QÜESTIIÓ, Vol 15, 1. pp. 3-20, 1991

ESTIMACIÓN NO PARAMÉTRICA DE LA FUNCIÓN DE DISTRIBUCIÓN

JUAN MANUEL VILAR FERNÁNDEZ

Departamento de Matemáticas

Sea X una variable aleatoria con función de distribución $F(x)$ y función de densidad $f(x)$, y $X_1, X_2, \cdots, X_n$ un conjunto de observaciones de la variable que pueden ser dependientes. Se definen dos estimadores no paramétricos generales (uno recursivo y otro no recursivo) de la función de distribución.
Bajo condiciones aceptables se obtiene el sesgo y la varianza y covarianza asintótica de los estimadores definidos. Finalmente se prueban propiedades de consistencia y normalidad asintótica.

Non Parametric Estimation of Distribution Function.

Keywords: Función de distribución, estimadores no paramétricos y condiciones de dependencia.

Clasificación AMS (1980): 62G05, 62H12.

1. INTRODUCCIÓN. DEFINICIÓN DE LOS ESTIMADORES

Sea X una variable aleatoria real con función de distribución $F(x)$, la estimación de esta función y la obtención de propiedades asintóticas óptimas es de gran interés para conocer el modelo probabilístico que sigue la variable o para la estimación de funciones relacionadas con la primera tales como la razón de supervivencia, $\overline{F}(x) = 1 - F(x) = P(X > x)$, o la función razón de fallo, $r(x) = f(x)/\overline{F}(x)$, muy utilizadas en las ciencias biomédicas e ingeniería.

–Juan Manuel Vilar Fernández. Departamento de Matemáticas. Facultad de Informática. Campus de Zapateira s/n. 15071, La Coruña.

–Article rebut el desembre de 1990.

En este trabajo se estudia la estimación no paramétrica recursiva y no recursiva de $F(x)$ que permite por técnicas sencillas y de fácil cálculo tener una idea de la forma funcional de $F(x)$ a partir del conocimiento de una muestra de la variable aleatoria X, sin presuponer ninguna hipótesis sobre la función de distribución o, a lo sumo, hipótesis muy generales sobre la regularidad de dicha función que permiten asegurar propiedades asintóticas del estimador utilizado. Y si el tamaño muestral es suficientemente grande puede obtenerse información local que no recogen las técnicas clásicas de estimación paramétrica.

Sea $X_1, X_2, \cdots, X_n$ una muestra de la variable aleatoria real X, con función de densidad $f(x)$ y función de distribución $F(x)$, verificando una condición de dependencia débil, hipótesis mucho más realista que la asunción de observaciones independientes, que no se cumple en muchas situaciones, sobre todo, si las observaciones son recogidas en forma secuencial, esto es, si se está estudiando una serie de tiempo. Para estimar $F(x)$ normalmente se utiliza la función de distribución empírica, $F_n(x)$, definida por:

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} U_i \tag{1}$$

siendo

$$U_i = I_{(X_i \le x)} = \begin{cases} 1 & \text{si } X_i \le x \\ 0 & \text{si } X_i > x \end{cases}$$

que en cierto sentido es bastante suave y en un contexto de datos dependientes ha sido estudiado por Roussas (1989), Sarda-Vieu (1989) e Izerman-Tran (1990).

Por otra parte es conocido que la mayoría de los estimadores no paramétricos de la función de densidad, $f(x)$, siguen el formato general:

$$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \delta_{h(n)}(x, X_i) \tag{2}$$

siendo $\{\delta_{h(n)}(x,u)\}$ una sucesión de funciones de ponderación definidas en $\mathbf{R} \times \mathbf{R}$ con valores en $\mathbf{R}$ y $h(n)$ la sucesión de parámetros de suavización, que nos mide la cantidad de suavización que se introduce en la estimación.

Por todo ello puede suponerse que un buen estimador de $F(x)$ viene dado por:

(3) $$\hat{F}_n(x) = \int_{-\infty}^{x} \hat{f}_n(t)\,dt = \frac{1}{n}\sum_{i=1}^{n} \hat{\delta}_{h(n)}(x, X_i)$$

siendo

$$\hat{\delta}_{h(n)}(x,u) = \int_{-\infty}^{x} \delta_{h(n)}(u,t)\,dt$$

En particular, si $\delta_{h(n)}(x,u) = \frac{1}{h(n)} K\left(\frac{x-u}{h(n)}\right)$, siendo $K(u)$ una función de densidad se obtiene el estimador núcleo (kernel), uno de los más estudiados y utilizados, otras elecciones de $\delta_{h(n)}(x,u)$ permite obtener el histograma o el estimador por desarrollos ortogonales. (Una relación exhaustiva de distintos tipos de funciones $\delta_{h(n)}(x,u)$ puede verse en Walter-Blum (1974)).

