, that is:

$$\boldsymbol{\theta} \mid \boldsymbol{\psi}_2 \sim \mathcal{N}(\mathbf{0}, Q^{-1}(\boldsymbol{\psi}_2))$$

- If we can assume conditional independence in $\boldsymbol{\theta}$, then this latent field is a **Gaussian Markov Random Field (GMRF)**.

Level 3: hyperparameters

- Finally, the third stage is formed by the **prior distribution** assigned to the hyperparameters:

$$\boldsymbol{\psi} = (\boldsymbol{\psi}_1, \boldsymbol{\psi}_2) \sim p(\boldsymbol{\psi})$$

GLMM

- Breslow and Clayton (1993) present a dataset where they account for the proportion of seeds that germinated on each of 21 plates arranged according to a 2 by 2 factorial layout by seed and type of root extract.

The variables are:

- **r**: number of germinated seeds per plate
- **n**: number of total seeds per plate
- **x1**: seed type (0: seed O. aegyptiaco 75; 1: seed O. aegyptiaco 73)
- **x2**: root extracted (0: bean; 1: cucumber)
- **plate**: indicator for the plate This dataset is located in the package **INLA**

r	n	x1	x2	plate
10	39	0	0	1
23	62	0	0	2
23	81	0	0	3
26	51	0	0	4
17	39	0	0	5
5	6	0	1	6

Example: mixed-effects model

- We assume the counts follow a conditionally independent Binomial likelihood function:

$$y_i \mid \pi_i \sim \text{Binomial}(n_i, \pi_i),\ i = 1, \ldots, 21$$

- We account for linear effects on **covariates** $x1_i$ and $x2_i$ for each individual, as well as a **random effect** on the individual level, the plate b_i.

$$\begin{aligned}
\eta_i &= \text{logit}(\pi_i) = \beta_0 + \beta_1 x1_i + \beta_2 x2_i + b_i \\
\beta_j &\sim \mathcal{N}(0, \tau_\beta^{-1}),\ \tau_\beta \text{ known},\ j = 0, 1, 2 \\
b_i &\sim \mathcal{N}(0, \tau_b^{-1})
\end{aligned}$$

So, in this case, $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, b_1, \ldots, b_{21})$. A Gaussian prior is assigned for each element of the **latent field**, so that $\boldsymbol{\theta} \mid \boldsymbol{\psi}$ is **Gaussian distributed**.

- To assign the prior of $\boldsymbol{\psi} = (\tau_b)$:

$$\log(\tau_b) \sim \text{logGamma}(1, 5 \cdot 10^{-5})$$

Gaussian Markov Random Fields (GMRFs)

- A GMRF is a random vector following a **multivariate normal distribution** with Markov properties.

$$\boldsymbol{i} \neq \boldsymbol{j},\ \boldsymbol{\theta_i} \mid \boldsymbol{\theta_{ij}},$$

  being $-\boldsymbol{ij}$ all elements other than $\boldsymbol{i}$ and $\boldsymbol{j}$.

- Rue et al. (2009) showed how conditional independence properties are encoded in the precision matrix, and how this can be exploited to improve computation involving these matrices.

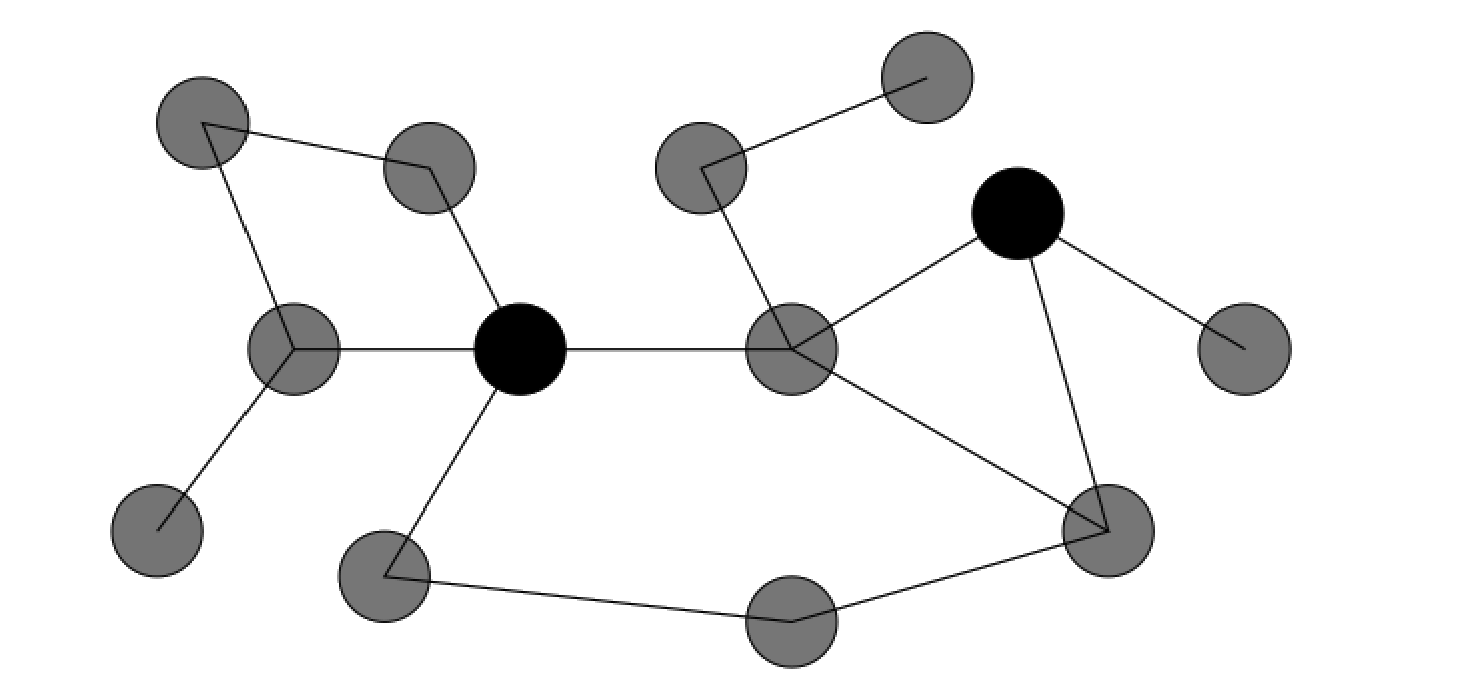
$$\boldsymbol{i} \neq \boldsymbol{j},\ \boldsymbol{\theta_i} \perp \boldsymbol{\theta_j} \mid \boldsymbol{\theta_{ij}},$$

