UNIVERSIDAD DE PANAMA

VICERRECTORIA DE INVESTIGACION Y POSTGRADO

PROGRAMA DE MAESTRIA EN MATEMATICA

"PREDICTORES LINEALES PARA SERIES TEMPORALES AUTORREGRESIVAS"

Por:

RAFAEL ANTONIO CAMARENA ACUÑA.

Tesis presentada como uno de los requisitos para optar por el grado de Maestro en Ciencias con Especialización en Estadística.

PANAMA, 1986.

UNIVERSIDAD DE PANAMA

VICERRECTORIA DE INVESTIGACION Y POSTGRADO

Aprobado por:

Director de Tesis ______________________
JORGE POLTRONIERI, Ph.D.

Miembro del Jurado ______________________
ELOY GIBBS, Ph. D.

Miembro del Jurado ______________________
ADELA ABAD, M.Sc.

Fecha 28 octubre 1986

DEDICATORIA

A mi esposa, Xenia, a nuestros hijos Rafael, Yasmín y Ambar. A mis padres Silvia y Manuel, quienes constituyen el pilar de mis ánimos de superación.

AGRADECIMIENTO

Quiero dejar constancia de un sincero agradecimiento al Doctor Jorge Poltronieri, Director de este trabajo.

Y a todos mis fieles amigos, que con sus palabras de estímulo y sanos consejos lograron que no desmayase en mí, el espíritu de lucha.

INDICE

INTRODUCCION

El uso de los procesos estocásticos como entes que modelizan fenómenos evólutivos que obedecen leyes probabilísticas, ha tenido un impacto reciente en los estudios que se hacen en el campo de las comunicaciones, teoría de control de la producción; y sobre todo, en lo que se refiere a la aplicación de métodos estadísticos en la econometría [10], [12].

Nos ocuparemos, aquí, del estudio de los procesos estocásticos denominados autorregresivos, que introdujo por primera vez G.U. Yule (1927) para modelizar series temporales asociadas a la observación de manchas solares, y cuya teoría amplió posteriormente Gilbert Walker (1931), quién la aplicó a datos atmosféricos. Sin embargo, subrayamos que el objetivo de esta investigación es, dado un modelo autorregresivo, obtener un procedimiento o fórmula (predictor) que nos lleve a predecir las componentes de una serie temporal con historia conocida.

No todo modelo que describe una serie temporal conduce siempre a un predictor. Es necesario entonces introducir una función de error (pérdida), para así caracterizar dicho predictor; en nuestro caso será el error cuadrático medio E.C.M., aunque es posible encontrar situaciones donde este no sea el procedimiento mas adecuado.

Para la determinación de tales predictores, hemos considerado dos casos: el modelo autorregresivo empleado tiene coeficientes conocidos; y el caso contrario.

En el primer caso hemos seguido la metodología de [1] y [7], y en el segundo nuevamente la de [7] que sigue muy de cerca los resultados de [6].

Para el desarrollo de este escrito, se han propuesto tres capítulos; el primero, llamado de preliminares, consta de todo un conjunto de resultados sobre la convergencia de procesos discretos en Z^+, así como de las interrelaciones naturales entre las formas de convergencia estocástica citadas.

En el segundo se introduce un resultado que sirve de base a la mayor parte de los resultados obtenidos en el tercer capítulo, a saber, la estacionaridad de un proceso a través de la noción de linealidad sobre un ruido blanco. Se establecen, además, criterios para poder determinar la estacionaridad de un proceso autorregresivo a partir del tipo de raíces de su ecuación característica asociada, ver [1], [3] y [7].

Finalmente nuestro tercer capítulo se ocupa de la obtención de predictores lineales para una serie autorregresiva, y de sus propiedades mas importantes.

Es de especial interés en este capítulo, el observar el comportamiento de un predictor lineal en el caso de parámetros desconocidos, de aquí que introduzcamos el concepto de orden de magnitud en probabilidad siguiendo a Mann y Wald (1943) por intermedio de [7] y verifiquemos que las propiedades óptimas de un predictor lineal, en el caso de parámetros conocidos, se mantienen asintóticamente (en probabilidad) en el caso de parámetros desconocidos.

CAPITULO I: PRELIMINARES

1.1. NOCION DE PROCESO ESTOCASTICO DISCRETO

Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y X e Y dos variables aleatorias reales (v.a.r.) centradas y definidas sobre dicho espacio. Denotemos con H el espacio de Hilbert (L_2) con producto interior $E\{XY\}$ y norma inducida $\|X\| = E^{\frac{1}{2}}\{X^2\}$, y con V al espacio vectorial de v.a.r. sobre Ω.

1.1.1. NOTACION

Podemos considerar un proceso estocástico real como la aplicación:

$$\beta : \Lambda \longmapsto V$$

$$\alpha \longmapsto \beta(\alpha) = Z_\alpha$$

donde $Z_\alpha(w) \in R$, $\alpha \in \Lambda$ y $w \in \Omega$.

<u>DEFINICION 1.1.1.1.</u>: Sea $\Lambda \neq \emptyset$ un conjunto arbitrario. Un proceso estocástico es una familia $\beta = \{Z_\alpha\}_{\alpha \in \Lambda}$ de v.a.r sobre Ω $(Z_\alpha \in V)$

<u>DEFINICION 1.1.1.2.</u>: Diremos que $\mathfrak{P}$ es un proceso estocástico discreto si Λ es discreto.

OBSERVACIONES:

* En lo que sigue, restringiremos nuestro estudio a la consideración de procesos en el subespacio H de V, es decir, procesos $\{X_t\}_{t\in\Lambda}$ tales que $E\{X_t^2\} < \infty$ para todo t
* Se supondrá además que $\Lambda = Z$ (o bien Z^+) y diremos que $\mathfrak{P}$ es un proceso doblemente infinito (o una sucesión aleatoria).
* Cuando Λ es contínuo, digamos $\Lambda = [ab] \subseteq R$ con $a < b$, $\mathfrak{P}$ será un proceso contínuo.
* Adoptaremos la notación $\{Z_t\}_t$ en vez de $\{Z_t\}_{t\in Z}$, para designar procesos discretos de parámetro temporal t.
* Para especificar completamente a un proceso discreto $\{Z_t\}_t$, bastaría conocer la distribución conjunta de cualquier subfamilia finita $Z_{t_1}, Z_{t_2}, \ldots, Z_{t_n}$, que notamos con $F(z_1,\ldots z_n;\ t_1\ldots,t_n) = P(Z_{t_1} \leq z_1,\ldots,Z_{t_n} \leq z_n)$, la cual está sujeta a las condiciones de consistencia de Kolmogorov.

1.1.2. TIPOS DE CONVERGENCIA

En el estudio de la convergencia de los procesos discretos ($\Lambda = Z^+$), requeriremos de las herramientas dadas a continuación.

DEFINICION 1.1.2.1.: Una sucesión de v.a.r $\{Z_t\}_t$ converge en probabilidad a la v.a.r Z', y lo notamos

$$Z_t \xrightarrow{P_r} Z'$$

si para todo $\epsilon > 0$

$$\lim_{t\to\infty} P\{|Z_t - Z'| > \epsilon\} = 0.$$

O equivalentemente, si para $\epsilon > 0$ y $\delta > 0$, arbitrarios, existe un N_ϵ tal que si $t > N_{\epsilon,\delta}$

$$P\{|Z_t - Z'| > \epsilon\} < \delta$$

DEFINICION 1.1.2.2.: Si $\{Z_t\}_t$ es una sucesión de v.a.r con funciones de distribución $\{F_{Z_t}\}_t$ entonces Z_t se dirá convergente en ley (o completamente) a la v.a.r. Z' con distribución $F_{Z'}$ y escribimos

$Z_t \xrightarrow{Ley} Z'$ si y solo sí $\lim_{t\to\infty} F_{z_t}(z) = F_{Z'}(z)$ en todo punto de continuidad z de $F_{Z'}$

<u>DEFINICION 1.1.2.3.:</u> Sea $\{Z_t\}_t$ una sucesión de v.a.r. sobre H. Diremos que Z_t converge en media cuadrática (m.c) a $Z' \in H$ si y solo sí

$$\| Z_t - Z' \|_{L_2} \underset{t \to \infty}{\longrightarrow} 0 \quad \text{o bien} \quad E\{ | Z_t - Z' |^2 \} \underset{t \to \infty}{\longrightarrow} 0$$

y lo notamos con $Z_t \xrightarrow{mc} Z'$.

<u>NOTACION:</u> Si dados los procesos $\{X_n\}_{n \in \mathbf{N}}$, $\{Z_t\}_t$ y $\{Y_t\}_t$ donde $X_n = \sum_{i=0}^{n} \theta_i Z_{t-i}$ con $\theta_i \in R$ y t fijo, cuando $X_n \xrightarrow{m,c} Y_t$ escribiremos $Y_t \overset{m,c}{=} \sum_{i=0}^{\infty} \theta_i Z_{t-i}$

o simplemente $Y_t = \sum_{i=0}^{\infty} \theta_i Z_{t-i}$

siendo esta una identidad en t casi segura.

<u>DEFINICION 1.1.2.4.:</u> Sea $\{Z_t\}_t$ una sucesión de v.a.r. y Z' una v.a.r. Z_t converge casi seguramente (c,s.) a Z' y lo notamos con $Z_t \xrightarrow{c,s} Z'$ si y sólo sí existe un evento $A \in \mathcal{A}$ tal que $P(A) = 1$ y para todo $w \in A$

$$Z_t(w) \underset{t \to \infty}{\longrightarrow} Z'(w)$$

El diagrama de abajo señala las relaciones naturales existentes entre los cuatro tipos de convergencia estocástica anteriormente citados.

$$\begin{tikzcd}
m.c \arrow[d] & & c.s \arrow[dll] \arrow[d] \\
P_r \arrow[rr] & & \text{Ley}
\end{tikzcd}$$

<u>TEOREMA 1.1.2.1:</u> $Z_t \xrightarrow{m,c} Z'$ entonces $Z_t \xrightarrow{P_r} Z'$

<u>DEMOSTRACION:</u> La desigualdad de Markov enuncia que para una v.a.r. $X \geqslant 0$, se tiene que $\forall\, r,c \in R^+$,

$$\frac{E\{X^r\}}{c^r} \geqslant P(X \geqslant c),$$ en particular si $r=1$ y

$X = (Z_t - Z')^2$

$$E\{(Z_t-Z')^2\} \geqslant c\,P\{(Z_t-Z')^2 \geqslant c\}\ .$$

Como para $t > N$, dado $\epsilon > 0$,

$$\frac{E\{(Z_t-Z')^2\}}{c} \leqslant \epsilon \quad \text{entonces } P\{(Z_t-Z')^2 \geqslant c\} \leqslant \epsilon\ .$$

<u>LEMA 1.1.2.1.:</u> Sean $\{X_t\}_t$ y $\{Y_t\}_t$ sucesiones de v.a.r. tales que

$$X_t - Y_t \xrightarrow{P_r} 0$$

Si existe una v.a.r. X tal que

$$X_t \xrightarrow{Ley} X,$$

entonces

$$Y_t \xrightarrow{Ley} X$$

DEMOSTRACION: Sean U,V v.a.r. con distribuciones respectivas $F_U(u)$, $F_V(v)$. Se pueden encontrar dos reales positivos ϵ, δ (desigualdad de Markov), tales que

$$P\{|U-V| \geq \epsilon\} \leq \delta$$

esto implica que

$$F_U(u-\epsilon) - \delta \leq F_V(u) \leq F_U(u+\epsilon) + \delta \qquad \text{para todo } u.$$

En efecto, puesto que por definición

$$\begin{aligned} F_U(u-\epsilon) - F_V(u) &= P\{U \leq u-\epsilon\} - P\{V \leq u\} \\ &\leq P\{U \leq u-\epsilon\} - P\{V \leq u \text{ y } U \leq u-\epsilon\} = \\ &= P\{U \leq u-\epsilon \text{ y } V > u\} \\ &\leq P\{(V-U) > \epsilon\} \\ &\leq P\{|V-U| > \epsilon\} \leq \delta \end{aligned}$$

pues $[U \leq u-\epsilon] \supseteq [V \leq u \text{ y } U \leq u-\epsilon]$ y

$[V > u \text{ y } V \leq u-\epsilon] \subseteq [V-U > \epsilon] \subseteq [|V-U| > \epsilon]$.

Análogamente, tenemos

$$
\begin{aligned}
F_V(u) - F_U(u+\epsilon) &= P\{V \leq u\} - P\{U \leq u+\epsilon\} \\
&\leq P\{V \leq u\} - P\{V \leq u \text{ y } U \leq u+\epsilon\} \\
&= P\{V \leq u \text{ y } U > u+\epsilon\} \\
&\leq P\{|U-V| > \epsilon\} \leq \delta .
\end{aligned}
$$

Consideremos ahora un punto x_o donde F_X es contínua. Entonces, para cualquier $\delta > 0$ existe $\eta > 0$ tales que

- $|F_X(x) - F_X(x_o)| < \frac{1}{4}\delta$ siempre que

  $|x - x_o| \leq \eta$ y

- F_X sea contínua en $x_o - \eta$ y $x_o + \eta$.

En particular para δ y η anteriores podemos encontrar un $t_1 > N_1$ de manera que

- $P\{|X_t - Y_t| > \eta\} < \frac{1}{2}\delta$

  dado que $X_t - Y_t \xrightarrow{P_r} 0$.

Así entonces por nuestro resultado inicial tenemos

$$F_{X_t}(x-\eta) - \frac{1}{2}\delta \leq F_{Y_t}(x) \leq F_{X_t}(x+\eta) + \frac{1}{2}\delta$$

para todo x.

Además, como $X_t \xrightarrow{\text{Ley}} X$, existe un N_2 tal que para $t > N_2$,

$$|F_{X_t}(x_o-\eta) - F_X(x_o-\eta)| < \tfrac{1}{4}\delta$$

y

$$|F_{X_t}(x_o+\eta) - F_X(x_o+\eta)| < \tfrac{1}{4}\delta$$

En consecuencia si $N = \text{máx}\,\{N_1, N_2\}$ y $t > N$

$$F_{X_t}(x_o+\eta) + \tfrac{1}{2}\delta < F_X(x_o+\eta) + \tfrac{3}{4}\delta \text{ y}$$

$$F_X(x_o-\eta) - \tfrac{3}{4}\delta < F_{X_t}(x_o-\eta) - \tfrac{1}{2}\delta$$

Por otro lado como $x_o + \eta$ y $x_o - \eta$ son puntos de continuidad de F_X, se tiene también que

- $F_X(x_o) - \delta < F_X(x_o-\eta) - \tfrac{3}{4}\delta$ y
- $F_X(x_o) + \delta > F_X(x_o+n) + \tfrac{3}{4}\delta$

Sabiendo que

$$F_{X_t}(x_o-\eta) - \tfrac{1}{2}\delta \leq F_{Y_t}(x_o) \leq F_{X_t}(x_o+\eta) + \tfrac{1}{2}\delta$$

se deduce el resultado

$$F_X(x_o) - \delta < F_X(x_o-\eta) - \tfrac{3}{4}\delta < F_{Y_t}(x_o) < F_X(x_o+\eta) + \tfrac{3}{4}\delta < F_X(x_o) + \delta$$

lo cual concluye la demostración.

<u>COROLARIO 1.1.2.1.:</u> $Z_t \xrightarrow{P_r} Z'$ implica $Z_t \xrightarrow{\text{Ley}} Z'$

<u>DEMOSTRACION:</u> Se sigue del lema 1.1.2.1. tomando $Y_t = Z'$.

