

Con el algoritmo de Metropolis-Hastings (MH), simular lo siguiente:

### Ejercicio 1

Sean  $x_i \sim Ga(\alpha, \beta)$   $i = 1, 2, \dots, n$ . Simular datos  $x_i$  con  $\alpha = 3$  y  $\beta = 100$  considerando los casos  $n = 4$  y 30.

Con  $\alpha \sim U(1, 4)$ ,  $\beta \sim exp(1)$  distribuciones a priori, se tiene la posterior

$$f(\alpha, \beta | \bar{x}) \propto \frac{\beta^{n\alpha}}{\Gamma(\alpha)^n} r_1^{\alpha-1} e^{-\beta(r_2+1)} \mathbb{1}_{1 \leq \alpha \leq 4} \mathbb{1}_{\beta > 1}$$

con  $r_2 = \sum_{i=1}^n x_i$  y  $r_1 = \prod_{i=1}^n x_i$ . En ambos casos, grafica los contornos para visualizar dónde está concentrada la posterior. Utilizar la propuesta

$$q\left(\begin{pmatrix} \alpha_p \\ \beta_p \end{pmatrix} \middle| \begin{pmatrix} \alpha \\ \beta \end{pmatrix}\right) = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

donde

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right)$$

### Solución:

Se implementa el algoritmo de Metropolis-Hastings con caminatas aleatorias para la posterior mostrada en el enunciado para cada caso, es decir,  $n = 4$  y  $n = 30$

Se procede definiendo una función que calcule la distribución posterior según ciertos valores de  $\alpha, \beta$ . Como es una función bivariada, y con objetivo de visualizar la distribución posterior, es que se grafican las curvas de nivel. Notemos que el soporte está en  $1 \leq \alpha \leq 4$  y  $\beta > 0$ . Tenemos las siguientes gráficas

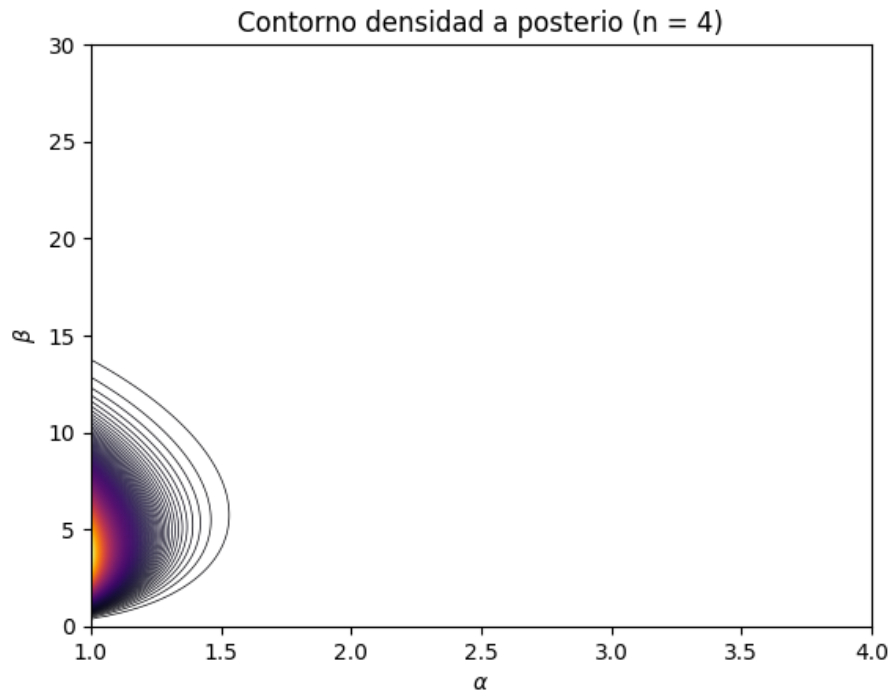


Figura 1: Grafica de contornos para  $n = 4$

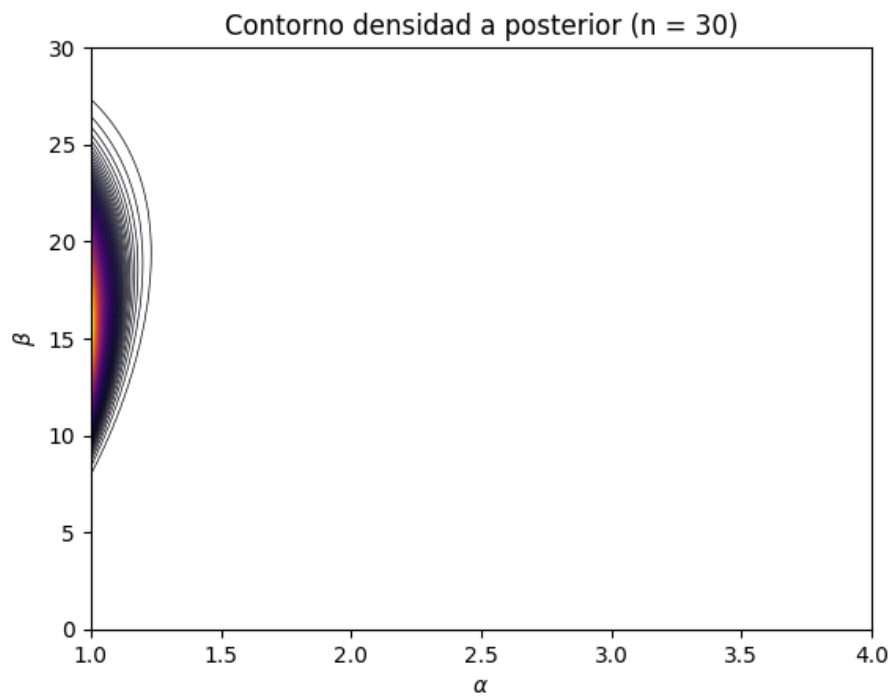


Figura 2: Grafica de contornos para  $n = 30$

Notemos que tenemos una concentración para  $\alpha$  pegados al extremo izquierdo del soporte, mientras que para  $\beta$  se aglomeran por 5 o 15 dependiendo de  $n$ .

Aplicando el método de Metropolis-Hastings con caminata aleatoria y punto inicial en  $\alpha = 3, \beta = 40$  y propuesta como es indicada en el enunciado

$$\varepsilon_1 \sim N(0, \sigma_1^2) \quad \varepsilon_2 \sim N(0, \sigma_2^2)$$

con  $\sigma_1 = 0.05$  y  $\sigma_2^2 = 0.5$

Como la propuesta es simétrica, la probabilidad de transición es solamente el mínimo entre 1 y el cociente de la densidad a posteriori  $f(\alpha_p, \beta_p | x) / f(\alpha, \beta | x)$  con  $\alpha_p, \beta_p$  el siguiente valor propuesto. De esta forma, la cadena seguirá una caminata aleatoria en el plano como se puede ver en la siguiente figura

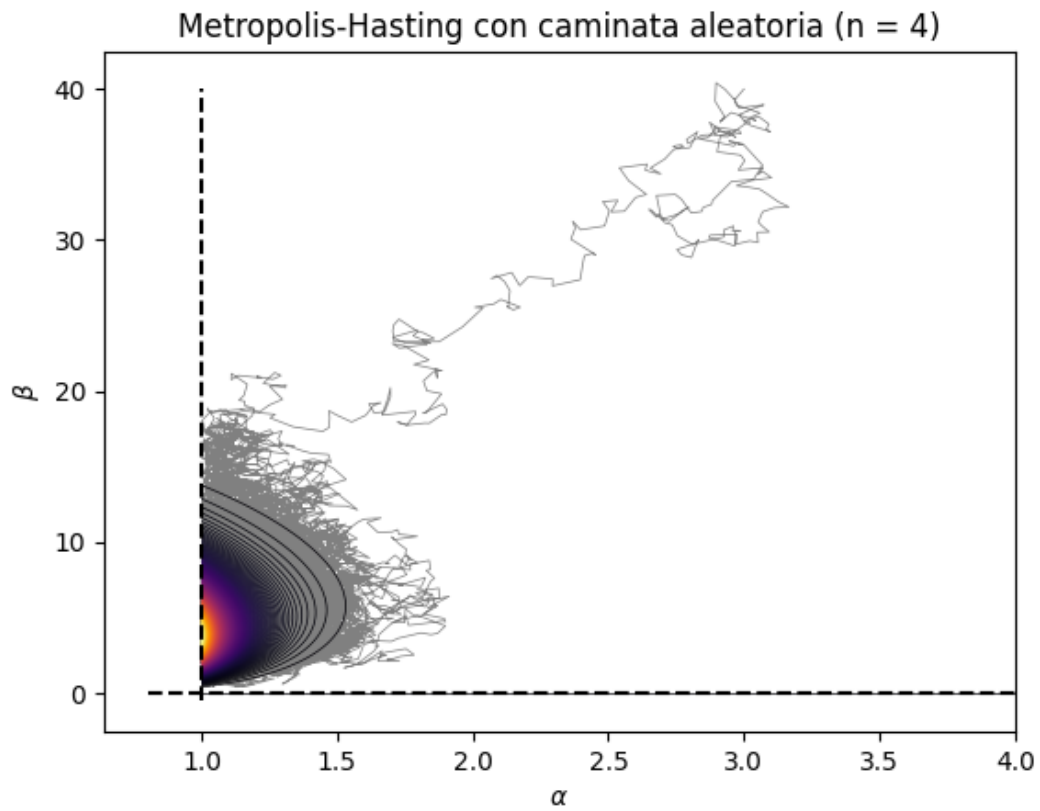


Figura 3: Metropolis-Hastings con caminata aleatoria para la distribución a posteriori con  $n = 4$  donde la distribución propuesta es una normal bivariada.

