



Universidad Nacional de Colombia

FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA

ESTADÍSTICA BAYESIANA

Parcial 3

Alejandro Urrego López
aurrego@unal.edu.co

Noviembre 2023

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

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?

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 [pág. 176]bernardo2009bayesian 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 $\mathbf{n} = (n_1, \dots, n_k)$ 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

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

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 gelman2013bayesian se discuten 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).

Resultados:

Candidato	Estimación	Q2.5 %	Q97.5 %	Registraduria
C. F. Galan	40.9	38.1	43.7	49.02
G. Bolivar	21.4	19.1	23.8	20.1
J. D. Oviedo	18.9	16.7	21.1	18.71

Tabla 2: Estimación Bayesiana

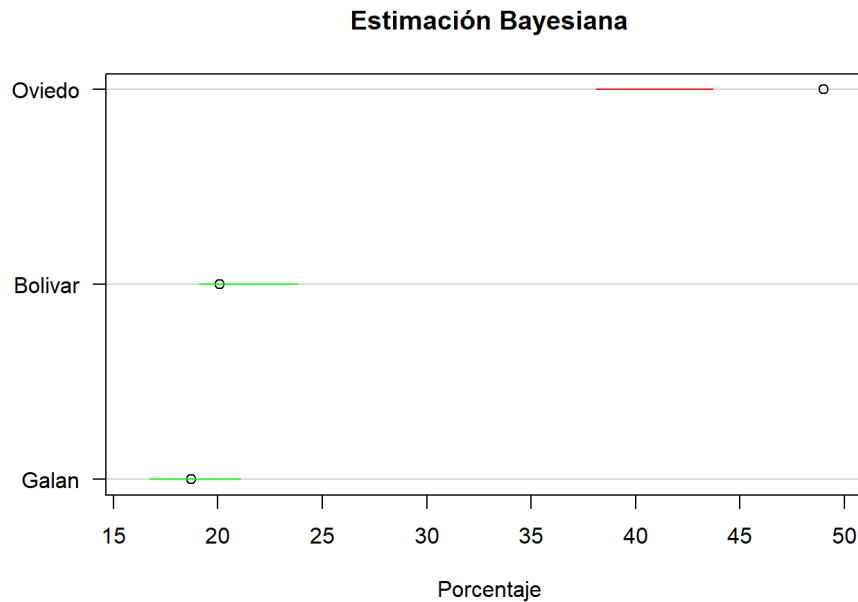


Figura 1: Estimación Bayesiana

Se puede observar que los resultados de Bolívar y Oviedo caen dentro del intervalo de credibilidad. Por lo tanto, la inferencia realizada sobre estos dos parámetros es correcta. Sin embargo, podemos ver que se está subestimando los resultados de Galán.

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.

Datos de diabetes

Considere la base de datos de diabetes dada en la Sección 9.3 de Hoff 2009 página 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 $\mathbf{y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2 \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$, 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 $\hat{\boldsymbol{\beta}} = \mathbf{E}(\boldsymbol{\beta} \mid \mathbf{y}_{\text{train}})$, 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:

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

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:

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

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

$$\mathbf{C}_\rho = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{bmatrix}.$$

Distribución previa:

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

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

Preguntas

Ajustar cada modelo utilizando los datos de entrenamiento ($\mathbf{y}_{\text{train}}, \mathbf{X}_{\text{train}}$) (incluir un anexo con todos los detalles).

1. Para cada modelo, generar $\hat{\mathbf{y}}_{\text{test}} = \mathbf{X}_{\text{test}} \hat{\boldsymbol{\beta}}$ usando los coeficientes de regresión estimados $\hat{\boldsymbol{\beta}} = \mathbf{E}(\boldsymbol{\beta} \mid \mathbf{y}_{\text{train}})$. Graficar $\hat{\mathbf{y}}_{\text{test}}$ frente \mathbf{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 pagina 170).
- En el encabezado de cada gráfica, incluir el error absoluto medio usando tres (3) cifras decimales.

