



UNL • FACULTAD DE  
INGENIERÍA QUÍMICA



# *Bootstrap – parte 1*

---

Curso de Estadística – MADATOP 2026

Andrea Bergesio – Fabricio Chiappini – Stefania D'lorio

- Datos:  $x_1, x_2, \dots, x_n$
- Representación de una muestra aleatoria:  $X_1, X_2, \dots, X_n$
- Modelo:  $X_i \sim F, i = 1, \dots, n$ , independientes
- Problema:  $F = ?$
- Frecuentemente, se supone  $X_i \sim F_\theta, i = 1, \dots, n$ , con  $\theta \in \Theta \subset \mathbb{R}^k$ .
- Hallar estimadores de  $\theta$ .

- La técnica jackknife fue desarrollada por Quenouille y Tukey, interesados en estimar sesgo y variabilidad de un estimador.
  - $\text{Var}_{\theta}(\hat{\theta}) = E_{\theta} \left( \left( \hat{\theta} - E_{\theta}(\hat{\theta}) \right)^2 \right) = E_{\theta} \left( \hat{\theta}^2 \right) - E_{\theta}^2 \left( \hat{\theta} \right)$
  - $\text{bias}_{\theta} = E_{\theta}(\hat{\theta}) - \theta$
- El estimador jackknife de un parámetro se obtiene eliminando sistemáticamente cada observación de un conjunto de datos, calculando la estimación del parámetro sobre las observaciones restantes y luego combinando estas observaciones.

- **Ejemplo 1.** Si el parámetro a estimar es la media poblacional de una variable aleatoria  $X$ , entonces para una muestra aleatoria  $X_1, \dots, X_n$ , el estimador natural es la media muestral:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Se procede como sigue: para cada  $i \in \{1, 2, \dots, n\}$ , calculamos la media  $\bar{X}_{(i)}$  de la submuestra jackknife formada por todos los datos excepto el  $i$ -ésimo:

$$\bar{X}_{(i)} = \frac{1}{n-1} \sum_{j=1, j \neq i}^n X_j, \quad i = 1, \dots, n.$$

Estas  $n$  réplicas jackknife  $\bar{X}_{(1)}, \dots, \bar{X}_{(n)}$  aproximan la distribución de la media muestral  $\bar{X}$ . Finalmente, el estimador jackknife se obtiene promediando estas réplicas:

$$\bar{X}_{(.)} = \frac{1}{n} \sum_{i=1}^n \bar{X}_{(i)}.$$

# Jackknife - Ejemplo 1

En el caso particular de la media, puede demostrarse que

$$\bar{X}_{(.)} = \frac{1}{n} \sum_{i=1}^n \bar{X}_{(i)} = \bar{X},$$

y como  $E[\bar{X}_{(.)}] = E[\bar{X}] = E[X]$ ,  $\bar{X}_{(.)}$  es insesgado.

Además,

$$\text{Var}(\bar{X}_{(.)}) = \text{Var}(\bar{X}) = \frac{\text{Var}(X)}{n}.$$

Sin embargo, estas propiedades no se cumplen en general cuando se desea estimar otros parámetros.

## Jackknife – Ejemplo 1

El estimador  $\bar{X}_{(.)}$  puede utilizarse para construir una estimación empírica del sesgo de  $\bar{X}$ :

$$\widehat{\text{bias}}(\bar{X})_{\text{jack}} = c (\bar{X}_{(.)} - \bar{X}),$$

con  $c > 0$ . En este caso, como  $\bar{X}_{(.)} = \bar{X}$ , el sesgo estimado es cero.

