



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

Análisis Bayesiano de los modelos cero-inflados

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Matemático

P R E S E N T A:

Marco Antonio Maqueda Hinojosa



**DIRECTORA DE TESIS:
Dra. Lizbeth Naranjo Albarrán
CD. MX. 2023**



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Índice general

1. Introducción	3
2. Paradigma Bayesiano y MCMC	5
2.1. Distribución a priori	5
2.2. Análisis conjugado	6
2.3. Distribución a priori no informativa	8
2.4. Predicción	10
2.5. Integración Monte Carlo	12
2.6. Métodos de Monte Carlo vía Cadenas de Markov (MCMC)	13
2.6.1. Algoritmo de Metropolis-Hastings	14
2.6.2. Muestreo de Gibbs	16
3. Aproximación de Laplace Anidada Integrada (INLA)	21
3.1. Modelos Gaussianos Latentes	21
3.2. Campos aleatorios Gaussianos de Markov (GMRF's)	22
3.3. Aproximación de Laplace	24
3.4. Método de INLA	26
3.5. El paquete R-INLA	34
4. Modelos lineales generalizados	37
4.1. Regresión lineal Bayesiana	37
4.1.1. Predicción	39
4.1.2. Criterios de selección de modelos	39
4.1.3. Análisis de la varianza	41
4.2. GLM's	42
4.3. Respuestas binarias	45
4.4. Respuestas para conteos	46
4.4.1. Regresión Poisson	46

ÍNDICE GENERAL

4.4.2. Regresión Binomial Negativa	49
5. Modelo de cero inflado con INLA	53
5.1. Introducción	53
5.2. Modelo de cero inflado Poisson	54
5.3. Modelo de cero inflado Binomial Negativa	54
6. Aplicación	59
6.1. Datos sobre la actividad física vigorosa	59
6.2. Análisis descriptivo de los datos	59
6.3. Resultados	64
7. Conclusiones	71
Bibliografía	73
A. Código de R	75

Capítulo 1

Introducción

La actividad física la define la OMS como cualquier movimiento corporal producido por los músculos esqueléticos, con el consiguiente consumo de energía. La actividad física vigorosa se caracteriza por ser una actividad física con alta intensidad, por ejemplo: natación, voleibol o fútbol a un nivel profesional.

En el presente trabajo se analizarán los datos recolectados de la Encuesta Nacional de Salud (ENSANUT) 2018, donde se tendrá particular interés en la información relacionada a la cantidad de horas en la que los encuestados realizan actividad física vigorosa. Lo anterior debido a que los propios datos indican que en la gran mayoría de encuestados este tipo de actividad física no forma parte de su rutina diaria, es decir, su registro en este rubro es 0 horas, y para aquellos encuestados que sí la realizan estos son más atípicos a mayor cantidad de horas.

El trabajo está dividido en 7 partes.

En el capítulo 2 se presentan aspectos generales del paradigma bayesiano. En particular se muestran los métodos MCMC para la simulación estocástica, muestreo de Gibbs y el algoritmo de metropolis-Hastings.

En el capítulo 3 se presenta el método de INLA, haciendo un recorrido sobre sus aspectos teóricos, sus beneficios al usarlo y una breve introducción de R-INLA, que es un paquete del software R con el que se implementarán los modelos en el presente trabajo.

En el capítulo 4 se presenta el tema de Modelos Lineales Generalizados. Haciendo un breve recorrido por el modelo de regresión lineal y sus características, y los Modelos Lineales Generalizados (GLM's). En particular, se presentan los modelos de regresión

Poisson y Binomial Negativa, que son modelos que sirven para conteos.

En el capítulo 5 se presenta el tema central del trabajo, el modelo cero inflado utilizando el método de INLA. Se hace un comentario de la importancia de estos modelos para datos de conteo donde existe una gran cantidad de ceros. Además, se presenta un ejemplo sencillo del Modelo cero inflado Poisson y Binomial Negativa con R-INLA.

En el capítulo 6 se presenta un ejemplo aplicado del modelo cero inflado utilizando INLA. La base de interés refiere a una encuesta donde se les pregunta a los encuestados acerca del número de días y horas, en los últimos 7 días, que realizaron actividad física vigorosa.

Finalmente, en el capítulo 7 se presentan las conclusiones del trabajo. Resaltando, por un lado, los beneficios de usar R-INLA y por otro lado comentando acerca del modelo cero inflado.

Capítulo 2

Paradigma Bayesiano y MCMC

En este capítulo, analizaremos conceptos fundamentales de la estadística Bayesiana, así como una introducción a dos métodos de simulación estocástica, muestreo de Gibbs y el algoritmo de Metropolis-Hastings.

A partir de ahora consideraremos $\underline{X} := (X_1, X_2, \dots, X_n)$ una muestra aleatoria de tamaño n y $\underline{x} := (x_1, x_2, \dots, x_n)$ una observación muestral, de tamaño n .

2.1. Distribución a priori

La estadística Bayesiana ha generado un gran interés debido a su propuesta para realizar inferencia estadística bajo incertidumbre, donde se resalta la posibilidad de contemplar tanto información observada como información obtenida a partir de experiencias previas.

Entonces los métodos bayesianos se enfocan en cómo se modifica la información acerca del valor del parámetro de acuerdo a los datos observados.

Definición 2.1.1. *Definimos a la familia paramétrica de densidades como:*

$$\mathcal{P} = \{p(x|\theta) : \theta \in \Theta\}$$

donde $p(\cdot|\theta)$ es una función de densidad y Θ un espacio paramétrico.

El objetivo de la estadística bayesiana es modelar la incertidumbre que tenemos sobre la variable aleatoria (o vector aleatorio) θ . Para esto se considera que θ tiene una distribución a priori o inicial $p(\theta)$ construida a partir de experiencia previa como:

datos de estudios anteriores, experiencia de un especialista de área, etc. Lo anterior se realiza antes de obtener los datos.

Por otro lado, del teorema de Bayes se tiene lo siguiente:

$$p(\theta|x_1, \dots, x_n) = \frac{p(\theta)p(x_1, \dots, x_n|\theta)}{\int p(\theta)p(x_1, \dots, x_n|\theta)d\theta}.$$

Entonces, dada una observación muestral \underline{x} , con la ayuda del teorema de Bayes podemos incorporar la información de \underline{x} junto a la distribución a priori, obteniendo así una descripción de la incertidumbre sobre el valor del parámetro mediante $p(\theta|x_1, \dots, x_n)$, a esta expresión se le conoce como la distribución a posteriori o final.

En conclusión, se tiene que $p(\theta|x_1, \dots, x_n)$, al igual que $p(\theta)$, es una distribución de probabilidad de θ , con la ventaja de que toma la información de $p(\theta)$ así como la información de la muestra \underline{x} . Esta distribución a posterior será importante para hacer inferencias sobre θ .

Notemos que $\int p(\theta)p(x_1, \dots, x_n|\theta)d\theta$ es un valor constante para θ . Por lo que es usual considerar el teorema de Bayes como sigue:

$$p(\theta|x_1, \dots, x_n) \propto p(\theta)\mathcal{L}(\theta|x_1, \dots, x_n),$$

donde

$$\mathcal{L}(\theta|x_1, \dots, x_n) = p(x_1, \dots, x_n|\theta),$$

es la función de verosimilitud.

2.2. Análisis conjugado

Como ya se mencionó, $p(\theta)$ y $p(\theta|\underline{x})$ son distribuciones de probabilidad de θ , en donde $p(\theta)$ considera una información inicial obtenida antes de aplicar la muestra y $p(\theta|\underline{x})$ incorpora la información de la distribución a priori con la información de la muestra, de tal forma que actualiza dicha información consiguiendo así una distribución a posteriori. Considerando lo anterior, es conveniente que $p(\theta)$ y $p(\theta|\underline{x})$ pertenezcan a la misma familia paramétrica, pues será de ayuda tanto para el análisis como para los cálculos computacionales.

Definición 2.2.1. Sea $\mathcal{P}\{p(\underline{x}|\theta) : \theta \in \Theta\}$ una familia paramétrica. Decimos que la colección (o clase) de distribuciones de probabilidad \mathcal{F} es una familia conjugada para \mathcal{P} si para todo $p(x_1, \dots, x_n|\theta) \in \mathcal{P}$ y $p(\theta) \in \mathcal{F}$ se cumple que $p(\theta|x_1, \dots, x_n) \in \mathcal{F}$.

Con la finalidad de garantizar que $p(\theta)$ y $p(\theta|x_1, \dots, x_n)$ pertenezcan a la misma familia de funciones de distribución, se considera a $p(\theta)$ tal que tenga la misma “estructura” de $p(x_1, \dots, x_n|\theta)$ vista como una función de θ .

A continuación, en la tabla [2.1](#), se presentan algunos modelos paramétricos univariados con sus respectivas familias conjugadas.

Familia paramétrica	Familia conjugada
Bernoulli(θ)	Beta($\theta \alpha, \beta$)
Geométrica(θ)	Beta($\theta \alpha, \beta$)
Exponencial(λ)	Gama($\lambda \alpha, \beta$)
Poisson(λ)	Gama($\lambda \alpha, \beta$)
Uniforme($0, \theta$)	Pareto($\theta \alpha, \beta$)
Normal(μ)	Normal($\mu \mu_0, \lambda_0$)

Tabla 2.1: Familias conjugadas.

Ejemplo 1. Sea $Y \sim \text{Poisson}(\mu)$, entonces su función de densidad es:

$$p(Y = y) = \exp(-\mu) \frac{\mu^y}{y!},$$

donde $y = 0, 1, \dots$ y $\mu > 0$.

Consideremos a $Y = \{Y_1, \dots, Y_n\}$ un conjunto de variables aleatorias tal que cada $Y_i \sim \text{Poisson}(\mu)$. Se puede tomar la distribución inicial de μ como una distribución gamma(α, β) con $\alpha, \beta > 0$, entonces su función de densidad está definida como sigue:

$$p(\mu) = \frac{\beta^\alpha}{\Gamma(\alpha)} \mu^{\alpha-1} \exp(-\beta\mu).$$

Se tiene que la función de verosimilitud de la distribución Poisson está dada por:

$$\mathcal{L}(\mu|y_1, \dots, y_n) \propto \prod_{i=1}^n \exp(-\mu) \mu^{y_i},$$

luego, la distribución final es:

$$\begin{aligned} p(\mu|y) &\propto \left[\prod_{i=1}^n \exp(-\mu) \mu^{y_i} \right] \mu^{\alpha-1} \exp(-\beta\mu) \\ &= \mu^{\alpha + \sum_{i=1}^n y_i - 1} \exp(-\mu(\beta + n)), \end{aligned}$$

es decir, la distribución final sigue una distribución gama $(\mu|\alpha + \sum_{i=1}^n y_i, \beta + n)$. En conclusión, la familia de distribuciones Poisson es conjugada para la familia de distribuciones gama.

2.3. Distribución a priori no informativa

La gran ventaja de utilizar estadística Bayesiana reside en que se permite añadir información a priori con información experimental. Por lo anterior, se presenta un gran problema cuando no existe una información inicial adecuada o simplemente no se cuenta con ella. En palabras, se dice que una distribución es no informativa si refleja bastante incertidumbre respecto a la información del parámetro.

Para tener de forma más clara el concepto de distribuciones muy informativas y poco informativas se da el siguiente ejemplo.

Sea x_1, x_2, \dots, x_n una muestra aleatoria de una distribución Binomial(k, θ), consideremos $n \in \mathbb{N}$ y la distribución inicial $\theta \sim \text{Beta}(a, b)$, con $a > 0$ y $b > 0$. Se sabe que la distribución final de θ es:

$$\theta|x_1, x_2, \dots, x_n \sim \text{Beta} \left(a + \sum_{i=1}^n x_i, b + kn - \sum_{i=1}^n x_i \right).$$

Además, el estimador puntual de θ es:

$$\hat{\theta} = \frac{a + \sum_{i=1}^n x_i}{a + b + nk}.$$

Entonces, para valores grandes de a y b , en comparación a la magnitud de $\sum_{i=1}^n x_i$ y nk ,

la información a priori tendrá más peso que la información muestral sobre las inferencias que se realicen sobre la distribución a posteriori. Por otro lado, para valores pequeños de a y b , cercanos a cero, se tendrá que la muestra tendrá un mayor peso, de este modo la distribución a priori tendrá una varianza mayor, por lo que las inferencias que se hagan se basarán principalmente en la información obtenida de los datos muestrales.

Una alternativa para tratar los casos cuando no se tiene una distribución a priori informativa es utilizar la Regla de Jeffreys, que construye una clase de distribuciones que busca invariancia ante transformaciones y, de esta forma, proveer la menor información a priori en relación a la información muestral vía la información esperada de Fisher. Para más información de la Regla de Jeffreys ver

Para el caso unidimensional se tiene lo siguiente.

Consideremos una variable aleatoria X y $f(x|\theta)$ su función de densidad. La información de Fisher es:

$$\mathbb{I}(\theta) = \mathbb{E}_{\theta} \left[\left(\frac{\partial}{\partial \theta} \log f(x|\theta) \right)^2 \right].$$

Bajo condiciones de regularidad se cumple que:

$$\mathbb{I}(\theta) = -\mathbb{E}_{\theta} \left[\frac{\partial^2}{\partial \theta^2} \log f(x|\theta) \right].$$

Se define la distribución a priori de Jeffreys como:

$$f(\theta) \propto \sqrt{\mathbb{I}(\theta)}.$$

Para el caso multidimensional se define la matriz de información de Fisher como sigue:

$$\mathbb{I}(\theta) = [\mathbb{I}_{ij}(\theta)]$$

donde

$$\mathbb{I}_{ij}(\theta) = -\mathbb{E}_{\theta} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(x|\theta) \right]$$

donde $\mathbb{I}_{ij}\theta$ son las entradas (i, j) de la matriz $\mathbb{I}(I)$, con $i, j \in 1, 2, \dots, k$ y $\theta \in \Theta \subset \mathbb{R}^2$

$$f(\theta) \propto \sqrt{\det \mathbb{I}(\theta)}.$$

Ejemplo 2. Sea $X \sim \text{Poisson}(\theta)$, usaremos la regla de Jeffreys para obtener una distribución inicial no informativa. Se tiene lo siguiente:

$$\begin{aligned} f(x|\theta) &= \frac{\exp(-\theta)\theta^x}{x!} \\ \log f(x|\theta) &= -\theta + x \log(\theta) - \log(x!) \\ \frac{d}{d\theta} \log f(x|\theta) &= -1 + x \frac{1}{\theta} = \frac{-\theta + x}{\theta}, \end{aligned}$$

entonces, calculando la información esperada de Fisher se tiene lo siguiente:

$$\begin{aligned} \mathbb{I}(\theta) &= \mathbb{E} \left[\left(\frac{d}{d\theta} \log f(x|\theta) \right)^2 \right] = \mathbb{E} \left[\left(\frac{-\theta + x}{\theta} \right)^2 \right] \\ &= \frac{\mathbb{E}[(-\theta + x)^2]}{\theta^2} = \frac{\text{Var}(x)}{\theta^2} \\ &= \frac{\theta}{\theta^2} = \frac{1}{\theta} \\ \sqrt{\mathbb{I}(\theta)} &\propto \theta^{-1/2}. \end{aligned}$$

Por lo tanto, se concluye que la distribución a priori no informativa de Jeffreys es $p(\theta) \propto \theta^{-1/2}$. Una observación importante es que la distribución obtenida es impropia, es decir, $\int p(\theta) d\theta = \infty$, sin embargo, se puede utilizar siempre que la distribución a posteriori sea propia, es decir, $\int p(\theta|x) d\theta < \infty$.

2.4. Predicción

Parte fundamental del análisis estadístico es poder predecir un valor de una futura observación X con ayuda de la información disponible. El modelo paramétrico $p(x|\theta)$, junto con la distribución a priori $p(\theta)$, inducen una distribución conjunta para (X, θ) , dada por:

$$p(x, \theta) = p(x|\theta)p(\theta).$$

Ahora, la distribución marginal

$$p(x) = \int p(x|\theta)p(\theta)d\theta$$

describe lo que conocemos de X dada la información inicial. A esta distribución se le conoce como la distribución predictiva a priori o inicial.

Siguiendo con este razonamiento, dada una muestra \underline{x} , $p(x|\theta)$ y la distribución a posteriori inducen una distribución conjunta para (X, θ) condicionada a los valores observados \underline{x} , esto es:

$$p(x, \theta|\underline{x}) = p(x|\theta, \underline{x})p(\theta|\underline{x}) = p(x|\theta)p(\theta|\underline{x}),$$

donde la última igualdad se da si existe independencia condicional de X y (X_1, \dots, X_n) dado θ . Entonces, la distribución

$$p(x|\underline{x}) = \int p(x|\theta)p(\theta|\underline{x})d\theta$$

describe el comportamiento de X dada toda la información disponible, a esta distribución se le conoce como distribución predictiva a posteriori o final.

Ejemplo 3. Se dará un ejemplo de predicción de una distribución Bernoulli. Sea X_1, \dots, X_n una muestra aleatoria de una distribución Bernoulli, es decir, son tomadas de la siguiente función de densidad:

$$p(x|\theta) = \theta^x(1 - \theta)^{1-x},$$

donde $x \in \{0, 1\}$ y $0 < \theta < 1$.

Ahora, supongamos que θ tiene una distribución inicial $\text{Beta}(\alpha_0, \beta_0)$. En la sección de análisis conjugado se mencionó que la distribución beta es conjugada para el modelo Bernoulli, entonces se tiene que la distribución final de θ sigue una distribución beta:

$$p(\theta|x_1, \dots, x_n) = \text{Beta}(\theta|\alpha_1, \beta_1),$$

donde $\alpha_1 = \alpha_0 + \sum_{i=1}^n x_i$ y $\beta_1 = \beta_0 + n - \sum_{i=1}^n x_i$.

Finalmente, la distribución predictiva final está dada como sigue:

$$p(x|x_1, \dots, x_n) = \frac{\Gamma(\alpha_1 + \beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)} \frac{\Gamma(\alpha_1 + x)\Gamma(\beta_1 + 1 - x)}{\Gamma(\alpha_1 + \beta_1 + 1)},$$

donde $x = \{0, 1\}$. Es decir, la distribución predictiva final es una distribución Beta-Binomial.

