

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

ESCUELA DE POSGRADO



PONTIFICIA  
**UNIVERSIDAD  
CATÓLICA**  
DEL PERÚ

**MÉTODOS DE SELECCIÓN DE VARIABLES BAJO EL  
ENFOQUE BAYESIANO PARA EL MODELO LINEAL  
NORMAL**

Tesis para optar el grado de Magíster en Estadística

**AUTOR**

Sthip Frank Blas Oyola

**ASESOR**

Dr. Cristian Luis Bayes Rodríguez

**JURADO**

Dr. Luis Hilmar Valdivieso Serrano  
Dr. Victor Giancarlo Sal Y Rosas Celi  
Dr. Cristian Luis Bayes Rodríguez

LIMA - PERÚ

2020

## Dedicatoria

A mis queridos padres, Regina Oyola y Aquilino Blas, cuyo apoyo favoreció a que, el día de hoy, culmine la maestría.

A mi compañera de vida Elizabeth Morales, por sus consejos, amor y comprensión.

A mis hermanos Haydee, Roldan y Lily, por la confianza y respaldo en todo lo necesario para lograr mis objetivos como persona y profesional.





## Agradecimientos

Agradezco a los profesores de la Maestría en Estadística de la Pontificia Universidad Católica del Perú por sus constantes capacitaciones y especialmente a mi asesor el Dr. Cristian Bayes por sus enseñanzas, su guía constante en el proceso de este trabajo.

¡ Arriba Alianza, vamos Huaral, Siempre juntos Chinonkilis 17 !



## Resumen

En muchos casos prácticos, al realizar un análisis de regresión, se cuenta con un gran número de potenciales variables explicativas de las cuáles sólo algunas serán importantes para explicar la variable respuesta. Por lo tanto, un problema importante para la construcción de un modelo de regresión es encontrar un adecuado conjunto de variables explicativas. A los métodos que lidian con este problema se les denomina métodos de selección de variables. En el presente proyecto de tesis, se estudiarán tres métodos de selección de variables bajo inferencia bayesiana para el modelo de regresión lineal normal los cuales fueron propuestos por [George y McCulloch \(1993\)](#), [Kuo y Mallick \(1998\)](#) y [Dellaportas et al. \(2002\)](#). Estos métodos, a diferencia de los métodos tradicionales, consideran la selección de variables dentro del mismo modelo, por ejemplo, introduciendo variables latentes que indiquen la presencia o ausencia de una variable explicativa. Se realizaron comparaciones de estos métodos bayesianos con los métodos Lasso y Stepwise por ser los más tradicionales. A través de un estudio con datos simulados, en diversos escenarios se observa que los métodos bayesianos permiten una adecuada selección de las variables explicativas. Adicionalmente se presentan los resultados de una aplicación con datos reales.

**Palabras-clave:** selección de variable, variable latente, método Lasso, método Stepwise, modelo de regresión lineal normal.

# Índice general

<b>Índice de figuras</b>	<b>VII</b>
<b>Índice de cuadros</b>	<b>X</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Objetivos	2
1.2. Organización del Trabajo	2
<b>2. Método SSVS</b>	<b>4</b>
2.1. Introducción	4
2.2. Método SSVS	4
2.3. Distribución a posteriori	6
2.4. Distribuciones condicionales completas	6
<b>3. Método KM</b>	<b>9</b>
3.1. Introducción	9
3.2. Método KM	9
3.3. Distribución a posteriori	10
3.4. Distribuciones condicionales completas	11
<b>4. Método GVS</b>	<b>14</b>
4.1. Introducción	14
4.2. Método GVS	14
4.3. Distribución a posteriori	16
4.4. Distribuciones condicionales completas	17
<b>5. Estudio de simulación</b>	<b>20</b>
5.1. Escenario 1: Veinte variables explicativas independientes.	20
5.1.1. Implementación SSVS	21
5.1.2. Implementación KM	23
5.1.3. Implementación GVS	25
5.1.4. Resumen de resultados	27
5.2. Escenario 2: Cuatro variables correlacionadas	29
5.2.1. Resumen de resultados	30
5.3. Escenario 3: Dos variables correlacionadas y una variable candidata con multicolinealidad	35

5.3.1. Resumen de resultados . . . . .	36
5.4. Escenario 4: Dos variables correlacionadas y una variable relevante con multicolinealidad . . . . .	41
5.4.1. Resumen de resultados . . . . .	42
<b>6. Aplicación con datos reales</b>	<b>48</b>
6.1. Descripción de los datos . . . . .	48
6.2. Implementación de los métodos bayesianos . . . . .	50
<b>7. Conclusiones y Sugerencias</b>	<b>55</b>
7.1. Conclusiones . . . . .	55
7.2. Sugerencias para investigaciones futuras . . . . .	55
<b>A. Código en R del estudio de simulación</b>	<b>56</b>
A.1. Método SSVS . . . . .	56
A.2. Método KM . . . . .	62
A.3. Método GVS . . . . .	67
A.4. Método Lasso . . . . .	75
A.5. Método Stepwise . . . . .	76
<b>B. Análisis de convergencia</b>	<b>77</b>
B.1. Convergencia de los parámetros en el estudio con datos simulados para veinte variables explicativas independientes . . . . .	77
B.2. Convergencia de los parámetros en el estudio con datos simulados para cuatro variables correlacionadas . . . . .	80
B.3. Convergencia de los parámetros en el estudio con datos simulados para dos variables correlacionadas y una variable candidata con multicolinealidad . . . . .	83
B.4. Convergencia de los parámetros en el estudio con datos simulados para dos variables correlacionadas y una variable relevante con multicolinealidad . . . . .	86
B.5. Convergencia de los parámetros en el estudio de la Esperanza de vida al nacer	89
<b>Bibliografía</b>	<b>92</b>

## Índice de figuras

5.1. Distribución a posteriori de $\beta$ en el estudio con datos simulados para veinte variables explicativas independientes empleando el método SSVS y considerando $(\phi_j, C_j) = (0.1, 500)$ . . . . .	23
5.2. Distribución a posteriori de $\beta$ en el estudio con datos simulados para veinte variables explicativas independientes empleando el método KM y considerando $D_0 = 16I$ . . . . .	25
5.3. Distribución a posteriori de $\beta$ en el estudio con datos simulados para veinte variables explicativas independientes empleando el método GVS y considerando $\Sigma_j = 100$ . . . . .	27
5.4. Distribución a posteriori de $\beta$ en el estudio con datos simulados para cuatro variables correlacionadas empleando el método SVSS y considerando $(\phi_j, C_j) = (0.1, 500)$ . . . . .	33
5.5. Distribución a posteriori de $\beta$ en el estudio con datos simulados para cuatro variables correlacionadas empleando el método KM y considerando $D_0 = 16I$ . . . . .	34
5.6. Distribución a posteriori de $\beta$ en el estudio con datos simulados para cuatro variables correlacionadas empleando el método GVS y considerando $\Sigma_j = 100$ . . . . .	35
5.7. Distribución a posteriori de $\beta$ en el estudio con datos simulados para dos variables correlacionadas y una variable candidata con multicolinealidad empleando el método SVSS y considerando $(\phi_j, C_j) = (0.1, 500)$ . . . . .	39
5.8. Distribución a posteriori de $\beta$ en el estudio con datos simulados para dos variables correlacionadas y una variable candidata con multicolinealidad empleando el método KM y considerando $D_0 = 16I$ . . . . .	40
5.9. Distribución a posteriori de $\beta$ en el estudio con datos simulados para dos variables correlacionadas y una variable candidata con multicolinealidad empleando el método GVS y considerando $\Sigma_j = 100$ . . . . .	41
5.10. Distribución a posteriori de $\beta$ en el estudio con datos simulados para dos variables correlacionadas y una variable relevante con multicolinealidad empleando el método SVSS y considerando $(\phi_j, C_j) = (0.1, 500)$ . . . . .	45
5.11. Distribución a posteriori de $\beta$ en el estudio con datos simulados para dos variables correlacionadas y una variable relevante con multicolinealidad empleando el método KM y considerando $D_0 = 16I$ . . . . .	46
5.12. Distribución a posteriori de $\beta$ en el estudio con datos simulados para dos variables correlacionadas y una variable relevante con multicolinealidad empleando el método GVS y considerando $\Sigma_j = 100$ . . . . .	47

6.1. Distribución a posteriori de $\beta$ en el estudio de la Esperanza de vida al nacer empleando el método SSVS y considerando $(\phi_j, C_j) = (0.1, 500)$ . . . . .	52
6.2. Distribución a posteriori de $\beta$ en el estudio de la Esperanza de vida al nacer empleando el método KM y considerando $D_0 = 16I$ . . . . .	53
6.3. Distribución a posteriori de $\beta$ en el estudio de la Esperanza de vida al nacer empleando el método GVS y considerando $\Sigma_j = 100$ . . . . .	54
B.1. Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros $\beta, \gamma, \sigma$ empleando el método SSVS y considerando $(\phi_j, C_j) = (0.1, 500)$ , en el estudio con datos simulados para veinte variables explicativas independientes. . . . .	77
B.2. Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros $\beta, \gamma, \sigma$ empleando el método KM y considerando $D_0 = 16I$ , en el estudio con datos simulados para veinte variables explicativas independientes. . . . .	78
B.3. Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros $\beta, \gamma, \sigma$ empleando el método GVS y considerando $\Sigma_j = 100$ , en el Estudio con datos simulados para veinte variables explicativas independientes. . . . .	79
B.4. Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros $\beta, \gamma, \sigma$ empleando el método SSVS y considerando $(\phi_j, C_j) = (0.1, 500)$ , en el estudio con datos simulados para cuatro variables correlacionadas. . . . .	80
B.5. Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros $\beta, \gamma, \sigma$ empleando el método KM y considerando $D_0 = 16I$ , en el estudio con datos simulados para cuatro variables correlacionadas. . . . .	81
B.6. Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros $\beta, \gamma, \sigma$ empleando el método GVS y considerando $\Sigma_j = 100$ , en el estudio con datos simulados para cuatro variables correlacionadas. . . . .	82
B.7. Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros $\beta, \gamma, \sigma$ empleando el método SSVS y considerando $(\phi_j, C_j) = (0.1, 500)$ , en el estudio con datos simulados dos variables correlacionadas y una variable candidata con multicolinealidad. . . . .	83
B.8. Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros $\beta, \gamma, \sigma$ empleando el método KM y considerando $D_0 = 16I$ , en el estudio con datos simulados para dos variables correlacionadas y una variable candidata con multicolinealidad. . . . .	84
B.9. Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros $\beta, \gamma, \sigma$ empleando el método GVS y considerando $\Sigma_j = 100$ , en el estudio con datos simulados para dos variables correlacionadas y una variable candidata con multicolinealidad. . . . .	85
B.10. Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros $\beta, \gamma, \sigma$ empleando el método SSVS y considerando $(\phi_j, C_j) = (0.1, 500)$ , en el estudio con datos simulados dos variables correlacionadas y una variable relevante con multicolinealidad. . . . .	86

B.11. Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros  $\beta$ ,  $\gamma$ ,  $\sigma$  empleando el método KM y considerando  $D_0 = 16I$ , en el estudio con datos simulados para dos variables correlacionadas y una variable relevante con multicolinealidad. . . . . 87

B.12. Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros  $\beta$ ,  $\gamma$ ,  $\sigma$  empleando el método GVS y considerando  $\Sigma_j = 100$ , en el estudio con datos simulados para dos variables correlacionadas y una variable relevante con multicolinealidad. . . . . 88

B.13. Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros  $\beta$ ,  $\gamma$ ,  $\sigma$  empleando el método SSVS y considerando  $(\phi_j, C_j) = (0.1, 500)$ , en el estudio de la esperanza de vida al nacer. . . . . 89

B.14. Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros  $\beta$ ,  $\gamma$ ,  $\sigma$  empleando el método KM y considerando  $D_0 = 16I$ , en el estudio de la esperanza de vida al nacer. . . . . 90

B.15. Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros  $\beta$ ,  $\gamma$ ,  $\sigma$  empleando el método GVS y considerando  $\Sigma_j = 1000$ , en el estudio de la esperanza de vida al nacer. . . . . 91





## Índice de cuadros

5.1. Parámetros estimados por mínimos cuadrados ordinarios en el estudio con datos simulados para veinte variables explicativas independientes. . . . .	21
5.2. Probabilidad de selección de los submodelos a partir de la muestra de Gibbs empleando el método SSVS en el estudio con datos simulados para veinte variables explicativas independientes. . . . .	22
5.3. Probabilidad de selección de los submodelos a partir de la muestra de Gibbs empleando el método KM en el estudio con datos simulados para veinte variables explicativas independientes. . . . .	24
5.4. Probabilidad de selección de los submodelos a partir de la muestra de Gibbs empleando el método GVS en el estudio con datos simulados para veinte variables explicativas independientes. . . . .	26
5.5. Mejores submodelos seleccionados por cada uno de los métodos bayesianos en el estudio con datos simulados para veinte variables explicativas independientes.	28
5.6. Coeficientes estimados del submodelo seleccionado por los métodos Lasso y StepAIC en el estudio con datos simulados para veinte variables explicativas independientes. . . . .	29
5.7. Parámetros estimados por mínimos cuadrados ordinarios en el estudio con datos simulados para cuatro variables correlacionadas. . . . .	30
5.8. Mejores submodelos seleccionados por cada uno de los métodos bayesianos en el estudio con datos simulados para cuatro variables correlacionadas. . . . .	31
5.9. Coeficientes estimados del submodelo seleccionado por los métodos Lasso y StepAIC en el estudio con datos simulados para cuatro variables correlacionadas.	32
5.10. Parámetros estimados por mínimos cuadrados ordinarios en el estudio con datos simulados para dos variables correlacionadas y una variable candidata con multicolinealidad. . . . .	36
5.11. Mejores submodelos seleccionados por cada uno de los métodos bayesianos en el estudio con datos simulados para dos variables correlacionadas y una variable candidata con multicolinealidad. . . . .	37
5.12. Coeficientes estimados del submodelo seleccionado por los métodos Lasso y StepAIC en el estudio con datos simulados para dos variables correlacionadas y una variable candidata con multicolinealidad. . . . .	38
5.13. Parámetros estimados por mínimos cuadrados ordinarios en el estudio con datos simulados para dos variables correlacionadas y una variable relevante con multicolinealidad. . . . .	42



5.14. Mejores submodelos seleccionados por cada uno de los métodos bayesianos en el estudio con datos simulados para dos variables correlacionadas y una variable relevante con multicolinealidad. . . . . 43

5.15. Coeficientes estimados del submodelo seleccionado por los métodos Lasso y StepAIC en el estudio con datos simulados para dos variables correlacionadas y una variable relevante con multicolinealidad. . . . . 44

6.1. Descripción de la muestra de países evaluados. . . . . 49

6.2. Análisis de correlación lineal. . . . . 49

6.3. Parámetros estimados por mínimos cuadrados ordinarios en el estudio de la Esperanza de vida al nacer. . . . . 50

6.4. Mejores submodelos seleccionados por cada uno de los métodos bayesianos en el estudio de la Esperanza de vida al nacer. . . . . 51

6.5. Coeficientes estimados del submodelo seleccionado por los métodos Lasso y StepAIC en el estudio de la Esperanza de vida al nacer. . . . . 51



## Capítulo 1

### Introducción

En diversos campos de las ciencias se desea estudiar el comportamiento de una variable respuesta a través de una o más variables explicativas. Este problema se aborda en el análisis de regresión. Para este análisis, habitualmente se dispone de un gran conjunto de variables explicativas, pero el incluir a todas las variables en la construcción del modelo de regresión genera problemas como la existencia de variables explicativas que no guardan relación con la variable respuesta (variables irrelevantes) o de variables explicativas redundantes (variables correlacionadas). La solución a estos problemas consiste en identificar las variables que mejor expliquen el comportamiento de la variable respuesta y en consecuencia ingresen al modelo de regresión. Los métodos que abordan este problema, se les denomina métodos de selección de variables. De esta manera, el objetivo de los métodos de selección de variables es buscar un subconjunto de variables explicativas, que mediante un modelo interpreten correctamente el comportamiento de la variable respuesta, y a la vez brinde un equilibrio entre la bondad de ajuste del modelo y la complejidad.

Dentro de los métodos clásicos de selección de variables se encuentra el método Stepwise. De acuerdo a [Huberty \(1989\)](#), el método fue propuesto por [Efron y Tibshirani \(1977\)](#) y fueron descritos en profundidad por [Draper y Smith \(1981\)](#) y [Jennrich \(1977\)](#). Este método parte de un modelo de regresión con intercepto y va agregando las variables explicativas que tienen mayor relación con la variable respuesta (procedimiento Forward), pero en cada etapa analiza si todas las variables introducidas deben permanecer en el modelo (procedimiento Backward) evitando así un sobreajuste. En cada etapa, para la elección del mejor modelo, se suele utilizar criterios de comparación de modelos, donde destacan, el criterio de información de Akaike (AIC) propuesto por [Akaike \(1973\)](#) y el criterio de información de Bayes (BIC) propuesto por [Schwarz \(1978\)](#). Dichos criterios se componen de un factor que mide la exactitud de la estimación (bondad de ajuste) y otro de penalización que mide la complejidad del modelo (número de parámetros), resolviendo así el problema de sobreajuste del modelo. Dado un conjunto de modelos candidatos, el mejor modelo es el que tiene el mínimo valor en los criterios. Este procedimiento se repite hasta que no se incluya ni se extraiga alguna variable explicativa.

Otro de los métodos de selección de variables es el método Lasso (por sus siglas en inglés Least absolute shrinkage and selection operator) propuesto por [Tibshirani \(1996\)](#). Lasso es un método de regresión penalizada, que estima los parámetros del modelo minimizando la

suma de cuadrados del error sujeto a una restricción para el vector de coeficientes del modelo. La función que minimiza es:

$$\min_{\beta} \left\{ \sum_{i=1}^n \left( y_i - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\},$$

siendo  $\lambda$  un parámetro de penalización, cuyo correspondiente término obliga a algunos coeficientes del modelo a tomar el valor cero y por consecuencia a excluir la variable explicativa asociada. Mediante validación cruzada se indentifica el valor de  $\lambda$  que minimiza el error cuadrático medio del modelo.

De otro lado, existen otros métodos de selección de variables propuestos bajo inferencia bayesiana, cuya estimación usualmente se realiza mediante el muestreador de Gibbs. En el presente proyecto de tesis, se estudiarán tres de estos métodos para el modelo de regresión lineal normal. El primero, es el método de selección de variables por búsqueda estocástica propuesto por [George y McCulloch \(1993\)](#), el cual introduce una variable latente como indicador de selección de la variable explicativa asociada dentro de la distribución a priori de los coeficientes del modelo de regresión. El segundo es el método de selección de variables por medio de distribuciones a priori incondicionales propuesto por [Kuo y Mallick \(1998\)](#), donde se propone incluir una variable latente indicadora de selección en el predictor lineal del modelo de regresión. Por último, se estudiará el método de selección de variables por muestreo de Gibbs propuesto por [Dellaportas et al. \(2002\)](#), el cual es una combinación de los dos métodos anteriores. Aquí la variable latente indicadora de selección está presente en el predictor lineal y en la distribución a priori de los coeficientes del modelo.

### 1.1. Objetivos

El objetivo general de la tesis es presentar tres métodos de selección de variables bajo inferencia bayesiana aplicado al modelo lineal normal, estudiar sus fundamentos y propiedades, compararlos y aplicarlos a un conjunto de datos reales. De manera específica:

- Revisar la literatura acerca de los métodos de selección de variables.
- Estudiar las propiedades de los métodos bajo estudio.
- Implementar computacionalmente los métodos en el software R.
- Realizar un estudio con datos simulados a efectos de evaluar el desempeño de los métodos.
- Aplicar los métodos a un conjunto de datos reales.

### 1.2. Organización del Trabajo

En el Capítulo 2, se presenta el método de selección de variables por búsqueda estocástica. En el Capítulo 3, se expone el método de selección de variables por medio de distribuciones a priori incondicionales. En el Capítulo 4, se desarrolla el método de selección de variables

por muestreo de Gibbs, el cual es una combinación de los dos métodos anteriores. En el Capítulo 5, se implementan y evalúan los tres métodos a un conjunto de datos simulados considerando diferentes escenarios y comparándolos con los métodos Lasso y Stepwise. En el Capítulo 6, se muestra la aplicación de los tres métodos a un conjunto de datos que describen la esperanza de vida al nacer, elaborado por el Programa de las Naciones Unidas para el Desarrollo (PNUD). La esperanza de vida al nacer es una estimación del promedio de años que viviría un grupo de personas nacidas el mismo año, si las condiciones de mortalidad del país evaluado se mantuvieran constantes. Finalmente en el Capítulo 7, se discuten algunas conclusiones obtenidas en este trabajo, analizando las ventajas y desventajas de los métodos. Incluimos finalmente dos anexos, en el primero A, se presentan los códigos utilizados en la implementación de los métodos en el software R (R Core Team, 2016), y en el segundo B, se muestran los gráficos de las distribuciones a posteriori y los gráficos de convergencia de los parámetros de los modelos.



## Capítulo 2

# Método de selección de variables por búsqueda estocástica

### 2.1. Introducción

El método de selección de variables por búsqueda estocástica, denominado de ahora en adelante SSVS (por sus siglas en inglés de Stochastic Search Variable Selection) fue propuesto por [George y McCulloch \(1993\)](#). Este método consiste en considerar el usual modelo de regresión lineal dentro de un modelo jerárquico de mixtura de normales, de modo que un conjunto de variables latentes permita identificar a los posibles subconjuntos de covariables que estarán presentes en el modelo.

Para la estimación de los parámetros del modelo, el método SSVS utiliza el algoritmo de Gibbs. De esta manera, como lo señalan [George y McCulloch \(1993\)](#), los subconjuntos de variables prometedoras se detecta por su frecuente aparición en la muestra de la distribución a posteriori, obtenida mediante este algoritmo.

### 2.2. Método SSVS

Consideremos el modelo de regresión lineal normal

$$\mathbf{y}|\boldsymbol{\beta}, \sigma^2 \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}), \quad (2.1)$$

donde  $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$  es un vector de dimensión  $n \times 1$  que contiene los valores de la variable dependiente para  $n$  sujetos,  $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)$  es una matriz  $n \times p$ ,  $\mathbf{x}_j = (x_{1j}, x_{2j}, \dots, x_{nj})^T$  es un vector que contiene los valores fijos de la  $j$ -ésima variable explicativa,  $j = 1, \dots, p$ ;  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$  es un vector con los coeficientes de regresión del modelo,  $\sigma^2 > 0$  e  $\mathbf{I}$  es la matriz identidad.

Para realizar una selección de variables explicativas, [George y McCulloch \(1993\)](#) proponen considerar el modelo lineal normal dado en (2.1) dentro de un modelo jerárquico más grande. En este, se considera que cada coeficiente de regresión tiene una distribución de mixtura de dos normales con media cero pero con varianzas diferentes, que puede ser representada en forma jerárquica como

$$\begin{aligned}
\beta_j | \gamma_j = 0 &\sim N(0, \tau_j^2) \\
\beta_j | \gamma_j = 1 &\sim N(0, C_j^2 \tau_j^2) \\
\gamma_j &\sim \text{Bernoulli}(p_j) ,
\end{aligned} \tag{2.2}$$

donde  $\gamma_j$  es una variable latente que facilitará la interpretación del modelo,  $0 < p_j < 1$  es la probabilidad a priori que la variable  $j$  sea incluida en el modelo,  $\tau_j^2 > 0$  y  $C_j^2 > 1$  son hiperparámetros a ser fijados previamente. Los autores proponen tomar un valor pequeño para  $\tau_j$  y un valor grande mayor que 1 para  $C_j$ , de modo que

- Si  $\gamma_j = 0$ , entonces  $\beta_j$  tendrá una distribución normal con media 0 y varianza pequeña, por lo que  $\beta_j$  tendrá un valor pequeño que puede ser considerado como 0.
- Si  $\gamma_j = 1$ , entonces  $\beta_j$  tendrá una distribución normal con media 0 y varianza grande, por lo que  $\beta_j$  probablemente tendrá un valor alejado de 0.

De esta manera  $\gamma_j$  puede ser interpretada como una variable latente que indica si la variable  $j$  está presente o no en el modelo. Luego, la distribución a priori conjunta de los coeficientes de regresión  $\boldsymbol{\beta}$  condicional a las variables latentes  $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_p)^T$  es dada por

$$p(\boldsymbol{\beta} | \boldsymbol{\gamma}) = \prod_{j=1}^p \phi_1(\beta_j | 0, \tau_j^2)^{1-\gamma_j} \phi_1(\beta_j | 0, C_j^2 \tau_j^2)^{\gamma_j} , \tag{2.3}$$

donde  $\phi_k(\mathbf{a} | \mathbf{b}, \mathbf{B})$  representa la función de densidad de una variable aleatoria normal  $k$ -variada evaluada en  $\mathbf{a}$ , con vector de medias  $\mathbf{b}$  y matriz de covarianza  $\mathbf{B}$ . La expresión (2.3) se puede reescribir como

$$p(\boldsymbol{\beta} | \boldsymbol{\gamma}) = \prod_{j=1}^p \phi_1(\beta_j | 0, a_j^2 \tau_j^2) ,$$

donde  $a_j$  es definido como

$$a_j = \begin{cases} 1 , & \text{si } \gamma_j = 0 , \\ C_j , & \text{si } \gamma_j = 1 . \end{cases}$$

Bajo esta reparametrización obtenemos que

$$\begin{aligned}
p(\boldsymbol{\beta} | \boldsymbol{\gamma}) &= \prod_{j=1}^p \frac{1}{\sqrt{2\pi} a_j \tau_j} \exp \left\{ -\frac{1}{2} \frac{\beta_j^2}{a_j^2 \tau_j^2} \right\} \\
&= \frac{1}{(2\pi)^{p/2} \prod_{j=1}^p a_j \tau_j} \exp \left\{ -\frac{1}{2} \sum_{j=1}^p \frac{\beta_j^2}{a_j^2 \tau_j^2} \right\}
\end{aligned} \tag{2.4}$$



$$= \frac{1}{(2\pi)^{p/2} |\mathbf{D}_\gamma|^{1/2}} \exp \left\{ -\frac{1}{2} \boldsymbol{\beta}^T \mathbf{D}_\gamma^{-1} \boldsymbol{\beta} \right\} ,$$

donde  $\mathbf{D}_\gamma$  es una matriz diagonal con elementos  $(a_1^2 \tau_1^2, a_2^2 \tau_2^2, \dots, a_p^2 \tau_p^2)$ . Por lo tanto, de la última expresión en (2.4) tenemos que

$$\boldsymbol{\beta} | \boldsymbol{\gamma} \sim N(0, \mathbf{D}_\gamma) . \quad (2.5)$$

Para la distribución a priori de  $\boldsymbol{\gamma}$ , [George y McCulloch \(1993\)](#) sugieren considerar componentes independientes entre sí, por lo que su distribución conjunta será dada por

$$p(\boldsymbol{\gamma}) = \prod_{j=1}^p p_j^{\gamma_j} (1 - p_j)^{1 - \gamma_j} . \quad (2.6)$$

Para el hiperparámetro  $p_j$  se sugiere considerar  $p_j = 0.5$ ,  $\forall j = 1, \dots, p$ , de manera que las variables explicativas tengan la misma probabilidad a priori de ser incluidas al modelo. Finalmente, para  $\sigma^2$  se considera la siguiente distribución a priori conjugada

$$\sigma^2 \sim \text{GammaInversa} \left( \frac{V}{2}, \frac{\delta V}{2} \right) , \quad (2.7)$$

donde  $V > 0$  y  $\delta > 0$ . La notación  $T \sim \text{GammaInversa}(c, d)$  representa que  $T$  sigue una distribución gamma inversa con media  $d/(c-1)$  y varianza  $d^2/[(c-1)^2(c-2)]$ .

### 2.3. Distribución a posteriori

Sea  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T, \sigma^2)^T$  el vector de parámetros del modelo, entonces la distribución a posteriori es dada por

$$p(\boldsymbol{\theta} | \mathbf{y}) \propto p(\mathbf{y} | \boldsymbol{\theta}) \times p(\boldsymbol{\theta}) ,$$

donde la función de verosimilitud  $p(\mathbf{y} | \boldsymbol{\theta})$  corresponde a la del modelo lineal normal definido en (2.1) y la distribución a priori  $p(\boldsymbol{\theta}) = p(\boldsymbol{\beta} | \boldsymbol{\gamma}) p(\boldsymbol{\gamma}) p(\sigma^2)$  se encuentra especificada en (2.5), (2.6) y (2.7). Entonces, tenemos que

$$p(\boldsymbol{\theta} | \mathbf{y}) \propto \phi_n(\mathbf{y} | \mathbf{X} \boldsymbol{\beta}, \sigma^2 \mathbf{I}) \phi_p(\boldsymbol{\beta} | \mathbf{0}, \mathbf{D}_\gamma) \prod_{j=1}^p p_j^{\gamma_j} (1 - p_j)^{1 - \gamma_j} g \left( \sigma^2 | \frac{V}{2}, \frac{\delta V}{2} \right) , \quad (2.8)$$

donde  $g$  representa la función de densidad de una variable aleatoria  $\sigma^2 \sim \text{GammaInversa} \left( \frac{V}{2}, \frac{\delta V}{2} \right)$ . Como podemos observar  $p(\boldsymbol{\theta} | \mathbf{y})$  no tiene forma conocida, razón por la cuál para realizar inferencias sobre este modelo, [George y McCulloch \(1993\)](#) proponen utilizar el muestreador de Gibbs. Para su implementación es necesario obtener las distribuciones condicionales completas de todos los parámetros. Estas serán presentadas en la siguiente sección.

### 2.4. Distribuciones condicionales completas

Para el cálculo de la distribución condicional completa de  $\boldsymbol{\beta}$  tenemos que

$$\begin{aligned}
p(\boldsymbol{\beta}|\boldsymbol{\gamma}, \sigma^2, \mathbf{y}) &\propto \phi_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})\phi_p(\boldsymbol{\beta}|\mathbf{0}, \mathbf{D}_\gamma) \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\} \exp\left\{-\frac{1}{2}\boldsymbol{\beta}^T\mathbf{D}_\gamma^{-1}\boldsymbol{\beta}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\boldsymbol{\beta}^T\left(\mathbf{D}_\gamma^{-1} + \frac{\mathbf{X}^T\mathbf{X}}{\sigma^2}\right)\boldsymbol{\beta} - 2\frac{\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y}}{\sigma^2}\right]\right\}.
\end{aligned}$$

En esta última expresión, notamos que la distribución condicional de  $\boldsymbol{\beta}$  es proporcional al núcleo de una distribución normal multivariada por tanto

$$\boldsymbol{\beta}|\boldsymbol{\gamma}, \sigma^2, \mathbf{y} \sim N(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta), \quad (2.9)$$

donde  $\boldsymbol{\Sigma}_\beta = (\mathbf{D}_\gamma^{-1} + \frac{1}{\sigma^2}\mathbf{X}^T\mathbf{X})^{-1}$  y  $\boldsymbol{\mu}_\beta = \frac{1}{\sigma^2}\boldsymbol{\Sigma}_\beta\mathbf{X}^T\mathbf{y}$ .

De otro lado, la distribución condicional completa de  $\sigma^2$  es proporcional a

$$\begin{aligned}
p(\sigma^2|\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{y}) &\propto \phi_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})g\left(\sigma^2\left|\frac{V}{2}, \frac{\delta V}{2}\right.\right) \\
&\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\} \left(\frac{1}{\sigma^2}\right)^{\frac{V}{2}+1} \exp\left\{-\frac{\delta V}{2}\frac{1}{\sigma^2}\right\} \\
&\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n+V}{2}+1} \exp\left\{-\frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \delta V}{2}\frac{1}{\sigma^2}\right\}.
\end{aligned} \quad (2.10)$$

De (2.10) se sigue que

$$\sigma^2|\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{y} \sim \text{GammaInversa}\left(\frac{n+V}{2}, \frac{\delta V + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2}\right). \quad (2.11)$$

Finalmente, la distribución condicional conjunta de  $\boldsymbol{\gamma}$  es proporcional a

$$p(\boldsymbol{\gamma}|\boldsymbol{\beta}, \sigma^2, \mathbf{y}) \propto \phi_p(\boldsymbol{\beta}|\mathbf{0}, \mathbf{D}_\gamma) \prod_{j=1}^p p_j^{\gamma_j} (1-p_j)^{1-\gamma_j},$$

la cual no tiene forma conocida, sin embargo como veremos ella si tiene una distribución condicional completa reconocible de cada variable latente  $\gamma_j$

$$p(\gamma_j|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}_{(j)}, \mathbf{y}) \propto p(\boldsymbol{\beta}|\boldsymbol{\gamma})p(\boldsymbol{\gamma}).$$

En efecto, considerando que  $\boldsymbol{\gamma}_{(j)}$  es un vector que contiene a todas las variables latentes con excepción de  $\gamma_j$ , la expresión anterior puede ser reescrita como



$$p(\gamma_j | \boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}_{(j)}, \mathbf{y}) \propto p(\boldsymbol{\beta} | \boldsymbol{\gamma}_{(j)}, \gamma_j) p(\boldsymbol{\gamma}_{(j)}, \gamma_j) .$$

Luego, por el supuesto de independencia a priori entre las variables latentes tenemos que

$$\begin{aligned} p(\gamma_j | \boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}_{(j)}, \mathbf{y}) &\propto p(\boldsymbol{\beta} | \boldsymbol{\gamma}_{(j)}, \gamma_j) p(\boldsymbol{\gamma}_{(j)}) p(\gamma_j) \\ &\propto p(\boldsymbol{\beta} | \boldsymbol{\gamma}_{(j)}, \gamma_j) p(\gamma_j) \\ &\propto p(\boldsymbol{\beta} | \boldsymbol{\gamma}_{(j)}, \gamma_j) p_j^{\gamma_j} (1 - p_j)^{1 - \gamma_j} , \end{aligned}$$

donde este resultado se puede expresar como

$$p(\gamma_j | \boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}_{(j)}, \mathbf{y}) = \begin{cases} p(\boldsymbol{\beta} | \boldsymbol{\gamma}_{(j)}, \gamma_j = 0) (1 - p_j) & \text{si } \gamma_j = 0 \\ p(\boldsymbol{\beta} | \boldsymbol{\gamma}_{(j)}, \gamma_j = 1) p_j & \text{si } \gamma_j = 1 \end{cases} .$$

Entonces, como  $\gamma_j$  solo toma valores 0 ó 1, su distribución condicional a posteriori es dada por

$$\gamma_j | \boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}_{(j)}, \mathbf{y} \sim \text{Bernoulli} \left( \frac{h_j}{h_j + f_j} \right) , \quad (2.12)$$

donde  $h_j = p(\boldsymbol{\beta} | \boldsymbol{\gamma}_{(j)}, \gamma_j = 1) p_j$  y  $f_j = p(\boldsymbol{\beta} | \boldsymbol{\gamma}_{(j)}, \gamma_j = 0) (1 - p_j)$  .

La probabilidad a posteriori de los posibles submodelos a seleccionar puede ser encontrada a partir de la distribución a posteriori de  $\boldsymbol{\gamma}$  dada por  $p(\boldsymbol{\gamma} | \boldsymbol{\beta})$ . De esta manera si deseáramos encontrar por ejemplo la probabilidad a posteriori de un modelo que incluya solamente a las dos primeras variables explicativas, ella estaría dada por

$$p(\boldsymbol{\gamma} = (1, 1, 0, \dots, 0)^T | \boldsymbol{\beta}) ,$$

la cual puede ser fácilmente calculada a partir de la muestra de Gibbs obtenida para  $\boldsymbol{\gamma}$ :  $\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2, \dots, \boldsymbol{\gamma}^L$

$$\frac{\sum_{l=1}^L I(\boldsymbol{\gamma}^l = (1, 1, 0, \dots, 0)^T)}{L} ,$$

dicho de otra manera, esta probabilidad se calcula como la frecuencia con la que aparece el submodelo en la muestra de Gibbs.