El estimador asi definido proporciona una mayor suavización respecto a la que introduce el estimador empírico, lo que conlleva ventajas adicionales. En este sentido, Reiss (1981), en un contexto de datos independientes, demuestra la deficiencia relativa del estimador empírico $(F_n(x))$ respecto al estimador núcleo $(\hat{F}_n(x))$, lo que significa que si $i(n)$ es el tamaño muestral necesario para que el Error Cuadrático Medio (ECM) de $F_n(x)$ sea menor o igual que el ECM $(\hat{F}_n(x))$, entonces es posible elegir $h(n)$ tal que $i(n) - n \geq Cn^{4/5} \longrightarrow \infty$, siendo C una constante positiva que depende de F.

También es de interés el resultado obtenido recientemente por Lejeune-Sarda (1989), quienes trabajando con el estimador núcleo, esto es, $\hat{\delta}_{h(n)}(x,X_i) = \int_{-\infty}^{(x-X_i)/h(n)} K(u)\,du$, demuestran que este estimador es el mismo que el que se obtiene ajustando localmente un polinomio a la función de distribución empírica minimizando la norma L^2 ponderada, esto es, el estimador $\hat{F}_{n,p}(x) = P_{p,x}(x)$, donde $P_{p,x}(x)$ es el polinomio de grado p que minimiza la expresión:

$$\int W\left(\frac{x-u}{h_n}\right)\left(F_n(u) - P_p(u)\right)^2\,du.$$

siendo $W(v)$ una función de peso.

En el apartado 2 de este trabajo se estudia el Error Cuadrático Medio de $\hat{F}_n(x)$ y la covarianza asintótica de $\hat{F}_n(x)$ y $\hat{F}_n(y)$, comparando los resultados con los obtenidos para la distribución empírica. Este estimador, en el supuesto de independencia, ha sido estudiado por Azzalini (1984) y Feraldo Roca-González-Manteiga (1984).

En el supuesto de que las observaciones se recojan secuencialmente es interesante trabajar con estimadores recursivos, lo que representa una ventaja desde un punto de vista computacional, sobre todo si el tamaño muestral es grande. Por ello, en el apartado 3 se estudia el Error Cuadrático Medio de la versión recursiva del estimador definido en (3) y cuya expresión es la siguiente:

$$\hat{F}_{n,\tau}(x) = \left(\sum_{i=1}^{n} h_i^{\tau}\right)^{-1} \left(\sum_{i=1}^{n} h_i^{\tau} \hat{\delta}_{h(i)}(x, X_i)\right) \tag{4}$$

siendo $\tau \in [0,1]$ un parámetro que influye en el Error Cuadrático Medio de forma análoga que $h(n)$ pero en menor medida, como se expondrá más adelante. Normalmente para τ se utilizan los valores 0, 1/2 y 1.

El estimador $\hat{F}_{n,\tau}(x)$ verifica la siguiente relación recursiva:

$$\hat{F}_{n+1,\tau}(x) = \left(\hat{F}_{n,\tau}(x) E_n + h(n+1)^{\tau} \hat{\delta}_{h(n+1)}(x, X_i)\right) E_{n+1}^{-1} \tag{5}$$

siendo $E_n = \sum_{i=1}^{n} h_i^{\tau}$.

Finalmente en el apartado 4 se obtiene la normalidad asintótica de los estimadores $\hat{F}_n(x)$ y $F_{n,z}(x)$.

LA ESTRUCTURA DE DEPENDENCIA

Aunque la mayoría de los estudios realizados sobre la estimación no paramétrica de curvas suponen que las observaciones son independientes, en muchas situaciones, esta hipótesis no es realista, sobre todo si se trabaja con series de tiempo, por ello en este trabajo se asume que las observaciones puedan ser dependientes, aunque exigiendo que la dependencia entre dos observaciones tienda a anularse al aumentar la distancia temporal entre ellos. Para ello se utiliza la condición de dependencia "fuertemente mixing" ($\alpha-$mixing) introducida por Rosenblatt (1956) y cuya definición es la siguiente:

"Sea $X(t)$, $t \in Z$, una sucesión de v.a. estrictamente estacionarias, definidas sobre el espacio de probabilidad (Ω, F, P) con valores en $\mathbb{R}$. Sea $F_{-\infty}^{0}$ y $F_n^{+\infty}$ las $\sigma-$álgebras generadas respectivamente por $X(t)$ con $t \leq 0$ y $X(t)$ con $t \geq n$. Entonces $X(t)$ es fuertemente mixing si:

(6) $\alpha(n) = \sup\left\{|P(A\cup B) - P(A)P(B)| : A \in F_{-\infty}^{0},\, B \in F_{n}^{+\infty}\right\} \longrightarrow 0$"

Esta condición es más débil que la mayoría de las utilizadas: uniformemente mixing, asintóticamente incorreladas o absolutamente regular. Verificándola entre otros los procesos lineales construidos a partir de sucesiones de variables aleatorias i.i.d. absolutamente continuas y funcionales no lineales de procesos gaussianos exigiendo condiciones no muy restrictivas a la densidad espectral. (Ver Bradley (1985)).