Sea el evento $A_t = [\sup_{t' \geq t} | Z_{t'} - Z' | \leq \epsilon]$ con $\epsilon > 0$ arbitrario, tenemos que:

(a) $A_t = \bigcap_{t' \geq t} [\, | Z_{t'} - Z' | \leq \epsilon]$

(b) $\{A_t\}_t$ es creciente.

De la definición 1.1.2.4, si $Z_t \xrightarrow{c,s} Z'$, dado $w \in A$ existe $t(\epsilon, w)$ tal que $w \in A_t$ y $w \in \bigcup_t A_t$; de donde $A \subseteq \bigcup_t A_t$ y $P(\bigcup_t A_t) = 1$. Ahora bien, como P es una medida positiva y por el resultado (b) se tiene

$$P(\bigcup_t A_t) = \lim_t P(A_t) = 1$$

lo que significa que

$$\sup_{t' \geq t} | Z_{t'} - Z' | \xrightarrow[t \to \infty]{P_r} 0$$

TEOREMA 1.1.2.2.: $Z_t \xrightarrow{c,s} Z'$ implica $Z_t \xrightarrow{p_r} Z'$

DEMOSTRACION: Sea $X_t = \sup\limits_{t' \geq t} |Z_{t'} - Z'|$, entonces

si $Z_t \xrightarrow{c,s} Z'$ sabemos que $X_t \xrightarrow[t \to \infty]{p_r} 0$.

Sea ahora $\epsilon > 0$

$$[X_t \leq \epsilon] = \bigcap_{t' \geq t} [|Z_{t'} - Z'| \leq \epsilon] \subseteq [|Z_t - Z'| \leq \epsilon]$$

como $P\{[X_t \leq \epsilon]\} \xrightarrow[t \to \infty]{} 1$ y $P\{[X_t \leq \epsilon]\} \leq P\{[|Z_t - Z'| \leq \epsilon]\}$

concluímos $P\{[|Z_t - Z'| \leq \epsilon]\} \xrightarrow[t \to \infty]{} 1$ o sea

$Z_t \xrightarrow{p_r} Z'$.

1.2. SERIE DE TIEMPO DISCRETA

Es frecuente, en la práctica, observar fenómenos cuyas componentes están ligadas entre sí mediante formas particulares de interdependencia (es lo que llamaríamos un <u>sistema</u>). Cuando esta interdependiencia es de naturaleza compleja, se hace necesario plantear el estado o situación del sistema desde un punto de vista aleatorio.

Un proceso estocástico será en consecuencia el modelo matemático llamado a representar la variación de los fenómenos temporales que producen la evolución del mundo.

Así entonces nuestro módelo tendrá la forma:

$$X_t = X(w,t)$$

donde t: es el parámetro (tiempo) del modelo

w: un suceso elemental.

Observemos que si $w = w_o$ $H_{w_o}(t) = X(w_o,t)$, es una función temporal que representa el estado del sistema bajo el suceso w_o. Llamaremos a esta función (no aleatoria) una <u>realización del proceso</u>, o simplemente una <u>serie temporal</u> de parámetro t.

De esta manera el proceso discreto $\{X_t\}_t$ engendra a la serie

$$X(w_o,t_1),\ X(w_o,t_2),\ldots,\ X(w_o,t_n);$$

que llamaremos <u>serie temporal discreta</u> o de parámetro discreto. Por consiguiente nos abocaremos, aquí, al estudio de procesos discretos como modelos probabilísticos en la interpretación de una serie temporal discreta, dichas series constituyen la motivación de este trabajo.

1.2.1. REALIZACION DE UN PROCESO ESTOCASTICO

DEFINICION 1.2.1.1: Sea $\{X_t\}_t$ un proceso discreto. Una realización del proceso X_t, dado $w_o \in \Omega$, es la función

$$H_{w_o} : Z \longrightarrow R$$

$$t \longrightarrow H_{w_o}(t) = X(w_o, t)$$

<u>OBSERVACIONES</u>

* Dado $w_o \in \Omega$, existe una única realización H_{w_o} asociada a $\{X_t\}_t$.
* Cuando $t = t_o$, $X_{t_o} = X(w,t_o)$, es una v.a.r. definida en Ω
* Se puede considerar entonces un proceso estocástico $\{Z_t\}_t$ como la familia de posibles realizaciones $\{Z(w,t)\}_{\substack{t \in \Lambda \\ w \in \Omega}}$

1.2.2. EJEMPLO

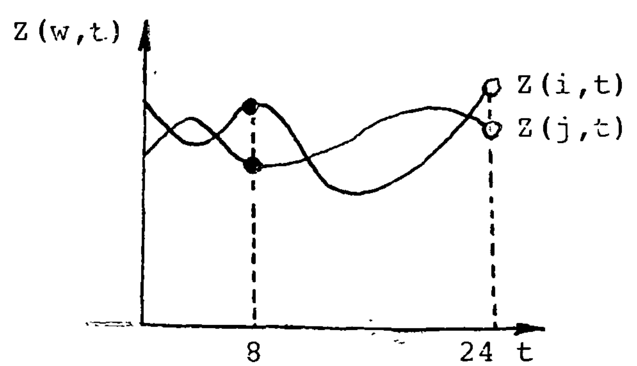
Sea el proceso Z_t = Temperatura de cierta localidad en el instante t.

Entonces si $t \in \{1,2,\ldots,24\} = \Lambda$,la serie

$$Z(i,1),\quad Z(i,2),\ldots,\quad Z(i,24)$$

representa la sucesión de temperaturas observadas el i-ésimo día, a intervalos (iguales) de una hora.

Gráficamente



Como en este caso Λ es finito, cada realización será una curva finita, sin embargo, en general consideramos una realización como una parte de la curva que la representa $(1 \leq t \leq T)$.

Por otro lado si t=8, Z(i,8) y, Z(j,8) son valores de la variable aleatoria Z_8 (puntos negros) que representan las temperaturas de dicha localidad después de 8 horas de observación y en dos días distintos i, j.

Notemos que, por ejemplo, Z_8 y Z_{24} pertenecen al mismo proceso pero en general no están igualmente distribuídos aunque usualmente supondremos procesos con distribución de probabilidades común.

Dado que para cualquier proceso Z_t,

$$Z_t : (\Omega, \mathcal{A}) \longrightarrow (R , B_R)$$

donde B_R denota el σ-álgebra de Borel sobre R ; tal que cualquiera de sus elementos es combinación operacional (uniones e intersecciones numerables) de intervalos de la forma

$$(a , b] = \{ Z(w,t) , w \in \Omega \ / \ a < Z(w,t) \leq b \}$$

se tendrá así que

$$P^{Z_t} \{ B \} = P \{ Z_t^{-1}(B) \in \mathcal{A} \} , B \in B_R$$

es la ley de probabilidades de Z_t.

1.2.3. MEDIA AUTOCOVARIANZA Y AUTOCORRELACION

Usualmente no se conocen las distribuciones finito dimensionales de un proceso, por lo que frecuentemente se recurre al uso de los momentos, generalmente los de primero y segundo orden, para tener alguna idea de la estructura probabilística del mismo.

Definimos estos momentos así:

- Función de medias

$$m(t) = E\{ Z_t \}$$

Función de autocovarianzas

$$\gamma(t_1,t_2)=\text{cov}\{Z_{t_1}, Z_{t_2}\} = E\{[Z_{t_1}-m(t_1)][Z_{t_2}-m(t_2)]\},$$

donde si $t = t_1 = t_2$, se tiene la Función de varianzas

$$\sigma_t^2 = \gamma(t) = \text{var}\{Z_t\} .$$

Por otro lado, como al analizar una serie temporal es a menudo imposible tomar mas de un valor $Z(w,t)$ para cada variable Z_t, confundiremos en lo que viene ambas notaciones. Así, podemos referirnos a la serie temporal Z_t, $1 \leqslant t \leqslant N$. Tenemos así los estimadores:

. $$\bar{Z} = m(t) = \frac{\sum_{t=1}^{N} Z_t}{N}$$

. $$c_k = \hat{\gamma}(k) = \frac{\sum_{t=1}^{N-k} (Z_t-\bar{Z})(Z_{t+k}-\bar{Z})}{N}$$

Los cuales llamaremos media total y autocovarianza para rezago k, y son los estimadores de $m(t)$ y $\gamma(k)$ respectívamente.

Además se puede medir (teóricamente) la correlación lineal entre cualquier par de observaciones a una distancia temporal k Z_t, Z_{t+k}, $1 \leqslant t \leqslant N-k$ usando el coeficiente de autocorrelación para un rezago k, definido por

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)}, \quad k = 0,1...$$

cuyo estimador usual es

$$r_k = \hat{\rho}(k) = \frac{\sum_{t=1}^{N-k} (z_t-\bar{z})(z_{t+k}-\bar{z})}{\sum_{t=1}^{N} (z_t-\bar{z})^2} = \frac{c_k}{c_o} \quad , \quad k \leq [N/4]$$

Notemos que los estimadores $\bar{z}$, c_k y r_k sólo tienen sentido práctico cuando la serie considerada no demuestra cambios tendenciales, puesto que de lo contrario serían funciones de t .

Debemos definir, entonces, estos estimadores bajo un modelo particular para series temporales donde la función de media no dependa del tiempo, y la función de autocovarianza (autocorrelación) no esté influenciada por alguna tendencia.

Los modelos probabilísticos en cuestión son llamados procesos estocásticos estacionarios.

CAPITULO 2: PROCESOS ESTACIONARIOS

2.1. ESTACIONARIDAD DE UN PROCESO

Cuando se estudia probabilísticamente una serie temporal, generalmente conviene que el proceso estocástico asociado mantenga estabilidad aleatoria de suerte que haya seguridad de que su comportamiento sea homogéneo alrededor de una media fija (constante en t). Es esta una de la propiedades deseables al analizar un proceso, tal propiedad (aunque no tan común) es la llamada estacionaridad de un proceso estocástico discreto.

2.1.1. ESTACIONARIDAD DEBIL Y FUERTE

DEFINICION 2.1.1.1.: Un proceso $\{X_t\}_t$ es llamado débilmente estacionario (estacionario de segundo orden o estacionario en covarianza) si satisface las siguientes condiciones:

a) $m = E\{X_t\}$

b) $\gamma(t, t+k) = \operatorname{cov}\{X_t, X_{t+k}\} = \gamma(k)$

para cualquier $t, t+k$.

Notemos que la definición anterior nos asegura que un proceso será estacionario débil si:

a) Su función de medias es constante en t.

b) Su función de autocovarianzas sólo depende de la distancia temporal k (rezago) entre las variables involucradas.

En particular si $k = 0$, $\gamma(0) = \text{var}\{X_t\} = \sigma^2 < \infty$ es decir, la varianza del proceso es constante y finita o sea $X_t \in H$.

Nuestro estudio posterior sobre procesos discretos se limitará al estudio de aquellos, que quedarán caracterizados por sus momentos de primero y segundo orden (los cuales llamamos estacionarios débiles), sin embargo, esta caracterización pierde rigor probabilístico, ya que podemos encontrar dos o mas procesos con distribuciones de probabilidad diferentes aunque sus momentos de primero y segundo orden sean comúnes. De allí entonces que empleeemos el término "débil".

EJEMPLO

Sea $\{e_t\}_t$ un proceso de media constante, digamos $E\{e_t\} = 0$, tal que

$$\gamma(k) = \begin{cases} \sigma_e^2 & \text{si } k = 0 \\ 0 & \text{si } k \neq 0 \end{cases}$$

es su función de autocovarianzas. e_t es un proceso estacionario débil que llamaremos proceso de errores o ruído blanco.

TEOREMA 2.1.1.1.: Para todo proceso $\{X_t\}_t$ débilmente estacionario con función de autocovarianza $\gamma(k) = \text{cov}\{X_t, X_{t+k}\}$ se tiene

a) $\gamma(0) \geq 0$

b) $\gamma(k) = \gamma(-k)$

c) $|\gamma(k)| \leq \gamma(0)$

d) $\gamma(k)$ es semidefinida positiva, o sea

$$\forall\ (\alpha_1,\ldots,\alpha_n) \in R^n \quad , \quad \forall\ (t_1, t_2,\ldots,t_n) \in Z^n$$

obtenemos $\displaystyle\sum_{j=1}^{n} \sum_{k=1}^{n} \alpha_i \alpha_k \ \gamma(t_j - t_k) \geq 0$

DEMOSTRACION: Los resultados (a) y (c) son conocidos del cálculo de probabilidades elemental.

El resultado (b) es inmediato toda vez que si $t,\ t+k \in Z$,

$$\gamma(k) = \text{cov}\{X_t, X_{t+k}\} = \text{cov}\{X_{t+k}, X_t\} = \gamma(-k).$$

Para probar (d) podemos suponer sin perder generalidad que $E\{X_t\} = 0$ para todo t (si $E\{X_t\} = m \neq 0$, tomamos el proceso centrado $\tilde{X}_t = X_t - m$).

Luego, dado el proceso $Y_t = \sum_{j=1}^{n} \alpha_j X_{t_j}$, tenemos

$$0 \leq \gamma_Y(0) = E\{\sum_{j=1}^{n} \alpha_j X_{t_j}\}^2 = E\{\sum_{j=1}^{n} \alpha_j X_{t_j} \sum_{i=1}^{n} \alpha_i X_{t_i}\} =$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j E\{ X_{t_i} X_{t_j} \} = \sum_{i,j=1}^{n} \alpha_i \alpha_j \gamma(t_j - t_i)$$

<u>TEOREMA 2.1.1.2.</u>: Sea $\{X_t\}_t$ un proceso débilmente estacionario y $\{Y_t\}_t$ tal que

$$Y_t = \sum_{i=1}^{k} \alpha_i X_{t-i} \quad \text{con } \alpha_i \in R \ , \ i = 1,\ldots,k$$

entonces Y_t es débilmente estacionario.

<u>DEMOSTRACION</u>: En efecto, si $E\{X_t\} = 0$ y $\gamma_X(0) = \sigma^2$ tenemos

1) $E\{Y_i\} = 0$

2) $\gamma_Y(t,t') = \text{cov}\{Y_t, Y_{t'}\} = \sum_{i=1}^{K} \sum_{j=1}^{K} \alpha_i \alpha_j \gamma_X(t-i, t'-j)$

$$= \sum_{i,j=1}^{K} \alpha_j \alpha_j \gamma_X \{(t'-t) + (i-j)\} = \gamma_Y(t'-t)$$

donde si $t' = t$.

$$\gamma_Y(0) = \text{var}\{Y_t\} = \sum_{i,j=1}^{k} \alpha_i \alpha_j \gamma_X(i-j).$$

Del teorema 2.1.1.1. se deduce que para cualesquier parejas X_t, $X_{t'}$ y X_{t+k}, $X_{t'+k}$, situadas a una misma distancia temporal k, podemos escribir

$\operatorname{cov}\{X_t, X_{t'}\} = \gamma(|t'-t|) = \operatorname{cov}\{X_{t+k}, X_{t'+k}\}.$

por lo que bastaría definir a γ sólo en Z^+.

Además, si $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ es una subfamilia finita del proceso X_t, la matríz de covarianzas del proceso.

$$V_n = \begin{bmatrix} \gamma(0) & \gamma(t_1-t_2) & \cdots & \gamma(t_1-t_n) \\ \gamma(t_2-t_1) & \gamma(0) & \cdots & \gamma(t_2-t_n) \\ \vdots & & & \vdots \\ \gamma(t_n-t_1) & \gamma(t_n-t_2) & \cdots & \gamma(0) \end{bmatrix}$$

es simétrica y de determinante no negativo.

