

**PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ**  
**ESCUELA DE POSGRADO**



**MODELOS DE REGRESIÓN CON MIXTURA DE ESCALA  
GAUSSIANA BAJO REGULARIZACIÓN BAYESIANA**

**TESIS PARA OPTAR POR EL GRADO ACADÉMICO DE MAESTRA  
EN ESTADÍSTICA**

**QUE PRESENTA:**

**Alejandrina Margarita Urbano Burgos**

**ASESOR:**

**Luis Hilmar Valdivieso Serrano**

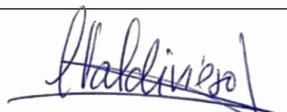
LIMA, 2024

## Informe de Similitud

Yo Luis Hilmar Valdivieso Serrano docente de la Escuela de Posgrado de la Pontificia Universidad Católica del Perú, asesor de la tesis titulada "MODELOS DE REGRESIÓN CON MIXTURA DE ESCALA GAUSSIANA BAJO REGULARIZACIÓN BAYESIANA", de la autora Urbano Burgos, Alejandrina Margarita dejo constancia de lo siguiente:

- El mencionado documento tiene un índice de puntuación de similitud de 5 %. Así lo consigna el reporte de similitud emitido por el software Turnitin el 17/07/2024.
- He revisado con detalle dicho reporte y confirmo que cada una de las coincidencias detectadas no constituyen plagio alguno.
- Las citas a otros autores y sus respectivas referencias cumplen con las pautas académicas.

Lima, 18 de Julio de 2024

Apellidos y nombres del asesor: Valdivieso Serrano, Luis Hilmar	
DNI: 07958730 ORCID: <a href="https://orcid.org/0000-0002-8975-7557">https://orcid.org/0000-0002-8975-7557</a>	Firma: 

# Dedicatoria

Dedico este trabajo a mi esposo y a mis hijos, por todo el apoyo que me brindaron durante el transcurso de mis estudios, por su generoso desprendimiento de tiempo, paciencia y comprensión que hicieron posible este trabajo en equipo. Una especial dedicatoria a mis padres, ya que esta tesis es fruto también de su amor, apoyo y sacrificio en toda mi formación académica.



# Agradecimientos

Quiero agradecer en primer lugar a Dios, quien termina todo lo que empieza. Nunca me ha dejado, me ha acompañado en cada paso de mis estudios y me regala el término de este trabajo de tesis a su tiempo.

También quiero agradecer a mi asesor Dr. Luis Valdivieso por toda su paciencia, por siempre estar dispuesto a apoyarme y haberme guiado en este trabajo hasta su culminación.

Quiero ofrecer un agradecimiento especial a Richard y Augusta, mis jefes, quienes me han apoyado desde el inicio de mis estudios en la maestría y han sido un factor clave para lograr el desarrollo de este trabajo y el término de mis estudios de posgrado. Todos mis estudios en la maestría se lo debo a ellos, quienes en todo momento me han brindado su apoyo incondicional. Un agradecimiento adicional es para Susy, quien con su apoyo y buena disposición hizo posible realizar una de las aplicaciones del presente trabajo.

El camino no ha sido fácil, pero Dios puso en mi camino a todas estas personas que lo hicieron posible.

# Resumen

La presente tesis busca estudiar las propiedades, estimación y aplicación a dos conjuntos de datos reales de diversas técnicas de regularización bayesiana sobre un modelo de regresión lineal múltiple con mixtura de escala Gaussiana, modelo que incluye al de una regresión logística. Estas técnicas de regresión penalizada bayesiana plantean distribuciones a priori que realizan la penalización, introduciendo el concepto de esparcidad, el cual se refiere al hecho de que solo un reducido número de variables tengan valores distintos de cero en sus coeficientes de regresión; es decir, es una especie de truncamiento de coeficientes llevados a cero que produce a su vez modelos más manejables e interpretables. De particular interés en este trabajo, fue la comparación de las técnicas de regularización bajo penalización y las derivadas de introducir las prioris de Horseshoe y de Horseshoe + a los coeficientes de regresión del modelo. Mostrando en la presente tesis, de manera explícita, cómo realizar un muestreo de Gibbs para la estimación de estos modelos, detallando no solo las distribuciones condicionales completas necesarias; sino también cómo es posible, mediante el uso del paquete bayesreg de R, optimizar algunas de estas propuestas de muestreo.

**Palabras-clave:** Regresión lineal múltiple, regresión logística, regresión penalizada, regresión penalizada bayesiana, priori, esparcidad, Horseshoe, Horseshoe +, distribución de mixtura de escala Gaussiana.

# Abstract

This thesis aims to study the properties, estimation and application to two real data sets of various Bayesian regularization techniques on a multiple linear regression model with Gaussian scale mixture, a model that includes a logistic regression. These Bayesian penalized regression techniques pose a priori distributions that perform the penalty, introducing the concept of sparsity, which refers to the fact that only a small number of variables have non-zero values in their regression coefficients; that is, it is a kind of truncation of coefficients taken to zero that in turn produces more manageable and interpretable models. Of particular interest in this work was the comparison of the penalty regularization techniques and those derived from introducing the Horseshoe and Horseshoe + priors to the regression coefficients of the model. In this thesis, we show explicitly how to perform Gibbs sampling for the estimation of these models, detailing not only the complete conditional distributions necessary, but also how it is possible, through the use of the bayesreg package of R, to optimize some of these sampling proposals.

**Keywords:** Multiple linear regression, logistic regression, penalized regression, Bayesian penalized regression, priori, sparsity, Horseshoe, Horseshoe +, Gaussian scale mixture distribution.

# Índice general

Dedicatoria	III
Agradecimientos	IV
Resumen	v
Abstract	VI
<b>1. Introducción</b>	<b>1</b>
1.1. Consideraciones preliminares . . . . .	1
1.2. Objetivos . . . . .	3
1.3. Organización del trabajo . . . . .	3
<b>2. Regresión con mixtura de escala Gaussiana y regularización</b>	<b>4</b>
2.1. Modelos de regresión lineal con MEG . . . . .	4
2.1.1. El modelo de regresión lineal Gaussiano . . . . .	5
2.1.2. El modelo de regresión lineal de Laplace . . . . .	6
2.1.3. El modelo de regresión lineal $t$ de Student . . . . .	8
2.1.4. Modelo de regresión logística binaria . . . . .	10
2.2. Regularización clásica en modelos de regresión lineal múltiple . . . . .	11
<b>3. Regularización bayesiana</b>	<b>14</b>
3.1. El modelo de regresión lineal normal bayesiano . . . . .	14
3.2. Regresión lineal múltiple con regularización bayesiana . . . . .	15
3.2.1. El modelo de regresión Lasso bayesiano . . . . .	15
3.2.2. El modelo de regresión Ridge Bayesiano . . . . .	17
3.2.3. El modelo de regresión Elastic Net Bayesiano . . . . .	18
3.3. El estimador de Horseshoe . . . . .	19
3.4. El estimador de Horseshoe + . . . . .	23

3.5.	Muestreo de Gibbs para el modelo de regresión con MEG . . . . .	25
3.5.1.	Muestreo del parámetro de intercepto $\beta_0$ . . . . .	26
3.5.2.	Muestreo del vector de parámetros de regresión $\beta$ . . . . .	27
3.5.3.	Muestreo de $\sigma^2$ . . . . .	28
3.5.4.	Muestreo de las variables latentes $\omega_i^2$ . . . . .	29
3.6.	Regresión logística binaria . . . . .	30
3.7.	Estimación bayesiana de los parámetros de contracción . . . . .	31
3.7.1.	Regresión ridge . . . . .	35
3.7.2.	Regresión lasso . . . . .	36
3.7.3.	Regresión Horseshoe . . . . .	37
3.7.4.	Regresión Horseshoe + . . . . .	38
<b>4.</b>	<b>Implementación de la estimación bayesiana en el modelo de estudio</b>	<b>40</b>
4.1.	Muestreo de los coeficientes de regresión $\beta$ . . . . .	40
4.1.1.	Algoritmo 1: mGrad-1 . . . . .	42
4.1.2.	Algoritmo 2: mGrad-2 . . . . .	42
4.2.	Muestreo del intercepto $\beta_0$ . . . . .	43
4.3.	Muestreo de $\sigma^2$ . . . . .	43
4.4.	Muestreo de los parámetros que inducen esparcidad . . . . .	43
4.4.1.	Muestreador Gamma Inversa-Gamma Inversa . . . . .	44
4.4.2.	Muestreo de rechazo . . . . .	44
<b>5.</b>	<b>Simulación</b>	<b>47</b>
5.1.	Método para estimar los parámetros . . . . .	47
5.2.	Criterios para evaluar la precisión de la simulación . . . . .	48
5.3.	Regresión lineal múltiple - covariables no correlacionados . . . . .	48
5.3.1.	Estimación por mínimos cuadrados . . . . .	49
5.3.2.	Estimación usando una priori de Horseshoe, Horseshoe + y Lasso bayesiano . . . . .	49
5.4.	Regresión lineal múltiple - Covariables correlacionados . . . . .	51
5.4.1.	Estimación por mínimos cuadrados . . . . .	51
5.4.2.	Estimación usando una priori de Horseshoe, Horseshoe + y Lasso Bayesiano . . . . .	52
5.5.	Regresión lineal con errores de Laplace . . . . .	53

5.5.1. Estimación usando una priori de Horseshoe, Horseshoe + y Lasso Bayesiano . . . . .	54
<b>6. Aplicaciones</b>	<b>57</b>
6.1. Aplicación a los datos BostonHousing . . . . .	57
6.1.1. Descripción de los datos . . . . .	57
6.1.2. Análisis descriptivo . . . . .	58
6.1.3. Especificación del modelo . . . . .	61
6.1.4. Métodos de aplicación . . . . .	62
6.1.5. Resultados . . . . .	62
6.2. Aplicación a un estudio de deserción universitaria . . . . .	66
6.2.1. Justificación de la aplicación . . . . .	66
6.2.2. Descripción de datos . . . . .	66
6.3. Análisis descriptivo . . . . .	70
6.3.1. Especificación del modelo . . . . .	72
6.3.2. Métodos de aplicación . . . . .	72
6.3.3. Resultados . . . . .	72
<b>7. Conclusiones</b>	<b>86</b>
7.1. Conclusiones . . . . .	86
7.2. Sugerencias para investigaciones futuras . . . . .	88
<b>A. Código R de las aplicaciones</b>	<b>90</b>
A.1. Aplicación con datos "BostonHousing" . . . . .	90
A.2. Aplicación a un estudio de deserción universitaria . . . . .	95
<b>B. Resultados adicionales de la aplicación a BostonHousing</b>	<b>109</b>
B.1. Análisis descriptivo . . . . .	109
<b>C. Resultados adicionales de la aplicación a la deserción</b>	<b>111</b>
C.1. Análisis descriptivo . . . . .	111
C.2. Resultados de regresión con todas las variables . . . . .	121
C.3. Resultados de regresión con las variables significativas . . . . .	133
C.4. Distribuciones a posteriori . . . . .	141
<b>Bibliografía</b>	<b>148</b>

# Capítulo 1

## Introducción

### 1.1. Consideraciones preliminares

Los modelos de regresión permiten investigar cómo ciertas covariables influyen en una población o muestra en estudio, y en los casos donde hay un gran número de ellas, el análisis es complejo.

Recientemente se han venido aplicando en diversas investigaciones la denominada, regresión penalizada, la cual ayuda a prevenir el sobreajuste, sobretodo cuando se tiene un gran número de predictores. Esta técnica logra seleccionar covariables que son relevantes para el modelo y se puede aplicar tanto a regresiones en datos de alta dimensión, donde el número de variables predictoras  $p$  es mayor que el tamaño de la muestra  $n$ ; como en datos donde el número de predictores  $p$  es menor que el tamaño de la muestra  $n$  (pero aún grande). La regresión penalizada puede ofrecer ventajas como evitar el sobreajuste y lograr modelos parsimoniosos en una regresión lineal múltiple, en comparación con otros métodos tradicionales de selección de variables como el stepwise (van Erp et al., 2019)

Los métodos clásicos de regresión penalizada básicamente buscan minimizar los errores de predicción a los cuales se les adiciona un término de penalización proporcional a la norma del vector de coeficientes de regresión, donde esta norma puede ser la norma  $L_1$  o norma  $L_2$  en el espacio euclídeo multidimensional. Así estos métodos nos brindan estimadores de los coeficientes de regresión relativamente pequeños debido a la penalización impuestas sobre ellos.

Otro caso especial de regresión penalizada, que se ha vuelto cada vez más popular, ver por ejemplo Bhadra et al. (2017), Griffin y Brown (2017) y Feng et al. (2017), es la penalización bayesiana. En esta se plantean distribuciones a priori que realizan la penalización, introduciendo el concepto de esparcidad, el cual se refiere al hecho de que solo un reducido

número de variables tengan valores distintos de cero en sus coeficientes de regresión; es decir, es una especie de truncamiento de coeficientes llevados a cero que produce a su vez modelos más manejables e interpretables (Perea, 2019).

En otras palabras, las prioris que se utilizan, inducen a la esparcidad llevando a cero los efectos pequeños y conservando grandes los verdaderos efectos.

En Carvalho et al. (2010), por ejemplo, propusieron un enfoque para lograr esparcidad mediante la introducción de una distribución a priori de Horseshoe, la cual está basada en la distribución half-Cauchy estándar  $C^+(0, 1)$ , cuyo rango son los reales positivos. Posteriormente Bhadra et al. (2015) introdujeron la priori de Horseshoe +.

En el presente trabajo, abordaremos la regularización bayesiana en una amplia gama de modelos lineales con mixtura de escala Gaussiana, utilizando las prioris de Horseshoe y de Horseshoe +.

Los modelos de regresión lineal con mixtura de escala Gaussiana, que en adelante llamaremos para abreviar como modelos de regresión lineal con MEG, surgen de la consideración del siguiente modelo jerárquico de regresión para un conjunto de datos  $\mathbf{y} = (y_1, \dots, y_n)^T \in \mathbb{R}^n$  de una variable respuesta y con  $p$  variables predictoras:

$$\begin{aligned} y_i \mid \mathbf{x}_i, \boldsymbol{\beta}, \beta_0, \omega_i^2, \sigma^2 &\sim \mathcal{N}(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}, \omega_i^2 \sigma^2), \quad i = 1, 2, \dots, n \\ \sigma^2 &\sim \pi_{\sigma^2}, \\ \omega_i^2 &\sim \pi_{\omega^2}, \end{aligned}$$

donde la variable respuesta en este modelo tiene una distribución condicionada normal de mixtura de escala, dado que su varianza es aleatoria, es decir, se trata de un modelo heterocedástico. Aquí  $\beta_0 \in \mathbb{R}$  es un intercepto,  $\boldsymbol{\beta} \in \mathbb{R}^p$  son los coeficientes de regresión,  $\mathbf{x}_i$  es el vector de variables predictoras,  $\omega_i^2$  son “parámetros” de mixtura latentes y  $\sigma^2$  es un “parámetro” de escala con distribución  $\pi$  sobre  $\mathbb{R}^+$ . Elecciones particulares de la distribución  $\pi_{\omega^2}$  pueden ser usadas para modelar diferentes distribuciones no Gaussianas.

En efecto, el parámetro de escala ( $\sigma^2 > 0$ ) y las variables latentes  $\boldsymbol{\omega}^2 = (\omega_1^2, \dots, \omega_n^2)^T$  son aquí usadas en un sistema Gaussiano con mixtura de escala para representar una variedad de modelos de regresión. La decisión de modelar la distribución generadora de los datos como una mixtura de escala Gaussiana con parámetros de mixtura  $\boldsymbol{\omega}^2 = (\omega_1^2, \dots, \omega_n^2)^T$  permite modelar un conjunto de modelos no-Gaussianos, cuando se proponen diferentes distribuciones no negativas para los  $\omega_i$ , en particular los modelos con errores Gaussianos (regresión lineal múltiple) y los modelos con errores de cola pesada, como los modelos con ruidos de Laplace y  $t$  Student. Estos últimos son modelados con mixturas de escala Gaussianas con densidades de

mixtura exponencial e inversa gamma, respectivamente. Adicionalmente este modelo puede extenderse a una regresión logística mediante una transformación de la variable respuesta y la consideración de una densidad de mixtura Pólya-Gamma para los  $\omega_i^2$  (Polson et al., 2013).

## 1.2. Objetivos

El objetivo general de la tesis es estudiar las propiedades, estimación y aplicación a conjuntos de datos reales de los modelos de regresión lineal con mixtura de escala Gaussiana (MEG) bajo regularización bayesiana y las prioris de Horseshoe y de Horseshoe +.

De manera específica:

- Revisar la literatura acerca de las diferentes propuestas de regularización bayesiana en los modelos de regresión.
- Estudiar la familia de los modelos de regresión con mixtura de escala Gaussiana.
- Proponer, estudiar propiedades, e implementar la estimación de los modelos de regresión con mixtura de escala Gaussiana usando los estimadores de Horseshoe y de Horseshoe +.
- Implementar métodos de inferencia bayesiana considerando métodos MCMC.
- Realizar estudios de simulación para la inferencia del modelo.
- Aplicar el modelo a conjunto de datos reales y comparar las estimaciones con otros modelos de regularización.

## 1.3. Organización del trabajo

En el capítulo 2, se presenta una introducción a los modelos de regresión lineal con MEG y a los métodos clásicos de regularización como el Lasso, Ridge y el Elastic Net. En el capítulo 3, se presentan los modelos Lasso, Ridge y Elastic Net bayesianos y la regresión con regularización bayesiana. Se detallan los modelos con priori de Horseshoe o Horseshoe + y se describe la inferencia bayesiana en los modelos de regresión con MEG. En el capítulo 4, se detalla la estimación del modelo utilizando simulación MCMC y se describen los algoritmos de estimación de los parámetros del modelo de regresión con MEG. El capítulo 5, propone un estudio de simulación. El capítulo 6, presenta dos aplicaciones a conjuntos de datos reales. Finalmente, en el capítulo 7, se resumen las principales conclusiones de este trabajo y se proponen algunas extensiones que se podrían derivar de este para investigaciones futuras.

## Capítulo 2

# Regresión con mixtura de escala Gaussiana y regularización

### 2.1. Modelos de regresión lineal con MEG

Se considerará en este trabajo el siguiente modelo de regresión para un conjunto de respuestas  $\mathbf{y} = (y_1, \dots, y_n)^T \in \mathbb{R}^n$  de una variable de interés  $y$  con  $p$  variables predictoras:

$$y_i | \mathbf{x}_i, \beta_0, \boldsymbol{\beta}, \omega_i, \sigma^2 \sim \mathcal{N}(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}, \omega_i^2 \sigma^2) \quad i = 1, 2, \dots, n \quad (2.1)$$

$$\sigma^2 \sim \pi_{\sigma^2} \quad (2.2)$$

$$\omega_i^2 \sim \pi_{\omega^2} \quad (2.3)$$

donde  $\beta_0 \in \mathbb{R}$  es un intercepto,  $\boldsymbol{\beta} \in \mathbb{R}^p$  son coeficientes de regresión,  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$  es un vector columna con  $p$  variables predictoras, las cuales consideramos, sin perder generalidad, fijas a lo largo de este trabajo, y  $\omega_i^2$  que son variables aleatorias latentes no negativas idénticamente distribuidas. La variable respuesta en este modelo tiene una distribución condicionada normal con mixtura de escala Gaussiana (MEG), pues su varianza es aleatoria, es decir se trata de un modelo heterocedástico.

La función de densidad de probabilidad marginal de la respuesta  $y$  con respecto a los  $\omega_i$  en (2.1) se puede escribir como:

$$\pi_{y_i}(y | \mathbf{x}_i, \beta_0, \boldsymbol{\beta}, \sigma^2) = \int_0^\infty N(y | \beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}, \omega^2 \sigma^2) \pi_{\omega^2}(\omega^2) d\omega^2, \quad (2.4)$$

donde  $N(\cdot | \mu, \sigma^2)$  denota la densidad de una distribución Gaussiana con media  $\mu \in \mathbb{R}$  y varianza  $\sigma^2 > 0$  y  $\pi_{\omega^2}$  es una función de densidad para los parámetros de mixtura  $\omega_i > 0$ . Elecciones particulares de la distribución  $\pi_{\omega^2}$  pueden ser usadas para modelar una gran variedad de distribuciones no Gaussianas. Por ejemplo si se selecciona para  $\pi_{\omega^2}$  la distribución inversa gamma,  $IGa(\delta/2, \delta/2)$ , se obtiene una variable respuesta con densidad  $t$  de Student de  $\delta > 0$  grados de libertad. Esta descomposición puede ser alternativamente escrita de la siguiente forma:

$$y_i | \mathbf{x}_i, \beta_0, \boldsymbol{\beta}, \sigma^2, \omega^2 \sim N(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}, \omega_i^2 \sigma^2),$$

$$\omega^2 | \delta \sim IGa(\delta/2, \delta/2).$$

Existen varias distribuciones que admiten la representación de mixtura de escala Gaussiana, por ejemplo la distribución de Laplace [Andrews y Mallows (1974)]. El parámetro de escala ( $\sigma^2 > 0$ ) y las variables latentes  $(\omega_1^2, \dots, \omega_n^2)$  son usados en un sistema con mixtura de escala Gaussiana para representar una variedad de modelos de regresión. La decisión de modelar la distribución generadora de los datos como una mixtura de escala Gaussiana (2.1) – (2.3) con parámetros de mixtura  $(\omega_1^2, \dots, \omega_n^2)$  permite modelar un conjunto de modelos no Gaussianos con errores de cola pesada, como los modelos con ruidos Laplace y  $t$  Student, los que son modelados con mixturas de escala Gaussianas con densidades de mixtura exponencial e inversa gamma, respectivamente.

En particular en esta sección se abordarán: los modelos de regresión lineal Gaussiano, Laplace,  $t$  de Student, y la regresión logística binaria, sin que ello exima a otros modelos que admitan esta representación o una similar.

### 2.1.1. El modelo de regresión lineal Gaussiano

Este caso se obtiene de fijar  $\omega_i = 1$  y asumir que  $\sigma^2$  es un parámetro fijo. Es decir,

$$y_i | \mathbf{x}_i, \beta_0, \boldsymbol{\beta} \sim \mathcal{N}(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}, \sigma^2)$$

Este modelo representa la relación entre una variable de interés  $y$  continua y un conjunto de  $p$  variables explicativas o regresoras  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ . En este modelo se supone que la función de regresión que relaciona la variable dependiente con las variables independientes es lineal. El modelo se puede escribir también como

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \sigma \epsilon_i$$

donde los  $\epsilon_i$  son errores aleatorios independientes de distribución normal estándar.

### 2.1.2. El modelo de regresión lineal de Laplace

La distribución de Laplace tiene colas más pesadas que la distribución Gaussiana y es comunmente usada para modelar datos contaminados y datos con valores atípicos.

Una variable aleatoria  $X$  tiene distribución de Laplace con parámetros  $\mu$  y  $\sigma^2$ , y se le denota por  $X \sim Laplace(\mu, \sigma^2)$ , si su función de densidad es:

$$f_X(x; \mu, \sigma) = \frac{\sqrt{2}}{2\sigma} \exp\left(-\frac{|x - \mu| \sqrt{2}}{\sigma}\right), x \in \mathbb{R}$$

Se cumple que

$$E(X) = \mu \text{ y } Var(X) = \sigma^2,$$

donde  $\mu$  es un parámetro de localización (real) y  $\sigma > 0$  es un parámetro de escala.

Esta distribución ofrece la ventaja sobre otras distribuciones de colas pesadas, tales como la  $t$  de Student, por tener todos sus momentos centrales finitos. El modelo de regresión lineal múltiple de Laplace se genera cuando  $\omega_i^2 \sim \exp(1)$  y  $\sigma^2$  es un parámetro fijo. Esta particular elección de distribución de mixtura asegura que la variable respuesta siga una distribución de Laplace. En efecto, esto es resultado de la siguiente proposición:

#### **PROPOSICIÓN 2.1**

Si  $Z | W = w \sim N(0, w)$  y  $W \sim \exp(1)$ , entonces  $Z \sim Laplace(0, 1)$ .

**Demostración:** La función de densidad de la v. a.  $Z$ , para sin pérdida de generalidad  $z > 0$ , viene dada por:

$$\begin{aligned}
f_Z(z) &= \int_0^\infty f_{z|w}(w) f_w(w) dw \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi w}} \exp\left(-\frac{z^2}{2w}\right) \exp(-w) dw \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi w}} \exp\left(-\frac{2w^2 + z^2}{2w}\right) dw \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi w}} \exp\left(-\frac{w^2 + \frac{z^2}{2}}{w}\right) dw \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi w}} \exp\left(-\frac{\left(w - \frac{z}{\sqrt{2}}\right)^2}{w} - \frac{2wz\frac{1}{\sqrt{2}}}{w}\right) dw \\
&= \exp\left(-2z\frac{1}{\sqrt{2}}\right) \int_0^\infty \frac{1}{\sqrt{2\pi w}} \exp\left(-\frac{z^2\left(w - \frac{z}{\sqrt{2}}\right)^2}{z^2 w}\right) dw \\
&= \exp\left(-2z\frac{1}{\sqrt{2}}\right) \int_0^\infty \frac{w}{\sqrt{2\pi w^3}} \exp\left(-\frac{z^2\left(w - \frac{z}{\sqrt{2}}\right)^2}{z^2 w}\right) dw
\end{aligned}$$

Si se define  $\mu = \frac{z}{\sqrt{2}}$  y  $\lambda = z^2$ , entonces podemos reescribir la última integral como

$$\frac{1}{\sqrt{\lambda}} \int_0^\infty w \frac{\sqrt{\lambda}}{\sqrt{2\pi w^3}} \exp\left(-\frac{\lambda(w - \mu)^2}{2\mu^2 w}\right) dw$$

Reconociendo que esta integral representa el cálculo del valor esperado de una variable  $G$  con distribución inversa Gaussiana de media  $\frac{z}{\sqrt{2}}$  y parámetro de escala  $\lambda = z^2$ , se tiene que:

$$\begin{aligned}
f_Z(z) &= \exp\left(-\sqrt{2}z\right) \frac{1}{z} \mathbb{E}(G) \\
&\quad \exp\left(-\sqrt{2}z\right) \frac{1}{z} \frac{z}{\sqrt{2}} \\
&\quad \frac{1}{\sqrt{2}} \exp\left(-\sqrt{2}z\right)
\end{aligned}$$

que es la función de densidad de una v. a. con distribución de Laplace. Dado que idéntico ejercicio se puede hacer con  $z < 0$  se tiene que  $Z \sim \text{Laplace}(\mu = 0, \sigma = 1)$ .

■

Por tanto, si representamos (2.1) por

$$y_i = \beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + \sigma \epsilon_i,$$

$$\epsilon_i | W = w \sim N(0, w)$$

$$\text{y } W \sim \text{exp}(1)$$

la proposición 2.1 nos dice que  $\epsilon_i \sim \text{Laplace}(0, 1)$ . Así, el modelo se podrá escribir como

$$y_i = \beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + \sigma \epsilon_i,$$

donde  $\epsilon_i \sim \text{Laplace}(0, 1)$ .

### 2.1.3. El modelo de regresión lineal $t$ de Student

La distribución  $t$  de Student es una alternativa a la distribución de Laplace en datos contaminados. La distribución  $t$  de Student tiene colas más pesadas que la distribución Gaussiana y es parametrizada por un parámetro de localización, escala y grados de libertad ( $\delta > 0$ ) que determina qué tan pesadas son las colas. Cuando este parámetro de grados de libertad es igual a 1 ( $\delta = 1$ ), la distribución  $t$  de Student se reduce a una distribución Cauchy, la cual tiene colas más pesadas que la distribución Gaussiana y de Laplace. Además, la distribución  $t$  de Student tiene colas más pesadas que la distribución de Laplace para todo ( $\delta \leq 5$ ), pero, a diferencia de la distribución de Laplace, la distribución  $t$  de Student tiene una varianza infinita cuando ( $\delta \leq 2$ ).

Una variable aleatoria  $X$  tiene distribución  $t$  de Student con  $\delta$  grados de libertad, si su función de densidad es

$$f_X(x; \delta) = \frac{\Gamma(\frac{\delta+1}{2})}{\sqrt{\delta\pi}\Gamma(\frac{\delta}{2})} \left(1 + \frac{x^2}{\delta}\right)^{-(\delta+1)/2},$$

En adelante usaremos la notación  $X \sim t(\delta)$ .

Una variable aleatoria  $X$  tiene distribución inversa gamma si su función de densidad de probabilidad está dada por:

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} (1/x)^{\alpha+1} \text{exp}(-\beta/x)$$

donde  $x > 0$ ,  $\alpha$  es un parámetro de forma y  $\beta$  un parámetro de escala. En adelante usaremos para esta variable aleatoria la notación  $X \sim \text{IGa}(\alpha, \beta)$ .

El modelo de regresión lineal múltiple  $t$  de Student, puede ser escrito como una distri-

bución Gaussiana con mixtura de escala, en donde la densidad de mixtura es la distribución inversa gamma:

$$\omega_i^2 \sim IGa\left(\frac{\delta}{2}, \frac{\delta}{2}\right)$$

y  $\sigma^2$  es un parámetro fijo. La elección de una distribución inversa gamma como distribución de las variables latentes  $\omega$  asegura que el modelo (2.1) pueda escribirse como uno con errores  $t$  de Student. Ello se basa en la siguiente proposición:

### **PROPOSICIÓN 2.2**

Si  $Z | W = w \sim N(0, w)$  y  $W \sim IGa\left(\frac{\delta}{2}, \frac{\delta}{2}\right)$ , entonces  $Z \sim t(\delta)$ .

#### **Demostración:**

La función de densidad de la v. a.  $Z$  viene dada por:

$$\begin{aligned} f_z(z) &= \int_0^\infty f_{z|w}(w) f_w(w) dw \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi w}} \exp\left(-\frac{z^2}{2w}\right) \frac{\delta/2}{\Gamma(\delta/2)} (w)^{-\delta/2-1} \exp\left(-\frac{\delta}{w}\right) dw \\ &= \frac{\delta/2^{\delta/2}}{\Gamma(\delta/2)\sqrt{2\pi}} \int_0^\infty \exp\left(-\frac{z^2 + \delta}{2w}\right) (w)^{-\delta/2-1} dw \\ &= \frac{\delta/2^{\delta/2}}{\Gamma(\delta/2)\sqrt{2\pi}} \Gamma\left(\frac{\delta+1}{2}\right) \left(\frac{z^2 + \delta}{2}\right)^{-\frac{\delta+1}{2}} \\ &= \frac{\delta/2^{\delta/2}}{\Gamma(\delta/2)\sqrt{2\pi}} \Gamma\left(\frac{\delta+1}{2}\right) \left(\frac{z^2 + \delta}{\delta}\right)^{-\frac{\delta+1}{2}} \frac{1}{2} 1^{-\frac{\delta+1}{2}} \delta^{-\frac{\delta+1}{2}} \\ &= \frac{\delta/2^{\delta/2}}{\Gamma(\delta/2)\sqrt{2\pi}} \Gamma\left(\frac{\delta+1}{2}\right) \left(1 + \frac{z^2}{\delta}\right)^{-\frac{\delta+1}{2}} \frac{1}{2} 1^{-\frac{\delta+1}{2}} \delta^{-\frac{\delta+1}{2}} \\ &= \frac{\Gamma\left(\frac{\delta+1}{2}\right)}{\Gamma(\delta/2)\sqrt{\pi}} \left(1 + \frac{z^2}{\delta}\right)^{-\frac{\delta+1}{2}} \delta^{\delta/2} \frac{1}{2} \frac{\delta/2}{2} 1^{-1/2} \frac{1}{2} 1^{-\delta/2} \frac{1}{2} 1^{-1/2} \delta^{-\delta/2} \delta^{-1/2} \\ &= \frac{\Gamma\left(\frac{\delta+1}{2}\right)}{\Gamma(\delta/2)\sqrt{\sigma\pi}} \left(1 + \frac{z^2}{\delta}\right)^{-\frac{\delta+1}{2}}, \end{aligned}$$

la cual corresponde a la densidad de una v. a. con distribución  $t$  de Student con  $\delta$  grados de libertad.

■

Por tanto, si representamos (2.1) por

$$\begin{aligned} y_i &= \beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + \sigma \epsilon_i, \\ \epsilon_i | W = w &\sim N(0, w) \\ \text{y } W &\sim IGa\left(\frac{\delta}{2}, \frac{\delta}{2}\right) \end{aligned}$$

la proposición 2.2 implica que  $\epsilon_i \sim t(\delta)$ . En consecuencia, el modelo (2.1) podrá escribirse como

$$y_i = \beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + \sigma \epsilon_i, \text{ donde } \epsilon_i \sim t(\delta).$$

Vale anotar que si bien la regresión logística no es literalmente un caso particular de (2.1), esta si puede representarse como un modelo con MEG, tal como lo veremos seguidamente.

#### 2.1.4. Modelo de regresión logística binaria

En el caso de datos binarios  $y_i \in \{0, 1\}$ , la relación entre las variables predictoras y la variable objetivo se suele representar mediante una regresión logística.

Polson et al. (2013) representan la función logística como una distribución con MEG bajo una densidad de mixtura Pólya-Gamma  $\omega_i^2 \sim PG(0, 1)$ .

Una variable aleatoria  $X$  se dice que tiene una distribución Pólya-Gamma con parámetros  $b > 0$  y  $c \in \mathcal{R}$ , denotado por  $X \sim PG(b, c)$ , si

$$X \stackrel{D}{=} \frac{1}{2\pi^2} \sum_{k=1}^{\infty} \frac{g_k}{(k - 1/2)^2 + c^2 / (4\pi^2)},$$

donde  $g_k \sim \text{Ga}(b, 1)$  son variables aleatorias con distribución gamma independientes, y  $\stackrel{D}{=}$  indica igualdad en la distribución.

El principal resultado en Polson et al. (2013) es que las probabilidades binomiales parametrizadas por los logaritmos de los ratios de odds pueden ser representadas como mixturas de escala Gaussianas con respecto a una distribución Pólya-Gamma. Su enfoque sostiene que para  $b > 0$ ,

$$\frac{\exp(\psi)^a}{(1 + \exp(\psi))^b} = 2^{-b} \exp(\kappa\psi) \int_0^{\infty} \exp(-\omega^2 \psi^2 / 2) p(\omega^2) d\omega^2, \quad (2.5)$$

donde  $\kappa = a - b/2$  y  $\omega^2 \sim PG(b, 0)$ .

Cuando  $\psi = \mathbf{x}^T \boldsymbol{\beta}$  es una función lineal de los predictores, el integrando es el núcleo de una probabilidad Gaussiana en  $\boldsymbol{\beta}$ .

Dado que para un modelo de regresión logística se cumple que

$$p(y_i|\boldsymbol{\beta}) = \frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})^{y_i}}{1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta})},$$

entonces (2.5) implica que

$$p(y_i|\boldsymbol{\beta}) = 2^{-1} \exp(\kappa_i \mathbf{x}_i^T \boldsymbol{\beta}) \int_0^\infty \exp(-\omega_i^2 (\mathbf{x}_i^T \boldsymbol{\beta})^2 / 2) p(\omega_i^2) d\omega_i^2 \quad (2.6)$$

donde  $\kappa_i = y_i - \frac{1}{2}$  y  $\omega_i^2 \sim \text{PG}(1, 0)$ .

Además Polson et al. (2013) muestran una estrategia simple para el muestreo de Gibbs de estos modelos, en las cuales la distribución Gaussiana produce los parámetros principales y la Pólya-Gamma genera una sola capa de variables latentes. El éxito de esta estrategia depende de la existencia de una forma sencilla y eficaz de simular variables aleatorias Pólya-Gamma. La representación de la suma de gammas en la distribución enunciada parece compleja y sugiere una aproximación finita. Pero Polson et al. (2013) describen un método de simulación Pólya-Gamma rápido y exacto que evita las dificultades que pueden resultar de truncar la suma infinita.

Con el fin de posteriormente comparar los métodos de regularización propuestos en este trabajo, repasaremos brevemente, para el caso de una regresión lineal, otros métodos de regularización, donde por el momento la restringiremos a una metodología clásica de estimación.

## 2.2. Regularización clásica en modelos de regresión lineal múltiple

Los métodos de regularización clásicos aplicados a los modelos de regresión tienen por objetivo la selección de variables importantes para el modelo y así evitar el sobreajuste en la predicción. Aplicar regresión penalizada ofrece beneficios en cuanto a que permite lograr modelos parsimoniosos y mejor interpretables que los que conseguimos con los métodos tradicionales de selección de variables (Derksen y Keselman, 1992) y (Tibshirani, 1996).

Estas técnicas, surgen de agregar un término de penalización a la minimización de la suma de los cuadrados de los residuos, con el propósito de reducir los coeficientes pequeños hacia cero y dejar los coeficientes grandes como tal. Ello se hace estimando los parámetros del modelo de regresión lineal múltiple mediante:

$$\underset{\beta_0, \boldsymbol{\beta}}{\text{minimizar}} \left\{ \frac{1}{2n} \left\| \mathbf{y} - \beta_0 \mathbf{1} - \tilde{\mathbf{X}} \boldsymbol{\beta} \right\|_2^2 + \lambda_c \|\boldsymbol{\beta}\|_q \right\},$$

donde  $\|\boldsymbol{\beta}\|_q = \left( \sum_{j=1}^p |\beta_j|^q \right)^{\frac{1}{q}}$  denota a la norma  $L^q$  de  $\boldsymbol{\beta}$  en el espacio euclídeo multidimensional;  $\mathbf{y} = (y_1, \dots, y_n)^T$ , es el vector de dimensión  $n$  que contiene las observaciones de la variable respuesta;  $\beta_0$  representa el intercepto;  $\mathbf{1}$  es un vector de unos de dimensión  $n$ ;  $\tilde{\mathbf{X}}$  es la matriz  $(n \times p)$  de las variables predictoras;  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$  es un vector de coeficientes de regresión de dimensión  $p$  y  $\lambda_c$  es un parámetro de penalización que mientras más grande sea, mayor será la esparcidad inducida en el modelo, es decir más variables con coeficientes de regresión serán llevadas a cero. Si  $\lambda_c = 0$  estamos frente a la solución ordinaria de mínimos cuadrados (van Erp et al., 2019).

La elección del valor de  $q$  indica el tipo de penalización inducida, de tal manera que si  $q = 1$  estamos frente a una regresión Lasso, la cual no solo contrae los coeficientes sino que realiza selección de variables induciendo esparcidad (Tibshirani, 1996). Si, por otro lado,  $q = 2$ , estamos ante una regresión Ridge, que contraresta los efectos adversos del problema de colinealidad en la estimación tradicional contrayendo ciertos coeficientes. Esta, sin embargo, no realiza selección de variables (Hoerl y Kennard, 1970).

El método Ridge incluye el término de penalización con el fin de contraer los coeficientes de regresión (reducirlos a valores cercanos a cero), cumpliéndose que a medida que aumentamos el valor de  $\lambda$  aumentaremos la penalización, y como consecuencia, se obtendrá una mayor contracción de los coeficientes. Es decir para cada valor que tome  $\lambda$  el modelo genera un vector de coeficientes distinto. El hecho de que este método no haga nulo ninguno de los coeficientes, es decir que no realice selección de variables, es una desventaja, sobretodo en contextos donde se tiene un gran número de variables predictoras (Carrasco, 2016).

Zou y Hastie (2005) propusieron otra técnica de regularización y selección de variables llamada el Elastic Net. Esta utiliza una función de penalidad que usa una combinación de las normas  $L_1$  y  $L_2$ :

$$\hat{\boldsymbol{\beta}}_{EN} = \underset{\boldsymbol{\beta}}{\text{mín}} \left\{ (\mathbf{y} - \tilde{\mathbf{X}} \boldsymbol{\beta})^T (\mathbf{y} - \tilde{\mathbf{X}} \boldsymbol{\beta}) + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2 \right\},$$

donde  $\lambda_1$  y  $\lambda_2$  son constantes no negativas que controlan el peso asignado a las restricciones o penalidades  $L_1$  y  $L_2$ .

A través de estudios de simulación y aplicaciones con data real, mostraron que el Elastic Net supera a Lasso y que es muy útil sobretodo cuando el número de predictores  $p$  es mucho mayor que el número de observaciones  $n$ , donde el método Lasso muestra un desempeño poco

## 2.2. REGULARIZACIÓN CLÁSICA EN MODELOS DE REGRESIÓN LINEAL MÚLTIPLE<sup>13</sup>

satisfactorio.

Además, Zou y Hastie, propusieron un algoritmo llamado LARS-EN para resolver la regularización Elastic Net eficientemente, cuya base es el algoritmo LARS de Efron et al. (2004), usado para la regresión Lasso.

Luego de haber revisado los modelos de regresión lineal con MEG, pasaremos a conocer cómo la inferencia bayesiana se aplica en estos modelos.



## Capítulo 3

# Regularización bayesiana

### 3.1. El modelo de regresión lineal normal bayesiano

Para el modelo de regresión lineal normal, el enfoque bayesiano, se resume de esta manera:

Dado el modelo de regresión lineal múltiple, matricialmente expresado como:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \beta_p \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \epsilon_n \end{bmatrix}$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (3.1)$$

en donde asumiremos que  $\boldsymbol{\epsilon} \sim N(0, \sigma^2 I)$  y en adelante que las covariables son fijas, la distribución de  $\mathbf{y}$ , condicionada a los parámetros  $\boldsymbol{\beta}$  y  $\sigma^2$ , es

$$\mathbf{y} \mid \boldsymbol{\beta}, \sigma^2 \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 I);$$

mientras que la verosimilitud viene dada por

$$L(\boldsymbol{\beta}, \sigma^2) = (2\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right).$$

Los valores que maximizan esta función son los estimadores de máxima verosimilitud de  $\boldsymbol{\beta}$  y

$\sigma^2$  y vienen dados por:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad \text{y} \quad \widehat{\sigma^2} = \frac{1}{n} (\mathbf{Y} - \mathbf{X} \widehat{\boldsymbol{\beta}})^T (\mathbf{Y} - \mathbf{X} \widehat{\boldsymbol{\beta}}).$$

En un enfoque bayesiano, se podrían considerar prioris no informativas independientes de estos parámetros, con la forma:

$$p(\boldsymbol{\beta}) \propto 1$$

$$p(\sigma^2) \propto \frac{1}{\sigma^2}.$$

Ella nos lleva, a la distribución a posteriori conjunta

$$p(\boldsymbol{\beta}, \sigma^2 \mid \mathbf{y}) \propto (\widehat{\sigma^2})^{-\left(\frac{n}{2}-1\right)} \exp\left(-\frac{1}{\widehat{\sigma^2}} \left( (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + (n-k)\widehat{\sigma^2} \right)\right)$$

y a la distribución marginal a posteriori de  $\boldsymbol{\beta}$

$$p(\boldsymbol{\beta} \mid \mathbf{y}) \propto \left( 1 + \frac{1}{n-k} (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})^T \frac{1}{\widehat{\sigma^2}} \mathbf{X}^T \mathbf{X} (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + (n-k)\widehat{\sigma^2} \right)^{-\frac{n}{2}}.$$

Naturalmente, también se podrían considerar otras prioris informativas, conjungas o no, para  $\boldsymbol{\beta}$  y  $\sigma^2$ .

## 3.2. Regresión lineal múltiple con regularización bayesiana

### 3.2.1. El modelo de regresión Lasso bayesiano

El modelo de regresión Lasso bayesiano, es un modelo de regresión lineal enmarcado en el análisis estadístico de inferencia bayesiana, al incluir una distribución a priori para los parámetros del modelo.

Dado el modelo de regresión lineal múltiple  $\mathbf{y} = \beta_0 \mathbf{1}_n + \tilde{\mathbf{X}} \boldsymbol{\beta} + \boldsymbol{\epsilon}$ , donde  $\mathbf{y}$  es un vector  $n \times 1$  que contiene la variable respuesta para cada individuo,  $\beta_0$  es la media general,  $\tilde{\mathbf{X}}$  es la matriz  $n \times p$  de regresores estandarizados y  $\boldsymbol{\epsilon}$  es el vector  $n \times 1$  de errores aleatorios independientes e idénticamente distribuidos normalmente con media cero y varianza constante  $\sigma^2$ , el método Lasso estima los parámetros de regresión  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$  buscando minimizar la suma de cuadrados de los errores y penalizando con la norma  $L_1$  el vector de coeficientes  $\boldsymbol{\beta}$ , de la siguiente manera:

$$\begin{aligned}\hat{\beta}_{\text{Lasso}} &= \underset{\beta}{\operatorname{argmin}} \left( (\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\beta)'(\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\beta) + \lambda \|\beta\|_1 \right) \\ &= \underset{\beta}{\operatorname{argmin}} \left( \sum_{i=1}^n \left( \tilde{y}_i - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right)\end{aligned}$$

donde,  $\tilde{\mathbf{y}} = \mathbf{y} - \tilde{y}\mathbf{1}_n$ ,  $\tilde{y}$  es la media aritmética de la variable respuesta y  $\lambda \geq 0$  indica qué tanto influye la penalización en la estimación (Allasia et al., 2016).

En este modelo podemos notar que si  $\lambda = 0$ , el estimador  $\hat{\beta}_{\text{Lasso}}$  se convierte en el estimador de mínimos cuadrados y que si  $\lambda$  toma un valor suficientemente grande  $\hat{\beta}_{\text{Lasso}}$  tiende a 0. Esto último trae consigo que sea de relativa importancia la selección de  $\lambda$ , y que se tenga un vector  $\hat{\beta}_{\text{Lasso}} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)^T$  distinto para cada valor de  $\lambda$  escogido. La selección de  $\lambda$  puede realizarse usándose métodos de validación cruzada (Efron et al., 2004).

Bajo el enfoque bayesiano el estimador  $\hat{\beta}_{\text{Lasso}}$  puede entenderse como la moda de la distribución a posteriori de  $\beta$ , es decir,  $\hat{\beta}_{\text{Lasso}} = \operatorname{argmax}_{\beta} p(\beta | \mathbf{y}, \sigma^2, \lambda)$ , cuando los parámetros de regresión tienen como priori a la siguiente distribución de Laplace con componentes independientes:

$$p(\beta | \sigma^2) = \prod_{j=1}^p \frac{\lambda}{2\sqrt{\sigma^2}} e^{-\frac{\lambda|\beta_j|}{\sqrt{\sigma^2}}} \propto e^{-\frac{\lambda}{\sqrt{\sigma^2}} \sum_{j=1}^p |\beta_j|}$$

y  $\sigma^2$  tiene como priori a una distribución no informativa  $p(\sigma^2) \propto \frac{1}{\sigma^2}$  o una gamma inversa para que se mantenga la conjugación.

Park y Casella (2008) desarrollaron el primer procedimiento explicado para la regresión Lasso bayesiana, utilizando una representación de la distribución de Laplace como una mixtura de escala Gaussianas (MEG), lo cual permite obtener la siguiente representación jerárquica del modelo completo, usando además un vector de variables latentes.

$$\begin{aligned}\mathbf{y} | \tilde{\mathbf{X}}, \beta, \sigma^2 &\sim \mathbf{N}_n(\tilde{\mathbf{X}}\beta, \sigma^2 \mathbf{I}_n) \\ \beta | \sigma^2, (\ell_1^2, \dots, \ell_p^2)^T &\sim \mathbf{N}_p(\mathbf{0}_p, \sigma^2 \operatorname{diag}(\ell_1^2, \dots, \ell_p^2)) \\ \sigma^2, (\ell_1^2, \dots, \ell_p^2)^T &\sim \pi(\sigma^2) \prod_{j=1}^p \frac{\lambda^2}{2} \exp\left(\frac{-\lambda^2 \ell_j^2}{2}\right)\end{aligned}$$

Luego de integrar sobre  $\ell_1^2, \dots, \ell_p^2$ , la priori condicional de  $\beta$  tiene la distribución condicional de Laplace deseada, donde  $\pi(\sigma^2)$  puede ser una priori impropia (no informativa) o bien gamma inversa.

Para obtener la distribución a posteriori, Casella y George (1992), proponen aplicar el muestreador de Gibbs.

Otros autores han también propuesto modelos jerárquicos similares, pero con diferentes algoritmos de simulación. Hans (2009) por ejemplo planteó un nuevo método para el muestreador de Gibbs en la regresión Lasso bayesiana, a través de una nueva caracterización de la distribución a posteriori, con el fin de mejorar las predicciones del modelo.

El análisis bayesiano resulta una mejor opción que el clásico, cuando se tiene un gran número de variables predictoras o existe una correlación lineal entre ellas. En el enfoque bayesiano, tener calculada la distribución a posteriori de los coeficientes, nos permite calcular el nivel de incertidumbre para cualquier modelo elegido, de esta manera la regresión Lasso bayesiana ofrece ventajas importantes frente a la regresión Lasso Clásica (Allasia et al., 2016).

