Computational Statistics

Introduction to R-INLA

Joaquin Cavieres

Georg-August-Universität Göttingen

Content

1. Introduction

Consider the following: $\mathbf{y}$ is the vector of observations with probability density function $p(\mathbf{y} \mid \boldsymbol{\theta})$. From that, generally we can estimate $\boldsymbol{\theta}$ in two ways:

1. Frequentist approach

$\boldsymbol{\theta}$ can be estimated by maximum likelihood.

2. Bayesian approach

$\boldsymbol{\theta}$ has a prior distribution $(p(\boldsymbol{\theta}))$ and we can estimate $\boldsymbol{\theta}$ based on Bayes theorem:

$$p(\boldsymbol{\theta} \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})} \propto p(\mathbf{y} \mid \boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{1}$$

where $p(\boldsymbol{\theta} \mid \mathbf{y})$ is the posterior distribution of $\boldsymbol{\theta}$.

Within Bayesian framework we can:

- Fit hierarchical models to consider complex structures of dependencies in the data.
- Assign explicit uncertainty to the parameters and latent variables.

Option 1: Markov Chain Monte Carlo (MCMC)

MCMC methods find a posterior stationary distribution based on sampling.

- The most used method for Bayesian inference.
- Commonly used in models with non tractable likelihood function.

Problem?

We need to specify appropriate priors, and the computational cost to calculate the posterior can sometimes be expensive.

Problem?

We need to specify appropriate priors, and the computational cost to calculate the posterior can sometimes be expensive.

Option 2: Integrated Nested Laplace Approximation (INLA)

This method was proposed by Rue et al. (2009) and basically, we can obtain the posterior distribution using numerical approximations. Advantage?

Sampling is not required.

- It is a Bayesian framework to fit Latent Gaussian Models.
- Available in R (see http://www.r-inla.org/)

Option 2: Integrated Nested Laplace Approximation (INLA)

This method was proposed by Rue et al. (2009) and basically, we can obtain the posterior distribution using numerical approximations. Advantage?

Sampling is not required.

- It is a Bayesian framework to fit Latent Gaussian Models.
- Available in R (see http://www.r-inla.org/)

Option 2: Integrated Nested Laplace Approximation (INLA)

This method was proposed by Rue et al. (2009) and basically, we can obtain the posterior distribution using numerical approximations. Advantage?

Sampling is not required.

- It is a Bayesian framework to fit Latent Gaussian Models.
- Available in R (see http://www.r-inla.org/)

INLA is an alternative method of estimation to the classical algorithms implemented in the MCMC method for Bayesian inference. The algorithms in the MCMC method are asymptotically exact, whereas INLA is an approximation method.

Types of models that can be fitted using `R-INLA`:

- Generalized Linear Models (GLM).
- Generalized Linear Mixed Models (GLMM).
- Time series models.
- Spatial models (areal, continuous and point patterns data).
- Spatio-temporal models.

2. Latent Gaussian Models

First, let consider $\mathbf{y} = (y_1, \ldots, y_n)$, a vector of n observations. The mean of this vector, μ_i, can be linked to a linear predictor η_i as follows:

$$\eta_i = \beta_0 + \sum_{j=1}^{J} \beta_j z_{ji} + \sum_{k=1}^{K} f_{(k)}(u_{ki}), \qquad i = 1, \ldots, n, \tag{2}$$

where β_0 is the intercept, β_j is the vector of parameters associated with covariates $\{\mathbf{z}_j\}$, and $f_{(k)}$ are the functions associated with the random effects $\{\mathbf{u}_k\}$.

The vector of latent effects $\mathbf{x}$ contains:

$$\mathbf{x} = (\eta_1, \ldots, \eta_n, \beta_0, \beta_1, \ldots) \tag{3}$$

For each linear effect and each component, we must consider a Gaussian prior with either a univariate or multivariate normal density to ensure that the additive $\mathbf{x}$ is Gaussian, as required by the LGM.

Likelihood

$$\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}_1 \sim \prod_{i \in \mathcal{I}} p(y_i \mid x_i, (\theta)_1)$$

LGM

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

Prior

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

For simplicity, we are considering $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$.

Since $(y_1, \ldots, y_n)$ are assumed as independent given $\mathbf{x}$ and $\boldsymbol{\theta}$, the likelihood can be re-written as:

$$p(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}) = \prod_{i=1} p(y_i \mid x_i, \boldsymbol{\theta}), \tag{4}$$

for $i = 1, \ldots n$.

Also, as **x** is assumed as a GMRF, the posterior distribution is:

$$p(\mathbf{x} \mid \boldsymbol{\theta}) \propto |\mathbf{Q}(\boldsymbol{\theta})|^{1/2} \exp\{-\frac{1}{2}\mathbf{x}^T\mathbf{Q}(\boldsymbol{\theta})\mathbf{x}\} \tag{5}$$

Finally, the joint posterior distribution $p(\mathbf{x}, \boldsymbol{\theta} \mid \mathbf{y})$ is:

$$p(\mathbf{x}, \boldsymbol{\theta} \mid \mathbf{y}) \propto p(\boldsymbol{\theta})|\mathbf{Q}(\boldsymbol{\theta})|^{1/2} \exp\{-\frac{1}{2}\mathbf{x}^T\mathbf{Q}(\boldsymbol{\theta})\mathbf{x}\} \prod_{i=1} p(y_i \mid x_i, \boldsymbol{\theta}) =$$
$$p(\boldsymbol{\theta})|\mathbf{Q}(\boldsymbol{\theta})|^{1/2} \exp\{-\frac{1}{2}\mathbf{x}^T\mathbf{Q}(\boldsymbol{\theta})\mathbf{x} + \sum_{i=1} \ln(p(y_i \mid x_i, \boldsymbol{\theta}))\} \quad (6)$$

3. INLA

How works `R-INLA`?

