TRABAJOS DE ESTADISTICA
Vol. 6, núm. 2, 1991, pp. 11 a 31

CONSISTENCIA DE UN ESTIMADOR NO PARAMETRICO, RECURSIVO, DE LA REGRESION BAJO CONDICIONES GENERALES

JUAN MANUEL VILAR FERNÁNDEZ
Dpto. de Matemáticas
Facultad de Informática
Castro de Elviña, 76
La Coruña, 15071

RESUMEN

Se define un estimador no paramétrico, recursivo, de la función de regresión $r(x) = E(Y/X = x)$, que se calcula a partir de un conjunto de n observaciones $\{(X_1, Y_i): i = 1, ..., n\}$ del vector aleatorio (X, Y). Bajo la hipótesis de que los datos son idénticamente distribuidos pero no necesariamente independientes, lo que permite utilizar el estimador definido para estimar la función de autorregresión de una serie de tiempo, se obtienen resultados sobre la consistencia puntual débil (en probabilidad) y fuerte (completa) del estimador definido.

Finalmente, se presentan ejemplos de utilización del estimador con datos simulados.

Palabras clave: función de regresión, estimación no paramétrica, estimación recursiva, consistencia.
Clasificación AMS (1980): 62G05.

SUMMARY

With a initial sample of size $n: \{(X_i, Y_i): i = 1, ..., n\}$ of a random vector (X, Y) a general recursive nonparametric estimator of a regression function, $r(x) = E(Y/X = x)$ is obtained. Assume the data are identically distributed but non necessarily independent. So, the defined estimator may be used to estimate the autoregression function of a time series.

Recibido: Abril 1990.
Revisado: Enero 1991.

Weak and strong pointwise consistency are obtained for such estimator. Finally some examples of simulation are presented.

Key words: regresión function, nonparametric estimation, recursive estimate, consistency.
Classification A.M.S.: 62G05.

1. INTRODUCCION

La estimación de la función de regresión $r(x) = E(Y/X = x)$, siendo (X, Y) dos variables aleatorias que toman valores en R^d y R, respectivamente, es de gran interés, ya que esta curva describe la relación que existe entre la variable explicativa (X) y la variable respuesta (Y), por lo que ha sido ampliamente estudiado.

Siendo $(X_i, Y_i)_{i=1}^n n$ observaciones que siguen el modelo $Y_i = r(X_i) + \varepsilon_i$, donde ε_i son los errores de observación, el problema de estimar la curva de regresión a partir de estos datos puede enfocarse, esencialmente, desde dos puntos de vista diferentes. Uno, es el enfoque paramétrico, más estudiado y utilizado, en él se asume que la función de regresión tiene una forma funcional específica, determinada por unos parámetros y, por tanto, el problema radica en estimar estos parámetros.

Un enfoque menos rígido es el no paramétrico, en el que no se presupone nada acerca de la forma funcional de la curva de regresión, a lo sumo, alguna propiedad de tipo general (es acotada, diferenciable...). Por ello, los métodos no paramétricos constituyen una herramienta útil y sencilla para explorar la relación entre las dos variables, pudiéndose utilizar para indicar el modelo paramétrico simple de la regresión cuando no se sabe nada acerca de éste; para localizar datos espúreos; y, para obtener predicciones de la variable respuesta sin necesidad de construir un modelo paramétrico.

Los procedimientos más flexibles de estimación no paramétrica de la regresión se han desarrollado en los últimos veinticinco años y pueden citarse entre los últimos trabajos sobre el tema los siguientes: Bierens (1983), Collomb (1985), Hardle (1988) y Vilar Fernández (1989a).

Puede verse en Hardle (1988) que la mayoría de los estimadores no paramétricos de la función de regresión se obtienen como una suma ponderada de los Y_i muestrales, ajustándose al siguiente formato:

$$r_n(x) = \sum_{i=1}^{n} W_{n,i}(x) Y_i \tag{1}$$

Siendo $\{W_{n,i}(x)\}_{i=1}^{n}$ una sucesión de pesos que depende del conjunto de datos $\{X_i\}_{i=1}^{n}$, siendo el peso $W_{n,j}(x)$ mayor cuanto más próximo está X_j de x.

En este trabajo se define una clase de estimadores no paramétricos de la regresión que se ajusta a la idea del modelo (1) y es de tipo recursivo. En el apartado II se estudian propiedades de consistencia débil y fuerte; y, en el apartado IV se exponen algunos ejemplo de su utilización trabajando con datos simulados.

A partir de un conjunto de n observaciones $\{(X_i, Y_i)\}_{i=1}^{n}$ se define el siguiente estimador de $r(x)$:

$$r_n(x) = \frac{\sum_{i=1}^{n} b_i^{d\tau} \delta_{m(i)}(x, X_i) Y_i}{\sum_{i=1}^{n} b_i^{d\tau} \delta_{m(i)}(x, X_i)} \tag{2}$$

Siendo $\{\delta_{m(n)}(x, u)\}$ una sucesión funcional definida en $R^d \times R^d$ con valores en R, verificando las siguientes propiedades:

D.1. Si $|x - u| < b_n$ entonces $\delta_{m(n)}(x, u) = 0(m(n)^d)$ y si $|x - u| \geqslant b_n$ entonces $\delta_{m(n)}(x, u) = 0(m(n)^d/n)$. Siendo $\delta_{m(n)}(x, u) \geqslant 0$.

D.2. Para toda función g integrable se verifica

$$\lim_{n \to \infty} \int \delta_{m(n)}(t, x) g(x)\, dx = g(t)$$

La sucesión $\{b_n\}$ que se tomará igual a $m(n)^{-1}$, llamada «parámetros de suavización» verifica las siguientes condiciones:

V.1. $b_n \to 0$, $nb_n^d \to \infty$ cuando $n \to \infty$.

V.2. Para $s = d\tau$ el $\lim_{n \to \infty} \sum_{i=1}^{n} \left(\frac{b_i}{b_n}\right)^s = \theta_s < \infty$.

La primera condición es clásica y la segunda no es muy fuerte, verificándola la sucesión de parámetros más utilizada que es de la forma $b_n = Bn^{-\gamma}$ con $0 < \gamma < 1/d$, pues en este caso $\theta_s = 1/(1 - \gamma s)$.

La elección del parámetro b_n es muy importante ya que influye mucho en la estimación. Valores grandes de b_n proporcionan estimaciones sobresuavizadas (sesgo grande) y valores pequeños de b_n proporcionan estimaciones poco suavizadas (varianza grande).

