



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

**“Revisión del enfoque
clásico-estocástico para el análisis
de las series de tiempo.”**

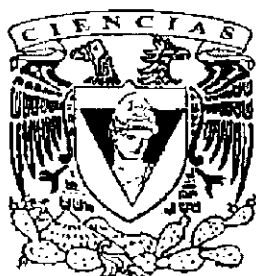
T E S I S

QUE PARA OBTENER EL TÍTULO DE

ACTUARIO

P R E S E N T A:

OLIVER MACÍAS PÉREZ



FACULTAD DE CIENCIAS
UNAM

**TUTOR:
ACT. JAIME VÁZQUEZ ALAMILLA**

2006



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD NACIONAL
AVENIDA DE
MEXICO

ACT. MAURICIO AGUILAR GONZÁLEZ
Jefe de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

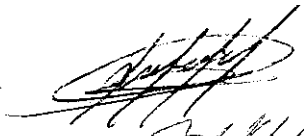
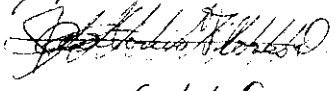



Comunicamos a usted que hemos revisado el trabajo escrito:
"Revisión del enfoque clásico-estocástico para el análisis
de las series de tiempo."

realizado por Oliver Macias Pérez

con número de cuenta 9200117-8 , quien cubrió los créditos de la carrera de: Actuaría

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis	Act. Jaime Vázquez Alamilla	
Propietario	M. en C. José Antonio Flores Díaz	
Propietario	Mat. Margarita Elvira Chávez Cano	
Suplente	M. en A.P. María del Pilar Alonso Reyes	
Suplente	Dr. Luis Antonio Rincón Solís	

Consejo Departamental de Matemáticas



Act. Jaime Vázquez Alamilla
CONSEJO DEPARTAMENTAL
DE
MATEMÁTICAS

Agradecimientos

A la querida Universidad Nacional Autónoma de México, institución a la cual orgullosamente pertenezco desde mi adolescencia y le debo toda mi formación profesional.

A la Facultad de Ciencias y todos mis profesores por transmitirme sus conocimientos con tanta pasión y esmero.

A Jaime Vázquez, mi maestro y asesor, por todo su apoyo, confianza y principalmente su amistad.

A mi mamá por todo su amor, dulzura y cuidado que de ella he recibido toda mi vida.

A mi hermano Rafael quien siempre ha estado a mi lado protegiéndome, empujando y alentando.

A mi querido abuelo Manuel por tantas risas y ternura compartida.


A mi padre y hermana Patricia.

A mi Arais, por todo lo vivido, lo soñado y los caminos que faltan por recorrer. Gracias por siempre estar a mi lado enseñándome a cada momento lo que es amor, comprensión y apoyo. Tu sonrisa, tus deseos de vivir y dedicación son la energía y motivos que me impulsan a seguir.

A mis hermanos de toda la vida, los exiliados Galo y Manuel, Iván y su familia, Carlos, Jorge, con quienes he crecido y aprendido.

A todos mis amigos de la facultad, a Pedro y Raymundo. A mis colegas del museo UNIVERSUM por todas las emotivas discusiones, ciclos de cine y encuentros deportivos.

A todos los que me faltan.

Autorizo a la Dirección General de Bibliotecas de la UNAM a digitalizar, almacenar, distribuir e imprimir el contenido de esta obra electrónica.
ACOMPAÑADO: Oliver Macías Pérez
FECHA: 3/Abril/2006
FIRMA: 

Índice general

Agradecimientos	I
Introducción	III
1. Series de tiempo	1
1.1. Modelo para las series de tiempo: procesos estocásticos	2
1.2. Componentes del modelo	7
1.3. Tipos de integración de las componentes del modelo	9
1.3.1. Criterios para detectar el tipo de integración de los componentes del modelo	11
1.4. Análisis de la tendencia	14
1.4.1. Métodos de estimación de la tendencia.	15
1.5. Análisis de la componente estacional	25
1.5.1. Desestacionalización	26
1.6. Análisis de la componente aleatoria.	30
1.6.1. Características de un proceso estocástico.	32
1.6.2. Características en el muestreo de una serie de tiempo .	33
1.6.3. Pruebas de aleatoriedad	34
2. Procesos estocásticos estacionarios	36
2.1. Definiciones básicas	36
2.2. Las funciones de autovarianza y autocorrelación	40
2.3. Tres procesos importantes: de ruido y ruido blanco, MA(q) y los procesos lineales	42
2.4. Procesos estacionarios ARMA(p,q)	46
2.5. Causalidad e invertibilidad en un proceso ARMA(p,q)	51
2.6. La función de autocovarianza (FAC) para los procesos causales ARMA(p, q)	56

2.6.1. Método 1	57
2.6.2. Método 2	57
3. Espacios de Hilbert	60
3.1. Espacios vectoriales con producto interior y sus propiedades .	60
3.2. Espacios de Hilbert	67
3.3. El Teorema de proyección	68
3.4. Conjuntos Ortonormales	75
4. Predicción de Procesos Estocásticos Estacionarios	82
4.1. El espacio $L^2(\Omega, \mathfrak{F}, P)$	83
4.2. Error cuadrático medio mínimo para procesos estocásticos esta- cionarios	84
4.3. Las ecuaciones de predicción	84
4.4. Algoritmo recursivos para el cálculo de los predictores lineales	86
4.4.1. El algoritmo Durbin-Levinson	87
4.4.2. El algoritmo de innovación	90
5. Apéndice A: Simulación	93
5.1. Programa 1	93
5.2. Programa 2	95
5.3. Programa 3	97
5.4. Programa 4	99
5.5. Programa 5	101
5.6. Programa 6	102
5.7. Programa 7	103
Comentarios Finales	106

Introducción

En el mundo actual, el riesgo y la incertidumbre ante el futuro se presentan cada vez más frecuentemente en diversos problemas, a los que distintos agentes tratan de adaptarse de diferentes maneras. Para reducir el grado de incertidumbre se suele recurrir a la elaboración de previsiones que tratan de anticipar la evolución de algún fenómeno. El disponer, en el presente, de un conocimiento sobre el futuro, aunque sea de forma aproximada, facilita la toma de decisiones en las que se incurre cuando se pretende modificar alguna determinada realidad.

Es en este marco de referencia donde se encuadra al *análisis de las series de tiempo*, como una de las técnicas más habituales que se utilizan en la previsión de los fenómenos de cualquier naturaleza.

La serie temporal, dada por los datos cuantitativos de una variable recogidos en periodos regulares de tiempo, constituye un punto de partida fundamental para el estudio de un determinado fenómeno y la elaboración de previsiones.

Las series de tiempo se encuentran presentes prácticamente en todas las disciplinas, encontrándose gran número de ejemplos en áreas como ingeniería, ciencias, sociología y economía.

Pero, los datos tratados directamente, pueden ser de difícil interpretación e inducir con ello a decisiones equivocadas. Se hace necesario, entonces, algún tipo de tratamiento o análisis de los mismos que ayuden a extraer conclusiones más fiables.

Sin embargo, la sola intención de trazar conclusiones para la toma de decisiones en fenómenos que se consideran aleatorios podría pensarse demasiado ambiciosa y tal vez deba dudarse si es alcanzable, por lo que, es necesario delimitar cuidadosamente el marco de trabajo, supuestos y alcances sobre los cuales se desarrolla el análisis.

Los datos de una variable en estudio a lo largo de un periodo de tiempo, se supondrán que contienen la información necesaria que encierra el mecanismo que motiva su ocurrencia y su desarrollo. Este supuesto conlleva a pensar que tal variable no es tan impredecible como se pensaba, esto es, tendrá pautas no del todo desconocidas, sino más bien regulares, las cuales se verán ocultas o escondidas por la parte aleatoria que encierra al fenómeno.

La anterior interpretación de la parte aleatoria de la serie es la visión característica del *enfoque clásico del análisis de las series de tiempo*, en donde se considera a la serie afectada por perturbaciones de carácter aleatorio atribuidas principalmente a errores de medición, y no se le considera como parte preponderante de la variable, por lo que se pretende encontrar las pautas del comportamiento de la serie eliminando tales perturbaciones.

No obstante, es evidente que al comportamiento aleatorio, impredecible y desconocido de la mayoría de las variables que dan lugar a las series de tiempo, no es plausible minimizarlo a simples errores de medición, por ejemplo, en series económicas, sino más bien reconocer o atribuirle un carácter propio e incluso de mayor relevancia a la parte aleatoria que conforma a la serie sobre las demás componentes regulares de la misma. Esta es la visión que caracteriza al *enfoque estocástico para las series de tiempo*, en el cual dado que se trabaja con variables del tipo aleatorio el modelo natural para representarlas es con la teoría de los procesos estocásticos.

Debe reconocerse que si bien estos enfoques se centran en la característica que el otro enfoque desprecia, persiguen los mismos fines, el conocimiento y entendimiento de la variable que da lugar a la serie, para proveer al investigador de las bases y herramientas necesarias para las posteriores tomas de decisiones generalmente relacionadas con el comportamiento futuro del fenómeno en cuestión.

Lo escrito en el párrafo anterior es la clave de todo el análisis de las series de tiempo, en él se resume la necesidad y los objetivos que la teoría que se desarrolla debe responder y, podrán ser atacados desde distintos ángulos pero la finalidad de estos enfoques es la misma.

Otro punto importante a resaltar es que generalmente se ha hecho la distinción entre estos enfoques no únicamente por sus propias características, sino que, además se le considera al enfoque clásico demasiado sencillo e intuitivo mientras que al enfoque estocástico se le considera más complejo pues requiere una formalidad matemática mayor.

Pues bien, dado lo anterior el presente trabajo tiene la finalidad de unificar ambos enfoques, pues como ya se mencionó, en ambos se persigue el mismo fin por lo que combinarlos resulta de gran interés para trazar conclusiones que sean lo más completas; por otro lado, lo que aquí se trata no son los principales resultados de ambos enfoques presentados por separado, sino la intención es realizar el análisis resultado de la combinación de ellos y mostrar que es posible tratarlos como un solo enfoque. No obstante, a lo largo de la tesis se dirá que algún resultado pertenece a algún enfoque, no con el fin

de separar sino para que el lector verifique que en efecto estos enfoques se complementan.

La estructura de la tesis, se presenta en la siguiente figura y es la siguiente, en el capítulo 1, *series de tiempo*, se presentan las bases y teoría necesaria sobre la cual se construye todo el modelo que engloba el análisis de la series de tiempo, asi como su vínculo natural con los procesos estocásticos, es en este capítulo donde se persigue encontrar las características regulares de la serie para tratar únicamente con la parte aleatoria.

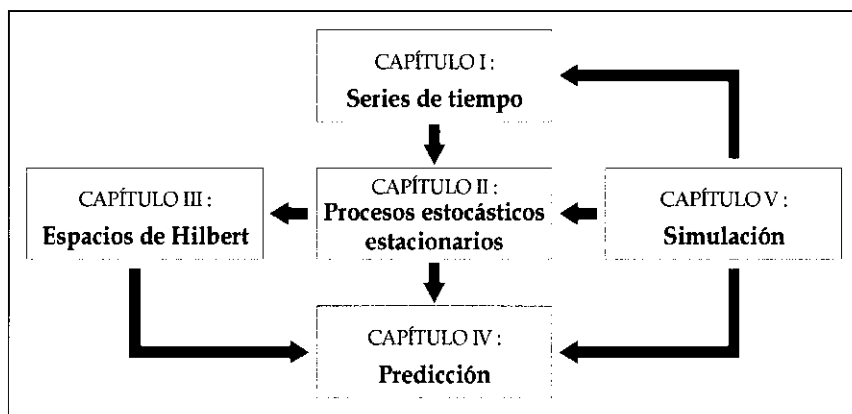


Figura 1: Estructura general de la tesis

Una vez que se tiene una estimación de la parte aleatoria de la serie, es natural necesitar modelos que la representen de la manera más adecuada, esto se realiza mediante el empleo de los procesos estocásticos, los cuales se estudian en el capítulo 2, *procesos estocásticos estacionarios*, para posteriormente trazar proyecciones acerca del comportamiento futuro de la serie.

En el capítulo 3, *espacios de Hilbert*, se presenta toda la herramienta de matemática formal que se requiere para el pronóstico de los valores de la serie, y en principio pareciera que se pierde toda relación con el tema principal de la tesis, las series de tiempo, sin embargo, este paseo por el álgebra lineal y topología resulta necesario para regresar con los resultados en la mano para el uso en el análisis de las series de tiempo.

En el capítulo 4, *predicción de procesos estocásticos estacionarios*, es donde se trata del pronóstico de las series de tiempo, usando las propiedades y características de los procesos estocásticos, pero además se debe mostrar que lo desarrollado en el capítulo 3, es aplicable al espacio vectorial donde se

desarrollan los procesos estocásticos estacionarios, el espacio L^2 , y con ello el pronóstico de la serie resultará ser inmediato.

Finalmente, en el capítulo 5, *simulación*, se presentan los algoritmos y explicación de los programas desarrollados que a lo largo de la tesis se fueron presentando mediante la simulación de series para darle un carácter más dinámico y visual a los resultados presentados, los cuales se ponen a disposición del lector.

Capítulo 1

Series de tiempo

Una serie de tiempo es un conjunto de observaciones x_t asociadas a una variable no determinista, cada una observada al tiempo t . Por tanto, una serie de tiempo es una serie de datos en los cuales se encuentra almacenada toda la información acerca del comportamiento pasado de dicha variable por algún intervalo de tiempo.

Generalmente, a la serie de tiempo se le representa por $\{x_t, t \in T_o \subset \mathbb{Z}\}$, donde se supone que el conjunto tiempo T_o sobre el cual se desarrolla la serie es un subconjunto de los números enteros, lo que significa que se estará trabajando con *series de tiempo discretas*.

Dentro de los objetivos del análisis de las series de tiempo se encuentran principalmente el conocimiento del comportamiento global del fenómeno que da lugar a la serie y el trazar pronósticos del comportamiento de la serie para la toma de decisiones. Tales objetivos, la teoría de las series de tiempo los pretende alcanzar usando únicamente a la misma variable sujeta a estudio, esto significa que *el fenómeno se explica a sí mismo*. Siendo lo anterior, la propiedad y gran diferencia que aleja al análisis de las series de tiempo de otras disciplinas, por ejemplo del análisis de regresión, teoría donde se utilizan a variables externas dentro del modelo para explicar el comportamiento de la variable en cuestión.

Es evidente que para intentar alcanzar tales objetivos, es necesario contar en primera instancia con modelos matemáticos apropiados para las series de tiempo que se ajusten a los datos de manera adecuada.

En este capítulo se siguieron las siguientes referencias bibliográficas, Ruiz Maya [10][9], Brockwell [2], Uriel [11], Rodríguez [8], Mood [5], Kendall [4] y Otero [6].

1.1. Modelo para las series de tiempo: procesos estocásticos

Dado que se trabaja con series asociadas a fenómenos no deterministas, resulta entonces natural suponer que cada observación x_t es el valor obtenido de una cierta variable aleatoria X_t definida en un espacio de probabilidad $(\Omega, \mathfrak{F}, P)$. Luego, la serie de tiempo $\{x_t, t \in T_o\}$ es la *realización* de una familia de variables aleatorias $\{X_t, t \in T_o\}$, respetándose así, la impredecible naturaleza de las observaciones.

Las consideraciones anteriores sugieren que se modelen los datos de una serie de tiempo como una realización (o parte de una realización) de un *proceso estocástico* $\{X_t, t \in T \subset \mathbb{Z}\}$ donde $T \supseteq T_o$, llegando así, a la conclusión de que los modelos apropiados que se utilizarán para las series se encuentran dentro de los procesos estocásticos, lo que convierte a esta teoría en algo central dentro del análisis de series de tiempo. Con la ayuda de la simulación es posible modelar una serie de tiempo como la que se muestra en la figura (1.1), la cual se obtiene simplemente agregando a una recta un factor no determinista.

Los procesos estocásticos son una generalización de las variables aleatorias cuando se manifiestan como función de un parámetro real.

Para comprender a un proceso estocástico, considérese $g(X; \theta)$ una función de dos argumentos, donde X es variable aleatoria definida sobre el espacio de probabilidad $(\Omega, \mathfrak{F}, P)$, y el parámetro $\theta \in \Theta$. A $g(X; \theta)$ se le conoce como *función aleatoria*.

Frecuentemente, en la función aleatoria $g(X; \theta)$ el parámetro representa el tiempo t que puede tomar valores en el campo de variación representado por $T \subset \mathbb{Z}$, siendo entonces la función aleatoria $g(X; t)$, de cuyo análisis, se deriva el concepto de *proceso estocástico*.

Cuando a los argumentos de la función aleatoria $g(X; t)$ se les permite total libertad de variación en donde están definidos, se dice que $g(X; t)$ es un proceso estocástico.

Abreviadamente, se utiliza alternativamente la notación $\{X(t)\}$ o $\{X_t\}$ para representar un proceso estocástico. Obsérvese que, también se puede considerar a un proceso estocástico como una sucesión o familia de variables aleatorias observables en cada $t \in T \subset \mathbb{Z}$. La anterior visión de un proceso estocástico es la que se adoptará como su definición, formalmente:

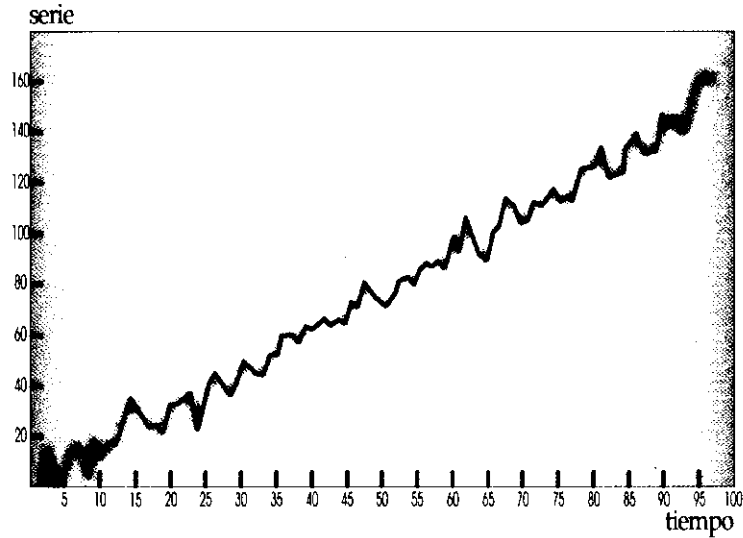


Figura 1.1: *Serie de tiempo obtenida por simulación mediante el empleo del programa 6 del apéndice.*

Definición 1.1 (Proceso estocástico) *Un proceso estocástico es una familia de variables aleatorias $\{X_t, t \in T \subset \mathbb{Z}\}$ definidas sobre un espacio de probabilidad $(\Omega, \mathfrak{S}, P)$.*

■

El proceso estocástico $\{X_t\}$ puede verse como una variable aleatoria cuando se fija el valor del parámetro t_o , se representa, entonces el proceso por $X_{t_o}(\bullet)$, y se denomina *estado* del proceso estocástico.¹

Por otro lado, cuando en un proceso estocástico $\{X_t\}$ se especifica el resultado ω_o de la variable aleatoria, se tiene, en este caso, que el proceso estocástico se convierte en $X_\bullet(\omega_o)$, siendo función de $T \subset \mathbb{Z}$, y se denomina *realización del proceso o trayectoria del proceso*.

Gráficamente la trayectoria de un proceso estocástico se ve como una serie de tiempo tal como se muestra en la figura (1.2).

Observación 1.1 *El concepto de realización de un proceso estocástico es el*

¹El estado de un proceso también se define como un proceso estocástico que sólo es función de Ω .

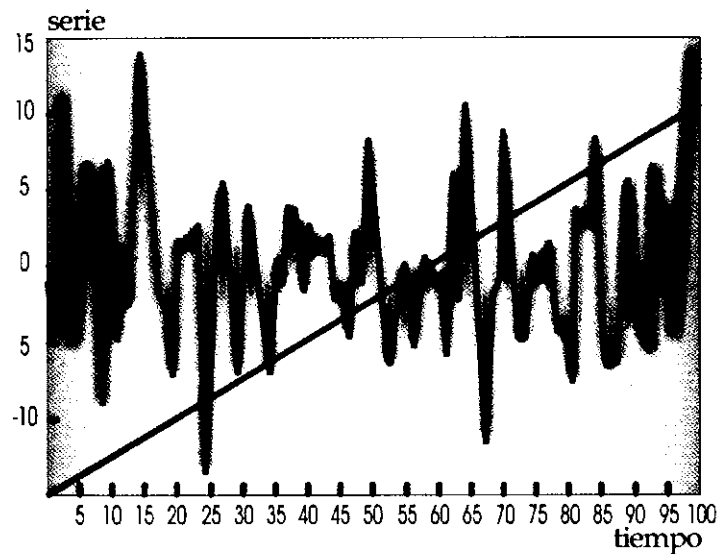


Figura 1.2: Trayectoria de un proceso estocástico

motivo por el cual se utilizan procesos estocásticos para modelar a las series de tiempo.

Observación 1.2 *El conjunto índice T sobre el cual se desarrolla el proceso estocástico $\{X_t\}$ para los fines del análisis de las series de tiempo se considera como un subconjunto de los números enteros, sin embargo, en una definición más general de un proceso estocástico, se permite total libertad sobre el conjunto índice T ; únicamente se pide que sea distinto del conjunto vacío.*

Los procesos estocásticos se clasifican, de acuerdo a la naturaleza discreta o continua de cada uno de sus argumentos como sigue,

Definición 1.2 *Se dice que $\{X_t, t \in T\}$ es un proceso estocástico de estado discreto (continuo) si las variables aleatorias X_t son discretas (continuas).*

■

Definición 1.3 (Función de distribución de un proceso estocástico $\{X_t\}$)
 Sea \mathcal{B} el conjunto de los vectores

$$\mathcal{B} = \left\{ \mathbf{t} = (t_1, \dots, t_n)' \in T^n : t_1 < t_2 < \dots < t_n, n = 1, 2, \dots \right\}$$

Entonces las **funciones de distribución** de un proceso estocástico $\{X_t, t \in T\}$ son las funciones $\{F_{\mathbf{t}}(\bullet), \mathbf{t} \in \mathcal{B}\}$ definidas por

$$F_{\mathbf{t}}(\mathbf{x}) = F(x_1, \dots, x_n; t_1, \dots, t_n) = P(X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n),$$

donde $\mathbf{x} = (x_1, \dots, x_n)' \in \mathbb{R}^n$.

■

Observación 1.3 De la definición anterior se sigue que toda función de distribución de un proceso estocástico $\{X_t, t \in T \subseteq \mathbb{Z}\}$ debe cumplir la condición de consistencia, si $m < n$ entonces

$$F(x_1, x_2, \dots, x_m; t_1, t_2, \dots, t_m) = F(x_1, x_2, \dots, x_m, \infty, \dots, \infty; t_1, t_2, \dots, t_m, \dots, t_n).$$

Frecuentemente, se desea decidir si una sucesión de variables aleatorias, con su respectivas funciones de distribución, constituyen un proceso estocástico. El siguiente teorema muestra que el criterio de decisión es la condición de consistencia de las funciones de distribución.

Teorema 1.1 (Teorema de extensión de Kolmogorov) Las funciones de distribución de probabilidad $\{F_{\mathbf{t}}(\bullet), \mathbf{t} \in \mathcal{B}\}$ son funciones de distribución de algún proceso estocástico si y sólo si para cada $n = 1, 2, \dots$, $\mathbf{t} = (t_1, \dots, t_n)' \in \mathcal{B}$ y $1 \leq i \leq n$,

$$\lim_{x_i \rightarrow \infty} F_{\mathbf{t}}(\mathbf{x}) = F_{\mathbf{t}(\mathbf{i})}(\mathbf{x}(\mathbf{i}))$$

donde $\mathbf{t}(\mathbf{i})$ y $\mathbf{x}(\mathbf{i})$ son los vectores formados por las $(n - 1)$ componentes resultantes al eliminar las i -ésimas componentes de \mathbf{t} y \mathbf{x} respectivamente.

Si bien se trabaja con series con un comportamiento global no determinista y por consecuencia se adoptan modelos no deterministas para ellas, no todo en la serie de tiempo es aleatorio. A las series de tiempo aquí se considerarán integradas por una parte regular, sistemática o determinista, y

por otra de carácter aleatorio, carente de regularidad y por tanto difícil de conocer su patrón de comportamiento.

La parte regular o determinista de la serie la conforman dos de las tres componentes que integran el modelo para las series de tiempo denominadas *componente tendencia* y *componente estacional*², las cuales serán representadas por m_t (una función de cambio lento), y por s_t (una función periódica con periodo conocido d), respectivamente. Mientras que se representará por Y_t a la *componente aleatoria*³ que está presente en toda serie de tiempo y, no se debe perder de vista que es esta parte aleatoria de la serie lo que obliga a considerar a la serie global aleatoria y a modelarla como la realización de un proceso estocástico.

De esta forma el modelo para la serie de tiempo $\{x_t, t \in T_o \subset \mathbb{Z}\}$ es mediante la realización de un proceso estocástico $\{X_t\}$ de la forma:

$$X_t = X(m_t, s_t, Y_t) \quad (1.1)$$

siendo esta expresión el *modelo general para representar a una serie de tiempo*.

A este modelo se le denomina *modelo clásico-estocástico* por ser el resultado de la combinación de los dos modelos usados en la literatura de las series de tiempo, a saber: el enfoque clásico o de descomposición y el enfoque estocástico. El primero de ellos, supone a la serie de tiempo formada por cuatro componentes: componente tendencia, componente estacional, componente ciclo y un componente aleatorio, en este enfoque se da gran importancia a la estimación de los tres primeros componentes antes citados, mientras que al componente aleatorio se le considera sin relevancia, pues suponen que existe sólo por errores de medición. Por su parte el enfoque estocástico, supone a la serie de tiempo formado por las mismas componentes que el enfoque clásico, excepto que no se considera al componente ciclo. En este enfoque se pone toda la atención en analizar a la componente aleatoria de la serie y se desprecia la importancia de las demás componentes.

Por tales motivos en esta tesis se pretende encontrar un equilibrio en estas dos corrientes dando la adecuada importancia a todas las componentes del modelo, utilizando los resultados más importantes de ambos enfoques y realizando la respectiva unificación de ellos para obtener así, un análisis de las series de tiempo lo más completo posible.

²Veáse la siguiente sección

³*Idem*

Para los fines del análisis de las series de tiempo es necesario separar las componentes que integran a la serie, esto es, separar los movimientos sistemáticos de las fluctuaciones irregulares, este es el trabajo que se desarrolla en el siguiente capítulo.

1.2. Componentes del modelo

Desde una perspectiva teórica y con la finalidad de comprender la evolución de un fenómeno determinado, el *modelo clásico-estocástico* (1.1) considera el comportamiento de una serie de tiempo como el resultado de la integración de tres componentes fundamentales: *la tendencia, la componente estacional y el componente de ruido aleatorio*.

A pesar de que suponer a las series de tiempo integradas por componentes es ampliamente aceptado y utilizado por muchos autores, las definiciones que cada uno de ellos utiliza para las componentes de la serie no son las mismas. Es por esto que, a continuación se aclara lo que en esta tesis se entenderá por cada una de las componentes que conforman a la serie de tiempo.

La tendencia

Se considera *tendencia al movimiento suave y regular de la serie de tiempo a largo plazo*. Es una componente que reviste gran interés pues, es quien refleja la dirección del movimiento de la variable en cuestión. Por tanto, con la tendencia es posible detectarse si, a largo plazo, la serie adopta una marcha persistente, ya sea de crecimiento, decrecimiento o estabilidad, aunque para observar esto generalmente se necesita contar con un horizonte temporal bastante amplio.

En muchas ocasiones, la posible predicción que pueda realizarse acerca de esta componente suele ser el objetivo principal en el análisis de las series de tiempo.

La componente estacional

La *componente estacional, si existe, se refiere al comportamiento, o patrón sistemático que es posible observar que se repite en la serie cada cierto número de observaciones*, en otras palabras, la componente estacional es un comportamiento de tipo periódico que se encuentra presente en la serie de tiempo.

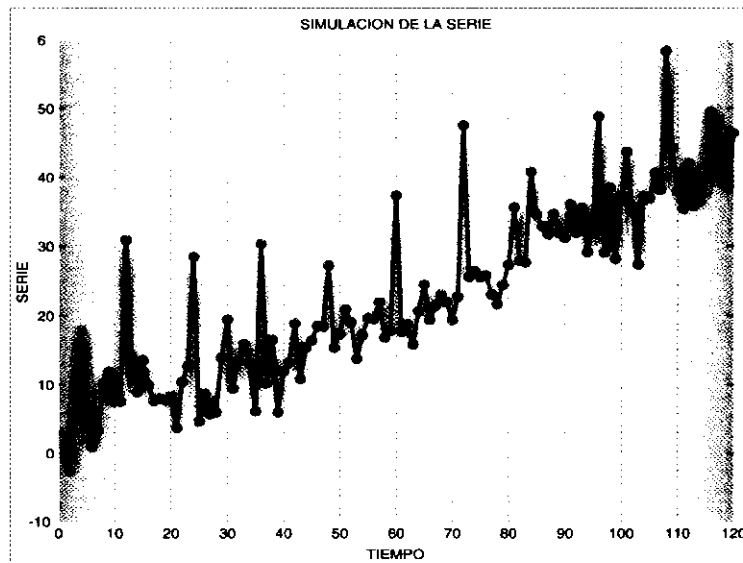


Figura 1.3: *Serie simulada con un comportamiento estacional obtenida mediante el empleo del programa 7 del apéndice.*

Generalmente, es a través de la representación gráfica de la serie cuando es factible determinar si está presente o no la componente estacional, sin embargo, hay casos en donde no es posible detectarla únicamente con la inspección visual, por lo que otros métodos deben realizarse para poder concluir con más fiabilidad sobre la presencia de la componente estacional en la serie.

En la figura (1.3) se muestra una serie de tiempo generada mediante simulación en donde se le ha forzado a tener un comportamiento estacional que se aprecia claramente.

La presencia de la componente estacional en muchas series es de manera inherente, por ejemplo en series de tiempo cuyos valores se encuentren relacionados con el turismo de forma mensual, tales como el número de reservaciones hoteleras, número de pasajeros de avión, etc., presentan sistemáticamente un incremento durante los meses de verano, el cual es causado únicamente porque es en esos meses cuando la mayoría de la población cuenta con vacaciones.

