

IMPLEMENTACIÓN CUADRÁTICA DEL ALGORITMO DE
INNOVACIONES APLICADO A UNA SERIE ESTACIONARIA

NADIA YADHIRA MARTÍNEZ MARTÍNEZ

TESIS

Presentada como Requisito Parcial para
Obtener el Grado de:

MAESTRO EN
ESTADÍSTICA APLICADA



UNIVERSIDAD AUTÓNOMA AGRARIA
“ANTONIO NARRO”
PROGRAMA DE GRADUADOS
Buenavista, Saltillo, Coahuila, México
Septiembre de 2010

Universidad Autónoma Agraria Antonio Narro
Dirección de Postgrado

IMPLEMENTACIÓN CUADRÁTICA DEL ALGORITMO DE
INNOVACIONES APLICADO A UNA SERIE ESTACIONARIA

TESIS

Por:

NADIA YADHIRA MARTÍNEZ MARTÍNEZ

Elaborada bajo la supervisión del comité particular de asesoría y aprobada
como requisito parcial, para optar al grado de

MAESTRO

EN ESTADÍSTICA APLICADA

C o m i t é P a r t i c u l a r

Asesor principal:

Dr. Rolando Cavazos Cadena

Asesor:

Dr. Mario Cantú Sifuentes

Asesor:

M. C. Félix de Jesús Sánchez Pérez

Dr. Jerónimo Landeros Flores

Director de Postgrado

Buenavista, Saltillo, Coahuila, Septiembre de 2010

AGRADECIMIENTOS

DEDICATORIA

COMPENDIO

IMPLEMENTACIÓN CUADRÁTICA DEL ALGORITMO DE INNOVACIONES APLICADO A UNA SERIE ESTACIONARIA

Por

NADIA YADHIRA MARTÍNEZ MARTÍNEZ

MAESTRÍA PROFESIONAL
EN ESTADÍSTICA APLICADA

UNIVERSIDAD AUTÓNOMA AGRARIA
ANTONIO NARRO

BUENAVISTA, SALTILLO, COAHUILA, Septiembre de 2010

Dr. Rolando Cavazos Cadena –Asesor–

Palabras clave: Proyección ortogonal, Errores de pronóstico, Procedimiento de Durbin-Levinson, Teorema de transformación, Procedimiento recursivo.

Este trabajo trata sobre el problema de pronóstico para una serie de tiempo estacionaria. Los propósitos principales son estudiar el comportamiento de los errores de pronóstico conforme el horizonte de planeación se incrementa, y encontrar una implementación eficiente del algoritmo de innovaciones, el cual permite expresar un pronóstico en términos de una base ortogonal del espacio de observaciones disponibles. Dicha implementación se logra combinando el algoritmo de Durbin-Levinson con un teorema de transformación de coordenadas especialmente diseñado para este trabajo. El procedimiento propuesto es de orden cuadrático en la carga aritmética, en contraste con el método de cálculo usual, el cual es de orden cúbico.

ABSTRACT

QUADRATIC IMPLEMENTATION OF THE INNOVATIONS ALGORITHM APPLIED TO A STATIONARY SERIES

BY

NADIA YADHIRA MARTÍNEZ MARTÍNEZ

MASTER

APPLIED STATISTICS

UNIVERSIDAD AUTÓNOMA AGRARIA
ANTONIO NARRO

BUENAVISTA, SALTILLO, COAHUILA, August, 2010

Dr. Rolando Cavazos Cadena –Advisor–

Key Words: Orthogonal projection, Forecasting errors, Durbin-Levinson Scheme, Change of coordinates, Recursive procedure.

This work is concerned with the forecasting problem for a stationary time series with discrete time parameter. The main objectives are (i) to analyze the behavior of the forecasting errors as the planning horizon increases, and (ii) to determine an efficient implementation of the innovations algorithm, which allows to express a forecast in terms of an orthogonal basis of the space of current observations. Such an implementation is obtained by combining the Durbin-Levinson algorithm with a transformation procedure obtained in this work. The proposed method requires $O(n^2)$ arithmetic operations, whereas the usual innovations algorithm has cubic order.

ÍNDICE DE CONTENIDO

1. INTRODUCCIÓN	1
1.1 El Tema de Trabajo	1
1.2 Proyecciones	4
1.3 Los Objetivos Principales	5
1.4 La Organización	8
2. ECUACIONES DE PRONÓSTICO	10
2.1 Ecuaciones y Errores de Pronóstico	10
2.2 Comportamiento de los Errores	15
2.3 Una Condición de Invertibilidad	19
2.4 Consecuencia de la Condición de Invertibilidad	22
3. ALGORITMO DE DURBIN-LEVINSON	26
3.1 Propiedades de los Coeficientes	26
3.2 Teorema de Recurrencia	29
3.3 El Algoritmo de Durbin-Levinson	32
3.4 Implementación del Procedimiento	36
3.5 El Orden del Algoritmo	39
3.6 EL Caso de Procesos Autorregresivos	40
4. ALGORITMO DE INNOVACIONES	48
4.1 Un Conjunto Ortogonal	48
4.2 Relaciones entre los Coeficientes	50
4.3 Algoritmo de Innovaciones	53
5. ALGORITMO DE ORDEN CUADRÁTICO	57
5.1 Relación entre dos Expresiones para un Pronóstico	57

5.2 Procedimiento de Transformación	61
5.3 Implementación del Algoritmo de Innovaciones	64
LITERATURA CITADA	66

CAPÍTULO 1

INTRODUCCIÓN

En este capítulo se hace una descripción del tema de este trabajo, se describen los objetivos que se pretende alcanzar, y se presenta una visión general de la organización del material subsecuente.

1.1. El Tema de Trabajo

Este trabajo se ubica en el área conocida como ‘análisis de una serie de tiempo’, y el tema principal es la *implementación eficiente de un procedimiento de pronóstico para una serie estacionaria*. De manera intuitiva, una series de tiempo es una sucesión $\{X_t\}$ de variables aleatorias definidas sobre un mismo espacio de probabilidad, donde la variable aleatoria X_t se interpreta como la observación que se realiza en el tiempo t , el cual, en el caso considerado en este trabajo, puede variar en un subconjunto de los números enteros. El rasgo fundamental de la sucesión $\{X_t\}$ es que, en contraste con el supuesto comúnmente adoptado en la teoría estadística clásica (Borovkov, 1999; Dudewicz y Mishra, 1998; Shao, 2010 y Wackerly *et al.*, 2009), no se supone la independencia de las variables X_t , ni que éstas tengan la misma distribución, características que permiten incluir en el estudio una gran variedad de observaciones que surgen en la práctica, como las ventas diarias de un almacén, la asistencia semanal a teatros, la población de un país, o la aparición de manchas solares (Brockwell y Davis, 1998; Shumway y Stoffer, 2006 y Chacón, 2010). Por otro lado, una serie de tiempo $\{X_t\}$ se

llama *estacionaria* cuando, para todo entero h , las propiedades ‘importantes’ de un segmento (X_1, X_2, \dots, X_n) son las mismas que las del segmento trasladado $(X_{1+h}, X_{2+h}, \dots, X_{n+h})$. Como no se harán supuestos sobre la distribución de las variables X_t , las propiedades ‘importantes’ se referirán a los momentos: concretamente, la serie es estacionaria si satisface los siguientes dos requerimientos:

- (i) Todas las variables X_t tienen los mismos momentos de orden uno y dos, esto es

$$E[X_t] \text{ y } E[X_t^2] \text{ son finitos y no dependen de } t.$$

- (ii) El segundo momento de la diferencia $X_s - X_t$, es el mismo que el segundo momento de $X_{s+h} - X_{t+h}$ para todo entero h . Estas dos condiciones equivalen a las siguientes propiedades (a) y (b):

(a) $E[X_t] = \mu$ no depende de t ;

(b) Para cada s y t , $\text{Cov}(X_s, X_t)$ está bien definida y depende sólo de la diferencia entre s y t .

Estas propiedades permiten definir la función de autocovarianza asociada a la serie estacionaria $\{X_t\}$ mediante

$$\gamma(h) = E[X_{t+h}X_t], \quad h = 0, \pm 1, \pm 2, \dots \quad (1.1.1)$$

Para una serie estacionaria, en el sentido recién definido, no es difícil ver que para cada entero h y cada entero positivo n ,

$$E[(X_1, X_2, \dots, X_n)'] = E[(X_{1+h}, X_{2+h}, \dots, X_{n+h})']$$

y

$$\text{Var}[(X_1, X_2, \dots, X_n)'] = \text{Var}[(X_{1+h}, X_{2+h}, \dots, X_{n+h})']$$

de manera que los segmentos (X_1, X_2, \dots, X_n) y $(X_{1+h}, X_{2+h}, \dots, X_{n+h})$ tienen el mismo valor esperado y la misma matriz de varianzas; en este sentido, ambos

segmentos tienen las mismas propiedades ‘importantes’, y el análisis de una serie de tiempo estacionaria estará basado en propiedades de segundo orden. Debido a que si $\{X_t\}$ es una serie estacionaria, entonces $\{X_t - \mu\}$ también lo es, en este trabajo supondremos, sin pérdida de generalidad, que la esperanza de X_t es nula. La idea de serie estacionaria es bastante restrictiva y, en general, una de tales series no es un modelo razonable para los datos que se observan comúnmente en la práctica; sin embargo, una serie estacionaria es la componente esencial del denominado *modelo clásico*, el cual si captura una amplia gama de situaciones reales (Chacón, 2010).

El problema de pronóstico para una serie estacionaria puede describirse como sigue: Dadas las observaciones X_1, \dots, X_n registradas hasta el tiempo n , determinar ‘la mejor’ aproximación para la variable X_{n+h} que se observará en el tiempo futuro $n+h$ en términos de los datos disponibles X_1, \dots, X_n . Un resultado conocido en la teoría estadística establece que la mejor aproximación a X_{n+h} en el sentido de minimizar el error cuadrático esperado es la esperanza condicional $g(X_1, \dots, X_n) = E[X_{n+h}|X_1, \dots, X_n]$; sin embargo, al no conocer la distribución de $\{X_t\}$, la función $g(\cdot)$ no se puede calcular, y en la teoría de series de tiempo, la mejor aproximación que se pretende encontrar es ‘la mejor aproximación lineal basada en los datos observados hasta el momento actual n ’:

$$\hat{g}(X_1, \dots, X_n) = a_1 X_1 + \dots + a_n X_n \quad (1.1.2)$$

donde a_1, \dots, a_n son constantes y

$$\min_{b_1, \dots, b_n} E[(X_{n+h} - (b_1 X_1 + \dots + b_n X_n))^2] = E[(X_{n+h} - (a_1 X_1 + \dots + a_n X_n))^2]. \quad (1.1.3)$$

tales números a_i siempre existen y se expresan solamente en términos de la función de autocovarianza $\gamma(\cdot)$; de hecho, a_1, \dots, a_n pueden determinarse por medio de diferenciación (Fulks, 1980; Khuri, 2002).

En un sentido general, *los problemas* de interés en este trabajo, se refieren a las propiedades de los coeficientes en (1.1.2) y del error de pronóstico en el lado derecho de (1.1.3), así como a métodos eficientes para calcular dichas cantidades.

1.2. Proyecciones

Puesto que este trabajo se centra en la determinación de las constantes a_1, \dots, a_n en (1.1.2), las cuales tienen la propiedad de optimalidad en (1.1.3), las ideas básicas de *espacio vectorial* y *producto interno* son de importancia fundamental en el desarrollo subsecuente; para un estudio de estos conceptos en espacios de dimensión finita, vea, por ejemplo, Graybill (2001), Harville (2008), Hoffman y Kunze (1975) y, particularmente, Strang (2003). La intuición del caso de dimensión finita es de gran importancia en el estudio de series de tiempo, aunque el entorno de trabajo es, naturalmente, el espacio

$$L^2 = L^2(\Omega, \mathcal{F}, P)$$

de todas las variables aleatorias $Y : \Omega \rightarrow R$ con segundo momento finito, y donde (Ω, \mathcal{F}, P) es el espacio de probabilidad donde la serie $\{X_t\}$ está definida. Este espacio es de dimensión infinita y el producto interno está especificado por

$$\langle Y_1, Y_2 \rangle = E[Y_1 Y_2], \quad Y_1, Y_2 \in L^2, \quad (1.2.1)$$

y la norma correspondiente está dada por

$$\|Y\| = \sqrt{\langle Y, Y \rangle} = E[Y^2]^{1/2}. \quad (1.2.2)$$

Con respecto a esta norma, L^2 es un espacio métrico completo, es decir, toda sucesión de Cauchy es convergente (Apostol, 1980; Rudin, 1984 y Royden, 2003). Considere ahora, una sucesión finita W_1, \dots, W_k de variables en L^2 , y sea \mathcal{W} el espacio vectorial generado por estas variables:

$$\mathcal{W} = \{c_1 W_1 + \dots + c_n W_n \mid c_1, c_2, \dots, c_n \in \mathbb{R}\}$$

el cual es un subespacio cerrado de L^2 (Rudin, 1984; Royden 2003). Dado $Y \in L^2$, existe una única variable aleatoria $P_{\mathcal{W}}Y$ que satisface las siguientes propiedades:

$$P_{\mathcal{W}}Y \in \mathcal{W} \quad \text{y} \quad \|Y - P_{\mathcal{W}}Y\| \leq \|Y - \mathbf{w}\| \quad \text{para todo } \mathbf{w} \in \mathcal{W}. \quad (1.2.3)$$

Estas condiciones son equivalentes a

$$P_{\mathcal{W}}Y \in \mathcal{W} \quad \text{y} \quad \langle P_{\mathcal{W}}Y, \mathbf{w} \rangle = \langle Y, \mathbf{w} \rangle, \quad \mathbf{w} \in \mathcal{W}. \quad (1.2.4)$$

(Brockwell y Davis, 1998; Strang, 2003); debido a la (bi-)linealidad del producto interno, estas condiciones pueden expresarse como

$$P_{\mathcal{W}}Y \in \mathcal{W} \quad \text{y} \quad \langle P_{\mathcal{W}}Y, W_i \rangle = \langle Y, W_i \rangle, \quad i = 1, 2, \dots, k. \quad (1.2.5)$$

La variable $P_{\mathcal{W}}Y$ se llaman la proyección (ortogonal) de Y sobre \mathcal{W} y es una transformación lineal; la equivalencia de (1.2.3–1.2.5) se conoce como *teorema de proyección*. La siguiente propiedad será útil: Si \mathcal{W}_0 y \mathcal{W}_1 son dos subespacios de L^2 , entonces

$$\mathcal{W}_0 \subset \mathcal{W}_1 \Rightarrow P_{\mathcal{W}_0}P_{\mathcal{W}_1} = P_{\mathcal{W}_0} \quad (1.2.6)$$

A partir de esta discusión se es claro que el mejor pronóstico $g(X_1, \dots, X_n)$ para X_{n+1} dadas las observaciones X_1, \dots, X_n , es la proyección de X_{n+1} sobre el espacio generado por X_1, \dots, X_n ; compare (1.1.2) y (1.1.3) con (1.2.3).

1.3. Los Objetivos Principales

Como ya se mencionó, dada una serie estacionaria, el interés de este trabajo se centra en la formulación del mejor pronóstico lineal de X_{n+h} dadas las observaciones X_1, \dots, X_n ; tal pronóstico es la proyección de X_{n+h} sobre el espacio generado por X_1, \dots, X_n y será denotado por \hat{X}_{n+h} , de manera que el error cuadrático de pronóstico es v_n^h está dado por

$$v_n^h = \|X_{n+h} - \hat{X}_{n+h}\|^2 = E[(X_{n+h} - \hat{X}_{n+h})^2]. \quad (1.3.1)$$

los objetivos de este trabajo se describen a continuación:

- (i) Analizar el comportamiento respecto a h de los errores cuadráticos de pronóstico v_n^h ; en particular se trata de investigar la validez de las siguientes relaciones:

$$v_n^h \leq v_n^{h+1}, \quad h = 1, 2, 3, \dots$$

y

$$\lim_{h \rightarrow \infty} v_n^h = \gamma(0).$$

La intuición detrás de estas propiedades es que a medida que el tiempo real de observación del dato X_{n+h} se aleja del tiempo actual n , será cada vez más difícil hacer un ‘buen pronóstico’, y que al considerar un dato que se observará en un tiempo muy distante del momento actual n , la influencia de los datos disponibles X_1, \dots, X_n se desvanece. Los resultados en esta dirección se establecen en el Capítulo 2, y muestran que las anteriores relaciones son correctas bajo condiciones adecuadas o cuando se reformulan apropiadamente, pero que, en general, no son ciertas para todas las series estacionarias.

- (ii) Estudiar la relación entre la expresión de un pronóstico \hat{X}_{n+h} en términos de bases diferentes para el espacio \mathcal{L}_n generado por X_1, \dots, X_n . Particularmente, se tiene interés en una base ortogonal U_1, \dots, U_n de \mathcal{L}_n , denominada base de innovaciones.

Este problema requiere de un análisis profundo que se desarrolla principalmente en los Capítulos 3 y 4, culminando con el teorema de transformación en el Capítulo 5.

- (iii) Determinar un algoritmo eficiente para encontrar los coeficientes del pronóstico \hat{X}_{n+h} expresado en términos de la base ortogonal de innovaciones.

Por lo tanto, el resultado en esta dirección se establece en el Teorema 5.3.1 del Capítulo 5, exhibiendo un algoritmo de orden cuadrático para implementar el método de innovaciones, en contraste con la implementación usual, que es de orden cúbico.

1.4. La Organización

La presentación de este trabajo ha sido organizada de la siguiente manera:

En el Capítulo 2, dada una serie de tiempo estacionaria se formulan las ecuaciones que determinan el pronóstico \hat{X}_{n+h} en términos de los datos actuales X_1, \dots, X_n y se encuentra una expresión simple para el correspondiente error de pronóstico v_n^h , analizando sus propiedades de monotonicidad y su comportamiento conforme h se incrementa sin límite.

En el Capítulo 3 se formula el procedimiento de Durbin-Levinson para encontrar los coeficientes φ_n en la representación del pronóstico \hat{X}_{n+1} como combinación lineal de X_1, X_2, \dots, X_n ; el esquema de cálculo es recursivo, y el vector de coeficientes φ_k se determina de una forma sencilla en términos de φ_{k-1} . Se discute la aplicación del algoritmo a procesos autorregresivos, lo cual conduce de forma natural, a estudiar la idea de *causalidad* de un proceso.

En el Capítulo 4 se estudia la representación de una proyección \hat{X}_{n+h} en términos de la base ortogonal de innovaciones U_1, \dots, U_n . Se formula el denominado algoritmo de innovaciones para determinar el vector de coeficientes θ_n correspondiente, y se analiza el orden del procedimiento, demostrando que es de orden cúbico respecto a la carga aritmética.

Finalmente, en el Capítulo 5 se presenta el principal resultado de este trabajo: se propone un algoritmo de orden cuadrático para determinar el vector de coeficientes θ_n en la representación del pronóstico \hat{X}_{n+1} en términos de las primeras n innovaciones. El procedimiento que se propone se basa en determinar la relación entre θ_n y los vectores de coeficientes φ_k en la representación de los pronósticos \hat{X}_{k+1} como combinación lineal de las variables originales X_1, \dots, X_k . Esta relación, conjuntamente con el algoritmo de Durbin Levinson, determinan el procedimiento propuesto.

Finalmente, los procedimientos numéricos que se utilizan en el trabajo son simples y se proporciona el código en lenguaje R de los métodos utilizados y, para facilitar la presentación, a partir del Capítulo 3 se supone que el horizonte de pronóstico h es igual a 1.

CAPÍTULO 2

ECUACIONES DE PRONÓSTICO

Dada una serie de tiempo estacionaria, en este capítulo se formulan las ecuaciones que determinan la mejor aproximación lineal para una observación futura en términos de los datos disponibles, y se encuentra una expresión simple para el correspondiente error de pronóstico. Se analizan dos ideas sobre el comportamiento de los errores que, desde un punto de vista intuitivo, parece natural esperar que sean válidas. El estudio que se realiza muestra que dichas ideas no son, en general, correctas, pero un análisis más profundo muestra que, bajo condiciones adecuadas o después de una reformulación apropiada, el comportamiento de los errores está de acuerdo con la intuición.

2.1. Ecuaciones y Errores de Pronóstico

Dada una serie estacionaria $\{X_t\}$ con función de autocovarianza $\gamma(\cdot)$, en esta sección se establecen las ecuaciones para determinar la proyección ortogonal de una variable X_{n+h} , donde $h > 0$, sobre el espacio generado por las variables X_1, \dots, X_n donde, por el momento, el entero n es fijo, y \mathcal{L}_n es el subespacio de L^2 generado por las variables X_t con $1 \leq t \leq n$:

$$\mathcal{L}_n(X_1, \dots, X_n) \equiv \mathcal{L}_n := \{a_1 X_1 + \dots + a_n X_n \mid (a_1, \dots, a_n) \in \mathbb{R}^n\}. \quad (2.1.1)$$

Para cualquier variable aleatoria Y con varianza finita, la proyección ortogonal de Y sobre \mathcal{L}_n se denotará mediante \hat{Y} :

$$\hat{Y} = P_{\mathcal{L}_n} Y, \quad (2.1.2)$$

de manera que \hat{Y} está caracterizada por los siguientes requerimientos:

$$\hat{Y} \in \mathcal{L}_n, \quad \text{y} \quad \langle Y, X_k \rangle = \langle \hat{Y}, X_k \rangle, \quad k = 1, 2, \dots, n; \quad (2.1.3)$$

vea el Teorema de proyección. En palabras, \hat{Y} es una combinación lineal de X_1, \dots, X_n , y los productos internos de \hat{Y} y Y con X_i coinciden, $i = 1, 2, \dots, n$. Ahora se determinará la forma que estas condiciones adoptan para el caso en que Y es la variable X_{n+h} que se observará h unidades después de haber registrado el valor de X_n ; primero, se estudiará el caso $h = 1$. Observe que la inclusión $\hat{X}_{n+1} \in \mathcal{L}_n$ significa que \hat{X}_{n+1} es una combinación lineal de X_1, \dots, X_n , de manera que existen constantes $\phi_{n1}, \phi_{n2}, \dots, \phi_{nn}$ tales que

$$\hat{X}_{n+1} = \phi_{n1}X_n + \phi_{n2}X_{n-1} + \dots + \phi_{nn}X_1,$$

o de forma más compacta,

$$\hat{X}_{n+1} = \sum_{j=1}^n \phi_{nj}X_{n+1-j}, \quad (2.1.4)$$

A continuación, note que la condición de que X_{n+1} y \hat{X}_{n+1} tengan los mismos productos internos con las variables X_r , $r = 1, 2, \dots, n$ puede escribirse en términos de la función de autocovarianza:

$$\begin{aligned} \langle X_{n+1}, X_r \rangle &= \langle \hat{X}_{n+1}, X_r \rangle \\ &= \left\langle \sum_{j=1}^n \phi_{nj}X_{n+1-j}, X_r \right\rangle \\ &= \sum_{j=1}^n \phi_{nj} \langle X_{n+1-j}, X_r \rangle, \quad r = 1, 2, \dots, n, \end{aligned} \quad (2.1.5)$$

igualdades que equivalen a

$$\gamma(n+1-r) = \sum_{j=1}^n \phi_{nj} \gamma(n+1-j-r), \quad r = 1, 2, \dots, n.$$

