EL BOOTSTRAP DE MODELOS MÁS GENERALES

1. El bootstrap de un modelo de regresión. Wild bootstrap.
2. Estimación bootstrap del error de una regla de predicción. Remuestreos bootstrap uniforme y suavizado en la estimación del optimismo esperado en los contextos de análisis de regresión y análisis discriminante. Estudios de simulación.
3. Introducción a la utilización del bootstrap en poblaciones finitas.
4. Validación de la aproximación bootstrap.

El mecanismo que genera los datos puede ser un MODELO DE PROBABILIDAD DESCONOCIDO más complejo que una función de distribución.

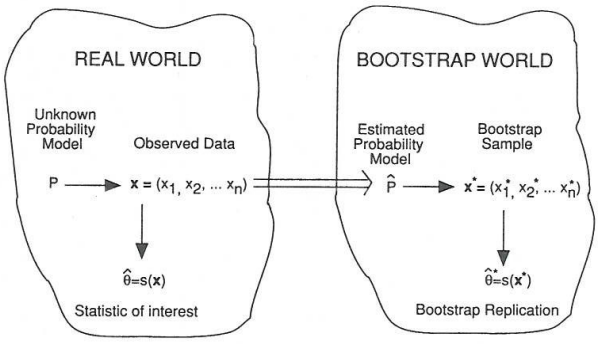


Figure 8.3. *Schematic diagram of the bootstrap applied to problems with a general data structure* $P \to \mathbf{x}$. *The crucial step* "$\Longrightarrow$" *produces an estimate* $\hat{P}$ *of the entire probability mechanism* P *from the observed data* $\mathbf{x}$. *The rest of the bootstrap picture is determined by the real world:* "$\hat{P} \to \mathbf{x}^*$" *is the same as* "$P \to \mathbf{x}$"; *the mapping from* $\mathbf{x}^* \to \hat{\theta}^*$, $s(\mathbf{x}^*)$, *is the same as the mapping from* $\mathbf{x} \to \hat{\theta}$, $s(\mathbf{x})$.

EL BOOTSTRAP DE UN MODELO DE REGRESIÓN

El bootstrap actúa sobre los datos observados imitando el mecanismo que los genera. Este mecanismo no tiene por qué ser siempre una función de distribución; puede ser una estructura más compleja, como por ejemplo un modelo de regresión.

Supongamos que los datos observados son de la forma

$$x_i = (\vec{c}_i^{\,t}, y_i),\ i = 1,\ldots,n \qquad \begin{cases} \vec{c}_i^{\,t} = (c_{i1},\ldots,c_{ip}) & \textit{parte explicativa conocida} \\ y_i & \textit{respuesta aleatoria asociada a } \vec{c}_i^{\,t} \end{cases}$$

y satisfacen el modelo de regresión lineal de diseño fijo y homocedástico

$$Y_i = \vec{c}_i^{\,t}\vec{\beta} + U_i,\ i = 1,\ldots,n \qquad \begin{cases} \vec{\beta}^{\,t} = (\beta_1,\ldots,\beta_p) & \textit{son parámetros desconocidos} \\ \vec{U}^{\,t} = (U_1,\ldots,U_n) & \textit{m.a.s. de } U\textit{, perturbación aleatoria} \\ \textit{con } F \textit{ desconocida, tal que } E_F(U) = 0; V_F(U) = \sigma^2 & \end{cases}$$

Con notación matricial:

$$\vec{Y} = \mathbf{c}\vec{\beta} + \vec{U}, \qquad \left(\vec{Y} = \begin{pmatrix} Y_1 \\ \cdots \\ Y_n \end{pmatrix}, \mathbf{c} = \begin{pmatrix} c_{11} & & c_{1p} \\ & \cdots & \\ c_{n1} & & c_{np} \end{pmatrix} \right)$$

Notemos que en este modelo, Y_i es la variable aleatoria respuesta (dependiente o endógena) y $\vec{c}_i^{\,t}$ es la variable explicativa (independiente o exógena) que suponemos conocida al observar la anterior. La componente desconocida del modelo es el par $P = (\vec{\beta}, F)$ que lo especifica, y constituye por lo tanto el mecanismo (más complejo que una función de distribución) que genera sus datos.

El objetivo habitual al considerar este tipo de modelos es hacer inferencia acerca de $\vec{\beta}$ para lo que será necesario conocer la distribución de estadísticos pivotales (basados en algún estimador adecuado de ese parámetro vectorial).

Sabemos que el estimador de mínimo error cuadrático medio de $\vec{\beta}$ es:

$$\hat{\vec{\beta}} = \arg\min_b \sum_{i=1}^n \left(Y_i - \vec{c}_i^{\,t} b\right)^2 = \left(\mathbf{c}^t\mathbf{c}\right)^{-1}\mathbf{c}^t\vec{Y}$$

Dicho estimador satisface:

$$E_F\left(\hat{\vec{\beta}}\right) = E_F\left(\left(\mathbf{c}^t\mathbf{c}\right)^{-1}\mathbf{c}^t\vec{Y}\right) = \left(\mathbf{c}^t\mathbf{c}\right)^{-1}\mathbf{c}^t E_F\left(\vec{Y}\right) = \left(\mathbf{c}^t\mathbf{c}\right)^{-1}\mathbf{c}^t\mathbf{c}\vec{\beta} = \vec{\beta}$$

$$V_F\left(\hat{\vec{\beta}}\right) = \left(\mathbf{c}^t\mathbf{c}\right)^{-1}\mathbf{c}^t V_F\left(\vec{Y}\right)\left(\left(\mathbf{c}^t\mathbf{c}\right)^{-1}\mathbf{c}^t\right)^t = \left(\mathbf{c}^t\mathbf{c}\right)^{-1}\mathbf{c}^t\sigma^2 I_n\mathbf{c}\left(\mathbf{c}^t\mathbf{c}\right)^{-1} = \sigma^2\left(\mathbf{c}^t\mathbf{c}\right)^{-1}$$

$$R\left(\vec{X};F\right) = \sqrt{n}\left(\hat{\vec{\beta}} - \vec{\beta}\right) \simeq N_p\left(\vec{0}, \sigma^2\left(\mathbf{c}^t\mathbf{c}\right)^{-1}\right)$$

(Teoría clásica)

Remuestreo bootstrap uniforme

$$P=\left(\vec{\beta};F\right)$$
$$\downarrow$$
$$\left\{x_i=\left(\vec{c}_i^{\,t},y_i\right),\ i=1,\ldots,n\right\}$$

$$P^*=\left(\hat{\vec{\beta}};\hat{F}\right)\quad\begin{cases}\hat{\vec{\beta}}=\left(\mathbf{c}^t\mathbf{c}\right)^{-1}\mathbf{c}^t\vec{y}\quad\left(\vec{y}^t=\left(y_1,\ldots,y_n\right)\right)\\ \hat{F}\ \textit{empírica de}\ \left\{\hat{u}_1,\ldots,\hat{u}_n\right\}\quad\left(\hat{u}_i=y_i-\vec{c}_i^{\,t}\hat{\vec{\beta}},\ i=1,\ldots,n\right)\end{cases}$$
$$\left(\textit{residuos centrados}:\ \hat{u}_i\equiv\hat{u}_i-\bar{\hat{u}}_i\right)$$

$$\downarrow$$

$$x_i^*=\left(\vec{c}_i^{\,t};y_i^*\right),\ i=1,\ldots,n\quad\begin{cases}\vec{c}_i^{\,t}=\left(c_{i1},\ldots,c_{ip}\right)\ \textit{parte eplicativa conocida}\\ y_i^*=\vec{c}_i^{\,t}\hat{\vec{\beta}}+u_i^*\quad\left(\vec{u}^{*t}=\left(u_1^*,\ldots,u_1^*\right)\ \textit{m.a.s. particular de}\ U^*\in\hat{F}\right)\end{cases}$$

Estos datos satisfacen el modelo:

$$\vec{Y}^*=\mathbf{c}\vec{\beta}+\vec{U}^*,\quad\left(\vec{Y}^*=\begin{pmatrix}Y_1^*\\ \cdots\\ Y_n^*\end{pmatrix},\mathbf{c}=\begin{pmatrix}c_{11} & & c_{1p}\\ & \cdots & \\ c_{n1} & & c_{np}\end{pmatrix}\right)$$

NOTA: Análogamente se procedería con los bootstrap paramétrico, simetrizado o suavizado, reemplazando $\hat{F}$ por los correspondientes estimadores de la función de distribución del error.

En este modelo bootstrap, el estimador de $\vec{\beta}$ de mínimo error cuadrático medio, basado en la réplica $\left(x_1^*,\ldots,x_n^*\right)$ es:

$$\hat{\vec{\beta}}^*=\arg\min_b\sum_{i=1}^n\left(Y_i^*-\vec{c}_i^{\,t}b\right)^2=\left(\mathbf{c}^t\mathbf{c}\right)^{-1}\mathbf{c}^t\vec{Y}^*$$

Notemos que este estimador satisface

$$E_{\hat{F}}\left(\hat{\vec{\beta}}^*\right)=E_{\hat{F}}\left(\left(\mathbf{c}^t\mathbf{c}\right)^{-1}\mathbf{c}^t\vec{Y}^*\right)=\left(\mathbf{c}^t\mathbf{c}\right)^{-1}\mathbf{c}^tE_{\hat{F}}\left(\vec{Y}^*\right)=\left(\mathbf{c}^t\mathbf{c}\right)^{-1}\mathbf{c}^t\mathbf{c}\hat{\vec{\beta}}=\hat{\vec{\beta}}$$
$$V_{\hat{F}}\left(\hat{\vec{\beta}}^*\right)=\left(\mathbf{c}^t\mathbf{c}\right)^{-1}\mathbf{c}^tV_{\hat{F}}\left(\vec{Y}^*\right)\left(\left(\mathbf{c}^t\mathbf{c}\right)^{-1}\mathbf{c}^t\right)^t=\left(\mathbf{c}^t\mathbf{c}\right)^{-1}\mathbf{c}^t\hat{\sigma}^2I_n\mathbf{c}\left(\mathbf{c}^t\mathbf{c}\right)^{-1}=\hat{\sigma}^2\left(\mathbf{c}^t\mathbf{c}\right)^{-1},$$

que son las versiones "plug-in" de las correspondientes características de $\hat{\vec{\beta}}$ (validación por "imitación" de este procedimiento de remuestreo bootstrap).

Considerando nuevas réplicas bootstrap podemos obtener tantos valores de $\hat{\vec{\beta}}^*$ como queramos, y por lo tanto, aproximar por Monte Carlo la distribución de

$$R\left(\vec{X}^*;\hat{F}\right)=\sqrt{n}\left(\hat{\vec{\beta}}^*-\hat{\vec{\beta}}\right)$$ (Teoría bootstrap)

OBSERVACIÓN

El bootstrap en este contexto de regresión (con el objetivo de aproximar la distribución del vector paramétrico desconocido), procede en dos fases sucesivas:

1. Primero, con la muestra inicial, imita el mecanismo que la genera.
2. A continuación, ejecuta el procedimiento correspondiente (en este caso, obtención de un valor del vector paramétrico desconocido) sobre réplicas bootstrap de la muestra inicial, es decir, sobre muestras del nuevo mecanismo imitador.

NOTA

- FREEDMAN (1981), prueba que

  aproxima bien

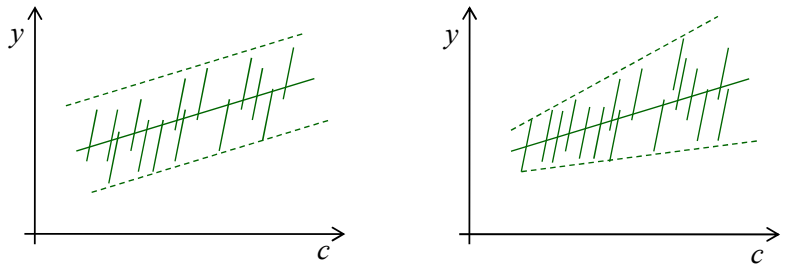
$$P_{\hat{F}}\left(\sqrt{n}\left(\hat{\vec{\beta}}^{*}-\hat{\vec{\beta}}\right)\le x\right)$$

$$P_{F}\left(\sqrt{n}\left(\hat{\vec{\beta}}-\vec{\beta}\right)\le x\right).$$

- NAVIDI (1989), prueba que esa aproximación es mejor que la normal.

WILD BOOTSTRAP

Si el modelo de regresión considerado es heterocedástico, es decir, la varianza del error es función de la parte predictora del dato (no constante, como antes),



entonces el bootstrap uniforme puede ser inadecuado puesto que no disponemos de n datos i.i.d. de un único error (perturbación aleatoria), sino de un dato único para cada uno de los n errores correspondientes.

El wild bootstrap (HÄRDLE-MAMMEN, 1993; HÄRDLE-MARRON, 1991), al permitir remuestrear un modelo a partir de un único dato observado procedente del mismo, mejora el comportamiento del bootstrap uniforme. En vez de considerar la empírica de los residuos centrados

$$U^{*}\in\begin{pmatrix}1/n & \cdots & 1/n\\ \hat{u}_1 & \cdots & \hat{u}_n\end{pmatrix}$$

considera para cada $i = 1,\ldots,n$ un modelo discreto para su correspondiente error que toma sólo dos valores

$$U_i^* \in \begin{pmatrix} \gamma & 1-\gamma \\ a & b \end{pmatrix}$$

adaptándose al único dato disponible de dicha variable aleatoria, $\hat{u}_i$, en el siguiente sentido:

$$\gamma a + (1-\gamma) b = 0$$
$$\gamma a^2 + (1-\gamma) b^2 = \hat{u}_i^2$$
$$\gamma a^3 + (1-\gamma) b^3 = \hat{u}_i^3$$

El resultado de resolver este sistema de ecuaciones es:

$$U_i^* \in \begin{pmatrix} \dfrac{5+\sqrt{5}}{10} & 1-\dfrac{5+\sqrt{5}}{10} \\ \dfrac{\hat{u}_i\left(1-\sqrt{5}\right)}{2} & \dfrac{\hat{u}_i\left(1+\sqrt{5}\right)}{2} \end{pmatrix}.$$

Tpc
Sección áurea

#REGRESION LINEAL (Y=a+bx+U: Modelo de diseño fijo y homocedástico). Estimación por Monte Carlo de la #densidad de los estimadores mínimo cuadráticos de a y b. Comparación con teoría clásica y bootstrap. #PERTURBACIÓN ALEATORIA NORMAL (0,1). Tambien EXPONENCIAL (1), aunque no adecuada, y t-STUDENT (2)

```
##########################################################################################################################################
n<-50
x<-seq(0,1,length=n)
Varx<-var(x)*(length(x)-1)/length(x)
```

x fijo, equiespaciado entre 0 y 1

#MONTE CARLO

```
N<-1000
a<-numeric(N)
b<-numeric(N)
for(h in 1:N)
{
 y<-2+3*x+rnorm(n,0,1)
 a[h]<-mean(y)-mean(x)*(cov(x,y)/Varx)
 b[h]<-cov(x,y)/Varx
}
plot(density(a),xlim=c(0,5),ylim=c(0,1.5))
lines(density(b))
```

#TEORÍA CLÁSICA

```
y<-2+3*x+rnorm(n,0,1)
a0<-mean(y)-mean(x)*(cov(x,y)/Varx)
b0<-cov(x,y)/Varx
#plot(x,y)
#abline(a=a0,b=b0)
u<-y-a0-b0*x
e<-u^2
Vare<-(n/(n-2))*mean(e)
z<-seq(0,5,length=500)
lines(z,dnorm(z,mean=a0,sd=sqrt((Vare/n)*(1+(mean(x)^2)/Varx))),col=2)
lines(z,dnorm(z,mean=b0,sd=(sqrt(Vare/(n*Varx)))),col=2)
```

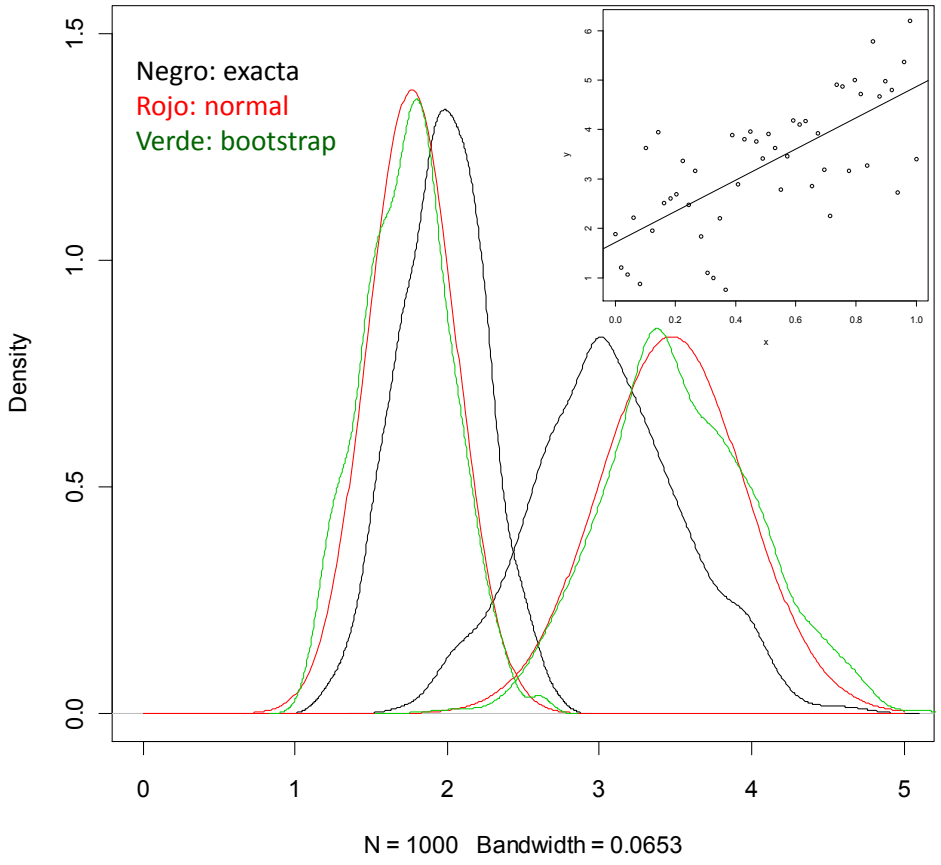
Recordemos que si la perturbación aleatoria es normal $N\left(0,\sigma^2\right)$, entonces los estimadores de a y b son normales e insesgados, con varianzas estimadas:

$$\widehat{Var\hat{a}} = \frac{\hat{\sigma}^2}{n}\left(1+\frac{\overline{x}^2}{s_x^2}\right)$$
$$\widehat{Var\hat{b}} = \frac{\hat{\sigma}^2}{n s_x^2}$$
$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n} u_i^2 \quad \left(u_i = y_i - \hat{a} - \hat{b}x_i\right)$$

