

Problemas de Intervalos de Confianza

Inferencia Estadística

13 de diciembre de 2022

- (Esta viñeta indica el texto de las resoluciones.)

1. Sea $X_1, \dots, X_n$ una muestra aleatoria simple de una variable aleatoria X con función de densidad

$$f_{\theta}(x) = \begin{cases} \frac{2x}{\theta^2} & 0 < x < \theta \\ 0 & \text{en el resto} \end{cases}$$

Resuelve los siguientes apartados:

a) Determina un estadístico suficiente y minimal.

- $f(x; \theta) = \frac{2x}{\theta^2} \cdot [x < \theta] \implies L(\vec{x}; \theta) = \frac{2^n}{\theta^{2n}} \prod x_i \cdot [x_{(n)} < \theta] \implies X_{(n)}$ es suficiente.
- Sean $\vec{x}$ y $\vec{y}$ muestras. Entonces $\left\{ \frac{L(\vec{x}; \theta)}{L(\vec{y}; \theta)} = \frac{\prod x_i \cdot [x_{(n)} < \theta]}{\prod y_i \cdot [y_{(n)} < \theta]} \right.$ es independiente de $\theta \iff x_{(n)} = y_{(n)}$
 $\left. \right\} \implies X_{(n)}$ mínimamente suficiente.

b) Calcula el estimador máximo-verosímil.

- Para $\theta > x_{(n)}$, $\ln L = \ln 2^n - 2n \ln \theta + \ln \prod x_i \implies \partial \ln L / \partial \theta = -2n/\theta < 0 \implies L$ decreciente $\implies \hat{\theta}_{MV} = X_{(n)}$

c) Calcula un intervalo de confianza a nivel $1 - \alpha = 0,95$ por el método de la función pivote para la realización muestral c(4.6, 4.4, 1.8, 3, 4, 2.7, 1.7, 4.3, 3.9, 2.2).

- La variable $Y = X/\theta$ tiene ojiva $F_Y(y) = \Pr[Y \leq y] = \Pr[X \leq y\theta] = F_X(y\theta) = \frac{(y\theta)^2}{\theta^2} = y^2$ independiente de θ , con $0 < y < 1$, luego $Y_{(n)} = X_{(n)}/\theta$ es un pivote, con ojiva $F(y) = y^{2n}$.
- Sean $\alpha_1 + \alpha_2 = \alpha$, $\alpha_1 = F(c_1)$, $1 - \alpha_2 = F(c_2)$. Entonces $c_1 = \sqrt[2n]{\alpha_1}$, $c_2 = \sqrt[2n]{1 - \alpha_2}$ y $1 - \alpha = \Pr[c_1 \leq X_{(n)}/\theta \leq c_2] = \Pr[X_{(n)}/c_2 \leq \theta \leq X_{(n)}/c_1]$, luego I.C. = $X_{(n)} \left[\frac{1}{\sqrt[2n]{1 - \alpha_2}}, \frac{1}{\sqrt[2n]{\alpha_1}} \right]$

$$\implies \text{amplitud} \propto \frac{1}{\sqrt[2n]{\alpha_1}} - \frac{1}{\sqrt[2n]{1 - \alpha_2}} = A(\alpha_1) := \frac{1}{\sqrt[2n]{\alpha_1}} - \frac{1}{\sqrt[2n]{1 - \alpha + \alpha_1}}$$

$$A(a1) := 1/a1^{1/(2*n)} - 1/(1-a+a1)^{1/(2*n)} \quad \$$$

$$\text{assume } (n > 1, a > 0, a < 1, a1 > 0, a1 \leq a) \quad \$$$

$$\text{define } (Aprima(a1), \text{diff}(A(a1), a1)) \quad ;$$

$$\text{minuendo} : \text{part}(Aprima(a1), 1);$$

$$\text{sustraendo} : \text{part}(Aprima(a1), 2, 1);$$

$$\text{is}(\text{minuendo} > \text{sustraendo}); \quad /* \text{unknown} */$$

$$\text{is}(\text{part}(\text{minuendo}, 1) > \text{part}(\text{sustraendo}, 1)); \quad /* \text{unknown} */$$

