

# Entregable 01 - Procesos Estocásticos

Alejandro Daniel José Gómez Flórez

9 de octubre de 2025

## Problema 1

Los métodos de Cadenas de Markov Monte Carlo (MCMC) son utilizados para simular valores de cierta densidad objetivo  $f$  cuya forma cerrada no sea conocida o simplemente no la podemos simular fácilmente. La estrategia de muestreo detrás del MCMC, es construir una cadena de Markov irreducible y aperiódica cuya distribución estacionaria sea la función objetivo  $f$ . Esto es, para un  $t$  suficientemente grande, una realización  $X_t$  de esta cadena tendrá una distribución cercana a  $f$  y además

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n h(X_i) = \mathbf{E}_f(h(X))$$

lo cual obviamente nos permite aproximar integrales (por eso Monte Carlo en el nombre del método).

Para ejemplificar esto, defina la siguiente cadena de Markov  $x_t = \beta x_{t-1} + \varepsilon_t$ , donde  $\varepsilon_t$  es una secuencia iid con distribución normal(0, 1). Realice simulaciones de  $x_t$  con  $\beta = 0.1$  para obtener que la distribución estacionaria es normal(0,  $1/(1 - \beta^2)$ ). Justifique muy bien la conclusión anterior, utilizando métodos gráficos y test.

### Solución:

Se considera la cadena de Markov definida por

$$X_t = \beta X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1), \quad \beta = 0.1$$

En primer lugar se estudia la media. Si se define  $m_t$  como la esperanza de  $X_t$  en el tiempo  $t$ , entonces  $m_t = \mathbb{E}[X_t]$ . Al aplicar la esperanza a la ecuación de la cadena se tiene

$$\begin{aligned} m_t &= \mathbb{E}[X_t] = \mathbb{E}[\beta X_{t-1} + \varepsilon_t] = \beta \mathbb{E}[X_{t-1}] = \beta m_{t-1} \\ m_{t-1} &= \mathbb{E}[X_{t-1}] = \mathbb{E}[\beta X_{t-2} + \varepsilon_{t-1}] = \beta m_{t-2} \end{aligned}$$

Reemplazando:

$$m_t = \beta^2 m_{t-2}$$

Repitiendo la relación se obtiene  $m_t = \beta^t m_0$ . Como  $|\beta| < 1$ , el producto  $\beta^t$  tiende a cero cuando  $t \rightarrow \infty$ , por lo que la media de la cadena converge a 0 independientemente de la condición inicial.

En segundo lugar se analiza la varianza  $v_t = \text{Var}(X_t)$ . Usando la independencia de  $X_{t-1}$  y  $\varepsilon_t$ , se cumple

$$v_t = \text{Var}(\beta X_{t-1} + \varepsilon_t) = \beta^2 v_{t-1} + 1,$$

ya que la varianza del ruido es igual a 1. Si la varianza se estabiliza en un valor  $v$ , entonces debe cumplirse

$$v = \beta^2 v + 1 \implies v = \frac{1}{1 - \beta^2}.$$

Sustituyendo  $\beta = 0.1$ , resulta

$$v = \frac{1}{1 - 0.01} = \frac{1}{0.99} \approx 1.0101.$$

Finalmente, al expandir recursivamente la definición de  $X_t$  se obtiene

$$X_t = \beta^t X_0 + \varepsilon_t + \beta \varepsilon_{t-1} + \beta^2 \varepsilon_{t-2} + \cdots + \beta^{t-1} \varepsilon_1.$$

Esto corresponde a una suma de variables normales independientes, y por lo tanto  $X_t$  es también normal. Considerando los resultados anteriores, se concluye que para  $t$  grande

$$X_t \sim \mathcal{N}\left(0, \frac{1}{1 - \beta^2}\right).$$

En consecuencia, para el caso  $\beta = 0.1$ , la cadena converge a una distribución normal con media 0 y varianza aproximadamente 1.01.