El parámetro τ está comprendido entre 0 y 1, e influye en el error medio cuadrático del estimador de forma análoga a como lo hace el parámetro de suavizado, aunque su efecto es mucho menor.

El estimador definido $r_n(x)$ es recursivo ya que verifica la siguiente relación:

$$r_{n+1}(x) = S_{n+1}^{-1}(r_n(x)S_n + b_{n+1}^{d\tau}\delta_{m(n+1)}(x, X_{n+1})Y_{n+1})$$

siendo

$$S_n = \sum_{i=1}^{n} b_i^{d\tau}\delta_{m(i)}(x, X_i)$$

La propiedad de recursividad es interesante desde un punto de vista computacional porque permite reducir los cálculos al aumentar el conjunto de observaciones, ya que para calcular nuevas estimaciones no hay que reinicializar el proceso sino que se obtienen como una combinación lineal de las últimas estimaciones y de los nuevos datos.

El estimador definido es un caso particular de la versión recursiva del estimador general (1), tomando como pesos:

$$W_{ii}(x) = \frac{b_i^{d\tau}\delta_{m(i)}(x, X_i)}{\sum_{j=1}^{n} b_j^{d\tau}\delta_{m(j)}(x, X_j)}$$

y, a su vez, es bastante general ya que distintas elecciones de la sucesión de funciones de ponderación $\{\delta_{m(n)}(x, u)\}$ permiten obtener las versiones recursivas de estimadores no paramétricos clásicos como el regresograma, el estimador núcleo o el de desarrollos ortogonales.

En la mayoría de los trabajos sobre estimación no paramétrica de la regresión se asume la hipótesis de que los datos son independientes, lo que impide que sean utilizados en el estudio de las series de tiempo, donde son de gran interés los estimadores recursivos ya que al obtenerse los datos con el transcurso del tiempo será normal que el conjunto de observaciones aumente. Por este motivo, en los resultados obtenidos en este trabajo se asume que los datos verifican alguna condición de dependencia de tipo mixing.

Fundamentalmente se trabajará con la condición de dependencia «fuertemente mixing» (α-mixing) que extiende al de independencia y era introducido por M. Rosenblatt en 1956, es el siguiente: «Sea $\{Z_t : t \in \mathbb{Z}\}$ una sucesión de v.a. y F_a^b la σ-álgebra generada por las v.a. X_t con $a \leqslant t \leqslant b$.

Llamemos $\alpha(n) = \sup\{|P(AB) - P(A)P(B)| : A \in F_{-\infty}^j ; B \in F_{j+n}^{\infty}\}$ entonces Z_t es α-mixing si la sucesión de coeficientes α-mixing $\{\alpha(n)\}$ es monótona decreciente convergiendo a cero.»

Este concepto introduce la idea de que la dependencia entre Z_n y Z_{t+n} se va anulando cuando n crece. Y de entre las condiciones de dependencia más utilizadas es la más débil, verificándola, entre otros, los procesos lineales construidos a partir de v.a. i.i.d. absolutamente continuas y funcionales no lineales de procesos gaussianos exigiendo condiciones no muy restrictivas a la densidad espectral (Bradley, 1986).

En este contexto una importante aplicación del estimador definido es su utilización para predecir los futuros valores de una serie de tiempo. Sea $Z_1, \ldots, Z_n$ una muestra de una serie de tiempo estacionaria y real $\{Z_t : t \in \mathbb{Z}\}$ y se desea estimar el valor de Z_{n+1}, utilizando como función de predicción la autorregresión de orden d:

$$r_d(\bar{z}) = E(Z_{t+1}/(Z_{t-k+1}, \ldots, Z_t) = \bar{z}) \quad \text{con } \bar{z} \in \mathbb{R}^d$$

que será la teórica en el caso de que el proceso sea markoviano de orden d. Su estimación utilizando (2) viene dada por:

$$r(\bar{z}) = \frac{\sum_{i=1}^{n} b_i^{d\tau} \delta_{m(i)}(\bar{z}, \bar{Z}_i) Z_{i+1}}{\sum_{i=1}^{n} b_i^{d\tau} \delta_{m(i)}(\bar{z}, \bar{Z}_i)} \tag{3}$$

donde

$$X_i = \bar{Z}_i = (Z_{1-d+1}, \ldots, Z_i), \quad Y_i = Z_{i+1}$$

Un estudio de este problema para el caso particular de que $\delta_{m(n)}(x, u) = b_n^{-d} K((x - u)/b_n)$, siendo $K(u)$ una función núcleo, puede verse en Vilar Fernández (1989b). En él se calcula el error medio cuadrático y la normalidad asintótica del estimador así como la influencia del parámetro τ en la estimación.

2. PROPIEDADES DE CONSISTENCIA

En este apartado se demuestran propiedades de consistencia puntual débil y fuerte del estimador $r_n(x)$ para casi todo (μ) $x \in \mathbb{R}^d$, es decir, sobre un subconjunto $S \subset \mathbb{R}^d$ tal que $\mu(S) = 1$, siendo μ la medida de probabilidad asociada a la v.a. X.

Se supone que el vector aleatoria (X, Y) es absolutamente continuo con función de densidad $f(x, y)$.

A. CONSISTENCIA DEBIL

Se verifica el siguiente resultado sobre convergencia en probabilidad del estimador $r_n(x)$.

Teorema 1

Bajo las siguientes hipótesis

H.1. $E|Y| < \infty$.

H.2. La sucesión (X_i, Y_i) es estrictamente estacionaria y fuertemente mixing, respecto a unos coeficientes $\alpha(h)$ tales que $\sum_h \alpha(h)^{\delta/(2+\delta)} < \infty$ para algún $\delta > 0$.

Entonces $r_n(x) \to r(x)$ en probabilidad c.p.d. (μ) (x).

Se utilizará la siguiente desigualdad en la demostración del teorema 1.

