Muestreo de Gibbs

José Antonio Perusquía Cortés

Estadística Bayesiana, Semestre 2026-I

Algoritmo

- Inicializa el algoritmo con puntos iniciales $\left(x_1^{(0)}, \ldots, x_p^{(0)}\right)$

- Al tiempo t:
  - Simular $X_1^{(t)} \sim f\left(\cdot \mid X_2^{(t-1)}, \ldots, X_p^{(t-1)} \right)$
  - Simular $X_2^{(t)} \sim f\left(\cdot \mid X_1^{(t)}, X_3^{(t-1)}, \ldots, X_p^{(t-1)} \right)$

    $\vdots$
  - Simular $X_p^{(t)} \sim f\left(\cdot \mid X_1^{(t)}, \ldots, X_{p-1}^{(t)} \right)$

Punto de cambio Poisson

- Considerar el siguiente modelo

$$X_i \sim \text{Po}(\lambda_1) \qquad i = 1, \ldots, M$$
$$X_i \sim \text{Po}(\lambda_2) \qquad i = M+1, \ldots, n$$

- Se considera a priori

$$f(\lambda_1, \lambda_2, M) = f(\lambda_1)(\lambda_2)f(M)$$

- Donde $\lambda_1 \sim Ga(\alpha_1, \beta_1)$, $\lambda_2 \sim Ga(\alpha_2, \beta_2)$ y $M \sim \text{Unif}\{1, \ldots, n-1\}$

- La posterior es proporcional a

$$f(\lambda_1, \lambda_2, M \mid \mathbf{x}^{(n)}) \propto \lambda_1^{\alpha_1 - 1 + \sum_{i=1}^{M} x_i} \exp(-(\beta_1 + M)\lambda_1)\lambda_2^{\alpha_2 - 1 + \sum_{i=M+1}^{n} x_i} \exp(-(\beta_2 + n - M)\lambda_2)$$

Punto de cambio Poisson

- Las condicionales de λ_1, λ_2

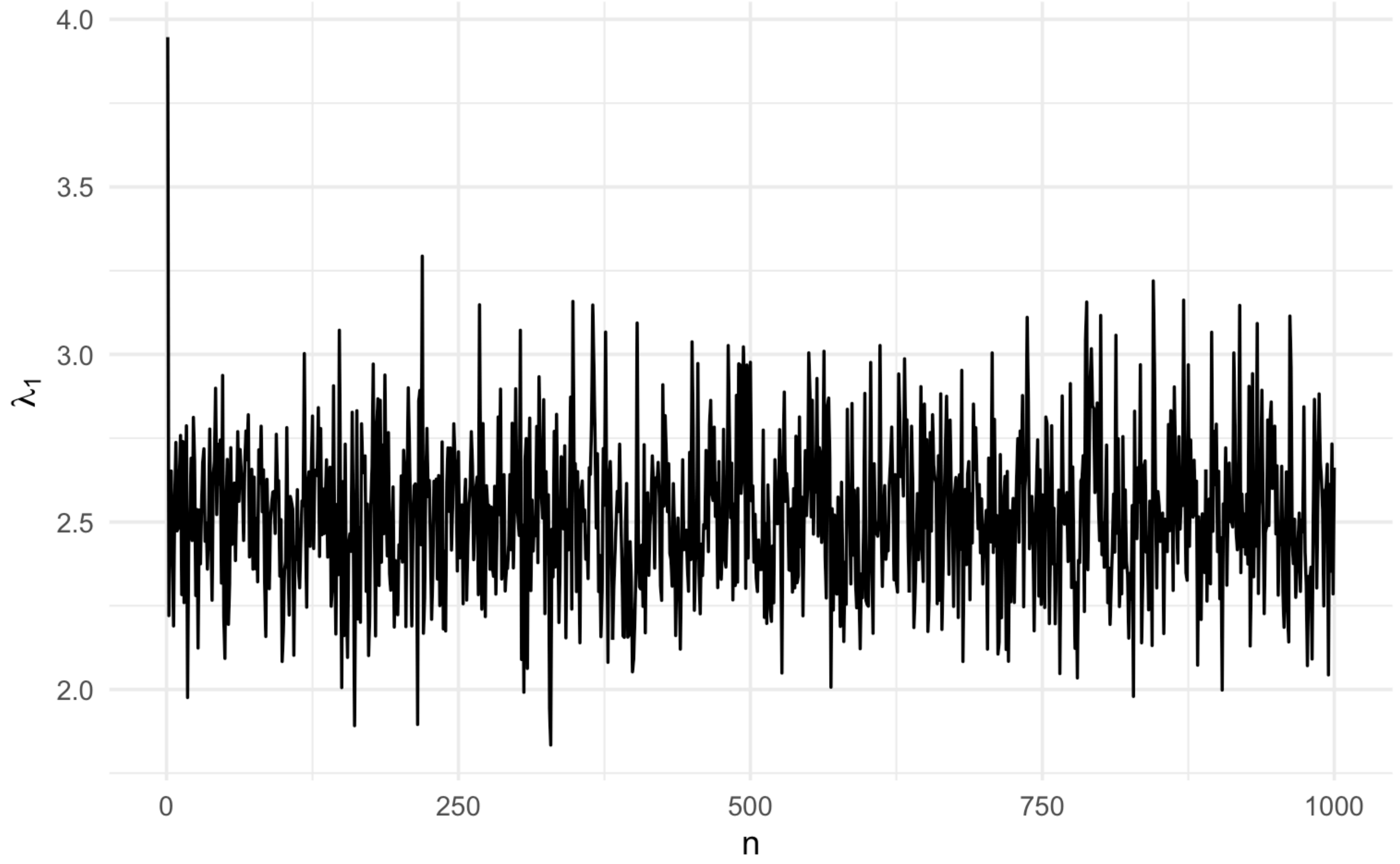
$$\lambda_1 \,|\, \mathbf{x}^{(n)}, M \sim \text{Ga}\left(\alpha_1 + \sum_{i=1}^{M} x_i, \beta_1 + M\right)$$

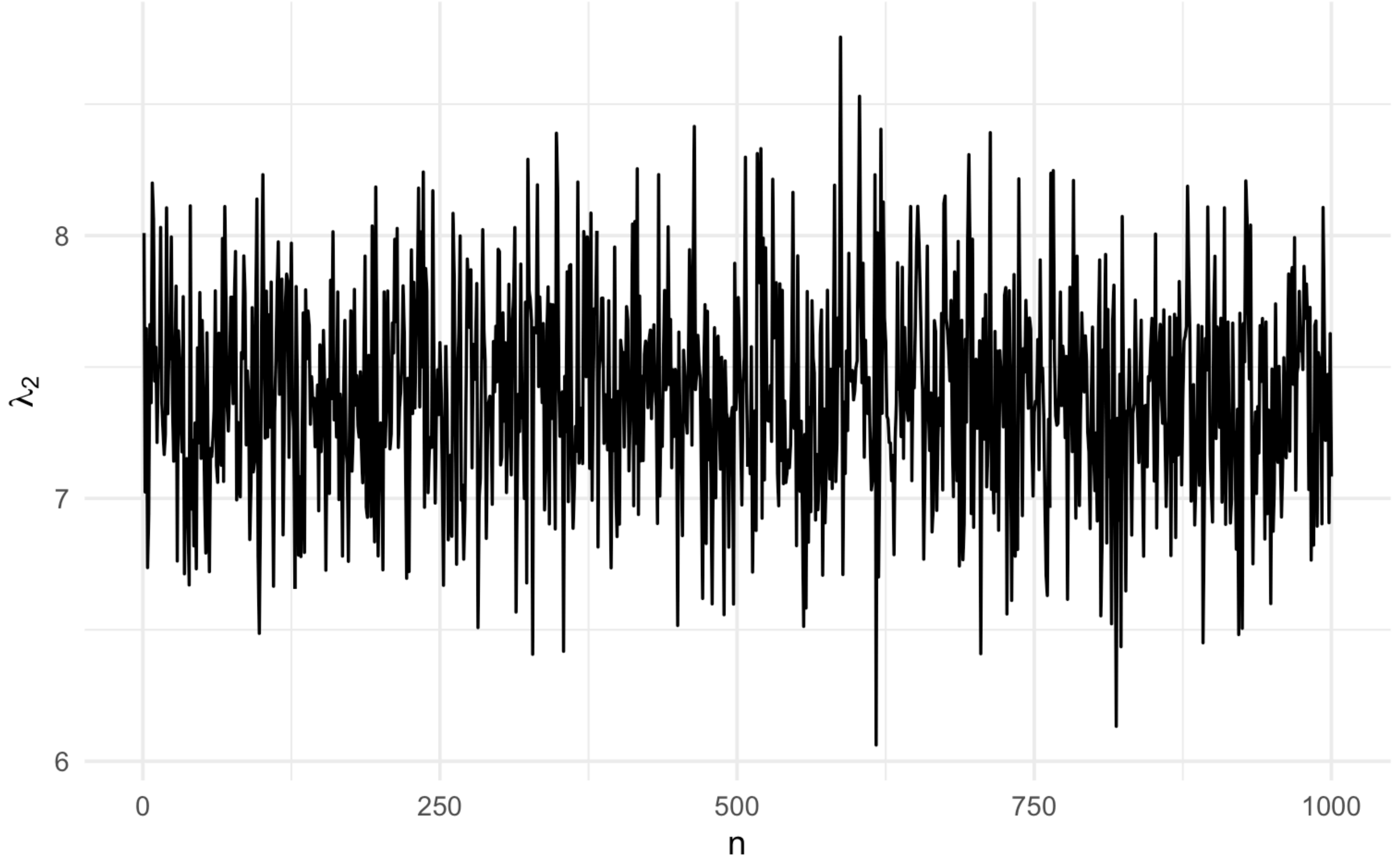
$$\lambda_2 \,|\, \mathbf{x}^{(n)}, M \sim \text{Ga}\left(\alpha_2 + \sum_{i=M+1}^{n} x_i, \beta_2 + n - M\right)$$

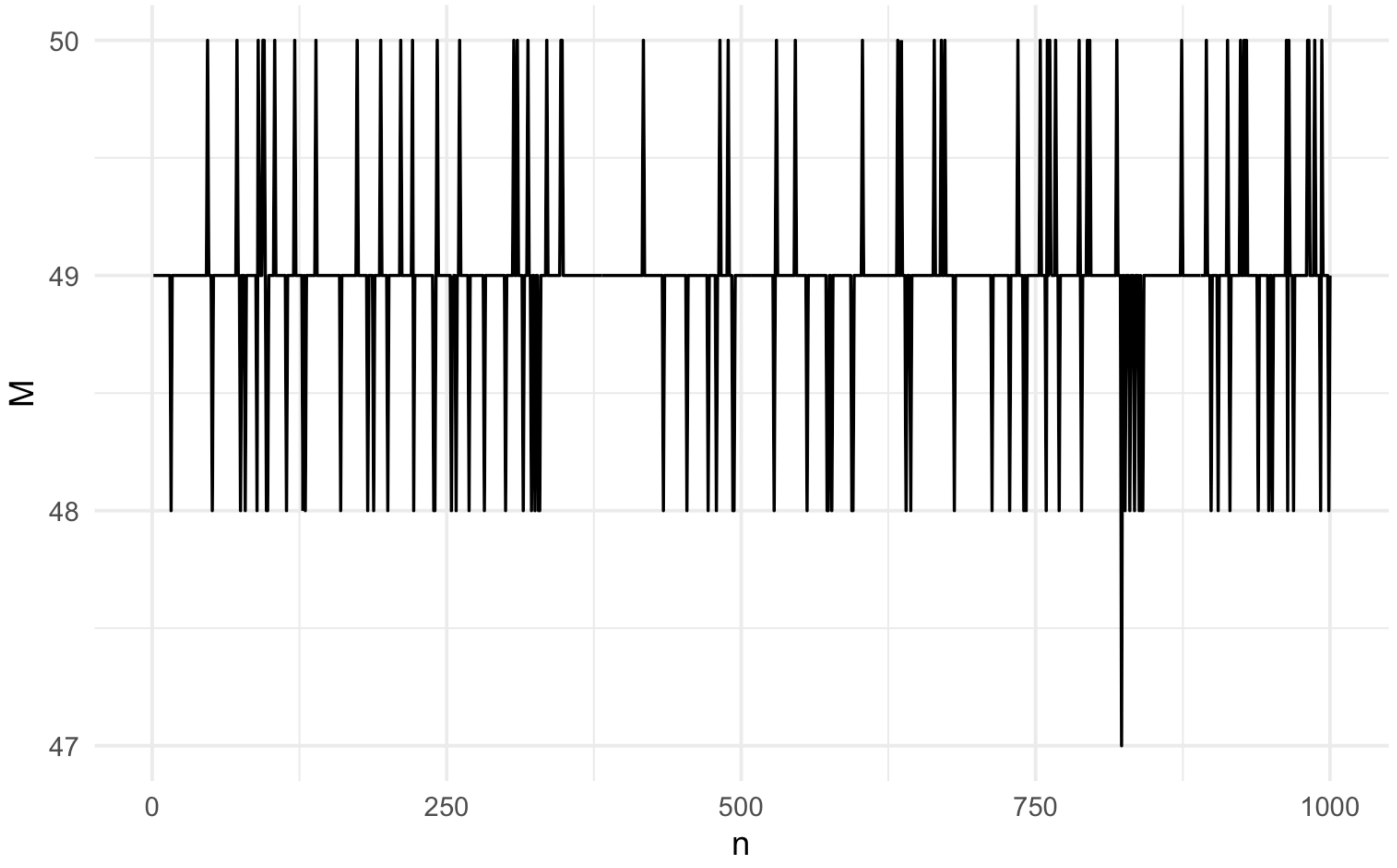
- Para cualquier valor de k en el soporte de M

$$\mathbb{P}\left(M = k \,|\, \lambda_1, \lambda_2, \mathbf{x}^{(n)}\right) \propto \left(\frac{\lambda_1}{\lambda_2}\right)^{\sum_{i=1}^{k} x_i} \exp((\lambda_2 - \lambda_1) \cdot k)$$

Punto de cambio Poisson







Condición de positividad

Definición

Una distribución con densidad $f(x_1, \ldots, x_n)$ y densidades marginales $f_i(x_i)$ satisface la condición de positividad si $f(x_1, \ldots, x_n) > 0$ para todo $x_1, \ldots, x_n$ tal que $f_i(x_i) > 0$.

