

Ejercicio 1

Simular $n = 5$ y $n = 50$ v.a's Bernoulli $Be(1/3)$; sea r el número de éxitos en cada caso

Solución

Sea $U \sim \text{Uniforme}(0, 1)$ y sea F_X la función de distribución acumulada de X . Entonces la variable aleatoria

$$Y = F_X^-(U)$$

tiene la misma distribución que X , es decir, $F_Y(x) = F_X(x)$ para todo $x \in \mathbb{R}$, donde F_X^- es la *función cuantil* (inversa generalizada) definida por:

$$F_X^-(u) = \inf\{x : F_X(x) \geq u\}.$$

Sea $X \sim \text{Bernoulli}(p)$, así $P(X = 1) = p$ y $P(X = 0) = 1 - p$.

La función de distribución acumulada $F_X(x)$ es:

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0, \\ 1 - p & \text{si } 0 \leq x < 1, \\ 1 & \text{si } x \geq 1. \end{cases}$$

Para $0 < u < 1$, tenemos que:

1. Si $u \leq 1 - p$, entonces el menor x tal que $F_X(x) \geq u$ es $x = 0$, pues $F_X(0) = 1 - p \geq u$.
2. Si $u \geq 1 - p$, entonces el menor x tal que $F_X(x) \geq u$ es $x = 1$, pues $F_X(1) = 1 \geq u$

Así la inversa generalizada $F_X^-(u)$ es:

$$F_X^-(u) = \begin{cases} 0 & \text{si } 0 < u \leq 1 - p, \\ 1 & \text{si } 1 - p < u < 1. \end{cases}$$

De esta forma para simular de la distribución:

1. Generar $U \sim \text{Uniforme}(0, 1)$
2. Calcular:

$$X = \begin{cases} 0 & \text{si } U \leq 1 - p, \\ 1 & \text{si } U > 1 - p. \end{cases}$$

Tamaño de muestra n	Número de éxitos r	Proporción $\hat{p} = \frac{r}{n}$	Error relativo
5	2	0.4000	20.00 %
50	18	0.3600	8.00 %

Cuadro 1: Resultados de la simulación Bernoulli con $p = \frac{1}{3}$

Ejercicio 2

Implementar el algoritmo Metropolis-Hastings para simular de la posterior

$$f(p|\bar{x}) \propto p^r(1-p)^{n-r} \cos(\pi p) \mathbf{1}_{[0, \frac{1}{2}]}(p)$$

Con los dos casos de n y r de arriba. Para ello poner la propuesta $(p'|p) = p' \sim Beta(r+1, n-r+1)$ y la distribución inicial de la cadena $\mu \sim U(0, \frac{1}{2})$.

Solución

Sean $X_i|p \sim Bernoulli(p)$, con $i = 1, 2, \dots, n$ con distribución previa $\pi(p) = \cos(\pi p)$ para $0 < p < \frac{1}{2}$. Entonces la verosimilitud es:

$$L(\mathbf{x}|p) = \prod_{i=1}^n P(X_i = x_i|p) = p^{\sum x_i} (1-p)^{n-\sum x_i}$$

La distribución posterior, aplicando el teorema de Bayes, está dada por:

$$\begin{aligned} f(p|\mathbf{x}) &\propto L(\mathbf{x}|p) \cdot \pi(p) \\ &\propto p^r(1-p)^{n-r} \cos(\pi p) \mathbf{1}_{0 < p < 1} \cdot \mathbf{1}_{0 < p < \frac{1}{2}} \\ &\propto p^r(1-p)^{n-r} \cos(\pi p) \mathbf{1}_{0 < p < \frac{1}{2}} \end{aligned}$$

donde $r = \sum_{i=1}^n x_i$ es el número total de éxitos.