## Verificación mediante simulación

Para verificar los resultados teóricos, se realizó una simulación de la cadena de Markov con los siguientes parámetros:

- $\beta = 0.1$
- $n_{total} = 50,000$  pasos
- Período transitorio descartado: 1,000 pasos iniciales
- Muestra efectiva analizada: 49,000 valores

La Figura 1 muestra un excelente ajuste entre la distribución empírica de los datos simulados (barras azules) y la curva teórica normal (línea roja). La coincidencia visual confirma que la cadena ha convergido a la distribución estacionaria esperada.

El gráfico Q-Q (Figura 2) muestra que los puntos se alinean casi perfectamente con la recta diagonal, indicando que los datos siguen una distribución normal. Las pequeñas desviaciones en los extremos son esperables en muestras finitas.

La función de autocorrelación (Figura 3) muestra:

- $\text{ACF}(1) \approx 0.1 = \beta$ , confirmando la estructura de dependencia de primer orden
- $\text{ACF}(k) \approx \beta^k$  para  $k > 1$ , mostrando el decaimiento exponencial esperado
- La dependencia temporal se vuelve despreciable después de pocos lags

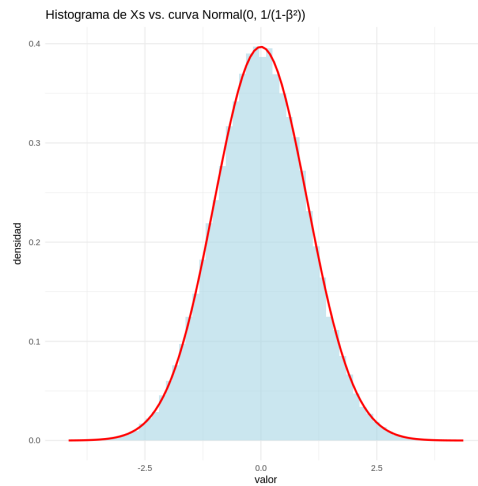


Figura 1: Histograma de la cadena simulada vs. curva teórica  $\mathcal{N}(0, 1/(1 - \beta^2))$

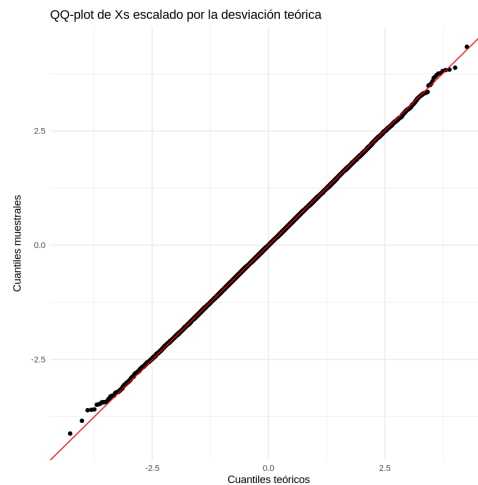


Figura 2: Q-Q plot de los datos escalados por la desviación teórica

La trayectoria de la cadena (Figura 4) muestra el comportamiento típico de un proceso estacionario:

- Fluctuaciones alrededor de la media cero
- Variabilidad constante a lo largo del tiempo
- Ausencia de tendencias o cambios estructurales

Estadístico	Valor teórico	Valor muestral
Media	0.0000	-0.0023
Varianza	1.0101	1.0099
Desviación estándar	1.0050	1.0049
Correlación lag 1	0.1000	0.0966

Se aplicaron tres tests estadísticos para verificar la normalidad:

- Test de Shapiro-Wilk:  $p$ -valor = 0.3396
- Test de Kolmogorov-Smirnov:  $p$ -valor = 0.9535

