



Universidad Nacional de Colombia

FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA

CASO DE ESTUDIO III

Estadística Bayesiana
Prof. Juan Camilo Sosa Martínez

Autores

Ana Sofía Bello Dueñas	Germán Camilo Vásquez Herrera
Estadística	Estadística
abellod@unal.edu.co	gvasquez@unal.edu.co

Bogotá, 26 de Noviembre de 2023

Índice

Alcaldía de Bogotá 2023	2
Preguntas	3
Datos de diabetes	5
Selección de modelos	5
Modelamiento	5
Modelo 1: Regresión clásica previa unitaria	6
Modelo 2: Regresión clásica previa g	6
Modelo 3: Regresión rígida	6
Modelo 4: Regresión con errores correlacionados	6
Preguntas	7
Punto 1	7
Punto 2	8
Punto 3	9
Punto 4	9
Referencias	9
Apéndice	10
Alcaldía de Bogotá	10
Cadena de logverosimilitud	10
Coeficientes de variación de Monte Carlo	10
Distribuciones	10
Datos de diabetes	12
Cadenas de logverosimilitud	12
Coeficientes de variación de Monte Carlo	12
Distribuciones Modelo 1	13
Distribuciones Modelo 2	14
Distribuciones Modelo 3	15
Distribuciones Modelo 4	17

Alcaldía de Bogotá 2023

En esta publicación de La Silla Vacía se “revisó el desempeño de cada encuestadora que hizo mediciones en las cinco ciudades principales frente a los resultados” de las elecciones regionales de Colombia de 2023. De acuerdo con este medio, Invamer fue una de las encuestadoras con menor error total en Bogotá. En esta hoja de cálculo se pueden consultar algunos detalles técnicos incluyendo los márgenes de error de las encuestas.

Así, Invamer S.A.S realizó una encuesta del 17 al 23 de octubre de 2023 para Noticias Caracol, Blu Radio y El Espectador, con el fin de medir la intención de voto en Bogotá, Medellín, Cali, Barranquilla, y Bucaramanga para las elecciones de alcaldes de 2023. En Bogotá se encuestaron 1200 hombres y mujeres de 18 años en adelante, de todos los niveles socio-económicos a nivel nacional, aptos para votar en las elecciones y que sean residentes de la ciudad. Se realizaron encuestas personales en el hogar de los encuestados a través de tablets y para las preguntas de intención de voto se utilizó tarjetón. La ficha técnica se puede descargar aquí. Se obtuvieron los resultados que se presentan en la Tabla 1.

Candidato	Cantidad	Proporción
C. F. Galán	493	0.411
G. Bolívar	257	0.214
J. D. Oviedo	227	0.189
D. Molano	48	0.040
R. Lara	41	0.034
J. L. Vargas	38	0.032
J. E. Robledo	28	0.023
N. Ramos	11	0.009
R. A. Quintero	3	0.003
Voto en Blanco	54	0.045
Total	1200	1.000

Tabla 1: Si las elecciones para la Alcaldía de Bogotá fueran mañana, y los candidatos fueran los que aparecen en este tarjetón, ¿por cuál de ellos votaría usted?

Aunque Invamer utilizó una clase particular de muestreo aleatorio sin reemplazo (ver Ficha Técnica), es posible considerar la muestra como una muestra aleatoria simple con reemplazo (IID), dado que el tamaño total de la muestra es muy pequeño en comparación con el tamaño del Universo. El principal interés científico y político en este problema se centra en estimar la proporción poblacional de votos que recibirá cada candidato de acuerdo con los datos proporcionados por Invamer.

Bajo las condiciones dadas anteriormente, dado que nuestra incertidumbre acerca de las respuestas de las 1200 personas en la encuesta es intercambiable, una versión particular del Teorema de De Finetti (Bernardo and Smith, 2000, pág. 176) garantiza que la única distribución muestral apropiada para datos de esta índole es la distribución multinomial.

Suponga que una población de interés tiene artículos de $k \geq 2$ tipos, y además, que la proporción de artículos de tipo j es $0 < \theta_j < 1$, para $j = 1, \dots, k$. Siendo $\theta = (\theta_1, \dots, \theta_k)$, note que las componentes de $\theta = (\theta_1, \dots, \theta_k)$ son tales que $\sum_{j=1}^k \theta_j = 1$. Ahora, suponga que se toma una muestra IID $\mathbf{y} = (y_1, \dots, y_n)$

de tamaño n de la población. Sea $\mathbf{n} = (n_1, \dots, n_k)$ el vector aleatorio que almacena los conteos asociados con cada tipo de artículo, así que n_j es el número de elementos en la muestra aleatoria de tipo j , para $j = 1, \dots, k$. En esta situación, se dice que n sigue una distribución multinomial con parámetros n y θ , la cual se define como sigue: $n|n, \theta \sim \text{Multinomial}(n, \theta)$ si y solo si

$$p(\mathbf{n}|n, \theta) = \frac{n!}{\prod_{j=1}^k n_j!} \prod_{j=1}^k \theta_j^{n_j}$$

siempre que $\sum_{j=1}^k n_j = n$ y $0 \leq n_j \leq n$ para todo $j = 1, \dots, k$.

Considere el modelo con distribución muestral $\mathbf{n}|n, \theta \sim \text{Multinomial}(n, \theta)$ y distribución previa jerárquica dada por

$$\theta|\alpha \sim \text{Dirichlet}(\alpha \mathbf{1}_k) \quad \text{y} \quad \alpha \sim \text{Gamma}(a, b)$$

donde $\mathbf{1}_k$ es el vector de unos de $k \times 1$ y a y b son los hiperparámetros del modelo (en el Cap. 8 de Gelman et. al (2013) se discuten los modelos más sofisticados que consideran la estrategia de muestreo).

Preguntas

Ajustar el modelo propuesto usando un muestreador de Gibbs con $a = b = 1$ (incluir un anexo con todos los detalles). Reportar visual y tabularmente las estimaciones puntuales, los intervalos de credibilidad al 95 % y los resultados oficiales de la Registraduría Nacional del Estado Civil para Galán, Bolívar y Oviedo, expresando todas las cifras en puntos porcentuales. Interpretar los resultados obtenidos (máximo 500 palabras).

Solución.