Para el algoritmo de Metropolis-Hastings con propuesta independiente $q(p'|p) = Beta(p'; r+1, n-r+1)$, el ratio de aceptación es:

$$\begin{aligned} \rho(p, p') &= \min \left\{ 1, \frac{f(p')q(p|p')}{f(p)q(p'|p)} \right\} \\ &= \min \left\{ 1, \frac{f(p')q(p)}{f(p)q(p')} \right\} \\ &= \min \left\{ 1, \frac{p'^r(1-p')^{n-r} \cos(\pi p') \cdot Beta(p; r+1, n-r+1)}{p^r(1-p)^{n-r} \cos(\pi p) \cdot Beta(p'; r+1, n-r+1)} \right\} \\ &= \min \left\{ 1, \frac{p'^r(1-p')^{n-r} \cos(\pi p')}{p^r(1-p)^{n-r} \cos(\pi p)} \left(\frac{p^r(1-p)^{n-r}}{p'^r(1-p')^{n-r}} \right) \right\} \\ &= \min \left\{ 1, \frac{\cos(\pi p')}{\cos(\pi p)} \mathbf{1}_{0 < p, p' < \frac{1}{2}} \right\} \end{aligned}$$

Se consideraron dos escenarios diferentes, utilizando los valores obtenidos en la simulación Bernoulli del problema anterior:

- **Caso 1:** $n = 5, r = 2$ (proporción muestral: $2/5 = 0,4$)
- **Caso 2:** $n = 50, r = 18$ (proporción muestral: $18/50 = 0,36$)

Para la inicialización del algoritmo Metropolis-Hastings, se utilizó una distribución uniforme en el intervalo $[0, \frac{1}{2}]$ como punto inicial de la cadena:

$$p^{(0)} \sim U(0, \frac{1}{2})$$

En la implementación práctica, esto se traduce en generar un valor aleatorio como:

$$p^{(0)} = 0,3482$$

Esta elección garantiza que el punto inicial se encuentre dentro del soporte de la distribución posterior mientras mantiene la aleatoriedad en la inicialización..

Ejercicio 3

Argumentar porque la cadena es f -irreducible y porque es ergódica. Implementar el algoritmo con los datos descritos y discutir los resultados.

Demostración

La cadena de Markov generada por el algoritmo Metropolis-Hastings es f -irreducible, lo que significa que puede alcanzar cualquier conjunto $A \subset [0, \frac{1}{2}]$ con probabilidad positiva $\int_A f(p|\mathbf{x})dp > 0$ en un número finito de pasos, pues la distribución propuesta Beta($r + 1, n - r + 1$) tiene soporte en $(0, 1)$, que contiene al soporte de la distribución objetivo $[0, \frac{1}{2}]$, así el ratio de aceptación es estrictamente positivo.

Además la cadena es fuertemente aperiódicadebido a que existe probabilidad positiva de generar propuestas fuera del soporte de la distribución objetivo, pues la distribución propuesta Beta($r + 1, n - r + 1$) genera valores en $(0, 1)$, así existe probabilidad positiva, $\mathbb{P}(p' > \frac{1}{2}) > 0$, de proponer valores fuera del soporte $[0, \frac{1}{2}]$.

De esta manera tenemos que es f -irreducible y fuertemente aperiodica, así es ergódica, por lo tanto la distribución objetivo $f(p|X)$ satisface el balance detallado, lo que la convierte en la distribución estacionaria única.

Para los datos obtenidos de la distribución Bernouilli,el algoritmo se ejecutó con $N = 10000$ iteraciones. Respecto al periodo de *burn-in*, el análisis de las cadenas revela que:

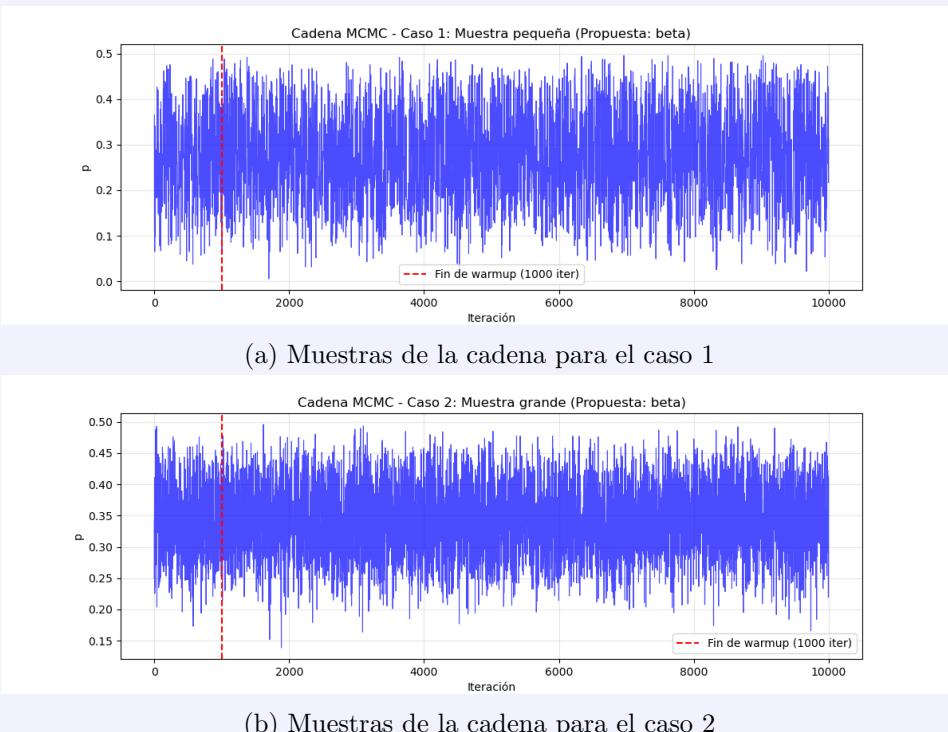
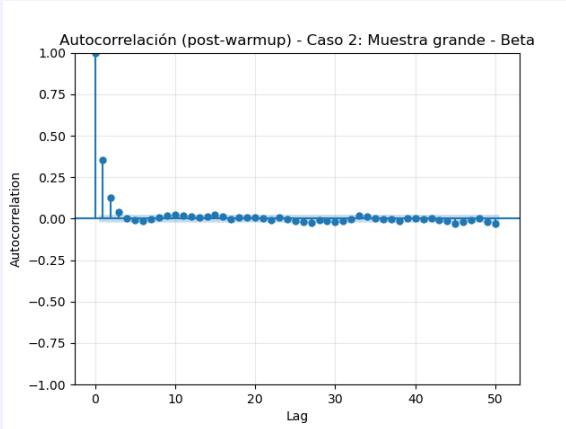


Figura 1: Serie de tiempo para los valores de la cadena en los Casos 1 y 2

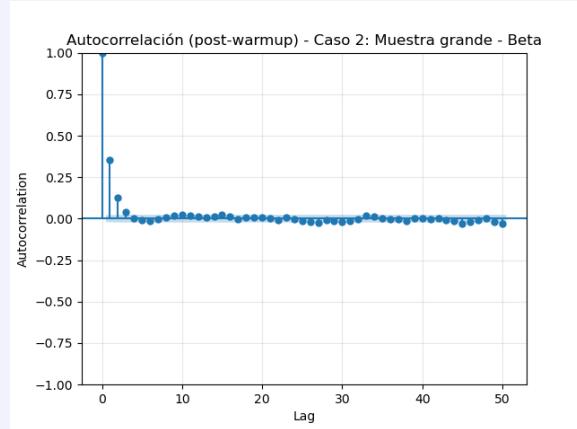
El análisis de las Figuras 1a y 1b muestra que ambas cadenas alcanzan la distribución estacionaria desde las primeras iteraciones, sin mostrar un periodo de adaptación prolongado, pues la elección de la distribución propuesta $q(p'|p) = \text{Beta}(r + 1, n - r + 1)$ resulta ser una excelente aproximación a la distribución posterior objetivo, facilitando una convergencia rápida.