- La condición de positividad implica que el soporte de la densidad conjunta f es el producto cartesiano de los soportes marginales.

Teorema de Hammersley - Clifford

Teorema

Sea $(X_1, \ldots, X_p)$ un vector multivariado con densidad conjunta $f(x_1, \ldots, x_p)$ que satisface la condición de positividad. Entonces para todo $(\xi_1, \ldots, \xi_p)$ en el soporte de f y cualquier permutación l en $\{1,2,\ldots,p\}$ se cumple que

$$f(x_1, \ldots, x_p) \propto \prod_{j=1}^{p} \frac{f_{l_j}(x_{l_j} \mid x_{l_1}, \ldots, x_{l_{j-1}}, \xi_{l_{j+1}}, \ldots, \xi_{l_p})}{f_{l_j}(\xi_{l_j} \mid x_{l_1}, \ldots, x_{l_{j-1}}, \xi_{l_{j+1}}, \ldots, \xi_{l_p})}.$$

- Tareita: Ver el caso (X_1, X_2, X_3)

Kernel de transición

Lema

El kernel de transición del algoritmo de Gibbs está dado por

$$K(\mathbf{x}^{(t-1)}, \mathbf{x}^{(t)}) = f_{X_1|X_{-1}}(x_1^{(t)} \mid x_2^{(t-1)}, \ldots, x_p^{(t-1)}) \cdots f_{X_p|X_{-p}}(x_p^{(t)} \mid x_1^{(t)}, \ldots, x_{p-1}^{(t)}) \,.$$

- No es reversible pero existen versiones reversibles del muestreo de Gibbs

Kernel de transición

Lema

La distribución conjunta $f(x_1, \ldots, x_p)$ es la distribución invariante de la cadena de Markov generada por el algoritmo de Gibbs.

- Demostración (ver las notas)

Kernel de transición

Lema

Si la distribución conjunta $f(x_1, \ldots, x_p)$ satisface la condición de positividad, el algoritmo Gibbs produce una cadena de Markov f-irreducible y recurrente.

- Demostración (ver las notas)
- Si el kernel es absolutamente continuo con respecto a la medida de Lebesgue entonces la cadena es Harris recurrente

Observaciones

- Las realizaciones no son independientes sino típicamente positivamente correlacionadas

- A mayor correlación la cadena se moverá lentamente por el espacio de estados (*slow mixing*)

Ejemplo normal bivariada

- Asumir que se desea simular de una distribución normal bivariada usando Gibbs

- La densidad conjunta es

$$f(x_1, x_2) \propto \exp\left(-\frac{(x_1 - (\mu_1 + \sigma_{12}/\sigma_{22}^2(x_2 - \mu_2)))^2}{2(\sigma_1^2 - (\sigma_{12})^2/\sigma_2^2)}\right)$$

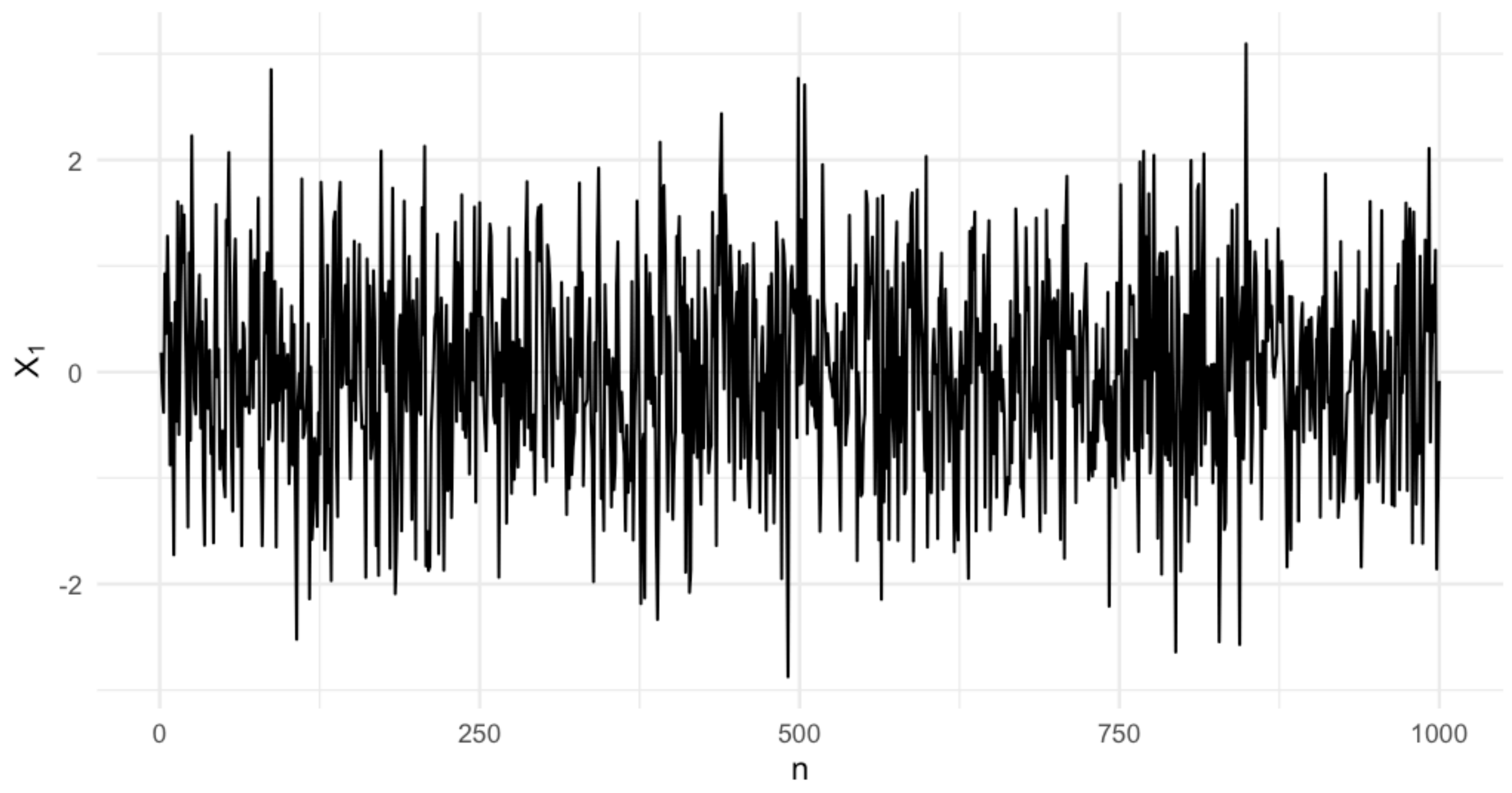
- Las marginales son

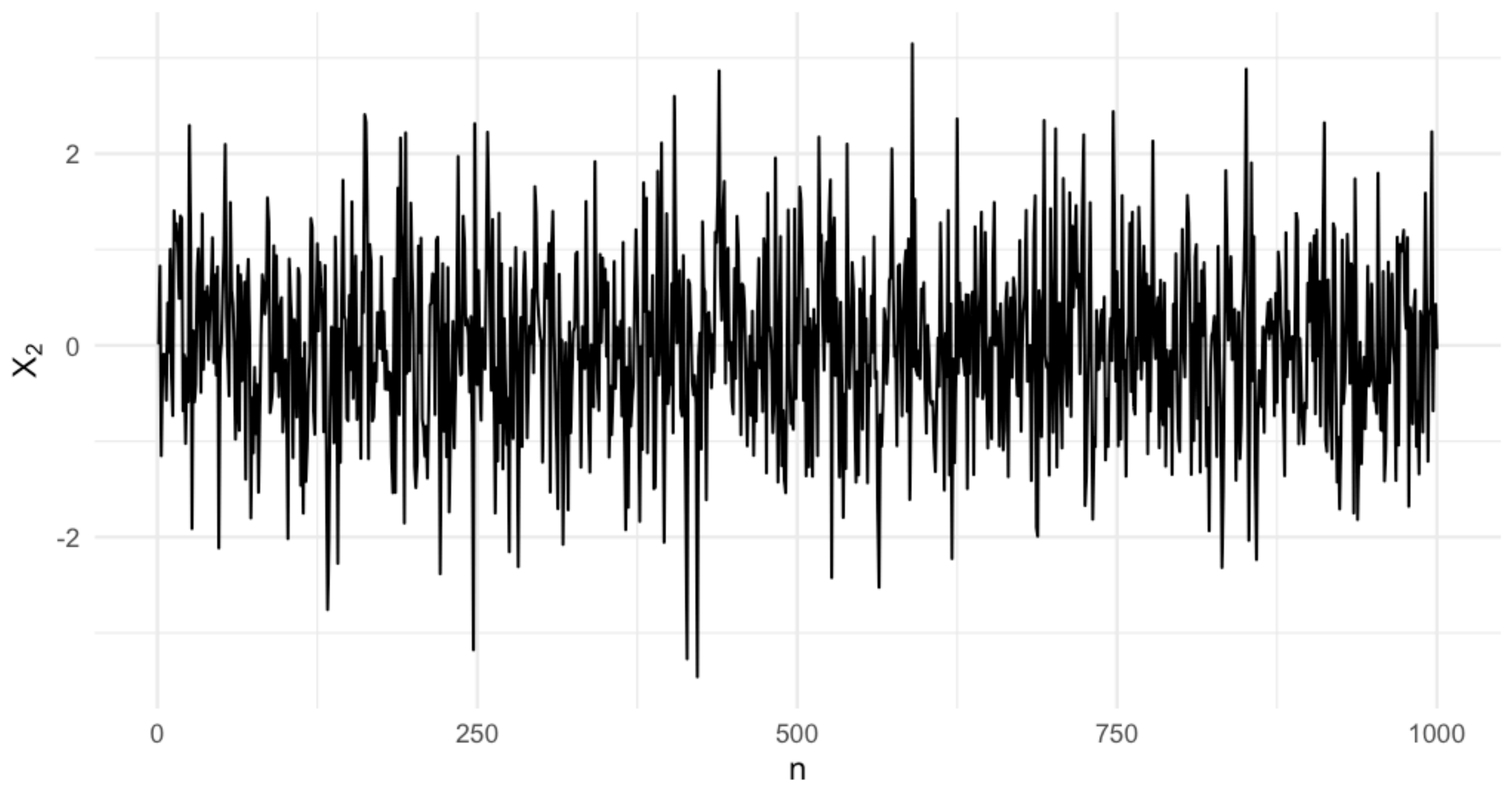
$$X_1 | X_2 = x_2 \sim \mathcal{N}\left(\mu_1 + \frac{\sigma_{12}}{\sigma_2^2}(x_2 - \mu_2), \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2}\right)$$

$$X_2 | X_1 = x_1 \sim \mathcal{N}\left(\mu_2 + \frac{\sigma_{12}}{\sigma_1^2}(x_1 - \mu_1), \sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2}\right)$$

