estimación intervalar

Carlos Carleos, Norberto Corral, Teresa López

10 de diciembre de 2022

Índice

1. definición

- $X \hookrightarrow F_\theta$, $\theta \in \Theta$
- $\vec{X} = X_1, \ldots, X_n$ muestra aleatoria simple de X
- $\alpha \in (0,1)$, α pequeño; $1-\alpha$ se llama **coeficiente de confianza**
- T_1 y T_2 estadígrafos tales que $\forall \theta \in \Theta$, $\Pr_\theta\{\vec{x} \in X(\Omega)^n \mid T_1(\vec{x}) \leqslant g(\theta) \leqslant T_2(\vec{x})\} \geqslant 1-\alpha$
- $[T_1(\vec{X}), T_2(\vec{X})]$ es una **horquilla** o **intervalo de confianza** (IC) para $g(\theta)$ a nivel $1-\alpha$

2. advertencia

- dada una realización muestral $\vec{x}_0$, en inferencia frecuentista (la nuestra) no se puede afirmar que $\Pr\{\theta \in [T_1(\vec{x}_0), T_2(\vec{x}_0)]\} \geqslant 1-\alpha$
- se trata de intervalos de *confianza*, no de *probabilidad*

3. ejemplo

$X \hookrightarrow N(\mu, 1) \implies \frac{\bar{X}-\mu}{1/\sqrt{n}} \hookrightarrow N(0,1) \implies \text{IC} = \bar{X} \pm z_{1-\frac{\alpha}{2}} \frac{1}{\sqrt{n}} = \left[\bar{X} - z_{1-\frac{\alpha}{2}} \frac{1}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}} \frac{1}{\sqrt{n}}\right]$ con $\Pr[N(0,1) \leqslant z_{1-\frac{\alpha}{2}}] = 1 - \frac{\alpha}{2}$

4. método pivotal

- la función $T(\vec{X}, \theta)$ es un **pivote** si su distribución de probabilidad no depende de θ (está completamente especificada)
- algoritmo para construir IC mediante un pivote
  1. elegir $\alpha_1 + \alpha_2 = \alpha$ (habitualmente, $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$)
  2. buscar c_1 y c_2 tales que $\Pr(T < c_1) \leqslant \alpha_1$ y $\Pr(T \leqslant c_2) \geqslant 1 - \alpha_2$
  3. expresar $c_1 \leqslant T(\vec{x}, \theta) \leqslant c_2$ como intervalo en torno a θ

4.1. ejemplo

$X \hookrightarrow N(\mu, \sigma) \implies$ pivotes:

- $\frac{\bar{X}-\mu}{\hat{S}/\sqrt{n}} \hookrightarrow t_{n-1} \implies \text{IC} = \bar{X} \pm t_{n-1,1-\frac{\alpha}{2}} \frac{\hat{S}}{\sqrt{n}} = \left[\bar{X} - t_{n-1,1-\frac{\alpha}{2}} \frac{\hat{S}}{\sqrt{n}}, \bar{X} + t_{n-1,1-\frac{\alpha}{2}} \frac{\hat{S}}{\sqrt{n}}\right]$ con $\Pr[t_{n-1} \leqslant t_{n-1,1-\frac{\alpha}{2}}] = 1 - \frac{\alpha}{2}$
- $\frac{(n-1)\hat{S}^2}{\sigma^2} \hookrightarrow \chi^2_{n-1} \implies \text{IC} = \left[\frac{(n-1)\hat{S}^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}, \frac{(n-1)\hat{S}^2}{\chi^2_{n-1,\frac{\alpha}{2}}}\right]$

4.2. pivote genérico

$X \hookrightarrow F_\theta$ absolutamente continua $\implies$ puede usarse el pivote $-\sum \ln F_\theta(X_i) \hookrightarrow \gamma(n,1)$ pues

- $F_X(X) \hookrightarrow U(0,1)$
- $-\ln U(0,1) \hookrightarrow \text{Exp}(1)$
- $\sum_{i=1}^n \text{Exp}(\lambda)$ independientes $\hookrightarrow \gamma(n, \lambda)$

