Temas de matemáticas

Una introducción a las series de tiempo

Luis Rincón
Verónica De Jesús Romo

Una introducción a las series de tiempo

Luis Rincón
Departamento de Matemáticas
Facultad de Ciencias UNAM
Circuito Exterior de CU
04510 México CDMX

—

Verónica De Jesús Romo
División Académica de Ciencias Básicas
Universidad Juárez Autónoma de Tabasco

La presente obra se financió con recursos del proyecto
PAPIME PE102321 “Estadística y Simulación” de la DGAPA, UNAM.

La presente obra se financió con recursos del proyecto PAPIME PE102321
"Estadística y Simulación" de la DGAPA, UNAM

Una introducción a las series de tiempo
1a edición digital, 2024

D.R. 2024. Universidad Nacional Autónoma de México
Facultad de Ciencias. México 04510, Ciudad de México
editoriales@ciencias.unam.mx
tienda.fciencias.unam.mx

ISBN: 978-607-587-025-0

Diseño de portada: Laura Uribe

Este libro ha sido dictaminado por pares académicos y sometido a aprobación del Comité Editorial de la Facultad de Ciencias de la UNAM.

Prohibida la reproducción total o parcial por cualquier medio, sin autorización escrita del titular de los derechos patrimoniales.

Impreso y hecho en México.

Prólogo

El presente texto contiene algunos elementos básicos para el estudio matemático de las series de tiempo univariadas. Está dirigido a todo estudiante del área científica con conocimientos medios en probabilidad y estadística a nivel licenciatura. En especial, los estudiantes de la carrera de Actuaría pueden obtener provecho de este material, quienes deben estudiar estos temas como parte de su formación académica.

Una serie de tiempo es un registro sistemático a lo largo del tiempo de una variable de interés. Estos registros sucesivos se llevan a cabo dentro de diversas áreas del conocimiento humano, como la ingeniería, la ecología, la economía, las finanzas, la epidemiología, la medicina y la demografía, entre muchas otras. El objetivo es comprender el comportamiento en el tiempo de la variable de interés para proponer modelos y pronósticos, y de esta manera llevar a cabo una mejor toma de decisiones dentro del área de estudio considerada.

El presente trabajo inicia con una introducción a los primeros elementos gráficos, numéricos y de modelación, asociados a una serie de tiempo. Posteriormente, se definen cuatro componentes principales que podrían estar presentes y ser identificados en una serie de tiempo. Estos componentes son: la tendencia, el componente cíclico, la estacionalidad y, finalmente, el elemento errático o aleatorio. Se revisa también el modelo autorregresivo AR, el modelo de promedios móviles MA, los modelos ARMA, ARIMA y SARIMA. Al final del trabajo se proporciona una breve introducción al problema de realizar pronósticos con base en los modelos elaborados. Nuestro tratamiento es tradicional en el sentido de estudiar la evolución de las variables como funciones del tiempo.

Es importante señalar que el texto es breve y está orientado con una perspectiva matemática, de modo que se ha buscado presentar procedimientos, enunciados y demostraciones con cierto grado de justificación matemática donde ha sido posible tal tratamiento sin recurrir a resultados más avanzados. Por otro lado, la mayor parte de la información se encuentra almacenada actualmente en bases de datos u hojas electrónicas, y el análisis

práctico de las series de tiempo se ha agilizado debido a que varios de los procedimientos se encuentran implementados en los programas de cómputo disponibles. Sin embargo, este trabajo no está asociado con ningún programa de cómputo. Aún así, se agregaron algunos códigos en lenguaje `R` que se utilizaron en la realización de algunos ejemplos. Tales códigos se pueden consultar al final del libro. Al lector interesado en acompañar este trabajo con el uso de `R` se le recomienda consultar el libro de Coghlan (2018) [12], el cual contiene una introducción práctica y elemental de la utilización de `R` en el análisis de series de tiempo. Otros textos más completos que se apoyan en el uso de `R` en su presentación son: Cowpertwait y Metcalfe (2009) [14], Cryer y Chan (2008) [15], y Shumway y Stoffer (2017) [33], por mencionar sólo algunos.

Agradecemos a la DGAPA UNAM por el apoyo otorgado a través del proyecto PAPIME PE102321 “Estadística y simulación”, gracias al cual fue posible la edición de este trabajo. Asimismo, agradecemos muy sinceramente a los árbitros de este trabajo por sus valiosos comentarios, así como a los Comités de Publicaciones del Departamento de Matemáticas y de la Facultad de Ciencias por su excelente labor editorial.

Los autores
Noviembre 2023

Contenido

1. Introducción **7**

1.1. Dos modelos elementales 9
1.2. Funciones básicas 14
1.3. Operadores 19
1.4. Estacionariedad 23
1.5. El correlograma 27
1.6. Ejercicios 29

2. Tendencia y estacionalidad **39**

2.1. Gráficas con respecto al tiempo 39
2.2. Transformaciones 45
2.3. Análisis de series con tendencia 54
2.4. Análisis de series con estacionalidad 62
2.5. Interpretación del correlograma 68
2.6. Ejercicios 71

3. Algunos modelos para series de tiempo **77**

3.1. Modelo MA 77
3.2. Modelo AR 91
3.3. Modelo ARMA 111
3.4. Modelo ARIMA 124
3.5. Resumen sobre causalidad e invertibilidad 131
3.6. Función de autocorrelación parcial (PACF) 133
3.7. Selección, ajuste y evaluación de modelos 134
3.8. Estimación de parámetros 136
3.9. Ejercicios 141

4. Construcción de modelos **153**
4.1. Identificación 154
4.2. Estimación 169
4.3. Verificación 185
4.4. Ejemplo con datos reales 190
4.5. Ejercicios 214

5. Algunos resultados sobre pronósticos **217**
5.1. La esperanza condicional como pronóstico 218
5.2. Pronóstico lineal para series estacionarias 226
5.3. Pronóstico lineal sin estacionariedad 233
5.4. Ejercicios 249

Apéndice **255**

A. Códigos en lenguaje R **255**
A.1. Ejemplo Box-Cox 255
A.2. Ejemplos generados por simulación 258
A.3. Ejemplo con datos reales 264

Bibliografía **275**

Índice alfabético **279**

Capítulo 1

Introducción

En este primer capítulo se revisan algunas definiciones y conceptos generales sobre series de tiempo. Una serie de tiempo es una sucesión de observaciones de una variable de interés realizada a lo largo del tiempo. Los tiempos de observación pueden ser cualesquiera tiempos ordenados $0 \leqslant t_1 < t_2 < t_3 < \cdots$ aunque, sin perder generalidad, supondremos que son: $1, 2, 3, \ldots$ De esta manera, tenemos que una serie de tiempo es una sucesión de observaciones numéricas: $x_1, x_2, x_3, \ldots$.

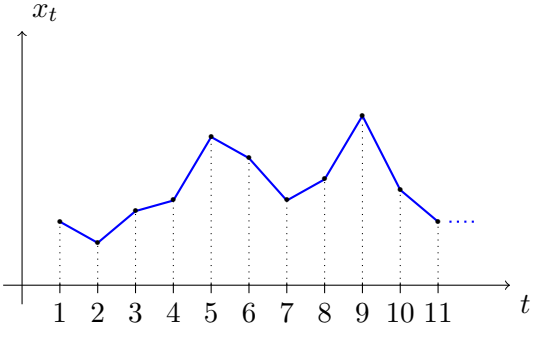

Figura 1.1: Gráfica de una serie de tiempo.

Estos valores pueden ser mediciones de variables de muy diversa índole. Pueden ser mediciones en el tiempo del precio del petróleo, un índice bursátil, la precipitación pluvial, o el nivel de un contaminante atmosférico;

entre otros muchos ejemplos. En la figura 1.1 se muestra una gráfica de una serie de tiempo. El análisis de este tipo de gráficas es de utilidad para tener un mejor entendimiento de algunas características de una serie en estudio.

Representaciones. Para representar al conjunto de observaciones de una serie de tiempo se puede usar la notación $\{x_t : t = 1, 2, \ldots\}$, o bien $\{x_t : t \geqslant 1\}$ o, por brevedad, también escribiremos simplemente $\{x_t\}$, en donde se usan letras minúsculas para indicar que se trata de cantidades numéricas. En la práctica, esta es una sucesión necesariamente finita de observaciones. En ocasiones es conveniente tomar como el primer tiempo de observación a $t = 0$.

Con el fin de buscar representaciones generales para su tratamiento abstracto, se usa el modelo de proceso estocástico a tiempo discreto $\{X_t : t \geqslant 1\}$ para denotar a una serie de tiempo. El uso de la letra mayúscula indica ahora que se trata de una sucesión de variables aleatorias. De esta manera, X_t denota la observación no especificada de la variable de interés al tiempo t. Las observaciones numéricas $x_1, x_2, \ldots$ son una realización particular del proceso estocástico $X_1, X_2, \ldots$ Extendiendo la variable tiempo t para que pueda tomar cualquier valor entero, se llega a la siguiente definición general.

Definición 1.1 *Una serie de tiempo es un proceso estocástico a tiempo discreto* $\{X_t : t \in \mathbb{Z}\}$.

Tanto a la colección finita de números $x_1, \ldots, x_n$ proveniente de observaciones concretas, como al proceso estocástico $\{X_t : t \in \mathbb{Z}\}$ que se utilice como modelo, se le llama serie de tiempo. El análisis de las series de tiempo consiste en el estudio de los modelos matemáticos para representar las sucesiones de observaciones con el objetivo de describir, interpretar, controlar y pronosticar el comportamiento de la variable en estudio. Un aspecto que es necesario resaltar es que las observaciones que aparecen en una serie de tiempo provienen típicamente de variables aleatorias dependientes, es decir, en general existe cierta dependencia entre las observaciones.

Distribuciones finito-dimensionales. Desde el punto de vista matemático,

una especificación completa de una serie de tiempo $\{X_t : t \geqslant 1\}$ se logra cuando se conocen las distribuciones finito-dimensionales del proceso. Recordaremos a continuación la definición de estas distribuciones.

