

Cómputo Científico

Tarea 8

MCMC: MH con Kérneles Híbridos y Gibbs Sampler

Iván Irving Rosas Domínguez

4 de diciembre de 2023

1. Aplique el algoritmo de Metrópolis-Hastings considerando como función objetivo la distribución normal bivariada:

$$f_{X_1, X_2}(\bar{x}) = \frac{1}{2\pi} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\bar{x} - \mu)^t \Sigma^{-1} (\bar{x} - \mu) \right\}$$

donde,

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Así, se tienen las siguientes distribuciones condicionales:

$$X_1|X_2 = x_2 \sim N \left(\mu_1 + \rho \frac{\sigma_1}{\sigma_2} (x_2 - \mu_2), \sigma_1^2 (1 - \rho^2) \right)$$

$$X_2|X_1 = x_1 \sim N \left(\mu_2 + \rho \frac{\sigma_2}{\sigma_1} (x_1 - \mu_1), \sigma_2^2 (1 - \rho^2) \right)$$

Considere las siguientes propuestas:

$$q_1((x'_1, x'_2)|(x_1, x_2)) = f_{X_1|X_2}(x'_1|x_2) \mathbb{1}_{\{x'_2=x_2\}}$$

$$q_2((x'_1, x'_2)|(x_1, x_2)) = f_{X_2|X_1}(x'_2|x_1) \mathbb{1}_{\{x'_1=x_1\}}$$

A partir del algoritmo MH usando Kerneles híbridos simule valores de la distribución normal bivariada, fijando $\sigma_1 = \sigma_2 = 1$, considere los casos $\rho = 0,8$ y $\rho = 0,95$.

Solución: Todos los contenidos de esta tarea están en el script *ComputoTarea8.py*. El mismo se encuentra comentado en las partes correspondientes a la implementación de los algoritmos.

Se comienza con el ejercicio 1 utilizando una semilla *seed* = 10 para replicar resultados.

Se crean dos funciones para este ejercicio: *dist_obj(x, rho)*, la cual calcula el valor de la distribución normal bivariada que se pide, ingresando el punto *x*, y con coeficiente de correlación ρ .

Asimismo, se crea la función *MH_gibbs1*, la cual se encarga de correr el algoritmo de Metrópolis-Hastings. Los parámetros que maneja son las *k* iteraciones, el coeficiente de correlación ρ de la normal bivariada, y *m* un vector inicial con el cual comenzar la cadena.

Una vez definidas las funciones, se coloca el punto inicial como $m = (5, 5)$, y se pone en marcha el algoritmo. Se tienen dos casos: $\rho = 0,8$ y $\rho = 0,95$.

$\rho = 0,8$: en este caso, se ejecutó el algoritmo con $k = 250,000$ iteraciones. Se programa una rejilla en la región $[-5, 5] \times [-5, 5]$ del plano para graficar los resultados en $2D$, y así mismo, se calcula el valor de la distribución objetivo sobre esta rejilla para obtener las curvas de nivel y contrastar con los resultados.

Se extrae también el logaritmo de la función de densidad objetivo para obtener un buen número para realizar el burn-in. Se obtiene el resultado de la figura 1. Observamos que la gráfica de la log-densidad objetivo rápidamente llega a una

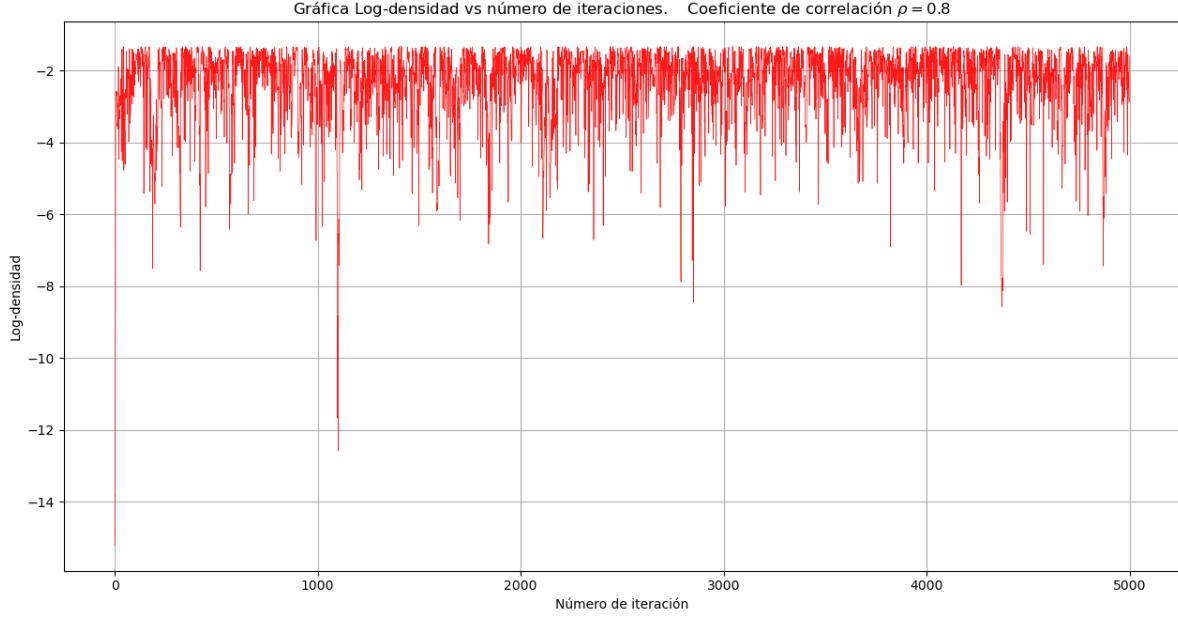


Figura 1: Log-densidad vs número de iteraciones. Coeficiente $\rho = 0,8$

sección que podemos intuir es estable. Esto nos hace tomar la decisión de tomar un Burn-in igual a 250 iteraciones. Graficamos el recorrido de la cadena, el cual se aprecia en la figura 2. En esta figura observamos el comportamiento de los kernels híbridos que actúan en solo una dirección: a cada iteración, de aceptar la propuesta, se avanza o bien en la dirección horizontal o bien en la vertical. Observamos una rápida convergencia a la distribución objetivo, que en este caso es una normal bivariada, con matriz de covarianzas la dada al inicio de este ejercicio. Graficamos ahora los histogramas de la variable x y la variable y . De acuerdo a la matriz de covarianzas, las variables marginales deben ser normales estándar. Las líneas dibujadas en ambos histogramas corresponden a las densidades de variables aleatorias normales estándar. Esto comprueba lo mencionado anteriormente.

Procedemos ahora con el ejemplo que utiliza $\rho = 0,95$. Análogamente a lo hecho antes, se obtienen los siguientes resultados: Tal y como en el primer caso, rápidamente la densidad llega a una zona de estabilidad. De la visualización de la figura 5, deducimos que un Burn-in de tamaño 250 es adecuado. Graficamos el recorrido de la cadena junto con las curvas de nivel de la densidad objetivo. Esto se presenta en la figura 6.

Presentamos también los histogramas de las variables x y y . A pesar de que el coeficiente de correlación ha cambiado (cosa que se refleja en la forma de la densidad conjunta en la figura 6) las marginales siguen siendo variables normales estándar. Esto se puede comprobar con las figuras 7 y 8.

En estos histogramas nuevamente se dibuja la función de densidad de las variables normales estándar. Tal y como se puede apreciar, estas encajan excelente con los histogramas, tal y como era de esperarse.

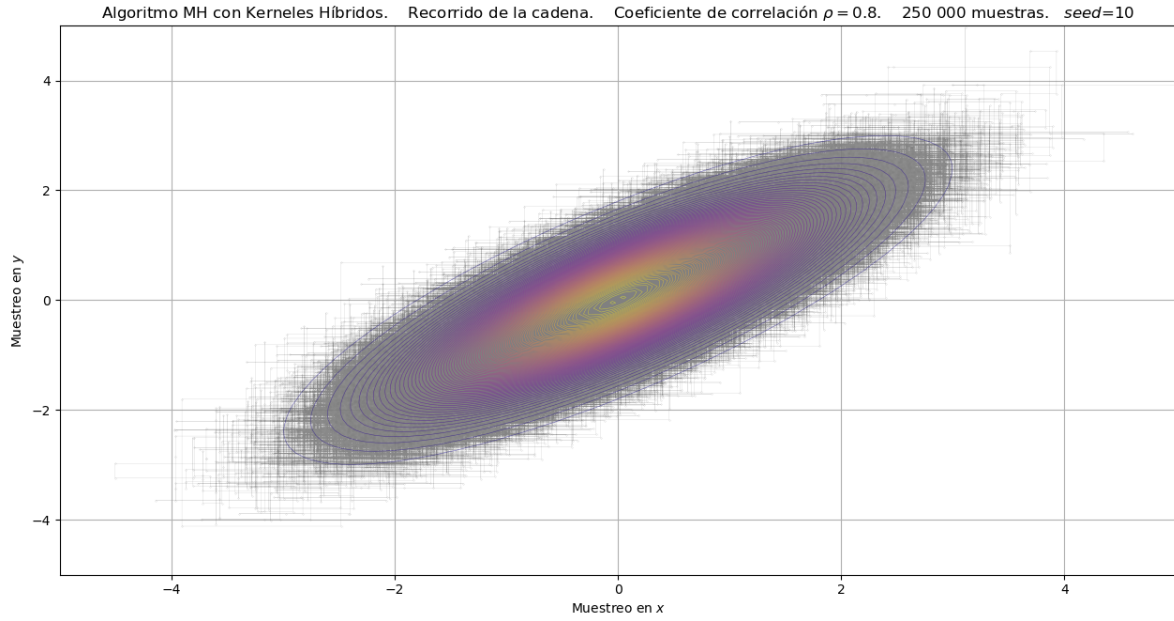


Figura 2: Recorrido de la cadena. Punto inicial $m = (5,5)$. Se puede apreciar una buena convergencia a la distribución objetivo.

