

Introduction to INLA

Joaquin Cavieres G.

Universidad de Valparaíso
Ph.D (c) in Statistics

October 15, 2020

Contents

- 1 Introduction
- 2 Latent Gaussian Models
- 3 INLA
- 4 Application

Commonly we have two paradigms for statistical modelling, for example:

Consider the following: \mathbf{y} is a set of observations with distribution of probability $\pi(\mathbf{y} | \boldsymbol{\theta})$. For the above we can estimate $\boldsymbol{\theta}$ of two ways:

Frequentist approach

$\boldsymbol{\theta}$ denotes **fixed** and **unknown** parameters and what can be estimated by maximum likelihood.

Bayesian approach

$\boldsymbol{\theta}$ denotes **random variables** with a **prior** $\pi(\boldsymbol{\theta})$ specification. We can estimate $\boldsymbol{\theta}$ based on the **posterior** like this:

$$\pi(\boldsymbol{\theta} | \mathbf{y}) = \frac{\pi(\mathbf{y} | \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\mathbf{y})} \propto \pi(\mathbf{y} | \boldsymbol{\theta})\pi(\boldsymbol{\theta}) \quad (1)$$

On the Bayesian framework we can use:

- Hierarchical models to consider complex structures of dependencies for the data
- Give an explicit uncertainty to model parameters and latent variables

Problem?

On the Bayesian framework we can use:

- Hierarchical models to consider complex structures of dependencies for the data
- Give an explicit uncertainty to model parameters and latent variables

Problem?

We need specify correct priors and the computational cost to calculate the posterior is expensive.

Option 1: Markov Chain Monte Carlo (MCMC)

The MCMC method use Markov chains to find a posterior stationary distribution based in sampling.

- The most used method for the Bayesian inference
- Commonly used in models with non tractable likelihood function.
- It's very easy to use but with slow convergence in complex models

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.

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?**

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?

Is not necessary sampling.

INLA

- A Bayesian framework to analyze a specific class of models called Latent Gaussian Models
- Available in R (see <http://www.r-inla.org/>)

Integrated Nested Laplace Approximation (INLA)

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

Types of models that we can use with INLA:

- Generalized Linear Models (GLM)
- Generalized Linear Mixed Models (GLMM)
- Time series models
- Spatial models
- Spatio-temporal models

Latent Gaussian Models

Likelihood

$$\mathbf{y} \mid \mathbf{x}, \theta_1 \sim \prod_{i \in I} \pi(y_i \mid x_i, \theta_1)$$

LGM

$$\mathbf{x} \mid \theta_2 \sim \mathcal{N}(\mu(\theta_2), \mathbf{Q}^{-1}(\theta_2))$$

Posterior

$$\pi(\mathbf{x}, \boldsymbol{\theta} \mid \mathbf{y}) \propto \pi(\boldsymbol{\theta}) \pi(\mathbf{x} \mid \boldsymbol{\theta}) \prod_{i \in I} \pi(y_i \mid x_i, \boldsymbol{\theta})$$

Hierarchical structure:

$$y \mid \mathbf{x}, \boldsymbol{\theta} \sim \prod_i \pi(y_i \mid \eta_i, \boldsymbol{\theta}) \quad \textit{Data}$$

$$\mathbf{x} \mid \boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}(\boldsymbol{\theta}^1)) \quad \textit{Gaussian}$$

$$\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta}) \quad \textit{Hyperparameters}$$

Here \mathbf{Q} is the precision matrix (sparse matrix) and generally \mathbf{x} is assumed as a Gaussian Markov Random Field (GMRF).

How works INLA?

From the posterior distribution

$$\pi(\mathbf{x}, \boldsymbol{\theta} \mid \mathbf{y}) \propto \pi(\boldsymbol{\theta})\pi(\mathbf{x} \mid \boldsymbol{\theta})\pi(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}) \quad (2)$$

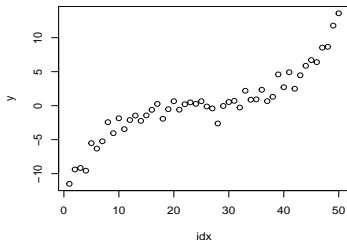
we can approximate the posterior marginals directly

$$\pi(x_i \mid y) \quad \text{and} \quad \pi(\theta_j \mid y) \quad (3)$$

How works INLA?

$$y_i = u(i) + \epsilon_i \quad i = 1, \dots, n$$

We assumed that $u(i)$ is a smooth function indexed by i and $\epsilon_i \sim \mathcal{N}(0, \tau_0)$ with known parameter τ_0 .



Figure

How we only are interested in posterior marginals, the INLA method can be summarized as:

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

Step 1: Laplace approximation to $\pi(\boldsymbol{\theta} \mid \mathbf{y})$

$$\tilde{\pi}(\boldsymbol{\theta} \mid \mathbf{y}) = \frac{\pi(\mathbf{x}, \boldsymbol{\theta} \mid \mathbf{y})}{\tilde{\pi}_G(\mathbf{x} \mid \boldsymbol{\theta}, \mathbf{y})} \Big|_{\mathbf{x}=\mathbf{x}^*(\boldsymbol{\theta})}$$

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

$$\pi(\mathbf{x}, \boldsymbol{\theta} \mid \mathbf{y}) = \pi(\boldsymbol{\theta})\pi(\mathbf{x} \mid \boldsymbol{\theta}) \prod_i \pi(y_i \mid x_i, \boldsymbol{\theta})$$

Step 2: Marginals of the latent field

Marginales of x_i :

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

We have three alternatives:

- Gaussian approximation
- Laplace approximation
- Simplified Laplace approximation

Step 2: Marginals of the latent field

- Gaussian distribution derived from $\hat{\pi}_G(x \mid \theta, y)$

$$\hat{\pi}(x_i \mid \theta, y) = \mathcal{N}(x_i; \mu_i(\theta), \sigma^2(\theta))$$

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

Step 2: Marginals of the latent field

- Laplace approximation

$$\hat{\pi}_{LA}(x_i \mid \boldsymbol{\theta}, y) \propto \frac{\pi(x, \boldsymbol{\theta}, y)}{\hat{\pi}_{GG}(x_{-1} \mid x_i, \boldsymbol{\theta}, y)} \Big|_{x_{-i}=x^*(x_i, \boldsymbol{\theta})}$$

The approximation is very good but expensive as n factorizations of $(n-1) \times (n-1)$ matrices are required to get the n marginals

Step 2: Marginals of the latent field

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

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

Step 3: Numerical integration

Considering Step 1:

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

and Step 2:

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

The numerical integration to approximate the marginales of the latent field:

$$\tilde{\pi}(x_i|\mathbf{y}) \approx \sum_k \tilde{\pi}(x_i|\theta_k, \mathbf{y}) \tilde{\pi}(\theta_k|\mathbf{y}) \Delta_k$$

where Δ_k is the area weight corresponding to θ_k .

