

# MOTIVACIÓN DEL PRINCIPIO BOOTSTRAP

1. Nota histórica.
2. El bootstrap uniforme.
3. Cálculo de la distribución bootstrap: distribución exacta y aproximada por Monte Carlo. Ejemplos.

## ¿QUÉ ES EL BOOTSTRAP?

Procedimiento estadístico que modeliza el muestreo de una población mediante el remuestreo de una muestra.

- **Ventajas**  
Sin hipótesis sobre el mecanismo que genera los datos.
- **Inconveniente**  
Técnica de “computación intensiva” (suavizada por la fuerza bruta del método de Monte Carlo).
- **Etimología**  
bootstrap = oreja.

El siguiente modismo anglosajón sugiere su filosofía:

**“to pull oneself up by one’s bootstraps”**

*(reponerse gracias a sus propios esfuerzos)*

## BREVE NOTA HISTÓRICA SOBRE EL BOOTSTRAP

---

### PREHISTORIA (precursores teóricos remotos)

- LAPLACE (1810). Teoría límite de primer orden.
- CHEBYSHEV (final s. XIX). Teoría límite de segundo orden.

### PRIMERAS CONTRIBUCIONES

- HUBBACK (1878-1968). Esquemas de muestreo espacial para ensayos agrícolas.
- MAHALANOBIS (años 30 y PGM). Precursor del bootstrap por bloques.

### OTRAS CONTRIBUCIONES

- GURNEY/McCARTHY/HARTIGAN (años 60, 70). Métodos de half-sampling para estimación de varianzas (U.S. Bureau of de Census).
- MARITZ/JARRET/SIMON (años 70, 80). Métodos de permutaciones relacionados con el bootstrap.

### ACTUALIDAD

- EFRON, B. (Stanford University). Fusionó la potencia de Monte Carlo con la resolución de problemas planteados de forma muy general.
- HALL, P. (Australian National University). Estadístico excepcional que dedica al bootstrap gran parte de su producción.

tpc

## MUNDO REAL

$X$  con distribución  $F$  desconocida  $\rightarrow \theta = \theta(F)$

$\downarrow$   $\uparrow$   
 $\vec{X} = (X_1, \dots, X_n)^t \in \mathcal{X} \rightarrow R(\vec{X}; F) ; \text{DISTRIBUCIÓN?}$

Observo  $\vec{x} = (x_1, \dots, x_n)^t$  y estimo  $F$  con  $\hat{F}$  a partir de  $\vec{x}$

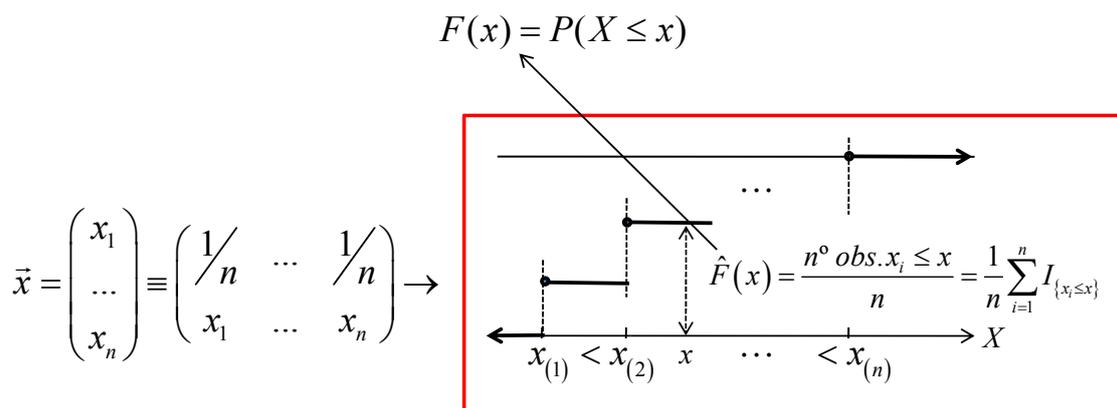
## MUNDO BOOTSTRAP

$X^*$  con distribución  $\hat{F}$  (conocida)

$\downarrow$   
 $\vec{X}^* = (X_1^*, \dots, X_n^*)^t \in \mathcal{X}^* = \mathcal{X} / \vec{x} \rightarrow R(\vec{X}^*; \hat{F})$  Como  $\hat{F}$  es conocida, esta distribución será conocida o aproximable por Monte Carlo

## BOOTSTRAP UNIFORME

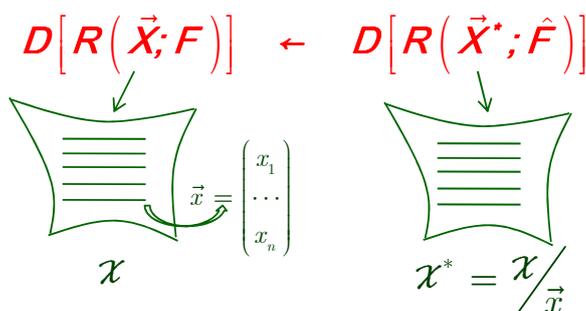
El bootstrap uniforme estima la distribución poblacional con la empírica.



- $\hat{F}$  es una función de distribución que asigna masa  $1/n$  a cada  $x_i$ ,  $i = 1, \dots, n$ .
- $X^*$  con distribución  $\hat{F}$  es la variable aleatoria discreta que toma valores  $x_i$ ,  $i = 1, \dots, n$  con probabilidad  $1/n$ .

Notemos que

$$EX^* = \frac{1}{n} \sum_{i=1}^n x_i; \quad VX^* = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$



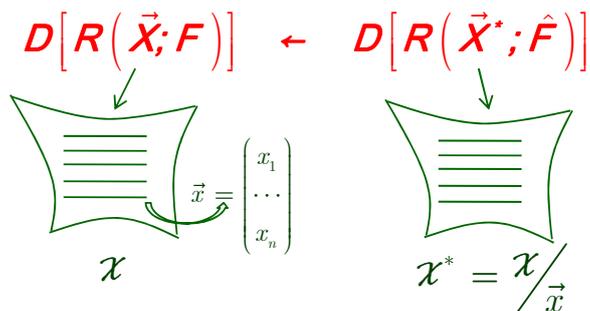
### Ejemplo 1:

$$\theta = \mu = \mu(F) = \int x dF(x)$$

$$R(\vec{X}; F) = \hat{\theta} = \theta(\vec{X}) = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X} \quad \text{¿Distribución?} \left[ \rightarrow N\left(\mu, \frac{\sigma^2}{n}\right) \right]$$

$$R(\vec{X}^*; \hat{F}) = \hat{\theta}^* = \theta(\vec{X}^*) = \frac{1}{n} \sum_{i=1}^n X_i^* = \bar{X}^* \quad \text{Distribución conocida o aproximable por Monte Carlo}$$

**NOTA.** Aproximamos la distribución de  $\bar{X}$ , variable de la que solo conocemos un valor, por la distribución de  $\bar{X}^*$ .



**Ejemplo 2:**

$$\theta = c_{1/2} = c_{1/2}(F)$$

$$R(\vec{X}; F) = \hat{\theta} = \theta(\vec{X}) = X_{(m)} \quad (\text{si } n = 2m - 1)$$

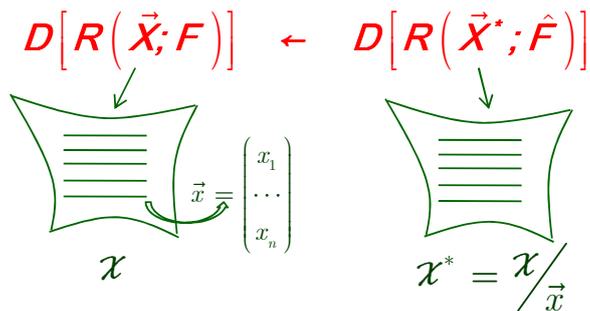
$$R(\vec{X}^*; \hat{F}) = \hat{\theta}^* = \theta(\vec{X}^*) = X_{(m)}^* \quad (\text{si } n = 2m - 1)$$

$$\text{¿Dist.?} \rightarrow N\left(c_{1/2}, \frac{1}{4nf^2(c_{1/2})}\right)$$

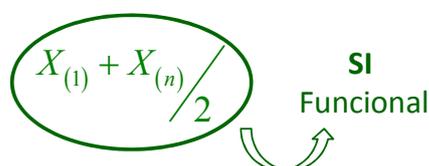
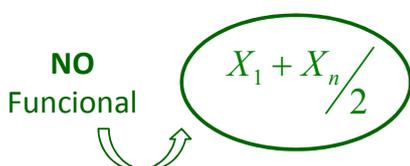
Distribución conocida o aproximable por Monte Carlo

**NOTA.** Aproximamos la distribución de  $X_{(m)}$ , variable de la que solo conocemos un valor, por la distribución de  $X_{(m)}^*$ .

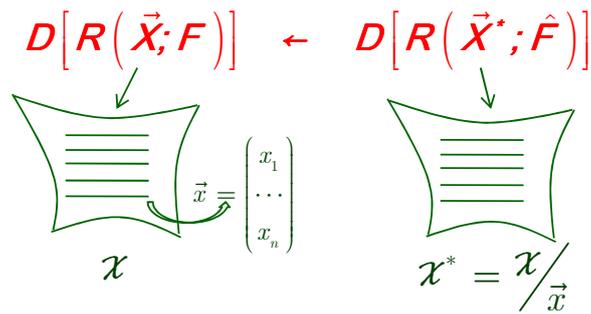
**CÁLCULO DE LA DISTRIBUCIÓN BOOTSTRAP**



- Puede calcularse exactamente (con dificultad si  $n$  es grande) si el número de posibles valores de  $R(\vec{X}^*; \hat{F})$  es igual al número de posibles m.a.s. de tamaño  $n$  extraídas de  $X^*$  sin tener en cuenta el orden. Por ejemplo, si  $X$  es continua y  $R(\vec{X}; F) = \hat{\theta} = \theta(\vec{X}) = \theta(\hat{F})$  es un estadístico funcional (toma el mismo valor sobre todas las permutaciones de una muestra dada), como una media o varianza muestral.



## CÁLCULO DE LA DISTRIBUCIÓN BOOTSTRAP



- La distribución bootstrap puede calcularse exactamente en algunos casos de modo sencillo, como por ejemplo cuando el estadístico de interés es la mediana muestral.

$$R(\vec{X}; F) = X_{(m)} \quad \text{si } n = 2m - 1$$

- La distribución bootstrap siempre se puede aproximar por Monte Carlo.

- **CÁLCULO EXACTO SI  $\#R(\vec{X}^*; \hat{F}) = \#m.a.s. \text{ de } X^*$  (SALVO EL ORDEN)**

¿Cómo se genera  $X^* = X / \vec{x}$  ?



Si  $N_i = n^\circ \text{ de veces que se obtiene } x_i \equiv \#\{X^* = x_i\}$ , entonces

$$(N_1, \dots, N_n) \in M_n\left(n, \left(\frac{1}{n}, \dots, \frac{1}{n}\right)\right)$$

$$P(N_1 = n_1, \dots, N_n = n_n) = \frac{n!}{n_1! \dots n_n!} \left(\frac{1}{n}\right)^{n_1} \dots \left(\frac{1}{n}\right)^{n_n} = \frac{n! n^{-n}}{n_1! \dots n_n!} \left[ n_i \geq 0, \sum_{i=1}^n n_i = n \right]$$

Por lo tanto, ésta es la probabilidad de cada m.a.s. particular de  $X^*$  con dist.  $\hat{F}$ , salvo permutaciones de sus elementos. En consecuencia, bajo la suposición anterior, es también la probabilidad de cada valor de la variable bootstrap  $R(\vec{X}^*; \hat{F})$ .

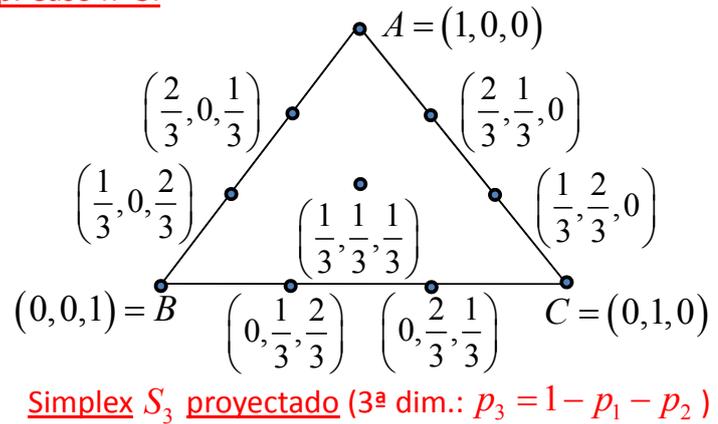
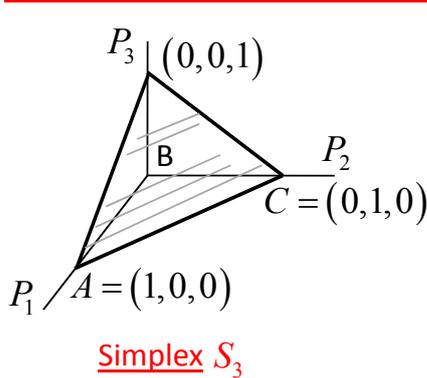
## VECTOR DE REMUESTRO BOOTSTRAP

Se llama vector de remuestro bootstrap al vector aleatorio

$$\vec{P}^* = (P_1^*, \dots, P_n^*) = \frac{1}{n} (N_1, \dots, N_n) \in \frac{1}{n} M_n \left( n, \left( \frac{1}{n}, \dots, \frac{1}{n} \right) \right).$$

Sus valores, también llamados vectores de remuestro (particulares), son masas de probabilidad sobre  $x_1, \dots, x_n$  cuyas coordenadas son (salvo 1/n) números enteros que suman n. Cada vector de remuestro caracteriza un elemento de  $\mathcal{X}^*$  (réplica bootstrap de  $(x_1, \dots, x_n)$ ). Así, si  $n=5$ ,  $(2/5, 0, 1/5, 2/5, 0) \equiv (x_1, x_1, x_3, x_4, x_4)$ .

### Vectores de remuestro bootstrap. Caso $n=3$ .



### Ejemplo de cálculo exacto de la distribución bootstrap

$$\theta = \mu = \mu(F) = \int x dF(x)$$

$$R(\bar{X}; F) = \hat{\theta} = \theta(\bar{X}) = \theta(\hat{F}) = \bar{X} \quad R(\bar{X}^*; \hat{F}) = \hat{\theta}^* = \theta(\bar{X}^*) = \theta(\hat{F}^*) = \bar{X}^*$$

$$D[R(\bar{X}; F) = \bar{X}] \leftarrow D[R(\bar{X}^*; \hat{F}) = \bar{X}^*]$$

$\mathcal{X}^*$ (salvo permutaciones)	$(N_1, N_2, N_3)$	$(P_1^*, P_2^*, P_3^*)$	$3!3^{-3}/n_1!n_2!n_3!$	$\bar{X}^*$
$(x_1, x_1, x_1)$	$(3, 0, 0)$	$(1, 0, 0)$	0,0370	$(3x_1)/3$
$(x_2, x_2, x_2)$	$(0, 3, 0)$	$(0, 1, 0)$	0,0370	$(3x_2)/3$
$(x_3, x_3, x_3)$	$(0, 0, 3)$	$(0, 0, 1)$	0,0370	$(3x_3)/3$
$(x_1, x_2, x_3)$	$(1, 1, 1)$	$(1/3, 1/3, 1/3)$	0,2222	$(x_1 + x_2 + x_3)/3$
$(x_1, x_1, x_2)$	$(2, 1, 0)$	$(2/3, 1/3, 0)$	0,1111	$(2x_1 + x_2)/3$
$(x_1, x_1, x_3)$	$(2, 0, 1)$	$(2/3, 0, 1/3)$	0,1111	$(2x_1 + x_3)/3$
$(x_1, x_2, x_2)$	$(1, 2, 0)$	$(1/3, 2/3, 0)$	0,1111	$(x_1 + 2x_2)/3$
$(x_2, x_2, x_3)$	$(0, 2, 1)$	$(0, 2/3, 1/3)$	0,1111	$(2x_2 + x_3)/3$
$(x_1, x_3, x_3)$	$(1, 0, 2)$	$(1/3, 0, 2/3)$	0,1111	$(x_1 + 2x_3)/3$
$(x_2, x_3, x_3)$	$(0, 1, 2)$	$(0, 1/3, 2/3)$	0,1111	$(x_2 + 2x_3)/3$

Como consecuencia de las consideraciones multinomiales anteriores, la distribución bootstrap puede considerarse continua:

$$(N_1, \dots, N_n) \in M_n \left( n, \left( \frac{1}{n}, \dots, \frac{1}{n} \right) \right)$$

$$P(N_1 = n_1, \dots, N_n = n_n) = \frac{n!}{n_1! \dots n_n!} \left( \frac{1}{n} \right)^{n_1} \dots \left( \frac{1}{n} \right)^{n_n} = \frac{n! n^{-n}}{n_1! \dots n_n!} \left[ n_i \geq 0, \sum_{i=1}^n n_i = n \right]$$

• Nº de átomos de  $(N_1, \dots, N_n)$  =  $\binom{2n-1}{n}$

- n=3
- 3 0 0 → •••××
  - 0 3 0 → ×••××
  - 0 0 3 → ××•••
  - 1 1 1 → •×•×•
  - 2 1 0 → ••×××
  - 2 0 1 → ••××•
  - 1 2 0 → •×••×
  - 0 2 1 → ×••×•
  - 1 0 2 → •××••
  - 0 1 2 → ×××••

$$n^\circ \text{ total} = \binom{2n-1}{n-1} = \binom{2n-1}{n}$$

[ × *separador*  
• *dato* ]

[ Nº de formas de seleccionar 2 cruces entre 5 posiciones. ]

• Átomo más probable ( $n_1 = \dots = n_n = 1$ ) : probabilidad =  $n! n^{-n}$

TABLE A.1. Number of atoms,  $\binom{2n-1}{n}$ , of bootstrap distribution, together with probability,  $n!/n^n$ , of most likely or modal atom. In each case the modal value is  $\hat{\theta}$ , the original statistic computed from the sample.

$n$	number of atoms	probability of most likely atom
2	3	0.5
3	10	0.2222
4	35	0.0940
5	126	0.0384
6	462	$1.5 \times 10^{-2}$
7	1,716	$6.1 \times 10^{-3}$
8	6,435	$2.4 \times 10^{-3}$
9	24,310	$9.4 \times 10^{-4}$
10	92,378	$3.6 \times 10^{-4}$
12	1,352,078	$5.4 \times 10^{-5}$
15	$7.8 \times 10^7$	$3.0 \times 10^{-6}$
20	$6.9 \times 10^{10}$	$2.3 \times 10^{-8}$

• **CÁLCULO EXACTO EN CASOS PARTICULARES: MEDIANA MUESTRAL**

$$\theta = c_{1/2} = c_{1/2}(F) = \inf \{t / P_F(X \leq t) \geq 1/2\}$$

Si  $n=2m-1$

$$R(\vec{X}; F) = \hat{\theta} = \theta(\vec{X}) = \theta(\hat{F}) = X_{(m)} \quad R(\vec{X}^*; \hat{F}) = \hat{\theta}^* = \theta(\vec{X}^*) = \theta(\hat{F}^*) = X_{(m)}^*$$

$$D[R(\vec{X}; F) = X_{(m)}] \leftarrow D[R(\vec{X}^*; \hat{F}) = X_{(m)}^*]$$

**NOTA.**  $X_{(m)}^*$  es una variable aleatoria discreta que toma los valores  $\{x_{(1)}, \dots, x_{(n)}\}$ .

$$Si \vec{x} = (x_1, \dots, x_7) = (4, 5, 7, 2, 6, 1, 9) \rightarrow \begin{matrix} X_{(4)}^* \\ (2, 2, 5, \boxed{6}, 6, 7, 7) \\ (1, 1, 4, \boxed{4}, 4, 6, 7) \\ (4, 5, 5, \boxed{6}, 7, 9, 9) \\ \dots \end{matrix}$$

$$P_*(X_{(m)}^* > x_{(k)}) \equiv P_{\hat{F}}(X_{(m)}^* > x_{(k)}) \equiv P_F \left( X_{(m)}^* > x_{(k)} / \vec{X} = \vec{x} \right) =$$

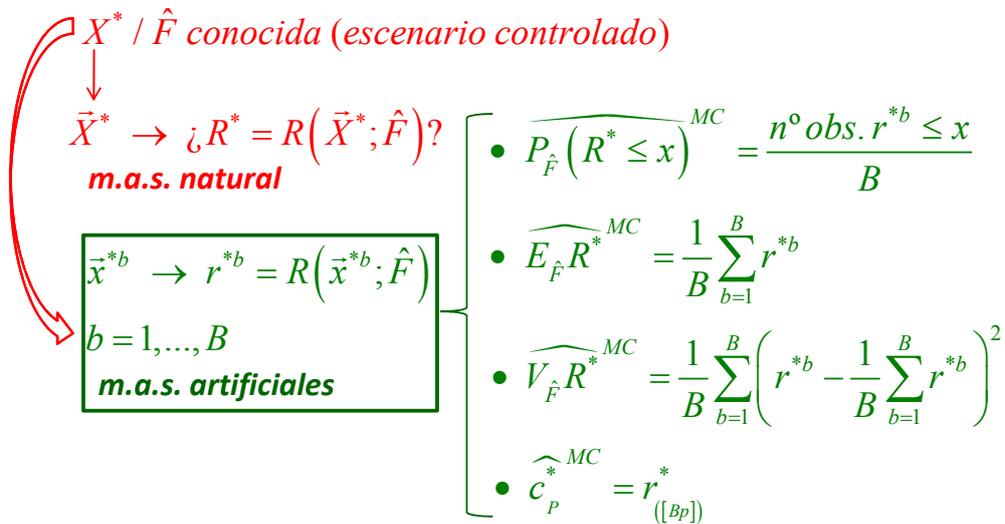
$$= P_F \left( \sum_{j=1}^k \# \{X^* = x_{(j)}\} \leq m-1 / \vec{X} = \vec{x} \right) = F_{Bin(n, \frac{k}{n})}(m-1).$$

$$P_*(X_{(m)}^* = x_{(k)}) = P_*(X_{(m)}^* > x_{(k-1)}) - P_*(X_{(m)}^* > x_{(k)}) = F_{Bin(n, \frac{k-1}{n})}(m-1) - F_{Bin(n, \frac{k}{n})}(m-1)$$

$$G_*(y) = P_*(X_{(m)}^* \leq y) = \begin{cases} \sum_{k=1}^r \left( F_{Bin(n, \frac{k-1}{n})}(m-1) - F_{Bin(n, \frac{k}{n})}(m-1) \right) & x_{(r)} \leq y < x_{(r+1)} \\ 0 & y < x_{(1)} \\ 1 & y \geq x_{(n)} \end{cases}$$

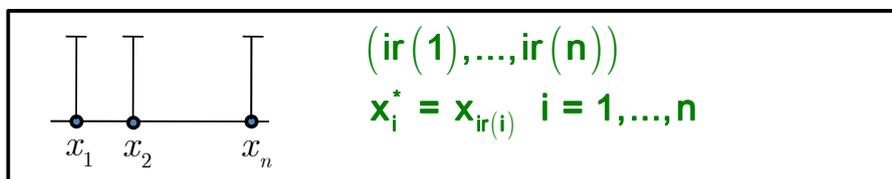
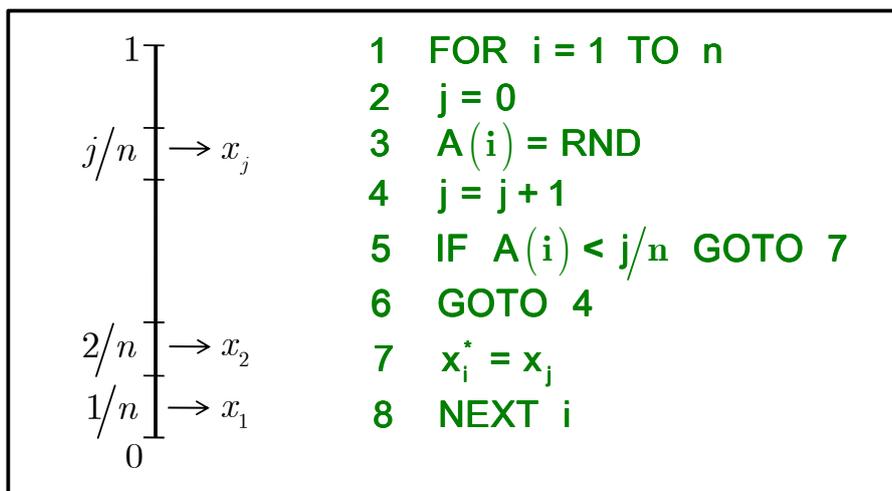
• **LA DISTRIBUCIÓN BOOTSTRAP SIEMPRE SE PUEDE APROXIMAR POR MONTE CARLO**

Si se conoce el mecanismo que genera los datos se podrá simular el mecanismo que genera los valores de cualquier variable  $R(\bar{X}; F)$ .



