

Bootstrap

Rogelio Ramos Quiroga

`rramosq@cimat.mx`

Noviembre 2020

Dónde empezó esto?

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

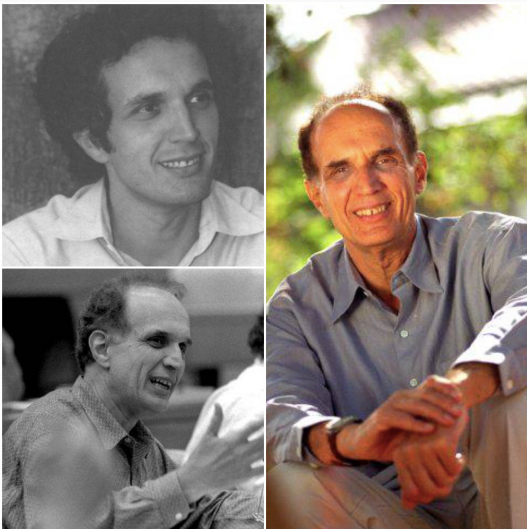
THE 1977 RIETZ LECTURE

BOOTSTRAP METHODS: ANOTHER LOOK AT THE JACKKNIFE

BY B. EFRON

Stanford University

Bradley Efron (USA, 1938 -)



Jackknife

Jackknife 1

- Sean x_1, x_2, \dots, x_n i.i.d. F . La media muestral $\bar{x} = \sum x_i / n$ estima la media $\mu = E(x)$ y, más aún, los datos proveen, no sólo información sobre μ , también dan información sobre la precisión de \bar{x} ,

$$\hat{\sigma}_{\bar{x}}^2 = \frac{1}{n} \hat{\sigma}^2 = \frac{1}{n(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2$$

- Un problema con esta fórmula para la precisión de \bar{x} , es que no es fácilmente extendible a otros estimadores. Pero puede reescribirse de forma que puede generalizarse. Sea $\bar{x}_{(i)}$ la media aritmética de x_1, x_2, \dots, x_n , pero sin incluir a x_i

$$\bar{x}_{(i)} = \frac{1}{n-1} \sum_{j \neq i} x_j = \frac{1}{n-1} \left[\sum_{j=1}^n x_j - x_i \right] = \frac{n\bar{x} - x_i}{n-1},$$

Jackknife 2

- Definamos $\bar{x}_{(.)}$ como la media de todas las $\bar{x}_{(i)}$'s,

$$\bar{x}_{(.)} = \frac{1}{n} \sum_{i=1}^n \bar{x}_{(i)} = \frac{n^2}{n(n-1)} \bar{x} - \frac{1}{n-1} \bar{x} = \bar{x}.$$

- El estimador **Jackknife**¹ de la desviación estándar de \bar{x} es

$$\hat{\sigma}_J = \left[\frac{n-1}{n} \sum_{i=1}^n (\bar{x}_{(i)} - \bar{x}_{(.)})^2 \right]^{1/2}$$

el cual, puede verse que es, precisamente, el estimador usual arriba mencionado

$$\hat{\sigma}_J^2 = \frac{1}{n(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2 = \hat{\sigma}_{\bar{x}}^2.$$

- Ahora, la expresión de arriba, si es generalizable.

¹Tukey, J. (1958), Bias and confidence in not quite large samples, (abstract), Ann. Math. Statist., 29, p. 614

Jackknife 3

- Sea $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$, un estimador de un parámetro θ . El estimador **Jackknife** de la desviación estándar de $\hat{\theta}$ se define como

$$\hat{\sigma}_J = \left[\frac{n-1}{n} \sum_{i=1}^n (\hat{\theta}_{(i)} - \hat{\theta}_{(.)})^2 \right]^{1/2}$$

donde

$$\hat{\theta}_{(i)} = \hat{\theta}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$$

y

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



Jackknife 4

```
# Jackknife básico
set.seed(4347)
n      = 40
med    = 10
des    = 1
data   = rnorm(n,med,des)
xb     = mean(data)           # 9.770336
err    = sd(data)/sqrt(n)    # 0.1405979

teti   = rep(0,n)
for(i in 1:n){
  dati  = data[-i]
  teti[i] = mean(dati) }

tetb   = mean(teti)           # 9.770336
sdJ    = sqrt(((n-1)/n)*sum((teti-tetb)^2)) # 0.1405979
```


Jackknife 5

- Una propiedad del Jackknife: Reducción de sesgo.
Supongamos que $\hat{\theta}$ es un estimador sesgado de θ , con sesgo de orden $1/n$

$$E(\hat{\theta}) = \theta + \left(\frac{a_1}{n} + \frac{a_2}{n^2} + \frac{a_3}{n^3} + \dots \right)$$

- El estimador Jackknife para el sesgo se define como

$$\hat{B}_J = (n-1) \left(\hat{\theta}_{(\cdot)} - \hat{\theta} \right)$$

así, el estimador Jackknife para θ se define como

$$\tilde{\theta} = \hat{\theta} - \hat{B}_J = \hat{\theta} - (n-1) \left(\hat{\theta}_{(\cdot)} - \hat{\theta} \right) = n\hat{\theta} - (n-1)\hat{\theta}_{(\cdot)}.$$

- Note que

$$E(\hat{\theta}_{(i)}) = \theta + \left(\frac{a_1}{n-1} + \frac{a_2}{(n-1)^2} + \frac{a_3}{(n-1)^3} + \dots \right)$$

Jackknife 6

- ... de modo que

$$E(\widehat{\theta}_{(.)}) = E\left(\frac{1}{n} \sum_{i=1}^n \widehat{\theta}_{(i)}\right) = \theta + \left(\frac{a_1}{n-1} + \frac{a_2}{(n-1)^2} + \frac{a_3}{(n-1)^3} + \dots\right)$$

- de aquí tenemos

$$\begin{aligned} E(\tilde{\theta}) &= nE(\widehat{\theta}) - (n-1)E(\widehat{\theta}_{(.)}) \\ &= \left[n\theta + a_1 + \frac{a_2}{n} + \frac{a_3}{n^2} + \dots\right] - \left[(n-1)\theta + a_1 + \frac{a_2}{(n-1)} + \frac{a_3}{(n-1)^2} + \dots\right] \\ &= \theta + a_2 \left[\frac{1}{n} - \frac{1}{n-1}\right] + a_3 \left[\frac{1}{n^2} - \frac{1}{(n-1)^2}\right] + \dots \end{aligned}$$

- el coeficiente dominante

$$\frac{1}{n} - \frac{1}{n-1} = -\frac{1}{n(n-1)} \approx -\frac{1}{n^2}$$

- de modo que si $\widehat{\theta}$ tiene sesgo de orden $1/n$, entonces $\tilde{\theta}$ tiene sesgo menor, de orden $1/n^2$.

Bootstrap

Bootstrap básico 1

Suponga que x_1, x_2, \dots, x_n son i.i.d. $N(\mu, \sigma^2)$ y queremos una estimación de la media, μ , y su error estándar

$$\text{errstd}(\hat{\mu}) = \sqrt{\widehat{\text{Var}}(\hat{\mu})},$$

es claro que $\hat{\mu} = \bar{x}$

$$\text{errstd}(\hat{\mu}) = \sqrt{\widehat{\text{Var}}(\bar{x})} \stackrel{!}{=} \sqrt{\frac{\hat{\sigma}^2}{n}} = \sqrt{\frac{s^2}{n}} = \sqrt{\frac{1}{n} \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}},$$

un punto importante en el cálculo anterior, es que conocemos la expresión para la varianza

$$\text{Var}(\hat{\mu}) = \text{Var}(\bar{x}) = \frac{\sigma^2}{n}.$$

El Bootstrap estima el error estándar usando simulación y, críticamente, no requiere que conozcamos esa (!) expresión explícita de la varianza.

Bootstrap básico 2

Para ello, simulamos B veces, con reemplazo, de $\{x_1, x_2, \dots, x_n\}$, para cada muestra bootstrap calculamos el estimador deseado y, finalmente, calculamos la desviación estándar muestral de esas B estimaciones, y ya está.

muestra	muestra bootstrap				estimador
1	x_1^1	x_2^1	\dots	x_n^1	$\hat{\mu}^1$
2	x_1^2	x_2^2	\dots	x_n^2	$\hat{\mu}^2$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
B	x_1^B	x_2^B	\dots	x_n^B	$\hat{\mu}^B$

El estimador Bootstrap² del error estándar es

$$\text{errstd}_b(\hat{\mu}) = \sqrt{\frac{1}{B-1} \sum_{b=1}^B (\hat{\mu}^b - \bar{\hat{\mu}})^2}, \quad \text{donde} \quad \bar{\hat{\mu}} = \frac{1}{B} \sum_{b=1}^B \hat{\mu}^b$$

²Efron, B. (1979) Bootstrap methods: another look at the jackknife, Ann. Statist., 7, 1, pp. 1-26.

Bootstrap básico 3

```
# Bootstrap no-paramétrico básico
set.seed(4347)
n      = 40
med    = 10
des    = 1
data   = rnorm(n, med, des)
xb     = mean(data)          # 9.770336
err    = sd(data)/sqrt(n)

B      = 5000
xbs    = rep(0,B)
for(i in 1:B){
  sel   = sample(1:n, size=n, replace=TRUE)
  xx    = data[sel]
  xbs[i] = mean(xx) }

sd(xbs)  # 0.1377988 = error estándar bootstrap
err      # 0.1405979 = error estándar
```

Bootstrap básico 4

Un ejemplo menos trivial es el cálculo del error estándar del estimador de Pearson del coeficiente de correlación³. Suponga

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) \sim \text{i.i.d. } N_2(\mu, \Sigma)$$

con

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}, \quad \rho = \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} \quad \text{y} \quad r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

r es el estimador de Pearson de ρ , queremos su error estándar

$$\text{errstd}(r) = \sqrt{\widehat{\text{Var}}(r)} \stackrel{?}{=} \dots$$

en este caso, no se cuenta con expresiones explícitas para la varianza de r , sin embargo, el procedimiento anterior es aplicable.

³Efron, B. (1982), The jackknife, the bootstrap and other resampling plans. SIAM.