2.5. Integración Monte Carlo

La integración Monte Carlo evalúa $\mathbb{E}[g(\theta)]$ obteniendo muestras de la distribución $p(\theta|x)$, que denotaremos como: $\{\theta^{(t)}, t = 1, \dots, n\}$ y así, aproximar:

$$\mathbb{E}[g(\theta)] \approx \frac{1}{n} \sum_{t=1}^n g(\theta^{(t)}),$$

donde $g(\theta)$ es una función de la variable aleatoria θ . Con lo anterior, se tiene que la media poblacional de $g(\theta)$ se estima por medio de una media muestral.

Cuando $\{\theta^{(t)}\}$ son independientes la Ley de los Grandes Números asegura que la aproximación puede hacerse tan precisa como se desee cuando la muestra es lo suficientemente grande. En general, elegir $\{\theta^{(t)}\}$ independientes de $p(\theta|x)$ no es posible, debido a que $p(\theta|x)$ puede no ser estándar. Pero las $\{\theta^{(t)}\}$ no necesitan ser independientes, pueden ser generadas por cualquier proceso que seleccione muestras por todo el soporte de $p(\theta|x)$ en las proporciones correctas. Una de las formas de realizar lo anterior es con una cadena de Markov, eligiendo a $p(\theta|x)$ como su distribución estacionaria. A lo anterior se le conoce como el método de Monte Carlo vía cadenas de Markov. Para revisar más respecto al tema de Integración Monte Carlo y los Métodos de Monte Carlo vía Cadenas de Markov revisar Robert, C.P. & G. Casella. (2010).

2.6. Métodos de Monte Carlo vía Cadenas de Markov (MCMC)

Es usual que en la práctica la distribución final y la distribución predictiva final tengan formas muy complicadas, esto debido a que las integrales requeridas no tienen una solución analítica, razón por la cual es necesario la implementación de métodos de simulación estocástica, como lo son los métodos MCMC o el método de INLA.

A continuación se presentarán definiciones elementales de procesos estocásticos que servirán para definir las técnicas MCMC. Puede encontrar más información y resultados en Rincón (2012).

Definición 2.6.1. *Se define un proceso estocástico como una colección de variables aleatorias indexadas por un conjunto T , denotado como $(X_t)_{t \in T}$, las variables aleatorias están definidas sobre el espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ que toman valores en el espacio medible (E, \mathcal{E}) , donde al conjunto E se le conoce como el espacio de estados.*

Nota: Dado un proceso estocástico $\{X_n : n \geq 0\}$ es común usar la notación de $\mathbb{P}(X_n = x_n)$ como $p(x_n)$, es decir, el subíndice indica también a la variable a la que se hace referencia.

Definición 2.6.2. *Una cadena de Markov es un proceso estocástico a tiempo discreto $\{X_n : n = 0, 1, \dots\}$ con espacio de estados E discreto que cumple la propiedad de Markov, es decir:*

$$p(x_{n+1}|x_0, \dots, x_n) = p(x_{n+1}|x_n).$$

Definición 2.6.3. *Sean i y j dos estados de una cadena de Markov, definimos la probabilidad de transición como:*

$$\mathbb{P}(X_{n+1} = j | X_n = i),$$

denotada como $p_{ij}(n, n+1)$ y representa la probabilidad de transición del estado i en el tiempo n , al estado j en el tiempo $n+1$. Variando los índices i y j sobre un conjunto

de estados, dígase $\{0, 1, 2, 3, \dots\}$ se puede obtener la matriz \mathbf{P} de probabilidades de transición. Lo anterior se conoce como probabilidades de transición en un paso. La probabilidad $\mathbb{P}(X_n = j | X_0 = 1)$, denotada como $p_{ij}(n)$, es conocida como la probabilidad de transición en n pasos.

Definición 2.6.4. Diremos que una distribución de probabilidad $\pi = (\pi_0, \pi_1, \dots)$ es estacionaria para una cadena de Markov $\{X_n : n = 0, 1, \dots\}$ con matriz de probabilidades de transición $\mathbf{P} = (p_{ij})$ si:

$$\pi_j = \sum_i \pi_i p_{ij}.$$

Definición 2.6.5. Sea $\{X_n : n = 0, 1, \dots\}$ una cadena de Markov con matriz de probabilidades de transición $\mathbf{P} = (p_{ij})$ y distribución inicial π^0 . Definimos la distribución límite o equilibrada de la cadena de Markov como:

$$\pi = \lim_{n \rightarrow \infty} \pi^0 \mathbf{P}^n = \lim_{n \rightarrow \infty} \sum_i \pi_i^0 p_{ij}(n).$$

Los métodos MCMC nos permiten generar observaciones de distribuciones multivariadas, de forma iterativa, que difícilmente podrían simularse. La idea de estas técnicas es construir una cadena de Markov que sea fácil de simular y cuya distribución de equilibrio corresponda a la distribución final de interés.

A continuación se mostrará una breve descripción de los dos modelos MCMC más utilizados, estos son muestreo de Gibbs y el algoritmo de Metropolis-Hastings.

2.6.1. Algoritmo de Metropolis-Hastings

Con el algoritmo de Metropolis-Hastings se construirá una cadena de Markov definiendo las probabilidades de transición como sigue.

Sea $Q(\theta^*, \theta)$ una probabilidad de transición (arbitraria), definimos:

$$\alpha(\theta^*, \theta) = \min \left\{ 1, \frac{p(\theta^* | X) Q(\theta | \theta^*)}{p(\theta | X) Q(\theta^* | \theta)} \right\}.$$

Dado un valor inicial $\theta^{(0)}$, la t -ésima iteración consta de:

- Se genera una observación θ^* de $Q(\theta^*, \theta^{(t)})$.
- Se genera una variable $u \sim U(0, 1)$, es decir, u sigue una distribución Uniforme en el intervalo $(0, 1)$.
- Si $\alpha(\theta^*, \theta^{(t)}) \geq u$ se considera $\theta^{(t+1)} = \theta^*$, en otro caso, $\theta^{(t+1)} = \theta^{(t)}$.

Con lo anterior, se genera una cadena de Markov con distribución de transición:

$$\mathbf{P}(\theta^{(t+1)}|\theta^{(t)}) = \alpha(\theta^{(t)}, \theta^{(t+1)})Q(\theta^{(t+1)}|\theta^{(t)})\mathbb{1}_{(\theta^{(t+1)} \neq \theta^{(t)})} + \left[1 - \int \alpha(\theta^{(t)}, \theta^*)Q(\theta^*|\theta^{(t)})d\theta^*\right] \mathbb{1}_{(\theta^{(t+1)} = \theta^{(t)})}.$$

donde $\mathbb{1}_\theta$ es la función indicadora.

En la práctica se destacan dos casos particulares, estos son:

- Caminata aleatoria. Consideremos $Q(\theta^*|\theta) = Q_1(\theta^* - \theta)$, donde Q_1 es una función de densidad simétrica centrada en el origen. Entonces,

$$\alpha(\theta^*, \theta) = \min \left\{ \frac{p(\theta^*|x)}{p(\theta|x)}, 1 \right\}.$$

- Independencia. Consideremos $Q(\theta^*|\theta) = Q_0(\theta^*)$, donde es una función de densidad sobre Θ . Entonces,

$$\alpha(\theta^*|\theta) = \min \left\{ \frac{\omega(\theta^*)}{\omega(\theta)}, 1 \right\},$$

con $\omega(\theta) = p(\theta|x)/Q_0(\theta)$.

Para ver más acerca del algoritmo de Metropolis-Hastings se recomienda revisar Gamerman et al. (2006), Robert et al. (2010) y Liang et al. (2010).

Ejemplo 4. Se tiene el interés de simular una distribución gama, es decir, una distribución de la forma

$$p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} \exp\{-\beta\theta\},$$

donde $\theta, \alpha, \beta > 0$. Particularmente se verá el caso donde $\alpha = 2.3$ y $\beta = 2.7$.

Para el ejemplo se simulan 100,000 muestras de una distribución normal con media α/β y varianza α/β^2 , en nuestro caso, $N(0.85, 0.30)$.

Se sabe que la media y varianza de una variable aleatoria X con distribución $\text{gama}(\alpha, \beta)$ están dada por:

$$\mathbb{E}[X] = \frac{\alpha}{\beta} \quad \text{y} \quad \text{Var}(X) = \frac{\alpha}{\beta^2},$$

entonces se espera obtener una media cercana a $\alpha/\beta = 2.3/2.7 \approx 0.8518$ y una varianza cercana a $\alpha/\beta^2 = 2.3/2.7^2 \approx 0.3155$.

Al simular la distribución se obtiene la siguiente media y varianza:

```
1 c(mean(vec), var(vec))
2 0.8491365 0.3111445
```

Finalmente, podemos notar que las aproximaciones con el algoritmo de Metropolis-Hastings resultaron muy cercanas a lo esperado. El código utilizado se puede encontrar en el Apéndice [A](#).

2.6.2. Muestreo de Gibbs

La técnica de muestreo de Gibbs permite simular una cadena de Markov, $\theta^{(1)}, \theta^{(2)}, \dots$ con distribución de equilibrio $p(\theta|x)$. En este método, cada nuevo valor de la cadena se obtiene con un proceso iterativo que sólo requiere generar muestras de distribuciones cuya dimensión es menor que la dimensión de θ , digamos d .

Sea $\theta = (\theta_1, \dots, \theta_k)$ una partición del vector θ , con $\theta_i \in \mathbb{R}^{d_i}$ y $\sum_{i=1}^k d_i = d$. Definimos las densidades condicionales completas como sigue:

Ejemplo 5. Buscaremos estimar los parámetros de una distribución normal donde la media μ y la varianza σ^2 son desconocidas. En lugar de elegir una distribución a priori no informativa se elige una a priori independiente informativa, tal que $p(\mu, \sigma^2) = p(\mu)p(\sigma^2)$. Para la media se elige:

$$\mu \sim \text{Normal}(\mu_0, \sigma_0^2),$$

y una a priori inversa Gama para la varianza:

$$\sigma^2 \sim \text{invGama}(a, b).$$

Se sigue que la distribución posterior conjunta está dada por:

$$\begin{aligned} p(\mu, \sigma^2 | y) &\propto (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right) \\ &\times (\sigma_0^2)^{-1/2} \exp\left(-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (\mu_i - \mu_0)^2\right) \\ &\times (\sigma^2)^{-(a+1)} \exp\left(-\frac{b}{\sigma^2}\right), \end{aligned}$$

donde $\sigma_0^2, a, b > 0$ y $-\infty < \mu_0 < \infty$. Luego se puede probar que la distribución condicional completa de la media μ es:

$$\begin{aligned}
p(\mu|\sigma^2, y) &\propto \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right) \times \exp\left(-\frac{1}{2\sigma_0^2} (\mu - \mu_0)^2\right) \\
&= \exp\left(-\frac{1}{2} \left(\frac{\sum_{i=1}^n y_i^2 + n\mu^2 - 2\mu \sum_{i=1}^n y_i}{\sigma^2} + \frac{\mu^2 + \mu_0^2 - 2\mu_0\mu}{\sigma_0^2} \right)\right) \\
&\propto \exp\left(-\frac{1}{2} \left[\mu^2 \left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2} \right) - 2\mu \left(\frac{\sum_{i=1}^n y_i}{\sigma^2} + \frac{\mu_0}{\sigma_0^2} \right) \right]\right).
\end{aligned}$$

Usando el hecho de que si

$$p(\theta) \propto \exp\left(-\frac{1}{2}(a\theta^2 - 2b\theta)\right),$$

entonces $\theta \sim \text{Normal}(b/a, 1/a)$, se puede llegar a que la distribución condicional completa sigue una distribución Gaussiana como sigue:

$$\mu|\sigma^2, y \sim \text{Normal}\left(\frac{\sum_{i=1}^n y_i}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}}, \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}}\right).$$

Ahora, respecto a la distribución condicional completa de la varianza σ^2 se tiene lo siguiente:

$$\begin{aligned}
 p(\sigma^2|\mu, y) &\propto (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right) \times (\sigma^2)^{-(a+1)} \exp\left(-\frac{b}{\sigma^2}\right) \\
 &\propto (\sigma^2)^{-n/2-(a+1)} \exp\left(-\frac{1}{\sigma^2} \left(b + \frac{\sum_{i=1}^n (y_i - \mu)^2}{2}\right)\right),
 \end{aligned}$$

de donde se concluye que sigue una distribución inversa gamma:

$$\sigma^2|\mu, y \sim \text{invGamma}\left(\frac{n}{2} + a, b + \frac{\sum_{i=1}^n (y_i - \mu)^2}{2}\right).$$

Para el ejemplo el objetivo será simular una distribución normal con media 10 y varianza 1.5. Se considerará la situación con valores iniciales alejados de los valores reales, por ejemplo $\mu^{(0)} = 1$ y $\sigma^{2(0)} = 5$. En general se seleccionan valores iniciales sobredispersos con respecto a algún conocimiento que se tenga de la distribución a priori. Finalmente se eligen valores que corresponderán a distribuciones a priori no informativas, estos valores son: $\mu_0 = 3$, $\sigma_0^2 = 10$, $a = 2.5$ y $b = 0.5$.

Finalmente, del muestreo de Gibbs se obtienen las siguientes media y varianza:

```

1 > mean(mu.sim);mean(sigma2.sim)
2 [1] 9.968869
3 [1] 1.601982

```

que podemos notar son muy cercanas a la media y varianza real.

En el Apéndice [A](#) se puede encontrar el código utilizado para realizar el muestreo de Gibbs del ejemplo anterior.

Capítulo 3

Aproximación de Laplace Anidada Integrada (INLA)

En este capítulo se analizarán conceptos fundamentales para entender el método de INLA (*Integrated Nested Laplace Approximation*) como: Modelos Gaussianos Latentes, campos aleatorios Gaussianos de Markov y aproximación de Laplace. Se dará un ejemplo del método paso a paso y se presentará una breve descripción del paquete R-INLA.

3.1. Modelos Gaussianos Latentes

La aproximación mediante el método INLA está restringida a una clase particular de modelos, estos son los Modelos Gaussianos Latentes (LGM's). Un modelo simple de un LGM es el modelo lineal generalizado bayesiano (los modelos lineales generalizados se estudiarán en el capítulo 4). Éste corresponde al predictor lineal:

$$\eta_i = \beta_0 + \beta_1 x_{i1}, \quad i = 1, 2, \dots, n;$$

donde β_0 es el intercepto, β_1 es la pendiente y x_{i1} es la covariable. La variable respuesta y_i se supone que sigue una distribución de la familia exponencial, donde su media $\mu_i = \mathbb{E}[Y_i]$ está asociada con η_i a través de una función liga $g(\cdot)$ tal que $\eta_i = g(\mu_i)$.

Cada η_i debe seguir una distribución normal dada por sus hiperparámetros en un LGM, esto es, β_0 y β_1 tienen una distribución a priori Gaussiana, $\beta_0 \sim N(\mu_0, \sigma_0^2)$ y

$\beta_1 \sim N(\mu_1, \sigma_1^2)$, lo que implica que $\eta_i \sim N(\mu_0 + \mu_1 x_{i1}, \sigma_0^2 + \sigma_1^2 x_{i1}^2)$.

El modelo general es:

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

que es un modelo donde se agregan más covariables y componentes del modelo $f_k(\cdot)$, estos últimos son usados para ajustar mejor el modelo lineal y/o introducir efectos aleatorios, por ejemplo, para analizar la sobredispersión de los datos ocasionado por heterogeneidad no observada o datos con correlación en datos longitudinales se puede considerar utilizar un efecto aleatorio en el modelo, que puede ser introducido considerando que las $f_k(\cdot)$ son independientes y siguen una distribución normal con media 0.

Para cada efecto lineal y cada componente del modelo, se debe considerar una distribución a priori Gaussiana que sea univariada o multivariada y de esta forma se cumpla que $\eta = (\eta_1, \dots, \eta_n)'$ sea Gaussiano como lo requieren los modelos LGM. En la siguiente sección se verá que existe una clase particular de modelos Gaussianos, estos son los campos aleatorios Gaussianos de Markov (GMRF), que serán más flexibles y eficientes al modelar efectos utilizados en un LGM.

3.2. Campos aleatorios Gaussianos de Markov (GMRF's)

Para entender el concepto de GMRF (*Gaussian Markov Random Field*) de manera más clara comenzaremos hablando del caso más simple, esto es, cuando η_i sigue un proceso autorregresivo de primer orden, denotado como AR(1).

El modelo AR(1) queda definido como sigue:

$$\begin{cases} \eta_1 \sim N(0, \sigma_\eta^2 / (1 - \rho^2)), \\ \eta_i | \eta_{i-1}, \dots, \eta_1 \sim N(\rho \eta_{i-1}, \sigma_\eta^2) \quad i = 2, \dots, n, \end{cases}$$

es decir, es un modelo donde las variables η_i , salvo η_1 , siguen una distribución normal con media $\rho \eta_{i-1}$ y varianza constante σ_η^2 , ρ es el coeficiente de correlación entre η_i y η_{i-1} .

Se puede probar que para cada $i = 1, \dots, n$ la distribución marginal de η_i es normal con media 0 y varianza $\sigma_\eta^2/(1 - \rho^2)$ (Brockwell 2016). Entonces se tiene que η es un proceso Gaussiano (proceso estocástico $\{\eta_t : t \geq 0\}$ a tiempo continuo tal que el vector $(\eta_{t_1}, \dots, \eta_{t_n})$ tiene distribución normal o Gaussiana multivariada, donde t_1, \dots, t_n es una colección finita de tiempos con vector de medias 0 y matriz de covarianzas Σ , es decir, $\eta \sim N(0, \Sigma)$).

Una ventaja importante de trabajar con los modelos AR(1) es que su matriz de precisión tiene una estructura sencilla, ésta es:

$$Q = \frac{1}{\sigma_\eta^2} \begin{pmatrix} 1 & -\rho & & & & \\ -\rho & 1 + \rho^2 & -\rho & & & \\ & \ddots & \ddots & \ddots & & \\ & & -\rho & 1 + \rho^2 & -\rho & \\ & & & -\rho & 1 & \end{pmatrix},$$

es decir, una matriz con entradas 0 salvo en las entradas de la diagonal principal, la diagonal por arriba de ésta así como la de abajo. La matriz Q tiene esa estructura debido a que para toda i , η_i solo depende η_{i-1} , es decir, η_i y η_{i-1} son condicionalmente independientes para $|i - j| > 1$.