### 3.2.2. El modelo de regresión Ridge Bayesiano

En la regresión Ridge bayesiana, el vector de parámetros  $\beta$  tiene como priori el siguiente producto de distribuciones normales independientes:

$$p(\beta) = \left( \frac{1}{2\pi\sigma_\beta^2} \right)^{\frac{p}{2}} \exp \left( -\frac{1}{2\sigma_\beta^2} \beta^T \beta \right),$$

donde  $\sigma_\beta^2$  es un hiperparámetro de varianza común de los efectos. Se tiene, por otro lado, la siguiente distribución a priori no informativa  $p(\sigma^2) \propto 1/\sigma^2$ . Luego la distribución a posteriori de  $\beta$  viene dada por

$$\begin{aligned} p(\beta | \sigma^2, \mathbf{y}) &\propto \exp \left( -\frac{1}{2\sigma_\beta^2} (\mathbf{y} - \tilde{\mathbf{X}}\beta)^T (\mathbf{y} - \tilde{\mathbf{X}}\beta) \right) \\ &\propto \exp \left( -\frac{1}{2\sigma_\beta^2} \beta^T \beta \right) p(\sigma^2). \end{aligned}$$

Luego de desarrollar y factorizar se obtiene que:

$$\begin{aligned} p(\beta | \sigma^2, \mathbf{y}) &\propto \exp \left( -\frac{1}{2\sigma_\beta^2} (\mathbf{y} - \tilde{\mathbf{X}}\beta)^T (\mathbf{y} - \tilde{\mathbf{X}}\beta) \right) \exp \left( -\frac{1}{2\sigma_\beta^2} \beta^T \beta \right) \\ &\propto \exp \left( -\frac{1}{2\sigma_\beta^2} \left[ -2\beta^T \tilde{\mathbf{X}}^T \mathbf{y} + \beta^T \left( \tilde{\mathbf{X}}^T \tilde{\mathbf{X}} + \frac{\sigma_\beta^2}{\sigma^2} I \right) \beta \right] \right) \end{aligned}$$

Se completa la forma cuadrática para obtener

$$p(\boldsymbol{\beta} \mid \sigma^2, \mathbf{y}) \propto \exp\left(-\frac{1}{2\sigma_\beta^2}(\boldsymbol{\beta} - \mathbf{a})^T \boldsymbol{\Sigma}^* (\boldsymbol{\beta} - \mathbf{a})\right), \text{ donde}$$

$$\mathbf{a} = \left(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}} + \frac{\sigma^2}{\sigma_\beta^2} \mathbf{I}\right)^{-1} \tilde{\mathbf{X}}^T \mathbf{y}$$

$$\boldsymbol{\Sigma}^* = \left(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}} + \frac{\sigma^2}{\sigma_\beta^2} \mathbf{I}\right)^{-1}$$

y por tanto  $\boldsymbol{\beta} \mid \sigma^2, \mathbf{y} \sim N(\mathbf{a}, \boldsymbol{\Sigma}^*)$ .

Con esta distribución final y una función de pérdida cuadrática, se obtiene el siguiente estimador Ridge bayesiano:

$$\hat{\boldsymbol{\beta}} = E(\boldsymbol{\beta} \mid \sigma^2, \mathbf{y}) = (\mathbf{X}^T \mathbf{X} + k\mathbf{I})^{-1} \mathbf{X}^T \mathbf{y}$$

que es el estimador de Ridge ordinario o clásico, cuando  $k = \sigma^2/\sigma_\beta^2$ .

### 3.2.3. El modelo de regresión Elastic Net Bayesiano

La técnica de regularización y selección de variables Elastic Net es un método flexible que mezcla la penalización con norma L1 y L2. Esta es generalmente usada cuando el número de predictores  $p$  es más grande que el tamaño de la muestra ( $p \gg n$ ).

En el método Elastic Net Bayesiano, se tiene como objetivo encontrar la moda de la distribución marginal a posteriori de  $\boldsymbol{\beta} \mid \mathbf{y}$  cuando se considera las siguientes priori:

$$p(\boldsymbol{\beta} \mid \sigma^2) \propto \exp\left(-\frac{1}{2\sigma^2} \left(\lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2\right)\right)$$

y  $p(\sigma^2) \propto 1/\sigma^2$  (Li y Lin, 2010).

De esta priori, podemos notar algunos casos especiales según los valores que tome  $\lambda_1$  y  $\lambda_2$ . Si  $\lambda_1 = 0$  y  $\lambda_2 > 0$ , la priori para  $\boldsymbol{\beta}$  es una distribución normal. En este caso estamos ante el modelo Ridge Bayesiano. Si  $\lambda_1 > 0$  y  $\lambda_2 = 0$  la priori para  $\boldsymbol{\beta}$  es la distribución de Laplace, en este caso estamos ante el modelo Lasso Bayesiano (López, 2012).

El modelo Elastic Net Bayesiano, según (Li y Lin, 2010), se resume jerárquicamente como:

$$\begin{aligned} \mathbf{y} \mid \boldsymbol{\beta}, \sigma^2 &\sim N\left(\tilde{\mathbf{X}}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n\right) \\ \boldsymbol{\beta} \mid \sigma^2 &\propto \exp\left(-\frac{1}{2\sigma^2}\left(\lambda_1\|\boldsymbol{\beta}\|_1 + \lambda_2\|\boldsymbol{\beta}\|_2^2\right)\right) \\ \sigma^2 &\propto \frac{1}{\sigma^2}. \end{aligned}$$

La distribución posteriori marginal para  $\boldsymbol{\beta} \mid \mathbf{y}$  es

$$p(\boldsymbol{\beta} \mid \mathbf{y}) = \int_0^\infty \frac{C(\lambda_1, \lambda_2, \sigma^2)}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\|\mathbf{y} - \tilde{\mathbf{X}}\boldsymbol{\beta}\|_2^2 + \lambda_1\|\boldsymbol{\beta}\|_1 + \lambda_2\|\boldsymbol{\beta}\|_2^2}{2\sigma^2}\right) \pi(\sigma^2) d\sigma^2,$$

donde  $C(\lambda_1, \lambda_2, \sigma^2)$  es una constante de normalización.

### 3.3. El estimador de Horseshoe

Carvalho et al. (2010) propusieron, por otro lado, un enfoque distinto a los anteriormente descritos para lograr esparcidad o contracción, al introducir ellos una distribución a priori de Horseshoe. Ellos describen las ventajas de su metodología sobre las ya existentes, enfatizando su robustez y adaptabilidad a diferentes patrones de esparcidad. Además, usando datos reales y simulados, muestran que el estimador de Horseshoe genera estimaciones similares a las obtenidas por un modelo bayesiano bajo una priori de mixtura discreta.

En el procedimiento que proponen, el objetivo es estimar un vector de coeficientes de regresión. El estimador se define como la media a posteriori de un modelo bayesiano con una priori conocida como la priori de Horseshoe. El nombre de “Horseshoe” se debe, como lo veremos más adelante, a la forma que toma esta priori para inducir esparcidad.

La distribución de Horseshoe está basada en la distribución half-Cauchy. Esta proviene de la distribución estándar de Cauchy, de la cual se ha tomado solo la parte positiva, es decir cuando su rango se restringe solo a valores positivos. La distribución half-Cauchy tiene media y varianza no definidas al tener colas pesadas. La distribución de half-Cauchy con parámetro de escala  $c$  y corrimiento  $x_0$  tiene como función de densidad a:

$$\begin{aligned} f(x; x_0, c) &= \frac{1}{\pi c \left[1 + \left(\frac{x-x_0}{c}\right)^2\right]} \\ &= \frac{1}{\pi} \left[ \frac{c}{(x-x_0)^2 + c^2} \right], \quad x > x_0 \end{aligned}$$

donde  $x_0$  indica la ubicación del pico de la distribución. Esta distribución la denotaremos por  $C^+(x_0, c)$

Cuando  $x_0 = 0$  y  $c = 1$  esta distribución se denomina la distribución half-Cauchy estándar y tiene como función de densidad a

$$f(x; 0, 1) = \frac{1}{\pi(1+x^2)}, \quad x > 0.$$

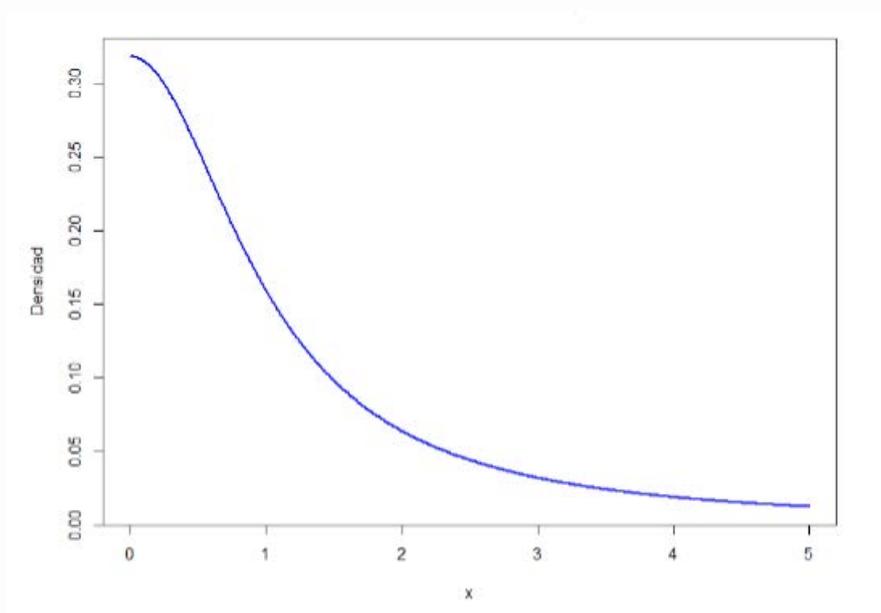


Figura 3.1: Función de densidad half-Cauchy estándar (elaboración propia)

Esta distribución es una distribución alternativa a la distribución Gamma Inversa como priori para los parámetros de algunos modelos jerárquicos bayesianos (Ghosh, 2014). También es usada en otros modelos jerárquicos como priori para la desviación estándar, mostrando ser flexible y tener un buen comportamiento para los fines del modelo (Molina y Ramírez, 2014).

El modelo de Horseshoe introducido por Carvalho et al. (2010) parte de considerar las variables aleatorias

$$Y_i \sim N(\theta_i, \sigma^2), \quad i = 1, 2, \dots, n \quad (3.2)$$

La distribución de Horseshoe se obtiene como la distribución incondicional de  $\theta_i$ , cuando

cada  $\theta_i$  tiene como priori la siguiente distribución:

$$\begin{aligned}\theta_i | \lambda_i &\sim N(0, \sigma^2 \lambda_i^2 \tau^2), \\ \lambda_i &\sim C^+(0, 1),\end{aligned}\tag{3.3}$$

siendo  $C^+(0, 1)$  la distribución half-Cauchy estándar. Aquí los  $\lambda_i$  se pueden considerarse como hiperparámetros de esparcidad local y  $\tau$  como un hiperparámetro de esparcidad global que busca llevar los  $\theta_i$  hacia el valor cero, mientras que los parámetros locales  $\lambda_i$  permiten que algunos de estos escapen a la contracción debido a la forma de la distribución half-Cauchy, la cual posee una cola pesada y concentra gran parte de su densidad en torno al valor cero. De esta manera, los parámetros  $\theta_i$  no significativos para el modelo asumen valores muy próximos a cero, y a su vez el modelo tiene la propiedad de que es robusto frente a los grandes efectos y permite que se manifiesten, es decir que algunos de los  $\theta_i$  se alejen de cero (Lozano, 2020).

En Bhadra et al. (2019) y Polson y Scott (2012), se puede ver que la introducción de los parámetros locales  $\lambda_i$ , mejora los resultados de predicción con nuevas muestras comparado con métodos anteriores que contienen solo un parámetro global.

El siguiente resultado nos provee de la media a posteriori de  $\theta_i$

### **PROPOSICIÓN 3.1**

$$E(\theta_i | y_i, \lambda_i) = (1 - \kappa_i) y_i,$$

donde  $\kappa_i = \frac{1}{1 + \lambda_i^2 \tau^2}$  es un factor de contracción correspondiente al parámetro  $\theta_i$ .

### **Demostración:**

Dado que

$$\begin{aligned}p(\theta_i | y_i, \lambda_i) &\propto p(y_i | \theta_i) p(\theta_i | \lambda_i) \\ &\propto \exp\left(-\frac{1}{2\sigma^2}(y_i - \theta_i)^2\right) \exp\left(-\frac{1}{2\sigma^2 \lambda_i^2 \tau^2} \theta_i^2\right) \\ &\propto \exp\left(-\frac{1}{2\sigma^2} \left(y_i^2 - 2\theta_i y_i + \theta_i^2 + \frac{\theta_i^2}{\lambda_i^2 \tau^2}\right)\right) \\ &\propto \exp\left(-\frac{1}{2\sigma^2} \left(\left(1 + \frac{1}{\lambda_i^2 \tau^2}\right) \theta_i^2 - 2\theta_i y_i + y_i^2\right)\right),\end{aligned}$$

se tiene completando cuadrados que

$$p(\theta_i | y_i, \lambda_i) \propto \exp\left(-\frac{1}{2\sigma^2} \left(1 + \frac{1}{\lambda_i^2 \tau^2}\right) \left(\theta_i^2 - 2\frac{\theta_i \lambda_i^2 \tau^2}{1 + \lambda_i^2 \tau^2} y_i + \frac{y_i^2 \lambda_i^2 \tau^2}{1 + \lambda_i^2 \tau^2}\right)\right)$$

Por lo que podemos concluir que:

$$\theta_i | y_i, \lambda_i \sim N \left( \frac{\lambda_i^2 \tau^2}{1 + \lambda_i^2 \tau^2} y_i, \frac{\sigma^2}{1 + \frac{1}{\lambda_i^2 \tau^2}} \right)$$

Así,

$$E(\theta_i | y_i, \lambda_i) = \frac{\lambda_i^2 \tau^2}{1 + \lambda_i^2 \tau^2} y_i$$

y fijando  $\kappa_i = \frac{1}{1 + \lambda_i^2 \tau^2}$  obtenemos que  $E(\theta_i | y_i, \lambda_i) = (1 - \kappa_i) y_i$ .

■

Puesto que a priori  $\lambda_i \sim C^+(0, 1)$ , no es difícil probar que el factor de contracción  $\kappa_i$  posee la siguiente distribución a priori:

$$p(\kappa_i) = \frac{\tau}{\pi} \left( \frac{1}{1 - (1 - \tau^2) \kappa_i} \right) \left( \frac{1}{\sqrt{\kappa_i} \sqrt{1 - \kappa_i}} \right) \quad (3.4)$$

En particular, cuando  $\tau = 1$ , esta distribución adquiere la forma de una herradura "horseshoe"; es decir, la priori half-Cauchy para los  $\lambda_i$  impone una distribución  $Beta(1/2, 1/2)$  como priori para los  $\kappa_i$ , la cual se ilustra en la Figura 3.2 y da nombre a la distribución de los  $\theta_i$

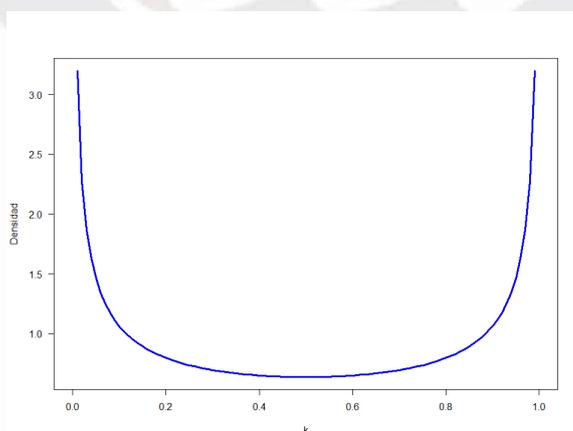


Figura 3.2: Función de densidad del factor de contracción para  $\tau = 1$

El factor de contracción  $\kappa_i$  se entiende como la fuerza con la que los  $\theta_i$  se ven atraídos hacia el cero. Como se observa en la Figura 3.2, los valores muy cercanos a 0 ó 1 alcanzan la

mayor densidad, donde  $\kappa_i = 0$  significa ausencia de contracción y  $\kappa_i = 1$  indica contracción total para su correspondiente parámetro  $\theta_i$ .

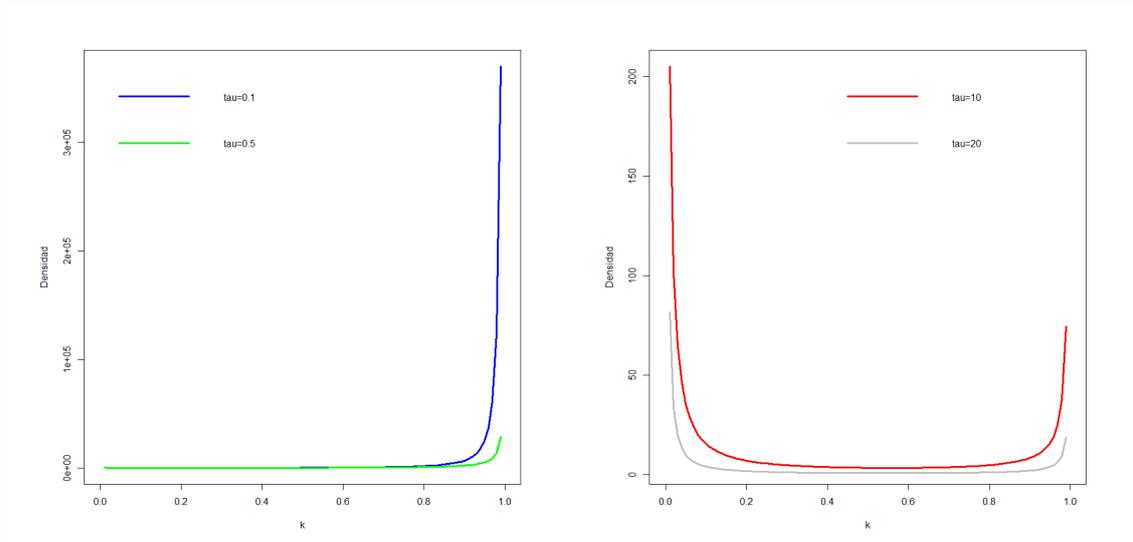


Figura 3.3: Función de densidad del factor de contracción para distintos valores de  $\tau$

Así también se puede observar en la Figura 3.3, que cuando el parámetro global  $\tau$  toma valores pequeños la densidad de  $\kappa_i$  se concentra en torno a 1, es decir, todos los parámetros  $\theta_i$  sufren una fuerte contracción; caso contrario, cuando  $\tau$  es grande, el valor de  $\kappa$  es cercano a cero, es decir la contracción en los coeficientes es débil. Por lo expuesto, la decisión previa sobre  $\tau$  influirá en  $\kappa_i$  (Lozano, 2020).

Este estimador presenta ventajas sobre los ya existentes, en cuanto a su robustez y adaptabilidad a diferentes patrones de esparcidad. Además, Carvalho et al. (2010) muestra dos teoremas: uno que detalla la robustez de la cola del estimador de Horseshoe y el otro que demuestra una tasa de convergencia para la estimación correcta de la densidad en situaciones de esparcidad.

### 3.4. El estimador de Horseshoe +

El modelo jerárquico anteriormente definido en (3.3) podría aplicarse a los coeficientes  $\beta_j$  de un modelo de regresión, escribiéndose

$$\begin{aligned}\beta_j | \lambda_j &\sim N(0, \lambda_j^2 \tau^2), \\ \lambda_j &\sim C^+(0, 1)\end{aligned}$$

Esto define al estimador de Horseshoe para  $\beta_j$ . Este estimador se puede definir alternatively para  $\tilde{\lambda}_j = \lambda_j \tau$  como

$$\begin{aligned}\beta_j | \tilde{\lambda}_j, \tau &\sim \mathcal{N}\left(0, \tilde{\lambda}_j^2\right), \\ \tilde{\lambda}_j | \tau &\sim C^+(0, \tau),\end{aligned}\tag{3.5}$$

donde  $C^+(0, \tau)$  denota a una distribución half-Cauchy.

El estimador Horseshoe + para  $\beta_j$  introducido por Bhadra et al. (2016) describe, a diferencia de (3.5), el siguiente modelo jerárquico, definido por el conjunto de distribuciones condicionales:

$$\begin{aligned}\beta_j | \tilde{\lambda}_j, \eta_j, \tau &\sim \mathcal{N}\left(0, \tilde{\lambda}_j^2\right) \\ \tilde{\lambda}_j | \eta_j, \tau &\sim C^+(0, \tau \eta_j) \\ \eta_j &\sim C^+(0, 1),\end{aligned}$$

donde se introduce otra priori half-Cauchy adicionando un nuevo parámetro  $\eta_j$  de esparcidad local. Este estimador posee propiedades distintas al estimador de Horseshoe, desde que la distribución marginal de  $\tilde{\lambda}_j | \tau$  ya no es  $C^+(0, 1)$ , sino dada por la siguiente proposición.

### **PROPOSICIÓN 3.2**

$$p(\tilde{\lambda} | \tau) = \frac{4}{\pi^2 \tau} \left( \frac{\log(\tilde{\lambda}/\tau)}{(\tilde{\lambda}/\tau)^2 - 1} \right)\tag{3.6}$$

#### **Demostración:**

Desde que

$$\begin{aligned}p(\tilde{\lambda} | \tau) &= \int_0^\infty p(\tilde{\lambda} | \tau, \eta) p(\eta) d\eta \\ &= \int_0^\infty \frac{1}{\pi} \frac{\tau \eta}{\tilde{\lambda}^2 + \tau^2 \eta^2} \frac{1}{\pi(1 + \eta^2)} d\eta = \frac{\tau}{\pi^2} \int_0^\infty \frac{\eta}{(\eta^2 + 1)(\tau^2 \eta^2 + \tilde{\lambda}^2)} d\eta,\end{aligned}$$

podemos hacer el cambio de variable  $u = \frac{\eta^2}{2}$  y aplicar fracciones parciales para obtener

$$\begin{aligned} p(\tilde{\lambda} | \tau) &= \frac{\tau}{\pi^2} \int_0^\infty \frac{1}{(2u+1)(2\tau^2u + \tilde{\lambda}^2)} du = \frac{\tau}{\pi^2} \int_0^\infty \left( \frac{\tau^2}{(\tau^2 - \tilde{\lambda}^2)(2\tau^2u + \tilde{\lambda}^2)} - \frac{1}{(\tau^2 - \tilde{\lambda}^2)(2u+1)} \right) du \\ &= \frac{\tau}{\pi^2} \left( \frac{\log(\tau^2\eta^2 + \tilde{\lambda}^2)}{2(\tau^2 - \tilde{\lambda}^2)} - \frac{\log(\eta^2 + 1)}{2(\tau^2 - \tilde{\lambda}^2)} \right) \Big|_0^\infty \\ &= \frac{1}{\pi^2\tau} \left( \frac{\log(\tilde{\lambda}/\tau)}{((\tilde{\lambda}/\tau)^2 - 1)} \right) \end{aligned}$$

■

La distribución marginal a priori que Bhadra et al. (2016) muestran para  $\beta$  es

$$p(\beta_j | \tau) = \int_0^\infty p(\beta_j | \lambda_j, \tau) p(\lambda_j | \tau) d\lambda_j,$$

en donde, si se considera  $\kappa_j$  como un parámetro de escala de esparcidad definido con  $\kappa_j = 1/(1 + \lambda_j^2\tau^2)$ , esta se podría escribir como:

$$p(\beta_j | \tau) = \int_0^1 p(\beta_j | \kappa_j, \tau) p(\kappa_j | \tau) d\kappa_j, \text{ con } p(\beta_j | \kappa_j, \tau) \sim \mathcal{N}\left(0, \frac{1 - \kappa_j}{\kappa_j}\right).$$

Si bien el parámetro de contracción global  $\tau$  lo hemos asumido hasta el momento conocido, éste podría también considerarse aleatorio. En un enfoque bayesiano completo, se puede asumir una priori haf-Cauchy estándar o una Uniforme  $(0, 1)$  para  $\tau$ . Otro enfoque sugiere apelar a un argumento asintótico que sugiere que el estimador empírico de Bayes de  $\tau$  debe fijarse como  $\hat{\tau} = p_n/n$  donde  $p_n$  es el número de entradas distintas de cero en  $\beta$  (van der Pas et al., 2014).

### 3.5. Muestreo de Gibbs para el modelo de regresión con MEG

Recodemos que el modelo de regresión lineal con mixtura de escala Gaussiana (MEG) plantea para cada  $i = 1, 2, \dots, n$  y  $j = 1, 2, \dots, p$  que:

$$y_i | \mathbf{x}_i, \beta_0, \boldsymbol{\beta}, \omega_i^2, \sigma^2 \sim \mathcal{N}(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}, \omega_i^2 \sigma^2), \quad (3.7)$$

$$\sigma^2 \sim \pi_{\sigma^2}, \quad (3.8)$$

$$\omega_i^2 \sim \pi_{\omega_i^2}, \quad (3.9)$$

$$\beta_0 \sim \pi_{\beta_0}, \quad (3.10)$$

$$\beta_j \mid \lambda_j^2, \tau^2, \sigma^2 \sim \mathcal{N}(0, \lambda_j^2 \tau^2 \sigma^2), \quad (3.11)$$

$$\lambda_j^2 \sim \pi_{\lambda_j^2}, \quad (3.12)$$

$$\tau^2 \sim \pi_{\tau^2}, \quad (3.13)$$

donde  $\beta_0 \in \mathbb{R}$  es un intercepto,  $\boldsymbol{\beta} \in \mathbb{R}^p$  son coeficientes de regresión,  $\mathbf{x}_i$  es un vector de  $p$  variables predictoras y las variables latentes  $\omega_i^2$  son no negativas. Para efectos de su estimación, este modelo jerárquico se podría segmentar en dos grupos: (i) Las ecuaciones (3.7) – (3.10) describen la distribución de la variable respuesta y (ii) las ecuaciones (3.11) – (3.13) describen las distribuciones a priori para los coeficientes de regresión. Los modelos estadísticos para ambos, los datos y las distribuciones a priori, serán construídos, como en (Andrews y Mallows, 1974), desde distribuciones Gaussianas con mixtura de escala intercambiables.

Para este modelo el cálculo analítico de las distribuciones a posteriori es intratable, por lo que se deben usar técnicas MCMC para obtener aproximaciones estocásticas de las densidades a posteriori correspondientes. Más aún si se eligen distribuciones a priori particulares para que todas las distribuciones condicionales completas correspondientes sean computables analíticamente, podríamos hacer uso del muestreador de Gibbs.

Dado que las covariables las estamos asumiendo, sin pérdida de generalidad fijas, omitiremos estas en adelante en la condicional en la ecuación (3.11).

### 3.5.1. Muestreo del parámetro de intercepto $\beta_0$

Makalic y Schmidt (2016a) usan la distribución a priori uniforme para el parámetro del intercepto (3.10), mientras que los parámetros de la regresión tienen una distribución a priori conjunta especificada por la MEG (3.11) – (3.13).

Concretamente, la distribución condicional completa para el parámetro del intercepto  $\beta_0$  es la distribución Gaussiana  $N(\bar{\mu}, \bar{\sigma}^2)$ , donde

$$\bar{\mu} = \left( \sum_{i=1}^n \frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\omega_i^2} \right) \left( \sum_{i=1}^n \frac{1}{\omega_i^2} \right)^{-1},$$

y

$$\bar{\sigma}^2 = \sigma^2 \left( \sum_{i=1}^n \frac{1}{\omega_i^2} \right)^{-1}$$

Esto se verifica realizando el siguiente desarrollo:

$$p(\beta_0 \mid \mathbf{y}, \boldsymbol{\beta}, \omega^2, \lambda) = \frac{p(\mathbf{y}, \beta_0, \boldsymbol{\beta}, \omega^2, \lambda)}{p(\mathbf{y}, \boldsymbol{\beta}, \omega^2, \lambda)}$$

$$\begin{aligned} &\propto p(\mathbf{y} \mid \beta_0, \boldsymbol{\beta}, \omega^2, \lambda) p(\boldsymbol{\beta} \mid \omega^2, \lambda) p(\lambda) \\ &\propto \prod_{i=1}^n p(y_i \mid \beta_0, \boldsymbol{\beta}, \omega_i^2, \lambda) \\ &\propto \prod_{i=1}^n \frac{1}{\sqrt{2\pi\omega_i^2\sigma^2}} \exp\left(-\frac{1}{2\omega_i^2\sigma^2}(y_i - \beta_0 - x_i^T \boldsymbol{\beta})^2\right) \\ &\propto \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n \frac{1}{\omega_i^2} (\beta_0 - (y_i - x_i^T \boldsymbol{\beta}))^2\right) \\ &\propto \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n \frac{1}{\omega_i^2} (\beta_0^2 - 2\beta_0(y_i - x_i^T \boldsymbol{\beta}) + (y_i - x_i^T \boldsymbol{\beta})^2)\right) \\ &\propto \exp\left(-\frac{1}{2\sigma^2} \left(\beta_0^2 \sum_{i=1}^n \frac{1}{\omega_i^2} - 2\beta_0 \sum_{i=1}^n \frac{y_i - x_i^T \boldsymbol{\beta}}{\omega_i^2} + \sum_{i=1}^n \omega_i^2 \left(\frac{y_i - x_i^T \boldsymbol{\beta}}{\omega_i^2}\right)^2\right)\right) \\ &\propto \exp\left(-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n \frac{1}{\omega_i^2}\right) \left(\beta_0^2 - 2\beta_0 \left(\sum_{i=1}^n \frac{1}{\omega_i^2}\right)^{-1} \left(\sum_{i=1}^n \frac{y_i - x_i^T \boldsymbol{\beta}}{\omega_i^2}\right) + \left(\frac{1}{\omega_i^2}\right)^{-2} \left(\frac{y_i - x_i^T \boldsymbol{\beta}}{\omega_i^2}\right)^2\right)\right) \\ &\propto \exp\left(-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n \frac{1}{\omega_i^2}\right) \left(\beta_0 - \left(\sum_{i=1}^n \frac{1}{\omega_i^2}\right)^{-1} \left(\sum_{i=1}^n \frac{y_i - x_i^T \boldsymbol{\beta}}{\omega_i^2}\right)\right)^2\right) \end{aligned}$$

lo cual implica que

$$\beta_0 \sim N\left(\left(\sum_{i=1}^n \frac{1}{\omega_i^2}\right)^{-1} \left(\sum_{i=1}^n \frac{y_i - x_i^T \boldsymbol{\beta}}{\omega_i^2}\right), \sigma^2 \left(\sum_{i=1}^n \frac{1}{\omega_i^2}\right)^{-1}\right).$$

A pesar de que se tiene aquí una distribución condicional completa para el intercepto  $\beta_0$ , esta no es usada en el muestreo que Schmidt y Makalic (2019) proponen en su paquete bayesreg. Ellos, en cambio, aplican un algoritmo computacionalmente más eficiente que logra una mejor mezcla de la cadena, el cual se detalla en el capítulo 4.

### 3.5.2. Muestreo del vector de parámetros de regresión $\boldsymbol{\beta}$

Por otro lado, la densidad condicional completa para el vector de parámetros de la regresión  $\boldsymbol{\beta}$  viene dada por la distribución Gaussiana p-variada  $N_p(\bar{\boldsymbol{\mu}}, \mathbf{A}_p^{-1})$ , donde

$$\bar{\boldsymbol{\mu}} = \mathbf{A}_p^{-1} \mathbf{X}^T \boldsymbol{\Omega}_n^{-1} (\mathbf{z} - \beta_0 \mathbf{1}_n), \quad \mathbf{A}_p = (\mathbf{X}^T \boldsymbol{\Omega}_n^{-1} \mathbf{X} + \boldsymbol{\Lambda}_p^{-1}),$$

$\mathbf{1}_n$  es un vector de unos,  $\mathbf{\Omega}_n = \sigma^2 \text{diag}(\omega_1^2, \dots, \omega_n^2)$  y  $\mathbf{\Lambda}_p = \sigma^2 \tau^2 \text{diag}(\lambda_1^2, \dots, \lambda_p^2)$ , siendo  $\mathbf{A}_p \in \mathbb{R}^{p \times p}$  una matriz simétrica, positiva, de precisión definida. Una derivación detallada de esta distribución condicional completa está disponible en Lindley y Smith (1972). Nosotros no profundizamos en ella, pues el mayor cuello de botella de estos algoritmos MCMC es el muestreo de los coeficientes de regresión desde la distribución Gaussiana  $p$ -variada  $N_p(\bar{\boldsymbol{\mu}}, \mathbf{A}_p^{-1})$ , cuando el número de predictores  $p$  es muy grande. Makalic y Schmidt (2016a) no recomiendan el cálculo directo de la matriz  $\mathbf{A}_p^{-1}$  porque presenta una precisión numérica deficiente y es computacionalmente costoso. En vez de esto, como se verá en el capítulo 4, Makalic y Schmidt (2016a), proponen una implementación que usa dos algoritmos para muestrear los coeficientes de regresión, donde la elección del algoritmo depende del tamaño de muestra  $n$  y del número de regresores  $p$ . Específicamente, Makalic y Schmidt (2016a) usan el algoritmo de Rue (2001), cuando el ratio ( $p/n < 2$ ), y usan el algoritmo de Anirban Bhattacharya y Mallick (2016) en caso contrario.

El algoritmo de Rue (2001) usa la factorización de Cholesky de la matriz de varianzas a posteriori condicional  $\mathbf{A}_p^{-1}$  y tiene complejidad computacional de orden  $O(p^3)$ . El algoritmo es eficiente si  $p$  no es muy grande comparado con  $n$ . En el caso donde  $p$  sea mucho más grande que  $n$ , el muestreador de Makalic y Schmidt (2016a) usa el algoritmo de Anirban Bhattacharya y Mallick (2016), la cual tiene complejidad computacional de orden  $O(n^2 p)$ , que es lineal en  $p$ . Anirban Bhattacharya y Mallick (2016) muestran que su algoritmo es órdenes de magnitud más eficiente que el algoritmo de Rue cuando ( $p/n > 2$ ). Por consiguiente, el muestreo de los coeficientes de regresión en la implementación propuesta por Makalic y Schmidt (2016a) es significativamente más rápido que las implementaciones alternativas (por ejemplo, Stan) y representa el estado actual del arte en términos de velocidad y precisión numérica.

### 3.5.3. Muestreo de $\sigma^2$

Para todos los modelos de errores considerados, el parámetro de escala  $\sigma^2$  será considerado con una distribución a priori invariante de escala  $\pi_{\sigma^2} \propto 1/\sigma^2$ . La distribución condicional completa para  $\sigma^2$  resulta ser entonces la distribución inversa gamma  $IGa(\tilde{\alpha}, \tilde{\beta})$ , donde

$$\tilde{\alpha} = \frac{n+p}{2} \quad \text{y} \quad \tilde{\beta} = \frac{1}{2} \left( \sum_{i=1}^n \frac{\epsilon_i^2}{\omega_i^2} + \sum_{j=1}^p \frac{\beta_j^2}{\tau_j^2 \lambda_j^2} \right).$$

y  $\epsilon_i = y_i - \beta_0 - \mathbf{x}_i^T \boldsymbol{\beta}$  denota a los residuales del modelo.

En efecto, esto se verifica realizando el siguiente desarrollo:

$$\begin{aligned}
p(\sigma^2 \mid \mathbf{y}, \beta_0, \boldsymbol{\beta}, \omega^2, \lambda, \tau) &= \frac{p(\mathbf{y}, \beta_0, \boldsymbol{\beta}, \omega^2, \sigma^2, \lambda, \tau)}{p(\mathbf{y}, \beta_0, \boldsymbol{\beta}, \omega^2, \lambda, \tau)} \\
&\propto p(\mathbf{y} \mid \beta_0, \boldsymbol{\beta}, \omega^2, \sigma^2, \lambda, \tau) p(\boldsymbol{\beta} \mid \sigma^2, \lambda, \tau) p(\sigma^2 \mid \lambda, \tau) p(\lambda \mid \tau) p(\tau) \\
&\propto \prod_{i=1}^n p(y_i \mid \beta_0, \boldsymbol{\beta}, \omega_i^2, \sigma^2) \prod_{j=1}^p p(\beta_j \mid \sigma^2, \lambda_j, \tau_j) \cdot \frac{1}{\sigma^2} \\
&\propto \prod_{i=1}^n \frac{1}{\sqrt{2\pi\omega_i^2\sigma^2}} \exp\left(-\frac{1}{2\omega_i^2\sigma^2}(y_i - \beta_0 - x_i^T \boldsymbol{\beta})^2\right) \prod_{j=1}^p \frac{1}{\sqrt{2\pi\lambda_j^2\tau_j^2\sigma^2}} \exp\left(-\frac{1}{2\lambda_j^2\tau_j^2\sigma^2}(\beta_j)^2\right) \cdot \frac{1}{\sigma^2} \\
&\propto \frac{1}{\sigma^n} \left(\prod_{i=1}^n \frac{1}{\omega_i}\right) \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n \frac{1}{\omega_i^2} (y_i - \beta_0 - x_i^T \boldsymbol{\beta})^2\right) \cdot \frac{1}{\sigma^p} \left(\prod_{j=1}^p \frac{1}{\lambda_j\tau_j}\right) \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^p \frac{\beta_j^2}{\lambda_j^2\tau_j^2}\right) \cdot \frac{1}{\sigma^2} \\
&\propto \frac{1}{\sigma^2} \binom{n+p+1}{\frac{n+p}{2}} \exp\left(-\frac{\frac{1}{2} \left(\sum_{i=1}^n \frac{(y_i - \beta_0 - x_i^T \boldsymbol{\beta})^2}{\omega_i^2} + \sum_{j=1}^p \frac{\beta_j^2}{\lambda_j^2\tau_j^2}\right)}{\sigma^2}\right)
\end{aligned}$$

lo cual implica que

$$\sigma^2 \sim IGa\left(\tilde{\alpha} = \frac{n+p}{2}, \tilde{\beta} = \frac{1}{2} \left(\sum_{i=1}^n \frac{\epsilon_i^2}{\omega_i^2} + \sum_{j=1}^p \frac{\beta_j^2}{\lambda_j^2\tau_j^2}\right)\right).$$

#### 3.5.4. Muestreo de las variables latentes $\omega_i^2$

Este muestreo dependerá de la familia particular de MEG utilizada. En el caso de la distribución  $t$  de Student, la distribución condicional completa para las variables latentes  $\omega_i^2$  es la distribución inversa gamma  $IGa(\tilde{\alpha}, \tilde{\beta}_i)$  donde

$$\tilde{\alpha} = \frac{\delta+1}{2} \quad y \quad \tilde{\beta}_i = \frac{1}{2} \left(\frac{\epsilon_i^2}{\sigma^2} + \delta\right),$$

siendo  $\sigma^2$  el parámetro de escala y  $\epsilon_i$  los residuales del modelo.

Esto se verifica observando que:

$$\begin{aligned}
p(\omega_i^2 \mid y_i, \beta_0, \boldsymbol{\beta}, \sigma^2, \lambda, \tau^2) &= \frac{p(y_i, \beta_0, \boldsymbol{\beta}, \omega_i^2, \sigma^2, \lambda, \tau^2)}{p(y_i, \beta_0, \boldsymbol{\beta}, \sigma^2, \lambda, \tau^2)} \\
&\propto p(y_i \mid \beta_0, \boldsymbol{\beta}, \omega_i^2, \sigma^2, \lambda, \tau^2) p(\boldsymbol{\beta} \mid \omega_i^2, \sigma^2, \lambda, \tau^2) p(\omega_i^2 \mid \sigma^2, \lambda, \tau^2) p(\sigma^2 \mid \lambda, \tau^2) p(\lambda \mid \tau^2) p(\tau^2) \\
&\propto \frac{1}{\sqrt{2\pi\omega_i^2\sigma^2}} \exp\left(-\frac{1}{2\omega_i^2\sigma^2}(y_i - \beta_0 - \mathbf{x}_i^T \boldsymbol{\beta})^2\right) \cdot \left(\frac{1}{\omega_i^2}\right)^{\delta/2+1} \exp\left(-\frac{\delta/2}{\omega_i^2}\right) \\
&\propto \frac{1}{\sigma} \left(\frac{1}{\omega_i^2}\right)^{\frac{\delta+1}{2}+1} \exp\left(-\frac{\frac{1}{2}\left(\frac{\epsilon_i^2}{\sigma^2} + \delta\right)}{\omega_i^2}\right)
\end{aligned}$$

lo cual significa que

$$\omega_i^2 \sim IGa\left(\tilde{\alpha} = \frac{\delta+1}{2}, \tilde{\beta} = \frac{1}{2}\left(\frac{\epsilon_i^2}{\sigma^2} + \delta\right)\right).$$

Análogamente, para el caso de la distribución de Laplace uno puede obtener que la distribución condicional completa de las variables latentes  $1/\omega_i^2$  es la distribución inversa Gaussiana  $IG(\tilde{\mu}_i, 2)$ , donde

$$\tilde{\mu}_i = \left(\frac{2\sigma^2}{\epsilon_i^2}\right)^{\frac{1}{2}}$$

siendo  $\sigma^2$  el parámetro de escala y  $(\epsilon_1, \dots, \epsilon_n)$  los residuales del modelo.

### 3.6. Regresión logística binaria

En el caso de la variable respuesta binaria  $\tilde{y}_i \in \{0, 1\}$ , la relación entre las variables predictoras y la variable objetivo es representada usando modelos de regresión logística.

La regresión logística es un caso asociado al modelo presentado en la sección 3.5 pero formalmente es otro modelo ya que la variable respuesta es binaria y no normal, pero con características similares, ya que como vimos en 2.1.4. éste es un modelo con MEG y, consecuentemente, se le puede aplicar la misma metodología.

Según (2.6), la implementación de la regresión logística bayesiana con la representación Pólya-gamma requiere solo de las variables latentes  $\boldsymbol{\omega}^2 = (\omega_1^2, \dots, \omega_n^2)^T$ . Para obtener el muestreador de Gibbs, podemos usar (2.6) y escribir

$$\begin{aligned}
p(Y_i \mid \boldsymbol{\beta}) &= \frac{\{\exp(x_i^T \boldsymbol{\beta})\}^{y_i}}{1 + \exp(x_i^T \boldsymbol{\beta})} \\
&\propto \exp(\kappa_i x_i^T \boldsymbol{\beta}) \int_0^\infty \exp\left\{-\omega_i^2 (x_i^T \boldsymbol{\beta})^2 / 2\right\} p(\omega_i^2) d\omega_i^2
\end{aligned}$$

donde  $\kappa_i = y_i - 1/2$ , y  $p(\omega_i^2)$  es la densidad de una variable aleatoria Pólya Gamma con parámetros  $(0, 1)$ . Combinando los términos para toda la data se obtiene la siguiente expresión para la posterior condicional de  $\beta$

$$\begin{aligned} p(\beta \mid \omega^2, \mathbf{y}) &\propto p(\beta)p(\mathbf{y} \mid \beta, \omega^2) \propto p(\beta) \prod_{i=1}^N p(y_i \mid \beta, \omega_i^2) = p(\beta) \prod_{i=1}^N \exp\left(\kappa_i x_i^T \beta - \omega_i^2 (x_i^T \beta)^2 / 2\right) \\ &\propto p(\beta) \prod_{i=1}^N \exp\left(\frac{\omega_i^2}{2} (x_i^T \beta - \kappa_i / \omega_i^2)^2\right) \\ &\propto p(\beta) \exp\left(-\frac{1}{2}(\mathbf{z} - \mathbf{X}\beta)^T \Omega^{-1}(\mathbf{z} - \mathbf{X}\beta)\right) \end{aligned}$$

donde  $\mathbf{z} = (\kappa_1/\omega_1^2, \dots, \kappa_n/\omega_n^2)$  y  $\Omega = \text{diag}(\omega_1^2, \dots, \omega_n^2)$ . Esta corresponde a una distribución Gaussiana en el vector respuesta  $\mathbf{z}$ . Así, si imponemos una distribución a priori Gaussiana o de Horseshoe  $p(\beta)$  para  $\beta$ , el muestreador de Gibbs resultará directo.

El éxito de esta estrategia depende de la existencia de una forma sencilla y eficaz de simular variables aleatorias Pólya-Gamma. La representación de la suma de gammas en la distribución enunciada parece compleja y sugiere una aproximación finita. Pero Polson et al. (2013) describen un método de simulación Pólya-Gamma rápido y exacto que evita las dificultades que pueden resultar de truncar la suma infinita. Un algoritmo eficiente para el muestreo desde la distribución Pólya-gamma fue recientemente propuesta por Jesse Windle y Scott (2014) y es usada en la implementación del paquete bayesreg de Makalic y Schmidt (2016a).

### 3.7. Estimación bayesiana de los parámetros de contracción

Las variables latentes  $\tau^2$  y  $(\lambda_1, \dots, \lambda_p)$  determinan el tipo de esparcidad que se aplica a los coeficientes de regresión  $\beta$ . El hiperparámetro  $\tau^2$  corresponde al parámetro de varianza global que controla la cantidad de contracción general de los coeficientes. Polson y Scott (2010) sugiere que la distribución a priori para  $\tau^2$  debe tener una masa sustancial a priori en la vecindad de cero para contraer los parámetros de regresión y suprimir el ruido. En ausencia de conocimiento a priori de expertos, Makalic y Schmidt (2016a) eligen la siguiente distribución a priori para el parámetro de contracción global  $\tau$ :

$$\tau \sim C^+(0, 1),$$

donde  $C^+(0, 1)$  es la distribución half-Cauchy estándar. Polson y Scott (2010) recomiendan esta como una distribución predeterminada sensata, la cual también la encontramos sugerida en Gelman (2006). Es interesante saber que la distribución half-Cauchy puede ser escrita como una mixtura de distribuciones inversa gamma, tal como nos dice la siguiente proposición:

### **PROPOSICIÓN 3.3**

Sea  $X$  una variable aleatoria no negativa, tal que  $X^2 \mid V = \nu, b \sim \text{IGa}(b, 1/\nu)$  y  $V \sim \text{IGa}(a, 1)$ . Entonces la función de densidad de  $X$  viene dada por

$$f(x) = \frac{2x^{2a-1}}{\text{Beta}(a, b)(1+x^2)^{a+b}}, \quad x > 0$$

donde  $\text{Beta}(a, b)$  denota a la función beta.

En particular, si  $a = b = \frac{1}{2}$ , esta define la distribución a priori half-Cauchy estándar.

#### **Demostración:**

Mostraremos, equivalentemente, que si  $Y \mid \mathbf{X} = \mathbf{x} \sim \text{IGa}(b, \frac{1}{x})$  y  $X \sim \text{IGa}(a, 1)$  entonces  $\sqrt{Y}$  tiene la densidad arriba dada. En efecto,

$$\begin{aligned} f_Y(y) &= \int_0^\infty f_{Y|X=x}(y) f_X(x) dx \\ &= \int_0^\infty \frac{\left(\frac{1}{x}\right)^b}{\Gamma(b)} \frac{1}{y^{b+1}} \exp\left(-\frac{1}{xy}\right) \frac{1}{\Gamma(a)} x^{a+1} \exp\left(-\frac{1}{x}\right) dx \\ &= \frac{1}{y^{b+1}\Gamma(a)\Gamma(b)} \int_0^\infty \frac{1}{x^{a+b+1}} \exp\left(-\frac{1}{x}\left(1+\frac{1}{y}\right)\right) dx \end{aligned}$$

Haciendo un cambio de variable  $u = \left(1 + \frac{1}{y}\right)\frac{1}{x}$  se tendrá que:

$$\begin{aligned} f_Y(y) &= \frac{1}{y^{b+1}\Gamma(a)\Gamma(b)} \int_\infty^0 \left(\frac{u}{1+\frac{1}{y}}\right)^{a+b+1} \exp(-u) \left(\frac{1+\frac{1}{y}}{u}\right)^2 \frac{1}{\left(1+\frac{1}{y}\right)} du \\ &= \frac{1}{y^{b+1}\Gamma(a)\Gamma(b)} \frac{1}{\left(1+\frac{1}{y}\right)^{a+b}} \int_0^\infty u^{a+b-1} \exp^{-u} du \\ &= \frac{y^{a+b}}{(1+y)^{a+b} y^{b+1}} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \\ &= \frac{1}{\text{Beta}(a, b)} \frac{y^{a-1}}{(1+y)^{a+b}} \end{aligned}$$

Se sabe que si  $Y$  es v.a. con función de densidad  $f_Y$ , se cumple que  $\sqrt{Y}$  tiene como función

de densidad a  $f_Y(y^2)2y$ , entonces la función de densidad de  $\sqrt{Y}$  es:

$$\begin{aligned} f_{\sqrt{y}}(y) &= \frac{1}{\text{Beta}(a, b)} \frac{y^{2a-2}}{(1+y^2)^{a+b}} 2y \\ &= \frac{1}{\text{Beta}(a, b)} \frac{2y^{2a-1}}{(1+y^2)^{a+b}} \end{aligned}$$

En particular, si  $a = b = \frac{1}{2}$ , esta corresponde a la distribución half-Cauchy como se aprecia seguidamente

$$\begin{aligned} f(y) &= \frac{1}{\text{Beta}(\frac{1}{2}, \frac{1}{2})} \frac{2}{(1+y^2)} \\ &= \frac{2}{\pi(1+y^2)} \end{aligned}$$

■

La proposición anterior nos dice entonces que toda priori  $\tau \sim C^+(0, 1)$ , podría generarse mediante

$$\tau^2 | V = \nu \sim \text{IGa}(1/2, 1/\nu) \text{ y } V \sim \text{IGa}(1/2, 1),$$

donde  $V > 0$  es una variable de mixtura.