Definición 1.2 *Las distribuciones finito-dimensionales de un proceso estocástico* $\{X_t : t \geqslant 1\}$ *son la colección de todas las funciones de distribución conjuntas:*

$$\begin{aligned} F(x_1) &= P(X_1 \leqslant x_1), \\ F(x_1, x_2) &= P(X_1 \leqslant x_1, X_2 \leqslant x_2), \\ F(x_1, x_2, x_3) &= P(X_1 \leqslant x_1, X_2 \leqslant x_2, X_3 \leqslant x_3), \\ &\vdots \end{aligned}$$

En muy pocas ocasiones es posible conocer explícitamente estas distribuciones conjuntas para una serie de tiempo. En lugar de una especificación completa, el primer y segundo momento de las variables de la serie han sido útiles para caracterizar algunas propiedades sobresalientes del proceso. Los primeros momentos son los valores $E(X_t)$ para $t = 1, 2, \ldots$, y los segundos momentos son las cantidades $E(X_s\, X_t)$ para $s, t = 1, 2, \ldots$ Por ejemplo, cuando se conoce que las distribuciones finito-dimensionales son normales multivariadas, es suficiente conocer las medias (primeros momentos) y las covarianzas (segundos momentos) para determinar completamente todas las distribuciones conjuntas. Más adelante regresaremos al análisis de las medias, varianzas y covarianzas como función de la variable tiempo. Veremos a continuación algunos modelos concretos de series de tiempo.

1.1. Dos modelos elementales

En esta sección definiremos dos modelos elementales de series de tiempo que son importantes desde el punto de vista teórico. Estos modelos son: el proceso aleatorio puro y la caminata aleatoria. A partir de estos modelos básicos se construyen otros modelos más elaborados.

Proceso aleatorio puro

Este es uno de los modelos más simples en donde las variables del proceso son independientes aunque todas tienen la misma distribución de probabilidad.

Definición 1.3 *Un proceso aleatorio puro o proceso de ruido es un proceso estocástico compuesto por variables aleatorias independientes e idénticamente distribuidas. Denotaremos a este proceso por* $\{Z_t : t \in \mathbb{Z}\}$.

Dada la idéntica distribución de las variables Z_t, podemos suponer que su media común es μ y su varianza común es σ^2. Supondremos que estas cantidades son finitas. De esta manera, tenemos que todo proceso aleatorio puro tiene las siguientes características: por lo recién mencionado, tenemos que $E(Z_t) = \mu$ y $\text{Var}(Z_t) = \sigma^2$, para cualquier $t \in \mathbb{Z}$. Considerando los casos $s = t$ y $s \neq t$, se puede comprobar que

$$E(Z_s\, Z_t) = \begin{cases} \sigma^2 + \mu^2 & \text{si} \quad s = t, \\ \mu^2 & \text{si} \quad s \neq t. \end{cases}$$

Usando este resultado y recordando la fórmula $\text{Cov}\,(X, Y) = E(XY) - E(X)E(Y)$, tenemos que la covarianza entre Z_s y Z_t es

$$\text{Cov}\,(Z_s, Z_t) = \begin{cases} \sigma^2 & \text{si} \quad s = t, \\ 0 & \text{si} \quad s \neq t. \end{cases}$$

Finalmente, a partir de la covarianza se puede encontrar que el coeficiente de correlación entre Z_s y Z_t es

$$\rho(Z_s, Z_t) = \begin{cases} 1 & \text{si} \quad s = t, \\ 0 & \text{si} \quad s \neq t. \end{cases}$$

Aquí se ha usado la definición $\rho(X, Y) = \text{Cov}\,(X, Y)/\sqrt{\text{Var}(X)\text{Var}(Y)}$. De esta manera hemos encontrado algunas características numéricas de todo proceso aleatorio puro. Con frecuencia supondremos que $\mu = 0$ como en el siguiente ejemplo.

Ejemplo 1.1 *Para un proceso aleatorio puro* $\{Z_t : t \in \mathbb{Z}\}$*, es frecuente suponer que* $Z_t \sim \mathrm{N}(0, \sigma^2)$*. A este proceso se le llama ruido blanco, y para denotarlo se usa el símbolo* $\mathrm{WN}(0, \sigma^2)$*. Las letras* WN *provienen del término en inglés white noise. Este proceso queda completamente caracterizado por el primer y segundo momentos. Por lo anteriormente mencionado, estas cantidades son:*

- $E(Z_t) = 0.$
- $E(Z_s\, Z_t) = \begin{cases} \sigma^2 & si \quad s = t, \\ 0 & si \quad s \neq t. \end{cases}$

En la figura 1.2 se muestra una parte de una trayectoria de un ruido blanco $\mathrm{WN}(0, 1)$*.*

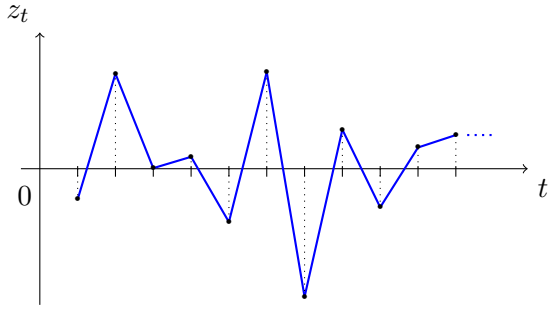

Figura 1.2: Trayectoria de un proceso de ruido blanco $\mathrm{WN}(0, 1)$.

■

Los modelos de series de tiempo que se construyen más adelante tienen un elemento azaroso dado por un proceso aleatorio puro.

Caminata aleatoria

Hemos mencionado que, aunque un proceso de ruido es sencillo en el sentido de que sus variables aleatorias son independientes unas de otras,

resulta de utilidad para construir otras series de tiempo a partir de él. Una caminata aleatoria es un ejemplo de un proceso que se construye a partir de un proceso de ruido de la siguiente forma.

Definición 1.4 *Sea* $\{Z_t : t \in \mathbb{Z}\}$ *un proceso de ruido con media* μ *y varianza* σ^2. *Un proceso estocástico* $\{X_t : t \geqslant 0\}$ *es una caminata aleatoria que inicia en el valor* x_0 *si*

$$\begin{aligned} X_0 &= x_0, \\ X_t &= x_0 + Z_1 + \cdots + Z_t, \quad \textit{para} \;\; t = 1, 2, \ldots \end{aligned} \tag{1.1}$$

Es claro que una caminata aleatoria se puede expresar de la siguiente manera recursiva:

$$X_t = X_{t-1} + Z_t, \quad \text{para} \;\; t = 1, 2, \ldots,$$

junto con la condición inicial $X_0 = x_0$. Escrita de esta forma, la caminata aleatoria se puede ver como un modelo autorregresivo. Estudiaremos este modelo en el capítulo 3. Por otro lado, no es muy complicado comprobar que, para una caminata aleatoria con las características anteriores, se cumplen las fórmulas que aparecen abajo. Se deja como ejercicio para el lector hacer los detalles de las demostraciones de estas fórmulas. En primer lugar, tenemos que, aplicando esperanza y varianza a (1.1),

$$\begin{aligned} E(X_t) &= x_0 + \mu t, \\ \text{Var}(X_t) &= \sigma^2 t. \end{aligned}$$

Desarrollando el producto $X_s X_t$ y aplicando esperanza, se encuentra que

$$E(X_s\, X_t) = \begin{cases} \sigma^2 t + (x_0 + \mu t)^2 & \text{si} \;\; s = t = 0, 1, \ldots \\ \sigma^2 s + (x_0 + \mu s)(x_0 + \mu t) & \text{si} \;\; s < t. \end{cases}$$

A partir de esta fórmula se puede encontrar la covarianza entre X_s y X_t, la cual resulta ser

$$\text{Cov}\,(X_s, X_t) = \begin{cases} \sigma^2 t & \text{si} \;\; s = t = 0, 1, \ldots \\ \sigma^2 s & \text{si} \;\; s < t. \end{cases}$$

Observe que se puede escribir $\text{Cov}\,(X_s, X_t) = \sigma^2\,\text{mín}\{s,t\}$. Finalmente, se puede comprobar que el coeficiente de correlación entre X_s y X_t es

$$\rho(s,t) = \rho\,(X_s, X_t) = \sqrt{s/t} \quad \text{si} \quad 1 \leqslant s \leqslant t.$$

Ejemplo 1.2 *Cuando el punto inicial es $x_0 = 0$ y el proceso de ruido $\{Z_t : t \in \mathbb{Z}\}$ es tal que*

$$\begin{aligned} P(Z_t = -1) &= 1/2, \\ P(Z_t = +1) &= 1/2, \end{aligned}$$

a la serie $\{X_t : t \geqslant 1\}$ dada por (1.1) se le llama caminata aleatoria simétrica simple. Este proceso inicia en el valor cero y salta una unidad, hacia arriba o hacia abajo, con idéntica probabilidad $1/2$ en cada tiempo. Es claro que este proceso tiene como espacio de estados el conjunto $\mathbb{Z}$. Además, $\mu = E(Z_t) = 0$ y $\sigma^2 = \text{Var}(Z_t) = E(Z_t^2) = 1$. En la figura 1.3 se muestra una trayectoria o realización de una caminata aleatoria simétrica simple.

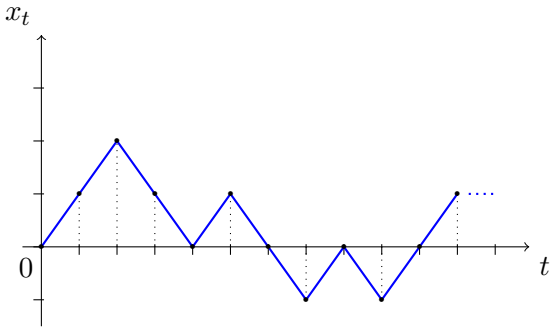

Figura 1.3: Trayectoria de una caminata aleatoria simétrica simple.

■

Ejemplo 1.3 *Cuando el proceso de ruido $\{Z_t : t \in \mathbb{Z}\}$ es $\text{WN}(0, \sigma^2)$, al proceso $\{X_t : t \geqslant 0\}$ dado por (1.1) se le llama caminata aleatoria con ruido blanco. En este caso, la caminata inicia en cualquier valor $x_0 \in \mathbb{R}$ y tiene*

como espacio de estados el conjunto de números reales $\mathbb{R}$. *En la figura 1.4 se muestra una trayectoria de este tipo de caminata aleatoria.*

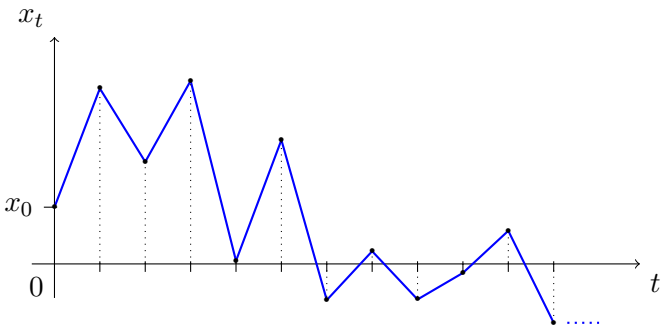

Figura 1.4: Trayectoria de una caminata aleatoria con ruido blanco.

