

PROPIEDADES ASINTÓTICAS DE LOS ESTIMADORES DE MÍNIMA DISTANCIA CON COVARIABLES

GONZÁLEZ MANTEIGA, W.*,
PRESEDO QUINDIMIL, M.A.†

Universidad de Santiago de Compostela

En este trabajo se obtienen propiedades de consistencia y normalidad asintótica para el estimador no paramétrico de la función de regresión ($m(x)$) resultante de la extensión de la metodología de la mínima distancia de Cramer-von Mises al contexto de la estimación de curvas. Se hacen algunas consideraciones acerca de la robustez del estimador resultante en base a la función de influencia local (LIF) y se realiza un estudio de Monte Carlo comparativo con otros métodos de estimación.

Asymptotic properties of minimum distance estimators dependent on covariables.

Keywords: Minimum distance estimation, bandwidth parameter, local influence function.

*Departamento de Estadística e Investigación Operativa. Facultad de Matemáticas. Universidad de Santiago de Compostela.

† Departamento de Estadística e Investigación Operativa. Facultad de Veterinaria. Universidad de Santiago de Compostela.

-Article rebut el desembre de 1990.

1. INTRODUCCIÓN

Dadas $(X_1, Y_1), \dots, (X_n, Y_n)$ observaciones independientes e idénticamente distribuidas generadas según el modelo de regresión

$$Y = m(X) + \xi$$

donde ξ representa un error aleatorio con función de distribución F_0 , X una variable aleatoria unidimensional e independiente de ξ , ha sido de notable interés en los últimos años el estudio de la estimación robusta no paramétrica de la curva de regresión $m(x) = E[Y/X = x]$.

Así, Härdle (1984) extiende a este contexto la metodología de la M estimación, introducida por Huber (1964) en el contexto de la estimación de un parámetro de localización, como una generalización del estimador de mínimos cuadrados de $m(x)$ propuesto independientemente por Nadaraya (1964) y Watson (1964), basado en el método núcleo introducido por Rosenblatt (1956) y Parzen (1962) para la estimación de la densidad.

Más adelante, otras alternativas robustas al estimador de mínimos cuadrados han sido propuestas por Härdle, Janssen y Serfling (1988) y Cheng y Cheng (1987), quienes trasladan al contexto de la estimación no paramétrica de la curva de regresión los conceptos de la L y R estimación, respectivamente. Pudiéndose ver en González Manteiga (1990) una perspectiva global de todos estos métodos.

Aunque los estimadores de los tipos M, L y R muestran un buen comportamiento en aquellas situaciones en las que la distribución del error aleatorio es simétrica, éstos no se comportan de un modo tan satisfactorio cuando no se verifica la hipótesis de simetría, siendo por tanto de interés la construcción de estimadores alternativos a los anteriormente señalados que eviten dicho problema.

En este trabajo proponemos como nuevo estimador de la curva de regresión el resultante de extender la metodología de la mínima distancia de Cramer-von Mises, tratada por Stute (1986a) en el contexto de la estimación paramétrica, al campo de la estimación de curvas; que, como puede verse a partir de los resultados obtenidos mediante el estudio de simulación aquí presentado, se comporta razonablemente bien bajo la hipótesis de simetría en el error aleatorio ξ , mejorando los resultados de los procedimientos clásicos (mínimos cuadrados, M estimación) cuando la distribución del error no es simétrica.

Asimismo, se obtienen propiedades de consistencia y normalidad asintótica para el estimador de mínima distancia propuesto, al tiempo que se hacen algunas consideraciones acerca de la robustez de éste en base a la función de influencia local (LIF). En la sección 4 se presenta un estudio de simulación comparativo con

otros métodos, y se finaliza con un apéndice incluyendo las hipótesis necesarias relativas a los resultados obtenidos.

2. DEFINICIONES Y RESULTADOS PREVIOS

El modelo de regresión no paramétrica estudiado viene dado por

$$(1) \quad Y = m(X) + \xi$$

donde ξ es una variable aleatoria con función de distribución F_0 conocida de media cero (no necesariamente simétrica), $m(\cdot)$ la función de regresión a estimar, X una variable aleatoria con soporte $W \subset \mathbf{R}$ e independiente de ξ y $F(y/x) = F_0(y - m(x))$ representando la distribución de la variable Y condicionado a $X = x$.

Dado que $m(x)$ es el valor de θ que minimiza la distancia de Cramer-von Mises

$$(2) \quad d(F(\cdot/x), F_\theta) = \int [F(y/x) - F_\theta(y)]^2 dF(y/x)$$

en la familia $\{F_\theta\}_{\theta \in \Theta \subset \mathbf{R}} = \{F_\theta/F_\theta(y) = F_0(y - \theta)\}_{\theta \in \Theta \subset \mathbf{R}}$ con Θ suficientemente grande para contener a todos los posibles valores de $m(\cdot)$, tiene sentido definir como nuevo estimador de $m(x)$ (bajo el supuesto de existencia) el que minimiza la estimación del funcional (2)

$$(3) \quad d(\hat{F}_n(\cdot/x), F_\theta) = \int [\hat{F}_n(y/x) - F_\theta(y)]^2 d\hat{F}_n(y/x)$$

siendo $\hat{F}_n(\cdot/x)$ un estimador piloto no paramétrico para $F(\cdot/x)$ construido a partir de una muestra $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$ de (X, Y) siguiendo el modelo (1).

Definición 2.1

Llamaremos estimador de mínima distancia de Cramer-von Mises de la curva de regresión $m(x)$, al valor $\hat{m}_n(x)$ que resulta de resolver el problema de minimización (3).

De esta forma, los estimadores de mínima distancia tratados por Stute (1986a) son ahora extendidos a la estimación local, permutándose el papel de la distribución empírica de aquel contexto por el de la empírica condicionada $\hat{F}_n(\cdot/x)$ adaptada al nuestro. En Parr y Schucany (1980) pueden verse otras medidas de discrepancia alternativas a la aquí considerada.

Por la relevancia adquirida en el campo de la estimación no paramétrica de curvas, en este trabajo consideraremos como elecciones para $\hat{F}_n(. / x)$ las dadas por:

$$(4) \quad \hat{F}_n(y/x) = \frac{\frac{1}{nh_n} \sum_{i=1}^n I_{\{Y_i \leq y\}} K\left(\frac{x - X_i}{h_n}\right)}{\frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)}$$

estimador de Nadaraya (1964) y Watson (1964) para la estimación de la regresión $E[I_{\{Y \leq y\}} / X = x]$, y

$$(5) \quad \hat{F}_n(y/x) = \frac{1}{nh_n} \sum_{i=1}^n I_{\{Y_i \leq y\}} K\left(\frac{F_n(x) - F_n(X_i)}{h_n}\right)$$

llamado estimador de los k puntos próximos simetrizado introducido por Yang (1981) y estudiado por Stute (1984).

En ambos casos K es la función núcleo, h_n el parámetro ventana y F_n es la distribución empírica de $\{X_1, \dots, X_n\}$.

