EPG3343 - Seminario de Estadística III

Clase 10

Jonathan Acosta

Pontificia Universidad Católica de Chile

Segundo Semestre, 2022

Esquema

1. INLA: Integrated Nested Laplace Approximations
   - Inferencia bayesiana aproximada con INLA

2. Regresión Bayesianos
   - Modelos de Regresión
   - Predicción

Esquema

1. INLA: Integrated Nested Laplace Approximations
   - Inferencia bayesiana aproximada con INLA

2. Regresión Bayesianos
   - Modelos de Regresión
   - Predicción

INLA

- El algoritmo INLA, propuesto por Rue et al.(2009), es un algoritmo determinista para la inferencia bayesiana.

INLA

- El algoritmo INLA, propuesto por Rue et al.(2009), es un algoritmo determinista para la inferencia bayesiana.
- A diferencia de los métodos de simulación, tales como MCMC, el INLA es especialmente diseñado para modelos latentes Gausianos.

INLA

- El algoritmo INLA, propuesto por Rue et al.(2009), es un algoritmo determinista para la inferencia bayesiana.
- A diferencia de los métodos de simulación, tales como MCMC, el INLA es especialmente diseñado para modelos latentes Gausianos.
- Comparado con el MCMC proporciona precisión en menor tiempo de computo.

La Aproximación de Laplace

La Aproximación de Laplace

- Suponga que el interés es calcular

$$\int f(x)dx = \int \exp(\log(f(x))dx$$

La Aproximación de Laplace

- Suponga que el interés es calcular

$$\int f(x)dx = \int \exp(\log(f(x))dx$$

- Considere la aproximación de Taylor de orden 2 de $\log(f(x))$ en $x = x_0$, es decir,

$$\log(f(x)) \approx \log(f(x_0)) + \frac{\partial f(x_0)}{\partial x}(x - x_0) + \frac{\partial^2 f(x_0)}{2\partial x}(x - x_0)^2$$

La Aproximación de Laplace

- Suponga que el interés es calcular

$$\int f(x)dx = \int \exp(\log(f(x))dx$$

- Considere la aproximación de Taylor de orden 2 de $\log(f(x))$ en $x = x_0$, es decir,

$$\log(f(x)) \approx \log(f(x_0)) + \frac{\partial f(x_0)}{\partial x}(x - x_0) + \frac{\partial^2 f(x_0)}{2\partial x}(x - x_0)^2$$

- Si x_0 corresponde a la moda de $\log(f(x))$, denotada por x^* entonces

$$\log(f(x)) \approx \log(f(x^*)) + \frac{\partial^2 f(x^*)}{2\partial x}(x - x^*)^2$$

La Aproximación de Laplace

- Suponga que el interés es calcular

$$\int f(x)dx = \int \exp(\log(f(x))dx$$

- Considere la aproximación de Taylor de orden 2 de $\log(f(x))$ en $x = x_0$, es decir,

$$\log(f(x)) \approx \log(f(x_0)) + \frac{\partial f(x_0)}{\partial x}(x - x_0) + \frac{\partial^2 f(x_0)}{2\partial x}(x - x_0)^2$$

- Si x_0 corresponde a la moda de $\log(f(x))$, denotada por x^* entonces

$$\log(f(x)) \approx \log(f(x^*)) + \frac{\partial^2 f(x^*)}{2\partial x}(x - x^*)^2$$

INLA

- Luego, la integral original se puede aproximar mediante

$$\int f(x)dx \approx \exp(\log(f(x^*) \int \exp\left(-\frac{(x-x^*)^2}{2\sigma^{2*}}\right) dx$$

donde $\sigma^{2*} = -\left(\frac{\partial^2 f(x^*)}{2\partial x}\right)^{-1}$.

INLA

- Luego, la integral original se puede aproximar mediante

$$\int f(x)dx \approx \exp(\log(f(x^*) \int \exp\left(-\frac{(x-x^*)^2}{2\sigma^{2*}}\right) dx$$

donde $\sigma^{2*} = -\left(\frac{\partial^2 f(x^*)}{2\partial x}\right)^{-1}$.

- Es decir, la integral se puede aproximar por una distribución Normal de media x^* y varianza σ^{2*}.

