ADIM: Bayesian inference using the integrated nested Laplace approximation (INLA)

Master’s Degree in Data Analysis, Process Improvement and Decision Support Engineering

Joaquín Martínez-Minaya, 2024-12-09

VAlencia BAyesian Research Group
Statistical Modeling Ecology Group
Grupo de Ingeniería Estadística Multivariante
jmarmin@eio.upv.es

Outline

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

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 i-th linear predictor that connects the data to the latent field.

Indeed each η_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:

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

Mixed Models Using INLA

- A study investigates agreement between devices measuring **respiratory rates** in **COPD patients** across **11 activities**, comparing a **chest-band device** to a **gold standard device** (Oxycon mobile).
- The dataset includes 21 subjects performing activities such as sitting, walking, and climbing stairs.

The variables are:

- **y**: respiratory rate (breaths per minute).
- **device**: measurement device (**oxicon**, **chest_band**).
- **replicate**: replicate measurements within each activity/subject.
- **act**: activity type (11 levels, e.g., sitting, climbing stairs).

subj	y	replicate	act	device
1	38.19294	1	Sitting	oxicon
1	40.65189	2	Sitting	oxicon
1	36.00310	3	Sitting	oxicon
1	32.39582	1	Lying	oxicon

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.

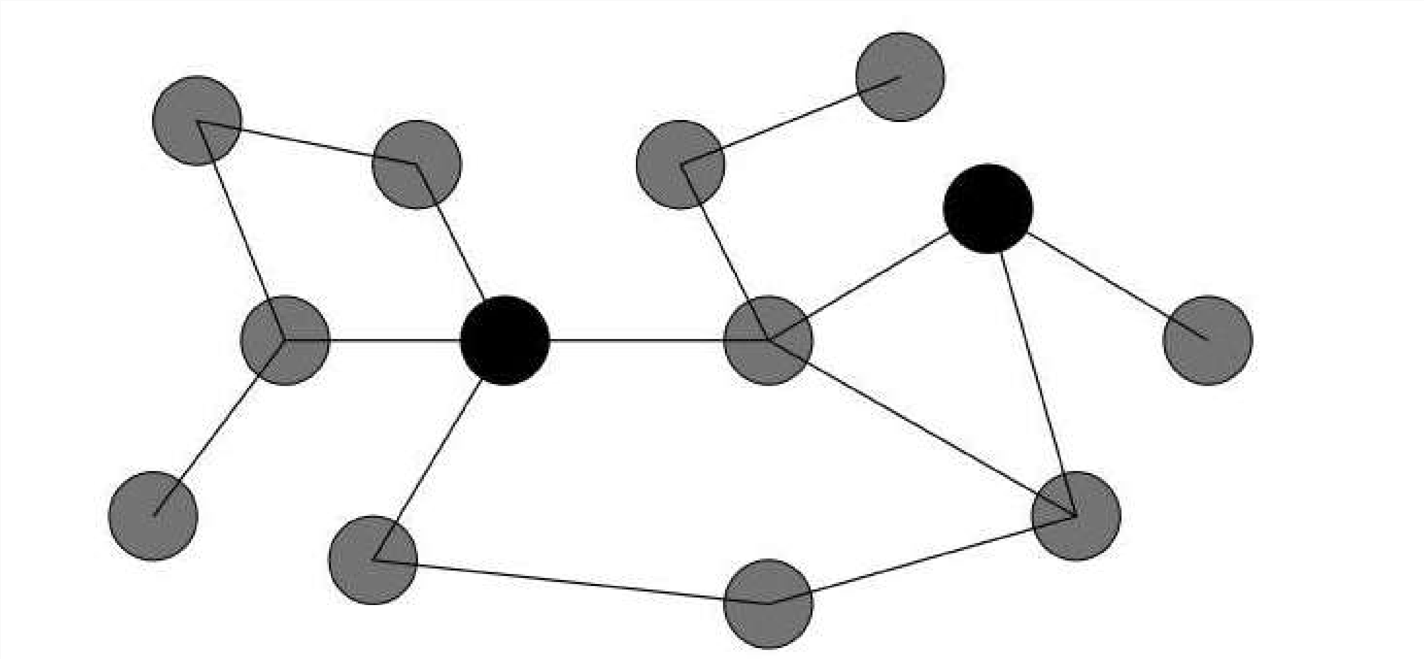
$$\boldsymbol{i} \neq \boldsymbol{j},\ \boldsymbol{\theta}_i \perp \boldsymbol{\theta}_j \mid \boldsymbol{\theta}_{ij},$$

