ALGEBRA DE FILTROS LINEALES Y EL ALGORITMO DE INNOVACIONES EN EL ANALISIS DE SERIES DE TIEMPO

JOSE LUIS FLORES AGUILAR

TESIS

PRESENTADA COMO REQUISITO PARCIAL
PARA OBTENER EL GRADO DE
MAESTRO EN CIENCIAS
EN ESTADISTICA EXPERIMENTAL

**Universidad Autonoma Agraria
Antonio Narro**

PROGRAMA DE GRADUADOS

Buenavista. Saltillo. Coah.

JUNIO DE 1999

UNIVERSIDAD AUTÓNOMA AGRARIA

ANTONIO NARRO

SUBDIRECCIÓN DE POSTGRADO

Álgebra de Filtros Lineales y el Algoritmo de Innovaciones en el Análisis de Series de Tiempo

TESIS

por

JOSÉ LUIS FLORES AGUILAR

Elaborada bajo la supervisión del Comité Particular de Asesoría y aprobada como requisito parcial para optar al grado de

MAESTRO EN CIENCIAS

en Estadística Experimental

COMITÉ PARTICULAR

Asesor Principal: ______________________
Dr. Rolando Cavazos Cadena

Asesor: ______________________
Dr. Heriberto Díaz Solís

Asesor: ______________________
M. C. Luis Rodríguez Gutiérrez

____
Dr. Ramiro López Trujillo
Subdirector Postgrado

Buenavista, Saltillo, Coahuila. Junio de 1999

AGRADECIMIENTOS

A mi esposa Silvia, por que la amo y admiro.

Al Dr. Rolando Cavazos, por ser un extraordinario maestro y por recordarme que siempre hay alguien que espera algo de amor y gratitud.

A mis Padres y hermanos, con mucho cariño.

A la Familia Polendo Luis, con todo mi afecto.

COMPENDIO

Álgebra de Filtros Lineales y el Algoritmo de Innovaciones en el Análisis de Series de Tiempo

por

JOSÉ LUIS FLORES AGUILAR

TESIS

MAESTRÍA EN CIENCIAS
en Estadística Experimental

UNIVERSIDAD AUTÓNOMA AGRARIA
ANTONIO NARRO

Buenavista, Saltillo, Coahuila, México. Junio de 1999

Dr. Rolando Cavazos Cadena — Asesor

Palabras Clave: Procesos Estacionarios, Filtros Lineales, Procesos ARMA, Algoritmo de Innovaciones, Construcción Recursiva de Pronósticos, Series de Laurent, Funciones Analíticas, Procedimiento de Ortogonalización de Gramm–Schmidt

Este trabajo trata sobre propiedades algebraicas de filtros lineales aplica-

dos al análisis de series de tiempo, y sobre la elaboración recursiva de pronósticos para observaciones futuras. Con relación al primer tema, se analiza la relación entre las operaciones de composición e inversión de filtros, y la multiplicación e inversión de funciones de variable compleja, aplicando los resultados al estudio de la noción de causalidad en procesos ARMA. Por otro lado, se demuestra la validez del procedimiento recursivo de pronósticos denominado algoritmo de innovaciones, via un análisis del procedimiento de ortogonalización de Gramm–Schmidt.

ABSTRACT

Algebra of Linear FIlters and the Innovations Algorithm
in Time Series Analysis

by

JOSÉ LUIS FLORES AGUILAR

THESIS

MASTER of SCIENCE
Experimental Statistics

UNIVERSIDAD AUTÓNOMA AGRARIA
ANTONIO NARRO

Buenavista, Saltillo, Coahuila, México. June, 1999

Dr. Rolando Cavazos Cadena — Advisor

Key Words: Stationary Processes, Linear Filters, ARMA Processes, Recursive Forecasting, Analytic Functions, Innovations Algorithm, Laurent Series, Gramm–Schmidt Orthogonalization Procedure

This work concerns algebraic properties of linear filters applied to time series analysis, and the construction of recursive forecasts. For the former topic,

the operations of composition and inversion of linear filters are asociated with the multiplication and inversion of analytic fucntions, and the results are applied to study the causality property for ARMA processes. On the other hand, for the second topic, the innovations algorithm is analyzed from a gemoetric perspective via the Gramm-Schmidt orthogonalization procedure.

CONTENIDO

Página

Capítulo 1

Perspectiva

En este capítulo se exponen los objetivos generales de esta tesis, y la forma en que se ha organizado la presentación del trabajo realizado. La intención fundamental es destacar la importancia que los problemas abordados tienen en el proceso de modelar una serie de tiempo, así como en la elaboración de pronósticos.

Introducción

El objeto de estudio de este trabajo es una sucesión de variables aleatorias que son observadas secuencialmente, las cuales se denotan mediante $\{X_t\}$ y conforman un proceso estocástico (Brockwell y Davis, 1991 y 1995); debido a que el índice t se interpreta como el momento en que el valor x_t asumido por X_t será observado, a dicha sucesión también se le refiere como *serie de tiempo*, terminología que se adoptará en lo sucesivo. En la actualidad, son inumerables los datos que se generan de manera cotidiana, y que pueden interpretarse como los valores que las componentes de una serie de tiempo asumen. Por ejemplo, cada día se hace público el tipo de cambio del peso frente al dolar norteamericano u otra divisa, quincenalmente se reporta el índice inflacionario, mensualmente se hace público el monto de las exportaciones e importaciones, a diario se informa sobre el precio del barril del petróleo en el mercado internacional, y los medios de comunicación reportan regularmente las cifras sobre la incidencia de delitos. Ejemplos similares de flujo de información se presentan en cada ámbito de la actividad humana y son abundantes en la época acutal. En cada caso, lo que se tiene es una sucesión de valores $x_1, x_2, \ldots, x_t, \ldots$, y la idea fundamental en el enfoque de series de tiempo, es interpretar dichas sucesión como una realización particular de una serie de variables aleatorias $X_1, X_2, \ldots, X_n, \ldots$ definidas en un mismo espacio de probabilidad.

Dada una serie de cifras $x_1, x_2, \ldots, x_n, \ldots$ que para un observador son relevantes, los problemas fundamentales son dos:

• Entender su estructura, es decir, averiguar si existen relaciones entre observaciones registradas en momentos distintos, a partir de las cuales sea posible establecer un modelo simple que 'explique' la evolución de las cifras.

• Formular pronósticos sobre el valor que asumirán los datos futuros, cuestión que, invariablemente, es interesante e importante

Sobre estas cuestiones trata el presente trabajo, y abordarlos constituye una tarea tan apasionante como la trascendencia que tienen los problemas considerados. Como se verá más adelante, la incursión en el área de series de tiempo, también representa una oportunidad de conjuntar técnicas de diversas áreas dentro de estadística matemática, como álgebra lineal, teoría de funciones de variable compleja, análisis de Fourier y teoría estadística en general.

Objetivos

Como ya se ha mencionado, uno de los temas que se abordarán en esta tesis está relacionado con la construcción de pronósticos. Considere, por ejemplo, el instante k en el cual ya se han observado los valores $x_1, x_2, \ldots, x_k$ asumidos por las variables $X_1, X_2, \ldots, X_k$. El problema de pronóstico consiste en utilizar los datos observados para emitir 'una aproximación' al valor que asumirá la variable X_{k+1}. En general, la formulación de un pronóstico implica una fuerte carga computacional que se incrementa conforme el número de observaciones crece. Por esta razón, es importante que un método de pronóstico utilice los cálculos realizados con anterioridad para simplificar la elaboración del siguiente pronóstico, esto es, es de gran importancia que un método sea de naturaleza recursiva. Por ejemplo, al emitir un pronóstico para la observación que tendrá en el instante 102, este debe elaborarse en base a las 101 observaciones disponibles, pero un método recursivo aporvechará los cálculos realizados cuando sólo se tenían 100 datos para simplificar la tarea. Uno de esos métodos recursivos es el denominado algoritmo de innovaciones, en cuyo análisis se centra la primera parte de este trabajo, el cual persigue alcanzar el siguiente propósito:

• *Establecer* una relación entre el algoritmo de innovaciones y el procedimiento algebraico de ortogonalización de Gramm–Schmidt, demostrando rigurosamente la validez del método de prónostico.

Por otro lado, una idea fundamental dentro de la estadística es aplicar una transformación a un objeto aleatorio determinado, con el propósito de obtener una familia de distribuciones dentro de la cual sea plausible encontrar una que describa 'adecuadamente' a un conjunto de datos. En el caso que nos ocupa los objetos aleatorios son procesos estocásticos *estacionarios* X_t, para los cuales sus rasgos esenciales, desde la óptica de segundo orden, son su función de medias y de autocovarianzas, sobre las cuales se supone que la estructura correspondiente permanece invariable al transcurrir el tiempo. El propósito en esta dirección es analizar propiedades algebraicas de las transformaciones denominadas filtros, las cuales transforman un proceso estacionario en otro que también tiene dicha propiedad, de manera especial, se pretende alcanzar los siguientes objetivos:

• *Determinar* rigurosamente la relación entre el problema de aplicar dos filtros en forma sucesiva, y la operación de multplicar series de potencias.

• *Demostrar* que un filtro tiene un inverso siempre que su serie de potencias correspondiente no se anule en el disco unitario, y *expresar* el filtro inverso en términos del recíproco de la serie de potencias asociada al filtro orignal;

• *Caracterizar* los filtros invertibles en términos de series de potencias, y utilizar el resultado obtenido para proporcionar una demostración de la existencia de procesos ARMA.

El tipo de procesos referidos en el objetivo precedente es de gran importancia en las aplicaciones de la teoría de series de tiempo (Anderson, 1981, Wei, 1990, Box y Jenkins, 1970).

Organización del Material

Para alcanzar los objetivos propuestos, el trabajo subsecuente se ha organizado de la siguiente manera: La exposición inicia en el Capítulo 2 presentando la idea de proceso estacionario desde la perspectiva de segundo orden. De manera especial, se enfatiza el hecho de que, a pesar de que los datos observados en las aplicaciones no pueden generalmente pensarse como realizaciones de un proceso que satisface la condición de estacionaridad, generalmente es posible realizar una transformación que conduzca a resultados importantes a partir del análisis de una serie estacionaria. En el Capítulo 3 se aborda el problema de construir pronósticos para una serie de tiempo desde una perspectiva de segundo orden. Bajo el supuesto de que la serie considerada tiene media nula, se analiza la relación que existe entre el procedimiento de Gramm-Schmidt, y el Algoritmo de innovaciones, demostrando rigurosamente la validez de éste último. Posteriormente, en el Capítulo 4 se introduce la clase de operadores denominados *filtros lineales*, los cuales actúan transformando una serie de tiempo en otra, de tal forma que en el proceso de cambio se preserva el rasgo de estacionaridad. En el Capítulo 5 se estudian dos problemas básicos del álgebra de filtros lineales, a saber, determinar la composición de dos filtros, y encontrar condiciones suficientes para que un filtro sea invertible, estableciendo el lazo existente entre estos problemas y la teoría de funciones analíticas. Finalmente, la presentación culmina en el Capítulo 6 con algunos comentarios breves.

Capítulo 2

Procesos Estacionarios

En este capítulo se introduce la noción de serie estacionaria desde la perspectiva de segundo orden, esto es, enfatizando la estructura de momentos de orden menor o igual a dos asociada a las variables que conforman el proceso, y se discute brevemente el desarrollo del trabajo en los siguientes capítulos.

Introducción

El objeto de estudio en este trabajo es una sucesión de variables aleatorias $\{X_t\}$, donde el índice t se interpreta como 'el tiempo', y asume valores enteros, mientras que el valor de X_t se observa en el 'instante' t. Una premisa formal básica, es que todas las variables aleatorias consideradas se encuentran definidas en un mismo espacio de probabilidad—y este caso la sucesión $\{X_t\}$ se denomina un *proceso estocástico*—supuesto que está en correspondencia con el hecho de que todas las X_t se asocian al mismo fenómeno. Debido a que el índice t se identifica como 'tiempo', es común referirse al proceso $\{X_t\}$ como una serie cronológica, o *serie de tiempo*, terminología que se adoptará en lo sucesivo.

Esta noción de serie de tiempo es sumamente genral y en el desarrollo de la teoría se restringe la atención a una clase de procesos estocásticos cuyas propiedades no se alteran conforme las diversas variables son observadas, enfatizando la idea básica de serie estacionaria (proceso estacionario) en el sentido de segundo orden. *La organización del material es la siguiente:* En la Sección 2 se introduce formalmente la idea de proceso estocástico, centrando la atención en las series estacionarias en la Sección 2, para las cuales la media de las variables que la conforman, así como las covarianzas entre pares de variables, no se alteran al trasladar el índice de tiempo; en este punto se introduce la idea central de función de autocovarianza. En la Sección 4 se abordan los problemas de estimación para el valor esperado de las variables que conforman un proceso estacionario, y para los valores de la función de autocovarianza de la serie, mientras que la exposición concluye en la Sección 5 con una descripción panorámica del desarrollo posterior de este trabajo.

Procesos Estocásticos

El propósito de esta sección es discutir la noción de *serie de tiempo* o *proceso estocástico*, alrededor de la cual gira el desarrollo de este trabajo.

Definición 2.1. Un proceso estocástico $\{X_t\}$ es una familia de variables aleatorias definidas en un mismo espacio de probabilidad $(\Omega, \mathcal{F}, P)$.

En este trabajo, se supone que el índice t representa el tiempo, y que la variable X_t se observa en el momento t. Además, el subíndice t asume cualquier valor entero, de tal forma que si el momento actual es t, X_s ya fue observada previamente si $s < t$, mientras que para $s > t$, X_s será registrada en el futuro.

Los propósitos básicos del análisis de series de tiempo son dos, a saber, (i) utilizar los datos observados hasta el tiempo n, esto es, $x_1, x_2, \ldots, x_n$, para formular inferencias sobre las propiedades de las variables $X_{n+1}, X_{n+2}, \ldots$ cuyos valores se observarán en el futuro, y (ii) Construir un modelo que describa adecuadamente los datos observados. Como ya se ha mencionado, este trabajo trata sobre el problema de pronóstico para una serie de tiempo general asumiendo que ésta tiene media constante, la cual sin pérdida de generalidad puede suponerse nula, y además se analiza un instrumento, denominado filtro (lineal), que permite generar una familia de modelos a partir de una serie de tiempo muy simple, denominada ruido blanco, la que será introducida en la siguiente sección. De manera general, los argumentos que se utilizan para abordar los temas de este trabajo, se ubican dentro de la categoría de 'segundo orden', lo cual significa que los modelos y métodos dependen sólo de características que pueden expresarse en términos de momentos de las variables aleatorias X_t de orden menor o igual a dos; como se supone que cada variable aleatoria tiene media nula, esto equivale a centrar la atención en las covarianzas de las componentes del proceso, esto es, en la función $\mathbb{K}$ dada por

$$\mathbb{K}(s,t) = \mathrm{Cov}\left[X_s, X_t\right]. \tag{2.1}$$

La Óptica de Segundo Orden

En esta sección se introduce formalmente la noción de *proceso estacionario* en el sentido de segundo orden—idea a la que también se le conoce con *estacionaridad débil*— la cual permite utilizar ideas geométrico–algebraicas en el análisis de series de tiempo.

Definición 3.1. Un proceso estocático $\{X_t\}$ es *débilmente estacionario* si satisface las siguientes condiciones:

(i) $E[X_t^2] < \infty$ para todo t, y

(ii) Para todos los índices s y t, y para todos los enteros h,

$$E[X_t] = E[X_{t+h}], \quad \text{y} \quad \mathrm{Cov}\left[X_s, X_t\right] = \mathrm{Cov}\left[X_{s+h}, X_{t+h}\right]. \tag{3.1}$$

En otras palabras, un proceso $\{X_t\}$ es estacionario en el sentido débil, si todas sus variables tienen la misma media, y la covarianza entre dos variables del proceso depende sólo de la diferencia entre sus índices; por ejemplo, la covarianza entre X_1 y X_5 es la misma que la que existe entre X_{12} y X_{16}, pues $5-1 = 16-12 = 4$.

Observación 3.1. En este trabajo, el adjetivo estacionario aplicado a un proceso estocático $\{X_t\}$ significa que la serie es *débilmente estacionaria*; otra idea de estacionaridad, la cual es mucho más restrictiva y no es frecuentemente utilizada en el estudio de series de tiempo, es la noción de estacionaridad estricta: Una serie $\{X_t\}$ es estrictamente estacionaria, si para cada entero positivo n y para cada entero h, los vectores $(X_1, X_2, \ldots, X_n)$ y $(X_{1+h}, X_{2+h}, \ldots, X_{n+h})$ tienen la misma distribución. Este concepto impone una fuerte restricción sobre la estructura probabiliística de la serie observada, y suponiendo que las variables X_t tienen segundo momento finito, una serie estrictamente estacionaris es, necesariamente, débilmente estacionaria (Brockwell y Davis, 1991 y 1996, Polendo–Luis, 1999).

Para procesos estacionarios, la función de covarianza en (2.1) adopta una forma mucho más simple.

Definición 3.2. Sea $\{X_t\}$ una serie (débilmente) estacionaria. La correspondiente *función de autocovarianza* se define mediante

$$\gamma_{\mathbf{X}}(h) = \operatorname{Cov}[X_t, X_{t+h}], \quad h = 0, \pm 1, \pm 2, \ldots \tag{3.2}$$

Note que la expresión anterior para $\gamma_{\mathbf{X}}(h)$ está bien determinada, en el sentido de que si s es cualquier otro índice, entonces $\gamma_{\mathbf{X}}(h) = \operatorname{Cov}[X_t, X_{t+h}] = \operatorname{Cov}[X_s, X_{s+h}]$, pues para un proceso estacionario, la covarianza entre dos variables X_p y X_q depende solamente de la diferencia entre p y q. En particular,

$$\operatorname{Cov}[X_s, X_t] = \gamma_{\mathbf{X}}(t - s). \tag{3.3}$$

A continuación se proporcionan dos ejemplos de series estacionarias. El primero de ellos es el más básico, y es el tipo de procesos sobre los cuales se construyen importantes modelos en la teoría de series de tiempo.

Ejemplo 3.1. Sea $\{Z_t\}$ una sucesión de variables aleatorias con segundo momento finito, y suponga que, para todo t, $E[Z_t] = 0$ y $E[Z_t^2] = \sigma^2 > 0$, donde σ es un número positivo dado, mientras que $\operatorname{Cov}[Z_t, Z_s] = 0, \quad s \neq t$. Esto significa que las variables aleatorias tienen media nula, varianza constante σ^2 y son no correlacionadas. Una sucesión con estas características se denomina ruido blanco con 'parámetros' 0 y σ^2, y se escribe

$$\{Z_t\} \sim \mathrm{WN}\left(0, \sigma^2\right).$$

No es difícil verificar que un ruido blanco es una serie estacionaria. Con esta finalidad, note que para todo s, t y h, $\mathrm{Cov}\left[Z_s, Z_t\right] = \mathrm{Cov}\left[Z_{s+h}, Z_{t+h}\right] = 0$ cuando $s \neq t$, mientras que

$$\mathrm{Cov}\left[Z_t, Z_t\right] = \mathrm{Var}\left[Z_t\right] = \sigma^2 = \mathrm{Var}\left[Z_{t+h}\right] = \mathrm{Cov}\left[Z_{t+h}, Z_{t+h}\right],$$

Luego,

$$\mathrm{Cov}\left[Z_t, Z_s\right] = \begin{cases} \sigma^2, & \text{si } t - s = 0 \\ 0, & \text{si } t - s \neq 0, \end{cases}$$

igualdad que muestra que la covarianza entre dos variables Z_t y Z_s depende sólamente de la diferencia $t - s$. Puesto que todas las variables Z_t tienen media nula, se desprende que el proceso $\{Z_t\}$ es estacionario. Note que la función de autocovarianza correspondiente al ruido blanco está dada por

$$\gamma_{\mathbf{Z}}(h) = \begin{cases} \sigma^2, & \text{si } h = 0 \\ 0, & \text{si } h \neq 0, \end{cases}$$

y que si la sucesión $\{Z_t\}$ es *i.i.d.*, con media cero y varianza común σ^2, entonces $\{Z_t\} \sim \mathrm{WN}\left(0, \sigma^2\right)$. □

Ejemplo 3.2. Suponga que $\{Z_t\} \sim \mathrm{WN}\left(0, \sigma^2\right)$ y defina un nuevo proceso mediante

$$X_t = Z_t - 2Z_{t-1}.$$

En este caso, $\{X_t\}$ es una serie estacionaria. En efecto, $E[X_t] = E[Z_t - 2Z_{t-1}] =$

$E[Z_t] - 2E[Z_{t-1}] = 0$ para todo t, mientras que

$$
\begin{aligned}
\mathrm{Cov}\,[X_s, X_t] &= \mathrm{Cov}\,[Z_s - 2Z_{s-1}, Z_t - 2Z_{t-1}] \\
&= \mathrm{Cov}\,[Z_s, Z_t] - 2\mathrm{Cov}\,[Z_{s-1}, Z_t] \\
&\qquad - 2\mathrm{Cov}\,[Z_s, Z_{t-1}] + 4\mathrm{Cov}\,[Z_{s-1}, Z_{t-1}] \\
&= \gamma_{\mathbf{Z}}(t-s) - 2\gamma_{\mathbf{Z}}(t-(s-1)) - 2\gamma_{\mathbf{Z}}((t-1)-s) \\
&\quad + 4\gamma_{\mathbf{Z}}((t-1)-(s-1)) \\
&= 5\gamma_{\mathbf{Z}}(t-s) - 2\gamma_{\mathbf{Z}}(t-s+1) - 2\gamma_{\mathbf{Z}}(t-1-s),
\end{aligned}
$$

donde se ha utilizado (3.3) para el proceso $\{Z_t\}$. Este argumento muestra que la covarianza entre X_t y X_s depende sólo de la diferencia $t-s$, y entonces $\{X_t\}$ es una serie estacionaria, y

$$
\begin{aligned}
\gamma_{\mathbf{X}}(t-s) &= \mathrm{Cov}\,[X_s, X_t] \\
&= 5\gamma_{\mathbf{Z}}(t-s) - 2\gamma_{\mathbf{Z}}(t-s+1) - 2\gamma_{\mathbf{Z}}(t-1-s)
\end{aligned}
$$

Por lo tanto, $\gamma_{\mathbf{X}}(h) = 5\gamma(h) - 2\gamma(h+1) - 2\gamma(h-1)$, y de forma explícita, el resultado del Ejemplo 3.1 muestra que

$$
\gamma_{\mathbf{X}}(h) = \begin{cases} 5, & \text{si } h = 0 \\ -2, & \text{si } h = \pm 1 \\ 0, & \text{en otro caso.} \end{cases}
$$

□

Para concluir esta sección se establecerán las propiedades básicas de una función de autocovarianza.

Lema 3.1. Sea $\gamma_{\mathbf{X}}(\cdot)$ la función de autocovarianza de una serie estacionaria $\{X_t\}$. En este caso, las siguientes propiedades son válidas.

(i) $\gamma_{\mathbf{X}}(0) = \mathrm{Var}\,[X_t]$ para todo t;

(ii) $|\gamma_{\mathbf{X}}(h)| \leq \gamma_{\mathbf{X}}(0)$ para todo h;

(iii) $\gamma_{\mathbf{X}}(h) = \gamma_{\mathbf{X}}(-h)$.

(iv) La función $\gamma_{\mathbf{X}}(\cdot)$ siempre es *no negativa definida*, esto es, para todos los enteros positivos n y para todas las constantes $a_1, a_2, \ldots, a_n$,

$$
\sum_{i=1}^{n} \sum_{j=1}^{n} a_i \gamma_{\mathbf{X}}(j-i) a_j \geq 0. \tag{3.5}
$$

(v) Más aún, suponga que ninguna combinación lineal de las variables X_t produce una variable aleatoria constante. En este caso, si el vector $[a_1, a_2, \ldots, a_n]$ no es nulo, la desigualdad es estricta en (3.5), esto es, la función $\gamma_{\mathbf{X}}(\cdot)$ es *positiva definida.*

Una demostración de ese resultado puede encontrarse, por ejemplo, en Anderson (1981), Polendo–Luis (1999), o en Brockwell y Davis (1991 y 1996).

Estimación de la Media y la Función de Autocovarianza

Los rasgos esenciales del proceso (débilmente) estacionario $\{X_t\}$ son el valor esperado de las variables que conforman la serie, y la función de autocovarianza $\gamma_{\mathbf{X}}(\cdot)$. Sin embargo, excepto para series obtenidas en simulaciones computacionales, estos rasgos de la serie no son conocidos en las aplicaciones, y en esta sección se aborda el problema de estimación para dichos parámetros. Si μ denota la esperanza común de las variables X_t, un estimador 'razonable' de este parámetro en el tiempo n, es la media muestral $\overline{X}_n$ de los datos $X_1, X_2, \ldots, X_n$ obtenidos desde el inicio del proceso de observaciones. Es bien conocido que $\overline{X}_n$ es un estimador insesgado de μ, y el resultado que se enucia en el Teorema 4.1 más adelante, establece la consistencia de este estimador bajo condiciones bastante generales.

Además de la media común de las variables que lo conforman, la otra característica esencial de un proceso estacionario $\{X_t\}$ es la función de autocovarianza $\gamma_{\mathbf{X}}(\cdot)$. Una manera natural de estimar dicha función por medio de los datos observados hasta el tiempo n se introduce a continuación.

Definición 4.1. Sea $\{X_t\}$ una serie de tiempo estacionaria. La *función de autocovarianza muestral* en el tiempo n se denota mediante $\hat{\gamma}_{\mathbf{X}}(\cdot)$ y se define mediante la siguiente expresión:

$$\widehat{\gamma}_{\mathbf{X}}(h) = \begin{cases} \dfrac{1}{n} \displaystyle\sum_{t=1}^{n-|h|} (X_{t+|h|} - \overline{X}_n)(X_t - \overline{X}_n), & \text{si } |h| < n, \\ 0, & \text{si } |h| \geq n. \end{cases}$$

Teorema 4.1. Suponga que la función de autocovariana de un proceso estacionario $\{X_t\}$ es absolutamente sumable, esto es,

$$\sum_k |\gamma(k)| < \infty$$

En este caso, $\overline{X}_n$ y $\widehat{\gamma}(\cdot)$ son estimadores consistentes de μ y $\gamma_X(\cdot)$. Más precisamente, conforme el número de observaciones n tiene a ∞, para cada $\varepsilon > 0$ se tiene que

$$\lim_{n\to\infty} P\left[\left|\overline{X}_n - \mu\right| > \varepsilon\right] = 0,$$

mientra que para cada entero h,

$$\lim_{n\to\infty} P\left[|\widehat{\gamma}(h) - \gamma_{\mathbf{X}}(h)| > \varepsilon\right] = 0.$$

Los detalles del argumento que conduce a establecer este resultado pueden encontrarse en Box y Jenkins (1970), Anderson (1981), y en Cavazos–Cadena (1994).

El Desarrollo Subsecuente

El propósito de esta sección es describir, de manera breve, el alcance y la importancia de los problemas que se considerarán en los siguientes capítulos. El primer problema que se considera en este trabajo, se refiere a la construcción de pronósticos para el valor que asumirá un término de una serie que será observado en el futuro. Considere, por ejemplo, el momento k, en el cual ya se han registrado los valores $y_1, y_2, \ldots, y_k$ asumidos por las variables $Y_1, Y_2, \ldots, Y_k$. Un pronóstico para el valor que tomará Y_{k+1}, la variable cuyo valor se conocerá en el siguiente momento de observación del proceso, es

$$\widehat{Y}_{k+1} = \mathbb{P}_k(Y_1, Y_2, \ldots, Y_k),$$

el cual expresa el pronóstico en términos de los datos disponibles en el momento k. En general, el cálculo de $\widehat{Y}_{k+1}$ implica una carga computacional considerable. Una vez que se observa el valor de Y_{k+1}, se debe encontrar un pronóstico para Y_{k+2}, el cual asumirá la forma

$$\widehat{Y}_{k+2} = \mathbb{P}_{k+1}(Y_1, Y_2, \ldots, Y_{k+1}).$$

Un problema concreto e importante en la evaluación de pronósticos, es disminuir la carga computacional requerida para su determinación. De manera más específica, se trata de aprovechar los cálculos realizados en la evaluación de $\widehat{Y}_{k+1}$ para facilitar el cómputo de $\widehat{Y}_{k+2}$, es decir, se trata de lograr una *implementación recursiva de un método de pronóstico.* Este problema se aborda en el desarrollo de esta tesis, realizando un estudio riguroso que conduce a establecer, desde una óptica que combina perspectivas geométricas y algebraicas, la validez de una técnica de pronósticos conocida como el Algoritmo de Innovaciones, la cual es sumamente poderosa y general, pues su aplicación no requiere de hipótesis de estacionaridad alguna.

Por otro lado, en este trabajo se analiza un tipo de operadores que transforman un proceso estocástico, los cuales se denominan *filtros*, y constituyen en instrumento fundamental en la construcción de modelos. De hecho, los denominados procesos autorregresivos y de promedios móviles, se construyen filtrando un ruido blanco. La importancia del procedimiento de filtrado, se debe a que a través de ellos se generan múltilpes formas para la función de autocovarianza del proceso generado, dentro de las cuales se puede seleccionar una que se ajuste a los datos observados. La intención del estudio es abordar temas sutiles, los cuales no son frecuentemente considerados a profundidad, y analizarlos utilizando las herramientas que proporciona la teoría de funciones analíticas (Alfhors, 1970).