Consideremos ahora la función de autocorrelación de un proceso estacionario débil, la cual definimos como

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} \quad , \quad k = 0,1,\ldots$$

por el Teorema 2.1.1.1., se ve fácilmente que $\rho(0) = 1$, $|\rho(k)| \leq 1$. Más aún, la matríz de autocorrelación

$$\Gamma_n = \gamma^{-1}(0)\, V_n \quad , \quad \gamma(0) \neq 0$$

goza de las mismas propiedades que V_n.

El siguiente teorema, para funciones de autocovarianzas, nos muestra el papel preponderante que juegan estas, en la determinación de un proceso estacionario débil.

<u>TEOREMA 2.1.1.3.:</u> Dada la familia numérica $\{\gamma(k)\}_k$ las siguientes proposiciones son equivalentes:

a) $\{\gamma(k)\}_k$ es la función de autocovarianzas de un proceso $\{X_t\}_t$, débilmente estacionario.

b) Existe una única función $F(\lambda)$ monótona no decreciente (en $[-\pi\ \pi]$), contínua a derecha (en $]-\pi\ \pi[$), y tal que $F(-\pi) = 0$, $F(\pi) = \gamma(0)$ y

$$\gamma(k) = \int_{-\pi}^{\pi} \cos \lambda k \ dF(\lambda) \qquad k = 0, \pm 1, \ldots$$

<u>DEMOSTRACION</u>: Ver [1] y [5]

<u>CONSECUENCIAS</u>:

* Se demuestra que $F(\lambda) = \lim\limits_{T} \left\{ \frac{1}{2\pi} \gamma(0) \ [\lambda + \pi] + \right.$

$$\left. + \frac{1}{\pi} \sum_{k=1}^{T-1} (1 - \frac{k}{T}) \gamma(k) \frac{\operatorname{sen} \lambda k}{k} \right\}$$, lo cual tiene sentido

si $\sum\limits_{k=1}^{\infty} \gamma(k) < \infty$. En cuyo caso $f(\lambda) = F'(\lambda)$,

siendo $f(\lambda)$ la densidad espectral o espectro de X_t

* $\rho(k) = \int_{-\pi}^{\pi} \cos \lambda k \ d\ G(\lambda)$ con $G(\lambda) = \frac{F(\lambda)}{\gamma(0)}$, $\gamma(0) \neq 0$.

* Cuando $f(\lambda) = \frac{1}{2\pi} \sum_{k=0}^{\infty} \gamma(k) \cos \lambda k < \infty$, si X_t es estacionario débil, existe la transformación inversa (transformada inversa de Fourier)

$$\gamma(k) = \int_{-\pi}^{\pi} \cos \lambda k \, d\, F(\lambda).$$

Esto significa que, dado un proceso estacionario débil, existe una correspondencia uno a uno entre su función de autocovarianzas (o autocorrelación) y su espectro (si existe).

De este modo, cuando sea conveniente (casos de series temporales con supuestas componentes periódicas con frecuencias λ, diversas), se caracterizarán dichos procesos por su media, varianza y su espectro en vez de emplear sus funciones de autocovarianzas (o autocorrelación)

Si una serie temporal se analiza mediante este criterio, se dice que el análisis es hecho en su dominio de frecuencia. En caso contrario el análisis será hecho en su dominio temporal. Este último es el que enfoque adoptado aquí.

<u>DEFINICION 2.1.1.2.:</u> Un proceso $\{X_t\}_t$ se dice fuertemente estacionario (estacionario en sentido escricto), si conserva sus leyes de probabilidad finito-dimensionales por cualquier traslación de índices.

Es decir

$\{X_t\}_t$ es estacionario fuerte si y solo sí

$\forall\ (t_1,t_2,\ldots,t_n),(t_1+h,\ t_2+h,\ldots,t_n+h)\in Z^n$, n finito.

$$F(x_1,x_2,\ldots,x_n;t_1,t_2,\ldots,t_n) = F(x_1,x_2,\ldots,x_n;t_1+h,t_2+h,\ldots,t_n+h).$$

OBSERVACIONES:

* Se puede considerar a un proceso estacionario fuerte como un proceso cuya estructura aleatoria aparece invariante en el tiempo.

* De la condición de compatibilidad en la estructura aleatoria de un proceso, se deduce que

  . $F(x;t) = F(x;t+h) \qquad t,t+h \in Z$

  . $F(x,x';t,t') = F(x,x';t+h,t'+h) = F(x,x';0,t'-t)$,

  tomando $h = -t$.

  De aquí que todo proceso estacionario fuerte es también estacionario débil.

EJEMPLO: Consideremos el proceso de errores o ruído $\{e_t\}_t$ al cual agregamos la condición de que los e_t sean independientes e idénticamente distribuídos. Entonces $\{e_t\}_t$ es estacionario fuerte (llamaremos a este proceso puramente aleatorio).

En efecto, dados $h,\ t_1,\ldots,\ t_m \in Z$

$$\begin{aligned} F(e_1,e_2,\ldots,e_m;t_1,t_2,\ldots t_m) &= F(e_1;t_1)\ F(e_2;t_2)\ldots F(e_m;t_m) \\ &= F(e_1;t_1+h)F(e_2;t_2+h)\ldots F(e_m;t_m+h) \\ &= F(e_1,e_2,\ldots,e_m;t_1+h,t_2+h,\ldots t_m+h) \end{aligned}$$

<u>DEFINICION 2.1.1.3.:</u> Un proceso $\{X_t\}_t$ se dice Gaussiano si , para $t_1, t_2, \ldots, t_n \in Z$, las variables $Z_{t_1}, Z_{t_2}, \ldots, Z_{t_n}$ tienen distribución normal n-variada.

<u>TEOREMA 2.1.1.4.:</u> Todo proceso Gaussiano $\{X_t\}_t$ débilmente estacionario es fuertemente estacionario.

<u>DEMOSTRACION:</u> Esto es inmediato puesto que cualquier familia $X_{t_1}, \ldots X_{t_k}$, con k finito, tendrá distribución normal k-variada con matríz de medias $M = (m_x \ldots, m_x)$ y matríz de covarianzas $V_k = (\gamma_X(t_i - t_j))_{i,j \in \{1,\ldots,k\}}$ comunes, toda vez que X_t es un proceso tal que $E\{X_t\} = m_X$, y $\gamma_X(t,t') = \gamma_X(t-t')$ para t,t' arbitrarios.

<u>COROLARIO 2.1.1.1.:</u> Si $\{X_t\}_t$ es Gaussiano, entonces, las nociones de estacionaridad débil y fuerte son equivalentes.

<u>NOTA:</u> En lo sucesivo designaremos a los procesos estacionarios débiles como estacionarios, a menos que, según la necesidad del caso, se deba especificar el tipo de estacionaridad.

2.1.2. PROCESOS LINEALES. OPERADORES DE BOX Y JENKINS.

DEFINICION 2.1.2.1: Un proceso $\{X_t\}_t$ se llama lineal sobre el ruido $\{e_t\}_t$ si existe una familia de reales $\{\psi_j\}_{j \in Z^+}$ de modo que

$$X_t = \sum_{j=0}^{\infty} \psi_j \, e_{t-j}$$

El teorema siguiente, nos presenta una caracterización de los procesos lineales.

TEOREMA 2.1.2.1: Sea el proceso $\{Y_t\}_t$ y la sucesión de reales $\{\alpha_j\}_{j \in Z^+}$ tales que

. $E\{Y_t^2\} \leq K$ para todo $t \in Z$ y algún K finito.

. $\sum_{j=0}^{\infty} |\alpha_j| < \infty$

Entonces existe el proceso $\{X_t\}_t$ de modo que

$$X_t = \sum_{j=0}^{\infty} \alpha_j \, Y_{t-j}$$

donde $E\{X_t^2\} < \infty$ para todo $t \in Z$ y X_t es unívocamente determinado excepto para un evento de probabilidad nula.

DEMOSTRACION: Sea $\epsilon > 0$, podemos escoger $N_\epsilon > 0$ tal que

$$\left(\sum_{j=N_\epsilon}^{\infty} |\alpha_j| \right) K < \epsilon$$

Sean, ahora, t fijo y $n > m > N_\epsilon$. Entonces

$$E\left\{ \left| \sum_{j=0}^{n} \alpha_j Y_{t-j} - \sum_{j=0}^{m} \alpha_j Y_{t-j} \right|^2 \right\} = E\left\{ \left| \sum_{j=m+1}^{n} \alpha_j Y_{t-j} \right|^2 \right\}.$$

$$\leq \sum_{j=m+1}^{n} |\alpha_j|^2 \cdot E\{ |Y_{t-j}|^2 \} +$$

$$+ 2 \sum_{j<k=m+1}^{n} |\alpha_j| |\alpha_k| E\{ |Y_{t-j}| |Y_{t-k}| \}$$

Puesto que de $(|Y_j| - |Y_k|)^2 \geq 0$, tenemos

$E\{ |Y_j| |Y_k| \} \leq K$ para todo j,k.

Se obtiene

$$E\left\{ \left| \sum_{j=0}^{n} \alpha_j Y_{t-j} - \sum_{j=0}^{m} \alpha_j Y_{t-j} \right|^2 \right\} \leq K \sum_{j=m+1}^{n} |\alpha_j|^2 + 2K \sum_{j<k=m+1}^{n} |\alpha_j| |\alpha_k|$$

$$\leq K \left(\sum_{j=N_\epsilon}^{\infty} |\alpha_j| \right)^2$$

Lo cual implica que $\left\{ \sum_{j=0}^{n} \alpha_j Y_{t-j} \right\}_{n \in N}$ es una sucesión de Cauchy en H, por lo que existe $X_t \in H$ o sea $E\{ X_t^2 \} < \infty$ tal que

$$X_t = \sum_{j=0}^{\infty} \alpha_j Y_{t-j}$$

Por otro lado, consideremos un proceso $\{X_t^*\}_t$ con las propiedades de $\{X_t\}_t$. Para $\epsilon > 0$ y t fijo se tiene

$$E\{(X_t - X_t^*)^2\} = E\{(X_t - \sum_{j=0}^{n} \alpha_j Y_{t-j} + \sum_{j=0}^{n} \alpha_j Y_{t-j} - X_t^*)^2\}$$

$$\leq [E|X_t - \sum_{j=0}^{n} \alpha_j Y_{t-j}|^2]^{\frac{1}{2}} + [E|\sum_{j=0}^{n} \alpha_j Y_{t-j} - X_t^*|^2]^{\frac{1}{2}}$$

ya que la desigualdad de Minkowski en H establece que

$$\{E(|X+Y|)^2\}^{\frac{1}{2}} \leq \{E(|X|^2)\}^{\frac{1}{2}} + \{E(|Y|^2)\}^{\frac{1}{2}}$$

Entonces como para $n > N_\epsilon$

$$E\{|X_t - \sum_{j=0}^{n} \alpha_j Y_{t-j}|^2\} < \epsilon \quad , y$$

$$E\{|X_t^* - \sum_{j=0}^{n} \alpha_j Y_{t-j}|^2\} < \epsilon \qquad \text{para } n > M_\epsilon$$

si tomamos $L_\epsilon = \max\{N_\epsilon, M_\epsilon\}$, deducimos

$E\{(X_t - X_t^*)^2\} \leq 4\epsilon$ para $n > L_\epsilon$. Luego

$E\{(X_t - X_t^*)^2\} = 0$ pues $\epsilon > 0$ es arbitrario, esto implica que $X_t = X_t^*$ casi seguramente para cada t. Así $\{X_t\}_t$ es único casi seguramente.

TEOREMA 2.1.2.2.: Sean las sucesiones de reales absolutamente sumables $\{\alpha_j\}_{j\in Z^+}$, $\{\beta_j\}_{j\in Z^+}$ y $\{Z_t\}_t$ un proceso estocástico tal que $E\{Z_t^2\}<\infty$ para todo $t \in Z$, definimos

$$X_t = \sum_{j=0}^{\infty} \alpha_j Z_{t-j}$$

y

$$Y_t = \sum_{k=0}^{\infty} \beta_k Z_{t-k}$$

Entonces

$$E\{X_t\} = \lim_{n\to\infty} \sum_{j=0}^{n} \alpha_j E\{Z_{t-j}\}$$

$$E\{X_t Y_t\} = \lim_{n\to\infty} \sum_{j=0}^{n} \sum_{k=0}^{n} \alpha_j\beta_k E\{Z_{t-j}Z_{t-k}\}$$

DEMOSTRACION: El teorema 2.1.2.1 nos permite construír los procesos $\{X_t^*\}_t$, $\{Y_t^*\}_t$ tales que

$$X_t^* = \sum_{j=0}^{\infty} |\alpha_j||Z_{t-j}|$$

$$Y_t^* = \sum_{j=0}^{\infty} |\beta_j||Z_{t-j}| \quad \text{para todo} \quad t \in Z$$

donde $E\{X_t^*\}^2 < \infty$ y $E\{Y_t^*\}^2 < \infty$

Tenemos que de $|X|^2 + 1 > |X|$ se obtiene

$$E\{|X_t^*|\} < E\{X_t^*\}^2$$

$$E\{|Y_t^*|\} < E\{Y_t^*\}^2$$

Además del teorema 1.1.2.1. $\sum_{j=0}^{n} \alpha_j Z_{t-j} \xrightarrow{P_r} X_t$

y $\sum_{j=0}^{n} \beta_j Z_{t-j} \xrightarrow{P_r} Y_t$. Por otro lado como

$$\left|\sum_{j=0}^{n} \alpha_j Z_{t-j}\right| \leq \sum_{j=0}^{n} |\alpha_j||Z_{t-j}| \leq X_t^*$$

$$\left|\sum_{j=0}^{n} \beta_j Z_{t-j}\right| \leq \sum_{j=0}^{n} |\beta_j||Z_{t-j}| \leq Y_t^*$$

y dado que $|XY| \leq \frac{1}{2}(X^2 + Y^2)$, se obtiene

$$\left|\left(\sum_{j=0}^{n} \alpha_j Z_{t-j}\right)\left(\sum_{k=0}^{n} \beta_k Z_{t-k}\right)\right| \leq \frac{1}{2}\left\{\left(\sum_{j=0}^{n} |\alpha_j||Z_{t-j}|\right)^2 + \left(\sum_{k=0}^{n} |\beta_k||Z_{t-k}|\right)^2\right\}$$

$$\leq \frac{1}{2}(X_t^*)^2 + \frac{1}{2}(Y_t^*)^2.$$

Así entonces, empleando el teorema de convergencia dominada (ver [9] pág 152) se tienen los límites usuales.

$$\lim_{n \to \infty} E\left\{\sum_{j=0}^{n} \alpha_j Z_{t-j}\right\} = E\{X_t\}$$

$$\lim_{n \to \infty} E\left\{\left(\sum_{j=0}^{n} \alpha_j Z_{t-j}\right)\left(\sum_{k=0}^{n} \beta_k Z_{t-k}\right)\right\} = \lim_{n\to\infty} \sum_{j,k=0}^{n} \alpha_j \beta_k E\{Z_{t-j} Z_{t-k}\} = E\{X_t Y_t\}$$

<u>TEOREMA 2.1.2.3.:</u> Sea $\{X_t\}_t$ un proceso lineal sobre el ruído $\{e_t\}_t$ con pesos $\psi_0, \psi_1, \ldots,$ en $\mathbb{R}$, entonces X_t es estacionario si $\sum_{j=0}^{\infty} \psi_j$ converge absolutamente.