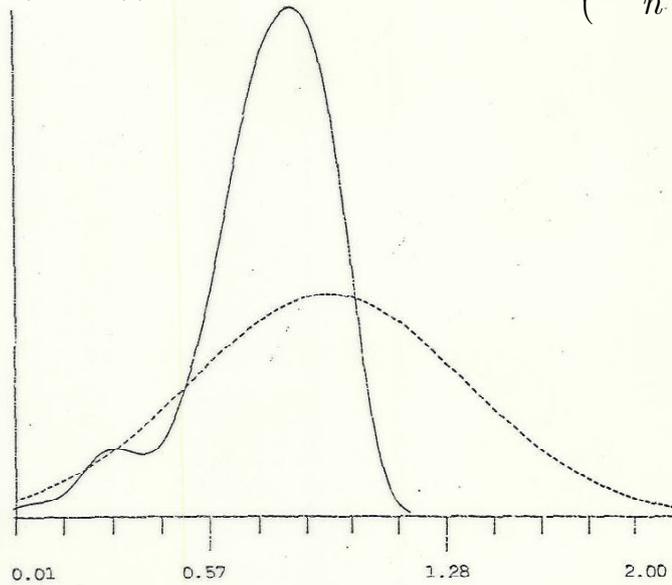
**NOTA.** Las V.A. que aportan esas 4 estimaciones, obtenidas reemplazando  $r^{*b}$  por  $R^{*b} = R(\bar{X}^{*b}; \hat{F})$ ,  $b = 1, \dots, B$ , son estadísticos sobre la m.a.s.  $(R^{*1}, \dots, R^{*B})$  que convergen C. S., consistencia fuerte, a los correspondientes 4 objetivos poblacionales.

La obtención de una muestra artificial de  $\hat{F}$  puede basarse en la simulación de una variable aleatoria uniforme en (0,1), o bien, en una uniforme discreta de tamaño n.



$$\begin{pmatrix} X \\ Y \end{pmatrix} \in N_2 \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right)$$

$$\left( (x_1, y_1), \dots, (x_n, y_n) \right)^t \rightarrow \sum_n \rightarrow \lambda_{ni} \xrightarrow{n \rightarrow \infty} N \left( \lambda_i, \frac{2}{n} \lambda_i^2 \right)$$



SEGUNDO AUTOVALOR  
 Aproximación normal: trazo discontinuo  
 Estimación densidad Bootstrap: trazo continuo.

**¡Las aproximaciones bootstrap y TCL pueden ser muy distintas!**

### ONE-SAMPLE PROBLEMS

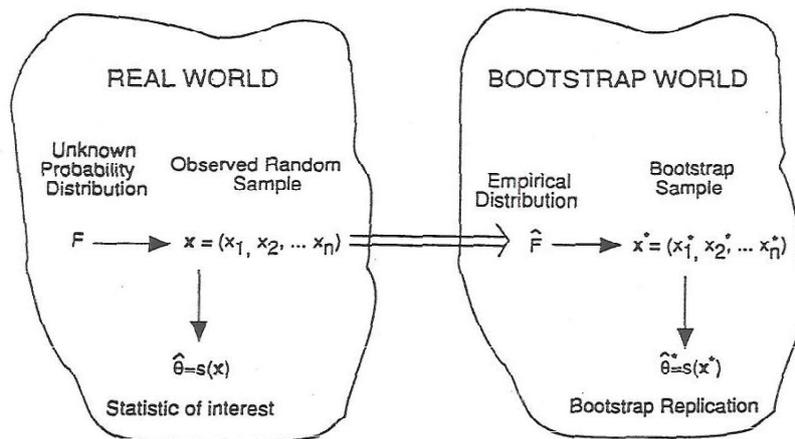


Figure 8.1. A schematic diagram of the bootstrap as it applies to one-sample problems. In the real world, the unknown probability distribution  $F$  gives the data  $x = (x_1, x_2, \dots, x_n)$  by random sampling; from  $x$  we calculate the statistic of interest  $\hat{\theta} = s(x)$ . In the bootstrap world,  $\hat{F}$  generates  $x^*$  by random sampling, giving  $\hat{\theta}^* = s(x^*)$ . There is only one observed value of  $\hat{\theta}$ , but we can generate as many bootstrap replications  $\hat{\theta}^*$  as affordable. The crucial step in the bootstrap process is " $\Rightarrow$ ", the process by which we construct from  $x$  an estimate  $\hat{F}$  of the unknown population  $F$ .

# EL BOOTSTRAP Y LA ESTIMACIÓN DEL SESGO Y LA PRECISIÓN DE UN ESTIMADOR

1. Estimación bootstrap del sesgo y la precisión de un estimador.
2. Estimación jackknife del sesgo y la precisión de un estimador.
3. Relación entre el bootstrap y el jackknife en la estimación del sesgo y la precisión de un estimador. Ejemplos.

## Aplicaciones:

- Aproximación por Monte Carlo de la distribución bootstrap de variables aleatorias de interés y características de las mismas. **P6**
- Comparación del bootstrap y el jackknife en la estimación de sesgos y varianzas. **P7**

## ESTIMACIÓN BOOTSTRAP DEL SESGO Y LA PRECISIÓN DE UN ESTIMADOR $\hat{\theta} = \theta(\bar{X})$ DE $\theta = \theta(F)$

$$R = R(\bar{X}; F) = \theta(\bar{X}) - \theta(F) \rightarrow \begin{aligned} \text{Sesgo}_F \hat{\theta} &= E_F R \\ \text{Var}_F \hat{\theta} &= \text{Var}_F R \end{aligned}$$

$$R^* = R(\bar{X}^*; \hat{F}) = \theta(\bar{X}^*) - \theta(\hat{F}) \rightarrow \begin{aligned} \widehat{\text{Sesgo}}_F \hat{\theta}^{boot} &= E_{\hat{F}} R^* \\ \widehat{\text{Var}}_F \hat{\theta}^{boot} &= \text{Var}_{\hat{F}} R^* \end{aligned}$$

$$\left[ \begin{array}{l} X^* / \hat{F} \text{ conocida (escenario controlado)} \\ \downarrow \\ \bar{x}^{*b} \rightarrow r^{*b} = R(\bar{x}^{*b}; \hat{F}), \quad b = 1, \dots, B \end{array} \right.$$

$$\bullet \widehat{\text{Sesgo}}_F \hat{\theta}^{boot MC} = \widehat{E}_{\hat{F}} R^* MC = \frac{1}{B} \sum_{b=1}^B r^{*b} = \frac{1}{B} \sum_{b=1}^B \theta(\bar{x}^{*b}) - \theta(\hat{F})$$

$$\bullet \widehat{\text{Var}}_F \hat{\theta}^{boot MC} = \widehat{V}_{\hat{F}} R^* MC = \widehat{V}_{\hat{F}} \left( \theta(\bar{X}^*) \right) MC = \frac{1}{B} \sum_{b=1}^B \left[ \theta(\bar{x}^{*b}) - \frac{1}{B} \sum_{b=1}^B \theta(\bar{x}^{*b}) \right]^2$$

Notemos que la última expresión anterior,

$$\widehat{\widehat{Var}}_F \hat{\theta}^{boot MC} = \widehat{V}_{\hat{F}} R^* MC = \widehat{V}_{\hat{F}} \left( \theta(\bar{X}^*) \right)^{MC} = \frac{1}{B} \sum_{b=1}^B \left[ \theta(\bar{x}^{*b}) - \frac{1}{B} \sum_{b=1}^B \theta(\bar{x}^{*b}) \right]^2$$

permite aproximar por Monte Carlo la varianza de un estimador sin conocer su expresión exacta (utilizando sólo el concepto de varianza).

- Si  $\hat{\theta} = \bar{X} \Rightarrow \text{var}_F \hat{\theta} = E_F \left[ \left( \hat{\theta} - E_F \hat{\theta} \right)^2 \right] = \frac{\sigma^2}{n}$  (expresión conocida)  
 $\widehat{\widehat{var}}_F \bar{X}^{boot} = \text{var}_{\hat{F}} \bar{X}^* = \text{var}_{\hat{F}} \frac{1}{n} \sum_{i=1}^n X_i^* = \frac{n}{n^2} \text{var}_{\hat{F}} X^* = \frac{1}{n} \left( \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right) = \frac{S^2}{n}$   
 que es la versión plug-in de la varianza conocida anterior.
- Si ahora  $\hat{\theta}$  es cualquier otro estimador de varianza desconocida, la estimación bootstrap exacta de ésta (su plug-in) también será desconocida, pero se podrá aproximar muy bien por Monte Carlo (consistencia fuerte y B tan grande como queramos).

### Ejemplo:

Supongamos que hemos observado 15 tiempos de vida de un organismo y queremos estimar la precisión de dos estimadores de su vida media: media y mediana muestral.

$$\bar{x} = (0,143; 0,182; 0,256; 0,260; 0,270; 0,437; 0,509; 0,611; 0,712; 1,04; 1,09; 1,15; 1,46; 1,88; 2,08)$$

$$\theta(F) = \mu \rightarrow \theta(\bar{X}) = \bar{X} \quad (=0,804 \text{ sobre } \bar{x})$$

$$\sigma(\bar{X}) = \left( \frac{\text{var}(X)}{n} \right)^{1/2} \rightarrow \hat{\sigma}(\bar{X}) = \left( \frac{1}{n^2} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2} = 0,155$$

$$\theta(F) = c_{1/2} \rightarrow \theta(\bar{X}) = c_{n,1/2} \quad (=0,611 \text{ sobre } \bar{x})$$

$$\sigma(c_{n,1/2})?? \rightarrow \hat{\sigma}(c_{n,1/2})??$$

**NOTA.** La Teoría clásica considera estimadores de máxima verosimilitud y aproxima su precisión por el inverso de la raíz cuadrada de la información de Fisher (contexto paramétrico) o bien, en contexto no paramétrico, considera la aproximación asintótica que proporciona el T.C.L. (expresión que depende de la densidad poblacional ...).

**El bootstrap proporciona una respuesta más general a la estimación de la precisión.**

```

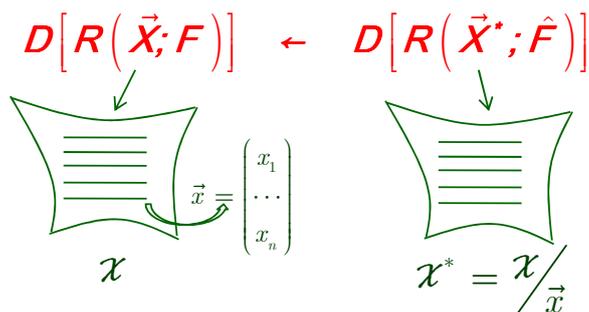
#Para la muestra de TIEMPOS DE VIDA, estima (plug-in) la precisión de la media muestral
#(desvmedia) y también aproxima por Monte Carlo la estimación bootstrap de dicha precisión
#(desvmediaboot) y también la precisión de la mediana muestral, de la cual no se conoce su
#expresión, (desvmedianaboot). También estima el sesgo bootstrap de esos dos estimadores
#(sesgmediaboot y sesgmedianaboot, respectivamente).
#####33

muestra<-c(0.143,0.182,0.256,0.260,0.270,0.437,0.509,0.611,0.712,1.04,1.09,1.15,1.46,1.88,2.08);
n=length(muestra);
varmedia<-(1/(n^2))*sum((muestra-mean(muestra))^2);
desvmedia=sqrt(varmedia)
remuestra<-numeric(n);
B=1000
media<-numeric(B)
remordenada<-numeric(n)
mediana<-numeric(B)
for (k in 1:B)
{
  for (i in 1:n)
  {
    j=1
    u=runif(1)
    while (u>=j/n)
    {
      j=j+1
    }
    remuestra[i]=muestra[j];
  }
  media[k]=mean(remuestra)
  remordenada=sort(remuestra)
  mediana[k]=remordenada[8]
}
varmediaboot=(1/B)*sum((media-mean(media))^2);
desvmediaboot=sqrt(varmediaboot);
varmedianaboot=(1/B)*sum((mediana-mean(mediana))^2);
desvmedianaboot=sqrt(varmedianaboot);
desvmedia
desvmediaboot
desvmedianaboot
sesgmediaboot=mean(media)-mean(muestra)
sesgmedianaboot=mean(mediana)-muestra[8]
sesgmediaboot
sesgmedianaboot

```

	$\bar{x} = 0,804$	$x_{(8)} = 0,611$
<b>Estimación precisión (plug-in)</b>	0,155579	?
<b>Estimación precisión (bootstrap)</b>	0,155591	0,250086
<b>Estimación sesgo (bootstrap)</b>	0,005875	0,055414

## ESTIMACIÓN JACKKNIFE DEL SESGO Y LA PRECISIÓN DE UN ESTIMADOR $\hat{\theta} = \theta(\vec{X})$ DE $\theta = \theta(F)$



$\mathcal{X}^*$  es el resultado de “exprimir” los datos de alguna manera:

- En el **bootstrap** se remuestra realizando muestreo aleatorio con reemplazamiento y tamaño fijo de la muestra inicial (m.a.s. de su función de distribución empírica). Sus vectores de remuestreo son  $\vec{P}^* = (n_1/n, \dots, n_n/n) / n_i \in Z^+, i = 1, \dots, n, \sum n_i = n$ .
- En el **jackknife** se remuestra omitiendo un dato en cada réplica de la muestra inicial. Sus vectores de remuestreo son  $\vec{P}_{(i)}^* = (1/n-1, \dots, 0, \dots, 1/n-1) /$ ,  $i = 1, \dots, n$ .

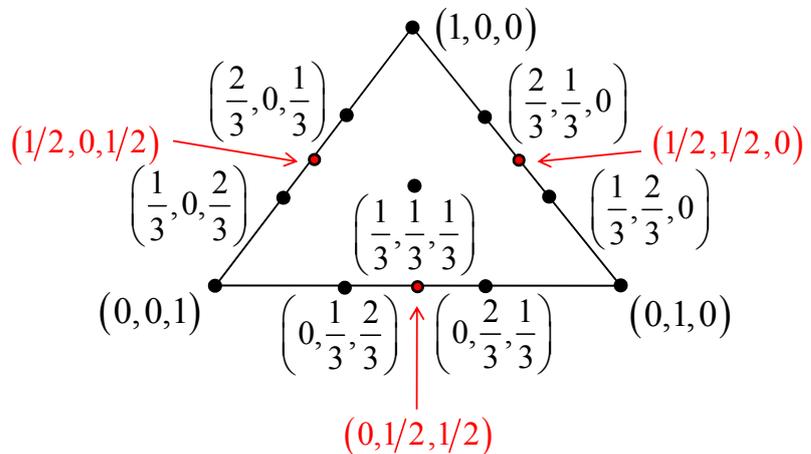
**Nota.** El *submuestreo aleatorio* realiza muestreo aleatorio sin reemplazamiento y tamaño variable de la muestra inicial.

$\mathcal{X}^*$ (salvo permutaciones)	$(P_1^*, P_2^*, P_3^*)$
<b>Bootstrap</b>	
$(x_1, x_1, x_1)$	$(1, 0, 0)$
$(x_2, x_2, x_2)$	$(0, 1, 0)$
$(x_3, x_3, x_3)$	$(0, 0, 1)$
$(x_1, x_2, x_3)$	$(1/3, 1/3, 1/3)$
$(x_1, x_1, x_2)$	$(2/3, 1/3, 0)$
$(x_1, x_1, x_3)$	$(2/3, 0, 1/3)$
$(x_1, x_2, x_2)$	$(1/3, 2/3, 0)$
$(x_2, x_2, x_3)$	$(0, 2/3, 1/3)$
$(x_1, x_3, x_3)$	$(1/3, 0, 2/3)$
$(x_2, x_3, x_3)$	$(0, 1/3, 2/3)$
<b>Jackknife</b>	
$(, x_2, x_3)$	$(0, 1/2, 1/2)$
$(x_1, , x_3)$	$(1/2, 0, 1/2)$
$(x_1, x_2, )$	$(1/2, 1/2, 0)$

### Vectores de remuestreo bootstrap y jackknife, n=3

$$S_3 = \left\{ \vec{p} = (p_1, p_2, p_3) / p_i \geq 0, \sum_{i=1}^3 p_i = 1 \right\}$$

(Conjunto de posibles ponderaciones de los datos)



Simplex  $S_3$  proyectado (3ª dim.:  $p_3 = 1 - p_1 - p_2$ )

El jackknife no suele utilizarse para aproximar la distribución de  $R(\vec{X}; F)$  sino para estimar características de esa variable aleatoria, como su sesgo y varianza.

$$R = R(\vec{X}; F) = \theta(\vec{X}) - \theta(F) \rightarrow \begin{aligned} \text{Sesgo}_F \hat{\theta} &= E_F R \\ \text{Var}_F \hat{\theta} &= \text{Var}_F R \end{aligned}$$

$$R^* = R(\vec{X}^*; \hat{F}) = \theta(\vec{X}^*) - \theta(\hat{F}), \quad \vec{X}^* \in \mathcal{X}^{jackk} / \vec{x} = \{ \vec{x}_{(i)} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n), i = 1, \dots, n \}$$

Si, razonando como en el bootstrap, definimos los estimadores jackknife del sesgo y la varianza de  $\hat{\theta}$  como:

$$ER^* = \frac{1}{n} \sum_{i=1}^n \theta(\vec{X}_{(i)}) - \theta(\vec{X})$$

$$VR^* = V(\theta(\vec{X}^*)) = \frac{1}{n} \sum_{i=1}^n \left[ \theta(\vec{X}_{(i)}) - \frac{1}{n} \sum_{i=1}^n \theta(\vec{X}_{(i)}) \right]^2$$

entonces, como las réplicas jackknife de  $\vec{x}$  son más parecidas a  $\vec{x}$  que, por ejemplo, las réplicas bootstrap, se tendría que

$$\frac{1}{n} \sum_{i=1}^n \theta(\bar{X}_{(i)}) - \theta(\bar{X}) < E_{\hat{F}} \hat{\theta}^* - \theta(\bar{X}) \quad (1)$$

$$\left( \theta(\bar{X}_{(i)}) - \frac{1}{n} \sum_{i=1}^n \theta(\bar{X}_{(i)}) \right)^2 < \left( \theta(\bar{X}^*) - E_{\hat{F}} \hat{\theta}^* \right)^2 \quad (2)$$

- Por (1),  $ER^*$  estaría infraestimando  $Sesgo_F \hat{\theta}$  (el segundo miembro de (1) es el plug-in de dicho sesgo: *estimación bootstrap*).
- Por (2),  $VR^*$  estaría infraestimando  $Var_F \hat{\theta}$  (la esperanza bajo la empírica del segundo miembro de (2) también es el plug-in de dicha varianza: *estimación bootstrap*).

Será necesario introducir un **factor de elevación** en ambos estimadores.

**QUENOULLI (1949)** introduce el método jackknife para estimar el sesgo de un estimador  $\hat{\theta}$  como:

$$\widehat{Sesgo}_F \hat{\theta}^{jackk} = (n-1) \left[ \frac{1}{n} \sum_{i=1}^n \theta(\bar{X}_{(i)}) - \theta(\bar{X}) \right]$$

↑  
(Factor elevación)

Su idea fue elegir el factor de elevación de modo que si  $\hat{\theta}$  es un estimador usual (media o varianza muestral) el estimador jackknife de su sesgo fuera insesgado en él.

- Si  $\hat{\theta} = \theta(\bar{X}) = \bar{X} = (1/n) \sum X_i$  ( $\theta = \mu$ ) entonces va a ocurrir que el estimador jackknife del sesgo de  $\hat{\theta}$  será insesgado en dicho sesgo (como estimador de  $\mu$ ) para cualquier factor de elevación, lo que no nos permitiría encontrar dicho factor:

$$\frac{1}{n} \sum_{i=1}^n \theta(\bar{X}_{(i)}) = \frac{1}{n} \left[ \frac{1}{n-1} \sum_{j \neq 1} X_j + \dots + \frac{1}{n-1} \sum_{j \neq n} X_j \right] = \frac{1}{n} \left[ \frac{\sum_{i=1}^n (n-1) X_i}{n-1} \right] = \bar{X}$$

- Si  $\hat{\theta} = \theta(\bar{X}) = S^2 = (1/n) \sum (X_i - \bar{X})^2$  ( $\theta = \sigma^2$ ) entonces puede probarse fácilmente que con el factor de elevación (n-1) el estimador jackknife del sesgo de  $\hat{\theta}$  es insesgado en dicho sesgo (como estimador de  $\sigma^2$ ).

$$E_F \left( \widehat{Sesgo}_F \hat{\theta}^{jackk} \right) = Sesgo \hat{\theta} \left[ = E_F \hat{\theta} - \sigma^2 = \frac{n-1}{n} \sigma^2 - \sigma^2 = -\frac{\sigma^2}{n} \right]$$

**TUCKEY (1958)** bautiza el método jackknife y lo utiliza para estimar la varianza de un estimador  $\hat{\theta}$  como:

$$\widehat{Var}_F \hat{\theta}^{jackk} = \frac{n-1}{n} \sum_{i=1}^n \left[ \theta(\bar{X}_{(i)}) - \frac{1}{n} \sum_{i=1}^n \theta(\bar{X}_{(i)}) \right]^2$$

**(Factor elevación)**

Nota

El jackknife es inconsistente estimando la varianza de otros estimadores, como por ejemplo, la mediana muestral.