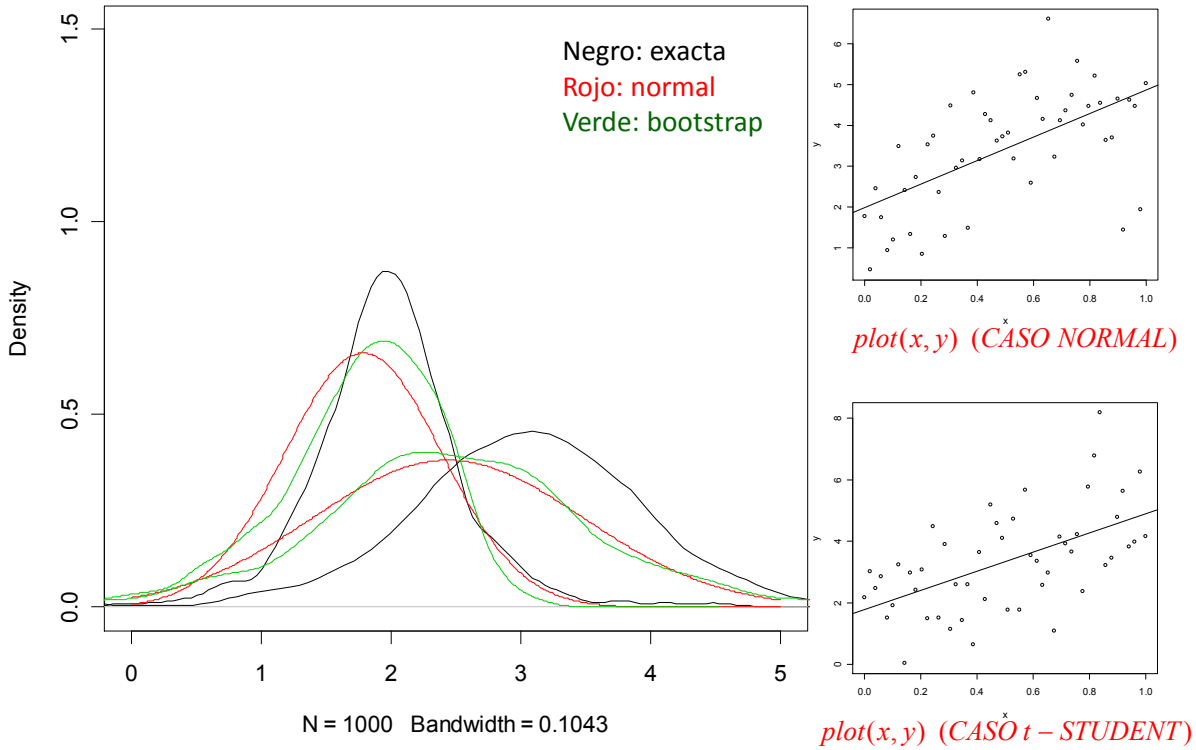
#BOOTSTRAP UNIFORME

```
muestra<-numeric(n)
muestra=u-mean(u)
yboot<-numeric(n)
B=1000
aboot<-numeric(B)
bboot<-numeric(B)
for (k in 1:B)
  {
   l<-sample(1:n,replace=TRUE)
   yboot=a0+b0*x+muestra[l]
   aboot[k]<-mean(yboot)-mean(x)*(cov(x,yboot)/Varx)
   bboot[k]<-cov(x,yboot)/Varx
   }
lines(density(aboot),col=3)
lines(density(bboot),col=3)
```

Densidad del estimador de la *ordenada en el origen* y la *pendiente* con $U \in N(0,1)$



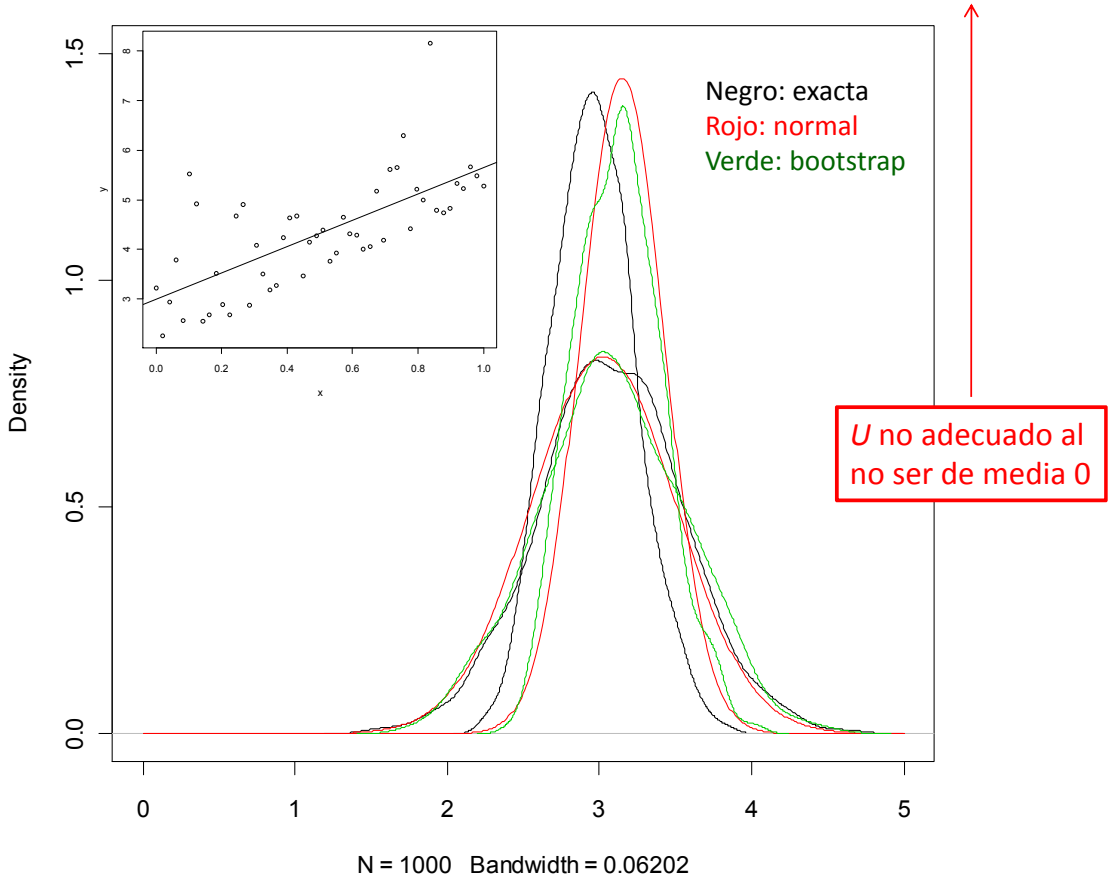
Densidad del estimador de la *ordenada en el origen* y la *pendiente* con $U \in t(2)$



$plot(x, y)$ $(CASO\ NORMAL)$

$plot(x, y)$ $(CASO\ t - STUDENT)$

Densidad del estimador de la *ordenada en el origen* y la *pendiente* con $U \in Exp(1)$



U no adecuado al no ser de media 0

OBSERVACIONES.

1. En la primera figura el bootstrap funciona prácticamente como la teoría clásica (óptima al ser normal de media 0 la perturbación aleatoria).
2. En la segunda, vemos cómo el bootstrap mejora ligeramente a la teoría clásica (no óptima al no ser normal –aunque sí de media 0- la perturbación).
3. En la tercera figura, al ser la media del error mayor que 0, el estimador de la ordenada en el origen es inadecuado (sesgado). En este caso particular los estimadores de la ordenada en el origen y de la pendiente tienen casi la misma media.
4. El desfase respecto de la distribución "exacta" con las teorías clásica y bootstrap en todas las figuras, es debido a la muestra particular considerada (más o menos buena).
5. Como puede verse en las figuras mostradas, aunque en general el bootstrap funciona de manera muy parecida a la teoría clásica, su gran ventaja respecto a ésta es que no necesita conocer las expresiones que proporcionan las varianzas de los estimadores de los parámetros del modelo, en este caso la ordenada en el origen y la pendiente (ver página anterior correspondiente).

#WILD BOOTSTRAP. REGRESION LINEAL (Y=a+bx+U: Modelo de diseño fijo y heterocedástico). Estimación por Monte Carlo #de la densidad de los estimadores mínimo cuadráticos de a y b. Comparación con teoría clásica y bootstraps uniforme y #Wild bootstrap. PERTURBACIÓN ALEATORIA NORMAL (0,x)

```
n<-50
x<-seq(0,1,length=n)
Varx<-var(x)*(length(x)-1)/length(x)
```

x fijo, equiespaciado entre 0 y 1

#MONTE CARLO

```
N<-1000
a<-numeric(N)
b<-numeric(N)
for(h in 1:N)
{
 y<-2+3*x+rnorm(n,0,x)
 a[h]<-mean(y)-mean(x)*(cov(x,y)/Varx)
 b[h]<-cov(x,y)/Varx
}
plot(density(a),xlim=c(0,5),ylim=c(0,4.5))
lines(density(b))
```

#TEORÍA CLÁSICA

```
#n=100
#x<-seq(0,1,length=n)
y<-2+3*x+rnorm(n,0,x)
a0<-mean(y)-mean(x)*(cov(x,y)/Varx)
b0<-cov(x,y)/Varx
#plot(x,y)
#abline(a=a0,b=b0)
u<-y-a0-b0*x
e<-u^2
Vare<-(n/(n-2))*mean(e)
z<-seq(0,5,length=500)
lines(z,dnorm(z,mean=a0,sd=sqrt((Vare/n)*(1+(mean(x)^2)/Varx))),col=2)
lines(z,dnorm(z,mean=b0,sd=(sqrt(Vare/(n*Varx)))),col=2)
```

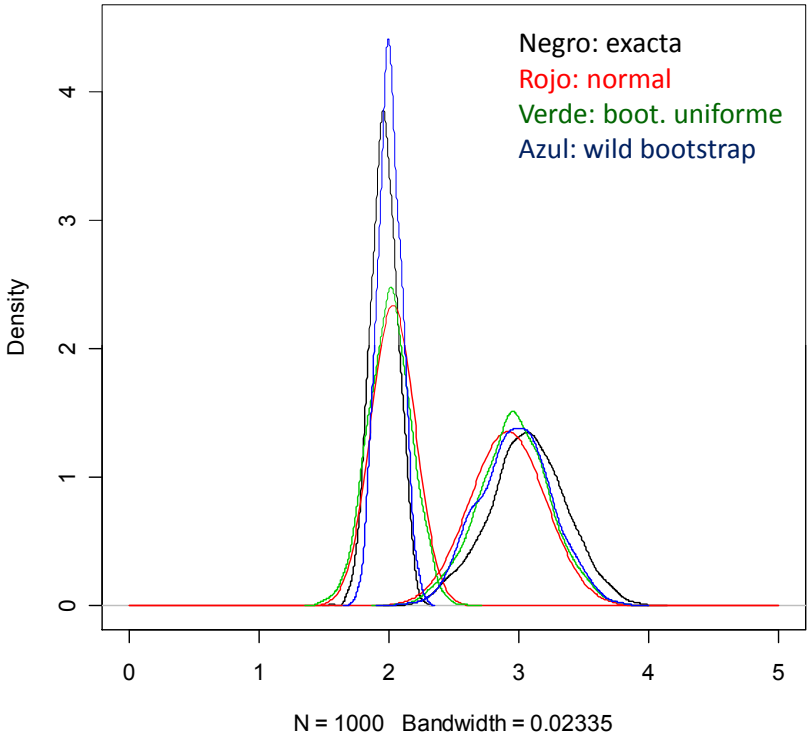
#BOOTSTRAP UNIFORME Y WILD BOOTSTRAP

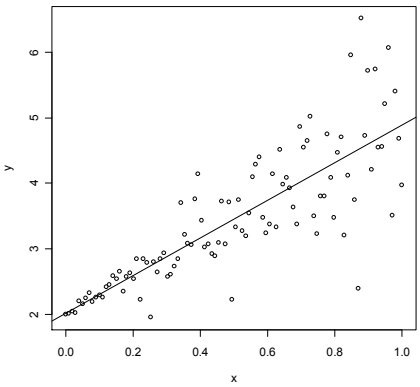
```
muestra<-numeric(n)
wmuestra<-numeric(n)
muestra=u-mean(u)
yboot<-numeric(n)
ywboot<-numeric(n)
B=1000
aboot<-numeric(B)
bboot<-numeric(B)
awboot<-numeric(B)
bwboot<-numeric(B)
for (k in 1:B)
  {
  l<-sample(1:n,replace=TRUE)
  yboot=a0+b0*x+muestra[l]
  aboot[k]<-mean(yboot)-mean(x)*(cov(x,yboot)/Varx)
  bboot[k]<-cov(x,yboot)/Varx
  for (i in 1:n)
    {
    r<-sample(c(u[i]*(1-sqrt(5))/2,u[i]*(1+sqrt(5))/2),replace=TRUE,
    +prob=c((5+sqrt(5))/10,1-(5+sqrt(5))/10));
    wmuestra[i]=r[1]
    }
    ywboot=a0+b0*x+wmuestra
    awboot[k]<-mean(ywboot)-mean(x)*(cov(x,ywboot)/Varx)
    bwboot[k]<-cov(x,ywboot)/Varx
  }
lines(density(aboot),col=3)
lines(density(bboot),col=3)
lines(density(awboot),col=4)
lines(density(bwboot),col=4)
```

b.u.

w.b.

Densidad del estimador de la *ordenada en el origen* y la *pendiente* con $U/x \in N(0,x)$





$plot(x,y)$ $(CASO\ N(0,x))$

ESTIMACIÓN BOOTSTRAP DEL ERROR DE UNA REGLA DE PREDICCIÓN

Consideraciones generales previas

1. Recordemos que las técnicas de remuestreo (el bootstrap en particular), al utilizar los datos observados para estimar el mecanismo que los genera (F, modelo de regresión, …), son capaces de replicar dichos datos tantas veces como se quiera, proporcionando otras tantas respuestas al problema planteado, al ejecutar el procedimiento estadístico correspondiente sobre dichas réplicas. Con el muestreo ordinario se tendría una única respuesta basada en los datos observados.
2. En un problema de ESTIMACIÓN, el objetivo, aunque desconocido, es FIJO (vimos muchos ejemplos). La única fuente de variabilidad en el correspondiente procedimiento estadístico son los datos. A diferencia de éste, en un problema de PREDICCIÓN (regresión, discriminación, clasificación, …), el objetivo, también desconocido, es ahora una VARIABLE ALEATORIA, por lo que existen dos fuentes de variabilidad: los datos y el propio objetivo. Estas dos fuentes de variabilidad hay que considerarlas a la hora de estimar el error (credibilidad) de una regla de predicción.
3. En lo que sigue veremos, en particular, cómo predecir con un modelo de regresión. En el epígrafe anterior vimos cómo estimar el vector de parámetros.

Una aplicación importante del bootstrap es la **estimación del error de una regla de predicción** (error al explicar, por ejemplo, peso con talla; discriminar-diagnosticar como sano o enfermo en base a unos síntomas; clasificar como hombre o mujer en base a una variable observada, …).

