

# Examen 1 2022

Javier Ortín

2025-12-12

## Cuestiones teóricas

**3 Sea  $X = U(0, \theta)$  con  $\theta$  desconocido. Se extrae una muestra aleatoria simple de tamaño  $n$ . Calcula la distribución asintótica de la función pivote  $n \left(1 - \frac{X_{(n)}}{\theta}\right)$  cuando  $n \rightarrow \infty$ .**

Aun siendo desconocida,  $\theta$  no deja de ser una constante positiva. Por tanto, tenemos:

$$\frac{X_{(n)}}{\theta} = \frac{\max_{1 \leq i \leq n} X_i}{\theta} = \max_{1 \leq i \leq n} \frac{X_i}{\theta} = \left(\frac{X}{\theta}\right)_{(n)}$$

Aplicando la transformación  $Y_i = \frac{1}{\theta} X_i$ , tenemos que la transformación es una m.a.s. de variables  $U(0, 1)$ . La función de distribución de su máximo vendrá dada por:

$$F_{Y_{(n)}}(x) = \begin{cases} 0 & \text{si } x < 0 \\ x^n & \text{si } x \in (0, 1) \\ 1 & \text{si } x \geq 1 \end{cases}$$

Aplicamos una transformación más. Sea  $T_n := 1 - Y_{(n)}$ ,

$$P(T_n \leq x) = P(1 - Y_{(n)} \leq x) = P(1 - x \leq Y_{(n)}) = 1 - F_{Y_{(n)}}(1 - x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - (1 - x)^n & \text{si } x \in (0, 1) \\ 1 & \text{si } x \geq 1 \end{cases}$$

Finalmente, aplicando la última transformación,  $P(nT_n \leq x) = P(T_n \leq \frac{x}{n})$ . Así,

$$F_n(x) := F_{nT_n}(x) = \begin{cases} 0 & \text{si } x < 0 \\ \left(1 - \frac{x}{n}\right)^n & \text{si } x \in (0, n) \\ 1 & \text{si } x \geq n \end{cases}$$

Veamos ahora cómo es la convergencia de la función de distribución. Si  $x \leq 0$ ,  $F_n(x) = 0 \xrightarrow{n \rightarrow \infty} 0$ . Si  $x > 0$ , por la propiedad arquimediana,  $\exists n_0 \in \mathbb{N}$  tal que  $x \in (0, n) \forall n \geq n_0$ . Por tanto,

$$\lim_{n \rightarrow \infty} F_n(x) = \lim_{n \rightarrow \infty} 1 - \left(1 - \frac{x}{n}\right)^n = 1 - e^{-x}$$

Se corresponde con la función de distribución de una exponencial de parámetro 1. Veamos una simulación del resultado para ver que estamos en lo cierto:

```

theta = runif(n=1,min=1,max=100)
n = 100
nr = 1e3
dist_pivote = replicate(nr, {
  muestra = runif(n=n, min=0, max=theta)
  n*(1-max(muestra)/theta)
})

```

Se puede apreciar una clara similitud entre la densidad empírica y la teórica, aun siendo un resultado asintótico.

Otra posible opción para resolver este ejercicio es simplemente observar que:

$$1 - \frac{X_{(n)}}{\theta} = \left(1 - \frac{X}{\theta}\right)_{(n)} \quad \text{con} \quad 1 - \frac{X}{\theta} \sim U(0, 1)$$

## Problemas prácticos

1) (4 puntos) Sea  $X$  una variable aleatoria  $\gamma(p, a)$  de la que se obtiene la siguiente muestra aleatoria simple  $\vec{x}$  de tamaño 20:

c(1.3, 1.2, 3.5, 4, 2.8, 5.1, 1.5, 2.3, 3.8, 4.3, 3, 2, 0.9, 2.5, 1.7, 1.6, 1.1, 1.8, 1.4, 1)

a) Halla un estadístico suficiente minimal para  $\theta = (p, a)$ .