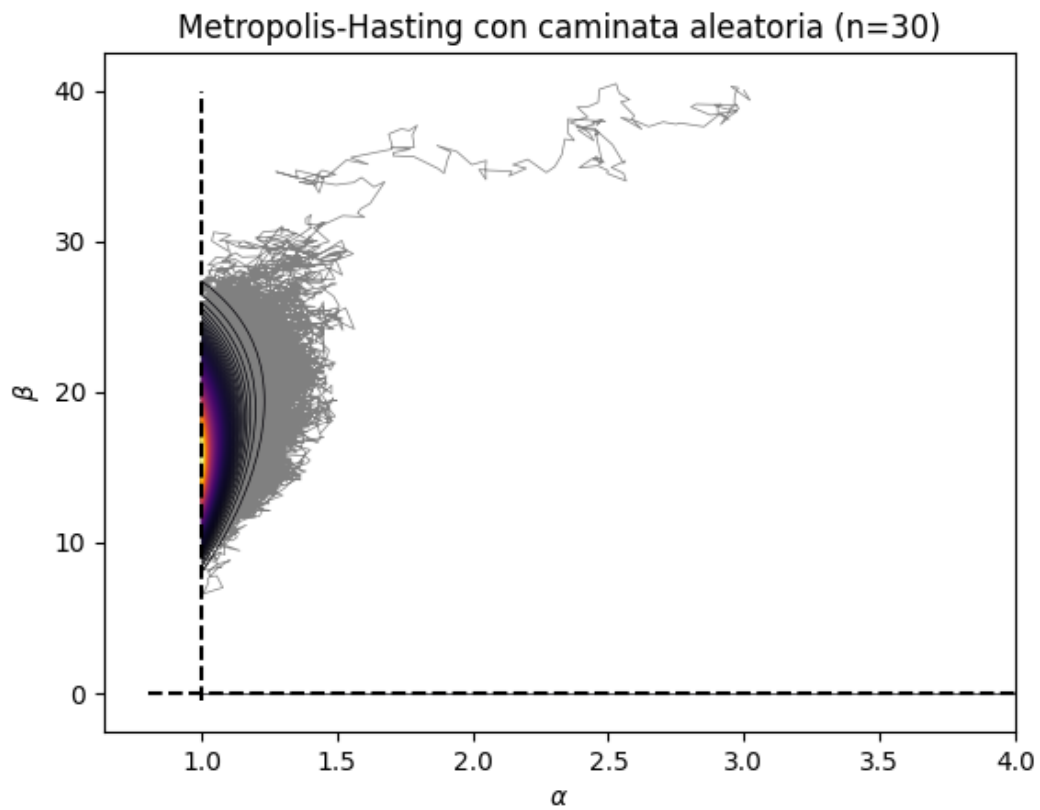


Figura 4: Metropolis-Hastings con caminata aleatoria para la distribución a posteriori con  $n = 30$  donde la distribución propuesta es una normal bivariada.

Posteriormente, se nos pide observar la evolución de la cadena de Markov

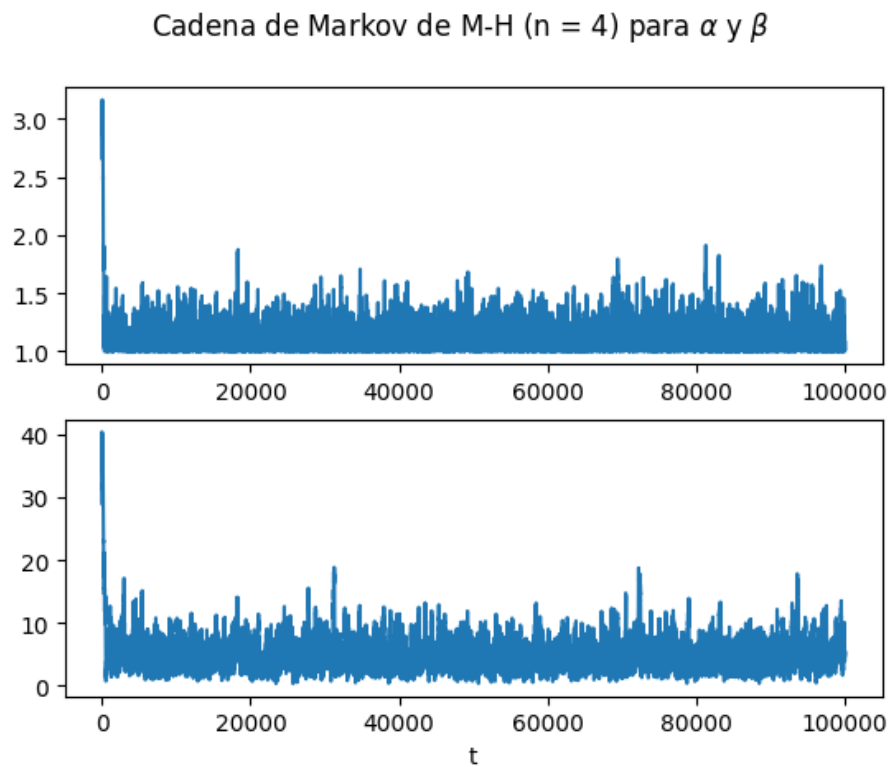


Figura 5: Cadena de Markov de Metropolis-Hastings

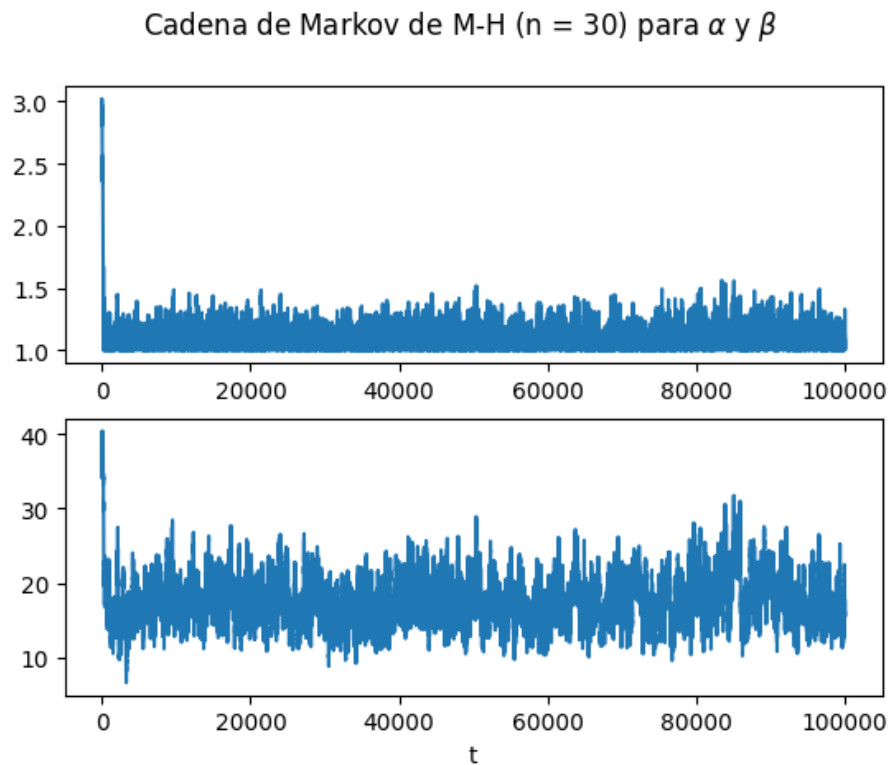


Figura 6: Cadena de Markov de Metropolis-Hastings

Notemos que tenemos una especie de convergencia hacia dentro de las curvas de nivel, es decir, se aproxima a simulaciones de la distribución objetivo. Para poder hablar de que la cadena producida es en efecto una muestra de la distribución objetivo, es necesario eliminar las primeras observaciones atípicas. Esto es llamado burn-in y para proceder calculamos el logaritmo de la densidad objetivo en cada punto de la cadena. Podemos ver en las siguientes figuras que tiende a estabilizarse