Supongamos una población (p+1)-dimensional

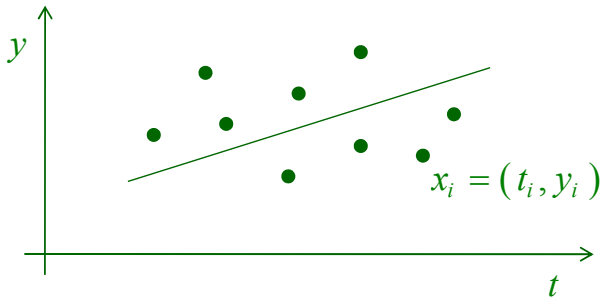
$$X = (\vec{T}, Y) \quad \begin{cases} \vec{T} = (T_1, \ldots, T_p)^t & parte\ predictora\ (explicativa) \\ Y & respuesta\ asociada\ a\ \vec{T} \end{cases}$$

- **REGLA DE PREDICCIÓN**

Sea $\vec{x} = (x_1, \ldots, x_n)^t$ una muestra aleatoria simple particular de X. Una regla de predicción basada en $\vec{x}$, $\eta(t, \vec{x})$, nos permitirá predecir la futura respuesta particular y_0 correspondiente a un valor t_0 observado, mediante $\eta(t_0, \vec{x})$.

$$\vec{x} \rightarrow y = \eta(t, \vec{x})$$

Ejemplo 1. Contexto de regresión

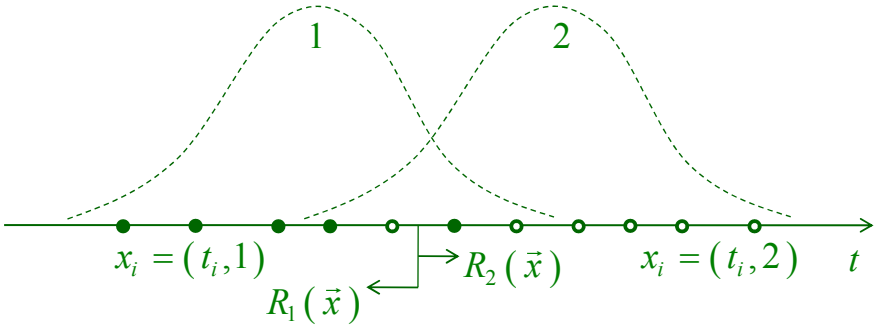


$$\vec{x} \rightarrow \eta(t,\vec{x}) = \bar{y} + \hat{\rho}_{yt}\frac{\hat{\sigma}_y}{\hat{\sigma}_t}(t-\bar{t})$$

(Recta de regresión de Y sobre T)

$$\eta(t,\vec{x}) = \sum_{i=1}^{n} y_i \frac{K_h(t-t_i)}{\sum_{j=1}^{n} K_h(t-t_j)}$$ es otra posible regla de predicción

Ejemplo 2. Contexto de discriminación entre dos poblaciones



$$\vec{x} \rightarrow y = \eta(t,\vec{x}) = \begin{cases} 1 & t \in R_1(\vec{x}) \\ 2 & t \in R_2(\vec{x}) \end{cases}$$ (Regla de discriminación)

Notemos que el hecho de predecir el valor de una variable aleatoria confiere a $\eta(t,\vec{x})$ una doble aleatoriedad: la de $\vec{X}$ y la de (Y_0, T_0). Un estimador $\hat{\theta}$ de una característica poblacional θ sólo tiene una fuente de aleatoriedad, la de $\vec{X}$.

- **FUNCIÓN DE PÉRDIDA**

La función de pérdida cuantifica el error cometido al predecir y_0 con $\eta(t_0,\vec{x})$

$$Q(y_0,\eta(t_0,\vec{x}))$$

Ejemplo 1. Regresión

$$Q(y_0,\eta(t_0,\vec{x})) = (y_0-\eta(t_0,\vec{x}))^2$$

Ejemplo2. Discriminación

$$Q(y_0,\eta(t_0,\vec{x})) = \begin{cases} 0 & y_0 = \eta(t_0,\vec{x}) \\ 1 & y_0 \neq \eta(t_0,\vec{x}) \end{cases}$$

- **ERROR VERDADERO DE LA REGLA DE PREDICCIÓN**

Es el valor esperado de la función de pérdida respecto de (Y_0,T_0) (segunda fuente de variabilidad de la regla de predicción) .

$$Err(\vec{x};F) = E_F Q(Y_0,\eta(T_0,\vec{x}))$$

F $(\vec{x}\ fijo)$

Ejemplo 1. Regresión

$$Err(\vec{x};F) = E_F\left[(Y_0-\eta(T_0,\vec{x}))^2\right]$$

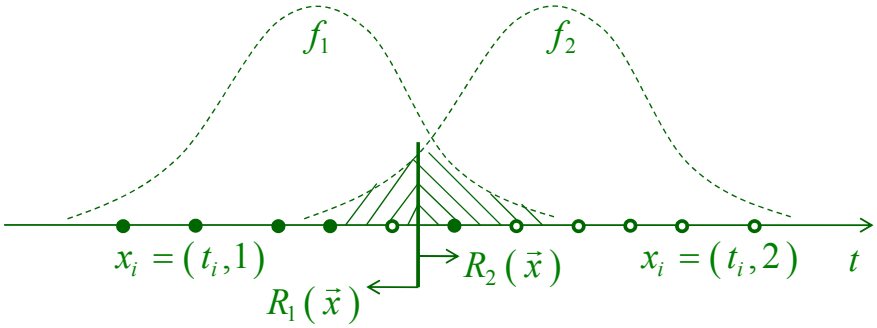
Error cuadrático medio (desconocido)

NOTA

Si η es la recta de regresión teórica, sabemos que dicho error es $\sigma_y^2(1-\rho_{yt}^2)$.

Así, si $(T,Y) \in N_2\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right) \rightarrow Err = 1-\rho^2$

Ejemplo2. Discriminación



$$\vec{x} \rightarrow y = \eta(t, \vec{x}) = \begin{cases} 1 & t \in R_1(\vec{x}) \\ 2 & t \in R_2(\vec{x}) \end{cases}$$

$$Q(y_0, \eta(t_0, \vec{x})) = \begin{cases} 0 & y_0 = \eta(t_0, \vec{x}) \\ 1 & y_0 \neq \eta(t_0, \vec{x}) \end{cases}$$

$$Err(\vec{x}; F) = E_F Q(Y_0, \eta(T_0, \vec{x})) = P_F(Y_0 \neq \eta(T_0, \vec{x})) = \frac{n_1}{n} \int_{R_1(\vec{x})} f_0(t)dt + \frac{n_2}{n} \int_{R_0(\vec{x})} f_1(t)dt$$

Probabilidad de clasificación incorrecta

- **ERROR APARENTE DE LA REGLA DE PREDICCIÓN**

Es la evaluación del error verdadero sobre los datos muestrales (evidentemente lo infravalorará, al haber sido construida la regla de predicción con dichos datos).

$$err(\vec{x}) = \frac{1}{n} \sum_{i=1}^{n} Q(y_i, \eta(t_i, \vec{x}))$$

Ejemplo 1. Regresión

$$err(\vec{x}) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \eta(t_i, \vec{x}))^2$$

¡la recta de regresión se construye minimizando esa expresión!

Ejempo 2. Discriminación

$$err(\vec{x}) = \frac{1}{n} \#\{y_i \neq \eta(t_i, \vec{x})\}$$

(Proporción de datos muestrales mal clasificados)

OPTIMISMO DE LA REGLA DE PREDICCIÓN

Es la diferencia entre el error verdadero y el aparente.

$$op(\vec{x};F) = Err(\vec{x};F) - err(\vec{x})$$

OPTIMISMO ESPERADO DE LA REGLA DE PREDICCIÓN

Es el valor esperado del optimismo respecto de $\vec{X}$ (primera fuente de variabilidad de la regla de predicción: variabilidad del conjunto de datos).

$$w(F) = E_F\left[\, op\left(\vec{X};F\right)\right]$$

En principio, sólo con los datos, el optimismo de cada regla particular sería 0 (pues los errores verdadero y aparente coincidirían), y por lo tanto también el optimismo esperado.

Notemos que el optimismo de una regla de predicción, $op\left(\vec{X};F\right)$, es una variable aleatoria del tipo

$$R\left(\vec{X};F\right)$$

cuya distribución (en particular su esperanza) podemos aproximar por bootstrap:

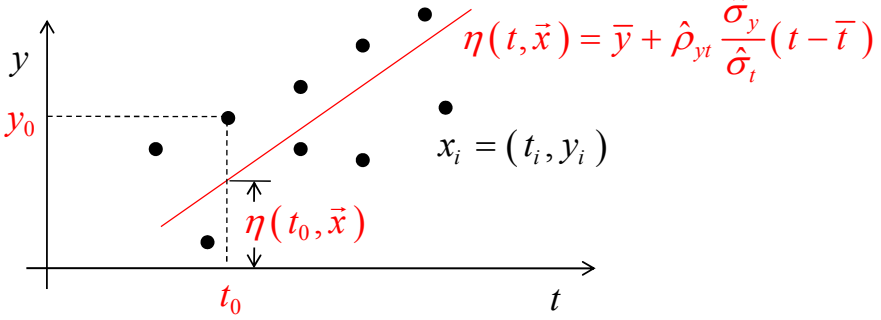
$$\widehat{w(F)}^{boot}(\vec{x}) = E_{\hat{F}}\left(op\left(\vec{X}^*;\hat{F}\right)\right) \simeq \frac{1}{B}\sum_{b=1}^{B} op\left(\vec{x}^{*b};\hat{F}\right).$$

Si consideramos ahora

$$err(\vec{x}) + \widehat{w(F)}^{boot}(\vec{x})$$

tendremos una estimación ideal, en base a $\vec{x}$, del error cometido al estimar Y_0 con $\eta(T_0,\vec{x})$.

Ejemplo 1. Regresión (revisión de conceptos) $(y_i \in \mathbb{R})$



$$Q(y_0, \eta(t_0,\vec{x})) = (y_0 - \eta(t_0,\vec{x}))^2$$

$$Err(\vec{x};F) = E_F\left[(Y_0 - \eta(T_0,\vec{x}))^2\right]$$

$$err(\vec{x}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \eta(t_i,\vec{x}))^2$$

$$op(\vec{x};F) = Err(\vec{x};F) - err(\vec{x}) \quad \rightarrow \quad w(F) = E_F\left[op(\vec{X};F)\right]$$

```
#Calcula el optimismo esperado de una regla de predicción (REGRESIÓN LINEAL)
#en un modelo de diseño fijo: t equiespaciado en (0,1) e y para cada t, N(2+3t,1)
#################################################################################################################
n<-100
t<-seq(0,1,length=n)
Vart<-var(t)*(length(t)-1)/length(t)
N=100
err<-numeric(N)
Err<-numeric(N)
op<-numeric(N)
for (j in 1:N)
  {
  y<-2+3*t+rnorm(n,0,1)
  a0<-mean(y)-mean(t)*(cov(t,y)/Vart)
  b0<-cov(t,y)/Vart
  u<-y-a0-b0*t
  e<-u^2
  err[j]<-mean(e)
  Er<-numeric(n)
  for (i in 1:n)
      {
      integrand<-function(z) {((z-a0-b0*t[i])^2)*dnorm(z,2+3*t[i],1)}
      p<-integrate(integrand,lower=-Inf,upper=Inf)
      Er[i]=p$value
      }
  Err[j]<-mean(Er)
  op[j]<-Err[j]-err[j]
  }

for (z in 1:5)
{
print(err[z])
print(Err[z])
print(op[z])
}
Eerr<-mean(err)
EErr<-mean(Err)
w<-mean(op)
Eerr
EErr
w
```

Error aparente de cada muestra $\vec{x}$ simulada

Error verdadero de cada muestra $\vec{x}$ simulada

$$Err(\vec{x};F) = \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\left(y - \hat{a} - \hat{b}t_i\right)^2 f_{N(2+3t_i,1)}(y)\,dy$$

Optimismo de cada muestra $\vec{x}$ simulada

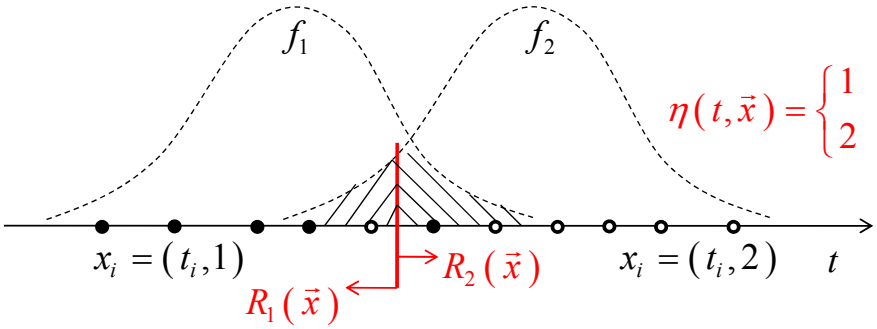
Optimismo esperado (sobre las muestras simuladas)

Salidas parciales

Regresión

n=100	Err	err	op
Muestra 1	1,021079	0,874768	0,146311
Muestra 2	1,046266	1,024506	0,021760
Muestra 3	1,014020	1,001032	0,012988
Muestra 4	1,009637	1,064847	-0,055210
Muestra 5	1,000223	1,093813	-0,093590
Media (N=100)	**1,024411**	**0,973489**	**w=0,050921**

Ejemplo 2. Discriminación entre dos poblaciones (revisión de conceptos) $(y_i = 0 \text{ ó } 1)$



$$\eta(t,\vec{x}) = \begin{cases} 1 & t \in R_1(\vec{x}) \\ 2 & t \in R_2(\vec{x}) \end{cases}$$

$$Q(y_0, \eta(t_0,\vec{x})) = \begin{cases} 0 & si \quad y_0 = \eta(t_0,\vec{x}) \\ 1 & si \quad y_0 \neq \eta(t_0,\vec{x}) \end{cases}$$

$$Err(\vec{x};F) = E_F Q(Y_0, \eta(T_0,\vec{x})) = P_F(Y_0 \neq \eta(T_0,\vec{x})) = \frac{n_1}{n} \int_{R_1(\vec{x})} f_0(t)dt + \frac{n_2}{n} \int_{R_0(\vec{x})} f_1(t)dt$$

$$err(\vec{x}) = \frac{1}{n} \#\{ y_i \neq \eta(t_i,\vec{x}) \}$$

$$op(\vec{x};F) = Err(\vec{x};F) - err(\vec{x}) \quad \rightarrow \quad w(F) = E_F\left[op(\vec{X};F) \right]$$

```
#Calcula el optimismo esperado de una regla de predicción (DISCRIMINACIÓN ENTRE DOS POBLACIONES). En este caso, dos
#poblaciones normales con igual varianza e iguales probabilidades a priori, situación en la que la función lineal discriminante de
#Fisher teórica , que asigna a la población cuya media dista menos (Mahalanobis si p>1 y predictores correlados, Euclídea  en otro
#caso) del nuevo predictor, es óptima en el sentido de clasificar en la población más probable a posteriori, minimizando la
#probabilidad de clasificación incorrecta. En la práctica se usa la función lineal estimada con las medias muestrales. (Peña, 2002)
##########################################################################################################################
n1=25
n2=25   #n=n1+n2
N=100
err<-numeric(N)
Err<-numeric(N)
op<-numeric(N)
for(j in 1:N)
 {
 t1<-rnorm(n1,4,1)
 t2<-rnorm(n2,6,1)
 m<-(mean(t1)+mean(t2))/2
 err1<-numeric(n1)
 err2<-numeric(n2)
 for (i in 1:n1)
 {
 if (t1[i]>=(mean(t1)+mean(t2))/2) {err1[i]=1}
 }
 for (k in 1:n2)
 {
 if (t2[k]<=(mean(t1)+mean(t2))/2) {err2[k]=1}
 }
 err[j]<-(n1/n)*mean(err1)+(n2/n)*mean(err2)
 integrand1<-function(z) {dnorm(z,4,1)}
 p1<-integrate(integrand1,lower=m,upper=Inf)
 Err1=p1$value
 integrand2<-function(z) {dnorm(z,6,1)}
 p2<-integrate(integrand2,lower=-Inf,upper=m)
 Err2=p2$value
 Err[j]=(n1/n)*Err1+(n2/n)*Err2
 op[j]=Err[j]-err[j]
 }
err=mean(err)
Err=mean(Err)
w=mean(op)
err; Err; w
```

$$\eta(t,\vec{x}) = \begin{cases} f_{N(4,1)} & t \le m = \dfrac{\overline{t_1}+\overline{t_2}}{2} \\ f_{N(6,1)} & t > m = \dfrac{\overline{t_1}+\overline{t_2}}{2} \end{cases} \qquad \vec{x} = (\vec{t_1}, \vec{t_2})$$

Error aparente de cada muestra $\vec{x} = (\vec{t_1}, \vec{t_2})$ simulada (proporción total de datos mal clasificados)

Error verdadero (probabilidad de mala clasificación)

$$Err(\vec{x}; F) = \frac{n_1}{n}\int_m^{\infty} f_{N(4,1)}(z)\,dz + \frac{n_2}{n}\int_{-\infty}^{m} f_{N(6,1)}(z)\,dz$$

Optimismo de cada muestra $\vec{x} = (\vec{t_1}, \vec{t_2})$ simulada

Optimismo esperado (sobre las muestras simuladas)

Discriminación

n=100	Err	err	op
Muestra 1	0,166113	0,140000	0,026113
Muestra 2	0,159029	0,100000	0,059029
Muestra 3	0,158657	0,220000	-0,061343
Muestra 4	0,158751	0,200000	-0,041249
Muestra 5	0,162218	0,200000	-0,037782
Media (N=100)	**0,161659**	**0,156800**	**w=0,004859**