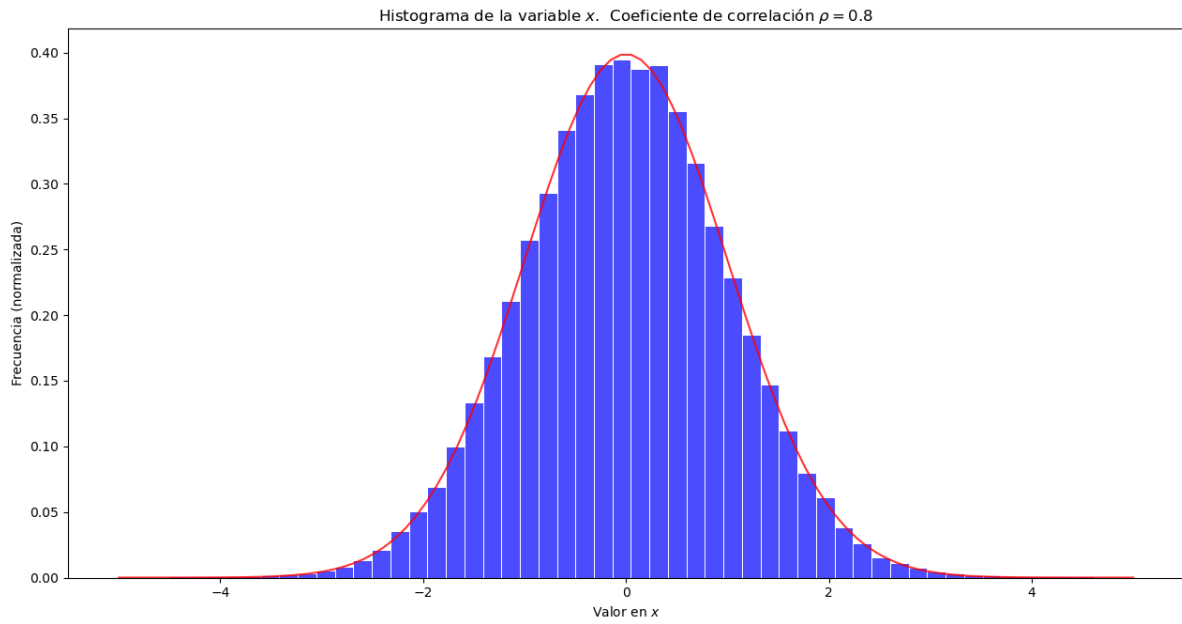


Figura 3: Histograma en x

2. Considere los tiempos de falla t_1, \dots, t_n con distribución $Weibull(\alpha, \lambda)$:

$$f(t_i|\alpha, \lambda) = \alpha \lambda t_i^{\alpha-1} e^{-t_i^\alpha \lambda}.$$

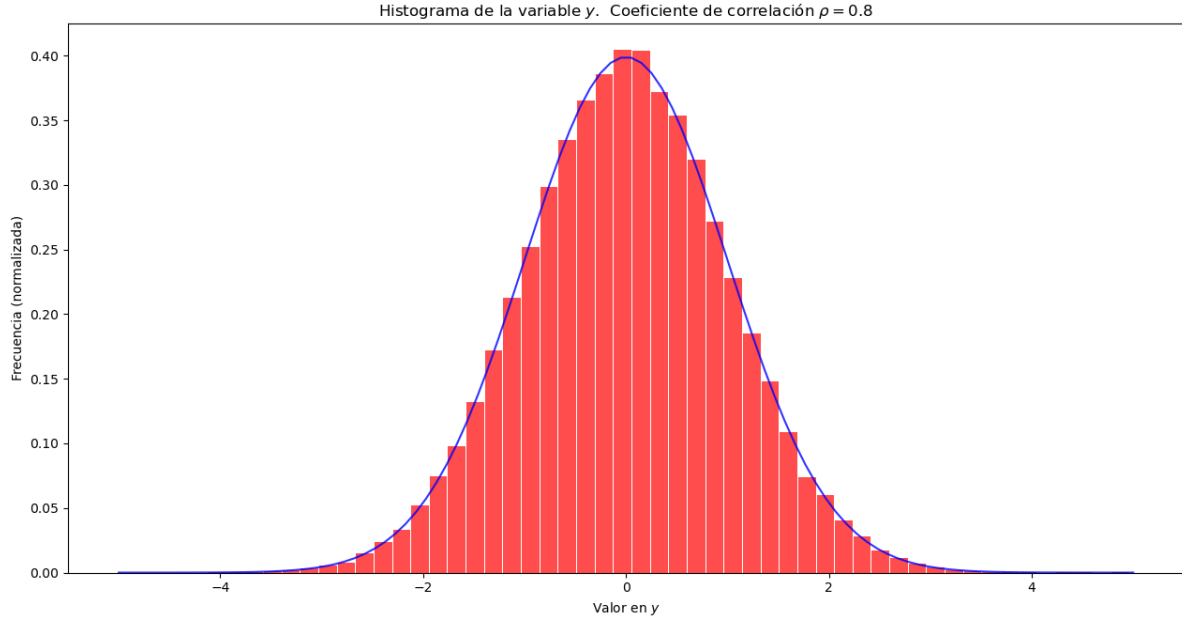


Figura 4: Histograma en y

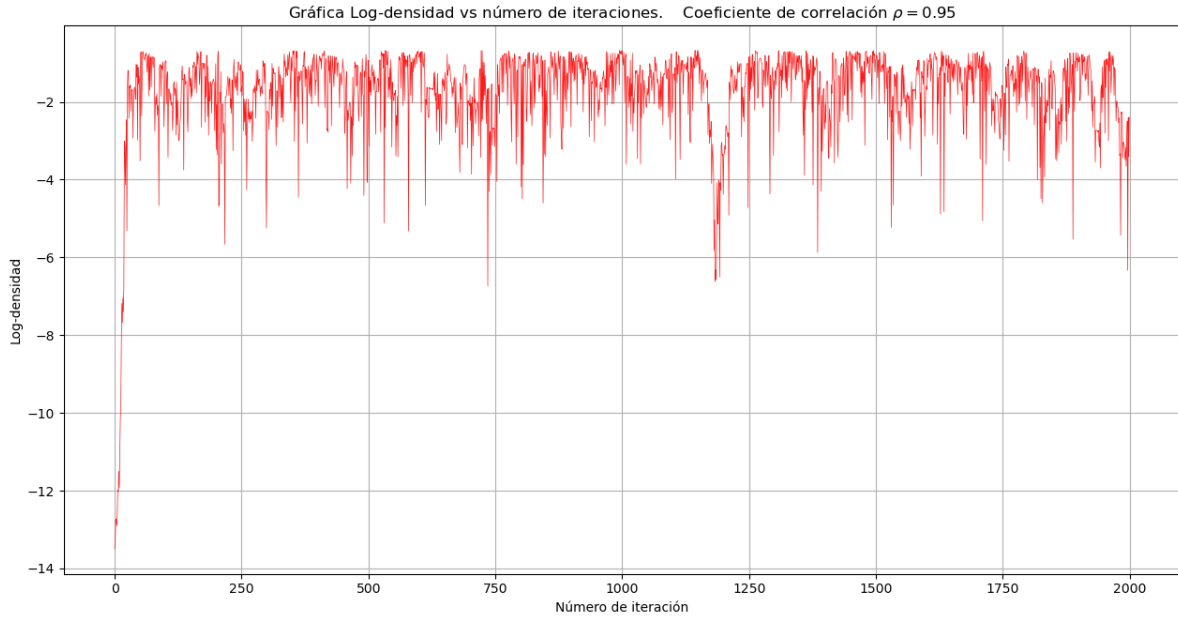


Figura 5: Log-densidad vs número de iteraciones. Caso $\rho = 0,95$

Se asumen como a priori $\alpha \sim \exp(c)$ y $\lambda|\alpha \sim \Gamma(\alpha, b)$, por lo tanto, $f(\alpha, \lambda) = f(\lambda|\alpha)f(\alpha)$. Así, para la distribución posterior se tiene:

$$f(\alpha, \lambda|\bar{t}) \propto f(\bar{t}|\alpha, \lambda)f(\alpha, \lambda).$$

A partir del algoritmo MH usando Kerneles híbridos simule valores de la distribución posterior $f(\alpha, \lambda|\bar{t})$, considerando

