

Un ejemplo de bootstrap suavizado

ANDRÉS M. ALONSO
Universidad Autónoma de Madrid, España

RESUMEN. En este trabajo se extienden resultados clásicos de normalidad asintótica en estadísticos definidos mediante funcionales diferenciables que satisfacen una condición local de Lipschitz y de validez del bootstrap utilizando la función de distribución empírica suavizada uniforme. En un estudio de simulación, se ilustra el adecuado comportamiento en muestras finitas del procedimiento bootstrap propuesto.

Key words and phrases. bootstrap, smoothed bootstrap, differentiable functionals.

2000 Mathematics Subject Classification. Primary 62F40. Secondary 62G09.

ABSTRACT. In this paper, we extend some classical results on the asymptotic normality for statistics defined by differentiable functionals satisfying a local Lipschitz condition. Also, we extend the validity of the bootstrap using a uniform smoothed empirical distribution function. A simulation study illustrate the correct performance in finite sample of the proposed bootstrap procedure.

1. Introducción

En muchos procedimientos estadísticos es necesario conocer determinadas características de la distribución muestral de los estadísticos o los estimadores empleados. Así, por ejemplo, en el contraste de hipótesis se necesitan los percentiles de la distribución muestral del estadístico de contraste, mientras que en problemas de estimación es esencial tener alguna medida de la exactitud (por ejemplo, el sesgo, la varianza o el error cuadrático medio) del estimador obtenido.

Un enfoque clásico para obtener medidas de exactitud de un estimador es calcularlas mediante análogos empíricos de las fórmulas explícitas obtenidas bajo un modelo determinado. Sin embargo, en la mayoría de los estadísticos es muy difícil o imposible obtener fórmulas exactas y explícitas de las medidas de exactitud. Los métodos de remuestreo reemplazan las derivaciones teóricas del enfoque clásico por la evaluación de los estadísticos en remuestras obtenidas a partir de los datos originales, y mediante estos valores se obtienen estimadores de las medidas de exactitud o de la distribución muestral del estadístico. Los métodos de remuestreo más populares en la literatura estadística son el *jackknife* de QUENOUVILLE (1949) y TUKEY (1958), y el *bootstrap* de EFRON (1979).

Una característica básica del método bootstrap es el principio *plug-in*, que puede interpretarse como la sustitución de la distribución subyacente F por un estimador \hat{F} de ésta. La selección más frecuente ha sido la función de distribución empírica $\hat{F}(x) = F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i \leq x\}}$, donde $I_{\{X_i \leq x\}} = 1$ si $X_i \leq x$ y 0 en otro caso (ver en EFRON (1979) y EFRON & TIBSHIRANI (1993)). Otra alternativa que ha sido empleada cuando F es absolutamente continua con densidad f es utilizar una \hat{F}_h asociada a un estimador de tipo kernel de la densidad (ver, por ejemplo, CUEVAS & ROMO (1997)).

En este trabajo, proponemos la utilización de otro estimador natural de F para el caso continuo, la distribución empírica suavizada uniforme

\hat{F}_n , definida por:

$$\hat{F}_n(x) = \begin{cases} 0 & \text{si } x < x_{(1)} \\ \frac{x-x_{(k)}}{n(x_{(k+1)}-x_{(k)})} + \frac{k}{n} & \text{si } x_{(k)} \leq x < x_{(k+1)} \\ 1 & \text{si } x \geq x_{(n)}, \end{cases} \quad (1)$$

donde $x_{(i)}$ es la i -ésima observación de la muestra ordenada, con $i \in \{1, \dots, n\}$. Recordemos que $F_n(x_{(i)}) = \frac{i}{n}$, por tanto $\hat{F}_n(x)$ es simplemente una interpolación lineal (superior) de F_n . Otro estimador análogo es:

$$\check{F}_n(x) = \begin{cases} 0 & \text{si } x < x_{(1)} \\ \frac{x-x_{(k)}}{n(x_{(k+1)}-x_{(k)})} + \frac{k-1}{n} & \text{si } x_{(k)} \leq x < x_{(k+1)} \\ 1 & \text{si } x \geq x_{(n)} \end{cases} \quad (2)$$

Notemos que $\|\hat{F}_n - F_n\| = \|\check{F}_n - F_n\| = \frac{1}{n}$, donde $\|\cdot\|$ denota la norma del supremo, y esta relación es válida para toda muestra de tamaño n . Dada esta propiedad intuimos que podremos extender a \hat{F}_n y a \check{F}_n , resultados como el teorema de Glivenko-Cantelli, la desigualdad de Dvoretzky-Kiefer-Wolfowitz, la normalidad asintótica de estadísticos basados en funcionales diferenciables, y la validez del bootstrap.

