Divulgaciones Matemáticas Vol. 16 No. 1(2008), pp. 155–171

Estimación semiparamétrica en procesos autorregresivos con régimen de Markov

Semiparametric estimation in autoregressive processes with Markov regime

Ricardo Ríos (rrios@euler.ciens.ucv.ve)
Departamento de Matemticas, Escuela de Matemáticas
Universidad Central de Venezuela, Caracas-Venezuela.
Luis Angel Rodríguez (larodri@uc.edu.ve)
Departamento de Matemáticas, Facultad de Ciencias y Tecnología,
Universidad de Carabobo, Valencia-Venezuela, y
Postgrado en Matemática, Facultad de Ciencias,
Universidad Central de Venezuela, Caracas-Venezuela.

Resumen

En este trabajo se considera la estimación semiparamétrica de los parámetros en procesos autorregresivos controlados por un régimen de Markov. Se estudia un estimador de mínimos cuadrados condicional modificado, demostrando la consistencia en probabilidad. Se calcula la velocidad de convergencia del estimador.
Palabras y frases clave: Procesos autorregresivos, cadenas de Markov ocultas, regresión de estructura variable, estimación por núcleos, mínimos cuadrados condicional.

Abstract

This work considers the semiparametric estimation of the parameters in autoregressive processes controlled by a Markov regime. A modified conditional least squares estimator is studied, proving its consistency in probability and calculating its convergence rate.
Key words and phrases: Autoregressive processes, hidden Markov chain, switching models, kernel estimation, conditional least squares.

Recibido 2006/03/12. Revisado 2006/07/03. Aceptado 2006/07/10.
MSC (2000): Primary 62M09; Secondary 62F12.

1 Introducción

En este trabajo se establece, para procesos autorregresivos con régimen de Markov, la consistencia y la velocidad de convergencia en probabilidad de un estimador de mínimos cuadrados modificado de los parámetros del proceso. Una ventaja práctica de realizar la inferencia estadística por un criterio de mínimos cuadrados frente a estimación por máxima verosimilitud es que no se requiere especificar una distribución para el proceso de ruido, contemplando la estimación no paramétrica de la densidad del ruido usando núcleos de convolución con los resíduos de la estimación a cada paso.

Un proceso autorregresivo con régimen de Markov (AR-RM) es un proceso a tiempo discreto definido por:

$$Y_n = f(Y_{n-1}, \theta_{X_n}) + \varepsilon_n \tag{1}$$

donde $\{X_n,\, n \in \mathbb{N}\}$ es una cadena de Markov no observada, homogénea con valores en el conjunto finito $\{1, \ldots, m\}$, matriz de transición $A = [a_{ij}]$ siendo $a_{ij} = \mathbb{P}(X_n = j | X_{n-1} = i)$. La familia de funciones $\mathcal{F} = \{f(\cdot, \theta) : \theta \in \Theta\}$ está parametrizada por el parámetro $\theta \in \Theta$ y Θ un subconjunto compacto de $\mathbb{R}^d$. Las variables aleatorias $\{\varepsilon_n\}$ se suponen centradas, independientes e idénticamente distribuidas con función de densidad Φ. Suponemos que $\{\varepsilon_n\}$, $\{X_n\}$ y Y_0 son independientes. El proceso $\{X_n\}$ no es observado y por lo tanto la inferencia se centra en el proceso observado $\{Y_n\}$.

El uso de un régimen de Markov oculto ofrece la posibilidad de modelar series temporales que cambian su comportamiento en el tiempo de manera marcada. Hamilton [6] utiliza un proceso AR-RM en el contexto económetrico, para el análisis anual de la serie del producto interno bruto de los Estados Unidos, con dos regímenes: contracción y expansión. Procesos autorregresivos lineales con régimen de Markov son usados en varias áreas de la ingeniería eléctrica, incluidas detección de fallas y control estocástico adaptativo, ver Douc *et al.* [4] y sus referencias.

Entre los trabajos más recientes en los que se desarrolla la estimación de los parámetros por el método de máxima verosimilitud para los procesos autorregresivos con régimen de Markov tenemos: Francq y Roussignol [5], Jensen y Petersen [7] y Douc *et al.* Sobre el problema del cálculo numérico del estimador de máxima verosimilitud consultar Ríos y Rodríguez [10] y sus referencias.

En Mevel [8] se considera, para un modelo de cadenas de Markov ocultas, el siguiente contraste

$$\tilde{S}_N(\psi) = \frac{1}{N}\sum_{n=1}^{N}(Y_n - \mathbb{E}_\psi(Y_n|Y_0^{n-1}))^2. \tag{2}$$

El estimador por mínimos cuadrados condicional (MCC) se define como

$$\tilde{\psi} = \underset{\psi\in\Psi}{\operatorname{argmin}}\tilde{S}_N(\psi). \tag{3}$$

Mevel demuestra la consistencia débil y la normalidad asintótica del estimador MCC. Para la estimación por MCC es necesario conocer la esperanza condicional $\mathbb{E}_\psi(Y_n|Y_0^{n-1})$, en nuestro caso la esperanza condicional del contraste depende de $\{Y_n\}$, ψ, $\{X_n\}$ y de la función de densidad Φ y como ésta es desconocida, $\mathbb{E}(Y_n|Y_0^{n-1})$ también lo es, por lo que el estimador $\tilde{\psi}$ no puede ser obtenido por minimización de $\tilde{S}_N(\psi)$. Nosotros reemplazaremos en la ecuación (2) la esperanza condicional por un estimador no paramétrico basado en la muestra $y_0, \ldots, y_N$ y estimaremos ψ_* minimizando este nuevo contraste.