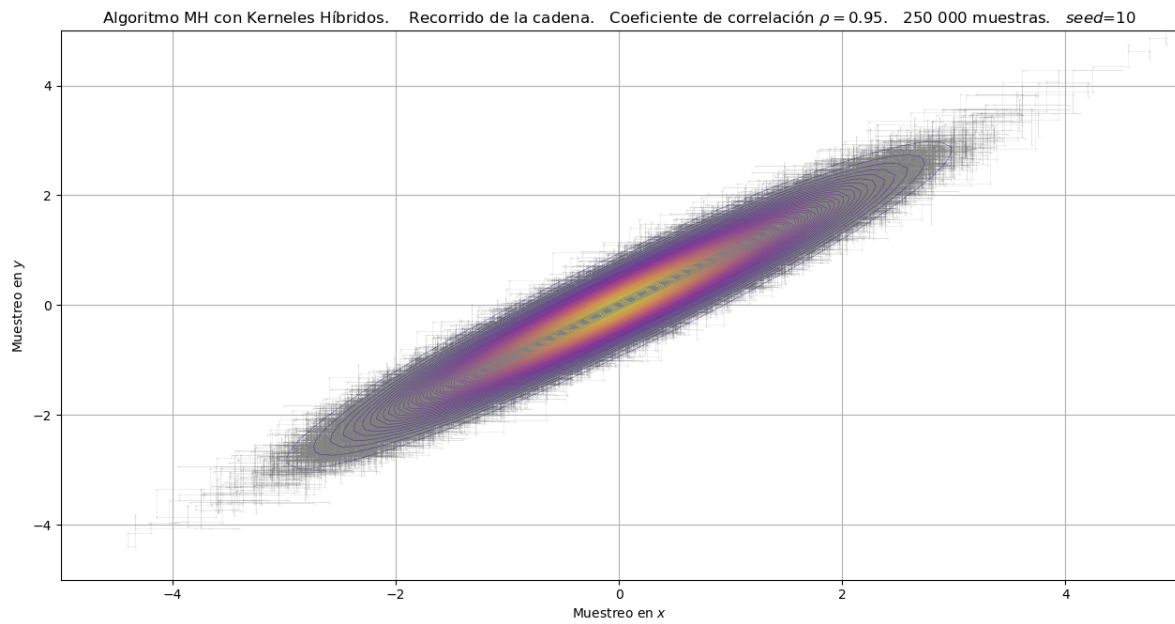


Figura 6: Recorrido de la cadena. Punto inicial $m = (5, 5)$. De nuevo se visualiza una convergencia adecuada

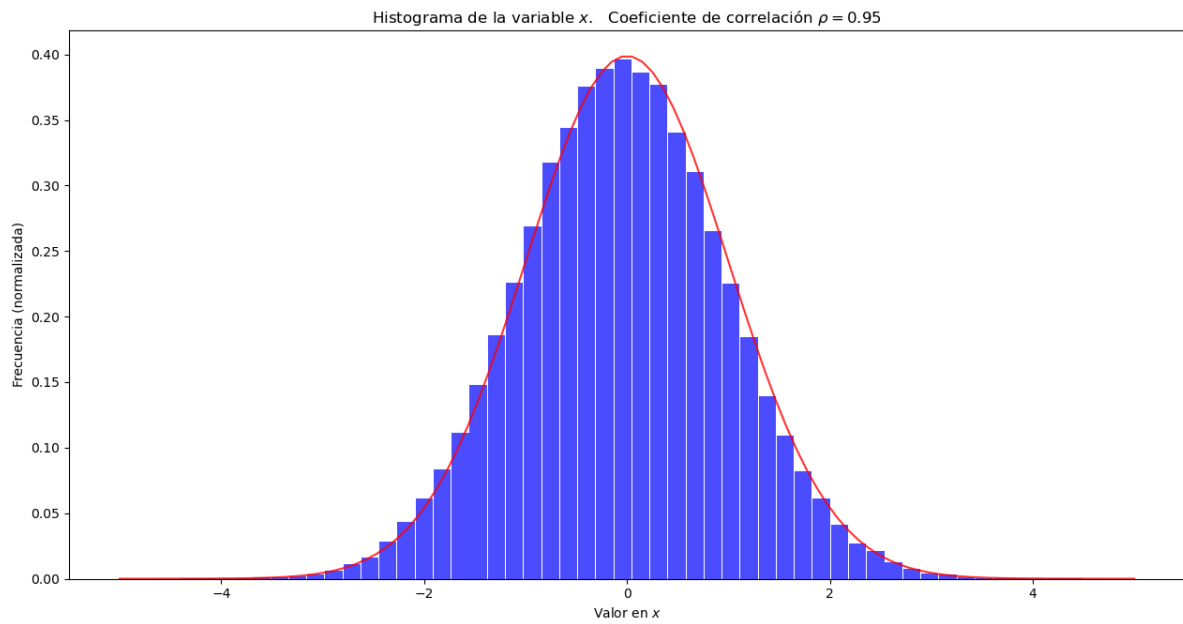


Figura 7: Histograma en x

las siguientes propuestas:

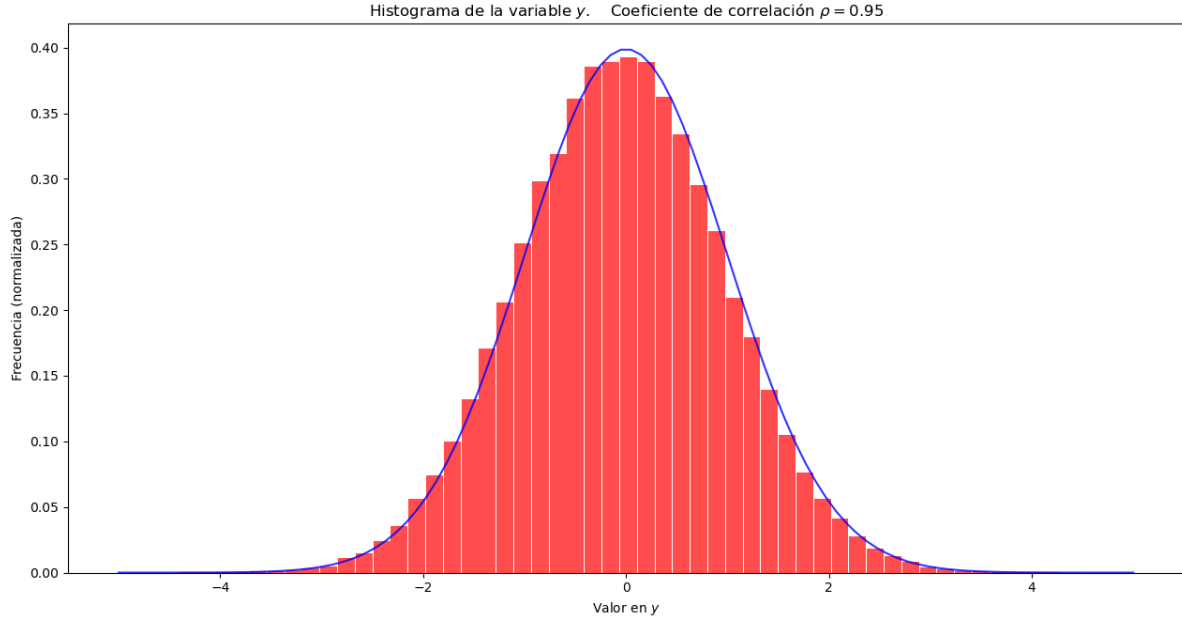


Figura 8: Histograma en y

Propuesta 1:

$$\lambda_p | \alpha, \bar{t} \sim \text{Gamma} \left(\alpha + n, b + \sum_{i=1}^n t_i^\alpha \right)$$

y dejando α fijo.

Propuesta 2:

$$\alpha_p | \lambda, \bar{t} \sim \text{Gamma}(n + 1, -\log(b) - \log(r_1) + c),$$

con $r_1 = \prod_{i=1}^n t_i$. y dejando λ fijo.

Propuesta 3:

$$\alpha_p \sim \exp(c) \text{ y } \lambda_p | \alpha_p \sim \text{Gamma}(\alpha_p, b).$$

Propuesta 4 (RWMH):

$$\alpha_p = \alpha + \varepsilon,$$

con $\varepsilon \sim N(0, \sigma)$ y dejando λ fijo.

Simular datos usando $\alpha = 1$ y $\lambda = 1$ con $n = 20$. Para la a priori usar $c = 1$ y $b = 1$.

Solución: Implementamos el algoritmo de Metrópolis Hastings utilizando Kernel Híbrido y Gibbs sampler. Observamos que en este caso tenemos 4 propuestas para realizar la implementación. En clase se argumentó que la primera de ellas, la cual consiste en muestrear el valor de $\lambda_p | \alpha$ es un Gibbs Sampler, por lo que la probabilidad de aceptar dicha propuesta siempre es 1. No obstante, las demás propuestas no necesariamente cumplen con esto, ya que por ejemplo la propuesta 4 consiste en un Random Walk Metrópolis Hastings.

Para implementar el algoritmo se coloca primero la semilla $seed = 20$ para replicar resultados. Posteriormente se crean las funciones que recrean la distribución objetivo y cada una de las cuatro propuestas que se mencionan en este problema. En particular, la función de la propuesta número 4 es tal que utiliza una variable aleatoria $Normal(0, 0.25)$ para

implementar el RWMH de dicha propuesta.

Se crea también la función que ejecutará el algoritmo de Metrópolis. Se otorga un peso igual a $1/4$ a cada uno de las cuatro propuestas, utilizando para ello una variable uniforme en $[0, 1]$ y ejecutando una de las cuatro propuestas dependiendo de si la variable uniforme cayó en alguno de los 4 subintervalos del $[0, 1]$ correspondiente a cada variable. A saber, la propuesta 1 correspondía a un valor de una uniforme en $[0, 0.25]$, la propuesta 2 correspondía a un valor de la uniforme en $[0.25, 0.5]$, etc.