El contenido de este trabajo se divide en las siguientes secciones: en la Sección 2 hacemos una breve presentación del método bootstrap; en la Sección 3 extendemos los resultados clásicos arriba mencionados a \hat{F}_n , y en la Sección 4 realizaremos un ejercicio de simulación en el que ejemplificaremos la similitud de resultados entre el bootstrap standard y el bootstrap suavizado propuesto.

2. El método bootstrap

El bootstrap de EFRON (1979) tiene como pieza clave la utilización “extrema” del principio de analogía (*plug-in*) que constituye uno de los métodos más simples utilizados para obtener un estimador de un parámetro poblacional $\theta = T(F)$, donde T es un funcional definido en

una clase convexa de funciones de distribución y F es la distribución subyacente de los datos. Un estimador plug-in o análogo es $\hat{\theta} = \theta(\hat{F})$, donde \hat{F} es un estimador de P . En lo que sigue combinamos el modo en que EFRON y TIBSHIRANI (1986), y SHAO & TU (1995) presentan el método bootstrap en una situación general:

Sea $\mathbf{X} = (X_1, X_2, \dots, X_n)$ un conjunto de datos generados por el modelo P , y sea $T(\mathbf{X})$ el estadístico cuya distribución $\mathcal{L}(T; \mathcal{P})$ deseamos estimar. El método bootstrap propone como estimador de $\mathcal{L}(T; \mathcal{P})$ la distribución $\mathcal{L}^*(T^*; \hat{\mathcal{P}})$ del estadístico $T^* = T(\mathbf{X}^*)$, donde \mathbf{X}^* es un conjunto de datos generado por el modelo estimado \hat{P} .

Concentrándonos en el caso de datos independientes e idénticamente distribuidos donde el modelo P es la función de distribución F que sigue X_i , la forma en que hemos presentado el bootstrap incluye el *bootstrap clásico* que corresponde a estimar F por la función de distribución empírica $\hat{F} = F_n$, el *bootstrap suavizado* para F continua que utiliza $\hat{F} = \hat{F}_h$ donde \hat{F}_h es la función de distribución asociada a un estimador de la función de densidad f , y el *bootstrap paramétrico* cuando se supone que la función de distribución pertenece a una familia paramétrica F_ϑ y se utiliza $\hat{F} = F_{\hat{\vartheta}}$, donde $\hat{\vartheta}$ es un estimador de ϑ .

Un punto importante a considerar es el cálculo de los estimadores bootstrap, ya que tenemos $\binom{n+m-1}{m-1}$ muestras distintas en el bootstrap estándar, e infinitas en el bootstrap suavizado o en el paramétrico. En general no se tienen fórmulas explícitas sobre los estimadores bootstrap, y el problema se aborda mediante (i) soluciones analíticas o aproximaciones como el método delta, aproximaciones jackknife, ó el método de punto de silla (ver resultados y referencias en el Capítulo 9 de DAVISON e HINKLEY (1997)), ó (ii) aproximaciones de Monte Carlo, cuya versión más simple mostramos a continuación:

Sea \hat{P} el modelo estimado utilizando los datos $\mathbf{X} = (X_1, X_2, \dots, X_n)$, y queremos obtener aproximaciones de los estimadores bootstrap del sesgo, la varianza

y la distribución muestral del estadístico $T(\mathbf{X})$ dados por:

$$b_{Boot} = E^*[T(\mathbf{X}^*)] - T(\mathbf{X}) \quad (3)$$

$$v_{Boot} = \text{Var}^*[T(\mathbf{X}^*)] \quad (4)$$

$$H_{Boot}(x) = \text{Pr}^*\{T(\mathbf{X}^*) \leq x\}, \quad (5)$$

donde \mathbf{X}^* es una muestra obtenida de \hat{P} , y E^* , Var^* y Pr^* denotan la esperanza, la varianza y la probabilidad bootstrap dada la muestra \mathbf{X} . Un procedimiento de Monte Carlo para estimar (3) - (5) es:

1. Generamos B muestras independientes $\mathbf{X}^{*(b)}$ a partir de \hat{P} , con $b = 1, 2, \dots, B$.
2. Calculamos $T^{*(b)} = T(\mathbf{X}^{*(b)})$ para $b = 1, 2, \dots, B$, y aproximamos b_{Boot} , v_{Boot} , y $H_{Boot}(x)$ por:

$$b_{Boot}^{(B)} = \frac{1}{B} \sum_{b=1}^B T^{*(b)} - T(\mathbf{X}) \quad (6)$$

$$v_{Boot}^{(B)} = \frac{1}{B} \sum_{b=1}^B \left(T^{*(b)} - \frac{1}{B} \sum_{b=1}^B T^{*(b)} \right)^2 \quad (7)$$

$$H_{Boot}^{(B)}(x) = \frac{1}{B} \sum_{b=1}^B I(T^{*(b)} \leq x). \quad (8)$$

La selección de B no es un problema sencillo, EFRON & TIBSHIRANI (1986) sugieren utilizar B entre 50 y 200 para estimadores de momentos, y al menos 1000 remuestras para estimadores de distribuciones o cuantiles.

3. Resultados clásicos extendidos a \hat{F}_n

En esta sección extenderemos resultados clásicos de la función de distribución empírica F_n (F_n^* , en el caso del bootstrap), para \hat{F}_n (\hat{F}_n^*).

Proposición 3.1. *Sea \hat{F}_n la función de distribución empírica suavizada uniforme definida por (1), obtenida de una muestra aleatoria simple*

de tamaño n de una variable aleatoria con función de distribución F . Entonces, se tiene que $\|\hat{F}_n - F\| \xrightarrow{cs} 0$.

Demostración: Tenemos que $\|\hat{F}_n - F\| = \|\hat{F}_n - F_n + F_n - F\| \leq \|\hat{F}_n - F_n\| + \|F_n - F\|$, y que $\|\hat{F}_n - F_n\| = \frac{1}{n}$. Por lo tanto, $\{\omega \in \Omega : \lim_{n \rightarrow \infty} \|F_n - F\| = 0\} = \{\omega \in \Omega : \lim_{n \rightarrow \infty} (\|F_n - F\| + \|\hat{F}_n - F_n\|) = 0\} \subset \{\omega \in \Omega : \lim_{n \rightarrow \infty} \|\hat{F}_n - F\| = 0\}$. Esto implica que $\Pr\{\omega \in \Omega : \lim_{n \rightarrow \infty} \|\hat{F}_n - F\| = 0\} \geq \Pr\{\omega \in \Omega : \lim_{n \rightarrow \infty} \|F_n - F\| = 0\} = 1$, por el teorema de Glivenko-Cantelli. \square

Notemos que esta proposición es válida para \tilde{F}_n , y en general para cualquier estimador \hat{F} que satisfaga $\|\hat{F} - F_n\| \leq a_n$ con $\{a_n\}_{n \in \mathbb{N}}$ una sucesión numérica tal que $\lim_{n \rightarrow \infty} a_n = 0$. Por otra parte, es evidente que al imponer esta condición tenemos que $\|\hat{F} - F_n\| \xrightarrow{cs} 0$.

Proposición 3.2. Sean \hat{F}_n y F como en la Proposición 2.1, entonces tenemos que $\Pr\{\omega \in \Omega : \sqrt{n}\|\hat{F}_n - F\| > t\} \leq 2e^{-2(t - \frac{1}{\sqrt{n}})^2}$, para todo $t \geq \frac{1}{\sqrt{n}}$.