Bootstrap básico 5

muestra	muestra bootstrap				estimador
1	$(x_1, y_1)^1$	$(x_2, y_2)^1$	\cdots	$(x_n, y_n)^1$	$\hat{\rho}^1$
2	$(x_1, y_1)^2$	$(x_2, y_2)^2$	\cdots	$(x_n, y_n)^2$	$\hat{\rho}^2$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
B	$(x_1, y_1)^B$	$(x_2, y_2)^B$	\cdots	$(x_n, y_n)^B$	$\hat{\rho}^B$

El estimador Bootstrap del error estándar es

$$\text{errstd}_b(r) = \sqrt{\frac{1}{B-1} \sum_{b=1}^B (\hat{\rho}^b - \bar{\hat{\rho}})^2}, \quad \text{donde} \quad \bar{\hat{\rho}} = \frac{1}{B} \sum_{b=1}^B \hat{\rho}^b$$

Bootstrap básico 6

```
set.seed(9320)
lsat  = c(576,635,558,578,666,580,555,661,651,605,653,575,545,572,594)
gpa   = c(339,330,281,303,344,307,300,343,336,313,312,274,276,288,296)
n     = length(gpa)
robs  = cor(lsat,gpa) # correlación observada = 0.7763745
# Cómo estimar su error estándar?
B     = 5000
b     = rep(0,B)
for(i in 1:B){
  sel  = sample(1:n,size=n,replace=TRUE)
  b[i] = cor(lsat[sel],gpa[sel]) }
s = sd(b)
# s = 0.1333642
```

Bootstrap básico 7

Para estimar el error estándar de r desde principios básicos, necesitamos la distribución de r . Fisher (1915)⁴:

$$f(r) = \frac{n-2}{\sqrt{2\pi}} \frac{\Gamma(n-1)}{\Gamma(n-1/2)} \frac{(1-\rho^2)^{(n-1)/2} (1-r^2)^{(n-4)/2}}{(1-\rho r)^{n-3/2}} {}_2F_1\left(\frac{1}{2}, \frac{1}{2}, \frac{2n-1}{2}, \frac{\rho r+1}{2}\right)$$

$${}_2F_1(a, b, c, z) = \sum_{j=0}^{\infty} \frac{(a)_j (b)_j}{(c)_j} \frac{z^j}{j!}$$

$$(q)_j = \begin{cases} 1 & j = 0 \\ q(q+1) \cdots (q+j-1) & j > 0 \end{cases}$$

donde ${}_2F_1(a, b, c, z)$ es la Función Hipergeométrica Gaussiana y $(q)_j$ denota el símbolo de Pochhammer. Calculamos la varianza evaluando $E(r)$ y $E(r^2)$ (cosa que, definitivamente, no queremos - o podemos - hacer!).

⁴Fisher, R.A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika, 10, 4, 507-521

Bootstrap básico 8

La aplicabilidad del Bootstrap es muy amplia, Efron, en una revisión en 2003⁵ comenta que “During the past 25 years an enormous amount of statistical research has investigated the validity of the bootstrap approach. For most models P and most statistics $\hat{\theta}$, we know that the bootstrap standard deviation $sd_*(\hat{\theta}^*)$ is a good estimator for the true standard deviation $sd(\hat{\theta})$, and likewise for other accuracy measures”.

⁵Efron, B. (2003). Second thoughts on the Bootstrap. Statistical Science, 18, 2, 135-140

Bootstrap paramétrico 1

El Bootstrap (no-paramétrico) muestrea sin reemplazo de $\{x_1, x_2, \dots, x_n\}$, esto es, obtiene muestras i.i.d. de la distribución empírica

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n I(x_i \leq x),$$

alternativamente, el Bootstrap paramétrico muestrea de la distribución paramétrica estimada. Por ejemplo, en el caso del coeficiente de correlación, bajo el supuesto de normalidad, se muestrea de la distribución normal bivariada

$$N_2(\hat{\mu}, \hat{\Sigma}), \quad \text{con} \quad \hat{\mu} = \frac{1}{n} \sum_{i=1}^n z_i \quad \text{y} \quad \hat{\Sigma} = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix}$$

donde $z_i = (x_i, y_i)$ y las s_{ij} 's son las varianzas y covarianzas muestrales.

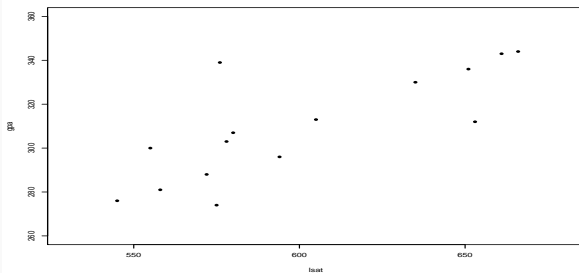
Bootstrap paramétrico 2

```
# Bootstrap Paramétrico.
set.seed(9320)
lsat = c(576,635,558,578,666,580,555,661,651,605,653,575,545,572,594)
gpa  = c(339,330,281,303,344,307,300,343,336,313,312,274,276,288,296)
robs = cor(lsat,gpa) # correlación observada = 0.7763745
plot(lsat,gpa,xlim=c(530,680),ylim=c(260,360),pch=16)
# (razonablemente como una normal bivariada)
X    = cbind(lsat,gpa)
mu   = colMeans(X)
Sig  = var(X)
# Supongamos que la población es:  $N(\mu, \text{Sig})$ , entonces, una forma
# de estimar el comportamiento distribucional de  $R_o$ , es muestrear
# de esta población.
library(mvtnorm)
n    = length(gpa)
B    = 5000
ros  = rep(0,B)
for(i in 1:B){
  mues = rmvnorm(n=n,mean=mu,sigma=Sig)
  ros[i] = cor(mues[,1],mues[,2]) }
s    = sd(ros)
# Estimación bootstrap del error estándar (bootstrap paramétrico)
# s = 0.119026
```

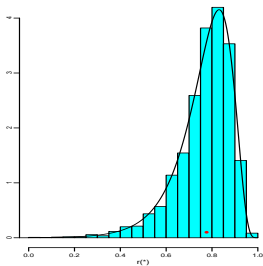
Bootstrap paramétrico 3

```
# Distribución del coeficiente de correlación de Pearson
library(hypergeo)
M = 201
rr = seq(0,1,length=M)
ff = rep(0,M)
for(i in 1:M){
  aa = ((n-2)*gamma(n-1))/(sqrt(2*pi)*gamma(n-.5))
  bb = (((1-robs^2)^((n-1)/2)) / ((1-robs*rr[i])^(n-1.5))
  cc = (1-rr[i]^2)^((n-4)/2)
  dd = hypergeo(.5,.5,(2*n-1)/2,(robs*rr[i]+1)/2)
  ff[i] = aa*bb*cc*dd }
par(mfrow=c(1,2),mar=c(4,4,1,1))
hist(ros,main="Réplicas (paramétricas) del Coef. de Corr.",
     cex.main=.9,xlab="r(*)",ylab="",cex.axis=.8,mgp=c(1.5,.5,0),
     col="cyan",nclass=20,xlim=c(0,1),probability=TRUE,ylim=c(0,4.5))
points(robs,.1,pch=16,col="red")
lines(rr,ff,lwd=2)
hist(b,main="Réplicas del Coef. de Corr.",
     cex.main=.9,xlab="r(*)",ylab="",cex.axis=.8,mgp=c(1.5,.5,0),
     col="cyan",nclass=20,xlim=c(0,1),probability=TRUE,ylim=c(0,4.5))
points(robs,.1,pch=16,col="red")
lines(rr,ff,lwd=2)
```

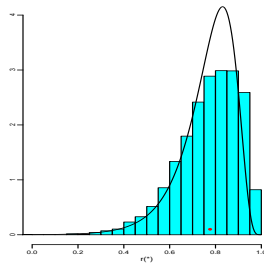
Bootstrap paramétrico 4



Réplicas (paramétricas) del Coef. de Corr.



Réplicas del Coef. de Corr.



Estimación plug-in

Estimación “plug-in” 1

Las versiones anteriores del Bootstrap estan basadas en simulación, sin embargo la versión más básica no requiere Monte Carlo. Sea θ una cantidad de interés acerca de una distribución F , podemos conceptualizar a θ como una función de F , $\theta = h(F)$, por ejemplo

θ	$=$	$h(F)$
μ	$=$	$\int x dF(x)$
σ^2	$=$	$\int x^2 dF(x) - (\int x dF(x))^2$
M	$=$	solución de $\int^M dF(x) = 1/2$

ahora, consideremos un estimador de θ , $\hat{\theta}_n$, basado en datos $\{x_1, x_2, \dots, x_n\}$. Para evaluar el desempeño de este estimador, podemos estar interesados en evaluar, por ejemplo, la distribución del error

$$\lambda_n(F) = P_F \left\{ \sqrt{n}[\hat{\theta}_n - h(F)] \leq a \right\},$$

Estimación “plug-in” 2

La idea clave de Efron consiste en considerar estimaciones de estas cantidades mediante el estimador “plug-in”, $\lambda_n(\hat{F})$, donde \hat{F} es la función de distribución empírica de la muestra. Para ilustrar cómo se puede estudiar el comportamiento de $\lambda_n(F)$ mediante $\lambda_n(\hat{F})$ consideremos un caso sencillo⁶. Supongamos $\theta = h(F) = E_F(x)$ y una muestra de tamaño $n = 2$, $\{x_1 = c, x_2 = d\}$ con $c < d$. El plug-in para $\lambda_n(F)$ es

$$\lambda_n(\hat{F}) = P_{\hat{F}} \left\{ \sqrt{n}[\hat{\theta}_n^* - h(\hat{F})] \leq a \right\}.$$

Veamos las diferentes componentes de esta expresión

$$h(\hat{F}) = E_{\hat{F}}(x^*) = \sum_{j=1}^n x_j P_{\hat{F}}(x^* = x_j) = \sum_{j=1}^n x_j \left(\frac{1}{n} \right) = \bar{x},$$

de modo que, con $n = 2$: $h(\hat{F}) = (c + d)/2$.

⁶Lehmann, E. (2004) Elements of Large Sample Theory. Springer

Estimación “plug-in” 3

El estimador de la media en el mundo bootstrap es

$\hat{\theta}_n^* = (x_1^* + x_2^*)/2$, y queremos la distribución de este estimador.