El soporte de una distribución gamma es  $(0, \infty)$ , que no depende de sus parámetros. Por tanto, si logramos factorizar su función de densidad adecuadamente, podríamos demostrar que forma parte de la familia exponencial.

$$f_X(\vec{x}_n; p, a) = \prod_{i=1}^n \frac{a^p}{\Gamma(p)} x_i^{p-1} \cdot \exp(-ax_i) \cdot I_{(0,+\infty)}(x_i)$$

Como  $\theta$  es un parámetro bidimensional, buscamos factorizar de la siguiente forma (funciones definidas sobre el soporte de la variable gamma):

$$f_X(\vec{x}_n; a, p) = h(\vec{x}_n) \cdot g(\theta) \cdot \exp\left(\sum_{k=1}^2 T_k(\vec{x}_n) \cdot Q_k(\theta_k)\right)$$

Basta considerar las siguientes funciones: ( $h(\vec{x}_n) = 1$ )

$$\underbrace{\left(\frac{a^p}{\Gamma(p)}\right)^n}_{g(\theta)} \cdot \exp\left(\underbrace{(p-1) \sum_{i=1}^n \log(x_i)}_{Q_1(\theta) \cdot T_1(\vec{x}_n)} - \underbrace{a \sum_{i=1}^n x_i}_{Q_2(\theta) \cdot T_2(\vec{x}_n)}\right)$$

Tal y como se demostró en teoría, al contener el espacio paramétrico natural intervalos reales, el siguiente estadístico es minimalmente suficiente:

$$T(\vec{x}_n) = \left(\sum_{i=1}^n \log(x_i), \sum_{i=1}^n x_i\right)$$

b) Halla estimadores de  $p$  y de  $a$  por el método de los momentos. Aplícalos a la muestra.

Por seguir  $X$  una distribución  $\gamma(p, a)$ , sabemos:

$$E[X] = \frac{p}{a}$$

$$Var[X] = \frac{p}{a^2} = E[X^2] - E[X]^2$$

$$E[X]^2 \cdot Var[X]^{-1} = \frac{p^2}{a^2} \cdot a^2 p = p$$

$$E[X] \cdot Var[X]^{-1} = \frac{p}{a} \cdot a^2 p = a$$

Aplicamos el estimador a la muestra:

```
observaciones = c(1.3, 1.2, 3.5, 4, 2.8, 5.1, 1.5, 2.3, 3.8, 4.3, 3,
                  2, 0.9, 2.5, 1.7, 1.6, 1.1, 1.8, 1.4, 1)
e_mom_gamma = function (muestra) {
  p_mom = mean(muestra)^2/var(muestra)
  a_mom = mean(muestra)/var(muestra)
  tibble(p=p_mom, a = a_mom)
}
(p_a_mom = e_mom_gamma(observaciones))
```

```
## # A tibble: 1 x 2
##       p       a
##   <dbl> <dbl>
## 1  3.57  1.53
```