Remuestreo bootstrap uniforme

$$\vec{x} \to X^* \in \hat{F} \equiv \begin{pmatrix} 1/n & \cdots & 1/n \\ x_1 & \cdots & x_n \end{pmatrix}$$

$$\vec{x}^{*b} = \left(x_1^{*b}, \ldots, x_n^{*b}\right)^t \qquad b = 1, \ldots, B$$

$$\eta\left(t^*, \vec{x}^{*b}\right)$$

$$Q\left(y_0^*, \eta\left(t_0^*, \vec{x}^{*b}\right)\right)$$

$$Err\left(\vec{x}^{*b}; \hat{F}\right) = E_{\hat{F}} Q\left(Y_0^*, \eta\left(T_0^*, \vec{x}^{*b}\right)\right) = \frac{1}{n}\sum_{i=1}^{n} Q\left(y_i, \eta\left(t_i, \vec{x}^{*b}\right)\right)$$

$$err\left(\vec{x}^{*b}\right) = \sum_{i=1}^{n} \frac{\#\left\{x_j^{*b} = x_i\right\}}{n} Q\left(y_i, \eta\left(t_i, \vec{x}^{*b}\right)\right)$$

$$op\left(\vec{x}^{*b}; \hat{F}\right) = \sum_{i=1}^{n}\left(\frac{1}{n} - \frac{\#\left\{x_j^{*b} = x_i\right\}}{n}\right) Q\left(y_i, \eta\left(t_i, \vec{x}^{*b}\right)\right)$$

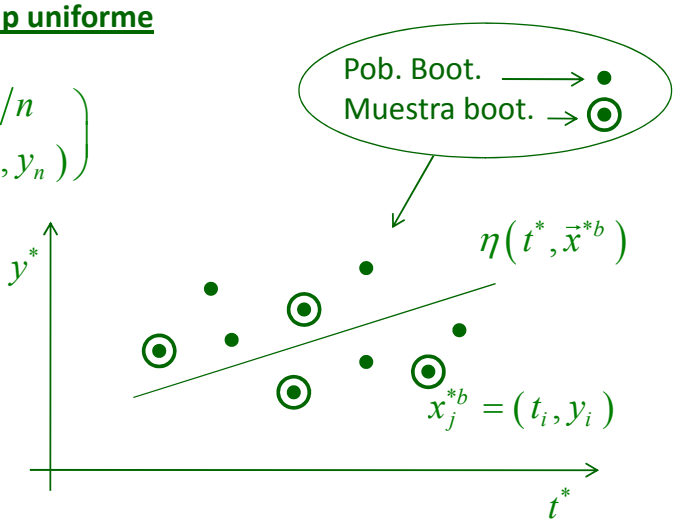
$$\widehat{w(F)}^{boot.unif.}(\vec{x}) = \frac{1}{B}\sum_{b=1}^{B} op\left(\vec{x}^{*b}; \hat{F}\right) \quad [+err(\vec{x})]$$

Est. boot. unif. error predicción de η

Ejemplo 1. Regresión. Bootstrap uniforme

$$\vec{x} \to X^* \in \hat{F} \equiv \begin{pmatrix} 1/n & \cdots & 1/n \\ (t_1, y_1) & \cdots & (t_n, y_n) \end{pmatrix}$$

$$\vec{x}^{*b} = \left(x_1^{*b}, \ldots, x_n^{*b}\right)^t \qquad b = 1, \ldots, B$$



$$Err\left(\vec{x}^{*b}; \hat{F}\right) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \eta\left(t_i, \vec{x}^{*b}\right)\right)^2$$

$$err\left(\vec{x}^{*b}\right) = \sum_{i=1}^{n} \frac{\#\left\{x_j^{*b} = x_i\right\}}{n}\left(y_i - \eta\left(t_i, \vec{x}^{*b}\right)\right)^2$$

```
#Calcula por simulación el optimismo esperado de una regla de predicción (REGRESIÓN LINEAL) en un modelo de diseño fijo: x
#equiespaciado en (0,1)  e y para cada x, N(2+3x,1).
#Estima mediante bootstrap uniforme el optimismo esperado en el modelo anterior, y lo compara con el real.
#########################################################################################################################

n<-100
t<-seq(0,1,length=n)
Vart<-var(t)*(length(t)-1)/length(t)
N=100
err<-numeric(N)
Err<-numeric(N)
op<-numeric(N)
opbu<-numeric(N)
for (j in 1:N)
{
y<-2+3*t+rnorm(n,0,1)
a0<-mean(y)-mean(t)*(cov(t,y)/Vart)
b0<-cov(t,y)/Vart
u<-y-a0-b0*t
e<-u^2
err[j]<-mean(e)      #error aparente para cada muestra
Er<-numeric(n)
for (i in 1:n)
  {
  integrand<-function(z) {((z-a0-b0*t[i])^2)*dnorm(z,2+3*t[i],1)}
  p<-integrate(integrand,lower=-Inf,upper=Inf)
  Er[i]=p$value
  }
Err[j]<-mean(Er)      #error verdadero para cada muestra
op[j]<-Err[j]-err[j]     #optimismo para cada muestra
  B=100
  Errbu<-numeric(B)
  errbu<-numeric(B)
  opbbu<-numeric(B)

for (b in 1:B)
  {
  l<-sample(1:n,replace=TRUE)
  tbu<-t[l]
  ybu<-y[l]
  Vartbu<-var(tbu)*(length(tbu)-1)/length(tbu)
  a0bu<-mean(ybu)-mean(tbu)*(cov(tbu,ybu)/Vartbu)
  b0bu<-cov(tbu,ybu)/Vartbu
  ubu<-y-a0bu-b0bu*t
  ebu<-ubu^2
  Errbu[b]<-mean(ebu)
  ubbu<-ybu-a0bu-b0bu*tbu
  ebbu<-ubbu^2
  errbu[b]<-mean(ebbu)
  opbbu[b]<-Errbu[b]-errbu[b]
  }
opbu[j]<-mean(opbbu)     #optimismo boot. uniforme para cada muestra
}
for (z in 1:5)
 {
 print(err[z])
 print(Err[z])
 print(op[z])
 print(opbu[z])
 }
Eerr<-mean(err)          #error aparente
EErr<-mean(Err)          #error verdadero
w<-mean(op)              #optimismo esperado
wbu<-mean(opbu)          #estimación boot. unif. del optimismo esperado
Eerr
EErr
w
wbu
```

Regresión

n=100	Err	err	op	Op. Boot.
Muestra 1	1,011531	1,465137	-0,453605	0,066038
Muestra 2	1,010755	1,016103	-0,005348	0,044817
Muestra 3	1,029795	0,938425	0,091369	0,034140
Muestra 4	1,065350	0,911482	0,153867	0,060072
Muestra 5	1,017091	0,876225	0,140864	0,034647
Media (N=100)	**1,023380**	**0,980672**	**w=0,042707**	**wb=0,040308**

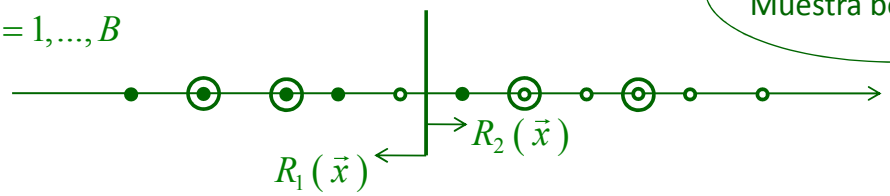
Ejemplo 2. Discriminación. Bootstrap uniforme

$$\vec{x} \to X^* \in \hat{F} \equiv \left(\begin{array}{ccc} & n_1/n & \cdots & n_2/n \\ \left(\begin{array}{ccc} 1/n_1 & \cdots & 1/n_1 \\ (t_1,1) & \cdots & (t_{n1},1) \end{array} \right) & \cdots & \left(\begin{array}{ccc} 1/n_2 & \cdots & 1/n_2 \\ (t_1,2) & \cdots & (t_{n2},2) \end{array} \right) \end{array} \right)$$

$$\vec{x}^{*b} = \left(x_1^{*b}, \ldots, x_n^{*b} \right)^t$$

$$b = 1, \ldots, B$$

Pob. Boot. →
Muestra boot. →



$$Err\left(\vec{x}^{*b}; \hat{F}\right) = \frac{1}{n}\sum_{i=1}^{n} Q\left(y_i, \eta\left(t_i, \vec{x}^{*b}\right)\right)$$

Prob. clas. mal con $\hat{F}$
Proporción datos iniciales mal

$$err\left(\vec{x}^{*b}\right) = \sum_{i=1}^{n} \frac{\#\left\{x_j^{*b} = x_i\right\}}{n} Q\left(y_i, \eta\left(t_i, \vec{x}^{*b}\right)\right)$$

Proporción datos $*$ mal
(Cada uno con su peso)

```
#Calcula  por simulación el optimismo esperado de una regla de predicción (DISCRIMINACIÓN ENTRE DOS POBLACIONES). En este caso, dos poblaciones
#normales con igual varianza e iguales probabilidades a priori, situación en la que la función lineal discriminante de Fisher teórica , que asigna a la población
#cuya media dista menos (Mahalanobis si p>1 y predictores correlados, Euclídea  en otro caso) del nuevo predictor, es óptima en el sentido de clasificar en la
#población más probable a posteriori, minimizando la probabilidad de clasificación incorrecta. En la práctica se usa la función lineal estimada con las medias
#muestrales. (Peña, 2002). Estima mediante bootstrap uniforme el optimismo esperado en el modelo anterior y lo compara con el real.
#########################################################################################################################################################
n1=25
n2=25
n=n1+n2
N=100
err<-numeric(N)
Err<-numeric(N)
op<-numeric(N)
errb<-numeric(N)
Errb<-numeric(N)
opb<-numeric(N)
for(j in 1:N)
{
  x1<-rnorm(n1,4,1)
  x2<-rnorm(n2,6,1)
  m<-(mean(x1)+mean(x2))/2
  err1<-numeric(n1)
  err2<-numeric(n2)
  for (i in 1:n1)
  {
  if (x1[i]>=m) {err1[i]=1}
  }
  for (k in 1:n2)
  {
  if (x2[k]<=m) {err2[k]=1}
  }
  err[j]<-(n1/n)*mean(err1)+(n2/n)*mean(err2)   #error aparente para cada muestra
  integrand1<-function(z) {dnorm(z,4,1)}
  p1<-integrate(integrand1,lower=m,upper=Inf)
  Err1=p1$value
  integrand2<-function(z) {dnorm(z,6,1)}
  p2<-integrate(integrand2,lower=-Inf,upper=m)
  Err2=p2$value
  Err[j]=(n1/n)*Err1+(n2/n)*Err2                #error verdadero para cada muestra
  op[j]=Err[j]-err[j]                           #optimismo para cada muestra
   B=100
   Errb<-numeric(B)
   errb<-numeric(B)
   opbb<-numeric(B)
  for (b in 1:B)
  {
   x<-numeric(n)
   for (p in 1:n)
   {
   l<-sample(c(0,1),prob=c((n1/n),(n2/n)),replace=TRUE)
   x[p]<-l[1]
   }
   sum<-sum(x)
   x1b<-numeric(sum)
   x2b<-numeric(n-sum)
   for (q in 1:sum)
    {
    z=1
    u=runif(1)
    while (u>z/n1)
    {z=z+1}
    x1b[q]=x1[z]
    }
   for (r in 1:(n-sum))
    {
    z=1
    u=runif(1)
    while (u>z/n2)
    {z=z+1}
    x2b[r]=x2[z]
    }
   mb<-(mean(x1b)+mean(x2b))/2
   err1b<-numeric(sum)
   err2b<-numeric(n-sum)
   for (s in 1:sum)
   {
   if (x1b[s]>=mb) {err1b[s]=1}
   }
   for (t in 1:(n-sum))
   {
   if (x2b[t]<=mb) {err2b[t]=1}
   }
   errb[b]<-(sum/n)*mean(err1b)+((n-sum)/n)*mean(err2b)
   Err1b<-numeric(n1)
   Err2b<-numeric(n2)
   for (a in 1:n1)
   {
   if (x1[a]>=mb) {Err1b[a]=1}
   }
   for (c in 1:n2)
   {
   if (x2[c]<=mb) {Err2b[c]=1}
   }
   Errb[b]<-(n1/n)*mean(Err1b)+(n2/n)*mean(Err2b)
   opbb[b]<-Errb[b]-errb[b]
  }
  opb[j]<-mean(opbb)        #optimismo boot. unif. para cada muestra
}
err=mean(err)    #error aparente
Err=mean(Err)    #error verdadero
w=mean(op)       #optimismo esperado
wb=mean(opb)     #estimación boot. unif. del optimismo esperado
err
Err
w
wb
```

Discriminación

n=100	Err	err	op	Op. Boot.
Muestra 1	0,161838	0,200000	-0,038161	0,009800
Muestra 2	0,159720	0,140000	0,019720	-0,016200
Muestra 3	0,167846	0,240000	-0,072153	0,009600
Muestra 4	0,158663	0,180000	-0,021336	0,004200
Muestra 5	0,161370	0,120000	0,041370	0,007600
Media (N=100)	**0,161379**	**0,157800**	**w=0,003579**	**wb=0,003876**

Remuestreo bootstrap suavizado

$$\vec{x} \to X^* \in \hat{F}_h$$

$$\downarrow$$

$$\vec{x}^{*b} = \left(x_1^{*b}, \ldots, x_n^{*b}\right)^t \qquad b = 1, \ldots, B$$

$$\longrightarrow \quad \eta\left(t^*, \vec{x}^{*b}\right)$$

$$Q\left(y_0^*, \eta\left(t_0^*, \vec{x}^{*b}\right)\right)$$

$$Err\left(\vec{x}^{*b}; \hat{F}_h\right) = E_{\hat{F}_h} Q\left(Y_0^*, \eta\left(T_0^*, \vec{x}^{*b}\right)\right) = \int Q\left(y^*, \eta\left(t^*, \vec{x}^{*b}\right)\right) d\hat{F}_h\left(t^*, y^*\right)$$

$$err\left(\vec{x}^{*b}\right) = \frac{1}{n}\sum_{i=1}^{n} Q\left(y_i^*, \eta\left(t_i^*, \vec{x}^{*b}\right)\right)$$

$$op\left(\vec{x}^{*b}; \hat{F}_h\right) = Err\left(\vec{x}^{*b}; \hat{F}_h\right) - err\left(\vec{x}^{*b}\right)$$

$$\widehat{w(F)}^{MC}_{boot.suav.}(\vec{x}) = \frac{1}{B}\sum_{b=1}^{B} op\left(\vec{x}^{*b}; \hat{F}_h\right) \quad [+err(\vec{x})]$$

Est. boot. suav. error predicción de η

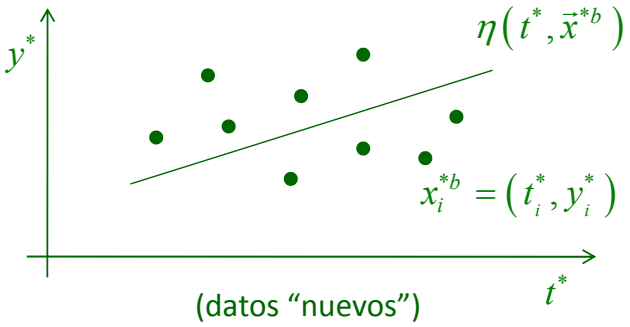
Ejemplo 1. Regresión. Bootstrap suavizado

$$\vec{x} \to X^* \in \hat{F}_h$$

$$\downarrow$$

$$\vec{x}^{*b} = \left(x_1^{*b}, \ldots, x_n^{*b} \right)^t$$

$$b = 1, \ldots, B$$



(datos "nuevos")

$$Err\left(\vec{x}^{*b}; \hat{F}_h\right) = \int \left(y^* - \eta\left(t^*, \vec{x}^{*b}\right)\right)^2 d\hat{F}_h\left(t^*, y^*\right)$$

$$err\left(\vec{x}^{*b}\right) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i^* - \eta\left(t_i^*, \vec{x}^{*b}\right)\right)^2$$

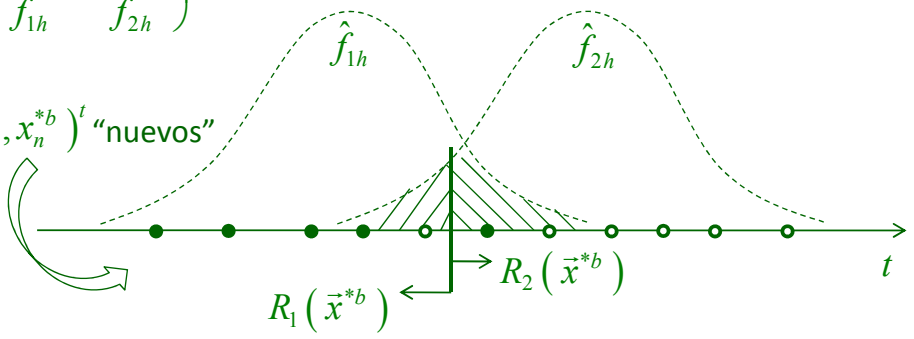
Ejemplo 2. Discriminación. Bootstrap suavizado

$$\vec{x} \to X^* \in \hat{F}_h \equiv \begin{pmatrix} n_1/n & n_2/n \\ \hat{f}_{1h} & \hat{f}_{2h} \end{pmatrix}$$

$$\downarrow$$

$$\vec{x}^{*b} = \left(x_1^{*b}, \ldots, x_n^{*b} \right)^t \text{ "nuevos"}$$

$$b = 1, \ldots, B$$



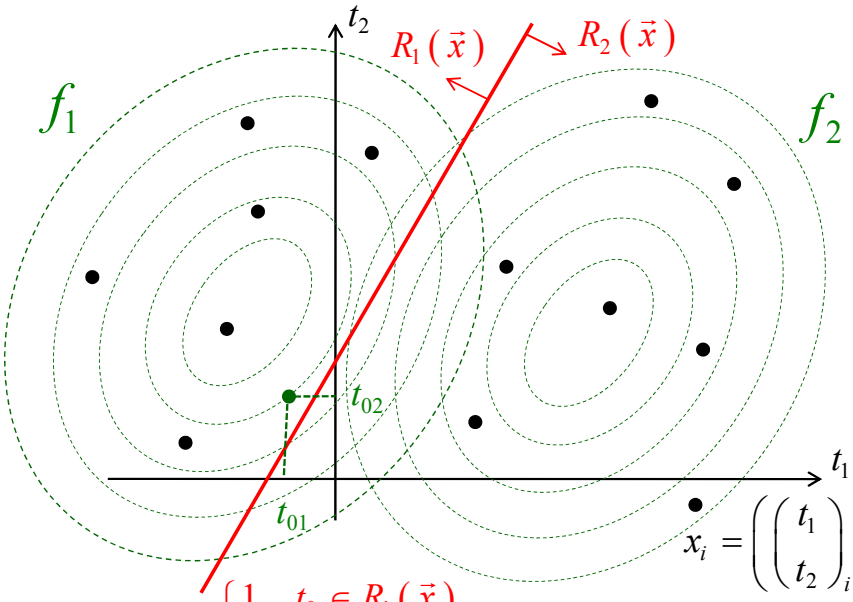
$$Err\left(\vec{x}^{*b}; \hat{F}_h\right) = \frac{n_1}{n}\int_{R_2\left(\vec{x}^{*b}\right)} \hat{f}_{1h}(t)\,dt + \frac{n_2}{n}\int_{R_1\left(\vec{x}^{*b}\right)} \hat{f}_{2h}(t)\,dt$$