INLA

- Luego, la integral original se puede aproximar mediante

$$\int f(x)dx \approx \exp(\log(f(x^*) \int \exp\left(-\frac{(x-x^*)^2}{2\sigma^{2*}}\right) dx$$

donde $\sigma^{2*} = -\left(\frac{\partial^2 f(x^*)}{2\partial x}\right)^{-1}$.

- Es decir, la integral se puede aproximar por una distribución Normal de media x^* y varianza σ^{2*}.

- Si la integral esta definida en (α, β) entonces

$$\int_\alpha^\beta f(x)dx \approx \sqrt{2\pi\sigma^{2*}} f(x^*)[\Phi(\beta) - \Phi(\alpha)]$$

INLA

Por ejemplo, para $X \sim \text{Gamma}(a, b)$ con densidad

$$f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx)$$

se tiene que

$$\begin{aligned}
\log(f(x)) &= (a-1)\log(x) - bx + \text{cte} \\
\frac{\partial \log(f(x))}{\partial x} &= \frac{a-1}{x} - b \\
\frac{\partial^2 \log(f(x))}{\partial x^2} &= -\frac{a-1}{x^2}.
\end{aligned}$$

Obteniendo para $a > 1$,

$$x^* = \frac{a-1}{b} \quad \wedge \quad \sigma^{2*} = \frac{a-1}{b^2}$$

INLA

- En el contexto bayesiano se asume que

$$\begin{aligned}
\boldsymbol{y} \mid \boldsymbol{\theta} &\sim p(\boldsymbol{y} \mid \boldsymbol{\theta}\boldsymbol{\psi}) = \prod_{i=1}^{n} p(y_i \mid \theta_i, \boldsymbol{\psi}) \\
\boldsymbol{\theta} \mid \boldsymbol{\psi} &\sim p(\boldsymbol{\theta} \mid \boldsymbol{\psi}) \\
\boldsymbol{\psi} &\sim p(\boldsymbol{\psi})
\end{aligned}$$

INLA

- En el contexto bayesiano se asume que

$$
\begin{aligned}
\boldsymbol{y} \mid \boldsymbol{\theta} &\sim p(\boldsymbol{y} \mid \boldsymbol{\theta}\boldsymbol{\psi}) = \prod_{i=1}^{n} p(y_i \mid \theta_i, \boldsymbol{\psi}) \\
\boldsymbol{\theta} \mid \boldsymbol{\psi} &\sim p(\boldsymbol{\theta} \mid \boldsymbol{\psi}) \\
\boldsymbol{\psi} &\sim p(\boldsymbol{\psi})
\end{aligned}
$$

- Donde $\boldsymbol{\theta}$ representa los parámetros de las covariables u otros efectos.

INLA

- En el contexto bayesiano se asume que

$$
\begin{aligned}
\boldsymbol{y} \mid \boldsymbol{\theta} &\sim p(\boldsymbol{y} \mid \boldsymbol{\theta}\boldsymbol{\psi}) = \prod_{i=1}^{n} p(y_i \mid \theta_i, \boldsymbol{\psi}) \\
\boldsymbol{\theta} \mid \boldsymbol{\psi} &\sim p(\boldsymbol{\theta} \mid \boldsymbol{\psi}) \\
\boldsymbol{\psi} &\sim p(\boldsymbol{\psi})
\end{aligned}
$$

- Donde $\boldsymbol{\theta}$ representa los parámetros de las covariables u otros efectos.
- Se asumirá que la distribución a priori de $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, Q^{-1}(\boldsymbol{\psi}))$, es decir, $Q(\boldsymbol{\psi})$ es la matriz de precisión.

INLA

- Se asumirá que las componentes de $\boldsymbol{\theta}$ son condicionalmente independientes, es decir,

$$p(\theta_i, \theta_j \mid \boldsymbol{\theta}_{-(ij)}) = p(\theta_i \mid \boldsymbol{\theta}_{-(ij)}) p(\theta_j \mid \boldsymbol{\theta}_{-(ij)})$$

INLA

- Se asumirá que las componentes de $\boldsymbol{\theta}$ son condicionalmente independientes, es decir,

$$p(\theta_i, \theta_j \mid \boldsymbol{\theta}_{-(ij)}) = p(\theta_i \mid \boldsymbol{\theta}_{-(ij)}) p(\theta_j \mid \boldsymbol{\theta}_{-(ij)})$$

- Esta especificación es conocida como *Gaussian Markov random field*.