Se simulan los datos de la distribución $Weibull(1, 1)$ utilizando *scipy*, y se selecciona como punto inicial al punto $m = (2, 2)$. Se fija un número de iteraciones $k = 250,000$. Se obtienen los siguientes resultados gráficos: comenzando con la log-densidad objetivo para obtener el Burn-in, se obtiene el contenido de la figura 9.

Nuevamente del comportamiento de esta gráfica concluimos que un valor de Burn-in igual a 250 es adecuado. Se

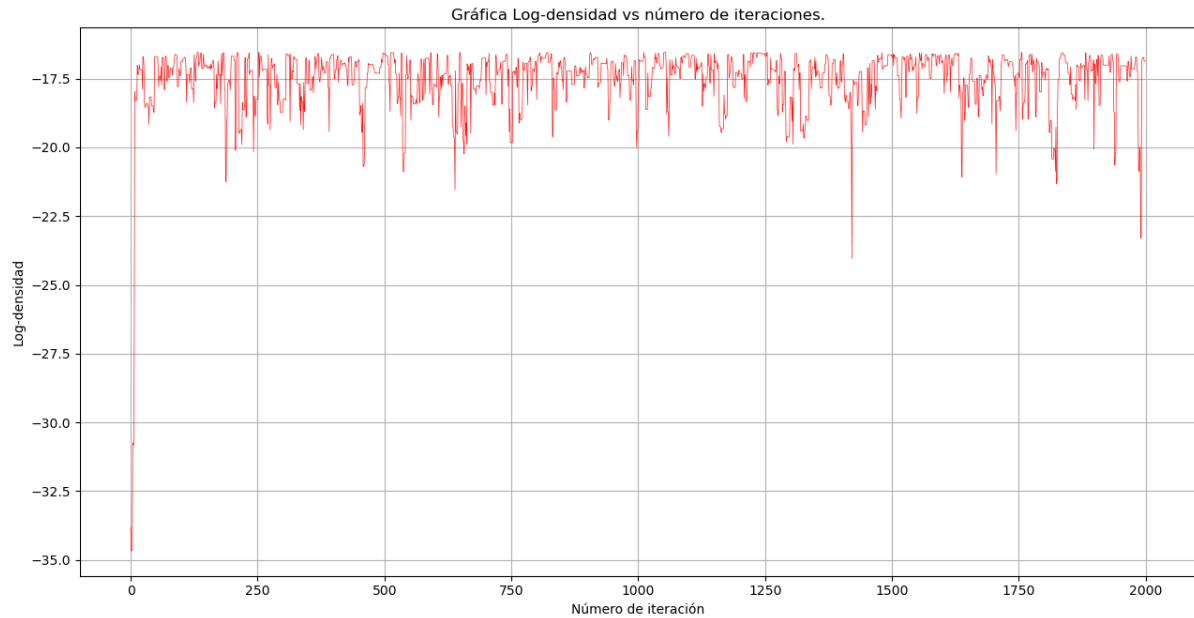


Figura 9: Log-densidad vs número de iteraciones en el ejercicio 2.

imprime ahora la gráfica del recorrido de la cadena, obteniendo el resultado de la figura 10.

De la figura se puede observar un buen comportamiento de la cadena en cuanto a su aproximación a la distribución objetivo. Para visualizar el comportamiento más de cerca de las variables λ y α , imprimimos el histograma de cada una de estas variables. Esto se presenta en las figuras 11 y 12. Observamos un comportamiento en la variable α que tiene una media de aproximadamente 1,7, mientras que en la variable λ se observa una media de aproximadamente 0,7.

3. Considere el ejemplo referente al número de fallas de bombas de agua en una central nuclear, donde p_i representa el número de fallas en el tiempo de operación t_i , con $i = 1, \dots, 10$.

Se considera el modelo $p_i \sim \text{Poisson}(\lambda_i t_i)$, (las λ_i son independientes entre sí), con distribuciones a priori $\lambda_i | \beta \sim \text{Gamma}(\alpha, \beta)$ y $\beta \sim \text{Gamma}(\gamma, \delta)$, por lo tanto:

$$f(\lambda_1, \dots, \lambda_n, \beta) = f(\lambda_1 | \beta) f(\lambda_2 | \beta) \cdot \dots \cdot f(\lambda_n | \beta) f(\beta).$$

Para la distribución posterior se tiene:

$$f(\lambda_1, \dots, \lambda_n, \beta | \bar{p}) \propto L(\bar{p}, \bar{\lambda}, \beta) f(\lambda_1, \dots, \lambda_n, \beta).$$

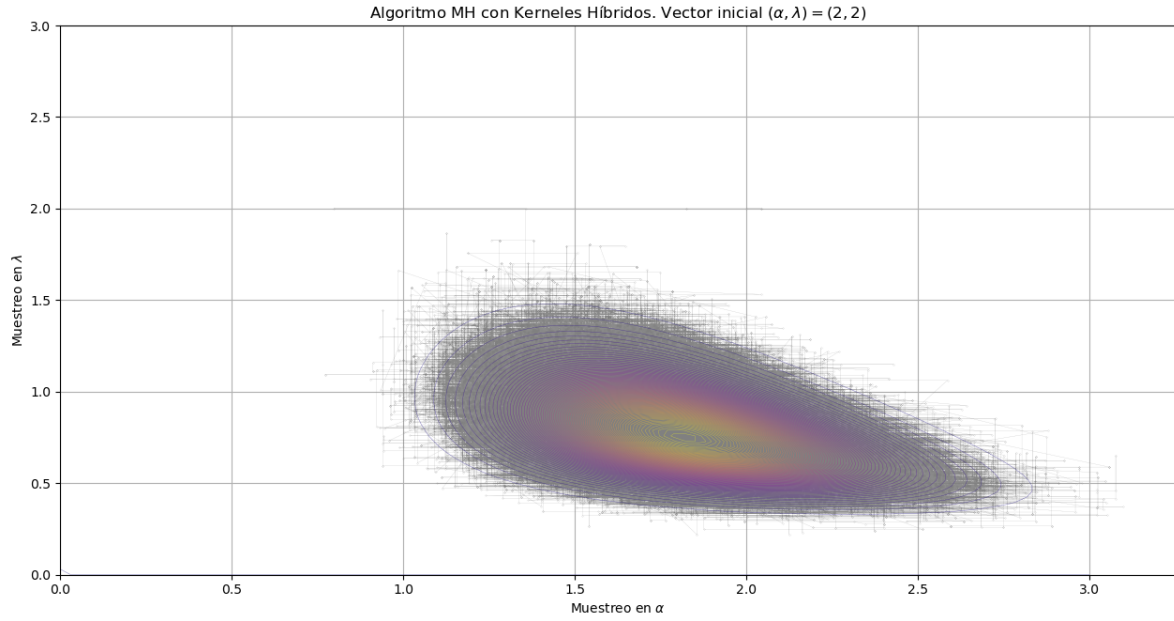


Figura 10: Recorrido de la cadena y curvas de nivel de la distribución objetivo. Observamos un buen desempeño.

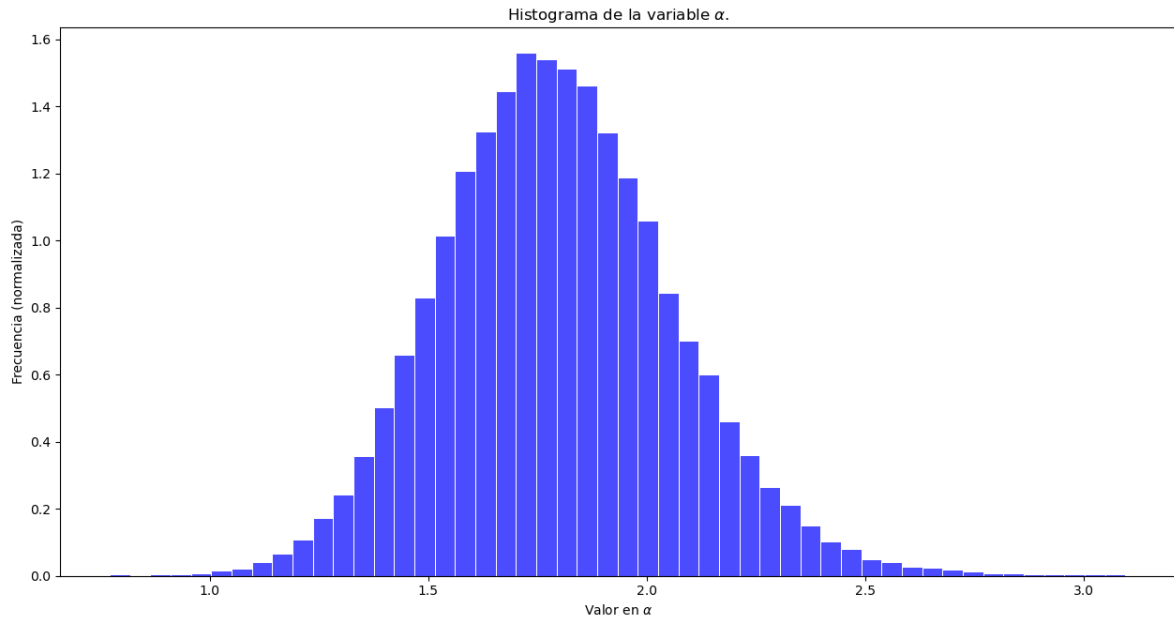


Figura 11: Histograma de la variable α

Simule valores de la distribución posterior $f(\lambda_1, \dots, \lambda_n, \beta | \bar{p})$, usando un kernel híbrido, considerando las propuestas:

$$\lambda_i | \bar{\lambda}_{-i}, \beta, \bar{t} \sim \text{Gamma}(p_i + \alpha, \beta + t_i).$$