Capítulo 3

El Algoritmo de Innovaciones

En este capítulo se aborda el problema de construir pronósticos para una serie de tiempo desde una perspectiva de segundo orden. Bajo el supuesto de que la serie considerada tiene media nula, el principal objetivo es estudiar en detalle la relación existente entre el procedimiento *algebraico* de ortogonalización de Gramm–Schmidt, y el denominado *algoritomo de innovaciones*, el cual es un método recursivo que permite establecer fórmulas explícitas para pronósticos sobre los datos que se observarán en cualquier horizonte futuro.

Introducción

Este capítulo trata sobre el problema de construir un pronóstico para una observación futura en una serie de tiempo. Desde una perspectiva de segundo orden, dicho pronóstico se construye tomando una combinación lineal adecuada de las observaciones disponibles, y construir la proyección se reduce a determinar los coeficientes involucrados (Anderson, 1981, Wei, 1990). El principal objetivo del capítulo es estudiar el procedimiento de ortogonalización clásico conocido como método de Gramm–Schmidt (Harville, 1997, Searle, 1982, Hoffman y Kunze, 1973), y establecer claramente su importancia en el problema de elaboración de pronósticos, demostrando rigurosamente la validez del procedimiento recursivo conocido como el algoritomo de innovaciones; este último es un instrumento poderoso y versátil para construir pronósticos, pues puede aplicarse a una serie cuya estructura de covarianza no es estacionaria (Brockwell y Davis, 1991 y 1996). *El plan de desarrollo del capítulo* es la siguiente: El punto de partida se ubica en la Sección 2 donde, en un contexto algebraico general, se analiza el problema de determinar la mejor aproximación a un vector dado en términos de un miembro de un subespacio, introduciéndo la idea clave de *proyección*, y destacando el papel central que desempeña el sistema de ecuaciones normales aún en un contexto abstracto. En la Sección 3 se estudia el problema de determinar el error de aproximación, y se analiza el papel que desempeñan los conjuntos de generadores ortogonales en el cálculo de una proyección. Posteriormente, en la Sección 4 se estudia el método de Gramm-Schmidt, el cual permite introducir el proceso de *innovaciones* de un proceso estocástico en la Sección 5. La Sección 6 es el núcleo de la presentación, pues en ella se establece la validez del algoritmo de innovaciones, y se señala cuidadosamente la naturaleza recursiva de los cálculos que permiten obtener pronósticos para la componente de una serie que se observará en el futuro inmediato. Este resultado se complementa en la Sección 7, donde se determina una fórmula de pronóstico para un dato que se observará en el futuro no necesariamente inmediato. Finalmente, la exposición culmina en la Sección 8 con algunos comentarios breves.

Proyección Sobre un Subespacio

Sean $\mathbf{w}$ y $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ vectores en un espacio vectorial $\mathcal{V}$, el cual se supone dotado con un producto interno $\langle \cdot, \cdot \rangle$; detalles sobre las propiedades de un producto interno que se utilizarán en los argumentos de este capítulo pueden encontrarse, por ejemplo, en Hoffman y Kunze (1973). Como es costumbre, la longitud (o norma) de un vector $\mathbf{t} \in \mathcal{V}$ se representa mediante $\|\mathbf{t}\|$, y es el único número no negativo que satisface $\|\mathbf{t}\|^2 = \langle \mathbf{t}, \mathbf{t} \rangle$. A través del desarrollo subsecuente, $\mathcal{L}$ denota al espacio generado por los vectores $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, i.e,

$$\mathcal{L} := \{\mathbf{v} \mid \mathbf{v} = \sum_{i=1}^{n} c_i \mathbf{v}_i,\ c_i \in \mathbb{R},\ i = 1, 2, \ldots, n\}. \tag{2.1}$$

Con esta notación el objetivo de esta sección es resolver el siguiente problema: *Caracterizar y determinar al vector* $\mathbf{v}^* \in \mathcal{L}$ *que dista lo menos posible del vector*

$\mathbf{w}$, esto es, $\mathbf{v}^*$ satisface

$$\mathbf{v}^* \in \mathcal{L} \quad \text{y} \quad \|\mathbf{w} - \mathbf{v}^*\|^2 = \min_{\mathbf{v} \in \mathcal{L}} \|\mathbf{w} - \mathbf{v}\|^2 . \tag{2.2}$$

Aunque este problema de aproximación está formulado en un contexto algebraico, su interpretación dentro de la teoría estadística es extremadamente importante. En efecto, (2.2) es un problema relevante en el análisis de los modelos lineales en general—incluyendo el análisis de experimentos y modelos de regresión—así como en el estudio de series de tiempo, aspecto con el cual se relaciona este trabajo. El siguiente resultado caracteriza la solución $\mathbf{v}^* \in \mathcal{L}$ de (2.2).

Teorema 2.1. (i) Si $\mathbf{v}^* \in \mathcal{L}$ es una solución del problema (2.2), entonces

$$\langle \mathbf{w}, \mathbf{v} \rangle = \langle \mathbf{v}^*, \mathbf{v} \rangle, \quad \mathbf{v} \in \mathcal{L}. \tag{2.3}$$

Recíprocamente,

(ii) Si un vector $\mathbf{v}^* \in \mathcal{L}$ satisface (2.3), entonces $\mathbf{v}^*$ es solución del problema de aproximación (2.2).

Consecuentemente,

(iii) El problema de aproximación (2.2) tiene a lo más una solución.

De acuerdo a este resultado, el problema (2.2) admite a lo más una solución $\mathbf{v}^*$, la cual es el único vector de $\mathcal{L}$ que satisface (2.3), propiedad que verbalmente se expresa como sigue: $\mathbf{w}$ y $\mathbf{v}^*$ tienen los mismos productos internos con cada vector del espacio $\mathcal{L}$.

Demostración. (i) Suponga que $\mathbf{v}^* \in \mathcal{L}$ satisface (2.2), y observe que para todo $\mathbf{v} \in \mathcal{L}$ y $a \in \mathbb{R}$ se satisface

$$\|\mathbf{w} - (\mathbf{v}^* + a\mathbf{v})\|^2 \geq \|\mathbf{w} - \mathbf{v}^*\|^2, \tag{2.4}$$

pues $\mathbf{v}^* + a\mathbf{v} \in \mathcal{L}$. Utilizando ahora la bilinealidad y simetría del producto interno, el lado izquierdo en esta desigualdad se expresa como

$$\begin{aligned}
\|\mathbf{w} - (\mathbf{v}^* + a\mathbf{v})\|^2 &= \|(\mathbf{w} - \mathbf{v}^*) - a\mathbf{v}\|^2 \\
&= \langle(\mathbf{w} - \mathbf{v}^*) - a\mathbf{v}, (\mathbf{w} - \mathbf{v}^*) - a\mathbf{v}\rangle \\
&= \langle(\mathbf{w} - \mathbf{v}^*), (\mathbf{w} - \mathbf{v}^*)\rangle - \langle a\mathbf{v}, (\mathbf{w} - \mathbf{v}^*)\rangle \\
&\quad + \langle a\mathbf{v}, a\mathbf{v}\rangle \; - \langle(\mathbf{w} - \mathbf{v}^*), a\mathbf{v}\rangle \\
&= \|\mathbf{w} - \mathbf{v}^*\|^2 + a^2\|\mathbf{v}\|^2 - 2a\langle(\mathbf{w} - \mathbf{v}^*), \mathbf{v}\rangle,
\end{aligned}$$

y combinando esta expresión con (2.4) se desprende que $\|\mathbf{w} - \mathbf{v}^*\|^2 + a^2\|\mathbf{v}\|^2 - 2a\langle(\mathbf{w} - \mathbf{v}^*), \mathbf{v}\rangle \geq \|\mathbf{w} - \mathbf{v}^*\|^2$. Cancelando ahora términos comunes en ambos lados de esta desigualdad, se obtiene

$$a^2\|\mathbf{v}\|^2 - 2a\langle\mathbf{w} - \mathbf{v}^*, \mathbf{v}\rangle \geq 0. \tag{2.5}$$

Recuerde que en este argumento $a \in \mathbb{R}$ es arbitrario, y defina la función $f\colon \mathbb{R} \to \mathbb{R}$ mediante

$$f(a) = a^2\|\mathbf{v}\|^2 - 2a\langle\mathbf{w} - \mathbf{v}^*, \mathbf{v}\rangle.$$

Claramente, esta función se anula en cero, mientras que, de acuerdo a (2.5), $f(a) \geq 0$ para todo $a \in \mathbb{R}$. Por lo tanto, $f(\cdot)$ alcanza su mínimo absoluto en $a = 0$, y entonces

$$\left.\frac{df(a)}{da}\right|_{a=0} = 0.$$

Debido a que la derivada de $f(\cdot)$ está dada por

$$\frac{df(a)}{da} = 2a\|\mathbf{v}\|^2 - 2\langle\mathbf{w} - \mathbf{v}^*, \mathbf{v}\rangle,$$

se desprende que $\langle\mathbf{w} - \mathbf{v}^*, \mathbf{v}\rangle = 0$, es decir, $\langle\mathbf{w}, \mathbf{v}\rangle = \langle\mathbf{v}^*, \mathbf{v}\rangle$. Esto completa la demostración de la parte (i), pues $\mathbf{v} \in \mathcal{L}$ es arbitrario en este argumento.

(ii) Suponga ahora que $\mathbf{v}^* \in \mathcal{L}$ satisface (2.3). Para establecer la parte (ii), se demostrará que $\mathbf{v}^*$ es solución del problema de optimización en (2.2). Con esta finalidad, observe que para cada $\mathbf{v} \in \mathcal{V}$, se tienen las siguientes igualdades:

$$\begin{aligned}
\|\mathbf{w}-\mathbf{v}\|^2 &= \|(\mathbf{w}-\mathbf{v}^*)+(\mathbf{v}^*-\mathbf{v})\|^2 \\
&= \langle (\mathbf{w}-\mathbf{v}^*)+(\mathbf{v}^*-\mathbf{v}), (\mathbf{w}-\mathbf{v}^*)+(\mathbf{v}^*-\mathbf{v})\rangle \\
&= \langle (\mathbf{w}-\mathbf{v}^*), (\mathbf{w}-\mathbf{v}^*)\rangle + 2\langle (\mathbf{w}-\mathbf{v}^*), (\mathbf{v}^*-\mathbf{v})\rangle + \langle (\mathbf{v}^*-\mathbf{v}), (\mathbf{v}^*-\mathbf{v})\rangle \\
&= \|\mathbf{w}-\mathbf{v}^*\|^2 + 2\langle (\mathbf{w}-\mathbf{v}^*), (\mathbf{v}^*-\mathbf{v})\rangle + \|\mathbf{v}^*-\mathbf{v}\|^2;
\end{aligned}$$

puesto que $\mathbf{v}^* - \mathbf{v} \in \mathcal{L}$, la condición (2.3) implica que $\langle (\mathbf{w}-\mathbf{v}^*), (\mathbf{v}^*-\mathbf{v})\rangle = 0$, y en este caso el argumento anterior implica

$$\|\mathbf{w}-\mathbf{v}\|^2 = \|\mathbf{w}-\mathbf{v}^*\|^2 + \|\mathbf{v}^*-\mathbf{v}\|^2 \geq \|\mathbf{w}-\mathbf{v}^*\|^2.$$

Puesto que $\mathbf{v} \in \mathcal{L}$ es arbitrario, a partir de esta desigualdad se desprende

$$\min_{\mathbf{v}\in\mathcal{L}} \|\mathbf{w}-\mathbf{v}\|^2 = \|\mathbf{w}-\mathbf{v}^*\|^2,$$

mostrando que $\mathbf{v}^* \in \mathcal{L}$ resuelve el problema (2.2).

(*iii*) Bajo el supuesto de que $\mathbf{v}^*$ y $\tilde{\mathbf{v}}^*$ son dos vectores de $\mathcal{L}$ que resuelven el problema (2.2), se demostrará que $\mathbf{v}^* = \tilde{\mathbf{v}}^*$. Para verificar esta afirmación, note que se ha demostrado en la parte (*i*) que cualquier solución al problema de aproximación debe satisfacer (2.3). Entonces

$$\langle \mathbf{w}, \mathbf{v}\rangle = \langle \mathbf{v}^*, \mathbf{v}\rangle, \quad \text{y} \quad \langle \mathbf{w}, \mathbf{v}\rangle = \langle \tilde{\mathbf{v}}^*, \mathbf{v}\rangle, \quad \mathbf{v} \in \mathcal{L},$$

y al tomar la diferencia de los miembros respectivos de ambas ecuaciones, se obtiene

$$0 = \langle \mathbf{v}^*, \mathbf{v}\rangle - \langle \tilde{\mathbf{v}}^*, \mathbf{v}\rangle = \langle \mathbf{v}^* - \tilde{\mathbf{v}}^*, \mathbf{v}\rangle, \quad \mathbf{v} \in \mathcal{L}.$$

Para concluir, note que $\mathbf{v}^* - \tilde{\mathbf{v}}^* \in \mathcal{L}$, pues tanto $\mathbf{v}^*$ como $\tilde{\mathbf{v}}^*$ son miembros de $\mathcal{L}$; luego, es posible poner $\mathbf{v} = \mathbf{v}^* - \tilde{\mathbf{v}}^*$ en la anterior igualdad, y al hacerlo se obtiene

$$0 = \langle \mathbf{v}^* - \tilde{\mathbf{v}}^*, \mathbf{v}^* - \tilde{\mathbf{v}}^*\rangle = \|\mathbf{v}^* - \tilde{\mathbf{v}}^*\|^2,$$

ecuación que equivale a $\mathbf{v}^* = \tilde{\mathbf{v}}^*$. □

De acuerdo al Teorema 2.1, el problema de aproximación básico establecido en (2.2) tiene a lo más una solución, la cual está caracterizada por la condición (2.3). Sin embargo, aunque esta información es sumamente valiosa, el teorema precedente no proporciona un método para construir $\mathbf{v}^*$ y, más aún, no garantiza la existencia de dicho vector. El siguiente teorema aborda estos dos últimos aspectos.

Teorema 2.2. Defina la matriz $\mathbf{G}$ de orden $n \times n$, y el vector $\mathbf{g} \in \mathbb{R}^n$ mediante

$$\mathbf{G}_{i\,j} = \langle \mathbf{v}_i, \mathbf{v}_j \rangle, \quad \mathbf{g}_i = \langle \mathbf{w}, \mathbf{v}_i \rangle, \quad i, j = 1, 2, \ldots, n. \tag{2.5}$$

Con esta notación, las siguientes afirmaciones (i) y (ii) son válidas.

(i) El sistema de ecuaciones lineales

$$\mathbf{G}\mathbf{x} = \mathbf{g} \tag{2.6}$$

es consistente.

(ii) Si $\mathbf{x} = [x_1, x_2, \ldots, x_n]'$ es una solución de (2.6), entonces el vector

$$\mathbf{v}^* := \sum_{j=1}^{n} x_j \mathbf{v}_j \tag{2.7}$$

satisface la condición (2.3), y por lo tanto, es la única solución del problema de aproximación (2.2).

Observación 2.1. (i) Este resultado permite trasladar el problema geométrico planteado en (2.2) hacia un contexto algebraico. En efecto, encontrar el vector $\mathbf{v}^*$ en (2.2) se reduce a resolver el sistema de ecuaciones lineales (2.6), el cual, de acuerdo a la parte (i), siempre tiene soluciones; después de encontrar cualquiera de ellas, sus componentes x_k se utilizan para combinar los vectores $\mathbf{v}_k$, generando al vector $\mathbf{v}^*$ como en (2.7) (Graybill, 1985).

(ii) La matriz $\mathbf{G}$ en (2.5) se conoce como *matriz de Gramm.* La razón por la que el problema (2.2), de naturaleza eminentemente geométrica, se reduce al problema algebraico de resolver (2.6), se debe a que $\mathbf{G}$ captura 'la geometría' del sistema de vectores $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$. En efecto, $\mathbf{G}_{i\,i} = \langle \mathbf{v}_i, \mathbf{v}_i \rangle = \|\mathbf{v}_i\|^2$ determina la longitud de $\mathbf{v}_i$, mientras que, si $\theta_{i\,j}$ es el ángulo entre $\mathbf{v}_i$ y $\mathbf{v}_j$, la igualdad $\mathbf{G}_{i\,j} = \langle \mathbf{v}_i, \mathbf{v}_j \rangle = \cos(\theta_{i\,j})\|\mathbf{v}_i\|\|\mathbf{v}_j\|$ muestra que todos los ángulos entre parejas de vectores $\mathbf{v}_i$ y $\mathbf{v}_j$ pueden calcularse a partir de $\mathbf{G}$. Por lo tanto, el conocimiento de $\mathbf{G}$ permite determinar el paralelepípedo $\mathcal{P}$ generado por los vectores $\mathbf{v}_k$, por lo que no es sorprendente que el volumen (n–dimensional) de $\mathcal{P}$ sea igual al (valor absoluto del) determinante de $\mathbf{G}$; en particular, la independencia lineal del sistema $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ equivale a que $\mathbf{G}$ sea una matriz no singular.

Demostración del Teorema 2.2. (i) Para verificar la consistencia del sistema (2.6), es suficiente con comprobar que

$$\text{Si } \mathbf{a} \in \mathbb{R}^n \text{ satisface } \mathbf{a}'\mathbf{G} = \mathbf{0}, \text{ entonces } \mathbf{a}'\mathbf{g} = 0, \tag{2.8}$$

afirmación que puede establecerse como sigue:

$$\begin{aligned}
\mathbf{a}'\mathbf{G} = \mathbf{0} &\Longrightarrow \mathbf{a}'\mathbf{G}\mathbf{a} = 0 \\
&\Longrightarrow \sum_{i=1}^{n}\sum_{j=1}^{n} a_i \mathbf{G}_{i\,j} a_j = 0 \\
&\Longrightarrow \sum_{i=1}^{n}\sum_{j=1}^{n} a_i \langle \mathbf{v}_i, \mathbf{v}_j \rangle a_j = 0 \quad \text{(vea (2.5))} \\
&\Longrightarrow \langle \sum_{i=1}^{n} a_i \mathbf{v}_i, \sum_{j=1}^{n} a_j \mathbf{v}_j \rangle = 0
\end{aligned}$$

donde se ha utilizado la bilinealidad del producto interno para establecer la última implicación. Luego,

$$\begin{aligned}
\mathbf{a}'\mathbf{G} = \mathbf{0} &\Longrightarrow \langle \sum_{i=1}^{n} a_i \mathbf{v}_i, \sum_{j=1}^{n} a_j \mathbf{v}_j \rangle = 0 \\
&\Longrightarrow \left\| \sum_{i=1}^{n} a_i \mathbf{v}_i \right\|^2 = 0 \\
&\Longrightarrow \sum_{i=1}^{n} a_i \mathbf{v}_i = 0 \\
&\Longrightarrow \langle \mathbf{w}, \sum_{i=1}^{n} a_i, \mathbf{v}_i \rangle = 0 \\
&\Longrightarrow \sum_{i=1}^{n} a_i \langle \mathbf{w}, \mathbf{v}_i \rangle = 0
\end{aligned}$$

y recordando que $\mathbf{g}_i = \langle \mathbf{w}, \mathbf{v}_i \rangle$, se desprende que

$$\mathbf{a}'\mathbf{G} = \mathbf{0} \Longrightarrow \sum_{i=1}^{n} a_i \langle \mathbf{w}, \mathbf{v}_i \rangle = 0 \Longrightarrow \sum_{i=1}^{n} a_i \mathbf{g}_i = 0 \Longrightarrow \mathbf{a}'\mathbf{g} = 0,$$

completando la verifición de la afirmación (2.8); como ya se ha mencionado, este argumento establece la consistencia del sistema (2.6) (Hoffman y Kunze, 1973).

(ii) Sea $\mathbf{x}$ una solución de (2.6) y defina $\mathbf{v}^*$ como en (2.7). En este caso, para cada $i = 1, 2, \ldots, n$,

$$\langle \mathbf{v}^*, \mathbf{v}_i \rangle = \langle \sum_{j=1}^{n} x_j \mathbf{v}_j, \mathbf{v}_i \rangle = \sum_{j=1}^{n} \langle \mathbf{v}_j, \mathbf{v}_i \rangle x_j = \sum_{j=1}^{n} \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_j = \sum_{j=1}^{n} \mathbf{G}_{ij} x_j = \mathbf{g}_i;$$

vea (2.6) para la última igualdad. Puesto que $\langle \mathbf{w}, \mathbf{v}_i \rangle = g_i$, se desprende que $\langle \mathbf{w}, \mathbf{v}_i \rangle = \langle \mathbf{v}^*, \mathbf{v}_i \rangle$, para todo entero i entre 1 y n, de donde, utilizando la linealidad del producto interno, se conculye que $\langle \mathbf{w}, \mathbf{v} \rangle = \langle \mathbf{v}^*, \mathbf{v} \rangle$ para todo vector que sea combinación linela de $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, esto es, para todo $\mathbf{v} \in \mathcal{L}$. □

El Teorema 2.2 muestra que siempre existe un vector $\mathbf{v}^*$ que satisface la condición (2.3), y por lo tanto es solución del problema de aproximación planteado al inicio de esta sección. Más aún, la determinación de $\mathbf{v}^*$ se reduce a resolver el sistema de ecuaciones lineales (2.6), y cualquiera de sus soluciones puede utilizarse para generar, a través de la fórmula (2.7), el vector $\mathbf{v}^* \in \mathcal{L}$ que aproxima mejor a $\mathbf{w}$; como $\mathbf{v}^*$ es único, por el Teorema 2.1(iii), no importa cual solución $\mathbf{x}$ se elija. Este hecho pone de manifiesto la importancia del sistema $\mathbf{Gx} = \mathbf{g}$, el cual expresa de forma algebraica el requerimiento geométrico $\langle \mathbf{w} - \mathbf{v}^*, \mathbf{v} \rangle = 0$ para todo $\mathbf{v} \in \mathcal{L}$, es decir, la perpendicularidad (normalidad, ortogonalidad) de $\mathbf{w} - \mathbf{v}^*$ al espacio $\mathcal{L}$, razón por la cual $\mathbf{Gx} = \mathbf{g}$ se conoce como el *sistema de ecuaciones normales.*

Definición 2.1. Si $\mathcal{L} = \mathcal{L}[\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]$ es el espacio generado por los vectores $\{\mathbf{v}_i\}$, entonces la proyección de $\mathbf{w}$ sobre $\mathcal{L}$ se denota por $P_{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n}\mathbf{w} \equiv P_{\mathcal{L}}\mathbf{w}$, y se define mediante

$$P_{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n}\mathbf{w} \equiv P_{\mathcal{L}}\mathbf{w} = \mathbf{v}^*,$$

donde $\mathbf{v}^* \in \mathcal{L}$ es el único vector que satisface (2.3).

El siguiente corolario establece dos propiedades fundamentales del operador de proyección $P_{\mathcal{L}}$.

Corolario 2.1. (i) La transformación $\mathbf{w} \mapsto P_{\mathcal{L}}\mathbf{w}$ es lineal en $\mathbf{w} \in \mathcal{V}$, y

(ii) $\mathbf{w} \in \mathcal{L} \iff P_{\mathcal{L}}\mathbf{w} = \mathbf{w}$.

(iii) $P_{\mathcal{L}}\mathbf{w} = \mathbf{0} \iff \mathbf{w}$ es ortogonal a cada $\mathbf{v} \in \mathcal{L}$.

Demostración. Dados dos vectores $\mathbf{w}_1, \mathbf{w}_2 \in \mathcal{V}$, denote mediante $\mathbf{v}_1^*$ y $\mathbf{v}_2^*$ a las proyecciones correspondientes sobre el espacio $\mathcal{L}$, i.e., $\mathbf{v}_1^* = P_{\mathcal{L}}\mathbf{w}_1$ y $\mathbf{v}_2^* = P_{\mathcal{L}}\mathbf{w}_2$. Para dos constantes reales arbitrarias a_1 y a_2, debe comprobarse que $a_1\mathbf{v}_1^* + a_2\mathbf{v}_2^* = P_{\mathcal{L}}[a_1\mathbf{w}_1 + a_2\mathbf{w}_2]$. Con este fin, observe que para cada $\mathbf{v} \in \mathcal{V}$, el Teorema 2.1$(i)$ y la linealidad del producto interno implican

$$\begin{aligned}\langle a_1\mathbf{w}_1 + a_2\mathbf{w}_2, \mathbf{v} \rangle &= a_1\langle \mathbf{w}_1, \mathbf{v} \rangle + a_2\langle \mathbf{w}_2, \mathbf{v} \rangle \\ &= a_1\langle \mathbf{v}_1^*, \mathbf{v} \rangle + a_2\langle \mathbf{v}_2^*, \mathbf{v} \rangle = \langle a_1\mathbf{v}_1^* + a_2\mathbf{v}_2^*, \mathbf{v} \rangle,\end{aligned}$$

y a partir del Teorema 2.1(ii) se desprende que

$$P_{\mathcal{L}}[a_1\mathbf{w}_1 + a_2\mathbf{w}_2] = a_1\mathbf{v}_1^* + a_2\mathbf{v}_2^* = a_1P_{\mathcal{L}}\mathbf{w}_1 + a_2P_{\mathcal{L}}\mathbf{w}_2,$$

lo cual establece la linealidad de $P_{\mathcal{L}}$, mientras que la parte (ii) se desprende de inmediato de la caracterización de $\mathbf{v}^* = P_{\mathcal{L}}\mathbf{w}$ como el único vector de $\mathcal{L}$ que satisface (2.3).

(iii) Suponga que $\mathbf{v}^* = P_{\mathcal{L}}\mathbf{w} = \mathbf{0}$. En este caso, (2.3) implica que para cada $\langle \mathbf{w}, \mathbf{v} \rangle = \langle \mathbf{v}^*, \mathbf{v} \rangle = \langle \mathbf{0}, \mathbf{v} \rangle = 0$, para cada $\mathbf{v} \in \mathcal{L}$, i.e., $\mathbf{w}$ es ortogonal a cada mimbro de $\mathcal{L}$. Recíprocamente, suponga que $\mathbf{w}$ satisface $\langle \mathbf{w}, \mathbf{v} \rangle = 0$ siempre que $\mathbf{v} \in \mathcal{L}$. En estas circunstancias es claro que $\langle \mathbf{w}, \mathbf{v} \rangle = 0 = \langle \mathbf{0}, \mathbf{v} \rangle$ para todo $\mathbf{v} \in \mathcal{L}$, de manera que (2.3) se satisface con $\mathbf{v}^* = \mathbf{0}$. Luego, $P_{\mathcal{L}}\mathbf{w} = \mathbf{0}$. □

La Distancia Mínima y Cálculo de la Proyección

El propósito de esta sección es discutir brevemente el cómputo de la distancia mínima de un vector $\mathbf{w}$ al subespacio $\mathcal{L}$, así como la determinación práctica del vector de proyección $\mathbf{v}^* = P_{\mathcal{L}}\mathbf{w}$.

Lema 3.1. La distancia entre $\mathbf{w}$ y su proyección $\mathbf{v}^* = P_{\mathcal{L}}\mathbf{w}$ está determinada por cualquiera de las siguientes expresiones (i)–(iii), donde la matriz de Gramm $\mathbf{G}$ y el vector $\mathbf{g}$ están dados en (2.5), y $\mathbf{x}$ es cualquier solución del sistema de ecuaciones normales (2.6).

(i) $\|\mathbf{w} - P_{\mathcal{L}}\mathbf{w}\|^2 = \|\mathbf{w}\|^2 - \|P_{\mathcal{L}}\mathbf{w}\|^2 = \|\mathbf{w}\|^2 - \|\mathbf{v}^*\|^2$;

(ii) $\|\mathbf{w} - P_{\mathcal{L}}\mathbf{w}\|^2 = \|\mathbf{w}\|^2 - \mathbf{x}'\mathbf{G}\mathbf{x}$;

(iii) $\|\mathbf{w} - P_{\mathcal{L}}\mathbf{w}\|^2 = \|\mathbf{w}\|^2 - \mathbf{x}'\mathbf{g}$.

Demostración. (i) Observe que a partir de (2.3) con $\mathbf{v} = \mathbf{v}^* = P_{\mathcal{L}}\mathbf{w} \in \mathcal{L}$ se obtiene

$$\langle \mathbf{w} - P_{\mathcal{L}}\mathbf{w}, P_{\mathcal{L}}\mathbf{w} \rangle = 0,$$

y entonces

$$
\begin{aligned}
\|\mathbf{w} - P_{\mathcal{L}}\mathbf{w}\|^2 &= \langle \mathbf{w} - P_{\mathcal{L}}\mathbf{w}, \mathbf{w} - P_{\mathcal{L}}\mathbf{w} \rangle \\
&= \langle \mathbf{w} - P_{\mathcal{L}}\mathbf{w}, \mathbf{w} \rangle - \langle \mathbf{w} - P_{\mathcal{L}}\mathbf{w}, P_{\mathcal{L}}\mathbf{w} \rangle \\
&= \langle \mathbf{w} - P_{\mathcal{L}}\mathbf{w}, \mathbf{w} \rangle \\
&= \langle \mathbf{w}, \mathbf{w} \rangle - \langle P_{\mathcal{L}}\mathbf{w}, \mathbf{w} \rangle = \|\mathbf{w}\|^2 - \langle P_{\mathcal{L}}\mathbf{w}, \mathbf{w} \rangle.
\end{aligned}
\tag{3.1}
$$

Para concluir, utilice de nueva cuenta (2.3), pero ahora con $\mathbf{v} = \mathbf{v}^* = P_{\mathcal{L}}\mathbf{w}$, para obtener via la simetría del producto interno,

$$\langle P_{\mathcal{L}}\mathbf{w}, \mathbf{w} \rangle = \langle \mathbf{v}^*, \mathbf{w} \rangle = \langle \mathbf{v}^*, \mathbf{v}^* \rangle = \|\mathbf{v}^*\|^2,$$

y combinando esta ecuación con la última igualdad en (3.1) se arriba a la conclusión deseada.

