
CASO DE ESTUDIO 3

ESTADÍSTICA BAYESIANA. 2023-II

Michel Mendivenson Barragán Zabala

Departamento de estadística
Universidad Nacional de Colombia
Bogotá, Colombia
mbarraganz@unal.edu.co

Gerardo Sebastian Gil Sanchez

Departamento de estadística
Universidad Nacional de Colombia
Bogotá, Colombia
ggil@unal.edu.co

27 de noviembre de 2023

1. 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.

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 $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$, note que las componentes de $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ 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

Intención de voto: Bogotá		
Candidato	Cantidad	Proporción
C. F. Galán	493	0.411
G. Bolivar	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 Blango	54	0.045
Total	1200	1.000

Cuadro 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?

$\mathbf{n} = (n_1, \dots, n_k)$ 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 \mathbf{n} sigue una distribución multinomial con parámetros n y $\boldsymbol{\theta}$, la cual se define como sigue: $\mathbf{n} \mid n, \boldsymbol{\theta} \sim \text{Multinomial}(n, \boldsymbol{\theta})$ si y solo si

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

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 $\mathbf{n} \mid n, \boldsymbol{\theta} \sim \text{Multinomial}(n, \boldsymbol{\theta})$ y distribución previa jerárquica dada por

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

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 modelos más sofisticados que consideran la estrategia de muestreo).

Resultados

Una vez ajustado el modelo propuesto se hacen estimaciones de la proporción de votos presentando los siguientes resultados:

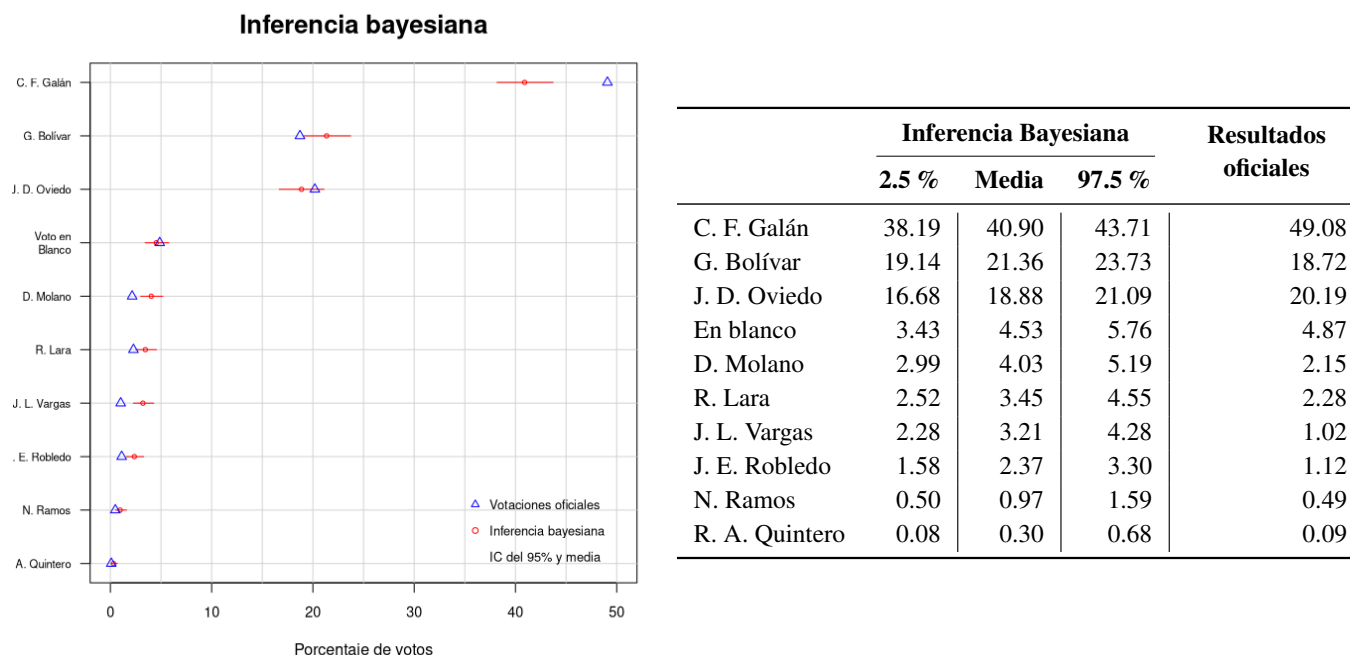


Figura 1: Ranking bayesiano y resultados oficiales (En puntos porcentuales) ¹.

