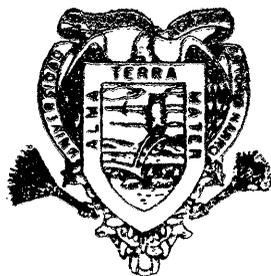


BONDAD DE AJUSTE Y CALCULO RECURSIVO
DE PRONOSTICOS EN EL ANALISIS DE
SERIES DE TIEMPO

SILVIA POLENDO LUIS

T E S I S

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

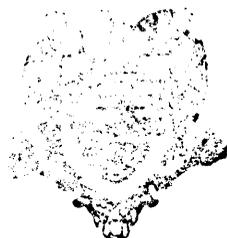


**Universidad Autónoma Agraria
Antonio Narro**

PROGRAMA DE GRADUADOS

Buenavista, Saltillo, Coah.

JUNIO DE 1999



**BIBLIOTECA
EGIDIO G. REBONATO
BANCO DE TESIS
U.A.A.A.N.**

11037

UNIVERSIDAD AUTÓNOMA AGRARIA

ANTONIO NARRO

SUBDIRECCIÓN DE POSTGRADO

Bondad de Ajuste y Cálculo Recursivo de Pronósticos
en el Análisis de Series de Tiempo

TESIS

por

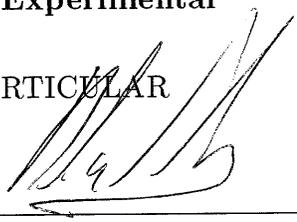
SILVIA POLENDO LUIS

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

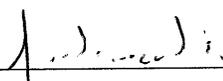
COMITÉ PARTICULAR

Asesor Principal:



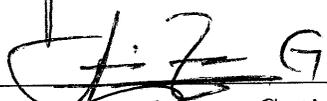
Dr. Rolando Cavazos Cadena

Asesor:

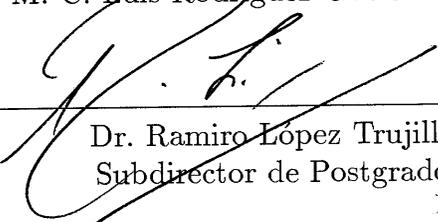


Dr. Heriberto Díaz Solís

Asesor:



M. C. Luis Rodríguez Gutiérrez



Dr. Ramiro López Trujillo
Subdirector de Postgrado

Buenavista, Saltillo, Coahuila. Junio de 1999

AGRADECIMIENTOS

Al Dr. Rolando Cavazos Cadena,

Por aceptar ser mi asesor, por su paciencia y por concederme el privilegio de su amistad.

A José Luis,

Por ser mi compañero y mi mejor amigo.

A mis padres,

Por su apoyo incondicional y su cariño.

A mis compañeros del Instituto Tecnológico de Saltillo,

Por brindarme la oportunidad de compartir el mismo espacio.

COMPENDIO

Bondad de Ajuste y Cálculo Recursivo de Pronósticos
en el Análisis de Series de Tiempo

por

SILVIA POLENDO LUIS

TESIS

MAESTRÍA EN CIENCIAS
en Estadística Experimental

UNIVERSIDAD AUTÓNOMA AGRARIA
ANTONIO NARRO

Buenavista, Saltillo, Coahuila, México. Junio de 1999

Rolando Cavazos Cadena — Asesor

Palabras Clave: Proceso estacionarios, Función de Autocovarianza, Procesos ARMA, Algoritmo de Durbin-Levinson, Pruebas no paramétricas, Estadístico de Máximos y Mínimos Locales, Distribución Asintótica.

Este trabajo trata sobre los problemas de elaboración de pronósticos y de bondad de ajuste en el estudio de series de tiempo. El método recursivo de

Durbin–Levinson es analizado desde una óptica geométrica, proporcionando una demostración rigurosa de su validez como generador de pronósticos . Además, se determina la distribución límite del estadístico de máximos y mínimos locales en una sucesión de datos, resultado que permite formular una prueba para la hipótesis de que los valores considerados provienen de variables aleatorias independientes con una misma distribución. Cuando este método se aplica a las estimaciones de los errores obtenidas a partir de un modelo para una serie de datos, se genera una prueba para decidir si el ajuste obtenido es apropiado.

ABSTRACT

Goodness of Fit and Recursive Forecasting
in Time Series Analysis

by

SILVIA POLENDO LUIS

THESIS

MASTER of SCIENCE
Experimental Statistics

UNIVERSIDAD AUTÓNOMA AGRARIA
ANTONIO NARRO

Buenavista, Saltillo, Coahuila, México. June, 1999

Rolando Cavazos Cadena — Advisor

Key Words: Stationary Processes, Autocovariance Function, ARMA Processes, Durbin–Levinson Algorithm, Nonparametric tests, Statistic of local maxima and minima, Asymptotic Distribution.

This work concerns (weakly) stationary stochastic processes endowed with discrete time parameter. A recursive forecasting algorithm, referred to as the

Durbin–Levinson procedure, is analyzed from a gemoetric perspective and it is formally verified. On the other hand, the statistic T of local maxima and minima in a given data set is studied, and its menan and variance are determined. These results are used to establish the asymptotic distribution of T , and a nonparamentric test is generated to assess the hypothesis that the data stem from independent and identically distributed random variables. When a model is proposed for a given observed series, and the above procedure is applied to the estimated noise sequence, a test for the goodness of fit of the model is obtained.

CONTENIDO

	Página Número
Panorámica General	1
Introducción	1
Objetivos	2
La Organización	4
Procesos Estacionarios	6
Introducción	6
La Idea Fundamental	8
Perspetiva de Segundo Orden	12
Estimación de la Media y la Función de Autocovarianza	17
Filtros Lineales y el Operador de Retardo	25
Procesos ARMA	28
La Descomposición Clásica	33
Los Problemas de Pronóstico y Bondad de Ajuste	36
Conclusión	38
Una Prueba de Bondad de Ajuste	39
Introducción	39
El Estadístico T_n	41
El Valor Esperado y la Varianza de T_n	43
Instrumentos Auxiliares	45
Demostración del Teorema 3.1	52
La Distribución Asintótica de T_n	54
Resultados Auxiliares	56
Demostración del Teorema 6.1	64
El Procedimiento de Prueba	67

Conclusión	68
Cálculo Recursivo de Pronósticos	70
Introducción	70
La Mejor Aproximación	71
El Error de Aproximación	78
Caracterización de los Coeficientes	84
Interpretación Estadística	89
El Algoritmo de Durbin–Levinson	91
Resultados Auxiliares	94
Demostración del Teorema 6.1	98
Un Ejemplo	102
Conclusión	105
Epílogo	107
Literatura Citada	109

Capítulo 1

Panorámica General

El propósito de este capítulo es describir, en términos generales, el contenido de esta tesis, plantear los objetivos que se pretenden alcanzar, y la manera en que se ha organizado la presentación del material subsecuente.

Introducción

Este trabajo gira alrededor de la idea de *serie de tiempo*, noción que se refiere a un conjunto de variables aleatorias cuyos valores son observados secuencialmente. En la actualidad, son innumerables los datos que, como producto de la actividad humana, se generan de manera cotidiana. Por ejemplo, cada quincena se hace público el índice de precios al consumidor o el nivel de inflación en nuestro país, información que conforma una sucesión $x_1, x_2, \dots, x_n, \dots$ que puede pensarse como una realización particular de una serie de variables aleatorias. Los datos que se generan secuencialmente son abundantes, y no se limitan a disciplina o tipo particular alguno. Por ejemplo, los productores de cultivos como manzana o durazno, están naturalmente interesados en el total de ‘horas-frío’ en la región donde se ubican sus huertos, e invariablemente, los servicios meteorológicos registran los datos necesarios y reportan sus cifras; desde luego, el nivel de precipitación pluvial, o el nivel de humedad son sólo otras de las cantidades que son de interés en las actividades agrícolas, y que se registran una y otra vez, digamos semanalmente, generando una serie de datos para los cuales es importante entender su estructura (el problema de modelado), así como proporcionar estimaciones sobre el valor que asumirán en el futuro (el problema de pronóstico).

El estudio de una serie de tiempo es un tema de enorme interés por su inmediata relación con situaciones concretas, y requiere del empleo de herramientas cuya complejidad está en concordancia con la trascendencia e importancia de los problemas de modelado y de pronóstico que son fundamentales en esta disciplina; vea, por ejemplo, Anderson (1976), Anderson (1981), Wei (1990), Brockwell y Davis (1991, 1996), o Box y Jenkins (1970). En términos generales, construir un modelo para una serie de datos interpretados como una serie de tiempo, consiste en plantear una ecuación que refleje las relaciones entre los datos que se observen sucesivamente; dicha ecuación debe ser lo más simple que sea posible y, particularmente, debe contener un número ‘pequeño’ de parámetros desconocidos que puedan estimarse a partir de los datos. Una vez que se ha planteado un modelo, es necesario juzgar la bondad de su ajuste a los datos observados y, al juzgarlo adecuado, se puede emprender la tarea de elaborar un pronóstico para las cifras que se observarán en el futuro, temas que, como se describe a continuación conforman la parte medular de este trabajo.

Objetivos

Suponga que $X_1, X_2, \dots, X_n, \dots$ es una sucesión de variables aleatorias, cuyo subíndice representa el tiempo, de manera que el valor x_t asumido por la variable X_t se observa en el instante t . Como ya se ha mencionado, construir un modelo para los datos observados $x_1, x_2, \dots, x_n, \dots$ es una etapa fundamental en el proceso de entender la estructura de las cifras observadas. Una posibilidad que puede presentarse es que las variables X_1, X_2, \dots, X_n que dan origen a los datos sean *independientes e idénticamente distribuidas (i.i.d.)*, de manera que el valor detectado en un momento determinado no aporta información sobre las cifras que fueron recabadas anteriormente, o sobre aquéllas que se harán disponibles en el futuro. En este caso, el analista no puede más que declarar que las variables son *i.i.d.*, y el ‘mejor pronóstico’ que puede elaborar sobre una observación futura X_s ,

es el valor esperado de $E[X_s]$, al menos cuando el error de pronóstico se mide a través del error cuadrático medio (Lehman, 1991, Dudewicz y Mishra, 1988, Mood *et. al.*, 1985). De esta forma, antes de emprender la tarea de construir un modelo complejo para una serie de tiempo, debe probarse la hipótesis de que las variables aleatorias que dan origen a los datos observados son independientes con una distribución común. Este problema también es relevante después de que un modelo ha sido propuesto para las cifras que se han registrado. Esencialmente, un modelo estocástico para los datos involucra a una serie de variables aleatorias ‘de ruido’ que provocan la aleatoriedad de las observaciones, y una manera de ver que el modelo propuesto describe adecuadamente la evolución de las mismas, es verificar que los ruidos estimados a partir del modelo forman una sucesión *i.i.d.* (Brockwell y Davis, 1991, 1996). Dentro de los métodos disponibles para este problema se encuentra el bien conocido enfoque basado en la distribución χ^2 (Serfling, 1980, Randles y Wolfe, 1983). En este trabajo, se analizará otra prueba de naturaleza no paramétrica, la cual se basa en el análisis del número total de máximos y mínimos locales observados en una sucesión Y_1, Y_2, \dots, Y_n , el cual se denota mediante T_n , la cual se describe brevemente en Brockwell y Davis (1996). Para generar una prueba de bondad de ajuste, se plantean los siguientes objetivos relacionados con la distribución de este estadístico.

- *Determinar* el valor esperado y la varianza del estadístico T_n .
- *Encontrar* la distribución asintótica de estadístico de máximos y mínimos locales.

Utilizando el resultado sobre la distribución límite de T_n , se alcanzará el siguiente propósito fundamental:

- *Formular* una prueba de significancia para la hipótesis de que las variables aleatorias de una sucesión son *i.i.d.*.

Por otro lado, en este trabajo también se aborda el cómputo de pronósticos mediante procedimientos recursivos. De forma específica, se analizará el llamado

algoritmo de Durbin–Levinson (Anderson, 1976), el cual es un procedimiento que se formula desde una perspectiva algebraica. La idea detrás del análisis de este método, es estudiarlo desde una óptica geométrica.

- *Interpretar* el algoritmo de Durbin–Levinson en términos de proyecciones ortogonales estableciendo de forma rigurosa la validez del método.

Finalmente, un propósito formativo detrás de esta incursión en el estudio de series de tiempo, es tratar de aglutinar los conocimientos sobre diversas áreas, como álgebra lineal y estadística no paramétrica, para lograr un mejor entendimiento de las técnicas aplicadas en el campo de procesos estocásticos estacionarios, disciplina que es tan compleja como apasionante.

La Organización

Para alcanzar los objetivos propuestos, el trabajo subsecuente ha organizado de la siguiente manera: La exposición técnica inicia en el Capítulo 2, en el cual se presenta 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, es posible realizar una transformación que conduzca a resultados importantes a partir del análisis de una serie estacionaria. La presentación de este capítulo también incluye una discusión, general y breve, del problema de construcción de modelos para una serie de datos, y se presenta la familia de procesos autorregresivos y de promedios móviles.

En el Capítulo 3 se considera el problema de bondad de ajuste de una serie de datos al supuesto de que las variables que les dan origen son *i.i.d.*. La prueba de significancia que se presenta es de naturaleza no paramétrica y, como ya se ha mencionado, se basa en el conteo del número total de máximos y mínimos locales que se presentan en la sucesión de datos observados hasta el tiempo n ; en este capítulo se alcanzan los primeros tres objetivos planteados. Posteriormente, en el

Capítulo 4 se estudia el problema de construir pronósticos para una observación futura desde un punto de vista geométrico, estableciendo la validez del algoritmo de *Durbin–Levinson* por medio del análisis de proyecciones ortogonales. Finalmente, la presentación culmina en el Capítulo 5 con conclusiones y algunos comentarios breves.

Capítulo 2

Procesos Estacionarios

El objetivo de este capítulo es introducir la idea de serie estacionaria desde la perspectiva de segundo orden, esto es, un proceso se considera estacionario cuando la estructura de sus momentos de orden menor o igual a dos no se altera al transcurrir el tiempo. Después de considerar los problemas de modelado y de pronóstico para una serie con la propiedad de estacionaridad, se describe la forma en que los resultados obtenidos pueden aplicarse en un contexto real, en donde usualmente las características de las variables que conforman el proceso se alteran conforme sus valores son observados. La presentación discute la importancia de los problemas de bondad de ajuste y de elaboración recursiva de pronósticos, los cuales serán abordados más adelante.

Introducción

El objeto central 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 representa ‘el tiempo’, es común referirse al proceso $\{X_t\}$ como una serie cronológica, o *serie de tiempo*, terminología que se adoptará en lo sucesivo.

Es claro que la noción de serie de tiempo es sumamente general. De hecho, cada vez que se observan datos secuencialmente, las cifras recabadas $\{x_t\}$ pueden

pensarse como los valores asumidos por variables aleatorias X_t . Por supuesto, existen innumerables situaciones prácticas en que este proceso de acopio secuencial de datos se presenta de modo natural, y el propósito de este capítulo es presentar los conceptos básicos que se relacionan con el estudio de dos problemas fundamentales en el estudio de una serie de tiempo, a saber, la construcción de un modelo para los datos, y la elaboración de un pronóstico para el valor que asumirá un dato que se observará en el futuro.

La exposición gira alrededor de la idea básica de proceso (serie) estacionaria y ha sido organizada como se describe a continuación: En la Sección 2 se discute la noción de *proceso estacionario*. La presentación tiene un carácter intuitivo, y se introduce el concepto de proceso estrictamente estacionario, idea que es sumamente demandante (restrictiva), y que conduce de forma natural a formular la idea de estacionaridad débil en la Sección 3, en donde se adopta la perspectiva de segundo orden en el análisis de una serie de tiempo; esencialmente, esto significa que las técnicas de estudio de un proceso se elaborarán utilizando la estructura de covarianzas entre las variables que lo conforman, así como sus valores esperados. En la Sección 4 se abordan los problemas de estimación para el valor esperado de las variables que conforman un proceso estacionario, y se discute el correspondiente problema para la función de autocovarianza del proceso. En la Sección 5 se presenta la noción de filtro lineal, idea que constituye el preámbulo para la presentación de los procesos autorregresivos y de promedios móviles en la Sección 6. Posteriormente, en la Sección 7 se aborda el problema de aplicar los resultados sobre series estacionarias en un contexto real, donde los datos observados no poseen, usualmente, la característica de estacionaridad, mientras que en la Sección 8 se discute la importancia de los problemas de bondad de ajuste y de cálculo recursivo de pronósticos que forman el núcleo de este trabajo. Finalmente, el capítulo concluye en la Sección 9 con algunos comentarios breves.

La Idea Fundamental

El propósito de esta sección es discutir la noción de *serie estacionaria*, idea que desempeña un papel central en el análisis de series de tiempo. En términos generales, un proceso estocástico $\{X_t\}$ es estacionario si sus características esenciales no varían conforme transcurre el tiempo. De manera un poco más precisa, la serie es estacionaria si ‘las propiedades importantes’ de un vector $(X_{t_1}, X_{t_2}, \dots, X_{t_k})$ son siempre las mismas que las correspondientes a $(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_k+h})$.

Discusión. Un propósito fundamental del análisis de series de tiempo es utilizar los datos observados hasta el tiempo n , esto es, x_1, x_2, \dots, x_n , para formular inferencias sobre las propiedades de las variables X_{n+1}, X_{n+2}, \dots cuyos valores se observarán en el futuro. Sin embargo, los datos x_1, x_2, \dots, x_n arrojan información sobre el vector (X_1, X_2, \dots, X_n) del cual son una realización, y una manera de ‘extrapolar’ las propiedades de éste último hacia variables que serán observadas posteriormente, es *suponer* que dichas características no se alterarán al transcurrir el tiempo. Por ejemplo, suponga que las variables X_1, X_2, \dots, X_n tienen una media común μ , la cual se desconoce. En este caso, \bar{x}_n —la media muestral de los datos observados hasta el tiempo n —es un estimador ‘razonable’ de μ pero, en general, no se puede pensar que \bar{x}_n también es una buena estimación de la media de una observación futura, *a menos que se esté dispuesto a suponer que el valor esperado de X_t permanece inalterado al transcurrir el tiempo*, condición que puede expresarse mediante la igualdad

$$E[X_t] = E[X_{t+h}] \quad \text{para todo } t \text{ y } h. \quad (2.1)$$

Para considerar otro caso similar, suponga que se trata de investigar la covarianza existente entre dos variables cuya observación se realiza en momentos separados por h unidades en el tiempo, como las parejas $(X_1, X_{1+h}), (X_2, X_{2+h}), (X_3, X_{3+h}), \dots$, y que se asume que, para variables seleccionadas entre X_1, X_2, \dots, X_n , la covarianza entre pares cuyo índice difiere en h unidades es una constante τ . En este

caso, los datos X_1, X_2, \dots, X_n pueden utilizarse para obtener un estimador de τ , denotado por $\hat{\tau}$, pero si se desea utilizar $\hat{\tau}$ como estimador ‘razonable’ de la covarianza entre X_{n+1} y X_{n+1+h} , es necesario suponer que la covarianza entre variables recabadas en instantes que difieren en h unidades de tiempo no se altera conforme las observaciones transcurren, esto es,

$$\text{Cov}[X_1, X_{1+h}] = \text{Cov}[X_{n+1}, X_{n+1+h}], \quad \text{para todo } n \text{ y } h. \quad (2.2)$$

Esta discusión puede resumirse de la siguiente forma: el supuesto de que los rasgos esenciales de una serie no se alteran al desplazar el índice del tiempo, permite utilizar los datos observados hasta el momento n para formular inferencias sobre características referentes a las variables que se observarán en el futuro. Sin embargo, este planteamiento conduce, de forma inmediata y natural, a un cuestionamiento: ¿Qué sucede con las series cuyas características de interés sí se alteran al transcurrir el tiempo? Por ejemplo, sea $\{X_t\}$ la serie de consumos semanales de agua en una comunidad. Si s denota una semana correspondiente al verano de un año dado, mientras que t es una semana del invierno correspondiente, debe esperarse que $E[X_s] > E[X_t]$, mientras que

$$\frac{1}{56} E \left[\sum_{i=1}^{56} X_i \right] < \frac{1}{56} E \left[\sum_{i=1}^{56} X_{i+560} \right];$$

en efecto, el lado izquierdo es el consumo promedio semanal de agua en el primer año de observación, mientras que el lado derecho es la cantidad correspondiente al undécimo año, y simplemente por el incremento poblacional, es razonable suponer que en éste último, el consumo de agua será mayor que en el primer año de observaciones. Este argumento muestra que el valor esperado de X_t si sufre alteraciones al avanzar el tiempo. Luego, aunque el supuesto de que los rasgos esenciales de una serie permanecen inalterados permite extrapolar conclusiones obtenidas de los datos disponibles hacia las observaciones que se recabarán posteriormente, dicha hipótesis excluye series interesantes e importantes, especialmente aquellas

que presentan componentes estacionales (periódicas) y tendencias ascendentes o descendentes. De hecho, es difícil pensar en una serie interesante que no presente, por lo menos, alguna de estas características y, a primera vista, parece ser que al asumir el carácter estacionario de un proceso, se excluyen de la discusión a la inmensa mayoría de las series que surgen en la práctica. Sin embargo, es posible, bajo supuestos adecuados y aplicando una transformación apropiada, reducir el estudio de una serie con tendencia o con componentes estacionales al análisis de un proceso estacionario (Box y Jenkins (1970)). De esta manera, estudiar series estacionarias es una empresa que tiene repercusiones importantes en el proceso de entender y analizar las series que se originan en las aplicaciones.

Desde luego, ‘la característica’ más importante de un vector aleatorio es su distribución, pues a través de ella se captura cualquier rasgo probabilístico que pueda ser de interés, observación que conduce a formular el siguiente concepto.

Definición 2.1. Una serie $\{X_t\}$ se denomina *estrictamente estacionaria* si, para cada entero positivo k , para todos los subíndices t_1, t_2, \dots, t_k , y para todo entero h , se tiene que

$$(X_{t_1}, X_{t_2}, \dots, X_{t_k}) \stackrel{d}{=} (X_{t_1+h}, X_{t_2+h}, \dots, X_{t_k+h}), \quad (2.3)$$

esto es, los vectores $(X_{t_1}, X_{t_2}, \dots, X_{t_k})$ y $(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_k+h})$ son siempre idénticamente distribuidos.

Esta noción de serie estacionaria es sumamente demandante. Para entender las implicaciones detrás del requerimiento (2.3), seleccione $k = 1$ y $t_1 = 1$. En este caso (2.3) se convierte en

$$X_1 \stackrel{d}{=} X_{1+h} \quad \text{para todo } h,$$

de tal manera que todas las variables X_t que conforman el proceso tiene la misma distribución. En particular, si el valor esperado de la función de distribución común está bien definido, entonces

$$E[X_t] \quad \text{no depende de } t.$$

Sin embargo, no sólo los valores esperados de las variables X_t coinciden bajo la estacionaridad estricta, sino que *absolutamente* todas sus características probabilísticas son las mismas, como su mediana o sus cuartiles. Seleccionando ahora $k = 2$, $t_1 = s$ y $t_2 = t$, se tiene que

$$(X_s, X_t) \stackrel{d}{=} (X_{s+h}, X_{t+h}) \quad \text{para cualquier } h,$$

y entonces la distribución de (X_s, X_t) permanece inalterada al ‘trasladar’ los índices s y t por la misma cantidad. Esto significa que cualquier característica que se exprese en términos de la función de distribución de (X_s, X_t) , no cambiará al desplazar el índice del tiempo. Por ejemplo, $P[X_s > X_t - 11] = P[X_{s+h} > X_{t+h} - 11]$, $\text{med}\{X_s - X_t\} = \text{med}\{X_{s+h} - X_{t+h}\}$ (Randles y Wolfe (1981)); considerando rasgos más sofisticados de una distribución conjunta, las distribuciones condicionales no se alterarán al trasladar el tiempo, i.e., siempre se tiene que $P[X_s \in A | X_t] = P[X_{s+h} \in A | X_{t+h}]$ para todo subconjunto $A \in \mathbb{R}$, y, especialmente, si las variables X_t tienen segundo momento finito, entonces, para todos los índices s y t ,

$$\text{Cov}[X_s, X_t] = \text{Cov}[X_{s+h}, X_{t+h}] \quad \text{para toda } h.$$

Estos comentarios muestran que una serie estacionaria en el sentido estricto introducido en la Definición 2.2 es, necesariamente, muy especial. En la teoría desarrollada para el análisis de series de tiempo, es frecuente encontrar que los métodos de estimación y ajuste de modelos no dependen de que *absolutamente todas* las características de un vector $(X_{t_1}, X_{t_2}, \dots, X_{t_k})$ permanezcan inalterados al desplazar el índice de tiempo. Más bien, lo que frecuentemente importa es que la media de las variables aleatorias sea constante y, especialmente, que la condición (2.2) sobre las covarianzas sea válida, consideración que conduce a formular una idea menos exigente de proceso estacionario, tarea que se emprende a continuación.

La Perspectiva de Segundo Orden

En esta sección se introduce una noción de proceso estacionario que impone menos requerimientos sobre la serie que la idea de estacionaridad estricta considerada previamente. En esencia, de lo que se trata es de enfatizar las propiedades de segundo orden de una serie, esto es, las características que pueden expresarse en términos de momentos de orden menor o igual a dos. Este punto de vista es suficiente para utilizar ideas geométrico–algebraicas en el análisis de series de tiempo, y constituye un concepto mucho menos exigente que la noción introducida en la Definición 2.1, esto es, más series serán estacionarias en el sentido presentado a continuación, que con respecto a la noción de estacionaridad estricta previamente introducida.

Definición 3.1. Un proceso estocástico $\{X_t\}$ 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 ,

$$E[X_t] = E[X_{t+h}], \quad \text{y} \quad \text{Cov}[X_s, X_t] = \text{Cov}[X_{s+h}, X_{t+h}]. \quad (3.1)$$

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 $5 - 1 = 16 - 12 = 4$. Comparando las ideas de estacionaridad débil y estricta es claro que si una serie $\{X_t\}$ es tal que todas las variables que la conforman tienen segundo momento finito, entonces la siguiente implicación es válida:

$$\{X_t\} \text{ es estrictamente estacionaria} \implies \{X_t\} \text{ es débilmente estacionaria.} \quad (3.2)$$

Para verificar esta afirmación, primero observe que para cualquier variable aleatoria se tiene que $E[|Y|] \leq \sqrt{E[Y^2]}$ (Dudwicz y Mishra (1988), Mood *et. al.* (1985), de donde se desprende que si Y tiene segundo momento finito, entonces su valor esperado está bien definido. Volviendo ahora a la implicación (3.2) suponga que la serie $\{X_t\}$ es estrictamente estacionaria, y que sus componentes X_t tienen segundo momento finito. En este caso, como acaba de ser verificado, la esperanza de cada variable está bien definida, y como se estableció en los comentarios siguientes a la Definición 2.1, se tiene que (3.1) ocurre, de manera que $\{X_t\}$ es estacionaria en el sentido débil.

Observación 3.1. En este trabajo, el adjetivo estacionario aplicado a un proceso estocástico $\{X_t\}$ significa que la serie es *débilmente estacionaria*.

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

$$\gamma_{\mathbf{X}}(h) = \text{Cov}[X_t, X_{t+h}], \quad h = 0, \pm 1, \pm 2, \dots \quad (3.3)$$

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 $\gamma_{\mathbf{X}}(h) = \text{Cov}[X_t, X_{t+h}] = \text{Cov}[X_s, X_{s+h}]$, 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,

$$\text{Cov}[X_s, X_t] = \gamma_{\mathbf{X}}(t - s). \quad (3.4)$$

A continuación se proporcionan dos ejemplos de series estacionarias.

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 $\text{Cov}[Z_t, Z_s] = 0$, $s \neq t$. 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

$$\{Z_t\} \sim \text{WN}(0, \sigma^2).$$

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

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

Luego,

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

igualdad que muestra que la covarianza entre dos variables Z_t y Z_s depende sólomente 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 $\{Z_t\} \sim \text{WN}(0, \sigma^2)$. □

Ejemplo 3.2. Suponga que $\{Z_t\} \sim \text{WN}(0, \sigma^2)$ 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

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

donde se ha utilizado (3.4) 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

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

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) $\gamma_{\mathbf{X}}(0) = \text{Var}[X_t]$ 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, \dots, 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 $[a_1, a_2, \dots, a_n]$ no es nulo, la desigualdad es estricta en (3.5), esto es, la función $\gamma_{\mathbf{X}}(\cdot)$ es *positiva definida*.