Un ejemplo más de la presencia de esta componente causada por factores

institucionales se encuentra en series de tiempo vinculadas con la educación, cuyas fluctuaciones intraanuales estarán fuertemente afectadas por el hecho institucional de que el año escolar se distribuye desde septiembre hasta junio.

Como puede apreciarse, un adecuado análisis de la componente estacional, es de vital importancia para los fines del análisis de las series de tiempo, ya que de no tomar en cuenta a la componente estacional o con un deficiente análisis, seguramente las predicciones que se realicen serán erróneas, y en consecuencia toda decisión basada en tales predicciones.

El componente aleatorio

El último componente que conforma a la serie de tiempo es el denominado *componente aleatorio* que está constituido por las variaciones de la serie cuyas leyes son desconocidas. Se caracteriza, por tanto, porque no responde a un comportamiento sistemático o regular, y en consecuencia no es posible su predicción.

Es debido a este componente que el modelo utilizado para el análisis de las series de tiempo, se apoye en la teoría de los procesos estocásticos, para así respetar la naturaleza aleatoria de las observaciones.

A pesar de que, el desconocimiento del mecanismo que genera a esta componente, vislumbra un panorama totalmente desalentador de cara la predicción de la serie, se verá más adelante que con los supuestos adoptados y con la ayuda de la teoría de los procesos estocásticos es posible realizar predicciones adecuadas del comportamiento de la serie.

En una serie de tiempo no tienen que estar presentes todas las componentes que supone el modelo clásico-estocástico, sin embargo, la componente de ruido aleatorio deberá suponerse siempre presente, pues se está trabajando con series de tiempo que no son deterministas, por lo que estarán afectadas por alguna perturbación de carácter aleatorio.

1.3. Tipos de integración de las componentes del modelo

En la sección anterior se vio una a una las tres componentes que integran al modelo clásico-estocástico para las series de tiempo, sin embargo, no se dijo nada acerca de la manera en que lo hacen.

Resulta conveniente suponer que la integración de las componentes es de manera funcional, es decir, a través de alguna función f . Naturalmente esto conduce a que el modelo para la serie de tiempo pueda adoptar múltiples formas, una por cada función f , donde cada una de ellas será más o menos sofisticado dependiendo que tan compleja sea la función adoptada.

De acuerdo a lo anterior, el modelo general (1.1) para la serie de tiempo $\{X_t\}$ bajo el esquema clásico-estocástico se convierte en el siguiente:

$$X_t = f(m_t, s_t, Y_t) \quad (1.2)$$

donde

- f , es cualquier función,
- m_t , es la tendencia,
- s_t , es el componente estacional,
- Y_t , es el componente de ruido aleatorio.

Generalmente, en la literatura del análisis de las series de tiempo, los tipos de integración que se utilizan para las componentes de la serie, y que también serán usados en el desarrollo de la tesis, son los siguientes:

$$\begin{aligned} \textit{integración aditiva} & : X_t = f(m_t, s_t, Y_t) = m_t + s_t + Y_t \\ \textit{integración multiplicativa} & : X_t = f(m_t, s_t, Y_t) = m_t \times s_t \times Y_t \end{aligned}$$

Las razones por las cuales sólo se utilizan estas dos integraciones radican en su simplicidad, pues involucran un tratamiento más sencillo, y por tanto, resultan ser más operativas.

La integración aditiva se utiliza cuando la variabilidad de la serie no muestra demasiados cambios, y es de esperarse que así continúe a lo largo del tiempo.

Por su parte, la integración multiplicativa resulta más adecuada en los casos en que la variabilidad de la serie presente notables cambios a través del tiempo. Igualmente es adecuada la integración multiplicativa en los casos en que los cambios en porcentajes reflejen adecuadamente los movimientos de la serie y de sus componentes. Así por ejemplo, en una serie sobre las ventas mensuales de un determinado producto, el decir que sistemáticamente las ventas bajan un 20% en el mes de agosto, significa que la componente estacional está integrado de forma multiplicativa.

Es conveniente señalar que en el caso de que el modelo de la serie suponga la integración multiplicativa, es posible transformarla, bajo logaritmos, a una integración del tipo aditiva.

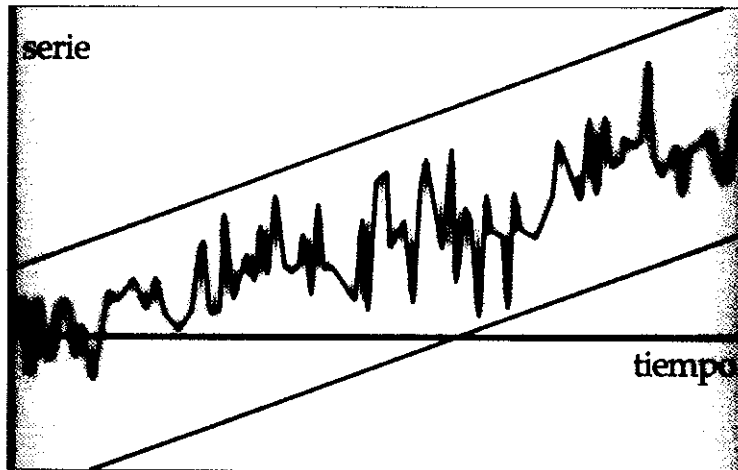


Figura 1.4: *Serie con integración aditiva en sus componentes*

1.3.1. Criterios para detectar el tipo de integración de los componentes del modelo

Aún cuando se han aceptado sólo dos maneras para la integración de las componentes del modelo para las series de tiempo, resulta necesario contar con criterios con los cuales determinar, con cierta precisión, el tipo de integración a utilizar en el modelo.

Los criterios a utilizar están basados en el comportamiento de la amplitud de las oscilaciones que en torno a su tendencia, presenta la serie de tiempo.

Mediante el gráfico de la serie

El primer criterio para detectar si las componentes se asocian de forma aditiva o multiplicativa, es mediante la inspección de la representación gráfica de la serie.

Como se señaló en la sección anterior, unas oscilaciones de la variable en torno a la tendencia más o menos de la misma amplitud representan un indicador de aditividad en las componentes (véase la figura 1.4), es decir, se admite la hipótesis de independencia entre la tendencia y las demás componentes.

Por el contrario, si la variable muestra oscilaciones cada vez mayores o menores, es decir, si la variabilidad de la serie no se mantiene constante en

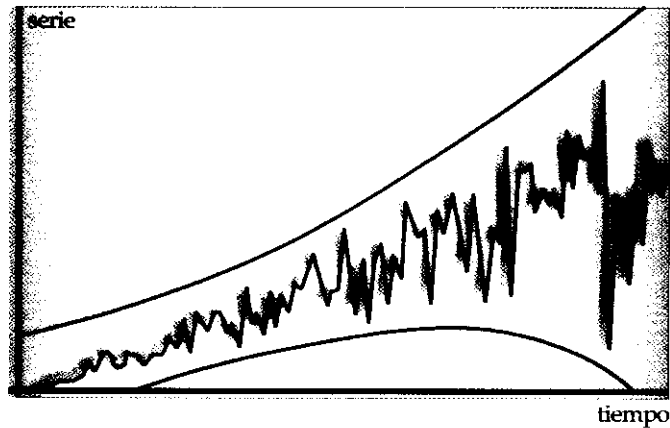


Figura 1.5: *Serie con integración multiplicativa en sus componentes*

el tiempo, un esquema multiplicativo (veáse la figura 1.5) es más adecuado. Aquí la hipótesis involucrada es la dependencia entre las componentes de la serie.

Si a través de la inspección gráfica de la serie no es posible detectar de forma concluyente cual es el tipo de integración de las componentes que más se corresponde con los datos, es necesario utilizar otro método.

Mediante el análisis del gráfico: media-dispersión.

Si con la inspección del gráfico de la serie no es posible determinar la asociación de las componentes de la misma, es seguramente debido a que las pautas que revelan esa información no se encuentran visibles.

Por ello, el *método del gráfico: media-dispersión*, propone “filtrar” la serie con el fin de separar la información útil de la que no lo es, y, después de conseguido esto, mediante un gráfico detectar la vinculación de los componentes.

Ese “filtro”, se consigue estudiando el comportamiento de la serie no en las mismas unidades de tiempo en las que se desarrolla, sino en periodos de tiempo más grandes. En otras palabras, la información sobre la integración de las componentes, el *método del gráfico: media-dispersión* la obtiene analizando el comportamiento de la serie desde una óptica más amplia que sobre la cual se va desarrollando normalmente la serie.

Estas ideas se plasman formalmente de la siguiente manera

Considere la serie de tiempo $\{X_t, t \in T_o \subseteq \mathbb{Z}\}$, luego, el *método del gráfico: media-dispersión*, pide dividir al conjunto T_o en d subconjuntos o periodos $\{T_i \subseteq T_o, i = 1 \dots d\}$ cada uno de los cuales contiene k observaciones de la serie (a las k observaciones se les conocen como estaciones). La partición del conjunto T_o , debe ser tal que en cada periodo T_i se espere que **la serie se vea afectada por los mismos factores y esperar así, un comportamiento similar para cada T_i** . Entre estos factores debe contemplarse a la componente estacional, por lo que en cada T_i también debe esperarse que la serie se vea afectada de igual manera por dicha componente.

Por ejemplo, cuando se trabaja con series de tipo mensual o trimestral, una posible partición del conjunto T_o , es en conjuntos T_i , que contengan cada uno de ellos doce y cuatro observaciones -estaciones- respectivamente, es decir, los T_i , contendrán los registros del comportamiento anual de la serie.

Por otro lado, de acuerdo a que la tendencia es un componente a largo plazo, no deben presentarse grandes variaciones de esta componente en cada periodo T_i , por lo que, es apropiado usar a la media de los datos pertenecientes a cada T_i como representantes de la tendencia que presenta la serie en cada estación T_i .

De la misma forma, la desviación estándar de las estaciones contenidas en cada periodo T_i , respecto a la media, constituye una medida sobre la dispersión o variabilidad que la serie presenta, respecto al componente tendencia. Por lo que, los movimientos correspondientes a los componentes estacional y aleatorio se verán reflejados en el valor de la desviación estándar de los valores de la serie en cada T_i .

Con base en esto, es posible calcular d medias y d desviaciones estándar, y construir el *gráfico media-dispersión* colocando en el eje de las abscisas los valores de las d medias, y en el eje de las ordenadas los valores correspondientes a las d desviaciones estándar..

Finalmente, si en el *gráfico media-dispersión*, es posible ver unos puntos más o menos alineados en torno a una recta, esto significa que a medida que el nivel de la tendencia crece, los demás componentes de la serie se mueven directa o inversamente proporcional (principalmente el componente estacional) a la tendencia; de donde, se hace evidente la existencia de una clara dependencia entre los componentes, y por lo tanto se debe concluir que las componentes se asocian de manera multiplicativa.

Si en cambio, en el *gráfico media-dispersión*, se concluye que los puntos se distribuyen aleatoriamente, lo que significaría la existencia de indepen-

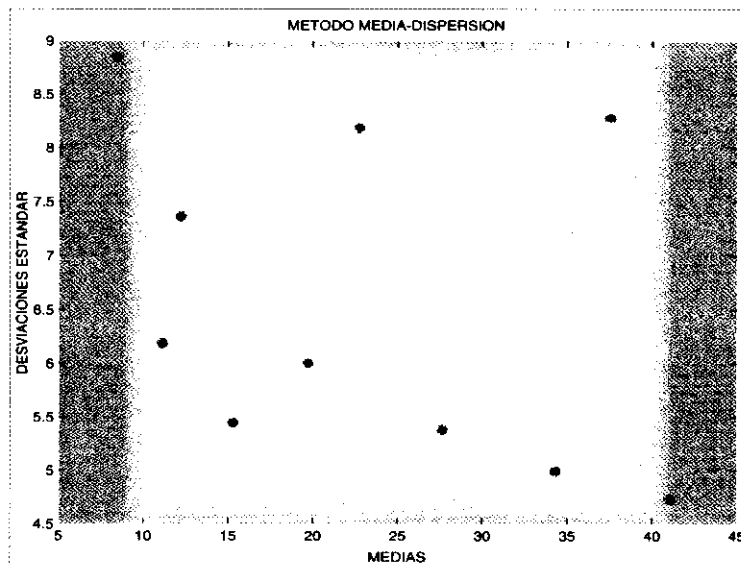


Figura 1.6: Resultado del método Media-Dispersión aplicado a la serie de la figura (1.3), utilizando simulación con la ayuda del programa 2 del apéndice.

dencia entre las componentes, en cuyo caso, la asociación aditiva para los componentes de la serie es la adecuada.

En 1.6 se muestra el resultado de aplicar el *método media dispersión* a la serie de tiempo de la figura (1.3) obtenida por simulación la cual además de tener un comportamiento estacional fue generada mediante un esquema aditivo como lo confirma la aplicación de éste método.

1.4. Análisis de la tendencia

Con el análisis de las series de tiempo se pretende extraer la componente regular que se observa en el pasado de la serie, obtener el mecanismo que la genera, para tener un mejor conocimiento de la misma en el tiempo y, una vez que se ha eliminado toda la parte determinista de la serie, proceder a estimar la parte aleatoria, y es en esta etapa donde los procesos estocásticos entran en acción con toda su fuerza como se verá en el capítulo 3. Cabe mencionar que si se intenta realizar un pronóstico, éste se hace bajo el supuesto de que las condiciones estructurales que conforman el fenómeno objeto de es-

tudio permanecen constantes, con la finalidad de predecir el comportamiento futuro, reduciendo de esta forma, la incertidumbre en la toma de decisiones.

Como ya se ha señalado, el enfoque clásico sugiere que el comportamiento de la serie es el resultado de la integración de tres componentes y, una vez que estos sean conocidas o estimadas, será entonces mediante su multiplicación o adición cuando se obtengan estimaciones sobre la variable que define a la serie de tiempo.

Es en *el análisis de la tendencia* cuando se estudian métodos con un único objetivo: *la estimación de la componente tendencia de la serie de tiempo*. Esto es con el fin de conocer las pautas de comportamiento a largo plazo (en algunos casos a corto plazo) de la variable.

Bajo el enfoque clásico-estocástico, es la tendencia la componente central de la serie de tiempo, quien indica y dirige su movimiento. Así habrán casos de series que presenten un comportamiento en la tendencia de forma creciente, decreciente, constante o también es posible encontrarse con series que en alguno de sus puntos manifiesten un cambio de tendencia.

De acuerdo a lo mencionado en la sección 1.1, la tendencia presenta características de alisamiento y continuidad, esto es, movimientos suaves y persistentes. Es por esto que en la mayoría de los análisis de las series de tiempo para estimar a la tendencia se utilizan funciones matemáticas dependientes del tiempo que cuentan con esas características, siendo más comúnmente utilizados los polinomios de grado k y la función exponencial.

Si bien estimar a la tendencia por medio de una función como las anteriores es lo más frecuente, también es posible realizar una estimación de ella mediante el empleo de medias móviles, que como se verá más adelante, su finalidad es suavizar a la serie eliminando las oscilaciones de la variable en el corto plazo.

Por tanto, en los siguientes apartados de esta sección se estudiarán métodos de estimación para los diferentes tipos de tendencia que pueda presentar una serie de tiempo.

1.4.1. Métodos de estimación de la tendencia.

En el estudio de la tendencia se emplean dos enfoques distintos, denominados, enfoque *global* y enfoque *local*, teniéndose así, un método de estimación para cada enfoque de la tendencia.

Con el enfoque global, la estimación del modelo que se haya propuesto para la tendencia se mantendría constante para todo el periodo. Igualmente,

la función que se utilice para estimar la tendencia se considera válida para todos los momentos del tiempo y, además otorga el mismo peso a todas las observaciones para la determinación de la tendencia. Este enfoque busca captar los aspectos más permanentes en la evolución de la variable, por lo que resulta adecuado emplear este enfoque cuando se requiera realizar predicciones a mediano y largo plazo.

En los métodos bajo el enfoque local, o bien solamente se utiliza una parte de las observaciones en la determinación de la tendencia en cada momento, o bien el peso que tiene cada observación es decreciente, al alejarse del momento de referencia. El enfoque local es adecuado cuando se trata de realizar previsiones a corto plazo, ya que, como la estimación de la tendencia va cambiando a medida que se añaden nuevas observaciones, los métodos se adaptan mejor a las circunstancias cambiantes.

Resulta conveniente que antes de proceder a la exposición de los distintos métodos de estimación de la tendencia, se formulen un conjunto de hipótesis de validez general tanto en los análisis de tendencia como de la componente estacional. Las hipótesis requeridas son las siguientes:

- a) Se asume una cierta estabilidad en la estructura del fenómeno estudiado, de tal forma que, si la serie presentara cambios bruscos de tendencia, sería más adecuado dividir el periodo de observación y estudiar cada parte por separado.
- b) Se asume que, dentro del año, la influencia de la componente estacional sobre el nivel medio de la serie es nulo, es decir, a lo largo de un año los efectos positivos se compensan con los negativos, resultando nulo el efecto neto. Esta consideración, además de que parece bastante razonable, favorecerá las interpretaciones.

1. Método bajo el enfoque global o método analítico

Para la aplicación del método analítico se requiere previamente que la componente estacional de la serie, en caso de estar presente, sea eliminada⁴, pues de lo contrario las oscilaciones que esta componente provoca perturbarían la identificación de la tendencia.

Este método trata de modelizar la tendencia m_t mediante funciones matemáticas dependientes del tiempo. Es decir, $m_t = m(t)$ donde en principio la función $m(t)$ puede adoptar múltiples expresiones unas más sencillas que otras,

⁴Los distintos métodos para eliminar la estacionalidad se detallan en 1.5

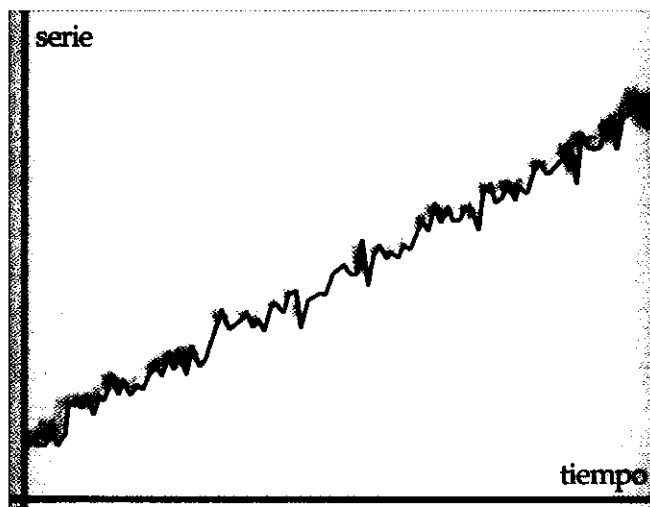


Figura 1.7: *Aspecto de una serie con tendencia lineal.*

y, se tratará de buscar aquella función que proporcione un mejor ajuste y cuyo tratamiento no resulte muy complicado.

Las funciones que se utilizarán son los polinomios de grado k en caso de que la serie cuente con una integración en sus componentes de forma aditiva, mientras que, en el caso de integración multiplicativa será la función exponencial la que estime a la tendencia.

Considérese la serie de tiempo $\{X_t\}$, con integración aditiva en sus componentes y sin presencia de la componente estacional, es decir:

$$X_t = m_t + Y_t$$

para este caso, el método analítico sugiere que la tendencia se represente genéricamente mediante un polinomio de grado k del tiempo del siguiente tipo

$$m_t = \sum_{j=0}^k b_j t^j \quad (1.3)$$

obsérvese que la estimación de la tendencia mediante (1.3) es una estimación determinista para los parámetros $\{b_j \in \mathbb{R}, j = 1..k\}$.

Asignando valores a k en (1.3) se obtienen modelos especiales para la serie de tiempo $\{X_t\}$. Cuando $k = 0$ se tiene que un modelo de media constante:

Definición 1.4 *La serie de tiempo $\{X_t\}$ es estimado por un modelo de tendencia media constante si*

$$X_t = b_0 + Y_t$$

donde $b_0 \in \mathbb{R}$ y Y_t es el componente aleatorio.

■

Es este modelo de tendencia constante el adecuado cuando los datos de la serie oscilan aleatoriamente con respecto a un valor medio.

Si $k = 1$, se obtiene un modelo de tendencia lineal:

Definición 1.5 *La serie de tiempo $\{X_t\}$ es estimado por un modelo de tendencia lineal si*

$$X_t = b_0 + b_1t + Y_t$$

donde $b_0, b_1 \in \mathbb{R}$ y Y_t es el componente aleatorio.

■

Cuando $k = 2$ el modelo es denominado de tendencia parabólica o de segundo grado:

Definición 1.6 *La serie de tiempo $\{X_t\}$ es estimado por un modelo de tendencia parabólica si*

$$X_t = b_0 + b_1t + b_2t^2 + Y_t$$

donde $b_0, b_1, b_2 \in \mathbb{R}$ y Y_t es el componente aleatorio.

■

Generalmente, a pesar de que el grado del polinomio para estimar a la tendencia no esta limitado, lo más usual es que sea alguno de los anteriores. Sin embargo, aún falta contar con criterios que señalen cuando suponer para la tendencia un modelo de media constante, lineal o parabólica. Algunos criterios se detallan a continuación.

El primer criterio para la detectar el grado del polinomio es mediante el gráfico de la serie, que la mayoría de los casos provee de información sobre el tipo de polinomio que es el indicado.

Resulta apropiado utilizar un modelo de media constante cuando en el gráfico de la serie se observe que la nube de puntos se distribuye en torno

a una recta paralela al eje de las abscisas. Por su parte, el modelo lineal es conveniente cuando gráficamente la nube de puntos se sitúan más o menos alineados en torno a una recta creciente o decreciente. Finalmente, en un modelo de tendencia parabólica, como su nombre lo indica, visualmente la nube de puntos correspondiente a los datos de la serie se agrupan alrededor de una parábola.

Por tanto, el distinguir entre un modelo de tendencia constante y uno de tendencia lineal es bastante claro gráficamente. No obstante, mediante el gráfico de la serie no siempre es posible decidir de manera concluyente entre un modelo de tendencia lineal y uno de tendencia parabólica, esto sucede cuando la nube de puntos se agrupa alrededor de algo que podría ser una recta o una parábola.

Para resolver este problema, es posible recurrir al gráfico de las primeras diferencias de la serie.

Definición 1.7 *El operador diferencia ∇ que a la serie de tiempo $\{X_t\}$ le asocia la primera diferencia como sigue:*

$$\nabla X_t = X_t - X_{t-1}.$$

■

En el caso de una tendencia lineal la primera diferencia de la serie oscila en torno a un valor medio constante, por lo que el gráfico de la primera diferencia se asemeja al gráfico de tendencia media constante.

Lo anterior se verifica fácilmente, pues partiendo de un modelo con tendencia lineal, se tiene que

$$\begin{aligned} X_t &= b_0 + b_1 t + Y_t \\ X_{t-1} &= b_0 + b_1(t-1) + Y_{t-1} \end{aligned}$$

y la primera diferencia adopta la siguiente expresión

$$\nabla X_t = X_t - X_{t-1} = b_1 + (Y_t - Y_{t-1})$$

Por tanto, el parámetro b_1 , pendiente en el modelo para la serie $\{X_t\}$ es el valor en torno al cual se distribuyen la nube de puntos correspondientes al gráfico de las primeras diferencias.

En cambio, en un modelo de tendencia parabólica, el gráfico correspondiente a la primera diferencia de la serie, distribuye la nube de puntos alrededor

de una recta, pues si se supone tendencia parabólica entonces,

$$\begin{aligned} X_t &= b_0 + b_1 t + b_2 t^2 + Y_t \\ X_{t-1} &= b_0 + b_1(t-1) + b_2(t-1)^2 + Y_{t-1} \end{aligned}$$

de donde

$$\nabla X_t = X_t - X_{t-1} = b_1 + b_2 + 2b_2 t + (Y_t - Y_{t-1})$$

es decir, en un modelo con tendencia parabólica es de esperar que su primera diferencia se comporte como un modelo con tendencia lineal.

Por tanto, el decidir si la tendencia se estima bajo un modelo lineal y un modelo parabólico, mediante el gráfico de las primeras diferencias de la serie, se reduce al nivel de decisión que existe entre un modelo lineal y uno de media constante y, como ya se apuntó esta decisión es bastante clara.

Una vez que ya es posible elegir entre los tres principales modelos para la tendencia de la serie de tiempo $\{X_t\}$, sólo resta la estimación de los parámetros $\{b_j\}$. Esta estimación será a través del método que siempre se utiliza para ajustar modelos lineales, el método de *mínimos cuadrados*.

Bajo el método de mínimos cuadrados la elección de los parámetros $\{b_j\}$ serán aquellos que minimicen

$$\sum_t (X_t - m_t)^2 = \sum_t \left(X_t - \sum_{j=0}^k b_j t^j \right)^2, \quad t = 1, \dots, n$$

La elección de los parámetros $\{b_j\}$ por mínimos cuadrados posee las buenas propiedades para que un estimador sea considerado como aceptable, tales como la insesgadez, la eficiencia y la consistencia, (ver Mood et al, 1974).

Es importante resaltar que lo hecho hasta el momento es bajo un esquema de integración aditiva para los componentes de la serie $\{X_t\}$. En el caso de una integración multiplicativa, se procede de la manera siguiente:

Considérese la serie de tiempo $\{X_t\}$, con integración multiplicativa en sus componentes y de nueva cuenta sin presencia de la componente estacional, es decir

$$X_t = m_t \times Y_t$$

en los esquemas multiplicativos se considera en general que la tendencia bien aproximada por una función exponencial del siguiente tipo

$$m_t = \exp \left(\sum_{j=0}^k b_j t^j \right) \quad (1.4)$$

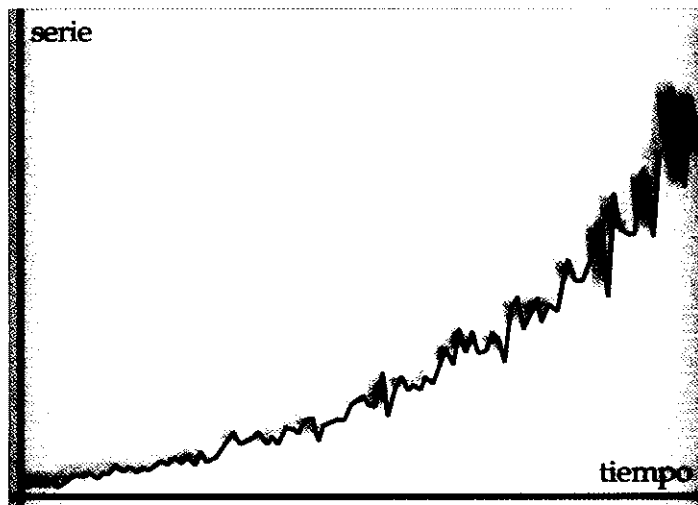


Figura 1.8: *Aspecto de una serie con tendencia exponencial*

al igual que en el caso de integración aditiva, la estimación de la tendencia mediante (1.4) es una estimación determinista para los parámetros $\{b_j \in \mathbb{R}, j = 1 \dots k\}$.

Por tanto, el modelo para la serie se ve de la siguiente forma

$$X_t = \exp\left(\sum_{j=0}^k b_j t^j\right) \times Y_t$$

$$X_t = \exp\left(\sum_{j=0}^k b_j t^j\right) \times \exp(\ln(Y_t))$$

agrupando se llega finalmente a

$$X_t = \exp\left(\sum_{j=0}^k b_j t^j + \ln(Y_t)\right) \quad (1.5)$$

Si en el modelo (1.5) se toman logaritmos neperianos se obtiene

$$\ln(X_t) = \sum_{j=0}^k b_j t^j + \ln(Y_t) \quad (1.6)$$

en donde se ve claramente que un esquema multiplicativo (1.5) se convierte en un esquema aditivo (1.6) cuyo tratamiento ya fue estudiado.

2. Método bajo el enfoque local ó método de las medias móviles

La estimación de la tendencia de una serie mediante una función matemática resulta apropiado si se observa un comportamiento continuado en los valores de la serie, ya sea de crecimiento, decrecimiento o constante. Pero, cuando el comportamiento de la serie va cambiando a lo largo del tiempo una función matemática no será conveniente, y en este caso, una hipótesis más realista es suponer que dicha función puede utilizarse para estimar a la tendencia en intervalos cortos de tiempo, de tal forma que se estimen funciones cambiantes en el tiempo. Esta es la idea que persigue e intenta conseguir *el método de medias móviles*.

Dado que la media de todos los valores de la serie (o simplemente media global), es una síntesis de la información contenida en la serie, puede utilizarse a la media global como el valor más representativo del comportamiento de la serie, siempre que su variabilidad no sea grande. Sin embargo, cuando la serie presenta tendencia con variaciones importantes en la evolución, la media global es claramente insuficiente para describir el fenómeno que se esté analizando. Para suplir estas insuficiencias es necesario introducir el concepto de *media móvil*, con el fin de que el promedio se vaya adaptando a las circunstancias cambiantes que se produzcan en la historia de la serie.

La media móvil es una media aritmética que se caracteriza porque toma un valor para cada momento en el tiempo y por que en su cálculo no entran todos los valores que conforman la historia de la serie. Para la aplicación de este método es necesario determinar su longitud u orden, es decir, el número de observaciones que intervienen en el cálculo de cada media móvil. Cuanto mayor sea la longitud, mejor se eliminarán las irregularidades de la serie, ya que al intervenir más observaciones en su cálculo se compensarán las fluctuaciones de este tipo. Por el contrario, cuando la longitud es pequeña, la media móvil refleja con mayor rapidez los cambios que puedan producirse en la evolución de la serie. Es conveniente, pues sopesar estos factores al decidir la longitud de la media móvil.