Aunque es evidente que nuestro modelo no logró predecir con precisión las proporciones de las votaciones finales, es fundamental considerar que la inferencia bayesiana se fundamentó en una muestra de 1200 personas. Este tamaño de muestra puede introducir cierto margen de error en las predicciones, y también se debe tener en cuenta que la discrepancia podría atribuirse a la diferencia temporal entre la recolección de la muestra y las elecciones.

Al analizar los resultados y las virtudes del modelo, detallados en la sección de información adicional, podemos evaluar su desempeño en términos de bondad de ajuste, así como en la variabilidad y estabilidad del modelo. Los análisis detallados en esa sección ofrecerán una comprensión más completa del comportamiento del modelo.

Este enfoque proporciona una visión equilibrada de las limitaciones del modelo y destaca la importancia de considerar las condiciones y el contexto de la muestra al interpretar los resultados.

2. 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.

¹Los resultados oficiales de las votaciones fueron obtenidos de la página de la Registraduría Nacional del Estado Civil.

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 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).

Modelamiento

Se considera un modelo de regresión de la forma $\mathbf{y} \mid \mathbf{X}, \beta, \sigma^2 \sim \mathcal{N}(\mathbf{X}\beta, \sigma^2 \mathbf{I}_n)$, 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, β 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 $\hat{\beta} = \mathbb{E}(\beta \mid \mathbf{y}_{\text{train}})$, se genera $\hat{\mathbf{y}}_{\text{test}} = \mathbf{X}_{\text{test}} \hat{\beta}$. 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:

$$p(\beta, \sigma^2, \lambda) = \mathcal{N}(\beta \mid \mathbf{0}_p, \frac{\sigma^2}{\lambda} \mathbf{I}_p) \cdot \text{Gl}(\sigma^2 \mid \nu_0/2, \nu_0 \sigma_0^2/2) \cdot \mathcal{G}(\lambda \mid a_\lambda, b_\lambda),$$

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

Modelo 4: Regresión con errores correlacionados

Distribución muestral:

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

donde C_ρ es una matriz con estructura autoregresiva de primer orden de la forma

$$C_\rho = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix}.$$

Distribución previa:

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

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

Resultados

Una vez ajustados los modelos obtenemos los siguientes resultados respecto a la validación y la bondad de ajuste:

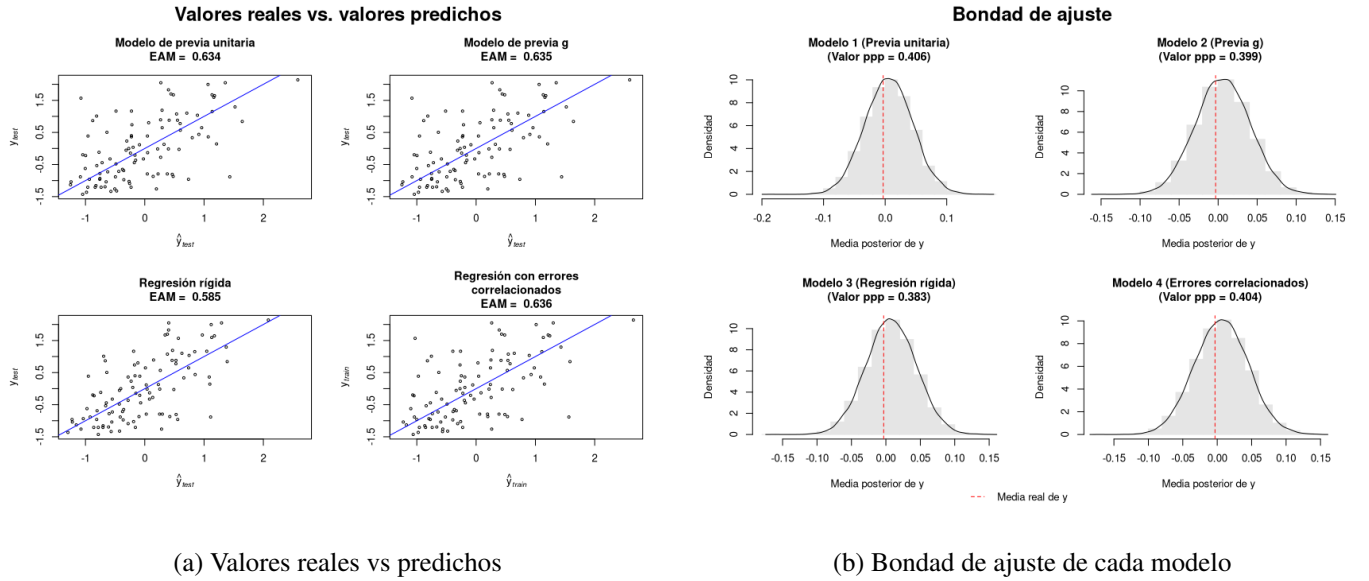


Figura 2: Resultados EAM y bondad de ajuste

De donde, si nos propusiéramos escoger un modelo, no tendríamos seguridad de cual de ellos escoger debido a que en términos de error absoluto medio el modelo 3 es mejor que los demás pero en términos de bondad de ajuste el modelo que se desempeña mejor es el modelo con errores correlacionados. Por eso, revisamos el Criterio de Información de la Devianza (DIC) de cada modelo, para poder tomar una decisión más informada:

Modelo	DIC	pDIC
Previa unitaria	784.849	64.594
Previa g	778.109	64.831
Regresión rígida	760.262	47.369
Regresión con errores correlacionados	783.766	63.401

Cuadro 2: DIC de cada modelo

Si bien es cierto que es posible llegar a pensar que aún cuando el DIC del modelo 3 de regresión rígida no parece arrojar un resultado correcto debido a que contradice el valor ppp encontrado para ese modelo, también es cierto que en cambio no contradice el error absoluto medio encontrado. Esta aparente contradicción se debe a que los dos valores cuantifican cosas distintas: Mientras que el EAM nos dice el ajuste del modelo a los valores particulares de la variable dependiente (Específicamente para los valores de prueba), el valor ppp por como se planteó el cálculo del mismo nos habla del comportamiento global respecto a un estadístico de prueba.

En todo caso como nuestro objetivo con estos modelos de regresión es, precisamente, predecir valores de y dado un valor para los regresores es mejor tomar decisiones respecto a medidas que nos indiquen precisamente cómo de bien predicen los modelos.

Teniendo todo esto en cuenta podemos concluir que el modelo a ser seleccionado luego de este breve análisis debe de ser el modelo de la regresión rígida.

Anexo. Información adicional

Parte 1 (Alcaldía de Bogotá)

Distribuciones condicionales completas

En el caso del modelo propuesto para el estudio de la encuesta hecha por Invamer S.A.S se tendrá que la distribución condicional completa del parámetro θ tiene forma cerrada mientras que la distribución condicional completa del parámetro α no es cerrada:

$$\theta \mid \alpha, \mathbf{n} \sim \text{Dirichlet}(\mathbf{n} + \alpha \mathbf{1}_k)$$

$$p(\alpha \mid \theta, \mathbf{n}) \propto \frac{\Gamma(k\alpha)}{[\Gamma(\alpha)]^{k+1}} \exp(-b\alpha) \prod_{j=1}^k \theta_j^{\alpha-1}$$

Tenga en cuenta además que $\alpha > 0$.

Implementación

Como la distribución condicional del parámetro θ es completa pero la distribución condicional del parámetro α no, se hace necesario usar una mezcla del muestreador de Gibbs y del algoritmo de Metropolis (Usualmente llamado Hasting-Gibbs por bloques) para muestrear del espacio de parámetros y tenga en cuenta que, en adición a esto, debido a la restricción sobre α será necesario usar una distribución de propuesta no simétrica:

$$\alpha^* \sim \text{Gamma} \left(\left[\frac{\alpha^{(b)}}{\delta} \right]^2, \frac{\alpha^{(b)}}{\delta^2} \right)$$

con $\alpha^{(b)}$ el valor generado en el paso inmediatamente anterior del algoritmo y δ^2 el parámetro de ajuste. Debido a esta parametrización de la distribución gamma se tendrá que $E(\alpha^*) = \alpha^{(b)}$ y $Var(\alpha^*) = \delta^2$.