$$\text{is}(\text{part}(\text{minuendo}, 1, 1) > \text{part}(\text{sustraendo}, 1, 1)); \quad /* \text{true} */$$

$$/* \implies \text{part}(\text{minuendo}, 1) < \text{part}(\text{sustraendo}, 1) */$$

$$/* \implies \text{minuendo} < \text{sustraendo} \implies Aprima < 0 \implies A \text{ decrece} */$$

$$\implies \alpha_1 = \alpha \text{ es } \acute{o}ptimo \implies \text{I.C.} = \left[X_{(n)}, \frac{X_{(n)}}{\sqrt[2n]{\alpha}} \right]$$

$$X \leftarrow c(4.6, 4.4, 1.8, 3, 4, 2.7, 1.7, 4.3, 3.9, 2.2)$$

$$Xn \leftarrow \max(X) \quad ; \quad n \leftarrow \text{length}(X) \quad ; \quad \text{alfa} \leftarrow 0.05$$

$$\text{print } (c(Xn, Xn/\text{alfa}^{1/(2*n)}))$$

```
[1] 4.600000 5.343297
```

d) Calcula un intervalo de confianza a nivel $1 - \alpha$ por el método *bootstrap* y compara su longitud con el intervalo del apartado anterior.

■ Haciendo los cálculos en R:

```
n <- 10 ; B <- 1e5 ; a <- .05
X <- c(4.6, 4.4, 1.8, 3, 4, 2.7, 1.7, 4.3, 3.9, 2.2)
T <- max # estimador
t <- T(X)
distri <- replicate (B, T(sample(X,TRUE)))
## intervalo pivotal
cat("pivotal:", t, t/a^(1/2/n), "\n")
## bs percentil/cuantil: no funciona con extremos
cat("bs cuantil: ")
print (ICc <- quantile (distri, c(a,1), names=FALSE))
## bs básico
cat("bs básico: ")
print (ICb <- t - quantile (distri-t, c(1,a), names=FALSE))
## bs paramétrico
distri <- replicate (B, T(rbeta(n,2,1)*t))
cat("bs paramétrico: ")
print (ICp <- t - quantile (distri-t, c(1,a), names=FALSE))
## t-bústrap paramétrico
dtT <- function (t) t*sqrt(n/((n+1)^2*(n+2)))
distri <- replicate (B, ((Tb<-T(rbeta(n,2,1)*t))-t)/dtT(Tb))
cat("t-bústrap: ")
print (ICt <- t - quantile (distri, c(1,a), names=FALSE) * dtT(t))
## estimador puntual
cat ("estimación puntual:", t, "\n Longitudes:\n")
## longitudes
cat("pivotal:", t*(1/a^(1/2/n)-1), "\n")
cat("bs cuantil:", diff(ICc), "\n")
cat("bs básico:", diff(ICb), "\n")
cat("bs paramétrico:", diff(ICp), "\n")
cat("t-bs paramétrico:", diff(ICt), "\n")
```

```
pivotal: 4.6 5.343297
```

```
bs cuantil: > [1] 4.3 4.6
```

```
bs básico: > [1] 4.6 4.9
```

```
bs paramétrico: > [1] 4.600000 5.244668
```

```
t-bústrap: > [1] 4.600002 5.348207
```

```
estimación puntual: 4.6
```

```
Longitudes:
```

```
pivotal: 0.7432972
```

bs cuantil: 0.3

bs básico: 0.3

bs paramétrico: 0.6446675

t-bs paramétrico: 0.7482052

El bústrap cuantil no funciona en absoluto.

El bústrap básico genera una distribución demasiado discreta y un intervalo muy estrecho.

El bústrap paramétrico da un intervalo más estrecho que el pivotal.

El t-bústrap paramétrico, con muchas repeticiones, da casi el mismo resultado que el pivote. (Mediante simulaciones, se puede comprobar que el bústrap básico y el paramétrico, a nivel 95 %, cubren respectivamente menos del 85 % y del 90 %.)