Con lo anterior, podemos concluir que los procesos AR(1) son procesos Gaussianos con la propiedad de ser condicionalmente independientes.

Definición 3.2.1. *Diremos que η es un GMRF si sigue una distribución normal multivariada y cumple la propiedad de ser condicionalmente independiente.*

Ahora, notemos que podemos escribir una formulación de un modelo jerárquico de tres etapas para cada LGM. Consideremos $y = (y_1, \dots, y_n)'$, asumimos que en la primera etapa de cada variable de y es condicionalmente independiente con una cierta distribución de la familia exponencial:

$$y \mid \eta, \theta_1 \sim \prod_{i=1}^n p(y_i \mid \eta_i, \theta_1),$$

dado η y los hiperparámetros θ_1 . Ahora, en la segunda etapa, se especifica η para que sea un campo Gaussiano aleatorio latente con función de densidad dada por:

$$p(\eta|\theta_2) \propto |Q_{\theta_2}|_+^2 \exp\left(-\frac{1}{2}\eta'Q_{\theta_2}\eta\right),$$

donde Q_{θ_2} es una matriz semi-definida positiva que depende de los hiperparámetros θ_2 y $|Q_{\theta_2}|_+^2$, conocida como la matriz de precisión, denota el producto de los valores propios no cero y describe la estructura de dependencia de los datos y su inversa, si es que existe, es una matriz de covarianza.

En la última etapa se considera que $\theta = (\theta_1, \theta_2)$ sigue una distribución a priori $\pi(\theta)$, que puede ser una función de distribución conjunta o producto de distribuciones. De este modo, la distribución conjunta a posteriori de η y θ es:

$$\begin{aligned} p(\eta, \theta|y) &\propto \pi(\theta)\pi(\eta|\theta_2) \prod_i p(y_i|\eta_i, \theta_1) \\ &\propto \pi(\theta)|Q_{\theta_2}|^{1/2} \exp\left(-\frac{1}{2}\eta'Q_{\theta_2}\eta + \sum_i \log \pi(y_i|\eta_i, \theta_1)\right). \end{aligned}$$

El método de INLA se aplica a la expresión anterior para obtener una aproximación de la distribución a posteriori de cada parámetro desconocido.

Para que todos los LGM's puedan ser ajustados de una manera óptima con el método INLA, se necesita que los siguientes supuestos se cumplan:

- a) El número de hiperparámetros θ debería ser pequeño, usualmente entre 2 y 5, pero no debe ser mayor a 20.
- b) Cuando n es muy grande, dígame 10^4 , entonces η debe ser un GMRF.
- c) Cada y_i sólo depende de un componente de η , por ejemplo, η_i .

Los supuestos anteriores ayudarán a que el método de INLA estime de manera más precisa y para tener un mejor rendimiento computacional.

3.3. Aproximación de Laplace

La aproximación de Laplace es usada para aproximar integrales de la forma:

$$I_n = \int_x \exp(nf(x))dx,$$

donde $n \rightarrow \infty$. Sea x_0 la moda de una función $f(x)$ dos veces diferenciable. A continuación, hacemos la expansión de Taylor de la función $f(x)$ alrededor de x_0 :

$$\begin{aligned} I_n &\approx \int_x \exp \left(n \left(f(x_0) + (x - x_0)f'(x_0) + \frac{1}{2}(x - x_0)^2 f''(x_0) \right) \right) dx \\ &= \exp(nf(x_0)) \int \exp \left(\frac{n}{2}(x - x_0)^2 f''(x_0) \right) dx \\ &= \exp(nf(x_0)) \sqrt{\frac{2\pi}{-nf''(x_0)}} \\ &= \hat{I}_n. \end{aligned}$$

Ahora, notemos que $f'(x_0) = 0$, ya que x_0 es la moda e I_n es una integral Gaussiana. Si $nf(x)$ es interpretada como la suma de log-verosimilitudes y x un parámetro desconocido, entonces la aproximación de Laplace será más exacta cuando $n \rightarrow \infty$.

A continuación, veremos la importancia de la aproximación de Laplace en el método de INLA con un ejemplo pequeño.

Sean $y_1 = 1$ y $y_2 = 2$ dos observaciones de una distribución Poisson con media λ_i , donde $i = 1, 2$. Consideremos la liga log para asociar λ_i con el predictor lineal η_i respectivamente, además η_i es un proceso AR(1). La distribución posterior queda definida como sigue:

$$\begin{aligned} p(\eta|y) &\propto \exp \left(-\frac{1}{2}\eta'Q_\theta\eta \right) \exp(\eta_1 + 2\eta_2 - \exp(\eta_1) - \exp(\eta_2)), \\ Q_\theta &= \tau \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}, \end{aligned}$$

donde $\theta = (\tau, \rho)$, con $\tau > 0$ es el parámetro de escala y ρ es el parámetro de autocorrelación. El objetivo es aproximar las marginales posteriores de η_1 y η_2 , el procedimiento a seguir es aplicar una secuencia de aproximaciones Gaussianas utilizando la aproximación de Laplace, por ejemplo, aproximar la marginal de η_1 de la siguiente manera:

$$p(\eta_1|y) = \frac{p(\eta_2|y)}{p(\eta_2|\eta_1)} \approx \frac{p(\eta_2|y)}{\tilde{p}(\eta_2|\eta_1)},$$

donde $\tilde{p}(\eta_2|\eta_1)$ es la aproximación Gaussiana de la condicional completa de η_2 para cada valor de η_1 , de manera análoga el caso de η_2 . Realizar el proceso de esta manera es más precisa, pues cada distribución condicional es más cercana a una distribución Gaussiana que su distribución conjunta.

En conclusión, la idea es aplicar la aproximación de Laplace a densidades casi Gaussianas. Para ver más sobre cómo funciona la aproximación de Laplace para el método de INLA se recomienda ver (Rue et al., 2017) y (Wang et al., 2018).

3.4. Método de INLA

La finalidad del método de INLA es aproximar las marginales posteriores $p(\eta_i|y)$ para $i = 1, \dots, n$ y $p(\theta_j|y)$ para $j = 1, \dots, m$. Si bien la aproximación de Laplace es complicada, pues se trabaja con producto de distribuciones Gaussianas y no Gaussianas, la estrategia que sigue el método de INLA busca reformular el problema y hacer casos más sencillos donde se pueda aplicar la aproximación de Laplace para distribuciones Gaussianas. La estrategia del método se puede resumir en tres pasos:

- 1) Proponer y aproximar $\tilde{p}(\theta|y)$ a la posterior conjunta $p(\theta|y)$.
- 2) Proponer una aproximación $\tilde{p}(\eta_i|\theta, y)$ a las marginales de la distribución condicional de η_i dados los datos y los hiperparámetros $p(\eta_i|\theta, y)$.
- 3) Explorar $\tilde{p}(\theta|y)$ y usarla para hacer integración numérica.

Usando el procedimiento anterior, las marginales posteriores aproximadas de interés tienen la siguiente forma:

$$\tilde{p}(\eta_i|y) = \sum_k \tilde{p}(\eta_i|\theta^{(k)}, y) \tilde{p}(\theta^{(k)}|y) \Delta\theta^{(k)},$$

$$\tilde{p}(\theta_j|y) = \int \tilde{p}(\theta|y) d\theta_{-j},$$

donde θ_{-j} denota el vector de θ con el elemento j -ésimo excluido, $\tilde{p}(\theta^{(k)}|y)$ son los valores de la densidad obtenida durante una exploración de $\tilde{p}(\theta|y)$.

A continuación, describiremos más concretamente los tres pasos del método INLA.

Paso 1. **Aproximando** $p(\theta|y)$

La aproximación de Laplace de la posterior conjunta de los hiperparámetros está dada por:

$$\tilde{p}(\theta|y) \propto \frac{p(\eta, \theta|y)}{\tilde{p}(\eta|\theta, y)} \Big|_{\eta=\eta^*(\theta)},$$

donde $\tilde{p}(\eta|\theta, y)$ es la aproximación Gaussiana del condicional completo de η y $\eta^*(\theta)$ es el modelo del condicional completo para η , dada una θ . La aproximación anterior es exacta si $p(\eta|\theta, y)$ es Gaussiana. Es importante recalcar que $\tilde{p}(\eta|\theta, y)$ tiene la siguiente forma:

$$\tilde{p}(\eta|\theta, y) \propto |R_\theta|^{1/2} \exp\left(-\frac{1}{2}(\eta - \mu_\theta)' R_\theta (\eta - \mu_\theta)\right),$$

donde μ_θ es la ubicación de la moda, $R_\theta = Q_\theta + \text{diag}(c_\theta)$ con c_θ un vector que contiene las segundas derivadas del log-verosimilitud en la moda con respecto a η_i (Rue y Held, 2005). Utilizar la aproximación de $\tilde{p}(\eta|\theta, y)$ tiene dos ventajas: 1) es un GMRF con la misma estructura de dependencia de Q_θ , por lo que es fácil evaluar $\tilde{p}(\eta|\theta, y)$ para cada θ ; y 2) la aproximación Gaussiana es precisa, ya que el impacto del condicionamiento en las observaciones es solo en la diagonal, lo que desplaza la media, reduce la varianza y podría introducir cierta asimetría en los márgenes. Más aún, no cambia la estructura de dependencia Gaussiana.

La aproximación de $\tilde{p}(\theta|y)$ puede ser mejorada usando transformaciones estabilizadoras de varianza de θ , por ejemplo utilizando la transformación logaritmo. Particularmente, el método INLA utiliza la siguiente transformación para $\theta = (\theta_1, \theta_2)$ del modelo AR(1):

$$\begin{cases} \theta_1 = \log(\kappa), & \kappa = \tau(1 - \rho^2), \\ \theta_2 = \log\left(\frac{1+\rho}{1-\rho}\right), \end{cases}$$

donde κ es la precisión marginal. Además, se puede utilizar la matriz Hessiana en la moda para construir combinaciones, o transformaciones, casi lineales independientes de θ . Estas transformaciones tienden a disminuir las colas largas y reducir la asimetría, lo que da posteriores mucho más simples y con mejor comportamiento.

Paso 2. **Aproximando** $p(\eta_i|\theta, y)$

Para aproximar $p(\eta_i|\theta, y)$ hay tres opciones usuales para proceder, éstas son:

1. Utilizar las marginales de la aproximación Gaussiana de $\tilde{p}(\eta|\theta, y)$.

$$\tilde{p}(\eta|\theta, y) \propto |R_\theta|^{1/2} \exp\left(-\frac{1}{2}(\eta - \mu_\theta)' R_\theta (\eta - \mu_\theta)\right).$$

El único problema adicional en cuanto a cálculos se refiere, es obtener $\tilde{p}_G(\eta_i|\theta, y)$, que representa las varianzas marginales de la matriz de precisión R_θ . Si bien este es el método más rápido con resultados aceptables no es del todo confiable pues la aproximación Gaussiana puede causar errores de asimetría.

2. Realizar la aproximación de Laplace de nuevo:

$$\tilde{p}_{LA}(\eta_i|\theta, y) \propto \frac{p(\eta, \theta|y)}{\tilde{p}(\eta_{-i}|\eta_i, \theta, y)} \Big|_{\eta_{-i}=\eta_{-i}^*(\eta_i, \theta)},$$

para toda $i = 1, \dots, n$, donde $\tilde{p}(\eta_{-i}|\eta_i, \theta, y)$ es la aproximación Gaussiana con la configuración modal $\eta_{-i}^*(\eta_i, \theta)$. Sin embargo este método necesita la localización de la moda y la factorización de una matriz de tamaño $(n - 1) \times (n - 1)$ para cada i , que, para n grande, resulta de un gran costo computacional.

3. Utilizar el método de aproximación de Laplace simplificado, denotado como $\tilde{p}_{SLA}(\eta_i|\theta, y)$ que consiste en aplicar una expansión de Taylor al numerador y denominador de $\tilde{p}_{LA}(\eta_i|\theta, y)$ hasta el tercer orden. Con lo anterior se busca corregir la aproximación Gaussiana con un menor costo computacional.

Para una descripción más detallada acerca de la aproximación Gaussiana, aproximación de Laplace y aproximación de Laplace simplificada ver Rue et al. (2009).

Paso 3. Explorando $\tilde{p}(\theta|y)$

Sin pérdida de generalidad, sea $\theta = (\theta_1, \dots, \theta_m) \in \mathbb{R}^m$. Se encuentra la moda de $\tilde{p}(\theta|y)$, denotada por θ^* y se calcula la matriz Hessiana negativa H en la configuración modal. Luego, se estandariza θ para obtener una nueva variable z , definida como:

$$z = (V\Lambda^{1/2})^{-1}(\theta - \theta^*),$$

donde $H^{-1} = V\Lambda V'$ es la eigen-descomposición.

Si la dimensión de θ es pequeña, por ejemplo $m \leq 2$, la z -parametrización es usada para construir una cuadrícula que cubre la mayor parte de la densidad de $\tilde{p}(\theta|y)$, sin embargo, este método tiene un gran costo computacional a medida que m aumenta. Otra alternativa es utilizar un método llamado Bayes empírico, que se utiliza para poder integrar sobre $p(\theta|y)$, este método subestimaré la incertidumbre pero dará resultados aceptables dado que la variabilidad en el campo latente no está dominada por la variabilidad de los hiperparámetros. Un tercer método utilizando la moda θ^* y la matriz H , se encuentran algunos puntos en el espacio m -dimensional para aproximar la función desconocida con una superficie de segundo orden y se utiliza un diseño cuadrático clásico como el diseño compuesto central (CCD). Un CCD contiene un diseño factorial o factorial fraccionado incorporado con puntos centrales aumentados con un grupo de puntos que permiten estimar la curvatura, utilizar este método requiere menor costo computacional comparada con los otros dos modelos y preserva la capacidad de capturar la variabilidad en el espacio θ , aún cuando el espacio de búsqueda es muy amplio como para usar una cuadrícula.

En la figura [3.1](#) se observa la distribución de puntos usando la técnica de la cuadrícula y el método CCD.

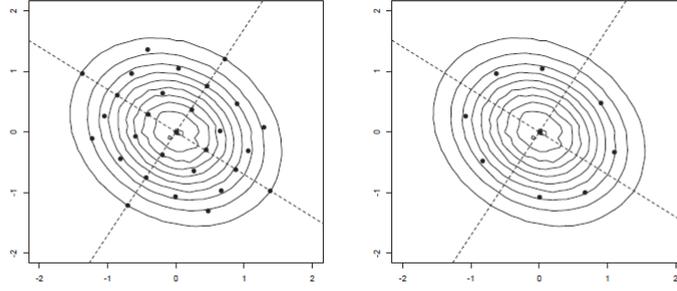


Figura 3.1: Ubicación de los puntos en un espacio θ de dos-dimensiones usando la cuadrícula (izquierda) y el método CCD (derecha) (obtenida de Wang et al., 2018).

Es común que si $m \leq 2$ entonces se utiliza el método de la cuadrícula, el caso contrario se usa el método CCD.

Aproximando $p(\theta_j|y)$

La distribución conjunta $p(\theta|y)$ puede ser aproximada como una distribución normal multivariada haciendo coincidir la moda y la curvatura en la moda de $\tilde{p}(\theta|y)$. Las verdaderas marginales pueden estar bastante sesgadas, por lo que se tienen que corregir las aproximaciones Gaussianas por la falta de sesgo. Esta corrección se realiza aproximando la distribución conjunta como una suma de mezclas de distribuciones normales con parámetros que pueden variar según diferentes ejes y sus direcciones. Entonces, las marginales se pueden calcular a través de la integración numérica de la distribución conjunta aproximada. El método anteriormente mencionado es el que usa INLA, si bien se obtienen resultados aceptables con un bajo costo computacional la integración numérica multidimensional se vuelve inestable cuando se ajustan modelos con un mayor número de hiperparámetros, lo que da como resultado marginales posteriores aproximadas con picos no deseados. Para solucionar este problema, Martins et al. (2013) propusieron un algoritmo libre de integración, donde la marginal posterior de cada θ_j se aproxima directamente mediante una mezcla de distribuciones normales como sigue:

$$\tilde{p}(\theta_j|y) = \begin{cases} N(0, \sigma_{j+}^2), & \theta_j > 0 \\ N(0, \sigma_{j-}^2), & \theta_j \leq 0, \end{cases}$$

donde los parámetros σ_{j-}^2 y σ_{j+}^2 se obtienen sin usar integración numérica.

Para ver más acerca de la teoría del método de INLA, así como más información de los tres pasos del método se recomienda ver Wang et al. (2018) y Rue et al. (2009). A continuación se presenta un ejemplo del modelo de INLA hecho paso a paso.