2. ERROR CUADRÁTICO MEDIO DE $F_n(x)$

En todo el trabajo se supondrá que la sucesión de funciones de ponderación $\delta_{h(n)}$ verifica las siguientes hipótesis:

D.1. $\delta_{h(n)}(x,u) \geq 0, \quad \int \delta_{h(n)}(x,u)\,du = 1$

D.2. $\delta_{h(n)}(x,x+u) = \delta_{h(n)}(x,x-u)$
$\delta_{h(n)}(x,x+u) = \delta_{h(n)}(0,u)$

D.3. $\delta_{h(n)}(x,u) = 0h(n)^{-1} \quad \text{si } |x-u| \leq h(n)$
$\delta_{h(n)}(x,u) = 0 \qquad \text{si } |x-u| > h(n)$

D.4. $h(n) \longrightarrow 0 \quad \text{y} \quad nh(n) \longrightarrow \infty$

Con el fin de poder comparar la conducta asintótica de los estimadores $F_n(x)$ y $\hat{F}(x)$ se obtiene, en primer lugar, el Error Cuadrático Medio de la distribución empírica (FDE).

Teorema 1.

Si el proceso $X(t)$ es $\alpha-$mixing con $\sum_t \alpha(t) < \infty$, se verifica:

i) $E(F_n(x)) = F(x)$

ii) (7) $\quad \operatorname{var}(F_n(x)) = \frac{1}{n}\left(F(x)(1-F(x)) + D_1^n\right) = \frac{1}{n}\,\sigma_1^2$

siendo D_1^n un término debido a la dependencia acotado por $2\sum_{k=1}^{n} \alpha(k) < \infty$.

Demostración:

El apartado i) es inmediato ya que el sesgo del estimador no se ve afectado por la dependencia de los datos. Respecto a ii) se obtiene:

$$\begin{aligned}\text{var}(F_n(x)) &= \frac{1}{n^2}\text{var}\ (\textstyle\sum_{i=1}^n U_i) = \\ &= \frac{1}{n^2}\left(nF(x)(1-F(x)) + 2\sum_{k=1}^{n-1}(n-k)\text{cov}(U_1 U_{1+k})\right)\end{aligned}$$

de la definición (6) se sigue $\text{cov}(U_1, U_{1+k}) \leq \alpha(k)$, sustituyendo

$$\text{var}(F_n(x)) \leq \tfrac{1}{n}\ \left(F(x)(1-F(x)) + 2\textstyle\sum_{k=1}^{n-1}\frac{n-k}{n}\alpha(k)\right) \leq \tfrac{1}{n}\ (F(x)(1-F(x)) + D_1^n)$$

■

En los dos siguientes teoremas se obtiene el sesgo y varianza del estimador $\hat{F}_n(x)$ definido en (3).

Teorema 2.

Si se verifica

H1. f es $(2r+1)$ veces diferenciable.

H2. $\int_{\mathbb{R}} g_n(t)\,dt$ es acotada, siendo $g_n(t) = \sup\limits_{|z-t|<h(n)} |f^{2r+1)}(z)|$

entonces

$$B(\hat{F}_n(x)) = E(\hat{F}_n(x)) - F(x) = \sum_{k=1}^{r}\frac{f^{(2k)}(x)}{(2k)!}D_{2k} + oh_n^{2r} \tag{8}$$

siendo

$$D_{2k} = \int_{-h_n}^{+h_n}\delta_{h(n)}(v,0)v^{2k}\,dv = 0(h_n^{2k})$$

Comentarios:

1. El sesgo no se ve afectado por la dependencia muestral. Y el estimador $\hat{F}_n$ es asintóticamente insesgado pero no insesgado como lo es la FDE: F_n.

2. Para el caso particular de utilizar el estimador núcleo, con $r = 1$ se obtiene la siguiente expresión del sesgo de $\hat{F}_n(x)$

$$B(\hat{F}_n(x)) = \frac{f''(x)}{2} K_2 h_n^2 + o h_n^2 \tag{9}$$

siendo $K_j = \int K(u)\, u^j\, du$.

Demostración:

Se realiza un desarrollo de Taylor de orden $2r$ en $f(u)$

$$\begin{aligned}
E(\hat{F}_n(x)) &= E(\hat{\delta}_{h(n)}(x, X_i) = \int \left(\int \delta_{h(n)}(u,t)\, dt \right) f(u)\, du = \\
&= F(x) + \sum_{k=1}^{2r} \int_{-\infty}^{x} \frac{f^{(k)}(t)}{k!} \left(\int \delta_{h(n)}(u,t)(u-t)^k\, du \right) dt + \\
&+ \int_{-\infty}^{x} \frac{f^{(2r+1)}(z)}{(r+1)!} \left(\int \delta_{h(n)}(u,t)(u-t)^{2r+1}\, du \right) dt
\end{aligned}$$

y de las hipótesis de la sucesión $\delta_{h(n)}$ y H2 se deduce (8).

■

Teorema 3.

Si se verifican las hipótesis del Teorema 2 ($r = 1$) y H3. Existe un $\nu > 0$ tal que $\sum_{k=1}^{\infty} \alpha(k)^{\nu/2+\nu} < \infty$ entonces $\text{var}(\hat{F}_n(x)) = \frac{1}{n}\, (F(x)(1 - F(x)) + f(x)$ $(A_n(x) - h_n) + 0h_n^2 + D_2^n) = \frac{1}{n}\, \sigma_2^2$ siendo

$$A_n(x) = \int_{-h_n}^{+h_n} \hat{\delta}^2_{h(n)}(x,u)\, du = 0 h_n \tag{10}$$

y D_2^n un término debido a la dependencia acotado por 16 $\sum_{k=1}^{n} \alpha(k)^{\nu/2+\nu}$.