Una forma más conveniente para este sistema se obtiene escribiendo

$$i = n + 1 - r;$$

con esta notación, cuando r toma cualquier valor entre 1 y n , i también lo hace y el anterior sistema de ecuaciones se escribe como

$$\gamma(i) = \sum_{j=1}^n \phi_{nj} \gamma(i-j), \quad i = 1, 2, \dots, n.$$

Definiendo los vectores $\gamma_n, \varphi_n \in \mathbb{R}^n$ y la matriz Γ_n de orden $n \times n$ mediante

$$\gamma_n = \begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(n) \end{bmatrix}, \quad \varphi_n = \begin{bmatrix} \phi_{n1} \\ \phi_{n2} \\ \vdots \\ \phi_{n,n} \end{bmatrix}, \quad \Gamma_n = [\gamma(i-j)]_{i,j=1,2,\dots,n}; \quad (2.1.6)$$

con esta notación, las ecuaciones anteriores equivalen a

$$\gamma_n = \Gamma_n \varphi_n, \quad (2.1.7)$$

sistema que siempre es consistente, por el teorema de proyección. Después de encontrar una solución φ_n para esta ecuación vectorial, el error cuadrático de pronóstico, denotado por v_n y definido mediante

$$v_n = \|X_{n+1} - \hat{X}_{n+1}\|^2, \quad (2.1.8)$$

puede determinarse de forma muy simple. En efecto, observando que

$$\langle X_{n+1} - \hat{X}_{n+1}, X_r \rangle = 0$$

para todo $r = 1, 2, \dots, n$ (vea la primera igualdad en (2.1.5)), como \hat{X}_{n+1} es una combinación lineal de X_1, \dots, X_n se desprende que

$$\langle X_{n+1} - \hat{X}_{n+1}, \hat{X}_{n+1} \rangle = 0, \quad (2.1.9)$$

lo que equivale a

$$\langle \hat{X}_{n+1}, X_{n+1} \rangle = \langle \hat{X}_{n+1}, \hat{X}_{n+1} \rangle = \|\hat{X}_{n+1}\|^2.$$

Con esto en mente,

$$\begin{aligned}
\|X_{n+1} - \hat{X}_{n+1}\|^2 &= \langle X_{n+1} - \hat{X}_{n+1}, X_{n+1} - \hat{X}_{n+1} \rangle \\
&= \langle X_{n+1} - \hat{X}_{n+1}, X_{n+1} \rangle - \langle X_{n+1} - \hat{X}_{n+1}, \hat{X}_{n+1} \rangle \\
&= \langle X_{n+1} - \hat{X}_{n+1}, X_{n+1} \rangle \quad (\text{por la ecuación (2.1.9)}) \\
&= \langle X_{n+1}, X_{n+1} \rangle - \langle \hat{X}_{n+1}, X_{n+1} \rangle \\
&= \gamma(0) - \langle \hat{X}_{n+1}, \hat{X}_{n+1} \rangle
\end{aligned}$$

donde, de nueva cuenta, se utilizó la ecuación (2.1.9) para establecer la última igualdad. Por lo tanto,

$$\|X_{n+1} - \hat{X}_{n+1}\|^2 = \gamma(0) - \langle \hat{X}_{n+1}, \hat{X}_{n+1} \rangle \quad (2.1.10)$$

Ahora se calculará la norma cuadrática de \hat{X}_{n+1} . Usando (2.1.4) se tiene que

$$\begin{aligned}
\langle \hat{X}_{n+1}, \hat{X}_{n+1} \rangle &= \left\langle \sum_{j=1}^n \phi_{nj} X_{n+1-j}, \sum_{i=1}^n \phi_{ni} X_{n+1-i} \right\rangle \\
&= \sum_{j=1}^n \sum_{i=1}^n \phi_{ni} \langle X_{n+1-j}, X_{n+1-i} \rangle \phi_{nj} \\
&= \sum_{j=1}^n \sum_{i=1}^n \phi_{ni} \gamma(i-j) \phi_{nj} \\
&= \boldsymbol{\varphi}'_n \boldsymbol{\Gamma}_n \boldsymbol{\varphi}_n
\end{aligned}$$

y recordando que $\boldsymbol{\Gamma}_n \boldsymbol{\varphi}_n = \boldsymbol{\gamma}_n$, se obtiene que $\langle \hat{X}_{n+1}, \hat{X}_{n+1} \rangle = \boldsymbol{\varphi}'_n \boldsymbol{\gamma}_n$; combinando esta relación con (2.1.10), se llega a la expresión

$$v_n = \|X_{n+1} - \hat{X}_{n+1}\|^2 = \gamma(0) - \boldsymbol{\varphi}'_n \boldsymbol{\gamma}_n.$$

De esta manera, al resolver la ecuación (2.1.7) es posible determinar, tanto la proyección \hat{X}_{n+1} por medio de (2.1.4), como el error de pronóstico v_n a través de la anterior expresión desplegada. Esta discusión se resume en el siguiente teorema.

Teorema 2.1.1. Suponga que $\{X_t\}$ es una serie estacionaria con función de autocovarianza $\gamma(\cdot)$. En este caso, la proyección \hat{X}_{n+1} de X_{n+1} sobre el espacio \mathcal{L}_n

generado por X_1, \dots, X_n está dada por

$$\hat{X}_{n+1} = \sum_{j=1}^n \phi_{nj} X_{n+1-j}$$

donde $\varphi_n = (\phi_{n1}, \dots, \phi_{nn})'$ es cualquier vector que satisfice el sistema de ecuaciones

$$\gamma_n = \Gamma_n \varphi_n,$$

y la notación es como en (2.1.6). Más aún, el error de pronóstico v_n está dado por

$$v_n = \|\hat{X}_{n+1} - X_{n+1}\|^2 = \gamma(0) - \varphi_n' \gamma_n. \quad (2.1.11)$$

Para concluir esta sección, ahora se discutirá brevemente la determinación de la proyección \hat{X}_{n+h} de X_{n+h} sobre \mathcal{L}_n . Como $\hat{X}_{n+h} \in \mathcal{L}_n$ existen constantes $\phi_{n1}^h, \phi_{n2}^h, \dots, \phi_{nn}^h$ tales que

$$\hat{X}_{n+h} = \phi_{n1}^h X_n + \phi_{n2}^h X_{n-1} + \dots + \phi_{nn}^h X_1, \quad (2.1.12)$$

y entonces las condiciones (2.1.3) implican que, para cada $i = 1, 2, \dots, n$,

$$\begin{aligned} \gamma(h+i-1) &= \langle X_{n+h}, X_{n+1-i} \rangle \\ &= \langle \hat{X}_{n+h}, X_{n+1-i} \rangle \\ &= \left\langle \sum_{j=1}^n \phi_{nj}^h X_{n+1-j}, X_{n+1-i} \right\rangle \\ &= \sum_{j=1}^n \phi_{nj}^h \langle X_{n+1-j}, X_{n+1-i} \rangle = \sum_{j=1}^n \phi_{nj}^h \gamma(i-j); \end{aligned}$$

Definiendo los vectores $\gamma_n^h, \varphi_n^h \in \mathbb{R}^n$ mediante

$$\gamma_n^h = \begin{bmatrix} \gamma(h) \\ \gamma(h+1) \\ \vdots \\ \gamma(h+n-1) \end{bmatrix} \quad \text{y} \quad \varphi_n^h = \begin{bmatrix} \phi_{n1}^h \\ \phi_{n2}^h \\ \vdots \\ \phi_{nn}^h \end{bmatrix}, \quad (2.1.13)$$

se desprende que

$$\gamma_n^h = \mathbf{\Gamma}_n \varphi_n^h \quad (2.1.14)$$

donde $\mathbf{\Gamma}_n$ es la matriz $n \times n$ definida en (2.1.6). Más aún, procediendo como en el caso $h = 1$, es posible obtener una fórmula simple para el error de pronóstico $\|X_{n+h} - \hat{X}_{n+h}\|^2$:

$$v_n^h := \|X_{n+h} - \hat{X}_{n+h}\|^2 = \|X_{n+h}\|^2 - \|\hat{X}_{n+h}\|^2 = \gamma(0) - \varphi_n^{h'} \gamma_n^h. \quad (2.1.15)$$

2.2. Comportamiento de los Errores

En esta sección se ejemplifica el cálculo de proyecciones mediante la solución del sistema (2.1.7) o (2.1.14), y se analizan dos aspectos simples que, a primera vista, parece natural esperar sobre los errores de pronóstico:

- (i) En el tiempo n , se han observado X_1, X_2, \dots, X_n , y a medida que $h > 0$ aumenta, el tiempo de observación $n + h$ de la variable X_{n+h} se aleja más del tiempo actual n . Esto puede generar la impresión de que cada vez será más difícil pronosticar X_{n+h} usando los datos X_1, \dots, X_n , lo cual se reflejará en el incremento del error de pronóstico $v_n^h = \|X_{n+h} - \hat{X}_{n+h}\|^2$ al crecer h . La expresión formal de esta primera impresión es

$$v_n^h \leq v_n^{h+1}, \quad h = 1, 2, 3, \dots \quad (2.2.1)$$

- (ii) Llevando al extremo la idea anterior, conforme h crece sin límite, parecería natural esperar que ‘la influencia’ de X_1, \dots, X_n sobre X_{n+h} tienda a desvanecerse conforme h se incrementa, en el sentido de que $\lim_{h \rightarrow \infty} \hat{X}_{n+h} = 0$. En este caso, usando que $v_n^h = \gamma(0) - \|\hat{X}_{n+h}\|^2$, parece natural esperar que

$$\lim_{h \rightarrow \infty} v_n^h = \gamma(0). \quad (2.2.2)$$

A partir de los siguientes ejemplos, se verá que *estas dos impresiones no son correctas en general*; sin embargo, se mostrará que la intuición va en el camino correcto bajo condiciones adecuadas, o después de una reformulación apropiada de las ideas.

Ejemplo 2.2.1. Considere el proceso de promedios móviles $X_t = Z_t + Z_{t-1} + Z_{t-5}$, donde la serie $\{Z_t\}$ es un ruido blanco con varianza unitaria, en este caso, la función de autocovarianza está dada por

$$\gamma(h) = \begin{cases} 3, & \text{si } h = 0 \\ 1, & \text{si } h = \pm 1 \text{ o } h = \pm 5, \\ 0, & \text{de otro modo;} \end{cases}$$

vea, por ejemplo, Brockwell y Davis (1998), Fuller (1998), o Shumway y Stoffer (2006). Dado $n = 2$, se determinarán, para cada $h > 0$, las proyecciones \hat{X}_{2+h} de X_{h+2} sobre \mathcal{L}_2 .

- (i) Para encontrar $\hat{X}_3 = \hat{X}_{2+1}$ se resuelve el sistema (2.1.7) (que equivale al sistema (2.1.14) con $h = 1$):

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \gamma(1) \\ \gamma(2) \end{bmatrix} = \begin{bmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(0) \end{bmatrix} \begin{bmatrix} \phi_{21} \\ \phi_{22} \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \phi_{21} \\ \phi_{22} \end{bmatrix}.$$

Este sistema tiene matriz invertible, y su única solución es $\varphi_2 = (3/8, -1/8)'$. Por lo tanto, $\hat{X}_3 = (3/8)X_2 - (1/8)X_1$, y el error de pronóstico es

$$v_2^1 = v_2 = \|X_3 - P_{\mathcal{L}_2}X_3\|^2 = 3 - (3/8, -1/8) \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 21/8.$$

- (ii) Las proyecciones $\hat{X}_4 = \hat{X}_{2+2}$ y $\hat{X}_5 = \hat{X}_{2+3}$ se determinan resolviendo (2.1.14) con $h = 2$ y $h = 3$, respectivamente. Observe ahora que $\gamma_2^h = (\gamma(h), \gamma(h+1))' = (0, 0)'$ para $h = 2, 3$, de modo que el sistema $\gamma_h = \mathbf{\Gamma}_2 \varphi_2^h$ tiene la única solución $\varphi_2^h = (0, 0)'$, y entonces $\hat{X}_4 = \hat{X}_5 = 0$ y los correspondientes errores de pronóstico son

$$v_2^2 = \|X_4 - \hat{X}_4\|^2 = \gamma(0) - \varphi_2^{2'} \gamma_2^2 = 3$$

$$v_2^3 = \|X_5 - \hat{X}_5\|^2 = \gamma(0) - \varphi_2^{3'} \gamma_2^3 = 3$$

(iii) La proyección $\hat{X}_6 = \hat{X}_{2+4} = P_{\mathcal{L}_2} X_6$ se encuentra observando que $\gamma_2^4 = (\gamma(4), \gamma(5))' = (0, 1)'$. En este caso, (2.1.14) con $h = 4$ se reduce

a

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \gamma(4) \\ \gamma(5) \end{bmatrix} = \begin{bmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(0) \end{bmatrix} \begin{bmatrix} \phi_{21}^4 \\ \phi_{22}^4 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \phi_{21}^4 \\ \phi_{22}^4 \end{bmatrix}.$$

cuya única solución es $\varphi_2^4 = (-1/8, 3/8)'$. La proyección \hat{X}_6 es $-X_2/8 + 3X_1/8$ y el correspondiente error de pronóstico es

$$v_2^4 = \|X_6 - P_{\mathcal{L}_2} X_6\|^2 = \gamma(0) - \varphi_2^{4'} \gamma_2^4 = 3 - 3/8 = 21/8$$

(iv) Para encontrar la proyección $\hat{X}_7 = \hat{X}_{2+5} = P_{\mathcal{L}_2} X_7$ observe que $\gamma_2^5 = (\gamma(5), \gamma(6))' = (1, 0)'$, de tal forma que (2.1.14) con $h = 5$ equivale a

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \gamma(5) \\ \gamma(6) \end{bmatrix} = \begin{bmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(0) \end{bmatrix} \begin{bmatrix} \phi_{21}^5 \\ \phi_{22}^5 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \phi_{21}^5 \\ \phi_{22}^5 \end{bmatrix}.$$

cuya única solución es $\varphi_2^5 = (3/8, -1/8)'$; se tiene que $\hat{X}_7 = (3X_2 - X_1)/8$ y el error de pronóstico es

$$v_2^5 = \|X_7 - \hat{X}_7\|^2 = \gamma(0) - \varphi_2^{5'} \gamma_2^5 = 3 - 3/8 = 21/8$$

(v) Para evaluar $\hat{X}_{2+h} = P_{\mathcal{L}_2} X_{2+h}$ cuando $h \geq 6$, note que la especificación de $\gamma(\cdot)$ implica que $\gamma_2^h = (\gamma(h), \gamma(h+1))' = (0, 0)'$. Por lo tanto, la solución de $\gamma_2^h = \Gamma_2 \varphi_2^h$ es $\varphi_2^h = (0, 0)'$, y $\hat{X}_{2+h} = 0$ para $h \geq 6$; el correspondiente error de pronóstico es $v_2^h = \|X_{2+h} - \hat{X}_{2+h}\|^2 = \|X_{2+h}\|^2 = \gamma(0) = 3$. \square

En el ejemplo anterior $n = 2$ y la sucesión de errores de pronóstico es

$$(v_2^1, v_2^2, v_2^3, v_2^4, v_2^5, v_2^6, v_2^7, \dots) = \left(\frac{21}{8}, 3, 3, \frac{21}{8}, \frac{21}{8}, 3, 3, \dots \right)$$

y es claro que v_2^h no es una función monótona creciente de h , de tal forma que la propiedad (2.2.1) no se satisface. Sin embargo, como $v_2^h = 3 = \gamma(0)$ para $h \geq 6$, en este ejemplo la convergencia en (2.2.2) *si ocurre*. Sin embargo, el siguiente ejemplo muestra que la validez de (2.2.2) tampoco puede garantizarse en el caso general.

Ejemplo 2.2.2. Considere la serie $\{X_t\}$ definida mediante

$$X_t = A \cos(\pi t/3) + B \sin(\pi t/3),$$

donde A y B son variables aleatorias no correlacionadas con varianza unitaria. En este caso, la correspondiente función de autocovarianza es

$$\gamma(h) = \cos(\pi h/3);$$

vea (Chacón, 2010) para una deducción detallada de este hecho. Ahora tome $n = 1$ y considere el correspondiente error de pronóstico

$$v_1^h = \|X_{1+h} - P_{\mathcal{L}_1} X_{1+h}\|^2 = \gamma(0) - \|P_{\mathcal{L}_1} X_{1+h}\|^2 = 1 - \|P_{\mathcal{L}_1} X_{1+h}\|^2$$

En este caso $P_{\mathcal{L}_1} X_{1+h} = \phi_{11}^h X_1$ y las ecuaciones (2.1.14) se reducen a $\gamma(h) = \gamma(0)\phi_{11}^h$, de donde $\phi_{11}^h = \gamma(h)/\gamma(0) = \cos(\pi h/3)$, de manera que

$$\|P_{\mathcal{L}_1} X_{1+h}\|^2 = \|\cos(\pi h/3)X_1\|^2 = \cos^2(\pi h/3)$$

y entonces $v_1^h = 1 - \cos^2(\pi h/3) = \sin^2(\pi h/3)$. Esta expresión es periódica respecto a h , y al variar h toma los valores $1/2, 1/2, 0, 1/2, 1/2, 0$ una y otra vez. Por lo tanto, v_1^h no converge a $\gamma(0) = 1$ conforme h tiene a ∞ . \square

Aunque (2.2.1) y (2.2.2) no son ciertas en general, en la siguiente sección se presenta una condición suficiente para que esta última relación ocurra, y se verá que, con una interpretación ligeramente distinta, una versión de (2.2.1) también es correcta.

2.3. Una Condición de Invertibilidad

La determinación del vector de coeficientes φ_n^h requiere resolver el sistema (2.1.14) cuya matriz es $\mathbf{\Gamma}_n = [\gamma(i-j)]_{i,j=1,\dots,n}$. En esta sección se presenta una condición sobre la función de autocovarianza $\gamma(\cdot)$ que asegura que dicha matriz $\mathbf{\Gamma}_n$ es invertible para cada n .

Teorema 2.3.1. Suponga que $\gamma(\cdot)$ es la función de autocovarianza de un proceso estacionario $\{X_t\}$ para la cual las siguientes condiciones (i) y (ii) se satisfacen:

- (i) $\gamma(0) > 0$;
- (ii) $\lim_{k \rightarrow \infty} \gamma(k) = 0$.

En estas circunstancias, para cada $n = 1, 2, 3, \dots$ la matriz $\mathbf{\Gamma}_n$ en (2.1.6) es invertible

Demostración. Note que $\mathbf{\Gamma}_1 = [\gamma(0)]$ es una matriz invertible, pues $\gamma(0) > 0$. El argumento ahora procede por contradicción. Suponga que $\mathbf{\Gamma}_r$ es una matriz singular para algún $r > 1$. En este caso, debido a que $\mathbf{\Gamma}_1$ es invertible, existe un entero $n \geq 1$ tal que

$$\mathbf{\Gamma}_1, \mathbf{\Gamma}_2, \dots, \mathbf{\Gamma}_n \text{ son invertibles, pero } \mathbf{\Gamma}_{n+1} \text{ es singular.} \quad (2.3.1)$$

Luego, existe un vector $\mathbf{b} = (b_1, \dots, b_n, b_{n+1})' \in \mathbb{R}^{n+1}$ tal que

$$\mathbf{\Gamma}_{n+1} \mathbf{b} = 0, \quad \mathbf{b} \neq 0. \quad (2.3.2)$$

Observando que $\mathbf{\Gamma}_{n+1}$ se representa en bloques como

$$\mathbf{\Gamma}_{n+1} = \begin{bmatrix} \mathbf{\Gamma}_n & \tilde{\gamma} \\ \tilde{\gamma}' & \gamma(0) \end{bmatrix},$$

la igualdad en (2.3.2) se escribe como

$$\begin{bmatrix} \mathbf{\Gamma}_n & \tilde{\gamma} \\ \tilde{\gamma}' & \gamma(0) \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{b}} \\ b_{n+1} \end{bmatrix} = 0, \quad \text{donde } \tilde{\mathbf{b}} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

Ahora se verá que $b_{n+1} \neq 0$. En efecto, si $b_{n+1} = 0$, la ecuación anterior implica que $\mathbf{\Gamma}\tilde{\mathbf{b}} = 0$, y entonces, por la invertibilidad de $\mathbf{\Gamma}_n$, se obtiene que $\tilde{\mathbf{b}} = 0$, lo que combinado con el supuesto de que $b_{n+1} = 0$ implica que $\mathbf{b} = 0$, en contradicción con (2.3.2). Sustituyendo a \mathbf{b} por $b_{n+1}^{-1}\mathbf{b}$ en (2.3.2) se puede suponer que

$$\mathbf{b} = (b_1, b_2, \dots, b_n, 1)'$$

Con esto en mente, note que

$$\left\| \sum_{s=1}^{n+1} b_s X_s \right\|^2 = \sum_{i,j=1}^{n+1} b_i \gamma(i-j) b_j = \mathbf{b}' \mathbf{\Gamma}_{n+1} \mathbf{b} = 0,$$

donde se utilizó la condición $\mathbf{\Gamma}_{n+1}\mathbf{b} = 0$ para establecer la última igualdad. Por lo tanto

$$\left\| \sum_{s=1}^n b_s X_s + X_{n+1} \right\| = 0 \quad (2.3.3)$$

lo cual significa que $X_{n+1} = -\sum_{s=1}^n b_s X_s$, esto, es

$$X_{n+1} \in \mathcal{L}(X_1, X_2, \dots, X_n). \quad (2.3.4)$$

Por otro lado, la estacionaridad del proceso $\{X_t\}$ implica que

$$\left\| \sum_{s=1}^n b_s X_{s+h} + X_{n+h+1} \right\| = 0 \text{ para todo } h,$$

y entonces $\sum_{s=1}^n b_s X_{s+h} + X_{n+h+1} = 0$, es decir $X_{n+h+1} = -\sum_{s=1}^n b_s X_{s+h}$, de donde se desprende que

$$X_{n+h+1} \in \mathcal{L}(X_{h+1}, X_{h+2}, \dots, X_{h+n}), \quad h = 1, 2, 3, \dots$$

Combinando esta relación con (2.3.4), se obtiene que

$$X_{n+h} \in \mathcal{L}(X_1, X_2, \dots, X_n), \quad h = 1, 2, 3, \dots$$

y entonces, para cada $h > 0$, existe un vector $\mathbf{b}^h = (b_1^h, \dots, b_n^h) \in \mathbb{R}^n$ tal que

$$X_{n+h} = \sum_{i=1}^n b_i^h X_i \quad (2.3.5)$$

A partir de esta igualdad se obtendrá una contradicción. Primero observe que debido a que $\mathbf{\Gamma}_n$ es invertible, existen constantes positivas m y M tales

$$m\|\mathbf{x}\|^2 \leq \mathbf{x}'\mathbf{\Gamma}_n\mathbf{x} \leq M\|\mathbf{x}\|^2, \quad \mathbf{x} \in \mathbb{R}^n;$$

vea, por ejemplo, Hoffman y Kunze (1980), Graybill (2001), o Lipschutz (1995). Por otro lado, a partir de (2.3.5), cálculos similares a los realizados en la sección precedente arrojan que $\gamma(0) = \|X_{n+h}\|^2 = \mathbf{b}^{h'}\mathbf{\Gamma}_n\mathbf{b}^h$, relación que al combinarse con las anteriores desigualdades desplegadas implica que

$$m\|\mathbf{b}^h\|^2 \leq \mathbf{b}^{h'}\mathbf{\Gamma}_n\mathbf{b}^h = \gamma(0) \quad \text{y} \quad \gamma(0) = \mathbf{b}^{h'}\mathbf{\Gamma}_n\mathbf{b}^h \leq \|\mathbf{b}^h\|^2.$$

Por lo tanto

$$\frac{\gamma(0)}{M} \leq \|\mathbf{b}^h\|^2 \leq \frac{\gamma(0)}{m}, \quad (2.3.6)$$

esto es, $\{\mathbf{b}^h, h = 1, 2, 3, \dots\}$ es una sucesión acotada. Para concluir, observe que

$$\begin{aligned} \gamma(0) &= \|X_{n+h}\|^2 \\ &= \text{Cov}(X_{n+h}, X_{n+h}) \\ &= \text{Cov}(X_{n+h}, \sum_{i=1}^n b_i^h X_i) \\ &= \sum_{i=1}^n b_i^h \text{Cov}(X_{n+h}, X_i) \\ &= \sum_{i=1}^n b_i^h \gamma(n+h-i) \end{aligned}$$

y entonces (2.3.6) permite establecer que

$$\gamma(0) \leq \sqrt{\frac{\gamma(0)}{m}} \sum_{i=1}^n |\gamma(n+h-i)|.$$

Tomando el límite conforme h tiende a ∞ , la condición $\lim_{k \rightarrow \infty} \gamma(k) = 0$ implica que $\gamma(0) \leq 0$, lo cual contradice el supuesto $\gamma(0) > 0$. Esto muestra que $\mathbf{\Gamma}_n$ es invertible para cada n , finalizando el argumento. \square

2.4. Consecuencia de la Condición de Invertibilidad

En esta sección se estudia una implicación de la invertibilidad de las matrices $\mathbf{\Gamma}_n$ sobre el comportamiento de los errores de pronóstico v_n^h . Como se mencionó en la Sección 2.2, desde un punto de vista intuitivo, parece natural esperar que $v_n^h = \|X_{n+h} - P_{\mathcal{L}_n} X_{n+h}\|$ converja a $\gamma(0)$ conforme h se incrementa, lo cual es un reflejo de la idea de que, a medida que h aumenta, la variable X_{n+h} que se observará en el tiempo $n+h$ tendrá cada vez ‘menos relación’ con las variables X_1, \dots, X_n que se han observado desde el tiempo 1 hasta el tiempo n . Sin embargo, los ejemplos en la Sección 2.2 muestran que en general dicha impresión no es correcta. A continuación se muestra que la convergencia (2.2.2) es válida bajo las condiciones del Teorema 2.3.1.