Ejemplo 6. Consideremos las observaciones $y = (y_1, \dots, y_n)$ y se consideran independientes e idénticamente distribuidas con $y_i | \mu, \sigma^2 \sim \text{Normal}(\mu, \sigma^2)$. Elegimos las distribuciones a priori independientes para μ y $\psi = 1/\sigma^2$ tal que:

$$\begin{aligned}\mu &\sim \text{Normal}(\mu_0, \sigma_0^2) \\ \psi &\sim \text{Gamma}(a, b).\end{aligned}$$

Entonces, se tiene lo siguiente:

$$y_i | \theta, \psi \sim \text{Normal}(\eta_i, \sigma^2),$$

donde $\eta_i = \theta = \mu$ y $\psi = 1/\sigma^2$. Ahora, se utilizará la distribución $\text{Gamma}(a, b)$ como distribución a priori para ψ . La primera parte se refiere a la distribución a posteriori del hiperparámetro ψ , que puede ser obtenido integrando sobre θ desde la distribución a posteriori conjunta como sigue:

$$p(\psi | y) \propto \int p(\theta, \psi | y) d\theta,$$

donde $p(\theta, \psi | y) \propto p(y | \theta, \psi) p(\theta) p(\psi)$. Para que el problema sea más sencillo podemos ver la distribución posterior como:

$$\begin{aligned}p(\psi | y) &= \frac{p(\theta, \psi | y)}{p(\theta | \psi, y)} \\ &\propto \frac{p(y | \theta, \psi) p(\theta) p(\psi)}{p(\theta | \psi, y)},\end{aligned}$$

donde las distribuciones involucradas son las siguientes:

$$\begin{aligned}
 y|\theta, \psi &\sim \prod_{i=1}^n \text{Normal}(\theta, 1/\psi) \\
 \theta &\sim \text{Normal}(\mu_0, \sigma_0^2) \\
 \psi &\sim \text{Gamma}(a, b) \\
 \theta|\psi, y &\sim \text{Normal}\left(\theta_n = \frac{\psi \sum_{i=1}^n y_i + \frac{\mu_0}{\sigma_0^2}}{n\psi + \frac{1}{\sigma_0^2}}, \sigma_n^2 = \frac{1}{n\psi + \frac{1}{\sigma_0^2}}\right).
 \end{aligned}$$

Se puede probar que $p(\psi|y)$ se puede escribir como sigue:

$$p(\psi|y) \propto \frac{1}{1/\sqrt{2\pi\sigma_n^2}} p(y|\theta, \psi) p(\theta) p(\psi)|_{\theta=\theta_n}.$$

Entonces, para evaluar la distribución posterior se eligen valores para ψ , se incluyen en el conjunto $\{\psi^{(j)}\}$ y para cada valor de dicho conjunto se obtiene la función de densidad como sigue:

$$p(\psi^{(j)}|y) \propto \frac{1}{1/\sqrt{2\pi\sigma_n^2}} p(y|\theta, \sigma^2) p(\theta) p(\psi^{(j)})|_{\theta=\theta_n, \psi=\psi^{(j)}}$$

donde

$$\theta_n = \frac{\psi^{(j)} \sum_{i=1}^n y_i + \frac{\mu_0}{\sigma_0^2}}{n\psi^{(j)} + \frac{1}{\sigma_0^2}} \quad \text{y} \quad \sigma_n^2 = \frac{1}{n\psi^{(j)} + \frac{1}{\sigma_0^2}}.$$

Para ver más detalles de cómo se obtuvo la aproximación de $p(\psi|y)$ ver Blangiardo et al (2015). Ahora, el segundo paso consiste en evaluar la distribución condicional completa $p(\theta|\psi, y)$ para cada $\psi \in \{\psi^{(j)}\}$ y $\theta \in \theta^{(l)}$. Lo anterior consiste en evaluar $p(\theta = \theta^{(l)}|\psi = \psi^{(j)}, y)$ usando la distribución $\text{Normal}(\theta_n, \sigma_n^2)$. Entonces, la distribución marginal posterior $p(\theta|y)$ puede ser obtenida integrando sobre ψ de la posterior conjunta $p(\theta, \psi|y)$ a través de una media ponderada finita como sigue:

$$p(\theta = \theta^{(l)}|y) \propto \sum_j p(\theta^{(l)}|\psi^{(j)}, y) p(\psi^{(j)}|y) \Delta_j,$$

donde $\Delta_j = 1/\sum_j p(\psi^{(j)}|y)$. Finalmente se presentan las gráficas de la distribución

posterior de σ^2 y μ .

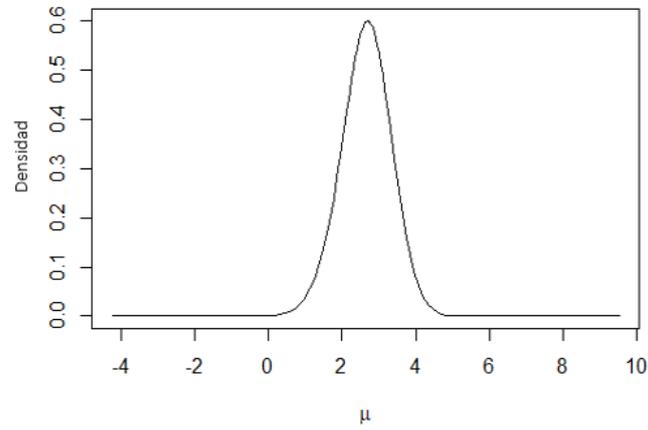


Figura 3.2: Distribución posterior de μ .

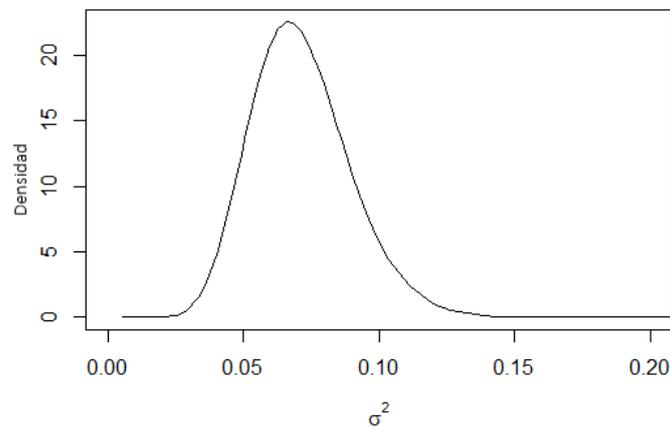


Figura 3.3: Distribución posterior de σ^2 .

El código donde se hacen los cálculos del método de INLA paso a paso se puede encontrar en el Apéndice [A](#).

Para tener más detalles de este ejemplo se recomienda revisar la sección 4.9 de Blangiardo et al (2015).

3.5. El paquete R-INLA

R-INLA es un paquete del software R para implementar aproximaciones de inferencia bayesiana utilizando el método de INLA ya visto.

A continuación se presentará de manera general la sintaxis de R-INLA.

Consideremos las covariables $x = (x_1, x_2)$ una función $f(\cdot)$ indexada por una tercera covariable z_1 ; entonces el predictor lineal visto como el modelo general de un LGM es presentado en R-INLA con el siguiente comando:

```
1 formula <- y~1 + x1 + x2 + f(z1, model='...'),
```

donde y , x_1 , x_2 y z_1 son las columnas del dataframe que contiene nuestros datos. Una observación es que el término 1 no es obligatorio, ya que se incluiría un intercepto por defecto. Ahora, en la función $f(\cdot)$ en el apartado `model="..."` especifica el tipo de función, por defecto se tiene `model="iid"`, que refiere a variables aleatorias independientes con distribución Gaussiana.

A continuación se muestra la lista de las otras opciones disponibles:

```
1 "linear"    "iid"     "mec"     "meb"     "rgeneric" "rw1"  "rw2"
2 "crw2"     "seasonal" "besag"   "besag2"  "bym"     "bym2"
3 "besagproper" "besagproper2" "fgn"    "fgn2"    "ar1"    "ar1c"
4 "ar"      "ou"     "intslope" "generic"  "generic0" "generic1"
5 "generic2" "generic3" "spde"    "spde2"   "spde3"   "iid1d"
6 "iid2d"    "iid3d"  "iid4d"   "iid5d"   "iidkd"   "2diid"  "z"
7 "rw2d"    "rw2diid" "slm"    "matern2d" "dmatern"  "copy"
8 "clinear" "sigm"    "revsigm" "log1exp" "logdist"
```

Finalmente, el modelo de INLA se corre utilizando la función `inla()` con el siguiente comando:

```
1 inla(formula, family="...", data),
```

donde “data” es un dataframe que contiene todas las variables de interés y “family” especifica la distribución de los datos, el paquete INLA provee de cerca de 80 distribuciones. A continuación se muestra una lista con algunas de las distribuciones disponibles:

```

1 "poisson" "xpoisson" "cenpoisson" "cenpoisson2" "gpoisson"
2 "binomial" "xbinomial" "pom" "gamma" "gammaturv"
3 "gammajw" "gammajwsurv" "gammacount" "qkumar" "qloglogistic"
4 "qloglogisticsurv" "beta" "betabinomial" "betabinomialna"
5 "cbinomial" "nbinomial" "nbinomial2" "simplex" "gaussian"
6 "zeroinflatedpoisson0" "zeroinflatedcenpoisson0"
7 "zeroinflatedbetabinomial0" "zeroinflateddbinomial0"
8 "zeroinflatednbinomial0"

```

A continuación se dará un ejemplo de un modelo sencillo de regresión lineal, se realizará con INLA y con MCMC en el Software R. El código utilizado se puede ver en el Apéndice [A](#).

Ejemplo 7. Se realizó un modelo de regresión lineal simple, es decir, se desea estimar los parámetros β_0 y β_1 del modelo $y = \beta_0 + \beta_1 x + \varepsilon$, se simuló una muestra aleatoria de tamaño 100 de una distribución uniforme para la variable dependiente, respecto a la variable se consideró multiplicar por 2 la variable x y después sumarle 1. Por otro lado se consideró que el error ε sigue una distribución normal con media 0 y varianza 0.1. Con lo anterior se buscará estimar los valores de β_0 y β_1 , que, por construcción, deben ser cercanos a 2 y 1, respectivamente.

Utilizando INLA se obtuvo el siguiente resumen del modelo:

```

1 Fixed effects:
2           mean      sd 0.025quant  0.5quant  0.975quant  mode
3 (Intercept) 1.019 0.024      0.972    1.019    1.066 1.019
4 x           1.982 0.038      1.907    1.982    2.057 1.982

```

El resumen incluye la media posterior, la desviación estándar y los cuartiles de los efectos fijos (que en este caso son los valores estimados de β_0 y β_1).

Utilizando MCMC se obtuvo lo siguiente:

```

1           mu.vect sd.vect   2.5%   25%   50%   75%  97.5%
2 beta_0      0.966  0.038  0.885  0.943  0.966  0.991  1.041
3 beta_1      2.040  0.080  1.877  1.994  2.042  2.088  2.199

```

Se puede observar que ambos modelos hacen una estimación muy buena de los parámetros de interés. Particularmente, se puede ver que utilizar MCMC nos da una estimación de parámetros con mayor desviación estándar, además, el intervalo de confianza al 95 % obtenido con INLA es más pequeño que el obtenido con MCMC.

Capítulo 4

Modelos lineales generalizados

4.1. Regresión lineal Bayesiana

Comenzaremos por hacer un breve recordatorio del modelo lineal usual. Sean $\{x_{i1}, \dots, x_{ip}\}$ con $i = 1, \dots, n$, n unidades observadas, donde cada una representa una medición del vector de covariables (x_1, \dots, x_p) . Consideremos a y la variable respuesta asociada a las unidades observadas ya mencionadas. Definimos el modelo de regresión lineal múltiple como sigue:

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

que visto de forma matricial se obtiene lo siguiente:

$$y = X\beta + \varepsilon,$$

donde

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Recordemos que, para este modelo, se hace un supuesto importante, esto es que el término de error ε en el modelo $y = X\beta + \varepsilon$ sigue una distribución normal con media 0 y varianza constante σ^2 , además para $i \neq j$, ε_i y ε_j son independientes. Además, aplicando estimación por máxima verosimilitud obtenemos los estimadores del modelo,

estos son:

$$\hat{\beta} = (X'X)^{-1}X'y$$

y

$$\hat{\sigma}^2 = \frac{1}{n}(y - X\hat{\beta})'(y - X\hat{\beta}),$$

donde $\hat{\beta}$ tiene la propiedad de ser un estimador insesgado. Por otro lado, $\hat{\sigma}^2$ no es un estimador insesgado, pero S^2 sí lo es:

$$S^2 = \frac{1}{n-p-1}(y - X\hat{\beta})'(y - X\hat{\beta}).$$

Ahora, veremos cómo hacer inferencia bayesiana en el modelo $y = X\beta + \varepsilon$. Se define a la precisión como $\tau = \sigma^{-2}$. Basándonos en los supuestos del modelo se tiene que:

$$y|\beta, \tau \sim N(X\beta, \tau^{-1}I),$$

donde se supone a β y τ independientes. De este modo, la distribución posterior conjunta de los parámetros desconocidos se ve como sigue:

$$\pi(\beta, \tau|X, y) \propto L(\beta, \tau|X, y)p(\beta)p(\tau),$$

donde $p(\beta)$ y $p(\tau)$ son las distribuciones a priori de β y τ respectivamente. Al utilizar el método de INLA se asume que el modelo es un LGM, entonces a β se le asigna un a priori Gaussiana. Para τ usualmente se asume un a priori con una varianza grande. Dos propuestas usuales de distribuciones para β y τ son:

$$\beta \sim N_{p+1}(c_0, V_0), \quad \tau \sim \text{Gamma}(a_0, b_0),$$

es decir, β sigue una distribución normal multivariada de dimensión $p + 1$ con media y varianza conocida y τ una distribución con media a_0/b_0 y varianza a_0/b_0^2 conocidas.

4.1.1. Predicción

Sea $\hat{x} = (\hat{x}_1, \dots, \hat{x}_p)'$ un vector de variables explicativas y sea \hat{y} la variable respuesta. Recordemos que la predicción Bayesiana está basada en la distribución posterior predictiva $p(\hat{y}|y)$.

Sea $\theta = (\beta, \tau)$, tenemos que:

$$\begin{aligned} p(\hat{y}|y) &= \frac{p(\hat{y}, y)}{p(y)} = p(y)^{-1} \int p(\hat{y}|\theta)p(y|\theta)p(\theta)d\theta \\ &= p(y)^{-1} \int p(\hat{y}|\theta)p(\theta|y)p(y)d\theta \\ &= \int p(\hat{y}|\theta)p(\theta|y)d\theta. \end{aligned}$$

La forma analítica de la distribución predictiva posterior en la mayoría de los modelos de regresión no está disponible. En el análisis bayesiano la predicción se puede realizar mediante simulación predictiva posterior, es decir, extrayendo muestras aleatorias de $p(\hat{y}|y)$.

4.1.2. Criterios de selección de modelos

Debido a la gran variedad de modelos que se pueden obtener en los diversos modelos propuestos por los datos es indispensable tener algún método que pueda medir qué tan consistentes son los datos con un modelo dado. A continuación se presentan algunos criterios de selección de modelos usuales.

Criterio de Información de Akaike

El Criterio de Información de Akaike (AIC) es el siguiente:

$$\text{AIC} = -2 \log(L(\hat{\theta}|x)) + 2k,$$

donde $L(\hat{\theta}|x)$ es función de verosimilitud evaluada en el estimador $\hat{\theta}$ y k es el número de parámetros estimados.

Notemos que $-2\log(L(\hat{\theta}|x))$ es una función decreciente, entonces, cuando la función

de verosimilitud es grande la expresión de la bondad de ajuste tiende a un valor pequeño. Lo anterior busca medir la falta de buen ajuste entre una distribución hipotética y una distribución teórica. Por otro lado, el término $2k$ penaliza en número de parámetros estimados, por lo que mide la complejidad del modelo. En conclusión, dado un modelo con muchos parámetros se tiene un mayor ajuste, sin embargo, hay mayor complejidad, por lo que el AIC es un equilibrio entre ambas cosas. Entonces, se buscará el modelo que tenga el AIC más pequeño.

Criterio de Información Bayesiano

El Criterio de Información Bayesiano (BIC) es el siguiente:

$$\text{BIC} = -2 \log(L(\hat{\theta}|x)) + 2k \log(n),$$

donde, a diferencia del AIC, el BIC penaliza de otra forma a los parámetros, esto es, mientras más complejo sea el modelo se tendrá que el BIC tendrá una mayor penalización que el AIC. De manera análoga, se buscará el modelo que tenga el BIC más pequeño.

Ordenada Predictiva Condicional

La Ordenada Predictiva Condicional se define como:

$$\text{CPO}_i = p(y_i | \mathbf{y}_{-i}),$$

donde \mathbf{y}_{-i} denota todos los datos menos la i -ésima observación. La Ordenada Predictiva Condicional expresa la probabilidad posterior de observar el valor de y_i cuando el modelo se ajusta a todos los datos excepto a y_i . Un valor alto implica un mejor ajuste del modelo a y_i .

Criterio de Información de la Devianza

Se define la devianza del modelo como:

$$D(\theta) = -2 \log(L(\hat{\theta}|x)).$$

En el modelo Bayesiano $D(\theta)$ es una variable aleatoria, entonces podemos considerar el

valor esperado de la devianza, $\mathbb{E}[D(\theta)]$ bajo la distribución posterior como una medida de ajuste. Para contar los parámetros, se introduce el concepto del número efectivo de parámetros p_D :

$$p_D = \mathbb{E}[D(\theta)] - D(\mathbb{E}[\theta]) = \bar{D} - D(\bar{\theta}),$$

entonces el Criterio de Información de la Devianza (DIC) es el siguiente:

$$\text{DIC} = \bar{D} + p_D.$$

Usando la misma idea del AIC y el BIC, se elegirá el modelo con menor DIC.

4.1.3. Análisis de la varianza

El análisis de la varianza, conocido como ANOVA, es una colección de modelos que es usado para determinar el grado de similitud o diferencia que hay entre dos o más grupos de datos.

Definición 4.1.1. *Dado un experimento, se define el factor de dicho experimento como la característica que se desea controlar, y el nivel como las categorías estudiadas del factor. Se le conoce como tratamiento a la combinación de niveles de los factores estudiados.*

Consideremos el caso de un único factor A, que tiene niveles $i = 1, 2, \dots, a$, además, se asumirá que la variable respuesta Y sigue una distribución normal y el efecto de la variable factor tiene una influencia en la media de Y . El modelo queda especificado como sigue:

$$y_{ij} = \mu_0 + \alpha_i + \varepsilon_{ij},$$

donde y_{ij} es el j -ésimo valor observado del i -ésimo nivel del factor, $\varepsilon_{ij} \sim N(0, \sigma^2)$ y α_i es el efecto del i -ésimo tratamiento y μ_0 es la media.

El modelo ANOVA anterior puede extenderse de forma sencilla añadiendo variables categóricas. Consideremos dos factores, A y B, donde A tiene “a” niveles y B tiene “b” niveles, consideraremos un diseño factorial de 2×2 , que consiste en cruzar los valores

de A y B. Supongamos que hay n replicas de cada celda de la forma ab . La observación y_{ijk} es la siguiente:

$$y_{ijk} = \mu_0 + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk},$$

donde $i = 1, \dots, a$, $j = 1, \dots, b$ y $k = 1, \dots, n$, μ_0 es la media, α_i es el efecto principal del nivel i del factor A, β_j es el efecto principal del nivel j del factor B, γ_{ij} es el efecto de interacción del i -ésimo nivel de A y el j -ésimo nivel de B y $\varepsilon_{ijk} \sim N(0, \sigma^2)$ es el término de error.