Suponiendo que $x < y$, se verifica:

$$\operatorname{cov}\left(\hat{F}_n(x), \hat{F}_n(y)\right) = \frac{1}{n}\left(F(x)(1-F(y)) + f(x)(B_n(x,y) - h_n) + 0h_n^2 + D_3^n\right)$$

siendo

$$B_n(x,y) = \int_{-h_n}^{+h_n} \hat{\delta}_{h(n)}(x,u)\, \hat{\delta}_{h(n)}(y,u)\, du = 0h_n \tag{11}$$

y D_3^n un término debido a la dependencia acotado por $16\sum_{k=1}^{n} \alpha(k)^{\nu/2+\nu}$.

Comentarios:

1. De las expresiones (10) y (11) se deduce que la dependencia de la muestra afecta a la varianza de $\hat{F}_n(x)$ y a la covarianza de $F_n(x), \hat{F}_n(y)$.

2. Dado que $A_n(x) - h_n$ es negativo, la varianza del estimador $\hat{F}_n(x)$ es menor que la de $F_n(x)$. Así, utilizando el kernel uniforme $(K(u) = 1/2$ si $|u| \leq 1)$ se obtiene que $A_n(x) - h_n = -(5/3)\, h_n$. Y si se utiliza el Kernel de Epanechnikov $\left(K(u) = 3/4(1-u)^2 \text{ si } |u| \leq 1\right)$ se obtiene que $A_n(x) - h_n = -(61/35)h_n$. Por ello, es mejor utilizar el Kernel de Epanechnikov que el uniforme, ya que tiene menor varianza.

3. De las expresiones (8) y (10) se deduce la expresión del Error Cuadrático Medio del estimador $\hat{F}(x)$, que en el supuesto de que f sea tres veces diferenciable $(r = 1)$ viene dada por:

$$\begin{aligned} \mathrm{ECM}(\hat{F}_n(x)) &= \left(\frac{f''(x)}{2} D_2 + 0h_n^2\right) + \frac{1}{n}\,(F(x)(1-F(x)) + \\ &+ f(x)(A_n(x) - h_n) + 0h_n^2 + D_2^n) \end{aligned}$$

   De donde se deduce la convergencia puntual en media cuadrática del estimador definido.

4. Considerando al $\mathrm{ECM}\,(\hat{F}_n(x))$ como una función de h, se puede calcular el parámetro h que minimice este error. De la expresión anterior se deduce que:

$$\mathrm{ECM}(\hat{F}_n(x)) = \psi(h) = C_1 n^{-1} - C_2 h n^{-1} + C_3 h^4 + C_4 h^2 n^{-1}$$

siendo C_1, C_2, C_3, y C_4 constantes positivas que dependen de la función teórica $F(x)$ y de las funciones de ponderación, $\delta_{h(n)}$, utilizadas. Minimizando la expresión anterior se obtiene que el parámetro de suavización óptimo tiene la forma: $h_{\text{ópt}} = C_0 n^{-1/3}$. Siendo C_0 una constante positiva, que depende de F, por ejemplo, utilizando el kernel de Epanechnikov vale, aproximadamente, $3'03 \left(\frac{f(x)}{f'(x)^2}\right)^{1/3}$.

Desde un punto de vista práctico la expresión de h obtenida no es útil ya que depende de la función teórica que deseamos estimar. Por ello, es interesante desarrollar técnicas para obtener dicho parámetro a partir de la muestra. En la actualidad estamos trabajando en este problema utilizando técnicas de validación cruzada mínimo cuadrático, global y local, adaptándolas al caso de datos dependientes.

Demostración

Por la estacionaridad del proceso se obtiene:

$$\text{(12)} \qquad \begin{aligned} \operatorname{cov}\Big(F_n(x), F_n(y)\Big) &= \frac{1}{n}\operatorname{cov}\Big(\hat{\delta}_{h(n)}(x, X_1), \delta_{h(n)}(y, X_1)\Big) + \\ &+ \frac{1}{n}\sum_{k=1}^{n-1}\frac{n-k}{n}\left(U_{x,y}(k) + V_{x,y}(k)\right) \end{aligned}$$

siendo

$$\begin{aligned} U_{x,y}(k) &= \operatorname{cov}\Big(\delta_{h(n)}(x, X_1),\ \delta_{h(n)}(y, X_{1+k})\Big) \\ V_{x,y}(k) &= \operatorname{cov}\Big(\hat{\delta}_{h(n)}(x, X_{1+k}) + \hat{\delta}_{h(n)}(y, X_1)\Big) \end{aligned}$$