Una estimación jackknife de la varianza de  $\bar{X}$  puede calcularse a partir de la varianza de las réplicas:

$$\widehat{\text{Var}}(\bar{X})_{\text{jack}} = \frac{n-1}{n} \sum_{i=1}^n (\bar{X}_{(i)} - \bar{X}_{(.)})^2 = \frac{1}{n(n-1)} \sum_{i=1}^n (X_i - \bar{X})^2.$$

# Jackknife

En general, para un estimador  $\hat{\theta}$  se puede demostrar que

- el estimador jackknife para el sesgo  $\hat{\theta}$  de está dado por

$$\widehat{\text{bias}}(\hat{\theta})_{\text{jack}} = (n - 1) \left( \bar{\hat{\theta}}_{(\cdot)} - \hat{\theta} \right),$$

- el estimador jackknife de la varianza de  $\hat{\theta}$  resulta

$$\widehat{\text{Var}}(\hat{\theta})_{\text{jack}} = \frac{n - 1}{n} \sum_{i=1}^n \left( \hat{\theta}_{(i)} - \bar{\hat{\theta}}_{(\cdot)} \right)^2,$$

donde  $\hat{\theta}_{(i)}$  es el estimador calculado eliminando la  $i$ -ésima observación y

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

## Jackknife – Ejemplo 2

Un estimador jackknife para  $\theta$  se puede definir como

$$\begin{aligned}\hat{\theta}_{\text{jack}} &= \hat{\theta} - \widehat{\text{bias}}(\hat{\theta})_{\text{jack}} \\ &= \hat{\theta} - (n-1) \left( \bar{\hat{\theta}}_{(\cdot)} - \hat{\theta} \right) \\ &= n\hat{\theta} - (n-1)\bar{\hat{\theta}}_{(\cdot)}\end{aligned}$$

- **Ejemplo 2.** Parámetro de interés:  $\theta = \max_F$ ,  
 $X_1, X_2, X_3 \sim F$  independientes.  
Datos: 1, 2, 100  
Estimación:  $\hat{\theta} = 100$   
Estimador jackknife: ?



*The Annals of Statistics*  
1979, Vol. 7, No. 1, 1-26

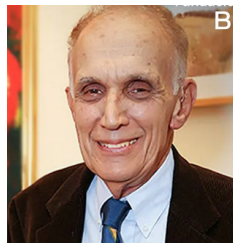
## THE 1977 RIETZ LECTURE

### BOOTSTRAP METHODS: ANOTHER LOOK AT THE JACKKNIFE

BY B. EFRON

*Stanford University*

We discuss the following problem: given a random sample  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  from an unknown probability distribution  $F$ , estimate the sampling distribution of some prespecified random variable  $R(\mathbf{X}, F)$ , on the basis of the observed data  $\mathbf{x}$ . (Standard jackknife theory gives an approximate mean and variance in the case  $R(\mathbf{X}, F) = \theta(\hat{F}) - \theta(F)$ ,  $\theta$  some parameter of interest.) A general method, called the “bootstrap,” is introduced, and shown to work satisfactorily on a variety of estimation problems. The jackknife is shown



# Función de distribución acumulada

## Función de distribución acumulada:

Sea  $X$  una variable aleatoria,

$$F(t) = P(X \leq t) = E [I_{(-\infty, t]}(X)].$$

## Función de distribución acumulada empírica:

Sean  $X_1, X_2, \dots, X_n$  i.i.d.,  $X_i \sim F$ ,

$$\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, t]}(X_i).$$

- $\hat{F}_n(t)$  es una función aleatoria.
- Representa la CDF de una variable discreta que asigna masa  $1/n$  a cada  $X_1, \dots, X_n$ .

- Para cada  $t$  fijo, por la Ley de los Grandes Números:

$$\hat{F}_n(t) \xrightarrow{c.s.} F(t).$$

- **Teorema de Glivenko–Cantelli:**

Sea  $X_1, \dots, X_n$  una muestra i.i.d. con función de distribución  $F$ , y sea  $\hat{F}_n$  la función de distribución empírica.

$$\sup_{t \in \mathbb{R}} \left| \hat{F}_n(t) - F(t) \right| \xrightarrow{c.s.} 0$$

Un funcional  $T(F)$  es una aplicación que se calcula sobre funciones de distribución.