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

- 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

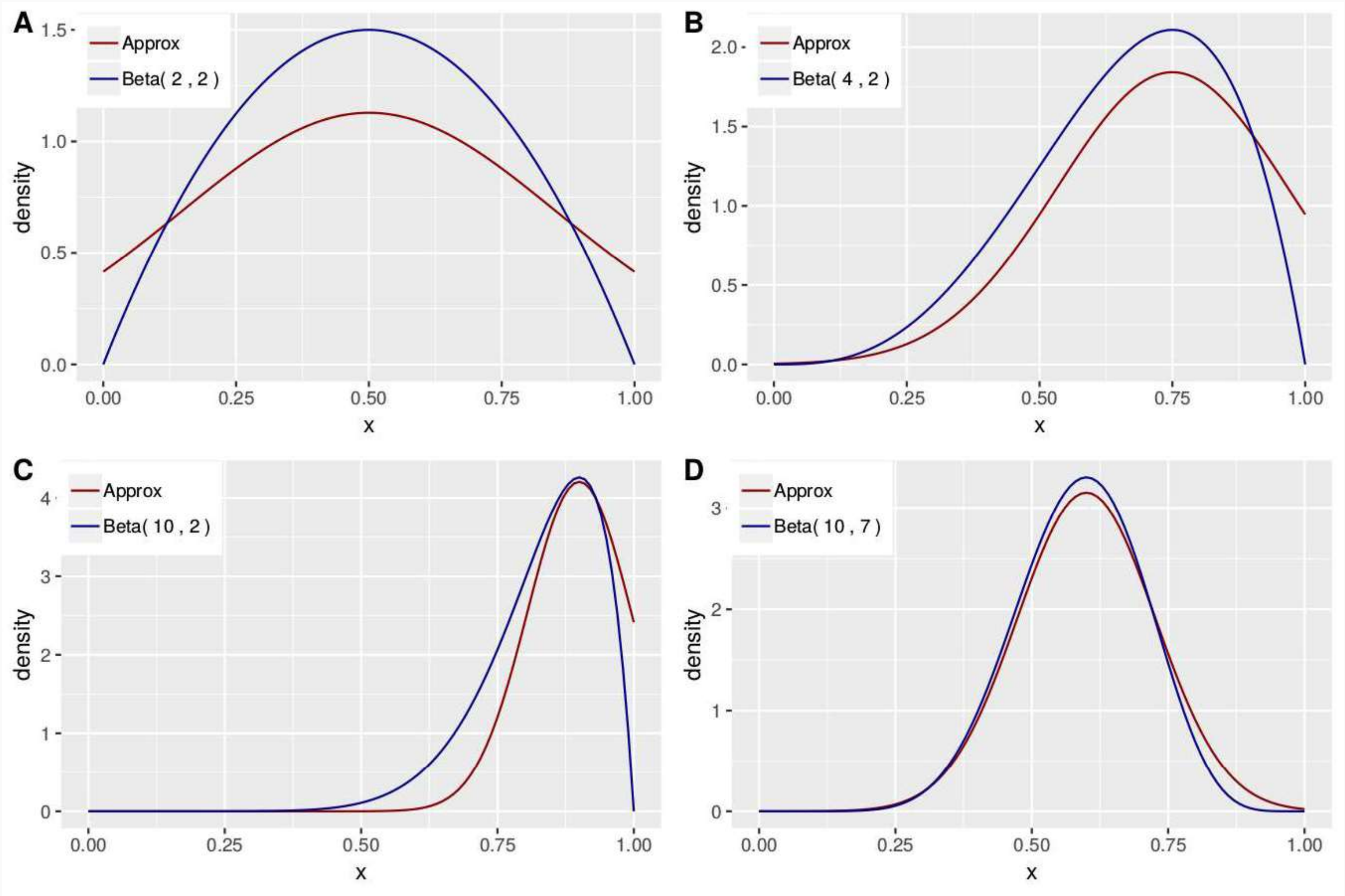
Laplace approximations

- **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

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

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

  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.

The new avenue in INLA

- Posterior means of $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ might be inaccurate based on the Gaussian assumption of the conditional posterior.

