

# Parcial1-2023

Javier Orín

2025-12-08

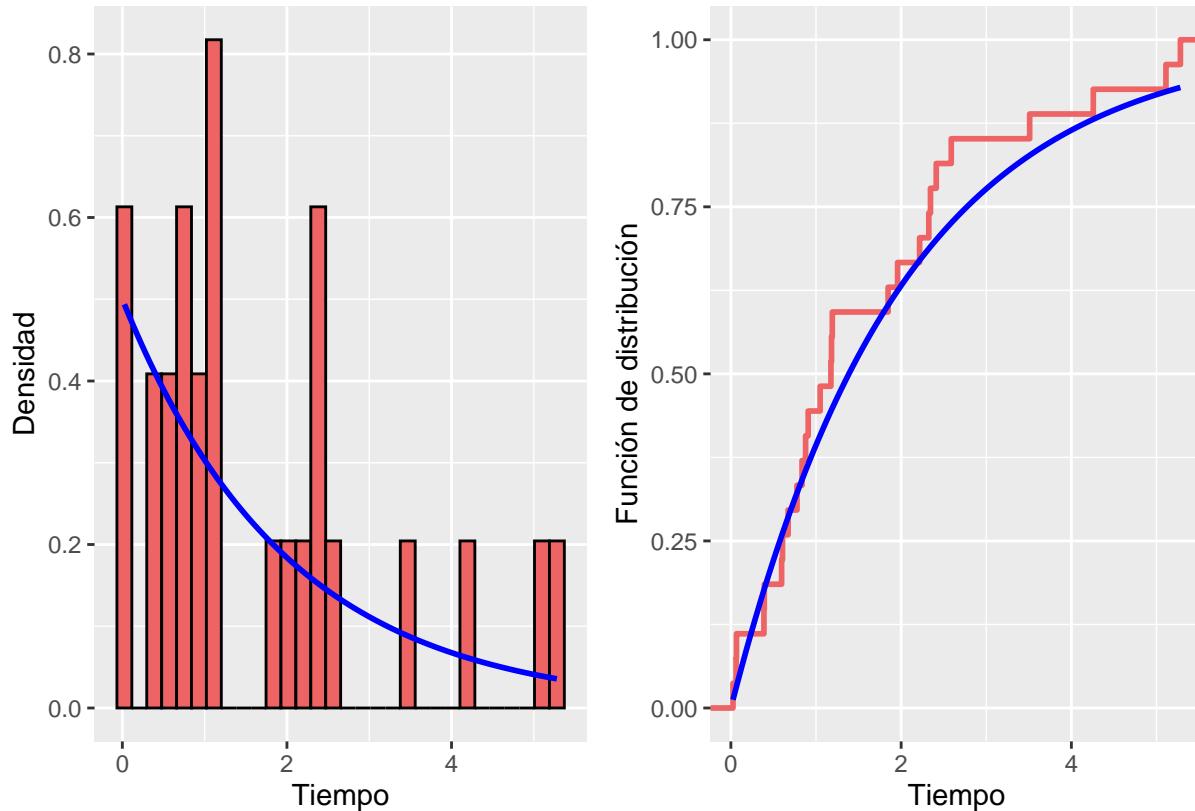
## Problema 2

1) (0,8 puntos) Basándote en los datos recogidos, estima puntualmente la varianza poblacional y estudia desde un punto de vista descriptivo si la distribución del tiempo ocurrido entre dos terremotos consecutivos sigue aproximadamente una distribución exponencial con varianza 4.

Si  $X \sim \varepsilon(\lambda)$ , entonces  $E(X) = \frac{1}{\lambda}$  y  $Var(X) = \frac{1}{\lambda^2}$ . Despejando,

$$Var(X) = 4 = \frac{1}{\lambda^2} \Rightarrow \lambda = \sqrt{\frac{1}{4}} = \frac{1}{2}$$

Cargamos el conjunto de datos y graficamos la distribución que menciona el enunciado junto con la densidad empírica:



Estimación muestral de la varianza:

```
var(Richter)
## [1] 2.148606
```

No se asemeja a la dada por el enunciado (4).

Simulemos muchas muestras  $\varepsilon(\lambda = 0.5)$  del mismo que tamaño que la del problema y veamos qué resultados hay:

```
nr = 1e4
n=length(Richter)
results=data.frame()
for (rep in 1:nr) {
  muestra = rexp(n=n, rate=0.5)
  results[rep, "ECM"] = mean((muestra-Richter)^2)
  results[rep, "err_abs"] = mean(abs(muestra-Richter))
  results[rep, "sesgo"] = mean(muestra-Richter)
}
summary(results)

##      ECM          err_abs        sesgo
##  Min.   : 1.409   Min.   :0.762   Min.   :-0.80202
##  1st Qu.: 4.361   1st Qu.:1.563   1st Qu.: 0.07854
##  Median : 5.669   Median :1.756   Median : 0.31834
##  Mean   : 6.185   Mean   :1.783   Mean   : 0.34169
##  3rd Qu.: 7.394   3rd Qu.:1.978   3rd Qu.: 0.57758
##  Max.   :30.593   Max.   :3.596   Max.   : 2.87263
```

El criterio del EMC es muestra de que los datos no siguen la distribución que indica el enunciado.