Se desarrolla, en primer lugar, la expresión (12) en el caso $x = y$ obteniendo:

$$\text{(13)} \qquad \operatorname{var}(\hat{F}_n(x)) = \tfrac{1}{n}\operatorname{var}(\hat{\delta}_{h(n)}(x, X_1)) + \tfrac{2}{n}\textstyle\sum_{k=1}^{n-1}\frac{n-k}{n}\,U_{x,x}(k)$$

pero

$$\text{(14)} \qquad \operatorname{var}(\hat{\delta}_{h(n)}(x, X_1)) = E(\delta^2_{h(n)}(x, X_1)) - \Big(E(\delta_{h(n)}(x, X_1))\Big)^2$$

y del resultado del Teorema 2, se sigue que:

$$\operatorname{var}(\delta_{h(n)}(x, X_1) = \int \acute{\delta}^2_{h(n)}(x, u) f(u)\, du - \left(F(x) + o\epsilon_n^2\right)^2 ,$$

haciendo un desarrollo de Taylor en el primer sumando, por las propiedades de $\acute{\delta}$ se sigue que $E(\acute{\delta}^2_{h(n)}(x, X_1)) = F(x) - f(x)h_n + f(x)A_n + 0h_n^2$ y sustituyendo en (14) se obtiene

$$\operatorname{var}(\acute{\delta}_{h(n)}(x, X_1)) = F(x)(1 - F(x)) + f(x)(A_n(x) + 0h_n^2 \tag{15}$$

Para acotar $U_{x,x}(k)$ se utiliza la **Desigualdad de Davydov:**

"Sean ξ y η dos variables aleatorias respecto a las σ-álgebras F_1^i y F_1^j, respectivamente, $i, j = 1, 2, 3, \cdots$ Siendo $E\,|\xi|^{2+\nu} < \infty$ y $E|\eta|^{2+\nu} < \infty$ para algún $\nu > 0$, entonces para $i \neq j$ se verifica:

$$\operatorname{cov}(\xi, \eta) \leq 8\ \left(E|\xi|^{2+\nu}\, E|\eta|^{2+\nu}\right)^{\nu/2+\nu} (\alpha|i-j|)^{\nu/2+\nu} \text{"} \tag{16}$$

(La demostración puede verse en Hall-Heyde (1980)).

Utilizando (16) se obtiene

$$|U_{x,x}(k)| \leq 8\ \left(E|\delta_{h(n)}(x, X_1)|^{2+\nu}\, E|\hat{\delta}_{h(n)}(x, X_{1+k})|^{2+\nu}\right)^{\nu/2+\nu} (\alpha|k|)^{\nu/2+\nu}$$

por un desarrollo análogo al realizado en (15), se sigue que

$$E|\hat{\delta}_{h(n)}(x, X_1)|^{2+\nu} = F(x - h_n) + 0h_{n'}$$

y por tanto

$$|U_{x,x}(k)| \leq 8\ (F(x - h_n) + 0h_n)^{2/2+\nu} (\alpha(k))^{\nu/2+\nu} \leq 8\ (\alpha(k))^{\nu/2+\nu} \tag{17}$$

Finalmente, sustituyendo (15) y (17) en (13) se obtiene el primer apartado del Teorema 3.

Supongamos ahora que $x < y$, de la expresión (12) se sigue

$$(18) \quad \operatorname{cov}\left(\hat{F}_n(x), \hat{F}_n(y)\right) = \tfrac{1}{n}\, U_{x,y}(0) + \tfrac{1}{n} \sum_{k=1}^{n-1} \tfrac{n-k}{n} \; (U_{x,y}(k) + V_{x,y}(k))$$

por desarrollos análogos a los realizados en la primera parte se obtiene:
(19)

$$\begin{aligned} U_{x,y}(0) &= E\left(\hat{\delta}_{h(n)}(x, X_1), \hat{\delta}_{h(n)}(y, X_1)\right) - E\left(\hat{\delta}_{h(n)}(x, X_1)\right) E\left(\hat{\delta}_{h(n)}(x, X_1)\right) \\ &= F(x)(1 - F(y)) + f(x)(B_n(x,y) - h_n) + 0h_n^2 \end{aligned}$$

Para acotar $U_{x,y}(k)$ y $V_{x,y}(k)$ se utiliza la desigualdad de Davydov:

$$\begin{aligned} |U_{x,y}(k)| &\leq 8\ \left(E|\hat{\delta}_{h(n)}(x, X_1)|^{2+\nu}\, E|\hat{\delta}_{h(n)}(y, X_{1+k})|^{2+\nu}\right)^{\nu/2+\nu} (\alpha|k|)^{\nu/2+\nu} \\ (20) \qquad &\leq 8\ (\alpha|k|)^{\nu/2+\nu} \end{aligned}$$

sustituyendo (19) y (20) en la expresión (18) se obtiene:

$$n \operatorname{cov}\left(\hat{F}_n(x), \hat{F}_n(y)\right) = F(x)(1 - F(y)) + f(x)(B_n(x,y) - h_n) + 0h_n^2 + D_3^n$$

siendo

$$D_3^n = \sum_{k=1}^{n-1} \frac{n-k}{n}\ (U_{x,y}(k) + V_{x,y}(k)) \leq 16 \sum_{k=1}^{n-1} (\alpha|k|)^{\nu/2+\nu} < \infty$$

■

3. ERROR CUADRÁTICO MEDIO DE $F_{n,\tau}(x)$

En este apartado se obtiene el Error Cuadrático Medio del estimador recursivo $\hat{F}_{n,\tau}(x)$, definido en (4).