Demostración: Tenemos que $\sqrt{n}\|\hat{F}_n - F\| \leq \sqrt{n}\|F_n - F\| + \frac{1}{\sqrt{n}}$. Por lo tanto, $\Pr\{\omega \in \Omega : \sqrt{n}\|\hat{F}_n - F\| > t\} \leq \Pr\{\omega \in \Omega : \sqrt{n}\|F_n - F\| > t - \frac{1}{\sqrt{n}}\}$, y si $t \geq \frac{1}{\sqrt{n}}$ aplicamos la desigualdad de Dvoretzky-Kiefer-Wolfowitz. \square

Notemos que esta proposición es válida para cualquier estimador \hat{F} que satisfaga $\|\hat{F} - F_n\| \leq a_n$, siendo entonces la cota $2e^{-2(t - \sqrt{na_n})^2}$ y $t \geq \sqrt{na_n}$.

Antes de obtener un resultado análogo a la normalidad asintótica de los estadísticos basados en funcionales diferenciables, introduzcamos la clase de funcionales sobre la cual probaremos este resultado.

Definición 3.3. Decimos que un funcional $T : \mathcal{F} \rightarrow \mathbb{R}$, definido sobre una clase convexa de distribuciones \mathcal{F} , satisface una condición de Lipschitz si existe $k > 0$ tal que $|T(G) - T(F)| \leq k\|G - F\|$ para todas $G, F \in \mathcal{F}$.

Definición 3.4. Decimos que un funcional $T : \mathcal{F} \rightarrow \mathbb{R}$ satisface una condición de Lipschitz localmente si existen k y $\varepsilon > 0$ tales que $|T(G) - T(F)| \leq k\|G - F\|$ para todas $G, F \in \mathcal{F}$ que verifican $\|G - F\| \leq \varepsilon$.

Notemos que si T satisface una condición de Lipschitz localmente con k y ε , y como $\exists n_0 > 0$ tal que $\frac{1}{n} < \varepsilon$ para todo $n > n_0$, obtenemos que $|T(\hat{F}_n) - T(F_n)| \leq k\|\hat{F}_n - F_n\| = \frac{k}{n}$, para todo $n > n_0$. Con esta desigualdad es inmediato el siguiente lema.

Lema 3.5. Sea $T : \mathcal{F} \rightarrow \mathbb{R}$ un funcional que satisface una condición de Lipschitz localmente, entonces $\sqrt{n}(T(\hat{F}_n) - T(F_n)) \xrightarrow{p} 0$.

Demostración: Si fijamos $\delta > 0$, existe un $n'_0 > 0$ tal que $\frac{k}{\sqrt{n}} < \delta$ para todo $n > n'_0$. Aplicando la desigualdad anterior, tenemos que $\sqrt{n}|T(\hat{F}_n) - T(F_n)| \leq \frac{k}{\sqrt{n}} < \delta$, para todo $n > \max(n_0, n'_0)$ y para toda muestra de tamaño n . Por lo tanto, $\lim_{n \rightarrow \infty} \Pr\{\omega \in \Omega : \sqrt{n}|T(\hat{F}_n) - T(F_n)| < \delta\} = 1$. \square

El lema anterior es válido para cualquier estimador \hat{F} que satisfaga $\|\hat{F} - F_n\| \leq a_n$ con $\lim_{n \rightarrow \infty} \sqrt{n}a_n = 0$, bastará tomar n_0 tal que $a_n < \varepsilon$ y n'_0 tal que $k\sqrt{n}a_n < \delta$, para todo $n > n_0$ y $n > n'_0$, respectivamente.

Proposición 3.6. Sea $T : \mathcal{F} \rightarrow \mathbb{R}$ un funcional diferenciable en $F \in \mathcal{F}$, que satisface una condición de Lipschitz localmente. Si

$$\sigma^2 = \int_{\mathbb{R}} T'(F; x)^2 dF(x) < +\infty,$$

entonces $\sqrt{n}(T(\hat{F}_n) - T(F)) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$.