Teorema 2.4.1. Suponga que $\{X_t\}$ es un proceso estacionario cuya función de autocovarianza $\gamma(\cdot)$ satisface las condiciones

$$\gamma(0) > 0 \quad \text{y} \quad \lim_{k \rightarrow \infty} \gamma(k) = 0. \quad (2.4.1)$$

En este caso,

$$v_n^h = \|X_{n+h} - P_{\mathcal{L}_n} X_{n+h}\| \rightarrow \gamma(0) \quad \text{conforme } h \rightarrow \infty.$$

Demostración. Recuerde que $P_{\mathcal{L}_n} X_{n+h} = \sum_{k=1}^n \phi_{n k}^h X_{n+1-k}$, donde

$$\boldsymbol{\varphi}_n^h = (\phi_{n 1}^h, \phi_{n 2}^h, \dots, \phi_{n n}^h)$$

satisface $\gamma_n^h = \mathbf{\Gamma}_n \varphi_n^h$, y el vector γ_n^h es como en (2.1.13). Como la matriz $\mathbf{\Gamma}_n$ es invertible, por el Teorema 2.3.1, se desprende que $\varphi_n^h = \mathbf{\Gamma}_n^{-1} \gamma_n^h$ y entonces la fórmula (2.1.15) puede escribirse como

$$v_n^h = \gamma(0) - \gamma_n^{h'} \mathbf{\Gamma}_n^{-1} \gamma_n^h;$$

puesto que $\gamma_n^h = (\gamma(h), \gamma(h+1), \dots, \gamma(h+n-1))' \rightarrow 0$ conforme h tiende a ∞ , por (2.4.1), tomando el límite cuando $h \rightarrow \infty$ en la anterior igualdad desplegada, se desprende que $\lim_{h \rightarrow \infty} v_n^h = \gamma(0)$. \square

Por otro lado, el otro resultado que intuitivamente pareció natural en la Sección 2.2 fue la relación de monotonicidad (2.2.1). Observando que la función de autocovarianza en el Ejemplo 2.2.1 satisface la condición (2.4.1), se desprende que en ese contexto las matrices $\mathbf{\Gamma}_n$ son invertibles, y entonces el análisis de dicho ejemplo muestra que (2.2.1) no puede garantizarse aún bajo condiciones que garantizan la invertibilidad de $\mathbf{\Gamma}_n$. Para concluir esta sección, se mostrará que la validez de la relación de monotonicidad depende *de la información disponible* en el tiempo n . Suponga que al momento n se han observado no sólo X_1, X_2, \dots, X_n , sino todas las variables X_t con $t \leq n$ y defina \mathcal{H}_n como la cerradura del espacio generado por las variables X_t con $-\infty < t \leq n$:

$$\mathcal{H}_n = \overline{\mathcal{L}(X_t, t \leq n)}. \quad (2.4.2)$$

Denotando por \tilde{X}_{n+h} a la proyección de X_{n+h} sobre \mathcal{H}_n , esto es,

$$\tilde{X}_{n+h} = P_{\mathcal{H}_n} X_{n+h}$$

se tiene que \tilde{X}_{n+h} es el mejor pronóstico lineal de la observación futura X_{n+h} en términos de la información disponible en el tiempo n representada por \mathcal{H}_n . El correspondiente error cuadrático de pronóstico es

$$\tilde{v}_n^h = \|X_{n+h} - \tilde{X}_{n+h}\|^2 = \|X_{n+h} - P_{\mathcal{H}_n} X_{n+h}\|^2.$$

Teorema 2.4.2. Con la notación precedente, se tiene que

$$\tilde{v}_n^h \leq \tilde{v}_n^{h+1}, \quad h = 1, 2, 3, \dots$$

Demostración. Se presentará un bosquejo del argumento: Observe que por la definición del operador de proyección y la especificación de la cerradura de un subespacio,

$$\begin{aligned} \tilde{v}_n^h &= \|X_{n+h} - \tilde{X}_{n+h}\|^2 \\ &= \min_k \min_{a_1, a_2, \dots, a_k} \left\| X_{n+h} - \sum_{i=1}^k a_i X_{n+1-i} \right\|^2, \end{aligned}$$

y aplicando la estacionaridad de $\{X_n\}$ se desprende que

$$\begin{aligned} \tilde{v}_n^h &= \min_k \min_{a_1, a_2, \dots, a_k} \left\| X_{n+h} - \sum_{i=1}^k a_i X_{n+1-i} \right\|^2 \\ &= \min_k \min_{a_1, a_2, \dots, a_k} \left\| X_h - \sum_{i=1}^k a_i X_{1-i} \right\|^2 \\ &= \|X_h - P_{\mathcal{L}(X_s, s \leq 0)} X_h\|^2 \\ &= \|X_h - P_{\mathcal{H}_0} X_h\|^2; \end{aligned}$$

por lo tanto,

$$\tilde{v}_n^h = \|X_h - P_{\mathcal{H}_0} X_h\|^2 = \|X_h\|^2 - \|P_{\mathcal{H}_0} X_h\|^2. \quad (2.4.3)$$

Procediendo de forma similar,

$$\begin{aligned} \tilde{v}_n^{h+1} &= \|X_{n+h+1} - \tilde{X}_{n+h+1}\|^2 \\ &= \min_k \min_{a_1, a_2, \dots, a_k} \left\| X_{n+h+1} - \sum_{i=1}^k a_i X_{n+1-i} \right\|^2, \end{aligned}$$

y entonces

$$\begin{aligned} \tilde{v}_n^{h+1} &= \min_k \min_{a_1, a_2, \dots, a_k} \left\| X_{n+h+1} - \sum_{i=1}^k a_i X_{n+1-i} \right\|^2 \\ &= \min_k \min_{a_1, a_2, \dots, a_k} \left\| X_h - \sum_{i=1}^k a_i X_{0-i} \right\|^2 \\ &= \|X_h - P_{\mathcal{L}(X_s, s \leq -1)} X_h\|^2 \\ &= \|X_h - P_{\mathcal{H}_{-1}} X_h\|^2. \end{aligned}$$

En consecuencia,

$$\tilde{v}_n^{h+1} = \|X_h - P_{\mathcal{H}_{-1}}X_h\|^2 = \|X_h\|^2 - \|P_{\mathcal{H}_{-1}}X_h\|^2. \quad (2.4.4)$$

Debido a que $\mathcal{H}_{-1} \subset \mathcal{H}_0$, por (2.4.2), se tiene la desigualdad $\|P_{\mathcal{H}_{-1}}X_h\|^2 \leq \|P_{\mathcal{H}_0}X_h\|^2$, relación que al combinarse con (2.4.3) y (2.4.4) implica que $\tilde{v}_n^{h+1} \geq \tilde{v}_n^h$, la cual es la conclusión deseada. \square

Los teoremas de esta sección muestran que, aunque las impresiones intuitivas mencionadas en la Sección 2.2 no son correctas en general, si son válidas bajo condiciones apropiadas o cuando se reformulan adecuadamente.

CAPÍTULO 3

ALGORITMO DE DURBIN-LEVINSON

En este capítulo se formula un procedimiento que permite encontrar los coeficientes en la representación del pronóstico \hat{X}_{n+1} como combinación lineal de X_1, X_2, \dots, X_n sin invertir la matriz $\mathbf{\Gamma}_n$. El esquema de cálculo es recursivo, y el vector de coeficientes φ_k se determina de una forma sencilla en términos de φ_{k-1} . Dicho procedimiento se conoce como algoritmo de Durbin-Levinson, y su aplicación a procesos autorregresivos conduce, naturalmente, a estudiar la idea de *causalidad* de un proceso.

3.1. Propiedades de los Coeficientes

El propósito de esta sección es establecer dos propiedades fundamentales de los coeficientes que permiten expresar la proyección de la variable aleatoria X_{n+1} sobre el espacio generado por las observaciones precedentes X_1, \dots, X_n ; dichas propiedades serán utilizadas para formular un procedimiento recursivo que permite determinar pronósticos sin realizar inversiones matriciales de manera explícita. Considere la expresión para la proyección de X_{n+1} sobre $\mathcal{L}(X_1, \dots, X_n)$:

$$P_{\mathcal{L}_n} X_{n+1} = P_{\mathcal{L}(X_1, \dots, X_n)} X_{n+1} = \sum_{k=1}^n \phi_{nk} X_{n+1-k} \quad (3.1.1)$$

Los resultados que se presentan a continuación representan los instrumentos claves para relacionar los anteriores coeficientes ϕ_{nk} con los números ϕ_{n+1j} que se utilizarían para proyectar X_{n+2} sobre $\mathcal{L}(X_1, \dots, X_{n+1})$. El primer objetivo de esta sección es ver que si se conocen los coeficientes en (3.1.1), entonces puede determinarse inmediatamente la proyección de X_{n+2} sobre $\mathcal{L}(X_2, X_3, \dots, X_{n+1})$.

Lema 3.1.1. La proyección X_{n+2} sobre $\mathcal{L}(X_2, \dots, X_{n+1})$ se determina por medio de los mismos coeficientes ϕ_{nk} que aparecen en (3.1.1). Más precisamente

$$P_{\mathcal{L}(X_2, \dots, X_{n+1})} X_{n+2} = \sum_{k=1}^n \phi_{nk} X_{n+2-k}. \quad (3.1.2)$$

Demostración. Por la caracterización de una proyección, la expresión (3.1.1) significa que

$$\begin{aligned} \gamma(i) &= \langle X_{n+1}, X_{n+1-i} \rangle \\ &= \left\langle \sum_{k=1}^n \phi_{nk} X_{n+1-k}, X_{n+1-i} \right\rangle \\ &= \sum_{k=1}^n \phi_{nk} \gamma(i-k), \quad i = 1, 2, \dots, n \end{aligned} \quad (3.1.3)$$

Por otro lado, la relación (3.1.2) equivale a

$$\langle X_{n+2}, X_{n+2-i} \rangle = \left\langle \sum_{k=1}^n \phi_{nk} X_{n+2-k}, X_{n+2-i} \right\rangle, \quad i = 1, 2, \dots, n$$

El lado izquierdo de esta igualdad es $\gamma(i)$, mientras que el lado derecho se reduce a $\sum_{k=1}^n \phi_{nk} \langle X_{n+2-k}, X_{n+2-i} \rangle = \sum_{k=1}^n \phi_{nk} \gamma(i-k)$; por lo que las anteriores relaciones desplegadas son equivalentes a (3.1.3), y por lo tanto son ciertas. \square

Como ilustración, suponga que se sabe que

$$P_{\mathcal{L}(X_1, X_2, X_3)} X_4 = 5X_1 + 3X_2 - 7X_3.$$

En este caso, el lema establece que $P_{\mathcal{L}(X_2, X_3, X_4)} X_5 = 5X_2 + 3X_3 - 7X_4$, es decir, los coeficientes en ambas expresiones son los mismos y aparecen en el mismo orden.

El siguiente lema muestra que la proyección de X_1 sobre el espacio $\mathcal{L}(X_2, X_3, \dots, X_{n+1})$ generado por X_2, \dots, X_{n+1} puede determinarse cuando se conocen los coeficientes en (3.1.1).

Lema 3.1.2.

- (i) La proyección X_1 sobre $\mathcal{L}(X_2, \dots, X_{n+1})$ se determina por medio de los mismos coeficientes ϕ_{nk} que aparecen en (3.1.1), pero en orden inverso.

Más precisamente

$$P_{\mathcal{L}(X_2, \dots, X_{n+1})}X_1 = \sum_{k=1}^n \phi_{n, n+1-k} X_{n+2-k}, \quad (3.1.4)$$

relación que, haciendo el cambio de variable $j = n + 1 - k$, equivale a

$$P_{\mathcal{L}(X_2, \dots, X_{n+1})}X_1 = \sum_{j=1}^n \phi_{nj} X_{j+1}. \quad (3.1.5)$$

- (ii) $\|X_1 - P_{\mathcal{L}(X_2, \dots, X_{n+1})}X_1\|^2 = \|X_{n+1} - P_{\mathcal{L}(X_1, \dots, X_n)}X_{n+1}\|^2 = v_n$

Demostración. Por la caracterización de una proyección, la expresión (3.1.4) significa que, para cada $i = 1, 2, \dots, n$

$$\begin{aligned} \gamma(i) &= \langle X_1, X_{i+1} \rangle \\ &= \left\langle \sum_{k=1}^n \phi_{n, n+1-k} X_{n+2-k}, X_{i+1} \right\rangle. \\ &= \sum_{k=1}^n \phi_{n, n+1-k} \gamma(n+1-i-k). \end{aligned} \quad (3.1.6)$$

Usando el cambio de variable $j = n + 1 - k$ y notando que cuando k varía entre 1 y n , j también lo hace, las anteriores relaciones equivalen a

$$\gamma(i) = \sum_{j=1}^n \phi_{nj} \gamma(j-i), \quad i = 1, 2, \dots, n$$

ecuaciones que son válidas pues, como se verificó en el lema precedente, representan la caracterización de $P_{\mathcal{L}(X_1, \dots, X_n)}X_{n+1}$, estableciendo la conclusión deseada.

(ii) Observe que

$$\begin{aligned}
\|X_1 - P_{\mathcal{L}(X_2, \dots, X_{n+1})} X_1\|^2 &= \|X_1\|^2 - \|P_{\mathcal{L}(X_2, \dots, X_{n+1})} X_1\|^2 \\
&= \gamma(0) - \left\| \sum_{j=1}^n \phi_{n j} X_{j+1} \right\|^2 \\
&= \gamma(0) - \sum_{i,j=1}^n \phi_{n i} \langle X_{i+1}, X_{j+1} \rangle \phi_{n j} \\
&= \gamma(0) - \sum_{i,j=1}^n \phi_{n i} \gamma(i-j) \phi_{n j} \\
&= \gamma(0) - \boldsymbol{\varphi}'_n \boldsymbol{\Gamma}_n \boldsymbol{\varphi}_n = \gamma(0) - \boldsymbol{\varphi}'_n \boldsymbol{\gamma}_n = v_n
\end{aligned}$$

vea (2.1.11). □

Para ejemplificar este resultado suponga, como antes, que

$$P_{\mathcal{L}(X_1, X_2, X_3)} X_4 = 5X_3 + 3X_2 - 7X_1.$$

En este caso el lema establece que $P_{\mathcal{L}(X_2, X_3, X_4)} X_1 = -7X_2 + 3X_3 + 5X_4$; así, los coeficientes invierten su papel.

3.2. Teorema de Recurrencia

Los resultados preliminares establecidos en la sección precedente serán ahora utilizados para analizar la relación entre los coeficientes de dos proyecciones ‘sucesivas’:

$$\begin{aligned}
P_{\mathcal{L}(X_1, X_2, \dots, X_{n-1})} X_n &= \sum_{i=1}^{n-1} \phi_{n-1 i} X_{n-i} \\
P_{\mathcal{L}(X_1, X_2, \dots, X_n)} X_{n+1} &= \sum_{k=1}^n \phi_{n k} X_{n+1-k}
\end{aligned} \tag{3.2.1}$$

El siguiente teorema establece la forma en que los coeficientes en ambas igualdades están relacionados, y representa el resultado clave para diseñar un procedimiento recursivo de cálculo de pronósticos.

Teorema 3.2.1. Suponga que Γ_n es una matriz invertible para cada entero positivo.

En este caso, las siguientes afirmaciones (i) y (ii) son válidas:

(i) La proyección $P_{\mathcal{L}(X_1, X_2, \dots, X_n)} X_{n+1}$ se expresa como

$$\begin{aligned} & P_{\mathcal{L}(X_1, X_2, \dots, X_n)} X_{n+1} \\ &= \sum_{k=1}^{n-1} [\phi_{nk} + \phi_{nn} \phi_{n-1, n-k}] X_{n+1-k} + \phi_{nn} [X_1 - P_{\mathcal{L}(X_2, \dots, X_n)} X_1]. \end{aligned} \quad (3.2.2)$$

Por lo tanto,

- (ii) $[\phi_{nk} + \phi_{nn} \phi_{n-1, n-k}] = \phi_{n-1, k}, \quad k = 1, 2, \dots, n-1.$
(ii) $\|P_{\mathcal{L}(X_1, X_2, \dots, X_n)} X_{n+1}\|^2 = \|P_{\mathcal{L}(X_1, X_2, \dots, X_{n-1})} X_n\|^2 + \phi_{nn}^2 v_{n-1}$

Demostración.

(i) Iniciando con la segunda igualdad en (3.2.1) se tiene que

$$\begin{aligned} & P_{\mathcal{L}(X_1, X_2, \dots, X_n)} X_{n+1} \\ &= \sum_{k=1}^{n-1} \phi_{nk} X_{n+1-k} + \phi_{nn} X_1 \\ &= \left[\sum_{k=1}^{n-1} \phi_{nk} X_{n+1-k} + \phi_{nn} P_{\mathcal{L}(X_2, \dots, X_n)} X_1 \right] \\ &\quad + \phi_{nn} [X_1 - P_{\mathcal{L}(X_2, \dots, X_n)} X_1]. \end{aligned} \quad (3.2.3)$$

Ahora, aplicando la expresión (3.1.5) con $n-1$ en lugar de n se obtiene que

$$P_{\mathcal{L}(X_2, \dots, X_n)} X_1 = \sum_{k=1}^{n-1} \phi_{n-1, k} X_{k+1} = \sum_{j=1}^{n-1} \phi_{n-1, n-j} X_{n+1-j};$$

luego,

$$\begin{aligned} & \left[\sum_{k=1}^{n-1} \phi_{nk} X_{n+1-k} + \phi_{nn} P_{\mathcal{L}(X_2, \dots, X_n)} X_1 \right] \\ &= \left[\sum_{k=1}^{n-1} \phi_{nk} X_{n+1-k} + \phi_{nn} \sum_{j=1}^{n-1} \phi_{n-1, n-j} X_{n+1-j} \right] \\ &= \sum_{k=1}^{n-1} [\phi_{nk} + \phi_{nn} \phi_{n-1, n-k}] X_{n+1-k} \end{aligned}$$

y (3.2.2) se desprende combinando esta igualdad con (3.2.3)

- (ii) Note que la sumatoria en (3.2.2) pertenece al espacio $\mathcal{L}(X_2, \dots, X_n)$ y que la variable aleatoria $X_1 - P_{\mathcal{L}(X_2, \dots, X_n)}X_1$ es ortogonal a dicho espacio, de manera que, aplicando el operador de proyección $P_{\mathcal{L}(X_2, \dots, X_n)}$ en ambos lados de (3.2.2), se desprende que

$$\begin{aligned} P_{\mathcal{L}(X_2, \dots, X_n)}X_{n+1} &= P_{\mathcal{L}(X_2, \dots, X_n)}P_{\mathcal{L}(X_1, X_2, \dots, X_n)}X_{n+1} \\ &= \sum_{k=1}^{n-1} [\phi_{n k} + \phi_{n n} \phi_{n-1 n-k}] X_{n+1-k} \end{aligned}$$

Observe ahora que

$$P_{\mathcal{L}(X_2, \dots, X_n)}X_{n+1} = \sum_{i=1}^{n-1} \phi_{n-1 i} X_{n+1-i},$$

por el Lema 3.1.1 aplicado con $n-1$ en lugar de n . Utilizando el supuesto de que $\mathbf{\Gamma}_n$ es invertible, se desprende que las variables X_2, \dots, X_{n+1} son linealmente independientes en L^2 , y entonces, las dos igualdades anteriormente desplegadas implican que

$$[\phi_{n k} + \phi_{n n} \phi_{n-1 n-k}] = \phi_{n-1 k}, \quad k = 1, 2, \dots, n-1$$

- (iii) Combinando las partes (i) y (ii) se obtiene

$$P_{\mathcal{L}(X_1, X_2, \dots, X_n)}X_{n+1} = \sum_{k=1}^{n-1} \phi_{n-1 k} X_{n+1-k} + \phi_{n n} [X_1 - P_{\mathcal{L}(X_2, \dots, X_n)}X_1]$$

donde la sumatoria pertenece a $\mathcal{L}(X_2, \dots, X_n)$ y el término entre corchetes es ortogonal a este espacio. Por lo tanto,

$$\begin{aligned} \|P_{\mathcal{L}(X_1, X_2, \dots, X_n)}X_{n+1}\|^2 &= \left\| \sum_{k=1}^{n-1} \phi_{n-1 k} X_{n+1-k} \right\|^2 \\ &\quad + \phi_{n n}^2 \|X_1 - P_{\mathcal{L}(X_2, \dots, X_n)}X_1\|^2 \end{aligned}$$

Note ahora que, por la estacionaridad del proceso $\{X_t\}$,

$$\left\| \sum_{k=1}^{n-1} \phi_{n-1 k} X_{n+1-k} \right\|^2 = \left\| \sum_{k=1}^{n-1} \phi_{n-1 k} X_{n-k} \right\|^2 = \|P_{\mathcal{L}(X_1, \dots, X_{n-1})}X_n\|^2$$

mientras que

$$\|X_1 - P_{\mathcal{L}(X_2, \dots, X_n)} X_1\|^2 = v_{n-1},$$

por el Lema 3.1.2(ii) aplicado con $n-1$ en lugar de n . Ahora, la conclusión se sigue combinando las tres últimas relaciones desplegadas. \square

3.3. El Algoritmo de Durbin-Levinson

En esta sección se formula un procedimiento recursivo para calcular los coeficientes ϕ_{nk} en la expresión para la proyección de X_{n+1} sobre $\mathcal{L}(X_1, \dots, X_n)$. Este procedimiento calcula el vector de coeficientes φ_n y el error de pronóstico v_n en términos del vector de coeficientes φ_{n-1} y el correspondiente error de v_{n-1} , evitando el problema de invertir la matriz $\mathbf{\Gamma}_n$. El esquema de cálculo en el siguiente teorema se conoce como el algoritmo de Durbin-Levinson. En el resto del capítulo, se utiliza la siguiente notación:

$$\begin{aligned} \hat{X}_1 &= 0, & v_0 &= \|X_1 - \hat{X}_1\|^2 = \gamma(0) \\ \hat{X}_n &= P_{\mathcal{L}(X_1, \dots, X_{n-1})} X_n, & v_{n-1} &= \|X_n - \hat{X}_n\|^2, \quad n = 2, 3, \dots \end{aligned} \quad (3.3.1)$$

Por conveniencia, a continuación se reescribe la expresión para \hat{X}_n :

$$\hat{X}_n = \sum_{k=1}^{n-1} \phi_{n-1,k} X_{n-k}. \quad (3.3.2)$$

Teorema 3.3.1. Suponga que $\{X_t\}$ es un proceso estacionario con función de autocovarianza $\gamma(\cdot)$ tal que $\gamma(0) > 0$ y $\gamma(h) \rightarrow 0$ conforme $h \rightarrow \infty$, de tal forma que las matrices $\mathbf{\Gamma}_n$ son no singulares. En este caso, los coeficientes ϕ_{nk} y los errores cuadráticos de pronóstico satisfacen las siguientes relaciones:

$$(i) \quad \phi_{11} = \frac{\gamma(1)}{\gamma(0)}, \quad v_0 = \gamma(0)$$

$$(ii) \quad \phi_{nn} = \frac{1}{v_{n-1}} \left[\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1j} \gamma(n-j) \right], \quad v_n = v_{n-1}[1 - \phi_{nn}^2]$$

$$(iii) \quad \phi_{nk} = \phi_{n-1k} - \phi_{nn}\phi_{n-1n-k}, \quad k = 1, 2, \dots, n-1$$

Demostración.