Teorema 4.

Si se verifican las hipótesis del Teorema 2 y además

H4. $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\left(\frac{h_i}{h_n}\right)^s = \theta_s$ con $s \le 2r+1$. Se obtiene:

$$(21)\quad B(\hat{F}_{n,\tau}(x)) = E(\hat{F}_{n,\tau}(x)) - F(x) = \sum_{k=1}^{r} \frac{f^{(2k)}(x)}{(2k)!} \frac{\sum_{i=1}^{n} h_i^{\tau} D_{2k}^{i}}{\sum_{i=1}^{n} h_i^{\tau}} + oh_n^{2r}$$

siendo $D_{2k}^{i} = \int_{-h_i}^{+h_i} \delta_{h(i)}(v,0)v^{2k}\,dv = 0(h_i^{2k})$.

Comentarios:

1. La hipótesis H4 la verifica la elección clásica del parámetro de suavización $h_n = Cn^{-\alpha}$, con $o < \alpha < 1$, en este caso, $\theta_s = 1/1-\alpha s$.
2. Al sesgo de $\hat{F}_{n,\tau}(x)$ no le afecta la dependencia de los datos, es asintóticamente insesgado pero de mayor sesgo que $\hat{F}_n(x)$, como se ve fácilmente en el caso particular del estimador núcleo, en el que se obtendría, para $r = 1$:

$$(22)\qquad B(\hat{F}_{n,\tau}(x)) = B(\hat{F}_n(x))H(\tau)$$

siendo $H(\tau) = \frac{\theta_{\tau+2}}{\theta_\tau}$, que para la ventana del tipo $h_n = Cn^{-\alpha}$, se obtiene que $H(\tau) = \frac{1-\alpha\tau}{1-\alpha\tau-2\alpha}$, que es una función creciente respecto a τ y acotada inferiormente por 1, para $\tau \in [0,1]$, por tanto, $B(\hat{F}_{n,\tau}(x)) > B(\hat{F}_n(x))$, aunque ambos son del mismo orden y, el sesgo de $\hat{F}_{n,\tau}(x)$ será mayor cuanto mayor sea τ.

Teorema 5.

Si se verifican las hipótesis H1-4, se obtiene:

$$(23)\qquad \mathrm{var}(\hat{F}_{n,z}(x)) = \mathrm{var}(\hat{F}_n(x))\,G(z) = \frac{1}{n}\,\sigma_3^2, \text{ siendo } G(\tau) = \theta_{2z-1}/\theta_z^2$$

Suponiendo que $x < y$, se verifica:

$$(24)\qquad \mathrm{cov}\left(\hat{F}_{n,\tau}(x), \hat{F}_{n,\tau}(y)\right) = \mathrm{cov}\left(\hat{F}_n(x), \hat{F}_n(y)\right)\, G(\tau)$$

Comentarios:

1. De la expresión (23) se deduce que la dependencia de los datos afecta a la varianza de $\hat{F}_{n,\tau}(x)$ y a la covarianza de $\hat{F}_{n,\tau}(x)$ y $\hat{F}_{n,\tau}(y)$, existiendo una estrecha relación con las expresiones obtenidas para el estimador $\hat{F}_n(x)$.

2. Para la elección de la ventana $h_n = Cn^{-\alpha}$, se obtiene que $G(z) = \frac{(1-\alpha z)^2}{1-2\alpha z+\alpha}$, que es una función decreciente respecto a la variable τ y que está acotada superiormente por 1, para $\tau \in [0,1]$. Por tanto, la varianza y covarianza del estimador recursivo, $\hat{F}_{n,\tau}(x)$, es menor que la del estimador no recursivo, $\hat{F}_n(x)$, aunque del mismo orden. Y éstas aumentan al tomar τ menores.

3. De los Teoremas 4 y 5 se deduce que el parámetro τ juega un papel similar al del parámetro de suavización h, ya que al aumentar τ aumenta el sesgo del estimador pero disminuye su varianza. Hecho que ya habíamos comprobado al estudiar la estimación no paramétrica de la densidad y regresión (Vilar Fernández, 1989). En cualquier caso por estudios de simulación que hemos realizado la influencia del parámetro τ es mucho menor que la de la ventana h.

4. Las expresiones (21) y (23) permiten obtener el Error Cuadrático Medio y la convergencia puntual en media cuadrática del estimador $\hat{F}_{n,\tau}(x)$, siendo la velocidad de convergencia del mismo orden que la del estimador no recursivo.

Demostración de los Teoremas 4 y 5.

Es análoga a la demostración de los Teoremas 2 y 3, utilizando el Lema de Toeplitz.

4. DISTRIBUCIÓN ASINTÓTICA

Para obtener la normalidad asintótica de los estimadores no paramétricos definidos $\hat{F}_n(x)$ y $\hat{F}_{n,\tau}(x)$ se utilizarán Teoremas Centrales del Límite para disposiciones triangulares fuertemente mixing, desarrollados por Bradley (1981).