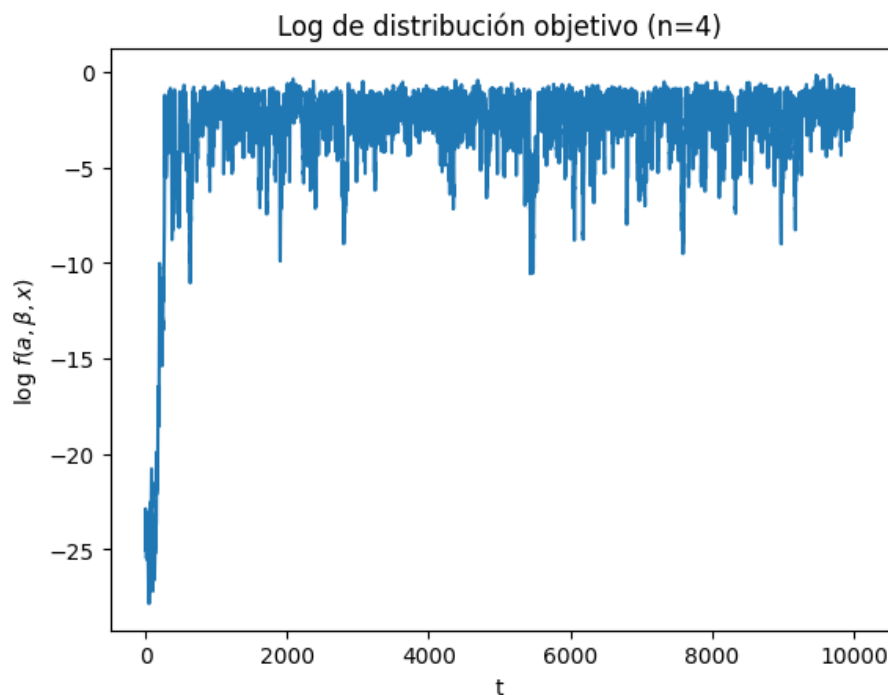


Figura 7: Grafica del logaritmo de la densidad objetivo evaluado en cada punto de la cadena de Markov

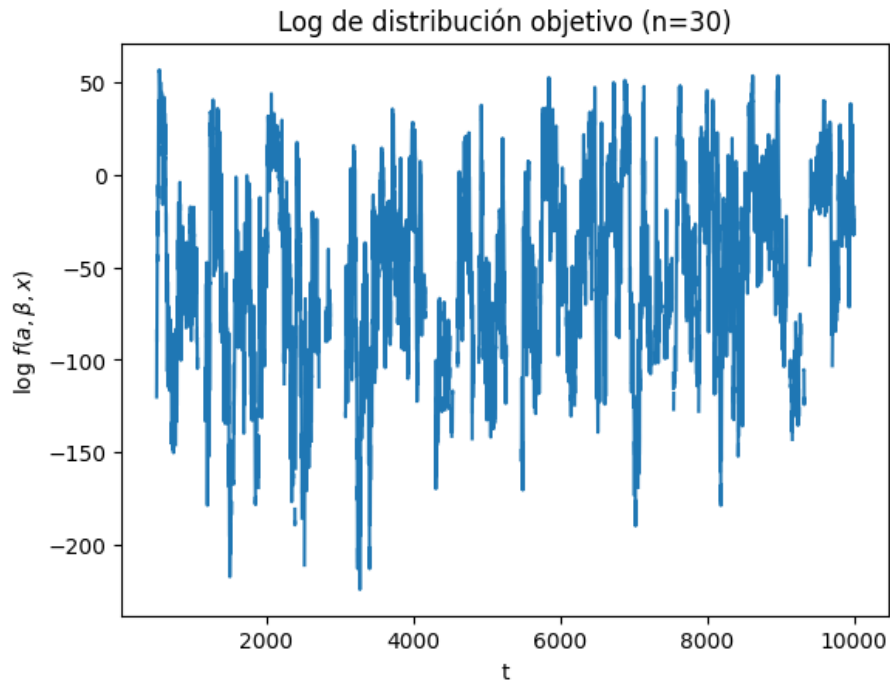


Figura 8: Grafica del logaritmo de la densidad objetivo evaluado en cada punto de la cadena de Markov

Para seleccionar el Burn-in notamos que para  $n = 4$  tenemos varios valles, entre ellos y más notable es el del extremos izquierdo, y un valle más pequeño pero notable por los 5000. Por tanto, elegimos un burn-in de 6000.

Posterior a remover las primeras 6000 iteraciones, hacemos un histograma de los valores observados para las marginales de  $\alpha$  y  $\beta$ .

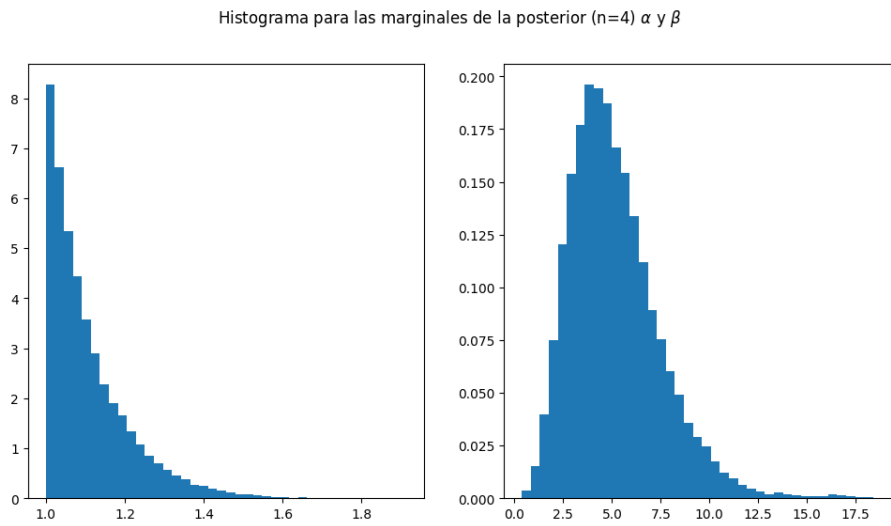


Figura 9: Histograma para  $\alpha$  y  $\beta$  en  $n = 4$  de M-H

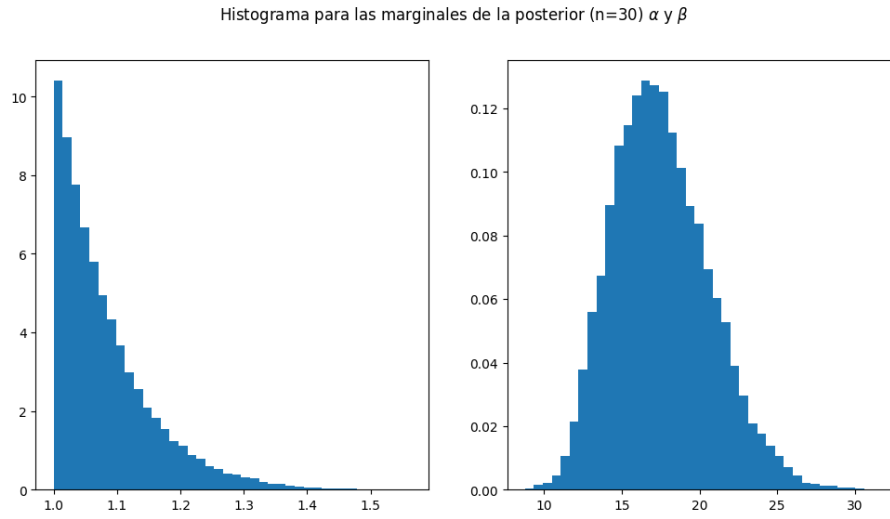


Figura 10: Histograma para  $\alpha$  y  $\beta$  en  $n = 30$  de M-H

Por último, consideramos otra distribución propuesta. Seleccionamos ahora un rectángulo uniforme. Es decir, sea  $e_1 \sim U(-\varepsilon_1, \varepsilon)$  y  $e_2 \sim U(-\varepsilon_2, \varepsilon_2)$  de forma que tomamos

$$\begin{pmatrix} \alpha_p \\ \beta_p \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}$$

Para este análisis nos restringimos a revisar la trayectoria dada por el método Metropolis-Hastings que se muestra en las figuras siguientes

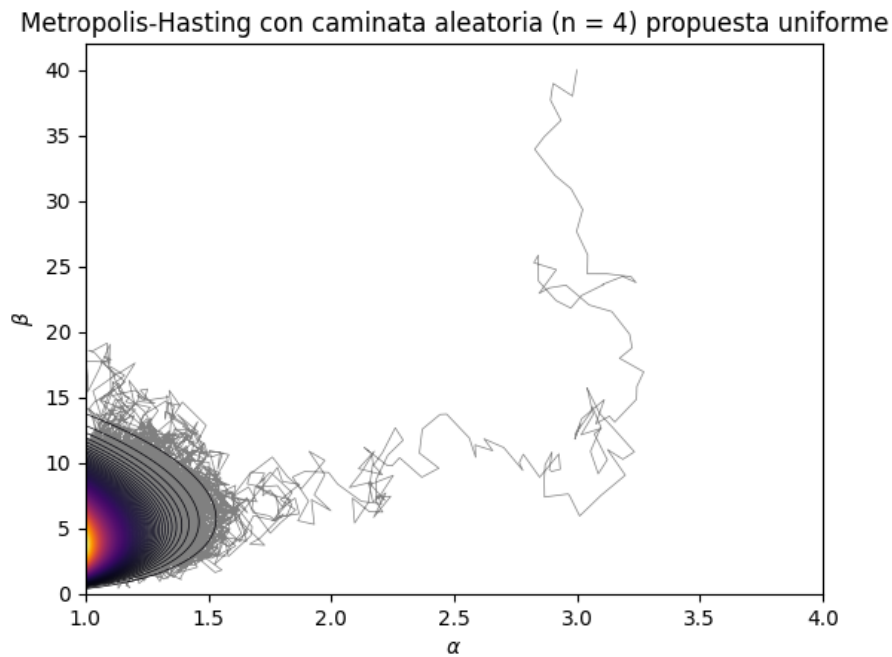


Figura 11: Metropolis-Hastings con caminata aleatoria para la distribución objetivo con  $n = 4$