Aunque otras elecciones alternativas a (4) y (5) son posibles (Härdle (1990)) la facilidad computacional en el método núcleo también influyó en nuestra elección. La posibilidad de obtención de órdenes optimales de convergencia con las distancias L^p , $1 \leq p \leq \infty$, como puede verse en el libro antes citado, es también un argumento de peso para la elección del método núcleo.

Suponiendo que $d(F(. / x), F_\theta)$ definida por (2) es derivable bajo el signo integral, $m(x)$ es solución de la ecuación

$$(6) \quad \int \frac{\partial}{\partial \theta} F_\theta(y) [F(y/x) - F_\theta(y)] dF(y/x) = 0$$

y por tanto un caso particular de la más general

$$(7) \quad \int T(y, \theta, F(. / x)) dF(y/x) = 0$$

en la que $T : \mathbf{R} \times \Theta \times \mathcal{F} \rightarrow \mathbf{R}$ (Θ el espacio de parámetros, \mathcal{F} el espacio de funciones de distribución) es un funcional general. Por tanto, (6) representa la particularización $T(y, \theta, F) = \frac{\partial}{\partial \theta} F_\theta(y) [F(y) - F_\theta(y)]$.

De este modo suponiendo $m(x) \in \Theta_0 \subset \Theta$, con Θ_0 compacto, se verifica el siguiente resultado:

Teorema 2.2

Para el modelo (1) se obtiene que:

- a) $\sqrt{nh_n}(\hat{F}_n(\cdot/x) - F_0(\cdot - m(x))) \xrightarrow{d} B_0$
 con B_0 un proceso Gaussiano en $[0,1]$ de media cero y covarianza

$$\begin{aligned} \text{cov}(B_0(y_1), B_0(y_2)) &= \\ &= [F_0(\min\{y_1, y_2\} - m(x)) - F_0(y_1 - m(x))F_0(y_2 - m(x))] \frac{\int K^2(u)du}{f(x)} \end{aligned}$$

con $\hat{F}_n(\cdot/x)$ definido por (4), supuesto que X posee densidad positiva en x , siempre que se cumplan las hipótesis B1, B2 y B3.

- b) $\sqrt{nh_n}(\hat{F}_n(\cdot/x) - F_0(\cdot - m(x))) \xrightarrow{d} B_0$
 con B_0 un proceso Gaussiano en $[0,1]$ de media cero y covarianza

$$\begin{aligned} \text{cov}(B_0(y_1), B_0(y_2)) &= \\ &= [F_0(\min\{y_1, y_2\} - m(x)) - F_0(y_1 - m(x))F_0(y_2 - m(x))] \int K^2(u)du \end{aligned}$$

con $\hat{F}_n(\cdot/x)$ definido por (5), supuesto que X posee distribución F continua, siempre que se cumplan las hipótesis C1, C2, C4 y C5.

Obsérvese que el apartado (b) es una particularización del teorema 1, más general, de Stute (1986b) a nuestro caso; mientras que el apartado (a) lo es de la proposición 3.3 de Dabrowska (1987) en el contexto de datos censurados.

Finalmente, con las hipótesis adicionales B'1-B'5 para el estimador (4) y C1 y C'1-C'4 para el estimador (5), se verifica

Teorema 2.3 (Horváth y Yandell, 1988)

- a) Si $\hat{F}_n(\cdot/x)$ es el estimador definido por (4) entonces

$$\overline{\lim}(\log n)^{-1/2} \sup_y \left| \sqrt{nh_n}(\hat{F}_n(y/x) - F_0(y - m(x))) \right| \leq \left[\frac{\int K^2(u)du}{2f(x)} \right]^{1/2}$$

de forma casi segura.

b) Si $\hat{F}_n(. / x)$ es el estimador definido por (5) entonces

$$\overline{\lim}(\log n)^{-1/2} \sup_y \left| \sqrt{nh_n}(\hat{F}_n(y/x) - F_0(y - m(x))) \right| \leq \left[\frac{\int K^2(u) du}{2} \right]^{1/2}$$

de forma casi segura.

Las hipótesis B1, B2 y C1, C2 son inherentes a la función núcleo K y al grado de suavización determinado por el parámetro h_n . Las hipótesis B3 y C3, C4, C5 son necesarias para el control del sesgo siempre existente en los estimadores no paramétricos de curvas. Finalmente las hipótesis, en general más restrictivas, B' y C' permiten la obtención del orden optimal de convergencia $\left(\frac{\log n}{nh_n}\right)^{1/2}$ respecto de la norma del supremo (Härdle (1990)).

3. CONSISTENCIA Y NORMALIDAD ASINTÓTICA

En lo que sigue, representaremos por $\hat{m}_n(x)$ el estimador de mínima distancia de Cramer-von Mises resultante de la optimización de (3) bajo una de las elecciones (4) ó (5) del estimador piloto no paramétrico $\hat{F}_n(. / x)$.

Suponiendo que $m(x) \in \Theta_0$, con Θ_0 un conjunto compacto contenido en Θ , y $F(. / x) = F_0(. - m(x)) \in \{F_\theta\}_{\theta \in \Theta}$, se verifica el siguiente resultado:

Teorema 3.1

Bajo las hipótesis B'1-B'5 ó bajo las hipótesis C'1 y C'1-C'4, respectivamente, según el estimador piloto no paramétrico $\hat{F}_n(. / x)$ sea el definido por (4) ó (5), se verifica que

$$\min_{\theta} d(\hat{F}_n(. / x), F_\theta) = O\left(\left(\frac{\log n}{nh_n}\right)^{1/2}\right)$$

de forma casi segura.

Demostración

Representando por $\hat{F}_n(. / x)$ una de las elecciones (4) ó (5) del estimador piloto no paramétrico, se tiene que:

$$\begin{aligned}
d(\hat{F}_n(\cdot/x), F_\theta) &= \int \left[\hat{F}_n(y/x) - F_0(y - \theta) \right]^2 d\hat{F}_n(y/x) \leq \\
&\leq \sup_y |\hat{F}_n(y/x) - F_0(y - \theta)| \leq \\
&\leq \sup_y |\hat{F}_n(y/x) - F_0(y - m(x))| + \\
&+ \sup_y |F_0(y - m(x)) - F_0(y - \theta)|
\end{aligned}$$

Teniendo en cuenta que $m(x) \in \Theta$,

$$\min_{\theta} d(\hat{F}_n(\cdot/x), F_\theta) \leq \sup_y |\hat{F}_n(y/x) - F_0(y - m(x))|$$

y por tanto la aplicación del teorema 2.3 completa la demostración

■

El resultado que sigue da condiciones suficientes para la consistencia del estimador $\hat{m}_n(x)$.

Teorema 3.2

Si $F_\theta(y) = F_0(y - \theta)$, con F_0 continua, bajo las hipótesis B'1-B'5 ó bajo las hipótesis C1 y C'1-C'4, respectivamente, según el estimador no paramétrico $\hat{F}_n(\cdot/x)$ sea el definido por (4) ó (5), se verifica que

$$\hat{m}_n(x) \longrightarrow m(x)$$

de forma casi segura.