4.3. comparación de intervalos

- motivo
  - pueden existir varios pivotes
  - con un mismo pivote, pueden escogerse α_1 y α_2 de varias maneras
- criterio
  - preferir el intervalo de amplitud (esperada) mínima
- ejemplo: $X \hookrightarrow U(0,\theta)$; dos procedimientos
  - pivote genérico:
    - $F(x) = \frac{x}{\theta} \implies -\sum \ln \frac{X_i}{\theta} \hookrightarrow \gamma(n,1)$
    - $a = -\sum \ln \frac{X_i}{\theta} = -\ln \prod \frac{X_i}{\theta} \implies e^{-a} = \frac{\prod X_i}{\theta^n} \implies \theta^n = \frac{\prod X_i}{e^{-a}} \implies \theta = \sqrt[n]{e^a \prod X_i}$
    - sean a y b tales que $\Pr[a \leqslant \gamma(n,1) \leqslant b] = 1-\alpha$
    - $\Pr[a \leqslant \gamma(n,1) \leqslant b] = \Pr\left[\sqrt[n]{e^a \prod X_i} \leqslant \theta \leqslant \sqrt[n]{e^b \prod X_i}\right] \implies$ amplitud $L = \sqrt[n]{e^b \prod X_i} - \sqrt[n]{e^a \prod X_i} = \sqrt[n]{\prod X_i}\left(\sqrt[n]{e^b} - \sqrt[n]{e^a}\right)$
    - $E\left[\sqrt[n]{\prod X_i}\right] = E\left[\prod \sqrt[n]{X_i}\right] = \prod E\left[\sqrt[n]{X_i}\right] = E^n\left[\sqrt[n]{X}\right]$
    - $E\left(X^{\frac{1}{n}}\right) = \int_0^\theta x^{\frac{1}{n}} \frac{1}{\theta} dx = \left|\frac{1}{\theta} \frac{x^{\frac{1}{n}+1}}{\frac{1}{n}+1}\right|_{x=0}^{x=\theta} = \frac{n}{n+1}\theta^{\frac{1}{n}}$
    - $E(L) = \theta\left(\frac{n}{n+1}\right)^n \left(e^{\frac{b}{n}} - e^{\frac{a}{n}}\right)$
  - pivote basado en un estimador, como $X_{(n)}$
    - $U = \frac{X_{(n)}}{\theta} \hookrightarrow F_{U(0,1)}(u)^n = u^n \ \forall u \in (0,1)$
    - $1-\alpha = \Pr\left[a \leqslant \frac{X_{(n)}}{\theta} \leqslant b\right] \implies a = \sqrt[n]{\alpha_1},\ b = \sqrt[n]{1-\alpha_2} \implies \Pr\left[\frac{X_{(n)}}{\sqrt[n]{1-\alpha_2}} \leqslant \theta \leqslant \frac{X_{(n)}}{\sqrt[n]{\alpha_1}}\right]$
    - $E(L) = \frac{n}{n+1}\theta\left(\frac{1}{\sqrt[n]{\alpha_1}} - \frac{1}{\sqrt[n]{1-\alpha_2}}\right)$
    - minimizar la amplitud esperada: $L(\alpha_1) = \frac{n}{n+1}\theta\left(\frac{1}{\sqrt[n]{\alpha_1}} - \frac{1}{\sqrt[n]{1-\alpha+\alpha_1}}\right)$
    - $\frac{\partial L}{\partial \alpha_1} = \frac{n}{n+1}\theta\frac{1}{n}\left(-\frac{1}{\alpha_1^{\frac{1}{n}+1}} + \frac{1}{(1-\alpha+\alpha_1)^{\frac{1}{n}+1}}\right) < 0$ si $\alpha_1 < 1-\alpha+\alpha_1 \iff 0 < 1-\alpha \implies L$ decreciente $\implies$ mínimo con $\alpha_1 = \alpha$ y $\alpha_2 = 0$

```
### X = U(0,zita)
longitudesEsperadas <- function (n, alfa, zita)
{
    ## pivote genérico con alfa1=alfa2
    a <- qgamma(alfa/2,n,1)
    b <- qgamma(1-alfa/2,n,1)
    L1 <- zita * (n/(n+1))^n * (exp(b/n)-exp(a/n))
    ## pivote genérico con alfa1 y alfa2 óptimos
    fL2 <- function (alfa1)
    {
        a <- qgamma(alfa1,n,1)
        b <- qgamma(1-alfa+alfa1,n,1)
        zita * (n/(n+1))^n * (exp(b/n)-exp(a/n))
    }
    L2 <- optimize (fL2, c(0,alfa)) $ objective
```

```
    ## pivote EMV
    L3 <- n/(n+1) * zita * (1/alfa^(1/n)-1)
    ## genérico Simétrico y Óptimo, y EMV
    c (genéricoS=L1, genéricoÓ=L2, EMV=L3)
}
n <- 100
alfa <- .05
zita <- 1
longitudesEsperadas (n, alfa, zita)
intervalos <- function (x, alfa)
{
    n <- length (x)
    ## pivote genérico...
    T <- prod (x)
    extremo <- function (a) (exp(a)*T)^(1/n)
    ## ...con alfa1=alfa2
    a <- qgamma (alfa/2, n, 1)
    b <- qgamma (1-alfa/2, n, 1)
    Gs1 <- extremo (a)
    Gs2 <- extremo (b)
    L1 <- Gs2 - Gs1
    ## ...con alfa1 y alfa2 óptimos
    fL2 <- function (alfa1)
    {
        a <- qgamma(alfa1,n,1)
        b <- qgamma(1-alfa+alfa1,n,1)
        zita * (n/(n+1))^n * (exp(b/n)-exp(a/n))
    }
    alfa1 <- optimize (fL2, c(0,alfa)) $ minimum
    a <- qgamma (alfa1, n, 1)
    b <- qgamma (1-alfa+alfa1, n, 1)
    Go1 <- extremo (a)
    Go2 <- extremo (b)
    L2 <- Go2 - Go1
    ## pivote EMV
    T <- max (x)
    E1 <- T
    E2 <- T / alfa^(1/n)
    L3 <- E2 - E1
    rbind (genéricoS = c(inf=Gs1, sup=Gs2, lon=L1),
           genéricoÓ = c(inf=Go1, sup=Go2, lon=L2),
           EMV       = c(ind=E1,  sup=E2,  lon=L3))
}
intervalos (runif(100,0,zita), alfa)
## cobertura
apply (replicate (1e5,
{
    x <- runif (100, 0, zita)
    intervalos.x <- intervalos (x, alfa)
    cubre <- apply (intervalos.x, 1,
                    function (inf.sup.lon)
                        inf.sup.lon["inf"] <= zita &
```

```
                        zita <= inf.sup.lon["sup"])
}),
1,
mean)
1 - alfa # confianza

### ejemplo de ejecución

> longitudesEsperadas (n, alfa, zita)
 genéricoS  genérico□́        EMV
0.39988706 0.39445253 0.03010946

> intervalos (runif(100,0,zita), alfa) # zita=1
                inf      sup        lon
genéricoS 1.0131663 1.498898 0.48573135
genérico□́ 0.9963855 1.475516 0.47913018
EMV       0.9925257 1.022709 0.03018326

genéricoS genérico□́       EMV # coberturas
  0.95001   0.95001   0.95082
> 1 - alfa # confianza
[1] 0.95
```