(i) La proyección $\hat{X}_2 = P_{\mathcal{L}(X_1)}X_2 = \phi_{11}X_1$ está caracterizada por $\langle \hat{X}_2, X_1 \rangle = \langle X_2, X_1 \rangle$, esto es $\langle \phi_{11}X_1, X_1 \rangle = \langle X_2, X_1 \rangle$, relación que es equivalente a $\phi_{11}\gamma(0) = \gamma(1)$, de donde se desprende que $\phi_{11} = \gamma(1)/\gamma(0)$. Por otro lado, la igualdad $v_0 = \gamma(0)$ es consecuencia de la convención $\hat{X}_1 = 0$; vea (3.3.1).

(ii) Se usará la igualdad

$$\begin{aligned} \hat{X}_{n+1} &= P_{\mathcal{L}(X_1, X_2, \dots, X_n)}X_{n+1} \\ &= \sum_{k=1}^{n-1} [\phi_{nk} + \phi_{nn}\phi_{n-1n-k}]X_{n+1-k} + \phi_{nn}[X_1 - P_{\mathcal{L}(X_2, \dots, X_n)}X_1] \end{aligned}$$

demostrada en el Teorema 3.2.1(ii); observando que la sumatoria pertenece a $\mathcal{L}(X_2, \dots, X_n)$ y que este espacio es ortogonal al vector $X_1 - P_{\mathcal{L}(X_2, \dots, X_n)}X_1$, después de tomar el producto interno por esta última diferencia en ambos lados de la anterior relación se obtiene

$$\begin{aligned} &\langle \hat{X}_{n+1}, X_1 - P_{\mathcal{L}(X_2, \dots, X_n)}X_1 \rangle \\ &= \phi_{nn} \langle X_1 - P_{\mathcal{L}(X_2, \dots, X_n)}X_1, X_1 - P_{\mathcal{L}(X_2, \dots, X_n)}X_1 \rangle \end{aligned} \tag{3.3.3}$$

Por otro lado, puesto que \hat{X}_{n+1} satisface $\langle \hat{X}_{n+1}, \mathbf{w} \rangle = \langle X_{n+1}, \mathbf{w} \rangle$ para

todo $\mathbf{w} \in \mathcal{L}(X_1, \dots, X_n)$, se tiene

$$\begin{aligned}
\langle \hat{X}_{n+1}, X_1 - P_{\mathcal{L}(X_2, \dots, X_n)} X_1 \rangle &= \langle X_{n+1}, X_1 - P_{\mathcal{L}(X_2, \dots, X_n)} X_1 \rangle \\
&= \gamma(n) - \langle X_{n+1}, P_{\mathcal{L}(X_2, \dots, X_n)} X_1 \rangle \\
&= \gamma(n) - \left\langle X_{n+1}, \sum_{j=1}^{n-1} \phi_{n-1j} X_{j+1} \right\rangle \\
&= \gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1j} \langle X_{n+1}, X_{j+1} \rangle \\
&= \gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1j} \gamma(n-j)
\end{aligned}$$

donde para establecer la tercera igualdad se usó la expresión (3.1.5) obtenida en el Lema 3.1.1. Utilizando ahora el Lema 3.1.2(ii) con $n-1$ en lugar de n se obtiene

$$\begin{aligned}
\langle X_1 - P_{\mathcal{L}(X_2, \dots, X_n)} X_1, X_1 - P_{\mathcal{L}(X_2, \dots, X_n)} X_1 \rangle \\
= \|X_1 - P_{\mathcal{L}(X_2, \dots, X_n)} X_1\|^2 = v_{n-1}
\end{aligned}$$

Combinando las dos últimas relaciones desplegadas con (3.3.3) se obtiene

$$\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1j} \gamma(n-j) = \phi_{nn} v_{n-1}$$

de donde se desprende la expresión deseada para ϕ_{nn} . La fórmula para v_n se obtiene como sigue: Recuerde la expresión

$$\|P_{\mathcal{L}(X_1, X_2, \dots, X_n)} X_{n+1}\|^2 = \|P_{\mathcal{L}(X_1, X_2, \dots, X_{n-1})} X_n\|^2 + \phi_{nn}^2 v_{n-1}$$

obtenida en el Teorema 3.2.1(iii), la cual, conjuntamente con (2.1.11) implica que

$$\begin{aligned}
v_n &= \gamma(0) - \|P_{\mathcal{L}(X_1, X_2, \dots, X_n)} X_{n+1}\|^2 \\
&= \gamma(0) - \|P_{\mathcal{L}(X_1, X_2, \dots, X_{n-1})} X_n\|^2 - \phi_{nn}^2 v_{n-1} \\
&= v_{n-1} - \phi_{nn}^2 v_{n-1} = (1 - \phi_{nn}^2) v_{n-1}
\end{aligned}$$

(iii) Esta parte fue establecida en el Teorema 3.2.1 (ii). \square

Este resultado permite encontrar los coeficientes ϕ_{nk} en la expresión

$$\hat{X}_{n+1} = \sum_{j=1}^n \phi_{nj} X_{n+1-j}$$

así como el error de pronóstico $v_n = \|X_{n+1} - \hat{X}_{n+1}\|^2$, de manera recursiva, sin invertir la matriz Γ_n . El procedimiento se describe como sigue:

Algoritmo de Durbin-Levinson

Entrada: Un vector $(\gamma(0), \gamma(1), \dots, \gamma(n))$, donde $\gamma(\cdot)$ es una función de autocovarianza.

Salida: El vector $\varphi_n = (\phi_{n1}, \dots, \phi_{nn})$ de coeficientes de la proyección \hat{X}_{n+1} , así como v_n , el correspondiente error cuadrático de pronóstico.

- Inicio: Defina $\phi_{1,1} = \gamma(1)/\gamma(0)$ y $v_1 = (1 - \phi_{1,1}^2)v_0$ y ponga $k = 1$
- Paso Recursivo: En términos de ϕ_{kj} , $j = 1, 2, \dots, k$ y v_k , calcule

$$(a) \quad \phi_{k+1, k+1} = v_k^{-1} \left[\gamma(k+1) - \sum_{j=1}^k \phi_{kj} \gamma(k+1-j) \right]$$

$$(b) \quad v_{k+1} = v_k [1 - \phi_{k+1, k+1}^2]$$

$$(c) \quad \phi_{k+1, j} = \phi_{kj} - \phi_{k+1, k+1} \phi_{k, n-j}, \quad j = 1, 2, \dots, k.$$

Además

$$(d) \quad \text{Incremente } k \text{ en una unidad.}$$

- Paso de Prueba: Si $k = n$ detenga el procedimiento.

Los números actuales v_k y ϕ_{kj} , $j = 1, 2, \dots, k$ son las cantidades deseadas.

Si $k < n$, vaya al paso anterior.

Este algoritmo es muy simple y, como se verá a continuación, admite un código sencillo

3.4. Implementación del Procedimiento

En esta sección se presenta una implementación del algoritmo de Durbin-Levinson en el lenguaje *R*. Con esta finalidad, se diseñó una función que lleva a cabo los cálculos, denominada `DurLev`. Dicha función acepta como parámetro a un vector que consiste de $\gamma(0), \gamma(1), \dots, \gamma(n)$, el cual, por comodidad, se descompuso en dos partes: `Gammazero`, en donde se coloca a $\gamma(0)$, y `Gamma` que acepta al vector $(\gamma(1), \dots, \gamma(n))$; esta descomposición está motivada por el hecho de que, en *R*, los índices de los vectores inician en 1. Para propósitos de ilustración, la función no sólo devuelve el vector φ_n y al error cuadrático de pronóstico v_n , sino que también conserva y devuelve a todos los vectores φ_k con $k < n$ y, desde luego, al correspondiente vector completo de errores (v_1, \dots, v_n) . Para almacenar todos estos vectores de longitud distinta, se usa una lista, denominada `phis` en la función. Después de leer los datos `Gammazero` y `Gamma`, la función calcula la longitud de `Gamma`, cuyo valor es n , el número de iteraciones que se llevarán a cabo, y dimensiona a `phis` como una lista de longitud $n + 1$, así como al vector `v` como uno de tipo numérico. Posteriormente se lleva a cabo la etapa de inicio, seguida por un ciclo `for` que efectúa la etapa iterativa. Al salir del ciclo, la componente $n + 1$ de la lista `phis` recibe al vector `v` de errores de pronóstico y la función devuelve la lista `phis`. Finalmente, note que se han incluido valores por defecto para `Gammazero` y `Gamma`, los cuales son 1.81 para `Gamazero` y el vector $(-.9, 0, 0, 0, 0, 0)$ para `Gamma`, los cuales corresponden a la función de autocovarianza del proceso $X_t = Z_t - .9Z_{t-1}$, donde $\{Z_t\}$ es un ruido blanco de varianza unitaria; este ejemplo

se utiliza en (Brockwell y Davis, 1998, p. 174). Por la presencia de valores por defecto para los parámetros, la función puede llamarse simplemente como

```
DurLev()
```

y esto será equivalente a hacerlo como

```
DurLev(1.81, c(-.9, 0, 0, 0, 0, 0))
```

El código fuente de la función es como sigue:

```
DurLev <- function(Gammazero= 1.81,
                   Gamma = c(-.9, 0, 0, 0, 0, 0)) {
  n <- length(Gamma)
  phis <- vector("list", n+1)
  v<-vector("numeric",n)
  phis[[1]] <- Gamma[1]/Gammazero
  v[1] <- Gammazero - Gamma[1]*Gamma[1]/Gammazero
  for(k in 2: n) {
    aux <- (Gamma[k]-sum(phis[[k-1]]*Gamma[(k-1):1]))/v[k-1]
    phis[[k]] <- phis[[k-1]] - aux*(phis[[k-1]][(k - 1):1])
    phis[[k]] <- c(phis[[k]], aux)
    v[k] <- v[k-1]*(1-aux*aux)
  }
  phis[[n+1]]<- v
  phis
}
```

Una llamada a la función guardando el resultado en un vector para uso futuro es como sigue:

```
> DurLev()-> phis
```

```

> phis
[[1]] [1] -0.4972376
[[2]] [1] -0.6605572 -0.3284538
[[3]] [1] -0.7404369 -0.4891009 -0.2431993
[[4]] [1] -0.7869838 -0.5827118 -0.3849145 -0.1913939
[[5]] [1] -0.8169085 -0.6428937 -0.4760223 -0.3144399
-0.1563513
[[6]] [1] -0.8373788 -0.6840619 -0.5383456 -0.3986110
-0.2633053 -0.1309253
[[7]] [1] 1.362486 1.215499 1.143607 1.101715 1.074782
1.056359

```

Los errores de pronóstico aparecen como el elemento $7 = 6+1$ de la lista, y son $v_1 = 1.362486$, $v_2 = 1.215499, \dots, v_6 = 1.056359$. El elemento 6 de la lista es el vector φ_6 , de modo que $\phi_{6,1} = -0.8373788$, $\phi_{6,2} = -0.6840619, \dots, \phi_{6,6} = -0.1309253$.

La salida de la función `DurLev` es difícil de interpretar si no se conoce la estructura de las componentes de la lista que la función devuelve. Para facilitar la lectura de la salida, se diseñó una función que da formato adecuado a la lista de salida, coloca los resultados en forma tabular, y pone encabezados que facilitan la interpretación. La función que organiza la presentación de resultados se llama `outputDurLev` y acepta como argumento a la lista generada por la función `DurLev`. La nueva función de formato utiliza la función `cat` de R para dirigir la salida a un archivo, en el cual se escriben los datos con el formato adecuado para insertar en el código fuente de una archivo escrito en PLAIN T_EX, redondeando los números tres cifras significativas. El archivo que produce la función fue insertado sin cambio alguno al presente texto, y el resultado es el siguiente:

Algoritmo de Durbin Levinson

$$\phi_{nk} = 0 \text{ para } k > n$$

n	v_n	ϕ_{n1}	ϕ_{n2}	ϕ_{n3}	ϕ_{n4}	ϕ_{n5}	ϕ_{n6}
1	1.362	-0.497					
2	1.215	-0.661	-0.328				
3	1.144	-0.740	-0.489	-0.243			
4	1.102	-0.787	-0.583	-0.385	-0.191		
5	1.075	-0.817	-0.643	-0.476	-0.314	-0.156	
6	1.056	-0.837	-0.684	-0.538	-0.399	-0.263	-0.131

Es claro que, con esta presentación, los resultados pueden interpretarse mucho más fácilmente.

3.5. El Orden del Algoritmo

En esta sección se discute brevemente la carga de operaciones requerida por el algoritmo de Durbin-Levinson, prestando atención al número de multiplicaciones y divisiones, así como a los requerimientos de memoria del procedimiento.

- En la etapa inicial se realizan una división para evaluar ϕ_{11} y dos multiplicaciones para calcular v_1 ; en total, 3 operaciones.
- En la etapa recursiva, para pasar de φ_k y v_k hacia φ_{k+1} y v_{k+1} se llevan a cabo
 - (a) k multiplicaciones y una división para evaluar $\phi_{k+1,k+1}$;
 - (b) dos multiplicaciones para calcular v_{k+1} , y
 - (c) k multiplicaciones para determinar las restantes componentes $\phi_{k+1,j}$, $j = 1, 2, \dots, k$.

En total se realizan $2k + 3$ multiplicaciones o divisiones.

Como la etapa iterativa se repite para $k = 1, 2, \dots, n-1$, todo el algoritmo requiere la realización de $3 + \sum_{k=1}^{n-1} (2k+3) = n(n+2)$ multiplicaciones o divisiones.

En este caso, se dice que el algoritmo de Durbin-Levinson es *de orden* n^2 , y se escribe $O(n^2)$. En cuanto a los requerimientos de memoria, se necesitan $n + 1$ lugares para almacenar $\gamma(0), \gamma(1), \dots, \gamma(n)$, n lugares adicionales para guardar v_1, \dots, v_n , y otros n lugares para almacenar $\phi_{nk}, k = 1, 2, \dots, n$, totalizando $3n + 1$ posiciones de memoria para los datos de entrada y salida. En esta descripción, la palabra ‘lugar’ se refiere a la cantidad de bytes necesaria para almacenar un número real con precisión determinada, y dicha cantidad depende del entorno de trabajo y del programa que ejecuta la tarea. Para propósitos de referencia futura, esta discusión se resume en el siguiente teorema.

Teorema 3.5.1.

- (i) Con respecto a las operaciones aritméticas que realiza, el algoritmo de Durbin-Levinson es *de orden* $O(n^2)$.
- (ii) El requerimiento de memoria del algoritmo de Durbin-Levinson es de orden $O(n)$; si se desea almacenar todos los coeficientes $\phi_{s,j}, 1 \leq j \leq s \leq n$ (como lo hace la función `DurLev` descrita anteriormente) el requerimiento de memoria naturalmente aumenta, y es de orden $O(n^2)$.

En la siguiente sección se estudia el algoritmo de Durbin-Levinson aplicado a la clase procesos autorregresivos.

3.6. EL Caso de Procesos Autorregresivos

En esta sección se estudia el comportamiento del procedimiento de Durbin-Levinson cuando se aplica a un proceso autorregresivo. La idea es observar el comportamiento de los coeficientes ϕ_{nk} , para finalmente establecer una conclusión general al final de la sección. Recuerde que un proceso autorregresivo de orden p —brevemente, un proceso $\text{AR}(p)$ — satisface una ecuación de la forma

$$X_t - \alpha_1 X_{t-1} - \dots - \alpha_p X_{t-p} = Z_t, \quad (3.6.1)$$

donde $\{Z_t\}$ es un ruido blanco con varianza $\sigma^2 > 0$, y $\alpha_1, \dots, \alpha_p$ son constantes dadas. Como se verá en los ejemplos que aparecen a continuación, los coeficientes de la proyección \hat{X}_{n+1} sobre el espacio generado por X_1, \dots, X_n se ‘estabilizan rápidamente’, esto es a partir de cierto n_0 , el coeficiente ϕ_{nk} ya no cambia conforme n aumenta; el propósito es demostrar que este comportamiento siempre se presentará para un proceso $\text{AR}(p)$, y caracterizar el número α_k^* para el cual $\phi_{nk} = \alpha_k^*$ a partir de cierto n en adelante. Como se verá al final de la sección, el valor α_k^* tiene que ver con la *representación causal* de un proceso; para detalles sobre la idea de causalidad, vea, por ejemplo, Chacón (2010).

Ejemplo 3.6.1.

- (i) Considere el proceso $\text{AR}(1)$

$$X_t + 2X_{t-1} = Z_t$$

donde $\{Z_t\}$ es un ruido blanco de varianza $\sigma^2 = 1$; esto corresponde al caso $p = 1$ y $\alpha_1 = -2$ en (3.6.1). La correspondiente función de autocovarianza es

$$\gamma(h) = \frac{1}{3}(-2)^{-h}, \quad h = 0, \pm 1, \pm 2, \dots;$$

vea, por ejemplo, Chacón (2010) o Vázquez (2010). Ahora se calcularán los primeros seis vectores φ_n . Para obtenerlos, se invoca a la función `DurLev` como sigue:

```
DurLev(1/3, ((-2)^(-(1:6)))/3) -> phis
```

La salida que se muestra a continuación se obtuvo por medio de la función `outputDurLev` usando

```
outputDurLev(phis)
```

e insertando directamente el archivo que produce:

Algoritmo de Durbin Levinson

$$\phi_{nk} = 0 \text{ para } k > n$$

n	v_n	ϕ_{n1}	ϕ_{n2}	ϕ_{n3}	ϕ_{n4}	ϕ_{n5}	ϕ_{n6}
1	0.250	-0.500					
2	0.250	-0.500	0.000				
3	0.250	-0.500	0.000	0.000			
4	0.250	-0.500	0.000	0.000	0.000		
5	0.250	-0.500	0.000	0.000	0.000	0.000	
6	0.250	-0.500	0.000	0.000	0.000	0.000	0.000

Un rasgo notable en esta tabla es la ‘estabilidad’ de los coeficientes ϕ_{nk} así como de los errores cuadráticos de pronóstico v_n .

(ii) Ahora considere el proceso autorregresivo de orden 2 determinado por

$$X_t - X_{t-1} + .21X_{t-2} = Z_t$$

donde el ruido blanco $\{Z_t\}$ tiene varianza 1. Usando un procedimiento estándar para determinar la función de autocovarianza $\gamma(\cdot)$, se tiene que

$$\gamma(h) = (-1.043261)(.3)^{|h|} + (4.345100)(.7)^{|h|}$$

de manera que $\gamma(0) = 3.301939$. La función DurLev fue aplicada como

```
DurLev(3.301939, (-1.043261)*(.3)^(-(1:6))
+ (4.345100)*((.7)^(-(1:6))) ) -> phis
```

y la salida fue formateada usando `outputDurLev(phis)`. El resultado es el siguiente:

Algoritmo de Durbin Levinson

$$\phi_{nk} = 0 \text{ para } k > n$$

n	v_n	ϕ_{n1}	ϕ_{n2}	ϕ_{n3}	ϕ_{n4}	ϕ_{n5}	ϕ_{n6}
1	1.047	0.826					
2	1.001	1.000	-0.210				
3	1.001	1.000	-0.210	-0.000			
4	1.001	1.000	-0.210	-0.000	-0.000		
5	1.001	1.000	-0.210	-0.000	-0.000		
6	1.001	1.000	-0.210	-0.000	-0.000	-0.000	-0.000

Como en el caso anterior, los coeficientes ϕ_{nk} y los errores cuadráticos de pronóstico ‘se estabilizan’. \square

A continuación se verá que el fenómeno observado en el ejemplo anterior es general para un proceso autorregresivo. Para establecer el resultado en esa dirección, es necesario introducir la idea de *causalidad*, la cual se analiza en detalle en (Chacón, 2010). Considere el proceso (3.6.1). El polinomio autorregresivo $\alpha(z)$ asociado a esa representación de $\{X_t\}$ se define como

$$\alpha(z) = 1 - \alpha_1 z - \dots - \alpha_p z^p.$$

En este caso, debe tenerse que $\alpha(z)$ no tiene raíces en el círculo unitario del plano complejo, es decir,

$$\alpha(z) \neq 0, \quad |z| = 1;$$

pues en otro caso no existe un proceso $\{X_t\}$ que satisfaga (3.6.1); vea, por ejemplo (Vázquez, 2010). Si $\alpha(z) \neq 0$ para todo número complejo z tal que $|z| \leq 1$ el polinomio $\alpha(z)$ se denomina causal; así, $\alpha(z)$ es causal cuando todas sus raíces tienen módulo mayor a 1. Para el caso del proceso AR(1) en el Ejemplo 3.6.1(i), $\alpha(z) = 1 + 2z$ tiene raíz $z = -1/2$, y por lo tanto este polinomio no es causal. En el Ejemplo 3.6.1(ii) se tiene $\alpha(z) = 1 - z + .21z^2 = (1 - .3z)(1 - .7z)$, y las

raíces son $1/.3 = 10/3$ y $1/.7 = 10/7$. Como estas raíces son mayores que 1, el polinomio $\alpha(z)$ es causal en este caso. La importancia de la idea de polinomio causal en esta discusión sobre la estabilidad de los coeficientes generados por el algoritmo de Durbin-Levinson, se debe a los siguientes resultados, que se desprenden directamente, por ejemplo, de los análisis en Chacón (2010).

Lema 3.6.1. Considere el proceso autorregresivo $\{X_t\}$ determinado por

$$X_t - \alpha_1 X_{t-1} - \cdots - \alpha_p X_{t-p} = Z_t,$$

donde $\{Z_t\}$ es un ruido blanco con varianza σ^2 . Si el polinomio $\alpha(z)$ es causal, entonces

$$\langle X_s, Z_t \rangle = 0 \quad \text{para cada } s < t.$$

Debido a que la relación de ortogonalidad es importante al construir proyecciones, el lema precedente es crucial en el estudio de pronósticos para un proceso autorregresivo. EL siguiente lema establece que todo proceso $\text{AR}(p)$, tiene una *representación causal*.

Lema 3.6.2. Si $\{X_t\}$ es un proceso autorregresivo que satisface

$$X_t - \alpha_1 X_{t-1} - \cdots - \alpha_p X_{t-p} = Z_t,$$

donde $\{Z_t\}$ es un ruido blanco con varianza σ^2 , entonces, existe un único polinomio causal

$$\tilde{\alpha}(z) = 1 - \tilde{\alpha}_1 z - \cdots - \tilde{\alpha}_p z^p$$

y un ruido blanco $\{\tilde{Z}_t\}$ con varianza $\tilde{\sigma}^2$ tal

$$X_t - \tilde{\alpha}_1 X_{t-1} - \cdots - \tilde{\alpha}_p X_{t-p} = \tilde{Z}_t. \quad (3.6.2)$$

Combinando los dos lemas precedentes, es posible establecer que, al aplicar el algoritmo de Durbin-Levinson a la función de autocovarianza de un proceso

autorregresivo, los coeficientes ϕ_{nk} y los errores cuadráticos de pronóstico v_n , ‘se estabilizarán’, y determinar los límites conforme n , el número de datos observados, aumenta. El resultado en esta dirección es el siguiente.