Considerando la representación anterior de  $\tau$ , la distribución condicional completa de  $\tau^2$  resulta ser entonces la distribución inversa gamma  $\text{IGa}(\tilde{\alpha}, \tilde{\beta})$ , donde

$$\tilde{\alpha} = \frac{p+1}{2} \text{ y } \tilde{\beta} = \frac{1}{\nu} + \frac{1}{2\sigma^2} \sum_{j=1}^p \frac{\beta_j^2}{\lambda_j^2}.$$

Esto se verifica realizando el siguiente desarrollo:

$$\begin{aligned}
p(\tau^2 \mid \mathbf{y}, \beta_0, \boldsymbol{\beta}, \omega^2, \sigma^2, \lambda, \nu) &= \frac{p(\mathbf{y}, \beta_0, \boldsymbol{\beta}, \omega^2, \sigma^2, \lambda, \tau^2, \nu)}{p(\mathbf{y}, \beta_0, \boldsymbol{\beta}, \omega^2, \sigma^2, \lambda, \nu)} \\
&\propto p(\mathbf{y} \mid \beta_0, \boldsymbol{\beta}, \omega^2, \sigma^2, \lambda, \tau^2, \nu) p(\boldsymbol{\beta} \mid \omega^2, \sigma^2, \lambda, \tau^2, \nu) p(\sigma^2 \mid \lambda, \tau^2, \nu) p(\lambda \mid \tau^2, \nu) p(\tau^2 \mid \nu) p(\nu) \\
&\propto \prod_{i=1}^n p(y_i \mid \beta_0, \boldsymbol{\beta}, \omega^2, \sigma^2, \lambda, \tau^2, \nu) \prod_{j=1}^p p(\boldsymbol{\beta}_j \mid \omega^2, \sigma^2, \lambda_j, \tau_j^2, \nu) \prod_{j=1}^p p(\lambda_j \mid \tau^2, \nu) \prod_{j=1}^p p(\tau_j^2 \mid \nu) \\
&\propto \prod_{i=1}^n \frac{1}{\sqrt{2\pi\omega_i^2\sigma^2}} \exp\left(-\frac{1}{2\omega_i^2\sigma^2}(y_i - \beta_0 - x_i^T \boldsymbol{\beta})^2\right) \prod_{j=1}^p \frac{1}{\sqrt{2\pi\lambda_j^2\tau_j^2\sigma^2}} \exp\left(-\frac{1}{2\lambda_j^2\tau_j^2\sigma^2}(\boldsymbol{\beta}_j)^2\right) \cdot \\
&\quad \left[ \frac{1}{\pi(1 + \lambda_j^2)} \right] \prod_{j=1}^p \left( \frac{1}{\tau_j^2} \right)^{\frac{1}{2}+1} \exp\left(\frac{-1/\nu}{\tau_j^2}\right) \\
&\propto \frac{1}{\sigma^p} \left( \prod_{j=1}^p \frac{1}{\lambda_j\tau_j} \right) \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^p \frac{\beta_j^2}{\lambda_j^2\tau_j^2}\right) \cdot \prod_{j=1}^p \left( \frac{1}{\tau_j^2} \right)^{\frac{1}{2}+1} \exp\left(\frac{-1/\nu}{\sum_{j=1}^p \tau_j^2}\right) \\
&\propto \left( \prod_{j=1}^p \frac{1}{\tau_j^2} \right)^{\frac{1}{2}} \prod_{j=1}^p \left( \frac{1}{\tau_j^2} \right)^{\frac{1}{2}+1} \cdot \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^p \frac{\beta_j^2}{\lambda_j^2\tau_j^2}\right) \exp\left(\frac{-1/\nu}{\sum_{j=1}^p \tau_j^2}\right) \\
&\propto \left( \frac{1}{\tau^2} \right)^{\left(\frac{p+1}{2}+1\right)} \exp\left(-\frac{\frac{1}{\nu} + \frac{1}{2\sigma^2} \sum_{j=1}^p \frac{\beta_j^2}{\lambda_j^2}}{\tau^2}\right)
\end{aligned}$$

lo cual implica que

$$\tau^2 \sim IGa\left(\tilde{\alpha} = \frac{p+1}{2}, \tilde{\beta} = \frac{1}{\nu} + \frac{1}{2\sigma^2} \sum_{j=1}^p \frac{\beta_j^2}{\lambda_j^2}\right).$$

De manera similar la distribución condicional completa para la variable de mixtura  $V$  es la distribución inversa gamma  $IGa(\tilde{\alpha}, \tilde{\beta})$ , donde

$$\tilde{\alpha} = 1 \quad y \quad \tilde{\beta} = 1 + \frac{1}{\tau^2}$$

Esto se verifica realizando el siguiente desarrollo:

$$\begin{aligned}
 p(\nu \mid \mathbf{y}, \beta_0, \boldsymbol{\beta}, \omega^2, \sigma^2, \lambda, \tau^2) &= \frac{p(\mathbf{y}, \beta_0, \boldsymbol{\beta}, \omega^2, \sigma^2, \lambda, \tau^2, \nu)}{p(\mathbf{y}, \beta_0, \boldsymbol{\beta}, \omega^2, \sigma^2, \lambda, \tau^2)} \\
 &\propto p(\mathbf{y} \mid \beta_0, \boldsymbol{\beta}, \omega^2, \sigma^2, \lambda, \tau^2, \nu) p(\boldsymbol{\beta} \mid \omega^2, \sigma^2, \lambda, \tau^2, \nu) p(\sigma^2 \mid \lambda, \tau^2, \nu) p(\lambda \mid \tau^2, \nu) p(\tau^2 \mid \nu) p(\nu) \\
 &\propto \prod_{i=1}^n p(y_i \mid \beta_0, \boldsymbol{\beta}, \omega^2, \sigma^2, \lambda, \tau^2, \nu) \prod_{j=1}^p p(\boldsymbol{\beta}_j \mid \omega^2, \sigma^2, \lambda_j, \tau_j^2, \nu) \prod_{j=1}^p p(\lambda_j \mid \tau^2, \nu) \prod_{j=1}^p p(\tau_j^2 \mid \nu) \prod_{i=1}^n p(\nu_i) \\
 &\propto \prod_{i=1}^n \frac{1}{\sqrt{2\pi\omega_i^2\sigma^2}} \exp\left(-\frac{1}{2\omega_i^2\sigma^2}(y_i - \beta_0 - \mathbf{x}_i^T \boldsymbol{\beta})^2\right) \prod_{j=1}^p \left(\frac{1}{\tau_j^2}\right)^{\frac{1}{2}+1} \exp\left(\frac{-1/\nu}{\tau_j^2}\right) \prod_{i=1}^n \left(\frac{1}{\nu_i}\right)^{\frac{1}{2}+1} \exp\left(\frac{-1}{\nu_i}\right) \\
 &\propto \prod_{j=1}^p \left(\frac{1}{\tau_j^2}\right)^{\frac{1}{2}+1} \exp\left(\frac{-1/\nu}{\tau_j^2}\right) \prod_{i=1}^n \left(\frac{1}{\nu_i}\right)^{\frac{1}{2}+1} \exp\left(\frac{-1}{\nu_i}\right) \\
 &\propto \prod_{i=1}^n \left(\frac{1}{\nu_i}\right)^{\frac{1}{2}+1} \cdot \exp\left(\frac{-1/\nu}{\sum_{j=1}^p \tau_j^2}\right) \exp\left(\frac{-1}{\sum_{i=1}^n \nu_i}\right) \\
 &\propto \left(\frac{1}{\nu}\right)^{(1+1)} \exp\left(-\frac{1 + \frac{1}{\tau^2}}{\nu}\right)
 \end{aligned}$$

lo cual implica que

$$V \sim IGa\left(\tilde{\alpha} = 1, \tilde{\beta} = 1 + \frac{1}{\tau^2}\right).$$

Los parámetros  $(\lambda_1, \dots, \lambda_p)$  corresponden a los componentes de contracción local que determinan el tipo de penalidad de contracción aplicada a los coeficientes de regresión. Según Polson y Scott (2012), los hiperparámetros de varianza local deben tener una distribución a priori con un polo en cero para garantizar la eficiencia predictiva en la recuperación de la verdadera distribución muestral de los parámetros de regresión  $\boldsymbol{\beta} \in \mathbb{R}^p$  y colas polinómicas para contraer agresivamente las variables de ruido y al mismo tiempo permitir que las señales grandes permanezcan. En las siguientes sub-secciones se consideran los detalles de muestreo para los estimadores Ridge bayesiano, Lasso bayesiano, Horseshoe y Horseshoe+ de estos parámetros de contracción local. De estos estimadores, sólo los estimadores de Horseshoe y Horseshoe+ poseen las dos propiedades de contracción deseables (Makalic y Schmidt, 2016a).

### 3.7.1. Regresión ridge

La regresión ridge bayesiana requiere solo un parámetro de contracción global  $\tau$  y no incluye parámetros de contracción local. Por lo que Makalic y Schmidt (2016a) configuran  $\lambda_j = 1$  para todo  $j = 1, \dots, p$ , lo cual implica que ya no es necesario el muestreo de estos parámetros.

### 3.7.2. Regresión lasso

El estimador Lasso bayesiano en Park y Casella (2008) y en Hans (2009), requieren usar una distribución a priori Laplace para los coeficientes de regresión por lo que la moda de la densidad a posteriori resultante puede ser representada como una distribución Gaussiana con mixtura de escala donde la densidad de mixtura es una distribución exponencial. Esto es precisamente lo que se vió en 3.2.1. Los parámetros  $(\lambda_1^2, \dots, \lambda_p^2)$  por lo tanto se podrían modelar (aunque dependiendo del factor de penalización) por:

$$\lambda_j^2 \sim \text{Exp}(1).$$

con media uno.

Según Park y Casella (2008), la distribución condicional completa para los parámetros de contracción local  $1/\lambda_j^2$ , en este caso, es la distribución inversa Gaussiana  $IG(\tilde{\mu}_j, \tilde{\lambda})$ , donde

$$\tilde{\mu}_j = \left( \frac{2\tau_j^2\sigma^2}{\beta_j^2} \right)^{\frac{1}{2}} y \quad \tilde{\lambda} = 2.$$

Esto se verifica realizando el siguiente desarrollo:

$$\begin{aligned} p(\lambda_j^2 \mid \mathbf{y}, \beta_0, \beta_j, \omega^2, \sigma^2, \tau_j^2) &= \frac{p(\mathbf{y}, \beta_0, \beta_j, \omega^2, \sigma^2, \lambda_j^2, \tau_j^2)}{p(\mathbf{y}, \beta_0, \beta_j, \omega^2, \sigma^2, \tau_j^2)} \\ &\propto p(\mathbf{y} \mid \beta_0, \beta_j, \omega^2, \sigma^2, \lambda_j^2, \tau_j^2) p(\beta_j \mid \omega^2, \sigma^2, \lambda_j^2, \tau_j^2) p(\sigma^2 \mid \lambda_j^2, \tau_j^2, \nu_j) p(\lambda_j^2 \mid \tau_j^2) p(\tau_j^2) \\ &\propto p(\beta_j \mid \omega^2, \sigma^2, \lambda_j^2, \tau_j^2) p(\lambda_j^2 \mid \tau_j^2) p(\tau_j^2) \\ &\propto \frac{1}{\sqrt{2\pi\lambda_j^2\tau_j^2\sigma^2}} \exp\left(-\frac{1}{2\lambda_j^2\tau_j^2\sigma^2}(\beta_j)^2\right) \exp(-\lambda_j^2) \end{aligned}$$

lo cual implica que

$$\frac{1}{\lambda_j^2} \sim IG\left(\tilde{\mu}_j = \left(\frac{2\tau_j^2\sigma^2}{\beta_j^2}\right)^{\frac{1}{2}}, \tilde{\lambda} = 2\right)$$

Entonces, Makalic y Schmidt (2016a), trabajan con una jerarquía de lasso bayesiana que mueve el parámetro de contracción global  $\tau^2$  hacia abajo en la jerarquía y al mismo nivel que los hiperparámetros de contracción local, los cuales permiten expresar el lasso bayesiano como un estimador de contracción global-local, como el Horseshoe y Horseshoe+. Otra ventaja de esta formulación es que alivia la necesidad de especificar hiperparámetros en el nivel más alto de la jerarquía.

### 3.7.3. Regresión Horseshoe

A diferencia de las distribuciones a priori lasso y ridge, la priori Horseshoe exhibe un polo en cero y colas polinómicas, propiedades importantes que garantizan un buen rendimiento en el dominio de big data (Polson y Scott, 2012). La distribución a priori para los hiperparámetros de contracción  $(\lambda_1, \dots, \lambda_p)$  es la distribución half-Cauchy estándar

$$\lambda_j \sim C^+(0, 1),$$

la cual, por la proposición 3.3, puede ser escrita de forma equivalente como:

$$\lambda_j^2 \mid \nu_j \sim IGa(1/2, 1/\nu_j) \text{ y } \nu_j \sim IGa(1/2, 1).$$

La distribución condicional completa para el parámetro de contracción local  $(1/\lambda_j^2)$  es la distribución exponencial  $Exp(\tilde{\beta}_j)$ , donde

$$\tilde{\beta}_j = \frac{1}{\nu_j} + \frac{\beta_j^2}{2\tau_j^2\sigma^2}.$$

Esto se verifica realizando el siguiente desarrollo:

$$\begin{aligned} p(\lambda_j^2 \mid \mathbf{y}, \beta_0, \beta_j, \omega^2, \sigma^2, \tau_j^2, \nu_j) &= \frac{p(\mathbf{y}, \beta_0, \beta_j, \omega^2, \sigma^2, \lambda_j^2, \tau_j^2, \nu_j)}{p(\mathbf{y}, \beta_0, \beta_j, \omega^2, \sigma^2, \tau_j^2, \nu_j)} \\ &\propto p(\mathbf{y} \mid \beta_0, \beta_j, \omega^2, \sigma^2, \lambda_j^2, \tau_j^2, \nu_j) p(\beta_j \mid \omega^2, \sigma^2, \lambda_j^2, \tau_j^2, \nu_j) p(\sigma^2 \mid \lambda_j^2, \tau_j^2, \nu_j) p(\lambda_j^2 \mid \tau_j^2, \nu_j) p(\tau_j^2 \mid \nu_j) p(\nu_j) \\ &\propto p(\beta_j \mid \omega^2, \sigma^2, \lambda_j, \tau_j^2, \nu_j) p(\lambda_j^2 \mid \tau_j^2, \nu_j) p(\nu_j) \\ &\propto \frac{1}{\sqrt{2\pi\lambda_j^2\tau_j^2\sigma^2}} \exp\left(-\frac{1}{2\lambda_j^2\tau_j^2\sigma^2}(\beta_j)^2\right) \left(\frac{1}{\lambda_j^2}\right)^{\frac{1}{2}+1} \exp\left(-\frac{1}{\lambda_j^2}\right) \left(\frac{1}{\nu_j}\right)^{\frac{1}{2}+1} \exp\left(-\frac{1}{\nu_j}\right) \\ &\propto \left(\frac{1}{\lambda_j^2}\right)^2 \exp\left(-\left(\frac{1}{\nu_j} + \frac{\beta_j^2}{2\tau_j^2\sigma^2}\right)\frac{1}{\lambda_j^2}\right) \end{aligned}$$

lo cual implica que

$$\frac{1}{\lambda_j^2} \sim Exp\left(\frac{1}{\nu_j} + \frac{\beta_j^2}{2\tau_j^2\sigma^2}\right)$$

De forma similar, las distribuciones condicionales completas para los hiperparámetros  $\lambda_j$  y  $\nu_j$  son distribuciones exponenciales para las cuales existen algoritmos de muestreo eficientes, incluso cuando el número  $p$  de predictores es grande.

### 3.7.4. Regresión Horseshoe +

El estimador Horseshoe + de Bhadra et al. (2016) es, como vimos, una extensión natural del estimador de Horseshoe para problemas de ultra-esparcidad. A diferencia del estimador de Horseshoe, el estimador de Horseshoe + tiene un error cuadrático medio a posteriori menor y ratios de concentración a posteriori más rápidos en términos de la métrica de divergencia Kullback-Leibler. De forma similar al estimador de Horseshoe, la distribución a priori para los hiperparámetros de contracción local  $(\lambda_1, \dots, \lambda_p)$  es la distribución half-Cauchy con media cero, donde:

$$\lambda_j \mid \eta_j \sim C^+(0, \eta_j) \text{ y } \eta_j \sim C^+(0, 1).$$

La diferencia entre los estimadores Horseshoe y el Horseshoe +, es que el Horseshoe + incorpora un nivel extra de parámetros  $(\eta_1, \dots, \eta_p)$ , donde cada  $\eta_j$  corresponde a la varianza a priori asociada con el parámetro  $\lambda_j$ . Teniendo en cuenta la expansión de parámetros propuesta en Makalic y Schmidt (2016b), la jerarquía de Horseshoe + equivale a:

$$\begin{aligned} \lambda_j^2 \mid \nu_j &\sim IGa(1/2, 1/\nu_j), \\ \nu_j \mid \eta_j^2 &\sim IGa(1/2, 1/\eta_j^2), \\ \eta_j^2 \mid \varsigma_j &\sim IGa(1/2, 1/\varsigma_j), \\ \varsigma_j &\sim IGa(1/2, 1). \end{aligned}$$

Esto permite un cálculo sencillo de las distribuciones condicionales completas para todos los parámetros. En particular, la distribución condicional completa para los parámetros de contracción local  $1/\lambda_j^2$  es la distribución exponencial  $Exp(\tilde{\beta}_j)$ , donde

$$\tilde{\beta}_j = \frac{1}{\nu_j} + \frac{\beta_j^2}{2\tau^2\sigma^2},$$

la cual es equivalente a la densidad posteriori condicional correspondiente en el estimador de Horseshoe. Las densidades posteriori condicionales para los otros hiperparámetros son:

$$\begin{aligned} 1/\nu_j &\sim Exp\left(\frac{1}{\eta_j^2} + \frac{1}{\lambda_j^2}\right), \\ 1/\eta_j^2 &\sim Exp\left(\frac{1}{\nu_j} + \frac{1}{\varsigma_j}\right), \\ 1/\varsigma_j &\sim Exp\left(1 + \frac{1}{\eta_j^2}\right). \end{aligned}$$

Esto se verifica realizando el siguiente desarrollo, para  $\nu_j$ :

$$\begin{aligned}
 p(\nu_j \mid \mathbf{y}, \beta_0, \beta_j, \omega^2, \sigma^2, \lambda_j^2, \tau_j^2, \eta_j^2, \varsigma_j) &= \frac{p(\mathbf{y}, \beta_0, \beta_j, \omega^2, \sigma^2, \lambda_j^2, \tau_j^2, \nu_j, \eta_j^2, \varsigma_j)}{p(\mathbf{y}, \beta_0, \beta_j, \omega^2, \sigma^2, \lambda_j^2, \tau_j^2, \eta_j^2, \varsigma_j)} \\
 &\propto p(\mathbf{y} \mid \beta_0, \beta_j, \omega^2, \sigma^2, \lambda_j^2, \tau_j^2, \nu_j, \eta_j^2, \varsigma_j) p(\beta_j \mid \omega^2, \sigma^2, \lambda_j^2, \tau_j^2, \nu_j, \eta_j^2, \varsigma_j) p(\sigma^2 \mid \lambda_j^2, \tau_j^2, \nu_j, \eta_j^2, \varsigma_j) \\
 &p(\lambda_j^2 \mid \tau_j^2, \nu_j, \eta_j^2, \varsigma_j) p(\tau_j^2 \mid \nu_j, \eta_j^2, \varsigma_j) p(\nu_j \mid \eta_j^2, \varsigma_j) p(\eta_j^2 \mid \varsigma_j) p(\varsigma_j) \\
 &\propto p(\lambda_j^2 \mid \tau_j^2, \nu_j, \eta_j^2, \varsigma_j) p(\nu_j \mid \eta_j^2, \varsigma_j) \\
 &\propto \left(\frac{1}{\lambda_j^2}\right)^{\frac{1}{2}+1} \exp\left(-\frac{\nu_j}{\lambda_j^2}\right) \left(\frac{1}{\nu_j}\right)^{\frac{1}{2}+1} \exp\left(-\frac{1}{\eta_j^2 \nu_j}\right) \\
 &\propto \left(\frac{1}{\nu_j}\right)^{\frac{1}{2}+1} \exp\left(-\left(\frac{1}{\eta_j^2} + \frac{1}{\lambda_j^2}\right) \frac{1}{\nu_j}\right)
 \end{aligned}$$

lo cual implica que

$$\frac{1}{\nu_j} \sim \text{Exp}\left(\frac{1}{\eta_j^2} + \frac{1}{\lambda_j^2}\right).$$

Ahora se muestra el desarrollo, para  $\eta_j^2$ :

$$\begin{aligned}
 p(\eta_j^2 \mid \mathbf{y}, \beta_0, \beta_j, \omega^2, \sigma^2, \lambda_j^2, \tau_j^2, \nu, \varsigma_j) &= \frac{p(\mathbf{y}, \beta_0, \beta_j, \omega^2, \sigma^2, \lambda_j^2, \tau_j^2, \nu, \eta_j^2, \varsigma_j)}{p(\mathbf{y}, \beta_0, \beta_j, \omega^2, \sigma^2, \lambda_j^2, \tau_j^2, \nu, \varsigma_j)} \\
 &\propto p(\mathbf{y} \mid \beta_0, \beta_j, \omega^2, \sigma^2, \lambda_j^2, \tau_j^2, \nu, \eta_j^2, \varsigma_j) p(\beta_j \mid \omega^2, \sigma^2, \lambda_j^2, \tau_j^2, \nu, \eta_j^2, \varsigma_j) p(\sigma^2 \mid \lambda_j^2, \tau_j^2, \nu, \eta_j^2, \varsigma_j) \\
 &p(\lambda_j^2 \mid \tau_j^2, \nu, \eta_j^2, \varsigma_j) p(\tau_j^2 \mid \nu, \eta_j^2, \varsigma_j) p(\nu \mid \eta_j^2, \varsigma_j) p(\eta_j^2 \mid \varsigma_j) p(\varsigma_j) \\
 &\propto p(\nu \mid \eta_j^2, \varsigma_j) p(\eta_j^2 \mid \varsigma_j) \\
 &\propto \left(\frac{1}{\nu_j}\right)^{\frac{1}{2}+1} \exp\left(-\frac{1}{\eta_j^2 \nu_j}\right) \left(\frac{1}{\eta_j^2}\right)^{\frac{1}{2}+1} \exp\left(-\frac{1}{\eta_j^2 \varsigma_j}\right) \\
 &\propto \left(\frac{1}{\eta_j^2}\right)^{\frac{1}{2}+1} \exp\left(-\left(\frac{1}{\nu} + \frac{1}{\varsigma}\right) \frac{1}{\eta_j^2}\right)
 \end{aligned}$$

lo cual implica que

$$\frac{1}{\eta_j^2} \sim \text{Exp}\left(\frac{1}{\nu_j} + \frac{1}{\varsigma_j}\right).$$

De la misma manera se puede hacer el desarrollo para  $\varsigma_j$ .

Luego de revisar cómo se aplica la inferencia bayesiana en los modelos MEG con regularización bayesiana pasaremos a revisar cómo se realiza la implementación de dicha estimación.

## Capítulo 4

# Implementación de la estimación bayesiana en el modelo de estudio

La distribución a posteriori del modelo jerárquico de regresión lineal con mixtura de escala Gaussiana es compleja y difícil de manejar, por lo que para su generación utilizaremos métodos MCMC, más concretamente el muestreador de Gibbs el cual repetidamente iterará las siguientes distribuciones condicionales completas:

1. Muestreo de los coeficientes de regresión desde  $p(\boldsymbol{\beta} \mid \beta_0, \sigma^2, \boldsymbol{\lambda}, \tau, \mathbf{y})$ ;
2. Muestreo de los parámetros de intercepto y varianza del modelo desde  $p(\beta_0, \mid \boldsymbol{\beta}, \boldsymbol{\lambda}, \tau, \mathbf{y})$  y  $p(\sigma^2 \mid \boldsymbol{\beta}, \boldsymbol{\lambda}, \tau, \mathbf{y})$ ;
3. Muestreo de los parámetros que inducen esparcidad desde  $p(\boldsymbol{\lambda}, \tau \mid \boldsymbol{\beta}, \sigma^2, \mathbf{y})$ .

Si bien en la sección 3.5 discutimos la estimación bayesiana del modelo de regresión lineal con MEG, mediante un muestreador de Gibbs, la implementación de este no es siempre eficiente, por lo que en este capítulo buscaremos optimizar esta, de ser necesario, en base a otras técnicas bayesianas.

### 4.1. Muestreo de los coeficientes de regresión $\boldsymbol{\beta}$

Si bien es cierto en la sección 3.5 se presentó la distribución condicional completa de los coeficientes de regresión  $\boldsymbol{\beta}$ , el paquete bayesreg implementado por Schmidt y Makalic (2019) no la emplea como parte del algoritmo de Gibbs, porque cuando se trata de una gran cantidad de covariables  $p$ , este es poco eficiente y computacionalmente costoso. En términos de eficiencia computacional, el mayor cuello de botella de este algoritmo de Gibbs

propuesto es el muestreo desde la distribución Gaussiana p-variada, vista en el capítulo anterior,  $N_p(\tilde{\boldsymbol{\mu}}, \mathbf{A}_p^{-1})$ , en donde el cálculo computacional de la inversa de la matriz  $\mathbf{A}_p^{-1}$  es costoso y presenta una pobre precisión numérica. En vez de ello, Schmidt y Makalic (2019) proponen usar los algoritmos de muestreo basados en gradiente auxiliar recientemente desarrollados. Estos son algoritmos de Metrópolis-Hastings adaptativos que aumentan la densidad de la variable objetivo con variables aleatorias auxiliares y adicionalmente usan una expansión de series de Taylor de primer orden en la verosimilitud y un paso de marginalización para construir la densidad de propuesta Metropolis-Hastings que se desempeña de ambas maneras como verosimilitud y como priori informativa. Este algoritmo es aplicable, en general, cuando la distribución a posteriori tiene la forma:

$$p(\boldsymbol{\beta}) \propto \exp(l(\boldsymbol{\beta}, \beta_0, \sigma^2))N(\boldsymbol{\beta} \mid \mathbf{0}, \mathbf{C})$$

donde  $l(\boldsymbol{\beta}; \beta_0, \sigma^2)$  es una función de log-verosimilitud y  $\mathbf{C}$  es la matriz de covarianza a priori para los coeficientes de regresión  $\boldsymbol{\beta}$ . Afortunadamente en nuestro modelo (3.7) – (3.13), la distribución a posteriori para los coeficientes  $\boldsymbol{\beta}$  tienen esta estructura, lo cual facilita la aplicación de los algoritmos muestreadores basados en gradiente auxiliar dentro de un esquema de muestreo de Gibbs. Adicionalmente, en el modelo, con una función de enlace estándar, la log-verosimilitud  $l(\boldsymbol{\beta}, \beta_0, \sigma^2)$  es una función convexa de los coeficientes de  $\boldsymbol{\beta}$ .

Schmidt y Makalic (2019) usan como punto de partida el muestreador gradiente marginal pre-condicionado general, el cual usa como propuesta (a denotarse con una barra horizontal) la siguiente:

$$\bar{\boldsymbol{\beta}} \mid \boldsymbol{\beta} \sim N\left(\mathbf{A}\left(\nabla f(\boldsymbol{\beta}) + \left(\frac{2}{\delta}\right)\mathbf{S}^{-1}\boldsymbol{\beta}^T\right), \left(\frac{2}{\delta}\right)\mathbf{A}\mathbf{S}^{-1}\mathbf{A} + \mathbf{A}\right), \quad (4.1)$$

donde  $\delta > 0$  es el tamaño del paso en el algoritmo Metropolis-Hasting,  $\mathbf{S}$  es una matriz preconditionante,  $\nabla f(\boldsymbol{\beta}) = \partial l(\boldsymbol{\beta}, \beta_0, \sigma^2) / \partial \boldsymbol{\beta}$  denota el gradiente, y

$$\mathbf{A} = \left(\mathbf{C}^{-1} + \left(\frac{2}{\delta}\right)\mathbf{S}^{-1}\right)^{-1}.$$

El tamaño del paso  $\delta$  debe ser escogido de tal manera que el 50% – 60% de las muestras sean aceptadas (Schmidt y Makalic, 2019). Dependiendo de cómo se elija la matriz preconditionante se derivan los siguientes algoritmos:

#### 4.1.1. Algoritmo 1: mGrad-1

El algoritmo más directo y recomendado usa la especificación  $\mathbf{S} = \mathbf{I}_p$  y se denotará como mGrad-1. En este, la matriz de covarianza  $\mathbf{C}$  es simplemente una matriz diagonal con entradas

$$C_{j,j} = \tau^2 \sigma^2 \lambda_j^2, \quad j = 1, \dots, p.$$

Esto junto con lo propuesto en (4.1) y el paso de aceptación Metropolis-Hastings simplifica la implementación. Este algoritmo que usa solo la información de primer orden trabaja de la siguiente manera:

1. Genera coeficientes propuestos de:

$$\bar{\beta}_j | \beta \sim N \left( \frac{C_{j,j} (\delta [\nabla f(\boldsymbol{\beta})]_j + 2\beta_j)}{2C_{j,j} + \delta}, \frac{\delta C_{j,j} (4C_{j,j} + \delta)}{(2C_{j,j} + \delta)^2} \right)$$

2. Genera  $u \sim U(0, 1)$ , y acepta la nueva propuesta si

$$u < \exp (l(\bar{\boldsymbol{\beta}}, \beta_0, \sigma^2) - l(\boldsymbol{\beta}, \beta_0, \sigma^2) + h_1(\boldsymbol{\beta}, \bar{\boldsymbol{\beta}}) - h_1(\bar{\boldsymbol{\beta}}, \boldsymbol{\beta})),$$

donde

$$h_1(\boldsymbol{\beta}, \bar{\boldsymbol{\beta}}) = \sum_{j=1}^p \left( \beta_j - \frac{C_{j,j} (4\bar{\beta}_j + \delta [\nabla f(\bar{\boldsymbol{\beta}})]_j)}{2(2C_{j,j} + \delta)} \right) \left( \frac{2C_{j,j} + \delta}{4C_{j,j} + \delta} \right) [\nabla f(\bar{\boldsymbol{\beta}})]_j$$

Este algoritmo mGrad-1 tienen una complejidad total computacional de orden  $O(pn)$  para el modelo, permitiendo así su aplicación en problemas con un número grande de variables.

#### 4.1.2. Algoritmo 2: mGrad-2

El algoritmo anterior utiliza solo la información de la verosimilitud de primer orden para generar los coeficientes propuestos. Una potencial mejora es utilizar la información de segundo orden. Schmidt y Makalic (2019) proponen el método de la expansión de segundo orden de las series de Taylor pero su desempeño es pobre y su ejecución es lenta ya que la distribución propuesta depende de la información de segundo orden. Por tal motivo los autores proponen como alternativa fijar  $\mathbf{S} = \mathbf{X}^T \mathbf{X}$ , de esta manera se mantienen la covarianza de las propuestas independientemente del estado y permite la pre-computación de  $\mathbf{S}^{-1}$ . El esfuerzo computacional de este algoritmo es  $O(p^3)$ , lo cual es computacionalmente más complejo que el m-Grad-1 (Schmidt y Makalic, 2019).

## 4.2. Muestreo del intercepto $\beta_0$

En la sección 3.5 se realizó el desarrollo del cálculo de la distribución condicional completa para el intercepto  $\beta_0$ , que es usada en el algoritmo de Gibbs, pero computacionalmente el paquete bayesreg utiliza sus propios algoritmos para lograr una mejor mezcla de las cadenas. Schmidt y Makalic (2019) proponen muestrear el intercepto  $\beta_0$  por separado usando el siguiente algoritmo que no depende del parámetro del tamaño del paso  $\delta$ :

1. Generar la propuesta desde

$$\bar{\beta}_0 | \beta_0 \sim N\left(\beta_0, \frac{2,5}{H(\beta_0)}\right)$$

donde  $H(\beta_0)$  es la segunda derivada del negativo de la verosimilitud con respecto a  $\beta_0$ .

2. Generar  $u \sim U(0, 1)$  y aceptar  $\bar{\beta}_0$  si  $u < \exp(l(\boldsymbol{\beta}, \bar{\beta}_0, \sigma^2) - l(\boldsymbol{\beta}, \beta_0, \sigma^2))$ . Con esto, Schmidt y Makalic (2019), obtuvieron como resultados niveles de aceptación en un rango del 50 % al 60 %.

## 4.3. Muestreo de $\sigma^2$

Tal como se desarrolló en la subsección 3.5.3, la distribución condicional completa para  $\sigma^2$  resulta ser la distribución inversa gamma  $IGa(\tilde{\alpha}, \tilde{\beta})$ , donde

$$\tilde{\alpha} = \frac{n+p}{2} \quad y \quad \tilde{\beta} = \frac{1}{2} \left( \sum_{i=1}^n \frac{\epsilon_i^2}{\omega_i^2} + \sum_{j=1}^p \frac{\beta_j^2}{\tau_j^2 \lambda_j^2} \right).$$

y  $\epsilon_i = y_i - \beta_0 - \mathbf{x}_i^T \boldsymbol{\beta}$  denota a los residuales del modelo.

El algoritmo implementado en el paquete bayesreg, utiliza esta distribución, teniendo en cuenta que el vector  $\boldsymbol{\beta}$ , el intercepto  $\beta_0$  ya fueron muestreados en los pasos anteriores con los algoritmos mencionados y los demás parámetros serán muestreados con los procedimientos que seguidamente veremos. Vale añadir, que para el muestreo de las variables latentes  $\omega_i^2$ , el paquete bayesreg utiliza simplemente la distribución condicional completa derivada en la subsección 3.5.4.

## 4.4. Muestreo de los parámetros que inducen esparcidad

Schmidt y Makalic (2019) presentan dos nuevos algoritmos para muestrear la condicional completa de  $\lambda_j^2$  y  $\tau$  para una priori de Horseshoe en los coeficientes del modelo de regresión. Uno es una generalización de un enfoque gamma-inversa gamma-inversa, tratado en la sección

3.7, y el otro es un muestreador de rechazo que aprovecha la log-concavidad de la distribución condicional de  $\log(\lambda_j)$ .

#### 4.4.1. Muestreador Gamma Inversa-Gamma Inversa

Tanto  $\tau$  como los  $\lambda_j^2$  se podrían muestrear por lo desarrollado en la sección 3.7, donde las prioris  $C^+(0, 1)$  para estos parámetros se podrían generar usando la proposición 3.3, demostrada en la sección 3.7 y una variable auxiliar de mixtura  $V$ . Este procedimiento, recordemos, consiste en

1. Primero muestrear de

$$Z | V = \nu \sim \text{IGa} \left( b, \frac{1}{\nu} \right)$$

2. Luego muestrear de la variable auxiliar

$$V \sim \text{IGa}(a, 1)$$

Este esquema que se aplica tanto a  $\tau$  como los  $\lambda_j^2$  y solo requiere muestras de distribuciones gamma inversa en vez de la tradicional distribución Gaussiana inversa substancialmente más compleja, lo que hace que la implementación sea más sencilla y comprensible (Schmidt y Makalic, 2019).

#### 4.4.2. Muestreo de rechazo

Los muestreadores de las distribuciones a priori anteriores tienen 100 % de nivel de aceptación pero presentan autocorrelación debido a su dependencia de variables auxiliares. Como alternativa, para el muestreo de estas distribuciones, surge el muestreo por rechazo, que sacrifica una tasa de aceptación reducida por la eliminación de la autocorrelación en las muestras. Recordemos que un muestreador de rechazo para una densidad objetivo  $p(x)$  trabaja primero con una muestra de una distribución propuesta  $q(x)$  y luego acepta esta muestra si  $p(x)/q(x) > u$ , donde  $u \sim U(0, 1)$ . La distribución propuesta debe satisfacer  $q(x) \geq p(x)$  para todos los  $x$  y debe ser sencilla de generar. Cuanto más cerca este  $q(x)$  de  $p(x)$ , el nivel de aceptación se hace mayor (Schmidt y Makalic, 2019).

Si  $\lambda$  sigue la distribución a priori  $C^+(0, 1)$ , entonces la densidad para la variable transformada  $\xi = \log \lambda$  es

$$p(\xi | a, b) \propto \exp(\xi(2a)) (1 + \exp(2\xi))^{-a-b}. \quad (4.2)$$

Esta distribución es log-cóncava, y  $-\log p(\xi | p, a, b) \asymp \xi$  cuando  $\xi \rightarrow \infty$ . Utilizan una

propuesta de densidad construida intercalando una densidad uniforme entre dos distribuciones exponenciales apropiadas. La moda de la densidad (4.2) viene dada por

$$\xi' = \frac{1}{2} \left[ \log \left( 2(a) + \sqrt{(-2a)^2} \right) - \log(4b) \right].$$

Colocan la densidad uniforme en el intervalo  $(L, R)$  que es elegido de tal manera que  $L < \xi' < R$ , y luego posiciona las dos distribuciones exponenciales a cada lado de la moda para encontrar los puntos de corte  $L$  y  $R$  para los tres componentes. Primero, se define

$$\begin{aligned} l(\xi) &= -\log p(\xi \mid m, p, a, b) \\ &= \exp(-2\xi)m + (p - 2a)\xi + (a + b) \log(1 + \exp(2\xi)) \end{aligned}$$

y

$$g(\xi) = -2a + \frac{2(a + b)\exp(2\xi)}{1 + \exp(2\xi)} - 2\exp(-2\xi)m + p$$

como la derivada de  $l(\xi)$ . Luego definen

$$\xi_L = \xi' - \frac{0,85}{\sqrt{p}}, \quad \xi_R = \xi' + \frac{1,3}{\sqrt{p}}.$$

Estos puntos se utilizan para construir los dos componentes exponenciales de la densidad propuesta, los puntos de corte están dados por

$$L = \xi_L - \frac{l(\xi_L) - l(\xi')}{g(\xi_L)}, \quad R = \xi_R - \frac{l(\xi_R) - l(\xi')}{g(\xi_R)}$$

Entonces la densidad propuesta es dada por

$$q(\xi) \propto \begin{cases} \exp(-g(\xi_L)(\xi - L)) & \text{si } -\infty < \xi < L \\ 1 & \text{si } L < \xi < R \\ \exp(-g(\xi_R)(\xi - R)) & \text{si } R < \xi < \infty \end{cases}.$$

Muestrear  $q(\xi)$  es sencillo como la normalización para cada uno de los componentes es directa:  $K_L = -1/g(\xi_L)$ ,  $K_C = R - L$ , y  $K_R = 1/g(\xi_R)$ , donde  $K_L, K_C$  and  $K_R$  denotan los términos de normalización para la izquierda, central y derecha respectivamente, y setea  $K = K_L + K_C + K_R$ .

El algoritmo propuesto es el siguiente:

1. Primero muestrear

$$u_1 \sim U(0, 1), u_2 \sim U(0, 1), u_3 \sim U(0, 1).$$

2. Después revisar  $u_1$  : (a) Si  $u_1 \in (0, K_L/K)$  entonces

$$x \leftarrow -\frac{\log(1-u_2)}{g(\xi_L)} + L, \quad q \leftarrow l(\xi_L) + g(\xi_L)(x - \xi_L)$$

(b) Si  $u_1 \in (K_L/K, (K_L + K_C)/K)$  entonces

$$x \leftarrow (R - L)u_2 + L, \quad q \leftarrow l(\xi')$$

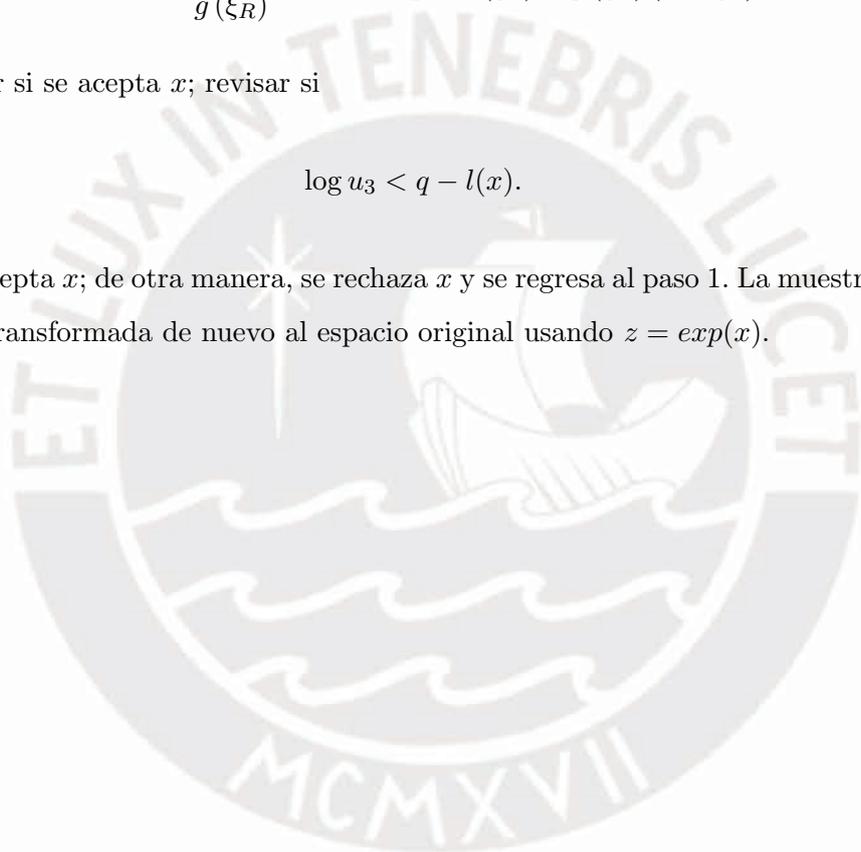
(c) Si  $u_1 \in ((K_L + K_C)/K, 1)$  entonces

$$x \leftarrow -\frac{\log(1-u_2)}{g(\xi_R)} + R, \quad q \leftarrow l(\xi_R) + g(\xi_R)(x - \xi_R)$$

3. Determinar si se acepta  $x$ ; revisar si

$$\log u_3 < q - l(x).$$

Si es así, se acepta  $x$ ; de otra manera, se rechaza  $x$  y se regresa al paso 1. La muestra aceptada  $x$  puede ser transformada de nuevo al espacio original usando  $z = \exp(x)$ .



# Capítulo 5

## Simulación

En esta sección se presenta un estudio de simulación para la estimación con regularización bayesiana de dos modelos de regresión lineal con MEG: un modelo de regresión lineal múltiple y un modelo con errores de Laplace.

Lo que se busca es comprobar si la estimación Horseshoe y Horseshoe + bayesiana en modelos de regresión lineal con mixtura de escala Gaussiana es capaz de recuperar los coeficientes de la regresión y comparar los resultados con los obtenidos por otros métodos de regularización bayesianos como el Lasso y también con un método sin regularización.

Para la regresión lineal múltiple, se plantean dos escenarios, en el primero se trabaja con cinco covariables no correlacionadas, y en el segundo igual número de covariables de las cuales dos covariables con coeficientes correspondientes distintos de cero, presentan alta correlación, es decir multicolinealidad. En ambos casos se tendrá una muestra de  $n = 100$  y se fija el número de muestras a posteriori a generar en 15000. Lo que se busca es observar también cómo se comportan los métodos frente a ambas situaciones, en especial en una situación de covariables correlacionadas y analizar los resultados en cada uno.

Para la regresión lineal con errores de Laplace, se plantea un escenario con  $n = 50$  y  $p = 5000$  variables y se fija el número de muestras a posteriori a generar en 1000. Lo que se busca es observar también cómo se comportan los métodos frente a la situación en que el número de variables predictoras es mucho mayor a la cantidad de datos disponibles.

### 5.1. Método para estimar los parámetros

Para la estimación de los coeficiente se usará el algoritmo de muestreo basado en el gradiente auxiliar m-Grad-1, el cual se encuentra implementado en el paquete bayesReg de R y tiene como base el método MCMC Metrópolis Hastings. Para la estimación del intercepto

$\beta_0$  se usará el procedimiento implementado en el paquete bayesReg de R, el cual propone valores para el intercepto desde una distribución normal específica, tal como se detalló en la sección 4.2.

Para la estimación de los hiperámetros se usará el algoritmo de muestreo Inversa Gamma-Inversa Gamma, el cual conduce a un muestreador Gibbs que es más eficiente considerando el indicador de muestras efectivas por minuto (Schmidt y Makalic, 2019). Esto se detalló en la subsección 4.3.1.

## 5.2. Criterios para evaluar la precisión de la simulación

Para medir qué tan bien el método logra recuperar los parámetros del modelo, se utiliza los siguientes indicadores: el sesgo, el sesgo relativo, y la raíz del error cuadrático medio (RMSE).

$$\text{Sesgo}(\hat{\theta}_j) = \frac{\sum_{i=1}^n \hat{\theta}_j^{(i)} - \theta_j}{n}$$

$$\text{SesgoRelativo}(\hat{\theta}_j) = \frac{\text{Sesgo}(\hat{\theta}_j)}{\theta_j} \times 100\%$$

$$\text{RMSE}(\hat{\theta}_j) = \sqrt{\frac{\sum_{i=1}^n (\hat{\theta}_j^{(i)} - \theta_j)^2}{n}}$$

donde  $\hat{\theta}_j^{(i)}$  es la estimación del parámetro  $\theta_j$  en la  $i$ -ésima muestra y  $n$  representa el número de muestras a posteriori generadas.

Para cada modelo se analiza también el error cuadrático de predicción esperado (MSE):

$$\text{MSE} = \sum_{j=1}^p \left( \frac{\sum_{i=1}^n \hat{\beta}_j^{(i)}}{n} - \beta_j \right)^2$$

## 5.3. Regresión lineal múltiple - covariables no correlacionados

Para simular un conjunto de datos para el modelo de regresión lineal Gaussiano con mixtura de escala (regresión lineal múltiple) con regularización bayesiana, se procedió de la siguiente manera:

- Tomando un tamaño de muestra igual a 100, se consideraron 5 covariables  $x_{ij}$  no correlacionadas. Las covariables fueron generadas de una distribución normal  $x_{ij} \sim N(0, 1)$ .

- Se fijó el valor para el intercepto  $\beta_0 = -1$
- Se fijaron valores los 5 coeficientes  $\beta_j$ , donde  $\beta_1 = 0$ ,  $\beta_2 = 0$ ,  $\beta_3 = 0$ ,  $\beta_4 = 1$  y  $\beta_5 = 1.3$ , los primeros tres coeficientes fueron fijados en cero ya que se comportarán como coeficientes espurios. Con esta definición se busca garantizar la existencia de coeficientes no significativos para el modelo, los cuales se espera que sean detectados por la estimación.
- Para los errores aleatorios, se considera  $\epsilon_i \sim N(0, 1)$ .

### 5.3.1. Estimación por mínimos cuadrados

En el cuadro 5.1, se muestran los parámetros estimados del modelo de regresión lineal múltiple por el método de mínimos cuadrados.

Parámetro	Valor verdadero	Valor estimado	Error estándar de estimación
$\beta_0$	-1	-0.99	0.0815
$\beta_1$	0	0.01	0.0875
$\beta_2$	0	0.03	0.0949
$\beta_3$	0	0.11	0.0758
$\beta_4$	1	1.08	0.0790
$\beta_5$	1.3	1.42	0.0794

Cuadro 5.1: Parámetros estimados por mínimos cuadrados en el estudio con datos simulados para covariables no correlacionadas.

Se obtuvo un valor de  $R^2$  ajustado de 0.8121, además los coeficientes significativos para el modelo fueron los de las covariables  $X_4$ ,  $X_5$  y el del intercepto  $\beta_0$ , que son los que generaron la variable dependiente  $Y$ .

### 5.3.2. Estimación usando una priori de Horseshoe, Horseshoe + y Lasso bayesiano

Al igual que antes los valores fijados para los 5 coeficientes  $\beta_j$ , son  $\beta_1 = 0$ ,  $\beta_2 = 0$ ,  $\beta_3 = 0$ ,  $\beta_4 = 1$  y  $\beta_5 = 1.3$

El cuadro 5.2 muestra los resultados para el estudio de simulación con las especificaciones descritas en las secciones anteriores y usando el paquete bayesReg de R con una priori de Horseshoe, Horseshoe + y Lasso Bayesiano. Se muestran los valores calculados del sesgo, sesgo relativo y la raíz del error cuadrático medio para todos los coeficientes en los tres métodos. Para los resultados mostrados se consideró un tamaño de muestra de 100. Para estimar los coeficientes la función generó 15000 muestras de la distribución a posteriori, usando un

burnin por defecto de 1000 y un nivel de thinning por defecto 5.

Método	Parámetro	Sesgo	Sesgo relativo	RMSE
Horseshoe	$\beta_0$	0.0142	-1.4285	0.0834
	$\beta_1$	0.0091		0.0613
	$\beta_2$	0.0108		0.0659
	$\beta_3$	0.0647		0.0941
	$\beta_4$	0.0673	6.7334	0.1047
	$\beta_5$	0.1134	8.7239	0.1389
Horseshoe +	$\beta_0$	0.1059	-10.5980	0.1535
	$\beta_1$	-0.0039		0.0736
	$\beta_2$	0.0011		0.0538
	$\beta_3$	0.0005		0.0690
	$\beta_4$	-0.0283	-2.8368	0.1096
	$\beta_5$	-0.2307	-17.7507	0.2535
Lasso Bayesiano	$\beta_0$	0.1041	-10.4120	0.1545
	$\beta_1$	-0.0122		0.1206
	$\beta_2$	0.0044		0.0890
	$\beta_3$	0.0071		0.1142
	$\beta_4$	-0.0496	-4.9659	0.1196
	$\beta_5$	-0.2534	-19.4935	0.2753

Cuadro 5.2: Parámetros estimados por modelo de regresión lineal Gaussiano con mixtura de escala (regresión lineal múltiple) con regularización bayesiana usando una priori de Horseshoe, Horseshoe + y Lasso Bayesiano en el estudio con datos simulados para covariables no correlacionadas,  $n=100$ .

Para los parámetros  $\beta_1$ ,  $\beta_2$  y  $\beta_3$  que eran los espurios, en los tres métodos, los sesgos son muy próximos a cero, lo que nos indica que han sido detectados correctamente.

Con relación a la covariable no espuria  $\beta_5$ . En los tres métodos, el estimador para  $\beta_5$  presenta el mayor sesgo comparado con el resto de coeficientes, lo que nos indica que el modelo no estima correctamente este coeficiente en los tres escenarios.

En el caso del estimador para el intercepto  $\beta_0$ , este presenta mayor sesgo, sesgo relativo y RMSE en los métodos Horseshoe + y Lasso Bayesiano, lo que indica que estos dos métodos tienen peor performance que con el estimador de Horseshoe.

Para todos los coeficientes, el método Lasso Bayesiano, presenta valores mayores de RMSE que los obtenidos usando la priori de Horseshoe y Horseshoe +.