A pesar de esta convergencia rápida, se decidió implementar un periodo de *burn-in* de 1000 iteraciones para cada cadena, lo cual representa aproximadamente el 10 % del total de iteraciones.

El análisis de autocorrelación de las cadenas MCMC, mostrado en la Figura 2, revela el siguiente comportamiento:



(a) Autocorrelación para el caso 1 de la cadena sin adelgazamiento



(b) Autocorrelación para el caso 2 de la cadena sin adelgazamiento

Figura 2: Autocorrelación de la cadena para los casos 1 y 2

Los correlogramas indican que no es necesario aplicar adelgazamiento a las cadenas, pues en ambos casos, la función de autocorrelación decae rápidamente, alcanzando valores dentro de las bandas de significancia estadística después de pocos lags, obteniendo así un tamaño de muestra efectivo de 9000 muestras.

Finalmente, a modo de confirmación visual, dado que la distribución original no está normalizada, las constantes de normalización calculadas mediante integración numérica fueron:

$$c_1 = \int_0^{1/2} p^2(1-p)^3 \cos(\pi p) dp = 5,317811 \times 10^{-3}$$

$$c_2 = \int_0^{1/2} p^{18}(1-p)^{32} \cos(\pi p) dp = 4,386223 \times 10^{-16}$$

Los histogramas de las muestras obtenidas mediante el algoritmo Metropolis-Hastings muestran la forma de las distribuciones posteriores para cada caso:

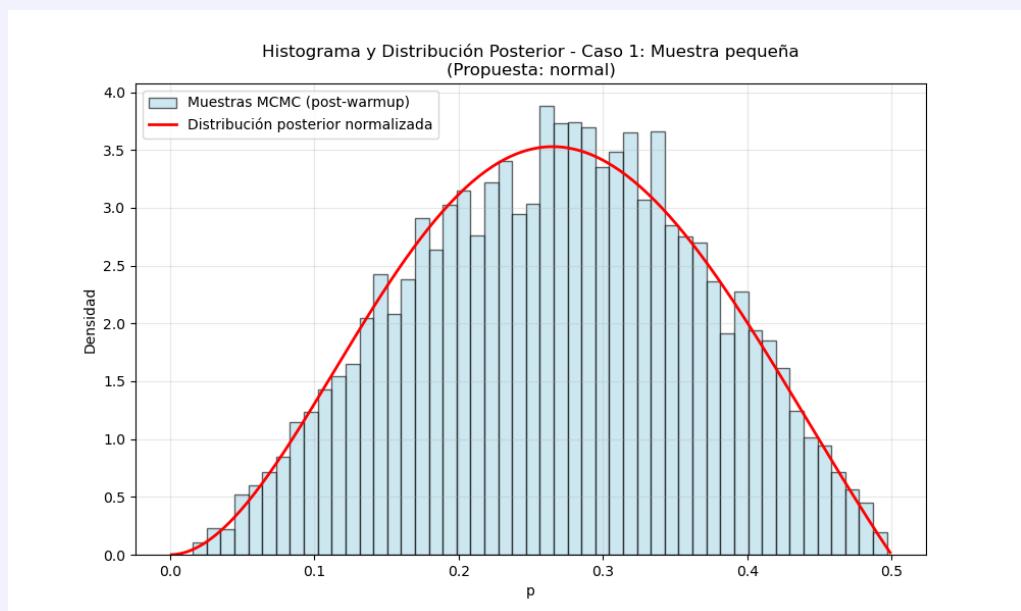


Figura 3: Histograma de la muestra simulada mediante MCMC para el caso 1 ($n = 5, r = 2$).