Teorema 3.6.1. Considere un proceso autorregresivo $\{X_t\}$ de orden p , y sea

$$X_t - \tilde{\alpha}_1 X_{t-1} - \cdots - \tilde{\alpha}_p X_{t-p} = \tilde{Z}_t, \quad \{\tilde{Z}_t\} \sim WN(0, \tilde{\sigma}^2). \quad (3.6.3)$$

la representación causal de $\{X_t\}$ cuya existencia está garantizada por el Lema 3.6.2. En este caso, para cada $n > p$,

(i)

$$P_{\mathcal{L}(X_1, \dots, X_n)} X_{n+1} = \tilde{\alpha}_1 X_n + \tilde{\alpha}_2 X_{n-1} + \cdots + \tilde{\alpha}_p X_{n-p}$$

Por lo tanto,

(ii) Si $n > p$,

$$v_n = \tilde{\sigma}^2, \quad \phi_{nk} = \tilde{\alpha}_k, \quad k = 1, 2, \dots, p, \quad \phi_{nk} = 0, \quad p < k \leq n.$$

Demostración.

(i) Sea $n > p$ un entero arbitrario. En este caso, $n - p \geq 1$, y se sigue que

$$(a) \quad \tilde{\alpha}_1 X_n + \tilde{\alpha}_2 X_{n-1} + \cdots + \tilde{\alpha}_p X_{n-p} \in \mathcal{L}(X_1, X_2, \dots, X_n)$$

Observe ahora que $X_{n+1} - [\tilde{\alpha}_1 X_n + \tilde{\alpha}_2 X_{n-1} + \cdots + \tilde{\alpha}_p X_{n-p}] = \tilde{Z}_{n+1}$, por (3.6.3), de manera que

$$\left\langle X_{n+1} - [\tilde{\alpha}_1 X_n + \tilde{\alpha}_2 X_{n-1} + \cdots + \tilde{\alpha}_p X_{n-p}], X_k \right\rangle = \left\langle \tilde{Z}_{n+1}, X_k \right\rangle;$$

Por otro lado, debido a la causalidad del polinomio $\tilde{\alpha}(z)$, el Lema 3.6.1 garantiza que $\left\langle \tilde{Z}_{n+1}, X_k \right\rangle = 0$ para todo $k \leq n$, y entonces

(b)

$$\left\langle X_{n+1} - [\tilde{\alpha}_1 X_n + \tilde{\alpha}_2 X_{n-1} + \cdots + \tilde{\alpha}_p X_{n-p}], X_k \right\rangle = 0, \quad k = 1, 2, \dots, n.$$

Combinando (a) y (b) se desprende que $[\tilde{\alpha}_1 X_n + \tilde{\alpha}_2 X_{n-1} + \cdots + \tilde{\alpha}_p X_{n-p}]$ satisface las condiciones para ser la proyección de X_{n+1} sobre $\mathcal{L}(X_1, \dots, X_n)$, y la conclusión deseada se desprende de la unicidad de dicha proyección.

(ii) Dado $n > p$, la proyección $P_{\mathcal{L}(X_1, \dots, X_n)} X_{n+1}$ se escribe en términos de las constantes $\phi_{n,k}$ como $P_{\mathcal{L}(X_1, \dots, X_n)} X_{n+1}$

$$= \phi_{n1} X_n + \phi_{n2} X_{n-1} + \cdots + \phi_{np} X_{n-p} + \phi_{n,p-1} X_{n-p-1} + \cdots + \phi_{nn} X_1$$

Por otro lado, la función de autocovarianza de un proceso autorregresivo satisface las condiciones $\gamma(0) > 0$ y $\gamma(h) \rightarrow 0$ si $h \rightarrow \infty$ (Chacón, 2010) de tal manera que la matriz de varianzas $\mathbf{\Gamma}_n$ es no singular, y entonces X_1, X_2, \dots, X_n son linealmente independientes. A partir de este punto, comparando la expresión de arriba para $P_{\mathcal{L}(X_1, \dots, X_n)} X_{n+1}$ con la establecida en la parte (i), se desprende que

$$\phi_{nk} = \tilde{\alpha}_k, \quad k = 1, 2, \dots, p, \quad \phi_{nk} = 0, \quad p < k \leq n.$$

Para concluir note que la parte (i) y la representación causal (3.6.3) conducen a

$$\begin{aligned} X_{n+1} - P_{\mathcal{L}(X_1, \dots, X_n)} X_{n+1} \\ = X_{n+1} - [\tilde{\alpha}_1 X_n + \tilde{\alpha}_2 X_{n-1} + \cdots + \tilde{\alpha}_p X_{n-p}] = \tilde{Z}_{n+1}, \end{aligned}$$

de modo que

$$v_N = \|X_{n+1} - P_{\mathcal{L}(X_1, \dots, X_n)} X_{n+1}\|^2 = \|\tilde{Z}_{n+1}\|^2 = \tilde{\sigma}^2,$$

concluyendo el argumento. □

En el caso del Ejemplo 3.6.1 (i) el proceso autorregresivo satisface

$$X_t + 2X_{t-1} = Z_t$$

y el polinomio $\alpha(z) = 1 + 2z$ no es causal. De acuerdo a los cálculos efectuado anteriormente, ϕ_{n1} se estabiliza alrededor de -0.500 , y a partir del Teorema 3.6.1(i)

se concluye que en la representación causal $X_t - \tilde{\alpha}_1 X_{t-1} = \tilde{Z}_t$ donde $\{\tilde{Z}_t\}$ es un ruido blanco de varianza $\tilde{\sigma}^2$, debe tenerse $\tilde{\alpha}_1 = -0.500$; además, como v_n se estabiliza en el valor 0.250, el teorema establece que $\tilde{\sigma}^2 = .250$. Así, la representación causal de $\{X_t\}$ es

$$\mathbf{x}_t + 0.500X_{t-1} = \tilde{Z}_t, \quad \{\tilde{Z}_t\} \sim WN(0, 0.250);$$

esta conclusión, obtenida numéricamente, puede verificarse directamente a partir de los resultados en Chacón (2010). En el caso (ii) del Ejemplo 3.6.1, el proceso satisface

$$X_t - X_{t-1} + .21X_{t-2} = Z_t$$

donde el ruido blanco $\{Z_t\}$ tiene varianza 1. En esta representación, el polinomio $\alpha(z)$ es causal, y el Teorema 3.6.1 asegura que $\phi_{n,1}$ y $\phi_{n,2}$ se estabilizan en $(\alpha_1, \alpha_2) = (1, -0.210)$, conclusión que coincide con el resultado numérico. Además, de acuerdo al teorema, v_n se estabiliza en $\sigma^2 = 1$; los cálculos indican que v_n llega al valor 1.001, y la diferencia de un milésimo se debe, desde luego, a errores de cálculo que surgen naturalmente, así como al hecho de que la función de covarianza de este proceso fue determinada numéricamente.

CAPÍTULO 4

ALGORITMO DE INNOVACIONES

En este capítulo se estudia la representación de una proyección \hat{X}_{n+1} en términos de una base ortogonal U_1, \dots, U_n del espacio generado por X_1, \dots, X_n , denominada la base de innovaciones. La variable U_i representa la información contenida en X_i que no puede ser obtenida a partir de las observaciones previas X_1, \dots, X_i . El objetivo es establecer un procedimiento recursivo para determinar los coeficientes de la representación de \hat{X}_{n+1} como combinación lineal de U_1, \dots, U_n , y determinar requerimientos de memoria y operaciones del algoritmo.

4.1. Un Conjunto Ortogonal

El problema de calcular proyecciones sobre un subespacio \mathcal{L} se facilita al disponer de un conjunto generador que sea ortogonal (Hoffman y Kunze, 1980; Graybill, 2001 y Harville, 2002). En esta sección se construye uno de tales conjuntos para $\mathcal{L}(X_1, X_2, \dots, X_n)$ donde, en adelante, n es un entero positivo fijo. Considere las variables aleatorias

$$U_1 = X_1, \quad U_k = X_k - P_{\mathcal{L}(X_1, \dots, X_{k-1})} X_k, \quad k = 1, 2, \dots, n. \quad (4.1.1)$$

En este caso, debido a que

$$U_i \in \mathcal{L}(X_1, \dots, X_i) \quad \text{y} \quad U_j \text{ es ortogonal a } \mathcal{L}(X_1, \dots, X_{j-1}),$$

se tiene la siguiente relación de ortogonalidad:

$$i < j \Rightarrow \langle U_i, U_j \rangle = 0.$$

Suponga ahora que la matriz de varianza $\mathbf{\Gamma}_n$ del vector (X_1, \dots, X_n) es no singular para cada n , de manera que X_1, \dots, X_n son linealmente independientes. Como U_k es una combinación lineal de X_1, \dots, X_k en la que el coeficiente de X_k es 1, se tiene que $U_k \neq 0$ para cada k . Combinando esta propiedad con la anterior relación de ortogonalidad se desprende que

$$U_1, \dots, U_n \text{ son linealmente independientes.}$$

y por lo tanto, como $U_1, \dots, U_n \in \mathcal{L}(X_1, \dots, X_n)$, un argumento directo de dimensionalidad permite establecer que

$$\mathcal{L}(U_1, \dots, U_n) = \mathcal{L}(X_1, \dots, X_n)$$

para cada n . Por otro lado, recuerde que el pronóstico para X_k en términos de X_1, \dots, X_{k-1} es $\hat{X}_k = P_{\mathcal{L}(X_1, \dots, X_{k-1})} X_k$, de manera que $U_k = X_k - \hat{X}_k$, y entonces

$$\|U_k\|^2 = \|X_k - \hat{X}_k\|^2 = v_{k-1};$$

vea (2.1.11). La discusión precedente se resume como sigue:

Lema 4.1.1. Dada una serie estacionaria $\{X_t\}$ defina las variables aleatorias U_k mediante (4.1.1). En este caso,

- (i) Las variables U_1, U_2, U_3, \dots son ortogonales;
- (ii) Se satisface la igualdad $\mathcal{L}(U_1, U_2, \dots, U_k) = \mathcal{L}(X_1, X_2, \dots, X_k)$ para cada k ;
- (iii) Para cada $k = 1, 2, 3, \dots$, $\|U_k\|^2 = v_{k-1}$.

Las conclusiones anteriores son válidas en general, sin suponer que las variables X_1, \dots, X_n, \dots sean linealmente independientes (Hoffman y Kunze, 1980; Graybill, 2001 y Harville, 2002). La variable U_k se denomina la *innovación* en el tiempo k . Para justificar este término, note que X_k se expresa como

$$X_k = P_{\mathcal{L}(X_1, \dots, X_{k-1})} X_k + [X_k - P_{\mathcal{L}(X_1, \dots, X_{k-1})} X_k]$$

y ubíquese en el momento en que X_k va a ser observada. En ese instante, ya se registraron X_1, \dots, X_{k-1} , y el término $P_{\mathcal{L}(X_1, \dots, X_{k-1})}X_k$ es combinación lineal de esas variables; por lo tanto, $P_{\mathcal{L}(X_1, \dots, X_{k-1})}X_k$ es conocido por el observador. En contraste, $[X_k - P_{\mathcal{L}(X_1, \dots, X_{k-1})}X_k]$ es ortogonal a X_1, \dots, X_{k-1} , y en ese sentido, representa ‘la nueva información’ que el observador obtendrá al registrar el valor de X_k . Por otro lado, como el conjunto ortogonal U_1, \dots, U_n genera el mismo espacio vectorial que X_1, \dots, X_n , se tiene que la proyección

$$\hat{X}_{n+1} = P_{\mathcal{L}(X_1, \dots, X_n)}X_{n+1} = P_{\mathcal{L}(U_1, \dots, U_n)}X_{n+1}$$

se expresa como

$$\begin{aligned} \hat{X}_{n+1} &= \theta_{n1}U_n + \theta_{n2}U_{n-1} + \dots + \theta_{nn-1}U_2 + \theta_{nn}U_1 \\ &= \sum_{k=1}^n \theta_{nk}U_{n+1-k}. \end{aligned} \tag{4.1.2}$$

Esta es la representación de \hat{X}_{n+1} en términos de las innovaciones U_1, \dots, U_n .

4.2. Relaciones entre los Coeficientes

En esta sección se estudia la determinación de los coeficientes de la expresión de \hat{X}_{n+1} como combinación lineal de las innovaciones U_1, \dots, U_k . Como punto de partida, suponga que, para cada $k = 1, \dots, n-1$, ya se han determinado los coeficientes θ_{kj} , $1 \leq j \leq k$ en las expresiones

$$\hat{X}_{k+1} = \sum_{j=1}^k \theta_{kj}U_{k+1-j} \tag{4.2.1}$$

así como los correspondientes errores cuadráticos de pronóstico

$$v_k = \|U_{k+1}\|^2 = \|X_{k+1} - \hat{X}_{k+1}\|^2. \tag{4.2.2}$$

El siguiente teorema muestra como pueden determinarse los coeficientes θ_{nk} y el error v_n en términos de estas cantidades.

Teorema 4.2.1. Suponga que para cada entero $k = 1, 2, \dots, n - 1$ se conocen los coeficientes $\theta_{k j}$, $1 \leq j \leq k$, así como v_k . En este caso,

(i) Los coeficientes $\theta_{n k}$ en la expresión

$$\hat{X}_{n+1} = \sum_{k=1}^n \theta_{n k} U_{n+1-k} \quad (4.2.3)$$

satisfacen

$$\begin{aligned} v_0 \theta_{n n} &= \gamma(n), \\ v_i \theta_{n n-i} &= \gamma(n-i) - \sum_{s=0}^{i-1} \theta_{i i-s} \theta_{n n-s} v_s, \quad i = 1, 2, \dots, n-1. \end{aligned}$$

(ii) Más aún,

$$v_n = \gamma(0) - \sum_{j=1}^n \theta_{n j}^2 v_{n-j}$$

Demostración. Como punto de partida, recuerde que la especificación de una proyección implica que

$$\langle X_{n+1}, U_j \rangle = \langle \hat{X}_{n+1}, U_j \rangle, \quad j = 1, 2, \dots, n, \quad (4.2.4)$$

mientras que haciendo el cambio de variable $j = n - k$, (4.2.3) equivale a

$$\hat{X}_{n+1} = \sum_{j=0}^{n-1} \theta_{n n-j} U_{j+1} \quad (4.2.5)$$

(i) Combinando las dos anteriores expresiones y aplicando la ortogonalidad de las variables U_i se obtiene

$$\langle X_{n+1}, U_1 \rangle = \langle \hat{X}_{n+1}, U_1 \rangle = \left\langle \sum_{j=0}^{n-1} \theta_{n n-j} U_{j+1}, U_1 \right\rangle = \theta_{n n} \langle U_1, U_1 \rangle,$$

y recordando que $U_1 = X_1$ y $\|X_1\|^2 = \gamma(0)$, se desprende que

$$\gamma(n) = \langle X_{n+1}, U_1 \rangle = \theta_{n n} \gamma(0) = \theta_{n n} v_0. \quad (4.2.6)$$

Ahora, seleccione un entero i entre 1 y $n - 1$, y tome el producto interno por U_{i+1} en ambos lados de (4.2.5) para obtener

$$\begin{aligned}\langle X_{n+1}, U_{i+1} \rangle &= \langle \hat{X}_{n+1}, U_{i+1} \rangle \\ &= \left\langle \sum_{j=0}^{n-1} \theta_{n n-j} U_{j+1}, U_{i+1} \right\rangle = \theta_{n n-i} \langle U_{i+1}, U_{i+1} \rangle\end{aligned}$$

Como $v_i = \langle U_{i+1}, U_{i+1} \rangle$ se tiene que

$$v_i \theta_{n n-i} = \langle X_{n+1}, U_{i+1} \rangle \quad (4.2.7)$$

A continuación observe que

$$U_{i+1} = X_{i+1} - P_{\mathcal{L}(X_1, \dots, X_i)} X_{i+1} = X_{i+1} - \sum_{s=0}^{i-1} \theta_{i i-s} U_{s+1}$$

de manera que

$$\begin{aligned}\langle X_{n+1}, U_{i+1} \rangle &= \left\langle X_{n+1}, X_{i+1} - \sum_{s=0}^{i-1} \theta_{i i-s} U_{s+1} \right\rangle \\ &= \langle X_{n+1}, X_{i+1} \rangle - \sum_{s=0}^{i-1} \theta_{i i-s} \langle X_{n+1}, U_{s+1} \rangle \\ &= \gamma(n-i) - \sum_{s=0}^{i-1} \theta_{i i-s} \langle X_{n+1}, U_{s+1} \rangle\end{aligned}$$

y combinando esta relación con (4.2.7) se obtiene

$$v_i \theta_{n n-i} = \gamma(n-i) - \sum_{s=0}^{i-1} \theta_{i i-s} \langle X_{n+1}, U_{s+1} \rangle \quad (4.2.8)$$

Para evaluar el producto interno en la sumatoria anterior, observe que

$$\begin{aligned}X_{n+1} &= [X_{n+1} - \hat{X}_{n+1}] + \hat{X}_{n+1} \\ &= [X_{n+1} - \hat{X}_{n+1}] + \sum_{j=0}^{n-1} \theta_{n n-j} U_{j+1} \\ &= U_{n+1} + \sum_{j=0}^{n-1} \theta_{n n-j} U_{j+1};\end{aligned}$$

vea (4.1.1) y (4.2.5). Usando la ortogonalidad de las variables U_k , se desprende que para cada $s < n$

$$\begin{aligned}\langle X_{n+1}, U_{s+1} \rangle &= \left\langle U_{n+1} + \sum_{j=0}^{n-1} \theta_{n n-j} U_{j+1}, U_{s+1} \right\rangle \\ &= \theta_{n n-s} \langle U_{s+1}, U_{s+1} \rangle = \theta_{n n-s} v_s;\end{aligned}$$

vea (4.2.2) para la última igualdad. Sustituyendo esta expresión en (4.2.8) se obtiene

$$v_i \theta_{n n-i} = \gamma(n-i) - \sum_{s=0}^{i-1} \theta_{i i-s} \theta_{n n-s} v_s, \quad i = 1, 2, \dots, n-1,$$

completando la demostración de la parte (i).

(ii) Debido a que las variables U_k son ortogonales, a partir de (4.2.3) y usando la notación en (4.2.2) se desprende que

$$\|\hat{X}_{n+1}\|^2 = \left\| \sum_{k=1}^n \theta_{n k} U_{n+1-k} \right\|^2 = \sum_{k=1}^n \theta_{n k}^2 \|U_{n+1-k}\|^2 = \sum_{k=1}^n \theta_{n k}^2 v_{n-k};$$

como $v_n = \|X_{n+1}\|^2 - \|\hat{X}_{n+1}\|^2 = \gamma(0) - \|\hat{X}_{n+1}\|^2$, la expresión deseada para v_n se obtiene de inmediato. \square

4.3. Algoritmo de Innovaciones

El Teorema 4.2.1 sugiere una manera natural de encontrar los coeficientes $(\theta_{n n}, \theta_{n n-1}, \dots, \theta_{n 1})$ en la expresión (4.2.3) para el pronóstico \hat{X}_{n+1} , así como el correspondiente error cuadrático $v_n = \|X_{n+1} - \hat{X}_{n+1}\|^2$.

Suponga que la función de autocovarianza es tal que, para cada n , la matriz Γ_n es invertible, de manera que $v_k > 0$ para todo k . Iniciando con

$$v_0 = \gamma(0),$$

el Teorema 4.2.1 permite calcular

$$(i) \theta_{11} = \gamma(1)/v_0 \text{ y } v_1 = \gamma(0) - \theta_{11}^2 v_0.$$

Usando estas cantidades, ahora se calculan los coeficientes θ_{22}, θ_{21} y v_1 como sigue:

$$(ii) \theta_{22} = \gamma(2)/v_0, \theta_{21} = [\gamma(1) - \theta_{11}\theta_{22}v_0]/v_1, \text{ y } v_2 = \gamma(0) - \theta_{22}^2 v_0 - \theta_{21}^2 v_1.$$

A partir de este punto, las fórmulas del Teorema 4.2.1 permiten calcular $\theta_{33}, \theta_{32}, \theta_{31}$ y v_3 , en ese orden. En general, dados $\theta_{kj}, 1 \leq j \leq k$, y v_0, \dots, v_k , es posible evaluar $\theta_{nn}, \theta_{n,n-1}, \dots, \theta_{n1}$ y v_n , en el orden indicado. Este procedimiento se describe formalmente a continuación:

Algoritmo de Innovaciones

Entrada: Un vector $(\gamma(0), \gamma(1), \dots, \gamma(n))$, donde $\gamma(\cdot)$ es una función de autocovarianza y $n \geq 2$

Salida: El vector $\theta_n = (\theta_{n1}, \dots, \theta_{nn})$ de coeficientes de la proyección \hat{X}_{n+1} en (4.2.3), así como v_n , el correspondiente error cuadrático de pronóstico.

- Inicio: Defina $v_0 = \gamma(0)$, $\theta_{11} = \gamma(1)/v_0$, $v_1 = \gamma(0) - \theta_{11}^2 v_0$ y $k = 1$.
- Paso Recursivo: Incremente k en 1 y calcule

$$\begin{aligned} \theta_{kk} &= \gamma(k)/v_0, \\ \theta_{kk-i} &= [\gamma(k-i) - \sum_{s=0}^{i-1} \theta_{ii-s} \theta_{kk-s} v_s]/v_i, \quad i = 1, 2, \dots, k-1, \\ v_k &= \gamma(0) - \sum_{j=1}^k \theta_{kj}^2 v_{k-j} \end{aligned}$$

- Paso de Prueba: Si $k = n$ detenga el procedimiento. Los números actuales v_k y $\theta_{kj}, j = 1, 2, \dots, k$ son las cantidades deseadas.

Si $k < n$, vaya al paso anterior.

Un aspecto importante en el algoritmo de innovaciones, es que para evaluar v_k y los números θ_{kj} , es necesario tener disponibles todas las cantidades θ_{si} y v_i , $1 \leq i \leq s < k$. Así, debe reservarse lugar para todos los números en el siguiente arreglo:

$$\begin{array}{cccccc}
 v_0 & & & & & \\
 \theta_{11} & v_1 & & & & \\
 \theta_{22} & \theta_{21} & v_2 & & & \\
 \theta_{33} & \theta_{33} & \theta_{31} & v_3 & & \\
 \vdots & \vdots & \vdots & \vdots & \ddots & \\
 \theta_{nn} & \theta_{nn-1} & \theta_{nn-2} & \cdots & \theta_{n1} & v_n
 \end{array}$$

totalizando $1 + 2 + \cdots + (n + 1) = (n + 1)(n + 2)/2$ lugares de memoria. Además se requieren $n + 1$ lugares para $\gamma(0), \gamma(1), \dots, \gamma(n)$ y uno más para la variable k . Así, el requerimiento de memoria es de $(n + 2)(n + 3)/2$ lugares numéricos, es decir, $O(n^2)$.

Ahora se discutirá la carga aritmética del algoritmo, contando el número de multiplicaciones y divisiones necesarias para ejecutarlo. En el paso recursivo, aplicado al entero k requiere

- (a) Una división para evaluar θ_{kk} ;
- (b) $2(i - 1)$ multiplicaciones y una división para calcular θ_{kk-i} , dando un total de $2i - 1$ operaciones; esto debe hacerse para $i = 1, 2, k - 1$, arrojando un total de $\sum_{i=1}^{k-1} (2i - 1) = k(k - 1) - (k - 1) = (k - 1)^2$ multiplicaciones o divisiones. Como el paso recursivo se realiza con valores $k = 2, \dots, n$, el total de operaciones en todos los pasos, es

$$\sum_{k=2}^n (k - 1)^2 = (n - 1)(n)(2n - 1)/6 \approx n^3/3.$$

Esta discusión se resume en el siguiente teorema:

Teorema 4.3.1. Para determinar los n coeficientes $\theta_{nn}, \dots, \theta_{n1}$ y el error cuadrático de pronóstico v_n , el algoritmo de innovaciones requiere $(n + 2)(n + 3)/2$ lugares numéricos de memoria, y la realización de $3 + (n - 1)(n)(2n - 1)/6 \approx n^3/3$ multiplicaciones o divisiones. Así, el algoritmo de innovaciones es de orden $O(n^2)$

en cuanto a requerimiento de memoria, y de orden $O(n^3)$ con respecto a la carga aritmética.

En el siguiente capítulo se diseña un método para determinar los coeficientes en (4.2.3) cuya carga aritmética es de orden cuadrático. Dicho resultado depende de la relación entre los coeficientes que se utilizan para representar la proyección \hat{X}_{n+1} en términos de las innovaciones U_1, \dots, U_n , y las cantidades que se usan en la representación como combinación lineal de las variables originales X_1, \dots, X_n .