Los posibles valores para (x_1^*, x_2^*) y sus respectivas probabilidades son

(x_1^*, x_2^*)	$P_{\hat{F}}(x_1^*, x_2^*)$	$\hat{\theta}_n^*$
(c, c)	1/4	c
(c, d)	1/4	$(c + d)/2$
(d, c)	1/4	$(c + d)/2$
(d, d)	1/4	d

Entonces, la distribución del estimador de la media, así como la del error (en el mundo bootstrap) son:

$\hat{\theta}_n^*$	$P_{\hat{F}}(\hat{\theta}_n^*)$	$\sqrt{n}(\hat{\theta}_n^* - h(\hat{F}))$	$P_{\hat{F}}\left[\sqrt{n}(\hat{\theta}_n^* - h(\hat{F}))\right]$
c	1/4	$(c - d)/\sqrt{2}$	1/4
$(c + d)/2$	1/2	0	1/2
d	1/4	$(d - c)/\sqrt{2}$	1/4

Estimación “plug-in” 4

Entonces, la distribución de la variable aleatoria $\sqrt{n}(\hat{\theta}_n^* - h(\hat{F}))$, está completamente determinada y, por lo tanto, podemos calcular cualquier probabilidad relacionada con ella y, consecuentemente, podemos aproximar las correspondientes probabilidades de error de interés:

$$\begin{aligned}\lambda_n(F) &= P_F \left\{ \sqrt{n}[\hat{\theta}_n - h(F)] \leq a \right\} \\ &\approx P_{\hat{F}} \left\{ \sqrt{n}[\hat{\theta}_n^* - h(\hat{F})] \leq a \right\} = \lambda_n(\hat{F}).\end{aligned}$$

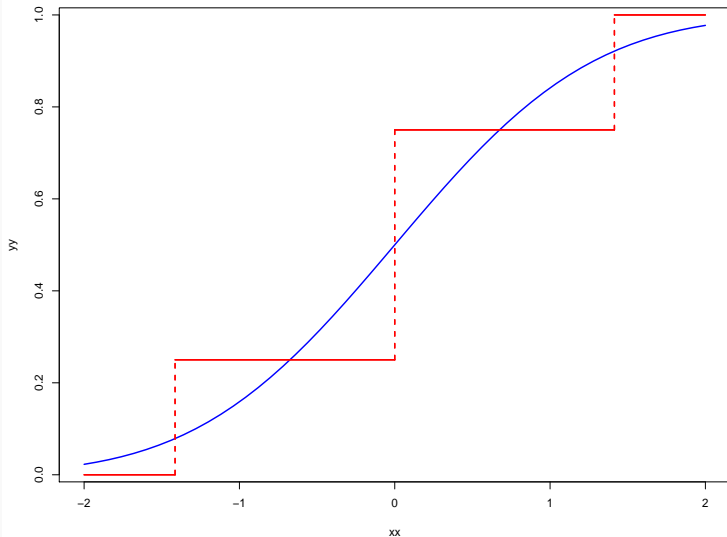
Note que podemos aproximar $\lambda_n(F)$ sin necesidad de recurrir a simulación, pero el cálculo explícito es factible sólo en algunos casos sencillos. En el caso (usual) de no poder hacer estos cálculos en forma cerrada, como conocemos la distribución \hat{F} , entonces podemos simular de ella y consecuentemente de $\hat{\theta}_n^*$, a este proceso es a lo que propiamente se le llama **bootstrap**.

Estimación “plug-in” 5

```
xx = seq(-2,2,len=51)
yy = pnorm(xx)
plot(xx,yy,lwd=2,col="blue",type="l")
segments(-2,0,-sqrt(2),0,lwd=2,col="red")
segments(-sqrt(2),0,-sqrt(2),1/4,lwd=2,col="red",lty=2)
segments(-sqrt(2),1/4,0,1/4,lwd=2,col="red")
segments(0,1/4,0,3/4,lwd=2,col="red",lty=2)
segments(0,3/4,sqrt(2),3/4,lwd=2,col="red")
segments(sqrt(2),3/4,sqrt(2),1,lwd=2,col="red",lty=2)
segments(sqrt(2),4/4,2,4/4,lwd=2,col="red")

set.seed(83838)
dat = c(3,5)
B = 1000
xb = rep(0,B)
for(i in 1:B){
  sel = sample(1:2,size=2,replace=TRUE)
  dd = dat[sel]
  xb[i] = mean(dd) }
zz = sqrt(2)*(xb-mean(dat))
table(zz)
-1.4142      0    1.4142
   236    518    246
```

Estimación “plug-in” 6



Consistencia

Consistencia del Bootstrap 1

Consideremos⁷ el caso de estimación Bootstrap de la media de una distribución. Supongamos x_1, x_2, \dots, x_n i.i.d. F , con $E(x) = \mu$, $\text{Var}(x) = \sigma^2$ y $E|x^3| = \rho$. Queremos darle sentido a la afirmación de que el comportamiento de $T_n = \sqrt{n}(\bar{x} - \mu)$ es aproximable mediante el comportamiento de $T_n^* = \sqrt{n}(\bar{x}^* - \bar{x})$. Definamos

$$H_n(t) = P_F(T_n \leq t) \quad \text{y} \quad H_n^*(t) = P_{\hat{F}}(T_n^* \leq t)$$

Queremos ver que $d(H_n^*, H_n) \xrightarrow{\text{a.s.}} 0$, donde

$$d(F, G) \equiv \sup_x |F(x) - G(x)|.$$

Sea $\hat{\sigma}^2 = \sum (x_i - \bar{x})^2 / n$. Observamos que \bar{x} y $\hat{\sigma}$ son la media y desviación estándar de la variable aleatoria x^* distribuida de acuerdo a la distribución empírica, y sean z_n y z_n^* versiones estandarizadas de T_n y T_n^* , respectivamente

$$z_n = \sqrt{n}(\bar{x} - \mu) / \sigma \quad \text{y} \quad z_n^* = \sqrt{n}(\bar{x}^* - \bar{x}) / \hat{\sigma}$$

⁷Wasserman, L. (2006). All of Nonparametric Statistics. Springer

Consistencia del Bootstrap 2

Ahora, denotemos $\Phi_a(x) \equiv \Phi(x/a)$, donde Φ es la función de distribución de la normal estándar. El resultado sobre la cota de Berry-Esseen, aplicado tanto a \bar{x} como a \bar{x}^* , nos dice que

$$\sup_z |P(z_n \leq z) - \Phi(z)| \leq \frac{C\rho}{\sqrt{n}\sigma^3} \quad \text{y} \quad \sup_z |P_{\hat{F}}(z_n^* \leq z) - \Phi(z)| \leq \frac{C\frac{1}{n} \sum |x_i - \bar{x}|^3}{\sqrt{n}\hat{\sigma}^3}$$

Note que

$$\begin{aligned} \sup_z |P_{\hat{F}}(z_n^* \leq z) - \Phi(z)| &= \sup_z |P_{\hat{F}}(\sqrt{n}(\bar{x}^* - \bar{x}) \leq z\hat{\sigma}) - \Phi(z\hat{\sigma}/\hat{\sigma})| \\ &= \sup_z |P_{\hat{F}}(\sqrt{n}(\bar{x}^* - \bar{x}) \leq z) - \Phi_{\hat{\sigma}}(z)| \\ &= d(H_n^*, \Phi_{\hat{\sigma}}) \end{aligned}$$

Por la desigualdad del triángulo, se tiene

$$d(H_n^*, H_n) \leq d(H_n^*, \Phi_{\hat{\sigma}}) + d(\Phi_{\hat{\sigma}}, \Phi_{\sigma}) + d(\Phi_{\sigma}, H_n) \equiv A + B + C$$

Consistencia del Bootstrap 3

- A está acotada y puede verse que converge a 0

$$A \leq \frac{C \frac{1}{n} \sum |x_i - \bar{x}|^3}{\sqrt{n} \hat{\sigma}^3} \xrightarrow{\text{a.s.}} 0$$

- B converge a 0, pues $\hat{\sigma} \xrightarrow{\text{a.s.}} \sigma$
- C converge a 0 por el Teorema Central del Límite

De modo que $d(H_n^*, H_n) \xrightarrow{\text{a.s.}} 0$. Las convergencias de A y B son en sentido estocástico y la de C en sentido de sucesiones de números reales.

Un resultado un poco más general acerca de la consistencia del bootstrap⁸: Supongamos que $E(x_1^2) < \infty$. Sea $T_n = g(\bar{x}_n)$, donde g es continuamente diferenciable en $\mu = E(x_1)$ y, además, $g'(\mu) \neq 0$. Entonces

$$\sup_u \left| P_{\hat{F}_n} \left(\sqrt{n} (T(\hat{F}_n^*) - T(\hat{F}_n)) \leq u \right) - P_F \left(\sqrt{n} (T(\hat{F}_n) - T(F)) \leq u \right) \right| \xrightarrow{\text{a.s.}} 0$$

⁸Wasserman, L. (2006). All of nonparametric statistics. Springer

Consistencia del Bootstrap 4

El bootstrap funciona bajo condiciones muy generales, sin embargo puede, por supuesto, fallar. La siguiente lista⁹ indica áreas en donde la aplicación del bootstrap no es automática y debe ponerse atención a métodos especiales:

- Muestras demasiado pequeñas
- Distribuciones con momentos infinitos
- Estimación de valores extremos
- Inferencia en poblaciones finitas
- Secuencias M -dependientes
- Procesos autoregresivos no estables
- Dependencias de largo plazo

⁹Chernick, M.R. (2008), Bootstrap methods: A guide for practitioners and researchers. Wiley

Intervalos de Confianza

Intervalos de confianza 1

Sea T un estadístico, estimador de θ . Supongamos que queremos un intervalo de confianza para θ , más aún, supongamos que conocemos la distribución de $T - \theta$ (p. ej. una cantidad pivotal) y podemos calcular sus cuantiles, i.e. podemos conocer a y b tales que

$$P(a \leq T - \theta \leq b) = 1 - \alpha$$

de aquí que

$$P(T - b \leq \theta \leq T - a) = 1 - \alpha$$

y, entonces, $(T - b, T - a)$ sería un intervalo del $100(1 - \alpha)\%$ de confianza para θ . En el mundo bootstrap aproximamos la distribución de $T - \theta$, mediante la distribución de $T^* - \hat{\theta}$, de aquí que una aproximación para a es $t_{(B+1)(\alpha/2)}^* - \hat{\theta}$ y para b sería $t_{(B+1)(1-\alpha/2)}^* - \hat{\theta}$.