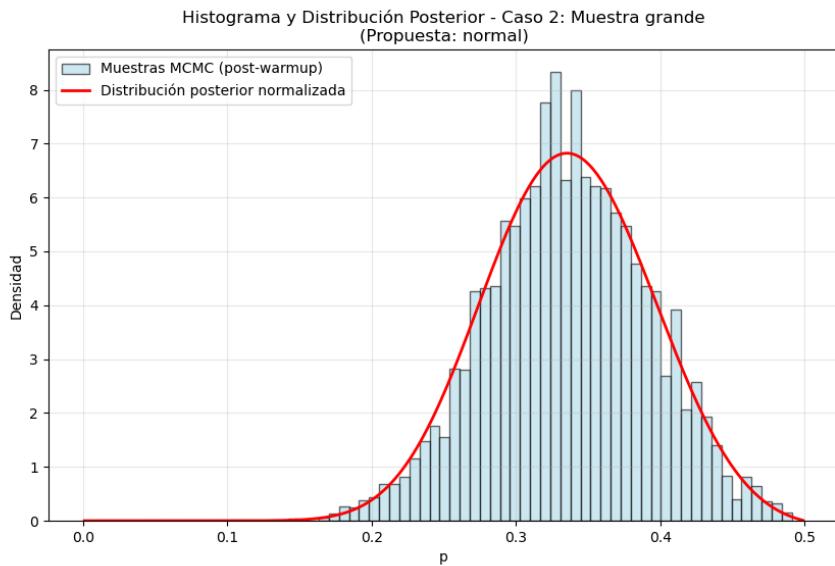


Figura 4: Histograma de la muestra simulada mediante MCMC para el caso 2 ($n = 50$, $r = 18$).

La distribución posterior mostrada en la Figura 3,muestrá una distribución mas dispersa en comparación a la mostrada en la Figura 4 la cual esta más concentrada alrededor de la media, reflejando la mayor información proporcionada por la muestra grande.

Caso	Tasa de Aceptación Global	Media Posterior	Constante normalización
$n = 5, r = 2$	0.4415	0.2663	$5,317 \times 10^{-3}$
$n = 50, r = 18$	0.7417	0.3359	$4,38 \times 10^{-16}$

Cuadro 2: Estadísticas de las cadenas MCMC para ambos casos.

Se observa una mayor tasa de aceptación en el caso 2, indicando que la propuesta Beta($r + 1$, $n - r + 1$) es más efectiva cuando el tamaño de muestra es grande, además la notable diferencia en las constantes de normalización entre ambos casos (10^{-3} vs 10^{-16}) refleja que la distribución posterior del caso 1 cubre un área significativa bajo la curva, indicando mayor incertidumbre y dispersión, mientras que en caso 2 esta más concentrada, resultando en un área total muy pequeña bajo la curva no normalizada.

Ejercicio 4

Implementar el algoritmo Metropolis-Hastings con la posterior de arriba tomando una propuesta diferente.

Demostración (Propuesta Normal)

Consideremos ahora una propuesta alternativa de tipo normal, independiente definida como:

$$q(p'|p) = q(p') \sim \mathcal{N}(\mu = 0,25, \sigma^2),$$

centrada en el punto medio del soporte $[0, \frac{1}{2}]$ con el fin de mejorar la exploración de la región de mayor probabilidad a posteriori.

En este caso, el cociente de aceptación general del algoritmo de Metropolis–Hastings se expresa como:

$$\rho(p, p') = \min \left\{ 1, \frac{f(p') q(p)}{f(p) q(p')} \right\}.$$

Sustituyendo las expresiones explícitas de f y q , se obtiene:

$$\rho(p, p') = \min \left\{ 1, \frac{p'^r (1 - p')^{n-r} \cos(\pi p') \phi(p; 0,25, \sigma^2)}{p^r (1 - p)^{n-r} \cos(\pi p) \phi(p'; 0,25, \sigma^2)} \right\} \mathbf{1}_{[0, \frac{1}{2}]}(p'),$$

donde $\phi(x; \mu, \sigma^2)$ denota la densidad de la normal:

$$\phi(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right),$$

Las propuestas fuera del intervalo $[0, \frac{1}{2}]$ son rechazadas automáticamente, pues la distribución posterior se anula fuera de dicho soporte.

