Métodos de Remuestreo

Tema 5. Estimación de sesgos mediante remuestreo y método Jackknife

basado en

B. Efron, R. Tibshirani (1993). An Introduction to the bootstrap.

O. Kirchkamp (2017). Resampling methods.

Curso 2018/19

Introducción

- Aparte del error estándar, existen otras medidas de precisión de los estimadores, como el **sesgo**, es decir, la diferencia entre la esperanza de un estimador $\hat{\theta}$ y el valor real del parámetro θ.
- El bootstrap es una técnica adecuada para estimar sesgos, aunque la técnica de ***jackknife*** también es útil en este caso.
- Se puede usar un estimador del sesgo para corregirlo aunque a veces no es una práctica demasiado efectiva.

Introducción

- El sesgo de un estimador $\hat{\theta} = s(\mathbf{x})$ de un parámetro $\theta = t(F)$ se define como

$$\boxed{\mathrm{Sesgo}_F(\hat{\theta}, \theta) = \mathrm{E}_F\left[s(\mathbf{x})\right] - t(F)}$$

- Es inevitable que aparezca variabilidad en el estimador de θ pero resulta mucho peor que esta variabilidad esté localizada en alguna parte concreta.

- Los estimadores insesgados tales que $\mathrm{E}_F(\hat{\theta}) = \theta$ resultan ser bastante importantes en la práctica.

- Los estimadores *plug-in* del tipo $\hat{\theta} = t(\hat{F})$ no son necesariamente insesgados pero suelen tener sesgos pequeños en comparación con sus errores estándar.

Bootstrap para el sesgo

- Se puede usar el bootstrap para calcular el sesgo de cualquier estimador $\hat{\theta} = s(\mathbf{x})$.
- El estimador bootstrap del sesgo se define como el estimador del sesgo donde se sustituye F por $\hat{F}$

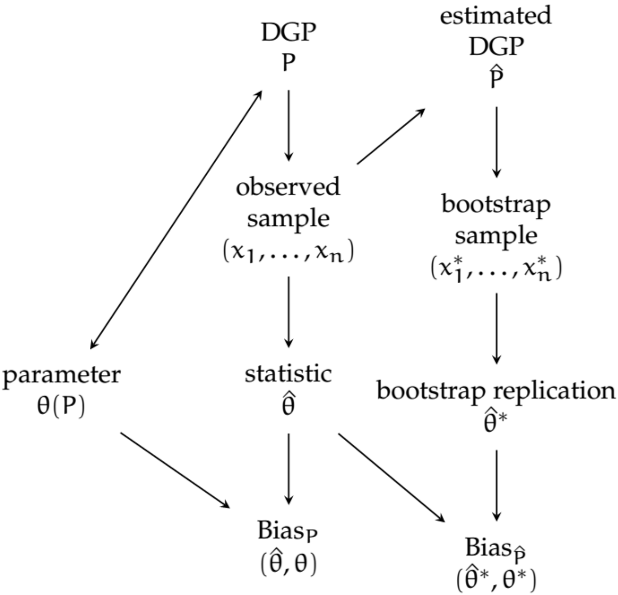
$$\mathrm{Sesgo}_{\hat{F}} = E_{\hat{F}}\left[s(\mathbf{x})\right] - t(\hat{F})$$

- En este caso $t(\hat{F})$ es el estimador *plug-in* de θ.
- En la práctica, el estimador bootstrap del sesgo basado en B réplicas es

$$\boxed{\widehat{\mathrm{Sesgo}}_B = \hat{\theta}^*(\cdot) - t(\hat{F})}$$

donde

$$\hat{\theta}^*(\cdot) = \sum_{b=1}^{B} \frac{\hat{\theta}^*(b)}{B}$$


DGP
P
estimated
DGP
$\hat{P}$
observed
sample
$(x_1, \ldots, x_n)$
bootstrap
sample
$(x_1^*, \ldots, x_n^*)$
parameter
$\theta(P)$
statistic
$\hat{\theta}$
bootstrap replication
$\hat{\theta}^*$
Bias_P
$(\hat{\theta}, \theta)$
$\text{Bias}_{\hat{P}}$
$(\hat{\theta}^*, \theta^*)$

Ejemplo

- Se tiene una serie de datos sobre el efecto de unos parches hormonales sobre 8 personas. Dichos parches difunden un medicamento en la sangre.