Ejemplo normal bivariada

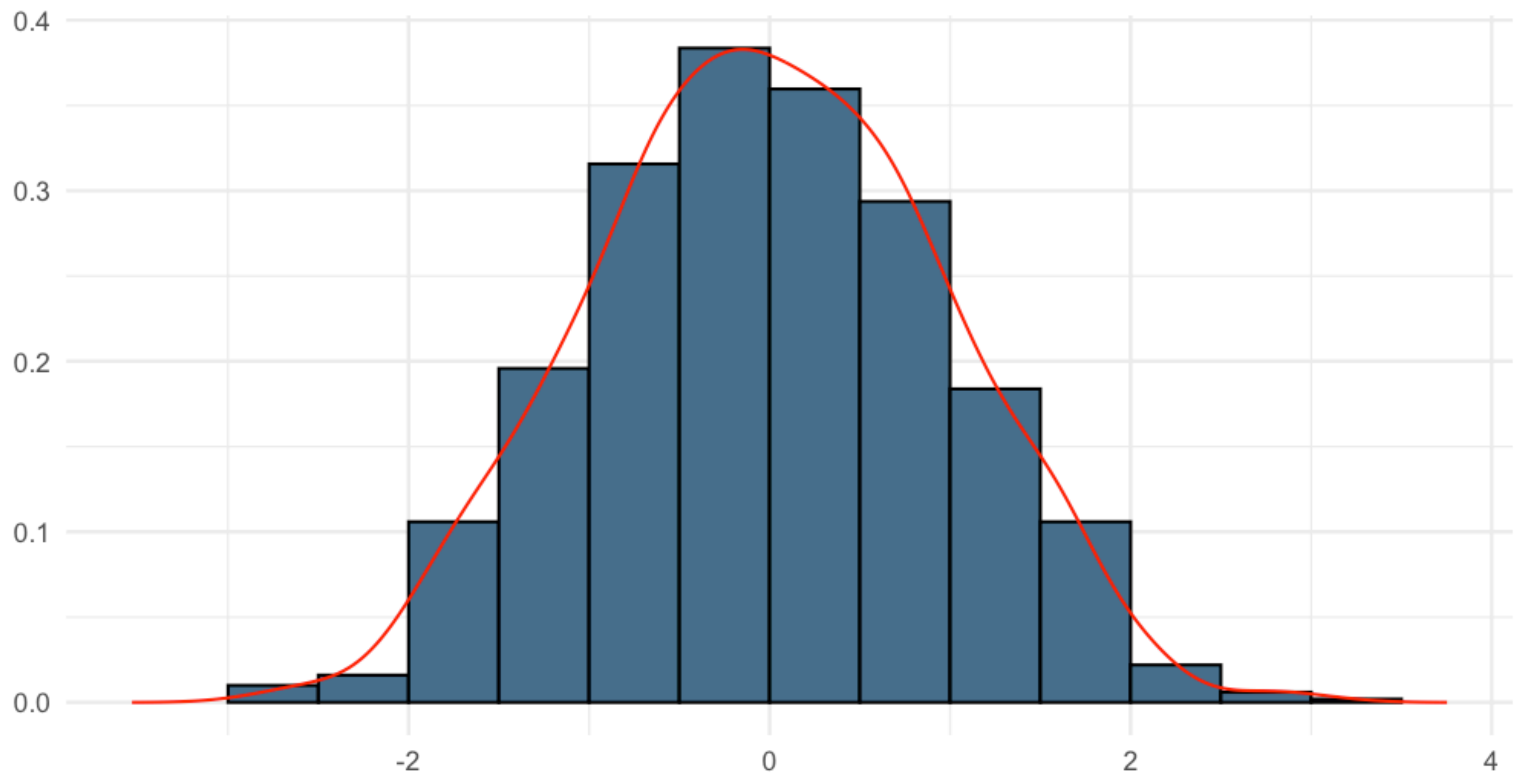
- Considerando $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$ y $\sigma_{12} = 0.3$

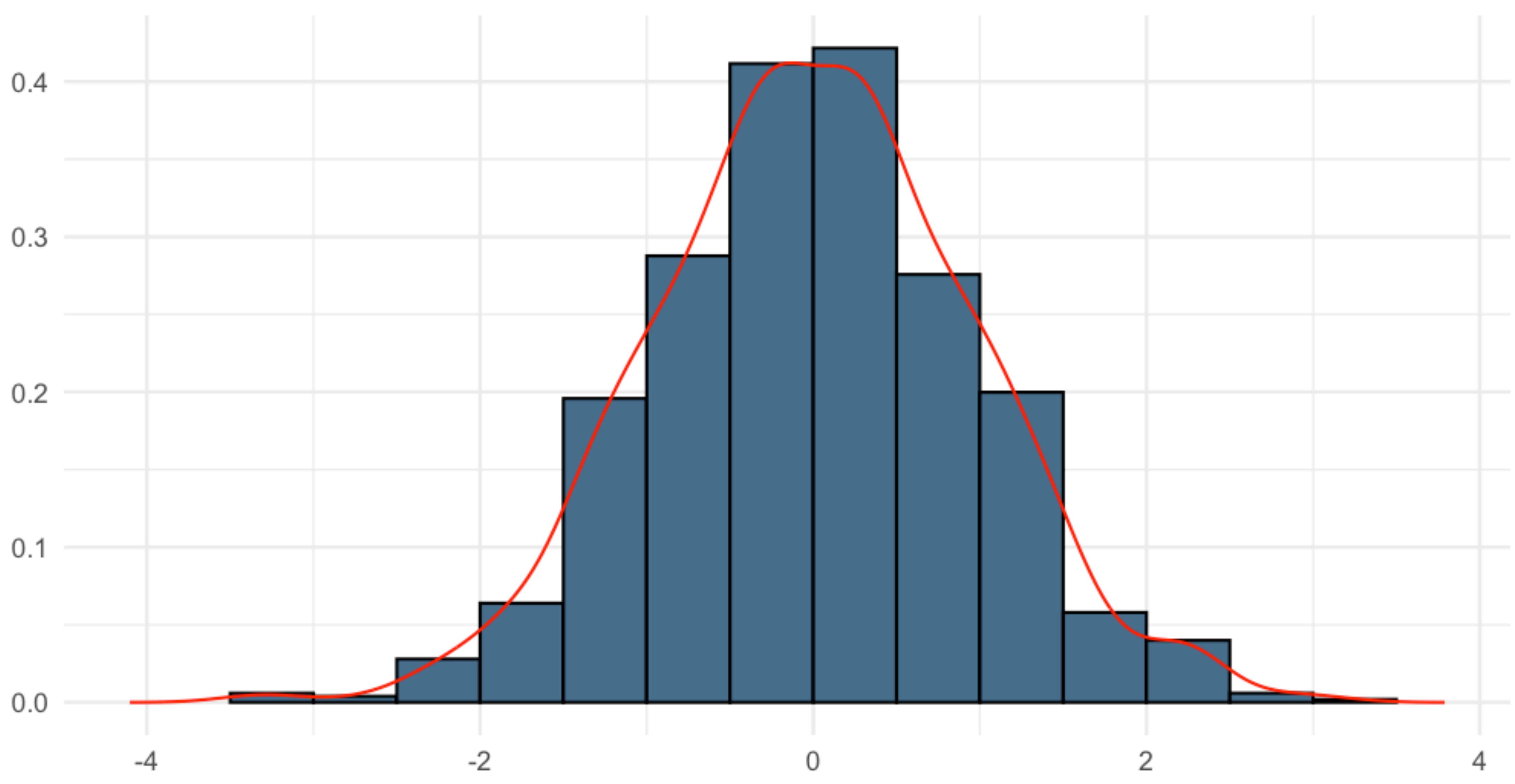




Ejemplo normal bivariada

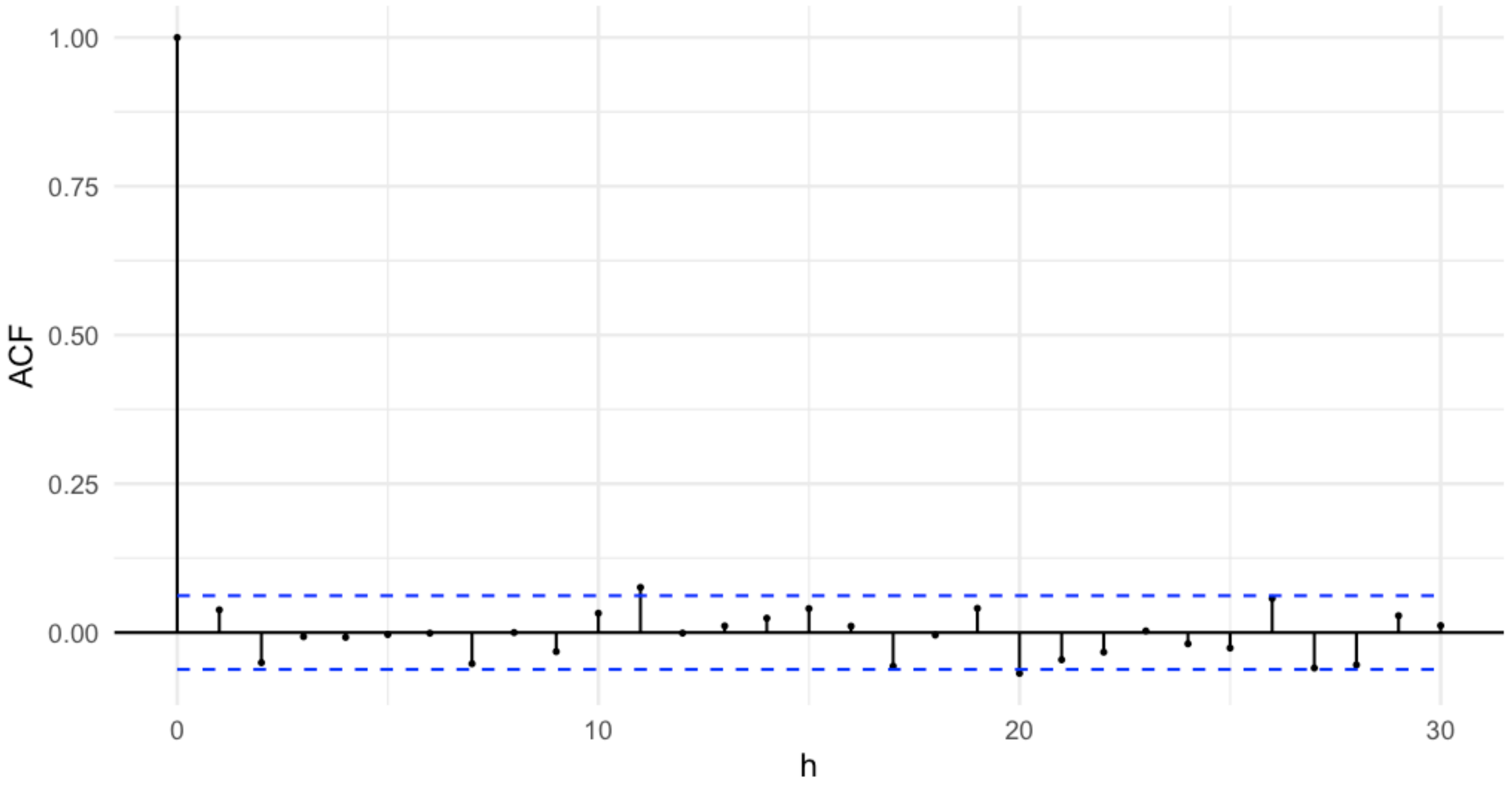
- Considerando $\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1$ y $\sigma_{12} = 0.3$