<u>DEMOSTRACION:</u> Como X_t es lineal sobre e_t, tenemos que

$$X_t = \sum_{j=0}^{\infty} \psi_j e_{t-j} \quad ;$$

luego por el Teorema 2.1.2.2. y dado que $\sum_{j=0}^{\infty} \psi_j$

. $E\{X_t\} = \sum_{j=0}^{\infty} \psi_j E(e_{t-j}) = 0$

. $\gamma(t,t+k) = E\{X_t X_{t+k}\} = \sigma^2 \sum_{j=k}^{\infty} \psi_j \psi_{j-k} = \gamma(k)$

En particular si $k = 0$

. $\operatorname{var}\{X_t\} = \gamma(0) = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2$

Observemos que esta dos últimas expresiones tienen sentido toda vez que las series

$$\sum_{j=k}^{\infty} \psi_j \psi_{j-k} \quad \text{y} \quad \sum_{j=0}^{\infty} \psi_j^2 \quad \text{son convergentes.}$$

<u>DEFINICION 2.1.2.2.:</u> Sea D el espacio de procesos estocásticos discretos sobre Ω . Llamaremos operadores de Box-Jenkins (ver [3], cap. 3) a los operadores de D en D, B (operador de retraso temporal) y F (operador de adelanto temporal), tales que

. $BZ_t = Z_{t-1}$. $FZ_t = Z_{t+1}$

PROPIEDADES DE B Y F.

a) $B = F^{-1}$

n) El modelo $X_t = \sum_{j=0}^{k} \psi_j Y_{t-j}$ con $\psi_0 = 1$

Se puede expresar en términos de B como

$$X_t = \left(\sum_{j=0}^{k} \psi_j B^j\right) Y_t \quad \text{donde} \quad B^j = \underbrace{BB\ldots B}_{j\text{-veces}}$$

y $B^0 = 1$

c) De modo mas general, se puede considerar el modelo

$X_t = \sum_{j=0}^{\infty} \psi_j \; Y_{t-j}$ y expresarlo en la forma

$$X_t = (\sum_{j=0}^{\infty} \psi_j \; B^j) Y_t \;=\; \lim_{k\to\infty} \; (\sum_{j=0}^{k} \psi_j \; B^j) Y_t \text{ (en m,c).}$$

d) B y F son operadores lineales en D.

Notemos que en (c), si $\{Y_t\}_t$ es estacionario y si

$\psi(B) = 1 + \sum_{j=1}^{\infty} \psi_j \; B^j$, tenemos el diagrama

$$Y_t \longrightarrow \psi(B) \longrightarrow X_t.$$

el cual corresponde a un modelo de filtro lineal (ver [2], [3]) donde $\psi(B)$ es llamada función de transferencia del filtro cuya entrada es Y_t y cuya salida es X_t.

Es de enorme importancia, en este caso, el conocer el operador $\psi(B)$, pues esto nos permite "controlar" la serie de salida para valores hipotéticos de entrada.

Algo análogo ocurre con la entrada si existe $\psi^{-1}(B)$.

<u>EJEMPLO</u>: Sea $\{a_t\}_t$ un ruído blanco y $\{X_t\}_t$ un proceso lineal sobre a_t tal que $X_t = a_t - \theta a_{t-1}$

Si hacemos $\psi_1 = -\theta$, $\psi_j = 0$, $j > 1$ tenemos

$\sum_{j=0}^{\infty} \psi_j = 1 - \theta < \infty$ y por Teorema 2.1.2.3. $\{X_t\}_t$ es estacionario.

Entonces escribimos $X_t = (1 - \theta B)\ a_t = \psi(B) a_t$,

de donde $a_t = (1 - \theta B)^{-1} X_t = (\sum_{j=0}^{\infty} \theta^j B^j) X_t$

siempre que $\sum_{j=0}^{\infty} \theta^j B^j$ converja para $|B| \leq 1$,

pero $\psi(B)^{-1} = (1 - \theta B)^{-1} = \sum_{j=0}^{\infty} \theta^j B^j$ existe si

$|\theta| < 1$ (pues $\sum_{j=0}^{\infty} \theta^j B^j \leq \frac{1}{1-\theta}$) en cuyo caso el

diagrama $a_t \longrightarrow \psi(B) \longrightarrow X_t$ es invertible, es decir, el proceso X_t es invertible.

<u>TEOREMA 2.1.2.4.:</u> Un proceso lineal será estacionario si la serie $\psi(B)$ converge para $|B| \leq 1$ e invertible si $\psi(B)^{-1}$ converge para $|B| \leq 1$

<u>DEMOSTRACION</u>: ver [3]

2.1.3. PROCESOS AUTORREGRESIVOS Y DE MEDIAS MOVILES DE ORDEN SUPERIOR.

DEFINICION 2.1.3.1.: Una sucesión temporal $\{w_t\}_t$ obedece una ecuación lineal de diferencias de orden p, si se puede escribir en la forma

$$b_0(t)w_t - b_1(t)w_{t-1} - \ldots - b_p(t)w_{t-p} = c(t),\ b_p(t) \neq 0$$

donde $b_i(t)$ $i = 0,\ldots,p$ y $c(t)$ denotan funciones de t.

En particular si sustituímos los w_t por v.a.r Y_t $t \in Z$, tenemos la ecuación de diferencias estocásticas de orden p.

$$b_0(t)\ Y_t - b_1(t)\ Y_{t-1} - \ldots - b_p(t)\ Y_{t-p} = Z_t$$

donde $\{Z_t\}_t$ es un proceso discreto.

DEFINICION 2.1.3.2.: Decimos que el proceso Y_t es autorregresivo de orden p (AR(p)) si satisface la ecuación de diferencias estocástica de orden p

$$Y_t - b_1\ Y_{t-1} - \ldots - b_p\ Y_{t-p} = a_t$$

donde $b_p \neq 0$ y a_t es un ruído blanco.

Debemos notar que la ecuación de diferencias estocástica en la definición 2.1.3.2 es equivalente a

$$(1-b_1 B-\ldots-b_p B^p)\ Y_t = a_t.$$

Asociamos a esta ecuación la ecuación homogénea

$$1 - b_1 x-\ldots-b_p x^p = 0,$$

llamada ecuación característica. El estudio de sus raíces es tema central en la transformación de un proceso autorregresivo en lineal y consecuentemente en el estudio de su estacionaridad.

Por otro lado, es importante observar que una forma auxiliar de la ecuación característica es la ecuación

$$z^p - b_1\ z^{p-1} -\ldots- b_p = 0$$

donde $z = x^{-1}$ $(x \neq 0)$, y sus raíces son unas inversas de otras.

El Teorema que sigue nos pone de manifiesto como el tipo de soluciones de una ecuación lineal de diferencias de orden superior es función de las raíces de su ecuación característica asociada.

<u>LEMA 2.1.3.1.</u>: La ecuación lineal de diferencias de orden p

$$b_0(t)w_t - b_1(t)w_{t-1}-\ldots-b_p(t)w_{t-p} = c(t),\ b_o(t) \neq 0,\ b_p(t) \neq 0,$$

admite solución única w_t, en un conjunto S de valores enteros consecutivos de t, si conocemos valores de w_t en p valores consecutivos de t.

<u>DEMOSTRACION:</u> Supongamos que los valores conocidos de w_t son $w_0, w_1, \ldots, w_{p-1}$ y que $p \in S$.

Entonces dado t=p, de la ecuación de diferencias el valor de w_p queda unívocamente determinado por:

$$w_p = b_0(p)^{-1} \{ c(p) + b_1(p)w_{p-1} + \ldots + b_p(p)w_0 \}$$

Sea $t \in S$, tal que $t = p+1, p+2, \ldots$, tenemos el sistema

$$b_0(p+1)w_{p+1} - b_1(p+1)w_p - \ldots - b_p(p+1)w_1 = c(p+1)$$

$$b_0(p+2)w_{p+2} - b_1(p+2)w_{p+1} - \ldots - b_p(p+2)w_2 = c(p+2)$$

$$\ldots \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad \ldots$$

Dicho sistema genera las soluciones

$$w_t = b_o(t)^{-1} \{ c(t) + b_1(t)w_{t-1} + \ldots + b_p(t)w_{t-p} \} , \quad t = p+1, \ldots$$

las cuales son únicas al resolver inductivamente para $w_{p+1}, w_{p+2}, \ldots,$ conocidos unívocamente los valores $w_0, w_1, \ldots, w_p$.

<u>LEMA 2.1.3.2.:</u> $w_t = \alpha x_i^t$ es solución de la ecuación de diferencias homogénea

$$* \; b_0 w_t - b_1 w_{t-1} - \ldots - b_p w_{t-p} = 0 \qquad b_0 \neq 0, \quad b_p \neq 0,$$

siendo x_i cualquier raíz de

$$** \quad b_0 x^p - b_1 x^{p-1} - \ldots - b_p = 0$$

<u>DEMOSTRACION:</u> Es inmediato ya que

$$\alpha x_i^{t-p} \{ b_0 x_i^p - b_1 x_i^{p-1} - \ldots - b_p \} = 0$$

<u>TEOREMA 2.1.3.1.:</u> Las soluciones de la ecuación de diferencias (*) se obtienen empleando las raíces de (**); las soluciones son suma de p términos donde:

1) Por cada raíz real y distinta, x_i, un término del tipo bx_i^t es incluído.

2) Por cada raíz real con multiplicidad p_i, un término de la forma

$$(b_1^{(i)} + b_2^{(i)}t + b_3^{(i)}t^2 + \ldots + b_{p_i}^{(i)} t^{p_i - 1}) x_i^t = P_{p_i - 1}(t)\, x_i^t$$

es incluído

3) Por cada par de complejos conjugados no repetidos x_i y $x_{i+1} = \bar{x}_i$, por ejemplo, un término del tipo

$$a x_i^t + b\, x_{i+1}^t$$

se incluye, donde a y b son complejos conjugados.

4) Por cada par de complejos conjugados x_i y x_{i+1} con multiplicidad q, incluímos el término

$$(a_1 + a_2 t + \ldots + a_q t^{q-1}) x_i^t + (b_1 + b_2 t + \ldots + b_q t^{q-1}) x_{i+1}^t ,$$

donde $a_j = \bar{b}_j$ $\quad j = 1,2,\ldots,q$.

Dichas soluciones son únicas.

DEMOSTRACION:

1. Supóngase que (**) tiene p soluciones distintas

x_i, $i = 1,\ldots,p$

Por el Lema 2.1.3.2. $a_i x_i^t$ $\quad i = 1,2,\ldots,p$ es solución de (*) y en general también lo es la combinación lineal

$$w_t = \sum_{i=1}^{p} k_i x_i^t \qquad \text{con } k_i \neq 0 ,$$

puesto que (*) es homogénea.

Además, como (*) es un caso particular de

$$b_0(t) w_t - \ldots - b_p(t) w_{t-p} = c(t) \qquad \text{tomando}$$

$b_i(t) = b_i$ y $c(t) = 0$ para toda t, tenemos por el Lema 2.1.3.1 que w_t es única dados los valores $w_0, w_1, \ldots, w_{p-1}$.

2. Sean x_i $i=1,\dots,q$ $(q \leq p)$, q - raíces distintas de (**) con multiplicidades respectivas p_i $i=1,\dots,q$ tales que $\sum_{i=1}^{q} p_i = p$

Entonces

$$Q_{p-m}(B) \quad (B - x_i)^m = 0 \quad \text{con } m = p_i \quad \text{y} \quad B^k a(t) = a(t-k).$$

Además

$$R_{p-m}(F) \quad (F - x_i^{-1})^m = 0 \quad \text{donde} \quad F^k a(t) = a(t+k).$$

Obsérvese que si $x_i \neq 0$ es raíz de la ecuación polinomial

$$b_0 B^0 - b_1 B^1 - \dots - b_p B^p = 0$$

entonces x_i^{-1} es raíz de la ecuación

$$b_0 F^p - b_1 F^{p-1} - \dots - b_p F^0 = 0 \quad , \text{ pues } F = B^{-1}.$$

Así tenemos que como

$$(F - x_i^{-1}) \, P_m(t) \, (x_i^{-1})^t = P_{m-1}(t) \, (x_i^{-1})^{t+1},$$

donde $P_m(t)$ es un polinomio de grado m en t,

$$(F - x_i^{-1})^m \, P_{m-1}(t) \, (x_i^{-1})^t = 0$$

(ver [1] cap. 5 pág. 249).

$$(x_i - B)^m \, P_{m-1}(t) \quad x_i^{-m-t} = 0$$

lo cual equivale a

$$(B - x_i)^m \, P_{m-1}(t) \, x_i^t = 0, \quad y \quad w_t = P_{m-1}(t) \, x_i^t$$

es solución de(*), que equivale a $[Q_{p-m}(B) \; (B - x_i)^m] \, w_t = 0$

Luego

$$w_t = \sum_{i=1}^{q} P_{p_i-1}(t) \, x_i^t = \sum_{i=1}^{q} (b_1^{(i)} + b_2^{(i)} t + \ldots + b_{p_i}^{(i)} t^{p_i-1}) \, x_i^t$$

es solución de(*), siendo ésta única debido al Lema 2.1.3.1.

3. $ax_i^t + bx_{i+1}^t$ es solución de(*), si suponemos p raíces distintas entonces por el Lema 2.1.3.2. $w_t = \sum_{i=1}^{p} k_i x_i^t$ es la solución general donde $a = k_i$, $b = k_{i+1}$. Como w_t es real, entonces si el resto de las raíces son reales tenemos que $ax_i^t + bx_{i+1}^t$ es real de donde a y b son conjugados. Notemos que en caso de haber mas raíces complejas este criterio se mantiene en vista de la unicidad de la solución general.

4. Como en 2, $\tilde{P}_{q-1}(t) \, x_i^t$ y $\tilde{Q}_{q-1}(t) \, x_{i+1}^t$ son soluciones de (*) en donde, ahora, $\tilde{P}_{q-1}(t)$ y $\tilde{Q}_{q-1}(t)$ son polinomios de grado q-1 con coeficientes complejos. Entonces, como en 3, $\tilde{P}_{q-1}(t) \, x_i^t + \tilde{Q}_{q-1}(t) \, x_{i+1}^t$ es solución de (*) con $\tilde{Q}_{q-1}(t) = \overline{\tilde{P}}_{q-1}(t)$, es decir, $\tilde{Q}_{q-1}(t)$ y $\tilde{P}_{q-1}(t)$ son po-

linomios cuyos coeficientes correspondientes son complejos conjugados.

Así, por el Lema 2.1.3.2. tenemos la solución general

$$w_t = \sum_{i=1}^{2k} \tilde{P}_{q_i-1}(t)\, x_i^t + \sum_{i=2k+1}^{p} Q_{p_i-1}(t)\, x_i^t \quad \text{con } k \leq [P/2]$$

Siendo x_i, $i = 1,\dots,2k$ las raíces complejas de (**) tales que $x_j = \bar{x}_{j+1}$, $q_j = q_{j+1}$ y $\tilde{P}_{q_j-1}(t) = \bar{\tilde{P}}_{q_{j+1}-1}(t)$ para $j = 1,2,\dots,2k-1$.

<u>DEFINICION 2.1.3.3.</u>: El proceso $\{Y_t\}_t$ es un proceso de medias móviles de orden p (MA(p)), si para cada $t \in Z$

$$Y_t = a_t + \psi_1 a_{t-1} + \dots + \psi_p a_{t-p}, \quad \psi_p \neq 0$$

donde a_t es un ruído blanco y $\psi_i \in R$ para $i = 1,\dots,p$.

<u>OBSERVACIONES</u>:

* Todo proceso MA(p) es lineal y estacionario (p finito) según Teorema 2.1.2.4.

* Todo proceso lineal sobre un ruído puede considerarse como un MA(∞).

* En general $\sum_{j=0}^{p} \psi_j \neq 1$, por lo cual la nominación medias

móviles no es estricta. Sin embargo se adopta aquí debido a su uso frecuente en las obras clásicas sobre este tema.