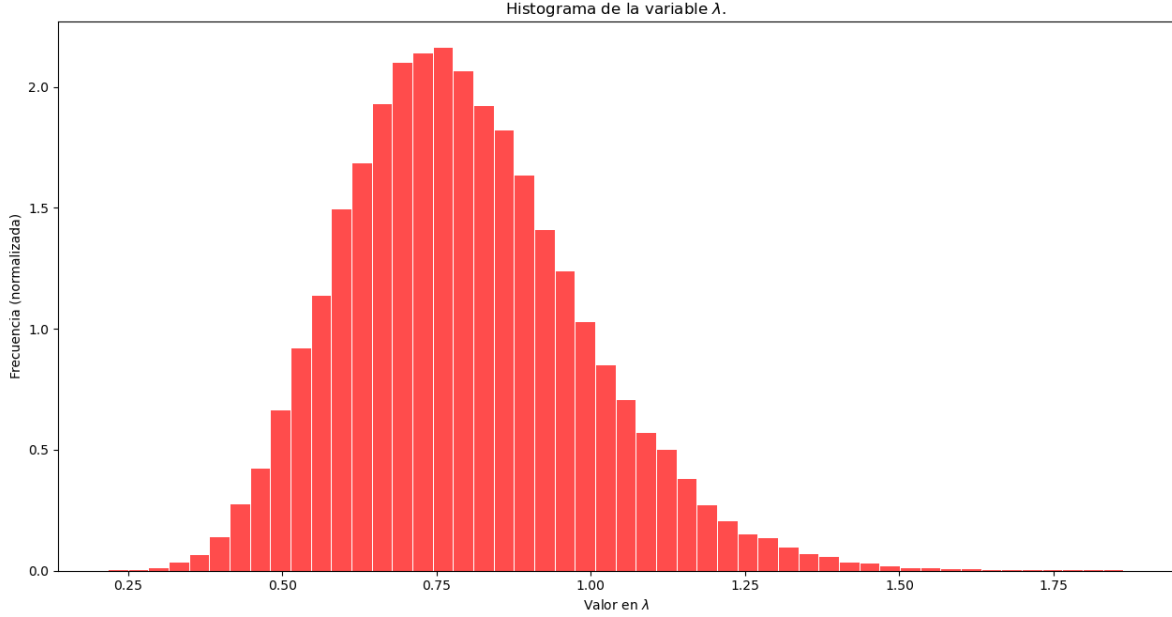


Figura 12: Histograma de la variable λ

| Bomba (i) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----------------------|-------|-------|-------|--------|------|-------|------|------|-----|-------|
| T. de uso t_i | 94.32 | 15.72 | 62.88 | 125.76 | 5.24 | 31.44 | 1.05 | 1.05 | 2.1 | 10.48 |
| # de fallas (p_i) | 5 | 1 | 5 | 14 | 3 | 17 | 1 | 1 | 4 | 22 |

Cuadro 1: Datos de bombas de agua en centrales nucleares (Robert y Casella, p.385) para el ejemplo 8.3.

$$\beta | \bar{\lambda}, \bar{t} \sim \text{Gamma} \left(n\alpha + \gamma, \delta + \sum_{i=1}^n \lambda_i \right).$$

Verifique que estas son propuestas Gibbs.

Use los datos de la tabla 1 con los parámetros a priori $\alpha = 1.8$, $\gamma = 0.01$ y $\delta = 1$.

Solución: para este último ejercicio, primero se fija la semilla en el valor $seed = 10$ para replicar resultados. Posteriormente, se definen las funciones a utilizar. En este caso, y dada la particularidad de las propuestas a usar, para implementar el algoritmo no es necesario hallar explícitamente la distribución objetivo, ya que las propuestas se componen de Kernels Gibbs, por lo que el cociente de Metrópolis-Hastings será siempre 1 y con ello las propuestas hechas siempre serán aceptadas. El que los Kernels presentados tengan esta propiedad se muestra al final de los resultados de este ejercicio.

Se define el vector inicial $m = (5, 5, \dots, 5) = (\lambda_1, \lambda_2, \dots, \lambda_{10}, \beta)$, y se definen dos funciones para ejecutar las propuestas. La primera de ellas consiste en realidad en una anidación de las propuestas para los 10 parámetros $\lambda_1, \dots, \lambda_{10}$.

Dicha función denominada *propuesta1* consiste en simular variables *gamma* dependiendo del parámetro i que sea ingresado, y que corresponde al valor de λ_i en cuestión.

La segunda función *propuesta2* consiste en simular valores para el parámetro β a partir de los datos que se tienen sobre $\lambda_1, \dots, \lambda_{10}$ y los datos.

Se crea la función *MH_gibbs3* que implementará el algoritmo. Dicha función crea una variable uniforme. Se divide el intervalo $[0,1]$ en 11 pedazos, y si la variable uniforme cae en el último de los 11 pedazos, se procede a actualizar el parámetro β de acuerdo a la propuesta 2, mientras que si cae en los primeros 11 pedazos, se utiliza nuevamente una variable aleatoria uniforme, pero esta vez discreta en el conjunto $\{1, \dots, 10\}$ para seleccionar el parámetro λ_i que será actualizado. Lo anterior asegura que los kérneles que se seleccionan tienen todos ellos igual peso entre sí.

Se cargan los datos del cuadro 1, y fijando $k = 250,000$ iteraciones, se procede a la ejecución del algoritmo. De manera completamente heurística, y a falta del cálculo de la distribución objetivo de manera explícita, se utiliza un Burn-in de tamaño 250 siendo congruente con lo visto en los otros ejercicios de esta tarea y de tareas pasadas.

Con tal Burn-in, se obtienen los histogramas de cada uno de los 11 parámetros $\lambda_1, \dots, \lambda_{10}, \beta$, los cuales son presentados en las figuras 13 en adelante.

Ciertamente la cantidad de dimensiones del vector de parámetros vuelve más difícil la visualización de recorrido de

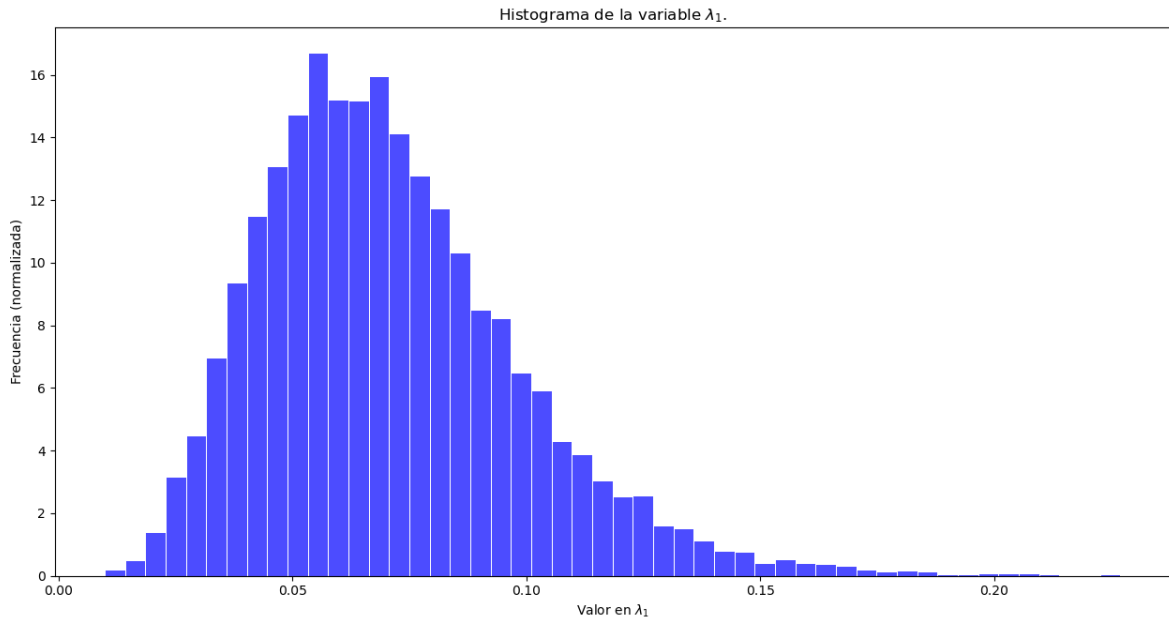


Figura 13: Histograma para λ_1

la cadena, así como el estudio de la densidad objetivo y por ende, del estudio de lo preciso de las simulaciones hechas aquí, por lo que los resultados anteriores deben ser sujetos a un análisis más detallado para conocer qué tan acertados son.