## Capítulo 3

# Método de selección de variables por medio de distribuciones a priori incondicionales

### 3.1. Introducción

El método de selección de variables por medio de distribuciones a priori incondicionales, denominado de ahora en adelante KM (por sus siglas del nombre de sus autores [Kuo y Mallick \(1998\)](#)), surge como una alternativa más simple al método SSVS. En lugar de construir un modelo jerárquico, proponen incluir una variable indicadora en el predictor lineal del modelo de regresión. La inclusión de la variable indicadora permite que los coeficientes de regresión puedan tomar valores iguales a cero, debido que mezcla una distribución discreta con una continua.

Para la estimación de los parámetros del modelo, el método KM también utiliza el algoritmo de Gibbs. De esta manera, los subconjuntos de variables prometedoras se detecta por su frecuente aparición en la muestra de la distribución a posteriori, obtenida mediante este algoritmo.

### 3.2. Método KM

[Kuo y Mallick \(1998\)](#) consideran el siguiente modelo de regresión lineal normal

$$y_i = \sum_{j=1}^p \beta_j \gamma_j x_{ij} + \epsilon_i, \quad \forall i = 1, \dots, n \quad (3.1)$$

donde  $y_i$  es el  $i$ -ésimo valor de la variable respuesta,  $x_{ij}$  es el  $i$ -ésimo valor de la  $j$ -ésima variable explicativa,  $\beta_j$  es el coeficiente de regresión de la  $j$ -ésima variable explicativa,  $\epsilon_i$  es el  $i$ -ésimo error aleatorio que se asume  $\epsilon_i \sim N(0, \sigma^2)$  y  $\gamma_j$  es una variable indicadora que toma valores 1 y 0. De manera que si  $\gamma_j = 1$  se incluye a la  $j$ -ésima variable explicativa en el modelo y si  $\gamma_j = 0$  se excluye esta del modelo.

El modelo dado en (3.1) puede ser reescrito como

$$y_i = \sum_{j=1}^p \vartheta_j x_{ij} + \epsilon_i, \quad \forall i = 1, \dots, n$$

con  $\vartheta_j = \beta_j \gamma_j$  y puede ser representado como

$$\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma} \sim N(\mathbf{X}\boldsymbol{\vartheta}, \sigma^2\mathbf{I}) , \quad (3.2)$$

donde  $\mathbf{y} = (y_1, \dots, y_n)^T$  es un vector de dimensión  $n \times 1$  que contiene los valores de la variable dependiente,  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_p)$  es una matriz  $n \times p$ ,  $\mathbf{x}_j = (x_{1j}, \dots, x_{nj})^T$  un vector que contiene los valores de la  $j$ -ésima variable explicativa,  $\boldsymbol{\vartheta} = (\beta_1\gamma_1, \dots, \beta_p\gamma_p)^T$  es un vector con los coeficientes de regresión del modelo,  $\sigma^2 > 0$  e  $\mathbf{I}$  es la matriz identidad.

Además por simplicidad, Kuo y Mallick (1998) proponen que a priori los parámetros  $\boldsymbol{\beta}$ ,  $\boldsymbol{\gamma}$  y  $\sigma^2$  sean independientes, esto con la finalidad de facilitar la elección de las distribuciones a priori de los parámetros del modelo.

La distribución a priori conjunta de los coeficientes de regresión  $\boldsymbol{\beta}$  es dada por

$$\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_0, \mathbf{D}_0) . \quad (3.3)$$

Kuo y Mallick (1998) sugieren considerar  $\boldsymbol{\beta}_0 = (0, \dots, 0)^T$  previo centrado y escalado de las variables  $\mathbf{y}$  y  $\mathbf{X}$  y para  $\mathbf{D}_0$  recomiendan una matriz diagonal con términos moderadamente grandes  $\mathbf{D}_0 = K^2\mathbf{I}$ .

Para la distribución a priori de  $\boldsymbol{\gamma}$ , Kuo y Mallick (1998) sugieren considerar las variables latentes  $\gamma_j$  como independientes entre sí, por lo tanto, su distribución conjunta será dada por

$$p(\boldsymbol{\gamma}) = \prod_{j=1}^p p_j^{\gamma_j} (1 - p_j)^{1-\gamma_j} , \quad (3.4)$$

donde para el hiperparámetro  $p_j$ , se sugiere considerar  $p_j = 0.5, \forall j = 1, \dots, p$ , de manera que las variables explicativas tengan la misma probabilidad a priori de ser incluidas en el modelo. Considerando las distribuciones a priori de  $\boldsymbol{\beta}$  y  $\boldsymbol{\gamma}$  dadas en (3.3) y (3.4), se puede representar la distribución a priori de  $\vartheta_j$ , considerada como coeficiente de regresión, como

$$\begin{aligned} \vartheta_j|\gamma_j = 0 &= 0, \\ \vartheta_j|\gamma_j = 1 &\sim N(\beta_{0j}, K^2) . \end{aligned}$$

En esta última expresión, notamos que la reparametrización de los coeficientes de regresión del modelo provienen de la mezcla de una distribución discreta con una continua; lo cual permite que estos tomen valores iguales a cero y en consecuencia la  $j$ -ésima variable explicativa sea excluida del modelo.

Finalmente, para  $\sigma^2$  se considera la siguiente distribución a priori conjugada

$$\sigma^2 \sim \text{GammaInversa} \left( \frac{V}{2}, \frac{\delta V}{2} \right) , \quad (3.5)$$

donde  $V > 0$  y  $\delta > 0$ .

### 3.3. Distribución a posteriori

Sea  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T, \sigma^2)^T$  el vector de parámetros del modelo, entonces la distribución a posteriori es dada por

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta}) \times p(\boldsymbol{\theta}) ,$$

donde la función de verosimilitud  $p(\mathbf{y}|\boldsymbol{\theta})$  corresponde a la del modelo lineal normal definido en (3.2) y la distribución a priori  $p(\boldsymbol{\theta}) = p(\boldsymbol{\beta})p(\boldsymbol{\gamma})p(\sigma^2)$  se encuentra especificada en (3.3), (3.4) y (3.5). Entonces, tenemos que

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto \phi_n(\mathbf{y}|\mathbf{X}\boldsymbol{\vartheta}, \sigma^2\mathbf{I})\phi_p(\boldsymbol{\beta}|\boldsymbol{\beta}_0, \mathbf{D}_0) \prod_{j=1}^p p_j^{\gamma_j} (1-p_j)^{1-\gamma_j} g\left(\sigma^2 \middle| \frac{V}{2}, \frac{\delta V}{2}\right) \quad (3.6)$$

no tiene forma conocida. Kuo y Mallick (1998) proponen por tanto utilizar un muestreador de Gibbs. Para su implementación es necesario obtener las distribuciones condicionales completas de todos los parámetros. Estas serán presentadas en la siguiente sección.

### 3.4. Distribuciones condicionales completas

Para el cálculo de la distribución condicional completa de  $\boldsymbol{\beta}$  tenemos que

$$\begin{aligned} p(\boldsymbol{\beta}|\boldsymbol{\gamma}, \sigma^2, \mathbf{y}) &\propto \phi_n(\mathbf{y}|\mathbf{X}\boldsymbol{\vartheta}, \sigma^2\mathbf{I}) \phi_p(\boldsymbol{\beta}|\boldsymbol{\beta}_0, \mathbf{D}_0) \\ &\propto \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\vartheta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\vartheta})\right\} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \mathbf{D}_0^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right\} . \end{aligned} \quad (3.7)$$

En (3.7) se puede observar que la distribución condicional a posteriori de  $\boldsymbol{\beta}$  se encuentra en términos de  $\boldsymbol{\vartheta}$ , por lo cual se define  $\mathbf{X}^* = (\gamma_1\mathbf{x}_1, \dots, \gamma_p\mathbf{x}_p)$  como una matriz  $n \times p$ , que nos permite escribir

$$\begin{aligned} \boldsymbol{\vartheta}^T \mathbf{X}^T &= (\beta_1\gamma_1, \dots, \beta_p\gamma_p)(\mathbf{x}_1, \dots, \mathbf{x}_p)^T \\ \boldsymbol{\vartheta}^T \mathbf{X}^T &= (\beta_1, \dots, \beta_p)(\gamma_1\mathbf{x}_1, \dots, \gamma_p\mathbf{x}_p)^T \\ \boldsymbol{\vartheta}^T \mathbf{X}^T &= \boldsymbol{\beta}^T \mathbf{X}^{*T} . \end{aligned}$$

Bajo esta reparametrización obtenemos que

$$p(\boldsymbol{\beta}|\boldsymbol{\gamma}, \sigma^2, \mathbf{y}) \propto \exp\left\{-\frac{1}{2}\left[\boldsymbol{\beta}^T \left(\mathbf{D}_0^{-1} + \frac{\mathbf{X}^{*T}\mathbf{X}^*}{\sigma^2}\right) \boldsymbol{\beta} - 2\boldsymbol{\beta}^T \left(\mathbf{D}_0^{-1}\boldsymbol{\beta}_0 + \frac{\mathbf{X}^{*T}\mathbf{y}}{\sigma^2}\right)\right]\right\} \quad (3.8)$$

es proporcional al núcleo de una distribución normal multivariada y por lo tanto

$$\boldsymbol{\beta}|\boldsymbol{\gamma}, \sigma^2, \mathbf{y} \sim N(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) , \quad (3.9)$$

donde  $\boldsymbol{\Sigma}_\beta = \left(\mathbf{D}_0^{-1} + \frac{\mathbf{X}^{*T}\mathbf{X}^*}{\sigma^2}\right)^{-1}$  y  $\boldsymbol{\mu}_\beta = \boldsymbol{\Sigma}_\beta \left(\mathbf{D}_0^{-1}\boldsymbol{\beta}_0 + \frac{\mathbf{X}^{*T}\mathbf{y}}{\sigma^2}\right)$  .

De otro lado, la distribución condicional completa de  $\sigma^2$  es proporcional a

$$\begin{aligned}
p(\sigma^2|\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{y}) &\propto \phi_n(\mathbf{y}|\mathbf{X}\boldsymbol{\vartheta}, \sigma^2\mathbf{I})g\left(\sigma^2\left|\frac{V}{2}, \frac{\delta V}{2}\right.\right) \\
&\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\vartheta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\vartheta})\right\} \left(\frac{1}{\sigma^2}\right)^{\frac{V}{2}+1} \exp\left\{-\frac{\delta V}{2} \frac{1}{\sigma^2}\right\} \\
&\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n+V}{2}+1} \exp\left\{-\frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\vartheta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\vartheta}) + \delta V}{2} \frac{1}{\sigma^2}\right\}. \tag{3.10}
\end{aligned}$$

Así la distribución condicional a posteriori de  $\sigma^2$  tiene la forma de una distribución gamma inversa:

$$\sigma^2|\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{y} \sim \text{GammaInversa}\left(\frac{n+V}{2}, \frac{\delta V + (\mathbf{y} - \mathbf{X}\boldsymbol{\vartheta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\vartheta})}{2}\right). \tag{3.11}$$

Finalmente, si bien la distribución condicional conjunta de  $\boldsymbol{\gamma}$  no tiene forma conocida, si podemos explicitar la distribución condicional de cada variable latente  $\gamma_j$

$$p(\gamma_j|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}_{(j)}, \mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2)p(\boldsymbol{\gamma}).$$

En efecto, considerando que  $\boldsymbol{\gamma}_{(j)}$  es un vector que contiene a todas las variables latentes con excepción de  $\gamma_j$ , la expresión anterior puede ser reescrita como

$$p(\gamma_j|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}_{(j)}, \mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}_{(j)}, \gamma_j, \sigma^2)p(\boldsymbol{\gamma}_{(j)}, \gamma_j).$$

Luego, por el supuesto de independencia a priori entre las variables latentes tenemos que

$$\begin{aligned}
p(\gamma_j|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}_{(j)}, \mathbf{y}) &\propto p(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}_{(j)}, \gamma_j, \sigma^2)p(\boldsymbol{\gamma}_{(j)})p(\gamma_j) \\
&\propto p(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}_{(j)}, \gamma_j, \sigma^2)p(\gamma_j) \\
&\propto p(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}_{(j)}, \gamma_j, \sigma^2)p_j^{\gamma_j}(1-p_j)^{1-\gamma_j},
\end{aligned}$$

donde este resultado se puede expresar como

$$p(\gamma_j|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}_{(j)}, \mathbf{y}) = \begin{cases} p(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}_{(j)}, \gamma_j = 0, \sigma^2)(1-p_j) & \text{si } \gamma_j = 0 \\ p(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}_{(j)}, \gamma_j = 1, \sigma^2)p_j & \text{si } \gamma_j = 1 \end{cases}.$$

Entonces, como  $\gamma_j$  solo toma valores 0 ó 1, su distribución condicional a posteriori es dada por

$$\gamma_j|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}_{(j)}, \mathbf{y} \sim \text{Bernoulli}\left(\frac{h_j}{h_j + f_j}\right), \tag{3.12}$$

donde  $h_j = p(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}_{(j)}, \gamma_j = 1, \sigma^2)p_j$  y  $f_j = p(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}_{(j)}, \gamma_j = 0, \sigma^2)(1 - p_j)$  .

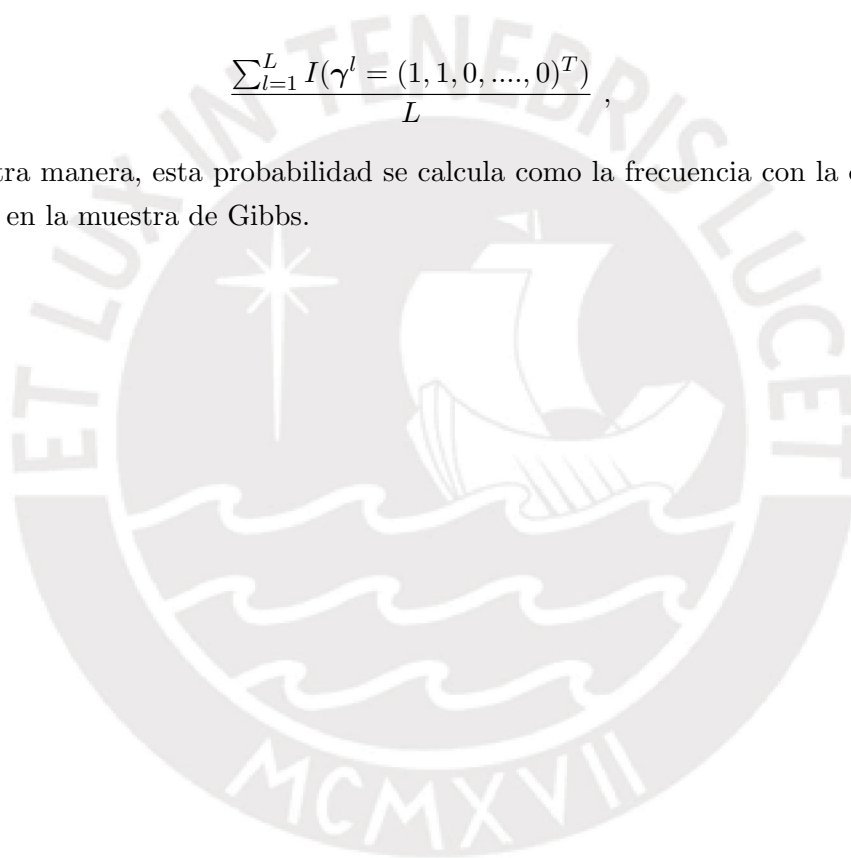
La probabilidad a posteriori de los posibles submodelos a seleccionar puede ser encontrada a partir de la distribución a posteriori de  $\boldsymbol{\gamma}$  dada por  $p(\boldsymbol{\gamma}|\mathbf{y})$ . De esta manera si deseáramos encontrar por ejemplo la probabilidad a posteriori de un modelo que incluya solamente a las dos primeras variables explicativas, ella estaría dada por

$$p(\boldsymbol{\gamma} = (1, 1, 0, \dots, 0)^T | \mathbf{y}) ,$$

esta puede ser fácilmente calculada a partir de la muestra de Gibbs obtenida para  $\boldsymbol{\gamma}$ :  $\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2, \dots, \boldsymbol{\gamma}^L$

$$\frac{\sum_{l=1}^L I(\boldsymbol{\gamma}^l = (1, 1, 0, \dots, 0)^T)}{L} ,$$

dicho de otra manera, esta probabilidad se calcula como la frecuencia con la que aparece el submodelo en la muestra de Gibbs.



## Capítulo 4

# Método de selección de variables por muestreo de Gibbs

### 4.1. Introducción

El método de selección de variables por muestreo de Gibbs, denominado de ahora en adelante GVS (por sus siglas en inglés de Gibbs Variable Selection) fue propuesto por [De-llaportas et al. \(2002\)](#), y ha sido desarrollado e implementado en el software BUGS por [Ntzoufras \(2002\)](#).

El método considera una combinación de los métodos SSVS y KM, donde a diferencia del método SSVS las variables explicativas correspondientes a  $\gamma_j = 0$  son realmente excluidas del modelo y a diferencia del método KM la distribución a priori de los coeficientes de regresión del modelo son dependientes a la variable latente  $\gamma_j$ .

Para la estimación de los parámetros del modelo, el método GVS utiliza el algoritmo de Gibbs. De esta manera, como lo señala [Ntzoufras \(2002\)](#), los subconjuntos de variables prometedoras se detecta por su frecuente aparición en la muestra de la distribución a posteriori obtenida mediante este algoritmo.

### 4.2. Método GVS

[Ntzoufras \(2002\)](#) considera el siguiente modelo de regresión lineal normal

$$y_i = \sum_{j=1}^p \beta_j \gamma_j x_{ij} + \epsilon_i, \quad \forall i = 1, \dots, n \quad (4.1)$$

donde  $y_i$  es el  $i$ -ésimo valor de la variable respuesta,  $x_{ij}$  es el  $i$ -ésimo valor de la  $j$ -ésima variable explicativa,  $\beta_j$  es el coeficiente de regresión de la  $j$ -ésima variable explicativa,  $\epsilon_i$  es el  $i$ -ésimo error aleatorio que se asume  $\epsilon_i \sim N(0, \sigma^2)$  y  $\gamma_j$  es una variable indicadora que toma valores 1 y 0. De manera que si  $\gamma_j = 1$  se incluye la  $j$ -ésima variable explicativa en el modelo y si  $\gamma_j = 0$  se excluye esta del modelo.

El modelo dado en (4.1) puede ser reescrito como

$$y_i = \sum_{j=1}^p \vartheta_j x_{ij} + \epsilon_i, \quad \forall i = 1, \dots, n$$

con  $\vartheta_j = \beta_j \gamma_j$  y puede ser representado como



$$\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma} \sim N(\mathbf{X}\boldsymbol{\vartheta}, \sigma^2\mathbf{I}), \quad (4.2)$$

donde  $\mathbf{y} = (y_1, \dots, y_n)^T$  es un vector de dimensión  $n \times 1$  que contiene los valores de la variable dependiente,  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_p)$  es una matriz  $n \times p$ ,  $\mathbf{x}_j = (x_{1j}, \dots, x_{nj})^T$  un vector que contiene los valores de la  $j$ -ésima variable explicativa,  $\boldsymbol{\vartheta} = (\beta_1\gamma_1, \dots, \beta_p\gamma_p)^T$  es un vector con los coeficientes de regresión del modelo,  $\sigma^2 > 0$  e  $\mathbf{I}$  es la matriz identidad.

Para realizar una selección de variables explicativas [Dellaportas et al. \(2002\)](#) propone considerar el modelo lineal normal dado en (4.2) dentro de un modelo jerárquico más grande, además de incluir la variable latente  $\gamma_j$  al modelo de regresión, lo que facilitará la selección e interpretación del modelo.

En este método, se considera que cada coeficiente de regresión tiene una distribución de mixtura de dos normales con media y varianza diferentes, que puede ser representada en forma jerárquica como

$$\begin{aligned} \beta_j|\gamma_j = 0 &\sim N(\mu_j, S_j) \\ \beta_j|\gamma_j = 1 &\sim N(0, \Sigma_j) \\ \gamma_j &\sim \text{Bernoulli}(p_j), \end{aligned} \quad (4.3)$$

donde  $\gamma_j$  es una variable latente que facilitará la interpretación del modelo,  $0 < p_j < 1$  es la probabilidad a priori que la variable  $j$  sea incluida en el modelo,  $\mu_j$ ,  $S_j$  y  $\Sigma_j$  son hiperparámetros a ser fijados previamente, de modo que

- Si  $\gamma_j = 0$ , entonces  $\beta_j$  tendrá una distribución normal con media  $\mu_j$  y varianza  $S_j$ , por lo que  $\beta_j$  tendrá un valor similar a  $\mu_j$ .
- Si  $\gamma_j = 1$ , entonces  $\beta_j$  tendrá una distribución normal con media 0 y varianza grande, por lo que  $\beta_j$  probablemente tendrá un valor alejado de 0.

De lo anterior, la distribución a priori conjunta de los coeficientes de regresión  $\boldsymbol{\beta}$  condicional a la variable latente  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_p)^T$  se representa como

$$\boldsymbol{\beta}|\boldsymbol{\gamma} \sim N(\boldsymbol{\beta}_\boldsymbol{\gamma}, \mathbf{D}_\boldsymbol{\gamma}), \quad (4.4)$$

donde  $\boldsymbol{\beta}_\boldsymbol{\gamma} = (a_1, \dots, a_p)^T$  y  $\mathbf{D}_\boldsymbol{\gamma}$  es una matriz diagonal con elementos  $(c_1, \dots, c_p)$ , siendo  $a_j$  y  $c_j$  definidos como

$$a_j = \begin{cases} \mu_j, & \text{si } \gamma_j = 0, \\ 0, & \text{si } \gamma_j = 1. \end{cases}$$

$$c_j = \begin{cases} S_j, & \text{si } \gamma_j = 0, \\ \Sigma_j, & \text{si } \gamma_j = 1. \end{cases}$$



Para la distribución a priori de  $\gamma$ , [Ntzoufras \(2002\)](#) sugiere considerar componentes independientes entre sí, por lo que su distribución conjunta será dada por

$$p(\gamma) = \prod_{j=1}^p p_j^{\gamma_j} (1 - p_j)^{1 - \gamma_j}. \quad (4.5)$$

Para el hiperparámetro  $p_j$  se sugiere considerar  $p_j = 0.5, \forall j = 1, \dots, p$ , de manera que las variables explicativas tengan la misma probabilidad a priori de ser incluidas al modelo.

Considerando las distribuciones a priori de  $\beta$  y  $\gamma$  dadas en (4.4) y (4.5), se puede representar la distribución a priori de  $\vartheta_j$ , considerada como coeficiente de regresión, como

$$\begin{aligned} \vartheta_j | \gamma_j = 0 &= 0, \\ \vartheta_j | \gamma_j = 1 &\sim N(0, \Sigma_j). \end{aligned}$$

En esta última expresión, notamos que la reparametrización de los coeficientes de regresión del modelo provienen de la mezcla de una distribución discreta con una continua; lo cual permite que tomen valores iguales a cero y en consecuencia la  $j$ -ésima variable explicativa sea excluida del modelo. De esta manera  $\gamma_j$  puede ser interpretada como una variable que indica si la  $j$ -ésima variable está presente o no en el modelo.

Finalmente, para  $\sigma^2$  se considera la siguiente distribución a priori conjugada

$$\sigma^2 \sim \text{GammaInversa} \left( \frac{V}{2}, \frac{\delta V}{2} \right), \quad (4.6)$$

donde  $V > 0$  y  $\delta > 0$ .

### 4.3. Distribución a posteriori

Sea  $\theta = (\beta^T, \gamma^T, \sigma^2)^T$  el vector de parámetros del modelo, entonces la distribución a posteriori es dada por

$$p(\theta | \mathbf{y}) \propto p(\mathbf{y} | \theta) \times p(\theta)$$

donde la función de verosimilitud  $p(\mathbf{y} | \theta)$  corresponde a la del modelo lineal normal definido en (4.2) y la distribución a priori  $p(\theta) = p(\beta | \gamma) p(\gamma) p(\sigma^2)$  se encuentra especificada en (4.4), (4.5) y (4.6). Entonces, tenemos que

$$p(\theta | \mathbf{y}) \propto \phi_n(\mathbf{y} | \mathbf{X}\theta, \sigma^2 \mathbf{I}) \phi_p(\beta | \beta_\gamma, \mathbf{D}_\gamma) \prod_{j=1}^p p_j^{\gamma_j} (1 - p_j)^{1 - \gamma_j} g \left( \sigma^2 \mid \frac{V}{2}, \frac{\delta V}{2} \right) \quad (4.7)$$

no tiene forma conocida, por tanto [Ntzoufras \(2002\)](#) proponen utilizar un muestreador de Gibbs. Para su implementación es necesario obtener las distribuciones condicionales de todos

los parámetros. Estas serán presentadas en la siguiente sección.

#### 4.4. Distribuciones condicionales completas

Para el cálculo de la distribución condicional completa de  $\beta$  tenemos que

$$\begin{aligned} p(\beta|\gamma, \sigma^2, \mathbf{y}) &\propto \phi_n(\mathbf{y}|\mathbf{X}\boldsymbol{\vartheta}, \sigma^2\mathbf{I}) \phi_p(\beta|\beta_\gamma, \mathbf{D}_\gamma) \\ &\propto \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\vartheta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\vartheta})\right\} \exp\left\{-\frac{1}{2}(\beta - \beta_\gamma)^T \mathbf{D}_\gamma^{-1}(\beta - \beta_\gamma)\right\}. \end{aligned} \quad (4.8)$$

En (4.8) se puede observar que la distribución condicional a posteriori de  $\beta$  se encuentra en términos de  $\boldsymbol{\vartheta}$ , por lo cual se define  $\mathbf{X}^* = (\gamma_1\mathbf{x}_1, \dots, \gamma_p\mathbf{x}_p)$  como una matriz  $n \times p$  que nos permite escribir

$$\begin{aligned} \boldsymbol{\vartheta}^T \mathbf{X}^T &= (\beta_1\gamma_1, \dots, \beta_p\gamma_p)(\mathbf{x}_1, \dots, \mathbf{x}_p)^T \\ \boldsymbol{\vartheta}^T \mathbf{X}^T &= (\beta_1, \dots, \beta_p)(\gamma_1\mathbf{x}_1, \dots, \gamma_p\mathbf{x}_p)^T \\ \boldsymbol{\vartheta}^T \mathbf{X}^T &= \beta^T \mathbf{X}^{*T}. \end{aligned}$$

Bajo esta reparametrización obtenemos que

$$p(\beta|\gamma, \sigma^2, \mathbf{y}) \propto \exp\left\{-\frac{1}{2}\left[\beta^T \left(\mathbf{D}_\gamma^{-1} + \frac{\mathbf{X}^{*T}\mathbf{X}^*}{\sigma^2}\right)\beta - 2\beta^T \left(\mathbf{D}_\gamma^{-1}\beta_\gamma + \frac{\mathbf{X}^{*T}\mathbf{y}}{\sigma^2}\right)\right]\right\} \quad (4.9)$$

es proporcional al núcleo de una distribución normal multivariada y por lo tanto

$$\beta|\gamma, \sigma^2, \mathbf{y} \sim N(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta), \quad (4.10)$$

donde  $\boldsymbol{\Sigma}_\beta = \left(\mathbf{D}_\gamma^{-1} + \frac{\mathbf{X}^{*T}\mathbf{X}^*}{\sigma^2}\right)^{-1}$  y  $\boldsymbol{\mu}_\beta = \boldsymbol{\Sigma}_\beta \left(\mathbf{D}_\gamma^{-1}\beta_\gamma + \frac{\mathbf{X}^{*T}\mathbf{y}}{\sigma^2}\right)$ .

De otro lado, la distribución condicional completa de  $\sigma^2$  es proporcional a

$$\begin{aligned} p(\sigma^2|\beta, \gamma, \mathbf{y}) &\propto \phi_n(\mathbf{y}|\mathbf{X}\boldsymbol{\vartheta}, \sigma^2\mathbf{I})g\left(\sigma^2\left|\frac{V}{2}, \frac{\delta V}{2}\right.\right) \\ &\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\vartheta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\vartheta})\right\} \left(\frac{1}{\sigma^2}\right)^{\frac{V}{2}+1} \exp\left\{-\frac{\delta V}{2} \frac{1}{\sigma^2}\right\} \\ &\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n+V}{2}+1} \exp\left\{-\frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\vartheta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\vartheta}) + \delta V}{2} \frac{1}{\sigma^2}\right\}. \end{aligned} \quad (4.11)$$

Así la distribución condicional a posteriori de  $\sigma^2$  tiene la forma de una distribución gamma inversa:

$$\sigma^2|\beta, \gamma, \mathbf{y} \sim \text{GammaInversa}\left(\frac{n+V}{2}, \frac{\delta V + (\mathbf{y} - \mathbf{X}\boldsymbol{\vartheta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\vartheta})}{2}\right). \quad (4.12)$$

Finalmente, si bien la distribución condicional conjunta de  $\boldsymbol{\gamma}$  no tiene forma conocida, si podemos explicitar la distribución condicional de cada variable latente  $\gamma_j$

$$p(\gamma_j | \boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}_{(j)}, \mathbf{y}) \propto p(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2) p(\boldsymbol{\beta} | \boldsymbol{\gamma}) p(\boldsymbol{\gamma}).$$

En efecto, considerando que  $\boldsymbol{\gamma}_{(j)}$  es un vector que contiene a todas las variables latentes con excepción de  $\gamma_j$ , la expresión anterior puede ser reescrita como

$$p(\gamma_j | \boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}_{(j)}, \mathbf{y}) \propto p(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\gamma}_{(j)}, \gamma_j, \sigma^2) p(\boldsymbol{\beta} | \boldsymbol{\gamma}_{(j)}, \gamma_j) p(\boldsymbol{\gamma}_{(j)}, \gamma_j).$$

Luego, por el supuesto de independencia a priori entre las variables latentes tenemos que

$$\begin{aligned} p(\gamma_j | \boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}_{(j)}, \mathbf{y}) &\propto p(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\gamma}_{(j)}, \gamma_j, \sigma^2) p(\boldsymbol{\beta} | \boldsymbol{\gamma}_{(j)}, \gamma_j) p(\boldsymbol{\gamma}_{(j)}) p(\gamma_j) \\ &\propto p(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\gamma}_{(j)}, \gamma_j, \sigma^2) p(\boldsymbol{\beta} | \boldsymbol{\gamma}_{(j)}, \gamma_j) p(\gamma_j) \\ &\propto p(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\gamma}_{(j)}, \gamma_j, \sigma^2) p(\boldsymbol{\beta} | \boldsymbol{\gamma}_{(j)}, \gamma_j) p_j^{\gamma_j} (1 - p_j)^{1 - \gamma_j}, \end{aligned}$$

donde este resultado se puede expresar como

$$p(\gamma_j | \boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}_{(j)}, \mathbf{y}) = \begin{cases} p(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\gamma}_{(j)}, \gamma_j = 0, \sigma^2) p(\boldsymbol{\beta} | \boldsymbol{\gamma}_{(j)}, \gamma_j = 0) (1 - p_j) & \text{si } \gamma_j = 0 \\ p(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\gamma}_{(j)}, \gamma_j = 1, \sigma^2) p(\boldsymbol{\beta} | \boldsymbol{\gamma}_{(j)}, \gamma_j = 1) p_j & \text{si } \gamma_j = 1 \end{cases}.$$

Entonces, como  $\gamma_j$  solo toma valores 0 ó 1, su distribución condicional a posteriori es dada por

$$\gamma_j | \boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}_{(j)}, \mathbf{y} \sim \text{Bernoulli} \left( \frac{h_j}{h_j + f_j} \right), \quad (4.13)$$

donde  $h_j = p(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\gamma}_{(j)}, \gamma_j = 1, \sigma^2) p(\boldsymbol{\beta} | \boldsymbol{\gamma}_{(j)}, \gamma_j = 1) p_j$  y

$$f_j = p(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\gamma}_{(j)}, \gamma_j = 0, \sigma^2) p(\boldsymbol{\beta} | \boldsymbol{\gamma}_{(j)}, \gamma_j = 1) (1 - p_j).$$

La probabilidad a posteriori de los posibles submodelos a seleccionar puede ser encontrada a partir de la distribución a posteriori de  $\boldsymbol{\gamma}$  dada por  $p(\boldsymbol{\gamma} | \mathbf{y}, \boldsymbol{\beta})$ . De esta manera si deseáramos encontrar por ejemplo la probabilidad a posteriori de un modelo que incluya solamente a las dos primeras variables explicativas, ella estaría dada por

$$p(\boldsymbol{\gamma} = (1, 1, 0, \dots, 0)^T | \mathbf{y}, \boldsymbol{\beta}),$$

la cual puede ser fácilmente calculada a partir de la muestra de Gibbs obtenida para  $\boldsymbol{\gamma}$ :  $\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2, \dots, \boldsymbol{\gamma}^L$

$$\frac{\sum_{l=1}^L I(\gamma^l = (1, 1, 0, \dots, 0)^T)}{L},$$

dicho de otra manera, esta probabilidad se calcula como la frecuencia con la que aparece el submodelo en la muestra de Gibbs.



## Capítulo 5

### Estudio de simulación

En el presente capítulo se analiza el comportamiento de los tres métodos de selección de variables anteriormente presentados, mediante un estudio de simulación. En un primer escenario se considera un modelo de regresión lineal normal con veinte variables explicativas independientes e idénticamente distribuidas, en un segundo escenario se consideran veinte variables explicativas de las cuales cuatro de ellas se encuentran correlacionadas, en un tercer escenario se consideran veinte variables explicativas de las cuales dos variables están correlacionadas y una variable candidata presenta multicolinealidad y un último escenario se consideran veinte variables explicativas de las cuales dos variables se encuentran correlacionadas y una variable relevante presenta multicolinealidad. El objetivo es evaluar los resultados de cada método bajo estos escenarios y evaluar los distintos valores de los hiperparámetros de los modelos recomendado por los autores. La implementación de los métodos se realizó a través del algoritmo de Gibbs, ejecutando 15000 iteraciones de las cuales fueron descartadas las primeras 2000 (ver apéndice A).

#### 5.1. Escenario 1: Veinte variables explicativas independientes.

Para este escenario, se toma como referencia el trabajo de [George y McCulloch \(1993\)](#) donde considera cinco variables explicativas independientes. Para esta tesis, se amplía el trabajo a veinte variables explicativas independientes e idénticamente distribuidas con un tamaño de muestra igual a  $n = 100$ . En vista de que las veinte variables explicativas van a competir para explicar el comportamiento de la variable respuesta  $\mathbf{y}$ , se genera intencionalmente  $\mathbf{y}$  a partir de solamente cinco de estas variables explicativas ( $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5$ ). Esto con la finalidad de conocer previamente las variables que deberían ser seleccionadas por los métodos y observar como afecta la presencia de un gran número de variables que no contribuyen al comportamiento de la variable respuesta. A continuación, se detallada la generación de las variables.

Las veinte variables explicativas, fueron generadas de una distribución normal con media 0 y varianza 1,

$$\mathbf{x}_j \sim N(0, 1), \forall j = 1, 2, \dots, 20 .$$

La variable respuesta  $\mathbf{y}$  fue generada de acuerdo al siguiente modelo

$$y_i = 2 x_{i1} + 2 x_{i2} + 2 x_{i3} + 2 x_{i4} + 2 x_{i5} + \epsilon_i, \forall i = 1, \dots, 100 ; \text{donde } \epsilon_i \sim N(0, 4) .$$

En el Cuadro 5.1, se muestran los parámetros estimados por el método de mínimos cuadrados ordinarios para el modelo de regresión con las veinte variables explicativas.