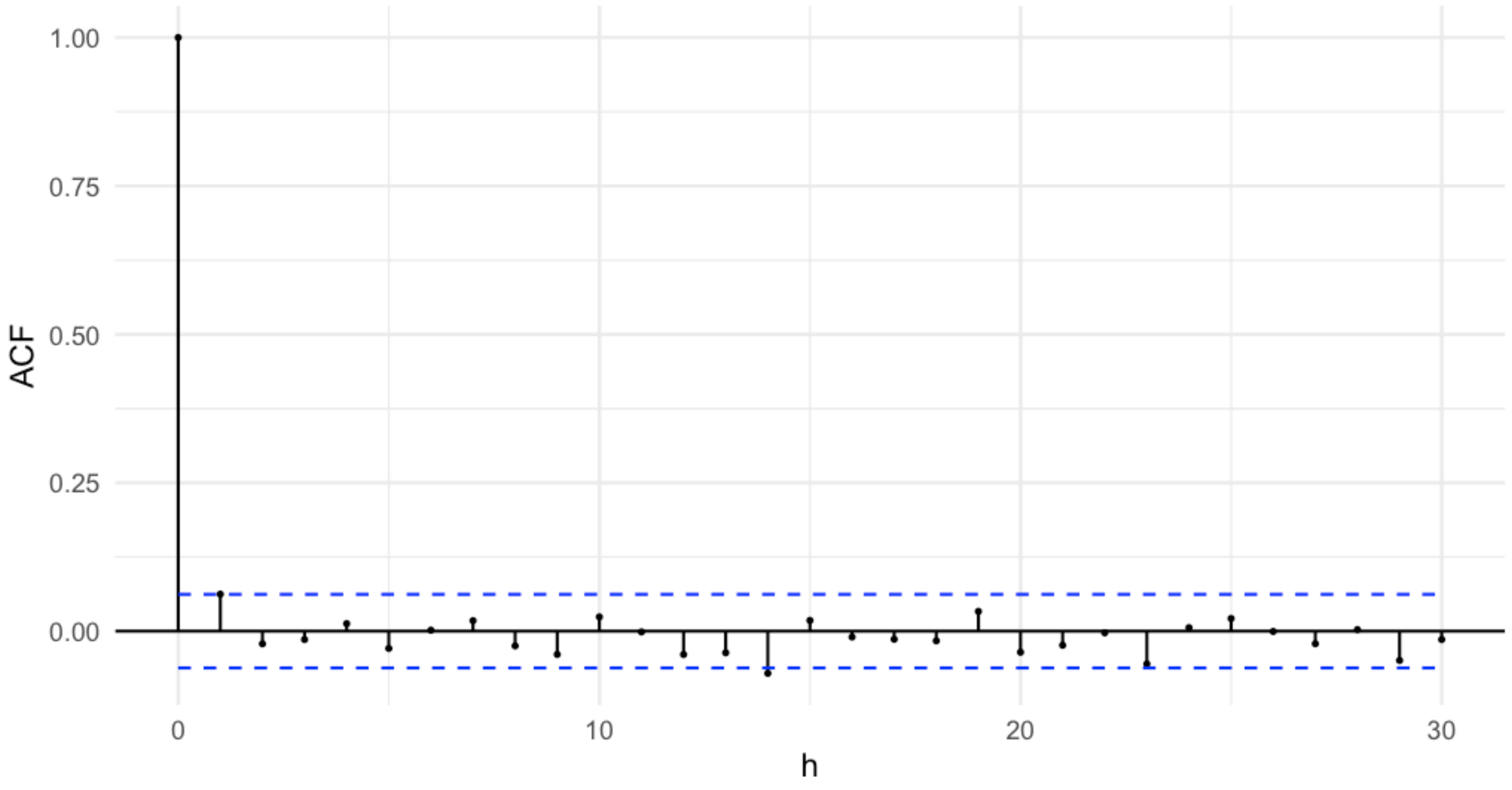




Ejemplo normal bivariada

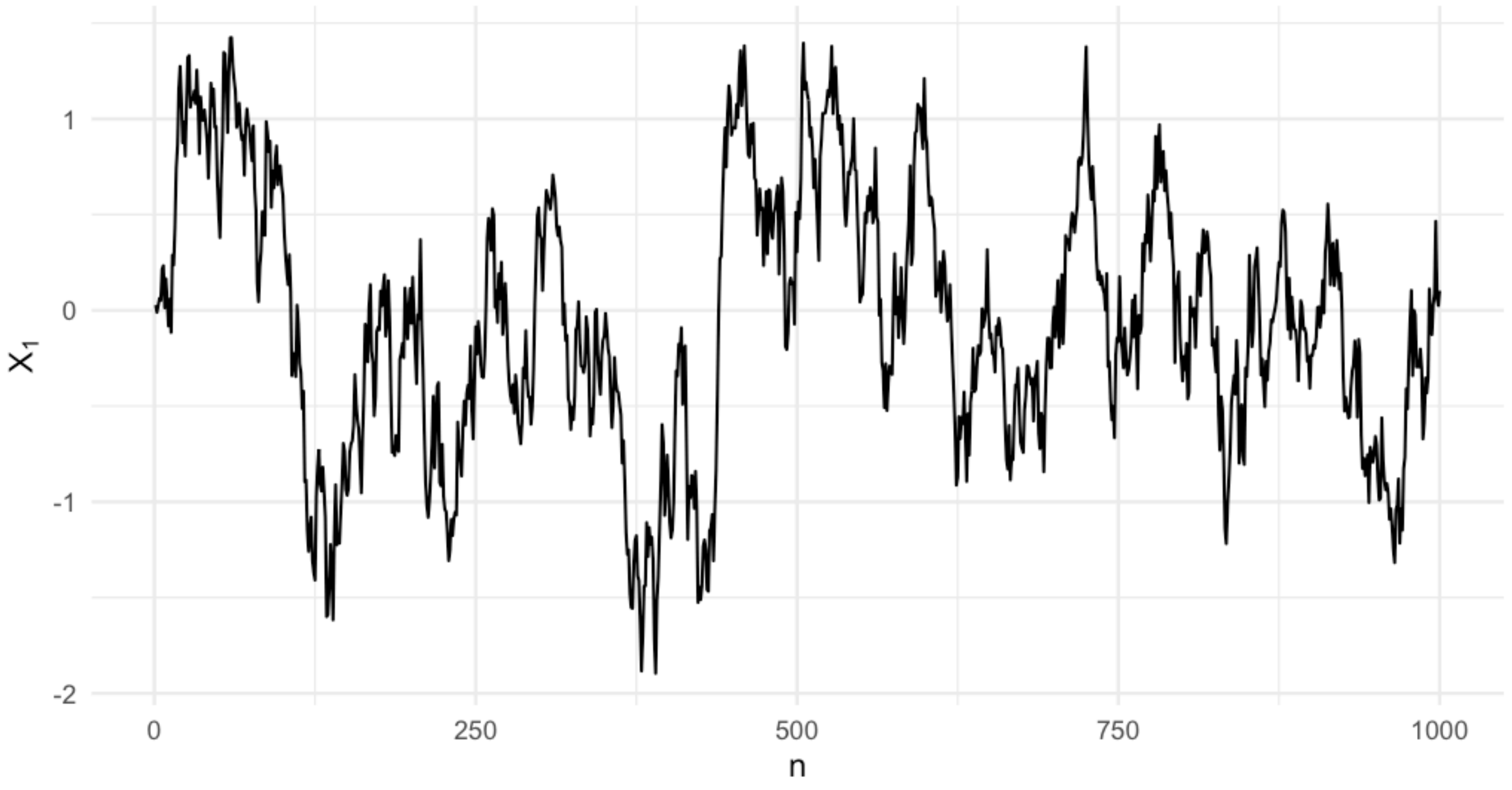
- Considerando $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$ y $\sigma_{12} = 0.3$

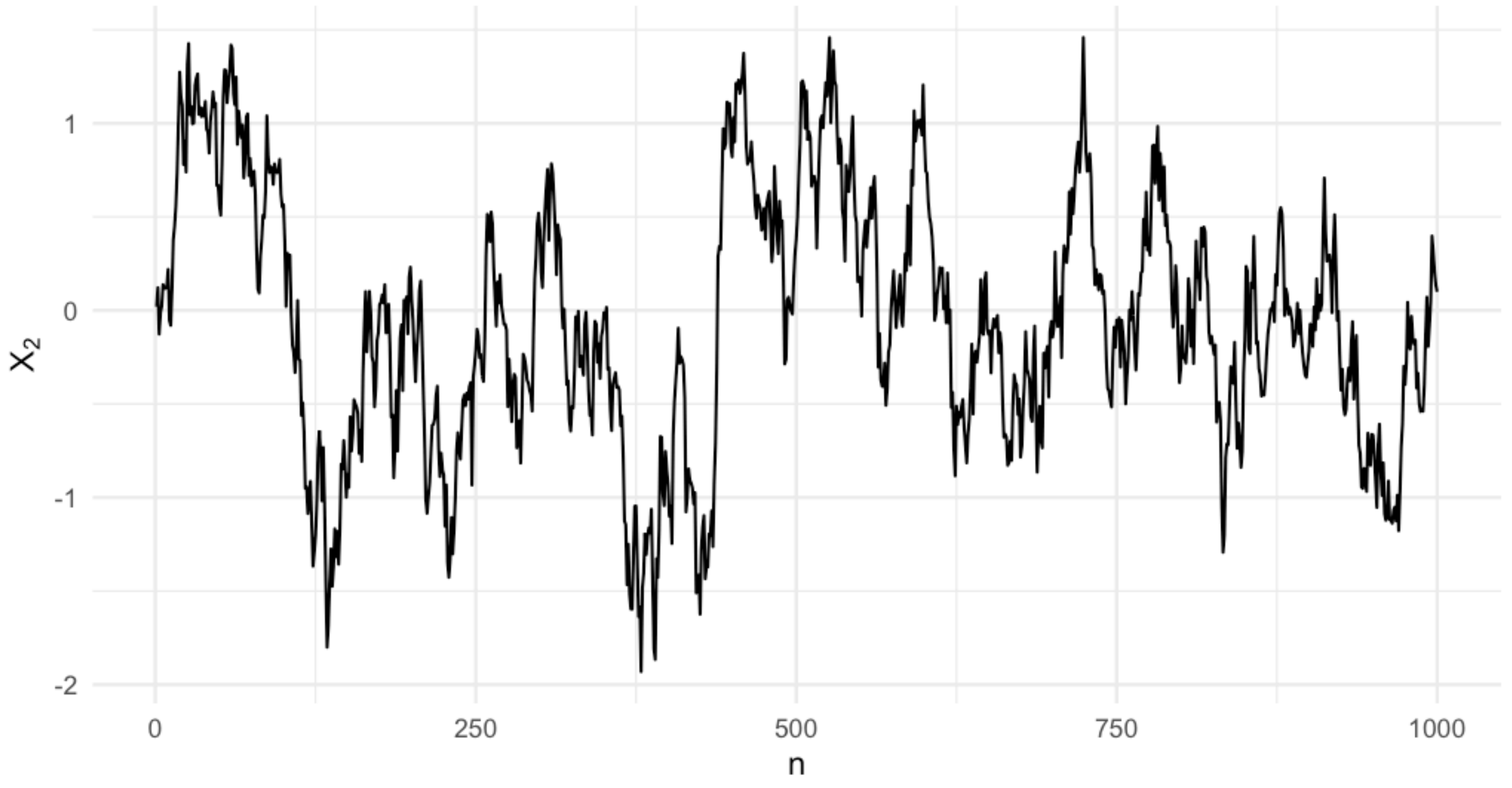




Ejemplo normal bivariada

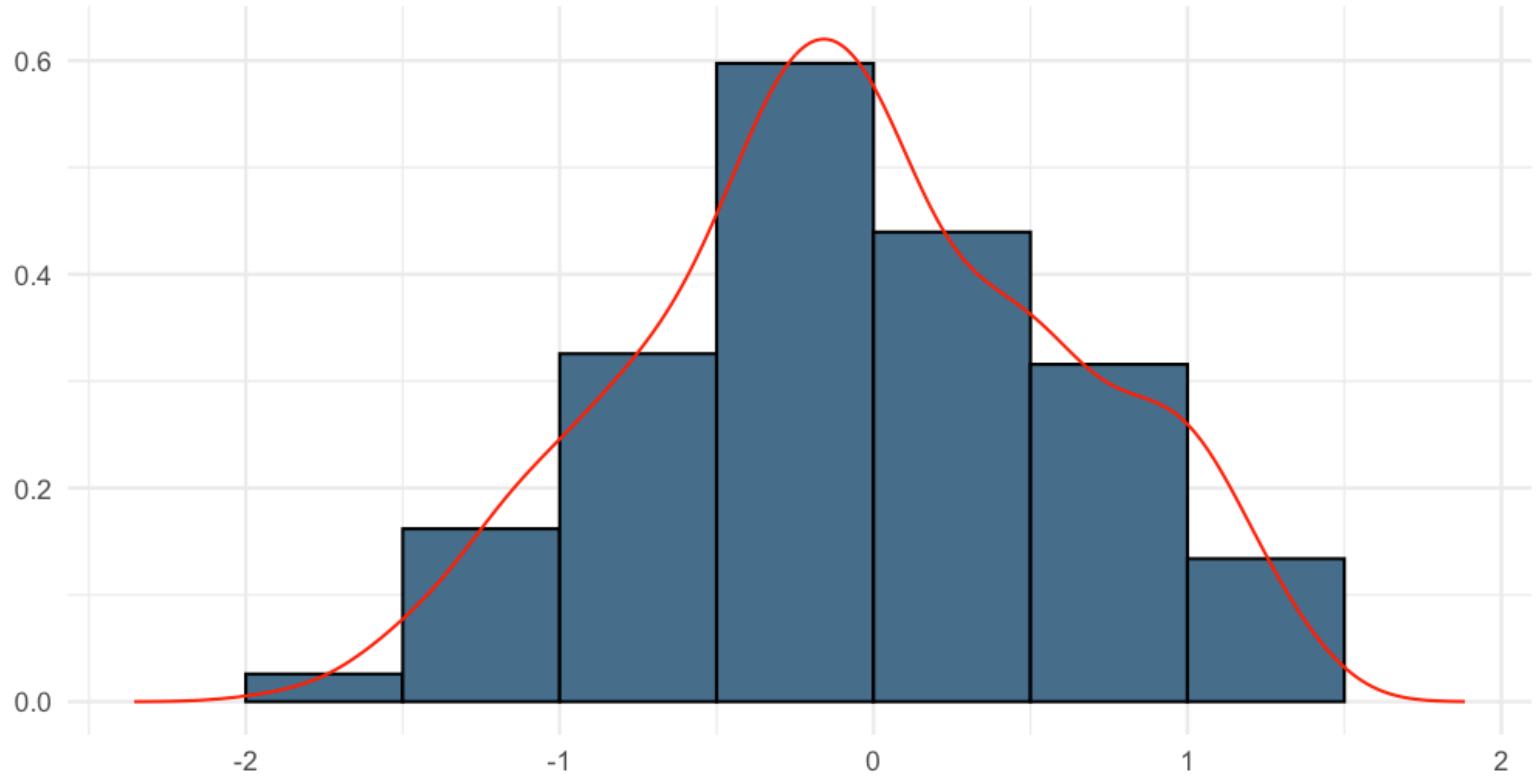
- Considerando $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$ y $\sigma_{12} = 0.99$