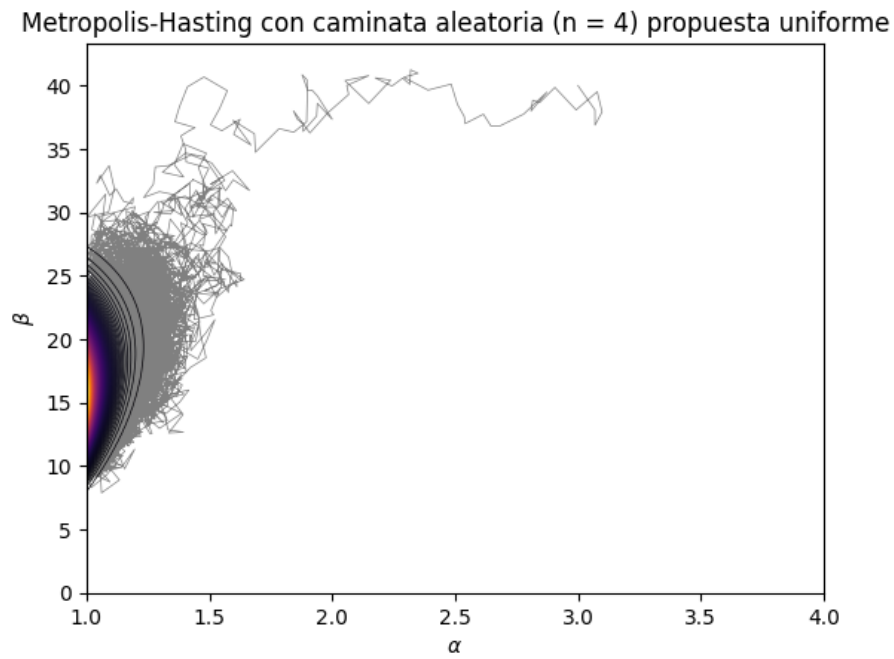


Figura 12: Metropolis-Hastings con caminata aleatoria para la distribución objetivo con  $n = 4$

Luego, la grafica para la evolución de la cadena de Markov es

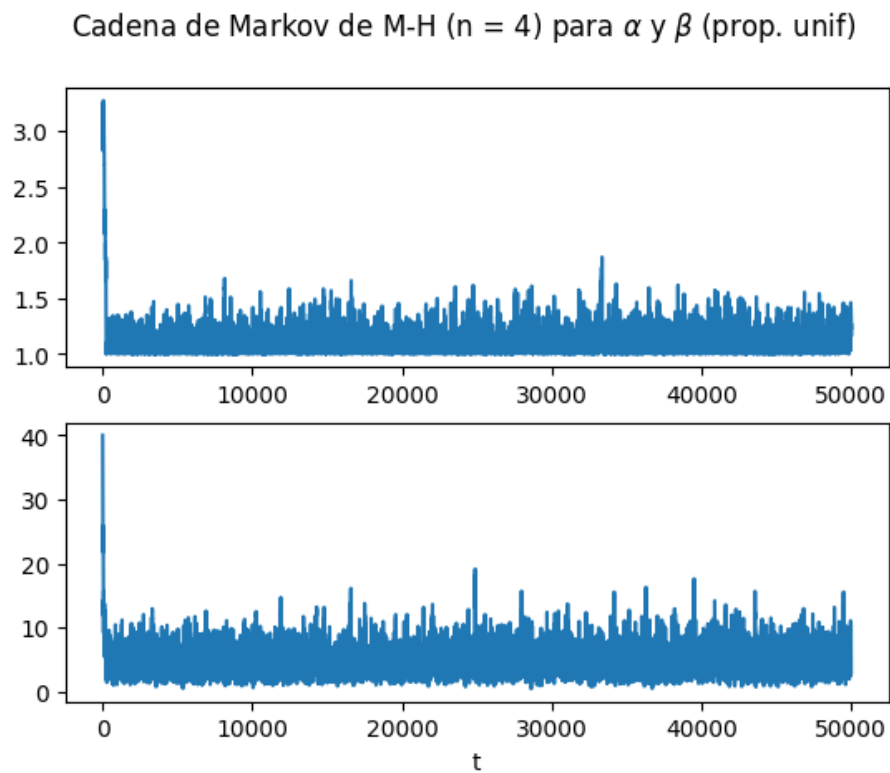


Figura 13: Cadena de Markov de Metropolis-Hastings para  $n = 4$



Cadena de Markov de M-H ( $n = 30$ ) para  $\alpha$  y  $\beta$  (prop. unif)

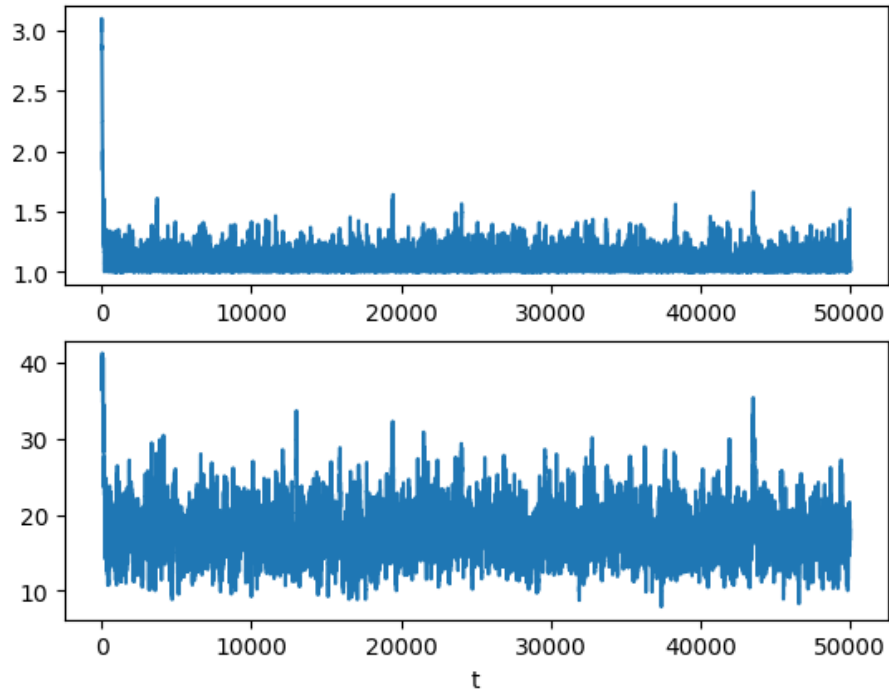


Figura 14: Cadena de Markov de Metropolis-Hastings para  $n = 30$

Nuevamente tomaremos un burn-in de 6000. Tras la eliminación inicial mencionada, mostramos la evolución de la cadena para los próximos 200 paso. Vemos que está tiene varios rechazos ahí donde la grafica es constante, por lo que se puede pensar que tarda más en converger que con la propuesta normal.

Histograma para las marginales de la posterior ( $n=4$ )  $\alpha$  y  $\beta$

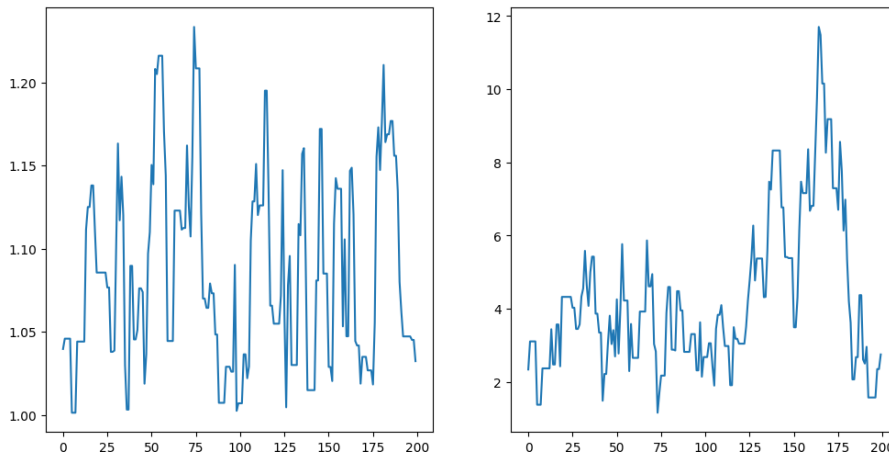


Figura 15: 200 pasos para la cadena de Markov de Metropolis-Hastings con burn-in ya realizado.