El factor de elevación, que coincide con el del estimador jackknife del sesgo, se elige con la misma idea anterior; en este caso, **que el estimador jackknife de la varianza de  $\hat{\theta} = \bar{X}$  sea insesgado en la varianza de  $\hat{\theta}$  como estimador de  $\theta = \mu$ .**

Comprobemos la afirmación anterior:

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

- $Var_F \hat{\theta} = Var_F \bar{X} = \frac{\sigma^2}{n}$

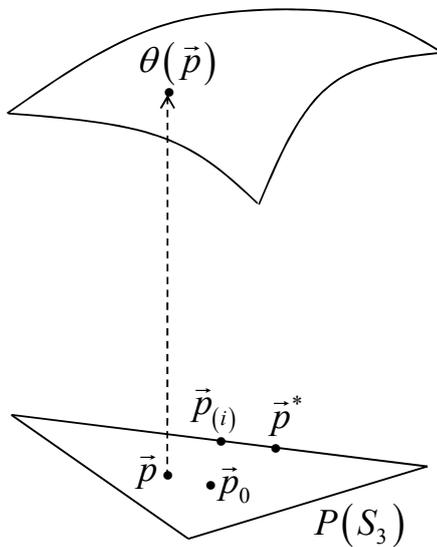
- $\widehat{Var}_F \hat{\theta}^{jackk} = \frac{n-1}{n} \sum_{i=1}^n \left[ \theta(\bar{X}_{(i)}) - \frac{1}{n} \sum_{i=1}^n \theta(\bar{X}_{(i)}) \right]^2 = *$

$$* \left[ \begin{aligned} \theta(\bar{X}_{(i)}) - \frac{1}{n} \sum_{i=1}^n \theta(\bar{X}_{(i)}) &= \frac{n\bar{X} - X_i}{n-1} - \frac{1}{n} \sum_{i=1}^n \left( \frac{n\bar{X} - X_i}{n-1} \right) = \\ &= \frac{n\bar{X} - X_i}{n-1} - \frac{1}{n(n-1)} (nn\bar{X} - n\bar{X}) = \frac{n\bar{X} - X_i}{n-1} - \frac{(n-1)\bar{X}}{n-1} = \frac{\bar{X} - X_i}{n-1} \\ &= \frac{n-1}{n} \sum_{i=1}^n \left( \frac{\bar{X} - X_i}{n-1} \right)^2 = \frac{1}{n(n-1)} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{S_c^2}{n} \end{aligned} \right.$$

- $E_F \left[ \widehat{Var}_F \hat{\theta}^{jackk} \right] = E_F \left[ \frac{S_c^2}{n} \right] = \frac{\sigma^2}{n} = Var_F \hat{\theta}$

# RELACIÓN ENTRE EL BOOTSTRAP Y EL JACKKNIFE EN LA ESTIMACIÓN DEL SESGO Y LA PRECISIÓN DE UN ESTIMADOR

Un estadístico funcional se puede considerar como una superficie aleatoria sobre  $S_n$ :



**Ejemplo:**  $\theta(F) = \int x dF(x)$

$$\bullet \vec{p}_0 \equiv \hat{F} \equiv \begin{pmatrix} 1/n & \cdots & 1/n \\ X_1 & \cdots & X_n \end{pmatrix} \rightarrow \theta(\vec{p}_0) = \theta(\hat{F}) = \int x d\hat{F}(x) = (1/n) \sum_{i=1}^n X_i$$

$$\bullet \vec{p}^* \equiv \hat{F}^* \equiv \begin{pmatrix} p_1^* & \cdots & p_n^* \\ X_1 & \cdots & X_n \end{pmatrix} \rightarrow \theta(\vec{p}^*) = \theta(\hat{F}^*) = \int x d\hat{F}^*(x) = \sum_{i=1}^n p_i^* X_i$$

$$\bullet \vec{p}_{(i)} \equiv \hat{F}_{(i)} \equiv \begin{pmatrix} 1/n-1 & 0 & 1/n-1 \\ X_1 & X_i & X_n \end{pmatrix} \rightarrow \theta(\vec{p}_{(i)}) = \theta(\hat{F}_{(i)}) = \int x d\hat{F}_{(i)}(x) = \sum_{j \neq i} \frac{1}{n-1} X_j$$

Fijada una m.a.s. particular  $\vec{x} = (x_1, \dots, x_n)^t$  esa superficie es determinista.

Diremos que  $\theta(\vec{p})$  es un **funcional lineal** si:

$$\theta(\vec{p}) = a + (\vec{p} - \vec{p}_0)^t b \quad \left( a \in \mathbb{R}, b \in \mathbb{R}^n / \sum_{i=1}^n b_i = 0 \right)$$

Diremos que  $\theta(\vec{p})$  es un **funcional cuadrático** si:

$$\theta(\vec{p}) = a + (\vec{p} - \vec{p}_0)^t b + \frac{1}{2} (\vec{p} - \vec{p}_0)^t c (\vec{p} - \vec{p}_0) \quad \left( \begin{array}{l} a \in \mathbb{R} \\ b \in \mathbb{R}^n / \vec{p}_0^t b = 0 \\ c = (c_{ij}) / \vec{p}_0^t c = 0 \\ c \ n \times n, \text{ simétrica} \end{array} \right)$$

## TEOREMA 1

Si  $\hat{\theta} = S^2$  (cualquier funcional cuadrático en general), entonces

$$\overbrace{\text{Sesgo}_F \hat{\theta}}^{jackk} = \frac{n}{n-1} \overbrace{\text{Sesgo}_F \hat{\theta}}^{boot}$$

**TEOREMA 2**

Si  $\hat{\theta} = \bar{X}$  (cualquier funcional lineal en general), entonces

$$\widehat{Var}_F \hat{\theta}^{jackk} = \frac{n}{n-1} \widehat{Var}_F \hat{\theta}^{boot}$$

**Ejemplo:**

- $\widehat{Var}_F \bar{X}^{jackk} = \frac{1}{n} \left( \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \right)$  insesgado en  $\frac{\sigma^2}{n}$
- $\widehat{Var}_F \bar{X}^{boot} = Var_{\hat{F}} \bar{X}^* = Var_{\hat{F}} \frac{1}{n} \sum_{i=1}^n X_i^* = \frac{n}{n^2} Var_{\hat{F}} X^* =$   
 $= \frac{1}{n} \left( \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right) = \frac{S^2}{n}$  sesgado en  $\frac{\sigma^2}{n}$

ESTUDIO DE SIMULACIÓN (jackknife / bootstrap como estimadores de varianzas)

$(X_1, \dots, X_n)$  m.a.s. de  $X \sim N(0,1)$ ;  $\hat{\theta}_d = \theta(X_1, \dots, X_n) = \frac{1}{n-2[nd]} \sum_{i=[nd+1]}^{n-[nd]} X_{(i)}$

	VARIANZA TEÓRICA	VARIANZA INCLUIVISE	VARIANZA BOOTSTRAP
MEDIA $d=0$	$1/n$	$S^2/n$ (coarsi-var) insesgado	$S_n^2/n$ (var) sesgado
MEDIANA $d \rightarrow 1/2$	$\pi/2n$	$\frac{n-1}{n} \sum_{i=1}^n (\hat{\theta}_{1/2}(\omega) - \hat{\theta}_{1/2}(\omega))^2$	$\sum_{k=1}^n (X_{(k)} - E_x)^2 P_k$ (1)
$\alpha$ $g/2$	$\frac{F_{\Gamma(g/2, 1/2)} \left( \frac{1}{2} (F_{N(0,1)}^{-1}(1-2d))^2 \right) + 2d (F_{N(0,1)}^{-1}(1-2d))^2}{n(1-2d)^2}$	$\left\{ \begin{array}{l} \text{Si } (n-1)d = g \text{ entero} \\ \bullet \frac{\sum_{i=1}^n (X_{(i)}(\omega) - \bar{X}_{(i)})^2}{n(n-1)(1-2d)^2} \quad (2) \\ \omega_i = \begin{cases} g+1 & i \leq g+1 \\ i & g+1 < i < n-g \\ n-g & i > n-g \end{cases} \\ \bullet \frac{n-1}{n} \sum_{i=1}^n (\hat{\theta}_d(\omega) - \hat{\theta}_d(\omega))^2 \quad (3) \end{array} \right.$	$\frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_{d,b}^* - \hat{\theta}_{d,1}^*)^2$

(1)  $P_k = \sum_{j=0}^{m-1} \left[ \binom{n}{j} \left(\frac{k-1}{n}\right)^j \left(1 - \frac{k-1}{n}\right)^{n-j} - \binom{n}{j} \left(\frac{k}{n}\right)^j \left(1 - \frac{k}{n}\right)^{n-j} \right] = P_x(\hat{\theta}_{1/2}^* = x_{(k)})$ ;  $E_x = \sum_{k=1}^n X_{(k)} \cdot P_k$

(2)=(3) si  $(n-1)d = g$  entero

Tables 1.1 and 1.2 present the results of a simulation study comparing the bootstrap and jackknife approximations to the variance of the sample mean, median and  $\alpha$ -trimmed means for a  $N(0, 1)$  population. We have used (2.6) and (2.12) and the fact that, in this case, the theoretical variance of the mean is  $\frac{1}{n}$  and the asymptotic variances of the sample median and  $\alpha$ -trimmed means are  $\frac{\pi}{2n}$  and

$$\frac{F_{\gamma(1,3/2)}\left(\frac{(\Phi^{-1}(1-\alpha))^2}{2}\right) + 2\alpha(\Phi^{-1}(1-\alpha))^2}{n(1-2\alpha)^2},$$

respectively, where  $F_{\gamma(1,3/2)}$  and  $\Phi$  are the distribution functions for the gamma and normal distributions, respectively. Both tables give the exact variance of the estimates and the means and mean squared errors —between parentheses— of the jackknife and bootstrap approximations for 1000 simulated samples of size 11 (Table 1.1) and 21 (Table 1.2) with  $B = 200$  bootstrap replications.

	Exact	Jackknife	Bootstrap
Mean Var.	0.0909	0.0904 (0.0016)	0.0821 (0.0014)
Median Var.	0.1427	0.1685 (0.0415)	0.1738 (0.0148)
0.1-mean Var.	0.0963	0.0978 (0.0025)	0.0989 (0.0021)
0.2-mean Var.	0.1040	0.1103 (0.0052)	0.1099 (0.0031)
0.3-mean Var.	0.1138	0.1267 (0.0110)	0.1222 (0.0050)
0.4-mean Var.	0.1263	0.1685 (0.0426)	0.1400 (0.0084)

Table 1.1

	Exact	Jackknife	Bootstrap
Mean Var.	0.0476	0.0475 (0.0002)	0.0452 (0.0002)
Median Var.	0.0747	0.1078 (0.0227)	0.0895 (0.0031)
0.1-mean Var.	0.0504	0.0506 (0.0003)	0.0520 (0.0003)
0.2-mean Var.	0.0545	0.0554 (0.0006)	0.0563 (0.0005)
0.3-mean Var.	0.0596	0.0636 (0.0014)	0.0625 (0.0008)
0.4-mean Var.	0.0661	0.0793 (0.0054)	0.0718 (0.0016)

Table 1.2

As expected, the results show a better behavior of the bootstrap approximation in mean squared error, improving with  $\alpha$ .

```
#Para la muestra de TIEMPOS DE VIDA, estima (plug-in) la precisión de la media muestral
#(desvmedia) y también estima mediante el jackknife dicha precisión
#(desvmediajackk) y también la precisión de la mediana muestral, de la cual no se conoce su
#expresión, (desvmedianaajackk). También estima el sesgo jackknife de esos dos estimadores
#(sesgomedianaajackk y sesgomedianaajackk, respectivamente).
#####33
```

```
muestra<-c(0.143,0.182,0.256,0.260,0.270,0.437,0.509,0.611,0.712,1.04,1.09,1.15,1.46,1.88,2.08);
n=length(muestra);
varmedia<-(1/(n^2))*sum((muestra-mean(muestra))^2);
desvmedia=sqrt(varmedia)
imuestra<-numeric(n-1);
media<-numeric(n)
remordenada<-numeric(n-1)
mediana<-numeric(n)
```

```
for (i in 1:n)
{
  imuestra=muestra[-i]
  media[i]=mean(imuestra)
  remordenada=sort(imuestra)
  mediana[i]=(remordenada[7]+remordenada[8])/2
}
```

```
varmediajackk=((n-1)/n)*sum((media-mean(media))^2);
desvmediajackk=sqrt(varmediajackk);
varmedianaajackk=((n-1)/n)*sum((mediana-mean(mediana))^2);
desvmedianaajackk=sqrt(varmedianaajackk);
desvmedia
desvmediaajackk
desvmedianaajackk
sesgomedianaajackk=(n-1)*(mean(media)-mean(muestra))
sesgomedianaajackk=(n-1)*(mean(mediana)-muestra[8])
sesgomedianaajackk
sesgomedianaajackk
```

	$\bar{x} = 0,804$	$x_{(8)} = 0,611$
Estimación precisión (plug-in)	0,155579	?
Estimación precisión (jackknife)	0,161040	0,183451
Estimación sesgo (jackknife)	0	-0,003733

**Recordemos que:**

	$\bar{x} = 0,804$	$x_{(8)} = 0,611$
Estimación precisión (plug-in)	0,155579	?
Estimación precisión (bootstrap)	0,155591	0,250086
Estimación sesgo (bootstrap)	0,005875	0,055414

# MODIFICACIONES DEL BOOTSTRAP UNIFORME

1. Bootstrap paramétrico.
2. Bootstrap simetrizado.
3. Bootstrap suavizado.
4. Bootstrap ponderado y sesgado.
5. Deficiencias del bootstrap uniforme.

## Aplicación:

- Aproximación de la distribución de un estadístico (de distribución conocida) mediante el bootstrap uniforme, paramétrico, simetrizado y suavizado. Comparación de las aproximaciones con la distribución exacta. **P8**

En contextos de inferencia con información parcial acerca del mecanismo que genera los datos, podemos utilizar mejores estimadores que  $\hat{F}$  para aproximar  $F$ .

## BOOTSTRAP PARAMÉTRICO

Si suponemos que  $F = F_\theta$  (modelo paramétrico), podemos estimarlo con  $F_{\hat{\theta}_{m.v.}}$ , siendo  $\hat{\theta}_{m.v.}$  el estimador de máxima verosimilitud de  $\theta$ , construido con los datos observados.

El planteamiento bootstrap resultante de reemplazar  $\hat{F}$  por  $F_{\hat{\theta}_{m.v.}}$  se conoce como *bootstrap paramétrico*.

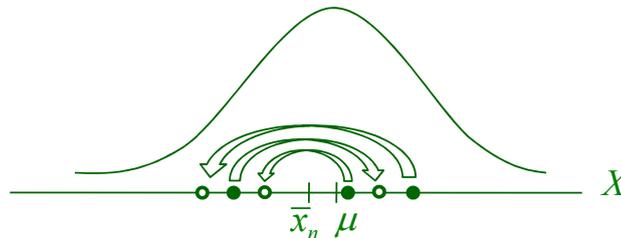
$$\begin{array}{ccc} X \in F_\theta & \rightarrow & D[R(\vec{X}; F_\theta)] \\ \uparrow (x_1, \dots, x_n) & & \\ X^* \in F_{\hat{\theta}_{m.v.}} & \rightarrow & D[R(\vec{X}^*; F_{\hat{\theta}_{m.v.}})] \end{array}$$

**NOTA.** La aproximación por Monte Carlo, en este caso, se hace a partir de un modelo continuo (si  $F_\theta$  es continua).

## BOOTSTRAP SIMETRIZADO

Si suponemos que  $F$  es un modelo simétrico ( $X$  simétrica respecto de  $\mu$ ), como los datos  $(x_1, \dots, x_n)$  no tienen por qué reflejar esta propiedad (simetría respecto de su  $\bar{x}_n$ ), podemos simetrizarlos duplicando la muestra inicial de la siguiente forma:

$$y_j = \begin{cases} x_j & j = 1, \dots, n \\ 2\bar{x}_n - x_{j-n} & j = n+1, \dots, 2n \end{cases}$$



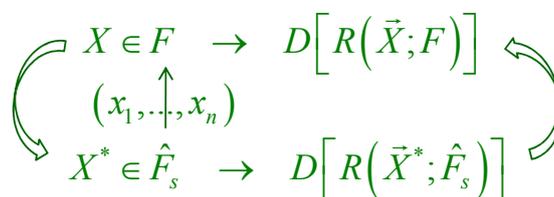
Notemos que:

$$\bar{y}_{2n} = \frac{1}{2n} \sum_{j=1}^{2n} y_j = \frac{1}{2n} \left[ \sum_{j=1}^n x_j + \sum_{j=n+1}^{2n} (2\bar{x}_n - x_{j-n}) \right] = \frac{1}{2n} [n\bar{x}_n + 2n\bar{x}_n - n\bar{x}_n] = \bar{x}_n$$

El planteamiento bootstrap resultante de reemplazar  $\hat{F}$  por  $\hat{F}_s$  (distribución empírica de la muestra simetrizada)

$$\hat{F}_s(x) = \frac{1}{2n} \sum_{j=1}^{2n} I_{\{y_j \leq x\}}$$

se conoce como *bootstrap simetrizado*.



### NOTA

- Al simetrizar “intervengo” en los datos: puedo, tengo información adicional en ese sentido y la utilizo.
- Esta simetrización de los datos altera mínimamente su modelo empírico, pues conserva la media (como vimos) y sus momentos centrales pares (los impares claramente se anulan).
- La aproximación por Monte Carlo se hace a partir de los nuevos  $2n$  datos.

## BOOTSTRAP SUAVIZADO

Si suponemos que  $F$  es un modelo absolutamente continuo con densidad  $f$ , podemos estimarlo mediante la integral del estimador no paramétrico de Parzen Rosemblat de  $f$ , es decir

$$\hat{F}_h(x) = \int_{-\infty}^x \hat{f}_h(t) dt$$

El planteamiento bootstrap resultante de reemplazar  $\hat{F}$  por  $\hat{F}_h$  se conoce como *bootstrap suavizado*.

$$\begin{array}{ccc} X \in F & \rightarrow & D[R(\bar{X}; F)] \\ \uparrow & & \uparrow \\ (x_1, \dots, x_n) & & \\ \downarrow & & \downarrow \\ X^* \in \hat{F}_h & \rightarrow & D[R(\bar{X}^*; \hat{F}_h)] \end{array}$$

Para aproximar por Monte Carlo la distribución de  $R(\bar{X}^*; \hat{F}_h)$  notemos que:

$$\hat{f}_h(t) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{t-x_i}{h}\right) = \frac{1}{n} \sum_{i=1}^n K_h(t-x_i) \quad \left[ K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right) \right]$$

Por lo tanto:

$$\hat{F}_h(x) = \int_{-\infty}^x \hat{f}_h(t) dt = \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^x K_h(t-x_i) dt = \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{x-x_i} K_h(z) dz$$

Por otro lado:

$Y \equiv$  variable aleatoria con función de densidad  $K(y)$

$$\bullet T = hY \rightarrow \left[ y = \frac{t}{h} \Rightarrow \frac{dy}{dt} = \frac{1}{h} \rightarrow K_h(t) = \frac{1}{h} K\left(\frac{t}{h}\right) \text{ es la densidad de } T \right]$$

$$\bullet X_0 \equiv \hat{F} \equiv \begin{pmatrix} 1/n & \cdots & 1/n \\ x_1 & \cdots & x_n \end{pmatrix}$$

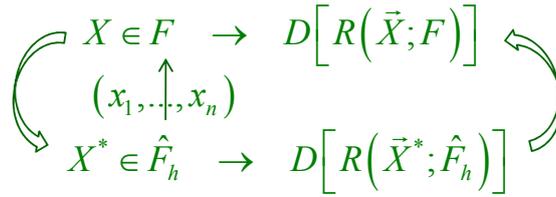
$$P(hY + X_0 \leq x) = \sum_{i=1}^n P\left(hY \leq x - X_0 \middle| X_0 = x_i\right) P(X_0 = x_i) = \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{x-x_i} K_h(z) dz$$

Es decir,

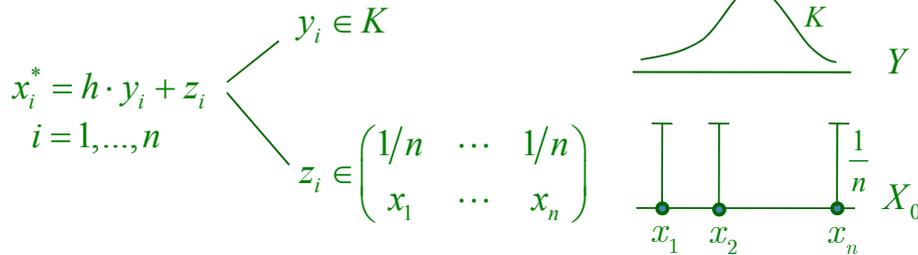
$$\hat{F}_h \text{ es la función de distribución de } hY + X_0$$

(Convolución de las distribuciones de los datos y del núcleo reescalado)

**Remuestreo bootstrap suavizado**



$\vec{x}^* = (x_1^*, \dots, x_n^*)$  m.a.s. de  $X^*$ ?



De esta forma generariamos tantos valores de  $R(\vec{X}^*; \hat{F}_h)$  como quisiéramos.

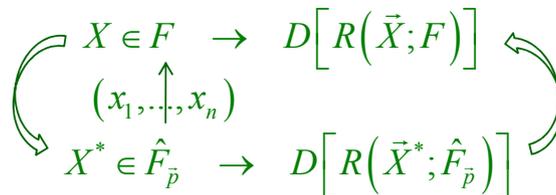
$$b = 1, \dots, B \rightarrow \vec{x}^{*b} = (x_1^*, \dots, x_n^*)^b \rightarrow r^{*b} = R(\vec{x}^{*b}; \hat{F}_h)$$

**BOOTSTRAP PONDERADO Y SESGADO**

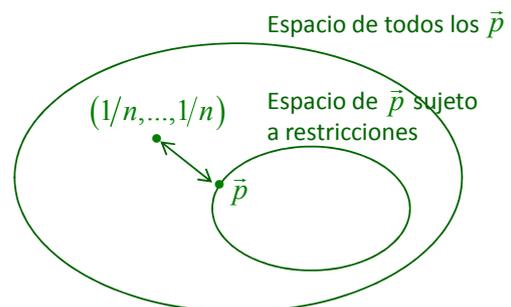
- El planteamiento bootstrap resultante de reemplazar  $\hat{F} = \begin{pmatrix} 1/n & \dots & 1/n \\ x_1 & \dots & x_n \end{pmatrix}$  por

$$\hat{F}_{\vec{p}} = \begin{pmatrix} p_1 & \dots & p_n \\ x_1 & \dots & x_n \end{pmatrix} \begin{cases} \sum_{i=1}^n p_i = 1 \\ p_i \geq 0, i = 1, \dots, n \end{cases}$$

se conoce como *bootstrap ponderado*.



- Se obtienen interesantes propiedades estadísticas (Hall, 98) si se elige  $\vec{p}$  minimizando la distancia a  $(1/n, \dots, 1/n)$  sujeto a un conjunto de restricciones estadísticas (*bootstrap sesgado*)



## DEFICIENCIAS DEL BOOTSTRAP UNIFORME

### Ejemplo de mal funcionamiento (estimando una función de densidad)

- Supongamos un contexto paramétrico  $F_\theta$  y un estadístico con distribución exacta conocida salvo por el valor de  $\theta$  (sabemos que en contexto paramétrico no hay problemas en ese sentido).

$$X \in U(0, \theta) \quad \left| \quad \begin{array}{l} f_\theta(x) = 1/\theta \quad x \in (0, \theta) \\ F_\theta(x) = x/\theta \quad x \in (0, \theta) \end{array} \right.$$

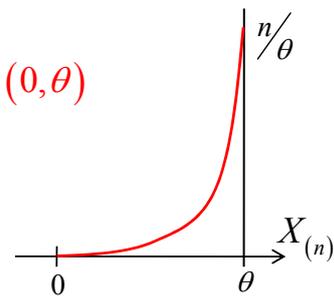
↓

$$\vec{X} = (X_1, \dots, X_n)^t \rightarrow \hat{\theta} = \theta(\vec{X}) = X_{(n)} \quad (\text{máximo muestral})$$

*m.a.s. de X*

$$g_\theta(x) = n f_\theta(x) F_\theta(x)^{n-1} = \frac{n}{\theta} \left( \frac{x}{\theta} \right)^{n-1} \quad x \in (0, \theta)$$

Distribución exacta de  $X_{(n)}$



- Fijemos  $\theta = 1$  (conocimiento total de  $F$ ) y obtengamos por Monte Carlo una muestra artificial de tamaño 50,  $\vec{x} = (x_1, \dots, x_{50})$ , de esa población (supongamos que su máximo valor fue  $x_{(50)} = 0,988$ ).
- Olvidémonos ahora de que conocemos  $F$  y utilicemos el bootstrap uniforme (no paramétrico) y el paramétrico sobre la muestra artificial para estimar la distribución de  $X_{(n)}$ , es decir, para "reconstruir  $g$ ". La distribución bootstrap la aproximamos por Monte Carlo ( $B=2000$ ). Dicho de otra forma, procedamos como si fuera una muestra natural en contextos no paramétrico o paramétrico.