$$\theta_i \perp \theta_j \mid \boldsymbol{\theta}_{ij} \leftrightarrow \boldsymbol{Q}_{ij} = 0$$

- This Markov assumption in the GMRF results in a **sparse precision matrix**. This sparseness aids extremely fast computation.

The pairwise Markov property

The two black nodes are conditionally independent given the gray nodes



Example: precision matrix in AR1

Covariance matrix (Σ)

0.8730	0.6957	0.5201	0.3460	0.1728
0.6957	1.3931	1.0417	0.6929	0.3460
0.5201	1.0417	1.5659	1.0417	0.5201
0.3460	0.6929	1.0417	1.3931	0.6957
0.1728	0.3460	0.5201	0.6957	0.8730

Precision matrix (Q)

1.9025	-0.9500	0.0000	0.0000	0.0000
-0.9500	1.9025	-0.9500	0.0000	0.0000
0.0000	-0.9500	1.9025	-0.9500	0.0000
0.0000	0.0000	-0.9500	1.9025	-0.9500
0.0000	0.0000	0.0000	-0.9500	1.9025

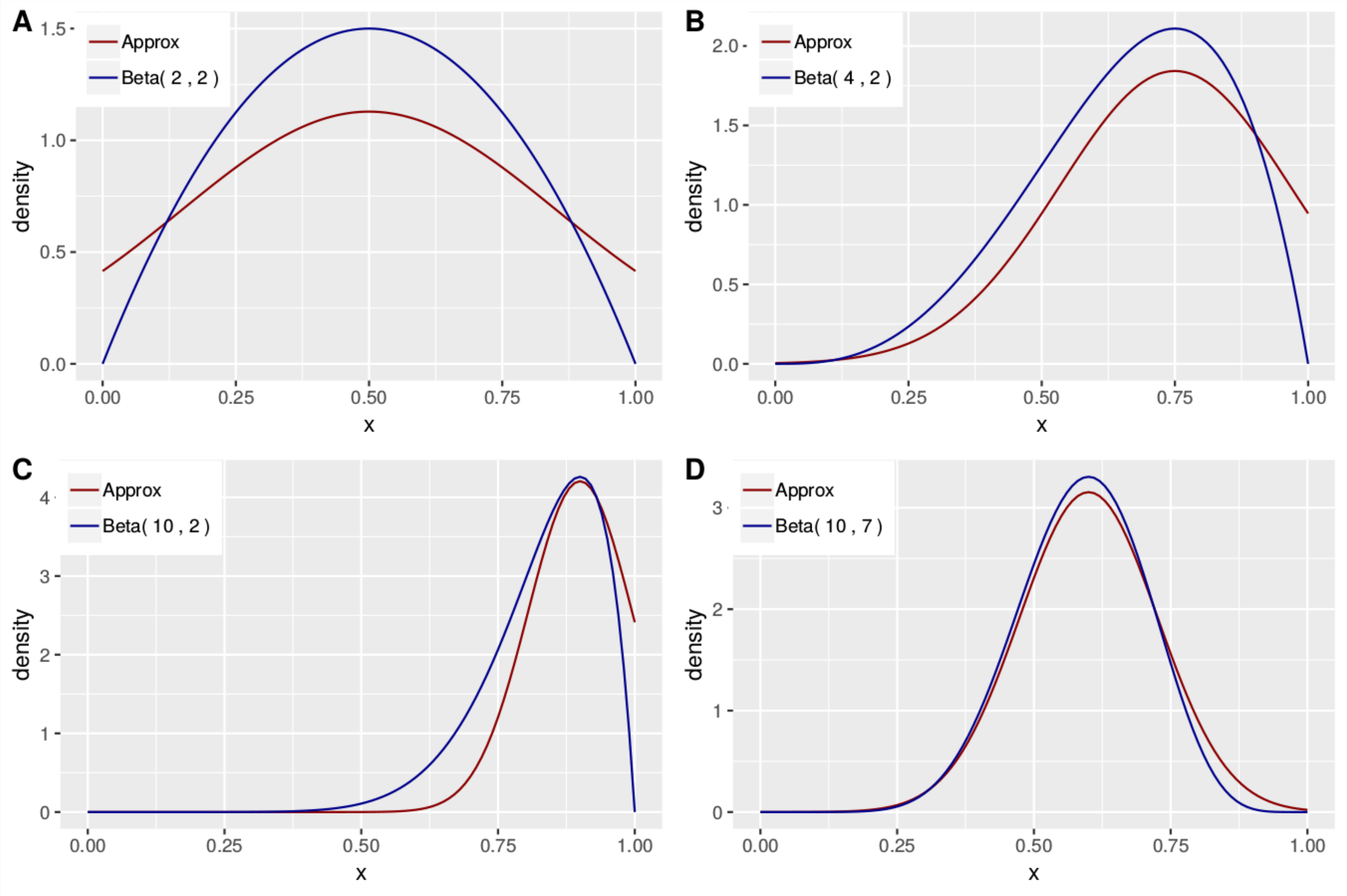
Laplace approximations

- **The Laplace approximation** is used to estimate any distribution $p(\theta)$ with a normal distribution.
- It uses the first three terms (quadratic function) **Taylor series expansion** around the mode θ^* of a function to approximate its log.
- Using the approximation, $p(\theta)$ can be approximated using a **Gaussian distribution** with mean the mode θ^* and variance the Fisher information, $\frac{-1}{\frac{d^2 \log(p(\theta^*))}{d\theta^2}}$.

$$p(\theta) \approx \mathcal{N}\left(\theta^*, \frac{-1}{\frac{d^2 \log(p(\theta^*))}{d\theta^2}}\right)$$

- It can be easily expanded to the multivariate case.