Lema 1 (Desigualdad de Davydov)

«Sean ξ y η dos v.a. respecto a las σ-álgebras F_1^i y F_1^j, respectivamente, $i, j = 1, 2, ..., n$. Siendo $E|\xi|^{2+\delta} < \infty$ y $E|\eta|^{2+\delta} < \infty$ para algún $\delta > 0$, entonces para $i \neq j$ se verifica

$$|\mathrm{cov}\,(\xi, \eta)| \leqslant 8(E|\xi|^{2+\delta}E|\eta|^{2+\delta})^{\delta/(2+\delta)}(\alpha|i-j|)^{\delta/(2+\delta)}.»$$

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

Demostración del Teorema 1

Se verifica que

$$r_n(x) = S_n^1(x)(S_n^2(x))^{-1} \tag{4}$$

siendo

$$S_n^1(x) = \frac{\sum_{i=1}^{n} b_i^{d\tau}\delta_{m(i)}(x, X_i)Y_i}{E\left(\sum_{i=1}^{n} b_i^{d\tau}\delta_{m(i)}(x, X_i)\right)} \quad ; \quad S_n^2(x) = \frac{\sum_{i=1}^{n} b_i^{d\tau}\delta_{m(i)}(x, X_i)}{E\left(\sum_{i=1}^{n} b_i^{d\tau}\delta_{m(i)}(x, X_i)\right)}$$

Se demuestra en primer lugar que $S_n^1(x) \xrightarrow{p} r(x)$ c.p.d. $(\mu)(x)$

$$S_n^1(x) = (S_n^1(x) - ES_n^1(x)) + ES_n^1(x) = A_n(x) + D_n(x) \tag{5}$$

de **D.2** se sigue que $E(\delta_{m(n)}(x, X_n)Y_n \to f(x)r(x)$ cuando $n \to \infty$ y por **V.2** aplicando el lema de Toeplitz respecto a la disposición triangular

$$a_{i,\,n} = b_i^{d\tau}/nb_n^{d\tau} \quad i = 1, 2, ..., n$$

se obtiene

$$\sum_{i=1}^{n} (b_i^{d\tau}/nb_n^{d\tau})E(\delta_{m(i)}(x, X_i)Y_i) \to r(x)f(x)$$

análogamente se demuestra que

$$\sum_{i=1}^{n} (b_i^{d\tau}/nb_n^{d\tau})E(\delta_{m(i)}(x, X_i)) \to f(x)$$

y de ambos resultados se concluye que

$$D_n(x) \to r(x) \tag{6}$$

Se demuestra ahora que la parte aleatoria $A_n(x)$ converge en probabilidad a cero:

$$A_n(x) = C_n^{-1}(x) \sum_{i=1}^{n} b_i^{d\tau} T_{i,\,n}(x) \tag{7}$$

siendo

$$C_n(x) = \sum_{i=1}^{n} b_i^{d\tau} E(\delta_{m(i)}(x, X_i))$$

$$T_{i,\,n}(x) = \delta_{m(i)}(x, X_i)Y_i - E(\delta_{m(i)}(x, X_i)Y_i)$$

Sea N fijo y llamemos

$$Y_i^1 = Y_i I_{|Y_i| \leq N} \quad ; \quad Y_i^2 = Y_i - Y_i^1 \tag{8}$$

sustituyendo en (7) tendremos $A_n(x) = A_n^1(x) + A_n^2(x)$.

Veamos que ambos sumandos convergen a cero: De la desigualdad de Tchebyshev se sigue que

$$P(|A_n^1(x)| > \varepsilon) \leqslant (\varepsilon^2 C_n^2(x))^{-1}\left(\sum_{i=1}^{n} b_i^{2d\tau} \operatorname{var}(T_{i,n}^1(x) + 2\sum_{i=1}^{n}\sum_{h=1}^{n-i}(b_i b_{i+h})^{\tau d}\right.$$
$$\left. \operatorname{cov}(T_{i,n}^1(x) T_{i+h,n}^1(x))\right) = V_n + C_n \tag{9}$$

De **D.1** se sigue que

$$V_n = (\varepsilon^2 C_n^2(x))^{-1}\left(\sum_{i=1}^{n} b_i^{2d\tau} \operatorname{var}(T_{i,n}^1(x)\right) \leqslant$$

$$\leqslant 0\left(\frac{N}{\varepsilon^2 C_n^2(x)} \frac{\sum_{i=1}^{n} b_i^{d\tau} E(\delta_{m(i)}(x, X_i) Y_i^1) b_i^{d(\tau-1)}}{\sum_{i=1}^{n} b_i^{d\tau} E(\delta_{m(i)}(x, X_i))}\right) \leqslant$$

$$\leqslant 0\left(\frac{b_n}{\sum_{i=1}^{n} b_i^{d\tau}} \frac{\sum_{i=1}^{n} b_i^{d\tau} E(\delta_{m(i)}(x, X_i) Y_i^1)}{\sum_{i=1}^{n} b_i^{d\tau} E(\delta_{m(i)}(x, X_i))}\right) \tag{10}$$

de **V.2** se sigue que el primer factor es de orden $0(nb_n^d)^{-1}$ que converge a cero por **V.1.**

Por un razonamiento análogo al efectuado en (6) se verifica que el segundo factor converge a $r_N(x) = E(Y^1/X = x)$, lo que permite concluir que V_n converge a cero.

Utilizando el lema 1 y **H.2** se sigue que

$$C_n \leqslant 0C_n^{-2}(x) \sum_{i=1}^{n}\sum_{h=1}^{n-1} (b_i b_{i+h})^{\tau d} \alpha(h)^{\delta/(2+\delta)} \leqslant$$

$$\leqslant 0\left(\frac{b_n^{\tau d}}{C_n(x)} \frac{\sum_{i=1}^{n} b_i^{d\tau}}{C_n(x)} \sum_h (\alpha(h))^{\delta/(2+\delta)}\right) \to 0 \tag{11}$$

De (10) y (11) se sigue que $A_n^1(x) \xrightarrow{p} 0$ c.p.d. $(\mu)(x)$.

Por otro lado, utilizando la desigualdad de Markov se obtiene que:

$$P(|A_n^2(x)| > \varepsilon) \leqslant \varepsilon^{-1} E(|A_n^2(x)|) \leqslant$$

$$\leqslant \frac{1}{\varepsilon} \frac{\sum_{i=1}^{n} b_i^{d\tau} E(\delta_{m(i)}(x, X_i) g_N(X_i))}{\sum_{i=1}^{n} b_i^{d\tau} E(\delta_{m(i)}(x, X_i))}$$

siendo

$$g_N(X_i) = E(|Y| I_{|Y|>N} / X = x).$$

De **H.1** y del teorema de la convergencia monótona se sigue que $E|g_N(x)| \to 0$ cuando $N \to \infty$, por tanto, tomando N suficientemente grande, se verifica que $P(|A_n^2(x)| > \varepsilon) \to 0$ y, por tanto, $A_n^2(x) \xrightarrow{p} 0$ *con lo que se concluye que* $A_n(x) \xrightarrow{p} 0$ c.p.d. $(\mu)(x)$.