Las medias móviles que se utilizarán para este método son las denominadas *medias móviles centradas* que se caracterizan porque el valor de la media móvil se asigna al punto medio del intervalo. Si la longitud es un número par entonces se plantea el problema de que el punto central no cor-

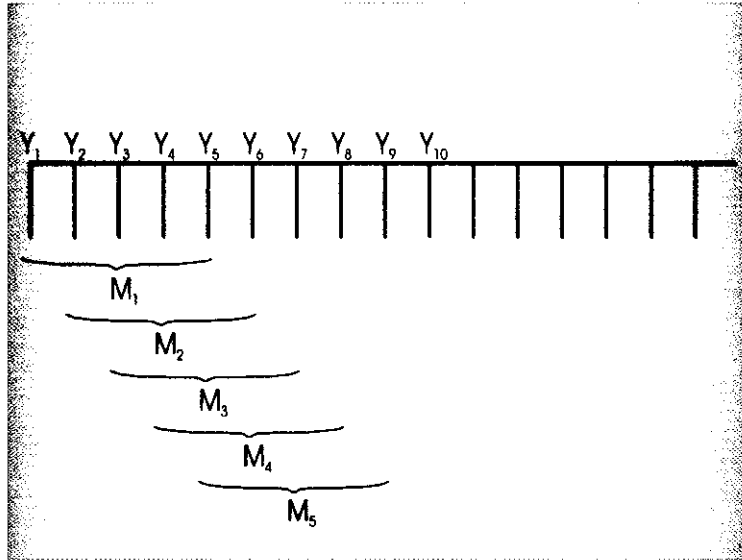


Figura 1.9: Ejemplo de una media móvil de orden 5.

responde a ningún valor de la serie, este problema se resuelve tomando una segunda media móvil como se verá más adelante.

Así, una media móvil centrada en t de longitud impar $(2p + 1)$, que se denota por $MM(2p + 1)_t$, viene dada por la siguiente expresión

$$MM(2p + 1)_t = \frac{\sum_{i=-p}^p X_{t+i}}{2p + 1} \quad (1.7)$$

como puede observarse, el subíndice asignado a la media móvil es el mismo que el de la observación central X_t . Obsérvese también que por construcción no pueden calcularse las medias móviles correspondientes a las p primeras y a las p últimas observaciones.

Por su parte, una media móvil descentrada, o centrada entre el período t y $t + 1$, es decir, en $t + 0.5$, de longitud par p , denotada por $MM(p)_{t+0.5}$, viene dada por la siguiente expresión

$$MM(p)_{t+0.5} = \frac{\sum_{i=-(p/2)+1}^{p/2} X_{t+i}}{p} \quad (1.8)$$

Para la estimación de la tendencia por medias móviles, considérese a la serie de tiempo $\{X_t\}$ con integración en sus componentes de manera aditiva,

entonces al promediar varios valores de la misma (es decir, el calcular medias móviles), se obtiene otra serie más suavizada que la anterior que es posible identificar con el componente tendencia. Lo anterior se justifica porque, si se acierta con el orden apropiado de la media móvil, tanto las variaciones dadas por los componentes estacional e irregular tenderán a anularse en promedio. Por lo tanto la serie resultante, la serie de las medias móviles, es una nueva serie que pretende ser un reflejo de las características más permanentes que subyacen a la variable que se está analizando y a las que se les ha denominado tendencia.

Sin embargo, ¿cómo elegir adecuadamente el valor p de la longitud de la serie?. Como ya se señaló, mientras mayor sea el valor de p de la longitud, mayor será el alisamiento conseguido mediante la media móvil, pues mejor se eliminarán las irregularidades de la serie, ya que intervienen más observaciones en su cálculo. Por el contrario, cuando en la media intervienen pocos datos, la media móvil refleja con mayor rapidez las variaciones de la serie a lo largo del tiempo. Dependiendo, por tanto, de la finalidad perseguida y sopesando ambos inconvenientes se determinará el valor de p de la longitud que es el más indicado. Si las variaciones de la serie se deben a la influencia del componente irregular, es conveniente seleccionar un valor p grande. Pero, si es el nivel de la serie el causante de las variaciones en la misma, un valor de p pequeño será más adecuado.

Cuando la serie presenta la componente estacional, se debe conseguir que estas oscilaciones queden eliminadas si es que se desea que la media móvil recoja fielmente la tendencia en cada momento. Para ello, la longitud de la media móvil más apropiado es el valor p que coincide precisamente con la periodicidad que presenta la serie de tiempo. Por ejemplo, en una serie mensual con periodicidad anual, el valor de p más indicado es 12.

Por lo tanto, la regla práctica para elegir correctamente el valor de la longitud p de la media móvil consiste en elegir un valor de p que coincida con la longitud media de las fluctuaciones de la serie que se desean eliminar.

En la siguiente figura (1.10) se exhibe el resultado de la estimación de la tendencia para la serie utilizada ya previamente en la figura (1.3) mediante el uso de medias móviles donde el orden p es de 12 observaciones.

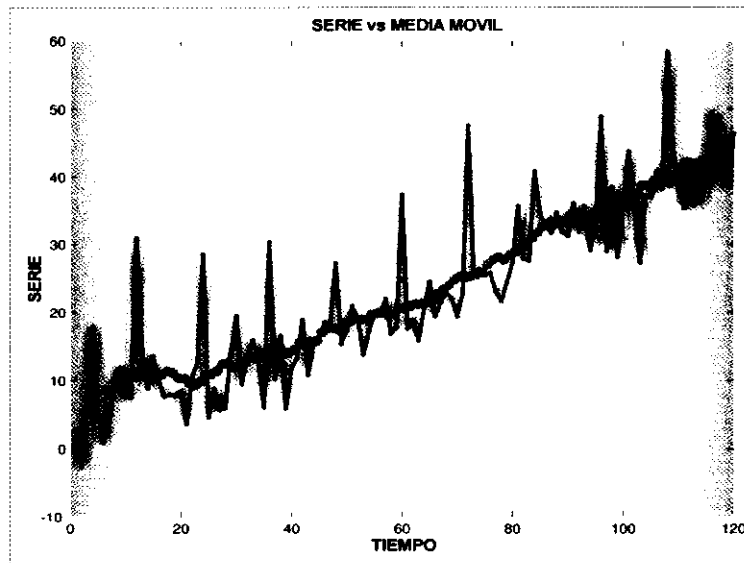


Figura 1.10: Estimación de la tendencia mediante medias móviles con $p=12$. La serie fue obtenida por simulación aplicando el programa 1 del apéndice.

1.5. Análisis de la componente estacional

En esta sección se realiza el análisis de la componente estacional con el objetivo de eliminar esta componente de la serie de tiempo, en otras palabras, se estudiará la *desestacionalización* de la serie.

La desestacionalización consiste en eliminar las fluctuaciones de carácter periódico que se presentan en la serie con el objetivo de hacer comparables los datos de la serie para momentos distintos del tiempo. Pues, en muchas ocasiones, la influencia del factor estacional actúa como una máscara que impide captar adecuadamente la evolución del fenómeno objeto de estudio. Por ejemplo, en una serie mensual sobre el índice de precios de consumo se observa una alta variación entre los meses de noviembre y diciembre, en todos los años, este hecho se debe analizar con cierta precaución, pues se sabe que tradicionalmente el mes de diciembre es un mes en el que los precios sufren un aumento considerable como consecuencia de los mayores gastos que acompañan a la fiesta de navidad, es decir, el dato correspondiente al mes de diciembre es estacionalmente alto. Luego, el mes de noviembre no sería comparable con el de diciembre si previamente no se elimina la com-

ponente estacional presente en esta serie que impide realizar un diagnóstico correcto de la realidad. Es este caso, la componente estacional actúa como un velo muy tupido que impide hacer cualquier tipo de consideración sobre la variación experimentada, es decir, no son comparables ambos datos. La desestacionalización de una serie de tiempo equivale a la eliminación de ese velo, lo que permite realizar comparaciones entre los valores de la serie.

Cabe señalar que la información que se obtiene dentro del análisis de la componente estacional, es de vital importancia tenerla presente de cara a la predicción de la serie de tiempo.

Usando la misma partición del conjunto tiempo de la serie, como en la Sección 1.3.1, los diversos métodos que se utilizan para el análisis de la componente estacional estiman esta componente de la serie a través de índices, expresando con ello las variaciones porcentuales de las diferentes estaciones dentro de cada periodo T_i respecto a la media del mismo periodo T_i . En este sentido, se supone que, para cualquier estación del periodo T_i , el efecto de la componente estacional radica en aumentar o disminuir el valor de la observación original mediante una proporción constante, comparado con lo que ocurriría si no tenemos en cuenta esta influencia estacional.

1.5.1. Desestacionalización

El método de la razón a la media móvil

Este método considera a la serie de tiempo $\{X_t\}$ con integración en sus componentes de manera multiplicativa, de la siguiente forma:

$$X_t = m_t \times s_t \times Y_t \quad (1.9)$$

el objetivo del método de la razón a la media móvil es aislar, o más precisamente estimar el componente estacional s_t de periodo conocido d .

Respecto al comportamiento del componente estacional s_t de la serie, a lo largo del tiempo, se supondrá estable. Este supuesto es el más simple, y en este caso, no se admite que el factor estacional pueda cambiar con el paso del tiempo. Dado que se conoce la longitud d del periodo estacional, y por la estabilidad de esta componente se verifica que para periodos sucesivos, la estacionalidad afecta de la misma forma a cada estación.

Para la aplicación de este método se procede por fases, las cuales son las siguientes:

1. *Obtención de una media móvil de longitud (orden) estacional.*

La expresión para el cálculo de una media móvil de longitud estacional, es decir de orden d , según (1.7) viene dada por

$$MM(d)_t = \frac{\sum_{i=-\left(\frac{d-1}{2}\right)}^{\left(\frac{d-1}{2}\right)} X_{t+i}}{d}$$

si d es impar. En el caso de que d sea par entonces de (1.8) se sigue que

$$MM(d)_{t+0,5} = \frac{\sum_{i=-\left(\frac{d}{2}\right)+1}^{\frac{d}{2}} X_{t+i}}{d}, \quad t = \frac{d}{2}, \frac{d}{2} + 1, \dots, n - \frac{d}{2} \quad (1.10)$$

En la mayoría de los casos la longitud estacional es par, que se corresponde con series de tipo mensual o trimestral. Como puede verse en la figura (1.11), el punto medio de las observaciones que intervienen en (1.10) está situado entre t y $t + 1$, es decir en el punto $t + 0.5$. Como el punto medio de estas medias móviles no se corresponde exactamente con ninguna observación, se les denomina medias móviles no centradas.

Sin embargo, el método necesita contar con una media móvil centrada, que en el caso de longitud impar no hay ningún problema, pero en los casos de longitud par la media móvil debe centrarse.

2. *Obtención de una media móvil centrada.*

Para obtener una media móvil centrada solo se necesita aplicarle a (1.10) a su vez una nueva media móvil de orden 2, que se denota por $MM(d, 2)_t$, tal como se aprecia en la figura (1.11) La expresión de esta media móvil centrada es la siguiente:

$$MM(d, 2)_t = \frac{MM(d)_{t-0,5} + MM(d)_{t+0,5}}{2}, \quad t = \frac{d}{2}, \frac{d}{2} + 1, \dots, n - \frac{d}{2} \quad (1.11)$$

Obsérvese que tal como se deseaba la media móvil resultante se encuentra centrada, y además que el número de términos de la serie que intervienen son $d + 1$, en otras palabras la media móvil es de orden $d + 1$.

Otro punto importante a resaltar es que la media móvil así obtenida, resulta ser una media aritmética ponderada ya que los elementos de la serie que intervienen en ella, en la primera media móvil (la de orden impar) son ponderados por el factor $\frac{1}{d}$, y con la nueva media móvil se tiene un nuevo factor de ponderación, el cual asigna un mayor peso, específicamente el doble

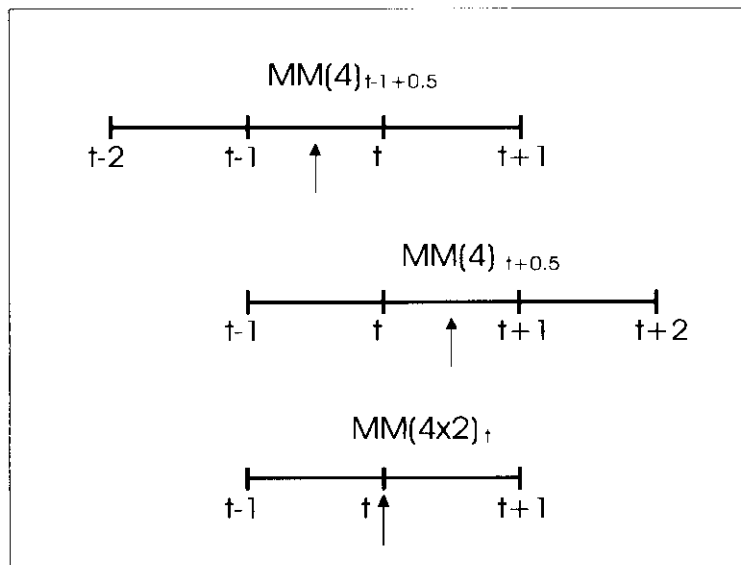


Figura 1.11: Construcción de una media móvil $MM(4,2)$.

a las observaciones centrales o cercanas al tiempo t , que a las observaciones extremas que intervienen en la media móvil $MM(d,2)_t$.

Por construcción, la primera media móvil de segundo orden o el primer valor de t para el cual esta definida $MM(d,2)_t$, es $\frac{d}{2} + 1$ y la última observación que se le asocia una $MM(d,2)_t$ es $n - \frac{d}{2}$.

3. Obtención de índices brutos de variación estacional (IBVE)

En la sección (1.4.1), se vió que mediante el uso de las medias móviles se pretende estimar o aproximarse al componente tendencia m_t de la serie de tiempo $\{X_t\}$, y, con el valor adecuado para el orden de la media móvil, se obtiene una serie mas suave que la original, esto significa que el efecto del componente aleatorio Y_t se ve disminuido en gran medida. Lo anterior también se verifica, aún con mayor fuerza con el empleo de medias móviles de segundo orden.

Por otro lado, al tomar una media móvil de segundo orden $MM(d,2)_t$ la influencia del componente estacional s_t quedará prácticamente eliminada, o atenuada en su mayor parte, ya que en cada media móvil de segundo orden intervienen dos datos de cada una de las estaciones.

En síntesis, a la expresión (1.11) se le considera como una adecuada estimación de la componente tendencia, y recordando la asociación supuesta para las componentes de la serie de tiempo, se tiene:

$$\frac{X_t}{MM(d, 2)_t} = \frac{m_t \times s_t \times Y_t}{m_t} = s_t \times Y_t \quad (1.12)$$

Con esto, se ha obtenido una estimación conjunta de los componentes estacional y aleatorio de la serie de tiempo, a esta estimación (1.12) se le denomina *Índices Brutos de Variación Estacional (IBVE_t)*, el término bruto obedece a la presencia de la componente aleatoria.

4. Obtención de índices de variación estacional sin normalizar (IVE_h)

Partiendo del supuesto de que es conocido el valor de d , la longitud estacional, entonces se define el número de años a para los datos que conforman a la serie de tiempo mediante la expresión, $a = \frac{n}{d}$. Luego, para cada estación h se dispone de $a - 1$ (IBVE_t), ya que se pierden $\frac{d}{2}$ datos al principio y final en la construcción de la $MM(d, 2)_t$, es decir, se pierde un dato por cada estación h .

En éstos términos, es posible para cada estación h agrupar a todos los (IBVE_t) pertenecientes a ella, y tomar como representante del valor de la estación a la *media* de todos ellos, esto es,

$$(IVE_h) = \sum_{t \in h} \frac{(IBVE_t)}{a - 1}, \quad h = 1, \dots, d \quad (1.13)$$

a la expresión anterior se le denomina *Índices de Variación Estacional sin Normalizar (IVE_h)*.

Al realizar un promedio de $a - 1$ datos, el componente irregular Y_t de la serie queda prácticamente eliminado si a es suficientemente grande. En todo caso, al promediar siempre se atenuará el efecto del componente aleatorio. Por esto, los (IVE_h), constituyen un índice de variación estacional en el cual se supone que el componente aleatorio ha desaparecido completamente.

En otras palabras se ha conseguido eliminar la presencia del componente aleatorio a los (IBVE_t) y tener así una estimación para el componente estacional s_t con los (IBVE_t). Sin embargo, como a esta estimación aún es necesario realizar unos pequeños ajustes.

5. Obtención de índices de variación estacional normalizados (IVEN_h)

Para clarificar el término y verificar la necesidad de *normalizar* supóngase que si se tiene una serie de tiempo $\{X_t\}$ con integración en sus componentes de manera multiplicativa en donde no existe estacionalidad, entonces se sigue que,

$$s_t = 1, \quad \forall t$$

característica que se convierte en condición que debe cumplir aquel que se considere como un buen estimador para s_t , es claro que en presencia de estacionalidad, ésta no debe afectar el nivel de la serie. Por lo tanto se vislumbra la exigencia a los estimadores para s_t que tengan *media 1, o que su suma sea d* .

Al implementar la condición anterior a los (IVE_h) se tiene

$$(IVEN_h) = (IVE_h) \frac{d}{\sum_{j=1}^d (IVE_j)} \quad (1.14)$$

llegando así, a la estimación de la tendencia s_t mediante los llamados *Indíces de Variación Estacional Normalizados* $(IVEN_h)$.

6. Desestacionalización de la serie $\{Xd_t\}$

Se ha llegado a la parte final y objetivo de esta sección, quitar el efecto de la componente estacional s_t de la serie de tiempo $\{X_t\}$, para con ello estar en condiciones de realizar comparaciones entre los distintos valores de la serie y poder analizar con mayor rigor la evolución que esta detraés de la serie.

La serie desestacionalizada se obtiene dividiendo (no olvidar la asociación supuesta para los componentes) cada valor de la serie $\{X_t\}$ por el $(IVEN_h)$ correspondiente, siempre y cuando t pertenezca a la estación h , entonces el valor de la serie desestacionalizada, a la que denotaremos $\{Xd_t\}$ viene dada por

$$Xd_t = \frac{X_t}{(IVEN_h)}, \quad \text{siempre que } t \in h \quad (1.15)$$

En la siguiente figura (1.12) se presenta el resultado de aplicar el método de desestacionalización presentado en esta apartado para la serie de la figura (1.3) y como puede apreciarse se tiene una serie más suave sin "picos".

1.6. Análisis de la componente aleatoria.

El objetivo de lo descrito en las secciones anteriores es producir datos desprovistos de *tendencia y estacionalidad*, esto es, en la serie solo debe quedar

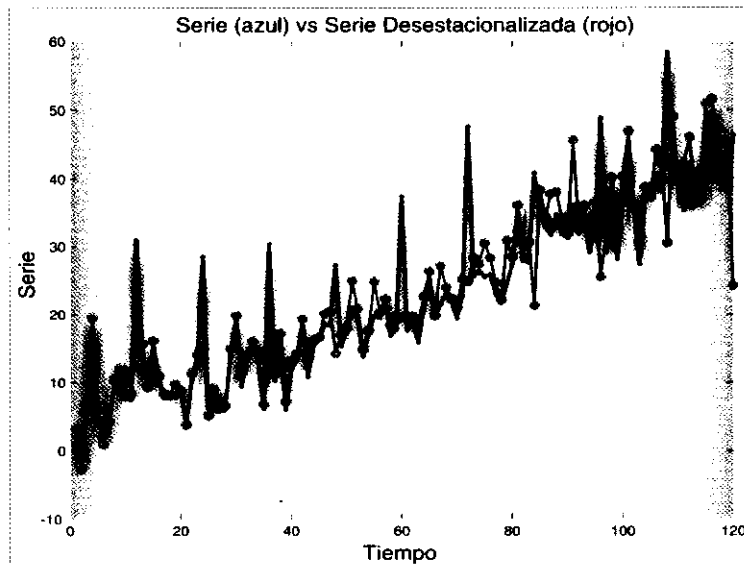


Figura 1.12: Serie desestacionalizada, y se obtuvo mediante simulación a través del empleo del programa 3 del apéndice.

la estimación de la componente aleatoria. Si esto efectivamente se consiguió, entonces el siguiente paso en el camino del análisis de la serie de tiempo, es adoptar el modelo más apropiado para el componente aleatorio.

Sin embargo, como se mencionó en la sección (1.2) se espera que las características de la componente aleatoria no sean tan "puras", es decir, que exista todavía información dentro de esta componente que ayude a comprender de mejor manera el fenómeno que se describe con la serie de tiempo.

Si se supone que la componente aleatoria esta conformada por observaciones de variables aleatorias no relacionadas entre si, es decir, independientes, entonces se debe aceptar que no queda nada por hacer, sino simplemente observar y disfrutar la presencia y evolución de un fenómeno puramente aleatorio.

Luego, resulta necesario pensar que existe *dependencia* en las observaciones de esta componente aleatoria. Por tales motivos, en esta sección se presentan algunas pruebas de hipótesis sencillas para verificar la hipótesis de que los residuales, esto es, la estimación de la componente aleatoria, son valores de variables aleatorias independientes e idénticamente distribuidas,

y en tales casos en que esta hipótesis sea cierta, se insiste en que el trabajo sobre la serie de tiempo, habrá concluido. En caso contrario, el caso deseado, en el cual esta hipótesis se rechaza, entonces será necesario revisar la teoría de los procesos estocásticos más a detalle para elegir modelos adecuados para tal componente.

1.6.1. Características de un proceso estocástico.

Previo al detalle de las pruebas de hipótesis de esta sección resultan necesarias unas definiciones. Sea $\{X_t, t \in T\}$ un proceso estocástico como en (def 1.1), se definen las *características del proceso*, como sigue:

Definición 1.8 La función esperanza de un proceso estocástico $\{X(t)\}$ se define por

$$\mu(t) = E [X(t)].$$

■

Definición 1.9 La función varianza de un proceso estocástico $\{X(t)\}$ se define por

$$\sigma^2(t) = E [(X(t) - \mu(t))^2].$$

■

Definición 1.10 Si $\{X_t, t \in T\}$ es un proceso tal que $Var(X_t) < \infty$ para cada $t \in T$, entonces la función de autocovarianza $\varphi_X(\bullet, \bullet)$ de $\{X_t\}$ se define por

$$\varphi_X(r, s) = Cov(X_r, X_s) = E [(X_r - E(X_r)) (X_s - E(X_s))]$$

donde $r, s \in T$.

■

Definición 1.11 Sea $\{X(t)\}$ un proceso estocástico, entonces para $r, s \in T$ se define la función de autocorrelación, $\rho_X(\bullet, \bullet)$ de $\{X_t\}$ como

$$\rho_X(r, s) = \frac{\varphi_X(r, s)}{\sigma(r)\sigma(s)}.$$

■

En la mayoría de las ocasiones se desconoce el proceso estocástico del cual proceden los valores de la serie de tiempo $\{x_t, t \in T_o \subset \mathbb{Z}\}$, y por consecuencia también resultan desconocidos las características del proceso, sin embargo, a partir de las observaciones x_t es posible realizar estimaciones de ellas a las cuales se les denomina características del proceso *en el muestreo*.

1.6.2. Características en el muestreo de una serie de tiempo

Sea $\{x_t, t \in T_o \subset \mathbb{Z}\}$ una serie de tiempo, se define:

Definición 1.12 La esperanza muestral de una serie de tiempo $\{x_t, t = 1, 2, \dots, n\}$ viene dada por

$$\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t .$$

■

Definición 1.13 La función de autocovarianza muestral $\hat{\varphi}(h)$ de una serie de tiempo $\{x_t, t = 1, 2, \dots, n\}$ es

$$\hat{\varphi}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x}) (x_t - \bar{x})$$

donde $-n < h < n$.

■

Definición 1.14 La función de autocorrelación muestral $\hat{\rho}(h)$ de una serie de tiempo $\{x_t, t = 1, 2, \dots, n\}$ es

$$\hat{\rho}(h) = \frac{\hat{\varphi}(h)}{\hat{\varphi}(0)}$$

donde $-n < h < n$.

■

1.6.3. Pruebas de aleatoriedad

En estas instancias se encuentra todo listo para conocer las pruebas de hipótesis propuestas para la revisión de la aleatoriedad, aquí se presentan solo dos pruebas, es claro que no son las únicas, sin embargo, se eligen estas pruebas sobre las demás debido a su sencillez, fácil manejo y sobretodo debido a que los conceptos que para ellas se ocupan, se manejarán en las secciones subsecuentes.

(a) Prueba de la función de autocorrelación en el muestreo.

Para un proceso estocástico $\{X_t, t \in T\}$ su función de autovarianza (def 1.10) es una medida de la relación que presentan las variables aleatorias X_t para diferentes momentos en el tiempo. Asimismo, la función de autocorrelación mide la relación lineal que existe entre las variables del proceso, también para diferentes momentos en el tiempo. En el capítulo 2 se revisará con todo rigor y formalidad la relación que existe entre un proceso estocástico y su función de autocovarianza, aquí solo se adelantará que tal relación es biunívoca, esto es, el conocimiento del proceso define a la función de autocovarianza, (lo cual no es novedad), sino lo que resulta sorprendente es la aseveración inversa, *conocer a la función de autocovarianza implica conocer al proceso estocástico que la define*. Es este resultado el ladrillo que permitirá construir todo el modelo para los procesos estocásticos a utilizar en el largo camino que falta por recorrer.

Considérese a la serie de tiempo $\{y_t, t = 1, 2, \dots, n\}$ observaciones provenientes de $\{Y_t, t = 1, 2, \dots, n\}$ variables aleatorias independientes e idénticamente distribuidas (iid) con varianza finita y función de autocorrelación en el muestreo $\hat{\varphi}(h)$, entonces se sigue que cuando n es lo suficientemente grande, $\hat{\rho}(h)$ converge en distribución a una normal $N(0, \frac{1}{n})$.

Por lo tanto, para la hipótesis nula, esta dada por

$$[H_0 : \{y_t, t = 1, 2, \dots, n\} \text{ son la realización de v.a. iid}]$$

bajo H_0 alrededor del 95% de los valores de la función de autocorrelación muestral deben caer en las bandas $\pm \frac{1.96}{\sqrt{n}}$.

Luego, debe rechazarse H_0 , si al calcular los valores de la función de autocorrelación muestral para $n \geq 40$, y se encuentra que más de dos o tres valores caen por fuera de la bandas, o que un valor cae muy lejano de dichas bandas.

(b) Prueba de Portmanteau

Al igual que en la prueba de la función de autocorrelación muestral, considérese a $\{Y_t, t = 1, 2, \dots, n\}$ variables aleatorias iid, luego el estadístico

$$Q = n \sum_{j=1}^h \hat{\rho}^2(j)$$

es la suma de cuadrados de variables aleatorias $\sqrt{n} \hat{\rho}(h)$, $j = 1, 2, \dots, h$ las cuales por lo visto en (a), son iid con distribución normal $N(0, 1)$, lo que significa que Q sigue una distribución *Ji - cuadrada* con h grados de libertad. Por lo tanto, si Q es demasiado grande, esto indica que las autocorrelaciones muestrales también son demasiado grandes y no corresponden a datos provenientes de *v.a. iid*.

Así, se rechaza la hipótesis nula H_0

$[H_0 : \{y_t, t = 1, 2, \dots, n\}$ son la realización de *v.a. iid*]

al nivel de significancia α si $Q > \chi_{1-\alpha}^2(h)$.

Capítulo 2

Procesos estocásticos estacionarios

Papel fundamental en el análisis de las series de tiempo lo tienen ciertos procesos estocásticos cuyas propiedades, o algunas de ellas no varían en el tiempo. Es claro que si se desean realizar predicciones, entonces debe asumirse que algo no cambia con el tiempo.

Dentro de las series de tiempo la meta es predecir a la serie que no es determinista, sino que contiene un componente aleatorio. Si este componente es estacionario, entonces es posible desarrollar técnicas de predicción, las cuales serán discutidas más adelante.

Para este capítulo se siguen a los autores, Brockwell y Davis [1] [2], Otero [6] y Karr [3].

2.1. Definiciones básicas

Dentro de la literatura de los procesos estocásticos generalmente se les clasifica como uno de los dos siguientes: *estacionarios o evolutivos*. Los procesos evolutivos se utilizan para describir fenómenos puramente aleatorios, cambiantes a todo momento; así que dado el fin aquí buscado, los procesos que se estudiarán serán los estacionarios.

La importancia de estudiar a los procesos estocásticos estacionarios radica en que este tipo de procesos describen la variación en el tiempo de un fenómeno aleatorio tal que ninguno de los factores que tienen influencia sobre él cambian a lo largo del tiempo.

De estacionariedad puede hablarse en dos sentidos, *estricta y débil*.

Procesos estacionarios en sentido estricto

Definición 2.1 Sea $\{X_t, t \in T \subset \mathbb{Z}\}$ proceso estocástico definido sobre el espacio de probabilidad $(\Omega, \mathfrak{F}, P)$ se le denomina estacionario en sentido estricto si el conjunto de las funciones de distribución $\{F_t(\bullet), t \in \mathcal{F}\}$ satisface

$$\begin{aligned} F_t(\mathbf{x}) &= F_{t+\tau}(\mathbf{x}) \quad \text{i.e.} \\ F(x_1, \dots, x_n; t_1, \dots, t_n) &= F(x_1, \dots, x_n; t_1 + \tau, \dots, t_n + \tau) \end{aligned}$$

donde $\tau \in \mathbb{R}$.

■

En los procesos estocásticos estacionarios en sentido estricto, el conjunto de las funciones de distribución no depende del origen del parámetro tiempo $t \in T$.

Proposición 2.1 En los procesos estocásticos estacionarios en sentido estricto, las distribuciones de primer orden (o unidimensionales) son iguales e independientes del tiempo.

Demostración:

Sea $\{X_t, t \in T\}$ un proceso estocástico estacionario en sentido estricto, y considérese las distribuciones de la forma $\{F_t(\bullet), t \in \mathbb{R}\}$, es decir las distribuciones unidimensionales o de primer orden. Entonces

$$\begin{aligned} F_t(x) &= P(X_t \leq x) \\ F_{t+\tau}(x) &= P(X_{t+\tau} \leq x) \quad \tau \in \mathbb{R}. \end{aligned}$$

al tratarse de un proceso estacionario en sentido amplio se cumplirá que

$$F_t(x) = F_{t+\tau}(x) \quad \text{para cualquier } \tau \in \mathbb{R}.$$

entonces para $\tau = -t$

$$F_t(x) = F_{t-t}(x) = F_0(x) = F(x)$$

que es independiente del tiempo.