5. EXTRA método de Neyman

- parámetro θ, confianza $1-\alpha$, estimador T
- para cada $\theta \in \Theta$ hallar $a(\theta)$ y $b(\theta)$ tales que $\Pr_\theta[a(\theta) \leqslant T \leqslant b(\theta)] \geqslant 1-\alpha$
- sea $C = \{(\theta, t) \mid \theta \in \Theta, t \in [a(\theta), b(\theta)]\} \subset \mathbb{R}^2$
- dada la realización muestral $\vec{x}_0$ y $t_0 = T(\vec{x}_0)$, se tiene la horquilla $\mathrm{IC} = \{\theta \mid (\theta, t_0) \in C\}$

6. método asintótico

6.1. pivote para EMV

n grande $\implies (\hat{\theta}_{\mathrm{MV}} - \theta)\sqrt{nI(\theta)} \hookrightarrow N(0,1)$

6.2. TCL

ejemplo:

- $X \hookrightarrow B(1,p) \overset{\text{TCL}}{\Longrightarrow} \frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}} \overset{\sim}{\hookrightarrow} N(0,1) \quad \implies \quad 1 - \alpha = \Pr\left[-z_{1-\frac{\alpha}{2}} \leqslant \frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}} \leqslant +z_{1-\frac{\alpha}{2}}\right]$ $= \Pr\left[\left|\frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}}\right| \leqslant z_{1-\frac{\alpha}{2}}\right] = \Pr\left[\frac{(\bar{X}-p)^2}{\frac{p(1-p)}{n}} \leqslant z^2_{1-\frac{\alpha}{2}}\right] \quad \implies \quad \mathrm{IC} =$ $\left[-\frac{z_{1-\frac{\alpha}{2}}\sqrt{z_{1-\frac{\alpha}{2}}{}^2+\left(4\,\bar{X}-4\,\bar{X}^2\right)n}-z_{1-\frac{\alpha}{2}}{}^2-2\,\bar{X}\,n}{2\,z_{1-\frac{\alpha}{2}}{}^2+2\,n}, \frac{z_{1-\frac{\alpha}{2}}\sqrt{z_{1-\frac{\alpha}{2}}{}^2+\left(4\,\bar{X}-4\,\bar{X}^2\right)n}+z_{1-\frac{\alpha}{2}}{}^2+2\,\bar{X}\,n}{2\,z_{1-\frac{\alpha}{2}}{}^2+2\,n}\right]$
- si n es grande, se puede considerar $\frac{\bar{X}-p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \overset{\sim}{\hookrightarrow} N(0,1) \implies \mathrm{IC} = \bar{X} \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}$

6.3. método delta

- cuando se trata de una trasformación de un estadígrafo con distribución asintótica asegurada por el TCL
- ejemplo
    - $X \hookrightarrow P(\lambda) \implies E(X) = \lambda,\ D(X) = \sqrt{\lambda}$
    - TCL $\implies \frac{\bar{X}-\lambda}{\sqrt{\frac{\lambda}{n}}} \overset{\sim}{\hookrightarrow} N(0,1)$

      versión primera despejando λ : $\frac{(\bar{X}-\lambda)^2}{\frac{\lambda}{n}} \overset{\sim}{\hookrightarrow} \chi^2_1 \implies \Pr\left[\frac{(\bar{X}-\lambda)^2}{\frac{\lambda}{n}} < a\right] = 1-\alpha \implies$

      $\Pr\left[\bar{X}^2 + \lambda^2 - 2\bar{X}\lambda < \frac{a}{n}\lambda\right] = 1-\alpha \implies \Pr\left[\lambda^2 + \left(-2\bar{X} - \frac{a}{n}\right)\lambda + \bar{X}^2 < 0\right] = 1-\alpha \implies$

      $\Pr\left[\lambda \in \left(\bar{X} + \frac{a}{2n} \pm \sqrt{\frac{a^2}{4n^2} + \frac{a\bar{X}}{n}}\right)\right] = 1-\alpha$

      versión segunda sustituyendo λ en el denominador por su estimador : $\frac{\bar{X}-\lambda}{\sqrt{\frac{\bar{X}}{n}}} \overset{\sim}{\hookrightarrow} N(0,1) \implies \text{IC} =$