1.2. Funciones básicas

Definiremos ahora algunas funciones que son de utilidad y que aparecen con frecuencia en el análisis de las series de tiempo.

Función media

Definición 1.5 *Sea* $\{X_t : t \in \mathbb{Z}\}$ *una serie de tiempo con esperanza finita. La función media es*

$$t \mapsto \mu(t) = E(X_t), \quad t \in \mathbb{Z}.$$

Es decir, para cada tiempo entero t, podemos calcular la esperanza de la variable aleatoria X_t, suponiendo su existencia, y obtener una función determinista que indica el comportamiento promedio de la serie al paso del tiempo. Cuando sea conveniente, al valor $E(X_t)$ lo denotaremos también como μ_t o $\mu(t)$. Esta cantidad es un número real que puede ser positivo,

negativo o cero. La gráfica de la función media indicará la tendencia de la serie de tiempo.

Ejemplo 1.4 *Para el proceso de ruido* $\{Z_t : t \in \mathbb{Z}\}$ *con media* μ *y varianza* σ^2*, la función media es constante* $E(Z_t) = \mu$*, mientras que para la caminata aleatoria definida en* (1.1)*, su media es la función lineal* $E(X_t) = x_0 + \mu t$*.* ▪

Para una realización $x_1, \ldots, x_n$ de una serie de tiempo, se define la media muestral como el número $\bar{x}$ que aparece a continuación. Note que esta cantidad es constante y depende del número de observaciones.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i. \tag{1.2}$$

Función varianza

Definición 1.6 *Sea* $\{X_t : t \in \mathbb{Z}\}$ *una serie de tiempo con varianza finita. La función varianza es*

$$t \mapsto \text{Var}(X_t), \quad t \in \mathbb{Z}.$$

Denotaremos a esta función por $\sigma^2_X(t)$, la cual es una función determinista que nunca toma un valor negativo. Para una realización $x_1, \ldots, x_n$ de una serie de tiempo se define la varianza muestral como el número no negativo

$$\text{var}(x) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2. \tag{1.3}$$

Nótese que esta varianza muestral difiere de la varianza muestral insesgada cuando se divide la suma entre $n - 1$ y las observaciones provienen de una muestra de variables aleatorias independientes e idénticamente distribuidas.

Función de autocovarianza

Más generalmente, para el caso de variables aleatorias se puede definir la función de covarianza de una serie de tiempo de la siguiente forma.

Definición 1.7 *Sea* $\{X_t : t \in \mathbb{Z}\}$ *una serie de tiempo con segundo momento finito. La función de autocovarianza es*

$$\begin{aligned}(s,t) \mapsto \gamma(s,t) &= \text{Cov}\,(X_s, X_t) \\ &= E[(X_s - \mu_s)(X_t - \mu_t)], \quad s,t \in \mathbb{Z}.\end{aligned}$$

Se llama autocovarianza pues se está calculando la covarianza de cada variable de la serie con las otras variables de la misma serie. También se usan las letras ACVF (*autocovariance function*) para referirse a esta función. Recordemos que la covarianza se puede calcular mediante la fórmula

$$\gamma(s,t) = \text{Cov}\,(X_s, X_t) = E(X_s X_t) - E(X_s)E(X_t).$$

Es conveniente recordar también que este número puede ser positivo, negativo o nulo, y representa una medida de la dependencia lineal entre dos variables aleatorias. Observe que se cumple la propiedad de simetría $\gamma(s,t) = \gamma(t,s)$. Cuando se toma $s = t$, se obtiene la función varianza, es decir,

$$\gamma(t,t) = \text{Var}(X_t) = E(X_t^2) - E^2(X_t).$$

En el caso de tener la realización $x_1, \ldots, x_n$ de una serie de tiempo, se define la autocovarianza muestral como la función

$$\begin{aligned}\tau \mapsto \text{cov}_x(\tau) &= \text{cov}[\ (x_1, \ldots, x_{n-\tau}), (x_{1+\tau}, \ldots, x_n)\ ] \\ &= \frac{1}{n-\tau} \sum_{t=1}^{n-\tau} (x_t - \bar{x})(x_{t+\tau} - \bar{x}),\end{aligned}$$

para $\tau = 0, 1, \ldots, n-1$. Observe que la autocovarianza muestral compara el vector $(x_1, \ldots, x_{n-\tau})$ con el vector $(x_{1+\tau}, \ldots, x_n)$. Los dos vectores tienen la misma longitud y el segundo tiene un desplazamiento de τ unidades adelante. Los posibles valores de τ hacen que esta comparación sea posible. Cuando $\tau = 0$ se recupera la varianza del vector completo, es decir,

$$\text{cov}_x(0) = \text{var}(x) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2.$$

Autocovarianza bajo una hipótesis adicional

En particular, nos interesará calcular $\gamma(t, t+\tau)$, en donde $\tau \in \mathbb{Z}$ es un desfase o diferencia en el valor de los subíndices de las variables. Cuando esta cantidad no depende de t, a la autocovarianza se le denota, más brevemente, por $\gamma(\tau)$. Observe que se usa la misma letra γ para la función, pero ahora con un único argumento. Bajo la hipótesis señalada tenemos entonces la función $\tau \mapsto \gamma(\tau) = \text{Cov}\,(X_t, X_{t+\tau})$.

En particular, $\gamma(0) = \text{Var}(X_t)$. Esto indica que la varianza de la serie de tiempo es constante. Además, se puede comprobar que la autocovarianza es una función par, es decir, $\gamma(\tau) = \gamma(-\tau)$. Por lo tanto, es suficiente especificar los valores de esta función para $\tau \geqslant 0$. Así lo haremos en lo sucesivo: sólo indicaremos los valores de $\gamma(\tau)$, para $\tau \geqslant 0$. Reiteramos que esto es posible sólo en situaciones en las que se cumpla la hipótesis de que $\gamma(t, t+\tau)$ no depende de t.

Función de autocorrelación

A continuación se define la función de autocorrelación para una serie de tiempo.

Definición 1.8 *Sea $\{X_t : t \in \mathbb{Z}\}$ una serie de tiempo con segundo momento finito. La función de autocorrelación es*

$$(s,t) \mapsto \rho(X_s, X_t) = \frac{\text{Cov}\,(X_s, X_t)}{\sqrt{\text{Var}(X_s)\,\text{Var}(X_t)}}, \quad s, t \in \mathbb{Z}.$$

Esta función se llama de autocorrelación, pues representa la correlación de cada variable de la serie con las otras variables de la misma serie. Por brevedad escribiremos $\rho(s,t)$ en lugar de $\rho(X_s, X_t)$. Como la autocovarianza es una función simétrica, también la autocorrelación lo es, esto es, $\rho(s,t) = \rho(t,s)$. Cuando se toma $s = t$ se obtiene la identidad conocida $\rho(t,t) = 1$. Se cumple también la identidad general $-1 \leqslant \rho(s,t) \leqslant 1$.

En el caso de tener una realización $x_1, \ldots, x_n$ de una serie de tiempo, se puede definir una función de autocorrelación muestral denotada por $r(\tau)$. Regresaremos a esta función en la sección 1.5 cuando definamos el correlograma.

Definición 1.9 *La función de autocorrelación muestral de una colección de observaciones $x_1, \ldots, x_n$ es*

$$\tau \mapsto r(\tau) = \frac{\frac{1}{n-\tau}\sum_{t=1}^{n-\tau}(x_t - \bar{x})(x_{t+\tau} - \bar{x})}{\frac{1}{n}\sum_{t=1}^{n}(x_t - \bar{x})^2}, \qquad \tau = 0, 1, \ldots, n-1. \quad (1.4)$$

En analogía con (1.5), se cumple que

$$r(\tau) = \frac{\text{cov}_x(\tau)}{\text{var}(x)} = \frac{\text{cov}_x(\tau)}{\text{cov}_x(0)}, \quad \tau = 0, 1, \ldots, n-1.$$

A esta función se le refiere como función ACF muestral.

Autocorrelación bajo una hipótesis adicional

Como en el caso de la autocovarianza, nos interesará calcular $\rho(t, t+\tau)$, para cada $\tau \in \mathbb{Z}$. En los casos donde esta cantidad no depende de t, a la autocorrelación se le denota, más brevemente, por $\rho(\tau)$. Nuevamente se usa la misma letra, pero ahora con un solo argumento. En esta situación, tenemos la función

$$\tau \mapsto \rho(\tau) = \frac{\text{Cov}\,(X_t, X_{t+\tau})}{\text{Var}(X_t)} = \frac{\gamma(\tau)}{\gamma(0)}. \quad (1.5)$$

A esta función se le denota también por las letras ACF del término en inglés *autocorrelation function*. La expresión (1.5) es la función ACF para un modelo teórico $\{X_t : t \in \mathbb{Z}\}$.

A partir de lo anterior y de las propiedades del coeficiente de correlación, es inmediato comprobar que $\rho(0) = 1$ y que $-1 \leqslant \rho(\tau) \leqslant 1$. Bajo la hipótesis indicada, como la autocovarianza es una función par, también la autocorrelación lo es, es decir, $\rho(\tau) = \rho(-\tau)$. Por lo tanto, será suficiente especificar los valores de esta función para $\tau \geqslant 0$. Así lo haremos en lo sucesivo: sólo indicaremos los valores de $\rho(\tau)$, para $\tau \geqslant 0$. Esto tiene sentido sólo en situaciones en las que se cumpla la hipótesis de que $\rho(t, t+\tau)$ no depende de t.

1.3. Operadores

En esta sección se definen algunos operadores que se aplican sobre los elementos de una serie de tiempo. Estas transformaciones toman una variable de la serie de tiempo y producen como resultado otra variable aleatoria, obteniendo así una nueva serie de tiempo.

Operador retraso

Este es un operador que toma una variable de una serie de tiempo y la lleva a otra variable de la misma serie pero algunas unidades de tiempo atrás.

Definición 1.10 *Sea $\{X_t : t \in \mathbb{Z}\}$ una serie de tiempo y sea $k \in \mathbb{Z}$. El operador retraso de k tiempos es*

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

En particular, $B^1 X_t = BX_t = X_{t-1}$ y $B^0 X_t = X_t$.

La letra B proviene del término en inglés *back*, que en este contexto se puede traducir como el verbo regresar. Esta es una forma conveniente de recordar la operación que se está efectuando. Se debe observar también la conveniencia de considerar que el espacio de tiempo es $\mathbb{Z}$, pues de este modo no hay problema alguno en aplicar el operador retraso a cualquier variable de la serie de tiempo. En particular, observemos que B^0 es el operador identidad y que, cuando k es un entero negativo, X_{t-k} es la

variable aleatoria X_t con $-k > 0$ unidades de tiempo hacia adelante.