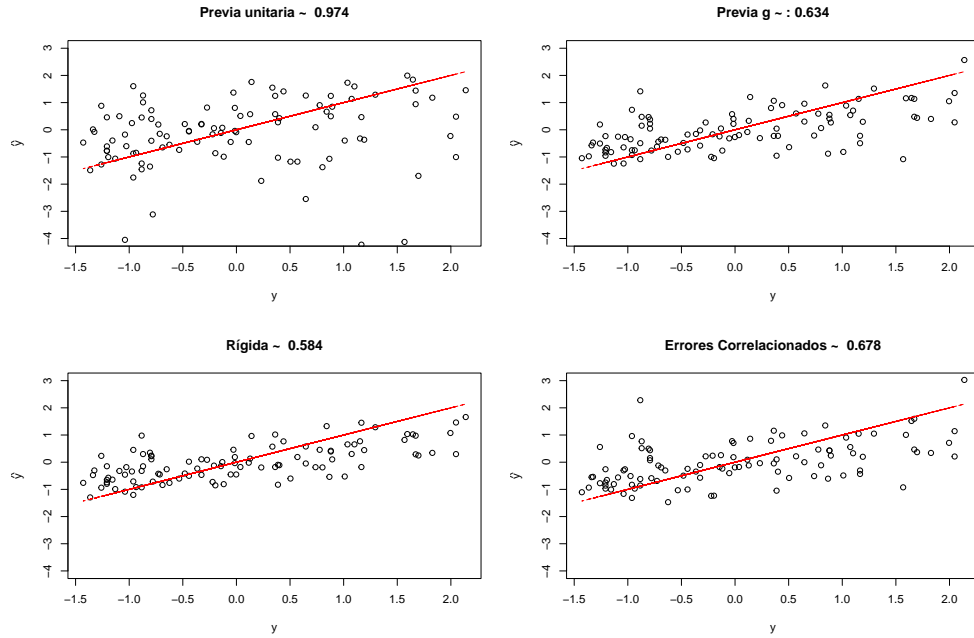


Figura 2: Error Absoluto Medio

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.