 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.

Introducción a INLA

Teoría y aplicaciones

Joaquin Cavieres G.

Modelo de regresión lineal

Consideremos el siguiente modelo de regresión lineal:

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i, \quad (1)$$

para $i = 1, \dots, n$, en donde n es el número de observaciones, p es el número de predictores para las covariables (x_1, \dots, x_p) y la variable y es la respuesta del modelo de regresión.

En ocasiones dentro del predictor pueden existir variables no lineales, por ejemplo en una regresión polinomial de grado 3 $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \varepsilon_i$, pero mantiene la linealidad ya que es lineal en los parámetros. Consideremos el modelo anterior en forma matricial:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2)$$

donde el vector $\boldsymbol{\varepsilon}$ se asume como independiente e idénticamente distribuido con media 0 y varianza constante. Generalmente este término tiene distribución $\mathcal{N}(0, \boldsymbol{\sigma}^2 \mathbf{I})$ con una varianza desconocida.

Mediante la metodología frecuentista los parámetros pueden estimarse directamente a través del método de máxima verosimilitud (MLE) y, si no hacemos ningún supuesto distribucional en el error aleatorio, se puede estimar mediante el método de mínimos cuadrados. En este caso la función de verosimilitud quedaría expresada como:

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\sigma}^2 \mid \mathbf{X}, \mathbf{y}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})} \quad (3)$$

Mediante las derivadas parciales podemos encontrar las expresiones para $\boldsymbol{\beta}$ y $\boldsymbol{\sigma}^2$,

$$M_1(\boldsymbol{\beta}, \boldsymbol{\sigma}^2) = \frac{\partial \log \mathcal{L}}{\partial \boldsymbol{\beta}} = -\frac{1}{2\sigma^2} \mathbf{X}^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \quad (4)$$

$$M_2(\boldsymbol{\beta}, \boldsymbol{\sigma}^2) = \frac{\partial \log \mathcal{L}}{\partial \boldsymbol{\sigma}^2} = -\frac{n}{2\sigma^3} + \frac{1}{2\sigma^3} \mathbf{X}^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \quad (5)$$

e igualando a 0 obtenemos los estimadores máximos verosímiles:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (6)$$

$$\hat{\sigma}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X} \hat{\beta})^T (\mathbf{y} - \mathbf{X} \hat{\beta}) \quad (7)$$

Comúnmente se utiliza el estimador de S^2 para la varianza ya que $\hat{\sigma}^2$ es sesgado.

$$S^2 = \frac{1}{n - p - 1} (\mathbf{y} - \mathbf{X} \hat{\beta}^T) (\mathbf{y} - \mathbf{X} \hat{\beta}) \quad (8)$$

Inferencia Bayesiana

Dentro de la estadística Bayesiana la varianza σ^2 es reemplazada por el parámetro *precision* τ , el cual es $\tau = 1/\sigma^2$.

Si consideramos los supuestos de un modelo de regresión lineal, entonces:

$$\mathbf{y} \mid \beta, \tau \sim \mathcal{N}(\mathbf{X}\beta, \tau^{-1}\mathbf{I}), \quad (9)$$

en donde $\pi(\beta)$ y $\pi(\tau)$ son las distribuciones a priori para β y τ .

La asignación de distribuciones a priori para los parámetros del modelo siempre ha sido materia de discusión en la estadística Bayesiana. Existe diversa literatura relacionada a distintos análisis y estimación utilizando distribuciones a priori para los parámetros en distintos contextos, pero en resumen nosotros podemos decir que: si se cuenta con información sobre los parámetros del modelo, entonces esta debe ser incorporada dentro de la distribución a priori, si no se cuenta con información previa sobre los parámetros del modelo entonces el objetivo principal es que la distribución a priori tenga una influencia mínima en la estimación de los parámetros (y su posterior inferencia). Por ejemplo, una prior no-informativa siempre se considera como un buen punto de partida ya que en muchas aplicaciones no se tiene información sobre los parámetros del modelo, el problema de usar una prior no informativa es que esta no es invariante a transformaciones previas de los parámetros. Una mayor discusión sobre la utilización de prior's se puede revisar en Gelman et al. (2014).

Como en INLA se asume un modelo latente (Latent Gaussian Model) entonces es sencillo asignar una distribución a priori Gaussiana para β . Para el hiperparámetro τ generalmente se asume una prior difusa (una distribución a priori con una gran varianza). Considerando lo anterior, y para nuestro ejemplo, entonces asignamos las siguientes distribuciones a priori:

$$\beta \sim \mathcal{N}(\mathbf{m}_0, \mathbf{T}_0), \quad \tau \sim \text{Gamma}(a_0, b_0),$$

donde β esta tiene una distribución a prior multivariante Gaussiana con \mathbf{m}_0 y \mathbf{T}_0 conocidos.

La distribución posterior para los parámetros del modelo es intratable analíticamente y el método de MCMC (a través de alguno de sus algoritmos) puede ayudar en la estimación a través del muestreo en sus iteraciones, en cambio INLA realiza aproximaciones hacia las marginales de los parámetros en forma eficiente y precisa (ver presentación en clases).