#### Bootstrap no paramétrico

$$X^* \equiv \begin{pmatrix} 1/50 & \dots & 1/50 \\ x_1 & \dots & x_{50} \end{pmatrix}$$

↓

$$\vec{x}^{*b} = (x_1^*, \dots, x_{50}^*)^{t,b} \rightarrow x_{(50)}^{*b}$$

$b = 1, \dots, 2000$

#### Bootstrap paramétrico

$$X^* \in U(0, \hat{\theta}_{m.v.} = 0,988)$$

↓

$$\vec{x}^{*b} = (x_1^*, \dots, x_{50}^*)^{t,b} \rightarrow x_{(50)}^{*b}$$

$b = 1, \dots, 2000$

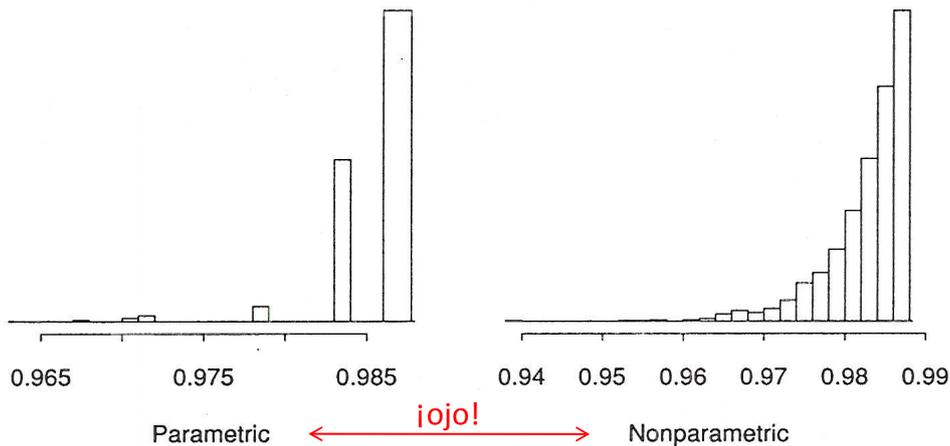
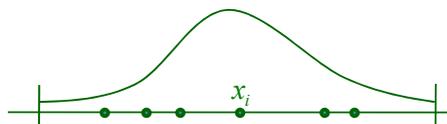


Figure 7.11. The left panel shows a histogram of 2000 bootstrap replications of  $\hat{\theta}^* = X_{(n)}$  obtained by sampling with replacement from a sample of 50 uniform numbers. The right panel shows 2000 parametric bootstrap replications obtained by sampling from the uniform distribution on  $(0, \hat{\theta})$ .

### OBSERVACIONES

- En general, el bootstrap uniforme va a funcionar mal aproximando la distribución de estadísticos que persigan extremos de un soporte, dado el carácter finito de la muestra (incluso si  $n \rightarrow \infty$ ).



- La variable aleatoria  $X_{(50)}^*$  toma como valores los datos muestrales, por lo que es discreta ( $n$  fijo), y por lo tanto inadecuada para aproximar a otra continua.
- Además, todas las réplicas bootstrap de la muestra inicial que contengan al máximo o al mínimo muestral ( $x_{(n)}$  ó  $x_{(1)}$ ) van a proporcionar ese valor para la variable  $X_{(50)}^*$ , lo que le otorgará un “protagonismo excesivo” (aunque  $n \rightarrow \infty$ ):

$$\begin{aligned}
 \bullet P\left(X_{(50)}^{*unif} = x_{(50)}\right) &= P\left(X_{(50)}^{*unif} \leq x_{(50)}\right) - P\left(X_{(50)}^{*unif} \leq x_{(49)}\right) = \\
 &= 1 - \left(1 - \frac{1}{n}\right)^n \xrightarrow{n \rightarrow \infty} 1 - e^{-1} = 0,632 \text{ ¡!} \\
 \bullet P\left(X_{(50)}^{*par} = x_{(50)}\right) &= 0
 \end{aligned}$$

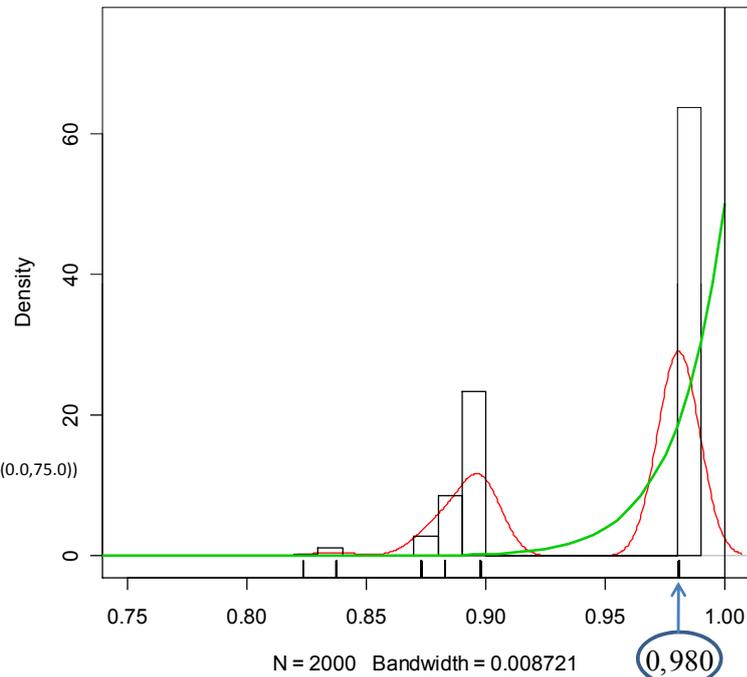
## BOOTSTRAP NO PARAMÉTRICO

Histograma de las 2000 réplicas bootstrap no paramétrico de  $X_{(50)}$

VERDE: Distribución exacta de  $X_{(50)}$

ROJO: Estimación no paramétrica de la densidad de  $X_{(50)}$ , sobre esas 2000 réplicas bootstrap

```
muestra<-runif(50)
n=length(muestra);
remuestra<-numeric(n);
B=2000
maximo<-numeric(B)
for (k in 1:B)
{
  for (i in 1:n)
  {
    j=1
    u=runif(1)
    while (u>=j/n)
    {
      j=j+1
    }
    remuestra[i]=muestra[j];
  }
  maximo[k]=max(remuestra)
}
plot(density(maximo),col=2,xlim=c(0.0,1.0),ylim=c(0.0,75.0))
hist(maximo,freq=FALSE,add=TRUE,lwd=3)
rug(maximo)
y<-seq(0,1,length=200)
lines(y,50*y^49,col=3,lwd=2)
max(muestra)
abline(v=1.0)
```



## OBSERVACIONES

- En realidad, no es muy adecuado utilizar la estimación no paramétrica de la densidad de  $X_{(50)}^*$ , pues tal variable (como ya se indicó) es discreta. Sin embargo, “puede hacerse”, pues la distribución que se busca es absolutamente continua.
- Todas las réplicas bootstrap de la muestra inicial que contengan al máximo muestral,  $x_{(50)}$ , van a proporcionar ese valor para la variable  $X_{(50)}^*$ . Por otro lado, las que contengan a  $x_{(49)}$  y no a  $x_{(50)}$  (menos que antes), proporcionarán para la variable  $X_{(50)}^*$  el valor  $x_{(49)}$ .

## NOTA

- Como se observa en la siguiente imagen, el funcionamiento del bootstrap paramétrico en este caso es mucho mejor (ya conocido).
- Para el bootstrap simetrizado son válidas en general las observaciones relativas al bootstrap no paramétrico, en este caso.

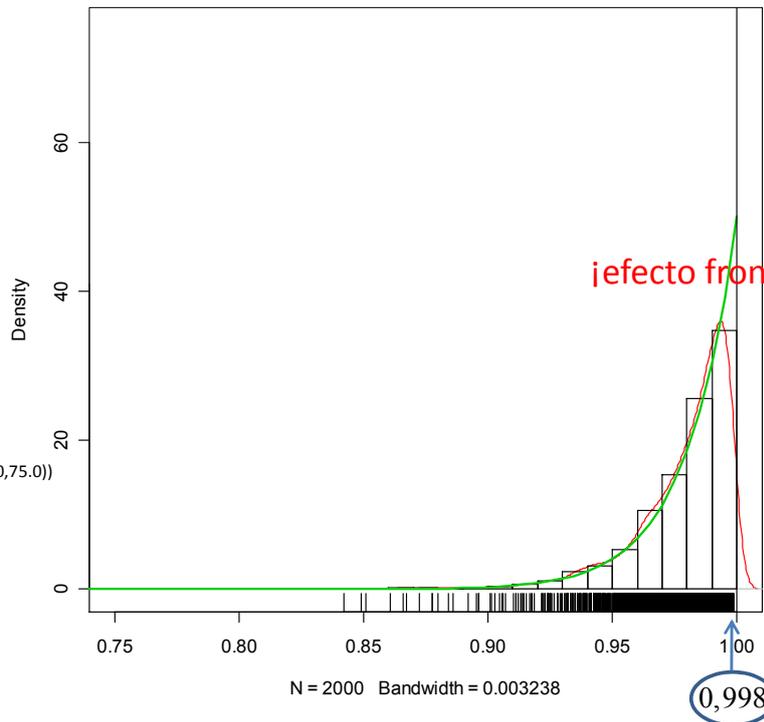
## BOOTSTRAP PARAMÉTRICO

Histograma de las 2000 réplicas bootstrap paramétrico de  $X_{(50)}$

VERDE: Distribución exacta de  $X_{(50)}$

ROJO: Estimación no paramétrica de la densidad de  $X_{(50)}$ , sobre esas 2000 réplicas bootstrap

```
muestra<-runif(50)
n=length(muestra);
remuestra<-numeric(n);
B=2000
maximo<-numeric(B)
for (k in 1:B)
{
  remuestra<-runif(n,0,max(muestra));
  maximo[k]=max(remuestra)
}
plot(density(maximo),col=2,xlim=c(0.75,1.0),ylim=c(0.0,75.0))
hist(maximo,freq=FALSE,add=TRUE,lwd=3)
rug(maximo)
y<-seq(0,1,length=200)
lines(y,50*y^49,col=3,lwd=2)
max(muestra)
abline(v=1.0)
```



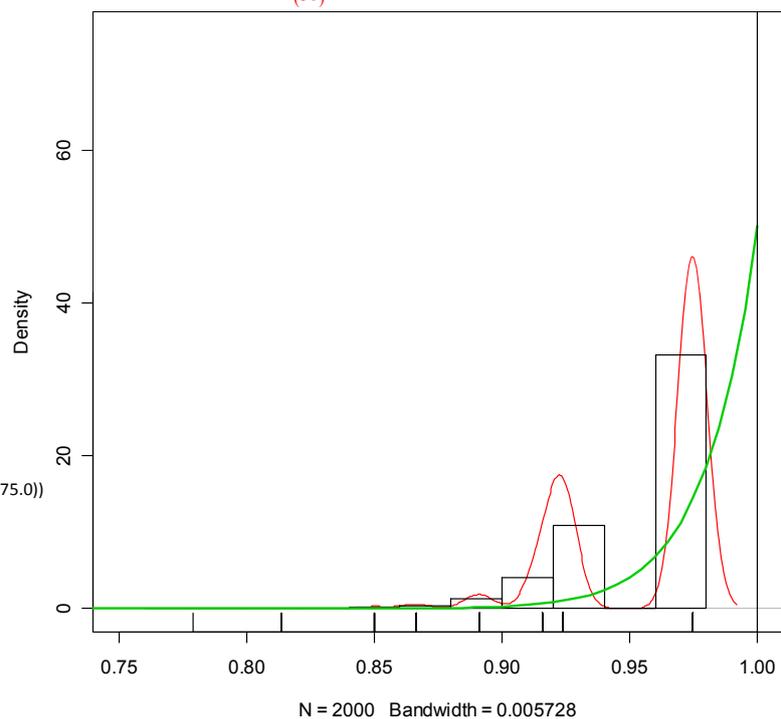
## BOOTSTRAP SIMETRIZADO

Histograma de las 2000 réplicas bootstrap simetrizado de  $X_{(50)}$

VERDE: Distribución exacta de  $X_{(50)}$

ROJO: Estimación no paramétrica de la densidad de  $X_{(50)}$ , sobre esas 2000 réplicas bootstrap

```
muestra<-runif(50)
muestrasim<-c(muestra,2*mean(muestra)-muestra)
n=length(muestrasim);
remuestra<-numeric(n);
B=2000
maximo<-numeric(B)
for (k in 1:B)
{
  for (i in 1:n)
  {
    j=1
    u=runif(1)
    while (u>=j/n)
    {
      j=j+1
    }
    remuestra[i]=muestrasim[j];
  }
  maximo[k]=max(remuestra)
}
plot(density(maximo),col=2,xlim=c(0.75,1.0),ylim=c(0.0,75.0))
hist(maximo,freq=FALSE,add=TRUE,lwd=3)
rug(maximo)
y<-seq(0,1,length=200)
lines(y,50*y^49,col=3,lwd=2)
max(muestra)
abline(v=1.0)
```

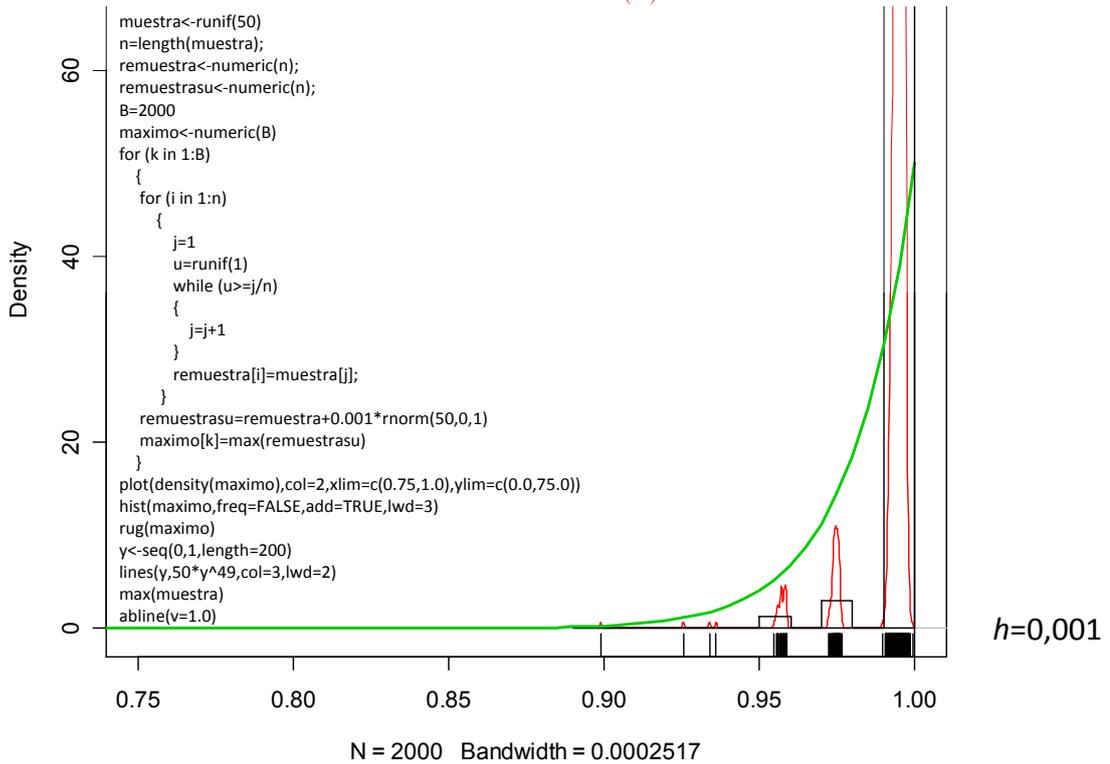


### BOOTSTRAP SUAVIZADO

Histograma de las 2000 réplicas bootstrap suavizado de  $X_{(50)}$

VERDE: Distribución exacta de  $X_{(50)}$

ROJO: Estimación no paramétrica de la densidad de  $X_{(50)}$ , sobre esas 2000 réplicas bootstrap

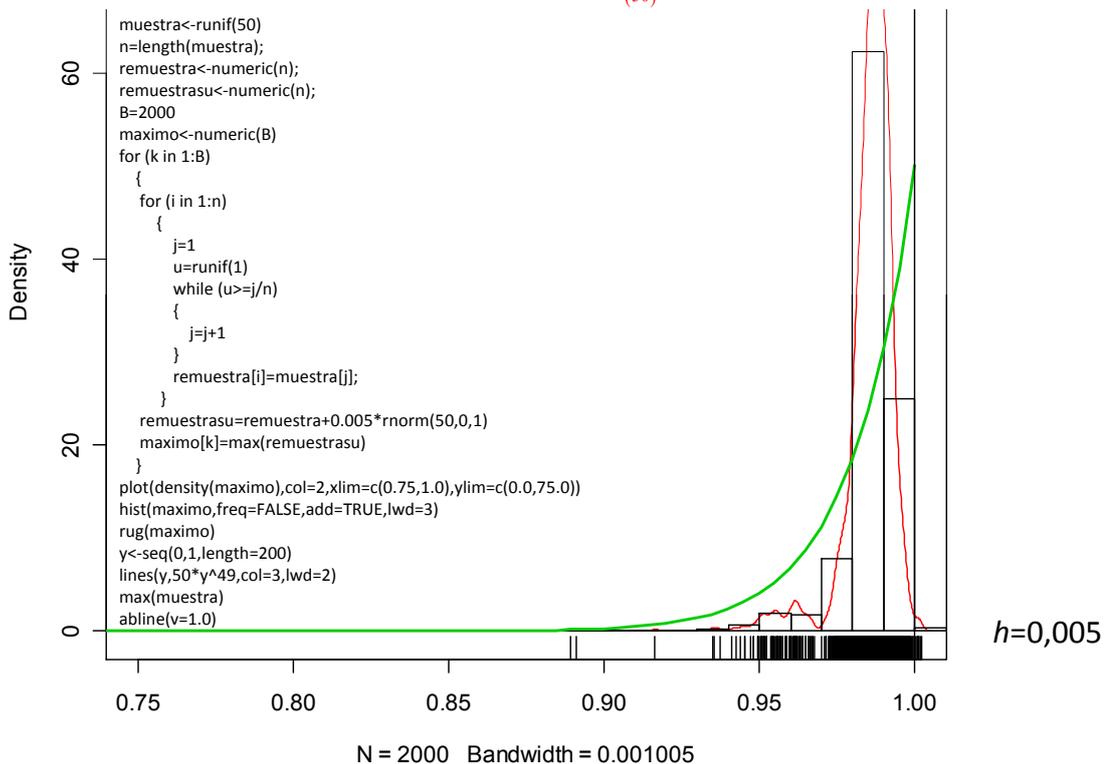


### BOOTSTRAP SUAVIZADO

Histograma de las 2000 réplicas bootstrap suavizado de  $X_{(50)}$

VERDE: Distribución exacta de  $X_{(50)}$

ROJO: Estimación no paramétrica de la densidad de  $X_{(50)}$ , sobre esas 2000 réplicas bootstrap

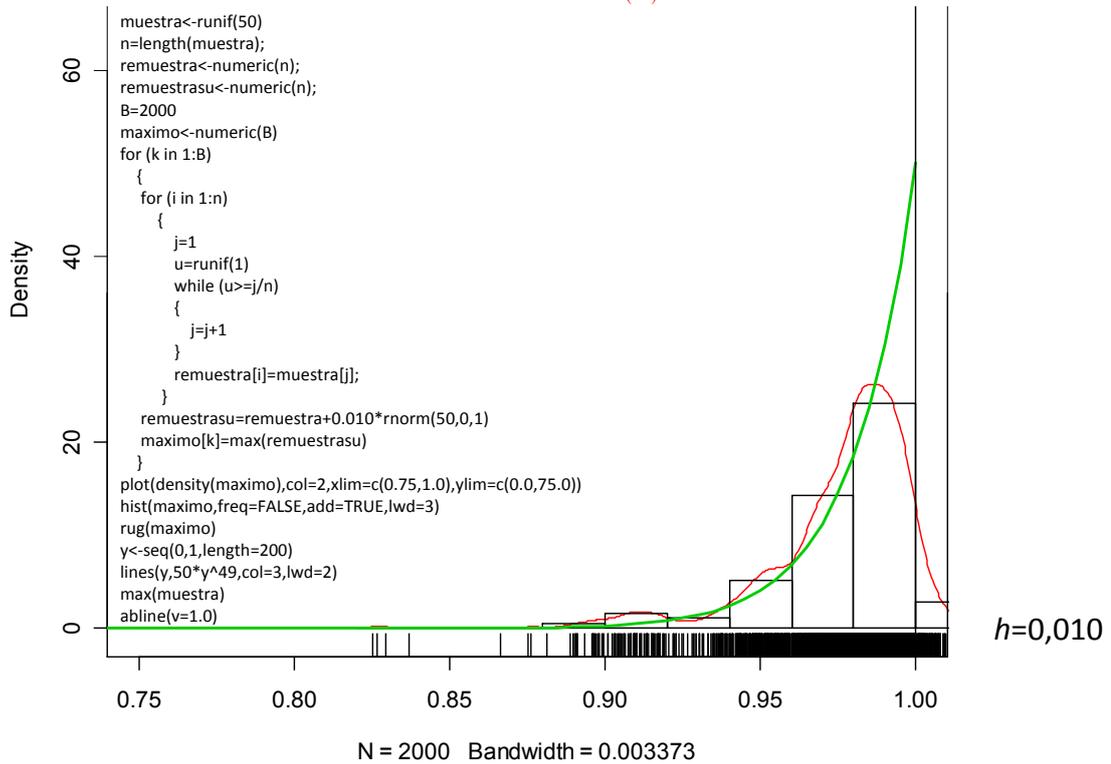


### BOOTSTRAP SUAVIZADO

Histograma de las 2000 réplicas bootstrap suavizado de  $X_{(50)}$

VERDE: Distribución exacta de  $X_{(50)}$

ROJO: Estimación no paramétrica de la densidad de  $X_{(50)}$ , sobre esas 2000 réplicas bootstrap

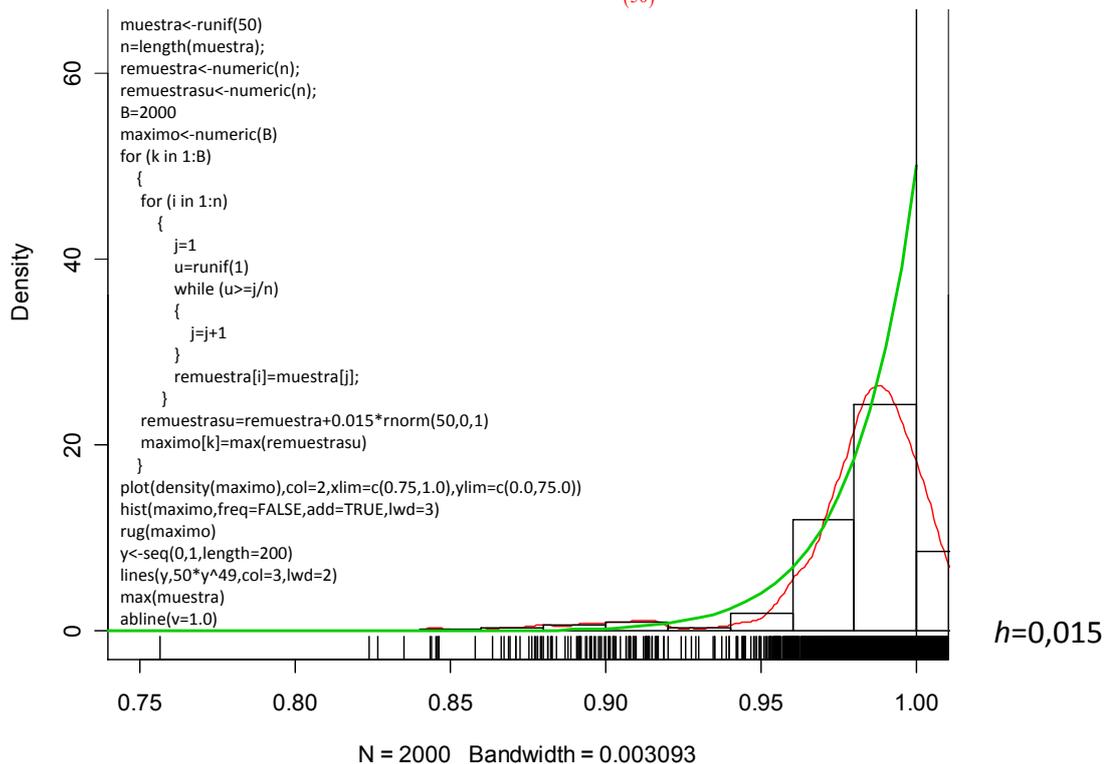


### BOOTSTRAP SUAVIZADO

Histograma de las 2000 réplicas bootstrap suavizado de  $X_{(50)}$

VERDE: Distribución exacta de  $X_{(50)}$

ROJO: Estimación no paramétrica de la densidad de  $X_{(50)}$ , sobre esas 2000 réplicas bootstrap