**2. (0,8 puntos)** Suponiendo que el tiempo sigue una distribución exponencial, determina de manera aproximada el tamaño de muestra necesario si el objetivo es estimar la varianza poblacional mediante el estimador  $T = (\bar{X}_n)^2$  con un error relativo máximo de estimación de 0,2 y confianza de 0,90. Se recuerda que el error relativo de estimación es igual al error absoluto de estimación entre el valor real a estimar.

Por definición del error relativo, buscamos encontrar un tamaño muestral  $n$  que cumpla:

$$P \left( \frac{|(\bar{X}_n)^2 - \frac{1}{\lambda^2}|}{(1/\lambda^2)} < 0.2 \right) \geq 0.9$$

Al considerar  $(X_1, \dots, X_n)$  m.a.s. de  $X \sim \varepsilon(\lambda)$ , aplicando la reproductividad y las propiedades de la distribución exponencial/gamma:

$$\frac{1}{n} \sum_{i=1}^n X_i \sim \frac{1}{n} \gamma(n, \lambda) \equiv \gamma(n, n\lambda)$$

De este modo,  $E(\bar{X}_n) = \frac{1}{\lambda}$ ,  $Var(\bar{X}_n) = \frac{1}{n\lambda^2}$ . Buscamos estimar el parámetro  $\frac{1}{\lambda^2}$  a partir del que estima la media muestral  $(\frac{1}{\lambda})$ . Podemos combinar el TLC con el método Delta:

$$\bar{X}_n \xrightarrow{L,n>>} N \left( \frac{1}{\lambda}, \frac{1}{\lambda\sqrt{n}} \right) \quad (\bar{X}_n)^2 \xrightarrow{L,n>>} N \left( \frac{1}{\lambda^2}, \frac{2}{\lambda^2\sqrt{n}} \right)$$

Hemos aplicado el método Delta para  $g(\theta) = \theta^2$ ,  $g'(\theta) = 2\theta$ .

Podemos tipificar la normal y hallar los cuartiles para  $N(0, 1)$ :

$$\frac{(\bar{X}_n)^2 - \frac{1}{\lambda^2}}{\left(\frac{2}{\lambda^2 n}\right)} \stackrel{n \text{ grande}}{\sim} N(0, 1)$$

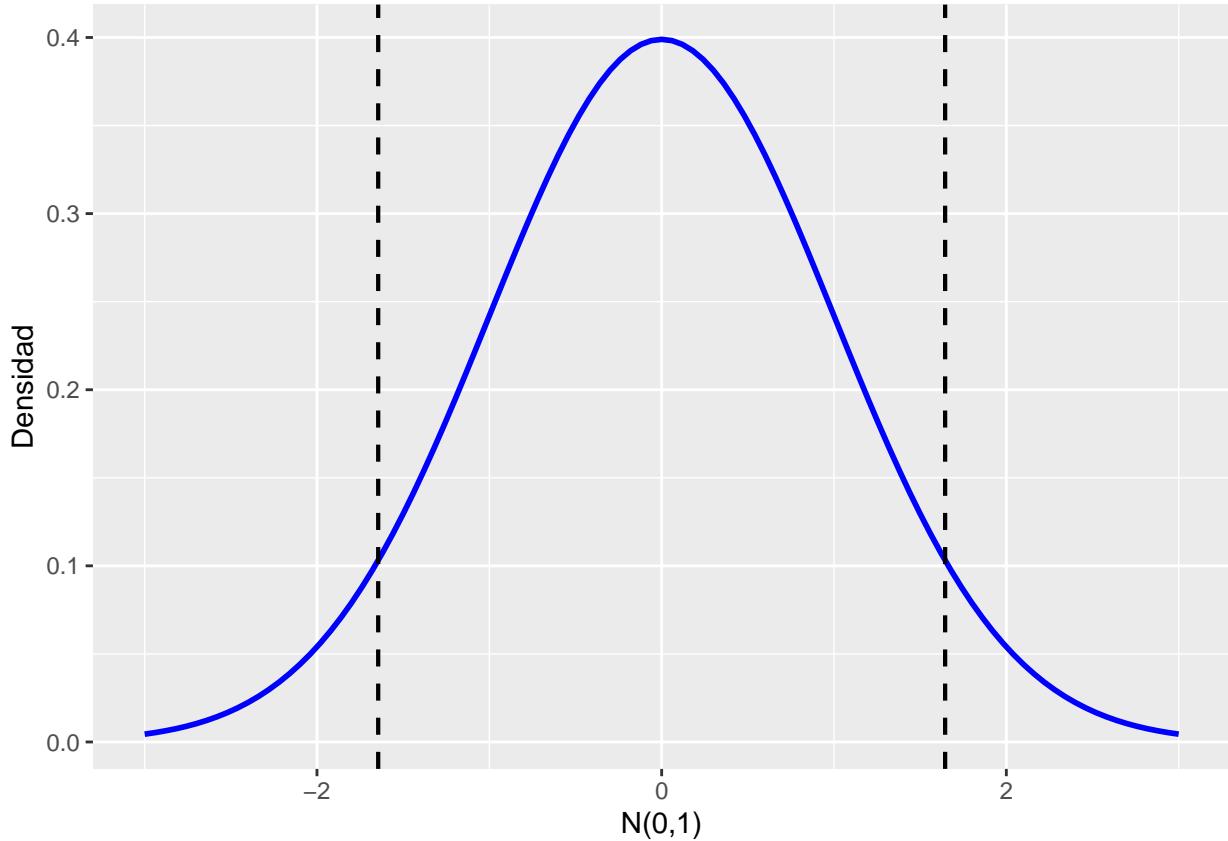
Despejamos  $\alpha$  para hallar los cuartiles deseados:

$$1 - \alpha = 0.9 \Rightarrow \alpha = 0.1 \quad \frac{\alpha}{2} = 0.05 \quad 1 - \frac{\alpha}{2} = 0.95$$

```
alpha = 0.1
(c1 = qnorm(p = alpha/2))

## [1] -1.644854
(c2 = qnorm(p=1-alpha/2))

## [1] 1.644854
```



Las líneas verticales muestran los cuantiles  $c_1$  y  $c_2$ . Despejamos el tamaño muestral del pivote:

$$\frac{|(\bar{X}_n)^2 - \frac{1}{\lambda^2}|}{\frac{1}{\lambda^2} \cdot \frac{2}{\sqrt{n}}} < c_2 \iff \frac{|(\bar{X}_n)^2 - \frac{1}{\lambda^2}|}{\frac{1}{\lambda^2}} < \frac{2 \cdot c_2}{\sqrt{n}} = \underbrace{0.2}_{\text{Error relativo}}$$

Despejamos para obtener el tamaño muestral:

$$n = \left( \frac{2 * c_2}{0.2} \right)^2$$

```
(n = (2*c2/0.2)^2)
```

```
## [1] 270.5543
```

Al considerar  $n = 271$  obtenemos un tamaño muestral  $n \geq 30$ , apto para el TLC.

3. (0,8 puntos) Suponiendo que el tiempo sigue una distribución exponencial, calcula un intervalo de confianza al nivel de confianza 0,90 para la varianza poblacional haciendo uso del método de la función pivotal.

Usaremos la función pivot  $T = \lambda \bar{X}_n$ , de distribución conocida:

$$\bar{X}_n \sim \gamma(n, \lambda n) \Rightarrow T = \lambda \bar{X}_n \sim \gamma(n, n)$$

En particular, nuestra muestra Richter tiene el siguiente tamaño:

```
(n = length(Richter))
```

```
## [1] 27
```

Calculamos los cuantiles correspondientes:

```
(c1 = qgamma(p=alpha/2, shape = n, rate = n))
```

```
## [1] 0.7058559
```

```
(c2 = qgamma(p=1-alpha/2, shape = n, rate = n))
```

```
## [1] 1.336171
```