El criterio de Mínimos Cuadrados Condicional Modificado (MCCM) se define entonces por

$$S_N(\psi) = \frac{1}{N}\sum_{n=1}^{N}(Y_n - \hat{\mathbb{E}}(Y_n|Y_0^{n-1}))^2 \tag{4}$$

donde $\hat{\mathbb{E}}(Y_n|Y_0^{n-1})$ es un estimador no paramétrico de $\mathbb{E}(Y_n|Y_0^{n-1})$ basado en $y_0, \ldots, y_N$. El estimador $\hat{\psi}$ de ψ_* es

$$\hat{\psi} = \underset{\psi\in\Psi}{\operatorname{argmin}}S_N(\psi). \tag{5}$$

Este estimador es considerado en [12, 11] para el modelo no lineal de errores estructurales en las variables. Las técnicas allí utilizadas son adaptadas en nuestro trabajo para demostrar la consistencia débil y obtener las velocidades de convergencia en probabilidad de los estimadores.

El artículo está organizado de la manera siguiente. El modelo y las hipótesis generales son presentados en la sección 2. En la sección 3 se presentan los resultados principales.

2 Hipótesis generales

2.1 Hipótesis sobre el modelo

Las condiciones (E1-E3) que definimos a continuación garantizan la existencia de una medida invariante para la cadena de Markov vectorial $\{(Y_n, X_n)\}$.

E1 La matriz de transición A es positiva, esto es, $a_{ij} \geq \delta$, para todo $i, j \in \{1, \ldots, m\}$ y para algún $\delta > 0$. Esta condición implica que la cadena es irreducible y aperiódica por lo tanto existe una única medida invariante $\mu = (\mu_1, \ldots, \mu_m)$ para la cadena oculta $\{X_n\}$.

E2 (Sublinealidad) Para $i = 1, ..., m$, existen constantes positivas λ_i y b_i tales que

$$|f(y, \theta_i)| \leq \lambda_i |y| + b_i.$$

E3 (Condición tipo radio espectral) Suponemos que $\sum_{i=1}^{m} \log \lambda_i \mu_i < 0$.

E3 Existe $p > 0$ tal que $\mathbb{E}(|\varepsilon_1|^p) < \infty$.

Las condiciones anteriores garantizan que la cadena de Markov extendida es geométricamente ergódica en el espacio de estados $\mathbb{R} \times \{1, \ldots, m\}$ (ver Yao y Attali [13]).

El parámetro $\psi = (\mu, A, \theta)$ pertenece al espacio de parámetros definido por $\Psi = [0,1]^m \times [0,1]^{m^2} \times \Theta^m$.

Para el estudio de teoremas límites, supondremos que los elementos de la familia de funciones $\mathcal{F}$ satisfacen las siguientes condiciones de Lipschitz (LP) y de acotación (AC)

LP Para $i = 1, ..., m$,

$$|f(y, \theta_i) - f(y, \theta'_i)| \leq K_1 |\theta_i - \theta'_i|.$$

AC Para $i = 1, ..., m$, existen constantes C_1, C_2 tales que

$$|f(y, \theta_i)| \leq C_1 \text{ y } \|\nabla_\psi f(y_n, \theta_i)\| \leq C_2,$$

donde ∇_ψ es el operador gradiente.

2.2 Hipótesis sobre la densidad y los estimadores

Suponemos que la función de densidad Φ es acotada y que existe $\rho > 0$ tal que,

D1 Φ tiene ρ derivadas continuas en un conjunto compacto $\mathcal{C}$.

D2 Existe $\alpha > 0$ tal que $\inf_{\varepsilon \in \mathcal{C}} \Phi(\varepsilon) > \alpha$.

Para la estimación no paramétrica utilizaremos núcleos de convolución $K : \mathbb{R} \to \mathbb{R}$ acotados y con soporte compacto. Además supondremos que

K1 $\int K(t)dt = 1$

K2 $0 < \left|\int t^{\rho} K(t)dt\right| < \infty$ y $\int t^{s} K(t)dt = 0$, $\forall s = 1, \dots, \rho - 1$.

K3 Existen constantes $\beta > 0$, $Ctte < \infty$, tales que

$$\forall \varepsilon, \varepsilon' \in \mathcal{C},\ |K(\varepsilon) - K(\varepsilon')| < Ctte|\varepsilon - \varepsilon'|^{\beta}.$$

La literatura sobre núcleos de convolución es extensa, dos ejemplos muy utilizados son el gaussiano $K_g(t) = 1/\sqrt{2\pi}\exp(-t^2/2)$ y el de Epanechnikov $K_e(t) = (3/(4\sqrt{5\sigma^2}))[1 - t^2/5\sigma^2]I_{\{t^2 \leq 5\sigma^2\}}$.

Sobre el parámetro de suavizado $h = h(N)$ impondremos la condición

PS $\lim_{N\to\infty} h(N) = 0$ y $\lim_{N\to\infty} Nh(N)/\log N = \infty$.

Suponemos que,

C1 La función $\psi \to \mathbb{E}_{\psi^*}(\mathbb{E}_{\psi}(Y_1|Y_0) - \mathbb{E}_{\psi^*}(Y_1|Y_0))^2$ admite un único mínimo en $\psi = \psi^*$,

C2 Se aceptan como ciertas las siguientes condiciones de momento

$$\mathbb{E}|Y_1 - \mathbb{E}(Y_1|Y_0)|^{4+2\delta} < \infty$$
$$\mathbb{E}\|\nabla_{\psi}\mathbb{E}(Y_1|Y_0)\|^{4+2\delta} < \infty.$$

3 Mínimos cuadrados condicional modificado (MCCM)

En esta sección se demuestra la consistencia del estimador MCCM. Se comienza construyendo el estimador de la esperanza condicional $\hat{\mathbb{E}}(Y_n|Y_0^{n-1})$, se establece la consistencia en probabilidad y la normalidad asintótica del estimador preliminar $\tilde{\psi}$. Por último se demuestra la consistencia en probabilidad del estimador semiparamétrico $\hat{\psi}$ y se establece su velocidad de convergencia.