Intervalos de confianza 2

Las expresiones para las aproximaciones bootstrap de los límites de un intervalo de confianza son

$$\begin{aligned}\widehat{t - b} &= \widehat{\theta} - [t_{(B+1)(1-\alpha/2)}^* - \widehat{\theta}] = 2\widehat{\theta} - t_{(B+1)(1-\alpha/2)}^* \\ \widehat{t - a} &= \widehat{\theta} - [t_{(B+1)(\alpha/2)}^* - \widehat{\theta}] = 2\widehat{\theta} - t_{(B+1)(\alpha/2)}^*\end{aligned}$$

Estos son los **límites de confianza bootstrap básicos**¹⁰. Ahora, acerca de la aproximación de la distribución de $T - \theta$, mediante la distribución de $T^* - \widehat{\theta}$: Si $F(u) = P(T - \theta \leq u)$ entonces su aproximación es

$$\widehat{F}_B(u) = \frac{\#\{t^* - \widehat{\theta} \leq u\}}{B} = \frac{1}{B} \sum_{i=1}^B I\{t_i^* - \widehat{\theta} \leq u\}$$

Conforme B aumenta, tenemos convergencia a la distribución empírica $\widehat{F}(u)$, y conforme n aumenta, tenemos convergencia a $F(u)$

$$\widehat{F}_B(u) \xrightarrow{B} \widehat{F}(u) \xrightarrow{n} F(u)$$

¹⁰Davison, A.C. and Hinkley, D.V. (1997) Bootstrap methods and their application. Cambridge Univ. Press

Intervalos de confianza 3

El **método de percentiles** es el más sencillo, sin embargo no se recomienda por no tener una justificación sólida. Supongamos B muestras bootstrap con los respectivos B valores de $\hat{\theta}_i^*$. El método consiste en ordenar esos B valores y calcular los correspondientes cuantiles

$$(\hat{\theta}_{\alpha/2}^*, \hat{\theta}_{1-\alpha/2}^*)$$

Este procedimiento puede justificarse si podemos asumir normalidad para los $\hat{\theta}_i^*$'s. En este caso, se tiene la ventaja de que los límites de confianza están en el mismo soporte que el parámetro de interés.

Intervalos de confianza 4

```
# Intervalos de confianza bootstrap
set.seed(9320)
lsat  = c(576,635,558,578,666,580,555,661,651,605,653,575,545,572,594)
gpa   = c(339,330,281,303,344,307,300,343,336,313,312,274,276,288,296)
robs  = cor(lsat,gpa) # correlación observada = 0.7763745
B     = 5000
b     = rep(0,B)
for(i in 1:B){
  sel  = sample(1:n,size=n,replace=TRUE)
  b[i] = cor(lsat[sel],gpa[sel]) }
s = sd(b)
# s = 0.1333642

# Intervalos de Confianza, 90%
# Método de percentiles:
icn = quantile(b,p=c(.05,.95))
      5%          95%
0.5200723 0.9456298

# Método Básico:
lininf = 2*robs-icn[2] # 0.607
limsup = 2*robs-icn[1] # 1.033
```


Intervalos de confianza 5

Los **intervalos BC_a (bias corrected and accelerated)**¹¹ son una mejora con respecto a los intervalos básico y de percentiles. Una mejora en el sentido de que son de “segundo orden”, esto es, la probabilidad de cobertura es de la forma c/n , mientras que los otros intervalos son de “primer orden” pues su cobertura es de la forma a/\sqrt{n} . Como su nombre indica, el BC_a , corrige por sesgo del plug-in, en forma automática. Los límites de confianza quedan en función de los cuantiles de $\hat{\theta}^*$, similar al método de percentiles, pero también dependen de dos parámetros más

a = aceleración

z_0 = corrección por sesgo

El intervalo BC_a , con cobertura $1 - 2\alpha$, es

$$BC_a = (\hat{\theta}_{lo}, \hat{\theta}_{up}) = (\hat{\theta}_{\alpha_1}^*, \hat{\theta}_{\alpha_2}^*)$$

¹¹Efron, B. and Tibshirani, R.J. (1993) An introduction to the bootstrap. Chapman & Hall.

Intervalos de confianza 6

donde

$$\alpha_1 = \Phi \left(z_0 + \frac{z_0 + z_\alpha}{1 - a(z_0 + z_\alpha)} \right)$$
$$\alpha_2 = \Phi \left(z_0 + \frac{z_0 + z_{1-\alpha}}{1 - a(z_0 + z_{1-\alpha})} \right)$$

note que si $a = 0$ y $z_0 = 0$ entonces $\alpha_1 = \alpha$ y $\alpha_2 = 1 - \alpha$, lo que nos daría el método de percentiles. Ahora, z_0 es un cuantil asociado a la proporción de réplicas bootstrap menores que $\hat{\theta}$ (el estimador original):

$$z_0 = \Phi^{-1} \left(\frac{\#\{\hat{\theta}_b^* < \hat{\theta}\}}{B} \right)$$

z_0 es una medida del sesgo de $\hat{\theta}^*$, esto es, la discrepancia entre la mediana de $\hat{\theta}^*$ y $\hat{\theta}$ (en unidades normales). Si $z_0 = 0$ entonces la mitad de los $\hat{\theta}^*$'s está por abajo de $\hat{\theta}$.

Intervalos de confianza 7

Ahora, acerca del parámetro a , su definición es en términos de los valores Jackknife del estimador $\hat{\theta}$. Sean

$\hat{\theta}_{(i)}$ = valor de $\hat{\theta}$ eliminando la i -ésima observación

$\hat{\theta}_{(.)}$ = promedio de $\hat{\theta}_i$'s (sobre los n datos)

entonces a se define como

$$a = \frac{\sum_{i=1}^n (\hat{\theta}_{(.)} - \hat{\theta}_{(i)})^3}{6 \left\{ \sum_{i=1}^n (\hat{\theta}_{(.)} - \hat{\theta}_{(i)})^2 \right\}^{3/2}}$$

a es llamado “parámetro de aceleración”, es una aproximación a la tasa de cambio del error estándar de $\hat{\theta}$ con respecto al valor verdadero de θ (\dots ?) ver Efron (1987)¹².

¹²Efron, B. (1987). Better bootstrap confidence intervals. (with discussion). JASA, 82, 171-200

Intervalos de confianza 8

```
# Intervalos de confianza BCa
set.seed(9320)
lsat  = c(576,635,558,578,666,580,555,661,651,605,653,575,545,572,594)
gpa   = c(339,330,281,303,344,307,300,343,336,313,312,274,276,288,296)
n     = length(gpa)
robs  = cor(lsat,gpa) # correlación observada = 0.7763745
B     = 5000;      b     = rep(0,B)
for(i in 1:B){
  sel  = sample(1:n,size=n,replace=TRUE)
  b[i] = cor(lsat[sel],gpa[sel]) }
z0    = qnorm(mean(b < robs))
rjack = rep(0,n)
for(i in 1:n){ rjack[i] = cor(lsat[-i],gpa[-i]) }
rm    = mean(rjack)
a     = sum( (rm-rjack)^3 ) / (6*( sum( (rm-rjack)^2 ) )^(1.5))
aen2  = .05
zal   = qnorm(aen2);      zaln  = qnorm(1-aen2)
alf1  = pnorm(z0 + (z0+zal)/(1-a*(z0+zal)))
alf2  = pnorm(z0 + (z0+zaln)/(1-a*(z0+zaln)))
inter = quantile(b,probs=c(alf1,alf2))
#      14.4% 99.999%
#      0.6158559 0.9975147
```

Estimación bootstrap del sesgo 1

- Supongamos $\hat{\theta}$ estimador de θ . El sesgo de $\hat{\theta}$ se define como $b = E(\hat{\theta} - \theta)$. El estimador Bootstrap del sesgo es $B^* = E^*(\hat{\theta}^* - \hat{\theta})$, o, usando simulación

$$\hat{B} = \frac{1}{R} \sum_{i=1}^R (\hat{\theta}_i^* - \hat{\theta})$$

- Más sencillo no puede ser, de hecho, esta sencillez de los métodos bootstrap hicieron que fueran inicialmente recibido con escepticismo:

“... because to use the data to generate more data seems analogous to a trick used by the fictional Baron Munchausen, who when he found himself at the bottom of a lake got out by pulling himself up by his bootstraps. In the simplest nonparametric problems we do literally sample from the data, and a common initial reaction is that this is a fraud. In fact it is not. It turns out that a wide range of statistical problems can be tackled this way, liberating the investigator from the need to oversimplify complex problems” (Davison y Hinkley).