2.1.4. TEOREMA DE CARACTERIZACION DE UN AR(p) ESTACIONARIO.

El Teorema que consideraremos en esta sección nos proveerá de las condiciones bajo las cuales un AR(p) puede considerarse como un MA(∞) estacionario.

Notemos que si X_t, satisface un modelo AR(p).

$$X_t = b_1 X_{t-1} + b_2 X_{t-2} + \ldots + b_p X_{t-p} + a_t$$

Si en esta ecuación de diferencias sustituímos consecutivamente

$$X_{t-i} = b_1 X_{t-i-1} + \ldots + b_p X_{t-i-p} + a_{t-1},$$

para cada $i = 1,2,\ldots$.

¿Qué condiciones nos permitirán asegurar que este procedimiento es convergente?

Es decir, bajo que condiciones

$$X_t = \sum_{j=0}^{\infty} w_j a_{t-j} \quad \text{con } w_0 = 1$$

TEOREMA 2.1.4.1.: Sea la ecuación polinomial

$$x^p - b_1 x^{p-1} - \ldots - b_p = 0,$$ cuyas raíces

x_i son tales que $|x_i| < 1$ con $b_p \neq 0$, y sea la sucesión de pesos $\{w_j\}_{j \in Z^+}$ definidas como solución de la ecuación homogénea de diferencias

$$w_j - b_1 w_{j-1} - \ldots - b_p w_{j-p} = 0 \quad , \quad j = p, \ p+1, \ldots$$

bajo las condiciones

$$w_0 = 1$$

$$w_1 = b_1 w_0$$

$$w_2 = b_1 w_1 + b_2 w_0$$

$$w_3 = b_1 w_2 + b_2 w_1 + b_3 w_0$$

....

$$w_{p-1} = b_1 w_{p-2} + b_2 w_{p-3} + \ldots + b_{p-1} w_0.$$

Sea además a_t un ruído. Entonces el límite en media cuadrática

$$X_t = \sum_{j=0}^{\infty} w_j a_{t-j}$$

es un proceso de medias móviles (infinito) estacionario. Más aún, X_t satisface la ecuación de diferencia estocástica

$$X_t - b_1 X_{t-1} - \ldots - b_p X_{t-p} = a_t$$

para casi toda realización de a_t.

DEMOSTRACION: Por el Teorema 2.1.3.1., si x_s con $s = 1,\ldots,d$

$$y \quad x_s = r_s e^{i\theta_s} \ , \quad x_{s+1} = r_s e^{-i\theta_s} \quad \text{con}$$

$s = d+1,\ldots, p-1$, denotan d raíces reales y p-d raíces complejas de la ecuación polinomial dada (ecuación característica); Y si

$$w_j = \sum_{s=1}^{p} c_s \, w_j^{(s)}$$

denota una solución de $w_j - \sum_{k=1}^{p} b_k \, w_{j-k} = 0$

para $j = p, p+1,\ldots,$ bajo las condiciones anteriores, entonces al acomodar convenientemente las expresiones en (2) y (4), resulta

$$w_j^{(s)} = (j)^{\ell} \, (x_s)^j \quad , \ s = 1,2,\ldots,d.$$

$$w_j^{(s)} = (j)^{\ell} \, (r_s)^j \cos j\,\theta_s, \ w_j^{(s+1)} = (j)^{\ell} \, (r_s)^j \operatorname{sen} j\,\theta_s$$

$s = d+1,\ldots, p-1$ y $0 \leq \ell \leq p$.

Notemos que cuando x_s es de multiplicidad $p_s \leq p$, podemos escribir

$$\left(\sum_{j=1}^{p_s} b_j t^{j-1} \right) x_s^t = \left(\sum_{j=1}^{p_s} b_j t^{j-1-\ell} \right) t^{\ell} x_s^t$$

en el caso real (ver (2) Teor. 2.1.3.1).

O bien

$$\left(\sum_{j=1}^{p_s} a_j t^{j-1}\right) x_s^t + \left(\sum_{j=1}^{p_s} b_j t^{j-1}\right) x_{s+1}^t =$$

$$\left(\sum_{j=1}^{p_s} (a_j+b_j) t^{j-1-\ell}\right) t^{\ell} r_s^t \cos t\theta_s - \left(\sum_{j=1}^{p_s} (a_j+b_j) t^{j-1-\ell}\right) t^{\ell} r_s^t \operatorname{sen} t\theta_s$$

en el caso complejo (ver (4) Teor. 2.1.3.1.).

Sea ahora λ tal que $1 > \lambda > \max_s \{|x_s|\}$

entonces existe un $M \geq 0$ tal que

$$(j)^{\ell} (|x_s|^j) < M \lambda^j \quad \text{para todo } j$$

(ver [7], pág. 91. ejerc. 24).

de donde

$$\sum_{j=0}^{\infty} |c_s (j)^{\ell} (x_s)^j| < |c_s| M \sum_{j=0}^{\infty} \lambda^j < \infty$$

y

$$\sum_{j=0} |w_j| < \infty$$

Por el Teorema 2.1.2.1. existe un proceso $\{X_t\}_{\tau}$ tal que

$$X_t = \sum_{j=0}^{\infty} w_j a_{t-j}$$

Además como w_j es absolutamente sumable se tiene del Teorema 2.1.2.2.

. $E\{X_t\} = 0$

. $\gamma_X(t,t+h) = E\{(\sum_{j=0}^{\infty} w_j a_{t-j})(\sum_{k=0}^{\infty} w_k a_{t+h-k})\} =$

$$= \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} w_j w_k E\{a_{t-j} a_{t+h-k}\} = \sum_{k=h}^{\infty} w_k w_{k-h} \sigma^2 = \gamma_x(h),$$

$h = 0,1,2,...$

De manera que X_t es estacionario.

Finalmente consideremos el proceso:

$$Z_t = b_0X_t - b_1 X_{t-1} -\dots - b_p X_{t-p} - a_t, \ b_0 = 1$$

donde

$$b_0X_t - b_1 X_{t-1} -\dots- b_p X_{t-p}=a_t(b_0w_0)+a_{t-1}(b_0 w_1 -b_1 w_0) +$$

$$+ a_{t-2}(b_0w_2-b_1w_1-b_2w_0)+\dots+a_{t-p+1}(b_0w_{p-1}-b_1w_{p-2}-\dots-b_{p-1}w_0) +$$

$$+ \sum_{j=p}^{\infty}(b_0w_j-b_1w_{j-1}-\dots-b_jw_{j-p})a_{t-j} = (b_0w_0)\ a_t +$$

$$+ \sum_{k=1}^{p-1}(b_0w_k- \sum_{j=1}^{k} b_jw_{k-j})a_{t-k} + \sum_{k=p}^{\infty} (b_0w_k - \sum_{j=1}^{k} b_jw_{k-j})a_{t-k}=$$

$= a_t$ (en m.c).

Asi tenemos que

$$E\{|Z_t|\}^2 = E\{|b_0X_t - b_1X_{t-1} - \ldots - b_p X_{t-p} - a_t|\}^2$$

$$= \lim_{L\to\infty} E\{| \sum_{k=1}^{p-1} (b_0w_k - \sum_{j=1}^{k} b_j w_{k-j})\, a_{t-k}$$

$$+ \sum_{k=p}^{L} (b_0w_k - \sum_{j=1}^{k} b_j w_{k-j})\, a_{t-k} |^2\}$$

$$= 0,$$

bajo las condiciones iniciales dadas sobre la ecuación homogénea de diferencias.

Entonces $Z_t = 0$ (casi seguramente), o sea

$b_0X_t - b_1X_{t-1} - \ldots - b_pX_{t-p} = a_t$ excepto en el conjunto

$A_t = [Z_t \neq 0]$ tal que $P(A_t) = 0$ para cada t, de allí

que $P\{\bigcup_t A_t\} = 0.$

<u>COROLARIO 2.1.4.1.:</u> Sea X_t un AR(p) con ecuación característica $x^p - \sum_{j=1}^{p} b_j x^{p-j} = 0$ $(b_p \neq 0)$,

cuyas raíces son en norma menores que 1.

Entonces X_t es independiente de $a_{t+1}, a_{t+2}, \ldots$ para todo t.

DEMOSTRACION: Sabemos que $|x_i| < 1$ $i = 1,\ldots,p$, donde x_i raíz de $x^p - \sum_{j=1}^{p} b_j x^{p-j} = 0$. Entonces las raíces z_i de $1 - \sum_{j=1}^{k-p} b_j Z^j = 0$ son tales que $z_i = x_i^{-1}$ y $|z_i| > 1$.

Sea z tal que $|z| < \min_i \{|z_i|\}$ entonces

$$\frac{1}{1-\sum_{j=1}^{p} b_j z^j} = \frac{1}{\prod_{i=1}^{p} (1-\frac{z}{z_i})} = \prod_{i=1}^{p} \sum_{k=0}^{\infty} (\frac{z}{z_i})^k = \sum_{i=0}^{\infty} w_i z^i$$

pues $|\frac{z}{z_i}| < 1$ $i = 1,\ldots,p$.

Así entonces $(1-\sum_{j=1}^{p} b_j z^j)(\sum_{i=0}^{\infty} w_i z^i) = \sum_{i=0}^{\infty} (w_i z^i - \sum_{j=1}^{p} b_j w_i z^{i+j}) =$

$$= w_0 + \sum_{k=1}^{p-1} (w_k - \sum_{j=1}^{k} b_j w_{k-j}) z^k + \sum_{k=p}^{\infty} (w_k - \sum_{j=1}^{p} b_j w_{k-j}) z^k = 1$$

De donde obtenemos el sistema de p ecuaciones (consistente para w_i $i = 0,\ldots,p-1$)

$$w_0 = 1$$

$$w_1 - b_1 w_0 = 0$$

....

$$w_{p-1} - b_1 w_{p-2} - \ldots - b_{p-1} w_0 = 0$$

el cual sirve de condicionante en la obtención de la solución de

$$w_t - b_1 w_{t-1} - \ldots - b_p w_{t-p} = 0 \;, \quad \text{para } t \geqslant p.$$

Por el Teorema 2.1.4.1 concluímos que

$$X_t = \sum_{j=0}^{\infty} w_j \, a_{t-j} \quad (X_t \text{ es lineal sobre } a_t)$$

lo que implica que X_t es independiente de a_s $s \geqslant t+1$.

3.1 PREDICCION LINEAL CON PARAMETROS CONOCIDOS.

En esta sección abordaremos el problema de obtener una previsión o predicción de los valores seriales

$$X_{n+1},\ X_{n+2},\ldots,\ X_{n+s},\ldots$$

a partir de los valores anteriores conocidos

$$X_n,\ X_{n-1},\ldots,\ X_1,\ldots$$

Esto lo haremos mediante el uso de funciones lineales del tipo

$$\hat{X}_{n+s}(\ldots,X_1,\ldots,X_n) = \sum_{j=0}^{h} c_j^{(s)} X_{n-j} \quad , \quad s = 1,2,\ldots$$

las cuales llamaremos predictores lineales de los X_{n+s}; además nos valdremos del error cuadrático medio de predicción (ECM) como único criterio para medir la 'bondad' de dichos estadísticos. Así diremos que $\hat{X}_{n+s}(\ldots,X_1,\ldots,X_n)$ es el mejor predictor lineal de X_{n+s}, si

$$\text{ECM}\left\{\hat{X}_{n+s}(\ldots,X_1,\ldots,X_n)\right\} = E\left\{[X_{n+s} - \hat{X}_{n+s}(\ldots,X_1,\ldots X_n)]^2\right\}$$

es mínimo dentro de la categoría de los predictores lineales de X_{n+s}.

Luego el problema consiste en la escogencia de las constantes $c_j^{(s)}$ de suerte que ECM sea mínimo.

Para la obtención de una función con las características apuntadas anteriormente, haremos las siguientes suposiciones en torno al proceso considerado.

1. $X_t = b_1 X_{t-1} + b_2 X_{t-2} + \ldots + b_p X_{t-p} + a_t$, $t = 0, \pm 1, \pm 2, \ldots$

   donde b_i, $i=1,\ldots,p$ y $\operatorname{var}\{a_t\} = \sigma^2$

   son conocidos,

2. X_t satisface las hipótesis del teorema 2.1.4.1, es decir, X_t es un AR(p) estacionario (lineal).

3. a_t es un ruído blanco independiente con $E\{a_t\} = 0$

De lo anterior escribimos

$$X_{n+s} = \sum_{j=0}^{\infty} w_j \, a_{n+s-j} \ ,$$

entonces si suponemos que $X^*_{n+s}(\ldots,X_1,\ldots,X_n) =$

$$= \sum_{j=s}^{\infty} w_j^* \, a_{n+s-j}$$ es el mejor predictor lineal

de X_{n+s}; de $\text{ECM}\{X^*_{n+s}(\ldots,X_1,\ldots,X_n)\} = \sum_{j=0}^{s-1} w_j^2 \sigma^2 +$

$+ \sum_{j=s}^{\infty} (w_j - w_j^*)^2 \sigma^2$, se deduce q' $w_j^* = w_j$ $j \geqslant s$, por lo que ECM$\{X^*_{n+s}(\ldots,X_1,\ldots,X_n)\} = (1+w_1^2+\ldots+w_{s-1}^2)\sigma^2$

y $X^*_{n+s}(\ldots,X_1,\ldots,X_n) = \sum_{j=s}^{\infty} w_j\, a_{n+s-j}$

En particular si s=1, ECM $\{X^*_{n+1}(\ldots,X_1,\ldots,X_n)\} = \sigma^2$

además, puesto que a_{n+s} es independiente de a_{n+s-j} y X_{n+s-j} para $j \geqslant 1$, se tiene

$$E\{a_{n+1}/ X_n,\ldots,X_1,\ldots\} = E\{a_{n+1}\} = 0 \quad ,$$

por lo que

$$E\{X_{n+1}/X_n,\ldots,X_1,\ldots\} = \sum_{j=1}^{p} b_j\, X_{n+1-j} = \hat{X}_{n+1}(\ldots,X_1,\ldots X_n).$$

$$\text{ECM}\{X^*_{n+1}(\ldots,X_1,\ldots X_n)\} = \sigma^2 + E\{[X^*_{n+1}(\ldots,X_1,\ldots,X_n) - \hat{X}_{n+1}(\ldots,X_1,\ldots,X_n)]^2\}$$

luego $\hat{X}_{n+1}(\ldots,X_1,\ldots,X_n)$ es el mejor predictor lineal de X_{n+1}.

Si s = 2 escribimos

$$X_{n+2} = b_1 \sum_{j=1}^{p} b_j\, X_{n+1-j} + \sum_{j=2}^{p} b_j\, X_{n+2-j} + b_1\, a_{n+1} + a_{n+2}.$$

$$= b_1 \hat{X}_{n+1}(\ldots,X_1,\ldots,X_n) + \sum_{j=2}^{p} b_j X_{n+2-j} + b_1\, a_{n+1} + a_{n+2}.$$

De donde

$$\hat{X}_{n+2}(\ldots,X_1,\ldots,X_n) = b_1 \hat{X}_{n+1}(\ldots,X_1,\ldots,X_n) + \sum_{j=2}^{p} b_j X_{n+2-j}$$

$$= E\{X_{n+2}/x_n,\ldots,X_1,\ldots\} \quad y$$

$$ECM\{\hat{X}_{n+2}(\ldots,X_1,\ldots,X_n)\} = (1+w_1^2)\sigma^2 \quad \text{es mínimo.}$$

Generalizando, se obtiene que

$$E\{X_{n+s}/X_n,\ldots,X_1,\ldots\} = \hat{X}_{n+s}(\ldots,X_1,\ldots,X_n).$$

$$= \sum_{j=1}^{s-1} b_j \hat{X}_{n+s-j} + \sum_{j=s}^{p} b_j X_{n+s-j} \quad , s \leq p$$

$$= \sum_{j=1}^{p} b_j \hat{X}_{n+s-j} \quad , \quad s > p$$

es el mejor predictor lineal de X_{n+s}.