Finaliza la demostración probando por un razonamiento análogo que $S_n^2(x) \xrightarrow{p} r(x)$ c.p.d. $(\mu)(x)$. □

B. CONSISTENCIA FUERTE

Se utiliza en este apartado el concepto de convergencia completa cuya definición es la siguiente:

«Una sucesión de v.a. $\{X_i : i \in \mathbb{N}\}$ converge completamente a cero ($X_n \xrightarrow{\text{co.}} 0$) si para todo $\varepsilon > 0$ se verifica que $\Sigma P(X_i | > \varepsilon) < \infty$.»

De esta definición es inmediato que aplicando el lema de Borel-Cantelli que la consistencia completa implica la consistencia casi segura.

La demostración de que $r_n(x) \xrightarrow{\text{co.}} r(x)$ se basa en la utilización de desigualdades de tipo exponencial análogas a las de Bernstein, pero para datos dependientes. Se realiza el estudio de la convergencia completa bajo dos tipos de dependencia: uniformemente mixing (φ-mixing) y fuertemente mixing (α-mixing); bajo el supuesto de que $\tau = 0$, que es lo más usual en la práctica.

Teorema 2

Bajo las siguientes hipótesis:

H.3. $|Y| \leqslant M < \infty$.

H.4. Se verifica una de las condiciones **A** o **B**:

A. La sucesión de pares (X_i, Y_i) es estrictamente estacionaria y φ-mixing respecto a los coeficientes $\varphi(h)$ tales que:

i. Existe una sucesión de números naturales $\{k_n\}$ tal que $n\varphi(k_n)/k_n \leqslant A$ para todo $n \in \mathbb{N}$, $1 \leqslant k_n \leqslant n$.

ii. $nb_n^d/k_n \log(n) \to \infty$ cuando $n \to \infty$.

B. La sucesión de pares (X_i, Y_i) es estrictamente estacionaria y geométricamente α-mixing respecto a los coeficientes $\alpha(h)$ que verifican $(nb_n^{d(1+\eta)})^{1/4}/\log(n) \to \infty$ para algún $\eta > 0$.

Entonces $r_n(x) \to r(x)$ completamente c.p.d. $(\mu)(x)$.

Se utilizarán las siguientes desigualdades de tipo exponencial en la demostración del teorema 2.

Lema 2 (Desigualdad de Collomb, 1985, para procesos φ-mixing)

«Para todo $n \in \mathbb{N}$ sea $\{\xi_{n,i} : i = 1, \ldots, n\}$ una sucesión de v.a. satisfaciendo $E(\xi_{n,i}) = 0$; $|\xi_{n,i}| \leqslant d_n$; $E|\xi_{n,i}| \leqslant g_n$; $E|\xi_{n,i}^2| \leqslant h_n$ para $i = 1, \ldots, n$. Son φ-mixing siendo la sucesión $\{\varphi(i) : i \in \mathbb{N}\}$ no dependiente de n. Entonces para todo $\varepsilon > 0$ se verifica:

$$P\left(|\sum_{i=1}^{n} \xi_{n,i}| \geqslant \varepsilon\right) \leqslant 2\exp(6v^2 n(h_n + d_n g_n \Sigma\, \varphi(i)) - v\varepsilon + 3e^{1/2} n\varphi(k_n)/k_n)$$

para todo $v > 0$, $k_n \in \mathbb{N}$ con $1 \leqslant k_n \leqslant n$, satisfaciendo $vk_n d_n \leqslant 1/4$.»

Lema 3 (Desigualdad de Doukhan, Leon y Portal (1984) para procesos α-mising)

«*Para todo* $n \in \mathbb{N}$ sea $\{\xi_{n,i} : i = 1, \ldots, n\}$ una sucesión de v.a. geométricamente α-mixing satisfaciendo $E(\xi_{n,i}) = 0$ y $|\xi_{n,i}| \leqslant 1$. Dado $0 < \eta < 1$, denotamos $v = 2/(1-\eta)$ y $\sigma = \sup\{\xi_{n,iv},\ i = 1, \ldots, n\}$ donde $\xi_{n,iv}^{v} = E|\xi_{n,i}|^{v}$.

Entonces existen constantes C_1 y C_2 las cuales dependen solamente de los coeficientes mixing $(\alpha(i))$, tales que:

$$P\left(| \sum_{i=1}^{n} \xi_{n,i}| \geqslant \varepsilon \right) \leqslant (C_1/\eta) \exp\left(\frac{-C_{2,n}\varepsilon^{1/2}}{n^{1/4}\sigma^{1/2}} \right)$$

donde $C_{2,n} = C_2$ si $n^{1/2}\sigma < 1$ y $C_{2,n} = C_2 n^{1/4}\sigma^{1/2}$ si $n^{1/2}\sigma > 1$.»

Demostración del teorema 2

A. Para procesos uniformemente mixing

Razonando como en el teorema 1, es suficiente con probar que $A_n(x) \xrightarrow{\text{co.}} 0$ c.p.d. $(\mu)(x)$.

Para ello se calcula una cota de

$$P(|A_n(x)| > \varepsilon) = P\left(| \sum_{j=1}^{n} Z_{j,n} | > \varepsilon \right)$$

siendo

$$Z_{i,n} = \left(\sum_{j=1}^{n} E(\delta_{m(j)}(x, X_j)) \right)^{-1} (Y_i \delta_{m(i)}(x, X_i) - E(Y_i \delta_{m(i)}(x, X_i)))$$

Por ser (X_i, Y_i) φ-mixing (α-mixing) también lo es Z_i respecto a los mismos coeficientes y se encuentra en las hipótesis del lema 2 con $d_n = 0(nb_n^d)^{-1}$; $g_n = 0(1/n)$; $h_n = 0(n^2 b_n^d)^{-1}$.