A continuación enunciamos algunas propiedades elementales del operador retraso. Para mayor claridad, en algunos casos Se usan paréntesis:

- $\underbrace{B \cdots B}_{k \text{ veces}} X_t = B^k X_t = X_{t-k}, \quad k \geqslant 0.$
- $B^{-k} X_t = X_{t+k}.$
- $B^{-1} B X_t = B^0 X_t = X_t.$
- $B(aX_t + b) = aB(X_t) + b, \quad a, b \text{ constantes}, \quad \text{(linealidad)}.$
- $B(c) = c, \quad c \text{ constante}.$
- $B^n B^m X_t = B^{n+m} X_t = X_{t-n-m}, \quad n, m \in \mathbb{Z}.$
- $(aB^n + bB^m)(X_t + Y_t) = aX_{t-n} + bX_{t-m} + aY_{t-n} + bY_{t-m},$
  en donde a y b son constantes.

Las identidades anteriores no son difíciles de verificar, su demostración se deja como ejercicio. A partir del operador retraso se puede construir el siguiente operador.

Definición 1.11 *Sean $a_0, a_1, \ldots, a_k$ números reales, en donde $k \geqslant 1$ es un entero y $a_k \neq 0$. Un polinomio de retrasos de orden k es un operador de la forma*

$$p(B) = a_0 + a_1 B + a_2 B^2 + \cdots + a_k B^k. \tag{1.6}$$

Por ejemplo, al aplicar el polinomio de retrasos (1.6) a la variable X_t de una serie de tiempo, se obtiene la variable aleatoria

$$\begin{aligned} p(B)X_t &= (a_0 + a_1 B + a_2 B^2 + \cdots + a_k B^k) X_t \\ &= a_0 X_t + a_1 X_{t-1} + a_2 X_{t-2} + \cdots + a_k X_{t-k}. \end{aligned}$$

De esta manera, si $\{X_t : t \in \mathbb{Z}\}$ es una serie de tiempo, entonces la nueva serie de tiempo es $\{p(B)X_t : t \in \mathbb{Z}\}$. Observe que se necesita que la serie original esté definida para todo tiempo $t \in \mathbb{Z}$, de tal modo que se pueda considerar cualquier retraso en todo momento. Para los polinomios de retraso (1.6) se cumplen las reglas usuales de las operaciones algebraicas de suma y producto de polinomios considerando que el operador B es una variable real susceptible de ser elevada a cualquier potencia entera.

Operador diferencia

Un ejemplo de un polinomio de retraso está dado por la expresión $p(B) = 1 - B^k$, para algún $k \geqslant 0$ entero. Esto lleva a definir el siguiente operador.

Definición 1.12 *Sea $\{X_t : t \in \mathbb{Z}\}$ una serie de tiempo y sea $k \geqslant 0$ un entero. Se define el operador*

$$\nabla_k X_t = (1 - B^k)X_t = X_t - X_{t-k}, \quad t \in \mathbb{Z}.$$

El siguiente operador es otro caso particular de un polinomio de retraso que nos será de utilidad.

Definición 1.13 *Sea $\{X_t : t \in \mathbb{Z}\}$ una serie de tiempo y sea $k \geqslant 0$ un entero. El operador diferencia de orden k es*

$$\nabla^k X_t = (1 - B)^k X_t, \quad t \in \mathbb{Z}. \tag{1.7}$$

Al operador ∇^k también se le llama la k-ésima diferencia, y este mismo nombre adquiere la nueva serie así definida,

$$Y_t := \nabla^k X_t, \quad t \in \mathbb{Z}.$$

Es decir, $\{Y_t : t \in \mathbb{Z}\}$ es la k-ésima diferencia de la serie $\{X_t : t \in \mathbb{Z}\}$. Debe tenerse cuidado en no confundir el operador diferencia ∇^k con el operador

antes definido ∇_k. El operador ∇^k, al que se le ha designado como operador diferencia, se usará con mayor frecuencia, y como una regla de nemotecnia el parámetro k colocado como exponente indica que el binomio $(1 - B)$ se eleva a la potencia k.

Para comprender mejor la definición del operador diferencia, a continuación se muestran explícitamente los operadores ∇^0, ∇^1 y ∇^2 como polinomios de retrasos.

- El operador ∇^0 es la diferencia cero y resulta ser la identidad, pues

$$\nabla^0 X_t = (1 - B)^0 X_t = X_t, \quad t \in \mathbb{Z}.$$

- El operador ∇^1 es la primera diferencia y es

$$\nabla^1 X_t = (1 - B)^1 X_t = X_t - BX_t = X_t - X_{t-1}, \quad t \in \mathbb{Z}.$$

- El operador ∇^2 es la segunda diferencia y es

$$\begin{aligned} \nabla^2 X_t &= (1 - B)^2 X_t \\ &= (1 - B)(1 - B)X_t \\ &= (1 - B)(X_t - X_{t-1}) \\ &= X_t - 2X_{t-1} + X_{t-2}, \quad t \in \mathbb{Z}. \end{aligned}$$

En el caso general, se puede desarrollar el término $(1-B)^k$ usando el teorema del binomio y obtener la siguiente expresión para la k-ésima diferencia de una serie de tiempo:

$$\begin{aligned} \nabla^k X_t &= (1 - B)^k X_t \\ &= \sum_{j=0}^{k} \binom{k}{j} (-B)^j X_t \\ &= \sum_{j=0}^{k} \binom{k}{j} (-1)^j X_{t-j} \\ &= \binom{k}{0} X_t - \binom{k}{1} X_{t-1} + \cdots + \binom{k}{k} (-1)^k X_{t-k}, \quad t \in \mathbb{Z}. \end{aligned}$$

Más adelante veremos que la serie diferenciada $Y_t := \nabla^k X_t$ puede adquirir ciertas características que la serie original no posee. En algunos casos también podremos escribir el operador inverso de ∇^k.

En la Tabla 1.1 se muestra un resumen de los operadores definidos. En la sección de ejercicios se pide demostrar algunas propiedades generales de estos operadores.

Nombre	Operador
Retraso de 1 tiempo	$BX_t = X_{t-1}$
Retraso de k tiempos	$B^k X_t = X_{t-k}$
Diferencia de retraso k	$\nabla_k X_t = (1 - B^k)X_t = X_t - X_{t-k}$
Diferencia de orden k	$\nabla^k X_t = (1 - B)^k X_t = \sum_{j=0}^{k} \binom{k}{j} (-1)^j X_{t-j}$

Tabla 1.1: Operadores.

En ocasiones aplicaremos estos operadores a los valores numéricos de una serie de tiempo, así por ejemplo, si x_t es el valor de una serie al tiempo t (o una función de t), cuando sea posible podremos escribir los términos numéricos $Bx_t = x_{t-1}$, $B^k x_t = x_{t-k}$, $\nabla_k x_t = (1 - B^k)x_t = x_t - x_{t-k}$ y $\nabla^k x_t = (1 - B)^k x_t$. Además, cuando estos operadores se aplican a una constante c, se debe entender que c representa una serie o función del tiempo que es constante.

1.4. Estacionariedad

La estacionariedad es una propiedad importante de algunas series de tiempo. Esta característica limita la aleatoriedad en la evolución de la serie en el tiempo, y esto ayuda a que el modelo matemático que se utilice para representarla sea más simple. Algunos cálculos sólo se pueden efectuar bajo la hipótesis de estacionariedad. En esta sección se revisan dos tipos de estacionariedad para series de tiempo.

Estacionariedad estricta

Definición 1.14 *Una serie de tiempo* $\{X_t : t \in \mathbb{Z}\}$ *es estrictamente estacionaria si para cualesquiera números enteros* $k \geqslant 1$ *y* $\tau \in \mathbb{Z}$*, se cumple que*

$$(X_1, \ldots, X_k) \stackrel{d}{=} (X_{1+\tau}, \ldots, X_{k+\tau}). \tag{1.8}$$

La identidad anterior establece que los dos vectores aleatorios indicados tienen la misma distribución de probabilidad. Esta propiedad también se puede denotar de la siguiente forma:

$$(X_1, \ldots, X_k) \sim (X_{1+\tau}, \ldots, X_{k+\tau}).$$

La estacionariedad estricta, llamada también estacionariedad fuerte, significa que al trasladar τ unidades de tiempo, hacia adelante o hacia atrás, los subíndices del vector aleatorio $(X_1, \ldots, X_k)$, la distribución del vector no cambia, es decir, su distribución permanece estacionaria en el tiempo. De esta manera, cuando se dice que una serie es estacionaria, significa que sus distribuciones finito-dimensionales son estacionarias en el tiempo.

Ejemplo 1.5 *Fácilmente se puede comprobar que el proceso de ruido* $\{Z_t : t \in \mathbb{Z}\}$ *definido en la página 10 es estrictamente estacionario. Por otro lado, la caminata aleatoria definida en* (1.1) *no es estrictamente estacionaria.* ▪

En particular, observemos que todas las variables aleatorias de una serie de tiempo estrictamente estacionaria tienen la misma distribución, es decir, $X_1 \sim X_2 \sim X_3 \sim \cdots$ Esto se obtiene de (1.8) tomando $k = 1$ y haciendo variar $\tau \geqslant 0$. También se cumple la idéntica distribución para los vectores aleatorios bidimensionales $(X_1, X_2) \sim (X_2, X_3) \sim (X_3, X_4) \sim \cdots$, y así sucesivamente. Más generalmente, se puede demostrar que la condición (1.8) es equivalente a

$$(X_{t_1}, \ldots, X_{t_k}) \stackrel{d}{=} (X_{t_1+\tau}, \ldots, X_{t_k+\tau}), \tag{1.9}$$

para cualesquiera tiempos $t_1 < \cdots < t_k$ y $\tau \in \mathbb{Z}$. Como se puede ver, el concepto de estacionariedad estricta impone condiciones muy fuertes sobre un proceso estocástico.

Estacionariedad débil

Una condición de estacionariedad menos fuerte que la estacionariedad estricta es la que se expone a continuación.