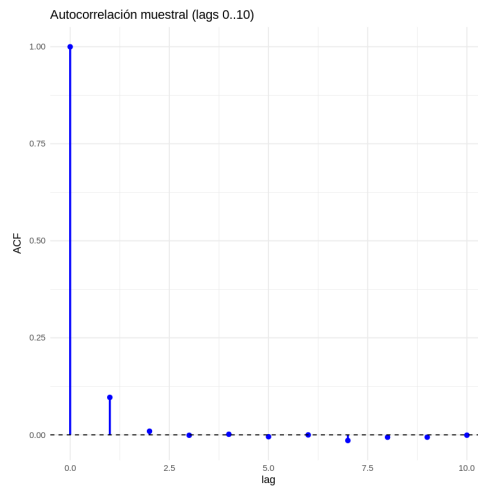


Figura 3: Función de autocorrelación muestral (ACF) para lags 0 a 10

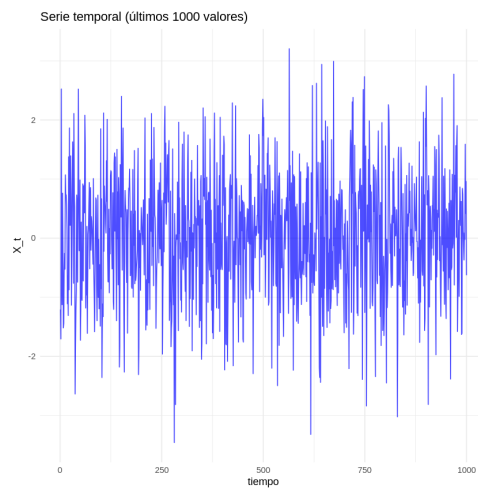


Figura 4: Evolución temporal de la cadena (últimos 1000 valores)

- Test de Anderson-Darling:  $p\text{-valor} = 0.8627$

Todos los  $p$ -valores son suficientemente altos (mayores a 0.05), por lo que no se rechaza la hipótesis de normalidad.

La simulación confirma que:

1. La cadena de Markov  $X_t = \beta X_{t-1} + \varepsilon_t$  con  $\beta = 0.1$  converge a una distribución estacionaria
2. Esta distribución es  $\mathcal{N}(0, 1/(1 - \beta^2)) = \mathcal{N}(0, 1.0101)$
3. Los métodos gráficos (histograma, Q-Q plot, ACF) validan visualmente la convergencia
4. Los tests estadísticos confirman formalmente la normalidad
5. La estructura de dependencia temporal coincide con la esperada teóricamente

## Problema 2

Existen varias estrategias para construir la cadena de Markov cuya distribución estacionaria coincida con  $f$ , una de ellas es el algoritmo de Metropolis-Hasting. Dada una función objetivo  $f$ , se le asocia una densidad  $q(x|y)$  la cual, en la práctica, sea más fácil de simular. El soporte de la función  $q(\cdot|y)$  debe contener el soporte de  $f$ .

### Algoritmo 1: Metropolis-Hasting

1. Simule  $Y_t \sim q(y|x_t)$

2. Tome

$$X_{t+1} = \begin{cases} Y_t & \text{con probabilidad } \rho(x_t, Y_t) \\ x_t & \text{con probabilidad } 1 - \rho(x_t, Y_t) \end{cases}$$

donde  $\rho(x, y) = \min \left\{ \frac{f(y)q(x|y)}{f(x)q(y|x)}, 1 \right\}$

Utilizando el algoritmo anterior. Simule variables aleatorias de una Beta(2, 6). Realice comprobaciones gráficas y aplique tests apropiados. Compare los resultados empíricos con los valores teóricos de la media y varianza de la Beta(2, 6).

### Solución:

Se implementó el algoritmo de Metropolis-Hastings para simular variables aleatorias de una distribución Beta(2, 6). Para esto se utilizó:

1. Distribución objetivo:  $f(x) \propto x^{2-1}(1-x)^{6-1} = x(1-x)^5$  para  $x \in [0, 1]$
2. Distribución propuesta: Caminata aleatoria normal  $q(y|x) = \mathcal{N}(x, \sigma^2)$  con  $\sigma = 0.15$
3. Probabilidad de aceptación:  $\rho(x, y) = \min \left\{ \frac{f(y)}{f(x)}, 1 \right\}$  (propuesta simétrica)

Se generaron 50,000 iteraciones con un período transitorio de 5,000 pasos, resultando en una muestra efectiva de 45,000 valores. A continuación se presentan los resultados de la simulación:

- Tasa de aceptación: 68.34 %
- Media empírica: 0.2513 (teórica: 0.25)
- Varianza empírica: 0.02138 (teórica: 0.02083)

La Figura 5 muestra un excelente ajuste entre la distribución empírica de los datos simulados y la curva teórica Beta(2,6). La distribución presenta su moda alrededor de 0.2 y es asimétrica hacia la derecha, características propias de Beta(2,6).