■

Proposición 2.2 *En los procesos estocásticos estacionarios en sentido estricto, las distribuciones de segundo orden (o bidimensionales) dependen de la diferencia $t_2 - t_1$.*

Demostración:

Sea $\{X_t, t \in T\}$ un proceso estocástico estacionario en sentido estricto, y considérese las distribuciones de la forma $\{F_{\mathbf{t}}(\bullet), \mathbf{t} \in \mathbb{R}^2\}$, es decir las distribuciones bidimensionales o de segundo orden:

$$\begin{aligned} F_{\mathbf{t}}(x) &= F(x_1, x_2; t_1, t_2) = P(X_{t_1} \leq x_1; X_{t_2} \leq x_2) \\ F_{\mathbf{t} + \tau}(x) &= F(x_1, x_2; t_1 + \tau, t_2 + \tau) = P(X_{t_1} \leq x_1 + \tau; X_{t_2} \leq x_2 + \tau) \end{aligned}$$

donde $\tau \in \mathbb{R}$. Al tratarse de un proceso estacionario en sentido amplio se verifica que

$$F_{\mathbf{t}}(x) = F_{\mathbf{t} + \tau}(x) \quad \text{para cualquier } \tau \in \mathbb{R}.$$

haciendo $\tau = -t_1$

$$\begin{aligned} F(x_1, x_2; t_1, t_2) &= F(x_1, x_2; t_1 - t_1, t_2 - t_1) \\ &= F(x_1, x_2; 0, t_2 - t_1) \\ &= F(x_1, x_2; t_2 - t_1). \end{aligned}$$

■

A consecuencia de las dos proposiciones anteriores las características de un proceso estocástico $\{X_t, t \in T\}$ estacionario en sentido estricto cumplen:

- Las funciones de esperanza y varianza del proceso no dependen de t , es decir, $\mu(t) = \mu, \sigma^2(t) < \infty$.
- Las funciones de autocovarianza y autocorrelación del proceso no dependen de t , esto es,

$$\begin{aligned} \varphi_X(r, s) &= \varphi_X(s - r), \\ \rho_X(r, s) &= \rho_X(s - r). \end{aligned}$$

Para ambas funciones haciendo $h = s - r$, se convierten en funciones de un sólo parámetro,

$$\begin{aligned} \varphi_X(r, s) &= \varphi_X(h) = Cov(X_{t+h}, X_t) \\ \rho_X(r, s) &= \rho_X(h) = Corr(X_{t+h}, X_t) \end{aligned}$$

La función de autocorrelación satisface además las siguientes propiedades, a saber:

- a) $\rho_X(0) = 1$
- b) $\rho_X(h) = \rho_X(-h)$
- c) $|\rho_X(h)| < \rho_X(0)$, es decir, $|\rho_X(h)| < 1$.

Como puede verse la definición de estacionariedad en sentido estricto implica que las características del proceso estocástico no sufren alteración al considerar tiempos históricos diferentes. Ésta es, sin duda, una propiedad y en casos se convierte en una restricción, demasiado fuerte para imponerla *a priori* a los procesos estocásticos destinados a representar por ejemplo, sistemas socioeconómicos, es decir, los datos económicos reales, violan la hipótesis de estacionariedad estricta.

Procesos estocásticos débilmente estacionarios o estacionarios en sentido amplio.

Las propiedades anteriores de las funciones esperanza, varianza, autocovarianza y autocorrelación, es decir, las características del proceso, no determinan unívocamente a un proceso estacionario en sentido estricto.

A la vista de esta circunstancia se extiende el concepto de estacionariedad en sentido estricto para considerar procesos en donde la estacionariedad no sea tan “fuerte”.

Definición 2.2 (*Proceso estacionario en sentido amplio*) *A un proceso estocástico $\{X_t, t \in T \subseteq \mathbb{Q}\}$ se le denomina estacionario en sentido amplio (o débilmente estacionario) si sus características satisfacen*

- a) $\mu(t) = \mu < \infty$
- b) $\sigma^2(t) = \sigma < \infty$,
- c) $\varphi_X(r, s) = \varphi_X(r + t, s + t)$

donde $r, s, t \in \mathbb{Q}$.

■

En los sucesivos cuando se diga que un proceso estocástico es estacionario, a menos que se especifique lo contrario, deberá entenderse estacionario en sentido amplio.

Se dirá que una serie de tiempo es estacionaria si el proceso estocástico (del cual la serie es una realización) es estacionario.

Relación entre estacionario y estacionario en sentido estricto

Por lo señalado en la nota (2.1) se sigue que si un proceso estocástico $\{X_t, t \in T\}$ es estacionario en sentido estricto entonces es estacionario. Sin embargo, la aseveración inversa no necesariamente es verdad. Para verificar esto, tómese por ejemplo una sucesión $\{X_t\}$ de variables aleatorias independientes tales que X_t se distribuye exponencial con media cero cuando t es impar, y X_t se distribuye normal estándar cuando t es par, entonces $\{X_t\}$ cumple con las características de un proceso estacionario según la definición (2.2), pero, dado que X_1 y X_2 tienen distribuciones diferentes $\{X_t\}$ no es un proceso estacionario en sentido estricto.

Existe un importante caso de un proceso en donde estacionariedad implica estacionariedad en sentido estricto, a saber:

Definición 2.3 *El proceso $\{X_t\}$ es una serie de tiempo Gaussiana si y sólo si todas las funciones de distribución de $\{X_t\}$ son normales multivariadas.*

■

2.2. Las funciones de autovarianza y autocorrelación

Las funciones de autovarianza y autocorrelación fueron definidas en (1.6.2), para un proceso estocástico $\{X_t, t \in T\}$, el cual en lo sucesivo se supondrá estacionario en sentido amplio. Las funciones de FAV y FAC proveen de una medida de los grados de dependencia de los valores de la serie de tiempo para diferentes momentos en el tiempo, y por tal motivo juegan un papel clave de cara a la predicción de la serie en términos de la única información disponible: el pasado y presente de la serie.

La importancia de estas funciones se ilustrará con el siguiente ejemplo, considérese una serie de tiempo $\{x_t, t = 1, \dots, n\}$ proveniente de un proceso estocástico estacionario Gaussiano $\{X_t, t \in T\}$. Luego, se intenta encontrar la función de x_n la cual sea la mejor predicción para x_{n+h} donde $h > 0$, el valor de la serie después de h tiempo. Obsérvese que bajo estas circunstancias se presenta el problema de definir lo que significa la mejor predicción, esto se resuelve adoptando la clásica y natural definición de mejor predicción como la función con error cuadrático medio mínimo.

Ahora para el caso Gaussiano, se sabe que la distribución condicional de X_{n+h} dado $X_n = x_n$ sigue una distribución normal¹

$$N(\mu + \rho(h)(x_n - \mu), \sigma^2(1 - \rho(h)^2))$$

donde μ y σ^2 son la media y varianza de $\{X_t, t \in T\}$. Por otro lado, utilizando el resultado de que la función $m(X_n)$ que minimiza el error cuadrático medio

$$E^2[X_{n+h} - m(X_n)]$$

es la esperanza condicional

$$m(X_n) = E(X_{n+h} | X_n) = \mu + \rho(h)(X_n - \mu)$$

y, el error cuadrático medio es

$$E^2[X_{n+h} - m(X_n)] = \sigma^2(1 - \rho(h)^2)$$

esta expresión muestra, al menos para el caso Gaussiano, que la predicción de X_{n+h} en términos de X_n , es más precisa cuando $|\rho(h)|$ este próximo a 1, esto es, cuando exista relación lineal entre los X_n .

Este resultado inmediatamente conduce a que en vez de buscar a la mejor función $m(X_n)$ para la predicción de X_{n+h} se considere a mejor predictor lineal, es decir, de la forma $\lambda(X_n) = aX_n + b$ entonces el problema se convierte en encontrar a y b tales que minimicen

$$E^2[X_{n+h} - \lambda(X_n)] = E^2[X_{n+h} - aX_n + b]$$

y finalmente se llega a

$$\lambda(X_n) = \mu + \rho(h)(X_n - \mu)$$

por lo tanto, para un proceso estocástico Gaussiano, se tiene que las estimaciones $m(X_n)$ y $\lambda(X_n)$, son las mismas. No obstante, en el caso general se tendrá siempre que $m(X_n)$ dará un menor error cuadrático medio que $\lambda(X_n)$ dado a su naturaleza de mejor predictor dentro de una clase de predictores más amplia.

Sin embargo, el hecho de que el mejor predictor lineal dependa únicamente de la media y de la FAC del proceso $\{X_t, t \in T\}$, significa que este predictor

¹Veáse Brockwell

puede calcularse sin mayor conocimiento de las distribuciones conjuntas de $\{X_t, t \in T\}$.

De este ejemplo se desprenden conclusiones importantes de cara al rumbo que tomará el análisis de las series de tiempo, la primera de ellas es la trascendencia de la FAV y de FAC, por tanto, para la estimación de la componente aleatoria para el modelo de la serie, se buscarán procesos estocásticos que mantengan una estrecha relación con sus FAV y FAC.

La segunda conclusión, es que de cara a la predicción de los valores de la serie, esta se realizará en términos no únicamente de x_n , sino de todo el pasado de la serie x_1, \dots, x_n , además de hacer uso del predictor lineal.

2.3. Tres procesos importantes: de ruido y ruido blanco, MA(q) y los procesos lineales

En apartados anteriores ya se ha mencionado la importancia de realizar una apropiada selección del modelo probabilístico para los datos que conforman a la serie de tiempo. Sin embargo, no se ha dicho nada sobre cuales son esos posibles “modelos apropiados”, motivo por el cual ahora se estudiarán algunos de ellos. También, por lo visto anteriormente se vislumbra que una propiedad muy importante para estos modelos probalísticos es que deberán ser estacionarios.

El modelo más simple para una serie de tiempo es en el cual no existe tendencia ni componente estacional y además las observaciones son la realización de un proceso donde las variables son independientes e idénticamente distribuidas de esperanza cero. A un proceso de tales características se le denomina **proceso de ruido iid**.

Definición 2.4 *Un proceso estocástico $\{Z_t, t \in T\}$ se dice que es un proceso de ruido iid si:*

sus funciones de distribución $\{F_{\mathbf{t}}(\bullet), \mathbf{t} \in \mathcal{F}\}$ (definición 1.3) satisfacen

$$\begin{aligned} F_{\mathbf{t}}(\mathbf{z}) &= P(Z_{t_1} \leq z_1, \dots, Z_{t_n} \leq z_n) \\ &= P(Z_{t_1} \leq z_1) \cdots P(Z_{t_n} \leq z_n) \\ &= F(z_1) \cdots F(z_n) \end{aligned}$$

donde $\mathbf{z} = (z_1, \dots, z_n)' \in \mathbb{R}^n$

para cada $t \in \mathcal{F}$ las variables Z_{t_1}, \dots, Z_{t_n} son idénticamente distribuidas la función esperanza del proceso estocástico $\mu(t) = 0$ para $t \in T$.

■

Observación 2.1 En un proceso de ruido iid el conocimiento de Z_{t_1}, \dots, Z_{t_n} no tiene valor para pronosticar el comportamiento de Z_{t+h} con $h \geq 0$.

Observación 2.2 Si $\{Z_t, t \in T\}$ es un proceso de ruido iid tal que la función de varianza cumpla

$$\sigma^2(t) = E[Z_t^2] < \infty$$

Entonces por la independencia se tiene que

$$\varphi_Z(t+h, t) = \begin{cases} \sigma^2, & \text{si } h = 0 \\ 0, & \text{si } h \neq 0 \end{cases} \quad (2.1)$$

luego la función de autocovarianza no depende de t y por tanto se tiene que el proceso de ruido iid es estacionario en sentido estricto, al que se denomina proceso de ruido estacionario y se denota por $\{Z_t\} \sim iid(0, \sigma^2)$.

Proceso de ruido blanco

Considérese $\{Z_t\}$ una sucesión de variables aleatorias incorrelacionadas, cada una con media cero y varianza σ^2 , entonces claramente se tiene que $\{Z_t\}$ define un proceso estacionario cuya función de autocovarianza tiene la misma expresión que (2.1). Un proceso estocástico con tales características es denominado *proceso de ruido blanco* con media cero y varianza σ^2 y se denota por $\{Z_t\} \sim WN(0, \sigma^2)$.

Observación 2.3 Obsérvese que todo proceso $\{Z_t\} \sim iid(0, \sigma^2)$ es $\{Z_t\} \sim WN(0, \sigma^2)$ pero la afirmación inversa no es verdad.

La importancia de los procesos iid y de ruido blanco es que proporcionan un camino sencillo para la construcción de series de tiempo $\{X_t\}$ que sean estacionarias.

Construcción del proceso $MA(q)$

Considérese $\{Z_t\} \sim iid(0, \sigma^2)$, proceso estacionario en sentido estricto y defínase

$$X_t = g(Z_t, Z_{t-1}, \dots, Z_{t-q})$$

para alguna función real $g(\mathbf{a})$ con $\mathbf{a} \in \mathbb{R}^{q+1}$. Entonces $\{X_t\}$ es estacionario en sentido estricto pues lo hereda de $\{Z_t\}$ ya que este último satisface

$$(Z_{t+h}, \dots, Z_{t+h-q})' \stackrel{d}{=} (Z_t, \dots, Z_{t-q})'$$

para $h \in \mathbb{Z}$.

Por construcción de $\{X_t\}$ se sigue que $\{X_t\}$ es q -dependiente, es decir, que X_s y X_t son independientes si $|t - s| > q$. Del mismo modo, se dice que una serie de tiempo estacionaria $\{X_t\}$ es q -correlacionada si $\varphi_X(h) = 0$ siempre que $|h| > q$.

Si la función real $g(\mathbf{a})$ con $\mathbf{a} \in \mathbb{R}^{q+1}$ es una función lineal entonces da lugar a uno de los procesos más importantes el denominado *proceso estocástico de medias móviles de orden q* , $MA(q)$, el cual según lo visto en los párrafos de arriba es q -correlacionado.

Definición 2.5 (Proceso $MA(q)$) *El proceso estocástico $\{X_t\}$ es un proceso de medias móviles de orden q si*

$$X_t = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} \quad (2.2)$$

donde $\{Z_t\} \sim WN(0, \sigma^2)$ y $\theta_1, \theta_2, \dots, \theta_q$ son constantes reales.

■

Cuando en (2.2) se tenga que $\{Z_t\} \sim iid(0, \sigma^2)$ entonces el proceso $MA(q)$ será estacionario en sentido estricto.

Si bien, más adelante se dirá mucho más de los procesos $MA(q)$, en este punto resulta necesario señalar la trascendencia de los procesos $MA(q)$, la cual se deriva de lo siguiente, por definición se tiene que un proceso $MA(q)$ es q -correlacionado, sin embargo, lo sorprendente es la aseveración inversa, *todo proceso estocástico que sea q -correlacionado será un proceso $MA(q)$* .

Este resultado es de vital importancia ya que generalmente cuando se trabaja con series de tiempo, es decir, con un conjunto de datos, sólo se dispondrá de un cierto número q de datos, los mismos que probablemente serán correlacionados, lo que implica modelarlos como la realización de algún proceso estocástico que sea q -correlacionado, esto es, con un proceso $MA(q)$

Procesos lineales

Los procesos estocásticos lineales corresponden a una familia muy extensa que incluye a los denominados procesos ARMA que resultarán (es el siguiente paso) la meta en lo referente a la búsqueda de los modelos apropiados para las series de tiempo con las que se ha venido trabajando a estas instancias, esto es, series desprovistas de la componente tendencia y periodicidad, formadas únicamente por la componente aleatoria.

Aquí sólo se dará una definición, sin embargo, se hablará más a detalle de estos procesos en la siguiente sección.

Definición 2.6 (Procesos lineales) *El proceso estocástico $\{X_t\}$ es un proceso lineal si para toda t , puede escribirse como*

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} \quad (2.3)$$

donde $\{Z_t\} \sim WN(0, \sigma^2)$ y $\{\psi_j\}$ es una sucesión de constantes tales que $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$.

■

De la definición anterior se tiene la siguiente pregunta lógica, que sentido tiene la expresión (2.3), sentido en cuanto a la convergencia de dicha expresión. Esto se resuelve observando que dada la condición $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$. y aplicando la desigualdad de *Lyapunov*² al proceso lineal definido en (2.3) se tiene que

$$\begin{aligned} E|X_t| &\leq \sum_{j=-\infty}^{\infty} (|\psi_{t-j}| E|Z_{t-j}|) \\ &\leq \left(\sum_{j=-\infty}^{\infty} |\psi_{t-j}| \right) \sigma \\ &\leq \infty \end{aligned}$$

esto significa que el proceso lineal $\{X_t\}$ definido en (2.3) está acotado, por lo que converge en L^1 y por lo tanto en probabilidad.

²Veáse Karr

2.4. Procesos estacionarios ARMA(p,q)

En esta sección se revisarán a una familia paramétrica de procesos estocásticos $\{X_t, t = 0, \pm 1, \dots\}$, los cuales se definen en términos de ecuaciones en diferencias lineales con coeficientes constantes, esta familia define a los denominados procesos ARMA.

Los procesos ARMA presentan una característica sumamente importante, mantienen una relación estrecha con las funciones de autocovarianza, a saber: dada una función de autocovarianza $\varphi(h)$ tal que $\lim_{h \rightarrow \infty} \varphi(h) = 0$ y para todo entero $k > 0$, es posible encontrar un elemento de la familia ARMA $\{X_t, t \in \mathbb{Z}\}$ con función de autocovarianza $\varphi_X(h)$ tal que coincida con $\varphi(h)$, i.e, se cumple que

$$\varphi_X(h) = \varphi(h) \text{ para } h = 0, 1, \dots, k$$

Por esto, la familia de los procesos ARMA serán considerados dentro de la construcción de modelos de las series de tiempo aquí propuesto, además de que su estructura lineal conlleva a una teoría práctica y sencilla dentro de la predicción que se revisará en el capítulo 4.

Definición 2.7 (proceso ARMA (p,q)) *El proceso estocástico $\{X_t, t \in \mathbb{Z}\}$ es un proceso de medias móviles autoregresivo de parámetros $p, q \in \mathbb{R}$, ó ARMA(p,q) si $\{X_t\}$ es estacionario y si para cada t ,*

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} \quad (2.4)$$

donde $\{Z_t\} \sim WN(0, \sigma^2)$ y $\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q$ son constantes.

■

Usualmente a la ecuación en diferencias (2.4) se le reescribe para tener una notación mas compacta, de la siguiente manera,

$$\phi(B)X_t = \theta(B)Z_t \quad (2.5)$$

donde $\phi_p(\cdot)$ y $\theta_q(\cdot)$ son polinomios de grados p y q respectivamente

$$\phi_p(z) = 1 - \phi_1 z - \dots - \phi_p z^p \quad (2.6)$$

y

$$\theta_q(z) = 1 + \theta_1 z - \dots + \theta_q z^q \quad (2.7)$$

estos polinomios $\phi_p(\cdot)$ y $\theta_q(\cdot)$ se les denomina como los polinomios de medias móviles autoregresivos de la ecuación en diferencias (2.4). Finalmente B es el operador de retardos definido por

$$B^j X_t = X_{t-j}, \quad j \in \mathbb{Z}$$

En la definición (2.7) no se dice mucho de (2.6) y (2.7) los polinomios de medias móviles autoregresivos, sólo se conoce son polinomios lineales con coeficientes constantes.

Por tanto, si se realizan diferentes combinaciones para estos polinomios, surgen las interrogantes acerca de la existencia y unicidad para $\{X_t\}$ y si cumple la condición de estacionariedad.

Proceso de medias móviles MA(q)

Considérese $\{X_t\}$ un proceso *ARMA* para el cual el polinomio (2.6) satisface $\phi(z) \equiv 1$, entonces el proceso $\{X_t\}$ tiene por expresión

$$X_t = \theta(B)Z_t \tag{2.8}$$

el cual es un proceso ya conocido y presentado en (2.2), *proceso de medias móviles de orden q* ó *MA(q)*.

Resulta claro que (2.8) tiene solución única, además ya se discutió el sentido de la expresión (2.8) en lo referente a la convergencia de un MA(q).

Para revisar que un MA(q) es estacionario, obsérvese que

$$\begin{aligned} E[X_t] &= E[\theta(B)Z_t] \\ &= E\left[\sum_{j=0}^q \theta_j Z_{t-j}\right] \quad \text{donde } \theta_0 = 1 \\ &= \sum_{j=0}^q \theta_j E[Z_{t-j}] \end{aligned}$$

y debido a que $Z_t \sim WN(0, \sigma)$ se llega finalmente a

$$E[X_t] = 0$$

Por otra parte, la función de autocovarianza para MA(q) es

$$Cov(X_{t+h}, X_t) = \begin{cases} \sigma^2 \sum_{j=0}^{q-|h|} \theta_j \theta_{j+|h|}, & \text{si } |h| \leq q, \\ 0, & \text{si } |h| > q \end{cases}$$

Estas características y según la definición (2.2) indican que un proceso $MA(q)$ es estacionario.

Considérese el caso particular $q=1$, entonces se tiene el proceso definido por

$$X_t = Z_t + \theta Z_{t-1},$$

donde $\{Z_t\} \sim WN(0, \sigma^2)$ y $\theta \in \mathbb{R}$. A $\{X_t\}$ se le denomina *proceso de promedios móviles o proceso $MA(1)$* .

Por su propia definición se desprende que las características de un proceso $MA(1)$ satisfacen

- a) $\mu(t) = 0$ para toda $t \in T$
- b) $\sigma^2(t) = E[X_t^2] = \sigma^2(1 + \theta^2) < \infty$

la función de autocovarianza tiene por expresión

$$\varphi_X(t+h, t) = \begin{cases} \sigma^2(1 + \theta^2), & \text{si } h = 0 \\ \sigma^2\theta, & \text{si } h = \pm 1 \\ 0, & \text{si } |h| > 1 \end{cases}$$

En la figura (2.1) se muestra una serie de tiempo obtenida por simulación de un proceso $MA(1)$ con coeficiente $\theta = -0,8$ y $\{Z_t\} \sim N(0, 1)$.

Proceso autoregresivo $AR(p)$

Considere $\{X_t\}$ un proceso $ARMA$ para el cual el polinomio (2.7) tiene por expresión $\theta(z) \equiv 1$, luego el proceso $\{X_t\}$ satisface

$$\phi(B)X_t = Z_t \tag{2.9}$$

este proceso es denominado *proceso autoregresivo de orden p , ó $AR(p)$* .

Tanto para este caso como el caso general de un proceso $\{X_t\} \sim ARMA$, debe revisarse a detalle la existencia y unicidad de una estacionaria solución X_t de la ecuación en diferencias. Aquí únicamente se examinará el caso de un $AR(1)$, que resulta sumamente ilustrativo. Sea

$$X_t = Z_t + \phi X_{t-1} \tag{2.10}$$

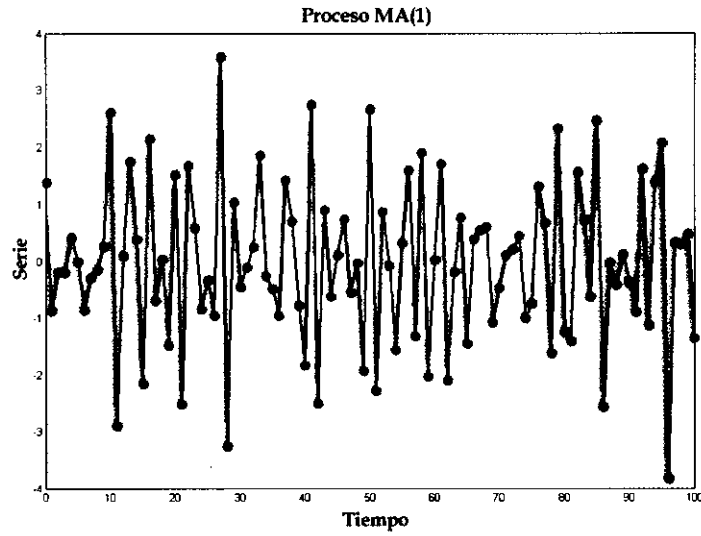


Figura 2.1: Simulación de una serie proveniente de un MA(1), a través del programa 4 del apéndice.

Iterando en (2.10) se obtiene

$$\begin{aligned}
 X_t &= Z_t + \phi Z_{t-1} + \phi^2 X_{t-2} \\
 &= \dots \\
 &= (Z_t + \phi Z_{t-1} + \dots + \phi^k Z_{t-k}) + \phi^{k+1} X_{t-k-1}
 \end{aligned}$$

Luego

$$\begin{aligned}
 X_t - (Z_t + \phi Z_{t-1} + \dots + \phi^k Z_{t-k}) &= X_t - \sum_{j=0}^k \phi^j Z_{t-j} \quad (2.11) \\
 &= \phi^{k+1} X_{t-k-1}
 \end{aligned}$$

Dado que tanto $\{X_t\}$ como $\{Z_t\}$ son procesos en $L^2(\Omega, \mathfrak{S}, P)$ la convergencia en media cuadrática define a la métrica que se tiene en $L^2(\Omega, \mathfrak{S}, P)$, por tanto

analizar la convergencia de (2.11) es equivalente a

$$\begin{aligned} \left\| X_t - \sum_{j=0}^k \phi^j Z_{t-j} \right\|^2 &= E \left[X_t - \sum_{j=0}^k \phi^j Z_{t-j} \right]^2 \\ &= E [\phi^{k+1} X_{t-k-1}]^2 \\ &= \phi^{2k+2} E [X_{t-k-1}]^2 \\ &= \phi^{2k+2} \|X_{t-k-1}\|^2 \end{aligned}$$

y como $\{X_t\}$ es estacionario, entonces tiene segundo momento finito, así

$$\left\| X_t - \sum_{j=0}^k \phi^j Z_{t-j} \right\|^2 = \phi^{2k+2} \|X_{t-k-1}\|^2 \rightarrow 0 \text{ si } k \rightarrow \infty$$

si $|\phi| < 1$.

Lo anterior significa que $\sum_{j=0}^{\infty} \phi^j Z_{t-j}$ es convergente en media cuadrática (y también en probabilidad), de donde se concluye

$$X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j} \quad (2.12)$$

lo que resuelve la existencia y unicidad para $AR(1)$.

Las características de un proceso $AR(1)$ son

$$\text{a) } \mu(t) = E[X_t] = \sum_{j=0}^{\infty} \phi^j E[Z_{t-j}] = 0$$

$$\text{b) } \sigma^2(t) = E[X_t^2] = \frac{\sigma^2}{1-\phi^2}$$

la función de autocovarianza tiene por expresión

$$\varphi_X(h) = \phi^h \frac{\sigma^2}{1-\phi^2}$$

esto muestra que $\{X_t\}$ definido por (2.12) es estacionario.

De igual manera como para los procesos $MA(1)$ se realizó una simulación del proceso $AR(1)$ con coeficiente $\phi = 0,9$, y $\{Z_t\} \sim N(0, 1)$ que puede verse en la figura (2.2).

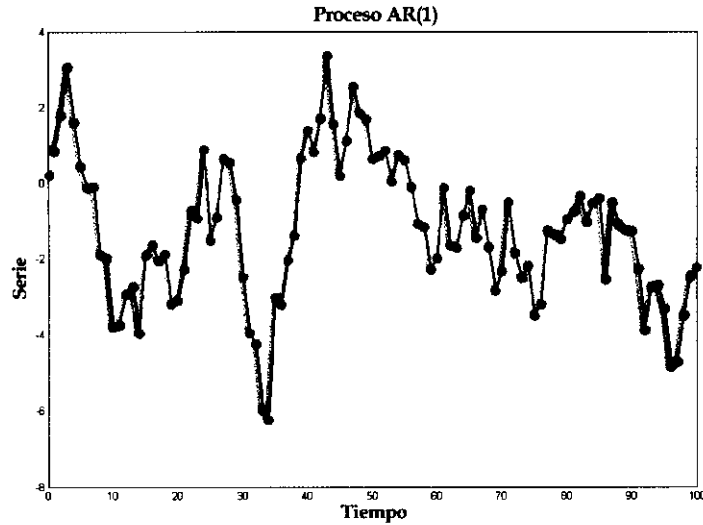


Figura 2.2: Simulación de una serie proveniente de un AR(1), a través del programa 5 del apéndice.

2.5. Causalidad e invertibilidad en un proceso ARMA(p,q)

A los procesos $\{X_t\} \sim ARMA(p, q)$ se les incorporará ahora un nuevo concepto, causalidad, que se encuentra ligado a la linealidad, (def. 2.6).

Definición 2.8 (Causalidad) Sea $\{X_t\} \sim ARMA(p, q)$ definido por las ecuaciones $\phi(B)X_t = \theta(B)Z_t$, se dice que $\{X_t\}$ es un proceso causal de $\{Z_t\}$ si existe una sucesión de constantes $\{\psi_j\}$ tales que la serie $\sum_{j=0}^{\infty} |\psi_j| < \infty$ sea absolutamente convergente y

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad t = 0, \pm 1, \dots \quad (2.13)$$

■

Debe resaltarse que la causalidad es una propiedad no únicamente del proceso $\{X_t\}$ sino más bien de la relación existente entre los dos procesos

$\{X_t\}$ y $\{Z_t\}$ que definen a un proceso *ARMA*. En ocasiones resulta conveniente visualizar la causalidad de la siguiente manera: se dirá que $\{X_t\}$ es causal cuando puede ser obtenido de $\{Z_t\}$ mediante la aplicación de un filtro lineal.

Por otro lado, debe clarificarse el sentido y significado de la suma en (2.13), para ello se tiene el siguiente resultado,

Proposición 2.3 *Sea $\{X_t\}$ sucesión de v.a. con $\sup_t E[X_t] < \infty$ y si existe una sucesión de constantes $\{\psi_j\}$ tales que la serie $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ sea absolutamente convergente, entonces la serie*

$$\psi(B)X_t = \sum_{j=-\infty}^{\infty} \psi_j B^j X_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} \quad (2.14)$$

converge absolutamente en probabilidad.

Si además, se tiene que $\sup_t E[X_t]^2 < \infty$ entonces la serie converge en media cuadrática.