Ejemplo

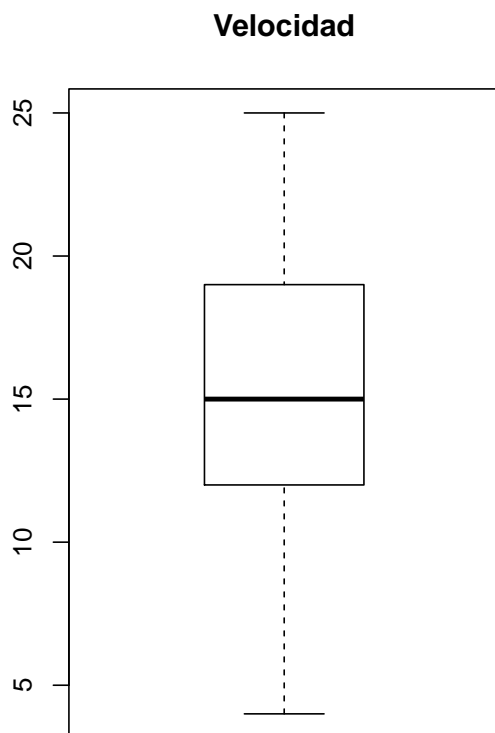
Vamos a utilizar un conjunto de datos disponibles en R.

```
head(cars) # Mostramos las primeras 6 observaciones
```

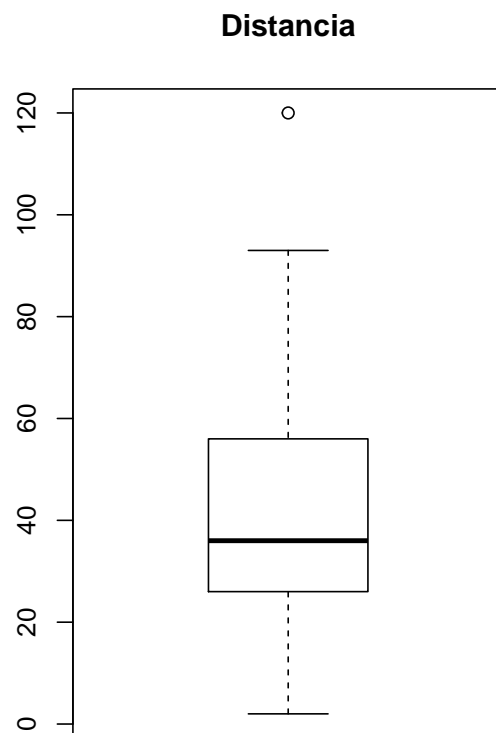
```
##   speed dist
## 1     4    2
## 2     4   10
## 3     7    4
## 4     7   22
## 5     8   16
## 6     9   10
```

Hacemos un gráfico `boxplot()` para identificar posibles outliers.

```
par(mfrow=c(1, 2))
boxplot(cars$speed, main="Velocidad", sub=paste("Outlier: ",
        boxplot.stats(cars$speed)$out))
boxplot(cars$dist, main="Distancia", sub=paste("Outlier rows: ",
        boxplot.stats(cars$dist)$out))
```



Outlier:



Outlier rows: 120

Vemos el grado de correlación existente entre las variables:

```
cor(cars$speed, cars$dist)
```

```
## [1] 0.8068949
```

Estimación: función `lm()`

Primero evaluaremos el modelo mediante la función `lm()` disponible en R.

```
formula = dist ~ speed
linear_model = lm(formula, data=cars)
round(coef(summary(linear_model)), 4)
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -17.5791      6.7584  -2.6011   0.0123
## speed        3.9324      0.4155   9.4640   0.0000
```

σ estimado:

```
round(summary(linear_model)$sigma, 4)
```

```
## [1] 15.3796
```

Estimación: INLA

En primera instancia asignamos una distribución a priori para $\beta \sim \mathcal{N}(0, 10^6)$ y para el parámetro $\log(\tau) \sim \log\text{Gamma}(1, 10^{-5})$. Mediante INLA tenemos:

```
library(INLA)
library(brinla)
inla_model = inla(formula, data = cars, control.compute = list(dic = TRUE, cpo = TRUE))
round(inla_model$summary.fixed, 4)
```

```
##              mean      sd 0.025quant 0.5quant 0.975quant      mode kld
## (Intercept) -17.5687 6.7465  -30.8684 -17.5691  -4.2786 -17.5693    0
## speed        3.9317 0.4148   3.1140  3.9317   4.7487  3.9318    0
```

Podemos ver la estimación para τ :

```
round(inla_model$summary.hyperpar, 4)
```

```
##              mean      sd 0.025quant 0.5quant
## Precision for the Gaussian observations 0.0044 9e-04   0.0029  0.0043
##              0.975quant      mode
## Precision for the Gaussian observations 0.0063 0.0042
```

Obtenemos σ para comparar con el modelo lineal frecuentista (`linear_mod`) previamente ajustado:

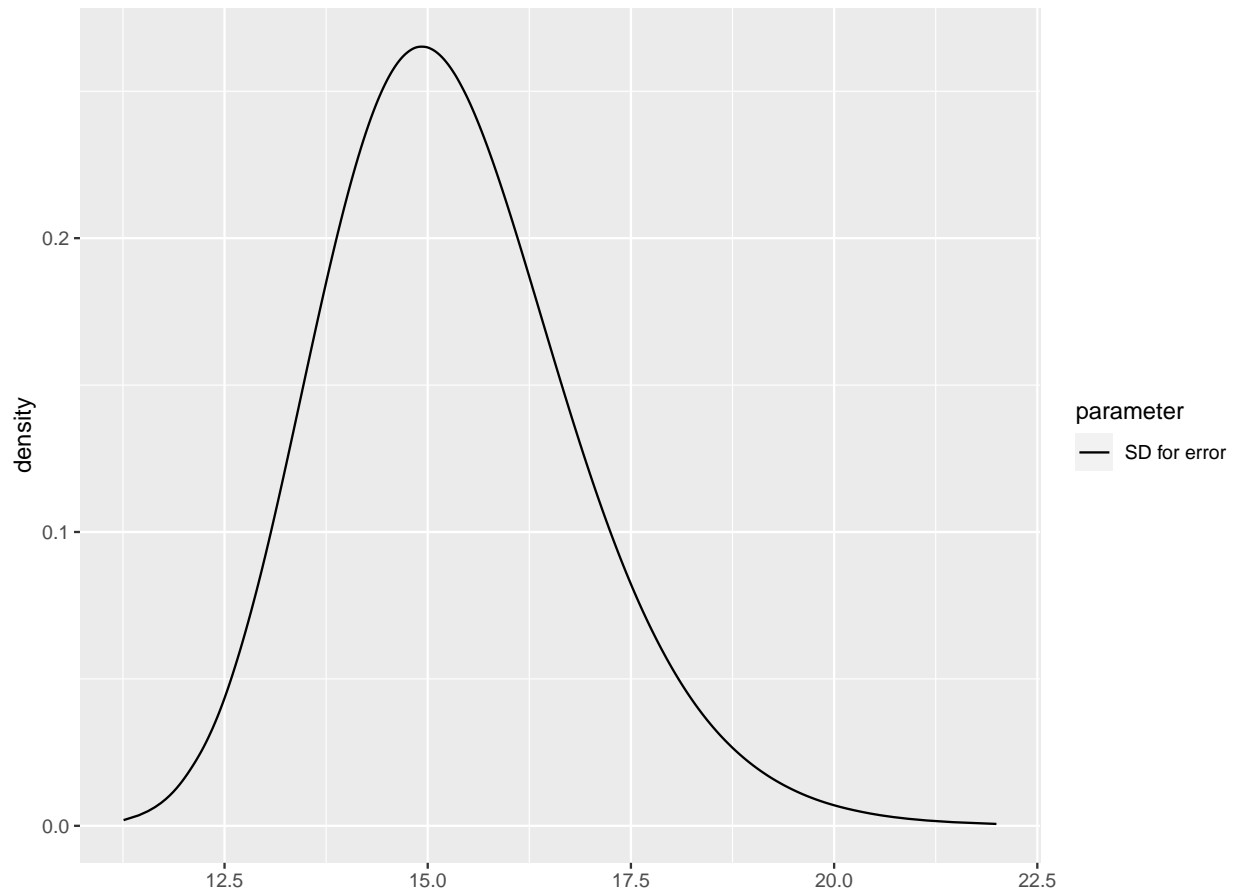
```
round(bri.hyperpar.summary(inla_model),4)
```

```
##                mean      sd  q0.025   q0.5  q0.975    mode
## SD for the Gaussian observations 15.2979 1.5536 12.6161 15.1683 18.7103 14.9249
```

Distribución posterior para σ

```
bri.hyperpar.plot(inla_model)
```

```
## Loading required package: ggplot2
```



Regresión logística con INLA

Teoría y aplicaciones

Joaquin Cavieres G.

1. Introducción

Dentro de la estadística Bayesiana el cálculo de los parámetros de interés generalmente suele realizarse mediante el método de MCMC. Este tipo de método utiliza diferentes algoritmos de estimación, por ejemplo, el algoritmo de Metropolis Hastings, el algoritmo de Gibbs Sampling o el algoritmo Hamiltoniano Monte Carlo. Estos algoritmos presentan buenas propiedades de convergencia y cuentan con diferentes formas de diagnósticos para evaluar sus estimaciones. Estos métodos de computación se encuentran disponibles en lenguajes de programación probabilística como **BUGS**, **JAGS** y **Stan**. Sin embargo, estos métodos generalmente presentan problemas de eficiencia computacional para modelos de estructuras complejas o modelos con una gran cantidad de observaciones, por lo que su utilización a veces se vuelve tediosa.

Por lo anterior es que Rue et al., (2009) propusieron una metodología basada en las aproximaciones de integrales de Laplace anidadas (Integrated Nested Laplace Approximation, **INLA**). **INLA** es un método de aproximación Bayesiana que frecuentemente es más eficiente que, por ejemplo, los métodos de MCMC tradicionales. Además, **INLA** está codificado plenamente en **R** y no es necesario comprender algún lenguaje de programación extra, los modelos están especificados previamente y contienen una amplia variedad de ejemplos sobre cada uno de ellos (ver <http://www.r-inla.org/>). Lo único necesario es tener un conocimiento medio de estadística Bayesiana para su aplicación.

En este documento se describirá brevemente la clase de modelos *Latent Gaussian Models* (LGM's) quienes representan el marco fundamental clave en **INLA** y el funcionamiento de esta librería para resolver un problema específico.

2. Latent Gaussian Models (LGM's)

Como se comentó anteriormente, **INLA** está construido en base a una clase específica de modelos, los Latent Gaussian Models (LGM's). Esta clase de modelos pueden ser utilizados en series temporales, modelos espacio-estado (state - space models), estadística espacial, modelos espacio - temporales, etc (Rue et al., 2009; Blangiardo and Cameletti, 2015).

En primera instancia consideremos a \mathbf{y} una variable aleatoria observada con una distribución de probabilidad de acuerdo a $\pi(\mathbf{y}; \boldsymbol{\theta})$, entonces, basandonos en las observaciones nosotros podemos estimar el parámetro $\boldsymbol{\theta}$. Bajo el paradigma frecuentista $\boldsymbol{\theta}$ es un vector de parámetros desconocidos y fijos que generalmente es estimado por el método de máxima verosimilitud. Si consideramos el paradigma Bayesiano en donde $\boldsymbol{\theta}$ denota a un vector que contiene a las variables aleatorias θ 's entonces podemos estimar este vector basado en la siguiente expresión:

$$\pi(\boldsymbol{\theta} \mid \mathbf{y}) \propto \pi(\mathbf{y} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta}) \quad (1)$$

donde $\pi(\boldsymbol{\theta})$ es la distribución a priori (prior) y $\pi(\mathbf{y} \mid \boldsymbol{\theta})$ es la función de verosimilitud para los datos observados.

Si tenemos una variable respuesta y_i con distribución perteneciente a la familia exponencial y su media μ_i es asociada con η_i mediante una función de enlace (link) $g(\cdot)$ tal que $\eta_i = g(\mu_i)$, entonces podemos modelar a y_i a través de un modelo lineal generalizado (GLM, por sus siglas en inglés) de la siguiente manera:

$$\eta_i = g(\mu_i) = \beta_0 + \sum_{j=1}^{n_\beta} \beta_j z_{ji}, \quad i = 1, \dots, n \quad (2)$$

donde β_0 y β_j son los coeficientes asociados a las covariables z_{ji} . La expresión anterior también puede ser extendida a un modelo aditivo generalizado (GAM, por sus siglas en inglés) como:

$$\eta_i = g(\mu_i) = \beta_0 + \sum_{l=1}^{n_f} f_l c_{li}, \quad i = 1, \dots, n \quad (3)$$

donde $g(\cdot)$ sigue siendo una función de enlace (*link*) y $f_l(\cdot)$ es el efecto no lineal asociado a las covariables c_{li} . Esta función $f_l(\cdot)$ puede adoptar distintas formas, como por ejemplo, un efecto no lineal como un random walk (Rue and Held, 2005), un Gaussian process (william and Rasmussen, 2006) o una función *smoothing* (Wang, 2011).