      $\bar{X} \pm z\sqrt{\frac{\bar{X}}{n}}$
    - método $\delta \implies \sqrt{\bar{X}} \overset{\sim}{\hookrightarrow} N\left(\sqrt{\lambda}, \frac{1}{2\sqrt{n}}\right) \implies \text{IC} = \left(\sqrt{\bar{X}} \pm \frac{z}{2\sqrt{n}}\right)^2$
    - nótese que los intervalos TCL2 y δ tienen la misma amplitud:
        - IC TCL: $\left(\bar{X} + z\sqrt{\frac{\lambda}{n}}\right) - \left(\bar{X} - z\sqrt{\frac{\lambda}{n}}\right) = 2z\sqrt{\frac{\lambda}{n}}$
        - IC δ: $\left(\sqrt{\bar{X}} + \frac{z}{2\sqrt{n}}\right)^2 - \left(\sqrt{\bar{X}} - \frac{z}{2\sqrt{n}}\right)^2 = \left(\bar{X} + \frac{z^2}{4n}\right) + 2\sqrt{\bar{X}}\frac{z}{2\sqrt{n}} - \left(\bar{X} + \frac{z^2}{4n}\right) - 2\sqrt{\bar{X}}\frac{z}{2\sqrt{n}} =$
          $2z\sqrt{\frac{\bar{X}}{n}}$

```
## X = P(l) => E(X)=l D(X)=raíz(l)
n = 30                                  # tamaño muestral
l = 5                                   # parámetro
a = 0.05                                # alfa, 1-confianza
resul = replicate (1e5,
{
    X = rpois (n, l)
    T = mean(X)
    ## TCL => (T - l) / raiz(l/n) = N(0,1) => (T - l)^2 / (l/n) = (Ji^2)_1
    ## {Pr[(Ji^2)_1 < ji] = 1-a} => IC: (T-l)^2 < ji l/n =>
    ## => T^2 + l^2 - 2 T l - ji l/n < 0 =>
    ## => l^2 + (-2T - ji/n) l + T^2 < 0 =>
    ji = qchisq(1-a,1)
    Itcl1 = T + ji/2/n + c(-1,1)*sqrt(ji^2/n^2/4+ji*T/n)
    ## TCL => T = N(l,raiz(l/n)) => IC = T +/- z raiz(l/n)
    ## => IC = T +/- z raiz(T/n)
    z = qnorm(1-a/2)
    Itcl2 = T + c(-1,1)*z*sqrt(T/n)
    ## delta => raiz(T) = N(raiz(l),1/[2raiz(n)])
    Idelta = (sqrt(T) + c(-1,1)*z/(2*sqrt(n)))^2
    ## longitudes y coberturas:
    c (Ltcl1 = diff(Itcl1), Ltcl2 = diff(Itcl2), Ldelta = diff(Idelta),
       Ctcl1 = Itcl1[1] <= l & l <= Itcl1[2],
       Ctcl2 = Itcl2[1] <= l & l <= Itcl2[2],
       Cdelta = Idelta[1] <= l & l <= Idelta[2])
})
```

```
apply(resul,1,mean)

### ejemplo de ejecución:

   Ltcl1    Ltcl2   Ldelta    Ctcl1    Ctcl2   Cdelta
1.604105 1.598978 1.598978 0.954893 0.952269 0.950206
```

7. remuestreo autosuficiente o bústrap (*bootstrap*)

7.1. definición

- población $X \hookrightarrow F_\theta$, estimador $\hat{\theta} = T$, realización muestral $\vec{x}_0 = (x_{01}, \dots, x_{0n})$
- **objetivo**: aproximar la distribución de T
- **idea**: considerar $\vec{x}_0$ como una población nueva, con distribución F_n (la ojiva empírica), es decir, $\forall i \in \{1, \dots, n\}$, $\Pr(x_{0i}) = \frac{1}{n}$
- algoritmo general: iterar B veces
  - obtener $\vec{x}^*$ muestra de tamaño n con reposición a partir de $\vec{x}_0$
  - calcular $t^* = T(\vec{x}^*)$

	Población	Muestra	Parámetro	Estimador
original	X	$\vec{x}$	θ	$t = T(\vec{x})$
bústrap	$\vec{x}_0$	$\vec{x}^*$	$t_0 = T(\vec{x}_0)$	$t^* = T(\vec{x}^*)$

- **étimo**: *bootstrap* es una parte del calzado (identificado a veces con la oreja) y se usa en una expresión inglesa que significa obtener algo en principio imposible; Pedro Gil aludía a la expresión duros a cuatro pesetas (€ a 80 ¢) para ilustrar el método

7.2. hipótesis

se supone que se cumple alguna de las siguientes:

hipótesis A F_n es una buena aproximación de F_θ, luego la distribución de T^* es similar a la de T

hipótesis B $T^* - t_0$ tiene distribución parecida a $T - \theta$

hipótesis C $\frac{T^* - t_0}{D(T^*)}$ tiene distribución parecida a $\frac{T-\theta}{D(T)}$ donde $D(T) = \sigma_T$ es el desvío típico de T

7.3. método gausiano

- (aunque es habitual estimar la varianza del estimador, este método no suele usarse para intervalos; aquí se presenta para introducir intuitivamente el bústrap percentil)
- supone $V(T) = V(T^*)$ (implicado por hipótesis A ó B) y distribución aproximadamente gausiana de T
- estimación bústrap de $V(T)$
  - generar B muestras bústrap $\vec{x}_1^*, \dots, \vec{x}_B^*$
  - aplicar el estimador a cada muestra: $\forall i \in \{1, \dots, B\}$, $t_i^* = T(\vec{x}_i^*)$
  - estimar la varianza $V(T^*)$ a partir de la (cuasi)varianza muestral $\hat{V}(T^*) = \hat{S}^2_{T^*} = \frac{\sum_{i=1}^B (t_i^* - \bar{t}^*)^2}{B-1}$ donde $\bar{t}^* = \frac{1}{n} \sum t_i^*$
- $T \overset{\sim}{\hookrightarrow} N(\theta, \sigma_T) \implies \text{IC} = T \pm z\hat{\sigma}_T = T(\vec{x}_0) \pm z\hat{\sigma}_T = t_0 \pm z\hat{\sigma}_T$

- $\pm z$ son los cuantiles de órdenes $\frac{\alpha}{2}$ y $1-\frac{\alpha}{2}$ de una gausiana típica, $N(0,1)$
- los extremos del IC son los cuantiles de órdenes $\frac{\alpha}{2}$ y $1-\frac{\alpha}{2}$ de una distribución gausiana $N\left(t_0, \hat{\sigma}_T\right)$

- $T^* \overset{\sim}{\hookrightarrow} N(t_0, \sigma_{T^*}) \implies 1-\alpha = \Pr^*(t_0 - z\hat{\sigma}_{T^*} \leqslant T^* \leqslant t_0 + z\hat{\sigma}_{T^*}) \implies \mathrm{IC}^* = t_0 \pm z\hat{\sigma}_{T^*}$
  - $\Pr_\theta$ indica la probabilidad asociada a F_θ y $\Pr^*$ indica la probabilidad asociada a F_n
  - los extremos del IC* son los cuantiles de órdenes $\frac{\alpha}{2}$ y $1-\frac{\alpha}{2}$ de una distribución gausiana $N\left(t_0, \sigma_{T^*}\right)$, teóricamente
  - los extremos del IC* son los cuantiles de órdenes $\frac{\alpha}{2}$ y $1-\frac{\alpha}{2}$ de la distribución bústrap, en la práctica