Demostración

De la convergencia monótona del operador esperanza y de

$$\sup_t E[|X_t|] < \infty$$

se sigue

$$\begin{aligned} E \left[\sum_{j=-\infty}^{\infty} |\psi_j| |X_{t-j}| \right] &= \lim_{n \rightarrow \infty} E \left[\sum_{j=-n}^n |\psi_j| |X_{t-j}| \right] \\ &\leq \lim_{n \rightarrow \infty} \left(\sum_{j=-n}^n |\psi_j| \right) \sup_t E[|X_{t-j}|] \end{aligned}$$

y como se sabe que la serie $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ se concluye

$$E \left[\sum_{j=-\infty}^{\infty} \psi_j X_{t-j} \right] \leq \infty$$

esto significa que el proceso (2.14) es convergente en L^1 y por tanto también en probabilidad.

Si se supone $\sup_t E[X_t^2] < \infty$, entonces resulta conveniente definir la suma parcial $S_n = \sum_{j=-n}^n \psi_j X_{t-j}$, luego, se utilizara el criterio de Cauchy (se verá en 3.2) para verificar la convergencia en L^2 , esto es, en media cuadrática. Sean $n > m > 0$, entonces

$$\begin{aligned} E[(S_n - S_m)^2] &= E\left[\left(\sum_{m < j \leq n} \psi_j X_{t-j}\right)^2\right] \\ &= \leq \sup_t E[X_t^2] \left(\sum_{m < j \leq n} \psi_j\right)^2 \\ &\rightarrow 0 \quad \text{cuando } n, m \rightarrow \infty \end{aligned}$$

pues al ser la serie $\sum_{j=-\infty}^{\infty} \psi_j$ convergente, se sigue que $\left(\sum_{j=-\infty}^{\infty} \psi_j\right)^2$ es también convergente, y por el criterio de Cauchy para series se sigue el resultado. ■

Obsérvese que las condiciones impuestas al proceso $\{X_t\}$ son equivalentes a pedir que $\{X_t\}$ sea estacionario.

Proposición 2.4 Sea $\{X_t\}$ un proceso estacionario con función de autocovarianzas $\varphi(\bullet)$ y si existe una sucesión de constantes $\{\psi_j\}$ tales que la serie $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ sea absolutamente convergente, entonces la serie (2.14) converge absolutamente en media cuadrática y en probabilidad. Si

$$Y_t = \psi(B)X_t \tag{2.15}$$

entonces el proceso $\{Y_t\}$ es estacionario con función de autocovarianza

$$\varphi_Y(h) = \sum_{j,k=0}^{\infty} \psi_j \psi_k \varphi(h - j + k)$$

Demostración

La convergencia en media cuadrática y en probabilidad de la serie (2.14) es una consecuencia directa de la proposición (2.3).

Para verificar la estacionariedad del proceso $\{Y_t\}$, se debe recurrir una vez más a la continuidad del operador esperanza, así

$$E[Y_t] = \lim_{n \rightarrow \infty} \sum_{j=-n}^n \psi_j E[X_{t-j}] = \left(\sum_{j=-\infty}^{\infty} |\psi_j|\right) E[X_t] = k$$

pues X_t es estacionario y la sucesión $\{\psi_j\}$ es convergente, por otro lado

$$\begin{aligned}
E[(Y_{t+h}, Y_t)] &= \lim_{n \rightarrow \infty} E \left[\left(\sum_{j=-n}^n \psi_j X_{t+h-j} \right) \left(\sum_{k=-n}^n \psi_k X_{t-k} \right) \right] \\
&= \sum_{j,k=-\infty}^{\infty} \psi_j \psi_k E[X_{t+h-j} X_{t-k}] \\
&= \sum_{j,k=-\infty}^{\infty} \psi_j \psi_k (\varphi(h-j+k) + E^2[X_t])
\end{aligned}$$

esta última expresión es consecuencia de la estacionariedad de X_t .

Luego, tanto $E[Y_t]$ como $E[(Y_{t+h}, Y_t)]$ son finitos e independientes de t , lo que implica que $\{Y_t\}$ es estacionario.

Finalmente la función de autocovarianza $\varphi_Y(h)$ esta dada por la expresión

$$\begin{aligned}
\varphi_Y(h) &= E[(Y_{t+h}, Y_t)] - E[Y_{t+h}] E[Y_t] \\
&= \sum_{j,k=-\infty}^{\infty} \psi_j \psi_k \varphi(h-j+k)
\end{aligned}$$

con lo que se completa la demostración de esta proposición. ■

A continuación se enunciará uno de los resultados claves y de mayor importancia de este capítulo.

Teorema 2.5 Sea $\{X_t\}$ un proceso $ARMA(p, q)$ para el cual los polinomios de medias móviles autoregresivos $\phi(\bullet)$ y $\theta(\bullet)$ no tienen ceros o raíces comunes. Entonces $\{X_t\}$ es un proceso causal si y sólo si $\phi(z) \neq 0 \forall z \in \mathbb{C}$ tal que $|z| \leq 1$. Los coeficientes $\{\psi_j\}$ en (2.13) son determinados por la relación

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, \quad |z| \leq 1.$$
■

Del teorema (2.5) se conoce

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, \quad |z| \leq 1 \tag{2.16}$$

donde $\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$ y $\phi(z) = 1 + \phi_1 z + \dots + \phi_p z^p$.

Con la finalidad de determinar los coeficientes ψ_j se puede reescribir (2.16) igualando los coeficientes de z^j y se llega

$$\psi_j - \sum_{0 < k \leq j} \phi_k \psi_{j-k} = \theta_j, \quad 0 \leq j < \max(p, q + 1) \quad (2.17)$$

donde $\theta_0 = 1$, $\theta_j = 0$, para $j > q$ y $\phi_0 = 1$, $\phi_j = 0$, para $j > p$, y

$$\psi_j - \sum_{0 < k \leq j} \phi_k \psi_{j-k} = 0, \quad j \geq \max(p, q + 1) \quad (2.18)$$

Las ecuaciones en (2.17) son conocidas como *condiciones iniciales o condiciones de frontera* para los coeficientes ψ_j las cuales se resuelven de manera recursiva. Por otro lado en (2.18) se tiene un conjunto de ecuaciones en diferencias³ para ψ_j , cuya solución general es de la forma

$$\psi_n = \sum_{i=1}^k \sum_{j=0}^{r_i-1} \alpha_{ij} n^j \xi_i^{-n}, \quad n \geq \max(p, q + 1) - p \quad (2.19)$$

donde ξ_i , $i = 1, \dots, k$ son las raíces de la ecuación característica de la ecuación en diferencias (2.18), $\phi(z)$ mientras que r_i es el número de repeticiones ó la multiplicidad de ξ_i , y por lo tanto se tiene $\sum_{i=1}^k r_i = p$. Por su parte las constantes p, α_{ij} así como los coeficientes ψ_j , $0 \leq j \leq \max(p, q + 1) - p$ se determinan de manera única por las $\max(p, q + 1)$ condiciones de frontera (2.17). De esta forma se completa la determinación de la sucesión de coeficientes $\{\psi_j\}$ para $\{X_j\} \sim ARMA(p, q)$ causal.

Ahora se introduce un nuevo concepto, el cual esta íntimamente relacionado con la causalidad.

Definición 2.9 (Invertibilidad) Sea $\{X_t\} \sim ARMA(p, q)$ definido por las ecuaciones $\phi(B)X_t = \theta(B)Z_t$, se dice que $\{X_t\}$ es un proceso invertible de $\{Z_t\}$ si existe una sucesión de constantes $\{\pi_j\}$ tales que la serie $\sum_{j=0}^{\infty} |\pi_j| < \infty$ sea absolutamente convergente y

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \quad t = 0, \pm 1, \dots \quad (2.20)$$

³Para una referencia acerca de las ecuaciones en diferencias y sus soluciones, véase Brockwell

■

Al igual que causalidad, la invertibilidad no es únicamente propiedad del proceso $\{X_t\}$ únicamente, sino de la relación entre los dos procesos $\{X_t\}$ y $\{Z_t\}$ que aparecen en la definición de un proceso ARMA. El siguiente teorema provee de las condiciones necesarias y suficientes para la invertibilidad del proceso ARMA $\{X_t\}$ y además indica cual será la expresión de los coeficientes π_j de (2.20)

Teorema 2.6 *Sea $\{X_t\}$ un proceso ARMA(p, q) para el cual los polinomios de medias móviles autoregresivos $\phi(\bullet)$ y $\theta(\bullet)$ no tienen ceros o raíces comunes. Entonces $\{X_t\}$ es invertible si y sólo si $\theta(z) \neq 0 \forall z \in \mathbb{C}$ tal que $|z| \leq 1$. Los coeficientes $\{\pi_j\}$ en (2.13) son determinados por la relación*

$$\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)}, \quad |z| \leq 1.$$

■

2.6. La función de autocovarianza (FAC) para los procesos causales ARMA(p, q)

Se han analizado las características y propiedades de los procesos estocásticos que serán los modelos adecuados para las series de tiempo aquí analizadas, los ARMA, recordándose que ésto es gracias a la correspondencia de los procesos ARMA y sus funciones de autocovarianza ARMA. Sin embargo, no se ha visto cual es la expresión de la FAC para un proceso $\{X_t\} \sim ARMA(p, q)$, y más aún como calcularla, por esto, en este apartado se discutirán estos temas. Para ello se proponen dos métodos, el primero de ellos es de mayor importancia pues permite determinar la expresión general de la función de autocovarianza para cualquier proceso ARMA causal. El segundo método proporciona una mayor sencillez para el cálculo de la FAC.

2.6.1. Método 1

Sea $\{X_t\} \sim ARMA(p, q)$ proceso causal, entonces de la definición de (2.13) se tiene $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ luego

$$\begin{aligned} \varphi_{X_t}(h) &= E[X_{t+h}X_t] \\ &= E\left[\left(\sum_{j=0}^{\infty} \psi_j Z_{t+h-j}\right)\left(\sum_{k=0}^{\infty} \psi_k Z_{t-k}\right)\right] \end{aligned} \quad (2.21)$$

y recordando de (2.1) la expresión de $\varphi_{Z_t}(h)$, se sigue que en (2.21) se tienen sumandos ceros, salvo para aquellos valores de j y k tales que den como resultado $E[(Z_t)^2] = \sigma^2$, así ((2.21)) se llega finalmente a

$$\varphi_{X_t}(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|} \quad (2.22)$$

siendo (2.22) la *expresión general de la función de autocovarianza para un proceso causal ARMA(p, q)*. Obsérvese que la expresión (2.22) se puede determinar fácilmente ya que sólo depende de los valores que puedan adoptar los coeficientes causales $\{\psi_j\}$, los cuales se vio como determinarlos en (2.19).

2.6.2. Método 2

Considérese $\{X_t\} \sim ARMA(p, q)$ causal estacionario que satisface la ecuación en diferencias

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} \quad (2.23)$$

multiplicando ambos lados de (2.23) por X_{t-k} y tomando esperanzas de ambos lados se tiene

$$E[X_{t-k}X_t - \dots - \phi_p X_{t-k}X_{t-p}] = E[X_{t-k}(Z_t + \dots + \theta_q Z_{t-q})] \quad (2.24)$$

y dado que $\{X_t\}$ es causal entonces $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ sustituyendo en (2.24)

$$E(X_{t-k}X_t) - \dots - \phi_p E(X_{t-k}X_{t-p}) = E\left[\sum_{j=0}^{\infty} \psi_{j-k} Z_{t-k-j} (Z_t + \dots + \theta_q Z_{t-q})\right] \quad (2.25)$$

obsérvese que en el lado izquierdo de (2.25) aparece la función de autocovarianza pues $\{X_t\}$ es un proceso de media cero, por lo que $E(X_{t-k}X_t) = \varphi_X(k)$. El lado derecho de (2.25) requiere una inspección mas minuciosa, dado que no se ha especificado nada acerca de la relación de k y q , se tendrá que sólo para aquellos valores de k menores o iguales que $\max(p, q + 1)$ en el lado derecho de la expresión (2.25) se tendrán la suma de combinaciones lineales de varianzas de Z_t . En los casos de k mayores de $\max(p, q + 1)$ por las propiedades de $\{Z_t\} \sim WN(0, \sigma)$ (las autocovarianzas en distintos momentos son cero) el valor de (2.25) será cero. Así, finalmente se tiene

$$\varphi_Y(k) - \phi_1\varphi_Y(k-1) - \dots - \phi_p\varphi_Y(k-p) = \sigma^2 \sum_{k \leq j \leq q} \theta_j \psi_{j-k} \quad (2.26)$$

siempre que $0 \leq k \leq \max(p, q + 1)$, y

$$\varphi_Y(k) - \phi_1\varphi_Y(k-1) - \dots - \phi_p\varphi_Y(k-p) = 0 \quad (2.27)$$

cuando $k > \max(p, q + 1)$.

En (2.27) se tienen un conjunto de ecuaciones en diferencias⁴ para $\varphi_Y(k)$ cuya solución general al igual que en (2.19) es

$$\varphi_Y(h) = \sum_{i=1}^k \sum_{j=0}^{r_i-1} \beta_{ij} h^j \xi_i^{-h}, \quad h \geq \max(p, q + 1) - p \quad (2.28)$$

donde $\xi_i, i = 1, \dots, k$ son las raíces de la ecuación característica de la ecuación en diferencias (2.27), $\phi(z)$ mientras que r_i es el numero de repeticiones ó la multiplicidad de ξ_i , y por lo tanto se tiene $\sum_{i=1}^k r_i = p$. Por su parte las constantes β_{ij} así como las covarianzas $\varphi(j), 0 \leq j < \max(p, q + 1) - p$ se determinan de manera única por las condiciones de frontera (2.26) después del cálculo de $\psi_0, \psi_1, \dots, \psi_q$ mediante (2.19).

Es importante, para los fines de este análisis hacer un alto y contemplar lo realizado y visto hasta ahora, este capítulo trató acerca de los modelos apropiados para la componente aleatoria, procesos estocásticos que no varíen en demasía, es decir, procesos que sean estacionarios.

⁴Para una referencia acerca de las ecuaciones en diferencias y sus soluciones, veáse Brockwell

Siguiendo esta línea y para estar acorde con las ideas centrales de no explosividad de la variable, mediante el supuesto de estacionariedad en las series, y de acuerdo a que la información con la que se dispone es únicamente su pasado se espera que resulte suficiente para conocer y comprender su desarrollo y posterior comportamiento del fenómeno que describe. Por tanto, es de vital importancia encontrar un intermediario entre los procesos o modelos que se utilizarán con la serie, dicho vínculo debe ser coherente y reflejar el papel crucial que desempeña la información disponible de la serie.

El vínculo natural y formal entre los modelos con los datos de la serie, y la elección de cada proceso a utilizar es mediante la función de autocovarianza que considera las interrelaciones de la serie con su pasado y futuro, convirtiéndose de esta forma en una herramienta sumamente importante, y como se vió en el capítulo, caracteriza al proceso, esto es, caracteriza y determina al modelo a emplear.

Dado lo anterior, el siguiente paso, es como calcular tan importante función, para ello se necesita enriquecer la teoría y desarrollar un mayor análisis que dió por resultado dos importantes métodos para alcanzar tales fines.

Sin embargo, es importante no perder de vista el camino alcanzado hasta este punto, ya se es posible determinar el proceso a emplear para modelar los datos, pero falta lo referente a la predicción, lo cual se alcanza en los dos capítulos posteriores.

Capítulo 3

Espacios de Hilbert

Este capítulo es el de mayor trascendencia en toda la tesis, pues en él se revisa la teoría que facilitará la culminación en el análisis de las series de tiempo: *la predicción*, así se pide al lector no pierda detalle de los temas discutidos aquí, sin embargo, debe aclararse el porque se hace esta solicitud precisamente en este capítulo.

Es en este lugar, donde, por su contenido, aparentemente se pierde conexión con todo lo desarrollado en capítulos anteriores, en el que se parte de un espacio vectorial complejo y se va enriqueciendo algebraicamente hasta llegar a conclusiones que al aterrizarlas en el enfoque de las series de tiempo, facilitan los resultados, y le dan un sabor especial. Por tanto, aquí se da un paseo por álgebra lineal, topología, y análisis matemático, para que al final se regrese con los resultados en la mano y listos para usarse en la predicción de las series de tiempo.

La teoría y análisis que se desarrolla en este capítulo se sigue a Brockwell y Davis [1].

3.1. Espacios vectoriales con producto interior y sus propiedades

Definición 3.1 (Espacio vectorial) *Un espacio vectorial es un conjunto no vacío \mathcal{H} en el cual se tienen dos operaciones entre sus elementos (llamados vectores), a saber, la suma de ellos $+$: $\mathcal{H} \times \mathcal{H} \rightarrow \mathcal{H}$ con la cual a cada \mathbf{x} , $\mathbf{y} \in \mathcal{H}$ se le asocia un nuevo vector suma $(\mathbf{x} + \mathbf{y}) \in \mathcal{H}$, y el producto de un vector de \mathcal{H} por un escalar (un elemento de un campo \mathcal{K}) \cdot : $\mathcal{K} \times \mathcal{H} \rightarrow \mathcal{H}$, con*

la cual, dado un $\mathbf{x} \in \mathcal{H}$ y un escalar $\lambda \in \mathcal{K}$, se le asocia un nuevo elemento $\lambda\mathbf{x} \in \mathcal{H}$, llamado producto del vector \mathbf{x} por el escalar λ , cumpliendo las siguientes propiedades:

La suma es conmutativa: $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$, para todo $\mathbf{x}, \mathbf{y} \in \mathcal{H}$

La suma es asociativa: $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$, para todo $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{H}$

Existe $0 \in \mathcal{H}$ tal que $\mathbf{x} + 0 = \mathbf{x}$, para todo $\mathbf{x} \in \mathcal{H}$

Para todo $\mathbf{x} \in \mathcal{H}$ existe $(-\mathbf{x}) \in \mathcal{H}$ tal que $\mathbf{x} + (-\mathbf{x}) = 0$ (inverso aditivo)

$\lambda(\mathbf{x} + \mathbf{y}) = \lambda\mathbf{x} + \lambda\mathbf{y}$ para todo $\mathbf{x}, \mathbf{y} \in \mathcal{H}$ y $\lambda \in \mathcal{K}$. (propiedad distributiva).

$(\lambda + \varphi)\mathbf{x} = \lambda\mathbf{x} + \varphi\mathbf{x}$ para todo $\mathbf{x} \in \mathcal{H}$ y $\lambda, \varphi \in \mathcal{K}$.

$1\mathbf{x} = \mathbf{x}$, para todo $\mathbf{x} \in \mathcal{H}$ y $1 \in \mathcal{K}$. (neutro multiplicativo).

■

Observación 3.1 Un campo cualquiera \mathcal{K} es en sí mismo un espacio vectorial, tomando como escalares a los mismos elementos de \mathcal{K} .

Definición 3.2 (Espacio vectorial complejo) El campo de los números complejos \mathbb{C} adquiere la estructura algebraica de espacio vectorial complejo si se toman por escalares a los mismos elementos de \mathbb{C} .

■

Definición 3.3 Un producto interior sobre un espacio vectorial complejo \mathcal{H} , es un mapeo $\langle \cdot, \cdot \rangle : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{C}$ que a cada par de elementos x e y en \mathcal{H} , le asocia el complejo $\langle x, y \rangle$, llamado el producto interior de x y y , tal que

a) $\langle x, y \rangle = \overline{\langle y, x \rangle}$, donde la barra significa conjugación compleja,

b) $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ para todo $x, y, z \in \mathcal{H}$,

c) $\langle \lambda x, y \rangle = \lambda \langle x, y \rangle$ para todo $x, y \in \mathcal{H}$ y $\lambda \in \mathbb{C}$,

d) $\langle x, x \rangle \geq 0$ para todo $x \in \mathcal{H}$,

e) $\langle x, x \rangle = 0$ si y sólo si $x = 0$.

■

Observación 3.2 Para todo $x \in \mathcal{H}$ el producto interior $\langle x, x \rangle$ es un complejo con parte real positiva y parte imaginaria nula, en otras palabras, $\langle x, x \rangle : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}^+$.

Un espacio vectorial con producto interior es el espacio euclidiano n -dimensional, es decir, el conjunto de todos los vectores

$$x = (x_1, x_2, \dots, x_n)' \in \mathbb{R}^n$$

este espacio es un espacio con producto interior al definirse el producto interior de la manera usual, como sigue:

$$\langle x, y \rangle = \sum_{i=1}^n x_i y_i.$$

■

Teorema 3.1 Sea \mathcal{H} un espacio vectorial complejo con producto interior entonces para todo $x, y, w \in \mathcal{H}$ y $\lambda \in \mathbb{C}$ se tiene

$$\mathbf{a}) \langle (x + y), (z + w) \rangle = \langle x, z \rangle + \langle x, w \rangle + \langle y, z \rangle + \langle y, w \rangle$$

$$\mathbf{b}) \langle x, \lambda y \rangle = \bar{\lambda} \langle x, y \rangle$$

$$\mathbf{c}) \langle x, 0 \rangle = 0 = \langle 0, x \rangle$$

$$\mathbf{d}) \langle x, z \rangle = \langle y, z \rangle \text{ si y sólo si } x = y$$

Demostración:

De los incisos **a)** y **b)** de la definición 3.3 se sigue que

$$\begin{aligned} \langle (x + y), (z + w) \rangle &= \langle x, (z + w) \rangle + \langle y, (z + w) \rangle \\ &= \overline{\langle (z + w), x \rangle} + \overline{\langle (z + w), y \rangle} \\ &= \overline{\langle z, x \rangle} + \overline{\langle w, x \rangle} + \overline{\langle z, y \rangle} + \overline{\langle w, y \rangle} \\ &= \langle x, z \rangle + \langle x, w \rangle + \langle y, z \rangle + \langle y, w \rangle \end{aligned}$$

quedando así probado el inciso **a)**.

Por otro lado, de nuevo por la def. 3.3 incisos **a)** y **c)**

$$\begin{aligned} \langle x, \lambda y \rangle &= \overline{\langle \lambda y, x \rangle} \\ &= \overline{\lambda \langle y, x \rangle} \\ &= \bar{\lambda} \langle x, y \rangle \end{aligned}$$

por tanto, el inciso **b)** del teorema queda probado.

El inciso **c)** se sigue de tomar el caso particular $\lambda = 0$ en el inciso **b)** del presente teorema y del inciso **c)** de la def. 3.3.

Por el inciso **b)** y **c)** de la def. 3.3 se tiene

$$\begin{aligned} \langle x, z \rangle &= \langle y, z \rangle \text{ si y sólo si} \\ \langle x, z \rangle - \langle y, z \rangle &= 0 \text{ si y sólo si} \\ \langle x - y, z \rangle &= 0 \text{ si y sólo si} \\ x - y &= 0 \text{ si y sólo si} \\ x &= y. \end{aligned}$$

Un concepto que se deriva de manera inmediata cuando se trabaja en espacios con producto punto, es, la *norma de un vector*. ■

Definición 3.4 La norma de un elemento x de un espacio vectorial complejo con producto interior se define como

$$\|x\| = \sqrt{\langle x, x \rangle}.$$

Observación 3.3 De la definición anterior y la observación 3.2 se tiene que la norma es una operación real positiva, i.e. $\|\cdot\| : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}^+$. ■

Ejemplo 3.1 La norma, en el espacio vectorial euclidiano n -dimensional del ejemplo 3.2, es definido como $\|x\| = \sqrt{\sum_{i=1}^n x_i^2}$

Un resultado de gran utilidad que asocia al producto punto de dos elementos cualesquiera del espacio con sus normas es el siguiente, ■

Teorema 3.2 (Desigualdad de Cauchy-Schwarz) Sea \mathcal{H} un espacio con producto interior, entonces

$$|\langle x, y \rangle| \leq \|x\| \|y\|$$

para toda $x, y \in \mathcal{H}$.

Demostración:

La desigualdad a probar es equivalente a:

$$\langle x, y \rangle^2 \leq \langle x, x \rangle \langle y, y \rangle.$$

Sean $x, y \in \mathcal{H}$ y $\lambda \in \mathbb{C}$, y como $\langle x, y \rangle$ es un complejo entonces se puede escribir de la forma $\langle x, y \rangle = be^{i\theta}$ donde $b = |\langle x, y \rangle|$ y $\theta \in (-\pi, \pi)$.

Utilizando las propiedades **a)** y **b)** del teorema 3.1 se tiene

$$\begin{aligned} \langle x - \lambda e^{i\theta} y, x - \lambda e^{i\theta} y \rangle &= \langle x, x \rangle - \lambda e^{-i\theta} \langle x, y \rangle - \lambda e^{i\theta} \langle y, x \rangle + \lambda^2 (e^{i\theta})(e^{-i\theta}) \langle y, y \rangle \\ &= \langle x, x \rangle - \lambda e^{-i\theta} (be^{i\theta}) - \lambda e^{i\theta} (be^{-i\theta}) + \lambda^2 \langle y, y \rangle \\ &= \langle x, x \rangle - 2\lambda b + \lambda^2 \langle y, y \rangle \end{aligned}$$

Para satisfacer que el producto interior sea no negativo es necesario asegurarse que el valor mínimo de la última expresión, que se puede considerar una función en $\lambda \in \mathbb{R}$, sea no negativo.

Utilizando cálculo elemental para la función

$$f(\lambda) = \langle x, x \rangle - 2\lambda b + \lambda^2 \langle y, y \rangle$$

se obtiene el punto crítico $\lambda = \frac{b}{\langle y, y \rangle}$, que por el criterio de la segunda derivada se comprueba que se trata de un mínimo.

Finalmente se necesita que la función en tal mínimo sea no negativa, luego:

$$0 \leq f\left(\frac{b}{\langle y, y \rangle}\right) = \langle x, x \rangle - 2\frac{b}{\langle y, y \rangle}b + \left(\frac{b}{\langle y, y \rangle}\right)^2 \langle y, y \rangle$$

si y sólo si

$$0 \leq \langle x, x \rangle - \frac{b^2}{\langle y, y \rangle}$$

que reescribiendo se llega finalmente al resultado deseado,

$$\begin{aligned} b^2 &\leq \langle x, x \rangle \langle y, y \rangle \\ \langle x, y \rangle^2 &\leq \langle x, x \rangle \langle y, y \rangle. \end{aligned}$$

■

Teorema 3.3 .(Desigualdad del triángulo) Si \mathcal{H} es un espacio con producto interior, entonces

$$\|x + y\| \leq \|x\| + \|y\|$$

para todo $x, y \in \mathcal{H}$.

Demostración:

$$\begin{aligned} \|x + y\|^2 &= \langle x + y, x + y \rangle \\ &= \langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y, y \rangle \end{aligned}$$

y, utilizando la desigualdad de Cauchy-Schwarz, se tiene

$$\|x + y\|^2 \leq \|x\|^2 + 2\|x\| \|y\| + \|y\|^2$$

es decir,

$$\|x + y\|^2 \leq (\|x\| + \|y\|)^2.$$

■

Hasta el momento solamente se ha definido la norma de un vector de un espacio con producto interior, pero, no se han enunciado propiedades de la norma, por tanto se presenta ahora la siguiente proposición.

Proposición 3.4 (Propiedades de la norma) *Sea \mathcal{H} un espacio vectorial complejo con producto interior y $\|x\|$ definida como en la def.3.4, entonces:*

- a) $\|x + y\| \leq \|x\| + \|y\|$ para todo $x, y \in \mathcal{H}$, (desigualdad del triángulo)
- b) $\|\lambda x\| = |\lambda| \|x\|$ para todo $x \in \mathcal{H}$ y todo $\lambda \in \mathbb{C}$,
- c) $\|x\| \geq 0$ para todo $x \in \mathcal{H}$,
- d) $\|x\| = 0$ si y sólo si $x = 0$.

Demostración:

La propiedad a) se sigue de la desigualdad del triángulo, mientras que las demás

propiedades son consecuencia de las definiciones 3.3 y 3.4.

■

Proposición 3.5 (Ley del paralelogramo) *Si \mathcal{H} es un espacio complejo con producto interior, entonces*

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2)$$

Demostración

Por definición se tiene

$$\begin{aligned} \|x + y\|^2 + \|x - y\|^2 &= [\langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y, y \rangle] + \\ &\quad [\langle x, x \rangle + \langle x, -y \rangle + \langle -y, x \rangle + \langle y, y \rangle] \end{aligned}$$

y asociando,

$$\begin{aligned} \|x + y\|^2 + \|x - y\|^2 &= [\langle x, y \rangle + \langle y, x \rangle - \langle x, y \rangle - \langle y, x \rangle] + \\ &\quad [\langle x, x \rangle + \langle x, x \rangle + \langle y, y \rangle + \langle y, y \rangle] \\ &= 2(\langle x, x \rangle + \langle x, x \rangle) \\ &= 2(\|x\|^2 + \|y\|^2). \end{aligned}$$

■

Definición 3.5 (Distancia) La distancia entre dos elementos x, y de un espacio vectorial complejo se define como

$$d(x, y) = \|y - x\|.$$

■

Proposición 3.6 (Propiedades de la distancia) Sea \mathcal{H} un espacio vectorial complejo con producto interior y distancia definida como en la def.3.5, entonces para todo $x, y, z \in \mathcal{H}$, se tiene

a) $d(x, y) \geq 0$, $d(x, y) = 0$ si y sólo si $x = y$.

b) $d(x, y) = d(y, x)$

c) $d(x, y) \leq d(x, z) + d(z, y)$.

Demostración:

La propiedad **a)** se sigue de la definición misma y de las propiedades **c)** y **d)** de la proposición 3.4. La propiedad **b)** se sigue de la propiedad **b)** de la proposición 3.4 $d(x, y) = \|y - x\| = \|(-1)(x - y)\| = |-1| \|x - y\| = \|x - y\| = d(y, x)$.

Finalmente la propiedad **c)** se sigue de la desigualdad del triángulo de la proposición 3.4, pues

$$\begin{aligned} d(x, y) &= \|x - y\| \\ &= \|(x - z) + (z - y)\| \\ &\leq \|x - z\| + \|z - y\| \\ &= d(x, z) + d(z, y). \end{aligned}$$

■

Definición 3.6 (Convergencia en norma) Una sucesión $\{x_n, n = 1, 2, 3, \dots\}$ de elementos del espacio complejo con producto interior \mathcal{H} se dice que es convergente en norma a $x \in \mathcal{H}$, si $\|x_n - x\| \rightarrow 0$, cuando $n \rightarrow \infty$.