Demostración

Este resultado se obtiene a partir de la adaptación del teorema 1 de Parr y Schucany (1980) a nuestro contexto. A lo largo de la demostración $\hat{F}_n(\cdot/x)$ representa una de las elecciones (4) ó (5) del estimador piloto no paramétrico.

$$\begin{aligned}
d(F(\cdot/x), F_{\hat{m}_n(x)}) &= d(F_0(\cdot - m(x)), F_0(\cdot - \hat{m}_n(x))) = \\
&= \int [F_0(y - m(x)) - F_0(y - \hat{m}_n(x))]^2 dF_0(y - m(x)) = \\
&= \int [F_0(y - m(x)) - \hat{F}_n(y/x)]^2 dF_0(y - m(x)) +
\end{aligned}$$

$$\begin{aligned}
& + \int [\hat{F}_n(y/x) - F_0(y - \hat{m}_n(x))]^2 dF_0(y - m(x)) + \\
& + 2 \int [F_0(y - m(x)) - \hat{F}_n(y/x)][\hat{F}_n(y/x) - F_0(y - \hat{m}_n(x))] dF_0(y - m(x)) = \\
& = I + II + III \leq \\
& \leq I + III + \int [\hat{F}_n(y/x) - F_0(y - \hat{m}_n(x))]^2 d\hat{F}_n(y/x) + \\
& + \int [\hat{F}_n(y/x) - F_0(y - \hat{m}_n(x))]^2 d(F_0(\cdot - m(x)) - \hat{F}_n(\cdot/x))(y) = \\
& = I + III + d(\hat{F}_n(\cdot/x), F_{\hat{m}_n(x)}) + IV
\end{aligned}$$

Un razonamiento análogo al utilizado en la demostración del teorema 3.1 muestra que

$$I \leq \sup_y |\hat{F}_n(y/x) - F_0(y - m(x))|$$

y

$$|III| \leq 2 \sup_y |\hat{F}_n(y/x) - F_0(y - m(x))|$$

mientras que, integrando por partes, se obtiene para el sumando IV que

$$|IV| \leq 2 \|\hat{F}_n(\cdot/x) - F_0(\cdot - \hat{m}_n(x))\|_V \sup_y |F_0(y - m(x)) - \hat{F}_n(y/x)|$$

donde $\|h\|_V$ representa la variación total de h .

Por tanto, la aplicación del teorema 2.3 muestra que

$$d(F(\cdot/x), F_{\hat{m}_n(x)}) \leq d(\hat{F}_n(\cdot/x), F_{\hat{m}_n(x)}) + o(1)$$

de forma casi segura.

Dado que $d(\hat{F}_n(\cdot/x), F_{\hat{m}_n(x)}) = \inf_{\theta} d(\hat{F}_n(\cdot/x), F_{\theta})$, se verifica que

$$d(\hat{F}_n(\cdot/x), F_{\hat{m}_n(x)}) \leq d(\hat{F}_n(\cdot/x), F_{m(x)})$$

por tanto razonando análogamente a la primera parte de la demostración,

$$d(\hat{F}_n(\cdot/x), F_{\hat{m}_n(x)}) \leq d(F(\cdot/x), F_{m(x)}) + o(1)$$

concluyéndose que

$$d(F(\cdot/x), F_{\hat{m}_n(x)}) \longrightarrow d(F(\cdot/x), F_{m(x)}) = 0$$

de forma casi segura.

La unicidad de $m(x)$, resultante de la optimización de (2), con $m(x)$ perteneciente a Θ_0 compacto, junto con la continuidad en θ de la aplicación $d(F(. / x), F_\theta)$ completan la demostración. ■

Finalmente, en el teorema 3.3 abajo indicado se establecen condiciones suficientes para la obtención de la normalidad asintótica del estimador $\hat{m}_n(x)$.

Teorema 3.3

Suponiendo que $\hat{m}_n(x) \xrightarrow{P} m(x)$, se verifica:

a) Bajo las hipótesis A1, A2 y B1–B3

$$\sqrt{nh_n}(\hat{m}_n(x) - m(x))H'(m(x)) \xrightarrow{d} N\left(0, \frac{\text{var}[Q(Y)/X = x]}{f(x)} \int K^2(u)du\right)$$

en cada punto x tal que $f(x) > 0$, con $\hat{F}_n(. / x)$ definido por (4).

b) Bajo las hipótesis A1, A2 y C1–C5

$$\sqrt{nh_n}(\hat{m}_n(x) - m(x))H'(m(x)) \xrightarrow{d} N\left(0, \text{var}[Q(Y)/X = x] \int K^2(u)du\right)$$

con $\hat{F}_n(. / x)$ definido por (5).

Demostración

Dado que $T(y, \theta, F(. / x)) = \frac{\partial}{\partial \theta} F_\theta(y)(F(y/x) - F_\theta(y))$, se tiene que

$$\begin{aligned} T(y, \theta, F(. / x) + th) &= T(y, \theta, F(. / x)) + th(y) \frac{\partial}{\partial \theta} F_\theta(y) = \\ &= T(y, \theta, F(. / x)) + DT(y, \theta, F(. / x))(th) \end{aligned}$$

donde DT es la diferencial en el sentido de Riesz

$$DT(y, \theta, F(. / x))(h) = \int h(w)\mu(y, \theta, F(. / x), dw)$$

con $\mu(y, \theta, F(. / x)) = \frac{\partial}{\partial \theta} F_\theta(y)\delta_y$ la medida derivada de Riesz.

Con esta notación, el apartado (b) del teorema es una particularización del teorema, más general, de González Manteiga (1990).

Respecto al apartado (a) del teorema, razonando de modo análogo al ya mencionado resultado de González Manteiga, se tiene que la distribución asintótica de $\sqrt{nh_n}(\hat{m}_n(x) - m(x))H'(m(x))$ es la misma que la de

$$-\int Q(y)dZ_n(y/x)$$

con $Z_n(\cdot/x)$ el proceso empírico definido por

$$Z_n(y/x) = \sqrt{nh_n}(\hat{F}_n(y/x) - F_0(y - m(x)))$$

cuya distribución límite es la referida en el apartado (a) del teorema 2.2.

$$-\int Q(y)dZ_n(y/x) = \sqrt{nh_n} \left\{ \frac{\frac{1}{nh_n} \sum_{i=1}^n Q(Y_i) K\left(\frac{x - X_i}{h_n}\right)}{\hat{f}_{h_n}(x)} - E[Q(Y)/X = x] \right\}$$

donde $\hat{f}_{h_n}(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)$ es el estimador de Rosenblatt (1956) y Parzen (1962) para la densidad marginal $f(x)$.

La aplicación de la proposición 6.3 de Collomb (1976) a las variables $Q(Y_i)$ completa la demostración del apartado (a) del teorema. ■

Son de interés las siguientes observaciones:

Observación 3.4

Extendiendo de forma natural el concepto de función de influencia (introducido por Hampel (1968, 1974)) al de influencia local

$$(8) \quad LIF(y, S, F(\cdot/x)) = \lim_{\varepsilon \rightarrow 0} \frac{S((1 - \varepsilon)F(\cdot/x) + \varepsilon \Delta_y(\cdot)) - S(F(\cdot/x))}{\varepsilon}$$

para $S = S(F(\cdot/x))$ un funcional generalizado, y aplicándolo al caso aquí tratado $S(F(\cdot/x)) = m(x) = E[Y/X = x] = \int y dF_0(y - m(x))$, con $m(x) = \theta$ la solución de (6), se obtiene

$$LIF(y, S, F(\cdot/x)) = \frac{\int [F_0(z - m(x)) - \Delta_y(z)][f_0(z - m(x))]^2 dz}{\int [f_0(z - m(x))]^3 dz}$$

supuesto que F_0 posee densidad f_0 .