Prop. votos candidatos Bogotá

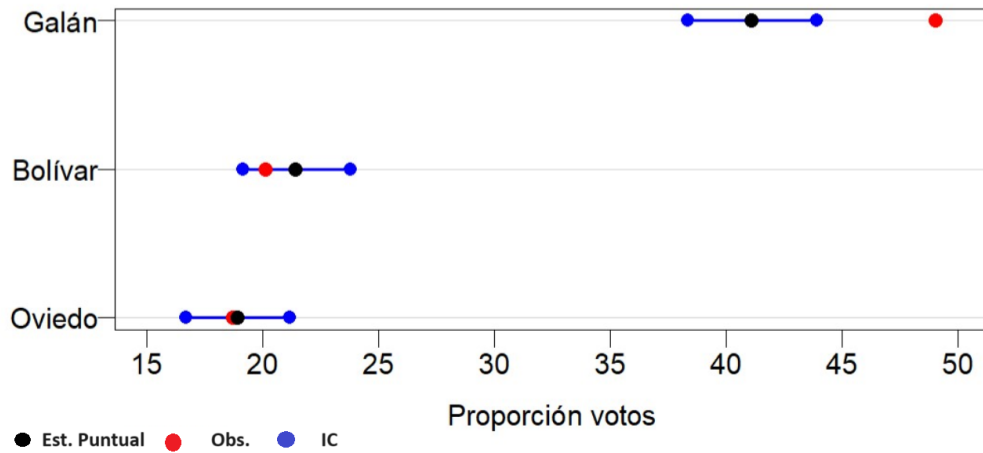


Figura 1: Estimación puntual, intervalo de credibilidad y valor observado de los porcentajes de votos de Galán, Bolívar y Oviedo

Candidato	R.O. Registraduría	Est. Puntual	L.I (2.5 %)	L.S. (97.5 %)
Galán	49.02 %	40.912 %	38.097 %	43.643 %
Bolívar	20.11 %	21.358 %	19.093 %	23.666 %
Oviedo	18.71 %	18.868 %	16.677 %	21.122 %

Tabla 2: Resultados de la registraduría, estimaciones puntuales e intervalos de credibilidad del porcentaje de votos para Galán, Bolívar y Oviedo

Asumir una distribución Dirichlet para cada proporción con un α igual para todas las proporciones indica que no hay preferencia o sesgo por algún candidato, todos tienen igual probabilidad de ser elegidos.

En general, aunque se asumió una previa no informativa los resultados son muy similares a los que obtuvo previamente Invamer S.A.S en su encuesta, también son muy similares a los resultados de la Registraduría, excepto por el resultado de Galán que dista casi en un 10 %.

Teniendo en cuenta esto, según el modelo propuesto, aproximadamente el 40.912 % de los votos serían para Galán, 21.358 % para Bolívar y 18.868 % para Oviedo.

Datos de diabetes

Considere la base de datos de diabetes dada en la Sección 9.3 de Hoff (2009, p. 161), que contiene datos asociados con 10 medidas basales x_1, \dots, x_{10} en un grupo de 442 pacientes diabéticos, así como una medida de progresión de la enfermedad y tomada un año después de las medidas basales. Los datos se pueden descargar de este enlace en `yX.diabetes.train` y `yX.diabetes.test`. A partir de estos datos, el objetivo es hacer un modelo predictivo para y y basado en x_1, \dots, x_{10} (tanto y como las x_j se encuentran estandarizadas). Si bien un modelo de regresión con diez variables no sería abrumadoramente complejo, se sospecha que la relación entre y y las x_j puede no ser lineal, así que se recomienda considerar términos de segundo orden de la forma x_j^2 y $x_j x_k$ para potenciar la capacidad predictiva del modelo. Así, las variables regresoras incluyen diez efectos principales x_j , $\binom{10}{2} = 45$ interacciones $x_j x_k$ y nueve términos cuadráticos x_j^2 (no es necesario considerar x_2^2 en el modelo porque $x_2 = \text{sexo}$ es binaria, y por lo tanto $x_2 = x_2^2$). Esto da un total de $p = 64$ variables regresoras (no es necesario considerar el intercepto porque todas las variables se encuentran estandarizadas).

Selección de modelos

Puede ocurrir que en un análisis de regresión haya un gran número de variables independientes x , aunque puede que la mayoría de estas variables no tengan una relación sustancial con la variable dependiente y . En estas situaciones, incluir todas las variables regresoras en el modelo de regresión conduce a modelos saturados poco parsimoniosos difíciles de interpretar con un rendimiento deficiente. Por lo tanto, se recomienda considerar en el modelo final solo aquellas variables x para las que exista evidencia sustancial de una asociación con y . Esto no solo produce análisis de datos más simples, sino que también proporciona modelos con mejores propiedades estadísticas en términos de predicción y estimación.

Modelamiento

Se considera un modelo de regresión de la forma $\mathbf{y}|\mathbf{X}, \boldsymbol{\beta}, \sigma^2 \sim \mathbf{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$, donde \mathbf{y} es un vector de $n \times 1$ que contiene los valores de la variable respuesta, \mathbf{X} es una matriz de $n \times p$ que contiene los valores de las variables regresoras, $\boldsymbol{\beta}$ es un vector de $p \times 1$ que contiene los parámetros desconocidos, y finalmente, \mathbf{I}_n es la matriz identidad de $n \times n$.

Para evaluar los modelos de regresión, se dividieron aleatoriamente a los 442 individuos con diabetes en 342 individuos de entrenamiento y 100 individuos de prueba, lo que provee un conjunto de datos de entrenamiento $(\mathbf{y}_{\text{train}}, \mathbf{X}_{\text{train}})$ y un conjunto de datos de prueba $(\mathbf{y}_{\text{test}}, \mathbf{X}_{\text{test}})$. Así, se ajustan los modelos usando $(\mathbf{y}_{\text{train}}, \mathbf{X}_{\text{train}})$, y luego, usando los coeficientes de regresión estimados $\hat{\boldsymbol{\beta}} = E(\boldsymbol{\beta}|\mathbf{y}_{\text{train}})$, se genera $\hat{\mathbf{y}}_{\text{test}} = \mathbf{X}_{\text{test}}\hat{\boldsymbol{\beta}}$. Luego, se evalúa el rendimiento predictivo del modelo comparando $\hat{\mathbf{y}}_{\text{test}}$ con \mathbf{y}_{test} por medio de una métrica apropiada.

Modelo 1: Regresión clásica previa unitaria

Distribución previa: Previa unitaria (*unit information prior*; Kass y Wasserman, 1995).

Modelo 2: Regresión clásica previa g

Distribución previa: Previa g (*g-prior* ; Zellner, 1986).

Modelo 3: Regresión rígida

Distribución previa:

$$p(\boldsymbol{\beta}, \sigma^2, \lambda) = \mathbf{N}_p \left(\boldsymbol{\beta} \mid \mathbf{0}_p, \frac{\sigma^2}{\lambda} \mathbf{I}_p \right) \times \text{GI} \left(\sigma^2 \mid \frac{\nu_0}{2}, \frac{\nu_0 \sigma_0}{2} \right) \times \text{G}(\lambda \mid a_\lambda, b_\lambda)$$

con $\nu_0 = 1$, $\sigma_0^2 = \hat{\sigma}_{OLS}^2$, $a_\lambda = 1$ y $b_\lambda = 2$.