■

Teorema 3.7 (Continuidad del producto interior) Si $\{x_n\}$ y $\{y_n\}$ son sucesiones de elementos de un espacio complejo con producto interior \mathcal{H} , tales que $\|x_n - x\| \rightarrow 0$, y $\|y_n - y\| \rightarrow 0$, entonces

a) $\|x_n\| \rightarrow \|x\|$ **b)** $\langle x_n, y_n \rangle \rightarrow \langle x, y \rangle$

Demostración:

Partiendo de la desigualdad del triángulo se sigue que,

$$\|x_n\| = \|(x_n - x) + x\| \leq \|x_n - x\| + \|x\|$$

es decir,

$$\|x_n\| - \|x\| \leq \|x_n - x\| \tag{3.1}$$

Por otro lado,

$$\|x\| = \|(x - x_n) + x_n\| \leq \|x - x_n\| + \|x_n\|$$

luego

$$\|x_n\| - \|x\| \geq -\|x - x_n\| \tag{3.2}$$

Entonces por 3.1 y por 3.2

$$0 \leq |\|x_n\| - \|x\|| \leq \|x_n - x\| \rightarrow 0$$

*quedando así probado lo establecido en **a**).*

Considérese

$$|\langle x_n, y_n \rangle - \langle x, y \rangle| = |\langle x_n, y_n - y \rangle + \langle x_n - x, y \rangle|$$

por la desigualdad del triángulo

$$|\langle x_n, y_n \rangle - \langle x, y \rangle| \leq |\langle x_n, y_n - y \rangle| + |\langle x_n - x, y \rangle|$$

y, por Cauchy_Schwarz se llega

$$0 \leq |\langle x_n, y_n \rangle - \langle x, y \rangle| \leq \|x_n\| \|y_n - y\| + \|x_n - x\| \|y\| \rightarrow 0$$

*por lo probado en **a**) y por las hipótesis de la proposición.*

■

3.2. Espacios de Hilbert

Cuando se trabaja con un espacio con producto interno que tenga la propiedad adicional de completéz, se dice que dicho espacio es un *Espacio de Hilbert*.

Definición 3.7 (Sucesión de Cauchy) Una sucesión $\{x_n \ n = 1, 2, 3, \dots\}$ de elementos del espacio con producto interior \mathcal{H} es llamada sucesión de Cauchy si $\|x_n - x_m\| \rightarrow 0$, cuando $m, n \rightarrow \infty$, es decir, para cada $\varepsilon > 0$, existe un entero positivo $N(\varepsilon)$ tal que $\|x_n - x_m\| \leq \varepsilon$ para todo $m, n > N(\varepsilon)$.

■

Definición 3.8 (Espacio de Hilbert) Un espacio de Hilbert \mathcal{H} es un espacio con producto interior y que además es completo, esto es, es un espacio con producto interior en el cual toda sucesión de Cauchy $\{x_n\}$ es convergente en norma a algún elemento $x \in \mathcal{H}$.

■

Proposición 3.8 Si $\{x_n\}$ es una sucesión de elementos pertenecientes a un Espacio de Hilbert \mathcal{H} , entonces $\{x_n\}$ converge en norma si y sólo si $\|x_n - x_m\| \rightarrow 0$ cuando $m, n \rightarrow \infty$.

Demostración:

La suficiencia de la proposición se sigue por definición de completitud en un Espacio de Hilbert, mientras que la necesidad que asevera la proposición es consecuencia de la desigualdad del triángulo, ya que si $\|x_n - x\| \rightarrow 0$, entonces $0 \leq \|x_n - x_m\| \leq (\|x_n - x\| + \|x - x_m\|) \rightarrow 0$, cuando $n, m \rightarrow \infty$.

■

3.3. El Teorema de proyección

En esta sección con la ayuda del teorema de proyección, que se verá un poco más adelante, se podrá tener una herramienta de vital importancia para poder enfrentar de una manera más natural, problemas de aproximación y predicción de un cierto tipo de variables aleatorias.

Sin embargo, previamente es necesario definir una nueva terminología.

Definición 3.9 (Subespacio cerrado) Un subespacio lineal \mathcal{M} de un espacio de Hilbert \mathcal{H} , es llamado un subespacio cerrado de \mathcal{H} si \mathcal{M} contiene todos sus puntos límites, esto es, si $\{x_n\} \in \mathcal{M}$ y $\|x_n - x\| \rightarrow 0$, entonces esto implica que $x \in \mathcal{M}$.

■

Definición 3.10 (Elementos ortogonales) Sea x, y elementos de \mathcal{H} , entonces se dice que x e y son ortogonales (y se denota por $x \perp y$), si y sólo si

$$\langle x, y \rangle = 0.$$

■

Definición 3.11 (Complemento ortogonal) El complemento ortogonal de un subconjunto \mathcal{M} de \mathcal{H} , se define como el conjunto \mathcal{M}^\perp formado de todos los elementos de \mathcal{H} , que son ortogonales a cada elemento de \mathcal{M} .

■

Proposición 3.9 Sea \mathcal{M} subespacio del espacio de Hilbert \mathcal{H} , entonces \mathcal{M}^\perp es un subespacio cerrado de \mathcal{H} .

Demostración:

Sea $\{x_n\} \in \mathcal{M}^\perp$ sucesión convergente en norma, es decir, $\|x_n - x\| \rightarrow 0$, entonces el problema a resolver es mostrar que $x \in \mathcal{M}^\perp$.

Por tanto, considérese cualquier elemento $y \in \mathcal{M}$ y constrúyase la sucesión $\{y_n = y, n = 1, 2, 3, \dots\}$, y obsérvese que $\|y_n - y\| \rightarrow 0$.

Luego, de la continuidad del producto interior, se tiene que

$$\langle x_n, y_n \rangle \rightarrow \langle x, y \rangle$$

es evidente que $\langle x_n, y_n \rangle = 0$ para toda $n = 1, 2, 3, \dots$ debido a la definición de elementos ortogonales, entonces $\{\langle x_n, y_n \rangle, n = 1, 2, \dots\}$ se trata de la sucesión constante cero, que converge en norma al elemento cero, en otras palabras $\langle x, y \rangle = 0$.

■

Teorema 3.10 (El Teorema de proyección) Sea \mathcal{M} un subespacio cerrado del espacio de Hilbert \mathcal{H} , y sea $x \in \mathcal{H}$, entonces

i) existe un único elemento $\hat{x} \in \mathcal{M}$ tal que

$$\|x - \hat{x}\| = \inf_{y \in \mathcal{M}} \|x - y\|,$$

y

ii) $\hat{x} \in \mathcal{M}$ y $\|x - \hat{x}\| = \inf_{y \in \mathcal{M}} \|x - y\|$ si y sólo si $(x - \hat{x}) \in \mathcal{M}^\perp$.

Al elemento \hat{x} se le denomina la proyección ortogonal de x sobre \mathcal{M} .

Demostración i):

Sea $x \in \mathcal{H}$, y defínase $d = \inf_{y \in \mathcal{M}} \|x - y\|^2$, que existe debido a que suponemos a $\mathcal{M} \neq \emptyset$. Ahora bien, con elementos $\{y_n\} \in \mathcal{M}$ es posible construir una sucesión tal que

$$\|x - y_n\|^2 \rightarrow d$$

como sigue:

Sea $y \in \mathcal{M}$ y calcúlese $\|y - x\|^2 = q$, donde $q = d + \varepsilon$, y , para encontrar a los siguientes elementos de la sucesión se procede de la siguiente manera, y_1 es tal que

$$\|x - y_1\|^2 = d + \frac{\varepsilon}{2}$$

y continuando de esa forma se llega finalmente

$$\|x - y_n\|^2 = d + \frac{\varepsilon}{2^n}$$

es inmediato que si $n \rightarrow \infty$ entonces $\|x - y_n\|^2 \rightarrow d$.

Sin embargo, hace falta mostrar explícitamente que la sucesión $\{y_n\}$ es convergente y, para tal fin mostraremos que la sucesión $\{y_n\}$ es de Cauchy:

Sean $m, n \in \mathbb{N}$ luego

$$0 \leq \|y_m - y_n\|^2 = \|(y_m - x) - (y_n - x)\|^2$$

se define $a = y_m - x$ y $b = y_n - x$ y por la ley del paralelogramo se tiene que

$$\|a - b\|^2 = 2(\|a\|^2 + \|b\|^2) - \|a + b\|^2$$

es decir

$$0 \leq \|y_m - y_n\|^2 = 2(\|y_m - x\|^2 + \|y_n - x\|^2) - \left\| 2 \left(\frac{y_m + y_n}{2} - x \right) \right\|^2$$

como \mathcal{M} es subespacio de \mathcal{H} , entonces $\frac{y_m + y_n}{2} \in \mathcal{M}$, y por propiedades de la norma

$$0 \leq \|y_m - y_n\|^2 = -4 \left\| \left(\frac{y_m + y_n}{2} \right) - x \right\|^2 + 2(\|y_m - x\|^2 + \|y_n - x\|^2)$$

tomando $m, n \rightarrow \infty$

$$0 \leq \|y_m - y_n\|^2 \rightarrow -4d + 2(2d) = 0$$

Por tanto $\{y_n\}$ es de Cauchy, por tal razón y por que \mathcal{M} es cerrado se asegura que existe $\hat{x} \in \mathcal{M}$ tal que $\|y_n - \hat{x}\|^2 \rightarrow 0$.

Por la continuidad del producto interior

$$\|x - \hat{x}\|^2 = \lim_{n \rightarrow \infty} \|x - y_n\|^2 = d$$

Para mostrar la unicidad, se hará por contradicción, esto es, supóngase que existe $\hat{y} \in \mathcal{M}$ tal que $\|x - \hat{x}\|^2 = \|x - \hat{y}\|^2 = d$, entonces aplicando el mismo argumento anterior por la ley del paralelogramo se tiene que

$$0 \leq \|\hat{x} - \hat{y}\|^2 = 2(\|\hat{x} - x\|^2 + \|\hat{y} - x\|^2) - 4\left\|\left(\frac{\hat{x} + \hat{y}}{2} - x\right)\right\|^2$$

$$0 \leq \|\hat{x} - \hat{y}\|^2 \rightarrow -4d + 2(2d) = 0$$

y por tanto $\hat{x} = \hat{y}$.

Para la prueba del inciso ii) considere $\hat{x} \in \mathcal{M}$, $\|x - \hat{x}\| = \inf_{y \in \mathcal{M}} \|x - y\|$ y se procederá por contradicción, esto es, supondremos $(x - \hat{x}) \notin \mathcal{M}^\perp$, lo que llevara a la siguiente contradicción: \hat{x} no es el elemento más cercano en \mathcal{M} a x .

El suponer $(x - \hat{x}) \notin \mathcal{M}^\perp$ implica que al menos existe un elemento $y \in \mathcal{M}$ tal que $\langle x - \hat{x}, y \rangle \neq 0$, con dicho elemento constrúyase

$$\tilde{x} = \hat{x} + \frac{ay}{\|y\|^2}$$

donde $a = \langle x - \hat{x}, y \rangle$, luego

$$\begin{aligned} \|x - \tilde{x}\|^2 &= \langle (x - \tilde{x} + \hat{x} - \hat{x}), (x - \tilde{x} + \hat{x} - \hat{x}) \rangle \\ &= \langle (x - \hat{x}) + (\hat{x} - \tilde{x}), (x - \hat{x}) + (\hat{x} - \tilde{x}) \rangle \\ &= \|x - \hat{x}\|^2 + \langle (x - \hat{x}), (\hat{x} - \tilde{x}) \rangle + \langle (\hat{x} - \tilde{x}), (x - \hat{x}) \rangle + \langle (\hat{x} - \tilde{x}), (\hat{x} - \tilde{x}) \rangle \\ &= \|x - \hat{x}\|^2 + 2\operatorname{Re}[\langle (\hat{x} - \tilde{x}), (x - \hat{x}) \rangle] + \|\hat{x} - \tilde{x}\|^2 \end{aligned}$$

y recordando como se definió \tilde{x}

$$\begin{aligned}
\|x - \tilde{x}\|^2 &= \|x - \hat{x}\|^2 + \left\| -\frac{ay}{\|y\|^2} \right\|^2 + 2 \operatorname{Re} \left[\left\langle \left(\hat{x} - \hat{x} - \frac{ay}{\|y\|^2} \right), (x - \hat{x}) \right\rangle \right] \\
&= \|x - \hat{x}\|^2 + \frac{|a|^2}{\|y\|^2} + 2 \operatorname{Re} \left[-\frac{a}{\|y\|^2} \langle y, (x - \hat{x}) \rangle \right] \\
&= \|x - \hat{x}\|^2 + \frac{|a|^2}{\|y\|^2} - 2 \operatorname{Re} \left[\frac{a\bar{a}}{\|y\|^2} \right] \\
&= \|x - \hat{x}\|^2 + \frac{|a|^2}{\|y\|^2} - 2 \operatorname{Re} \left[\frac{a\bar{a}}{\|y\|^2} \right]
\end{aligned}$$

del cálculo complejo se sigue que $a\bar{a} = |a|^2$,

$$\|x - \tilde{x}\|^2 = \|x - \hat{x}\|^2 + \frac{|a|^2}{\|y\|^2} - 2 \operatorname{Re} \left[\frac{|a|^2}{\|y\|^2} \right]$$

y de las notas 3.2 y 3.3 se sigue que la parte real (denotada por Re) de $\left[\frac{|a|^2}{\|y\|^2} \right]$ es el mismo número, de modo que

$$\|x - \tilde{x}\|^2 = \|x - \hat{x}\|^2 - \frac{|a|^2}{\|y\|^2} < \|x - \hat{x}\|^2$$

luego \hat{x} no corresponde a su definición.

Inversamente,

Si $\hat{x} \in \mathcal{M}$ y $(x - \hat{x}) \in \mathcal{M}^\perp$, entonces tómese $y \in \mathcal{M}$ luego

$$\begin{aligned}
\|x - y\|^2 &= \|x - y + (\hat{x} - \hat{x})\|^2 = \|x - \hat{x} + (\hat{x} - y)\|^2 \\
&= \langle (x - \hat{x}) + (\hat{x} - y), (x - \hat{x}) + (\hat{x} - y) \rangle \\
&= \|x - \hat{x}\|^2 + \langle (x - \hat{x}), (\hat{x} - y) \rangle + \overline{\langle (x - \hat{x}), (\hat{x} - y) \rangle} + \|\hat{x} - y\|^2
\end{aligned}$$

pero, como $(\hat{x} - y) \in \mathcal{M}$ se tiene

$$\|x - y\|^2 = \|x - \hat{x}\|^2 + \|\hat{x} - y\|^2$$

Es decir, todo elemento $y \in \mathcal{M}$ “dista de $x \in \mathcal{H}$ ” más que lo que “dista \hat{x} de x ”, por lo tanto, se concluye que \hat{x} es el único elemento por (i), que satisface

$$\|x - \hat{x}\| = \inf_{y \in \mathcal{M}} \|x - y\|.$$

■

El teorema de proyección asegura la existencia de un elemento \hat{x} en un subespacio cerrado \mathcal{M} de un espacio de Hilbert \mathcal{H} , que es el más “próximo” en norma a cualquier elemento $x \in \mathcal{H}$, y más aún, el teorema brinda la manera de encontrar dicho $\hat{x} \in \mathcal{M}$ para cada $x \in \mathcal{H}$, \hat{x} es aquel que satisfaga las llamadas *Ecuaciones de Predicción*:

$$\langle x - \hat{x}, y \rangle = 0, \forall y \in \mathcal{M}, x \in \mathcal{H}.$$

Corolario 3.11 (El mapeo proyección de \mathcal{H} sobre \mathcal{M}) *Si \mathcal{M} es un subespacio cerrado del espacio de Hilbert \mathcal{H} y sea I el mapeo identidad en \mathcal{H} , entonces existe un único mapeo $P_{\mathcal{M}}$ de \mathcal{H} sobre \mathcal{M} tal que, $I - P_{\mathcal{M}}$, mapea \mathcal{H} sobre \mathcal{M}^{\perp} , (i.e. $P_{\mathcal{M}^{\perp}} = I - P_{\mathcal{M}}$).*

A $P_{\mathcal{M}}$ se le denomina el mapeo proyección de \mathcal{H} sobre \mathcal{M} .

Demostración:

Por el Teorema de Proyección se sigue que para cada $x \in \mathcal{H}$, existe un único $\hat{x} \in \mathcal{M}$ tal que $x - \hat{x} \in \mathcal{M}^{\perp}$ entonces el mapeo requerido es

$$P_{\mathcal{M}}(x) = \hat{x}.$$

■

Proposición 3.12 (Propiedades de los mapeo proyecciones) *Sea \mathcal{H} espacio de Hilbert y sea $P_{\mathcal{M}}$ el mapeo proyección sobre un subespacio cerrado \mathcal{M} .*

Entonces,

i) $P_{\mathcal{M}}(\alpha x + \beta y) = \alpha P_{\mathcal{M}}(x) + \beta P_{\mathcal{M}}(y)$, donde $x, y \in \mathcal{H}$, $\alpha, \beta \in \mathbf{C}$

ii) $\|x\|^2 = \|P_{\mathcal{M}}(x)\|^2 + \|P_{\mathcal{M}^{\perp}}(x)\|^2$,

iii) cada $x \in \mathcal{H}$ tiene una única representación como suma de un elemento de \mathcal{M} y de

un elemento de \mathcal{M}^{\perp} , i.e.

$$x = P_{\mathcal{M}}(x) + P_{\mathcal{M}^{\perp}}(x) \tag{3.3}$$

iv) $P_{\mathcal{M}}(x_n) \rightarrow P_{\mathcal{M}}(x)$ si $\|x_n - x\| \rightarrow 0$,

v) $x \in \mathcal{M}$ si y sólo si $P_{\mathcal{M}}(x) = x$,

vi) $x \in \mathcal{M}^{\perp}$ si y sólo si $P_{\mathcal{M}}(x) = 0$, y

vii) $\mathcal{M}_1 \subseteq \mathcal{M}_2$ si y sólo si $P_{\mathcal{M}_1}(P_{\mathcal{M}_2}(x)) = P_{\mathcal{M}_1}(x)$

Demostración:

i) Puesto que \mathcal{M} es un subespacio de \mathcal{H} , se sigue que dados $u, v \in \mathcal{M}$ la operación combinación lineal es una operación cerrada en \mathcal{M} , entonces $\alpha P_{\mathcal{M}}(x) + \beta P_{\mathcal{M}}(y) \in \mathcal{M}$, con $\alpha, \beta \in \mathbb{C}$.

Por otro lado,

$$\alpha x + \beta y - \alpha P_{\mathcal{M}}(x) + \beta P_{\mathcal{M}}(y) = \alpha(x - P_{\mathcal{M}}(x)) + \beta(y - P_{\mathcal{M}}(y))$$

donde $x - P_{\mathcal{M}}(x) \in \mathcal{M}^\perp$ al igual que $y - P_{\mathcal{M}}(y) \in \mathcal{M}^\perp$, entonces por la proposición P4, se tiene que \mathcal{M}^\perp es un subespacio lineal de \mathcal{H} , y por tanto $\alpha(x - P_{\mathcal{M}}(x)) + \beta(y - P_{\mathcal{M}}(y)) \in \mathcal{M}^\perp$.

Las dos propiedades anteriores son las que identifican a $\alpha P_{\mathcal{M}}(x) + \beta P_{\mathcal{M}}(y)$ como el mapeo proyección de $\alpha x + \beta y$ sobre \mathcal{M} .

El inciso i) muestra que el mapeo proyección abre sumas y saca escalares, es decir, es lineal.

ii) Obsérvese que

$$\begin{aligned} \|P_{\mathcal{M}}(x)\|^2 + \|P_{\mathcal{M}^\perp}(x)\|^2 &= \langle \hat{x}, \hat{x} \rangle + \langle x - \hat{x}, x - \hat{x} \rangle \\ &= \langle \hat{x}, \hat{x} \rangle + \langle x, x \rangle - \langle x, \hat{x} \rangle - \langle \hat{x}, x \rangle + \langle \hat{x}, \hat{x} \rangle \\ &= \|x\|^2 - [\langle \hat{x}, x \rangle - \langle \hat{x}, \hat{x} \rangle] - [\langle x, \hat{x} \rangle - \langle \hat{x}, \hat{x} \rangle] \\ &= \|x\|^2 - \langle \hat{x}, x - \hat{x} \rangle - \langle x - \hat{x}, \hat{x} \rangle \end{aligned}$$

y, por la ortogonalidad de $P_{\mathcal{M}}$ y $P_{\mathcal{M}^\perp}$ se concluye

$$\|x\|^2 = \|P_{\mathcal{M}}(x)\|^2 + \|P_{\mathcal{M}^\perp}(x)\|^2$$

El inciso ii) se puede interpretar como una generalización del teorema de Pitágoras para espacios de Hilbert.

iii) Una representación de $x \in \mathcal{H}$ como la que se pide es claramente

$$x = P_{\mathcal{M}}(x) + (I - P_{\mathcal{M}})x$$

para mostrar la unicidad supongamos que existe otra representación de $x = y + z$ tal que $y \in \mathcal{M}$, $z \in \mathcal{M}^\perp$, entonces

$$\begin{aligned} 0 &= y + z - [P_{\mathcal{M}}(x) + (I - P_{\mathcal{M}})x] \\ &= y - P_{\mathcal{M}}(x) + z - (I - P_{\mathcal{M}})x \end{aligned}$$

Tomando el producto interior en cada lado con $y - P_{\mathcal{M}}(x)$ se tiene

$$\begin{aligned} 0 &= \langle 0, y - P_{\mathcal{M}}(x) \rangle && \text{por c) del teorema 3.1} \\ &= \langle y - P_{\mathcal{M}}(x) + z - (I - P_{\mathcal{M}})x, y - P_{\mathcal{M}}(x) \rangle \\ &= \|y - P_{\mathcal{M}}(x)\|^2 && \text{pues } z - (I - P_{\mathcal{M}})x \in \mathcal{M}^{\perp} \end{aligned}$$

por lo tanto $y = P_{\mathcal{M}}(x)$. Análogamente $z = (I - P_{\mathcal{M}})x$.

iv) Por el inciso ii)

$$0 \leq \|P_{\mathcal{M}}(x_n - x)\|^2 \leq \|x_n - x\|^2 \rightarrow 0 \text{ cuando } \|x_n - x\| \rightarrow 0$$

v) $x \in \mathcal{M}$ si y sólo si la única representación $x = y + z$ con $y \in \mathcal{M}$, $z \in \mathcal{M}^{\perp}$ es tal que $y = x$ y $z = 0$, esto es, si y sólo si $P_{\mathcal{M}}(x) = x$.

vi) Análogo a v) con $y = 0$ y $z = x$

vii) Por el inciso iii) se sigue que $x = P_{\mathcal{M}_2}(x) + (I - P_{\mathcal{M}_2})x$, cuyo mapeo proyección a \mathcal{M}_1 es

$$P_{\mathcal{M}_1} = P_{\mathcal{M}_1}P_{\mathcal{M}_2}(x) + P_{\mathcal{M}_1}(I - P_{\mathcal{M}_2})x$$

puesto que $P_{\mathcal{M}_1} = P_{\mathcal{M}_1}P_{\mathcal{M}_2}(x)$ para toda $x \in \mathcal{H}$ si y sólo si $P_{\mathcal{M}_1}(y) = 0$ para toda $y \in \mathcal{M}_2^{\perp}$, es decir, si y sólo si, $\mathcal{M}_2^{\perp} \subseteq \mathcal{M}_1^{\perp}$, si y sólo si $\mathcal{M}_2 \subseteq \mathcal{M}_1$.

■

3.4. Conjuntos Ortonormales

Definición 3.12 (El span cerrado) . Sea $\{x_t, t \in T\}$ subconjunto de un espacio de Hilbert \mathcal{H} , entonces el span cerrado, denotado por $\overline{\text{sp}} \{x_t, t \in T\}$, se define como el subespacio cerrado más pequeño de \mathcal{H} , que contiene a cada elemento $x_t, t \in T$.

■

Observación 3.4 Si $\{x_1, \dots, x_n\}$ es un subconjunto finito de un espacio de Hilbert \mathcal{H} , entonces el $\overline{\text{sp}} \{x_1, \dots, x_n\}$ es el conjunto de todas las combinaciones lineales $y = \alpha_1 x_1 + \dots + \alpha_n x_n$, donde $\alpha_1, \dots, \alpha_n \in \mathbb{C}$.

Definición 3.13 (Conjunto ortonormal) . Un conjunto $\{e_t, t \in T\}$ de elementos de un espacio con producto interior es llamado ortonormal si para cada $s, t \in T$

$$\langle e_s, e_t \rangle = \begin{cases} 1 & \text{si } s = t \\ 0 & \text{si } s \neq t \end{cases}$$

■

Observación 3.5 Obsérvese que de la definición se sigue que los elementos de un conjunto ortonormal tienen norma 1.

Observación 3.6 Si $\{\Phi_n, n \in \mathbb{N} \mid \Phi_n \sim N(0, 1)\}$ es cualquier sucesión de variables aleatorias independientes entonces constituyen un conjunto ortonormal para el espacio $L^2(\Omega, \mathfrak{F}, P)$.¹

Teorema 3.13 Si $\{e_1, \dots, e_k\}$ es un subconjunto ortonormal del espacio de Hilbert \mathcal{H} y $\mathcal{M} = \overline{\text{sp}}\{e_1, \dots, e_k\}$ entonces:

$$\begin{aligned} \text{i)} \quad P_{\mathcal{M}}(x) &= \sum_{i=1}^k \langle x, e_i \rangle e_i \quad \text{para todo } x \in \mathcal{H} & (3.4) \\ \text{ii)} \quad \|P_{\mathcal{M}}(x)\|^2 &= \sum_{i=1}^k |\langle x, e_i \rangle|^2 \quad \text{para todo } x \in \mathcal{H} \\ \text{iii)} \quad \left\| x - \sum_{i=1}^k \langle x, e_i \rangle e_i \right\| &\leq \left\| x - \sum_{i=1}^k c_i e_i \right\| \quad \text{para todo } x \in \mathcal{H} \end{aligned}$$

y para todo $c_1, \dots, c_k \in \mathbb{C}$ (o \mathbb{R} si \mathcal{H} es real)

Demostración:

Puesto que $P_{\mathcal{M}}(x) \in \mathcal{M}$ debe satisfacer las ecuaciones de predicción (y además es el único) para probar **i)** basta con verificar esto, bajo la expresión de $P_{\mathcal{M}}(x)$ según el inciso **i)** Sea $y \in \mathcal{M}$ que, por $y = \sum_{j=1}^k c_j e_j$, de las ecuaciones de predicción se tiene

$$\begin{aligned} \langle x - P_{\mathcal{M}}(x), y \rangle &= 0 \\ \left\langle x - P_{\mathcal{M}}(x), \sum_{j=1}^k c_j e_j \right\rangle &= 0 \end{aligned}$$

¹Veáse la sección 4.1

de las propiedades del producto interior se llega a que la igualdad anterior es equivalente al siguiente sistema de k ecuaciones

$$\langle x - P_{\mathcal{M}}(x), e_j \rangle = 0, \quad j = 1, \dots, k$$

si y sólo si

$$\langle P_{\mathcal{M}}(x), e_j \rangle = \langle x, e_j \rangle, \quad j = 1, \dots, k$$

Si $P_{\mathcal{M}}(x) = \sum_{i=1}^k \langle x, e_i \rangle e_i$ entonces

$$\langle P_{\mathcal{M}}(x), e_j \rangle = \left\langle \sum_{i=1}^k \langle x, e_i \rangle e_i, e_j \right\rangle, \quad j = 1, \dots, k$$

y, por la linealidad del producto interior

$$\langle P_{\mathcal{M}}(x), e_j \rangle = \sum_{i=1}^k [\langle x, e_i \rangle \langle e_i, e_j \rangle], \quad j = 1, \dots, k$$

Obsérvese que la expresión anterior se simplifica bastante pues por la definición de conjunto ortogonal $\langle e_i, e_j \rangle = 0$ si $i \neq j$. Y se llega finalmente a

$$\langle P_{\mathcal{M}}(x), e_j \rangle = \langle x, e_j \rangle, \quad j = 1, \dots, k$$

quedando así probado el inciso **i**).

ii) Por el inciso **i**)

$$\begin{aligned} \|P_{\mathcal{M}}(x)\|^2 &= \left\langle \sum_{i=1}^k \langle x, e_i \rangle e_i, \sum_{j=1}^k \langle x, e_j \rangle e_j \right\rangle \\ &= \sum_{i=1}^k \left\langle \langle x, e_i \rangle e_i, \sum_{j=1}^k \langle x, e_j \rangle e_j \right\rangle \\ &= \sum_{i=1}^k \sum_{j=1}^k \langle \langle x, e_i \rangle e_i, \langle x, e_j \rangle e_j \rangle \\ &= \sum_{i=1}^k \sum_{j=1}^k \left[\langle x, e_i \rangle \overline{\langle x, e_j \rangle} \langle e_i, e_j \rangle \right] \end{aligned}$$

y, de nuevo por la definición de conjunto ortogonal se llega finalmente a

$$\|P_{\mathcal{M}}(x)\|^2 = \sum_{i=1}^k |\langle x, e_i \rangle|^2$$

Finalmente, el inciso **iii**) se sigue del teorema de proyección.

■

Corolario 3.14 (Desigualdad de Bessel) Si $\{e_1, \dots, e_k\}$ es un subconjunto ortonormal del espacio de Hilbert \mathcal{H} y $x \in \mathcal{H}$ entonces

$$\sum_{i=1}^k |\langle x, e_i \rangle|^2 \leq \|x\|^2$$

Demostración:

La desigualdad de Bessel es consecuencia directa del teorema anterior inciso *ii*) y de la propiedad *ii*) de la proposición 3.12.

■

Definición 3.14 (Conjunto ortonormal completo) Sea $\{e_t, t \in T\}$ un subconjunto ortonormal del espacio de Hilbert \mathcal{H} , si $\mathcal{H} = \overline{\text{sp}} \{e_t, t \in T\}$ entonces se dice que $\{e_t, t \in T\}$ es un conjunto ortonormal completo o una base ortonormal para \mathcal{H} .

■

Definición 3.15 (Separabilidad) El espacio de Hilbert \mathcal{H} es separable si $\mathcal{H} = \overline{\text{sp}} \{e_t, t \in T\}$ con $\{e_t, t \in T\}$ un conjunto ortonormal numerable o finito.