Se utilizarán las siguientes definiciones:

$$\begin{aligned} \rho(A,B) &= \sup\{\text{Correl}\,(f,g) : f \in L^2(A);\ g \in L^2(B)\} \\ \rho(n) &= \rho\left(F_1^k; F_{k+n}^{\infty}\right) \text{ y } \rho^* = \lim_{n\to\infty} \rho(n) \end{aligned}$$

Para $\epsilon > 0$ y $0 \leq \rho < 1$, se define $g(\epsilon, \rho) = \dfrac{1 + \rho^{2\epsilon/(2+\epsilon)} + 2\rho^{2/(2+\epsilon)}}{2^{\epsilon/2}(1-p)^{(2+E)/2}}$

Teorema 6.

Si se verifican las hipótesis del Teorema 3 y además H5. Existe un ϵ, $0 \leq \epsilon < 1$ tal que $g(\epsilon, \rho^*) < 1$.

Se obtiene

$$\sqrt{n}\left(\hat{F}_n(x) - E\hat{F}_n(x)\right) \xrightarrow{d} \mathrm{N}(0, \sigma_2) \tag{25}$$

Y si además se verifica H6. $nh_n^4 \longrightarrow 0$.

Entonces

$$\sqrt{n}\left(\hat{F}_n(x) - F(x)\right) \xrightarrow{d} \mathrm{N}(o, \sigma_2) \tag{26}$$

La normalidad asintótica del estimador recursivo $\hat{F}_{n,\tau}(x)$ se expone en el siguiente teorema:

Teorema 7.

Si se verifican las hipótesis del Teorema 5, H5 y además H7. $nh_n^{2\tau} \longrightarrow \infty$ cuando $n \rightarrow \infty$. Se obtiene

$$\sqrt{n}\left(\hat{F}_{n,\tau}(x) - E\hat{F}_{n,\tau}(x)\right) \xrightarrow{d} \mathrm{N}(0, \sigma_3) \tag{27}$$

Y si además se verifica H6, entonces

$$\sqrt{n}\left(\hat{F}_{n,\tau}(x) - F(x)\right) \xrightarrow{d} \mathrm{N}(0, \sigma_3) \tag{28}$$

Comentarios:

1. Las hipótesis H6 y H7 no son restrictivas ya que la elección del parámetro de suavización óptima, esto es, que minimiza el Error Cuadrático Medio, es de la forma $h_n = Cn^{-1/3}$ que verifica las mencionadas hipótesis.

2. La hipótesis H5 si es más restrictiva respecto a la condición de dependencia de los datos muestrales. Trivialmente se verifica si $\rho^* = 0$, que es la conocida condición de que el proceso sea "asintóticamente incorrelado", condición intermedia entre uniformemente mixing y fuertemente mixing y, que para procesos Gaussianos estacionarios es equivalente a la fuertemente mixing (Kolmogorov-Rozanov, 1960). La hipótesis H5 no exige que ρ^* sea cero pero si ha de ser próximo a cero.

3. Otra metodología para obtener la normalidad asintótica de los estimadores no paramétricos definidos es la clásica utilizada en un contexto de dependencia, que se basa en considerar la suma de variables aleatorias como una suma de grandes bloques separados por pequeños bloques de órdenes tales que permitan probar que los primeros son asintóticamente independientes y los segundos asintóticamente nulos. Lo que permite aplicar Teoremas Centrales del Límite clásicos, como el de Lindenberg. Este método ha sido utilizado entre otros por Masry (1986) estudiando estimadores no paramétricos recursivos de la función de densidad y aunque permite utilizar hipótesis más débiles sobre la dependencia de la muestra se le exigen hipótesis más restrictivas a la sucesión de parámetros de suavización.

Demostración del Teorema 6.

Se considera la disposición triangular de variables aleatorias siguiente:

$$\begin{aligned} Z_i^n = \delta_{h(n)}(x, X_i) - E\,\delta_{h(n)}(x, X_i) \quad i &= 1, 2, \cdots n, \\ n &= 1, 2, \cdots \end{aligned}$$

que verifica, para cualquier n, $\{Z_i^n\}_{i=1}^n$ es estacionaria y fuertemente mixing, siendo $EZ_i^n = 0$, y de la expresión (15) se sigue que

$$\mathrm{var}(Z_i^n) = \mathrm{var}(\delta_{h(n)}(x, X_i)) = F(x)(1 - F(x)) + 0h_n \leq 1$$

Por otra parte, utilizando el desarrollo realizado en (17) se verifica que dado $\epsilon > 0$, $E|Z_i^n|^{2+\epsilon} = F(x - h_n) + 0h_n < \infty$.

Sea $S_n = \sum_{i=1}^n Z_i^n = n\left(F_n(x) - EF_n(x)\right)$ y $s_n^2 = E\,(S_n^2) = n^2 \mathrm{var}\left(F_n(x)\right) = 0\,(n) \longrightarrow \infty$ según el resultado obtenido en el Teorema 3.

De todo lo anterior se sigue que se cumplen las hipótesis del Teorema 4 de Bradley (pp. 4, 1981) y, por tanto, se verifica que $S_n/s_n \longrightarrow \mathrm{N}(0,1)$, de las definiciones de S_n y s_n se deduce la primera conclusión del teorema.