INLA

- Se asumirá que las componentes de $\boldsymbol{\theta}$ son condicionalmente independientes, es decir,

$$p(\theta_i, \theta_j \mid \boldsymbol{\theta}_{-(ij)}) = p(\theta_i \mid \boldsymbol{\theta}_{-(ij)}) p(\theta_j \mid \boldsymbol{\theta}_{-(ij)})$$

- Esta especificación es conocida como *Gaussian Markov random field*.

- Luego,

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\psi} \mid \boldsymbol{y}) &\propto p(\boldsymbol{\psi}) p(\boldsymbol{\theta} \mid \boldsymbol{\psi}) p(\boldsymbol{y} \mid \boldsymbol{\theta}, \boldsymbol{\psi}) \\ &\propto p(\boldsymbol{\psi}) \sqrt{|Q(\boldsymbol{\psi})|} \exp\left(-\frac{1}{2} \boldsymbol{\theta}^\top Q(\boldsymbol{\psi}) \boldsymbol{\theta} \right. \\ &\qquad \left. + \sum_{i=1}^{n} \log(p(y_i \mid \theta_i, \boldsymbol{\psi}))\right) \end{aligned}$$

Esquema

1. INLA: Integrated Nested Laplace Approximations
   - Inferencia bayesiana aproximada con INLA

2. Regresión Bayesianos
   - Modelos de Regresión
   - Predicción

Regresión Lineal Bayesiana

- Sea $\boldsymbol{y} = (y_1, \ldots, y_n)$ un vector de n observaciones (respuestas).

Regresión Lineal Bayesiana

- Sea $\boldsymbol{y} = (y_1, \ldots, y_n)$ un vector de n observaciones (respuestas).
- Sea $\boldsymbol{X} = (\mathbf{1}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_M)$ la matriz de diseño de orden $n \times (M + 1)$.

Regresión Lineal Bayesiana

- Sea $\boldsymbol{y} = (y_1, \ldots, y_n)$ un vector de n observaciones (respuestas).
- Sea $\boldsymbol{X} = (\mathbf{1}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_M)$ la matriz de diseño de orden $n \times (M+1)$.
- El modelo de regresión lineal asume que

$$y_i \sim \mathcal{N}(\mu_i, \sigma^2) \quad \text{donde} \quad \mu_i = \beta_0 + \sum_{j=1}^{M} \beta_j X_{ij}$$

Regresión Lineal Bayesiana

- Sea $\boldsymbol{y} = (y_1, \dots, y_n)$ un vector de n observaciones (respuestas).
- Sea $\boldsymbol{X} = (\mathbf{1}, \boldsymbol{x}_1, \dots, \boldsymbol{x}_M)$ la matriz de diseño de orden $n \times (M+1)$.
- El modelo de regresión lineal asume que

$$y_i \sim \mathcal{N}(\mu_i, \sigma^2) \quad \text{donde} \quad \mu_i = \beta_0 + \sum_{j=1}^{M} \beta_j X_{ij}$$

- Equivalentemente,

$$\mathbb{E}[Y_i \mid \boldsymbol{\beta}, \boldsymbol{x}_i) = \beta_0 + \sum_{j=1}^{M} \beta_j X_{ij} = \boldsymbol{X}\boldsymbol{\beta}$$

Regresión Lineal Bayesiana

- Sea $\boldsymbol{y} = (y_1, \ldots, y_n)$ un vector de n observaciones (respuestas).
- Sea $\boldsymbol{X} = (\mathbf{1}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_M)$ la matriz de diseño de orden $n \times (M+1)$.
- El modelo de regresión lineal asume que

$$y_i \sim \mathcal{N}(\mu_i, \sigma^2) \quad \text{donde} \quad \mu_i = \beta_0 + \sum_{j=1}^{M} \beta_j X_{ij}$$

- Equivalentemente,

$$\mathbb{E}[Y_i \mid \boldsymbol{\beta}, \boldsymbol{x}_i) = \beta_0 + \sum_{j=1}^{M} \beta_j X_{ij} = \boldsymbol{X}\boldsymbol{\beta}$$

- En el enfoque clásico $\boldsymbol{\beta}$ y σ^2 son constantes y se estiman mediante máxima verosimilitud.