Para concluir, comprobamos que en efecto las propuestas hechas antes corresponden a Kérneles de Gibss. En efecto. Recordemos que $\bar{\lambda}_{-i}$ representa al conjunto de datos $\lambda_1, \dots, \lambda_n$, pero con el i -ésimo dato suprimido. Utilizando que los parámetros λ_i son independientes entre sí, calculamos:

$$f(\lambda_i | \bar{\lambda}_{-i}, \beta, \bar{t}) = \frac{f(\bar{\lambda}, \beta | \bar{t})}{f(\bar{\lambda}_{-i}, \beta | \bar{t})} \propto \frac{\mathcal{L}(\bar{t} | \bar{\lambda}, \beta) f(\lambda_1 | \beta) \cdot \dots \cdot f(\lambda_n | \beta) f(\beta)}{\mathcal{L}(\bar{t} | \bar{\lambda}_{-i}, \beta) f(\lambda_1 | \beta) \cdot \dots \cdot f(\lambda_{i-1} | \beta) f(\lambda_{i+1} | \beta) \cdot \dots \cdot f(\lambda_n | \beta) f(\beta)} = \frac{\mathcal{L}(\bar{t} | \bar{\lambda}, \beta)}{\mathcal{L}(\bar{t} | \bar{\lambda}_{-i}, \beta)} f(\lambda_i | \beta).$$

Y utilizando que $p_i \sim \text{Poisson}(\lambda_i t_i)$, que $\lambda_i | \beta \sim \text{Gamma}(\alpha, \beta)$, y que los elementos p_i y t_i son, para nuestros propósitos, constantes, se tiene que

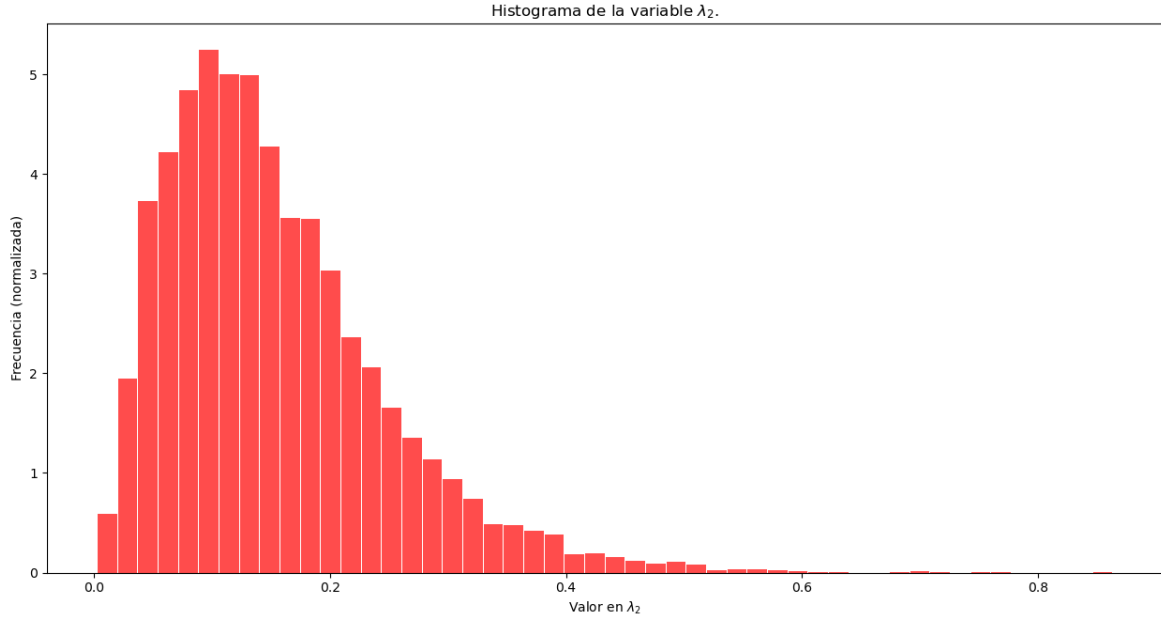


Figura 14: Histograma para λ_2

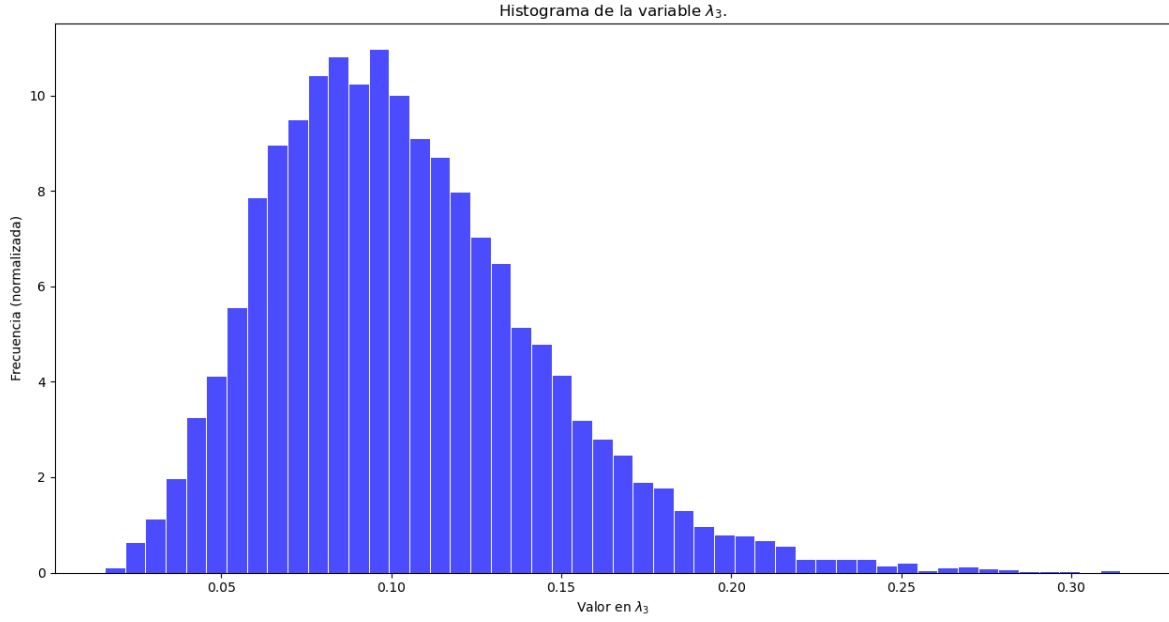


Figura 15: Histograma para λ_3

$$\begin{aligned}
 \frac{\mathcal{L}(\bar{t}|\bar{\lambda}, \beta)}{\mathcal{L}(\bar{t}|\bar{\lambda}_{-i}, \beta)} f(\lambda_i|\beta) &= \left(\prod_{j=1}^n \frac{(\lambda_j t_j)^{p_j} e^{-\lambda_j t_j}}{p_j!} \right) \left(\prod_{j \neq i} \frac{(\lambda_j t_j)^{p_j} e^{-\lambda_j t_j}}{p_j!} \right)^{-1} \left(\frac{(\beta \lambda_i)^{\alpha-1} \beta e^{-\beta \lambda_i}}{\Gamma(\alpha)} \right) \\
 &= \frac{(\lambda_i t_i)^{p_i} e^{-\lambda_i t_i}}{p_i!} \left(\frac{(\beta \lambda_i)^{\alpha-1} \beta e^{-\beta \lambda_i}}{\Gamma(\alpha)} \right) \\
 &\propto \lambda_i^{p_i + \alpha - 1} e^{-\lambda_i (t_i + \beta)} \\
 &\propto f_G(\lambda_i),
 \end{aligned}$$