- Se mide el nivel de la hormona que aparece después de usar tres parches diferentes: un parche *placebo* (sin hormona), un parche *viejo* y uno *nuevo*.

- Se trata de estudiar su *bioequivalencia*.

- El criterio que se utiliza en la agencia estatal norteamericana de medicamentos (*FDA*) es que

$$\frac{|E(nuevo) - E(viejo)|}{E(viejo) - E(placebo)} \leq 0{,}20$$

Ejemplo

- Es decir, la *FDA* exige que el nuevo tipo de parche se ajuste a la cantidad de hormona que liberaba el antiguo (respecto al placebo) en no más del 20 %.
- Se denomina como
  $\mathbf{z} \equiv$ (medida parche viejo $-$ medida placebo)
  $\mathbf{y} \equiv$ (medida parche nuevo $-$ medida parche viejo)
- Se asume que las parejas $x_i = (z_i, y_i)$ se obtienen a partir de una distribución bivariante $F \to \mathbf{x} = (x_1, x_2, \ldots, x_8)$.

Ejemplo

- El parámetro en este caso es entonces

$$\theta = t(F) = \frac{E_F(y)}{E_F(z)} = \frac{E(nuevo) - E(viejo)}{E(viejo) - E(placebo)}$$

- En este caso $t(\cdot)$ es una función que tiene como objeto las parejas de $\mathbf{x}$ y da como resultado el ratio de las esperanzas.

- El estimador *plug-in* de θ es

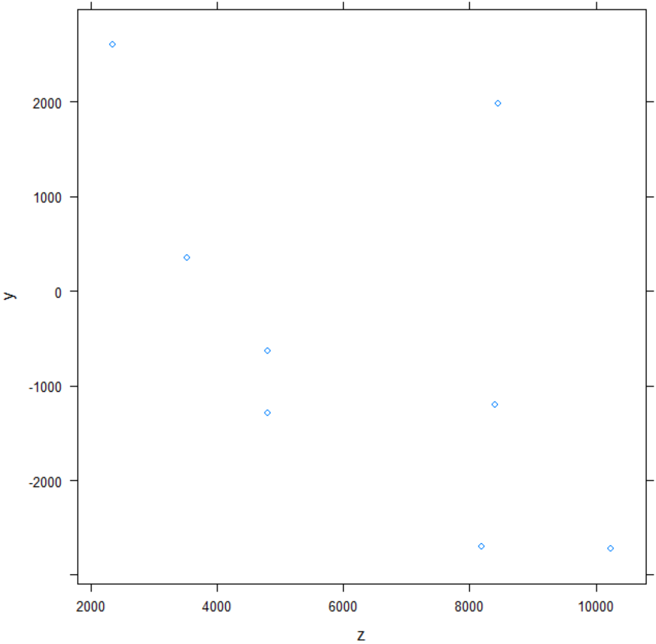
$$\hat{\theta} = t(\hat{F}) = \frac{\overline{y}}{\overline{z}} = \frac{\sum_i y_i/8}{\sum_i z_i/8}$$

Ejemplo

- En el ejemplo,

$$\frac{E(nuevo) - E(viejo)}{E(viejo) - E(placebo)} = \frac{E(y)}{E(z)} \approx \frac{\overline{y}}{\overline{z}} = \frac{-452{,}25}{6342{,}375} \approx -0{,}0713$$

```r
data(patch, package="bootstrap")
lattice::xyplot(y ~ z, data=patch)
```


2000
1000
0
-1000
-2000
y
2000
4000
6000
8000
10000
z

Ejemplo

- Si se aplica directamente la librería `bootstrap` se tiene que:

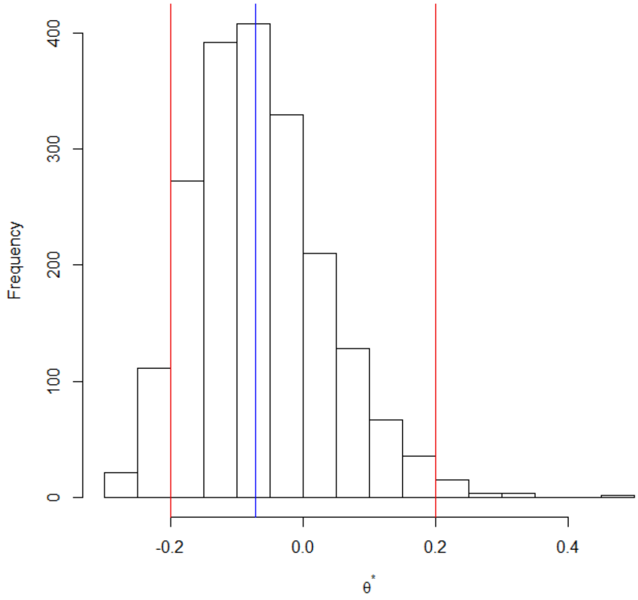
```
teta = function(ind){
      Y = patch[ind,"y"]
      Z = patch[ind,"z"]
      mean(Y)/mean(Z) }

patch.boot = bootstrap(1:8, 2000, teta)
names(patch.boot)
```

```
(ratio = teta(1:8))  # estimador plug-in original
```

```
[1] -0.0713061
```

```
hist(patch.boot$thetastar, xlab=expression(theta^"*"),
main="")
abline(v=c(-0.2, 0.2), col="red2")
abline(v=ratio, col="blue")
```


Frequency
0
100
200
300
400
-0.2
0.0
0.2
0.4
θ*

Ejemplo

- Se calcula el estimador bootstrap del sesgo:

```
# Estimador bootstrap del sesgo (bias)
mean(patch.boot$thetastar) - ratio
```

```
[1] 0.006098727
```

```
# Se obtine el error estandar bootstrap
sd(patch.boot$thetastar)
```

```
[1] 0.1005338
```

Ejemplo

- Alternativamente:

```
N = dim(patch)[1]
B = 5000

thetaHatBoot = replicate(B,
{ind = sample(1:N, replace=TRUE) ;
with(patch[ind,], mean(y)/mean(z))})

# Estimador plug-in original
thetaHat = with(patch, mean(y)/mean(z))

# Estimador bootstrap del sesgo (bias)
(bias = mean(thetaHatBoot) - thetaHat)
```

```
[1] 0.009508933
```

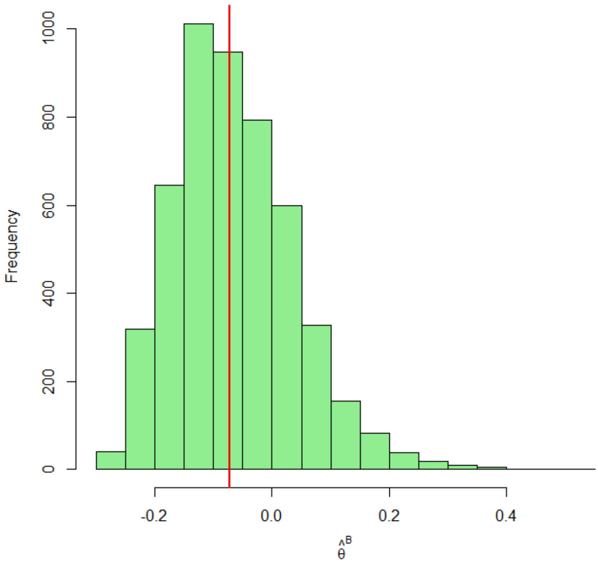
Ejemplo

```
print(paste("Probabilidad que theta.hat > 0.2 =",
sum(thetaHatBoot>0.2)/B))
```

```
[1] Probabilidad que theta.hat > 0.2 = 0.19
```

```
hist(thetaHatBoot,
     main=paste("Sesgo=",
     round(abs(thetaHat-mean(thetaHatBoot)),2),";
     Stdev=", round(sd(thetaHatBoot),2)),
     col='lightgreen', xlab=expression(hat(theta)^B))
abline(v=thetaHat, col="red", lwd=2)
```