Modelo 4: Regresión con errores correlacionados

Distribución muestral:

$$\mathbf{y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2, \rho \sim \text{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{C}_\rho)$$

donde \mathbf{C}_ρ es una matriz con estructura autorregresiva de primer orden de la forma

$$\mathbf{C}_\rho = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix} \quad (1)$$

Distribución previa:

$$p(\boldsymbol{\beta}, \sigma^2, \rho) = \prod_{j=1}^p \text{N}(\beta_j \mid 0, \tau_0^2) \times \text{GI}(\sigma^2 \mid \nu_0/2, \nu_0 \sigma_0^2) \times \text{U}(\rho \mid a_\rho, b_\rho)$$

con $\tau_0^2 = 50$, $\nu_0 = 1$, $\sigma_0^2 = \sigma_{OLS}^2$, $a_\rho = 0$ y $b_\rho = 1$.

Preguntas

Punto 1

Para cada modelo, generar $\hat{\mathbf{y}}_{\text{test}} = \mathbf{X}_{\text{test}}\hat{\boldsymbol{\beta}}$ usando los coeficientes de regresión estimados $\hat{\boldsymbol{\beta}} = E(\boldsymbol{\beta}|\hat{\mathbf{y}}_{\text{train}})$. Graficar \hat{y}_{test} frente y_{test} y calcular el error absoluto medio $\frac{1}{n} \sum_i |y_{\text{test},i} - \hat{y}_{\text{test},i}|$ correspondiente.

Nota:

- Incluir todos los gráficos en una sola figura con 2×2 paneles (todos los paneles deben tener la misma escala en los ejes).
- En cada gráfico, superponer la recta $\hat{y}_{\text{test}} = y_{\text{test}}$ como punto de referencia (ver el segundo panel de la Figura 9.7 de Hoff 2009, p. 170).
- En el encabezado de cada gráfica, incluir el error absoluto medio usando tres (3) cifras decimales

Solución.

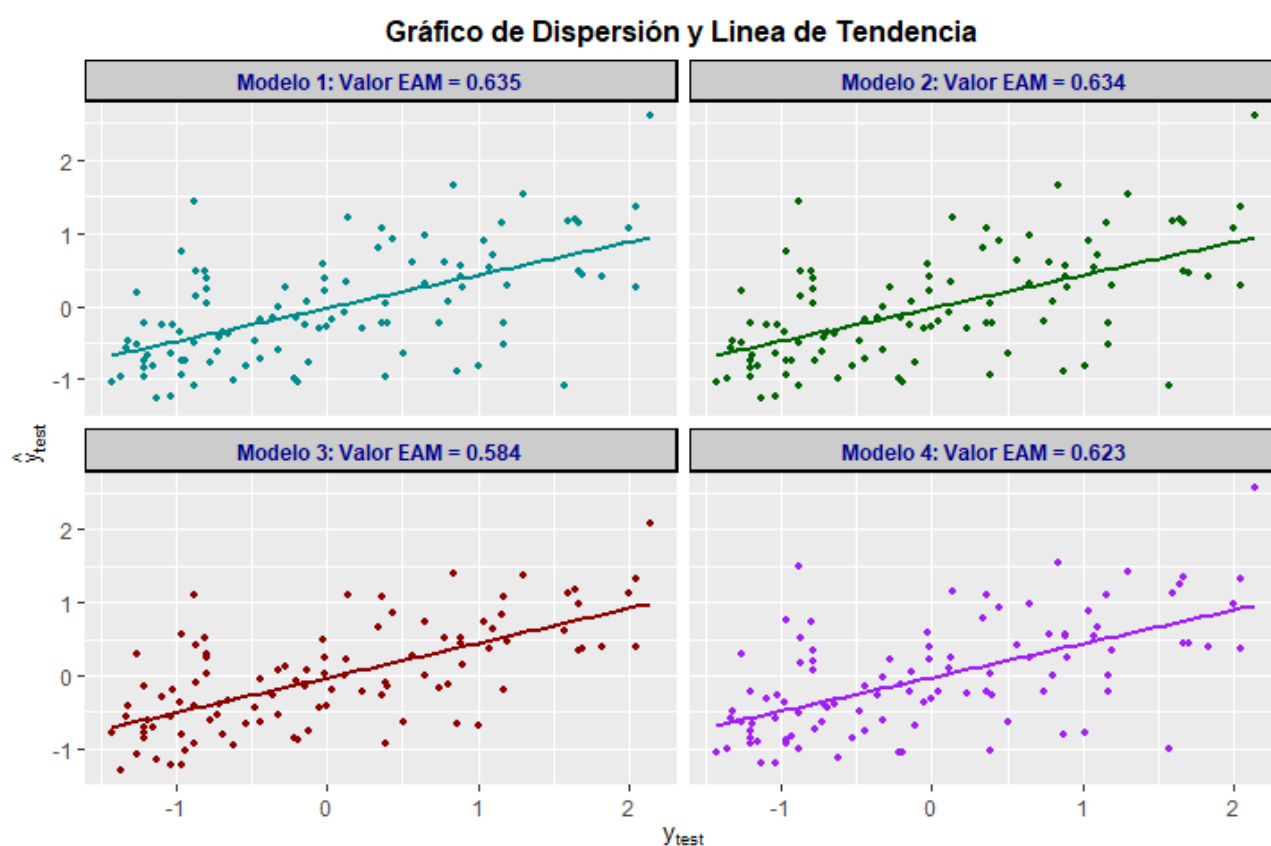


Figura 2: Gráficos de dispersión de valores observados frente a las estimaciones en los datos de prueba

Punto 2

Para cada modelo, chequear la bondad de ajuste usando la media como estadístico de prueba. Graficar la distribución predictiva posterior por medio de un histograma.

Nota:

- Incluir todos los gráficos en una sola figura con 2×2 paneles (todos los paneles deben tener la misma escala en los ejes)
- En cada gráfico, superponer el valor observado de la media como punto de referencia.
- En el encabezado de cada gráfica, incluir el valor p predictivo posterior correspondiente usando tres (3) cifras decimales.

Solución.