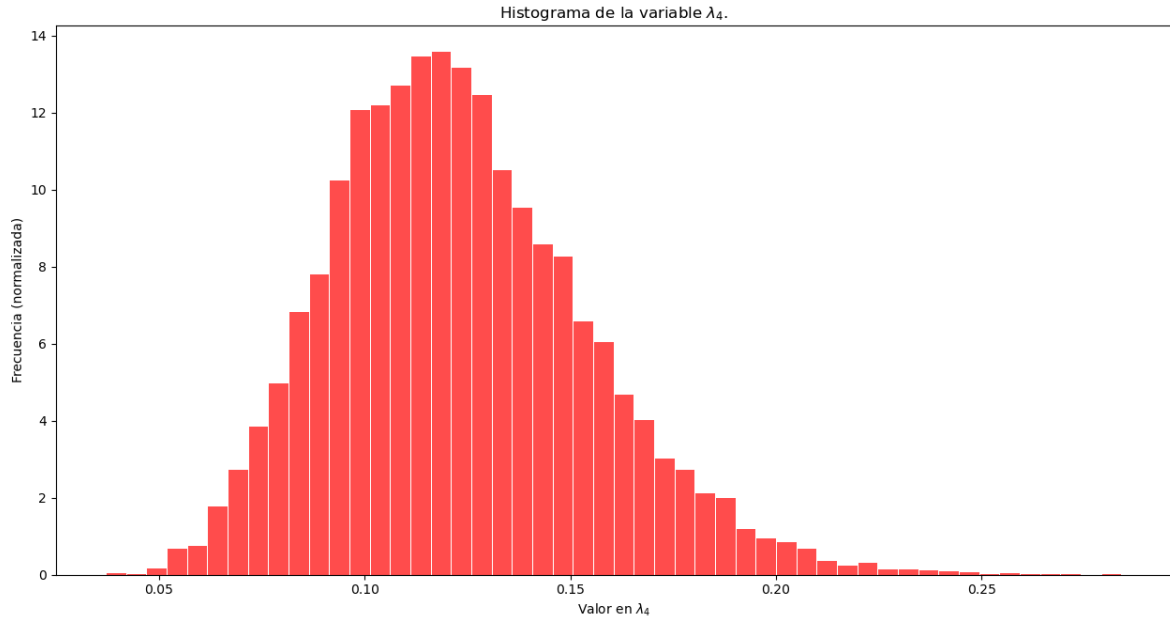


Figura 16: Histograma para λ_4

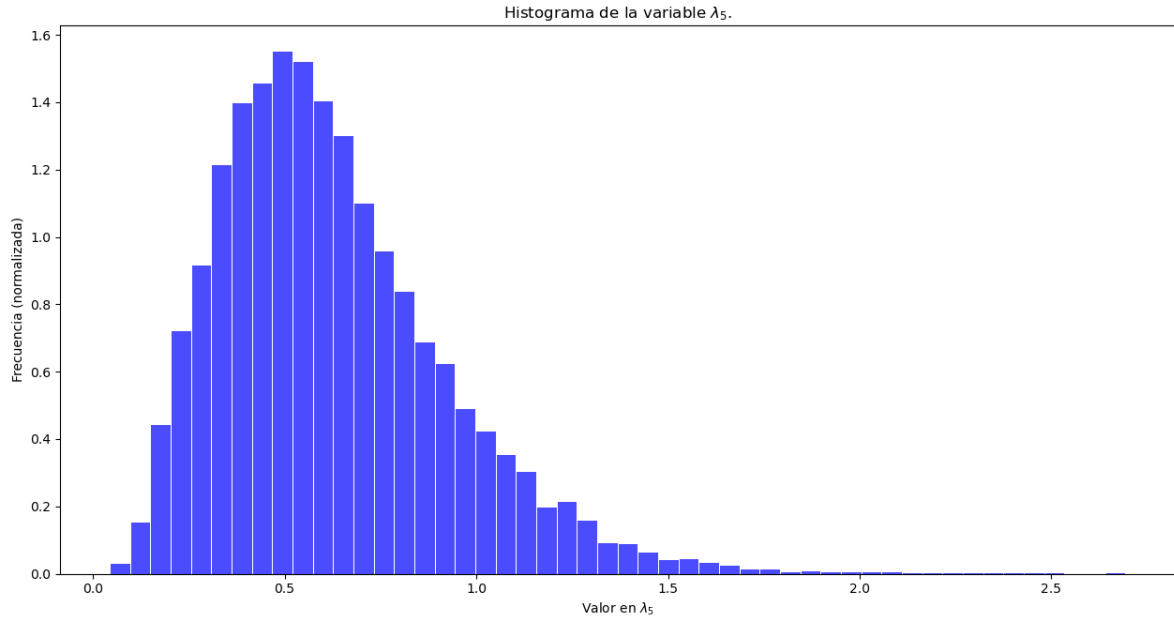


Figura 17: Histograma para λ_5

donde $G \sim \Gamma(p_i + \alpha, t_i + \beta)$, por lo que

$$\lambda_i | \bar{\lambda}_{-i}, \beta, \bar{t} \sim \Gamma(p_i + \alpha, t_i + \beta),$$

justo como buscábamos.

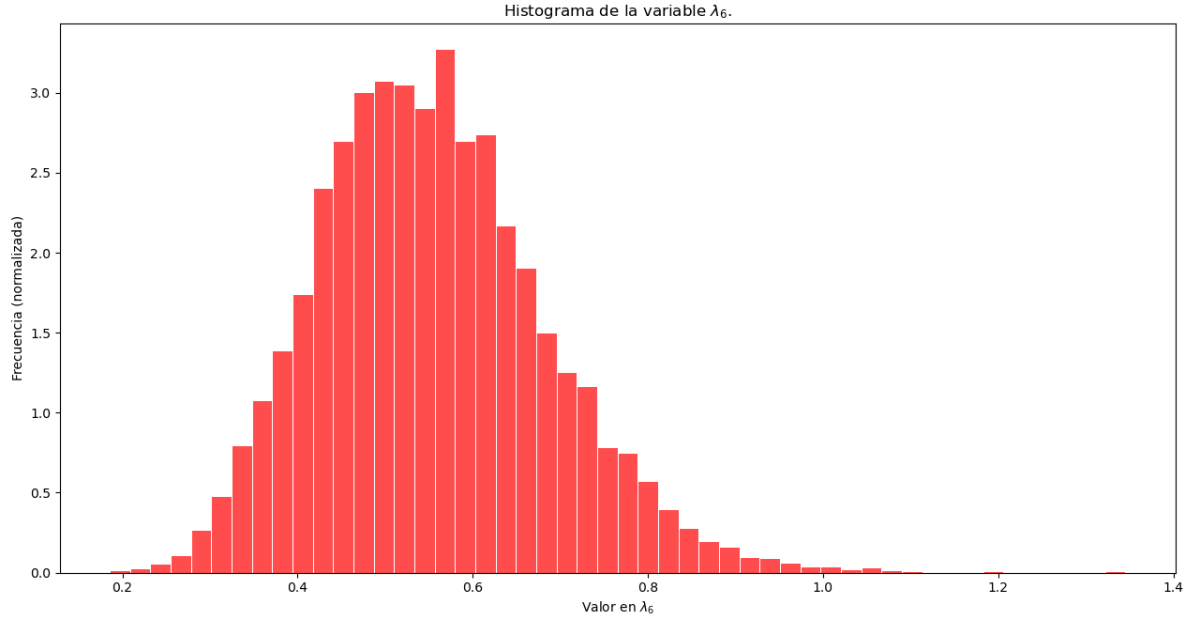


Figura 18: Histograma para λ_6

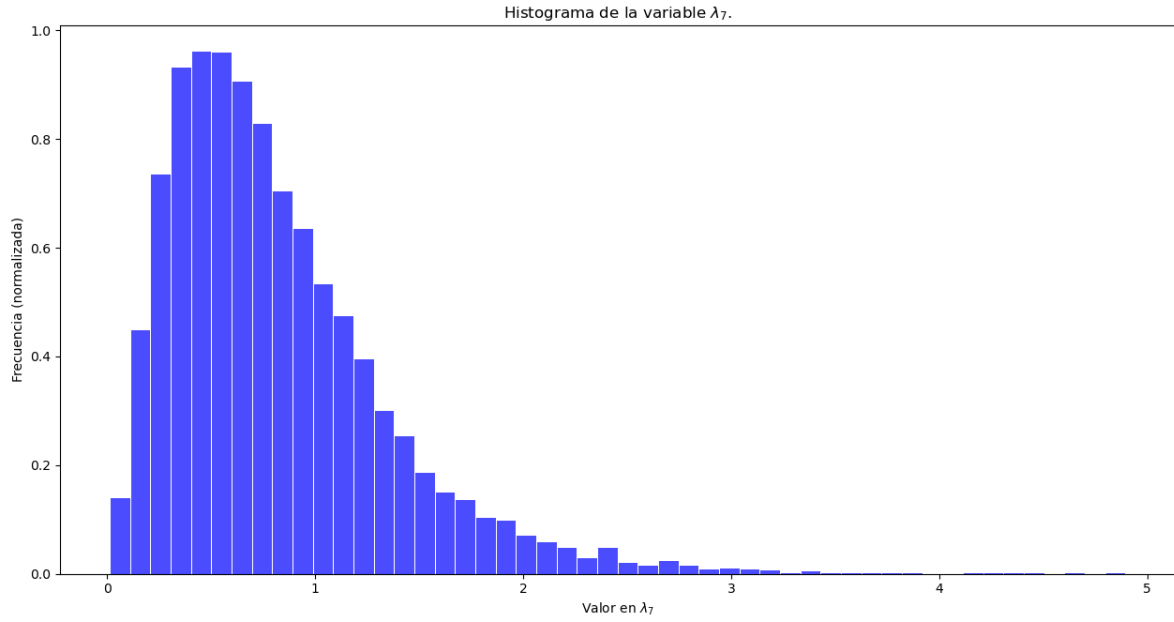


Figura 19: Histograma para λ_7

Finalmente, en el caso de β , notamos que

$$f(\lambda_1, \dots, \lambda_n, \beta | \bar{p}) \propto \mathcal{L}(\bar{p}, \bar{\lambda}, \beta) f(\lambda_1, \dots, \lambda_n, \beta) \propto f(\lambda_1 | \beta) \cdot \dots \cdot f(\lambda_n | \beta) f(\beta),$$

y sabiendo $\lambda_i | \beta \sim \text{Gamma}(\alpha, \beta)$, y que $\beta \sim \Gamma(\gamma, \delta)$, usando los parámetros que no involucren β como constantes, se