Demostración. A partir de (3.3) se obtiene que $\gamma(0) = \text{Cov}[X_t, X_t] = \text{Var}[X_t]$ para todo t , estableciendo la parte (i). Recuerde ahora la desigualdad

$$|\text{Cov}[X_t, X_{t+h}]| \leq \sqrt{\text{Var}[X_t]} \sqrt{\text{Var}[X_{t+h}]}$$

la cual, via (3.3) y la parte (i), se reduce a $\gamma(h) \leq \sqrt{\gamma(0)} \sqrt{\gamma(0)} = \gamma(0)$. Para verificar la parte (iii), note que la ecuación (3.4) implica que $\gamma_{\mathbf{X}}(h) = \text{Cov}[X_t, X_{t+h}] = \text{Cov}[X_{t+h}, X_t] = \gamma(-h)$, y ahora se comprobarán las partes (iv) y (v). Con este fin, note que siempre se tiene que $\text{Var}[\sum_{i=1}^n a_i X_i] \geq 0$. Luego,

$$\begin{aligned} 0 &\leq \text{Var} \left[\sum_{i=1}^n a_i X_i \right] = \text{Cov} \left[\sum_{i=1}^n a_i X_i, \sum_{j=1}^n a_j X_j \right] \\ &= \sum_{i=1}^n \sum_{j=1}^n a_i \text{Cov}[X_i, X_j] a_j = \sum_{i=1}^n \sum_{j=1}^n a_i \gamma(j-i) a_j \end{aligned}$$

donde la segunda igualdad se debe a la bilinealidad de la covarianza, y es utilizó (3.4) para obtener la última igualdad. Este argumento demuestra (3.5). Para concluir, note que si ninguna combinación no nula de las variables X_t produce una variable aleatoria constante, entonces $0 < \text{Var}[\sum_{i=1}^n a_i X_i]$ siempre que las constantes a_i no formen un vector nulo, de manera que en este caso $\sum_{i=1}^n \sum_{j=1}^n a_i \gamma(j-i) a_j$ es estrictamente mayor a cero. \square

Observación 3.1. Dada la función de autocovarianza $\gamma_{\mathbf{X}}(\cdot)$ y un entero positivo n , defina la matriz \mathbf{G} de orden $n \times n$ mediante

$$G = [\gamma_{\mathbf{X}}(j-i)]_{i,j=1,2,\dots,n}.$$

En este caso, para todo vector $\mathbf{a} = [a_1, a_2, \dots, a_n]' \in \mathbb{R}^n$, no es difícil verificar que

$$\mathbf{a}' \mathbf{G} \mathbf{a} = \sum_{i=1}^n \sum_{j=1}^n a_i \gamma_{\mathbf{X}}(j-i) a_j,$$

y entonces, de acuerdo a la parte (iv) del Lema 3.1, \mathbf{G} es, invariablemente, una matriz no negativa definida, mientras que la función de autocovarianza $\gamma_{\mathbf{X}}(\cdot)$ es positiva definida si y sólo si la matriz \mathbf{G} también posee dicha característica en el sentido utilizado en álgebra lineal (Hoffman y Kunze, 1973, Grossman, 1983, Searle, 1982, Harville, 1997).

Estimación de la Media y la Función de Autocovarianza

Desde la óptica de segundo orden, las características esenciales de un proceso 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.

Denotando mediante μ a la media común de las variables X_t , un estimador 'razonable' de este parámetro es la media muestral \bar{X}_n de los datos X_1, X_2, \dots, X_n obtenidos desde el inicio del proceso de observaciones hasta el momento actual n . Es bien conocido que \bar{X}_n es un estimador insesgado de μ , y a continuación se analizará su consistencia en el sentido de segundo orden, esto es, se presentarán condiciones bajo las cuales $\text{Var}[\bar{X}_n] \rightarrow 0$ conforme el número de observaciones disponibles se incrementa.

Teorema 4.1. Suponga que la función de autocovarianza $\gamma_{\mathbf{X}}(\cdot)$ de una serie estacionaria es absolutamente sumable, esto es, $\sum_{t=-\infty}^{\infty} |\gamma_{\mathbf{X}}(t)| < \infty$. En estas circunstancias,

$$\lim_{n \rightarrow \infty} n \text{Var}[\bar{X}_n] = \sum_{t=-\infty}^{\infty} \gamma_{\mathbf{X}}(t), \quad (4.1)$$

y entonces $\text{Var}[\bar{X}_n] \rightarrow 0$ conforme $n \rightarrow \infty$. Consecuentemente, para cada $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P[|\bar{X}_n - \mu| > \varepsilon] = 0, \quad (4.2)$$

esto es, \bar{X}_n converge en probabilidad a μ conforme el número de observaciones se incrementa.

Demostración. Como punto de partida, observe que a partir de la bilinealidad de la covarianza se desprende

$$\begin{aligned} \text{Var} \left[\sum_{i=1}^n X_i \right] &= \text{Cov} \left[\sum_{i=1}^n X_i, \sum_{j=1}^n X_j \right] \\ &= \sum_{i=1}^n \sum_{j=1}^n \text{Cov} [X_i, X_j] = \sum_{i=1}^n \sum_{j=1}^n \gamma(j-i). \end{aligned} \quad (4.3)$$

A continuación, note que cuando los índices i y j varían entre 1 y n , la diferencia $j-i = h$ asume valores entre $-(n-1)$ y $n-1$; más aún, un valor h en este rango puede obtenerse de $n - |h|$ formas. Para comprobar esta afirmación, suponga primero que $h > 0$. En este caso $j-i = h$ ocurre cuando la pareja (i, j) es cualquiera de las siguientes:

$$(1, 1+h), (2, 2+h), (3, 3+h), \dots, (n-h, h);$$

desde luego, en esta lista hay $n-h$ parejas. Similarmente, si $h < 0$, la igualdad $j-i = h$ ocurre sólo si la pareja (i, j) se encuentra entre

$$(-h+1, 1), (-h+2, 2), (-h+3, 3), \dots, (n, n+h).$$

Observando que $n+h = n-|h|$, pues $h < 0$, se desprende que la igualdad $j-i = h$ ocurre para $n-|h|$ parejas. Finalmente, la posibilidad $j-i = 0$ es válida en el caso de que (i, j) pertenece a $\{(t, t), t = 1, 2, \dots, n\}$, conjunto que consiste de n pares. Esta discusión muestra que cuando i y j varían entre 1 y n , el número de pares (i, j) para los cuales se tiene que $j-i = h$ está dado por $n-|h|$ para cualquier h entre $-(n-1)$ y $(n-1)$. Por lo tanto, la doble sumatoria en la extrema derecha de (4.3) se reduce a $\sum_{h=-(n-1)}^{n-1} (n-|h|)\gamma_{\mathbf{X}}(h)$, y entonces

$$\text{Var} \left[\sum_{i=1}^n X_i \right] = \sum_{h=-(n-1)}^{n-1} (n-|h|)\gamma_{\mathbf{X}}(h).$$

Observe ahora que $\text{Var} [\bar{X}_n] = n^{-2} \text{Var} [\sum_{i=1}^n X_i]$, igualdad que al combinarse con la precedente produce

$$n \text{Var} [\bar{X}_n] = \frac{1}{n} \sum_{h=-(n-1)}^{n-1} (n - |h|) \gamma_{\mathbf{X}}(h) = \sum_{h=-(n-1)}^{n-1} \left(1 - \frac{|h|}{n}\right) \gamma_{\mathbf{X}}(h). \quad (4.4)$$

A partir de esta ecuación, (4.1) puede demostrarse mediante el siguiente argumento: Para cada h entre $-(n-1)$ y $(n-1)$ se tiene que

$$\left| \left(1 - \frac{|h|}{n}\right) \gamma_{\mathbf{X}}(h) \right| = \left(1 - \frac{|h|}{n}\right) |\gamma_{\mathbf{X}}(h)| \leq |\gamma_{\mathbf{X}}(h)|,$$

esto es, cada término en la sumatoria que aparece en la extrema derecha de (4.4) está dominado por el correspondiente valor de la función de autocovarianza; puesto que ésta última se supone absolutamente sumable, i.e., $\sum_{h=-\infty}^{\infty} |\gamma_{\mathbf{X}}(h)| < \infty$, el teorema de convergencia dominada implica que

$$\lim_{n \rightarrow \infty} \sum_{h=-(n-1)}^{n-1} \left(1 - \frac{|h|}{n}\right) \gamma_{\mathbf{X}}(h) = \sum_{h=-\infty}^{\infty} \gamma_{\mathbf{X}}(h),$$

(Ash, 1975) y tomando límite en ambos lados de (4.4) se desprende

$$\lim_{n \rightarrow \infty} n \text{Var} [\bar{X}_n] = \sum_{h=-\infty}^{\infty} \gamma_{\mathbf{X}}(h),$$

estableciendo (4.1). Puesto que el lado derecho en esta convergencia es finito, se concluye que

$$\lim_{n \rightarrow \infty} \text{Var} [\bar{X}_n] = 0. \quad (4.5)$$

Esta ecuación implica la validez de (4.2). En efecto, a partir de la desigualdad de Chebichev (Mood *et. al.* (1985), Dudewic y Mishra (1988)), se obtiene que para cada $\varepsilon > 0$

$$P [|\bar{X}_n - \mu| > \varepsilon] \leq \frac{\text{Var} [\bar{X}_n]}{\varepsilon^2}$$

de manera que usando (4.5) se desprende

$$\lim_{n \rightarrow \infty} P [|\bar{X}_n - \mu| > \varepsilon] \leq \lim_{n \rightarrow \infty} \frac{\text{Var} [\bar{X}_n]}{\varepsilon^2} = 0,$$

estableciendo (4.2) y concluyendo la demostración del Teorema 4.1. \square

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:

$$\hat{\gamma}_{\mathbf{X}}(h) = \begin{cases} \frac{1}{n} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \bar{X}_n)(X_t - \bar{X}_n), & \text{si } |h| < n, \\ 0, & \text{si } |h| \geq n. \end{cases} \quad (4.6)$$

Esta fórmula da origen a una pregunta interesante. Para plantearla, primero suponga que μ , la media común de las variables X_t es conocida. Puesto que $\gamma_{\mathbf{X}}(h)$ es el valor de la covarianza entre dos variables observadas en dos instantes separados por h unidades de tiempo, en el momento n es natural estimar $\gamma_{\mathbf{X}}(h)$, con $0 \leq h < n$, mediante la covarianza de los datos que han sido detectados en momentos que difieren en h unidades. Hay $n - |h|$ parejas con estas características, a saber, $(X_1, X_{1+h}), (X_2, X_{2+h}), \dots, (X_{n-h}, X_n)$. Utilizando estos datos, y ya que μ se supone conocida, $\gamma(h)$ podría estimarse mediante

$$\frac{1}{n-h} \sum_{t=1}^{n-h} (X_{t+h} - \mu)(X_t - \mu).$$

Sin embargo, en las aplicaciones μ no se conoce, de manera que μ se reemplaza por \bar{X}_n para obtener

$$\frac{1}{n-h} \sum_{t=1}^{n-h} (X_{t+h} - \bar{X}_n)(X_t - \bar{X}_n). \quad (4.7)$$

Esta expresión, a la que se arriba de manera intuitiva, difiere de la fórmula en (4.6) pues los divisores— n en (4.6), y $n - h$ en la igualdad precedente—no coinciden.

Luego, la siguiente pregunta surge de inmediato: ¿Hay alguna razón de fondo para utilizar el divisor n en (4.6)? La respuesta es afirmativa. En efecto, si se utiliza el divisor n —como en la Definición forma 4.2—la función de covarianza muestral es *positiva definida*, pero esta propiedad no es necesariamente válida cuando el divisor se cambia a $n - h$, como en (4.7). A continuación se demuestran estos hechos.

Lema 4.1. (i) La función de autocovarianza definida en (4.6) es, siempre, no negativa definida.

(ii) Más aún, si $X_i \neq X_j$ para alguna pareja (i, j) con $i, j \in \{1, 2, \dots, b\}$, entonces $\hat{\gamma}_{\mathbf{X}}(\cdot)$ es positiva definida.

Demostración. (i) Dado un entero positivo fijo k , defina la matriz $\hat{\mathbf{G}}$ de orden $k \times k$ mediante

$$\hat{\mathbf{G}}_{ij} = \hat{\gamma}_{\mathbf{X}}(j - i), \quad i, j = 1, 2, \dots, k. \quad (4.8)$$

Para demostrar la parte (i), es suficiente comprobar que $\hat{\mathbf{G}}$ es una matriz no negativa definida, propiedad que equivale a que \mathbf{G} se pueda factorizar como

$$\hat{\mathbf{G}} = \mathbf{T}\mathbf{T}'. \quad (4.9)$$

Defina $Y_i = X_i - \bar{X}_n$ para $i = 1, 2, \dots, n$, y considere la matriz \mathbf{T} de orden $k \times (k+n)$ determinada por la siguiente expresión:

$$\mathbf{T} = \begin{bmatrix} Y_1 & Y_2 & Y_3 & \cdots & Y_{n-1} & Y_n & 0 & 0 & \cdots & 0 & 0 \\ 0 & Y_1 & Y_2 & \cdots & Y_{n-2} & Y_{n-1} & Y_n & 0 & \cdots & 0 & 0 \\ 0 & 0 & Y_1 & \cdots & Y_{n-3} & Y_{n-2} & Y_{n-1} & Y_n & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & Y_1 & Y_2 & Y_3 & Y_4 & \cdots & Y_n & 0 \\ 0 & 0 & 0 & \cdots & 0 & Y_1 & Y_2 & Y_3 & \cdots & Y_{n-1} & Y_n \end{bmatrix}$$

Verbalmente, esta matriz se construye formando las filas como sigue: primero, la fila uno se obtiene colocando Y_1, Y_2, \dots, Y_n en sucesión, y agregando k ceros. Después, este vector se traslada sucesivamente una unidad a la derecha, incorporando un cero al inicio cada vez que se realiza un traslado, y eliminando el cero final, de

manera que el tamaño de la fila permanece constante e igual a $n + k$. Usando la definición de $\hat{\gamma}_{\mathbf{X}}(\cdot)$ en (4.6), no es difícil ver que

$$\hat{\mathbf{G}} = \frac{1}{n} \mathbf{T} \mathbf{T}',$$

y entonces $\hat{\mathbf{G}}$ es no negativa definida.

(ii) Suponga ahora que $X_i \neq X_j$ para alguna pareja (i, j) con $i \neq j$, $i, j \in \{1, 2, \dots, n\}$. En este caso, se tiene que $X_i \neq \bar{X}_n$ para algún índice i , de manera que $Y_i \neq 0$, y no es difícil comprobar que las filas de la matriz \mathbf{T} definida anteriormente son linealmente independientes, de manera que el rango de \mathbf{T} es k y, consecuentemente, también $\hat{\mathbf{G}}$ tiene rango k . Puesto que la matriz $\hat{\mathbf{G}}$ es siempre no negativa definida, este hecho implica su positividad. \square

A continuación se mostrará que si la expresión (4.7) se utiliza para definir la función de covarianza muestral, entonces ésta no tiene que poseer la propiedad de no negatividad. Defina

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

Lema 4.2. Es posible que la función $\tilde{\gamma}_{\mathbf{X}}(h)$ no posea la propiedad de ser no negativa definida.

Demostración. Para demostrar el Lema es suficiente con proporcionar un ejemplo de un conjunto de datos para los cuales la función $\tilde{\gamma}(\cdot)$ no tenga la propiedad de ser no negativa definida. Dicho ejemplo se presenta a continuación: Suponga que $n = 3$ y que

$$X_1 = x_1 = 1, \quad X_2 = x_2 = 0, \quad X_3 = x_3 = -1.$$

En este caso,

$$\bar{x}_3 = \frac{1}{3} [1 + 0 + (-1)] = 0,$$

y no es difícil ver que

$$\tilde{\gamma}_{\mathbf{X}}(0) = \frac{1}{3-0} [(x_1 - \bar{x}_3)^2 + (x_2 - \bar{x}_3)^2 + (x_3 - \bar{x}_3)^2] = \frac{2}{3},$$

mientras que

$$\begin{aligned} \tilde{\gamma}_{\mathbf{X}}(1) &= \tilde{\gamma}_{\mathbf{X}}(-1) \\ &= \frac{1}{3-1} [(x_1 - \bar{x}_3)(x_2 - \bar{x}_3) + (x_2 - \bar{x}_3)(x_3 - \bar{x}_3)] \\ &= \frac{1}{2} [(1)(0) + (0)(-1)] = 0, \end{aligned}$$

y

$$\tilde{\gamma}_{\mathbf{X}}(2) = \tilde{\gamma}_{\mathbf{X}}(-2) = \frac{1}{3-2} [(x_1 - \bar{x}_3)(x_3 - \bar{x}_3)] = \frac{1}{3-2} [(1)(-1)] = -1.$$

A continuación, forme la matriz

$$\tilde{\mathbf{G}} = \begin{bmatrix} \tilde{\gamma}_{\mathbf{X}}(0) & \tilde{\gamma}_{\mathbf{X}}(1) & \tilde{\gamma}_{\mathbf{X}}(2) \\ \tilde{\gamma}_{\mathbf{X}}(1) & \tilde{\gamma}_{\mathbf{X}}(0) & \tilde{\gamma}_{\mathbf{X}}(1) \\ \tilde{\gamma}_{\mathbf{X}}(2) & \tilde{\gamma}_{\mathbf{X}}(1) & \tilde{\gamma}_{\mathbf{X}}(0) \end{bmatrix} = \begin{bmatrix} 2/3 & 0 & -1 \\ 0 & 2/3 & 0 \\ -1 & 0 & 2/3 \end{bmatrix}.$$

En este caso,

$$[1, 0, 1] \tilde{\mathbf{G}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 2\frac{2}{3} + 2(-1) = -\frac{2}{3} < 0,$$

lo cual muestra que, con estos datos, la función $\tilde{\gamma}_{\mathbf{X}}(\cdot)$ es no posee la propiedad de ser no negativa definida. \square

Una pregunta importante relacionada con la estimación de la función de autocovarianza se refiere a la consistencia del estimador $\hat{\gamma}_{\mathbf{X}}(h)$:

¿Es cierto que $\lim_{n \rightarrow \infty} P[|\hat{\gamma}_{\mathbf{X}}(h) - \gamma_{\mathbf{X}}(h)| > \varepsilon] = 0$?

Bajo condiciones bastante generales, la respuesta a esta pregunta es afirmativa, aunque el argumento que conduce a demostrar este hecho es más complicado que el utilizado para establecer el Teorema 4.1. Los detalles pueden encontrarse en Box y Jenkins (1970) y Anderson (1981) para una clase de procesos conocidos como filtros lineales con coeficientes sumables, y en Cavazos-Cadena (1994) para procesos

con coeficientes cuadrado sumables. Por otro lado, es oportuno mencionar que la varianza de $\tilde{\gamma}_{\mathbf{X}}(h)$ como estimador de $\gamma_{\mathbf{X}}(h)$ es mayor que la correspondiente a $\hat{\gamma}_{\mathbf{X}}(h)$, constituyendo este aspecto un punto más a favor del empleo del estimador $\hat{\gamma}_{\mathbf{X}}(h)$.

Después del desarrollo realizado en esta sección y la precedente, es oportuno reflexionar sobre el trayecto recorrido hasta ahora: Se ha analizado la noción de proceso (débilmente) estacionario y se ha mostrado que los datos disponibles permiten estimar sus características esenciales, a saber, la media común de las variables que conforman la serie, y su función de autocovarianza. Para relacionar estos resultados con los problemas de modelado y de pronóstico en el estudio de series de tiempo, el siguiente paso consiste en determinar una clase amplia de modelos para los cuales se pueda determinar su función de autocovarianza; lo que se pretende es que dichos modelos contengan un número ‘pequeño de parámetros desconocidos’, pero que al variar éstos, se obtenga una clase amplia de funciones de autocovarianza, de manera que entre ellas se pueda encontrar una ‘buena aproximación’ a la función muestral obtenida a partir de los datos. En otras palabras, un enfoque básico al problema de modelado de series de tiempo, consiste en las siguientes etapas:

- (i) Disponer de una clase \mathcal{M} de modelos;
- (ii) Encontrar la función de autocovarianza para cada modelo $\mathbf{m} \in \mathcal{M}$;
- (iii) Determinar un modelo $\mathbf{m}^* \in \mathcal{M}$ cuya función de autocovarianza ‘se ajuste’, en cierto sentido, a los valores calculados de la función de autocovarianza muestral $\hat{\gamma}_{\mathbf{X}}(\cdot)$.

El modelo \mathbf{m}^* en la etapa (iii) es el que se adopta para describir los datos observados, y su función de autocovarianza (y su media) se utilizan para abordar el problema de pronóstico. Por lo tanto, el siguiente paso en este capítulo es

describir la clase de modelos \mathcal{M} a la que se ha hecho referencia, y la siguiente sección representa el primer paso en esta dirección.

Filtros Lineales y el Operador de Retardo

El propósito de esta sección es introducir la idea de filtro lineal. En términos generales, un filtro lineal es un tipo especial de operador que permite transformar un proceso estacionario en otro que también tiene esta característica. Durante el proceso de cambio, la función de autocovarianza original se altera, pero la correspondiente función para el proceso transformado puede determinarse con facilidad. Como punto inicial, es conveniente introducir la noción de operador de retardo.

Definición 5.1. Sea $\{X_t\}$ una serie dada.

(i) El operador de retardo se denota por B y actúa sobre la serie de la siguiente forma:

$$BX_t = X_{t-1}. \quad (5.1)$$

(ii) Las potencias de B actúan sobre variables individuales de la manera natural:

$$B^n X_t = X_{t-n}, \quad t, n = 0, \pm 1, \pm 2, \dots \quad (5.2)$$

(iii) Si $\theta(z) = \theta_0 + \theta_1 z + \dots + \theta_q z^q$ es un polinomio en la variable z , el operador

$$\theta(B) = \theta_0 B^0 + \theta_1 B + \dots + \theta_q B^q$$

actúa sobre la serie $\{X_t\}$ como sigue:

$$\begin{aligned} \theta(B)X_t &= \theta_0 X_t + \theta_1 BX_t + \dots + \theta_q B^q X_t \\ &= \theta_0 X_t + \theta_1 X_{t-1} + \dots + \theta_q X_{t-q}. \end{aligned} \quad (5.3)$$

Para ejemplificar la notación relacionada con el operador de retardo, note que $B^3 X_t = X_{t-3}$, $B^0 X_t = X_t$, mientras que $B^{-3} X_t = X_{t-(-3)} = X_{t+3}$. Usando

la notación introducida en la Definición 5.1, el proceso $\{X_t\}$ en el Ejemplo 3.2 está determinado como $X_t = (1 - 2B)Z_t = Z_t - 2Z_{t-1}$; alternativamente, si $\theta(z)$ es el polinomio dado por $\theta(z) = 1 - 2z$, entonces $X_t = \theta(B)Z_t$. La actuación de un polinomio $\theta(z)$ ‘evaluado en B ’ dada en (5.3), es un caso particular de la idea de filtro lineal presentada a continuación.

Definición 5.2. Sea $\{a_n \mid n = 0, \pm 1, \pm 2, \dots\}$ una sucesión sumable de números reales, i.e.,

$$\sum_{n=-\infty}^{\infty} |a_n| < \infty, \quad (5.4)$$

y sea $\{X_t\}$ una serie estacionaria dada. Si la serie $\{Y_t\}$ se define mediante

$$Y_t = \sum_k a_k X_{t-k} = \sum_k a_k B^k X_t, \quad (5.5)$$

entonces se dice que $\{Y_t\}$ se obtiene aplicando el filtro lineal $\{a_k\}$ a la serie original $\{X_t\}$.

El siguiente teorema es el resultado fundamental referente a la acción de filtro lineales.

Teorema 5.1. Sea $\{a_n\}$ una sucesión absolutamente sumable. Si el proceso $\{X_t\}$ es estacionario, entonces la serie $\{Y_t\}$ determinada a través de (5.5) también es estacionaria.

Demostración. El primer paso consiste en demostrar que Y_t en (5.5) tiene esperanza bien definida. Con este fin, note que $E[|X_t|] \leq \sqrt{E[X_t^2]} = \sqrt{\gamma(0)}$ (Dudewicz y Mishra, 1988), y que

$$|Y_t| = \left| \sum_k a_k X_{t-k} \right| \leq \sum_k |a_k| |X_{t-k}|.$$

Utilizando ahora el teorema de convergencia monótona (Ash, 1975), la desigualdad anterior implica

$$E[|Y_t|] \leq \sum_k |a_k| E[|X_{t-k}|] \leq \sum_k |a_k| \sqrt{\gamma(0)},$$

de manera que el valor esperado de cada variable Y_t es finito. Más aún, si

$$\mu_{\mathbf{X}} = E[X_t],$$

entonces el teorema de convergencia dominada enunciado en Ash (1975) implica

$$\begin{aligned} E[Y_t] &= E \left[\sum_k a_k X_{t-k} \right] \\ &= \sum_k a_k E[X_{t-k}] \\ &= \sum_k a_k \mu_{\mathbf{X}} \end{aligned} \quad (5.6)$$

de manera que $E[Y_t]$ es constante. Por otro lado,

$$\begin{aligned} \text{Cov}[Y_t, Y_{t+h}] &= \text{Cov} \left[\sum_k a_k X_{t-k}, \sum_j a_j X_{t+h-j} \right] \\ &= \sum_k \sum_j a_k \text{Cov}[X_{t-k}, X_{t+h-j}] a_j \\ &= \sum_k \sum_j a_k \gamma_{\mathbf{X}}(k - j + h) a_j \end{aligned} \quad (5.7)$$

de donde se desprende que $\text{Cov}[Y_t, Y_{t+h}]$ depende sólo de h y no de t , estableciendo el carácter estacionario de $\{Y_t\}$. \square

Corolario 5.1. Defina el proceso $\{Y_t\}$ mediante

$$Y_t = \sum_k a_k Z_{t-k},$$

donde $\{Z_t\} \sim \text{WN}(0, \sigma^2)$ y la sucesión $\{a_k\}$ es absolutamente sumable. En este caso $\{Y_t\}$ es una serie estacionaria. Más aún $E[Y_t] = 0$, y

$$\gamma_{\mathbf{Y}}(h) = \sigma^2 \sum_k a_k a_{k+h}. \quad (5.8)$$

Demostración. A partir de (5.5) con $\mu_{\mathbf{X}} = \mu_{\mathbf{Z}} = E[Z_t] = 0$, se desprende que todas las variables Y_t tienen media nula. Por otro lado, utilizando (5.7) con la función de autocovarianza $\gamma_{\mathbf{Z}}(\cdot)$ de $\{Z_t\}$ reemplazando a $\gamma_{\mathbf{X}}(\cdot)$, se obtiene

$$\text{Cov}[Y_t, Y_{t+h}] = \sum_k \sum_j a_k \gamma_{\mathbf{Z}}(k - j + h) a_j.$$

Recordando ahora que $\gamma_{\mathbf{Z}}(h) = 0$ si $h \neq 0$, mientras que $\gamma_{\mathbf{Z}}(0) = \sigma^2$ (vea el Ejemplo 3.1), se desprende que $\gamma_{\mathbf{Z}}(k - j + h) = 0$ cuando $j \neq k + h$, y $\gamma_{\mathbf{Z}}(k - j + h) = \sigma^2$ si $j = k + h$. Por lo tanto,

$$\text{Cov}[Y_t, Y_{t+h}] = \sum_k \sum_j a_k \gamma_{\mathbf{Z}}(k - j + h) a_j = \sum_k a_k a_{k+h} \sigma^2,$$

esto es, $\gamma_{\mathbf{Y}}(h) = \sigma^2 \sum_k a_k a_{k+h}$, la cual es la expresión que se deseaba establecer.

□

Procesos ARMA

En esta sección se introducen varios tipos especiales de procesos estacionarios, los cuales pertenecen a la categoría de filtros lineales. La exposición culmina presentando la noción de procesos autorregresivos de promedios móviles, comúnmente referidos como procesos ARMA.

Definición 6.1. [Proceso de Promedios Móviles] Una serie $\{X_t\}$ es un proceso de promedios móviles de orden q , si existe un proceso $\{Z_t\} \sim \text{WN}(0, \sigma^2)$ y constantes $\theta_1, \theta_2, \dots, \theta_q$ tal que

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \dots + \theta_q Z_{t-q}. \quad (6.1)$$

En este caso se escribe $\{X_t\} \sim \text{MA}(q)$, expresión que se lee como ‘ $\{X_t\}$ es un proceso de promedios móviles de orden q ’.

Note que definiendo el polonomio $\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q$, la igualdad (6.1) equivale a $X_t = \theta(B)Z_t$; vea la Definición 5.1. Utilizando el Corolario 5.1, particularmente la igualdad (5.8), se obtiene la siguiente expresión para la función de autocovarianza de un proceso $\text{MA}(q)$.

Lema 6.1. Si el proceso $\{X_t\} \sim \text{MA}(q)$, entonces

$$\gamma_{\mathbf{X}}(h) = \begin{cases} \sum_{i=0}^{q-|h|} \theta_i \theta_{i+|h|}, & \text{si } |h| \leq q \\ 0, & \text{si } |h| > q. \end{cases}$$

De acuerdo a este resultado, la función de autocovarianza de un proceso MA(q) se anula en argumentos h con valor absoluto mayor a q . Este resultado tiene un recíproco interesante: Si la función de autocovarianza de una serie estacionaria $\{X_t\}$ es tal que $\gamma_X(h) = 0$ para todo h con $|h| > q$, entonces la serie $\{X_t\}$ es un proceso de promedios móviles de orden q (Brockewll y Davis, 1991).

Definición 6.2. [Proceso Autorregresivo.] Una serie $\{X_t\}$ es un proceso autorregresivo orden p , si existe un proceso $\{Z_t\} \sim \text{WN}(0, \sigma^2)$ y constantes $\phi_1, \phi_2, \dots, \phi_p$ tal que

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + Z_t. \quad (6.2)$$

En este caso se escribe $\{X_t\} \sim \text{AR}(p)$, expresión que se lee como ‘ $\{X_t\}$ es un proceso autorregresivo de orden p ’.