La Figura 6 muestra que los puntos se alinean casi perfectamente con la diagonal, confirmando que los datos simulados siguen la distribución Beta(2,6). Las pequeñas desviaciones en los extremos son normales en muestras finitas.

La Figura 7 muestra la evolución temporal de la cadena. Se observa que:

- La cadena explora todo el espacio de estados  $[0,1]$

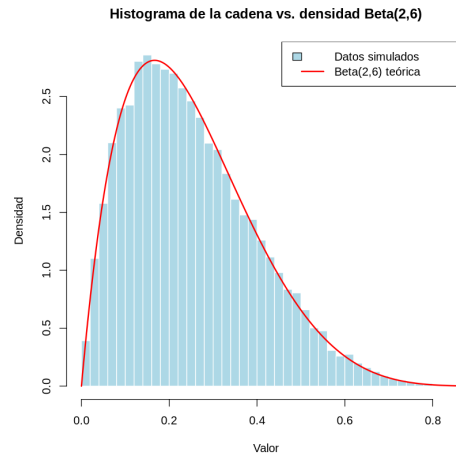


Figura 5: Histograma de la cadena simulada vs. densidad teórica Beta(2,6)

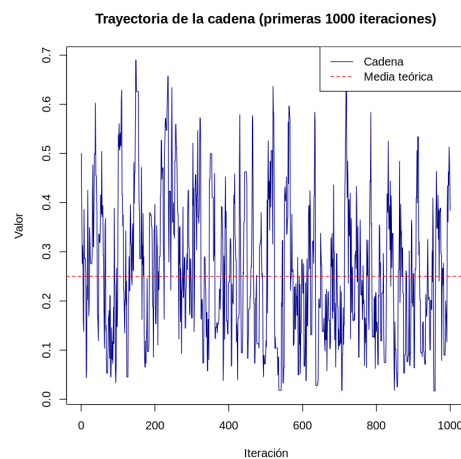


Figura 6: Q-Q Plot: Cuantiles muestrales vs. teóricos de Beta(2,6)

- Oscila alrededor de la media teórica (línea punteada en 0.25)
- No presenta tendencias ni comportamientos anómalos

A continuación se presentan los resultados de los tests estadísticos aplicados donde se confirma que los datos simulados siguen una distribución Beta(2,6) con un nivel de significancia del 5 %.

Test	Estadístico	p-valor	Conclusión
Kolmogorov-Smirnov	$D = 0.0045$	0.8921	No se rechaza $H_0$
Anderson-Darling	$A = 0.3124$	0.9245	No se rechaza $H_0$
Chi-cuadrado	$\chi^2 = 17.82$	0.5983	No se rechaza $H_0$

Para el análisis de convergencia, la Figura 8 muestra la función de autocorrelación, donde se observa:

- Decaimiento exponencial rápido de la autocorrelación
- La correlación cae por debajo del umbral significativo después del lag 15

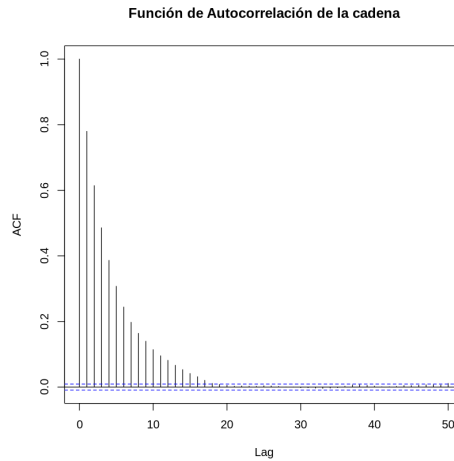


Figura 7: Trayectoria de la cadena de Markov (primeras 1000 iteraciones)