Como en el caso anterior, la cadena se inicializa con un valor aleatorio:

$$p^{(0)} \sim U(0, \frac{1}{2}),$$

garantizando que el punto inicial se encuentre dentro del soporte de la posterior.

Se analizaron los mismos escenarios que en la propuesta Beta:

- **Caso 1:** $n = 5, r = 2$ (proporción muestral $2/5 = 0,4$),
- **Caso 2:** $n = 50, r = 18$ (proporción muestral $18/50 = 0,36$).

En ambos casos, la propuesta fue $q(p') = \mathcal{N}(0,25, \sigma^2)$ con $\sigma = 0,15$, truncada implícitamente a $[0, 0,5]$.

La Figura 5 muestra la evolución temporal de las cadenas para ambos escenarios, donde se aprecia que las cadenas alcanzan la región estacionaria de la distribución posterior rápidamente, sin un periodo de adaptación prolongado. No obstante, se aplicó un periodo de *burn-in* de 1000 iteraciones (10 % del total) por consistencia metodológica.

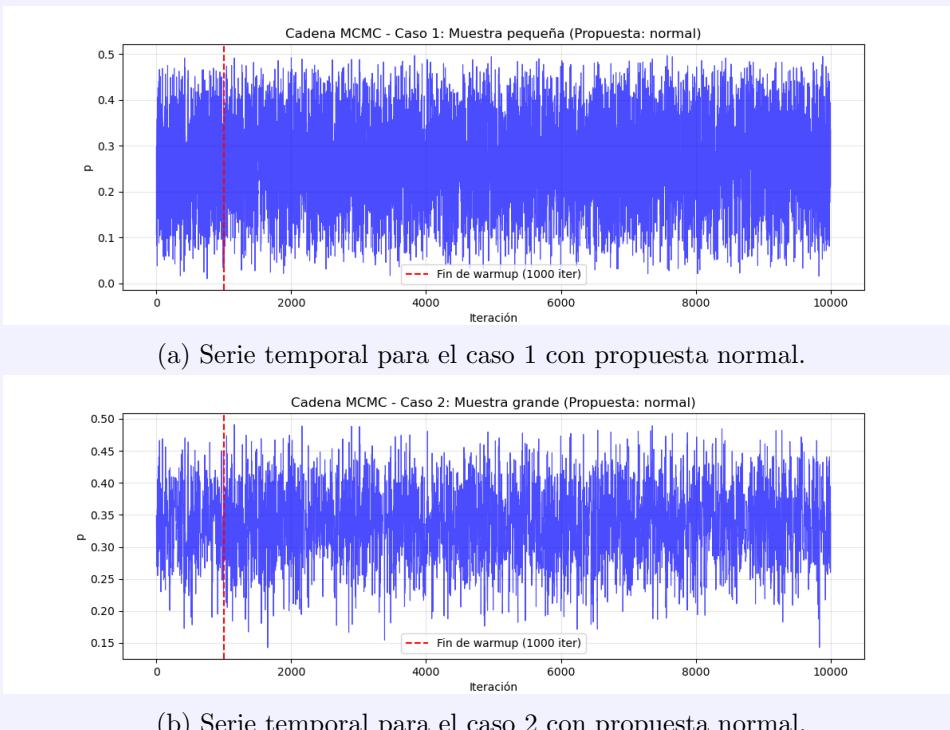
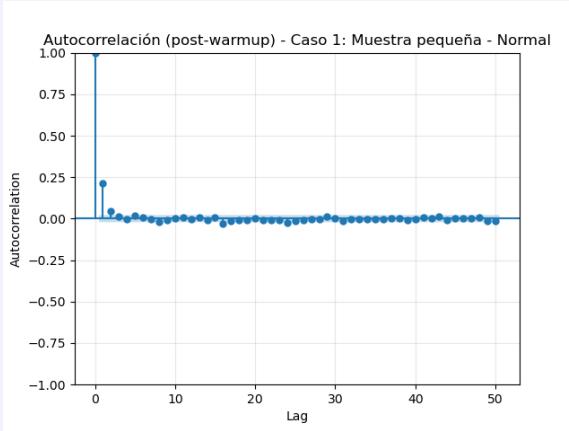
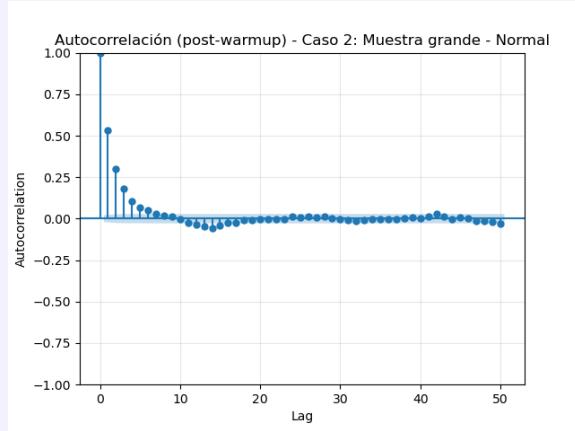