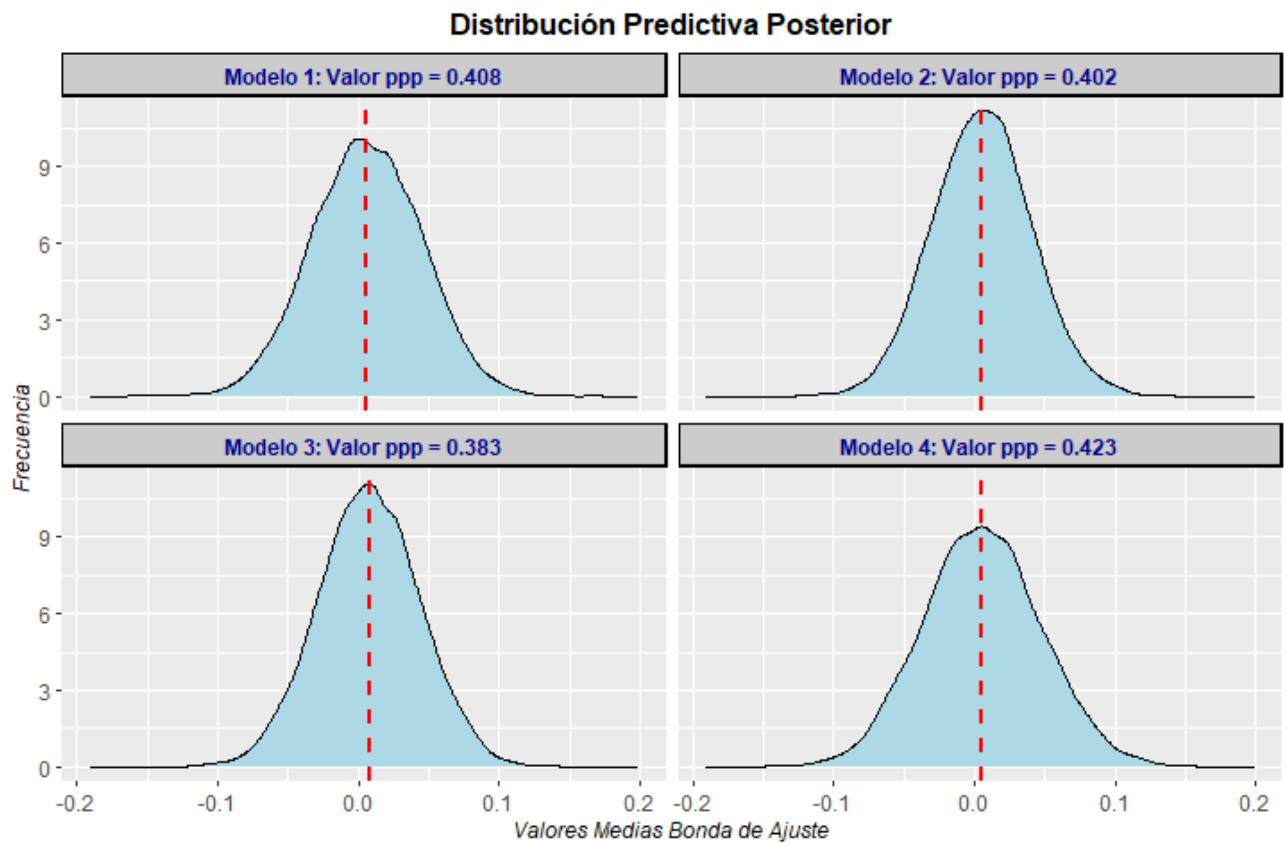


Figura 3: Bondad de ajuste de los modelos

Punto 3

Para cada modelo, calcular el DIC. Presentar los resultados tabularmente usando tres (3) cifras decimales.

Solución.

Modelo	Modelo 1	Modelo 2	Modelo 3	Modelo 4
DIC	784.833	777.773	759.527	782.023

Tabla 3: DIC de los modelos

Punto 4

Interpretar los resultados obtenidos en los numerales anteriores (máximo 500 palabras).

Solución.

La bondad de ajuste muestra que los modelos, en general, presenta un buen comportamiento. La media real se encuentra donde hay mayor densidad de las medias simuladas por los modelos, además, los *ppp*'s toman valores que no son ni muy altos ni muy bajos, es decir, los modelos no subestiman ni sobrestiman la media.

En cuanto a los gráficos de dispersión, todos los modelos parecen dar buenas estimaciones de la variable respuesta y , sin embargo, el modelo de regresión rígida es quien presenta menor error absoluto medio, es decir, da las mejores estimaciones. Esto último también podemos confirmarlo con el valor del DIC que resulta ser más pequeño para el modelo 3 (regresión rígida).

Por lo tanto, se recomienda usar el modelo de regresión rígida para predecir y basado en x_1, \dots, x_{10} teniendo en cuenta sus interacciones y respectivos términos cuadráticos.

Referencias

Sosa, J. C., and Aristizabal, J. P. (2022). Some Developments in Bayesian Hierarchical Linear Regression Modeling: Algunos desarrollos en modelos de regresión lineal jerárquicos Bayesianos. *Revista Colombiana de Estadística -Applied Statistics*, 45(2), 231-255. <http://doi.org/10.15446/rce.v45n2.98988>

Apéndice

Alcaldía de Bogotá

Cadena de log-verosimilitud

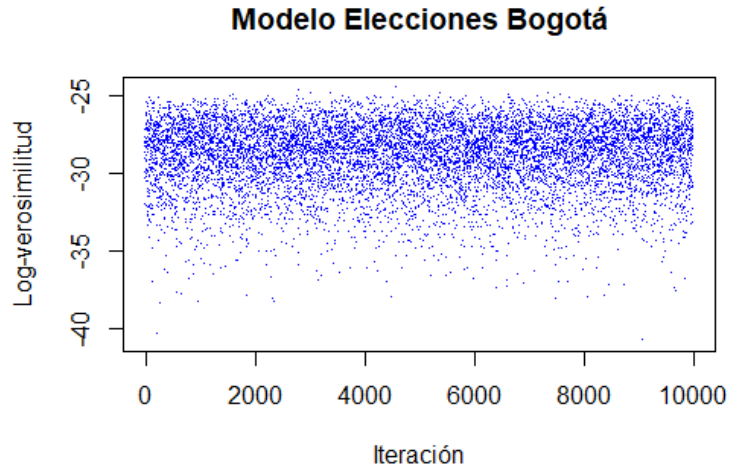


Figura 4: Cadena de log-verosimilitud modelo de Elecciones en Bogotá

Coefficientes de variación de Monte Carlo

Estadístico	CV MC (%)	Estadístico	CV MC
θ_1	0.0337 %	θ_6	0.1557 %
θ_2	0.0553 %	θ_7	0.1847 %
θ_3	0.0599 %	θ_8	0.2919 %
θ_4	0.1391 %	θ_9	0.5256 %
θ_5	0.1518 %	θ_{10}	0.1264 %
α	0.3798 %		