Sesgo= 0.01 ; Stdev= 0.1



Estimador del sesgo mejorado

- Se presenta un problema en el cálculo directo del sesgo:

$$\mathrm{Sesgo}_{\hat{F}} = \widehat{\mathrm{Sesgo}}_B = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^{*b} - s(\mathbf{x})$$

- Ya que en $\frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^{*b}$ puede que algunas observaciones aparezcan más frecuentemente que otras, mientras que en $s(\mathbf{x})$ todas las observaciones en $\mathbf{x}$ tienen el mismo peso.

- Sería conveniente, entonces, hacer una correción para considerar los distintos pesos.

Estimador del sesgo mejorado

- Se considera la noción de ***vector de remuestreo***.
- Se denomina como P_j^* a la proporción de una muestra bootstrap $\mathbf{x}^* = (x_1^*, x_2^*, \ldots, x_n^*)$ que es igual al dato j-esimo original:

$$P_j^* = \frac{\#\left\{x_i^* = x_j\right\}}{n}$$

para $j = 1, 2, \ldots, n$ y cada $i = 1, 2, \ldots, n$.

- El *vector de remuestreos*

$$\mathbf{P}^* = (P_1^*, P_2^*, \ldots, P_n^*)$$

tiene componentes no negativos que suman 1.

Estimador del sesgo mejorado

- Por ejemplo, en el caso del los datos de los parches, si una muestra bootstrap fuera $\mathbf{x}^* = (x_1, x_6, x_6, x_5, x_7, x_1, x_3, x_8)$ el vector de remuestreos sería

$$\mathbf{P}^* = \left(\frac{2}{8}, 0, \frac{1}{8}, 0, \frac{1}{8}, \frac{2}{8}, \frac{1}{8}, \frac{1}{8}\right)$$

- Una réplica bootstrap $\hat{\theta}^* = s(\mathbf{x}^*)$ se puede considerar una función del vector de remuestreos $\mathbf{P}^*$.

- Por ejemplo en el caso de $\hat{\theta} = \overline{y}/\overline{z}$

$$\hat{\theta}^* = \frac{\overline{y}^*}{\overline{z}^*} = \frac{\sum_{j=1}^{8} P_j^* y_j}{\sum_{j=1}^{8} P_j^* z_j}$$

Estimador del sesgo mejorado

- Con esta notación, los datos $\mathbf{x}$ se consideran fijos y se considera que las únicas cantidades aleatorias son los P_j^*.
- De este modo se escribe

$$\hat{\theta}^* = T\left(\mathbf{P}^*\right)$$

para indicar que $\hat{\theta}^*$ es una función del vector de remuestreos.

- Se denomina como $\mathbf{P}^0$ el vector de longitud n tal que todos sus elementos son iguales a $1/n$

$$\mathbf{P}^0 = \left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)$$

Estimador del sesgo mejorado

- El valor de $T\left(\mathbf{P}^0\right)$ es el valor de $\hat{\theta}^*$ cuando cada $P_j^* = \frac{1}{n}$, es decir cuando cada dato original x_j ocurre exactamente una vez en la muestra bootstrap $\mathbf{x}^*$.
- Pero esto implica que $\mathbf{x}^* = \mathbf{x}$ excepto en las permutaciones de los elementos. Así el valor del estadístico original es igual al valor observado del estadístico:

$$T\left(\mathbf{P}^0\right) = \hat{\theta} = t(\hat{F})$$

Estimador del sesgo mejorado

- Las B muestras bootstrap $\mathbf{x}^{*1}, \mathbf{x}^{*2}, \ldots, \mathbf{x}^{*B}$ dan lugar a los correspondientes vectores de remuestreo $\mathbf{P}^{*1}, \mathbf{P}^{*2}, \ldots, \mathbf{P}^{*B}$.