Aplicando dicho lema con $v = (k_n d_n)^{-1}$, de **H.4.A** se sigue que

$$P(|A_n(x)|) > \varepsilon) \leqslant 2 \exp(-C\varepsilon n b_n^d / k_n).$$

Y del criterio logarítmico de convergencia de series se sigue que $\sum_n P(|A_n(x)| > \varepsilon)$ converge si $nb_n^d/k_n \log(n) \to \infty$ cuando $n \to \infty$, que es **H.4.A.ii.** □

B. Para procesos fuertemente mixing

En el supuesto de que la dependencia fuese fuertemente mixing, utilizando el lema 3 se obtiene que: $\sigma = 0(nb_n^{d((1+\eta)/2)})$ y, por tanto, $P(|A_n(x)| > \varepsilon) \leqslant 0 \exp(-n^{1/4} b_n^{d(1+\eta)/4)})$ y, utilizando el criterio de convergencia de series anterior se deduce que $A_n(x) \xrightarrow{\text{co.}} 0$ c.p.d. $(\mu)(x)$ si $n^{1/4}b_n^{d(1+\eta)/4}/\log(n) \to \infty$ que es **H.4.B.** □

Comentarios

1. En el supuesto de dependencia φ-mixing, son elecciones de interés de k_n verificando **H.4.A** las siguientes, propuestas por G. Collomb:

• Si (X_i, Y_i) son m-dependientes, esto es, $\varphi_j = 0$ si $j \geqslant m + 1$, entonces $k_n = m + 1$ y la condición **H.4.A.ii** es: $nb_n^d/\log(n) \to \infty$.

• Si (X_i, Y_i) son geométricamente φ-mixing, esto es, $\varphi_j = \alpha\beta^j$ $0 < \alpha < \infty$, $0 < \beta < 1$, entonces $k_n = C\log(n)$ con $C > -1/\log(\beta)$ y la condición **H.4.A.ii** es: $nb_n^d/(\log(n))^2 \to \infty$.

• Si los coeficientes φ-mixing son de la forma $\varphi_j = \alpha j^{-\beta}$ $0 < \alpha < \infty$, $0 < \beta < 1$, entonces $k_n = n^{1/(1+\beta)}$ la condición **H.4.A.ii** es: $n^{\beta/(1+\beta)}b_n^d/\log(n) \to \infty$.

2. La condición de ser φ-mixing es bastante restrictiva, sobre todo si se trabaja con series de tiempo ya que los procesos autorregresivos o de tipo ARMA no lo son. De hecho como señalaron Ibragimov-Linnick (1971), los procesos estacionarios gaussianos son φ-mixing si y sólo si son m-dependientes. Trabajando con procesos markovianos la condición es bastante buena y puede verse en Rosenblatt (1971) que bajo condiciones no muy restrictivas (condición en norma L^p/condición de Doeblin) el proceso es geométricamente φ-mixing. En cualquier caso como indicaron Collomb y Hardle (1984) el contexto natural de trabajar en series de tiempo es el de los procesos fuertemente mixing, que como indicamos en el apartado 1 es muy poco restrictivo.

3. En la hipótesis **H.4.B** se consideran procesos α-mixing con coeficientes exponenciales, condición no muy restrictiva y que cumplen, por ejemplo, los procesos ARMA generados a partir de ruido continuo. Se puede debilitar esta condición no exigiendo que los coeficientes $\alpha(j)$ sean de tipo exponencial, para ello habría que utilizar una desigualdad de M. Carbon (1983) en lugar del lema 3, pero se exigirá una condición más restrictiva al parámetro de suavización b_n. En este supuesto la hipótesis **H.4.B**′ sería:

«La sucesión de pares (X_i, Y_i) es α-mixing y

i. existe una sucesión de números naturales k_n, con $1 \leqslant k_n \leqslant n$ tal que $(n/k_n)\alpha(k)^{2k/5n} < A < \infty$ y $\dfrac{\hat{\alpha}(k)}{k_n b_n^d} < \infty$ con $\hat{\alpha}(k) = \sum_{i=1}^{k(n)} \alpha(i)$.

ii. $(nb_n^d)/(k_n \log(n)) \to \infty$.

Que para el caso de coeficientes exponenciales $\alpha(k) = A\rho^j$, $A > 0$, $0 < \rho < 1$, eligiendo $k_n = Cn^{(1+\beta)/2}$, con $C \in (1, 2)$ tal que k_n sea entero para algún β, la condición **ii** es $b_n^{d(1-\beta)/2}/\log(n) \to \infty$, más restrictiva que la dada anteriormente.

4. Utilizando un resultado de Glick (1974) se deducen de los teoremas 1 y 2 propiedades de consistencia global del estimador $r_n(x)$:

Corolario. «Sea $|Y| < \gamma < \infty$. Entonces bajo las hipótesis de los teoremas anteriores (1 ó 2) se otiene que $\int |r_n(x) - r(x)|\mu(dx) \to 0$ cuando $n \to \infty$ en probabilidad o casi segura, respectivamente.»

3. CONCLUSIONES

Se ha definido un estimador no paramétrico recursivo de la función de regresión $r_n(x)$, bastante general ya que comprende a diferentes estimadores no paramétricos clásicos. Y se han estudiado propiedades de consistencia débil y fuerte en un contexto de datos dependientes. Se ha utilizado la hipótesis de dependencia α-mixing que verifican los procesos ARMA, lo que permite adaptar el estimador definido para estimar la función de autorregresión de una serie de tiempo univariante, la cual es utilizada normalmente como función de predicción.

En el estudio de las propiedades de consistencia que hemos realizado uno de los puntos importantes es observar bajo qué condiciones se verifica que:

$$\lim_{n \to \infty} \frac{\sum_{i=1}^{n} b_i^{d\tau} E(g(X)\delta_{m(i)}«x, X_i))}{\sum_{i=1}^{n} b_i^{d\tau} E(\delta_{m(i)}(x, X_i))}$$

Para ello se ha utilizado la hipótesis **D.2.** Siguiendo la metodología de

Greblicki y Pawlak (1987) puede demostrarse que dicha condición se cumple en el supuesto del estimador núcleo y para $\tau = 0$ si:

1. $$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} h_i^d I_{h_i > \varepsilon}}{\sum_{i=1}^{n} h_i^d} = 0 \quad \text{para todo } \varepsilon > 0$$

2. $C_1 H(x) \leqslant K(x) \leqslant C_2 H(x)$ con $C_1 > 0$ y $H(x)$ una función medible, no negativa, no decreciente definida en $\mathbb{R}^+$, con $0 < H(0+) < \infty$ - $\lim t^d H(t)\, 0$. Esta condición es muy débil ya que no exige ninguna condición a la medida de probabilidad asociada a la v.a. X y permite trabajar con todo tipo de funciones núcleo, incluso las de soporte no acotado como el gaussiano o $\exp(-|x|)$, y también con funciones núcleo no integrables como $K(x) = 1/e$ si $|x| \leqslant e$ y $K(x) = 1/(|x| \ln(x))$ si $|x| \geqslant e$.