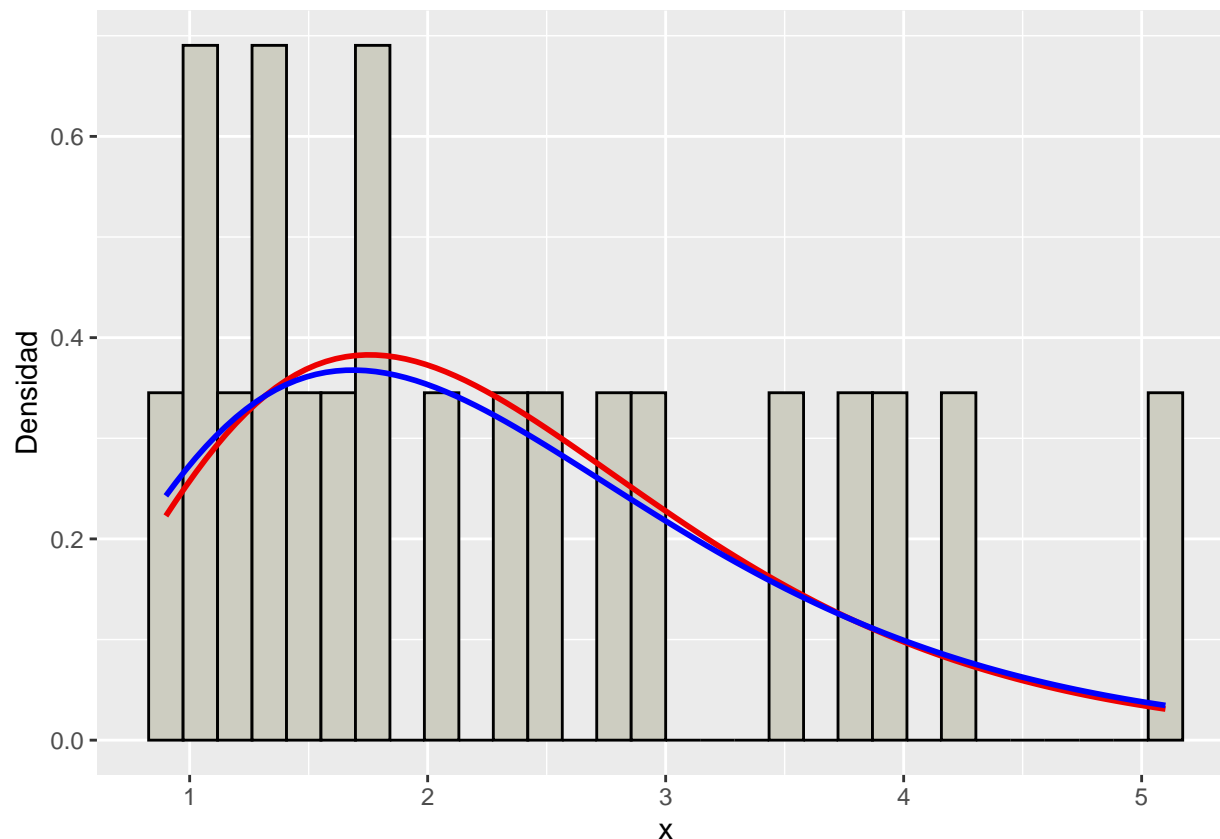
c) Halla las estimaciones maximo-verosímiles de  $p$  y de  $a$ . Compáralas con las del apartado anterior.

La función de densidad no puede ser derivada de manera usual, luego hemos de recurrir a optimizaciones numéricas.

```
p = as.numeric(p_a_mom[1,"p"])
a = as.numeric(p_a_mom[1,"a"])

suma = sum(observaciones)
logsuma = sum(log(observaciones))
n = length(observaciones)
log_ver_gamma = function (theta) {
  p = theta[1]
  a = theta[2]
  # Multiplicamos por (-1) porque 'optim' por defecto minimiza
  -(p*n*log(a) - n * log(gamma(p)) -a * suma + (p-1)*logsuma)
}
# Parámetros p y a, respectivamente
(gamma_emv = optim(par=c(p,a), fn=log_ver_gamma, gr=NULL)$par)

## [1] 3.997510 1.708325
```



Se muestra la estimación por el método de los momentos en azul, y la del EMV en rojo.