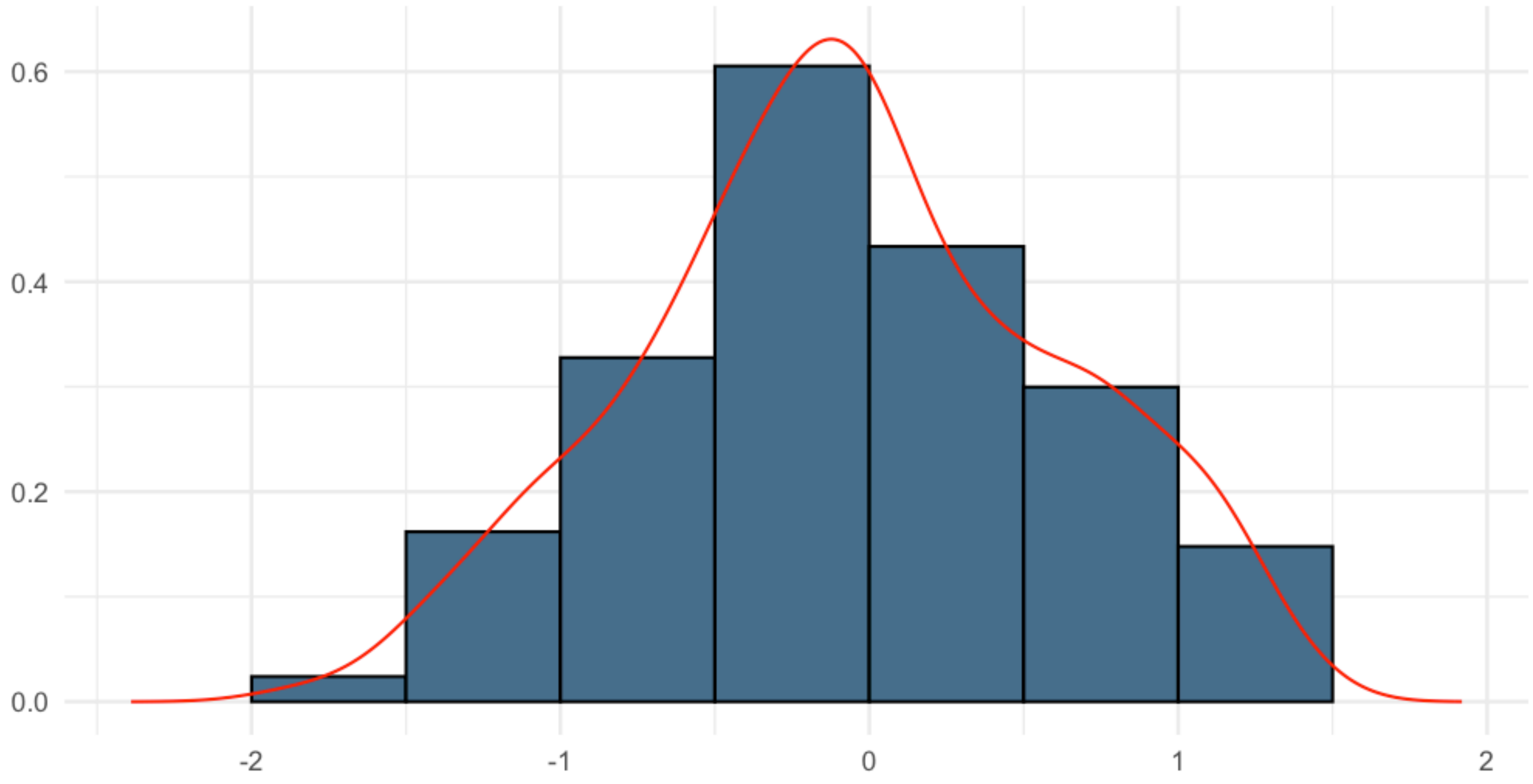




Ejemplo normal bivariada

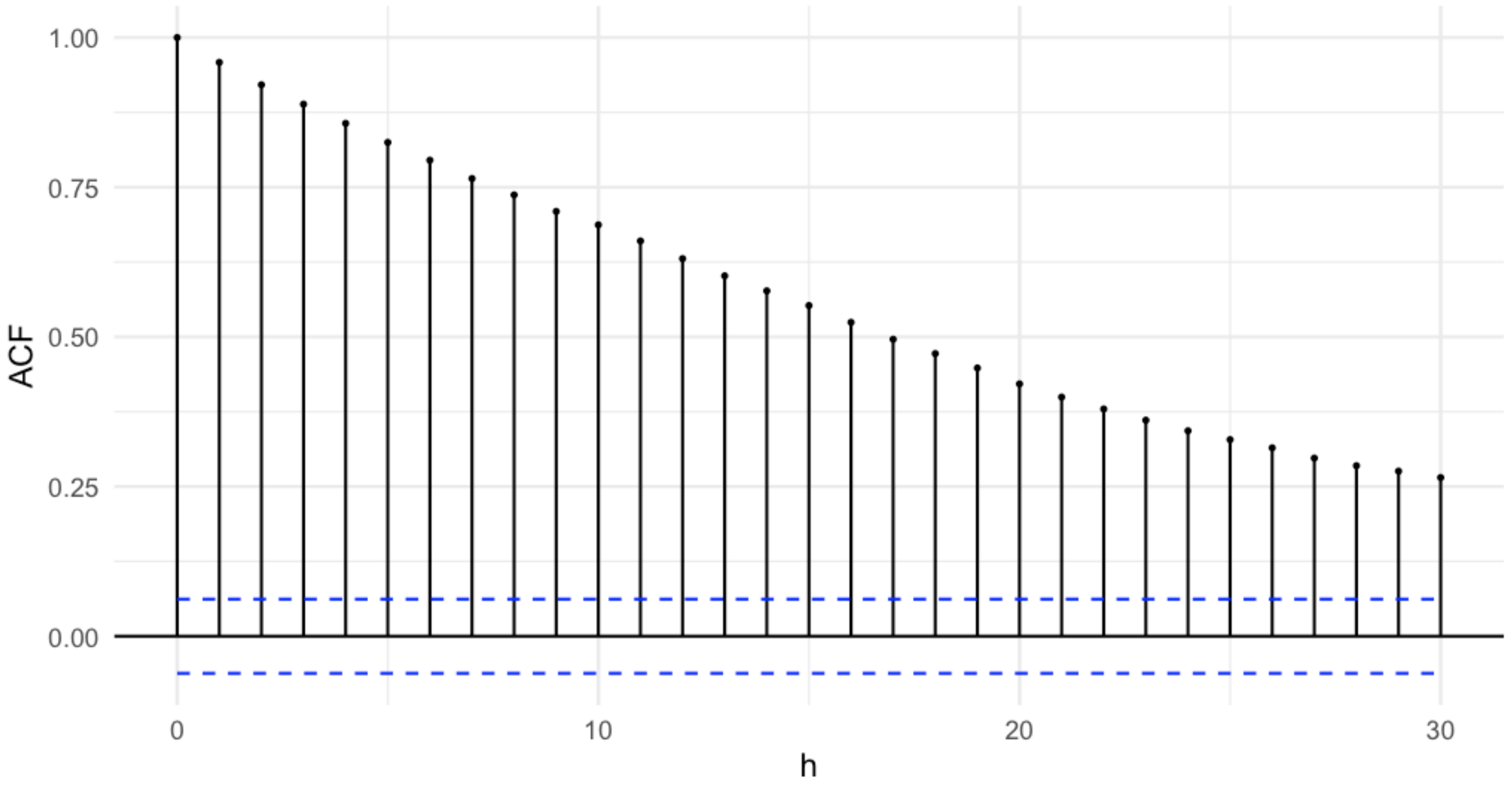
- Considerando $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$ y $\sigma_{12} = 0.99$

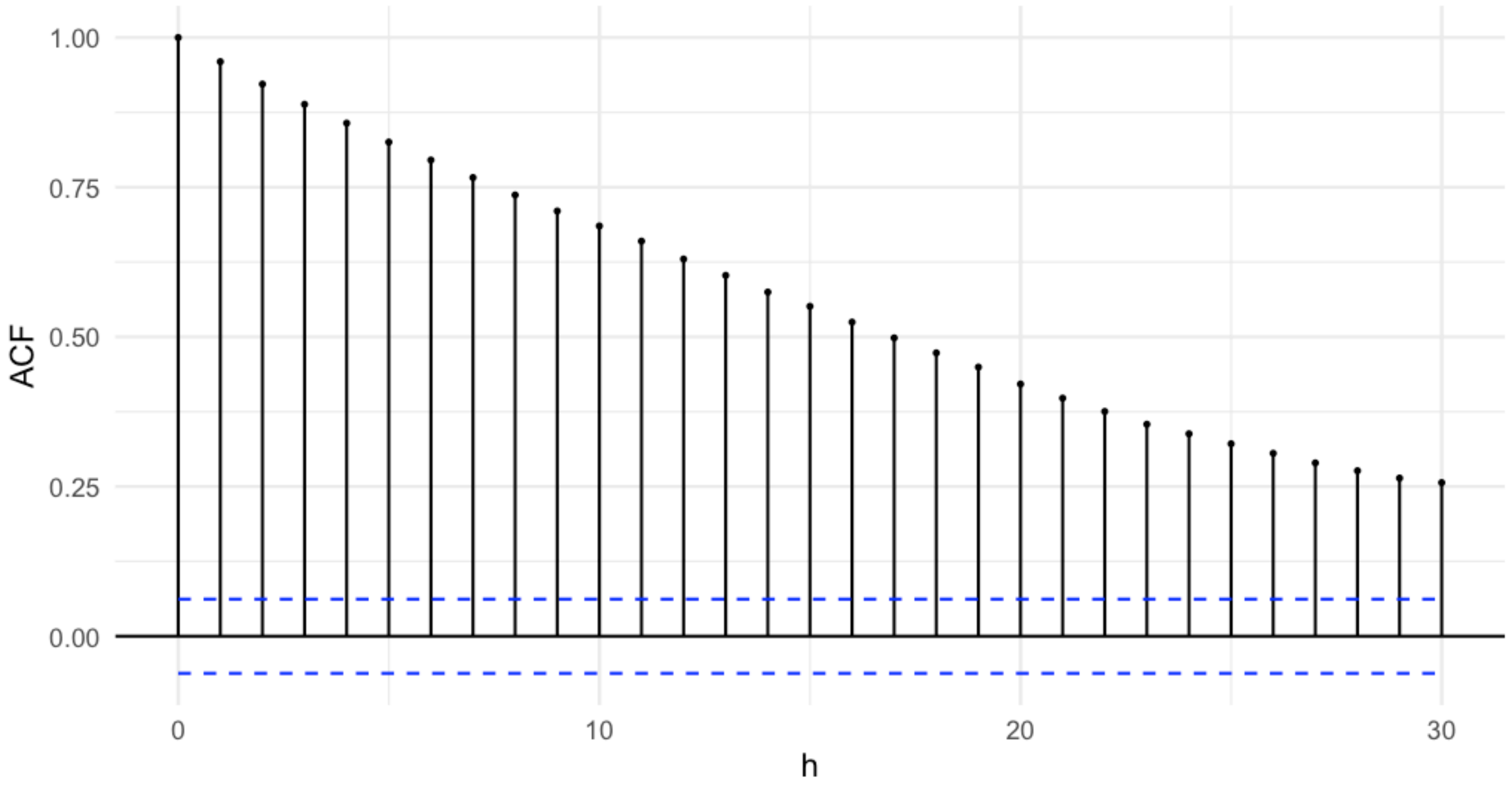




Ejemplo normal bivariada

- Considerando $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$ y $\sigma_{12} = 0.99$





Compleción

Definición

Dada una densidad de probabilidad f, una densidad g que satisfaga

$$\int g(x,z)dz = f(x)$$

se dice una compleción de f.

- Técnica de aumentación
- Elegir g tal que sus condicionales sean fáciles de simular
- Ejemplo conocido es el **slice sampler**

Slice sampler

Lema

Si $f(\theta) = \prod_{i=1}^{k} f_i(\theta)$ donde $f_i(\theta)$ son funciones positivas (no necesariamente densidades)

entonces f se puede completar como

$$\prod_{i=1}^{k} \mathbb{I}(0 \leq u_i \leq f_i(\theta))$$

Slice sampler

Algoritmo

Dados valores iniciales $u_1^{(0)}, \ldots, u_k^{(0)}, \theta^{(0)}$ al tiempo t:

- Simular $u_1^{(t)} \sim \text{Unif}\,(0, f_1(\theta^{(t-1)}))$
- Simular $u_2^{(t)} \sim \text{Unif}\,(0, f_2(\theta^{(t-1)}))$

  $\vdots$

- Simular $u_k^{(t)} \sim \text{Unif}\,(0, f_k(\theta^{(t-1)}))$
- Simular $\theta^{(t)} \sim \text{Unif}\,(A^{(t)})$ donde $A^{(t)} = \{y : f_i(y) \geq u_i^{(t)}, i = 1, \ldots, k\}$

Ejemplo normal truncada

- La densidad de la distribución normal truncada en $[\mu^*, \infty)$ es proporcional a

$$f(x) \propto \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) \mathbb{I}_{(\mu^*,\infty)}^{(x)} = f_1(x) f_2(x)$$

donde

$$f_1(x) = \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right), \quad f_2(x) = \mathbb{I}_{(\mu^*,\infty)}^{(x)}$$

- Se puede completar como

$$g(x,z) \propto \mathbb{I}_{(\mu^*,\infty)}^{(x)} \mathbb{I}_{(0,\exp\left(-(x-\mu)^2/2\sigma^2\right))}^{(z)}$$

Ejemplo normal truncada

- El algoritmo se traduce a simular variables uniformes en los soportes adecuados

- El soporte de Z está bien definido

$$\mathbb{1}^{(z)}_{\left(0, \exp\left(-(x-\mu)^2/2\sigma^2\right)\right)}$$

- Para X se tienen dos condiciones:

1. $x \geq \mu*$

2. Como $0 \leq z \leq \exp\left(-(x-\mu)^2/2\sigma^2\right)$ entonces $x \leq \mu + \sqrt{-2\sigma^2 \log(z)}$

Ejemplo normal truncada

- El algoritmo queda definido mediante las siguientes iteraciones

1. $x^{(t)} \mid z^{(t-1)} \sim \text{Unif}\left(\mu^*, \mu + \sqrt{-2\sigma^2 \log(z^{(t-1)})}\right)$