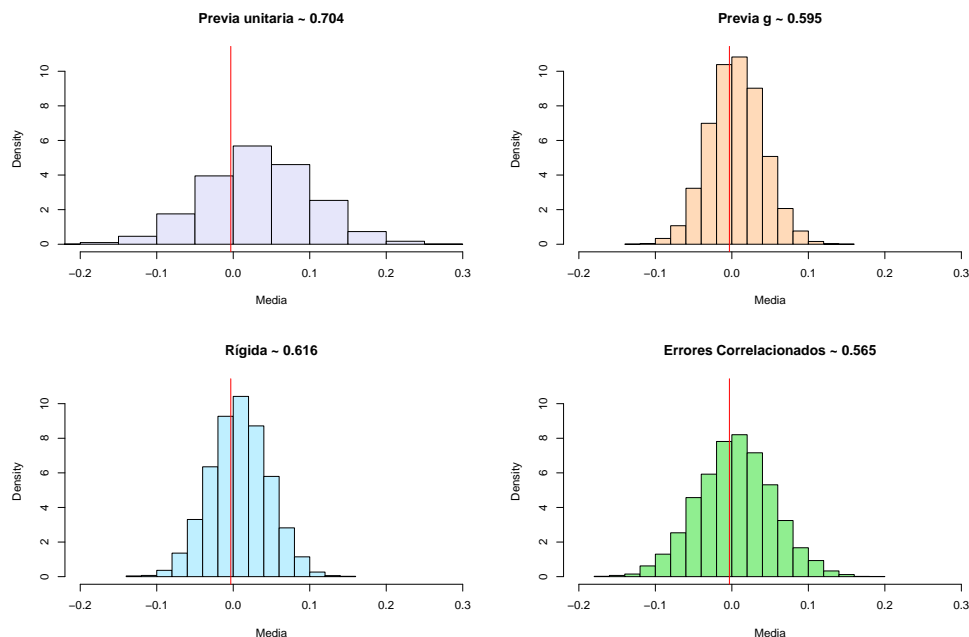


Figura 3: Media Distribución Predicativa Posterior

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

	Modelo 1	Modelo 2	Modelo 3	Modelo 4
DIC	1144.967	777.575	768.112	840.542

Tabla 3: DIC

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

Se puede observar que el modelo de regresión clásica previa g , la regresión rígida y el modelo con errores correlacionados, junto con sus respectivas distribuciones predictivas posteriores, capturan de manera adecuada el comportamiento de la media. En otras palabras, los tres modelos permiten realizar inferencias sobre la media de la población. No obstante, se nota que el modelo de regresión clásica previa unitaria presenta dificultades para reproducir dicho comportamiento.

Adema al calcular el error absoluto medio, se evidencia que el modelo de regresión rígida presenta un error absoluto medio más bajo. Esto se refleja en el hecho de que el modelo tiene un DIC más bajo, llevándonos a la conclusión de que es el mejor en términos de precisión predictiva.

Anexos:

Punto 1.

Muestreador de Gibbs con paso de Metropolis:

1. Se encuentran las distribuciones condicionales completas de $\Theta = \{\theta, \alpha\}$:

$$(\theta | \text{resto}) \sim \text{Dirichlet}(n_j + \alpha)$$

$$(\alpha | \text{resto}) \propto \frac{\Gamma\left(\sum_{j=1}^k \alpha_j\right)}{\prod_{j=1}^k \Gamma(\alpha_j)} \prod_{j=1}^k \theta_j^{\alpha_j - 1} \alpha^{-(a+1)} \exp\left\{-\frac{b}{\alpha}\right\}$$

2. Se inicializa $\Theta^{[0]}$, asignándole un valor en el espacio paramétrico para θ y α . Este valor se puede muestrear de la distribución previa de ambos parámetros. En este caso, a α se le asigna un valor igual a uno y se muestrea de la distribución previa para θ .
3. Se actualiza el valor de los parámetros del modelo, es decir, $\Theta^{[b]}$ a $\Theta^{[b+1]}$ para $b \in \{0, 1, \dots, B-1\}$, tomando muestras de las distribuciones condicionales completas, es decir:

$$\theta^{[b+1]} \sim \text{Dirichlet}(n_j + \alpha^{[b]})$$

En este caso, como la distribución condicional completa de α es desconocida, se añade un paso de Metrópolis para muestrearla.

- a) Se simula un $\alpha^* \sim J(\alpha | \alpha^{[b]})$, donde J se denomina distribución de propuestas. Esta distribución es simétrica y, por lo tanto, está centrada en el valor más actual del parámetro, es decir, $\alpha^{[b]}$. En este caso,

$$\alpha^* \sim N(\alpha^{[b]}, \delta)$$

donde δ es el parámetro de ajuste y se elige para que el algoritmo funcione de manera eficiente.

Como $\alpha \in (0, \infty)$, se le aplica una transformación logarítmica, es decir, $\eta = \ln(\alpha)$, y en lugar de generar θ^* se genera η^* para que se respete la restricción paramétrica. Por lo tanto,

$$\eta^* \sim N(\ln \alpha^{[b]}, \delta)$$

- b) Se calcula la tasa de aceptación corregida, dado que se le aplicó una transformación a α . Por lo tanto, la tasa de aceptación viene dada por

$$r = \frac{p(\eta^* | y)}{p(\eta^{[b]} | y)} = \frac{p(\exp \eta^* | y)}{p(\exp \eta^{[b]} | y)} \frac{\exp \eta^*}{\exp \eta^{[b]}}$$

Para brindar estabilidad al proceso, se calcula primero en escala logarítmica, es decir,

$$r = \exp \{ \ln(p(\eta^* | y)) - \ln(p(\exp \eta^* | y)) \}$$

- c) Se actualiza a α de la siguiente forma

$$\alpha^{[b+1]} = \begin{cases} \exp \eta^*, & \text{con probabilidad } \min(1, r); \\ \alpha^{[b]}, & \text{con probabilidad } 1 - \min(1, r). \end{cases}$$