Finalmente podemos generalizar las estructuras previas en un LGM's de la siguiente manera:

$$\mathbf{x} = \{\boldsymbol{\beta}, f_l(\cdot), \boldsymbol{\eta}\} \quad (4)$$

en donde cada elemento de \mathbf{x} tiene una prior Gaussiana. INLA permite estimar modelos estadísticos complejos utilizando distribuciones pertenecientes a la familia exponencial (www.r-inla.org/models/likelihoods)

A través de una estructura jerárquica podemos representar un modelo general como:

$$\mathbf{y} \mid \mathbf{x}, \theta_1 \sim \prod_{i=1}^n \pi(y_i \mid x_i, \theta_1) \quad (5)$$

$$\mathbf{x} \mid \theta_2 \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{Q}^{-1}) \quad (6)$$

$$\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta}) \quad (7)$$

donde \mathbf{y} es condicionalmente independiente a \mathbf{x} y θ_1 , el *latent field* \mathbf{x} es un Gaussian Markov Random Field (GMRF) con matriz de precisión \mathbf{Q} y los hiperparámetros $\boldsymbol{\theta} = (\theta_1, \theta_2)$ tienen prior Gaussianas. La eficiencia computacional de INLA, a través del uso de los LGM's, se ve incrementada notablemente si se consideran modelos con dependencia espacial, espacio-temporal o modelos espacio-estado.

Nota: Ejecutando `names(inla.models())$likelihood` en R se pueden ver los diferentes modelos disponibles en INLA. En este documento sólo se presentará el clásico modelo de regresión logística como caso motivacional para el uso de esta librería.

3. Estimación de un GLM Bayesiano con respuesta binomial

La regresión logística es un caso particular de los GLM y nos permite estimar la probabilidad de que un evento aleatorio. El ejemplo más sencillo de este tipo de problemas es la estimación de la probabilidad de que al lanzar una moneda esta sea “cara”, con sólo dos posibles resultados “cara” y “sello”, por tanto asumimos que Y proviene de una distribución bernoulli con $\mathbb{P}(Y = 1) = \pi$ con densidad dada por:

$$f(y) = \exp\left\{y \log\left(\frac{\pi}{1-\pi}\right) + \log(1-\pi)\right\} \quad (8)$$

El parámetro canónico es θ , que es igual al *logit* de π , esto es $\log(\frac{\pi}{1-\pi})$ y un parámetro de dispersión $\phi = 1$. Lo anterior obtiene como resultado que π es la media y la varianza es $\pi(1-\pi)$ para un modelo de la forma:

$$\mathbf{y} \sim \text{Bernoulli}(\boldsymbol{\pi}) \quad (9)$$

$$\text{logit}(\boldsymbol{\pi}) = \mathbf{x}^T \boldsymbol{\beta} = \beta_0 + \beta_1 x_i + \dots, + \beta_p x_{ip} \quad (10)$$

Para ejemplificar como estimar los parámetros en un modelo logístico con INLA vamos a considerar los datos de los sobrevivientes del Titanic los cuales contienen la siguiente información:

- **ID Pasajero**: ID correspondiente a cada pasajero
- **Sobreviviente**: Si el pasajero sobrevivió al naufragio (0 (no) y 1 (si))
- **Clase**: Clase en la cual el pasajero se encontraba dentro del barco (primera clase (1), segunda clase (2) y tercera clase (3))
- **Sexo**: Sexo del pasajero
- **Edad**: Edad del pasajero
- **Num_par1**: Número de hermanos (as) o cónyuges fuera del barco
- **Num_par2**: Número de padres / hijos a bordo
- **Embarque**: Puerto de embarque (C = Cherbourg; Q = Queenstown; S = Southampton)

```
#=====
#                               GLM logístico con INLA
#=====
library(INLA)
library(dplyr)
library(ggplot2)
library(tictoc)
library(brinla)
library(ggregplot)
```

```
data = read.csv("data_titanic.csv", header=T)
head(data)
```

```
##   ID Sobreviviente Clase   Sexo Edad Num_par1 Num_par2 Embarque
## 1  1             0     3  male  22         1         0         S
## 2  2             1     1 female  38         1         0         C
## 3  3             1     3 female  26         0         0         S
## 4  4             1     1 female  35         1         0         S
## 5  5             0     3  male  35         0         0         S
## 6  6             0     3  male  NA         0         0         Q
```

```
summary(data)
```

```
##           ID           Sobreviviente           Clase           Sexo           Edad
## Min.      : 1.0   Min.      :0.0000   Min.      :1.000   female:314   Min.      : 0.42
## 1st Qu.:223.5   1st Qu.:0.0000   1st Qu.:2.000   male  :577   1st Qu.:20.12
## Median :446.0   Median :0.0000   Median :3.000                                Median :28.00
## Mean    :446.0   Mean    :0.3838   Mean    :2.309                                Mean    :29.70
## 3rd Qu.:668.5   3rd Qu.:1.0000   3rd Qu.:3.000                                3rd Qu.:38.00
## Max.    :891.0   Max.    :1.0000   Max.    :3.000                                Max.    :80.00
##                                     NA's    :177
##           Num_par1           Num_par2           Embarque
## Min.      :0.000   Min.      :0.0000           : 2
## 1st Qu.:0.000   1st Qu.:0.0000   C:168
## Median :0.000   Median :0.0000   Q: 77
## Mean    :0.523   Mean    :0.3816   S:644
## 3rd Qu.:1.000   3rd Qu.:0.0000
## Max.    :8.000   Max.    :6.0000
##
```

```
#data$Sobreviviente = as.factor(data$Sobreviviente)
data$Sexo = as.factor(data$Sexo)
data$Clase = as.factor(data$Clase)
data$Embarque = as.factor(data$Embarque)
```

```
# En la variable "Edad" tenemos 177 registros con valores "NA"
# Reemplazaremos estos valores NA por la media de las edades (sólo para completar
# datos faltantes)
data = data %>% mutate(Edad = replace(Edad,is.na(Edad), median(Edad, na.rm = T)))
```

```
# Seleccionaremos las variables más influyentes mediante el algoritmo stepwise
modelo = glm(Sobreviviente ~1 , data = data, na.action=na.exclude, family=binomial())
summary(modelo, correlation=F)
```

```
##
## Call:
## glm(formula = Sobreviviente ~ 1, family = binomial(), data = data,
##      na.action = na.exclude)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.9841  -0.9841  -0.9841   1.3839   1.3839
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.47329    0.06889  -6.87  6.4e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1186.7  on 890  degrees of freedom
## Residual deviance: 1186.7  on 890  degrees of freedom
## AIC: 1188.7
##
## Number of Fisher Scoring iterations: 4
```

```
# Funcion step
step(modelo, scope=~ Clase + Sexo + Edad + Num_par1 + Num_par2 + Embarque)
```

```
## Start:  AIC=1188.66
## Sobreviviente ~ 1
##
##           Df Deviance    AIC
## + Sexo      1    917.8  921.8
## + Clase     2   1083.1 1089.1
## + Embarque  3   1157.0 1165.0
## + Num_par2  1   1180.8 1184.8
## + Edad      1   1182.9 1186.9
## <none>       1186.7 1188.7
## + Num_par1  1   1185.5 1189.5
##
## Step:  AIC=921.8
## Sobreviviente ~ Sexo
##
##           Df Deviance    AIC
## + Clase     2    826.89  834.89
```

```

## + Embarque 3 897.34 907.34
## + Num_par1 1 904.69 910.69
## + Num_par2 1 914.43 920.43
## <none> 917.80 921.80
## + Edad 1 917.25 923.25
## - Sexo 1 1186.66 1188.66
##
## Step: AIC=834.89
## Sobreviviente ~ Sexo + Clase
##
##          Df Deviance      AIC
## + Edad 1 805.53 815.53
## + Num_par1 1 819.15 829.15
## + Embarque 3 818.54 832.54
## <none> 826.89 834.89
## + Num_par2 1 825.06 835.06
## - Clase 2 917.80 921.80
## - Sexo 1 1083.11 1089.11
##
## Step: AIC=815.53
## Sobreviviente ~ Sexo + Clase + Edad
##
##          Df Deviance      AIC
## + Num_par1 1 791.23 803.23
## + Embarque 3 797.58 813.58
## + Num_par2 1 801.92 813.92
## <none> 805.53 815.53
## - Edad 1 826.89 834.89
## - Clase 2 917.25 923.25
## - Sexo 1 1046.99 1054.99
##
## Step: AIC=803.23
## Sobreviviente ~ Sexo + Clase + Edad + Num_par1
##
##          Df Deviance      AIC
## <none> 791.23 803.23
## + Embarque 3 785.71 803.71
## + Num_par2 1 790.78 804.78
## - Num_par1 1 805.53 815.53
## - Edad 1 819.15 829.15
## - Clase 2 902.20 910.20
## - Sexo 1 1044.35 1054.35
##
## Call: glm(formula = Sobreviviente ~ Sexo + Clase + Edad + Num_par1,
## family = binomial(), data = data, na.action = na.exclude)
##
## Coefficients:

```

```
## (Intercept)      Sexomale      Clase2      Clase3      Edad      Num_par1
##      4.00350      -2.73959      -1.18051      -2.35213      -0.03957      -0.35456
##
## Degrees of Freedom: 890 Total (i.e. Null); 885 Residual
## Null Deviance:      1187
## Residual Deviance: 791.2      AIC: 803.2
```

```
#=====
#
# GLM frecuentista
#=====
glm = glm(Sobreviviente ~ Sexo + Clase + Edad + Num_par1, data = data,
          family=binomial())

# Parámetros del modelo
round(coef(summary(glm)), 4)
```

```
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept)   4.0035     0.3972  10.0789   0e+00
## Sexomale      -2.7396     0.1941 -14.1177   0e+00
## Clase2        -1.1805     0.2614  -4.5168   0e+00
## Clase3        -2.3521     0.2434  -9.6643   0e+00
## Edad          -0.0396     0.0078  -5.0760   0e+00
## Num_par1      -0.3546     0.1036  -3.4235   6e-04
```

```
#=====
#
# GLM con INLA
#=====
inla = inla(Sobreviviente ~ Sexo + Clase + Edad + Num_par1, family='binomial',
            Ntrials = 1,
            data = data,
            control.family=list(link='logit'),
            control.compute=list(dic=TRUE, cpo=TRUE, waic=TRUE))

# Parámetros del modelo
round(inla$summary.fixed[, c(1:5)], 4)
```