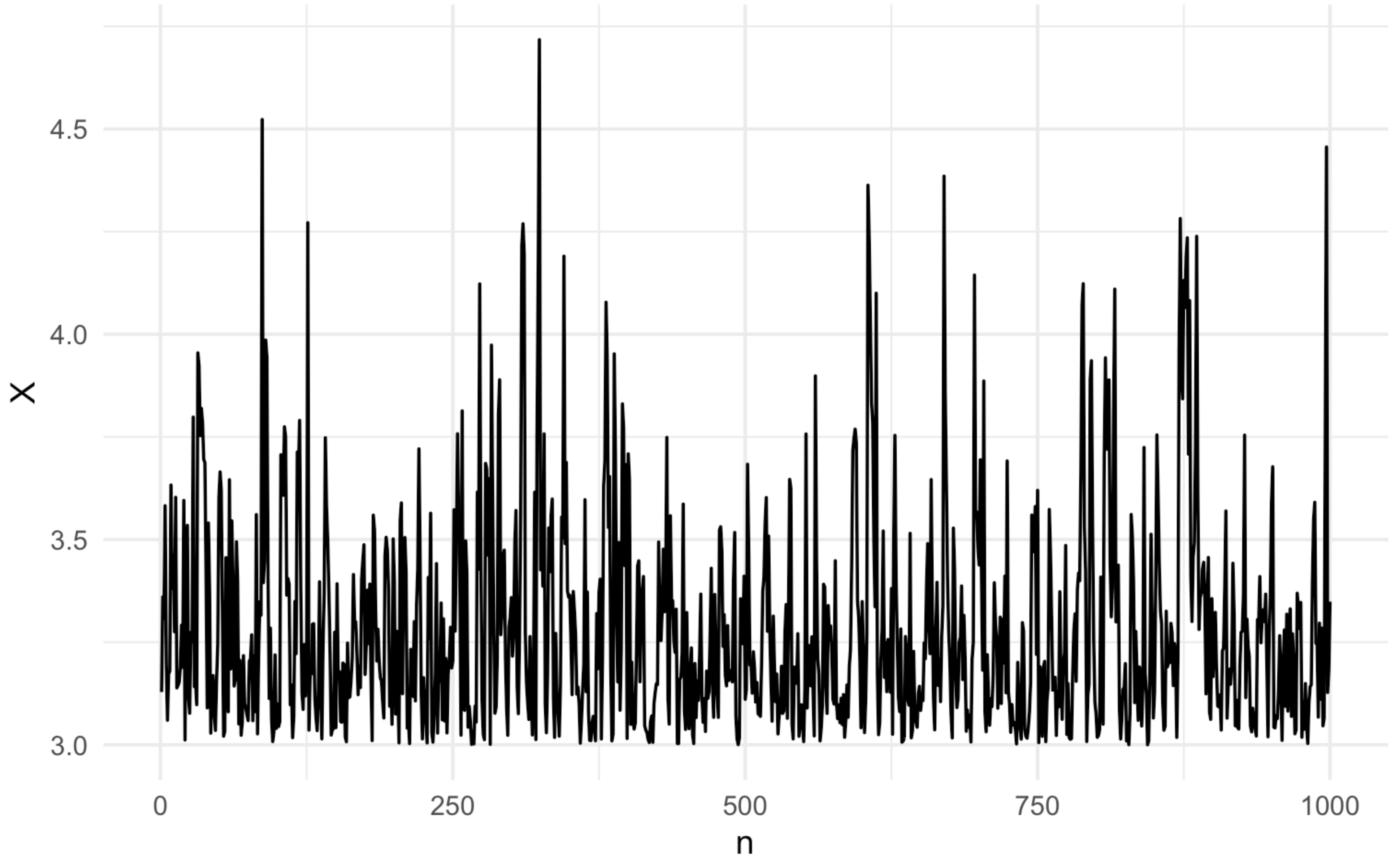
2. $z^{(t)} \mid x^{(t)} \sim \text{Unif}\left(0, \exp\left(-(x^{(t)} - \mu)^2/2\sigma^2\right)\right)$

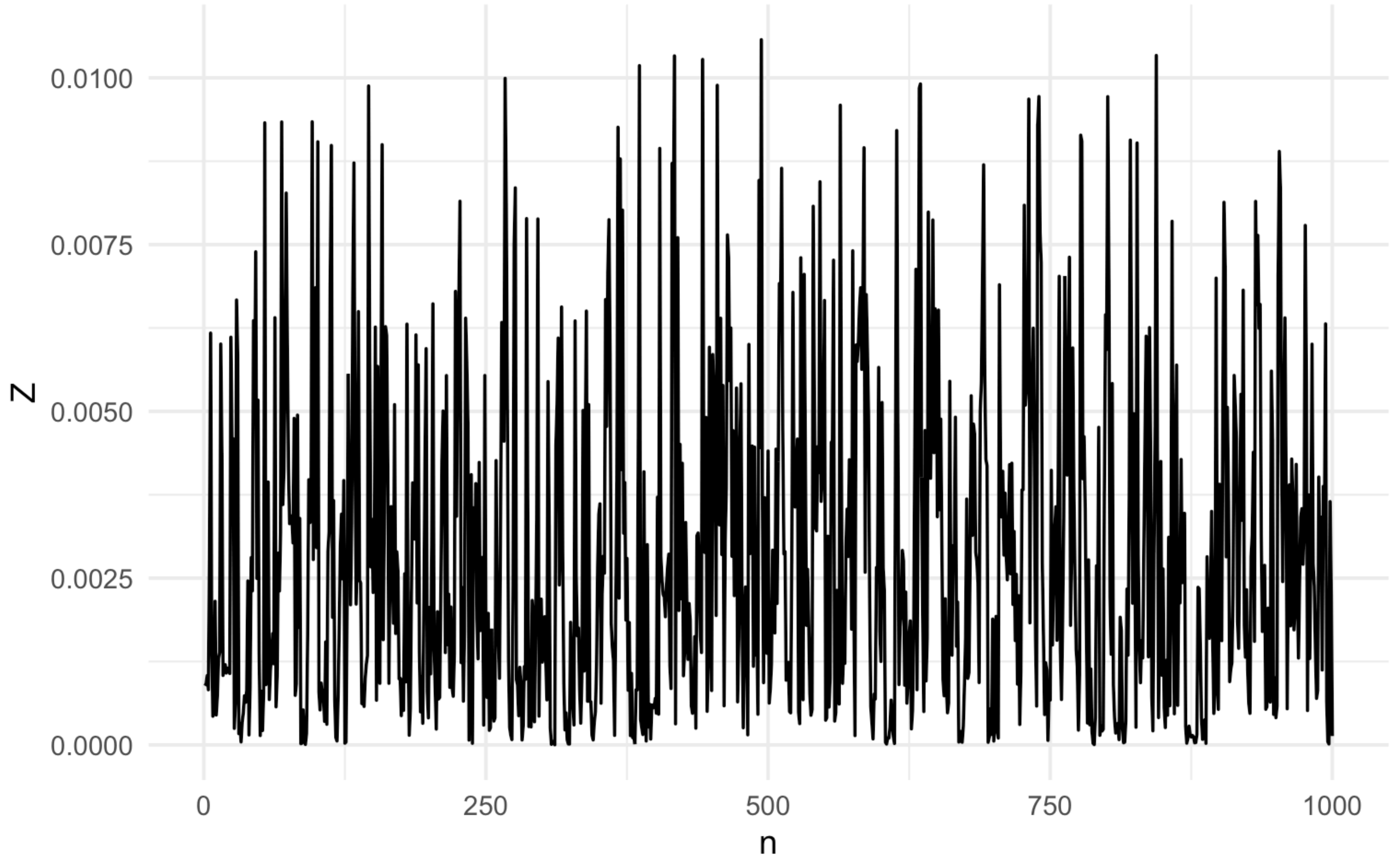
- Con $z^{(0)}$ elegido de tal forma que $\mu + \sqrt{-2\sigma^2 \log(z^{(0)})} > \mu^*$ lo que implica que

$$z^{(0)} < \exp\left(\frac{-(\mu^* - \mu)^2}{2\sigma^2}\right)$$

Ejemplo normal truncada

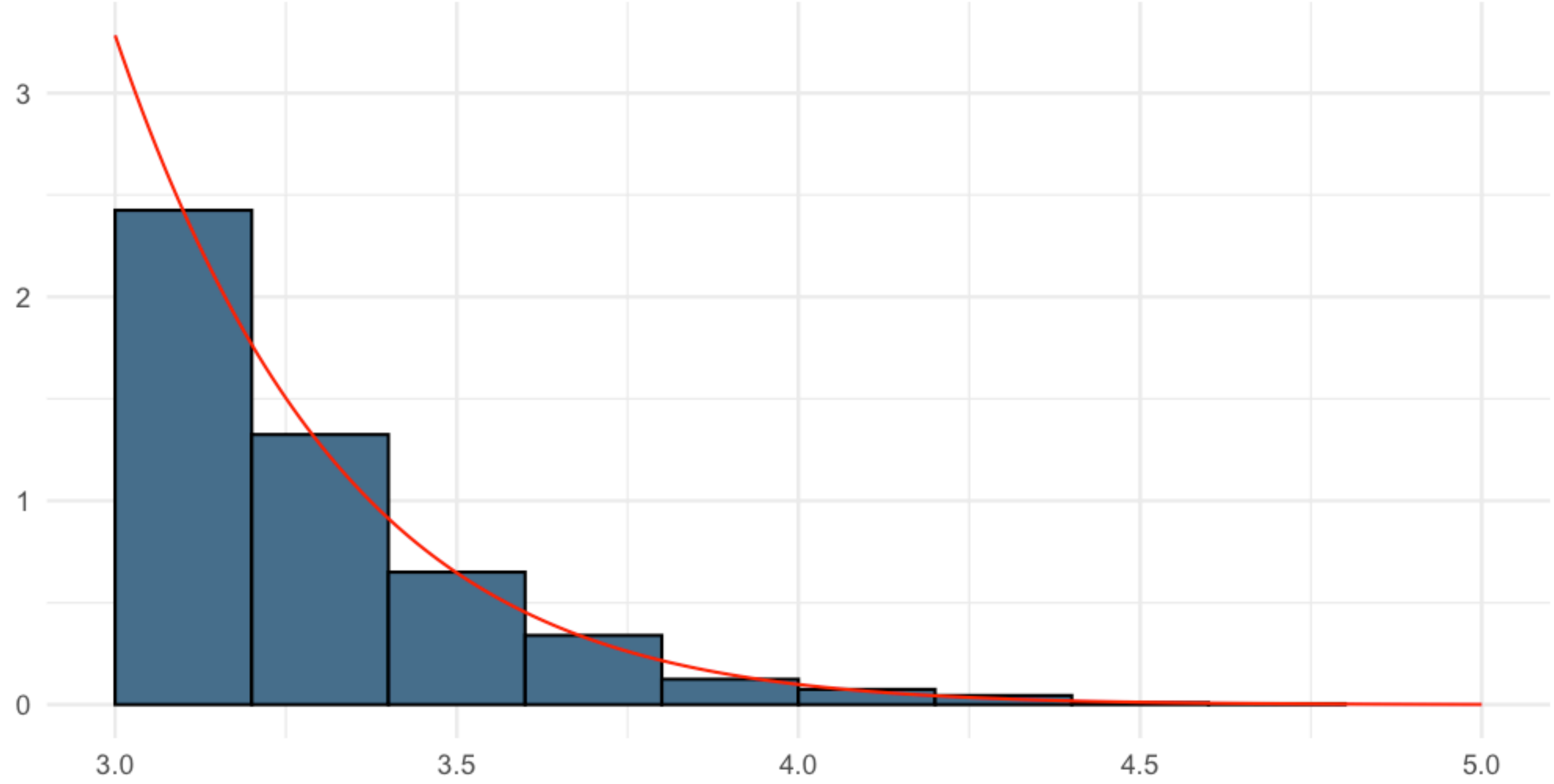
- Tomando $\mu = 0, \mu* = 3, \sigma^2 = 1$ y puntos iniciales $x^{(0)} = 0$ y $z^{(0)} = 0.0055$ e iterando 1000 veces el algoritmo