(ii) A partir de la expresión (2.7) para la proyección $\mathbf{v}^*$, se tiene

$$\|\mathbf{v}^*\|^2 = \langle \mathbf{v}^*, \mathbf{v}^* \rangle = \langle \sum_{i=1}^n x_i \mathbf{v}_i, \sum_{j=1}^n x_j \mathbf{v}_j \rangle = \sum_{i=1}^n \sum_{j=1}^n x_i \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_j = \mathbf{x}'\mathbf{G}\mathbf{x},$$

ecuación que, a través de la parte (i), implica $\|\mathbf{w} - P_{\mathcal{L}}\mathbf{w}\|^2 = \|\mathbf{w}\|^2 - \mathbf{x}'\mathbf{G}\mathbf{x}$.

(iii) La conclusión se obtiene combinando la parte (ii) con el sistema de ecuaciones normales; vea (2.6). □

Hasta este punto, el desarrollo del capítulo ha tenido como referencia un marco de trabajo abstracto, es decir, un espacio vectorial general dotado con un producto interno, característica que, en cierto sentido, representa una ventaja. En efecto, los resultados establecidos pueden aplicarse en diversos contextos, y en este momento es oportuno señalar su rasgo principal: El problema gemétrico (2.2) se reduce, *aún dentro de un contexto abstracto*, a resolver el sistema de ecuaciones normales (2.6); dicho sistema, que surge constante y naturalmente en la metodología estadística, tiene solución única cuando los vectores $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ son linealmente independientes, pues en este caso $\mathbf{G}$ es una matriz no singular y

$\mathbf{x} = \mathbf{G}^{-1}\mathbf{g}$ (vea la Observación 2.1(ii)). Por otro lado, el caso en que $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ no es un sistema linealmente independiente se presenta con frecuencia, especialmente en el área de diseños experimentales; en estas circunstancias la matriz $\mathbf{G}$ es singular, de tal forma que $\mathbf{Gx} = \mathbf{g}$, siendo un sistema consistente, tiene múltiples soluciones. Sin embargo, en este último caso, el Teorema 2.2 garantiza que el vector $\mathbf{v}^*$ obtenido mediante la expresión (2.7) siempre será el mismo sin importar cual sea la solución $\mathbf{x}$ que se utilice en su construcción. Por lo tanto, lo realmente importante es encontrar *una* solución al sistema de ecuaciones normales, por lo que es suficiente seleccionar $\mathbf{x} = \mathbf{G}^c\mathbf{g}$, donde $\mathbf{G}^c$ es una *inversa condicional* de $\mathbf{G}$. En la literatura se reportan diversos diversos métodos para resolver (2.6), como el método de triangulación de Cholesky (Graybill, 1985), o el Método de Hausholder (Harville, 1997). En el apéndice de este trabajo se incluye un procedimento que permite determinar una inversa condicional para una matriz arbitraria $\mathbf{G}$; aunque dicho método no será utilizado más adelante, ha sido incluido por tener un interés en si mismo, pues utiliza sólamente operaciones elementales de fila, y no se ha detectado una formulación explícita de esta técnica en la literatura.

Después de estos comentarios, es oportuno tomar un respiro en el desarrollo abstracto, e ilustrar la aplicación de los Teoremas 2.1 y 2.2, así como del Lema 3.1, en un caso concreto.

Ejemplo 3.1. Suponga que el espacio $\mathcal{V}$ es $\mathbb{R}^4$, el cual está dotado con el producto interno canónico, es decir,

$$\langle \mathbf{a}, \mathbf{b} \rangle = \sum_{i=1}^{4} a_i b_i, \quad \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} \in \mathbb{R}^4.$$

En este contexto, considere los vectores

$$\mathbf{v}_1 = \begin{bmatrix} 5 \\ 3 \\ 3 \\ 5 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 0 \\ 0 \\ 2 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 3 \\ 5 \\ 5 \\ 3 \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} 0 \\ 3 \\ 3 \\ -2 \end{bmatrix}.$$

Con estos datos, se busca determinar la proyección $P_{\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3}\mathbf{w}$, así como el error de aproximación $\|\mathbf{w} - P_{\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3}\mathbf{w}\|$. Para alcanzar estos objetivos, el punto de partida sugerido por el Teorema 2.2 es establecer el sistema de ecuaciones normales. Utilizando las especificaciones de la la matriz $\mathbf{G}$ y el vector $\mathbf{g}$ en (2.5) con los datos actuales, no es difícil ver que el sistema (2.6) adopta la forma

$$\begin{bmatrix} 68 & 20 & 60 \\ 20 & 8 & 12 \\ 60 & 12 & 68 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 8 \\ -4 \\ 24 \end{bmatrix}, \tag{3.2}$$

el cual tiene múltiples soluciones, dos de las cuales son las siguientes:

$$\mathbf{x} = \begin{bmatrix} 1 \\ -3 \\ 0 \end{bmatrix}, \quad \text{y} \quad \tilde{\mathbf{x}} = \begin{bmatrix} -3/2 \\ 1 \\ 3/2 \end{bmatrix}. \tag{3.3}$$

• Utilizando la solución $\mathbf{x}$, se obtiene que la proyección de $\mathbf{w}$ sobre $\mathcal{L}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ está determinada por

$$\mathbf{v}^* = 1\mathbf{v}_1 - 3\mathbf{v}_2 + 0\mathbf{v}_3 = \begin{bmatrix} 5 \\ 3 \\ 3 \\ 5 \end{bmatrix} - 3\begin{bmatrix} 2 \\ 0 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \\ 3 \\ -1 \end{bmatrix} \tag{3.4}$$

(vea (2.7)), mientras que el error de aproximación puede determinarse mediante cualquiera de las expresiones en el Lemma 3.1:

(a) Primero se empleará la fórmula $\|\mathbf{w} - P_{\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3}\mathbf{w}\|^2 = \|\mathbf{w}\|^2 - \|\mathbf{v}^*\|^2$. Puesto que

$$\|\mathbf{w}\|^2 = \left\| \begin{bmatrix} 0 \\ 3 \\ 3 \\ -2 \end{bmatrix} \right\|^2 = 22, \quad \text{y} \quad \|\mathbf{v}^*\|^2 = \left\| \begin{bmatrix} -1 \\ 3 \\ 3 \\ -1 \end{bmatrix} \right\|^2 = 20,$$

se obtiene $\|\mathbf{w} - P_{\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3}\mathbf{w}\|^2 = 22 - 20 = 2$.

(b) Para aplicar el Lema 3.1 (ii), primero note que

$$\begin{aligned} \mathbf{x}'\mathbf{G}\mathbf{x} &= [1, -3, 0] \begin{bmatrix} 68 & 20 & 60 \\ 20 & 8 & 12 \\ 60 & 12 & 68 \end{bmatrix} \begin{bmatrix} 1 \\ -3 \\ 0 \end{bmatrix} \\ &= (1)^2(68) + 2(1)(-3)(20) + (-3)^2(8) = 68 - 120 + 72 = 20, \end{aligned}$$

de donde se desprende que $\|\mathbf{w} - P_{\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3}\mathbf{w}\|^2 = \|\mathbf{w}\|^2 - \mathbf{x}'\mathbf{G}\mathbf{x} = 22 - 20 = 2$.

(c) Note ahora que 'el lado derecho' de las ecuaciones normales (3.2) es el vector $\mathbf{g}$, de manera que

$$\mathbf{x}'\mathbf{g} = [1, -3, 0]\begin{bmatrix} 8 \\ -4 \\ 24 \end{bmatrix} = 1(8) + (-3)(-4) = 20$$

y la fórmula en el Lema 3.1(iii) produce $\|\mathbf{w} - P_{\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3}\mathbf{w}\|^2 = \|\mathbf{w}\|^2 - \mathbf{x}'\mathbf{G}\mathbf{x} = 22 - 20 = 2$.

Como lo establece el Lemma 3.1, las tres expresiones para el cuadrado del error de aproximación coinciden. Por otro lado, el vector $\tilde{\mathbf{x}}$ en (3.3) también satisface el sistema de ecuaciones normales, de manera que la proyección $\mathbf{v}^*$ puede encontrarse combinando los vectores $\mathbf{v}_i$ con las componentes de la solución $\tilde{\mathbf{x}}$; al hacerlo, se obtiene

$$\begin{aligned}
\mathbf{v}^* &= -\frac{3}{2}\mathbf{v}_1 + 1\mathbf{v}_2 + \frac{3}{2}\mathbf{v}_3 \\
&= -\frac{3}{2}\begin{bmatrix} 5 \\ 3 \\ 3 \\ 5 \end{bmatrix} + \begin{bmatrix} 2 \\ 0 \\ 0 \\ 2 \end{bmatrix} + \frac{3}{2}\begin{bmatrix} 3 \\ 5 \\ 5 \\ 3 \end{bmatrix} \\
&= \begin{bmatrix} -15/2 \\ -9/2 \\ -9/2 \\ -15/2 \end{bmatrix} + \begin{bmatrix} 4/2 \\ 0 \\ 0 \\ 4/2 \end{bmatrix} + \begin{bmatrix} 9/2 \\ 15/2 \\ 15/2 \\ 9/2 \end{bmatrix} \\
&= \begin{bmatrix} [-15 + 4 + 9]/2 \\ [-9 + 0 + 15]/2 \\ [-9 + 0 + 15]/2] \\ [-15 + 4 + 9]/2 \end{bmatrix} = \begin{bmatrix} -2/2 \\ 6/2 \\ 6/2 \\ -2/2 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \\ 3 \\ -1 \end{bmatrix}
\end{aligned}$$

expresión que, en concordancia con el Teorema 2.2, coincide con (3.4). □

Como ya se ha mencionado, el sistema de ecuaciones normales es extremadamente importante en el análisis estadístico, y es conveniente distinguir un caso en el cual es posible determinar sus soluciones de manera simple.

Lema 3.2. Suponga que los vectores $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ son ortogonales, esto es,

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0 \quad i \neq j. \tag{3.5}$$

Bajo esta condición, todas las soluciones del sistema de ecuaciones normales $\mathbf{Gx} = \mathbf{g}$ están determinadas por la siguiente expresión:

$$x_i = \begin{cases} \dfrac{\langle \mathbf{w}, \mathbf{v}_i \rangle}{\langle \mathbf{v}_i, \mathbf{v}_i \rangle} = \dfrac{\langle \mathbf{w}, \mathbf{v}_i \rangle}{\|\mathbf{v}_i\|^2}, & \text{si } \mathbf{v}_i \neq \mathbf{0} \\ \text{un número arbitrario}, & \text{si } \mathbf{v}_i = \mathbf{0}, \end{cases} \tag{3.6}$$

Más aún, la proyección de $\mathbf{w}$ sobre $\mathcal{L}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ está dado por

$$\mathbf{v}^* = \sum_{i:\mathbf{v}_i \neq \mathbf{0}} x_i \mathbf{v}_i. \tag{3.7}$$

Demostración. Bajo la condición (3.5), la definición de la matriz de Gramm y del vector $\mathbf{g}$ en (2.5) muestra que $\mathbf{G}$ es una matriz diagonal con componente $\langle \mathbf{v}_i, \mathbf{v}_i \rangle = \|\mathbf{v}_i\|^2$ el la i–ésima posición de su línea diagonal principal. Luego, las ecuaciones normales en (2.6) se reducen a

$$\|\mathbf{v}_i\|^2 x_i = g_i = \langle \mathbf{w}, \mathbf{v}_i \rangle, \quad i = 1, 2, \ldots, n;$$

cuando $\mathbf{v}_i = \mathbf{0}$, x_i en esta igualdad puede ser cualquier número, mientras que si $\mathbf{v}_i \neq \mathbf{0}$, entonces $x_i = \dfrac{\langle \mathbf{w}, \mathbf{v}_i \rangle}{\|\mathbf{v}_i\|^2}$, estableciendo (3.6), mientras que utilizando la expresión (2.7) se desprende que $\mathbf{v}^* = \sum_{i:\mathbf{v}_i \neq \mathbf{0}} x_i \mathbf{v}_i$. □

Continuación del Ejemplo 3.1. De acuerdo al Lema 3.2, el problema de determinar una proyección sobre un subespacio $\mathcal{L}$ se simplifica sustancialemte cuando se dispone de un conjunto ortogonal de generadores. Considere los datos del Ejemplo 3.1 considerado anteriormente, esto es,

$$\mathbf{v}_1 = \begin{bmatrix} 5 \\ 3 \\ 3 \\ 5 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 0 \\ 0 \\ 2 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 3 \\ 5 \\ 5 \\ 3 \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} 0 \\ 3 \\ 3 \\ -2 \end{bmatrix}.$$

A continuación se construira la proyección de $\mathbf{w}$ sobre $\mathcal{L} = \mathcal{L}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ utilizando el Lema 3.2; para hacerlo, el requisito indispensable es disponer de vectores ortogonales que generen a $\mathcal{L}$. Considere los siguientes vectores $\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2$:

$$\tilde{\mathbf{v}}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \tilde{\mathbf{v}}_2 = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}. \tag{3.8}$$

Observando que

$$\tilde{\mathbf{v}}_1 = \frac{1}{8}[\mathbf{v}_1 + \mathbf{v}_3], \quad \tilde{\mathbf{v}}_2 = -\frac{1}{3}[\mathbf{v}_1 - 4\mathbf{v}_2],$$

se desprende que $\mathcal{L}\{\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2\} \subset \mathcal{L}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. Además, las igualdades

$$\mathbf{v}_1 = 4\tilde{\mathbf{v}}_1 + \tilde{\mathbf{v}}_2, \quad \mathbf{v}_2 = \tilde{\mathbf{v}}_1 + \tilde{\mathbf{v}}_2, \quad \text{y} \quad \mathbf{v}_3 = 4\tilde{\mathbf{v}}_1 - \tilde{\mathbf{v}}_2,$$

muestran que $\mathcal{L}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} \subset \mathcal{L}\{\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2\}$. Por lo tanto, $\mathbf{v}_1, \mathbf{v}_2$ y $\mathbf{v}_3$ generan el mismo subespacio que $\tilde{\mathbf{v}}_1$ y $\tilde{\mathbf{v}}_2$. Por otro lado, las expresiones (3.8) implican, de forma inmediata, que $\langle \tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2 \rangle = 0$, de manera que es posible aplicar el Lema 3.1 con $\tilde{\mathbf{v}}_1$ y $\tilde{\mathbf{v}}_2$ en lugar de $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$. Al hacerlo, se desprende que la proyección de $\mathbf{w}$ sobre $\mathcal{L}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} = \mathcal{L}\{\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2\}$ es $\mathbf{v}^* = x_1\tilde{\mathbf{v}}_1 + x_2\tilde{\mathbf{v}}_2$, donde

$$\begin{aligned} x_1 &= \frac{\langle \mathbf{w}, \tilde{\mathbf{v}}_1 \rangle}{\|\tilde{\mathbf{v}}_1\|^2} = \frac{4}{4} = 1 \\ x_2 &= \frac{\langle \mathbf{w}, \tilde{\mathbf{v}}_2 \rangle}{\|\tilde{\mathbf{v}}_2\|^2} = -\frac{8}{4} = -2 \end{aligned}$$

y por lo tanto

$$\mathbf{v}^* = \tilde{\mathbf{v}}_1 - 2\tilde{\mathbf{v}}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - 2\begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \\ 3 \\ -1 \end{bmatrix},$$

resultado que coincide con el encontrado anteriormente. Note, sin embargo, que en este caso no es necesario resolver ningun conjunto de ecuaciones, pues la ortogonalidad de los generadores hace posible obtener los coeficientes en la expresión para la proyección por medio de (3.6). □

El ejemplo anterior ilustra el hecho de que la tarea de determinar una proyección sobre un subespacio $\mathcal{L}$ se simplifica si se dispone de vectores ortogonales que generan a $\mathcal{L}$. Sin embargo, para poder utilizar el Lema 3.2 de forma eficiente, se debe contar con un conjunto ortogonal de generadores, y su construcción se aborda en la siguiente sección.

La Técnica de Ortogonalización de Gramm–Schmidt

En esta sección se presenta un procedimento clásico que permite obtener un conjunto ortogonal de vectores a partir de un conjunto arbitrario; al establecer dicho método de manera formal, se sentarán las bases para analizar el problema de pronóstico más adelante.

A través de esta sección, $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ son vectores fijos pertenecientes a un espacio vectorial dotado de un producto interno, y el espacio generado por los primeros k vectores $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ se denota mediante $\mathcal{L}_k$, esto es,

$$\mathcal{L}_k := \mathcal{L}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}. \tag{4.1}$$

El siguiente teorema describe un procedimiento para generar vectores ortogonales a partir de $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ (Hoffman, 1973, Searle, 1981, Graybill, 1985).

Teorema 4.1. [El Proceso de Ortogonalización de Gramm–Schmidt.] Sea $\mathcal{L}_0$ el espacio vectorial que consiste solamente del vector nulo, y defina los vectores $\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \ldots, \tilde{\mathbf{v}}_n$ mediante

$$\tilde{\mathbf{v}}_k = \mathbf{v}_k - P_{\mathcal{L}_{k-1}}\mathbf{v}_k, \quad k = 1, 2, \ldots, n. \tag{4.2}$$

En este caso, las siguientes afirmaciones (i)—(iii) son válidas.

(i) $\mathcal{L}_k = \mathcal{L}\{\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \ldots, \tilde{\mathbf{v}}_k\}$ para cada $k = 1, 2, \ldots, n$;

(ii) Los vectores $\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \ldots, \tilde{\mathbf{v}}_n$ son ortogonales.

(iii) Los vectores $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ son dependientes si y sólo si $\tilde{\mathbf{v}}_k = \mathbf{0}$ para algún k.

Demostración. (i) Observe que $P_{\mathcal{L}_0}\mathbf{v}_1 = \mathbf{0}$, pues $\mathcal{L}_0$ es el espacio nulo, y entonces

$$\tilde{\mathbf{v}}_1 = \mathbf{v}_1 - P_{\mathcal{L}_0}\mathbf{v}_1 = \mathbf{v}_1. \tag{4.3}$$

Defina ahora el conjunto $\mathcal{C}$ como sigue:

$$\mathcal{C} = \{k \mid \mathcal{L}_k \neq \mathcal{L}\{\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \ldots, \tilde{\mathbf{v}}_k\}, \tag{4.4}$$

de tal forma que $\mathcal{C}$ consiste de todos los enteros k para los cuales la igualdad $\mathcal{L}_k = \mathcal{L}\{\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \ldots, \tilde{\mathbf{v}}_k\}$ no se satisface. Luego, para establecer la parte (i) es claramente suficiente demostrar que

$$\mathcal{C} = \emptyset. \tag{4.5}$$

Parra arribar a esta conclusión, note que a partir de (4.3) se desprende que

$$\mathcal{L}_1 = \mathcal{L}\{\mathbf{v}_1\} = \mathcal{L}\{\tilde{\mathbf{v}}_1\}, \tag{4.6}$$

relación que equivale a $1 \notin \mathcal{C}$; vea (4.4). El argumento continua ahora por el método de reducción al absurdo. *Suponga* que $\mathcal{C} \neq \emptyset$, y sea

$$m = \min \mathcal{C}.$$

En este caso, la definición de $\mathcal{C}$ implica que

$$\mathcal{L}_m \neq \mathcal{L}_m\{\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \ldots, \tilde{\mathbf{v}}_m\}, \quad \text{y} \quad \mathcal{L}_t = \mathcal{L}\{\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \ldots, \tilde{\mathbf{v}}_t\}, \quad t = 1, 2, \ldots, m-1; \tag{4.7}$$

como ya se ha observado, $1 \notin \mathcal{C}$, de modo que $m > 1$. Para obtener una contradicción, primero observe que

$$\tilde{\mathbf{v}}_m = \mathbf{v}_m - P_{\mathcal{L}_{m-1}}\mathbf{v}_m \in \mathbf{v}_m + \mathcal{L}_{m-1} \subset \mathcal{L}_m;$$

además, debido a (4.7),

$$\mathcal{L}\{\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \ldots, \tilde{\mathbf{v}}_{m-1}\} = \mathcal{L}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{m-1}\} \subset \mathcal{L}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\} = \mathcal{L}_m,$$

y entonces $\tilde{\mathbf{v}}_i \in \mathcal{L}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$, $\quad i = 1, 2, \ldots, m$, de donde se sigue

$$\mathcal{L}\{\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \ldots, \tilde{\mathbf{v}}_m\} \subset \mathcal{L}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}. \tag{4.8}$$

Para obtener la inclusión inversa, seleccione $\mathbf{v} \in \mathcal{L}_m = \mathcal{L}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ En este caso, existen constantes $c_1, c_2, \ldots, c_m$ tales que

$$
\begin{aligned}
\mathbf{v} &= \sum_{i=1}^{m} c_i \mathbf{v}_i \\
&= \sum_{i=1}^{m-1} c_i \mathbf{v}_i + c_m \mathbf{v}_m \\
&= \left[\sum_{i=1}^{m-1} c_i \mathbf{v}_i - c_m P_{\mathcal{L}_{m-1}} \mathbf{v}_m \right] + c_m \left[\mathbf{v}_m - P_{\mathcal{L}_{m-1}} \mathbf{v}_m \right] \\
&= \left[\sum_{i=1}^{m-1} c_i \mathbf{v}_i - c_m P_{\mathcal{L}_{m-1}} \mathbf{v}_m \right] + c_m \tilde{\mathbf{v}}_m,
\end{aligned}
$$

donde la última igualdad se desprende de la definición de $\tilde{\mathbf{v}}_{m-1}$. Observe ahora que el término entre corchetes es un miembro de $\mathcal{L}_{m-1} = \mathcal{L}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{m-1}\} = \mathcal{L}\{\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \ldots, \tilde{\mathbf{v}}_{m-1}\}$ (vea (4.7)), de manera que el argumento precedente implica

$$
\begin{aligned}
\mathbf{v} &= \left[\sum_{i=1}^{m-1} c_i \mathbf{v}_i - c_m P_{\mathcal{L}_{m-1}} \mathbf{v}_m \right] + c_m \tilde{\mathbf{v}}_m \\
&\in \mathcal{L}\{\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \ldots, \tilde{\mathbf{v}}_{m-1}\} + c_m \tilde{\mathbf{v}}_m \\
&\subset \mathcal{L}\{\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \ldots, \tilde{\mathbf{v}}_m\}
\end{aligned}
$$

Puesto que $\mathbf{v} \in \mathcal{L}_m$ es arbitrario en este argumento, se desprende

$$\mathcal{L}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\} \subset \mathcal{L}\{\tilde{\mathbf{v}}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\},$$

inclusión que al combinarse con (4.8) implica

$$\mathcal{L}_m = \mathcal{L}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\} = \mathcal{L}\{\tilde{\mathbf{v}}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}.$$

Esta última igualdad muestra que $m \notin \mathcal{C}$ (vea (4.4)), hecho que se opone a la definición de m como el elemento mínimo de $\mathcal{C}$. En resumen, se ha mostrado que el supuesto $\mathcal{C} \neq \emptyset$ conduce a una contradicción. Por lo tanto (4.5) ocurre, y como ya se ha mencionado, esto concluye la demostración de la parte (i).

(ii) Sea k y s enteros diferentes entre 1 y n, y sin pérdida de generalidad suponga que $k > s$. A partir de la definición de $\tilde{\mathbf{v}}_k$ en (4.2), se obtiene que

$$\begin{aligned}\langle \tilde{\mathbf{v}}_k, \tilde{\mathbf{v}}_s \rangle &= \langle \mathbf{v}_k - P_{\mathcal{L}_{k-1}} \mathbf{v}_k, \tilde{\mathbf{v}}_s \rangle \\ &= \langle \mathbf{v}_k, \tilde{\mathbf{v}}_s \rangle - \langle P_{\mathcal{L}_{k-1}} \mathbf{v}_k, \tilde{\mathbf{v}}_s \rangle\end{aligned}$$

Para concluir, observe que para todo $\mathbf{v} \in \mathcal{L}_{k-1}$

$$\langle \mathbf{v}_k, \tilde{\mathbf{v}} \rangle - \langle P_{\mathcal{L}_{k-1}} \mathbf{v}_k, \mathbf{v} \rangle = 0,$$

igualdad que se obtiene a partir del Lema 3.2; además, debido a que

$$\tilde{\mathbf{v}}_s \in \mathcal{L}\{\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \ldots, \tilde{\mathbf{v}}_{k-1}\} = \mathcal{L}_{k-1}$$

(recuerde que $s < k$ y que la parte (i) ya ha sido demostrada), las dos últimas igualdades desplegadas implican

$$\langle \tilde{\mathbf{v}}_k, \tilde{\mathbf{v}}_s \rangle = \langle \mathbf{v}_k, \tilde{\mathbf{v}}_s \rangle - \langle P_{\mathcal{L}_{k-1}} \mathbf{v}_k, \tilde{\mathbf{v}}_s \rangle = 0,$$

mostrando que dos vectores distintos $\tilde{\mathbf{v}}_k$ y $\tilde{\mathbf{v}}_s$ son ortogonales.

(iii) Observe e que el conjunto $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ es independiente si y sólo si para algún entero k entre 1 y n

$$\mathbf{v}_k \in \mathcal{L}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k-1}\} = \mathcal{L}_{k-1}$$

(Hoffman y Kunze, 1973). Por el Lema 3.2, esta inclusión es equivalente a $\mathbf{v}_k = P_{\mathcal{L}_{k-1}} \mathbf{v}_k$, condición que, via (4.2), coincide con $\tilde{\mathbf{v}}_k = \mathbf{0}$. □

Para calcular el vector $\tilde{\mathbf{v}}_k$, $k > 1$, el procedimiento de Gramm-Schmidt prescribe restar del vector $\mathbf{v}_k$ su proyección sobre el espacio $\mathcal{L}_{k-1}$ generado por los vectores precedentes $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k-1}$. Una ventaja del método, es que en el momento de determinar $\mathbf{v}_k$ se dispone de un conjunto ortogonal de generadores del espacio $\mathcal{L}_{k-1}$, a saber, $\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \ldots, \tilde{\mathbf{v}}_{k-1}$, de manera que $P_{\mathcal{L}_{k-1}} \mathbf{v}_k$ puede encontrarse utilizando el Lema 3.2 como se establece a continuación.

Corolario 4.1. En el procedimiento de Gramm-Schmidt, para cada $k > 1$ se tiene que

$$P_{\mathcal{L}_{k-1}}\mathbf{v}_k = \sum_{i:1\leq i\leq k-1,\ \tilde{\mathbf{v}}_i\neq \mathbf{0}} \frac{\langle \mathbf{v}_k, \tilde{\mathbf{v}}_i\rangle}{\|\tilde{\mathbf{v}}_i\|^2}\tilde{\mathbf{v}}_i.$$

Por lo tanto,

$$\tilde{\mathbf{v}}_k = \mathbf{v}_k - \sum_{i:\ 1\leq i\leq k-1,\ \tilde{\mathbf{v}}_i\neq \mathbf{0}} \frac{\langle \mathbf{v}_k, \tilde{\mathbf{v}}_i\rangle}{\|\tilde{\mathbf{v}}_i\|^2}\tilde{\mathbf{v}}_i \qquad (4.9).$$

En el siguiente ejemplo se ilustra la aplicación de la fórmula (4.9) en un caso específico.

Ejemplo 4.1. En el espacio vectorial $\mathcal{V} = \mathbb{R}^4$, considere los vectores

$$\mathbf{v}_1 = \begin{bmatrix} 5\\ 3\\ 3\\ 5\end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 2\\ 0\\ 0\\ 2\end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 3\\ 5\\ 5\\ 3\end{bmatrix},$$

los cuales fueron introducidos en el Ejemplo 3.1. Los vectores $\tilde{\mathbf{v}}_i$ generados por el procedimiento de Gramm–Schmidt son calculan como sigue:

- El punto de partida es definir $\tilde{\mathbf{v}}_i = \mathbf{v}_1 = [5, 3, 3, 5]'$; note que $\|\tilde{\mathbf{v}}_1\|^2 = 68$.