Estimación bootstrap del sesgo 2

- Ejemplo: Datos de un estudio sobre Bioequivalencia (tomado de Efron y Tibshirani (1993)).
- 8 participantes del estudio usaron parches medicinales que liberan cierta hormona en el torrente sanguíneo (la hormona es producida en forma natural en el cuerpo humano y, se supone, los parches subsanan deficiencias en su producción natural).
- Cada individuo probó cada uno de tres parches bajo estudio: Placebo, Nuevo, Viejo. Los nuevos y viejos se refieren al mismo tipo de parche pero manufacturado en una planta nueva y una vieja. En USA la agencia FDA regula la fabricación de medicamentos. Los parches nuevos tienen que ser aprobados por la FDA para su venta, pero no se requiere que cumplan todo el proceso de aprobación, sólo se requiere que muestren bioequivalencia con los antiguos. El criterio de aprobación es que

$$\frac{|E(\text{nuevo}) - E(\text{viejo})|}{E(\text{viejo}) - E(\text{placebo})} \leq 0.20$$

- Los datos del estudio son

	Individuo							
parche	1	2	3	4	5	6	7	8
placebo	9243	9671	11792	13357	9055	6290	12412	18806
old	17649	12013	19979	21816	13850	9806	17208	29044
new	16449	14614	17274	23798	12560	10157	16570	26325

Estimación bootstrap del sesgo 3

- El parámetro de interés es

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

- De los datos se obtiene que el plug-in es $\hat{\theta} = -0.0713$. La estimación del sesgo es

$$\hat{B} = \frac{1}{R} \sum_{i=1}^R \hat{\theta}_i^* - \hat{\theta} = -0.0625 - (-0.0713) = 0.0088$$

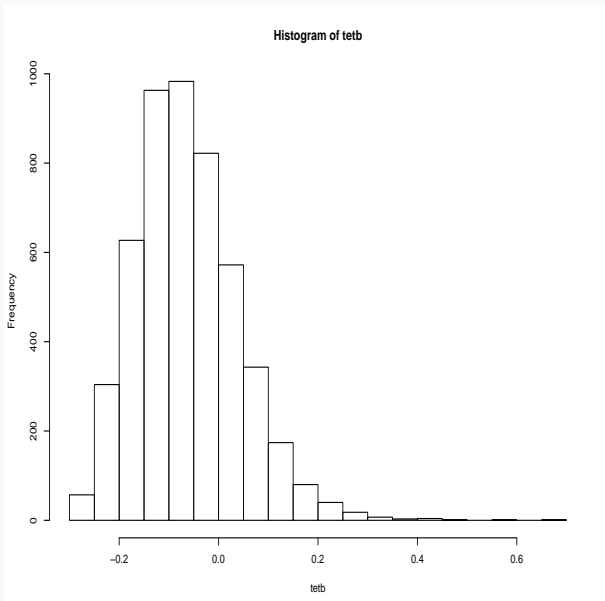
un sesgo bastante pequeño (sesgo/error est = 0.085, como regla general, es ok con que sea menor a 0.25).

- Un intervalo de confianza para θ arroja $(-0.234, 0.167)$, adentro del 0.20 requerido.

Estimación bootstrap del sesgo 4

```
# Estimación del Sesgo (tomado de Efron y Tibshirani p.127)
placebo = c(9243,9671,11792,13357,9055,6290,12412,18806)
old     = c(17649,12013,19979,21816,13850,9806,17208,29044)
new     = c(16449,14614,17274,23798,12560,10157,16570,26325)
z       = old-placebo
y       = new-old
tetg    = mean(y)/mean(z)
n       = length(z); B       = 5000; tetb = rep(0,B)
set.seed(828)
for(i in 1:B){
  sel     = sample(1:n, size=n,replace=TRUE)
  tetb[i] = mean(y[sel])/mean(z[sel]) }
mean(tetb)
sd(tetb)
sesgo = mean(tetb)-tetg
sesgo/sd(tetb)
hist(tetb)
round( quantile(tetb,probs=seq(.9,.99,by=.01)), 4)
#      90%      91%      92%      93%      94%      95%      96%      97%      98%      99%
# 0.0707 0.0794 0.0870 0.0955 0.1053 0.1161 0.1316 0.1533 0.1773 0.2267
quantile(tetb,probs=c(.025,.975))
#           2.5%           97.5%
# -0.2342787  0.1673464
```


Estimación bootstrap del sesgo 5



Bootstrap y Big Data

Bootstrap con altos volúmenes de datos 1

Una observación inmediata acerca del Bootstrap es que puede ser computacionalmente demandante. Si tenemos n grande y el procedimiento base es costoso, entonces Bootstrap implica B veces ese procedimiento costoso basado en n muestras completas. Algunas alternativas que se han propuesto en la literatura, para aminorar el problema:

- Bootstrap m -de- n ¹³, el cual requiere muestrear sólo $m < n$ puntos.
- Otra alternativa es usar Submuestreo¹⁴

Sin embargo, el primero, requiere del conocimiento explícito de las tasas de convergencia del procedimiento base; y, aparentemente, el segundo procedimiento es no robusto a la especificación de hiperparámetros.

¹³Bickel, P.J., Gotze, F., and van Zwet, W. (1997) Resampling fewer than n observations: Gains, losses, and remedies for losses. *Statistica Sinica*, 7, 1-31.

¹⁴Politis, D., Romano, J., and Wolf, M. (1999) Subsampling. Springer.

Bootstrap con altos volúmenes de datos 2

Una tercera opción es llamada BLB¹⁵ (Bag of Little Bootstraps) y consiste en hacer múltiples bootstraps en subconjuntos de datos y luego promediarlos. Supongamos datos $\{x_1, x_2, \dots, x_n\}$ y particionemoslos en s grupos de tamaño b (en realidad, no tienen que formar una partición, incluso pueden formarse al azar sin tener que ser disjuntos, pero para fines de exposición, supongamos que $n = bs$ y los n datos se pueden separar en s grupos), supongamos que queremos el error estándar de cierta estimación $\hat{\theta}$.

Subconjunto	Datos	Empírica	Objetivo
1	$x_1^1, x_2^1, \dots, x_b^1$	\hat{F}_b^1	$\text{errstd}(\hat{\theta})^1$
2	$x_1^2, x_2^2, \dots, x_b^2$	\hat{F}_b^2	$\text{errstd}(\hat{\theta})^2$
\vdots	\vdots	\vdots	\vdots
s	$x_1^s, x_2^s, \dots, x_b^s$	\hat{F}_b^s	$\text{errstd}(\hat{\theta})^s$

¹⁵Kleiner, A., Talwalkar, A., Sarkar, P., and Jordan, M.I. (2012). The big data bootstrap. Proceedings of the 29th International Conference on Machine Learning, Edinburgh, Scotland, UK

Bootstrap con altos volúmenes de datos 3

La forma de calcular las cantidades objetivo de interés para cada subconjunto, es vía simulación bootstrap, pero de la siguiente forma: Consideremos los datos del subconjunto j : $x_1^j, x_2^j, \dots, x_b^j$ muestreamos de la empírica \hat{F}_b^j B muestras de tamaño n (note $n > b$). Este muestreo es equivalente a generar un vector aleatorio multinomial

$$(n_1, n_2, \dots, n_b) \sim \text{Multinomial}(n; 1/b, 1/b, \dots, 1/b)$$

luego estimamos, para cada una de las B muestras, el parámetro de interés, $\hat{\theta}$, reponderando cada vez las muestras

$$\frac{n_1}{n} x_1^j, \frac{n_2}{n} x_b^j, \dots, \frac{n_b}{n} x_b^j$$

(casi todos los procedimientos en Estadística pueden considerar datos ponderados). Con estas B estimaciones, estimamos el error estándar, $\text{errstd}(\hat{\theta})^j$.

Bootstrap con altos volúmenes de datos 4

Finalmente, promediamos

$$\text{errstd}_{BLB}(\hat{\theta}) = \frac{1}{s} \sum_{j=1}^s \text{errstd}(\hat{\theta})^j$$

Aquí hemos usado el error estándar para ejemplificar el procedimiento, pero puede ser cualquier otra medida de la calidad de un estimador. Resumiendo, BLB requiere sólo cálculos repetidos en subconjuntos de tamaño pequeño, en vez del bootstrap original que requiere cálculos repetidos con conjuntos grandes de datos. Los autores del BLB muestran consistencia del procedimiento, en un artículo paralelo¹⁶.

¹⁶Kleiner, A., Talwalkar, A., Sarkar, P., and Jordan, M.I (2012). A scalable bootstrap for massive data, <http://arxiv.org/abs/1112.5016>

Inferencia bayesiana y el bootstrap 1

- Consideremos el bootstrap paramétrico, ilustrado con el ejemplo del cálculo del error estándar del coeficiente de correlación de Pearson.
- Tenemos $(x_1, y_1), \dots, (x_n, y_n)$, i.i.d. $N_2(\mu, \Sigma)$. El coeficiente de Pearson es

$$r = \frac{[\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})]^{1/2}}{[\sum_{i=1}^n (x_i - \bar{x})^2]^{1/2} [\sum_{i=1}^n (y_i - \bar{y})^2]^{1/2}}$$

- La versión paramétrica del bootstrap usa la distribución $N_2(\hat{\mu}, \hat{\Sigma})$.
- Supongamos ahora una previa para ρ , $\pi(\rho)$, y la posterior

$$\pi(\rho|r) \propto \pi(\rho)f(r|\rho)$$

donde $f(r|\rho)$ es la densidad de r (hoja Bootstrap básico 7), parametrizada con el MLE $\hat{\rho} = 0.776$ (curva sólida en hoja Bootstrap paramétrico 4).

Inferencia bayesiana y el bootstrap 2

- ¿Cómo hacer cálculos con la posterior?. Sea $A \subset [-1, 1]$ un subconjunto del espacio paramétrico para ρ .

$$P(\rho \in A|r) = \int_A \pi(\rho|r) d\rho = \frac{\int_A \pi(\rho) f(r|\rho) d\rho}{\int_{-1}^1 \pi(\rho) f(r|\rho) d\rho}$$

- Sea $R(\rho)$ el cociente de la verosimilitud y la densidad bootstrap (de la cual podemos muestrear)

$$R(\rho) = \frac{f(r|\rho)}{f(\rho|r)}$$

entonces

$$P(\rho \in A|r) = \frac{\int_A \pi(\rho) R(\rho) f(\rho|r) d\rho}{\int_{-1}^1 \pi(\rho) R(\rho) f(\rho|r) d\rho}$$

Inferencia bayesiana y el bootstrap 3

- Más general, si $t(\rho)$ es cualquier función (razonable) de ρ , su esperanza posterior es

$$E(t(\rho)|r) = \frac{\int_{-1}^1 t(\rho)\pi(\rho)R(\rho)f(\rho|r)d\rho}{\int_{-1}^1 \pi(\rho)R(\rho)f(\rho|r)d\rho}$$

- Ahora, como r_1^*, \dots, r_B^* son i.i.d. de $f(\rho|r)$, obtenemos un estimador de esa esperanza (muestreo por importancia)

$$\hat{E}(t(\rho|r)) = \frac{\sum_{i=1}^B t_i \pi_i R_i}{\sum_{i=1}^B \pi_i R_i}$$

- En otras palabras, bootstrap nos puede ayudar a tener estimaciones de cantidades de interés para inferencia bayesiana.