Si se define el polinomio $\phi(z)$ mediante $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p$, la ecuación (6.2) es equivalente a $\phi(B)X_t = Z_t$. En este caso, $\phi(z)$ se denomina el polinomio autorregresivo del proceso, y puede demostrarse que existe una serie $\{X_t\}$ que satisfaga (6.2) si y sólo si $\phi(z) \neq 0$ para todo número complejo con $|z| = 1$, esto es, siempre que $\phi(z)$ no posea raíces en el círculo unitario. En este último caso, no se incurre en pérdida de generalidad alguna al suponer que $\phi(z) \neq 0$ para todo z con $|z| \leq 1$, y la función $\psi(z) = 1/\phi(z)$ es analítica en un conjunto abierto que contiene al disco unitario. Luego, $\psi(z)$ admite una expansión en series de potencia alrededor de cero cuyo radio de convergencia es mayor o igual a uno, esto es, existe un número $R > 1$ tal que

$$\frac{1}{\phi(z)} = \psi(z) = \sum_{k=0}^{\infty} \psi_k z^k, \quad |z| \leq R.$$

Con esta notación, la única solución $\{X_t\}$ a la ecuación (6.2) es

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

de manera que $\{X_t\}$ se obtiene aplicando el filtro lineal $\{\psi_k\}$ al ruido blanco $\{Z_t\}$. A continuación se describe un procedimiento para determinar la función de autocovarianza del proceso AR(p) en (6.2).

Lema 6.2. Borockwell y Davis (1991). Suponga que es un proceso autorregresivo de orden p como en (6.2), donde se asume que el polinomio autorregresivo $\phi(z)$ no se anula para todo número complejo z tal que $|z| \leq 1$. En este caso, la función de autocovarianza de $\{X_t\}$ se determina a través del siguiente procedimiento de dos etapas, donde $\sigma^2 > 0$ es la varianza de las variables aleatorias que conforman el ruido blanco.

Etapla 1. Encuentre $\gamma(0), \gamma(1), \dots, \gamma(p)$ resolviendo el sistema de Yule–Walker

$$\begin{aligned} \gamma(0) - \sum_{k=1}^p \gamma(k)\phi_k &= \sigma^2 \\ \gamma(i) - \sum_{k=1}^p \gamma(|i-k|)\phi_k &= 0, \quad i = 1, 2, \dots, p; \end{aligned} \tag{6.3}$$

Etapla 2. Los valores $\gamma(p+1), \gamma(p+2), \dots$ se determinan recursivamente por medio de la siguiente relación: $\gamma(i) = -\sum_{k=1}^p \gamma(i-k)\phi_k, i > p$.

Un análisis de este resultado, y particularmente un estudio completo sobre el sistema de Yule–Walker (6.3), se encuentra en Alemán Valerio (1992). Un resultado importante sobre procesos autorregresivos, es que éstos pueden utilizarse como modelos que ‘se ajustan’ a cualquier función de autocovarianza γ^+ que sea positiva definida, esto es, dado un entero $p > 0$, siempre es posible determinar un proceso AR(p) cuya función de autocovarianza coincide con γ^+ hasta el orden p , propiedad que se establece formalmente en el siguiente resultado.

Lema 6.3. Sea $\gamma(\cdot)$ una función de autocovarianza positiva definida. Dado un entero positivo p , existe un proceso autorregresivo $\{X_t\}$ de orden p tal que

su función de autocovarianza $\gamma_{\mathbf{X}}(\cdot)$ coincide con $\gamma(\cdot)$ en el conjunto $\{-p, -(p-1), \dots, 0, \dots, p-1, p\}$, i.e.,

$$\gamma_{\mathbf{X}}(h) = \gamma(h), \quad h = 0, \pm 1, \pm 2, \dots, \pm p.$$

Dicho proceso AR(p) se construye por medio del siguiente método.

(i) Defina la matriz \mathbf{G} de orden $p \times p$ y el vector $\mathbf{g} \in \mathbb{R}^p$ mediante

$$\mathbf{G} = [\gamma(i-j)]_{i,j=1,2,\dots,p}, \quad \text{y} \quad \mathbf{g} = [\gamma(1), \gamma(2), \dots, \gamma(p)]',$$

y observe que la matriz \mathbf{G} es no singular, pues $\gamma(\cdot)$ es positiva definida.

(ii) Determine $\tilde{\sigma}^2$ y $\tilde{\phi}_1, \tilde{\phi}_2, \dots, \tilde{\phi}_p$ mediante

$$\tilde{\sigma}^2 = \gamma(0) - \mathbf{g}'\mathbf{G}^{-1}\mathbf{g}$$

y

$$[\tilde{\phi}_1, \tilde{\phi}_2, \dots, \tilde{\phi}_p]' = \mathbf{G}^{-1}\mathbf{g}.$$

En este caso, si $\{Z_t\} \sim \text{WN}(0, \tilde{\sigma}^2)$, la serie $\{X_t\}$ es el proceso AR(p) dado por

$$X_t = \tilde{\phi}_1 X_{t-1} + \tilde{\phi}_2 X_{t-2} + \dots + \tilde{\phi}_p X_{t-p} + Z_t.$$

Este resultado, cuya demostración puede encontrarse en Brockwell y Davis (1991), muestra que función de autocovarianza positiva definida siempre puede pensarse como proveniente de un proceso autorregresivo ‘hasta cierto orden’ p elegido de antemano, resultado que muestra el papel relevante que los procesos AR(p) tienen en el proceso de modelado de una serie de tiempo. La siguiente clase de proceso combina las ideas en las dos definiciones precedentes.

Definición 6.3. [Proceso Autorregresivo y de Promedios Móviles.] Una serie $\{X_t\}$ es un proceso autorregresivo de promedio móvil de orden (p, q) , si existe un proceso $\{Z_t\} \sim \text{WN}(0, \sigma^2)$ y constantes $\phi_1, \phi_2, \dots, \phi_p$ y $\theta_1, \theta_2, \dots, \theta_q$ tales que

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \dots + \theta_q Z_{t-q}. \quad (6.4)$$

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

Si $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p$, y $\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q$, entonces la igualdad (6.4) se escribe en forma compacta como $\phi(B)X_t = \theta(B)Z_t$. Un proceso $\{X_t\}$ que satisfaga (6.4) existe si y sólo si sólo si $\phi(z) \neq 0$ para todos los números complejos sobre el círculo unitario, y como en el caso de los procesos AR(p), no se incurre en pérdida de generalidad alguna al suponer que $\phi(z) \neq 0$ para todo z con $|z| \leq 1$; en estas circunstancias, definiendo

$$\psi(z) = \frac{\theta(z)}{\phi(z)},$$

se tiene que $\psi(z)$ se expande en serie de potencias alrededor del origen, y que dicha expansión es válida en un conjunto abierto que contiene al disco unitario, i.e.,

$$\psi(z) = \frac{\theta(z)}{\phi(z)} = \sum_{k=0}^{\infty} \psi_k z^k, \quad |z| < R,$$

donde $R > 1$. En este caso, la única solución $\{X_t\}$ a la ecuación (6.2) es

$$X_t = \sum_{k=0}^{\infty} \psi_k Z_{t-k} = \sum_{k=0}^{\infty} \psi_k B^k Z_t, \quad (6.5)$$

esto es, $\{X_t\}$ se obtiene aplicando el filtro $\{\psi_k\}$ al ruido blanco $\{Z_t\}$; la función de autocovarianza de un proceso ARMA(p, q) puede determinarse mediante un procedimiento similar al utilizado para calcular la función correspondiente a un proceso AR(p). El conjunto de procesos ARMA(p, q) es la clase amplia de procesos \mathcal{M} a la que se hizo referencia al final de la Sección 5, y el resultado establecido en el Lema 6.3 hace plausible que dentro de esta clase puedan encontrarse modelos cuya función de autocovarianza ‘se aproxime’ a la calculada a partir de los datos muestrales. El procedimiento de ajustar un proceso ARMA(p, q) se encuentra descrito en la literatura y no será un objeto directo de estudio en este trabajo, el cual se concentra en los problemas de bondad de ajuste y de pronóstico. Para detalles

sobre aspectos no cubiertos explícitamente, se remite al lector a Brockwell y Davis (1991), donde puede encontrarse una demostración de que el proceso ARMA (p, q) en (6.4) está unívocamente determinado a través de (6.5).

La Descomposición Clásica

En esta sección se discute un modelo clásico para series de tiempo, el cual incluye componentes estacionales, de tendencia, y la influencia aleatoria de una serie estacionaria. Al idea es poner de manifiesto la utilidad de ajustar procesos ARMA (p, q) , aún en el caso de que el proceso originalmente observado no tenga la propiedad de estacionaridad. Con este fin, suponga que los datos que se registran sucesivamente, i.e, $x_1, x_2, \dots, x_n, \dots$ son una realización de un proceso $\{X_t\}$ con la siguiente estructura:

$$X_t = s_t + \kappa_t + W_t, \quad (7.1)$$

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

$$s_t = s_{t-d}, \quad \text{para algún entero } d > 0, \quad (7.2)$$

esto es, s_t tiene período d , mientras que κ_t es una función polinomial a la cual se le conoce como la función de tendencia:

$$\kappa_t \quad \text{es un polinomio de grado } g. \quad (7.3)$$

La representación (7.1) 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 dicha serie no es estacionaria. Como se sugirió en la Sección 2, esta representación refleja características de los datos que se observan comúnmente en las aplicaciones.

El carácter no estacionario de la serie $\{X_t\}$ parece indicar que la metodología delineada anteriormente no es adecuada para analizarla. Sin embargo, a

continuación se comprobará que una transformación adecuada conduce a un proceso estacionario $\{Y_t\}$, y que un modelo y un método de pronóstico para este último proceso, se interpretan de inmediato en términos de los datos originales. Para ver que esto es así, defina el polinomio $\Delta(z)$ mediante

$$\Delta(z) = (1 - z)^g(1 - z^d). \quad (7.4)$$

En este caso

$$\Delta(B)s_t = (1 - B^g)(1 - B^d)s_t = (1 - B^g)[s_t - s_{t-d}] = (1 - B^g)[0] = 0,$$

donde la segunda igualdad utiliza el hecho de que s_t es una función de período d . Además, debido a que κ_t es un polinomio de grado g , y a que $\Delta(z)$ tiene grado $g + 1$, no es difícil ver que

$$\Delta(B)\kappa_t = 0,$$

y por lo tanto,

$$\Delta(B)X_t = \Delta(B)s_t + \Delta(B)\kappa_t + \Delta(B)W_t = \Delta(B)W_t.$$

Así, el proceso $\{Y_t\} = \{\Delta(B)X_t\}$ es estacionario, pues coincide con el proceso que se obtiene aplicando $\Delta(B)$, un polinomio en el operador de retardo, a la serie $\{W_t\}$. Por lo tanto, se puede ajustar un proceso ARMA (p, q) a la sucesión $\{Y_t\}$; suponga que el modelo que produce un ajuste adecuado es

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \theta(B)Z_t = \sum_{k=1}^p \phi_k Y_{t-k} + \theta(B)Z_t, \quad (7.5)$$

igualdad que equivale a

$$Y_t = \sum_{i=1}^p \phi_i \Delta(B)X_{t-i} + \theta(B)Z_t,$$

pues $Y_t = \Delta(B)X_t$. Observando ahora que el polinomio $\Delta(z)$ en (7.4) se expresa como $\Delta(z) = 1 + \sum_{i=1}^{g+1} a_i z^i$ para algunas constantes a_i , se obtiene que $Y_t =$

$\Delta(B)X_t = X_t + \sum_{i=1}^{g+1} a_i B^i X_t$, igualdad que al combinarse en la anterior ecuación desplegada permite obtener

$$X_t = \sum_{k=1}^p \phi_k \Delta(B)X_{t-k} - \sum_{i=1}^{g+1} a_i B^i X_{t-i} + \theta(B)Z_t, \quad (7.6)$$

de manera que al seleccionar un modelo para $\{Y_t\}$ se está, implícitamente, obteniendo uno para las serie $\{X_t\}$. Por otro lado, debido a que $\Delta(z)$ es un polinomio de grado $g + 1$, $Y_t = \Delta(B)X_t$ expresa en términos de $X_t, X_{t-1}, \dots, X_{t-(g+1)}$. Además, el proceso de observaciones se inicia en $t = 1$, de manera que las variables Y_t que pueden ser explícitamente calculadas son Y_{g+2}, Y_{g+3}, \dots . En consecuencia, en cualquier tiempo $n > g + 2$, un método de pronóstico para la variable Y_{n+1} debe utilizar solamente los datos conocidos Y_{g+2}, \dots, Y_n . Sea \hat{Y}_{n+1} un pronóstico para Y_{n+1} dado por

$$\hat{Y}_{n+1} = \mathbb{P}(Y_{g+2}, \dots, Y_n), \quad (7.7)$$

donde \mathbb{P} es una función determinada a través del modelo ARMA (p, q) para el proceso $\{Y_t\}$. Debido a que $Y_t = X_t + \sum_{i=1}^{g+1} a_i B^i X_{t-i}$, se tiene que

$$\hat{Y}_{n+1} = \mathbb{P} \left(X_{g+2} + \sum_{i=1}^{g+1} a_i B^i X_{g+2-i}, \dots, X_n + \sum_{i=1}^{g+1} a_i B^i X_{n-i} \right).$$

Note ahora que $\hat{Y}_{n+1} = \hat{X}_{n+1} + \sum_{i=1}^{g+1} a_i B^i X_{n+1-i}$, pues todas las variables X_t en la sumatoria son conocidas en el tiempo n , y combinando esta igualdad con la precedente se obtiene

$$\hat{X}_{n+1} = \mathbb{P} \left(X_{g+2} + \sum_{i=1}^{g+1} a_i B^i X_{g+2-i}, \dots, X_n + \sum_{i=1}^{g+1} a_i B^i X_{n-i} \right) - \sum_{i=1}^{g+1} a_i B^i X_{n+1-i}. \quad (7.8)$$

Esta discusión puede resumirse como sigue: Al considerar una serie $\{X_t\}$ no estacionaria generada de acuerdo al modelo clásico (7.1), se tiene que ésta no es estacionaria. Sin embargo, definiendo el polinomio $\Delta(z)$ como en (7.4), el proceso transformando $\{Y_t\} = \{\Delta(B)X_t\}$ es estacionario, y hay una relación estrecha entre los problemas de modelado y de pronóstico para los procesos $\{X_t\}$ y $\{Y_t\}$:

(i) Al construir un modelo ARMA (p, q) para la serie transformada $\{Y_t\}$, se obtiene, implícitamente, un modelo par el proceso original $\{X_t\}$; vea (7.5) y (7.6).

(ii) Un método de pronóstico para la serie $\{Y_t\}$, genera una fórmula para construir pronósticos de observaciones futuras del proceso original $\{X_t\}$.

A partir de estos comentarios, se desprende que, aunque la mayoría de las series que surgen en las aplicaciones no son estacionarias, un análisis de procesos que si tienen la propiedad de estacionaridad, tiene repercusiones importantes para los problemas de modelado y pronóstico de procesos que surgen en un contexto práctico; la idea de aplicar un polinomio en el operador de retardo a una serie $\{X_t\}$ para obtener un proceso estacionario se debe a Box y Jenkins (1970) .

Los Problemas de Pronóstico y Bondad de Ajuste

El propósito de esta sección es discutir el alcance y la importancia de los problemas que se considerarán en el desarrollo subsecuente. Sea $\{Y_t\}$ una serie dada con $E[Y_t] = 0$ para todo t , y considere la hipótesis

$$\mathcal{H}_0: \text{Las variables } Y_i \text{ son } i.i.d.$$

Cuando las variables Y_t tienen segundo momento finito, este supuesto implica que el proceso $\{Y_t\}$ es un ruido blanco, y que no es posible obtener una relación entre las variables Y_t que permita establecer pronósticos para observaciones futuras en términos de datos actuales. Así, una primera etapa en el análisis de un proceso, es decidir si los datos observados apoyan o no la hipótesis \mathcal{H}_0 ; cuando se rechaza este supuesto, es razonable emprender la tarea de construir un modelo que se ajuste a los datos observados. Un propósito fundamental de este trabajo, es estudiar una prueba de significancia no paramétrica para el supuesto \mathcal{H}_0 , empresa que se realiza en el siguiente capítulo, donde se analiza un procedimiento distinto al método usual conocido como la prueba χ^2 (Serfling , 1980). La importancia del estudio que se realiza, no se restringe a las etapas iniciales del estudio de una serie.

Por ejemplo, suponga que la hipótesis \mathcal{H}_0 ha sido rechazada, y que se determina un modelo ARMA(p, q), el cual se piensa que describe el comportamiento de los datos observados de forma adecuada; vea (7.5) En este caso, es posible obtener estimaciones \widehat{Z}_t de los valores asumidos por el ruido, y una forma de decidir si el modelo (7.5) es adecuado, es probar el supuesto

$$\widetilde{\mathcal{H}}_0: \text{Las variables } \widehat{Z}_t \text{ son } i.i.d.$$

Cuando $\widetilde{\mathcal{H}}_0$ no se rechaza al aplicar el procedimiento de prueba, se tiene que los datos observados sustentan la validez del modelo (7.5). En resumen, una prueba de significancia para el supuesto de que un proceso consta de variables *i.i.d.* es relevante, tanto en las etapas iniciales del análisis de un proceso, como al juzgar la adecuación de un modelo para la serie.

El otro problema fundamental 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, \dots, y_k asumidos por las variables Y_1, Y_2, \dots, Y_k . Un pronóstico para el valor que tomará Y_{k+1} , la variable cuyo valor se conocerá en la siguiente observación, es

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

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

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

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 la validez de una técnica conocida como el procedimiento de Durbin–Levinson desde una óptica geométrica.

Conclusión

En este capítulo se ha presentado la idea de proceso (débilmente) estacionario, noción que se formuló en la Definición 3.1. Además, se elaboró sobre el concepto de filtro lineal, y sobre la clase de series de tiempo conocidas como procesos ARMA (p, q) . Dentro de este tipo de modelos puede buscarse uno que describa adecuadamente a los datos observados, en el sentido de que la función de autocovarianza derivada del modelo supuesto se aproxime a la calculada a partir de los datos. Un aspecto que es conveniente enfatizar, es que la mayoría de las series que surgen en las aplicaciones no son estacionarias, pero que a través de una transformación adecuada, es frecuentemente posible reducir su análisis al estudio de un proceso estacionario. Finalmente, se ha destacado la importancia de las pruebas de significancia para la hipótesis de que una serie de datos proviene de variables aleatorias independientes con distribución común, tanto antes de buscar un modelo para la serie, como para juzgar la bondad de un modelo obtenido a partir de datos muestrales. Este problema, conjuntamente con la elaboración recursiva de pronósticos constituye la espina dorsal del desarrollo de este trabajo.

Capítulo 3

Una Prueba de Bondad de Ajuste

Este capítulo trata sobre un problema importante en el proceso de modelado de una serie de tiempo, a saber, decidir si un modelo estimado a partir de datos muestrales puede considerarse adecuado. El análisis que se realiza se refiere a una técnica de prueba para la hipótesis de que la sucesión estimada de errores consta de variables aleatorias independientes e idénticamente distribuidas. El procedimiento que se analiza es de naturaleza no paramétrica, y se basa en determinar, de forma aproximada, la distribución del estadístico de máximos y mínimos locales que presentan los datos conforme el número de observaciones crece. La principal contribución técnica consiste en establecer, de forma detallada y elemental, la normalidad asintótica de una suma de variables aleatorias que no son independientes.

Introducción

El tema de este capítulo es la bondad de ajuste de una sucesión de datos a las condiciones de independencia e identidad de distribución. Este problema surge, de manera natural, cuando al modelar una serie de tiempo se busca que los errores ‘estimados’ formen un ruido blanco. De manera más específica, el estudio que se realiza a continuación persigue desarrollar la teoría de un método de prueba para la hipótesis de que una sucesión consta de variables aleatorias independientes e idénticamente distribuidas. Debido a que este supuesto no involucra forma especial alguna para la distribución común de las observaciones, es necesario que el procedimiento sea de naturaleza no paramétrica (Randles y Wolfe, 1981, Hollander y Wolfe, 1983), y entre las diversas maneras de abordar este problema, se analizará una técnica basada en el conteo del número de máximos y mínimos locales que una

sucesión contiene. Para describir el procedimiento, considere variables aleatorias Y_1, Y_2, \dots, Y_n , donde Y_k se observa en el instante k . En este caso, un máximo local ocurre en el tiempo i cuando la observación Y_i es mayor que los dos datos adyacentes, esto es,

$$Y_i > Y_{i-1}, \quad Y_i > Y_{i+1}, \quad i = 2, 3, \dots, n-1,$$

mientras que en el momento j se presenta un mínimo local si

$$Y_j < Y_{j-1}, \quad Y_j < Y_{j+1}, \quad j = 2, 3, \dots, n-1.$$

Suponiendo que las variables aleatorias Y_1, Y_2, \dots, Y_n son independientes e idénticamente distribuidas, el objetivo de este capítulo es estudiar la distribución de T_n , el número total de máximos y mínimos locales dentro de dicha sucesión. Como producto del análisis que se realiza a continuación, será posible prescribir un conjunto de valores dentro del cual, siempre que las variables Y_i se ajusten a la condición de ser *i.i.d.*, es ‘normal’ que T_n esté incluido. De esta forma, el hecho de que en un caso particular se observe que T_n asume un valor fuera de ese rango, se considerará como evidencia de que las variables Y_k no son independientes, o que no tienen distribución común. El análisis de este problema se basa en el Teorema Central de Límite, establecido en el Teorema 6.2 más adelante, pero la argumentación requerida no es una aplicación directa de dicho resultado clásico. De hecho, la *principal contribución* técnica de este capítulo, consiste en demostrar la normalidad asintótica de una suma de variables idénticamente distribuidas, cuando éstas *no son completamente independientes*.

La *organización* del capítulo es la siguiente: En la Sección 2 se introduce formalmente el estadístico T_n , el número total de máximos y mínimos locales dentro de una serie de variables aleatorias, el cual se expresa como una suma de términos con distribución de Bernoulli, mientras que en el Teorema 3.1 de la Sección 3 se establecen expresiones para su valor esperado y varianza. Este resultado se

demuestra en la Sección 5, después de incluir las herramientas necesarias en la Sección 4. En la Sección 6 se inicia el estudio de la distribución límite de T_n , y en el Teorema 6.1 se establece su normalidad asintótica conforme el número de observaciones se incrementa. Debido a que T_n no se expresa como una suma de variables aleatorias *i.i.d.*, es necesario desarrollar resultados auxiliares en la Sección 7, los cuales permiten demostrar el Teorema 6.1 en la Sección 8. La aplicación del resultado sobre la normalidad asintótica de T_n al problema de bondad de ajuste se presenta en la Sección 9 y, finalmente, la exposición concluye en el Sección 10 con algunos comentarios breves.

El Estadístico T_n

El objetivo de esta sección es describir formalmente la construcción del estadístico T , el número total de máximos y mínimos locales que se observan en una sucesión de variables aleatorias Y_1, Y_2, \dots, Y_n . Como punto de partida, es conveniente recordar la notación de indicadores que es común en la literatura estadística (Mood *et. al.*, 1985, Dudewicz y Mishra, 1988). Dado un espacio de probabilidad (Ω, \mathcal{F}, P) , sea A un evento, esto es un subconjunto de Ω que pertenece a \mathcal{F} . En este caso, el indicador de A se denota mediante $I[A]$, y es la variable aleatoria que asume sólo los valores 0 y 1 de acuerdo a la siguiente regla:

$$I[A](\omega) = \begin{cases} 1, & \text{si } \omega \in A \\ 0, & \text{si } \omega \notin A \end{cases} \quad (2.1)$$

De esta forma, el valor que $I[A]$ toma puede utilizarse para determinar si A ocurre o no: $I[A] = 1$ sólo cuando el resultado del experimento pertenece a A (lo cual significa que “ A ocurrió”), mientras que $I[A] = 0$ indica que dicho resultado no es un miembro de A . Note que $I[A]$ asume el valor 1 con probabilidad $P[A]$, de manera que

$$E[I[A]] = P[A]. \quad (2.2)$$

Suponga ahora que A_1, A_2, \dots, A_K son K eventos, y considere la sucesión correspondiente de variables indicadoras $I[A_1], I[A_2], \dots, I[A_K]$. Cuando el experimento

aleatorio se realiza y se observa el resultado ω , se obtiene

$$I[A_1](\omega), I[A_2](\omega), \dots, I[A_K](\omega),$$

sucesión que consta de ceros y unos; el número total de unos es $\sum_{i=1}^K I[A_i]$, y a partir de (2.1) se desprende que esta sumatoria es el número de eventos A_i a los cuales pertenece el resultado del experimento. Esta breve discusión puede resumirse estableciendo que la variable aleatoria $\sum_{i=1}^K I[A_i]$ es el número total de eventos A_i a los que pertenece el resultado del experimento aleatorio subyacente. Después de este preámbulo, a continuación se determinará una expresión para el número de máximos y mínimos locales dentro de una sucesión de variables aleatorias.

Definición 2.1. Sean Y_1, Y_2, \dots, Y_n variables aleatorias definidas en un mismo espacio de probabilidad (Ω, \mathcal{F}, P) .

(i) Para cada $i = 2, 3, \dots, n-1$, las variables aleatorias Ψ_i^+ y Ψ_i^- están dadas por

$$\Psi_i^+ = I[Y_{i-1} < Y_i, Y_i > Y_{i+1}], \quad y \quad \Psi_i^- = I[Y_{i-1} > Y_i, Y_i < Y_{i+1}],$$

mientras que

$$\Psi_i = \Psi_i^+ + \Psi_i^-.$$

(ii) La variable aleatoria T_n se define mediante

$$T_n = \sum_{i=2}^{n-1} \Psi_i.$$

De acuerdo a la discusión precedente, $\Psi_i^+ = 1$ si la sucesión Y_1, Y_2, \dots, Y_n alcanza un máximo local en la observación i , mientras que $\Psi_i^- = 1$ indica que en dicha observación se tiene un mínimo local. Por lo tanto, T_n representa el número total de máximos o mínimos locales que ocurren dentro de la sucesión observada Y_1, Y_2, \dots, Y_n . La notación introducida en la definición precedente será

de gran utilidad en el análisis subsecuente, el cual se desarrollará bajo la siguiente condición.

Hipótesis 2.1. Las variables aleatorias Y_1, Y_2, \dots, Y_n son independientes e idénticamente distribuidas, y su distribución común es continua.

El Valor Esperado y la Varianza de T_n

El supuesto presentado en la Hipótesis 2.1, el cual se mantendrá en el desarrollo del capítulo, permite determinar características esenciales de la distribución de T_n , como su media y su varianza, para abordar posteriormente el problema de determinar su distribución asintótica, esto es, cuando el número de observaciones crece sin límite. Note que la Hipótesis 2.1 es de naturaleza no paramétrica, en el sentido de que no impone forma especial alguna a la distribución común de las variables aleatorias Y_1, Y_2, \dots, Y_n , y sólo se supone que ésta tiene densidad; esta característica permite utilizar los resultados que se obtienen más adelante en el estudio de series de tiempo, en cuyo análisis no se especifica una forma especial para la distribución del ruido.

Teorema 3.1. Bajo la Hipótesis 2.1, el valor esperado del estadístico T_n es

$$E[T_n] = \frac{2(n-2)}{3},$$

mientras que su varianza está determinada por

$$\text{Var}[T_n] = \frac{16n-29}{90}.$$

La demostración de este teorema requiere varios preliminares técnicos, los cuales se presentan en la siguiente sección. Por el momento, es oportuno establecer el siguiente resultado auxiliar.

Lema 3.1. Suponga que la Hipótesis 2.1 ocurre, y sean Ψ_i^+ y Ψ_i^- las variables aleatorias introducidas en la Definición 2.1. En este caso,

$$E[\Psi_i^+] = E[\Psi_i^-] = \frac{1}{3}, \quad i = 2, \dots, n-1.$$

Demostración. De acuerdo a la Definición 2.1(i), $\Psi_i^+ = I[Y_{i-1} < Y_i, Y_i > Y_{i+1}]$, de manera que (2.2) implica que

$$E[\Psi_i^+] = P[Y_{i-1} < Y_i, Y_i > Y_{i+1}]. \quad (3.1)$$

Para evaluar la probabilidad en el lado derecho de esta igualdad, denote mediante $f(\cdot)$ a la densidad común de las variables Y_1, Y_2, \dots, Y_n , y observe que, debido a la Hipótesis 2.1, la densidad conjunta de las variables aleatorias Y_{i-1}, Y_i, Y_{i+1} es $f(y_{i-1})f(y_i)f(y_{i+1})$, de manera que

$$\begin{aligned} P[Y_{i-1} < Y_i, Y_i > Y_{i+1}] &= \int \int \int_{y_{i-1} < y_i, y_i > y_{i+1}} f(y_{i-1})f(y_i)f(y_{i+1}) dy_{i-1} dy_i dy_{i+1} \\ &= \int_{-\infty}^{\infty} f(y_i) \left[\int_{-\infty}^{y_i} f(y_{i-1}) dy_{i-1} \left[\int_{-\infty}^{y_i} f(y_{i+1}) dy_{i+1} \right] \right] dy_i \end{aligned}$$

donde la integral triple se ha expresado en términos de integrales iteradas. Denotando mediante $F(\cdot)$ a la función de distribución correspondiente a la densidad $f(\cdot)$, se desprende que

$$\begin{aligned} P[Y_{i-1} < Y_i, Y_i > Y_{i+1}] &= \int_{-\infty}^{\infty} f(y_i) \left[\int_{-\infty}^{y_i} f(y_{i-1}) dy_{i-1} \left[\int_{-\infty}^{y_i} f(y_{i+1}) dy_{i+1} \right] \right] dy_i \\ &= \int_{-\infty}^{\infty} f(y_i) \left[\int_{-\infty}^{y_i} f(y_{i-1}) dy_{i-1} [F(y_i)] \right] dy_i \\ &= \int_{-\infty}^{\infty} f(y_i)F(y_i) \left[\int_{-\infty}^{y_i} f(y_{i-1}) dy_{i-1} \right] dy_i \\ &= \int_{-\infty}^{\infty} f(y_i)F(y_i) [F(y_i)] dy_i \\ &= \int_{-\infty}^{\infty} f(y_i)(F(y_i))^2 dy_i. \end{aligned}$$

Combinando esta igualdad con (3.1), y recordando que $F'(x) = f(x)$, se desprende que

$$E[\Psi_i^+] = \int_{-\infty}^{\infty} f(y_i)(F(y_i))^2 dy_i = \frac{1}{3} F(y_i)^3 \Big|_{-\infty}^{\infty} = \frac{1}{3}.$$

De manera similar puede establecerse que $E[\Psi_i^-] = \frac{1}{3}$. □

Instrumentos Auxiliares

En esta sección se presentan las herramientas que, adicionalmente al Lema 3.1, se utilizarán para demostrar el Teorema 3.1. Los resultados se presentan en la forma de los Lemas 4.1–4.4.