- Para calcular $\mathbf{v}_2$ se utiliza (4.9) con $k = 2$. En el caso presente se tiene que $\mathbf{v}_1 \neq \mathbf{0}$, de manera que

$$\begin{aligned}\tilde{\mathbf{v}}_2 &= \mathbf{v}_2 - \frac{\langle \mathbf{v}_2, \tilde{\mathbf{v}}_1\rangle}{\|\tilde{\mathbf{v}}_1\|^2}\tilde{\mathbf{v}}_1\\ &= \begin{bmatrix} 2\\ 0\\ 0\\ 2\end{bmatrix} - \frac{20}{68}\begin{bmatrix} 5\\ 3\\ 3\\ 5\end{bmatrix}\\ &= \begin{bmatrix} 2\\ 0\\ 0\\ 2\end{bmatrix} - \frac{5}{17}\begin{bmatrix} 5\\ 3\\ 3\\ 5\end{bmatrix} = \frac{1}{17}\begin{bmatrix} 9\\ -15\\ -15\\ 9\end{bmatrix};\end{aligned}$$

observe que $\|\tilde{\mathbf{v}}_2\|^2 = 612/17^2$.

• Puesto que $\tilde{\mathbf{v}}_1$ y $\tilde{\mathbf{v}}_2$ son diferentes de cero, para $k = 3$ la expresión (4.9) implica

$$\tilde{\mathbf{v}}_3 = \mathbf{v}_3 - \frac{\langle \mathbf{v}_3, \tilde{\mathbf{v}}_1 \rangle}{\|\tilde{\mathbf{v}}_1\|^2}\tilde{\mathbf{v}}_1 - \frac{\langle \mathbf{v}_3, \tilde{\mathbf{v}}_2 \rangle}{\|\tilde{\mathbf{v}}_2\|^2}\tilde{\mathbf{v}}_2. \tag{4.10}$$

Para calcular $\tilde{\mathbf{v}}_3$, note que $\langle \mathbf{v}_3, \tilde{\mathbf{v}}_2 \rangle = -96/17$, y entonces

$$\langle \mathbf{v}_3, \tilde{\mathbf{v}}_2 \rangle / \|\tilde{\mathbf{v}}_2\|^2 = [-96/17]/[612/17^2] = -8/3,$$

de donde se desprende que

$$\frac{\langle \mathbf{v}_3, \tilde{\mathbf{v}}_2 \rangle}{\|\tilde{\mathbf{v}}_2\|^2}\tilde{\mathbf{v}}_2 = -\frac{8}{3}\frac{1}{17}\begin{bmatrix} 9 \\ -15 \\ -15 \\ 9 \end{bmatrix} = -\frac{8}{17}\begin{bmatrix} 3 \\ -5 \\ -5 \\ 3 \end{bmatrix}.$$

Por otro lado, usando que $\langle \mathbf{v}_3, \tilde{\mathbf{v}}_1 \rangle = 60$, y $\|\tilde{\mathbf{v}}_1\|^2 = 68$, se obtiene $\langle \mathbf{v}_3, \tilde{\mathbf{v}}_1 \rangle / \|\tilde{\mathbf{v}}_1\|^2 = 15/17$, y entonces de donde se desprende que

$$\frac{\langle \mathbf{v}_3, \tilde{\mathbf{v}}_1 \rangle}{\|\tilde{\mathbf{v}}_1\|^2}\tilde{\mathbf{v}}_1 = \frac{15}{17}\begin{bmatrix} 5 \\ 3 \\ 3 \\ 5 \end{bmatrix}.$$

Luego,

$$\frac{\langle \mathbf{v}_3, \tilde{\mathbf{v}}_1 \rangle}{\|\tilde{\mathbf{v}}_1\|^2}\tilde{\mathbf{v}}_1 + \frac{\langle \mathbf{v}_3, \tilde{\mathbf{v}}_2 \rangle}{\|\tilde{\mathbf{v}}_2\|^2}\tilde{\mathbf{v}}_2 = \frac{15}{17}\begin{bmatrix} 5 \\ 3 \\ 3 \\ 5 \end{bmatrix} + \left(-\frac{8}{17}\right)\begin{bmatrix} 3 \\ -5 \\ -5 \\ 3 \end{bmatrix} = \frac{1}{17}\begin{bmatrix} 51 \\ 85 \\ 85 \\ 51 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \\ 5 \\ 3 \end{bmatrix} = \mathbf{v}_3,$$

y entonces $\tilde{\mathbf{v}}_3 = \mathbf{0}$; vea (4.10). Observe que la igualdad $\tilde{\mathbf{v}}_3 = \mathbf{0}$ significa que los vectores $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ son linealmente dependientes, por el Teorema 4.1(*iii*). Una vez que se ha construido un sistema ortogonal de vectores que generan $\mathcal{L} = \mathcal{L}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, se puede calcular fácilmente la proyección de cualquier vector $\mathbf{w}$ sobre $\mathcal{L}$. Como en el Ejemplo 3.1 y en su continuación al final de la Sección 3, considere el vector $\mathbf{w} = [0, 3, 3, -2]'$. De acuerdo al Lema 3.2, la proyección de $\mathbf{w}$ sobre $\mathcal{L} = \mathcal{L}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} = \mathcal{L}\{\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2\}$ se construye a través de

$$P_{\mathcal{L}}\mathbf{w} = \frac{\langle \mathbf{w}, \tilde{\mathbf{v}}_1 \rangle}{\|\tilde{\mathbf{v}}_1\|^2}\tilde{\mathbf{v}}_1 + \frac{\langle \mathbf{w}, \tilde{\mathbf{v}}_2 \rangle}{\|\tilde{\mathbf{v}}_2\|^2}\tilde{\mathbf{v}}_2.$$

Notando que

$$\frac{\langle \mathbf{w}, \tilde{\mathbf{v}}_1\rangle}{\|\tilde{\mathbf{v}}_1\|^2} = \frac{8}{68} = \frac{2}{17}$$
$$\frac{\langle \mathbf{w}, \tilde{\mathbf{v}}_2\rangle}{\|\tilde{\mathbf{v}}_2\|^2} = \frac{-[108/17]}{[612/17^2]} = -3$$

se arriba la siguiente expresión:

$$P_{\mathcal{L}}\mathbf{w} = \frac{2}{17}\tilde{\mathbf{v}}_1 + (-3)\tilde{\mathbf{v}}_2 = \frac{2}{17}\begin{bmatrix}5\\3\\3\\5\end{bmatrix} + (-3)\frac{1}{17}\begin{bmatrix}9\\-15\\-15\\9\end{bmatrix} = \begin{bmatrix}-1\\3\\3\\-1\end{bmatrix},$$

la cual coincide con aquella obtenida en la Sección 3; sin embargo, hay una diferencia de fondo entre el argumento actual y aquel empleado anteriormente, a saber, el sistema ortogonal que se utilizó ahora ha sido construido mediante un método general que puede aplicarse en cualquier contexto, mientras que en la Sección 3 dicho conjunto fue dado de forma arbitraria. □

Innovaciones de un Proceso Estocástico

Esta sección trata sobre la relación existente entre el problema de pronóstico para un proceso estocástico y los resultados precedentes, particularmente, sobre el papel del método de Gramm-Schmidt. Como punto de partida, sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad fijo en el cual está definido un proceso estocástico $\{X_t\}$, donde se supone que las variables que lo conforman tienen segundo momento finito y esperanza nula. Sea $\mathcal{V}$ el espacio vectorial generado por las variables aleatorias X_t, de tal forma que $\mathcal{V}$ consiste de todas las combinaciones lineales $\sum_{i=1}^{n} c_i X_{t_i}$, donde las constantes c_i son números reales, los enteros t_i son todos diferentes, y n es un número natural arbitrario. Note que para dos variables aleatorias definidas en $(\Omega, \mathcal{F}, P)$ siempre se tiene que $(X+Y)^2 \leq 2[X^2+Y^2]$, y a partir de este hecho no es difícil verificar que $E[Y^2] < \infty$ siempre que $Y \in \mathcal{V}$; más aún, recordando que cada variable X_t tiene media cero, se desprende que $E[Y] = 0$ si $Y \in \mathcal{V}$. El producto interno en este espacio se define mediante

$$\langle X, Y\rangle = E[XY] = \mathrm{Cov}\,[X,Y], \quad X, Y \in \mathcal{V}, \tag{5.1}$$

de manera que 'la norma' de $Y \in \mathcal{V}$ está dada por

$$\|Y\| = \sqrt{E[Y^2]} = \sqrt{\mathrm{Var}\,[Y]}, \quad Y \in \mathcal{V}, \tag{5.2}$$

es decir, con la definición de producto interno en (5.1), la desviación estándar de Y es precisamente la longitud de Y.

El problema básico de pronóstico para el proceso $\{X_t\}$ consiste en determinar, para cada tiempo n, la mejor aproximación a X_{n+1} en términos de $X_1, X_2, \ldots, X_n$. En la práctica, este objetivo corresponde a la intención de determinar, después de observar el valor asumido por la n–ésima variable aleatoria X_n, un pronóstico para el valor que tomará la variable X_{n+1}, la cual se manifestará en el tiempo $n+1$, es decir, en el futuro inmediato. Como se observó en el Capítulo 2, desde la perspectiva de segundo orden en el análisis de series de tiempo, este objetivo conduce al siguiente problema: *Determinar la combinación* $\sum_{i=1}^n b_i X_i$ *que satisface*

$$\left\| X_{n+1} - \sum_{i=1}^n b_i X_i \right\|^2 = \min_{\beta_1, \beta_2, \ldots, \beta_n} \left\| X_{n+1} - \sum_{i=1}^n \beta_i X_i \right\|^2 .$$

Entre las variables linealmente generadas a partir de $X_1, X_2, \ldots, X_n$, la combinación $\sum_{i=1}^n b_i X_i$ es aquella que se encuentra 'más cerca' de la variable aleatoria X_{n+1}, y es el pronóstico que para dicha variable se emitiriá en el momento n. Si $\mathcal{L} = \mathcal{L}\{X_1, X_2, \ldots, X_n\}$ denota el espacio generado por $X_1, X_2, \ldots, X_n$, el problema de pronóstico puede plantearse como el de encontrar la variable $Y^* \in \mathcal{L}$ que satisface

$$\|X_{n+1} - Y^*\|^2 = \min_{Y \in \mathcal{L}} \|X_{n+1} - Y\|^2 , \tag{5.3}$$

formulación que pone de manifiesto que la construcción de un pronóstico es una versión del problema de aproximación estudiado en las secciones precedentes; su solución única es la proyección de X_{n+1} sobre $\mathcal{L}$, i.e.,

$$Y^* = P_{\mathcal{L}} X_{n+1} = P_{X_1, X_2, \ldots, X_n} X_{n+1};$$

la variable aleatoria Y^* puede pensarse como la parte de X_{n+1} que está (linealmente) determinada por $X_1, X_2, \ldots, X_n$. Puesto que los valores asumidos por cada una de estas variables son todos concidos en el tiempo n, el valor de Y^*—una de sus combinaciones lineales—también se conoce en el momento n, y es el pronóstico del valor que asumirá X_{n+1} en el tiempo $n+1$. Por supuesto, la predicción no tiene porqué resultar exacta, y la diferencia $X_{n+1} - Y^*$, determinada en el momento $n+1$, se interpreta como la parte de X_{n+1} que no se conocía aneriormente, esto es, como 'la novedad', o *innovación*, que X_{n+1} aporta al conocimiento de la serie. En las aplicaciones, es común utilizar una notación ligeramente diferente a la empleada hasta ahora.

Definición 5.1. Considere un proceso estocástico $\{X_t\}_{t=1,2,\ldots}$ con segundos momentos finitos y esperanzas nulas, y sea n un entero positivo.

(i) El pronóstico para X_n en términos de $X_1, X_2, \ldots, X_{n-1}$ es

$$\hat{X}_n = P_{X_1,X_2,\ldots,X_{n-1}} X_n,$$

donde

$$\hat{X}_1 = 0.$$

(ii) La innovación en el tiempo n es $U_n = X_n - \hat{X}_n$.

Si se piensa en X_t como el vector $\mathbf{v}_t$ de las secciones anteriores, entonces $\mathbf{v}_t - P_{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_{t-1}} \mathbf{v}_t = \tilde{\mathbf{v}}_t$ no es otra cosa que $X_t - P_{X_1,X_2,\ldots,X_{t-1}} X_t = X_t - \hat{X}_t = U_t$. Después de esta observación, el Teorema 4.1 puede formularse, en el contexto actual, de las siguiente manera.

Teorema 5.1. Bajo las condiciones sobre $\{X_t\}$ establecidas en la Definición 5.1, las siguientes afirmaciones (i)–(iii) son válidas para el proceso de innovaciones $\{U_t\}$:

(i) $\mathcal{L}\{U_1, U_2, \ldots, U_n\} = \mathcal{L}\{X_1, X_2, \ldots, X_n\}$, donde n es cualquier entero positivo.

(ii) Las innovaciones $U_1, U_2, \ldots, U_n, U_{n+1}, \ldots$ son ortogonales, i.e., para todos los enteros $s, t \geq 1$ con $s \neq t$,

$$\langle U_t, U_s \rangle = \operatorname{Cov}[U_t, U_s] = 0,$$

y

(iii) Las variables aleatorias $\{X_t\}$ son linealmente dependientes si y sólo si U_k es la variable aleatoria cero para algún k.

Las propiedades enunicadas en este teorema son extremadamente importantes. En particular, las partes (i) y (ii) establecen que las innovaciones $U_1, U_2, \ldots, U_n$ forman un sistema ortogonal que genera el mismo espacio que $X_1, X_2, \ldots, X_n$, Luego, al elaborar en el tiempo n el pronóstico $\hat{X}_{n+1}$—la proyección de X_{n+1} sobre $\mathcal{L}\{X_1, X_2, \ldots, X_n\}$—es conveniente usar la representación en términos de $U_1, U_2, \ldots, U_n$:

$$\hat{X}_{n+1} = \theta_{n\,1}U_n + \theta_{n\,2}U_{n-1} + \theta_{n\,3}U_{n-2} + \cdots + \theta_{n\,n}U_1 = \sum_{k=1}^{n} \theta_{n\,k}U_{n+1-k}. \qquad (5.4)$$

En esta expresión se ha utilizado un doble subíndice en los coeficientes de las innovaciones. El primero de ellos indica el tiempo en el cual se está elaborando el pronóstico—en este caso, n—mientras que el segundo, esto es, k, es un indicador de 'la novedad' de la innovación correspondiente. Por ejemplo, U_n es la más reciente de las innovaciones en el tiempo n, por lo que su coeficiente $\theta_{n\,1}$ tiene $k = 1$, U_{n-1} ocupa el segundo lugar entre las innovaciones recientes, asi que su coeficiente es $\theta_{n\,2}$, el cual se asocia a $k = 2$. La ortogonalidad de las innovaciones, conjuntamente con el Lema 3.2, permite escribir la siguiente expresión para $\theta_{n\,k}$, válida bajo el supuesto de que $\|U_{n+1-k}\|^2 = \mathrm{Var}\,[U_{n+1-k}] \neq 0$:

$$\theta_{n\,k} = \frac{\langle X_{n+1}, U_{n+1-k}\rangle}{\|U_{n+1-k}\|^2} = \frac{\mathrm{Cov}\,[X_{n+1}, U_{n+1-k}]}{\mathrm{Var}\,[U_{n+1-k}]}. \qquad (5.5)$$

En última instancia, los valores para estos coeficientes dependen de la estructura de covarianza del proceso $\{X_t\}$. Sin embargo, determinarlos no es una tarea sencilla. Afortunadamente, es posible calcularlos por medio de un procedimiento recursivo, tema que se analiza a continuación. Antes de concluir, note que el cuadrado del error de pronóstico (aproximación) que se comente en el tiempo n es

$\|X_{n+1} - \hat{X}_{n+1}\|$, el cual será denotado por v_n, convención que se establece ahora formalmente para propósitos de referencia futura:

$$v_n = \|X_{n+1} - \hat{X}_{n+1}\|^2 = \|U_{n+1}\|^2 = E[U_{n+1}^2] = \mathrm{Var}\,[U_{n+1}]. \tag{5.6}$$

El Algoritmo de Innovaciones

En esta sección se presenta un método recursivo para calcular los coeficientes $\theta_{n\,k}$ en la expresión (5.4). El resultado se establece precisamente en el Teorema 6.1 líneas más adelante, y es sumamente general, pues es válido para series de tiempo cuya estructura de covarianza no es necesariamente estacionaria, es decir, no se requiere que $\mathrm{Cov}\,[X_{s+h}, X_{t+h}]$ sea independiente de h. A través de esta sección se escribe

$$\mathbb{K}(s,t) = \mathrm{Cov}\,[X_s, X_t]\,. \tag{6.1}$$

Teorema 6.1. Sea $\{X_t\}$ un proceso estocástico cuyas componentes tienen media nula y varianza finita, y son linealmente independientes. En este caso, los coeficientes $\theta_{n\,k}$ en (5.4) y (5.5), y los cuadrados del error de pronóstico v_n en (5.6), se determinan recursivamente a través del siguiente procedimiento:

$$v_0 = \mathbb{K}(1,1), \tag{6.2}$$

mientras que para $n \geq 1$,

$$\begin{aligned} \theta_{n\,n} &= v_0^{-1}\mathbb{K}(n+1,1) \\ \theta_{n\,n-k} &= v_k^{-1}\left(\mathbb{K}(n+1,k+1) - \sum_{j=0}^{k-1}\theta_{k\,k-j}\theta_{n\,n-j}v_j\right), \quad k = 1,2,\ldots n-1 \end{aligned} \tag{6.3}$$

y

$$v_n = \mathbb{K}(n+1,n+1) - \sum_{j=0}^{n-1}\theta_{n,n-j}^2 v_j. \tag{6.4}$$

Demostración. Primero observe que $\hat{X}_1 = 0$ implica que $U_1 = X_1 - \hat{X}_1 = X_1$, de manera que a partir de (5.6) y (6.1) se obtiene

$$v_0 = \mathrm{Var}\,[U_1] = \mathrm{Var}\,[X_1] = \mathrm{Cov}\,[X_1, X_1] = \mathbb{K}(1, 1,),$$

demostrando (6.2). A continuación, seleccione un entero $n \geq 1$, y observe que a partir de la igualdad (5.4) se desprende que

$$\begin{aligned}\langle \hat{X}_{n+1}, U_1\rangle &= \langle \theta_{n\,1}U_n + \theta_{n\,2}U_{n-1} + \theta_{n\,3}U_{n-2} + \cdots + \theta_{n\,n}U_1, U_1\rangle \\ &= \theta_{n\,1}\langle U_n, U_1\rangle + \theta_{n\,2}\langle U_{n-1}, U_1\rangle + \theta_{n\,3}\langle U_{n-2}, U_1\rangle + \cdots + \theta_{n\,n}\langle U_1, U_1\rangle\end{aligned}$$

y usando la ortogonalidad de las innovaciones se obtiene

$$\langle \hat{X}_{n+1}, U_1\rangle = \theta_{n\,n}\langle U_1, U_1\rangle,$$

de manera que

$$\theta_{n\,n} = \frac{\langle \hat{X}_{n+1}, U_1\rangle}{\langle U_1, U_1\rangle} = \frac{\langle \hat{X}_{n+1}, U_1\rangle}{\|U_1\|^2} = \frac{\langle \hat{X}_{n+1}, U_1\rangle}{v_0}.$$

Por otro lado, debido a que

$$\hat{X}_{n+1} = P_{X_1, X_2, \ldots, X_n}X_{n+1}, \quad \text{y} \quad U_1 = X_1 \in \mathcal{L}\{X_1, X_2, \ldots, X_n\},$$

a partir del Teorema 2.1(i) se desprende que

$$\langle \hat{X}_{n+1}, U_1\rangle = \langle X_{n+1}, U_1\rangle = \langle X_{n+1}, X_1\rangle = \mathbb{K}(n+1, 1),$$

hecho que al combinarse con la expresión anterior para $\theta_{n\,n}$ implica

$$\theta_{n\,n} = v_0^{-1}\mathbb{K}(n+1, 1),$$

verificando la primera igualdad en (6.3). A continuación observe que la ecuación (5.4) permite escribir

$$\hat{X}_{n+1} = \sum_{k=1}^{n} \theta_{n\,k}U_{n+1-k} = \sum_{j=0}^{n-1} \theta_{n\,n-j}U_{j+1}, \tag{6.5}$$

donde la segunda igualdad se obtuvo a través del cambio de variable $j = n - k$. Seleccione ahora *un entero positivo* $k < n$. En este caso, tomado el producto interno con U_{k+1}, se desprende que

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

donde se utilizó la ortogonalidad de las innovaciones para obtener la segunda igualdad, de tal manera que

$$\theta_{n\,n-k} = \frac{\langle \hat{X}_{n+1}, U_{k+1} \rangle}{\langle U_{k+1}, U_{k+1} \rangle} = \frac{\langle \hat{X}_{n+1}, U_{k+1} \rangle}{v_k}; \tag{6.6}$$

vea (5.6). Para simplificar esta expresión, recuerde que $\hat{X}_{n+1} = P_{X_1, X_2, \ldots, X_n} X_{n+1}$ es tal que $\langle \hat{X}_{n+1}, Y \rangle = \langle X_{n+1}, Y \rangle$ para todo $Y \in \mathcal{L}\{X_1, X_2, \ldots, X_n\}$, y como $U_{k+1} = X_{k+1} - \hat{X}_{k+1} \in \mathcal{L}\{X_1, X_2, \ldots, X_{k+1}\} \subset \mathcal{L}\{X_1, X_2, \ldots, X_n\}$ (pues $k < n$) se concluye que

$$\begin{aligned}
\langle \hat{X}_{n+1}, U_{k+1} \rangle &= \langle X_{n+1}, U_{k+1} \rangle \\
&= \langle X_{n+1}, X_{k+1} - \hat{X}_{k+1} \rangle \\
&= \langle X_{n+1}, X_{k+1} \rangle - \langle X_{n+1}, \hat{X}_{k+1} \rangle \\
&= \mathbb{K}(n+1, k+1) - \langle X_{n+1}, \sum_{j=0}^{k-1} \theta_{k\,k-j} U_{j+1} \rangle
\end{aligned}$$

donde la última igualdad utiliza la notación en (6.1) y la igualdad (6.5) con $n = k$. Por lo tanto,

$$\langle \hat{X}_{n+1}, U_{k+1} \rangle = \mathbb{K}(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k\,k-j} \langle X_{n+1}, U_{j+1} \rangle. \tag{6.7}$$

Por otro lado, para todo $j \in \{0, 1, \ldots, k-1\}$, la condición $k < n$ implica que

$$U_{j+1} = X_{j+1} - P_{X_1, X_2, \ldots, X_j} X_{j+1} \in \mathcal{L}\{[X_1, X_2, \ldots, X_n\},$$

y entonces, a partir del Teorema 2.1 se obtiene

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

donde se ha usado la expresión (6.5) y la última igualdad se desprende de la ortogonalidad de las innovaciones. Por lo tanto, (5.6) permite escribir $\langle X_{n+1}, U_{j+1}\rangle = \theta_{n,n-j} v_j$, y a partir de (6.7) se obtiene

$$\langle \hat{X}_{n+1}, U_{k+1}\rangle = \mathbb{K}(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k\,k-j} \theta_{n\,n-j} v_j,$$

igualdad que al combinarse con (6.6) produce

$$\theta_{n\,n-k} = v_k^{-1} \left[\mathbb{K}(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k\,k-j} \theta_{n\,n-j} v_j \right],$$

concluyendo la demostración de la segunda igualdad en (6.3), pues k en este argumento es cualquier entero positivo menor a n. Finalmente, observe que la expresión dada en el Teorema 3.1 para el error de aproximación $\|X_{n+1} - P_{X_1,X_2,\ldots,X_n} X_{n+1}\|$ adopta, en el caso actual, la forma

$$\begin{aligned}
v_n &= \|X_{n+1} - P_{X_1,X_2,\ldots,X_n} X_{n+1}\|^2 \\
&= \|X_{n+1}\|^2 - \|P_{X_1,X_2,\ldots,X_n} X_{n+1}\|^2 \\
&= \mathrm{Var}\left[X_{n+1}\right] - \left\| \sum_{j=0}^{n-1} \theta_{n\,n-j} U_{j+1} \right\|^2 \\
&= \mathbb{K}(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n\,n-j}^2 \|U_{j+1}\|^2 \\
&= \mathbb{K}(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n\,n-j}^2 v_j
\end{aligned}$$

donde se ha utilizado la ortogonalidad de las innovaciones, conjuntamente con (6.1), (6.2) y (6.5). Esto demuestra (6.4) y finaliza la demostración del teorema. □

Observación 6.1. (*i*) El Teorema 6.1 permite calcular las cantidades v_n y $\{\theta_{n\,k}\}$ de una forma recursiva; esta característica significa que al momento de evaluar una de dichas cantidades, ésta puede encontrarse por medio de términos previamente determinados. Para verificar esta afirmación, suponga que n es un entero positivo, y que conocen v_s y $\{\theta_{s\,j}\}$ para $s < n$. En este caso, $\theta_{n\,n}$ se calcula por medio de (6.3) directamente a partir de la función de covarianza del proceso, mientras que $\theta_{n\,n-1}$ se calcula por medio de (6.3), para obtener

$$\theta_{n\,n-1} = v_1^{-1}\left(\mathbb{K}(n+1,2) - \theta_{1,1}\theta_{n\,n}v_0\right);$$

note que en el lado derecho de esta expresión aparecen la función de covarianza, v_1 y $\theta_{1\,1}$—valores determinados antes de enfrentar el problema de calcular v_n y $\{\theta_{n\,k}\}$—así como $\theta_{n\,n}$, cantidad que ya ha sido determinada. Esto significa que $\theta_{n\,n}$ se evalúa a partir de vlaores previamente determinados. Similarmente,

$$\theta_{n\,n-2} = v_1^{-1}\left(\mathbb{K}(n+1,3) - \theta_{2,2}\theta_{n\,n}v_0 - \theta_{2,1}\theta_{n\,n-1}v_1\right)$$

se evalúa sólo a través de cantidades conocidas. Una vez que se ha concluido la determinación de todos los números $\theta_{n\,n-k}$, con $k = 0, 1, 2, \ldots, n-1$, entonces v_n se calcula por medio de (6.4). Esta discusión muestra que las cantidades en el Teorema 6.1 deben evaluarse, en cada etapa del proceso, en el siguiente orden (Brockwell y Davis (1991, 1996)):

$$\theta_{n\,n},\ \theta_{n\,n-1},\ \theta_{n\,n-2},\ \theta_{n\,n-3},\ \ldots,\ \theta_{n\,2},\ \theta_{n\,1},\ v_n.$$

(*ii*) Como ya se ha mencionado, el Teorema 6.1 no requiere que la serie $\{X_t\}$ sea estacionaria. Sin embargo, este último caso se presenta en las aplicaciones con frecuencia. En etas circunstancias, si $\gamma(\cdot)$ es la función de autocovarianza del proceso, entonces los únicos 'cambios' en el Teorema 6.1 adaptado al caso estacionario son los siguientes: Reemplazar $\mathbb{K}(1,1)$, y $\mathbb{K}(n+1,n+1)$ por $\gamma(0)$, y cambiar $\mathbb{K}(n+1,k+1)$ y $\mathbb{K}(n+1,n+1)$ por $\gamma(n-k)$ y $\gamma(0)$, respectivamente.

Pronóstico Para una Dato Lejano

En la sección anterior se estudió el problema de formular un pronóstico para la observación X_{n+1} a partir de $X_1, X_2, \ldots, X_n$. El problema que se considera a continuación es similar, pero ahora se trata de emitir un pronóstico para el dato X_{n+h} que se conocerá h unidades de tiempo más adelante. Desde luego, como se estableció en la Sección 2, la mejor aproximación a X_{n+h} en términos de una combinación lineal de $X_1, X_2, \ldots, X_n$ es la proyección de X_{n+h} sobre $\mathcal{L}\{X_1, X_2, \ldots, X_n\}$, y el problema real es proporcionar una fórmula explícita para $P_{X_1,X_2,\ldots,X_n} X_{n+h}$, así como para el error de pronóstico $\|X_{n+h} - P_{X_1,X_2,\ldots,X_n} X_{n+h}\|^2$. El resultado en esta dirección es el siguiente:

Teorema 7.1. Suponga que $\{X_t\}$ un proceso estocástico cuyas componentes tienen media nula y varianza finita, y además son linealmente independientes. En este caso, la proyección de X_{n+h} sobre $\mathcal{L}\{X_1, X_2, \ldots, X_n\}$ está dada por

$$P_{X_1,X_2,\ldots,X_n} X_{n+h} = \sum_{j=h}^{n+h-1} \theta_{n+h-1\,j} U_{n+h-j}, \tag{7.1}$$

dponde h y n son enteros positvos, mientras que el error de pronóstico es

$$\|X_{n+h} - P_{X_1,X_2,\ldots,X_n} X_{n+h}\|^2 = \mathbb{K}(n+h, n+h) - \sum_{j=h}^{n+h-1} \theta^2_{n+h-1\,j} v_{n+h-j-1}. \tag{7.2}$$

Antes de proceder a demostrar este resultado, es conveniente puntualizar que la fórmula (7.1) para $P_{X_1,X_2,\ldots,X_n} X_{n+h}$ involucra a las innovaciones U_{n+h-j}, donde j varía entre h y $n+h-1$, de tal forma que sólo aparecen $U_n, U_{n-1}, \ldots, U_1$, las innovaciones que ya se han determinado en el tiempo n, mientras que los coeficientes $\theta_{n+h-1,k}$ involucrados son los que se obtienen implenentando el algoritmo de innovaciones hasta la etapa $n+h-1$. Por otro lado, suponga, por el momento que la expresión (7.1) es válida. En este caso, el Lema 3.1 implica que

$$\begin{aligned}
\|X_{n+h} - P_{X_1,X_2,\ldots,X_n} X_{n+h}\|^2 &= \|X_{n+h}\|^2 - \|P_{X_1,X_2,\ldots,X_n} X_{n+h}\|^2 \\
&= \operatorname{Var}[X_{n+h}] - \left\| \sum_{j=h}^{n+h-1} \theta_{n+h-1\,j} U_{n+h-j} \right\|^2 \\
&= \operatorname{Var}[X_{n+h}] - \sum_{j=h}^{n+h-1} \theta^2_{n+h-1\,j} \|U_{n+h-j}\|^2
\end{aligned}$$

donde la tercera igualdad se debe a la ortogonalidad de las innovaciones, y se utilizó (5.2) en la segunda igualdad. Observando que

$$\operatorname{Var}[X_{n+h}] = \operatorname{Cov}[X_{n+h}, X - n + h] = \mathbb{K}(n+h, n+h),$$

la formula (5.6) permite escribir

$$\|X_{n+h} - P_{X_1,X_2,\ldots,X_n} X_{n+h}\|^2 = \mathbb{K}(n+h, n+h) - \sum_{j=h}^{n+h-1} \theta^2_{n+h-1\,j} v_{n+h-j-1}$$

expresión que coincide con (7.2).