De esta forma los resultados relativos al teorema 3.3 pueden reescribirse como

$$\sqrt{nh_n}(\hat{m}_n(x) - m(x)) \xrightarrow{d} N\left(0, \int LIF(y, S, F(. / x))^2 dF(y/x) \int K^2(u) du\right)$$

para $\hat{F}_n(. / x)$ definido por (5) y de forma análoga para $\hat{F}_n(. / x)$ definido por (4) corrigiéndolo por el factor $f(x)$.

Trasladando al contexto de la estimación local el concepto de la B robustez, el estimador en mínima distancia local es un estimador robusto en la distribución $F(. / x)$ cuando la función de influencia local está acotada. Así por ejemplo si $F_0 = \Phi$, la distribución normal standard, $LIF(y, S, F(. / x)) = \sqrt{3\pi}(\Phi[\sqrt{2}(y - m(x))] - 1/2)$ es acotada como función de y , siendo por tanto el estimador en mínima distancia robusto en $F(. / x) = \Phi(. - m(x))$.

Observación 3.5

Una posibilidad, en general, de robustecer el estimador de mínima distancia consistiría en la sustitución de la ecuación integral (6) por la más general

$$(9) \quad \int \frac{\partial}{\partial \theta} F_\theta(y) J(F(y/x))(F(y/x) - F_\theta(y)) dF(y/x) = 0$$

con J una función no negativa tal que $\int_0^1 J(u) du = 1$, y posterior estimación de (9) a través de

$$(10) \quad \int \frac{\partial}{\partial \theta} F_\theta(y) J(\hat{F}_n(y/x))(\hat{F}_n(y/x) - F_\theta(y)) d\hat{F}_n(y/x) = 0$$

con $\hat{F}_n(. / x)$ definido por (4) ó (5).

$J(u) = 1$ sería el caso particular antes tratado, $J(u) = \frac{1}{1-2\epsilon} I_{\{\epsilon < u < 1-\epsilon\}}$ una distancia trimmed Cramer-von Mises, etc.

Suponiendo que J es una función dos veces continuamente diferenciable con derivadas acotadas, y sustituyendo las hipótesis A1 y A2 del teorema 3.3 por las correspondientes A'1 y A'2 (ver apéndice), se verifica dicho teorema con

$$Q(y) = - \int_{-\infty}^y \frac{\partial}{\partial \theta} \Big|_{\theta=m(x)} F_\theta(z) J(F_0(z - m(x))) dF_0(z - m(x))$$

es decir

$$\sqrt{nh_n}(\hat{m}_n(x) - m(x)) \xrightarrow{d} N\left(0, \int LIF(y, S, F(. / x))^2 dF(y/x) \int K^2(u) du\right)$$

para $\hat{F}_n(\cdot/x)$ definido por (5) y de forma análoga para $\hat{F}_n(\cdot/x)$ definido por (4) corrigiéndolo por el factor $f(x)$, con $\hat{m}_n(x)$ verificando la ecuación (10) y

$$\begin{aligned} LIF(y, S, F(\cdot/x)) &= \\ &= \frac{\int [F_0(z - m(x)) - \Delta_y(z)][f_0(z - m(x))]^2 J(F_0(z - m(x))) dz}{\int [f_0(z - m(x))]^3 J(F_0(z - m(x))) dz} \end{aligned}$$

en el supuesto de que F_0 sea absolutamente continua con función de densidad f_0 .

Obsérvese que el cambio de las hipótesis A1 y A2 por las correspondientes A'1 y A'2 está únicamente ligado a la nueva función $\frac{\partial}{\partial \theta} F_\theta(y) J(F_0(y - m(x)))$ alternativa a $\frac{\partial}{\partial \theta} F_\theta(y)$.

Observación 3.6

Sea $\hat{F}_n(y/x) = \sum_{i=1}^n I_{\{Y_i \leq y\}} W_{n,i}(x)$ un estimador no paramétrico de $F(\cdot/x)$, donde $W_{n,i}(x)$ representa una función de ponderación (así por ejemplo, los estimadores (4) y (5) corresponden a las elecciones $W_{n,i}(x) = K\left(\frac{x - X_i}{h_n}\right) / \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)$ y $W_{n,i}(x) = \frac{1}{nh_n} K\left(\frac{F_n(x) - F_n(X_i)}{h_n}\right)$, respectivamente).

Dado que, por la observación 3.4, la varianza asintótica del estimador de mínima distancia $\hat{m}_n(x)$ es proporcional a $\int LIF(y, S, F(\cdot/x))^2 dF(y/x)$ (siendo el factor de proporcionalidad $\int K^2(u) du / f(x)$ ó $\int K^2(u) du$, respectivamente, según el estimador no paramétrico elegido sea (4) ó (5)), tiene sentido definir como estimador de este factor

$$\int LIF_n(y, S, \hat{F}_n(\cdot/x))^2 d\hat{F}_n(y/x) = \sum_{j=1}^n LIF_n(Y_j, S, \hat{F}_n(\cdot/x))^2 W_{n,j}(x)$$

donde $LIF_n(Y_j, S, \hat{F}_n(\cdot/x))$ es el estimador de la función de influencia local que resulta de sustituir en la expresión (8) ε por $W_{n,j}(x)$ y $F(\cdot/x)$ por $\hat{F}_n^{(j)}(\cdot/x)$

$$\begin{aligned} LIF_n(Y_j, S, \hat{F}_n(\cdot/x)) &= \\ &= \frac{S[(1 - W_{n,j}(x))\hat{F}_n^{(j)}(\cdot/x) + W_{n,j}(x)\Delta_{Y_j}(\cdot)] - S(\hat{F}_n^{(j)}(\cdot/x))}{W_{n,j}(x)} \end{aligned}$$

con $\hat{F}_n^{(j)}(\cdot/x) = \sum_{i \neq j} I_{\{Y_i \leq y\}} W_{n,i}^{(j)}(x)$ el estimador no paramétrico de $F(\cdot/x)$ construido a partir de las observaciones $\{(X_i, Y_i)\}_{i \neq j}$.

En el supuesto de que el estimador no paramétrico $\hat{F}_n(\cdot/x)$ elegido sea el definido por (4), se obtiene

$$LIF_n(Y_j, S, \hat{F}_n(\cdot/x)) = \frac{S(\hat{F}_n(\cdot/x)) - S(\hat{F}_n^{(j)}(\cdot/x))}{W_{n_j}(x)}$$

siendo entonces el estimador para la varianza asintótica

$$\sum_{i=1}^n \frac{(S(\hat{F}_n(\cdot/x)) - S(\hat{F}_n^{(j)}(\cdot/x)))^2}{W_{n_j}(x)} \frac{\int K^2(u)du}{\hat{f}(x)}$$

con $\hat{f}(x)$ un estimador no paramétrico de $f(x)$.