7.4. método percentil

- supone hipótesis A
- basado en que
  - $1-\alpha = \Pr(a \leqslant T \leqslant b) \approx \Pr^*(a \leqslant T^* \leqslant b)$
  - T aproximadamente insesgado, luego $a \leqslant \theta \leqslant b$ y
    - $\hat{a} = T_1 = \hat{T}_{\frac{\alpha}{2}} = \widehat{\text{cuantil}}\ \frac{\alpha}{2}$ de T
    - $\hat{b} = T_2 = \hat{T}_{1-\frac{\alpha}{2}} = \widehat{\text{cuantil}}\ 1-\frac{\alpha}{2}$ de T
  - existe una trasformación g tal que $g(T)$ es simétrica centrada en $g(\theta)$; Efron y Tibshirani, An introduction to the bootstrap, 13.3:
    - en inferencia clásica existe un amplio catálogo de trasformaciones para conseguir acercar la distribución de un estimador a la situación ideal $T \hookrightarrow N(\theta, \sigma_T)$
    - el método percentil extiende la utilidad del método gausiano sin necesidad de conocer dicho catálogo
- los extremos del IC* son los cuantiles de órdenes $\frac{\alpha}{2}$ y $1-\frac{\alpha}{2}$ de la distribución bústrap (generalización de la idea del método gausiano)
  - obtener a y b como cuantiles de órdenes respectivos $\frac{\alpha}{2}$ y $1-\frac{\alpha}{2}$ de la distribución de T^*
  - $\mathrm{IC}^* = [a, b]$
- el bústrap percentil no funciona bien con estimadores como $X_{(1)}$ y $X_{(n)}$

7.4.1. teorema

1. enunciado
   - si existe g creciente tal que $g(T) - g(\theta)$ tiene la misma distribución Ψ que $g(T^*) - g(T)$, simétrica respecto al origen,
   - entonces el intervalo «percentil» [a,b] cumple la cobertura probabilística
2. demostración
   - por la simetría, $\Psi(-x) = 1 - \Psi(x)$
   - sea ψ_α el cuantil de orden α de Ψ, luego $\psi_\alpha = -\psi_{1-\alpha}$
   - $1-\alpha = \Pr_\theta\left[-\psi_{1-\frac{\alpha}{2}} \leqslant g(T) - g(\theta) \leqslant \psi_{1-\frac{\alpha}{2}}\right] \approx \Pr^*\left[-\psi_{1-\frac{\alpha}{2}} \leqslant g(T^*) - g(T) \leqslant \psi_{1-\frac{\alpha}{2}}\right]$
   - extremo inferior del intervalo: $\frac{\alpha}{2} = \Pr^*[g(T^*) - g(T) < -\psi_{1-\frac{\alpha}{2}}] = \Pr^*[g(T^*) < g(T) - \psi_{1-\frac{\alpha}{2}}] = \Pr^*\left[T^* < g^{-1}\left(g(T) - \psi_{1-\frac{\alpha}{2}}\right)\right]$ luego $g^{-1}\left(g(T) - \psi_{1-\frac{\alpha}{2}}\right) = F_{T^*}^{-1}\left(\frac{\alpha}{2}\right)$
   - análogamente para el extremo superior: $g^{-1}\left(g(T) + \psi_{1-\frac{\alpha}{2}}\right) = F_{T^*}^{-1}\left(1-\frac{\alpha}{2}\right)$

- $\Pr\left[-\psi_{1-\frac{\alpha}{2}} \leqslant g(T)-g(\theta) \leqslant \psi_{1-\frac{\alpha}{2}}\right] = \Pr\left[-g(T)-\psi_{1-\frac{\alpha}{2}} \leqslant -g(\theta) \leqslant -g(T)+\psi_{1-\frac{\alpha}{2}}\right] =$
  $\Pr\left[g(T)+\psi_{1-\frac{\alpha}{2}} \geqslant g(\theta) \geqslant g(T)-\psi_{1-\frac{\alpha}{2}}\right] = \Pr\left[g^{-1}\big(g(T)+\psi_{1-\frac{\alpha}{2}}\big) \geqslant \theta \geqslant g^{-1}\big(g(T)-\psi_{1-\frac{\alpha}{2}}\big)\right] =$
  $\Pr\left[g^{-1}\big(g(T)-\psi_{1-\frac{\alpha}{2}}\big) \leqslant \theta \leqslant g^{-1}\big(g(T)+\psi_{1-\frac{\alpha}{2}}\big)\right] = \Pr\left[F_{T^*}^{-1}\left(\frac{\alpha}{2}\right) \leqslant g(T)-g(\theta) \leqslant F_{T^*}^{-1}\left(1-\frac{\alpha}{2}\right)\right]$
  $\implies \text{IC} = \left[F_{T^*}^{-1}\left(\frac{\alpha}{2}\right), F_{T^*}^{-1}\left(1-\frac{\alpha}{2}\right)\right]$