Para encontrar el valor apropiado de δ^2 antes del ciclo de quemado del algoritmo se corre el mismo para parámetros de ajuste 0.1, 0.2, 0.3, 0.4, 0.5 y 0.6 obteniendo los resultados presentados en la tabla 3.

δ^2	0.1	0.2	0.3	0.4	0.5	0.6
Tasa de aceptación	59.7 %	46.2 %	42.3 %	38 %	31.9 %	31.1 %

Cuadro 3: Tasa de aceptación para distintos δ^2

Por lo que se decide tomar $\delta^2 = 0.3$ para la implementación del algoritmo.

Finalmente, el algoritmo se itera 105000 veces de las cuales 5000 iteraciones serán tomadas como el calentamiento del algoritmo y para las 100000 restantes se escogen solamente 10000 muestras usando un muestreo sistemático con un intervalo de 10 para disminuir la autocorrelación de la cadena contando al final con una tasa de aceptación del 42.2 % aproximadamente.¹

Convergencia y no autocorrelación Hasting-Gibbs por bloques

Una vez implementado el muestreador, nos enfocamos en estudiar la convergencia de la cadena en el espacio de α para analizar si nuestro algoritmo quedó bien implementado:

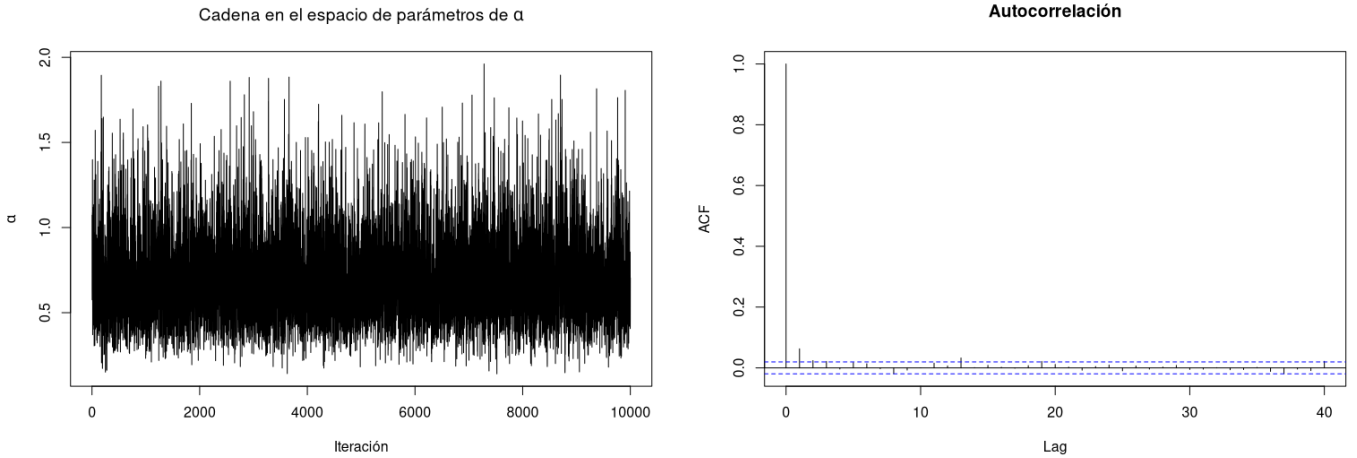


Figura 3: Convergencia y autocorrelación de α

Revisando la figura 3 se observa que la cadena generada no exhibe evidencia de no convergencia debido a la estacionariedad de la primera figura y además, debido al muestreo sistemático, nuestra cadena exhibe el comportamiento deseado en cuanto a autocorrelación se refiere. Por otro lado, en la siguiente tabla,

Parámetro	$\theta_{\text{Galán}}$	$\theta_{\text{Bolívar}}$	θ_{Oviedo}	θ_{Molano}	θ_{Lara}	θ_{Vargas}	θ_{Robledo}	θ_{Ramos}	θ_{Quintero}	$\theta_{\text{En blanco}}$	α
CV	0.035 %	0.055 %	0.06 %	0.141 %	0.151 %	0.159 %	0.187 %	0.295 %	0.537 %	0.135 %	0.434 %

Cuadro 4: Coeficientes de variación de Monte Carlo de los parámetros

Donde, por ejemplo, $\theta_{\text{Galán}}$ es el parámetro de proporción asociado a la proporción de votos del candidato C. F. Galán. Podemos comprobar que para todos los parámetros tenemos coeficientes de variación de Monte Carlo bajos lo cual

¹Para más información sobre la implementación se recomienda revisar los documentos adjuntos a este archivo

indica que nuestra cadena es estable y con poca varianza relativa a la media de los parámetros por lo que el método implementado es confiable para generar resultados basados en inferencia bayesiana.