Inferencia bayesiana y el bootstrap 4

- Las láminas anteriores indican una liga entre el bootstrap paramétrico y bayesiana.
- De las primeras conexiones publicadas está: Rubin, D.B. (1981). The bayesian bootstrap. *The Annals of Statistics*, 9, 1, 130-134. Aquí se presenta la conexión del bootstrap noparamétrico con bayesiana.
- Un artículo, probablemente más influyente (aunque con la desgracia de aparecer en pleno boom MCMC), es: Newton, M.A. & Raftery, A.E. (1994). Approximate bayesian inference with the weighted likelihood bootstrap. *JRSS-B*, 56, 1, 3-48.

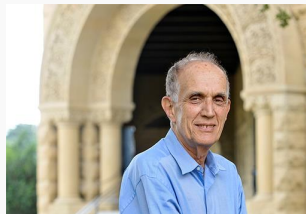
El bootstrap y otras técnicas

- Bagging
- Random forests
- Boosting
- Ajuste de p -valores en pruebas múltiples
- Evaluación de incertidumbre en predicciones de NN's
- Problemas de calibración en Química analítica

Cox 2017 & Efron 2019



International Prize in Statistics Awarded to Stanford's Bradley Efron



Acerca del nombre ...¹⁷

Acknowledgments. I am grateful to Professors Rupert Miller and David Hinkley for numerous discussions, suggestions and references, and to Joseph Verducci for help with the numerical computations. The referees contributed several helpful ideas, especially concerning the connection with Hartigan's work, and the large sample theory. I also wish to thank the many friends who suggested names more colorful than *Bootstrap*, including *Swiss Army Knife*, *Meat Axe*, *Swan-Dive*, *Jack-Rabbit*, and my personal favorite, the *Shotgun*, which, to paraphrase Tukey, "can blow the head off any problem if the statistician can stand the resulting mess."

¹⁷Efron, B. (1979) Bootstrap methods: another look at the jackknife, *Ann. Statist.*, 7, 1, pp. 1-26.

Ejemplos

Regresión noparamétrica 1

```
# Regresión Noparamétrica
# file = kidney Efron Hastie.txt
dat = read.table( file=file.choose(), header=TRUE) # 157 x 2
attach(dat)

plot(age,tot,ylim=c(-7,7),type="n", main="Intervalos Bootstrap")
set.seed(84848)
n = dim(dat)[1]
ed = seq(20,85,by=5)
m = length(ed)
B = 1000
pred = matrix(0,B,m)
for(i in 1:B){
  sel = sample(1:n,size=n,replace=TRUE)
  eda = age[sort(sel)]
  tt = tot[sort(sel)]
  bb = loess(tt~eda)
  pp = predict(bb)
  pred[i,] = predict(bb,data.frame(eda=ed))
  lines(age, pp, col = gray(.8)) }
```

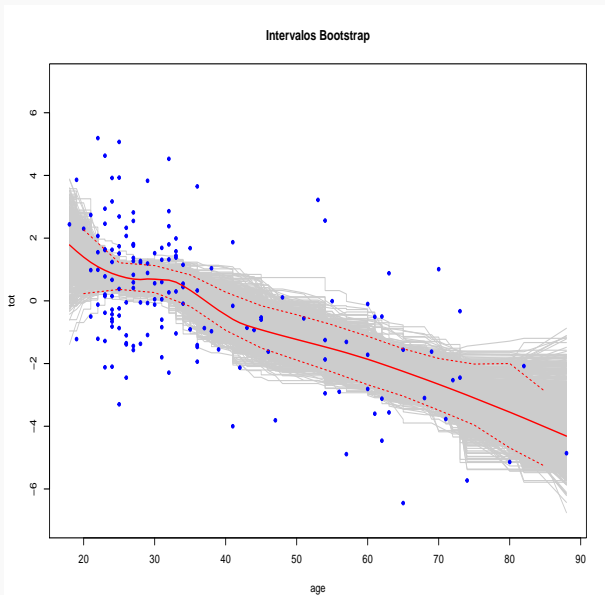
Regresión no paramétrica 2

```
points(age,tot,pch=20,col="blue")
bb = loess(tot~age)
pp = predict(bb)
pp2 = predict(bb,data.frame(age=ed))
lines(age,pp, col = 2, lwd=2)

q025 = function(x){quantile(x,p=.025,na.rm=TRUE)}
q975 = function(x){quantile(x,p=.975,na.rm=TRUE)}
liminf = apply(pred,2,q025)
limsup = apply(pred,2,q975)

lines(ed,liminf, col = 2, lty=2)
lines(ed,limsup, col = 2, lty=2)
```


Regresión noparamétrica 3



Regresión 1

```
# Regresión
# Los datos del archivo EmisionesNOX.csv corresponden
# a un estudio sobre emisiones con camionetas diesel.
# Los nombres de las variables son: NITROX, HUMEDAD, TEMP, PRESION.

datos <- read.csv( "c:\\...\\EmisionesNOX.csv", header=TRUE) # 44 x 4

histo <- function(x, ...){
  usr <- par("usr")
  on.exit(par(usr))
  par( usr=c(usr[1:2],0,1.5) )
  h      <- hist(x,plot=FALSE)
  cortes <- h$breaks
  nc     <- length(cortes)
  y      <- h$counts
  y      <- y/max(y)
  rect(cortes[-nc], 0, cortes[-1], y, col="cyan")}

pairs(datos, diag.panel=histo, cex.labels=.9, panel=panel.smooth)

attach(datos)
out <- lm(NITROX ~ HUMEDAD+TEMP+PRESION)
summary(out)
```

Regresión 2

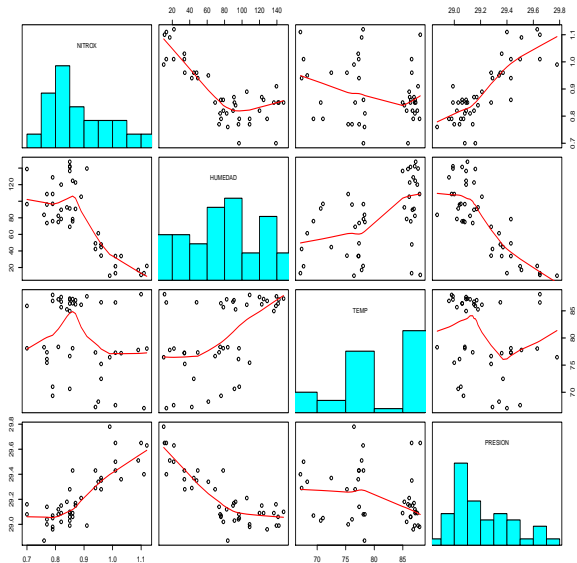
```
#               Estimate Std. Error t value Pr(>|t|)
# (Intercept) -7.3362273   2.3048269  -3.183   0.00282 **
# HUMEDAD     -0.0008559   0.0004614  -1.855   0.07096 .
# TEMP         0.0012073   0.0016698   0.723   0.47388
# PRESION      0.2803706   0.0792100   3.540   0.00103 **

# Bootstrapeando residuales
set.seed(82887)
res = out$residuals
yg  = out$fitted.values
n   = length(res)
p   = 4
B   = 5000
bet = matrix(0,B,p)
for(i in 1:B){
  sel = sample(1:n,size=n,replace=TRUE)
  reb = res[sel]
  yb  = yg + reb
  oub = lm(yb ~ HUMEDAD+TEMP+PRESION)
  bet[i,] = oub$coeff }
```

Regresión 3

```
sqrt(diag(var(bet)))  
# 2.2205360647 0.0004435263 0.0015919215 0.0763126055  
# Bootstrapeando observaciones  
set.seed(828)  
n = length(NITROX)  
p = 4  
B = 5000  
bet = matrix(0,B,p)  
for(i in 1:B){  
  sel = sample(1:n,size=n,replace=TRUE)  
  db = datos[sel,]  
  oub = lm(NITROX ~ HUMEDAD+TEMP+PRESION, data=db)  
  bet[i,] = oub$coeff }  
  
sqrt(diag(var(bet)))  
# 2.5110654926 0.0004857757 0.0012690729 0.0856142622
```

Regresión 4



Poisson 1

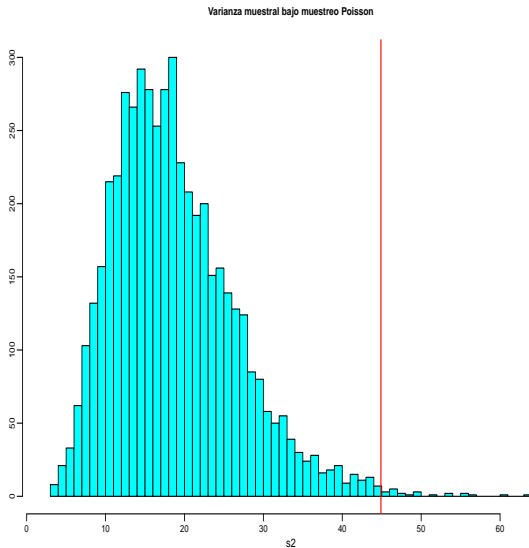
```
# Bootstrap paramétrico bajo el modelo Poisson
# (Tomado de Casella y Berger, p.250)

larvas = c(19,32,29,13,8,12,16,20,14,17,22,18,23)
xbar    = mean(larvas)    # 18.7
s2      = var(larvas)     # 44.9
n       = length(larvas)
# Distribución de  $S^2$ 
set.seed(74747)
B       = 5000
s2b     = rep(0,B)
for(i in 1:B){ s2b[i] = var( rpois(n,xbar) ) }

hist(s2b,main="Varianza muestral bajo muestreo Poisson",
     cex.main=.9,xlab="s2",ylab="",cex.axis=.8,mgp=c(1.5,.5,0),
     col="cyan",nclass=50,xlim=c(2,65) ) # ,
     probability=TRUE,ylim=c(0,4.5))
abline(v=s2,lwd=2,col="red")

1 - sum( s2b <= s2 ) / B # .0044 proporción de realizaciones de  $s^2$  que
                        # sobrepasan el valor observado de la varianza
```