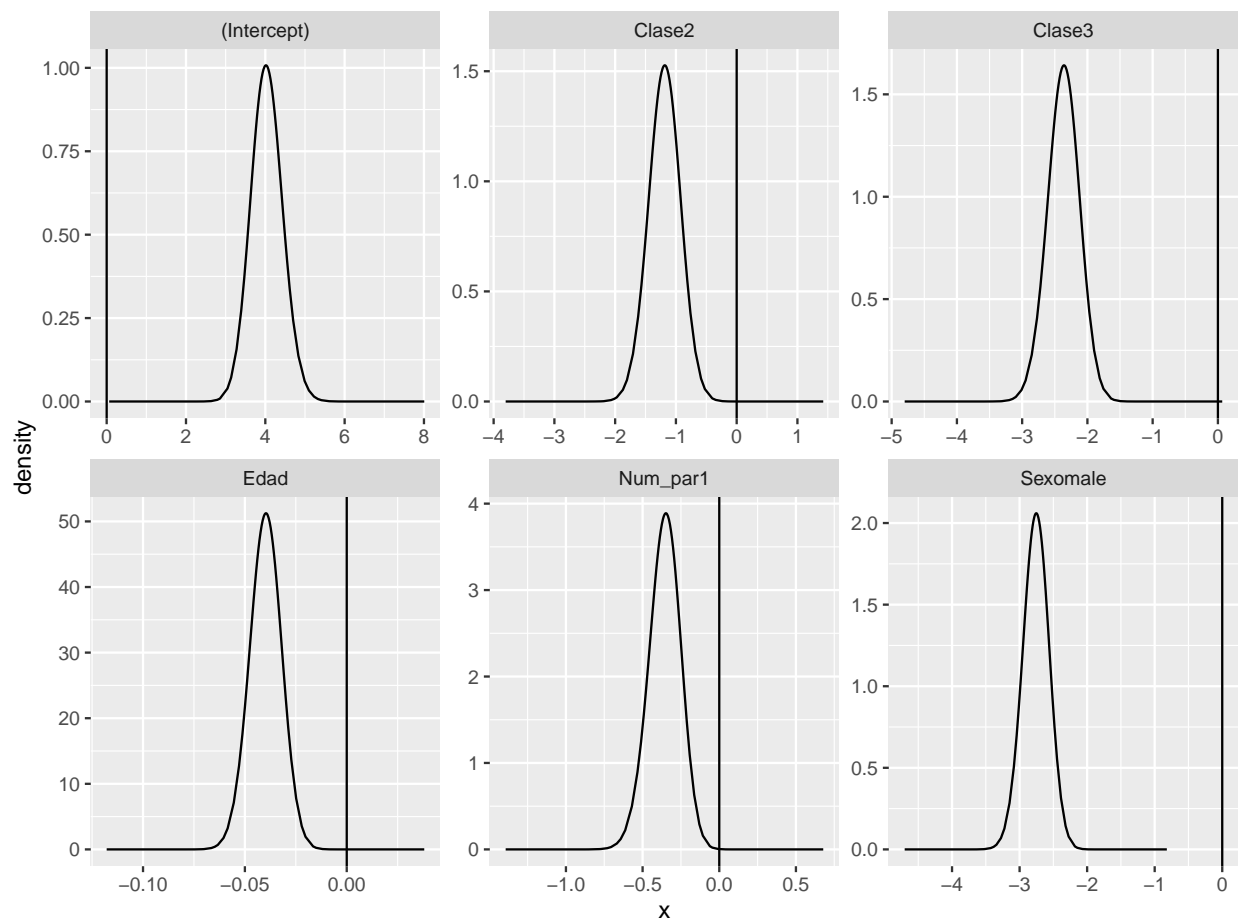
```
##           mean      sd 0.025quant 0.5quant 0.975quant
## (Intercept)  4.0355 0.3971      3.2755      4.0290      4.8341
## Sexomale     -2.7607 0.1940     -3.1490     -2.7581     -2.3868
## Clase2       -1.1882 0.2614     -1.7067     -1.1865     -0.6801
## Clase3       -2.3670 0.2434     -2.8530     -2.3641     -1.8971
## Edad         -0.0399 0.0078     -0.0554     -0.0398     -0.0248
## Num_par1     -0.3576 0.1035     -0.5698     -0.3544     -0.1635
```

Como se aprecia en el cuadro anterior, las estimaciones de los parámetros son similares a las obtenidas con el GLM frecuentista, sólo que INLA además presenta los intervalos de credibilidad (quantiles) del 0.25 y 0.75.

El siguiente gráfico presenta las distribuciones posteriores para los parámetros relevantes dentro del modelo. La línea recta presenta el valor 0 otorgado como valor de entrada a priori (Gaussiana) por INLA pero estos pueden ser cambiados según el conocimiento previo del analista. Una consideración especial sobre este punto es que las precisiones son representadas en forma interna en escala logarítmica. Esto es eficiente en términos de estimación ya que el parámetro no se encuentra acotado. La especificación de las priors son generalmente sobre las ‘hyperpriors’ que gobiernan el comportamiento del LGM’s. Por ejemplo, dentro de un modelo espacial es conveniente usar distribuciones a priori para los parámetros de la función de Matérn, generalmente los hiperparámetros τ y κ .

La declaración de las priors dentro de INLA puede realizarse argumentando `hyper` dentro de la función `control.family`.

```
# Posteriores para los parámetros del modelo
bri.fixed.plot(inla)
```