Figura 5: Evolución de las cadenas MCMC bajo propuesta normal en ambos casos.

El análisis de autocorrelación (Figura 6) muestra una rápida caída de la dependencia entre muestras, lo que hace innecesario aplicar adelgazamiento y asegura un tamaño efectivo cercano a las 9000 observaciones por cadena.



(a) Autocorrelación — Caso 1.



(b) Autocorrelación — Caso 2.

Figura 6: Correlogramas de las cadenas con propuesta normal.

Las distribuciones posteriores simuladas mediante MCMC se muestran en las Figuras 7 y 8. Se observa que:

- El caso con $n = 5$ exhibe una mayor dispersión, indicando mayor incertidumbre.
- El caso con $n = 50$ muestra una concentración notable alrededor de la media posterior, reflejando la ganancia de información al aumentar el tamaño muestral.

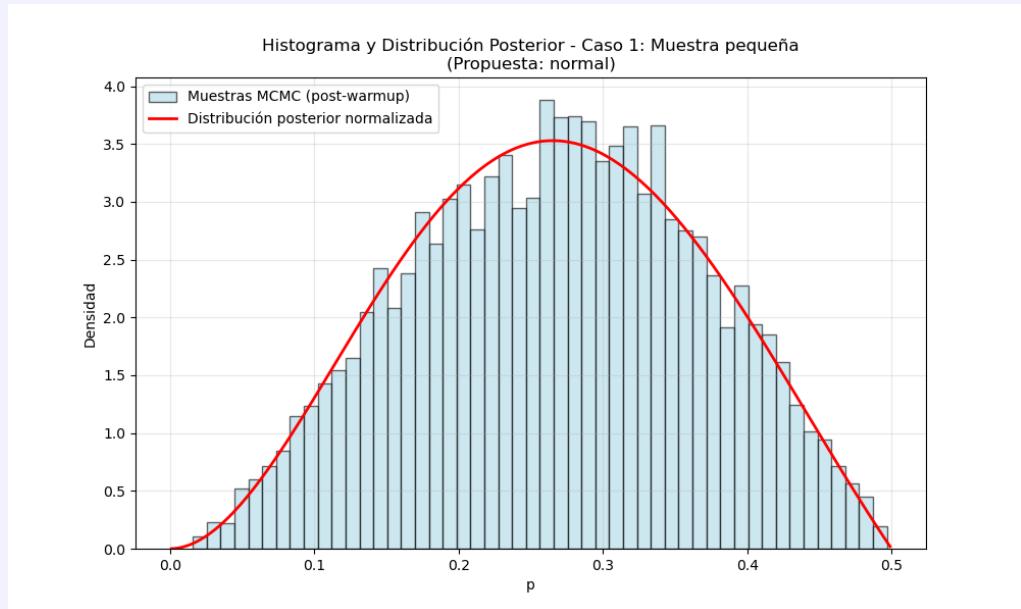


Figura 7: Distribución posterior simulada para el caso 1 ($n = 5, r = 2$) con propuesta normal.