Coef. Regresión	Coef. Estimados	Desv. Estándar	p-valor
$\beta_1$	1.8930	0.2116	0.0000
$\beta_2$	1.8699	0.2216	0.0000
$\beta_3$	1.8166	0.2023	0.0000
$\beta_4$	1.8893	0.1877	0.0000
$\beta_5$	2.1703	0.2082	0.0000
$\beta_6$	0.1436	0.1883	0.4480
$\beta_7$	0.2262	0.1966	0.2530
$\beta_8$	-0.0335	0.2026	0.8690
$\beta_9$	-0.0226	0.2542	0.9290
$\beta_{10}$	0.2651	0.2122	0.2150
$\beta_{11}$	-0.1727	0.2045	0.4010
$\beta_{12}$	-0.1308	0.1873	0.4870
$\beta_{13}$	0.0749	0.1931	0.6990
$\beta_{14}$	-0.0585	0.2098	0.7810
$\beta_{15}$	-0.0624	0.1985	0.7540
$\beta_{16}$	-0.0837	0.1979	0.6740
$\beta_{17}$	-0.0443	0.1967	0.8220
$\beta_{18}$	0.1884	0.1842	0.3090
$\beta_{19}$	0.0276	0.1989	0.8900
$\beta_{20}$	0.0996	0.1792	0.5800

Cuadro 5.1: Parámetros estimados por mínimos cuadrados ordinarios en el estudio con datos simulados para veinte variables explicativas independientes.

Adicionalmente a ello, el error estándar residual obtenido es  $\hat{\sigma} = 1.801$ . Se observa del Cuadro 5.1, que los valores  $p$  de las variables  $(x_1, x_2, x_3, x_4, x_5)$  son menores a 0.05, lo que indica que son significativas.

### 5.1.1. Implementación SSVS

Para la implementación del método SSVS, se tomará  $p_j = 0.5$ ,  $V = 0.01$  y  $\delta = 0.01$ ; y, para  $\tau_j$  como [George y McCulloch \(1993\)](#) sugieren, se considerará que este es proporcional al error estándar de cada coeficiente esto es

$$\tau_j = \phi_j \hat{\sigma}_{\beta_j} ,$$

donde  $\hat{\sigma}_{\beta_j}$  es el error estándar estimado por el método de mínimos cuadrados ordinarios considerando todas las variables explicativas. Luego, tomaremos en cuenta los mismos valores que [George y McCulloch \(1993\)](#) proponen para  $(\phi_j, C_j) = \{(1, 5), (0.1, 100), (0.1, 500)\}$ .

Estos valores surgen de un estudio de simulación donde la elección de  $\tau_j$  y  $C_j$  deben ser apropiadas, de modo que, cuando  $\gamma_j = 0$  los  $\beta_j$  provenientes de  $N(0, \tau_j^2)$  deben tomar un



valor cercano a 0, y cuando  $\gamma_j = 1$ , los  $\beta_j$  provenientes de  $N(0, C_j^2 \tau_j^2)$  deben tomar un valor alejado de 0. Esta definición hace que los  $\beta_j$  cercanos a 0 puedan ser excluidos del modelo de regresión. George y McCulloch (1993) consideran que este enfoque tiene la característica de ser invariante bajo el cambio de escala de las variables explicativas, además a medida que disminuye  $\phi_j$  y aumenta  $C_j$  las distribuciones  $N(0, \tau_j^2)$  y  $N(0, C_j^2 \tau_j^2)$  originan los valores deseados para  $\beta_j$ .

En el Cuadro 5.2, se presenta el número de submodelos seleccionados, y los submodelos con mayor frecuencia de aparición en la muestra de Gibbs para cada valor sugerido de  $(\phi_j, C_j)$ ; ordenados de manera descendente por su probabilidad de selección.

Modelos	(1,5)	(0.1,100)	(0.1,500)
Nº <i>Submodelos</i>	4577	1212	167
$x_1, x_2, x_3, x_4, x_5$	<b>0.0176</b>	<b>0.1371</b>	<b>0.6541</b>
$x_1, x_2, x_3, x_4, x_5, x_{10}$	0.0078	0.0452	0.0485
$x_1, x_2, x_3, x_4, x_5, x_7$	0.0066	0.0337	0.0252
$x_1, x_2, x_3, x_4, x_5, x_{13}$	0.0065	0.0343	0.0239
$x_1, x_2, x_3, x_4, x_5, x_{11}$	0.0058	0.0201	0.0214
$x_1, x_2, x_3, x_4, x_5, x_{18}$	0.0076	0.0220	0.0202
$x_1, x_2, x_3, x_4, x_5, x_9$	0.0062	0.0127	0.0196
$x_1, x_2, x_3, x_4, x_5, x_{15}$	0.0049	0.0158	0.0164
$x_1, x_2, x_3, x_4, x_5, x_{20}$	0.0055	0.0209	0.0160
$x_1, x_2, x_3, x_4, x_5, x_8$	0.0045	0.0126	0.0156

Cuadro 5.2: Probabilidad de selección de los submodelos a partir de la muestra de Gibbs empleando el método SSVS en el estudio con datos simulados para veinte variables explicativas independientes.

Del Cuadro 5.2 observamos que el submodelo  $(x_1, x_2, x_3, x_4, x_5)$ , para cualquier elección de los valores de los hiperparámetros  $(\phi_j, C_j)$ , presenta la mayor probabilidad de selección, lo que indica que el método SSVS lo selecciona como el mejor. Cuando se emplea la sugerencia  $(\phi_j, C_j) = (0.1, 500)$ , el método presenta un mejor desempeño, ya que selecciona un menor número de submodelos posibles (167) y el submodelo elegido para la creación de la variable respuesta  $\mathbf{y}$  presenta una mayor probabilidad de selección (0.6541).

En la Figura 5.1, se presenta la distribución a posteriori de los coeficientes de regresión, donde observamos que la distribución de los coeficientes de las variables explicativas excluidas del modelo se encuentran centradas en 0.

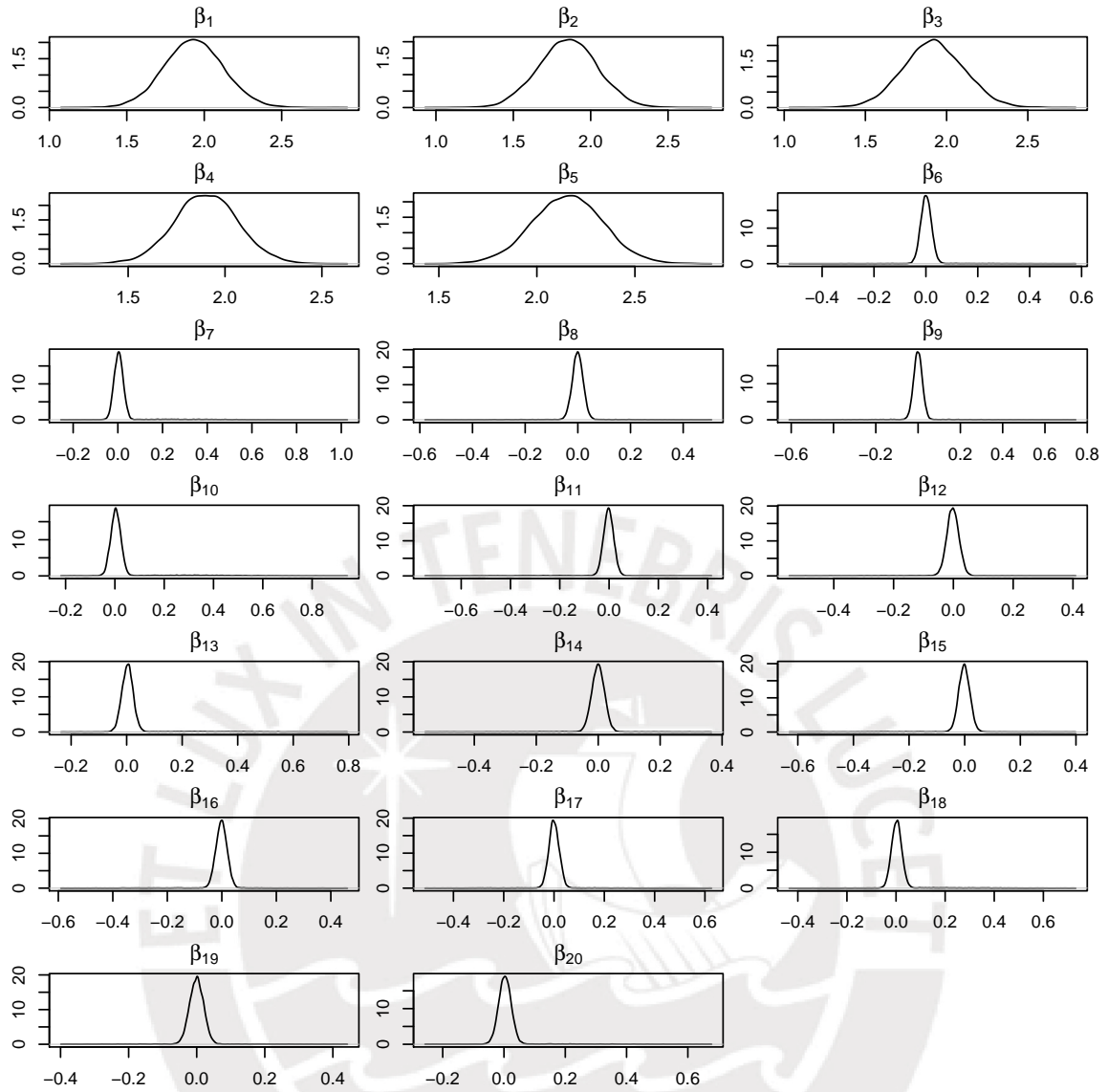


Figura 5.1: Distribución a posteriori de  $\beta$  en el estudio con datos simulados para veinte variables explicativas independientes empleando el método SSVS y considerando  $(\phi_j, C_j) = (0.1, 500)$

### 5.1.2. Implementación KM

Para la implementación del método KM, se consideran los hiperparámetros  $p_j = 0.5$ ,  $\beta_0 = (0, \dots, 0)^T$ ,  $V = 0.01$  y  $\delta = 0.01$  y dos valores para la matriz de varianza covarianza de la distribución a priori de  $\beta$ , que fueron sugeridos por [Kuo y Mallick \(1998\)](#), los cuales son  $D_0 = I$  y  $D_0 = 16I$ . Para la elección de  $K$ , [Kuo y Mallick \(1998\)](#) asumen un modelo de regresión con intercepto y la estandarización de las variables  $\mathbf{x}_j$ ,  $\forall j = 1, \dots, p$ , e  $\mathbf{y}$ . Basados en lo anterior y en un estudio de simulación, proponen un  $D_0$  con términos moderadamente grande; puesto que, a mayor valor en los términos de la matriz de varianza y covarianza los coeficientes del modelo tomán valores alejados del 0, permitiendo una mejor selección de las variables explicativas. De acuerdo a lo anterior, consideran que  $K$  debe tomar un valor entre 0.5 y 4. Por razones prácticas, recomiendan trabajar con dos escenarios  $K = 1$  ( $D_0 = I$ ) y  $K = 4$  ( $D_0 = 16I$ ).



En el Cuadro 5.3, se presenta el número de submodelos seleccionados, y los submodelos con mayor frecuencia de aparición en la muestra de Gibbs para cada valor sugerido de  $D_0$ ; ordenados de manera descendente por su probabilidad de selección.

Modelos	$D_0 = I$	$D_0 = 16I$
Nº <i>Submodelos</i>	3586	590
$x_1, x_2, x_3, x_4, x_5$	<b>0.0228</b>	<b>0.3626</b>
$x_1, x_2, x_3, x_4, x_5, x_{10}$	0.0166	0.0644
$x_1, x_2, x_3, x_4, x_5, x_{13}$	0.0110	0.0393
$x_1, x_2, x_3, x_4, x_5, x_7$	0.0155	0.0329
$x_1, x_2, x_3, x_4, x_5, x_{20}$	0.0085	0.0326
$x_1, x_2, x_3, x_4, x_5, x_{18}$	0.0073	0.0288
$x_1, x_2, x_3, x_4, x_5, x_{11}$	0.0052	0.0263
$x_1, x_2, x_3, x_4, x_5, x_{15}$	0.0051	0.0222
$x_1, x_2, x_3, x_4, x_5, x_{12}$	0.0046	0.0194
$x_1, x_2, x_3, x_4, x_5, x_9$	0.0061	0.0192

Cuadro 5.3: Probabilidad de selección de los submodelos a partir de la muestra de Gibbs empleando el método KM en el estudio con datos simulados para veinte variables explicativas independientes.

Del Cuadro 5.3, observamos que para los dos escenarios el submodelo con mayor probabilidad de selección es  $(x_1, x_2, x_3, x_4, x_5)$ . Además, cuando se considera  $D_0 = 16I$ , el método KM presenta un mejor desempeño, ya que selecciona un número menor de submodelos (590) y el submodelo elegido para la creación de la variable respuesta  $y$  presenta una mayor probabilidad de selección (0.3626).

De acuerdo a los resultados, el método selecciona un gran número de submodelos posibles 590 de  $2^{20}$ , esto hace que los submodelos tengan una baja presencia en la muestra de Gibbs (menor a 0.5), lo que indica que el método KM se ve afectado por el gran número de variables explicativas.

En la Figura 5.2, se presenta la distribución a posteriori de los coeficientes de regresión, donde observamos que la distribución de los coeficientes de las variables explicativas excluidas del modelo se encuentran centradas en 0.

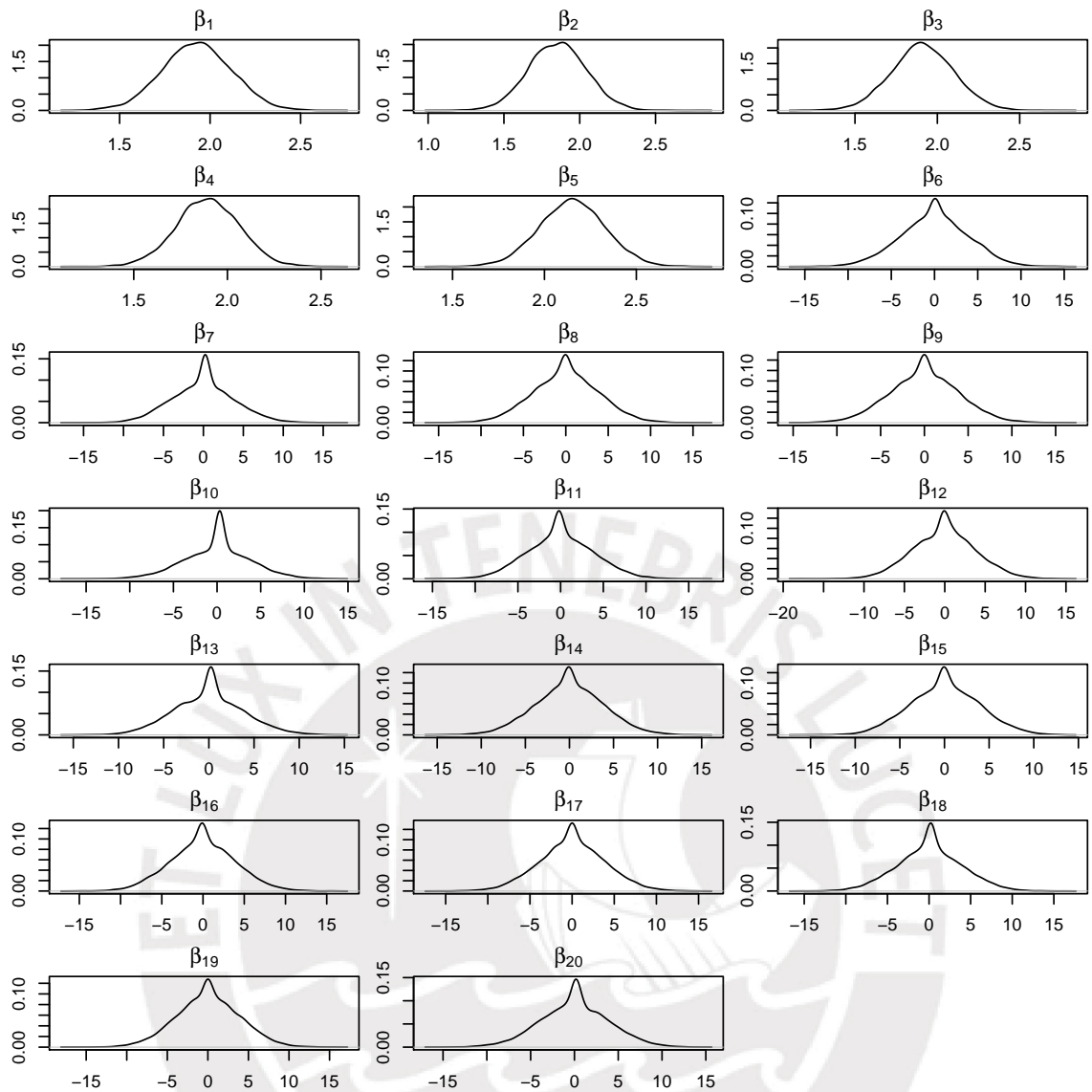


Figura 5.2: Distribución a posteriori de  $\beta$  en el estudio con datos simulados para veinte variables explicativas independientes empleando el método KM y considerando  $D_0 = 16I$

### 5.1.3. Implementación GVS

Para la implementación del método GVS, se toma  $p_j = 0.5$ ,  $V = 0.01$ ,  $\delta = 0.01$ ,  $\mu_j = 0$ , para  $S_j$  como [Ntzoufras \(2002\)](#) sugieren, se considerará que este es proporcional a la varianza de la distribución a priori de los  $\beta_j$  seleccionados, esto es

$$S_j = \omega \Sigma_j ,$$

donde recomienda  $\omega = 0.01$ . Luego, tomaremos en cuenta los mismos valores que [Ntzoufras \(2002\)](#) proponen para  $\Sigma_j = \{1, 10, 100\}$ . De acuerdo a [Ntzoufras \(2002\)](#), estos valores surgen de un estudio de simulación y se basan en la propiedad de la distribución a priori de  $\beta$  introducida en el método SSVS por [George y McCulloch \(1993\)](#), donde mencionan que los  $\beta_j$

incluidos al modelo deben de provenir de una distribución normal con varianza grande y los no incluidos de una distribución normal con varianza pequeña.

En el Cuadro 5.4, se presenta el número de submodelos, y los submodelos con mayor frecuencia de aparición en la muestra de Gibbs para cada valor sugerido de  $\Sigma_j$ ; ordenados de manera descendente por su probabilidad de selección.

Modelos	$\Sigma_j = 1$	$\Sigma_j = 10$	$\Sigma_j = 100$
N° Submodelos	3610	829	207
$x_1, x_2, x_3, x_4, x_5$	<b>0.0232</b>	<b>0.2626</b>	<b>0.6461</b>
$x_1, x_2, x_3, x_4, x_5, x_{10}$	0.0181	0.0607	0.0452
$x_1, x_2, x_3, x_4, x_5, x_{13}$	0.0128	0.0402	0.0317
$x_1, x_2, x_3, x_4, x_5, x_7$	0.0132	0.0418	0.0311
$x_1, x_2, x_3, x_4, x_5, x_{18}$	0.0071	0.0259	0.0245
$x_1, x_2, x_3, x_4, x_5, x_{11}$	0.0061	0.0262	0.0207
$x_1, x_2, x_3, x_4, x_5, x_{20}$	0.0090	0.0274	0.0207
$x_1, x_2, x_3, x_4, x_5, x_9$	0.0042	0.0221	0.0141
$x_1, x_2, x_3, x_4, x_5, x_{12}$	0.0064	0.0192	0.0134
$x_1, x_2, x_3, x_4, x_5, x_{14}$	0.0054	0.0165	0.0134

Cuadro 5.4: Probabilidad de selección de los submodelos a partir de la muestra de Gibbs empleando el método GVS en el estudio con datos simulados para veinte variables explicativas independientes.

Del Cuadro 5.4, observamos que para todos los escenarios el submodelo con mayor probabilidad de selección es  $(x_1, x_2, x_3, x_4, x_5)$ . Además, cuando se considera  $\Sigma_j = 100$ , el método GVS presenta un mejor desempeño, ya que selecciona un número menor de submodelos (207) y el submodelo elegido para la creación de la variable respuesta  $\mathbf{y}$  presenta una mayor probabilidad de selección (0.6461).

En la Figura 5.3, se presenta la distribución a posteriori de los coeficientes de regresión, donde observamos que la distribución de los coeficientes de las variables explicativas excluidas del modelo se encuentran centradas en 0.

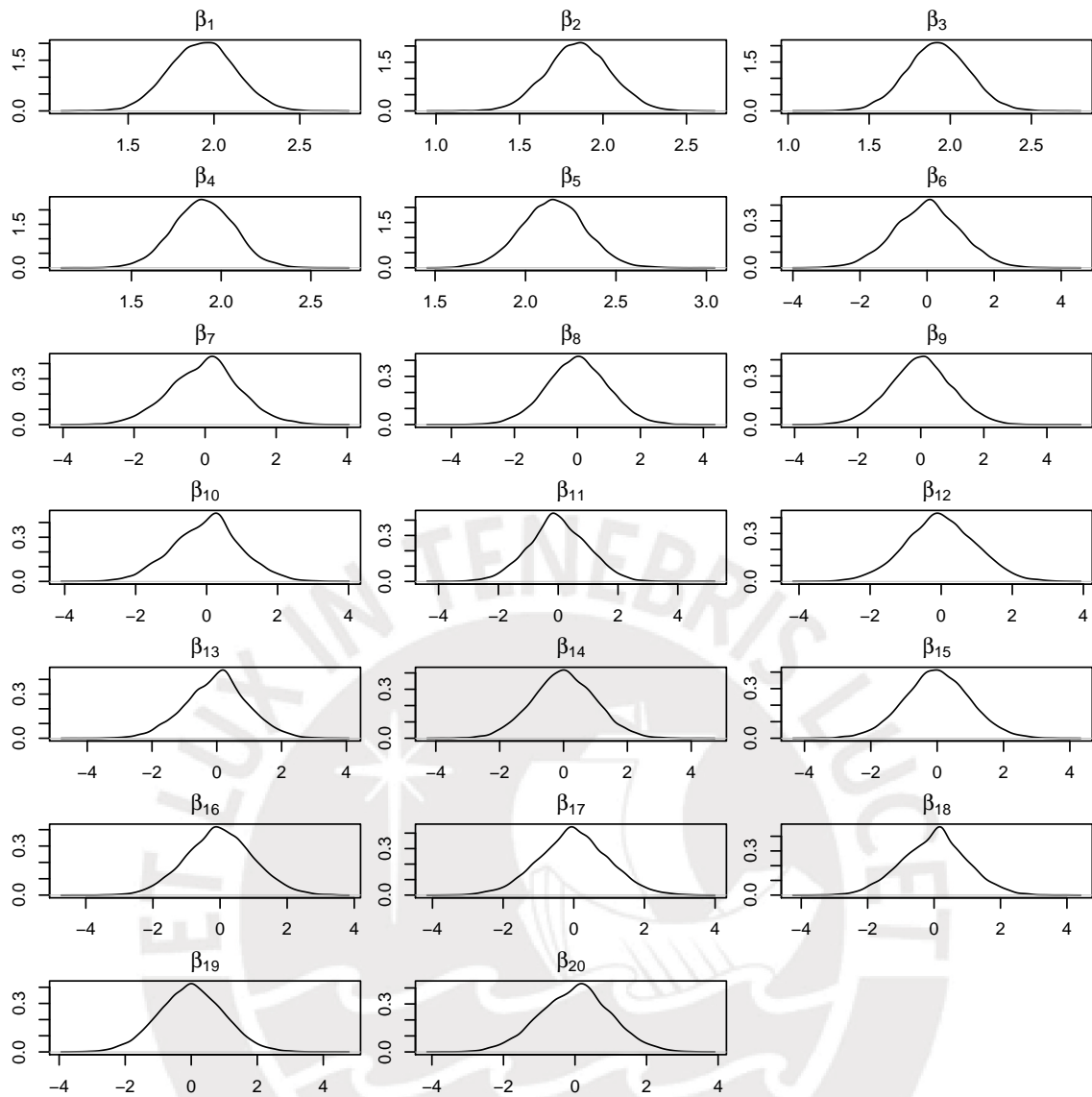


Figura 5.3: Distribución a posteriori de  $\beta$  en el estudio con datos simulados para veinte variables explicativas independientes empleando el método GVS y considerando  $\Sigma_j = 100$ .

#### 5.1.4. Resumen de resultados

En el Cuadro 5.5, se muestra un resumen con los mejores resultados observados en cada uno de los métodos implementados. El cuadro contiene los 10 submodelos con mayor frecuencia de aparición en la muestra de Gibbs, ordenados de manera descendente por su probabilidad de selección.

Método	SSVS	KM	GVS
Parámetros	$(\phi_j, C_j) = (0.1, 500)$	$\mathbf{D}_0 = 16\mathbf{I}$	$\Sigma_j = 100$
N° <i>Sudmodelos</i>	167	590	207
$x_1, x_2, x_3, x_4, x_5$	<b>0.6541</b>	<b>0.3626</b>	<b>0.6461</b>
$x_1, x_2, x_3, x_4, x_5, x_{10}$	0.0485	0.0644	0.0452
$x_1, x_2, x_3, x_4, x_5, x_7$	0.0252	0.0329	0.0311
$x_1, x_2, x_3, x_4, x_5, x_{13}$	0.0239	0.0393	0.0317
$x_1, x_2, x_3, x_4, x_5, x_{11}$	0.0214	0.0263	0.0207
$x_1, x_2, x_3, x_4, x_5, x_{18}$	0.0202	0.0288	0.0245
$x_1, x_2, x_3, x_4, x_5, x_9$	0.0196	0.0192	0.0141
$x_1, x_2, x_3, x_4, x_5, x_{15}$	0.0164	0.0222	0.0125
$x_1, x_2, x_3, x_4, x_5, x_{20}$	0.0160	0.0326	0.0207
$x_1, x_2, x_3, x_4, x_5, x_8$	0.0156	0.0169	0.0130

Cuadro 5.5: Mejores submodelos seleccionados por cada uno de los métodos bayesianos en el estudio con datos simulados para veinte variables explicativas independientes.

Los resultados del Cuadro 5.5 muestran que los tres métodos bayesianos seleccionan al modelo  $(x_1, x_2, x_3, x_4, x_5)$  con una alta probabilidad de selección en comparación al resto, lo cual resulta correcto, debido a que dicho submodelo fue empleado para la generación de la variable respuesta  $y$ . Asimismo, observamos que el método SSVS presenta un mejor desempeño en comparación al resto, ya que selecciona al modelo  $(x_1, x_2, x_3, x_4, x_5)$  con una mayor probabilidad de selección 0.6541 y el número de submodelos que aparecen en la muestra Gibbs es menor 167. Por otro lado, ocurre lo contrario con el método KM el cual se ve afectado por el gran número de variables explicativas, repercutiendo en la probabilidad de selección del submodelo correcto.

En el presente trabajo, a modo de comparación se ejecutó los métodos Lasso y Stepwise. Se emplearon las funciones *cv.glmnet* y *stepAIC* de las librerías *glmnet* Tibshirani et al. (2008) y *MASS* Ripley et al. (2018) del software R (ver apéndice A). Los coeficientes estimados del submodelo seleccionado por los métodos son mostrados en el Cuadro 5.6.

Coef. Regresión	Lasso	StepAIC
$\beta_0$	0.0000	0.0000
$\beta_1$	<b>1.4816</b>	<b>1.9705</b>
$\beta_2$	<b>1.5376</b>	<b>1.8608</b>
$\beta_3$	<b>1.4852</b>	<b>1.8844</b>
$\beta_4$	<b>1.5530</b>	<b>1.9019</b>
$\beta_5$	<b>1.9233</b>	<b>2.1331</b>
$\beta_6$	0.0000	0.0000
$\beta_7$	<b>0.0320</b>	0.0000
$\beta_8$	0.0000	0.0000
$\beta_9$	0.0000	0.0000
$\beta_{10}$	0.0000	<b>0.3051</b>
$\beta_{11}$	0.0000	0.0000
$\beta_{12}$	0.0000	0.0000
$\beta_{13}$	<b>0.0086</b>	0.0000
$\beta_{14}$	0.0000	0.0000
$\beta_{15}$	0.0000	0.0000
$\beta_{16}$	0.0000	0.0000
$\beta_{17}$	0.0000	0.0000
$\beta_{18}$	0.0000	0.0000
$\beta_{19}$	0.0000	0.0000
$\beta_{20}$	0.0000	0.0000

Cuadro 5.6: Coeficientes estimados del submodelo seleccionado por los métodos Lasso y StepAIC en el estudio con datos simulados para veinte variables explicativas independientes.

Los resultados indican que los métodos Lasso y Stepwise seleccionan a los submodelos  $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_7, \mathbf{x}_{13})$  y  $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_{10})$ , respectivamente.

En base a los resultados para este escenario, los métodos bayesianos seleccionan al submodelo correcto  $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5)$ . Ocurre lo contrario con los métodos tradicionales Lasso y Stepwise, dado que los submodelos seleccionados contienen variables no significativas tales como  $\mathbf{x}_7$ ,  $\mathbf{x}_{10}$  y  $\mathbf{x}_{13}$ .

## 5.2. Escenario 2: Cuatro variables correlacionadas

En este escenario, se considera lo desarrollado por [George y McCulloch \(1993\)](#) donde dos de sus cinco variables explicativas se encuentran correlacionados. Para este trabajo, se generan veinte variables aleatorias de un tamaño de muestra igual a  $n = 100$ . De los cuales  $(\mathbf{x}_6, \mathbf{x}_1)$  y  $(\mathbf{x}_7, \mathbf{x}_2)$  tiene una alta correlación lineal 0.9791 y 0.9869 respectivamente. Además, se genera intencionalmente la variable respuesta  $\mathbf{y}$  a partir de  $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5)$ . A continuación, se detalla la generación de las variables:

$$\mathbf{x}_j \sim N(0, 1), \forall j = 1, 2, \dots, 20 - \{6, 7\} .$$



$$x_{i6} = x_{i1} + 0.21W_i, \forall i = 1, \dots, 100 ; \text{ donde } W_i \sim N(0, 1) .$$

$$x_{i7} = x_{i2} + 0.18Z_i, \forall i = 1, \dots, 100 ; \text{ donde } Z_i \sim N(0, 1) .$$

La variable respuesta  $\mathbf{y}$  fue generada de acuerdo al siguiente modelo

$$y_i = 2 x_{i1} + 2 x_{i2} + 2 x_{i3} + 2 x_{i4} + 2 x_{i5} + \epsilon_i, \forall i = 1, 2, \dots, 100 ; \text{ donde } \epsilon_i \sim N(0, 4) .$$

En el Cuadro 5.1, se muestran los parámetros estimados por el método de mínimos cuadrados ordinarios para el modelo de regresión con las veinte variables explicativas.

Coef. Regresión	Coef. Estimados	Desv. Estándar	p-valor
$\beta_1$	2.7433	1.3703	0.0487
$\beta_2$	1.7265	1.4022	0.2218
$\beta_3$	2.0187	0.2643	0.0001
$\beta_4$	2.2965	0.2388	0.0001
$\beta_5$	2.0026	0.2283	0.0001
$\beta_6$	-0.2914	1.3133	0.8250
$\beta_7$	0.3356	1.3386	0.8027
$\beta_8$	0.0502	0.2570	0.8458
$\beta_9$	0.2562	0.2258	0.2600
$\beta_{10}$	0.0606	0.2287	0.7920
$\beta_{11}$	0.0177	0.2654	0.9471
$\beta_{12}$	0.0514	0.2215	0.8172
$\beta_{13}$	-0.1355	0.2587	0.6020
$\beta_{14}$	-0.1416	0.2747	0.6078
$\beta_{15}$	0.1583	0.2079	0.4487
$\beta_{16}$	0.0278	0.2705	0.9186
$\beta_{17}$	-0.2852	0.2387	0.2356
$\beta_{18}$	-0.1303	0.2340	0.5791
$\beta_{19}$	0.3679	0.2463	0.1391
$\beta_{20}$	0.3110	0.2713	0.2551

Cuadro 5.7: Parámetros estimados por mínimos cuadrados ordinarios en el estudio con datos simulados para cuatro variables correlacionadas.

Adicionalmente a ello, el error estándar residual obtenido es  $\hat{\sigma} = 2.211$ . Se observa del Cuadro 5.7, que los valores  $p$  de las variables ( $\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_4$  y  $\mathbf{x}_5$ ) son menores a 0.05, lo que indica que son significativas.

### 5.2.1. Resumen de resultados

Considerando los resultados obtenidos en el escenario 1, se optó por utilizar los hiperparámetros  $(\phi_j, C_j) = (0.1, 500)$ ,  $\mathbf{D}_0 = 16\mathbf{I}$ ,  $\Sigma_j = 100$  los cuales conllevan a un mejor desempeño



de los métodos bayesianos SSVS, KM y GVS, respectivamente.

En el Cuadro 5.8, se muestra un resumen con los mejores resultados observados en cada uno de los métodos implementados. Este contiene los 10 submodelos con mayor frecuencia de aparición en la muestra de Gibbs, ordenados de manera descendente por su probabilidad de selección.

Método	SSVS	KM	GVS
Parámetros	$(\phi_j, C_j) = (0.1, 500)$	$\mathbf{D}_0 = 16\mathbf{I}$	$\Sigma_j = 100$
N° Submodelos	198	1361	508
$x_1, x_2, x_3, x_4, x_5$	<b>0.3834</b>	<b>0.1216</b>	<b>0.3069</b>
$x_1, x_3, x_4, x_5, x_7$	0.2261	0.0781	0.1881
$x_2, x_3, x_4, x_5, x_6$	0.0845	0.0222	0.0452
$x_3, x_4, x_5, x_6, x_7$	0.0722	0.0077	0.0226
$x_1, x_2, x_3, x_4, x_5, x_7$	0.0264	0.0422	0.0426
$x_1, x_2, x_3, x_4, x_5, x_6$	0.0198	0.0384	0.0252
$x_1, x_3, x_4, x_5, x_6, x_7$	0.0109	0.0222	0.0214
$x_1, x_2, x_3, x_4, x_5, x_9$	0.0102	0.0138	0.0131
$x_1, x_2, x_3, x_4, x_5, x_{19}$	0.0095	0.0170	0.0164
$x_2, x_3, x_4, x_5, x_6, x_7$	0.0080	0.0069	0.0069
$x_1, x_2, x_3, x_4, x_5, x_{20}$	0.0073	0.0160	0.0148

Cuadro 5.8: Mejores submodelos seleccionados por cada uno de los métodos bayesianos en el estudio con datos simulados para cuatro variables correlacionadas.

Se observa que los tres métodos bayesianos seleccionan al submodelo correcto  $(x_1, x_2, x_3, x_4, x_5)$  con una probabilidad no mayor a 0.4. Debido a la alta correlación lineal de las variables explicativas  $(x_6, x_1)$  y  $(x_7, x_2)$  los métodos bayesianos presentan a los submodelos  $(x_1, x_3, x_4, x_5, x_7)$ ,  $(x_2, x_3, x_4, x_5, x_6)$  y  $(x_3, x_4, x_5, x_6, x_7)$  como potenciales candidatos a seleccionar por considerarlas equivalentes a  $(x_1, x_2, x_3, x_4, x_5)$ . Esto se debe a que los métodos bayesianos emplean las distribuciones de las variables explicativas para el cálculo de probabilidad a posteriori de los submodelos. En las Figuras 5.4, 5.5 y 5.6 se presenta la distribución a posteriori de los coeficientes de regresión, donde observamos que la distribución de los coeficientes de las variables explicativas excluidas del modelo se encuentran centradas en 0.

De la misma manera que en el escenario anterior, se ejecutó los métodos Lasso y Stepwise en el software R. Por su parte, la estimación de los coeficientes de la regresión de los submodelos seleccionados se muestran en el Cuadro 5.9.