4. SIMULACIONES

Hemos realizado un estudio de simulación comparativo del estimador de mínima distancia propuesto en este artículo con otros métodos más conocidos de estimación, como pueden ser el método de mínimos cuadrados o la M estimación.

El estudio de simulación ha consistido en la generación de muestras $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$ según el modelo de regresión (1), siendo la curva de regresión aquí considerada $m(x) = 4 \sin(x)$, con X una variable aleatoria uniformemente distribuida en el intervalo $[0, 2\pi]$. Para el error aleatorio ξ hemos considerado distintas elecciones de la función de distribución F_0 (con notación abreviada entre paréntesis): distribución normal standard ($N(0,1)$), mixturas de normales ($F_0(t) = (1-p)\Phi(t) + p\Phi(t/10)$, para valores de $p = 0, 10$ ($0.90N(0,1)$) y $p = 0.20$ ($0.80N(0,1)$)), la distribución de Laplace (LAP), la distribución logística (LOG) y la distribución $\chi_4^2 - 4$ (CHI).

La simulación de estas distribuciones se ha llevado a cabo a partir de la utilización de las subrutinas GGUBS, GGNML y GGEXN de la librería IMSL (con parámetro semilla 11063), junto con la implementación de algunos algoritmos de generación de muestras (así por ejemplo, la distribución logística se ha simulado según el algoritmo de aceptación-rechazo descrito en la página 471 del libro de Devroye (1986)).

Para cada una de las muestras generadas hemos obtenido el estimador de mínimos cuadrados de Nadaraya y Watson

$$\hat{m}_n(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)}$$

el M estimador $\hat{m}_n(x)$, solución con respecto a θ de la ecuación

$$\sum_{i=1}^n \psi(Y_i - \theta) K\left(\frac{x - X_i}{h_n}\right) = 0$$

con $\psi(u) = \Phi(u) - \frac{1}{2}$, si la distribución del error es la normal standard o bien una mixtura de normales, y $\psi(u) = F_0(u) - \frac{1}{2}$ en los demás casos considerados, y el estimador de mínima distancia $\hat{m}_n(x)$, que se obtiene como solución de la ecuación (6) sustituyendo $F(\cdot/x)$ por el estimador piloto no paramétrico (4), siendo la familia paramétrica de funciones de distribución considerada $\{F_\theta\}_{\theta \in \Theta} = \{F_\theta/F_\theta(y) = \Phi(y - \theta)\}_{\theta \in \Theta}$, si la distribución del error es la normal standard o bien una mixtura de normales, y $\{F_\theta\}_{\theta \in \Theta} = \{F_\theta/F_\theta(y) = F_0(y - \theta)\}_{\theta \in \Theta}$ en los otros casos considerados.

El método de aproximación de soluciones utilizado para el cálculo del M estimador y del estimador de mínima distancia, es el conocido como método de dicotomía.

Como función núcleo hemos considerado el llamado núcleo de Epanechnikov, definido por $K(u) = 3/4(1 - u^2)I_{[-1,1]}(u)$, habiendo merecido mayor atención por nuestra parte el problema de la elección del parámetro ventana h_n . Para lo cual hemos hecho un amplio estudio de la expresión asintótica de la función *MISE* (Mean Integrated Squared Error)

$$MISE(h_n) = E \left[\int (\hat{m}_n(x) - m(x))^2 f(x) dx \right]$$

para cada una de las elecciones $\hat{m}_n(x)$, eligiendo como h_n óptimo el resultado de minimizar la expresión resultante en cada caso.

CUADRO 1

Resultados para la media y la varianza de las evaluaciones del *ASE* correspondientes a 1000 muestras de tamaño $n = 25$.

Distribución	Estimador (h_n)	Media	Varianza
N(0,1)	LS (0.911)	$3.60661 \cdot 10^{-1}$	$2.29902 \cdot 10^{-2}$
	ME (0.919)	$3.87901 \cdot 10^{-1}$	$2.70509 \cdot 10^{-2}$
	MD (0.928)	$6.01899 \cdot 10^{-1}$	$8.81574 \cdot 10^{-2}$
0.90N(0,1)	LS (0.911)	2.23748	6.09287
	ME (0.919)	$5.86186 \cdot 10^{-1}$	$1.52207 \cdot 10^{-1}$
	MD (0.928)	1.46784	2.35341
0.80N(0,1)	LS (0.911)	3.85052	$1.11638 \cdot 10^1$
	ME (0.919)	$9.16714 \cdot 10^{-1}$	$6.09620 \cdot 10^{-1}$
	MD (0.928)	2.37092	4.80853
LOG	LS (1.156)	$8.27417 \cdot 10^{-1}$	$2.00105 \cdot 10^{-1}$
	ME (1.135)	$8.24646 \cdot 10^{-1}$	$1.89409 \cdot 10^{-1}$
	MD (1.137)	1.37997	$6.75133 \cdot 10^{-1}$
LAP	LS (1.046)	$5.49622 \cdot 10^{-1}$	$7.16991 \cdot 10^{-2}$
	ME (0.965)	$5.13637 \cdot 10^{-1}$	$7.02340 \cdot 10^{-2}$
	MD (0.981)	$7.77757 \cdot 10^{-1}$	$2.20432 \cdot 10^{-1}$
CHI	LS (1.381)	$1.73897 \cdot 10^1$	$2.26032 \cdot 10^1$
	ME (1.273)	$1.62862 \cdot 10^1$	8.30696
	MD (1.367)	$1.34472 \cdot 10^1$	8.26628

Por último, como medida de discrepancia entre el estimador $\hat{m}_n(x)$ y la curva de regresión $m(x)$, hemos considerado la dada por la función *ASE* (Average Squared Error)

$$ASE(h_n) = \frac{1}{n} \sum_{i=1}^n (\hat{m}_n(X_i) - m(X_i))^2$$

La elección de esta medida se debe únicamente al hecho de su facilidad de cálculo a partir de la muestra inicial y a la equivalencia asintótica existente entre la función *ASE* y la función *MISE* en muchos casos de interés, como puede verse por ejemplo en Härdle y Marron (1985).

La comparación de los distintos estimadores $\hat{m}_n(x)$, obtenidos según el procedimiento descrito, ha consistido por tanto en la generación de muestras de tamaño n de cada uno de los casos considerados, obteniéndose a continuación el promedio y la varianza de las evaluaciones del *ASE* realizadas.

CUADRO 2

Resultados para la media y la varianza de las evaluaciones del *ASE* correspondientes a 100 muestras de tamaño $n = 50$.