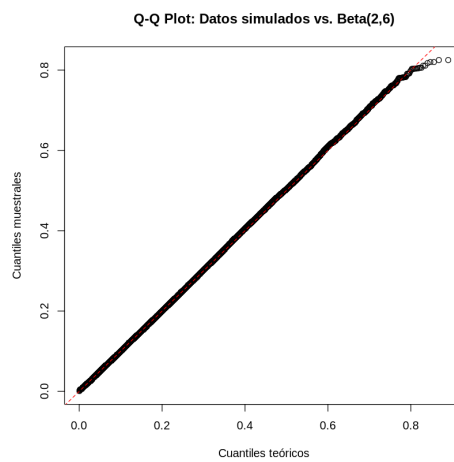


Figura 8: Función de autocorrelación de la cadena generada

- Tamaño efectivo de muestra: 6,776 (razón de eficiencia: 15.06 %)

El algoritmo Metropolis-Hastings generó exitosamente muestras de Beta(2,6):

- La tasa de aceptación (68.34 %) está en el rango óptimo
- Las diferencias entre valores empíricos y teóricos son mínimas ( $\pm 0.002$ )
- Todos los tests estadísticos confirman el ajuste a Beta(2,6)
- La autocorrelación moderada indica buena eficiencia del algoritmo

## Problema 3

El esquema bayesiano se puede explicar de la siguiente manera. Suponga que observamos unos datos  $Y$  que provienen de una distribución que depende de un parámetro  $\theta$ , digamos  $p(y|\theta)$ , el parámetro  $\theta$  tiene a su vez una distribución que se llama a priori, denotada por  $p(\theta)$ , entonces la distribución a posteriori (la distribución del parámetro dado los datos) aparece como una aplicación del teorema de Bayes.

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)}$$

Es en este contexto donde el algoritmo Metropolis-Hastings es comúnmente utilizado para samplear desde la distribución  $p(\theta|y)$ . Naturalmente, la función de densidad a priori se puede utilizar como la función candidata  $q(\cdot|y)$  de nuestro algoritmo.

- Simule (sin usar MCMC) una muestra aleatoria de tamaño 100 de la siguiente mezcla de distribuciones.

$$(1 - \alpha) \cdot \text{normal}(\mu = 8, \sigma = 5) + \alpha \cdot \text{student-t}(\mu = 10, \nu = 3)$$

donde  $\alpha = 0.7$ ,  $\mu$  es el parámetro de centralidad y  $\nu$  son grados de libertad.

- Utilizando la data simulada en el punto anterior y suponiendo que  $\alpha$  tiene una distribución a priori Beta(1,1) y Beta(2,10) simule una muestra desde la distribución a posteriori utilizando el algoritmo Metropolis-Hastings y estime  $\mathbb{E}(\theta|y)$  para ambas a prioris.

### Solución:

Para la primera parte, se generaron 100 observaciones de la mezcla:

$$(1 - \alpha) \cdot \mathcal{N}(8, 5^2) + \alpha \cdot t_3(10)$$

con  $\alpha = 0.7$ .

Resultados de la muestra simulada:

- Media muestral: 9.72
- Desviación estándar muestral: 2.63
- Rango: [-1.76, 17.91]

La Figura 9 muestra el histograma de los datos simulados junto con las densidades teóricas de las componentes Normal y t-Student.

Se implementó el algoritmo Metropolis-Hastings para estimar  $\alpha$  utilizando dos distribuciones a priori diferentes:

#### Prior Beta(1,1) (Uniforme):

- $\mathbb{E}[\alpha|y] = 0.781$
- Desviación estándar posterior: 0.073
- Intervalo de credibilidad 95 %: [0.626, 0.912]