Para los tres métodos, las covariables con mayor importancia asignada por el algoritmo son las covariables  $\beta_4$  y  $\beta_5$ , y solo ellas muestran un nivel de significancia del 95 %, es decir, el 95 % de los intervalos de credibilidad no incluyen el cero. Ninguna estimación de las otras covariables obtiene algún nivel de significancia, lo que indica que este modelo solo selecciona las covariables  $\beta_4$  y  $\beta_5$ , lo cual es correcto porque son las covariables no espurias del modelo.

Con el modelo de regresión clásico, se obtuvo un error cuadrático de predicción esperado (MSE) de 0.0376, con el estimador de Horseshoe se obtuvo 0.0220, con el estimador de Horseshoe + se obtuvo 0.0653 y con la estimación usando el método Lasso bayesiano se obtuvo 0.0777.

En conclusión el modelo MEG con la priori de Horseshoe, es el que mejor performance presenta en esta situación de covariables no correlacionadas, ya que presenta el menor valor de error cuadrático de predicción esperado y los valores de RMSE más bajos de los tres métodos, además de que detecta correctamente las covariables no espurias.

## 5.4. Regresión lineal múltiple - Covariables correlacionados

En este caso se procedió como en el caso anterior; pero generando la quinta covariable como

$$x_{i5} = x_{i4} + 0.13 \times z_i$$

donde  $z_i \sim N(0, 1)$ . La correlación entre ambas variables fue de 0.9916.

### 5.4.1. Estimación por mínimos cuadrados

En el cuadro 5.3, se muestran los parámetros estimados del modelo de regresión lineal múltiple por el método de mínimos cuadrados.

Parámetro	Valor verdadero	Valor estimado	Desviación estándar	IC al 95 %
$\beta_0$	-1	-1.0535	0.0997	1.2515 -0.8555
$\beta_1$	0	0.1438	0.1067	-0.0681 0.3557
$\beta_2$	0	-0.1481	0.0973	-0.3414 0.0452
$\beta_3$	0	-0.0865	0.1068	-0.2986 0.1255
$\beta_4$	1	2.2644	0.7642	0.7469 3.7819
$\beta_5$	1.3	0.1031	0.7482	-1.3825 1.5887

Cuadro 5.3: Parámetros estimados por mínimos cuadrados en el estudio con datos simulados para covariables correlacionadas.

Se obtuvo un valor de  $R^2$  ajustado de 0.8382, además solo el coeficiente  $X_4$  y el intercepto  $\beta_0$  resultaron significativos al 95 %. La covariable  $X_5$  resultó no significativa para el modelo, a pesar que fijamos un valor de 1.3 para este coeficiente en la generación de la variable dependiente  $\mathbf{Y}$ .

### 5.4.2. Estimación usando una priori de Horseshoe, Horseshoe + y Lasso Bayesiano

El siguiente cuadro 5.4 muestra los resultados para el estudio de simulación con las especificaciones descritas en el escenario de variables correlacionadas y usando el paquete bayesReg de R con una priori de Horseshoe, Horseshoe + y Lasso Bayesiano. Se muestran el sesgo, sesgo relativo y la raíz del error cuadrático medio para los tres métodos. Para los resultados mostrados se consideró un tamaño de muestra de 100, para estimar los coeficientes la función generó 15000 muestras de la distribución a posteriori, usando un burnin por defecto de 1000 y un nivel de thinning por defecto en 5.

Método	Parámetro	Sesgo	Sesgo relativo	RMSE
Horseshoe	$\beta_0$	0.0284	-2.8446	0.1089
	$\beta_1$	-0.0690		0.1070
	$\beta_2$	0.0860		0.1335
	$\beta_3$	-0.0248		0.0826
	$\beta_4$	1.0098	100.9883	1.2085
	$\beta_5$	-1.0082	-77.5557	1.1942
Horseshoe +	$\beta_0$	0.0256	-2.5689	0.1075
	$\beta_1$	-0.0550		0.0951
	$\beta_2$	0.0701		0.1200
	$\beta_3$	-0.0182		0.0706
	$\beta_4$	1.0143	101.4390	1.2261
	$\beta_5$	-1.0075	-77.5026	1.2074
Lasso Bayesiano	$\beta_0$	0.0433	-4.3336	0.1152
	$\beta_1$	-0.1171		0.1485
	$\beta_2$	0.1487		0.1872
	$\beta_3$	-0.0532		0.1242
	$\beta_4$	0.6372	63.7267	0.8753
	$\beta_5$	-0.6678	-51.3716	0.8867

Cuadro 5.4: Parámetros estimados por modelo de regresión lineal Gaussiano con mixtura de escala (regresión lineal múltiple) con regularización bayesiana usando una priori de Horseshoe, Horseshoe + y Lasso Bayesiano en el estudio con datos simulados para covariables correlacionadas,  $n=100$ .

En los tres métodos, con relación al sesgo, los estimadores de los parámetros de regresión presentan sesgos relativamente bajos a excepción de  $\beta_4$  y  $\beta_5$ . Lo que indica que los métodos no estiman correctamente los coeficientes no espurios cuando corresponden a las covariables correlacionadas.

En la estimación de Horseshoe y Horseshoe +, los parámetros  $\beta_1$ ,  $\beta_2$  y  $\beta_3$  que eran los espurios, los sesgos son muy próximos a cero, lo que nos indica que han sido detectados correctamente por la estimación. En cambio en la estimación de Lasso Bayesiano, solo el

sesgo asociado al estimador de  $\beta_3$  es muy próximo a cero, lo que nos indica que la estimación Lasso Bayesiana no detecta correctamente los coeficientes espurios.

En la estimación de Lasso Bayesiano, los estimadores de los coeficientes de regresión presentan sesgos mayores que los otros métodos, a excepción de los correspondientes a  $\beta_4$  y  $\beta_5$ . Es decir para los coeficientes no espurios presenta un sesgo menor, lo que nos indica que este método, en esta situación estima mejor los valores de las covariables no espurias.

En los tres métodos, respecto al RMSE se observa que los estimadores para los parámetros  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$  y  $\beta_3$  presentan valores aceptables, en general son relativamente bajos, a diferencia de los estimadores de  $\beta_4$  y  $\beta_5$ , que presentan valores mayores.

Cabe resaltar, que respecto al RMSE se observa que en la estimación de Lasso Bayesiana, presenta valores mayores a los obtenidos usando los otros dos métodos, a excepción de los asociados a los estimadores de los parámetros  $\beta_4$  y  $\beta_5$ , que presenta valores de RMSE menores.

En la estimación de Horseshoe y Horseshoe +, la covariable con mayor importancia asignada por el algoritmo es la covariable  $\beta_4$  y solo ella muestra un nivel de significancia, del 95 % en la estimación de Horseshoe y en la estimación de Horseshoe +. Ninguna estimación de las otras covariables obtiene algún nivel de significancia, lo que indica que estos dos métodos solo seleccionan la covariable  $\beta_4$ .

En cambio, en la estimación de Lasso Bayesiana, la covariables con mayor importancia asignada por el algoritmo son  $\beta_4$ ,  $\beta_5$ ,  $\beta_1$  y  $\beta_2$ . La covariable  $\beta_3$  no presenta nivel de significancia, lo que indica que este modelo solo descarta la covariable  $\beta_3$ .

En este escenario, con el modelo de regresión clásico, se obtuvo un error cuadrático de predicción esperado (MSE) de 2.4288, con el estimador de Horseshoe se obtuvo 2.0480, con el estimador de Horseshoe + se obtuvo 2.05196 y con la estimación usando el método Lasso bayesiano se obtuvo 0.1152 . Entonces el estimador de Lasso Bayesiano obtuvo el menor valor de error cuadrático de predicción esperado de los tres métodos

## 5.5. Regresión lineal con errores de Laplace

En este escenario se considera un conjunto de datos de tamaño  $n = 50$  y un gran número de covariables  $p = 50000$  no correlacionadas. Los datos a generar tienen distribución con errores de Laplace.

Para simular este conjunto de datos para el modelo de regresión lineal Gaussiano con mixtura de escala con errores de Laplace con regularización bayesiana, en este escenario, se procedió de la siguiente manera:

- Tomando un tamaño de muestra igual a 50, se consideraron 50000 covariables  $x_{ij}$ . Las covariables fueron generadas de una distribución normal con media 0 y varianza 1.
- Se fijó el valor para el intercepto  $\beta_0 = 0$
- Se fijaron valores para los 5 primeros coeficientes  $\beta_j$ , donde  $\beta_1 = 5$ ,  $\beta_2 = 5$ ,  $\beta_3 = 1$ ,  $\beta_4 = 1$  y  $\beta_5 = 1$ , todos los demás fueron fijados en cero, ya que se comportarán como los coeficientes espurios. Con esta definición se busca garantizar la existencia de coeficientes no significativos para el modelo, los cuales se espera que sean detectados por la estimación.
- Se fija el ratio señal-ruido, que es la relación entre la varianza de los datos y la varianza de los errores, en 8.
- Se calcula el valor de la varianza de los residuos. Y con este valor se calcula el parámetro de escala  $b$  de la distribución Laplace.
- Se fija la media  $\mu$  de la distribución Laplace en cero.
- Se simulan los valores de los errores aleatorios de una distribución de Laplace con media  $\mu$  y parámetro de escala  $b$ .

### 5.5.1. Estimación usando una priori de Horseshoe, Horseshoe + y Lasso Bayesiano

El cuadro 5.5 muestra los resultados para el estudio de simulación con las especificaciones descritas en el escenario de 5000 variables predictoras y tamaño de datos  $n = 50$  y errores de Laplace, usando el paquete bayesReg de R con una priori de Horseshoe, Horseshoe + y Lasso Bayesiano. Se muestran el sesgo, sesgo relativo y la raíz del error cuadrático medio para los tres métodos. Para los resultados mostrados se generó 1000 muestras de la distribución a posteriori, usando un burnin por defecto de 1000 y un nivel de thinning por defecto 5.

En los tres métodos, con relación al sesgo, los estimadores de los parámetros de regresión presentan valores de sesgo bajos para los coeficientes espurios  $\beta_6, \dots, \beta_{5000}$ , lo que indica que estos métodos detectan bien los coeficientes espurios en este escenario  $p > n$ .

En cuanto a los estimadores de los coeficientes  $\beta_1, \dots, \beta_5$ , los métodos de Horseshoe y Horseshoe+, presentan menor sesgo en la estimación, que el método Lasso, pero cabe resaltar que las estimaciones que presentan mayor sesgo son las de  $\beta_3, \beta_4$  y  $\beta_5$ . En cambio en el método de Lasso bayesiano, los estimadores que presentan mayor sesgo son los estimadores de  $\beta_1$  y  $\beta_2$ . En cuanto al intercepto, en los tres métodos vemos que el sesgo, en valor absoluto, es

mayor que los sesgos de los otros coeficientes espurios, lo que nos indica que a pesar de que el intercepto se fijó en cero, este no es del todo bien estimado por los tres métodos.

En la estimación de Lasso Bayesiano, los estimadores de los coeficientes de regresión no espurios  $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$  presentan sesgos mayores que los otros métodos, esto nos indica que este método, en esta situación no estima bien los valores de las covariables no espurias.

En los tres métodos, respecto al RMSE se observa que los estimadores para los coeficientes espurios  $\beta_6, \dots, \beta_{5000}$  presentan valores relativamente bajos, lo que indica que estos métodos detectan bien los coeficientes espurios en este escenario  $p > n$ .

En cuanto a los estimadores de los coeficientes  $\beta_1, \dots, \beta_5$ , en los métodos de Horseshoe y Horseshoe+, los valores de RMSE son menores que los de Lasso bayesiano, lo que nos indica que este método, en este escenario  $p > n$ , no estima bien los valores de las covariables no espurias. Además en este método los estimadores de los coeficiente  $\beta_1$  y  $\beta_2$  y del intercepto presentan mayor RMSE que los otros dos métodos.

En cuanto al intercepto, en los tres métodos vemos que el RMSE es mayor que el de los otros coeficientes espurios, lo que nos indica que a pesar de que el intercepto se fijó en cero, este no es bien estimado por los tres métodos.

En general, respecto al RMSE se observa que en la estimación de Lasso Bayesiano, presenta valores mayores que los obtenidos usando los otros dos métodos.

En la estimación de Horseshoe y Horseshoe + las covariables con mayor importancia asignada por el algoritmo son  $\beta_1$  y  $\beta_2$ . Ninguna estimación de las otras covariables obtiene algún nivel de significancia, lo que indica que estos dos métodos solo seleccionan las covariables  $\beta_1$  y  $\beta_2$ . En la estimación de Lasso Bayesiano ninguna covariable obtiene algún nivel de significancia, lo que indica que descarta todas para este escenario.

En este escenario, con el estimador de Horseshoe se obtuvo un error cuadrático de predicción esperado (MSE) de 3.8221, con el estimador de Horseshoe + se obtuvo 7.0338 y con la estimación usando el método Lasso bayesiano se obtuvo 54.6717. Entonces el estimador de Horseshoe obtuvo el menor valor de error cuadrático de predicción esperado de los tres métodos.

Todos los métodos en los escenarios descritos han convergido.

Método	Parámetro	Sesgo	Sesgo relativo	RMSE
Horseshoe	$\beta_0$	-0.5243		0.7529
	$\beta_1$	-0.6242	-12.4842	1.0126
	$\beta_2$	-0.0424	-0.8488	0.4442
	$\beta_3$	-0.9980	-99.8037	0.9989
	$\beta_4$	-0.9771	-97.7140	0.9867
	$\beta_5$	-0.9693	-96.9340	0.9855
	$\beta_6$	-0.0004		0.0217
	$\beta_7$	-0.0192		0.1322
	$\beta_8$	-0.0101		0.0833
	$\beta_9$	-0.0004		0.0226
	...	...	...	...
$\beta_{5000}$	-0.0022		0.0349	
Horseshoe +	$\beta_0$	-0.8050		1.0002
	$\beta_1$	-0.6518	-13.0373	1.0850
	$\beta_2$	-0.4669	-9.3389	1.2792
	$\beta_3$	-0.9965	-99.6576	0.9972
	$\beta_4$	-0.9777	-97.7713	0.9850
	$\beta_5$	-0.9957	-99.5772	0.9967
	$\beta_6$	-0.0007		0.0180
	$\beta_7$	-0.0045		0.0415
	$\beta_8$	-0.0112		0.0823
	$\beta_9$	0.0015		0.0243
	...	...	...	...
$\beta_{5000}$	-0.0026		0.0369	
Lasso Bayesiano	$\beta_0$	-1.4574		1.7795
	$\beta_1$	-4.9512	-99.0255	4.9523
	$\beta_2$	-4.9477	-98.9547	4.9492
	$\beta_3$	-1.0017	-100.1730	1.0086
	$\beta_4$	-0.9935	-99.3551	0.9979
	$\beta_5$	-1.0030	-100.3033	1.0102
	$\beta_6$	0.0026		0.0964
	$\beta_7$	0.0023		0.1305
	$\beta_8$	-0.0031		0.1155
	$\beta_9$	-0.0017		0.0921
	...	...	...	...
$\beta_{5000}$	0.0006	-51.3716	0.1018	

Cuadro 5.5: Parámetros estimados por modelo de regresión lineal con MEG con errores de Laplace usando una priori de Horseshoe, Horseshoe + y Lasso Bayesiano en el escenario de  $p > n$ .

# Capítulo 6

## Aplicaciones

### 6.1. Aplicación a los datos BostonHousing

En esta primera parte se desarrolla una aplicación del modelo regresión lineal Gaussiano con mixtura de escala con regularización bayesiana para respuestas continuas.

Para esto usamos la base de datos “BostonHousing”, que son datos de viviendas en 506 zonas censales en el Área Estadística Metropolitana Estándar de Boston (SMSA) del censo de 1970. Los datos, tomados del estudio de Harrison y Rubinfeld (1978), se encuentran en el paquete de R “mlbench”.

La variables de interés es “medv” que denota al valor medio de las viviendas ocupadas por sus propietarios en miles de dólares. La base de datos contiene también otras 13 covariables, como 2 variables de atributos estructurales, ocho variables sobre el vecindario, dos variables de accesibilidad y una variable de contaminación del aire. La variable de contaminación utilizada es la concentración de óxidos de nitrógeno (NOX).

El objetivo es estudiar la relación entre la variable de respuesta y estas covariables, e identificar cuáles tienen efectos significativos y cuáles no en el modelo, es decir cuáles podrían ser contraídas para lograr la esparcidad.

#### 6.1.1. Descripción de los datos

Las covariables consideradas son:

Covariables relacionadas con la estructura de las viviendas

- *rm* : Promedio de número de cuartos por vivienda.
- *age* : Proporción de unidades ocupadas por sus propietarios construídas antes de 1940.

Covariables relacionadas con el vecindario

- $b$  : Igual a  $1000(B - 0.63)^2$ , donde  $B$  es la proporción de personas de raza negra por localidad.
- $lstat$  : Proporción de la población que pertenece al estatus más bajo.
- $crim$  : Tasa de criminalidad per cápita por localidad.
- $zn$  : Proporción de terreno residencial zonificado para lotes de más de 25,000 pies cuadrados.
- $indus$  : Proporción de acres de negocios no minoristas por ciudad.
- $tax$  : Tasa del impuesto sobre el valor total de la propiedad (\$/\$10,000).
- $ptratio$  : Ratio alumnos-maestro por localidad.
- $chas$  : Variable dummy (= 1 si la zona limita con el río Charles; 0 en caso contrario).

Covariables relacionadas con la accesibilidad

- $dis$  : Media ponderada de distancias a cinco centros de empleo de Boston.
- $rad$  : índice de accesibilidad a carreteras radiales.

Covariables relacionadas con la contaminación

- $nox$ : concentración de óxidos de nitrógeno en ppm (concentración promedio anual en partes por cien millones).

### 6.1.2. Análisis descriptivo

La variable de interés es el valor medio de las viviendas ocupadas por sus propietarios en miles de dólares, la media de esta variable es 22.53 y la mediana 21.20

Para observar la distribución de la variable respuesta de forma gráfica se tiene su histograma, en el cual se puede distinguir la forma de la distribución de la variable valor medio de las viviendas ocupadas por sus propietarios en miles de dólares, vemos que una gran cantidad

de datos se concentran en el rango entre 15 y 25 mil dólares.

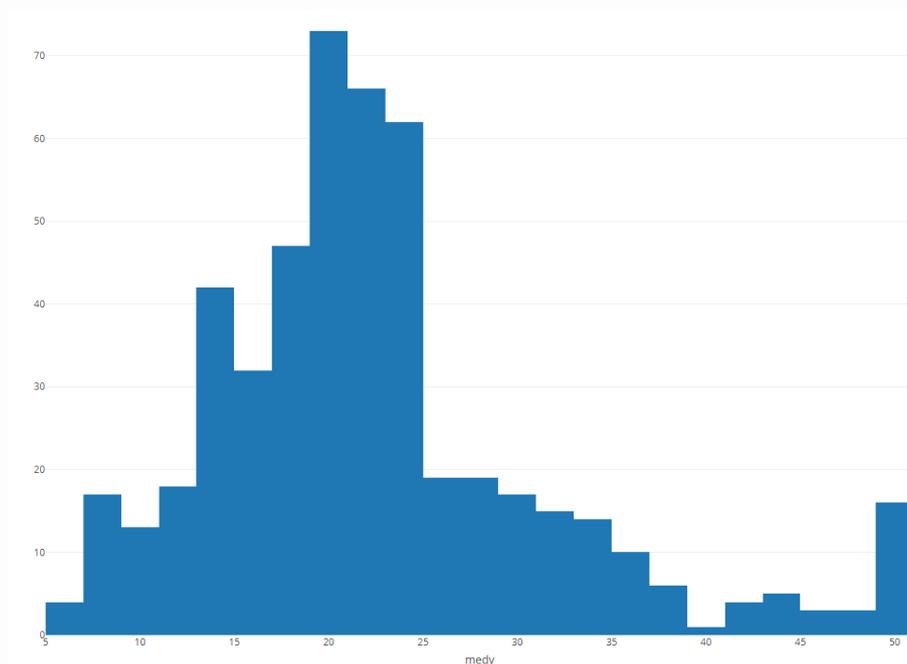


Figura 6.1: Histograma de la distribución del valor medio de las viviendas ocupadas por sus propietarios en miles de dólares

Asimismo se observa en el histograma anterior que el valor medio de las viviendas ocupadas por sus propietarios en miles de dólares (medv), no parece tener una distribución normal, y tiene una marcada cola a la derecha, por lo que es posible que, al hacer la regresión clásica, no se cumpla con el supuesto de normalidad. Esto se verifica en el qqplot que se muestra en los apéndices.

Las correlaciones entre las covariables y la variable respuesta se pueden resumir a través



negativa con “*crim*” tasa de criminalidad per cápita por localidad y “*age*” proporción de unidades ocupadas por sus propietarios construidas antes de 1940.

Por otro lado, analizando el valor medio de las viviendas ocupadas por sus propietarios en miles de dólares según si la zona limita o no con el río, se revisa el siguiente boxplot en el que se observa que las viviendas que limitan con el río presentan un mayor valor medio en miles de dólares. Además los valores medios de las viviendas que limitan con el río presentan mayor dispersión y asimetría. Adicionalmente los valores medios de las viviendas que no limitan con el río presentan una gran cantidad de outliers, es decir existen viviendas, en este grupo, cuyos valores medios son muy superiores.

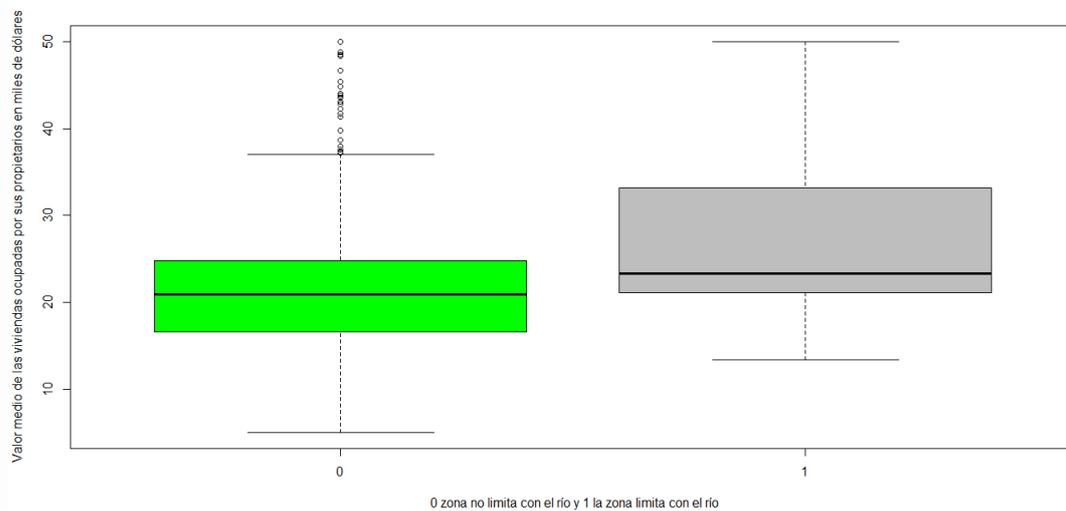


Figura 6.3: Valor medio de las viviendas ocupadas por sus propietarios en miles de dólares según si la zona limita o no con el río

### 6.1.3. Especificación del modelo

En el modelo de regresión lineal Gaussiano con MEG asumiremos que la variable respuesta  $y_i$ , que representa el valor medio de las viviendas ocupadas por sus propietarios para la zona censal (localidad)  $i$ , satisface

$$y_i \mid \mathbf{x}_i, \boldsymbol{\beta}, \beta_0, \omega_i, \sigma^2 \sim \mathcal{N}(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}, \sigma^2),$$

donde  $\beta_0$  es un intercepto,  $\boldsymbol{\beta}$  son los coeficientes de regresión y  $\mathbf{x}_i$  es el vector de variables predictoras.

Harrison y Rubinfeld (1978) propone la siguiente especificación que usaremos en esta

aplicación:

$$\log(\text{medv}_i) = \beta_0 + \beta_1 \text{rm}_i^2 + \beta_2 \text{age}_i + \beta_3 \log(\text{dis})_i + \beta_4 \log(\text{rad})_i + \beta_5 \text{tax}_i + \beta_6 \text{ptratio}_i \\ + \beta_7 1000(B - 0 \cdot 63)_i^2 + \beta_8 \log(\text{lstat})_i + \beta_9 \text{crim}_i + \beta_{10} \text{zn}_i + \beta_{11} \text{indus}_i + \beta_{12} \text{chas}_i + \beta_{13} \text{nox}_i + \epsilon$$

donde identificaremos a las covariables para una localidad  $i$  por:  $\mathbf{x}_{i1}$  como el promedio del número de cuartos por vivienda al cuadrado ( $\text{rm}^2$ ),  $\mathbf{x}_{i2}$  como la proporción de unidades ocupadas por sus propietarios construídas antes de 1940 ( $\text{age}$ ),  $\mathbf{x}_{i3}$  como el logaritmo de la media ponderada de distancias a cinco centros de empleo de Boston ( $\log(\text{dis})$ ),  $\mathbf{x}_{i4}$  como el logaritmo del índice de accesibilidad a carreteras radiales ( $\log(\text{rad})$ ),  $\mathbf{x}_{i5}$  como la tasa del impuesto sobre el valor total de la propiedad ( $\text{tax}$ ),  $\mathbf{x}_{i6}$  como el ratio alumnos-maestro por localidad ( $\text{ptratio}$ ),  $\mathbf{x}_{i7}$  es igual a  $1000(B - 0 \cdot 63)^2$ , donde  $B$  es la proporción de personas de raza negra por localidad,  $\mathbf{x}_{i8}$  como el porcentaje del estatus más bajo de la población ( $\text{lstat}$ ),  $\mathbf{x}_{i9}$  como la tasa de criminalidad per cápita ( $\text{crim}$ ),  $\mathbf{x}_{i10}$  como la proporción de terreno residencial zonificado para lotes de más de 25,000 pies cuadrados ( $\text{zn}$ ),  $\mathbf{x}_{i11}$  como la proporción de acres de negocios no minoristas por ciudad ( $\text{indus}$ ),  $\mathbf{x}_{i12}$  como la variable dummy (= 1 si la zona limita con el río Charles; 0 en caso contrario) ( $\text{chas}$ ) y finalmente  $\mathbf{x}_{i13}$  como la concentración de óxidos de nitrógeno en ppm ( $\text{nox}$ ).

#### 6.1.4. Métodos de aplicación

Se ajustará este modelo aplicando regresión clásica, los métodos de regularización bayesiana con priori de Horseshoe, Horseshoe + y el estimador Lasso Bayesiano.

El objetivo es ajustar un modelo de regresión lineal Gaussiano con MEG para la respectiva variable dependiente, donde se asume una distribución normal para los errores.

Todos los métodos se implementaron en el software R, usando la función “bayesreg” del paquete “bayesreg” de R (Makalic y Schmidt, 2016a). Para los métodos bayesianos se ejecutaron 15000 iteraciones, de las cuales se descartaron las primeras 2000.

#### 6.1.5. Resultados

Usando la data BostonHousing se procedió a estimar los parámetros del modelo. Primero aplicando regresión múltiple clásica (errores Gaussianos), luego se aplicaron los métodos de regularización bayesiana, estimador de Horseshoe, de Horseshoe + y estimador Lasso Bayesiano en el modelo de mixtura de escala con errores de Laplace.

Para la regresión clásica se muestran los intervalos de confianza. Además se aplicó regresión “stepwise”, que seleccionó las mismas covariables mostradas en el siguiente cuadro.

De las 13 covariables, descartó solo tres: “*zn*” que es la proporción de terreno residencial zonificado para lotes de más de 25,000 pies cuadrados, “*indus*” que es la proporción de acres de negocios no minoristas por ciudad y “*age*” que es la proporción de unidades ocupadas por sus propietarios construidas antes de 1940.

Covariable	Parámetro	$\hat{\beta}$	Error Estándar	IC (95 %)	p-value
Intercepto	$\beta_0$	4.799	0.1770	(4.4510; 5.1464)	< 2e-16
<i>rm</i> <sup>2</sup>	$\beta_1$	0.0064	0.0013	(0.0038; 0.0090)	1.21e-06
<i>age</i>	$\beta_2$	0.0002	0.0005	(-0.0008; 0.0012)	0.6798
<i>log(dis)</i>	$\beta_3$	-0.2006	0.0341	(-0.2676; -0.1335)	7.54e-09
<i>log(rad)</i>	$\beta_4$	0.0976	0.0192	(0.0598; 0.1353)	5.25e-07
<i>tax</i>	$\beta_5$	-0.0004	0.0001	(-0.0006; -0.0001)	0.0007
<i>ptratio</i>	$\beta_6$	-0.0302	0.0049	(-0.0399; -0.0204)	2.52e-09
<i>b</i>	$\beta_7$	0.0003	0.0001	(0.0001 0.0005)	0.0004
<i>lstat</i>	$\beta_8$	-0.3715	0.0250	(-0.4207; -0.3223)	< 2e-16
<i>crim</i>	$\beta_9$	-0.0119	0.0012	(-0.0143; -0.0094)	< 2e-16
<i>zn</i>	$\beta_{10}$	-0.00006	0.0005	(-0.0010; 0.0009)	0.8981
<i>indus</i>	$\beta_{11}$	0.0002	0.0023	(-0.0043; 0.0049)	0.9060
<i>chas1</i>	$\beta_{12}$	0.0880	0.0332	(0.0228; 0.1533)	0.0082
<i>nox</i>	$\beta_{13}$	-0.8455	0.1524	(-1.1450; -0.5460)	4.75e-08

Cuadro 6.1: Parámetros estimados por modelo de regresión lineal con mixtura de escala Gaussiana: regresión lineal múltiple clásica

La regresión clásica muestra un coeficiente de determinación de 0.80.

En cuanto a los métodos bayesianos, dado que la variable respuesta muestra una cola a la derecha relativamente pesada, hemos considerado un modelo de regresión lineal con MEG de errores de Laplace, que permiten modelar mejor los datos con valores atípicos, como muestran estos datos. Para este modelo aplicaremos los métodos de regularización bayesiana, estimador de Horseshoe, de Horseshoe + y estimador Lasso Bayesiano. La convergencia del proceso bayesiano se evaluó observando el comportamiento de las cadenas para cada parámetro y el Rhat, los cuales junto a las autocorrelaciones mostraron evidencias de convergencia.

Las distribuciones a posteriori y las cadenas se presentan en el apéndice B.

En los cuadros siguientes se muestra el detalle de la estimación aplicando los métodos bayesianos señalados (Horseshoe, Horseshoe + y Lasso bayesiano), se aprecian las estimaciones de los parámetros de regresión, los errores estándar, los cuantiles de los parámetros que definen los intervalos de credibilidad y el Rhat de los coeficientes.

Covariable	Parámetro	$\hat{\beta}$	Error Estándar	IC (95 %)	Rhat
Intercepto	$\beta_0$	3.8578	0.1622	(3.5524; 4.1897)	1.0000
$rm^2$	$\beta_1$	0.0142	0.0012	(0.0117; 0.0166)	1.0002
$age$	$\beta_2$	-0.0005	0.0004	(-0.0014; 0.0001)	1.0000
$\log(dis)$	$\beta_3$	-0.1431	0.0272	(-0.1986; -0.0917)	1.0000
$\log(rad)$	$\beta_4$	0.0470	0.0159	(0.0170; 0.0790)	0.9999
$tax$	$\beta_5$	-0.0003	0.00009	(-0.0005; -0.0001)	1.0000
$ptratio$	$\beta_6$	-0.0259	0.0035	(-0.0330; -0.0190)	1.0000
$b$	$\beta_7$	0.0005	0.0001	(0.0003; 0.0007)	1.0001
$lstat$	$\beta_8$	-0.2429	0.0243	(-0.2902; -0.1953)	0.9999
$crim$	$\beta_9$	-0.0105	0.0018	(-0.0144; -0.0074)	0.9999
$zn$	$\beta_{10}$	0.0001	0.0003	(-0.0004; 0.0008)	1.0001
$indus$	$\beta_{11}$	0.0009	0.0013	(-0.0014; 0.0039)	0.9999
$chas1$	$\beta_{12}$	0.0565	0.0251	(0.0074; 0.1062)	1.0000
$nox$	$\beta_{13}$	-0.5181	0.1299	(-0.7722; -0.2647)	0.9999

Cuadro 6.2: Parámetros estimados por modelo de regresión lineal con mixtura de escala Gaussiana con errores de Laplace con regularización bayesiana usando priori de Horseshoe

Covariable	Parámetro	$\hat{\beta}$	Error Estándar	IC (95 %)	Rhat
Intercepto	$\beta_0$	3.8670	0.1649	(3.5501; 4.1988)	1.0002
$rm^2$	$\beta_1$	0.0142	0.0012	(0.0117; 0.0165)	0.9999
$age$	$\beta_2$	-0.0004	0.0004	(-0.0014; 0.0001)	1.0000
$\log(dis)$	$\beta_3$	-0.1415	0.0267	(-0.1952; -0.0910)	0.9999
$\log(rad)$	$\beta_4$	0.0454	0.0162	(0.0137; 0.0778)	0.9999
$tax$	$\beta_5$	-0.0003	0.0001	(-0.0005; -0.0001)	0.9999
$ptratio$	$\beta_6$	-0.0260	0.0035	(-0.0330; -0.0191)	1.0002
$b$	$\beta_7$	0.0005	0.0001	(0.0003; 0.0007)	1.0000
$lstat$	$\beta_8$	-0.2457	0.0244	(-0.2923; -0.1966)	1.0002
$crim$	$\beta_9$	-0.0104	0.0018	(-0.0143; -0.0073)	1.0006
$zn$	$\beta_{10}$	0.0001	0.0002	(-0.0003; 0.0007)	0.9999
$indus$	$\beta_{11}$	0.0007	0.0012	(-0.0013; 0.0036)	1.0001
$chas1$	$\beta_{12}$	0.0524	0.0263	(0.0010; 0.1038)	0.9999
$nox$	$\beta_{13}$	-0.5177	0.1301	(-0.7678; -0.2581)	0.9999

Cuadro 6.3: Parámetros estimados por modelo de regresión lineal con mixtura de escala Gaussiana con errores de Laplace con regularización bayesiana usando priori de Horseshoe +

Covariable	Parámetro	$\hat{\beta}$	Error Estándar	IC (95 %)	Rhat
Intercepto	$\beta_0$	3.8548	0.1627	(3.5480; 4.1814)	0.9999
$rm^2$	$\beta_1$	0.0143	0.0012	(0.0118; 0.0165)	0.9999
<i>age</i>	$\beta_2$	-0.0006	0.0004	(-0.0015; 0.0001)	1.0000
$\log(dis)$	$\beta_3$	-0.1461	0.0270	(-0.2013; -0.0944)	0.9999
$\log(rad)$	$\beta_4$	0.0502	0.0155	(0.0208; 0.0812)	0.9999
<i>tax</i>	$\beta_5$	-0.0003	0.00009	(-0.0005; -0.0001)	1.0002
<i>ptratio</i>	$\beta_6$	-0.0258	0.0035	(-0.0330; -0.0191)	0.9999
<i>b</i>	$\beta_7$	0.0005	0.0001	(0.0003; 0.0007)	1.0007
<i>lstat</i>	$\beta_8$	-0.2393	0.0244	(-0.2872; -0.1926)	0.9999
<i>crim</i>	$\beta_9$	-0.0105	0.0018	(-0.0143; -0.0074)	1.0000
<i>zn</i>	$\beta_{10}$	0.0002	0.0003	(-0.0004; 0.0009)	0.9999
<i>indus</i>	$\beta_{11}$	0.0013	0.0015	(-0.0016; 0.0043)	1.0003
<i>chas1</i>	$\beta_{12}$	0.0605	0.0242	(0.0144; 0.1090)	1.0002
<i>nox</i>	$\beta_{13}$	-0.5207	0.1264	(-0.7681; -0.2739)	1.0000

Cuadro 6.4: Parámetros estimados por modelo de regresión lineal con mixtura de escala Gaussiana con errores de Laplace con regularización bayesiana usando Lasso bayesiano

El método bayesiano que aplica el estimador de Horseshoe + selecciona las mismas 10 covariables que la regresión clásica. En cuanto a los métodos bayesianos de Horseshoe y Lasso, sin embargo seleccionaron 11 covariables significativas. Estos dos métodos bayesianos añaden en la selección a la covariable *age* que es la proporción de unidades ocupadas por sus propietarios construídas antes de 1940. Los tres métodos bayesianos muestran errores estándar relativamente pequeños.

Finalmente, analizando la precisión predictiva esperada, según el criterio WAIC, el modelo con el estimador Lasso bayesiano muestra un ( $WAIC = -183,57$ ), el modelo aplicando la priori de Horseshoe muestra un ( $WAIC = -183,19$ ) y el modelo aplicando la priori de Horseshoe + muestra un ( $WAIC = -182,9$ ). Por lo que en esta aplicación el modelo de regresión lineal con MEG de errores de Laplace y regularización bayesiana muestra ligeramente mejores resultados para el estimador Lasso bayesiano, seguido por el que aplica el estimador de Horseshoe. Cabe resaltar que los métodos Lasso y de Horseshoe seleccionan una variable adicional que la regresión clásica *age*, como covariable de importancia, y además con signo negativo, en comparación con la regresión clásica que le asigna signo positivo aunque no es variable significativa. Por lo que entendemos que estos dos métodos bayesianos consideran que la antigüedad de las viviendas afecta negativamente al valor de la vivienda.

Cabe resaltar que se aplicaron también los tres métodos para el modelo con errores  $t$  de Student, resultando que se obtienen similares resultados, pero con un mayor WAIC.

Adicionalmente se aplicaron previamente los mismos métodos pero para modelos con errores Gaussianos, resultando que los modelos con errores de Laplace y  $t$  de Student presentaron un mejor ajuste que los modelos con errores Gaussianos, esto probablemente porque la variable respuesta muestra una marcada asimetría positiva y colas pesadas.

## 6.2. Aplicación a un estudio de deserción universitaria

En esta sección se desarrollará una aplicación con datos reales para los modelos de regresión MEG bajo regularización bayesiana. El objetivo es estimar la variable retiro intersemestral de los alumnos ingresantes matriculados a una facultad de una universidad privada del país, al término de su primer semestre de estudios. Para ello se usará una base de datos que comprende información de los alumnos ingresantes a esta facultad en los ciclos 2018-2, 2019-1, 2021-2, 2022-1, 2022-2 y 2023-1 que se matricularon en el mismo ciclo en el que fueron admitidos y que además, nunca antes se habían matriculado en la universidad.

En este conjunto de datos hemos aplicado el modelo de regresión logística binaria con mixtura de escala Gaussiana y otros modelos para realizar comparaciones entre ellos.

### 6.2.1. Justificación de la aplicación

La universidad está interesada en implementar estrategias para reducir la deserción estudiantil temprana en sus estudiantes, por lo que es importante analizar a los estudiantes ingresantes, en este caso el estudio se centra en la estimación del retiro intersemestral de estos alumnos, es decir estimar la probabilidad de que no se matriculen en el siguiente semestre considerando un conjunto de variables que se sospecha están relacionadas con la variable objetivo. Este modelo ayudará a realizar esa estimación, lo que proveerá de una herramienta útil para la universidad, con el fin de identificar a los alumnos que no se matricularán en el semestre inmediato y así aplicar con ellos estrategias de retención.

### 6.2.2. Descripción de datos

Los datos para la aplicación corresponden a los alumnos ingresantes matriculados en una determinada facultad, de los semestres 2018-2, 2019-1, 2021-2, 2022-1, 2022-2 y 2023-1.

Los datos no incluyeron los semestres 2019-2, 2020-1, 2020-2 y 2021-1, porque estos están influenciados por la pandemia que se desarrolló en ese periodo, hecho que afectó la deserción de los alumnos en la universidad.

La base de datos se divide en dos, una de entrenamiento en donde se obtendrá el modelo y una de prueba en donde se pondrá a prueba los modelos obtenidos. La base de entrenamiento contiene los datos de los semestres 2018-2, 2019-1, 2021-2 y 2022-1 y la base de prueba contiene los datos de los semestres 2022-2 y 2023-1.

Definimos a la variable dicotómica respuesta de interés como el “retiro intersemestral” (*i.retintersem*). Esta asume el valor de “1” (Sí) cuando “No” se matriculó en el semestre siguiente o el valor de “0” (No) cuando “Sí” se matriculó en el semestre siguiente.

Dado que la universidad es privada, se consideraron variables asociadas a las condiciones de buen pagador del alumno, las cuales fueron construídas con bastante cuidado y definiendo criterios, dada la complejidad del tema económico relacionado a pagos y boletas y a otras variables que se asocian como el hecho de que un alumno se matricule en pocos créditos o cuando es otra institución la que paga sus estudios.

En esta aplicación buscaremos identificar las variables o factores que determinan el retiro intersemestral de los ingresantes matriculados en una facultad de una universidad privada del país. Para ello, las posibles covariables a considerar en este modelo serán:

- *cicmat* : Ciclo de matrícula que corresponde al ciclo en el que el alumno fue admitido y se matriculó. Está expresado como la concatenación de cicloano (año del ciclo), guión y ciclo ('01', primer semestre, o '02', segundo semestre).
- *especialidad* : Código de especialidad del alumno
  - 0220 = FÍSICA
  - 0230 = MATEMÁTICAS
  - 0240 = QUÍMICA
  - 0250 = INGENIERÍA CIVIL
  - 0260 = INGENIERÍA INDUSTRIAL
  - 0270 = INGENIERÍA MECÁNICA
  - 0275 = INGENIERÍA MECATRÓNICA
  - 0280 = INGENIERÍA DE MINAS
  - 0290 = INGENIERÍA ELECTRÓNICA
  - 0300 = INGENIERÍA INFORMÁTICA
  - 0310 = INGENIERÍA BIOMÉDICA
  - 0931 = INGENIERÍA DE LAS TELECOMUNICACIONES
  - 1901 = INGENIERÍA GEOLÓGICA
  - 2521 = INGENIERÍA AMBIENTAL Y SOSTENIBLE

- *modalidad\_ingreso* : Código de modalidad de admisión bajo la cual ingresó el alumno a la universidad
  - 2 = ADM. COLE PREFERENTE. Admisión por colegio preferente
  - 7 = BECA COAR. Beca COAR
  - 8 = ING. PRE. UNIV1. Ingreso por la pre de la universidad 1
  - 9 = ING. PRE. UNIV1 TERCIO. Ingreso por la pre de la universidad 1 tercio superior
  - 10= ING. PRE. UNIV2. Ingreso por la pre de la universidad 2
  - 12= DIPLOMA BACH. UNIV1. Diplomas bachillerato universidad 1
  - 13= DIPLOMAS BACH. UNIV2. Diplomas bachillerato universidad 2
  - 14= EV. TALENTO Y EX. Evaluación de talento y exonerados
  - 17= ING. BECA 18. Ingreso por beca 18
  - 18= ING. CONVENIO UNIV2. Ingreso por convenio con universidad 2
  - 20= ING. ESPECIAL. Ingreso por modalidad especial
  - 24= ING.FE Y ALEGRÍA. Ingreso de colegios Fe y Alegría
  - 26= ING.TERCIO SUP. UNIV1. Ingreso por ser tercio superior universidad 1
  - 27= ING.TERCIO SUP. UNIV2. Ingreso por ser tercio superior universidad 2
  - 32= PRIMERA OPORTUNIDAD Ingreso en primera oportunidad
  - 34= TRANSFERENCIAS UNIV2. Transferencias universidad 2
  
- *cond\_colenolimacallao* : Condición de colegio fuera de Lima y Callao, es una variable dicotómica que indica si el colegio de procedencia del alumno se encuentra fuera de Lima y Callao y toma los valores de:
  - 1 = Si proviene de colegio fuera de Lima y Callao
  - 0 = No proviene de colegio fuera de Lima y Callao
  
- *sexo* : Identifica el sexo del alumno y toma los valores de:
  - 0 = Femenino
  - 1 = Masculino
  
- *tipocicloing* : Tipo de ciclo de ingreso, indica el tipo de ciclo de ingreso del alumno y toma los valores de
  - 1 = Si ingresó en el primer semestre
  - 2 = Si ingresó en el segundo semestre
  
- *I\_BENEFICIO* : Condición de tener beca o descuento es una variable dicotómica que toma los valores de
  - 1 = sí tiene beca o descuento

0 = no tiene beca o descuento

- *ESCALAPE* : Escala de pago del alumno en la universidad y toma los valores desde  $G1$  hasta  $G9$ .
- *CRAEST* : Es uno de los dos indicadores de rendimiento académico en el primer ciclo matriculado que tenemos en la base, es el coeficiente estandarizado que indica rendimiento del alumno al finalizar el primer ciclo.
- *PRC\_APDO\_UE* : Es el otro indicador de rendimiento académico en el primer ciclo matriculado que tenemos en la base, se expresa como el porcentaje de créditos aprobados acumulados en pregrado en su facultad.
- *PRC\_RET* : Indicador de retiro de cursos en el primer ciclo matriculado, se expresa como el porcentaje de créditos retirados en relación a los matriculados en el ciclo
- *total\_cremat* : Total de créditos matriculados en el ciclo, considerando que los cursos en la universidad tienen asignado un número de créditos según planes de estudio.
- *cond.buenpagador* : Condición de buen pagador o puntualidad en el pago en el ciclo, considerando las boletas de pago desde la segunda hasta la quinta, esta variable dicotómica toma los valores de
  - 1 = si fue buen pagador en las boletas desde la 2 a la 5
  - 0 = no fue buen pagador en las boletas desde la 2 a la 5
- *i\_cond.buenpag\_bol2* : Indicador si fue buen pagador en la boleta 2, es una variable dicotómica que toma los valores de:
  - 1 = si fue buen pagador en la boleta 2
  - 0 = no fue buen pagador en la boleta 2
- *i\_cond.buenpag\_bol3* : Indicador si fue buen pagador en la boleta 3, es una variable dicotómica que toma los valores de:
  - 1 = si fue buen pagador en la boleta 3
  - 0 = no fue buen pagador en la boleta 3
- *i\_cond.buenpag\_bol4* : Indicador si fue buen pagador en la boleta 4, es una variable dicotómica que toma los valores de:
  - 1 = si fue buen pagador en la boleta 4
  - 0 = no fue buen pagador en la boleta 4

- *i\_cond.buenpag\_bol5* : Indicador si fue buen pagador en la boleta 5, es una variable dicotómica que toma los valores de:  
 1 = si fue buen pagador en la boleta 5  
 0 = no fue buen pagador en la boleta 5

### 6.3. Análisis descriptivo

La variable de interés es el indicador de retiro intersemestral de los ingresantes matriculados en una facultad de una universidad privada del país, la cual indica si el estudiante ingresante no se matriculará en el siguiente semestre.

En la data de entrenamiento se tienen 3780 casos en donde existe un 9.23% de retiro.

El siguiente gráfico muestra la distribución marginal de esta variable según los semestres de estudio

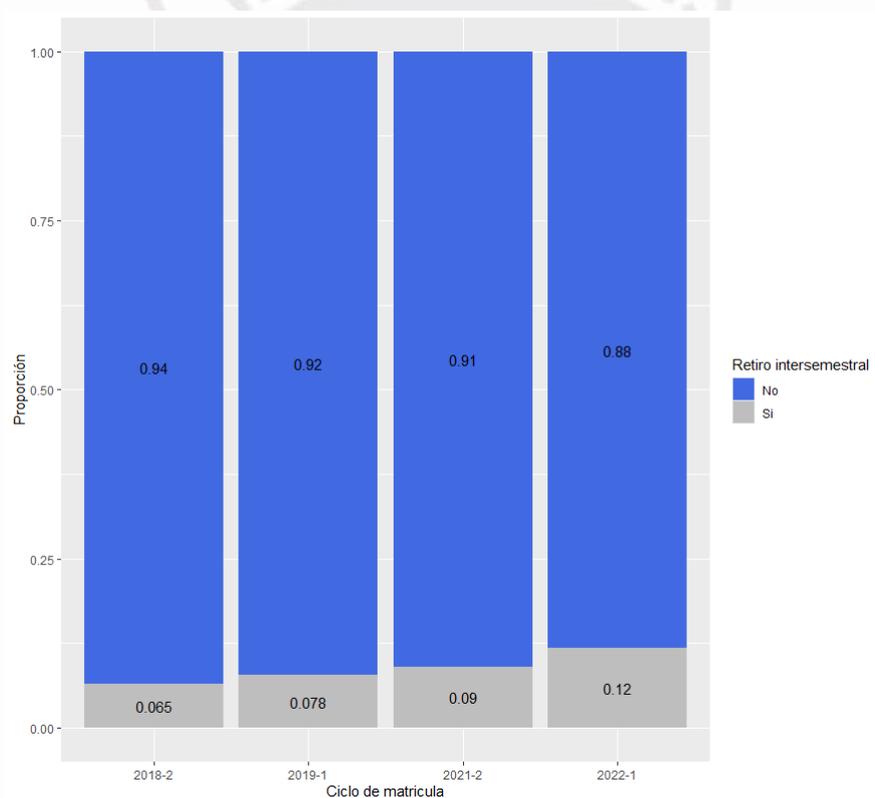


Figura 6.4: Distribución de la variable de interés retiro intersemestral por ciclo de matrícula en la base de entrenamiento

Como se aprecia el porcentaje de retiro se ha ido incrementando hasta lograr el mayor valor (12%) en el ciclo de matrícula 2022-1. Este mayor valor puede explicarse a que en el 2022-2 la universidad retornó a la presencialidad, hecho que ocasionó que muchos alumnos no decidieran matricularse en ese semestre ya que se encontraban viviendo lejos de la universidad.

Con respecto a la variable tipo de ciclo de ingreso, se observa que la proporción de los alumnos que se retiran es similar tanto para los que ingresaron en el primer semestre como para los que ingresaron en el segundo semestre, tal como se aprecia en la siguiente tabla e imagen, por tal motivo consideramos que al no existir una relación marginal clara de esta variable con la variable objetivo la retiraremos de la lista de posibles predictoras.