Poisson 2



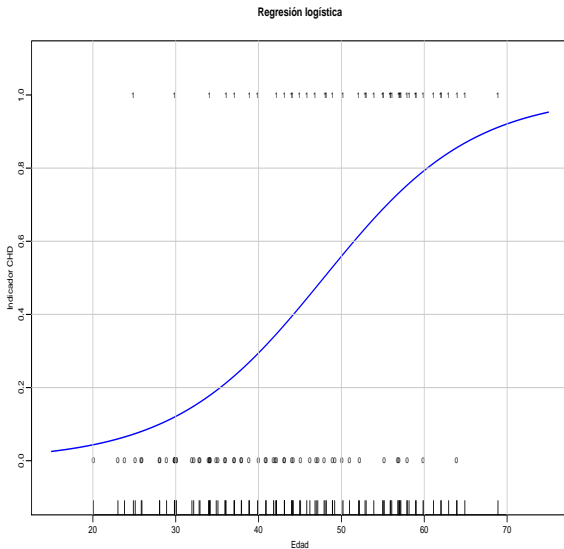
ROC 1

```
# Ejemplo Curvas ROC
# Hosmer, D.W. & Lemeshow, S.(1989) Applied logistic regression. Wiley
# Edad y Coronaria (daño significativo en coronaria)
edad = c(
  20, 23, 24, 25, 25, 26, 26, 28, 28, 29, 30, 30, 30, 30, 30, 30, 32,
  32, 33, 33, 34, 34, 34, 34, 34, 35, 35, 36, 36, 36, 37, 37, 37, 38,
  38, 39, 39, 40, 40, 41, 41, 42, 42, 42, 42, 43, 43, 43, 44, 44, 44,
  44, 45, 45, 46, 46, 47, 47, 47, 48, 48, 48, 49, 49, 49, 50, 50, 51,
  52, 52, 53, 53, 54, 55, 55, 55, 56, 56, 56, 57, 57, 57, 57, 57, 57,
  58, 58, 58, 59, 59, 60, 60, 61, 62, 62, 63, 64, 64, 65, 69)
coro = c(
  0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,1,0,0,0,0,0,1,0,0,1,0,0,
  0,0,1,0,1,0,0,0,0,0,1,0,0,1,0,0,1,1,0,1,0,1,0,0,1,0,1,1,0,0,1,0,1,0,
  0,1,1,1,1,0,1,1,1,1,1,0,0,1,1,1,1,0,1,1,1,1,0,1,1,1,1,1,0,1,1,1)
n = length(coro)
# Gráfica de los datos
edadj = jitter(edad)      # solo con fines de graficación
plot(edadj,coro,xlab="Edad",ylab="Indicador CHD",ylim=c(-.1,1.1),
     mgp=c(1.5,.5,0),cex.axis=.8,cex.lab=.8,cex.main=1,xlim=c(15,75),
     cex=.7, main="Regresión logística",pch=ifelse(coro==1,"1","0"))
rug(edadj)
out = glm(coro ~ edad, family=binomial)
b = out$coeff
```


ROC 2

```
# Agregamos curva logística a la gráfica original
xx = seq(15,75,length=200)
X = cbind(rep(1,n),xx)
p = 1/( 1+exp( -as.vector(X%*%b) ) )
lines(xx,p,lwd=2,col="blue")
abline(h=seq(.2,.8,by=.2),v=seq(20,70,by=10),col=gray(.8))
```

ROC 3



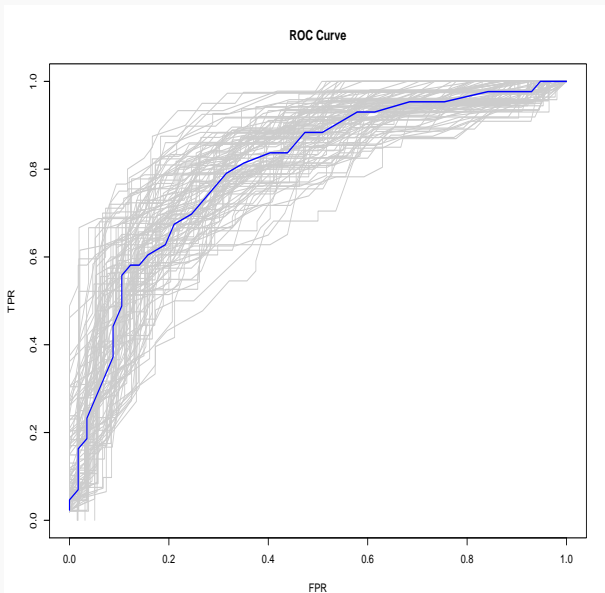
ROC 4

```
### CURVA ROC Bootstrap
set.seed(828298)
plot(0,0,xlim=c(0,1),ylim=c(0,1),xlab="FPR",ylab="TPR",type="n",
     main="ROC Curve")
B      = 101
n      = length(coro)
M      = 101; ROC = matrix(0,M,2)
for(b in 1:B){
  sel   = sample(1:n,size=n,replace=TRUE)
  out   = glm(coro[sel] ~ edad[sel], family=binomial)
  prob  = out$fitted.values
  umbral = seq(min(prob),max(prob),length=M)
  for(i in 1:M){
    yg    = ifelse(prob >= umbral[i], 1, 0)
    yg    = factor(yg,levels=c(0,1))
    aa    = table(coro[sel],yg)
    TN    = aa[1,1]
    FP    = aa[1,2]
    FN    = aa[2,1]
    TP    = aa[2,2]
    ROC[i,1] = FP/(FP+TN)
    ROC[i,2] = TP/(FN+TP) }
  lines(ROC[,1],ROC[,2],type="l",col=gray(.8)) }
```

ROC 5

```
out = glm(coro ~ edad, family=binomial)
prob = out$fitted.values
M = 101
umbral = seq(min(prob),max(prob),length=M)
ROC = matrix(0,M,2)
for(i in 1:M){
  yg = ifelse(prob >= umbral[i], 1, 0)
  yg = factor(yg,levels=c(0,1))
  aa = table(coro,yg)
  TN = aa[1,1]
  FP = aa[1,2]
  FN = aa[2,1]
  TP = aa[2,2]
  ROC[i,1] = FP/(FP+TN)
  ROC[i,2] = TP/(FN+TP) }
lines(ROC[,1],ROC[,2],type="l",xlab="FPR",ylab="TPR",lwd=2,col="blue")
```

ROC 6



La mediana 1

```
# Ejemplo Bootstrap de la Mediana
# mediana de LSAT
set.seed(8484)
mobs  = median(lsat)
n      = length(lsat)
B      = 5000
b      = rep(0,B)
for(i in 1:B){
  sel  = sample(1:n,size=n,replace=T)
  b[i] = median(lsat[sel]) }
s = sd(b)
# Estimación bootstrap del error estándar
# s = 19.08892

hist(b,main="Réplicas Bootstrap de la Mediana",
     cex.main=.9,xlab="m(*)",ylab="",cex.axis=.8,mgp=c(1.5,.5,0),
     col="cyan",nclass=20)
points(mobs,100,pch=16,col="red")
```

La mediana 2

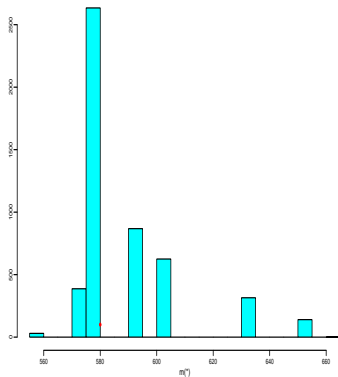
```
# Bootstrap suavizado
set.seed(87641)
h      = 10
for(i in 1:B){
  sel  = sample(1:n,size=n,replace=T)
  pert = runif(n,min=-h,max=h)
  b[i] = median(lsat[sel]+pert) }
s <- sd(b)      # 18.3366

hist(b,main="Réplicas Suavizadas, Bootstrap de la Mediana",
     cex.main=.9,xlab="m(*)",ylab="",cex.axis=.8,mgp=c(1.5,.5,0),
     col="cyan",nclass=20)
points(mobs,100,pch=16,col="red")

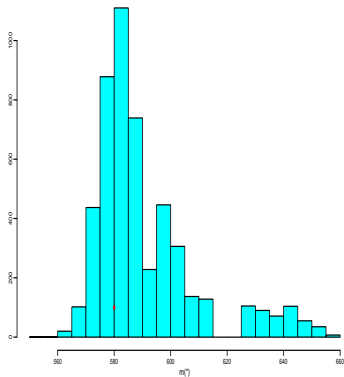
# SE(median) = 1.2533 x SE( x_bar ) = sqrt(pi/2) * se( X_bar)
sqrt(pi/2) * sd(lsat) / sqrt(n) # 13.525 (bajo normalidad)
```

La mediana 3

Réplicas Bootstrap de la Mediana



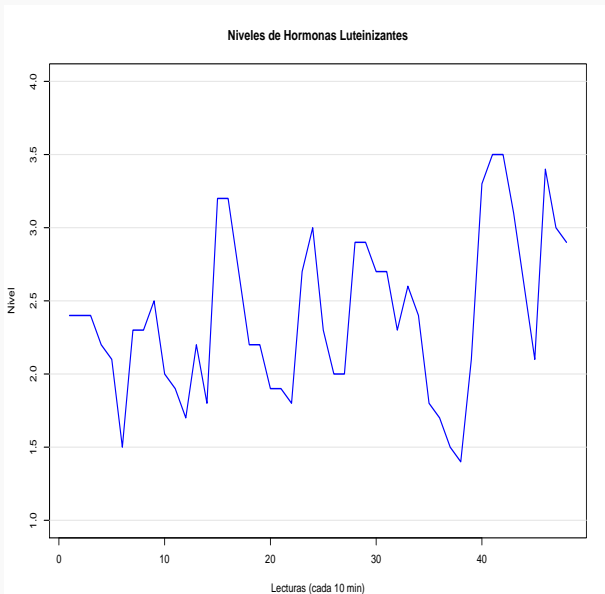
Réplicas Suavizadas, Bootstrap de la Mediana