En el modelo $y_{ij} = \mu_0 + \alpha_i + \varepsilon_{ij}$, si α_i es considerado para los modelos de un factor, entonces son vistas como efectos fijos. Por otro lado, si son considerados como una muestra aleatoria de una distribución, por ejemplo $\alpha_i \sim N(0, \sigma_\alpha^2)$, entonces son considerados como efectos aleatorios.

En el sentido Bayesiano, todas las variables son vistas como variables aleatorias, para los efectos fijos se le suele considerar como $\alpha_i \sim N(0, \sigma_0^{-1})$ donde σ_0^{-1} es fija, y para efectos aleatorios correspondientes al factor A se le suele asignar una distribución a priori $\alpha_i | \sigma_\alpha^2 \sim N(0, \sigma_\alpha^2)$ donde σ_α^2 es una distribución a priori en un hiperparámetro.

4.2. GLM's

Los modelos lineales generalizados (GLM), son modelos que generalizan la regresión lineal ordinaria al permitir que los modelos se relacionen con la variable de respuesta a través de una función de liga, o también llamada de enlace, y permitir que la magnitud de la varianza de cada medición sea una función de su valor predicho. Así, estos modelos permiten describir variables de respuesta que tienen una distribución de error diferente a la normal. Evitan tener que seleccionar ciertas transformaciones de los datos para lograr los objetos posiblemente conflictivos de normalidad, linealidad y/u homogeneidad de varianza.

Sean y_1, y_2, \dots, y_n un conjunto de observaciones de las variables aleatorias Y_1, Y_2, \dots, Y_n . A continuación se presentan los componentes de un GLM:

- Componente aleatorio: las variables independientes Y_i 's que se asumen generadas de una distribución de la familia exponencial.

- Predictor lineal: la combinación lineal de los predictores, es decir:

$$\theta_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} = x_i' \beta,$$

donde $\beta = (\beta_0, \beta_1, \dots, \beta_p)'$, $x_i = (1, x_{i1}, \dots, x_{ip})'$ y x_{ij} , con $j = 1, \dots, p$ es el valor de la j -ésima covariable para la i -ésima observación.

- Función liga: el valor esperado de la variable respuesta, $\mu = \mathbb{E}[Y_i]$ y el predictor lineal son relacionados a través de la función liga, o enlace, $g(\cdot)$:

$$g(\mu_i) = \theta_i.$$

A continuación, en la tabla [4.1](#) se muestran algunas funciones liga comunes, así como su función inversa y su función de varianza.

Familia	$\theta = g(\mu)$	$\mu = g^{-1}(\theta)$	función de varianza
Normal	μ	$\mu = \theta$	1
Poisson	$\log(\mu)$	$\exp(\theta)$	μ
Binomial	$\log(\mu/(1 - \mu))$	$\exp(\theta)/(1 + \exp(\theta))$	$\mu(1 - \mu)$
Gamma	μ^{-1}	θ^{-1}	μ^2
Gaussiana inversa	μ^{-2}	$\theta^{-1/2}$	μ^3

Tabla 4.1: Funciones liga

Como recordatorio, se hablará un poco acerca de las distribuciones que pertenecen a la familia exponencial.

Definición 4.2.1. *Se dirá que $f(x; \theta)$ pertenece a la familia exponencial si su función de densidad se puede escribir de la siguiente manera:*

$$f(x; \theta) = a(\theta)b(x)\exp\{c(\theta)d(x)\}, \quad \theta \in \Theta, \quad -\infty < x < \infty$$

donde $a(\theta)$ y $c(\theta)$ son funciones de θ y $b(x)$ y $d(x)$ son funciones de x .

Algunos ejemplos de distribuciones que pertenecen a la familia exponencial son los siguientes:

1. Sea $X \sim \exp(\theta)$, es decir, X tiene una distribución exponencial, entonces su función de densidad queda definida como sigue:

$$f(x; \theta) = \theta \exp(-\theta x) \mathbb{1}_{(0, \infty)}.$$

X pertenece a la familia exponencial, ya que $a(\theta) = \theta$, $b(x) = 1_{(0, \infty)}$, $c(\theta) = -\theta$ y $d(x) = x$.

2. Sea $X \sim \text{Poisson}(\theta)$, es decir, X tiene una distribución Poisson, entonces su función de densidad queda definida como sigue:

$$f(x; \theta) = \frac{\exp(-\theta) \theta^x}{x!} \mathbb{1}_{\{0, 1, \dots\}} = \exp(-\theta) \exp(x \log(\theta)) \frac{1}{x!} \mathbb{1}_{\{0, 1, \dots\}}.$$

X pertenece a la familia exponencial, ya que $a(\theta) = \exp(-\theta)$, $b(x) = 1/x!$, $c(\theta) = \log(\theta)$ y $d(x) = x$.

3. Sea $X \sim \text{Geométrica}(\theta)$, es decir, X tiene una distribución Geométrica, entonces su función de densidad queda definida como sigue:

$$f(x; \theta) = \theta(1 - \theta)^x \mathbb{1}_{\{0, 1, \dots\}} = \theta \mathbb{1}_{\{0, 1, \dots\}} \exp(x \log(1 - \theta)).$$

X pertenece a la familia exponencial, ya que $a(\theta) = \theta$, $b(x) = \mathbb{1}_{\{0, 1, \dots\}}$, $c(\theta) = -\theta$ y $d(x) = x$.

4. Sea $X \sim \text{Bernoulli}(\theta)$, es decir, X tiene una distribución Bernoulli, entonces su función de densidad queda definida como sigue:

$$f(x; \theta) = \theta^x (1 - \theta)^{1-x} = (1 - \theta) \mathbb{1}_{\{0, 1\}} \exp\left(x \log\left(\frac{\theta}{1 - \theta}\right)\right).$$

X pertenece a la familia exponencial, ya que $a(\theta) = 1 - \theta$, $b(x) = \mathbb{1}_{\{0,1\}}$, $c(\theta) = -\theta/(1 - \theta)$ y $d(x) = x$. Una observación importante es que $f(x; \theta)$ se puede ver también como sigue:

$$\begin{aligned} f(x; \theta) &= \theta^x (1 - \theta)^{1-x} \\ &= \exp \{ \log(\theta^x (1 - \theta)^{1-x}) \} \\ &= \exp \{ x \log(\theta) + (1 - x) \log(1 - \theta) \} \\ &= \exp \left\{ x \log \left(\frac{\theta}{1 - \theta} \right) + \log(1 - \theta) \right\}. \end{aligned}$$

4.3. Respuestas binarias

En diversas aplicaciones la variable de respuesta toma uno de los dos únicos valores posibles que representan el éxito y el fracaso de la variable de interés. Un caso particular de los Modelos Lineales Generalizados es la regresión logística, que se utiliza para estimar la probabilidad de una respuesta binaria basada en una o más variables predictoras.

Consideremos una variable aleatoria $Y \sim \text{Bernoulli}(\pi)$, tal que su probabilidad de éxito es π , es decir, $\mathbb{P}(Y = 1) = \pi$. Su función de densidad queda definida como sigue:

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

Entonces, como se mencionó antes, se tiene que la distribución Bernoulli pertenece a la familia exponencial, donde el parámetro canónico θ es igual a $\log(\pi/(1 - \pi))$, a lo anterior se le conoce como la función liga logit de π , que conduce al modelo de regresión logística.

Un modelo logístico con k variables explicativas es el siguiente:

$$\text{logit}(\pi_i) = \log \left(\frac{\pi}{1 - \pi} \right) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}.$$

Al despejar π se tiene que:

$$\pi = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik})}{1 - \exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik})},$$

si consideramos $\alpha = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}$, entonces lo anterior queda simplificado como:

$$\pi = \frac{\exp(\alpha)}{1 + \exp(\alpha)}.$$

4.4. Respuestas para conteos

En diversos estudios se tiene el objetivo de contabilizar el número de ocurrencias de una variable de interés. Entonces, para este tipo particular de datos, se tiene que las observaciones son valores enteros donde el interés es contar más que agrupar o clasificar. Una observación importante es que la distribución de conteos es discreta, lo que lleva a la posibilidad de que sea sesgada. Si se tuviera un interés en aplicar un modelo de regresión lineal se podrían presentar algunos problemas, principalmente se puede presentar el hecho de que el modelo produzca valores negativos, situación que no tiene sentido al estar realizando conteos.

A continuación se presentarán dos modelos utilizados para realizar conteos.

4.4.1. Regresión Poisson

Sea $Y \sim \text{Poisson}(\mu)$, donde μ representa la media de las ocurrencias del evento de interés. Esta distribución tiene como característica que $\mathbb{E}[Y] = \mu = \text{Var}(Y)$. Usualmente μ se representa a través de una tasa.

Una señal importante que nos indica que es adecuado trabajar un modelo de regresión Poisson es cuando se intenta modelar a μ , se pueden tomar en cuenta distintos niveles de exposición.

Consideremos las variables Y_1, Y_2, \dots, Y_N independientes, donde cada Y_i representa el número de eventos ocurridos de un total de expuestos n_i y sean x_i las covariables

asociadas al evento de interés.

La esperanza de Y_i puede verse como:

$$\mathbb{E}(Y_i) = \mu_i = n_i\theta_i.$$

La tasa de ocurrencia es de la forma:

$$\theta_i = \exp(x_i'\beta).$$

Para el modelo de regresión Poisson se considera la liga logaritmo, de este modo, aplicando la función logaritmo a la expresión anterior se obtiene el modelo de regresión Poisson que se define como sigue:

$$\log(\mu_i) = \log(n_i) + \beta_0 + \beta_1x_{i1} + \cdots + \beta_px_{ip},$$

donde $\log(n_i)$ se le conoce como *offset*, que representa una constante conocida que especifica los diferentes niveles de exposición y se incorpora en la estimación de los parámetros. Hay ocasiones en que no se logra que el modelo se ajuste de manera adecuada a los datos, dos motivos recurrentes son los siguientes:

- El modelo no está bien especificado, esto es, hay variables predictorias que no se han incluido en el modelo o bien, no existen relaciones lineales.
- Existe sobredispersión, es decir, $\mathbb{E}[Y] < \text{Var}(Y)$.

Al revisar la construcción del modelo de regresión Poisson podemos notar que este es muy restrictivo en el hecho de que $\mathbb{E}[Y] = \mu = \text{Var}(Y)$, lo anterior ocasiona que fallen los conteos. A continuación se dará un ejemplo del modelo de regresión Poisson utilizando INLA.

Ejemplo 8. Se trabajará con la base “crabs”, que es una base que contiene información acerca de la anidación de cangrejos herradura. En este estudio, cada cangrejo herradura hembra en el estudio tenía un cangrejo macho unido a ella en su nido. El estudio

investigó posibles factores que afectan al hecho de que la cangreja tenía otros machos, llamados satélites, que residían cerca de ella. Las variables explicativas que se supone afectan esto son: el color del cangreja (COLOR), la condición de la columna vertebral (SPINE), el peso (WEIGHT) y el ancho del caparazón (WIDTH). La variable de respuesta para cada cangreja es su número de satélites (SATELLITES).

A continuación se muestra en la tabla [4.2](#) las variables de interés y una breve descripción de ellas.

Nombre de la variable	Descripción	Valor
SATELLITES	número de satélites por cada cangrejo hembra	enteros
COLOR	color del cangrejo	1 = media luz 2 = medio 3 = medio obscuro 4 = obscuro
SPINE	condición de la columna del cangrejo	1 = ambos bien 2 = uno desgastado o roto 3 = ambos desgastados o rotos
WEIGHT	peso del cangrejo	kilogramos
WIDTH	ancho del caparazón	centímetros

Tabla 4.2: Tabla de valores de la base crabs

El modelo se hará considerando las variables COLOR, SPINE y WIDTH. Al ajustar el modelo con INLA se obtiene lo siguiente:

	mean	sd	0.025quant	0.5quant	0.975quant	mode
1 (Intercept)	-0.049	0.254	-0.550	-0.048	0.446	-0.046
2 COLOR2	-0.269	0.168	-0.591	-0.272	0.067	-0.277
3 COLOR3	-0.523	0.194	-0.900	-0.525	-0.138	-0.528
4 COLOR4	-0.543	0.225	-0.987	-0.543	-0.102	-0.542
5 SPINE2	-0.162	0.211	-0.589	-0.157	0.241	-0.148
6 SPINE3	0.092	0.120	-0.139	0.091	0.330	0.089

8 WIDTH	0.548	0.073	0.403	0.548	0.690	0.549
---------	-------	-------	-------	-------	-------	-------

Podemos observar que la media posterior del parámetro WIDTH es 0.548 y su intervalo al 95 % de confianza es (0.403, 0.690), es decir, no está considerado el 0 como posible valor para el parámetro, entonces podemos afirmar que hay un efecto positivo del ancho del caparazón de las hembras en la cantidad de machos que hay alrededor de ella. Caso contrario con las variables COLOR y SPINE, donde los parámetros no resultan ser significativos, en el sentido bayesiano, ya que en un intervalo del 95 % se considera el 0 como posible valor, por lo que no se puede determinar si esos efectos son positivos o negativos en función de los datos.

4.4.2. Regresión Binomial Negativa

Sean $\nu_i \sim \text{Gamma}(\alpha, \alpha)$, con $i = 1, \dots, n$, variables aleatorias. Entonces la función de densidad de cada ν_i se puede ver como sigue:

$$f(\nu) \propto x^{\alpha-1} \exp(-\alpha\nu) \mathbb{1}_{\{\nu>0\}}.$$

Notemos que, condicionado a ν_i , la variable Y_i sigue una distribución Poisson con media $\nu_i\mu_i$, es decir, $Y_i|\nu_i \sim \text{Poisson}(\nu_i\mu_i)$. Específicamente, se desea hacer notar que una distribución Poisson con una distribución a priori gamma se ve como una distribución binomial negativa. Entonces, marginalmente Y_i sigue una distribución binomial negativa dada por:

$$P(Y_i = y; \alpha, \mu_i) = \frac{\Gamma(y + \alpha)}{\Gamma(\alpha)y!} \left(\frac{\alpha}{\mu_i + \alpha} \right)^\alpha \left(\frac{\mu_i}{\mu_i + \alpha} \right)^y,$$

donde $y \in \{0, 1, 2, 3, \dots\}$.

Consideremos las variables Y_1, Y_2, \dots, Y_N independientes, donde cada Y_i representa el número de eventos ocurridos de un total de expuestos n_i . Para el modelo de regresión binomial negativa se considera la liga logaritmo, de este modo, el modelo se define como sigue:

$$\log(\mu_i) = \log(n_i) + \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}.$$

En conclusión, el modelo de regresión binomial negativa es una alternativa útil respecto a la regresión Poisson pues permite la sobredispersión. Para este caso se tiene que la media y varianza marginal son: $\mathbb{E}[Y_i] = \mu_i$ y $\text{Var}(Y_i) = \mu_i + \mu_i^2/\alpha$ respectivamente. Se puede notar que el parámetro α cuantifica la sobredispersión. Es usual que para el modelo de regresión binomial negativa se defina el parámetro de dispersión $\phi = 1/\alpha$, entonces, para valores muy grandes de α se tiene que $\phi \rightarrow 0$, lo que implica que no hay sobredispersión, por lo que el modelo binomial negativo se reduce al modelo Poisson.

A continuación se presenta la adaptación del ejemplo **Ejemplo 8**, ahora utilizando el modelo de regresión binomial negativa y haciendo una comparación respecto al modelo Poisson.

Ejemplo 9. Al ajustar el modelo con INLA considerando las variables COLOR, SPINE y WIDTH se obtiene lo siguiente:

	mean	sd	0.025quant	0.5quant	0.975quant	mode
1 (Intercept)	-0.313	0.575	-1.442	-0.314	0.820	-0.315
2 COLOR2	-0.320	0.378	-1.087	-0.311	0.400	-0.295
3 COLOR3	-0.596	0.414	-1.429	-0.589	0.200	-0.576
4 COLOR4	-0.579	0.473	-1.521	-0.575	0.340	-0.568
5 SPINE2	-0.244	0.379	-0.978	-0.248	0.511	-0.255
6 SPINE3	0.040	0.244	-0.448	0.042	0.513	0.047
7 WIDTH	0.697	0.177	0.355	0.694	1.052	0.689

Podemos observar que la media posterior del parámetro WIDTH es 0.697 y su intervalo al 95 % de confianza es (0.355, 1.052), es decir, no está considerado el 0 como posible valor para el parámetro, entonces podemos afirmar que hay un efecto positivo del ancho del caparazón de las hembras en la cantidad de machos que hay alrededor de ella. Caso contrario con las variables COLOR y SPINE, donde los parámetros no resultan ser significativos, en el sentido bayesiano, ya que en un intervalo del 95 % se considera el 0 como posible valor, por lo que no se puede determinar si esos efectos son positivos o negativos en función de los datos. Por otro lado, podemos notar que las medias posteriores de los parámetros en ambos modelos no difieren demasiado, salvo por la del intercepto. Finalmente, se calcula el DIC para ambos modelos, se obtiene que el DIC para el modelo Poisson es 918.98 y para el modelo binomial negativa se obtiene un DIC de 761.18, por lo que se opta por elegir el modelo de regresión binomial negativa. El código de los ejemplos 8 y 9 se pueden encontrar en el Apéndice [A](#).

Finalmente, la pregunta natural que surge es saber en qué momento se debe usar un modelo u otro. Como ya se comentó en la especificación del modelo de regresión

Poisson, éste se caracteriza por deber cumplir que $\mathbb{E}[Y] = \mu = \text{Var}(Y)$, es decir, no contempla la existencia de sobredispersión excesiva. La imposición inapropiada de esta restricción puede resultar en que los errores estándar estimados excesivamente pequeños de las estimaciones de los parámetros. Por lo anterior, ante la presencia de excesiva sobredispersión en los datos que se analizan la regresión Binomial Negativa es la opción muy conveniente para poder “relajar” las condiciones del modelo de regresión Poisson.

Para más información relacionada a la especificación y selección de modelos de regresión Poisson y Binomial Negativa revisar Hilbe (2011).