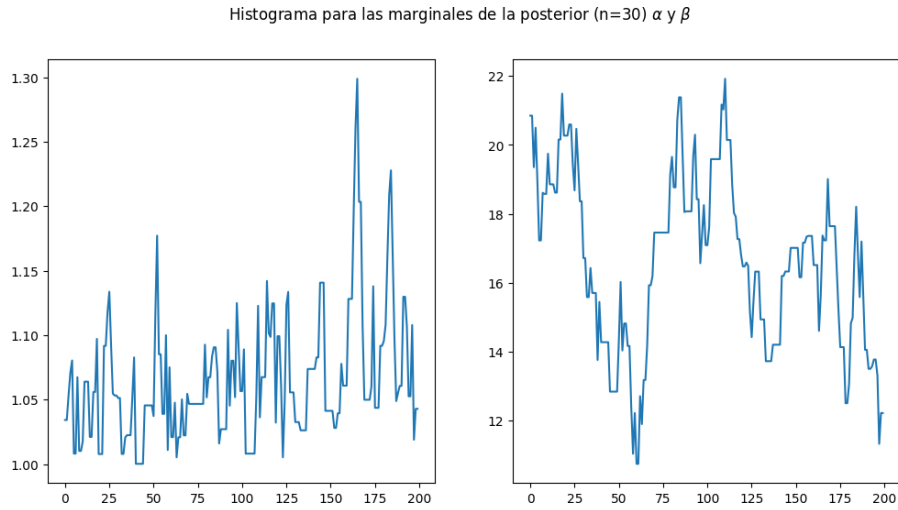


Figura 16: 200 pasos para la cadena de Markov de Metropolis-Hastings con burn-in ya realizado.

### Ejercicio 2

Simular de la distribución  $\text{Gamma}(\alpha, 1)$  con la propuesta  $\text{Gamma}([a], 1)$ , donde  $[a]$  denota la parte entera de  $a$ .

Además, realizar el siguiente experimento: poner como punto inicial  $x_0 = 900$  y graficar la evolución de la cadena, es decir,  $f(X_t)$  vs  $t$ .

### Solución:

Para este inciso implementamos Metropolis-Hastings para simular de la distribución  $\text{Gamma}(a, 1)$  donde  $a = 7.73$  donde consideraremos primeramente la distribución propuesta  $\text{Gamma}([a], 1)$  y  $[\cdot]$  es la función parte entera.

Observemos que tenemos una propuesta independiente. Es decir,  $q(y|x) = q(y)$ . Luego, la cadena de Markov cambiará de estado al propuesto con probabilidad  $\rho$  de la forma

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

$$= \min \left\{ 1, \frac{\frac{1}{\Gamma(a)} y^{a-1} e^{-y}}{\frac{1}{\Gamma(a)} x^{a-1} e^{-x}} \cdot \frac{\frac{1}{\Gamma([a])} x^{[a]-1} e^{-x}}{\frac{1}{\Gamma([a])} y^{[a]-1} e^{-y}} \right\} \quad (2)$$

$$= \min \left\{ 1, \left( \frac{y}{x} \right)^{a-[a]} \right\} \quad (3)$$

Así, simulamos de la  $\text{Gamma}(7, 1)$  y hacemos M-H con el punto inicial  $x_0 = 900$ . Notemos de (3) si  $x$  es muy grande, entonces el cociente es pequeño y por tanto la probabilidad de aceptar la propuesta  $y$  es muy baja. Luego, esperamos que en los primeros pasos tengamos que el estado de la cadena no cambia, es decir se mantiene en 900. Por tanto, es necesario hacer un burn-in para poder usar los datos como simulación de la distribución objetivo. Veamos la cadena en las 30000 iteraciones calculadas

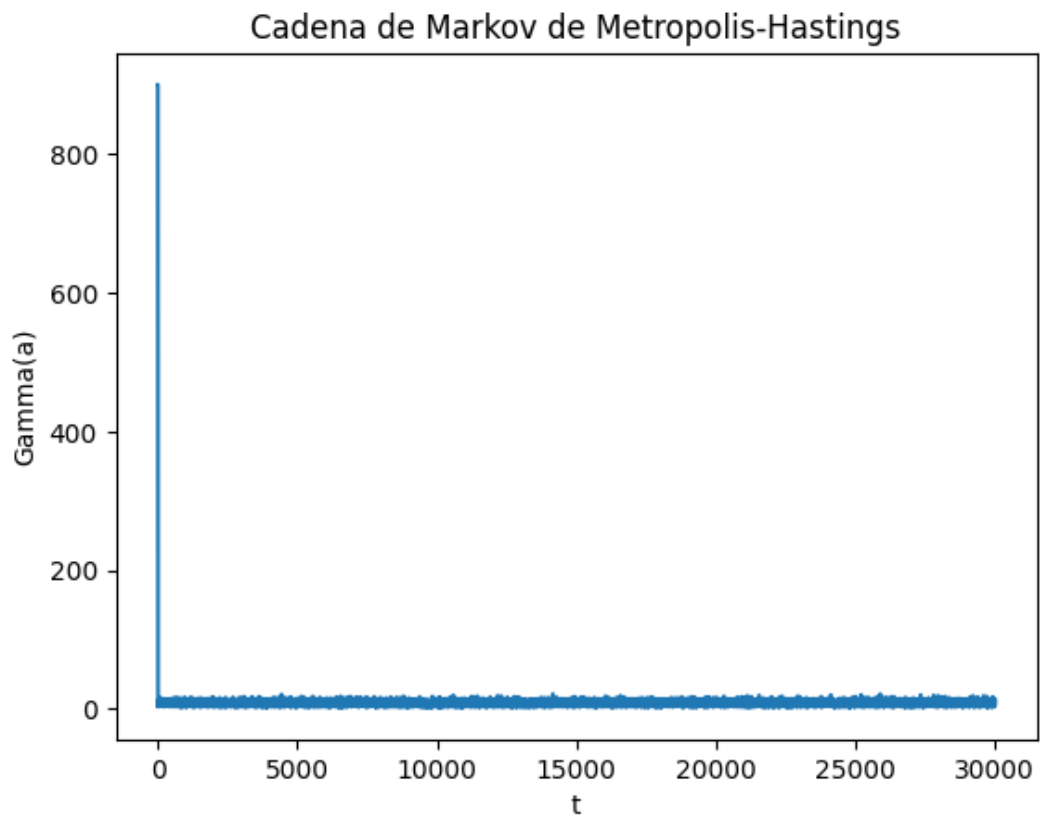


Figura 17: Cadena de Markov para M-H con distribución objetivo  $\text{Gamma}(a,1)$

Para tener más claro la cantidad de Burn-in grafiquemos la cadena para los primeros 50 pasos

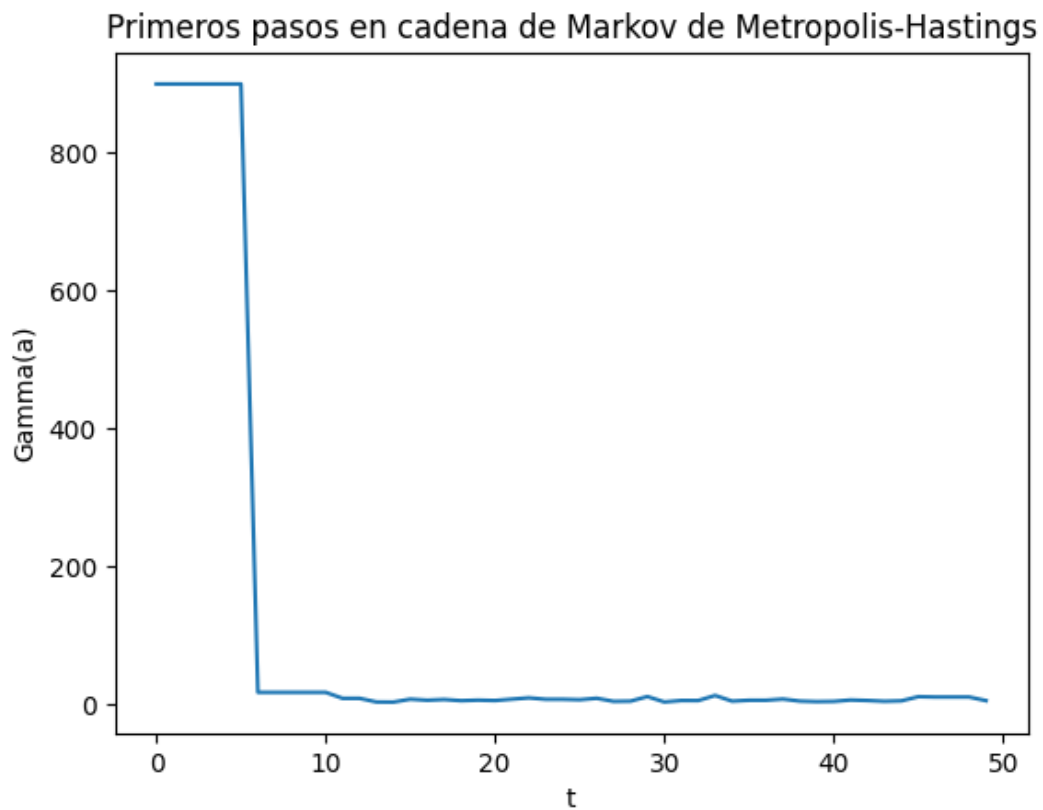


Figura 18: Cadena de Markov para M-H en los primeros 50 pasos.