Demostración del Teorema 7.1. Es claramente suficiente con verificar la igualdad (7.1). El punto de partida para lograr este propósito es la ecuación

$$P_{X_1,X_2,\ldots,X_{n+h-1}} X_{n+h} = \sum_{k=1}^{n+h-1} \theta_{n+h-1\,k} U_{n+h-k}, \tag{7.3}$$

obtenida de (5.4) reemplazando n por $n+h-1$. A Continuación, se utilizará el siguiente hecho sobre proyecciones: Si $\mathcal{L}$ y $\mathcal{M}$ son dos subespacios de $\mathcal{V}$, entonces

$$\mathcal{L} \subset \mathcal{M} \Longrightarrow P_{\mathcal{L}} P_{\mathcal{M}} = P_{\mathcal{L}}. \tag{7.4}$$

La validación de este hecho puede encontrarse, por ejemplo, en Grossman (1983), o Harville (1997). A continuación, observe que a partir de la implicación (7.4) con $\mathcal{L} = \mathcal{L}\{X_1, X_2, \ldots, X_n\}$ y $\mathcal{M} = \mathcal{M}\{X_1, X_2, \ldots, X_{n+h-1}\}$, la igualdad (7.3) implica

$$\begin{aligned} P_{X_1,X_2,\ldots,X_n} X_{n+h} &= P_{X_1,X_2,\ldots,X_n} P_{X_1,X_2,\ldots,X_{n+h-1}} X_{n+h} \\ &= P_{X_1,X_2,\ldots,X_n} \left(\sum_{k=1}^{n+h-1} \theta_{n+h-1\,k} U_{n+h-k} \right) \\ &= \sum_{k=1}^{n+h-1} \theta_{n+h-1\,k} P_{X_1,X_2,\ldots,X_n} (U_{n+h-k}) \end{aligned} \tag{7.5}$$

donde se ha utilizado la linealidad de un operador proyección establecida en el Corolario 2.1(i). Por otro lado, recuerde que las innovaciones $U_1, U_2, \ldots, U_n$ generan el mismo espacio que $X_1, X_2, \ldots, X_n$; debido a que dos innovaciones U_s y U_t con índice distinto son ortogonales, se desprende que $P_{X_1,X_2,\ldots,X_n} U_k = P_{U_1,U_2,\ldots,U_n} U_k = \mathbf{0}$ cuando $k > n$, de manera que

$$P_{X_1,X_2,\ldots,X_n} (U_{n+h-k}) = \mathbf{0}, \quad \text{si } n + h - k > n,$$

esto es, $P_{X_1,X_2,\ldots,X_n} (U_{n+h-k})$ es el vector nulo cuando $h > k$, de manera que

$$\sum_{k=1}^{n+h-1} \theta_{n+h-1\,k} P_{X_1,X_2,\ldots,X_n} (U_{n+h-k}) = \sum_{k=h}^{n+h-1} \theta_{n+h-1\,k} P_{X_1,X_2,\ldots,X_n} (U_{n+h-k}).$$

igualdad que combinada con (7.5) produce

$$P_{X_1,X_2,\ldots,X_n} X_{n+h} = \sum_{k=h}^{n+h-1} \theta_{n+h-1\,k} P_{X_1,X_2,\ldots,X_n} (U_{n+h-k}).$$

Para concluir, note que $n + h - k \leq n$ cuando $k \geq h$, y en este caso $U_{n+h-k} \in \mathcal{L}\{U_1, U_2, \ldots, U_n\} = \mathcal{L}\{X_1, X_2, \ldots, X_n\}$, de manera que el Corolario 2.1(iii) implica que $P_{X_1,X_2,\ldots,X_n} (U_{n+h-k}) = U_{n+h-k}$, por lo que la anterior igualdad desplegada equivale a $P_{X_1,X_2,\ldots,X_n} X_{n+h} = \sum_{k=h}^{n+h-1} \theta_{n+h-1\,k} U_{n+h-k}$, expresión que coincide con (7.1) y finaliza la demostración del Teorema. $\square$

Conclusión

En este capítulo se ha estudiado un problema básico en el análisis de una serie de tiempo, a saber, la emisión de un pronóstico para el valor que asumirá una observación futura en base al conocimiento que proporcionan los datos actuales. El procedimiento analizado fue el algoritmo de innovaciones, el cual es un instrumento muy general para la formulación de pronósticos, pues no requiere establecer el supuesto de que la serie observada sea estacionaria, y posibilita la emisión de un pronóstico para la observación que se tendrá en el futuro inmediato, o h unidades de tiempo más adelante. La principal contribución de este capítulo consistió en determinar de manera precisa la relación existente entre el algoritmo de innovaciones y el procedimiento de ortogonalización de Gramm-Schmidt, y el análisis realizado mostró que, al adoptar una perspectiva de segundo orden, el problema de pronóstico se plantea como un problema gemoétrico que puede abordarse mediante métodos algebraicos.

Capítulo 4

Filtros Lineales

El objetivo de este capítulo es introducir una clase de operdores que actúa transformando una serie de tiempo en otra, de tal forma que en el proceso de cambio se preserva la propiedad de estacionaridad. Los operadores que se analizan se denominan *filtros*, pues al aplicarlos a una serie $\{X_t\}$, su efecto es dejar pasar algunas de sus componentes e impedir el paso de otras. Por otro lado, los resultados de este capítulo son la base para discutir, posteriormente, la existencia de los denominados procesos ARMA (p, q)

Introducción

En este capítulo se aplica una idea fundamental en estadística, a saber, aplicar transformaciones a un objeto aleatorio determinado. El propósito que se persigue es obtener una familia de nuevos objetos aleatorios, dentro de la cual sus características esenciales varían, haciendo posible obtener un mejor ajuste a un conjunto de datos que cuando la atención se restringe a un miembro individual. En el caso que nos ocupa los objetos aleatorios son procesos estocásticos $\{X_t\}$, y sus rasgos esenciales, desde la óptica de segundo orden, son su función de medias y de autocovarianzas. La idea esencial, es que al transformar el proceso de una manera adecuada, se obtengan series transformadas $\{Y_t\}$, y que dentro de esas series, pueda encontrarse una cuyas funciones de medias y autocovarianzas proporcionen una 'buena aproximación' a las características correspondientes obtenidas a partir de datos muestrales.

La organización del capítulo es la siguiente: En la Sección 2 se presenta una discusión general de la utilidad de realizar transformaciones sobre una variable o vector aleatorio, y de las ventajas potenciales que esta operación tiene. Posteriormente, en la Sección 3 se discute la motivación detrás de la noción de filtro, y se introduce su definición formal, mientras que en la Sección 4 se demuestra que la serie transformada $\{Y_t\}$ está bien definida, estableciendo de forma rigurosa sus propiedades básicas. En la Sección 5 se presenta el operador de retardo, un tipo especial de filtro, en términos de cuyas potencias es posible expresar un filtro arbitrario como una serie de potencias. La razón para denominar '*filtro*' a un operador de transfromación introducido en este capítulo se discute en la Sección 6, mientras que en la Sección 7 se analiza la relación entre filtros lineales y la propiedad de estacionaridad, concluyendo en la Sección 8 con una representación integral para la función de autocovarianza de un proceso que se obtiene filtrando un ruido blanco.

La idea Fundamental

La idea de aplicar transformaciones a una variable o vector aleatorio es de gran importancia en la disciplina estadística (Dudewicz y Mishra, 1988, Lehmman, 1991, Mood *et al.,* 1985). En efecto, de esta forma se generan familias de distribuciones cuyas propiedades pueden estudiarse a partir de uno solo de sus miembros. Por ejemplo, considere una variable aleatoria X cuya densidad está dada por

$$f(x) = \frac{1}{2}e^{-|x|}, \quad x \in \mathbb{R}. \tag{2.1}$$

En este caso,

$$E[X] = 0, \quad \operatorname{Var}[X] = 2, \quad \text{y} \quad M_X(t) = \frac{1}{1-t^2}, \quad |t| < 1, \tag{2.2}$$

iguldades que pueden verificarse como sigue: la densidad $f(\cdot)$ es simétrica respecto al origen, de donde se desprende que X tiene esperanza nula, $E[X^2] = \int_{\mathbb{R}} x^2 f(x)\,dx = 2\int_0^\infty x^2 f(x)\,dx = \int_0^\infty x^2 e^{-x}\,dx = 2$, mientras que la expresión para la función generadora de momentos de X se establece después de algunos cálculos directos. Considere ahora la transformación

$$Y = T_{\alpha,\beta}X = \alpha + \beta X, \tag{2.3}$$

donde $\alpha \in \mathbb{R}$ y $\beta \in (0, \infty)$ En este caso, por medio de la fórmula del cambio de variable se obtiene que la densidad de Y es

$$f(x) = \frac{1}{2\beta} e^{-|y-\alpha|/\beta}, \quad y \in \mathbb{R}, \tag{2.4}$$

a partir de la cual pueden determinarse características de interés para la variable aleatoria Y. Sin embargo, es más simple utilizar (2.2) para obtener, via (2.3) y de forma directa, que $E[Y] = \alpha$, $\mathrm{Var}[Y] = 2\beta^2$, y $M_Y(t) = e^{\alpha t}/(1 - \beta^2 t^2)$ para $|t| < 1/\beta$. De hecho, el conjunto de densidades (2.4), conocida como la familia de Laplace, se obtiene mediante la transformación de escala y traslación en (2.3), y este hecho permite reducir el análisis de la familia completa al estudio de cualquiera de sus miembros, por ejemplo, la función en (2.1); una vez que cualquier rasgo de interés se determina para dicha densidad, la contraparte correspondiente a un miembro arbitrario de la clase de densidades (2.4) puede encontrarse fácilmente. Además, al variar α y β, $f(\cdot)$ adopta múltiples 'formas', dentro de las cuales puede encontrarse una que se ajuste mejor a un conjunto de datos observados, que si se restringe la atención a una sóla pareja de parámetros (α, β).

Sin duda alguna, la familia de distribuciones más importante que se genera a través de una clase de transformaciones, es la familia normal (Graybill, 1985). Sea $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)$ un vector cuyas componentes son variables aleatorias independientes con distribución normal estándar. En este caso, no es difícil ver que

$$E[\mathbf{Z}] = \mathbf{0}, \quad \mathrm{Var}[\mathbf{Z}] = \mathbf{I}, \quad M_{\mathbf{Z}}(\mathbf{t}) = e^{-\mathbf{t}'\mathbf{t}/2}, \quad \mathbf{t} \in \mathbb{R}^n. \tag{2.5}$$

Si $\mathbf{C}$ es cualquier matriz no negativa definida, sea $\mathbf{C}^{1/2}$ la única matriz no negativa definida cuyo cuadrado es $\mathbf{C}$; note que $\mathbf{C}^{1/2}$ se obtiene a través del teorema espectral (Hoffman y Kunze (1975), Harville (1997)). Ahora, dado un vector $\mathbf{m} \in \mathbb{R}^n$, defina la transformación

$$\mathbf{X} = T_{\mathbf{m},\ \mathbf{C}}\mathbf{Z} = \mathbf{m} + \mathbf{C}^{1/2}\mathbf{Z}. \tag{2.6}$$

Por definición, $\mathbf{X}$ tiene la distribución normal n–dimensional con 'parámetros' $\mathbf{m}$ y $\mathbf{C}$, la cual se denota mediante $\mathcal{N}(\mathbf{m}, \mathbf{C})$. A partir de (2.5) se desprende que $E[\mathbf{X}] = \mathbf{m}$, $\mathrm{Var}\,[\mathbf{X}] = \mathbf{C}^{1/2}\mathrm{Var}\,[\mathbf{Z}]\mathbf{C}^{1/2} = \mathbf{C}^{1/2}\mathbf{I}\mathbf{C}^{1/2} = \mathbf{C}$, de manera que los parámetros de la distribución de $\mathbf{X}$ son su media y su varianza. Además, usando sólo las propiedades estándar de la función generadora de momentos, (2.5) y (2.6) permiten determinar que $M_{\mathbf{X}}(\mathbf{t}) = e^{\mathbf{m}'\mathbf{t}+\mathbf{t}'\mathbf{C}\mathbf{t}/2}$ para todo $\mathbf{t} \in \mathbb{R}^n$. En general, propiedades de la distribución $\mathcal{N}(\mathbf{m}, \mathbf{C})$ se pueden establecer vía la transformación (2.6). Por ejemplo, si $\mathbf{Y} \sim \mathcal{N}(\mathbf{m}, \mathbf{C})$, entonces $(\mathbf{Y}-\mathbf{m})'\mathbf{C}^{-1}(\mathbf{Y}-\mathbf{m})$ tiene la misma distribución que $\mathbf{Z}'\mathbf{Z}$ la cual es χ^2_n, mientras que $\mathbf{Y}$ asume valores en el hiperplano $\mathbf{m} + \mathcal{H}$, donde $\mathcal{H}$ es el espacio generado por las columnas de $\mathbf{C}$, el cual coincide con el que expanden las columnas de $\mathbf{C}^{1/2}$. En resumen, la idea de transformar un objeto aleatorio a través de un tipo de operador, conduce a obtener una familia de distribuciones cuyo análisis se simplifica al poner atención a las propiedades de la distribución original y, además, al variar los parámetros de la transformación las densidades adoptan múltiples formas dentro de las cuales se puede seleccionar una que se ajuste a un conjunto de datos. Estas ideas son las que está detrás de la noción de filtro lineal, la cual es el objeto de estudio de este capítulo.

Filtros Lineales

El propósito de esta sección es introducir una clase de operadores que, al actuar sobre una serie de tiempo determinada, produce un nuevo proceso cuyas características pueden ser analizadas en términos de la serie original; especialmente, la intención es que cuando el proceso inicial sea estacionario, el proceso transformado también posea dicho rasgo. Para motivar la definición formal que se presentará a continuación, sea $\{X_t\}$ una serie dada, y considere un instante específico n. En ese momento, los valores asumidos por las variables aleatorias X_t, con $t \leq n$, ya han sido observados. La idea es combinar linealmente estos datos para generar una nueva variable aleatoria Y_n. Denote mediante a_k al coeficiente que se utilizará para 'ponderar' el valor de X_{n-k}, esto es, el dato observado k unidades de tiempo antes del momento actual. Esta notación se presenta en el siguiente arreglo, donde el coeficiente de cada variable se escribe debajo de la misma:

$$\begin{array}{ccccccc} \dots & X_{n-r} & \dots & X_{n-3} & X_{n-2} & X_{n-1} & X_n \\ \dots & a_r & \dots & a_3 & a_2 & a_1 & a_0 \end{array} \tag{3.1}$$

Multiplicando cada variable por el coeficiente correspondiente, y sumando los resultados, se obtiene

$$Y_n = \cdots + a_r X_{n-r} + \cdots + a_3 X_{n-3} + a_2 X_{n-2} + a_1 X_{n-1} + a_0 X_n,$$

expresión que en notación más compacta equivale a

$$Y_n = \sum_{k=1}^{\infty} a_k X_{n-k}. \tag{3.2}$$

La pregunta importante ahora, es cómo debe construirse la variable Y_{n+s} en el momento posterior $n+s$, en el cual se han observado los datos X_t con $t \leq n + s$. Recuerde que un objetivo fundamental es que las variables Y_t que se están construyendo formen una serie estacionaria cuando el proceso $\{X_t\}$ tenga dicha propiedad, y note que en este caso la serie $\dots, X_{n+s-3}, X_{n+s-2}, X_{n+s-1}, X_{n+s}$, es 'una réplica probabilística' de la sucesión $\dots, X_{n-3}, X_{n-2}, X_{n-1}, X_n$, de manera que el análogo de la tabla (3.1) en el momento $n+s$ es

$$\begin{array}{ccccccc} \dots & X_{n+s-r} & \dots & X_{n+s-3} & X_{n+s-2} & X_{n+s-1} & X_{n+s} \\ \dots & a_r & \dots & a_3 & a_2 & a_1 & a_0 \end{array} \tag{3.3}$$

y multiplicando cada variable en este arreglo por el coeficiente correspondiente y sumando los resultados se obtiene

$$Y_{n+s} = \cdots + a_r X_{n+s-r} + \cdots + a_3 X_{n+s-3} + a_2 X_{n+s-2} + a_1 X_{n+s-1} + a_0 X_{n+s},$$

esto es,

$$Y_{n+s} = \sum_{k=0}^{\infty} a_k X_{n+s-k}. \tag{3.4}$$

Las ecuaciones (3.3) y (3.4) sugieren que para cada entero t, la variable Y_t debe definirse mediante

$$Y_t = \sum_{k=0}^{\infty} a_k X_{t-k}. \tag{3.5}$$

Esta expresión, motivada por la intención de que $\{Y_t\}$ sea un proceso estacionario siempre que $\{X_t\}$ tenga esta propiedad, es la base para formular la definición de filtro lineal renglones más adelante. Desde un punto de vista formal, la expresión (3.7) que aparece a continuación es más general, en el sentido de que se permite que la ponderación incluya variables que se observarán posteriormente, esto es, en la sucesión de coeficientes $\{a_k\}$ es posible que k asuma un valor negativo, de manera que, por ejemplo, la suma (3.8) para la variable Y_t 'construida' en el momento t incluye un término de la forma $a_{-3}X_{t-(-3)} = a_{-3}X_{t+3}$, y la variable X_{t+3} será directamente observada sólo tres unidades más adelante, esto es, en el futuro. Por otro lado, es importante señalar que es necesario imponer condiciones sobre los coeficientes $\{a_k\}$ que se utilizarán para ponderar los datos X_t, de tal forma que la suma resultante tenga sentido, es decir, esté bien definida; dicha restricción se establece en la expresión (3.6) a continuación.

Definición 3.1. Sea $\{a_k \mid k = 0, \pm1, \pm2, \pm3, \ldots\}$ una sucesión de números reales, la cual es absolutamente sumable, i.e.,

$$\sum_{k=-\infty}^{\infty} |a_k| < \infty, \tag{3.6}$$

y suponga que el proceso $\{X_t\}$ es acotado en la media de orden 2, esto es, existe una constante finita M tal que

$$E[X_t^2] \leq M \quad \text{para todo } t. \tag{3.7}$$

Bajo estas condiciones, defina la sucesión $\{Y_t\}$ mediante

$$Y_t = \sum_{k=-\infty}^{\infty} a_k X_{t-k}. \tag{3.8}$$

En este caso, *el filtro lineal* asociado a la sucesión $\{a_k\}$ es el operador

$$\{X_t\} \mapsto \{Y_t\},$$

el cual transforma la serie $\{X_t\}$ en el proceso $\{Y_t\}$ determinado por (3.8).

El siguiente paso en el desarrollo de la teoría de filtros lineales, es verificar que la definición de las componentes del proceso transformado se encuentran bien definidas a través de (3.8), y analizar el sentido en que converge la serie que define a las variables Y_t.

El Teorema Básico

El propósito de esta sección es establecer el resultado fundamental sobre filtro lineales, a saber, que las variables Y_t en (3.8) heredan de la serie original $\{X_t\}$ la propiedad de tener segundo momento acotado, y analizar la convergencia de la serie que define a las componentes del proceso $\{Y_t\}$. En esta última dirección, se mostrará que dicha convergencia es puntual (con probabilidad uno), en la media de orden uno, y en la media cuadrática, resultados que son fundamentales para el desarrollo subsecuente (Anderson, 1976, Box y Jenkins, 1970, Cavazos–Cadena, 1994).

Teorema 4.1. Suponga que la sucesión $\{a_k\}$ y la serie $\{X_t\}$ satisfacen las condiciones (3.6) y (3.7), respectivamente. En estas circunstancias,

(i) Para cada entero t,

$$E\left[\sum_{k=-\infty}^{\infty} |a_k|\,|X_{t-k}|\right] \leq \sqrt{M} \sum_{k=-\infty}^{\infty} |a_k| < \infty,$$

donde M es la constante en (3.7). Consecuentemente,

(ii) Con probabilidad uno, la serie $\sum_{k=-\infty}^{\infty} a_k X_{t-k}$ converge absolutamente, i.e., $\sum_{k=-\infty}^{\infty} |a_k X_{t-k}| < \infty$, y entonces la variable aleatoria Y_t en (3.8) está bien definida.

(iii) $E[|Y_t|] \leq \sqrt{M} \sum_{k=-\infty}^{\infty} |a_k|$.

(iv) La serie en (3.8) converge a Y_t en la media de orden uno, es decir, para todas las sucesiones $\{D(n)\}$ y $\{C(n)\}$ de enteros positivos que crecen hacia ∞,

$$\lim_{n\to\infty} E\left[\left|Y_t - \sum_{k=-C(n)}^{D(n)} a_k X_{t-k}\right|\right] = 0.$$

Similarmente,

(v) $E[Y_t^2] \leq M \left(\sum_{k=-\infty}^{\infty} |a_k|\right)^2$, y

(vi) La serie en (3.8) converge a Y_t con respecto a la media cuadrática, esto es,

$$\lim_{n\to\infty} E\left[\left(Y_t - \sum_{k=-C(n)}^{D(n)} a_k X_{t-k}\right)^2\right] = 0,$$

siempre que las sucesiones $\{D(n)\}$ y $\{C(n)\}$ de enteros positivos crezcan hacia ∞.

Demostración. (i) El argumento inicia observando que la desigualdad de Cauchy-Schwartz implica que

$$E[|X_t|] \leq \left(E[X_t^2]\right)^{1/2} \leq \sqrt{M}, \tag{4.1}$$

donde se utilizó la condición (3.7) para obtener la segunda desigualdad. Por otro lado, el teorema de convergencia monótona permite concluir que

$$\begin{aligned} E\left[\sum_{k=-\infty}^{\infty} |a_k|\,|X_{t-k}|\right] &= E\left[\lim_{n\to\infty} \sum_{k=-n}^{n} |a_k|\,|X_{t-k}|\right] \\ &= \lim_{n\to\infty} E\left[\sum_{k=-n}^{n} |a_k|\,|X_{t-k}|\right] \\ &= \lim_{n\to\infty} \sum_{k=-n}^{n} |a_k|\, E\left[|X_{t-k}|\right] \\ &\leq \lim_{n\to\infty} \sum_{k=-n}^{n} |a_k|\sqrt{M} \\ &= \sqrt{M} \sum_{k=-\infty}^{\infty} |a_k| < \infty, \end{aligned}$$

donde la desigualdad se debe a (4.1).

(ii) Puesto que $E\left[\sum_{k=-\infty}^{\infty} |a_k|\, |X_{t-k}|\right] < \infty$, por la parte (i), se desprende que

$$P\left[\sum_{k=-\infty}^{\infty} |a_k|\, |X_{t-k}| < \infty\right] = 1.$$

Luego, si Ω' es el evento entre corchetes en esta igualdad, se tiene que $P[\Omega'] = 1$, y $\sum_{k=-\infty}^{\infty} |a_k|\, |X_{t-k}(\omega)| < \infty$ siempre que $\omega \in \Omega'$. Este hecho significa que la serie $\sum_{k=-\infty}^{\infty} a_k\, X_{t-k}(\omega)$ converge absolutamente hacia un límite finito, esto es, sin importar el orden en que se agreguen los términos de la serie, y por lo tanto $Y_t(\omega)$ está bien definida mediante la expresión

$$Y_t(\omega) = \sum_{k=-\infty}^{\infty} a_k\, X_{t-k}(\omega), \quad \omega \in \Omega'. \tag{4.2}$$

Debido que Ω' tiene probabilidad uno, Y_t puede definirse de manera arbitraria cuando $\omega \notin \Omega'$ sin alterar sus propiedades esenciales.

(iii) A partir de (4.2) se obtiene que

$$|Y_t(\omega)| \leq \sum_{k=-\infty}^{\infty} |a_k|\, |X_{t-k}(\omega)|, \quad \omega \in \Omega';$$

puesto que $P[\Omega'] = 1$, esta desigualdad implica que

$$\begin{aligned} E[|Y_t|] &\leq E\left[\sum_{k=-\infty}^{\infty} |a_k|\, |X_{t-k}(\omega)|\right] \\ &= \sum_{k=-\infty}^{\infty} |a_k|\, E\left[|X_{t-k}(\omega)|\right] \\ &\leq \sqrt{M} \sum_{k=-\infty}^{\infty} |a_k| \end{aligned}$$

donde la igualdad se debe al teorema de convergencia monótona, y se utilizó (4.1) para obtener la segunda desigualdad.

(iv) La expresión (4.2) para $Y_t(\omega)$ implica que

$$Y_t(\omega) - \sum_{k=-C(n)}^{D(n)} a_k X_{t-k}(\omega) = \sum_{k=-\infty}^{-C(n)-1} a_k X_{t-k}(\omega) + \sum_{k=D(n)+1}^{\infty} a_k X_{t-k}(\omega), \tag{4.3}$$

y entonces

$$\left| Y_t(\omega) - \sum_{k=-C(n)}^{D(n)} a_k X_{t-k}(\omega) \right| \leq \sum_{k=-\infty}^{-C(n)-1} |a_k|\,|X_{t-k}(\omega)| + \sum_{k=D(n)+1}^{\infty} |a_k|\,|X_{t-k}(\omega)|.$$

El hecho crucial es que esta relación es válida para todo ω en el evento Ω', el cual tiene probabilidad uno. Por lo tanto, via el teorema de convergencia monótona y (4.1), se desprende que

$$\begin{aligned} E\left[\left| Y_t - \sum_{k=-C(n)}^{D(n)} a_k X_{t-k} \right| \right] &\leq \sum_{k=-\infty}^{-C(n)-1} |a_k|\, E[|X_{t-k}|] + \sum_{k=D(n)+1}^{\infty} |a_k|\, E[|X_{t-k}|] \\ &\leq \sqrt{M} \left[\sum_{k=-\infty}^{-C(n)-1} |a_k| + \sum_{k=D(n)+1}^{\infty} |a_k| \right] \end{aligned} \tag{4.4}$$

Puesto que la condición (3.6) es válida, y tanto $C(n)$ como $D(n)$ crecen sin límite conforme n tiende a ∞, se desprende que

$$\lim_{n\to\infty} \left[\sum_{k=-\infty}^{-C(n)-1} |a_k| + \sum_{k=D(n)+1}^{\infty} |a_k| \right] = 0,$$

y combinando esta convergencia con la desigualdad (4.4) se concluye que

$$\lim_{n\to\infty} E\left[\left| Y_t - \sum_{k=-C(n)}^{D(n)} a_k X_{t-k} \right| \right] \leq \sqrt{M} \lim_{n\to\infty} \left[\sum_{k=-\infty}^{-C(n)-1} |a_k| + \sum_{k=D(n)+1}^{\infty} |a_k| \right] = 0.$$

(v) El punto de partida es la ecuación (4.2); utilizándola, se desprende que para cada $\omega \in \Omega'$

$$\begin{aligned} Y_t(\omega)^2 &= \left| \sum_{k=-\infty}^{\infty} a_k\, X_{t-k}(\omega) \right| \times \left| \sum_{j=-\infty}^{\infty} a_j\, X_{t-j}(\omega) \right| \\ &\leq \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} |a_j||a_k||X_{t-j}(\omega)||X_{t-k}(\omega)| \end{aligned}$$

y entonces, recordando que $P[\Omega'] = 1$, se desprende que

$$\begin{aligned} E[Y_t^2] &\leq E\left[\sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} |a_j||a_k||X_{t-j}||X_{t-k}| \right] \\ &= \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} |a_j||a_k| E\left[|X_{t-j}||X_{t-k}|\right] \end{aligned} \tag{4.5}$$

Por otro lado, de la desigualdad de Cauchy–Schwartz se obtiene $E\left[|X_{t-j}||X_{t-k}|\right]$ $\leq \sqrt{E[X_{t-j}^2]}\sqrt{E[X_{t-k}^2]} \leq \sqrt{M}\sqrt{M} = M$, y entonces

$$E[Y_t^2] \leq \sum_{k=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} |a_j||a_k|M$$
$$= M\sum_{k=-\infty}^{\infty}|a_k|\sum_{j=-\infty}^{\infty}|a_j| = M\left(\sum_{k=-\infty}^{\infty}|a_k|\right)^2 < \infty$$

(vi) Observe que para dos variables aleatorias L y M siempre se satisface $(L+M)^2 \leq 2[L^2+M^2]$. Luego, aplicando la ecuación (4.3) se obtiene que para todo $\omega \in \Omega'$,

$$\left|Y_t(\omega) - \sum_{k=-C(n)}^{D(n)} a_k X_{t-k}(\omega)\right|$$

$$\leq 2\left(\sum_{k=-\infty}^{-C(n)-1} a_k X_{t-k}(\omega)\right)^2 + 2\left(\sum_{k=D(n)+1}^{\infty} a_k X_{t-k}(\omega)\right)^2,$$

y como $P[\Omega'] = 1$, se concluye que

$$E\left[\left|Y_t - \textstyle\sum_{k=-C(n)}^{D(n)} a_k X_{t-k}\right|^2\right]$$

$$\leq 2E\left[\left(\sum_{k=-\infty}^{-C(n)-1} a_k X_{t-k}\right)^2\right] + 2E\left[\left(\sum_{k=D(n)+1}^{\infty} a_k X_{t-k}\right)^2\right]$$
$$\leq 2M\left(\sum_{k=-\infty}^{-C(n)-1} |a_k|\right)^2 + 2M\left(\sum_{k=D(n)+1}^{\infty} |a_k|\right)^2$$

donde la segunda desigualdad se establece a través de procedimientos análogos a los utilizados para demostrar la parte (v). Puesto que la serie $\sum_k a_k$ es absolutamente sumable, se tiene que, conforme n crece sin límite,

$$\left(\sum_{k=-\infty}^{-C(n)-1} |a_k|\right)^2 + \left(\sum_{k=D(n)+1}^{\infty} |a_k|\right)^2 \to 0,$$

y combinando esta convergencia con la anterior desigualdad desplegada, se obtiene

$$\lim_{n\to\infty} E\left[\left|Y_t - \sum_{k=-C(n)}^{D(n)} a_k X_{t-k}\right|^2\right] = 0,$$

concluyendo la demostración del teorma. □

El Operador de Retardo y Series de Potencias

En la expresión (3.8) para la variable Y_t, el coeficiente a_k multiplica a X_{t-k}; pensando en el instante t como el momento actual, X_{t-k} es el dato observado k unidades de tiempo previas al presente. En esta sección se introduce un tipo especial de filtro que permite representar, de una manera cómoda, la operación de retrazar el momento en que las variables $\{X_t\}$ son observadas.