Probab. de clas. mal con $\hat{F}_h$

$$err\left(\vec{x}^{*b}\right) = \frac{1}{n}\sum_{i=1}^{n} Q\left(y_i^*, \eta\left(t_i^*, \vec{x}^{*b}\right)\right)$$

Proporcion datos $*$ mal clasificados

Ejemplo 2. Discriminación. Predictor bidimensional.



$$x_i = \left(\begin{pmatrix} t_1 \\ t_2 \end{pmatrix}_i , y_i \right) \quad (y_i = 1,2)$$

$$\eta(t_0, \vec{x}) = \begin{cases} 1 & t_0 \in R_1(\vec{x}) \\ 2 & t_0 \in R_2(\vec{x}) \end{cases}$$

$$Q(y_0, \eta(t_0, \vec{x})) = \begin{cases} 0 & y_0 = \eta(t_0, \vec{x}) \\ 1 & y_0 \neq \eta(t_0, \vec{x}) \end{cases}$$

$Err(\vec{x}; F)$ Probabilidad mala clasificación

$err(\vec{x})$ Proporción de datos mal clasificados

Ejemplo 2. Discriminación. Predictor bidimensional). ESTUDIO DE SIMULACIÓN.

MODELO POBLACIONAL

$$X = (T, Y) \Big/ \begin{array}{l} f(t) = \frac{1}{2} f_{N_2\left(\begin{pmatrix} -\frac{1}{2} \\ 0 \end{pmatrix}, I\right)}(t) + \frac{1}{2} f_{N_2\left(\begin{pmatrix} \frac{1}{2} \\ 0 \end{pmatrix}, I\right)}(t) \\ P(Y=1) = P(Y=2) = \frac{1}{2} \end{array}$$

REGLA DE DISCRIMINACIÓN

$$\eta(t, \vec{x}) = \begin{cases} 1 & (t - \vec{t}_1) S^{-1} (t - \vec{t}_1)^t < (t - \vec{t}_2) S^{-1} (t - \vec{t}_2)^t \\ 2 & (t - \vec{t}_1) S^{-1} (t - \vec{t}_1)^t \geq (t - \vec{t}_2) S^{-1} (t - \vec{t}_2)^t \end{cases}$$

$$\left(S = \sum_{g=1}^{G} \frac{n_g - 1}{n - G} \left(\frac{1}{n_g - 1} \sum_{i=1}^{n_g} (t_{ig} - \bar{t}_g)(t_{ig} - \bar{t}_g)^t \right); \ aquí, G = 2 \right)$$

FUNCIÓN NÚCLEO

$$k(z) = \begin{cases} 1/4 & z \in [-1,1] \cdot [-1,1] \\ 0 & en\ otro\ caso \end{cases}$$

PARÁMETRO VENTANA

$$h_n = \frac{1}{2} n^{-\frac{1}{3}} S^{-\frac{1}{2}}$$

N=14
N=100
B=200

TABLE I

	Err	err	op	$\hat{w}_{CV}$	$\hat{w}_{BT}$	$\hat{w}_{BTS}$	$\hat{w}_{BTB}$	$\hat{w}_{BTBS}$
	0.319	0.142	0.176	0.000	0.052	0.063	-0.000	0.032
	0.316	0.142	0.173	0.000	0.033	0.051	0.001	0.095
	0.330	0.214	0.116	0.071	0.076	0.055	0.000	0.021
	0.337	0.357	-0.019	0.071	0.093	0.048	-0.001	-0.060
	0.309	0.142	0.166	0.142	0.061	0.086	-0.000	0.133
	0.338	0.214	0.124	0.142	0.081	0.115	0.001	0.095
	0.380	0.071	0.308	0.071	0.056	0.136	-0.002	0.158
	0.324	0.357	-0.033	0.000	0.081	0.078	-0.020	-0.046
	0.321	0.071	0.249	0.071	0.048	0.124	-0.001	0.142
	0.360	0.422	-0.067	0.071	0.104	0.097	0.006	-0.086
Mean	0.349	0.252	0.097	0.100	0.077	0.091	0.000	0.050
St.D.	0.057	0.128	0.126	0.089	0.027	0.027	0.007	0.073
Corr.				0.005	-0.663	-0.245	0.011	0.749
M.S.E.				0.023	0.021	0.018	0.025	0.009

NOTA

- Con n=14 el bootstrap suavizado funciona mejor que el uniforme.
- Mejor comportamiento validación cruzada que bootstrap pero con mayor M.S.E. (mayor varianza).

N=20
N=100
B=200

TABLE II

	Err	err	op	$\hat{w}_{CV}$	$\hat{w}_{BT}$	$\hat{w}_{BTS}$	$\hat{w}_{BTB}$	$\hat{w}_{BTBS}$
	0.324	0.150	0.174	0.050	0.033	0.087	-0.002	0.088
	0.318	0.400	-0.081	0.100	0.068	0.084	-0.000	-0.022
	0.318	0.200	0.118	0.050	0.041	0.076	0.005	0.093
	0.312	0.300	0.012	0.000	0.051	0.066	-0.004	0.016
	0.332	0.200	0.132	0.000	0.032	0.081	0.014	0.126
	0.326	0.250	0.076	0.100	0.058	0.064	-0.001	0.073
	0.310	0.350	-0.039	0.150	0.075	0.085	0.007	0.056
	0.521	0.300	0.221	0.100	0.070	0.091	-0.000	0.074
	0.339	0.300	0.039	0.200	0.070	0.062	0.012	0.079
	0.331	0.050	0.281	0.000	0.036	0.008	-0.004	0.201
Mean	0.333	0.277	0.056	0.053	0.060	0.074	-0.001	0.044
St.D.	0.029	0.095	0.103	0.049	0.018	0.023	0.006	0.055
Corr.				-0.251	-0.715	-0.364	0.001	0.771
M.S.E.				0.015	0.013	0.013	0.014	0.005

NOTA

- Con n=20 el bootstrap uniforme funciona mejor que el suavizado.
- Mejor comportamiento validación cruzada que bootstrap pero con mayor M.S.E. (mayor varianza).

INTRODUCCIÓN A LA UTILIZACIÓN DEL BOOTSTRAP EN POBLACIONES FINITAS

Basándonos en que sabemos cómo utilizar el bootstrap en la estimación del vector de parámetros de un modelo de regresión lineal (primer epígrafe de este tema) veremos a continuación su utilidad en la estimación del error que se comete (esperanza y varianza del mismo) al estimar la función de distribución de una variable de interés Y, definida sobre una población finita $P = \{1,...,N\}$, con un estimador adecuado. Supondremos para ello que se dispone de información auxiliar conocida X, y que dicha variable de interés satisface un modelo de regresión lineal (modelo superpoblacional).

EJEMPLO

Consideremos:

N plantaciones de caña de azúcar $(P = \{1,...,N\})$

$Y \rightarrow$ cosecha producida en cada plantación (variable de interés)

$X \rightarrow$ superficie de la plantación (conocida)

$$F_N(t) = P(Y \le t) = \frac{1}{N}\sum_{k \in P} I(Y_k \le t)$$

Planteamiento del problema: MODELO DE SUPERPOBLACIÓN

$$\xi : Y = a + bX + U, \quad U \in G \text{ desconocida} / E_G U = 0,\ V_G U = \sigma^2$$

$\downarrow$

$Y_1,...,Y_N \quad m.a.s. \text{ del modelo } \xi \rightarrow$ Población finita de interés

$[Y_k = a + bx_k + U_k, \quad k \in P,\ (U_k \ iid\ U)]$

$\downarrow$

$\{(Y_i, x_i), i \in S \subset P, \#S = n\} \rightarrow$ Muestra observada (sr: sin reemplazamiento)

Observemos que la función de distribución de la variable de interés puede descomponerse en dos componentes, dependiendo respectivamente de la muestra observada (parte conocida) y de lo que llamaremos la comuestra (parte desconocida):

$$F_N(t) = \frac{1}{N}\sum_{k \in P} I(Y_k \le t) = \frac{1}{N}\left[\sum_{i \in S} I(Y_i \le t) + \sum_{j \in P/S} I(Y_j \le t)\right]$$

Estimador de CHAMBERS/DUNSTAN (1986) para $F_N(t)$

$$E_\xi(F_N(t)) = \frac{1}{N}\left[\sum_{i\in S} I(I_i \le t) + \sum_{j\in P/S} E_\xi\left(I\left(I_j \le t\right)\right)\right]$$

$$E_\xi\left(I\left(I_j \le t\right)\right) = P_\xi\left(Y_j \le t\right) = P_\xi\left(a + bx_j + U_j \le t\right) =$$

$$= P_\xi\left(U_j \le t - a - bx_j\right) = G\left(t - a - bx_j\right) \simeq \hat{G}\left(t - \hat{a} - \hat{b}x_j\right) =$$

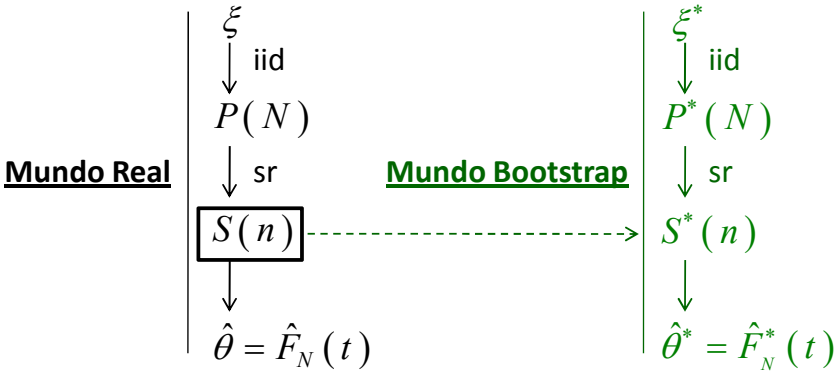
$$= \frac{1}{n}\sum_{i\in S} I\left(\hat{u}_i \le t - \hat{a} - \hat{b}x_j\right) \quad \left(\hat{u}_i = Y_i - \hat{a} - \hat{b}x_i, \quad i \in S\right)$$

(Mínimos cuadrados)

$$\hat{F}_N(t) = \frac{1}{N}\left[\sum_{i\in S} I(Y_i \le t) + \sum_{j\in P/S} \hat{G}\left(t - \hat{a} - \hat{b}x_j\right)\right]$$

CHAMBERS/DORFMAN/HALL (1992) obtuvieron expresiones asintóticas para la esperanza y varianza de la variable aleatoria $\left(\hat{F}_N(t) - F_N(t)\right)$. (Con el planteamiento que hacemos, esa variable es aleatoria en sus dos términos).

Planteamiento bootstrap (LOMBARDÍA/GONZÁLEZ/PRADA (2002))



1. Modelo bootstrap $\longrightarrow$
$$\xi^* : Y^* = \hat{a} + \hat{b}x + U^* \quad \left(U^* \in \hat{G}\right)$$
$$\hat{G}(u) = \frac{1}{n}\sum_{i=1}^{n} I\left(\hat{u}_i - \bar{\hat{u}} \le u\right)$$
$$\hat{u}_i = Y_i - \hat{a} - \hat{b}x_i; \quad \bar{\hat{u}} = \frac{1}{n}\sum_{i=1}^{n}\hat{u}_i$$
(Análogamente boot. suavizado)
2. Población finita bootstrap $\rightarrow \left\{\left(Y_k^*, x_k\right), \; k \in P^*\right\}$ *iid de* 1.
3. Muestra bootstrap $\longrightarrow \left\{\left(Y_i^*, x_i\right), \; i \in S^*\right\}$ *sr de* 2.

$$F_N^*(t) = \frac{1}{N}\left[\sum_{i\in S^*} I\left(Y_i^* \le t\right) + \sum_{j\in P^*/S^*} I\left(Y_j^* \le t\right)\right]$$

$$\hat{F}_N^*(t) = \frac{1}{N}\left[\sum_{i\in S^*} I\left(Y_i^* \le t\right) + \sum_{j\in P^*/S^*} \hat{G}^*\left(t - \hat{a}^* - \hat{b}^* x_j\right)\right]$$

$$\left[\hat{G}^*(u) = \frac{1}{n}\sum_{i=1}^{n} I\left(\hat{u}_i^* - \bar{\hat{u}}^* \le u\right);\ \hat{u}_i^* = Y_i^* - \hat{a}^* - \hat{b}^* x_i;\quad \bar{\hat{u}}^* = \frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^*\right]$$

LOMBARDÍA/GONZÁLEZ/PRADA (2002) obtuvieron expresiones asintóticas para la esperanza y la varianza de la variable aleatoria $\left(\hat{F}_N^*(t) - F_N^*(t)\right)$, con la misma estructura que las obtenidas por CHAMBERS/DORFMAN/HALL (1992), reemplazando elementos teóricos (mundo real) por sus elementos empíricos correspondientes (mundo bootstrap). Este resutado sirve para validar en este contexto el remuestreo bootstrap considerado (validación por "imitación").

N		$\varepsilon \in N(0,1)$			$\varepsilon \in t-Student$			$\varepsilon \in Mixtura$		
n	$F_N(t)$	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75
	MSE_n	5.987	7.980	5.460	5.121	7.402	5.294	6.188	7.148	5.402
100	$AMSE_n$	5.198	7.101	5.188	5.946	6.977	5.259	5.830	7.310	5.036
25	MSE	4.992	6.402	4.515	5.328	5.906	5.150	4.927	7.350	4.418
	$AMSE$	4.331	5.929	4.323	5.267	5.871	4.712	5.177	7.012	4.393
	MSE_*	4.305	5.701	4.203	4.832	5.222	4.304	4.381	6.419	4.213
	MSE_n	1.010	1.608	1.102	1.009	1.501	1.200	1.200	1.600	1.109
500	$AMSE_n$	1.111	1.423	1.020	0.984	1.429	1.110	1.211	1.497	1.075
125	MSE	0.969	1.249	0.918	0.800	1.612	0.944	1.100	1.502	1.010
	$AMSE$	0.927	1.189	0.849	0.885	1.213	0.991	1.084	1.455	0.944
	MSE_*	0.900	1.100	0.900	1.001	1.200	0.903	1.000	1.402	1.002
	MSE_n	0.400	0.502	0.305	0.300	0.404	0.317	0.403	0.504	0.403
500	$AMSE_n$	0.370	0.474	0.340	0.328	0.476	0.370	0.404	0.499	0.358
250	MSE	0.328	0.429	0.304	0.301	0.585	0.329	0.405	0.505	0.304
	$AMSE$	0.329	0.422	0.302	0.306	0.428	0.343	0.375	0.490	0.329
	MSE_*	0.300	0.400	0.300	0.300	0.400	0.300	0.400	0.500	0.300

Tabla 2.1. *$MSE \times 10^3$, aproximado por Monte Carlo y asintótico de $(F_n(t) - F_N(t))$ y $(\hat{F}(t) - F_N(t))$, respectivamente, y para este último también la estimación bootstrap.*

Consideraciones técnicas estudio simulación

1. La mixtura (perturbación aleatoria) es de poblaciones equiprobables N(-2,1) y N(2,1).
2. Se obtuvieron I=500 muestras iniciales de tamaño n (para MSE Monte Carlo). Para cada una se generaron B=50 poblaciones finitas bootstrap. De cada una de ellas se extrajeron R=100 muestras bootstrap de tamaño n.

$$MSE_* \simeq \frac{1}{I}\frac{1}{B}\frac{1}{R}\sum_{i=1}^{I}\sum_{b=1}^{B}\sum_{r=1}^{R}\left(\hat{F}_N^{*ibr}(t) - F_N^{*ib}(t)\right)^2$$

VALIDACIÓN DE LA APROXIMACIÓN BOOTSTRAP

La validación teórica de la aproximación bootstrap puede hacerse de diferentes formas:

1. Probando que la distribución bootstrap de la variable de interés converge hacia el mismo límite que la distribución ordinaria.
2. Probando que una distancia funcional entre ambas distribuciones converge hacia cero.
3. Construyendo los desarrollos de Edgeworth de ambas distribuciones para comprobar el orden de la convergencia a cero de la diferencia entre las mismas.
4. Probando que características análogas de ambas distribuciones tienen la misma estructura, reemplazando momentos poblacionales por muestrales (validación por imitación).