3.1 Construcción del estimador $\hat{\mathbb{E}}(Y_n|Y_0^{n-1})$

En esta sección se construye un estimador de la esperanza condicional

$$\mathbb{E}(Y_n|Y_0^{n-1})$$

. Esta esperanza condicional se calcula mediante la fórmula

$$\mathbb{E}(Y_n|Y_0^{n-1}) = \sum_{i=1}^{m} f(y_{n-1},\theta_i)\alpha_n(i), \tag{6}$$

y $\alpha_n(i) = p(X_n = i|Y_0^{n-1})$ es evaluada recursivamente por la fórmula forward de Baum (ver [8]),

$$\alpha_n(j) = \sum_{i=1}^{N} \alpha_{n-1}(i)a_{ij}\Phi(y_n - f(y_{n-1},\theta_i)). \tag{7}$$

Para la función de densidad Φ, elegimos el estimador de núcleo:

$$\hat{\Phi}(\varepsilon) = \frac{1}{Nh}\sum_{n=1}^{N} K\left(\frac{\varepsilon-\varepsilon_n}{h}\right)$$

donde K satisface [K1-K3]. Como se observa en la ecuación (6) la esperanza condicional $\mathbb{E}(Y_n|Y_0^{n-1})$ sólo depende de Φ a través de las probabilidades $\{\alpha_k(i)\}$. Estimaremos $\mathbb{E}(Y_n|Y_0^{n-1})$ por

$$\hat{\mathbb{E}}(Y_n|Y_0^{n-1}) = \sum_{i=1}^{m} f(y_{n-1},\theta_i)\hat{\alpha}_n(i)$$

y

$$\hat{\alpha}_n(j) = \sum_{i=1}^{N} \hat{\alpha}_{n-1}(i)a_{ij}\hat{\Phi}(y_n - f(y_{n-1},\theta_i)).$$

En la sección siguiente demostramos algunas propiedades asintóticas de este estimador.

3.2 Propiedades asintóticas de $\hat{\alpha_n}$

El siguiente teorema nos permite establecer el comportamiento asintótico del estimador del filtro α_n.

Proposición 1. *Para el proceso AR-RM definido en (1) bajo las hipótesis de estabilidad [E1-E3], las condiciones sobre el núcleo K [K1-K3] y la condición [PS], se verifica que*

$$\|\hat{\alpha}_n(j) - \alpha_n(j)\|_\infty = O_p\left(v_n\right),$$

con $v_n = (\log n/n)^v$, $v = \rho/(2\rho+1)$. *La norma infitita* $\|\cdot\|_\infty$ *se entiende como el supremo de una función, calculado sobre un compacto.*

Demostración: Como la sucesión $\{\varepsilon_n\}$ es i.i.d es conocido que:

$$\|\hat{\Phi} - \Phi\|_\infty = O_p(v_n),$$

por ejemplo ver Ango-Nze y Ríos [1].

En lo que sigue demostramos que la norma $\|\hat{\alpha}_n(j) - \alpha_n(j)\|_\infty$ esta acotada por $C\|\hat{\Phi} - \Phi\|_\infty$, para esto procedemos por inducción en n. Para $n = 2$,

$$\begin{aligned} |\hat{\alpha}_2(j) - \alpha_2(j)| &= |\sum_{i=1}^{m} a_{ij}(\hat{\Phi}(\varepsilon_{i,1}) - \Phi(\varepsilon_{i,1}))\mu_i| \\ &\leq m\|\hat{\Phi} - \Phi\|_\infty. \end{aligned}$$

Suponiendo cierto el resultado para $n-1$ tenemos,

$$\begin{aligned} |\hat{\alpha}_n(j) - \alpha_n(j)| &= |\sum_{i=1}^{m} a_{ij}\left(\hat{\Phi}(\varepsilon_{i,n})\hat{\alpha}_{n-1}(i) - \Phi(\varepsilon_{i,n})\alpha_{n-1}(i)\right)| \\ &= |\sum_{i=1}^{m} a_{ij}\Big((\hat{\alpha}_{n-1}(i) - \alpha_{n-1}(i))\hat{\Phi}(\varepsilon_{i,n}) \\ &\quad + (\hat{\Phi}(\varepsilon_{i,n}) - \Phi(\varepsilon_{i,n}))\alpha_{n-1}(i)\Big)| \\ &\leq m(M_2\|\hat{\alpha}_{n-1} - \alpha_{n-1}\| + \|\hat{\Phi} - \Phi\|_\infty). \end{aligned}$$

■

En la próxima sección demostramos la consistencia y la velocidad de convergencia del estimador $\tilde{\psi}$.

3.3 Propiedades asintóticas del estimador preliminar $\tilde{\psi}$

El estimador preliminar $\tilde{\psi}$ es consistente y asintóticamente normal como se demuestra en la siguiente proposición.

Proposición 2. *Para el proceso AR-RM definido en (1) bajo las hipótesis de estabilidad [E1-E3], las condiciones [C1-C2], tenemos que $\tilde{\psi}$ definido en (3) satisface,*

- **i** *El estimado $\tilde{\psi}$ converge en probabilidad a ψ_*.*
- **ii** *Se tiene que $\sqrt{n}(\tilde{\psi} - \psi_*) \to \mathcal{N}(0, \Sigma)$ donde Σ es una matriz definida positiva.*

En el siguiente lema se verifican que las hipótesis necesarias para la convergencia débil de un estimador de mínimo contraste (ver [3], §3, pág. 93) se satisfacen en este caso.