Example: approximating the beta distribution



3. Putting all the pieces together: INLA

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

INLA: Aim

Marginals of the latent field and hyperparameters

$$p(\theta_i \mid \boldsymbol{y}) = \int p(\theta_i \mid \boldsymbol{\psi}, \boldsymbol{y}) \cdot p(\boldsymbol{\psi} \mid \boldsymbol{y}) d\boldsymbol{\psi} \ , \ i = 1, \ldots, n$$

$$p(\psi_j \mid \boldsymbol{y}) = \int p(\boldsymbol{\psi} \mid \boldsymbol{y}) d\boldsymbol{\psi}_{-j} \, , \ j = 1, \ldots, m$$

- As a result, we have to numerically approximate:

1. The **joint posterior distribution of the hyperparmeters** $p(\boldsymbol{\psi} \mid \boldsymbol{y})$, needed to calculate the posterior hyperparameters marginals $p(\psi_j \mid \boldsymbol{y})$, and the posterior marginals of the latent field $p(\theta_i \mid \boldsymbol{y})$.

2. The **marginals of the full conditional distribution** of $\boldsymbol{\theta}$, $p(\theta_i \mid \boldsymbol{\psi}, \boldsymbol{y})$, needed to compute the posterior marginals of the latent field $p(\theta_i \mid \boldsymbol{y})$.

Hyperparameters: joint posterior distribution

- The approximation is computed as follows

$$\tilde{p}(\boldsymbol{\psi} \mid \boldsymbol{y}) := \left. \frac{p(\boldsymbol{\theta}, \boldsymbol{\psi} | \boldsymbol{y})}{p_G(\boldsymbol{\theta} \mid \boldsymbol{\psi}, \boldsymbol{y})} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*(\boldsymbol{\psi})},$$

- where:
    - $p_G(\boldsymbol{\theta} \mid \boldsymbol{\psi}, \boldsymbol{y})$ is the Gaussian approximation to the full conditional of $\boldsymbol{\theta}$, $p(\boldsymbol{\theta} \mid \boldsymbol{\psi}, \boldsymbol{y})$ given by the **Laplace method**, and,
    - $\boldsymbol{\theta}^*(\boldsymbol{\psi})$ is the mode of the full conditional of $\boldsymbol{\theta}$ for a given $\boldsymbol{\psi}$.
    - Note: this approximation is exact if $p(\boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{\psi})$ is Gaussian.

Full posterior marginals for the latent field

Gaussian approximation

- Conditional posterior distributions $p(\theta_i \mid \boldsymbol{\psi}, \boldsymbol{y})$ are approximated directly as the marginals from $p_G(\boldsymbol{\theta} \mid \boldsymbol{\psi}, \boldsymbol{y})$.
- It is the **fastest to compute** but with possible **errors** in the location of the posterior mean.

Laplace approximation

- The vector $\boldsymbol{\theta}$ is rewriten as $\boldsymbol{\theta} = (\theta_i, \boldsymbol{\theta}_{-i})$, and the Laplace approximation is used for each element of the latent field

$$\tilde{p}(\theta_i \mid \boldsymbol{\psi}, \boldsymbol{y}) := \left. \frac{p(\boldsymbol{\theta}, \boldsymbol{\psi} | \boldsymbol{y})}{p_{LG}(\boldsymbol{\theta}_{-i} \mid \theta_i, \boldsymbol{\psi}, \boldsymbol{y})} \right|_{\boldsymbol{\theta}_{-i} = \boldsymbol{\theta}^*_{-i}(\theta_i, \boldsymbol{\psi})},$$

where $p_{LG}(\boldsymbol{\theta}_{-i} \mid \theta_i, \boldsymbol{\psi}, \boldsymbol{y})$ is the Laplace Gaussian approximation to $p(\boldsymbol{\theta}_{-i} \mid \theta_i, \boldsymbol{\psi}, \boldsymbol{y})$ and $\boldsymbol{\theta}_{-i}$ is its mode.

- The **most accurate** but **time consuming**.

Full posterior marginals for the latent field

Simplified Laplace approximation

- Based on a Taylor's series expansion of third order.
- **Fast to compute** and usually **accurate enough**.

Final step: integration

- The INLA algorithm uses Newton-like methods to explore the joint posterior distribution for the hyperparameters $\tilde{p}(\boldsymbol{\psi}|\boldsymbol{y})$ to find **suitable points** for the numerical integration.
- Posterior marginals for the **latent variables** $\tilde{p}(\theta_i|\boldsymbol{y})$ are then computed via numerical integration as:

$$\tilde{p}(\theta_i \mid \boldsymbol{y}) = \int \tilde{p}(\theta_i \mid \boldsymbol{\psi}, \boldsymbol{y})\tilde{p}(\boldsymbol{\psi} \mid \boldsymbol{y})\mathrm{d}\boldsymbol{\psi} \approx \sum_{k=1}^{K} \tilde{p}(\theta_i \mid \boldsymbol{\psi}^{(k)}, \boldsymbol{y})\tilde{p}(\boldsymbol{\psi}^{(k)} \mid \boldsymbol{y})\Delta_k$$