Mostramos a continuación algunos resultados notables relativos a la validación de los bootstraps suniforme, suavizado y simetrizado, utilizando los tres primeros procedimientos indicados. El cuarto fue utilizado en el epígrafe relativo al bootstrap en poblaciones finitas.

VALIDACIÓN APROXIMACIÓN BOOTSTRAP UNIFORME

BICKEL-FREEDMAN (1981)

$$\left(R\left(\vec{X}; F_0 \right) = \sqrt{n}\left(\bar{X}_n - \mu \right) \right)$$

Si $X_1, X_2, \ldots$ son variables aleatorias iid con varianza finita σ^2, entonces para casi toda sucesión muestral $X_1, X_2, \ldots$, dado $\left(X_1, \ldots, X_n \right)$, la distribución condicionada de

$$\sqrt{m}\left(\bar{X}_m^* - \bar{X}_n \right)$$

converge a una $N\left(0, \sigma^2 \right)$ si n y m tienden a infinito. Notemos que esa es también la distribución límite de

$$\sqrt{n}\left(\bar{X}_n - \mu \right).$$

NOTA. Los mismos autores prueban un resultado análogo para la mediana (si esta es única, ν, la población tiene densidad f con derivada positiva en un entorno de la misma, y $n = m$), con ley límite

$$N\left(0, \frac{1}{4 f^2(\nu)} \right).$$

BICKEL-FREEDMAN (1981)

$$\left(R\left(\vec{X};F_0 \right) = \sqrt{n}\left(\bar{X}_n - \mu \right) \right)$$

$$d_2\left(P_{F_1}\left\{ R\left(\vec{X}^*;F_1 \right) \le \cdot \Big/ X_1,\ldots,X_n \right\}, P_{F_0}\left\{ R\left(\vec{X};F_0 \right) \le \cdot \right\} \right) \xrightarrow{C.S} 0,$$

siendo d_2 una métrica de Mallows.

SINGH (1981)

$$\left(R\left(\vec{X};F_0 \right) = \sqrt{n}\left(\bar{X}_n - \mu \right) \right)$$

$$d_\infty\left(P_{F_1}\left\{ R\left(\vec{X}^*;F_1 \right) \le \cdot \Big/ X_1,\ldots,X_n \right\}, P_{F_0}\left\{ R\left(\vec{X};F_0 \right) \le \cdot \right\} \right) = O\left(n^{-\frac{1}{2}} \right) C.S.$$

HALL (1988)

$$\left(R\left(\vec{X};F_0 \right) = \frac{\sqrt{n}\left(\bar{X}_n - \mu \right)}{S_n} \right)$$

$$P_{F_0}\left\{ \sqrt{n}\frac{\left(\bar{X}_n - \mu \right)}{S_n} \le x \right\} = \Phi(x) + n^{-\frac{1}{2}}\left(\frac{k_3}{6\sigma^3}\left(1+2x^2 \right) \right)\phi(x) + O\left(n^{-1} \right)$$

$$P_{F_1}\left\{ \sqrt{n}\frac{\left(\bar{X}_n^* - \bar{X}_n \right)}{S_n^*} \le x \right\} = \Phi(x) + n^{-\frac{1}{2}}\left(\frac{k_{3n}}{6S_n^3}\left(1+2x^2 \right) \right)\phi(x) + O_P\left(n^{-1} \right)$$

$$\Downarrow$$

$$\begin{array}{c} \textit{Aproximación normal}: \quad O\left(n^{-1/2} \right) \\ \textit{Aproximación bootstrap}: \quad O_P\left(n^{-1} \right) \end{array}$$

NOTA. Si F_0 simétrica, como $\mu_3 = 0$ entonces: $\begin{cases} \textit{normal}: \ O\left(n^{-1} \right) \\ \textit{bootstrap}: \ O_P\left(n^{-1} \right) \end{cases}$

VALIDACIÓN APROXIMACIÓN BOOTSTRAP SUAVIZADO

CAO (1990)

$$\left(R\left(\vec{X}; F_0 \right) = (nh)^{1/2} \left(\hat{f}_h(x) - f(x) \right) \right)$$

$$\sup_z \left| P\left((nh)^{1/2} \left(\hat{f}_h(x) - f(x) \right) \le z \right) - \Phi\left(\frac{z - \hat{B}}{\hat{V}^{1/2}} \right) \right| = O\left(n^{-\frac{1}{5}} \right)$$

$$\sup_z \left| P\left((nh)^{1/2} \left(\hat{f}_h(x) - f(x) \right) \le z \right) - P^*\left((nh)^{1/2} \left(\hat{f}_h^*(x) - \hat{f}_g(x) \right) \le z \right) \right| = O_P\left(n^{-\frac{2}{9}} \right)$$

Mejor comportamiento aproximación bootstrap que normal: $n^{-\frac{2}{9}} < n^{-\frac{1}{5}}$

VALIDACIÓN APROXIMACIÓN BOOTSTRAP SIMETRIZADO

CAO/PRADA (1993)

$$\left(R\left(\vec{X}; F_0 \right) = \frac{\sqrt{n}\left(\bar{X}_n - \mu \right)}{S_n}, \quad F_0 \text{ simétrica} \right)$$

$$P_{F_0}\left\{ \sqrt{n}\frac{\left(\bar{X}_n - \mu \right)}{S_n} \le x \right\} = \Phi(x) + n^{-\frac{1}{2}}\left(\frac{k_3}{6\sigma^3}\left(1 + 2x^2\right) \right)\phi(x) + O\left(n^{-1} \right)$$

$$P_{F_1}\left\{ \sqrt{n}\frac{\left(\bar{X}_n^* - \bar{X}_{2n} \right)}{S_{2n}^*} \le x \right\} = \Phi(x) + n^{-\frac{1}{2}}\left(\frac{k_{3(2n)}}{6S_{2n}^3}\left(1 + 2x^2\right) \right)\phi(x) + O_P\left(n^{-1} \right)$$

$$\text{Aproximación normal}: \quad O\left(n^{-1} \right)$$

$$\text{Aproximación bootstrap}: \quad O_P\left(n^{-\frac{3}{2}} \right)$$

LA ITERACIÓN DEL PRINCIPIO BOOTSTRAP

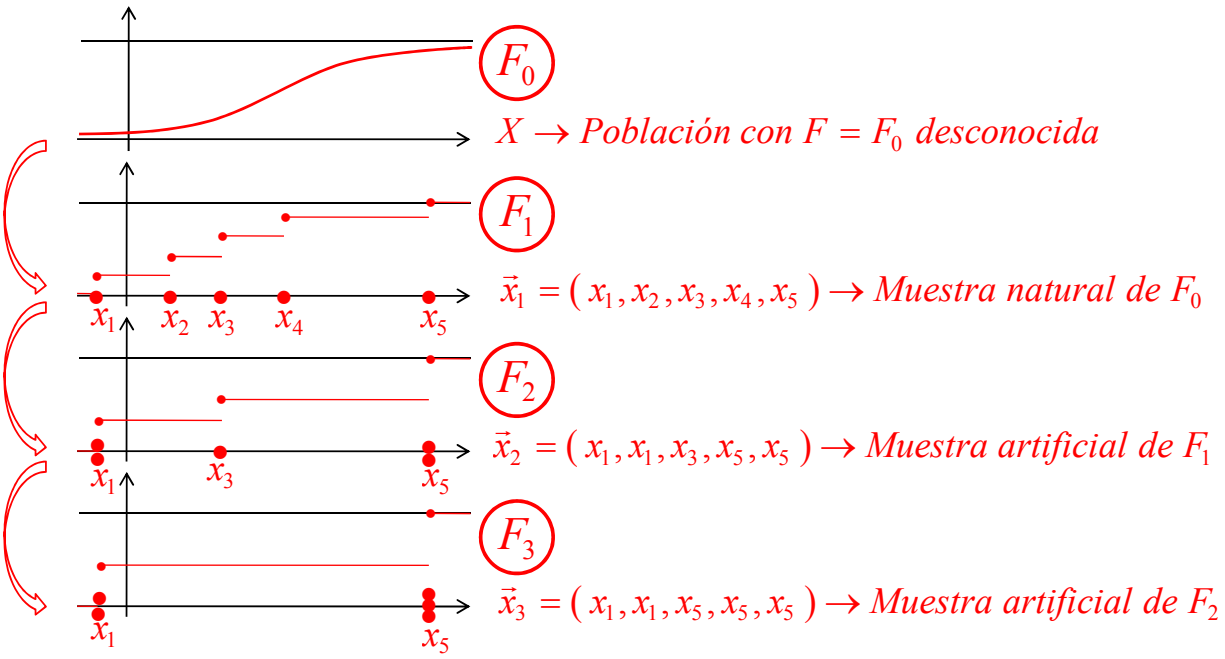
1. Motivación del bootstrap iterado. Resultados teóricos generales.
2. Bootstrap iterado y corrección del sesgo de un estimador. Estudio de simulación.
3. Bootstrap iterado y corrección del error de recubrimiento de un intervalo de confianza. Estudio de simulación.

Aplicación:

- Corrección del sesgo de un estimador y del error de recubrimiento de un intervalo de confianza mediante el bootstrap iterado. **P**

MOTIVACIÓN DEL BOOTSTRAP ITERADO. RESULTADOS TEÓRICOS GENERALES

Consideraciones preliminares



En el esquema anterior, cada F_i , $i = 1, 2, \ldots$, es una estimación cualquiera de F_{i-1} basada en $\vec{x}_i$ (se muestra la empírica de esos datos, pero podría ser cualquier otra estimación).

En ciertas condiciones generales, veremos que si la relación entre F_0 y F_1 es aproximadamente la que existe entre F_1 y F_2 ; la relación entre F_0 , F_1 y F_2 es aproximadamente la que existe entre F_1 , F_2 y F_3 ; y así, sucesivamente, entonces el proceso de inferencia acerca de una característica de F_0 se puede mejorar, sucesivamente, tanto como se quiera.

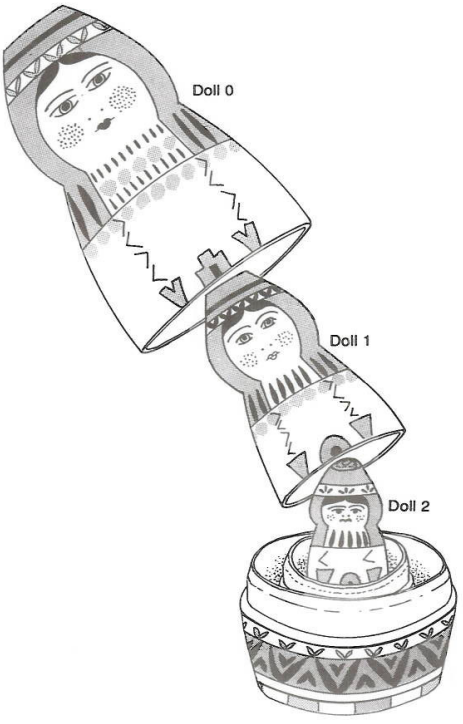


FIGURE 1.1. A Russian matryoshka doll.

Para motivar esta idea, en la que se basa el bootstrap iterado, Hall (1992) propone el siguiente ejemplo o símil físico. Consideremos un juego de muñecas rusas, en el que cada muñeca i, i=1,2,…, es una imitación a escala reducida de la muñeca i-1. Supongamos que la muñeca 0 (modelo inicial de las demás) es desconocida y que queremos estimar su "número de pecas", n_0 . Sea n_i el número de pecas de la muñeca i.

Notemos que al disminuir el tamaño de las muñecas se produce una pérdida "aleatoria" de información.

Suponiendo que la relación entre las muñecas 0 y 1 es aproximadamente la que existe entre las correspondientes 1 y 2, podemos construir un estimador inicial de n_0 , razonando de la siguiente forma:

1. **Construcción de un estimador inicial**

$$n_0 - tn_1 = 0 \quad (n_0 = tn_1)$$
$$\uparrow$$
$$n_1 - \hat{t}_{01} n_2 = 0 \Rightarrow \hat{t}_{01} = \frac{n_1}{n_2} \Rightarrow \left[\hat{n}_0 = \hat{t}_{01} n_1 = \frac{n_1}{n_2} n_1 = \frac{n_1^2}{n_2} \right]$$

Como hemos aproximado la solución de la primera ecuación por la de la segunda, se tendrá que

$$\boxed{n_0 - \hat{t}_{01} n_1 \simeq 0} \qquad (1)$$

NOTA. Considerando la primera muñeca desconocida, la segunda una muestra de la primera y la tercera una réplica de la segunda, hemos utilizado la "idea bootstrap" para construir este estimador inicial. La primera ecuación se denomina POBLACIONAL (depende de la muñeca desconocida) y la segunda MUESTRAL (depende de muñecas conocidas, observadas o replicadas).

Supongamos ahora que la relación entre las muñecas 0 , 1 y 2 es aproximadamente la que existe entre las correspondientes 1, 2 y 3.

2. **Corrección (multiplicativa) del estimador anterior**

$$n_0 - t\frac{n_1}{n_2} n_1 = 0 \quad (t : \textit{corrector multiplicativo de la estimación inicial})$$
$$\uparrow$$
$$n_1 - \hat{t}_{02} \frac{n_2}{n_3} n_2 = 0 \Rightarrow \hat{t}_{02} = \frac{n_1 n_3}{n_2^2} \Rightarrow \left[\hat{n}_0 = \hat{t}_{02} \frac{n_1}{n_2} n_1 = \frac{n_1 n_3}{n_2^2} \frac{n_1}{n_2} n_1 = \frac{n_1^3 n_3}{n_2^3} \right]$$

Al introducir un corrector de la estimación anterior y estimar dicho corrector (aproximando como antes la solución de la nueva ecuación poblacional por la de su correspondiente ecuación muestral), se tendrá que

$$\boxed{n_0 - \hat{t}_{02} \hat{t}_{01} n_1 \simeq 0} \qquad (2)$$

y cabe esperar que haya más proximidad a 0 en (2) que en (1).

NOTA. Tras estimar inicialmente, corregir la estimación y estimar la corrección , hemos utilizado dos veces la "idea bootstrap" (bootstrap iterado).

Podemos considerar nuevas correcciones suponiendo que la relación entre las muñecas 0 , 1, 2 y 3 es aproximadamente la que existe entre las correspondientes 1, 2, 3 y 4, y así, sucesivamente.

3. <u>Segunda corrección (multiplicativa) del estimador anterior</u>

$$n_0 - t\frac{n_1 n_3}{n_2^2}\frac{n_1}{n_2}n_1 = 0 \quad (t: \textit{nuevo corrector multiplicativo})$$

$$\uparrow$$

$$n_1 - \hat{t}_{03}\frac{n_2 n_4}{n_3^2}\frac{n_2}{n_3}n_2 = 0 \Rightarrow \hat{t}_{03} = \frac{n_1 n_3^3}{n_2^3 n_4} \Rightarrow$$

$$\Rightarrow \left[\hat{n}_0 = \hat{t}_{03}\frac{n_1 n_3}{n_2^2}\frac{n_1}{n_2}n_1 = \frac{n_1 n_3^3}{n_2^3 n_4}\frac{n_1 n_3}{n_2^2}\frac{n_1}{n_2}n_1 = \frac{n_1^4 n_3^4}{n_2^6 n_4}\right]$$

De esta manera, como antes, se tendrá que

$$n_0 - \hat{t}_{03}\hat{t}_{02}\hat{t}_{01}n_1 \simeq 0 \qquad (3)$$

cabiendo esperar que haya más proximidad a 0 en (3) que en (2).

NOTA. Para llegar aquí, hemos utilizado tres veces la "idea bootstrap".

Muchos problemas de inferencia estadística, en la línea de estimar alguna característica de una variable aleatoria, se pueden plantear del siguiente modo:

$$\text{¿}t\text{?}/E_{F_0}[f_t(F_0, F_1)] = 0$$ <u>Ecuación poblacional</u> (depende de F_0)

<u>Ejemplos</u>:

- $f_t(F_0, F_1) = \theta(F_1) + t - \theta(F_0) \quad [\theta(F_1)\ \textit{estimador de}\ \theta(F_0)]$

  <u>Estimador insesgado</u> de $\theta(F_0)$ <u>(t solución de la ec. anterior)</u>

- $f_t(F_0, F_1) = I_{\{\theta(F_1)-t \le \theta(F_0) \le \theta(F_1)+t\}} - (1-\alpha)$

  <u>Int. confianza simétrico</u> <u>para $\theta(F_0)$ **de cobertura exacta** (1- α)</u>
  <u>(t solución de la ec. anterior)</u>

$$\hat{t}_0/E_{F_1}\left[f_{\hat{t}_0}(F_1, F_2)\right] = 0$$ <u>Ecuación muestral</u> (se puede resolver)

Iteración del principio bootstrap. Planteamiento general

$$\begin{cases} E_{F_0}\left[f_t(F_0, F_1) \right] = 0 \quad (I) \\ E_{F_1}\left[f_{\hat{t}_{01}}(F_1, F_2) \right] = 0 \quad (II) \end{cases} \rightarrow \quad \hat{t}_{01} = T_1(F_1, F_2)$$

$$\boxed{E_{F_0}\left[f_{T_1(F_1,F_2)}(F_0, F_1) \right] \simeq 0} \longrightarrow (1)$$

$$U(F_1, F_2; t) = \begin{cases} (1+t)T_1(F_1, F_2) & \text{(corrección multiplicativa)} \\ t + T_1(F_1, F_2) & \text{(corrección aditiva)} \end{cases}$$

$$\begin{cases} E_{F_0}\left[f_{U(F_1,F_2;t)}(F_0, F_1) \right] = 0 \quad (I) \\ E_{F_1}\left[f_{U(F_2,F_3;\hat{t}_{02})}(F_1, F_2) \right] = 0 \quad (II) \end{cases} \rightarrow \quad \hat{t}_{02} = T_2(F_1, F_2, F_3)$$

$$\boxed{E_{F_0}\left[f_{U(F_1,F_2;T_2(F_1,F_2,F_3))}(F_0, F_1) \right] \simeq 0} \longrightarrow (2)$$

NOTA. Cabe esperar que (2) sea más próximo a 0 que (1).