2. Sea X una variable con distribución $F_\theta(x) = x^\theta$ para $x \in (0, 1)$. Calcular un intervalo de confianza a nivel $1 - \alpha = 0,9$ para una muestra aleatoria simple de tamaño n .

- $X \hookrightarrow F_X(x) = x^\theta \implies f(x) = \theta x^{\theta-1} \sim \beta(\theta, 1) \implies L(\vec{x}, \theta) = \theta^n (\prod x_i)^{\theta-1} \implies l := \ln L = n \ln \theta + (\theta - 1)(-T)$ con $T := -\ln \prod x_i = -\sum \ln x_i \implies l' = n/\theta - T \implies \hat{\theta}_{MV} = n/T$
- $X \hookrightarrow \beta(\theta, 1) \implies -\ln X \hookrightarrow \text{Exp}(\theta) = \gamma(1, \theta) \implies T \hookrightarrow \gamma(n, \theta) \implies \theta T \hookrightarrow \gamma(n, 1)$ es un pivote.
- $\alpha_1 + \alpha_2 := \alpha, \quad \Pr[\gamma(n, 1) < c_1] := \alpha_1, \quad \Pr[\gamma(n, 1) > c_2] := \alpha_2 \implies 1 - \alpha = \Pr[c_1 \leq \theta T \leq c_2] = \Pr[c_1/T \leq \theta \leq c_2/T]$
- Como $\gamma(n, 1)$ es acampanada, suele tomarse $\alpha_1 = \alpha_2 = \alpha/2 = 0,05$

3. La duración de cierto tipo de válvulas sigue una distribución $\gamma(1, \lambda)$. Se desea determinar qué duración θ es superada por el 90 % de las válvulas. Para ello se dispone de la duración de 9 válvulas: 15,5 7,3 12,5 22 14 12 30 10 10,5. Determinar un intervalo de confianza para θ con $1 - \alpha = 0,95$.

- $X \hookrightarrow \gamma(1, \lambda) = \text{Exp}(\lambda) \implies 0,9 = \Pr[X > \theta] = e^{-\lambda\theta} \implies \lambda = (-\ln 0,9)/\theta$
- $X \hookrightarrow \gamma(1, \lambda) \implies \sum X_i \hookrightarrow \gamma(n, \lambda) = \gamma(n, [-\ln 0,9]/\theta) \implies (-\ln 0,9) \sum X_i/\theta \hookrightarrow \gamma(n, 1)$ pivote.
- $n = 9 \quad \alpha_1 = \alpha_2 = \alpha/2 = 0,025 \implies (c_1, c_2) =$

```
print (c (c1 = c1 <- qgamma(0.025,9,1), c2 = c2 <- qgamma(0.975,9,1)))
```

```
      c1      c2
4.115373 15.763189
```

$\implies \text{I.C.} = (-\ln 0,9) \sum x_i \cdot [1/c_2, 1/c_1] =$

```
X <- c(15.5, 7.3, 12.5, 22, 14, 12, 30, 10, 10.5)
print (-log(0.9) * sum(X) / c(c2,c1))
```

```
[1] 0.8943138 3.4255064
```

4. El número de partículas alfa emitidas, en un segundo, por una sustancia radiactiva sigue una distribución $P(\lambda)$. Se dispone de una muestra aleatoria simple de 10 segundos $(x_1, \dots, x_{10})$ durante los que se detectaron tres partículas alfa. Calcula un intervalo de confianza para λ con coeficiente de confianza $1 - \alpha = 0,95$.

- $\sum X_i \hookrightarrow P(n\lambda)$ que es discreta y no puede proporcionar un pivote exacto, así que buscaremos uno asintótico, aunque el tamaño muestral sea pequeño.