Retiro intersemestral	Tipo de ciclo de ingreso			
	primer semestre		segundo semestre	
	N	%	N	%
No	2607	90 %	824	92 %
Sí	282	10 %	67	8 %
Total	2889	76 %	891	24 %

Cuadro 6.5: Tabla de la distribución de la variable de interés retiro intersemestral por tipo de ciclo de ingreso

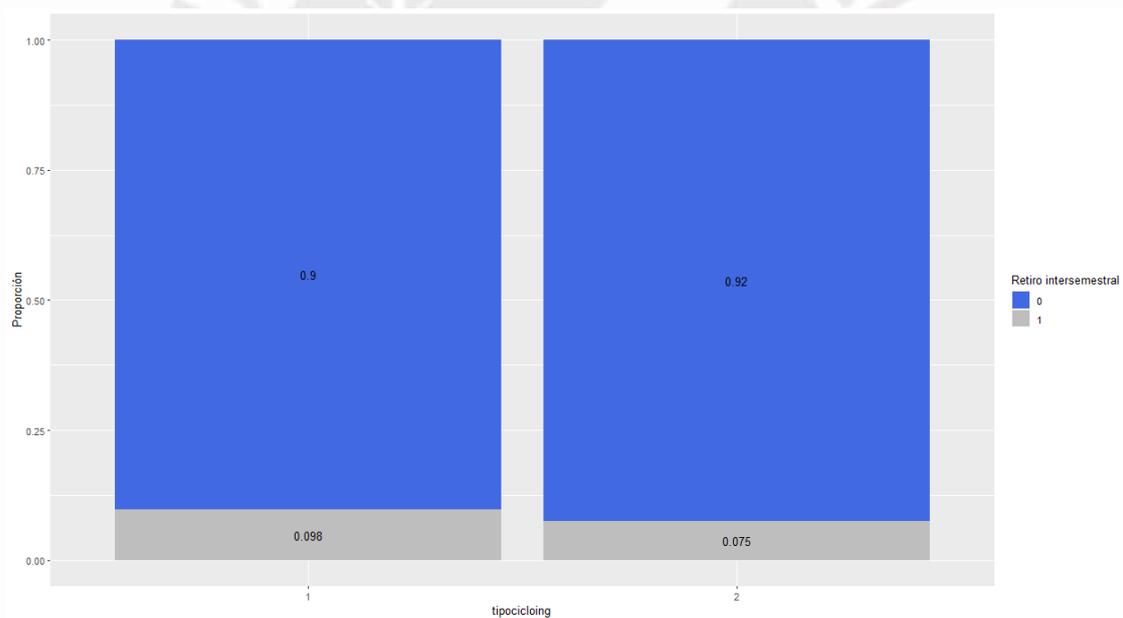


Figura 6.5: Distribución de la variable de interés retiro intersemestral por tipo de ciclo de ingreso

Dada su extensión hemos incluido el análisis descriptivo de las covariables del estudio en el apéndice.

Cabe resaltar que en la data de prueba se tiene 1730 casos en donde existe un 10% de retiro.

### 6.3.1. Especificación del modelo

En el modelo de regresión lineal MEG bajo regularización bayesiana se plantea para la variable respuesta  $Y_i = "i\_retintersem"$ , retiro intersemestral de los ingresantes matriculados en una facultad el siguiente modelo de regresión logística binaria, que la relaciona esta variable con las variables predictoras anteriormente definidas:

$$p(Y_i = 1 | \mathbf{x}_i, \beta_0, \boldsymbol{\beta}) = \frac{1}{1 + \exp(-(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}))}$$

Recordemos que podemos tratar este como un modelo de regresión lineal con MEG, desde que por (2.6), se tiene la representación

$$p(y_i | \boldsymbol{\beta}) = 2^{-1} \exp(\kappa_i \mathbf{x}_i^T \boldsymbol{\beta}) \int_0^\infty \exp(-\omega_i^2 (\mathbf{x}_i^T \boldsymbol{\beta})^2 / 2) p(\omega_i^2) d\omega_i^2$$

donde  $\kappa_i = y_i - \frac{1}{2}$  y  $p(\omega_i^2)$  es la densidad de una variable aleatoria Pólya Gamma con parámetros  $(0, 1)$ . Todo esto se explicó con detalle en la sección 3.6.

### 6.3.2. Métodos de aplicación

Se ajustará este modelo aplicando regresión logística clásica en primer lugar, luego los métodos de regularización bayesiana con priori de Horseshoe, Horseshoe + y el estimador Lasso Bayesiano.

Todos los métodos se implementaron en el software R, usando la función “bayesreg” del paquete “bayeseg” de R (Makalic y Schmidt, 2016a). Para los métodos bayesianos se ejecutaron 15000 iteraciones, de las cuales se descartaron las primeras 2000.

### 6.3.3. Resultados

Usando la base de datos de entrenamiento se procedió a estimar los parámetros del modelo. Primero aplicando regresión logística binaria, luego se aplicaron los métodos de regularización bayesiana, estimador de Horseshoe, de Horseshoe + y estimador Lasso Bayesiano en el modelo de mixtura de escala consideran un modelo logístico para la variable respuesta.

Detallando la categoría base para las variables categóricas con más de dos categorías:

- *cicmat*: 2018-2
- *especialidad*: FÍSICA
- *modalidad\_ingreso*: ADM. COLE PREFERENTE

- *ESCALAPE(escaladepago): G1*

En cuanto a la aplicación de la regresión logística clásica, primero se corrió el modelo con todas las 16 covariables presentadas. Este primer modelo presentaba un  $AIC = 1649$  y devianza residual igual a 1545. En este modelo los estimadores de los coeficientes que resultaron significativos fueron: especialidad en las categorías, Ingeniería Civil, Ingeniería Industrial, Ingeniería Mecatrónica, Ingeniería de Minas, Ingeniería Electrónica, Ingeniería Informática e Ingeniería de las Telecomunicaciones; modalidad de ingreso en sus categorías DIPLOMA BACH. UNIV1. e ING.TERCIO SUP. UNIV1; escala de pago en las categorías, G3, G4, G5, G6, G7, G8 y G9; CRAEST; PRC\_APDO\_UE; PRC\_RET; i.cond\_buenpag\_bol4 en su categoría “Sí” e i.cond\_buenpag\_bol5 en su categoría “Sí”.

Luego se corrió el modelo con las 8 covariables que resultaron significativas en el modelo 1, este segundo modelo presentaba una reducción en el AIC,  $AIC = 1636.3$ , y un incremento ligero en la devianza residual igual a 1552.3. En este modelo, el estimador del intercepto resultó significativo y los estimadores que resultaron significativos al 95 % fueron: especialidad en las categorías, Ingeniería Civil, Ingeniería Industrial, Ingeniería Mecatrónica, Ingeniería de Minas, Ingeniería Electrónica, Ingeniería Informática e Ingeniería de las Telecomunicaciones; modalidad de ingreso en las categorías, DIPLOMA BACH. UNIV1, ING. BECA 18 e ING. TERCIO SUP. UNIV1; escala de pago en las categorías, G3, G4, G5, G6, G7, G8 y G9; CRAEST; PRC\_APDO\_UE; PRC\_RET; i.cond\_buenpag\_bol4 en su categoría “Sí” e i.cond\_buenpag\_bol5 en su categoría “Sí”.

Como las variables CRAEST y PRC\_APDO\_UE, relacionadas con el rendimiento académico, están correlacionadas, se corrió un tercer modelo, considerando solo como covariables a estas dos, resultando que las dos salen significativas para el modelo. Esto nos indica que no podemos prescindir de ninguna de ellas, ya que ambas aportan información al modelo. Además al correr un cuarto modelo sin considerar una de ellas, en este caso la variable CRAEST, se obtuvo un modelo con un  $AIC = 1651.9$  y una devianza residual de 1569.9, es decir se obtuvo un notable incremento en el coeficiente AIC y en la devianza residual, esto nos indica que este es el peor modelo de los analizados, por lo que procedemos a escoger el segundo modelo con las 8 covariables, por el principio de parsimonia, ya que es un modelo menos complejo que el modelo con todas las covariables y según el índice de Akaike (AIC) es el mejor modelo de los analizados.

La regresión logística clásica con todas las covariables y solo con las variables significativas muestran un pseudo  $R^2$ , mediante la medida de Hosmer y Lemeshow  $R_L^2$ , muy cercano de 0.33.

En cuanto a los métodos bayesianos, hemos aplicado para la estimación en la regresión logística, los métodos de regularización bayesiana con los estimadores de Horseshoe, de Horseshoe + y el estimador Lasso Bayesiano.

La convergencia del proceso bayesiano se evaluó observando el comportamiento de las cadenas para cada parámetro y el Rhat, los cuales junto a las autocorrelaciones mostraron evidencias de convergencia.

Las distribuciones a posteriori y las cadenas se presentan en el apéndice C.4.

Con referente al modelo de regresión logística con MEG y regularización bayesiana usando una priori de Horseshoe, este modelo seleccionó las siguientes covariables significativas: modalidad de ingreso en su categoría DIPLOMA BACH. UNIV1 (con 75 % de credibilidad) y las covariables CRAEST, PRC\_APDO\_UE, PRC\_RET, i\_cond\_buenpag\_bol4 en su categoría “Sí” e i\_cond\_buenpag\_bol5 en su categoría “Sí”. Las estimaciones de los coeficientes del resto de covariables y categorías presentan coeficientes con valores de casi cero lo que confirma la inducción de esparcidad ya que todo lo que no es relevante o no aporta información para la estimación de la variable respuesta se redujo a cero. Cabe resaltar que todas las covariables seleccionadas en este modelo también fueron seleccionadas por el modelo clásico, con la salvedad de que el modelo clásico selecciona un número mayor de variables. Este modelo con la priori de Horseshoe presenta un  $WAIC = 817.43$ , un pseudo  $R^2 = 0.3178$  y un tiempo de ejecución promedio de 29.14 minutos.

Cabe resaltar que adicionalmente se corrió el modelo solo con las variables significativas, obteniendo así un modelo menos complejo, con un  $WAIC = 813.8$  y un pseudo  $R^2 = 0.3145$ . En este la covariable modalidad de ingreso en su categoría DIPLOMA BACH. UNIV1 resultó significativa bajo el criterio considerando el intervalo de credibilidad al 95 % a diferencia del modelo con todas las covariables que la seleccionó pero bajo el criterio considerando el intervalo de credibilidad al 75 %. Además otras categorías de esta covariable resultaron significativas pero bajo el criterio considerando el intervalo de credibilidad al 75 %, son ING.TERCIO SUP. UNIV1 y TRANSFERENCIAS UNIV2.

En cuanto a la interpretación de los coeficientes estimados de las covariables significativas para el modelo con la priori de Horseshoe, nos dice que ser un alumno que ha ingresado bajo la modalidad de ingreso DIPLOMA BACH. UNIV1 presenta riesgo de retiro intersemestral, esto puede explicarse a que este grupo de alumnos usualmente recibe ofertas de estudios en otras universidades del país o extranjeras lo que puede influir en su decisión de trasladarse o empezar sus estudios en otro lugar. Adicionalmente en cuanto a la condición de buen pagador, nos indica que si el alumno fue buen pagador en las boletas 4 o 5 no presenta riesgo de retiro

intersemestral. Con respecto a las variables cuantitativas, nos indica que tener un mayor CRAEST o PRC\_APDO\_UE no presenta riesgo de retiro, en cambio mientras mayor sea el valor de PRC\_RET para un alumno, este presenta riesgo de retirarse en el siguiente semestre.

En cuanto al modelo de regresión logística con mixtura de escala Gaussiana con regularización bayesiana usando una priori de Horseshoe +, este modelo seleccionó las siguientes covariables, bajo el criterio considerando el intervalo de credibilidad al 75 %, se seleccionó solo modalidad de ingreso en su categoría DIPLOMA BACH. UNIV1; y bajo el criterio, considerando el intervalo de credibilidad al 95 %, se seleccionaron como significativas: CRAEST, PRC\_APDO\_UE, PRC\_RET, i\_cond\_buenpag\_bol4 en su categoría “Sí” e i\_cond\_buenpag\_bol5 en su categoría “Sí”. Las estimaciones de los coeficientes del resto de covariables y categorías presentan coeficientes con valores de casi cero lo que confirma la inducción de esparcidad ya que todo lo que no es relevante o no aporta información para la estimación de la variable objetivo se redujo a cero. Cabe resaltar que el modelo con la priori de horseshoe + seleccionó las mismas variables que el modelo con la priori de horseshoe. Este modelo con la priori de horseshoe + presenta un  $WAIC = 817.49$ , un pseudo  $R^2 = 0.3159$  y un tiempo de ejecución promedio de 29.11 minutos.

En cuanto a la interpretación de los coeficientes estimados de las covariables significativas para el modelo con la priori de Horseshoe + es la misma que la mencionada en el modelo anterior con la priori de Horseshoe.

En cuanto al modelo de regresión logística con mixtura de escala Gaussiana con regularización bayesiana usando una priori de Lasso, este modelo seleccionó las siguientes covariables, bajo el criterio considerando el intervalo de credibilidad al 75 %, se seleccionaron como significativas: especialidad en sus categorías ING. CIVIL, ING. INDUSTRIAL e ING. DE LAS TELECOMUNICACIONES; modalidad de ingreso en sus categorías ING. PRE. UNIV2, DIPLOMA BACH. UNIV2, ING. TERCIO SUP. UNIV1; escala de pago en su categoría G4; y bajo el criterio, considerando el intervalo de credibilidad al 95 %, se seleccionaron como significativas: modalidad de ingreso en su categoría DIPLOMA BACH. UNIV1, CRAEST, PRC\_APDO\_UE, PRC\_RET, i\_cond\_buenpag\_bol4 en su categoría “Sí” e i\_cond\_buenpag\_bol5 en su categoría “Sí”. Cabe resaltar que este modelo además de seleccionar las covariables seleccionadas por el modelo con la priori de Horseshoe + y el modelo con la priori de Horseshoe, selecciona un grupo adicional, es decir se queda con más covariables significativas para el modelo, lo que hace que se a un modelo más complejo que los dos

anteriores. Además se observa que varios de los coeficientes estimados de las categorías no significativas no son cercanos a cero. Este modelo con la priori de Lasso presenta un  $WAIC = 820.46$ , un pseudo  $R^2 = 0.3282$  y un tiempo de ejecución promedio de 28 minutos.

En cuanto a la interpretación de los coeficientes estimados de las covariables significativas para el modelo con la priori de Lasso, nos dice que ser estudiante de las especialidades ING. CIVIL, ING. INDUSTRIAL e ING. DE LAS TELECOMUNICACIONES no presenta riesgo de retiro intersemestral; además el haber ingresado bajo la modalidad de ingreso ING. PRE. UNIV2 o DIPLOMA BACH. UNIV2 no presenta riesgo de retiro intersemestral, esto se explica a que este grupo corresponde a los alumnos que han ingresado a la universidad denominada UNIV2 con la intención de estudiar la carrera ING. BIOMÉDICA, de estudios compartidos con la universidad UNIV1, entonces podemos concluir que no se suelen retirar ya que su objetivo claramente es continuar; pero el haber ingresado bajo la modalidad de ingreso ING. TERCIO SUP. UNIV1 y DIPLOMA BACH. UNIV1 presenta riesgo de retiro intersemestral, esto puede explicarse en cuanto al primero porque usualmente a los alumnos de este grupo se les asigna escalas de pago muy altas que puede afectar su decisión de continuar con sus estudios, y para el segundo porque este grupo de alumnos usualmente recibe ofertas de estudios en otras universidades del país o extranjeras lo que puede influir en su decisión de trasladarse o empezar sus estudios en otro lugar. Adicionalmente en cuanto a la escala de pago nos dice que el grupo que pertenece a la escala G4 presenta riesgo de retiro intersemestral y que este grupo es significativamente diferente a los que pertenecen a las otras escalas. Con respecto a la condición de buen pagador, nos indica que si el alumno fue buen pagador en las boletas 4 o 5 no presenta riesgo de retiro intersemestral. Con respecto a las variables cuantitativas, nos indica que tener un mayor CRAEST o PRC\_APDO\_UE no presenta riesgo de retiro, en cambio mientras mayor sea el valor de PRC\_RET para un alumno, este presenta riesgo de retirarse en el siguiente semestre.

En conclusión, los métodos bayesianos seleccionan como significativas un grupo menor de covariables que las seleccionadas por la regresión logística clásica, con la priori de Horseshoe y Horseshoe + se selecciona el mismo conjunto de 6 covariables, en cambio con la priori de Lasso se selecciona un grupo de 8 covariables, todas incluidas en el grupo seleccionado por el modelo clásico.

Los métodos bayesianos con prioris de Horseshoe y Horseshoe + muestran errores estándar menores que los obtenidos por el modelo con priori de Lasso. Cabe resaltar que el modelo con priori de Horseshoe + muestra valores de errores estándar menores e intervalos de credibilidad

más pequeños que los del modelo con priori de Horseshoe.

Finalmente, analizando la precisión predictiva esperada, según el criterio WAIC calculado para los modelos obtenidos aplicando los métodos bayesianos, el modelo con priori de Horseshoe presenta el menor valor ( $WAIC = 817.43$ ), por lo que este modelo tiene el mejor ajuste seguido por el que aplica el estimador de Horseshoe + ( $WAIC = 817.52$ ).

Además el modelo con priori de Horseshoe presenta mayor valor de pseudo  $R^2$  ( $R^2 = 0.3178$ ) que el modelo con priori de Horseshoe + ( $R^2 = 0.3158$ ) lo que indica que Horseshoe presenta mejor ajuste que Horseshoe + para estos datos. En cambio el modelo con priori de Lasso presenta el mayor valor de pseudo  $R^2$  ( $R^2 = 0.3282$ ) de los tres modelos, lo que nos indica un mejor ajuste que los otros dos, pero recordamos que este selecciona dos covariables adicionales, lo que lo hace un modelo más complejo.

Adicionalmente, por el principio de parsimonia, es conveniente trabajar con el modelo más simple y con buen ajuste, en este caso como lo indicamos, el modelo obtenido con la priori de Lasso selecciona dos covariables adicionales pero no mejora el ajuste (indicado por el valor WAIC) comparado con los dos métodos bayesianos aplicados, lo que nos lleva a concluir que para esta aplicación nos conviene apostar por el modelo obtenido con el estimador de Horseshoe, ya que presenta el menor valor WAIC y valores de errores estándar pequeños. Además presenta un valor de  $R^2$  ajustado mayor al modelo con priori de Horseshoe + y como veremos líneas abajo presenta un valor de AUC alto, destacando sobre el modelo MEG con priori de Lasso.

	Priori de Horseshoe	Priori de Horseshoe +	Priori de Lasso
$WAIC$	817.43	817.52	820.46
$R^2$	0.3178	0.3158	0.3282

Cuadro 6.6: WAIC y pseudo  $R^2$  de los modelos bayesianos evaluados en la aplicación a un estudio de deserción universitaria

En los cuadros siguientes se muestra el detalle de la estimación aplicando un modelo MEG con priori de Horseshoe con todas las variables, se aprecian las estimaciones de los parámetros de regresión, los errores estándar, los cuantiles de los parámetros que definen los intervalos de credibilidad y el Rhat de los coeficientes. Cabe resaltar que para que una variable sea considerada “significativa”, su intervalo de credibilidad al 95 % para la exponenciación del coeficiente estimado no debería contener a 1, hecho que se señala en la covariable como “\*\*\*”. Adicionalmente, el paquete “bayesreg”, considera también la posibilidad de tener variables significativas considerando intervalos de credibilidad al 75 %, esto se señala como “\*”.

Covariable	Parámetro	$\hat{\beta}$	$\exp(\hat{\beta})$	exp(Error Estándar)	exp(IC(95 %))	Rhat
Intercepto	$\beta_0$	2.3256	10.2336	44.1027	(0.6457 173.5283)	1.0000
cicmat2019-1	$\beta_{11}$	-0.0114	0.9885	0.0971	(0.7082 1.0889)	0.9999
cicmat2021-2	$\beta_{12}$	-0.0026	0.9973	0.1358	(0.7068; 1.2395)	0.9999
cicmat2022-1	$\beta_{13}$	0.0110	1.0102	0.1274	(0.9025; 1.4021)	0.9999
especialidad MA- TEMÁTICAS	$\beta_{21}$	0.0150	1.0152	1.3014	(0.3998; 5.5013)	1.0001
especialidad QUÍMICA	$\beta_{22}$	0.0021	1.0021	0.3853	(0.5466; 2.0571)	1.0001
especialidad ING. CIVIL	$\beta_{23}$	-0.0530	0.9483	0.1281	(0.5402; 1.0424)	1.0003
especialidad ING. INDUSTRIAL	$\beta_{24}$	-0.0601	0.9416	0.1331	(0.51679; 1.0387)	0.9999
especialidad ING. MECÁNICA	$\beta_{25}$	0.0020	1.0020	0.1714	(0.7845; 1.4568)	1.0002
especialidad ING. MECATRÓNICA	$\beta_{26}$	0.0001	1.0001	0.1052	(0.8193; 1.2320)	1.0000
especialidad ING. DE MINAS	$\beta_{27}$	-0.0050	0.9950	0.1945	(0.5629; 1.3255)	1.0009
especialidad ING. ELECTRÓNICA	$\beta_{28}$	-0.0011	0.9988	0.1618	(0.6748; 1.3095)	0.9999
especialidad ING. INFORMÁTICA	$\beta_{29}$	0.0072	1.0072	0.1374	(0.9011; 1.4401)	1.0000
especialidad ING. BIOMÉDICA	$\beta_{2(10)}$	0.0230	1.0233	0.3197	(0.8930; 2.1462)	1.0006
especialidad ING. DE LAS TELE- COMUNIC.	$\beta_{2(11)}$	-0.0505	0.9506	0.2019	(0.3316; 1.1232)	0.9999

Cuadro 6.7: Parámetros estimados por modelo de regresión logística con mixtura de escala Gaussiana con regularización bayesiana usando una priori de Horsehoe

Covariable	Parámetro	$\hat{\beta}$	$exp(\hat{\beta})$	exp(Error Estándar)	exp(IC(95 %))	Rhat
especialidad ING. GEOLÓGICA	$\beta_{2(12)}$	0.0015	1.0015	0.2243	(0.6975; 1.5770)	1.0002
especialidad ING. AMBIENTAL Y SOSTENIBLE	$\beta_{2(13)}$	0.0151	1.0152	0.3662	(0.7603; 2.1961)	1.0000
modalidad de ingreso BECA COAR	$\beta_{31}$	-0.0221	0.9781	3.0441	(0.000; 11.9330)	1.0001
modalidad ingreso ING. PRE. UNIV1	$\beta_{32}$	-0.0173	0.9828	0.1156	(0.6542; 1.1078)	0.9999
modalidad ingreso ING. PRE. UNIV1 TERCIO	$\beta_{33}$	0.0023	1.0023	0.2661	(0.6393; 1.6826)	1.0000
modalidad ingreso ING. PRE. UNIV2	$\beta_{34}$	-0.1484	0.8620	0.3586	(0.0581; 1.4641)	0.9999
modalidad ingreso DIPLOMA BACH. UNIV1 (*IC al 75 %)	$\beta_{35}$	0.7669	2.1531	1.0533	(0.9750; 5.1042)	1.0000
modalidad ingreso DIPLOMA BACH. UNIV2	$\beta_{36}$	-0.1518	0.8591	0.4078	(0.0000; 1.5985)	1.0002
modalidad ingreso EV. TALENTO Y EX.	$\beta_{37}$	-0.0087	0.9913	0.1015	(0.7120; 1.1102)	0.9999

Cuadro 6.8: Parámetros estimados por modelo de regresión logística con mixtura de escala Gaussiana con regularización bayesiana usando una priori de Horseshoe

Covariable	Parámetro	$\hat{\beta}$	$\exp(\hat{\beta})$	exp(Error Estándar)	exp(IC(95 %))	Rhat
modalidad ingreso ING. BECA 18	$\beta_{38}$	0.0612	1.0632	0.8795	(0.6853; 4.1331)	0.9999
modalidad ingreso ING. CONVENIO UNIV2.	$\beta_{39}$	0.0414	1.0422	0.5061	(0.7502; 2.7343)	0.9999
modalidad ingreso ING. ESPECIAL.	$\beta_{3(10)}$	0.0090	1.0091	1.8075	(0.2112; 7.2969)	0.9999
modalidad ingreso ING. FE Y ALEGRÍA.	$\beta_{3(11)}$	0.0301	1.0305	1.3681	(0.3372; 5.7004)	1.0004
modalidad ingreso ING.TERCIO SUP. UNIV1	$\beta_{3(12)}$	0.0900	1.0942	0.1951	(0.9580; 1.7230)	1.0001
modalidad ingreso ING.TERCIO SUP. UNIV2	$\beta_{3(13)}$	0.0190	1.0192	0.4356	(0.6858; 2.3936)	1.0002
modalidad ingreso PRIMERA OPORTUNIDAD	$\beta_{3(14)}$	0.0003	1.0003	0.1317	(0.7911; 1.3074)	1.0003
modalidad ingreso TRANSFERENCIAS UNIV2.	$\beta_{3(15)}$	0.0977	1.1027	0.5961	(0.8117; 3.1485)	1.0001
cond colenolimacallao Sí	$\beta_{41}$	0.0126	1.0127	0.1048	(0.8963; 1.3072)	1.0001
LBENEFICIO Sí	$\beta_{51}$	0.0247	1.0251	0.2588	(0.7748; 1.7896)	1.0001
escala de pago G2	$\beta_{61}$	-0.0028	0.9971	0.0917	(0.7227; 1.0822)	1.0000

Cuadro 6.9: Parámetros estimados por modelo de regresión logística con mixtura de escala Gaussiana con regularización bayesiana usando una priori de Horseshoe

Covariable	Parámetro	$\hat{\beta}$	$exp(\hat{\beta})$	exp(Error Estándar)	exp(IC(95 %))	Rhat
escala de pago G3	$\beta_{62}$	0.0012	1.0012	0.0913	(0.907; 1.266)	1.0000
escala de pago G4	$\beta_{63}$	0.0018	1.0018	0.0961	(0.927; 1.304)	0.9999
escala de pago G5	$\beta_{64}$	0.0015	1.0015	0.0972	(0.914; 1.295)	0.9999
escala de pago G6	$\beta_{65}$	-0.0000	0.9999	0.1229	(0.788; 1.271)	1.0007
escala de pago G7	$\beta_{66}$	0.0024	1.0024	0.1784	(0.845; 1.544)	0.9999
escala de pago G8	$\beta_{67}$	0.0024	1.0024	0.1784	(0.845; 1.544)	0.9999
escala de pago G9	$\beta_{68}$	0.0045	1.0045	0.2579	(0.831; 1.842)	0.9999
CRAEST **	$\beta_8$	-0.0647	0.9373	0.0173	(0.903; 0.971)	1.0000
PRC_APDO_UE **	$\beta_9$	-1.6729	0.1876	0.1183	(0.070; 0.534)	1.0000
PRC_RET **	$\beta_{10}$	3.9627	52.6002	17.6104	(28.866; 97.899)	1.0001
total_cremat	$\beta_{11}$	0.0004	1.0004	0.0609	(0.889; 1.128)	0.9999
cond_buenpagador Sí	$\beta_{12(1)}$	0.0004	1.0004	0.0609	(0.889; 1.128)	1.0003
i_cond_buenpag_bol2 Sí	$\beta_{13(1)}$	0.0090	1.0090	0.1569	(0.806; 1.421)	0.9999
i_cond_buenpag_bol3 Sí	$\beta_{14(1)}$	0.0004	1.0004	0.1336	(0.780; 1.304)	0.9999
i_cond_buenpag_bol4 Sí **	$\beta_{15(1)}$	-0.6798	0.5066	0.1313	(0.330; 0.845)	1.0000
i_cond_buenpag_bol5 Sí **	$\beta_{16(1)}$	-0.9583	0.3835	0.0869	(0.242; 0.583)	1.0002
sexo 1 (Masculino)	$\beta_{18(1)}$	-0.0130	0.9870	0.0958	(0.761; 1.137)	0.9999

Cuadro 6.10: Parámetros estimados por modelo de regresión logística con mixtura de escala Gaussiana con regularización bayesiana usando una priori de Horseshoe

Una aplicación con esta misma data de deserción univeritaria se realizó en un estudio previo desarrollado por la universidad propietaria de los datos. El modelo que hallaron en SPSS Modeler fue un modelo XGBoost, el cual seleccionó algunas de las mismas covariables seleccionadas por nuestro modelo con priori de Horseshoe, tales como: CRAEST, PRC\_APDO\_UE, PRC\_RET, i\_cond\_buenpag\_bol5 e i\_cond\_buenpag\_bol4. también consideró a la modalidad de ingreso como significativa pero en las categorías: ING.TERCIO SUP. UNIV1. y ING.

PRE. UNIV1., que difiere al modelo con priori de Horseshoe que marcó como significativa la categoría DIPLOMA BACH. UNIV1. Adicionalmente el modelo XGBoost seleccionó la variable especialidad en su categoría INGENIERÍA INFORMÁTICA pero nuestro modelo con priori de Horseshoe no la seleccionó.

Por otro lado este modelo XGBoost presentó un AUC de 0.89, valor similar a los obtenidos por los modelos con priori de Horseshoe, Horseshoe + y de Lasso, como mostraremos en los siguientes cuadros y gráficos.

Los modelos analizados se aplicaron a la base de datos reservada como de prueba y se realizaron las estimaciones, obteniendo así las métricas AUC (mide el rendimiento de un clasificador binario) de evaluación de los modelos y los gráficos de las curvas ROC (representación gráfica de la sensibilidad frente a la especificidad), en general se obtiene un AUC de 0.89, que indica que los modelos tienen un buen poder predictivo o de clasificación, destacando entre ellos los modelos con las estimaciones de Horseshoe + y Horseshoe.

	Logístico clásico	Priori de Horseshoe	Priori de Horseshoe +	Priori de Lasso
AUC	0.8918	0.8966	0.8975	0.8954

Cuadro 6.11: Métrica AUC de los modelos evaluados en la aplicación a un estudio de deserción universitaria

En los gráficos ROC se observa que está señalado en rojo el punto de corte óptimo, que es aquel que maximiza la proporción de verdaderos positivos (sensibilidad), al mismo tiempo minimiza el ratio de falsos positivos (1 – especificidad).

Los puntos de corte óptimos se resumen en un cuadro en donde también se muestran los valores de sensibilidad y especificidad según punto de corte 0.5 y el mejor punto de corte calculado para cada modelo.

El modelo estimado con la priori de Horseshoe + presenta mayor valor de sensibilidad con su punto de corte óptimo, en cambio los modelos clásico y con priori de Lasso presentan mayor especificidad con su punto de corte óptimo.

Modelo	Punto de corte	Sensibilidad	Especificidad
Logístico clásico	0.50	0.5028	0.9690
	0.1381	0.8343	0.8121
Horseshoe	0.50	0.4696	0.9742
	0.1309	0.8343	0.8096
Horseshoe +	0.50	0.4641	0.9396
	0.1230	0.8398	0.8005
Lasso	0.50	0.5028	0.9690
	0.1343	0.8343	0.8121

Cuadro 6.12: Análisis de sensibilidad y especificidad según punto de corte 0.5 y óptimo

Adicionalmente se muestran las tablas de contingencia generadas por estos dos puntos de corte en cada modelo. Como podemos observar al escoger el punto de corte óptimo obtenemos un mejor poder de clasificación, superior al obtenido con el punto de corte 0.5.

Modelo	Real	Estimado (corte=0.5)		Estimado (mejor punto de corte)	
		NO se retiró 0	Se retiró 1	NO se retiró 0	Se retiró 1
Logístico clásico	NO se retiró 0	1501	48	1258	291
	Se retiró 1	90	91	30	151
Horsehoe	NO se retiró 0	1509	40	1254	295
	Se retiró 1	96	85	30	151
Horsehoe+	NO se retiró 0	1509	40	1240	309
	Se retiró 1	97	84	29	152
Lasso	NO se retiró 0	1509	40	1261	288
	Se retiró 1	97	84	34	147

Cuadro 6.13: Matrices de confusión para el modelo de regresión logística, bajo cada método de estimación

Adicionalmente, se replicó este estudio (no mostrado aquí) con un modelo que consideraba como data de entrenamiento, a solo los semestres “2021 – 2” y “2022 – 1”. Este seleccionó casi las mismas covariables, con la salvedad de que presentó un menor valor de AUC, por lo que se decidió continuar trabajando con toda la data de entrenamiento.

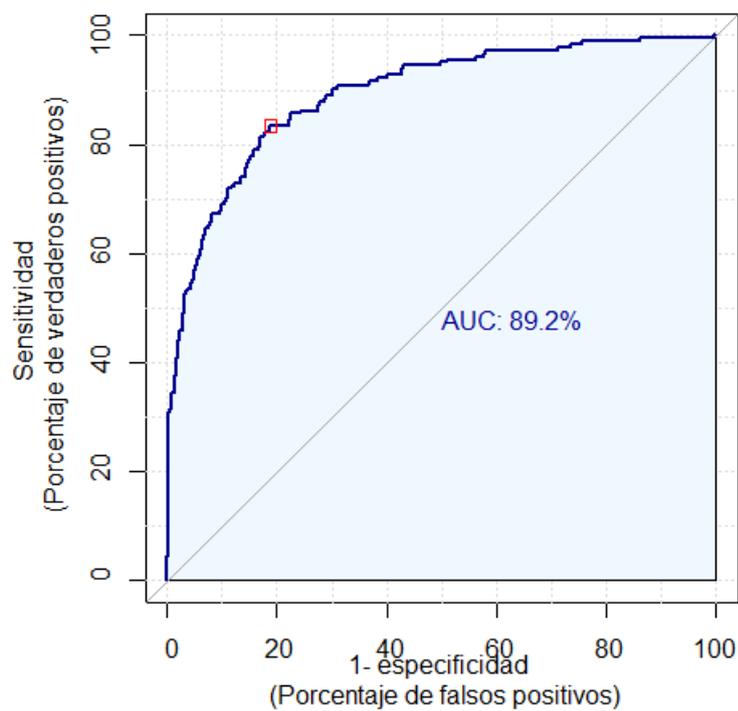


Figura 6.6: Gráfico de la curva ROC para el modelo de regresión logística clásico

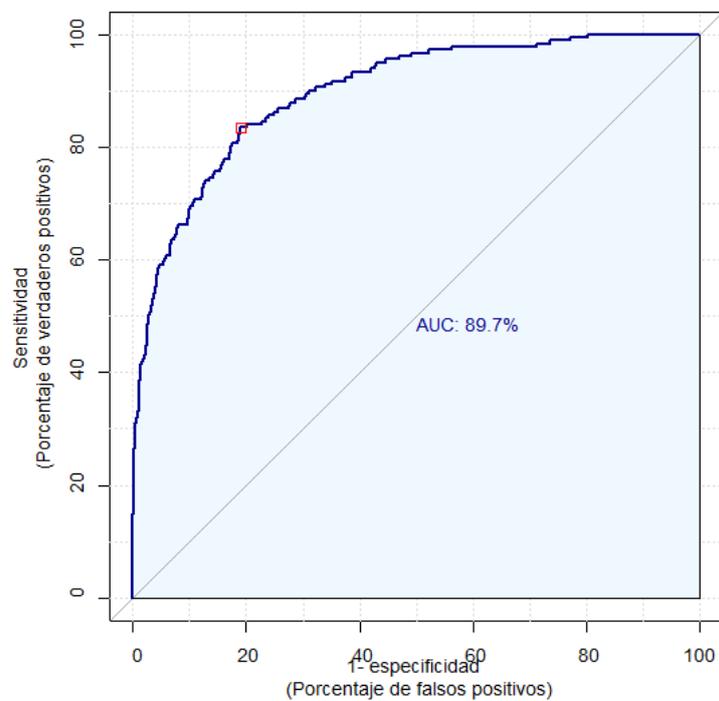


Figura 6.7: Gráfico de la curva ROC para el modelo MEG con priori de Horseshoe

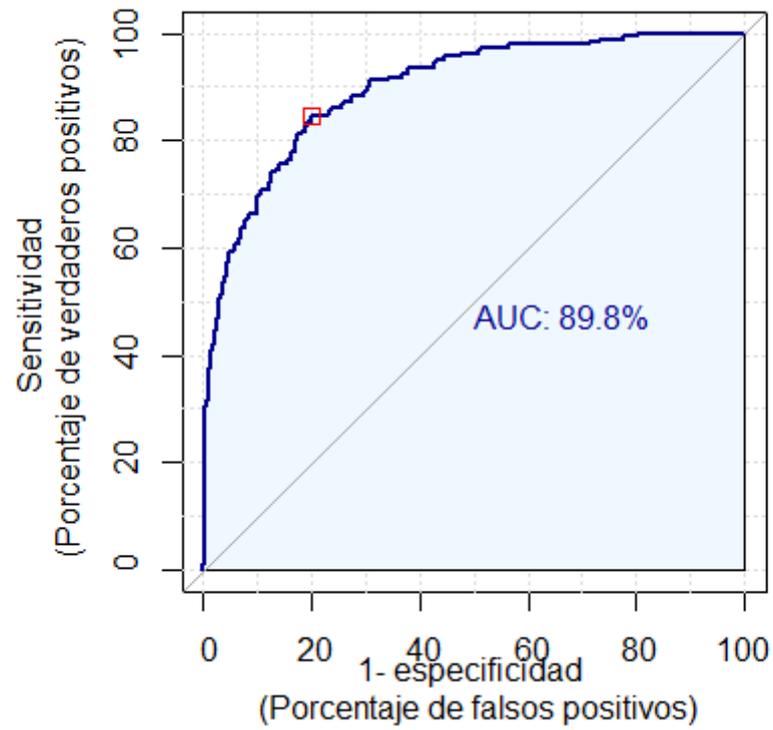


Figura 6.8: Gráfico de la curva ROC para el modelo MEG con priori de Horseshoe +

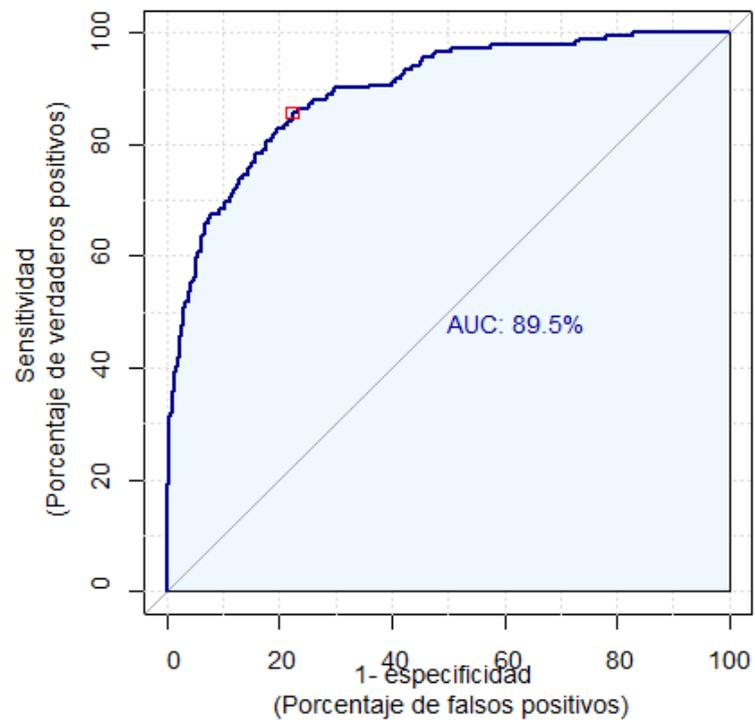


Figura 6.9: Gráfico de la curva ROC para el modelo MEG con priori de Lasso

# Capítulo 7

## Conclusiones

### 7.1. Conclusiones

En este trabajo hemos desarrollado las propiedades, estimación y aplicación a dos conjuntos de datos reales de diversas técnicas de regularización bayesiana sobre un modelo de regresión lineal múltiple con mixtura de escala Gaussiana, modelo que incluye al de una regresión logística. De particular interés en este trabajo, fue la comparación de las técnicas de regularización bajo penalización y las derivadas de introducir las prioris de Horseshoe y de Horseshoe + a los coeficientes de regresión del modelo. Mostramos aquí, de manera explícita, cómo realizar un muestreo de Gibbs para la estimación de estos modelos, detallando no solo las distribuciones condicionales completas necesarias; sino también cómo es posible, mediante el uso del paquete bayesreg de R, optimizar algunas de estas propuestas de muestreo.

En cuanto a los estudios de simulación realizados y aplicaciones, obtuvimos las siguientes conclusiones:

- En el estudio con datos simulados para covariables no correlacionadas, se observa que los tres técnicas bayesianas seleccionaron el modelo correcto, pero la estimación usando la priori de Horseshoe obtuvo un menor error cuadrático medio y menores sesgos y RMSE para los estimadores de los coeficientes, lo que nos indica que ofrece una estimación más precisa de los parámetros  $\beta$  que los otros dos métodos bayesianos e incluso que el de la regresión clásica.
- En el estudio con datos simulados para covariables correlacionadas, se observa que las tres técnicas bayesianas descartan la covariable no espuria altamente correlacionada. En este escenario el método Lasso presenta menor error cuadrático medio y menor sesgo al estimar la covariable seleccionada, pero no estima correctamente las covariables espurias

como sí lo hace los métodos de estimación de Horseshoe y Horseshoe +.

- En la simulación con covariables correlacionadas, las tres técnicas bayesianas aplicadas proporcionan mejores resultados en comparación a la regresión clásica (menor error cuadrático medio). Todas, eso sí indican que la covariable correlacionada no debería ser seleccionada en el modelo.
- En la aplicación con los datos de la base "BostonHousing", los modelos con errores de Laplace y de  $t$  de Student mostraron un mejor ajuste que los modelos con errores Gaussianos, dado que la variable respuesta presenta datos atípicos y cola a la derecha. Entre estos, el de Laplace presentó un mejor ajuste, destacando muy marginalmente cuando se emplea el estimador Lasso bayesiano. Tanto el estimador con priori de Horseshoe y el del Lasso bayesiano seleccionaron una variable adicional que las seleccionadas por la regresión múltiple clásica y la priori de Horseshoe +, que es la variable *age*: proporción de unidades ocupadas por sus propietarios construídas antes de 1940, lo que incrementa la complejidad del modelo pero aún así presentan un mejor ajuste.
- El relación a la aplicación sobre la deserción universitariaa datos de una Universidad privada del país, el modelo de regresión logístico clásico y el Lasso bayesiano seleccionaron 8 de las 16 covariables (aunque con algunas categorías divergentes) , en cambio los modelos logísticos con mixtura de escala Gaussiana y priori Horseshoe y Horseshoe +, seleccionaron solo 6 de estas 8 covariables, con algunas diferencias en ciertas categorías. Todos los modelos comparados seleccionaron las variables cuantitativas relacionadas con el rendimiento académico como muy significativas, de igual manera las variables relacionadas con el pago de las boletas 4 y 5 por derechos académicos. En cuanto a la modalidad o canal de ingreso, los modelos con priori de Horseshoe y Horseshoe + indicaron que la categoría DIPLOMA BACH. UNIV1 (Diplomas bachillerato universidad 1) es significativamente diferente a las demás y con un coeficiente estimado positivo, es decir este grupo de alumnos es un grupo de riesgo importante para el modelo.
- En la aplicación de deserción universitaria, el estimador con priori de Horseshoe + mostró el menor error estándar de estimación de los tres modelos con regularización bayesiana, seguido por el modelo con priori de Horseshoe. En cambio el estimador Lasso bayesiano mostró mayores errores estándar de estimación e intervalos de credibilidad más anchos que los de los modelos con priori de Horseshoe y Horseshoe +. Según el criterio WAIC, la estimación que presentó un mejor ajuste fué la realizada con la priori de Horseshoe, seguido muy de cerca por el de la priori de Horseshoe +. La técnica

Lasso presentó el peor ajuste de los tres técnicas de regularización bayesiana, además de seleccionar dos covariables más que los otros métodos.

- Al aplicar estas estimaciones con los datos de los semestres (2018–2, 2019–1, 2021–2 y 2022–1) a los datos de validación correspondientes a los semestres (2022–2 y 2023–1), se obtuvo un alto poder predictivo bajo todos los métodos, con altos valores de las métricas AUC de las curvas ROC, siendo los mayores las obtenidas con las estimaciones bajo las prioris de Horseshoe + y Horseshoe, seguido muy de cerca por el de la estimación con el Lasso bayesiano y por último por el del modelo logístico clásico. Estos análisis se realizaron tanto para un punto de corte uniforme (0.5) como para uno óptimo por método.

## 7.2. Sugerencias para investigaciones futuras

El desarrollo de estas técnicas, da pie naturalmente a poder extender este trabajo a la incorporación de otras técnicas de regularización bayesiana. Aparte de las técnicas de penalización, se han desarrollado recientemente diversas posibilidades de introducir una distribución a priori que realice la penalización en modelos de regresión. Una de estas técnicas que podría ser interesante de analizar en este contexto es la de los estimadores de Horseshoe generalizados, propuestos por Schmidt y Makalic (2019).

Parte de la motivación de este trabajo, fue la de comparar algunas técnicas de clasificación realizadas por Machine Learning para el problema de la deserción universitaria, con las de la regresión logística bajo regularización bayesiana. Estas no fueron aquí incorporadas, por razones de tiempo y confiabilidad. Dada la gran cantidad de técnicas de clasificación existentes, podría ser muy interesante un estudio que se enfoque en este tipo de comparaciones.

Desde el punto de vista de la principal aplicación en este trabajo; es decir, del de la deserción universitaria, un tema que nos quedó pendiente es el de cómo lidiar con data desbalanceada, sobre todo si estamos restringidos, bajo nuestro modelo, a un enlace logístico. En esta misma aplicación, si bien el rendimiento académico mostró ser importante, este solo se mide al finalizar el semestre de estudios, momento que puede ser muy tarde para ejecutar alguna acción que prevenga que los alumnos se retiren el siguiente semestre, por lo que se propone a futuro incluir covariables relacionadas con rendimiento académico temprano, como las notas de las evaluaciones parciales de los alumnos. Aparte de la posibilidad de diseñar otras posibles covariables que puedan afectar a la probabilidad de retiro, sería interesante poder también

extender este estudio a otras facultades o a la Universidad y poder ampliar el estudio con información de nuevos semestres, dado que en nuestra aplicación tuvimos la restricción de tener que omitir algunos semestres por el problema del COVID19.