Lema 1. *Para el proceso AR-RM definido en (1) bajo las hipótesis de estabilidad [E1-E3], las condiciones [C1-C2], tenemos que:*

i. *Para todo $\psi \in \Psi$, existe una función $\tilde{S}(\psi, \psi*)$ tal que $\tilde{S}_N(\psi) \to \tilde{S}(\psi, \psi*)$.*

ii. *La función $\tilde{S}(\psi, \psi*)$ admite un único mínimo en $\psi = \psi*$.*

iii. *Existen sucesiones $\{\rho_k\}$ y $\{\eta_k\}$ tales que*

$$\lim_{N\to\infty} \mathbb{P}\left(\sup\left\{|\tilde{S}_N(\psi) - \tilde{S}_N(\psi')|,\ |\psi - \psi'| \leq \rho_k\right\} \geq \eta_k\right) = 0.$$

Demostración de (i).

Aplicando el teorema ergódico a la cadena de Markov estacionaria $\{(Y_n, X_n, \alpha_n)\}$,

$$\frac{1}{N}\sum_{n=1}^{N}(Y_n - \mathbb{E}_\psi(Y_n|Y_0^{n-1}))^2 \to \tilde{S}(\psi, \psi*) := \mathbb{E}_{\psi^*}(Y_1 - \mathbb{E}_\psi(Y_1|Y_0))^2.$$

Demostración de (ii).

Tenemos que

$$\mathbb{E}(Y_1 - \mathbb{E}_\psi(Y_1|Y_0))^2 = \mathbb{E}(Y_1 - \mathbb{E}_{\psi*}(Y_1|Y_0))^2 + \mathbb{E}(\mathbb{E}_{\psi*}(Y_1|Y_0) - \mathbb{E}_\psi(Y_1|Y_0))^2$$

y por hipótesis (C1) la función $\mathbb{E}(\mathbb{E}_{\psi*}(Y_1|Y_0) - \mathbb{E}_\psi(Y_1|Y_0))^2$ admite un único mínimo en $\psi = \psi^*$ y así $\tilde{S}(\psi, \psi*)$.

Demostración de (iii).

$$\begin{aligned}
&\tilde{S}_N(\psi) - \tilde{S}_N(\psi') \\
&= \frac{1}{N}\left[\sum_{n=1}^{N}(Y_n - \mathbb{E}_\psi(Y_n|Y_0^{n-1}))^2 - (Y_n - \mathbb{E}_{\psi'}(Y_n|Y_0^{n-1}))^2\right] \\
&= \frac{1}{N}\sum_{n=1}^{N}[\mathbb{E}_{\psi'}(Y_n|Y_0^{n-1}) - \mathbb{E}_\psi(Y_n|Y_0^{n-1})] \\
&\qquad\qquad \cdot[2Y_n - (\mathbb{E}_{\psi'}(Y_n|Y_0^{n-1}) + \mathbb{E}_\psi(Y_n|Y_0^{n-1}))]
\end{aligned}$$

Como la esperanza de $2Y_n - (\mathbb{E}_{\psi'}(Y_n|Y_0^{n-1}) + \mathbb{E}_{\psi'}(Y_n|Y_0^{n-1}))$ es acotada nos basta demostrar que el término $\mathbb{E}_{\psi'}(Y_n|Y_0^{n-1}) - \mathbb{E}_{\psi}(Y_n|Y_0^{n-1})$ esta acotado por $Ctte\|\psi - \psi'\|$ y así el punto (iii) es cierto. En efecto,

$$\begin{aligned}
&|\mathbb{E}_{\psi'}(Y_n|Y_0^{n-1}) - \mathbb{E}_{\psi}(Y_n|Y_0^{n-1})| \\
&= |\sum_{i=1}^{m}(f(y_{n-1},\theta_i)\alpha_n(i) - f(y_{n-1},\theta'_i)\alpha'_n(i))| \\
&= |\sum_{i=1}^{m}(f(y_{n-1},\theta_i) - f(y_{n-1},\theta'_i))\alpha_n(i) + f(y_{n-1},\theta'_i)(\alpha_n(i) - \alpha'_n(i))| \\
&\leq |\sum_{i=1}^{m} K_1(\theta_i - \theta'_i) + Ctte(\alpha_n(i) - \alpha'_n(i))| \\
&= m\,(K_1\|\theta - \theta'\| + Ctte\|\alpha_n - \alpha'_n\|)
\end{aligned}$$

para controlar el término $\|\alpha_n - \alpha'_n\|$ procedemos por inducción en n. Para $n = 2$,

$$\begin{aligned}
|\alpha_2(j) - \alpha'_2(j)| &= |\sum_{i=1}^{m} a_{ij}\Phi(y_1 - f(y_0,\theta_i))\mu_i - a'_{ij}\Phi(y_1 - f(y_0,\theta'_i))\mu'_i| \\
&\leq |\sum_{i=1}^{m}(\Phi(y_1 - f(y_0,\theta_i)) - \Phi(y_1 - f(y_0,\theta_i)))\mu_i \\
&\qquad\qquad + \Phi(y_1 - f(y_0,\theta'_i))(\mu_i - \mu'_i)| \\
&\leq \sum_{i=1}^{m} K_2|f(y_0,\theta'_i) - f(y_0,\theta_i)| + M_2|\mu_i - \mu'_i| \\
&\leq \sum_{i=1}^{m} K_2K_1|\theta'_i - \theta_i| + M_2|\mu_i - \mu'_i| \\
&\leq m(K_2K_1\|\theta - \theta'\| + M_2\|\mu - \mu'\|).
\end{aligned}$$

Si suponemos cierto para $n-1$ entonces para n tenemos,

$$\begin{aligned}
|\alpha_n(j) - \alpha'_n(j)| &= |\sum_{i=1}^{m} a_{ij}\Phi(y_n - f(y_{n-1},\theta_i))\alpha_{n-1}(i) \\
&\quad - a'_{ij}\Phi(y_n - f(y_{n-1},\theta'_i))\alpha'_{n-1}(i)| \\
&\leq \sum_{i=1}^{m} K_2K_1|\theta'_i - \theta_i| + M_2|\alpha_{n-1}(i) - \alpha'_n(i)| \\
&\leq m(K_2K_1\|\theta - \theta'\| + M_2\|\alpha_n - \alpha'_n\|).
\end{aligned}$$

■

El siguiente lema será necesario para establecer la parte (ii) de la proposición 2.