Demostración: Bajo las condiciones de diferenciability y σ^2 finito, tenemos que $\sqrt{n}(T(F_n) - T(F)) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$, y por el Lema 2.5 tenemos $\sqrt{n}(T(\hat{F}_n) - T(F_n)) \xrightarrow{p} 0$. Bastará aplicar el teorema de Slutsky

en $\sqrt{n}(T(\hat{F}_n) - T(F_n) + T(F_n) - T(F))$. \checkmark

Para demostrar la validez del bootstrap utilizando \hat{F}_n , necesitaremos obtener una relación entre F_n^* , función de distribución empírica de una muestra artificial o remuestra tomada de F_n , y \hat{F}_n^* , función de distribución empírica de una muestra artificial tomada de \hat{F}_n asociada a la anterior. Notemos que una remuestra de \hat{F}_n puede interpretarse como que primero se toma una remuestra de F_n , y después por cada $x_{(k)}^*$ tomamos un valor en $(x_{(k-1)}^*, x_{(k)}^*]$ de manera uniforme siendo $x_{(0)}^* = x_{(1)}^*$ (Si remuestreamos de \hat{F}_n tomaremos un valor en $[x_{(k)}^*, x_{(k+1)}^*)$ siendo $x_{(n+1)}^* = x_{(n)}^*$).

Lema 3.7. (a) $\|\hat{F}_n^* - F_n^*\| \xrightarrow{p} 0$ (b) $\sqrt{n}\|\hat{F}_n^* - F_n^*\| \xrightarrow{p} 0$.

Demostración: (a) Notemos que \hat{F}_n^* y F_n^* difieren en los distintos subintervalos en a lo sumo cantidades de la forma $\frac{j}{n}$. Fijamos $0 < \varepsilon < 1$, y sea \tilde{j} el primer valor tal que $\frac{\tilde{j}}{n} \geq \varepsilon$, o simplemente $\tilde{j} = \lceil n\varepsilon \rceil$.

- La probabilidad de que el valor k -ésimo sea seleccionado exactamente j veces en una remuestra es: $\Pr\{E_k^j\} = \binom{n}{j} (\frac{1}{n})^j (1 - \frac{1}{n})^{n-j}$, y por lo tanto la probabilidad de que sea seleccionado \tilde{j} ó más veces es: $\Pr\{\cup_{j=\tilde{j}}^n E_k^j\} = \sum_{j=\tilde{j}}^n \binom{n}{j} (\frac{1}{n})^j (1 - \frac{1}{n})^{n-j}$, que podemos acotar por $\sum_{j=\tilde{j}}^n \frac{1}{j!}$.
- $\Pr\{\|\hat{F}_n^* - F_n^*\| \geq \varepsilon\} = \Pr\{\cup_{k=1}^n \cup_{j=\tilde{j}}^n E_k^j\} \leq \sum_{k=1}^n \Pr\{\cup_{j=\tilde{j}}^n E_k^j\} \leq n \sum_{j=\tilde{j}}^n \frac{1}{j!}$.
- Notemos que $\frac{n}{\lceil n\varepsilon \rceil}$ converge a $\frac{1}{\varepsilon}$, por lo tanto, para n suficientemente grande será menor que $\frac{2}{\varepsilon}$, y también lo serán los términos de la forma $\frac{n}{j}$ con $j \geq \lceil n\varepsilon \rceil$.
- Finalmente, $n \sum_{j=\tilde{j}}^n \frac{1}{j!} < \frac{2}{\varepsilon} \sum_{j=\tilde{j}}^n \frac{1}{(j-1)!}$, que converge a cero por ser “resto” de la serie convergente $\sum_{j=1}^{\infty} \frac{1}{(j-1)!}$. Por lo tanto, el límite $\lim_{n \rightarrow \infty} \Pr\{\|\hat{F}_n^* - F_n^*\| \geq \varepsilon\} = 0$.

(b) La prueba es básicamente como en (a), basta tomar \tilde{j} como el primer valor tal que $\frac{\tilde{j}}{n} \geq \frac{\varepsilon}{\sqrt{n}}$, ó $\tilde{j} = \lceil \sqrt{n}\varepsilon \rceil$, y en el tercer punto utilizar que $\frac{n}{\lceil \sqrt{n}\varepsilon \rceil (\lceil \sqrt{n}\varepsilon \rceil - 1)}$ converge a $\frac{1}{\varepsilon^2}$. \square

Lema 3.8. *Sea $T : \mathcal{F} \rightarrow \mathbb{R}$ un funcional que satisface una condición de Lipschitz, entonces $\sqrt{n}(T(\hat{F}_n^*) - T(F_n^*)) \xrightarrow{p} 0$.*