Definición 5.1. El *operador de retardo*, denotado mediante B, transforma una serie $\{X_t\}$ en otro proceso $\{R_t\}$ definido mediante

$$R_t = BX_t = X_{t-1}, \tag{5.1}$$

esto es, $\{R_t\}$ se compone de las mismas variables X_t, *pero observadas con una unidad de retrazo.*

Si se aplica B a la serie compuesta por las variables R_t, entonces se obtiene $BR_t = R_{t-1}$, expresión que equivale a $B[BX_t] = X_{(t-1)-1}$, o en notación más compacta, $B^2X_t = X_{t-2}$. Más generalmente, para cada $k = 1, 2, \ldots$, la actuación de B^k sobre las variables X_t está dada por

$$B^kX_t = X_{t-k}, \tag{5.2}$$

de tal manera que si $\{W_t\}$ se obtiene aplicando el operador B^k a la serie $\{X_t\}$, entonces $W_t = B^kX_t = X_{t-k}$, igualdad que significa que el proceso $\{W_t\}$ consiste de las mismas variables X_t, pero observadas con k unidades de retraso. Hasta ahora se ha pensado en el operador B como un medio para indicar 'un retardo' en el momento de observar los miembros de una serie, y k en la expresión (5.2) es un número positivo. Sin embargo, el lado derecho en dicha ecuación tiene sentido aún cuando k sea negativo. Por ejemplo, si $k = -3$, entonces entonces $B^{-3}X_t = X_{t-(-3)} = X_{t+3}$, lo cual significa que la actuación de B^{-3} sobre los miembros de la serie, consiste en *adelantar* el tiempo en que las variables X_t son observadas. También es posible poner $k = 0$, y obtener que $B^0 X_t = X_t$, y entonces la actuación de B^0 sobre la serie $\{X_t\}$ deja inalterado el proceso. En resumen, el operador B actúa sobre series de acuerdo a (5.1), y la acción de sus potencias está determinada mediante (5.2), donde k es cualquier entero: si la serie $\{Y_t\}$ se obtiene a través de $Y_t = B^k X_t = X_{t-k}$, entonces (i) para $k > 0$, las variables X_t se observan con un retrazo de k unidades, (ii) si $k < 0$, la aplicación de B^k significa que las componentes del proceso $\{X_t\}$ se observan con un adelanto de k unidades, mientras que (iii) el operador B^0 es la identidad, en el sentido de que su aplicación deja inalterado el momento en que se observan las variables X_t.

La acción del filtro asociado a una sucesión sumable de números reales $\{a_k\}$ pude representarse de una forma conveniente utilizando la notación (5.2). En efecto, de acuerdo a (3.8) y (5.2),

$$\sum_k a_k B^k X_t = \sum_k a_k X_{t-k} = Y_t. \tag{5.3}$$

Esta igualdad permite denotar al filtro correspondiente a $\{a_k\}$ mediante

$$\sum_k a_k B^k, \tag{5.4}$$

la cual es una suma formal de potencias del operador de retardo cuya actuación sobre una serie $\{X_t\}$ está dada por la primera igualdad en (5.3). Además de la serie (5.4), en el desarrollo subsecuente se utilizará otra serie formal asociada a la sucesión $\{a_k\}$; ésta última es una serie de variable compleja, la cual es

$$\alpha(z) = \sum_k a_k z^k; \tag{5.5}$$

si en esta ecuación se reemplaza z por el operador B se tiene que

$$\alpha(B) = \sum_k a_k B^k \tag{5.6}$$

es el filtro lineal asociado a $\{a_k\}$.

Observación 5.1. En este momento es oportuno discutir la convergencia de la serie formal (5.5), teniendo en mente que la sucesión $\{a_k\}$ es absolutamente sumable; vea (3.6). Los resultados que se presentarán a continuación desempeñan un papel relevante en el estudio de la composición e invertibilidad de filtros lineales que se emprenderá en el siguiente capítulo.

(i) Note que la serie $\alpha(z)$ converge (absolutamente) para todo número complejo z con $|z| = 1$. En efecto,

$$\sum_k |a_k z^k| = \sum_k |a_k|\,|z|^k = \sum_k |a_k| < \infty.$$

(ii) Sin embargo, es interesante, e importante, observar que es posible que la serie (5.5) sólo converja cuando $|z| = 1$. Para ver que esta posibilidad es factible, considere la sucesión $\{a_k\}$ dada por

$$a_k = \frac{1}{(1+|k|)^2} \quad \text{para todo entero } k.$$

En este caso,

$$\begin{aligned}\sum_k a_k z^k &= a_0 + \sum_{k=1}^{\infty} a_k z^k + \sum_{k=-\infty}^{-1} a_k z^k \\ &= a_0 + \sum_{k=1}^{\infty} \frac{1}{(1+|k|)^2} z^k + \sum_{k=-\infty}^{-1} \frac{1}{(1+|k|)^2} z^k;\end{aligned}$$

en el extremo derecho de la última igualdad, la primera serie converge para $|z| \leq 1$, mientras que la segunda, la cual involucra potencias negativas de z, converge sólo cuando $|z| \geq 1$. Puesto que $\sum_k a_k z^k$ es convergente sólo cuando *ambas* series en la extrema derecha tienen esta propiedad, se tiene que 'la región' de convergencia es $\{z\colon |z| \leq 1\} \cap \{z \,:\, |z| \geq 1\} = \{z \,|\, |z| = 1\}$, es decir, $\sum_k a_k z^k$ converge sólo cuando el número complejo z tiene módulo (norma) igual a uno.

(iii) Suponga ahora que la serie que define a $\alpha(z)$ converge en un conjunto abierto, el cual forsozamente contiene al disco unitario (ve la parte (i) líneas arriba), i.e.,

$$\alpha(z) = \sum_k a_k z^k, \quad r < |z| < R, \text{ donde } r < 1 < R. \tag{5.7}$$

En este caso, la función $\alpha(z)$ es analítica en el anillo $\mathcal{A} = \{z\colon r < |z| < R\}$, y la expansión en serie para $\alpha(z)$ en (5.7) es la representación (o desarrollo) de Laurent para $\alpha(z)$ en la región $\mathcal{A}$ (Alfhors (1970), Nevannlina y Patero (1975), Churchill (1976)). Los coeficientes a_k en (5.7) están unívocamente asociados a $\alpha(z)$, y están determinados por la siguiente expresión integral:

$$a_k = \frac{1}{2\pi i} \int_{\mathcal{C}} \xi^{-k+1} \alpha(\xi)\, d\xi, \tag{5.8}$$

donde $\mathcal{C}$ denota al círculo unitario, y la integral es en el sentido positivo, esto es, contrario al de las manecillas del reloj.

(iv) Suponga que $\alpha(z)$ es como en (5.7), y que además,

$$\alpha(z) \neq 0, \quad \text{si } r < |z| < R. \tag{5.9}$$

En estas circunstancias, el recíproco de $\alpha(z)$ está bien definido: defina

$$\beta(z) = \frac{1}{\alpha(z)}, \quad r < |z| < R, \tag{5.10}$$

y note que $\beta(z)$ es analítica en el anillo $\mathcal{A}$, y que se expande en serie de Laurent, esto es,

$$\beta(z) = \sum_k b_k z^k, \quad r < |z| < R, \tag{5.11}$$

Los coeficientes en esta serie son únicos y se determinan a través de una expresión integral análoga a (5.8), la cual es

$$b_k = \frac{1}{2\pi i} \int_{\mathcal{C}} \xi^{-k+1} \beta(\xi)\, d\xi = \frac{1}{2\pi i} \int_{\mathcal{C}} \frac{\xi^{-k+1}}{\alpha(\xi)}\, d\xi. \tag{5.12}$$

Encontrar el desarrollo en serie de Laurent correspondiente al recíproco de una función $\alpha(z)$, donde ésta función no se anula en un anillo que contiene al círculo unitario, es de gran importancia en el estudio de filtros lineales, pues este problema está íntimamente relacionado con el proceso de invertir el efecto de un filtro. Aunque los coeficientes de la expansión de $\beta(z) = 1/\alpha(z)$ se determinan de manera única por medio de (5.12), en algunos caso de interés la función $\alpha(z)$ es un polinomio cuyas raíces se conocen, y la fórmula (5.12) es demasiado compleja comparada con el procedimento alternativo delineado en el siguiente lema y su corolario, resultados en los que se presenta una fórmula explícita para obtener el desarrollo en serie de Laurent del recíproco de un polinomio, en donde la expansión es válida en un anillo que contienen al círculo unitario.

Lema 5.1. Suponga que $0 \neq |a| \neq 1$. En este caso,

$$\text{Si } |a| < 1, \text{ entonces } \quad \frac{1}{1-az} = \sum_{k=0}^{\infty} a^k z^k, \quad |z| < \frac{1}{|a|},$$

mientras que

$$\text{Si } |a| > 1, \quad \frac{1}{1-az} = -\sum_{k=1}^{\infty} \frac{1}{a^k z^k} = -\sum_{k=-\infty}^{-1} a^k z^k, \quad |z| > \frac{1}{|a|}.$$

Demostración. La igualdad clave es

$$\frac{1}{1-r} = \sum_{k=0}^{\infty} r^k, \quad \text{si } |r| < 1. \tag{5.13}$$

Suponga ahora que $|a| < 1$. En este caso,

$$|z| < \frac{1}{|a|} \Longrightarrow |az| < 1 \Longrightarrow \frac{1}{1-az} = \sum_{k=0}^{\infty} (az)^k = \sum_{k=0}^{\infty} a^k z^k,$$

donde la segunda implicación utiliza (5.13) con $r = az$. Por otro lado, cuando $|a| > 1$, se tiene que

$$|z| > \frac{1}{|a|} \Longrightarrow \frac{1}{|az|} < 1 \Longrightarrow \frac{1}{1-az} = -\left(\frac{1}{az}\right) \frac{1}{1-\dfrac{1}{az}} = -\left(\frac{1}{az}\right) \sum_{k=0}^{\infty} \left(\frac{1}{az}\right)^k,$$

donde se utilizó (5.13) con $r = 1/az$. Por lo tanto,

$$\frac{1}{1-az} = -\left(\frac{1}{az}\right)\sum_{k=0}^{\infty}\left(\frac{1}{az}\right)^k = -\sum_{k=1}^{\infty}\frac{1}{a^k z^k} = -\sum_{k=-\infty}^{-1} a^k z^k.$$

□

Ejemplo 5.1. (i) Si $\alpha(z) = 1 - 2z$ entonces

$$\frac{1}{\alpha(z)} = \frac{1}{1-2z} = -\sum_{k=-\infty}^{-1} 2^k z^k, \quad |z| > \frac{1}{2}.$$

(ii) Si $\alpha(z) = 1 - z/2$, se tiene que

$$\frac{1}{\alpha(z)} = \frac{1}{1-z/2} = \sum_{k=0}^{\infty}\frac{1}{2^k} z^k, \quad |z| < 2.$$

(iii) En los dos casos anteriores, la expansión se obtiene aplicando directamente el Lema 5.1, el cual se refiere al desarrollo en serie de Laurent del recíproco de un polinomio de grado uno. Sin embargo, frecuentemente se está interesado en polinomios de grado mayor. En este caso, el Lema 5.1 aún puede ser útil. Considere, por ejemplo, el polinomio $\alpha(z) = 1 - 2.5z + z^2$, el cual se factoriza como $\alpha(z) = (1-2z)(1-z/2)$, de manera que

$$\frac{1}{\alpha(z)} = \frac{1}{(1-2z)(1-z/2)} = \frac{4/3}{1-2z} + \frac{-1/3}{1-z/2}.$$

y aplicando las expansiones obtenidas previamente, se desprende que

$$\frac{1}{\alpha(z)} = -\frac{4}{3}\sum_{k=-\infty}^{-1} 2^k z^k - \frac{1}{3}\sum_{k=0}^{\infty}\frac{1}{2^k} z^k, \quad \frac{1}{2} < |z| < 2.$$

□

Este ejemplo muestra que determinar el desarrollo en serie del recíproco de un polinomio $\alpha(z)$, puede ser una tarea simple si se conocen sus raíces; la idea es obtener la expansión de Laurent de $1/\alpha(z)$ desarrollando este recíproco en fracciones parciales. Sin embargo, una dificultad surge en este enfoque, a saber, $\alpha(z)$ puede tener raíces múltiples. El siguiente corolario aborda este problema.

Corolario 5.1. Suponga que $|a| \neq 1$, y para cada entero k defina

$$(k)_0 = 1, \quad \text{y} \quad (k)_m = k(k-1)\cdots(k-m+1), \quad m = 1, 2, 3, \ldots,$$

y

$$\binom{k}{m} = \frac{(k)_m}{m!}, \quad m = 0, 1, 2, \ldots$$

En este caso, para cada entero positivo $m \geq 0$, las siguientes afirmaciones (i)–(iii) son válidas:

$$(i) \qquad \frac{d^m}{dr^m}\left(\frac{1}{1-r}\right) = \frac{m!}{(1-r)^{m+1}} = \sum_{k=0}^{\infty}(k)_m r^{k-m}.$$

$$(ii) \qquad \text{Si } |a| < 1 \quad = \frac{1}{(1-az)^{m+1}} = \sum_{k=0}^{\infty}\binom{k}{m}a^{k-m}z^{k-m}, \quad |z| < \frac{1}{|a|}.$$

$$(iii) \qquad \text{Si } |a| > 1 \quad \frac{1}{(1-az)^{m+1}} = -\sum_{k=-\infty}^{-1}\binom{k}{m}a^{k-m}z^{k-m}, \quad |z| > \frac{1}{|a|}.$$

(iv) Si $\alpha(z)$ es un polinomio que no se anula en el círculo unitario, el desarrollo en serie de Laurent del recíproco $\beta(z) = 1/\alpha(z)$ se obtiene mediante el siguiente procedimiento:

(a) Factorice $\alpha(z)$ como

$$\alpha(z) = A\prod_{i=1}^{\ell}(1 - a_i z)^{m_i+1},$$

donde $1/a_1, \ldots 1/a_\ell$ son las raíces distintas de $\alpha(z)$, y la multiplicidad de $1/a_i$ es $m_i + 1$.

(b) Desarrolle el recíproco de $\alpha(z)$ en fracciones parciales como

$$\beta(z) = \frac{1}{\alpha(z)} = \frac{1}{\prod_{i=1}^{\ell}(1 - a_i z)^{m_i+1}} = \sum_{i=1}^{\ell}\sum_{j=0}^{m_i}\frac{A_{j\,i}}{(1 - a_i z)^{j+1}}.$$

(c) La expansión de Laurent se obtiene desarrollando $\dfrac{1}{(1-a_i z)^{j+1}}$ como se indica en las partes (ii) y (iii), multiplicando las expansiones por $A_{j\,i}$ y sumando los resultados.

Terminología

EL concepto principal de este capítulo, involucra la transformación de una serie en otra utilizando un operador denotado por $\sum_k a_k B^k$, donde B es el operador de retardo, y los coeficientes a_k confortman una sucesión sumable. Dicho operador ha sido llamado *filtro*, terminología que no ha sido 'justificada' o explicada. La intención de esta sección es proporcionar una idea intuitiva de la razón detrás del nombre 'filtro' para el operador $\sum_k a_k B^k$. En realidad, si los coeficientes a_k se seleccionan de forma adecuada, entonces dicho operador *efectivamente filtra* los datos, eliminado algunas de sus componentes, y dejando sólamente aquellas que son de interés para el observador.

La Descomposición Clásica. Suponga que una serie de datos $x_1, x_2, \ldots, x_n, \ldots$ se registran sucesivamente, donde se supone que las observaciones son una realización de un proceso $\{X_t\}$ con la siguiente estructura (Box y Jenkins (1970):

$$X_t = s_t + \kappa_t + W_t, \tag{6.1}$$

donde las variables aleatorias W_t tienen media nula y conforman un proceso estacionario, la sucesión s_t, denominada la componente estacional, satisface

$$s_t = s_{t-d}, \quad \text{para algún entero } d > 0, \quad \text{y} \quad \sum_{j=0}^{d-1} s_{t+j} = 0, \tag{6.2}$$

de manera que s_t tiene período d, mientras que κ_t es una función lineal (un polinomio de grado uno en t) el cual se conoce como la función de tendencia:

$$\kappa_t = C_0 + C_1 t. \tag{6.3}$$

La representación (6.1) se denomina la descomposición clásica del proceso $\{X_t\}$; note que en este caso $E[X_t] = s_t + \kappa_t$ no es constante, y entonces la serie no es estacionaria. El modelo (6.1) es 'natural' en muchas situaciones práctica. Por ejemplo, suponga que un incremento unitario en el índice de tiempo representa un mes, y que X_t es la demanda de nieve que recibe una fábrica de helados en el mes t. En este caso, la componente estacional s_t toma en cuenta que la demanda de nieve es mayor en el verano que durante el invierno, mientras que k_t representa la tendencia global que manifiestan los pedidos que la fabrica recibe. Debido al incremento poblacional y al afianzamiento de la marca comercializada en el mercado, es natural suponer que κ_t es creciente, y para propósitos de ilistración se ha supuesto que κ_t es una función lineal. Desde luego, esta función de tendencia representa la evolución mercantil de la empresa al transcurrir el tiempo, y es muy importante estimarla. Sin embargo, los datos registrados X_t, además de incluir la inevitable componente aleatoria, incluyen además de κ_t, a la componente estacional s_t. A continuación se verá que, aplicando un operador adecuado de la forma $\sum_k a_k B^k$, dicha componente puede ser *filtrada* (eliminada), dejando solamente la componente de tendecia más un inevitable ruido aleatorio.

Proposición 6.1. Considere la representación (6.1) de la serie $\{X_t\}$, donde se supone que (6.2) y (6.3) son válidas. Defina las constantes a_k como sigue:

$$a_k = \begin{cases} \dfrac{2}{2d}, & \text{si } k = 0, \\ \dfrac{1}{2d}, & \text{si } k \pm 1, \pm 2, \ldots, \pm(d-1), \\ 0, & \text{si } |k| \geq d. \end{cases}$$

En este caso,

$$\sum_{k=-(d-1)}^{(d-1)} a_k B^k X_t = \kappa_t + \frac{1}{2d}\left[2W_t + \sum_{k=1}^{d-1}(W_{t-k} + W_{t+k})\right]. \tag{6.4}$$

El aspecto notable en esta expresión, es que a pesar de que la componente estacional aparecen en la ecuación (6.1), después de aplicar el operador

$$\sum_{k=-(d-1)}^{(d-1)} a_k B^k$$

a la serie $\{X_t\}$, el proceso transformado contiene solamente a la componente de tendencia combinada con un ruido aleatorio inevitable. En otras palabras, el operador deja intacta a la función κ_t, mientras que 'impide el paso' o '*filtra*' a la componente estacional s_t. Estas consideraciones proporcionan una justificación intuitiva para llamar filtro a un operador del tipo $\sum_k a_k B^k$.

La Proposición 6.1 puede demostrarse de manera simple utilizando (7.1)–(7.3); puesto que en este trabajo no se continuará el análisis de la construcción de filtros, el argumento se omite. Un desarrollo de este tema puede encotrarse en Alemán–Valerio (1992).

Estacionaridad

La intención detrás de la construcción de un filtro lineal fue que al transformar un proceso estacionario, la serie obtenida conservara esa propiedad. En esta Sección se verifica que dicho objetivo se alcanza.

Teorem 7.1. Sea $\{X_t\}$ un proceso estacionario y, dada una sucesión $\{a_k\}$ absolutamente sumable, sea $\sum_k a_k B^k$ el correspondiente filtro lineal. En este caso, si

$$Y_t = \sum_k a_k B^k X_t = \sum_k a_k X_{t-k},$$

se tiene que la serie $\{Y_t\}$ también es estacionaria. Más precisamente,

$$E[Y_t] = \mu_{\mathbf{X}} \sum_{k=-\infty}^{\infty} a_k, \tag{7.1}$$

y

$$\operatorname{Cov}[X_t, X_{t+h}] = \gamma_{\mathbf{Y}}(h) == \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} a_k \gamma_{\mathbf{X}}(h+k-j) a_j. \tag{7.2}$$

Demostración. A partir del Teorema 4.1(iv) se obtiene que el valor esperado de Y_t está bien definido, y que

$$\begin{aligned} E[Y_t] &= \lim_{n\to\infty} E\left[\sum_{k=-n}^{k=n} a_k X_{t-k}\right] = \lim_{n\to\infty} \sum_{k=-n}^{k=n} a_k E\left[X_{t-k}\right] \\ &= \lim_{n\to\infty} \sum_{k=-n}^{k=n} a_k \mu_{\mathbf{X}} = \mu_{\mathbf{X}} \sum_{k=-\infty}^{\infty} a_k \end{aligned}$$

donde $\mu_{\mathbf{X}}$ es el valor (constante) de $E[X_t]$. Por lo tanto, $E[Y_t]$ no depende de t. Para concluir, observe que la parte (vi) del Teorema 4.1 implica, a través de la bilinealidad de la covarianza, que

$$\begin{aligned} \operatorname{Cov}\left[Y_t, Y_{t+h}\right] &= \lim_{n\to\infty} \operatorname{Cov}\left[\sum_{k=-n}^{n} a_k Y_{t-k},\ \sum_{j=-n}^{n} a_j Y_{t+h-j}\right] \\ &= \lim_{n\to\infty} \sum_{k=-n}^{n} \sum_{j=-n}^{n} a_k \operatorname{Cov}\left[Y_{t-k},\ Y_{t+h-j}\right] a_j \\ &= \lim_{n\to\infty} \sum_{k=-n}^{n} \sum_{j=-n}^{n} a_k \operatorname{Cov}\left[Y_{t-k},\ Y_{t+h-j}\right] a_j \\ &= \lim_{n\to\infty} \sum_{k=-n}^{n} \sum_{j=-n}^{n} a_k \gamma_{\mathbf{X}}(h+k-j) a_j \\ &= \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} a_k \gamma_{\mathbf{X}}(h+k-j) a_j \end{aligned}$$

y de esta expresión se desprende que $\operatorname{Cov}\left[Y_t, Y_{t+h}\right]$ no depende de t. Por lo tanto, $\{Y_t\}$ es un proceso estacionario, y

$$\gamma_{\mathbf{Y}}(h) = \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} a_k \gamma_{\mathbf{X}}(h+k-j) a_j,$$

concluyendo la demostración del teorema. □

Corolario 7.1. Si $\{Z_t\} \sim \mathrm{WN}\left(0, \sigma^2\right)$ y $Y_t = \sum_k a_k B^k Z_t = \sum_k a_k Z_{t-k}$, entonces

$$E[Y_t] = 0, \quad \text{y} \quad \gamma_{\mathbf{Y}}(h) = \sigma^2 \sum_{k=-\infty}^{\infty} a_k a_{k+h}.$$

Demostración. El corolario se desprende de inmediato de (7.1) y (7.2). En efecto, cuando la serie $\{X_t\}$ se reemplaza por el ruido blanco $\{Z_t\}$ se tiene que $\mu_{\mathbf{Z}} = 0$ de manera que $\mu_{\mathbf{Y}} = \mu_{\mathbf{Z}} \sum_k a_k = 0$. Por otro lado, puesto que $\gamma_{\mathbf{Z}}(0) = \sigma^2$ y $\gamma_{\mathbf{Z}}(t) = 0$ si $t \neq 0$, entonces $\gamma_{\mathbf{Z}}(h + k - j) = 0$ a menos que $h + k = j$, y en este caso, $\gamma_{\mathbf{Z}}(0) = \sigma^2$, de manera que

$$\begin{aligned}\gamma_{\mathbf{Y}}(h) &= \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} a_k \gamma_{\mathbf{Z}}(h + k - j) a_j \\ &= \sum_{k=-\infty}^{\infty} a_k \sigma^2 a_{k+h} \\ &= \sigma^2 \sum_{k=-\infty}^{\infty} a_k a_{k+h}\end{aligned}$$

□

Representación Integral

Suponga que $\{Z_t\}$ es un ruido blanco, y que el proceso $\{X_t\}$ se obtiene aplicando un filtro lineal a $\{Z_t\}$. En esta sección se demuestra que la función de autocovarianza del proceso transformado puede representarse mediante una expresión integral; el resultado será utilizado en el capítulo siguiente con relación a la idea de causalidad.

Teorema 8.1 Sea $\{a_k\}$ una sucesión absolutamente sumable de números reales. Si $\alpha(z)$ es la serie de potencias en (5.5) y $\alpha(B)$ es el filtro lineal correspondiente, defina

$$\mathbf{X}_t = \alpha(B) X_t = \sum_k a_k B^k Z_t = \sum_k a_k Z_{t-k}.$$

En este caso, para todo entero h, se tiene que

$$\gamma_{\mathbf{X}}(h) = \frac{\sigma^2}{2\pi} \int_0^{2\pi} e^{-ih} |\alpha(e^{i\xi})|^2 \, d\xi.$$

Demostración. EL argumento utiliza el siguiente resultado sobre expansiones de Fourier (Apostol, 1981, Churchill, 1976, Rudin, 1976): Si $f(\xi)$ y $g(\xi)$ son dos funciones definidas en $[0, 2\pi]$, y

$$\begin{aligned} f(\xi) &= \sum_k f_k e^{ik\xi} \\ g(\xi) &= \sum_k g_k e^{ik\xi}, \end{aligned} \tag{8.1}$$

entonces,

$$\frac{1}{2\pi}\int_0^{2\pi} f(\xi)\overline{g(\xi)}\, d\xi = \sum_k f_k g_k. \tag{8.2}$$

Ahora, defina

$$f(\xi) = \alpha(e^{i\xi}) = \sum_k a_k e^{ik\xi},$$

de manera que $f_k = a_k$ para todo k, y

$$\begin{aligned} g(\xi) &= e^{ih\xi}\alpha(e^{i\xi}) \\ &= \sum_k a_k e^{i(k+h)\xi} \\ \\ &= \sum_k a_{j-h} e^{i(j)\xi} \end{aligned}$$

de donde se desprende que $g_k = a_{k-h}$ para todo k. Aplicando ahora (8.2), se desprende que

$$\begin{aligned} \frac{1}{2\pi}\int_0^{2\pi} e^{-ih\xi}|\alpha(e^{i\xi})|^2\, d\xi &= \frac{1}{2\pi}\int_0^{2\pi} \alpha(e^{i\xi})\overline{e^{ih\xi}\alpha(e^{i\xi})}\, d\xi \\ &= \sum_k f_k g_k \\ &= \sum_k a_k a_{k-h} \\ &= \sum_k a_k a_{k+h}, \end{aligned}$$

y utilizando el Corolario 7.1, se obtiene que

$$\frac{\sigma^2}{2\pi}\int_0^{2\pi} e^{-ih\xi}|\alpha(e^{i\xi})|^2\, d\xi = \sigma^2 \sum_k a_k a_{k+h} = \gamma_{\mathbf{X}}(h).$$

□

El Teorema 8.1 permite responder a una pregunta interesante: Suponga que $\{Z_t\} \sim \mathrm{WN}\left(0, \sigma^2\right)$, ¿Bajo qué condiciones el proceso $\{X_t\}$ obtenido aplicando un filtro $\{Z_t\}$ también es un ruido blanco?