Iteración del principio bootstrap. Resultados teóricos generales (Hall, 92)

En situaciones de regularidad, expresiones como

$$E_{F_0}\left[f_{T_1(F_1,F_2)}(F_0, F_1) \right] \simeq 0 \longrightarrow (1)$$

son habitualmente series de potencias en $n^{-1/2}$ o n^{-1} (desarrollos de Edgeworth …). Pues bien, cada iteración reduce el orden de magnitud del error por un factor de al menos $n^{-1/2}$:

PROPOSICIÓN 1

Si para un funcional suave c,

$$E_{F_0}\left[f_{T_1(F_1,F_2)}(F_0, F_1) \right] = c(F_0) n^{-j/2} + O\left(n^{-(j+1)/2} \right) \rightarrow (1)$$

y además

$$\frac{\partial}{\partial t} E_{F_0}\left[f_{U(F_1,F_2;t)}(F_0, F_1) \right]\Big|_{t=0}$$

es asintóticamente (cuando n tiende a infinito) una constante no nula, entonces

$$E_{F_0}\left[f_{U(F_1,F_2;T_2(F_1,F_2,F_3))}(F_0, F_1) \right] = O\left(n^{-(j+1)/2} \right) \rightarrow (2)$$

PROPOSICIÓN 2

Si para un funcional suave c,

$$E_{F_0}\left[f_{T_1(F_1,F_2)}(F_0,F_1)\right]=c(F_0)n^{-j}+O\left(n^{-(j+1)}\right) \quad \rightarrow (1)$$

y además

$$d(F_0)=\frac{\partial}{\partial t}E_{F_0}\left[f_{U(F_1,F_2;t)}(F_0,F_1)\right]\Big|_{t=0}$$

es asintóticamente una constante no nula, y

$$E\{\Delta e(\Delta)\}=O\left(n^{-1/2}\right),$$

siendo

$$\Delta=n^{1/2}\left\{c(F_1)d(F_1)^{-1}-c(F_0)d(F_0)^{-1}\right\},$$

$$e(\Delta)=\frac{\partial}{\partial t}E_{\Delta}\left[f_{U(F_1,F_2;t)}(F_0,F_1)\right]\Big|_{t=0},$$

entonces

$$E_{F_0}\left[f_{U(F_1,F_2;T_2(F_1,F_2,F_3))}(F_0,F_1)\right]=O\left(n^{-(j+1)}\right) \quad \rightarrow (2)$$

BOOTSTRAP ITERADO Y CORRECCIÓN DEL SESGO DE UN ESTIMADOR. ESTUDIO DE SIMULACIÓN

Primera fase

$$t \,/\, E_{F_0}\left[\theta(F_1)+t-\theta(F_0)\right]=0 \rightarrow\ t=\theta(F_0)-E_{F_0}\theta(F_1) \quad \textit{(-sesgo)}$$

$$\hat{t}_{01} \,/\, E_{F_1}\left[\theta(F_2)+\hat{t}_{01}-\theta(F_1)\right]=0 \rightarrow\ \hat{t}_{01}=\theta(F_1)-E_{F_1}\theta(F_2)$$

$$\hat{\theta}_1=\theta(F_1)+\hat{t}_{01}=2\theta(F_1)-E_{F_1}\theta(F_2)$$

(lo estimo (boot) y se lo resto al estimador)

Ejemplo:

$$\theta(F_0)=\mu^3;\quad \theta(F_1)=\bar{X}^3 \quad \rightarrow \quad \hat{\theta}_1=2\bar{X}^3-E_{F_1}\left(\bar{X}^{*3}\right)\simeq 2\bar{X}^3-\frac{1}{B}\sum_{b=1}^{B}\bar{X}_b^{*3}$$

Aproximación Monte carlo:

$$(X_1,\ldots,X_n)\rightarrow \bar{X}^3 \;\Big|\; \begin{array}{l}(X_1,\ldots,X_n)_1^*\rightarrow \bar{X}_1^{*3}\\ \cdots \\ (X_1,\ldots,X_n)_B^*\rightarrow \bar{X}_B^{*3}\end{array} \quad \rightarrow E_{F_1}\left(\bar{X}^{*3}\right)\simeq\frac{1}{B}\sum_{b=1}^{B}\bar{X}_b^{*3}$$

Estimador insesgado bootstrap aproximado por Monte Carlo de μ^3

Segunda fase

$$t \,/\, E_{F_0}\left[\theta(F_1)+\left(\underbrace{\theta(F_1)-E_{F_1}\theta(F_2)}_{\hat{t}_{01}}+t\right)-\theta(F_0)\right]=0$$

corrector aditivo del sesgo estimado

$$t = E_{F_0}\left(E_{F_1}\theta(F_2)\right)+\theta(F_0)-2E_{F_0}\theta(F_1)$$

$$\hat{t}_{02} \,/\, E_{F_1}\left[\theta(F_2)+\left(\theta(F_2)-E_{F_2}\theta(F_3)+\hat{t}_{02}\right)-\theta(F_1)\right]=0$$

$$\hat{t}_{02} = E_{F_1}\left(E_{F_2}\theta(F_3)\right)+\theta(F_1)-2E_{F_1}\theta(F_2)$$

$$\hat{\theta}_2 = \theta(F_1)+\hat{t}_{01}+\hat{t}_{02} = 3\theta(F_1)-3E_{F_1}\theta(F_2)+E_{F_1}\left(E_{F_2}\theta(F_3)\right)$$

Estimador insesgado bootstrap de $\theta(F_0)$ con una corrección aditiva del sesgo estimado, a su vez estimada por bootstrap

Ejemplo:

$$\theta(F_0)=\mu^3;\quad \theta(F_1)=\bar{X}^3 \quad\rightarrow\quad \hat{\theta}_2 = 3\bar{X}^3-3E_{F_1}\left(\bar{X}^{*3}\right)+E_{F_1}\left(E_{F_2}\left(\bar{X}^{**3}\right)\right)$$

Aproximación Monte Carlo

$$\vec{X}\rightarrow\bar{X}^3 \left|\begin{array}{l} \vec{X}_1^*\rightarrow\bar{X}_1^{*3}\left|\begin{array}{l}\vec{X}_{11}^{**}\rightarrow\bar{X}_{11}^{**3}\\ \cdots\\ \vec{X}_{1C}^{**}\rightarrow\bar{X}_{1C}^{**3}\end{array}\right. \frac{1}{C}\sum_{c=1}^{C}\bar{X}_{1c}^{**3} \\ \cdots \qquad\qquad \rightarrow E_{F_1}\left(E_{F_{2\partial}}\bar{X}^{**3}\right)\simeq\frac{1}{B}\sum_{b=1}^{B}\left(\frac{1}{C}\sum_{c=1}^{C}\bar{X}_{bc}^{**3}\right) \\ \vec{X}_B^*\rightarrow\bar{X}_B^{*3}\left|\begin{array}{l}\vec{X}_{11}^{**}\rightarrow\bar{X}_{B1}^{**3}\\ \cdots\\ \vec{X}_{1C}^{**}\rightarrow\bar{X}_{BC}^{**3}\end{array}\right. \frac{1}{C}\sum_{c=1}^{C}\bar{X}_{Bc}^{**3} \end{array}\right.$$

$$\hat{\theta}_2 \simeq 3\bar{X}^3-3\frac{1}{B}\sum_{b=1}^{B}\bar{X}_b^{*3}+\frac{1}{B}\sum_{b=1}^{B}\left(\frac{1}{C}\sum_{c=1}^{C}\bar{X}_{bc}^{**3}\right)$$

Aproximación por Monte Carlo del estimador insesgado bootstrap de μ^3 con una corrección aditiva del sesgo estimado, a su vez estimada por bootstrap

Se deduce fácilmente por inducción la expresión del estimador bootstrap con sesgo corregido tras j-1 correcciones aditivas (j iteraciones):

$$\hat{\theta}_j = \sum_{i=1}^{j+1} \binom{j+1}{i} E_{F_1}\theta(F_i)(-1)^{i+1} \quad j \geq 1,$$

donde

$$E_{F_1}\theta(F_i) = E_{F_1}\left(E_{F_2}\left(\cdots E_{F_{i-1}}\theta(F_i)\right)\right).$$

ESTUDIO DE SIMULACIÓN

Consideremos como F_0 dos posibles poblaciones F_θ de media 2:

- $X \in N(\mu, \sigma^2);\ \mu = 2, \sigma^2 = 5\ (simétrica))$
- $X \in b\chi_1^2 = \Gamma\left(\frac{1}{2b}, \frac{1}{2}\right);\ b = \mu = 2,\ (asimétrica)$

Objetivo:

$$\theta(F_0) = \left(\int x dF_0(x)\right)^3 = \mu^3 \quad (2^3 = 8\ \ en\ ambos\ escenarios)$$

Estimadores:

$$\theta(F_1) = \begin{cases} \theta(\hat{F}) = \left(\int x d\hat{F}(x)\right)^3 = \bar{X}^3 & si \quad F_1 = \hat{F} \\ \theta(\hat{F}_{2n}) = \left(\int x d\hat{F}_{2n}(x)\right)^3 = \bar{X}^3 & si \quad F_1 = \hat{F}_{2n} \\ \theta(F_{\hat{\theta}_{mv}}) = \left(\int x dF_{\hat{\theta}_{mv}}(x)\right)^3 = \bar{X}^3 & si \quad F_1 = F_{\hat{\theta}_{mv}} \end{cases}$$

Sesgo exacto:

$$E_{F_0}\left(\theta(F_1)\right) = E_{F_0}\left(\bar{X}^3\right) = E_{F_0}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\mu + \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)\right)^3\right] =$$

$$= E_{F_0}\left[\mu^3 + n^{-3}\left(\sum_{i=1}^{n}(X_i-\mu)\right)^3 + 3\mu n^{-2}\left(\sum_{i=1}^{n}(X_i-\mu)\right)^2 + 3\mu^2 n^{-1}\sum_{i=1}^{n}(X_i-\mu)\right] =$$

$$= \mu^3 + \underbrace{n^{-2}\gamma + 3n^{-1}\mu\sigma^2}_{\text{sesgo}} \quad \left[\gamma = E\left((X-\mu)^3\right);\ \sigma^2 = E\left((X-\mu)^2\right)\right]$$

- **Caso normal** (n=10)

$$N(2,5) \ \rightarrow \ 8 + 3\cdot\frac{1}{10}\cdot 2\cdot 5 = 8 + \boxed{3} = 11$$

- **Caso gamma** (n=10)

$$\Gamma\left(\frac{1}{4},\frac{1}{2}\right) \ \rightarrow \ 8 + \frac{1}{10^2}\cdot 8\cdot 2^3 + 3\cdot\frac{1}{10}\cdot 2\cdot 2\cdot 2^2 = 8 + \boxed{4{,}8 + 0{,}64} = 13{,}44$$

5,44

Denotando por

$$\hat{\theta}_1^i = 2\theta\left(F_1^i\right) - E_{F_1}\theta\left(F_2^i\right) \quad \left| \begin{array}{l} i = NP \quad si \quad F_j = \hat{F}\ (j=1,2) \\ i = S \quad si \quad F_j = \hat{F}_{2n}\ (j=1,2), \\ i = P \quad si \quad F_j = F_{\hat{\theta}_{mv}}\ (j=1,2) \end{array} \right.$$

puede probarse que

$$E_{F_0}\hat{\theta}_1^i = \left| \begin{array}{cc} \mu^3 + 3n^{-2}\left(\mu\sigma^2 - \gamma\right) + 6n^{-3}\gamma + 2n^{-4}\gamma & si \quad i = NP \\ \mu^3 + 3n^{-2}\mu\sigma^2 & si \quad i = S, P\ (normal), \\ \left(\mu^3 + 3n^{-1}\mu\sigma^2 + n^{-2}\gamma\right)\left(1 - 6n^{-1} - 8n^{-2}\right) & si \quad i = P\ (b\chi_1^2) \end{array} \right.$$

de lo que se deduce que, salvo en el último caso, la reducción esperada del sesgo es de orden $O\left(n^{-1}\right)$. El cálculo de $E_{F_0}\hat{\theta}_j^i,\ j>1$ es notablemente más complicado.

Estudiaremos por simulación el comportamiento de los estimadores $\hat{\theta}_j^i,\ j=1,2,$ en las dos poblaciones y los tres casos bootstrap considerados.

TABLE I

Bootstrap estimates of μ^3 for $N(2,5)$ from samples of size 10 using nonparametric (NP), symmetrized nonparametric (S) or parametric (P) resampling with various resampling volumes B (average of 500 basic samples), with standard deviations in parentheses. $\hat{\theta}_1^*$ and $\hat{\theta}_2^*$ are the results of applying first- and second- order additive bootstrap corrections to the initial estimate $\theta(F_1^*) = \bar{X}^3$. The exact expected values of $\theta(F_1^*)$ and $\hat{\theta}_1^*$ are shown in parentheses in the heading ($*$=NP,S or P).

B	$\hat{E}_{F_0}\theta(F_1^*)$ (11)	$\hat{E}_{F_0}\hat{\theta}_1^{NP}$ (8.3)	$\hat{E}_{F_0}\hat{\theta}_2^{NP}$	$\hat{E}_{F_0}\hat{\theta}_1^{S}$ (8.3)	$\hat{E}_{F_0}\hat{\theta}_2^{S}$	$\hat{E}_{F_0}\hat{\theta}_1^{P}$ (8.3)	$\hat{E}_{F_0}\hat{\theta}_2^{P}$
100	10.956	8.258	7.947	8.380	8.138	8.334	8.041
	(10.772)	(9.887)	(9.921)	(10.163)	(10.415)	(10.104)	(10.327)
250	10.956	8.355	8.071	8.383	8.138	8.360	8.097
	(10.772)	(9.880)	(9.793)	(9.896)	(9.812)	(9.979)	(9.988)
500	10.956	8.361	8.089	8.426	8.223	8.385	8.138
	(10.772)	(9.905)	(9.839)	(10.041)	(10.081)	(9.972)	(9.937)
750	10.956	8.316	7.998	8.371	8.116	8.361	8.100
	(10.772)	(9.904)	(9.823)	(10.009)	(9.995)	(10.007)	(10.001)

TABLE II

As for Table I, but for a $2\chi_1^2$ population.

B	$\hat{E}_{F_0}\theta(F_1^*)$ (13.44)	$\hat{E}_{F_0}\hat{\theta}_1^{NP}$ (6.9568)	$\hat{E}_{F_0}\hat{\theta}_2^{NP}$	$\hat{E}_{F_0}\hat{\theta}_1^{S}$ (8.48)	$\hat{E}_{F_0}\hat{\theta}_2^{S}$	$\hat{E}_{F_0}\hat{\theta}_1^{P}$ (4.3008)	$\hat{E}_{F_0}\hat{\theta}_2^{P}$
100	13.312	7.711	8.459	7.993	7.603	4.647	10.857
	(16.723)	(10.620)	(12.392)	(11.518)	(13.076)	(8.659)	(16.428)
250	13.312	7.350	7.799	7.737	7.093	4.113	10.185
	(16.723)	(10.111)	(10.732)	(10.352)	(10.681)	(6.410)	(13.725)
500	13.312	7.175	7.510	7.906	7.464	4.317	10.495
	(16.723)	(9.641)	(9.906)	(10.127)	(10.071)	(6.299)	(14.093)
750	13.312	7.358	7.815	7.735	7.153	4.318	10.511
	(16.723)	(9.628)	(9.774)	(9.819)	(9.501)	(5.992)	(13.713)

ESTUDIO DE SIMULACIÓN CORRECCIÓN SESGO. OBSERVACIONES

- La mejora en la corrección del sesgo es siempre clara (no significativamente en el bootstrap paramétrico de $2\chi_1^2$ como era de esperar).
- En general, existe poca sensibilidad a B, sobre todo en el caso normal, por lo que es aconsejable tomar valores pequeños.
- En el caso normal,
  1. Influye muy poco el tipo de remuestreo utilizado (estimaciones todas muy parecidas).
  2. Una iteración del bootstrap (doble bootstrap) reduce prácticamente a cero el sesgo.
- En el caso $2\chi_1^2$,
  1. El bootstrap no paramétrico (uniforme) funciona bien.
  2. El simetrizado, mal (segunda estimación peor que la primera), por la asimetría de la población.
  3. El paramétrico mal, como ya se indicó.