# Apéndice A

## Código R de las aplicaciones

Se muestran aquí el código R utilizado para la aplicación

### A.1. Aplicación con datos "BostonHousing"

```
library(bayesreg)
library(MASS)
library(corrplot)
library(plotly)
library(mlbench)
library(rstan)
library(MASS)
library(psych)
library(ggpubr)

datos1=BostonHousing #datos BostonHousing

##### analisis de los datos#####
head(BostonHousing)
str(datos1)
summary(datos1)
#histograma
plot_ly(data = datos1, x = ~medv, type = "histogram")
#qqplot
ggqqplot(datos1$medv, add="qqline", xlab="\n Theoretical quantiles",
          ylab="Sample quantiles", shape=1, col="blue",
          title="qqplot de la variable medv")
#diagrama de correlacion
corr_matrix<-cor(datos1[,-4])
corrplot(corr_matrix, type="upper")
```

```

summary(datos1$medv)
#diagramas de dispersion
plot(datos1)
pairs(~medv + rm + lstat + ptratio + indus, data = datos1)

Set1<-data.frame(datos1$medv, datos1$rm, datos1$lstat, datos1$ptratio, datos1$indus );
Set1

pairs.panels(x = Set1, ellipses = FALSE, lm = TRUE, method = "pearson",
hist.col = "#CD1076")

#boxplot
plot(x = datos1$chas, y = datos1$medv ,main = "Valor medio de las viviendas
ocupadas en miles de d'olares por zona l\imite al r\'io",
      xlab = " 0 zona no limita con el r\'io y 1 la zona limita con el r\'io",
      ylab = "Valor medio de las viviendas ocupadas por sus propietarios en
      miles de d'olares", col = c( "green", "grey"))
boxplot(datos1$medv)

#transformacion propuesta por Harrison y Rubinfeld (1978)

datos_t <- data.frame(logmedv=log (datos1$medv) , rmcuad=(datos1$rm)^2,
age=datos1$age, logdis=log(datos1$dis) , lograd=log(datos1$rad),
tax=datos1$tax , ptratio=datos1$ptratio , b=datos1$b,
lstat=log(datos1$lstat),
crim=datos1$crim , zn=datos1$zn, indus=datos1$indus, chas=datos1$chas,
nox=datos1$nox)

##### Estimacion por maxima verosimilitud #####

boston_t.lm <- lm(datos_t$logmedv~.,datos_t) # Regular least-squares
summary(boston_t.lm)
#calculo de intervalos de confianza para regresion multiple
confint(object=boston_t.lm , level=0.95)
#stepwise
modelo_vacio=lm(datos_t$logmedv~1,datos_t)
regstep<-step(modelo_vacio, scope = list(lower=modelo_vacio, upper=boston_t.lm),
direction = "both");
summary(regstep)

##### Estimacion priori Horseshoe model Laplace #####

```

```

boston_t.hs <- bayesreg(datos_t$logmedv~.,datos_t,model= 'Laplace', prior='hs',
n.samples=15000, burnin=2000 ) # Horseshoe regression
summary(boston_t.hs)
boston_t.hs.s <- summary(boston_t.hs)
boston_t.hs.s$se.coef
#calculo Rhat
beta=boston_t.hs$beta;
beta_o=boston_t.hs$beta0;
Rhat(beta_o[2001:M])
for (i in 1:13) {
  Rhat(beta[i,2001:M])
}

##### Estimacion priori Horseshoe + model Laplace #####

boston_t.hsmas <- bayesreg(datos_t$logmedv~.,datos_t,prior='hs+',model= 'Laplace',
n.samples=15000, burnin=2000) # Horseshoe + regression
boston_t.hsmas.s <- summary(boston_t.hsmas)
#calculo Rhat
beta=boston_t.hsmas$beta;
beta_o=boston_t.hsmas$beta0;
Rhat(beta_o[2001:M])
for (i in 1:13) {
  Rhat(beta[i,2001:M])
}

##### Estimacion Lasso bayesiano + model Laplace #####

boston_t.lasso <- bayesreg(datos_t$logmedv~.,datos_t,model='Laplace',
prior='lasso',n.samples=15000, burnin=2000) # lasso regression
boston_t.lasso.s <- summary(boston_t.lasso)
#calculo Rhat
beta=boston_t.lasso$beta;
beta_o=boston_t.lasso$beta0;
Rhat(beta_o[2001:M])
for (i in 1:13) {
  Rhat(beta[i,2001:M])
}

##### Estimacion priori Horseshoe model t Student #####

boston_t.hs_tStudent <- bayesreg(datos_t$logmedv~.,datos_t,prior='hs',model= 't',
n.samples=15000, burnin=2000) # Horseshoe regression

```

```

boston_t.hs_tStudent.s <- summary(boston_t.hs_tStudent)

##### Estimacion priori Horseshoe + model t Student #####

boston_t.hsmas_tStudent <- bayesreg(datos_t$logmedv~.,datos_t,prior='hs+',
model= 't',n.samples=15000, burnin=2000) # Horseshoe + regression
boston_t.hsmas_tStudent.s <- summary(boston_t.hsmas_tStudent)

##### Estimacion Lasso bayesiano + model t Student #####

boston_t.lasso_tStudent <- bayesreg(datos_t$logmedv~.,datos_t,
model='t', prior='lasso',n.samples=15000, burnin=2000) # lasso regression
boston_t.lasso.s_tStudent <- summary(boston_t.lasso_tStudent)

##### Graficos#####

#####Estimacion priori Horseshoe #####

### Graficos de resumen - Beta ###
beta=boston_t.hs$beta;
beta_o=boston_t.hs$beta0;
M=boston_t.hs$n.samples
par(mfrow=c(3,3))
plot(density(beta_o[2001:M]),main ='Posteriori density of Beta0',xlab=NA,ylab=NA)
for (i in 1:8) {
  plot(density(beta[i,2001:M]),main ='Posteriori density of Beta',xlab=NA,ylab=NA)
}
dev.off()
par(mfrow=c(2,3))
for (i in 9:13) {
  plot(density(beta[i,2001:M]),main ='Posteriori density of Beta',xlab=NA,ylab=NA)
}
dev.off()
#graficos de convergencia#
par(mfrow=c(3,3))
ts.plot(beta_o[2001:M],main='Trace of beta0',xlab=NA,ylab=NA)
for (i in 1:8) {
  ts.plot(beta[i,2001:M],main='Trace of beta1',xlab=NA,ylab=NA)
}
dev.off()
par(mfrow=c(2,3))

```

```

for (i in 9:13) {
  ts.plot(beta[i,2001:M],main='Trace of beta1',xlab=NA,ylab=NA)
}
dev.off()

#####Estimacion priori Horseshoe + #####

### Graficos de resumen - Beta ###
beta=boston_t.hsmas$beta;
beta_o=boston_t.hsmas$beta0;
M=boston_t.hsmas$n.samples
par(mfrow=c(3,3))
plot(density(beta_o[2001:M]),main='Posteriori density of Beta0',xlab=NA,ylab=NA)
for (i in 1:8) {
  plot(density(beta[i,2001:M]),main='Posteriori density of Beta',xlab=NA,ylab=NA)
}
dev.off()
par(mfrow=c(2,3))
for (i in 9:13) {
  plot(density(beta[i,2001:M]),main='Posteriori density of Beta',xlab=NA,ylab=NA)
}
dev.off()

#graficos de convergencia#
par(mfrow=c(3,3))
for (i in 1:8) {
  ts.plot(beta[i,2001:M],main='Trace of beta1',xlab=NA,ylab=NA)
}
dev.off()
par(mfrow=c(2,3))
for (i in 9:13) {
  ts.plot(beta[i,2001:M],main='Trace of beta1',xlab=NA,ylab=NA)
}
dev.off()

#####Estimacion Lasso bayesiano #####

### Graficos de resumen - Beta ###
beta=boston_t.lasso$beta;
beta_o=boston_t.lasso$beta0;
M=boston_t.lasso$n.samples
par(mfrow=c(3,3))
for (i in 1:8) {

```

```

  plot(density(beta[i,2001:M]),main = 'Posteriori density of Beta',xlab=NA,ylab=NA)
}
dev.off()
par(mfrow=c(2,3))
for (i in 9:13) {
  plot(density(beta[i,2001:M]),main = 'Posteriori density of Beta',xlab=NA,ylab=NA)
}
dev.off()
#graficos de convergencia#
par(mfrow=c(3,3))
for (i in 1:8) {
  ts.plot(beta[i,2001:M],main='Trace of beta1',xlab=NA,ylab=NA)
}
dev.off()
par(mfrow=c(2,3))
for (i in 9:13) {
  ts.plot(beta[i,2001:M],main='Trace of beta1',xlab=NA,ylab=NA)
}
dev.off()

```

## A.2. Aplicación a un estudio de deserción universitaria

```

library(bayesreg)
library(foreign)
library(dplyr)
library(ggplot2)
library(gridExtra)
library(corrplot)
library(psych)
library(plotly)
library(mlbench)
library(rstan)
library(MASS)
library(ggpubr)

datos_piloto <- read.spss("G:/Mi unidad/tesis MAESTRIA ESTADISTICA/2024/aplicacion/
ingresantes matriculados para base piloto APL 2024 v2.sav", to.data.frame = TRUE)
head(datos_piloto)

datos_piloto2= datos_piloto[-19]
datos_piloto2= datos_piloto2[-8]
datos_piloto2= datos_piloto2[-17]

```

```

head(datos_piloto2)

datos_piloto2$i_retintersem=factor(datos_piloto2$i_retintersem)
datos_piloto2$cond_colenolimacallao=factor(datos_piloto2$cond_colenolimacallao)
datos_piloto2$I_BENEFICIO=factor(datos_piloto2$I_BENEFICIO)
datos_piloto2$ESCALAPE=factor(datos_piloto2$ESCALAPE)
datos_piloto2$cond_buenpagador=factor(datos_piloto2$cond_buenpagador)
datos_piloto2$sexo=factor(datos_piloto2$sexo)
datos_piloto2$tipocicloing=factor(datos_piloto2$tipocicloing)
datos_piloto2$i_cond_buenpag_bol2=factor(datos_piloto2$i_cond_buenpag_bol2)
datos_piloto2$i_cond_buenpag_bol3=factor(datos_piloto2$i_cond_buenpag_bol3)
datos_piloto2$i_cond_buenpag_bol4=factor(datos_piloto2$i_cond_buenpag_bol4)
datos_piloto2$i_cond_buenpag_bol5=factor(datos_piloto2$i_cond_buenpag_bol5)

##### analisis de los datos#####

str(datos_piloto2)
summary(datos_piloto2)
attach(datos_piloto2)

#####descriptivos univariados#####

#i_retintersem
tabla=table(datos_piloto2$i_retintersem)
tabla
prop.table(tabla)
ggplot(datos_piloto2, aes(x=i_retintersem , y = after_stat(count/sum(count)))) +
  geom_bar(fill = "royalblue") +
  geom_text(
    aes(label=signif(..count.. /sum(..count..) , digits=2), vjust=-0.5),
    stat="count",
  ) +
  labs(x= "Retiro intersemestral", y="Proporci'on")

#sexo
tabla=table(datos_piloto2$sexo)
tabla
prop.table(tabla)

ggplot(datos_piloto2, aes(x=sexo , y = after_stat(count/sum(count)))) +
  geom_bar(fill = "royalblue") +
  geom_text(

```

```

    aes(label=signif(..count.. /sum(..count..) , digits=2), vjust=-0.5),
    stat="count",
  ) +
  labs(x= "sexo", y="Proporci\ 'on")

#especialidad

tabla=table(datos_piloto2$especialidad)
tabla
prop.table(tabla)
ggplot(datos_piloto2, aes(x=especialidad , y = after_stat(count/sum(count)))) +
  geom_bar(fill = "royalblue") +
  geom_text(
    aes( label=signif(..count.. /sum(..count..) , digits=2), vjust=-0.5),
    stat="count",
  ) +
  coord_flip()+
  labs(x= "especialidad", y="Proporci\ 'on")

#modalidad de ingreso

tabla=table(datos_piloto2$modalidad_ingreso)
tabla
prop.table(tabla)
ggplot(datos_piloto2, aes(x=modalidad_ingreso , y = after_stat(count/sum(count)))) +
  geom_bar(fill = "royalblue") +
  geom_text(
    aes( label=signif(..count.. /sum(..count..) , digits=2), vjust=-0.5),
    stat="count",
  ) +
  coord_flip()+
  labs(x= "modalidad_ingreso", y="Proporci\ 'on")

#escala de pago

tabla=table(datos_piloto2$ESCALAPE)
tabla
prop.table(tabla)
ggplot(datos_piloto2, aes(x=ESCALAPE , y = after_stat(count/sum(count)))) +
  geom_bar(fill = "royalblue") +
  geom_text(
    aes( label=signif(..count.. /sum(..count..) , digits=2), vjust=-0.5),

```

```

    stat="count",
  ) +
  coord_flip()+
  labs(x= "ESCALAPE", y="Proporci\ 'on")

#tipo ciclo ingreso
tabla=table(datos_piloto2$tipocicloing)
tabla
prop.table(tabla)

#####descriptivos bivariados#####

#por ciclo de matricula
tabla11=table(datos_piloto2$i_retintersem,datos_piloto2$cicmat)
addmargins(tabla11)

tabla2=prop.table(table(datos_piloto2$i_retintersem,datos_piloto2$cicmat), margin=2)

ggplot(datos_piloto2, aes(x=cicmat, fill=i_retintersem)) +
  geom_bar(position="fill") +
  geom_text(
    aes(label=signif(..count.. / tapply(..count.., ..x.., sum)[as.character(..x..)], digits=2)),
    stat="count",
    position=position_fill(vjust=0.5)) +
  scale_fill_manual(values = c("royalblue", "grey"))+
  guides(fill = guide_legend(title = "Retiro intersemestral")) +
  labs(x= "Ciclo de matricula", y="Proporci\ 'on")

#por sexo

tabla3 <- prop.table(table(datos_piloto2$i_retintersem, datos_piloto2$sexo),margin=2)
tabla3
ggplot(datos_piloto2, aes(x=sexo, fill=i_retintersem)) +
  geom_bar(position="fill") +
  geom_text(
    aes(label=signif(..count.. / tapply(..count.., ..x.., sum)[as.character(..x..)], digits=2)),
    stat="count",
    position=position_fill(vjust=0.5)) +
  scale_fill_manual(values = c("royalblue", "grey"))+
  guides(fill = guide_legend(title = "Retiro intersemestral")) +
  labs(x= "Sexo", y="Proporci\ 'on")

```

```
#por especialidad
```

```
prop.table(table(datos_piloto2$especialidad))
table(datos_piloto2$i_retintersem, datos_piloto2$especialidad)
tabla5 <- prop.table(table(datos_piloto2$i_retintersem, datos_piloto2$especialidad),margin=2)
tabla5
ggplot(datos_piloto2, aes(x=especialidad, fill=i_retintersem)) +
  geom_bar(position="fill") +
  geom_text(
    aes(label=signif(..count.. / tapply(..count.., ..x.., sum)[as.character(..x..)], digits=2)),
    stat="count",
    position=position_fill(vjust=0.5)) +
  scale_fill_manual(values = c("royalblue", "grey"))+
  guides(fill = guide_legend(title = "Retiro intersemestral")) +
  coord_flip()+
  labs(x= "Especialidad", y="Proporci'on")
```

```
#por modalidad de ingreso
```

```
prop.table(table(datos_piloto2$modalidad_ingreso))
table(datos_piloto2$i_retintersem, datos_piloto2$modalidad_ingreso)
tabla5 <- prop.table(table(datos_piloto2$i_retintersem, datos_piloto2$modalidad_ingreso),margin=2)
tabla5
ggplot(datos_piloto2, aes(x=modalidad_ingreso, fill=i_retintersem)) +
  geom_bar(position="fill") +
  geom_text(
    aes(label=signif(..count.. / tapply(..count.., ..x.., sum)[as.character(..x..)], digits=2)),
    stat="count",
    position=position_fill(vjust=0.5)) +
  scale_fill_manual(values = c("royalblue", "grey"))+
  guides(fill = guide_legend(title = "Retiro intersemestral")) +
  coord_flip()+
  labs(x= "Modalidad de ingreso", y="Proporci'on")
```

```
#por condicion de tener beca o descuento
```

```
prop.table(table(datos_piloto2$I_BENEFICIO))
table(datos_piloto2$i_retintersem, datos_piloto2$I_BENEFICIO)
tabla4 <- prop.table(table(datos_piloto2$i_retintersem, datos_piloto2$I_BENEFICIO),margin=2)
tabla4
ggplot(datos_piloto2, aes(x=I_BENEFICIO, fill=i_retintersem)) +
  geom_bar(position="fill") +
```

```

geom_text(
  aes(label=signif(..count.. / tapply(..count.., ..x.., sum)[as.character(..x..)], digits=2)),
  stat="count",
  position=position_fill(vjust=0.5)) +
scale_fill_manual(values = c("royalblue", "grey"))+
guides(fill = guide_legend(title = "Retiro intersemestral")) +
labs(x= "Condicion de tener beca o descuento", y="Proporci'on")

#por escala de pension

prop.table(table(datos_piloto2$ESCALAPE))
table(datos_piloto2$i_retintersem, datos_piloto2$ESCALAPE)
tabla4 <- prop.table(table(datos_piloto2$i_retintersem, datos_piloto2$ESCALAPE),margin=2)
tabla4
ggplot(datos_piloto2, aes(x=ESCALAPE, fill=i_retintersem)) +
  geom_bar(position="fill") +
  geom_text(
    aes(label=signif(..count.. / tapply(..count.., ..x.., sum)[as.character(..x..)], digits=2)),
    stat="count",
    position=position_fill(vjust=0.5)) +
  scale_fill_manual(values = c("royalblue", "grey"))+
  guides(fill = guide_legend(title = "Retiro intersemestral")) +
  coord_flip()+
  labs(x= "Escala de pago", y="Proporci'on")

#por condicion de buen pagador en boletas desde las 2 a la 5

prop.table(table(datos_piloto2$cond_buenpagador))
table(datos_piloto2$i_retintersem, datos_piloto2$cond_buenpagador)
tabla4 <- prop.table(table(datos_piloto2$i_retintersem, datos_piloto2$cond_buenpagador),
margin=2)
tabla4
ggplot(datos_piloto2, aes(x=cond_buenpagador, fill=i_retintersem)) +
  geom_bar(position="fill") +
  geom_text(
    aes(label=signif(..count.. / tapply(..count.., ..x.., sum)[as.character(..x..)], digits=2)),
    stat="count",
    position=position_fill(vjust=0.5)) +
  scale_fill_manual(values = c("royalblue", "grey"))+
  guides(fill = guide_legend(title = "Retiro intersemestral")) +
  labs(x= "Condicion de buen pagador considerando las boletas 2 hasta 5 en el ciclo",
y="Proporci'on")

```

```
#i_cond_buenpag_bol2 - i_cond_buenpag_bol5
```

```
prop.table(table(datos_piloto2$i_retintersem, datos_piloto2$i_cond_buenpag_bol2),margin=2)
prop.table(table(datos_piloto2$i_retintersem, datos_piloto2$i_cond_buenpag_bol3),margin=2)
prop.table(table(datos_piloto2$i_retintersem, datos_piloto2$i_cond_buenpag_bol4),margin=2)
prop.table(table(datos_piloto2$i_retintersem, datos_piloto2$i_cond_buenpag_bol5),margin=2)
```

```
plot1 <- ggplot(datos_piloto2, aes(x=i_cond_buenpag_bol2, fill=i_retintersem)) +
  geom_bar(position="fill") +
  geom_text(
    aes(label=signif(..count.. / tapply(..count.., ..x.., sum)[as.character(..x..)], digits=2)),
    stat="count",
    position=position_fill(vjust=0.5)) +
  scale_fill_manual(values = c("royalblue", "grey"))+
  guides(fill = guide_legend(title = "Retiro intersemestral")) +
  labs(x= "i_cond_buenpag_bol2", y="Proporci\ 'on")
```

```
plot2 <- ggplot(datos_piloto2, aes(x=i_cond_buenpag_bol3, fill=i_retintersem)) +
  geom_bar(position="fill") +
  geom_text(
    aes(label=signif(..count.. / tapply(..count.., ..x.., sum)[as.character(..x..)], digits=2)),
    stat="count",
    position=position_fill(vjust=0.5)) +
  scale_fill_manual(values = c("royalblue", "grey"))+
  guides(fill = guide_legend(title = "Retiro intersemestral")) +
  labs(x= "i_cond_buenpag_bol3", y="Proporci\ 'on")
```

```
plot3 <- ggplot(datos_piloto2, aes(x=i_cond_buenpag_bol4, fill=i_retintersem)) +
  geom_bar(position="fill") +
  geom_text(
    aes(label=signif(..count.. / tapply(..count.., ..x.., sum)[as.character(..x..)], digits=2)),
    stat="count",
    position=position_fill(vjust=0.5)) +
  scale_fill_manual(values = c("royalblue", "grey"))+
  guides(fill = guide_legend(title = "Retiro intersemestral")) +
  labs(x= "i_cond_buenpag_bol4", y="Proporci\ 'on")
```

```
plot4 <- ggplot(datos_piloto2, aes(x=i_cond_buenpag_bol5, fill=i_retintersem)) +
  geom_bar(position="fill") +
  geom_text(
    aes(label=signif(..count.. / tapply(..count.., ..x.., sum)[as.character(..x..)], digits=2)),
```

```

  stat="count",
  position=position_fill(vjust=0.5)) +
  scale_fill_manual(values = c("royalblue", "grey"))+
  guides(fill = guide_legend(title = "Retiro intersemestral")) +
  labs(x= "i_cond_buenpag_bol5", y="Proporci\ 'on")

grid.arrange(plot1, plot2, plot3, plot4, ncol=2, nrow=2)

#tipocicloing
table(datos_piloto2$i_retintersem,datos_piloto2$tipocicloing)
prop.table(table(datos_piloto2$i_retintersem, datos_piloto2$tipocicloing),margin=2)

ggplot(datos_piloto2, aes(x=tipocicloing, fill=i_retintersem)) +
  geom_bar(position="fill") +
  geom_text(
    aes(label=signif(..count.. / tapply(..count.., ..x.., sum)[as.character(..x..)], digits=2)),
    stat="count",
    position=position_fill(vjust=0.5)) +
  scale_fill_manual(values = c("royalblue", "grey"))+
  guides(fill = guide_legend(title = "Retiro intersemestral")) +
  labs(x= "tipocicloing", y="Proporci\ 'on")

#####variables cuantitativas#####
#boxplot

p1<-ggplot(datos_piloto2, aes(x=i_retintersem, y=CRAEST, fill=i_retintersem)) +
  scale_fill_manual(values=c("royalblue", "grey")) +
  geom_boxplot(outlier.colour="red", outlier.shape=8,
              outlier.size=4)+
  guides(fill = guide_legend(title = "Retiro intersemestral"))+
  labs(x= "Retiro intersemestral", y="CRAEST")

p2<-ggplot(datos_piloto2, aes(x=i_retintersem, y=PRC_APDO_UE, fill=i_retintersem)) +
  scale_fill_manual(values=c("royalblue", "grey")) +
  geom_boxplot(outlier.colour="red", outlier.shape=8,
              outlier.size=4)+
  guides(fill = guide_legend(title = "Retiro intersemestral"))+
  labs(x= "Retiro intersemestral", y="Porcentaje de creditos aprobados acumulados
  en pregrado en su facultad")

grid.arrange(p1, p2, ncol=2, nrow=1)

```

```
#####imputacion de valores perdidos de las variables CRAEST #####

##### y PRC_APDO_UE #####

datos_piloto2[is.na(datos_piloto2$CRAEST), "CRAEST" ] <- mean(datos_piloto2$CRAEST, na.rm=T)

datos_piloto2[is.na(datos_piloto2$PRC_APDO_UE), "PRC_APDO_UE" ] <- mean(datos_piloto2$PRC_APDO_UE,
na.rm=T)

summary(datos_piloto2)

#correlaciones de variables numericas
data_num=data.frame(datos_piloto2[,8:11])
summary(data_num)
corr_matrix<-cor(data_num)
corrplot(corr_matrix, type="upper")

#####diagramas de dispersion#####

pairs(~ CRAEST + PRC_APDO_UE + PRC_RET + total_cremat, data = data_num)
Set1<-data.frame(CRAEST , PRC_APDO_UE , PRC_RET , total_cremat ); Set1
pairs.panels(x = Set1, ellipses = FALSE, lm = TRUE, method = "pearson", hist.col = "#CD1076")

#####retirando la covariable tipo de ciclo#####
datos_piloto2= datos_piloto2[-18]
head(datos_piloto2)

#####MODELO 1 GLM#####

modelo1 <- glm(i_retintersem ~ ., data = datos_piloto2, family = binomial("logit"))
summary(modelo1)
par(mfcol = c(2, 2))
plot(modelo1)

modelo2= glm(i_retintersem ~ especialidad + modalidad_ingreso + ESCALAPE +
CRAEST + PRC_APDO_UE + PRC_RET + i_cond_buenpag_bol4 + i_cond_buenpag_bol5 , data = datos_piloto2,
family = binomial("logit"))
summary(modelo2)

table(datos_piloto2$modalidad_ingreso)

#anova(modelo2, test = "Chi")
```

```

#anova(modelo1,modelo2)

modelo3= glm(i_retintersem ~ especialidad + modalidad_ingreso + ESCALAPE + PRC_APDO_UE + PRC_RET
+ i_cond_buenpag_bol4 + i_cond_buenpag_bol5 , data = datos_piloto2, family = binomial("logit"))
#anova(modelo3, test = "Chi")
summary(modelo3)

#modelo escogido es el modelo 2
#modelo2: i_retintersem ~ especialidad + modalidad_ingreso + ESCALAPE + CRAEST
#+ PRC_APDO_UE + PRC_RET + i_cond_buenpag_bol4 + i_cond_buenpag_bol5

modelo2
#calculo de odds para modelo 2 seleccionado
exp(coef(modelo2))
#calculo de intervalos de confianza para modelo 2 seleccionado
exp(confint(modelo2,level = 0.95))

modelo.null <- glm(i_retintersem ~ 1, binomial, datos_piloto2)
anova(modelo.null, modelo2, test = "Chi")

R2=(modelo1>null.deviance - modelo1$deviance) / modelo1>null.deviance
R2
modelo1
#calculo de odds para modelo 2 seleccionado
exp(coef(modelo1))
#calculo de intervalos de confianza para modelo 2 seleccionado
exp(confint(modelo1,level = 0.95))

##### analisis de variables cuantitativas relacionadas#####
modelo_analisis <- glm(i_retintersem ~ datos_piloto2$CRAEST + datos_piloto2$PRC_APDO_UE ,
data = datos_piloto2, family = binomial("logit"))
summary(modelo_analisis)
model.lasso_analisis <- bayesreg( i_retintersem ~ datos_piloto2$CRAEST +
datos_piloto2$PRC_APDO_UE , datos_piloto2 ,model= 'binomial', prior='lasso',
n.samples=15000, burnin=2000 )
summary(model.lasso_analisis)
##### los dos salen significativos#####

##### Estimacion priori Horseshoe model binomial #####

t <- proc.time() # Inicia el cronometro
start_time=Sys.time()

```

```

model.hs <- bayesreg( i_retintersem ~ . , datos_piloto2 ,model= 'binomial', prior='hs',
n.samples=15000, burnin=2000)
model.hs2 <- bayesreg( i_retintersem ~ modalidad_ingreso + CRAEST + PRC_APDO_UE + PRC_RET +
i_cond_buenpag_bol4 + i_cond_buenpag_bol5 , datos_piloto2 ,model= 'binomial', prior='hs',
n.samples=15000, burnin=2000 )
# Horseshoe regression

proc.time()-t # Detiene el cronometro
end_time=Sys.time();
diferencia_time=end_time - start_time
diferencia_time

summary(model.hs,display.OR = TRUE)
summary(model.hs,display.OR = FALSE)

summary(model.hs)
summary(model.hs2)
model_retiro.hs.s <- summary(model.hs)
model_retiro.hs.s$p.r2
model_retiro.hs.s$mu.coef
model_retiro.hs.s$med.OR
model_retiro.hs.s$se.0
model_retiro.hs.s$CI.OR

##### calculo Rhat
beta=model.hs$beta;
beta_o=model.hs$beta0;
M=model.hs$n.samples
Rhat(beta_o[2001:M])
for (i in 1:51) {
  Rhat(beta[i,2001:M])
}

##### Graficos#####

#####Estimacion priori Horseshoe #####

### Graficos de resumen - Beta ###

par(mfrow=c(3,3))
plot(density(beta_o[2001:M]),main ='Posteriori density of Beta0',xlab=NA,ylab=NA)
for (i in 1:8) {

```

```

    plot(density(beta[i,2001:M]),main = 'Posteriori density of Beta',xlab=NA,ylab=NA)
  }
dev.off()

#####graficos de convergencia#####
par(mfrow=c(3,3))

ts.plot(beta_o[2001:M],main='Trace of beta0',xlab=NA,ylab=NA)
for (i in 1:8) {
  ts.plot(beta[i,2001:M],main='Trace of beta1',xlab=NA,ylab=NA)
}
dev.off()
par(mfrow=c(3,3))
for (i in 9:17) {
  ts.plot(beta[i,2001:M],main='Trace of beta1',xlab=NA,ylab=NA)
}
dev.off()
par(mfrow=c(3,3))
for (i in 18:26) {
  ts.plot(beta[i,2001:M],main='Trace of beta1',xlab=NA,ylab=NA)
}
dev.off()
par(mfrow=c(3,3))
for (i in 27:35) {
  ts.plot(beta[i,2001:M],main='Trace of beta1',xlab=NA,ylab=NA)
}
dev.off()
par(mfrow=c(3,3))
for (i in 36:44) {
  ts.plot(beta[i,2001:M],main='Trace of beta1',xlab=NA,ylab=NA)
}
dev.off()
par(mfrow=c(3,3))
for (i in 45:51) {
  ts.plot(beta[i,2001:M],main='Trace of beta1',xlab=NA,ylab=NA)
}
dev.off()

##### Estimacion priori Horseshoe + model binomial #####

t2 <- proc.time() # Inicia el cronometro
start_time2=Sys.time()

```

```

model.hs_mas <- bayesreg( i_retintersem ~ . , datos_piloto2 ,model= 'binomial', prior='hs+',
n.samples=15000, burnin=2000 )
# Horseshoe + regression

proc.time()-t2 # Detiene el cronometro
end_time2=Sys.time();
diferencia_time2=end_time2 - start_time2
diferencia_time2

summary(model.hs_mas)
model_retiro.hsmas.s <- summary(model.hs_mas)

summary(model.hs_mas,display.OR = TRUE)
summary(model.hs_mas,display.OR = FALSE)

model_retiro.hsmas.s$p.r2
model_retiro.hsmas.s$mu.coef
model_retiro.hsmas.s$med.OR
model_retiro.hsmas.s$se.coef
model_retiro.hsmas.s$se.0
model_retiro.hsmas.s$CI.OR

#####calculo Rhat
beta=model.hs_mas$beta;
beta_o=model.hs_mas$beta0;
M=model.hs_mas$n.samples
Rhat(beta_o[2001:M])
for (i in 1:51) {
  Rhat(beta[i,2001:M])
}

##### Estimacion priori Lasso model binomial #####

t3 <- proc.time() # Inicia el cronometro
start_time3=Sys.time()
model.lasso <- bayesreg( i_retintersem ~ . , datos_piloto2 ,model= 'binomial', prior='lasso',
n.samples=15000, burnin=2000 )
# Horseshoe + regression

proc.time()-t3 # Detiene el cronometro
end_time3=Sys.time();
diferencia_time3=end_time3 - start_time3

```

```
diferencia_time3
```

```
summary(model.lasso)
model.lasso.s <- summary(model.lasso)
summary(model.lasso,display.OR = TRUE)
summary(model.lasso,display.OR = FALSE)
model.lasso.s$p.r2
model.lasso.s$mu.coef
model.lasso.s$med.OR
model.lasso.s$se.coef
model.lasso.s$se.0
model.lasso.s$CI.OR
```

```
#####calculo Rhat
```

```
beta=model.lasso$beta;
beta_o=model.lasso$beta0;
M=model.lasso$n.samples
Rhat(beta_o[2001:M])
for (i in 1:51) {
  Rhat(beta[i,2001:M])
}
```



## Apéndice B

# Resultados adicionales de la aplicación a BostonHousing

### B.1. Análisis descriptivo

El siguiente gráfico qqplot muestra que la variable respuesta, valor medio de las viviendas ocupadas por sus propietarios en miles de dólares, parece no tener una distribución normal.

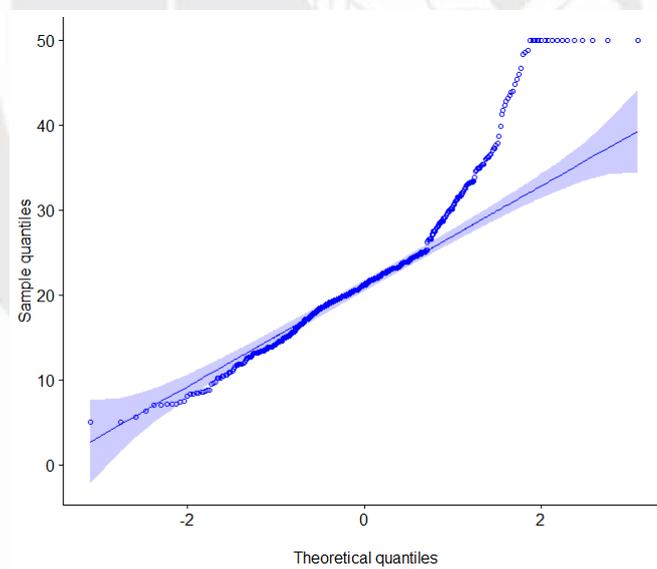


Figura B.1: Qqplot de la variable valor medio de las viviendas ocupadas por sus propietarios en miles de dólares

Algunos diagramas de dispersión de pares entre la variable respuesta y las variables que presentan una alta o moderada correlación con ella que permiten comprender el conjunto de

datos gráficamente:

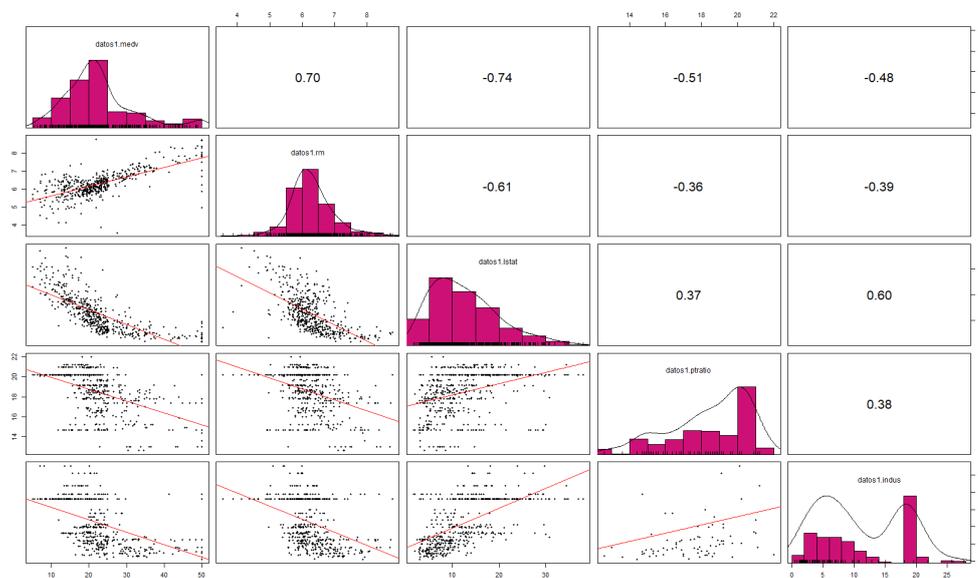


Figura B.2: Gráfico de dispersión de 5 covariables de la data “BostonHousing”

En el gráfico anterior se verifica la correlación directa o positiva fuerte entre las variables “*rm*” promedio de número de cuartos por vivienda y “*medv*” valor medio de las viviendas ocupadas por sus propietarios en miles de dólares, lo que significa que esta variable contribuirá en gran medida a la precisión en la estimación del valor de la vivienda. También se verifica la correlación negativa fuerte entre las variables “*lstat*” porcentaje del estatus más bajo de la población y “*medv*” valor medio de las viviendas ocupadas por sus propietarios en miles de dólares, lo que significa que esta variable también contribuirá en gran medida a la precisión en la estimación del valor de la vivienda. Además se puede observar una ligera correlación negativa entre las variables “*ptratio*” ratio alumnos-maestro por localidad y “*medv*” valor medio de las viviendas ocupadas por sus propietarios en miles de dólares; y entre “*indus*” proporción de acres de negocios no minoristas por ciudad y “*medv*” valor medio de las viviendas ocupadas por sus propietarios en miles de dólares.

Además, en la diagonal de la Figura anterior, se puede ver el histograma de cada una de las variables mencionadas. Y por encima de la diagonal, se observan los valores del coeficiente de correlación.

## Apéndice C

# Resultados adicionales de la aplicación a la deserción

### C.1. Análisis descriptivo

Analizando ahora la distribución de la variable sexo en la base de entrenamiento, vemos que la población mayoritariamente es masculina, 75 % son hombres y 25 % son mujeres.

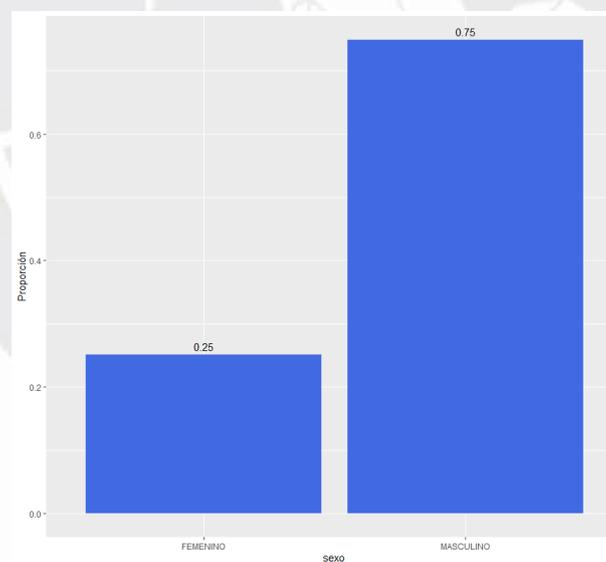


Figura C.1: Distribución de la variable sexo en la base de entrenamiento

Ahora si vemos la distribución de la variable de interés retiro intersemestral por sexo, se observa que el porcentaje de alumnos que no se matriculan en el semestre siguiente es similar tanto en hombres como en mujeres y es de aproximadamente 9 %.

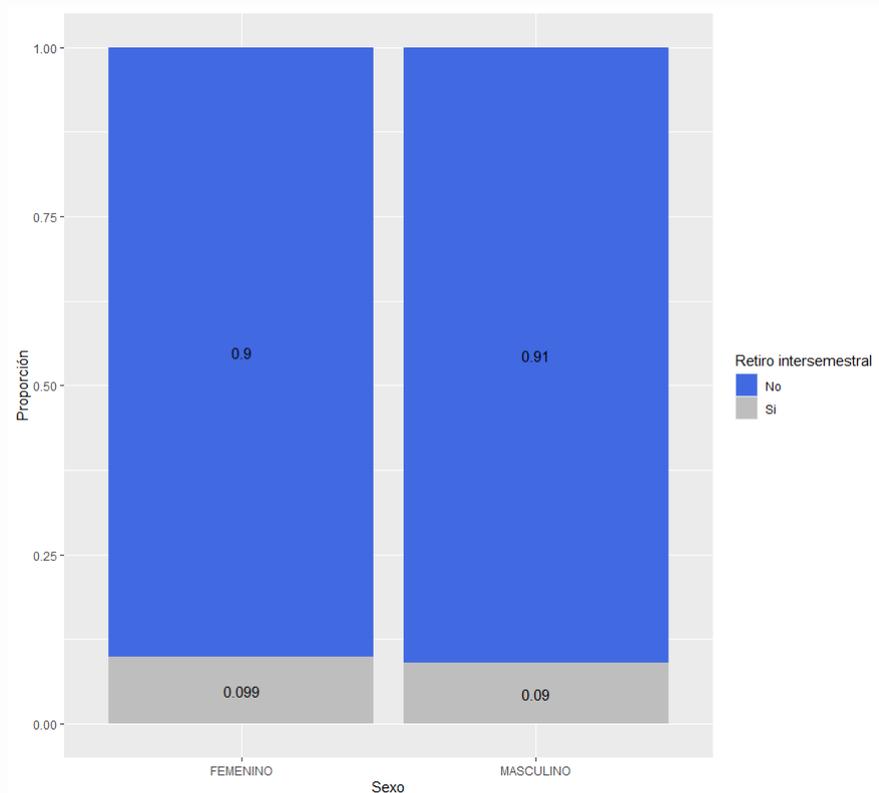


Figura C.2: Distribución de la variable de interés retiro intersemestral por sexo en la base de entrenamiento

Revisando la distribución de la especialidad del alumno en la población de entrenamiento tenemos que las especialidades Ingeniería Informática, Ingeniería Mecatrónica, Ingeniería Industrial e Ingeniería Civil son las que tienen mayor cantidad de estudiantes.

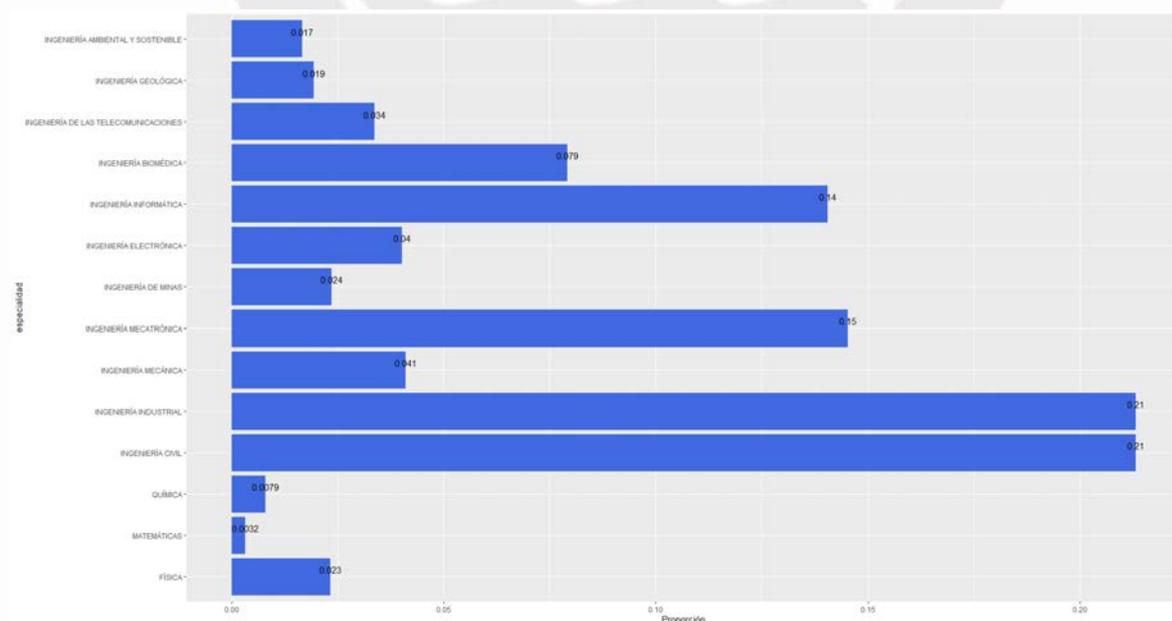


Figura C.3: Distribución de la variable especialidad en la base de entrenamiento

Y si analizamos la distribución de la variable de interés retiro intersemestral por especialidad vemos que las especialidades con mayor porcentaje de retiros son Ingeniería Ambiental y Sostenible, Ingeniería Geológica, Ingeniería Informática, Ingeniería Mecánica y Física. De estas destaca Ingeniería Informática pues tiene una considerable cantidad de alumnos en la población (531) de los cuales el 11 % se retiran (60).

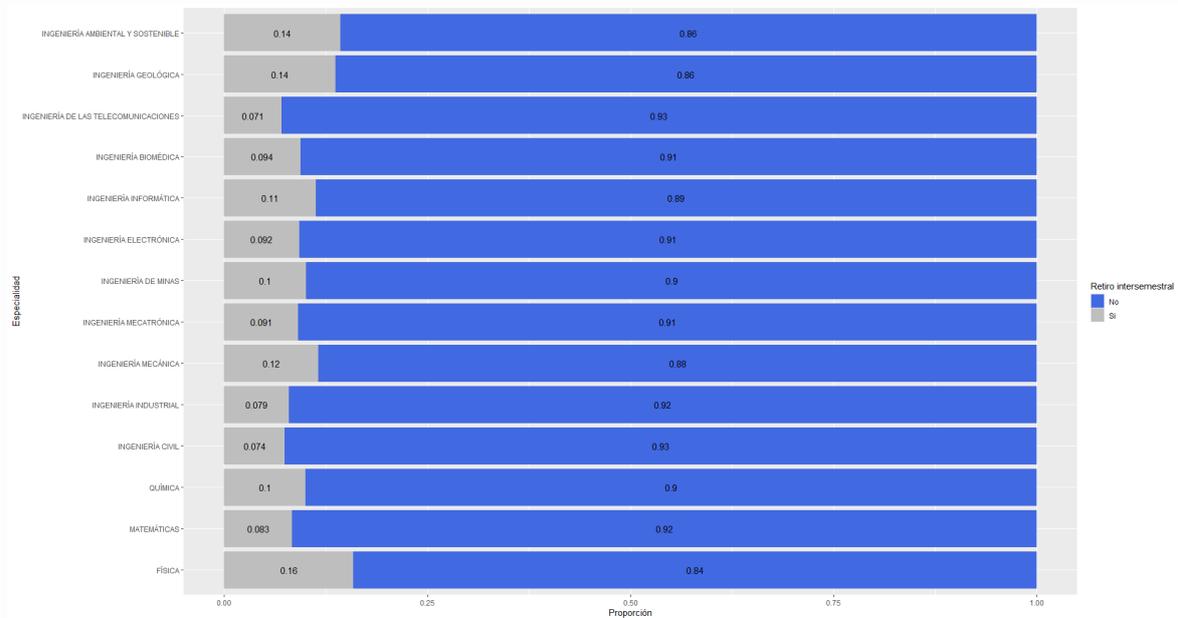


Figura C.4: Distribución de la variable de interés retiro intersemestral por especialidad en la base de entrenamiento

Analizando la distribución de la modalidad de admisión o de ingreso del alumno en la población de entrenamiento tenemos que las modalidades con mayor cantidad de estudiantes son “Evaluación de talento y exonerados”, “Ingreso por ser tercio superior universidad 1”, “Ingreso por la pre de la universidad 1” e “Ingreso en primera oportunidad”. Entre estas 4 modalidades suman el 82 % de los datos.

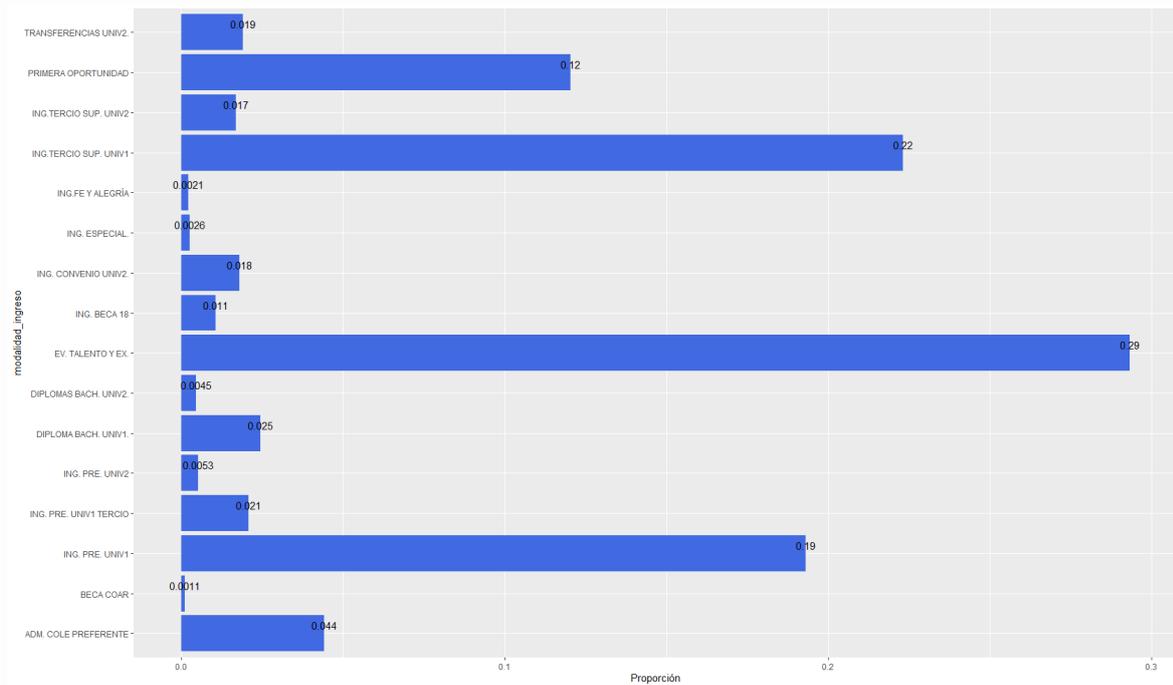


Figura C.5: Distribución de la variable modalidad de ingreso a la universidad en la base de entrenamiento

Considerando la distribución de la variable de interés retiro intersemestral por modalidad de ingreso vemos que la modalidad con mayor porcentaje de retiro intersemestral es “Diplomas bachillerato universidad 1” (19%) y dentro de las modalidades con mayor cantidad de alumnos, la modalidad “Ingreso por ser tercio superior universidad 1” presenta el mayor porcentaje de retiros (13%).

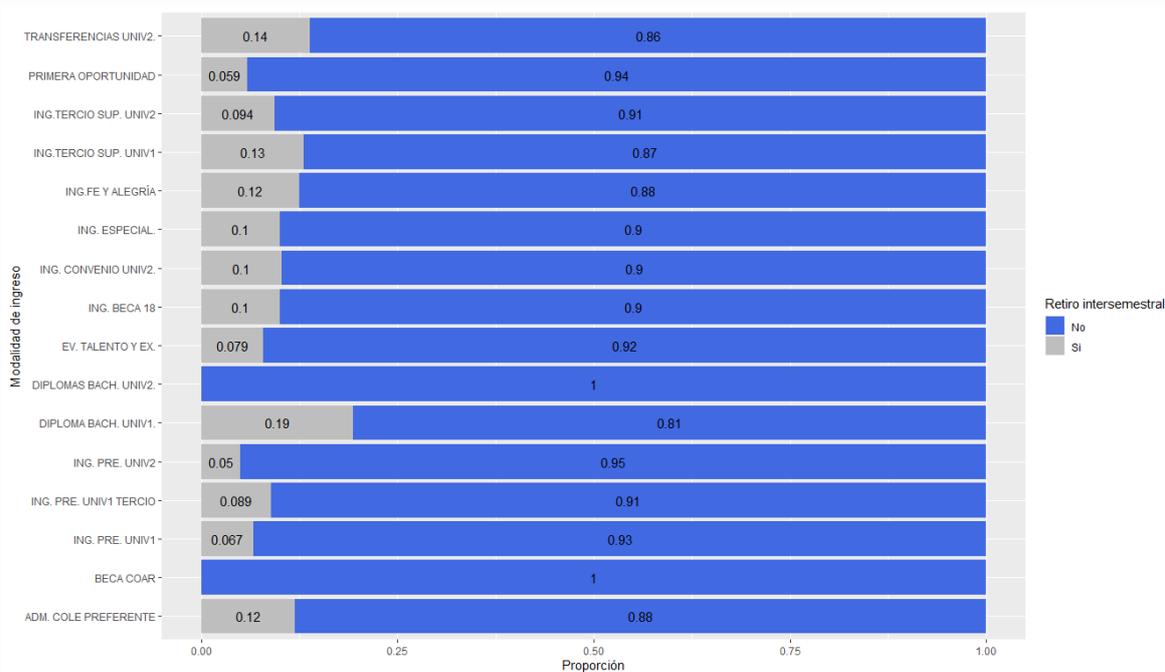


Figura C.6: Distribución de la variable de interés retiro intersemestral por modalidad de ingreso a la universidad en la base de entrenamiento

Analizando la variable condición de tener beca o descuento, solo el 5% de la población tiene algún beneficio como tener beca o descuento, y observando distribución de la variable de interés retiro intersemestral por condición de tener beca o descuento, se concluye que el porcentaje de alumnos que no se matriculan en el semestre siguiente es mayor en el grupo que tiene algún beneficio y es de 18%.