■

Teorema 3.15 Sea \mathcal{H} un espacio de Hilbert separable donde $\{e_1, e_2, \dots\}$ es la base ortonormal de \mathcal{H} , entonces

i) el conjunto de todas las combinaciones lineales finitas de $\{e_1, e_2, \dots\}$ es densa en \mathcal{H} , es decir, para toda $x \in \mathcal{H}$ y $\varepsilon > 0$ existe $k \in \mathbb{N}$ y constantes c_1, \dots, c_k tales que

$$\left\| x - \sum_{i=1}^k c_i e_i \right\| < \varepsilon$$

ii) $x = \sum_{i=1}^{\infty} \langle x, e_i \rangle e_i$ para toda $x \in \mathcal{H}$, esto es, $\|x - \sum_{i=1}^n \langle x, e_i \rangle e_i\| \rightarrow 0$ si $n \rightarrow \infty$,

iii) $\|x\|^2 = \sum_{i=1}^{\infty} |\langle x, e_i \rangle|^2$ para toda $x \in \mathcal{H}$,

- iv) (Identidad de Parseval) $\langle x, y \rangle = \sum_{i=1}^{\infty} \langle x, e_i \rangle \langle e_i, y \rangle$ para toda $x, y \in \mathcal{H}$ y,*
- v) $x = 0$ si y sólo si $\langle x, e_i \rangle = 0$ para toda $i = 1, 2, \dots$*

Demostración:

Sea $S = \bigcup_{j=1}^{\infty} \overline{\text{sp}} \{e_1, e_2, \dots, e_j\}$, el conjunto de todas las combinaciones lineales de $\{e_1, e_2, \dots\}$ entonces la cerradura ² \bar{S} de S es un subespacio cerrado³ de \mathcal{H} (i.e. $\bar{S} \subset \mathcal{H}$) que contiene a $\{e_i, i = 1, 2, \dots\}$. Por otro lado, por hipótesis se tiene que $\mathcal{H} = \overline{\text{sp}} \{e_t, t \in T\}$, esto es, \mathcal{H} es el cerrado más pequeño que contiene a $\{e_i, i = 1, 2, \dots\}$, luego $\mathcal{H} \subset \bar{S}$. Y por tanto se concluye que $\bar{S} = \mathcal{H}$.

Entonces, por ser $\mathcal{H} = \bar{S}$ un subespacio cerrado cada $x \in \mathcal{H}$ es un punto de contacto de S , es decir, para todo $\varepsilon > 0$ existe $y \in S$, donde $y = \sum_{i=1}^k c_i e_i$ para alguna $k \in \mathbb{N}$ y constantes c_1, \dots, c_k tal que $\|y - x\| < \varepsilon$.

ii) Por la Desigualdad de Bessel's $\sum_{i=1}^k |\langle x, e_i \rangle|^2 \leq \|x\|^2$ para toda $k \in \mathbb{N}$ y por lo tanto

$$\sum_{i=1}^{\infty} |\langle x, e_i \rangle|^2 \leq \|x\|^2 \quad (3.5a)$$

lo que significa que la serie compleja $\sum_{i=1}^{\infty} \langle x, e_i \rangle e_i$ es absolutamente convergente.

Por el inciso *iii)* del teorema 3.13 y el inciso *i)* que se acaba de probar se concluye que para cada $\varepsilon > 0$ existe $k \in \mathbb{N}$ tal que

$$\left\| x - \sum_{i=1}^k \langle x, e_i \rangle e_i \right\| < \varepsilon$$

Ahora, por el inciso *i)* del teorema 3.13 $\sum_{i=1}^n \langle x, e_i \rangle e_i = P_{\mathcal{M}}(x)$ donde $\mathcal{M} =$

²Si S es subconjunto del espacio normado $(\mathcal{H}, \|\cdot\|)$, se define la **cerradura** de S como el conjunto $\bar{S} = \{x \mid x \in \mathcal{H} \text{ y } x \text{ es punto de contacto de } S\}$, donde $x \in \mathcal{H}$ es **punto de contacto** de S si para $\varepsilon > 0$ se tiene que $\{y \in \mathcal{H} \mid \|y - x\| < \varepsilon\} \cap S \neq \emptyset$.

³Para verificar que \bar{S} es un subespacio cerrado de \mathcal{H} , sean $\varepsilon > 0$ y $\{x_n\} \in \bar{S}$ sucesión tal que $\|x_n - x\| \rightarrow 0$, esto implica que para $\varepsilon' = \frac{\varepsilon}{2}$ existe $N \in \mathbb{N}$, tal que si $k \geq N$ se tiene $\|x_k - x\| < \varepsilon'$, y, por ser x_k puntos de contacto de S entonces para ε' existe $y \in S$ tal que $\|y - x_k\| < \varepsilon'$.

Luego por el inciso c) de la proposición 3.6 se tiene que $\|y - x\| \leq \|y - x_k\| + \|x_k - x\| < \varepsilon' + \varepsilon' = \varepsilon$.

Por lo tanto x es punto de contacto para S , es decir, $x \in \bar{S}$, luego \bar{S} es subespacio cerrado de \mathcal{H} .

$\overline{sp} \{e_1, \dots, e_n\}$ y si se toma $k \leq n$, $\sum_{i=1}^n \langle x, e_i \rangle e_i \in \mathcal{M}$, luego

$$\left\| x - \sum_{i=1}^n \langle x, e_i \rangle e_i \right\| < \varepsilon \quad \text{para toda } n \geq k \quad (3.6)$$

iii) Retomando las ideas del inciso anterior y por 3.6, para $n \geq k$ se cumple

$$\|x\|^2 \leq \left\| x - \sum_{i=1}^n \langle x, e_i \rangle e_i \right\|^2 + \left\| \sum_{i=1}^n \langle x, e_i \rangle e_i \right\|^2$$

$$\|x\|^2 < \varepsilon^2 + \sum_{i=1}^n |\langle x, e_i \rangle|^2 \quad \text{por el inciso ii) del teorema 3.13}$$

$$\|x\|^2 < \varepsilon^2 + \sum_{i=1}^{\infty} |\langle x, e_i \rangle|^2$$

Dado que la elección de $\varepsilon > 0$ fue arbitraria se concluye que

$$\|x\|^2 \leq \sum_{i=1}^{\infty} |\langle x, e_i \rangle|^2$$

Finalmente por 3.5a se obtiene

$$\|x\|^2 = \sum_{i=1}^{\infty} |\langle x, e_i \rangle|^2$$

iv) Por el inciso ii) de este teorema se tiene que $\|x - \sum_{i=1}^n \langle x, e_i \rangle e_i\| \rightarrow 0$ si $n \rightarrow \infty$, para cada $x \in \mathcal{H}$. Usando la continuidad del producto interior para

$x, y \in \mathcal{H},$

$$\begin{aligned}
\langle x, y \rangle &= \lim_{n \rightarrow \infty} \left\langle \sum_{i=1}^n \langle x, e_i \rangle e_i, \sum_{j=1}^n \langle y, e_j \rangle e_j \right\rangle \\
&= \lim_{n \rightarrow \infty} \sum_{i=1}^n \left\langle \langle x, e_i \rangle e_i, \sum_{j=1}^n \langle y, e_j \rangle e_j \right\rangle \\
&= \lim_{n \rightarrow \infty} \sum_{i=1}^n \sum_{j=1}^n \langle \langle x, e_i \rangle e_i, \langle y, e_j \rangle e_j \rangle \\
&= \lim_{n \rightarrow \infty} \sum_{i=1}^n \sum_{j=1}^n \left[\langle x, e_i \rangle \overline{\langle y, e_j \rangle} \langle e_i, e_j \rangle \right] \\
&= \lim_{n \rightarrow \infty} \sum_{i=1}^n \langle x, e_i \rangle \langle e_i, y \rangle \\
&= \sum_{i=1}^{\infty} \langle x, e_i \rangle \langle e_i, y \rangle
\end{aligned}$$

v) *Este resultado es consecuencia inmediata del inciso ii) de este teorema.*

■

Capítulo 4

Predicción de Procesos Estocásticos Estacionarios

Es en este capítulo donde se analiza y resuelve el problema central, *la predicción* de los valores del proceso estocástico estacionario $\{X_t, t \geq n + 1\}$ en términos de $\{X_1, \dots, X_n\}$ de estos últimos se tiene la información de los valores de la realización del proceso, esto es, la serie de tiempo. Para estos fines se pretende utilizar únicamente el pasado para predecir el subsecuente comportamiento futuro del proceso.

Ahora bien, se debe decir algo acerca del proceder para encontrar a los predictores del proceso, así como también sus bondades. De cara a la predicción para la series de tiempo, no se pierda de vista que se esta buscando la expresión para encontrar al mejor predictor de la serie el cual será denotado por \hat{X}_{n+1} , es decir, el predictor para los siguientes momentos de la variable, esto se consigue mediante el empleo de la teoría desarrollada en el capítulo anterior con la ayuda del *teorema de proyección*.

En este análisis, siguiendo la terminología y lo desarrollado en el capítulo (3) y, puesto que se trabaja con procesos estocásticos que están definidos en subespacios \mathcal{M} cerrados de $L^2(\Omega, \mathfrak{F}, P)$, el mejor predictor en el subespacio de X_{n+h} es definido como aquel elemento en \mathcal{M} con mínima distancia en media cuadrática de X_{n+h} .

Es claro que esta definición del mejor predictor no es única, sin embargo, los procesos con segundo momento finito conllevan a una teoría de predicción que resulta bastante simple, elegante y bastante apropiada para su uso en la práctica.

Como se vió a detalle que la proyección $P_{\mathcal{M}(X_1, \dots, X_n)} X_{n+h}$ es la mejor función de X_1, \dots, X_n así como la mejor combinación lineal para predecir a X_{n+h} . Por tales motivos los predictores buscados para los procesos estocásticos estacionarios se encontrarán a partir de este resultado, es decir serán, los mejores predictores lineales.

En este capítulo las referencias bibliograficas son Brockwell y Davis [1] [2] y Karr [3].

4.1. El espacio $L^2(\Omega, \mathfrak{F}, P)$

Considérese el espacio de probabilidad $(\Omega, \mathfrak{F}, P)$ y la colección \mathcal{C} de todas las variables aleatorias X definidas sobre Ω y que satisfagan la siguiente condición,

$$E[X^2] < \infty.$$

es decir, con segundo momento finito.

Con la usual definición de multiplicación por reales y la adición de variables aleatorias, se sigue que \mathcal{C} es un espacio vectorial, donde el elemento neutro o elemento cero, es la variable aleatoria que es idénticamente cero sobre Ω .

En $L^2(\Omega, \mathfrak{F}, P)$, se define el producto interior $\langle \bullet \rangle$ para cualesquiera dos elementos $X, Y \in \mathcal{C}$ como

$$\langle X, Y \rangle = E[XY] \tag{4.1}$$

De las características que tiene el operador esperanza, se verifica que $\langle X, Y \rangle$ satisface todas las propiedades de un producto interior, sin embargo, si se tiene que $\langle X, X \rangle = 0$, es decir si $E[X^2] = 0$, no se puede concluir que $X(\varpi) = 0$ para todo $\varpi \in \Omega$, sino sólomente se puede decir que $P(X(\varpi) = 0) = 1$.

Esta dificultad se puede evitar definiendo una relación de equivalencia en donde se dice que X y Y son equivalentes si $P(X = Y) = 1$. Esta relación de equivalencia particiona a \mathcal{C} en clases de variables aleatorias, de tal forma que, cualesquiera dos variables aleatorias en la misma clase son iguales con probabilidad uno.

El Espacio $L^2(\Omega, \mathfrak{F}, P)$ es la colección de todas esas clases de equivalencia con producto interior definido como en 4.1.

En L^2 , la convergencia en norma de una sucesión $\{X_n\}$ de elementos en L^2 al limite X , significa

$$\|X_n - X\|^2 = E [|X_n - X|^2] \rightarrow 0, \text{ cuando } n \rightarrow \infty.$$

La convergencia en norma de $\{X_n\}$ a X , en L^2 , es conocida como convergencia en media cuadrática.

Por tanto, como se ha visto *el Espacio L^2* $(\Omega, \mathfrak{F}, P)$ es un espacio de Hilbert.

4.2. Error cuadrático medio mínimo para procesos estocásticos estacionarios

Sea $\{X_t, t = 0, \pm 1, \dots\}$ un proceso estocástico estacionario sobre un espacio de probabilidad $(\Omega, \mathfrak{F}, P)$, con media cero y función de autocovarianza $\varphi(\cdot)$, y considere el problema de encontrar la combinación lineal $\widehat{X}_{n+1} = \sum_{j=1}^n \phi_{nj} X_{n+1-j}$ tal que sea la mejor aproximación a X_{n+1} en el sentido que $E \left[\left| X_{n+1} - \sum_{j=1}^n \phi_{nj} X_{n+1-j} \right|^2 \right]$ sea mínimo. Tal problema es fácilmente resuelto con la ayuda del teorema de proyección tomando como espacio de Hilbert $\mathcal{H} = L^2(\Omega, \mathfrak{F}, P)$ y $\mathcal{M} = \left\{ \sum_{j=1}^n \alpha_j X_{n+1-j} : \alpha_1, \dots, \alpha_n \in \mathbb{R} \right\}$. Por la definición de producto interior en $L^2(\Omega, \mathfrak{F}, P)$, y por el teorema de proyección la solución buscada es $\widehat{X}_{n+1} = P_{\mathcal{M}}(X_{n+1})$. Luego, las ecuaciones de predicción son

$$\left\langle X_{n+1} - \sum_{j=1}^n \phi_{nj} X_{n+1-j}, Y \right\rangle = 0 \text{ para toda } Y \in \mathcal{M},$$

las cuales por la linealidad del producto interior, son equivalentes a las n ecuaciones

$$\left\langle X_{n+1} - \sum_{j=1}^n \phi_{nj} X_{n+1-j}, X_k \right\rangle = 0, \quad k = n, n-1, \dots, 1 \quad (4.2)$$

4.3. Las ecuaciones de predicción

Sea $\{X_t\}$ un proceso estocástico estacionario definido en $L^2(\Omega, \mathfrak{F}, P)$ y considérese el subespacio cerrado $\mathcal{H}_n = \overline{\text{sp}} \{X_1, \dots, X_n \mid n \geq 1\}$ se definen los predictores de un paso del proceso, como sigue

Definición 4.1 (Predictores paso uno) Los predictores de paso uno \hat{X}_{n+1} para X_{n+1} del proceso $\{X_t\}$ se definen por la siguiente expresión:

$$\hat{X}_{n+1} = \begin{cases} 0 & \text{si } n = 0 \\ P_{\mathcal{H}_n}(X_{n+1}) & \text{si } n \geq 1 \end{cases}$$

■

Dado que $\hat{X}_{n+1} \in \mathcal{H}_n$, satisface para $n \geq 1$

$$\hat{X}_{n+1} = \phi_{n1}X_n + \dots + \phi_{nn}X_1, \quad (4.3)$$

donde los coeficientes $\phi_{n1} + \dots + \phi_{nn}$ satisfacen las ecuaciones (4.2) y recordando que en L^2 el producto interior es $\langle X, Y \rangle = E[XY]$, y por las linealidad del operador esperanza, esas ecuaciones adoptan la forma,

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

es decir

$$\sum_{i=1}^n \phi_{ni} \varphi(i-j) = \varphi(j), \quad j = 1, \dots, n$$

o de manera equivalente,

$$\Gamma_n \Phi_n = \varphi_n \quad (4.4)$$

donde Γ_n es la matriz de varianzas y covarianzas del proceso, $\Gamma_n = [\varphi(i-j)]_{i,j=1,\dots,n}$, φ_n y Φ_n son los vectores columna de autocovarianzas y de los coeficientes buscados para determinar al predictor \hat{X}_{n+1} ,

$$\varphi_n = \begin{bmatrix} \varphi(1) \\ \cdot \\ \cdot \\ \cdot \\ \varphi(n) \end{bmatrix}, \quad \text{y } \Phi_n = \begin{bmatrix} \phi_{n1} \\ \cdot \\ \cdot \\ \cdot \\ \phi_{nn} \end{bmatrix}$$

El teorema de proyección visto en la sección (3.3) garantiza la existencia de al menos una solución para el sistema de ecuaciones definido en (4.4), dado la expresión de \hat{X}_{n+1} en (4.3), a estas ecuaciones se les denomina *ecuaciones de predicción de un paso*.

Obsérvese que el sistema (4.4) se debe resolver para encontrar Φ_n , y es posible que este sistema presente mas de una solución posible, sin embargo, en este caso, cuando se se sustituyan todas esas soluciones en (4.3) para determinar al predictor, dan como resultado el mismo predictor para \hat{X}_{n+1} , esto por la unicidad que asegura el teorema de proyección.

Existe una única solución de (4.4) si y sólo si la matriz Γ_n es no singular, en cuyo caso la solución es fácilmente determinada por

$$\Phi_n = \Gamma_n^{-1} \varphi_n \quad (4.5)$$

las condiciones que se deben de cumplir para que la matriz Γ_n sea no singular, se especifican en el siguiente resultado

Proposición 4.1 *Si $\varphi(0) > 0$ y $\varphi(h) \rightarrow 0$ cuando $h \rightarrow \infty$ entonces la matriz de autocovarianzas $\Gamma_n = [\varphi(i-j)]_{i,j=1,\dots,n}$ de $(X_1, \dots, X_n)'$ es no singular para cada n .*

■

4.4. Algoritmo recursivos para el cálculo de los predictores lineales

En esta sección se revisaran dos métodos recursivos para determinar y calcular los predictores de paso uno \hat{X}_{n+1} , $n \geq 1$, definido en (4.3), y se mostrará como pueden ser usados para calcular los predictores de paso h . La predicción recursiva es de suma importancia en la práctica debido a que calcular a \hat{X}_{n+1} , $n \geq 1$, de manera directa de la definición implica para cada valor de n , resolver grandes sistemas de ecuaciones lineales. Además, cada vez que las observaciones de la serie aumentan el mismo procedimiento debe ser repetido.

Por tales motivos, resulta necesario disponer de algún mecanismo que permita el cálculo de los predictores de paso uno sin tener que realizar inversiones de matrices como dice la definición, mas aún el algoritmo recursivo que aqui se presenta utiliza las n observaciones con las que cuenta para encontrar el predictor basado en $n + 1$ observaciones, esto para cada $n = 1, 2, \dots$

4.4.1. El algoritmo Durbin-Levinson

Dado que $\hat{X}_{n+1} = P_{\mathcal{H}_n}(X_{n+1}) \in \mathcal{H}_n$, es posible expresar a \hat{X}_{n+1} de la siguiente forma,

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

con error cuadrático medio de predicción v_n ,

$$v_n = E \left[\left(X_{n+1} - \hat{X}_{n+1} \right)^2 \right]$$

donde claramente se tiene $v_0 = \varphi(0)$.

El algoritmo especificado en la siguiente proposición, es conocido como *algoritmo de Durbin-Levinson*, es un esquema recursivo para el cálculo de Φ_n .

Proposición 4.2 (Durbin-Levinson) *Sea $\{X_t\}$ un proceso estocástico estacionario de media cero y función de autocovarianza $\varphi(\cdot)$ tal que $\varphi(0) > 0$ y $\varphi(h) \rightarrow 0$ cuando $h \rightarrow \infty$, entonces los coeficientes ϕ_{nj} y los errores cuadráticos medios de predicción v_n satisfacen,*

$$\phi_{11} = \frac{\varphi(1)}{\varphi(0)} \tag{4.6}$$

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

$$\begin{bmatrix} \phi_{n1} \\ \vdots \\ \phi_{n,n-1} \end{bmatrix} = \begin{bmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{bmatrix} - \phi_{nn} \begin{bmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{bmatrix}$$

y

$$\begin{aligned} v_0 &= \varphi(0) \\ v_n &= v_{n-1} [1 - \phi_{nn}^2] \end{aligned} \tag{4.7}$$

Demostración

Considere a los subespacios cerrados \mathcal{H}_1 y \mathcal{H}_2 , definidos por

$$\mathcal{H}_1 = \overline{\text{sp}}\{X_2, \dots, X_n \mid n \geq 1\}$$

y

$$\mathcal{H}_2 = \overline{\text{sp}}\{X_1 - P_{\mathcal{H}_1}(X_1)\}$$

los cuales son subespacios cerrados y ortogonales de $\mathcal{H}_n = \overline{\text{sp}}\{X_1, \dots, X_n \mid n \geq 1\}$, mas aún, para cada $Y \in L^2$ se tiene por lo visto en (3.3) que $P_{\mathcal{H}_n}(Y) = P_{\mathcal{H}_1}(Y) + P_{\mathcal{H}_2}(Y)$, es decir

$$\hat{X}_{n+1} = P_{\mathcal{H}_1}(X_{n+1}) + P_{\mathcal{H}_2}(X_{n+1})$$

luego por (3.4) se sigue la siguiente expresión

$$\hat{X}_{n+1} = P_{\mathcal{H}_1}(X_{n+1}) + a(X_1 - P_{\mathcal{H}_1}(X_1)) \quad (4.8)$$

donde

$$a = \frac{\langle X_{n+1}, X_1 - P_{\mathcal{H}_1}(X_1) \rangle}{\|X_1 - P_{\mathcal{H}_1}(X_1)\|^2} \quad (4.9)$$

Por otro lado, debido por la estacionariedad de $\{X_t\}$ los vectores $(X_n, X_{n-1}, \dots, X_1)$ y (X_2, \dots, X_{n+1}) tienen la misma matriz de varianzas y covarianzas que (X_1, \dots, X_n) , por lo que

$$P_{\mathcal{H}_1}(X_1) = \sum_{j=1}^{n-1} \phi_{n-1,j} X_{j+1} \quad (4.10)$$

y

$$P_{\mathcal{H}_1}(X_{n+1}) = \sum_{j=1}^{n-1} \phi_{n-1,j} X_{n+1-j} \quad (4.11)$$

así

$$\|X_1 - P_{\mathcal{H}_1}(X_1)\|^2 = \|X_{n+1} - P_{\mathcal{H}_1}(X_{n+1})\|^2 = \|X_n - \hat{X}_n\|^2 = v_{n-1} \quad (4.12)$$

de las ecuaciones (4.8), (4.10) y (4.11) se obtiene la siguiente expresión para el predictor de un paso

$$\hat{X}_{n+1} = aX_1 + \sum_{j=1}^{n-1} [\phi_{n-1,j} - a\phi_{n-1,n-j}] X_{n+1-j} \quad (4.13)$$

ahora, de (4.9) y (4.10) se tiene

$$\begin{aligned} a &= \left(\langle X_{n+1}, X_1 \rangle - \sum_{j=1}^{n-1} \phi_{n-1,j} \langle X_{n+1}, X_1 \rangle \right) v_{n-1}^{-1} \\ &= \left[\varphi(h) - \sum_{j=1}^{n-1} \phi_{n-1,j} \varphi(n-j) \right] v_{n-1}^{-1} \end{aligned}$$

Por (4.5) y la proposición (4.1) aunado con el supuesto de que la función de autocovarianza satisface $\varphi(h) \rightarrow 0$ cuando $h \rightarrow \infty$, entonces debe cumplirse que la representación para el predictor \hat{X}_{n+1} dada por

$$\hat{X}_{n+1} = \phi_{n1}X_n + \dots + \phi_{nn}X_1, \quad (4.14)$$

es única, por lo tanto al comparar las expresiones (4.13) y (4.14) se sigue que

$$\phi_{nn} = a$$

y

$$\phi_{n,j} = \phi_{n-1,j} - a\phi_{n-1,n-j}, \quad j = 1, \dots, n-1 \quad (4.15)$$

lo que prueba lo establecido en (4.6).

Solo resta probar lo establecido en (4.7), para tales fines obsérvese que el error cuadrático medio para el predictor \hat{X}_{n+1} es

$$\begin{aligned} v_n &= \left\| X_{n+1} - \hat{X}_{n+1} \right\|^2 \\ &= \left\| X_{n+1} - P_{\mathcal{H}_1}(X_{n+1}) + P_{\mathcal{H}_2}(X_{n+1}) \right\|^2 \\ &= \left\| X_{n+1} - P_{\mathcal{H}_1}(X_{n+1}) \right\|^2 + \left\| P_{\mathcal{H}_2}(X_{n+1}) \right\|^2 - 2 \langle X_{n+1} - P_{\mathcal{H}_1}(X_{n+1}), P_{\mathcal{H}_2}(X_{n+1}) \rangle \end{aligned}$$

de (4.12), de la ortogonalidad de \mathcal{H}_1 y \mathcal{H}_2 , y del hecho que $P_{\mathcal{H}_2}(X_{n+1}) = a(X_1 - P_{\mathcal{H}_1}(X_1))$ se tiene

$$v_n = v_{n-1} + a^2 v_{n-1} - 2a \langle X_{n+1}, X_1 - P_{\mathcal{H}_1}(X_1) \rangle$$

finalmente de (4.9) se obtiene

$$v_n = v_{n-1}(1 - a^2)$$

lo que completa la prueba de la proposición. ■

4.4.2. El algoritmo de innovación

La idea central sobre la que se basa el algoritmo Durbin-Levinson es descomponer el espacio \mathcal{H}_n en dos espacios ortogonales \mathcal{H}_1 y \mathcal{H}_2 , sin embargo, es posible descomponer a \mathcal{H}_n en más de dos espacios ortogonales para obtener resultados similares, esta es la idea sobre la cual se desarrolla el segundo método de recursión predictiva que a continuación se presenta.

En el algoritmo de innovación no se necesita que el proceso $\{X_t\}$ sea estacionario, solo que sea de media cero y función de autocovarianza

$$\varphi(i, j) = \langle X_i, X_j \rangle = E(X_i X_j)$$

Como antes, considerese $\mathcal{H}_n = \overline{\text{sp}} \{X_1, \dots, X_n \mid n \geq 1\}$, el predictor de paso uno \widehat{X}_{n+1} tal como se definió en (4.3), y de nueva cuenta sea $v_n = \|X_{n+1} - \widehat{X}_{n+1}\|$ el error cuadrático medio del predictor \widehat{X}_{n+1} .

Se definen los subespacios \mathcal{H}'_j , $j = 1, \dots, n$ como sigue

$$\mathcal{H}'_j = \overline{\text{sp}} \left\{ X_1 - \widehat{X}_1, \dots, X_j - \widehat{X}_{j+1} \mid j = 1, \dots, n \right\}$$

donde $\widehat{X}_1 := 0$.

Nótese que bajo tales definiciones de los subespacios \mathcal{H}'_j , se sigue que $\mathcal{H}'_n = \mathcal{H}_n$, por lo que en lo sucesivo no se hará distinción entre \mathcal{H}_n y \mathcal{H}'_n . Así,

$$\mathcal{H}_n = \overline{\text{sp}} \left\{ X_1 - \widehat{X}_1, \dots, X_n - \widehat{X}_{n+1} \mid n \geq 1 \right\}$$

y como el conjunto $\left\{ X_1 - \widehat{X}_1, \dots, X_n - \widehat{X}_{n+1} \right\}$ es ortogonal, se tiene que la expresión para el predictor de paso uno es

$$\widehat{X}_{n+1} = \sum_{j=1}^n \lambda_{nj} \left(X_{n+1-j} - \widehat{X}_{n+1-j} \right)$$

el problema ahora a resolver es como calcular los valores de λ_{nj} , $j = 1, \dots, n$ y de v_n , $n = 1, 2, \dots$, esto se establece en algoritmo recursivo que se presenta a continuación.

Proposición 4.3 (*Algoritmo de innovación*) Sea $\{X_t\}$ un proceso estocástico de media cero, con matriz de varianzas y covarianzas $[\varphi(i, j)]_{i,j=1}^n$ no

singular $\forall n \in \mathbb{N}$, entonces el predictor de paso uno \widehat{X}_{n+1} , $n \geq 0$ y sus errores cuadráticos medios v_n , $n \geq 1$, están dados por

$$\widehat{X}_{n+1} = \begin{cases} 0 & \text{si } n = 0 \\ \sum_{j=1}^n \lambda_{nj} (X_{n+1-j} - \widehat{X}_{n+1-j}), & \text{si } n \geq 1 \end{cases} \quad (4.16)$$

y

$$\begin{aligned} v_0 &= \varphi(1, 1) & (4.17) \\ \lambda_{n, n-k} &= v_k^{-1} \left(\varphi(n+1, k+1) - \sum_{j=0}^{n-1} \lambda_{k, k-j} \lambda_{n, n-j} v_j \right), \quad k = 0, \dots, n-1 \\ v_n &= \varphi(n+1, n+1) - \sum_{j=0}^{n-1} \lambda_{n, n-j}^2 v_j \end{aligned}$$

Demostración

Dado que el conjunto $\{X_1 - \widehat{X}_1, \dots, X_n - \widehat{X}_n\}$ es ortogonal y calculando el producto interior en ambos lados de (4.16) con $X_{k+1} - \widehat{X}_{k+1}$, $0 \leq k \leq n$, se tiene

$$\langle \widehat{X}_{n+1}, X_{k+1} - \widehat{X}_{k+1} \rangle = \lambda_{n, n-k} v_k \quad (4.18)$$

Por la ortogonalidad de $(X_{n+1} - \widehat{X}_n)$ con $(X_{k+1} - \widehat{X}_{k+1})$, es posible calcular los coeficientes $\lambda_{n, n-k}$, $k = 0, \dots, n-1$ están dados por

$$\langle (X_{n+1} - \widehat{X}_{n+1}), (X_{k+1} - \widehat{X}_{k+1}) \rangle = \langle X_{n+1}, X_{k+1} - \widehat{X}_{k+1} \rangle - \langle \widehat{X}_{n+1}, X_{k+1} - \widehat{X}_{k+1} \rangle$$

luego por (4.18)

$$\langle X_{n+1}, X_{k+1} - \widehat{X}_{k+1} \rangle = \lambda_{n, n-k} v_k$$

es decir

$$\lambda_{n, n-k} = v_k^{-1} \langle X_{n+1}, X_{k+1} - \widehat{X}_{k+1} \rangle \quad (4.19)$$

Utilizando la representación (4.16) con $n = k$, y por las propiedades del producto interior la expresión (4.19) se convierte en

$$\lambda_{n, n-k} = v_k^{-1} \left[\varphi(n+1, k+1) - \sum_{j=0}^{n-1} \lambda_{k, k-j} \langle X_{n+1}, X_{j+1} - \widehat{X}_{j+1} \rangle \right] \quad (4.20)$$

de nueva cuenta por (4.19) la expresión (4.20) se convierte

$$\lambda_{n,n-k} = v_k^{-1} \left(\varphi(n+1, k+1) - \sum_{j=0}^{n-1} \lambda_{k,k-j} \lambda_{n,n-j} v_j \right)$$

como lo asegura la proposición. Por el teorema de proyección visto en (3.3) y por la proposición (3.12) el cálculo de los errores cuadráticos medios

$$\begin{aligned} v_n &= \left\| X_{n+1} - \widehat{X}_{n+1} \right\|^2 = \|X_{n+1}\|^2 - \left\| \widehat{X}_{n+1} \right\|^2 \\ &= \varphi(n+1, n+1) - \sum_{j=0}^{n-1} \lambda_{n,n-j}^2 v_j \end{aligned}$$

con lo queda demostrada la proposición.