3.1.1. NOTACION

Dados los valores $\ldots,X_1,\ldots,X_n$ de una serie temporal X_t, notamos con:

* $\hat{X}_{n+s}$ $s > o$, al predictor lineal de x_{n+s} en vez de $\hat{X}_{n+s}(\ldots,X_1,\ldots,X_n)$ cuando los datos sean obvios.

* $E_n\{X_{n+s}\}$, a la esperanza condicional de X_{n+s} en el instante n, en lugar de $E\{X_{n+s}/X_n,\ldots,X_1,\ldots\}$; y con $var_n\{X_{n+s}\}$, a la varianza condicional; en lugar de $var\{X_{n+s}/X_n,\ldots,X_1,\ldots\}$

* e_{n+s} , al error de predicción $X_{n+s}-\hat{X}_{n+s}$, esto es, $e_{n+s} = X_{n+s}-\hat{X}_{n+s}$. Por el Teorema 2.1.4.1,

$$\hat{X}_{n+s} = \sum_{j=s}^{\infty} w_j\, a_{n+s-j} \quad \text{y} \quad e_{n+s} = \sum_{j=0}^{s-1} w_j\, a_{n+s-j}$$

3.1.2. PROPIEDADES

1. $E_n\{e_{n+s}\} = 0$, luego $\hat{X}_{n+s} = E_n\{X_{n+s}\}$ es el mejor estimador lineal e insesgado de X_{n+s}

2. $\hat{X}_{n+s}$ es solución de la ecuación de diferencias

   $w_t - b_1 w_{t-1} -\ldots- b_p w_{t-p} = 0$,

   bajo las condiciones iniciales: $w_n = X_n$, $w_{n-1} = X_{n-1},\ldots, w_1 = X_1,\ldots$

3. Si las raíces x_i (supongámoslas distintas) de $x^p - b_1 x^{p-1} -\ldots-b_p = 0$, son tales que $|x_i| < 1$.

Por el Teorema 2.1.3.1 la solución general de la ecuación de diferencias es del tipo:

$$w_\theta = \sum_{i=1}^{p} k_i \, x_i^{\theta} \quad , \quad \theta > n$$

Entonces $\hat{X}_\theta \longrightarrow 0$ si $\theta \rightarrow \infty$ (ver [10]).

4. X_{n+s}, tiende a no tomar en cuenta la información contenida en los datos cuando $s \rightarrow \infty$. Esto es así pues se demuestra que $E_n \{ X_{n+s} \} \underset{s \rightarrow \infty}{\longrightarrow} E \{ X_{n+s} \}$

Asimismo de $\hat{X}_{n+s} = w_s \, a_{n+s} + \ldots + w_t \, a_{n+s-t} + \ldots$

donde $w_t \longrightarrow 0$ si $t \rightarrow \infty$, podemos afirmar que las épocas alejadas de n no intervienen en la previsión mas que con un peso débil. De aquí que en la práctica sólo se requiera una serie finita, digamos $X_n, \ldots, X_1$, para pronosticar X_{n+s}.

5. Si a_t es un proceso puramente aleatorio tal que $a_t \sim N(0, \sigma^2)$, se tiene que

$$\hat{X}_{n+s} - z_\alpha \sigma \sqrt{\sum_{j=0}^{s-1} w_j^2} \quad ; \quad \hat{X}_{n+s} + z_\alpha \sigma \sqrt{\sum_{j=0}^{s-1} w_j^2}$$

es un intervalo de confianza para X_{n+s}

donde $P\{-z_{\alpha} < Z < z_{\alpha}\} = \alpha$ y $Z = \dfrac{e_{n+s}}{\sqrt{E\{e^2_{n+s}\}}} \sim N(0,1)$

3.2. PREDICCION LINEAL CON PARAMETROS DESCONOCIDOS.

Al estimar los parámetros de un modelo AR(p) nos enfrentaremos al problema de aproximación, es decir, obtendremos estimadores que tendrán propiedades óptimas como: insesgamiento, eficiencia y consistencia, sólo asintóticamente. En consecuencia, los predictores lineales aquí presentados serán óptimos desde un punto vista asintótico.

Sin embargo debemos destacar que la noción de aproximación a la que nos referimos, es probabilística, por ello se introduce en la sección que sigue el concepto de orden de magnitud en probabilidad, el cual se debe a H.B. Mann y A. Wald (1943).

3.2.1. ORDEN DE MAGNITUD Y ORDEN DE MAGNITUD EN PROBABILIDAD.

Sea $\{b_n\}_n$ una sucesión de reales y $\{c_n\}_n$ una sucesión de reales positivos.

<u>DEFINICION 3.2.1.1.:</u> Decimos que b_n tiene menor orden de magnitud que c_n y lo escribimos

$$b_n = o\,(c_n)$$

si

$$\lim_{n\to\infty} \frac{b_n}{c_n} = 0$$

<u>DEFINICION 3.2.1.2:</u> Decimos que b_n es a lo sumo de orden c_n y escribimos

$$b_n = O\,(c_n)$$

si existe un real M tal que $c_n^{-1} \mid b_n \mid \leq M$ para todo n.

<u>DEFINICION 3.2.1.3.:</u> Una sucesión X_n de v.a.r es de menor orden, en probabilidad, que c_n y escribimos

$$X_n = o_p\,(c_n)$$

si

$$\frac{X_n}{c_n} \xrightarrow{P_r} 0 \quad \text{cuando} \quad n \to \infty.$$

DEFINICION 3.1.2.4: Una sucesión X_n de v.a.r. es a lo sumo de orden c_n, en probabilidad, y lo escribimos

$$X_n = O_p(c_n)$$

si, para cualquier $\epsilon > 0$, existe un real positivo M_ϵ tal que

$$P\{ | X_n | \geq M_\epsilon \, c_n \} \leq \epsilon$$

para todo n.

DEFINICION 3.1.2.5.: Si $X_n = (X_{jn})_{j=1,\ldots,k}$ es un vector aleatorio k-dimensional, X_n es a lo sumo de orden c_n en probabilidad y escribimos

$$X_n = O_p(c_n)$$

si,

$$X_{jn} = O_p(c_n) \quad j=1,\ldots,k.$$

Diremos que X_n es de menor orden que c_n en probabilidad y escribimos

$$X_n = o_p(c_n)$$

si

$$X_{jn} = o_p(c_n) \quad j=1,\ldots,k.$$

<u>TEOREMA 3.2.1.1</u>: Sean $\{b_n\}_n$ y $\{c_n\}_n$ dos sucesiones de reales positivos, y sean $\{X_n\}_n$ y $\{Y_n\}_n$ dos sucesiones de variables aleatorias.

(i) Si $X_n = o_p(b_n)$ y $Y_n = o_p(c_n)$, entonces

. $X_n Y_n = o_p(b_n c_n)$,

. $|X_n|^s = o_p(b_n^s)$ $\quad s > o$

. $X_n + Y_n = o_p(\max\{b_n, c_n\})$.

(ii) Si $X_n = O_p(b_n)$ y $Y_n = O_p(c_n)$, entonces

. $X_n Y_n = O_p(b_n c_n)$

. $|X_n|^s = O_p(b_n^s)$ $\quad s > o$,

. $X_n + Y_n = O_p(\max\{b_n, c_n\})$.

(iii) Si $X_n = o_p(b_n)$ y $Y_n = O_p(c_n)$, entonces

. $X_n Y_n = o_p(b_n c_n)$

<u>DEMOSTRACION:</u> Demostraremos sólo (ii), dejando (i) y (iii) como ejercicio al lector.

. Sea $\epsilon = \frac{1}{2}\delta > o$, existen M_ϵ y N_ϵ , reales positivos, tales que

$$P\{|X_n| \geq M_\epsilon\, b_n\} \leq \epsilon \quad y \quad P\{|Y_n| \geq N_\epsilon\, c_n\} \leq \epsilon \qquad \forall n.$$

Entonces como $\left| \dfrac{X_n}{M_\epsilon\, b_n} \quad \dfrac{Y_n}{N_\epsilon\, c_n} \right| > 1$ implica

$$\left|\frac{X_n}{M_\epsilon\, b_n}\right| > 1 \quad o \quad \left|\frac{Y_n}{N_\epsilon\, c_n}\right| > 1 \text{ (o ambos)}$$

Tenemos

$$P\{|X_nY_n| \geq M_\epsilon \cdot N_\epsilon \cdot b_nc_n\} \leq P\{|X_n| \geq M_\epsilon\, b_n\} + P\{|Y_n| \geq N_\epsilon\, c_n\}$$

$$\leq \tfrac{1}{2}\delta + \tfrac{1}{2}\delta = \delta$$

el resultado se obtiene al tomar $M_\epsilon \cdot N_\epsilon = K_\delta$.

. El segundo resultado se obtiene del hecho de que dado $\epsilon > 0$ arbitrario y $M_\epsilon > 0$

$$P\{|X_n| \geq M_\epsilon \cdot b_n\} = P\{|X_n|^s \geq M_\epsilon^s \cdot b_n^s\} \leq \epsilon$$

para todo n.

. Finalmente, sea $\epsilon = \tfrac{1}{2}\delta > 0$. Entonces si

$q_n = \max\{b_n, c_n.\}$

$$P\{|X_n| \geq M_\epsilon\, q_n\} \leq \tfrac{1}{2}\delta$$

$$P\{|Y_n| \geq N_\epsilon\, q_n\} \leq \tfrac{1}{2}\delta \text{ , para todo n.}$$

Además

$$P\{|X_n| \geq M_\epsilon \cdot q_n\} + P\{|Y_n| \geq N_\epsilon q_n\} \leq \delta.$$

Si excogemos $L \geq \max\{M_\epsilon , N_\epsilon\}$, tenemos

$$P\{|X_n| \geq M_\epsilon \cdot q_n\} \geq P\{|X_n| \geq L_\epsilon \cdot q_n\}$$

$$P\{|Y_n| \geq N_\epsilon \cdot q_n\} \geq P\{|Y_n| \geq L_\epsilon \cdot q_n\}$$

de donde

$$P\{|X_n| \geq M_\epsilon \cdot q_n\} + P\{|Y_n| \geq N_\epsilon \cdot q_n\} \geq P\{|X_n| \geq L_\epsilon \cdot q_n\} +$$

$$+ P\{|Y_n| \geq L_\epsilon \cdot q_n\} \geq P\left\{\left|\frac{X_n}{q_n}\right| + \left|\frac{Y_n}{q_n}\right| \geq L_\epsilon\right\} \geq P\left\{\left|\frac{X_n+Y_n}{q_n}\right| \geq L_\epsilon\right\} =$$

$$= P\{|X_n + Y_n| \geq L_\epsilon \cdot q_n\}$$

Luego

$P\{|X_n + Y_n| \geq L_\delta \cdot q_n\} \leq \delta$, de lo cual obtenemos nuestro resultado.

OBSERVACIONES

* Se ha demostrado (Mann y Wald (1943)) que el álgebra de O y o es equivalente al álgebra de Op y op. En consecuencia, existe una versión similar del Teorema 3.2.1.1 para O y o (ver [7], pág. 184).

* Por el Teorema de Markov, si X es una v.a.r tal que. $E\{X^r\} < \infty$ para algún $r > o$.

  $X = Op(1)$

* Toda variable aleatoria del tipo op también es del tipo Op. Lo contrario no es necesariamente válido.

* $X_n \xrightarrow{P_r} X$ si y sólo sí $X_n - X = op(1)$, de aquí que $\hat{\theta}$ es estimador consistente de θ si y sólo sí $\hat{\theta}-\theta=o_p(1)$.

* Dadas $X_n = Op(b_n)$ y $k_n = O(c_n)$, donde b_n y c_n son reales positivos para cada n en N, entonces

$$X_n k_n = Op\ (b_n c_n)$$

  Además se verifica fácilmente que si

  $x_n = o_p(b_n)$ y $k_n = O(c_n)$ se tendrá

  $X_n k_n = o_p(b_n c_n)$.

<u>TEOREMA 3.2.1.2.</u>: Sea $\{X_n\}_n$ una sucesión de variables aleatorias y $\{a_n\}_n$ una sucesión de reales positivos tal que

$$E\{X_n^2\} = O(a_n^2)$$

Entonces

$$X_n = Op(a_n)$$

<u>DEMOSTRACION:</u> Por hipótesis existe un real $M_1 > o$ tal que

$$E\{X_n^2\} \leqslant M_1^2 a_n^2 \quad \text{para todo } n.$$

Por la desigualdad de Chebyshev, para cualquier $M_2 > o$,

$$P\{|X_n| \geqslant M_2 a_n\} \leqslant \frac{E\{X_n\}^2}{M_2^2 a_n^2}$$

de donde

$$P\{|X_n| \geqslant M_2 a_n\} \leqslant \frac{M_1^2}{M_2^2}$$

Entonces, dado $\epsilon > o$, si escogemos $M_2 \geqslant M_1 \cdot \epsilon^{-\frac{1}{2}}$

obtenemos

$$P\{|X_n| \geqslant M_2 a_n\} \leqslant \epsilon$$

con lo que se verifica el resultado.

<u>COROLARIO 3.2.1.1.:</u> Sea $\{X_n\}_n$ una sucesión de variables aleatorias tales que

$$E\{(X_n - E\{X_n\})^2\} = O(a_n^2)$$

y

$$E\{X_n\} = O(a_n)$$

siendo $\{a_n\}_n$ una sucesión de reales positivos.

Entonces,

$$X_n = O_p(a_n)$$

<u>DEMOSTRACION</u>:

$$E\{X_n^2\} = E\{(X_n - E\{X_n\})^2\} + (E\{X_n\})^2 = O(a_n^2),$$

el resultado se sigue del Teorema 3.2.1.2.

3.2.2. AUTORREGRESION EN MUESTRAS GRANDES.

Sea Y_t un proceso satisfaciendo el modelo AR(p) general

$$(Y_t - \mu) = \theta_1 (Y_{t-1} - \mu) + \ldots + \theta_p (Y_{t-p} - \mu) + a_t,$$

el cual puede reescribirse como

$$Y_t = \Theta_0 + \Theta_1 Y_{t-1} + \ldots + \Theta_p Y_{t-p} + a_t \quad t=p+1,\ldots,n,$$

Tomando $\Theta_0 = \mu(1-\Theta_1-\ldots-\Theta_p)$ con $E\{Y_t\} = \mu$

para todo t.