Retiro intersemestral	Condición de tener beca o descuento			
	no tiene beca o descuento		tiene beca o descuento	
	N	%	N	%
No	3268	91 %	163	82 %
Sí	314	9 %	35	18 %
Total	3582	95 %	198	5 %

Cuadro C.1: Tabla de la distribución de la variable de interés retiro intersemestral por Condición de tener beca o descuento en la base de entrenamiento

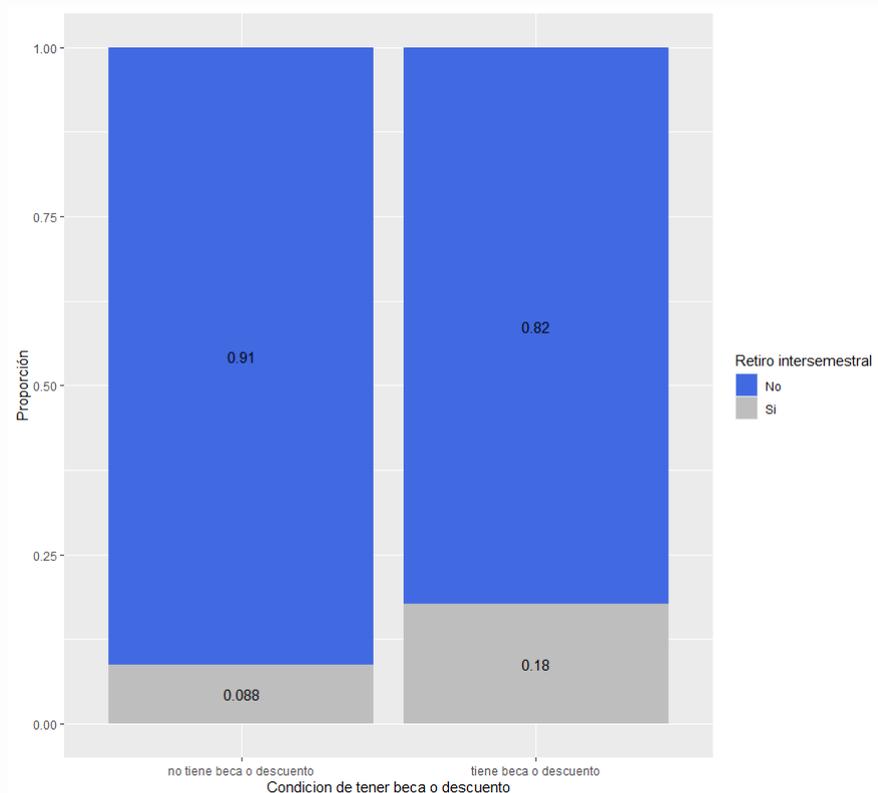


Figura C.7: Distribución de la variable de interés retiro intersemestral por condición de tener beca o descuento en la base de entrenamiento

Revisando la distribución de la variable escala de pago del alumno en la población de entrenamiento tenemos que las escalas G2 a G5 son las que tienen mayor cantidad de estudiantes, entre estas 4 escalas suman el 71 % de los datos.

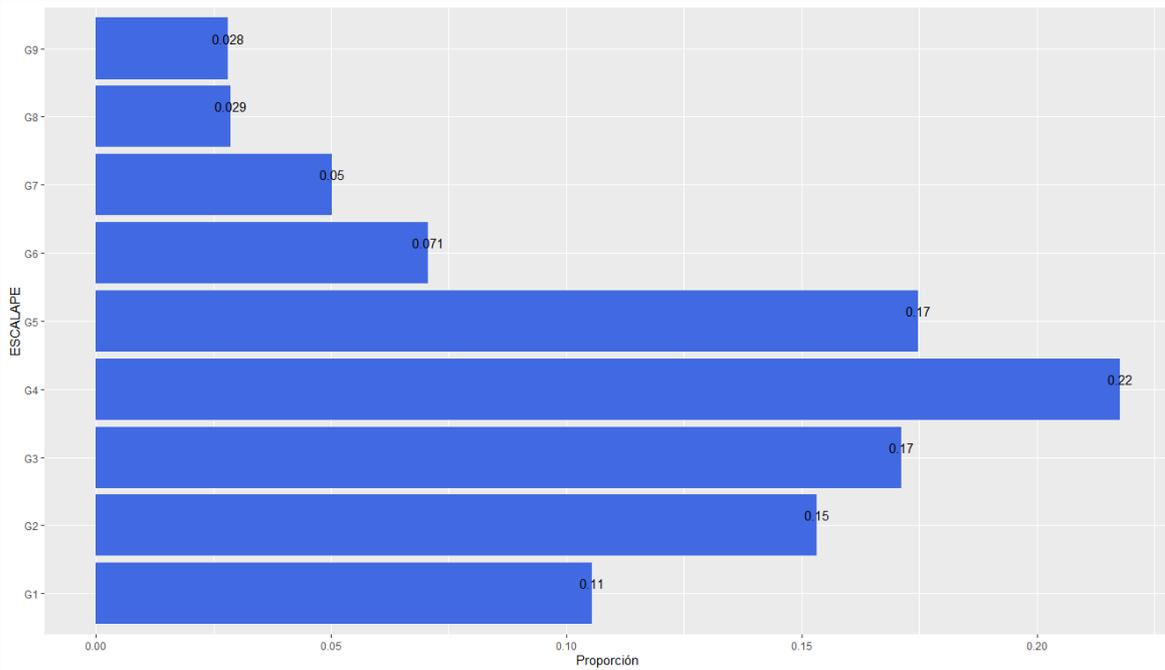


Figura C.8: Distribución de la variable escala de pago del alumno en la base de entrenamiento

Observando la distribución de la variable de interés retiro intersemestral por escala de pago del alumno, vemos que las escalas altas G8 y G9 presentan el mayor porcentaje de retiros (11%). La escala con menor proporción de retiros es la escala más baja, G1 (7.5%).

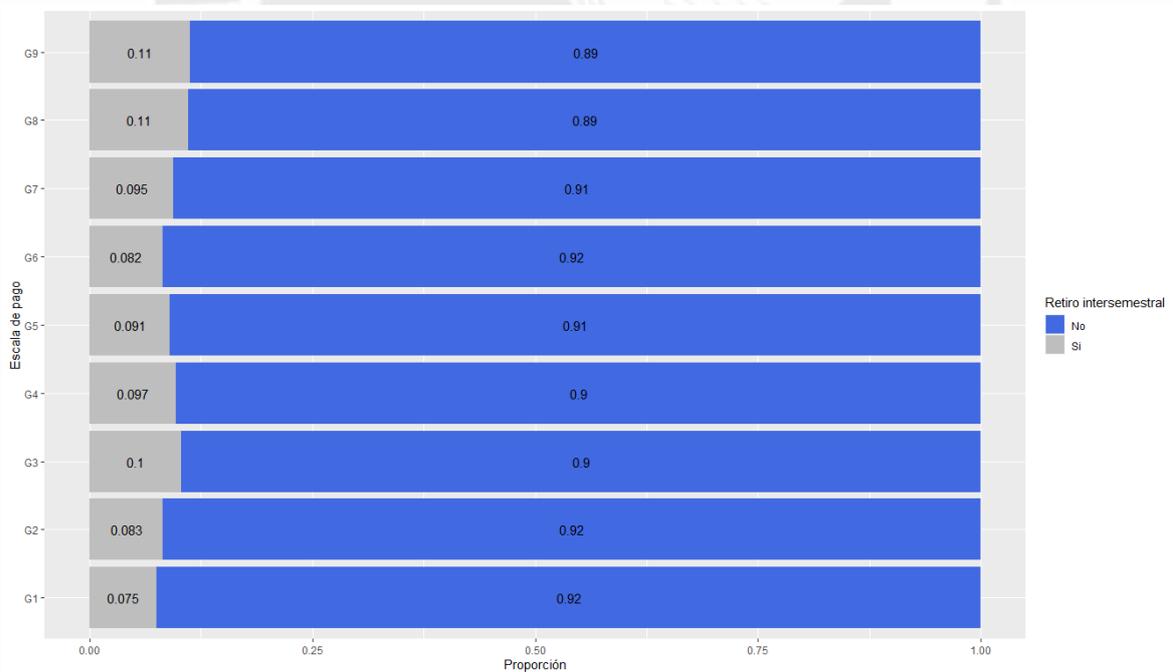


Figura C.9: Distribución de la variable de interés retiro intersemestral por escala de pago del alumno en la base de entrenamiento

Se tiene en la base de datos cuatro variables adicionales dicotómicas que se refieren a la

condición de buen pagador en cada una de las boletas 2, 3, 4 y 5. Revisando la distribución de la variable de interés retiro intersemestral por estas cuatro variables, se observa que entre el 23% al 28% de los alumnos que no fueron buenos pagadores en alguna de sus boletas se retiraron (no se matriculan en el siguiente semestre).

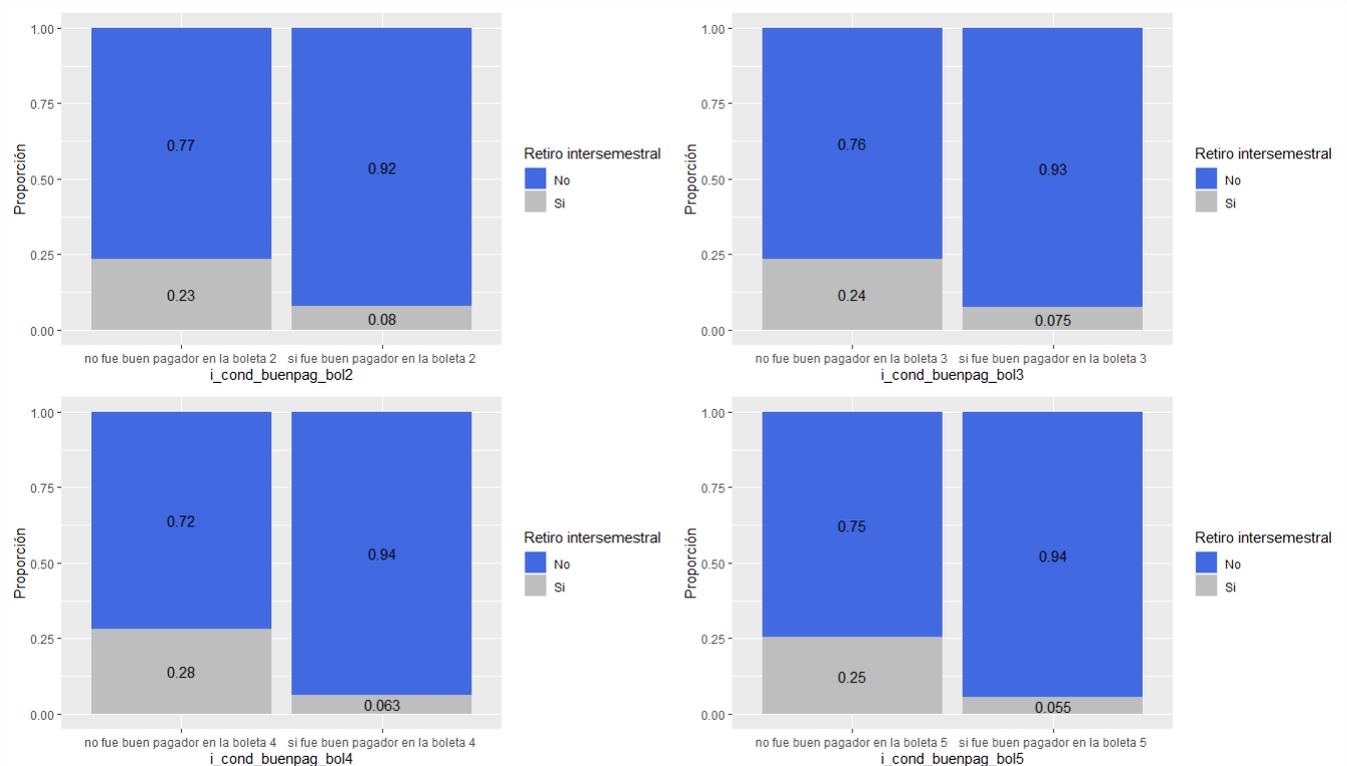


Figura C.10: Distribución de la variable de interés retiro intersemestral por condición de buen pagador en cada una de las boletas 2 a 5 en la base de entrenamiento

Analizando la distribución de las variables cuantitativas relacionadas al rendimiento académico en el primer ciclo de matrícula por la variable de interés retiro intersemestral se observa que el grupo de los alumnos que se retiraron presentan un menor valor de la mediana tanto en la variable CRAEST como el porcentaje de créditos aprobados acumulados en pregrado en la facultad.

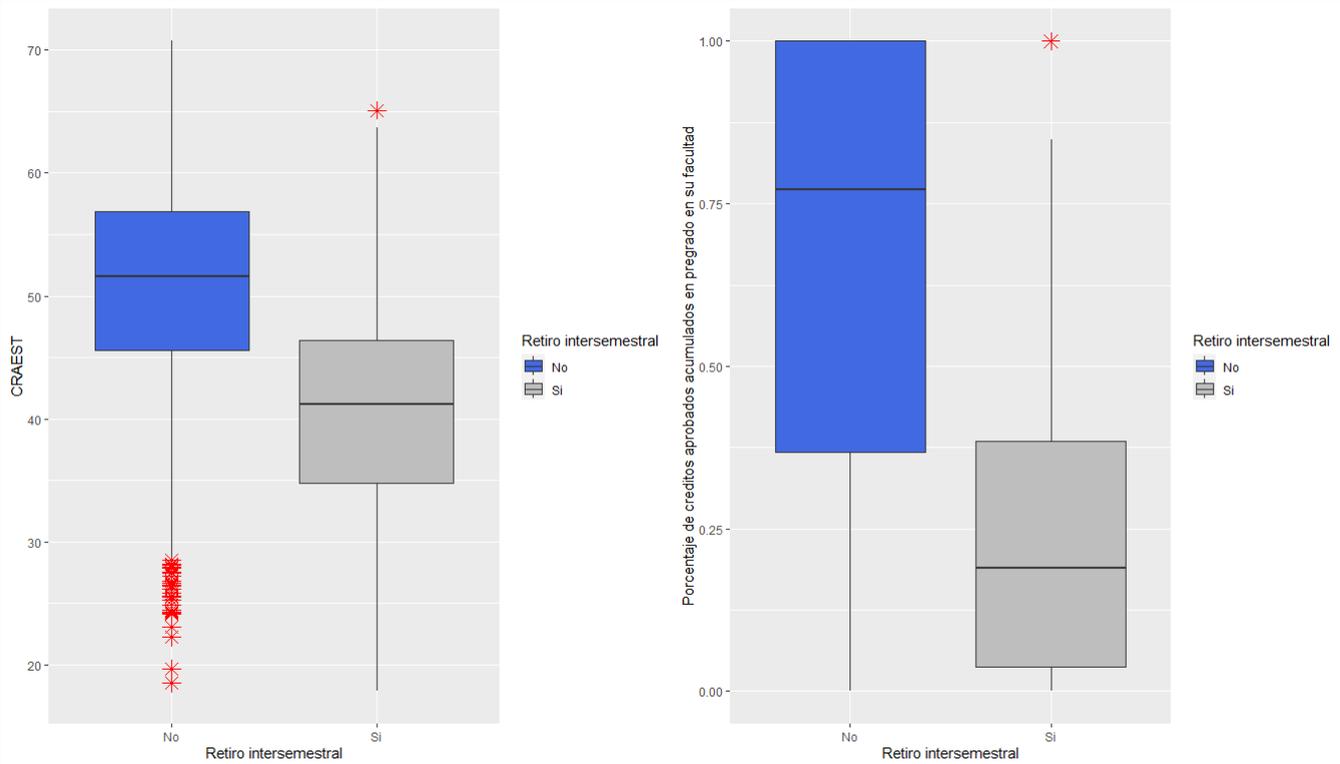


Figura C.11: Distribución de las variables CRAEST y porcentaje de créditos aprobados acumulados en pregrado en la facultad por la variable de interés retiro intersemestral en la base de entrenamiento

Las correlaciones entre las covariables cuantitativas CRAEST, porcentaje de créditos aprobados acumulados en pregrado en su facultad (PRC\_APDO\_UE), porcentaje de créditos retirados en relación a los matriculados en el ciclo (PRC\_RET) y total de créditos matriculados en el ciclo (total\_cremat) se muestran en el siguiente gráfico, en el que se observa que solo las variables relacionadas con el rendimiento académico en el primer ciclo CRAEST y porcentaje de créditos aprobados acumulados en pregrado en su facultad (PRC\_APDO\_UE) están fuertemente correlacionadas positivamente con un coeficiente de correlación de 0.88.

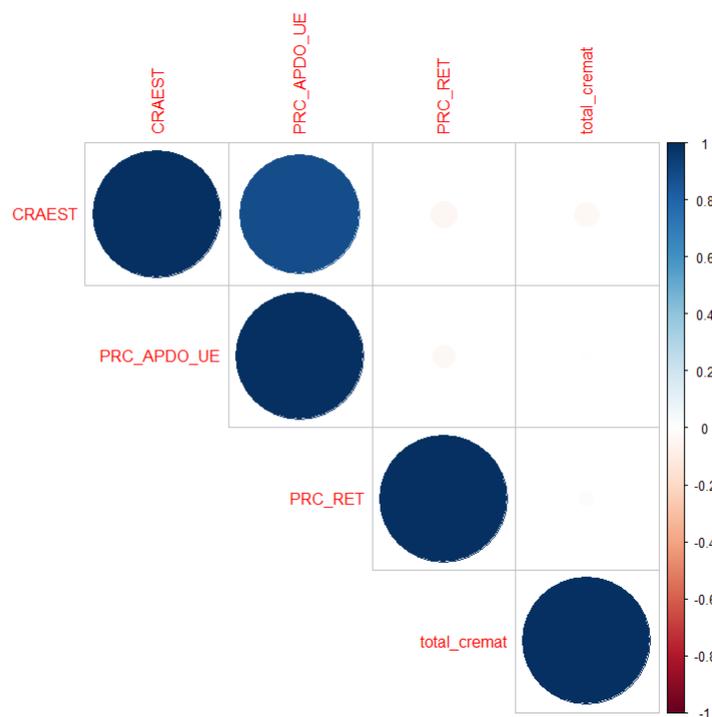


Figura C.12: Gráfico de matriz de correlaciones de las variables cuantitativas de la base de entrenamiento

A continuación se presentan los diagramas de dispersión de pares entre las variables cuantitativas de la base, se observa una alta correlación positiva solo entre las variables relacionadas con el rendimiento académico en el primer ciclo, CRAEST y porcentaje de créditos aprobados acumulados en pregrado en su facultad (PRC\_APDO\_UE), lo que significa que estas variables causan multicolinealidad al estar correlacionadas. También se verifica la correlación negativa débil o baja (-0.12) entre las variables CRAEST y porcentaje de créditos retirados (PRC\_RET). Los demás pares de variables presentan correlaciones negativas casi nulas.

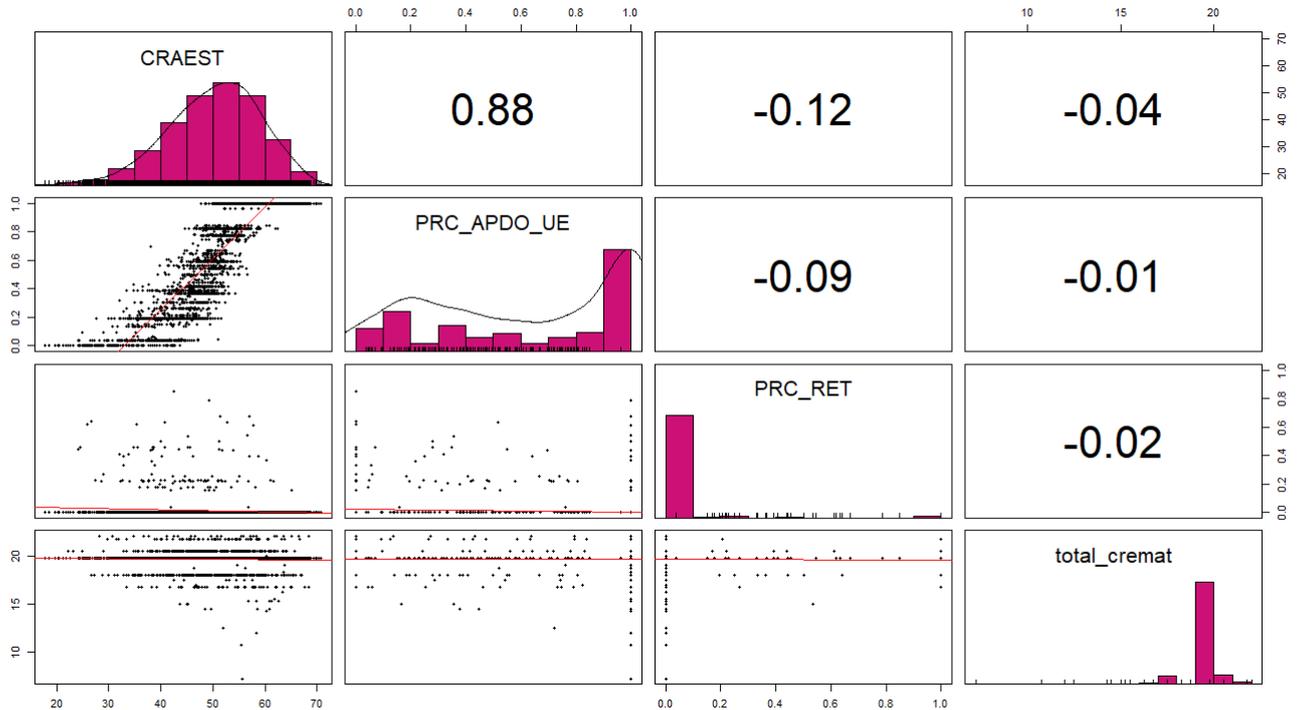


Figura C.13: Gráfico de dispersión de las 4 variables cuantitativas de la base de entrenamiento

Además, en la diagonal de la Figura anterior, se puede ver el histograma de cada una de las variables mencionadas. En estos, se observa que el porcentaje de créditos retirados (PRC\_RET) toma valores mayoritariamente cercano a 0 y la variable total de créditos matriculados toma valores mayoritariamente alrededor de 20. Asimismo, se observan los valores del coeficiente de correlación de cada par de variables.

## C.2. Resultados de regresión con todas las variables

A continuación se presentan los resultados de la ejecución del modelo logístico con todas las covariables, se muestran los estimadores de los coeficientes, el error estándar, los ratios de odds de los coeficientes y los intervalos de confianza de los ratios de odds.

En el siguiente cuadro se indican las variables “significativas”, al 95 % como “\*”, al 99 % como “\*\*” y al 99.9 % como “\*\*\*”.

Covariable	Parámetro	$\hat{\beta}$	Error Estándar	$\hat{OR} = \exp(\hat{\beta})$ razón de odds	$\exp(IC(95\%))$	p-value
Intercepto	$\beta_0$	4.384	2.458	80.16	(0.51; 7602.62)	0.074 .
cicmat2019-1	$\beta_{11}$	-0.250	0.301	0.778	(0.433; 1.419)	0.4069
cicmat2021-2	$\beta_{13}$	0.047	0.340	1.048	(0.536; 2.046)	0.8896
cicmat2022-1	$\beta_{14}$	0.042	0.312	1.043	(0.569; 1.946)	0.8923
especialidad MA- TEMÁTICAS	$\beta_{21}$	0.112	1.368	1.119	(0.041; 11.130)	0.9343
especialidad QUÍMICA	$\beta_{22}$	-0.756	0.808	0.469	(0.077; 2.238)	0.3698
especialidad ING.CIVIL	$\beta_{23}$	-1.340	0.398	0.261	0.122; 0.589)	0.0007 ***
especialidad ING. INDUSTRIAL	$\beta_{24}$	-1.441	0.404	0.236	(0.109; 0.537)	0.0003 ***
especialidad ING. MECÁNICA	$\beta_{25}$	-0.843	0.475	0.430	(0.169; 1.107)	0.0763
especialidad ING. MECATRÓNICA	$\beta_{26}$	-0.993	0.410	0.370	(0.168; 0.852)	0.0155 *
especialidad ING. DE MINAS	$\beta_{27}$	-1.266	0.575	0.281	(0.088; 0.860)	0.0278 *
especialidad ING. ELECTRÓNICA	$\beta_{28}$	-1.179	0.514	0.307	(0.1111; 0.845)	0.0219 *
especialidad ING. INFORMÁTICA	$\beta_{29}$	-0.801	0.404	0.448	(0.207; 1.021)	0.0475 *
especialidad ING. BIOMÉDICA	$\beta_{2(10)}$	-0.401	0.700	0.668	(0.153; 2.508)	0.5662
especialidad ING. TELECOMUNIC.	$\beta_{2(11)}$	-1.631	0.563	0.195	(0.062; 0.579)	0.0038 **

Cuadro C.2: Parámetros estimados por modelo de regresión lineal con mixtura de escala Gaussiana: regresión logística clásica

Covariable	Parámetro	$\hat{\beta}$	Error Estándar	$\hat{OR} = \exp(\hat{\beta})$ razón de odds	$\exp(\text{IC}(95\%))$	p-value
especialidad ING. GEOLÓGICA	$\beta_{2(12)}$	-0.884	0.559	0.412	(0.135; 1.227)	0.1138
especialidad ING. AMBIENTAL Y SOSTENIBLE	$\beta_{2(13)}$	-0.430	0.637	0.650	(0.179; 2.219)	0.5000
modalidad ingreso BECA COAR	$\beta_{31}$	-9.513	712.245	0.00002	(- ínf; 1.043e+31)	0.9893
modalidad ingreso ING. PRE. UNIV1	$\beta_{32}$	0.156	0.350	1.169	(0.596; 2.364)	0.6551
modalidad ingreso ING. PRE. UNIV1 TERCIO	$\beta_{33}$	0.542	0.560	1.720	(0.541; 4.963)	0.3323
modalidad ingreso ING. PRE. UNIV2	$\beta_{34}$	-1.905	1.461	0.148	(0.004; 1.904)	0.1922
modalidad ingreso DIPLOMA BACH. UNIV1	$\beta_{35}$	1.261	0.489	3.529	(1.344; 9.218)	0.0099 *
modalidad ingreso DIPLOMAS BACH. UNIV2	$\beta_{36}$	-13.09	325.895	0.000002	(2.388e-36; 57.278)	0.9679
modalidad ingreso EV. TALENTO Y EX.	$\beta_{37}$	0.260	0.348	1.297	(0.664; 2.614)	0.4540
modalidad ingreso ING. BECA 18	$\beta_{38}$	1.598	0.875	4.947	(8.261e-01; 26.450)	0.0678

Cuadro C.3: Parámetros estimados por modelo de regresión lineal con mixtura de escala Gaussiana: regresión logística clásica

Covariable	Parámetro	$\hat{\beta}$	Error Estándar	$\hat{OR} = \exp(\hat{\beta})$ razón de odds	$\exp(IC(95\%))$	p-value
modalidad ingreso ING. CONVENIO UNIV2.	$\beta_{39}$	0.458	0.775	1.581	(0.365; 7.900)	0.5542
modalidad ingreso ING. ESPECIAL.	$\beta_{3(10)}$	1.840	2.433	6.296	(0.077; 214.755)	0.4494
modalidad ingreso ING. FE Y ALEGRÍA	$\beta_{3(11)}$	2.016	1.245	7.515	(0.332; 67.144)	0.1054
modalidad ingreso ING. TERCIO SUP. UNIV1	$\beta_{3(12)}$	0.634	0.324	1.886	(1.017; 3.643)	0.049 *
modalidad ingreso ING. TERCIO SUP. UNIV2	$\beta_{3(13)}$	0.251	0.817	1.285	(0.260; 6.749)	0.7582
modalidad ingreso PRIMERA OPORTUNIDAD	$\beta_{3(14)}$	0.494	0.379	1.639	(0.784; 3.485)	0.1921
modalidad ingreso TRANSFERENCIAS UNIV2	$\beta_{3(15)}$	0.516	0.750	1.676	(0.403; 7.968)	0.4913
cond_colenolimacallao Sí	$\beta_{41}$	0.125	0.157	1.134	(0.829; 1.540)	0.4255
LBENEFICIO Sí	$\beta_{51}$	0.149	0.396	1.161	(0.518; 2.463)	0.7064
escala de pago G2	$\beta_{61}$	0.408	0.298	1.504	(0.843; 2.732)	0.1715
escala de pago G3	$\beta_{62}$	0.781	0.287	2.184	(1.257; 3.889)	0.0065 **
escala de pago G4	$\beta_{63}$	0.780	0.289	2.182	(1.253; 3.903)	0.0069 **
escala de pago G5	$\beta_{64}$	0.725	0.302	2.064	(1.153; 3.780)	0.0163 *
escala de pago G6	$\beta_{65}$	0.806	0.370	2.240	(1.077; 4.624)	0.0294 *

Cuadro C.4: Parámetros estimados por modelo de regresión lineal con mixtura de escala Gaussiana: regresión logística clásica

Covariable	Parámetro	$\hat{\beta}$	Error Estándar	$\hat{OR} = \exp(\hat{\beta})$ razón de odds	$\exp(IC(95\%))$	p-value
escala de pago G7	$\beta_{66}$	1.054	0.398	2.871	(1.299; 6.228)	0.0080 **
escala de pago G8	$\beta_{67}$	0.965	0.439	2.625	(1.081; 6.115)	0.02802 *
escala de pago G9	$\beta_{68}$	1.179	0.454	3.251	(1.302; 7.809)	0.0095 *
CRAEST	$\beta_8$	-0.067	0.016	0.934	(0.905; 0.964)	3.28e-05 ***
PRC_APDO_UE	$\beta_9$	-1.885	0.464	0.151	(0.060; 0.375)	4.98e-05 ***
PRC_RET	$\beta_{10}$	4.093	0.342	59.952	(31.226; 120.108)	<2e-16 ***
total_cremat	$\beta_{11}$	-0.086	0.115	0.9167	(0.742; 1.166)	0.4532
cond_buenpagador Sí	$\beta_{12(1)}$	0.422	0.342	1.525	(0.796; 3.056)	0.2168
i_cond_buen pag_bol2 Sí	$\beta_{13(1)}$	0.233	0.278	1.262	(0.738; 2.210)	0.4028
i_cond_buen pag_bol3 Sí	$\beta_{14(1)}$	-0.033	0.272	0.9666	(0.569; 1.659)	0.9008
i_cond_buen pag_bol4 Sí	$\beta_{15(1)}$	-1.010	0.2684	0.364	(0.212; 0.610)	0.0001 ***
i_cond_buen pag_bol5 Sí	$\beta_{16(1)}$	-1.352	0.301	0.258	(0.139; 0.456)	7.30e-6 ***
sexo (masculino)	$\beta_{18(1)}$	-0.092	0.164	0.9113	(0.662; 1.261)	0.5714

Cuadro C.5: Parámetros estimados por modelo de regresión lineal con mixtura de escala Gaussiana: regresión logística clásica

En los cuadros siguientes se muestra el detalle de la estimación aplicando un modelo MEG con priori de Horseshoe + con todas las variables, se aprecian las estimaciones de los parámetros de regresión, los errores estándar, los cuantiles de los parámetros que definen los intervalos de credibilidad y el Rhat de los coeficientes.

Cabe resaltar aquí que para que una variable sea considerada “significativa”, su intervalo de credibilidad al 95 % para la exponenciación del coeficiente estimado no debería contener a 1,

hecho que se señala como “\*\*\*”. Adicionalmente, el paquete “bayesreg”, considera también la posibilidad de tener variables significativas considerando intervalos de credibilidad al 75 %, esto se señala como “\*”.

Covariable	Parámetro	$\hat{\beta}$	$exp(\hat{\beta})$	exp(Error Estándar)	exp(IC(95 %))	Rhat
Intercepto	$\beta_0$	2.3065	10.039	28.929	(0.7863; 114.1891)	1.0000
cicmat2019-1	$\beta_{11}$	-0.0013	0.9986	0.0841	(0.7406; 1.0705)	0.9999
cicmat2021-2	$\beta_{12}$	-0.0003	0.9997	0.1119	(0.7374; 1.1761)	1.0000
cicmat2022-1	$\beta_{13}$	0.0016	1.0016	0.1155	(0.9380; 1.3907)	0.9999
especialidad MATEMÁTICAS	$\beta_{21}$	0.0018	1.0018	1.0179	(0.4907; 4.4809)	1.0001
especialidad QUÍMICA	$\beta_{22}$	0.0003	1.0003	0.2976	(0.6227; 1.7893)	1.0001
especialidad ING. CIVIL	$\beta_{23}$	-0.0071	0.9929	0.1237	(0.5454; 1.0304)	1.0001
especialidad ING. INDUSTRIAL	$\beta_{24}$	-0.0100	0.9900	0.1237	(0.5246; 1.0294)	0.9999
especialidad ING. MECÁNICA	$\beta_{25}$	0.0002	1.0002	0.1331	(0.8314; 1.3532)	1.0003
especialidad ING. MECATRÓNICA	$\beta_{26}$	0.0000	1.0000	0.0801	(0.8581; 1.1723)	1.0000
especialidad ING. DE MINAS	$\beta_{27}$	-0.0003	0.9996	0.1571	(0.6308; 1.2469)	1.0004
especialidad ING. ELECTRÓNICA	$\beta_{28}$	-0.0001	0.9999	0.1282	(0.7225; 1.2254)	1.0000
especialidad ING. INFORMÁTICA	$\beta_{29}$	0.0008	1.0008	0.1233	(0.9266; 1.4100)	1.0003
especialidad ING. BIOMÉDICA	$\beta_{2(10)}$	0.0048	1.0049	0.3294	(0.9323; 2.2240)	1.0000

Cuadro C.6: Parámetros estimados por modelo de regresión logística con mixtura de escala Gaussiana con regularización bayesiana usando una priori de Horseshoe +

Covariable	Parámetro	$\hat{\beta}$	$exp(\hat{\beta})$	exp(Error Estándar)	exp(IC(95 %))	Rhat
especialidad ING. DE LAS TELECOMUNIC.	$\beta_{2(11)}$	-0.0051	0.9948	0.1913	(0.3413; 1.0913)	0.9999
especialidad ING. GEOLÓGICA	$\beta_{2(12)}$	0.0001	1.0001	0.1826	(0.7655; 1.4814)	1.0009
especialidad ING. AMBIENTAL Y SOSTENIBLE	$\beta_{2(13)}$	0.0014	1.0014	0.3176	(0.8331; 2.0781)	0.9999
modalidad de ingreso BECA COAR	$\beta_{31}$	-0.0033	0.9966	1.9181	(0.0000; 7.5190)	0.9999
modalidad ingreso ING. PRE. UNIV1	$\beta_{32}$	-0.0026	0.9973	0.1011	(0.6784; 1.0749)	1.0000
modalidad ingreso ING. PRE. UNIV1 TERCIO	$\beta_{33}$	0.0001	1.0001	0.2221	(0.6902; 1.5612)	1.0004
modalidad ingreso ING. PRE. UNIV2	$\beta_{34}$	-0.0223	0.9779	0.3341	(0.0692; 1.3792)	1.0002
modalidad ingreso DIPLOMA BACH. UNIV1 (*IC al 75 %)	$\beta_{35}$	0.7727	2.1656	1.0419	(0.9786; 5.0631)	1.0001
modalidad ingreso DIPLOMA BACH. UNIV2	$\beta_{36}$	-0.0262	0.9741	0.3745	(0.0000; 1.4681)	0.9999

Cuadro C.7: Parámetros estimados por modelo de regresión logística con mixtura de escala Gaussiana con regularización bayesiana usando una priori de Horseshoe +

Covariable	Parámetro	$\hat{\beta}$	$exp(\hat{\beta})$	exp(Error Estándar)	exp(IC(95 %))	Rhat
modalidad ingreso EV. TALENTO Y EX.	$\beta_{37}$	-0.0012	0.9987	0.0896	(0.7288; 1.0804)	0.9999
modalidad ingreso ING. BECA 18	$\beta_{38}$	0.0081	1.0081	0.7878	(0.7696; 3.8579)	1.0001
modalidad ingreso ING. CONVENIO UNIV2.	$\beta_{39}$	0.0058	1.0058	0.4409	(0.8137; 2.5422)	1.0000
modalidad ingreso ING. ESPECIAL.	$\beta_{3(10)}$	0.0006	1.0006	1.3183	(0.2996; 5.4677)	1.0010
modalidad ingreso ING. FE Y ALEGRÍA.	$\beta_{3(11)}$	0.0035	1.0035	1.0046	(0.3997; 4.3380)	1.0000
modalidad ingreso ING. TERCIO SUP. UNIV1	$\beta_{3(12)}$	0.0403	1.0411	0.1934	(0.9674; 1.7258)	0.9999
modalidad ingreso ING. TERCIO SUP. UNIV2	$\beta_{3(13)}$	0.0022	1.0022	0.3590	(0.7645; 2.1721)	1.0000
modalidad ingreso PRIMERA OPORTUNIDAD	$\beta_{3(14)}$	0.0000	1.0000	0.1062	(0.8231; 1.2397)	0.9999
modalidad ingreso TRANSFERENCIAS UNIV2.	$\beta_{3(15)}$	0.0217	1.0219	0.5716	(0.8646; 3.1055)	1.0002
cond colenolima-callao Sí	$\beta_{41}$	0.0024	1.0024	0.0904	(0.9087; 1.2632)	0.9999
I.BENEFICIO Sí	$\beta_{51}$	0.0050	1.0050	0.2155	(0.8252; 1.6702)	0.9999
escala de pago G2	$\beta_{61}$	-0.0003	0.9996	0.0815	(0.7395; 1.0593)	1.0001
escala de pago G3	$\beta_{62}$	0.0000	1.0000	0.0663	(0.927; 1.1871)	1.0000

Cuadro C.8: Parámetros estimados por modelo de regresión logística con mixtura de escala Gaussiana con regularización bayesiana usando una priori de Horseshoe +

Covariable	Parámetro	$\hat{\beta}$	$exp(\hat{\beta})$	exp(Error Estándar)	exp(IC(95 %))	Rhat
escala de pago G4	$\beta_{63}$	0.0001	1.0001	0.0814	(0.9397; 1.2591)	1.0005
escala de pago G5	$\beta_{64}$	0.0001	1.0001	0.0814	(0.9291; 1.2482)	0.9999
escala de pago G6	$\beta_{65}$	0.0000	1.0000	0.0959	(0.8360; 1.2121)	0.9999
escala de pago G7	$\beta_{66}$	0.0002	1.0002	0.1519	(0.8889; 1.4846)	0.9999
escala de pago G8	$\beta_{67}$	0.0000	1.0000	0.1527	(0.8131; 1.4118)	1.0001
escala de pago G9	$\beta_{68}$	0.0005	1.0005	0.2094	(0.8752; 1.6962)	1.0005
CRAEST **	$\beta_8$	-0.0648	0.9372	0.0165	(0.9046; 0.9696)	1.0000
PRC_APDO_UE **	$\beta_9$	-1.6617	0.1898	0.1179	(0.0723; 0.5345)	0.9999
PRC_RET **	$\beta_{10}$	3.9706	53.0169	17.3482	(29.4047; 97.4097)	1.0000
total_cremat	$\beta_{11}$	0.0000	1.0000	0.0510	(0.9159; 1.116)	0.9999
cond_buenpagador Sí	$\beta_{12(1)}$	0.0004	1.0004	0.1610	(0.8012; 1.4325)	1.0002
i_cond_buenpag_bol2 Sí	$\beta_{13(1)}$	0.0018	1.0018	0.1282	(0.8425; 1.3453)	1.0008
i_cond_buenpag_bol3 Sí	$\beta_{14(1)}$	0.0002	1.0002	0.1120	(0.8213; 1.2606)	1.0000
i_cond_buenpag_bol4 Sí **	$\beta_{15(1)}$	-0.6856	0.5037	0.1216	(0.3332; 0.8101)	1.0000
i_cond_buenpag_bol5 Sí **	$\beta_{16(1)}$	-0.9519	0.3859	0.0818	(0.2512; 0.5720)	0.9999
sexo 1 (Masculino)	$\beta_{18(1)}$	-0.0020	0.9979	0.0791	(0.7941; 1.1044)	1.0007

Cuadro C.9: Parámetros estimados por modelo de regresión logística con mixtura de escala Gaussiana con regularización bayesiana usando una priori de Horseshoe +

En los cuadros siguientes se muestra el detalle de la estimación aplicando un modelo MEG con priori de Lasso con todas las variables, se aprecian las estimaciones de los parámetros de regresión, los errores estándar, los cuantiles de los parámetros que definen los intervalos de credibilidad y el Rhat de los coeficientes.

Covariable	Parámetro	$\hat{\beta}$	$\exp(\hat{\beta})$	exp(Error Estándar)	exp(IC(95 %))	Rhat
Intercepto	$\beta_0$	2.8070	16.5606	156.7096	(0.7813; 615.0833)	0.9999
cicmat2019-1	$\beta_{11}$	-0.1058	0.8995	0.1337	(0.6247; 1.1492)	1.0000
cicmat2021-2	$\beta_{12}$	0.0103	1.0103	0.2130	(0.6735; 1.5087)	0.9999
cicmat2022-1	$\beta_{13}$	0.0483	1.0494	0.1698	(0.7979; 1.4638)	0.9999
especialidad MATEMÁTICAS	$\beta_{21}$	0.4786	1.6138	3.3721	(0.1231; 13.3417)	0.9999
especialidad QUÍMICA	$\beta_{22}$	-0.0282	0.9721	0.7430	(0.2379; 3.1504)	0.9999
especialidad ING. CIVIL *	$\beta_{23}$	-0.3866	0.6793	0.1569	(0.4074; 1.0227)	1.0000
especialidad ING. INDUSTRIAL *	$\beta_{24}$	-0.4251	0.6536	0.1606	(0.3800; 1.0097)	0.9999
especialidad ING. MECÁNICA	$\beta_{25}$	0.0242	1.0245	0.3266	(0.5519; 1.8322)	0.9999
especialidad ING. MECATRÓNICA	$\beta_{26}$	-0.0688	0.9334	0.2094	(0.5665; 1.3874)	1.0000
especialidad ING. DE MINAS	$\beta_{27}$	-0.2469	0.7812	0.3251	(0.3087; 1.5834)	1.0007
especialidad ING. ELECTRÓNICA	$\beta_{28}$	-0.1513	0.8595	0.2929	(0.4066; 1.5549)	1.0003
especialidad ING. INFORMÁTICA	$\beta_{29}$	0.0570	1.0586	0.2476	(0.6760; 1.6467)	1.0000
especialidad ING. BIOMÉDICA	$\beta_{2(10)}$	0.1867	1.2053	0.5399	(0.6014; 2.7181)	0.9999
especialidad ING. DE LAS TELECOMUNIC. *	$\beta_{2(11)}$	-0.6198	0.5380	0.2274	(0.2122; 1.1039)	1.0001
especialidad ING. GEOLÓGICA	$\beta_{2(12)}$	0.0001	1.0032	0.4288	(0.4309; 2.1120)	0.9999

Cuadro C.10: Parámetros estimados por modelo de regresión logística con mixtura de escala Gaussiana con regularización bayesiana usando una priori de Lasso

Covariable	Parámetro	$\hat{\beta}$	$exp(\hat{\beta})$	exp(Error Estándar)	exp(IC(95 %))	Rhat
especialidad ING. AMBIENTAL Y SOSTENIBLE	$\beta_{2(13)}$	0.0014	1.2786	0.6587	(0.5386; 3.1209)	1.0001
modalidad de ingreso BECA COAR	$\beta_{31}$	-0.0033	0.2209	11.2996	(0.000; 44.2946)	1.0000
modalidad ingreso ING. PRE. UNIV1	$\beta_{32}$	-0.0026	0.8985	0.1846	(0.5882; 1.312)	1.0000
modalidad ingreso ING. PRE. UNIV1 TERCIO	$\beta_{33}$	0.0001	1.1387	0.5715	(0.4806; 2.721)	1.0007
modalidad ingreso ING. PRE. UNIV2 *	$\beta_{34}$	-0.0223	0.2416	0.3705	(0.0134; 1.4657)	0.9999
modalidad ingreso DIPLOMA BACH. UNIV1 **	$\beta_{35}$	0.7727	2.7189	1.1731	(1.2298; 5.8285)	1.0001
modalidad ingreso DIPLOMA BACH. UNIV2 *	$\beta_{36}$	-0.0262	0.1006	0.3925	(0.00003; 1.5387)	0.9999
modalidad ingreso EV. TALENTO Y EX.	$\beta_{37}$	-0.0012	0.9741	0.1952	(0.6583; 1.4237)	1.0000
modalidad ingreso ING. BECA 18	$\beta_{38}$	0.7248	2.0643	1.8248	(0.5693; 7.7228)	1.0001
modalidad ingreso ING. CONVENIO UNIV2.	$\beta_{39}$	0.2530	1.2879	0.8001	(0.4975; 3.634)	1.0001

Cuadro C.11: Parámetros estimados por modelo de regresión logística con mixtura de escala Gaussiana con regularización bayesiana usando una priori de Lasso

Covariable	Parámetro	$\hat{\beta}$	$exp(\hat{\beta})$	exp(Error Estándar)	exp(IC(95 %))	Rhat
modalidad ingreso ING. ESPECIAL.	$\beta_{3(10)}$	0.4116	1.5093	7.5598	(0.0691; 29.7037)	1.0001
modalidad ingreso ING. FE Y ALEGRÍA.	$\beta_{3(11)}$	0.6918	1.9973	4.3699	(0.1313; 17.2616)	1.0001
modalidad ingreso ING. TERCIO SUP. UNIV1 *	$\beta_{3(12)}$	0.3239	1.3825	0.2937	(0.9487; 2.100)	1.0002
modalidad ingreso ING. TERCIO SUP. UNIV2	$\beta_{3(13)}$	0.1211	1.1288	0.7484	(0.3882; 3.3223)	1.0003
modalidad ingreso PRIMERA OPORTUNIDAD	$\beta_{3(14)}$	0.0850	1.0887	0.2855	(0.6933; 1.8127)	1.0001
modalidad ingreso TRANSFERENCIAS UNIV2.	$\beta_{3(15)}$	0.3647	1.4402	0.8310	(0.5915; 3.8492)	0.9999
cond colenolima-callao Sí	$\beta_{41}$	0.0579	1.0596	0.1295	(0.8624; 1.370)	1.0000
I.BENEFICIO Sí	$\beta_{51}$	0.0772	1.0803	0.3193	(0.670; 1.9219)	1.0000
escala de pago G2	$\beta_{61}$	-0.0315	0.9689	0.1945	(0.6661; 1.4289)	1.0010
escala de pago G3	$\beta_{62}$	0.1966	1.2173	0.2621	(0.8843; 1.9117)	1.0000
escala de pago G4 *	$\beta_{63}$	0.1947	1.2150	0.2594	(0.8967; 1.9139)	0.9999
escala de pago G5	$\beta_{64}$	0.1555	1.1682	0.2577	(0.8513; 1.8618)	0.9999
escala de pago G6	$\beta_{65}$	0.0608	1.0627	0.3514	(0.6130; 1.9906)	0.9999
escala de pago G7	$\beta_{66}$	0.2175	1.2429	0.4844	(0.6839; 2.5829)	0.9999
escala de pago G8	$\beta_{67}$	0.1523	1.1645	0.5005	(0.593; 2.5551)	1.0003
escala de pago G9	$\beta_{68}$	0.3132	1.3679	0.6253	(0.6861; 3.1374)	1.0000

Cuadro C.12: Parámetros estimados por modelo de regresión logística con mixtura de escala Gaussiana con regularización bayesiana usando una priori de Lasso

Covariable	Parámetro	$\hat{\beta}$	$exp(\hat{\beta})$	exp(Error Estándar)	exp(IC(95 %))	Rhat
CRAEST **	$\beta_8$	-0.0658	0.9362	0.0149	(0.9071; 0.9655)	1.0001
PRC_APDO_UE **	$\beta_9$	-1.5650	0.2090	0.1036	(0.0861 0.4925)	1.0000
PRC_RET **	$\beta_{10}$	3.8793	48.3928	16.7083	(26.1316; 91.6283)	1.0003
total_cremat	$\beta_{11}$	-0.0237	0.9765	0.0736	(0.8322; 1.121)	0.9999
cond_buenpagador Sí	$\beta_{12(1)}$	-0.0326	0.9679	0.2316	(0.5699; 1.4779)	1.0001
i_cond_buenpag_bol2 Sí	$\beta_{13(1)}$	0.0630	1.0650	0.2243	(0.7605; 1.6399)	1.0000
i_cond_buenpag_bol3 Sí	$\beta_{14(1)}$	-0.0043	0.9957	0.1867	(0.6998; 1.4318)	0.9999
i_cond_buenpag_bol4 Sí **	$\beta_{15(1)}$	-0.6405	0.5270	0.1341	(0.3273; 0.8532)	0.9999
i_cond_buenpag_bol5 Sí **	$\beta_{16(1)}$	-0.8934	0.4092	0.1035	(0.2547; 0.6605)	1.0000
sexo 1 (Masculino)	$\beta_{18(1)}$	-0.0436	0.9572	0.1162	(0.7397; 1.1953)	1.0000

Cuadro C.13: Parámetros estimados por modelo de regresión logística con mixtura de escala Gaussiana con regularización bayesiana usando una priori de Lasso

### C.3. Resultados de regresión con las variables significativas

A continuación presentamos los resultados de la ejecución del modelo logístico 2 escogido, con las covariables significativas, se muestran los estimadores de los coeficientes, el error estándar, los ratios de odds de los coeficientes y los intervalos de confianza de los ratios de odds.