Capítulo 5

Modelo de cero inflado con INLA

5.1. Introducción

Como ya se mencionó, la especificación de modelos lineales generalizados, como Poisson o Binomial Negativa, para datos de conteo puede ser inapropiada debido al exceso de ceros en los datos en comparación con lo esperado del modelo. Para tratar este problema se propone especificar los modelos cero inflados.

En esta sección, se presentarán los modelos cero inflados basados en una distribución Poisson (ZIP) y los modelos cero inflados basados en la distribución Binomial Negativa (ZINB).

Estos modelos son una mezcla de dos componentes: una masa puntual en cero y una distribución de conteo, por ejemplo la Poisson o Binomial. De esta forma, dichos modelos distinguen entre ceros estructurales, para unidades donde el cero es el único valor observable, y ceros muestrales, para unidades en las que observamos un cero, pero también se pueden haber registrado otros valores.

En particular, hay dos procesos de generación de datos donde, para determinar el proceso, se puede ver como un ensayo Bernoulli como sigue:

Para la i -ésima observación el primer proceso genera ceros con probabilidad ϕ y, por otro lado, el segundo proceso genera resultados de una distribución paramétrica usadas para conteos, por ejemplo la distribución Poisson o Binomial Negativa, con probabilidad $1 - \phi$. Entonces, el modelo queda expresado como:

$$Y_i \sim \begin{cases} 0 & \text{con probabilidad } \phi \\ g(Y_i|x_i) & \text{con probabilidad } 1 - \phi, \end{cases}$$

donde $g(\cdot)$ es una función de densidad. De este modo, se tiene que la probabilidad de $p(Y_i = y_i|x_i)$ queda definida como sigue:

$$p(Y_i = y_i|x_i) = \phi \mathbb{1}_{\{y_i=0\}} + (1 - \phi)g(y_i|x_i).$$

5.2. Modelo de cero inflado Poisson

Sea y_i , con $i = 1, \dots, n$, entonces la función de probabilidad que define el modelo cero inflado Poisson es el siguiente:

$$p(y_i|\mu_i) = \phi \mathbb{1}_{\{y_i=0\}} + (1 - \phi) \frac{\exp(-\mu_i)\mu_i^{y_i}}{y_i!}.$$

Ahora, la probabilidad de observar un cero en la i -ésima observación es $\phi + (1 - \phi)\exp(-\mu_i)$ y la media y varianza son las siguientes:

$$\mathbb{E}[y_i] = (1 - \phi)\mu_i,$$

$$\text{Var}(y_i) = (1 - \phi)\mu_i + \frac{\phi}{1 + \phi}(1 - \phi)\mu_i.$$

5.3. Modelo de cero inflado Binomial Negativa

Sea y_i , con $i = 1, \dots, n$, y sea n_i , con $i = 1, \dots, k$, entonces la función de probabilidad que define el modelo cero inflado Binomial Negativa es el siguiente:

$$p(y_i|\mu_i) = \phi \mathbb{1}_{\{y_i=0\}} + (1 - \phi) \binom{n_i}{y_i} \mu_i^{y_i} (1 - \mu_i)^{n_i - y_i}.$$

Ahora, la probabilidad de observar un cero en la i -ésima observación es: $\phi + (1 - \phi)(1 - \mu_i)^{n_i - y_i}$ y la media y varianza son las siguientes:

$$\mathbb{E}[y_i] = (1 - \phi)n_i\mu_i,$$

$$\text{Var}(y_i) = (1 - \phi)n_i\mu_i(1 - \mu_i(1 - \phi n_i)).$$

A continuación se dará un ejemplo comparando el modelo Poisson y el modelo Binomial Negativa.

Ejemplo 10. Se trabajará con la base “articles” que se encuentra en la librería “brinla” en R, que a su vez son tomados de Long (1997). El estudio relacionado a la base “articles” busca examinar cómo afecta el número de artículos publicados por un científico considerando variables como: género, el estado civil, el número de hijos pequeños, el prestigio del programa de posgrado y el número de artículos publicados por el mentor del científico. El código de este ejemplo se puede encontrar en el Apéndice [A](#).

A continuación se muestra en la tabla [5.1](#) las variables de interés y una breve descripción de ellas.

Nombre de la variable	Descripción	Valor
fem	género	0= hombre; 1=mujer
ment	número de artículos publicados por el mentor	enteros
phd	prestigio del programa de posgrado	numérico
mar	estado civil	0 = soltero; 1= casado
kid5	número de niños pequeños	enteros
art	número de artículos	enteros

Tabla 5.1: Descripción de datos.

Ahora, se procede a mostrar la frecuencia del número de artículos publicados por científicos y la proporción de estos.

1	0	1	2	3	4	5	6	7	8	9	10	11	12	16	19
2	275	246	178	84	67	27	17	12	1	2	1	1	2	1	1

CAPÍTULO 5. MODELO DE CERO INFLADO CON INLA

1	0	1	2	3	4	5	6	7	8
2	0.301	0.269	0.195	0.092	0.073	0.030	0.019	0.013	0.001
3	9	10	11	12	16	19			
4	0.002	0.001	0.001	0.002	0.001	0.001			

Se nota que la proporción de científicos que no publican artículos es 0.301, es decir, hay una gran presencia de ceros en la variable respuesta.

Se ajusta un modelo de cero inflado Poisson y cero inflado Binomial Negativa con R-INLA y se obtienen los siguientes resultados:

Para el modelo cero inflado Poisson se obtuvieron los siguientes valores:

1		mean	sd	0.025quant	0.5quant	0.975quant	mode
2	(Intercept)	0.552	0.114	0.328	0.552	0.774	0.552
3	fem1	-0.231	0.059	-0.347	-0.231	-0.117	-0.231
4	mar1	0.132	0.066	0.002	0.132	0.262	0.132
5	kid5	-0.170	0.043	-0.256	-0.170	-0.086	-0.170
6	phd	0.003	0.028	-0.053	0.003	0.059	0.003
7	ment	0.022	0.002	0.017	0.022	0.026	0.022

De lo anterior se puede notar que la variable “phd” no es significativa en el sentido bayesiano, pues en un intervalo del 95 % de credibilidad se considera el 0 como probable valor a considerar.

Ahora, para el modelo cero inflado Binomial Negativa se obtuvieron los siguientes valores:

1		mean	sd	0.025quant	0.5quant	0.975quant	mode
2	(Intercept)	0.274	0.139	0.000	0.274	0.546	0.274
3	fem1	-0.218	0.072	-0.360	-0.218	-0.076	-0.218
4	mar1	0.149	0.082	-0.011	0.149	0.310	0.149
5	kid5	-0.176	0.053	-0.280	-0.176	-0.073	-0.176
6	phd	0.015	0.036	-0.056	0.015	0.085	0.015
7	ment	0.029	0.003	0.022	0.029	0.036	0.029

De lo anterior se puede notar que la variable “phd” y “mar1” no son significativas en el sentido bayesiano.

Comparando los resultados de ambos modelos se puede notar que la media posterior estimada de cada variable es muy parecida, sin embargo, la desviación estándar con el modelo de Binomial Negativa es más grande respecto al modelo Poisson. Se prosigue a

calcular el DIC para ambos modelos y se obtiene que el DIC para el modelo cero inflado Poisson es 3255.584, mientras que el DIC para el modelo cero inflado Binomial Negativa es 3137.209, por lo que, es mejor elegir el modelo cero inflado Binomial Negativa.

Con lo anterior, el análisis general del ejemplo es: las mujeres científicas publican menos artículos que los científicos hombres, el tener niños pequeños afecta en publicar menos artículos, mientras que el tener un mentor que publique una gran cantidad de artículos impacta positivamente en el número de artículos que publica el científico. Finalmente, como se eligió el modelo cero inflado Binomial Negativa entonces las variables “estado civil” y “prestigio del programa de posgrado” no son significativas para el modelo.

Una conclusión importante es que para los datos de conteo sobredispersos, los errores estándar estimados de los parámetros a menudo son muy pequeños en los modelos de Poisson, pues ya se vio que la distribución Poisson asume que su varianza es igual a su media. Por lo anterior, usar la distribución Poisson puede ser muy restrictiva, por lo que la distribución binomial negativa suele ser una mejor opción en la práctica.

Se puede ver una comparación más detallada sobre el modelo cero inflado Poisson y cero inflado Binomial negativa en R. N. Amalia et al. (2021).

Capítulo 6

Aplicación

6.1. Datos sobre la actividad física vigorosa

La base de datos con la que se trabajará fue tomada de la Encuesta Nacional de Salud (ENSANUT) 2018. Los datos corresponden a un cuestionario realizado a 44,446 adolescentes y adultos entre los 15 y 69 años, donde se les pregunta acerca del número de días y horas, en los últimos 7 días, que realizaron actividad física vigorosa, moderada o caminar. Además de datos más generales como son su estado de residencia, edad, sexo, etc.

Para el presente trabajo se tiene intención de localizar variables que tengan exceso de ceros, en particular, se observa que el número horas, en los últimos 7 días, en los que un individuo realizó actividad física vigorosa tiene una gran cantidad de ceros, esto es, existe una gran cantidad de personas que no realiza actividad física vigorosa a la semana.

6.2. Análisis descriptivo de los datos

A continuación se muestra el número de horas que los individuos encuestados realizan actividad física durante una semana.

En la Figura [6.1](#) se puede observar que existe una gran proporción de ceros, traducido a que gran parte de los encuestados no realiza actividad física vigorosa a la semana. Lo que indica que esta variable respuesta es una candidata para un modelo cero-inflado.

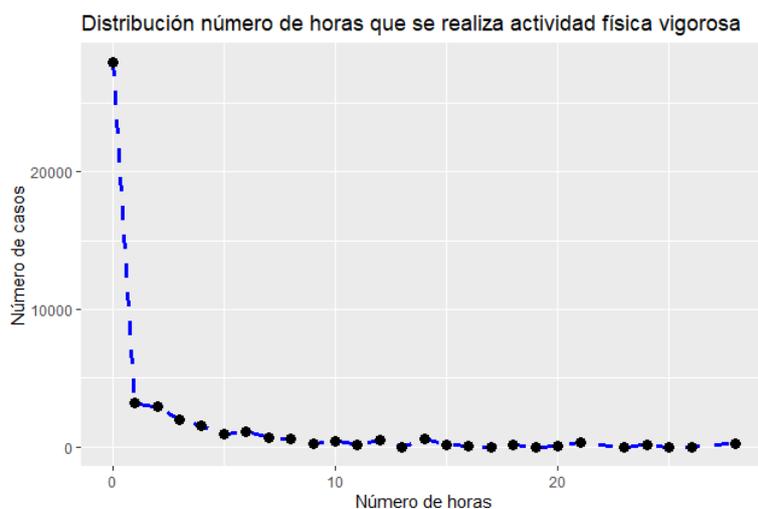


Figura 6.1: Distribución de horas que se realiza actividad física vigorosa.

En la Figura 6.2 podemos observar que la mayor proporción de los encuestados es para el grupo de edad 15-19 años. Entre los grupos de edad de 20 y 44 años se mantiene una cantidad similar de encuestados y para casos mayores a 45 va disminuyendo hasta llegar al grupo de edad con menos encuestados, el de 65-69 años.

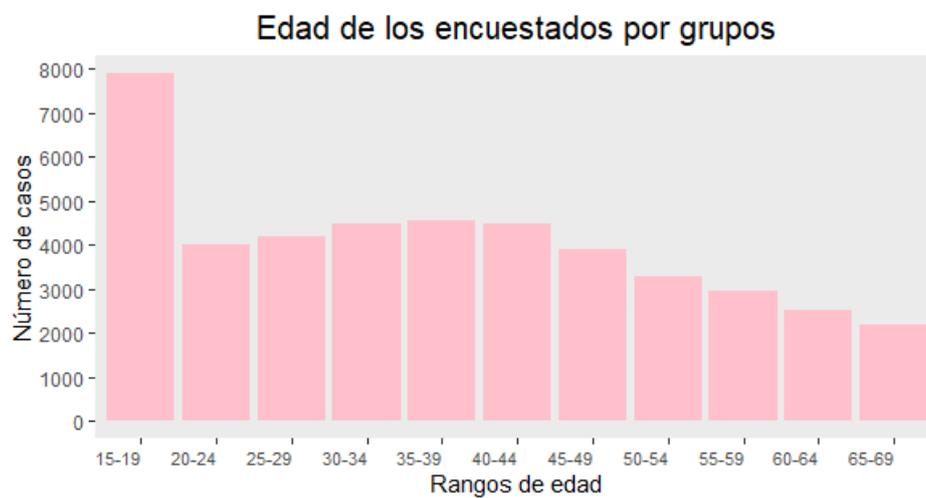


Figura 6.2: Encuestados por grupos de edad.

En la Figura 6.3 se puede observar que hay una proporción, casi del doble, mayor de encuestados en el grupo de 15-19 años respecto de los demás. Además, a partir del rango de 35-39 los encuestados van disminuyendo.

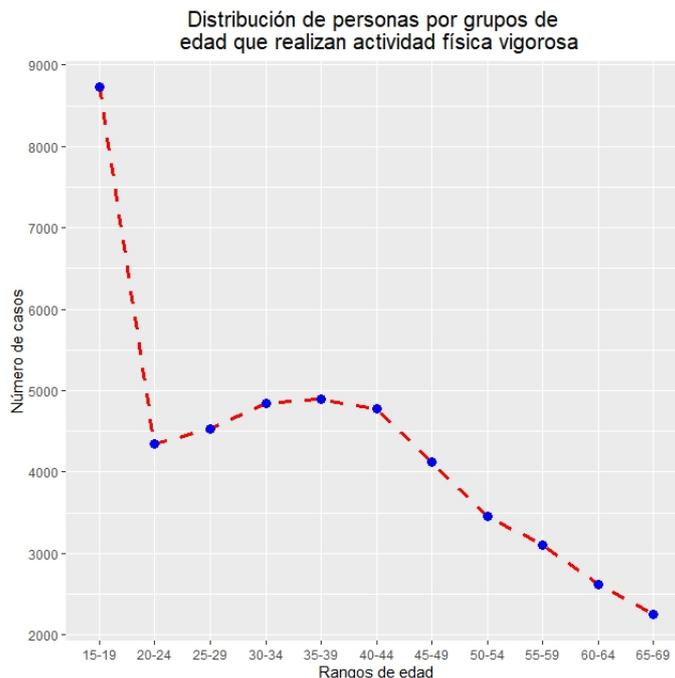


Figura 6.3: Distribución por grupos de edad que realizan actividad física vigorosa.

En la gráfica [6.4](#) se presenta el dominio al que pertenecen los encuestados, esto es, si pertenecen a una zona rural o urbana. Se puede observar que la mayor proporción de los encuestados corresponde a residentes de una zona urbana. Un poco más de la cuarta parte corresponde a residentes de una zona rural.

Dominio de los encuestados

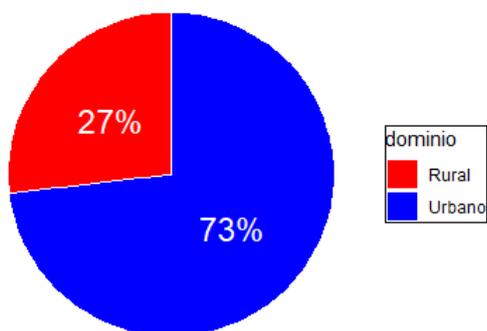


Figura 6.4: Proporción por dominio al que pertenecen los encuestados.

En la gráfica [6.5](#) se presenta el sexo que tienen los encuestados. Se puede observar

que la mayor proporción de los encuestados son mujeres.

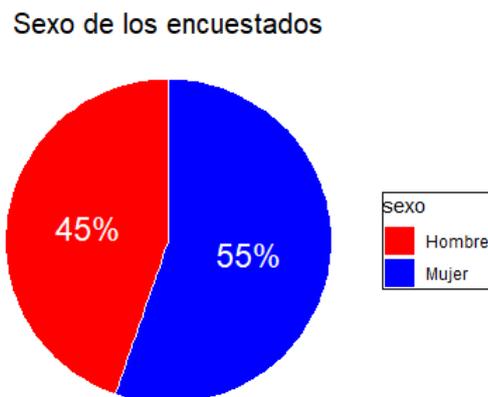


Figura 6.5: Proporción por sexo al que pertenecen los encuestados.

En la gráfica 6.6 se presentan los encuestados que realizan actividad física separada por sexo y grupos de edad. Podemos notar que los grupos de edad 15-19, 20-24, 50-54 y 60-64 son los que presentan mayor diferencia de encuestados por sexo. Siendo los tres últimos donde se concentra la mayor cantidad de mujeres encuestadas.

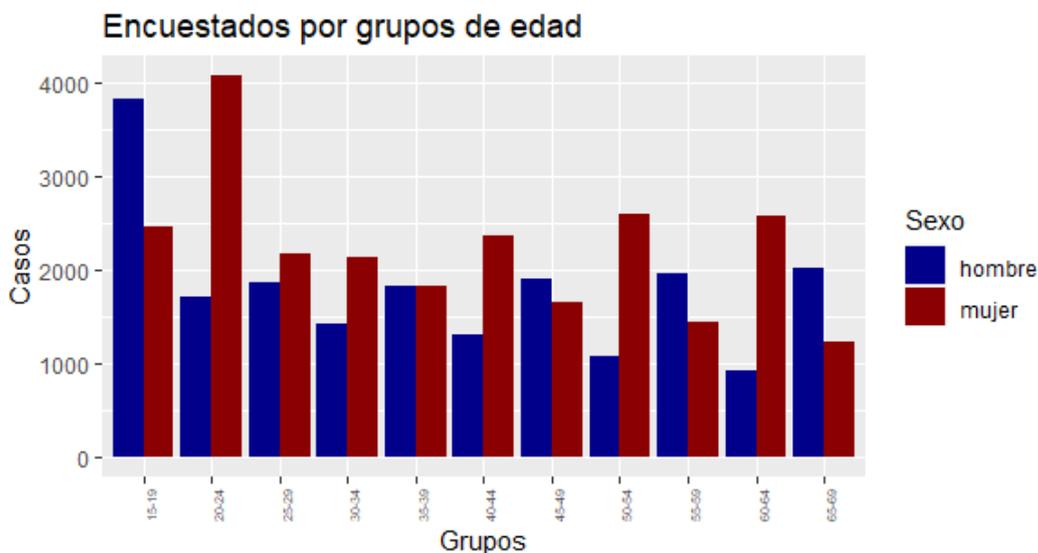


Figura 6.6: Encuestados por sexo y grupos de edad que realizan actividad física.