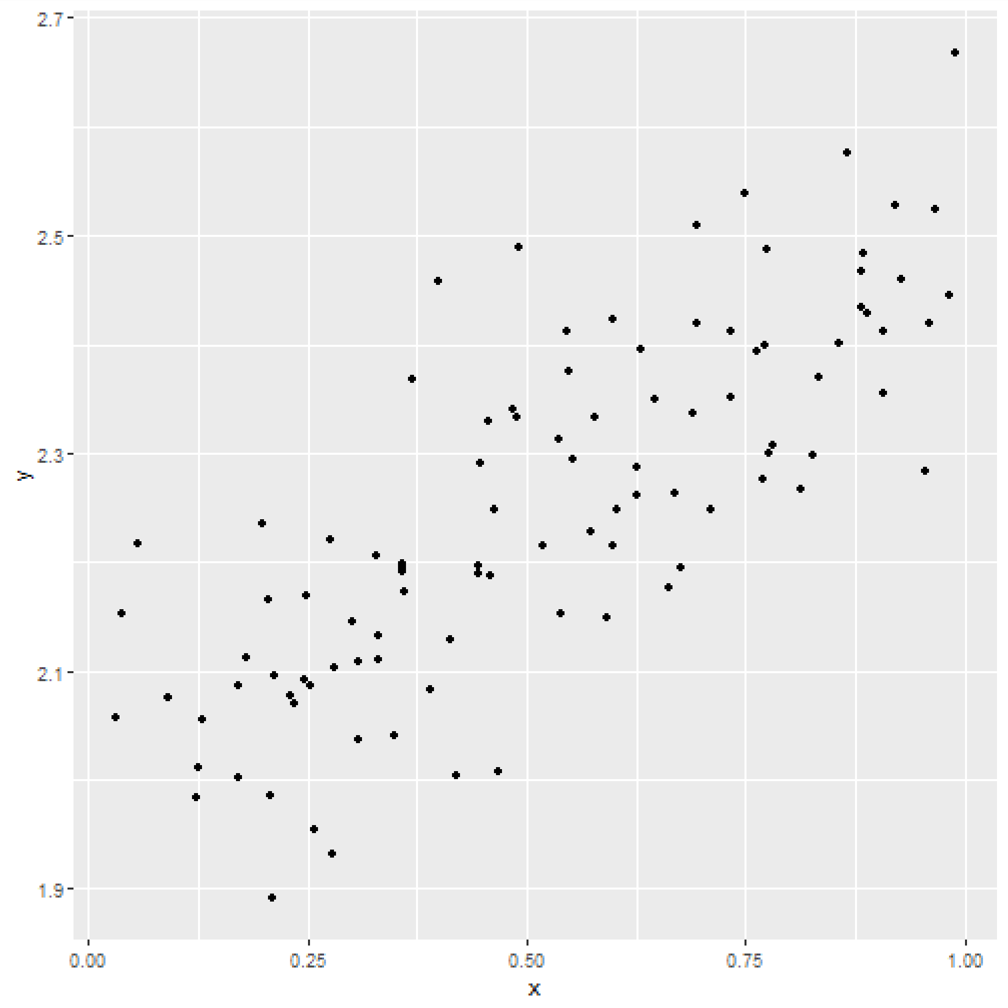
- Posterior marginals for the **hyperparameters** ψ_j are approximated using the integrations points previously constructed.

4. R-INLA

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Data

y	x	id
2.109177	0.3077661	1
1.954976	0.2576725	2
2.294048	0.5523224	3
2.217938	0.0563832	4
2.007082	0.4685493	5
2.339932	0.4837707	6



Fitting the model using `R-INLA`

Defining the formula

```r
formula <- y ~ 1 + x # 1 is refered to the intercept term
formula <- y ~ 1 + f(x, model = "linear")
```

Calling `R-INLA`

```r
model1 <- inla(formula,
               family       = 'gaussian',
               data         = data,
               control.inla = list(strategy = 'simplified.laplace'))
```

Posterior distributions

Posterior distribution of the parameters

	mean	sd	0.025quant	0.5quant	0.975quant	mode	kld
(Intercept)	1.9946	0.0226	1.9502	1.9946	2.0389	1.9946	0
x	0.4935	0.0388	0.4172	0.4935	0.5698	0.4935	0

Posterior distributions of the hyperparameters

	mean	sd	0.025quant	0.5quant	0.975quant	mode
Precision for the Gaussian observations	99.4865	13.9906	73.9147	98.8369	128.7726	97.5136

Families

```r
inla.list.models(section = "likelihood")
```

```
## Section [likelihood]
##       agaussian                     The aggregated Gaussian likelihoood
##       beta                          The Beta likelihood
##       betabinomial                  The Beta-Binomial likelihood
##       betabinomialna                The Beta-Binomial Normal approximation likelihood
##       bgev                          The blended Generalized Extreme Value likelihood
##       binomial                      The Binomial likelihood
##       cbinomial                     The clustered Binomial likelihood
##       cenpoisson                    Then censored Poisson likelihood
##       cenpoisson2                   Then censored Poisson likelihood (version 2)
##       circularnormal                The circular Gaussian likelihoood
##       coxph                         Cox-proportional hazard likelihood
##       dgp                           Discrete generalized Pareto likelihood
##       exponential                   The Exponential likelihood
##       exponentialsurv               The Exponential likelihood (survival)
##       fmri                          fmri distribution (special nc-chi)
```

Latent effects

```r
inla.list.models(section = "latent")
```

```
## Section [latent]
##       2diid                          (This model is obsolute)
##       ar                             Auto-regressive model of order p (AR(p))
##       ar1                            Auto-regressive model of order 1 (AR(1))
##       ar1c                           Auto-regressive model of order 1 w/covariates
##       besag                          The Besag area model (CAR-model)
##       besag2                         The shared Besag model
##       besagproper                    A proper version of the Besag model
##       besagproper2                   An alternative proper version of the Besag model
##       bym                            The BYM-model (Besag-York-Mollier model)
##       bym2                           The BYM-model with the PC priors
##       clinear                        Constrained linear effect
##       copy                           Create a copy of a model component
##       crw2                           Exact solution to the random walk of order 2
##       dmatern                        Dense Matern field
##       fgn                            Fractional Gaussian noise model
```