Para probar la segunda parte, basta tener en cuenta que

$$\sqrt{n}\left(F_n(x)-F(x)\right)=\sqrt{n}\left(\hat{F}_n(x)-E\hat{F}_n(x)\right)+\sqrt{n}\left(E\hat{F}_n(x)-F(x)\right)$$

el primer sumando converge en distribución a una $N(0,\sigma_2)$ y del teorema 2 y la hipótesis H6 se deduce la convergencia a cero del segundo sumando.

■

Demostración del Teorema 7.

La demostración del Teorema 7 es análoga a la del teorema 6, en este caso, se considera la siguiente disposición triangular de variables aleatorias:

$$\begin{array}{rcl} Y_i^n = h_i^\tau \delta_{h(i)}(x,X_i) - Eh_i^\tau \delta_{h(i)}(x,X_i) \quad i & = & 1,2,\cdots n \\ n & = & 1,2,\cdots \end{array}$$

que verifica, para todo $n=1,2,\cdots$ que $\{Y_i^n\}_{i=1}^n$ es fuertemente mixing pero no es estacionaria.

Sea

$$S_n=\sum_{i=1}^n Y_i^n \quad \text{y} \quad s_n^2=E\left(S_n^2\right)=\text{var}\left(\left(\sum_{i=1}^n h_i^\tau\right)\hat{F}_{n,\tau}(x)\right)=$$

$$\left(\sum_{i=1}^n h_i^\tau\right)^2\frac{1}{n}\sigma_3^2=\left(\frac{1}{n}\sum_{i=1}^n\frac{h_i^\tau}{h_n^\tau}\right)^2 h_n^{2\tau}n\sigma_3^2=0(nh_n^{2\tau}) \longrightarrow \infty \quad \text{por H7.}$$

Razonando como en el Teorema 6 se obtiene que $\{Y_i^n\}$ verifica las hipótesis del Teorema 3 de Bradley (1981), que es válido para disposiciones triangulares no necesariamente estacionarias y, por tanto, se verifica que $S_n/s_n \longrightarrow N(0,1)$.

Siendo

$$\frac{S_n}{s_n}=\frac{\left(\sum_{i=1}^n h_i^\tau\right)\left(\hat{F}_{n,\tau}(x)-E\hat{F}_{n,\tau}(x)\right)}{n^{-1/2}\left(\sum_{i=1}^n h_i^\tau\right)\sigma_3}=\frac{n^{1/2}}{\sigma_3}\left(\hat{F}_{n,\tau}(x)-E\hat{F}_{n,\tau}(x)\right)$$

De donde se concluye la primera parte del Teorema 7. El segundo resultado se obtiene igual que en el Teorema 6.

■

5. BIBLIOGRAFÍA

[1] **Azzalini, A.** (1981). "A note on the estimation of a distribution function and quantiles by a kernel method". *Biometrika*,68, 1, 226-8.

[2] **Bradley, R.C.** (1981). "Central Limit Theorems under Weak Depedence". *Journal of Multivariate Analysis*,11, pp. 1-16.

[3] **Bradley, R.C.** (1985). "Basic Properties of strong mixing conditions". Del libro "Dependence in Probability and Statistics". Birkhauser.

[4] **Faraldo Roca – Gonzalez Manteiga** (1984). "Obtención del sesgo, varianza y error cuadrático medio de una familia axomática de estimadores no paramétricos para funciones de distribución". Actas del XIV Congreso de la SEIO, Granada.

[5] **Hall – Heyde** (1980). "Martingale limit their and its application". Academic Press.

[6] **Izenman – Tran** (1990). "Kernel estimation of the survival function and hazard rate weak dependence". *Journal of statistical planning and Inference*,24, pp. 233-247.

[7] **Kolmogorov – Rozanov** (1960). "On strong mixing conditions for stationary Gaussian processes". *Ther. Probability Appl.*,5, 204-208.

[8] **Lejeune – Sarda** (1989). "Smooth estimators of distribution and density functions". Preprint.

[9] **Masry, E.** (1986). "Recursive probability density estimation for weakly dependent stationary processes". IEEE, vol. IT-32, 2, 254-267.

[10] **Reiss, R.D.** (1981). "Nonparametric estimation of smooth distribution functions". *Scand. J. Statist.*,8, pp. 116-119.

[11] **Roussas, G.** (1989). "Some asymptotic properties of an estimate of the survival function under dependence conditions". *Statistics & Probability Letters*,8, pp. 235-243.

[12] **Sarda – Vieu** (1989). "Empirical distribution function for mixing random variables. Application in nonparametric Hazard estimation". *Statistics*,20, pp. 559-571.

[13] **Sarda** (1990). "Estimating smooth distribution functions". Preprint.

[14] **Vilar Fernández, J.** (1989). "Estimación recursiva, tipo núcleo, de la función de autorregresión para datos dependientes". *Estadística Española*, vol. 31, n°. 121, pp. 207-227.

[15] **Walter-Blum** (1979). "Probability density estimation using delta sequences". *The Annals of Statistics*, vol. 7, n°. 2, pp. 328-340.