Dependiendo la semilla o la realización, la cantidad de veces que se rechaza la propuesta manteniendo la posición inicial puede variar. En nuestro caso tenemos que dicho cambio está un poco antes del paso 10. Sin embargo, haciendo experimentos se pudo encontrar casos que superaban los 50 pasos. Por lo pronto y para este caso, tomamos un burn-in de 50.

De la cadena de Markov, ya aplicado el burn-in podemos obtener un histograma y comparar con la densidad objetivo

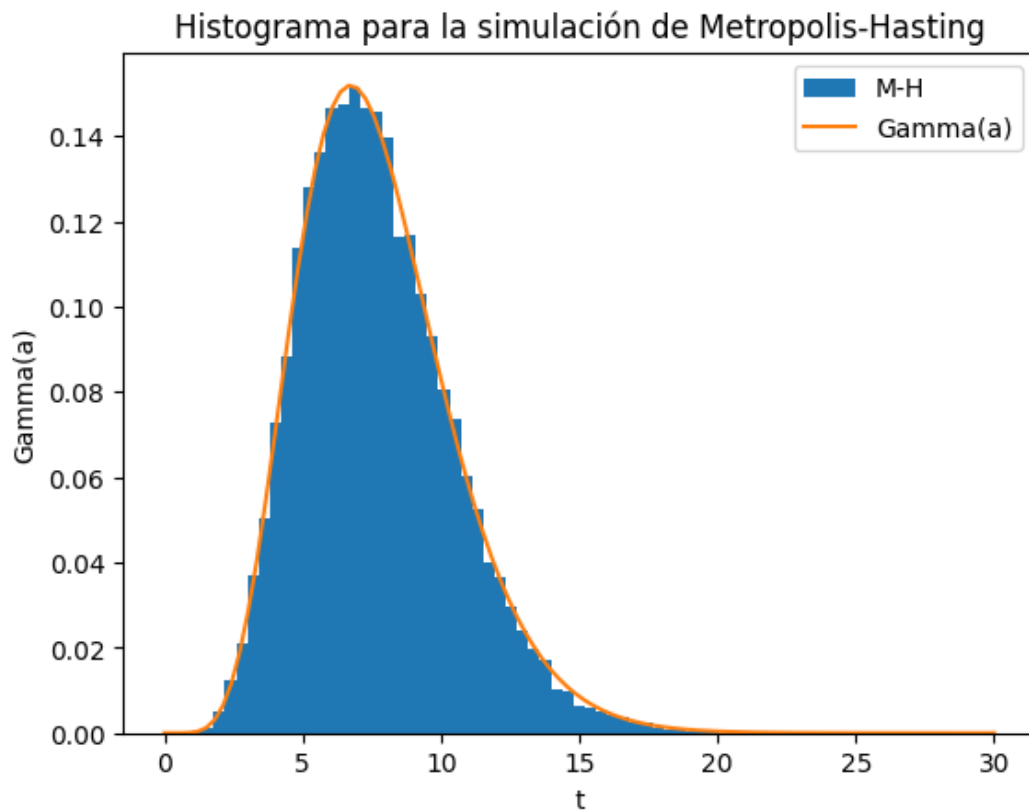


Figura 19: Histograma de la simulación obtenida por el método de Metropolis-Hastings

La cadena es bastante óptima, se aproxima bastante rápido a la distribución objetivo, lo cual tiene sentido ya que existe cierta continuidad en los parámetros de la distribución gamma y estos se encuentran cercanos, pues la distancia para el parámetro  $\alpha$  no supera uno.

Se propone ahora usar la distribución propuesta como la uniforme en  $[0,30]$  ya que el soporte para la gamma está concentrado en este intervalo. Luego, por el método M-H tenemos que eliminar el burn-in, que para este caso y por construcción se puede permitir ser más bajo ya que el punto inicial bajará con alta probabilidad. Como podemos ver en las gráficas para la cadena de Markov