Hyperpriors

```r
inla.list.models(section = "prior")
```

```
## Section [prior]
##       betacorrelation                 Beta prior for the correlation
##       dirichlet                       Dirichlet prior
##       expression:                     A generic prior defined using expressions
##       flat                            A constant prior
##       gamma                           Gamma prior
##       gaussian                        Gaussian prior
##       invalid                         Void prior
##       jeffreystdf                     Jeffreys prior for the doc
##       linksnintercept                 Skew-normal-link intercept-prior
##       logflat                         A constant prior for log(theta)
##       loggamma                        Log-Gamma prior
##       logiflat                        A constant prior for log(1/theta)
##       logitbeta                       Logit prior for a probability
##       logtgaussian                    Truncated Gaussian prior
##       logtnormal                      Truncated Normal prior
```

5. Model Selection

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Model selection scores in `R-INLA`

- When use different covariates and random effects, we need some measures to select the best model:
  - **DIC**: deviance information criteria

$$DIC = 2 * E(D(\boldsymbol{\theta})) - D(E(\boldsymbol{\theta}))$$

  - **WAIC**: within-sample predictive score

$$WAIC = \sum_i var_{post}(log(p(y_i|\boldsymbol{\theta})))$$

  - **LCPO**: leave-one-out cross-validation score

$$CPO_i = p(y_i \mid y_{-i})$$

$$LCPO = -\overline{\log(CPO_i)}$$

6. Examples

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

GLMM

- Breslow and Clayton (1993) present a dataset where they account for the proportion of seeds that germinated on each of 21 plates arranged according to a 2 by 2 factorial layout by seed and type of root extract.

The variables are:

- **r**: number of germinated seeds per plate
- **n**: number of total seeds per plate
- **x1**: seed type (0: seed O. aegyptiaco 75; 1: seed O. aegyptiaco 73)
- **x2**: root extracted (0:bean; 1:cucumber);
- **plate**: indicator for the plate This dataset is located in the package **INLA**

r	n	x1	x2	plate
10	39	0	0	1
23	62	0	0	2
23	81	0	0	3
26	51	0	0	4
17	39	0	0	5
5	6	0	1	6

Example: mixed-effects model

- We assume the counts follow a conditionally independent Binomial likelihood function:

$$y_i \mid \pi_i \sim \text{Binomial}(n_i, \pi_i),\ i = 1, \dots, 21$$

- We account for linear effects on **covariates** $x1_i$ and $x2_i$ for each individual, as well as a **random effect** on the individual level, the plate b_i.

$$\begin{aligned}
\eta_i &= \text{logit}(\pi_i) = \beta_0 + \beta_1 x1_i + \beta_2 x2_i + b_i \\
\beta_j &\sim \mathcal{N}(0, \tau_\beta^{-1}),\ \tau_\beta \text{ known},\ j = 0, 1, 2 \\
b_i &\sim \mathcal{N}(0, \tau_b^{-1})
\end{aligned}$$

So, in this case, $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, b_1, \dots, b_{21})$. A Gaussian prior is assigned for each element of the **latent field**, so that $\boldsymbol{\theta} \mid \boldsymbol{\psi}$ is **Gaussian distributed**.

- To assing the prior of $\boldsymbol{\psi} = (\tau_b)$:

$$\log(\tau_b) \sim \text{logGamma}(1, 5 \cdot 10^{-5})$$

Bayesian splines

- GLMM with independent random effect does not cover situations in which relationship between the response variable and the covariate is not linear.

- In INLA, we can do this by means of the **random walk** of order 1 and 2.

  - **First order Random Walk (RW1)**

$$\Delta x_j = x_j - x_{j+1} \sim \mathcal{N}\left(0, \sigma^2 = \frac{1}{\tau}\right)$$

  - **Second order Random Walk (RW2)**

$$\Delta^2 x_i = x_i - 2x_{i+1} + x_{i+2} \sim \mathcal{N}\left(0, \sigma^2 = \frac{1}{\tau}\right)$$

  - The prior for the hyperparameter τ is reparametrized in terms of their logarithm:

$$\log(\tau) \sim \text{logGamma}(1, 5 \cdot 10^{-5}) \ .$$

Smoothing time series of binomial data

- The number of **occurrences of rainfall** over 1 mm in the Tokyo area for each calendar year during two years (1983-84) are registered.

- It is of interest to estimate the underlying probability π_t of rainfall for calendar day t which is, a priori, assumed to change gradually over time.

- For each day $t = 1, \ldots, 366$ of the year we have the number of days that rained y_t and the number of days that were observed n_t.

Dataset

y	n	time
0	2	1
0	2	2
1	2	3
1	2	4
0	2	5
1	2	6

Smoothing time series of binomial data. The model

- A conditionally independent **binomial likelihood** function:

$$y_t \mid \pi_t \sim \mathrm{Binomial}(n, \pi_t),\ t = 1, \ldots, 366$$

with (usual) logit link function:

$$\pi_t = \frac{\exp(\eta_t)}{1 + \exp(\eta_t)}$$

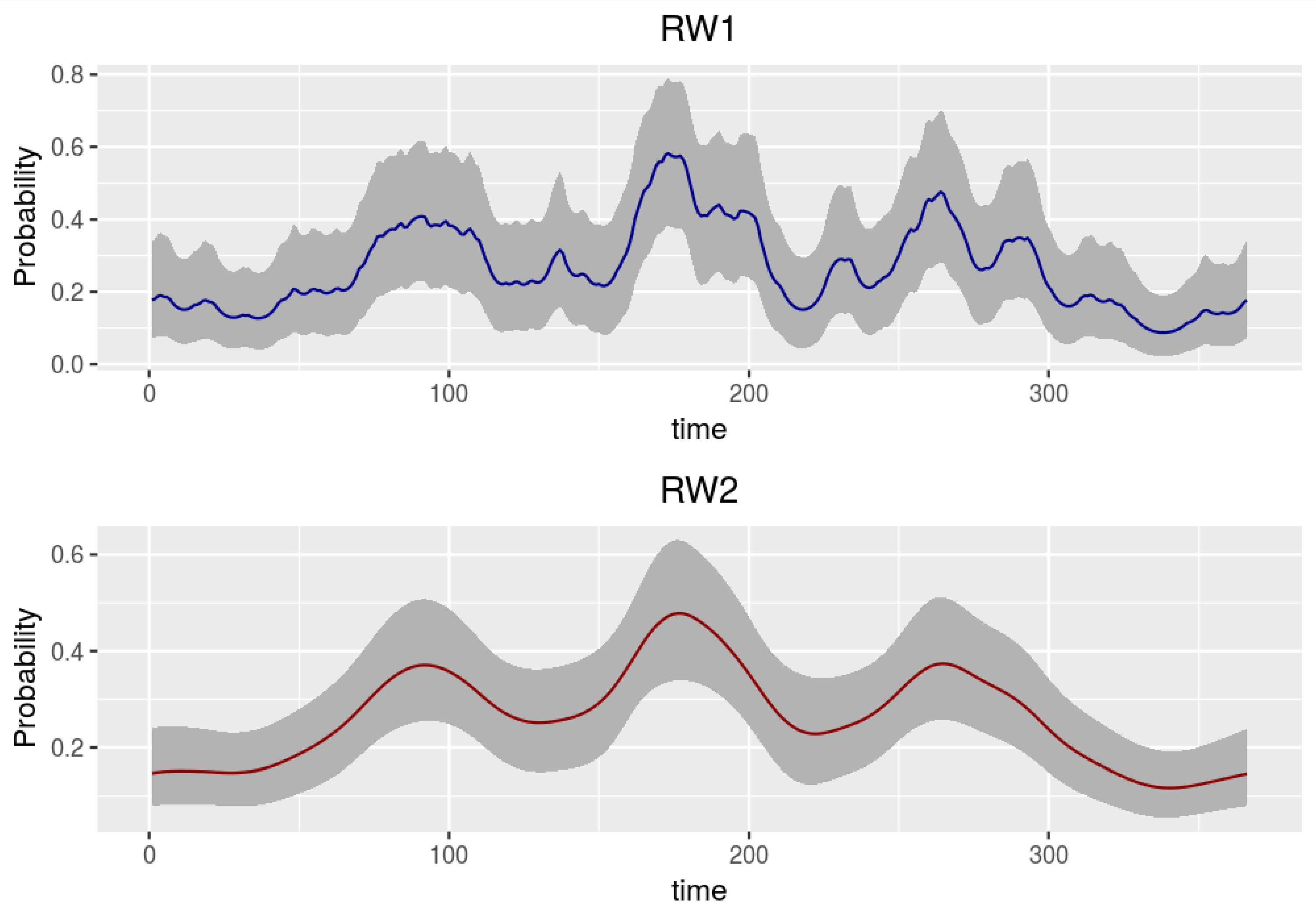
- We assume that (instead of a linear predictor), $\eta_t = f_t$, where f_t follows a circular **random walk** of second order (RW2) model with precision τ:

$$\Delta^2 f_i = f_i - 2f_{i+1} + f_{i+2} \sim \mathcal{N}(0, \tau^{-1}).$$

The fact that we use a circular model here means that in this case f_1 is a neighbor of f_{366}. So, in this case $\boldsymbol{\theta} = (f_1, \ldots, f_{366})$ and again $\boldsymbol{\theta}|\boldsymbol{\psi}$ is **Gaussian distributed**.

- To assing the prior of $\boldsymbol{\psi} = (\tau)$:

Posterior distribution of the probability



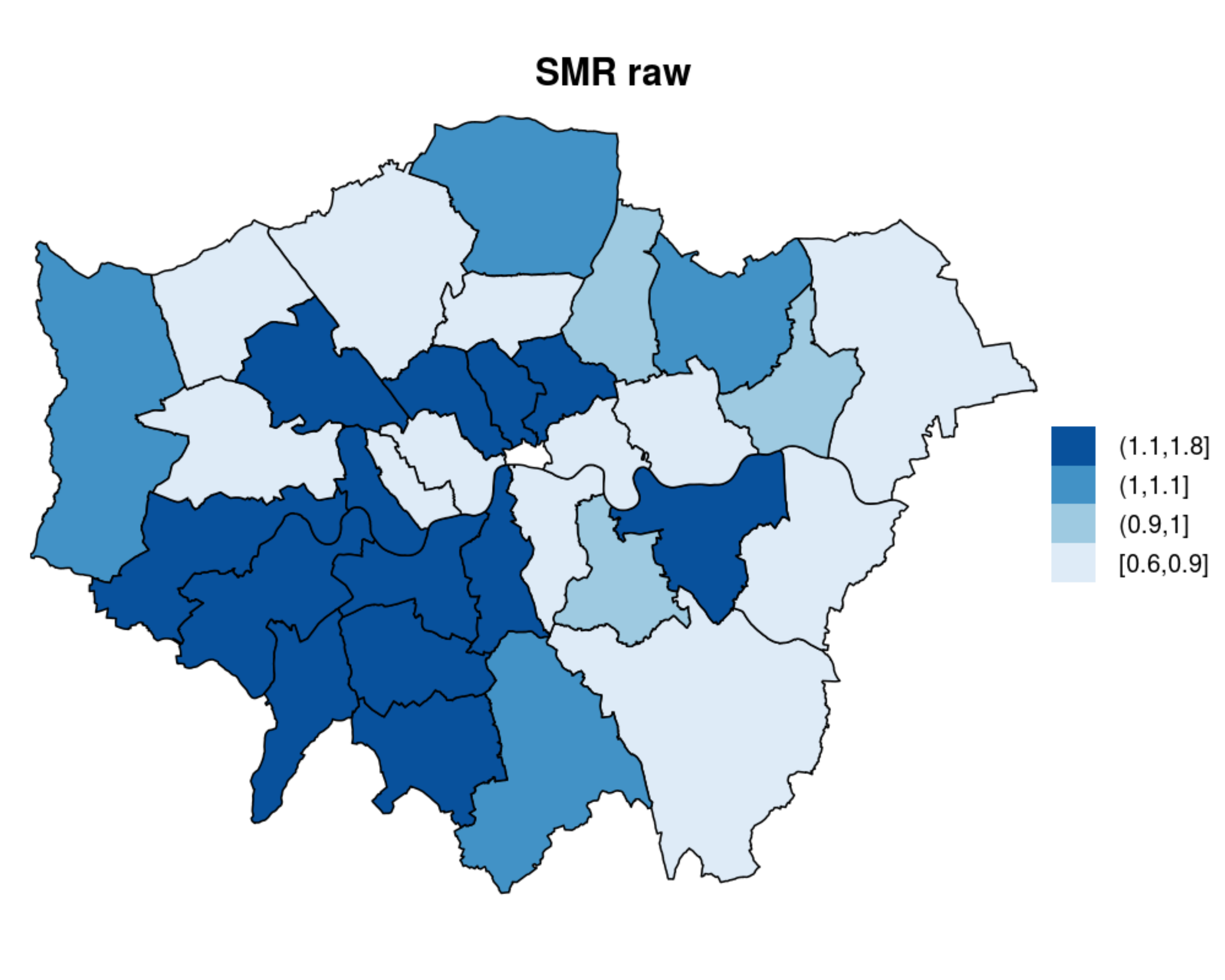
Disease mapping

- Congdon (2007) study suicide mortality in 32 London boroughs (excluding the City of London) in the period 1989–1993 for male and female combined, using a disease mapping model and an ecological regression model.

- The variables are:
  - **N**: which contains the name of boroughs
  - **O**: the number of observed suicides in the period under study
  - **E**: the number of expected cases of suicides (E),
  - **x1**: an index of social deprivation, and
  - **x2**: an index of social fragmentation (X2), which represents the lack of social connections and of sense of community.

NAME	y	E	x1	x2
Barking and Dagenham	75	80.7	0.87	-1.02
Barnet	145	169.8	-0.96	-0.33
Bexley	99	123.2	-0.84	-1.43
Brent	168	139.5	0.13	-0.10
Bromley	152	169.1	-1.19	-0.98
Camden	173	107.2	0.35	1.77

Standarized Mortality Ratio (SMR): raw data



The model

- A conditional independent **Poisson** likelihood function is assumed:

$$y_i \sim \text{Poisson}(\lambda_i), \quad \lambda_i = E_i \rho_i, \quad \log(\rho_i) = \eta_i \ , i = 1, \dots 32$$

- We assume that $\eta_i = \beta_0 + u_i + v_i$, being $\boldsymbol{u}$ the **independent random effect** and $\boldsymbol{v}$ the **spatially structured random effect**:

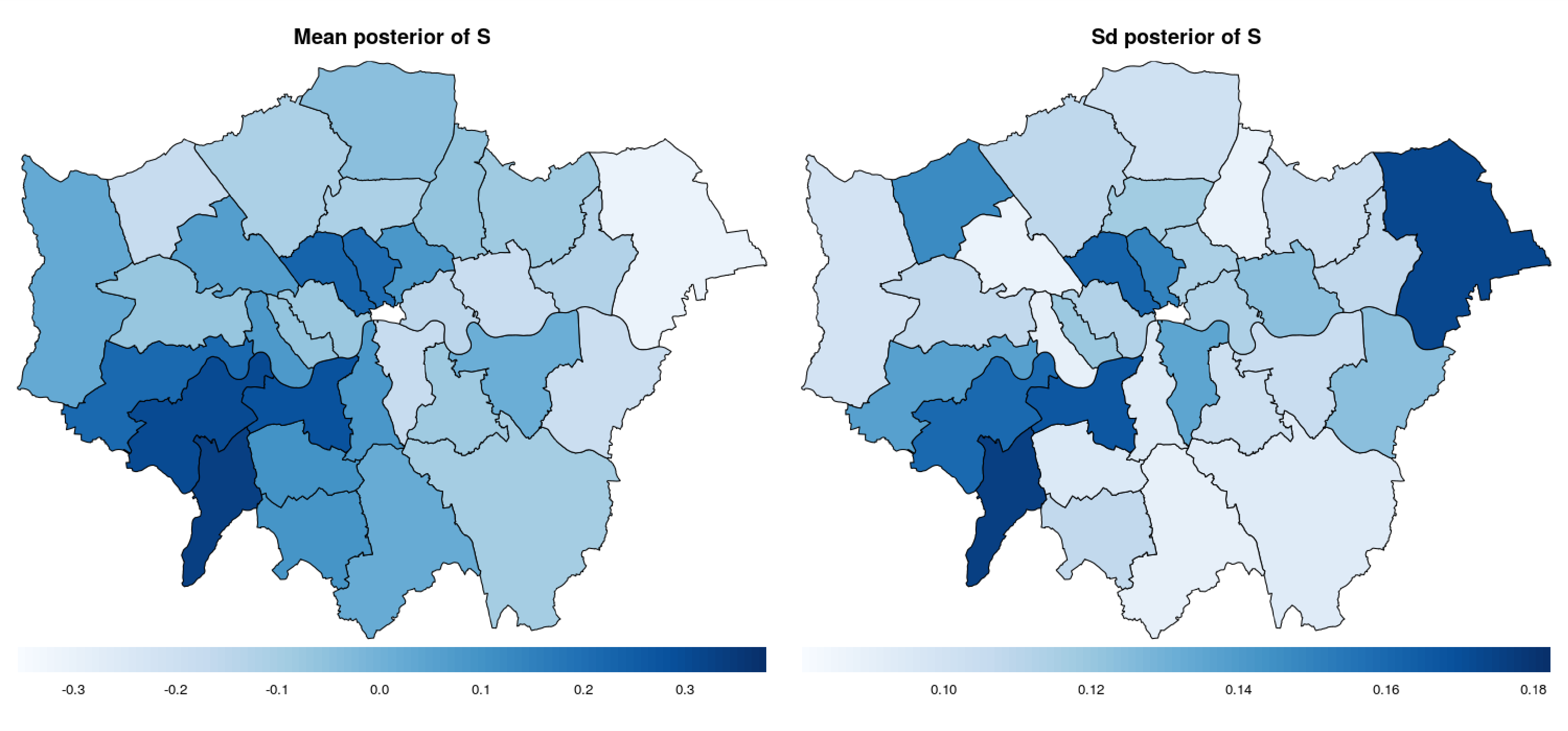
$$u_i \sim \mathcal{N}\left(0, \tau_{\boldsymbol{u}}^{-1}\right), \ v_i \mid \boldsymbol{v}_{-i} \sim \mathcal{N}\left(\frac{1}{n_i} \sum_{i \sim j} v_j, \frac{1}{n_i \tau_{\boldsymbol{v}}}\right) \ .$$