Capítulo 5

Apéndice A: Simulación

A lo largo de la tesis se fue desarrollando herramientas para la aplicación de todos los resultados que se fueron presentando, por lo que aquí se presentan los programas realizados en el programa MATLAB incluyendo la simulación de las series utilizadas, además de que se utilizaron algunos resultados de Ross [7].

El simular las distintas herramientas y conclusiones que se desarrollaron a lo largo de este trabajo, revisten de dinamismo y aportan claridad a las temas expuestos.

Cada programa indica en su cuerpo su objetivo así como fue el proceder para su realización, por lo que sólo se presentan los programas fuentes idénticos al que se usaron para el desarrollo de la tesis.

5.1. Programa 1

```
function media_movil
disp('ESTE PROGRAMA ES PARA CALCULAR LOS VALORES DE UNA MEDIA MOVIL DE
ORDEN H (MMH)')
%(el orden se refiere al numero de integrantes de la media),
disp('.')
disp('PARA ESTE PROGRAMA SE NECESITAN COMO PARAMETROS DE ENTRADA LOS
VALORES DE LOS VECTORES DE TIEMPO t')
disp('.')
disp('Y EL VECTOR QUE CONTENGA LOS VALORES DE LA SERIE')
```

```

disp('EL PROGRAMA REGRESA UNA MATRIZ CON LOS VALORES DE LA SERIE ORIGINAL
LOS VALORES DE LA MEDIA MOVIL ASI COMO LA GRAFICA CORRESPONDIENTE')

```

```

disp(' .')
disp('INTRODUZCA EL VECTOR DE TIEMPO t')
disp(' .')
t=input("");
disp(' .')
disp('INTRODUZCA EL VECTOR CON LOS VALORES DE LA SERIE DE TIEMPO')
disp(' .')
Yt=input("");
Yt=Yt';
t=t';
disp(' .')
disp('INTRODUZCA EL ORDEN DE LA MEDIA MOVIL')
disp(' .')
H=input("");
if length(t)==length(Yt)
    p=fix(H/2);
    if rem(H,2)==1 %se revisa si es media movil centrada (orden impar) o no centrada (orden par)
        MMHt=zeros(1,length(t));
        for i=p+1:length(t)-p %indice para los valores de t
            MMHt(1,i)=sum(Yt(i-p:i+p))/H;
            %for j=i-p:i+p-1
            %MMHt(1,i)=Yt(j)+Yt(j+1);
            %end
            %MMHt(1,i)=MMHt(1,i)/H;
        plot(t(i),MMHt(1,i),'o','color','r','MarkerFaceColor',[.49 1 .63],'MarkerSize',7);
        hold on
    end
    tMM=[t(1,p+1:(length(t)-p))];
    MMHt2=[MMHt(1,p+1:(length(t)-p))];
    plot(t,Yt,'o','color','b','MarkerFaceColor',[.49 1 .63],'MarkerSize',3);
    hold on
    plot(t,Yt,'color','b','LineWidth',2)
    hold on
    plot(tMM,MMHt2,'color','m','LineWidth',2)
    [t',Yt',MMHt']
else

```



```

%tMM=zeros(1,length(t));
%MMHt=zeros(1,length(t));
for i=p:length(t)-p
%for j=i-p+1:i+p-1
% tMM(1,i)=t(j)+t(j+1);
% MMHt(1,i)=Yt(j)+Yt(j+1);
%end
MMHt(1,i)=sum(Yt(i-p+1:i+p))/H;
tMM(1,i)=sum(t(i-p+1:i+p))/H;
plot(tMM(1,i),MMHt(1,i),'o','color','r','MarkerFaceColor',[.49 1 .63],'MarkerSize',7);
hold on
end
tMM=[tMM(1,p:(length(t)-p))];
MMHt=[MMHt(1,p:(length(t)-p))];
plot(t,Yt,'o','color','b','MarkerFaceColor',[.49 1 .63],'MarkerSize',3);
hold on
plot(t,Yt,'color','b','LineWidth',2)
hold on
plot(tMM,MMHt,'color','m','LineWidth',2)
disp(['t Yt'])
[t',Yt']
disp(['t_MM MMHt'])
[tMM',MMHt']
end
else
disp('ERROR, NO CONCUERDAN LAS OBSERVACIONES CON EL VECTOR DE TIEMPO')
end

```

5.2. Programa 2

```

function [T,M,D]=media_dispersion_tesis%(X)
disp(' ESTE PROGRAMA DETERMINA LA ASOCIACION DE LOS COMPONENTES DE LA
SERIE ')
disp(' MEDIANTE EL METODO MEDIA DISPERSION')
disp(' ')
disp(' INGRESE LOS VALORES DE LA SERIE X(t) TECLEANDOS O COPIANDOS DE
UN ARCHIVO DE TEXTO SEPARADOS POR " ; " SIN FORMATO')

```

```

disp(' ')
disp(' Y DEBEN ESTAR ENTRE CORCHETES AL INICIO Y AL FINAL')
disp(' ')
disp(' INGRESE LOS VALORES DE X(t) ')
X=input("");
X=X';
disp(' ')
disp(' INGRESE EL NUMERO DE OBSERVACIONES QUE FORMA A LA ESTACION (EL VAL-
OR DE K) ')
k=input(""); % EL NUMERO DE OBSERVACIONES QUE FORMA A LA ESTACION O EL NU-
MERO DE ELEMENTOS DE CADA CONJUNTO Ti
d=length(X)/k; % SE CALCULA EL NUMERO DE ESTACIONES QUE HAY EN TOTAL
for i=1:k % AQUI COMIENZAN A CALCULARSE LOS Ti, Y LA MANERA DE PROCEDER ES
CONSTRUIR UNA MATRIZ T CUYOS
    T(1,i)=X(i); % RENGLONES SEAN LOS CONJUNTOS Ti, POR LO QUE ESTA MATRIZ TEN-
DRA TANTOS RENGLONES COMO ESTACIONES
end % Y TENDRA K COLUMNAS. HASTA AQUI SE CONSTRUYO EL PRIMER RENGLON DE
LA MATRIZ T O PRIMER CONJUNTO Ti
for i=2:d % COMIENZA LA COSTRUCCION DE LOS DEMAS RENGLONES O ESTACIONES,
MOVIENDO i SE CONSTRUYEN LOS RENGLONES DE T
    for j=(i-1)*k+1:i*k % CON j SE MUEVEN LAS COLUMNAS DE LA MATRIZ
        T(i,j-(i-1)*k)=X(j);
    end
end
M=mean(T'); %SE CALCULA EL VECTOR DE MEDIAS CADA UNA CORRESPONDIENTE A
CADA SUBCONJUNTO Ti
D=std(T'); %SE CALCULA EL VECTOR DE DESVIACIONES ESTANDAR CORRESPONDI-
ENTES A CADA ESTACION Ti
figure
for i=1:length(M)
    plot(M(i),D(i),'o','color','r','MarkerFaceColor',[.49 1 .63],'MarkerSize',7); %AQUI SE GRAFICAN
LOS PUNTOS CORRESPONDIENTES A LOS VALORES DE LA SERIE
    %ENCERRADOS EN CIRCULO
    hold on
end
title('METODO MEDIA-DISPERSION')
xlabel('MEDIAS'); ylabel('DESVIACIONES ESTANDAR');

```

5.3. Programa 3

```
function [MM,MMC,IBVE,IVE,IVEN,t,Yt,Y_descs]=media_movil
disp(' .')
disp('ESTE PROGRAMA ES PARA CALCULAR LOS VALORES DE UNA MEDIA MOVIL DE
ORDEN H (MMH)')
%(el orden se refiere al numero de integrantes de la media).
disp('PARA ESTE PROGRAMA SE NECESITAN COMO PARAMETROS DE ENTRADA LOS
VALORES DE LOS VECTORES DE TIEMPO t')
disp('Y EL VECTOR QUE CONTENGA LOS VALORES DE LA SERIE')
disp('EL PROGRAMA REGRESA UNA MATRIZ CON LOS VALORES DE LA SERIE ORIGINAL
LOS VALORES DE LA MEDIA MOVIL ASI COMO LA GRAFICA CORRESPONDIENTE')
disp(' .')
disp('INTRODUZCA EL VECTOR DE TIEMPO t')
disp(' .')
t=input("");
disp(' .')
disp('INTRODUZCA EL VECTOR CON LOS VALORES DE LA SERIE DE TIEMPO')
disp(' .')
Yt=input("");
Yt=Yt';
t=t';
disp(' .')
disp('INTRODUZCA EL ORDEN DE LA MEDIA MOVIL')
disp(' .')
H=input("");
figure;
if length(t)==length(Yt)
p=fix(H/2);
if rem(H,2)==1 %se revisa si es media movil centrada (orden impar) o no centrada (orden par)
MMHt=zeros(1,length(t));
tMM=t;
for i=p+1:1:length(t)-p %indice para los valores de t
MMHt(1,i)=sum(Yt(i-p:i+p))/H;
figure;
plot(t(i),MMHt(1,i),'o','color','r','MarkerFaceColor',[.49 1 .63],'MarkerSize',7);
hold on
end
```

```

tMM2=[t(1,p+1:(length(t)-p))];
MMHt2=[MMHt(1,p+1:(length(t)-p))];
plot(t,Yt,'o','color','b','MarkerFaceColor',[.49 1 .63],'MarkerSize',3);
hold on
plot(t,Yt,'color','b','LineWidth',2)
hold on
plot(tMM2,MMHt2,'color','m','LineWidth',2)
else
tMM=zeros(1,length(t));
MMHt=zeros(1,length(t));
for i=p:length(t)-p
MMHt(1,i)=sum(Yt(i-p+1:i+p))/H;
tMM(1,i)=sum(t(i-p+1:i+p))/H;
plot(tMM(1,i),MMHt(1,i),'o','color','r','MarkerFaceColor',[.49 1 .63],'MarkerSize',7);
hold on
end
tMM2=[tMM(1,p:(length(t)-p))];
MMHt2=[MMHt(1,p:(length(t)-p))];
plot(t,Yt,'o','color','b','MarkerFaceColor',[.49 1 .63],'MarkerSize',3);
hold on
plot(t,Yt,'color','b','LineWidth',2)
hold on
plot(tMM2,MMHt2,'color','m','LineWidth',2)
end
title('Serie vs Media Movil')
xlabel('Tiempo'); ylabel('Serie');
hold on
t=t';
Yt=Yt';
tMM=tMM';
MMHt=MMHt';
disp(' .')
%disp('MM = [tiempo serie t_mediamovil MediaMovil]');
MM=[tMM2',MMHt2'];
disp('Desca continuar? 1/0')
disp(' .')
res=input("");
if res==1;

```

```

disp('Indique el nuevo orden para la MM')
disp('.')
H2=input("");
MMC=media_movil_recurso(MM,H2);
else MMC=MM;
end
%IBVE=indices_IBVE(MMC,t,Yt);
[IBVE,IVE,IVEN]=indices_IBVE_nuevo_intento(MMC,t,Yt,H);
%cálculo de la serie desestacionalizada
for i=0:H-1
for j=i+1:H:length(t) %j=i*H+1:(i+1)*H
Y_dscs(j)=Yt(j)/IVEN(i+1);
end
end
Y_dscs=Y_dscs';
%length(Y_dscs)
%length(t)
figure;
plot(t,Yt,'o','color','b','MarkerFaceColor',[.49 1 .63],'MarkerSize',3);
hold on
plot(t,Yt,'color','b','LineWidth',2)
hold on
plot(t,Y_dscs,'o','color','m','MarkerFaceColor',[.49 1 .63],'MarkerSize',7);
hold on
plot(t,Y_dscs,'color','m','LineWidth',2)
title('Serie (azul) vs Serie Desestacionalizada (rojo)')
xlabel('Tiempo'); ylabel('Serie');
hold on
else
disp('ERROR, NO CONCUERDAN LAS OBSERVACIONES CON EL VECTOR DE TIEMPO')
end
end

```

5.4. Programa 4

```

function [X,Z]=simulación_MA
disp(' ESTE PROGRAMA SIMULA UNA REALIZACION DE UN PROCESO DE MEDIAS MOVILES-
MA(1)')

```

```

disp(' ')
disp(' CON MODELO  $X(t)=Z(t)+O(1)*Z(t-1)$  DONDE  $Z(t)$  ES  $N(0,VAR(Y))$ ')
disp(' ')
disp(' SE HARA MEDIANTE LA GENERACION DE NUMEROS ALEATORIOS ')
disp(' ')
disp(' INGRESE LA VARIANZA DESEADA PARA LA SIMULACION ')
var=input(' ');
disp(' ')
disp(' ')
E = 0:1:100; % el tiempo a escala de decimales
randn('state',sum(100*clock));
Alea=randn(1,length(E));
figure;1
q=1;
factor=-.8;
Z=zeros(1,length(E));
X=zeros(1,length(E));
X(1,1)=1;
for i=1:length(E)
    if i==1
        Z(1,i)=sqrt(var)*Alea(1,i);
        plot(E(i),Z(1,i),'o','color','r','MarkerFaceColor',[.49 1 .63],'MarkerSize',7);
        hold on
    else
        Z(1,i)=sqrt(var)*Alea(1,i);
        X(1,i)=Z(i)+factor*Z(1,i-1);
        plot(E(i),X(1,i),'o','color','r','MarkerFaceColor',[.49 1 .63],'MarkerSize',7);
        hold on
    end
end
plot(E,X,'color','b','LineWidth',2)
title('Proceso MA(1)')
xlabel('Tiempo'); ylabel('Serie');
hold on
X=X';
Z=Z';

```

5.5. Programa 5

```
function [X,Z]=simulación_AR
disp(' ESTE PROGRAMA SIMULA UNA REALIZACION DE UN PROCESO AUOTOREGRESIVO-
AR(1)')
disp(' .')
disp(' CON MODELO  $X(t)=Z(t)+O(1)*X(t-1)$  DONDE  $Z(t)$  ES  $N(0,VAR(Y))$ ')
disp(' .')
disp(' SE HARA MEDIANTE LA GENERACION DE NUMEROS ALEATORIOS ')
disp(' .')
disp(' INGRESE LA VARIANZA DESEADA PARA LA SIMULACION ')
var=input("");
disp(' .')
disp(' .')
E = 0:1:100;% el tiempo a escala de decimales
randn('state',sum(100*clock));
Alea=randn(1,length(E));
figure:1
q=1;
factor=.9;
Z=zeros(1,length(E));
X=zeros(1,length(E));
X(1,1)=1;
for i=1:length(E)
    if i==1
        Z(1,i)=sqrt(var)*Alea(1,i);
        plot(E(i),Z(1,i),'o','color','r','MarkerFaceColor',[.49 1 .63],'MarkerSize',7);
    hold on
    else
        Z(1,i)=sqrt(var)*Alea(1,i);
        X(1,i)=Z(i)+factor*X(1,i-1);
        plot(E(i),X(1,i),'o','color','r','MarkerFaceColor',[.49 1 .63],'MarkerSize',7);
    hold on
end
end
plot(E,X,'color','b','LineWidth',2)
title('Proceso AR(1)')
xlabel('Tiempo'); ylabel('Serie');
```

```

hold on
X=X';
Z=Z';

```

5.6. Programa 6

```

function serie_simulada=simul_tesis_1
disp(' ESTE PROGRAMA SE CREA CON LA INTENCION DE SIMULAR UNA SERIE DE TIEM-
PO')
disp(' .')
disp(' CON MODELO X(t)=f(t)+ Y(t) Y DADO QUE SE NECESITA QUE Y(t) SEA UNA SUCE-
SION DE VAR ALEA IID N(0,VAR(Y))')
disp(' .')
disp(' SE HARÁ MEDIANTE LA GENERACION DE NUMEROS ALEATORIOS QUE PARA
GARANTIZAR LA INDEPENDENCIA')
disp(' .')
disp(' SE FORZARA AL GENERADOR QUE CAMBIE DE ESTADO DEPENDIENDO DEL TIEM-
PO ')
disp(' .')
disp(' INGRESE LA VARIANZA DESEADA PARA LA SIMULACION ')
var=input("");
disp(' ELLJA EL MODELO DE ASOCIACION PARA LA SERIE')
disp(' 1.- ADITIVO 2.- MULTIPLICATIVO')
mod_asoc=input("");
disp(' ELLJA EL TIPO DE SIMULACION PARA SERIE')
disp(' 1.- CRECIENTE 2.-CONSTANTE 3.-DECRECIENTE')
tipo=input("");
E = 0:1:100;% el tiempo a escala de decimales
randn('state',sum(100*clock));
Alca=randn(length(E));
for i=1:length(E)
Y(i) =sqrt(var) *Alca(i);
if mod_asoc==1
if tipo==1
serie_simulada(i)=5+.02*E(i)+.001*(E(i)^2)+Y(i);% series crecientes
elseif tipo==2
serie_simulada(i)=100+Y(i);%series de media constante

```



```

else
serie_simulada(i)=20-.02*E(i)-.001*(E(i)^2)+Y(i);
end
else
if tipo==1
serie_simulada(i)=exp(5+.02*E(i)+.001*(E(i)^2)+Y(i)); % series crecientes
elseif tipo==2
serie_simulada(i)=exp(100+Y(i)); %series de media constante
else
serie_simulada(i)=exp(20-.02*E(i)-.001*(E(i)^2)+Y(i));
end
end
plot(E(i),serie_simulada(i),'o','color','r','MarkerFaceColor',[.49 1 .63],'MarkerSize',7);
hold on
end
plot(E,serie_simulada,'color','b','LineWidth',2)
title('Serie Simulada')
xlabel('Tiempo'); ylabel('Serie');
hold on
serie_simulada=serie_simulada';

```

5.7. Programa 7

```

function simul=simul_tesis_2
% CON ESTE PROGRAMA SE GENERARAN SERIES CON UN COMPORTAMIENTO
%ESTACIONAL Y POR PERIODOS CRECIENTES Y DECRECIENTES
disp(' ESTE PROGRAMA SIMULA UNA SERIE DE TIEMPO CON UN COMPORTAMIENTO
ESTACIONAL')
disp(' .')
disp(' INGRESE LA VARIANZA DESEADA PARA LA SIMULACION ')
var=input("");
disp(' ELIJA EL MODELO DE ASOCIACION PARA LA SERIE')
disp(' .')
disp(' 1.- ADITIVO 2.- MULTIPLICATIVO')
disp(' .')
mod_asoc=input("");
if mod_asoc==1;

```

```

disp(' ELIJA EL TIPO DE SIMULACION PARA SERIE')
disp(' .')
disp(' 1.- CRECIENTE 2.-CONSTANTE 3.-DECRECIENTE')
tipo=input("");
else
disp(' ELIJA EL TIPO DE SIMULACION PARA SERIE')
disp(' .')
disp(' 1.- CRECIENTE 2.-DECRECIENTE')
disp(' .')
tipo=input("");
end
disp(' INDIQUE EL ORDEN DEL PERIODO ESTACIONAL')
perio_estac=input("");
E = 0:1:99;
%cont=1;
randn('state',sum(100*clock));
Alea=randn(length(E));
for i=1:length(E)
randn('state',sum(100*clock))
Y(i) =sqrt(var) * Alea(i)*;
if mod_asoc==1
if tipo==1
for i=0:fix(perio_estac)-1
for j=i+1:H:length(t) %j=i*H+1:(i+1)*H
Y_dcesos(j)=Yt(j)/IVEN(i+1);
end
end
serie_simulada(i)=5+.2*E(i)+.001*(E(i)^2)+Y(i); % series crecientes

elseif tipo==2
serie_simulada(i)=100+Y(i); %series de media constante
else
serie_simulada(i)=20-.2*E(i)-.001*(E(i)^2)+Y(i);
end
else
if tipo==1
serie_simulada(i)=exp((500+.2*E(i)+.01*(E(i)^2)+Y(i))/100); % series crecientes

```

```

else
serie_simulada(i)=exp((500-.2*E(i)-.01*(E(i)^2)+Y(i))/100);
end
end

if i~=10 & i~=20 & i~=30 & i~=40 & i~=50 & i~=60 & i~=70 & i~=80 & i~=90 & i~=99
serie(i)=100+.002*E(i)+.0001*(E(i)^2)+Y(i);
%serie(i)=(100+.002*E(i))*Y(i);
plot(E(i),serie(i),'o','color','r','MarkerFaceColor',[.49 1 .63],'MarkerSize',7);
hold on
else
serie(i)=(100+.002*E(i))*Y(i)*2.5;
plot(E(i),serie(i),'o','color','r','MarkerFaceColor',[.49 1 .63],'MarkerSize',7);
hold on

end
end
plot(E,serie,'color','b','LineWidth',2)
serie=serie';
t=E';
simul=[t,serie];

```

Comentarios Finales

Este trabajo surgió del gusto hacia los temas de probabilidad y estadística, entre los cuales se encuentra el área denominada series de tiempo, campo relativamente nuevo y que cada vez es más usado en la práctica en diferentes disciplinas, cuya esencia es *la exploración y análisis de los datos de una variable asociada a un fenómeno aleatorio recogidos a través del tiempo, con el objetivo de comprender su comportamiento pasado, presente y futuro, para la toma de decisiones.*

Cabe señalar que este objetivo no es nuevo, ya otras ramas de la probabilidad han perseguido tales fines, así que la cuestión inmediata es ¿qué hace diferente al análisis de las series de tiempo respecto a otras disciplinas?, la respuesta radica en la base sobre la que construye toda la teoría, misma que se puede resumir de la siguiente manera: toda la información de que se dispone y que resulta necesaria para el conocimiento de la variable es su pasado, mediante sus observaciones, así que basta explorar a la serie para conocerla, dicho en otras palabras, se supone que el fenómeno se explica a sí mismo. De esta forma el análisis de las series de tiempo, sostiene que en cada ocurrencia o manifestación, el fenómeno está revelándose ante quien lo estudia, así que el investigador no necesita de nada más que las observaciones para desenmascarar las pautas del comportamiento de la variable.

Es evidente que tal propósito puede resultar demasiado pretencioso, y aparecen interrogantes naturales, ¿es realmente posible alcanzarlo?, ¿bajo que condiciones?, ¿en qué tipo de fenómenos aleatorios esto será posible?, y principalmente: ¿cuál es el sustento teórico y cómo hacerlo?

Es en este contexto donde la tesis se desarrolló, tratando de contribuir en la respuesta de la última pregunta, centrándose principalmente en presentar y desarrollar argumentos teóricos sobre los cuales se construye una visión del análisis de las series de tiempo.

Para alcanzar tal meta se hace uso de la teoría ya desarrollada en la literatura, encontrándose dos corrientes, el enfoque clásico y el estocástico, que a pesar de tener el mismo objetivo, lo pretenden conseguir de manera distinta, y usualmente son catalogados como diferentes, pero realmente la gran diferencia radica en el grado de complejidad que cada uno requiere, siendo más intuitivo el enfoque clásico mientras que el enfoque estocástico

requiere de mayor formalidad matemática.

Ambos enfoques presentan el mismo modelo para las series de tiempo, las suponen compuestas de una tendencia, una componente estacional y una componente aleatoria, pero el enfoque clásico se centra de manera exhaustiva en determinar o detectar el comportamiento de la tendencia y del componente estacional, y acerca del componente aleatorio no imprime mayor énfasis, esto lo consigue mediante el empleo de ciertos métodos simples y argumentación matemática no demasiado compleja, como se vió en el capítulo 1, sin embargo, esta visión es compatible con el enfoque estocástico, donde se usa el mismo modelo para la serie, sólo que se centra principalmente en el comportamiento de la componente aleatoria auxiliándose de los procesos estocásticos. Es por este motivo que este enfoque requiere mayores herramientas matemáticas y probabilísticas. No obstante, tal diferencia o brecha entre los enfoques puede ser fácilmente subsanada realizando los adecuados vínculos y generalizaciones entre las exigencias que ambos requieren, y de esta manera tratarse como un único enfoque, esto se consigue imponiendo desde el inicio al modelo para las series de tiempo las condiciones necesarias para que los resultados que cada corriente sostiene sigan siendo válidos, es este el propósito principal de la tesis y que se desarrolló en los capítulos 1 y 2.

Posteriormente, para completar los objetivos del análisis de las series de cara a la predicción, se presenta el capítulo 3, es ahí donde se estudia lo necesario para realizar pronósticos, aunque a primera vista no lo parezca, esto es debido a que a pesar de que existen varias maneras de realizar predicciones mediante el empleo de procesos estocásticos estacionarios, en este trabajo se auxilia de una teoría muy general para espacios vectoriales de mucha riqueza algebraica y formalidad matemática, los espacios de Hilbert, la cual aplicada en particular en el contexto del espacio $L^2(\Omega, \mathfrak{F}, P)$ permite mediante el empleo de un sólo teorema, el llamado teorema de proyección, poder trazar pronósticos para las series como se revisó en el capítulo 4.

Una característica adicional a la unificación realizada es que permite llevar a cabo un análisis completo y serio para la serie de datos, además de que si se desea puede hacerse a distintos niveles de profundidad sin perderse sustento o bases teóricas del modelo. Esto constituye otro objetivo que se persiguió al realizar este trabajo de investigación.

Por otro lado, debe reconocerse que se puede todavía desarrollar más análisis principalmente en ajuste y determinación de los parámetros de los modelos propuestos, por ejemplo los parámetros de los procesos ARMA a utilizar, sin embargo, esto ya sobrepasa los objetivos del trabajo, pues se

centraron únicamente en la unificación y estudio de los principales resultados de los dos enfoques existentes y que dan lugar al enfoque *clásico-estocástico* para modelar series de tiempo.

Ω

Bibliografía

- [1] Brockwell, Peter J. and Davis, Richard A. *Time Series: Theory and Methods*. 2nd edition. Springer-Verlag New York Inc. 1991.
- [2] Brockwell, Peter J. and Davis, Richard A. *Introduction to Time Series and Forecasting*. Springer-Verlag New York Inc. 1996
- [3] Karr, Alan F. *Probability*. Springer-Verlag. New York, Inc. 1993.
- [4] Kendall, Maurice and Ord, J. Keith. *Time Series*. 3rd edition. Edward Arnold.1990.
- [5] Mood, A.M., Graybill, F.A. and Boes, D.C. *Introduction to the Theory of Statistics*. 3rd edition. New York. McGraw-Hill. 1974.
- [6] Otero, José M.a. *Econometría Series temporales y predicción*. Madrid. AC.1993
- [7] Ross, S. *Simulation*, 2nd. edition. Academic Press, Inc. 1999.
- [8] Rodríguez, Carmen M. *Análisis de Series Temporales*. Madrid. Hespérides. 2000
- [9] Ruiz-Maya,Luis P. y Martín, F. Javier. *Estadística I: Probabilidad*. Madrid AC. 1995
- [10] Ruiz-Maya,Luis P. y Martín, F. Javier. *Estadística II: Inferencia*. Madrid AC. 1995
- [11] Uriel, E. *Análisis de datos: Series Temporales y Análisis Multivariante*. Madrid. AC. 1995

Índice alfabético

- Algoritmo
 - Durbin-Levinson, 87
 - innovación, 90
- Análisis de
 - componente aleatoria, 30
 - componente estacional, 25, 26
 - tendencia, 14
 - estimación de, 15
- Características muestrales, 33
 - autocovarianza muestral, 33
- Características muestrales
 - autocorrelación muestral, 33
 - esperanza muestral, 33
- Causalidad, 51
- Componentes de la serie de tiempo, 7
 - aleatoria, 9
 - estacional, 7
 - tendencia, 7
 - tipos de integración de, 9
 - métodos para los, 14
- Conjunto
 - ortonormal, 76
 - completo, 78
- Convergencia en media cuadrática, 84
- Desigualdad
 - de Bessel, 78
 - de Cauchy-Schwartz, 63
 - del triángulo, 64
- Distancia, 66
- Ecuaciones de Predicción, 73, 84
- Espacio $L^2(\Omega, \mathfrak{G}, P)$, 49, 83
- Espacios
 - de Hilbert, 60, 68
 - separable, 78
 - vectorial
 - complejo, 61
 - vectoriales, 60
- Función
 - de autocorrelación, 33, 40
 - de autocovarianza, 32, 40
 - de ARMA(p,q), 57
 - esperanza, 32
 - varianza, 32
- Identidad de Parseval, 79
- Invertibilidad, 55
- Ley del paralelogramo, 65
- Mapeo proyección, 73
- Medias móviles, 22
- Norma, 63
 - convergencia en, 66
 - espacio euclidiano, 63
- Ortogonal
 - complemento, 69
 - elemento, 69
- Predictores paso uno, 85

- Procesos estocásticos
 - AR(p), 48
 - ARMA(p,q), 46
 - características de, 32
 - de ruido blanco, 43
 - de ruido iid, 43
 - definición de, 3
 - estacionarios, 36
 - débilmente o amplio, 39
 - estricto, 37
 - estado discreto, continuo, 4
 - funciones de distribución, 5
 - lineales, 45
 - MA(q), 44, 47
 - realización de, 4
- Producto interior, 61
 - continuidad del, 66
 - en $L^2(\Omega, \mathfrak{S}, P)$, 83
- Pruebas, 34
 - de autocorrelación en el muestreo, 34
 - de Portmanteau, 35
- Series de tiempo, 1
 - Gaussiana, 40
 - modelo general para, 6
 - modelos para, 2
- Span cerrado, 75
- Subespacio
 - cerrado, 69
- Sucesión
 - de Cauchy, 68
- Teorema
 - de extensión de Kolmogorov, 5
 - de Proyección, 70