Tabla 4: Coeficientes de variación MC modelo Elecciones en Bogotá

Distribuciones

Distribución posterior

Sea $\Theta = (\theta_1, \dots, \theta_{10}, \alpha)$, entonces

$$p(\boldsymbol{\Theta}|\mathbf{n}) \propto \text{Multinomial}(\mathbf{n} \mid \boldsymbol{\theta}, n) \times \text{Dirichlet}(\boldsymbol{\theta} \mid \alpha \mathbf{1}_k) \times \text{Gamma}(\alpha \mid a, b)$$

Distribuciones condicionales completas

- Para $\boldsymbol{\theta}$

$$\begin{aligned} p(\boldsymbol{\theta}|-) &\propto \text{Multinomial}(\mathbf{n} \mid \boldsymbol{\theta}, n) \times \text{Dirichlet}(\boldsymbol{\theta} \mid \alpha \mathbf{1}_k) \\ &\propto \prod_{j=1}^k \theta_j^{n_j} \times \prod_{j=1}^k \theta_j^{\alpha-1} \\ &= \prod_{j=1}^k \theta_j^{n_j+\alpha-1} \end{aligned}$$

entonces

$$\boldsymbol{\theta}|- \sim \text{Dirichlet}(\mathbf{n} + \alpha \mathbf{1}_k)$$

Dado que $\boldsymbol{\theta}|-$ presenta una distribución cerrada, entonces se obtuvo muestras de dicha distribución aplicando el muestreador de Gibbs, con 1000 muestras de calentamiento y realizando un muestreo sistemático cada 10 muestras. De modo que, dados $\{\boldsymbol{\theta}^{(b)}, \alpha^{(b)}\}$ se actualiza $\boldsymbol{\theta}$ muestreando $\boldsymbol{\theta}^{(b+1)}$ de una distribución $\text{Dirichlet}(\mathbf{n} + \alpha^{(b)} \mathbf{1}_k)$.

- Para α

$$\begin{aligned} p(\alpha|-) &\propto \text{Dirichlet}(\boldsymbol{\theta} \mid \alpha \mathbf{1}_k) \times \text{Gamma}(\alpha \mid a, b) \\ &\propto \frac{\Gamma(k\alpha)}{\Gamma(\alpha)^k} \prod_{j=1}^k \theta_j^{\alpha-1} \times \alpha^{a-1} \exp\{-b\alpha\} \\ &= \frac{\Gamma(k\alpha)}{\Gamma(\alpha)^k} \prod_{j=1}^k \theta_j^{\alpha} \times \alpha^{a-1} \exp\{-b\alpha\} \end{aligned}$$

la cual no tiene una forma cerrada, por lo que procedió a utilizar el algoritmo de Metrópolis para encontrar muestras de dicha distribución. Para actualizar $\alpha^{(b)}$ por medio de Metrópolis se requiere el estado actual de $\boldsymbol{\theta}^{(b+1)}$. Se utilizó un valor de $\delta^2 = 0.75$ y se obtuvo una tasa de aceptación del 45.19%

Datos de diabetes

Cadenas de log-verosimilitud

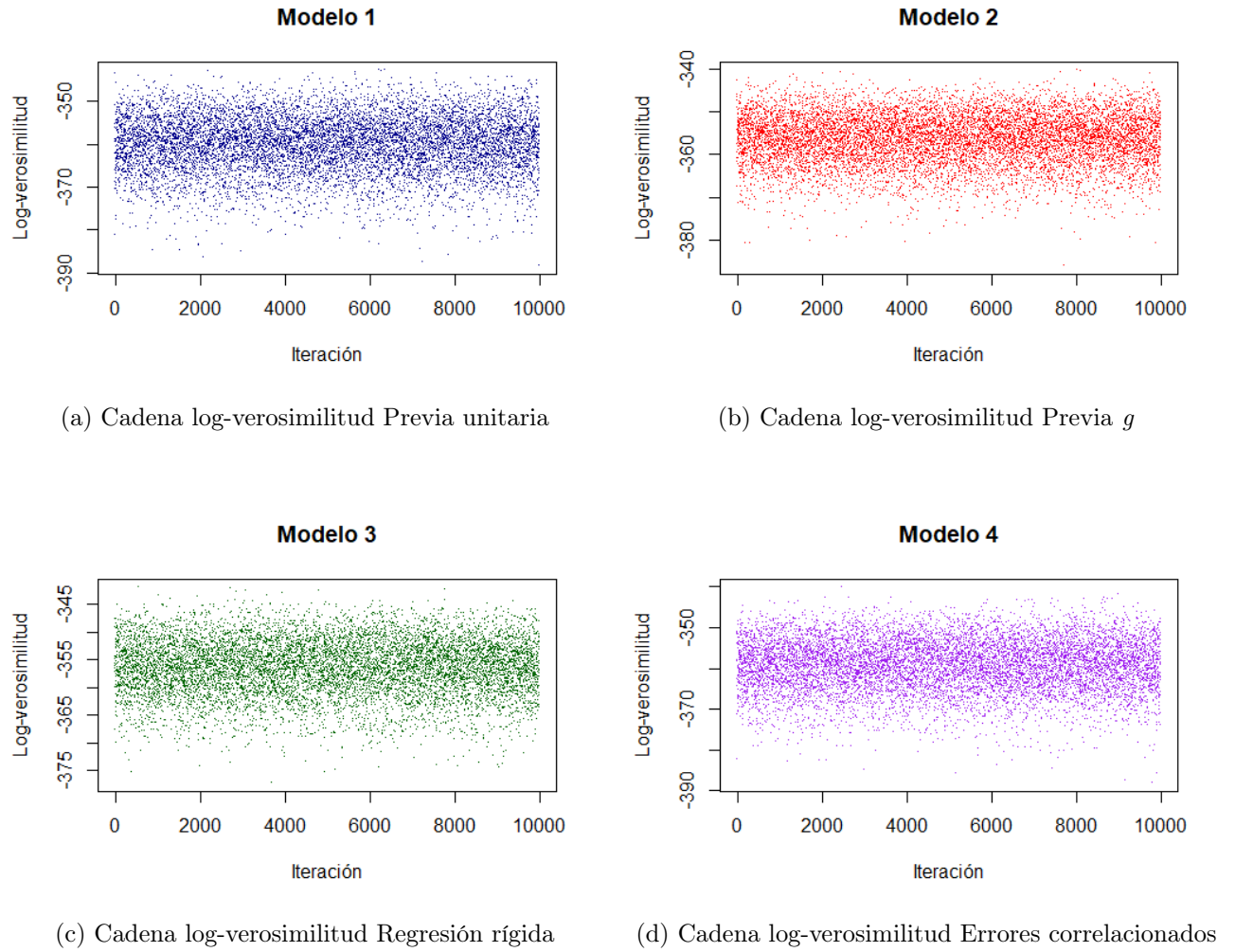


Figura 5: Cadenas de log-verosimilitud de los modelos

Coefficientes de variación de Monte Carlo

Modelo	Min.	Cuartil 1	Mediana	Media	Cuartil 3	Máy.
Previa unitaria	0.2217 %	0.9613 %	1.2675 %	2.5073 %	2.2100 %	30.6384 %
Previa g	0.2043 %	0.8353 %	1.1522 %	2.2713 %	1.9129 %	21.9721 %
Rígida	0.1838 %	0.9454 %	1.9944 %	2.9579 %	2.4867 %	26.7263 %
E. Correlacionados	0.2194 %	0.8779 %	1.3474 %	2.3067 %	2.0077 %	23.3419 %