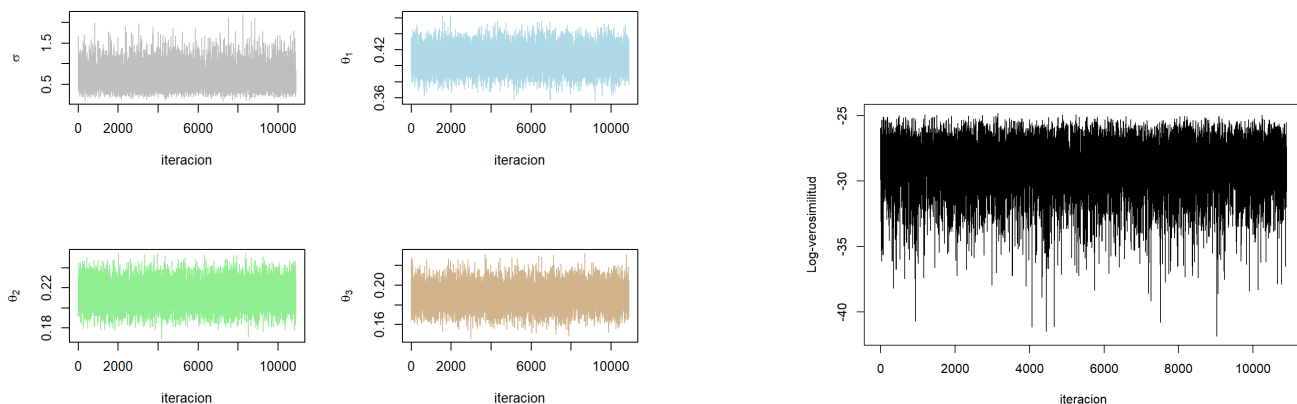
- d) Para establecer δ , se realizan 1000 iteraciones previas calculando la tasa de aceptación, es decir,

$$\text{acs} = \frac{N}{1000}$$

con N como el número de veces en que $\alpha^{[b+1]} = \exp \eta^*$. Este proceso se repite hasta que $0.3 < \text{acs} < 0.5$.

4. Una vez establecido δ , se repite el paso 3 B veces, hasta convergencia.

Cadenas de Márkov y Log-verosimilitud



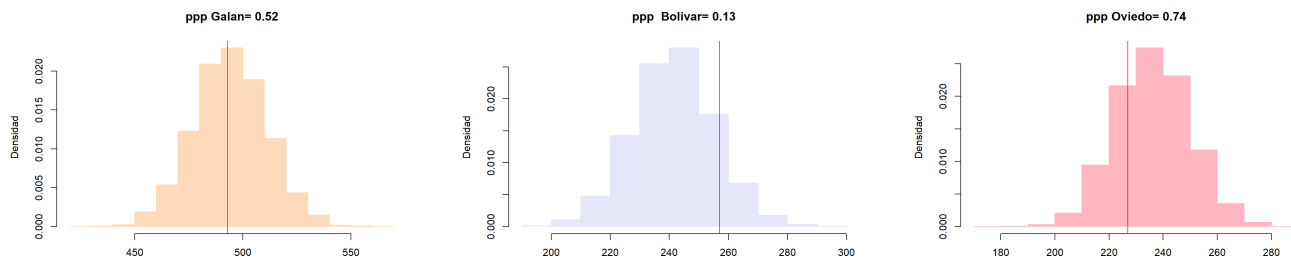
Cadenas de Márkov

Log-verosimilitud

Tamaños efectivos de muestra y Coeficientes de variación de Montecarlo

Parámetro	Tamaño Efectivo	C.V de Montecarlo
θ_1	10260.14	0.000
θ_2	10900.00	0.000
θ_3	10900.00	0.000
θ_4	10900.00	0.000
θ_5	10900.00	0.000
θ_6	10900.00	0.000
θ_7	11486.92	0.000
θ_8	10900.00	0.000
θ_9	10900.00	0.000
θ_{10}	11389.45	0.000
α	10118.45	0.003

ppp e intervalos de credibilidad



Punto 2.