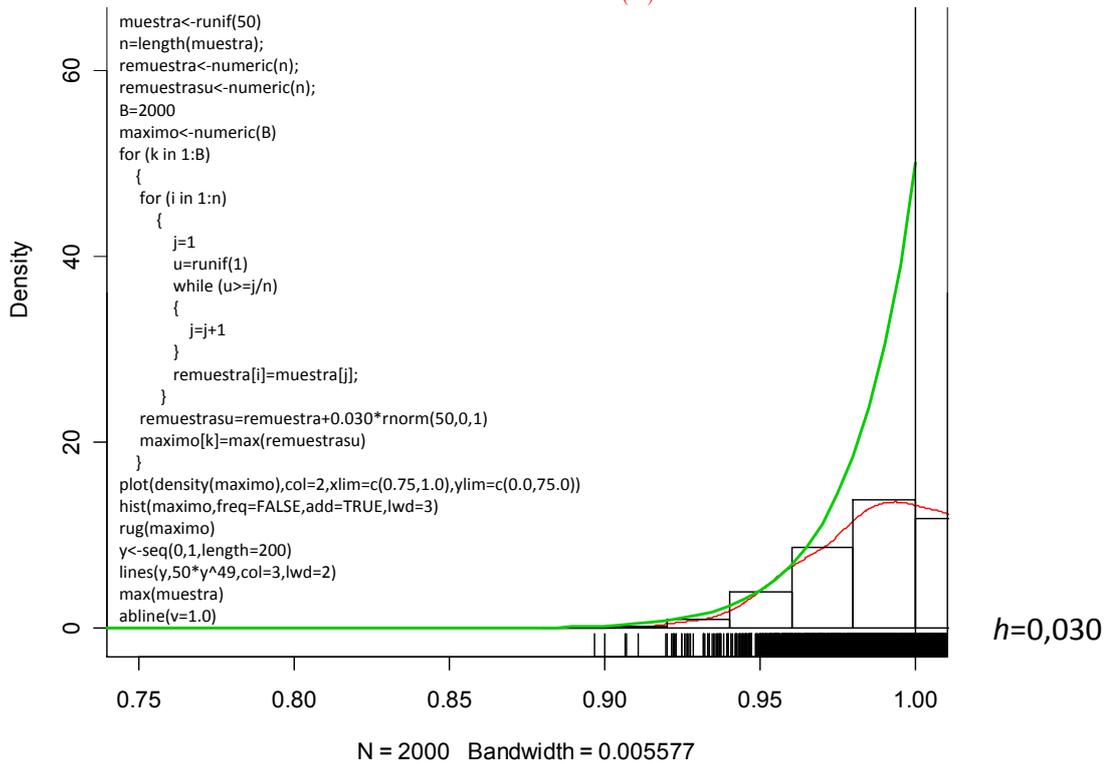


## BOOTSTRAP SUAVIZADO

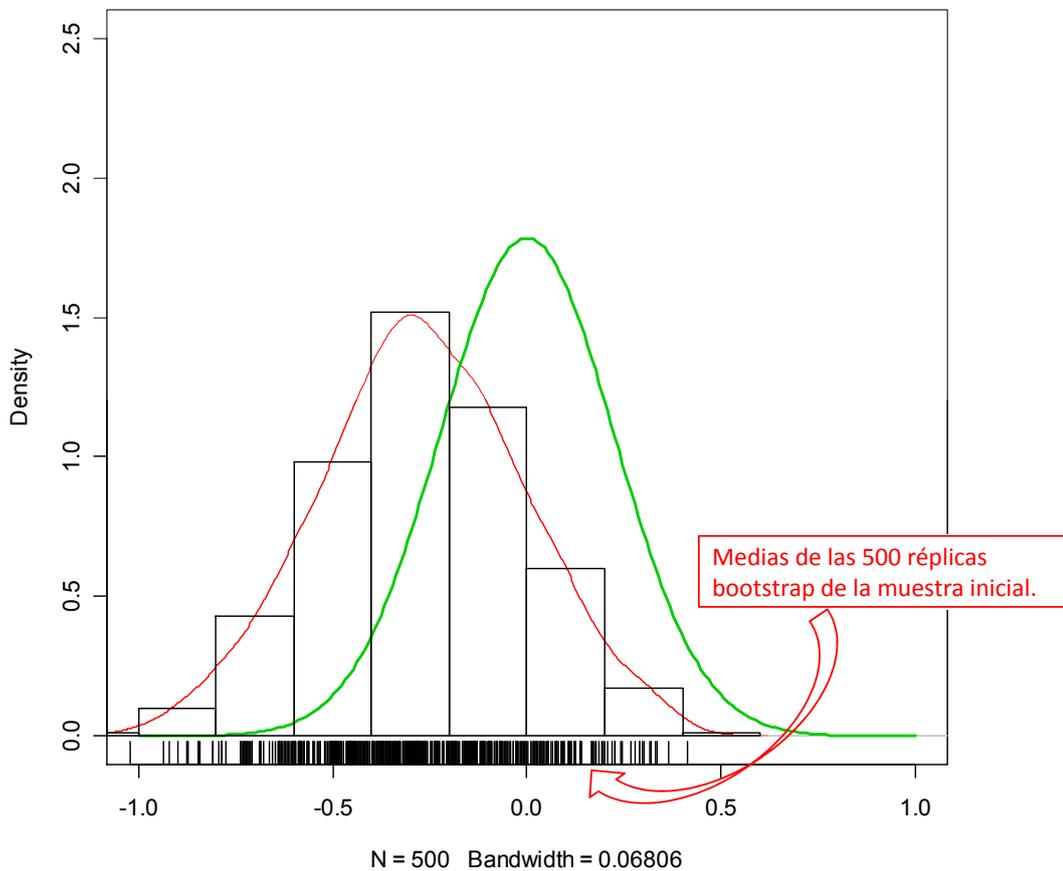
Histograma de las 2000 réplicas bootstrap suavizado de  $X_{(50)}$

VERDE: Distribución exacta de  $X_{(50)}$

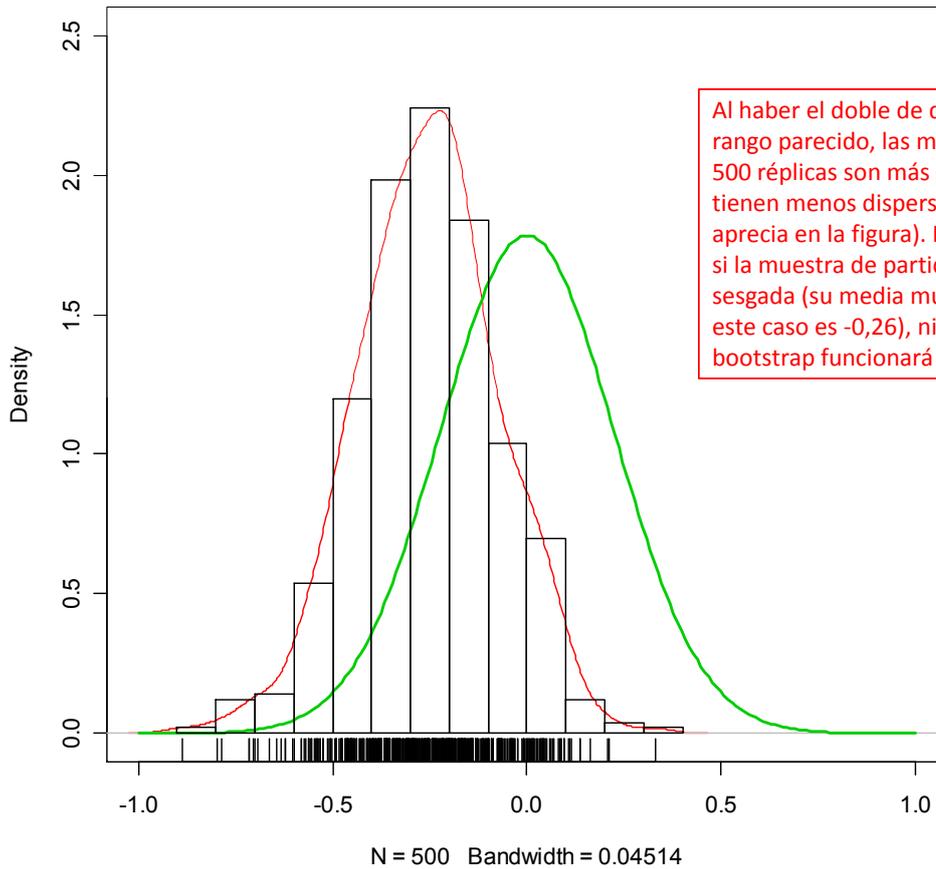
ROJO: Estimación no paramétrica de la densidad de  $X_{(50)}$ , sobre esas 2000 réplicas bootstrap



**Bootstrap no paramétrico** aproximando la densidad de la media muestral sobre una muestra de tamaño 20 (sesgada: media=-0,26) de una población normal (0,1).



**Bootstrap simetrizado (no paramétrico)** aproximando la densidad de la media muestral sobre una muestra de tamaño 20 (mala) de una población normal (0,1).



Al haber el doble de datos en un rango parecido, las medias de las 500 réplicas son más parecidas, tienen menos dispersión (se aprecia en la figura). No obstante, si la muestra de partida es sesgada (su media muestral en este caso es -0,26), ningún bootstrap funcionará bien.

## Comparación aproximaciones normal y bootstrap de una función de distribución

```
m=13
n=2*m-1
g=function(x,m=13,F=pexp,f=dexp){choose(2*m-1,m)*m*F(x)^(m-1)*(1-F(x))^(m-1)*f(x)}
curve(g(x,m=13,F=pexp,f=dexp),from=0,to=1.7)
G=function(lim){integrate(g,lower=0,upper=lim)$value}
x=seq(0,1.7,length=100)
Gf=numeric(length(x))
for (i in 1:length(x)){
  Gf[i]=G(x[i])
}
plot(x,Gf,type="l",ylim=c(0,1),lwd=2)
```

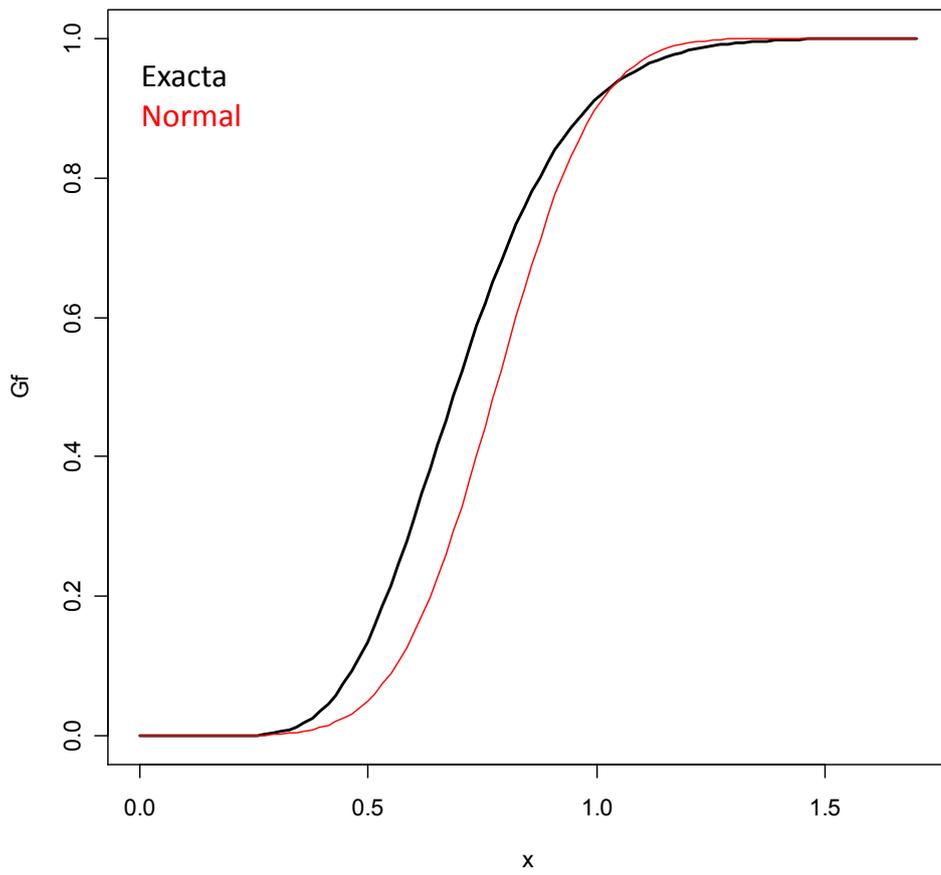
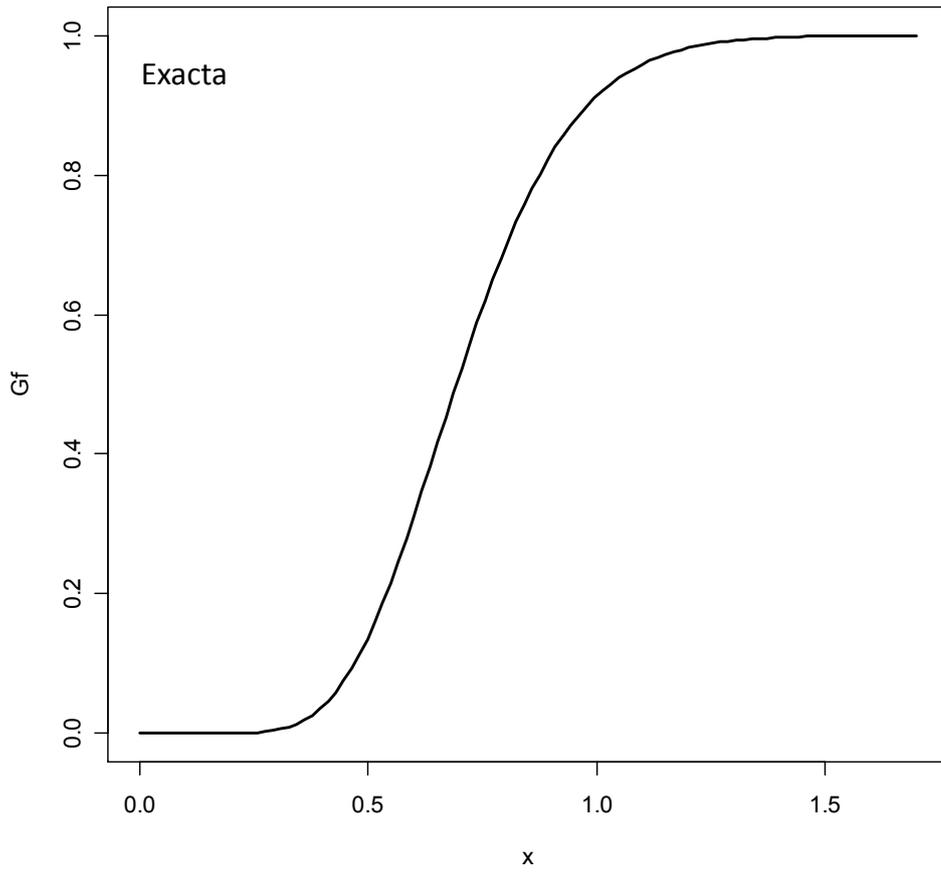
Distribución exacta de la mediana muestral de una población Exponencial (1), con n=25

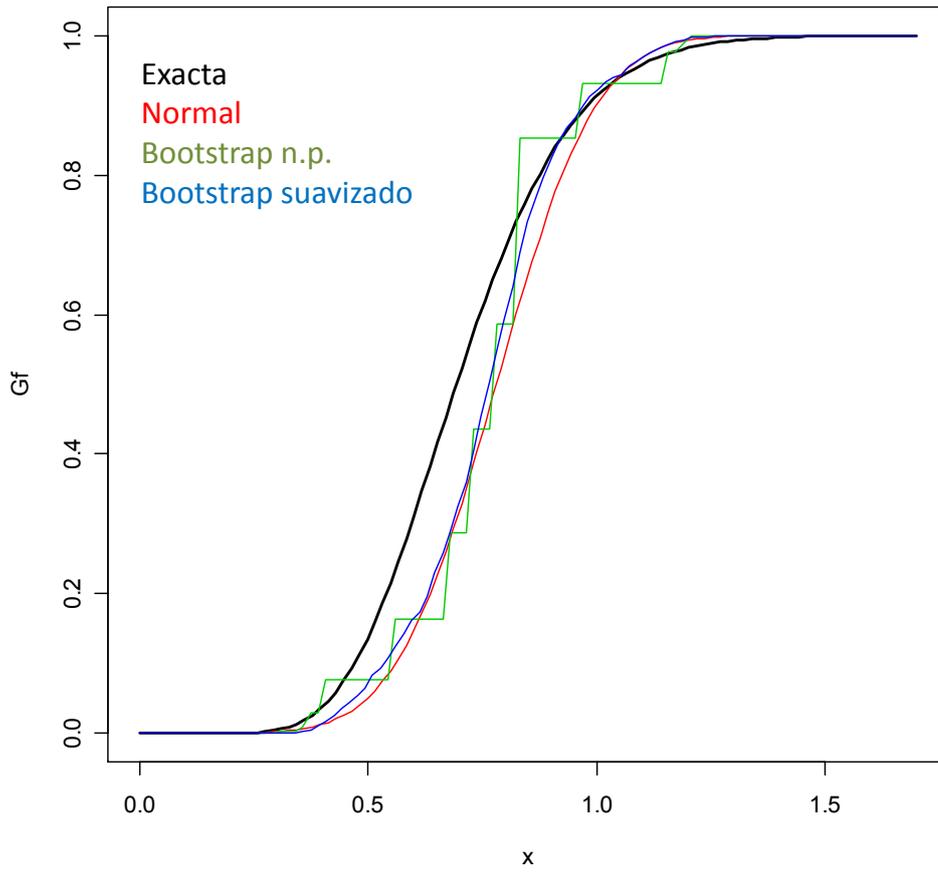
```
muestra<-rexp(n)
xmed=median(muestra)
lines(x.pnorm(x,mean=xmed,sd=1/sqrt(4*n*(density(x,n=1,from=xmed,to=xmed)$y^2)),col=2)
```

Aproximación normal con una muestra generada de esa población (n=25)

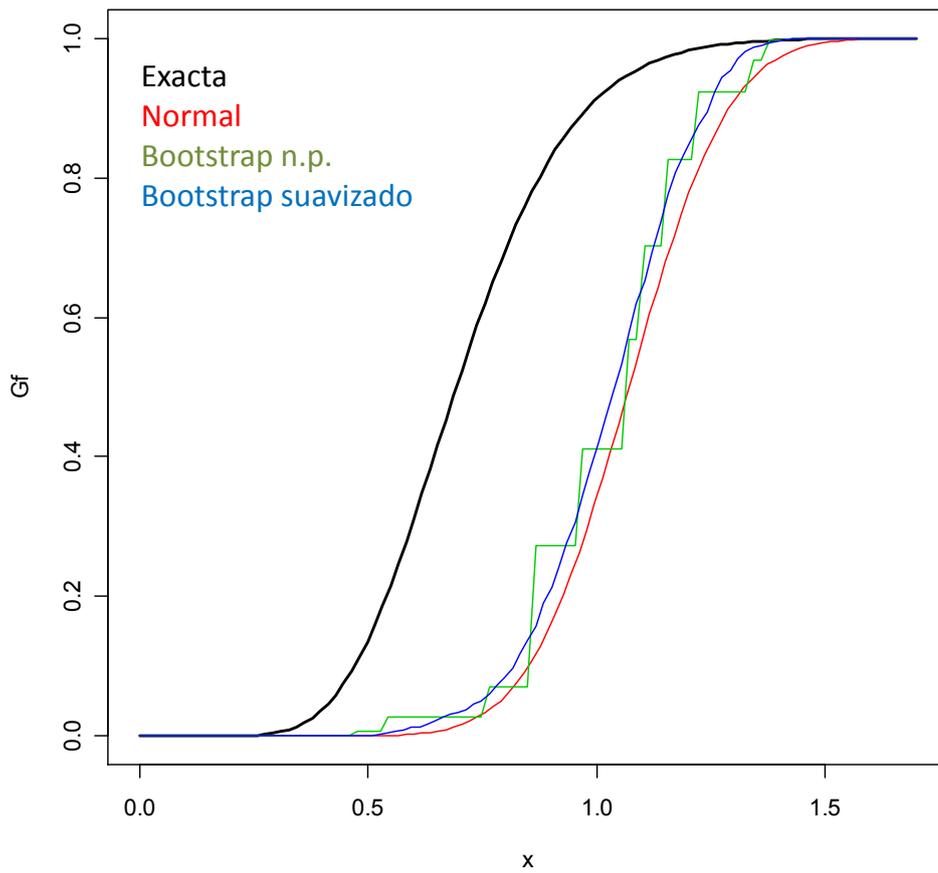
```
n=length(muestra)
remuestranp<-numeric(n)
remuestrasu<-numeric(B)
B=2000
mediananp<-numeric(B)
medianasu<-numeric(B)
for (k in 1:B)
{
  for (i in 1:n)
  {
    j=1
    u=runif(1)
    while (u>=j/n)
    {
      j=j+1
    }
    remuestranp[i]=muestra[j];
  }
  mediananp[k]=median(remuestranp)
  remuestrasu=remuestranp+0.1*rnorm(50,0,1)
  medianasu[k]=median(remuestrasu)
}
Bnp=ecdf(mediananp)
curve(Bnp,from=0,to=1.7,col=3,add=TRUE)
Bsu=ecdf(medianasu)
curve(Bsu,from=0,to=1.7,col=4,add=TRUE)
legend("topleft",legend=c("Exacta", "Normal", "Bootstrap n.p.", "Bootstrap suav."),lwd=2,col = c(1,2,3,4))
```