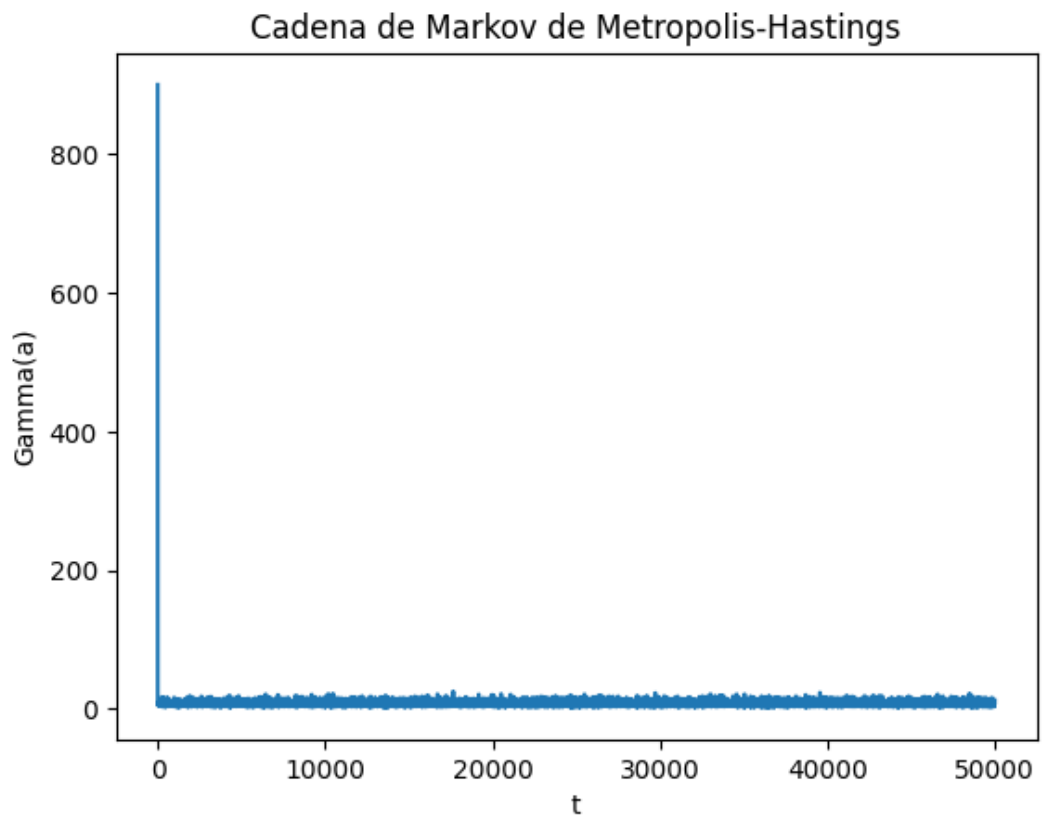


Figura 20: Cadena de Markov para la propuesta uniforme

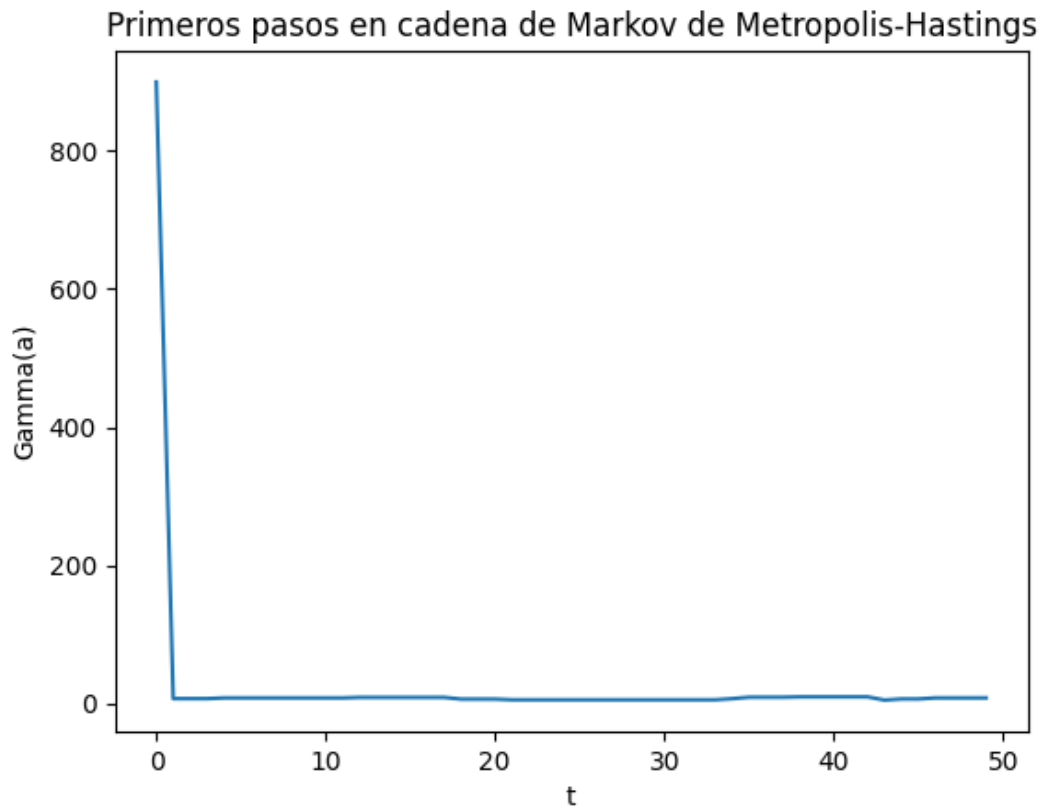


Figura 21: Primeros valores de la cadena de Markov para M-H con propuesta uniforme

Por último, eliminando el burn-in y tomando el histograma con la simulación es

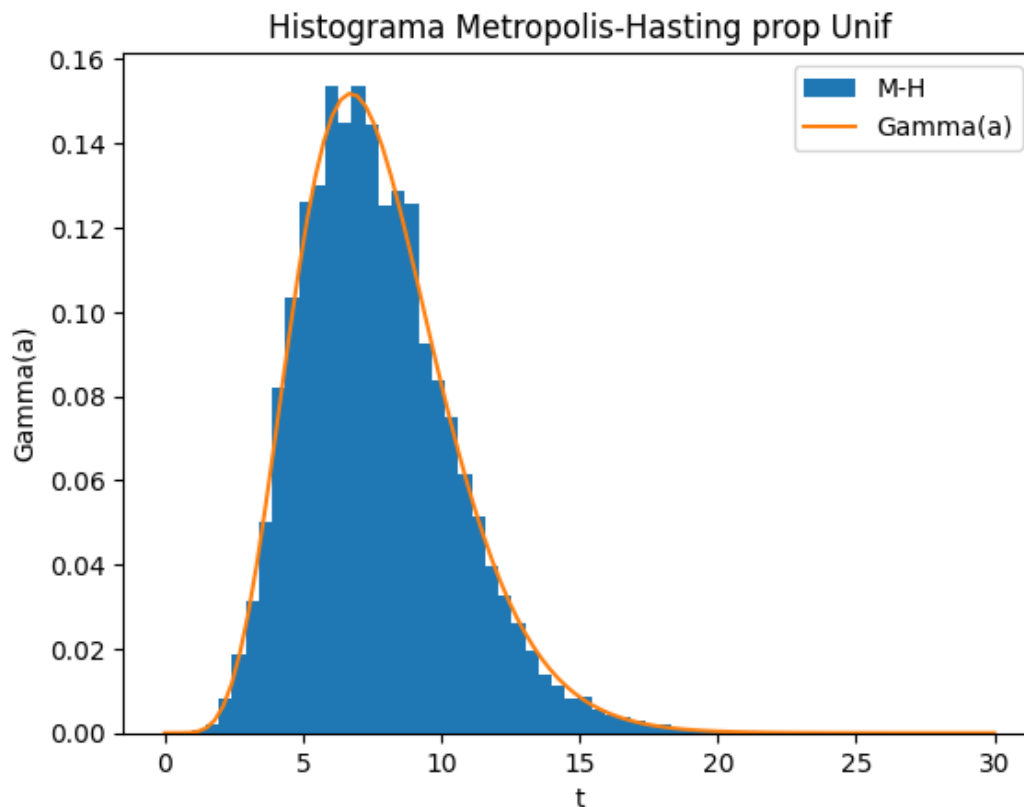


Figura 22: Histograma para simulación M-H con propuesta uniforme.

### Ejercicio 3

Implementar Random Walk Metropolis Hasting (RWMH) donde la distribución objetivo es  $\mathcal{N}_2(\mu, \Sigma)$ , con

$$\mu = \begin{pmatrix} 3 \\ 5 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 1 & 0,9 \\ 0,9 & 1 \end{pmatrix}.$$

Utilizar como propuesta  $\varepsilon_i \sim \mathcal{N}_2(0, \sigma^2 I)$  ¿Cómo elegir  $\sigma$  para que la cadena sea eficiente? ¿Qué consecuencias tiene la elección de  $\sigma$ ?

Como experimento, elige como punto inicial  $x_0 = \begin{pmatrix} 1000 \\ 1 \end{pmatrix}$  y comenta los resultados.

### Solución:

Análogo a los ejercicios previos, tenemos que implementar un Metropolis-Hastings con caminata aleatoria. Tenemos como distribución una normal bivariada como se describe en el enunciado. Haremos tres experimentos. Empezamos la caminata aleatoria desde puntos distantes como (40,1) (100,1) y (1000,1). Tomamos  $\sigma = 0.5$  ya que tomar uno mayor hace que los saltos sean muy bruscos por lo que no consigue converger a la distribución objetivo de forma óptima.

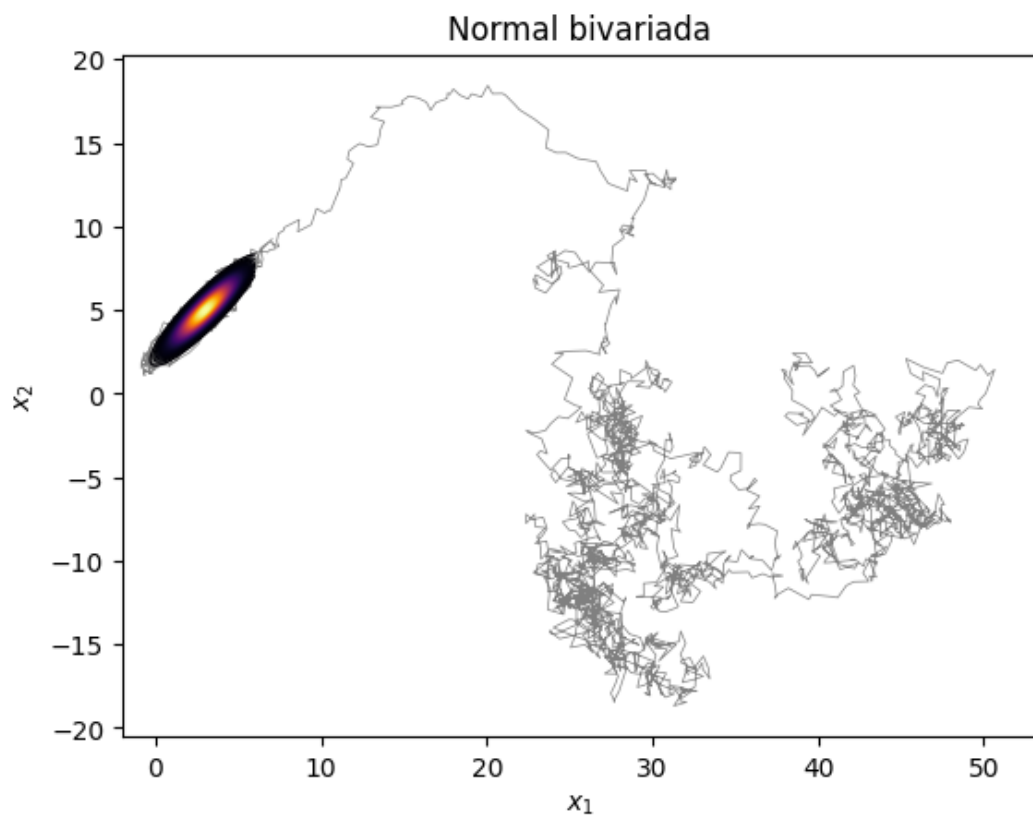


Figura 23: M-H con punto inicial (40,1)