Muchas veces un parámetro de interés puede escribirse como:

$$\theta = T(F)$$

- **Media:**

$$T(F) = E_F(X) = \int t \, dF(t)$$

- **Varianza:**

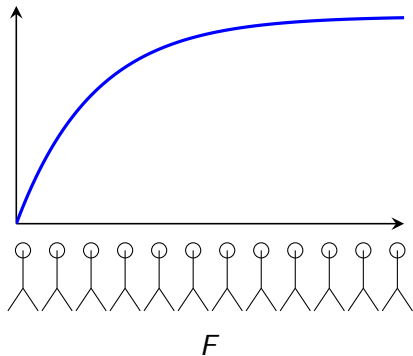
$$T(F) = E_F[(X - E_F(X))^2] = \int t^2 \, dF(t) - \left( \int t \, dF(t) \right)^2$$

- **Mediana:**

$$T(F) = F^{-1}(0.5) = \inf\{x \in \mathbb{R} : F(x) \geq 0.5\}$$

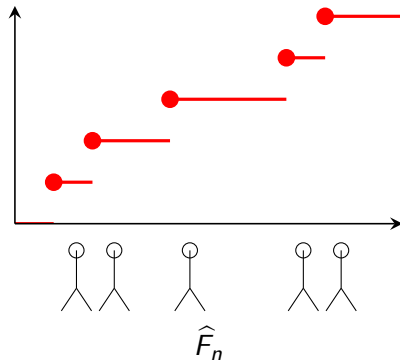
# El método plug-in como fábrica de estimadores

Mundo Ideal



$$\theta = T(F)$$

Mundo Real



$$\hat{\theta} = T(\hat{F}_n)$$

## Planteo del problema

Supongamos que deseamos estimar  $\theta = T(F)$  basándonos en una muestra  $X_1, X_2, \dots, X_n$  iid, cada una con distribución  $F$ .

Sea  $\hat{\theta} = h(X_1, \dots, X_n)$  un estimador de  $\theta$ .

Para hacer inferencia, nos interesa conocer la distribución de  $\hat{\theta}$ :  $G_F(u) = P_F(\hat{\theta} \leq u)$ , o alguna característica de su distribución.

En general, estas cantidades no son fáciles de obtener. En particular, la distribución de  $\hat{\theta}$  suele ser complicada.

## Idea

Una opción es estimar  $G_F(u)$ , o alguna característica como por ejemplo la  $\text{Var}_F(\hat{\theta})$ .

## Dificultad

Solo contamos con una muestra, es decir, observamos un único valor de  $\hat{\theta}$ .

# Bootstrap ideal

## Pregunta

¿Cómo podemos estimar  $G_F$ ?

## Idea: estimador plug-in

El estimador bootstrap ideal de  $G_F(u)$  es el estimador plug-in:

$$\hat{G}_F(u) = G_{\hat{F}}(u),$$

donde  $\hat{F}$  es la distribución empírica (por simplicidad, omitimos el subíndice  $n$ ).

## ¿Qué significa hacer plug-in?

Teníamos,

$$G_F(u) = P_F(\hat{\theta} \leq u) = P_F(h(X_1, \dots, X_n) \leq u)$$

Luego, definimos

$$\hat{G}_F(u) = P_{\hat{F}}(\hat{\theta}^* \leq u) = P_{\hat{F}}(h(X_1^*, \dots, X_n^*) \leq u),$$

donde  $X_1^*, \dots, X_n^*$  son iid con distribución  $\hat{F}$ .

El asterisco indica que ahora la distribución poblacional es  $\hat{F}$  (en lugar de  $F$ ).

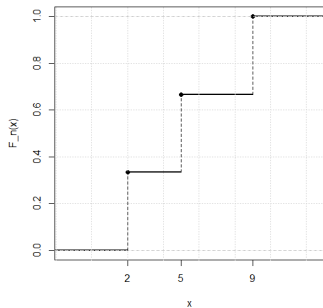
# Ejemplo 1. Varianza de la mediana muestral

## Objetivo

Calcular el estimador bootstrap **ideal** de la varianza de la mediana muestral.

Sean  $\theta = \text{med}_F(X)$  y  $\hat{\theta} = \text{med}_{\hat{F}}(X)$ .