Lema 4.1. Con la misma notación que en el Lema 3.1, las siguientes afirmaciones (i) y (ii) son válidas para cada $i = 2, 3, \dots, n - 2$:

$$(i) \quad E[\Psi_i^+ \Psi_{i+1}^+] = 0, \quad \text{y} \quad E[\Psi_i^+ \Psi_{i+1}^-] = \frac{5}{24}.$$

Similarmente,

$$(ii) \quad E[\Psi_i^- \Psi_{i+1}^-] = 0, \quad \text{y} \quad E[\Psi_i^- \Psi_{i+1}^+] = \frac{5}{24}.$$

Demostración. (i) Primero observe que $\Psi_i^+ = I[Y_{i-1} < Y_i, Y_i > Y_{i+1}]$, y $\Psi_{i+1}^+ = I[Y_i < Y_{i+1}, Y_{i+1} > Y_i]$. Luego, cuando $\Psi_i^+ = 1$, se tiene que $Y_i > Y_{i+1}$, y en este caso $\Psi_{i+1}^+ = 0$. Por lo tanto, $\Psi_i^+ \Psi_{i+1}^+ = 0$, y entonces $E[\Psi_i^+ \Psi_{i+1}^+] = 0$. Para establecer la igualdad restante, note que

$$\begin{aligned} \Psi_i^+ \Psi_{i+1}^- &= I[Y_{i-1} < Y_i, Y_i > Y_{i+1}] I[Y_i > Y_{i+1}, Y_{i+1} < Y_{i+2}] \\ &= I[Y_{i-1} < Y_i, Y_i > Y_{i+1}, Y_{i+1} < Y_{i+2}]. \\ &= I[Y_{i-1} < Y_i, (Y_i \wedge Y_{i+2}) > Y_{i+1}], \end{aligned}$$

donde $a \wedge b$ denota el mínimo de los números a y b . Para evaluar el valor esperado de $E[\Psi_i^+ \Psi_{i+1}^-]$, primero se calculará la esperanza condicional dados los valores de Y_i y Y_{i+2} :

$$\begin{aligned} E[\Psi_i^+ \Psi_{i+1}^- | Y_i = y_i, Y_{i+2} = y_{i+2}] &= E[I[Y_{i-1} < Y_i, (Y_i \wedge Y_{i+2}) > Y_{i+1}] | Y_i = y_i, Y_{i+2} = y_{i+2}] \\ &= E[I[Y_{i-1} < y_i, (y_i \wedge y_{i+2}) > Y_{i+1}] | Y_i = y_i, Y_{i+2} = y_{i+2}] \\ &= E[I[Y_{i-1} < y_i, (y_i \wedge y_{i+2}) > Y_{i+1}]] \end{aligned}$$

donde la última igualdad se debe a la independencia de las variables aleatorias Y_k .

Luego,

$$\begin{aligned}
 E[\Psi_i^+ \Psi_{i+1}^- | Y_i = y_i, Y_{i+2} = y_{i+2}] &= E[I[Y_{i-1} < y_i, (y_i \wedge y_{i+2}) > Y_{i+1}]] \\
 &= P[Y_{i-1} < y_i, (y_i \wedge y_{i+2}) > Y_{i+1}] \\
 &= P[[Y_{i-1} < y_i]P[Y_{i+1} < (y_i \wedge y_{i+2})]] \\
 &= F(y_i)F(y_i \wedge y_{i+2})
 \end{aligned}$$

Por lo tanto, $E[\Psi_i^+ \Psi_{i+1}^- | Y_i, Y_{i+2}] = F(Y_i)F(Y_i \wedge Y_{i+2})$, y el teorema de la doble esperanza implica que

$$E[\Psi_i^+ \Psi_{i+1}^-] = E[E[\Psi_i^+ \Psi_{i+1}^- | Y_i, Y_{i+2}]] = E[F(Y_i)F(Y_i \wedge Y_{i+2})],$$

de donde se obtiene la siguiente expresión:

$$E[\Psi_i^+ \Psi_{i+1}^-] = \int \int F(y_i)F(y_i \wedge y_{i+2})f(y_i)f(y_{i+2}) dy_i dy_{i+2}. \quad (4.1)$$

La integral en el lado derecho se extiende sobre todos los valores posibles de y_i y y_{i+2} . Para evaluarla se descompondrá la región de integración en dos partes:

$$\begin{aligned}
 &\int \int_{y_i < y_{i+2}} F(y_i)F(y_i \wedge y_{i+2})f(y_i)f(y_{i+2}) dy_i dy_{i+2} \\
 &= \int \int_{y_i < y_{i+2}} F(y_i)F(y_i)f(y_i)f(y_{i+2}) dy_i dy_{i+2} \\
 &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{y_{i+2}} F(y_i)^2 f(y_i) dy_i \right] f(y_{i+2}) dy_{i+2} \\
 &= \int_{-\infty}^{\infty} \left[\frac{F(y_i)^3}{3} \Big|_{y_i=-\infty}^{y_{i+2}} \right] f(y_{i+2}) dy_{i+2} \\
 &= \int_{-\infty}^{\infty} \frac{F(y_{i+2})^3}{3} f(y_{i+2}) dy_{i+2} = \frac{F(y_{i+2})^4}{12} \Big|_{-\infty}^{\infty} = \frac{1}{12}.
 \end{aligned}$$

Por otro lado,

$$\begin{aligned}
& \int \int_{y_i \geq y_{i+2}} F(y_i)F(y_i \wedge y_{i+2})f(y_i)f(y_{i+2}) dy_i dy_{i+2} \\
&= \int \int_{y_i \geq y_{i+2}} F(y_i)F(y_{i+2})f(y_i)f(y_{i+2}) dy_i dy_{i+2} \\
&= \int_{-\infty}^{\infty} F(y_i)f(y_i) \left[\int_{-\infty}^{y_i} F(y_{i+2})f(y_{i+2}) dy_{i+2} \right] dy_i \\
&= \int_{-\infty}^{\infty} F(y_i)f(y_i) \left[\frac{F(y_{i+2})^2}{2} \Big|_{y_{i+2}=-\infty}^{y_i} \right] dy_i \\
&= \int_{-\infty}^{\infty} F(y_i)f(y_i) \left[\frac{F(y_i)^2}{2} \right] dy_i \\
&= \int_{-\infty}^{\infty} \frac{F(y_i)^3}{2} f(y_i) dy_i = \frac{F(y_i)^4}{8} \Big|_{-\infty}^{\infty} = \frac{1}{8}.
\end{aligned}$$

Combinando las dos últimas igualdades desplegadas con (4.1) se obtiene

$$E[\Psi_i^+ \Psi_{i+1}^-] = \frac{1}{12} + \frac{1}{8} = \frac{5}{24}.$$

Este argumento completa la deomstración de la parte (i), mientras que la parte (ii) se establece mediante argumentos similares. \square

Lema 4.2. Con la misma notación que en el Lema 3.1, las siguientes afirmaciones

(i) y (ii) son válidas para cada $i = 2, 3, \dots, n-3$:

$$(i) \quad E[\Psi_i^+ \Psi_{i+2}^+] = \frac{2}{15};$$

$$(ii) \quad E[\Psi_i^+ \Psi_{i+2}^-] = \frac{11}{120}.$$

Similarmente,

$$(iii) \quad E[\Psi_i^- \Psi_{i+2}^-] = \frac{2}{15}, \quad \text{y} \quad E[\Psi_i^- \Psi_{i+2}^+] = \frac{11}{120}.$$

Demostración. (i) Como punto de partida, observe que

$$\begin{aligned}
\Psi_i^+ \Psi_{i+2}^+ &= I[Y_{i-1} < Y_i, Y_i > Y_{i+1}] I[Y_{i+1} < Y_{i+2}, Y_{i+2} > Y_{i+3}] \\
&= I[Y_{i-1} < Y_i, Y_i > Y_{i+1}, Y_{i+1} < Y_{i+2}, Y_{i+2} > Y_{i+3}] \\
&= I[Y_{i-1} < Y_i, (Y_i \wedge Y_{i+2}) > Y_{i+1}, Y_{i+2} > Y_{i+3}]
\end{aligned}$$

Dados los valores y_i y y_{i+2} que las variables aleatorias Y_i y Y_{i+2} asumen, la esperanza condicional de $\Psi_i^+ \Psi_{i+2}^+$ está determinada por

$$\begin{aligned}
& E[\Psi_i^+ \Psi_{i+2}^+ | Y_i = y_i, Y_{i+2} = y_{i+2}] \\
&= E[I[Y_{i-1} < Y_i, (Y_i \wedge Y_{i+2}) > Y_{i+1}, Y_{i+2} > Y_{i+3}] | Y_i = y_i, Y_{i+2} = y_{i+2}] \\
&= E[I[Y_{i-1} < y_i, (y_i \wedge y_{i+2}) > Y_{i+1}, y_{i+2} > Y_{i+3}] | Y_i = y_i, Y_{i+2} = y_{i+2}] \\
&= P[Y_{i-1} < y_i, (y_i \wedge y_{i+2}) > Y_{i+1}, y_{i+2} > Y_{i+3} | Y_i = y_i, Y_{i+2} = y_{i+2}] \\
&= P[Y_{i-1} < y_i, (y_i \wedge y_{i+2}) > Y_{i+1}, y_{i+2} > Y_{i+3}] \\
&= P[Y_{i-1} < y_i] P[Y_{i+1} < (y_i \wedge y_{i+2})] P[Y_{i+3} < y_{i+2}] \\
&= F(y_i) F(y_i \wedge y_{i+2}) F(y_{i+2})
\end{aligned}$$

donde las últimas igualdades se deben a la independencia de las variables aleatorias Y_k . Luego, la esperanza condicional de $\Psi_i^+ \Psi_{i+2}^+$ dados Y_i y Y_{i+2} está dada por, $E[\Psi_i^+ \Psi_{i+2}^+ | Y_i, Y_{i+2}] = F(Y_i) F(Y_i \wedge Y_{i+2}) F(Y_{i+2})$, y el teorema de la doble esperanza implica que

$$\begin{aligned}
E[\Psi_i^+ \Psi_{i+2}^+] &= E [E[\Psi_i^+ \Psi_{i+2}^+ | Y_i, Y_{i+2}]] \\
&= E[F(Y_i) F(Y_i \wedge Y_{i+2}) F(Y_{i+2})] \\
&= \int \int F(y_i) F(y_i \wedge y_{i+2}) F(y_{i+2}) f(y_i) f(y_{i+2}) dy_i dy_{i+2}.
\end{aligned} \tag{4.2}$$

Como en la demostración del lema precedente, la integral en esta igualdad se evaluará descomponiendo la región de integración en dos partes:

$$\begin{aligned}
& \int \int_{y_i < y_{i+2}} F(y_i) F(y_i \wedge y_{i+2}) F(y_{i+2}) f(y_i) f(y_{i+2}) dy_i dy_{i+2} \\
&= \int \int_{y_i < y_{i+2}} F(y_i) F(y_i) F(y_{i+2}) f(y_i) f(y_{i+2}) dy_i dy_{i+2} \\
&= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{y_{i+2}} F(y_i)^2 f(y_i) dy_i \right] F(y_{i+2}) f(y_{i+2}) dy_{i+2} \\
&= \int_{-\infty}^{\infty} \left[\frac{F(y_i)^3}{3} \Big|_{y_i=-\infty}^{y_{i+2}} \right] F(y_{i+2}) f(y_{i+2}) dy_{i+2} \\
&= \int_{-\infty}^{\infty} \frac{F(y_{i+2})^4}{3} f(y_{i+2}) dy_{i+2} = \frac{F(y_{i+2})^5}{15} \Big|_{-\infty}^{\infty} = \frac{1}{15}.
\end{aligned}$$

De manera similar,

$$\int \int_{y_i > y_{i+2}} F(y_i)F(y_i \wedge y_{i+2})F(y_{i+2})f(y_i)f(y_{i+2}) dy_i dy_{i+2} = \frac{1}{15},$$

y combinando estas igualdades con (4.2) se desprende que

$$E[\Psi_i^+ \Psi_{i+2}^+] = \frac{1}{15} + \frac{1}{15} = \frac{2}{15}.$$

(ii) Note que

$$\begin{aligned} \Psi_i^+ \Psi_{i+2}^- &= I[Y_{i-1} < Y_i, Y_i > Y_{i+1}]I[Y_{i+1} > Y_{i+2}, Y_{i+2} < Y_{i+3}] \\ &= I[Y_{i-1} < Y_i, Y_i > Y_{i+1}, Y_{i+1} > Y_{i+2}, Y_{i+2} < Y_{i+3}] \\ &= I[Y_{i-1} < Y_i, Y_i > Y_{i+1} > Y_{i+2}, Y_{i+2} < Y_{i+3}]. \end{aligned}$$

La esperanza de $E[\Psi^+ \Psi_{i+2}^-]$ se evaluará utilizando un argumento de esperanza condicional: Para $y_i > y_{i+2}$, se tiene que

$$\begin{aligned} E[\Psi_i^+ \Psi_{i+2}^- | Y_i = y_i, Y_{i+2} = y_{i+2}] &= E[I[Y_{i-1} < Y_i, Y_i > Y_{i+1} > Y_{i+2}, Y_{i+2} < Y_{i+3}] | Y_i = y_i, Y_{i+2} = y_{i+2}] \\ &= E[I[Y_{i-1} < y_i, y_i > Y_{i+1} > y_{i+2}, y_{i+2} < Y_{i+3}] | Y_i = y_i, Y_{i+2} = y_{i+2}] \\ &= E[I[Y_{i-1} < y_i, y_i > Y_{i+1} > y_{i+2}, y_{i+2} < Y_{i+3}]] \\ &= P[Y_{i-1} < y_i, y_i > Y_{i+1} > y_{i+2}, y_{i+2} < Y_{i+3}] \\ &= P[Y_{i-1} < y_i]P[y_i > Y_{i+1} > y_{i+2}]P[y_{i+2} < Y_{i+3}] \\ &= F(y_i)[F(y_i) - F(y_{i+2})][1 - F(y_{i+2})] \end{aligned}$$

donde las últimas igualdades se deben a la independencia de las variables aleatorias Y_k . Luego,

$$E[\Psi_i^+ \Psi_{i+2}^- | Y_i, Y_{i+2}] = F(Y_i)[F(Y_i) - F(Y_{i+2})][1 - F(Y_{i+2})]I[Y_i > Y_{i+2}],$$

de manera que

$$\begin{aligned} E[\Psi_i^+ \Psi_{i+2}^-] &= E[E[\Psi_i^+ \Psi_{i+2}^- | Y_i, Y_{i+2}]] \\ &= E[F(Y_i)[F(Y_i) - F(Y_{i+2})][1 - F(Y_{i+2})]I[Y_i > Y_{i+2}]] \\ &= \int \int_{y_i > y_{i+2}} F(y_i)[F(y_i) - F(y_{i+2})][1 - F(y_{i+2})]f(y_i)f(y_{i+2}) dy_i dy_{i+2} \end{aligned} \tag{4.3}$$

y para evaluar la integral en esta expresión observe primero que

$$\begin{aligned}
& \int_{y_{i+2}}^{\infty} F(y_i)[F(y_i) - F(y_{i+2})]f(y_i) dy_i \\
&= \int_{y_{i+2}}^{\infty} F(y_i)^2 f(y_i) - F(y_{i+2}) \int_{y_{i+2}}^{\infty} F(y_i) f(y_i) dy_i \\
&= \frac{F(y_i)^3}{3} \Big|_{y_i=y_{i+2}}^{\infty} - F(y_{i+2}) \frac{F(y_i)^2}{2} \Big|_{y_i=y_{i+2}}^{\infty} \\
&= \frac{1 - F(y_{i+2})^3}{3} - F(y_{i+2}) \frac{1 - F(y_{i+2})^2}{2}
\end{aligned}$$

Por lo tanto,

$$\begin{aligned}
& \int_{-\infty}^{\infty} \left[\int_{y_{i+2}}^{\infty} F(y_i)[F(y_i) - F(y_{i+2})]f(y_i) dy_i \right] [1 - F(y_{i+2})]f(y_{i+2}) dy_{i+2} \\
&= \int_{-\infty}^{\infty} \left[\frac{1 - F(y_{i+2})^3}{3} - F(y_{i+2}) \frac{1 - F(y_{i+2})^2}{2} \right] [1 - F(y_{i+2})]f(y_{i+2}) dy_{i+2} \\
&= \int_0^1 \left[\frac{1 - u^3}{3} - u \frac{1 - u^2}{2} \right] [1 - u] du
\end{aligned}$$

donde se realizó el cambio de variable $u = F(y_{i+2})$ para obtener la última expresión. En consecuencia,

$$\begin{aligned}
& \int_{-\infty}^{\infty} \left[\int_{y_{i+2}}^{\infty} F(y_i)[F(y_i) - F(y_{i+2})]f(y_i) dy_i \right] [1 - F(y_{i+2})]f(y_{i+2}) dy_{i+2} \\
&= \int_0^1 \left[\frac{1 - u^3}{3} - u \frac{1 - u^2}{2} \right] [1 - u] du \\
&= \int_0^1 \left[\frac{1 - u^3 - u + u^4}{3} - \frac{u - u^3 - u^2 + u^4}{2} \right] du \\
&= \frac{1 - 1/4 - 1/2 + 1/5}{3} - \frac{1/2 - 1/4 - 1/3 + 1/5}{2} \\
&= \frac{1 - 1/4 - 1/2 + 1/5}{3} - \frac{1/2 - 1/4 - 1/3 + 1/5}{2} \\
&= \frac{9}{60} - \frac{7}{120} = \frac{11}{120}.
\end{aligned}$$

(iii) Esta parte se obtiene a través de argumentos similares a los empleados en la demostración de (i) y (ii). □

Lema 4.3. Para cada $i = 2, 3, \dots, n - 2$,

$$(i) \quad \text{Cov} [\Psi_i^+, \Psi_{i+1}^+] = -\frac{1}{9} \quad \text{y} \quad \text{Cov} [\Psi_i^+, \Psi_{i+1}^-] = \frac{7}{72}.$$

$$(ii) \quad \text{Cov} [\Psi_i^-, \Psi_{i+1}^+] = \frac{7}{72} \quad \text{y} \quad \text{Cov} [\Psi_i^+, \Psi_{i+1}^-] = -\frac{1}{9}.$$

En consecuencia,

$$(iii) \quad \text{Cov} [\Psi_i, \Psi_{i+1}] = -\frac{1}{36}.$$

Demostración. (i) Usando los resultados en los Lemas 3.1 y 3.2, se desprende que

$$\begin{aligned} \text{Cov} [\Psi_i^+, \Psi_{i+1}^+] &= E[\Psi_i^+ \Psi_{i+1}^+] - E[\Psi_i^+]E[\Psi_{i+1}^+] \\ &= 0 - \frac{1}{3} \times \frac{1}{3} = -\frac{1}{9} \end{aligned}$$

y

$$\begin{aligned} \text{Cov} [\Psi_i^+, \Psi_{i+1}^-] &= E[\Psi_i^+ \Psi_{i+1}^-] - E[\Psi_i^+]E[\Psi_{i+1}^-] \\ &= \frac{5}{24} - \frac{1}{3} \times \frac{1}{3} \\ &= \frac{15}{72} - \frac{8}{72} = \frac{7}{72} \end{aligned}$$

(ii) Esta parte se desprende mediante argumentos similares a los empleados en la parte (i).

(iii) Recordando que $\Psi_k = \Psi_k^+ + \Psi_k^-$, las propiedades establecidas en los dos incisos precedentes, conjuntamente con la bilinealidad de la covarianza, implican que

$$\begin{aligned} \text{Cov}[\Psi_i, \Psi_{i+1}] &= \text{Cov} [\Psi_i^+ + \Psi_i^-, \Psi_{i+1}^+ + \Psi_{i+1}^-] \\ &= \text{Cov} [\Psi_i^+ \Psi_{i+1}^+] + \text{Cov} [\Psi_i^+ \Psi_{i+1}^-] + \text{Cov} [\Psi_i^- \Psi_{i+1}^+] + \text{Cov} [\Psi_i^- \Psi_{i+1}^-] \\ &= -\frac{1}{9} + \frac{7}{72} - \frac{1}{9} + \frac{7}{72} \\ &= -\frac{2}{9} + \frac{7}{36} \\ &= -\frac{8}{36} + \frac{7}{36} = -\frac{1}{36} \end{aligned}$$

□

Lema 4.4. Para cada $i = 2, 3, \dots, n - 3$,

$$(i) \quad \text{Cov} [\Psi_i^+, \Psi_{i+2}^+] = \frac{1}{45} \quad \text{y} \quad \text{Cov} [\Psi_i^+, \Psi_{i+2}^-] = -\frac{7}{360}.$$

$$(ii) \quad \text{Cov} [\Psi_i^-, \Psi_{i+2}^+] = \frac{1}{45} \quad \text{y} \quad \text{Cov} [\Psi_i^+, \Psi_{i+2}^-] = -\frac{7}{360}.$$

En consecuencia,

$$(iii) \quad \text{Cov} [\Psi_i, \Psi_{i+2}] = \frac{1}{180}.$$

Demostración. (i) A partir de los Lemas 3.1 y 3.2, se obtiene

$$\begin{aligned} \text{Cov} [\Psi_i^+, \Psi_{i+2}^+] &= E[\Psi_i^+ \Psi_{i+2}^+] - E[\Psi_i^+]E[\Psi_{i+2}^+] \\ &= \frac{2}{15} - \frac{1}{3} \times \frac{1}{3} = \frac{6}{45} - \frac{5}{45} = \frac{1}{45} \end{aligned}$$

y

$$\begin{aligned} \text{Cov} [\Psi_i^+, \Psi_{i+2}^-] &= E[\Psi_i^+ \Psi_{i+2}^-] - E[\Psi_i^+]E[\Psi_{i+2}^-] \\ &= \frac{11}{120} - \frac{1}{3} \times \frac{1}{3} \\ &= \frac{11}{120} - \frac{1}{9} = \frac{33}{360} - \frac{40}{360} = -\frac{7}{360} \end{aligned}$$

(ii) Esta parte puede establecerse a través de los mismos argumentos utilizados para demostrar la parte (i).

(iii) La bilinealidad de la covarianza, combinada con las expresiones de las dos partes previamente demostradas, implica

$$\begin{aligned} &\text{Cov}[\Psi_i, \Psi_{i+2}] \\ &= \text{Cov} [\Psi_i^+ + \Psi_i^-, \Psi_{i+2}^+ + \Psi_{i+2}^-] \\ &= \text{Cov} [\Psi_i^+ \Psi_{i+2}^+] + \text{Cov} [\Psi_i^+ \Psi_{i+2}^-] + \text{Cov} [\Psi_i^- \Psi_{i+2}^+] + \text{Cov} [\Psi_i^- \Psi_{i+2}^-] \\ &= \frac{1}{45} - \frac{7}{360} - \frac{7}{360} + \frac{1}{45} \\ &= \frac{2}{45} - \frac{7}{180} = \frac{8}{180} - \frac{7}{180} = \frac{1}{180} \end{aligned}$$

□

Demostración del Teorema 3.1

El propósito de esta sección es demostrar la validez de las expresiones para la media y la varianza de T_n enunciadas en la Sección 3. El argumento

depende de los resultados en los Lemma 3.1, y 4.1–4.4 de la sección precedente. Observe $\Psi_i = \Psi_i^+ + \Psi_i^-$, de manera que a partir del Lema 3.1 se obtiene que $E[\Psi_i] = E[\Psi_i^+] + E[\Psi_i^-] = 1/3 + 1/3$. Por conveniencia, esta y dos igualdades más se reproducen a continuación:

$$\begin{aligned} E[\Psi_i] &= \frac{2}{3}, \quad i = 2, 3, \dots, n-1 \\ \text{Cov}[\Psi_i, \Psi_{i+1}] &= -\frac{1}{36}, \quad i = 2, 3, \dots, n-2, \\ \text{Cov}[\Psi_i, \Psi_{i+2}] &= \frac{1}{180}, \quad i = 2, 3, \dots, n-3; \end{aligned} \quad (5.1)$$

vea los Lemas 4.3 y 4.4.

Demostración del Teorema 3.1. (i) De acuerdo a la Definición 2.1(ii), $T_n = \sum_{i=2}^{n-1} \Psi_i$, de manera que, a partir de (5.1) se sigue que

$$E[T_n] = \sum_{i=2}^{n-1} E[\Psi_i] = \sum_{i=2}^{n-1} \frac{2}{3} = \frac{2(n-2)}{3}.$$

(ii) Observe que la varianza de T_n está dada por

$$\text{Var}[T_n] = \text{Var}\left[\sum_{i=2}^{n-1} \Psi_i\right] = \sum_{i=2}^{n-1} \text{Var}[\Psi_i] + 2 \sum_{2 \leq i < j \leq n-1} \text{Cov}[\Psi_i, \Psi_j]. \quad (5.2)$$

Para evaluar esta expresión, primero observe que cada variable Ψ_i tiene distribución de Bernoulli con parámetro $\frac{2}{3}$, esto es, Ψ_i toma sólo los valores 0 y 1, y $P[\Psi_i = 1] = \frac{2}{3}$. Por lo tanto, $\text{Var}[\Psi_i] = \frac{2}{3} \times \frac{1}{3} = \frac{2}{9}$. Luego,

$$\sum_{i=2}^{n-1} \text{Var}[\Psi_i] = \sum_{i=2}^{n-1} \frac{2}{9} = \frac{2(n-2)}{9}. \quad (5.3)$$

Para calcular la segunda suma en (5.1), observe los siguientes hechos (a)–(c):

(a) Considere una pareja (i, j) , donde $i, j \in \{2, 3, \dots, n-1\}$ y son tales que $j \geq i+3$. En este caso, Ψ_i depende de Y_{i-1}, Y_i, Y_{i+1} , mientras que Ψ_j es función de Y_{j-1}, Y_j, Y_{j+1} ; como $j \geq i+3$ se tiene que $j-1 \geq i+2 > i+1$, y entonces Ψ_i y Ψ_j dependen de grupos disjuntos de variables Y_k , y entonces la Hipótesis 2.1 implica que Ψ_i y Ψ_j son variables aleatorias independientes. En consecuencia,

$$\text{Cov}[\Psi_i, \Psi_j] = 0, \quad \text{si } j \geq i+3.$$

Por lo tanto,

$$\sum_{2 \leq i < j \leq n-1} \text{Cov} [\Psi_i, \Psi_j] = \sum_{i=2}^{n-2} \text{Cov} [\Psi_i, \Psi_{i+1}] + \sum_{i=2}^{n-3} \text{Cov} [\Psi_i, \Psi_{i+2}]. \quad (5.4)$$

(b) La primera suma en la extrema derecha de la anterior igualdad tiene $(n - 3)$ términos, y cada uno de ellos es igual a $-\frac{1}{36}$, por (5.1). Luego,

$$\sum_{i=2}^{n-2} \text{Cov} [\Psi_i, \Psi_{i+1}] = -\frac{n-3}{36}. \quad (5.5)$$

(b) Similarmente, (5.1) implica que

$$\sum_{i=2}^{n-3} \text{Cov} [\Psi_i, \Psi_{i+2}] = \frac{n-4}{180}. \quad (5.6)$$

Combinando (5.2)–(5.6), se desprende que

$$\begin{aligned} \text{Var} [T_n] &= \frac{2(n-2)}{9} + 2 \left[-\frac{n-3}{36} + \frac{n-4}{180} \right] \\ &= \frac{2n-4}{9} + 2 \left[\frac{-5n+15}{180} + \frac{n-4}{180} \right] \\ &= \frac{2n-4}{9} + 2 \left[\frac{-4n+11}{180} \right] \\ &= \frac{40n-80}{180} + \frac{22-8n}{180} \\ &= \frac{32n-58}{180} = \frac{16n-29}{90}. \end{aligned}$$

Esto completa la demostración del Teorema 3.1 □

La Distribución Asintótica de T_n

El Estadístico T_n introducido en la Definición 2.1 tiene una distribución discreta, la cual es difícil de determinar de manera exacta. Por esta razón, es importante encontrar un procedimiento que permita aproximar dicha distribución. El siguiente teorema establece que, conforme el número de observaciones crece, la distribución de la variable aleatoria T_n , apropiadamente normalizada, converge hacia la distribución normal estándar.