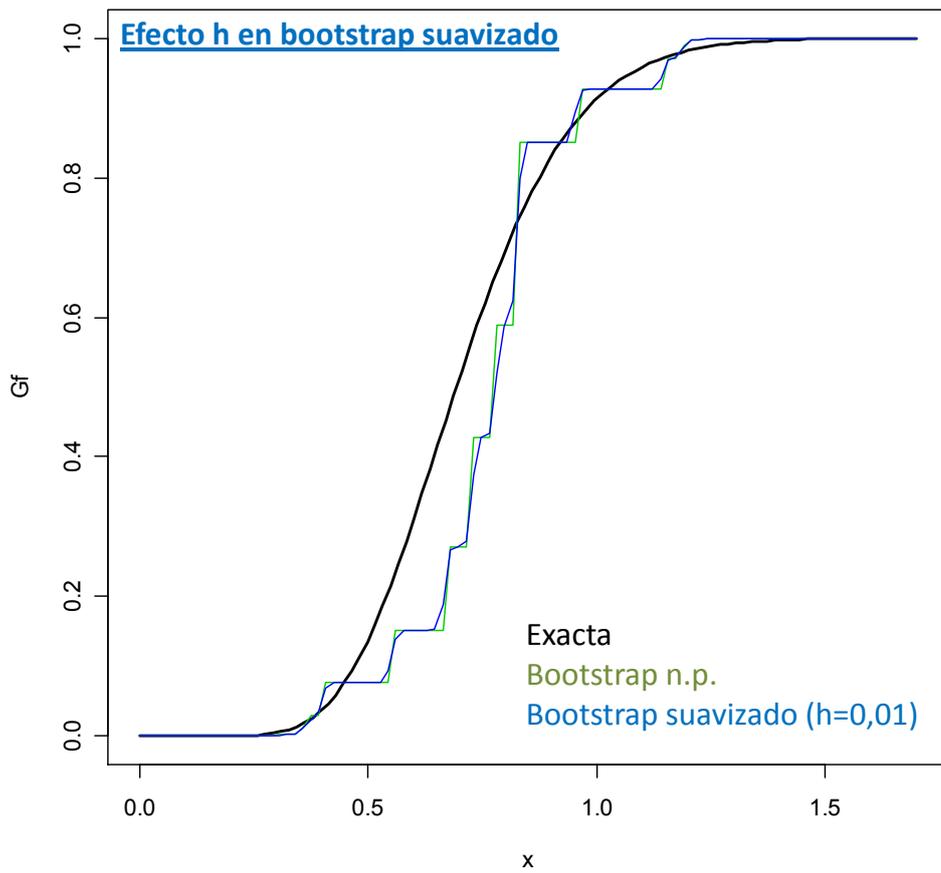
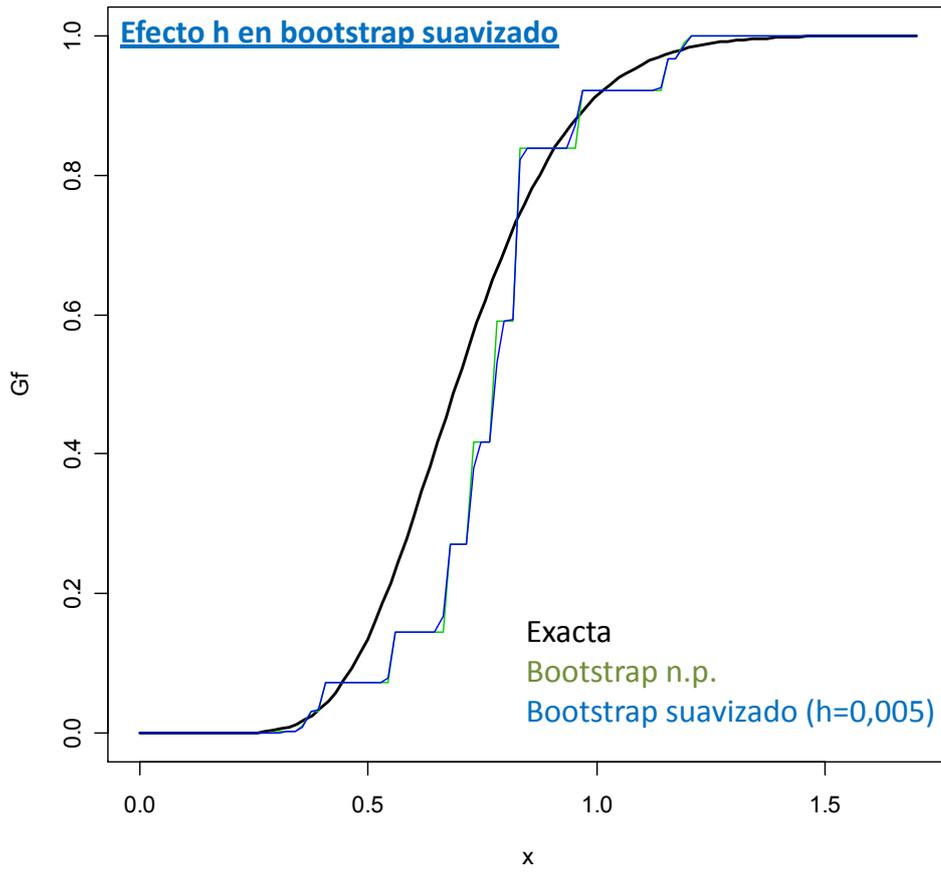
Aproximaciones bootstrap no paramétrico y suavizado con esa misma muestra

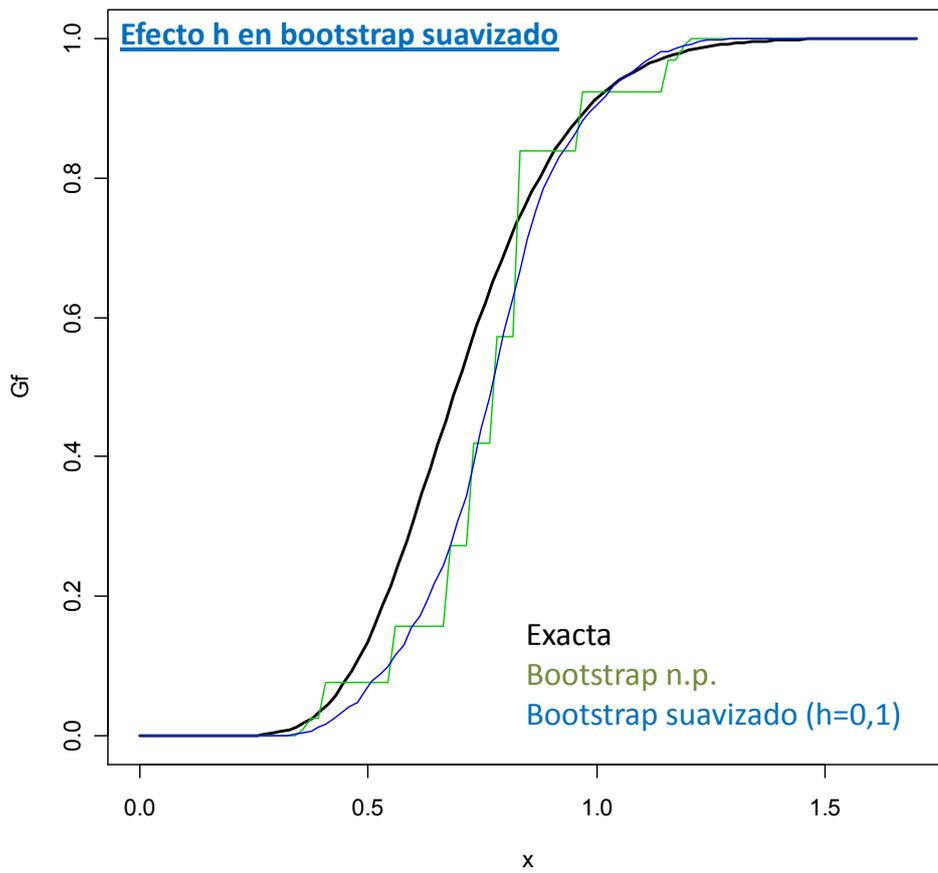
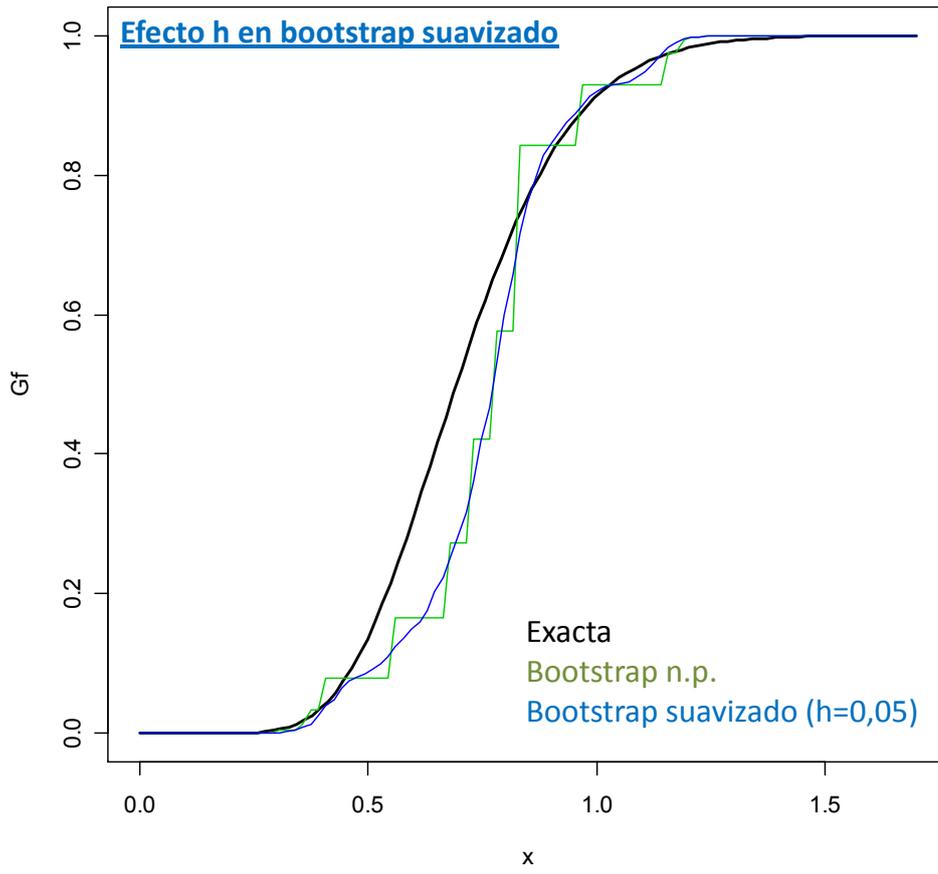




Con otra muestra generada de igual tamaño (claramente peor)







# INTERVALOS DE CONFIANZA BOOTSTRAP

1. Consideraciones y resultados teóricos generales.
2. Intervalos de confianza asintóticos normales unilaterales y bilaterales. Errores de recubrimiento.
3. Intervalos de confianza bootstrap unilaterales: métodos percentil, percentil-t y percentil-t simetrizado. Errores de recubrimiento.
4. Intervalos de confianza bootstrap bilaterales: métodos percentil, percentil-t y percentil-t simetrizado. Errores de recubrimiento.
5. Estudios de simulación.

## Aplicación:

- Construcción de intervalos de confianza bootstrap unilaterales y bilaterales. Comparación con la aproximación normal. **P9**

---

## CONSIDERACIONES Y RESULTADOS TEÓRICOS GENERALES

- Recordemos que para construir intervalos de nivel de confianza  $1 - \alpha$  hay que conocer la distribución de ciertas variables aleatorias (*variables pivotaes*).
- Restringiéndonos al contexto no paramétrico (planteamiento realista) la respuesta a este problema es siempre aproximada y además, fácil sólo cuando el objetivo son medias o funciones suaves de medias (en otro caso la respuesta puede ser muy complicada). El planteamiento clásico está basado en resultados del tipo

$$\sqrt{n} \frac{\bar{X} - \mu}{\sigma} \approx N(0,1)$$

Veremos cómo el bootstrap nos permitirá de modo fácil y siempre aproximar la distribución de estadísticos pivotaes, y además, en algunos casos, los intervalos de confianza bootstrap tendrán menor error de recubrimiento que los contruidos con la teoría clásica.

## DESARROLLOS DE EDGEWORTH

Sea  $X$  una variable aleatoria y  $\alpha(t) = E_F(e^{itX})$  su función característica.

Llamaremos  $j$ -ésimo cumulante de  $X$  (lo denotaremos  $k_j$ ) al coeficiente de  $(it)^j / j!$  en el siguiente desarrollo:

$$\log \alpha(t) = \sum_{j=1}^{\infty} k_j \frac{(it)^j}{j!}$$

Es decir,

$$\alpha(t) = \exp \left\{ \sum_{j=1}^{\infty} k_j \frac{(it)^j}{j!} \right\}$$

Si  $\varphi(t)$  denota la función característica de una variable aleatoria  $N(0,1)$

$$\varphi(t) = \exp \left\{ \sum_{j=1}^{\infty} \gamma_j \frac{(it)^j}{j!} \right\}$$

$\uparrow$   
j-ésimo cumulante de la  $N(0,1)$

el cociente entre ambas funciones características es

$$\frac{\alpha(t)}{\varphi(t)} = \exp \left\{ \sum_{j=1}^{\infty} (k_j - \gamma_j) \frac{(it)^j}{j!} \right\},$$

es decir,

$$\alpha(t) = \exp \left\{ \sum_{j=1}^{\infty} (k_j - \gamma_j) \frac{(it)^j}{j!} \right\} \varphi(t).$$

Por la fórmula de inversión:

$$F(x) - F(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-itx} - e^{-ity}}{it} \alpha(t) dt.$$

Haciendo tender  $y$  a  $-\infty$ , se tiene el desarrollo de Edgeworth de  $F$  en  $x$ :

$$F(x) = \Phi(x) + \sum_{j=1}^{\infty} (k_j - \gamma_j) \frac{(-1)^j}{j!} \Phi^{(j)}(x)$$

## RELACIÓN ENTRE MOMENTOS Y CUMULANTES

$$\bullet \alpha(t) = E_F(e^{itX}) \stackrel{\uparrow}{=} E_F\left(\sum_{j=0}^{\infty} X^j \frac{(it)^j}{j!}\right) = 1 + \underbrace{\sum_{j \geq 1} (E_F X^j)}_x \frac{(it)^j}{j!}$$

$e^z = \sum_{j=0}^{\infty} \frac{z^j}{j!}$

$$\bullet \log \alpha(t) = \sum_{j \geq 1} k_j \frac{(it)^j}{j!} \stackrel{\uparrow}{=} \sum_{k \geq 1} (-1)^{k+1} \frac{1}{k} \left( \sum_{j \geq 1} (E_F X^j) \frac{(it)^j}{j!} \right)^k$$

$\log(1+x) = \sum_{k \geq 1} (-1)^{k+1} \frac{1}{k} x^k$

*cumulante de orden j*       $\alpha_j = \text{momento de orden } j$

La relación entre momentos y cumulantes se tiene identificando los coeficientes de  $(it)^j/j!$  en esta última expresión:

$$k_1 = \alpha_1 = E_F X$$

$$k_2 = \alpha_2 - \alpha_1^2 = V_F X$$

$$k_3 = \alpha_3 - 3\alpha_2\alpha_1 + 2\alpha_1^3 = E_F (X - E_F X)^3$$

$$k_4 = \alpha_4 - 4\alpha_3\alpha_1 - 3\alpha_2^2 + 12\alpha_2\alpha_1^2 - 6\alpha_1^4 = E_F (X - E_F X)^4 - 3(V_F X)^2$$

...

$$\text{Ejemplo: } X \in N(\mu, \sigma^2) \rightarrow \log \varphi(t) = \log \exp \left\{ \underbrace{\mu}_{\gamma_1} (it) + \underbrace{\sigma^2}_{\gamma_2} \frac{(it)^2}{2} \right\}$$

$\gamma_i = 0$  si  $i > 2$

### OBSERVACIONES

- $k_j$  es un polinomio homogéneo en los momentos, de grado  $j$ .  
(análogamente  $\alpha_j$  es un polinomio homogéneo en los  $k_j$ , de grado  $j$ .)
- $\frac{k_3}{\sigma^3} \rightarrow$  coeficiente de sesgo de  $X$ .
- $\frac{k_4}{\sigma^4} \rightarrow$  coeficiente de curtosis de  $X$ .

## TEOREMA DE CRAMER

Sean  $X_1, \dots, X_n$  variables aleatorias independientes e idénticamente distribuidas (con media  $\mu$  y varianza  $\sigma^2$ ) tales que

$$E(|X|^{j+2}) < \infty \quad y \quad \limsup_{|t| \rightarrow \infty} |\alpha(t)| < 1.$$

Entonces:

$$P\left(\sqrt{n} \frac{\bar{X} - \mu}{\sigma} \leq y\right) = \Phi(y) + n^{-\frac{1}{2}} p_1(y) \phi(y) + \dots + n^{-\frac{j-1}{2}} p_{j-1}(y) \phi(y) + O\left(n^{-\frac{j}{2}}\right)$$

uniformemente en  $y \in \mathbb{R}$ . Los polinomios  $p_i$  dependen de los momentos de  $X$  de orden menor o igual que  $i+2$ , son de grado  $3i-1$  y tienen paridad opuesta a su índice ( $p_1(y) = p_1(-y)$ ;  $p_2(y) = -p_2(-y)$  ...). En este desarrollo:

$$p_1(y) = -\frac{1}{6} \left(\frac{k_3}{\sigma^3}\right) (y^2 - 1) \quad \xrightarrow{\gamma_1}$$
$$p_2(y) = -y \left[ \frac{1}{24} \left(\frac{k_4}{\sigma^4}\right) (y^2 - 3) + \frac{1}{72} \left(\frac{k_3}{\sigma^3}\right)^2 (y^4 - 10y^2 + 15) \right] \quad \xrightarrow{\gamma_2}$$

## OBSERVACIONES

- La condición de continuidad de Cramer

$$\limsup_{|t| \rightarrow \infty} |\alpha(t)| < 1,$$

falla si  $X$  es de tipo "lattice", es decir

$$x \in \text{soporte de } X \Leftrightarrow x = jh + a, \quad h > 0, \quad -\infty < a < \infty \quad y \quad j \text{ entero.}$$

- Si  $X$  es absolutamente continua, entonces  $\limsup_{|t| \rightarrow \infty} |\alpha(t)| = 0$ .

- Notemos que el teorema de Cramer aporta información relevante sobre la velocidad de convergencia implícita en el Teorema de Lévy-Lindeberg:

$$\left( \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \xrightarrow{D} N(0,1) \right)$$

Consecuencia del teorema anterior es lo siguiente:

$$\begin{aligned}
 P\left(\sqrt{n}\frac{\bar{X}-\mu}{\sigma} \leq y\right) &= \Phi(y) + \\
 &+ n^{-\frac{1}{2}}\left(-\frac{1}{6}\frac{k_3}{\sigma^3}(y^2-1)\right)\phi(y) + \gamma_1 \\
 &+ n^{-1}\left(-y\left[\frac{1}{24}\frac{k_4}{\sigma^4}(y^2-3) + \frac{1}{72}\left(\frac{k_3}{\sigma^3}\right)^2(y^4-10y^2+15)\right]\right)\phi(y) + \gamma_2 \\
 &+ O\left(n^{-\frac{3}{2}}\right).
 \end{aligned}$$

Aporta a la normal una corrección de primer orden por SESGO (si  $k_3 = 0$  ese término de error no existe).

Aporta a la normal una corrección de segundo orden por CURTOSIS (si  $k_3 = k_4 = 0$  ese término de error no existe).

Tpc  
peso errores  
con n=100

$P\left(\sqrt{n}\frac{\bar{X}^* - \bar{X}}{\hat{\sigma}} \leq y\right)$  Tiene un desarrollo de Edgeworth análogo, reemplazando los momentos poblacionales  $\alpha_j$  por los muestrales  $\hat{\alpha}_j$  (implícitos en  $k_j$ ).

Como siempre se verifica que  $\hat{\alpha}_j = \alpha_j + O_p\left(n^{-\frac{1}{2}}\right)$ , se tiene:

$$P\left(\sqrt{n}\frac{\bar{X}-\mu}{\sigma} \leq y\right) - P\left(\sqrt{n}\frac{\bar{X}^* - \bar{X}}{\hat{\sigma}} \leq y\right) = O_p\left(n^{-1}\right) \text{ uniformemente en } y.$$

Veámoslo:

$$P\left(\sqrt{n}\frac{\bar{X}-\mu}{\sigma} \leq y\right) = \Phi(y) + n^{-\frac{1}{2}}(\dots k_3 \dots) + n^{-1}(\dots k_3, k_4 \dots) + \dots$$

$$P\left(\sqrt{n}\frac{\bar{X}^* - \bar{X}}{\hat{\sigma}} \leq y\right) = \Phi(y) + n^{-\frac{1}{2}}(\dots \hat{k}_3 \dots) + n^{-1}(\dots \hat{k}_3, \hat{k}_4 \dots) + \dots$$

Restando ambas expresiones desaparece  $\Phi(y)$  y como  $\hat{k}_3 = k_3 + O_p\left(n^{-\frac{1}{2}}\right)$  y  $n^{-\frac{1}{2}} \times n^{-\frac{1}{2}} = n^{-1}$ , se obtiene el resultado indicado.

### CONSIDERACIONES IMPORTANTES

- El bootstrap *captura adecuadamente el efecto de la asimetría*: recordemos que el primer término de error (orden  $n^{-1/2}$ ) en el desarrollo de Edgeworth anterior desaparece si el coeficiente de sesgo o asimetría es 0.

$$P\left(\sqrt{n}\frac{\bar{X}-\mu}{\sigma}\leq y\right)-P\left(\sqrt{n}\frac{\bar{X}^*-\bar{X}}{\hat{\sigma}}\leq y\right)=O_P(n^{-1}),$$

mientras que, evidentemente, para la aproximación normal se tiene que

$$iP\left(\sqrt{n}\frac{\bar{X}-\mu}{\sigma}\leq y\right)-\Phi(y)=O(n^{-1/2})!$$

- Notemos también que si  $X$  es simétrica, el *bootstrap uniforme* no aporta mejora sobre la aproximación normal (ya de orden  $O(n^{-1})$ ). Es importante indicar que si utilizamos el *bootstrap simetrizado* ( $\hat{k}_3 = 0$ ), entonces, razonando como antes, el bootstrap recupera su ventaja:

$$O_P(n^{-3/2}) \text{ frente a } O(n^{-1})$$

### TEOREMA (BHATTACHARYA-GHOST)

Sea  $\hat{\theta} = \theta(\hat{F})$  el estimador bootstrap de  $\theta = \theta(F)$  y supongamos que

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow[n \rightarrow \infty]{} N(0, \sigma^2).$$

Bajo condiciones de regularidad generales (Bhattacharya and Ghost, 1978) se verifica que

$$P\left(\sqrt{n}\frac{\hat{\theta}-\theta}{\sigma}\leq x\right)=\Phi(x)+n^{-\frac{1}{2}}p_1(x)\phi(x)+n^{-1}p_2(x)\phi(x)+\dots$$

$$P\left(\sqrt{n}\frac{\hat{\theta}-\theta}{\hat{\sigma}}\leq x\right)=\Phi(x)+n^{-\frac{1}{2}}q_1(x)\phi(x)+n^{-1}q_2(x)\phi(x)+\dots$$

donde  $p_j$  y  $q_j$  son polinomios de grado  $3j-1$  y paridad opuesta a sus índices.

En particular se tiene que:

$$\begin{aligned}
 P\left(\sqrt{n}\frac{\bar{X}-\mu}{\hat{\sigma}} \leq y\right) &= \Phi(y) + \\
 &+ n^{-\frac{1}{2}}\left(\frac{1}{6}\frac{k_3}{\sigma^3}(2y^2+1)\right)\phi(y) + \gamma_1 \\
 &+ n^{-1}\left(y\left[\frac{1}{12}\frac{k_4}{\sigma^4}(y^2-3) - \frac{1}{18}\left(\frac{k_3}{\sigma^3}\right)^2(y^4+2y^2-3) - \frac{1}{4}(y^2+3)\right]\right)\phi(y) + \gamma_2 \\
 &+ O\left(n^{-\frac{3}{2}}\right). \gamma_2
 \end{aligned}$$

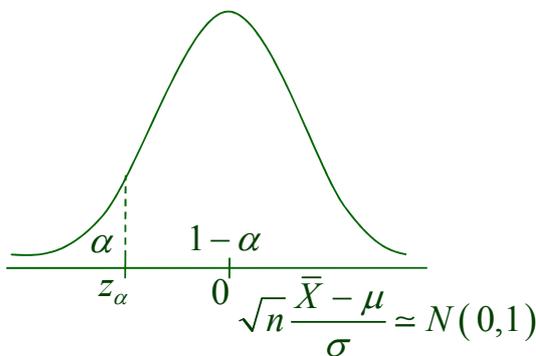
**NOTA**

- En contexto más general caben, por lo tanto, las mismas consideraciones realizadas para el caso de la media muestral no studentizada.
- En particular también se tiene el teorema anterior.