Distribución	Estimador (h_n)	Media	Varianza
N(0,1)	LS (0.793)	$2.32570 \cdot 10^{-1}$	$1.17130 \cdot 10^{-2}$
	ME (0.800)	$2.51153 \cdot 10^{-1}$	$1.45101 \cdot 10^{-2}$
	MD (0.807)	$3.25330 \cdot 10^{-1}$	$1.48095 \cdot 10^{-2}$
0.90N(0,1)	LS (0.793)	1.29616	$9.78654 \cdot 10^{-1}$
	ME (0.800)	$3.52154 \cdot 10^{-1}$	$5.37792 \cdot 10^{-2}$
	MD (0.807)	1.02833	1.69160
0.80N(0,1)	LS (0.793)	2.35373	3.30253
	ME (0.800)	$5.04819 \cdot 10^{-1}$	$1.57465 \cdot 10^{-1}$
	MD (0.807)	1.92398	3.00408
LOG	LS (1.006)	$4.67457 \cdot 10^{-1}$	$5.37638 \cdot 10^{-2}$
	ME (0.988)	$4.60421 \cdot 10^{-1}$	$5.41632 \cdot 10^{-2}$
	MD (0.990)	$7.04061 \cdot 10^{-1}$	$1.65645 \cdot 10^{-1}$
LAP	LS (0.911)	$3.43350 \cdot 10^{-1}$	$2.05068 \cdot 10^{-2}$
	ME (0.840)	$3.04922 \cdot 10^{-1}$	$1.49760 \cdot 10^{-2}$
	MD (0.854)	$3.81895 \cdot 10^{-1}$	$2.42988 \cdot 10^{-2}$
CHI	LS (1.202)	$1.71485 \cdot 10^1$	$1.19329 \cdot 10^1$
	ME (1.108)	$1.65790 \cdot 10^1$	5.48375
	MD (1.190)	$1.48003 \cdot 10^1$	5.08592

En el cuadro 1 se recogen los resultados obtenidos a partir de la simulación de 1000 muestras de tamaño $n = 25$. En los cuadros 2 y 3 pueden observarse los resultados relativos a la simulación de 100 muestras de tamaño $n = 50$ y $n = 100$, respectivamente. En todos ellos, por LS, ME y MD, hacemos mención a los estimadores de Nadaraya-Watson, M estimador y estimador de mínima distancia, respectivamente, representándose entre paréntesis el valor del parámetro ventana h_n considerado.

A la vista de los resultados obtenidos, podemos hacer las siguientes consideraciones:

- i) Al aumentar el tamaño muestral observamos unos mejores resultados en cuanto al valor del *ASE* obtenido, pero una comparación de los resultados proporcionados por los distintos estimadores muestra que su comportamiento es similar para $n = 25$, $n = 50$ y $n = 100$.

CUADRO 3

Resultados para la media y la varianza de las evaluaciones del ASE correspondientes a 100 muestras de tamaño $n = 100$.

Distribución	Estimador (h_n)	Media	Varianza
N(0,1)	LS (0.690)	$1.38861 \cdot 10^{-1}$	$1.83667 \cdot 10^{-3}$
	ME (0.697)	$1.48678 \cdot 10^{-1}$	$1.81357 \cdot 10^{-3}$
	MD (0.703)	$1.70767 \cdot 10^{-1}$	$2.20834 \cdot 10^{-3}$
0.90N(0,1)	LS (0.690)	$7.25093 \cdot 10^{-1}$	$2.63198 \cdot 10^{-1}$
	ME (0.697)	$1.73863 \cdot 10^{-1}$	$6.67517 \cdot 10^{-3}$
	MD (0.703)	$5.24150 \cdot 10^{-1}$	$4.89897 \cdot 10^{-1}$
0.80N(0,1)	LS (0.690)	1.28971	$6.14347 \cdot 10^{-1}$
	ME (0.697)	$2.26094 \cdot 10^{-1}$	$1.00721 \cdot 10^{-2}$
	MD (0.703)	$8.83665 \cdot 10^{-1}$	$7.82162 \cdot 10^{-1}$
LOG	LS (0.876)	$3.17001 \cdot 10^{-1}$	$1.76898 \cdot 10^{-2}$
	ME (0.860)	$3.15909 \cdot 10^{-1}$	$1.62037 \cdot 10^{-2}$
	MD (0.862)	$4.27930 \cdot 10^{-1}$	$3.96429 \cdot 10^{-2}$
LAP	LS (0.793)	$2.13369 \cdot 10^{-1}$	$5.48842 \cdot 10^{-3}$
	ME (0.731)	$1.84955 \cdot 10^{-1}$	$4.19726 \cdot 10^{-3}$
	MD (0.743)	$2.19602 \cdot 10^{-1}$	$6.41025 \cdot 10^{-3}$
CHI	LS (1.046)	$1.66082 \cdot 10^1$	5.05569
	ME (0.965)	$1.64769 \cdot 10^1$	3.30376
	MD (1.036)	$1.54213 \cdot 10^1$	2.98150

- ii) Ante la ausencia de contaminación en la muestra y siendo además la distribución del error simétrica (N(0,1), LOG, LAP), los tres estimadores considerados producen resultados similares, observándose el mejor comportamiento del estimador de Nadaraya-Watson bajo normalidad en el error (lo que cabría esperar, dada la bondad de los estimadores de mínimos cuadrados en esta situación).
- iii) Es de destacar el buen comportamiento del M estimador considerado, produciendo éste los mejores resultados salvo en aquellas situaciones en las que la distribución del error es la normal standard o la distribución chi-cuadrado. Ello es debido, a nuestro juicio, a que el haber elegido como función $\psi(u) = F_0(u) - 1/2$ (salvo para las mixturas normales) supone estar suministrando al estimador una cantidad de información adicional imposible de determinar en situaciones prácticas.

- iv) En cuanto al estimador de mínima distancia, vemos que, si la distribución del error no es simétrica (CHI), éste mejora los resultados producidos por el estimador de Nadaraya-Watson y el M estimador en esta situación. Además se observa el buen comportamiento de éste ante la presencia de contaminación en la muestra, mejorando los resultados producidos por el estimador de mínimos cuadrados (lo que cabría sospechar dadas las buenas propiedades de robustez del estimador de mínima distancia).

Por último, en las figuras 1, 2 y 3 se muestran representaciones gráficas de los estimadores de Nadaraya-Watson, M estimador y estimador de mínima distancia de Cramer-von Mises, respectivamente, para una muestra simulada según el procedimiento descrito, siendo la distribución del error aleatorio $F_0(t) = (1 - 0.20)\Phi(t) + 0.20\Phi(t/10)$. Pudiéndose apreciar en la figura 1 la sensibilidad del estimador de Nadaraya-Watson a la presencia de observaciones anómalas (outliers) en la muestra, frente al buen comportamiento del M estimador y del estimador de mínima distancia aquí considerados, como puede observarse en las figuras 2 y 3, respectivamente.