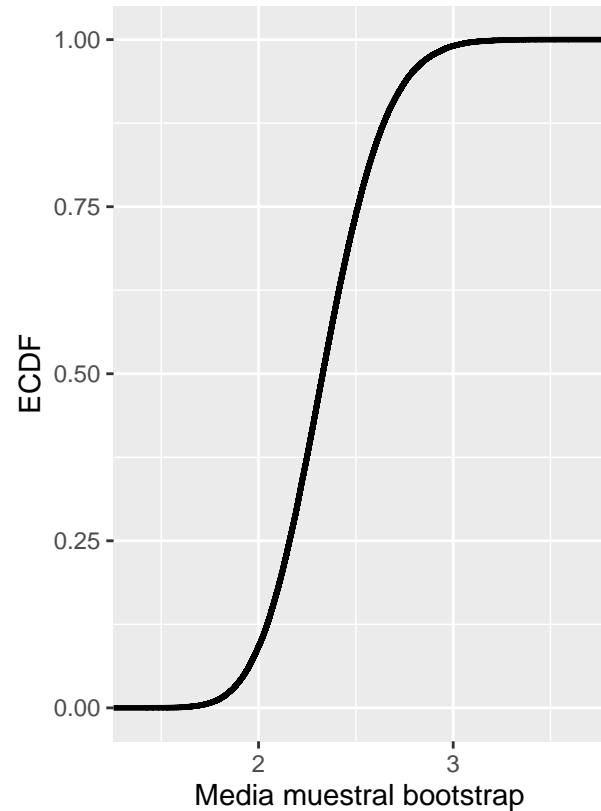
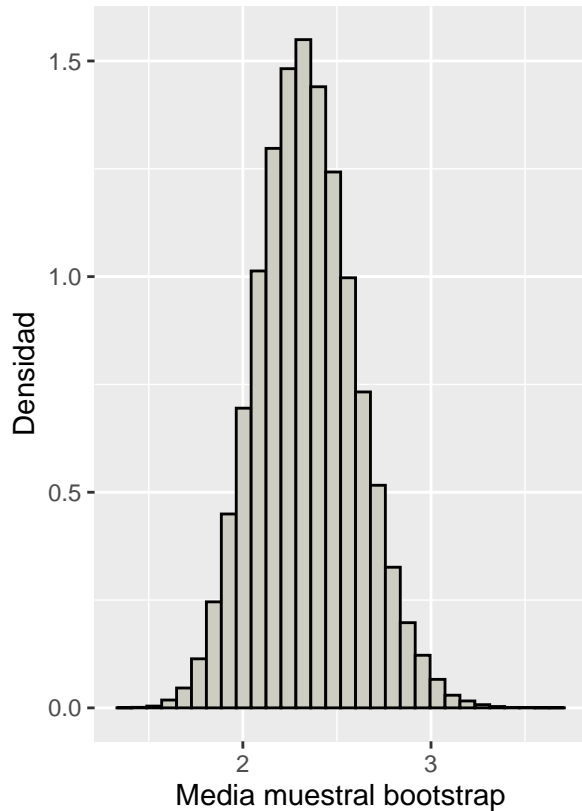
**d) (1 punto) Calcula mediante bootstrap un intervalo de confianza para el parámetro  $\mu = \frac{p}{a}$  a nivel  $1 - \alpha = 0.95$ .**

En este caso, queremos estimar  $E[X]$ , no  $p$  ni  $a$ . Lo haremos por medio de la media muestral. Al trabajar con bootstrap, consideraremos que nuestra muestra de observaciones constituye la población bootstrap, de la que se obtienen remuestreos.

Comenzemos utilizando bootstrap paramétrico para generar las muestras según una distribución  $\gamma(p_{MV}, a_{MV})$ . Usaremos además bootstrap percentil para ver la distribución del parámetro directamente.

```
B = 1e5
p_mv = gamma_emv[1]
a_mv = gamma_emv[2]
dist_media = replicate(B, {
  mean(rgamma(n=n, shape=p_mv, rate=a_mv))
})
```

Vamos a graficar la función de distribución y la densidad empíricas del estadístico para intentar dar un intervalo de poca longitud dentro de los de su confianza:



Debido a su alta simetría, consideraremos un intervalo simétrico:

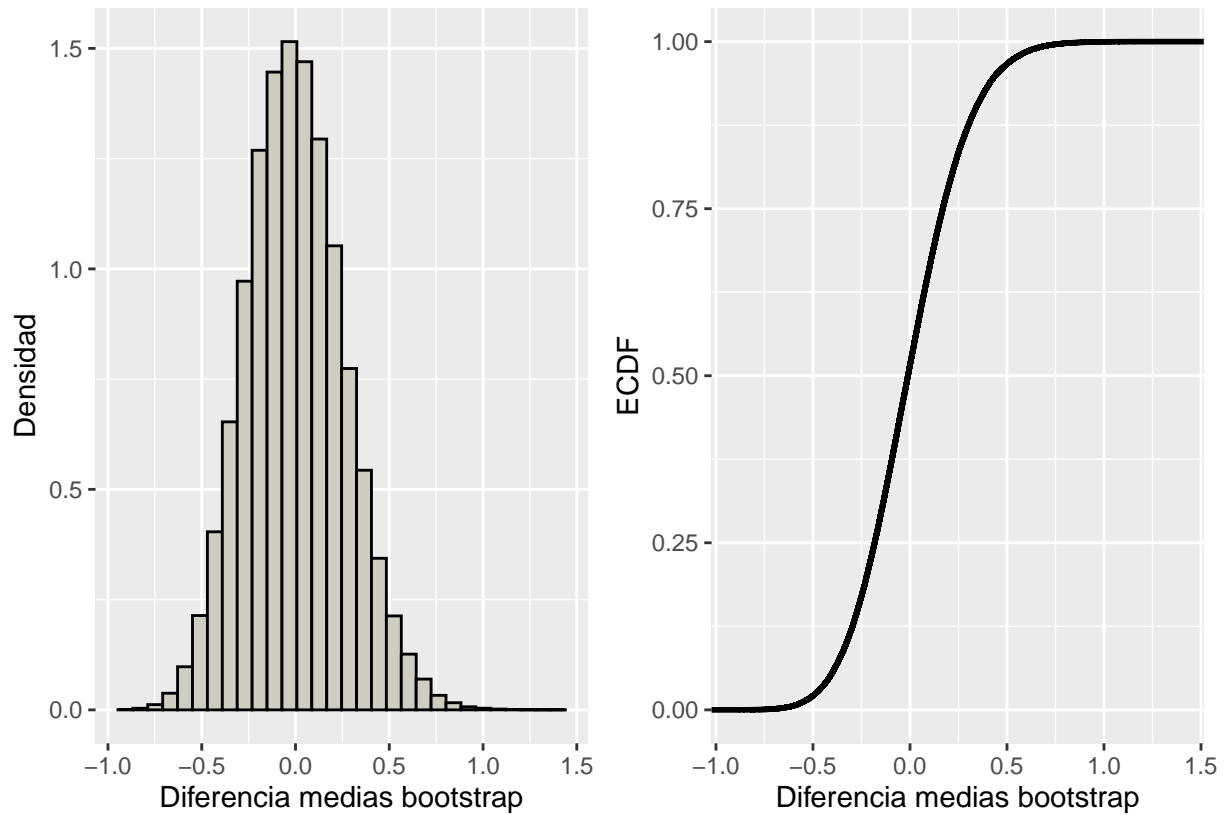
```
alpha = 0.05 # 1- alpha = 0.95
(intervalo = quantile(dist_media, c(alpha/2, 1-alpha/2)))
```

```
##      2.5%      97.5%
## 1.855282 2.882979
```

Usemos ahora bootstrap básico, generando también las muestras paramétricamente.

```
B = 1e5
p_mv = gamma_emv[1]
a_mv = gamma_emv[2]
media = mean(observaciones)
diff_media = replicate(B, {
  mean(rgamma(n=n, shape=p_mv, rate=a_mv) - media)
})
```

En vista del procedimiento anterior, parece intuitivo que guarde mucha simetría. Veamos si es así:

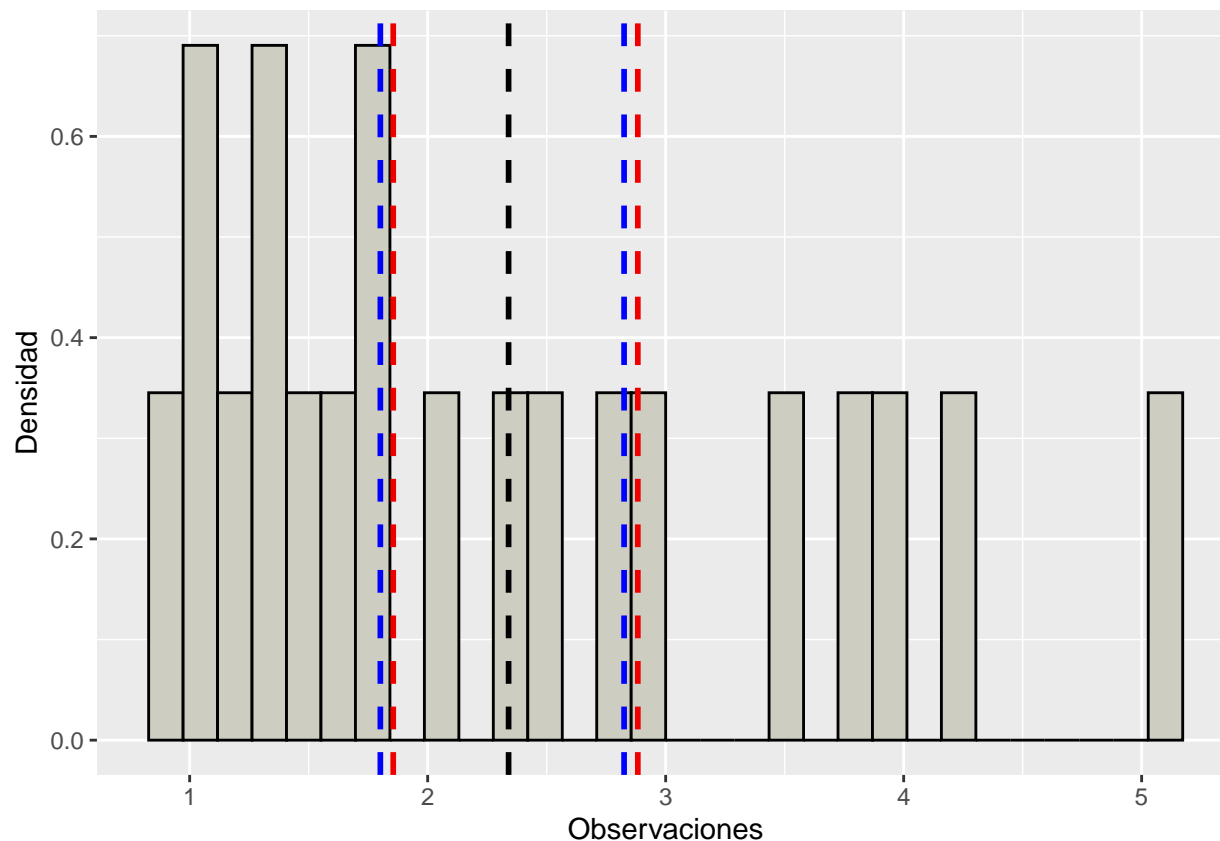


En este caso, la media es todavía más representativa. Por tanto, tomaremos un intervalo centrado:

```
cuantiles = quantile(diff_media, c(alpha/2, 1-alpha/2))
des_pivote = function (q) media - q
(intervalo_basico = c(i1=des_pivote(max(cuantiles)), i2=des_pivote(min(cuantiles))))
```

```
##      i1      i2
## 1.801161 2.825323
```

Comparando con las observaciones muestrales:



Media muestral de las observaciones en negro. Se compara con las medias de los extremos de bootstrap: azul para el percentil y rojo para el básico. Visualmente, parece que ambos métodos tienen longitudes similares, aunque con un ligero desplazamiento entre ambos. Veamos aritméticamente si se cumple la intuición:

```
l_p = max(intervalo) - min(intervalo)
l_b = max(intervalo_basico) - min(intervalo_basico)
paste("Longitud percentil: ", l_p)
```

```
## [1] "Longitud percentil: 1.02769702115521"
```

```
paste("Longitud básico: ", l_b)
```

```
## [1] "Longitud básico: 1.02416253270797"
```

```
l_b < l_p
```

```
## [1] TRUE
```

Son muy similares. De hecho, al no haber utilizado “set.seed” cada vez que genero este PDF obtengo resultados comparativos distintos (a veces el percentil da menor longitud que el básico, pero en ocasiones se cumple lo contrario).