Teorema 6.1. Sea T_n el número de máximos y mínimos locales que se presentan la sucesión de observaciones Y_1, Y_2, \dots, Y_n , y sean μ_n y σ_n^2 la media y la varianza de T_n , respectivamente, esto es,

$$\mu_n = \frac{2n-4}{9}, \quad \text{y} \quad \sigma_n^2 = \frac{16n-29}{90}.$$

Con esta notación, bajo la Hipótesis 2.1 se tiene que

$$\frac{T_n - \mu_n}{\sigma_n} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1). \quad (6.1)$$

Observación 6.1. En el enunciado del teorema precedente, $\mathcal{N}(0, 1)$ es la distribución normal estándar, y en este momento es oportuno recordar que la convergencia (6.1) tiene el siguiente significado: Para todo $a \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} P \left[\frac{T_n - \mu_n}{\sigma_n} \leq a \right] = \int_{-\infty}^a \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \Phi(a),$$

donde $\Phi(\cdot)$ es la función de distribución correspondiente a la distribución normal estándar; vea, por ejemplo, Serffling(1985). Luego, dados dos números a y b , con $a < b$,

$$\lim_{n \rightarrow \infty} P \left[a < \frac{T_n - \mu_n}{\sigma_n} \leq b \right] = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \Phi(b) - \Phi(a),$$

y entonces, para n ‘grande’, la probabilidad del evento $\left[a < \frac{T_n - \mu_n}{\sigma_n} \leq b \right]$ puede ser aproximada por $\Phi(b) - \Phi(a)$, evitando la tarea de calcular la distribución de T_n de manera exacta.

La demostración del Teorema 6.1 se presentará en la Sección 8, y se obtiene combinando los instrumentos técnicos desarrollados más adelante, con el Teorema Central de Límite enunciado a continuación.

Teorema 6.2. Sean $X_1, X_2, \dots, X_n \dots$ variables aleatorias independientes idénticamente distribuidas con segundo momento finito. Defina $S_n = \sum_{i=1}^n X_i$ y denote mediante μ_n y σ_n^2 a la media y la varianza de S_n , respectivamente. En este caso,

$$\frac{S_n - \mu_n}{\sigma_n} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Comparando los enunciados de los Teoremas 6.1 y 6.2, es natural preguntarse por qué no es posible obtener la convergencia (6.1) aplicando directamente el Teorema Central de Límite. La razón es que a pesar de que T_n se expresa como la suma de las variables aleatorias Ψ_i , las cuales tienen la misma distribución, éstas no son independientes.

Resultados Auxiliares

Esta sección contiene los instrumentos técnicos que se utilizarán para demostrar el Teorema 6.1. El punto de partida, es la introducción de variables aleatorias que permiten la aplicación directa del Teorema Central de Límite.

Definición 7.1. Sea K un entero positivo fijo mayor a 3, y defina la variable aleatoria A_r mediante

$$A_r = \sum_{i=K(r-1)+3}^{Kr} \Psi_i, \quad r = 1, 2, \dots \quad (7.1)$$

Teorema 7.1. (i) Bajo la Hipótesis 2.1, las variables aleatorias son independientes e idénticamente distribuidas.

(ii) Cada variable A_r tiene la misma distribución que T_K . Por lo tanto,

$$E[A_r] = \frac{2(K-2)}{3}, \quad y \quad \text{Var}[A_r] = \frac{16K-29}{90}.$$

Demostración. (i) Recuerde que cada variable Ψ_i depende de Y_{i-1}, Y_i , y Y_{i+1} ; vea la Definición 2.1. Como la variable aleatoria A_r se construye a partir de $\Psi_{(K-1)r+3}, \Psi_{(K-1)r+4}, \Psi_{(K-1)r+5}, \dots, \Psi_{Kr}$, se tiene que

$$A_r \text{ es función de } Y_{K(r-1)+2}, Y_{K(r-1)+3}, \dots, Y_{Kr+1}. \quad (7.2)$$

Considere ahora dos variables aleatorias A_r y A_s con $s > r$. En este caso,

$$A_s \text{ es función de } Y_{K(s-1)+2}, Y_{K(s-1)+3}, \dots, Y_{Ks+1}; \quad (7.3)$$

Como $s > r$, se tiene que $s - 1 \geq r$, y entonces $K(s - 1) + 2 \geq Kr + 2 > Kr + 1$. Luego, (7.2) y (7.3) implican que A_s y A_r dependen de subconjuntos disjuntos de las variables $Y_1, Y_2, \dots, Y_n \dots$. Como esta afirmación es válida para todos los índices $s > r$, se desprende que A_1, A_2, \dots son independientes, por la Hipótesis 2.1. Para concluir, observe que A_r es una suma de las $K - 2$ variables $\Psi_{K(r-1)+3}, \Psi_{K(r-1)+4}, \dots, \Psi_{Kr}$, y que combinando la Hipótesis 2.1 con la Definición 2.1 se desprende que

$$(\Psi_{K(r-1)+3}, \Psi_{K(r-1)+4}, \dots, \Psi_{Kr}) \quad \text{y} \quad (\Psi_2, \Psi_3, \dots, \Psi_{K-1})$$

tienen la misma distribución, y entonces lo mismo ocurre con

$$A_r = \Psi_{K(r-1)+3} + \Psi_{K(r-1)+4} + \dots + \Psi_{Kr} \quad \text{y} \quad T_K = \Psi_2 + \Psi_3 + \dots + \Psi_{K-1};$$

vea, por ejemplo, Randles y Wolfe (1981). En resumen, se ha demostrado que las variables aleatorias son independientes, y que cada una de ellas tiene la misma distribución que T_K , estableciendo la parte (i).

(ii) Observe que $E[A_r] = E[T_K]$ y $\text{Var}[A_r] = \text{Var}[T_K]$ para cada r , pues A_r y T_K tienen la misma distribución. Utilizando el Teorema 3.1 con K en lugar de n , se desprende que

$$E[A_r] = \frac{2(K-2)}{3}, \quad \text{y} \quad \text{Var}[A_r] = \frac{16K-29}{90},$$

concluyendo la demostración del teorema. □

Aplicando el Teorema Central de Límite, se obtiene el siguiente resultado.

Corolario 7.1. Defina la variable aleatoria U_M mediante $U_M = \sum_{r=1}^M A_r$, y sea U_M^* la correspondiente variable aleatoria estandarizada, esto es,

$$U_M^* = \frac{U_M - E[U_M]}{\sqrt{\text{Var}[U_M]}}. \quad (7.4)$$

En este caso, conforme M converge a ∞ , $U_M^* \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$.

En el resto de esta sección, se estudia la aproximación de la variable aleatoria T_n , apropiadamente estandarizada, por medio de sumas U_M . Para establecer el resultado principal en esta dirección, es conveniente introducir primero la siguiente notación: Dado un entero $n > K$, $m(n)$ denota la parte entera de del cociente $\frac{n-1}{K}$, esto es,

$$m(n) = \left[\frac{n-1}{K} \right],$$

mientras que

$$T_n^* = \frac{T_n - E[T_n]}{\sqrt{\text{Var}[T_n]}} = \frac{T_n - \mu_n}{\sigma_n}, \quad (7.5)$$

es la versión estandarizada de T_n .

Teorema 7.2. Defina la variable aleatoria D_n mediante

$$D_n = T_n^* - U_{m(n)}^*. \quad (7.6)$$

En este caso, existe una constante C_0 y una sucesión $\{\delta_n\}$ que converge a cero, tales que

$$\text{Var}[D_n] \leq \frac{C_0}{K} + \delta_n.$$

Este es el resultado principal en la ruta hacia la determinación de la distribución asintótica de T_n . El argumento para demostrar este teorema depende de los siguientes dos lemas.

Lema 7.1. Defina la variable aleatoria V_n mediante

$$V_n = (T_n - E[T_n]) - (U_{m(n)} - E[U_{m(n)}]).$$

En este caso, para cada $n > K$, se tiene que $E[V_n] = 0$ y

$$\text{Var}[V_n] \leq m(n) + K.$$

Demostración. Primero note que

$$U_{m(n)} - E[U_{m(n)}] = \sum_{r=1}^{m(n)} (A_r - E[A_r]) = \sum_{r=1}^{m(n)} \sum_{i=K(r-1)+3}^{Kr} \left[\Psi_i - \frac{2}{3} \right]$$

Comparando esta expresión con

$$T_n - E[T_n] = \sum_{k=2}^{n-1} \Psi_k - \frac{2(n-2)}{3} = \sum_{k=2}^{n-1} \left[\Psi_k - \frac{2}{3} \right],$$

se desprende que

$$\begin{aligned} V_n &= (T_n - E[T_n]) - (U_{m(n)} - E[U_{m(n)}]) \\ &= \left[\Psi_2 - \frac{2}{3} \right] + \sum_{r=2}^{m(n)-1} \sum_{i=K(r-1)+1}^{K(r-1)+2} \left[\Psi_i - \frac{2}{3} \right] \\ &\quad + \sum_{i=Km(n)+1}^{n-1} \left[\Psi_i - \frac{2}{3} \right], \end{aligned} \tag{7.7}$$

de donde es claro que V_n tiene esperanza nula. Denote a los términos de esta expresión mediante

$$B_1 = \left[\Psi_2 - \frac{2}{3} \right], \quad B_r = \sum_{i=K(r-1)+1}^{K(r-1)+2} \left[\Psi_i - \frac{2}{3} \right], \quad r = 2, 3, \dots, m(n) - 1,$$

y

$$C = \sum_{i=Km(n)+1}^{n-1} \left[\Psi_i - \frac{2}{3} \right],$$

de tal manera que con esta notación (7.7) puede expresarse como

$$V_n = B_1 + B_2 + \dots + B_{m(n)-1} + C, \tag{7.8}$$

Observe ahora los siguientes hechos que se desprenden de la Definición 2.1:

(a) B_1 depende de Y_1, Y_2, Y_3 y, debido a que Ψ_1 tiene distribución de Bernoulli, con parámetro $2/3$, se tiene que $\text{Var}[B_1] = 2/9 \leq 1$;

(b) Para $r = 2, 3, \dots, m(n)-1$, B_r es una función de $Y_{K(r-1)}, Y_{K(r-1)+1}, Y_{K(r-1)+2}, Y_{K(r-1)+3}$; debido a que B_r involucra la suma de dos variables Ψ_i consecutivas, el Teorema 3.1 implica que

$$\text{Var}[B_r] = \frac{16(4) - 29}{90} \leq 1.$$

(c) Las variables aleatorias $B_1, B_2, \dots, B_{m(n)-1}$ son independientes, pues se expresan en términos de subconjuntos disjuntos de variables Y_1, Y_2, \dots, Y_n .

(d) Observe que $m(n)$ es la parte entera del cociente $(n-1)/K$, de manera que

$$Km(n) \leq n-1 < K(m(n)+1) \quad (7.9)$$

de donde se desprende que $n-1-Km(n) < K$. Note ahora que C es la suma de $n-1-Km(n)$ variables consecutivas Ψ_i a las cuales se les ha restado su media, de manera que, utilizando el el Teorema 3.1 con $n-1-Km(n)+2$ en vez de n , se desprende que

$$\text{Var}[C] = \frac{16(n-1-Km(n)+2)-29}{90} < \frac{16(K+3)-29}{90} = \frac{16K+19}{90} < K.$$

Además, C se expresa en términos de $Y_{Km(n)}, Y_{Km(n)+1}, \dots, Y_n$, por lo que es independiente de todas las variables B_i ; vea (a)–(c). A partir de estas observaciones, (7.7) implica que

$$\begin{aligned} \text{Var}[V_n] &= \text{Var}[B_1 + B_2 + \dots + B_{m(n)-1} + C] \\ &= \sum_{i=1}^{m(n)-1} \text{Var}[B_i] + \text{Var}[C] \\ &\leq \sum_{i=1}^{m(n)-1} 1 + K < m(n) + K. \end{aligned}$$

□

Lema 7.2. Sea V_n como en el Lema 7.1. En este caso,

$$T_n^* - U_{m(n)}^* = \frac{V_n}{\sigma_n} + \left(\frac{\tilde{\sigma}_n}{\sigma_n} - 1 \right) U_{m(n)}^*,$$

donde

$$\sigma_n^2 = \text{Var}[T_n], \quad \text{y} \quad \tilde{\sigma}_n^2 = \text{Var}[U_n].$$

Demostración. Usando (7.6) se desprende que

$$\begin{aligned}
 \frac{V_n}{\sigma_n} &= \frac{T_n - E[T_n]}{\sigma_n} - \frac{U_{m(n)} - E[U_{m(n)}]}{\sigma_n} \\
 &= T_n^* - \frac{U_{m(n)} - E[U_{m(n)}]}{\sigma_n} \\
 &= T_n^* - \frac{\tilde{\sigma}_n U_{m(n)} - E[U_{m(n)}]}{\sigma_n \tilde{\sigma}_n} \\
 &= T_n^* - \frac{\tilde{\sigma}_n}{\sigma_n} U_{m(n)}^*
 \end{aligned}$$

de manera que

$$T_n^* - U_{m(n)}^* = \frac{V_n}{\sigma_n} + \left(\frac{\tilde{\sigma}_n}{\sigma_n} - 1 \right) U_{m(n)}^*.$$

□

El siguiente resultado es la última etapa previa a la demostración del Teorema 7.2.

Lema 7.3. Con la notación de los dos lemas previos,

$$(i) \quad \frac{m(n) + k}{\sigma_n^2} \leq \frac{6}{K} + \frac{6}{m(n)}$$

y

$$(ii) \quad \frac{\tilde{\sigma}_n^2}{\sigma_n^2} \rightarrow 1 - \frac{29}{16K} \quad \text{conforme } n \rightarrow \infty.$$

Demostración. (i) Recuerde que σ_n^2 , la varianza de T_n , esta dada por

$$\sigma_n^2 = \frac{16n - 29}{90};$$

vea el Teorema 3.1. Luego, debido a que $n > K > 3$,

$$\sigma_n^2 = \frac{16(n-1) - 13}{90} > \frac{15(n-1)}{90} = \frac{n-1}{6},$$

de manera que

$$\frac{m(n) + k}{\sigma_n^2} < 6 \frac{m(n) + k}{n-1}.$$

Para concluir, recuerde que $m(n)$ es la parte entera de $(n - 1)/K$, y entonces $n - 1 \geq m(n)K$, por lo que la anterior desigualdad desplegada implica que

$$\frac{m(n) + k}{\sigma_n^2} < 6 \frac{m(n) + k}{m(n)K} = \frac{6}{K} + \frac{6}{m(n)}.$$

(ii) Primero observe que la varianza de $U_{n(m)}$ puede evaluarse de la siguiente manera:

$$\begin{aligned} \text{Var} [U_{m(n)}] &= \text{Var} \left[\sum_{r=1}^{m(n)} A_r \right] \\ &= \sum_{r=1}^{m(n)} \text{Var} [A_r] \\ &= \sum_{r=1}^{m(n)} \frac{16K - 29}{90} \end{aligned}$$

donde se ha utilizado la independencia de las variables A_i , y la tercera igualdad se debe al Teorema 7.1. Luego,

$$\tilde{\sigma}_n^2 = \text{Var} [U_{m(n)}] = \frac{16m(n)K - m(n)29}{90}.$$

Por lo tanto, utilizando la expresión para σ_n^2 establecida en el Teorema 3.1, se desprende que, conforme n tiende a infinito,

$$\frac{\tilde{\sigma}_n^2}{\sigma_n^2} = \frac{16m(n)K - m(n)29}{16n - 29} \rightarrow 1 - \frac{29}{16K}, \quad (7.10)$$

donde se han utilizado las siguientes convergencias derivadas de (7.9):

$$\frac{m(n)K}{n} \rightarrow 1, \quad \text{y} \quad \frac{m(n)}{n} \rightarrow \frac{1}{K}.$$

Demostración del Teorema 7.2. Primero observe que para dos variables aleatorias W y Z ,

$$\text{Var} [W + Z] \leq 3(\text{Var} [W] + \text{Var} [Z]). \quad (7.11)$$

Para verificar esta igualdad, puede suponerse, sin pérdida de generalidad, que $\text{Var}[Z] \geq \text{Var}[W]$. En este caso,

$$\begin{aligned}
 \text{Var}[W + Z] &= \text{Var}[W] + \text{Var}[Z] + 2\text{Cov}[W, Z] \\
 &\leq \text{Var}[W] + \text{Var}[Z] + 2\sqrt{\text{Var}[Z]\text{Var}[W]} \\
 &\leq \text{Var}[W] + \text{Var}[Z] + 2\sqrt{\text{Var}[Z]\text{Var}[Z]} \\
 &\leq \text{Var}[W] + \text{Var}[Z] + 2\text{Var}[Z] \\
 &= \text{Var}[W] + 3\text{Var}[Z] \leq 3[\text{Var}[W] + \text{Var}[Z]]
 \end{aligned}$$

Para continuar, recuerde la igualdad establecida en el Lema 7.2:

$$T_n^* - U_{m(n)}^* = \frac{V_n}{\sigma_n} + \left(\frac{\tilde{\sigma}_n}{\sigma_n} - 1 \right) U_{m(n)}^*,$$

de tal manera que, utilizando (7.11), se obtiene que $D_n = T_n^* - U_{m(n)}^*$ satisface

$$\begin{aligned}
 \text{Var}[D_n] &\leq 3 \left[\text{Var} \left[\frac{V_n}{\sigma_n} \right] + \text{Var} \left[\left(\frac{\tilde{\sigma}_n}{\sigma_n} - 1 \right) U_{m(n)}^* \right] \right] \\
 &\leq 3 \frac{\text{Var}[V_n]}{\sigma_n^2} + 3 \left(\frac{\tilde{\sigma}_n}{\sigma_n} - 1 \right)^2 \text{Var} \left[U_{m(n)}^* \right] \quad (7.12) \\
 &\leq 3 \left[\frac{6}{K} + \frac{6}{m(n)} \right] + 3 \left(\frac{\tilde{\sigma}_n}{\sigma_n} - 1 \right)^2
 \end{aligned}$$

donde para obtener la última desigualdad se utilizó el Lemma 7.3, así como el hecho de que la varianza de $U_{m(n)}^*$ es la unidad. Observe ahora que a partir del Lemma 7.3(ii) se desprende que

$$\left| \frac{\tilde{\sigma}_n}{\sigma_n} - 1 \right| \left| \frac{\tilde{\sigma}_n}{\sigma_n} + 1 \right| = \left| \frac{\tilde{\sigma}_n^2}{\sigma_n^2} - 1 \right| \leq \frac{\Delta}{K}$$

donde Δ es una constante (vea también (7.10)). Por lo tanto,

$$\left| \frac{\tilde{\sigma}_n}{\sigma_n} - 1 \right| \leq \frac{\Delta}{K \left| \frac{\tilde{\sigma}_n}{\sigma_n} + 1 \right|} \leq \frac{\Delta}{K}$$

de manera que

$$\left(\frac{\tilde{\sigma}_n}{\sigma_n} - 1 \right)^2 \leq \frac{\Delta^2}{K^2} \leq \frac{\Delta^2}{K},$$

pues $K > 1$ (de hecho, K se supone mayor que 3). Combinando esta desigualdad con (7.12) se desprende que

$$\text{Var}[D_n] \leq \frac{18 + \Delta^2}{K} + \frac{18}{m(n)},$$

y la conclusión del Teorema 7.2 se obtiene definiendo

$$C_0 = 18 + \Delta^2, \quad \text{y} \quad \delta_n = \frac{18}{m(n)};$$

note que $\delta_n \rightarrow 0$ conforme $n \rightarrow \infty$. □

Demostración del Teorema 6.1

En esta sección se demuestra que la estandarización T_m^* de la variable aleatoria T_n tiene una distribución que converge a la distribución normal estándar; vea (6.1). Como punto de partida, observe que, con la notación en el Teorema 7.2, se tiene que

$$T_n^* = U_{m(n)}^* + D_n. \tag{8.1}$$

Considere ahora el evento $[T_n \leq a]$. Dado un número $\varepsilon > 0$, a partir de la ecuación (8.1) se desprende que

$$U_{m(n)}^* \leq a - \varepsilon \text{ y } D_n \leq \varepsilon \quad \Rightarrow \quad T_n \leq a,$$

de donde se desprende la siguiente inclusión:

$$[U_{m(n)}^* \leq a - \varepsilon] \cap [D_n \leq \varepsilon] \subset [T_n \leq a],$$

y por lo tanto,

$$P \left[[U_{m(n)}^* \leq a - \varepsilon] \cap [D_n \leq \varepsilon] \right] \leq P[T_n \leq a]. \tag{8.2}$$

Observe ahora que

$$\begin{aligned} P \left[[U_{m(n)}^* \leq a - \varepsilon] \cap [D_n \leq \varepsilon] \right] &= P[U_{m(n)}^* \leq a - \varepsilon] \\ &\quad - P \left[[U_{m(n)}^* \leq a - \varepsilon] \cap [D_n > \varepsilon] \right] \\ &\geq P[U_{m(n)}^* \leq a - \varepsilon] - P[D_n > \varepsilon], \end{aligned}$$

donde se ha utilizado que $P \left[[U_{m(n)}^* \leq a - \varepsilon] \cap [D_n > \varepsilon] \right] \leq P[D_n > \varepsilon]$ para obtener la segunda desigualdad. Luego, utilizando (8.2) se arriba a

$$P[T_n^* \leq a] \geq P[U_{m(n)}^* \leq a - \varepsilon] - P[D_n > \varepsilon]. \quad (8.3)$$

De manera similar, considere ahora el evento $[T_n^* > a - \varepsilon]$, y observe que partir de la ecuación (8.1) se desprende que

$$U_{m(n)}^* > a \text{ y } D_n > -\varepsilon \quad \Rightarrow \quad T_n^* > a - \varepsilon,$$

de donde se obtiene la siguiente inclusión:

$$[U_{m(n)}^* > a] \cap [D_n > -\varepsilon] \subset [T_n^* > a - \varepsilon],$$

y en consecuencia

$$P \left[[U_{m(n)}^* > a] \cap [D_n > -\varepsilon] \right] \leq P[T_n^* > a - \varepsilon]. \quad (8.4)$$

Ahora note que

$$\begin{aligned} P \left[[U_{m(n)}^* > a] \cap [D_n > -\varepsilon] \right] &= P[U_{m(n)}^* > a] - P \left[[U_{m(n)}^* > a] \cap [D_n \leq -\varepsilon] \right] \\ &\geq P[U_{m(n)}^* > a] - P[D_n \leq -\varepsilon] \end{aligned}$$

donde la segunda desigualdad se desprende de

$$P \left[[U_{m(n)}^* > a] \cap [D_n \leq -\varepsilon] \right] \leq P[D_n \leq -\varepsilon].$$

Combinando este hecho con (8.4) se concluye que $P[T_n^* > a - \varepsilon] \geq P[U_{m(n)}^* > a] - P[D_n \leq -\varepsilon]$, desigualdad que equivale a

$$P[T_n^* \leq a - \varepsilon] \leq P[U_{m(n)}^* \leq a] + P[D_n \leq -\varepsilon];$$

en este argumento, a es un número arbitrario, y reemplazando a por $a + \varepsilon$, se obtiene

$$P[T_n^* \leq a] \leq P[U_{m(n)}^* \leq a + \varepsilon] + P[D_n \leq -\varepsilon]. \quad (8.5)$$

Después de este preámbulo, el resultado sobre la distribución asintótica de T_n^* puede demostrarse de la siguiente forma.

Demostración del Teorema 6.1. Primero observe que la variable aleatoria D_n en el Teorema 7.1 tiene esperanza nula, y note que a partir de la desigualdad de Chebychev se obtiene

$$P[D_n < -\varepsilon] + P[D_n > \varepsilon] = P[|D_n| > \varepsilon] \leq \frac{\text{Var}[D_n]}{\varepsilon^2} \leq \frac{1}{\varepsilon^2} \left[\frac{C_0}{K} + \delta_n \right].$$

Por lo tanto,

$$\lim_{n \rightarrow \infty} [P[D_n < -\varepsilon] + P[D_n > \varepsilon]] \leq \lim_{n \rightarrow \infty} \frac{1}{\varepsilon^2} \left[\frac{C_0}{K} + \delta_n \right] = \frac{C_0}{\varepsilon^2 K}, \quad (8.6)$$

donde se ha utilizado que $\lim_{n \rightarrow \infty} \delta_n = 0$; vea el Teorema 7.1. Para continuar, observe que $\lim_{n \rightarrow \infty} m(n) = \infty$, hecho que se desprende directamente de la definición de $m(n)$ como la parte entera de $(n-1)/K$. Por lo tanto, el Corolario 7.1 implica que, para todo número real b ,

$$\lim_{n \rightarrow \infty} P[U_{m(n)}^* \leq b] = \Phi(b). \quad (8.7)$$

A continuación, combine (8.3) y (8.5) para obtener, para todo $a \in \mathbb{R}$ y $\varepsilon > 0$,

$$P[U_{m(n)}^* \leq a - \varepsilon] - P[D_n > \varepsilon] \leq P[T_n^* \leq a] \leq P[U_{m(n)}^* \leq a + \varepsilon] + P[D_n \leq -\varepsilon],$$

y tomando límite conforme n crece a infinito, (8.6) y (8.7) implican

$$\Phi(a - \varepsilon) - \frac{C_0}{K\varepsilon^2} \leq \lim_{n \rightarrow \infty} P[T_n^* \leq a] \leq \Phi(a + \varepsilon) + \frac{C_0}{K\varepsilon^2};$$

en esta relación K es un entero mayor a 3, y ε es un número positivo arbitrario, Tomando límite conforme K tiende a infinito, se arriba a

$$\Phi(a - \varepsilon) \leq \lim_{n \rightarrow \infty} P[T_n^* \leq a] \leq \Phi(a + \varepsilon),$$

y después de tomar el límite cuando ε decrece a cero, se obtiene, usando la continuidad de la función de distribución $\Phi(\cdot)$,

$$\Phi(a) \leq \lim_{n \rightarrow \infty} P[T_n^* \leq a] \leq \Phi(a),$$

relación que equivale a

$$\lim_{n \rightarrow \infty} P[T_n^* \leq a] = \Phi(a),$$

y puesto que $a \in \mathbb{R}$ es arbitrario, esta convergencia equivale a $T_n^* \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$, concluyendo la demostración de Teorema 6.1. \square

El Procedimiento de Prueba

El resultado sobre la normalidad asintótica del estadístico estandarizado T_n^* presentado en el Teorema 6.1, permite establecer un procedimiento de prueba para la Hipótesis 2.1, de acuerdo a la cual las observaciones Y_1, Y_2, \dots, Y_n son independientes y tiene una distribución común, la cual es continua.

Dado un número $\alpha \in [0, 1]$, de acuerdo al Teorema 6.1, se tiene que

$$\lim_{n \rightarrow \infty} P[-z_{\alpha/2} < T_n^* < z_{\alpha/2}] = \Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2}) = 1 - \alpha,$$

donde $z_{\alpha/2}$ es el percentil superior de orden $\alpha/2$ de la distribución normal estándar. De acuerdo a esta convergencia, siempre que la cantidad de datos n sea ‘grande’, la probabilidad del evento $[-z_{\alpha/2} < T_n^* < z_{\alpha/2}]$ es, aproximadamente, $1 - \alpha$; seleccionando α como un número ‘pequeño’ se tiene que T_n^* se ubicará en el intervalo $[-z_{\alpha/2}, z_{\alpha/2}]$ con una probabilidad ‘cercana’ a 1, de manera que el observador considerará como un hecho normal el que la siguiente inclusión ocurra:

$$T_n^* \in [-z_{\alpha/2}, z_{\alpha/2}]. \quad (9.1)$$

Sin embargo, cuando esta inclusión falle, el observador sospechara que el supuesto que conduce a considerar a (9.1) como ‘normal’, a saber la Hipótesis 2.1, no se satisface. Luego, si (9.1) falla, este hecho será atribuido a alguna de las siguientes razones:

(i) Los datos observados y_1, y_2, \dots, y_n no son realizaciones de una sucesión de variables independientes;

(ii) Los datos y_1, y_2, \dots, y_n no son los valores asumidos por variables aleatorias Y_1, Y_2, \dots, Y_n con una distribución continua común, o

(iii) Los datos y_1, y_2, \dots, y_n si son los valores que toman por variables aleatorias Y_1, Y_2, \dots, Y_n independientes e idénticamente distribuidas, pero su distribución común no es continua.

En la práctica, especialmente cuando los datos y_1, y_2, \dots, y_n se piensan como los valores asumidos por variables aleatorias Y_1, Y_2, \dots, Y_n para las cuales se desea validar la hipótesis de que conforman un ruido blanco, frecuentemente se esta dispuesto a asumir sin duda alguna que las Y_i tienen densidad, así que si (9.1) no se observa, se atribuirá el hecho a alguna de las dos primeras razones enunciadas anteriormente. Esta discusión permite enunciar el siguiente procedimiento de prueba:

- Suponga que cada una de las variables aleatorias Y_1, Y_2, \dots, Y_n tiene densidad, y considere la siguiente hipótesis:

$$\mathcal{H}_0: Y_1, Y_2, \dots, Y_n \text{ son } i.i.d.$$

Sea T_n^* la versión estandarizada del estadístico de máximos y mínimos locales introducido en la Definición 2.1. En este caso, una prueba de la hipótesis \mathcal{H}_0 con nivel de significancia (asintótico) α está determinada como sigue:

$$\text{Rechace } \mathcal{H}_0 \text{ si } T_n^* \notin [z_{\alpha/2}, z_{\alpha/2}].$$

Este método de prueba de la hipótesis \mathcal{H}_0 , tiene un nivel de significancia aproximadamente igual a α , a condición de que el número de datos, esto es, n , sea ‘grande’.