Regresión clásica previa unitaria - Muestreador de Gibbs

- Se encuentran las distribuciones condicionales completas de $\Theta = \{\beta, \sigma^2\}$:

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

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

- Se inicializa $\Theta^{[0]}$, asignándole un valor en el espacio paramétrico para β y σ^2 . Este valor se puede muestrear de la distribución previa de ambos parámetros.

- Se actualiza el valor de los parámetros del modelo, es decir, $\Theta^{[b]}$ a $\Theta^{[b+1]}$ para $b \in \{0, 1, \dots, B-1\}$, tomando muestras de las distribuciones condicionales completas, es decir:

$$\begin{aligned}\beta^{[b+1]} &\sim \mathcal{N}_p \left((\Sigma_0^{-1} + \frac{1}{(\sigma^2)^{[b]}} \mathbf{X}^T \mathbf{X})^{-1} (\Sigma_0^{-1} \beta_0 + \frac{1}{(\sigma^2)^{[b]}} \mathbf{X}^T \mathbf{y}), (\Sigma_0^{-1} + \frac{1}{(\sigma^2)^{[b]}} \mathbf{X}^T \mathbf{X})^{-1} \right) \\ \sigma^{2[b+1]} &\sim \text{Gl} \left(\frac{\nu_0 + n}{2}, \frac{\nu_0 \sigma_0^2 + \text{SSR}(\beta^{[b+1]})}{2} \right)\end{aligned}$$

- Se repite el paso 3 B veces, hasta convergencia.

Regresión clásica previa g - Monte Carlo

- Se encuentran las distribuciones posterior:

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

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

- Se toma una muestra de σ^2 , es decir:

$$\sigma^2 \sim p(\sigma^2 \mid \mathbf{X}, \mathbf{y})$$

- Con esta muestra se simula β , es decir,

$$\beta \sim p(\beta \mid \sigma^2, \mathbf{X}, \mathbf{y})$$

- Se repite el paso 3 el número de veces que se considere necesario para tener una muestra lo suficientemente grande de la distribución posterior.

Regresión clásica rígida - Muestreador de Gibbs

- Se encuentran las distribuciones condicionales completas de $\Theta = \{\beta, \sigma^2\}$:

$$\begin{aligned}\beta \mid \text{resto} &\sim \mathcal{N}_p \left(\left(\frac{\lambda}{\sigma^2} \mathbf{I}_p + \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{X} \right)^{-1} \left(\frac{1}{\sigma^2} \mathbf{X}^T \mathbf{y} \right), \left(\frac{\lambda}{\sigma^2} \mathbf{I}_p + \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{X} \right)^{-1} \right) \\ \sigma^2 \mid \text{resto} &\sim \text{Gl} \left(\frac{\nu_0 + p + n}{2}, \frac{\nu_0 \sigma_0^2 + \text{SSR}(\beta) + \sum_{i=1}^p \beta_i^2}{2} \right) \\ \lambda \mid \text{resto} &\sim \text{G} \left(\mathbf{a}_\lambda + \frac{p}{2}, \mathbf{b}_\lambda + \frac{\lambda}{2} \sum_{i=1}^p \beta_i^2 \right)\end{aligned}$$

- Se inicializa $\Theta^{[0]}$, asignándole un valor en el espacio paramétrico para β , σ^2 y λ . Este valor se puede muestrear de la distribución previa de ambos parámetros.
- Se actualiza el valor de los parámetros del modelo, es decir, $\Theta^{[b]}$ a $\Theta^{[b+1]}$ para $b \in \{0, 1, \dots, B-1\}$, tomando muestras de las distribuciones condicionales completas, es decir:

$$\begin{aligned}\beta \mid \text{resto} &\sim \mathcal{N}_p \left(\left(\frac{\lambda^{[b]}}{(\sigma^2)^{[b]}} \mathbf{I}_p + \frac{1}{(\sigma^2)^{[b]}} \mathbf{X}^T \mathbf{X} \right)^{-1} \left(\frac{1}{(\sigma^2)^{[b]}} \mathbf{X}^T \mathbf{y} \right), \left(\frac{\lambda^{[b]}}{(\sigma^2)^{[b]}} \mathbf{I}_p + \frac{1}{(\sigma^2)^{[b]}} \mathbf{X}^T \mathbf{X} \right)^{-1} \right) \\ \sigma^2 \mid \text{resto} &\sim \text{Gl} \left(\frac{\nu_0 + p + n}{2}, \frac{\nu_0 \sigma_0^2 + \text{SSR}(\beta^{[b+1]}) + \sum_{i=1}^p (\beta_i^{[b+1]})^2}{2} \right) \\ \lambda \mid \text{resto} &\sim \text{G} \left(\mathbf{a}_\lambda + \frac{p}{2}, \mathbf{b}_\lambda + \frac{\lambda}{2} \sum_{i=1}^p (\beta_i^{[b+1]})^2 \right)\end{aligned}$$

4. se repite el paso 3 B veces, hasta convergencia.

Regresión con errores correlacionados - Muestreador de Gibbs con paso de Metropolis

1. Se encuentran las distribuciones condicionales completas de $\Theta = \{\beta, \sigma^2, \rho\}$:

$$\begin{aligned}\beta \mid \text{resto} &\sim \mathbf{N}(\beta_n, \Sigma_n) \\ \Sigma_n &= \left(\Sigma_0^{-1} + \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{C}_\rho \mathbf{X} \right)^{-1} \\ \mu_n &= \Sigma_n \left(\Sigma_0^{-1} \beta_0 + \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{C}_\rho^{-1} \mathbf{y} \right) \\ \sigma^2 \mid \text{resto} &\sim \text{IG} \left(\frac{\nu_0 + n}{2}, \frac{\nu_0 \sigma_0^2 + \text{SSR}_\rho}{2} \right) \\ \text{SSR}_\rho &= (\mathbf{y} - \mathbf{X}\beta)^T \mathbf{C}_\rho^{-1} (\mathbf{y} - \mathbf{X}\beta)\end{aligned}$$

2. Se inicializa $\Theta^{[0]}$, asignándole un valor en el espacio paramétrico para θ y α . Este valor se puede muestrear de la distribución previa de ambos parámetros.
3. Se actualiza el valor de los parámetros del modelo, es decir, $\Theta^{[b]}$ a $\Theta^{[b+1]}$ para $b \in \{0, 1, \dots, B-1\}$, tomando muestras de las distribuciones condicionales completas, es decir:

$$\begin{aligned}^{[b+1]} &\sim \mathbf{N}(\beta_n, \Sigma_n) \\ \Sigma_n &= \left(\Sigma_0^{-1} + \frac{1}{(\sigma^2)^{[b]}} \mathbf{X}^T \mathbf{C}_\rho \mathbf{X} \right)^{-1} \\ \mu_n &= \Sigma_n \left(\Sigma_0^{-1} \beta_0 + \frac{1}{\sigma^2^{[b]}} \mathbf{X}^T \mathbf{C}_\rho^{-1} \mathbf{y} \right) \\ \sigma^{2[b+1]} &\sim \text{IG} \left(\frac{\nu_0 + n}{2}, \frac{\nu_0 \sigma_0^2 + \text{SSR}_\rho}{2} \right) \\ \text{SSR}_\rho &= (\mathbf{y} - \mathbf{X}\beta^{[b+1]})^T \mathbf{C}_\rho^{-1} (\mathbf{y} - \mathbf{X}\beta^{[b+1]})\end{aligned}$$

En este caso, como la distribución condicional completa de ρ es desconocida, se añade un paso de Metrópolis para muestrearla.

- a) Se simula un $\rho^* \sim \mathbf{J}(\rho \mid \rho^{[b]})$, donde \mathbf{J} se denomina distribución de propuestas. Esta distribución es simétrica y, por lo tanto, está centrada en el valor más actual del parámetro, es decir, $\rho^{[b]}$. En este caso,

$$\rho^* \sim \mathbf{N}(\rho^{[b]}, \delta)$$

donde δ es el parámetro de ajuste y se elige para que el algoritmo funcione de manera eficiente.