Lema 2. *Bajo las hipótesis de estabilidad [E1-E3] suponiendo las las condiciones [C1-C2] el proceso AR-RM definido en (1) satisface las siguientes propiedades,*

i $\sqrt{N}\nabla_\psi \tilde{S}_N(\psi_*) \to \mathcal{N}(0,\Sigma_1)$.

ii $\nabla^2_\psi \tilde{S}_N(\psi_*) \to \left[2\mathbb{E}\left(\frac{\partial \mathbb{E}_\psi(Y_n|Y_0^{n-1})}{\partial \psi_{k'}} \frac{\partial \mathbb{E}_\psi(Y_n|Y_0^{n-1})}{\partial \psi_k}\right)\right]$

iii $R_N = o_p(1)$ *con*

$$R_N = \int_0^1 \{\nabla^2_\psi \tilde{S}_N(s\tilde{\psi} + (1-s)\psi) - \nabla^2_\psi \tilde{S}_N(\psi_*)\}ds.$$

Demostración:

Para demostrar (i) observemos que

$$\nabla_\psi \tilde{S}_N(\psi) = -\frac{2}{N}\sum_{n=1}^{N}(Y_n - \mathbb{E}_\psi(Y_n|Y_0^{N-1}))\nabla_\psi \mathbb{E}_\psi(Y_n|Y_0^{n-1}).$$

Si definimos las variables $Z_n = (Y_n - \mathbb{E}_\psi(Y_n|Y_0^{N-1}))\nabla_\psi \mathbb{E}_\psi(Y_n|Y_0^{n-1})$ podemos probar que

- $\mathbb{E}(Z_n) = 0$.
- $\mathbb{E}(Z_n|\sigma(Y_0^N)) = 0$.

de donde la sucesión $\{Z_n\}$ es una diferencia de martingalas que satisface,

- $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n Z_k Z_k^t = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n cov(Z_k) = \Sigma_1$.
- $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n \mathbb{E}\|\frac{Z_k}{\sqrt{n}}\|^{2+\delta} = 0$.

estas afirmaciones las obtenemos observando que por Cauchy-Schwarz,

$$\mathbb{E}\|Z_k\|^{2+\delta} \leq \sqrt{\mathbb{E}|Y_k - \mathbb{E}_\psi(Y_k|Y_0^{k-1})|^{4+2\delta}}\sqrt{\mathbb{E}\|\nabla_\psi \mathbb{E}_\psi(Y_k|Y_0^{k-1})\|^{4+2\delta}}$$

y las cantidades de la derecha están acotadas por la hipótesis (C2). Entonces satisfechas las condiciones para el TCL para diferencia de martingalas (ver [2] § 6, teorema 6.16, pág. 116) se tiene que

$$\frac{1}{\sqrt{n}}\sum_{k=1}^n Z_k \to \mathcal{N}(0,\Sigma_1).$$

Para demostrar (ii) tenemos que

$$\frac{\partial \tilde{S}_N(\psi)}{\partial \psi_{k'} \partial \psi_k} = -\frac{2}{N}\sum_{n=1}^{N}(Y_n - \mathbb{E}_\psi(Y_n|Y_0^{N-1}))\frac{\partial \mathbb{E}_\psi(Y_n|Y_0^{n-1})}{\partial \psi_{k'} \partial \psi_k} + \frac{2}{N}\sum_{n=1}^{N}\frac{\partial \mathbb{E}_\psi(Y_n|Y_0^{n-1})}{\partial \psi_{k'}}\frac{\partial \mathbb{E}_\psi(Y_n|Y_0^{n-1})}{\partial \psi_k} \tag{8}$$

y aplicando el teorema ergódico obtenemos

- $-\frac{2}{N}\sum_{n=1}^{N}(Y_n - \mathbb{E}_\psi(Y_n|Y_0^{N-1}))\frac{\partial \mathbb{E}_\psi(Y_n|Y_0^{n-1})}{\partial \psi'_k \partial \psi_k} \to 0$
- $\frac{2}{N}\sum_{n=1}^{N}\frac{\partial \mathbb{E}_\psi(Y_n|Y_0^{n-1})}{\partial \psi_{k'}}\frac{\partial \mathbb{E}_\psi(Y_n|Y_0^{n-1})}{\partial \psi_k} \to 2\mathbb{E}\left(\frac{\partial \mathbb{E}_\psi(Y_n|Y_0^{n-1})}{\partial \psi_{k'}}\frac{\partial \mathbb{E}_\psi(Y_n|Y_0^{n-1})}{\partial \psi_k}\right)$.