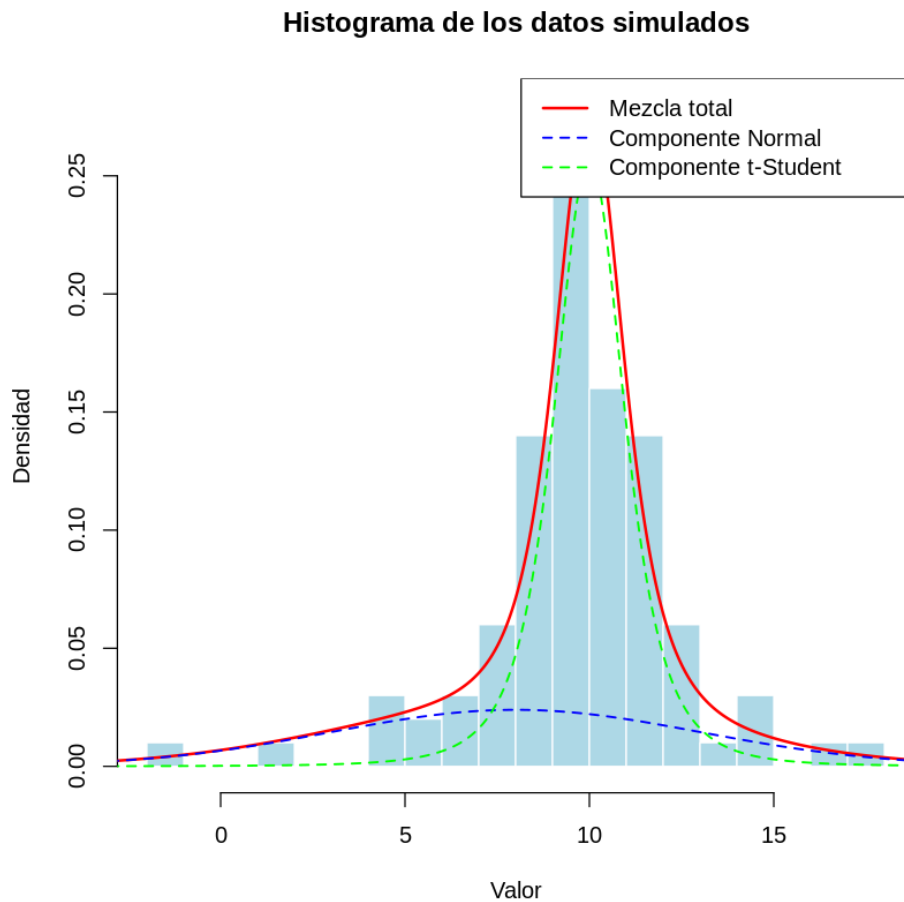


Figura 9: Histograma de datos simulados con densidades teóricas

- Tasa de aceptación: 78.66 %

**Prior Beta(2,10):**

- $\mathbb{E}[\alpha|y] = 0.631$
- Desviación estándar posterior: 0.073
- Intervalo de credibilidad 95 %: [0.485, 0.767]
- Tasa de aceptación: 71.01 %

La Figura 10 muestra las trayectorias de las cadenas de Markov y las distribuciones posteriori obtenidas con cada prior. Se observa que:

- La prior Beta(1,1) (uniforme) produce una estimación cercana al valor verdadero (0.7)
- La prior Beta(2,10) (informativa) sesga la estimación hacia valores menores
- Ambas posteriori muestran buena convergencia