De los teoremas 1 y 2 se deduce que el estimador no paramétrico $r_n(x)$ presenta un buen comportamiento asintótico en la estimación de la mayoría de los modelos de regresión, incluso bajo condiciones de dependencia, ya que no se le exige ninguna hipótesis a la función de regresión $r(x)$. En el estudio de la consistencia en media cuadrática (Vilar Fernández, 1989b) será necesario utilizar condiciones de regularidad de $r(x)$ (que sea k veces diferenciable) que son usuales en la estimación no paramétrica de curvas, pudiendo aumentar la velocidad de convergencia en media cuadrática al aumentar el valor de k.

Comparando el estimador $r_n(x)$ con el estimador no paramétrico no recursivo $\hat{r}_n(x)$, cuya forma general es la siguiente:

$$\hat{r}_n(x) = \frac{\sum_{i=1}^{n} \delta_{m(n)}(x, X_i) Y_i}{\sum_{i=1}^{n} \delta_{m(n)}(x, X_i)}$$

se puede afirmar que, bajo condiciones generales, ambos estimadores verifican las mismas propiedades de consistencia y con el mismo orden de convergencia, pero el error de estimación al utilizar $r_n(x)$ es mayor que el de $\hat{r}_n(x)$. Por contra, el primero tiene la propiedad de ser recursivo.

En ambos estimadores la velocidad de convergencia va a depender de las siguientes causas: Tipo de estimador (forma de la función núcleo, parámetro de suavización); regularidad de la función de regresión; y dimensión de la variable regresora (d). Siendo muy importante esta última ya que la velocidad de convergencia en media cuadrática de los dos estimadores es $n^{-1/4+d}$, lo que exige que al aumentar la dimensión de la variable regresora ha de aumentar el tamaño muestral para obtener resultados aceptables.

En la realización del estudio de simulación que se expone en el apartado siguiente se ha observado que el tiempo de cómputo al utilizar r_n o $\hat{r}_n$ es similar. Ambos estimadores se obtienen rápidamente una vez que se ha decidido el parámetro de suavización a utilizar, siendo esta etapa la más costosa, ateniéndose al tiempo de computación, si se calcula el parámetro de suavización a partir de los datos muestrales utilizando técnicas de validación cruzada modificadas para el supuesto de dependencia, problema en el que estamos trabajando en la actualidad. Por ello, en el supuesto de un muestreo secuencial es conveniente utilizar el estimador recursivo.

4. SIMULACIONES

Para ilustrar el comportamiento del estimador definido en el apartado 1 se han simulado muestras de tamaño 100 y 200 para el estudio de la función de regresión y de autorregresión de datos con distintos tipos de dependencia.

Se han elegido modelos de regresión suficientemente regulares, lo que permite obtener mejores estimaciones. En las tres primeras figuras se considera un modelo de regresión racional que se estudia bajo distintas condiciones de dependencia, obteniéndose el mejor resultado en la figura 3, donde la dependencia es más débil. En la figura 4 se estudia un modelo sinusoidal y en las dos últimas se estudian dos modelos de autorregresión.

Las curvas se estudian en 240 puntos distribuidos equidistantemente en un intervalo de $(-2\sigma, 2\sigma)$, siendo σ la desviación típica de X. Las series de tiempo utilizadas se han generado a partir de ruido gaussiano. Y las estimaciones se realizan utilizando el método descrito en (2) y (3).

Siendo $\delta_{m(n)}(x, u) = b_n^{-d}K((x-u)/b_n)$ y $K(u)$ la función núcleo Epanechnikov y el parámetro de suavización es $b_n = Cn^{-1/5}$, C un valor elegido según la muestra, de forma empírica, tal que minimice el Error Medio Cuadrático:

$$\text{E.M.C.}(r) = \frac{1}{N_p} \sum_{i=1}^{N_p} (r(x_i) - r_n(x_i))^2$$

Los modelos estudiados son los siguientes:

Figura 1. Modelo de regresión $Y = \dfrac{5X}{1 + X^2} + \varepsilon$, siendo X la serie de tiempo ARMA(2, 2) siguiente:

$$(1 - 0{,}5B + 0{,}3B^2)X_t = (1 + 0{,}2B + 0{,}2B^2)a_t$$

ε *y* a_t son $N(0, 1)$.

Eligiendo el tamaño muestral $n = 200$, $C = 1$, $\tau = 0{,}5$ se ha obtenido un E.M.C.$(r) = 0{,}04371655$.

Figura 2. Modelo de regresión $Y = \dfrac{4X}{2 + X^2} + \varepsilon$, siendo X la serie de tiempo AR(3) siguiente:

$$(1 - 0{,}7B + 0{,}25B^2 - 0{,}175B^3)X_t = a_t$$

ε y a_t son $N(0, 1)$.

Eligiendo el tamaño muestral $n = 200$, $C = 2$ y $\tau = 0$ se ha obtenido un E.M.C.$(r) = 0{,}04238389$.

Figura 3. Modelo de regresión $Y = \dfrac{7X}{3 + X^2} + \varepsilon$, siendo X la serie de tiempo MA(3) siguiente:

$$X_t = (1 + 0{,}5B - 0{,}62B^2 - 0{,}33B^3)a_t$$

ε y a_t son $N(0, 1)$.

Eligiendo al tamaño muestral $n = 200$, $C = 2$ y $\tau = 1$ se ha obtenido un E.M.C.$(r) = 0{,}01104943$.

Figura 4. Modelo de regresión $Y = \operatorname{sen}(3X + 1) + \varepsilon$, siendo X la serie de tiempo ARMA(2, 1) siguiente:

$$(1 - 0{,}3B - 0{,}1B^2)X_t = (1 + 0{,}7B)a_t$$

ε es una $N(0m\,0{,}5)$ y a_t es $N(0, 1)$.

Eligiendo el tamaño muestral $n = 100$, $C = 1$ y $\tau = 0$ se ha obtenido un E.M.C.$(r) = 0{,}06496141$.

Figura 5. Modelo de autorregresión $X_t = 0{,}7X_{t-1} + \varepsilon_t$ siendo ε_t ruido normal: $N(0, 0{,}5)$.

Eligiendo el tamaño muestral $n = 100$, $C = 1{,}4$ y $\tau = 0$ se ha obtenido un E.M.C.$(r) = 0{,}04641599$.