Coef. Regresión	Lasso	Stepwise
$\beta_0$	0.0000	0.0000
$\beta_1$	<b>1.8683</b>	<b>2.3967</b>
$\beta_2$	<b>1.0355</b>	<b>2.0761</b>
$\beta_3$	<b>1.6368</b>	<b>2.0118</b>
$\beta_4$	<b>1.9221</b>	<b>2.2109</b>
$\beta_5$	<b>1.6596</b>	<b>2.1003</b>
$\beta_6$	0.0000	0.0000
$\beta_7$	<b>0.5400</b>	0.0000
$\beta_8$	0.0000	0.0000
$\beta_9$	0.0000	<b>0.3154</b>
$\beta_{10}$	0.0000	0.0000
$\beta_{11}$	0.0000	0.0000
$\beta_{12}$	0.0000	0.0000
$\beta_{13}$	0.0000	0.0000
$\beta_{14}$	0.0000	0.0000
$\beta_{15}$	0.0000	0.0000
$\beta_{16}$	0.0000	0.0000
$\beta_{17}$	0.0000	0.0000
$\beta_{18}$	0.0000	0.0000
$\beta_{19}$	0.0000	<b>0.3187</b>
$\beta_{20}$	0.0000	<b>0.3574</b>

Cuadro 5.9: Coeficientes estimados del submodelo seleccionado por los métodos Lasso y StepAIC en el estudio con datos simulados para cuatro variables correlacionadas.

Para este escenario de simulación, los métodos Lasso y Stepwise seleccionan a los submodelos  $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_7)$  y  $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_9, \mathbf{x}_{19}, \mathbf{x}_{20})$ , respectivamente.

De acuerdo a los resultados, los métodos bayesianos se ven ligeramente afectados por la alta correlación de las variables explicativas, sin embargo la información que presentan contribuyen a una mejor toma de decisiones, debido a que muestran los potenciales submodelos a seleccionar. Por otra parte, los métodos Lasso y Stepwise incluyen variables explicativas no significativas dentro de los submodelos seleccionados.

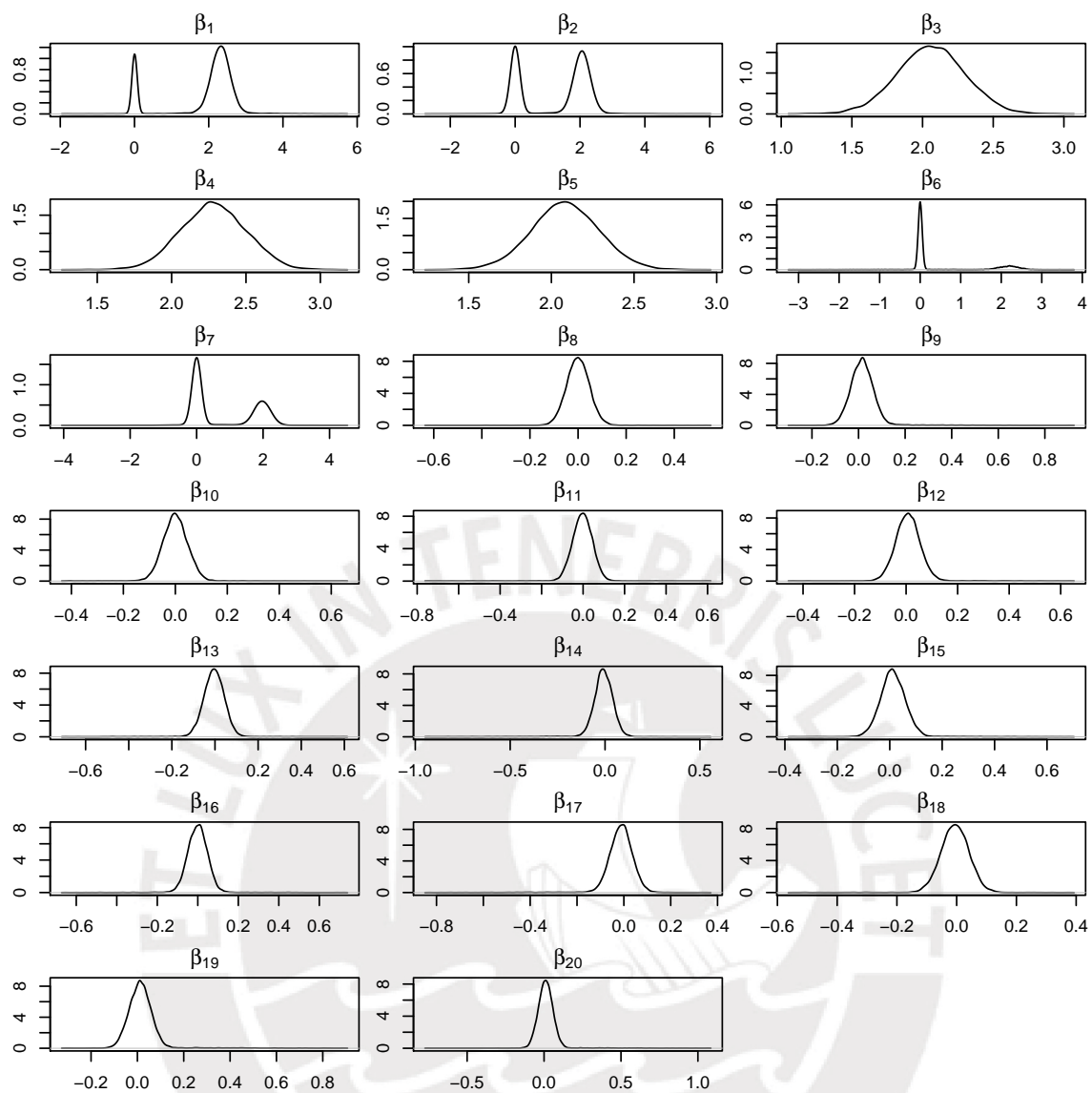


Figura 5.4: Distribución a posteriori de  $\beta$  en el estudio con datos simulados para cuatro variables correlacionadas empleando el método SVSS y considerando  $(\phi_j, C_j) = (0.1, 500)$ .

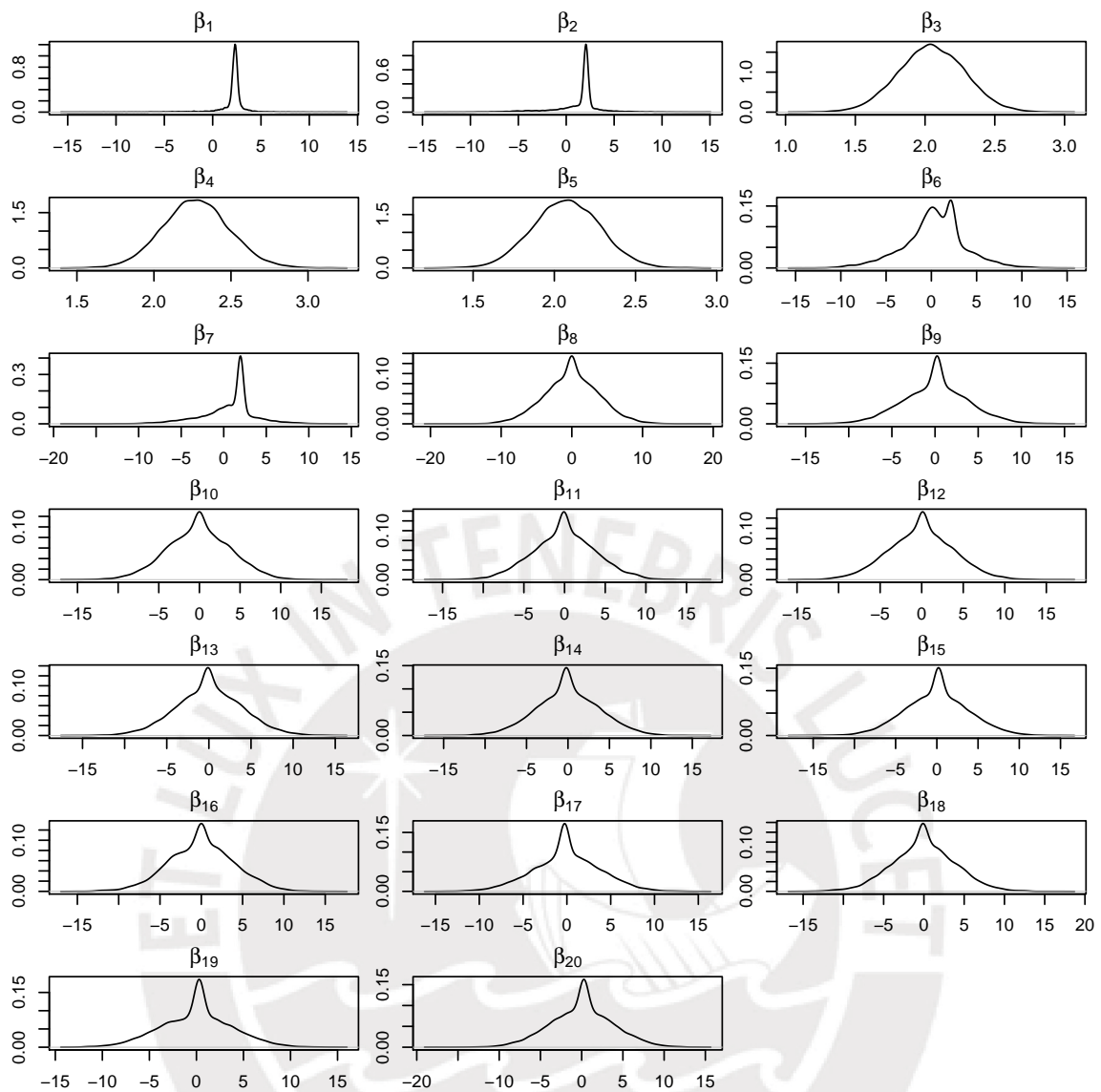


Figura 5.5: Distribución a posteriori de  $\beta$  en el estudio con datos simulados para cuatro variables correlacionadas empleando el método KM y considerando  $D_0 = 16I$ .

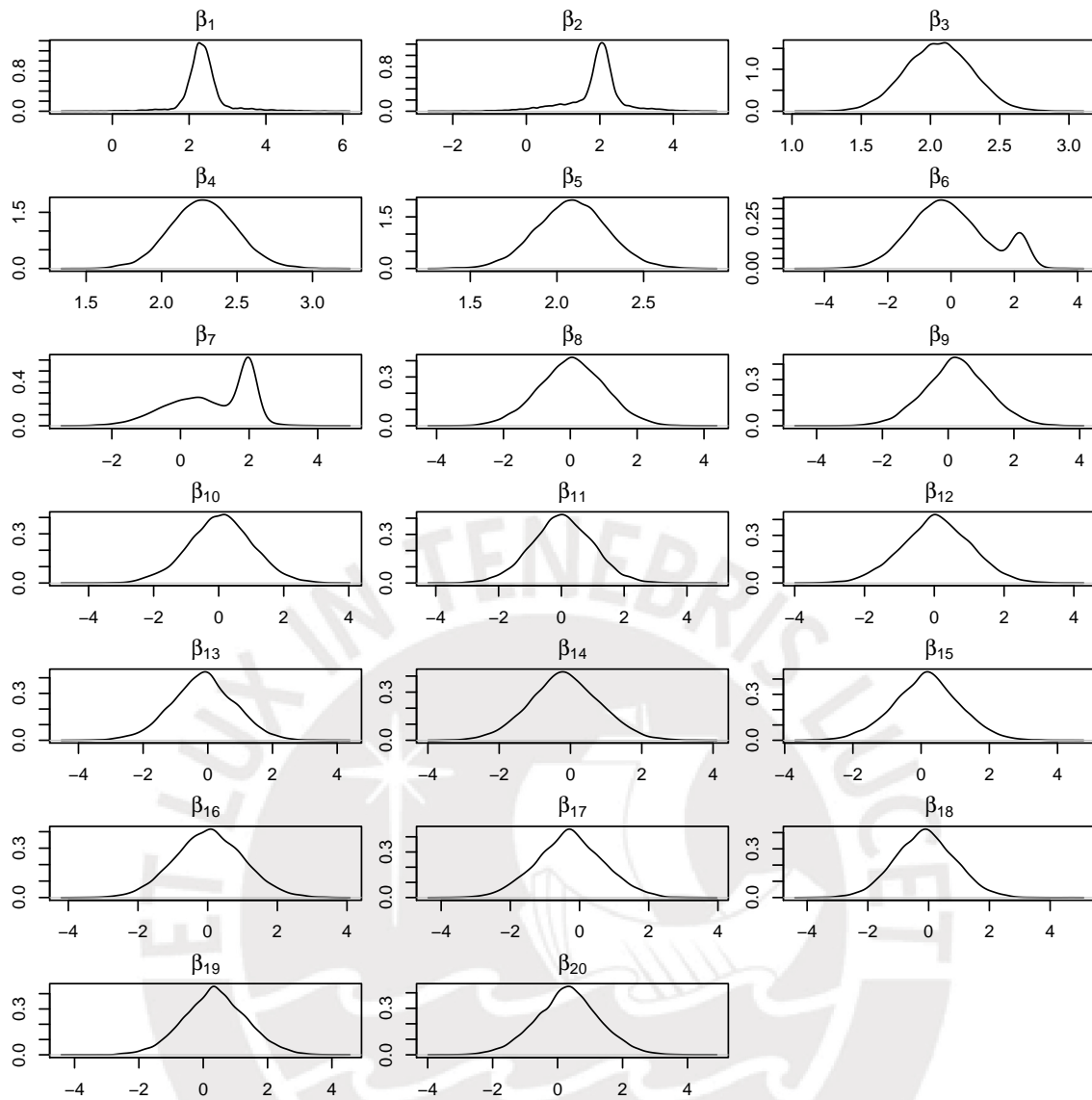


Figura 5.6: Distribución a posteriori de  $\beta$  en el estudio con datos simulados para cuatro variables correlacionadas empleando el método GVS y considerando  $\Sigma_j = 100$ .

### 5.3. Escenario 3: Dos variables correlacionadas y una variable candidata con multicolinealidad

En este escenario de elaboración propia, se tomó como referencia los escenarios anteriores y se realizó algunas variantes. Se generan veinte variables aleatorias de un tamaño de muestra igual a  $n = 100$ . De los cuales  $(\mathbf{x}_6, \mathbf{x}_1)$  tiene una alta correlación lineal 0.9791 y  $\mathbf{x}_7$  presenta multicolinealidad con  $(\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_8)$ . Además, la variable respuesta  $\mathbf{y}$  es generada intencionalmente por  $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5)$ . A continuación, se detalla la generación de las variables:

$$\mathbf{x}_j \sim N(0, 1), \forall j = 1, 2, \dots, 20 - \{6, 7\}.$$

$$x_{i6} = x_{i1} + 0.21W_i, \forall i = 1, \dots, 100 ; \text{ donde } W_i \sim N(0, 1) .$$

$$x_{i7} = x_{i2} + x_{i3} + x_{i8} + 0.18Z_i, \forall i = 1, \dots, 100 ; \text{ donde } Z_i \sim N(0, 1) .$$

La variable respuesta  $\mathbf{y}$  fue generada de acuerdo al siguiente modelo

$$y_i = 2 x_{i1} + 2 x_{i2} + 2 x_{i3} + 2 x_{i4} + 2 x_{i5} + \epsilon_i, \forall i = 1, \dots, 100 ; \text{ donde } \epsilon_i \sim N(0, 4) .$$

En el Cuadro 5.10, se muestran los parámetros estimados por el método de mínimos cuadrados ordinarios para el modelo de regresión con las veinte variables explicativas.

Coef. Regresión	Coef. Estimados	Desv. Estándar	p-valor
$\beta_1$	2.7433	1.3703	0.0487
$\beta_2$	1.7265	1.4022	0.2218
$\beta_3$	1.6831	1.3403	0.2129
$\beta_4$	2.2965	0.2388	0.0001
$\beta_5$	2.0026	0.2283	0.0001
$\beta_6$	-0.2914	1.3133	0.8250
$\beta_7$	0.3356	1.3386	0.8027
$\beta_8$	-0.2855	1.3359	0.8313
$\beta_9$	0.2562	0.2258	0.2600
$\beta_{10}$	0.0606	0.2287	0.7920
$\beta_{11}$	0.0177	0.2654	0.9471
$\beta_{12}$	0.0514	0.2215	0.8172
$\beta_{13}$	-0.1355	0.2587	0.6020
$\beta_{14}$	-0.1416	0.2747	0.6078
$\beta_{15}$	0.1583	0.2079	0.4487
$\beta_{16}$	0.0278	0.2705	0.9186
$\beta_{17}$	-0.2852	0.2387	0.2356
$\beta_{18}$	-0.1303	0.2340	0.5791
$\beta_{19}$	0.3679	0.2463	0.1391
$\beta_{20}$	0.3110	0.2713	0.2551

Cuadro 5.10: Parámetros estimados por mínimos cuadrados ordinarios en el estudio con datos simulados para dos variables correlacionadas y una variable candidata con multicolinealidad.

Adicionalmente a ello, el error estándar residual obtenido es  $\hat{\sigma} = 2.211$ . Se observa del Cuadro 5.10, que los valores  $p$  de las variables ( $\mathbf{x}_1$ ,  $\mathbf{x}_4$  y  $\mathbf{x}_5$ ) son menores a 0.05, lo que indica que son significativas.

### 5.3.1. Resumen de resultados

Tomando referencia al escenario anterior, se optó por trabajar con los hiperparámetros que conllevan a un mejor desempeño de los métodos bayesianos.

En el Cuadro 5.11, se muestra un resumen con los mejores resultados observados en cada uno de los métodos implementados. Este contiene los 10 submodelos con mayor frecuencia de aparición en la muestra de Gibbs, ordenados de manera descendente por su probabilidad de selección.

Método	SSVS	KM	GVS
Parámetros	$(\phi_j, C_j) = (0.1, 500)$	$\mathbf{D_0} = 16\mathbf{I}$	$\Sigma_j = 100$
N° Submodelos	121	1303	484
$x_1, x_2, x_3, x_4, x_5$	<b>0.6964</b>	<b>0.1779</b>	<b>0.3692</b>
$x_2, x_3, x_4, x_5, x_6$	0.1293	0.0198	0.0304
$x_1, x_2, x_3, x_4, x_5, x_6$	0.0320	0.0495	0.0382
$x_1, x_4, x_5, x_7, x_8$	-	0.0539	0.1511
$x_4, x_5, x_6, x_7, x_8$	-	0.0075	0.0262
$x_1, x_2, x_3, x_4, x_5, x_{20}$	0.0099	0.0267	0.0218
$x_1, x_2, x_3, x_4, x_5, x_{19}$	0.0124	0.0231	0.0211
$x_1, x_4, x_5, x_6, x_7, x_8$	-	0.0132	0.0182
$x_1, x_2, x_3, x_4, x_5, x_{17}$	0.0087	0.0200	0.0165
$x_1, x_2, x_3, x_4, x_5, x_9$	0.0116	0.0242	0.0159
$x_1, x_2, x_3, x_4, x_5, x_{15}$	0.0073	0.0142	0.0138

Cuadro 5.11: Mejores submodelos seleccionados por cada uno de los métodos bayesianos en el estudio con datos simulados para dos variables correlacionadas y una variable candidata con multicolinealidad.

De la misma manera que en el escenario anterior, se ejecutó los métodos Lasso y Stepwise en el software R. De los resultados del Cuadro 5.12, los submodelos seleccionados por los métodos Lasso y Stepwise son  $(x_1, x_2, x_3, x_4, x_5, x_7)$  y  $(x_1, x_2, x_3, x_4, x_5, x_9, x_{19}, x_{20})$ , respectivamente.



Coef. Regresión	Lasso	Stepwise
$\beta_0$	0.0000	0.0000
$\beta_1$	<b>1.6978</b>	<b>2.3967</b>
$\beta_2$	<b>1.0619</b>	<b>2.0761</b>
$\beta_3$	<b>1.0891</b>	<b>2.0118</b>
$\beta_4$	<b>1.9088</b>	<b>2.2109</b>
$\beta_5$	<b>1.5320</b>	<b>2.1003</b>
$\beta_6$	0.0000	0.0000
$\beta_7$	<b>0.3716</b>	0.0000
$\beta_8$	0.0000	0.0000
$\beta_9$	0.0000	<b>0.3154</b>
$\beta_{10}$	0.0000	0.0000
$\beta_{11}$	0.0000	0.0000
$\beta_{12}$	0.0000	0.0000
$\beta_{13}$	0.0000	0.0000
$\beta_{14}$	0.0000	0.0000
$\beta_{15}$	0.0000	0.0000
$\beta_{16}$	0.0000	0.0000
$\beta_{17}$	0.0000	0.0000
$\beta_{18}$	0.0000	0.0000
$\beta_{19}$	0.0000	<b>0.3187</b>
$\beta_{20}$	0.0000	<b>0.3574</b>

Cuadro 5.12: Coeficientes estimados del submodelo seleccionado por los métodos Lasso y StepAIC en el estudio con datos simulados para dos variables correlacionadas y una variable candidata con multicolinealidad.

Para este escenario, los tres métodos bayesianos seleccionan al submodelo correcto  $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5)$  anteriormente definido para la creación de la variable respuesta  $\mathbf{y}$ . El método SSVS no se ve afectado por la correlación de las variables  $(\mathbf{x}_1, \mathbf{x}_6)$ , ni por la multicolinealidad de la variable  $\mathbf{x}_7$ , a diferencia de los métodos KM y GVS donde se ve un ligero efecto en su desempeño ya que seleccionan un gran número de submodelos posibles y la probabilidad de selección del submodelo correcto es mejor a 0.5. Por su parte, los métodos tradicionales Lasso y Stepwise se alejan considerablemente del submodelo correcto.

En las Figuras 5.7, 5.8 y 5.9 se presenta la distribución a posteriori de los coeficientes de regresión, donde observamos que la distribución de los coeficientes de las variables explicativas excluidas del modelo se encuentran centradas en 0.

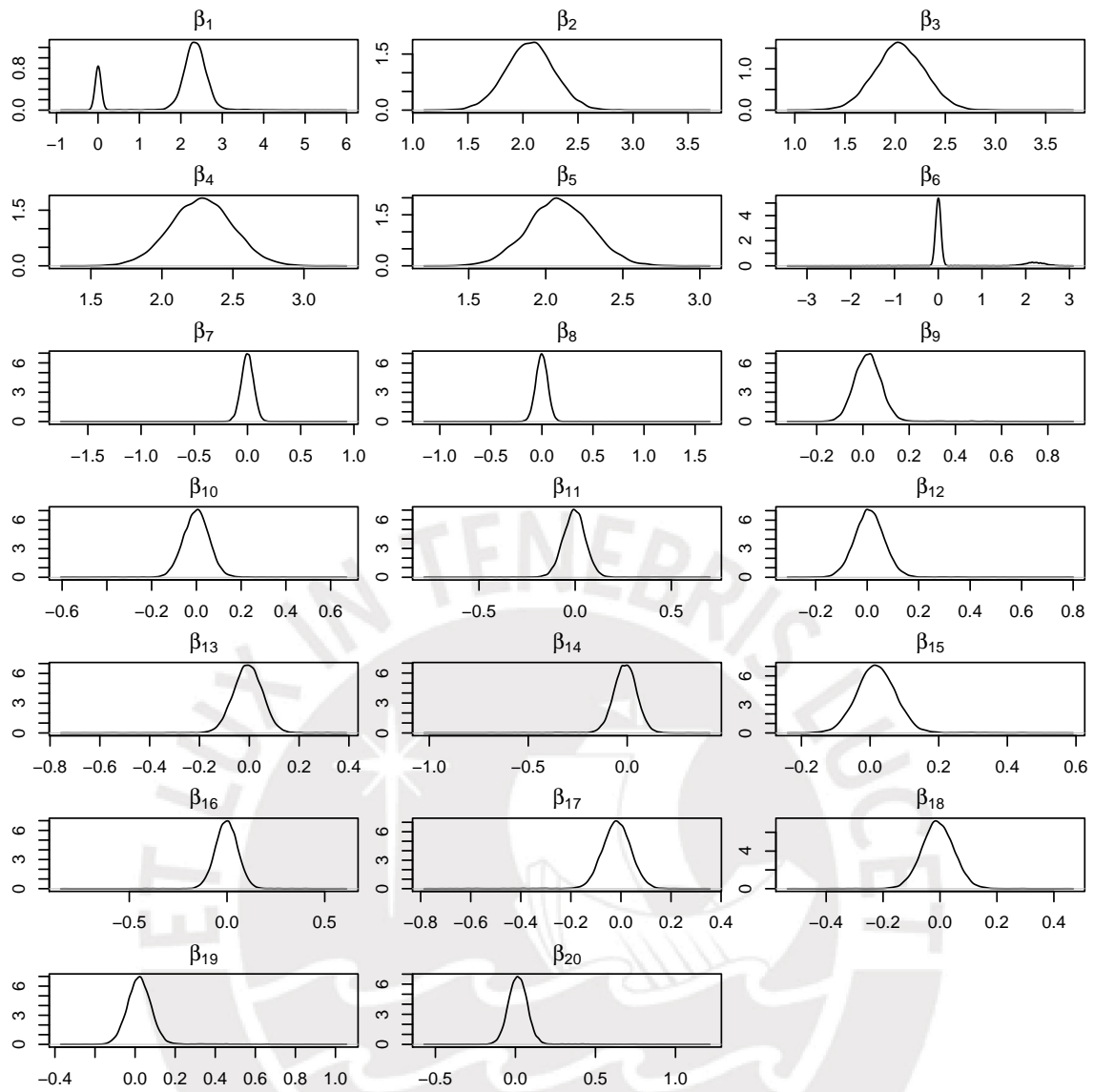


Figura 5.7: Distribución a posteriori de  $\beta$  en el estudio con datos simulados para dos variables correlacionadas y una variable candidata con multicolinealidad empleando el método SVSS y considerando  $(\phi_j, C_j) = (0.1, 500)$ .

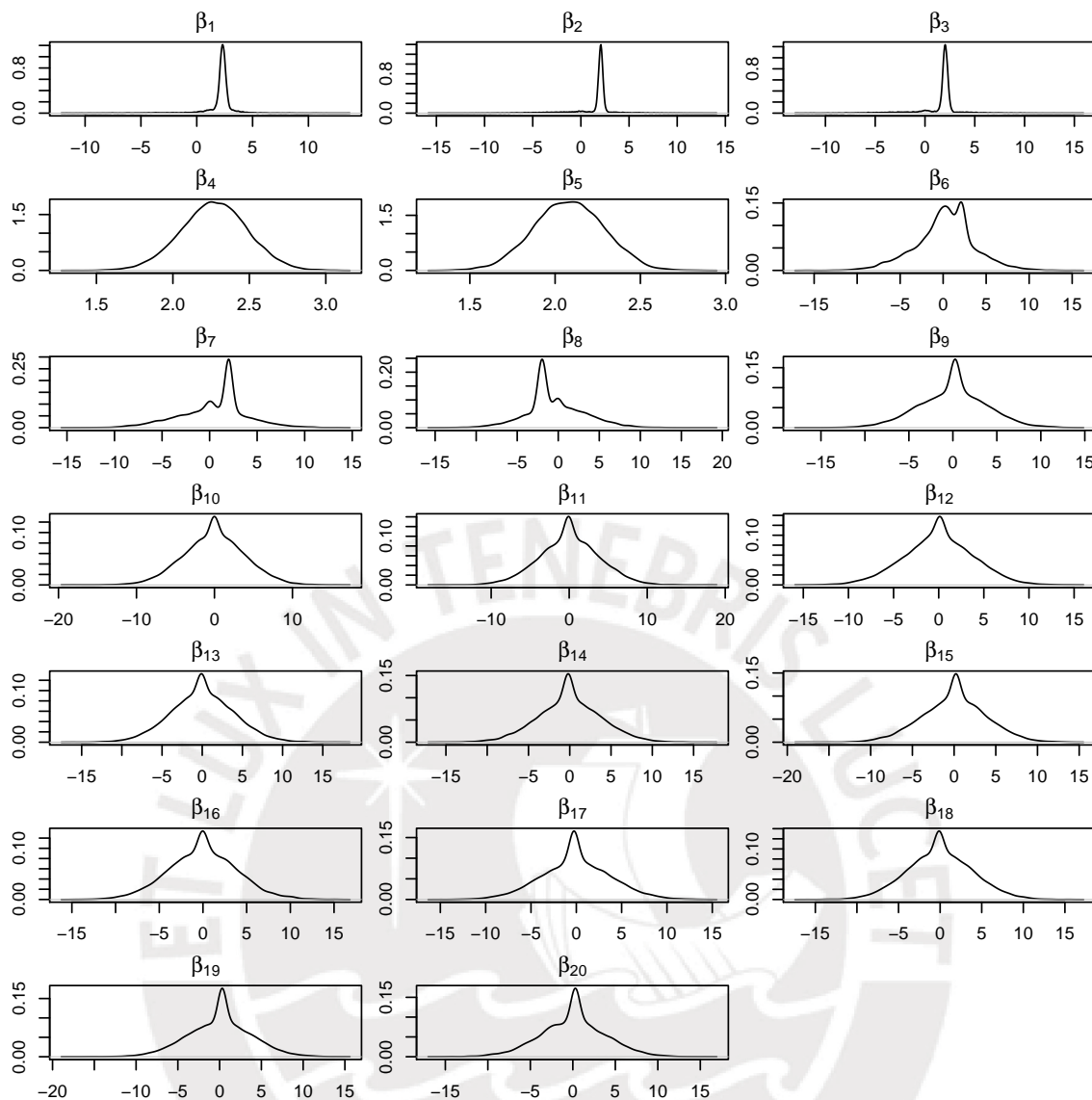


Figura 5.8: Distribución a posteriori de  $\beta$  en el estudio con datos simulados para dos variables correlacionadas y una variable candidata con multicolinealidad empleando el método KM y considerando  $D_0 = 16I$ .

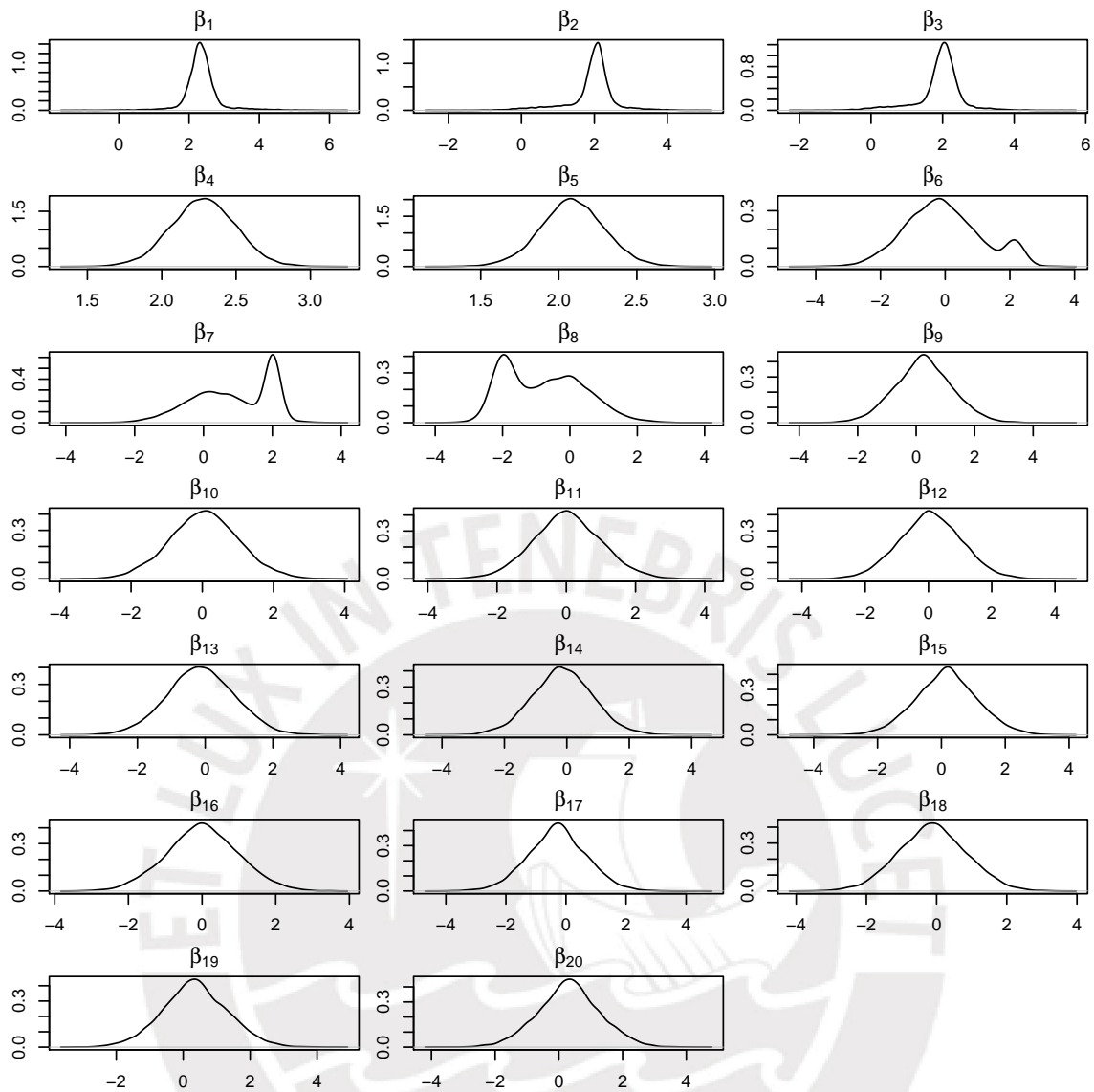


Figura 5.9: Distribución a posteriori de  $\beta$  en el estudio con datos simulados para dos variables correlacionadas y una variable candidata con multicolinealidad empleando el método GVS y considerando  $\Sigma_j = 100$ .

#### 5.4. Escenario 4: Dos variables correlacionadas y una variable relevante con multicolinealidad

En este escenario de elaboración propia, se tomó como referencia los escenarios anteriores y se realizó algunas variantes. Se generan veinte variables aleatorias de un tamaño de muestra igual a  $n = 100$ . De los cuales  $(\mathbf{x}_6, \mathbf{x}_1)$  tienen una alta correlación lineal 0.9791 y la variable  $\mathbf{x}_2$  presenta multicolinealidad con  $(\mathbf{x}_3, \mathbf{x}_7, \mathbf{x}_8)$ . Además, la variable respuesta  $\mathbf{y}$  es generada intencionalmente por  $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5)$ . A continuación, se detalla la generación de las variables:

$$\mathbf{x}_j \sim N(0, 1), \forall j = 1, 2, \dots, 20 - \{6, 7\}.$$

$$x_{i6} = x_{i1} + 0.21W_i, \forall i = 1, \dots, 100 ; \text{ donde } W_i \sim N(0, 1) .$$

$$x_{i2} = x_{i3} + x_{i7} + x_{i8} + 0.18Z_i, \forall i = 1, \dots, 100 ; \text{ donde } Z_i \sim N(0, 1) .$$

La variable respuesta  $\mathbf{y}$  fue generada de acuerdo al siguiente modelo

$$y_i = 2 x_{i1} + 2 x_{i2} + 2 x_{i3} + 2 x_{i4} + 2 x_{i5} + \epsilon_i, \forall i = 1, \dots, 100 ; \text{ donde } \epsilon_i \sim N(0, 4) .$$

En el Cuadro 5.13, se muestran los parámetros estimados por el método de mínimos cuadrados ordinarios para el modelo de regresión con las veinte variables explicativas.

Coef. Regresión	Coef. Estimados	Desv. Estándar	p-valor
$\beta_1$	2.7433	1.3703	0.0487
$\beta_2$	1.7265	1.4022	0.2218
$\beta_3$	1.6831	1.3403	0.2129
$\beta_4$	2.2965	0.2388	0.0001
$\beta_5$	2.0026	0.2283	0.0001
$\beta_6$	-0.2914	1.3133	0.8250
$\beta_7$	0.3356	1.3386	0.8027
$\beta_8$	-0.2855	1.3359	0.8313
$\beta_9$	0.2562	0.2258	0.2600
$\beta_{10}$	0.0606	0.2287	0.7920
$\beta_{11}$	0.0177	0.2654	0.9471
$\beta_{12}$	0.0514	0.2215	0.8172
$\beta_{13}$	-0.1355	0.2587	0.6020
$\beta_{14}$	-0.1416	0.2747	0.6078
$\beta_{15}$	0.1583	0.2079	0.4487
$\beta_{16}$	0.0278	0.2705	0.9186
$\beta_{17}$	-0.2852	0.2387	0.2356
$\beta_{18}$	-0.1303	0.2340	0.5791
$\beta_{19}$	0.3679	0.2463	0.1391
$\beta_{20}$	0.3110	0.2713	0.2551

Cuadro 5.13: Parámetros estimados por mínimos cuadrados ordinarios en el estudio con datos simulados para dos variables correlacionadas y una variable relevante con multicolinealidad.