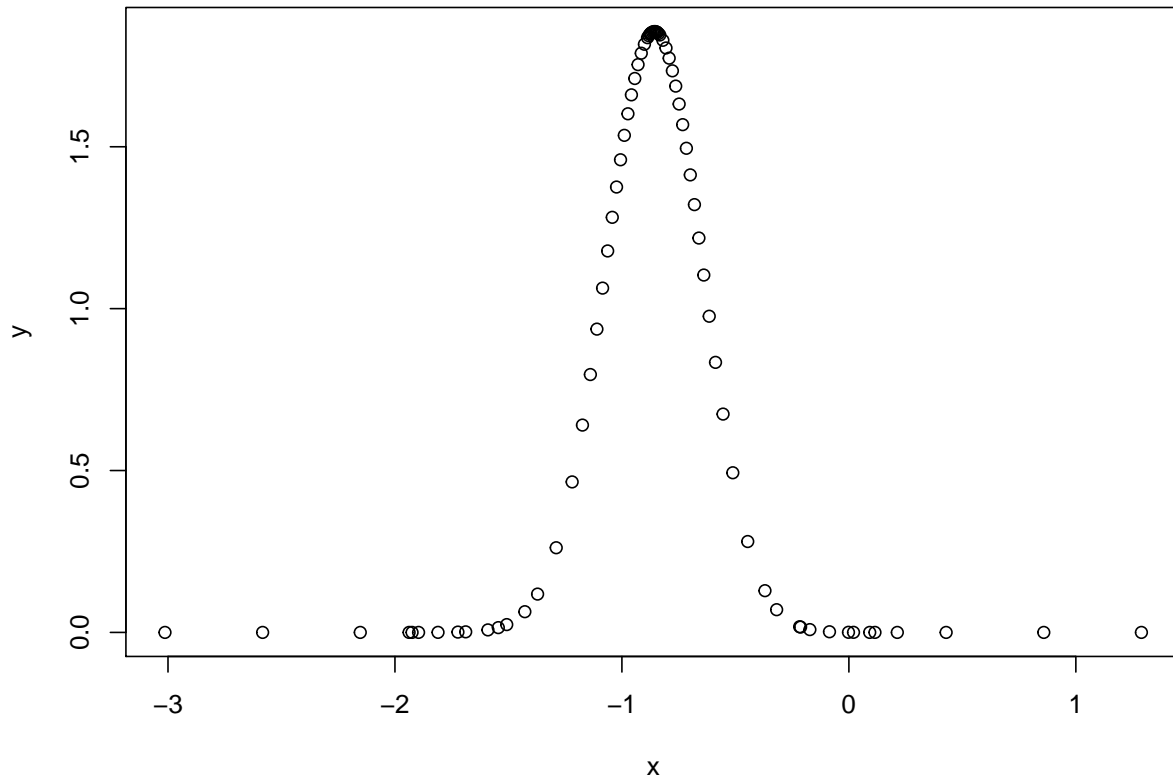


La predicción para un valor específico se puede hacer de forma sencilla declarando NA en donde se quiere obtener el nuevo valor.

```
#=====
#                               Predicción
#=====
newdata = data
newdata[7, "Sobreviviente"] = NA

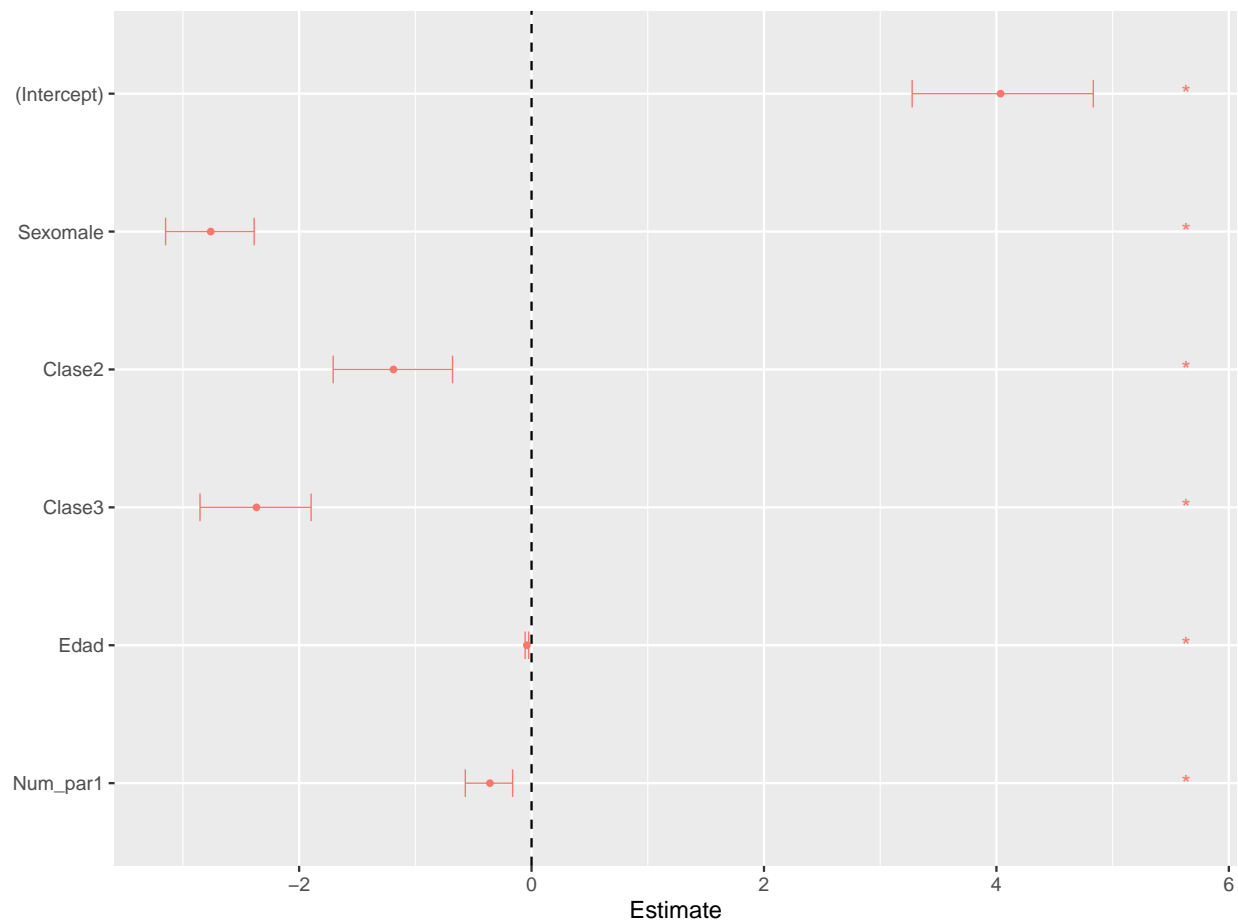
# re-run the model
res.predict = inla(Sobreviviente ~ Sexo + Clase + Edad + Num_par1, family='binomial', Ntrials = 1,
                   data = newdata,
                   control.family=list(link='logit'),
                   control.compute=list(dic=TRUE, cpo=TRUE, waic=TRUE, config = TRUE))

# marginal posterior for the linear predictor
eta7 = res.predict$marginals.linear.predictor[[7]]
plot(eta7)
```



La siguiente figura muestra los intervalos de credibilidad para los parámetros asociados a cada variable dentro del modelo. Estas representaciones son directas ya que los parámetros son tratados como variables aleatorias.

```
#=====
#                               Plot de intervalos de credibilidad
#=====
Efxplot(inla)
```



4. Conclusión

Se presenta un breve análisis relacionado a la teoría detrás de los LGM's, su aplicación y resolución con INLA. Se presentaron las estimaciones para un problema con respuesta binomial y resuelto a través de un GLM frecuentista y un GLM Bayesiano con INLA . Los valores fueron similares en ambas metodologías.

INLA considera a los LGM's para su funcionalidad y presenta buenas propiedades de eficiencia computacional. Esta eficiencia se ve reflejada de mejor manera cuando se consideran modelos con estructuras complejas y con dependencia espacial y/o temporal en sus observaciones.

5. Bibliografía

Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian Models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* 71, 319-392.

Blangiardo, M., & Cameletti, M. (2015). *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley & Sons.

Wang, X., Ryan, Y. Y., & Faraway, J. J. (2018). *Bayesian regression modeling with INLA*. CRC Press.

Rue, H., & Held, L. (2005). *Gaussian Markov random fields: theory and applications*. CRC press.

Williams, C. K., & Rasmussen, C. E. (2006). *Gaussian processes for machine learning* (Vol. 2, No. 3, p. 4). Cambridge, MA: MIT press.

Wang, Y. (2011). *Smoothing splines: methods and applications*. CRC Press.