- Por el T.C.L., $\bar{X} \xrightarrow{d} N(\lambda, \sqrt{\lambda/n})$ luego $1 - \alpha = \Pr[|\bar{X} - \lambda|/\sqrt{\lambda/n} \leq z_{1-\alpha/2}] = \Pr[(\bar{X} - \lambda)^2/(\lambda/n) < z^2] \implies \text{I.C.} =$

```
expresion : (Xm - l)^2 / (l/n) - z $
print (expand (expresion)) $
soluciones : solve (expresion, l) $
print (soluciones) $
```

$$(-z)^2 + \frac{1}{n} + \frac{Xm^2}{n} - 2Xm \frac{z}{n}$$

$$[1 = -\frac{\sqrt{z^2 + 4Xm \frac{z}{n}} - z - 2Xm \frac{z}{n}}{2n}, 1 = \frac{\sqrt{z^2 + 4Xm \frac{z}{n}} + z + 2Xm \frac{z}{n}}{2n}]$$

- En el enunciado, $n = 10$, $\bar{X} = 3/10$, luego I.C. =

```
n : 10 $ Xm : 3/10 $ z : 1.96 $
expresion : (Xm - l)^2 / (l/n) - z $
soluciones : ev (solve (expresion, l), numer) $
print (soluciones) $
```

```
[1 = 0.1364584163082282, 1 = 0.6595415836917717]
```

- Otra opción sería usar bústrap paramétrico (básico o percentil, porque el t-bústrap produce intervalos infinitos debido a estimaciones nulas de la varianza):

```
n <- 10 ; T0 <- 3/10 ; a <- 0.05 ; B <- 1000
distri <- replicate (B,
{
  Xb <- rpois (n, T0)
  Tb <- mean(Xb)
  Tb - T0
})
T0 - quantile(distri, c(1-a/2,a/2), names=FALSE) # básico
```

```
[1] -0.1 0.6
```

```
distri <- replicate (B,
{
  Xb <- rpois (n, T0)
  mean(Xb)
})
quantile(distri, c(a/2,1-a/2), names=FALSE) # percentil
```

```
[1] 0.0 0.7
```

5. En el lenguaje de la silvicultura se denomina *bosque de Poisson* de intensidad λ a aquél que satisface:

- a) El número de árboles existentes en una parcela de área A contenida en el bosque es una variable aleatoria con ley de Poisson de parámetro λA .

- b) Si $A_1, \dots, A_n$ son parcelas disjuntas dentro del bosque, las variables aleatorias «número de árboles en cada parcela» son independientes.

Resuelve las siguientes cuestiones referentes a cierto bosque de Poisson de intensidad λ :

- a) Fijado un punto P cualquiera en el bosque, probar que la ley de probabilidad de la variable aleatoria «distancia de P al árbol más próximo» tiene densidad $f(x) = 2\pi\lambda x \exp(-\pi\lambda x^2)$ si $x > 0$.
- Sea D dicha distancia. $F_D(x) = \Pr[D \leq x] = 1 - \Pr[D > x] = 1 - \Pr[X = 0] = 1 - \exp(-\lambda\pi x^2)$, siendo X el número de árboles en el círculo con centro P y radio x , luego $X \hookrightarrow P(\lambda\pi x^2)$.
 - Entonces $f_D(x) = F'_D(x) = [1 - \exp(-\lambda\pi x^2)]' = -\exp(-\lambda\pi x^2) \cdot (-\lambda\pi 2x)$
 - Por otro lado, la forma de F_D sugiere que la distribución de D está relacionada con la exponencial; considerando $Y := D^2$ se tiene que $F_Y(y) = \Pr[D^2 \leq y] = \Pr[D \leq \sqrt{y}] = 1 - \exp(-\lambda\pi\sqrt{y}^2) = 1 - \exp(-\lambda\pi y) \implies D^2 \hookrightarrow \text{Exp}(\lambda\pi)$
- b) Fijados 10 puntos al azar suficientemente alejados en el bosque se midieron las distancias en metros $X_1, \dots, X_{10}$ de cada punto al árbol más próximo, obteniéndose los siguientes resultados: (7.11, 3.24, 2.31, 5.38, 4.29, 5.63, 4.91, 6.08, 7.89, 3.53)

Con dicha información, obtener:

- 1) Una estimación máximo-verosímil para $\theta := 1/\lambda$.