- Se define como $\overline{\mathbf{P}}^{*}$ la media de los vectores anteriores

$$\overline{\mathbf{P}}^{*} = \frac{1}{B} \sum_{b=1}^{B} \mathbf{P}^{*b}$$

- De este modo, el mejor estimador bootstrap del sesgo es

$$\boxed{\overline{\mathrm{Sesgo}}_B = \hat{\theta}^{*}(\cdot) - T\left(\overline{\mathbf{P}}^{*}\right)}$$

donde

$$\hat{\theta}^{*}(\cdot) = \sum_{b=1}^{B} \frac{\hat{\theta}^{*}(b)}{B}$$

Estimador del sesgo mejorado

- En el ejemplo de los parches se remuestrean 400 vectores y se obtiene

$$\overline{\mathbf{P}}^* = (0{,}1178; 0{,}1187; 0{,}1313; 0{,}1259; 0{,}1219; 0{,}1275; 0{,}1306; 0{,}1213)$$

- De modo que

$$T\left(\overline{\mathbf{P}}^*\right) = \frac{\sum_{j=1}^{8} \overline{P}_j^* y_j}{\sum_{j=1}^{8} \overline{P}_j^* z_j} = -0{,}075$$

- con lo que

$$\begin{aligned} \overline{\mathrm{Sesgo}}_B &= \hat{\theta}^*(\cdot) - T\left(\overline{\mathbf{P}}^*\right) = \\ &= -0{,}067 - (-0{,}075) = 0{,}0080 \end{aligned}$$

Correción del sesgo

```
z = patch$oldpatch - patch$placebo
y = patch$newpatch - patch$oldpatch
n = length(y)

otramedia = function(p, x) sum(x*p)

elratio = function(p, x, y){
   otramedia(p, x)/otramedia(p, y)
}

p.hat = rep(1/n, n)

# Estimador plug-in de teta
(theta.hat = elratio(p.hat, y, z))
```

```
[1] -0.0713061
```

Correción del sesgo

```
nboot = 1000
theta.star = double(nboot)

p.bar = rep(0, n)

for (i in 1:nboot){
    k.star = sample(n, replace=TRUE)
    p.star = tabulate(k.star, n)/n
    theta.star[i] = elratio(p.star, y, z)
    p.bar = p.bar + p.star
}

p.bar = p.bar/nboot
(theta.bar = elratio(p.bar, y, z))
```

```
[1] -0.07330194
```

Correción del sesgo

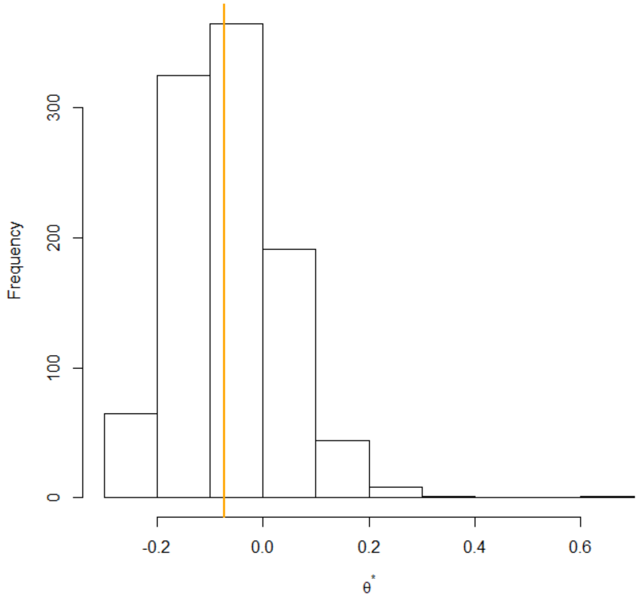
```r
hist(theta.star, xlab=expression(theta^"*"), main="")
abline(v=theta.bar, col="orange", lwd=2)

# Estimador bootstrap del sesgo mejorado
(bias.hat = mean(theta.star) - theta.bar)
```

```
[1] 0.008591554
```

```r
# Estimador de teta corregido del sesgo
theta.hat - bias.hat
```

```
[1] -0.07989765
```


Frequency
0
100
200
300
-0.2
0.0
0.2
0.4
0.6
θ*

Correción del sesgo

- ¿Para qué se quiere estimar el sesgo de un estimador $\hat{\theta}$?
  La razón original es para corregir el estimador $\hat{\theta}$ de modo que sea menos sesgado.