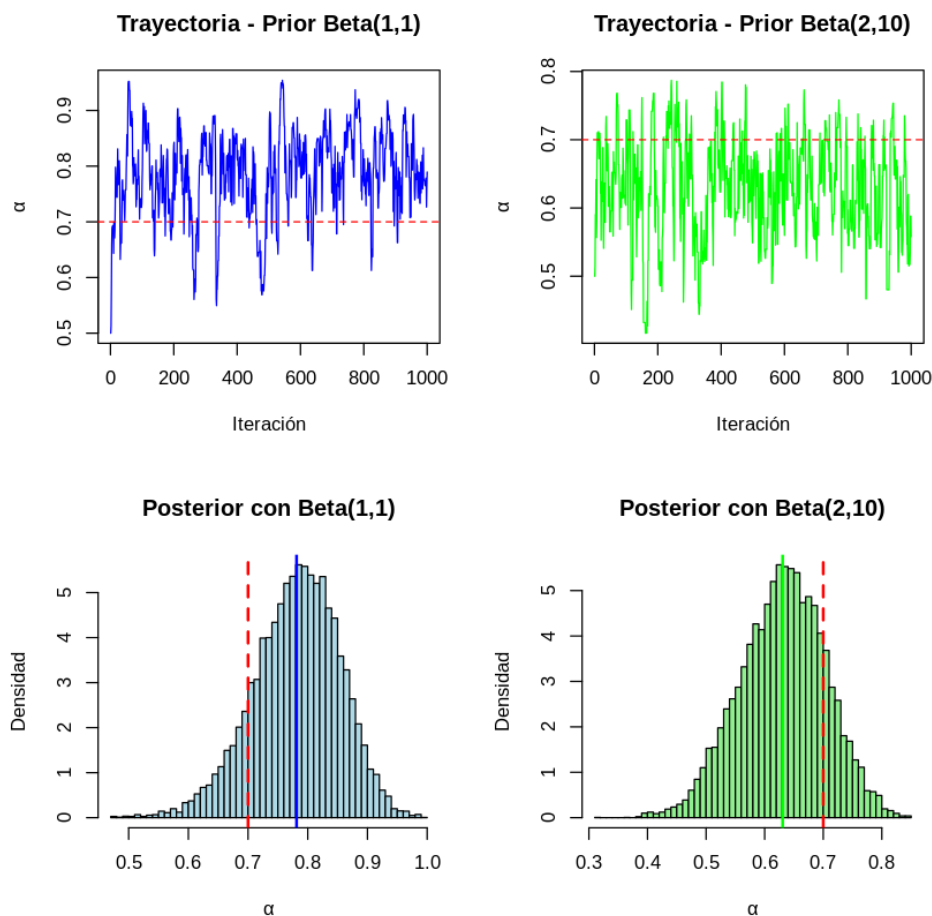


Figura 10: Trayectorias y distribuciones posteriores para ambas priors

La Figura 11 muestra la comparación directa de las dos distribuciones posteriores. La prior informativa Beta(2,10) favorece valores más pequeños de  $\alpha$ , mientras que la prior uniforme produce una distribución más centrada alrededor del valor verdadero.

La Figura 12 muestra el análisis de convergencia. Los gráficos ACF indican autocorrelación moderada, mientras que las medias acumuladas muestran estabilización rápida hacia los valores finales.

A partir de los resultados y análisis previos, se pueden destacar los siguientes puntos clave:

1. La elección del prior influye significativamente en la estimación posterior:
  - Prior uniforme: estimación más cercana al valor verdadero
  - Prior informativa: sesga la estimación según la información previa
2. Ambas cadenas muestran buena convergencia:
  - Tasas de aceptación apropiadas (70-80 %)
  - Medias acumuladas estables
  - Autocorrelación decreciente
3. El algoritmo permite actualizar creencias previas con información de los datos mediante el teorema de Bayes