3. **EXTRA** necesidad de la simetría

   - sean α_1 y α_2 tales que $\alpha_1+\alpha_2=\alpha$
   - sean c_1 y c_2 tales que $\alpha_1=\Psi(c_1)$; $1-\alpha_2=\Psi(c_2)$
   - $1-\alpha = \Pr_\theta\left[c_1 \leqslant g(T)-g(\theta) \leqslant c_2\right] \approx \Pr^*\left[c_1 \leqslant g(T^*)-g(T) \leqslant c_2\right]$
   - extremo inferior del intervalo: $\frac{\alpha}{2} = \Pr^*[g(T^*)-g(T) < c_1] = \Pr^*[g(T^*) < g(T)+c_1] = \Pr^*\big[T^* < g^{-1}\big(g(T)+c_1\big)\big]$ luego $g^{-1}\big(g(T)+c_1\big) = F_{T^*}^{-1}\left(\frac{\alpha}{2}\right)$
   - análogamente para el extremo superior: $g^{-1}\big(g(T)+c_2\big) = F_{T^*}^{-1}\left(1-\frac{\alpha}{2}\right)$
   - $\Pr\left[c_1 \leqslant g(T)-g(\theta) \leqslant c_2\right] = \Pr\left[-g(T)+c_1 \leqslant -g(\theta) \leqslant -g(T)+c_2\right] =$
     $\Pr\left[g(T)-c_1 \geqslant g(\theta) \geqslant g(T)-c_2\right] = \Pr\left[g^{-1}\big(g(T)-c_1\big) \geqslant \theta \geqslant g^{-1}\big(g(T)-c_2\big)\right] =$
     $\Pr\left[g^{-1}\big(g(T)-c_2\big) \leqslant \theta \leqslant g^{-1}\big(g(T)-c_1\big)\right] = \big\langle$ **aquí necesitamos** $c_1=-c_2$ $\big\rangle =$
     $\Pr\left[F_{T^*}^{-1}\left(\frac{\alpha}{2}\right) \leqslant g(T)-g(\theta) \leqslant F_{T^*}^{-1}\left(1-\frac{\alpha}{2}\right)\right] \implies \text{IC} = \left[F_{T^*}^{-1}\left(\frac{\alpha}{2}\right), F_{T^*}^{-1}\left(1-\frac{\alpha}{2}\right)\right]$

7.4.2. EXTRA corrección de sesgo

- el método percentil suele suponer $\Psi \sim N(0,\sigma)$
- generalización: $\frac{g(T)-g(\theta)}{\sigma} \hookrightarrow N(-z_0,1)$
- z_0 y σ desconocidos ; z_0 estimable:
  - $p = \Pr^*\left[T^* \leqslant T\right] \implies$ estimable mediante bústrap como $\hat{p} = F_{T^*}^{-1}(T)$
  - $p = \Pr^*\left[T^* \leqslant T\right] = \Pr^*\left[\frac{gT^*-gT}{\sigma} \leqslant 0\right] = \Phi(z_0) \implies \hat{z}_0 = \Phi^{-1}\left(\hat{p}\right)$
- $1-\alpha = \Pr\left[z_{\alpha_1} \leqslant \frac{gT-g\theta}{\sigma}+z_0 \leqslant z_{1-\alpha_2}\right] = \Pr\left[g^{-1}\big(gT+\sigma(z_0+z_{\alpha_1})\big) \leqslant \theta \leqslant g^{-1}\big(gT+\sigma(z_0+z_{1-\alpha_2})\big)\right]$
  $\implies \text{I.C.} = \left[g^{-1}\big(gT+\sigma(z_0+z_{\alpha_1})\big)\ ,\ g^{-1}\big(gT+\sigma(z_0+z_{1-\alpha_2})\big)\right]$
- extremo inferior: $\Pr^*\left[T^* \leqslant g^{-1}\big(gT+\sigma(z_0+z_{\alpha_1})\big)\right] = \Pr^*\left[\frac{gT^*-gT}{\sigma}+z_0 \leqslant 2z_0+z_{\alpha_1}\right] =$
  $\Pr^*\left[N(0,1) \leqslant 2z_0+z_{\alpha_1}\right] = \Phi(2z_0+z_{\alpha_1}) \implies g^{-1}\big(gT+\sigma(z_0+z_{\alpha_1})\big) = F_{T^*}^{-1}\big(\Phi(2z_0+z_{\alpha_1})\big)$
- extremo superior, análogamente: $g^{-1}\big(gT+\sigma(z_0+z_{1-\alpha_2})\big) = F_{T^*}^{-1}\big(\Phi(2z_0+z_{1-\alpha_2})\big)$

7.5. método básico

- supone hipótesis B
- basado en que $1-\alpha = \Pr(a \leqslant T-\theta \leqslant b) \approx \Pr^*(a \leqslant T^*-t_0 \leqslant b)$
- obtener a y b como cuantiles de órdenes respectivos $\frac{\alpha}{2}$ y $1-\frac{\alpha}{2}$ de la distribución de T^*-t_0
- $1-\alpha = \Pr(a \leqslant T-\theta \leqslant b) = \Pr(a-T \leqslant -\theta \leqslant b-\theta) = \Pr(T-a \geqslant -\theta \geqslant T-b) = \Pr(T-b \leqslant \theta \leqslant T-a)$
  $\implies \text{IC} = [T-b, T-a] = [t_0-b, t_0-a]$