Corolario 8.1 Suponga que $\{a_k\}$ una sucesión absolutamente sumable de números reales y sean $\alpha(z)$ y $\alpha(B)$ como en el Teorema 8.1. Defina

$$\mathbf{X}_t = \alpha(B)X_t = \sum_k a_k B^k Z_t = \sum_k a_k Z_{t-k}.$$

En este caso, $\{X_t\}$ es un ruido blanco si

$$|\alpha(e^{i\xi})| \equiv C \tag{8.3}$$

es constante en el intervalo $[0, 2\pi]$, y en este caso,

$$\{X_t\} \sim \mathrm{WN}\left(0, C^2\sigma^2\right).$$

Demostración. La propiedad que caracteriza a un ruido blanco es que su función de autocovariana se anula para agumentos diferentes de cero. Suponga que (8.3) ocurre, y sea h un entero arbitrario. En este caso el Teorema 8.1 implica que

$$\begin{aligned}
\gamma_{\mathbf{X}}(h) &= \frac{\sigma^2}{2\pi}\int_0^{2\pi} e^{-ih\xi}|\alpha(e^{i\xi})|^2\, d\xi \\
&= \frac{\sigma^2}{2\pi}\int_0^{2\pi} e^{-ih\xi}|C^2\, d\xi \\
&= \frac{C^2\sigma^2}{2\pi}\int_0^{2\pi} e^{-ih\xi}|\, d\xi
\end{aligned}$$

Cuando $h \neq 0$, la última integral se anula y entonces $\gamma_{\mathbf{x}}(h) = 0$, mientras que si $h = 0$ la integral asume el valor 2π, de manera que $\gamma(0) = C^2\sigma^2$, de tal forma que $\{X_t\} \sim \mathrm{WN}\left(0, C^2\sigma^2\right)$. □

Capítulo 5

Composición e Inversión

En este capítulo se estudian dos problemas básicos del álgebra de filtros lineales, a saber, determinar la composición de dos filtros, y encontrar condiciones suficientes para que un filtro sea invertible. Los resultados en estos temas establecen una relación entre los problemas algebraicos referentes a filtros, y operaciones con funciones analíticas de variable compleja.

Introducción

El objetivo de este capítulo es estudiar la composición de dos filtros lineales, y la construcción de la transformación inversa de un filtro. Los resultados que se obtienen en estas dos vertientes, pueden resumirse de la siguiente forma:

(i) La composición de dos filtros lineales es, de nueva cuenta, un filtro lineal, y la sucesión correspondiente al filtro compuesto se obtiene mediante la operación algebraica de convolución de dos sucesiones; cuando los filtros originlaes inducen funciones analíticas en un anillo abierto, esta conclusión es equivalente a que los coeficientes de la composición de dos filtros se obtienen multiplicando las dos series de Laurent correspondientes a los filtros originales.

(ii) Con relación a la inversión de un filtro, se demuestra que un filtro es invertible siempre que la serie de Laurent correspondiente converja en un anillo abierto alrededor del círculo unitario, en la cual dicha serie no se anule; en este caso, la serie de Laurent asociada al filtro inverso es el recíproco de la serie de Laurent correspondiente al filtro original.

Estos resultados establecen una relación estrecha entre las operaciones de composición e inversión de filtros y las propiedades de funciones de variable compleja. *La organización del material es la siguiente*: En la Sección 2 se trata el problema de determinar la composición de dos filtro lineales, estableciendo que dicha operación produce un nuevo filtro cuyos coeficientes se obtienen convolucionando las sucesiones correspondientes a los dos filtros originales. En la Sección 3 se explora la relación entre la multiplicación de funciones analíticas y la composición de filtros, mientras que en la Sección 4 se proporcionan condiciones suficientes para que un filtro admita una transformación inversa.

Composición de Filtros Lineales

Suponga que $\psi = \{\psi_k\}$ y $\varphi = \{\varphi_k\}$ son dos sucesiones absolutamente sumables y sean $\psi(B)$ y $\varphi(B)$ los filtros correspondientes, esto es, dado un proceso $\{X_t\}$ para el cual sus componentes tienen segundo momento acotado, entonces para todo t

$$\psi(B)X_t = \sum_k \psi_k X_{t-k}, \quad \text{y} \quad \varphi(B)X_t = \sum_k \varphi_k X_{t-k}. \tag{2.1}$$

El propósito de esta sección es estudiar la composición de los dos filtros *arbitrarios* $\psi(B)$ y $\varphi(B)$, estableciendo que la transformación obtenida aplicando los dos filtros en forma sucesiva es, de nueva cuenta, un filtro lineal. En forma más precisa, dada una serie $\{W_t\}$, aplicando $\psi(B)$ se obtiene una nueva serie $\{X_t\}$ dada por

$$X_t = \psi(B)W_t = \sum_k \psi_k W_{t-k}, \tag{2.2}$$

y si a continuación se transforma este proceso por medio del filtro $\varphi(B)$, se obtiene el proceso $\{Y_t\}$ determinado a través de

$$Y_t = \varphi(B)X_t = \sum_k \varphi_k X_{t-k}. \tag{2.3}$$

Así, este procedimiento de dos etapas que conduce de $\{W_t\}$ a $\{Y_t\}$ puede resumirse escribiendo

$$Y_t = \varphi(B)X_t = \varphi(B)[\psi(B)W_t], \tag{2.4}$$

de manera que $\{Y_t\}$ se obtiene aplicando *la composición* de los filtros $\psi(B)$ y $\varphi(B)$ al proceso original $\{W_t\}$. Dicha composición se denota mediante $\varphi(B) \circ \psi(B)$, y transforma, directamente, a la serie original $\{W_t\}$ en el proceso $\{Y_t\}$:

$$\varphi(B) \circ \psi(B) \quad \text{convierte} \quad \{W_t\} \quad \text{en} \quad \{Y_t\}. \tag{2.5}$$

De acuerdo al resultado principal de esta sección, establecido en el Teorema 2.1 líneas más adelante, la composición $\varphi(B) \circ \psi(B)$ es también un filtro. Para formular este resultado, es conveniente introducir formalemente una operación entre sucesiones, la cual es familiar en la teoría estadística de variables aleatorias discretas Dudewicz y Mishra (1988), Mood *et al.*, (1985).

Definición 2.1. Sean $\psi = \{\psi_k\}$ y $\varphi = \{\varphi_k\}$ dos sucesiones absolutamente sumables. En este caso, la *convolución* de φ y ψ se denota mediante $\varphi \star \psi$ y es la sucesión determinada mediante

$$[\varphi \star \psi]_k = \sum_s \varphi_s \psi_{k-s}. \tag{2.6}$$

Teorema 2.1. Si $\varphi(B)$ y $\psi(B)$ son dos filtro lineales, entonces su composición $\varphi(B) \circ \psi(B)$ también es un filtro lineal. De forma precisa,

$$\varphi(B) \circ \psi(B) = [\varphi \star \psi](B), \tag{2.7}$$

esto es, la composición de $\varphi(B)$ y $\psi(B)$ es el filtro asociado a $\varphi \star \psi$, la convolución de las sucesiones que dan origen a $\varphi(B)$ y $\psi(B)$, respectivamente.

La demostración de este resultado depende de los siguientes lemas técnicos.

Lema 2.1. [Ash (1975), Rudin (1976).] Si $\psi = \{\psi_k\}$ y $\varphi = \{\varphi_k\}$ son dos sucesiones abolutamente sumables, entonces su convolución $\varphi \star \psi$ también tiene dicha propiedad.

Demostración. Recuerde que la suma de una serie de términos no negativos no depende del orden en que se agreguen los sumandos. Con esto en mente, observe que

$$\begin{aligned}
\sum_k |[\varphi \star \psi]_k| &= \sum_k \left|\sum_s \varphi_s \psi_{k-s}\right| \\
&\leq \sum_k \sum_s |\varphi_s \psi_{k-s}| = \sum_k \sum_s |\varphi_s|\,|\psi_{k-s}| \\
&= \sum_s \sum_k |\varphi_s|\,|\psi_{k-s}| = \sum_s |\varphi_s| \left[\sum_k |\psi_{k-s}|\right] \\
&= \sum_s |\varphi_s| \left[\sum_j |\psi_j|\right] = \left[\sum_s |\varphi_s|\right] \left[\sum_j |\psi_j|\right] < \infty,
\end{aligned}$$

donde la última desigualdad se desprende de la sumabilidad absoluta de las sucesiones φ y ψ. Este argumento muestra que la convolución $\varphi \star \psi$ es absoultamente sumable y concluye la demostración del lema. □

Lema 2.2. Sean $\psi = \{\psi_k\}$ y $\varphi = \{\varphi_k\}$ son dos sucesiones abolutamente sumables, y suponga que el proceso $\{W_t\}$ es tal que todas sus componentes tienen segundo momento acotado por una constante M. En este caso, existe un evento Ω_1 tal que $P[\Omega_1] = 1$, y

$$\sum_k \sum_s |\varphi_k \psi_s W_{t-k-s}(\omega)| < \infty, \quad \omega \in \Omega_1.$$

Demostración. Puesto que $E[W_t^2] \leq M$, se desprende que $E[|W_t|] \leq \sqrt{E[W_t^2]} \leq \sqrt{M}$; vea el argumento en la demostración del Teorema 4.1(i) del Capítulo 4. Por lo tanto, usando el Teorema de convergencia monótona, se desprende que

$$\begin{aligned}
E\left[\sum_k \sum_s |\varphi_k \psi_s W_{t-k-s}|\right] &= \sum_k \sum_s E\left[|\varphi_k \psi_s W_{t-k-s}|\right] \\
&= \sum_k \sum_s |\varphi_k|\,|\psi_s| E\left[|W_{t-k-s}|\right] \\
&\leq \sum_k \sum_s |\varphi_k|\,|\psi_s|\sqrt{M}
\end{aligned}$$

y puesto que las sucesiones φ y ψ son absolutamente sumables, se concluye que

$$E\left[\sum_k \sum_s |\varphi_k \psi_s W_{t-k-s}|\right] \leq \sqrt{M} \left[\sum_k |\varphi_k|\right] \left[\sum_s |\psi_s|\right] < \infty$$

Defina ahora

$$\Omega_1 = \left\{ \omega \,\middle|\, \sum_k \sum_s |\varphi_k \psi_s W_{t-k-s}(\omega)| < \infty \right\}.$$

En este caso, la anterior de desigualdad desplegada implica que $P[\Omega_1] = 1$, mientras que de la definición de Ω_1 se desprende que $\sum_k \sum_s |\varphi_k \psi_s W_{t-k-s}(\omega)| < \infty$ siempre que $\omega \in \Omega_1$.

Demostración del Teorema 2.1 Como punto de partida, note que el Lema 3.1 muestra que el filtro $\varphi \star \psi(B)$ está bien definido, pues la sucesión asociada a $\varphi \star \psi$ es absolutamente sumable. Para completar la demostración del teorema, es suficiente mostrar que para un proceso $\{W_t\}$ con segundos momentos acotados, si $\{X_t\}$ y $\{Y_t\}$ son las series dadas en (2.2) y (2.3), entonces

$$Y_t = \sum_k [\varphi \star \psi]_k B^k W_t,$$

igualdad que equivale a

$$Y_t = \sum_k [\varphi \star \psi]_k W_{t-k}. \tag{2.8}$$

Para establecer esta relación, primero observe que existe un evento Ω' cuya probabilidad es uno, y tal que para todo t

$$X_t(\omega) = \sum_s \psi_s W_{t-s}(\omega), \quad \omega \in \Omega', \tag{2.9}$$

donde

$$\sum_s |\psi_s W_{t-s}|(\omega) < \infty, \quad \omega \in \Omega'; \tag{2.10}$$

vea el Teorema 4.1(ii) del Capítulo 4. Similarmente, existe otro evento Ω'' de probabilidad uno, para el cual se satisfacen las siguientes relaciones para todo t:

$$Y_t(\omega) = \sum_k \varphi_k X_{t-k}(\omega), \quad \omega \in \Omega'', \tag{2.11}$$

donde la serie en el lado derecho converge absolutamente, i.e.,

$$\sum_k |\psi_k X_{t-k}|(\omega) < \infty, \quad \omega \in \Omega''. \tag{2.12}$$

Defina ahora el evento $\widetilde{\Omega}$ mediante

$$\widetilde{\Omega} = \Omega' \cap \Omega'' \cap \Omega_1$$

donde Ω_1 es como en el Lema 2.2. Puesto que cada evento en la definición de $\widetilde{\Omega}$ tiene probabilidad uno, se desprende que $P[\widetilde{\Omega}] = 1$, y que las igualdades en (2.9)–(2.12) son válidas si $\omega \in \widetilde{\Omega}$. En particular, las series que determinan a $X_t(\omega)$ y $Y_t(\omega)$ en (2.9) y (2.11) convergen absolutamente y $\sum_k \sum_s |\varphi_k \psi_s W_{t-k-s}(\omega)| < \infty$. Seleccione $\omega \in \widetilde{\Omega}$ y observe que

$$Y_t(\omega) = \sum_k \varphi_k X_{t-k}(\omega) = \sum_k \varphi_k \left[\sum_s \psi_s W_{t-k-s}(\omega)\right].$$

Puesto que la (doble) serie en el extremo derecho es absolutamente sumable, los términos que la conforman pueden agregarse en cualquier orden sin alterar el valor de la suma, de tal forma que

$$\begin{aligned}
Y_t(\omega) &= \sum_k \varphi_k \left[\sum_s \psi_s W_{t-k-s}(\omega)\right] \\
&= \sum_k \sum_s \varphi_k \psi_s W_{t-k-s}(\omega) \\
&= \sum_j \sum_{(k,s):k+s=j} \varphi_k \psi_s W_{t-k-s}(\omega) \\
&= \sum_j \sum_k \varphi_k \psi_{j-k} W_{t-j}(\omega) \\
&= \sum_j W_{t-j}(\omega) \sum_k \varphi_k \psi_{j-k} \\
&= \sum_j W_{t-j}(\omega) [\varphi \star \psi]_j
\end{aligned}$$

donde la última igualdad se debe a la definición de convolución. Por lo tanto,

$$Y_t(\omega) = \sum_j [\varphi \star \psi]_j W_{t-j}(\omega), \quad \omega \in \widetilde{\Omega},$$

puesto que $\widetilde{\Omega}$ es un evento de probabilidad uno, se desprende que

$$Y_t = \sum_j [\varphi \star \psi]_j W_{t-j} = \sum_j [\varphi \star \psi]_j B^j W_t = [\varphi \star \psi](B) W_t,$$

concluyendo la demostración del teorema. □

Note que la convolución de dos sucesiones absolutamente sumables es una operación conmutativa, esto es, $\varphi \star \psi = \psi \star \varphi$. Por lo tanto,

$$\varphi(B) \circ \psi(B) = [\varphi \star \psi](B) = [\psi \star \varphi](B) = \psi(B) \circ \varphi(B),$$

igualdad que conduce a establecer la siguiente conclusión.

Corolario 2.1. La composición de dos filtros lineales es una operación conmutativa.

Convolución y Multiplicación de Series

De acuerdo al Teorema 2.1, la composición de dos filtros $\varphi(B)$ y $\psi(B)$ es, de nueva cuenta un filtro, el cual está asociado a la convolución $\varphi \star \psi$:

$$\varphi(B) \circ \psi(B) = [\varphi \star \psi](B). \tag{3.1}$$

Luego, encontrar la composición de dos filtros se reduce al problema de encotrar la convolución de las sucesiones que les dan origen. En esta sección se muestra que para determinar la convolución de dos sucesiones, es suficiente calcular el producto de dos series de potencias de variable compleja z.

Lema 3.1.[Alfhors, 1970, Nevannlina y Pattero, 1975, Churchill, 1976.] Suponga que $\psi = \{\psi_k\}$ y $\varphi = \{\varphi_k\}$ son dos sucesiones abolutamente sumables, y sean $\varphi(z)$ y $\psi(z)$ las dos series de potencias asociadas, las cuales convergen en dos anillos $\mathcal{A}_\varphi$ y $\mathcal{A}_\psi$ que contienen al disco unitario, esto es,

$$\varphi(z) = \sum_k \varphi_k z^k, \quad z \in \mathcal{A}_\varphi,$$
$$\psi(z) = \sum_k \psi_k z^k, \quad z \in \mathcal{A}_\psi.$$

En este caso, la serie de potencias asociada a la convolución $\varphi \star \psi$ es el producto de $\varphi(z)$ y $\psi(z)$ en la región $\mathcal{A}_\varphi \cap \mathcal{A}_\psi$, esto es,

$$[\varphi \star \psi](z) = \sum_k [\varphi \star \psi]_k z^k = \varphi(z)\psi(z), \quad z \in \mathcal{A}_\varphi \cap \mathcal{A}_\psi.$$

Demostración. Las expansiones de Laurent para $\varphi(z)$ y $\psi(z)$ convergen absolutamente en $\mathcal{A}_\varphi \cap \mathcal{A}_\psi$, de manera que en la serie doble que aparece a continuación los términos pueden agregarse en un orden arbitrario sin alterar el resultado, siempre que $z \in \mathcal{A}_\varphi \cap \mathcal{A}_\psi$:

$$\begin{aligned}
\varphi(z)\psi(z) &= \sum_k \varphi_k z^k \sum_s \psi_s z^s = \sum_k \sum_s \varphi_k z^k \psi_s z^s \\
&= \sum_k \sum_s \varphi_k \psi_s z^{k+s} = \sum_j \sum_{(k,s):k+s=j} \varphi_k \psi_s z^{k+s} \\
&= \sum_j z^j \sum_k \varphi_k \psi_{j-k} = \sum_j [\varphi \star \psi]_j z^j
\end{aligned}$$

y por lo tanto,

$$\varphi(z)\psi(z) = [\varphi \star \psi](z), \quad z \in \mathcal{A}_\varphi \cap \mathcal{A}_\psi.$$

□

Este resultado permite, en algunos casos, determinar la composición de dos filtros a través del siguiente procedimiento:

(i) Determine las series de potencias correspondientes a los filtros involucrados y multiplíquelas;

(ii) Obtenga la expansión en serie de Laurent del producto, y

(iii) En la anterior expansión de Laurent reemplace la variable compleja z por el operador B, determinando de esta forma la composición de los filtros iniciales.

Este procedimiento se ilustra ahora mediante dos ejemplos.

Ejemplo 3.1. Considere los $\alpha(B)$ y $\beta(B)$ filtros cuya acción sobre una serie $\{X_t\}$ está determinada por

$$Y_t = \alpha(B)X_t = \sum_{k=0}^{\infty} \frac{1}{3^k} X_{t-k},$$

$$Y_t = \beta(B)X_t = \sum_{k=0}^{\infty} \frac{(-1)^k}{3^k} X_{t-k}.$$

El problema es determinar la composición $\alpha(B) \circ \beta(B)$. Con este fin, note que

$$\alpha(B)X_t = \sum_{k=0}^{\infty} \frac{1}{3^k} X_{t-k} = \sum_{k=0}^{\infty} \frac{1}{3^k} B^k X_t$$
$$\beta(B)X_t = \sum_{k=0}^{\infty} \frac{(-1)^k}{3^k} X_{t-k} = \sum_{k=0}^{\infty} \frac{(-1)^k}{3^k} B^k X_t$$

Por lo tanto, las series de Laurent para los filtros considerados son

$$\alpha(z) = \sum_{k=0}^{\infty} \frac{1}{3^k} z^k = \sum_{k=0}^{\infty} \frac{z}{3^k} = \frac{1}{1 - z/3}, \quad |z| < 3,$$

y

$$\beta(z) = \sum_{k=0}^{\infty} \frac{(-1)^k}{3^k} z^k = \sum_{k=0}^{\infty} \left(\frac{-z}{3}\right)^k = \frac{1}{1-(-z/3)} = \frac{1}{1+z/3}, \quad |z| < 3,$$

De acuerdo al Lema 3.1, la convolución de las sucesiónes $\alpha = \{\alpha_k\}$ y $\beta = \{\beta_k\}$ está determinada por

$$[\alpha \star \beta](z) = \alpha(z)\beta(z) = \left(\frac{1}{1 - z/3}\right)\left(\frac{1}{1 + z/3}\right) = \left(\frac{1}{1 - z^2/9}\right),$$

expresión que es válida siempre que $|z| < 3$. Por lo tanto, los términos de la sucesión $[\alpha \star \beta]$ se obtienen determinando la serie de Laurent para la expresión en el lado derecho. En este caso, dicho desarrollo se obtiene via la ecuación (5.13) en el Capítulo 4:

$$[\alpha \star \beta](z) = \alpha(z)\beta(z) = \left(\frac{1}{1 - z^2/9}\right) = \sum_{k=0}^{\infty} \frac{z^{2k}}{9^k},$$

de manera que $[\alpha \star \beta]_k = 0$ si $k < 0$ o si k es impar, mientras que $[\alpha \star \beta]_{2k} = 1/9^k$ siempre que k sea un entero no negativo. La composición de los filtros $\alpha(B)$ y $\beta(B)$ se determina reemplazando z por B en el lado derecho de la última igualdad desplegada:

$$\alpha(B) \circ \beta(B) = [\alpha \star \beta](B) = \sum_{k=0}^{\infty} \frac{B^{2k}}{9^k},$$

de manera que

$$[\alpha(B)\circ\beta(B)]X_t=\sum_{k=0}^{\infty}\frac{B^{2k}X_t}{9^k}=\sum_{k=0}^{\infty}\frac{1}{9^k}X_{t-2k}.$$

Ejemplo 3.2. Suponga ahora que $\alpha(B)$ es como en el Ejemplo 3.1, esto es,

$$\alpha(B)X_t=\sum_{k=0}^{\infty}\frac{1}{3^k}X_{t-k},$$

mientras que $\beta(B)$ está determinado mediante

$$\beta(B)X_t=\sum_{k=0}^{\infty}\frac{1}{2^k}X_{t-k}.$$

Como antes, el problema es determinar la composición $\alpha(B)\circ\beta(B)$. Para alcanzar este objetivo, note que

$$\alpha(B)X_t=\sum_{k=0}^{\infty}\frac{1}{3^k}X_{t-k}=\sum_{k=0}^{\infty}\frac{1}{3^k}B^kX_t$$
$$\beta(B)X_t=\sum_{k=0}^{\infty}\frac{1}{2^k}X_{t-k}=\sum_{k=0}^{\infty}\frac{1}{2^k}B^kX_t$$

Por lo tanto, las series de Laurent para los filtros considerados son

$$\alpha(z)=\sum_{k=0}^{\infty}\frac{1}{3^k}z^k=\sum_{k=0}^{\infty}\frac{z}{3^k}=\frac{1}{1-z/3},\quad |z|<3,$$

y

$$\beta(z)=\sum_{k=0}^{\infty}\frac{1}{2^k}z^k=\sum_{k=0}^{\infty}\left(\frac{z}{2}\right)^k=\frac{1}{1-z/2}\quad |z|<2.$$

Como en el ejemplo precedente, la convolución de las sucesiónes $\alpha=\{\alpha_k\}$ y $\beta=\{\beta_k\}$ está determinada por el producto de las series:

$$[\alpha\star\beta](z)=\alpha(z)\beta(z)=\left(\frac{1}{1-z/3}\right)\left(\frac{1}{1-z/2}\right),$$

expresión que es válida siempre que $|z|<2$. Para obtener la expansión de Laurent, se utilizará el procedimiento delineado en el Corolario 5.1(iv), basado en el desarrollo en fracciones parciales:

$$\begin{aligned}\left(\frac{1}{1-z/3}\right)\left(\frac{1}{1-z/2}\right)&=\frac{-2}{1-z/3}+\frac{3}{1-z/2}\\&=-2\sum_{k=0}^{\infty}\frac{z^k}{3^k}+3\sum_{k=0}^{\infty}\frac{z^k}{2^k}\end{aligned}$$

y entonces

$$[\alpha \star \beta](z) = \sum_{k=0}^{\infty} \left(\frac{3}{2^k} - \frac{2}{3^k} \right) z^k.$$

Por lo tanto,

$$\alpha(B) \circ \beta(B) = [\alpha \star \beta](B) = \sum_{k=0}^{\infty} \left(\frac{3}{2^k} - \frac{2}{3^k} \right) B^k.$$

El Problema de Inversión

Sea $\alpha(B) = \sum_k a_k B^k$ un filtro determinado. El problema que se considera a continuación es determinar si existe un filtro $\beta(B)$ tal que, para todo t,

$$\beta(B) \circ \alpha(B) X_t = X_t$$

siempre que las componentes de la serie $\{X_t\}$ tengan segundo momento acotado. En este caso, si $\alpha(B)$ transforma al proceso $\{X_t\}$ en la serie $\{Y_t\}$, i.e., $Y_t = \alpha(B)X_t$, al aplicar el filtro $\beta(B)$ al nuevo proceso, se obtiene la igualdad $\beta(B)Y_t = X_t$, lo cual significa que $\beta(B)$ 'deshace' la transformación que el filtro $\alpha(B)$ realizó. El filtro $\beta(B)$ es el inverso de $\alpha(B)$. El principal resultado de esta sección, proporciona una condición suficiente para que un filtro $\alpha(B)$ admita un inverso.

Teorema 4.1. Suponga que el filtro $\alpha(B) = \sum_k a_k B^k$ es tal que la serie de Laurent correspondiente converge en un anillo que contiene al círculo unitario, esto es,

$$\alpha(z) = \sum_k a_k z^k, \quad r < |z| < R,$$

donde $r < 1 < R$. En este caso, si

$$\alpha(z) \neq 0 \quad \text{cuando} \quad |z| = 1,$$

se tiene que $\alpha(B)$ admite una transformación inversa, la cual es el filtro $\beta(B)$ correspondiente a la expansión de Laurent de $1/\alpha(z)$, i.e.,

$$\beta(z) = \sum_k b_k z^k = \frac{1}{\alpha(z)}.$$

Demostración. El argumento inicia observando que el hecho de que $\alpha(z) \neq 0$ cuando $|z| = 1$, implica que $\alpha(z) \neq 0$ para un conjunto abierto que contiene al círculo unitario. Luego, existen números positivos r' y R' tales que $r' < 1 < R'$, tales que $\alpha(z) \neq 0$ si $r' < |z| < R'$, por lo que no se incurre en pérdida de generalidad alguna si se supone que $\alpha(z) \neq 0$ en el anillo $r < |z| < R$. En este caso, el recíproco de $1/\alpha(z)$ es una función analítica en el anillo $r < |z| < R$, y se desarrolla en serie de Laurent (Alfhors (1970), Apostol (1981), Rudin (1976)), i.e.,

$$\beta(z) = \frac{1}{\alpha(z)} = \sum_k b_k z^k.$$

En este caso, $\beta(z)\alpha(z) = 1$, de manera que

$$\beta(B) \circ \alpha(B) = 1B^0 = I,$$

y por lo tanto el Teorema 2.1 y el Lema 3.1 implican que

$$[\beta(B) \circ \alpha(B)]X_t = X_t$$

para todo proceso $\{X_t\}$. □

De acuerdo a este resultado, cuando la serie $\alpha(z)$ converge en un anillo abierto que contiene al círculo unitario $\mathcal{C}$, y $\alpha(z) \neq 0$ si $z \in \mathcal{C}$, obtener el inverso del filtro $\alpha(z)$ es un problema equivalente a expander en serie de Laurent a la función $\beta(z) = 1/\alpha(z)$. Esta idea se ilustra a continuación.

Ejemplo 4.1. Considere el filtro $\alpha(B) = 1 - 2.5B + B^2$. En este caso, $\alpha(z) = 1 - 2.5z + z^2 = (1 - 2z)(1 - z/2)$ se anula sólo para $z = 2$ y $z = 1/2$, y por lo tanto, $\alpha(z) \neq 0$ sobre el círculo unitario. Por lo tanto, el filtro inverso $\beta(B)$ está determinado por la igualdad

$$\beta(z) = \frac{1}{\alpha(z)} = \frac{1}{(1 - 2z)(1 - z/2)},$$

donde se debe desarrollar en serie de Laurent el lado derecho de esta igualdad. Desarrollando en fracciones parciales, y usando el Lema 5.1 se obtiene

$$\begin{aligned}\beta(z) &= \frac{1}{(1-2z)(1-z/2)}\\ &= \frac{4/3}{1-2z} - \frac{1/3}{1-z/2}\\ &= -\frac{4}{3}\sum_{k=1}^{\infty}\frac{1}{(2z)^k} - \frac{1}{3}\sum_{k=0}^{\infty}\frac{z^k}{2^k}\\ &= -\frac{4}{3}\sum_{k=-\infty}^{1}(2z)^k - \frac{1}{3}\sum_{k=0}^{\infty}\frac{z^k}{2^k}\end{aligned}$$

y por lo tanto, el filtro inverso de $\alpha(B) = I - 2.5B + B^2$, es

$$\beta(B) = -\frac{4}{3}\sum_{k=-\infty}^{1} 2^k B^k - \frac{1}{3}\sum_{k=0}^{\infty}\frac{1}{2^k}B^k.$$

Procesos ARMA

En estas sección se introduce un tipo importante de procesos estacionarios con media nula, los cuales pertenecen a la categoría de filtros lineales y se conocen como procesos ARMA. Dichas series desempeñan un papel central en el proceso de construcción de modelos para una serie de datos observados, debido a la siguiente propiedad: Dada una función de autocovarianza muestral $\widehat{\gamma}$, existe un proceso ARMA cuya función de autocovariaza coincide con $\widehat{\gamma}$ hasta un orden arbitrario menor que el número de observaciones (Brockwell y Davis (1991, 1996), Polendo–Luis (1999)). Esta es una propiedad sumamente relevante, pues desde la perspectiva de segundo orden, la función de autocovarianza de un proceso captura los rasgos esenciales del proceso observado.