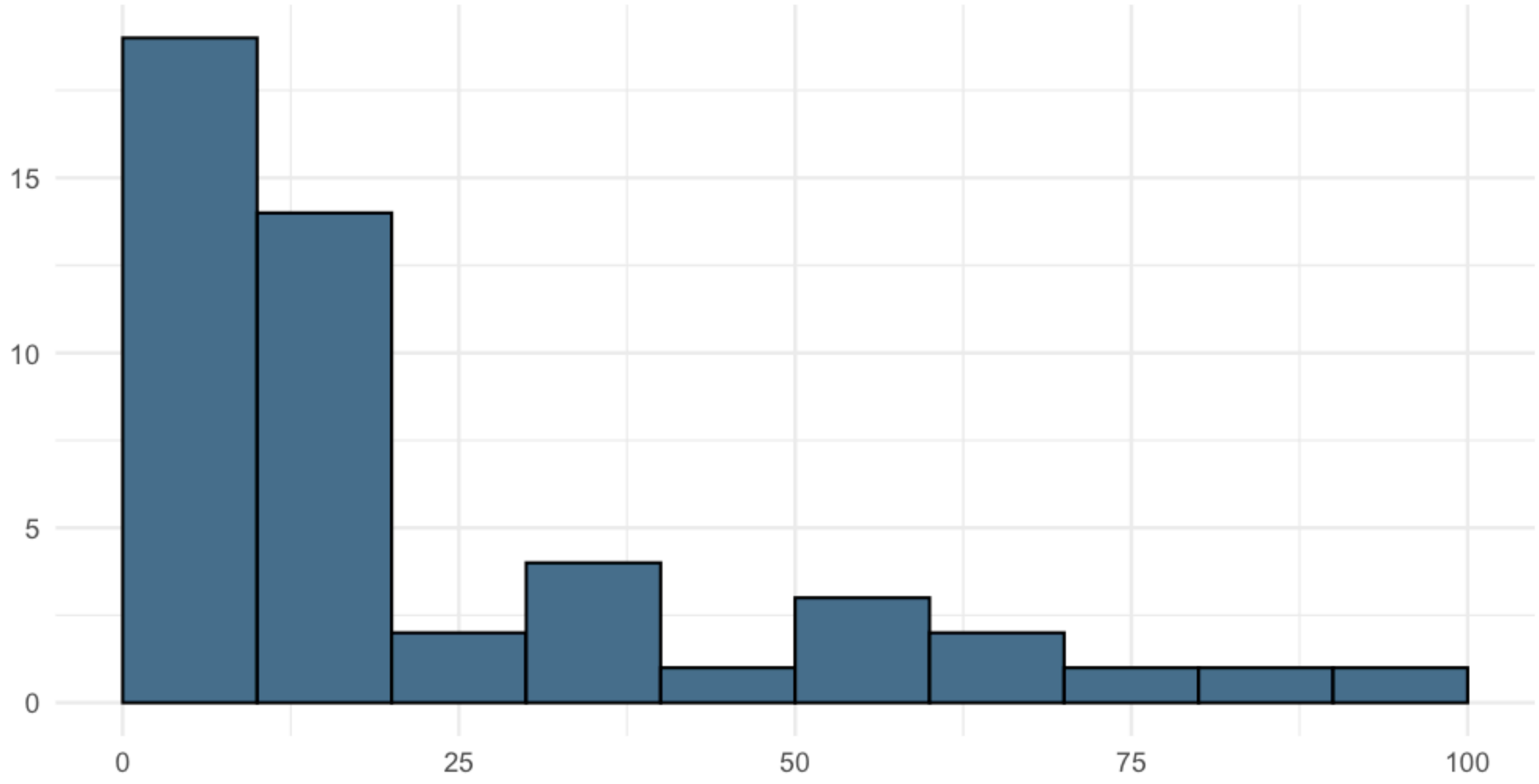
Ejemplo normal truncada

- Tomando $\mu = 0$, $\mu* = 3$, $\sigma^2 = 1$ y puntos iniciales $x^{(0)} = 0$ y $z^{(0)} = 0.0055$ e iterando 1000 veces el algoritmo



Ejemplo

- Se tienen 48 tiempos de supervivencia de pacientes con mieloma múltiple



- Tiene sesgo positivo y claramente no es normal

Ejemplo

- ¿Cómo se puede hacer para que los datos sean normales?

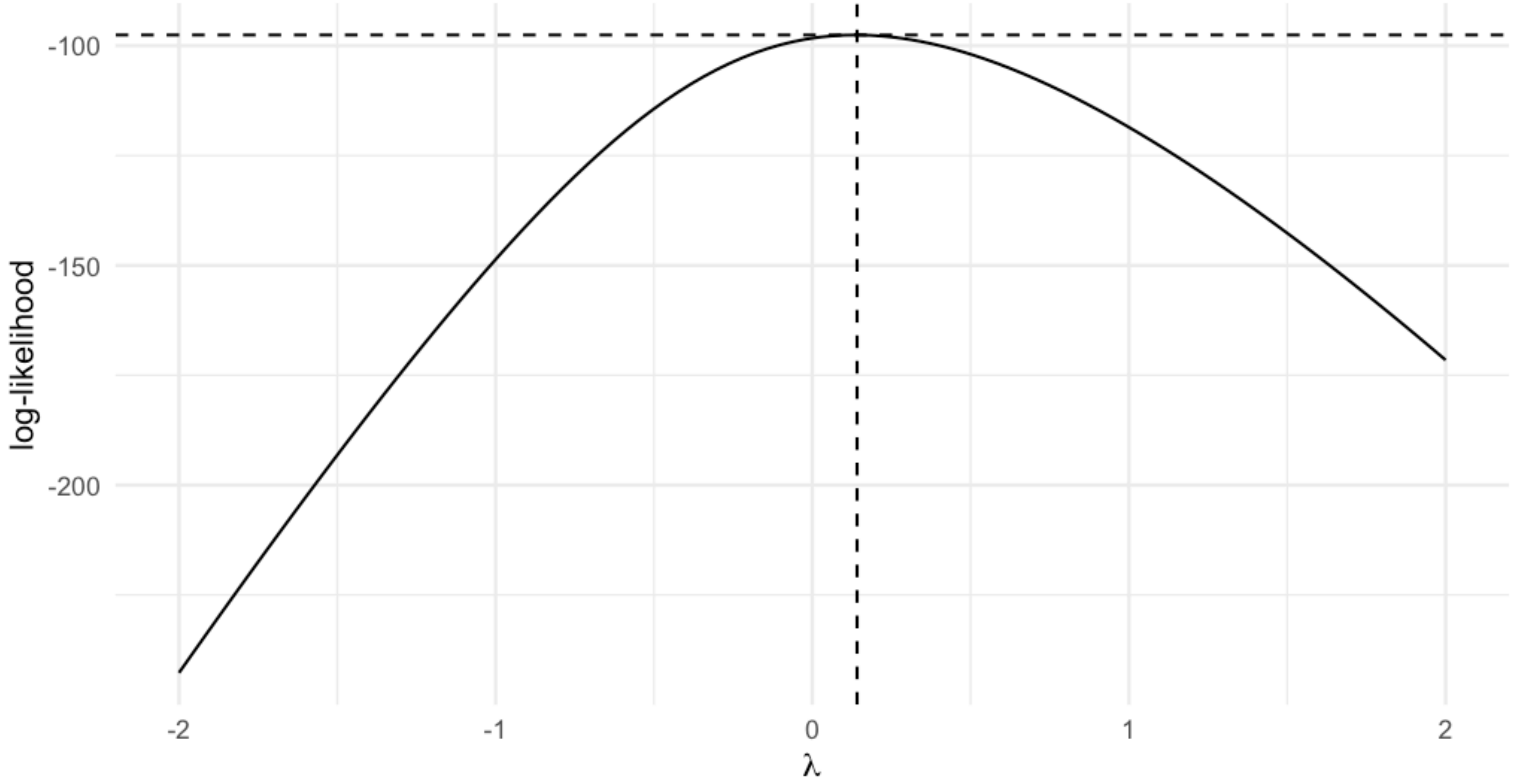
- Box y Cox propusieron la transformación

$$w_i = \begin{cases} \dfrac{y_i^{\lambda} - 1}{\lambda} & \text{si} \quad \lambda \neq 0 \\ \log(y_i) & \text{si} \quad \lambda = 0 \end{cases}$$

- Aplicaciones a modelos de regresión y a series de tiempo

Ejemplo

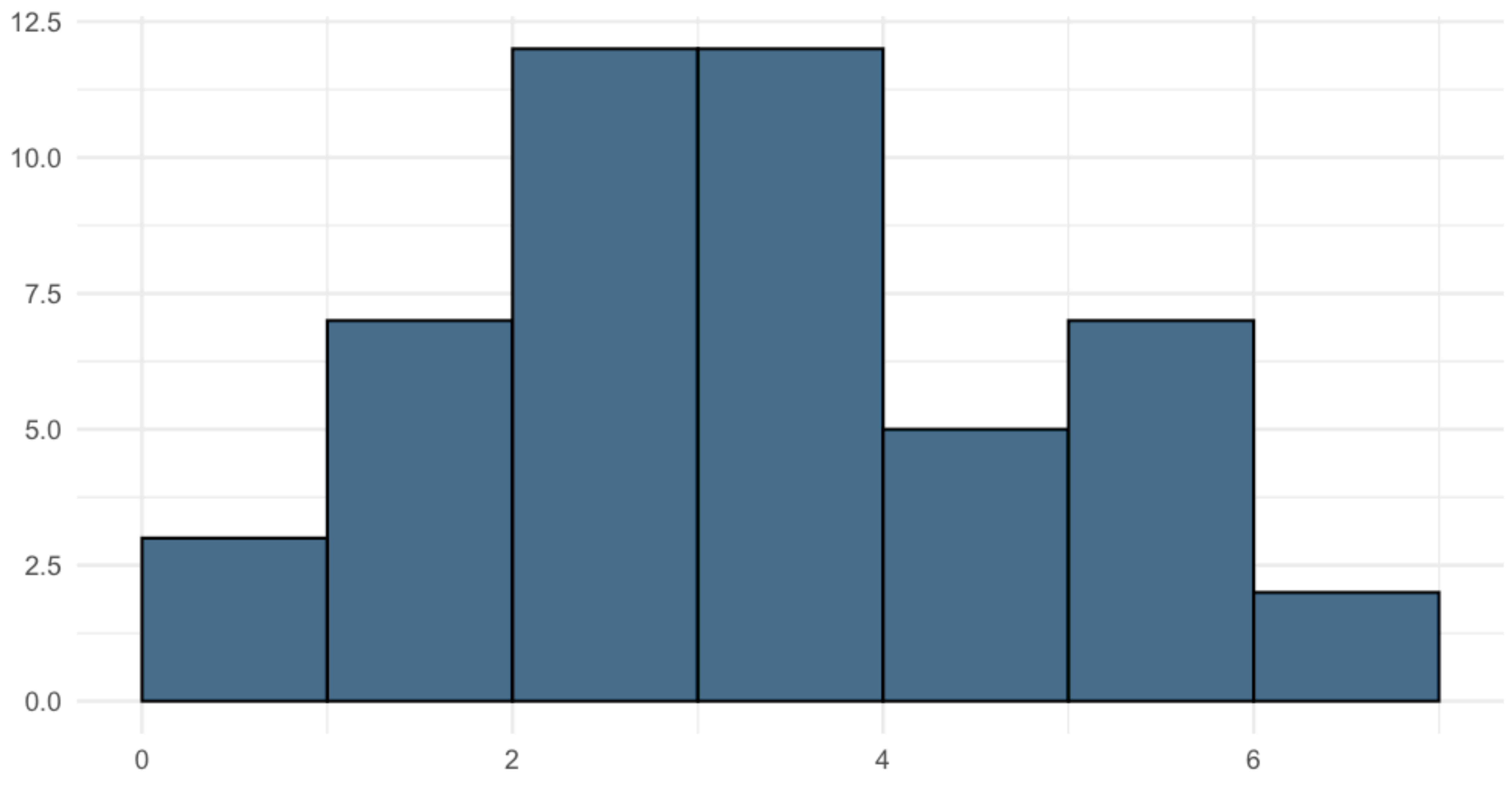
- En el enfoque clásico es otro parámetro a estimar mediante la maximización de la (log) verosimilitud perfilada, i.e., $\mathscr{L}(\lambda) = \mathscr{L}(\widehat{\mu}, \widehat{\sigma}^2, \lambda)$



- El estimador máximo verosímil es $\widehat{\lambda} = 0.1414$

Ejemplo

- Con esta $\widehat{\lambda}$



Prueba	Estadístico	p-valor
Anderson-Darling	0.47426	0.2308
Lilliefors	0.12077	0.07713
Shapiro-Wilk	0.96828	0.2169

- Para esta transformación $\widehat{\mu}_w = 3.2994$ y $\widehat{\sigma}_w = 1.6129$

Ejemplo

- Bajo un enfoque bayesiano se puede hacer inferencia sobre λ, μ, σ^2 conjuntamente

- Utilizando la distribución no-informativa

$$f(\lambda, \mu, \sigma^2) \propto \frac{1}{\sigma}$$

- La verosimilitud está dada por

$$f(\mathbf{y} \mid \lambda, \mu, \sigma^2) = \prod_{i=1}^{n} \phi\left(\frac{y_i^{\lambda} - 1}{\lambda} \mid \mu, \sigma^2\right) y_i^{\lambda-1}$$

Ejemplo

- La posterior es

$$f(\lambda, \mu, \sigma^2 \mid \mathbf{y}) \propto \left(\frac{1}{\sigma^2}\right)^{\left(\frac{n+1}{2}\right)} \prod_{i=1}^{n} y_i^{\lambda} \prod_{i=1}^{n} \exp\left(-\frac{(w_i(\lambda) - \mu)^2}{2\sigma^2}\right)$$

- Las condicionales de μ y σ^2 son conocidas

$$\mu \mid \sigma^2, \lambda, \mathbf{y} \sim \mathcal{N}\left(\bar{w}(\lambda), \frac{\sigma^2}{n}\right)$$

$$\sigma^2 \mid \mu, \lambda, \mathbf{y} \sim \text{InvGa}\left(\frac{n-1}{2}, \frac{\sum_i (w_i(\lambda) - \mu)^2}{2}\right)$$

Ejemplo

- El ‘problema’ es la condicional de λ

$$
\begin{aligned}
f(\lambda \mid \mu, \sigma^2 \mathbf{y}) &\propto \prod_{i=1}^{n} y_i^{\lambda} \prod_{i=1}^{n} \exp\left(-\frac{(w_i(\lambda) - \mu)^2}{2\sigma^2} \right) \\
&= \exp\left(\lambda \sum_i \log(y_i) \right) \prod_{i=1}^{n} \exp\left(-\frac{(w_i(\lambda) - \mu)^2}{2\sigma^2} \right)
\end{aligned}
$$