Como $\rho \in (a, b)$, se le aplica una transformación, es decir, $\eta = \text{logit} \left(\frac{\rho - a}{b - a} \right)$, y en lugar de generar ρ^* se genera η^* para que se respete la restricción paramétrica. Por lo tanto,

$$\eta^* \sim \mathbf{N} \left(\text{logit} \left(\frac{\rho - a}{b - a} \right), \delta \right)$$

- b) Se calcula la tasa de aceptación corregida, dado que se le aplicó una transformación a ρ . Por lo tanto, la tasa de aceptación viene dada por

$$r = \frac{p(\eta^* \mid y)}{p(\eta^{[b]} \mid y)} = \frac{p(h(\eta^*) \mid y)}{p(h(\eta^{[b]}) \mid y)} \frac{h'(\eta^*)}{h'(\eta^{[b]})}$$

Para brindar estabilidad al proceso, se calcula primero en escala logarítmica, es decir,

$$r = \exp \{ \ln(p(\eta^* \mid y)) - \ln(p(\eta^{[b]} \mid y)) \}$$

c) Se actualiza a ρ de la siguiente forma

$$\rho^{[b+1]} = \begin{cases} h(\eta^*), & \text{con probabilidad } \min(1, r); \\ \alpha^{[b]}, & \text{con probabilidad } 1 - \min(1, r). \end{cases}$$

con $h = (b - a)(\expit(x) + \frac{a}{b-a})$ y $h'(x) = (b - a)\frac{e^{-x}}{(1+e^{-x})^2}$, es decir, la inversa de la transformacion aplicada a ρ y su respectiva derivada.

d) Para establecer δ , se realizan 1000 iteraciones previas calculando la tasa de aceptación, es decir,

$$\text{acs} = \frac{N}{1000}$$

con N como el número de veces en que $\alpha^{[b+1]} = \exp \eta^*$. Este proceso se repite hasta que $0.3 < \text{acs} < 0.5$.

e) Una vez establecido δ , se repite el paso 3 B veces, hasta convergencia.

Resumen Coeficientes de Variación de Montecarlo

	Minimo	Q2.5 %	Mediana	Media	Q97.5 %	Maximo
Modelo 1	5.050e-05	7.817e-05	2.482e-04	3.030e-03	1.151e-03	4.316e-02
Modelo 2	0.0003984	0.0006258	0.0019857	0.0235277	0.0094916	0.3565136
Modelo 3	6.090e-06	2.799e-04	3.401e-04	1.175e-03	3.801e-04	5.670e-02
Modelo 4	0.0003473	0.0006042	0.0019778	0.0179075	0.0076465	0.2472472

Tabla 4: Coeficientes de Variación de Montecarlo

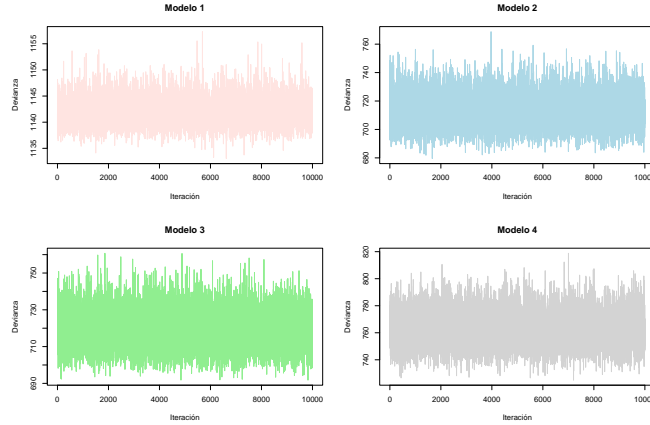


Figura 6: Devianza

Referencias

- Hoff, P. D. (2009). *A first course in Bayesian statistical methods*. Springer.
- Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., Rubin, D. (2013). *Bayesian Data Analysis*. CRC Press.
- Bernardo, J. M., Smith, A. F. (2000). *Bayesian theory*. John Wiley & Sons.