Definición 1.15 *Una serie de tiempo* $\{X_t : t \in \mathbb{Z}\}$ *es débilmente estacionaria si satisface las siguientes dos condiciones:*

1. *La función media* $\mu(t)$ *es constante, es decir, no depende de* t*.*
2. *La función de autocovarianza* $\gamma(t, t+\tau)$ *no depende de* t*.*

De manera implícita, en la definición anterior estamos suponiendo que toda serie de tiempo débilmente estacionaria posee media y covarianza finitas. A la estacionariedad débil se le llama también estacionariedad de segundo orden. La estacionariedad débil es el tipo de estacionariedad al que nos referiremos con mayor frecuencia, de modo que para hacer el lenguaje escrito más corto, diremos simplemente que una serie es estacionaria para indicar que es estacionaria en el sentido débil. Utilizaremos la expresión completa cuando deseemos hacer énfasis en el sentido débil de la estacionariedad o cuando estemos analizando los dos tipos de estacionariedad al mismo tiempo.

Notación. En vista de la definición 1.15, conviene aquí reiterar la notación que se estableció en la sección 1.2 para la función de autocovarianza de una serie que satisface cierta hipótesis adicional que ahora podemos identificar como consecuencia de la propiedad de estacionariedad: puesto que la función covarianza $\tau \mapsto \gamma(t, t+\tau)$ no depende del parámetro t para una serie de tiempo estacionaria, a esta función se le denota por $\gamma(\tau)$, es decir,

$$\gamma(\tau) := \gamma(t, t+\tau) = \text{Cov}\,(X_t, X_{t+\tau}).$$

Observe que se trata del mismo nombre de la función γ pero se escribe sólo como función de la variable τ de la que depende. En particular, recordemos que $\gamma(0) = \text{Var}(X_t)$ para toda $t \in \mathbb{Z}$. En consecuencia, toda serie de tiempo estacionaria cumple que, tanto la función media $\mu(t) = E(X_t)$, como la función varianza $\text{Var}(X_t)$, son constantes en el tiempo. El siguiente

resultado establece la relación entre los dos conceptos de estacionariedad que hemos definido.

Proposición 1.1 *Suponiendo segundo momento finito, la estacionariedad estricta en una serie de tiempo implica su estacionariedad débil.*

Demostración. Por la estacionariedad estricta, para cualquier $t \in \mathbb{Z}$, $X_t \sim X_{t+\tau}$, en donde $\tau \geqslant 0$ es cualquier entero. Por lo tanto, $E(X_t) = E(X_{t+\tau})$. Es decir, la función media $\mu(t)$ es constante. Esta constante la denotaremos por la letra μ. Esto demuestra la primera parte de lo requerido. Nuevamente, por la estacionariedad estricta, se cumple que $(X_t, X_{t+\tau}) \sim (X_1, X_{1+\tau})$, para cualesquiera $t \in \mathbb{Z}$ y $\tau \geqslant 0$. Por lo tanto,

$$\begin{aligned} \gamma(t, t+\tau) &= \text{Cov}\,(X_t, X_{t+\tau}) \\ &= E(X_t\, X_{t+\tau}) - E(X_t)\, E(X_{t+\tau}) \\ &= E(X_1\, X_{1+\tau}) - \mu^2. \end{aligned}$$

Es decir, la función de autocovarianza $\gamma(t, t+\tau)$ no depende de t. ∎

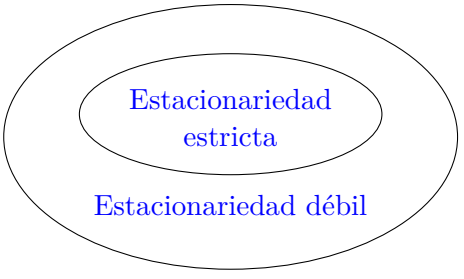

Figura 1.5: Relación entre la estacionariedad estricta y débil.

La relación recién demostrada se puede representar mediante el diagrama que aparece en la figura 1.5. Todo punto dentro de la región más pequeña representa una serie de tiempo estrictamente estacionaria, ese punto pertenece a la región más grande y la serie también es débilmente estacionaria. El recíproco, sin embargo, no se cumple en general, es decir,

la estacionariedad débil no implica la estacionariedad estricta. Véase el ejercicio 24 en la página 36. No obstante, existen condiciones adicionales bajo las cuales la implicación se puede cumplir. Véanse los ejercicios 24 y 25.

1.5. El correlograma

Sea $\{X_t : t \in \mathbb{Z}\}$ una serie de tiempo estacionaria. La función de autocorrelación $\rho(\tau)$ está dada por la expresión (1.5) y está definida para valores de τ en $\mathbb{Z}$. Esta función indica la correlación o grado de dependencia lineal entre las variables aleatorias X_t y $X_{t+\tau}$. Debido a la propiedad de estacionariedad, esta función no depende realmente del valor de t, sólo depende del desfase de tiempo τ en el subíndice de las variables. A los valores τ también se les llama rezagos (o *lags*, por su término en inglés).

- Recordemos que, para series estacionarias, la función de autocorrelación $\rho(\tau)$ se puede escribir en términos de la función de autocovarianza $\gamma(\tau)$ de la siguiente manera

$$\rho(\tau) = \frac{\text{Cov}\,(X_t, X_{t+\tau})}{\text{Var}(X_t)} = \frac{\gamma(\tau)}{\gamma(0)}, \quad \tau \in \mathbb{Z}.$$

- Para el caso de observaciones $x_1, \ldots, x_n$, la función de autocorrelación muestral $r(\tau)$ se define como aparece en (1.4) y se cumple que

$$r(\tau) = \frac{\text{cov}_x(\tau)}{\text{cov}_x(0)} = \frac{\text{cov}_x(\tau)}{\text{var}(x)}, \quad \tau = 0, 1, \ldots, n-1.$$

Definición 1.16

- *Se le llama correlograma a la gráfica de la función de autocorrelación $\rho(\tau)$.*
- *Se le llama correlograma muestral a la gráfica de la función de autocorrelación muestral $r(\tau)$.*

Por brevedad, a las gráficas de $\rho(\tau)$ o de $r(\tau)$ se les llama simplemente correlograma. Recordemos que a ambas funciones se les llama función ACF (*autocorrelation function*), la primera teórica, la segunda muestral. Así que, sin distinguir si se trata de la situación teórica o de la situación muestral, también podemos decir que el correlograma es la gráfica de la función ACF.

Los distintos modelos teóricos para la serie de tiempo $\{X_t : t \in \mathbb{Z}\}$ producirán distintos correlogramas. Como un ejemplo, en la figura 1.6 se muestra un correlograma de un modelo que no se ha especificado. La función $\rho(\tau)$ toma valores en el intervalo $[-1, 1]$ y en la gráfica mostrada se han unido los valores con líneas rectas para una mejor visualización del comportamiento de la función. Observe además que la gráfica puede extenderse hacia la parte negativa del eje horizontal reflejando la trayectoria respecto del eje vertical, pues la correlación es una función par, es decir, $\rho(\tau) = \rho(-\tau)$, $\tau \in \mathbb{Z}$.

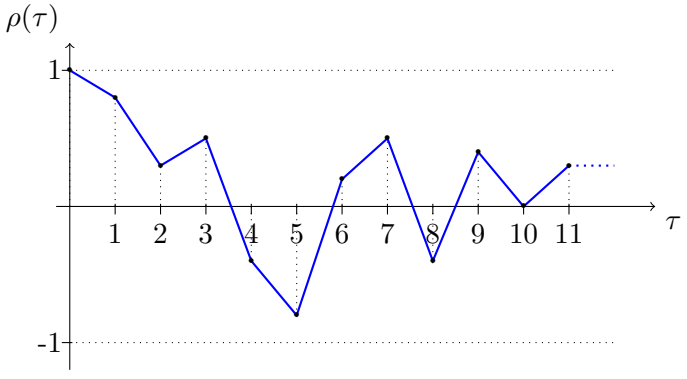

Figura 1.6: Correlograma de una serie de tiempo.

La gráfica de la autocorrelación muestral $\tau \mapsto r(\tau)$ es un elemento importante en el análisis de las series de tiempo. La idea es comparar sus características con las de una función de autocorrelación teórica $\tau \mapsto \rho(\tau)$ para los distintos modelos que se estudian más adelante. Cuando se presente alguna semejanza entre estas dos funciones, tendremos elementos para argumentar que el modelo teórico es un posible candidato para representar

a la serie observada.

En la sección 2.5 del siguiente capítulo se somenta sobre la interpretación del correlograma.

1.6. Ejercicios

Dos modelos elementales

1. Sea $\{Z_t : t \in \mathbb{Z}\}$ un proceso de ruido con media μ y varianza σ^2. Demuestre que:

   a) $E(Z_s\, Z_t) = \begin{cases} \sigma^2 + \mu^2 & \text{si } \ s = t, \\ \mu^2 & \text{si } \ s \neq t. \end{cases}$

   b) $\text{Cov}\,(Z_s, Z_t) = \begin{cases} \sigma^2 & \text{si } \ s = t, \\ 0 & \text{si } \ s \neq t. \end{cases}$

2. Sea $\{X_t : t = 0, 1, \ldots\}$ una caminata aleatoria definida como en (1.1), es decir,

   $$X_t = X_{t-1} + Z_t, \quad t = 1, 2, \ldots$$

   con $X_0 = x_0$ y en donde $\{Z_t : t = 1, 2, \ldots\}$ es un proceso de ruido con media μ y varianza σ^2. Demuestre que:

   a) $E(X_t) = x_0 + \mu t$.

   b) $\text{Var}(X_t) = \sigma^2 t$.

   c) $E(X_s\, X_t) = \begin{cases} \sigma^2 t + (x_0 + \mu t)^2 & \text{si } \ s = t = 0, 1, \ldots \\ \sigma^2 s + (x_0 + \mu s)(x_0 + \mu t) & \text{si } \ s < t. \end{cases}$

   d) $\gamma(s,t) = \text{Cov}\,(X_s, X_t) = \begin{cases} \sigma^2 t & \text{si } \ s = t = 0, 1, \ldots \\ \sigma^2 s & \text{si } \ s < t. \end{cases}$

   e) $\rho(s,t) = \rho\,(X_s, X_t) = \sqrt{s/t} \quad \text{si } \ 1 \leqslant s \leqslant t.$

   Usando la notación $\text{mín}\{s,t\} = s \wedge t$ y $\text{máx}\{s,t\} = s \vee t$, observe que se puede escribir:

   c) $E(X_s\, X_t) = \sigma^2 \cdot (s \wedge t) + [x_0 + \mu \cdot (s \wedge t)][x_0 + \mu \cdot (s \vee t)]$.

d) $\gamma(s,t) = \text{Cov}\,(X_s, X_t) = \sigma^2 \cdot (s \wedge t)$.

e) $\rho(s,t) = \rho(X_s, X_t) = \dfrac{1}{\sqrt{s\,t}} \cdot (s \wedge t), \quad s \cdot t \neq 0.$

Funciones básicas

3. *Función media.*

   Sea $x_1, \ldots, x_n$ una serie de tiempo observada. Demuestre que:

   a) $\displaystyle\sum_{i=1}^{n}(x_i - \bar{x}) = 0.$

   b) $\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}.$

4. *Función media.*

   Sean $x_1, \ldots, x_n$ y $y_1, \ldots, y_n$ dos series de tiempo observadas. Defina la nueva serie $z_i = ax_i + by_i$ para $i = 1, \ldots, n$, en donde a y b son dos constantes. Demuestre que $\bar{z} = a\bar{x} + b\bar{y}$.