Supongamos  $n = 3$  y observamos:  $x_1 = 5$ ,  $x_2 = 2$ ,  $x_3 = 9$ .



$$\hat{\theta} = \inf\{x : \hat{F}(x) \geq 0.5\} = 5$$



# Ejemplo 1. Varianza de la mediana muestral

Nos interesa estimar:

$$\text{Var}_F(\hat{\theta})$$

El estimador bootstrap ideal es:

$$\text{Var}_{\hat{F}_n}(\hat{\theta}^*)$$

## Clave

Si los datos tienen distribución  $\hat{F}$ , entonces  $\hat{\theta}^*$  es una variable aleatoria discreta que solo puede tomar los valores:

5, 2, 9.

## Idea

En este caso simple podemos calcular explícitamente la distribución de  $\hat{\theta}^*$ .

## Ejemplo 1. Varianza de la mediana muestral

Hay  $3^3 = 27$  posibles ternas ordenadas  $(X_1^*, X_2^*, X_3^*)$  con valores en  $\{2, 5, 9\}$ . Cada una ocurre con probabilidad  $1/27$ .

### Resultados posibles

Muestra	$\hat{\theta}^*$ (mediana)	Probabilidad
$\{2, 5, 9\}$	5	$6/27$
$\{2, 2, 5\}$	2	$3/27$
$\{2, 2, 9\}$	2	$3/27$
$\{5, 5, 2\}$	5	$3/27$
$\{5, 5, 9\}$	5	$3/27$
$\{9, 9, 2\}$	9	$3/27$
$\{9, 9, 5\}$	9	$3/27$
$\{2, 2, 2\}$	2	$1/27$
$\{5, 5, 5\}$	5	$1/27$
$\{9, 9, 9\}$	9	$1/27$

## Ejemplo 1: Varianza de la mediana muestral

### Distribución de la mediana bootstrap

$$P_{\hat{F}}(\hat{\theta}^* = 2) = \frac{1}{27}(3 + 3 + 1) = \frac{7}{27}$$

$$P_{\hat{F}}(\hat{\theta}^* = 5) = \frac{1}{27}(6 + 3 + 3 + 1) = \frac{13}{27}$$

$$P_{\hat{F}}(\hat{\theta}^* = 9) = \frac{1}{27}(3 + 3 + 1) = \frac{7}{27}$$

### Función de distribución acumulada

$$\hat{G}_F(u) = P_{\hat{F}}(\hat{\theta}^* \leq u) = \begin{cases} 0 & \text{si } u < 2 \\ \frac{7}{27} & \text{si } 2 \leq u < 5 \\ \frac{20}{27} & \text{si } 5 \leq u < 9 \\ 1 & \text{si } u \geq 9 \end{cases}$$

# Ejemplo 1: Varianza de la mediana muestral

## Varianza bootstrap ideal

El estimador bootstrap ideal de la varianza es:

$$\text{Var}_{\hat{F}}(\hat{\theta}^*) = E_{\hat{F}}[(\hat{\theta}^*)^2] - \left(E_{\hat{F}}[\hat{\theta}^*]\right)^2$$

## Cuenta

$$E_{\hat{F}}[(\hat{\theta}^*)^2] = 2^2 \cdot \frac{7}{27} + 5^2 \cdot \frac{13}{27} + 9^2 \cdot \frac{7}{27} = \frac{920}{27} \approx 34.07$$

$$E_{\hat{F}}[\hat{\theta}^*] = 2 \cdot \frac{7}{27} + 5 \cdot \frac{13}{27} + 9 \cdot \frac{7}{27} = \frac{142}{27} \approx 5.26$$

$$\text{Var}_{\hat{F}}(\hat{\theta}^*) \approx 34.07 - 5.26^2 = 6.4024$$

## Ejemplo 2: Varianza de la media muestral

### Objetivo

Calcular el estimador bootstrap ideal de la varianza de la media muestral.

Sean  $\theta = E_F(X)$ , y  $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n X_i$ .

Nos interesa estimar:  $\text{Var}_F(\hat{\theta})$ .

Estimador bootstrap ideal:  $\text{Var}_{\hat{F}}(\hat{\theta}^*)$ .

Sabemos, para cualquier  $F$ :  $\text{Var}_F\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \text{Var}_F(X)$ .