CAPÍTULO 5

ALGORITMO DE ORDEN CUADRÁTICO

En este capítulo se presenta el principal resultado de este trabajo: se formula un algoritmo de orden $O(n^2)$ para determinar el vector de coeficientes $\boldsymbol{\theta}_n$ en la representación del pronóstico \hat{X}_{n+1} en términos de las primeras n innovaciones. El procedimiento que se propone se basa en determinar la relación entre $\boldsymbol{\theta}_n$ y los vectores de coeficientes $\boldsymbol{\varphi}_k$ en la representación de los pronósticos \hat{X}_{k+1} como combinación lineal de las variables originales X_1, \dots, X_k . Esta relación, conjuntamente con el algoritmo de Durbin Levinson, determinan el procedimiento propuesto.

5.1. Relación entre dos Expresiones para un Pronóstico

El propósito de esta sección es desarrollar la idea principal en que se sustenta la formulación de un procedimiento de segundo orden para determinar el vector de coeficientes $\boldsymbol{\theta}_n$ que aparece al expresar el pronóstico \hat{X}_{n+1} en términos de las innovaciones U_1, \dots, U_n . El resultado principal relaciona dicho vector con la representación de los pronósticos \hat{X}_{k+1} en términos de las variables originales.

Considere las dos representaciones para la proyección \hat{X}_{n+1} de X_{n+1} sobre $\mathcal{L}(X_1, \dots, X_n)$:

$$\begin{aligned}\hat{X}_{n+1} &= \sum_{j=1}^n \theta_{n,j} U_{n+1-j} \\ \hat{X}_{n+1} &= \sum_{j=1}^n \phi_{n,j} X_{n+1-j}\end{aligned}\tag{5.1.1}$$

donde, como antes, U_{k+1} es la innovación en el tiempo $k+1$. El principal resultado en esta dirección es el siguiente.

Teorema 5.1.1. Considere una serie estacionaria $\{X_t\}$ tal que la matriz Γ_k es invertible para cada entero positivo k ; vea (2.1.6). Sea n un entero positivo, y suponga que, para algún entero k entre 1 y n , se tiene la siguiente representación:

$$P_{\mathcal{L}(X_1, \dots, X_k)} X_{n+1} = \sum_{i=1}^k c_i X_{k+1-i}, \quad (5.1.2)$$

donde c_1, \dots, c_k son constantes. En este caso,

- (i) El coeficiente $\theta_{n, n+1-k}$ en (5.1.1) está dado por

$$\theta_{n, n+1-k} = c_1.$$

- (ii) La proyección X_{n+1} sobre el espacio $\mathcal{L}(X_1, \dots, X_{k-1})$ está dada por

$$P_{\mathcal{L}(X_1, \dots, X_{k-1})} X_{n+1} = \sum_{i=1}^{k-1} d_i X_{k-i},$$

donde

$$(d_1, d_2, \dots, d_{k-1}) = (c_2, c_3, \dots, c_k) + c_1(\phi_{k-1, 1}, \phi_{k-1, 2}, \dots, \phi_{k-1, k-1})$$

Demostración. Primeramente, observe que (5.1.2) puede escribirse como

$$\begin{aligned} P_{\mathcal{L}(X_1, \dots, X_k)} X_{n+1} &= c_1 X_k + \sum_{i=2}^k c_i X_{k+1-i} \\ &= c_1 X_k + \sum_{j=1}^{k-1} c_{j+1} X_{k-j} \\ &= c_1 [X_k - P_{\mathcal{L}(X_1, \dots, X_{k-1})} X_k] \\ &\quad + \left[\sum_{j=1}^{k-1} c_{j+1} X_{k-j} + c_1 P_{\mathcal{L}(X_1, \dots, X_{k-1})} X_k \right]. \end{aligned}$$

El término entre corchetes que multiplica a c_1 es precisamente la innovación U_k , mientras que $P_{\mathcal{L}(X_1, \dots, X_{k-1})}X_k = \sum_{j=1}^{k-1} \phi_{k-1,j}X_{k-j}$. Por lo tanto,

$$\begin{aligned} P_{\mathcal{L}(X_1, \dots, X_k)}X_{n+1} &= c_1U_k + \left[\sum_{j=1}^{k-1} c_{j+1}X_{k-j} + c_1 \sum_{j=1}^{k-1} \phi_{k-1,j}X_{k-j} \right] \\ &= c_1U_k + \left[\sum_{j=1}^{k-1} [c_{j+1} + c_1\phi_{k-1,j}]X_{k-j} \right] \end{aligned} \quad (5.1.3)$$

Por otro lado, la primera igualdad en (5.1.1) puede escribirse como

$$P_{\mathcal{L}(X_1, \dots, X_n)}X_{n+1} = \sum_{j=1}^n \theta_{n,j}U_{n+1-j} = \sum_{s=1}^n \theta_{n,n+1-s}U_s$$

Recuerde ahora dos hechos fundamentales:

- (a) Si \mathcal{M}_0 y \mathcal{M}_1 son dos subespacios tales que $\mathcal{M}_0 \subset \mathcal{M}_1$, entonces

$$P_{\mathcal{M}_0}P_{\mathcal{M}_1} = P_{\mathcal{M}_0};$$

vea, por ejemplo, Hoffman y Kunze (1980) o Lipschutz (1995).

- (b) $\mathcal{L}(X_1, \dots, X_s) = \mathcal{L}(U_1, \dots, U_s)$, igualdad que se estableció en el capítulo precedente.

Usando estas dos propiedades, y recordando que k es un entero entre 1 y n , se desprende que

$$\begin{aligned} P_{\mathcal{L}(X_1, \dots, X_k)}X_{n+1} &= P_{\mathcal{L}(X_1, \dots, X_k)}P_{\mathcal{L}(X_1, \dots, X_n)}X_{n+1} \\ &= P_{\mathcal{L}(U_1, \dots, U_k)} \left[\sum_{j=1}^n \theta_{n,j}U_{n+1-j} \right] \\ &= \sum_{j=1}^n \theta_{n,j}P_{\mathcal{L}(U_1, \dots, U_k)}U_{n+1-j} \\ &= \sum_{j=n+1-k}^n \theta_{n,j}U_{n+1-j} \end{aligned}$$

donde se usó que $P_{\mathcal{L}(U_1, \dots, U_k)}U_s = 0$ si $s > k + 1$ y $P_{\mathcal{L}(U_1, \dots, U_k)}U_s = U_s$ cuando $s \leq k$. El mismo argumento con $k - 1$ en vez de k muestra que

$$P_{\mathcal{L}(X_1, \dots, X_{k-1})}X_{n+1} = \sum_{j=n+2-k}^n \theta_{n,j}U_{n+1-j}.$$

Combinando estas dos últimas relaciones desplegadas se obtiene que

$$\begin{aligned} P_{\mathcal{L}(X_1, \dots, X_k)} X_{n+1} &= \sum_{j=n+1-k}^n \theta_{n,j} U_{n+1-j} \\ &= \theta_{n, n+1-k} U_k + \sum_{j=n+2-k}^n \theta_{n,j} U_{n+1-j} \end{aligned}$$

y entonces

$$P_{\mathcal{L}(X_1, \dots, X_k)} X_{n+1} = \theta_{n, n+1-k} U_k + P_{\mathcal{L}(X_1, \dots, X_{k-1})} X_{n+1}$$

Comparando esta expresión con (5.1.3), la ortogonalidad y no nulidad de las innovaciones implica que $\theta_{n, n+1-k} = c_1$, como se establece en la parte (i), mientras que

$$P_{\mathcal{L}(X_1, \dots, X_{k-1})} X_{n+1} = \sum_{j=1}^{k-1} [c_{j+1} + c_1 \phi_{k-1,j}] X_{k-j} = \sum_{j=1}^{k-1} d_j X_{k-j},$$

donde $d_j = c_{j+1} + c_1 \phi_{k-1,j}$ para $j = 1, 2, \dots, k-1$, como se establece en la parte (ii). \square

Conociendo los vectores $\varphi_k = (\phi_{k,1}, \dots, \phi_{k,k})$, $k = 1, 2, \dots, n$, este teorema permite calcular el vector $\theta_n = (\theta_{n,1}, \theta_{n,2}, \dots, \theta_{n,n})$, cuyas componentes son los coeficientes en la representación de \hat{X}_{n+1} en términos de las innovaciones U_1, \dots, U_n .

Iniciando con

$$P_{\mathcal{L}(X_1, \dots, X_n)} X_{n+1} = \sum_{i=1}^n \phi_{n,i} X_{n+1-i}$$

el Teorema 5.1.1 con $k = n$ y $c = (c_1, \dots, c_n) = \varphi_n$ establece que $\theta_{n, n+1-n} = \theta_{n,1} = c_1 = \phi_{n,1}$, y permite calcular el vector $d = (d_1, d_2, \dots, d_{n-1})$ tal que

$$P_{\mathcal{L}(X_1, \dots, X_{n-1})} X_{n+1} = \sum_{i=1}^{n-1} d_i X_{n-i}$$

Otra aplicación del teorema precedente con $k = n-1$ y con el vector d en lugar de c arroja que $\theta_{n, n+1-(n-1)} = \theta_{n,2} = d_1$, y un vector $\tilde{d} = (\tilde{d}_1, \tilde{d}_2, \dots, \tilde{d}_{n-2})$ que

satisface

$$P_{\mathcal{L}(X_1, \dots, X_{n-2})} X_{n+1} = \sum_{i=1}^{n-2} \tilde{d}_i X_{n-1-i}$$

A partir de aquí se obtiene $\theta_{n3} = \tilde{d}_1$ y otro vector \hat{d} tal que

$$P_{\mathcal{L}(X_1, \dots, X_{n-3})} X_{n+1} = \sum_{i=1}^{n-3} \tilde{d}_i X_{n-2-i}$$

Continuando con este proceso, se llega finalmente a determinar todos los coeficientes $\theta_{n1}, \theta_{n2}, \dots, \theta_{nn}$.

5.2. Procedimiento de Transformación

En esta sección se describe formalmente el algoritmo sugerido por el Teorema 5.1.1 para encontrar los coeficientes $\theta_{n1}, \dots, \theta_{nn}$ en la representación de \hat{X}_{n+1} en términos de las innovaciones U_1, \dots, U_n .

Algoritmo de Transformación

Entrada: Un entero positivo n y los vectores φ_k tales que

$$\hat{X}_{k+1} = (X_1, \dots, X_k) \varphi_k$$

para cada $k = 1, \dots, n$.

Salida: El vector $\theta_n = (\theta_{n1}, \dots, \theta_{nn})$ tal que $\hat{X}_{k+1} = (U_1, \dots, U_k) \theta_n$.

Inicio: Ponga $k = 1$ y defina el vector $\alpha = \varphi_n$

Etapas de cálculo:

- (a) Si $k = n$ ponga $\theta_{nn} = \alpha_1$
- (b) Si $k < n$, ponga $\theta_{n, n+1-k} = \alpha_1$, y sustituya α por el vector

$$(\alpha_2, \dots, \alpha_{n+1-k}) + \alpha_1 \varphi_{n-k}$$

e incremente k por una unidad.

Etapa de salida:

- (a) Si $k > n$ salga del procedimiento mostrando el vector θ_n .
- (b) Si $k \leq n$ vaya a la etapa de cálculo.

Este procedimiento fue codificado en *R* mediante una función como se muestra a continuación

```
DurLevtoInn <- function(phis) {
  n<- length(phis)
  alpha<- phis[[n]]
  theta<- vector("numeric", n)
  for(k in n:2){
    theta[n+1-k] <- alpha[1]
    alpha <- alpha[2: k] + alpha[1]*phis[[k-1]]
  }
  theta[n] <-alpha
  theta
}
```

La función acepta como único parámetro una lista denominada `phis`, tal que la componente `phis[[k]]` es el vector φ_k ; el número n de datos observados está implícito en la longitud de la lista.

Ejemplo 5.2.1. Suponga que se desea encontrar al vector θ_5 en la representación de \hat{X}_6 en términos de innovaciones, y que el proceso $\{X_t\}$ tiene función de autocovarianza $\gamma(0) = 1.81$, $\gamma(h) = -.9$ si $|h| = 1$, y $\gamma(h) = 0$ para $|h| > 1$, correspondiente a un proceso de promedios móviles de orden 1. Para usar la función anterior, es necesario disponer de la lista

$$\text{phis} = (\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5)$$

Para generar esta lista, primero se llama a la función `DurLev` con `Gammazero = 1.81`, y `Gamma=(-.9, 0, 0, 0, 0)` (un vector de longitud cinco, pues $n = 5$).

```
DurLev(1.81, c(-.9, 0, 0, 0, 0)) -> phis
```

Como la salida de la función `DurLev` es la lista $(\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \mathbf{v})$ donde \mathbf{v} es el vector de errores de pronóstico, la llamada función `DurLevtoInn` se hará con la sublista

```
phis[1 : 5] = (phi_1, phi_2, phi_3, phi_4, phi_5)
> DurLevtoInn(phis[1:5])
[1] -.8169085 -0.000000 0.000000 -0.000000 -0.000000
```

Esta salida muestra que $\theta_{5k} = 0$ para $k > 1$, como debe ser en este caso, pues $\gamma(\cdot)$ es la función de autocovarianza de un proceso de promedios móviles de orden 1, y que $\theta_{51} = -.8169085$; este valor coincide con los resultados obtenidos en Brockwell y Davis (1998). \square

Para concluir la sección, se analiza el orden del procedimiento de transformación para obtener el vector $\boldsymbol{\theta}_n$.

Teorema 5.2.1. El procedimiento de transformación descrito anteriormente es de orden cuadrático, tanto en el requerimiento de memoria, como en la carga aritmética.

Demostración. Debido a que el procedimiento requiere utilizar los vectores $\varphi_1, \dots, \varphi_n$ cuyas longitudes suman $1 + 2 + \dots + n = n(n+1)/2$, es claro que el requerimiento de memoria tiene orden $O(n^2)$. Por otro lado, en la etapa de cálculo se realizan $n - k$ multiplicaciones al realizar el producto $\alpha_1 \varphi_{n-k}$, y esto se hace para $k = 1, 2, \dots, n - 1$. Por lo tanto se realizan

$$(n-1) + (n-1) + \dots + 1 = n(n-1)/2$$

multiplicaciones, así que el procedimiento también es de orden cuadrático respecto a la carga aritmética. \square

5.3. Implementación del Algoritmo de Innovaciones

En esta sección se describe un procedimiento para determinar el vector

$$\boldsymbol{\theta}_n = (\theta_{n1}, \theta_{n2}, \dots, \theta_{nn},)$$

en la representación

$$\hat{X}_{n+1} = \sum_{k=1}^n \theta_{nk} X_{n+1-k}$$

así como el error de pronóstico

$$v_n = \|X_{n+1} - \hat{X}_{n+1}\|^2$$

El siguiente procedimiento combina el algoritmo de Durbin-Levinson el esquema de transformación de la sección precedente para resolver el problema, cuya solución es el principal objetivo del algoritmo de innovaciones.

Algoritmo combinado para determinar $\boldsymbol{\theta}_n$ y v_n

Paso 1: Utilice el método de Durbin-Levinson para determinar v_n y los vectores $\boldsymbol{\varphi}_k$, $k = 1, 2, \dots, n$, en las representaciones $\hat{X}_{k+1} = (X_1, \dots, X_k)\boldsymbol{\varphi}_k$.

Paso 2: Aplique el método de transformación descrito en la sección precedente para determinar $\boldsymbol{\theta}_n$ a partir de $\boldsymbol{\varphi}_k$, $k = 1, 2, \dots, n$.

El siguiente teorema condensa el trabajo realizado previamente, y se enuncia para enfatizar la importancia del resultado.

Teorema 5.3.1. El algoritmo combinado descrito anteriormente es de orden cuadrático, tanto en los requerimientos de memoria como en la carga aritmética.

Demostración. Debido a que tanto el algoritmo de Durbin-Levinson como el método de transformación son de orden cuadrático en requerimientos de memoria y carga aritmética, se desprende que el esquema combinado también es de orden $O(n^2)$. \square

Este resultado indica claramente la ventaja del algoritmo combinado respecto a la formulación original del algoritmo de innovaciones. Ambos procedimientos encuentran θ_n y v_n , pero el algoritmo combinado es de orden $O(n^2)$ y por lo tanto requiere realizar menos operaciones aritméticas que el algoritmo de innovaciones original, el cual tiene orden $O(n^3)$.

LITERATURA CITADA

- [1]. T. M. Apostol (1980), *Mathematical Analysis*, Addison Wesley, Reading, Massachusetts.
- [2]. A. A. Borovkov (1999), *Mathematical Statistics*, Gordon and Breach, New York
- [3]. P. J. Brockwell y R. A. Davis (1998), *Time Series: Theory and Methods*, Springer-Verlag, New York.
- [4]. J. C. Chacón Hernández (2010), *Un Criterio Integral de Causalidad Para Procesos ARMA*, Tesis de Maestría en Estadística Aplicada, Universidad Autónoma Agraria Antonio Narro , Saltillo Coah., MÉXICO.
- [5]. E. Dudewicz y S. Mishra (1998). *Mathematical Statistics*, Wiley, New York.
- [6]. W. A. Fuller (1998), *Introduction to Statistical Time Series* , Wiley, New York.
- [7]. W. Fulks (1980), *Cálculo Avanzado*, Limusa, México, D. F.
- [8]. F. A. Graybill (2000), *Theory and Application of the Linear Model*, Duxbury, New York.
- [9]. F. A. Graybill (2001), *Matrices with Applications in Statistics* Duxbury, New York.
- [10]. D. A. Harville (2008), *Matrix Algebra Form a Statistician's Perspective*, Springer-Verlaf, New York.
- [11]. K. Hoffman y R. Kunze (1975), *Linear Algebra*, Prentice-Hall, New York.
- [12]. A. I. Khuri (2002), *Advanced Calculus with Applications in Statistics*, Wiley, New York.
- [13]. S. Lipschutz (1995), *Linear Algebra*, McGraw-Hill, New York.
- [14]. W. Rudin (1984), *Real and Complex Analysis*, McGraw-Hill, New York.

- [15]. H. L. Royden (2003), Real Analysis, *MacMillan*, London.
- [16]. J. Shao (2010), Mathematical Statistics, *Springer*, New York.
- [17]. R. H. Shumway y D. S. Stoffer (2006), Time Series Analysis and Its Applications With R Examples, *Springer-Verlag*, New York. Edition
- [18]. G. Strang, (2003), Linear Algebra, *Prentice-Hall*, New York.
- [19]. S. Vázquez Baxcajay (2010), Ejemplos sobre la Teoría Espectral para Series de Tiempo Estacionarias, Tesis de Maestría en Estadística Aplicada, Universidad Autónoma Agraria Antonio Narro , Saltillo Coah., MÉXICO.
- [20]. D. Wackerly, W. Mendenhall y R. L. Scheaffer (2009), Mathematical Statistics with Applications, *Prentice-Hall*, New York.

Una Implementación de Orden Cuadrático del Algoritmo de Innovaciones en Series de Tiempo

Nadia Y. Martínez Martínez, Tesis de Maestría en Estadística Aplicada, Universidad Autónoma Agraria Antonio Narro, Saltillo, COAH., MÉXICO

nadia.mtz.mtz@gmail.com

Resumen. Este trabajo trata sobre el problema de pronóstico para una serie de tiempo estacionaria. El principal resultado es la formulación de un método de orden cuadrático para implementar el algoritmo de innovaciones, cuya codificación usual tiene orden cúbico. El procedimiento propuesto combina el esquema recursivo de Durbin Levinson, con un teorema de transformación de representaciones para un pronóstico.

A Second Order Implementation for the Innovations Algorithm in Time Series Analysis

Abstract. This work concerns the forecasting problem for a stationary time series. The main result is a second order implementation of the innovations algorithm, whose usual codification has cubic order. The proposed method is obtained by combining the Durbin-Levinson algorithm with a new theorem on the relation between two representations of a forecast, one in terms of the innovations, and the other one in terms of the original observations.

1. Introducción

Este trabajo se ubica en el área conocida como *análisis de una serie de tiempo*, y el *principal objetivo* es presentar una implementación eficiente de un procedimiento de pronóstico conocido como algoritmo de innovaciones. De manera intuitiva, una serie de tiempo es una sucesión $\{X_t\}$ de variables aleatorias definidas sobre un mismo espacio de probabilidad, donde la variable aleatoria X_t se interpreta como la observación que se realiza en el tiempo t , el cual, en el caso considerado en este trabajo, puede variar en un subconjunto de los números enteros. El rasgo fundamental de la sucesión $\{X_t\}$ es que, en contraste con el supuesto comúnmente adoptado en la teoría estadística clásica (Borovkov, 1999; Dudewicz y Mishra, 1998; Shao, 2010; Wackerly *et al.*, 2009), no se supone la independencia de las variables X_t , ni que éstas tengan la misma distribución, características que permiten incluir en el estudio una gran variedad de observaciones que surgen en la práctica, como las ventas diarias de una almacén, la asistencia semanal a teatros, la población de un país, o la aparición de manchas solares (Brockwell y Davis, 1998; Shumway y Stoffer, 2008 y Chacón, 2010). Las series que se analizan son estacionarias, en el sentido de que las propiedades relevantes de un segmento (X_1, X_2, \dots, X_n) son las mismas que las del segmento trasladado $(X_{1+h}, X_{h+2}, \dots, X_{n+h})$ para cada entero h ; como no se hace supuesto alguno sobre la distribución de los datos X_t , las propiedades importantes de la serie involucran sus momentos. Formalmente, una serie es estacionaria si

- (a) $E[X_t] = \mu$ no depende de t ;
- (b) Para cada s y t , $\text{Cov}(X_s, X_t)$ está bien definida y depende sólo de la diferencia entre s y t .

Estas propiedades permiten definir la función de autocovarianza asociada a la serie estacionaria $\{X_t\}$ mediante

$$\gamma(h) = E[X_{t+h}X_t], \quad h = 0, \pm 1, \pm 2, \dots \quad (1.1)$$

Aunque la idea de serie estacionaria es bastante restrictiva y, en general, una de tales series no es un modelo razonable para los datos que se observan comúnmente en la práctica, las series estacionarias son parte esencial del denominado *modelo clásico*, el cual si captura una amplia gama de situaciones reales (Chacón, 2010).

El problema de pronóstico para una serie estacionaria puede describirse como sigue: Dadas las observaciones X_1, \dots, X_n registradas hasta el tiempo n , determinar ‘la mejor’ aproximación

para la variable X_{n+h} que se observará en el tiempo futuro $n + h$ en términos de los datos disponibles X_1, \dots, X_n . Un resultado conocido en la teoría estadística establece que la mejor aproximación a X_{n+h} en el sentido de minimizar el error cuadrático esperado es la esperanza condicional $g(X_1, \dots, X_n) = E[X_{n+h} | X_1, \dots, X_n]$; sin embargo, al no conocer la distribución de $\{X_t\}$, la función $g(\cdot)$ no se puede calcular, y en la teoría de series de tiempo, la aproximación que se pretende encontrar es ‘la mejor aproximación lineal’ basada en los datos observados hasta el momento actual n :

$$\hat{g}(X_1, \dots, X_n) = a_1 X_1 + \dots + a_n X_n \quad (1.2)$$

donde a_1, \dots, a_n son constantes y

$$\min_{b_1, \dots, b_n} E[(X_{n+h} - (b_1 X_1 + \dots + b_n X_n))^2] = E[(X_{n+h} - (a_1 X_1 + \dots + a_n X_n))^2]. \quad (1.3)$$

tales números a_i siempre existen y se expresan solamente en términos de la función de autocovarianza $\gamma(\cdot)$; de hecho, a_1, \dots, a_n pueden determinarse por medio de diferenciación (Fulks, 1980; Khuri, 2002) e invirtiendo una matriz. Sin embargo, se conocen métodos alternativos y más eficientes para determinar pronósticos. En este trabajo el principal interés se centra en el denominado *algoritmo de innovaciones*, y el problema que se analiza es diseñar una implementación de dicho procedimiento que realice menos operaciones que la codificación usual. Este objetivo se alcanza combinando un algoritmo conocido como método de Durbin-Levinson, con un resultado de transformación entre representaciones de un pronóstico que se establece en este trabajo como Teorema 7.1, el cual representa la principal aportación técnica del artículo.