- $L(\vec{x}, \theta) = \prod f(x_i, \theta) = \prod 2\pi x_i \exp(-\pi x_i^2/\theta)/\theta = 2^n \pi^n \prod x_i \exp(-\pi \sum x_i^2/\theta)/\theta^n \implies l := \ln L = n \ln 2 + n \ln \pi + \sum \ln x_i - \pi \sum x_i^2/\theta - n \ln \theta \implies l' = \pi \sum x_i^2/\theta^2 - n/\theta \implies \hat{\theta}_{\text{MV}} = \pi \sum x_i^2/n$
- Otra forma: $X^2 \hookrightarrow \text{Exp}(\pi/\theta) \implies \widehat{\pi/\theta} = n/\sum x_i^2 \implies \widehat{1/\theta} = n/(\pi \sum x_i^2) \implies \hat{\theta} = \pi \sum x_i^2/n$ por la equivariancia del estimador máximo-verosímil.
- estimación =

```
X <- c(7.11, 3.24, 2.31, 5.38, 4.29, 5.63, 4.91, 6.08, 7.89, 3.53)
n <- length(X)
pi * sum(X^2) / n
```

[1] 88.34746

- 2) Un intervalo de confianza para λ con coeficiente de confianza 0,99.

- $X^2 \hookrightarrow \text{Exp}(\pi\lambda) \implies \sum X_i^2 \hookrightarrow \gamma(n, \pi\lambda) \implies \pi\lambda \sum X_i^2 \hookrightarrow \gamma(n, 1)$ es un pivote.
- $1 - \alpha = \Pr[c_1 \leq \pi\lambda \sum X_i^2 \leq c_2] = \Pr[c_1/(\pi \sum X_i^2) \leq \lambda \leq c_2/(\pi \sum X_i^2)]$
- $\alpha_i = \alpha/2 \implies$

```
a <- 0.01
qgamma (c(a/2, 1-a/2), n) / (pi * sum(X^2)) # I.C.
```

[1] 0.004207164 0.022636104

- c) Exponer otro método de recogida de datos para estimar el parámetro λ .

- Fijar varias parcelas $P_1, \dots, P_n$ de área A cada una y contar el número de árboles en cada una para obtener una muestra de una $P(\lambda A)$.
- Fijar un punto C como centro de un círculo de radio R y considerar las distancias cuadráticas a C de los árboles inscritos, como una muestra aleatoria simple de una variable con distribución exponencial de parámetro λ truncada a $[0, R]$.

6. La vida media de un virus sigue una distribución $N(\mu, 3)$. Se toman 10 cultivos de dicho virus, y se obtiene que el tiempo medio de supervivencia del virus es de 12 horas.

- a) Calcular un intervalo de confianza, con $1 - \alpha = 0,99$, para la media de la distribución.

- pivote $(\bar{X} - \mu)/(3/\sqrt{n}) \hookrightarrow N(0,1) \implies \text{I.C.} = \bar{X} \pm z_{1-\alpha/2} \cdot \frac{3}{\sqrt{n}}$
 $12 + c(-1,1) * \text{qnorm}(.995) * 3 / \text{sqrt}(10)$
 $[1] \quad 9.556354 \quad 14.443646$

b) ¿Qué tamaño de muestra sería necesario para que el intervalo tuviese longitud menor que 1?

- $1 > \text{longitud} = 2 \cdot z_{1-\alpha/2} \cdot \frac{3}{\sqrt{n}} \iff n > (6 \cdot z_{1-\alpha/2})^2$
 $(6 * \text{qnorm} (.995))^2$
 $[1] \quad 238.8563$
 $\implies n \geq 239$

7. El departamento de control de calidad de una empresa desea estimar la proporción, p , de piezas defectuosas que se producen en la fábrica. Para ello elige al azar una muestra de treinta piezas, de las cuales cinco presentan algún defecto. Calcula la estimación máximo-verosímil de p .

- $X_i := \mathbb{1}(\text{pieza } i.\text{a} \text{ defectuosa}) \implies X \hookrightarrow B(1,p) \implies L(\vec{x},p) = p^{\sum x_i} (1-p)^{n-\sum x_i} \implies \hat{p}_{\text{MV}} = \frac{\sum x_i}{n} = \bar{x} \implies \text{estimación} = 5/30 \approx 17\%$