5. *Propiedades de la función varianza.*

   Sea $\{X_t : t \in \mathbb{Z}\}$ una serie de tiempo con función media μ_t y con segundo momento finito. Sea c una constante. A partir de la definición $\text{Var}(X_t) := E(X_t - \mu_t)^2$ y usando las propiedades de la esperanza, demuestre las siguientes propiedades de la varianza:

   a) $\text{Var}(X_t) \geqslant 0$.

   b) $\text{Var}(c) = 0$.

   c) $\text{Var}(c\,X_t) = c^2\,\text{Var}(X_t)$.

   d) $\text{Var}(X_t + c) = \text{Var}(X_t)$.

   e) $\text{Var}(X_t) = E(X_t^2) - \mu_t^2$.

   f) En general, $\text{Var}(X_s + X_t) \neq \text{Var}(X_s) + \text{Var}(X_t)$.

6. *Propiedades de la función de autocovarianza.*

   Sea $\{X_t : t \in \mathbb{Z}\}$ una serie de tiempo con función media μ_t y con segundo momento finito. Sea c una constante. A partir de la definición

$$\gamma(s,t) = \text{Cov}(X_s, X_t) := E\left[(X_s - \mu_s)(X_t - \mu_t)\right],$$

y usando las propiedades de la esperanza y la varianza, demuestre las siguientes propiedades de la función de autocovarianza:

a) $\gamma(s,t) = E(X_s X_t) - \mu_s\,\mu_t$.

b) $\gamma(s,t) = \gamma(t,s)$.

c) $\gamma(t,t) = \text{Var}(X_t)$.

d) $\text{Cov}(c, X_t) = 0$.

e) $\text{Cov}(cX_s, X_t) = c\,\gamma(s,t)$.

f) $\text{Cov}(X_s + c, X_t) = \gamma(s,t)$.

g) $\text{Cov}(X_s + X_t, X_u) = \gamma(s,u) + \gamma(t,u)$.

h) Si X_s y X_t son independientes, entonces $\gamma(s,t) = 0$.

i) En general, $\gamma(s,t) = 0 \implies X_s$ y X_t son independientes.

7. Sea $\{X_t : t \in \mathbb{Z}\}$ una serie de tiempo tal que $E(X_1) = -1$, $E(X_2) = 4$ y $E(X_1\,X_2) = 2$. Encuentre:

a) $\text{Cov}(X_1, X_2)$.

b) $\text{Cov}(2X_1, X_2 - 1)$.

c) $\text{Cov}(X_1 + 4, X_2 - 1)$.

d) $\text{Cov}(2X_1 - 3X_2, 6X_1 + 3X_2)$.

8. Sea $\{X_t : t \in \mathbb{Z}\}$ una serie de tiempo con segundo momento finito. Sean $a_1, \ldots, a_n, b_1, \ldots, b_m$ constantes. Demuestre que

$$\text{Cov}(\sum_{i=1}^{n} a_i X_{t_i}, \sum_{j=1}^{m} b_j X_{s_j}) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \text{Cov}(X_{t_i}, X_{s_j}),$$

para cualesquiera tiempos $t_1, \ldots, t_n, s_1, \ldots, s_m \in \mathbb{Z}$, $n, m \geqslant 1$ enteros. Obtenga también el caso particular

$$\text{Var}(\sum_{i=1}^{n} a_i X_{t_i}) = \sum_{i=1}^{n} a_i^2 \text{Var}(X_{t_i}) + 2\sum_{i=2}^{n}\sum_{j=1}^{i-1} a_i a_j \text{Cov}(X_{t_i}, X_{t_j}).$$

9. *Propiedades de la función de autocorrelación.*

Sea $\{X_t : t \in \mathbb{Z}\}$ una serie de tiempo con función media μ_t y con segundo momento finito. Sea c una constante. A partir de la definición

$$\rho(s,t) = \rho(X_s, X_t) := \frac{\operatorname{Cov}(X_s, X_t)}{\sqrt{\operatorname{Var}(X_s)\operatorname{Var}(X_t)}},$$

y usando las propiedades de la esperanza y la covarianza, demuestre las siguientes propiedades de la función de autocorrelación.

a) $\rho(s,t) = \rho(t,s)$.
b) $\rho(t,t) = 1$.
c) $\rho(cX_s, X_t) = \operatorname{signo}(c) \cdot \rho(s,t), \quad c \neq 0$.
d) $\rho(X_s + c, X_t) = \rho(s,t)$.
e) En general, $\rho(X_s + X_t, X_u) \neq \rho(s,u) + \rho(t,u)$.
f) Si X_s y X_t son independientes, entonces $\rho(s,t) = 0$.
g) En general, $\rho(s,t) = 0 \implies X_s$ y X_t son independientes.

10. Sea $\{X_t : t \in \mathbb{Z}\}$ una serie de tiempo tal que $E(X_1) = 0$, $\operatorname{Var}(X_1) = 2$, $E(X_2) = 0$, $\operatorname{Var}(X_2) = 8$ y $\operatorname{Cov}(X_1, X_2) = 2$. Encuentre:

    a) $\rho(X_1, X_2)$.
    b) $\rho(X_1 - 2, 4X_2)$.
    c) $\rho(X_1 + 1, X_2 - 2)$.
    d) $\rho(X_1 - 2X_2, 2X_1 + 3X_2)$.

11. La función de autocorrelación de una caminata aleatoria $\{X_t : t = 0, 1, \ldots\}$, definida como en (1.1), es

$$\rho(s,t) = \rho(X_s, X_t) = \sqrt{s/t} \quad \text{si} \quad 1 \leqslant s \leqslant t.$$

    a) Para $t = 10$ fijo, grafique la función $\rho(s, 10)$ para $s = 1, \ldots, 10$. Interprete la gráfica.
    b) Para $s = 10$ fijo, grafique la función $\rho(10, t)$ para $t = 10, 11, \ldots$. Interprete la gráfica.

12. *Simulación de una caminata aleatoria.*

    Mediante un programa de cómputo obtenga 100 valores al azar $z_1, \ldots, z_{100}$ de la distribución $\mathrm{N}(0,1)$.

a) Asigne un valor a x_0 y calcule los valores sucesivos $x_t = x_{t-1} + z_t$ para $t = 1, \ldots, 100$.

b) Grafique la trayectoria $\{x_t : t = 0, 1, \ldots, 100\}$.

c) Interprete la gráfica en términos de los resultados teóricos $E(X_t) = x_0$ y $\text{Var}(X_t) = t$.

13. *Transformación lineal de las funciones básicas.*

Sea $x_1, \ldots, x_n$ una serie de tiempo observada, y sean a y b dos constantes. Defina la serie $y_i = ax_i + b$ para $i = 1, \ldots, n$. Demuestre que:

a) $\bar{y} = a\,\bar{x} + b$.

b) $\text{var}(y) = a^2 \cdot \text{var}(x)$.

c) $\text{cov}_y(\tau) = a^2 \cdot \text{cov}_x(\tau)$.

d) $r_y(\tau) = r_x(\tau), \quad a \neq 0$.

Operadores

14. *Propiedades del operador retraso B^k.*

Sean $\{X_t : t \in \mathbb{Z}\}$ y $\{X_t : t \in \mathbb{Z}\}$ dos series de tiempo. Sean $k, n, m \in \mathbb{Z}$ y a, b, c constantes reales. Demuestre que:

a) $B^{-k}X_t = X_{t+k}$.

b) $B^{-1}BX_t = X_t$.

c) $B^k(a\,X_t + b) = a\,B^k(X_t) + b$.

d) $B^k(c) = c$.

e) $B^nB^mX_t = B^{n+m}X_t = X_{t-n-m}$.

f) $B^k(X_t + Y_t) = B^kX_t + B^kY_t$.

g) $(a\,B^n + b\,B^m)(X_t) = a\,B^n(X_t) + b\,B^m(X_t)$.

15. *Propiedades del operador ∇_k.*

Sean $\{X_t : t \in \mathbb{Z}\}$ y $\{X_t : t \in \mathbb{Z}\}$ dos series de tiempo. Sea $k \geqslant 0$ un entero y sean a y c dos constantes reales. Demuestre que:

a) $\nabla_0(X_t) = 0$.

b) $\nabla_k(a) = 0$.

c) $\nabla_k(aX_t) = a\,\nabla_k(X_t)$.

d) $\nabla_k(X_t + c) = \nabla_k X_t$.

e) $\nabla_k(X_t + Y_t) = \nabla_k X_t + \nabla_k Y_t$.

16. *Propiedades del operador diferencia* ∇^k.

Sean $\{X_t : t \in \mathbb{Z}\}$ y $\{X_t : t \in \mathbb{Z}\}$ dos series de tiempo. Sea $k \geqslant 0$ un entero, y sean a y c dos constantes reales. Demuestre que:

a) $\nabla^k(a) = \begin{cases} a & \text{si } \ k = 0, \\ 0 & \text{si } \ k = 1, 2, \ldots \end{cases}$

b) $\nabla^k(aX_t) = a\,\nabla^k(X_t)$.

c) $\nabla^k(X_t + c) = \begin{cases} X_t + c & \text{si } \ k = 0, \\ \nabla^k X_t & \text{si } \ k = 1, 2, \ldots \end{cases}$

d) $\nabla^k(X_t + Y_t) = \nabla^k X_t + \nabla^k Y_t$.