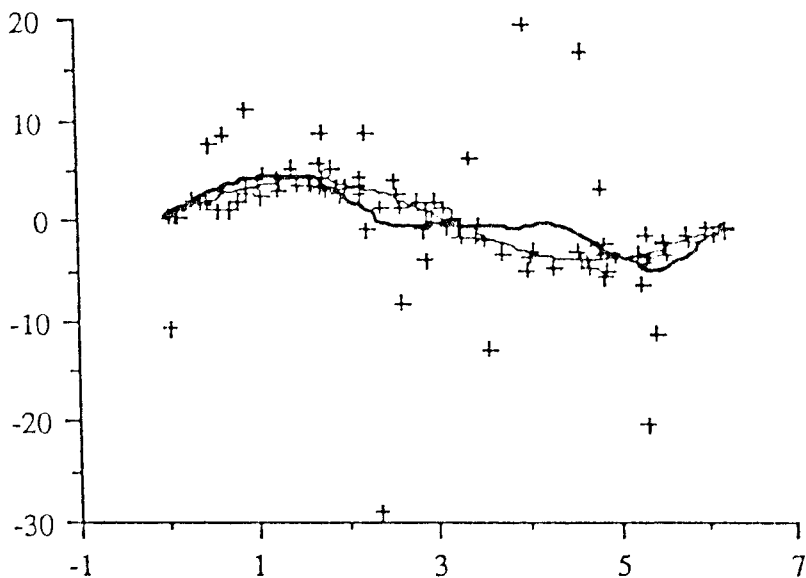


Figura 1. Representación gráfica de la curva de regresión $m(x) = 4 \sin(x)$, con trazo delgado, y del estimador de Nadaraya-Watson, con trazo grueso.

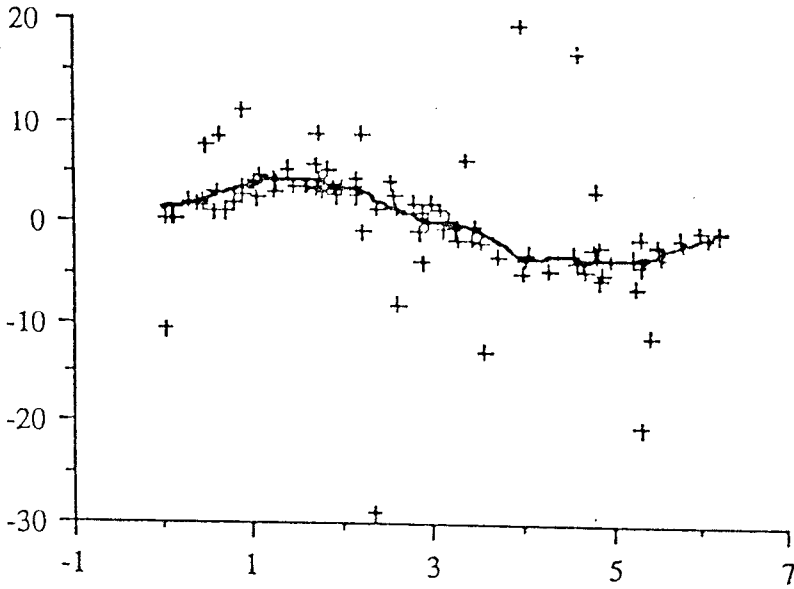


Figura 2. Representación gráfica del M estimador no paramétrico de la curva de regresión $m(x) = 4 \sin(x)$.

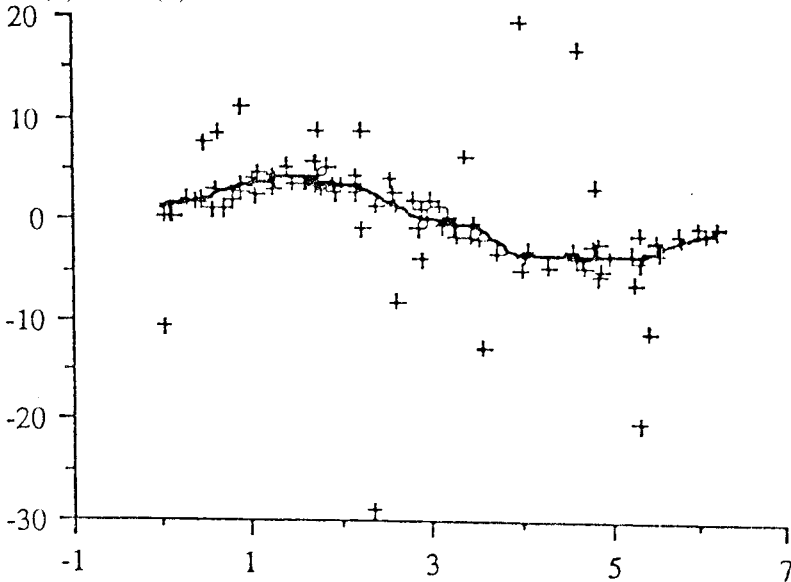


Figura 3. Representación gráfica del estimador de mínima distancia de Cramer-von Mises de la curva de regresión $m(x) = 4 \sin(x)$.

5. APÉNDICE

A) Hipótesis relativas al modelo (1).

A1) La función $\theta \rightarrow \frac{\partial}{\partial \theta} F_\theta(y)$ es continua y uniformemente acotada en $\theta \in \Theta_0$ e $y \in \mathbf{R}$, con $F_\theta(y) = F_0(y - \theta)$.

A2) La función $H(\theta) = \int T(y, \theta, F_0(\cdot - m(x))) dF_0(y - m(x))$ es continuamente diferenciable en Θ_0 , con

$$T(y, \theta, F_0(\cdot - m(x))) = \frac{\partial}{\partial \theta} F_\theta(y) [F_0(y - m(x)) - F_\theta(y)]$$

y

$$\sup_{z \in W_0} E \left[\left| \frac{\partial}{\partial \theta} F_\theta(y) - \frac{\partial}{\partial \sigma} F_\sigma(y) \right|^2 / X = z \right] \leq \text{const} \times |\theta - \sigma|^{1+\eta}$$

para algún $\eta > 0$, siendo W_0 un compacto contenido en W al que pertenece x .

A'1) La función $\theta \rightarrow \frac{\partial}{\partial \theta} F_\theta(y) J(F_0(y - m(x)))$ es continua y uniformemente acotada en $\theta \in \Theta_0$ e $y \in \mathbf{R}$, con $F_\theta(y) = F_0(y - \theta)$.

A'2) La función $H(\theta) = \int T(y, \theta, F_0(\cdot - m(x))) dF_0(y - m(x))$ es continuamente diferenciable en Θ_0 , con

$$T(y, \theta, F_0(\cdot - m(x))) = \frac{\partial}{\partial \theta} F_\theta(y) J(F_0(y - m(x))) (F_0(y - m(x)) - F_\theta(y))$$

y

$$\begin{aligned} \sup_{z \in W_0} E \left[\left| \frac{\partial}{\partial \theta} F_\theta(y) J(F(y/x)) - \frac{\partial}{\partial \sigma} F_\sigma(y) J(F(y/x)) \right|^2 / X = z \right] &\leq \\ &\leq \text{cte} \times |\theta - \sigma|^{1+\eta} \end{aligned}$$

para algún $\eta > 0$, siendo W_0 un compacto contenido en W al que pertenece x .

B) Hipótesis relativas al estimador (4).

B1) K es una densidad de media cero.

B2) $h_n \rightarrow 0$, $nh_n \rightarrow \infty$ y $nh_n^5 \rightarrow 0$.