- Si el sesgo estimado bootstrap es

$$\widehat{\mathrm{sesgo}}_B = \hat{\theta}^*(\cdot) - \hat{\theta}$$

- Entonces el estimador corregido del sesgo, digamos $\overline{\theta}$, es

$$\begin{aligned} \overline{\theta} &= \hat{\theta} - \widehat{\mathrm{sesgo}}_B \\ &= 2\hat{\theta} - \hat{\theta}^*(\cdot) \end{aligned}$$

Correción del sesgo

- Con esto se dice que si $\hat{\theta}^*(\cdot)$ es mayor que $\hat{\theta}$ entonces el estimador de sesgo corregido $\overline{\theta}$ debe ser menor que $\hat{\theta}$.
- La correción del sesgo a veces no es conveniente, dado que la estimación del sesgo puede tener una variabilidad alta, con cual, el estimador corregido tendrá también un error estándar alto.

Jackknife

- Dado un conjunto de datos $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ la muestra i-esima $\mathbf{x}_{(i)}$ se define como el conjunto de datos original menos la observación i-esima

  $$\mathbf{x}_{(i)} = (x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$$

  para $i = 1, 2, \ldots, n$.

- La réplica i-esima $\hat{\theta}_{(i)}$ del estadístico $\theta = s(\mathbf{x})$ es el valor de $s(\cdot)$ evaluado en $\mathbf{x}_{(i)}$

  $$\hat{\theta}_{(i)} = s\left(\mathbf{x}_{(i)}\right)$$

  para $i = 1, 2, \ldots, n$.

Jackknife

- En el caso de los estimadores *plug-in* $\hat{\theta} = t(\hat{F})$, $\hat{\theta}_{(i)}$ es igual a $t(\hat{F}_{(i)})$ donde $\hat{F}_{(i)}$ es la distribución empírica sin la observación *i*-esima

- El estimador jackknife del sesgo se define como

$$\boxed{\widehat{\text{Sesgo}}_{jack} = (n-1)\left(\hat{\theta}_{(\cdot)} - \hat{\theta}\right)}$$

- donde

$$\hat{\theta}_{(\cdot)} = \frac{\sum_{i=1}^{n} \hat{\theta}_{(i)}}{n}$$

```
data(patch, package="bootstrap")
n = nrow(patch)
y = patch$y
z = patch$z
theta.hat = mean(y) / mean(z)
```

Jackknife

```
theta.hat
```

```
[1] -0.0713061
```

```
# Jackknife
jacks = sapply(1:n,
function(i) with(patch[-i,], mean(y)/mean(z)))
(biasJack = (n-1)*(mean(jacks) - theta.hat))
```

```
[1] 0.008002488
```

```
# Alternativamente
theta.jack = numeric(n)
for (i in 1:n){
  theta.jack[i] = mean(y[-i]) / mean(z[-i])
}
(sesgo = (n-1)*(mean(theta.jack) - theta.hat))
```

```
[1] 0.008002488
```

Jackknife

- El jackknife funciona bien si el estadístico es *suave* (smooth). Por ejemplo, la mediana **no** lo es...

- Se puede definir también el estimador jackknife del error estándar

$$\hat{se}_{jack} = \sqrt{\frac{n-1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_{(i)} - \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{(i)}\right)^2}$$

```r
se = sqrt((n-1) *
mean((theta.jack - mean(theta.jack))^2))
se
```

```
[1] 0.1055278
```

Ejemplo jackknife

- Se puede considerar un ejemplo simulado respecto al cálculo de correlaciones.

```
library(bootstrap)
datos = matrix(rnorm(30), ncol=2)
n = 15
teta = function(ind, datos){
   cor(datos[ind,1], datos[ind,2])
}

results = jackknife(1:n, teta, datos)
results
```

```
$jack.se
[1] 0.4306432

$jack.bias
[1] 0.04536305
```