Adicionalmente a ello, el error estándar residual obtenido es  $\hat{\sigma} = 2.211$ . Se observa del Cuadro 5.13 que los valores  $p$  de las variables ( $\mathbf{x}_1$ ,  $\mathbf{x}_4$  y  $\mathbf{x}_5$ ) son menores a 0.05, lo que indica que son significativas.

#### 5.4.1. Resumen de resultados

De acuerdo a los resultados de los escenarios anteriores, se optó por seguir trabajando con los hiperparámetros que conllevan a un mejor desempeño de los métodos bayesianos.

En el Cuadro 5.14, se muestra un resumen con los mejores resultados observados en cada uno de los métodos implementados. Este contiene los 10 submodelos con mayor frecuencia de aparición en la muestra de Gibbs, ordenados de manera descendente por su probabilidad de selección.

Método	SSVS	KM	GVS
Parámetros	$(\phi_j, C_j) = (0.1, 500)$	$D_0 = 16I$	$\Sigma_j = 100$
N° Submodelos	121	1303	484
$x_1, x_2, x_3, x_4, x_5$	<b>0.7465</b>	<b>0.2284</b>	<b>0.5321</b>
$x_2, x_3, x_4, x_5, x_6$	0.0652	0.0365	0.0472
$x_1, x_2, x_3, x_4, x_5, x_6$	0.0287	0.0511	0.0495
$x_1, x_2, x_3, x_4, x_5, x_9$	0.0132	0.0233	0.0235
$x_1, x_2, x_3, x_4, x_5, x_{20}$	0.0121	0.0292	0.0282
$x_1, x_2, x_3, x_4, x_5, x_{19}$	0.0112	0.0286	0.0264
$x_1, x_2, x_4, x_5, x_7, x_8$	0.0100	0.0075	0.0028
$x_1, x_2, x_3, x_4, x_5, x_{14}$	0.0097	0.0169	0.0178
$x_1, x_2, x_3, x_4, x_5, x_{17}$	0.0091	0.0292	0.0245
$x_1, x_2, x_3, x_4, x_5, x_{15}$	0.0090	0.0163	0.0165

Cuadro 5.14: Mejores submodelos seleccionados por cada uno de los métodos bayesianos en el estudio con datos simulados para dos variables correlacionadas y una variable relevante con multicolinealidad.

De la misma manera que en el escenario anterior, se ejecutó los métodos Lasso y Stepwise en el software R. Del Cuadro 5.15, observamos que los submodelos seleccionados por los métodos Lasso y Stepwise son  $(x_1, x_2, x_3, x_4, x_5)$  y  $(x_1, x_2, x_3, x_4, x_5, x_9, x_{19}, x_{20})$ , respectivamente.

Coef. Regresión	Lasso	Stepwise
$\beta_0$	0.0000	0.0000
$\beta_1$	<b>1.7886</b>	<b>2.4033</b>
$\beta_2$	<b>1.873</b>	<b>1.9693</b>
$\beta_3$	<b>1.6316</b>	<b>2.0470</b>
$\beta_4$	<b>1.8307</b>	<b>2.2065</b>
$\beta_5$	<b>1.5900</b>	<b>2.1017</b>
$\beta_6$	0.0000	0.0000
$\beta_7$	0.0000	0.0000
$\beta_8$	0.0000	0.0000
$\beta_9$	0.0000	<b>0.3163</b>
$\beta_{10}$	0.0000	0.0000
$\beta_{11}$	0.0000	0.0000
$\beta_{12}$	0.0000	0.0000
$\beta_{13}$	0.0000	0.0000
$\beta_{14}$	0.0000	0.0000
$\beta_{15}$	0.0000	0.0000
$\beta_{16}$	0.0000	0.0000
$\beta_{17}$	0.0000	0.0000
$\beta_{18}$	0.0000	0.0000
$\beta_{19}$	0.0000	<b>0.3029</b>
$\beta_{20}$	0.0000	<b>0.3640</b>

Cuadro 5.15: Coeficientes estimados del submodelo seleccionado por los métodos Lasso y StepAIC en el estudio con datos simulados para dos variables correlacionadas y una variable relevante con multicolinealidad.

De los resultados, se observa que los tres métodos bayesianos dan la mayor probabilidad de selección al submodelo  $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5)$ , el cual fue utilizado para la creación de la variable respuesta  $\mathbf{y}$ . Además, los métodos bayesianos SSVS y GVS no se ven afectados por la multicolinealidad de la variable relevante  $\mathbf{x}_2$  ni por la alta correlación de  $(\mathbf{x}_1, \mathbf{x}_6)$ ; sin embargo, el método KM se ve ligeramente afectado en su desempeño, dado que selecciona un gran número de submodelos 1303 y elige al submodelo correcto con una probabilidad menor a 0.25. Al igual que en los escenarios anteriores, los métodos bayesianos nos brindan mejores resultados a diferencia de los métodos Lasso y Stepwise.

En las Figuras 5.10, 5.11 y 5.12 se presenta la distribución a posteriori de los coeficientes de regresión, donde observamos que la distribución de los coeficientes de las variables explicativas excluidas del modelo se encuentran centradas en 0.



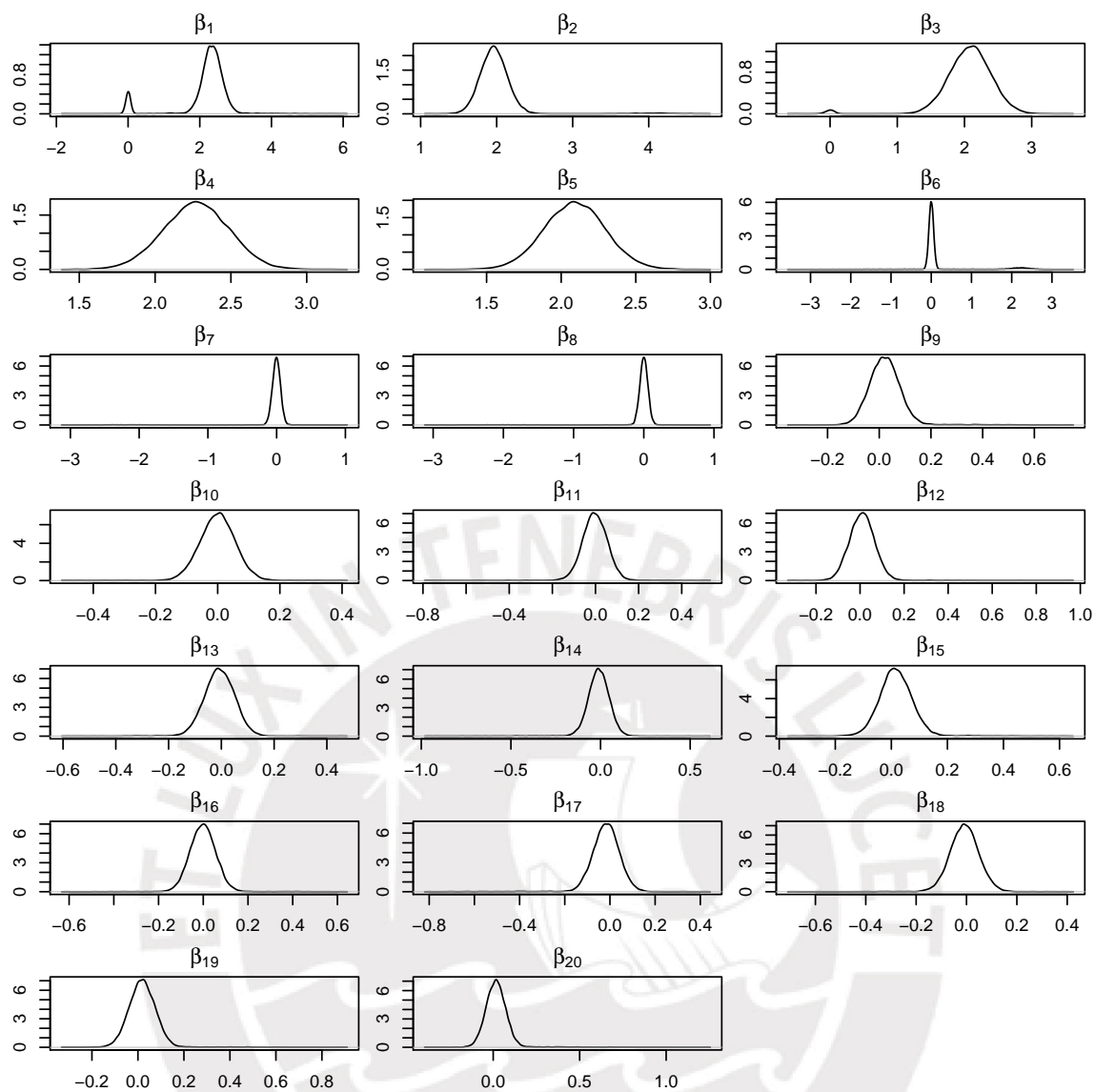


Figura 5.10: Distribución a posteriori de  $\beta$  en el estudio con datos simulados para dos variables correlacionadas y una variable relevante con multicolinealidad empleando el método SVSS y considerando  $(\phi_j, C_j) = (0.1, 500)$ .

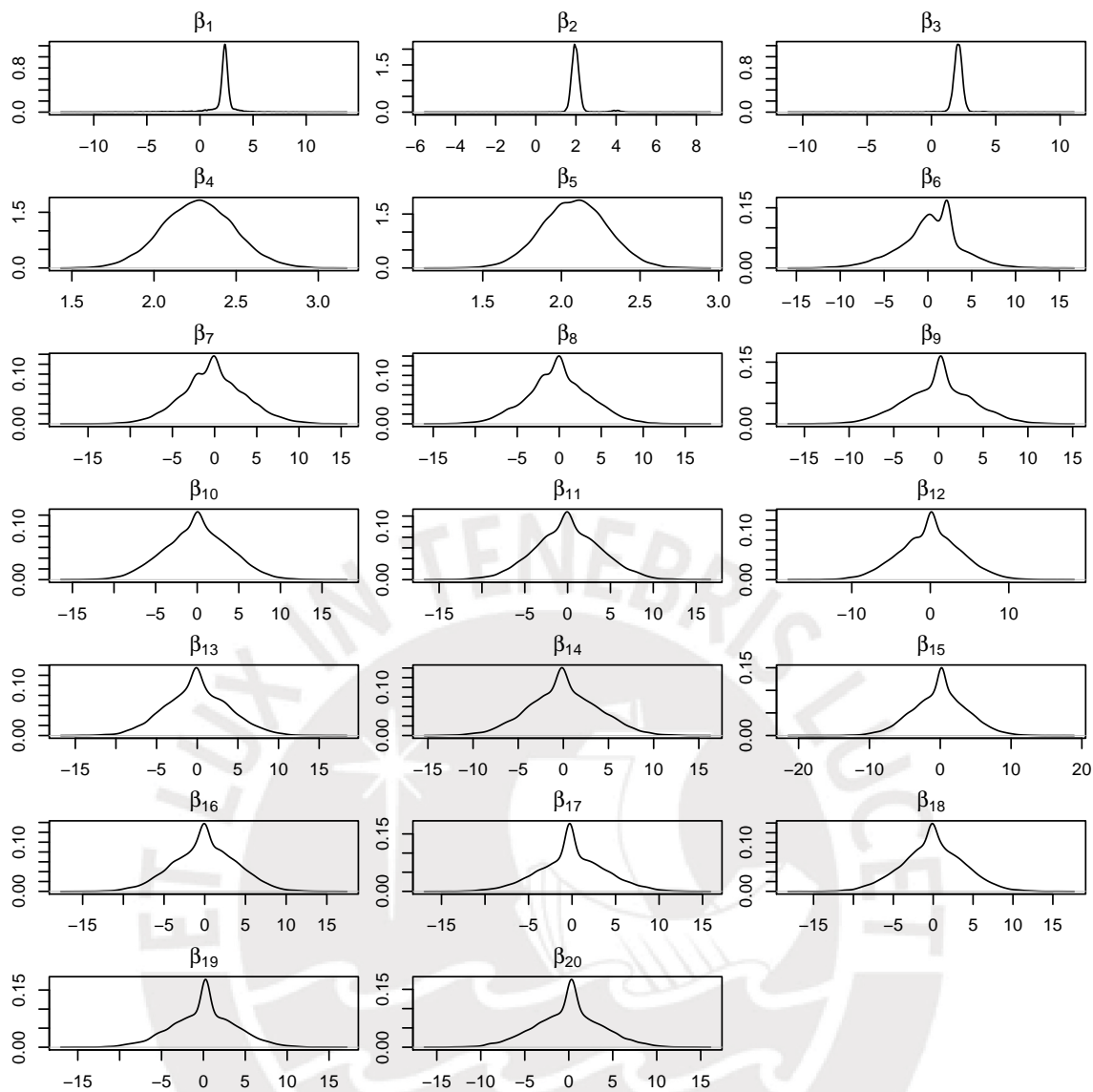


Figura 5.11: Distribución a posteriori de  $\beta$  en el estudio con datos simulados para dos variables correlacionadas y una variable relevante con multicolinealidad empleando el método KM y considerando  $D_0 = 16I$ .

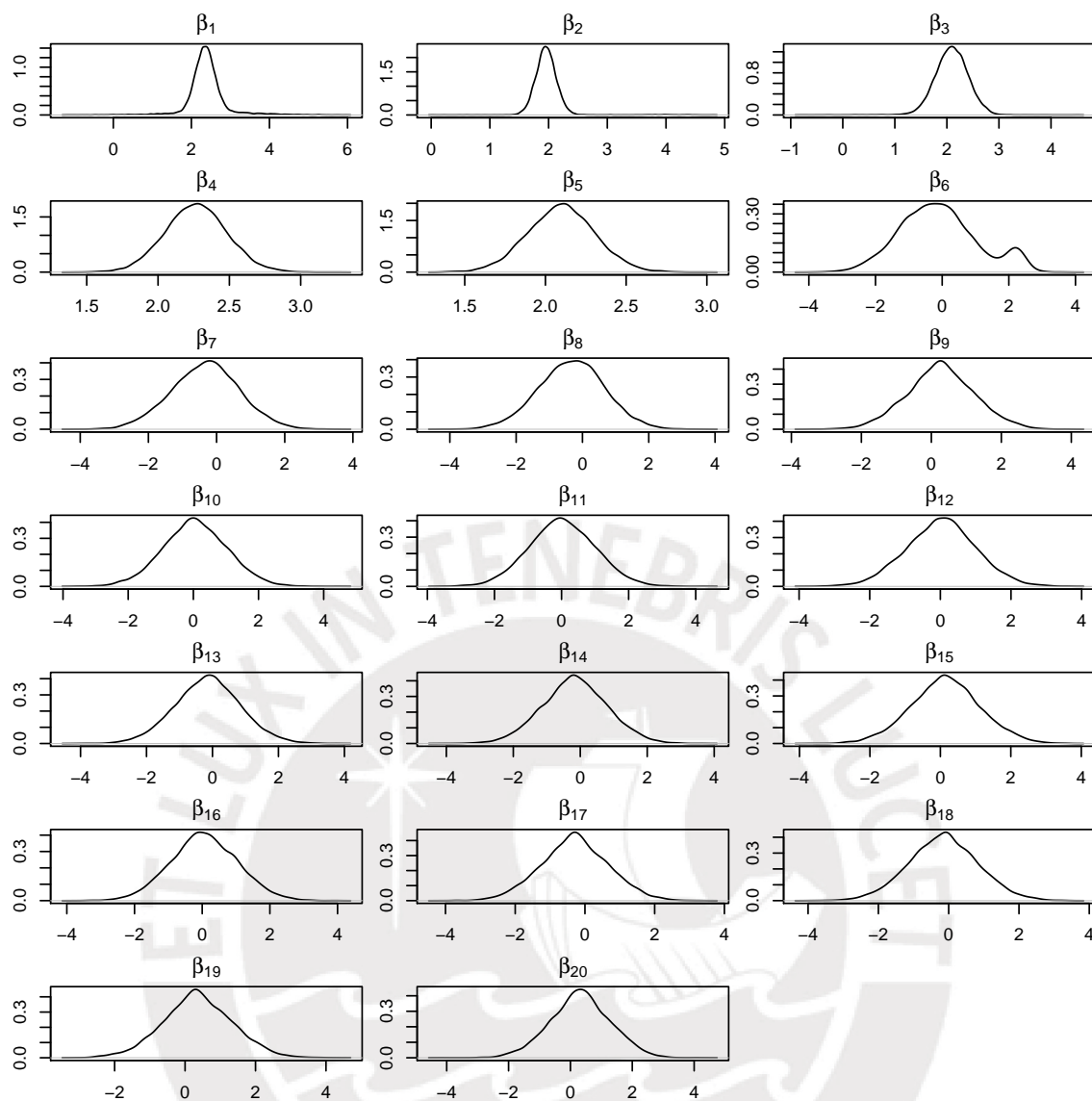


Figura 5.12: Distribución a posteriori de  $\beta$  en el estudio con datos simulados para dos variables correlacionadas y una variable relevante con multicolinealidad empleando el método GVS y considerando  $\Sigma_j = 100$ .

## Capítulo 6

### Aplicación con datos reales

En este capítulo se ilustrará el comportamiento de los tres métodos de selección de variables anteriormente presentados, empleando los datos de la Esperanza de vida al nacer de 189 países obtenidos de la United Nations Development Programme (UNDP, 2017), de los que se excluyeron 29 países debido que había un porcentaje alto de datos faltantes. Se analizará también de manera conjunta los métodos Lasso y Stepwise. El objetivo de la aplicación es conocer el conjunto de variables explicativas seleccionadas por los métodos bayesianos y los tradicionales.

La implementación de los métodos bayesianos se realizó a través del algoritmo de Gibbs. Se ejecutaron 15000 iteraciones, de las anteriores se descartaron las primeras 2000 para evitar el efecto de los valores iniciales. Cabe mencionar que las variables fueron estandarizadas antes de ingresar a los métodos de selección de variables.

#### 6.1. Descripción de los datos

Los datos empleados en esta aplicación fueron obtenidos de la United Nations Development Programme, las cuales fueron medidas en el año 2017 y en donde se tienen registradas las siguientes 9 variables explicativas: años esperados de escolaridad ( $\mathbf{x}_1$ ), promedio de años de escolaridad ( $\mathbf{x}_2$ ), ingreso nacional bruto (INB) per cápita en miles de dólares ( $\mathbf{x}_3$ ), índice de desigualdad de género ( $\mathbf{x}_4$ ), tasa de mortalidad materna (muerte por 100,000 nacidos vivos) ( $\mathbf{x}_5$ ), tasa de natalidad de los adolescentes (nacimientos por 1,000 mujeres de 15 a 19 años) ( $\mathbf{x}_6$ ), proporción de mujeres con escaños en el parlamento ( $\mathbf{x}_7$ ), porcentaje de mujeres con al menos educación secundaria ( $\mathbf{x}_8$ ), tasa de participación de mujeres en la fuerza laboral (porcentaje de la población femenina entre 15-64 años) ( $\mathbf{x}_9$ ).

En el Cuadro 6.1 se presenta una breve descripción de las variables obtenidas para la población final de 160 países, las cuales serán incluidas en el modelo. Del análisis observamos que la esperanza de vida al nacer fluctúa entre Japón con 83.908 años en la primera posición y Sierra Leona con 52.214 años en la última posición, la media es de 72.65 años y una desviación estándar de 7.52 años, lo cual refleja una alta dispersión para la esperanza de vida entre los países. Respecto al ingreso nacional bruto per cápita, tasa de mortalidad materna y tasa de natalidad de los adolescentes, observamos que presentan altas desviaciones, lo que indica la presencia de valores extremos.

Otra variable importante por su alta correlación (coeficiente Pearson 0.81) con la esperanza de vida al nacer, es el número de años esperados de escolaridad con una media de 13.51;

siendo Australia el país con el mayor valor (22.92 años). Además, es importante destacar la alta correlación inversa de la esperanza de vida al nacer con el índice de desigualdad de género y la tasa de mortalidad materna, ambos con un coeficiente de Pearson de -0.85, esto nos indica que a menor valor de dichas variables la esperanza de vida de un país tiende a aumentar.

VARIABLES	Media	Desv. Estándar	Mínimo	Máximo
Esperanza de vida al nacer	72.65	7.52	52.21	83.90
Años esperados de escolaridad	13.51	2.86	5.36	22.92
Promedio de años de escolaridad	8.76	3.10	1.46	14.08
Ingreso Nacional Bruto per cápita	18.75	19.29	0.66	116.82
Índice de desigualdad de género	0.34	0.17	0.03	0.83
Tasa de mortalidad materna	151.11	221.01	3.00	1360.00
Tasa de natalidad de los adolescentes	44.54	38.74	1.64	191.98
% mujeres con escaños en el parlamento	22.33	11.19	0.00	55.66
% mujeres con al menos educación secundaria	61.07	29.51	1.68	100.00
Tasa participación de mujeres en la fuerza laboral	51.56	15.39	6.00	86.00

Cuadro 6.1: Descripción de la muestra de países evaluados.

	y	x <sub>1</sub>	x <sub>2</sub>	x <sub>3</sub>	x <sub>4</sub>	x <sub>5</sub>	x <sub>6</sub>	x <sub>7</sub>	x <sub>8</sub>	x <sub>9</sub>
y	1									
x <sub>1</sub>	0.81	1								
x <sub>2</sub>	0.77	0.82	1							
x <sub>3</sub>	0.66	0.64	0.6	1						
x <sub>4</sub>	-0.85	-0.85	-0.85	-0.72	1					
x <sub>5</sub>	-0.85	-0.69	-0.73	-0.49	0.74	1				
x <sub>6</sub>	-0.77	-0.7	-0.69	-0.58	0.81	0.76	1			
x <sub>7</sub>	0.23	0.29	0.2	0.11	-0.35	-0.14	-0.09	1		
x <sub>8</sub>	0.71	0.75	0.95	0.54	-0.8	-0.71	-0.67	0.13	1	
x <sub>9</sub>	-0.2	-0.02	-0.06	0.01	-0.01	0.26	0.25	0.22	-0.08	1

Cuadro 6.2: Análisis de correlación lineal.

En el cuadro 6.3, se muestran los parámetros estimados del modelo de regresión por el método de mínimos cuadrados ordinarios.

Coef. Regresión	Coef. Estimados	Desv. Estándar	p-valor
$\beta_1$	0.1948	0.0638	0.0027
$\beta_2$	0.1161	0.1229	0.3466
$\beta_3$	0.1101	0.0462	0.0185
$\beta_4$	-0.3530	0.0969	0.0004
$\beta_5$	-0.4447	0.0547	0.0000
$\beta_6$	0.0255	0.0633	0.6880
$\beta_7$	0.0054	0.0377	0.8855
$\beta_8$	-0.1940	0.1073	0.0725
$\beta_9$	-0.0966	0.0357	0.0077

Cuadro 6.3: Parámetros estimados por mínimos cuadrados ordinarios en el estudio de la Esperanza de vida al nacer.

Adicionalmente a ello, el error estándar residual obtenido es  $\hat{\sigma} = 0.3871$ . Del Cuadro 6.3, observamos que las variables ( $\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_9$ ) presentan valores  $p$  menores a 0.05, lo cual nos indica que son significativas.

## 6.2. Implementación de los métodos bayesianos

- Para la implementación del método SSVS, se considera  $p_j = 0.5$ ,  $V = 0.01$  y  $\delta = 0.01$  y la sugerencia de  $(\phi_j, C_j) = (0.1, 500)$  como valor de los hiperparámetros de la distribución a priori de  $\beta$  por brindar mejores resultados en el estudio de simulación.
- Para la implementación del método KM, se considera  $p_j = 0.5$ ,  $\beta_0 = (0, \dots, 0)^T$ ,  $V = 0.01$ ,  $\delta = 0.01$  y la sugerencia de  $D_0 = 16I$  como valor para la matriz de varianza y covarianza de la distribución a priori de  $\beta$  por brindar mejores resultados en el estudio de simulación.
- Para la implementación del método GVS, se considera  $p_j = 0.5$ ,  $V = 0.01$ ,  $\delta = 0.01$ ,  $\mu_j = 0$ ,  $S_j = \omega \Sigma_j$  con  $\omega = 0.01$  y la sugerencia de  $\Sigma_j = 100$  para la varianza de la distribución a priori de los  $\beta_j$  seleccionados por brindar mejores resultados en el estudio de simulación.

En el Cuadro 6.4, se presenta el número de submodelos seleccionados, y los submodelos con mayor frecuencia de aparición en la muestra de Gibbs por cada uno de los métodos bayesianos, ordenados de manera descendente por su probabilidad de selección.

Método	SSVS	KM	GVS
Modelos	$(\phi_j, C_j) = (0.1, 500)$	$\mathbf{D_0 = 16I}$	$\Sigma_j = 100$
N° Submodelos	57	59	41
$x_1, x_4, x_5$	<b>0.3300</b>	<b>0.3614</b>	<b>0.4647</b>
$x_1, x_3, x_4, x_5$	0.1486	0.1575	0.0671
$x_4, x_5$	0.0524	0.0945	0.2653
$x_3, x_4, x_5$	0.0887	0.0714	0.0705
$x_1, x_4, x_5, x_9$	0.1197	0.1151	0.0538
$x_1, x_4, x_5, x_8$	0.0488	0.0325	0.0171
$x_4, x_5, x_9$	0.0178	0.0125	0.0142
$x_1, x_3, x_4, x_5, x_9$	0.0476	0.0377	0.0061
$x_3, x_4, x_5, x_9$	0.0186	0.0065	0.0054
$x_1, x_3, x_5$	0.0119	0.0128	0.0048

Cuadro 6.4: Mejores submodelos seleccionados por cada uno de los métodos bayesianos en el estudio de la Esperanza de vida al nacer.

Del Cuadro 6.4, se observa que los tres métodos bayesianos dan la mayor probabilidad de selección al submodelo  $(x_1, x_4, x_5)$ . Siendo el método GVS el que asigna mayor probabilidad de selección 0.4647. En las Figuras 6.1, 6.2 y 6.3 se presentan la distribuciones a posteriori de los coeficientes de regresión, donde se observa que la distribución de los coeficientes de las variables explicativas excluidas del modelo se encuentran centradas en 0.

Del mismo modo que en el estudio de simulación, se ejecutó los métodos Lasso y Stepwise en el software R. Del Cuadro 6.5, observamos que los submodelos seleccionados por los métodos Lasso y Stepwise son  $(x_1, x_3, x_4, x_5)$  y  $(x_1, x_3, x_4, x_5, x_8, x_9)$ , respectivamente.

Coef. Regresión	Lasso	Stepwise
$\beta_0$	0.0000	0.0000
$\beta_1$	<b>0.1503</b>	<b>0.2161</b>
$\beta_2$	0.0000	<b>0.1103</b>
$\beta_3$	<b>0.0551</b>	<b>-0.3513</b>
$\beta_4$	<b>-0.2722</b>	<b>-0.3513</b>
$\beta_5$	<b>-0.4061</b>	<b>-0.4387</b>
$\beta_6$	0.0000	0.0000
$\beta_7$	0.0000	0.0000
$\beta_8$	0.0000	<b>-0.1096</b>
$\beta_9$	0.0000	<b>-0.0906</b>

Cuadro 6.5: Coeficientes estimados del submodelo seleccionado por los métodos Lasso y StepAIC en el estudio de la Esperanza de vida al nacer.

De los resultados en esta aplicación, observamos que los métodos bayesianos, Lasso y Stepwise presentan diferentes resultados, sin embargo los métodos bayesianos seleccionan un submodelo más parsimonioso donde todas sus variables son significativas. Por otro lado, el método Lasso elige al segundo submodelo bayesiano con mayor probabilidad de selección, mientras el método Stepwise selecciona a un submodelo que no aparece en la muestra de Gibbs. Además,



de acuerdo a las estimaciones por mínimos cuadrados y a un nivel de significancia del 5%, la variable que adiciona el método Stepwise ( $\mathbf{x}_8$ ) a su submodelo es no significativo. Por lo tanto, concluimos que las variables explicativas del submodelo elegido por los métodos bayesianos son las más indicadas para explicar el comportamiento de la esperanza de vida al nacer de los países, las cuales son: años esperados de escolaridad ( $\mathbf{x}_1$ ), índice de desigualdad género ( $\mathbf{x}_4$ ) y tasa de mortalidad materna ( $\mathbf{x}_5$ ).

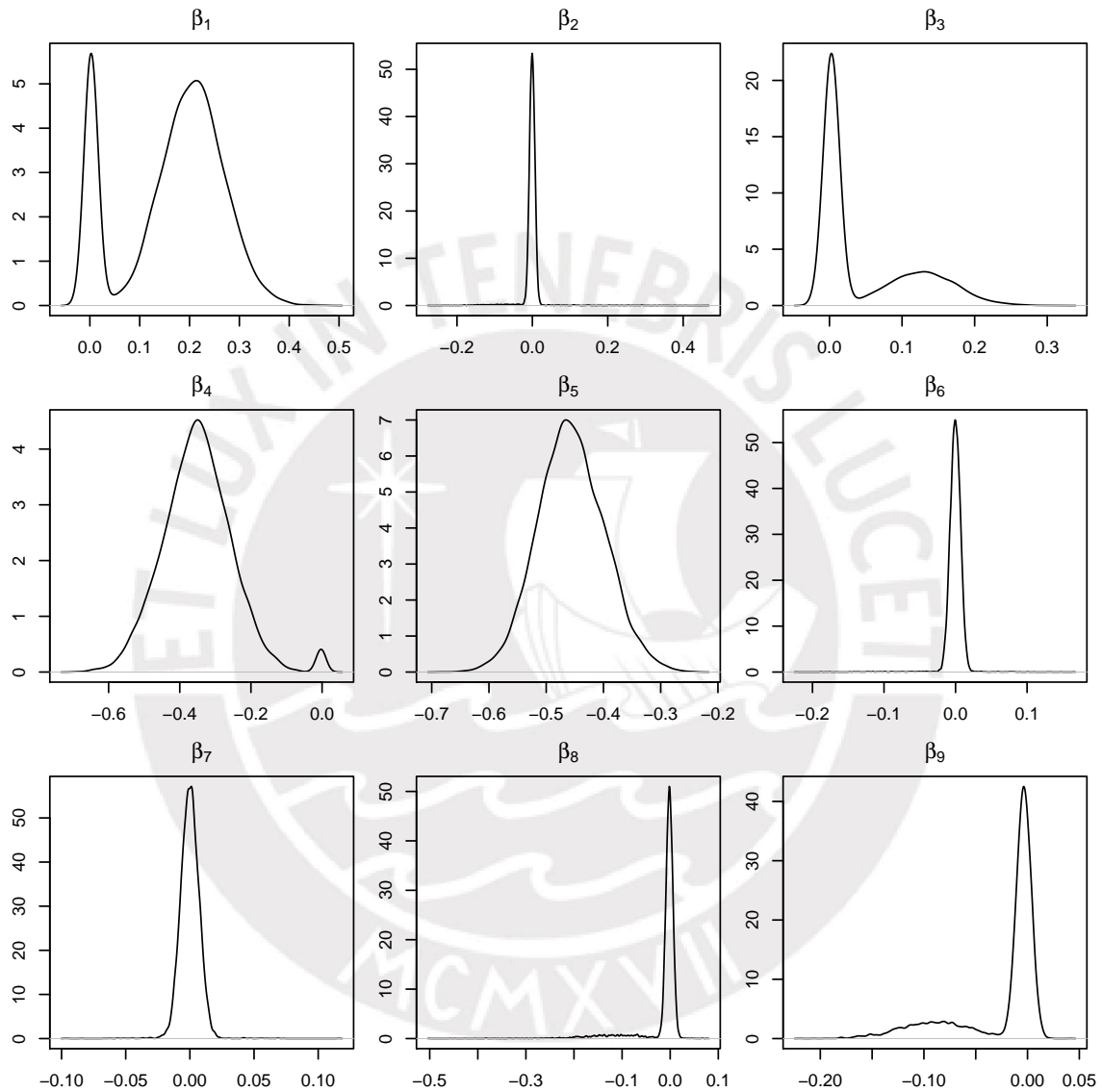


Figura 6.1: Distribución a posteriori de  $\beta$  en el estudio de la Esperanza de vida al nacer empleando el método SSVS y considerando  $(\phi_j, C_j) = (0.1, 500)$ .

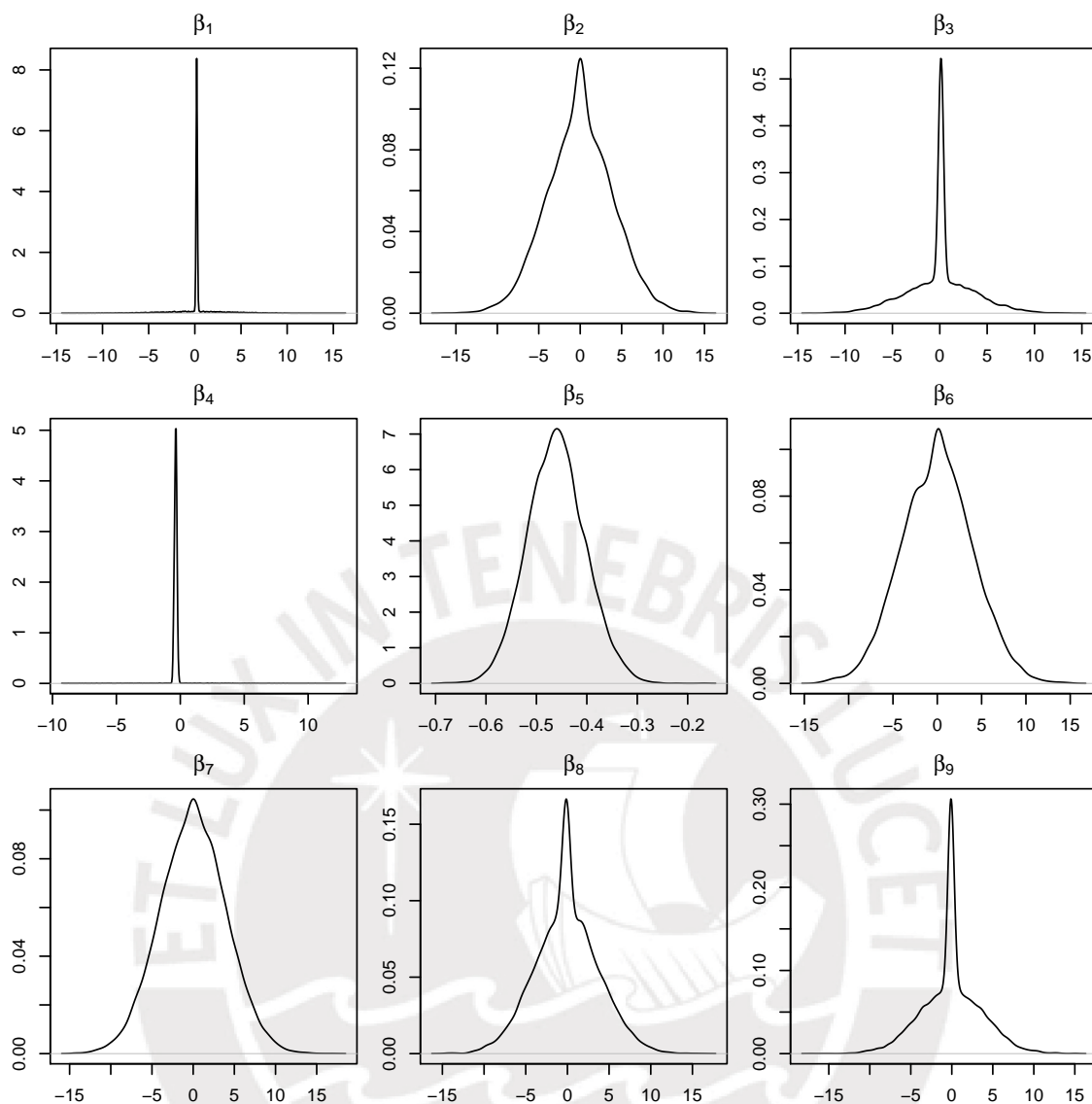


Figura 6.2: Distribución a posteriori de  $\beta$  en el estudio de la Esperanza de vida al nacer empleando el método KM y considerando  $D_0 = 16I$ .