Para la demostración de (iii), en virtud de la ecuación (8)

$$\frac{\partial \tilde{S}_N(\psi(s))}{\partial \psi_{k'} \partial \psi_k} - \frac{\partial \tilde{S}_N(\psi_*)}{\partial \psi_{k'} \partial \psi_k} = F_1 + F_2 + F_3,$$

donde $\psi(s) = s\tilde{\psi} + (1-s)\psi$,

$$F_1(s) = \frac{2}{N}\sum_{n=1}^{N}\frac{\partial \mathbb{E}_{\psi(s)}(Y_n|Y_0^{n-1})}{\partial \psi_{k'}}\frac{\partial \mathbb{E}_{\psi(s)}(Y_n|Y_0^{n-1})}{\partial \psi_{k'}} - \frac{\partial \mathbb{E}_{\psi}(Y_n|Y_0^{n-1})}{\partial \psi_{k'}}\frac{\partial \mathbb{E}_{\psi}(Y_n|Y_0^{n-1})}{\partial \psi_{k'}}$$

$$F_2(s) = \frac{2}{N}\sum_{n=1}^{N} Y_n\left(\frac{\partial^2 \mathbb{E}_{\psi(s)}(Y_n|Y_0^{n-1})}{\partial \psi_{k'} \partial \psi_k} - \frac{\partial^2 \mathbb{E}_{\psi}(Y_n|Y_0^{n-1})}{\partial \psi_{k'} \partial \psi_k}\right),$$

$$F_3(s) = \frac{2}{N}\sum_{n=1}^{N}\mathbb{E}_{\psi(s)}(Y_n|Y_0^{n-1})\frac{\partial \mathbb{E}_{\psi(s)}(Y_n|Y_0^{n-1})}{\psi_{k'}} - \mathbb{E}_{\psi}(Y_n|Y_0^{n-1})\frac{\partial \mathbb{E}_{\psi}(Y_n|Y_0^{n-1})}{\psi_{k'}}.$$

Escribimos

$$\frac{\partial \mathbb{E}_{\psi(s)}(Y_n|Y_0^{n-1})}{\partial \psi_{k'}} - \frac{\partial \mathbb{E}_{\psi(s)}(Y_n|Y_0^{n-1})}{\partial \psi_k} = \frac{\partial^2 \mathbb{E}_{\overline{\psi}}(Y_n|Y_0^{n-1})}{\partial \psi_{k'} \partial \psi_k}s(\tilde{\psi} - \psi_*),$$

donde $\overline{\psi}$ es un punto en el segmento que une a $\psi(s)$ con ψ_*, lo que nos permite expresar $F_1(s)$ como

$$s\frac{\partial^2 \mathbb{E}_{\overline{\psi}}(Y_n|Y_0^{n-1})}{\partial \psi_{k'} \partial \psi_k}(\tilde{\psi} - \psi_*)\left(\frac{\partial \mathbb{E}_{\psi(s)}(Y_n|Y_0^{n-1})}{\partial \psi_{k'}} + \frac{\partial \mathbb{E}_{\psi_*}(Y_n|Y_0^{n-1})}{\partial \psi_k}\right).$$

La consistencia de $\tilde{\psi}$, la compacidad de Θ y la condición de regularidad [AC] permiten demostrar que $\int_0^1 F_1(s)ds = o_p(1)$. Análogamente para F_2 y F_3. ■

Demostración proposición 2

La parte (i) de la proposición es una consecuencia directa del lema 2.

Para la demostración de la parte (ii), construimos el desarrollo de Taylor de $\nabla\tilde{S}_N$ alrededor de ψ_*,

$$\nabla_\psi \tilde{S}_N(\tilde{\psi}) = \nabla_\psi \tilde{S}_N(\psi_*) + (\tilde{\psi} - \psi_*)\nabla^2_\psi \tilde{S}_N(s\tilde{\psi} + (1-s)\psi)$$

de la definición de $\tilde{\psi}$ se tiene que $\nabla_\psi \tilde{S}_n(\tilde{\psi}) = 0$, por lo tanto la ecuación anterior es equivalente a la expresión,

$$\sqrt{N}(\tilde{\psi} - \psi_*) = -\sqrt{N}\nabla_\psi \tilde{S}_N(\psi_*)[\nabla^2_\psi \tilde{S}_N(\psi_*) + R_N]^{-1},$$

como una consecuencia del Lema 2 obtenemos el resultado.

3.4 Consistencia y velocidad de convergencia del estimador $\hat{\psi}$

En el teorema siguiente demostramos la convergencia en probabilidad del estimador global $\hat{\psi}$ a ψ^*.

Teorema 1. *Para el proceso AR-RM definido en (1) bajo las hipótesis de estabilidad [E1-E3], las condiciones sobre el núcleo K [K1-K4], las condiciones [C1-C2] y la condición [PS], se verifica que el estimador $\hat{\psi} \to \psi^*$ en probabilidad.*

Demostración: Para demostrar la consistencia de $\hat{\psi}$ comenzaremos demostrando que $S_N(\psi) - \tilde{S}_N(\psi) = o_p(1)$.

$$\begin{aligned}
&S_N(\psi) - \tilde{S}_N(\psi) \\
&\quad = \frac{1}{N}\left[\sum_{n=1}^{N}(Y_n - \hat{\mathbb{E}}_\psi(Y_n|Y_0^{n-1}))^2 - (Y_n - \mathbb{E}_\psi(Y_n|Y_0^{n-1}))^2\right] \\
&\quad = \frac{1}{N}\sum_{n=1}^{N}[\mathbb{E}_\psi(Y_n|Y_0^{n-1}) - \hat{\mathbb{E}}_\psi(Y_n|Y_0^{n-1})] \\
&\qquad\qquad \cdot[2Y_n - (\hat{\mathbb{E}}_\psi(Y_n|Y_0^{n-1}) + \mathbb{E}_\psi(Y_n|Y_0^{n-1}))].
\end{aligned}$$

Como vimos en la proposición 2 $|\mathbb{E}_\psi(Y_n|Y_0^{n-1}) - \hat{\mathbb{E}}_\psi(Y_n|Y_0^{n-1})| = o_p(1)$ y como la esperanza del término $2Y_n - (\hat{\mathbb{E}}_\psi(Y_n|Y_0^{n-1}) + \mathbb{E}_\psi(Y_n|Y_0^{n-1}))$ es acotada entonces se obtiene que $|S_N(\psi) - \tilde{S}_N(\psi)| = o_p(1)$, como $\tilde{S}_n(\psi)$ es un contraste de acuerdo con la proposición 1 entonces $S_n(\psi)$ también es un contraste y argumentando como para $\tilde{\psi}$, deducimos que $\hat{\psi}$ es consistente.