Podemos, inclusive, reescribir matricialmente este último modelo como

$$\mathbf{Y} = \mathbf{X}\,\boldsymbol{\Theta} + \mathbf{a}$$

donde

$$\mathbf{Y} = \begin{bmatrix} Y_{p+1} \\ \vdots \\ Y_n \end{bmatrix}_{(n-p)\times 1}, \quad \boldsymbol{\Theta} = \begin{bmatrix} \Theta_0 \\ \vdots \\ \Theta_p \end{bmatrix}_{(p+1)\times 1}, \quad \mathbf{a} = \begin{bmatrix} a_{p+1} \\ \vdots \\ a_n \end{bmatrix}_{(n-p)\times 1}$$

corresponden a los vectores de las observaciones, los parámetros (desconocidos), y los errores (ruído) respectivamente.
Y

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_{p+1} \\ \vdots \\ \mathbf{x}_n \end{bmatrix} = \begin{bmatrix} 1 & Y_p & Y_{p-1} & \cdots & Y_1 \\ 1 & Y_{p+1} & Y_p & \cdots & Y_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & Y_{n-1} & Y_{n-2} & \cdots & Y_{n-p} \end{bmatrix}_{(n-p)\times(p+1)}$$

donde $\mathbf{x}_t = [1, Y_{t-1}, \ldots, Y_{t-p}] \quad t = p+1,\ldots,n$

Supongamos que se desea estimar los parámetros del modelo, bajo la forma matricial, usando las técnicas de estimación de un modelo de regresión usual con las siguientes condiciones

1. $E\{\mathbf{a}\} = \mathbf{0}$, $\operatorname{var}\{\mathbf{a}\} = \sigma^2 \mathbf{I}$

2. $Y_p, Y_{p-1}, \ldots, Y_1$ son valores constantes o dados.

3. $\tilde{Y}_t = Y_t - \mu$ satisface las hipótesis del Teorema 2.1.4.1. siendo a_t un ruído blanco independiente.

Así entonces de

$$\sum_{t=p+1}^{n} \mathbf{x}_t Y_t = \sum_{t=p+1}^{n} \mathbf{x}_t' \mathbf{x}_t \boldsymbol{\theta} + \sum_{t=p+1}^{n} \mathbf{x}_t' a_t$$

se obtiene el estimador de mínimos cuadrados

$$(*) \quad \hat{\boldsymbol{\theta}} = \mathbf{A}_n^{-1} \mathbf{v}_n \qquad \text{con} \qquad \mathbf{A}_n = \frac{1}{n-p} \sum_{t=p+1}^{n} \mathbf{x}_t' \mathbf{x}_t$$

$$\text{y} \qquad \mathbf{v}_n = \frac{1}{n-p} \sum_{t=p+1}^{n} \mathbf{x}_t' Y_t$$

Sin embargo, dado que $\mathbf{X}$ es una matríz aleatoria, el Teorema de Gauss-Markov no es válido y $\hat{\boldsymbol{\theta}}$ no sería un estimador óptimo de $\boldsymbol{\theta}$. Observemos además que otro de los factores que impide aplicar los resultados de la Teoría de mínimos cuadrados ordinarios es el hecho de que a_t y Y_{t-j} no son independientes si $j \leq 0$.

Durbin (1959) pudo comprobar que las propiedades de insesgamiento, eficacia y consistencia de $\hat{\theta}$ sólo son válidas en muestras grandes.

LEMA 3.2.2.1. Sea $\{X_t\}_t$ el proceso definido por

$$X_t = \sum_{j=-\infty}^{\infty} \alpha_j \, a_{t-j}$$

donde $\{\alpha_j\}_j$ es absolutamente sumable, y las a_t forman un ruído blanco independiente tal que

$E\{a_t\} = 0$, $\operatorname{var}\{a_t\} = \sigma^2$ y $E\{a_t^4\} = \sigma^4$

Entonces para $h \geqslant q \geqslant 0$ fijos y $\tilde{\gamma}(k) = \frac{1}{n-k} \sum_{t=1}^{n-k} X_t X_{t+k}$

$$\lim_{n \to \infty} (n-q) \operatorname{cov}\{\tilde{\gamma}(h), \tilde{\gamma}(q)\} = (\eta-3)\gamma(h)\,\gamma(q) +$$

$$+ \sum_{p=-\infty}^{\infty} \left[\gamma(p)\gamma(p-h+q) + \gamma(p+q)\,\gamma(p-h) \right].$$

DEFINICION 3.2.2.1: Diremos que el proceso $\{X_t\}_t$ es m-dependiente si $s-r > m$, donde m es un entero positivo, implica que los conjuntos

$$\ldots, X_{r-2}, X_{r-1}, X_r \quad \text{y} \quad X_s, X_{s+1}, X_{s+2}, \ldots$$

son independientes.

LEMA 3.2.2.2.: Sea $\{X_t\}_t$ una sucesión aleatoria m- dependiente con $E\{X_t\} = 0$, $E\{X_t^2\} = \sigma_t^2 < \infty$,

y $E\{|X_t|^3\} \leq \beta^3$. Tal que

$$\lim_{p \to \infty} p^{-1} \sum_{j=1}^{p} s_{t+j} = s \neq 0 \quad \text{para todo} \quad t = 0, 1, 2, \ldots$$

donde

$$s_t = E\{X_{t+m}^2\} + 2 \sum_{j=1}^{m} E\{X_{t+m-j} X_{t+m}\}$$

Entonces

$$n^{-\frac{1}{2}} \sum_{t=1}^{n} X_t \xrightarrow{\text{Ley}} N(0,s), \quad n \to \infty$$

LEMA 3.2.2.3.: Sea $\{\xi_n\}_n$ una sucesión de variables aleatorias asociada a la sucesión de funciones de distribución $\{F_{\xi_n}(z)\}_n$, y definida por

$$\xi_n = S_{kn} + D_{kn} \quad \text{para} \quad k = 1,2,\ldots \quad y \quad n = 1,2,\ldots$$

Sea $D_{kn} \xrightarrow[k \to \infty]{p_r} 0$ para todo n, y

$$F_{S_{kn}}(z) \xrightarrow{\text{Ley}} F_{\psi_k}(z) \qquad n \to \infty$$

$$F_{\psi_k}(z) \xrightarrow{\text{Ley}} F_{\xi}(z) \qquad k \to \infty$$

Entonces,

$$F_{\xi_n}(z) \xrightarrow{\text{Ley}} F_{\xi}(z) \qquad n \to \infty$$

NOTA: Para la demostración de los Lemas 3.2.2.1., 3.2.2.2. y 3.2.2.3 ver [7] cap. 6.

Si Y_t satisface el modelo AR(P)

$$(Y_t - \mu) = \theta_1 (Y_{t-1} - \mu) + \ldots + \theta_p (Y_{t-p} - \mu) + a_t$$

con $\quad E\{Y_t\} = \mu$

Cuando $x^p - \sum_{i=1}^{p} \theta_i x^{p-i} = 0$, tiene todas sus raíces, en norma, menores que 1 por el Teorema 2.1.4.1

$$Y_t = \mu + \sum_{j=0}^{\infty} \theta_j a_{t-j}$$

Si además pedimos que $E\{a_t^4\} = \eta\sigma^4$; del Lema 3.2.2.1 tenemos que al tomar $h = q$

$$\lim_{n \to \infty} (n-h) \operatorname{var}\{ \tilde{\tilde{\gamma}}_{\tilde{Y}}(h) \} = (\eta-3)\, \gamma^2(h) + \sum_{k=-\infty}^{\infty} [\gamma^2(k) + \gamma(k+h)\, \gamma(k-h)] < \infty$$

Luego como $E\{\tilde{\gamma}_{\tilde{Y}}(h)\} = \gamma(h)$

$$\lim_{n \to \infty} E\{(\tilde{\gamma}_{\tilde{Y}}(h) - \gamma(h))^2\} = 0, \quad \text{0 sea} \quad \tilde{\gamma}_{\tilde{Y}}(h) \xrightarrow{p_r} \gamma(h)$$

debido al Teorema 1.1.2.1.

Entonces $\mathbf{A}_n \xrightarrow{p_r} \mathbf{A}$, donde si a_{ij} denota el ij-ésimo elemento de $\mathbf{A}$

$$(**) \quad a_{ij} = \begin{cases} \gamma(|i-j|) + \mu^2, & i,j = 2,3,\ldots,p+1 \\ 1, & i=j=1 \\ \mu, & \text{otros.} \end{cases}$$

<u>TEOREMA 3.2.2.1:</u> Sea $\{Y_t\}_t$ un proceso AR(p) con representación

$$Y_t = \theta_o + \sum_{i=1}^{p} \theta_i Y_{t-i} + a_t$$

donde las raíces de

$$x^p - \sum_{i=1}^{p} \theta_i x^{p-i} = 0$$

son menores que uno (en norma) y los a_t forman un ruído blanco independiente tal que

$$E\{a_t\} = 0, \quad E\{a_t^2\} = \sigma^2 \quad y \quad E\{a_t^4\} = \eta\sigma^4 .$$

Entonces

$$n^{\frac{1}{2}} (\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) \xrightarrow{\text{Ley}} N(\mathbf{0},\mathbf{A}^{-1}\sigma^2).$$

donde $\hat{\boldsymbol{\theta}}$ se define como en (*) y $\mathbf{A}_n \xrightarrow{P_r} \mathbf{A}$.

<u>DEMOSTRACION:</u> Y_t satisface las hipótesis del Lema 3.2.2.1, luego $\mathbf{A}_n \xrightarrow{P_r} \mathbf{A}$ con $\mathbf{A}$ una matríz $(p+1)\mathrm{x}(p+1)$ definida como en (**).

Del modelo de regresión dado al inicio de esta sección, tenemos

$$\boldsymbol{\theta} = \Big(\sum_{t=p+1}^{n} \mathbf{X}_t'\mathbf{X}_t\Big)^{-1} \Big(\sum_{t=p+1}^{n} (Y_t-a_t)\mathbf{X}_t'\Big) = \Big(\sum_{t=p+1}^{n} \mathbf{X}_t'\mathbf{X}_t\Big)^{-1}\Big(\sum_{t=p+1}^{n} Y_t\mathbf{X}_t' -$$

$$- \sum_{t=p+1}^{n} a_t\mathbf{X}_t'\Big) = \mathbf{A}_n^{-1}(\mathbf{v}_n - \boldsymbol{\epsilon}_n), \quad \text{con} \quad \boldsymbol{\epsilon}_n = \frac{1}{n-p} \sum_{t=p+1}^{n} a_t\mathbf{X}_t',$$

por lo cual escribimos

$$n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) = n^{\frac{1}{2}} \mathbf{A}_n^{-1} \boldsymbol{\epsilon}_n .$$

Sea $\boldsymbol{\lambda}' = [\lambda_0, \lambda_1, \ldots, \lambda_p]$ un vector arbitrario en R^{p+1}, entonces

$$n^{\frac{1}{2}} \boldsymbol{\lambda}' \boldsymbol{e}_n = \frac{n^{\frac{1}{2}}}{n-p} \left(\lambda_0 \sum_{t=p+1}^{n} a_t + \sum_{k=1}^{p} \lambda_k \sum_{t=p+1}^{n} Y_{t-k} a_t \right)$$

$$= n^{-\frac{1}{2}} \left(\lambda_0 \sum_{t=p+1}^{n} a_t + \sum_{k=1}^{p} \lambda_k \sum_{t=p+1}^{n} Y_{t-k}\, a_t \right) +$$

$$+ \left(\lambda_0 \sum_{t=p+1}^{n} a_t + \sum_{k=1}^{p} \lambda_k \sum_{t=p+1}^{n} Y_{t-k} a_t \right) O\,(n^{-\frac{1}{2}})$$

$$= n^{-\frac{1}{2}} \left(\lambda_0 \sum_{t=p+1}^{n} a_t + \sum_{k=1}^{p} \lambda_k \sum_{t=p+1}^{n} Y_{t-k} a_t \right) + O_p(n^{-\frac{1}{2}}) \; .$$

pues

- $\lambda_0 \sum_{t=p+1}^{n} a_t + \sum_{k=1}^{p} \lambda_k \sum_{t=p+1}^{n} Y_{t-k} a_t = O_p(1)$

- $\frac{n^{\frac{1}{2}}}{n-p} = (1 + O(1))\, n^{-\frac{1}{2}} = n^{-\frac{1}{2}} + O(n^{-\frac{1}{2}})$.

- $O_p(1) \; O(n^{-\frac{1}{2}}) = O_p(n^{-\frac{1}{2}})$.

Bajo las hipótesis del Teorema 2.1.4.1. Y_t adopta la forma $Y_t = \mu + \sum_{j=0}^{\infty} w_j a_{t-j}$, la cual corresponde a un proceso estacionario (lineal) tal que $E\{Y_t\} = \mu$,

$$\operatorname{var}\{Y_t\} = \sum_{j=0}^{\infty} w_j w_{j+h}\, \sigma^2 = \gamma(h)$$

Así tenemos que

$$n^{\frac{1}{2}} \lambda' \epsilon_n = n^{-\frac{1}{2}} \sum_{t=p+1}^{n} (\lambda_o a_t + \sum_{i=1}^{p} \lambda_i V_{t-i,k}\, a_t) +$$

$$+ n^{-\frac{1}{2}} \sum_{t=p+1}^{n} (\sum_{i=1}^{p} \lambda_i W_{t-i,k}\, a_t) + O_p(n^{-\frac{1}{2}})$$

donde, para k fijo, definimos

$$V_{t,k} = \sum_{j=0}^{k} w_j\, a_{t-j} \quad , \quad W_{t,k} = \sum_{j=k+1}^{\infty} w_j\, a_{t-j}$$

Por otro lado

$$\mathrm{var}\left\{ (n-p)^{-\frac{1}{2}} \sum_{i=1}^{p} \lambda_i (\sum_{t=p+1}^{n} W_{t-i,k} a_t) \right\} = (n-p)^{-1} \left\{ \mathrm{var}\, [\sum_{i=1}^{p} (\right.$$

$$\sum_{t=p+1}^{n} \lambda_i W_{t-i,k} a_t)] + \mathrm{cov}[\sum_{i=1}^{p} (\sum_{t=p+1}^{n} \lambda_i W_{t-i,k} a_t), \sum_{j=1}^{p} (\sum_{t=p+1}^{n} \lambda_j W_{t-j,k} a_t)]$$

$$= \sigma^4 \sum_{i=1}^{p} \sum_{j=k+1}^{\infty} \lambda_i^2\, w_j^2 + \sum_{i \neq j=1}^{p} \lambda_i \lambda_j \sigma^2 \gamma_w(|i-j|) =$$

$$= \sum_{i,j=1}^{p} \lambda_i \lambda_j\, \sigma^2 \gamma_w\, (|i-j|).$$

siendo $\gamma_w(h) = \sigma^2 \sum_{j=k+1}^{\infty} w_j\, w_{j+h}$, $h = |i-j|$.

Ahora bien, como $\sum_{i,j=1}^{p} \lambda_i \lambda_j \sigma^2 \gamma_w(|i-j|) \to o$

si $k \to \infty$, y

$$\mathrm{var}\left\{ n^{-\frac{1}{2}} \sum_{i=1}^{p} \lambda_i \left(\sum_{t=p+1}^{n} W_{t-i,k} a_t \right) \right\} \leq \sum_{i,j=1}^{p} \lambda_i \lambda_j \sigma^2 \gamma_w(|i-j|),$$

entonces

$$H_{kn} = n^{-\frac{1}{2}} \sum_{i=1}^{p} \lambda_i \left(\sum_{t=p+1}^{n} W_{t-i,k} \, a_t \right) \xrightarrow{p_r} o \text{ si } k \to \infty$$

dado que $E\{H_{kn}\} = 0$, para todo n.