Definición 5.1. [Proceso Autorregresivo y de Promedios Móviles (ARMA).] Sea $\{X_t\}$ cuyas componentes tienen media nula. En este caso, $\{X_t\}$ es un proceso autorregresivo de promedio móvil de orden (p, q), si existe un proceso $\{Z_t\} \sim$

WN $(0, \sigma^2)$ y constantes $\phi_1, \phi_2, \ldots, \phi_p$ y $\theta_1, \theta_2, \ldots, \theta_q$ tales que

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q}. \quad (5.1)$$

En este caso, se escribe $\{X_t\} \sim$ ARMA (p, q), expresión que se lee como '$\{X_t\}$ es un proceso ARMA (p, q)'.

Defina los polinomios $\varphi(z)$ y $\theta(z)$ mediante

$$\begin{aligned} \phi(z) &= 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p \\ \theta(z) &= 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q, \end{aligned} \quad (5.2)$$

los cuales se denominan polinomios autorregresivo y de promedios móviles, respectivamente. Con esta notación, igualdad (5.1) se escribe en forma compacta como

$$\phi(B) X_t = \theta(B) Z_t. \quad (5.3)$$

Dado un ruido blanco $\{Z_t\}$, el siguiente resultado trata sobre la existencia de proceso $\{X_t\}$ que satisface (5.1) o (5.3).

Teorema 5.1. [Existencia y Unicidad de procesos ARMA.] Suponga que que el polinomio autoregresivo $\varphi(z)$ no tiene raíces en el círculo unitario. En este caso, existe un único proceso $\{X_t\}$ que satsiface (5.1) y (5.3).

Demostración. Por hipótesis, $\varphi(z) \neq 0$ siempre que $|z| = 1$. Luego, existen números r y R que satisfacen $r < 1 < R$, y tales que $\varphi(z) \neq 0$ siempre que $r < |z| < R$. Por lo tanto, el filtro $\varphi(B)$ tiene un inverso $\beta(B)$ determinado a través de la igualdad

$$\beta(z) = \sum_k b_k z^k = \frac{1}{\varphi(z)}, \quad r < |z| < R.$$

• Existencia: Defina el proceso $\{X_t\}$ mediante

$$X_t = \beta(B) \circ \theta(B) Z_t. \quad (5.4)$$

En este caso, ya que $\varphi(B)\beta(B) = I$, se desprende que

$$\begin{aligned}\varphi(B)X_t &= \varphi(B) \circ \beta(B) \circ \theta(B)Z_t \\ &= [\varphi(B) \circ \beta(B)] \circ \theta(B)Z_t \\ &= [I] \circ \theta(B)Z_t \\ &= \theta(B)Z_t,\end{aligned}$$

i.e, $\{X_t\}$ satisface (5.3), y por lo tanto, también su equivalente (5.1).

• Unicidad: Suponga que $\{X_t\}$ satisface (5.3). Aplicando el filtro $\beta(B)$ en ambos lados de (5.3) se obtiene

$$\beta(B) \circ \varphi(B)X_t = \beta(B) \circ \theta(B)Z_t,$$

y como $\beta(B) \circ \varphi(B) = I$, se desprende que X_t está dada por la expresión (5.4). □De acuerdo al terema precedente, un proceso ARMA existe siempre que el correspondiente polinomio autorregresivo satisfaga $\varphi(z) \neq 0$ cuando $|z| = 1$, condición que se asumirá en el desarrollo subsecuente.

Causalidad

Dado un filtro lineal $\alpha(B)$ su acción sobre una serie $\{Z_r\}$ produce un nuevo proceso $\{X_t\}$ determinado por

$$X_t = \alpha(B)Z_t = \sum_k a_k B^k Z_t = \sum_k a_k Z_{t-k}. \tag{6.1}$$

Frecuentemente, el proceso $\{\mathbf{Z}_r\}$ se interpreta como un ruido blanco cuyo término Z_r se manifiesta en el tiempo r. En la anterior igualdad, la variable X_t es el dato que el observador registra en el momento presente t, y se expresa como una combinación de las componentes del ruido $\{Z_r\}$. Note, sin embargo, que si $a_k \neq 0$ para un entero $k < 0$, entonces $a_k Z_{t-k}$ es una variable aleatoria cuyo efecto se manifestará en el futuro. Por ejemplo, si $k = -2$, $a_k Z_{t-k} = a_{-2}Z_{t+2}$ tiene un efecto que se presenta en el momento $t + 2$, dos unidades de tiempo más adelante del momento presente. Desde el punto de vista de un modelador, esta característica significa que el dato registrado en el momento actual t depende de perturbaciones aleatorias que aún no se manifiestan, y no es considerada, por regla general, como un rasgo aceptable. Por el contrario, si $a_k = 0$ para todo $k < 0$, entonces X_t en (6.1) se expresa como una combinación de variables aleatorias Z_r con $r \leq t$, esto es, de variables cuyos efectos se manifiestan en el pasado o en momento presente.

Definición 6.1. Un filtro lineal $\alpha(B) = \sum_k a_k B^k$ se denomina causal, si y sólo si $a_k = 0$ para todo $k < 0$.

El propósito de esta sección es mostrar que un proceso ARMA (p, q) se expresa aplicando un filtro causal a un ruido blanco adecuado. Por conveniencia, estos resultados se presentan en dos partes establecidas a continuación como los Teoremas 6.1 y 6.2.

Teorema 6.1. Suponga que el proceso $\{X_t\}$ es un proceso ARMA (p, q) el cual satisface

$$\varphi(B)X_t = \theta(B)Z_t,$$

donde los polinomios $\varphi(z)$ y $\theta(z)$ están dados en (5.2). Suponga que

$$\varphi(z) \neq 0 \quad \text{si } |z| \leq 1.$$

En este caso, el filtro

$$\alpha(B) = \frac{\theta(B)}{\varphi(B)} = \sum_k a_k B^k$$

es causal, y por lo tanto

$$X_t = \sum_{k=0}^{\infty} a_k B^k Z_t = \sum_{k=0}^{\infty} a_k Z_{t-k}.$$

Demostración. Como el polinomio $\varphi(z)$ no se anula en el disco unitario, existe un número $R > 1$ tal que $\varphi(z) \neq 0$ siempre que $|z| < R$. En este caso, el recíproco de $\beta(z) = 1/\varphi(z)$ es una función analítica en en la región $|z| < R$, de manera que

$$\beta(z) = \frac{1}{\varphi(z)} = \sum_{k=0}^{\infty} b_k z^k, \quad |z| < R,$$

y enotonces

$$\alpha(z) = \beta(z)\theta(z) = \frac{\theta(z)}{\varphi(z)} = [1 + \sum_{j=1}^{q} \theta_j z^j] \sum_{k=0}^{\infty} b_k z^k, \quad |z| < R,$$

y esta expansión no involucra potencias negativas de z. Por lo tanto, realizando la multiplicación se obtiene que $\alpha(z) \sum_{k=0}^{\infty} a_k z^k$ en un disco que contiene al círculo unitario, y el filtro correspondiente $\alpha(B)$ es causal. □

Cuando el polinomo autorregresivo $\varphi(z)$ no se anula en el disco unitario, el proceso $\{X_t\} \sim$ ARMA (p, q) es tal que X_t se expresa en términos de $\{Z_t\}$ mediante un filtro causal, lo cual implica que X_t es una combinación de las variables Z_r con $t \leq r$. A continuación se analiza la posibilidad de expresar a un proceso ARMA mediante un filtro causal cuando $\varphi(z) = 0$ para algún z con $|z| < 1$.

Teorema 6.2. Suponga que el proceso $\{X_t\}$ es un proceso ARMA (p, q) el cual satisface

$$\varphi(B)X_t = \theta(B)Z_t,$$

donde los polinomios $\varphi(z)$ y $\theta(z)$ están dados en (5.2) y $\varphi(z) \neq 0$ si $|z| = 1$. Suponga que

$$\varphi(z) = 0 \quad \text{para algún } z \text{ que satisface } |z| < 1.$$

En este caso, existe un polinomio $\tilde{\varphi}(z)$ de grado p, y un ruido blanco $\{\tilde{Z}_t\}$ tal que

$$\tilde{\varphi}(B)X_t = \theta(B)\tilde{Z}_t, \quad \text{y} \quad \tilde{\varphi}(z) \neq 0, \quad \text{para } |z| \leq 1.$$

Demostración. Sena $m_1, m_2, \ldots m_K$ las raíces de $\varphi(z)$ que tienen módulo menor a uno, y denote mediante $M_1, M_2, \ldots, M_{p-K}$ a las raíces cuya norma es mayor a uno. En este caso, $\varphi(z)$ se factoriza como

$$\varphi(z) = \prod_{i=1}^{K}(1 - z/m_i) \prod_{j=1}^{p-K}(1 - z/M_i). \tag{6.2}$$

Defina el polinomio $\tilde{\varphi}(z)$ mediante

$$\tilde{\varphi}(z) = \prod_{i=1}^{K}(1 - m_i z) \prod_{j=1}^{p-K}(1 - z/M_i). \tag{6.3}$$

el cual tiene raíces $M_1, M_2, \ldots, M_{p-K}$ y $1/m_1, 1/m_2, \ldots 1/m_K$, las cuales tienen, invariablemente módulo mayor a uno, esto es,

$$\tilde{\varphi}(z) \neq 0 \quad \text{si } |z| \leq 1.$$

Defina hora la función racional $\alpha(z)$ mediante

$$\alpha(z) = \frac{\prod_{i=1}^{K}(1 - m_i z)}{\prod_{i=1}^{K}(1 - z/m_i)} \tag{6.4}$$

y note que (6.2)–(6.4) implican que

$$\tilde{\varphi}(z) = \alpha(z)\varphi(z),$$

de donde se desprende que

$$\begin{aligned} \tilde{\varphi}(B)X_t &= \alpha(B)\varphi(B)X_t \\ &= \alpha(B)\theta(B)Z_t \\ &= \theta(B)\alpha(B)Z_t, \end{aligned} \tag{6.5}$$

donde se ha utilizado que la composición de filtros es conmutativa; vea el Corolario 2.1 de la Sección 2 del Capítulo 4. Definiendo el proceso $\{\tilde{Z}_t\}$ mediante

$$\tilde{\mathbf{Z}}_t = \alpha(B)Z_t, \tag{6.6}$$

la igualdad (6.5) equivale a

$$\tilde{\varphi}(B)X_t = \theta(B)\tilde{Z}_t,$$

y para concluir la demostración es suficiente con mostrar que $\{\tilde{Z}_t\}$ es un ruido blanco. Con este fin, se utilizará el Corolario 8.1 del Capítulo 5. Note que

$$\begin{aligned} |\alpha(e^{i\xi})| &= \left| \frac{\prod_{k=1}^{K}(1 - m_k e^{i\xi})}{\prod_{k=1}^{K}(1 - e^{i\xi}/m_k)} \right| \\ &= \prod_{k=1}^{K} |m_k| \left| \frac{\prod_{k=1}^{K}(1 - m_k e^{i\xi})}{\prod_{k=1}^{K}(m_k - e^{i\xi})} \right| \\ &= \prod_{k=1}^{K} |m_k| \left| \frac{\prod_{k=1}^{K}(e^{-i\xi} - m_k)}{\prod_{k=1}^{K}(m_k - e^{i\xi})} \right| \end{aligned}$$

Para continuar, observe que las raíces de φ se presentan en pares conjugados, pues $\varphi(z)$ tiene coeficientes reales. Esta observación implica que $\prod_{k=1}^{K}(e^{-i\xi} - m_k) = \prod_{k=1}^{K}(e^{-i\xi} - \overline{m}_k)$, ecuación que al combinarse con la anterior desigualdad desplegada, implica

$$\begin{aligned}
|\alpha(e^{i\xi})| &= \prod_{k=1}^{K} |m_k| \left| \frac{\prod_{k=1}^{K}(e^{-i\xi} - \overline{m}_k)}{\prod_{k=1}^{K}(m_k - e^{i\xi})} \right| \\
&= \prod_{k=1}^{K} |m_k| \left| \frac{\prod_{k=1}^{K} \overline{(m_k - e^{i\xi})}}{\prod_{k=1}^{K}(m_k - e^{i\xi})} \right| \\
&= \prod_{k=1}^{K} |m_k|,
\end{aligned}$$

esto es $|\alpha(e^{i\xi})|$ es constante, y entonces $\{\tilde{Z}_t\}$ es un ruido blanco; vea el Corolario 8.1 del Capítulo 5. □

Capítulo 6

Epílogo

En esta trabajo se ha realizado un análisis riguroso del algortmo de innovaciones, instrumento que se utiliza en la elaboración de pronóasticos, y se han estudiado las relaciones existentes entre el álgebra de los filtros lineales por un lado, y propiedades de funciones analíticas.

La presentación formal inció introduciendo la idea de proceso (débilmente) estacionario en el Capítulo 2, la cual sirvió de preámbulo para abordar, en el Capítulo 3, el estudio del algoritmo de innovaciones. El análisis realizado, mostró que dicho algoritmo puede considerarse como una versión del procedimiento de ortogonalización de Gramm-Schmidt, el cual es clásico dentro de la teoría general de espacios vectoriales dotados con un producto interno. Posteriormente, el estudio de filtro lineales inició en el Capítulo 4, en el cual se demostró que, para una serie absolutamente sumable, el filtro correspondiente está bien definido, y que la convergencia de la serie que lo define es válida en diversos sentidos: con probabilidad uno, en la media de orden uno, y en la media cuadrática. El análisis realizado, mostró de manera explícita la importancia de la teoría general del análisis matemático en el área de series de tiempo.

Por otro lado, en el Capítulo 5 se abordaron los problemas de determinar la composición de dos filtros, y el filtro inverso a uno determinado. estableciendo que estos problemas se reducen, finalmente, a multiplicar series de Laurent, y a obtener la expansión de Laurent del recíproco de una función analítica. De manera especial, se mostró que obtener el inverso de un filtro asociado a un polinomio, se reduce a expander en fracciones parciales a una función racional, a saber, el recíproco del polinomio.

Finalmente, desde un punto de vista formativo, el desarrollo de este trabajo permitió conjuntar ideas estudiadas en diversas áreas, como álgebra lineal, modelos lineales, y teoría de funciones analíticas, para abordar el estudio de los problemas considerados en el análisis de series de tiempo.

Literatura Citada

Alemán–Valerio, J. V. (1992), Aspectos Teóricos de Series de Tiempo Estacionarias, *Tesis de Maestría en Ciencias en Estadística Experimental*, Universidad Autónoma Agraria Antonio Narro, Saltillo, Coahuila, México.

Anderson, T. W. (1981), *The Statistical Analysis of Time Series*, *Wiley*, New York.

Alfhors, V. (1970), Complex Analysisi, *McGraw–Hill*, New York.

Apostol, T. M. (1981) Mathematical Analysis, *Addison–Wesley*, New York.

Ash, R. B. (1975), *Real Analysis and Probability*, *Academic Press*, New York.

Box, G.E.P., y G.M. Jenkins (1970), *Time Series Analysis: Forecasting and Control*, *Holden–Day*, San Francisco.

Brockewll, P.J., y Davis, R.A (1991), *Time Series: Theory and Methods*, *Springer–Verlag*, New York.

Brockewll, P.J., y Davis, R.A (1996), *Introduction to Time Series and Forecasting*, *Springer–Verlag*, New York.

Cavazos–Cadena, R. (1994), The asymptotic distribution of sample autocorrelations for a class of linear filters, *Journal of Multivariate Analysis*, **48**, 249–274.

Churchill, R. V. (1976), Introduction to Complex Analysisi, *McGraw–Hill*, New York.

Dudewicz, E., y N. Mishra, (1988), *Modern Mathematical Statistics*, *Wiley*, New York.

Graybill, F. A. (1985), *Theory and Application of the Linear Model, 1st. Edition*, *Wadsworth*, Belmont, CA

Grossman, S. I. (1983), *Algebra Lineal*, *Grupo Editorial Iberoamérica*, Tercera Edición, México D. F.

Harville, D. A. (1997), *Matrix Algebra from a Statistician Perspective*, *Springer–*

Verlag, New York

Hoffman, K. y R. Kunze. (1973) *Algebra Lineal*, *Prentice Hall Hispanoamericana*, México D. F.

Lehmman, E. L. (1991), *Testing Statistical Hypothesis*, *Wiley*, N.Y.

Mood, A. M., F. A. Graybill y D. C. Boes (1985), *Introduction to the Theory of Statistics*, *McGraw–Hill*, New York.

Nevannlina, R. y V. Pattero (1975), *Introduction to the Theory of Complex Analysis*, *McGraw–Hill*, N.Y.

Polendo–Luis, S. (1999), Bondad de Ajuste y Cálculo Recursivo de Pronósticos el Análisis de Series de Tiempo, *Tesis de Maestría en Ciencias en Estadística Experimental*, Universidad Autónoma Agraria Antonio Narro, Saltillo Coahuila, México.

Rudin, W. (1976), Real and Complex Analysis, *McGraw–Hill*, N.Y.

Searle, S. R. (1982), *Matrix Algebra Useful for Statistics*, *McGraw–Hill*, New York.

Wei, W. W. S. (1990), *Time Series Analysis: Univariate and Multivariate Analysis*, *Addison Wesley*, Reading, Massachusetts.

APÉNDICE

Construcción de Inversas Condicionales

Introducción

El objetivo de esta sección es introducir un procedimiento basado en operaciones elementales de fila para construir una inversa condicional de una matriz dada. Sea $\mathbf{A}$ una matriz de orden $m \times n$. Una matriz $\mathbf{M}$ de orden $n \times m$ es es una inversa condicional de $\mathbf{A}$ si

$$\mathbf{AMA} = \mathbf{A}. \tag{1}$$

Esta definición es un poco lacónica, y con el propósito de obtener un mejor entendimiento de la noción de inversa condicional, se establecerá la siguiente caracterización.

Lema 1.1. La matriz $\mathbf{M}$ es una inversa condicional de $\mathbf{A}$ si y sólo si para cada vector $\mathbf{g} \in \mathcal{I}(\mathbf{A})$ el sistema $\mathbf{Ax} = \mathbf{g}$ tiene como una de sus soluciones a $\mathbf{x} = \mathbf{Mg}$.

Demostración.

• Suponga que $\mathbf{M}$ es una inversa condicional de $\mathbf{A}$ y sea $\mathbf{g} \in \mathcal{I}(\mathbf{A})$ un vector arbitrario. En este caso $\mathbf{g} = \mathbf{Ay}$ para algún vector $\mathbf{y}$, y como $\mathbf{M}$ satisface (1), se tiene que $\mathbf{AMAy} = \mathbf{Ay}$, y combinando esta igualdad con $\mathbf{g} = \mathbf{Ay}$ se obtiene que $\mathbf{AMg} = \mathbf{g}$, i.e., $\mathbf{x} = \mathbf{Mg}$ es una solución de $\mathbf{Ax} = \mathbf{g}$.

• Suponga ahora que para cada $\mathbf{g} \in \mathcal{I}(\mathbf{A})$ el sistema

$$\mathbf{Ax} = \mathbf{g} \quad \text{tiene como una de sus soluciones a} \quad \mathbf{x} = \mathbf{Mg}. \tag{2}$$

En este caso se verificará que $\mathbf{M}$ es una inversa condicional de $\mathbf{A}$. Para ver esto note que (2) equivale a

$$\mathbf{AMg} = \mathbf{g} \quad \text{para todo} \quad \mathbf{g} \in \mathcal{I}(\mathbf{A}),$$

y debido a que un elemento general de $\mathcal{I}(\mathbf{A})$ se expresa como $\mathbf{g} = \mathbf{Ay}$, la ecuación anterior es equivalente a

$$\mathbf{AMAy} = \mathbf{Ay} \quad \textbf{para todo} \quad \mathbf{y} \in \mathbb{R}^n,$$

de donde se desprende que $\mathbf{AMA} = \mathbf{A}$, esto es, que $\mathbf{M}$ es una inversa condicional de $\mathbf{A}$. □

El Lema 1.1 muestra expícitamente la relación entre la idea de inversa condicional y el proceso de solución de un sistema de ecuaciones lineales. Debido a que las operaciones elementales de fila fueron diseñadas para resolver ecuaciones lineales, es de esperar que esas mismas operaciones sean de utilidad para determinar una inversa condicional.

Inversa Condicional y Operaciones Elementales de Fila

Considere la matriz $\mathbf{A}$ de orden $m \times n$ de la sección anterior y forme la matriz

$$\mathbf{V} = [\,\mathbf{A}\,|\,\mathbf{I}\,],$$

A continuación realice operaciones elementales de fila sobre la matriz $\mathbf{V}$ hasta llevar al bloque $\mathbf{A}$ a su forma escalón reducida por filas $\mathbf{R}$, de manera que

$$\mathbf{V} = [\,\mathbf{A}\,|\,\mathbf{I}\,] \sim [\,\mathbf{R}\,|\,\mathbf{P}\,].$$

La matriz $\mathbf{P}$ es inversible, pues es un producto de matrices elementales y, más aún,

$$\mathbf{R} = \mathbf{PA},$$

de manera que

$$\mathbf{Ax} = \mathbf{g} \iff \mathbf{Rx} = \mathbf{Pg}. \tag{2}$$

Sea r el rango de $\mathbf{A}$, y observe que en este caso las filas $r+1, r+2, \ldots, m$ de la matriz $\mathbf{R}$ son nulas. Sean $k_1 < k_2 < \ldots < k_r$ las posiciones de los unos líderes de las filas no nulas de $\mathbf{R}$ y sea $\mathcal{F}:=\{1,2,\ldots,n\}\setminus\{k_1,k_2,\ldots,k_r\}$. Además, denote mediante $\mathbf{p}_i'$ a la i–ésima fila de $\mathbf{P}$. Si el sistema $\mathbf{Ax}=\mathbf{g}$ es consistente, i.e., si $\mathbf{g}\in\mathcal{I}(\mathbf{A})$, entonces $\mathbf{Ax}=\mathbf{g}$ tiene las mismas soluciones que

$$\mathbf{x}_{k_i}+\sum_{j\in\mathcal{F}} R_{i\,j}x_j=\mathbf{p}_i'\mathbf{g},\quad i=1,2,\ldots,r; \tag{3}$$

las soluciones de este sistema se obtienen asignando valores arbitrarios a x_j para cada $j\in\mathcal{F}$, y despejando x_{k_i} $i=1,2,\ldots,r$ en (3). La posibilidad más simple es asignar el valor cero a cada x_j con $j\in\mathcal{F}$, y entonces obtener directamente los valore de x_{k_i}. La solución que se obtiene es

$$x_j=0,\quad j\in\mathcal{F}\quad \text{y}\quad x_{k_i}=\mathbf{p}_i'\mathbf{g},\quad i=1,2,\ldots,r. \tag{4}$$

Esta discusión puede resumirse como sigue: Si $\mathbf{g}\in\mathcal{I}(\mathbf{A})$, entonces el vector $\mathbf{x}$ cuyas componentes están dadas en (4) es solución del sistema $\mathbf{Ax}=\mathbf{g}$. Ahora se utilizará este resultado para construir una inversa condicional de $\mathbf{A}$. Para esto, usando el Lema 1.1, sólo es necesario escribir el vector $\mathbf{x}$ en (4) como $\mathbf{x}=\mathbf{Mg}$, y un momento de reflexión muestra que la matriz $\mathbf{M}$ está determinada como sigue:

(a) Para cada $j\in\mathcal{F}$, defina la j–ésima fila de $\mathbf{M}$ como el vector nulo; de esta manera, la j–ésima componente de de $\mathbf{Mg}$ se anula, de acuerdo a (4).

(b) Para $i=1,2,\ldots,r$, defina la fila k_i de $\mathbf{M}$ como $\mathbf{p}_i'$, de modo que la componente k_i de $\mathbf{Mg}$ es $\mathbf{p}_i'\mathbf{g}$, lo cual también está de acuerdo con (4).

Luego, para todo $\mathbf{g}\in\mathcal{I}(\mathbf{A})$, el vector $\mathbf{x}=\mathbf{Mg}$ es solución del sistema $\mathbf{Ax}=\mathbf{g}$, y entonces la matriz $\mathbf{M}$ construida de acuerdo a (a) y (b) es una inversa condicional de $\mathbf{A}$. Para propósitos de referencia futura, este resultado se establece ahora formalmente.

Teorema 1.2. Sea $\mathbf{A}$ una matriz de orden $m\times n$. Entonces, la matriz $\mathbf{M}$ construida de acuerdo al siguiente procedimiento es una inversa condicional de $\mathbf{A}$.

1. Forme la matriz $\mathbf{V} = [\,\mathbf{A}\,|\,\mathbf{I}\,]$ y realice operaciones elementales de fila sobre $\mathbf{V}$ hasta llevar a la matriz $\mathbf{A}$ a la forma escalón reducida por filas $\mathbf{R}$; esto produce

$$[\,\mathbf{A}\,|\,\mathbf{I}\,] \sim [\,\mathbf{R}\,|\,\mathbf{P}\,].$$

2. Sean $k_1, k_2, \ldots, k_r$ las posiciones de los unos líderes en las filas no nulas de $\mathbf{R}$, donde r = rango de $\mathbf{A}$. Defina la fila k_i de $\mathbf{M}$ como la fila i de $\mathbf{P}$, $i = 1, 2, \ldots, r$, y ponga todas las otras $n - r$ filas de $\mathbf{M}$ igual a la fila nula.

A continuación se ilustra la aplicación de este procedimiento.

Ejemplo 1.2. Considere la matriz

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 0 & 1 \\ -1 & -2 & 1 & 1 \\ 1 & 2 & 1 & 3 \end{bmatrix}$$

Para encontrar una inversa condicional de $\mathbf{A}$ el procedimiento del Teorema 1.2 se aplica como sigue:

1. Se forma la matriz (aumentada) $\mathbf{V} = [\,\mathbf{A}\,|\,\mathbf{I}\,]$:

$$\mathbf{V} = \begin{bmatrix} 1 & 2 & 0 & 1 & 1 & 0 & 0 \\ -1 & -2 & 1 & 1 & 0 & 1 & 0 \\ 1 & 2 & 1 & 3 & 0 & 0 & 1 \end{bmatrix}$$

y se realizan operaciones elementales de fila hasta reducir el bolque $\mathbf{A}$ a la forma escalón reducida por filas.

$$\begin{bmatrix} 1 & 2 & 0 & 1 & 1 & 0 & 0 \\ -1 & -2 & 1 & 1 & 0 & 1 & 0 \\ 1 & 2 & 1 & 3 & 0 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 & 1 & 1 & 0 \\ 0 & 0 & 1 & 2 & -1 & 0 & 1 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 2 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -2 & -1 & 1 \end{bmatrix} = [\,\mathbf{R}\,|\,\mathbf{P}\,]$$

2. La matriz escalón reducida por filas **R** equivalente a **A** tiene rango $r = 2$ y los unos líderes en las filas no nulas ocupan las posiciones

$$k_1 = 1 \quad \text{y} \quad k_2 = 3.$$

La matriz **P** está dada por

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -2 & -1 & 1 \end{bmatrix}$$

La inversa condicional **M** se construye como sigue: Primero observe que **M** es de orden 4×3:

$$\mathbf{M} = \begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \end{bmatrix}.$$

Ahora, la fila $k_1 = 1$ de **M** es la fila 1 de **P**:

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 \\ * & * & * \\ * & * & * \\ * & * & * \end{bmatrix}.$$

Además, la fila $k_2 = 3$ de **M** es la fila 2 de **P**:

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 \\ * & * & * \\ 1 & 1 & 0 \\ * & * & * \end{bmatrix},$$

mientras que las restantes filas de **M** son todas nulas. Por lo tanto **M** está dada por

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

de acuerdo al Teorema 1.2, esta matriz **M** es una inversa condicional de **A**. Para verificar esta afirmación observe que

$$
\begin{aligned}
\mathbf{AMA} &= \begin{bmatrix} 1 & 2 & 0 & 1 \\ -1 & -2 & 1 & 1 \\ 1 & 2 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 & 1 \\ -1 & -2 & 1 & 1 \\ 1 & 2 & 1 & 3 \end{bmatrix} \\
&= \left\{ \begin{bmatrix} 1 & 2 & 0 & 1 \\ -1 & -2 & 1 & 1 \\ 1 & 2 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right\} \begin{bmatrix} 1 & 2 & 0 & 1 \\ -1 & -2 & 1 & 1 \\ 1 & 2 & 1 & 3 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 & 1 \\ -1 & -2 & 1 & 1 \\ 1 & 2 & 1 & 3 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 2 & 0 & 1 \\ -1 & -2 & 1 & 1 \\ 1 & 2 & 1 & 3 \end{bmatrix} = \mathbf{A}
\end{aligned}
$$

de manera que, efectivamente, $\mathbf{M}$ es una inversa condicional de $\mathbf{A}$.