La presentación ha sido organizada de la siguiente manera. En la Sección 2 se discute la noción de proyección ortogonal en un espacio vectorial con producto interno y su relación con el problema de pronóstico, mientras que estas ideas son usadas en la Sección 3 para formular las ecuaciones de pronóstico y para encontrar los errores correspondientes. A continuación, en la Sección 4 se discute el procedimiento para calcular pronósticos conocido como método de Durbin-Levinson, y se demuestra que es de segundo orden. Posteriormente, en la Sección 5 se introduce la base ortogonal de innovaciones y en la Sección 6 se formula el algoritmo clásico correspondiente, probando que es de orden cúbico. Finalmente, en la Sección 7 se presenta un método que relaciona las expresiones para un pronóstico en términos de las bases de innovaciones y de observaciones originales, y la exposición concluye en la Sección 8 formulando una implementación del algoritmo de innovaciones que tiene orden cuadrático; el método propuesto combina el algoritmo de Durbin-Levinson con el resultado de transformación previamente obtenido.

2. Proyecciones

Puesto que este trabajo se centra en la determinación de las constantes a_1, \dots, a_n en (1.2), las cuales tienen la propiedad de optimalidad en (1.3), las ideas básicas de *espacio vectorial* y *producto interno* son de importancia fundamental en el desarrollo subsecuente; para un estudio de estos conceptos en espacios de dimensión finita, vea, por ejemplo, Graybill (2000, 2001), Harville (2008), Hoffman y Kunze (1975) y, particularmente, Strang (2003). La intuición del caso de dimensión finita es de gran importancia en el estudio de series de tiempo, aunque el entorno de trabajo es, naturalmente, el espacio

$$L^2 = L^2(\Omega, \mathcal{F}, P)$$

de todas las variables aleatorias $Y : \Omega \rightarrow \mathcal{R}$ con segundo momento finito, y donde (Ω, \mathcal{F}, P) es el espacio de probabilidad donde la serie $\{X_t\}$ está definida. Este espacio es de dimensión infinita y el producto interno está especificado por

$$\langle Y_1, Y_2 \rangle = E[Y_1 Y_2], \quad Y_1, Y_2 \in L^2, \quad (2.1)$$

y la norma correspondiente está dada por

$$\|Y\| = \sqrt{\langle Y, Y \rangle} = E[Y^2]^{1/2}. \quad (2.2)$$

Con respecto a esta norma, L^2 es un espacio métrico completo, es decir, toda sucesión de Cauchy es convergente (Apostol, 1980; Rudin, 1984; Royden, 2003). Considere ahora, una sucesión finita W_1, \dots, W_k de variables en L^2 , y sea \mathcal{W} el espacio vectorial generado por estas variables:

$$\mathcal{W} = \{c_1 W_1 + \dots + c_n W_n \mid c_1, c_2, \dots, c_n \in \mathbb{R}\}$$

el cual es un subespacio cerrado de L^2 (Rudin, 1984; Royden 2003). Dado $Y \in L^2$, existe una única variable aleatoria $P_{\mathcal{W}}Y$ que satisface las siguientes propiedades:

$$P_{\mathcal{W}}Y \in \mathcal{W} \quad \text{y} \quad \|Y - P_{\mathcal{W}}Y\| \leq \|Y - \mathbf{w}\| \quad \text{para todo } \mathbf{w} \in \mathcal{W}. \quad (2.3)$$

Estas condiciones son equivalentes a

$$P_{\mathcal{W}}Y \in \mathcal{W} \quad \text{y} \quad \langle P_{\mathcal{W}}Y, \mathbf{w} \rangle = \langle Y, \mathbf{w} \rangle, \quad \mathbf{w} \in \mathcal{W}. \quad (2.4)$$

(Brockwell y Davis, 1998; Strang, 2003); debido a la (bi-)linealidad del producto interno, estas condiciones pueden expresarse como

$$P_{\mathcal{W}}Y \in \mathcal{W} \quad \text{y} \quad \langle P_{\mathcal{W}}Y, W_i \rangle = \langle Y, W_i \rangle, \quad i = 1, 2, \dots, k. \quad (2.5)$$

La variable $P_{\mathcal{W}}Y$ se llama la proyección (ortogonal) de Y sobre \mathcal{W} y es una transformación lineal; la equivalencia de (2.3–2.5) se conoce como *teorema de proyección*. La siguiente propiedad será útil: Si \mathcal{W}_0 y \mathcal{W}_1 son dos subespacios de L^2 , entonces

$$\mathcal{W}_0 \subset \mathcal{W}_1 \Rightarrow P_{\mathcal{W}_0}P_{\mathcal{W}_1} = P_{\mathcal{W}_0} \quad (2.6)$$

A partir de esta discusión se es claro que el mejor pronóstico $g(X_1, \dots, X_n)$ para X_{n+1} dadas las observaciones X_1, \dots, X_n , es la proyección de X_{n+1} sobre el espacio generado por X_1, \dots, X_n ; compare (1.2) y (1.3) con (2.3).

3. Ecuaciones y Errores de Pronóstico

Dada una serie estacionaria $\{X_t\}$ con función de autocovarianza $\gamma(\cdot)$, en esta sección se establecen las ecuaciones para determinar la proyección ortogonal de una variable X_{n+h} , donde $h > 0$, sobre el espacio generado por las variables X_1, \dots, X_n donde, por el momento, el entero n es fijo, y el \mathcal{L}_n es el subespacio de L^2 generado por las variables X_t con $1 \leq t \leq n$:

$$\mathcal{L}_n(X_1, \dots, X_n) \equiv \mathcal{L}_n := \{a_1 X_1 + \dots + a_n X_n \mid (a_1, \dots, a_n) \in \mathbb{R}^n\}. \quad (3.1)$$

Para cualquier variable aleatoria Y con varianza finita, la proyección ortogonal de Y sobre \mathcal{L}_n se denotará mediante \hat{Y} :

$$\hat{Y} = P_{\mathcal{L}_n}Y, \quad (3.2)$$

de manera que, como se estableció en la sección precedente, \hat{Y} está caracterizada por los siguientes requerimientos:

$$\hat{Y} \in \mathcal{L}_n, \quad \text{y} \quad \langle Y, X_k \rangle = \langle \hat{Y}, X_k \rangle, \quad k = 1, 2, \dots, n; \quad (3.3)$$

en palabras, \hat{Y} es una combinación lineal de X_1, \dots, X_n , y los productos internos de \hat{Y} y Y con X_i coinciden, $i = 1, 2, \dots, n$. Ahora se determinará la forma que estas condiciones adoptan para el caso en que Y es la variable X_{n+1} que se observará una unidad de tiempo después de haber registrado el valor de X_n ; primero, se estudiará el caso $h = 1$. Observe que la inclusión $\hat{X}_{n+1} \in \mathcal{L}_n$ significa que \hat{X}_{n+1} es una combinación lineal de X_1, \dots, X_n , de manera que existen constantes $\phi_{n1}, \phi_{n2}, \dots, \phi_{nn}$ tales que

$$\hat{X}_{n+1} = \phi_{n1}X_n + \phi_{n2}X_{n-1} + \dots + \phi_{nn}X_1,$$

o de forma más compacta,

$$\hat{X}_{n+1} = \sum_{j=1}^n \phi_{nj} X_{n+1-j}, \quad (3.4)$$

A continuación, note que la condición de que X_{n+1} y \hat{X}_{n+1} tengan los mismos productos internos con las variables X_r , $r = 1, 2, \dots, n$ puede escribirse en términos de la función de autocovarianza:

$$\begin{aligned} \langle X_{n+1}, X_r \rangle &= \langle \hat{X}_{n+1}, X_r \rangle \\ &= \left\langle \sum_{j=1}^n \phi_{nj} X_{n+1-j}, X_r \right\rangle \\ &= \sum_{j=1}^n \phi_{nj} \langle X_{n+1-j}, X_r \rangle, \quad r = 1, 2, \dots, n, \end{aligned} \quad (3.5)$$

igualdades que equivalen a

$$\gamma(n+1-r) = \sum_{j=1}^n \phi_{nj} \gamma(n+1-j-r), \quad r = 1, 2, \dots, n.$$

Una forma más conveniente para este sistema se obtiene escribiendo

$$i = n+1-r;$$

con esta notación, cuando r toma cualquier valor entre 1 y n , i también lo hace y el anterior sistema de ecuaciones se escribe como

$$\gamma(i) = \sum_{j=1}^n \phi_{nj} \gamma(i-j), \quad i = 1, 2, \dots, n.$$

Definiendo los vectores $\gamma_n, \varphi_n \in \mathbb{R}^n$ y la matriz Γ_n de orden $n \times n$ mediante

$$\gamma_n = \begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(n) \end{bmatrix}, \quad \varphi_n = \begin{bmatrix} \phi_{n1} \\ \phi_{n2} \\ \vdots \\ \phi_{nn} \end{bmatrix}, \quad \Gamma_n = [\gamma(i-j)]_{i,j=1,2,\dots,n}; \quad (3.6)$$

con esta notación, las ecuaciones anteriores equivalen a

$$\gamma_n = \Gamma_n \varphi_n, \quad (3.7)$$

sistema que siempre es consistente, por el teorema de proyección. Después de encontrar una solución φ_n para esta ecuación vectorial, el error cuadrático de pronóstico, denotado por v_n y definido mediante

$$v_n = \|X_{n+1} - \hat{X}_{n+1}\|^2, \quad (3.8)$$

puede determinarse de forma muy simple. En efecto, observando que

$$\langle X_{n+1} - \hat{X}_{n+1}, X_r \rangle = 0$$

para todo $r = 1, 2, \dots, n$ (vea la primera igualdad en 3.5), como \hat{X}_{n+1} es una combinación lineal de X_1, \dots, X_n se desprende que

$$\langle X_{n+1} - \hat{X}_{n+1}, \hat{X}_{n+1} \rangle = 0, \quad (3.9)$$

lo que equivale a

$$\langle \hat{X}_{n+1}, X_{n+1} \rangle = \langle \hat{X}_{n+1}, \hat{X}_{n+1} \rangle = \|\hat{X}_{n+1}\|^2.$$

Con esto en mente,

$$\begin{aligned} \|X_{n+1} - \hat{X}_{n+1}\|^2 &= \langle X_{n+1} - \hat{X}_{n+1}, X_{n+1} - \hat{X}_{n+1} \rangle \\ &= \langle X_{n+1} - \hat{X}_{n+1}, X_{n+1} \rangle - \langle X_{n+1} - \hat{X}_{n+1}, \hat{X}_{n+1} \rangle \\ &= \langle X_{n+1} - \hat{X}_{n+1}, X_{n+1} \rangle \text{ (por la ecuación (3.9))} \\ &= \langle X_{n+1}, X_{n+1} \rangle - \langle \hat{X}_{n+1}, X_{n+1} \rangle \\ &= \gamma(0) - \langle \hat{X}_{n+1}, \hat{X}_{n+1} \rangle \end{aligned}$$

donde, de nueva cuenta, se utilizó la ecuación (3.9) para establecer la última igualdad. Por lo tanto,

$$\|X_{n+1} - \hat{X}_{n+1}\|^2 = \gamma(0) - \langle \hat{X}_{n+1}, \hat{X}_{n+1} \rangle \quad (3.10)$$

Ahora se calculará la norma cuadrática de \hat{X}_{n+1} . Usando (3.4) se tiene que

$$\begin{aligned} \langle \hat{X}_{n+1}, \hat{X}_{n+1} \rangle &= \left\langle \sum_{j=1}^n \phi_{nj} X_{n+1-j}, \sum_{i=1}^n \phi_{ni} X_{n+1-i} \right\rangle \\ &= \sum_{j=1}^n \sum_{i=1}^n \phi_{ni} \langle X_{n+1-j}, X_{n+1-i} \rangle \phi_{nj} \\ &= \sum_{j=1}^n \sum_{i=1}^n \phi_{ni} \gamma(i-j) \phi_{nj} \\ &= \boldsymbol{\varphi}'_n \boldsymbol{\Gamma}_n \boldsymbol{\varphi}_n \end{aligned}$$

y recordando que $\boldsymbol{\Gamma}_n \boldsymbol{\varphi}_n = \boldsymbol{\gamma}_n$, se obtiene que $\langle \hat{X}_{n+1}, \hat{X}_{n+1} \rangle = \boldsymbol{\varphi}'_n \boldsymbol{\gamma}_n$; combinando esta relación con (3.10), se llega a la expresión

$$v_n = \|X_{n+1} - \hat{X}_{n+1}\|^2 = \gamma(0) - \boldsymbol{\varphi}'_n \boldsymbol{\gamma}_n.$$

De esta manera, al resolver la ecuación (3.7) es posible determinar, tanto la proyección \hat{X}_{n+1} por medio de (3.4), como el error de pronóstico v_n a través de la anterior expresión desplegada. Esta discusión se resume en el siguiente teorema.

Teorema 3.1. *Suponga que $\{X_t\}$ es una serie estacionaria con función de autocovarianza $\gamma(\cdot)$. En este caso, la proyección \hat{X}_{n+1} de X_{n+1} sobre el espacio \mathcal{L}_n generado por X_1, \dots, X_n está dada por*

$$\hat{X}_{n+1} = \sum_{j=1}^n \phi_{nj} X_{n+1-j}$$

donde $\boldsymbol{\varphi}_n = (\phi_{n1}, \dots, \phi_{nn})'$ es cualquier vector que satisface el sistema de ecuaciones

$$\boldsymbol{\gamma}_n = \boldsymbol{\Gamma}_n \boldsymbol{\varphi}_n,$$

y la notación es como en (3.6). Más aún, el error de pronóstico v_n está dado por

$$v_n = \|\hat{X}_{n+1} - X_{n+1}\|^2 = \gamma(0) - \boldsymbol{\varphi}'_n \boldsymbol{\gamma}_n. \quad (3.11)$$

4. El Algoritmo de Durbin-Levinson

En esta sección se formula un procedimiento recursivo para calcular los coeficientes ϕ_{nk} en la expresión para la proyección de X_{n+1} sobre $\mathcal{L}(X_1, \dots, X_n)$. Este procedimiento calcula el vector de coeficientes φ_n y el error de pronóstico v_n en términos del vector de coeficientes φ_{n-1} y el correspondiente error de v_{n-1} , evitando el problema de invertir la matriz $\mathbf{\Gamma}_n$. El esquema de cálculo en el siguiente teorema se conoce como el algoritmo de Durbin-Levinson. En el resto del capítulo, se utiliza la siguiente notación:

$$\begin{aligned}\hat{X}_1 &= 0, & v_0 &= \|X_1 - \hat{X}_1\|^2 = \gamma(0) \\ \hat{X}_n &= P_{\mathcal{L}(X_1, \dots, X_{n-1})} X_n, & v_{n-1} &= \|X_n - \hat{X}_n\|^2, \quad n = 2, 3, \dots\end{aligned}\quad (4.1)$$

Por conveniencia, a continuación se reescribe la expresión para \hat{X}_n :

$$\hat{X}_n = \sum_{k=1}^{n-1} \phi_{n-1,k} X_{n-k}. \quad (4.2)$$

Teorema 4.1. *Suponga que $\{X_t\}$ es un proceso estacionario con función de autocovarianza $\gamma(\cdot)$ tal que $\gamma(0) > 0$ y $\gamma(h) \rightarrow 0$ conforme $h \rightarrow \infty$, de tal forma que las matrices $\mathbf{\Gamma}_n$ son no singulares. En este caso, los coeficientes ϕ_{nk} y los errores cuadráticos de pronóstico satisfacen las siguientes relaciones:*

$$(i) \quad \phi_{11} = \frac{\gamma(1)}{\gamma(0)}, \quad v_0 = \gamma(0)$$

$$(ii) \quad \phi_{nn} = \frac{1}{v_{n-1}} \left[\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j} \gamma(n-j) \right], \quad v_n = v_{n-1} [1 - \phi_{nn}^2]$$

$$(iii) \quad \phi_{nk} = \phi_{n-1,k} - \phi_{nn} \phi_{n-1,n-k}, \quad k = 1, 2, \dots, n-1$$

Una demostración de este teorema puede encontrarse en Brockwell y Davis (1998), o en Fuller (1998). Este resultado permite encontrar los coeficientes ϕ_{nk} en

$$\hat{X}_{n+1} = \sum_{j=1}^n \phi_{n,j} X_{n+1-j},$$

así como el error de pronóstico $v_n = \|X_{n+1} - \hat{X}_{n+1}\|^2$, de manera recursiva, sin invertir la matriz $\mathbf{\Gamma}_n$. El procedimiento se describe como sigue:

Algoritmo de Durbin-Levinson

Entrada: Un vector $(\gamma(0), \gamma(1), \dots, \gamma(n))$, donde $\gamma(\cdot)$ es una función de autocovarianza.

Salida: El vector $\varphi_n = (\phi_{n1}, \dots, \phi_{nn})$ de coeficientes de la proyección \hat{X}_{n+1} , así como v_n , el correspondiente error cuadrático de pronóstico.

- Inicio: Defina $\phi_{1,1} = \gamma(1)/\gamma(0)$ y $v_1 = (1 - \phi_{1,1}^2)v_0$ y ponga $k = 1$
- Paso Recursivo: En términos de ϕ_{kj} , $j = 1, 2, \dots, k$ y v_k , calcule

$$(a) \quad \phi_{k+1,k+1} = v_k^{-1} \left[\gamma(k+1) - \sum_{j=1}^k \phi_{kj} \gamma(k+1-j) \right]$$

$$(b) \quad v_{k+1} = v_k [1 - \phi_{k+1 k+1}^2]$$

$$(c) \quad \phi_{k+1 j} = \phi_{k j} - \phi_{k+1 k+1} \phi_{k n-j}, \quad j = 1, 2, \dots, k.$$

Además

$$(d) \quad \text{Incremente } k \text{ en una unidad.}$$

• Paso de Prueba: Si $k = n$ detenga el procedimiento.

Los números actuales v_k y $\phi_{k j}$, $j = 1, 2, \dots, k$ son las cantidades deseadas.

Si $k < n$, vaya al paso anterior.

A continuación se analiza el orden del algoritmo, esto es el requerimiento de memoria y la cantidad de operaciones aritméticas necesarias para obtener los coeficientes $\phi_{n k}$; se prestará atención al número de multiplicaciones y divisiones, pues estas operaciones consumen un mayor tiempo de cómputo que las sumas y restas.

• En la etapa inicial se realizan una división para evaluar $\phi_{1 1}$ y dos multiplicaciones para calcular v_1 ; en total, 3 operaciones

• En la etapa recursiva, para pasar de φ_k y v_k hacia φ_{k+1} y v_{k+1} se llevan a cabo

(a) k multiplicaciones y una división para evaluar $\phi_{k+1 k+1}$;

(b) dos multiplicaciones para calcular v_{k+1} , y

(c) k multiplicaciones para determinar las restantes componentes $\phi_{k+1, j}$, $j = 1, 2, \dots, k$.

En total se realizan $2k + 3$ multiplicaciones o divisiones.

Como la etapa iterativa se repite para $k = 1, 2, \dots, n - 1$, todo el algoritmo requiere la realización de $3 + \sum_{k=1}^{n-1} (2k + 3) = n(n + 2)$ multiplicaciones o divisiones. En este caso, se dice que el algoritmo de Durbin-Levinson es de orden n^2 , y se escribe $O(n^2)$. En cuanto a los requerimientos de memoria, se necesitan $n + 1$ lugares para almacenar $\gamma(0), \gamma(1), \dots, \gamma(n)$, n lugares adicionales para guardar v_1, \dots, v_n , y otros n lugares para almacenar $\phi_{n k}$, $k = 1, 2, \dots, n$, totalizando $3n + 1$ posiciones de memoria para los datos de entrada y salida. En esta descripción, la palabra ‘lugar’ se refiere a la cantidad de bytes necesaria para almacenar un número real con precisión determinada, y dicha cantidad depende del entorno de trabajo y del programa que ejecuta la tarea. Para propósitos de referencia futura, esta discusión se resume en el siguiente teorema.

Teorema 4.2. (i) Con respecto a las operaciones aritméticas que realiza, el algoritmo de Durbin-Levinson es de orden $O(n^2)$.

(ii) El requerimiento de memoria del algoritmo de Durbin-Levinson es de orden $O(n)$; si se desea almacenar todos los coeficientes $\phi_{s, j}$, $1 \leq j \leq s \leq n$ el requerimiento de memoria naturalmente aumenta, y es de orden $O(n^2)$.

5. Un Conjunto Ortogonal

El problema de calcular proyecciones sobre un subespacio \mathcal{L} se facilita al disponer de un conjunto generador que sea ortogonal (Hoffman y Kunze, 1980; Graybill, 2001 y Harville, 2002). En esta sección se construye uno de tales conjuntos para $\mathcal{L}(X_1, X_2, \dots, X_n)$ donde, en adelante, n es un entero positivo fijo. Considere las variables aleatorias

$$U_1 = X_1, \quad U_k = X_k - P_{\mathcal{L}(X_1, \dots, X_{k-1})} X_k, \quad k = 1, 2, \dots, n. \quad (5.1)$$

En este caso, debido a que

$$U_i \in \mathcal{L}(X_1, \dots, X_i) \quad \text{y} \quad U_j \text{ es ortogonal a } \mathcal{L}(X_1, \dots, X_{j-1}),$$

se tiene la siguiente relación de ortogonalidad:

$$i < j \Rightarrow \langle U_i, U_j \rangle = 0.$$

Suponga ahora que la matriz de varianzas $\mathbf{\Gamma}_n$ del vector (X_1, \dots, X_n) es no singular para cada n , de manera que X_1, \dots, X_n son linealmente independientes. Como U_k es una combinación lineal de X_1, \dots, X_k en la que el coeficiente de X_k es 1, se tiene que $U_k \neq 0$ para cada k . Combinando esta propiedad con la anterior relación de ortogonalidad se desprende que

$$U_1, \dots, U_n \text{ son linealmente independientes.}$$

y por lo tanto, como $U_1, \dots, U_n \in \mathcal{L}(X_1, \dots, X_n)$, un argumento directo de dimensionalidad permite establecer que

$$\mathcal{L}(U_1, \dots, U_n) = \mathcal{L}(X_1, \dots, X_n)$$

para cada n . Por otro lado, recuerde que el pronóstico para X_k en términos de X_1, \dots, X_{k-1} es $\hat{X}_k = P_{\mathcal{L}(X_1, \dots, X_{k-1})} X_k$, de manera que $U_k = X_k - \hat{X}_k$, y entonces

$$\|U_k\|^2 = \|X_k - \hat{X}_k\|^2 = v_{k-1};$$

vea (3.11). La discusión precedente se resume como sigue:

Lema 5.1. *Dada una serie estacionaria $\{X_t\}$ defina las variables aleatorias U_k mediante (5.1). En este caso,*

- (i) *Las variables U_1, U_2, U_3, \dots son ortogonales;*
- (ii) *Se satisface la igualdad $\mathcal{L}(U_1, U_2, \dots, U_k) = \mathcal{L}(X_1, X_2, \dots, X_k)$ para cada k ;*
- (iii) *Para cada $k = 1, 2, 3, \dots$, $\|U_k\|^2 = v_{k-1}$.*

La variable U_k se denomina la *innovación* en el tiempo k . Para justificar este término, note que X_k se expresa como

$$X_k = P_{\mathcal{L}(X_1, \dots, X_{k-1})} X_k + [X_k - P_{\mathcal{L}(X_1, \dots, X_{k-1})} X_k]$$

y ubíquese en el momento en que X_k va a ser observada. En ese instante, ya se registraron X_1, \dots, X_{k-1} , y el término $P_{\mathcal{L}(X_1, \dots, X_{k-1})} X_k$ es combinación lineal de esas variables; por lo tanto, $P_{\mathcal{L}(X_1, \dots, X_{k-1})} X_k$ es conocido para el observador. En contraste, $[X_k - P_{\mathcal{L}(X_1, \dots, X_{k-1})} X_k]$ es ortogonal a X_1, \dots, X_{k-1} , y en ese sentido, representa ‘la nueva información’ que el observador obtendrá al registrar el valor de X_k . Por otro lado, como el conjunto ortogonal U_1, \dots, U_n genera el mismo espacio vectorial que X_1, \dots, X_n , se tiene que la proyección

$$\hat{X}_{n+1} = P_{\mathcal{L}(X_1, \dots, X_n)} X_{n+1} = P_{\mathcal{L}(U_1, \dots, U_n)} X_{n+1}$$

se expresa como

$$\begin{aligned} \hat{X}_{n+1} &= \theta_{n1} U_n + \theta_{n2} U_{n-1} + \dots + \theta_{nn-1} U_2 + \theta_{nn} U_1 \\ &= \sum_{k=1}^n \theta_{nk} U_{n+1-k}. \end{aligned} \tag{5.2}$$

Esta es la representación de \hat{X}_{n+1} en términos de las innovaciones U_1, \dots, U_n .