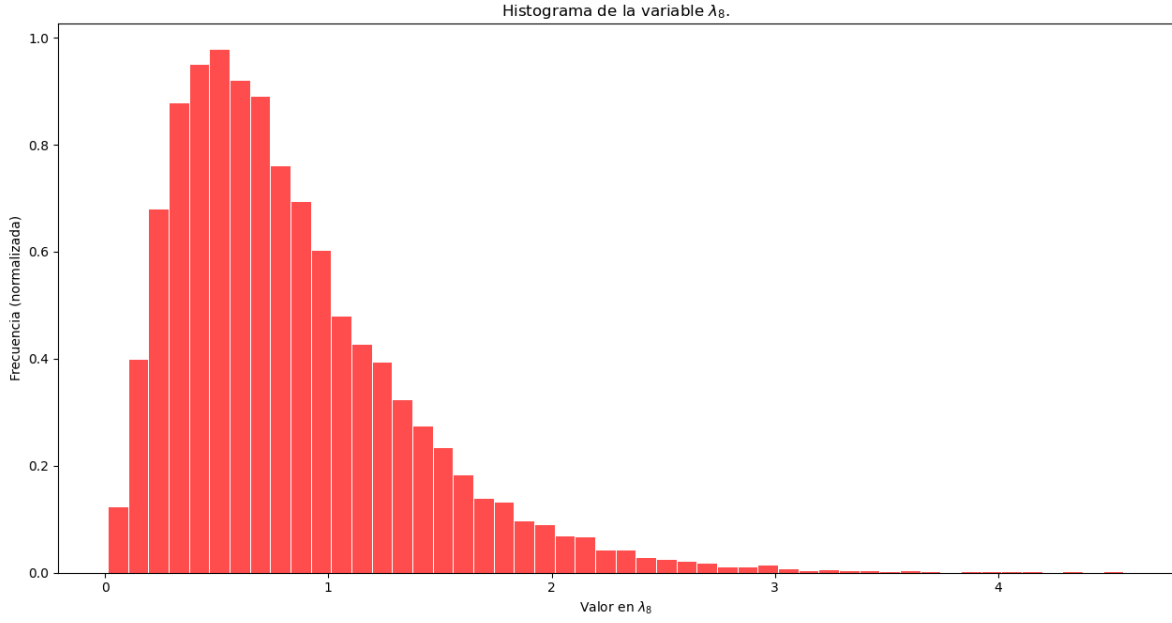


Figura 20: Histograma para λ_8

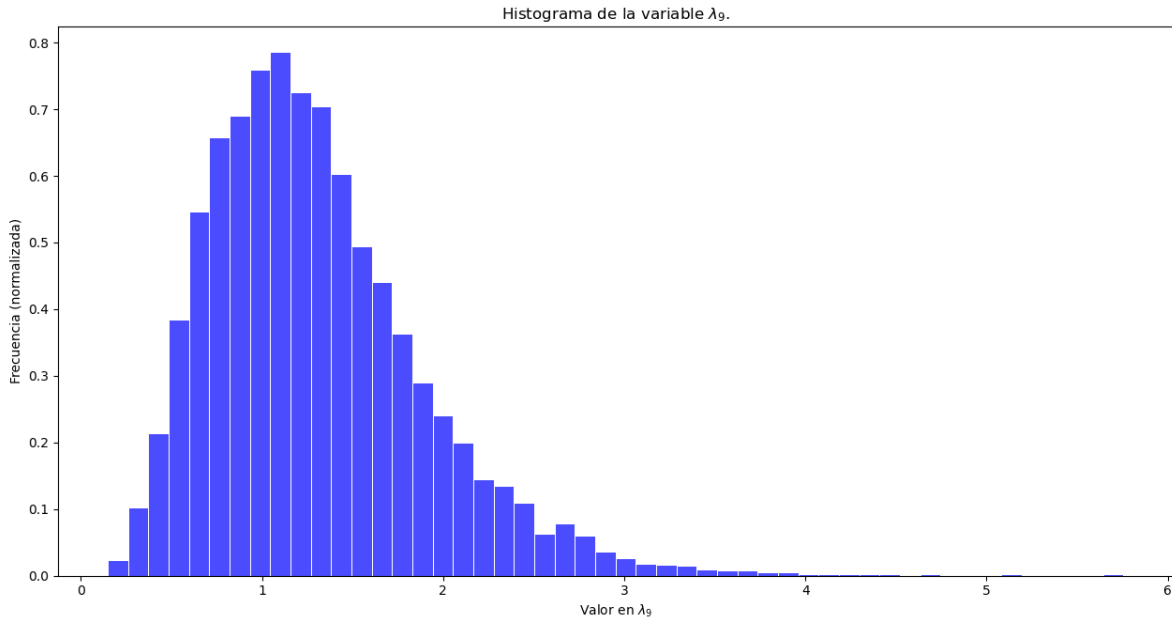


Figura 21: Histograma para λ_9

tiene que

$$f(\lambda_1, \dots, \lambda_n, \beta | \bar{p}) \propto \frac{\beta^{n\alpha+\gamma-1} \delta^\gamma e^{-\beta(\sum_{i=1}^n \lambda_i + \delta)} \prod_{i=1}^n \lambda_i^{\alpha-1}}{\Gamma(\alpha)^n \Gamma(\gamma)} \propto \beta^{n\alpha+\gamma-1} e^{-\beta(\sum_{i=1}^n \lambda_i + \delta)},$$

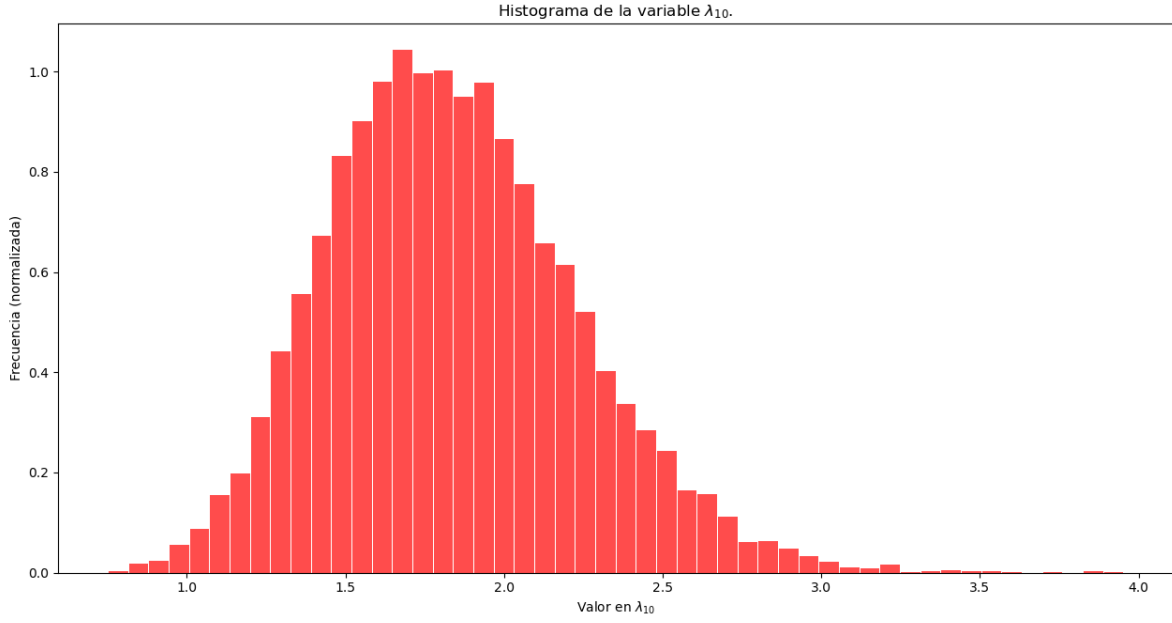


Figura 22: Histograma para λ_{10}

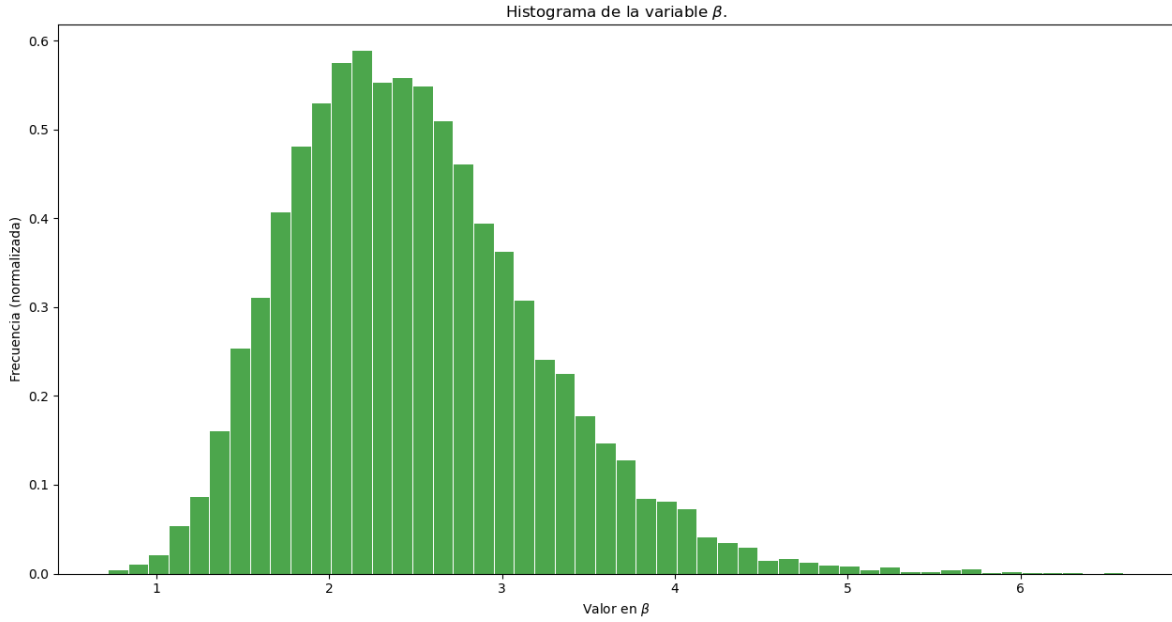


Figura 23: Histograma para β

y esta última función es a su vez proporcional a la función de densidad de una variable $\Gamma(n\alpha + \gamma, \sum_{i=1}^n \lambda_i + \delta)$. Se concluye entonces que

$$\beta | \bar{\lambda}, \bar{t} \sim \Gamma(n\alpha + \gamma, \sum_{i=1}^n \lambda_i + \delta).$$

Se sigue que las propuestas anteriores son Kerneles Gibbs.