In this case $\boldsymbol{\theta} = (v_1, \dots, v_{32}, u_1, \dots, u_{32})$, and $\boldsymbol{\theta} \mid \boldsymbol{\psi}$ is Gaussian distributed.

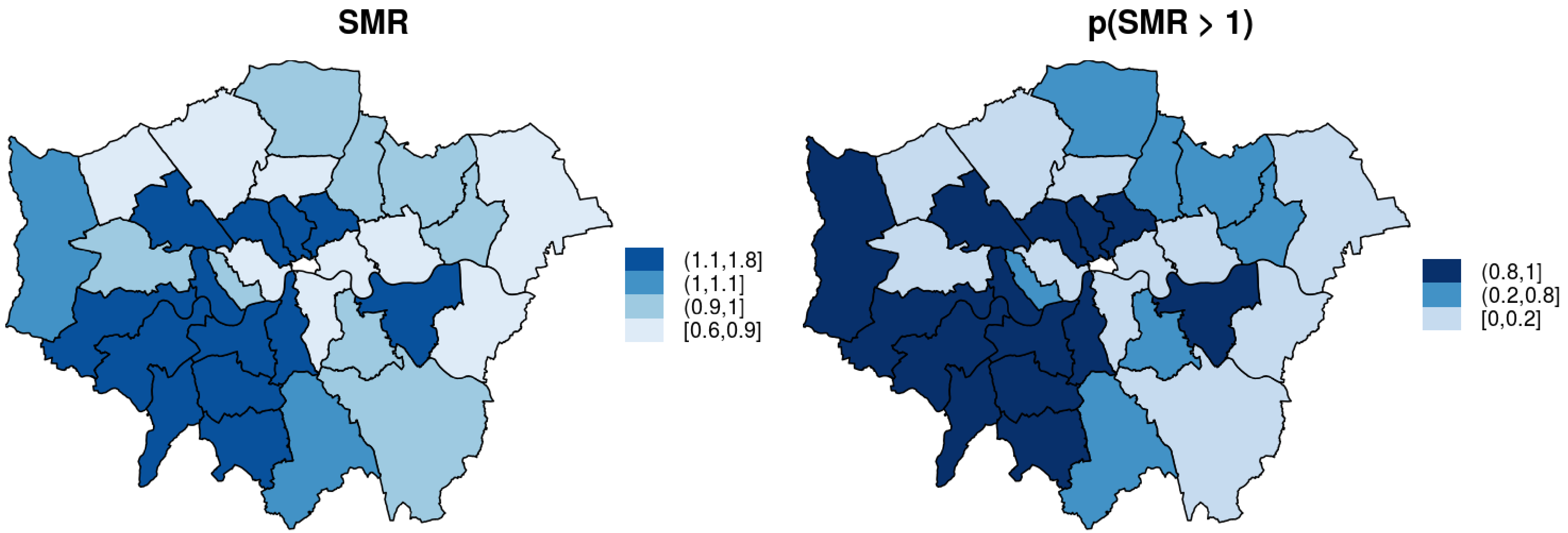
- **Hyperpriors** for $\tau_{\boldsymbol{u}}$ and $\tau_{\boldsymbol{v}}$ are reparametrized in terms of their logarithm:

$$\log(\tau_{\boldsymbol{v}}) \sim \text{logGamma}(1, 0.001) \ , \ \log(\tau_{\boldsymbol{u}}) \sim \text{logGamma}(1, 0.001) \ .$$

Posterior distribution of the spatial effect



Posterior distribution for the SMR and P(SMR > 1)



7. References

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

This material has been constructed based on:

- Blangiardo, M., & Cameletti, M. (2015). Spatial and spatio-temporal Bayesian models with R-INLA. John Wiley & Sons.
- Rue, H., & Held, L. (2005). Gaussian Markov random fields: theory and applications. Chapman and Hall/CRC.
- Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the royal statistical society: Series b (statistical methodology), 71(2), 319-392.
- Wang, X., Ryan, Y. Y., & Faraway, J. J. (2018). Bayesian Regression Modeling with INLA. Chapman and Hall/CRC.
- Tutorials by Haakon Bakka
- A gentle INLA tutorial by Kathryn Morrison
- INLA book by Virgilio Gómez-Rúbio

The End

UNIVERSITAT POLITÈCNICA DE VALÈNCIA