■

Teorema 2. *Para el proceso AR-RM definido en (1) bajo las hipótesis de estabilidad [E1-E3], las condiciones sobre el núcleo K [K1-K4], las condiciones [C1-C2] y la condición [PS], tenemos que*

$$\hat{\psi} - \psi_* = O_p\left(v_N\right)$$

con $v_N = \left(\log N/N\right)^v$ *y* $v = 1/(2\rho+2)(2\rho+1)$.

Para la demostración de este teorema establecemos el siguiente lema.

Lema 3. *Para el proceso AR-RM definido en (1) bajo las hipótesis de estabilidad [E1-E3], las condiciones sobre el núcleo K [K1-K4], las condiciones [C1-C2] y la condición [PS], se satisfacen las propiedades siguientes:*

i. *La velocidad de convergencia de*

$$\|\nabla_\psi \tilde{S}_N(\psi_*) - \nabla_\psi S_N(\psi_*)\| = O_p\left(v_N\right)$$

con $v_N \left(\log N/N\right)^v$ *y* $v = 1/(2\rho+2)(2\rho+1)$.

ii. $\nabla^2_\psi S_N(\psi_*) \to \left[2\mathbb{E}\left(\frac{\partial \hat{\mathbb{E}}_\psi(Y_n|Y_0^{n-1})}{\partial \psi'_k}\frac{\partial \hat{\mathbb{E}}_\psi(Y_n|Y_0^{n-1})}{\partial \psi_k}\right)\right]$.

iii. $R_n = o_p(1)$ *con*

$$R_N = \int_0^1 \nabla^2_\psi S_N(s\hat{\psi} + (1-s)\psi) - \nabla^2_\psi S_N(\psi_*) ds.$$

Demostración (i).

Escribimos

$$\nabla_\psi \tilde{S}_N(\psi_*) - \nabla_\psi S_N(\psi_*) = T_1 + T_2 + T_3$$

donde

$$\begin{aligned}
T_1 &= 2/N \textstyle\sum_{n=1}^N (Y_n - \mathbb{E}(Y_n|Y_0^{n-1}))(\nabla_\psi \hat{\mathbb{E}}(Y_n|Y_0^{n-1}) - \nabla_\psi \mathbb{E}(Y_n|Y_0^{n-1}))\\
T_2 &= 2/N \textstyle\sum_{n=1}^N (\hat{\mathbb{E}}(Y_n|Y_0^{n-1}) - \mathbb{E}(Y_n|Y_0^{n-1}))(\nabla_\psi \hat{\mathbb{E}}(Y_n|Y_0^{n-1})\\
&\qquad\qquad - \nabla_\psi \mathbb{E}(Y_n|Y_0^{n-1}))\\
T_3 &= 2/N \textstyle\sum_{n=1}^N (\hat{\mathbb{E}}(Y_n|Y_0^{n-1}) - \mathbb{E}(Y_n|Y_0^{n-1}))\nabla_\psi \mathbb{E}(Y_n|Y_0^{n-1}).
\end{aligned}$$

De las expresiones que definen los términos T_1, T_2 y T_3 observamos que la velocidad de convergencia está gobernada por el comportamiento de las expresiones,

- $\hat{\mathbb{E}}(Y_n|Y_0^{n-1}) - \mathbb{E}(Y_n|Y_0^{n-1})$.
- $\nabla_\psi \hat{\mathbb{E}}(Y_n|Y_0^{n-1}) - \nabla_\psi \mathbb{E}(Y_n|Y_0^{n-1})$.

En la sección § 3.2 demostramos que

$$\|\hat{\alpha}_n - \alpha_n\|_\infty = O_p\left((\log N/N)^{v_1}\right),$$

$v_1 = \rho/(2\rho+1)$ y como

$$\hat{\mathbb{E}}(Y_n|Y_0^{n-1}) - \mathbb{E}(Y_n|Y_0^{n-1}) = \sum_{i=1}^{m} f(Y_n, \theta_i)(\hat{\alpha}_n(i) - \alpha_n(i))$$

y estamos suponiendo que las funciones de regresión son acotadas entonces

$$\|\hat{\Phi} - \Phi\|_\infty = O_p\left(\left(\frac{\log N}{N}\right)^{v_1}\right).$$

Por otra parte,

$$\begin{aligned}
\nabla_\psi \hat{\mathbb{E}}(Y_n|Y_0^{n-1}) - \nabla_\psi \mathbb{E}(Y_n|Y_0^{n-1}) &= \sum_{i=1}^{m} \nabla_\psi f(y_n, \theta_i)(\hat{\alpha}_n^\psi(i) - \alpha_n^\psi(i)) \\
&+ \sum_{i=1}^{m} f(y_n, \theta_i)(\nabla_\psi \hat{\alpha}_n^\psi(i) - \nabla_\psi \alpha_n^\psi(i))
\end{aligned}$$

y como $\|\nabla_\psi f(y_n, \theta_i)\| \leq C_2$ entonces

$$\|\sum_{i=1}^{m} \nabla_\psi f(y_n, \theta_i)(\hat{\alpha}_n^\psi(i) - \alpha_n^\psi(i))\| \leq mC_2 \|\hat{\alpha}_n - \alpha_n\|.$$