Sea

$$Z_{tk} = \lambda_o \, a_t + \sum_{i=1}^{p} \lambda_i \, V_{t-i,k} \, a_t \, , \quad t=p+1, \ldots$$

puede verificarse, sin dificultad, que:

. Z_{tk} es estacionario y (k+p) - dependiente

. $E\{Z_{tk}\} = 0$, $E\{Z_{tk}^2\} = \gamma_{Z_{tk}}(0) = \lambda_o^2 \sigma^2 +$

$$+ \sum_{i,j=1}^{p} \lambda_i \lambda_j \gamma_{V_{tk}}(|i-j|) \sigma^2 = s \neq 0, \text{ para todo t.}$$

donde $\gamma_{V_{tk}}(|i-j|) = \sum_{s=o}^{k} w_s w_{s+h} \sigma^2$

y

$$\gamma_{V_{tk}}(|i-j|) \longrightarrow \gamma(|i-j|) \quad \text{si} \quad k \longrightarrow \infty$$

. $E\{|Z_{tk}|^3\} \leqslant E^{\frac{1}{2}}\{Z_{tk}^4\}\, E^{\frac{1}{2}}\{Z_{tk}^2\} < \infty$, pues $E\{Z_{tk}^4\} < \infty$

Entonces si escribimos

$$n^{\frac{1}{2}} \lambda' \epsilon_n = n^{-\frac{1}{2}} \sum_{t=p+1}^{n} Z_{tk} + H_{kn} + O_p(n^{-\frac{1}{2}}) \quad ,$$

por los Lemas 3.2.2.2. y 3.2.2.3, tenemos

1. $n^{-\frac{1}{2}} \sum_{t=p+1}^{n} Z_{tk} \xrightarrow{\text{Ley}} N(o,s) \quad \text{si} \quad n \longrightarrow \infty$

2. $n^{\frac{1}{2}} \lambda' \epsilon_n \xrightarrow{\text{Ley}} N(o,s') \quad \text{si} \quad n \longrightarrow \infty$

donde $s \longrightarrow s' = \lambda_o \sigma^2 + \sum_{i,j=1}^{p} \lambda_i \lambda_j \gamma(|i-j|)\, \sigma^2 = \lambda' A \lambda\, \sigma^2$

si $k \longrightarrow \infty$.

De esto último, $n^{\frac{1}{2}} \epsilon_n \xrightarrow{\text{Ley}} N(0, A\,\sigma^2)$ y

$n^{\frac{1}{2}} A^{-1} \epsilon_n \xrightarrow{\text{Ley}} N(0, A^{-1}\sigma^2)$, por lo que el Teorema

queda probado pues $A_n \xrightarrow{p_r} A$ implica

$$n^{\frac{1}{2}} A_n^{-1} \epsilon_n \xrightarrow{\text{Ley}} N(0, A^{-1}\sigma^2)$$

<u>COROLARIO 3.2.2.1.:</u> Sean $\mathbf{A}_n$ y $\hat{\boldsymbol{\theta}}$ definidos como en el Teorema 3.2.2.1, entonces

(a) $\mathbf{A}_n = O_p(1)$

(b) $E\{\hat{\boldsymbol{\theta}}\} = \boldsymbol{\theta} + O(n^{-\frac{1}{2}})$

(c) $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = O_p\,(n^{-\frac{1}{2}})$

(d) $\mathrm{var}\,\{\hat{\boldsymbol{\theta}}\} = \sigma^2\, \mathbf{A}^{-1} + O(n^{-1})$

<u>DEMOSTRACION</u>:

(a) Es inmediato, toda vez que $\mathbf{A}_n$ se compone de variables aleatorias con medias finitas.

(b) Sea $\hat{\theta}_i - \theta_i$ $i=0,\ldots,p$, la i-ésima componente de $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$, por el Teorema 3.2.2.1

$$\frac{E\{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\}}{n^{-\frac{1}{2}}} \longrightarrow 0 \quad \text{si } n \longrightarrow \infty \text{, de donde}$$

$$E\{\hat{\theta}_i\} = \theta_i + O(n^{-\frac{1}{2}}) \quad \text{para todo i.}$$

Luego $E\{\hat{\boldsymbol{\theta}}\} = \boldsymbol{\theta} + O(n^{-\frac{1}{2}})$.

(c) $E\{|\hat{\theta}_i - \theta_i|\} = O(n^{-\frac{1}{2}})$, entonces debido al Teorema 3.2.2.1 , $|\hat{\theta}_i - \theta_i|^{\frac{1}{2}} = O_p(n^{-\frac{1}{4}})$, por lo que del Teorema 3.2.1.1. se obtiene

$$\hat{\theta}_i - \theta_i = O_p(n^{-\frac{1}{2}}) \quad \text{para todo} \quad i = 0,\dots,p$$

En consecuencia $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = O_p(n^{-\frac{1}{2}})$.

(d) Sabemos que por el Teorema 3.2.2.1.

$$\frac{\operatorname{var}\{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\} - \sigma^2 \boldsymbol{A}^{-1}}{n^{-1}} \longrightarrow 0 \quad \text{si} \quad n \to \infty$$, así entonces

$$\sigma_{ij} = \sigma^2 a^{ij} + O(n^{-1}) \quad , \quad i,j=1,\dots,p+1$$

siendo σ_{ij} la ij-ésima componente en $\operatorname{var}\{\hat{\boldsymbol{\theta}}\}$ y a^{ij} la ij-ésima componente en $\boldsymbol{A}^{-1}$. Por lo cual

$$\operatorname{var}\{\hat{\boldsymbol{\theta}}\} = \sigma^2 \boldsymbol{A}^{-1} + O(n^{-1})$$

3.2.3. PROPIEDADES ASINTOTICAS DE LOS PREDICTORES

Reconsideremos ahora el problema de obtener un predictor lineal de X_{n+s} $s > 0$, eficiente, dadas las observaciones previas $X_n, X_{n-1},\dots,X_1$, tales que

$$X_t = \hat{\theta}_0 + \hat{\theta}_1 X_{t-1} + \dots + \hat{\theta}_p X_{t-p} + a_t \quad , \quad t=p+1,\dots$$

donde θ_i $i=0,\dots,p$ son obtenidos de (*) sección 3.2.2.

Entonces un predictor lineal de X_{n+s} es

$$E_n\{X_{n+s}\} = \dot{X}_{n+s} = \hat{\theta}_0 + \sum_{j=1}^{p} \hat{\theta}_j \dot{X}_{n+s-j} \quad \text{con}$$

$$\dot{X}_t = X_t \quad \text{si} \quad t \leq n.$$

Notemos que por el Corolario 3.2.2.1 - (c)

$$\hat{\theta}_i = \theta_i + O_p(n^{-\frac{1}{2}}) \quad i=0,\dots,p, \quad \text{de donde}$$

$$\dot{X}_{n+s} = \theta_0 + \sum_{j=1}^{p} \theta_j \hat{X}_{n+s-j} + O_p(n^{-\frac{1}{2}}) \quad \text{y} \quad \hat{X}_t = E_n\{X_t\}.$$

Así, para muestras grandes (n grande)

$\dot{X}_{n+s} \approx E_n\{X_{n+s}\}$, es decir, $\dot{X}_{n+s}$ tiende (en probabilidad) a ser eficiente cuando n crece.

Más aún, como $\dot{e}_{n+s} = X_{n+s} - \dot{X}_{n+s} = \sum_{j=0}^{s-1} w_j a_{n+s-j} +$

$+ \sum_{j=s}^{\infty} (w_j - \hat{w}_j) a_{n+s-j}$ denota el error de predicción, suponiendo parámetros desconocidos, donde w_j se obtiene de

$$\hat{w}_0 = 1$$

$$\hat{w}_1 - \hat{\theta}_1 \hat{w}_0 = 0$$

$$\vdots$$

$$\hat{w}_{p-1} - \dots - \hat{\theta}_{p-1}\hat{w}_0 = 0$$

$$\hat{w}_t - \dots - \hat{\theta}_p \hat{w}_{t-p} = 0 \qquad t \geqslant p.$$

Tenemos que $w_j - \hat{w}_j = O_p(n^{-\frac{1}{2}})$, entonces

$$\dot{e}_{n+s} = \sum_{j=0}^{s-1} w_j \, a_{n+s-j} + O_p(n^{-\frac{1}{2}})$$

$$= e_{n+s} + O_p(n^{-\frac{1}{2}})$$

Asimismo

$$E.C.M\{\dot{X}_{n+s}\} = (1 + w_1^2 + \dots + w_{s-1}^2)\,\sigma^2 + O(n^{-1}) \qquad y$$

$$E\{\dot{e}_{n+s}\} = O(n^{-\frac{1}{2}}).$$

CONCLUSIONES

Según [1], siguiendo los lineamientos de Mann y Wald (1943), la distribución condicional de $Y_{p+1},\ldots,Y_n$ para $Y_p,\ldots,Y_1$ fijos (ver Teorema 3.2.2.1) se determina a partir de la distribución conjunta de $a_{p+1},\ldots,a_n$ con $a_t \sim N(0, \sigma^2)$, $t=p+1,\ldots,n$.

Entonces el estimador de máxima verósimilitud resultante, coincide con el estimador de mínimos cuadrados $\hat{\boldsymbol{\theta}}$, además [6] demuestra que

$$\hat{\boldsymbol{\theta}} \xrightarrow{\text{Ley}} N(\mathbf{0}, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}).$$

Por lo que, si σ^2 es conocido

$$\hat{\theta}_i - z_{\alpha} \sqrt{\sigma^{ii}} \ ; \ \hat{\theta}_i + z_{\alpha} \sqrt{\sigma^{ii}} \qquad \text{y}$$

$$\dot{Y}_{n+s} - z'_{\alpha}\, \sigma \sqrt{\sum_{j=0}^{s-1} w_j^2} \ ; \ \dot{Y}_{n+s} - z'_{\alpha}\, \sigma \sqrt{\sum_{j=0}^{s-1} w_j^2}$$

denotan intervalos de confianza asintóticos, a un nivel α, para θ_i y Y_{n+s} respectivamente.

Esto último queda sustentado con el hecho de que

$$Z = \frac{\hat{\theta}_i - \theta_i}{\sqrt{\sigma^{ii}}} \xrightarrow{\text{Ley}} N(0,1).$$

Y dado que, $\dot{Y}_{n+s} \xrightarrow{\text{Ley}} N(Y_{n+s}, \sum_{j=0}^{s-1} w_j^2 \sigma^2)$. También

$$Z' = \frac{\dot{Y}_{n+s} - Y_{n+s}}{\sigma\sqrt{\sum_{j=0}^{s-1} w_j^2}} \xrightarrow{\text{Ley}} N(0,1).$$

Notemos que si σ^2 es desconocido, se puede emplear el estimador $\hat{\sigma}^{ii}$ de σ^{ii}, correspondiente al i-ésimo elemento diagonal en $\hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1}$, donde

$\hat{\sigma}^2 = \frac{1}{n-2p-1} \sum_{t=p+1}^{n} (Y_t - \mathbf{X}_t \hat{\boldsymbol{\theta}})^2$ es un estimador consistente de σ^2. En efecto, si $\tilde{\sigma}^2 = \frac{1}{n-p} \sum_{t=p+1}^{n} a_t^2$

Tenemos del Corolario 3.2.2.1

$$\hat{\sigma}^2 - \tilde{\sigma}^2 = O_p(n^{-\delta}) \ , \quad 0 < \delta < 1$$

es decir

$$\hat{\sigma}^2 - \tilde{\sigma}^2 = o_p(1).$$

Además var $\{\tilde{\sigma}^2\} = o(1)$, luego $\hat{\sigma}^2 - \sigma^2 = o_p(1)$.

Asimismo, podemos estimar los w_j mediante los $\hat{w}_j$, los cuales son estimadores consistentes obtenidos de las ecuaciones

$$\hat{w}_0 = 1$$

$$\hat{w}_1 - \hat{\theta}_1 \hat{w}_0 = 0$$

$$\cdots\cdots$$

$$\hat{w}_{p-1} - \cdots - \hat{\theta}_{p-1}\hat{w}_0 = 0$$

$$\hat{w}_t - \cdots - \hat{\theta}_p \hat{w}_{t-p} = 0, \qquad t=p,\ p+1, \ldots$$

Se prueba en [1] y [3] que las propiedades de los parámetros estimados por el método de máxima verosimilitud son similares (asintóticamente) a las que se verifican por mínimos cuadrados. Específicamente, en muestras grandes, toda la teoría inferencial de los modelos regresivos es válida cuando los parámetros son estimados por máxima verosimilitud. En consecuencia, todos los procedimientos inferenciales empleados en la teoría de regresión usual son aplicables, asintóticamente, en un modelo de autorregresión, ver [11].

Todo lo anteriormente planteado es válido sólo cuando $E\{a_t a_{t+k}\} = 0$, $k \neq 0$. (errores incorrelacionados). Sin embargo, si por él contrario tenemos el modelo AR(1) (ver [10])

$$\begin{cases} X_t = \theta\, X_{t-1} + a_t & t \quad 2,\ldots,n. \\ a_t = b\, a_{t-1} + \eta_t & \text{con } \eta_t \text{ un proceso aleatorio puro.} \end{cases}$$

Entonces, eliminando a_t tenemos

$$X_t = (b+\theta)\, X_{t-1} - b\,\theta\, X_{t-2} + \eta_t$$

de donde

$$\sum_{t=2}^{n} X_t X_{t-1} = (b+\theta) \sum_{t=2}^{n} X_{t-1}^2 - b\theta \sum_{t=2}^{n} X_{t-1} X_{t-2} + \sum_{t=2}^{n} \eta_t X_{t-1}$$

Y

por el Teorema 3.2.2.1 con $\mu = 0$,

$$\hat{\theta} = b+\theta - b\theta \frac{\sum_{t=2}^{n} X_{t-1} X_{t-2}}{\sum_{t=2}^{n} X_{t-1}^2} + \frac{\sum_{t=2}^{n} \eta_t X_{t-1}}{\sum_{t=2}^{n} X_{t-1}^2} ,$$

por lo cual $\hat{\theta} \xrightarrow{P_r} \frac{b+\theta}{1+b\theta}$, es decir, en este caso $\hat{\theta}$ tiene un error sistemático $\frac{b(1-\theta^2)}{1+b\theta}$. Esto último imposibilita la construcción del intervalo de confianza para θ, y genera predictores no adecuados, en el sentido de que el E.C.M. estimado tiende a aumentar en probabilidad.

BIBLIOGRAFIA

[1] Anderson, T.W. The Statical Analisis of Time Series. John Wiley and Sons. New York, 1971.

[2] Barbut, M. et Forgeaud, C. Eléments D'analyse Mathematique des Chroniques. Collection Hachette Université. Paris, 1971.

[3] Box, G.E.P and Jenkins, G.M. Time Series Analisis, Forecasting and Control. Holden Day (Edición revisada). San Franciso, 1976.

[4] Chatfield, C. The Analisis of Time Series. Theory and Practice. Chapman and Hall (traducido parcialmente por C.I.E.N.E.S.) London, 1976.

[5] Dossou, S. et Ettinger, Gbete P. Contribution a L'etude des Processus Arma. Universite Paul Sabatier (Toulouse), Publications du Laboratoire de Statistique, No. 05.76, Noviembre 1976.

[6] Durbin, J. Estimation of parameters in Time-Series Regresion Models. Journal of the Royal Statistical Society, Series B, vol.22, No. 1, 1960, p. 139-153.

[7] Fuller, Wayne A. Introduction to Statistical Time Series. John Wiley and Sons. New York, 1976.

[8] Hannan, E.J. Time Series Analisis. Butler and Tanner Ltd. London, 1960.

[9] Loéve, M. Probability Theory. D. van Nostrand Company, Inc. New York, 1955.

[10] Malinvaud, E. Métodos Estadísticos de la Econometría. Ediciones Ariel, (traducción al castellano, de la obra original: 'Methodes statistiques de L'économétrie', por Luis Barbe Durán). Barcelona, 1967.

[11] Morettin, P. De Castro, C. Modelos para Previsão de Séries Temporais. I.M.P.A.(13° colóquio Brasileiro de Matemática). Río de Janeiro, 1981.

[12] Vegas Pérez, A. Estadística-Aplicaciones Econométricas y Actuariales. Ediciones Pirámide, S.A. Madrid, 1981.