BOOTSTRAP ITERADO Y CORRECCIÓN DEL ERROR DE COBERTURA DE UN INTERVALO DE CONFIANZA. ESTUDIO DE SIMULACIÓN

Primera fase (intervalo bilateral simétrico)

$$t \,/\, E_{F_0}\left[I_{\{\theta(F_1)-t\le\theta(F_0)\le\theta(F_1)+t\}} - (1-\alpha) \right] = 0 \rightarrow$$

$$\rightarrow\ t = (1-\alpha)\,cuantil\ de\,\left|\theta(F_1)-\theta(F_0)\right|\ bajo\ F_0 \quad \textit{(semiamplitud)}$$

$$\hat{t}_{01} \,/\, E_{F_1}\left[I_{\{\theta(F_2)-\hat{t}_{01}\le\theta(F_1)\le\theta(F_2)+\hat{t}_{01}\}} - (1-\alpha) \right] = 0 \rightarrow$$

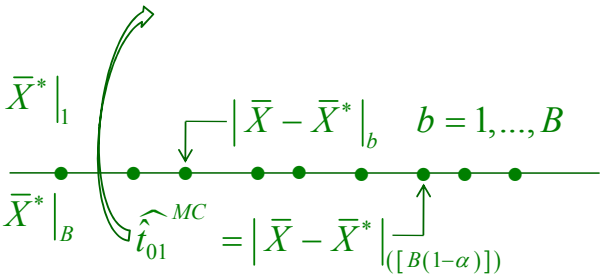
$$\rightarrow\ \hat{t}_{01} = \hat{t}_{01}(F_1, F_2) = (1-\alpha)\,cuantil\ de\,\left|\theta(F_2)-\theta(F_1)\right|\ bajo\ F_1$$

$$\left(\theta(F_1)-\hat{t}_{01},\ \theta(F_1)+\hat{t}_{01}\right) \quad \textit{(semiamplitud bootstrap)}$$

Ejemplo. Aproximación Monte carlo:

$$\begin{aligned}\theta(F_0) &= \mu\\ \theta(F_1) &= \bar{X}\\ \theta(F_2) &= \bar{X}^*\end{aligned}$$

$$\vec{X} \rightarrow \begin{cases} \vec{X}_1^* \rightarrow \left|\bar{X}-\bar{X}^*\right|_1 \\ \vec{X}_B^* \rightarrow \left|\bar{X}-\bar{X}^*\right|_B \end{cases}$$



OBSERVACIÓN

- Análogamente se construiría un intervalo unilateral:

$$\left(-\infty,\ \theta(F_1)+\hat{t}_{01}\right)$$

donde $\hat{t}_{01}$ es ahora el $(1-\alpha)$ cuantil de $(\theta(F_1)-\theta(F_2))$ bajo F_1.

- Recordemos que el *método percentil* estudiado proporcionaba el intervalo

$$\left(-\infty,\ \theta(F_1)-n^{-\frac{1}{2}}x_\alpha^*\right) \quad \left(error = O\left(n^{-\frac{1}{2}}\right)\right)$$

$$\left[P\left(n^{\frac{1}{2}}\left(\bar{X}-\mu\right) \ge x_\alpha \right) = 1-\alpha \Rightarrow P\left(\mu \le \bar{X} - n^{-\frac{1}{2}}x_\alpha \right) = 1-\alpha \rightarrow \left(-\infty, \bar{X} - n^{-\frac{1}{2}}x_\alpha \right) \right]$$

$$x_\alpha^* = \alpha\ cuantil\ de\ \ n^{\frac{1}{2}}\left(\bar{X}^*-\bar{X}\right)$$

Pues bien, ambos intervalos **coinciden**:

$$\left[\hat{t}_{01} = (1-\alpha)\,cuantil\ de\left(\bar{X}-\bar{X}^*\right) = -\alpha\ cuantil\ de\left(\bar{X}^*-\bar{X}\right) = -n^{-\frac{1}{2}}x_\alpha^* \right]$$

NOTA En el caso bilateral no ocurre eso, pues el intervalo percentil no tiene por qué ser simétrico.

Segunda fase (intervalo bilateral simétrico). **Corrección ADITIVA**

$$t \,/\, E_{F_0}\left[I_{\{|\theta(F_1)-\theta(F_0)|\le \hat{t}_{01}(F_1,F_2)+t\}} - (1-\alpha) \right] = 0 \rightarrow$$

corrector aditivo de la semiamplitud boot.

$$\rightarrow\ t = (1-\alpha)\,cuantil\ de\ \left[|\theta(F_1)-\theta(F_0)| - \hat{t}_{01}(F_1,F_2) \right]\ bajo\ F_0$$

$$\hat{t}_{02} \,/\, E_{F_0}\left[I_{\{|\theta(F_1)-\theta(F_0)|\le \hat{t}_{01}(F_1,F_2)+\hat{t}_{02}\}} - (1-\alpha) \right] = 0 \rightarrow$$

$$\rightarrow\ \hat{t}_{02} = \hat{t}_{02}(F_1,F_2,F_3) = (1-\alpha)\,cuantil\ de\ \left[|\theta(F_2)-\theta(F_1)| - \hat{t}_{01}(F_2,F_3) \right]\ bajo\ F_1$$

$$\left(\theta(F_1) - \left(\hat{t}_{01} + \hat{t}_{02} \right), \theta(F_1) + \left(\hat{t}_{01} + \hat{t}_{02} \right) \right)$$

Intervalo de confianza simétrico bootstrap de $\theta(F_0)$ ***con una corrección aditiva de su semiamplitud, a su vez estimada por bootstrap***

Ejemplo:

$$\theta(F_0) = \mu;\quad \theta(F_1) = \bar{X};\quad \theta(F_2) = \bar{X}^*;\quad \theta(F_3) = \bar{X}^{**}$$

Aproximación Monte Carlo de

$$\hat{t}_{02} = (1-\alpha)\,cuantil\ de\ \left[|\theta(F_2)-\theta(F_1)| - \hat{t}_{01}(F_2,F_3) \right]\ bajo\ F_1$$

$$\vec{X} \rightarrow \begin{cases} \vec{X}_1^* \rightarrow \begin{cases} \vec{X}_{11}^{**} \rightarrow \left| \bar{X}^* - \bar{X}^{**} \right|_{11} \\ \cdots \\ \vec{X}_{1C}^{**} \rightarrow \left| \bar{X}^* - \bar{X}^{**} \right|_{1C} \end{cases} \\ \vec{X}_b^* \rightarrow \left| \bar{X} - \bar{X}^* \right|_b \\ \vec{X}_B^* \rightarrow \begin{cases} \vec{X}_{B1}^{**} \rightarrow \left| \bar{X}^* - \bar{X}^{**} \right|_{B1} \\ \cdots \\ \vec{X}_{BC}^{**} \rightarrow \left| \bar{X}^* - \bar{X}^{**} \right|_{BC} \end{cases} \end{cases}$$

$\left| \bar{X}^* - \bar{X}^{**} \right|_{1c}$ $\quad c = 1,\ldots,C$

$\simeq \hat{t}_{01}(F_2,F_3)_1$

$\left| \bar{X} - \bar{X}^* \right|_b - \hat{t}_{01}(F_2,F_3)_b$ $\quad b = 1,\ldots,B$

$\widehat{t_{02}(F_1,F_2,F_3)}^{MC}$

$\left| \bar{X}^* - \bar{X}^{**} \right|_{Bc}$ $\quad c = 1,\ldots,C$

$\simeq \hat{t}_{01}(F_2,F_3)_B$

$$\left(\theta(F_1) - \left(\widehat{\hat{t}_{01}}^{MC} + \widehat{\hat{t}_{02}}^{MC} \right), \theta(F_1) + \left(\widehat{\hat{t}_{01}}^{MC} + \widehat{\hat{t}_{02}}^{MC} \right) \right)$$

Segunda fase (intervalo bilateral simétrico). **Corrección por RECUBRIMIENTO**

En la primera fase construimos

$$\left(\theta(F_1) \mp \hat{t}_{01}(F_1, F_2)_{(1-\alpha)}\right),$$

siendo

$$\hat{t}_{01}(F_1, F_2)_{(1-\alpha)} = (1-\alpha)\,cuantil\ de\ \left|\theta(F_2) - \theta(F_1)\right|.$$

Notemos que

$$\pi(1-\alpha) = P_{F_0}\left[\theta(F_0) \in \left(\theta(F_1) \mp \hat{t}_{01}(F_1, F_2)_{(1-\alpha)}\right)\right] \quad (\simeq 1-\alpha)$$

se puede estimar mediante

$$\hat{\pi}(1-\alpha) = P_{F_1}\left[\theta(F_1) \in \left(\theta(F_2) \mp \hat{t}_{01}(F_2, F_3)_{(1-\alpha)}\right)\right].$$

La **corrección por recubrimiento** consiste en estimar $\hat{\pi}(1-\alpha)$ para diferentes valores de α, y seleccionar aquél, $\hat{\alpha}$, para el cual $\hat{\pi}(1-\hat{\alpha}) = 1-\alpha$.

Ejemplo:

$$\theta(F_0) = \mu; \quad \theta(F_1) = \bar{X}; \quad \theta(F_2) = \bar{X}^*; \quad \theta(F_3) = \bar{X}^{**}$$

Aproximación Monte Carlo de

$$\hat{\pi}(1-\alpha) = P_{F_1}\left[\theta(F_1) \in \left(\theta(F_2) \mp \hat{t}_{01}(F_2, F_3)_{(1-\alpha)}\right)\right]$$

$$\vec{X} \rightarrow \left|\begin{array}{l} \vec{X}_1^* \\ \vec{X}_b^* \rightarrow \left|\begin{array}{l} \vec{X}_{b1}^{**} \\ \vec{X}_{bc}^{**} \\ \vec{X}_{bC}^{**} \end{array}\right. \rightarrow \left|\bar{X}^* - \bar{X}^{**}\right|_{bc} \\ \vec{X}_B^* \end{array}\right. \qquad c = 1, \ldots, C$$

$$\hat{t}_{01}(F_2, F_3)_{1-\alpha}^{b} = \left|\bar{X}^* - \bar{X}^{**}\right|_{b([C(1-\alpha)])} \qquad b = 1, \ldots, B$$

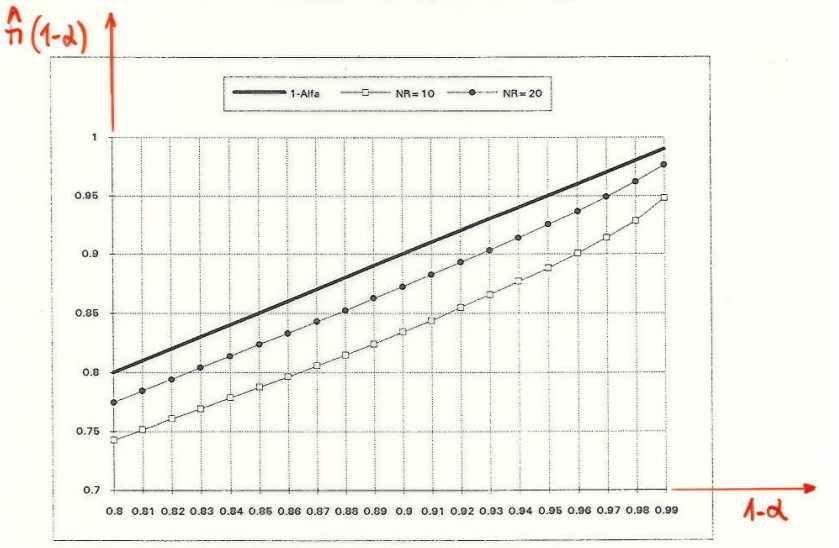
$$\widehat{\hat{\pi}(1-\alpha)}^{MC} = \frac{1}{B}\sum_{b=1}^{B} I_{\left\{\bar{X} \in \left(\bar{X}_b^* \mp \hat{t}_{01}(F_2, F_3)_{1-\alpha}^{b}\right)\right\}}$$

Si $\hat{\alpha}$ es tal que $\hat{\pi}(1-\hat{\alpha}) = 1-\alpha \rightarrow$

$$\left(\theta(F_1) \mp \hat{t}_{01}(F_1, F_2)_{(1-\hat{\alpha})}\right)$$

(Semiamplitud boot corregida por recubrimiento)

$1-\alpha$	$n=10$	$n=20$
0.80	0.742	0.774
0.81	0.751	0.784
0.82	0.760	0.794
0.83	0.769	0.804
0.84	0.778	0.813
0.85	0.787	0.823
0.86	0.796	0.833
0.87	0.805	0.842
0.88	0.814	0.852
0.89	0.824	0.862
0.90	0.834	0.872
0.91	0.844	0.882
0.92	0.854	0.893
0.93	0.865	0.903
0.94	0.876	0.914
0.95	0.888	0.925
0.96	0.900	0.936
0.97	0.913	0.948
0.98	0.928	0.962
0.99	0.948	0.976



Estimación de $\hat{\pi}(1-\alpha)$ en la corrección por recubrimiento, para una muestra particular.

$$X \in N(2,5);\ \theta(F_0) = \mu.$$

OBSERVACIÓN

En este caso, para cubrir 0,95 con una muestra de tamaño 10, hay que "buscar" cobertura 0,99; con una muestra de tamaño 20, 0,97. El bootstrap siempre infraestima.

TABLE IV

True coverage of nominal 95% iterated bootstrap confidence intervals for the mean of $N(2,5)$ estimated from samples of size 10 using nonparametric (*NP*), symmetrized nonparametric (*S*) or parametric (*P*) resampling with various resampling volumes *B* (average of 500 basic samples). For each resampling method, the coverage of the initial boostrap interval is shown together with those of intervals estimated by applying additive (*add*) or confidence level (*lev*) bootstrap corrections.

B	*NP*	*NP* (*add*)	*NP* (*lev*)	*S*	*S* (*add*)	*S* (*lev*))	*P*	*P* (*add*)	*P* (*lev*)
100	0.890	0.926	0.944	0.888	0.938	0.944	0.912	0.936	0.946
200	0.900	0.938	0.950	0.898	0.936	0.942	0.912	0.932	0.946
300	0.898	0.942	0.944	0.896	0.940	0.954	0.916	0.934	0.950
400	0.906	0.948	0.950	0.898	0.938	0.950	0.916	0.934	0.942
500	0.900	0.938	0.954	0.900	0.930	0.948	0.916	0.930	0.940

TABLE V

As for Table IV, but for a $2\chi_1^2$ population.

B	*NP*	*NP* (*add*)	*NP* (*lev*))	*S*	*S* (*add*))	*S* (*lev*))	*P*	*P* (*add*)	*P* (*lev*)
100	0.848	0.896	0.896	0.848	0.870	0.878	0.902	0.942	0.946
200	0.844	0.898	0.892	0.850	0.882	0.898	0.906	0.950	0.958
300	0.850	0.908	0.906	0.846	0.880	0.888	0.908	0.950	0.970
400	0.844	0.900	0.898	0.852	0.884	0.892	0.914	0.950	0.964
500	0.850	0.912	0.910	0.846	0.884	0.894	0.908	0.954	0.966

ESTUDIO DE SIMULACIÓN CORRECCIÓN COBERTURA I.C. OBSERVACIONES

- La mejora en la reducción del error de cobertura con la corrección es siempre clara (mejor recubrimiento (lev) que aditiva (add) en el caso normal, y equiparable en el caso $2\chi_1^2$).
- Una sóla corrección resulta suficiente en el caso normal así como, con bootstrap paramétrico (P), en el caso $2\chi_1^2$.
- En general, existe poca sensibilidad a B, por lo que es aconsejable tomar valores pequeños.
- En el caso normal, las correcciones tratan por igual a las tres metodologías. Sin corregir, los bootstraps no paramétrico (NP) y simetrizado (S) funcionan peor que el paramétrico (P) por la menor información utilizada.
- En el caso $2\chi_1^2$, excluyendo el bootstrap simetrizado (S) por inadecuado (aunque está en la línea del no paramétrico), es clara la mejora del paramétrico sobre el no paramétrico por la mayor información utilizada.

TABLE VII

True coverage, mean width and standard deviation of width of nominal 95% confidence intervals for the mean of $N(2,5)$ and $2\chi_1^2$ estimated from samples of size 10, using nonparametric (*NP*), symmetrized nonparametric (*S*) or parametric (*P*), by the iterated bootstrap methods (*add* and *lev*), the acceletared bias-corrected (*abc*) or bias-corrected (*bc*) percentile methods, and the bootstrap-t method (*b-t*). Coverages and widths are means of the results for 500 basic samples with resampling volume B=500.

	$N(2,5)$			$2\chi_1^2$		
	coverage	width	sd width	coverage	width	sd width
NP	0.900	2.574	0.577	0.850	2.745	1.419
add NP	0.938	3.133	0.723	0.912	3.786	2.188
lev NP	0.954	3.309	0.772	0.912	3.836	2.169
abc NP	0.906	2.614	0.579	0.876	2.999	1.667
b-t NP	0.952	3.381	0.819	0.930	6.357	7.048
S	0.900	2.584	0.575	0.846	2.825	1.490
add S	0.930	3.053	0.713	0.884	3.189	1.586
lev S	0.946	3.216	0.768	0.894	3.371	1.707
abc S	0.892	2.590	0.577	0.846	2.841	1.501
b-t S	0.946	3.187	0.737	0.902	3.304	1.620
P	0.916	2.719	0.609	0.908	3.477	1.403
add P	0.930	3.052	0.709	0.954	4.604	1.850
lev P	0.942	3.172	0.758	0.968	5.585	2.241
bc P	0.908	2.731	0.613	0.940	3.540	1.424
b-t P	0.932	2.977	0.675	0.968	6.025	3.181

OBSERVACIÓN. El bootstrap t (*b-t*), que al no requerir iteración de la metodología tiene menor coste computacional, se comporta como las correcciones consideradas.