- **Variational Bayes correction** to Gaussian means by Van Niekerk and Rue (2021) can be used to efficiently calculate an improved mean for the marginal posteriors of the linear predictors, by improving the posterior means of the latent field.

- It is improved based on the following variational function:

$$E_{q(\boldsymbol{\theta}|\boldsymbol{y})}[-\log(p(\boldsymbol{\theta} \mid \boldsymbol{y}))] + KLD[q(\boldsymbol{\theta} \mid \boldsymbol{y}) \mid\mid p(\boldsymbol{\theta})]] \, ,$$

1. $q(.)$ a member of the variational class,

2. $p(\boldsymbol{\theta} \mid \boldsymbol{y})$ the posterior distribution of the latent field, and

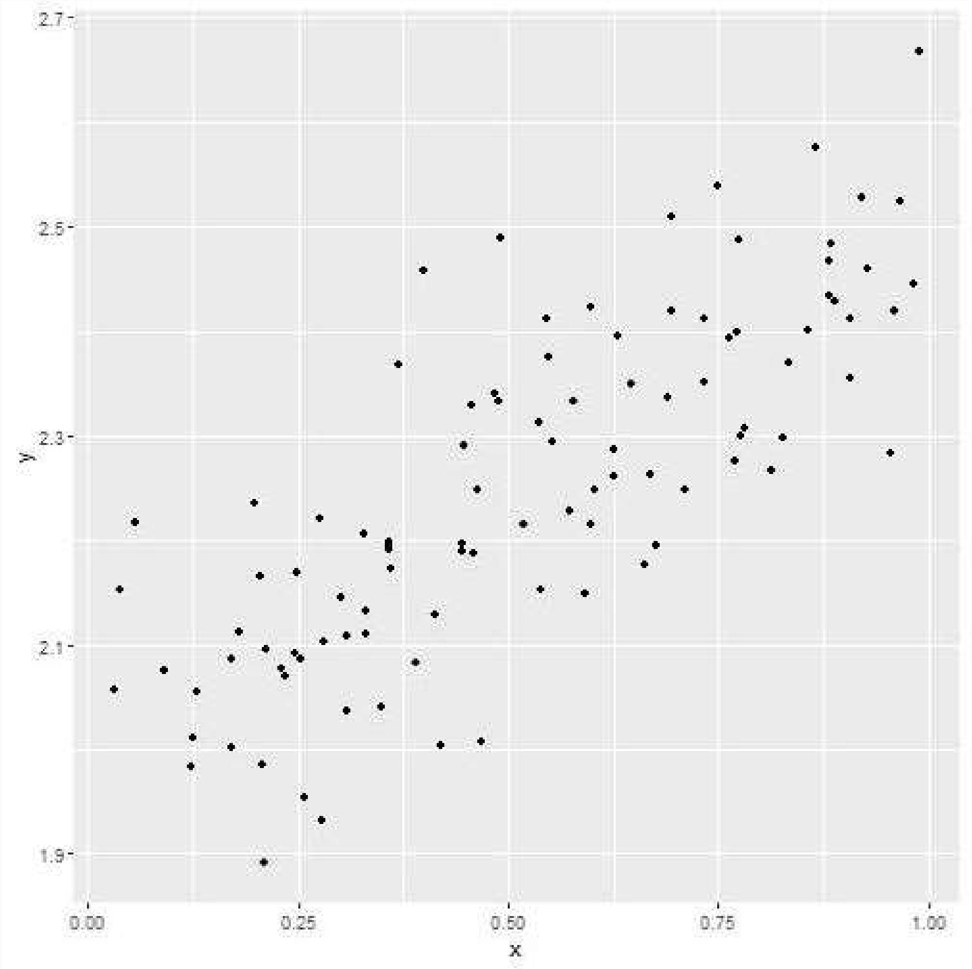
3. $p(\boldsymbol{\theta})$ the posterior distribution of the latent field.