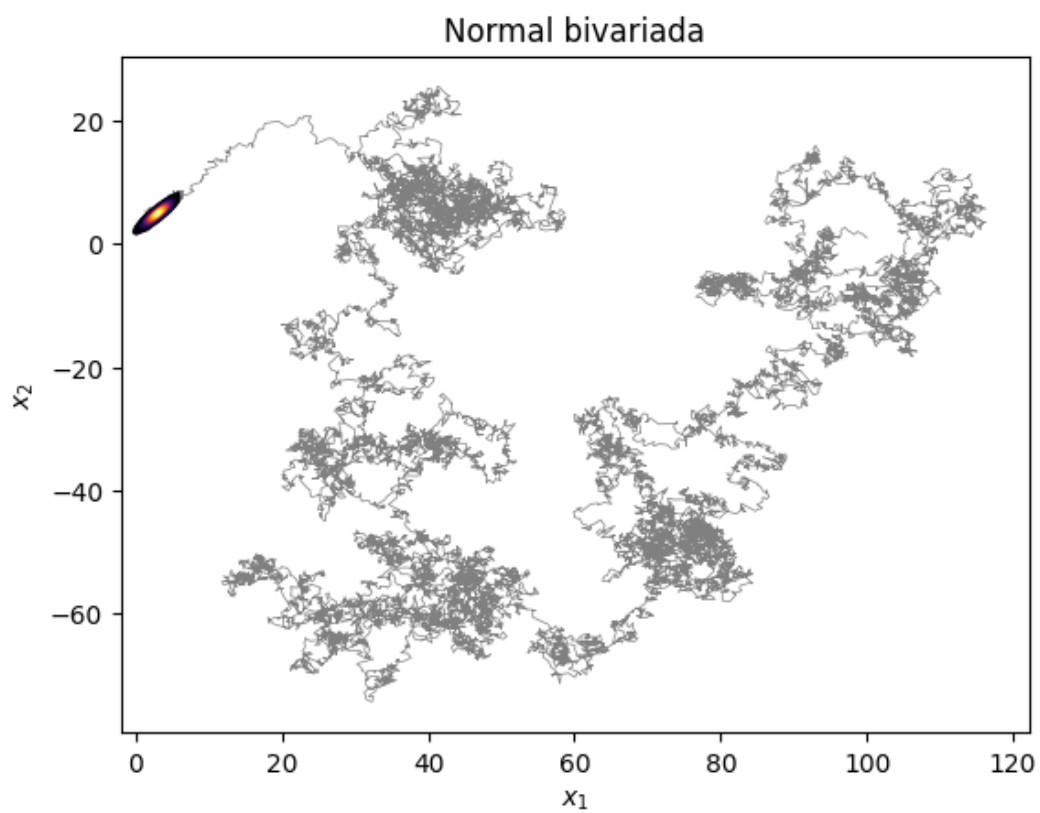


Figura 24: M-H con punto inicial (100,1)



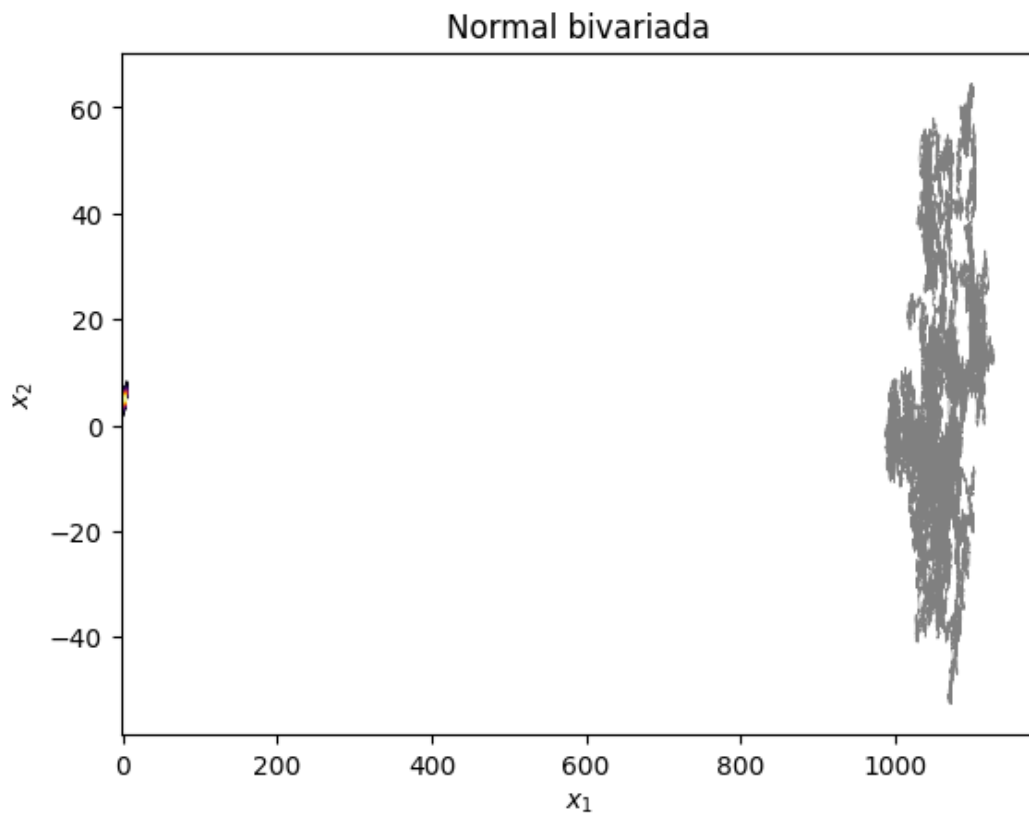


Figura 25: M-H con punto inicial (1000,1)

Veamos que cuando el punto inicial es cercano se puede converger a la distribución objetivo con 50000 iteraciones. Aún en el caso con punto inicial (100,1) tenemos que dicha caminata bordea la distribución hasta que finalmente converge a la objetivo. Sin embargo para el punto (1000,1) no bastan las 50000 iteraciones. Se puede modificar el valor de  $\sigma$  pero ponerlo más grande ocasiona que se vaya más lejos y tenga una convergencia aún más lenta.

Por otro lado, si cambiamos la propuesta como un cuadrado uniforme como en el ejercicio uno, tenemos un comportamiento similar al de la propuesta normal. Con el punto inicial (100,1) tenemos una convergencia tardía como mostramos en la siguiente grafica

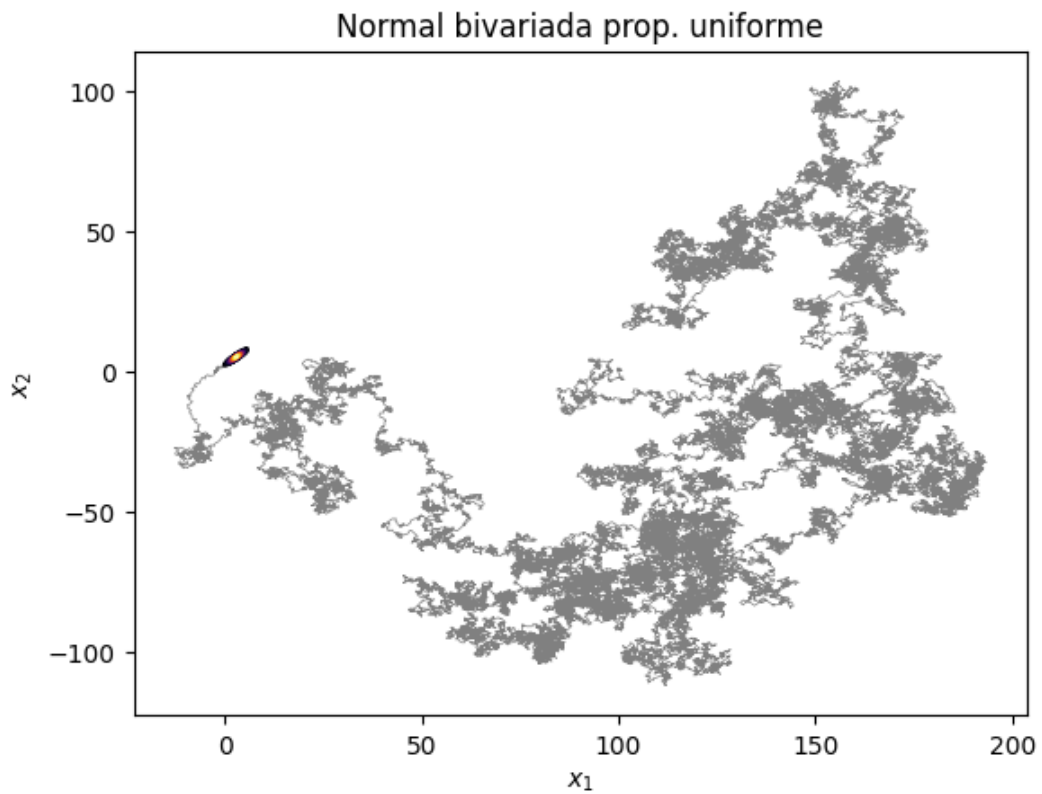


Figura 26: Metropolis-Hastings para la normal bivariada con propuesta uniforme y punto inicial (100,1)

Nuevamente, aumentar el valor del tamaño de la uniforme ( $\epsilon$ ) produce pasos grandes y por tanto la convergencia no es precisa, pues oscila de su distribución objetivo. Bajar el valor de  $\epsilon$  implica que tiene pasos pequeños y no llega con la cantidad de iteraciones propuestas.

## Referencias

- [1] Robert, C. P., Casella, G., and Casella, G. (1999). Monte Carlo statistical methods (Vol. 2). New York: Springer.
- [2] Wasserman, L. (2004). All of statistics: a concise course in statistical inference (p. 413). New York: Springer.