6. Determinación de los Coeficientes de las Innovaciones

En esta sección se enuncia un resultado clásico sobre la determinación de los coeficientes de la expresión de \hat{X}_{n+1} como combinación lineal de las innovaciones U_1, \dots, U_k . Como punto de partida, suponga que, para cada $k = 1, \dots, n-1$, ya se han determinado los coeficientes θ_{kj} , $1 \leq j \leq k$ en las expresiones

$$\hat{X}_{k+1} = \sum_{j=1}^k \theta_{kj} U_{k+1-j} \quad (6.1)$$

así como los correspondientes errores cuadráticos de pronóstico

$$v_k = \|U_{k+1}\|^2 = \|X_{k+1} - \hat{X}_{k+1}\|^2. \quad (6.2)$$

El siguiente teorema muestra como pueden determinarse los coeficientes θ_{nk} y el error v_n en términos de estas cantidades.

Teorema 6.1. *Suponga que para cada entero $k = 1, 2, \dots, n-1$ se conocen los coeficientes θ_{kj} , $1 \leq j \leq k$, así como v_k . En este caso,*

(i) *Los coeficientes θ_{nk} en la expresión*

$$\hat{X}_{n+1} = \sum_{k=1}^n \theta_{nk} U_{n+1-k} \quad (6.3)$$

satisfacen

$$\begin{aligned} v_0 \theta_{nn} &= \gamma(n), \\ v_i \theta_{n, n-i} &= \gamma(n-i) - \sum_{s=0}^{i-1} \theta_{i, i-s} \theta_{n, n-s} v_s, \quad i = 1, 2, \dots, n-1. \end{aligned}$$

(ii) *Más aún,*

$$v_n = \gamma(0) - \sum_{j=1}^n \theta_{nj}^2 v_{n-j}$$

Una demostración de este resultado se encuentra, por ejemplo, en Brockwell y Davis (1998), o Fuller (1998). El Teorema 6.1 sugiere una manera natural de encontrar los coeficientes $\theta_{nn}, \theta_{n, n-1}, \dots, \theta_{n1}$ en la expresión (6.3) para el pronóstico \hat{X}_{n+1} , así como el correspondiente error cuadrático $v_n = \|X_{n+1} - \hat{X}_{n+1}\|^2$.

Suponga que la función de autocovarianza es tal que, para cada n , la matriz $\mathbf{\Gamma}_n$ es invertible, de manera que $v_k > 0$ para todo k . Iniciando con

$$v_0 = \gamma(0),$$

el Teorema 6.1 permite calcular

$$(i) \theta_{11} = \gamma(1)/v_0 \text{ y } v_1 = \gamma(0) - \theta_{11}^2 v_0.$$

Usando estas cantidades, ahora se calculan los coeficientes θ_{22}, θ_{21} y v_1 como sigue:

$$(ii) \theta_{22} = \gamma(2)/v_0, \quad \theta_{21} = [\gamma(1) - \theta_{11} \theta_{22} v_0]/v_1, \quad \text{y} \quad v_2 = \gamma(0) - \theta_{22}^2 v_0 - \theta_{21}^2 v_1.$$

A partir de este punto, las fórmulas del Teorema 6.1 permiten calcular $\theta_{33}, \theta_{32}, \theta_{31}$ y v_3 , en ese orden. En general, dados θ_{kj} , $1 \leq j \leq k$, y v_0, \dots, v_k , es posible evaluar $\theta_{nn}, \theta_{n, n-1}, \dots, \theta_{n1}$ y v_n , en el orden indicado. Este procedimiento se describe formalmente a continuación:

Algoritmo de Innovaciones

Entrada: Un vector $(\gamma(0), \gamma(1), \dots, \gamma(n))$, donde $\gamma(\cdot)$ es una función de autocovarianza y $n \geq 2$

Salida: El vector $\theta_n = (\theta_{n1}, \dots, \theta_{nn})$ de coeficientes de la proyección \hat{X}_{n+1} en (6.3), así como v_n , el correspondiente error cuadrático de pronóstico.

- Inicio: Defina $v_0 = \gamma(0)$, $\theta_{11} = \gamma(0)/v_0$, $v_1 = \gamma(0) - \theta_{11}^2 v_0$ y $k = 1$.
- Paso Recursivo: Incremente k en 1 y calcule

$$\begin{aligned}\theta_{kk} &= \gamma(k)/v_0, \\ \theta_{kk-i} &= [\gamma(k-i) - \sum_{s=0}^{i-1} \theta_{i-s} \theta_{kk-s} v_s] / v_i, \quad i = 1, 2, \dots, k-1, \\ v_k &= \gamma(0) - \sum_{j=1}^k \theta_{kj}^2 v_{k-j}\end{aligned}$$

- Paso de Prueba: Si $k = n$ detenga el procedimiento. Los números actuales v_k y θ_{kj} , $j = 1, 2, \dots, k$ son las cantidades deseadas.

Si $k < n$, vaya al paso anterior.

Un aspecto importante en el algoritmo de innovaciones, es que para evaluar v_k y los números θ_{kj} , es necesario tener disponibles todas las cantidades θ_{si} y v_i , $1 \leq i \leq s < k$. Así, debe reservarse lugar para todos los números en el siguiente arreglo:

$$\begin{array}{ccccccc} v_0 & & & & & & \\ \theta_{11} & v_1 & & & & & \\ \theta_{22} & \theta_{21} & v_2 & & & & \\ \theta_{33} & \theta_{32} & \theta_{31} & v_3 & & & \\ \vdots & \vdots & \vdots & \vdots & \ddots & & \\ \theta_{nn} & \theta_{nn-1} & \theta_{nn-2} & \cdots & \theta_{n1} & v_n & \end{array}$$

totalizando $1 + 2 + \dots + (n+1) = (n+1)(n+2)/2$ lugares de memoria. Además se requieren $n+1$ lugares para $\gamma(0), \gamma(1), \dots, \gamma(n)$ y uno más para la variable k . Así, el requerimiento de memoria es de $(n+2)(n+3)/2$ lugares numéricos, es decir, $O(n^2)$.

Ahora se discutirá la carga aritmética del algoritmo, contando el número de multiplicaciones y divisiones necesarias para ejecutarlo. En el paso recursivo, aplicado al entero k requiere

- Una división para evaluar θ_{kk} ;
- $2(i-1)$ multiplicaciones y una división para calcular θ_{kk-i} , dando un total de $2i-1$ operaciones; esto debe hacerse para $i = 1, 2, \dots, k-1$, arrojando un total de $\sum_{i=1}^{k-1} (2i-1) = k(k-1) - (k-1) = (k-1)^2$ multiplicaciones o divisiones. Como el paso recursivo se realiza con valores $k = 2, \dots, n$, el total de operaciones en todos los pasos, es $\sum_{k=2}^n (k-1)^2 = (n-1)n(2n-1)/6 \approx n^3/3$.

Esta discusión se resume en el siguiente teorema:

Teorema 6.2. *Para determinar los n coeficientes $\theta_{nn}, \dots, \theta_{n1}$ y el error cuadrático de pronóstico v_n , el algoritmo de innovaciones requiere $(n+2)(n+3)/2$ lugares numéricos de memoria, y la realización de $3 + (n-1)n(2n-1)/6 \approx n^3/3$ multiplicaciones o divisiones. Así, el algoritmo de innovaciones es de orden $O(n^2)$ en cuanto a requerimiento de memoria, y de orden $O(n^3)$ con respecto a la carga aritmética.*

Como ya se mencionó, el principal objetivo de este trabajo es diseñar una implementación del algoritmo de innovaciones que sea más eficiente, en el sentido de que carga aritmética sea de orden menor a n^3 . Dicho resultado depende de la relación entre los coeficientes que se utilizan para representar la proyección \hat{X}_{n+1} en términos de las innovaciones U_1, \dots, U_n , y las cantidades que se usan en la representación como combinación lineal de las variables originales X_1, \dots, X_n .

7. Relación entre dos Expresiones para un Pronóstico

EL propósito de esta sección es desarrollar la idea principal en que se sustenta la formulación de un procedimiento de segundo orden para determinar el vector de coeficientes θ_n que aparece al expresar el pronóstico \hat{X}_{n+1} en términos de las innovaciones U_1, \dots, U_n . El resultado principal relaciona dicho vector con la representación de los pronósticos \hat{X}_{k+1} en términos de las variables originales.

Considere las dos representaciones para la proyección \hat{X}_{n+1} de X_{n+1} sobre $\mathcal{L}(X_1, \dots, X_n)$:

$$\begin{aligned}\hat{X}_{n+1} &= \sum_{j=1}^n \theta_{nj} U_{n+1-j} \\ \hat{X}_{n+1} &= \sum_{j=1}^n \phi_{nj} X_{n+1-j}\end{aligned}\tag{7.1}$$

donde, como antes, U_{k+1} es la innovación en el tiempo $k+1$. El siguiente resultado es la principal contribución técnica de este trabajo.

Teorema 7.1. *Considere una serie estacionaria $\{X_t\}$ tal que la matriz Γ_k es invertible para cada entero positivo k ; vea (3.6). Sea n un entero positivo, y suponga que, para algún entero k entre 1 y n , se tiene la siguiente representación:*

$$P_{\mathcal{L}(X_1, \dots, X_k)} X_{n+1} = \sum_{i=1}^k c_i X_{k+1-i},\tag{7.2}$$

donde c_1, \dots, c_k son constantes. En este caso,

(i) El coeficiente $\theta_{n, n+1-k}$ en (7.1) está dado por

$$\theta_{n, n+1-k} = c_1.$$

(ii) La proyección X_{n+1} sobre el espacio $\mathcal{L}(X_1, \dots, X_{k-1})$ está dada por

$$P_{\mathcal{L}(X_1, \dots, X_{k-1})} X_{n+1} = \sum_{i=1}^{k-1} d_i X_{k-i},$$

donde

$$(d_1, d_2, \dots, d_{k-1}) = (c_2, c_3, \dots, c_k) + c_1(\phi_{k-1, 1}, \phi_{k-1, 2}, \dots, \phi_{k-1, k-1})$$

Demstración. Primeramente, observe que (7.2) puede escribirse como

$$\begin{aligned}P_{\mathcal{L}(X_1, \dots, X_k)} X_{n+1} &= c_1 X_k + \sum_{i=2}^k c_i X_{k+1-i} \\ &= c_1 X_k + \sum_{j=1}^{k-1} c_{j+1} X_{k-j} \\ &= c_1 [X_k - P_{\mathcal{L}(X_1, \dots, X_{k-1})} X_k] \\ &\quad + \left[\sum_{j=1}^{k-1} c_{j+1} X_{k-j} + c_1 P_{\mathcal{L}(X_1, \dots, X_{k-1})} X_k \right].\end{aligned}$$

El término entre corchetes que multiplica a c_1 es precisamente la innovación U_k , mientras que $P_{\mathcal{L}(X_1, \dots, X_{k-1})} X_k = \sum_{j=1}^{k-1} \phi_{k-1, j} X_{k-j}$. Por lo tanto,

$$\begin{aligned} P_{\mathcal{L}(X_1, \dots, X_k)} X_{n+1} &= c_1 U_k + \left[\sum_{j=1}^{k-1} c_{j+1} X_{k-j} + c_1 \sum_{j=1}^{k-1} \phi_{k-1, j} X_{k-j} \right] \\ &= c_1 U_k + \left[\sum_{j=1}^{k-1} [c_{j+1} + c_1 \phi_{k-1, j}] X_{k-j} \right] \end{aligned} \quad (7.3)$$

Por otro lado, la primera igualdad en (7.1) puede escribirse como

$$P_{\mathcal{L}(X_1, \dots, X_n)} X_{n+1} = \sum_{j=1}^n \theta_{n, j} U_{n+1-j} = \sum_{s=1}^n \theta_{n, n+1-s} U_s$$

Recuerde ahora dos hechos fundamentales:

(a) Si \mathcal{M}_0 y \mathcal{M}_1 son dos subespacios tales que $\mathcal{M}_0 \subset \mathcal{M}_1$, entonces

$$P_{\mathcal{M}_0} P_{\mathcal{M}_1} = P_{\mathcal{M}_0};$$

vea, por ejemplo, Hoffman y Kunze (1980) o Lipschutz (1995).

(b) $\mathcal{L}(X_1, \dots, X_s) = \mathcal{L}(U_1, \dots, U_s)$, igualdad que se estableció en el capítulo precedente.

Usando estas dos propiedades, y recordando que k es un entero entre 1 y n , se desprende que

$$\begin{aligned} P_{\mathcal{L}(X_1, \dots, X_k)} X_{n+1} &= P_{\mathcal{L}(X_1, \dots, X_k)} P_{\mathcal{L}(X_1, \dots, X_n)} X_{n+1} \\ &= P_{\mathcal{L}(U_1, \dots, U_k)} \left[\sum_{j=1}^n \theta_{n, j} U_{n+1-j} \right] \\ &= \sum_{j=1}^n \theta_{n, j} P_{\mathcal{L}(U_1, \dots, U_k)} U_{n+1-j} \\ &= \sum_{j=n+1-k}^n \theta_{n, j} U_{n+1-j} \end{aligned}$$

donde se usó que $P_{\mathcal{L}(U_1, \dots, U_k)} U_s = 0$ si $s > k + 1$ y $P_{\mathcal{L}(U_1, \dots, U_k)} U_s = U_s$ cuando $s \leq k$. El mismo argumento con $k - 1$ en vez de k muestra que

$$P_{\mathcal{L}(X_1, \dots, X_{k-1})} X_{n+1} = \sum_{j=n+2-k}^n \theta_{n, j} U_{n+1-j}.$$

Combinando estas dos últimas relaciones desplegadas se obtiene que

$$\begin{aligned} P_{\mathcal{L}(X_1, \dots, X_k)} X_{n+1} &= \sum_{j=n+1-k}^n \theta_{n, j} U_{n+1-j} \\ &= \theta_{n, n+1-k} U_k + \sum_{j=n+2-k}^n \theta_{n, j} U_{n+1-j} \end{aligned}$$

y entonces

$$P_{\mathcal{L}(X_1, \dots, X_k)} X_{n+1} = \theta_{n, n+1-k} U_k + P_{\mathcal{L}(X_1, \dots, X_{k-1})} X_{n+1}$$

Comparando esta expresión con (7.3), la ortogonalidad y no nulidad de las innovaciones implica que $\theta_{n,n+1-k} = c_1$, como se establece en la parte (i), mientras que

$$P_{\mathcal{L}(X_1, \dots, X_{k-1})} X_{n+1} = \sum_{j=1}^{k-1} [c_{j+1} + c_1 \phi_{k-1,j}] X_{k-j} = \sum_{j=1}^{k-1} d_j X_{k-j},$$

donde $d_j = c_{j+1} + c_1 \phi_{k-1,j}$ para $j = 1, 2, \dots, k-1$, como se establece en la parte (ii). \square

Conociendo los vectores $\varphi_k = (\phi_{k,1}, \dots, \phi_{k,k})$, $k = 1, 2, \dots, n$, este teorema permite calcular el vector $\theta_n = (\theta_{n,1}, \theta_{n,2}, \dots, \theta_{n,n})$, cuyas componentes son los coeficientes en la representación de \hat{X}_{n+1} en términos de las innovaciones U_1, \dots, U_n . Iniciando con

$$P_{\mathcal{L}(X_1, \dots, X_n)} X_{n+1} = \sum_{i=1}^n \phi_{n,i} X_{n+1-i}$$

el Teorema 7.1 con $k = n$ y $c = (c_1, \dots, c_n) = \varphi_n$ establece que $\theta_{n,n+1-n} = \theta_{n,1} = c_1 = \phi_{n,1}$, y permite calcular el vector $d = (d_1, d_2, \dots, d_{n-1})$ tal que

$$P_{\mathcal{L}(X_1, \dots, X_{n-1})} X_{n+1} = \sum_{i=1}^{n-1} d_i X_{n-i}$$

Otra aplicación del teorema precedente con $k = n-1$ y con el vector d en lugar de c arroja que $\theta_{n,n+1-(n-1)} = \theta_{n,2} = d_1$, y un vector $\tilde{d} = (\tilde{d}_1, \tilde{d}_2, \dots, \tilde{d}_{n-2})$ que satisface

$$P_{\mathcal{L}(X_1, \dots, X_{n-2})} X_{n+1} = \sum_{i=1}^{n-2} \tilde{d}_i X_{n-1-i}$$

A partir de aquí se obtiene $\theta_{n,3} = \tilde{d}_1$ y otro vector \hat{d} tal que

$$P_{\mathcal{L}(X_1, \dots, X_{n-3})} X_{n+1} = \sum_{i=1}^{n-3} \hat{d}_i X_{n-2-i}$$

Continuando con este proceso, se llega finalmente a determinar todos los coeficientes $\theta_{n,1}, \theta_{n,2}, \dots, \theta_{n,n}$. Estas ideas se presentan en el siguiente algoritmo:

Algoritmo de Transformación

Entrada: Un entero positivo n y los vectores φ_k tales que

$$\hat{X}_{k+1} = (X_1, \dots, X_k) \varphi_k$$

para cada $k = 1, \dots, n$.

Salida: El vector $\theta_n = (\theta_{n,1}, \dots, \theta_{n,n})$ tal que $\hat{X}_{k+1} = (U_1, \dots, U_k) \theta_n$.

Inicio: Ponga $k = 1$ y defina el vector $\alpha = \varphi_n$

Eta de cálculo: (a) Si $k = n$ ponga $\theta_{n,n} = \alpha_1$

(b) Si $k < n$, ponga $\theta_{n,n+1-k} = \alpha_1$, y sustituya α por el vector

$$(\alpha_2, \dots, \alpha_{n+1-k}) + \alpha_1 \varphi_{n-k}$$

e incremente k por una unidad.

Eta de salida: (a) Si $k > n$ salga del procedimiento mostrando el vector θ_n .

(b) Si $k \leq n$ vaya a la etapa de cálculo.

Para concluir la sección, se analiza el orden del procedimiento de transformación para obtener el vector θ_n .

Teorema 7.2. *El procedimiento de transformación descrito anteriormente es de orden cuadrático, tanto en el requerimiento de memoria, como en la carga aritmética.*

Demostración. Debido a que el procedimiento requiere utiliza los vectores $\varphi_1, \dots, \varphi_n$ cuyas longitudes suman $1 + 2 + \dots + n = n(n+1)/2$, es claro que el requerimiento de memoria tiene orden $O(n^2)$. Por otro lado, en la etapa de cálculo se realizan $n - k$ multiplicaciones al realizar el producto $\alpha_1 \varphi_{n-k}$, y esto se hace para $k = 1, 2, \dots, n - 1$. Por lo tanto se realizan

$$(n-1) + (n-1) + \dots + 1 = n(n-1)/2$$

multiplicaciones, así que el procedimiento también es de orden cuadrático respecto a la carga aritmética. \square

8. Implementación Cuadrática del Algoritmo de Innovaciones

En esta sección se describe un procedimiento para determinar el vector

$$\theta_n = (\theta_{n1}, \theta_{n2}, \dots, \theta_{nn})$$

en la representación

$$\hat{X}_{n+1} = \sum_{k=1}^n \theta_{nk} X_{n+1-k}$$

así como el error de pronóstico

$$v_n = \|X_{n+1} - \hat{X}_{n+1}\|^2$$

El siguiente procedimiento combina el algoritmo de Durbin-Levinson y el esquema de transformación de la sección precedente para resolver el problema, cuya solución es el principal objetivo del algoritmo de innovaciones.

Algoritmo combinado para determinar θ_n y v_n

Paso 1: Utilice el método de Durbin-Levinson para determinar v_n y los vectores φ_k , $k = 1, 2, \dots, n$, en las representaciones $\hat{X}_{k+1} = (X_1, \dots, X_k)\varphi_k$.

Paso 2: Aplique el método de transformación descrito en la sección precedente para determinar θ_n a partir de φ_k , $k = 1, 2, \dots, n$.

El siguiente teorema condensa el trabajo realizado previamente, y se enuncia para enfatizar la importancia del resultado.

Teorema 8.1. *El algoritmo combinado descrito anteriormente es de orden cuadrático, tanto en los requerimientos de memoria como en la carga aritmética.*

Demostración. Debido a que tanto el algoritmo de Durbin-Levinson como el método de transformación son de orden cuadrático en requerimientos de memoria y carga aritmética, se desprende que el esquema combinado también es de orden $O(n^2)$. \square

Este resultado indica claramente la ventaja del algoritmo combinado respecto a la formulación original del algoritmo de innovaciones. Ambos procedimientos encuentran θ_n y v_n , pero

el algoritmo combinado es de orden $O(n^2)$ y por lo tanto requiere realizar menos operaciones aritméticas que el algoritmo de innovaciones original, el cual tiene orden $O(n^3)$.

Referencias

- [1]. T. M. Apostol (1980), *Mathematical Analysis*, Addison Wesley, Reading, Massachusetts.
- [2]. A. A. Borovkov (1999), *Mathematical Statistics*, Gordon and Breach, New York
- [3]. P. J. Brockwell y R. A. Davis (1998), *Time Series: Theory and Methods*, Springer-Verlag, New York.
- [4]. J. C. Chacón Hernández (2010), *Un Criterio Integral de Causalidad Para Procesos ARMA*, Tesis de Maestría en Estadística Aplicada, Universidad Autónoma Agraria Antonio Narro, Saltillo Coah., MÉXICO.
- [5]. E. Dudewicz y S. Mishra (1998). *Mathematical Statistics*, Wiley, New York.
- [6]. W. A. Fuller (1998), *Introduction to Statistical Time Series*, Wiley, New York.
- [7]. W. Fulks (1980), *Cálculo Avanzado*, Limusa, México, D. F.
- [8]. F. A. Graybill (2000), *Theory and Application of the Linear Model*, Duxbury, New York.
- [9]. F. A. Graybill (2001), *Matrices with Applications in Statistics* Duxbury, New York.
- [10]. D. A. Harville (2008), *Matrix Algebra Form a Statistician's Perspective*, Springer-Verlag, New York.
- [11]. K. Hoffman y R. Kunze (1975), *Linear Algebra*, Prentice-Hall, New York.
- [12]. A. I. Khuri (2002), *Advanced Calculus with Applications in Statistics*, Wiley, New York.
- [13]. S. Lipschutz (1995), *Linear Algebra*, McGraw-Hill, New York.
- [14]. W. Rudin (1984), *Real and Complex Analysis*, McGraw-Hill, New York.
- [15]. H. L. Royden (2003), *Real Analysis*, MacMillan, London.
- [16]. J. Shao (2010), *Mathematical Statistics*, Springer, New York.
- [17]. R. H. Shumway y D. S. Stoffer (2006), *Time Series Analysis and Its Applications With R Examples*, Springer-Verlag, New York. Edition
- [18]. G. Strang, (2003), *Linear Algebra*, Prentice-Hall, New York.
- [19]. D. Wackerly, W. Mendenhall y R. L. Scheaffer (2009), *Mathematical Statistics with Applications*, Prentice-Hall, New York.