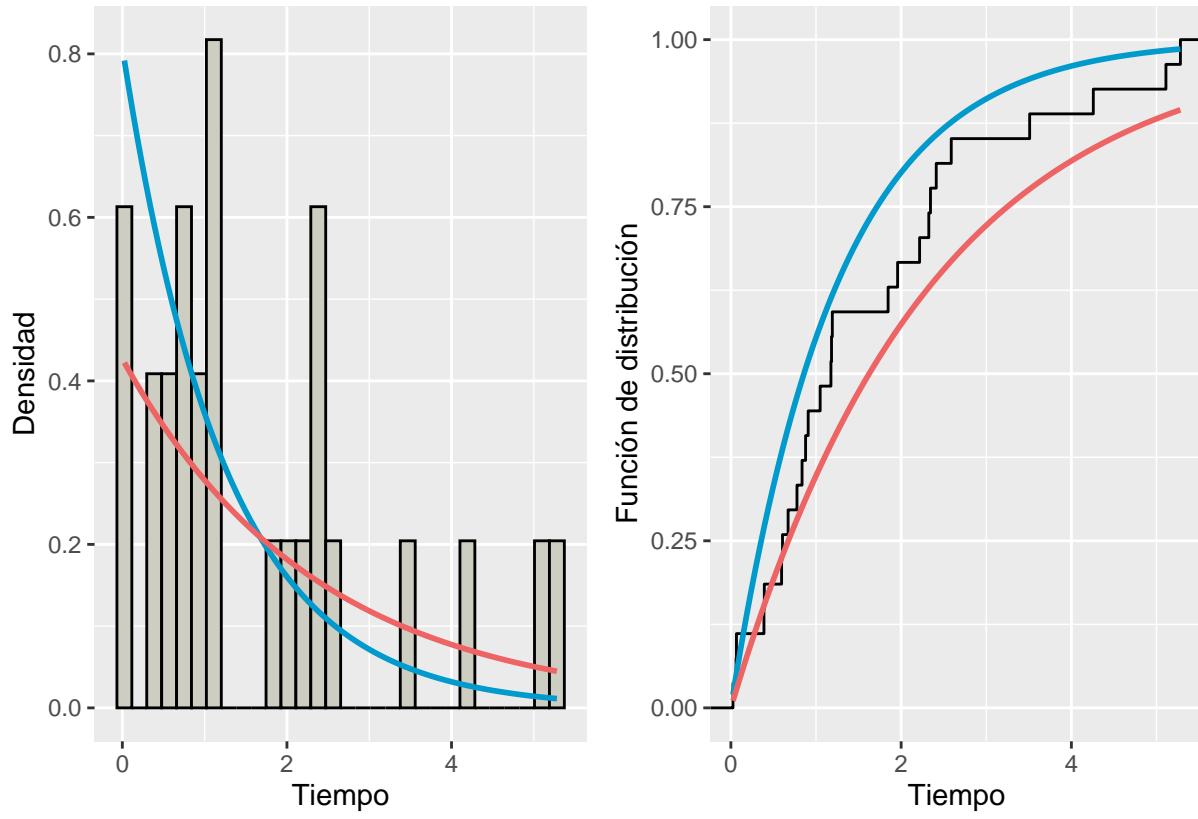
Deshacemos el pivote. En este caso, el error no es relativo:

$$c_1 < \lambda \bar{X}_n < c_2 \iff \frac{c_1}{\bar{X}_n} < \lambda < \frac{c_2}{\bar{X}_n} \iff \left( \frac{\bar{X}_n}{c_2} \right)^2 < \frac{1}{\lambda^2} < \left( \frac{\bar{X}_n}{c_1} \right)^2$$

```
media = mean(Richter)
des_pivote = function(q) (media/q)^2
(intervalo=tibble(i1=des_pivote(c2), i2=des_pivote(c1)))
```

```
## # A tibble: 1 x 2
##       i1     i2
##   <dbl> <dbl>
## 1 1.53  5.49
```

Vamos a graficar las distribuciones exponenciales cuyas varianzas son los extremos en intervalo de confianza que acabamos de calcular:



Otro posible pivote hubiera sido, por ejemplo,

$$2\lambda \sum_{i=1}^n X_i \sim \gamma \left( n, \frac{1}{2} \right) \equiv \chi_{2n}^2$$

En ambos casos se obtiene el mismo intervalo de confianza.

4. (0,8 puntos) Calcula un intervalo de confianza bootstrap para la varianza poblacional al nivel de confianza 0,90, y estima mediante simulación la cobertura de dicho intervalo para el actual tamaño muestral. Puedes suponer que el verdadero valor del parámetro es su estimación máximo-verosímil.

Hallemos el EMV de  $\lambda$  en primer lugar:

$$\log L(\lambda; \vec{x}_n) = n \log(\lambda) - \lambda \sum_{i=1}^n x_i \Rightarrow \frac{\partial}{\partial \lambda} \log L(\lambda; \vec{x}_n) = \frac{n}{\lambda} - \sum_{i=1}^n x_i \Rightarrow \lambda_{MV} = (\bar{X}_n)^{-1}$$

Es en efecto un máximo, pues:

$$\frac{\partial^2}{\partial \lambda^2} \log L(\lambda; \vec{x}_n) = \frac{-n}{\lambda^2} < 0$$

Por la equivarianza de los estimadores maximo-verosímiles, tenemos que  $g(\bar{X}_n^{-1})$  es un EMV de  $g(\lambda)$  para cualquier función definida en el espacio paramétrico. En particular, para  $g(\theta) = \theta^{-2}$ ,  $(\bar{X}_n)^2$  es un EMV de  $\frac{1}{\lambda^2}$ .

Pasemos ahora a la simulación bootstrap.