**INTERVALOS DE CONFIANZA ASINTÓTICOS NORMALES PARA  $\mu$**   
**TEORÍA CLÁSICA**

Sea  $\vec{X} = (X_1, \dots, X_n)^t$  una muestra aleatoria simple de tamaño n de  $F$  desconocida

**INTERVALOS UNILATERALES**



$$P\left(\sqrt{n}\frac{\bar{X}-\mu}{\sigma} \leq z_\alpha\right) = \alpha$$

$$P\left(\bar{X} - \mu \leq \frac{\sigma}{\sqrt{n}}z_\alpha\right) = \alpha$$

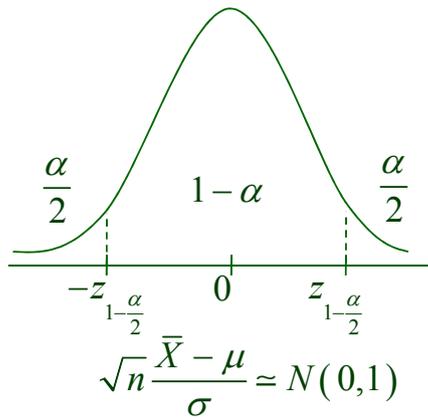
$$P\left(\mu \geq \bar{X} - \frac{\sigma}{\sqrt{n}}z_\alpha\right) = \alpha$$

$$P\left(\mu \leq \bar{X} - \frac{\sigma}{\sqrt{n}}z_\alpha\right) = 1 - \alpha$$

$$P\left(\mu \in \left(-\infty, \bar{X} - \frac{\sigma}{\sqrt{n}}z_\alpha\right)\right) = 1 - \alpha$$

(Nota:  $z_\alpha < 0$ )

## INTERVALOS BILATERALES



$$P\left(-z_{1-\frac{\alpha}{2}} \leq \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \leq z_{1-\frac{\alpha}{2}}\right) \approx 1 - \alpha$$

$$P\left(-\frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \leq \bar{X} - \mu \leq \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right) \approx 1 - \alpha$$

$$P\left(\frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \geq \mu - \bar{X} \geq -\frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right) \approx 1 - \alpha$$

$$P\left(\bar{X} - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \leq \mu \leq \bar{X} + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right) \approx 1 - \alpha$$

$$P\left(\mu \in \left(\bar{X} \mp \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right)\right) \approx 1 - \alpha$$

## INTERVALOS DE CONFIANZA ASINTÓTICOS NORMALES PARA $\mu$ ERRORES DE RECUBRIMIENTO

Veamos la utilidad de los teoremas anteriores para calcular el error de recubrimiento (cobertura) de los intervalos de confianza normales para  $\mu$  de una distribución bajo sus hipótesis.

$$J_0 = \left(-\infty, \bar{X} - \frac{\sigma}{\sqrt{n}} z_\alpha\right) \quad \text{I.C.}(\mu) \text{ de nivel } \approx (1 - \alpha) \text{ unilateral}$$

$$I_0 = \left(\bar{X} \mp \frac{\sigma}{\sqrt{n}} z_\alpha\right) \quad \text{I.C.}(\mu) \text{ de nivel } \approx (1 - \alpha) \text{ bilateral}$$

Por el teorema de Cramer:

$$\begin{aligned} \bullet \quad P(\mu \in J_0) &= P\left(\mu \leq \bar{X} - \frac{\sigma}{\sqrt{n}} z_\alpha\right) = 1 - P\left(\mu \geq \bar{X} - \frac{\sigma}{\sqrt{n}} z_\alpha\right) = \\ &= 1 - P\left(\frac{\sigma}{\sqrt{n}} z_\alpha \geq \bar{X} - \mu\right) = 1 - P\left(\sqrt{n} \frac{\bar{X} - \mu}{\sigma} \leq z_\alpha\right) = 1 - \alpha + O\left(n^{-\frac{1}{2}}\right) \end{aligned}$$

pues

$$P\left(\sqrt{n} \frac{\bar{X} - \mu}{\sigma} \leq z_\alpha\right) = \Phi(z_\alpha) + n^{-\frac{1}{2}} p_1(z_\alpha) \phi(z_\alpha) + O(n^{-1})$$

Análogamente, razonando con el intervalo bilateral:

$$\begin{aligned}
 \bullet \quad \boxed{P(\mu \in I_0)} &= P\left(\bar{X} - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \leq \mu \leq \bar{X} + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right) = \\
 &= P\left(-z_{1-\frac{\alpha}{2}} \leq \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \leq z_{1-\frac{\alpha}{2}}\right) = \\
 &= P\left(\sqrt{n} \frac{\bar{X} - \mu}{\sigma} \leq z_{1-\frac{\alpha}{2}}\right) - P\left(\sqrt{n} \frac{\bar{X} - \mu}{\sigma} \leq -z_{1-\frac{\alpha}{2}}\right) = \\
 &= 1 - \frac{\alpha}{2} + n^{\frac{1}{2}} p_1\left(z_{1-\frac{\alpha}{2}}\right) \phi\left(z_{1-\frac{\alpha}{2}}\right) + O(n^{-1}) - \left(\frac{\alpha}{2} + n^{\frac{1}{2}} p_1\left(-z_{1-\frac{\alpha}{2}}\right) \phi\left(-z_{1-\frac{\alpha}{2}}\right) + O(n^{-1})\right) = \\
 &= \boxed{1 - \alpha + O(n^{-1})} \quad (\text{pues } p_1(y) = p_1(-y))
 \end{aligned}$$

### **OBSERVACIÓN**

- La aproximación normal es mejor en el caso bilateral.
- Análogo resultado se obtiene si  $\sigma^2$  es desconocida (estimándola con  $\hat{\sigma}^2$ ) utilizando el teorema de Bhattacharya-Ghost.

## **INTERVALOS DE CONFIANZA BOOTSTRAP UNILATERALES: métodos percentil y percentil-t. ERRORES DE RECUBRIMIENTO**

Sea  $\vec{X} = (X_1, \dots, X_n)^t$  una muestra aleatoria simple de tamaño  $n$  de  $F$  desconocida

**OBJETIVO:** Construir un intervalo de confianza unilateral de nivel  $(1 - \alpha)$  para  $\theta = \theta(F)$  a partir de un estimador  $\hat{\theta} = \theta(\vec{X})$  con varianza asintótica  $\sigma^2/n$  (en las hipótesis de los teoremas previos anteriores).

### **MÉTODO PERCENTIL**

Si se conociera la distribución de  $\sqrt{n}(\hat{\theta} - \theta)$  bajo  $F$ , entonces, si  $x_\alpha$  es el cuantil de orden  $\alpha$  de esa distribución,

$$\alpha = P_F\left(\sqrt{n}(\hat{\theta} - \theta) \leq x_\alpha\right) = P_F\left(\theta \geq \hat{\theta} - n^{\frac{1}{2}} x_\alpha\right) \Rightarrow P_F\left(\theta \leq \hat{\theta} - n^{\frac{1}{2}} x_\alpha\right) = 1 - \alpha$$

y por lo tanto:

$$J = \left(-\infty, \hat{\theta} - n^{\frac{1}{2}} x_\alpha\right) \quad \text{I.C. para } \theta \text{ de nivel } (1 - \alpha)$$

El método percentil aproxima la distribución de  $R(\bar{X}, F) = \sqrt{n}(\hat{\theta} - \theta)$  bajo  $F$  por la de  $R(\bar{X}^*, \hat{F}) = \sqrt{n}(\hat{\theta}^* - \hat{\theta})$  bajo  $\hat{F}$ , es decir

$$P_F(\sqrt{n}(\hat{\theta} - \theta) \leq y) \approx P_{\hat{F}}(\sqrt{n}(\hat{\theta}^* - \hat{\theta}) \leq y),$$

y por lo tanto, si  $x_\alpha^*$  es el cuantil de orden  $\alpha$  de  $\sqrt{n}(\hat{\theta}^* - \hat{\theta})$ ,

$$J_1 = \left( -\infty, \hat{\theta} - n^{-\frac{1}{2}} x_\alpha^* \right) \text{ I.C. BOOT. PERCENTIL para } \theta \text{ de nivel } (1 - \alpha)$$

En la práctica,  $x_\alpha^*$  se aproxima por Monte Carlo a partir de  $\bar{x}$ . Así, si  $\theta = \mu$  y  $\hat{\theta} = \bar{X}$

$$X^* \equiv \begin{pmatrix} 1/n & \dots & 1/n \\ x_1 & \dots & x_n \end{pmatrix}$$

↓

$$\bar{x}^{*b} \rightarrow r^{*b} = R(\bar{x}^{*b}; \hat{F}) = \sqrt{n}(\bar{x}^{*b} - \bar{x})$$

$b = 1, \dots, B$

$x_\alpha^* \approx x_\alpha^{*MC} = r^{*([B\alpha])}$

### ERROR DE COBERTURA MÉTODO PERCENTIL

Consideremos  $\theta = \mu$  y  $\hat{\theta} = \bar{X}$ .

Por el teorema de Cramer sabemos que

$$P_F(\sqrt{n}(\bar{X} - \mu) \leq y) = \Phi\left(\frac{y}{\sigma}\right) + O\left(n^{-\frac{1}{2}}\right).$$

Además, también sabemos que

$$\hat{\alpha}_j = \alpha_j + O_p\left(n^{-\frac{1}{2}}\right).$$

Análogamente:

$$P_{\hat{F}}(\sqrt{n}(\bar{X}^* - \bar{X}) \leq y) = \Phi\left(\frac{y}{\hat{\sigma}}\right) + O_p\left(n^{-\frac{1}{2}}\right).$$

Por lo tanto,

$$P_F(\sqrt{n}(\bar{X} - \mu) \leq y) - P_{\hat{F}}(\sqrt{n}(\bar{X}^* - \bar{X}) \leq y) = O_p\left(n^{-\frac{1}{2}}\right).$$

En consecuencia,

$$\begin{aligned}
 \boxed{P_F(\mu \in J_1)} &= P_F\left(\mu \leq \bar{X} - n^{\frac{1}{2}}x_\alpha^*\right) = P_F\left(\sqrt{n}(\mu - \bar{X}) \leq -x_\alpha^*\right) = \\
 &= P_F\left(\sqrt{n}(\bar{X} - \mu) \geq x_\alpha^*\right) = 1 - P_F\left(\sqrt{n}(\bar{X} - \mu) \leq x_\alpha^*\right) = \\
 &= 1 - P_{\hat{F}}\left(\sqrt{n}(\bar{X}^* - \bar{X}) \leq x_\alpha^*\right) + O_P\left(n^{-\frac{1}{2}}\right) = \\
 &= \boxed{1 - \alpha + O_P\left(n^{-\frac{1}{2}}\right)}
 \end{aligned}$$

Notemos que este error de recubrimiento coincide con el de la teoría clásica:

$$\begin{aligned}
 P_F(\mu \in J_0) &= 1 - \alpha + O\left(n^{-\frac{1}{2}}\right) \\
 P_F(\mu \in J_1) &= 1 - \alpha + O_P\left(n^{-\frac{1}{2}}\right)
 \end{aligned}$$

### MÉTODO PERCENTIL-t

Si se conociera la distribución de la variable aleatoria

$$\sqrt{n} \frac{\hat{\theta} - \theta}{\hat{\sigma}}$$

bajo  $F$ , entonces, si  $x_\alpha$  es el cuantil de orden  $\alpha$  de esa distribución,

$$\alpha = P_F\left(\sqrt{n} \frac{\hat{\theta} - \theta}{\hat{\sigma}} \leq x_\alpha\right) = P_F\left(\theta \geq \hat{\theta} - \hat{\sigma} n^{\frac{1}{2}} x_\alpha\right) \Rightarrow P_F\left(\theta \leq \hat{\theta} - \hat{\sigma} n^{\frac{1}{2}} x_\alpha\right) = 1 - \alpha$$

y por lo tanto:

$$J = \left(-\infty, \hat{\theta} - \hat{\sigma} n^{\frac{1}{2}} x_\alpha\right) \text{ I.C. para } \theta \text{ de nivel } (1 - \alpha).$$

El método percentil-t aproxima la distribución de  $R(\bar{X}; F) = \sqrt{n} \frac{\hat{\theta} - \theta}{\hat{\sigma}}$  bajo  $F$

por la de  $R(\bar{X}^*; \hat{F}) = \sqrt{n} \frac{\hat{\theta}^* - \hat{\theta}}{\hat{\sigma}^*}$  bajo  $\hat{F}$ , es decir

$$P_F \left( \sqrt{n} \frac{\hat{\theta} - \theta}{\hat{\sigma}} \leq y \right) = P_{\hat{F}} \left( \sqrt{n} \frac{\hat{\theta}^* - \hat{\theta}}{\hat{\sigma}^*} \leq y \right),$$

y por lo tanto, si  $x_\alpha^*$  es el cuantil de orden  $\alpha$  de  $\sqrt{n} \frac{\hat{\theta}^* - \hat{\theta}}{\hat{\sigma}^*}$ ,

$$J_2 = \left( -\infty, \hat{\theta} - \hat{\sigma} n^{-\frac{1}{2}} x_\alpha^* \right) \quad \text{I.C. BOOT. PERCENTIL-t para } \theta \text{ de nivel } (1 - \alpha)$$

En la práctica,  $x_\alpha^*$  se aproxima por Monte Carlo a partir de  $\bar{x}$ . Así, si  $\theta = \mu$  y  $\hat{\theta} = \bar{X}$

$$X^* \equiv \begin{pmatrix} 1/n & \dots & 1/n \\ x_1 & \dots & x_n \end{pmatrix}$$

↓

$$\bar{x}^{*b} \rightarrow r^{*b} = R(\bar{x}^{*b}; \hat{F}) = \sqrt{n} \frac{\bar{x}^{*b} - \bar{x}}{\hat{\sigma}^{*b}} \quad b = 1, \dots, B$$

$x_\alpha^* \approx \hat{x}_\alpha^{*MC} = r^*([B\alpha])$

### ERROR DE COBERTURA MÉTODO PERCENTIL-t

Consideremos  $\theta = \mu$  y  $\hat{\theta} = \bar{X}$ .

Por el teorema de Bhattacharya-Ghost, sabemos que

$$P_F \left( \sqrt{n} \frac{\bar{X} - \mu}{\hat{\sigma}} \leq y \right) = \Phi(y) + n^{-\frac{1}{2}} \left( \frac{1}{6} \underbrace{\left( \frac{k_3}{\sigma^3} \right)}_{\gamma_1} (2y^2 + 1) \right) \phi(y) + O(n^{-1}).$$

Reemplazando momentos poblacionales por muestrales se tiene una expresión análoga para  $\sqrt{n} \frac{\bar{X}^* - \bar{X}}{\hat{\sigma}^*}$ .

Además, también sabemos que

$$\hat{\alpha}_j = \alpha_j + O_p \left( n^{-\frac{1}{2}} \right).$$

Restando ambas expresiones se deduce por lo tanto:

$$P_F \left( \sqrt{n} \frac{\bar{X} - \mu}{\hat{\sigma}} \leq y \right) - P_{\hat{F}} \left( \sqrt{n} \frac{\bar{X}^* - \bar{X}}{\hat{\sigma}^*} \leq y \right) = O_p(n^{-1}).$$

En consecuencia,

$$\begin{aligned}
 \boxed{P_F(\mu \in J_2)} &= P_F\left(\mu \leq \bar{X} - \hat{\sigma} n^{-\frac{1}{2}} x_\alpha^*\right) = P_F\left(\sqrt{n}(\mu - \bar{X}) \leq -\hat{\sigma} x_\alpha^*\right) = \\
 &= P_F\left(\sqrt{n} \frac{\bar{X} - \mu}{\hat{\sigma}} \geq x_\alpha^*\right) = 1 - P_F\left(\sqrt{n} \frac{\bar{X} - \mu}{\hat{\sigma}} \leq x_\alpha^*\right) = \\
 &= 1 - P_{\hat{F}}\left(\sqrt{n} \frac{\bar{X}^* - \bar{X}}{\hat{\sigma}^*} \leq x_\alpha^*\right) + O_P(n^{-1}) = \\
 &= \boxed{1 - \alpha + O_P(n^{-1})}
 \end{aligned}$$

Notemos que este error de recubrimiento mejora al de la teoría clásica:

$$\begin{aligned}
 P_F(\mu \in J_0) &= 1 - \alpha + O\left(n^{-\frac{1}{2}}\right) \\
 P_F(\mu \in J_2) &= 1 - \alpha + O_P(n^{-1})
 \end{aligned}$$

Variables pivotaes "aproximadas" para una media poblacional	Intervalos de confianza unilaterales para la media poblacional de nivel $1 - \alpha$	Error de recubrimiento $P_F(\mu \in J_i) - (1 - \alpha)$
$\sqrt{n} \frac{\bar{X} - \mu}{\hat{\sigma}} \approx N(0, 1)$	$J_0 = \left(-\infty, \bar{X} - \hat{\sigma} n^{-\frac{1}{2}} z_\alpha\right)$	$O\left(n^{-\frac{1}{2}}\right)$
$\sqrt{n}(\bar{X} - \mu) \approx \sqrt{n}(\bar{X}^* - \bar{X})$	$J_1 = \left(-\infty, \bar{X} - n^{-\frac{1}{2}} x_\alpha^*\right)$	$O_P\left(n^{-\frac{1}{2}}\right)$
$\sqrt{n} \frac{\bar{X} - \mu}{\hat{\sigma}} \approx \sqrt{n} \frac{\bar{X}^* - \bar{X}}{\hat{\sigma}^*}$	$J_2 = \left(-\infty, \bar{X} - \hat{\sigma} n^{-\frac{1}{2}} x_\alpha^*\right)$	$O_P(n^{-1})$

## CONSIDERACIONES IMPORTANTES

- Notemos que el método **percentil** invierte su esfuerzo en calcular una corrección para la escala de  $\hat{\theta}$ , por lo que no proporciona una corrección efectiva para el sesgo.
- El método **percentil-t**, al estar el estadístico ya corregido por su escala, invierte su esfuerzo en corregir el sesgo y de ahí su mejor comportamiento (recordemos que el principal término del error cometido, respecto de la normal, al aproximar la distribución de un estadístico es debido al sesgo, ver teoremas de Cramer y de Bhattacharya-Ghost).

## **INTERVALOS DE CONFIANZA BOOTSTRAP BILATERALES: métodos percentil, percentil-t y percentil-t simetrizado. ERRORES DE RECUBRIMIENTO**

---

Sea  $\vec{X} = (X_1, \dots, X_n)'$  una muestra aleatoria simple de tamaño  $n$  de  $F$  desconocida

**OBJETIVO:** Construir un intervalo de confianza bilateral de nivel  $(1 - \alpha)$  para  $\theta = \theta(F)$  a partir de un estimador  $\hat{\theta} = \theta(\vec{X})$  con varianza asintótica  $\sigma^2/n$  (en las hipótesis de los teoremas previos anteriores).

### MÉTODO PERCENTIL

Si se conociera la distribución de  $\sqrt{n}(\hat{\theta} - \theta)$  bajo  $F$ , entonces, si  $x_\alpha$  es el cuantil de orden  $\alpha$  de esa distribución,

$$P_F \left( x_{\alpha/2} \leq \sqrt{n}(\hat{\theta} - \theta) \leq x_{1-\alpha/2} \right) = 1 - \alpha$$

y por lo tanto:

$$I = \left( \hat{\theta} - n^{-\frac{1}{2}} x_{1-\alpha/2}, \hat{\theta} - n^{-\frac{1}{2}} x_{\alpha/2} \right) \quad \text{I.C. para } \theta \text{ de nivel } (1 - \alpha)$$

El método percentil aproxima la distribución de  $R(\bar{X}, F) = \sqrt{n}(\hat{\theta} - \theta)$  bajo  $F$  por la de  $R(\bar{X}^*, \hat{F}) = \sqrt{n}(\hat{\theta}^* - \hat{\theta})$  bajo  $\hat{F}$ , es decir

$$P_F(\sqrt{n}(\hat{\theta} - \theta) \leq y) \approx P_{\hat{F}}(\sqrt{n}(\hat{\theta}^* - \hat{\theta}) \leq y),$$

y por lo tanto, si  $x_\alpha^*$  es el cuantil de orden  $\alpha$  de  $\sqrt{n}(\hat{\theta}^* - \hat{\theta})$ ,

$$I_1 = \left( \hat{\theta} - n^{-\frac{1}{2}} x_{1-\alpha/2}^*, \hat{\theta} - n^{-\frac{1}{2}} x_{\alpha/2}^* \right) \text{ I.C. BOOT. PERCENTIL para } \theta \text{ de nivel } (1-\alpha)$$

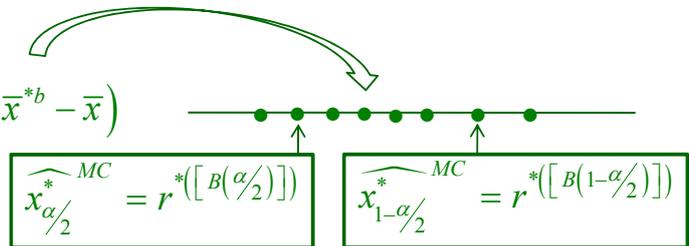
En la práctica,  $x_\alpha^*$  se aproxima por Monte Carlo a partir de  $\bar{x}$ . Así, si  $\theta = \mu$  y  $\hat{\theta} = \bar{X}$

$$X^* \equiv \begin{pmatrix} 1/n & \dots & 1/n \\ x_1 & \dots & x_n \end{pmatrix}$$

↓

$$\bar{x}^{*b} \rightarrow r^{*b} = R(\bar{x}^{*b}; \hat{F}) = \sqrt{n}(\bar{x}^{*b} - \bar{x})$$

$b = 1, \dots, B$



### ERROR DE COBERTURA MÉTODO PERCENTIL

Consideremos  $\theta = \mu$  y  $\hat{\theta} = \bar{X}$ .