Tenemos que $\nabla_\psi \alpha_n^\psi(j)$ está compuesto de las siguientes derivadas parciales:

$$\frac{\partial \alpha_n^\psi(j)}{\partial \theta_l} =$$

$$\frac{\partial \alpha_{n-1}^\psi(j)}{\partial \theta_l} a_{lj} \phi(y_n - f(y_{n-1}, \theta_l)) + \alpha_{n-1}^\psi(l) a_{lj} \Phi'(y_n - f(y_{n-1}, \theta_l)) \frac{\partial f(y_{n-1}, \theta_l)}{\partial \theta_l}$$

y

$$\frac{\partial \alpha_n^{\psi}(j)}{\partial a_{ls}}$$
$$= \begin{cases} \frac{\partial \alpha_{n-1}^{\psi}(j)}{\partial a_{ls}} a_{ls}\Phi(y_n - f(y_{n-1},\theta_l)) + \alpha_n^{\psi}(l)\Phi(y_n - f(y_{n-1},\theta_l)) & j = s \\ 0 & j \neq s \end{cases}$$

Las fórmulas son análogas para $\hat{\alpha}_n^{\psi}$ sustituyendo Φ por $\hat{\Phi}$ y Φ por $\hat{\Phi}'$. Al sustituir en $\nabla_{\psi}\hat{\mathbb{E}}(Y_n|Y_0^{n-1}) - \nabla_{\psi}\mathbb{E}(Y_n|Y_0^{n-1})$ aparecen expresiones con términos del tipo $\hat{\Phi} - \Phi$ y $\hat{\Phi}' - \Phi'$. Es conocido (ver por ejemplo [9], §3) que

$$\|\hat{\Phi}' - \Phi'\|_{\infty} = O_p\left(\left(\frac{\log N}{N}\right)^{v_2}\right),$$

con $v_2 = (\rho - 1)/(2\rho + 1)$. Por lo tanto obtenemos que:

$$T_3 = O_p\left(\left(\frac{\log N}{N}\right)^{v_1}\right)$$

y

$$T_1 = O_p\left(\left(\frac{\log N}{N}\right)^{v_3}\right)$$

con $v_3 = 1/(2\rho + 2)(2\rho + 1)$ y para T_2 tenemos

$$T_2 = O_p\left(\left(\frac{\log N}{N}\right)^{v_3}\right).$$

Demostración de (ii).

Para demostrar (ii) procedemos como en la parte (ii) del lema 2. ∎

Demostración del teorema 3: Expandiendo $\nabla_{\psi}S_n$ alrededor de ψ_* y se obtiene que

$$(\hat{\psi} - \psi_*) = -\nabla_{\psi}S_N(\psi_*)[\nabla^2_{\psi}S_N(\psi_*) + R_N]^{-1},$$

donde $R_N = \nabla^2_{\psi}S_N(s\hat{\psi} + (1-s)\psi_*) - \nabla^2_{\psi}S_N(\psi_*)$. La ecuación anterior se puede reescribir como

$$\begin{aligned} \hat{\psi} - \psi_* & \\ = \quad & -\nabla_{\psi}\tilde{S}_N(\psi_*)[\nabla^2_{\psi}S_N(\psi_*) + R_N]^{-1} \\ & \qquad + (\nabla_{\psi}\tilde{S}_N(\psi_*) - \nabla_{\psi}S_N(\psi_*))[\nabla^2_{\psi}S_N(\psi_*) + R_N]^{-1} \\ = \quad & O_p(N^{-1/2}) + (\nabla_{\psi}\tilde{S}_N(\psi_*) - \nabla_{\psi}S_N(\psi_*))[\nabla^2_{\psi}S_N(\psi_*) + R_N]^{-1} \end{aligned}$$

para concluir utilizamos el lema 3 se obtiene que,

$$\hat{\psi} - \psi = O_p\left(\left(\frac{\log N}{N}\right)^v\right).$$

eligiendo $v = v_3$. ■

Agradecimientos

A los organizadores de XXX aniversario del Postgrado en Matemática de la Facultad de Ciencias, de la Universidad Central de Venezuela, y a los árbitros por sus precisas sugerencias.

Referencias

[1] P. Ango-Nze and R. Ríos. Density estimation in $\mathbb{L}^\infty$ norm for mixing process. J. Statist. Plann. Inference. **83**(1), 75–90, 2000.

[2] H. J. Bierens. *Topics in advanced econometrics.* Cambridge University Press, 1994.

[3] D. Dacunha-Castelle and M. Duflo. *Probability and Statistics. Volume II.* Springer-Verlag, Berlin, 1986.

[4] R. Douc, E. Moulines and T. Rydén. Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. Ann. Statist. **32**(5), 2254–2304, 2004.

[5] C. Francq and M. Roussignol. Ergocity of autoregressive process with Markov-switching and consitency of the maximum likelihood estimator. Statistics. **32**, 151–173, 1998.

[6] J. D. Hamilton. A new approach to the economic analysis of non stationary time series and the business cycle. Econometrica **57**, 357–384, 1989.

[7] J. L Jensen and N. V. Petersen. Asymptotic normality of the maximum likelihood estimator in state space models. Ann. Statist. **27**(2), 514–535, 1999.

[8] L. Mevel. *Statistique asymtotique pour les modeles Markov cachés.* Tesis de doctorado, Université Rennes I 1997.

[9] R. Ríos. *Utilisation de techniques non paramétriques et semi paramétriques en statistique de données dépendantes.* Tesis de doctorado, Université París IX 1996.

[10] R. Ríos and L. A. Rodríguez. Estimation of the number of states for AR Gaussian processes with Markov regime. Preprint, 2006.

[11] F. Comte and M.-L. Taupin. Semi-parametric estimation in (auto)-regressive β-mixing model with errors-in-variables. Mathematical Methods of Statistics **10**, 121-160, 2001.

[12] Marie-Luce Taupin. Semi-parametric estimation in the nolinear structural errors-in-variable model. Ann. Statist. **29**(1), 66–93, 2001.

[13] J. Yao and J. G. Attali. On stability of nonlinear AR process with Markov switching. Adv. Applied Probab. **32**(2), 394–407, 1999.