Demostración: Tenemos que si

$$\sqrt{n}\|\hat{F}_n^* - F_n^*\| < \frac{\varepsilon}{k},$$

entonces $\sqrt{n}|T(\hat{F}_n^*) - T(F_n^*)| \leq \varepsilon$, por lo tanto

$$\Pr\{\sqrt{n}|T(\hat{F}_n^*) - T(F_n^*)| \leq \varepsilon\} \geq \Pr\{\sqrt{n}\|\hat{F}_n^* - F_n^*\| \leq \frac{\varepsilon}{k}\},$$

que converge a 1, por el Lema 2.7. \square

A diferencia del Lema 2.5, aquí pedimos que el funcional T satisfaga una condición de Lipschitz global, que es más restrictiva. Notemos que en lugar de $\sqrt{n}\|\hat{F}_n - F_n\| \rightarrow 0$, solo tenemos que $\sqrt{n}\|\hat{F}_n^* - F_n^*\| \xrightarrow{p} 0$.

Proposición 3.9. *Sea $T : \mathcal{F} \rightarrow \mathbb{R}$ un funcional diferenciable en $F \in \mathcal{F}$ que satisface una condición de Lipschitz. Si $\sigma^2 = \int_{\mathbb{R}} T'(F; x)^2 dF(x) < +\infty$, entonces $\sqrt{n}(T(\hat{F}_n^*) - T(\hat{F}_n)) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$.*

Demostración: Bajo las condiciones de diferenciabilidad y σ^2 finito, tenemos que $\sqrt{n}(T(F_n^*) - T(F_n)) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$, por el Lema 2.5 tenemos que $\sqrt{n}(T(\hat{F}_n) - T(F_n)) \xrightarrow{p} 0$, y por el Lema 2.7 (b), tenemos que $\sqrt{n}(T(\hat{F}_n^*) - T(F_n^*)) \xrightarrow{p} 0$. Bastará aplicar el teorema de Slutsky en $\sqrt{n}(T(\hat{F}_n^*) - T(F_n^*) + T(F_n^*) - T(F_n) + T(F_n) - T(\hat{F}_n))$. \square

Esta proposición establece la validez del bootstrap en funcionales diferenciables utilizando remuestras de \hat{F}_n . Podemos extender los resultados del Lema 2.7, y por consiguiente el Lema 2.8 y la Proposición 2.9, a cualquier estimador \hat{F} que satisfaga $\|\hat{F} - F_n\| \leq a_n$ con $na_n \leq 1$. Por otra parte, algunos estimadores de F , como \hat{F}_n y \tilde{F}_n , no son del tipo

kernel con $\int_{\mathbb{R}} tK(t)dt = 0$ que se pide en la condición K0 de CUEVAS & ROMO (1997).

4. Ejercicio de simulación

Con este ejercicio pretendemos ejemplificar la similitud de los resultados obtenidos mediante el bootstrap clásico y el bootstrap suavizado. Utilizaremos el estimador *media α -truncada*, definido por el funcional siguiente $T(F) = \int_{\alpha}^{1-\alpha} F^{-1}(t)dt/(1-2\alpha)$, que es diferenciable y continuo para toda $F \in \mathcal{F}$ (ver, por ejemplo Hampel et al. (1986)). En el Apéndice 1 se presentan los listados de las funciones MATLAB que utilizamos en la simulación. El ejercicio se efectúa según los siguientes pasos:

1. Se generan $N = 10000$ muestras de tamaño $n = 25, 50$ y 100 de una variable aleatoria con distribución $N(0,1)$. En cada una de las muestras se calcula la media α -truncada tomando $\alpha = 0.1$. Estos 10000 valores permiten obtener una estimación, mediante la función de distribución empírica F_n , de la función de distribución F del estadístico media α -truncada.
2. Se generan $N_1 = 100$ muestras de tamaño n , sobre las cuales aplicaremos bootstrap standard y bootstrap suavizado con \hat{F}_n calculando $B = 1000$ remuestras en cada muestra. En cada remuestra se calcula la media α -truncada. En la Figura 1, se presentan las funciones de distribución por ambos procedimientos que denotaremos $F_b^{(i)}$ y $F_s^{(i)}$ con $i \in \{1, 2, \dots, N_1\}$, respectivamente (en trazo continuo aparece la función de distribución obtenida en el paso 1).
3. Se estima la norma del supremo de $F - F_b^{(i)}$ y de $F - F_s^{(i)}$. Para estimar $\|F - F^{(i)}\|$, recordemos que tenemos B valores de x para $F^{(i)}(x)$ y N valores para $F(x)$, que no coinciden en general. De manera que interpolamos la función F para los valores de x correspondientes a $F^{(i)}$, con lo que tenemos B valores de $F^{(i)}$ y de F para los mismos valores x . A los N_1 vectores $B \times 1$ resultantes se