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

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

Calling `R-INLA`

```
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	0.0225	1.9503	1.9946	2.0389	1.9946	0
x	0.4935	0.0388	0.4174	0.4935	0.5697	0.4935	0

Posterior distributions of the hyperparameters

	mean	sd	0.025quant	0.5quant	0.975quant	mode
Precision for the Gaussian observations	99.4863	14.0682	73.8685	98.8272	128.9193	97.5018

Families

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

```
## Section [likelihood]
##      0binomial                      New 0-inflated Binomial
##      0binomialS                     New 0-inflated Binomial Swap
##      0poisson                       New 0-inflated Poisson
##      0poissonS                      New 0-inflated Poisson Swap
##      agaussian                      The aggregated Gaussian likelihoood
##      bell                           The Bell likelihood
##      beta                           The Beta likelihood
##      betabinomial                   The Beta-Binomial likelihood
##      betabinomialna                 The Beta-Binomial Normal approximation likelih
##      bgev                           The blended Generalized Extreme Value likeliho
##      binomial                       The Binomial likelihood
##      cbinomial                      The clustered Binomial likelihood
##      cennbinomial2                  The CenNegBinomial2 likelihood (similar to cen
##      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
```

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 mod
##      bym                           The BYM-model (Besag-York-Mollier model)
##      bym2                          The BYM-model with the PC priors
##      cgeneric                      Generic latent model specified using C
##      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
##      fgn2                          Fractional Gaussian noise model (alt 2)
##      generic                       A generic model
```

Hyperpriors

```
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
##      laplace                      Laplace prior
##      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
##      minuslogsqrtruncnormal       (obsolete)
##      mvnorm                       A multivariate Normal prior
```

5. Model Selection

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. Example

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:

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

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

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

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

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:

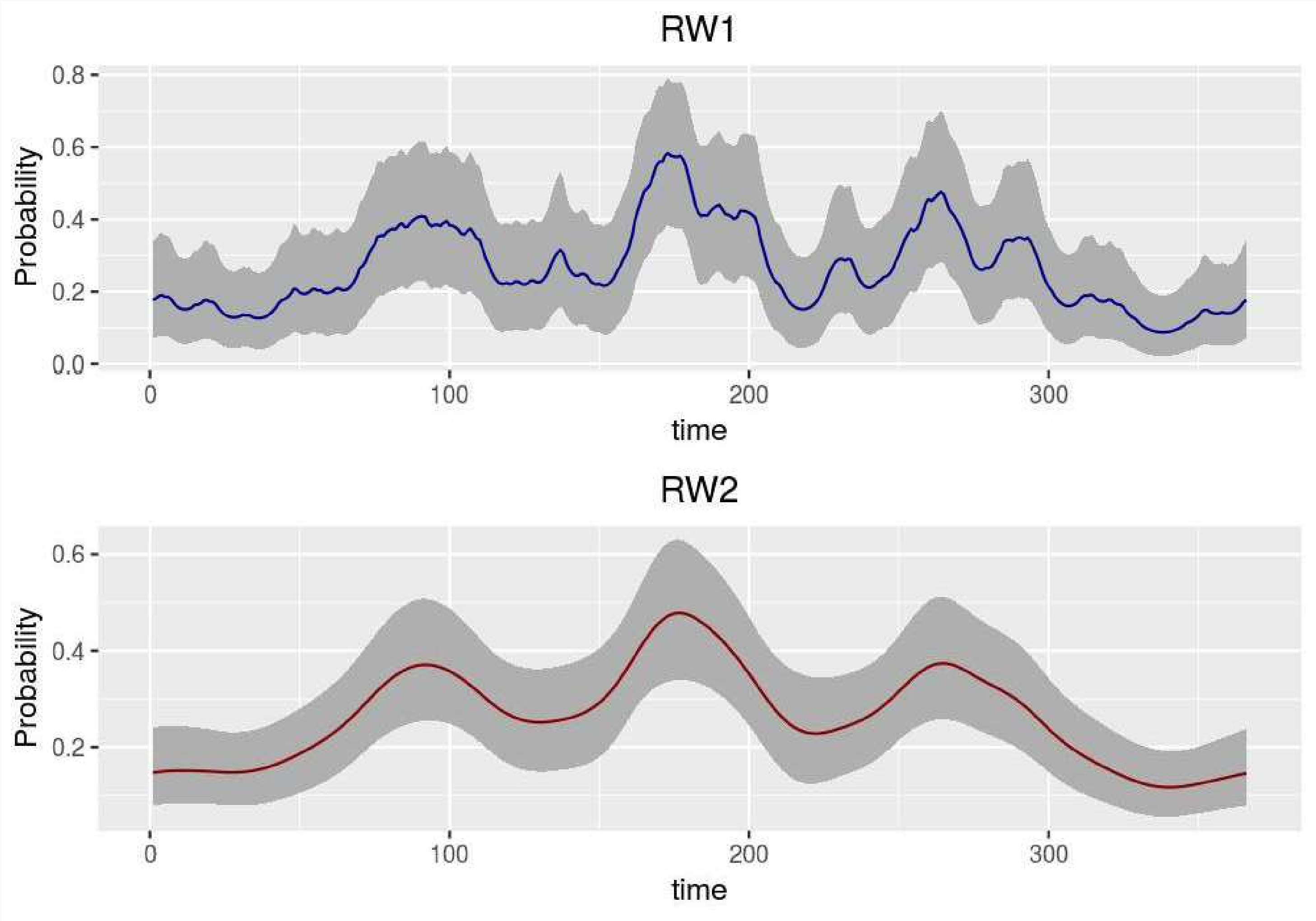
$$\Delta^2 f_i = f_i - 2f_{i+1} + f_{i+2} \sim \mathcal{N}(0, \tau^{-1}).$$

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 $\boldsymbol{\theta} = (f_1, \dots, f_{366})$ and again $\boldsymbol{\theta}|\boldsymbol{\psi}$ is **Gaussian distributed**.

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

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

Posterior distribution of the probability



7. References

This material has been constructed based on:

- Gómez-Rubio, V. (2020). Bayesian inference with INLA. Chapman and Hall/CRC.
- Parker, R. A., Scott, C., Inácio, V., & Stevens, N. T. (2020). Using multiple agreement methods for continuous repeated measures data: a tutorial for practitioners. BMC Medical Research Methodology, 20, 1-14.
- 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

ADIM: Bayesian inference using the integrated nested Laplace approximation (INLA)

Master's Degree in Data Analysis, Process Improvement and Decision Support Engineering

Joaquín Martínez-Minaya, 2024-12-09
VAlencia BAyesian Research Group
Statistical Modeling Ecology Group
Grupo de Ingeniería Estadística Multivariante

jmarmin@eio.upv.es

UNIVERSITAT POLITÈCNICA DE VALÈNCIA