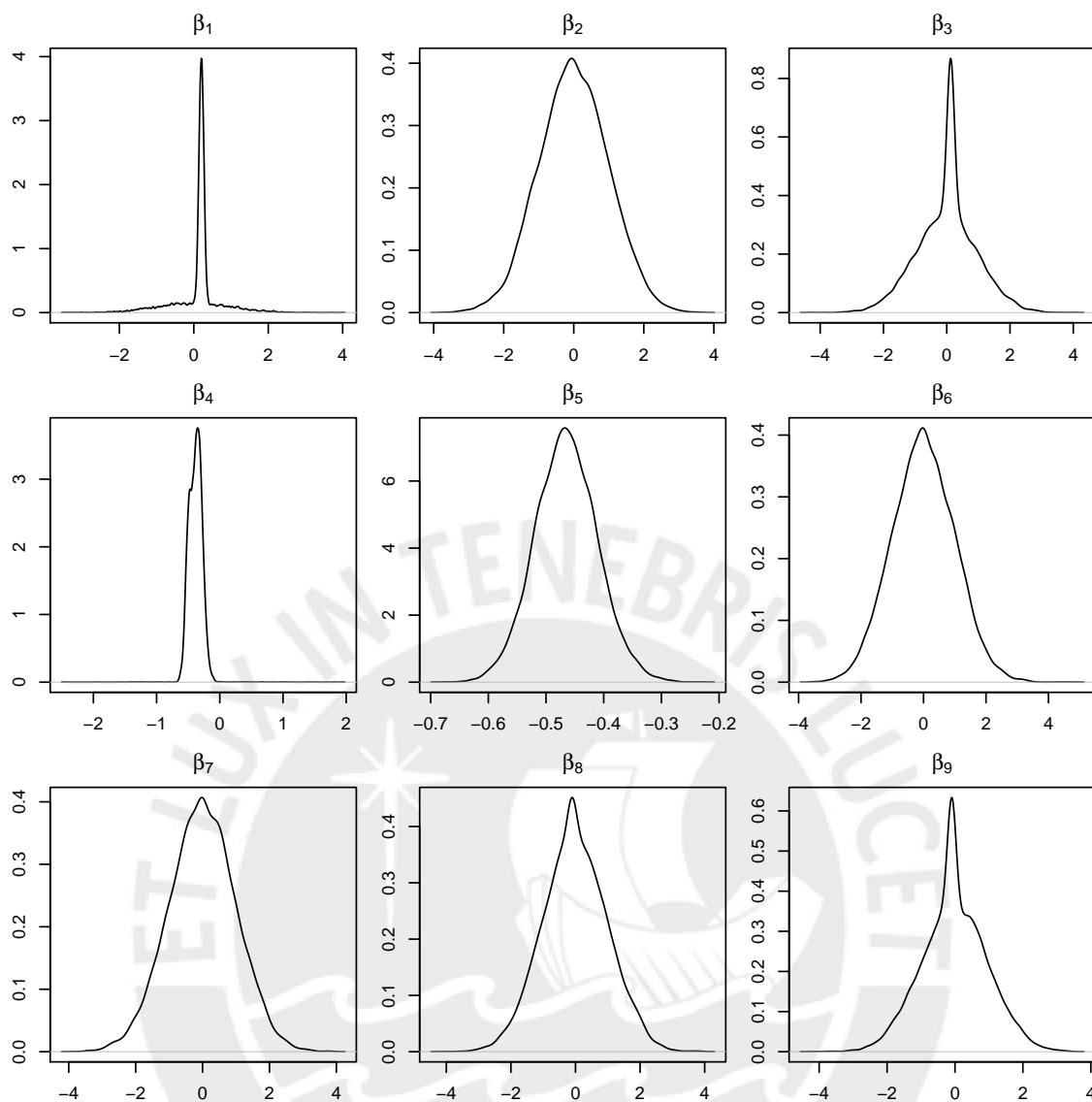


Figura 6.3: Distribución a posteriori de  $\beta$  en el estudio de la Esperanza de vida al nacer empleando el método GVS y considerando  $\Sigma_j = 100$ .

## Capítulo 7

# Conclusiones y Sugerencias

### 7.1. Conclusiones

- En este trabajo, se presenta en forma detallada tres métodos bayesianos para la selección de variables en el análisis de regresión. Estos métodos seleccionan a las variables explicativas incorporando una variable latente al modelo de regresión como indicador de selección. Asimismo, se describieron las distribuciones a priori y se obtuvieron las distribuciones condicionales completas, con las que el algoritmo de Gibbs fue implementado en el software R.
- En el estudio de simulación, se observó que los métodos bayesianos SSVS, KM y GVS tienen un mejor desempeño (medido en número de submodelos y probabilidad de selección del submodelo correcto) cuando se considera una varianza grande para la distribución a priori conjunta de los coeficientes de regresión seleccionadas. Los hiperparámetros que nos llevan a dicha conclusión son  $(\phi_j, C_j) = (0.1, 500)$ ,  $\mathbf{D}_0 = 16\mathbf{I}$  y  $\Sigma_j = 100$  para los métodos SSVS, KM y GVS, respectivamente.
- Al comparar los resultados de los métodos bayesianos, Lasso y Stepwise en el estudio de simulación, podemos mencionar que los métodos bayesianos brindan mejores resultados, ya que eligen submodelos más parsimoniosos y presentan a otros potenciales submodelos, facilitando realizar un mejor análisis y toma de decisiones. A diferencia de los métodos Lasso y Stepwise que no presentan este tipo de información.
- Basados en los resultados de la aplicación a los datos de la esperanza de vida al nacer de los países. Los métodos bayesianos, seleccionaron un submodelo más parsimonioso donde todas las variables explicativas guardan una fuerte relación con la esperanza de vida al nacer de los países; a diferencia de las variables explicativas seleccionadas por los métodos Lasso y Stepwise.

### 7.2. Sugerencias para investigaciones futuras

- Estudiar la posible aplicación de los métodos bayesianos SSVA, KM y GVS a otros tipos de modelos de regresión, por ejemplo, al modelo lineal generalizado.
- Implementar la estimación de los modelos estudiados utilizando el método Monte Carlo Hamiltoniano implementado en el programa Stan ([Carpenter et al., 2016](#)).

## Apéndice A

### Código en R del estudio de simulación

#### A.1. Método SSVS

```
#####  
### Metodo Stochastic Search Variable Selection (SSVS) ###  
##### 20 Variables Explicativas Independientes #####  
#####  
  
## Limpiando entorno de trabajo  
rm(list = ls())  
  
## Carga Librerias  
library('mnormt')  
library('invgamma')  
library('LaplacesDemon')  
  
## Obteniendo datos simulados  
set.seed('20190123')  
p=20  
n=100  
mu=rep(0,p)  
I=diag(p)  
X=rmnorm(n,mu,I)  
E=rnorm(n,0,2)  
Y= 2*X[,1] + 2*X[,2] + 2*X[,3] + 2*X[,4] + 2*X[,5] + E  
  
## Estimacion por minimos cuadrados ordinarios  
lm_ssvs1=lm(Y~X-1)  
sum_lm_ssvs1=summary(lm_ssvs1)  
  
## Estableciendo parametros  
  
M=15000  
beta=matrix(0,M,p)
```

```

beta[1,]=sum_lm_ssvs1$coefficients[,1]
gamma=matrix(0,M,p)
gamma[1,]=rep(1,p)
pexito=0.5
pexitopost=matrix(0,M,p)
pexitopost[1,]=rep(1,p)
sigma=numeric(M-1)
sigma[1]=sum_lm_ssvs1$sigma
model=numeric(M-1)
C=500
phi=0.1
t=sum(sum_lm_ssvs1$coefficients[,2])*phi/p
Vgamma=0.01
Deltgamma=0.01

## Algoritmo de Gibbs
for(h in 2:M){
  ### Generando Beta ###
  A=c((gamma[h-1,1]*C+(1-gamma[h-1,1])*1)*t, (gamma[h-1,2]*C+(1-gamma[h-1,2])*1)*t,
      (gamma[h-1,3]*C+(1-gamma[h-1,3])*1)*t, (gamma[h-1,4]*C+(1-gamma[h-1,4])*1)*t,
      (gamma[h-1,5]*C+(1-gamma[h-1,5])*1)*t, (gamma[h-1,6]*C+(1-gamma[h-1,6])*1)*t,
      (gamma[h-1,7]*C+(1-gamma[h-1,7])*1)*t, (gamma[h-1,8]*C+(1-gamma[h-1,8])*1)*t,
      (gamma[h-1,9]*C+(1-gamma[h-1,9])*1)*t, (gamma[h-1,10]*C+(1-gamma[h-1,10])*1)*t,
      (gamma[h-1,11]*C+(1-gamma[h-1,11])*1)*t, (gamma[h-1,12]*C+(1-gamma[h-1,12])*1)*t,
      (gamma[h-1,13]*C+(1-gamma[h-1,13])*1)*t, (gamma[h-1,14]*C+(1-gamma[h-1,14])*1)*t,
      (gamma[h-1,15]*C+(1-gamma[h-1,15])*1)*t, (gamma[h-1,16]*C+(1-gamma[h-1,16])*1)*t,
      (gamma[h-1,17]*C+(1-gamma[h-1,17])*1)*t, (gamma[h-1,18]*C+(1-gamma[h-1,18])*1)*t,
      (gamma[h-1,19]*C+(1-gamma[h-1,19])*1)*t, (gamma[h-1,20]*C+(1-gamma[h-1,20])*1)*t)
  D=diag(A^2,p,p)
  varbeta=solve(solve(D)+(sigma[h-1]^(-2))*(t(X)%*%X))
  mubeta=varbeta%*%((sigma[h-1]^(-2))*(t(X)%*%Y))
  beta[h,]=rmnorm(1,mubeta,varbeta)

  ### Generando Sigma ###
  SRR=t(Y)%*%Y-2*t(beta[h,])%*%t(X)%*%Y+t(beta[h,])%*%t(X)%*%X%*%beta[h,]
  sigma[h]=(rinvgamma(1, (n+Vgamma)/2, (SRR+Vgamma*Deltgamma)/2))^(1/2)

  ### Generando Gamma ###
  gamma1=gamma[h-1,]
  W=sample(1:p,p) ## Considerando la recomendacion de Ntzoufras (2002)
  ## de no generar gamma_i de manera consecutiva

```

```

for(i in W){
  gamma1[i]=1
  Agamma=c((gamma1[1]*C+(1-gamma1[1])*1)*t, (gamma1[2]*C+(1-gamma1[2])*1)*t,
            (gamma1[3]*C+(1-gamma1[3])*1)*t, (gamma1[4]*C+(1-gamma1[4])*1)*t,
            (gamma1[5]*C+(1-gamma1[5])*1)*t, (gamma1[6]*C+(1-gamma1[6])*1)*t,
            (gamma1[7]*C+(1-gamma1[7])*1)*t, (gamma1[8]*C+(1-gamma1[8])*1)*t,
            (gamma1[9]*C+(1-gamma1[9])*1)*t, (gamma1[10]*C+(1-gamma1[10])*1)*t,
            (gamma1[11]*C+(1-gamma1[11])*1)*t, (gamma1[12]*C+(1-gamma1[12])*1)*t,
            (gamma1[13]*C+(1-gamma1[13])*1)*t, (gamma1[14]*C+(1-gamma1[14])*1)*t,
            (gamma1[15]*C+(1-gamma1[15])*1)*t, (gamma1[16]*C+(1-gamma1[16])*1)*t,
            (gamma1[17]*C+(1-gamma1[17])*1)*t, (gamma1[18]*C+(1-gamma1[18])*1)*t,
            (gamma1[19]*C+(1-gamma1[19])*1)*t, (gamma1[20]*C+(1-gamma1[20])*1)*t)
  Dgamma=diag(Agamma^2,p,p)
  pBgamma1=dmnorm(beta[h,],mu,Dgamma)*pexito
  numeradorgamma=pBgamma1

  gamma1[i]=0
  Agamma=c((gamma1[1]*C+(1-gamma1[1])*1)*t, (gamma1[2]*C+(1-gamma1[2])*1)*t,
            (gamma1[3]*C+(1-gamma1[3])*1)*t, (gamma1[4]*C+(1-gamma1[4])*1)*t,
            (gamma1[5]*C+(1-gamma1[5])*1)*t, (gamma1[6]*C+(1-gamma1[6])*1)*t,
            (gamma1[7]*C+(1-gamma1[7])*1)*t, (gamma1[8]*C+(1-gamma1[8])*1)*t,
            (gamma1[9]*C+(1-gamma1[9])*1)*t, (gamma1[10]*C+(1-gamma1[10])*1)*t,
            (gamma1[11]*C+(1-gamma1[11])*1)*t, (gamma1[12]*C+(1-gamma1[12])*1)*t,
            (gamma1[13]*C+(1-gamma1[13])*1)*t, (gamma1[14]*C+(1-gamma1[14])*1)*t,
            (gamma1[15]*C+(1-gamma1[15])*1)*t, (gamma1[16]*C+(1-gamma1[16])*1)*t,
            (gamma1[17]*C+(1-gamma1[17])*1)*t, (gamma1[18]*C+(1-gamma1[18])*1)*t,
            (gamma1[19]*C+(1-gamma1[19])*1)*t, (gamma1[20]*C+(1-gamma1[20])*1)*t)
  Dgamma=diag(Agamma^2,p,p)
  pBgamma1=dmnorm(beta[h,],mu,Dgamma)*(1-pexito)
  denominadorgamma=pBgamma1
  pexitopost[h,i]=numeradorgamma/(denominadorgamma+numeradorgamma)
  gamma1[i]=rbinom(1,1,pexitopost[h,i])
}
gamma[h,]=gamma1
}

## Submodelos Seleccionados
m=2000
for(j in m:M){
  model[j]=1+1*gamma[j,1]+2*gamma[j,2]+4*gamma[j,3]+8*gamma[j,4]+16*gamma[j,5]+
            32*gamma[j,6]+64*gamma[j,7]+(2^7)*gamma[j,8]+(2^8)*gamma[j,9]+
            (2^9)*gamma[j,10] +(2^10)*gamma[j,11]+(2^11)*gamma[j,12]+

```



```

(2^12)*gamma[j,13]+(2^13)*gamma[j,14]+(2^14)*gamma[j,15]+
(2^15)*gamma[j,16]+(2^16)*gamma[j,17]+(2^17)*gamma[j,18]+
(2^18)*gamma[j,19]+(2^19)*gamma[j,20]}

## Probabilidad a posteriori de los submodelos
model_ssvs=data.frame(table(model))
model_ssvs=subset(model_ssvs,model_ssvs$model!=0)
model_ssvs$Porce=model_ssvs$Freq/(M-m+1)
model_ssvs$Porce=round(model_ssvs$Porce,4)
model_ssvs=model_ssvs[order(-model_ssvs$Porce),]
head(model_ssvs)

## Grafico de densidad de Beta
pdf("density_beta_ssvs_1.pdf")
par(mfrow=c(7,3),mai=c(0.28,0.25,0.25,0.1))
plot(density(beta[2001:M,1]),main =expression(beta[1]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,2]),main =expression(beta[2]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,3]),main =expression(beta[3]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,4]),main =expression(beta[4]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,5]),main =expression(beta[5]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,6]),main =expression(beta[6]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,7]),main =expression(beta[7]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,8]),main =expression(beta[8]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,9]),main =expression(beta[9]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,10]),main =expression(beta[10]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,11]),main =expression(beta[11]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,12]),main =expression(beta[12]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,13]),main =expression(beta[13]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,14]),main =expression(beta[14]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,15]),main =expression(beta[15]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,16]),main =expression(beta[16]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,17]),main =expression(beta[17]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,18]),main =expression(beta[18]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,19]),main =expression(beta[19]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,20]),main =expression(beta[20]),xlab=NA,ylab=NA)
dev.off()

## Grafico de cadena de simulacion de Beta
pdf("mcmc_beta_ssvs_1.pdf")
par(mfrow=c(7,3),mai=c(0.28,0.25,0.25,0.1))
ts.plot(beta[2001:M,1],main=expression(beta[1]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,2],main=expression(beta[2]),xlab=NA,ylab=NA)

```

```
ts.plot(beta[2001:M,3],main=expression(beta[3]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,4],main=expression(beta[4]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,5],main=expression(beta[5]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,6],main=expression(beta[6]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,7],main=expression(beta[7]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,8],main=expression(beta[8]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,9],main=expression(beta[9]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,10],main=expression(beta[10]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,11],main=expression(beta[11]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,12],main=expression(beta[12]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,13],main=expression(beta[13]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,14],main=expression(beta[14]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,15],main=expression(beta[15]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,16],main=expression(beta[16]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,17],main=expression(beta[17]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,18],main=expression(beta[18]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,19],main=expression(beta[19]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,20],main=expression(beta[20]),xlab=NA,ylab=NA)
dev.off()

## Grafico de densidad de Gamma
pdf("density_gamma_ssvs_1.pdf")
par(mfrow=c(7,3),mai=c(0.28,0.25,0.25,0.1))
plot(density(gamma[2001:M,1]),main =expression(gamma[1]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,2]),main =expression(gamma[2]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,3]),main =expression(gamma[3]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,4]),main =expression(gamma[4]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,5]),main =expression(gamma[5]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,6]),main =expression(gamma[6]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,7]),main =expression(gamma[7]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,8]),main =expression(gamma[8]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,9]),main =expression(gamma[9]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,10]),main =expression(gamma[10]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,11]),main =expression(gamma[11]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,12]),main =expression(gamma[12]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,13]),main =expression(gamma[13]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,14]),main =expression(gamma[14]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,15]),main =expression(gamma[15]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,16]),main =expression(gamma[16]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,17]),main =expression(gamma[17]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,18]),main =expression(gamma[18]),xlab=NA,ylab=NA)
```

```
plot(density(gamma[2001:M,19]),main =expression(gamma[19]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,20]),main =expression(gamma[20]),xlab=NA,ylab=NA)
dev.off()
```

```
## Grafico de cadena de simulacion de Gamma
pdf("mcmc_gamma_ssvs_1.pdf")
par(mfrow=c(7,3),mai=c(0.28,0.25,0.25,0.1))
ts.plot(gamma[2001:M,1],main=expression(gamma[1]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,2],main=expression(gamma[2]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,3],main=expression(gamma[3]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,4],main=expression(gamma[4]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,5],main=expression(gamma[5]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,6],main=expression(gamma[6]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,7],main=expression(gamma[7]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,8],main=expression(gamma[8]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,9],main=expression(gamma[9]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,10],main=expression(gamma[10]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,11],main=expression(gamma[11]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,12],main=expression(gamma[12]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,13],main=expression(gamma[13]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,14],main=expression(gamma[14]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,15],main=expression(gamma[15]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,16],main=expression(gamma[16]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,17],main=expression(gamma[17]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,18],main=expression(gamma[18]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,19],main=expression(gamma[19]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,20],main=expression(gamma[20]),xlab=NA,ylab=NA)
dev.off()
```

```
## Grafico de densidad y cadena de simulacion de Sigma
pdf("sigma_ssvs_1.pdf")
par(mfrow=c(2,1))
plot(density(sigma),main =expression(sigma),xlab=NA,ylab=NA)
ts.plot(sigma[2001:M],main=expression(sigma),xlab=NA,ylab=NA)
dev.off()
```

## A.2. Método KM

```
#####
### Metodo unconditional priors for variable selection (KM) ###
##### 20 Variables Explicativas Independientes #####
#####

## Limpiando entorno de trabajo
rm(list = ls())

## Carga Librerias
library('mnormt')
library('invgamma')
library('LaplacesDemon')

## Obteniendo datos simulados
set.seed('20190123')
p=20
n=100
mu=rep(0,p)
I=diag(p)
X=rmnorm(n,mu,I)
E=rnorm(n,0,2)
Y= 2*X[,1] + 2*X[,2] + 2*X[,3] + 2*X[,4] + 2*X[,5] + E

## Estimacion por minimos cuadrados ordinarios
lm_km1=lm(Y~X-1)
sum_lm_km1=summary(lm_km1)

## Estableciendo parametros
M=15000
beta=matrix(0,M,p)
beta0=rep(0,p)
beta[1,]=sum_lm_km1$coefficients[,1]
gamma=matrix(0,M,p)
gamma[1,]=rep(1,p)
pexito=0.5
pexitopost=matrix(0,M,p)
pexitopost[1,]=rep(1,p)
sigma=numeric(M)
sigma[1]=sum_lm_km1$sigma
model=numeric(M-1)
Vgamma=0.01
```

```

Deltgamma=0.01
k=4

## Algoritmo de Gibbs
for(h in 2:M){
  ### Generando Beta ###
  X2=X%*%diag(gamma[h-1,])
  D0=k^2*I
  varbeta=solve(solve(D0)+(sigma[h-1]^(-2))*t(X2)%*%X2)
  mubeta=varbeta%*%(solve(D0)%*%beta0+(sigma[h-1]^(-2))*t(X2)%*%Y)
  beta[h,]=rmnorm(1,mubeta,varbeta)

  ### Generando Sigma ###
  A=beta[h,]*gamma[h-1,]
  theta=t(t(A))
  sigma[h]=(rinvgamma(1, (n+Vgamma)/2,
    (Deltgamma + t(Y-X%*%theta)%*%(Y-X%*%theta))/2))^^(1/2)

  ### Generando Gamma ###
  gamma1=gamma[h-1,]
  W=sample(1:p,p) ## Considerando la recomendacion de Ntzoufras (2002)
  ## de no generar gamma_i de manera consecutiva

  for(i in W){
    gamma1[i]=1
    mu2=c(t(X%*%t(t(beta[h,]*gamma1))))
    var2=(sigma[h]^2)*diag(n)
    pBgamma1=dmnorm(Y,mu2,var2)*pexito
    numeradorgamma=pBgamma1

    gamma1[i]=0
    mu2=c(t(X%*%t(t(beta[h,]*gamma1))))
    var2=(sigma[h]^2)*diag(n)
    pBgamma1=dmnorm(Y,mu2,var2)*(1-pexito)
    denominadorgamma=pBgamma1
    pexitopost[h,i]=numeradorgamma/(denominadorgamma+numeradorgamma)
    gamma1[i]=rbinom(1,1,pexitopost[h,i])
  }
  gamma[h,]=gamma1
}

## Submodelos Seleccionados

```



```

m=2000
for(j in m:M){
  model[j]=1+1*gamma[j,1]+2*gamma[j,2]+4*gamma[j,3]+8*gamma[j,4]+16*gamma[j,5]+
    32*gamma[j,6]+64*gamma[j,7]+(2^7)*gamma[j,8]+(2^8)*gamma[j,9]+
    (2^9)*gamma[j,10] +(2^10)*gamma[j,11]+(2^11)*gamma[j,12]+
    (2^12)*gamma[j,13]+(2^13)*gamma[j,14]+(2^14)*gamma[j,15]+
    (2^15)*gamma[j,16]+(2^16)*gamma[j,17]+(2^17)*gamma[j,18]+
    (2^18)*gamma[j,19]+(2^19)*gamma[j,20]}

## Probabilidad a posteriori de los submodelos
model_km=data.frame(table(model))
model_km=subset(model_km,model_km$model!=0)
model_km$Porce=model_km$Freq/(M-m+1)
model_km$Porce=round(model_km$Porce,4)
model_km=model_km[order(-model_km$Porce),]
head(model_km)

## Grafico de densidad de Beta
pdf("density_beta_km_1.pdf")
par(mfrow=c(7,3),mai=c(0.28,0.25,0.25,0.1))
plot(density(beta[2001:M,1]),main =expression(beta[1]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,2]),main =expression(beta[2]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,3]),main =expression(beta[3]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,4]),main =expression(beta[4]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,5]),main =expression(beta[5]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,6]),main =expression(beta[6]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,7]),main =expression(beta[7]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,8]),main =expression(beta[8]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,9]),main =expression(beta[9]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,10]),main =expression(beta[10]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,11]),main =expression(beta[11]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,12]),main =expression(beta[12]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,13]),main =expression(beta[13]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,14]),main =expression(beta[14]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,15]),main =expression(beta[15]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,16]),main =expression(beta[16]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,17]),main =expression(beta[17]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,18]),main =expression(beta[18]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,19]),main =expression(beta[19]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,20]),main =expression(beta[20]),xlab=NA,ylab=NA)
dev.off()

```

```
## Grafico de cadena de simulacion de Beta
pdf("mcmc_beta_km_1.pdf")
par(mfrow=c(7,3),mai=c(0.28,0.25,0.25,0.1))
ts.plot(beta[2001:M,1],main=expression(beta[1]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,2],main=expression(beta[2]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,3],main=expression(beta[3]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,4],main=expression(beta[4]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,5],main=expression(beta[5]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,6],main=expression(beta[6]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,7],main=expression(beta[7]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,8],main=expression(beta[8]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,9],main=expression(beta[9]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,10],main=expression(beta[10]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,11],main=expression(beta[11]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,12],main=expression(beta[12]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,13],main=expression(beta[13]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,14],main=expression(beta[14]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,15],main=expression(beta[15]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,16],main=expression(beta[16]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,17],main=expression(beta[17]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,18],main=expression(beta[18]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,19],main=expression(beta[19]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,20],main=expression(beta[20]),xlab=NA,ylab=NA)
dev.off()

## Grafico de densidad de Gamma
pdf("density_gamma_km_1.pdf")
par(mfrow=c(7,3),mai=c(0.28,0.25,0.25,0.1))
plot(density(gamma[2001:M,1]),main =expression(gamma[1]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,2]),main =expression(gamma[2]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,3]),main =expression(gamma[3]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,4]),main =expression(gamma[4]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,5]),main =expression(gamma[5]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,6]),main =expression(gamma[6]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,7]),main =expression(gamma[7]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,8]),main =expression(gamma[8]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,9]),main =expression(gamma[9]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,10]),main =expression(gamma[10]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,11]),main =expression(gamma[11]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,12]),main =expression(gamma[12]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,13]),main =expression(gamma[13]),xlab=NA,ylab=NA)
```



```
plot(density(gamma[2001:M,14]),main =expression(gamma[14]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,15]),main =expression(gamma[15]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,16]),main =expression(gamma[16]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,17]),main =expression(gamma[17]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,18]),main =expression(gamma[18]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,19]),main =expression(gamma[19]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,20]),main =expression(gamma[20]),xlab=NA,ylab=NA)
dev.off()
```

```
## Grafico de cadena de simulacion de Gamma
```

```
pdf("mcmc_gamma_km_1.pdf")
```

```
par(mfrow=c(7,3),mai=c(0.28,0.25,0.25,0.1))
```

```
ts.plot(gamma[2001:M,1],main=expression(gamma[1]),xlab=NA,ylab=NA)
```

```
ts.plot(gamma[2001:M,2],main=expression(gamma[2]),xlab=NA,ylab=NA)
```

```
ts.plot(gamma[2001:M,3],main=expression(gamma[3]),xlab=NA,ylab=NA)
```

```
ts.plot(gamma[2001:M,4],main=expression(gamma[4]),xlab=NA,ylab=NA)
```

```
ts.plot(gamma[2001:M,5],main=expression(gamma[5]),xlab=NA,ylab=NA)
```

```
ts.plot(gamma[2001:M,6],main=expression(gamma[6]),xlab=NA,ylab=NA)
```

```
ts.plot(gamma[2001:M,7],main=expression(gamma[7]),xlab=NA,ylab=NA)
```

```
ts.plot(gamma[2001:M,8],main=expression(gamma[8]),xlab=NA,ylab=NA)
```

```
ts.plot(gamma[2001:M,9],main=expression(gamma[9]),xlab=NA,ylab=NA)
```

```
ts.plot(gamma[2001:M,10],main=expression(gamma[10]),xlab=NA,ylab=NA)
```

```
ts.plot(gamma[2001:M,11],main=expression(gamma[11]),xlab=NA,ylab=NA)
```

```
ts.plot(gamma[2001:M,12],main=expression(gamma[12]),xlab=NA,ylab=NA)
```

```
ts.plot(gamma[2001:M,13],main=expression(gamma[13]),xlab=NA,ylab=NA)
```

```
ts.plot(gamma[2001:M,14],main=expression(gamma[14]),xlab=NA,ylab=NA)
```

```
ts.plot(gamma[2001:M,15],main=expression(gamma[15]),xlab=NA,ylab=NA)
```

```
ts.plot(gamma[2001:M,16],main=expression(gamma[16]),xlab=NA,ylab=NA)
```

```
ts.plot(gamma[2001:M,17],main=expression(gamma[17]),xlab=NA,ylab=NA)
```

```
ts.plot(gamma[2001:M,18],main=expression(gamma[18]),xlab=NA,ylab=NA)
```

```
ts.plot(gamma[2001:M,19],main=expression(gamma[19]),xlab=NA,ylab=NA)
```

```
ts.plot(gamma[2001:M,20],main=expression(gamma[20]),xlab=NA,ylab=NA)
```

```
dev.off()
```

```
## Grafico de densidad y cadena de simulacion de Sigma
```

```
pdf("sigma_km_1.pdf")
```

```
par(mfrow=c(2,1))
```

```
plot(density(sigma),main =expression(sigma),xlab=NA,ylab=NA)
```

```
ts.plot(sigma[2001:M],main=expression(sigma),xlab=NA,ylab=NA)
```

```
dev.off()
```

## A.3. Método GVS

```
#####
##### Metodo Gibbs Variable Selection (GVS) #####
##### 20 Variables Explicativas Independientes ###
#####
## Limpiando entorno de trabajo
rm(list = ls())

## Carga Librerias
library('mnormt')
library('invgamma')
library('LaplacesDemon')

## Obteniendo datos simulados
set.seed('20190123')
p=20
n=100
mu=rep(0,p)
I=diag(p)
X=rmnorm(n,mu,I)
E=rnorm(n,0,2)
Y= 2*X[,1] + 2*X[,2] + 2*X[,3] + 2*X[,4] + 2*X[,5] + E

## Estimacion por minimos cuadrados ordinarios
lm_gvs1=lm(Y~X-1)
sum_lm_gvs1=summary(lm_gvs1)

## Estableciendo parametros
M=15000
beta=matrix(0,M,p)
beta0=rep(0,p)
beta[1,]=rep(0,p)
gamma=matrix(0,M,p)
gamma[1,]=rep(1,p)
pexito=0.5
pexitopost=matrix(0,M,p)
pexitopost[1,]=rep(1,p)
sigma=numeric(M)
sigma[1]=sum_lm_gvs1$sigma
model=numeric(M-1)
C=100
omega=0.01
```

```

t=rep(C*omega,p)
Vgamma=0.01
Deltgamma=0.01

## Algoritmo de Gibbs
for(h in 2:M){

  ### Generando Beta ###
  A=c((gamma[h-1,1]*C+(1-gamma[h-1,1])*t[1]),
      (gamma[h-1,2]*C+(1-gamma[h-1,2])*t[2]),
      (gamma[h-1,3]*C+(1-gamma[h-1,3])*t[3]),
      (gamma[h-1,4]*C+(1-gamma[h-1,4])*t[4]),
      (gamma[h-1,5]*C+(1-gamma[h-1,5])*t[5]),
      (gamma[h-1,6]*C+(1-gamma[h-1,6])*t[6]),
      (gamma[h-1,7]*C+(1-gamma[h-1,7])*t[7]),
      (gamma[h-1,8]*C+(1-gamma[h-1,8])*t[8]),
      (gamma[h-1,9]*C+(1-gamma[h-1,9])*t[9]),
      (gamma[h-1,10]*C+(1-gamma[h-1,10])*t[10]),
      (gamma[h-1,11]*C+(1-gamma[h-1,11])*t[11]),
      (gamma[h-1,12]*C+(1-gamma[h-1,12])*t[12]),
      (gamma[h-1,13]*C+(1-gamma[h-1,13])*t[13]),
      (gamma[h-1,14]*C+(1-gamma[h-1,14])*t[14]),
      (gamma[h-1,15]*C+(1-gamma[h-1,15])*t[15]),
      (gamma[h-1,16]*C+(1-gamma[h-1,16])*t[16]),
      (gamma[h-1,17]*C+(1-gamma[h-1,17])*t[17]),
      (gamma[h-1,18]*C+(1-gamma[h-1,18])*t[18]),
      (gamma[h-1,19]*C+(1-gamma[h-1,19])*t[19]),
      (gamma[h-1,20]*C+(1-gamma[h-1,20])*t[20]))
  D=diag(A,p,p)
  A1=c((gamma[h-1,1]*0+(1-gamma[h-1,1])*beta0[1]),
      (gamma[h-1,2]*0+(1-gamma[h-1,2])*beta0[2]),
      (gamma[h-1,3]*0+(1-gamma[h-1,3])*beta0[3]),
      (gamma[h-1,4]*0+(1-gamma[h-1,4])*beta0[4]),
      (gamma[h-1,5]*0+(1-gamma[h-1,5])*beta0[5]),
      (gamma[h-1,6]*0+(1-gamma[h-1,6])*beta0[6]),
      (gamma[h-1,7]*0+(1-gamma[h-1,7])*beta0[7]),
      (gamma[h-1,8]*0+(1-gamma[h-1,8])*beta0[8]),
      (gamma[h-1,9]*0+(1-gamma[h-1,9])*beta0[9]),
      (gamma[h-1,10]*0+(1-gamma[h-1,10])*beta0[10]),
      (gamma[h-1,11]*0+(1-gamma[h-1,11])*beta0[11]),
      (gamma[h-1,12]*0+(1-gamma[h-1,12])*beta0[12]),
      (gamma[h-1,13]*0+(1-gamma[h-1,13])*beta0[13]),

```

```

(gamma[h-1,14]*0+(1-gamma[h-1,14])*beta0[14]),
(gamma[h-1,15]*0+(1-gamma[h-1,15])*beta0[15]),
(gamma[h-1,16]*0+(1-gamma[h-1,16])*beta0[16]),
(gamma[h-1,17]*0+(1-gamma[h-1,17])*beta0[17]),
(gamma[h-1,18]*0+(1-gamma[h-1,18])*beta0[18]),
(gamma[h-1,19]*0+(1-gamma[h-1,19])*beta0[19]),
(gamma[h-1,20]*0+(1-gamma[h-1,20])*beta0[20]))
beta00=c(A1)
X2=X%%diag(gamma[h-1,])
varbeta=solve(solve(D) + (sigma[h-1]^(-2))*t(X2)%%X2)
mubeta=varbeta%%(solve(D)%%beta00+(sigma[h-1]^(-2))*t(X2)%%Y)
beta[h,]=rmnorm(1,mubeta,varbeta)

### Generando Sigma ###
A3=beta[h,]*gamma[h-1,]
theta=t(t(A3))
sigma[h]=(rinvgamma(1, (n+Vgamma)/2, (Deltgamma +
t(Y-X%%theta)%%(Y-X%%theta))/2))^^(1/2)

### Generando Gamma ###
gamma1=gamma[h-1,]
W=sample(1:p,p) ## Considerando la recomendacion de Ntzoufras (2002)
## de no generar gamma_i de manera consecutiva

for(i in W){
  gamma1[i]=1
  mu2=c(t(X%%t(t(beta[h,]*gamma1))))
  var2=(sigma[h]^2)*diag(n)
  pYgamma1=dmnorm(Y,mu2,var2)
  Agamma1=c((gamma1[1]*C+(1-gamma1[1])*t[1]),
            (gamma1[2]*C+(1-gamma1[2])*t[2]),
            (gamma1[3]*C+(1-gamma1[3])*t[3]),
            (gamma1[4]*C+(1-gamma1[4])*t[4]),
            (gamma1[5]*C+(1-gamma1[5])*t[5]),
            (gamma1[6]*C+(1-gamma1[6])*t[6]),
            (gamma1[7]*C+(1-gamma1[7])*t[7]),
            (gamma1[8]*C+(1-gamma1[8])*t[8]),
            (gamma1[9]*C+(1-gamma1[9])*t[9]),
            (gamma1[10]*C+(1-gamma1[10])*t[10]),
            (gamma1[11]*C+(1-gamma1[11])*t[11]),
            (gamma1[12]*C+(1-gamma1[12])*t[12]),
            (gamma1[13]*C+(1-gamma1[13])*t[13]),