Conclusión

En este capítulo se ha analizado la distribución límite del estadístico de máximos y mínimos locales asociado a una sucesión Y_1, Y_2, \dots, Y_n . El análisis se

desarrollo bajo el supuesto de que las variables aleatorias Y_k tienen la misma distribución continua y son independientes. Después de definir de manera rigurosa la construcción de T_n en la Sección 2, se establecieron fórmulas para el valor esperado y la varianza de T_n en la Sección 3. El argumento utilizado para demostrar dichas fórmulas requirió un cálculo cuidadoso de las probabilidades de que los máximos o mínimos locales se localicen en instantes de tiempo separados por una o dos observaciones. El cómputo de dichas probabilidades utilizó expresiones integrales y fue presentado en la Sección 4. En este punto del trabajo, es natural preguntarse por qué no se utilizaron argumentos de simetría para realizar los cálculos de la Sección 4 en lugar de los, aparentemente, más complicados argumentos en base a Cálculo Integral. La razón es que, con excepción de los cálculos en el Lemma 3.1, no parece ser menos complicado utilizar argumentos de simetría que realizar el cómputo de las probabilidades relevantes de manera directa; este hecho se confirma con los valores obtenidos para las esperanzas establecidas en los Lemas de la Sección 4. Un aspecto sobresaliente de los resultados de este capítulo, es que se logró establecer, de forma rigurosa y con argumentos elementales, la normalidad asintótica del estadístico T_n , el cual es una suma de variables aleatorias que, aunque poseen la misma distribución, *no son independientes*.

Capítulo 4

Cálculo Recursivo de Pronósticos

Este capítulo trata sobre el problema de construir pronósticos en el análisis de una serie de tiempo estacionaria. De manera más específica, el principal objetivo es mostrar como los cálculos que se realizan para elaborar un pronóstico en un tiempo n , pueden simplificarse si se hace un uso adecuado de los datos obtenidos al realizar un pronóstico en el momento $n - 1$. El resultado central que se presenta es el algoritmo recursivo de Durbin–Levinson, y la principal contribución técnica derivada del análisis subsecuente consiste en establecer, de manera explícita, una caracterización, en términos geométricos, de la solución de un sistema de ecuaciones normales.

Introducción

Este capítulo trata sobre el problema de construir un pronóstico para una observación futura en una serie de tiempo estacionaria. 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. El problema se aborda desde una perspectiva algebraica, y el principal objetivo es proporcionar una demostración rigurosa del algoritmo recursivo de Durbin–Levinson, el cual es un instrumento poderoso para calcular pronósticos.

El estudio que se realiza a continuación tiene como primera etapa una análisis del sistema de ecuaciones normales, y la principal contribución consiste en establecer, de manera explícita y con argumentos de naturaleza geométrica, una

caracterización de las componentes de una solución del sistema, la cual es la base para demostrar, posteriormente, el método recursivo de Durbin–Levinson.

La organización del capítulo es la siguiente: En la Sección 2 se plantea el problema de determinar la mejor aproximación a un vector dado en términos de una combinación lineal de un conjunto determinado de vectores, y se muestra como se arriba al sistema de ecuaciones normales, mientras que en la Sección 3 se determina el correspondiente error de aproximación. La Sección 4 contiene el resultado central, de acuerdo al cual un coeficiente que forma parte de la solución a un sistema de n ecuaciones normales, se expresa en términos de una solución a un sistema que consiste de $n - 1$ ecuaciones. La Sección 5 analiza la interpretación de los resultados precedentes en términos estadísticos, mientras que en la Sección 6 se establece el algoritmo recursivo de Durbin–Levinson. La demostración de este resultado se presenta en la Sección 8, después de establecer los resultados técnicos preliminares en la Sección 7. En la Sección 9 se presenta un ejemplo, cuidadosamente desarrollado, sobre la aplicación de dicho método recursivo, y la exposición concluye en la Sección 10 con algunos comentarios breves. Por otro lado, en el desarrollo del capítulo se hace un uso intensivo de las propiedades fundamentales de un producto interno en un espacio vectorial abstracto, las cuales pueden encontrarse, por ejemplo, en Hoffman y Kunze (1973), Harville (1997), o Searle (1982).

La Mejor Aproximación

Sean \mathbf{w} y $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ vectores en un espacio vectorial \mathcal{V} dotado con un producto interno, el cual se denotará mediante $\langle \cdot, \cdot \rangle$. En este caso, 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

$$\|\mathbf{t}\|^2 = \langle \mathbf{t}, \mathbf{t} \rangle. \quad (2.1)$$

El problema que se aborda en esta sección, consiste en obtener la mejor aproxi-

mación a \mathbf{w} en términos de los vectores $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$, propósito que se plantea formalmente a continuación:

Determine constantes c_1, c_2, \dots, c_n tales que

$$\left\| \mathbf{w} - \sum_{i=1}^n c_i \mathbf{v}_i \right\|^2 = \min_{\beta_1, \beta_2, \dots, \beta_n} \left\| \mathbf{w} - \sum_{i=1}^n \beta_i \mathbf{v}_i \right\|^2. \quad (2.2)$$

Este problema algebraico general tiene interpretaciones importantes dentro de la teoría estadística, y el análisis que se realizará líneas más adelante dará cuenta de ello. Como punto de partida, se caracterizarán las constantes c_1, c_2, \dots, c_n que resuelven el problema (2.2).

Teorema 2.1. Si c_1, c_2, \dots, c_n es una solución del problema (2.2), entonces las siguientes n ecuaciones lineales se satisfacen.

$$\sum_{j=1}^n \langle \mathbf{v}_k, \mathbf{v}_j \rangle c_j = \langle \mathbf{w}, \mathbf{v}_k \rangle, \quad k = 1, 2, \dots, n. \quad (2.3)$$

Demostración. Usando la bilinealidad del producto interno, se desprende que

$$\begin{aligned} \left\| \mathbf{w} - \sum_{i=1}^n \beta_i \mathbf{v}_i \right\|^2 &= \left\langle \mathbf{w} - \sum_{i=1}^n \beta_i \mathbf{v}_i, \mathbf{w} - \sum_{j=1}^n \beta_j \mathbf{v}_j \right\rangle \\ &= \langle \mathbf{w}, \mathbf{w} \rangle - \left\langle \sum_{i=1}^n \beta_i \mathbf{v}_i, \mathbf{w} \right\rangle \\ &\quad - \left\langle \mathbf{w}, \sum_{j=1}^n \beta_j \mathbf{v}_j \right\rangle + \left\langle \sum_{j=1}^n \beta_j \mathbf{v}_j, \sum_{j=1}^n \beta_j \mathbf{v}_j \right\rangle \\ &= \langle \mathbf{w}, \mathbf{w} \rangle - \sum_{i=1}^n \beta_i \langle \mathbf{v}_i, \mathbf{w} \rangle \\ &\quad - \sum_{j=1}^n \beta_j \langle \mathbf{w}, \mathbf{v}_j \rangle + \sum_{i=1}^n \sum_{j=1}^n \beta_i \beta_j \langle \mathbf{v}_i, \mathbf{v}_j \rangle \end{aligned}$$

A partir de esta expresión, la derivada de $\left\| \mathbf{w} - \sum_{i=1}^n \beta_i \mathbf{v}_i \right\|^2$ respecto a β_k puede

determinarse como sigue:

$$\begin{aligned}
\frac{d}{d\beta_k} \left\| \mathbf{w} - \sum_{i=1}^n \beta_i \mathbf{v}_i \right\|^2 &= \frac{d}{d\beta_k} \langle \mathbf{w}, \mathbf{w} \rangle - \sum_{i=1}^n \frac{d}{d\beta_k} \beta_i \langle \mathbf{v}_i, \mathbf{w} \rangle \\
&\quad - \sum_{j=1}^n \frac{d}{d\beta_k} \beta_j \langle \mathbf{w}, \mathbf{v}_j \rangle + \sum_{i=1}^n \sum_{j=1}^n \frac{d}{d\beta_k} \beta_i \beta_j \langle \mathbf{v}_i, \mathbf{v}_j \rangle \\
&= 0 - \langle \mathbf{v}_k, \mathbf{w} \rangle \\
&\quad - \langle \mathbf{w}, \mathbf{v}_k \rangle + \sum_{i=1}^n \sum_{j=1}^n \frac{d}{d\beta_k} \beta_i \beta_j \langle \mathbf{v}_i, \mathbf{v}_j \rangle
\end{aligned} \tag{2.4}$$

donde se ha utilizado que $\frac{d}{d\beta_k} \beta_s = 0$ para $s \neq k$, mientras que $\frac{d}{d\beta_k} \beta_k = 1$. Por otro lado, observe que

$$\frac{d}{d\beta_k} \beta_i \beta_j = \begin{cases} \beta_j, & \text{si } i = k \text{ y } j \neq k; \\ \beta_i, & \text{si } j = k \text{ y } i \neq k; \\ 2\beta_k, & \text{si } i = j = k \end{cases}$$

y combinando esta igualdad con (2.4) se desprende que

$$\begin{aligned}
\frac{d}{d\beta_k} \left\| \mathbf{w} - \sum_{i=1}^n \beta_i \mathbf{v}_i \right\|^2 &= -\langle \mathbf{v}_k, \mathbf{w} \rangle - \langle \mathbf{w}, \mathbf{v}_k \rangle \\
&\quad + \sum_{j:j \neq k} \beta_j \langle \mathbf{v}_k, \mathbf{v}_j \rangle + \sum_{i:i \neq k} \beta_i \langle \mathbf{v}_i, \mathbf{v}_k \rangle + 2\beta_k \langle \mathbf{v}_k, \mathbf{v}_k \rangle.
\end{aligned}$$

Finalmente, la propiedad de simetría del producto interno, esto es, $\langle \mathbf{a}, \mathbf{b} \rangle = \langle \mathbf{b}, \mathbf{a} \rangle$ para $\mathbf{a}, \mathbf{b} \in \mathcal{V}$, implica que

$$\begin{aligned}
\frac{d}{d\beta_k} \left\| \mathbf{w} - \sum_{i=1}^n \beta_i \mathbf{v}_i \right\|^2 &= -2\langle \mathbf{w}, \mathbf{v}_k \rangle + \sum_{j:j \neq k} \beta_j \langle \mathbf{v}_j, \mathbf{v}_k \rangle \\
&\quad + \sum_{i:i \neq k} \beta_i \langle \mathbf{v}_i, \mathbf{v}_k \rangle + 2\beta_k \langle \mathbf{v}_k, \mathbf{v}_k \rangle. \\
&= -2\langle \mathbf{w}, \mathbf{v}_k \rangle + 2 \sum_{j:j \neq k} \beta_j \langle \mathbf{v}_j, \mathbf{v}_k \rangle + 2\beta_k \langle \mathbf{v}_k, \mathbf{v}_k \rangle. \\
&= -2\langle \mathbf{w}, \mathbf{v}_k \rangle + 2 \sum_{j=1}^n \langle \mathbf{v}_k, \mathbf{v}_j \rangle \beta_j
\end{aligned}$$

Para concluir, observe que si las constantes c_1, c_2, \dots, c_n resuelven el problema (2.2), necesariamente debe tenerse que $\frac{d}{d\beta_k} \left\| \mathbf{w} - \sum_{i=1}^n \beta_i \mathbf{v}_i \right\|^2_{\beta_1=c_1, \dots, \beta_n=c_n} = 0$, de

manera que la expresión anterior implica

$$0 = -2\langle \mathbf{w}, \mathbf{v}_k \rangle + 2 \sum_{j=1}^n \langle \mathbf{v}_k, \mathbf{v}_j \rangle c_j,$$

de donde se obtiene que

$$\sum_{j=1}^n \langle \mathbf{v}_k, \mathbf{v}_j \rangle c_j = \langle \mathbf{w}, \mathbf{v}_k \rangle,$$

lo cual concluye la demostración de (2.3), pues $k \in \{1, 2, \dots, n\}$ es arbitrario. \square

El sistema (2.3) permite encontrar la combinación lineal de los vectores \mathbf{v}_k , $k = 1, 2, \dots, n$, que dista lo menos posible de \mathbf{w} , y es importante determinar en que casos dicho sistema es consistente, i.e., cuando (2.3) tiene solución, y las circunstancias bajo las cuales ésta es única.

Teorema 2.2. (i) El sistema (2.3) siempre es consistente.

(ii) La solución de (2.3) es única si y sólo si v_1, v_2, \dots, v_n son independientes.

(iii) Si c_1, c_2, \dots, c_n y d_1, d_2, \dots, d_n satisfacen (2.3), entonces

$$\sum_{i=1}^n c_i \mathbf{v}_i = \sum_{i=1}^n d_i \mathbf{v}_i.$$

De acuerdo a este Teorema, (2.3) siempre tiene solución, propiedad que está acorde con el hecho, claro desde el punto de vista geométrico, de que siempre es posible encontrar dentro del espacio vectorial generado por $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ un vector cuya distancia a \mathbf{w} sea lo menor posible.

Demostración del Teorema 2.2. Sean $\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_n$ las filas de la matriz principal del sistema (2.3). Para verificar que el sistema (2.3) es consistente, es suficiente comprobar que

$$a_1 \mathbf{f}_1 + a_2 \mathbf{f}_2 + \dots + a_n \mathbf{f}_n = 0 \implies a_1 \langle \mathbf{w}, \mathbf{v}_1 \rangle + a_2 \langle \mathbf{w}, \mathbf{v}_2 \rangle + \dots + a_n \langle \mathbf{w}, \mathbf{v}_n \rangle = 0. \quad (2.5)$$

Para validar esta implicación, observe que la k -ésima fila de (2.3) es

$$\mathbf{f}_k = (\langle \mathbf{v}_k, \mathbf{v}_1 \rangle, \langle \mathbf{v}_k, \mathbf{v}_2 \rangle, \dots, \langle \mathbf{v}_k, \mathbf{v}_n \rangle),$$

de manera que

$$\begin{aligned}
 \sum_{k=1}^n a_k \mathbf{f}_k &= \sum_{k=1}^n a_k (\langle \mathbf{v}_k, \mathbf{v}_1 \rangle, \langle \mathbf{v}_k, \mathbf{v}_2 \rangle, \dots, \langle \mathbf{v}_k, \mathbf{v}_n \rangle) \\
 &= \left(\sum_{k=1}^n a_k \langle \mathbf{v}_k, \mathbf{v}_1 \rangle, \sum_{k=1}^n a_k \langle \mathbf{v}_k, \mathbf{v}_2 \rangle, \dots, \sum_{k=1}^n a_k \langle \mathbf{v}_k, \mathbf{v}_n \rangle \right) \\
 &= \left(\left\langle \sum_{k=1}^n a_k \mathbf{v}_k, \mathbf{v}_1 \right\rangle, \left\langle \sum_{k=1}^n a_k \mathbf{v}_k, \mathbf{v}_2 \right\rangle, \dots, \left\langle \sum_{k=1}^n a_k \mathbf{v}_k, \mathbf{v}_n \right\rangle \right)
 \end{aligned}$$

donde se ha utilizado la linealidad del producto interno. Luego,

$$\begin{aligned}
 \sum_{k=1}^n a_k \mathbf{f}_k = \mathbf{0} &\implies \left(\left\langle \sum_{k=1}^n a_k \mathbf{v}_k, \mathbf{v}_1 \right\rangle, \left\langle \sum_{k=1}^n a_k \mathbf{v}_k, \mathbf{v}_2 \right\rangle, \dots, \left\langle \sum_{k=1}^n a_k \mathbf{v}_k, \mathbf{v}_n \right\rangle \right) = \mathbf{0} \\
 &\implies \left\langle \sum_{k=1}^n a_k \mathbf{v}_k, \mathbf{v}_i \right\rangle \text{ para todo } i \\
 &\implies \sum_{i=1}^n a_i \left\langle \sum_{k=1}^n a_k \mathbf{v}_k, \mathbf{v}_i \right\rangle = 0 \\
 &\implies \left\langle \sum_{k=1}^n a_k \mathbf{v}_k, \sum_{i=1}^n a_i \mathbf{v}_i \right\rangle = 0 \\
 &\implies \left\| \sum_{k=1}^n a_k \mathbf{v}_k \right\|^2 = 0
 \end{aligned}$$

y como $\mathbf{0}$ es el único vector en \mathcal{V} con norma nula, se desprende que

$$\begin{aligned}
 \sum_{k=1}^n a_k \mathbf{f}_k = \mathbf{0} &\implies \sum_{k=1}^n a_k \mathbf{v}_k = \mathbf{0} \\
 &\implies \left\langle \mathbf{w}, \sum_{k=1}^n a_k \mathbf{v}_k \right\rangle = \langle \mathbf{w}, \mathbf{0} \rangle = 0 \\
 &\implies \sum_{k=1}^n a_k \langle \mathbf{w}, \mathbf{v}_k \rangle = 0
 \end{aligned}$$

lo cual comprueba la implicación (2.5).

(ii) Para establecer esta parte se utilizará el siguiente hecho: (2.3) tiene solución única si y sólo si el sistema homogéneo correspondiente tiene sólo la solución nula. De esta manera, el análisis inicia estudiando el sistema homogéneo asociado

a (2.3), el cual está dado por

$$\sum_{i=1}^n \langle \mathbf{v}_k, \mathbf{v}_i \rangle x_i = 0 \quad k = 1, 2, \dots, n$$

o equivalentemente,

$$\langle \mathbf{v}_k, \sum_{i=1}^n x_i \mathbf{v}_i \rangle = 0 \quad \text{para todo } k. \quad (2.6)$$

Observe ahora que

$$\begin{aligned} (2.6) \text{ ocurre} &\implies \sum_{k=0}^n x_k \langle \mathbf{v}_k, \sum_{i=1}^n x_i \mathbf{v}_i \rangle = 0 \\ &\implies \langle \sum_{k=0}^n x_k \mathbf{v}_k, \sum_{i=1}^n x_i \mathbf{v}_i \rangle = 0 \\ &\implies \left\| \sum_{i=0}^n x_i \mathbf{v}_i \right\|^2 = 0 \\ &\implies \sum_{k=0}^n x_k \mathbf{v}_k = 0; \end{aligned}$$

recíprocamente,

$$\sum_{k=0}^n x_k \mathbf{v}_k = 0 \implies \langle \mathbf{v}_k, \sum_{i=1}^n x_i \mathbf{v}_i \rangle = 0 \text{ para todo } k$$

Esta discusión puede demostrarse la siguiente equivalencia:

$$\langle \mathbf{v}_k, \sum_{i=1}^n x_i \mathbf{v}_i \rangle = 0 \text{ para todo } k \iff \sum_{k=0}^n x_k \mathbf{v}_k = 0. \quad (2.7)$$

La conclusión puede alcanzarse ahora como sigue: El sistema (2.6) tiene sólo la solución nula si y sólo si $\sum_{k=0}^n x_k \mathbf{v}_k = 0$ implica que $x_i = 0$ para todo i , lo cual equivale a la independencia de $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$.

(iii) Suponga que c_1, c_2, \dots, c_n y d_1, d_2, \dots, d_n satisfacen el sistema (2.3), esto es, para todo $k = 1, 2, \dots, n$,

$$\sum_{i=1}^n \langle \mathbf{v}_k, \mathbf{v}_i \rangle c_i = \langle \mathbf{w}, \mathbf{v}_k \rangle, \quad \text{y} \quad \sum_{i=1}^n \langle \mathbf{v}_k, \mathbf{v}_i \rangle d_i = \langle \mathbf{w}, \mathbf{v}_k \rangle.$$

En este caso, se desprende que

$$\sum_{i=1}^n \langle \mathbf{v}_k, \mathbf{v}_i \rangle (c_i - d_i) = \sum_{i=1}^n \langle \mathbf{v}_k, (c_i - d_i) \mathbf{v}_i \rangle = 0, \quad k = 1, 2, \dots, n,$$

y definiendo $x_k = c_k - d_k$, (2.7) implica, que $\sum_{k=0}^n (c_i - d_i) \mathbf{v}_i = 0$, lo cual equivale a $\sum_{k=0}^n c_i \mathbf{v}_i = \sum_{k=0}^n d_i \mathbf{v}_i$. \square

De acuerdo al teorema precedente, aún en el caso en que el sistema (2.3) no tenga solución única, el vector que se obtiene al combinar los vectores $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ utilizando una solución particular, es siempre el mismo; dicha combinación es el vector más cercano a \mathbf{w} dentro del espacio generado por $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$.

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

$$P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w} \equiv P_{\mathcal{L}} \mathbf{w} = \sum_{i=1}^n c_i \mathbf{v}_i,$$

donde c_1, c_2, \dots, c_n son constantes que satisfacen (2.3).

A partir del Teorema 2.1 y de la parte (iii) del Teorema 2.2, se desprende la siguiente caracterización de una proyección.

Corolario 2.1. El vector $P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}$ satisface las siguientes condiciones (i)–(ii) para todo $k = 1, 2, \dots, n$:

$$(i) \langle P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}, \mathbf{v}_k \rangle = \langle \mathbf{w}, \mathbf{v}_k \rangle;$$

$$(ii) \langle \mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}, \mathbf{v}_k \rangle = 0.$$

Más aún,

(iii) Si $\mathbf{v}^* \in \mathcal{V}$ satisface $\langle \mathbf{w} - \mathbf{v}^*, \mathbf{v}_k \rangle = 0$ para todo k , o bien $\langle \mathbf{v}^*, \mathbf{v}_k \rangle = \langle \mathbf{w}, \mathbf{v}_k \rangle$, entonces $\mathbf{v}^* = P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}$.

De acuerdo a este corolario, la proyección de \mathbf{w} sobre el espacio vectorial generado por $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$, es el único vector \mathbf{v}^* tal que $\mathbf{w} - \mathbf{v}^*$ tiene

un producto interno nulo con cada vector \mathbf{v}_k . Recordando que dos vectores $\mathbf{a}, \mathbf{b} \in \mathcal{V}$ son normales (perpendiculares, ortogonales), si $\langle \mathbf{a}, \mathbf{b} \rangle = 0$, puede establecerse que $P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}$ es el único vector de \mathcal{V} con la propiedad de que $\mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}$ es *normal* a cada \mathbf{v}_k . Note que, usando la Definición 2.1, las igualdades $\langle \mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}, \mathbf{v}_k \rangle = 0$ para todo k , son equivalentes a (2.3); por esta razón las ecuaciones (2.3) se conocen como *el sistema de las ecuaciones normales* (Graybill, 1985).

El Error de Aproximación

En la sección anterior se mostró que determinar la ‘mejor’ aproximación a un vector \mathbf{w} en términos de una combinación lineal de vectores dados $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ se reduce a encontrar la solución del sistema de ecuaciones normales (2.3). Sin embargo, aún hay un problema importante que resolver, a saber, determinar el error que se comete al aproximar \mathbf{w} por medio de $P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}$. Dicho error de aproximación es $\|\mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}\|$, y el siguiente resultado permite calcularlo fácilmente una vez que las ecuaciones normales han sido resueltas.

Teorema 3.1. El error de aproximación $\|\mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}\|$ está determinado por cualquiera de las siguientes expresiones (i)–(iii):

$$(i) \quad \|\mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}\|^2 = \|\mathbf{w}\|^2 - \|P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}\|^2;$$

$$(ii) \quad \|\mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}\|^2 = \|\mathbf{w}\|^2 - \sum_{i=1}^n \sum_{j=1}^n c_i \langle \mathbf{v}_i, \mathbf{v}_j \rangle c_j.$$

Más aún,

$$(iii) \quad \|\mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}\|^2 = \|\mathbf{w}\|^2 - \sum_{i=1}^n c_i \langle \mathbf{w}, \mathbf{v}_i \rangle. \quad (3.1)$$

Demostración. (i) Primero observe que

$$\langle P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}, \mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w} \rangle = 0. \quad (3.2)$$

En efecto, de acuerdo al Corolario 2.1, $\langle \mathbf{v}_k, \mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w} \rangle = 0$ para todo k , de manera que

$$\sum_{k=1}^n c_k \langle \mathbf{v}_k, \mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w} \rangle = \left\langle \sum_{k=1}^n c_k \mathbf{v}_k, \mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w} \right\rangle = 0,$$

y seleccionado c_1, c_2, \dots, c_n como una solución del sistema de ecuaciones normales, se desprende, via la Definición 2.1, que $\langle P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}, \mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w} \rangle = 0$.

Note, además, que (3.2) implica que

$$\langle P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}, \mathbf{w} \rangle = \langle P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}, P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w} \rangle. \quad (3.3)$$

Después de esta discusión, la parte (i) puede demostrarse como sigue:

$$\begin{aligned} \|\mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}\|^2 &= \langle \mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}, \mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w} \rangle \\ &= \langle \mathbf{w}, \mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w} \rangle \\ &\quad - \langle P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}, \mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w} \rangle \\ &= \langle \mathbf{w}, \mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w} \rangle \end{aligned}$$

donde (3.2) se utilizó para obtener la última igualdad. Luego,

$$\begin{aligned} \|\mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}\|^2 &= \langle \mathbf{w}, \mathbf{w} \rangle - \langle \mathbf{w}, P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w} \rangle \\ &= \langle \mathbf{w}, \mathbf{w} \rangle - \langle P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}, P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w} \rangle \\ &= \|\mathbf{w}\|^2 - \|P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}\|^2 \end{aligned}$$

donde la segunda igualdad se desprende de (3.3).

(ii) Observe que la norma de $P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n}$ está dada por

$$\begin{aligned} \|P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}\|^2 &= \langle P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}, P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w} \rangle \\ &= \left\langle \sum_{i=1}^n c_i \mathbf{v}_i, \sum_{j=1}^n c_j \mathbf{v}_j \right\rangle \\ &= \sum_{i=1}^n \sum_{j=1}^n c_i \langle \mathbf{v}_i, \mathbf{v}_j \rangle c_j \end{aligned}$$

y aplicando la parte (i) se obtiene

$$\begin{aligned}\|\mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}\|^2 &= \|\mathbf{w}\|^2 - \|P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}\|^2 \\ &= \|\mathbf{w}\|^2 - \sum_{i=1}^n \sum_{j=1}^n c_i \langle \mathbf{v}_i, \mathbf{v}_j \rangle c_j\end{aligned}$$

(iii) A partir de las ecuaciones normales, se desprende que, para cada i ,

$$\sum_{j=1}^n \langle \mathbf{v}_i, \mathbf{v}_j \rangle c_j = \langle \mathbf{w}, \mathbf{v}_i \rangle;$$

luego,

$$\begin{aligned}\sum_{i=1}^n \sum_{j=1}^n c_i \langle \mathbf{v}_i, \mathbf{v}_j \rangle c_j &= \sum_{i=1}^n c_i \left[\sum_{j=1}^n \langle \mathbf{v}_i, \mathbf{v}_j \rangle c_j \right] \\ &= \sum_{i=1}^n c_i \langle \mathbf{w}, \mathbf{v}_i \rangle\end{aligned}$$

y la conclusión se obtiene combinando esta igualdad con la parte (ii); en efecto,

$$\begin{aligned}\|\mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}\|^2 &= \|\mathbf{w}\|^2 - \sum_{i=1}^n \sum_{j=1}^n c_i \langle \mathbf{v}_i, \mathbf{v}_j \rangle c_j \\ &= \|\mathbf{w}\|^2 - \sum_{i=1}^n c_i \langle \mathbf{w}, \mathbf{v}_i \rangle\end{aligned}$$

□

Observe que las expresiones en el Teorema 3.1 permiten obtener el error de aproximación a partir de la solución c_1, c_2, \dots, c_n de las ecuaciones normales; de particular interés es la fórmula (3.1), la cual determina el error de aproximación restando de $\|\mathbf{w}\|^2$ la cantidad que se obtiene multiplicando la constante c_i por el lado derecho de la i -ésima ecuación normal, y sumando los resultados. Estas ideas se ilustran a continuación.

Ejemplo 3.1. Considere los siguientes vectores en $\mathcal{V} = \mathbb{R}^4$:

$$\mathbf{w} = \begin{bmatrix} 8 \\ 2 \\ 2 \\ 8 \end{bmatrix}, \quad \mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix}.$$

El problema que se considera a continuación consiste en determinar, tanto la proyección $P_{\mathbf{v}_1, \mathbf{v}_2} \mathbf{w}$, como el correspondiente error de aproximación, el cual está dado por $\|\mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2} \mathbf{w}\|$.

• El sistema de ecuaciones normales está determinada como sigue: La primera de las ecuaciones es

$$\langle \mathbf{v}_1, \mathbf{v}_1 \rangle c_1 + \langle \mathbf{v}_1, \mathbf{v}_2 \rangle c_2 = \langle \mathbf{w}, \mathbf{v}_1 \rangle,$$

la cual se reduce a

$$4c_1 + 4c_2 = 20.$$

La segunda ecuación en el sistema (2.3) es

$$\langle \mathbf{v}_2, \mathbf{v}_1 \rangle c_1 + \langle \mathbf{v}_2, \mathbf{v}_2 \rangle c_2 = \langle \mathbf{w}, \mathbf{v}_2 \rangle,$$

igualdad que para los datos actuales se convierte en

$$4c_1 + 6c_2 = 26.$$

Luego, el sistema de ecuaciones normales es, en el caso presente,

$$\begin{aligned} 4c_1 + 4c_2 &= 20 \\ 4c_1 + 6c_2 &= 26 \end{aligned} \tag{3.4}$$

Este sistema tiene solución única dada por

$$c_1 = 2, \quad c_2 = 3, \tag{3.5}$$

y entonces la proyección de \mathbf{w} sobre el espacio generado por \mathbf{v}_1 y \mathbf{v}_2 es

$$P_{\mathbf{v}_1, \mathbf{v}_2} \mathbf{w} = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 = 2 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + 3 \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ 5 \\ 8 \end{bmatrix} \tag{3.6}$$

• Después de calcular $P_{\mathbf{v}_1, \mathbf{v}_2} \mathbf{w}$ la diferencia $\mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2} \mathbf{w}$ está dada por

$$\mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2} \mathbf{w} = \begin{bmatrix} 8 \\ 2 \\ 2 \\ 8 \end{bmatrix} - \begin{bmatrix} 5 \\ 2 \\ 5 \\ 8 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ -3 \\ 0 \end{bmatrix},$$

de manera que el error de aproximación se determina mediante

$$\|\mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2} \mathbf{w}\|^2 = \left\| \begin{bmatrix} 3 \\ 0 \\ -3 \\ 0 \end{bmatrix} \right\|^2 = 18.$$

Ahora se obtendrá el mismo error de aproximación utilizando cada una de las expresiones en el Teorema 3.1. Primero observe que $\|\mathbf{w}\|^2 = 8^2 + 2^2 + 2^2 + 8^2 = 136$.