17. Sean $n \geqslant 0$ y $m \geqslant 0$ enteros. Diga falso o verdadero.

a) $\nabla^0 = 1$. (Operador identidad)

b) $\nabla_0 = 1$.

c) $\nabla^1 = \nabla_1$.

d) $\nabla^2 = \nabla_2$.

e) $\nabla_n + \nabla_m = \nabla_{n+m}$.

f) $\nabla^n + \nabla^m = \nabla^{n+m}$.

g) $\nabla_n X_{t+n} + \nabla_{-n} X_t = 0$.

h) $\nabla^n = \underbrace{\nabla_1 \cdots \nabla_1}_{n \text{ veces}}, \quad n \geqslant 1.$

18. *Eliminación de la tendencia por diferenciación.*

Sea $\{X_t : t = 1, 2, \ldots\}$ una serie de tiempo de la forma

$$X_t = m_t + Z_t, \quad t \in \mathbb{Z},$$

en donde m_t es una función determinista de t (este es el componente de tendencia) y $\{Z_t : t \in \mathbb{Z}\}$ es un proceso de ruido (este es el componente aleatorio). Demuestre que:

a) Si $m_t = a_0 + a_1 t$, entonces $Y_t := \nabla X_t = a_1 + \nabla Z_t$ para $t \geqslant 2$. Es decir, el operador ∇ elimina la tendencia lineal de la serie. Observe que la serie diferenciada $\{Y_t\}$ empieza a partir de $t = 2$.

b) Si $m_t = a_0 + a_1 t + a_2 t^2$, entonces $Y_t := \nabla^2 X_t = 2a_2 + \nabla^2 Z_t$ para $t \geqslant 3$. Es decir, el operador ∇^2 elimina la tendencia cuadrática de la serie. Observe que la serie diferenciada $\{Y_t\}$ empieza a partir de $t = 3$.

Más generalmente, se puede demostrar que si $m_t = a_0 + a_1 t + \cdots + a_k t^k$, entonces

$$\nabla^k X_t = k!\, a_k + \nabla^k Z_t, \quad \text{para} \quad t \geqslant k.$$

Es decir, el operador ∇^k elimina la tendencia polinomial de grado k de la serie.

Estacionariedad

19. Demuestre que el proceso de ruido $\{Z_t : t \in \mathbb{Z}\}$ definido en la página 10 es estrictamente estacionario.

20. Demuestre que toda serie de tiempo conformada por variables aleatorias independientes y con idéntica distribución es estrictamente estacionaria.

21. Sea $\{X_t : t \in \mathbb{Z}\}$ una serie de tiempo estrictamente estacionaria. Diga falso o verdadero:

    a) Las variables $X_1, X_2, \ldots$ tienen la misma distribución.

    b) Las variables $X_1, X_2, \ldots$ son independientes.

22. Sea $\{X_t :\in \mathbb{Z}\}$ una serie de tiempo estrictamente estacionaria. Demuestre que

$$(X_1, X_3) \sim (X_2, X_4).$$

Un procedimiento similar se puede usar para demostrar que la condición (1.8) que aparece en la definición de serie de tiempo estacionaria es equivalente a la siguiente afirmación aparentemente más general: para cualesquiera tiempos $t_1 < \cdots < t_k$ y $\tau \in \mathbb{Z}$,

$$(X_{t_1}, \ldots, X_{t_k}) \sim (X_{t_1+\tau}, \ldots, X_{t_k+\tau}).$$

23. *Paridad.*

Sea $\{X_t : t \in \mathbb{Z}\}$ una serie de tiempo tal que su función de autocovarianza $\tau \mapsto \gamma(\tau) = \mathrm{Cov}\,(X_t, X_{t+\tau})$ no depende de t. Este es el caso de las series estacionarias. Demuestre que:

a) $\gamma(\tau)$ es una función par, es decir, $\gamma(\tau) = \gamma(-\tau)$, $\tau \in \mathbb{Z}$.

b) $\rho(\tau)$ es una función par, es decir, $\rho(\tau) = \rho(-\tau)$, $\tau \in \mathbb{Z}$.

24. *Estacionariedad débil no implica estacionariedad estricta.*

Sea $\{Z_t : t \in \mathbb{Z}\}$ un proceso de ruido blanco $\mathrm{WN}(0, 1)$, es decir, las variables Z_t son independientes con idéntica distribución $\mathrm{N}(0, 1)$. Para cada $t \in \mathbb{Z}$ defina

$$X_t = \begin{cases} Z_t & \text{si } t \text{ es par,} \\ (Z_{t-1}^2 - 1)/\sqrt{2} & \text{si } t \text{ es impar,} \end{cases}$$

Demuestre que la serie $\{X_t : t \in \mathbb{Z}\}$ es débilmente estacionaria pero no es estacionaria en el sentido estricto.

25. *Caso particular cuando la estacionariedad débil implica la estacionariedad estricta.*

Sea $\{X_t : t = 1, 2, \ldots\}$ una serie de tiempo débilmente estacionaria con distribución gausiana, es decir, el vector aleatorio $(X_1, \ldots, X_k)$ tiene una distribución normal multivariada para cada $k \in \mathbb{N}$. Demuestre que la serie es estacionaria en el sentido estricto.

26. Sea $\{X_t : t \in \mathbb{Z}\}$ una serie de tiempo estacionaria.

a) Demuestre que la función de autocovarianza $\gamma(\tau)$ es una función par, es decir, $\gamma(\tau) = \gamma(-\tau)$, para cualquier $\tau \in \mathbb{Z}$.

b) Sean $\tau_1, \ldots, \tau_n \in \mathbb{Z}$ tiempos distintos. Demuestre que la matriz $(\gamma(\tau_i - \tau_j))$ es positiva semidefinida, es decir, para cualesquiera números reales $a_1, \ldots, a_n$ se cumple

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i\, a_j\, \gamma(\tau_i - \tau_j) \geqslant 0.$$

27. *La caminata aleatoria no es una serie estacionaria.*

Considere la caminata aleatoria $\{X_t : t = 0, 1, \ldots\}$ que inicia en $X_0 = 0$ y que está dada por $X_t = X_{t-1} + Z_t$, para $t = 1, 2, \ldots$, en donde $\{Z_t : t = 1, 2, \ldots\}$ es un proceso estocástico tal que:

a) $E(Z_t) = 0$.

b) $\text{Var}(Z_t) = \sigma^2$.

c) $\text{Cov}\,(Z_s, Z_t) = 0$, para $s \neq t$.

Demuestre que la serie $\{X_t : t = 1, 2, \ldots\}$ no es estacionaria pero la serie de diferencias $\nabla X_t = X_t - X_{t-1}$ es estrictamente estacionaria.

28. *Transformación lineal de una serie estacionaria.*

Sea $\{X_t : t \in \mathbb{Z}\}$ una serie de tiempo estacionaria con función media $\mu_X(t)$ y función de autocovarianza $\gamma_X(\tau)$. Defina la serie $Y_t = aX_t + c$, para $t \in \mathbb{Z}$, en donde $a \neq 0$ y c son dos constantes.

a) Encuentre $\mu_Y(t)$ en términos de $\mu_X(t)$.

b) Encuentre $\text{Cov}\,(Y_t, Y_{t+\tau})$ en términos de $\gamma_X(\tau)$.

c) Compruebe que $\{Y_t : t \in \mathbb{Z}\}$ es estacionaria.

En particular, las series de tiempo $\{-X_t : t \in \mathbb{Z}\}$ y $\{X_t + c : t \in \mathbb{Z}\}$ son estacionarias.

29. *Suma de series estacionarias.*

Sean $\{X_t : t \in \mathbb{Z}\}$ y $\{Y_t : t \in \mathbb{Z}\}$ dos series de tiempo estacionarias. Suponga que las series son independientes, aunque de manera menos restrictiva se puede suponer solamente que las series no están correlacionadas. Demuestre que la serie $\{X_t + Y_t : t \in \mathbb{Z}\}$ es estacionaria.

El correlograma

30. Sea $\{X_t : t \in \mathbb{Z}\}$ una serie de tiempo compuesta por variables aleatorias independientes con varianza finita. Demuestre que la función de autocorrelación es

$$\rho(\tau) = \begin{cases} 1 & \text{si } \tau = 0, \\ 0 & \text{si } \tau \neq 0. \end{cases}$$

31. Sea $x_1, \ldots, x_n$ una serie de tiempo de datos colineales, es decir, $x_t = at + b$, para $t = 1, \ldots, n$, en donde $a \neq 0$ y b son constantes. Demuestre que la función de autocorrelación muestral es

$$r_x(\tau) = 1, \quad \text{para} \quad \tau = 0, 1, \ldots, n-1.$$

32. Considere la siguiente serie de tiempo:

$x_1 = 3.7$	$x_4 = 8.2$	$x_7 = 12$	$x_{10} = 17$
$x_2 = 5$	$x_5 = 9.5$	$x_8 = 14$	$x_{11} = 18$
$x_3 = 6$	$x_6 = 11$	$x_9 = 16$	$x_{12} = 20$

    a) Elabore una gráfica de la serie.

    b) Muestre el correlograma de la serie.

33. Elabore un programa de cómputo que genere de manera independiente 50 valores al azar $x_1, \ldots, x_{50}$ de la distribución unif$(-1, 1)$. Grafique la serie obtenida y muestre el correlograma.

Capítulo 2

Tendencia y estacionalidad

Nuestro objetivo en este capítulo es estudiar algunos elementos que caracterizan el comportamiento de una serie de tiempo observada. Estos elementos ayudan a la identificación y propuesta de los distintos modelos teóricos. Una hipótesis común consiste en que el comportamiento de los datos de una serie de tiempo es resultado de una combinación de los siguientes cuatro componentes:

- La tendencia.
- El componente cíclico.
- El componente estacional.
- El componente irregular.

En las siguientes secciones se definen estos componentes y se explica su relación con el concepto de estacionariedad.

2.1. Gráficas con respecto al tiempo

Con el objetivo de visualizar el comportamiento y características de una serie de tiempo, las gráficas de sus trayectorias se vuelven indispensables. En esta sección se describen los elementos más importantes de estas gráficas. Además, se estudian los principales componentes que las caracterizan. Lo anterior se complementa con algunos ejemplos de cómo se observan las

series de tiempo que son dominadas por cada uno de estos componentes.

Una forma natural de visualizar los datos que se generan periódicamente es graficando las observaciones contra el tiempo. En estas gráficas es común unir las observaciones consecutivas con líneas para visualizar el comportamiento de los datos en forma de una trayectoria. Además, se intentan identificar ciertas características o regularidades que determinan algún patrón o relación entre los datos; tales elementos nos ayudarán a proponer modelos útiles para replicar su comportamiento. Aunque las gráficas no son la única herramienta que se utiliza para proponer un modelo, sí es la mejor forma para identificar visualmente los componentes y elementos que caracterizan a una serie de tiempo. Algunos patrones presentes en el comportamiento de las series de tiempo se enuncian a continuación.

- La tendencia

  Este es el componente de largo plazo que representa el crecimiento o disminución en los valores de la serie cuando se observa globalmente sobre un periodo amplio de tiempo. En las figuras 2.1 y 2.2 se observan algunos ejemplos gráficos de tendencia lineal creciente, lineal decreciente, constante y no lineal. El índice de producción industrial de Estados Unidos en el periodo de 1950 a 2020 que se muestra en la figura 2.3 es un ejemplo de serie de tiempo con componente de tendencia lineal creciente.