- B3) Las funciones f , función de densidad marginal de la variable X que se supone existente, $F_0(\cdot - m(x))$, $h(x) = E[T(Y, \theta, F_0(\cdot - m(x)))/X = x]$ y $g(x) = E[Q(Y)/X = x]$, con $Q(y) = -\int_{-\infty}^y \frac{\partial}{\partial \theta} |_{\theta=m(x)} F_\theta(z) dF_0(z - m(x))$ y $T(y, \theta, F_0(\cdot - m(x)))$ como en la hipótesis (A2), tienen derivadas de segundo orden continuas en un entorno de x , uniformemente acotadas respecto de $\theta \in \Theta_0$.
- B'1) $f(x) > 0$, donde f es la densidad de X supuesta existente.
- B'2) $\sup |f'(u)| < \infty$.
- B'3) Si $h(y, u) = \frac{\partial}{\partial u} H(y, u)$ con H la función de distribución conjunta de (Y, X) entonces $\sup_{y,u} h(y, u) < \infty$, $\sup_{y,u} |\frac{\partial}{\partial u} h(y, u)| < \infty$ y además $\sup_{y,u} |\frac{\partial^2}{\partial u^2} h(y, u)| < \infty$.
- B'4) $\int |u|K^2(u)du < \infty$, $\int K^2(u)du < \infty$ y K de variación acotada.
- B'5) $h_n \log n \rightarrow 0$ y $nh_n^3(\log n)^{-6} \rightarrow \infty$.
- C) Hipótesis relativas al estimador (5).
- C1) K es una densidad dos veces continuamente diferenciable con soporte compacto y verificando $\int K(u)du = 1$ y $\int uK(u)du = 0$.
- C2) $h_n \rightarrow 0$, $nh_n^3 \rightarrow \infty$ y $nh_n^5 \rightarrow 0$.
- C3) Las funciones $h(u) = E[T(Y, \theta, F_0(\cdot - m(x)))/F(X) = F(x) = u]$ y $g(u) = E[Q(Y)/F(X) = F(x) = u]$, con $Q(y)$ y $T(y, \theta, F_0(\cdot - m(x)))$ como en la hipótesis (B3), tienen derivadas de segundo orden continuas en un entorno de $F(x) = u$, uniformemente acotadas respecto de $\theta \in \Theta_0$.
- C4) $\sup_{|t-s| \leq \sigma} |F_0(t - m(z)) - F_0(s - m(z))| = O((\ln \delta^{-1})^{-1})$ cuando $\delta \rightarrow 0$, uniformemente en un entorno U de x (esto se verifica por ejemplo cuando $F_0(\cdot - m(x))$ es una función Hölder continua de orden positivo).
- C5) $\sup_{z \in U} \sup_y |F_0''(y - m(z))| < \infty$, con U entorno de x .
- C'1) $\inf f(u) > 0$ en un entorno U de x .
- C'2) $\sup |f'(u)| < \infty$ en un entorno U de x .
- C'3) $\sup_{u \in U} \sup_y h(y, u) < \infty$, $\sup_{u \in U} \sup_y |\frac{\partial}{\partial u} h(y, u)| < \infty$ y además $\sup_{u \in U} \sup_y |\frac{\partial^2}{\partial u^2} h(y, u)| < \infty$.
- C'4) $h_n \log n \rightarrow 0$ y $nh_n^3(\log n)^{-6} \rightarrow \infty$.

6. AGRADECIMIENTOS

Nuestro agradecimiento a dos referees anónimos. Su lectura crítica contribuyó notablemente en la mejor presentación de este artículo. Asimismo agradecemos al profesor Ricardo Fraiman por sus comentarios en la elaboración de la versión revisada.

7. REFERENCIAS

- [1] **Cheng, P.E. and Cheng, K.F.** (1987). "Robust nonparametric estimation of a regression function". *Sankhyā* B 49, 9–22.
- [2] **Collomb, G.** (1976). "Estimation non parametrique de la régression para la méthode du noyau". Tesis Doctoral. Univ. de Toulouse. Francia.
- [3] **Dabrowska, D.M.** (1987). "Nonparametric regression with censored survival time data". *Scand. J. Statist.* 14, 181–197.
- [4] **Devroye, L.** (1986). "Non uniform random variate generation". Springer-Verlag.
- [5] **González Manteiga, W.** (1990). "Asymptotic normality of generalized functional estimators dependent on covariables". *J. Statist. Plann. Inference*, 24, 377–390.
- [6] **Hampel, F.R.** (1968). "Contributions to the theory of robust estimation". Ph. D. thesis. University of California, Berkely.
- [7] **Hampel, F.R.** (1974). "The influence curve and its role in robust estimation". *JASA*, 69, 383–393.
- [8] **Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A.** (1986). "Robust statistics: The approach based on influence functions". Wiley.
- [9] **Härdle, W.** (1984). "Robust regression function estimation". *J. Multivariate Anal.* 14, 169–180.
- [10] **Härdle, W.** (1990). "Applied nonparametric regression". Cambridge University Press.
- [11] **Härdle, W. and Gasser, T.** (1984). "Robust nonparametric function fitting". *J. Rōy. Statist. Soc.*, B 46, 42–51.
- [12] **Härdle, W., Janssen, P. and Serfling, R.** (1988). "Strong uniform consistency rates for estimators of conditional functionals". *Ann. Statist.*, 16, 1428–1449.
- [13] **Härdle, W., and Marron, S.** (1985). "Optimal bandwidth selection in nonparametric regression function estimation". *Ann. Statist.*, 13, 1465–1481.

- [14] **Horváth, L. and Yandell, B.S.** (1988). "Asymptotics of conditional empirical processes". *J. Multivariate Anal.*, 26, 184–206.
- [15] **Huber, P.J.** (1964). "Robust estimation of a location parameter". *Ann. Math. Statist.*, 35, 73–101.
- [16] **Huber, P.J.** (1981). "Robust statistics". Wiley.
- [17] **Nadaraya, E.A.** (1964). "On estimating regression". *Theory Probab. Appl.*, 9, 141–142.
- [18] **Parr, W.C. and Schucany, W.R.** (1980). "Minimum distance and robust estimation". *JASA*, 75, 616–624.
- [19] **Parzen, E.** (1962). "On estimation of a probability density function". *Ann. Math. Statist.*, 31, 1065–1076.
- [20] **Rosenblatt, M.** (1956). "Remarks on some nonparametric estimators of a density function". *Ann. Math. Statist.*, 27, 832–837.
- [21] **Stute, W.** (1984). "Asymptotic normality of nearest neighbor regression function estimates". *Ann. Statist.*, 12, 917–926.
- [22] **Stute, W.** (1986a). "Parameter estimation in smooth empirical processes". *Stochast. Proc. and their applications*, 22, 223–244.
- [23] **Stute, W.** (1986b). "Conditional empirical processes". *Ann. Statist.*, 14, 638–647.
- [24] **Watson, G.S.** (1964). "Smooth regression analysis". *Sankhyā A* 26, 359–372.
- [25] **Yang, S.S.** (1981). "Linear functions of concomitants of order statistics with applications to nonparametric estimation of a regression function". *JASA*, 76, 658–662.