Covariable	Parámetro	$\hat{\beta}$	Error Estándar	$\hat{OR} = \exp(\hat{\beta})$ razón de odds	$\exp(IC(95\%))$	p-value
Intercepto	$\beta_0$	2.515	0.780	12.37	(2.648; 56.622)	0.0012
especialidad MATEMÁTICAS	$\beta_{11}$	0.317	1.309	1.374	(0.054; 12.46)	0.8081
especialidad QUÍMICA	$\beta_{12}$	-1.020	0.808	0.360	(0.063; 1.600)	0.2068
especialidad INGENIERÍA CIVIL	$\beta_{13}$	-1.372	0.395	0.253	(0.119; 0.568)	0.0005
especialidad INGENIERÍA INDUSTRIAL	$\beta_{14}$	-1.414	0.397	0.243	(0.114; 0.546)	0.0003
especialidad INGENIERÍA MECÁNICA	$\beta_{15}$	-0.885	0.472	0.412	(0.164; 1.054)	0.0607
especialidad INGENIERÍA MECATRÓNICA	$\beta_{16}$	-1.024	0.406	0.358	(0.165; 0.819)	0.0116
especialidad INGENIERÍA DE MINAS	$\beta_{17}$	-1.281	0.573	0.277	(0.087; 0.843)	0.0254
especialidad INGENIERÍA ELECTRÓNICA	$\beta_{18}$	-1.194	0.509	0.302	(0.110; 0.825)	0.0191
especialidad INGENIERÍA INFORMÁTICA	$\beta_{19}$	-0.830	0.400	0.436	(0.203; 0.986)	0.0382

Cuadro C.14: Parámetros estimados por modelo de regresión lineal con mixtura de escala Gaussiana: regresión logística clásica con variables significativas

Covariable	Parámetro	$\hat{\beta}$	Error Estándar	$\hat{OR} = \exp(\hat{\beta})$ razón de odds	$\exp(IC(95\%))$	p-value
especialidad INGENIERÍA BIOMÉDICA	$\beta_{1(10)}$	-0.449	0.684	0.637	(0.150; 2.31)	0.511
especialidad INGENIERÍA DE LAS TELECOMUNICACIONES	$\beta_{1(11)}$	-1.639	0.559	0.194	(0.062; 0.57)	0.0033
especialidad INGENIERÍA GEOLÓGICA	$\beta_{1(12)}$	-0.974	0.555	0.3773	(0.124; 1.114)	0.0795
especialidad INGENIERÍA AMBIENTAL Y SOSTENIBLE	$\beta_{1(13)}$	-0.498	0.577	0.607	(0.188; 1.849)	0.3875
modalidad de ingreso BECA COAR	$\beta_{21}$	-9.241	719.095	0.00009	(- ínf; 4.3015e+31)	0.9897
modalidad de ingreso ING. PRE. UNIV1	$\beta_{22}$	0.253	0.339	1.288	(0.672; 2.551)	0.4542
modalidad de ingreso ING. PRE. UNIV1 TERCIO	$\beta_{23}$	0.428	0.556	1.535	(0.484; 4.387)	0.441

Cuadro C.15: Parámetros estimados por modelo de regresión lineal con mixtura de escala Gaussiana: regresión logística clásica con variables significativas

Covariable	Parámetro	$\hat{\beta}$	Error Estándar	$\hat{OR} = \exp(\hat{\beta})$ razón de odds	$\exp(IC(95\%))$	p-value
modalidad de ingreso ING. PRE. UNIV2	$\beta_{24}$	-1.876	1.471	0.153	(0.004; 1.979)	0.2022
modalidad de ingreso DIPLOMA BACH. UNIV1	$\beta_{25}$	1.492	0.455	4.448	(0.672; 2.551)	0.0010
modalidad de ingreso DIPLOMAS BACH. UNIV2	$\beta_{26}$	-12.878	326.346	0.000002	(5.3469e-36; 73.197)	0.9685
modalidad de ingreso EV. TALENTO Y EX.	$\beta_{27}$	0.369	0.319	1.446	(0.787; 2.765)	0.2477
modalidad de ingreso ING. BECA 18	$\beta_{28}$	1.848	0.716	6.350	(1.360; 23.842)	0.0098
modalidad de ingreso ING. CONVENIO UNIV2.	$\beta_{29}$	0.478	0.771	1.613	(0.365; 7.90)	0.5351
modalidad de ingreso ING. ESPECIAL.	$\beta_{2(10)}$	1.904	2.569	6.717	(0.076; 230.092)	0.4584
modalidad de ingreso ING. FE Y ALEGRÍA	$\beta_{2(11)}$	2.011	1.177	7.473	(0.357; 56.335)	0.0876
modalidad de ingreso ING. TERCIO SUP. UNIV1	$\beta_{2(12)}$	0.709	0.316	2.032	(1.115; 3.865)	0.0248

Cuadro C.16: Parámetros estimados por modelo de regresión lineal con mixtura de escala Gaussiana: regresión logística clásica con variables significativas

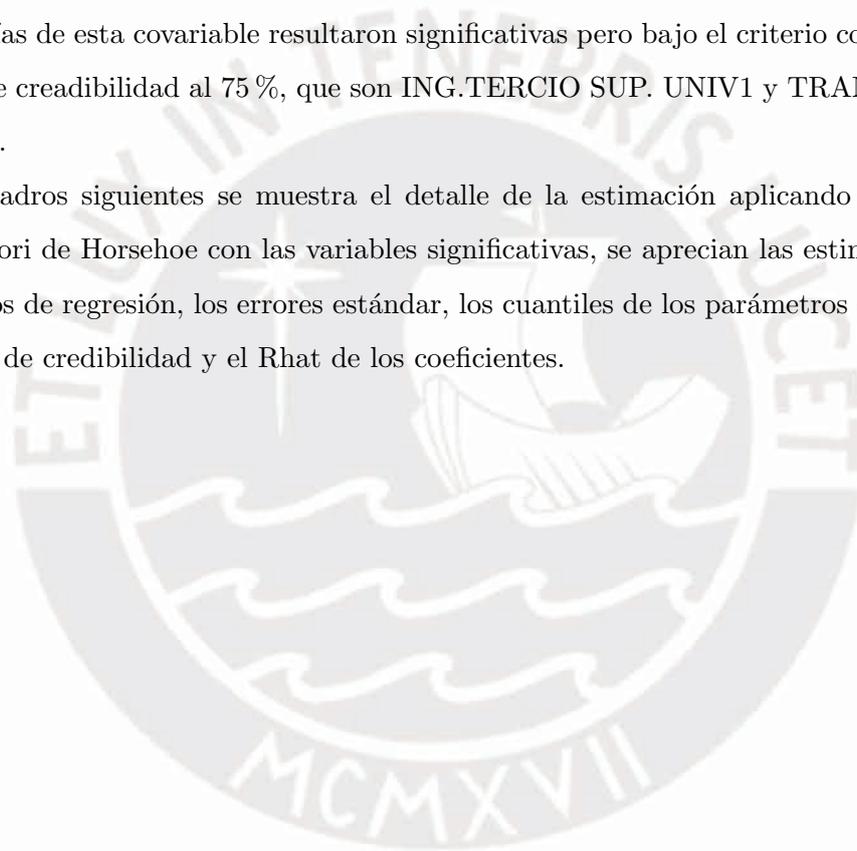
Covariable	Parámetro	$\hat{\beta}$	Error Estándar	$\hat{OR} = \exp(\hat{\beta})$ razón de odds	$\exp(\text{IC}(95\%))$	p-value
modalidad de ingreso ING. TERCIO SUP. UNIV2	$\beta_{2(13)}$	0.373	0.808	1.453	(0.300; 7.509)	0.6437
modalidad de ingreso PRIMERA OPORTUNIDAD	$\beta_{2(14)}$	0.505	0.374	1.657	(0.800; 3.491)	0.1768
modalidad de ingreso TRANSFERENCIAS UNIV2	$\beta_{2(15)}$	0.589	0.741	1.802	(0.441; 8.437)	0.42702
escala de pago G2	$\beta_{31}$	0.402	0.296	1.495	(0.842; 2.702)	0.1744
escala de pago G3	$\beta_{32}$	0.725	0.283	2.066	(1.197; 3.654)	0.0105
escala de pago G4	$\beta_{33}$	0.707	0.282	2.028	(1.178; 3.584)	0.0124
escala de pago G5	$\beta_{34}$	0.646	0.294	1.907	(1.081; 3.441)	0.0282
escala de pago G6	$\beta_{35}$	0.670	0.359	1.955	(0.960; 3.946)	0.0618
escala de pago G7	$\beta_{36}$	0.957	0.388	2.605	(1.200; 5.548)	0.0137
escala de pago G8	$\beta_{37}$	0.933	0.436	2.544	(1.054; 5.887)	0.0322
escala de pago G9	$\beta_{38}$	0.960	0.434	2.612	(1.085; 6.029)	0.0272
CRAEST	$\beta_4$	-0.064	0.015	0.937	(0.909; 0.966)	3.08e-05
PRC_APDO_UE	$\beta_5$	-1.929	0.441	0.145	(0.060; 0.342)	1.23e-05
PRC_RET	$\beta_6$	4.172	0.311	64.907	(35.872; 121.923)	< 2e-16
i_cond_buen pag_bol4 Sí	$\beta_{71}$	-0.716	0.180	0.488	(0.343; 0.697)	7.44e-05
i_cond_buen pag_bol5 Sí	$\beta_{81}$	-1.028	0.167	0.357	(0.257; 0.497)	8.27e-10

Cuadro C.17: Parámetros estimados por modelo de regresión lineal con mixtura de escala Gaussiana: regresión logística clásica con variables significativas

Este modelo muestra un pseudo  $R^2$  mediante la medida de Hosmer y Lemeshow  $R_L^2$  de 0.33.

En cuanto al modelo de regresión logística con MEG con regularización bayesiana usando una priori de Horseshoe, se corrió solo para las variables que resultaron significativas, estas son: modalidad de ingreso en su categoría DIPLOMA BACH. UNIV1, bajo el criterio considerando el intervalo de credibilidad al 75%; y CRAEST, PRC\_APDO\_UE, PRC\_RET, i.cond\_buenpag\_bol4 en su categoría “Sí” e i.cond\_buenpag\_bol5 en su categoría “Sí”, bajo el criterio, considerando el intervalo de credibilidad al 95%. El modelo obtenido tuvo un  $WAIC = 813.8$  y un pseudo  $R^2 = 0.3145$ . En este la covariable modalidad de ingreso en su categoría DIPLOMA BACH. UNIV1 resultó significativa bajo el criterio considerando el intervalo de credibilidad al 95% a diferencia del modelo con todas las covariables, además otras categorías de esta covariable resultaron significativas pero bajo el criterio considerando el intervalo de credibilidad al 75%, que son ING.TERCIO SUP. UNIV1 y TRANSFERENCIAS UNIV2.

En los cuadros siguientes se muestra el detalle de la estimación aplicando un modelo MEG con priori de Horseshoe con las variables significativas, se aprecian las estimaciones de los parámetros de regresión, los errores estándar, los cuantiles de los parámetros que definen los intervalos de credibilidad y el Rhat de los coeficientes.



Covariable	Parámetro	$\hat{\beta}$	$exp(\hat{\beta})$	exp(Error Estándar)	exp(IC(95 %))	Rhat
Intercepto	$\beta_0$	2.1708	8.765	38.9468	(0.7058; 153.3773)	0.9999
modalidad de ingreso BECA COAR	$\beta_{11}$	-0.3987	0.671	3.0441	(0.000; 11.933)	1.0001
modalidad ingreso ING. PRE. UNIV1	$\beta_{12}$	-0.0434	0.957	0.1156	(0.6542; 1.1078)	0.9999
modalidad ingreso ING. PRE. UNIV1 TERCIO	$\beta_{13}$	0.014	1.014	0.2661	(0.6393; 1.6826)	1.0000
modalidad ingreso ING. PRE. UNIV2	$\beta_{14}$	-0.485	0.615	0.3586	(0.0581; 1.4641)	0.9999
modalidad ingreso DIPLOMA BACH. UNIV1 **	$\beta_{15}$	1.036	2.818	1.0533	(0.9750; 5.1042)	1.0000
modalidad ingreso DIPLOMA BACH. UNIV2	$\beta_{16}$	-1.054	0.348	0.4078	(0.000; 1.5985)	1.0002
modalidad ingreso EV. TALENTO Y EX.	$\beta_{17}$	-0.026	0.973	0.1015	(0.7120; 1.1102)	0.9999
modalidad ingreso ING. BECA 18	$\beta_{38}$	0.3927	1.480	0.8795	(0.6853; 4.1331)	0.9999
modalidad ingreso ING. CONVENIO UNIV2.	$\beta_{39}$	0.403	1.497	0.5061	(0.7502; 2.7343)	0.9999

Cuadro C.18: Parámetros estimados por modelo de regresión logística con mixtura de escala Gaussiana con regularización bayesiana usando una priori de Horseshoe con variables significativas

Covariable	Parámetro	$\hat{\beta}$	$exp(\hat{\beta})$	exp(Error Estándar)	exp(IC(95 %))	Rhat
modalidad ingreso ING. ESPECIAL.	$\beta_{3(10)}$	0.084	1.088	1.8075	(0.2112; 7.2969)	0.9999
modalidad ingreso ING. FE Y ALEGRÍA.	$\beta_{3(11)}$	0.164	1.179	1.3681	(0.3372; 5.7004)	1.0004
modalidad ingreso ING.TERCIO SUP. UNIV1 * (IC al 75 %)	$\beta_{3(12)}$	0.268	1.307	0.1951	(0.9580; 1.723)	1.0001
modalidad ingreso ING.TERCIO SUP. UNIV2	$\beta_{3(13)}$	0.269	1.309	0.4356	(0.6858; 2.3936)	1.0002
modalidad ingreso PRIMERA OPORTUNIDAD	$\beta_{3(14)}$	0.009	1.009	0.1317	(0.7911; 1.3074)	1.0003
modalidad ingreso TRANSFERENCIAS UNIV2.	$\beta_{3(15)}$	0.641	1.899	0.5961	(0.8117; 3.1485)	1.0001
CRAEST **	$\beta_4$	-0.062	0.939	0.0173	(0.9034; 0.9713)	1.0000
PRC_APDO_UE **	$\beta_5$	-1.697	0.183	0.1183	(0.0706; 0.5346)	1.0000
PRC_RET **	$\beta_6$	3.990	54.092	17.6104	(28.8668; 97.8997)	1.0001
i.cond.buenpag_bol4 Sí **	$\beta_{7(1)}$	-0.669	0.512	0.1313	(0.3301; 0.8450)	1.0000
i.cond.buenpag_bol5 Sí **	$\beta_{8(1)}$	-0.944	0.388	0.0869	(0.2423; 0.5832)	1.0002

Cuadro C.19: Parámetros estimados por modelo de regresión logística con mixtura de escala Gaussiana con regularización bayesiana usando una priori de Horseshoe con variables significativas

## C.4. Distribuciones a posteriori

En las siguientes figuras se presentan los gráficos de la distribución a posteriori condicional y cadenas de la estimación de  $\beta$  usando una priori de Horseshoe en la aplicación a la deserción universitaria.

Los gráficos de cadenas para los parámetros  $\beta$  muestran evidencia de convergencia.

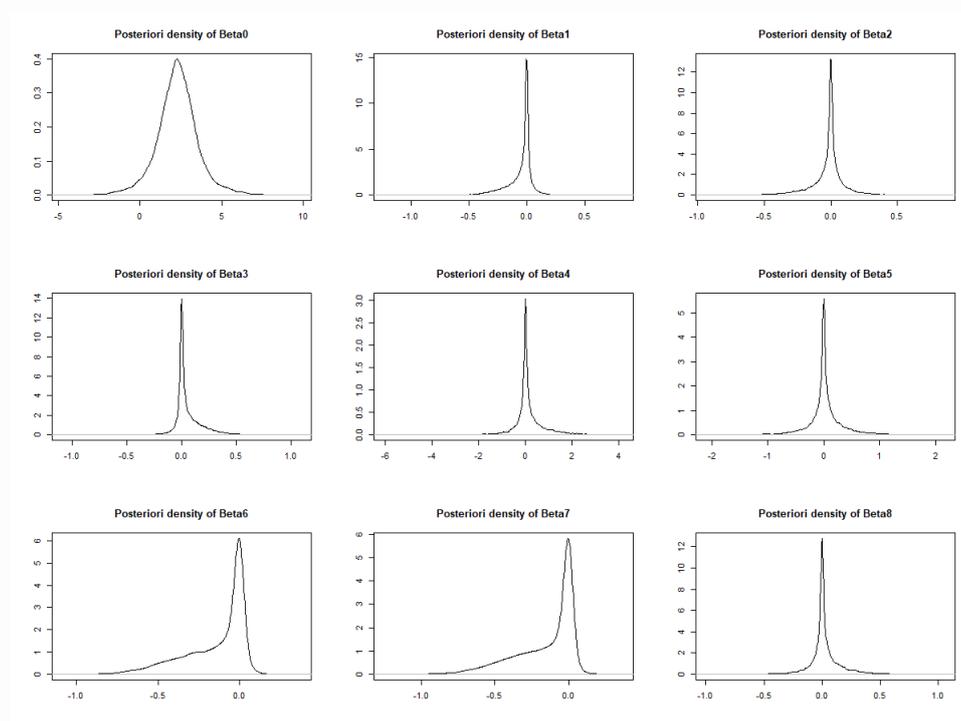


Figura C.14: Distribución a posteriori condicional de  $\beta$ ,  $\beta_0$  al  $\beta_8$ , en la aplicación con datos reales

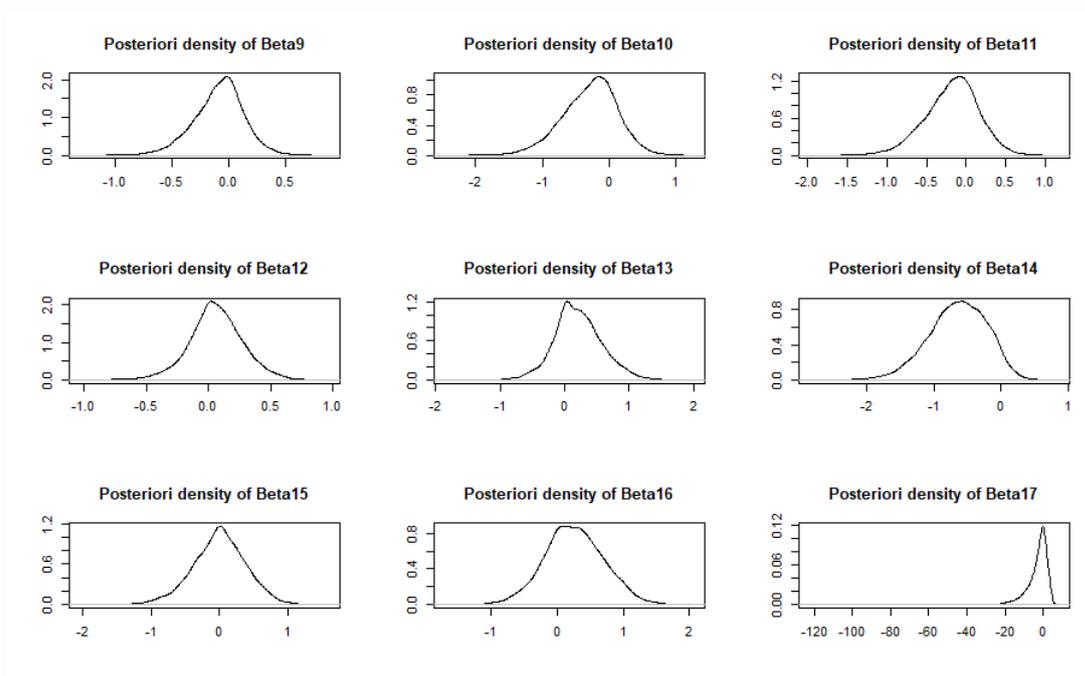


Figura C.15: Distribución a posteriori condicional de  $\beta$ ,  $\beta_9$  al  $\beta_{17}$

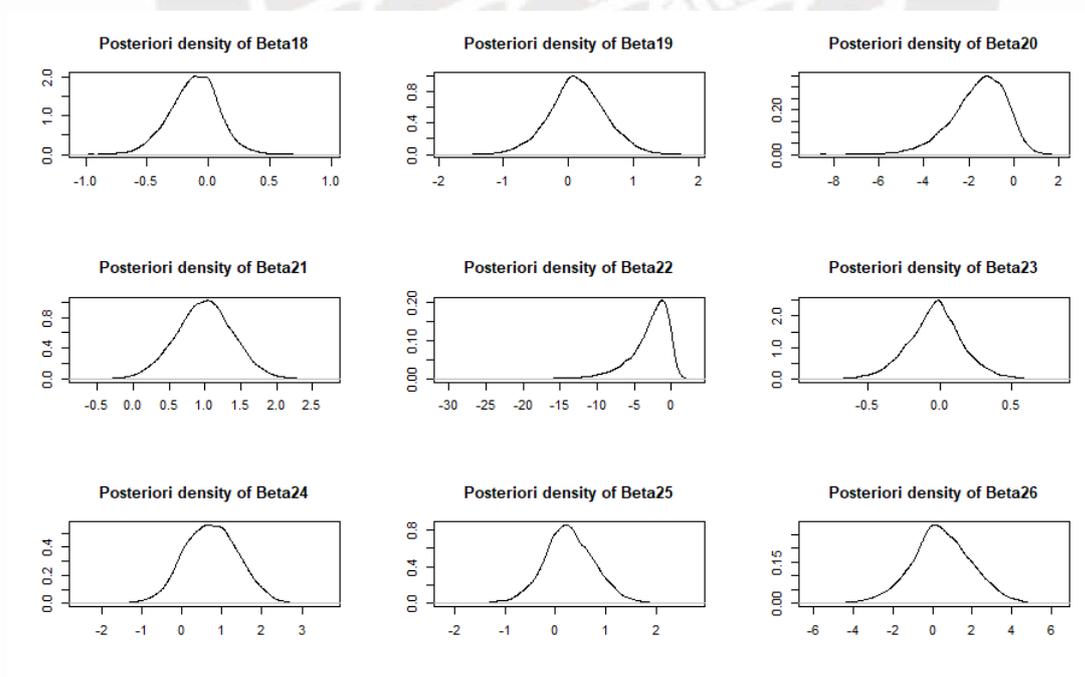


Figura C.16: Distribución a posteriori condicional de  $\beta$ ,  $\beta_{18}$  al  $\beta_{26}$

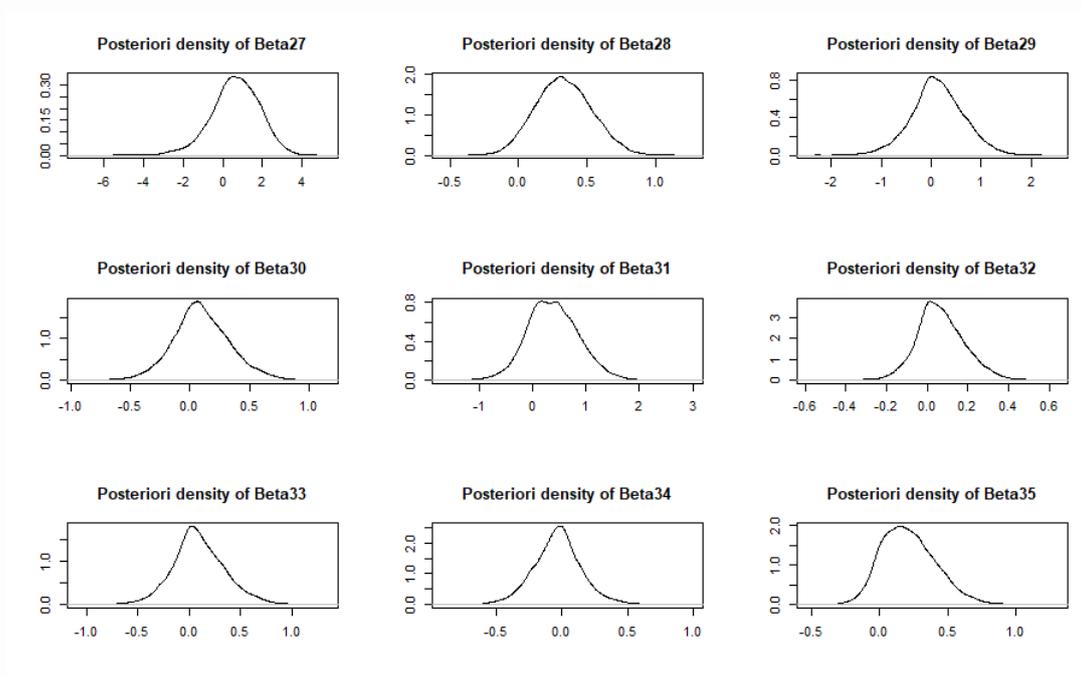


Figura C.17: Distribución a posteriori condicional de  $\beta$ ,  $\beta_{27}$  al  $\beta_{35}$

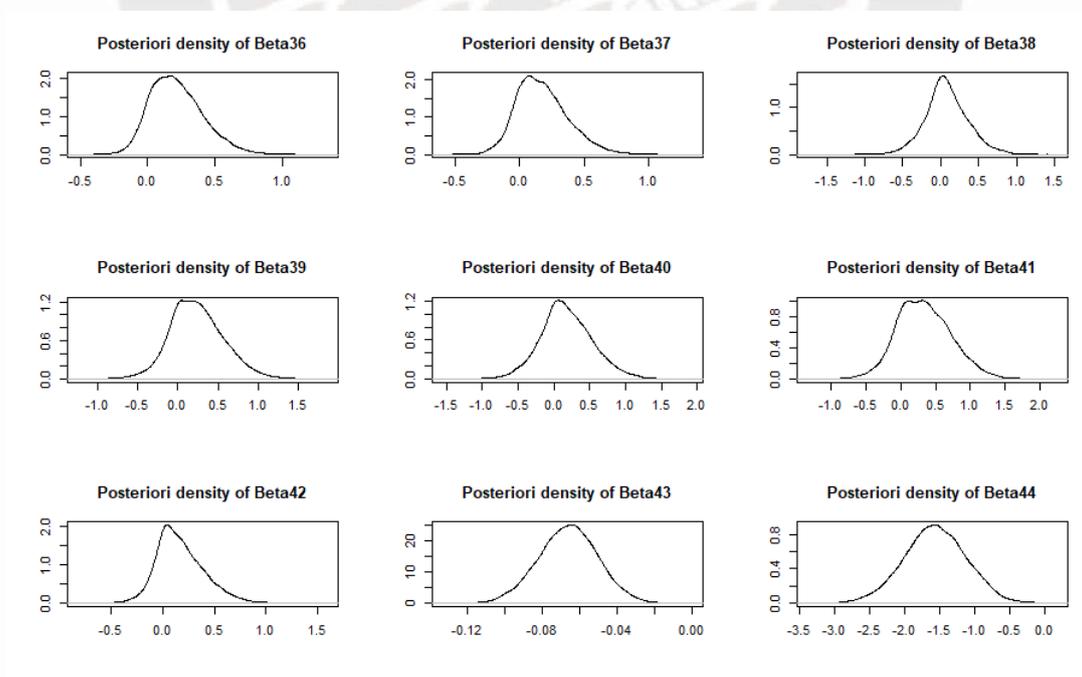


Figura C.18: Distribución a posteriori condicional de  $\beta$ ,  $\beta_{36}$  al  $\beta_{44}$

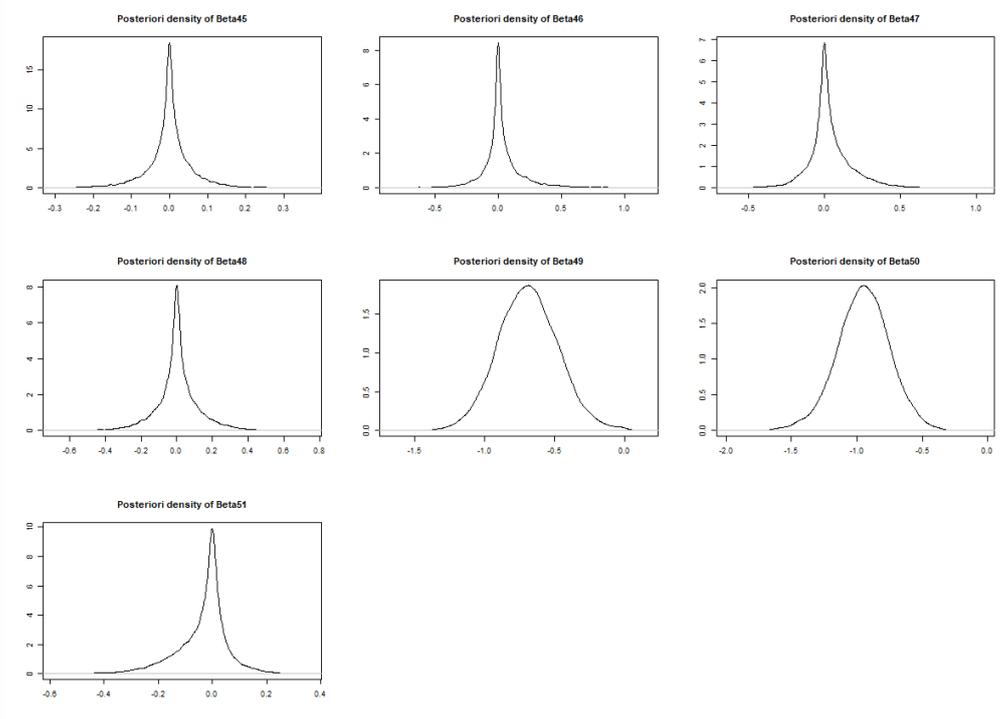


Figura C.19: Distribución a posteriori condicional de  $\beta$ ,  $\beta_{45}$  al  $\beta_{51}$

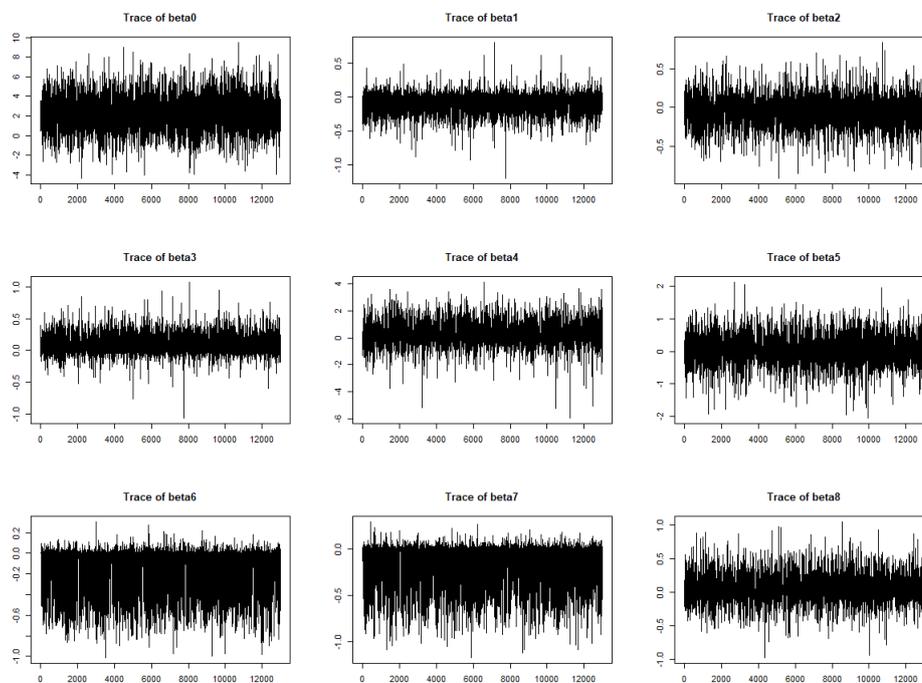


Figura C.20: Gráfico de cadenas de  $\beta$ ,  $\beta_0$  al  $\beta_8$ , aplicando el estimador de Horseshoe

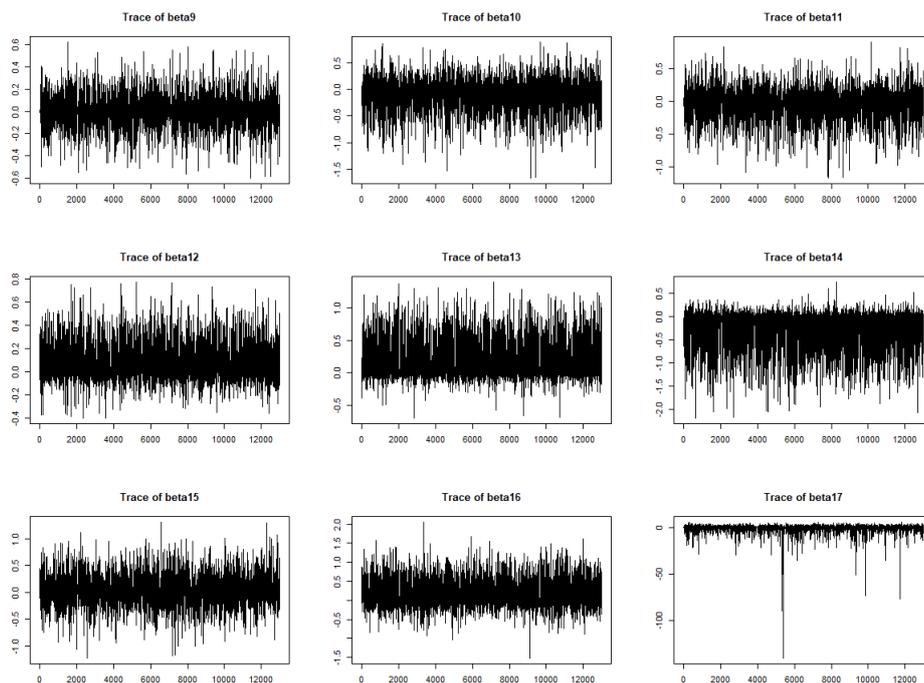


Figura C.21: Gráfico de cadenas de  $\beta$ ,  $\beta_9$  al  $\beta_{17}$ , aplicando el estimador de Horseshoe

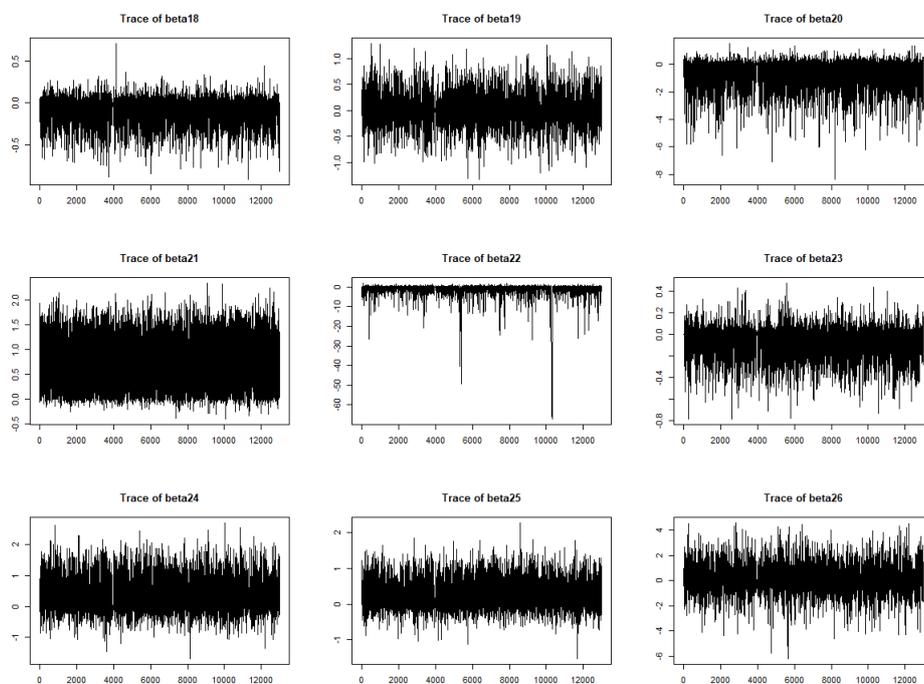


Figura C.22: Gráfico de cadenas de  $\beta$ ,  $\beta_{18}$  al  $\beta_{26}$ , aplicando el estimador de Horseshoe

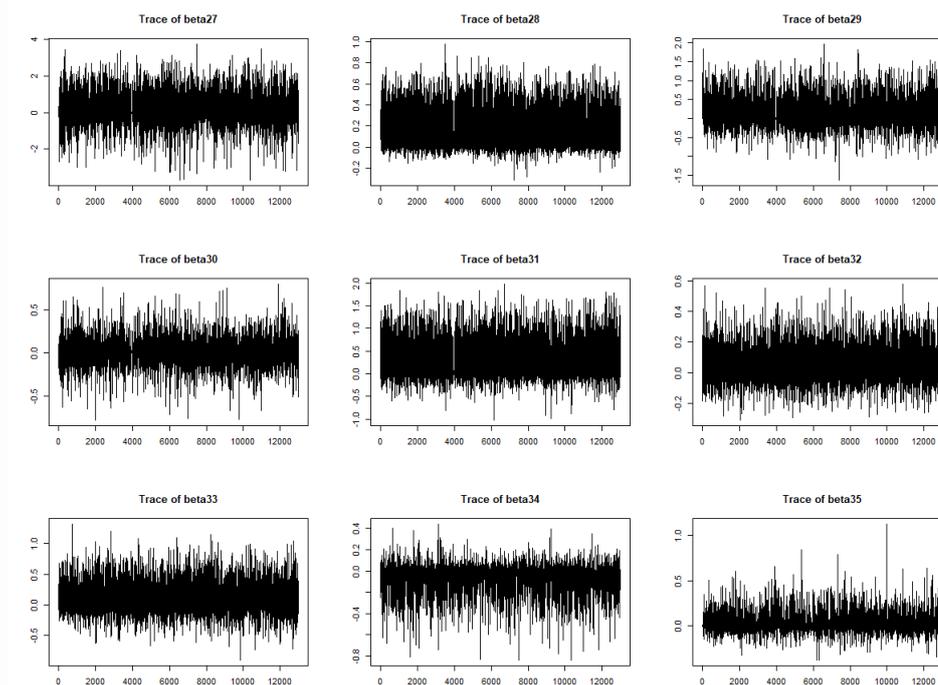


Figura C.23: Gráfico de cadenas de  $\beta$ ,  $\beta_{27}$  al  $\beta_{35}$ , aplicando el estimador de Horseshoe

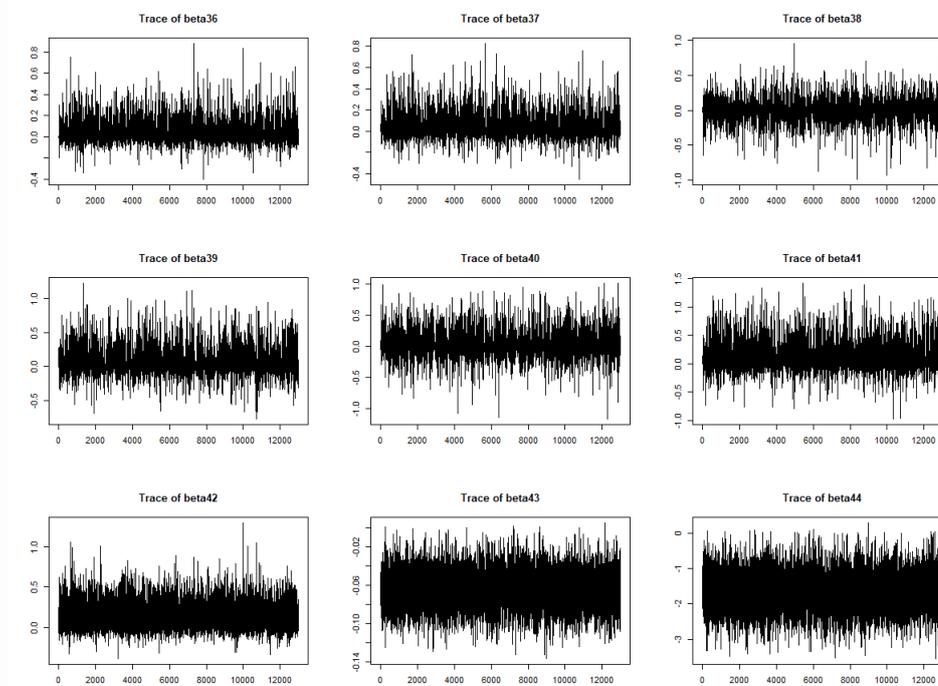


Figura C.24: Gráfico de cadenas de  $\beta$ ,  $\beta_{36}$  al  $\beta_{44}$ , aplicando el estimador de Horseshoe

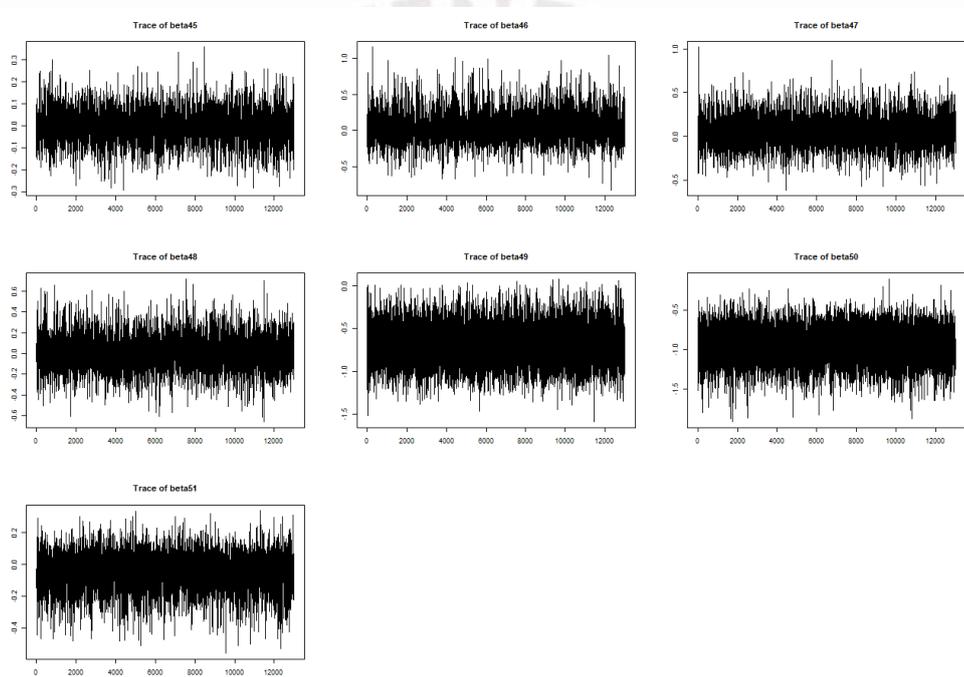


Figura C.25: Gráfico de cadenas de  $\beta$ ,  $\beta_{45}$  al  $\beta_{51}$ , aplicando el estimador de Horseshoe

# Bibliografía

- Allasia, M. B., Branco, M. D. y Quaglino, M. B. (2016). Regresión lasso bayesiana. ajuste de modelos lineales penalizados mediante la asignación de priores normales con mezcla de escala., *Instituto de Investigaciones Teóricas y Aplicadas en Estadística* .
- Andrews, D. F. y Mallows, C. L. (1974). Scale mixtures of normal distributions, *Journal of the Royal dataistical Society. Series B (Methodological)* **36(1)**: 99–102.  
**URL:** <http://www.jstor.org/stable/2984774>
- Anirban Bhattacharya, A. C. y Mallick, B. K. (2016). Fast sampling with gaussian scale mixture priors in high-dimensional regression, *Biometrika* **103**: 985â991.
- Bhadra, A., Datta, J., Li, Y., Polson, N. G. y Willard, B. (2019). Prediction risk for the horseshoe regression, *Journal of Machine Learning Research* **20(2)**: 1 – 39.  
**URL:** <https://jmlr.csail.mit.edu/papers/volume20/18-321/18-321.pdf>
- Bhadra, A., Datta, J., Polson, N. G. y Willard, B. (2016). The horseshoe+ estimator of ultra-sparse signals, *Bayesian Analysis* .
- Bhadra, A., Datta, J., Polson, N. G. y Willard, B. (2017). The horseshoe+ estimator of ultra-sparse signals, *Bayesian Analysis* **12**: 1105 â 1131.
- Bhadra, A., Datta, J., Polson, N. G. y Willard, B. T. (2015). The horseshoe+ estimator of ultra-sparse signals, *arXiv: dataistics Theory* .  
**URL:** <https://api.semanticscholar.org/CorpusID:88513176>
- Carrasco, M. (2016). *Técnicas de Regularización en regresión: Implementación y aplicaciones*. [https://idus.us.es/bitstream/handle/11441/43746/Carrasco%20Carrasco, %20Mar%C3%ADa%20TFG.pdf](https://idus.us.es/bitstream/handle/11441/43746/Carrasco%20Carrasco,%20Mar%C3%ADa%20TFG.pdf).
- Carvalho, C., Polson, N. y Scott, J. (2010). The horseshoe estimator for sparse signals, *Biometrika* **97(2)**: 465 – 480.  
**URL:** <https://ani.stat.fsu.edu/debdeep/HS-BKA.pdf>

- Casella, G. y George, E. (1992). Explaining the gibbs sampler, *The American Statistician* **46**(3): 167 – 174.
- Derksen, S. y Keselman, H. J. (1992). Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables, *British Journal of Mathematical and dataistical Psychology* **45**: 265 â 282.
- Efron, B., Hastie, T., Johnstone, I. y Tibshirani, R. (2004). Least angle regression, *Ann ghostist* **32**(2): 407 – 499.
- Feng, X.-N., Wang, Y., Lu, B. y Song, X.-Y. (2017). Bayesian regularized quantile structural equation models, *Journal of Multivariate Analysis* **154**: 234 – 248.  
**URL:** <https://www.sciencedirect.com/science/article/pii/S0047259X16301324>
- Gelman, A. (2006). Prior distributions for variance parameters en modelos jerárquicos, *Bayesian Analysis* **1**(3): 515–533.
- Ghosh, I. (2014). The kumaraswamy-half-cauchy distribution: properties and applications, *Mathematics and Statistics, Austin Peay State University* **13**(2): 122 – 134.
- Griffin, J. y Brown, P. (2017). Hierarchical shrinkage priors for regression models, *Bayesian Analysis* **12**: 135 – 159.
- Hans, C. (2009). Bayesian lasso regression, *Biometrika* **96**(4): 835 – 845.  
**URL:** <https://econpapers.repec.org/RePEc:oup:biomet:v:96:y:2009:i:4:p:835-845>
- Harrison, D. y Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air, *Journal of Environmental Economics and Management* **5**: 81–102.  
**URL:** <https://api.semanticscholar.org/CorpusID:55571328>
- Hoerl, E. y Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems, *Technometrics* **12**(1): 55 – 67.  
**URL:** <https://www.jstor.org/stable/1267351>
- Jesse Windle, N. G. P. y Scott, J. G. (2014). Sampling pólya-gamma random variates: alternate and approximate techniques.  
**URL:** <https://arxiv.org/pdf/1405.0506.pdf>
- Li, Q. y Lin, N. (2010). The bayesian elastic net, *International Society for Bayesian Analysis* **5**(1): 151 â 170.

- Lindley, D. V. y Smith, A. F. M. (1972). Bayes estimates for the linear model, *Journal of the Royal dataistical Society. Series B (Methodological)* **34**(1): 1–41.  
**URL:** <https://www.jstor.org/stable/2985048>
- López, M. (2012). Aplicación del elastic net lasso y modelos relacionados en selección genómica basados en marcadores moleculares, *Institución de enseñanza e investigación en ciencias agrícolas* .
- Lozano, J. H. (2020). *Distribución previa horseshoe en la regresión lineal: estudio de su comportamiento como método de selección de variables en condiciones de alta dimensionalidad*, Universidad Autónoma de Madrid'.  
**URL:** <https://libros.uam.es/tfm/catalog/download/1149/2031/2309?inline=1>
- Makalic, E. y Schmidt, D. (2016a). High-dimensional bayesian regularized regression with the bayesreg package, *arXiv: Computation* .
- Makalic, E. y Schmidt, D. (2016b). A simple sampler for the horseshoe estimator, *IEEE Signal Procesing Letters* **23**(1): 179–182.
- Molina, J. y Ramírez, I. (2014). Análisis de distribuciones a priori de los parámetros de escala del modelo zip, *Comunicaciones en Estadística. Universidad Santo Tomás* **7**(2): 115 – 129.
- Park, T. y Casella, G. (2008). The bayesian lasso, *Journal of the American Statistical Association* **103**(482): 681 – 686.
- Perea, J. (2019). *Técnicas de Regularización en el Aprendizaje Estadístico*.  
[http://e-spacio.uned.es/fez/eserv/bibliuned:masterMatavanz-Jrperea/Perea\\_Luque\\_JuanRafael\\_TFM.pdf](http://e-spacio.uned.es/fez/eserv/bibliuned:masterMatavanz-Jrperea/Perea_Luque_JuanRafael_TFM.pdf).
- Polson, N. G. y Scott, J. G. (2010). Shrink globally, act locally: Sparse bayesian regularization and prediction, *Bayesian dataistics* **9**.
- Polson, N. G. y Scott, J. G. (2012). Local shrinkage rules, levy processes and regularized regression, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **74**(2): 287 – 311.  
**URL:** <https://doi.org/10.48550/arXiv.1010.3390>
- Polson, N. G., Scott, J. G. y Windle, J. (2013). Bayesian inference for logistic models using poly-gamma latent variables, **108**(504): 1339–1349.

- Rue, H. (2001). Fast sampling of gaussian markov random fields, *Journal of the Royal dataistical Society. Series B (dataistical Methodology)* **63**(2): 325–338.  
**URL:** <https://www.jstor.org/stable/2680602>
- Schmidt, D. y Makalic, E. (2019). Bayesian generalized horseshoe estimation of generalized linear models, *ECML PKDD 2019: Machine Learning and Knowledge Discovery in Databases* pp. 598 – 613.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso, *Journal of the Royal dataistical Society. Series B (Methodological)* **58**(1): 267 â 288.  
**URL:** <https://www.jstor.org/stable/2346178>
- van der Pas, S. L., Kleijn, B. J. K. y van der Vaart, A. W. (2014). The horseshoe estimator: Posterior concentration around nearly black vectors, *Electronic Journal of Statistics* **8**(2): 2585–2618.
- van Erp, S., Oberski, D. L. y Mulder, J. (2019). Shrinkage priors for bayesian penalized regression, *Journal of Mathematical Psychology* pp. 31 – 50.
- Zou, H. y Hastie, T. (2005). Regularization and variable selection via the elastic net, *Journal Royal dataistical Society: Series B* **67**(2): 301 â 320.