Regresión Lineal Bayesiana

- Ahora en el enfoque bayasiano se asumirá que $\boldsymbol{\beta}$ y σ^2 son variables aleatorias, por lo que, debemos asignar una distribución priori.

Regresión Lineal Bayesiana

- Ahora en el enfoque bayasiano se asumirá que $\boldsymbol{\beta}$ y σ^2 son variables aleatorias, por lo que, debemos asignar una distribución priori.
- Por ejemplo, una típica elección es

$$\begin{aligned} \beta_j &\sim \mathcal{N}(0, 10^6), \quad j = 1, \ldots, M \\ \log(\tau) = \log(1/\sigma^2) &\sim \text{logGamma})(1, 10^{-5}) \end{aligned}$$

Regresión Lineal Bayesiana

- Ahora en el enfoque bayasiano se asumirá que $\boldsymbol{\beta}$ y σ^2 son variables aleatorias, por lo que, debemos asignar una distribución priori.

- Por ejemplo, una típica elección es

$$
\begin{aligned}
\beta_j &\sim \mathcal{N}(0, 10^6), \quad j = 1, \ldots, M \\
\log(\tau) = \log(1/\sigma^2) &\sim \text{logGamma})(1, 10^{-5})
\end{aligned}
$$

- El objetivo es obtener la distribución posterior de $\boldsymbol{\beta}$ y σ^2

Regresión Lineal Bayesiana

- Asumiendo prioris no informativa en $\boldsymbol{\beta}$ y σ^2,

$$p(\beta_j) = 1, \quad j = 1, \ldots, M, \quad \wedge \quad p(\log(\sigma^2)) = 1.$$

Regresión Lineal Bayesiana

- Asumiendo prioris no informativa en $\boldsymbol{\beta}$ y σ^2,

$$p(\beta_j) = 1, \quad j = 1, \ldots, M, \quad \wedge \quad p(\log(\sigma^2)) = 1.$$

- La distribución posterior de $\boldsymbol{\beta}$ es una normal multivariada:

$$\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{y}, \boldsymbol{X} \sim \mathcal{N}\left((\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y},\ \sigma^2 (\boldsymbol{x}^\top)^{-1}\right)$$

Regresión Lineal Bayesiana

- Asumiendo prioris no informativa en $\boldsymbol{\beta}$ y σ^2,

$$p(\beta_j) = 1, \quad j = 1, \dots, M, \quad \wedge \quad p(\log(\sigma^2)) = 1.$$

- La distribución posterior de $\boldsymbol{\beta}$ es una normal multivariada:

$$\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{y}, \boldsymbol{X} \sim \mathcal{N}\left((\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y},\ \sigma^2 (\boldsymbol{x}^\top)^{-1}\right)$$

- Como esta posterior depende de σ^2, integrando obtenemos la distribución marginal posterior de β_j, es decir,

$$p(\beta_j \mid \boldsymbol{y}, \boldsymbol{X}) = \int_0^\infty p(\beta_j \mid \sigma^2, \boldsymbol{y}, \boldsymbol{X}) d\sigma^2$$

Regresión Lineal Bayesiana

- Asumiendo prioris no informativa en $\boldsymbol{\beta}$ y σ^2,

$$p(\beta_j) = 1, \quad j = 1, \ldots, M, \quad \wedge \quad p(\log(\sigma^2)) = 1.$$

- La distribución posterior de $\boldsymbol{\beta}$ es una normal multivariada:

$$\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{y}, \boldsymbol{X} \sim \mathcal{N}\left((\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y},\ \sigma^2 (\boldsymbol{x}^\top)^{-1}\right)$$

- Como esta posterior depende de σ^2, integrando obtenemos la distribución marginal posterior de β_j, es decir,

$$p(\beta_j \mid \boldsymbol{y}, \boldsymbol{X}) = \int_0^\infty p(\beta_j \mid \sigma^2, \boldsymbol{y}, \boldsymbol{X}) d\sigma^2$$

- Esto nos lleva a,

$$\beta_j \mid \boldsymbol{y}, \boldsymbol{X} \sim t_{n-M}\left(\widehat{\beta}_j, se(\widehat{\beta}_j)^2\right)$$

donde $\widehat{\beta}_j$ es EMV de β_j y $se(\widehat{\beta}_j)$ su respectiva desviación estándar.