Figura 6. Modelo de autorregresión $X_t = \operatorname{sen}(2X_{t-1} + 2) + \varepsilon_t$ siendo ε_t ruido normal: $N(0, 0{,}5)$.

Eligiendo el tamaño muestral $n = 100$, $C = 0{,}9$ y $\tau = 0$ se ha obtenido un E.M.C.$(r) = 0{,}01679199$.

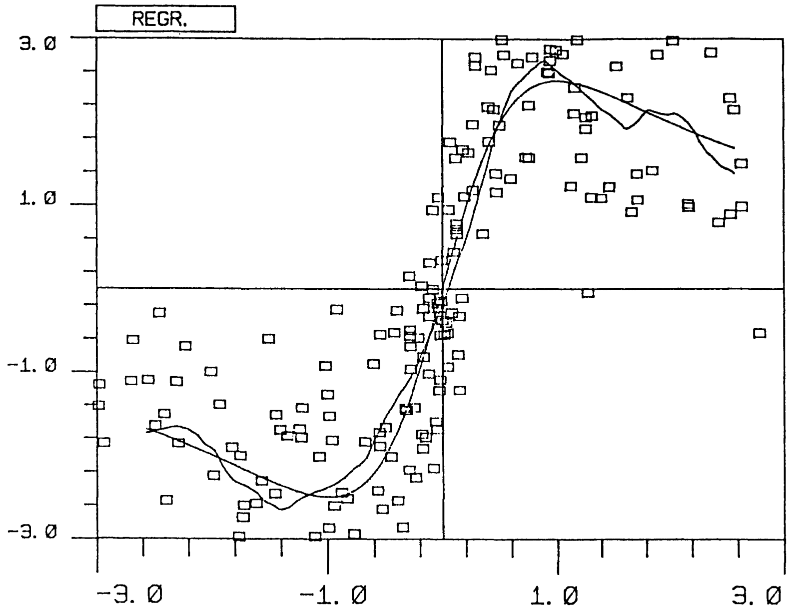

FIG. 1

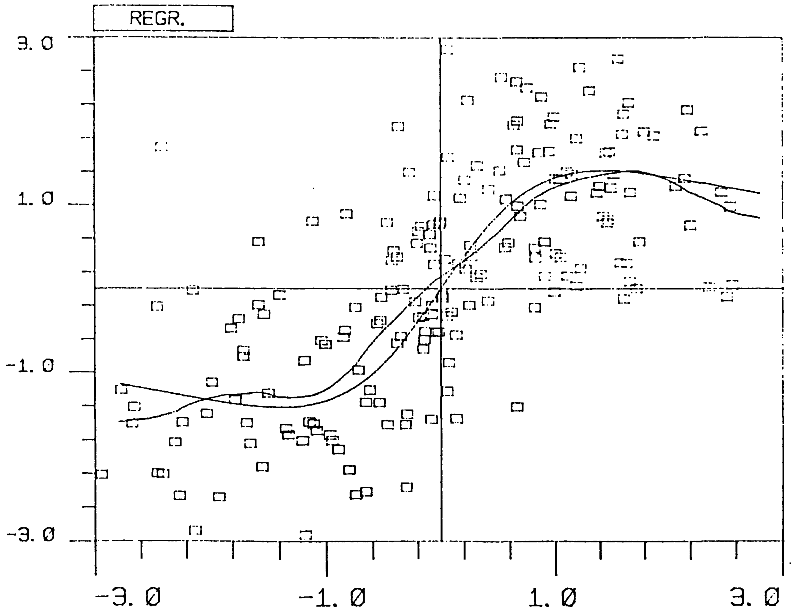

FIG. 2

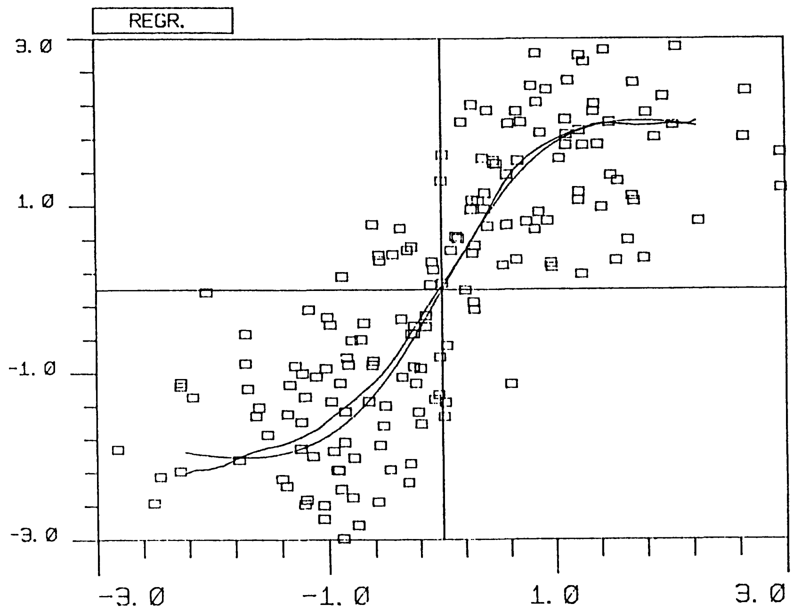

FIG. 3

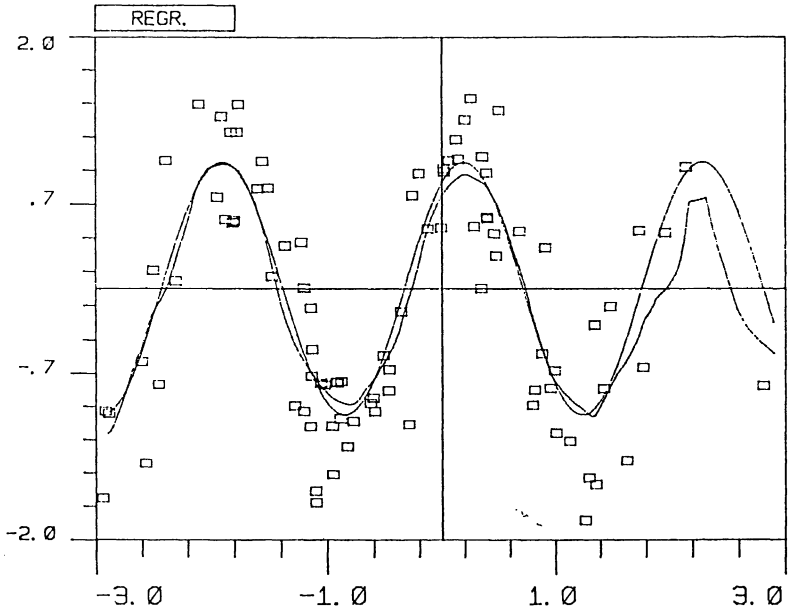

FIG. 4

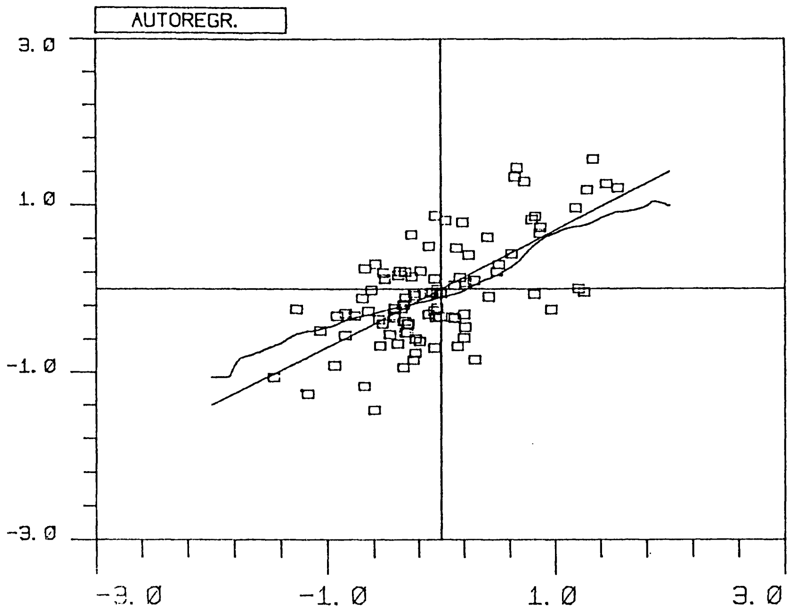

FIG. 5

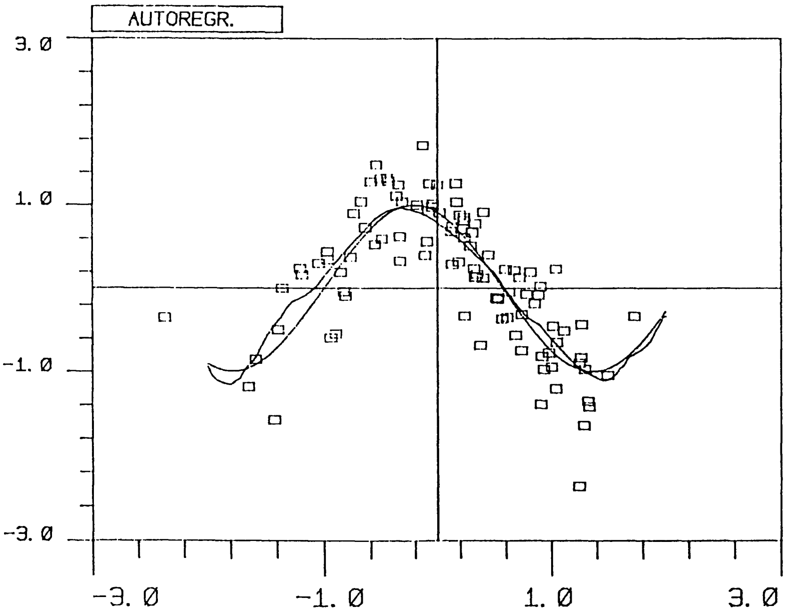

FIG. 6

BIBLIOGRAFIA

BIERENS, H. J. (1983): «Uniform consistency of kernel estimators of a regression function under generalized conditions», *J.A.S.A.*, vol. 78, pp. 699-707.

BRADLEY, R. C. (1986): *Basic properties of strong mixing conditions,* Ed. Birkhauser.

CARBON, M. (1983): «Bernstein's inequality for strong mixing non stationary processes. Applications», *C. R. Acad. Sci. Paris Ser. I Math.*, 297, pp. 303-306.

COLLOMB, G., y HARDLE, W. (1984): «Strong uniform convergence rates in robust nonparametric time series analysis and prediction: kernel regression estimation from dependent observations», *Stochastic Process. Appl.*, pp. 77-89.

COLLOMB, G. (1985): «Nonparametric regression: an up-to-date Bibliography», *Statistics,* 16, pp. 309-324.

COLLOMB, G. (1985): «Non Parametric Time Series Analysis and Prediction: Uniform almost sure convergence of the window and k-NN Autorregression Estimates», *Statistics,* 16, 2, pp. 297-307.