Tabla 5: Resumen coeficientes de variación de MC para los 64 β 's de los modelos

Modelo	σ^2	λ	ρ
Previa unitaria	0.0863 %	-	-
Previa g	0.0766 %	-	-
Rígida	0.0774 %	0.1718 %	-
E. Correlacionados	0.0867 %	-	0.6620 %

Tabla 6: Coeficientes de variación de MC para los demás parámetros de los modelos

Distribuciones Modelo 1

Distribución posterior

Sea $\Theta = (\beta, \sigma^2)$, entonces

$$p(\Theta|\mathbf{y}) \propto \mathbf{N}_n(\mathbf{y} \mid \mathbf{X}\beta, \sigma^2 \mathbf{I}_n) \times \mathbf{N}_p(\beta \mid \beta_0, \Sigma_0) \times \text{GI}\left(\sigma^2 \mid \frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\right)$$

con $\beta_0 = \hat{\beta}_{OLS} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$, $\Sigma_0 = n \hat{\sigma}_{OLS}^2 (\mathbf{X}^T \mathbf{X})^{-1}$, $\nu_0 = 1$ y $\sigma^2 = \hat{\sigma}_{OLS}^2$

Distribuciones condicionales completas

- Para β

$$\begin{aligned} p(\beta| -) &\propto \mathbf{N}_n(\mathbf{y} \mid \mathbf{X}\beta, \sigma^2 \mathbf{I}_n) \times \mathbf{N}_p(\beta \mid \beta_0, \Sigma_0) \\ &\propto \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta)\right\} \times \exp\left\{-\frac{1}{2}(\beta - \beta_0)^T \Sigma_0^{-1}(\beta - \beta_0)\right\} \\ &\propto \exp\left\{-\frac{1}{2\sigma^2}(\beta^T \mathbf{X}^T \mathbf{X} \beta - 2\beta^T \mathbf{X}^T \mathbf{y})\right\} \times \exp\left\{-\frac{1}{2}(\beta^T \Sigma_0^{-1} \beta - 2\beta^T \Sigma_0^{-1} \beta_0)\right\} \\ &= \exp\left\{-\frac{1}{2}(\beta^T (\Sigma_0^{-1} + \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{X}) \beta - 2\beta^T (\Sigma_0^{-1} \beta_0 + \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{y}))\right\} \end{aligned}$$

entonces

$$\beta| - \sim \mathbf{N}_p\left(\left(\Sigma_0^{-1} + \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{X}\right)^{-1} \left(\Sigma_0^{-1} \beta_0 + \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{y}\right), \left(\Sigma_0^{-1} + \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{X}\right)^{-1}\right)$$

- Para σ^2

$$\begin{aligned} p(\sigma^2 | -) &\propto \mathbf{N}_n(\mathbf{y} | \mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n) \times \text{GI} \left(\sigma^2 \mid \frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2} \right) \\ &\propto (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \times (\sigma^2)^{-(\frac{\nu_0}{2}+1)} \exp \left\{ -\frac{\nu_0 \sigma_0^2 / 2}{\sigma^2} \right\} \\ &= (\sigma^2)^{-\left(\frac{n+\nu_0}{2}+1\right)} \exp \left\{ -\frac{1}{\sigma^2} \left(\frac{\nu_0 \sigma_0^2 + \text{SSE}(\boldsymbol{\beta})}{2} \right) \right\} \end{aligned}$$

donde $\text{SSE}(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$, entonces

$$\sigma^2 | - \sim \text{GI} \left(\frac{n + \nu_0}{2}, \frac{\nu_0 \sigma_0^2 + \text{SSE}(\boldsymbol{\beta})}{2} \right)$$

Dado que las dos distribuciones condicionales completas tienen forma cerrada, se obtienen muestras de la distribución posterior utilizando el muestreador de Gibbs, con 1000 muestras de calentamiento y realizando un muestreo sistemático cada 10 muestras.

Entonces, dados $\{\boldsymbol{\beta}^{(b)}, \sigma^{2(b)}\}$:

- Actualizar $\boldsymbol{\beta}$: Muestreamos $\boldsymbol{\beta}^{(b+1)}$ de su distribución condicional completa (Normal p -variada) usando el estado actual de $\sigma^{2(b)}$.
- Actualizar σ^2 : Muestreamos $\sigma^{2(b+1)}$ de su distribución condicional completa (Gamma Inversa) usando el estado actual de $\boldsymbol{\beta}^{(b+1)}$.

Distribuciones Modelo 2

Distribución posterior

Sea $\boldsymbol{\Theta} = (\boldsymbol{\beta}, \sigma^2)$, entonces

$$p(\boldsymbol{\Theta} | \mathbf{y}) \propto \mathbf{N}_n(\mathbf{y} | \mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n) \times \mathbf{N}_p(\boldsymbol{\beta} | \boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0) \times \text{GI} \left(\sigma^2 \mid \frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2} \right)$$

con $\boldsymbol{\beta}_0 = \mathbf{0}_p$ y $\boldsymbol{\Sigma}_0 = g\sigma^2(\mathbf{X}^T \mathbf{X})^{-1}$.