Parte 2 (Datos de Diabetes)

Distribuciones condicionales completas

A continuación se presentan las distribuciones condicionales completas de cada uno de los modelos propuestos:

■ Modelo previa unitaria:

$$\beta \mid \text{resto} \sim N_p \left((\Sigma_0^{-1} + \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{X})^{-1} (\Sigma_0^{-1} \beta_0 + \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{y}), (\Sigma_0^{-1} + \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{X})^{-1} \right)$$

y

$$\sigma^2 \mid \text{resto} \sim GI \left(\frac{\nu_0 + n}{2}, \frac{\nu_0 \sigma_0^2 + SSR(\beta)^1}{2} \right)$$

Especificando los siguientes valores para la previa: $\beta_0 = \hat{\beta}_{ols}$, $\sigma_0^2 = n \hat{\sigma}_{ols}^2 (\mathbf{X}^T \mathbf{X})^{-1}$, $\nu_0 = 1$ y $\sigma^2 = \hat{\sigma}_{ols}^2$.

■ Modelo previa g:

$$\beta \mid \text{resto} \sim N_p \left(\frac{g}{g+1} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}, \frac{g}{g+1} \frac{1}{\sigma^2} (\mathbf{X}^T \mathbf{X})^{-1} \right)$$

y

$$\sigma^2 \mid \text{resto} \sim GI \left(\frac{\nu_0 + n}{2}, \frac{\nu_0 \sigma_0^2 + SSR_g^2}{2} \right)$$

■ Modelo regresión rígida:

$$\beta \mid \text{resto} \sim N_p (\sigma^2 (\lambda \mathbf{I}_p + \mathbf{X}^T \mathbf{X})^{-1}, (\lambda \mathbf{I}_p + \mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y})$$

y

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

y finalmente

$$\sigma^2 \mid \text{resto} \sim GI \left(\frac{n + p + \nu_0}{2}, \frac{SSR(\beta) + \beta^T \beta + \nu_0 \sigma_0^2}{2} \right)$$

■ Modelo regresión con errores correlacionados:

$$\beta \mid \text{resto} \sim N_p(\beta_n, \Sigma_n)$$

$$\text{con } \Sigma_n = (\Sigma_0^{-1} + \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{C}_\rho^{-1} \mathbf{X})^{-1} \text{ y } \beta_n = \Sigma_n (\Sigma_0^{-1} \beta_0 + \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{C}_\rho^{-1} \mathbf{y}) \text{ y}$$

¹Con $SSR(\beta)$ la suma de cuadrados residuales de la regresión.

²Con $SSR_g = \mathbf{y}^T \left(\mathbf{I}_n - \frac{g}{g+1} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right) \mathbf{X}^T \mathbf{y}$

$$\sigma^2 \mid \text{resto} \sim GI \left(\frac{\nu_0 + n}{2}, \frac{\nu_0 \sigma_0^2 + SSR_\rho^3}{2} \right)$$

Sin embargo, ρ no tiene una distribución condicional completa cerrada.

Convergencia y coeficientes de variación de Monte Carlo

Para ajustar el modelo de la previa unitaria y el modelo de regresión rígida se utiliza el método del muestreador de Gibbs, para el modelo de la previa g se utiliza un método de Monte Carlo puro y finalmente para el modelo de los errores correlacionados debido a que ρ no presenta una distribución condicional cerrada se usa una combinación de Hasting-Gibbs como el utilizado en la parte 1 del presente caso de estudio. Ahora bien, a continuación se presentará alguna información general sobre los algoritmos implementados sobre todo para revisar su convergencia y determinar si estos métodos fueron implementados correctamente. Note, sin embargo, que como el modelo de previa g es ajustado usando un método de Monte Carlo puro no será necesario revisar nada de lo mencionado para ese modelo, primero veamos que los modelos si alcanzaron la estacionariedad revisando las cadenas de log-verosimilitud de cada uno:

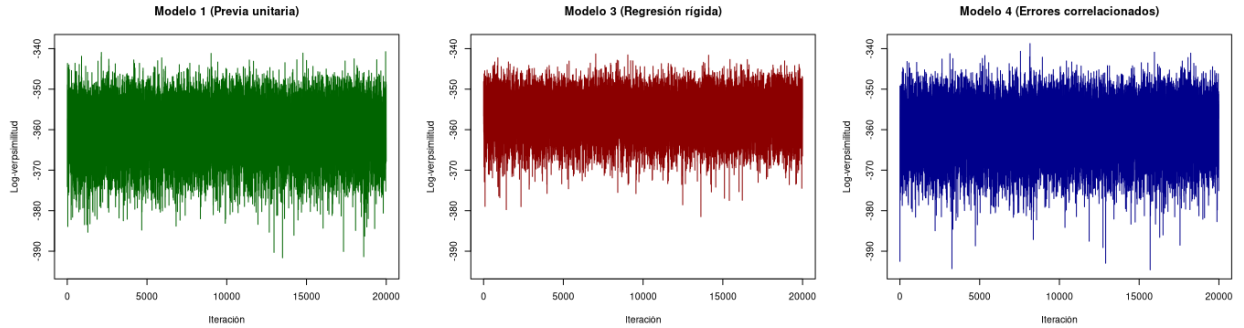


Figura 4: Log-verosimilitud de los modelos

Donde se puede notar claramente que en las figuras de cada modelo no hay evidencia presente que indique la no estacionariedad o no convergencia de los modelos. También se revisan los coeficientes de variación de Monte Carlo:

³Con $SSR_\rho = (\mathbf{y} - \mathbf{X}\beta)^T \mathbf{C}_\rho^{-1} (\mathbf{y} - \mathbf{X}\beta)$