```

```

(gamma1[14]*C+(1-gamma1[14])*t[14]),
(gamma1[15]*C+(1-gamma1[15])*t[15]),
(gamma1[16]*C+(1-gamma1[16])*t[16]),
(gamma1[17]*C+(1-gamma1[17])*t[17]),
(gamma1[18]*C+(1-gamma1[18])*t[18]),
(gamma1[19]*C+(1-gamma1[19])*t[19]),
(gamma1[20]*C+(1-gamma1[20])*t[20]))
Dgamma=diag(Agamma1,p,p)
Agamma2=c((gamma1[1]*0+(1-gamma1[1])*beta0[1]),
(gamma1[2]*0+(1-gamma1[2])*beta0[2]),
(gamma1[3]*0+(1-gamma1[3])*beta0[3]),
(gamma1[4]*0+(1-gamma1[4])*beta0[4]),
(gamma1[5]*0+(1-gamma1[5])*beta0[5]),
(gamma1[6]*0+(1-gamma1[6])*beta0[6]),
(gamma1[7]*0+(1-gamma1[7])*beta0[7]),
(gamma1[8]*0+(1-gamma1[8])*beta0[8]),
(gamma1[9]*0+(1-gamma1[9])*beta0[9]),
(gamma1[10]*0+(1-gamma1[10])*beta0[10]),
(gamma1[11]*0+(1-gamma1[11])*beta0[11]),
(gamma1[12]*0+(1-gamma1[12])*beta0[12]),
(gamma1[13]*0+(1-gamma1[13])*beta0[13]),
(gamma1[14]*0+(1-gamma1[14])*beta0[14]),
(gamma1[15]*0+(1-gamma1[15])*beta0[15]),
(gamma1[16]*0+(1-gamma1[16])*beta0[16]),
(gamma1[17]*0+(1-gamma1[17])*beta0[17]),
(gamma1[18]*0+(1-gamma1[18])*beta0[18]),
(gamma1[19]*0+(1-gamma1[19])*beta0[19]),
(gamma1[20]*0+(1-gamma1[20])*beta0[20]))
betagama=c(Agamma2)
pBgamma1=dmnorm(beta[h,],betagama,Dgamma)
numeradorgamma=pYgamma1*pBgamma1*pexito

gamma1[i]=0
mu2=c(t(X%*%t(t(beta[h,]*gamma1))))
var2=(sigma[h]^2)*diag(n)
pYgamma1=dmnorm(Y,mu2,var2)
Agamma1=c((gamma1[1]*C+(1-gamma1[1])*t[1]),
(gamma1[2]*C+(1-gamma1[2])*t[2]),
(gamma1[3]*C+(1-gamma1[3])*t[3]),
(gamma1[4]*C+(1-gamma1[4])*t[4]),
(gamma1[5]*C+(1-gamma1[5])*t[5]),
(gamma1[6]*C+(1-gamma1[6])*t[6]),

```

```

(gamma1[7]*C+(1-gamma1[7])*t[7]),
(gamma1[8]*C+(1-gamma1[8])*t[8]),
(gamma1[9]*C+(1-gamma1[9])*t[9]),
(gamma1[10]*C+(1-gamma1[10])*t[10]),
(gamma1[11]*C+(1-gamma1[11])*t[11]),
(gamma1[12]*C+(1-gamma1[12])*t[12]),
(gamma1[13]*C+(1-gamma1[13])*t[13]),
(gamma1[14]*C+(1-gamma1[14])*t[14]),
(gamma1[15]*C+(1-gamma1[15])*t[15]),
(gamma1[16]*C+(1-gamma1[16])*t[16]),
(gamma1[17]*C+(1-gamma1[17])*t[17]),
(gamma1[18]*C+(1-gamma1[18])*t[18]),
(gamma1[19]*C+(1-gamma1[19])*t[19]),
(gamma1[20]*C+(1-gamma1[20])*t[20]))
Dgamma=diag(Agamma1,p,p)
Agamma2=c((gamma1[1]*0+(1-gamma1[1])*beta0[1]),
(gamma1[2]*0+(1-gamma1[2])*beta0[2]),
(gamma1[3]*0+(1-gamma1[3])*beta0[3]),
(gamma1[4]*0+(1-gamma1[4])*beta0[4]),
(gamma1[5]*0+(1-gamma1[5])*beta0[5]),
(gamma1[6]*0+(1-gamma1[6])*beta0[6]),
(gamma1[7]*0+(1-gamma1[7])*beta0[7]),
(gamma1[8]*0+(1-gamma1[8])*beta0[8]),
(gamma1[9]*0+(1-gamma1[9])*beta0[9]),
(gamma1[10]*0+(1-gamma1[10])*beta0[10]),
(gamma1[11]*0+(1-gamma1[11])*beta0[11]),
(gamma1[12]*0+(1-gamma1[12])*beta0[12]),
(gamma1[13]*0+(1-gamma1[13])*beta0[13]),
(gamma1[14]*0+(1-gamma1[14])*beta0[14]),
(gamma1[15]*0+(1-gamma1[15])*beta0[15]),
(gamma1[16]*0+(1-gamma1[16])*beta0[16]),
(gamma1[17]*0+(1-gamma1[17])*beta0[17]),
(gamma1[18]*0+(1-gamma1[18])*beta0[18]),
(gamma1[19]*0+(1-gamma1[19])*beta0[19]),
(gamma1[20]*0+(1-gamma1[20])*beta0[20]))
betagama=c(Agamma2)
pBgamma1=dmnorm(beta[h,],betagama,Dgamma)
denominadorgamma=pYgamma1*pBgamma1*(1-pexito)
pexitopost[h,i]=numeradorgamma/(denominadorgamma+numeradorgamma)
gamma1[i]=rbinom(1,1,pexitopost[h,i])
}
gamma[h,]=gamma1

```



```

}

## Submodelos Seleccionados
m=2000
for(j in m:M){
  model[j]=1+1*gamma[j,1]+2*gamma[j,2]+4*gamma[j,3]+8*gamma[j,4]+16*gamma[j,5]+
    32*gamma[j,6]+64*gamma[j,7]+(2^7)*gamma[j,8]+(2^8)*gamma[j,9]+
    (2^9)*gamma[j,10] +(2^10)*gamma[j,11]+(2^11)*gamma[j,12]+
    (2^12)*gamma[j,13]+(2^13)*gamma[j,14]+(2^14)*gamma[j,15]+
    (2^15)*gamma[j,16]+(2^16)*gamma[j,17]+(2^17)*gamma[j,18]+
    (2^18)*gamma[j,19]+(2^19)*gamma[j,20]}

## Probabilidad a posteriori de los submodelos
model_gvs=data.frame(table(model))
model_gvs=subset(model_gvs,model_gvs$model!=0)
model_gvs$Porce=model_gvs$Freq/(M-m+1)
model_gvs$Porce=round(model_gvs$Porce,4)
model_gvs=model_gvs[order(-model_gvs$Porce),]
head(model_gvs)

## Grafico de densidad de Beta
pdf("density_beta_gvs_1.pdf")
par(mfrow=c(7,3),mai=c(0.28,0.25,0.25,0.1))
plot(density(beta[2001:M,1]),main =expression(beta[1]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,2]),main =expression(beta[2]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,3]),main =expression(beta[3]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,4]),main =expression(beta[4]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,5]),main =expression(beta[5]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,6]),main =expression(beta[6]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,7]),main =expression(beta[7]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,8]),main =expression(beta[8]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,9]),main =expression(beta[9]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,10]),main =expression(beta[10]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,11]),main =expression(beta[11]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,12]),main =expression(beta[12]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,13]),main =expression(beta[13]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,14]),main =expression(beta[14]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,15]),main =expression(beta[15]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,16]),main =expression(beta[16]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,17]),main =expression(beta[17]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,18]),main =expression(beta[18]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,19]),main =expression(beta[19]),xlab=NA,ylab=NA)
plot(density(beta[2001:M,20]),main =expression(beta[20]),xlab=NA,ylab=NA)

```



```
dev.off()

## Grafico de cadena de simulacion de Beta
pdf("mcmc_beta_gvs_1.pdf")
par(mfrow=c(7,3),mai=c(0.28,0.25,0.25,0.1))
ts.plot(beta[2001:M,1],main=expression(beta[1]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,2],main=expression(beta[2]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,3],main=expression(beta[3]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,4],main=expression(beta[4]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,5],main=expression(beta[5]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,6],main=expression(beta[6]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,7],main=expression(beta[7]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,8],main=expression(beta[8]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,9],main=expression(beta[9]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,10],main=expression(beta[10]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,11],main=expression(beta[11]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,12],main=expression(beta[12]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,13],main=expression(beta[13]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,14],main=expression(beta[14]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,15],main=expression(beta[15]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,16],main=expression(beta[16]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,17],main=expression(beta[17]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,18],main=expression(beta[18]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,19],main=expression(beta[19]),xlab=NA,ylab=NA)
ts.plot(beta[2001:M,20],main=expression(beta[20]),xlab=NA,ylab=NA)
dev.off()

## Grafico de densidad de Gamma
pdf("density_gamma_gvs_1.pdf")
par(mfrow=c(7,3),mai=c(0.28,0.25,0.25,0.1))
plot(density(gamma[2001:M,1]),main =expression(gamma[1]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,2]),main =expression(gamma[2]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,3]),main =expression(gamma[3]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,4]),main =expression(gamma[4]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,5]),main =expression(gamma[5]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,6]),main =expression(gamma[6]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,7]),main =expression(gamma[7]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,8]),main =expression(gamma[8]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,9]),main =expression(gamma[9]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,10]),main =expression(gamma[10]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,11]),main =expression(gamma[11]),xlab=NA,ylab=NA)
```

```
plot(density(gamma[2001:M,12]),main =expression(gamma[12]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,13]),main =expression(gamma[13]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,14]),main =expression(gamma[14]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,15]),main =expression(gamma[15]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,16]),main =expression(gamma[16]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,17]),main =expression(gamma[17]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,18]),main =expression(gamma[18]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,19]),main =expression(gamma[19]),xlab=NA,ylab=NA)
plot(density(gamma[2001:M,20]),main =expression(gamma[20]),xlab=NA,ylab=NA)
dev.off()
```

```
## Grafico de cadena de simulacion de Gamma
pdf("mcmc_gamma_gvs_1.pdf")
par(mfrow=c(7,3),mai=c(0.28,0.25,0.25,0.1))
ts.plot(gamma[2001:M,1],main=expression(gamma[1]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,2],main=expression(gamma[2]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,3],main=expression(gamma[3]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,4],main=expression(gamma[4]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,5],main=expression(gamma[5]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,6],main=expression(gamma[6]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,7],main=expression(gamma[7]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,8],main=expression(gamma[8]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,9],main=expression(gamma[9]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,10],main=expression(gamma[10]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,11],main=expression(gamma[11]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,12],main=expression(gamma[12]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,13],main=expression(gamma[13]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,14],main=expression(gamma[14]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,15],main=expression(gamma[15]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,16],main=expression(gamma[16]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,17],main=expression(gamma[17]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,18],main=expression(gamma[18]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,19],main=expression(gamma[19]),xlab=NA,ylab=NA)
ts.plot(gamma[2001:M,20],main=expression(gamma[20]),xlab=NA,ylab=NA)
dev.off()
```

```
## Grafico de densidad y cadena de simulacion de Sigma
pdf("sigma_gvs_1.pdf")
par(mfrow=c(2,1))
plot(density(sigma),main =expression(sigma),xlab=NA,ylab=NA)
ts.plot(sigma[2001:M],main=expression(sigma),xlab=NA,ylab=NA)
```

```
dev.off()
```

#### A.4. Método Lasso

```
#####
### Metodo tradicional de Seleccion de variables - Lasso ###
##### 20 Variables Explicativas Independientes #####
#####

## Limpiando entorno de trabajo
rm(list = ls())

## Carga librerias
library(glmnet)
library('mnormt')
library('invgamma')
library('LaplacesDemon')

## Obteniendo datos simulados
set.seed('20190123')
p=20
n=100
mu=rep(0,p)
I=diag(p)
X=rmnorm(n,mu,I)
E=rnorm(n,0,2)
Y= 2*X[,1] + 2*X[,2] + 2*X[,3] + 2*X[,4] + 2*X[,5] + E

## Algoritmo Lasso
cv.glmmod = cv.glmnet(x=X, y=Y, alpha=1, intercept = FALSE)
best.lambda1 = cv.glmmod$lambda.min
best.lambda2= cv.glmmod$lambda.1se
MSE=cv.glmmod$cvsd
plot(cv.glmmod)

## Seleccion del submodelo
coef(cv.glmmod,s=best.lambda1) # Modelo con menor error cuadratico medio
coef(cv.glmmod,s=best.lambda2) # Mejor modelo con menor ECM y menor numero de variables
```

## A.5. Método Stepwise

```
#####
### Metodo tradicional de Seleccion de variables - Stepwise ##
##### 20 Variables Explicativas Independientes #####
#####

## Limpiando entorno de trabajo
rm(list = ls())

## Carga Librerias
library(MASS)
library('mnormt')
library('invgamma')
library('LaplacesDemon')

## Obteniendo datos simulados
set.seed('20190123')
p=20
n=100
mu=rep(0,p)
I=diag(p)
X=rmnorm(n,mu,I)
E=rnorm(n,0,2)
Y= 2*X[,1] + 2*X[,2] + 2*X[,3] + 2*X[,4] + 2*X[,5] + E

## Creacion de modelo sin interseco
model1 = lm(Y~X[,1]+X[,2]+X[,3]+X[,4]+X[,5]
            +X[,6]+X[,7]+X[,8]+X[,9]+X[,10]
            +X[,11]+X[,12]+X[,13]+X[,14]+X[,15]
            +X[,16]+X[,17]+X[,18]+X[,19]+X[,20]-1)

## Algoritmo StepAIC
model1.stepAIC <- stepAIC(model1, direction=c("both"))

## Seleccion del submodelo
model1.stepAIC$coefficients
model1.stepAIC$aov
summary(model1.stepAIC)
```

## Apéndice B

### Análisis de convergencia

#### B.1. Convergencia de los parámetros en el estudio con datos simulados para veinte variables explicativas independientes

##### Método SSVS

En la Figura B.1 se presentan los gráficos de las distribuciones a posteriori condicional y cadenas de las estimaciones de los parámetros  $\beta$ ,  $\gamma$  y  $\sigma$ ; aplicando la sugerencia de  $(\phi_j, C_j) = (0.1, 500)$  para los parámetros de la distribución a priori de  $\beta_j$ .

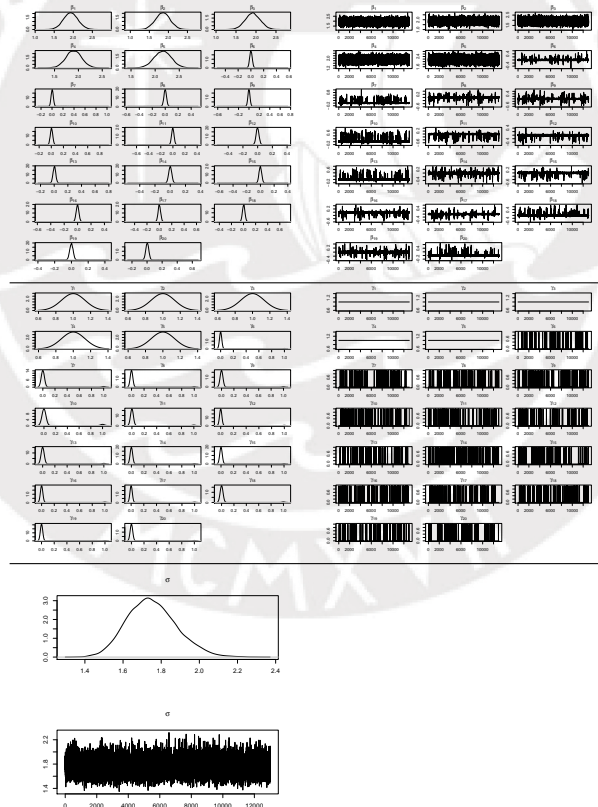


Figura B.1: Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros  $\beta$ ,  $\gamma$ ,  $\sigma$  empleando el método SSVS y considerando  $(\phi_j, C_j) = (0.1, 500)$ , en el estudio con datos simulados para veinte variables explicativas independientes.

Método KM

En la Figura B.2 se presentan los gráficos de las distribuciones a posteriori condicional y cadenas de las estimaciones de los parámetros  $\beta$ ,  $\gamma$  y  $\sigma$ ; aplicando la sugerencia de  $D_0 = 16I$  para la matriz de varianza y covarianza de la distribución a priori de  $\beta$ .

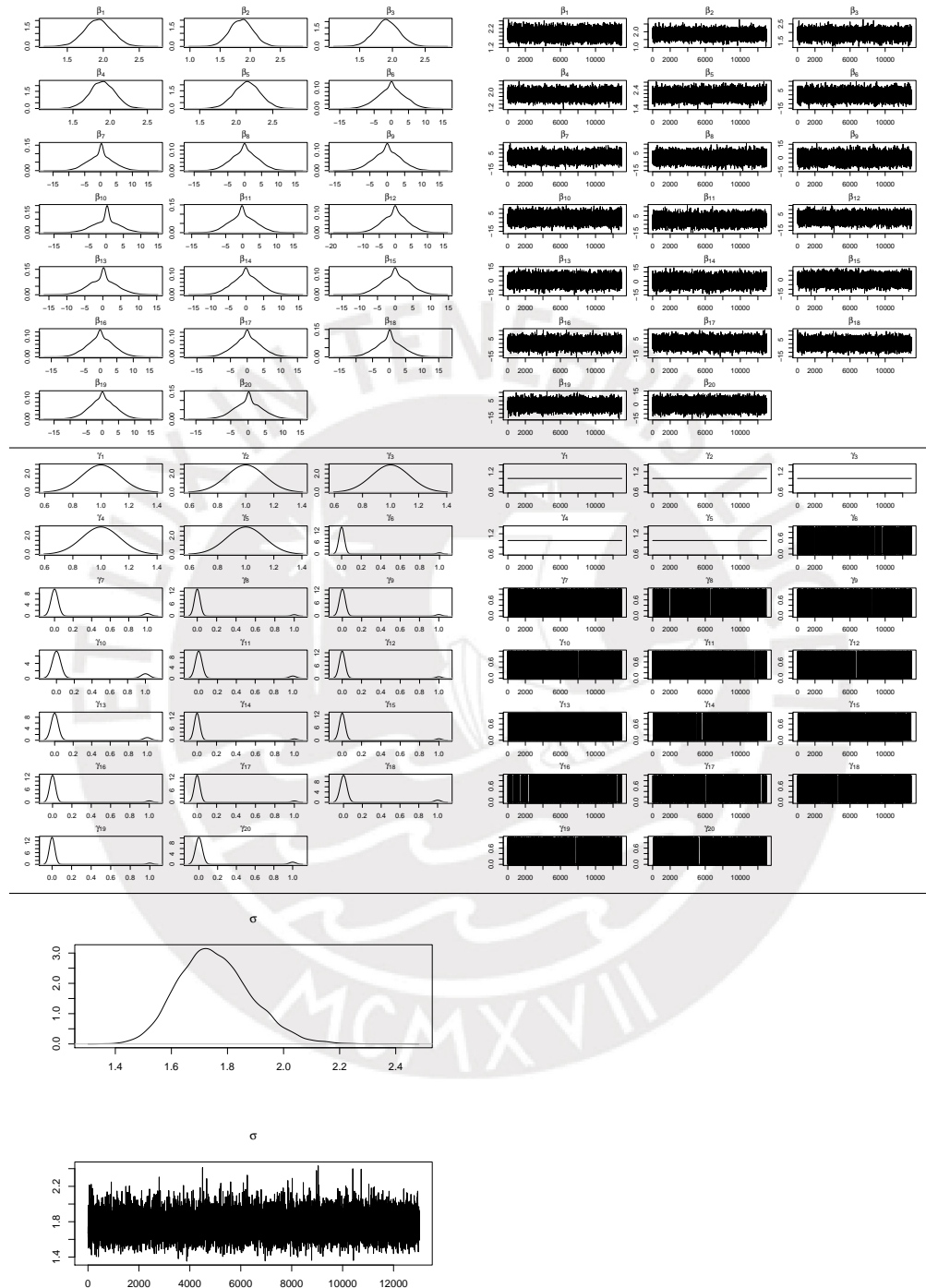


Figura B.2: Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros  $\beta$ ,  $\gamma$ ,  $\sigma$  empleando el método KM y considerando  $D_0 = 16I$ , en el estudio con datos simulados para veinte variables explicativas independientes.



Método GVS

En la Figura B.3 se presentan los gráficos de las distribuciones a posteriori condicional y cadenas de las estimaciones de los parámetros  $\beta$ ,  $\gamma$  y  $\sigma$ ; aplicando la sugerencia de  $\Sigma_j = 100$  para la varianza de la distribución a priori de los  $\beta_j$  seleccionados.

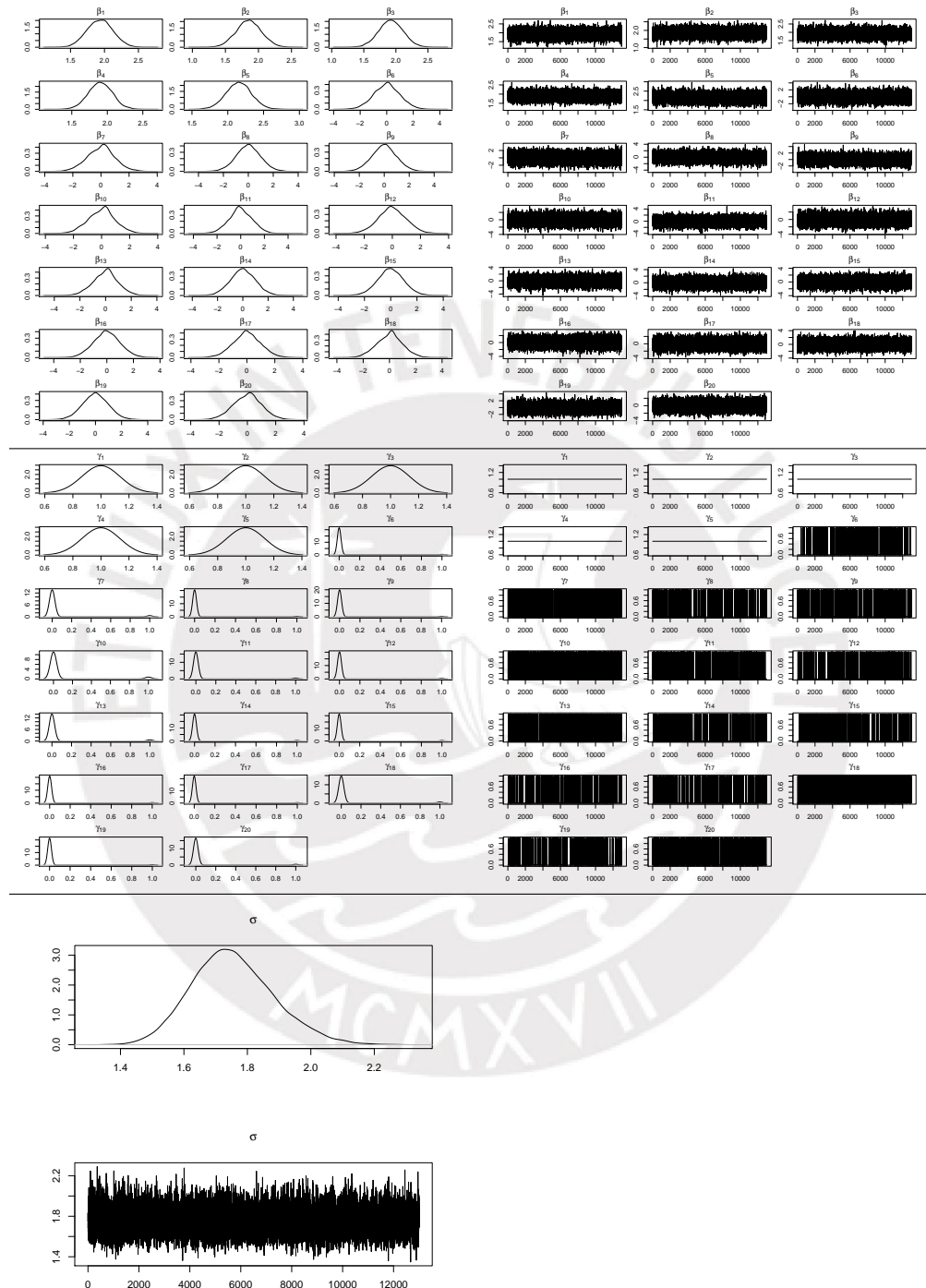


Figura B.3: Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros  $\beta$ ,  $\gamma$ ,  $\sigma$  empleando el método GVS y considerando  $\Sigma_j = 100$ , en el Estudio con datos simulados para veinte variables explicativas independientes.



## B.2. Convergencia de los parámetros en el estudio con datos simulados para cuatro variables correlacionadas

### Método SSVS

En la Figura B.4 se presentan los gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros  $\beta$ ,  $\gamma$  y  $\sigma$ ; aplicando la sugerencia de  $(\phi_j, C_j) = (0.1, 500)$  para los parámetros de la distribución a priori de  $\beta_j$ .



Figura B.4: Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros  $\beta$ ,  $\gamma$ ,  $\sigma$  empleando el método SSVS y considerando  $(\phi_j, C_j) = (0.1, 500)$ , en el estudio con datos simulados para cuatro variables correlacionadas.

Método KM

En la Figura B.5 se presentan los gráficos de las distribuciones a posteriori condicional y cadenas de las estimaciones de los parámetros  $\beta$ ,  $\gamma$  y  $\sigma$ ; aplicando la sugerencia de  $D_0 = 16I$  para la matriz de varianza y covarianza de la distribución a priori de  $\beta$ .

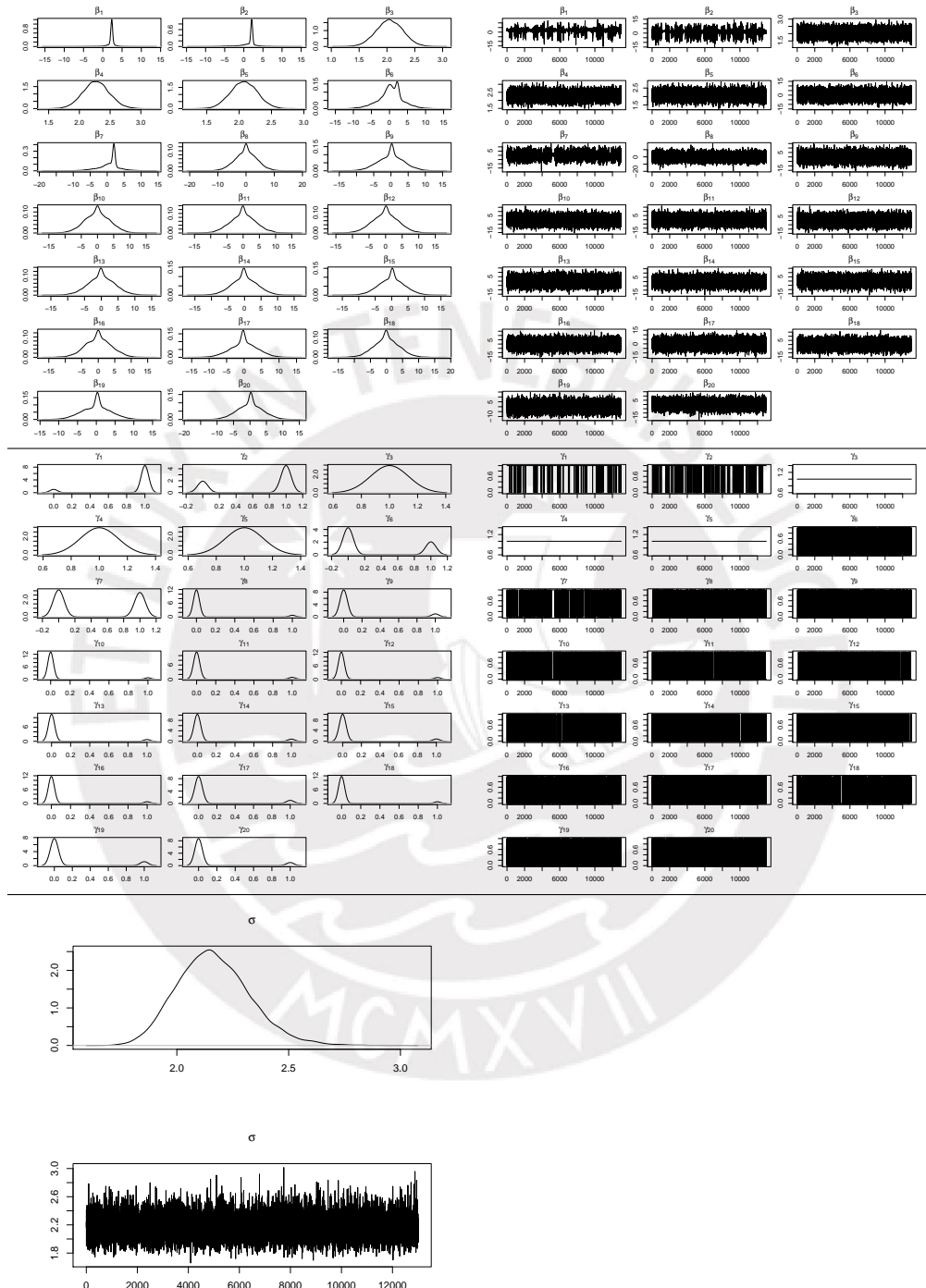


Figura B.5: Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros  $\beta$ ,  $\gamma$ ,  $\sigma$  empleando el método KM y considerando  $D_0 = 16I$ , en el estudio con datos simulados para cuatro variables correlacionadas.

Método GVS

En la Figura B.6 se presentan los gráficos de las distribuciones a posteriori condicional y cadenas de las estimaciones de los parámetros  $\beta$ ,  $\gamma$  y  $\sigma$ ; aplicando la sugerencia de  $\Sigma_j = 100$  para la varianza de la distribución a priori de los  $\beta_j$  seleccionados.

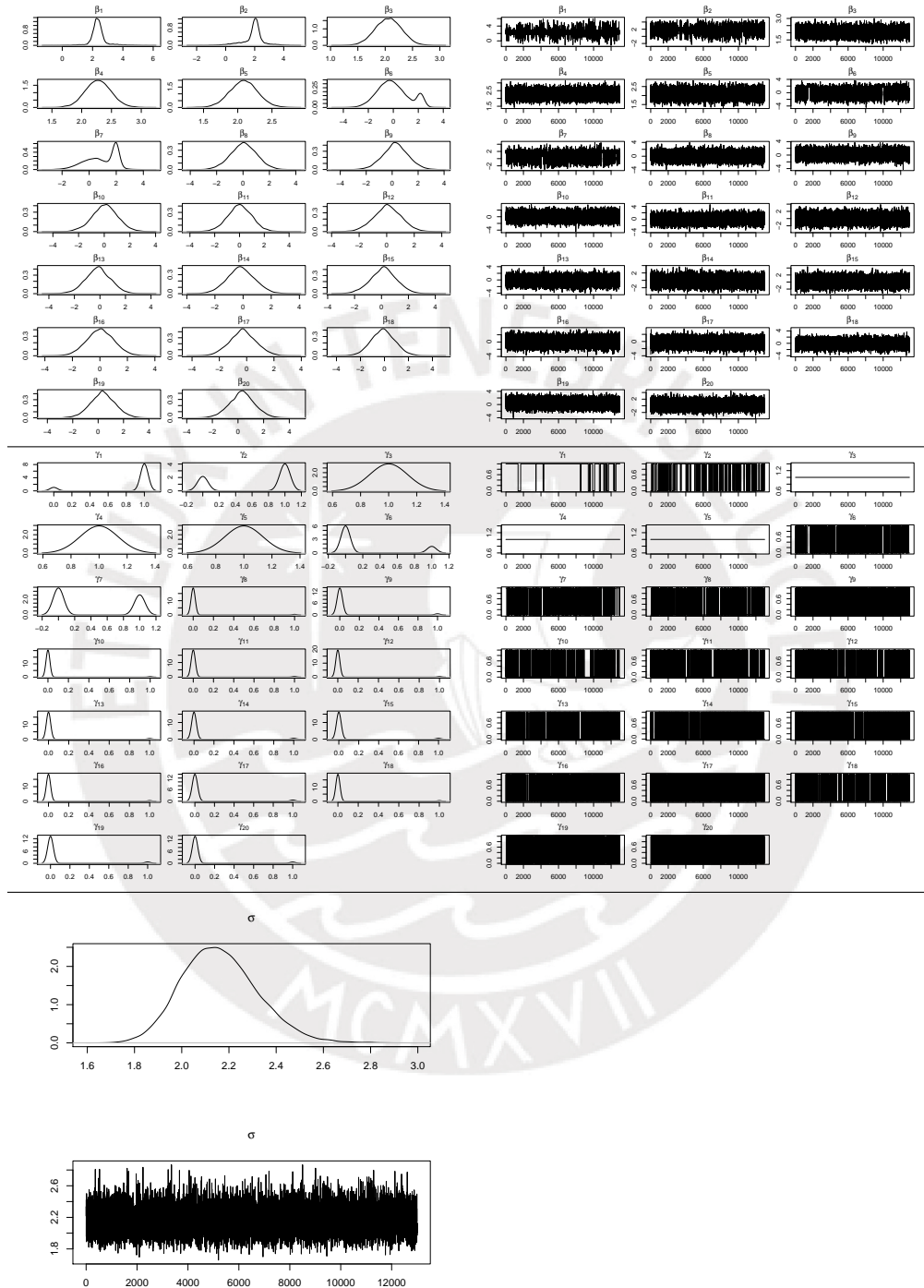


Figura B.6: Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros  $\beta$ ,  $\gamma$ ,  $\sigma$  empleando el método GVS y considerando  $\Sigma_j = 100$ , en el estudio con datos simulados para cuatro variables correlacionadas.

### B.3. Convergencia de los parámetros en el estudio con datos simulados para dos variables correlacionadas y una variable candidata con multicolinealidad

#### Método SSVS

En la Figura B.7 se presentan los gráficos de las distribuciones a posteriori condicional y cadenas de las estimaciones de los parámetros  $\beta$ ,  $\gamma$  y  $\sigma$ ; aplicando la sugerencia de  $(\phi_j, C_j) = (0.1, 500)$  para los parámetros de la distribución a priori de  $\beta_j$ .