(i)

$$\|\mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2} \mathbf{w}\|^2 = \|\mathbf{w}\|^2 - \|P_{\mathbf{v}_1, \mathbf{v}_2} \mathbf{w}\|^2 = 136 - \left\| \begin{bmatrix} 5 \\ 2 \\ 5 \\ 8 \end{bmatrix} \right\|^2 = 136 - 118 = 18.$$

(ii) Para utilizar la segunda fórmula en el Teorema 3.1, observe que la matriz del sistema de ecuaciones normales es $\begin{bmatrix} \langle \mathbf{v}_1, \mathbf{v}_1 \rangle & \langle \mathbf{v}_1, \mathbf{v}_2 \rangle \\ \langle \mathbf{v}_2, \mathbf{v}_1 \rangle & \langle \mathbf{v}_2, \mathbf{v}_2 \rangle \end{bmatrix} = \begin{bmatrix} 4 & 4 \\ 4 & 6 \end{bmatrix}$, de tal manera que

$$\begin{aligned} \sum_{i=1}^2 \sum_{j=1}^2 c_i \langle \mathbf{v}_i, \mathbf{v}_j \rangle c_j &= [c_1, c_2] \begin{bmatrix} \langle \mathbf{v}_1, \mathbf{v}_1 \rangle & \langle \mathbf{v}_1, \mathbf{v}_2 \rangle \\ \langle \mathbf{v}_2, \mathbf{v}_1 \rangle & \langle \mathbf{v}_2, \mathbf{v}_2 \rangle \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \\ &= [2, 3] \begin{bmatrix} 4 & 4 \\ 4 & 6 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} \\ &= 2(4)2 + 2(4)3 + 3(4)2 + 3(6)3 \\ &= 16 + 24 + 24 + 54 = 118 \end{aligned}$$

y entonces el error de aproximación se determina mediante

$$\|\mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2} \mathbf{w}\|^2 = \|\mathbf{w}\|^2 - \sum_{i=1}^2 \sum_{j=1}^2 c_i \langle \mathbf{v}_i, \mathbf{v}_j \rangle c_j = 136 - 118 = 18.$$

(iii) La tercera fórmula en el Teorema 3.1 es la más simple y la más interesante. Se encontró en (3.5), mientras que los lados derechos de las ecuaciones normales son

$$\langle \mathbf{w}, \mathbf{v}_1 \rangle = 20, \quad \text{y} \quad \langle \mathbf{w}, \mathbf{v}_2 \rangle = 26,$$

mientras que la solución está dada en (3.5). Por lo tanto

$$c_1 \langle \mathbf{w}, \mathbf{v}_1 \rangle + c_2 \langle \mathbf{w}, \mathbf{v}_2 \rangle = 2(20) + 3(26) = 118,$$

y entonces

$$\|\mathbf{w} - P_{\mathbf{v}_1, \mathbf{v}_2} \mathbf{w}\|^2 = \|\mathbf{w}\|^2 - [c_1 \langle \mathbf{w}, \mathbf{v}_1 \rangle + c_2 \langle \mathbf{w}, \mathbf{v}_2 \rangle] = 136 - 118 = 18.$$

□

El sistema de ecuaciones normales es, sin duda alguna, el más importante conjunto de ecuaciones lineales que surge en el análisis estadístico. Especialmente cuando n es grande, resolverlo es una tarea complicada, y se han diseñado métodos especiales para resolverlo, como por ejemplo, el método de triangulación de Cholesky (Graybill (1985)). Aunque el interés de esta presentación no radica en encontrar de manera explícita las soluciones de (2.3), es conveniente señalar dos casos en los es posible resolver las ecuaciones normales con gran facilidad.

Ejemplo 3.2. (i) Considere el caso $n = 1$, de manera que se desea aproximar a \mathbf{w} por medio de un múltiplo del vector \mathbf{v}_1 . En este caso, (2.3) se reduce a una única ecuación:

$$c_1 \langle \mathbf{v}_1, \mathbf{v}_1 \rangle = \langle \mathbf{w}, \mathbf{v}_1 \rangle,$$

y entonces $c_1 = \frac{\langle \mathbf{w}, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}$. La proyección $P_{\mathbf{v}_1} \mathbf{w}$ es

$$P_{\mathbf{v}_1} \mathbf{w} = c_1 \mathbf{v}_1 = \frac{\langle \mathbf{w}, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 = \frac{\langle \mathbf{w}, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2} \mathbf{v}_1,$$

y, de acuerdo al Teorema 3.1(iii), el error de aproximación es

$$\|\mathbf{w} - P_{\mathbf{v}_1} \mathbf{w}\|^2 = \|\mathbf{w}\|^2 - c_1 \langle \mathbf{w}, \mathbf{v}_1 \rangle = \|\mathbf{w}\|^2 - \frac{\langle \mathbf{w}, \mathbf{v}_1 \rangle^2}{\|\mathbf{v}_1\|^2}.$$

(ii) Suponga que $n > 1$ es arbitrario, pero que los vectores \mathbf{v}_k son ortogonales, esto es,

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

Bajo esta condición, la i -ésima ecuación del sistema (2.3) es

$$c_i \langle \mathbf{v}_i, \mathbf{v}_i \rangle = \langle \mathbf{w}, \mathbf{v}_i \rangle,$$

y entonces

$$c_i = \frac{\langle \mathbf{w}, \mathbf{v}_i \rangle}{\langle \mathbf{v}_i, \mathbf{v}_i \rangle} = \frac{\langle \mathbf{w}, \mathbf{v}_i \rangle}{\|\mathbf{v}_i\|^2};$$

la proyección $P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}$ es

$$P_{\mathbf{v}_1} \mathbf{w} = \sum_{i=1}^n c_i \mathbf{v}_i = \sum_{i=1}^n \frac{\langle \mathbf{w}, \mathbf{v}_i \rangle}{\langle \mathbf{v}_i, \mathbf{v}_i \rangle} \mathbf{v}_i = \sum_{i=1}^n \frac{\langle \mathbf{w}, \mathbf{v}_i \rangle}{\|\mathbf{v}_i\|^2} \mathbf{v}_i,$$

y el error de aproximación está dado por

$$\begin{aligned} \|\mathbf{w} - P_{\mathbf{v}_1} \mathbf{w}\|^2 &= \|\mathbf{w}\|^2 - \sum_{i=1}^n c_i \langle \mathbf{w}, \mathbf{v}_i \rangle \\ &= \|\mathbf{w}\|^2 - \sum_{i=1}^n \frac{\langle \mathbf{w}, \mathbf{v}_i \rangle}{\|\mathbf{v}_i\|^2} \langle \mathbf{w}, \mathbf{v}_i \rangle \\ &= \|\mathbf{w}\|^2 - \sum_{i=1}^n \frac{\langle \mathbf{w}, \mathbf{v}_i \rangle^2}{\|\mathbf{v}_i\|^2} \end{aligned}$$

□

Caracterización de los Coeficientes

Como ya se ha mencionado, resolver el sistema de ecuaciones normales es una tarea complicada cuando el número de ecuaciones es grande. Sin embargo, frecuentemente es posible obtener expresiones para las constantes c_i que las resuelven (2.3) en términos de proyecciones, y el propósito de esta sección es establecer dichas fórmulas.

Teorema 4.1. Sean \mathbf{w} y $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ vectores en un espacio vectorial \mathcal{V} dotado con un producto interno $\langle \cdot, \cdot \rangle$. Sea $P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}$ la proyección de \mathbf{w} sobre el espacio generado por $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$, de manera que

$$P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w} = \sum_{i=1}^n c_i \mathbf{v}_i, \quad (4.1)$$

donde las constantes c_i satisfacen el sistema de ecuaciones normales (2.3). Suponga que para algún entero $k \in \{1, 2, \dots, n\}$, el vector \mathbf{v}_k no pertenece al espacio generado por los restantes vectores \mathbf{v}_i , esto es,

$$\mathbf{v}_k \notin \mathcal{L}\{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n\}. \quad (4.2)$$

En este caso, el coeficiente c_k en (4.1) está determinado por la siguiente expresión:

$$c_k = \frac{\langle \mathbf{w}, \mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k \rangle}{\|\mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k\|^2}; \quad (4.3)$$

equivalentemente,

$$c_k = \frac{\langle \mathbf{w} - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{w}, \mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k \rangle}{\|\mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k\|^2}. \quad (4.4)$$

Observación 4.1. Hay un aspecto notable en el Teorema 4.1 que se resalta a continuación: El problema original es determinar la proyección de \mathbf{w} sobre el espacio generado por un conjunto de n vectores $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$, y para alcanzar este objetivo es suficiente determinar las constantes c_i , $i = 1, 2, \dots, n$. Como se desprende de (4.3) o (4.4), una constante c_k puede determinarse si se sabe como obtener proyecciones sobre el espacio generado por los $n - 1$ vectores \mathbf{v}_i , $i = 1, \dots, k - 1, k + 1, \dots, n$. Luego, el contenido esencial del Teorema 4.1, es que el problema de construir una proyección sobre el espacio generado por un conjunto de n vectores, puede abordarse si se sabe como construir una proyección sobre un espacio que se genera a partir de $n - 1$ vectores.

Demostración del Teorema 4.1. Como punto de partida, recuerde que, por el Corolario 2.1,

$$\langle P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}, \mathbf{v}_k \rangle = \langle \mathbf{w}, \mathbf{v}_k \rangle \quad \text{para todo } k = 1, 2, \dots, n.$$

Luego, si $\mathbf{v} = \sum_{i=1}^n a_i \mathbf{v}_i$ es una combinación lineal arbitrario de $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$, entonces

$$\begin{aligned} \langle P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}, \mathbf{v} \rangle &= \langle P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}, \sum_{i=1}^n a_i \mathbf{v}_i \rangle \\ &= \sum_{i=1}^n a_i \langle P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}, \mathbf{v}_i \rangle \\ &= \sum_{i=1}^n a_i \langle \mathbf{w}, \mathbf{v}_i \rangle \\ &= \langle \mathbf{w}, \sum_{i=1}^n a_i \mathbf{v}_i \rangle, \end{aligned}$$

de donde se desprende

$$\langle P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}, \mathbf{v} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle, \quad \text{para todo } \mathbf{v} \in \mathcal{L}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}. \quad (4.5)$$

A continuación, observe que

$$\begin{aligned} P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w} &= \sum_{j=1}^n c_j \mathbf{v}_j = \sum_{j:j \neq k}^n c_j \mathbf{v}_j + c_k \mathbf{v}_k \\ &= \left[\sum_{j:j \neq k}^n c_j \mathbf{v}_j + c_k P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k \right] \\ &\quad + c_k [\mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k] \end{aligned} \quad (4.6)$$

Note ahora que $\mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k$ es perpendicular a todos los vectores \mathbf{v}_j con $j \neq k$ (por el Corolario 2.1), esto es, $\langle \mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k, \mathbf{v}_j \rangle = 0$ cuando $j \neq k$. Luego,

$$\begin{aligned} &\langle \mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k, \sum_{j:j \neq k} a_j \mathbf{v}_j \rangle \\ &= \sum_{j:k \neq j} a_j \langle \mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k, \mathbf{v}_j \rangle = 0; \end{aligned}$$

Debido a que el primer término entre corchetes en (4.6) es una combinación lineal de los vectores \mathbf{v}_j , con $j \neq k$, se desprende que

$$\left\langle \left[\sum_{j:j \neq k}^n c_j \mathbf{v}_j + c_k P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k \right], [\mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k] \right\rangle = 0$$

Tomando el producto interno con $\mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k$ en la ecuación (4.6) y usando la igualdad precedente, se desprende que

$$\begin{aligned} &\langle P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}, \mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k \rangle \\ &= \left\langle \left[\sum_{j:j \neq k}^n c_j \mathbf{v}_j + c_k P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k \right], [\mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k] \right\rangle \\ &\quad + \langle c_k [\mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k], [\mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k] \rangle \\ &= c_k \langle [\mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k], [\mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k] \rangle \\ &= c_k \|\mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k\|^2. \end{aligned}$$

Para continuar, observe que $\mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k$, es un vector no nulo, por la condición (4.2), de manera que es posible despejar c_k para obtener

$$c_k = \frac{\langle P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}, \mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k \rangle}{\| \mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k \|^2}.$$

Puesto que el vector $\mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k$ es una combinación lineal de los vectores \mathbf{v}_j con $j \neq k$, (4.5) implica que

$$\langle P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n} \mathbf{w}, \mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k \rangle = \langle \mathbf{w}, \mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k \rangle$$

de manera que la expresión anterior para c_k se reduce a

$$c_k = \frac{\langle \mathbf{w}, \mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k \rangle}{\| \mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k \|^2},$$

estableciendo (4.3). Para verificar (4.4), recuerde que $\mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k$ es perpendicular a cualquier combinación lineal de $\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n$, de manera que

$$\langle P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{w}, \mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k \rangle = 0,$$

pues $P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{w}$ es una combinación lineal de los vectores $\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n$. Por lo tanto,

$$\begin{aligned} & \langle \mathbf{w} - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{w}, \mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k \rangle \\ &= \langle \mathbf{w}, \mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k \rangle \\ &\quad - \langle P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{w}, \mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k \rangle \\ &= \langle \mathbf{w}, \mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k \rangle \end{aligned}$$

y entonces (4.3) implica que

$$c_k = \frac{\langle \mathbf{w} - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{w}, \mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k \rangle}{\| \mathbf{v}_k - P_{\mathbf{v}_1, \dots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n} \mathbf{v}_k \|^2},$$

estableciendo (4.4) y concluyendo la demostración del Teorema. \square

Ejemplo 4.1. Sean los vectores \mathbf{w} , \mathbf{v}_1 y \mathbf{v}_2 como en el Ejemplo 3.1. A continuación se utilizará el Teorema 4.1 para determinar la proyección de \mathbf{w} sobre el espacio generado por \mathbf{v}_1 y \mathbf{v}_2 . Desde luego, dicha proyección es

$$P_{\mathbf{v}_1, \mathbf{v}_2} \mathbf{w} = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2,$$

y el problema consiste en encontrar los coeficientes c_1 y c_2 . De acuerdo a (4.3),

$$c_1 = \frac{\langle \mathbf{w}, \mathbf{v}_1 - P_{\mathbf{v}_2} \mathbf{v}_1 \rangle}{\|\mathbf{v}_1 - P_{\mathbf{v}_2} \mathbf{v}_1\|^2}. \quad (4.7)$$

Para determinar este coeficiente, primero debe encontrarse $P_{\mathbf{v}_2} \mathbf{v}_1$, y el Ejemplo 3.2(i) permite escribir

$$\begin{aligned} P_{\mathbf{v}_2} \mathbf{v}_1 &= \frac{\langle \mathbf{v}_1, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2} \mathbf{v}_2 \\ &= \frac{4}{6} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix}, \end{aligned}$$

y por lo tanto,

$$\mathbf{v}_1 - P_{\mathbf{v}_2} \mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{4}{6} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 1 \\ 1/3 \\ -1/3 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 \\ 3 \\ 1 \\ -1 \end{bmatrix}.$$

Luego,

$$\langle \mathbf{w}, \mathbf{v}_1 - P_{\mathbf{v}_2} \mathbf{v}_1 \rangle = \frac{1}{3} [8(1) + 2(3) + 2(1) + 8(-1)] = \frac{8}{3}.$$

Puesto que

$$\|\mathbf{v}_1 - P_{\mathbf{v}_2} \mathbf{v}_1\|^2 = \frac{1}{9} [1^2 + 3^2 + 1^2 + (-1)^2] = \frac{12}{9} = \frac{4}{3},$$

(4.7) implica

$$c_1 = \frac{8/3}{4/3} = 2. \quad (4.8)$$

A continuación se determinará el coeficiente c_2 utilizando el Teorema 4.1:

$$c_2 = \frac{\langle \mathbf{w}, \mathbf{v}_2 - P_{\mathbf{v}_1} \mathbf{v}_2 \rangle}{\|\mathbf{v}_2 - P_{\mathbf{v}_1} \mathbf{v}_2\|^2}. \quad (4.9)$$

Como antes, primero debe encontrarse $P_{\mathbf{v}_1} \mathbf{v}_2$, vector que, de acuerdo al Ejemplo 3.2(i), está dado por

$$\begin{aligned} P_{\mathbf{v}_1} \mathbf{v}_2 &= \frac{\langle \mathbf{v}_1, \mathbf{v}_2 \rangle}{\|\mathbf{v}_1\|^2} \mathbf{v}_1 \\ &= \frac{4}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \end{aligned}$$

y por lo tanto,

$$\mathbf{v}_2 - P_{\mathbf{v}_1} \mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 0 \\ -1 \\ 0 \\ 1 \end{bmatrix}.$$

Luego,

$$\langle \mathbf{w}, \mathbf{v}_2 - P_{\mathbf{v}_1} \mathbf{v}_2 \rangle = [8(0) + 2(-1) + 2(0) + 8(1)] = 6.$$

Puesto que

$$\|\mathbf{v}_2 - P_{\mathbf{v}_1} \mathbf{v}_2\|^2 = [0^2 + (-1)^2 + 0^2 + (1)^2] = 2,$$

a partir de (4.9) se obtiene

$$c_1 = \frac{6}{2} = 3;$$

note que estos valores para los coeficientes c_1 y c_2 coinciden con los determinados en el Ejemplo 3.1. Por lo tanto,

$$P_{\mathbf{v}_1, \mathbf{v}_2} \mathbf{w} = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 = 2 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + 3 \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ 5 \\ 8 \end{bmatrix}.$$

Interpretación Estadística

El objetivo de esta sección es trasladar los resultados algebraicos establecidos anteriormente al contexto de variables aleatorias, proporcionando los instrumentos para demostrar el algoritmo de Durbin–Levinson presentado más adelante. Dado un espacio de probabilidad fijo (Ω, \mathcal{F}, P) , considere el conjunto \mathcal{V} de variables

aleatorias definidas en Ω que tiene segundo momento finito, esto es, una variable aleatoria X definida en Ω pertenece a \mathcal{V} si y sólo si $E[X^2] < \infty$. Utilizando que $(X + Y)^2 \leq 2[X^2 + Y^2]$, no es difícil ver que \mathcal{V} es un espacio vectorial, a condición de que se identifiquen todas las variables aleatorias que satisfacen $E[X^2] = 0$ con la variable aleatoria nula. El producto interno en este espacio \mathcal{V} se define mediante

$$\langle X, Y \rangle = E[XY]. \quad (5.1)$$

Regresión Lineal. Suponga que las variables aleatorias Y y X_1, X_2, \dots, X_n son miembros de \mathcal{V} . El problema de encontrar la mejor aproximación a Y por medio de una combinación lineal de X_1, X_2, \dots, X_n se reduce a determinar constantes c_1, c_2, \dots, c_n tales que

$$\left\| Y - \sum_{i=1}^n c_i X_i \right\|^2 = \min_{\beta_1, \beta_2, \dots, \beta_n} \left\| Y - \sum_{i=1}^n \beta_i X_i \right\|^2, \quad (5.2)$$

o equivalentemente,

$$E \left[\left(Y - \sum_{i=1}^n c_i X_i \right)^2 \right] = \min_{\beta_1, \beta_2, \dots, \beta_n} E \left[\left(Y - \sum_{i=1}^n \beta_i X_i \right)^2 \right]. \quad (5.3)$$

De acuerdo al Teorema 2.1, las constantes c_1, c_2, \dots, c_n se determinan resolviendo el sistema de ecuaciones (2.3), i.e.,

$$\sum_{j=1}^n \langle X_k, X_j \rangle c_j = \langle Y, X_k \rangle, \quad k = 1, 2, \dots, n,$$

las cuales, utilizando (5.1), se reducen a

$$\sum_{j=1}^n E[X_k, X_j] c_j = E[Y, X_k], \quad k = 1, 2, \dots, n. \quad (5.4)$$

La proyección de Y sobre el espacio generado por X_1, X_2, \dots, X_n es

$$P_{X_1, X_2, \dots, X_n} Y = \sum_{i=1}^n c_i X_i, \quad (5.5)$$

la cual, en el contexto presente, se conoce como la *función de regresión lineal* de Y respecto a X_1, X_2, \dots, X_n ; en las aplicaciones, es frecuente que X_1 sea la variable aleatoria constante 1, de modo que la función de regresión es

$$P_{X_1, X_2, \dots, X_n} Y = c_1 + \sum_{i=2}^n c_i X_i,$$

y la constante c_1 se conoce como la *ordenada al origen*.

El Problema de Pronóstico. Suponga que X_1, X_2, \dots, X_n se observan secuencialmente, esto es, X_i se manifiesta en el tiempo i , mientras que Y será observada en el tiempo $n + 1$. En este caso, de manera natural surge el problema de pronosticar, en el tiempo n , el valor aleatorio que se obtendrá en el tiempo $n + 1$, es decir, Y . Al buscar el ‘mejor’ pronóstico para Y en términos de una combinación lineal de X_1, X_2, \dots, X_n , se arriba al problema de encontrar constantes c_1, c_2, \dots, c_n que satisfacen (5.2) (o (5.3)), y determinarlas se reduce a resolver (5.4). La proyección de Y sobre el espacio generado por X_1, X_2, \dots, X_n , es el mejor pronóstico lineal para Y en términos de las variables X_i , $i = 1, 2, \dots, n$. Observe que si todas las variables aleatorias X_i tienen media nula, el sistema (5.4) puede escribirse como

$$\sum_{j=1}^n \text{Cov}[X_k, X_j] c_j = \text{Cov}[Y, X_k], \quad k = 1, 2, \dots, n. \quad (5.6)$$

Esta discusión ha mostrado que la teoría algebraica presentada en las secciones precedentes tiene interpretaciones interesantes en el campo de la Estadística. En la siguiente sección se utiliza este enfoque para establecer uno de los métodos más importantes de construcción de pronósticos para series de tiempo estacionarias.

El Algoritmo de Durbin–Levinson

Sea $X_1, X_2, \dots, X_n, X_{n+1} \dots$ una serie de tiempo estacionaria con media nula. El propósito de esta sección es presentar un procedimiento que permite calcular la proyección de X_{n+1} sobre $\mathcal{L}\{X_1, X_2, \dots, X_n\}$, en términos de la proyección

de X_n sobre $\mathcal{L}\{X_1, X_2, \dots, X_{n-1}\}$. Para describir el resultado de manera precisa, sean $\phi_{n1}, \phi_{n2}, \dots, \phi_{nn}$ los coeficientes de X_n, X_{n-1}, \dots, X_1 en la expresión para el pronóstico de X_{n+1} ;

$$P_{X_1, X_2, \dots, X_n} X_{n+1} = \phi_{n1} X_n + \phi_{n2} X_{n-1} + \dots + \phi_{nn} X_1 = \sum_{k=1}^n \phi_{nk} X_{n+1-k}. \quad (6.1)$$

Similarmente, la proyección de X_n sobre X_1, X_2, \dots, X_{n-1} es

$$P_{X_1, X_2, \dots, X_{n-1}} X_n = \phi_{n-11} X_{n-1} + \phi_{n-12} X_{n-2} + \dots + \phi_{n-1, n-1} X_1 = \sum_{k=1}^{n-1} \phi_{n-1k} X_{n-k}. \quad (6.2)$$

El error de pronóstico en el tiempo n es

$$v_n = E \left[(X_{n+1} - Pr_{X,n} X_{n+1})^2 \right]. \quad (6.3)$$

EL principal objetivo es determinar cómo los coeficientes $\{\phi_{nk} | k = 1, 2, \dots, n\}$ pueden expresarse en términos de $\{\phi_{n-1k} | k = 1, 2, \dots, n-1\}$. El siguiente procedimiento, conocido como el algoritmo de Durbin–Levinson, permite alcanzar este propósito, así como determinar los errores de pronóstico en cada etapa.

Teorema 6.1. Teorema. Sea $X_1, X_2, \dots, X_n, \dots$ una serie de tiempo estacionaria con media nula y función de covarianza $\gamma(\cdot)$, y suponga que las variables aleatorias $X_1, X_2, \dots, X_n, X_{n+1}, \dots$ son independientes. Defina

$$v_0 = \gamma(0) \quad (6.4)$$

En este caso, se tiene que

$$\phi_{11} = \frac{\gamma(1)}{\gamma(0)}, \quad (6.5)$$

y

$$v_1 = v_0 [1 - \phi_{11}^2]. \quad (6.6)$$

Más aún, para $n \geq 2$,

$$\phi_{nn} = \left[\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j} \gamma(n-j) \right] v_{n-1}^{-1}, \quad (6.7)$$

$$\begin{bmatrix} \phi_{n1} \\ \vdots \\ \phi_{nn-1} \end{bmatrix} = \begin{bmatrix} \phi_{n-11} \\ \vdots \\ \phi_{n-1n-1} \end{bmatrix} - \phi_{nn} \begin{bmatrix} \phi_{n-1n-1} \\ \vdots \\ \phi_{n-11} \end{bmatrix} \quad (6.8)$$

y

$$v_n = v_{n-1}[1 - \phi^2]. \quad (6.9)$$

La parte más interesante de este teorema se conforma con las igualdades (6.7)–(6.9), las cuales permiten determinar los coeficientes $\{\phi_{nk}\}$ y el error de pronóstico v_n , en términos de los coeficientes $\{\phi_{n-1k}\}$ y v_{n-1} . Por el momento, observe que (6.5) y (6.6) se obtiene directamente del Ejemplo 3.1(i). En efecto, la proyección de X_2 sobre X_1 es

$$P_{X_1}X_2 = \frac{\langle X_1, X_2 \rangle}{\langle X_1, X_1 \rangle} X_1 = \frac{E[X_1X_2]}{E[X_1^2]} X_1,$$

donde la segunda igualdad se debe a (5.1). Debido a que la serie $\{X_t\}$ tiene media nula, la expresión anterior se reduce a

$$P_{X_1}X_2 = \frac{\text{Cov}[X_1X_2]}{\text{Cov}[X_1, X_1]} X_1 = \frac{\gamma(1)}{\gamma(0)} X_1,$$

donde se ha utilizado la definición de la función de autocovarianza. Luego, comparando con (6.1), se desprende que la expresión (6.5) es válida, mientras que aplicando el Teorema 3.2, se obtiene que

$$\begin{aligned} v_1 &= \|X_2 - P_{X_1}X_2\|^2 \\ &= \|X_2\|^2 - \|P_{X_1}X_2\|^2 \\ &= \|X_2\|^2 - \|\phi_{11}X_1\|^2 \\ &= E[X_2^2] - \phi_{11}^2 E[X_1^2] \\ &= \gamma(0) - \phi_{11}^2 \gamma(0) \end{aligned}$$

de manera que la definición (6.4) implica

$$v_1 = v_0[1 - \phi_{11}^2],$$

estableciendo (6.6). El argumento que conduce a demostrar (6.7)–(6.9) es más complicado, y requiere de las herramientas auxiliares presentadas a continuación.

Resultados Auxiliares

Esta sección contiene los preliminares técnicos que se utilizarán para demostrar el Teorema 6.1, los cuales se presentan a continuación en los Lemas 7.1 y 7.2; ambos resultados utilizan la hipótesis de que la serie $\{X_t\}$ es estacionaria.