Distribución condicional completa de $\boldsymbol{\beta}$

Para $\boldsymbol{\beta}$ el procedimiento es exactamente el mismo que el del Modelo 1, teniendo en cuenta que $\boldsymbol{\beta}_0 = \mathbf{0}_p$ y $\boldsymbol{\Sigma}_0 = g\sigma^2(\mathbf{X}^T \mathbf{X})^{-1}$, entonces

$$\begin{aligned}
\beta| - &\sim \mathbf{N}_p \left(\left(\Sigma_0^{-1} + \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{X} \right)^{-1} \left(\Sigma_0^{-1} \beta_0 + \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{y} \right), \left(\Sigma_0^{-1} + \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{X} \right)^{-1} \right) \\
&\sim \mathbf{N}_p \left(\left(\frac{1}{g\sigma^2} \mathbf{X}^T \mathbf{X} + \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{X} \right)^{-1} \left(\frac{1}{\sigma^2} \mathbf{X}^T \mathbf{y} \right), \left(\frac{1}{g\sigma^2} \mathbf{X}^T \mathbf{X} + \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{X} \right)^{-1} \right) \\
&\sim \mathbf{N}_p \left(\left(\frac{1+g}{g\sigma^2} (\mathbf{X}^T \mathbf{X}) \right)^{-1} \left(\frac{1}{\sigma^2} \mathbf{X}^T \mathbf{y} \right), \left(\frac{1+g}{g\sigma^2} (\mathbf{X}^T \mathbf{X}) \right)^{-1} \right) \\
&\sim \mathbf{N}_p \left(\frac{g}{1+g} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}, \frac{g}{g+1} \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \right)
\end{aligned}$$

Dado que la distribución condicional completa tiene forma cerrada, se obtiene muestra de dicha distribución utilizando el muestreador de Gibbs, con 1000 muestras de calentamiento y realizando un muestreo sistemático cada 10 muestras. Entonces, dados $\beta^{(b)}$ actualizamos β muestreando $\beta^{(b+1)}$ de su distribución condicional completa (Normal p -variada) usando el valor de σ_j^2 tal que $j = b$.

Distribución posterior de σ^2

$$\begin{aligned}
p(\sigma^2 | \mathbf{X}, \mathbf{y}) &\propto p(\mathbf{y} | \sigma^2, \mathbf{X}) \times p(\sigma^2) \\
&= \int_{\mathbb{R}^p} p(\mathbf{y} | \beta, \sigma^2, \mathbf{X}) p(\beta | \sigma^2, \mathbf{X}, \mathbf{y}) \cdot d\beta \times p(\sigma^2) \\
&\propto \text{GI} \left(\sigma^2 \left| \frac{\nu_0 + n}{2}, \frac{\nu_0 \sigma_0^2 + \text{SSR}_g}{2} \right. \right)
\end{aligned}$$

$$\text{con } \text{SSR}_g = \mathbf{y}^T \left(\mathbf{I}_n - \frac{g}{g+1} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right).$$

Al obtener la distribución posterior, se obtienen finalmente $j = 10000$ valores de σ^2 , los cuales se utilizan para actualizar β .

Distribuciones Modelo 3

Distribución posterior

Sea $\Theta = (\beta, \sigma^2, \lambda)$, entonces

$$p(\Theta | \mathbf{y}) \propto \mathbf{N}_n(\mathbf{y} | \mathbf{X}\beta, \sigma^2 \mathbf{I}_n) \times \mathbf{N}_p \left(\beta \left| \mathbf{0}_p, \frac{\sigma^2}{\lambda} \mathbf{I}_p \right. \right) \times \text{GI} \left(\sigma^2 \left| \frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2} \right. \right) \times \text{G}(\lambda | a_\lambda, b_\lambda)$$

Distribuciones condicionales completas

■ Para β

$$\begin{aligned}
p(\beta|-) &\propto \mathbf{N}_n(\mathbf{y} \mid \mathbf{X}\beta, \sigma^2 \mathbf{I}_n) \times \mathbf{N}_p\left(\beta \mid \mathbf{0}_p, \frac{\sigma^2}{\lambda} \mathbf{I}_p\right) \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta)\right\} \times \exp\left\{-\frac{\lambda}{2\sigma^2}\beta^T\beta\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}(\beta^T \mathbf{X}^T \mathbf{X} \beta - 2\beta^T \mathbf{X}^T \mathbf{y})\right\} \times \exp\left\{-\frac{\lambda}{2\sigma^2}\beta^T\beta\right\} \\
&= \exp\left\{-\frac{1}{2\sigma^2}(\beta^T (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I}_p) \beta - 2\beta^T \mathbf{X}^T \mathbf{y})\right\}
\end{aligned}$$

entonces

$$\beta|- \sim \mathbf{N}_p\left(\left(\frac{1}{\sigma^2}(\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I}_p)\right)^{-1} \left(\frac{1}{\sigma^2} \mathbf{X}^T \mathbf{y}\right), \left(\frac{1}{\sigma^2}(\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I}_p)\right)^{-1}\right)$$

■ Para σ^2

$$\begin{aligned}
p(\sigma^2|-) &\propto \mathbf{N}_n(\mathbf{y} \mid \mathbf{X}\beta, \sigma^2 \mathbf{I}_n) \times \text{GI}\left(\sigma^2 \mid \frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\right) \times \mathbf{N}_p\left(\beta \mid \mathbf{0}_p, \frac{\sigma^2}{\lambda} \mathbf{I}_p\right) \\
&\propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta)\right\} \times (\sigma^2)^{-(\frac{\nu_0}{2}+1)} \exp\left\{-\frac{\nu_0 \sigma_0^2/2}{\sigma^2}\right\} \\
&\quad \times (\sigma^2)^{-\frac{p}{2}} \exp\left\{-\frac{\lambda}{2\sigma^2}\beta^T\beta\right\} \\
&= (\sigma^2)^{-\left(\frac{n+p+\nu_0}{2}+1\right)} \exp\left\{-\frac{1}{\sigma^2} \left(\frac{\nu_0 \sigma_0^2 + \lambda \beta^T \beta + \text{SSE}(\beta)}{2}\right)\right\}
\end{aligned}$$

donde $\text{SSE}(\beta) = (\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta)$, entonces

$$\sigma^2|- \sim \text{GI}\left(\frac{n+p+\nu_0}{2}, \frac{\nu_0 \sigma_0^2 + \lambda \beta^T \beta + \text{SSE}(\beta)}{2}\right)$$

■ Para λ

$$\begin{aligned}
p(\lambda|-) &\propto \mathbf{N}_p\left(\beta \mid \mathbf{0}_p, \frac{\sigma^2}{\lambda} \mathbf{I}_p\right) \times \text{G}(\lambda \mid a_\lambda, b_\lambda) \\
&\propto (\lambda^{-1})^{-\frac{p}{2}} \exp\left\{-\frac{\lambda}{2\sigma^2}\beta^T\beta\right\} \times \lambda^{(a_\lambda-1)} \exp\{-b_\lambda \lambda\} \\
&= \lambda^{\left(\frac{p}{2}+a_\lambda-1\right)} \exp\left\{-\left(b_\lambda + \frac{1}{2\sigma^2}\beta^T\beta\right) \lambda\right\}
\end{aligned}$$

entonces

$$\lambda|- \sim \text{G}\left(a_\lambda + \frac{p}{2}, b_\lambda + \frac{1}{2\sigma^2}\beta^T\beta\right)$$

Dado que las tres distribuciones condicionales completas tienen forma cerrada, se obtienen muestras de la distribución posterior utilizando el muestreador de Gibbs, con 1000 muestras de calentamiento y realizando un muestreo sistemático cada 10 muestras.