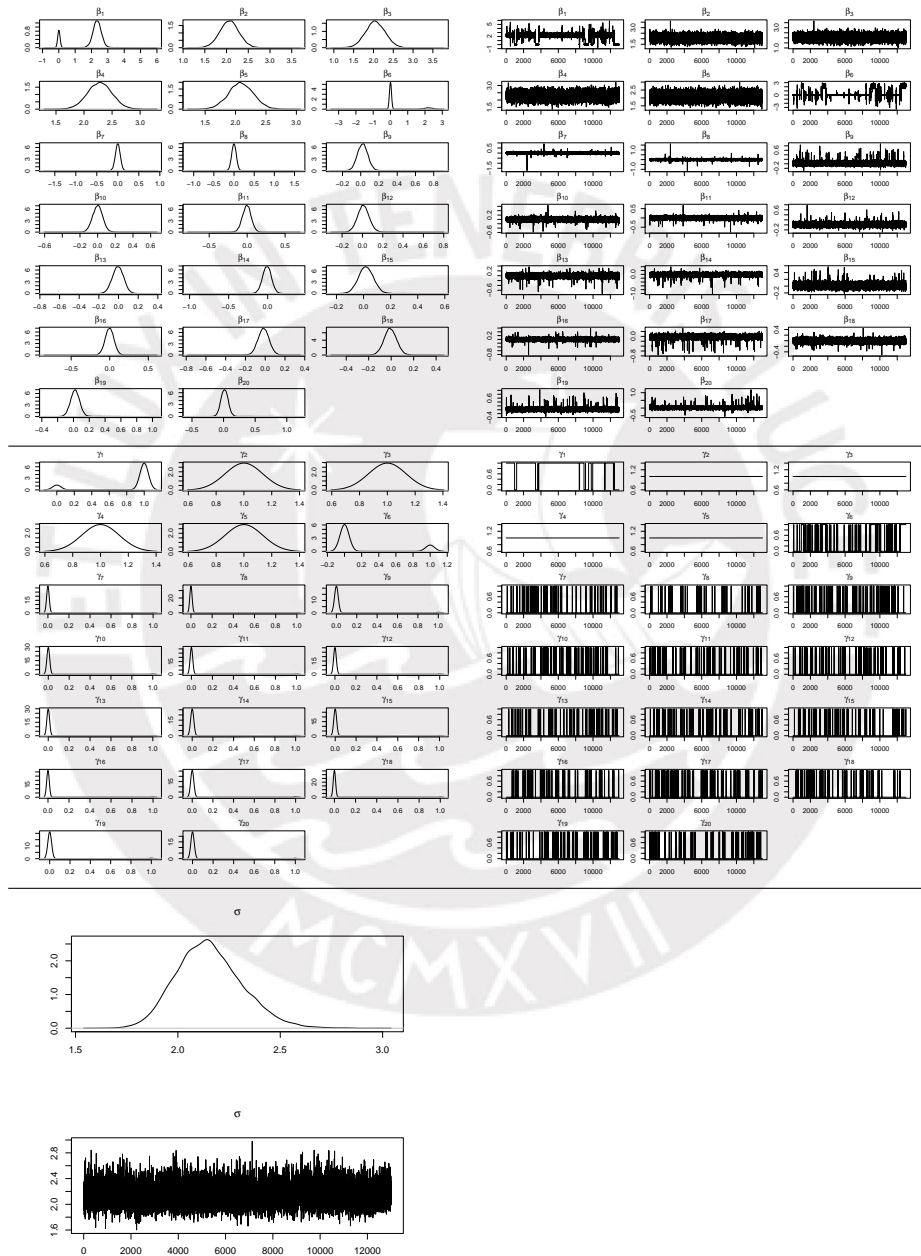


Figura B.7: Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros  $\beta$ ,  $\gamma$ ,  $\sigma$  empleando el método SSVS y considerando  $(\phi_j, C_j) = (0.1, 500)$ , en el estudio con datos simulados dos variables correlacionadas y una variable candidata con multicolinealidad.

Método KM

En la Figura B.8 se presentan los gráficos de las distribuciones a posteriori condicional y cadenas de las estimaciones de los parámetros  $\beta$ ,  $\gamma$  y  $\sigma$ ; aplicando la sugerencia de  $D_0 = 16I$  para la matriz de varianza y covarianza de la distribución a priori de  $\beta$ .

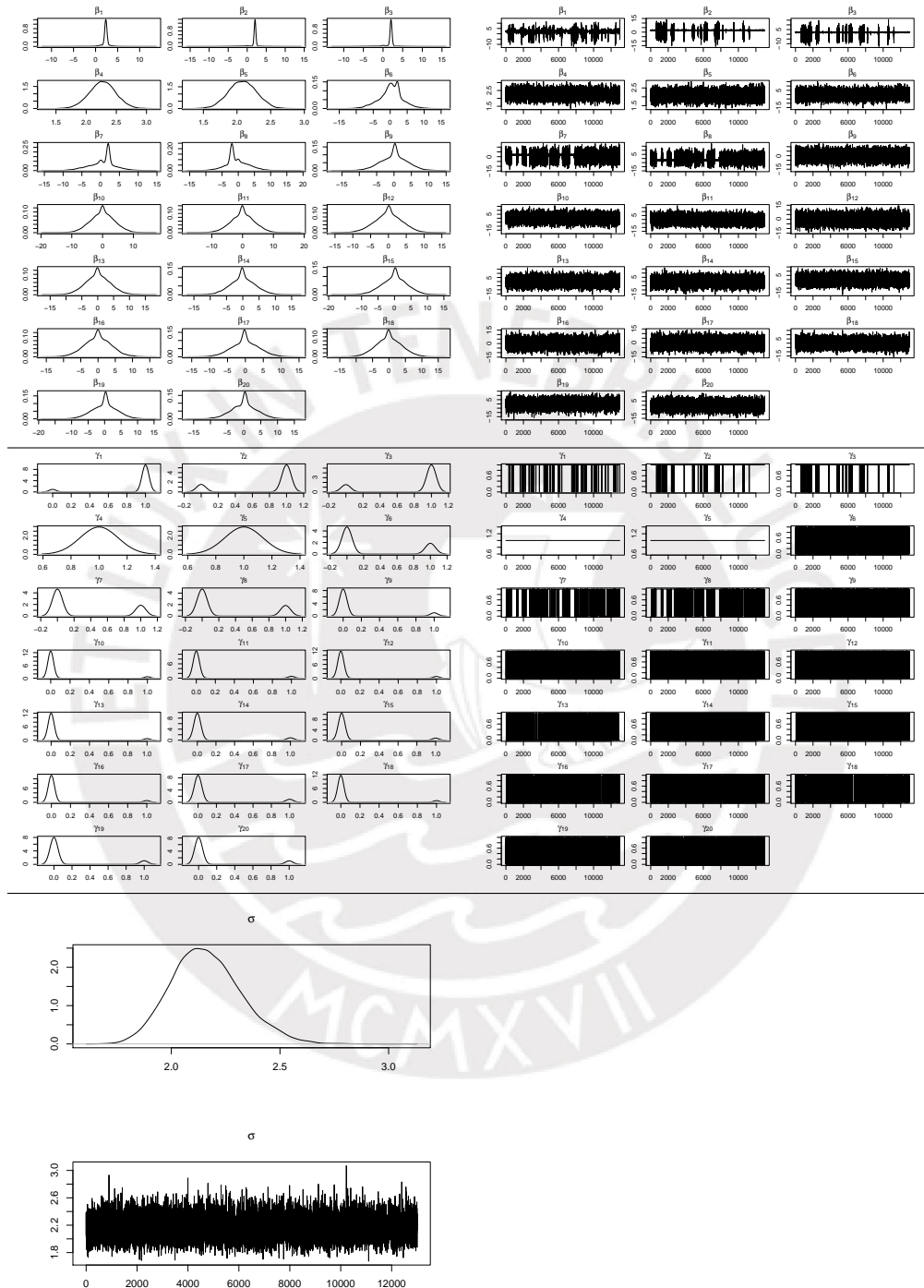


Figura B.8: Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros  $\beta$ ,  $\gamma$ ,  $\sigma$  empleando el método KM y considerando  $D_0 = 16I$ , en el estudio con datos simulados para dos variables correlacionadas y una variable candidata con multicolinealidad.

Método GVS

En la Figura B.9 se presentan los gráficos de las distribuciones a posteriori condicional y cadenas de las estimaciones de los parámetros  $\beta$ ,  $\gamma$  y  $\sigma$ ; aplicando la sugerencia de  $\Sigma_j = 100$  para la varianza de la distribución a priori de los  $\beta_j$  seleccionados.

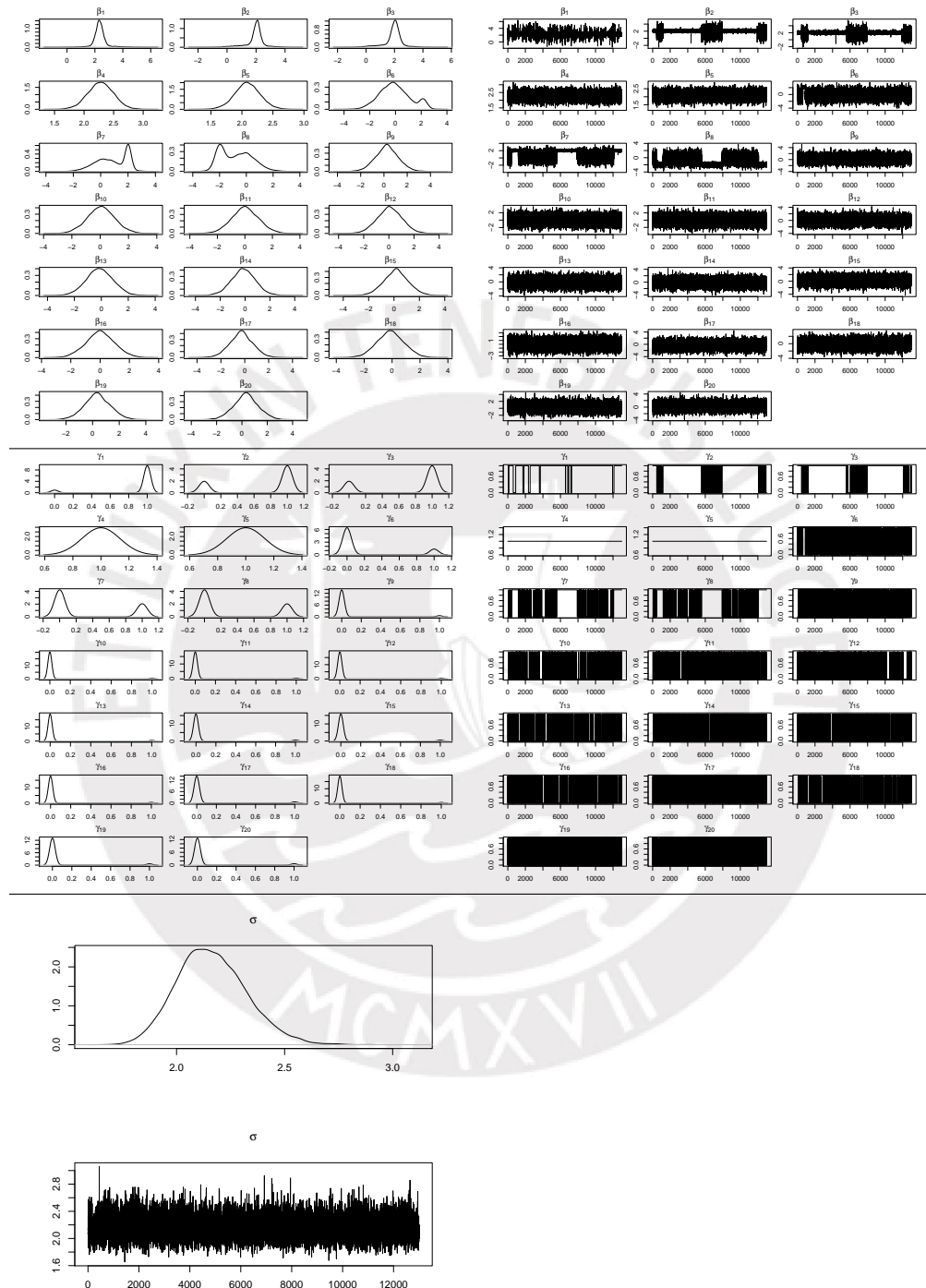


Figura B.9: Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros  $\beta$ ,  $\gamma$ ,  $\sigma$  empleando el método GVS y considerando  $\Sigma_j = 100$ , en el estudio con datos simulados para dos variables correlacionadas y una variable candidata con multicolinealidad.

### B.4. Convergencia de los parámetros en el estudio con datos simulados para dos variables correlacionadas y una variable relevante con multicolinealidad

#### Método SSVS

En la Figura B.10 se presentan los gráficos de las distribuciones a posteriori condicional y cadenas de las estimaciones de los parámetros  $\beta$ ,  $\gamma$  y  $\sigma$ ; aplicando la sugerencia de  $(\phi_j, C_j) = (0.1, 500)$  para los parámetros de la distribución a priori de  $\beta_j$ .

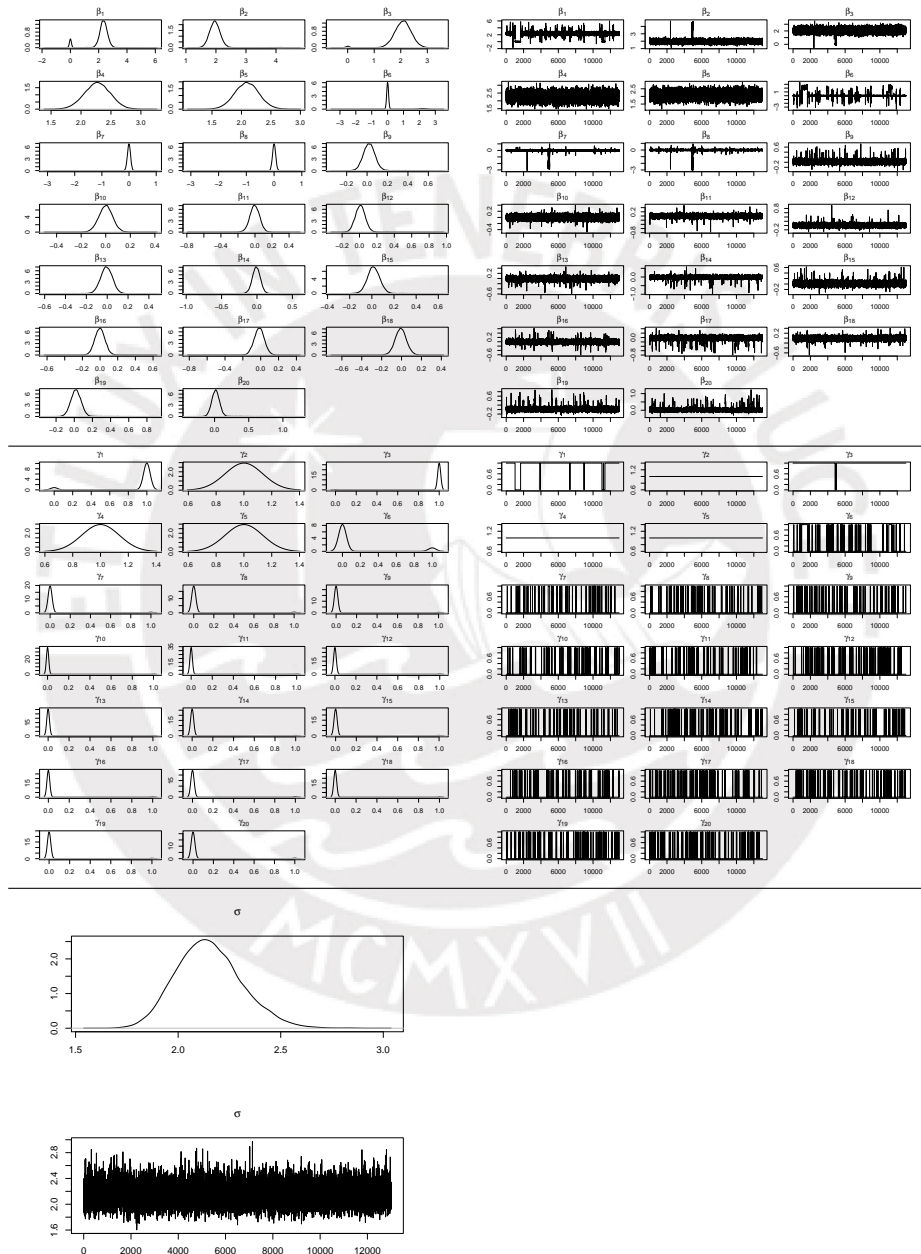


Figura B.10: Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros  $\beta$ ,  $\gamma$ ,  $\sigma$  empleando el método SSVS y considerando  $(\phi_j, C_j) = (0.1, 500)$ , en el estudio con datos simulados dos variables correlacionadas y una variable relevante con multicolinealidad.



Método KM

En la Figura B.11 se presentan los gráficos de las distribuciones a posteriori condicional y cadenas de las estimaciones de los parámetros  $\beta$ ,  $\gamma$  y  $\sigma$ ; aplicando la sugerencia de  $D_0 = 16I$  para la matriz de varianza y covarianza de la distribución a priori de  $\beta$ .



Figura B.11: Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros  $\beta$ ,  $\gamma$ ,  $\sigma$  empleando el método KM y considerando  $D_0 = 16I$ , en el estudio con datos simulados para dos variables correlacionadas y una variable relevante con multicolinealidad.

Método GVS

En la Figura B.12 se presentan los gráficos de las distribuciones a posteriori condicional y cadenas de las estimaciones de los parámetros  $\beta$ ,  $\gamma$  y  $\sigma$ ; aplicando la sugerencia de  $\Sigma_j = 100$  para la varianza de la distribución a priori de los  $\beta_j$  seleccionados.

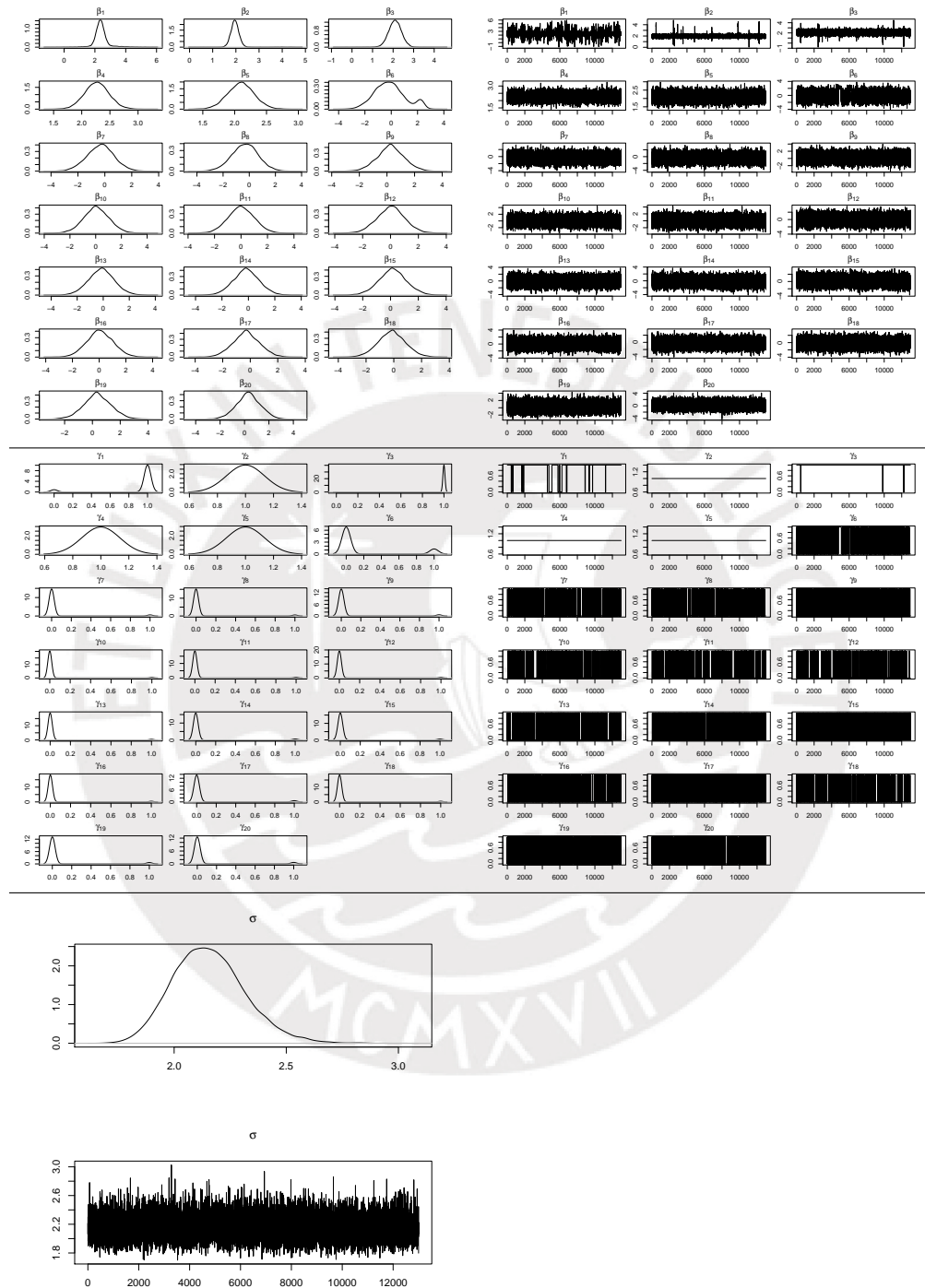


Figura B.12: Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros  $\beta$ ,  $\gamma$ ,  $\sigma$  empleando el método GVS y considerando  $\Sigma_j = 100$ , en el estudio con datos simulados para dos variables correlacionadas y una variable relevante con multicolinealidad.

### B.5. Convergencia de los parámetros en el estudio de la Esperanza de vida al nacer

#### Método SSVS

En la Figura B.13 se presentan los gráficos de las distribuciones a posteriori condicional y cadenas de las estimaciones de los parámetros  $\beta$ ,  $\gamma$  y  $\sigma$ ; aplicando la sugerencia de  $(\phi_j, C_j) = (0.1, 500)$  para los parámetros de la distribución a priori de  $\beta_j$ .

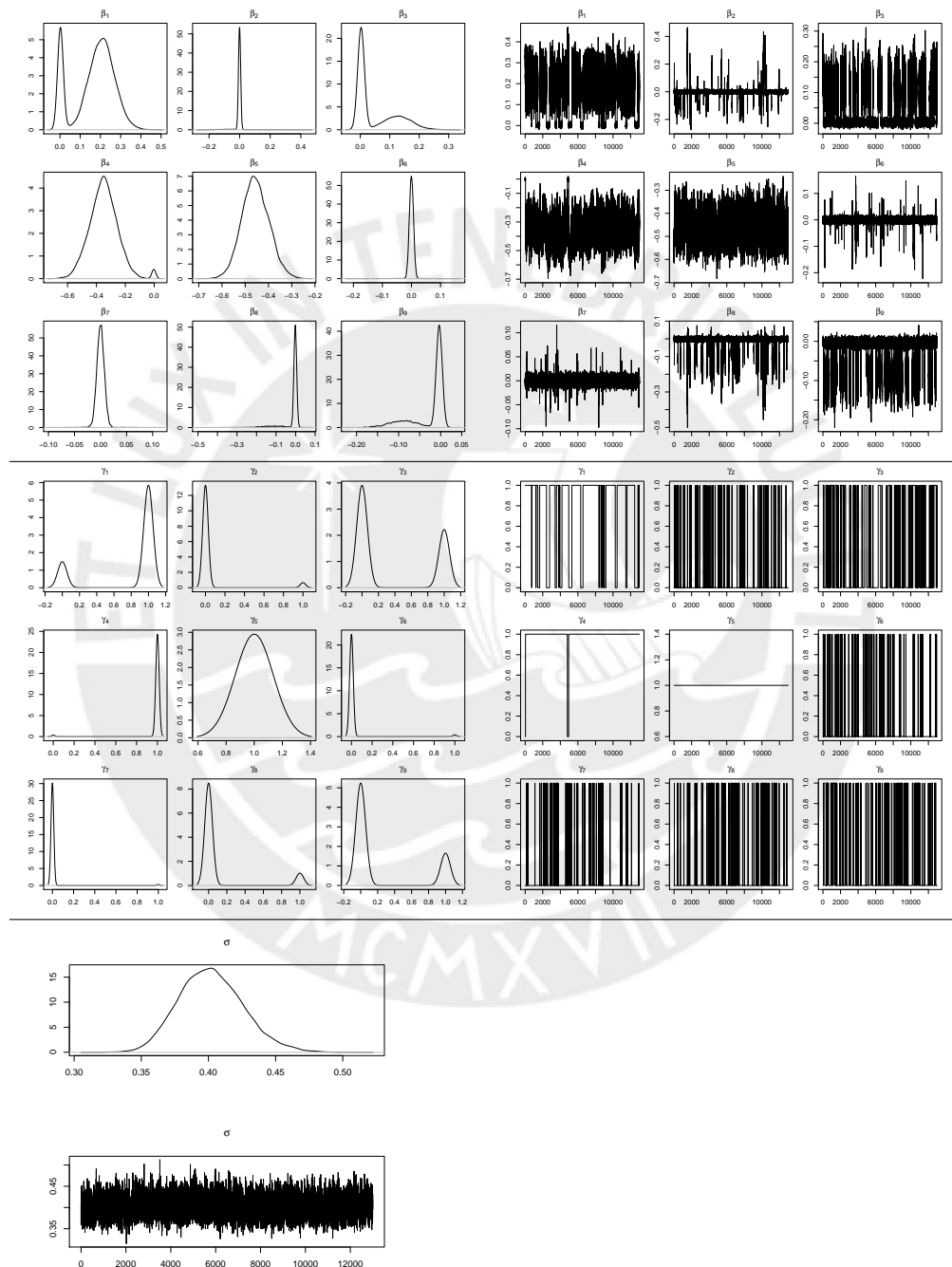


Figura B.13: Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros  $\beta$ ,  $\gamma$ ,  $\sigma$  empleando el método SSVS y considerando  $(\phi_j, C_j) = (0.1, 500)$ , en el estudio de la esperanza de vida al nacer.

Método KM

En la Figura B.14 se presentan los gráficos de las distribuciones a posteriori condicional y cadenas de las estimaciones de los parámetros  $\beta$ ,  $\gamma$  y  $\sigma$ ; aplicando la sugerencia de  $D_0 = 16I$  para la matriz de varianza y covarianza de la distribución a priori de  $\beta$ .

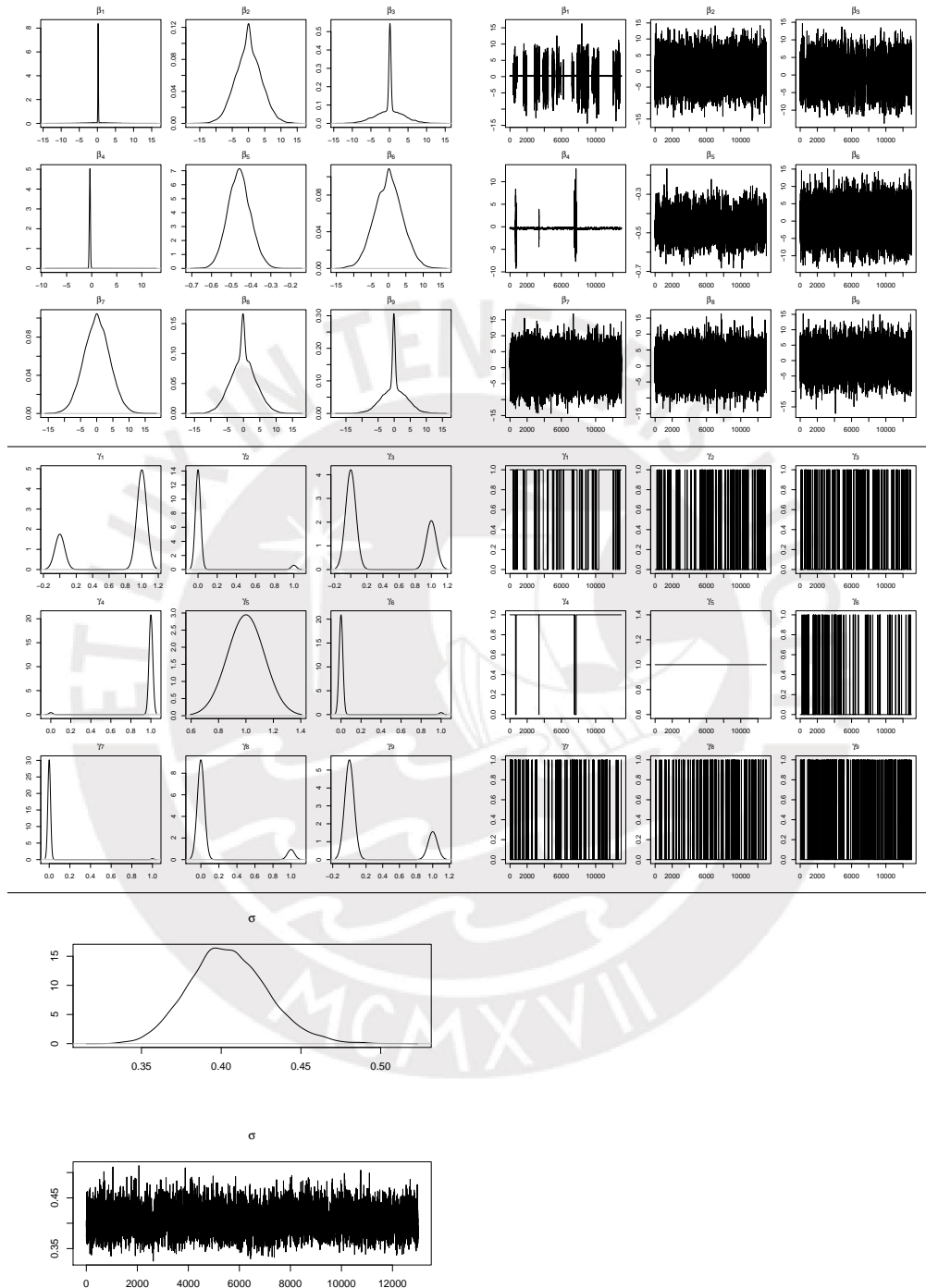


Figura B.14: Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros  $\beta$ ,  $\gamma$ ,  $\sigma$  empleando el método KM y considerando  $D_0 = 16I$ , en el estudio de la esperanza de vida al nacer.

Método GVS

En la Figura B.15 se presentan los gráficos de las distribuciones a posteriori condicional y cadenas de las estimaciones de los parámetros  $\beta$ ,  $\gamma$  y  $\sigma$ ; aplicando la sugerencia de  $\Sigma_j = 100$  para la varianza de la distribución a priori de los  $\beta_j$  seleccionados.

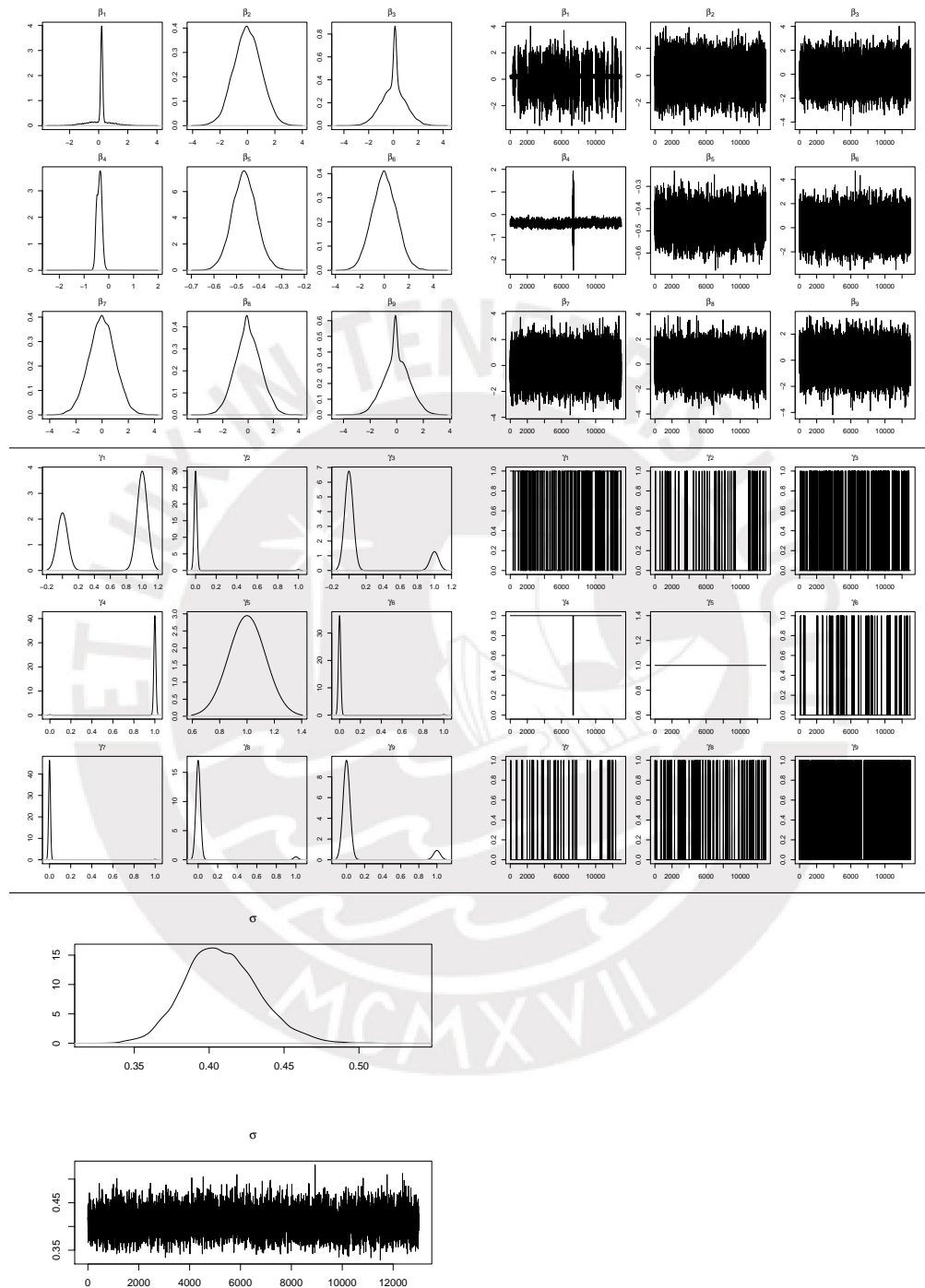


Figura B.15: Gráficos de las distribuciones a posteriori condicional y cadenas de los parámetros  $\beta$ ,  $\gamma$ ,  $\sigma$  empleando el método GVS y considerando  $\Sigma_j = 1000$ , en el estudio de la esperanza de vida al nacer.

## Bibliografía

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *Second international symposium on information theory*, *Académiai Kiadó, Budapest* pp. 267–281.
- Carpenter, B., Gelman, A., Hoffman, A., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P. y Riddell, A. (2016). Stan: A probabilistic programming language, *Journal of Statistical Software* **76**.
- Dellaportas, P., Forster, J. y I., N. (2002). On bayesian model and variable selection using mcmc, *Statistics and Computing* **12**: 27–36.
- Draper, N. y Smith, H. (1981). *Applied Regression Analysis*, 2nd edn, Wiley, New York.
- Efroymson, M. A. (1960). Multiple regression analysis, *Mathematical methods for digital computers* pp. 191–203.
- George, E. y McCulloch, R. (1993). Variable selection via gibbs sampling, *Journal of the American Statistical Association* **88**: 881–889.
- Huberty, C. J. (1989). Advances in social science methodology, *The American Statistician* **1**: 43–70.
- Jennrich, R. I. (1977). Stepwise regression, *Statistical methods for digital computers* **3**: 76–96.
- Kuo, L. y Mallick, B. (1998). Variable selection for regression models, *Sankhya B* **60**: 65–81.
- Ntzoufras, I. (2002). Gibbs variable selection using bugs, *Journal of Statistical Software* **7(7)**: 1–19.
- R Core Team (2016). R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria.  
**URL:** <https://www.r-project.org/>
- Ripley, B., Venables, B., Bates, D., Hornik, K., Gebhardt, A. y Firth, D. (2018). Mass: Support functions and datasets for venables and ripley’s mass, R package version 7.3-51.1.  
**URL:** <https://CRAN.R-project.org/package=MASS>
- Schwarz, G. (1978). Estimating the dimension of a model, *The Annals of Statistics* **6 (2)**: 461–464.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society B* **58**: 267–288.
- Tibshirani, R., Jerone, F. y Trevor, H. (2008). glmnet: Lasso and elastic-net regularized generalized linear models, R package version 2.0-10.  
**URL:** <https://CRAN.R-project.org/package=glmnet>
- UNDP (2017). Datos sobre el desarrollo humano (1980-2015).  
**URL:** <http://hdr.undp.org/es/data>