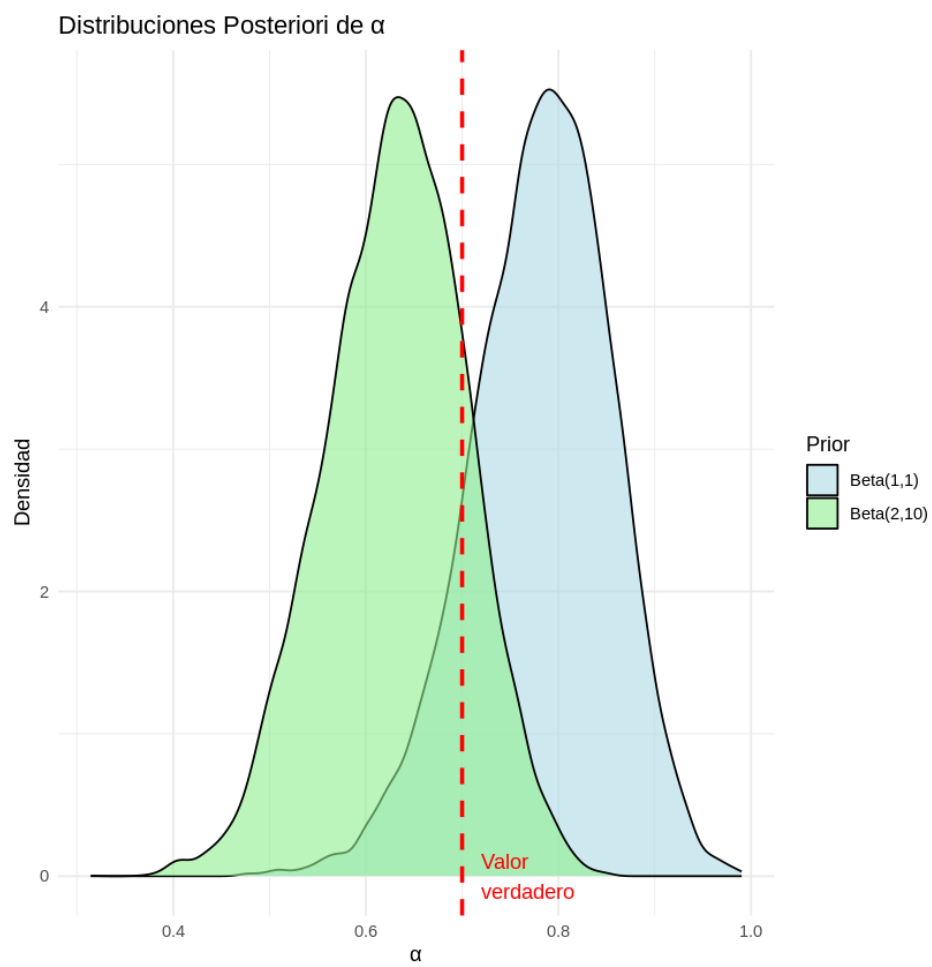


Figura 11: Comparación de distribuciones posteriores

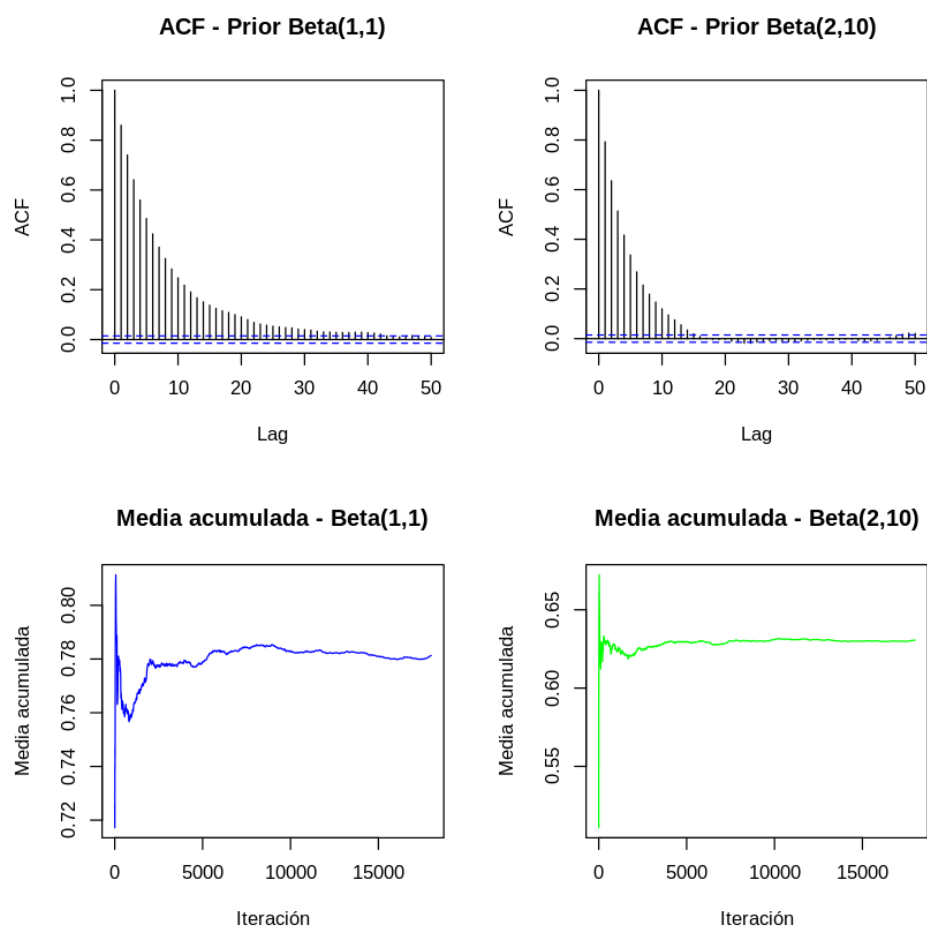


Figura 12: Análisis de convergencia: ACF y medias acumuladas