Por el teorema de Cramer sabemos que

$$P_F(\sqrt{n}(\bar{X} - \mu) \leq y) = \Phi\left(\frac{y}{\sigma}\right) + O\left(n^{-\frac{1}{2}}\right).$$

Además, también sabemos que

$$\hat{\alpha}_j = \alpha_j + O_p\left(n^{-\frac{1}{2}}\right).$$

Análogamente:

$$P_{\hat{F}}(\sqrt{n}(\bar{X}^* - \bar{X}) \leq y) = \Phi\left(\frac{y}{\hat{\sigma}}\right) + O_p\left(n^{-\frac{1}{2}}\right).$$

Por lo tanto,

$$P_F(\sqrt{n}(\bar{X} - \mu) \leq y) - P_{\hat{F}}(\sqrt{n}(\bar{X}^* - \bar{X}) \leq y) = O_p\left(n^{-\frac{1}{2}}\right).$$

En consecuencia,

$$\begin{aligned}
 \boxed{P_F(\mu \in I_1)} &= P_F\left(\bar{X} - n^{-\frac{1}{2}}x_{1-\alpha/2}^* \leq \mu \leq \bar{X} - n^{-\frac{1}{2}}x_{\alpha/2}^*\right) = P_F\left(x_{\alpha/2}^* \leq \sqrt{n}(\bar{X} - \mu) \leq x_{1-\alpha/2}^*\right) = \\
 &= P_F\left(\sqrt{n}(\bar{X} - \mu) \leq x_{1-\alpha/2}^*\right) - P_F\left(\sqrt{n}(\bar{X} - \mu) \leq x_{\alpha/2}^*\right) = \\
 &= \left[ P_{\hat{F}}\left(\sqrt{n}(\bar{X}^* - \bar{X}) \leq x_{1-\alpha/2}^*\right) + O_P\left(n^{-\frac{1}{2}}\right) \right] - \left[ P_{\hat{F}}\left(\sqrt{n}(\bar{X}^* - \bar{X}) \leq x_{\alpha/2}^*\right) + O_P\left(n^{-\frac{1}{2}}\right) \right] = \\
 &= P_{\hat{F}}\left(x_{\alpha/2}^* \leq \sqrt{n}(\bar{X}^* - \bar{X}) \leq x_{1-\alpha/2}^*\right) + O_P\left(n^{-\frac{1}{2}}\right) = \boxed{1 - \alpha + O_P\left(n^{-\frac{1}{2}}\right)}
 \end{aligned}$$

Notemos que este error de recubrimiento es peor que el de la teoría clásica:

$$\begin{aligned}
 P_F(\mu \in I_0) &= 1 - \alpha + O(n^{-1}) \\
 P_F(\mu \in I_1) &= 1 - \alpha + O_P\left(n^{-\frac{1}{2}}\right)
 \end{aligned}$$

### **MÉTODO PERCENTIL-t**

Si se conociera la distribución de la variable aleatoria

$$\sqrt{n} \frac{\hat{\theta} - \theta}{\hat{\sigma}}$$

bajo  $F$ , entonces, si  $x_\alpha$  es el cuantil de orden  $\alpha$  de esa distribución,

$$P_F\left(x_{\alpha/2} \leq \sqrt{n} \frac{\hat{\theta} - \theta}{\hat{\sigma}} \leq x_{1-\alpha/2}\right) = 1 - \alpha$$

y por lo tanto:

$$I = \left(\hat{\theta} - \hat{\sigma} n^{-\frac{1}{2}} x_{1-\alpha/2}, \hat{\theta} - \hat{\sigma} n^{-\frac{1}{2}} x_{\alpha/2}\right) \quad \text{I.C. para } \theta \text{ de nivel } (1 - \alpha).$$

El método percentil-t aproxima la distribución de  $R(\bar{X}; F) = \sqrt{n} \frac{\hat{\theta} - \theta}{\hat{\sigma}}$  bajo  $F$

por la de  $R(\bar{X}^*; \hat{F}) = \sqrt{n} \frac{\hat{\theta}^* - \hat{\theta}}{\hat{\sigma}^*}$  bajo  $\hat{F}$ , es decir

$$P_F \left( \sqrt{n} \frac{\hat{\theta} - \theta}{\hat{\sigma}} \leq y \right) = P_{\hat{F}} \left( \sqrt{n} \frac{\hat{\theta}^* - \hat{\theta}}{\hat{\sigma}^*} \leq y \right),$$

y por lo tanto, si  $x_\alpha^*$  es el cuantil de orden  $\alpha$  de  $\sqrt{n} \frac{\hat{\theta}^* - \hat{\theta}}{\hat{\sigma}^*}$ ,

$$I_2 = \left( \hat{\theta} - \hat{\sigma} n^{\frac{1}{2}} x_{1-\alpha/2}^*, \hat{\theta} - \hat{\sigma} n^{\frac{1}{2}} x_{\alpha/2}^* \right) \text{ I.C. BOOT. PERCENTIL - } t \text{ de nivel } (1-\alpha)$$

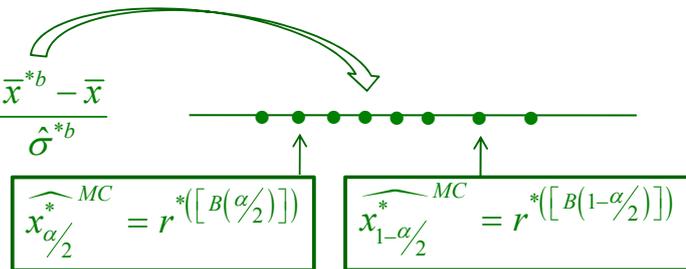
En la práctica,  $x_\alpha^*$  se aproxima por Monte Carlo a partir de  $\bar{x}$ . Así, si  $\theta = \mu$  y  $\hat{\theta} = \bar{X}$

$$X^* \equiv \begin{pmatrix} 1/n & \dots & 1/n \\ x_1 & \dots & x_n \end{pmatrix}$$

↓

$$\bar{x}^{*b} \rightarrow r^{*b} = R(\bar{x}^{*b}; \hat{F}) = \sqrt{n} \frac{\bar{x}^{*b} - \bar{x}}{\hat{\sigma}^{*b}}$$

$b = 1, \dots, B$



### ERROR DE COBERTURA MÉTODO PERCENTIL-t

Consideremos  $\theta = \mu$  y  $\hat{\theta} = \bar{X}$ .

Por el teorema de Bhattacharya-Ghost, sabemos que

$$P_F \left( \sqrt{n} \frac{\bar{X} - \mu}{\hat{\sigma}} \leq y \right) = \Phi(y) + n^{-\frac{1}{2}} \left( \frac{1}{6} \frac{k_3}{\sigma^3} (2y^2 + 1) \right) \phi(y) + O(n^{-1}).$$

Reemplazando momentos poblacionales por muestrales se tiene una expresión análoga para  $\sqrt{n} \frac{\bar{X}^* - \bar{X}}{\hat{\sigma}^*}$ .

Además, también sabemos que

$$\hat{\alpha}_j = \alpha_j + O_p \left( n^{-\frac{1}{2}} \right).$$

Restando ambas expresiones se deduce por lo tanto:

$$P_F \left( \sqrt{n} \frac{\bar{X} - \mu}{\hat{\sigma}} \leq y \right) - P_{\hat{F}} \left( \sqrt{n} \frac{\bar{X}^* - \bar{X}}{\hat{\sigma}^*} \leq y \right) = O_p(n^{-1}).$$

En consecuencia,

$$\begin{aligned}
 \boxed{P_F(\mu \in I_2)} &= P_F\left(\bar{X} - \hat{\sigma}n^{-\frac{1}{2}}x_{1-\alpha/2}^* \leq \mu \leq \bar{X} - \hat{\sigma}n^{-\frac{1}{2}}x_{\alpha/2}^*\right) = P_F\left(x_{\alpha/2}^* \leq \sqrt{n}\frac{\bar{X} - \mu}{\hat{\sigma}} \leq x_{1-\alpha/2}^*\right) = \\
 &= P_F\left(\sqrt{n}\frac{\bar{X} - \mu}{\hat{\sigma}} \leq x_{1-\alpha/2}^*\right) - P_F\left(\sqrt{n}\frac{\bar{X} - \mu}{\hat{\sigma}} \leq x_{\alpha/2}^*\right) = \\
 &= \left[ P_{\hat{F}}\left(\sqrt{n}\frac{\bar{X}^* - \bar{X}}{\hat{\sigma}} \leq x_{1-\alpha/2}^*\right) + O_P(n^{-1}) \right] - \left[ P_{\hat{F}}\left(\sqrt{n}\frac{\bar{X}^* - \bar{X}}{\hat{\sigma}} \leq x_{\alpha/2}^*\right) + O_P(n^{-1}) \right] = \\
 &= P_{\hat{F}}\left(x_{\alpha/2}^* \leq \sqrt{n}\frac{\bar{X}^* - \bar{X}}{\hat{\sigma}} \leq x_{1-\alpha/2}^*\right) + O_P(n^{-1}) = \boxed{1 - \alpha + O_P(n^{-1})}
 \end{aligned}$$

Notemos que este error de recubrimiento coincide con el de la teoría clásica:

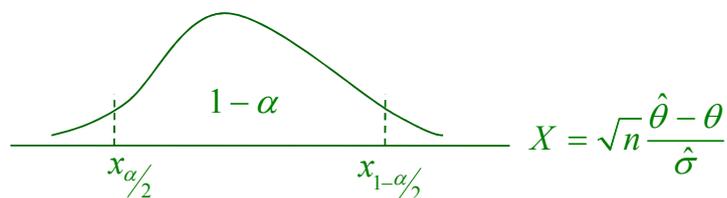
$$P_F(\mu \in I_0) = 1 - \alpha + O(n^{-1})$$

$$P_F(\mu \in I_2) = 1 - \alpha + O_P(n^{-1})$$

### MÉTODO PERCENTIL-t SIMETRIZADO

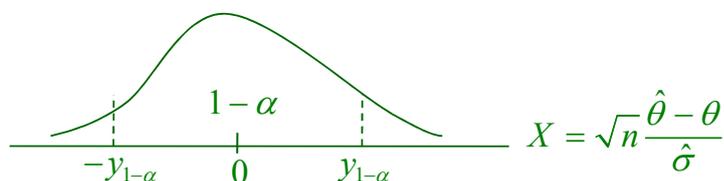
Es análogo al método percentil-t, pero difiere en la forma de estimar los cuantiles:

- El método percentil-t (anterior) considera intervalos de “colas iguales”:



- Los intervalos que considera el método percentil-t simetrizado no tienen por qué ser de “colas iguales”, al estar definidos a partir del cuantil  $1 - \alpha$  de  $|X|$ .

$$P(|X| \leq y_{1-\alpha}) = 1 - \alpha$$



Conociéndose la distribución de la variable aleatoria

$$\sqrt{n} \frac{\hat{\theta} - \theta}{\hat{\sigma}}$$

bajo  $F$ , si  $y_{1-\alpha}$  es el cuantil de orden  $1 - \alpha$  de su valor absoluto,

$$P_F \left( \sqrt{n} \left| \frac{\hat{\theta} - \theta}{\hat{\sigma}} \right| \leq y_{1-\alpha} \right) = 1 - \alpha$$

entonces:

$$I = \left( \hat{\theta} \mp \hat{\sigma} n^{-\frac{1}{2}} y_{1-\alpha} \right) \text{ I.C. para } \theta \text{ de nivel } (1 - \alpha).$$

El método percentil-t simetrizado aproxima la distribución de  $R(\bar{X}; F) = \sqrt{n} \left| \frac{\hat{\theta} - \theta}{\hat{\sigma}} \right|$  bajo  $F$  por la de  $R(\bar{X}^*; \hat{F}) = \sqrt{n} \frac{\hat{\theta}^* - \hat{\theta}}{\hat{\sigma}^*}$  bajo  $\hat{F}$ , es decir

$$P_F \left( \sqrt{n} \left| \frac{\hat{\theta} - \theta}{\hat{\sigma}} \right| \leq y \right) \approx P_{\hat{F}} \left( \sqrt{n} \left| \frac{\hat{\theta}^* - \hat{\theta}}{\hat{\sigma}^*} \right| \leq y \right),$$

y por lo tanto, si  $y_{1-\alpha}^*$  es el cuantil de orden  $1 - \alpha$  de  $\sqrt{n} \left| \frac{\hat{\theta}^* - \hat{\theta}}{\hat{\sigma}^*} \right|$ ,

$$I_3 = \left( \hat{\theta} \mp \hat{\sigma} n^{-\frac{1}{2}} y_{1-\alpha}^* \right) \text{ I.C. BOOT. PERCENTIL - t SIMETRIZADO de nivel } (1 - \alpha)$$

En la práctica,  $y_{1-\alpha}^*$  se aproxima por Monte Carlo a partir de  $\bar{x}$ . Así, si  $\theta = \mu$  y  $\hat{\theta} = \bar{X}$

$$X^* \equiv \begin{pmatrix} 1/n & \dots & 1/n \\ x_1 & \dots & x_n \end{pmatrix}$$

↓

$$\bar{x}^{*b} \rightarrow r^{*b} = R(\bar{x}^{*b}; \hat{F}) = \sqrt{n} \left| \frac{\bar{x}^{*b} - \bar{x}}{\hat{\sigma}^{*b}} \right|$$

$b = 1, \dots, B$



$$y_{1-\alpha}^* \approx \widehat{y_{1-\alpha}^*}^{MC} = r^{*([B(1-\alpha)])}$$

## ERROR DE COBERTURA MÉTODO PERCENTIL-t SIMETRIZADO

Consideremos  $\theta = \mu$  y  $\hat{\theta} = \bar{X}$ .

Por el teorema de Bhattacharya-Ghost, sabemos que

$$P_F \left( \sqrt{n} \frac{\bar{X} - \mu}{\hat{\sigma}} \leq y \right) = \Phi(y) + n^{-\frac{1}{2}} q_1(y) \phi(y) + n^{-1} q_2(y) \phi(y) + O \left( n^{-\frac{3}{2}} \right).$$

siendo  $q_1$  un polinomio de orden par simétrico. Se tiene entonces:

$$\begin{aligned} P_F \left( \sqrt{n} \left| \frac{\bar{X} - \mu}{\hat{\sigma}} \right| \leq y \right) &= P_F \left( \sqrt{n} \frac{\bar{X} - \mu}{\hat{\sigma}} \leq y \right) - P_F \left( \sqrt{n} \frac{\bar{X} - \mu}{\hat{\sigma}} \leq -y \right) = \\ &= \left[ \Phi(y) + n^{-\frac{1}{2}} q_1(y) \phi(y) + n^{-1} q_2(y) \phi(y) + O \left( n^{-\frac{3}{2}} \right) \right] - \\ &= - \left[ \underbrace{\Phi(-y)}_{=1-\Phi(y)} + n^{-\frac{1}{2}} \underbrace{q_1(-y)}_{=q_1(y)} \underbrace{\phi(-y)}_{=\phi(y)} + n^{-1} \underbrace{q_2(-y)}_{=-q_2(y)} \underbrace{\phi(-y)}_{=\phi(y)} + O \left( n^{-\frac{3}{2}} \right) \right] = \\ &= 2\Phi(y) - 1 + 2n^{-1} q_2(y) \phi(y) + O \left( n^{-\frac{3}{2}} \right). \end{aligned}$$

Como reemplazando momentos poblacionales por muestrales (cuyo orden de aproximación conocemos,  $O(n^{-1/2})$ ) se tiene una expresión análoga para la correspondiente variable bootstrap:

$$P_F \left( \sqrt{n} \left| \frac{\bar{X} - \mu}{\hat{\sigma}} \right| \leq y \right) - P_F \left( \sqrt{n} \left| \frac{\bar{X}^* - \bar{X}}{\hat{\sigma}^*} \right| \leq y \right) = O_p \left( n^{-\frac{3}{2}} \right).$$

En consecuencia,

$$\begin{aligned} P_F(\mu \in I_3) &= P_F \left( \bar{X} - \hat{\sigma} n^{-\frac{1}{2}} y_{1-\alpha}^* \leq \mu \leq \bar{X} + \hat{\sigma} n^{-\frac{1}{2}} y_{1-\alpha}^* \right) = P_F \left( \sqrt{n} \left| \frac{\bar{X} - \mu}{\hat{\sigma}} \right| \leq y_{1-\alpha}^* \right) = \\ &= P_{\hat{F}} \left( \sqrt{n} \left| \frac{\bar{X}^* - \bar{X}}{\hat{\sigma}^*} \right| \leq y_{1-\alpha}^* \right) + O_p \left( n^{-\frac{3}{2}} \right) = 1 - \alpha + O_p \left( n^{-\frac{3}{2}} \right). \end{aligned}$$

Notemos que este error de recubrimiento mejora al de la teoría clásica:

$$\begin{aligned} P_F(\mu \in I_0) &= 1 - \alpha + O(n^{-1}) \\ P_F(\mu \in I_3) &= 1 - \alpha + O_p \left( n^{-\frac{3}{2}} \right) \end{aligned}$$

Variables pivotaes "aproximadas" para una media poblacional	Intervalos de confianza bilaterales para la media poblacional de nivel $1 - \alpha$	Error de recubrimiento $P_F(\mu \in I_j) - (1 - \alpha)$
$\sqrt{n} \frac{\bar{X} - \mu}{\hat{\sigma}} \approx N(0,1)$	$I_0 = \left( \bar{X} \mp \hat{\sigma} n^{\frac{1}{2}} z_{1-\alpha/2} \right)$	$O(n^{-1})$
$\sqrt{n}(\bar{X} - \mu) \approx \sqrt{n}(\bar{X}^* - \bar{X})$	$I_1 = \left( \bar{X} - n^{\frac{1}{2}} x_{1-\alpha/2}^*, \bar{X} - n^{\frac{1}{2}} x_{\alpha/2}^* \right)$	$O_P\left(n^{-\frac{1}{2}}\right)$
$\sqrt{n} \frac{\bar{X} - \mu}{\hat{\sigma}} \approx \sqrt{n} \frac{\bar{X}^* - \bar{X}}{\hat{\sigma}^*}$	$I_2 = \left( \bar{X} - \hat{\sigma} n^{\frac{1}{2}} x_{1-\alpha/2}^*, \bar{X} - \hat{\sigma} n^{\frac{1}{2}} x_{\alpha/2}^* \right)$	$O_P(n^{-1})$
$\sqrt{n} \left  \frac{\bar{X} - \mu}{\hat{\sigma}} \right  \approx \sqrt{n} \left  \frac{\bar{X}^* - \bar{X}}{\hat{\sigma}^*} \right $	$I_3 = \left( \bar{X} \mp \hat{\sigma} n^{\frac{1}{2}} y_{1-\alpha}^* \right)$	$O_P\left(n^{-\frac{3}{2}}\right)$

- Notemos que el método **percentil-t simetrizado** (cuarta aproximación: estadístico positivo corregido por su escala) es mejor que el **percentil-t** (tercera aproximación: estadístico sólo corregido por su escala) y éste a su vez mejor que el método **percentil** (segunda aproximación: estadístico sin correcciones para sus sesgo y escala).

## ESTUDIO DE SIMULACIÓN

### Comparativa aproximaciones normal y bootstrap. P9

$$\begin{aligned}
 X &\in \text{Exp}\left(\frac{1}{100}\right) & \mu = EX = 100 \\
 \downarrow \\
 \vec{x}_i &= \begin{pmatrix} x_1 \\ \dots \\ x_n \end{pmatrix}_i \equiv X_1^* \equiv \begin{pmatrix} 1/n & \dots & 1/n \\ x_1 & \dots & x_n \end{pmatrix}_i \rightarrow \vec{x}_i^{*b} = \begin{pmatrix} x_1^* \\ \dots \\ x_n^* \end{pmatrix}_i, \quad b = 1, \dots, B \rightarrow I_0^i, I_1^i, I_2^i, I_3^i \\
 & & & & & (I_0^i \text{ sólo depende de } \vec{x}_i) \\
 i &= 1, \dots, N \quad (N \text{ m.a.s. artificiales})
 \end{aligned}$$

$$i = 1, \dots, N \rightarrow \begin{cases} \left. \begin{array}{l} \text{si } 100 \in I_0^i \rightarrow 1 \\ \text{si } 100 \notin I_0^i \rightarrow 0 \end{array} \right\} \rightarrow \frac{1}{N} \sum_{i=1}^N 1, 0 \dots = \widehat{cob.}^{MC} \text{ de } I_0 \\ \left. \begin{array}{l} \text{si } 100 \in I_1^i \rightarrow 1 \\ \text{si } 100 \notin I_1^i \rightarrow 0 \end{array} \right\} \rightarrow \frac{1}{N} \sum_{i=1}^N 1, 0 \dots = \widehat{cob.}^{MC} \text{ de } I_1 \\ \left. \begin{array}{l} \text{si } 100 \in I_2^i \rightarrow 1 \\ \text{si } 100 \notin I_2^i \rightarrow 0 \end{array} \right\} \rightarrow \frac{1}{N} \sum_{i=1}^N 1, 0 \dots = \widehat{cob.}^{MC} \text{ de } I_2 \\ \left. \begin{array}{l} \text{si } 100 \in I_3^i \rightarrow 1 \\ \text{si } 100 \notin I_3^i \rightarrow 0 \end{array} \right\} \rightarrow \frac{1}{N} \sum_{i=1}^N 1, 0 \dots = \widehat{cob.}^{MC} \text{ de } I_3 \end{cases}$$

**#Estimación por Monte Carlo de la cobertura de los intervalos bilaterales para la media: clásico, boot.percentil, boot.percentil-t #y boot.percentil-t simetrizado, en un estudio de simulación. Escenario exponencial (rate=1/100).**

#####

```
t1=proc.time()
alpha=0.1
n=100
N=1000
cobertural0=numeric(N)
cobertural1=numeric(N)
cobertural2=numeric(N)
cobertural3=numeric(N)
muestra=numeric(n)
for (z in 1:N)
{
  muestra<-rexp(n,rate=0.01)
  media=mean(muestra)
  desv=sd(muestra)
  l0=c(media-desv*(qnorm(1-(alpha/2))/sqrt(n)),media+desv*(qnorm(1-(alpha/2))/sqrt(n)))
  if ((l0[1]<100)&(100<l0[2]))
    {cobertural0[z]=1} → CLÁSICO

  remuestra<-numeric(n)
  B=1000
  percentil<-numeric(B)
  percentilord<-numeric(B)
  percentilt<-numeric(B)
  percentiltord<-numeric(B)
  percentilts<-numeric(B)
  percentiltsord<-numeric(B)
  for (k in 1:B)
  {
    for (i in 1:n)
    {
      j=1
      u=runif(1)
      while (u>=j/n)
      {
        j=j+1
      }
      remuestra[i]=muestra[j];
    }
    percentil[k]=sqrt(n)*(mean(remuestra)-media)
    percentilord[k]=sqrt(n)*(mean(remuestra)-media)/sd(remuestra)
    percentilts[k]=sqrt(n)*(abs(mean(remuestra)-media))/sd(remuestra)
  }

  percentilord=sort(percentil)
  percentilord=sort(percentilt)
  percentiltsord=sort(percentilts)
}
```

$1 - \alpha = 0.9$	Clásico	Percentil	Percentil-t	Percentil-t simetrizado	Time user
n=100 B=1000 N=1000	0.890 (3º)	0.887 (4º)	0.896 (2º)	0.900 (1º)	3.18 h.

```
cuantil1=percentilord[floor(B*(1-(alpha/2)))]
cuantil1s=percentilord[floor(B*(alpha/2))]
l1=c((media-cuantil1)/sqrt(n),media-cuantil1s/sqrt(n))
if ((l1[1]<100)&(100<l1[2]))
  {cobertural1[z]=1} → PERCENTIL

cuantil2=percentiltord[floor(B*(1-(alpha/2)))]
cuantil2s=percentiltord[floor(B*(alpha/2))]
l2=c((media-desv*cuantil2)/sqrt(n),media-desv*cuantil2s/sqrt(n))
if ((l2[1]<100)&(100<l2[2]))
  {cobertural2[z]=1} → PERCENTIL-T

cuantil3=percentiltsord[floor(B*(1-alpha))]
l3=c((media-desv*cuantil3)/sqrt(n),media-desv*cuantil3/sqrt(n))
if ((l3[1]<100)&(100<l3[2]))
  {cobertural3[z]=1} → PERCENTIL-T SIMETRIZADO
}
```

**COBERTURAS ESTIMADAS POR MONTE CARLO PARA**

$$X \in \text{Exp}\left(\frac{1}{100}\right) \quad (\mu = 100)$$

$$1 - \alpha = 0,9$$

$$N = 100$$

n	B	$I_0$	$I_1$	$I_2$	$I_3$
10	100	0,78	0,40	0,55	0,60
	500		0,51	0,66	0,65
25	100	0,82	0,57	0,65	0,71
	500		0,68	0,70	0,76
100	100	0,84	0,71	0,70	0,83
	500		0,76	0,82	0,88
200	100	0,92	0,82	0,82	0,94
	500		0,86	0,92	0,89