INLA will not attempt to estimate the full posterior distribution but will instead focus on the marginals of the latent effects and hyperparameters.

Since we are only interested in posterior marginals, the `R-INLA` method can be summarized as:

- Step 1: Find a Laplace approximation to $p(\boldsymbol{\theta} \mid \mathbf{y})$.
- Step 2: Find an approximation to $p(\mathbf{x}_i \mid \boldsymbol{\theta}, \mathbf{y})$:
  - Gaussian approximation
  - Laplace approximation
  - Simplified Laplace approximation
- Step 3: Numerical integration
  - Grid strategy
  - Central composite design (CCD)

Step 1: Posterior marginal for a hyperparameter $\boldsymbol{\theta}_k$

$$p(\boldsymbol{\theta}_k \mid \mathbf{y}) = \int p(\boldsymbol{\theta} \mid \mathbf{y}) d\boldsymbol{\theta}_{-k},$$

where $\boldsymbol{\theta}_{-k}$ is a vector of hyperparameters $\boldsymbol{\theta}$ without the element $\boldsymbol{\theta}_k$.

Hence, using the posterior approximation of $\boldsymbol{\theta}$, $\tilde{p}(\boldsymbol{\theta} \mid \mathbf{y})$, allows us to compute the marginals of $\mathbf{x}$ and hyperparameters $\boldsymbol{\theta}$ such that:

$$\tilde{p}(\boldsymbol{\theta} \mid \mathbf{y}) = \left.\frac{p(\mathbf{x}, \boldsymbol{\theta} \mid \mathbf{y})}{\tilde{p}_G(\mathbf{x} \mid \boldsymbol{\theta}, \mathbf{y})}\right|_{\mathbf{x}=\mathbf{x}^*(\boldsymbol{\theta})}$$

and evaluate it at the mode $\mathbf{x}^*(\boldsymbol{\theta})$.

Step 2: Approximating the marginals fof the latent effects

$$\tilde{p}\left(\mathbf{x}_i|\mathbf{y}\right) = \int p\left(\mathbf{x}_i|\boldsymbol{\theta},\mathbf{y}\right)\tilde{p}(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}$$

There are three alternatives:

- Gaussian approximation
- Laplace approximation
- Simplified Laplace approximation

- Gaussian distribution ($\hat{p}_G(\mathbf{x} \mid \boldsymbol{\theta}, \mathbf{y})$):

$$\hat{p}(\mathbf{x}_i \mid \boldsymbol{\theta}, \mathbf{y}) = \mathcal{N}(\mathbf{x}_i; \mu_i(\boldsymbol{\theta}), \sigma^2(\boldsymbol{\theta}))$$

with mean $\mu_i(\boldsymbol{\theta})$ and marginal variance $\sigma^2(\boldsymbol{\theta})$.

- Laplace approximation

$$\hat{p}_{LA}(\mathbf{x}_i \mid \boldsymbol{\theta}, \mathbf{y}) \propto \left. \frac{p(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\hat{p}_{GG}(\mathbf{x}_{-1} \mid \mathbf{x}_i, \boldsymbol{\theta}, \mathbf{y})} \right|_{\mathbf{x}_{-i}=\mathbf{x}^\star(\mathbf{x}_i,\boldsymbol{\theta})}$$

The approximation is highly accurate but computationally expensive requiring n factorizations of $(n-1) \times (n-1)$ matrices to obtain the n marginals.

- Simplified Laplace approximation
  - Based on a series expansion up to third order of the numerator and denominator of $\hat{p}_{LA}(\mathbf{x}_i \mid \boldsymbol{\theta}, \mathbf{y})$
  - Corrects the Gaussian approximation for error in location and lack of skewness (fit a Skew-Normal density!).
  - Significantly faster: $\mathcal{O}(n\log(n))$ for each i.

This is default option when using R-INLA but this choice can be modified.

Step 3: Numerical integration

From Step 1:

$$\tilde{p}\left(\mathbf{x}_i|\mathbf{y}\right) = \int \tilde{p}\left(\mathbf{x}_i|\boldsymbol{\theta}, \mathbf{y}\right)\tilde{p}(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}$$

and Step 2:

$$\tilde{p}\left(\boldsymbol{\theta}_j|\mathbf{y}\right) = \int \tilde{p}(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}_{-j}$$

The numerical integration to approximate the marginals of the latent effects:

$$\tilde{p}\left(\mathbf{x}_i|\mathbf{y}\right) \approx \sum_k \tilde{p}\left(\mathbf{x}_i|\boldsymbol{\theta}_k, \mathbf{y}\right)\tilde{p}\left(\boldsymbol{\theta}_k|\mathbf{y}\right)\Delta_k,$$

where Δ_k is the area weight corresponding to $\boldsymbol{\theta}_k$.

Books about INLA (R-INLA):

- Wang et al. (2018): Bayesian Regression Modeling with INLA
- Gómez-Rubio (2020): Bayesian inference with INLA.
- Krainski et al. (2018): Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA.

See you next week!

References I

Gómez-Rubio, V. (2020). *Bayesian inference with INLA*. Chapman and Hall/CRC.

Krainski, E., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., and Rue, H. (2018). *Advanced spatial modeling with stochastic partial differential equations using R and INLA*. Chapman and Hall/CRC.

Rue, H., Martino, S., and 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., Yue, Y. R., and Faraway, J. J. (2018). *Bayesian regression modeling with INLA*. Chapman and Hall/CRC.