DOUKHAN, P.; LEON, J., y PORTAL, F. (1984): «Vitesse de convergencc dans le théoreme central limite pour des variables aléatoires mélangeantes a valeurs dans un espace de Hilbert», *C. R. Acad. Sci. Paris Ser. I Math.*, 298, pp. 305-308.

GLICK, N. (1974): «Consistency conditios for probability estimators and integrals for density estimators», *Utilitas Math.*, 5, pp. 61-74.

GREBLICKI, W.; KRZYZAK, A., y PAWLAK, M. (1984): «Distribution-free pointwise consistency of kernel regression estimate», *The Annals of Statistics*, vol. 12 N 4, pp. 1570-1575.

GREBLICKI, W., y PAWLAK, M. (1987): «Necesary and sufficient consistency conditions for a recursive kernel regression estimate», *Journal of Multivariate Analysis*, 23, 67-76.

HALL-HEYDE (1980): *Martingale limit theory and its applications*, Nueva York, Academic Press.

HARDLE, W. (1988): *Applied nonparametric regression*, Springer Verlag.

IBRAGIMOV, I. A., y LINNIK Y. V. (1971): *Independent and stationary sequences of random variables*, Wolters-Noordhoff, Groningen.

ROSENBLATT, M. (1971): *Markov Processes, structure and asymptotic bahavior*, Springer Verlag.

VILAR FERNANDEZ, J. M. (1989): «Estimación no paramétrica de curvas notables para datos dependientes», *Trabajos de Estadística*, vol. 4, n.º 2, pp. 69-89.

VILAR FERNANDEZ, J. M. (1989): «Estimación recursiva, tipo núcleo, de la función de autorregresión para datos dependientes», *Estadística Española*, vol. 31, n.º 131.