Series de tiempo 1

```
# Bootstrapping Time Series
library(bootstrap)
data(lutenhorm)
dat = lutenhorm[,4]
n = length(dat)
plot(1:n,dat,ylim=c(1,4),type="n",col="blue",lwd=2,ylab="Nivel",
     main="Niveles de Hormonas Luteinizantes",xlab="Lecturas (cada 10 min)")
abline(h=seq(1,4,by=.5),col=gray(.9))
lines(1:n,dat,col="blue",lwd=2)
out = ar(dat,order=1)
out$ar           # 0.5755245
sqrt(out$asy.var.coef) # 0.1205758
set.seed(7654)
B = 300
xbs = rep(0,B)
for(i in 1:B){
  sel = sample(1:(n-2),size=12,replace=T)
  xx = NULL
  for(j in 1:12){ xx = c(xx,dat[(sel[j]:(sel[j]+2))]) }
  out = ar(xx,aic=FALSE,order.max=1)
  xbs[i] = out$ar }
sd(xbs) # 0.1405133
```

Series de tiempo 2



Smooth bootstrap 1

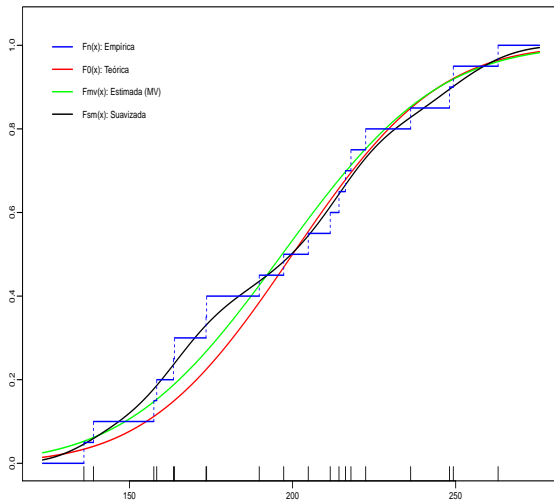
```
# Smooth Bootstrap: Distribuciones Empírica, Teórica,  
# Suavizada y Estimada (MV)  
  
n      = 20  
med    = 200  
des    = 35  
delt   = .10  
edf    = (1:n)/n  
set.seed(5963471)  
dat    = sort(rnorm(n,mean=med,sd=des))  
xb     = mean(dat)  
ss     = sd(dat)  
ran    = max(dat) - min(dat)  
lmin   = min(dat) - delt*ran  
lmax   = max(dat) + delt*ran  
xx     = seq(lmin,lmax,length=M)  
yy     = pnorm(xx,mean=med,sd=des)  
yyE    = pnorm(xx,mean=xb,sd=ss)  
aa     = rep(0,n)  
for(i in 1:n){ aa  = aa + pnorm(xx,mean=dat[i],sd=10) }  
ysm    = aa/n
```

Smooth bootstrap 2

```
plot(dat,edf, xlim=c(lmin,lmax), ylim=c(0,1.05), mgp=c(1.5,.5,0),
     xlab="", ylab="", cex.axis=.8, cex.lab=.8,type="n",
     main="Funciones de Distribución",cex.main=1)
rug(dat)
lines(xx,yy,col="red",lwd=2)
lines(xx,yyE,col="green",lwd=2)
lines(xx,ysm,col="black",lwd=2)
segments(lmin,0,dat[1],0,col="blue",lwd=2)
for(i in 1:(n-1)){ segments(dat[i],edf[i],dat[i+1],edf[i],
                           col="blue",lwd=2) }
segments(dat[n],1,lmax,1,col="blue",lwd=2)
segments(dat[1],0,dat[1],edf[1],col="blue",lty=2)
for(i in 2:n){
  segments(dat[i],edf[i-1],dat[i],edf[i],col="blue",lty=2) }
legend(lmin,1.05,legend=c("Fn(x): Empírica","F0(x): Teórica",
  "Fmv(x): Estimada (MV)","Fsm(x): Suavizada"),
      col=c("blue","red","green","black"),lwd=2, cex=.8,bty="n")
```

Smooth bootstrap 3

Funciones de Distribución



Tasa de momios

```
# Tasa de Momios

set.seed(438844)
n11 = 834
n12 = 170
n21 = 760
n22 = 222
np1 = n11+n21
np2 = n12+n22
B = 5000
p1 = n11/np1
p2 = n12/np2
s1 = rbinom(B,size=np1,prob=p1)
s2 = rbinom(B,size=np2,prob=p2)
odds = s1*(np2-s2)/(s2*(np1-s1))
mean(odds) # 1.439826
mean(log(odds)) # 0.358274
sqrt(var(log(odds))) # 0.1117214
sqrt(1/n11 + 1/n12 + 1/n21 + 1/n22) # 0.1135856
```

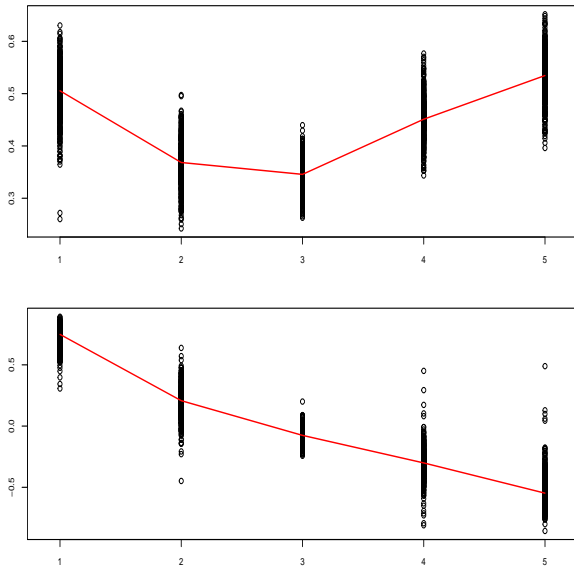
Componentes principales 1

```
# Componentes Principales
# Datos de calificaciones de exámenes a libro cerrado/abierto
library(bootstrap)
data(scor)
dim(scor) # 88 x 5 (ver Mardia, p.4)
out = princomp(scor)
lo = out$loadings
cp1 = abs(lo[,1])
cp2 = lo[,2]
n = dim(scor)[1]
B = 500
c1 = matrix(0,B,5)
c2 = matrix(0,B,5)
for(i in 1:B){
  sel = sample(1:n,size=n,replace=T)
  out = princomp(scor[sel,])
  aa = out$loadings
  c1[i,] = aa[,1]
  c2[i,] = aa[,2] }
```

Componentes principales 2

```
par(mfrow=c(2,1),mar=c(4,4,1,1))
c1 = abs(c1)
yl = range(c1)
plot(rep(1,B),c1[,1],xlim=c(0.9,5.1),ylim=yl,xlab="",ylab="",
      cex.axis=.8)
for(j in 2:5){
  points(rep(j,B),c1[,j])}
lines(1:5,cp1,lwd=2,col="red")
ss = sign(c2[,1])
c2 = ss*c2
yl = range(c2)
plot(rep(1,B),c2[,1],xlim=c(0.9,5.1),ylim=yl,xlab="",ylab="",
      cex.axis=.8)
for(j in 2:5){
  points(rep(j,B),c2[,j])}
lines(1:5,cp2,lwd=2,col="red") }
```


Componentes principales 3



Tarea (1)

1. Los siguientes 15 datos forman una muestra aleatoria de una distribución Gama con parámetro de forma $\alpha = 3$ y parámetro de escala $\beta = 2$ (la media es $\alpha\beta$ y la varianza $\alpha\beta^2$)

14.18 10.99 3.38 6.76 5.56 1.26 4.05 4.61
1.78 3.84 4.69 2.12 2.39 16.75 4.19

Encuentre un intervalo de confianza para la media de la distribución.

2. En la librería `boot` de R acceder los datos `cd4`, los cuales son conteos de células CD4 en pacientes HIV-positivos al inicio y después de 1 año de tratamiento con un antiviral.
 - a. Construya un intervalo bootstrap para el coeficiente de correlación entre los conteos base y los conteos después del tratamiento.
 - b. Calcule el coeficiente estimado de correlación corregido por sesgo usando el jackknife.

Tarea (2)

3. Sea $\hat{F}(x)$ la función de distribución empírica. Muestre que $E[\hat{F}(x)] = F(x)$, $\text{Var}[\hat{F}(x)] = F(x)[1 - F(x)]$ y comente porque $\hat{F}(x)$ puede considerarse como normalmente distribuida.
4. ¿Cuántas simulaciones se requieren para estimar cuantiles? Para darse una idea ejecute el siguiente programa en R varias veces:

```
par(mfrow=c(4,4),mar=c(2,2,2,2))
cuantiles <- matrix(0,16,4)
n <- c(40,100,400,1000)
p <- c(0.025,0.05,0.95,0.975)
for(i in 1:4){
  y <- rnorm(1000)
  for(j in 1:4){
    cuantiles[(j-1)*4+i,] <- quantile(y[1:n[j]],probs=p)
    qqnorm(y[1:n[j]],main=paste("B = ",n[j]),cex.main=.9)
    abline(h=quantile(y[1:n[j]],p),lty=2,col="red")  } }
```

¿Cuántas simulaciones se requieren para obtener estimaciones razonables de los cuantiles del 5% y 95%?, ¿para los cuantiles del 2.5% y 97.5%?

Tarea (3)

5. Considere la matriz de calificaciones de 88 estudiantes sobre 5 materias (`data(scor)` en el paquete `bootstrap` de R). Sea S la matrix de varianzas y covarianzas correspondiente y sea $\hat{\lambda}_1$ el máximo valor propio de S . Si Σ es la matriz de varianza poblacional, es claro que $\hat{\lambda}_1$ estima a λ_1 , la máxima raíz de Σ .
- a. Use bootstrap para estimar el error estándar de $\hat{\lambda}_1$ y construya un intervalo de confianza para λ_1 .
 - b. Sea $F(x; \theta)$ cierta distribución. Sea $\hat{\theta}$ un estimador de θ . El sesgo de $\hat{\theta}$ es $b = E(\hat{\theta}) - \theta$, lo cual puede escribirse como $b = E(\hat{\theta} | F) - t(F)$. Podemos usar bootstrap para estimar este sesgo de la siguiente forma:

$$\tilde{b} = E(\hat{\theta} | \hat{F}) - t(\hat{F}) \approx \frac{1}{B} \sum_{i=1}^B \hat{\theta}^* - \hat{\theta}$$

Use estas ideas para estimar el sesgo de $\hat{\lambda}_1$.