Entonces, dados $\{\beta^{(b)}, \sigma^{2(b)}, \lambda^{(b)}\}$:

- Actualizar β : Muestreamos $\beta^{(b+1)}$ de su distribución condicional completa (Normal p -variada) usando el estado actual de $\sigma^{2(b)}$ y $\lambda^{(b)}$.
- Actualizar σ^2 : Muestreamos $\sigma^{2(b+1)}$ de su distribución condicional completa (Gamma Inversa) usando el estado actual de $\beta^{(b+1)}$ y $\lambda^{(b)}$.
- Actualizar λ : Muestreamos $\lambda^{(b+1)}$ de su distribución condicional completa (Gamma) usando el estado actual de $\beta^{(b+1)}$ y $\sigma^{2(b+1)}$.

Distribuciones Modelo 4

Distribución posterior

Sea $\Theta = (\beta, \sigma^2, \rho)$, entonces

$$p(\Theta|\mathbf{y}) \propto \mathbf{N}_n(\mathbf{y} \mid \mathbf{X}\beta, \sigma^2 \mathbf{C}_\rho) \times \mathbf{N}_p(\beta \mid \mathbf{0}_p, \Sigma_0) \times \text{GI}\left(\sigma^2 \mid \frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\right) \times \text{U}(\rho|a_\rho, b_\rho)$$

Distribuciones condicionales completas

- Para β

$$\begin{aligned} p(\beta|-) &\propto \mathbf{N}_n(\mathbf{y} \mid \mathbf{X}\beta, \sigma^2 \mathbf{C}_\rho) \times \mathbf{N}_p(\beta \mid \mathbf{0}_p, \Sigma_0) \\ &\propto \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\beta)^T \mathbf{C}_\rho^{-1}(\mathbf{y} - \mathbf{X}\beta)\right\} \times \exp\left\{-\frac{1}{2}\beta^T \Sigma_0^{-1}\beta\right\} \\ &\propto \exp\left\{-\frac{1}{2\sigma^2}(\beta^T \mathbf{X}^T \mathbf{C}_\rho^{-1} \mathbf{X} \beta - 2\beta^T \mathbf{X}^T \mathbf{C}_\rho^{-1} \mathbf{y})\right\} \times \exp\left\{-\frac{1}{2}\beta^T \Sigma_0^{-1}\beta\right\} \\ &= \exp\left\{-\frac{1}{2}\left(\beta^T \left(\frac{1}{\sigma^2} \mathbf{X}^T \mathbf{C}_\rho^{-1} \mathbf{X} + \Sigma_0^{-1}\right) \beta - 2\beta^T \left(\frac{1}{\sigma^2} \mathbf{X}^T \mathbf{C}_\rho^{-1} \mathbf{y}\right)\right)\right\} \end{aligned}$$

entonces

$$\beta|-\sim \mathbf{N}_p\left(\left(\frac{1}{\sigma^2} \mathbf{X}^T \mathbf{C}_\rho^{-1} \mathbf{X} + \Sigma_0^{-1}\right)^{-1} \left(\frac{1}{\sigma^2} \mathbf{X}^T \mathbf{C}_\rho^{-1} \mathbf{y}\right), \left(\frac{1}{\sigma^2} \mathbf{X}^T \mathbf{C}_\rho^{-1} \mathbf{X} + \Sigma_0^{-1}\right)^{-1}\right)$$

- Para σ^2

$$\begin{aligned}
p(\sigma^2|-) &\propto \mathbf{N}_n(\mathbf{y} \mid \mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{C}_\rho) \times \text{GI}\left(\sigma^2 \mid \frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\right) \\
&\propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{C}_\rho^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\} \times (\sigma^2)^{-(\frac{\nu_0}{2}+1)} \exp\left\{-\frac{\nu_0 \sigma_0^2/2}{\sigma^2}\right\} \\
&= (\sigma^2)^{-\left(\frac{n+\nu_0}{2}+1\right)} \exp\left\{-\frac{1}{\sigma^2} \left(\frac{\nu_0 \sigma_0^2 + \text{SSE}_\rho}{2}\right)\right\}
\end{aligned}$$

donde $\text{SSE}_\rho = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{C}_\rho^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$, entonces

$$\sigma^2|- \sim \text{GI}\left(\frac{n + \nu_0}{2}, \frac{\nu_0 \sigma_0^2 + \text{SSE}_\rho}{2}\right)$$

Dado que las dos distribuciones condicionales completas tienen forma cerrada, se obtienen muestras de dichas distribuciones utilizando el muestreador de Gibbs, con 1000 muestras de calentamiento y realizando un muestreo sistemático cada 10 muestras.

Entonces, dados $\{\boldsymbol{\beta}^{(b)}, \sigma^{2(b)}, C_{\rho^{(b)}}\}$:

- Actualizar $\boldsymbol{\beta}$: Muestreamos $\boldsymbol{\beta}^{(b+1)}$ de su distribución condicional completa (Normal p -variada) usando el estado actual de $\sigma^{2(b)}$ y $C_{\rho^{(b)}}$.
- Actualizar σ^2 : Muestreamos $\sigma^{2(b+1)}$ de su distribución condicional completa (Gamma Inversa) usando el estado actual de $\boldsymbol{\beta}^{(b+1)}$ y $C_{\rho^{(b)}}$.

- Para ρ

$$\begin{aligned}
p(\rho|-) &\propto \mathbf{N}_n(\mathbf{y} \mid \mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{C}_\rho) \times \text{U}(\rho|a_\rho, b_\rho) \\
&\propto |\mathbf{C}_\rho|^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{C}_\rho^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\}
\end{aligned}$$

la cual no tiene una forma cerrada, por lo que procedió a utilizar el algoritmo de Metrópolis para encontrar muestras de dicha distribución. Para actualizar $\rho^{(b)}$ por medio de Metrópolis se requiere el estado actual de $\boldsymbol{\beta}^{(b+1)}$ y $\sigma^{2(b+1)}$. Se utilizó un valor de $\delta = 0.3$ y se obtuvo una tasa de aceptación del 43.13 %. Además con cada actualización de $\rho^{(b)}$, se construye la matriz y $C_{\rho^{(b)}}$.