GLM Bayesiano

- La regresión lineal es un caso particular de los modelos lineales generalizados, donde se asume que Y_i pertenece a la familia exponencial.

GLM Bayesiano

- La regresión lineal es un caso particular de los modelos lineales generalizados, donde se asume que Y_i pertenece a la familia exponencial.

- Si η representa el parámetro natural de la familia exponencial, entonces se asume que

$$\eta_i = g(\mu_i) = \beta_0 + \sum_{j=1}^{M} \beta_j X_{ij} = \boldsymbol{X\beta}$$

GLM Bayesiano

- La regresión lineal es un caso particular de los modelos lineales generalizados, donde se asume que Y_i pertenece a la familia exponencial.
- Si η representa el parámetro natural de la familia exponencial, entonces se asume que

$$\eta_i = g(\mu_i) = \beta_0 + \sum_{j=1}^{M} \beta_j X_{ij} = \boldsymbol{X\beta}$$

- Por ejemplo, en el caso binomial (Bernoulli) se tiene que

$$\begin{aligned} Y_i &\sim \mathrm{Bin}(n_i, \pi_i) \\ \eta_i = \mathrm{logit}(\pi_i) &= \beta_0 + \sum_{j=1}^{M} \beta_j X_{ij} = \boldsymbol{X\beta} \end{aligned}$$

GLM Bayesiano

- La regresión lineal es un caso particular de los modelos lineales generalizados, donde se asume que Y_i pertenece a la familia exponencial.

- Si η representa el parámetro natural de la familia exponencial, entonces se asume que

$$\eta_i = g(\mu_i) = \beta_0 + \sum_{j=1}^{M} \beta_j X_{ij} = \boldsymbol{X}\boldsymbol{\beta}$$

- Por ejemplo, en el caso binomial (Bernoulli) se tiene que

$$\begin{aligned} Y_i &\sim \text{Bin}(n_i, \pi_i) \\ \eta_i = \text{logit}(\pi_i) &= \beta_0 + \sum_{j=1}^{M} \beta_j X_{ij} = \boldsymbol{X}\boldsymbol{\beta} \end{aligned}$$

- Finalmente se asume una distribución a priori para $\beta_0, \beta_1, \ldots, \beta_M$, la cual típicamente es una Normal con varianza grande.

Predicción

- Suponga que $y^\star$ representa un valor futuro o nuevo.

Predicción

- Suponga que $y^\star$ representa un valor futuro o nuevo.
- Así, la densidad de esta nueva observación, dada la muestra queda como:

$$
\begin{aligned}
p(y^\star \mid \boldsymbol{y}) &= \frac{p(\boldsymbol{y}, y^\star)}{p(\boldsymbol{y})} \\
&= \frac{1}{p(\boldsymbol{y})} \int p(y^* \mid \boldsymbol{\theta}) p(\boldsymbol{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
&= \frac{1}{p(\boldsymbol{y})} \int p(y^* \mid \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \boldsymbol{y}) p(\boldsymbol{y}) d\boldsymbol{\theta} \\
&= \int p(y^* \mid \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \boldsymbol{y}) d\boldsymbol{\theta}
\end{aligned}
$$

Bibliografia

Banerjee S., Carlin B., and Gelfand A. (2015) *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton: Chapman Hall/CRC.

Cressie, N. (1993) *Statistics for Spatial Data*. New York: Wiley.

Matheron, G., (1965) Les variables régionalisés et leur estimation. Masson, Paris.

Shabenberger, O., Gotway, C. A. (2005) *Statistical Methods for Spatial Data Analysis*. Boca Raton, FL: Chapman & Hall/CRC.

¿Alguna Consulta?