les calcula la norma del supremo. En la Tabla 1, se muestran los estadísticos media y desviación standard de estos valores.

TABLA 1. Media y desviación standard de $\|F - F^{(i)}\|$

| Tamaño muestral | Media (SD) $\ F - F_b^{(i)}\ $ | Media (SD) $\ F - F_s^{(i)}\ $ |
|-----------------|--------------------------------|--------------------------------|
| n = 25 | 0.3309 (0.0215) | 0.3441 (0.0208) |
| n = 50 | 0.3188 (0.0214) | 0.3242 (0.0227) |
| n = 100 | 0.3068 (0.0185) | 0.3053 (0.0188) |

Los resultados obtenidos, apuntan a que con el estadístico media α -truncada las estimaciones con bootstrap standard y suavizado son similares, particularmente las medias de $\|F - F_b^{(i)}\|$ y $\|F - F_s^{(i)}\|$ son prácticamente iguales en los tres tamaños muestrales. Por otra parte, como era de esperar las diferencias con F van disminuyendo con n (notemos la menor variabilidad de las funciones de distribución en la Figura 1).

Conclusiones

En este trabajo se extiende el teorema de Glivenko-Cantelli, la desigualdad de Dvoretzky-Kiefer-Wolfowitz, la normalidad asintótica y la validez del bootstrap en estadísticos basados en funcionales diferenciables tipo Lipschitz utilizando como estimador de la función de distribución F a la distribución empírica suavizada uniforme \hat{F}_n . \hat{F}_n es un estimador natural de F cuando ésta es absolutamente continua y no impone restricciones de existencia y acotación de las derivadas de la función de densidad f como en CUEVAS & ROMO (1997) ni necesita seleccionar parámetros de suavizado como en el caso de estimadores kernel. Se ilustra el correcto comportamiento en muestras finitas del procedimiento propuesto.

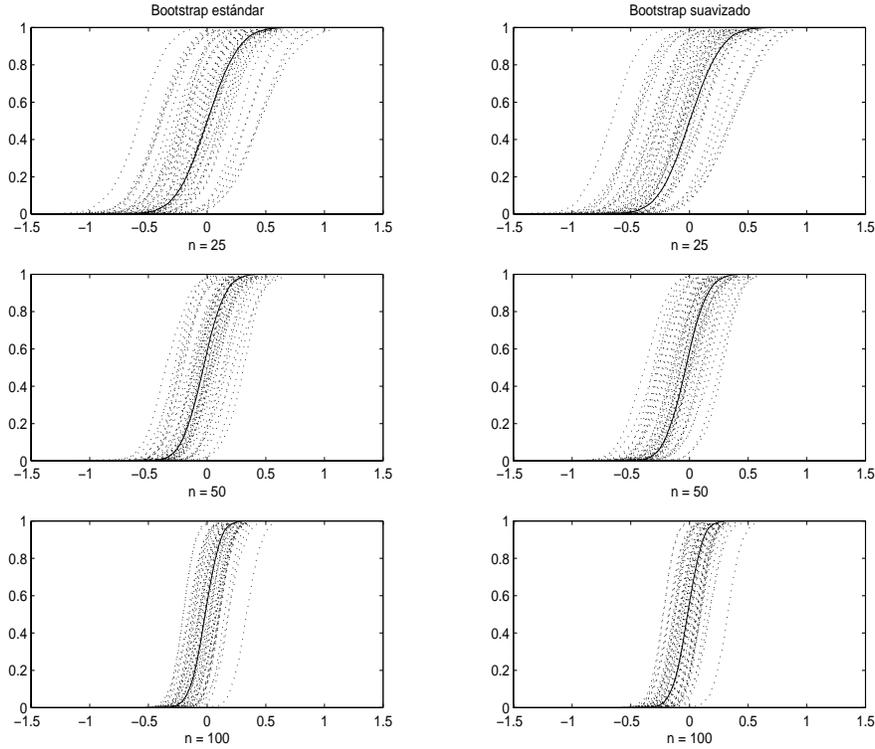


FIGURA 1. – Función de distribución F , -- Funciones de distribución $F_b^{(i)}$ y $F_s^{(i)}$.

