Tema 6. Estimación de sesgos mediante remuestreo

basado en
B. Efron, R. Tibshirani (1993). An Introduction to the bootstrap.
O. Kirchkamp (2019). Resampling methods.

Curso 2022/2023

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 $\hat{\theta}$ pero resulta inconveniente 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 $\hat{F}$

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

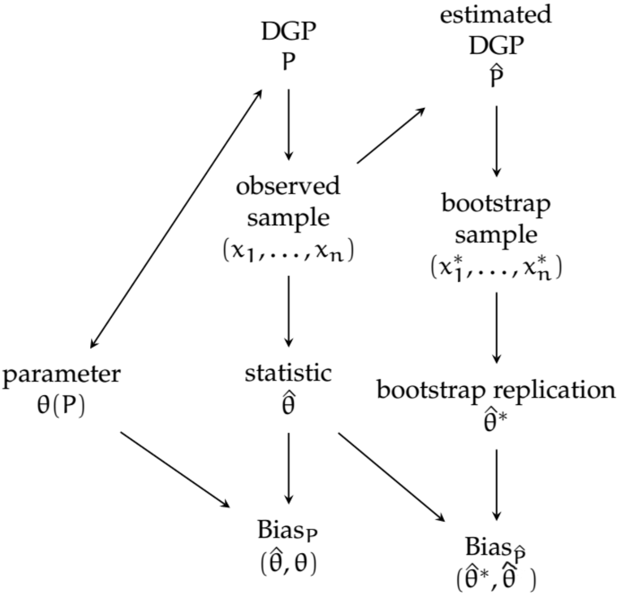
  donde $t(\hat{F})$ es el estimador *plug-in* de θ.

- 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
$(\hat{\theta}, \theta)$
$\mathrm{Bias}_{\hat{P}}$
$(\hat{\theta}^*, \hat{\theta})$

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 $F \rightarrow \mathbf{x} = (x_1, x_2, \ldots, x_8)$.

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

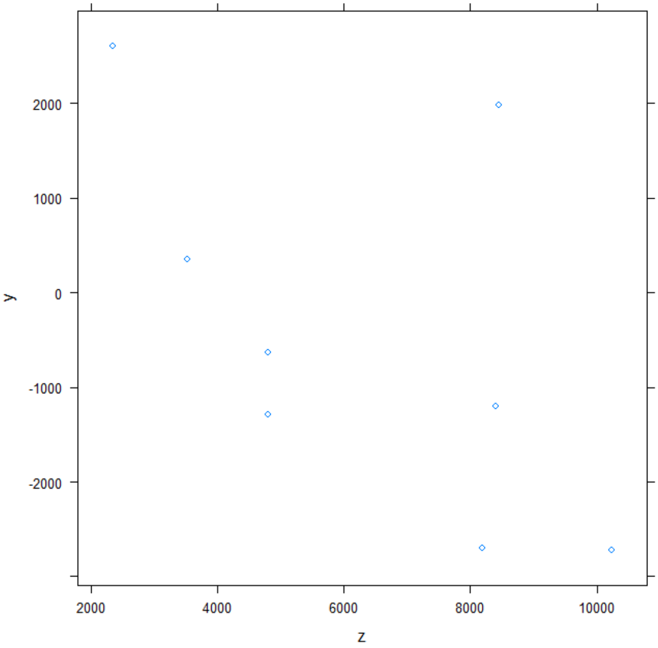
$$\hat{\theta} = t(\hat{F}) = \frac{\overline{y}}{\overline{z}} = \frac{\sum_i y_i/8}{\sum_i z_i/8}$$

Ejemplo

- En el ejemplo,

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

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


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

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 *corrección* 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:

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

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

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

Corrección 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

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

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

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

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 corrección 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.