7.6. método t ó estudentizado

- supone hipótesis C
- étimo: si $X \hookrightarrow N(\theta, \sigma)$, $T = \bar{X}$ y $\hat{D}(T) = \frac{\hat{S}}{\sqrt{n}}$, entonces $\frac{T-\theta}{\hat{D}(T)} \hookrightarrow t_{n-1}$
    - Efron y Tibshirani, en *An introduction to the bootstrap*, 12.4, sugieren que sirve de aproximación para distribuciones arbitrarias: en 1908, Gosset *el Estudiante* derivó la aproximación $\frac{T-\theta}{\hat{D}(T)} \overset{\sim}{\hookrightarrow} t_{n-1}$ para el caso $T = \bar{X}$
- a menudo $\frac{T^*-t_0}{\hat{D}(T^*)}$ es más estable que $T^* - t_0$
- basado en que $1-\alpha = \Pr\left(a \leqslant \frac{T-\theta}{\hat{D}(T)} \leqslant b\right) \approx \Pr^*\left(a \leqslant \frac{T^*-t_0}{\hat{D}(T^*)} \leqslant b\right)$
- para calcular $\hat{D}(T^*)$ requiere expresión explícita de la varianza de T (véase ejemplo siguiente) o usar sobre la muestra bústrap $\vec{x}^*$
    - otro bústrap (anidado)
    - remuestreo herramental
- obtener a y b como cuantiles de órdenes respectivos $\frac{\alpha}{2}$ y $1-\frac{\alpha}{2}$ de la distribución de $\frac{T^*-t_0}{\hat{D}(T^*)}$
- $1-\alpha = \Pr\left(a \leqslant \frac{T-\theta}{\hat{D}(T)} \leqslant b\right) = \Pr\left(T - b\hat{D}(T) \leqslant \theta \leqslant T - a\hat{D}(T)\right) \implies \mathrm{IC} = [T - b\hat{D}(T), T - a\hat{D}(T)] = [t_0 - bd_0, T - ad_0]$
- d_0 se puede obtener como desvío típico de T^* o, si está disponible, a través de la expresión explícita de la varianza de T

7.7. método paramétrico

- los métodos anteriores son **no paramétricos** ; el remuestreo se realiza a partir de la distribución empírica F_n
- en el bústrap **paramétrico** se sustituye el remuestreo F_n por el remuestreo a partir de $F_{\hat{\theta}}$, la supuesta distribución de X con los parámetros sustituidos por estimaciones
- el resto de pasos son los mismos que en bústrap no paramétrico
- ventajas
    - no paramétrico: evita establecer una familia paramétrica de distribuciones para X
    - paramétrico:
        - produce resultados más precisos que las fórmulas clásicas asintóticas y puede usarse en problemas para los que no existe fórmula (Efron y Tibshirani, An introduction to the bootstrap, 6.5)
        - funciona también cuando el estimador es $X_{(1)}$ y $X_{(n)}$ (en tales casos, un remuestreo no paramétrico produciría una mayoría de remuestras con el mismo valor del estimador)

7.8. ejemplo

- $X \hookrightarrow N(\mu, \sigma)$
- se busca IC para σ^2
- $\frac{(n-1)\hat{S}^2}{\sigma^2} \hookrightarrow \chi^2_{n-1} \implies V\left[\frac{(n-1)\hat{S}^2}{\sigma^2}\right] = 2(n-1) \implies V(\hat{S}^2) = \frac{2\sigma^4}{n-1} \implies D(\hat{S}^2) = \sigma^2\sqrt{\frac{2}{n-1}}$

```
a = 0.05 # alfa, 1 - confianza
n = 30 # tamaño muestral
mu = 0
sigma2 = 100 # parámetro de interés
sigma = sqrt(sigma2)
longitudesYcoberturas = function (x0)
{
  T = var # estimador = cuasivarianza = S^2
  t0 = T(x0)
  B = 1e3 # número de muestras bústrap
  distri = replicate (B, # distribuciones bústrap
  {
    xB = sample (x0, replace=TRUE) # muestra bústrap x*
    tB = T(xB)                     # t*
    c (percentil = tB,
       básico = tB - t0,
       t = (tB - t0) / (tB*sqrt(2/(n-1))))
  })
  cuantiles = apply (distri, 1, function (x) quantile(x,c(a/2,1-a/2)))
  Ipercentil = cuantiles[,"percentil"]
  Ibásico = t0 - rev(cuantiles[,"básico"])
  It = t0 - rev(cuantiles[,"t"]) * (t0*sqrt(2/(n-1)))
  intervalos = rbind (Iteórico = (n-1)*t0/qchisq(c(1-a/2,a/2),n-1),
                      Ipercentil, Ibásico, It)
  longitudes = apply (intervalos, 1, diff)
  coberturas = apply (intervalos, 1, function (x) x[1]<=sigma2&sigma2<=x[2])
  cbind (intervalos, longitudes, coberturas)
}
x0 = rnorm (n, mu, sigma)
longitudesYcoberturas(x0)
resul = replicate (1e4,
{
  x0 = rnorm (n, mu, sigma)
  lYc = longitudesYcoberturas(x0)
  list (longitudes = lYc[,"longitudes"],
        coberturas = lYc[,"coberturas"])
})
apply(do.call(rbind,resul["longitudes",]),2,mean)
apply(do.call(rbind,resul["coberturas",]),2,mean)

### ejemplo de ejecución

## > x0 = rnorm (n, mu, sigma)
## > longitudesYcoberturas(x0)
##                2.5%    97.5% longitudes coberturas
## Iteórico   65.36465 186.2409  120.87627          1
## Ipercentil 58.29280 148.3789   90.08614          1
## Ibásico    57.73288 147.8190   90.08614          1
## It         71.57701 182.1941  110.61711          1

## > apply(do.call(rbind,resul["longitudes",]),2,mean)
##   Iteórico Ipercentil    Ibásico         It
```

```
##   117.4148    92.0845    92.0845   117.5555
## > apply(do.call(rbind,resul["coberturas",]),2,mean)
##   Iteórico Ipercentil    Ibásico         It
##     0.9475     0.8811     0.8785     0.9130
```