En la gráfica 6.7 se presentan los encuestados que realizan actividad física separado por sexo y grupos de edad. Se puede observar que para los grupos de edad 20-24, 40-

44, 50-54 y 60-64 las mujeres realizan más actividad física que los hombres. Además, podemos notar que para el grupo más extremo, 15-19, los hombres hacen actividad física en mayor proporción que las mujeres.

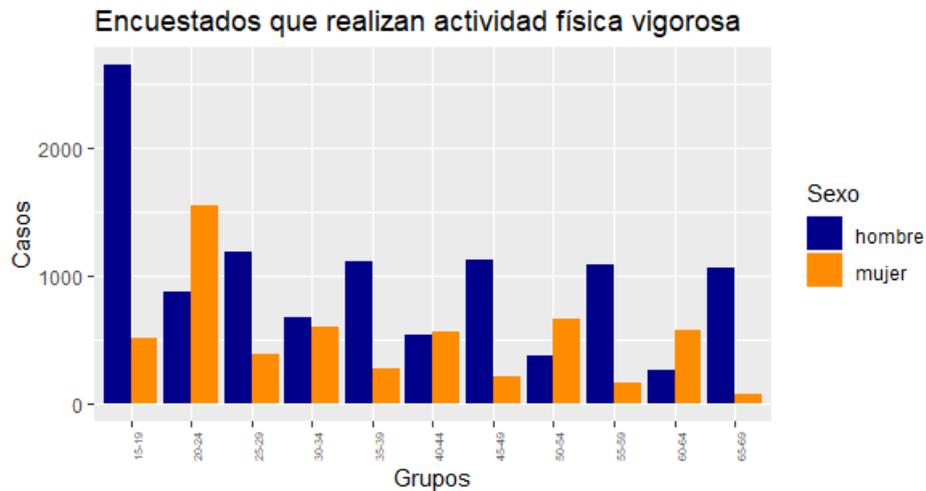


Figura 6.7: Encuestados por sexo y edad que realizan actividad física vigorosa.

En la gráfica [6.8](#) podemos observar que en todos los estados los hombres encuestados realizan más actividad física vigorosa. Se puede observar que la menor diferencia se da en la Ciudad de México, sólo 58 casos de diferencia. Por otro lado, la mayor diferencia se da en San Luis Potosí, con 184 casos.

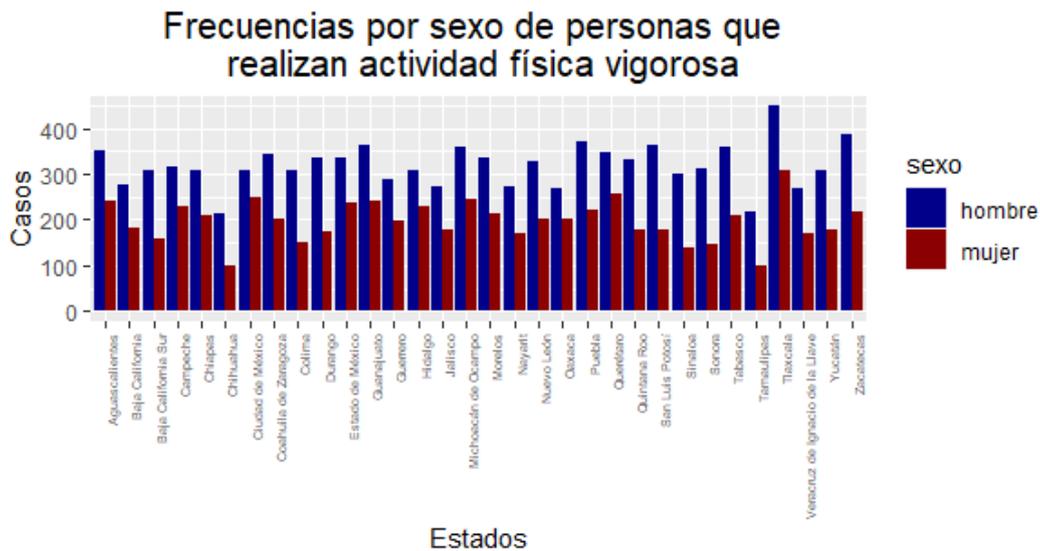


Figura 6.8: Frecuencias por sexo y estado que realizan actividad física vigorosa.

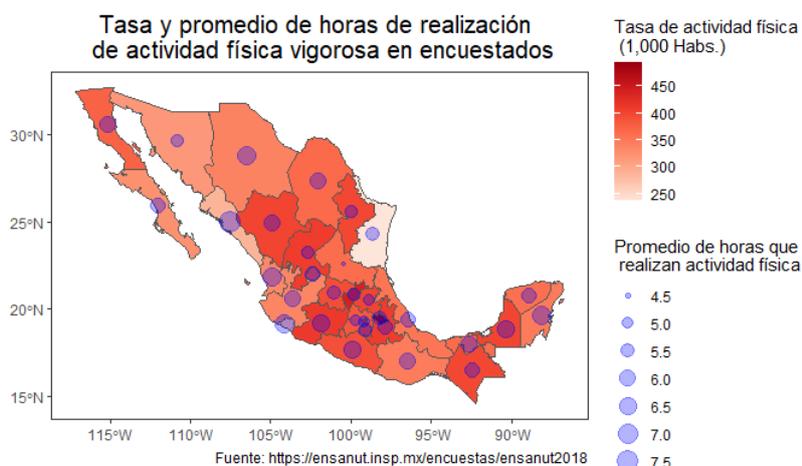


Figura 6.9: Mapa de tasa de actividad física vigorosa en encuestados y promedio de horas en que realizan actividad física vigorosa por estados

En la figura [6.9](#) se puede observar que los estados donde hay una mayor tasa de realización de actividad física vigorosa son: Tlaxcala, Querétaro y Ciudad de México todos por arriba de 430 por cada 1000 habitantes. Por otro lado, los estados donde hay una menor tasa son: Tamaulipas, Chihuahua y Sinaloa, con 237, 275 y 287 por cada 1000 habitantes respectivamente. Además, los estados donde los encuestados tienen un promedio de horas de actividad física mayor son: Sinaloa, Nayarit, Chihuahua y Colima, todos por arriba de 6.5 horas por semana. La mayoría de los estados se encuentran en un rango de 5 y 6 horas por semana, el estado que menos horas presenta es San Luis Potosí, con 4.5 horas por semana.

6.3. Resultados

A continuación se ajustan los modelos cero inflado Poisson y cero inflado binomial negativa para los datos de actividad física vigorosa. La variable respuesta es el número de horas realizadas de actividad física y las covariables son la edad, sexo, entidad y dominio al que pertenecen los individuos. Donde las variables sexo, entidad, dominio y edad son variables categóricas. La variable edad se construyó en 11 grupos como se muestra a continuación:

Años	Grupo
15-19	1
20-24	2
25-29	3
30-34	4
35-39	5
40-44	6
45-49	7
50-54	8
55-59	9
60-64	10
65-69	11

Adicional, las entidades quedaron clasificadas como sigue:

Número	Entidad	Número	Entidad
1	Aguascalientes	17	Morelos
2	Baja California	18	Nayarit
3	Baja California Sur	19	Nuevo León
4	Campeche	20	Oaxaca
5	Chiapas	21	Puebla
6	Chihuahua	22	Querétaro
7	Coahuila	23	Quintana Roo
8	Colima	24	San Luis Potosí
9	Ciudad de México	25	Sinaloa
10	Durango	26	Sonora
11	Guanajuato	27	Tabasco
12	Guerrero	28	Tamaulipas
13	Hidalgo	29	Tlaxcala
14	Jalisco	30	Veracruz
15	Estado de México	31	Yucatán
16	Michoacán	32	Zacatecas

Entonces, al aplicar el modelo cero inflado Poisson se obtienen los siguientes resultados:

Modelo cero inflado poisson

	mean	sd	0.025quant	0.5quant	0.975quant	mode	kld
1 (Intercept)	1.7692	0.0185	1.7326	1.7693	1.8054	1.7694	0
2 EDADF2	0.1043	0.0118	0.0810	0.1043	0.1275	0.1043	0
3 EDADF3	0.0991	0.0122	0.0751	0.0991	0.1229	0.0991	0
4 EDADF4	0.0915	0.0122	0.0676	0.0915	0.1154	0.0916	0
5 EDADF5	0.1008	0.0123	0.0766	0.1008	0.1250	0.1009	0
6 EDADF6	0.1298	0.0127	0.1049	0.1298	0.1546	0.1299	0
7 EDADF7	0.2043	0.0134	0.1779	0.2043	0.2305	0.2043	0
8 EDADF8	0.1859	0.0152	0.1560	0.1860	0.2157	0.1860	0
9 EDADF9	0.2446	0.0164	0.2123	0.2447	0.2768	0.2448	0
10 EDADF10	0.2603	0.0188	0.2232	0.2603	0.2970	0.2605	0
11 EDADF11	0.2612	0.0230	0.2157	0.2613	0.3060	0.2615	0
12 SEXOF2	-0.4442	0.0075	-0.4589	-0.4442	-0.4296	-0.4442	0
13 ENTF2	0.0625	0.0259	0.0116	0.0625	0.1133	0.0625	0
14 ENTF3	-0.0080	0.0262	-0.0595	-0.0080	0.0432	-0.0080	0
15 ENTF4	0.1086	0.0244	0.0607	0.1086	0.1565	0.1086	0
16 ENTF5	-0.0425	0.0253	-0.0922	-0.0425	0.0072	-0.0425	0
17 ENTF6	0.1178	0.0254	0.0679	0.1178	0.1676	0.1178	0
18 ENTF7	-0.0153	0.0255	-0.0653	-0.0153	0.0346	-0.0153	0
19 ENTF8	0.0987	0.0283	0.0430	0.0987	0.1541	0.0988	0
20 ENTF9	-0.0898	0.0258	-0.1406	-0.0898	-0.0392	-0.0898	0
21 ENTF10	-0.0111	0.0253	-0.0608	-0.0111	0.0386	-0.0110	0
22 ENTF11	-0.0561	0.0248	-0.1049	-0.0561	-0.0075	-0.0561	0
23 ENTF12	0.0647	0.0253	0.0150	0.0647	0.1144	0.0647	0
24 ENTF13	-0.1729	0.0264	-0.2248	-0.1729	-0.1211	-0.1728	0
25 ENTF14	0.0715	0.0260	0.0204	0.0715	0.1225	0.0716	0
26 ENTF15	-0.1091	0.0257	-0.1595	-0.1091	-0.0588	-0.1091	0
27 ENTF16	0.0784	0.0239	0.0315	0.0784	0.1254	0.0784	0
28 ENTF17	-0.0372	0.0253	-0.0869	-0.0372	0.0125	-0.0372	0
29 ENTF18	0.1641	0.0253	0.1145	0.1641	0.2136	0.1641	0
30 ENTF19	-0.0875	0.0261	-0.1387	-0.0875	-0.0364	-0.0875	0
31 ENTF20	0.0058	0.0260	-0.0454	0.0058	0.0568	0.0058	0
32 ENTF21	-0.0281	0.0247	-0.0765	-0.0281	0.0204	-0.0281	0
33 ENTF22	-0.0952	0.0251	-0.1444	-0.0952	-0.0461	-0.0952	0
34 ENTF23	0.1354	0.0246	0.0871	0.1354	0.1837	0.1354	0
35 ENTF24	-0.3202	0.0275	-0.3742	-0.3202	-0.2664	-0.3201	0
36 ENTF25	0.2103	0.0248	0.1616	0.2103	0.2589	0.2103	0
37 ENTF26	-0.1255	0.0272	-0.1789	-0.1255	-0.0723	-0.1254	0
38 ENTF27	-0.0264	0.0247	-0.0750	-0.0264	0.0221	-0.0264	0
39 ENTF28	-0.1002	0.0305	-0.1603	-0.1001	-0.0405	-0.0999	0

41	ENTF29	-0.0383	0.0234	-0.0843	-0.0383	0.0077	-0.0383	0
42	ENTF30	-0.0415	0.0269	-0.0944	-0.0415	0.0112	-0.0414	0
43	ENTF31	-0.0259	0.0261	-0.0772	-0.0259	0.0252	-0.0259	0
44	ENTF32	-0.1291	0.0250	-0.1782	-0.1291	-0.0800	-0.1291	0
45	DOMINIOF2	0.1507	0.0077	0.1356	0.1507	0.1658	0.1508	0

Se puede observar que no todas las variables son significativas en el sentido bayesiano, esto es, en un intervalo del 95% se considera el 0 como posible valor de los coeficientes estimados. Las variables no significativas para el modelo poisson son: ENTF3, ENTF5, ENTF7, ENTF10, ENTF17, ENTF20, ENTF21, ENTF27, ENTF29, ENTF30 y ENTF31.

CAPÍTULO 6. APLICACIÓN

Por otro lado, al aplicar el modelo cero inflado Binomial Negativa se obtienen los siguientes resultados: **Modelo cero inflado binomial negativa**

	mean	sd	0.025quant	0.5quant	0.975quant	mode	kld
1 (Intercept)	1.7461	0.0642	1.6211	1.7457	1.8732	1.7448	0
2 EDADF2	-0.0771	0.0421	-0.1595	-0.0772	0.0056	-0.0773	0
3 EDADF3	-0.2106	0.0420	-0.2928	-0.2106	-0.1280	-0.2107	0
4 EDADF4	-0.1898	0.0414	-0.2709	-0.1899	-0.1084	-0.1900	0
5 EDADF5	-0.2758	0.0415	-0.3570	-0.2759	-0.1943	-0.2760	0
6 EDADF6	-0.3066	0.0423	-0.3895	-0.3067	-0.2233	-0.3069	0
7 EDADF7	-0.3680	0.0447	-0.4554	-0.3681	-0.2799	-0.3684	0
8 EDADF8	-0.5106	0.0481	-0.6046	-0.5108	-0.4158	-0.5112	0
9 EDADF9	-0.6059	0.0506	-0.7047	-0.6062	-0.5060	-0.6066	0
10 EDADF10	-0.7663	0.0544	-0.8723	-0.7666	-0.6589	-0.7672	0
11 EDADF11	-1.1634	0.0590	-1.2782	-1.1637	-1.0470	-1.1643	0
12 SEXOF2	-1.2085	0.0231	-1.2535	-1.2086	-1.1629	-1.2088	0
13 ENTF2	0.0462	0.0853	-0.1210	0.0461	0.2138	0.0458	0
14 ENTF3	-0.2770	0.0824	-0.4387	-0.2770	-0.1154	-0.2771	0
15 ENTF4	0.0693	0.0825	-0.0925	0.0693	0.2313	0.0692	0
16 ENTF5	-0.0888	0.0827	-0.2510	-0.0889	0.0735	-0.0890	0
17 ENTF6	-0.1744	0.0832	-0.3377	-0.1745	-0.0110	-0.1745	0
18 ENTF7	-0.0678	0.0827	-0.2301	-0.0679	0.0945	-0.0679	0
19 ENTF8	-0.3237	0.0883	-0.4967	-0.3239	-0.1501	-0.3242	0
20 ENTF9	0.1076	0.0842	-0.0574	0.1075	0.2730	0.1073	0
21 ENTF10	-0.0572	0.0845	-0.2228	-0.0573	0.1088	-0.0575	0
22 ENTF11	-0.1262	0.0805	-0.2842	-0.1262	0.0317	-0.1261	0
23 ENTF12	0.0407	0.0852	-0.1262	0.0406	0.2082	0.0404	0
24 ENTF13	-0.1707	0.0838	-0.3350	-0.1707	-0.0061	-0.1709	0
25 ENTF14	-0.1189	0.0838	-0.2832	-0.1190	0.0456	-0.1191	0
26 ENTF15	-0.1344	0.0813	-0.2940	-0.1344	0.0251	-0.1344	0
27 ENTF16	0.0625	0.0809	-0.0963	0.0625	0.2212	0.0624	0
28 ENTF17	-0.0832	0.0829	-0.2458	-0.0832	0.0796	-0.0833	0
29 ENTF18	-0.0655	0.0837	-0.2297	-0.0656	0.0989	-0.0657	0
30 ENTF19	-0.1537	0.0836	-0.3176	-0.1538	0.0104	-0.1539	0
31 ENTF20	-0.1439	0.0838	-0.3083	-0.1440	0.0205	-0.1441	0
32 ENTF21	-0.0788	0.0814	-0.2386	-0.0788	0.0810	-0.0788	0
33 ENTF22	-0.0398	0.0826	-0.2019	-0.0399	0.1225	-0.0400	0
34 ENTF23	-0.0876	0.0814	-0.2475	-0.0876	0.0722	-0.0876	0
35 ENTF24	-0.5033	0.0821	-0.6645	-0.5033	-0.3423	-0.5033	0
36 ENTF25	-0.1857	0.0811	-0.3450	-0.1857	-0.0266	-0.1856	0
37 ENTF26	-0.3860	0.0824	-0.5477	-0.3860	-0.2243	-0.3860	0
38 ENTF27	-0.0976	0.0804	-0.2556	-0.0976	0.0601	-0.0975	0

40	ENTF28	-0.6775	0.0857	-0.8457	-0.6776	-0.5092	-0.6777	0
41	ENTF29	0.1181	0.0799	-0.0387	0.1181	0.2747	0.1180	0
42	ENTF30	-0.1879	0.0855	-0.3556	-0.1880	-0.0199	-0.1882	0
43	ENTF31	-0.1897	0.0834	-0.3533	-0.1897	-0.0259	-0.1898	0
44	ENTF32	-0.1545	0.0811	-0.3138	-0.1545	0.0047	-0.1545	0
45	DOMINIOF2	0.0371	0.0255	-0.0129	0.0370	0.0872	0.0370	0

De manera análoga, se puede observar que las variables EDAD2, ENTF2, ENTF4, ENTF5, ENTF7, ENTF9, ENTF10, ENTF11, ENTF12, ENTF14, ENTF15, ENTF16, ENTF17, ENTF18, ENTF19, ENTF20, ENTF21, ENTF22, ENTF23, ENTF27, ENTF29, ENTF32 y DOMINIOF2 no son significativas en el sentido bayesiano, pues en un intervalo del 95 % se considera el 0 como posible valor del coeficiente estimado.