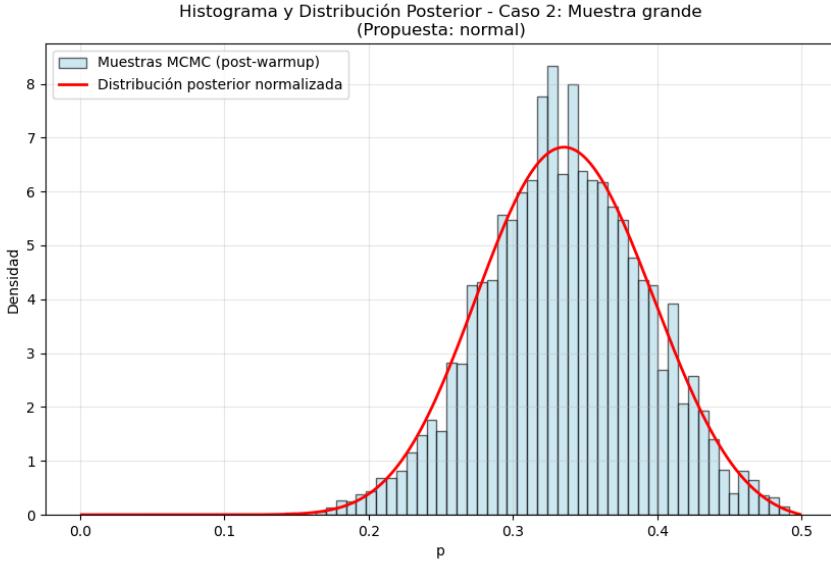


Figura 8: Distribución posterior simulada para el caso 2 ($n = 50, r = 18$) con propuesta normal.

Caso	Tasa de Aceptación Global	Media Posterior	Constante de Normalización
$n = 5, r = 2$	0.7649	0.2675	$5,317 \times 10^{-3}$
$n = 50, r = 18$	0.3884	0.3359	$4,38 \times 10^{-16}$

Cuadro 3: Estadísticas resumen de las cadenas MCMC bajo propuesta normal.

Se observa una tasa de aceptación superior al 70 % en el caso con menor información, lo cual resulta coherente, ya que la mayor parte de la masa de la distribución posterior se concentra alrededor de ese punto. En consecuencia, una propuesta normal centrada en 0.25, lo que favorece una alta tasa de aceptación y una exploración adecuada del soporte de la distribución. En contraste, cuando $n = 50$ la tasa de aceptación disminuye, pues la posterior se concentra en una región alejada de la media de la propuesta normal. Dado que la propuesta es simétrica, genera con mayor frecuencia valores fuera de la zona de alta probabilidad, reduciendo así la eficiencia del muestreo.

De esta manera se tiene que la aproximación normal resulta adecuada cuando la masa de la distribución posterior está centrada en una región conocida y la forma es aproximadamente simétrica. Sin embargo, si la media se aleja , la propuesta normal deja de ser apropiada. En tales casos, una distribución Beta es más conveniente, pues respeta el soporte $[0, 1]$ y se ajusta mejor a la forma real de la posterior.

Dado que los datos fueron simulados, los resultados reflejan la inferencia basada en una muestra particular y no necesariamente la probabilidad real $p = \frac{1}{3}$. El algoritmo estima el parámetro a partir de la información contenida en esa muestra, no obstante como se ve en el Cuadro 2 y 3 conforme el tamaño muestral aumenta, la estimación converge al valor teórico, validando la coherencia del procedimiento.