mismo ejemplo con bústrap paramétrico:

```
a = 0.05 # alfa, 1 - confianza
n = 30 # tamaño muestral
mu = 0
sigma2 = 100 # parámetro de interés
sigma = sqrt(sigma2)
longitudesYcoberturas = function (x0)
{
  T = var # estimador = cuasivarianza = S^2
  t0 = T(x0)
  B = 1e3 # número de muestras bústrap
  muE = mean(x0) ; sigmaE = sd(x0) # PARÁMETROS ESTIMADOS para remuestreo
  distri = replicate (B,           # distribuciones bústrap
  {
    xB = rnorm (n, muE, sigmaE) # muestra bústrap x* PARAMÉTRICA
    tB = T(xB)                  # t*
    c (percentil = tB,
       básico = tB - t0,
       t = (tB - t0) / (tB*sqrt(2/(n-1))))
  })
  cuantiles = apply (distri, 1, function (x) quantile(x,c(a/2,1-a/2)))
  Ipercentil = cuantiles[,"percentil"]
  Ibásico = t0 - rev(cuantiles[,"básico"])
  It = t0 - rev(cuantiles[,"t"]) * (t0*sqrt(2/(n-1)))
  intervalos = rbind (Iteórico = (n-1)*t0/qchisq(c(1-a/2,a/2),n-1),
                      Ipercentil, Ibásico, It)
  longitudes = apply (intervalos, 1, diff)
  coberturas = apply (intervalos, 1, function (x) x[1]<=sigma2&sigma2<=x[2])
  cbind (intervalos, longitudes, coberturas)
}
x0 = rnorm (n, mu, sigma)
longitudesYcoberturas(x0)
resul = replicate (1e4,
{
  x0 = rnorm (n, mu, sigma)
  lYc = longitudesYcoberturas(x0)
  list (longitudes = lYc[,"longitudes"],
        coberturas = lYc[,"coberturas"])
})
apply(do.call(rbind,resul["longitudes",]),2,mean)
apply(do.call(rbind,resul["coberturas",]),2,mean)

### ejemplo de ejecución

## > x0 = rnorm (n, mu, sigma)
## > longitudesYcoberturas(x0)
##               2.5%     97.5% longitudes coberturas
```

```
## Iteórico   40.24300 114.66280   74.41979          1
## Ipercentil 35.33351 100.31189   64.97838          1
## Ibásico    26.58481  91.56319   64.97838          0
## It         40.13180 113.93423   73.80243          1

## > apply(do.call(rbind,resul["longitudes",]),2,mean)
##   Iteórico Ipercentil    Ibásico         It
##   116.7928   101.3657   101.3657   116.1387
## > apply(do.call(rbind,resul["coberturas",]),2,mean)
##   Iteórico Ipercentil    Ibásico         It
##     0.9538     0.9287     0.8893     0.9510
```

8. determinar tamaño muestral

- error ϵ de un intervalo
  - semiamplitud de un intervalo
  - si IC $= T \pm z\sigma_T$, entonces $\epsilon = z\sigma_T$
- objetivo: hallar mínimo tamaño muestral n necesario para
  - obtener un error menor que ϵ
  - dada cierta confianza $1-\alpha$
- $\Pr[\theta \in T \pm z\sigma_T] \geqslant 1-\alpha \implies$ ¿n?

8.1. σ conocida

8.1.1. caso gausiano

- $X \hookrightarrow N(\mu, \sigma)$
- $\theta = \mu = E(X)$
- $T = \bar{X} \implies \text{IC} = \bar{X} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$
- $z_{1-\frac{\alpha}{2}} \sigma \sqrt{n} \leqslant \epsilon \implies n \geqslant \frac{z^2_{1-\frac{\alpha}{2}} \sigma^2}{\epsilon^2}$

8.1.2. caso general

1. TCL
   - como caso gausiano, pero asegurando $n > 30$, 50, 100… según la asimetría de la población
2. desigualdad de Chebichev
   - $\Pr\left[|X-\mu| > k\sigma\right] \leqslant \frac{1}{k^2}$
   - $\Pr\left[|T-\mu_T| \leqslant \epsilon\right] \geqslant \frac{\sigma_T^2}{\epsilon^2} \geqslant 1-\alpha$
   - $\Pr\left[\left|\bar{X}-\mu\right| \leqslant \epsilon\right] \geqslant \frac{\frac{\sigma^2}{n}}{\epsilon^2} \geqslant 1-\alpha \implies n \geqslant \frac{\sigma^2}{\alpha\epsilon^2}$

8.2. σ desconocida

- estimar σ a partir de una muestra piloto
- acotar σ
  - caso gausiano: si $\sigma \leqslant \sigma_0$ entonces $n \geqslant \frac{z^2_{1-\frac{\alpha}{2}}\sigma_0^2}{\epsilon^2}$
  - caso bernuli:
    - $X \hookrightarrow B(1,p)$, n grande $\Longrightarrow$ $\bar{X} \overset{\sim}{\hookrightarrow} N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$ $\Longrightarrow$ $\text{IC} = \bar{X} \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}$ $\Longrightarrow$ $z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} \leqslant \epsilon \Longrightarrow z^2_{1-\frac{\alpha}{2}}\frac{p(1-p)}{n} \leqslant \epsilon^2 \Longrightarrow n \geqslant z^2_{1-\frac{\alpha}{2}}\frac{p(1-p)}{\epsilon^2}$
    - $\forall p \in (0,1)$, $p(1-p) \leqslant \frac{1}{4} \Longrightarrow$ tomar $n \geqslant z^2_{1-\frac{\alpha}{2}}\frac{1}{4\epsilon^2}$

```
unoMenosAlfas <- c (.9, .95, .99)       #confianzas
epsilones <- c (.05, .02, .01, 0.001)   #errores
tabla <- outer (unoMenosAlfas, epsilones,
                function (alfa1, eps)
                {
                    alfa <- 1 - alfa1
                    z <- qnorm (1-alfa/2)
                    ceiling (z^2 / (4 * eps^2))
                })
dimnames(tabla) <- list (conf=unoMenosAlfas, err=epsilones)
tabla
```

```
      err
conf    0.05 0.02  0.01   0.001
  0.9    271 1691  6764  676386
  0.95   385 2401  9604  960365
  0.99   664 4147 16588 1658725
```

8.3. consecuencias

n crece si

- aumenta la confianza $1-\alpha$
- disminuye el error ϵ
- aumenta la dispersión σ