Calculando el DIC para ambos modelos se obtiene que, para el modelo Poisson, el $DIC = 185,019.6$, para el modelo binomial negativa se obtiene un $DIC = 143,442.3$, por lo que, siguiendo el criterio para elegir los datos, tenemos que el modelo binomial negativa es una mejor opción. Ahora, calculando el CPO para el modelo Poisson y binomial negativa se obtiene un $CPO = 18032.45$ y un $CPO = 18606.28$ respectivamente, lo que nos indica que el modelo binomial negativa es una mejor opción. Como se mencionó en el subcapítulo [4.4.2](#), la media y la varianza son iguales en el modelo Poisson, mientras que en el modelo binomial negativa no, lo que ocasiona que éste último resulte una mejor opción a utilizar en la práctica.

Capítulo 7

Conclusiones

A lo largo de este trabajo se han expuesto el método de INLA y se ha analizado las ventajas de utilizarlo, donde se resalta el hecho de una estimación más precisa y el tener un rendimiento computacional más óptima. Se presentan los métodos cero inflado Binomial Negativa y Poisson, como un método de estimación para datos donde existe una gran cantidad de ceros.

Además, podemos notar la importancia de aplicar los modelos cero inflados en los datos donde existe una gran cantidad de ceros, ya que se le da un tratamiento más eficiente a este tipo de datos, pues se propone un modelo mixto de dos componentes donde, por un lado, se tiene un modelo perteneciente a la familia exponencial y, por el otro, una función de masa concentrada en cero.

Como se pudo notar en el ejemplo [5.3](#) el modelo cero inflado binomial negativa resulta más efectivo que el modelo cero inflado Poisson debido a que el modelo cero inflado Binomial Negativa considera la sobredispersión en los datos, hecho que, en una gran cantidad de datos, resulta muy eficiente.

Para el ejemplo aplicado, la encuesta ENSANUT 2018, se concluye que el modelo más eficiente a utilizar es el modelo cero inflado Binomial Negativa de acuerdo al criterio DIC y CPO. Por otro lado, considerando el modelo Binomial Negativa como el mejor modelo, se tienen las siguiente conclusiones: pertenecer al rango de edad 15-19 años está asociado a una mayor cantidad de horas de realización de actividad física vigorosa, el ser hombre está asociado positivamente a una mayor cantidad de horas de realización de actividad física vigorosa y, finalmente, pertenecer a los estados Baja California, Campeche, Ciudad de México, Guerrero, Michoacán y Tlaxcala están aso-

ciados positivamente a una mayor cantidad de horas de actividad de actividad física vigorosa, siendo Tlaxcala la mayor de ellas.

Finalmente, como conclusión general se tiene que la utilización de INLA es una propuesta alternativa muy importante a considerar debido a sus grandes beneficios computacionales y de precisión. Adicional, los modelos cero inflados, tanto Poisson o Binomial Negativa, resultan muy eficientes al momento de proponer modelos a datos con exceso de ceros. Particularmente el modelo Binomial Negativa será más eficiente siempre que se tengan datos con una gran sobredispersión.

Bibliografía

1. AGRESTI, A. (2002). *Categorical data analysis*. John Wiley & Sons, New York.
2. ALBERT, J. (2009). *Bayesian Computation with R*. Springer, New York.
3. BLANGIARDO, M. & M. CAMELETTI. (2015). *Spatial and Spatio-temporal Bayesian Models with R-INLA*. John Wiley & Sons, United Kingdom.
4. BROCKWELL, P. & R. DAVIS. (2016). *Introduction to Time Series and Forecasting*. Springer.
5. CAMERON, C. & P. TRIVEDI. (2013). *Regression Analysis of Count Data*. Cambridge University Press.
6. GAMERMAN, D. & H. F. LOPES. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall/CRC, Florida.
7. GILKS, W.R. & D. J. SPIEGELHALTER. (1996). *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.
8. GOMEZ-RUBIO, V. (2020). *Bayesian Inference with INLA*. Chapman and Hall/CRC, Florida.
9. GUTIÉRREZ-PEÑA, E. (1997). *Métodos Computacionales de la Inferencia Bayesiana*. Monografías, México: IIMAS, UNAM.
10. HILBE, JOSEPH. (2011). *Negative Binomial Regression*. University Press, Cambridge.
11. LIANG, F, & C. LIU & R. CARROLL. (2010). *Advanced Markov Chain Monte Carlo Methods: learning from past samples*. John Wiley & Sons, United Kingdom.
12. LONG, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks: Sage Publications.

BIBLIOGRAFÍA

13. ORGANIZACIÓN MUNDIAL DE LA SALUD. (2022). *Actividad física*.
URL: <https://www.who.int/es/news-room/fact-sheets/detail/physical-activity>
14. RINCÓN, L. (2012). *Introducción a los procesos estocásticos*. Facultad de Ciencias, UNAM. México.
15. R. N. AMALIA ET AL. (2021). *A Study of ZIP and ZINB Regression Modeling for Count Data with Excess Zeros*. Journal of Physics: Conference Series.
16. ROBERT, C.P. & G. CASELLA. (2010). *Introducing Monte Carlo Methods with R*. Springer.
17. RUE, H. & L. HELD (2005). *Gaussian Markov Random Fields: Theory and Applications*. Chapman & Hall, London.
18. RUE, H. & S. MARTINO (2007). *Approximate Bayesian inference for hierarchical Gaussian Markov random field models*. Journal of Statistical Planning and Inference 137(10), 3177–3192.
19. RUE, H., S. MARTINO & N. CHOPIN. (2009). *Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion)*. Journal of the Royal Statistical Society: Series B (Methodological) 71(2), 319–392.
20. RUE, H., A. RIEBLER, S. H. SORBYE, J. B. SIMPSON & F. LINDGREN. (2017). *Bayesian computing with INLA: a review*. Annual Review of Statistics and Its Application.
21. WANG, X, Y. RYAN & J. FARAWAY. (2018). *Bayesian regression modelling with INLA*. Chapman & Hall/CRC.

Apéndice A

Código de R

Código del ejemplo 4

```
1 #Ejemplo Metropolis-Hastings
2 #Distribucion gamma
3
4 fgamma<-function (n, a, b)
5 {
6   mu <- a/b
7   sig <- sqrt(a/(b * b))
8   vec <- vector("numeric", n)
9   x <- a/b
10  vec[1] <- x
11  for (i in 2:n) {
12    can <- rnorm(1, mu, sig)
13    aprob <- min(1,
14                (dgamma(can, a, b)/dgamma(x, a, b))/
15                (dnorm(can, mu, sig)/dnorm(x, mu, sig)))
16    u <- runif(1)
17    if (u < aprob)
18      x <- can
19    vec[i] <- x
20  }
21  vec
22 }
23
24 vec<-fgamma(10000,2.3,2.7)
25 c(mean(vec),var(vec))
```

Código del ejemplo 5

```
1 set.seed(44566)
2 mu <- 10
```

BIBLIOGRAFÍA

```
3 sigma2 <- 1.5
4 n <- 100
5 y <- rnorm(n=n, mean=mu, sd=sqrt(sigma2))
6
7 mu.init <- 1
8 sigma2.init <- 5
9 mu.0 <- 3
10 sigma2.0 <- 10
11 a <- 2.5
12 b <- 0.5
13
14 m <- 5000
15 mu.sim <- c()
16 sigma2.sim <- c()
17 #Valores iniciales
18 mu.sim[1] <- mu.init
19 sigma2.sim[1] <- sigma2.init
20 #Algoritmo
21 for(i in 2:m){
22   mu.sim[i] <- rnorm(1,
23                     mean = (sum(y)/sigma2.sim[i-1]
24                             + mu.0/sigma2.0) /
25                             (n/sigma2.sim[i-1] + 1/sigma2.0),
26                     sd = sqrt(1/(n/sigma2.sim[i-1]
27                               + 1/sigma2.0)))
28   sigma2.sim[i] <- 1/rgamma(1, shape=n/2+a,
29                            scale = 1 /
30                            (sum((y-mu.sim[i-1])^2)/2 + b))
31 }
32
33 t0 <- 1000
34 mu.sim <- mu.sim[-c(1:t0)]
35 sigma2.sim <- sigma2.sim[-c(1:t0)]
36
37 mean(mu.sim)
38 mean(sigma2.sim)
```

Código del ejemplo 6

```
1 #INLA paso a paso
2 # Datos
3 y <- c(1.2697,7.7637,2.2532,3.4557,4.1776,6.4320,-3.6623,7.7567,
4       5.9032,7.2671,-2.3447,8.0160,3.5013,2.8495,0.6467,3.2371,
5       5.8573,-3.3749,4.1507,4.3092,11.7327,2.6174,9.4942,-2.7639,
```

```
6     -1.5859,3.6986,2.4544,-0.3294,0.2329,5.2846)
7 n <- length(y)
8 ybar <- mean(y)
9
10 #Definimos los par metros de la distribuci n a priori
11 mu0 <- -3
12 sigma2_0 <- 4
13 a <- 1.6
14 b <- 0.4
15
16 #Seleccionamos los puntos de la cuadr cula H para el hiperparametro
    psi
17 H <- 25
18 psi.min <- 0.001
19 psi.max <- 0.3
20 psi.grid <- seq(psi.min,psi.max,length.out=H)
21 hprior <- dgamma(psi.grid,shape=a,rate=b)
22
23 # Calculamos cantidades de p(psi|y)
24 theta.n <- sigma2.n <- lik <- num <- den <- prior <- c()
25 for (h in 1:H) {
26     theta.n[h] <- (psi.grid[h]*n*ybar + mu0/sigma2_0) /
27         (psi.grid[h]*n + 1/sigma2_0)
28     sigma2.n[h] <- 1 / (n*psi.grid[h] + 1/sigma2_0)
29     prior[h] <- dnorm(theta.n[h], mu0, sd=sqrt(sigma2_0))
30     lik[h] <- prod(dnorm(y, theta.n[h], sd=1/sqrt(psi.grid[h])))
31     num[h] <- hprior[h] * prior[h] * lik[h]
32     den[h] <- dnorm(theta.n[h], theta.n[h], sd=sqrt(sigma2.n[h]))
33 }
34
35
36 # Posterior marginal no normalizada de psi
37 post.psi <- num/den
38 # Se normaliza la densidad
39 f.psi <- approxfun(psi.grid, post.psi,
40                   yleft=min(psi.grid), yright=max(psi.grid))
41 const <- integrate(f.psi, min(psi.grid), max(psi.grid))
42 post.psi <- post.psi/const$value
43
44 # Seleccionamos los puntos de la cuadr icula J para el hiperparametro
    theta
45 J <- 50
46 min.theta <- -8
```

BIBLIOGRAFÍA

```
47 max.theta <- 5
48 theta.grid <- seq(min.theta,max.theta,length.out=J)
49 # Distribuciones condicionales completas theta | psi,y
50 full.cond.theta <- matrix(NA,J,H)
51 for (j in 1:J) {
52   for (h in 1:H) {
53     full.cond.theta[j,h] <- dnorm(theta.grid[j], theta.n[h],
54                                   sd=sqrt(sigma2.n[h]))
55   }
56 }
57
58 # Posterior conjunta ponderada para psi y theta
59 Delta <- 1/sum(post.psi)
60 joint.post.theta.psi <- matrix(NA,J,H)
61 for (h in 1:H) {
62   joint.post.theta.psi[,h] <- full.cond.theta[,h] * post.psi[h]*Delta
63 }
64
65 # Integramos sobre psi para obtener la marginal posterior de theta
66 marg.post.theta <- rowSums(joint.post.theta.psi)
67 # Se normaliza la densidad
68 f.theta <- approxfun(theta.grid,marg.post.theta,
69                      yleft=min(theta.grid), yright=max(theta.grid)
70                      )
71 const <- integrate(f.theta,min(theta.grid),max(theta.grid))
72 marg.post.theta <- marg.post.theta/const$value
73
74 plot(post.psi)
75 plot(marg.post.theta)
```

Código del ejemplo 7

```
1 ### Modelo con INLA
2 set.seed(125)
3 x <- runif(100)
4
5 y <- 1+2*x+rnorm(n=100,sd=0.1)
6
7 formula = y~1+x
8 mod = inla(formula,
9             data=data.frame(x=x,y=y),
10            family = "gaussian")
11 summary(mod)
12
13
14 ### Modelo con MCMC
15
16 samplesize <- 10 # tamaño
17 b_length <- sort(rnorm(samplesize)) # variable explicativa
18
19 int_true <- 1 # intercepto real
20 slope_true <- 2 # pendiente real
21 mu <- int_true + slope_true * b_length # medias reales de la dist.
22 sigma <- 0.1 # sd real de la dist.
23
24 b_mass <- rnorm(samplesize,
25                mean = mu, sd = sigma) # variable respuesta
26
27 snakes1 <- data.frame(b_length = b_length, b_mass = b_mass)
28 head(snakes1)
29
30 jagsdata_s1 <- with(snakes1,
31                    list(b_mass = b_mass, b_length = b_length,
32                         N = length(b_mass)))
33
34 lm1_jags <- function(){
35   # Verosimilitud:
36   for (i in 1:N){
37     b_mass[i] ~ dnorm(mu[i], tau) # tau es la precision (1/var)
38     mu[i] <- beta_0 + beta_1 * b_length[i]
39   }
40   # A prioris:
41   beta_0 ~ dnorm(0, 0.01) # intercepto
```

BIBLIOGRAFÍA

```
42 beta_1 ~ dnorm(0, 0.01) # pendiente
43 sigma ~ dunif(0, 100) # sd
44 tau <- 1 / (sigma * sigma)
45 }
46 init_values <- function(){
47   list(beta_0 = rnorm(1), beta_1 = rnorm(1), sigma = runif(1))
48 }
49
50 params <- c("beta_0", "beta_1", "sigma")
51
52 fit_lm1 <- jags(data = jagsdata_s1, inits = init_values,
53               parameters.to.save = params,
54               model.file = lm1_jags, n.chains = 1,
55               n.iter = 12000, n.burnin = 2000,
56               n.thin = 10, DIC = F)
57 fit_lm1
```

Código del ejemplo 8 y 9

```
1 data(crab)
2 str(crab)
3
4 #Poisson
5 crab.poi.inla <- inla(SATELLITES ~ COLOR + SPINE + WIDTH, data = crab,
6                       family = "poisson", control.compute =
7                       list(dic = TRUE, cpo = TRUE))
8 round(crab.poi.inla$summary.fixed, 3)
9
10 #Binomial negativa
11
12 crab.bn.inla <- inla(SATELLITES ~ COLOR + SPINE + WIDTH, data = crab,
13                    family = "nbinomial", control.compute =
14                    list(dic = TRUE, cpo = TRUE))
15 round(crab.bn.inla$summary.fixed, 3)
16
17 # El mejor modelo con DIC
18
19 c(crab.poi.inla$dic$dic, crab.bn.inla$dic$dic)
```

Código del ejemplo 10

```
1
2 #Tabla de articulos publicados
3 table(articles$art)
4
5 #Proporcion de articulos publicados
6 round(prop.table(table(articles$art)),3)
7
8 # Modelo Poisson
9 articles.inla1 <- inla(art ~ fem + mar + kid5 + phd + ment, data =
10                       articles, family = "zeroinflatedpoisson1",
11                       control.compute =
12                           list(dic = TRUE, cpo = TRUE))
13 round(articles.inla1$summary.fixed, 3)
14 round(articles.inla1$summary.hyperpar, 4)
15
16 #Modelo Binomial N
17 articles.inla2 <- inla(art ~ fem + mar + kid5 + phd + ment, data =
18                       articles, family = "zeroinflatednbinomial1",
19                       control.compute =
20                           list(dic = TRUE, cpo = TRUE))
21 round(articles.inla2$summary.fixed, 3)
22 round(articles.inla2$summary.hyperpar, 4)
23
24 #Comparacion de modelos con DIC
25 c(articles.inla1$dic$dic, articles.inla2$dic$dic)
26
27 #sobredispersión de BN
28 overdisp_post <- inla.tmarginal(fun = function(x) 1/x, marg =
29                               articles.inla2$marginals.hyperpar
30                               [[1]])
31 round(inla.emarginal(fun=function(x) x, marg=overdisp_post), 4)
32
33 round(inla.qmarginal(c(0.025, 0.975), overdisp_post), 4)
```

Código del modelo cero inflado Poisson y Binomial Negativa

```
1
2 datosact <- read.csv("C:/actvigorosa.csv")
3 ENTF <- datosact[,"ENT"]
4 SEXOF <- datosact[,"SEXO"]
5 DOMINIOF <- datosact[,"DOMINIO"]
6 EDADF <- datosact[, "EDADGRUPOS"]
7
8 datosact$ENTF <- as.factor(datosact$ENT)
9 datosact$SEXOF <- as.factor(datosact$SEXO)
10 datosact$DOMINIOF <- as.factor(datosact$DOMINIO)
11 datosact$EDADF <- as.factor(datosact$EDADGRUPOS)
12
13 str(datosact)
14 head(datosact)
15
16 modelo.inla1 <- inla( horasint ~ EDADF + SEXOF + ENTF + DOMINIOF, data
17 =
18     datosact, family = "zeroinflatedpoisson1",
19     control.compute =
20     list(dic = TRUE, cpo = TRUE))
21 round(modelo.inla1$summary.fixed, 4)
22 round(modelo.inla1$summary.hyperpar, 4)
23
24 #Modelo Binomial Negativa
25 modelo.inla2 <- inla(horasint ~ EDADF + SEXOF + ENTF + DOMINIOF, data
26 =
27     datosact, family = "zeroinflatednbinomial1",
28     control.compute =
29     list(dic = TRUE, cpo = TRUE))
30 round(modelo.inla2$summary.fixed, 4)
31 round(modelo.inla2$summary.hyperpar, 4)
32
33 #Comparacion de modelos con DIC
34 c(modelo.inla1$dic$dic, modelo.inla2$dic$dic)
35
36 #Comparacion de modelos con CPO
37 c(sum(modelo.inla1$cpo$cpo),sum(modelo.inla2$cpo$cpo))
```