Referencias

- [1] CUEVAS, A. & ROMO, J. *Differentiable functionals and smoothed bootstrap*, *Annals of the Institute of Statistical Mathematics* **49** (1997), 355–370.
- [2] DAVISON, A. C. & HINKLEY, D. V. *Bootstrap Methods and their Applications*, Cambridge University Press, Cambridge, 1997.
- [3] EFRON, B. *Bootstrap methods: Another look at the jackknife*, *The Annals of Statistics* **7** (1979), 1–26.
- [4] EFRON, B. & TIBSHIRANI, R. J. *Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy*, *Statistical Science* **1** (1986), 54–77.

- [5] EFRON, B. & TIBSHIRANI, R. J. *Introduction to the bootstrap*, Chapman & Hall, New York, 1993.
- [6] HAMPEL, F. AND RONCHETTI, E. AND ROUSSEEUW, P. & STAHEL, W. *Robust Statistics. The Approach based on Influence Functions*, Jonh Wiley & Sons, New York, 1986.
- [7] QUENOUILLE, M. *Approximation test of correlation in time series*, Journal of the Royal Statistical Society. Series B **11** (1949), 18–84.
- [8] SHAO, J. & TU, D. *The Jackknife and Bootstrap*, Springer–Verlag, New York, 1995.
- [9] TUKEY, J. *Bias and confidence in not quite large samples*, The Annals of Mathematical Statistics **29** (1958), 614.

Apéndice 1

Función alpha_trimmed.m

```
function R = alpha_trimmed(alpha, N, n, filename);

% Rutina para generar una aproximaci\on de la %
% funci\on de distribuci\on de la media alpha %
% truncada de una v.a. N(0,1).                %

x = sort(randn(n,N));
x_alpha = x(round(alpha*n):round(n-alpha*n),:);
clear x;
y_alpha = sort(sum(x_alpha)/(n*(1-2*alpha)));
clear x_alpha;
F = (1/N):(1/N):1;
save(filename,'y_alpha','F')
```

Función bootstrap.m

```
function R = bootstrap(x, n, B, alpha)

% Rutina para bootstrap standard en el %
% caso de media alpha-truncada.        %

y = zeros(n,B);
for i = 1:B
    orden = ceil(n*rand(n,1));
    y(:,i) = x(orden);
end

y = sort(y);
```

```
x_alpha = y(round(alpha*n):round(n - alpha*n),:);
clear y;
y_alpha = sort(sum(x_alpha)/(n*(1-2*alpha)));
clear x_alpha;
R = y_alpha;
```

Función smoothed_bootstrap.m

```
function R = smoothed_bootstrap(x, n, B, alpha)

% Rutina para bootstrap suavizado en el %
% caso de media alpha-truncada.      %

% Funci\on de distribuci\on          %
x = sort(x)';
x = [x(1)-eps,x];
F = 0:(1/n):1;

% Remuestreo sobre F                  %
y = zeros(n,B);
for i = 1:B
    uniforme = rand(n,1);
    y(:,i) = interp1(F,x,uniforme);
end

% C\alculo de media alpha-truncada.   %
y = sort(y);
x_alpha = y(round(alpha*n):round(n - alpha*n),:);
clear y;
y_alpha = sort(sum(x_alpha)/(n*(1-2*alpha)));
clear x_alpha;
R = y_alpha;
```

(Recibido en noviembre 2001. Versión revisada en febrero de 2002)

ANDRÉS M. ALONSO, DEPARTAMENTO DE MATEMÁTICAS
 UNIVERSIDAD AUTÓNOMA DE MADRID, MADRID 28049, ESPAÑA
e-mail: andres.alonso@uam.es