- No tiene una distribución cerrada
- Se puede usar un slice sampler o si es log-cóncava un ARS o ....
- Metropolis - Hastings

Ejemplo

- A grandes rasgos el algoritmo de Metropolis-Hastings genera una cadena de Markov de tal forma que al tiempo t

1. Se genera un valor $\lambda*$ de una distribución instrumental
2. Se acepta $\lambda^{(t)} = \lambda*$con probabilidad α
3. En caso contrario $\lambda^{(t)} = \lambda^{(t-1)}$

- Los detalles se verán más adelante

Ejemplo

- El algoritmo queda definido como:

1. Definir valores iniciales $\mu^{(0)}, \sigma^{2(0)}, \lambda^{(0)}$ e hiperparámetros requeridos

2. Al tiempo t:

- Simular $\mu^{(t)} \sim \mathcal{N}\left(\bar{w}(\lambda^{(t)}), \frac{\sigma^{2(t-1)}}{n}\right)$

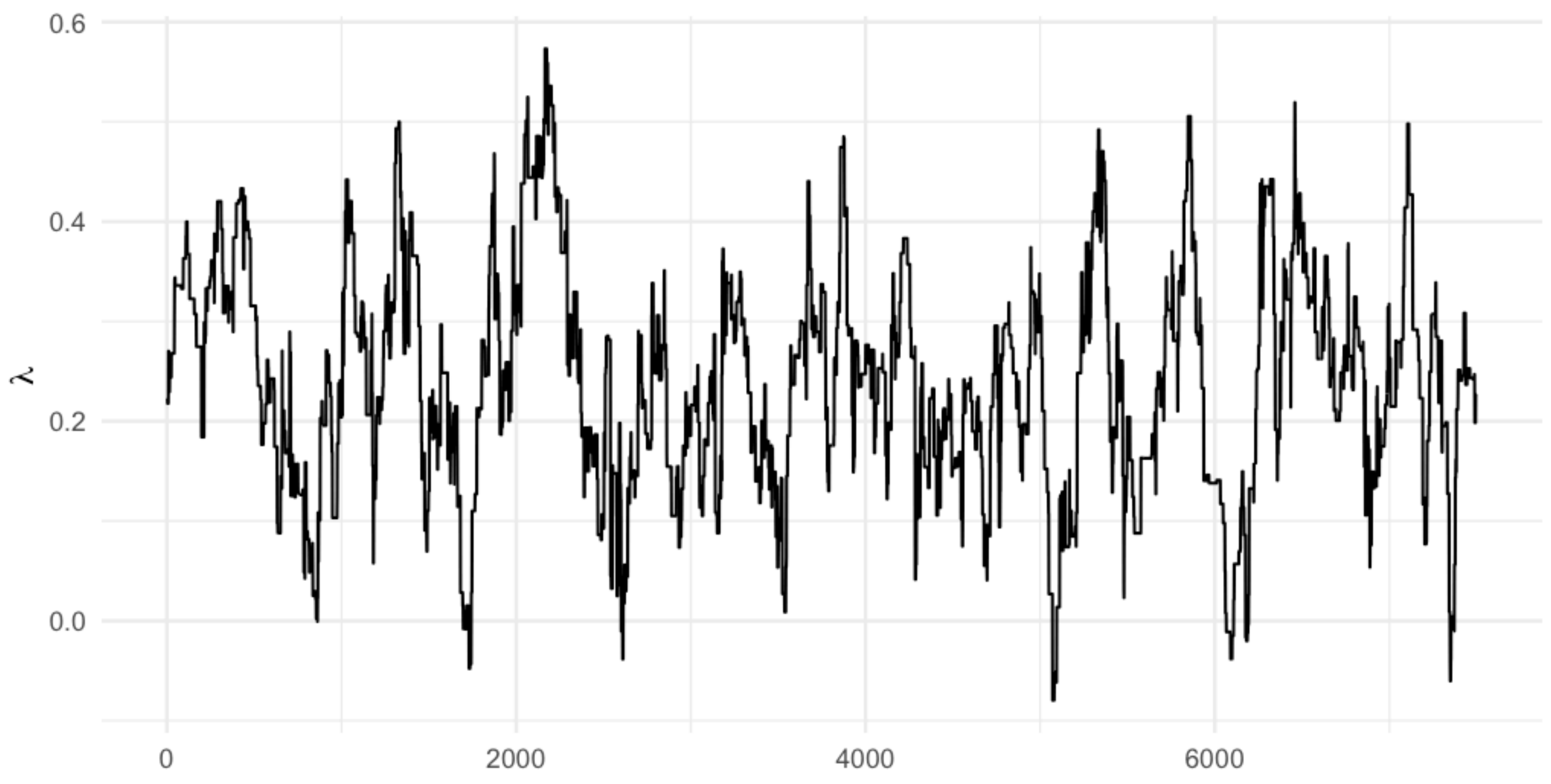
- Simular $\sigma^{2(t)} \sim \text{InvGa}\left(\frac{n-1}{2}, \frac{\sum_i (w_i(\lambda^{(t-1)}) - \mu^{(t)})^2}{2}\right)$

- Simular $\lambda^{(t)}$ utilizando un Metropolis-Hastings

3. Repetir N veces

Ejemplo

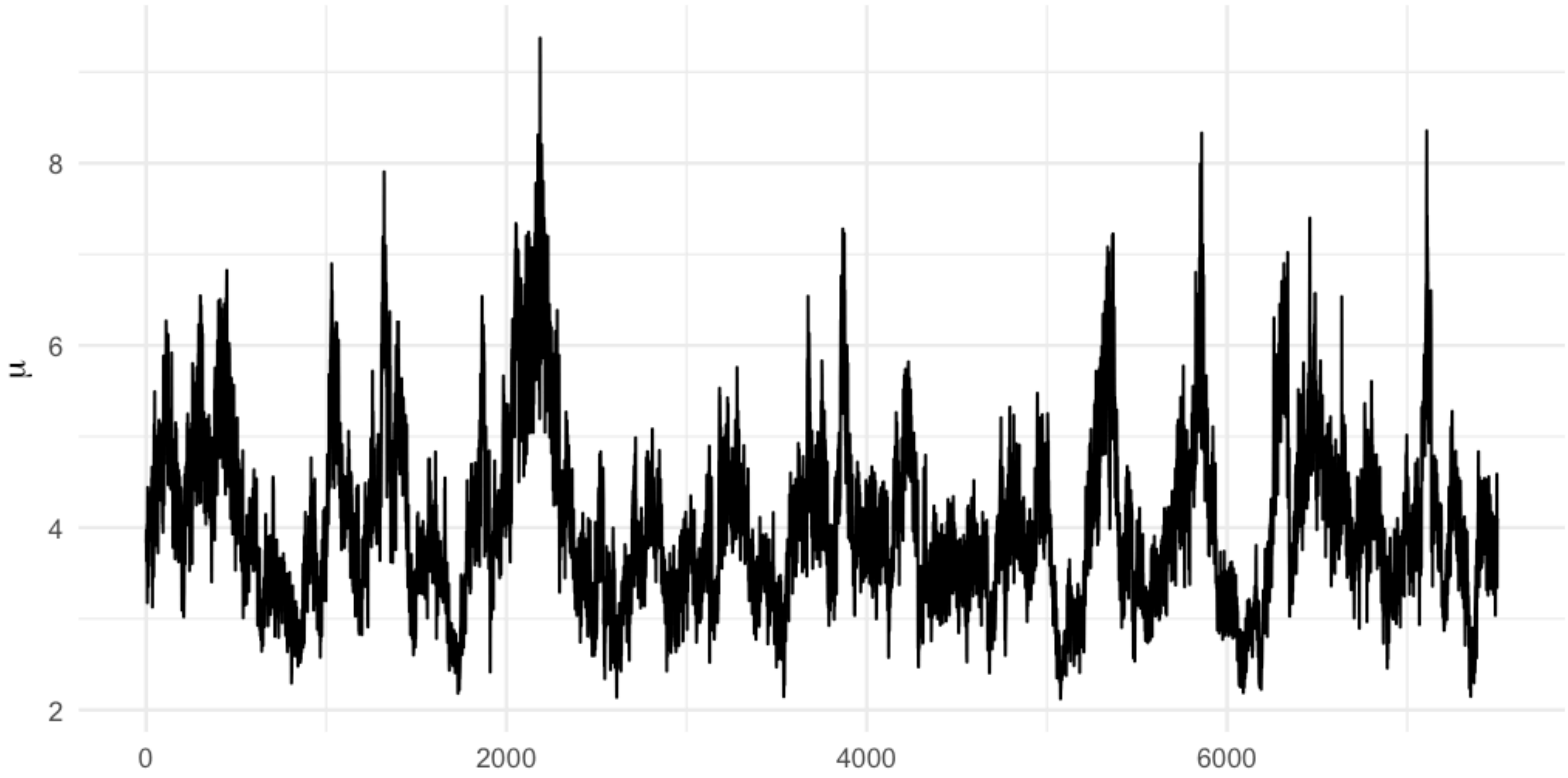
- Para λ



Min.	Q1	Mediana	Media	Q3	Max.
-0.07959	0.16291	0.24326	0.24034	0.30959	0.57317

Ejemplo

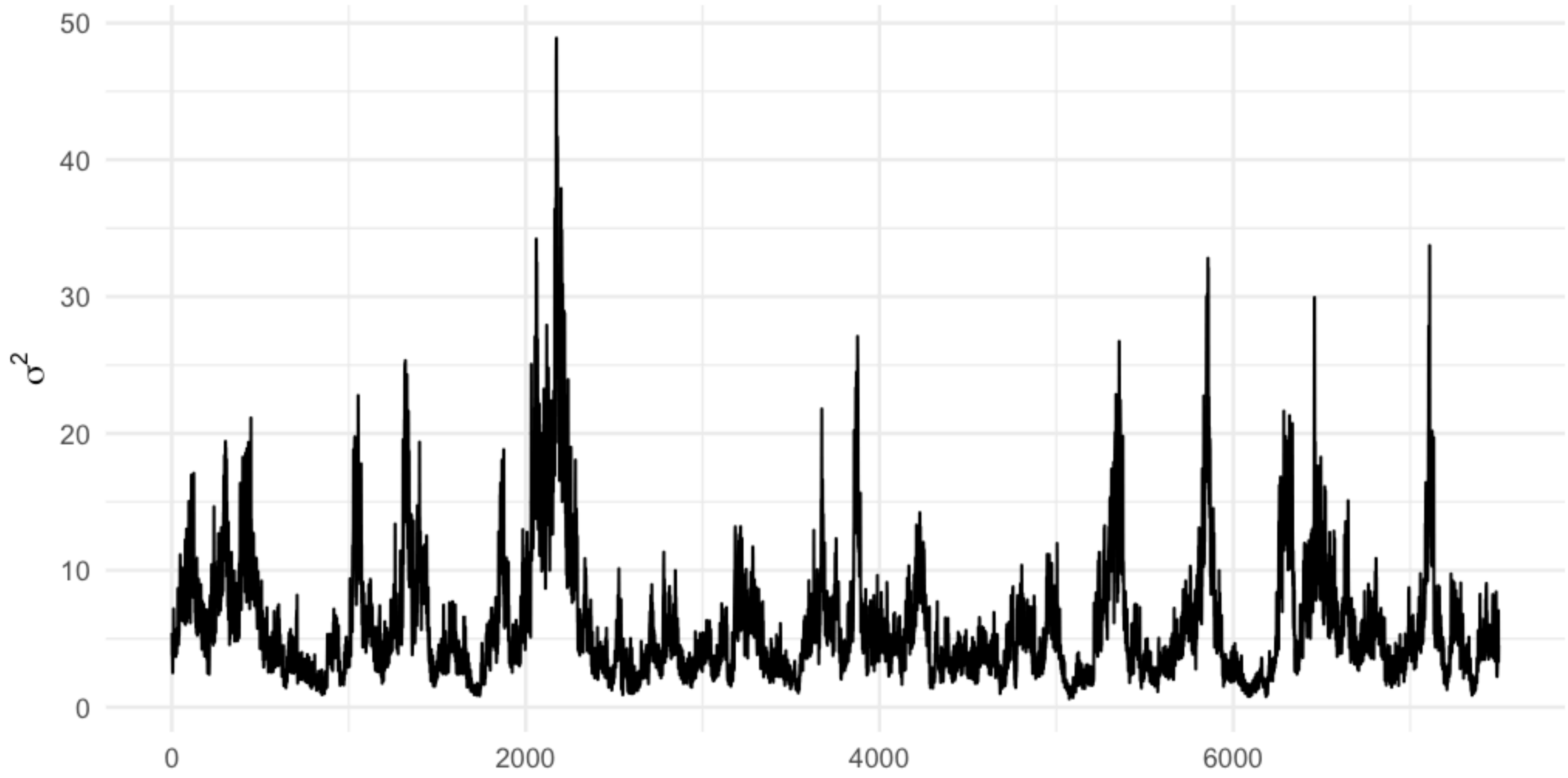
- Para μ



Min.	Q1	Mediana	Media	Q3	Max.
2.121	3.392	3.908	4.051	4.530	9.372

Ejemplo

- Para σ^2



Min.	Q1	Mediana	Media	Q3	Max.
0.5983	3.0050	4.6209	5.9167	7.1654	48.8956