Lema 7.1. Sean $\{\phi_{n-1 k} \mid k = 1, 2, \dots, n-1\}$ los coeficientes de X_{n-1}, \dots, X_1 en la proyección de X_n sobre $\mathcal{L}\{X_1, X_2, \dots, X_{n-1}\}$, esto es,

$$\begin{aligned} P_{X_1, X_2, \dots, X_{n-1}} X_n \\ = \phi_{n-1 1} X_{n-1} + \phi_{n-1 2} X_{n-2} + \dots + \phi_{n-1 n-1} X_1 = \sum_{k=1}^{n-1} \phi_{n-1 k} X_{n-k}; \end{aligned}$$

vea (6.2). En este caso, la proyección de X_{n+1} sobre $\mathcal{L}\{X_2, X_3, \dots, X_n\}$ se obtiene utilizando los mismos coeficientes $\{\phi_{n-1 k} \mid k = 1, 2, \dots, n-1\}$ en el mismo orden, esto es,

$$P_{X_2, X_3, \dots, X_n} X_{n+1} = \phi_{n-1 1} X_n + \phi_{n-1 2} X_{n-1} + \dots + \phi_{n-1 n-1} X_2. \quad (7.1)$$

Más aún, el error de aproximación al proyectar X_{n+1} sobre X_2, X_3, \dots, X_n es

$$\|X_{n+1} - P_{X_2, X_3, \dots, X_n} X_{n+1}\|^2 = v_{n-1}. \quad (7.2)$$

Demostración. El argumento descansa en la teoría presentada en la Sección 2. Dada una sucesión de k vectores $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$, y otro vector \mathbf{w} , la proyección de \mathbf{w} sobre el espacio generado por $\mathcal{L}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$, se construye resolviendo las ecuaciones normales (2.3). Dicho sistema depende de los productos internos

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle, \quad \text{y} \quad \langle \mathbf{w}, \mathbf{v}_i \rangle, \quad i, j = 1, 2, \dots, k. \quad (7.3)$$

Considere ahora el problema de determinar la proyección de X_n sobre el espacio generado por las variables aleatorias X_1, X_2, \dots, X_{n-1} , el cual corresponde a seleccionar $k = n - 1$, $\mathbf{w} = X_n$, mientras que $\mathbf{v}_i = X_{n-i}$, para $i = 1, 2, \dots, n - 1$. En este caso,

$$\begin{aligned}\langle \mathbf{v}_i, \mathbf{v}_j \rangle &= \langle X_{n-i}, X_{n-j} \rangle = \text{Cov} [X_{n-i}, X_{n-j}] = \gamma(|i - j|), \\ \langle \mathbf{w}, \mathbf{v}_i \rangle &= \langle X_n, X_{n-i} \rangle = \text{Cov} [X_n, X_{n-i}] = \gamma(i);\end{aligned}\tag{7.4}$$

estas cantidades permiten formular el sistema de ecuaciones normales, y al resolverlo se obtienen las constantes

$$c_i = \phi_{n-1,i},$$

de manera que

$$P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{n-1}} \mathbf{w} = \sum_{i=1}^{n-1} c_i \mathbf{v}_i,\tag{7.5}$$

esto es,

$$P_{X_1, X_2, \dots, X_{n-1}} X_n = \sum_{i=1}^{n-1} \phi_{n-1,i} X_{n-i},$$

y el error de aproximación es

$$\|X_n - P_{X_1, X_2, \dots, X_{n-1}}\|^2 = v_{n-1}.\tag{7.6}$$

Ahora se plantea el problema de determinar la proyección de X_{n+1} sobre las $n - 1$ variables que le preceden, esto es, X_2, X_3, \dots, X_n , lo que equivale a seleccionar $k = n - 1$, $\mathbf{w} = X_{n+1}$ y $\mathbf{v}_i = X_{n+1-i}$, $i = 1, 2, \dots, n - 1$. En este caso,

$$\begin{aligned}\langle \mathbf{v}_i, \mathbf{v}_j \rangle &= \langle X_{n+1-i}, X_{n+1-j} \rangle = \text{Cov} [X_{n+1-i}, X_{n+1-j}] = \gamma(|i - j|), \\ \langle \mathbf{w}, \mathbf{v}_i \rangle &= \langle X_{n+1}, X_{n+1-i} \rangle = \text{Cov} [X_{n+1}, X_{n+1-i}] = \gamma(i);\end{aligned}$$

comparando estas cantidades con aquellas en (7.4), se desprende que en el caso presente el sistema de ecuaciones normales es el mismo que antes, y por lo tanto, las constantes que las resuelven son idénticas a las del caso precedente, esto es, $c_i = \phi_{n-1,i}$, $i = 1, 2, \dots, n - 1$, y el error de aproximación es el mismo, i.e.,

$$\|X_{n+1} - P_{X_2, X_3, \dots, X_n}\|^2 = v_{n-1}$$

lo cual establece (7.2), y

$$P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{n-1}} \mathbf{w} = \sum_{i=1}^{n-1} c_i \mathbf{v}_i$$

igualdad que en el caso que nos ocupa equivale a

$$P_{X_2, X_3, \dots, X_n} X_{n+1} = \sum_{i=1}^{n-1} \phi_{n-1 i} X_{n+1-i},$$

esto es,

$$P_{X_2, X_3, \dots, X_n} X_{n+1} = \phi_{n-1 1} X_n + \phi_{n-1 2} X_{n-1} + \dots + \phi_{n-1 n-1} X_2,$$

concluyendo la demostración del lema. \square

Lema 7.2. Sean $\{\phi_{n-1 k} \mid k = 1, 2, \dots, n-1\}$ los coeficientes de X_{n-1}, \dots, X_1 en la proyección de X_n sobre $\mathcal{L}\{X_1, X_2, \dots, X_{n-1}\}$, i.e.,

$$\begin{aligned} P_{X_1, X_2, \dots, X_{n-1}} X_n &= \phi_{n-1 1} X_{n-1} + \phi_{n-1 2} X_{n-2} + \dots + \phi_{n-1 n-1} X_1 \\ &= \sum_{k=1}^{n-1} \phi_{n-1 k} X_{n-k}. \end{aligned}$$

vea (6.2). En este caso, la proyección de X_1 sobre $\mathcal{L}\{X_2, X_3, \dots, X_n\}$ se obtiene utilizando los mismos coeficientes $\{\phi_{n-1 k} \mid k = 1, 2, \dots, n-1\}$ en el orden inverso; más precisamente,

$$P_{X_2, X_3, \dots, X_n} X_1 = \phi_{n-1 1} X_2 + \phi_{n-1 2} X_3 + \dots + \phi_{n-1 n-1} X_n, \quad (7.7)$$

y el error de aproximación es

$$\|X_1 - P_{X_2, X_3, \dots, X_n} X_1\|^2 = v_{n-1}. \quad (7.8)$$

Demostración. Como en la demostración del lema precedente, el argumento des-
cansa en las ecuaciones normales (2.3). Al abordar el problema de determinar la

proyección de X_n sobre X_1, X_2, \dots, X_{n-1} , el sistema (2.3) se construye seleccionando $\mathbf{v}_i = X_{n-i}$, $i = 1, 2, \dots, n-1 = k$, y $\mathbf{w} = X_n$, y calculando los productos internos

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \langle X_{n-i}, X_{n-j} \rangle = \text{Cov} [X_{n-i}, X_{n-j}] = \gamma(|i-j|),$$

$$\langle \mathbf{w}, \mathbf{v}_i \rangle = \langle X_n, X_{n-i} \rangle = \text{Cov} [X_n, X_{n-i}] = \gamma(i).$$

lo cual conduce a determinar los coeficientes $\{\phi_{n-1 k}\}$, así como el error de aproximación

$$\|X_n - P_{X_1, X_2, \dots, X_{n-1}} X_n\|^2 = v_{n-1}.$$

Considere el problema de encontrar la proyección de X_1 sobre X_2, X_3, \dots, X_n , lo que equivale a seleccionar $k = n-1$, así como

$$\mathbf{w} = X_1 \quad \text{y} \quad \mathbf{v}_i = X_{i+1}, \quad i = 1, 2, \dots, n-1. \quad (7.9)$$

En este caso,

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \langle X_{i+1}, X_{j+1} \rangle = \text{Cov} [X_{i+1}, X_{j+1}] = \gamma(|i-j|),$$

$$\langle \mathbf{w}, \mathbf{v}_i \rangle = \langle X_1, X_{i+1} \rangle = \text{Cov} [X_1, X_{i+1}] = \gamma(i);$$

Luego, en ambos casos el sistema de ecuaciones normales es el mismo, y por lo tanto también las soluciones son idénticas y el error de aproximación coincide, i.e.,

$$\|X_1 - P_{X_2, X_3, \dots, X_n} X_1\|^2 = v_{n-1},$$

y $c_i = \phi_{n-1 i}$, para todo i , de tal forma que

$$P_{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{n-1}} \mathbf{w} = \sum_{i=1}^{n-1} c_i \mathbf{v}_i$$

igualdad que con los datos en (7.9) equivale a

$$P_{X_2, X_3, \dots, X_n} X_1 = \sum_{i=1}^{n-1} \phi_{n-1 i} X_{i+1},$$

y en forma desarrollada se obtiene

$$P_{X_2, X_3, \dots, X_n} X_1 = \phi_{n-1 1} X_2 + \phi_{n-1 2} X_3 + \dots + \phi_{n-1 n-1} X_n,$$

concluyendo la demostración del lema. \square

Demostración del Teorema 6.1

Después de establecer los instrumentos técnicos de la sección anterior, la demostración del algoritmo de Durbin–Levinson se completa a continuación.

Demostración del Teorema 6.1. Como ya se validaron las igualdades (6.5) y (6.6), es suficiente verificar (6.7) y (6.8). Para iniciar, observe que

$$\begin{aligned}
 P_{X_1, X_2, \dots, X_n} X_{n+1} &= \sum_{i=1}^n \phi_{ni} X_{n+1-i} \\
 &= \sum_{i=1}^{n-1} \phi_{ni} X_{n+1-i} + \phi_{nn} X_1 \\
 &= \left[\sum_{i=1}^{n-1} \phi_{ni} X_{n+1-i} + \phi_{nn} P_{X_2, X_3, \dots, X_n} X_1 \right] \\
 &\quad + \phi_{nn} [X_1 - P_{X_2, X_3, \dots, X_n} X_1]
 \end{aligned} \tag{8.1}$$

Tomando el producto interno por $X_1 - P_{X_2, X_3, \dots, X_n} X_1$ en los extremos de (8.1), se desprende que

$$\begin{aligned}
 \langle P_{X_1, X_2, \dots, X_n} X_{n+1}, X_1 - P_{X_2, X_3, \dots, X_n} X_1 \rangle \\
 = \phi_{nn} \langle X_1 - P_{X_2, X_3, \dots, X_n} X_1, X_1 - P_{X_2, X_3, \dots, X_n} X_1 \rangle,
 \end{aligned}$$

donde se ha utilizado que $X_1 - P_{X_2, X_3, \dots, X_n} X_1$ tiene producto interno nulo con cualquier combinación lineal de X_2, X_3, \dots, X_n . Aplicando ahora el Corolario 2.1, se obtiene que X_{n+1} y $P_{X_1, X_2, \dots, X_n} X_{n+1}$ tienen los mismos productos internos con toda combinación lineal de X_1, X_2, \dots, X_n , de manera que

$$\langle P_{X_1, X_2, \dots, X_n} X_{n+1}, X_1 - P_{X_2, X_3, \dots, X_n} X_1 \rangle = \langle X_{n+1}, X_1 - P_{X_2, X_3, \dots, X_n} X_1 \rangle.$$

Combinando esta ecuación con la anterior igualdad desplegada, se obtiene

$$\begin{aligned}
 \langle X_{n+1}, X_1 - P_{X_2, X_3, \dots, X_n} X_1 \rangle \\
 = \phi_{nn} \langle X_1 - P_{X_2, X_3, \dots, X_n} X_1, X_1 - P_{X_2, X_3, \dots, X_n} X_1 \rangle.
 \end{aligned} \tag{8.2}$$

Ahora puede verificarse (6.7) de la siguiente forma: Primero, note que el Lemma 7.2 implica que

$$P_{X_2, X_3, \dots, X_n} X_1 = \sum_{i=1}^{n-1} \phi_{n-1, i} X_{i+1}$$

de manera que

$$\begin{aligned} \langle X_{n+1}, X_1 - \sum_{i=1}^{n-1} \phi_{n-1, i} X_{i+1} \rangle &= \langle X_{n+1}, X_1 \rangle - \sum_{i=1}^{n-1} \langle X_{n+1}, \phi_{n-1, i} X_{i+1} \rangle \\ &= \gamma(n) - \sum_{i=1}^{n-1} \phi_{n-1, i} \langle X_{n+1}, X_{i+1} \rangle \\ &= \gamma(n) - \sum_{i=1}^{n-1} \phi_{n-1, i} \gamma(n-i) \end{aligned}$$

y

$$\langle X_1 - P_{X_2, X_3, \dots, X_n} X_1, X_1 - P_{X_2, X_3, \dots, X_n} X_1 \rangle = \|X_1 - P_{X_2, X_3, \dots, X_n} X_1\|^2 = v_{n-1};$$

combinando estas dos últimas igualdades con (8.2) se desprende que

$$\gamma(n) - \sum_{i=1}^{n-1} \phi_{n-1, i} \gamma(n-i) = \phi_{n n} v_{n-1},$$

y entonces

$$\phi_{n n} = \left[\gamma(n) - \sum_{i=1}^{n-1} \phi_{n-1, i} \gamma(n-i) \right] v_{n-1}^{-1},$$

verificando (6.7) Para comprobar (6.8), primero se demostrará que

$$P_{X_2, X_3, \dots, X_n} X_{n+1} = \left[\sum_{i=1}^{n-1} \phi_{n i} X_{n+1-i} + \phi_{n n} P_{X_2, X_3, \dots, X_n} X_1 \right]. \quad (8.3)$$

Para verificar esta afirmación, es suficiente con establecer que

$$\langle X_{n+1} - \left[\sum_{i=1}^{n-1} \phi_{n i} X_{n+1-i} + \phi_{n n} P_{X_2, X_3, \dots, X_n} X_1 \right], X_j \rangle = 0, \quad j = 2, 3, \dots, n; \quad (8.4)$$

vea el Corolario 2.1. Con esta finalidad, note que a partir de (8.1) se desprende

$$\begin{aligned}
X_{n+1} - \left[\sum_{i=1}^{n-1} \phi_n i X_{n+1-i} + \phi_n n P_{X_2, X_3, \dots, X_n} X_1 \right] \\
= -[X_{n+1} - P_{X_1, X_2, \dots, X_n} X_{n+1}] + \phi_n n [X_1 - P_{X_2, X_3, \dots, X_n} X_1].
\end{aligned}$$

Note ahora que, aplicando el Corolario 2.1, $X_{n+1} - P_{X_1, X_2, \dots, X_n} X_{n+1}$ es perpendicular a cada X_i con $i \leq n$: en particular

$$\langle X_{n+1} - P_{X_1, X_2, \dots, X_n} X_{n+1}, X_j \rangle = 0, \quad j = 1, 2, \dots, n.$$

Además, $X_1 - P_{X_2, X_3, \dots, X_n} X_1$ es perpendicular a X_j para todo $j \geq 2$, i.e.,

$$\langle X_1 - P_{X_2, X_3, \dots, X_n} X_1, X_j \rangle = 0, \quad j = 1, 2, \dots, n,$$

y combinando las tres últimas igualdades se desprende que la afirmación (8.4) es válida. Usando ahora el Lema 7.1 se obtiene que

$$P_{X_2, X_3, \dots, X_n} X_{n+1} = \sum_{i=0}^{n-1} \phi_{n-1, i} X_{n+1-i},$$

mientras que el Lemma 7.2 implica

$$P_{X_2, X_3, \dots, X_n} X_1 = \sum_{k=0}^{n-1} \phi_{n-1, k} X_{k+1} = \sum_{i=0}^{n-1} \phi_{n-1, n-i} X_{n+1-i},$$

donde la segunda igualdad se obtiene haciendo el cambio de variable $k = n - i$.

Combinando estas dos últimas igualdades con (8.3) se desprende que

$$\begin{aligned}
\sum_{i=0}^{n-1} \phi_{n-1, i} X_{n+1-i} &= \sum_{i=1}^{n-1} \phi_n i X_{n+1-i} \\
&\quad + \phi_n n \sum_{i=0}^{n-1} \phi_{n-1, n-i} X_{n+1-i} \\
&= \sum_{i=1}^{n-1} [\phi_n i + \phi_n n \phi_{n-1, n-i}] X_{n+1-i}
\end{aligned}$$

y usando la independencia de X_1, X_2, \dots, X_n , esta igualdad implica que

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

de donde se concluye

$$\phi_{ni} = \phi_{n-1,i} - \phi_{nn}\phi_{n-1,n-i}, \quad i = 1, 2, \dots, n-1,$$

igualdad que en notación vectorial equivale a

$$\begin{bmatrix} \phi_{n1} \\ \vdots \\ \phi_{nn-1} \end{bmatrix} = \begin{bmatrix} \phi_{n-11} \\ \vdots \\ \phi_{n-1n-1} \end{bmatrix} - \phi_{nn} \begin{bmatrix} \phi_{n-1n-1} \\ \vdots \\ \phi_{n-11} \end{bmatrix}$$

comprobando (6.8). Para concluir, se demostrará (6.9). Aplicando el Teorema 3.1, se obtiene que

$$v_n = \|X_{n+1}\|^2 - \|P_{X_1, X_2, \dots, X_n} X_{n+1}\|^2. \quad (8.5)$$

A continuación, observe la expresión (8.1) para la proyección de X_{n+1} sobre el espacio generado por las n variables aleatorias X_1, X_2, \dots, X_n ; los dos vectores entre corchetes en la extrema derecha de (8.1) son perpendiculares, de tal forma que

$$\begin{aligned} & \|P_{X_1, X_2, \dots, X_n} X_{n+1}\|^2 \\ &= \left\| \sum_{i=1}^{n-1} \phi_{ni} X_{n+1-i} + \phi_{nn} P_{X_2, X_3, \dots, X_n} X_1 \right\|^2 + \phi_{nn}^2 \|X_1 - P_{X_2, X_3, \dots, X_n} X_1\|^2 \end{aligned}$$

y a partir de (8.5) se desprende

$$\begin{aligned} v_n &= \|X_{n+1}\|^2 - \left\| \sum_{i=1}^{n-1} \phi_{ni} X_{n+1-i} + \phi_{nn} P_{X_2, X_3, \dots, X_n} X_1 \right\|^2 \\ &\quad - \phi_{nn}^2 \|X_1 - P_{X_2, X_3, \dots, X_n} X_1\|^2 \\ &= \|X_{n+1}\|^2 - \|P_{X_2, X_3, \dots, X_n}\|^2 \\ &\quad - \phi_{nn}^2 \|X_1 - P_{X_2, X_3, \dots, X_n} X_1\|^2 \end{aligned}$$

donde se utilizó (8.3) para obtener la segunda igualdad. Aplicando de nueva cuenta el Teorema 3.1, se sigue que

$$\begin{aligned} v_n &= \|X_{n+1} - P_{X_2, X_3, \dots, X_n}\|^2 \\ &\quad - \phi_{nn}^2 \|X_1 - P_{X_2, X_3, \dots, X_n} X_1\|^2 \end{aligned}$$

y aplicando los Lemas 7.1 y 7.2 (vea, especialmente, (7.6) y (7.8)), se arriba a

$$v_n = v_{n-1} - \phi_{nn}^2 v_{n-1} = v_{n-1}[1 - \phi_{nn}^2],$$

expresión que coincide con (6.9). \square

Después de demostrar la validez del método recursivo de Durbin–Levinson, es conveniente puntualizar que el instrumento fundamental que condujo a establecer dicho método es el resultado, presentado en la Sección 4, sobre la caracterización de las constantes que resuelven las ecuaciones normales.

Un Ejemplo

En esta sección se ilustra el procedimiento recursivo de Durbin–Levinson, y para ello se realizarán, de forma detallada, los cálculos que conducen a determinar los coeficientes $\{\phi_{nk}\}$ en términos de $\{\phi_{n-1k}\}$. Aunque el método descrito en el Teorema 6.1 puede implementarse fácilmente en computadora, la idea detrás de los cálculos que se realizan a continuación, es ilustrar los detalles de cómputo en un caso específico.

Ejemplo 9.1. Se utilizará los siguientes datos sobre la función de autocovarianza:

$$\gamma(0) = 5$$

y

$$\gamma(1) = 4, \quad \gamma(2) = 3,$$

$$\gamma(3) = 2, \quad \gamma(4) = 1.$$

A continuación se calcularán los coeficientes $\{\phi_{nk}\}$ para $n = 1, 2, 3, 4$. El punto inicial es

$$v_0 = \gamma(0) = 5.$$

• $n = 1$.

$$\phi_{11} = \frac{\gamma(1)}{\gamma(0)} = \frac{4}{5}.$$

$$\mathbf{v}_1 = v_0[1 - \phi_{11}^2] = 5 \left[1 - \left(\frac{4}{5} \right)^2 \right] = \frac{9}{5}.$$

• $n = 2$.

$$\begin{aligned} \phi_{22} &= \left[\gamma(2) - \sum_{j=1}^{2-1} \phi_{2-1j} \gamma(2-j) \right] v_{2-1}^{-1} \\ &= \left[\gamma(2) - \sum_{j=1}^1 \phi_{1j} \gamma(2-j) \right] v_1^{-1} \\ &= \left[3 - \frac{4}{5}(4) \right] \left(\frac{9}{5} \right)^{-1} \\ &= \left[-\frac{1}{5} \right] \left(\frac{5}{9} \right) \\ &= -\frac{1}{9} \end{aligned}$$

$$\phi_{21} = \phi_{11} - \phi_{22}[\phi_{11}] = \frac{4}{5} - \frac{-1}{9} \left(\frac{4}{5} \right) = \frac{40}{45} = \frac{8}{9}.$$

Así, para $n = 2$,

$$\phi_{22} = -\frac{1}{9}, \quad \phi_{21} = \frac{8}{9}.$$

$$v_2 = v_1[1 - \phi_{22}^2] = \frac{9}{5} \left[1 - \left(\frac{-1}{9} \right)^2 \right] = \frac{9}{5} \left[\frac{80}{81} \right] = \frac{16}{9}.$$

• $n = 3$.

$$\begin{aligned} \phi_{33} &= \left[\gamma(3) - \sum_{j=1}^{3-1} \phi_{3-1j} \gamma(3-j) \right] v_{3-1}^{-1} \\ &= \left[\gamma(3) - \sum_{j=1}^2 \phi_{2j} \gamma(3-j) \right] v_2^{-1} \\ &= [\gamma(3) - [\phi_{21} \gamma(2) + \phi_{22} \gamma(1)]] v_2^{-1} \\ &= \left[2 - \left[\frac{8}{9}(3) + \frac{-1}{9}(4) \right] \right] \left(\frac{16}{9} \right)^{-1} \\ &= \left(\frac{-2}{9} \right) \left(\frac{9}{16} \right) \\ &= -\frac{1}{8} \end{aligned}$$

Después de evaluar ϕ_{33} , se está en posibilidad de calcular ϕ_{31} y ϕ_{32} :

$$\begin{aligned} \begin{bmatrix} \phi_{31} \\ \phi_{32} \end{bmatrix} &= \begin{bmatrix} \phi_{21} \\ \phi_{22} \end{bmatrix} - \phi_{33} \begin{bmatrix} \phi_{22} \\ \phi_{21} \end{bmatrix} \\ &= \begin{bmatrix} 8/9 \\ -1/9 \end{bmatrix} - \left(\frac{-1}{8}\right) \begin{bmatrix} -1/9 \\ 8/9 \end{bmatrix} \\ &= \begin{bmatrix} 8/9 \\ -1/9 \end{bmatrix} + \begin{bmatrix} -1/72 \\ 1/9 \end{bmatrix} = \begin{bmatrix} 63/72 \\ 0 \end{bmatrix} = \begin{bmatrix} 7/8 \\ 0 \end{bmatrix} \end{aligned}$$

Así, para $n = 3$ se tiene

$$\phi_{31} = \frac{63}{72}, \quad \phi_{32} = 0, \quad \phi_{33} = -\frac{1}{8},$$

y

$$v_3 = \frac{16}{9} \left[1 - \left(\frac{-1}{8}\right)^2 \right] = \frac{16}{9} \left[\frac{63}{64} \right] = \frac{7}{4}.$$

• $n = 4$.

$$\begin{aligned} \phi_{44} &= \left[\gamma(4) - \sum_{j=1}^{4-1} \phi_{4-1j} \gamma(4-j) \right] v_{4-1}^{-1} \\ &= \left[\gamma(4) - \sum_{j=1}^3 \phi_{3j} \gamma(4-j) \right] v_3^{-1} \\ &= [\gamma(4) - [\phi_{31}\gamma(3) + \phi_{32}\gamma(2) + \phi_{33}\gamma(1)]] v_3^{-1} \\ &= \left[1 - \left[\frac{63}{72}(2) + 0(3) + \frac{-1}{8}(4) \right] \right] \left(\frac{7}{4}\right)^{-1} \\ &= \left[1 - \left[\frac{63}{36} - \frac{1}{2} \right] \right] \left(\frac{4}{7}\right) \\ &= \left[1 - \frac{45}{36} \right] \left(\frac{4}{7}\right) = -\frac{1}{4} \left(\frac{4}{7}\right) = -\frac{1}{7}, \end{aligned}$$

y ahora se pueden evaluar ϕ_{41} , ϕ_{42} y ϕ_{43} :

$$\begin{aligned} \begin{bmatrix} \phi_{41} \\ \phi_{42} \\ \phi_{43} \end{bmatrix} &= \begin{bmatrix} \phi_{31} \\ \phi_{32} \\ \phi_{33} \end{bmatrix} - \phi_{44} \begin{bmatrix} \phi_{33} \\ \phi_{32} \\ \phi_{31} \end{bmatrix} \\ &= \begin{bmatrix} 7/8 \\ 0 \\ -1/8 \end{bmatrix} - \frac{-1}{7} \begin{bmatrix} -1/8 \\ 0 \\ 7/8 \end{bmatrix} \\ &= \begin{bmatrix} 7/8 \\ 0 \\ -1/8 \end{bmatrix} + \begin{bmatrix} -1/56 \\ 0 \\ 1/8 \end{bmatrix} \\ &= \begin{bmatrix} 48/56 \\ 0 \\ 0 \end{bmatrix} \end{aligned}$$

Luego, para $n = 4$ se tiene

$$\phi_{44} = -\frac{1}{7}, \quad \phi_{43} = 0, \quad \phi_{42} = 0, \quad \phi_{41} = \frac{48}{56},$$

y

$$v_4 = v_3[1 - \phi_{44}^2] = \frac{7}{4} \left[1 - \left(\frac{-1}{7} \right)^2 \right] = \frac{7}{4} \left[\frac{48}{49} \right] = \frac{12}{7}.$$

Conclusión

En este capítulo se ha estudiado la construcción de pronósticos para series de tiempo estacionarias. Este problema, de naturaleza eminentemente estadística, fue abordado desde una perspectiva geométrica, en la cual la idea de proyección ortogonal en un espacio vectorial desempeña un papel central, y se enfatizó su relación con el sistema de ecuaciones normales. El estudio mostró como dicho sistema, bien conocido en el análisis de experimentos o problemas de regresión, es el núcleo de la construcción de pronósticos. En efecto, la caracterización de un coeficiente de proyección dada en el Teorema 4.1, permitió establecer el procedimiento recursivo para evaluar los coeficientes de una proyección; este resultado se presentó formalmente en el Teorema 6.1, y su implementación se ilustró mediante un Ejemplo específico en la Sección 9. Finalmente, es oportuno mencionar que existe otro

método recursivo de pronóstico que puede aplicarse a series que no son estacionarias. El costo de esta mayor generalidad es una mayor carga computacional que la necesaria para implementar el procedimiento de Durbin–Levinson. Dicho procedimiento alternativo, denominado algoritmo de innovaciones, se fundamenta en el proceso de ortogonalización de Gram–Schmidt y ha sido estudiado en detalle en Flores–Aguilar (1999).

Capítulo 5

Epílogo

En esta trabajo se han estudiado dos problemas fundamentales en el análisis de series de tiempo, a saber, la construcción de una prueba de significancia no paramétrica para juzgar la bondad de ajuste de un modelo a los datos muestrales, y el cálculo recursivo de pronósticos para observaciones futuras. La exposición formal inició presentando la idea de proceso (débilmente) estacionario en el Capítulo 2, continuando con exposición breve sobre filtros lineales que sirvió de preámbulo para la introducción de la clase de modelos conocida como la familia de procesos ARMA (p, q) . Dentro de este conjunto, puede buscarse un proceso que describa adecuadamente a los datos observados, en el sentido de que la función de autocovarianza derivada del modelo se aproxima a la que se calcula a partir de los datos. Un aspecto que es oportuno enfatizar, es que la mayoría de las series que surgen en las aplicaciones no son estacionarias pero, utilizando una transformación adecuada, con frecuencia es posible reducir su análisis al estudio de una serie estacionaria. La prueba de bondad de ajuste que se presentó en el Capítulo 3 utilizó el estadístico T_n , el cual es el número de máximos y mínimos locales que se presentan en una sucesión Y_1, Y_2, \dots, Y_n ; el análisis se desarrolló bajo el supuesto de que dichas variables tienen la misma distribución continua y son independientes, y la principal contribución técnica consistió en demostrar, mediante argumentos elementales, la normalidad asintótica del estadístico T_n , el cual es una suma de variables aleatorias que, aunque poseen la misma distribución, *no son independientes*.

Por otro lado, en el Capítulo 4 se utilizó la idea geométrica de proyección ortogonal para analizar las soluciones de las ecuaciones normales, obteniendo una

caracterización de sus soluciones que permitió establecer, de forma rigurosa, la validez del procedimiento recursivo de Durbin–Levinson para calcular pronósticos.

Finalmente, desde un punto de vista formativo, el desarrollo de este trabajo permitió conjuntar ideas estudiadas en diversas áreas, como álgebra lineal, estadística no paramétrica y modelos lineales, 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.
- Anderson, T. W., y A. M. Walker (1964), On the asymptotic distribution of the autocorrelations of a sample from a linear stochastic process, *Annals of Mathematical Statistics*, **35**, 1296–1303.
- Anderson, O. D. (1976), *Time Series Analysis and Forecasting. The Box-Jenkins Approach*, Butterworths, London.
- 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.
- Dudewicz, E., y N. Mishra, (1988), *Modern Mathematical Statistics*, Wiley, New York.
- Flores-Aguilar, J. L. (1999), Álgebra de Filtros Lineales y el Algoritmo de Innovaciones en el Análisis de Series de Tiempo, *Tesis de Maestría en Ciencias en Estadística Experimental*, Universidad Autónoma Agraria Antonio Narro, Saltillo Coahuila, México.

- 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
- Hollander, M. y D. A. Wolfe (1981), *Nonparametric Statistical Methods*, *Wiley*, New York
- Hoffman, K. y R. Kunze. (1973) *Algebra Lineal*, *Prentice Hall Hispanoamericana*, México D. F.
- Lehmmann, 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.
- Randles, R. H. y D. A. Wolfe (1981), *Introduction to the Theory of Nonparametric Statistics*, *Wiley*, New York.
- Searle, S. R. (1982), *Matrix Algebra Useful for Statistics*, *McGraw-Hill*, New York.
- Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, 3rd. Edition, *McGraw-Hill*, New York.
- Wei, W. W. S. (1990), *Time Series Analysis: Univariate an Multivairate Analysis*, *Addison Wesley*, Reading, Massachusetts.