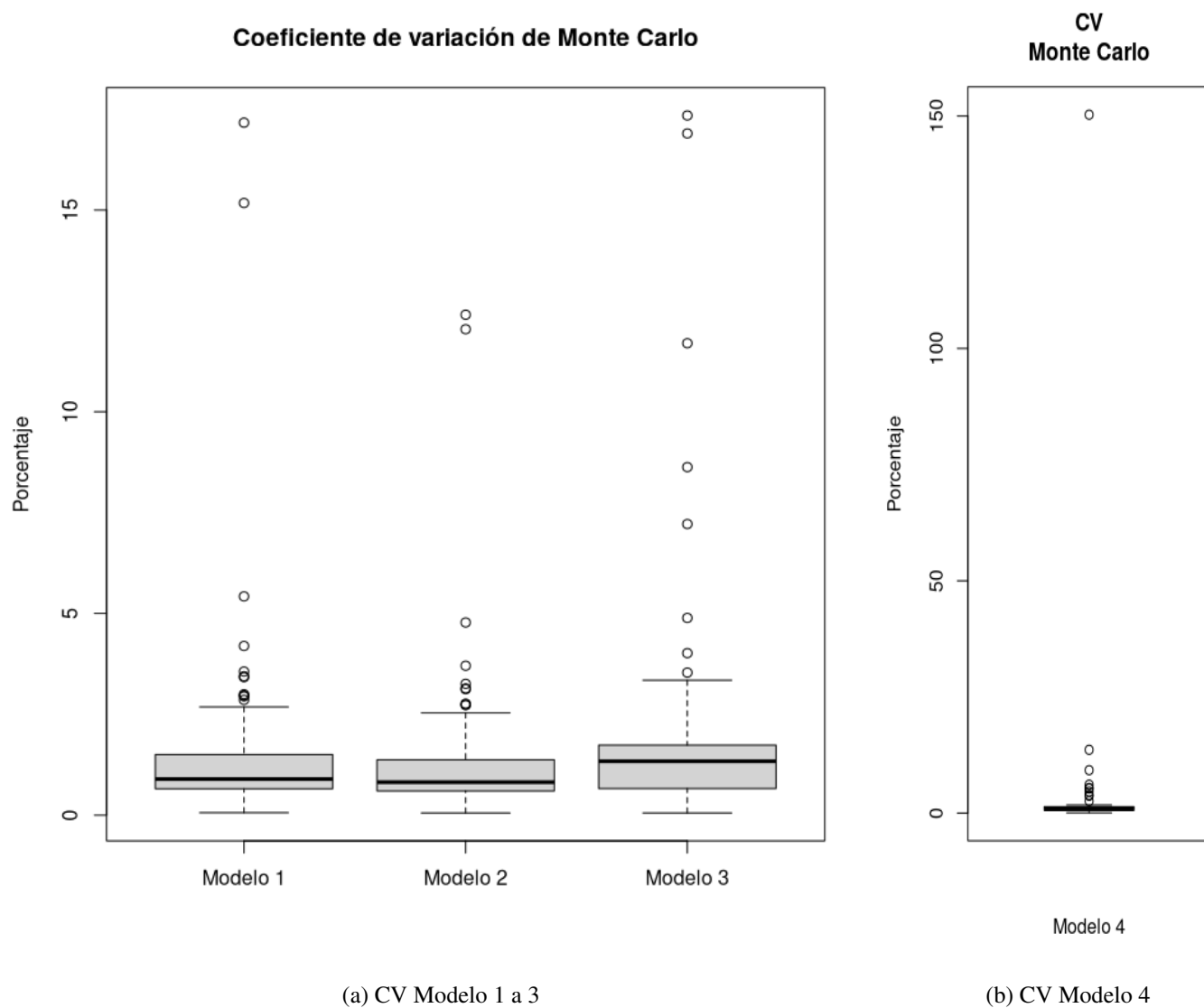


Figura 5: Boxplot de coeficientes de Monte Carlo

	Mínimo	25 %	Media	75 %	Máximo
Modelo 1	0.060	0.654	1.749	1.501	17.165
Modelo 2	0.055	0.597	1.509	1.372	12.404
Modelo 3	0.055	0.660	2.157	1.734	17.342
Modelo 4	0.061	0.592	3.812	1.362	150.271

Cuadro 5: Coeficientes de variación de Monte Carlo (En porcentajes)

Los cuales, en su mayoría se encuentran dentro de los límites aceptables por lo que es posible afirmar que nuestra cadena es estable y junto con lo anterior nos es permitido decir que son modelos adecuados para hacer inferencia bayesiana.

Comentarios sobre la implementación del modelo de errores correlacionados

De la misma forma en como se implementó el modelo en la sección 2, para el caso del modelo con errores correlacionados se prueba el algoritmo antes del ciclo de quemado para conocer qué valor δ satisface el criterio de una tasa de aceptación de entre 30 % y 50 % tomándose la decisión de utilizar $\delta = 0,28$ dando como resultado, luego de iterar el algoritmo 201000 veces (Incluyendo las iteraciones del ciclo de calentamiento) una tasa de aceptación de aproximadamente 45,35 %. Además de esto, de nuevo siguiendo la metodología desarrollada en la parte 1, la distribución de propuesta utilizada para ρ en la implementación del modelo fue simular primero $a \sim U(\rho^{(b)} - \delta, \rho^{(b)} + \delta)$ y definir a ρ^* como sigue:

$$\rho^* = \min\{|\rho|, 2 - |\rho|\}$$

Esto con el objetivo de que la distribución de propuesta sea simétrica y por ende podamos usar una tasa de aceptación r de cálculo más sencillo. Además para todos los modelos implementados (Menos para el 2 por tratarse de un método Monte Carlo) se hicieron 201000 iteraciones del algoritmo con un ciclo de quemado de 1000 y un muestreo sistemático de ancho 10 para reducir la auto correlación de la cadena para obtener al final una muestra de 20000 elementos del espacio de parámetros.

Como una nota final cabe destacar que inicialmente con una muestra de tamaño 10000 se encontraron coeficientes de variación de Monte Carlo inusualmente altos para algunos parámetros en algunos modelos por lo que se decide duplicar el tamaño de las muestras para todos los modelos para disminuir en cierto grado los coeficientes de variación. Esos fueron los resultados que se presentaron en la tabla 5.

Referencias

Bernardo, J. M. and Smith, A. F. (2000). *Bayesian theory*. John Wiley & Sons.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2013). *Bayesian Data Analysis*. CRC Press.

Hoff, P. D. (2009). *A first course in Bayesian statistical methods*, volume 580. Springer.