### Aplicando bootstrap

$$\text{Var}_{\hat{F}}(\hat{\theta}^*) = \frac{1}{n} \text{Var}_{\hat{F}}(X^*) = \frac{1}{n} (E_{\hat{F}}[(X^*)^2] - E_{\hat{F}}[X^*]^2) = \frac{1}{n} \left( \frac{1}{n} \sum_{i=1}^n X_i^2 - \left( \frac{1}{n} \sum_{i=1}^n X_i \right)^2 \right).$$

- El ejemplo 2 es muy particular,  $\hat{\theta}$  es una función lineal de  $X$ .
- Casi siempre, calcular el estimador bootstrap ideal requiere enumerar todos los posibles valores de  $\hat{\theta}^*$ , calcular sus probabilidades, enumerar todas las muestras posibles  $\{X_1^*, \dots, X_n^*\}$
- El número de muestras posibles crece muy rápidamente con  $n$ , haciendo el cálculo del bootstrap ideal impracticable.

De hecho, existen:  $\binom{2n-1}{n-1}$  muestras distintas  $\{X_1^*, \dots, X_n^*\}$ . Por ejemplo, para  $n = 30$ , el número de muestras (distintos elementos en la original) las muestras es  $\approx 5.91 \times 10^{16}$ .

# Bootstrap (bootstrap simulado)

## Motivación

El estimador bootstrap ideal  $\hat{G}_F$  de  $G_F$ , si bien es conceptualmente calculable (depende de  $\hat{F}$ ), es impracticable, ya que requiere enumerar un número exorbitante de muestras.

## Idea

Aproximar  $\hat{G}_F$  mediante simulación.

## Definición

El estimador bootstrap de  $G_F(u)$  se basa en generar  $B$  realizaciones bootstrap:

$$\hat{\theta}_1^*, \dots, \hat{\theta}_B^* \text{ iid } \sim G_{\hat{F}}.$$

Se define:

$$G_{\text{boot}}(u) = \hat{G}_{\text{boot}}(u) = \frac{\#\{\hat{\theta}_b^* : \hat{\theta}_b^* \leq u\}}{B}.$$

# Bootstrap (bootstrap simulado)

Cómo obtener las realizaciones  $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$ ?

Repetimos el siguiente procedimiento para  $b = 1, \dots, B$ :

- Muestrear con reposición  $n$  observaciones de  $\{X_1, \dots, X_n\}$ , obteniendo  $\{X_1^{*(b)}, \dots, X_n^{*(b)}\}$
- Calcular

$$\hat{\theta}_b^* = h(X_1^{*(b)}, \dots, X_n^{*(b)})$$

Podemos calcular  $G_{\text{boot}}$  o bien, por ejemplo, definir el estimador de la varianza bootstrap de  $\hat{\theta}$  como

$$V_{\text{boot}}(\hat{\theta}) = \frac{1}{B} \sum_{b=1}^B (\hat{\theta}_b^* - \overline{\hat{\theta}^*})^2$$

donde

$$\overline{\hat{\theta}^*} = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^*.$$



En resumen, tenemos



- Efron, B. (1979). *Bootstrap Methods: Another Look at the Jackknife*. The Annals of Statistics, 7(1), 1–26.
- Efron, B. y Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Chapman & Hall/CRC.
- Bickel, P. J. y Freedman, D. A. (1981). *Some Asymptotic Theory for the Bootstrap*. The Annals of Statistics, 9(6), 1196–1217.
- van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.
- Breiman, L. (1996). *Bagging Predictors*. Machine Learning, 24(2), 123–140.
- Hastie, T., Tibshirani, R., Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.

**Bradley Efron: "The bootstrap is rarely the star of statistics, but it is the best supporting actor"**

(1.5). The use of the term bootstrap derives from the phrase *to pull oneself up by one's bootstrap*, widely thought to be based on one of the eighteenth century Adventures of Baron Munchausen, by Rudolph Erich Raspe. (The Baron had fallen to the bottom of a deep lake. Just when it looked like all was lost, he thought to pick himself up by his own bootstraps.) It is not the same as the term “bootstrap” used in computer science meaning to “boot” a computer from a set of core instructions, though the derivation is similar.