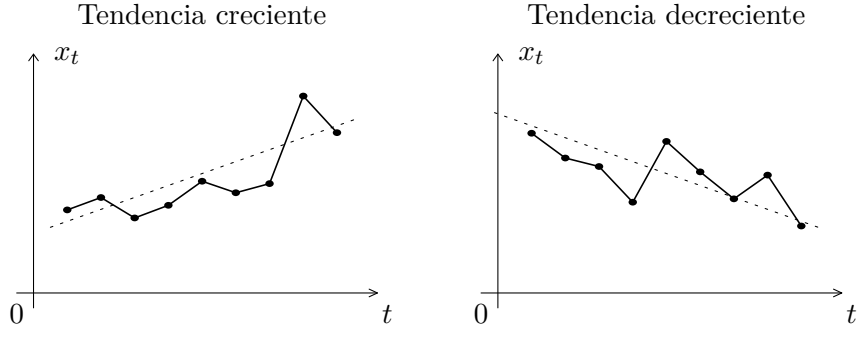

Figura 2.1: Tendencia lineal.

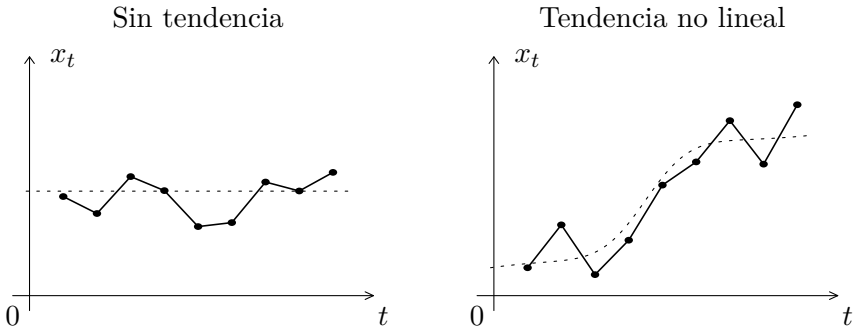

Figura 2.2: Otros tipos de tendencia.

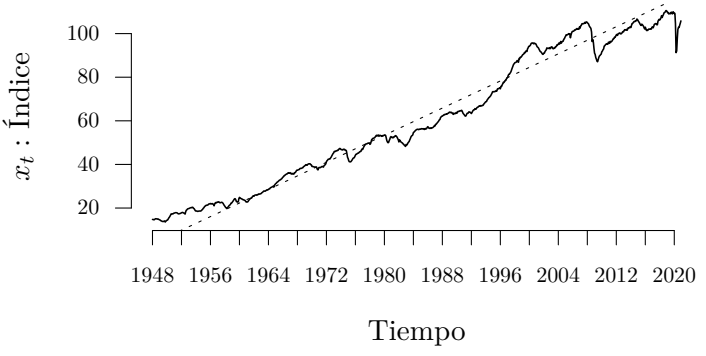

Figura 2.3: Índice mensual de producción industrial de EE.UU. tomando como referencia el año 2012, que equivale a 100. Fuente: Banco de la Reserva Federal de EE.UU. Elaboración propia.

- El componente cíclico

  Este componente puede ser visto como una alteración o desviación en forma de onda alrededor de la tendencia. Se trata de un patrón de cambio que usualmente no se repite con una frecuencia fija. Por ejemplo, ciertas variables económicas y sociales fluctúan con cierto intervalo de tiempo de acuerdo al "ciclo económico", mostrando fases de expansión y contracción. La figura 2.4 muestra un ejemplo

del componente cíclico de una serie de tiempo. Observe que en esta gráfica los dos ciclos mostrados tiene la misma duración, pero, en general, estos ciclos pueden ser de longitud distinta, como los que aparecen en la serie de datos reales de la figura 2.6. Cuando los ciclos abarcan el periodo de un año a este componente se le llama estacional.

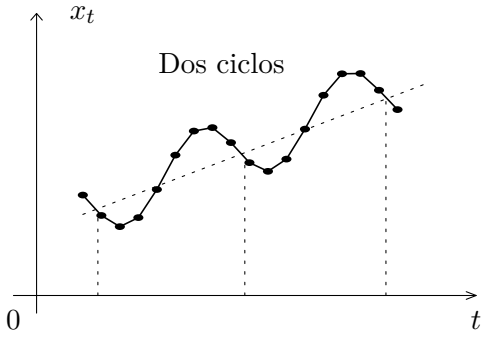

Figura 2.4: Componente cíclico.

- El componente estacional

  Consiste en un patrón de cambio que se caracteriza por ser observado con una frecuencia fija y conocida, generalmente derivado de fenómenos relacionados con un calendario, normalmente, periodos menores a un año. Por ejemplo, estaciones del año, ciclo escolar, días de la semana, etc.

  Como se ha mencionado, este patrón ocurre en intervalos de tiempo fijo a diferencia del componente cíclico, que ocurre en intervalos de tiempo irregulares, normalmente más amplios y que varían de ciclo a ciclo. Por ejemplo, la serie de tiempo de las temperaturas promedio mensuales que se muestra en la figura 2.5 tiene un componente estacional. Por otro lado, la serie de tiempo de la tasa de desempleo mensual que aparece en la figura 2.6 tiene un componente cíclico. Sin embargo, es necesario mencionar que en muchos contextos los términos estacional y cíclico se usan de manera indistinta.

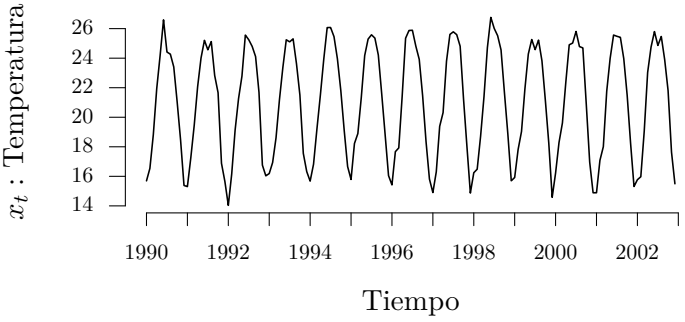

Figura 2.5: Temperatura promedio mensual en México en grados Celsius. Fuente: Portal del clima del Banco Mundial. Elaboración propia.

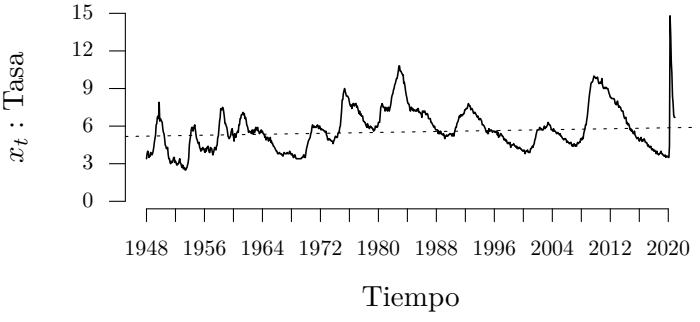

Figura 2.6: Tasa de desempleo mensual de EE.UU. entre 1948 y 2020. Fuente: *Federal Reserve Economic Data.* Elaboración propia.

- El componente irregular o aleatorio

  Esta es la variabilidad de la serie de tiempo después de retirar las otras componentes. Este componente es impredecible y es la consecuencia de múltiples fenómenos no controlables que inciden en el comportamiento de la variable en estudio. La naturaleza azarosa del fenómeno en cuestión permitirá en algunas ocasiones suponer características

específicas relacionadas con la media y la varianza de este componente. Un ejemplo de componente irregular se muestra en la figura 2.7.

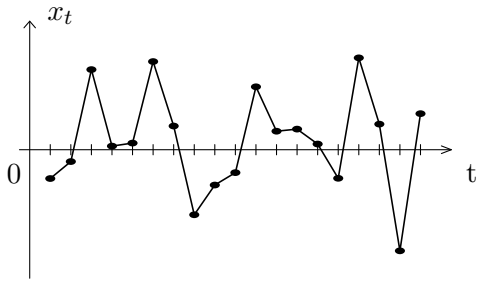

Figura 2.7: Componente irregular o aleatorio.

Integración de componentes

La siguiente suma llamada modelo aditivo representa una manera de combinar los cuatro componentes básicos:

$$X_t = m_t + c_t + s_t + Z_t, \quad t \in \mathbb{Z},$$

en donde m_t, c_t y s_t (en minúsculas) son fluctuaciones deterministas y desconocidas que representan la tendencia, el componente cíclico y el componente estacional, respectivamente. El último término, Z_t (en mayúscula), es un proceso de ruido que representa el componente irregular o aleatorio.

Otra forma de integrar los cuatro componentes es a través de su producto, conocido como modelo multiplicativo:

$$X_t = m_t \cdot c_t \cdot s_t \cdot Z_t, \quad t \in \mathbb{Z}.$$

Sin embargo, cuando sea posible, se puede aplicar la función logaritmo a la igualdad anterior y obtener un modelo aditivo de los logaritmos de los componentes, aunque el proceso de ruido será diferente. También se pueden

considerar modelos en los cuales algunos componentes se sumen y otros se multipliquen, como en el siguiente caso

$$X_t = m_t \cdot s_t + Z_t, \quad t \in \mathbb{Z}.$$

Es importante mencionar que no todas las series presentan los cuatro componentes mencionados. Por ejemplo, algunas series sólo presentan los componentes de tendencia e irregularidad, o estacionalidad e irregularidad. La gráfica de los datos respecto al tiempo nos permite identificar visualmente cuáles son los componentes presentes en los datos de manera preliminar y subjetiva.

2.2. Transformaciones

Supongamos que se tiene un conjunto de datos observados u obtenidos de algún experimento. En ocasiones es necesario que estos datos tengan ciertas características y en algunos casos se pueden aplicar ciertas transformaciones algebraicas a los datos con el objetivo de que adquieran las características deseadas. Por ejemplo, es posible que se requiera una varianza constante, que no exista tendencia, un componente estacional constante, o bien normalidad en la distribución de los datos. El objetivo de esta sección es presentar algunas transformaciones útiles para tales fines.

Transformación de centralización

Por su simplicidad e importancia, mencionaremos primero la así llamada transformación de centralización. Particularmente en las series de tiempo estacionarias, es posible que la media constante $E(X_t) = \mu$ no sea cero. La sencilla transformación dada por $Y_t = X_t - \mu$ produce una serie con media $E(Y_t) = 0$ y varianza $\text{Var}(Y_t) = \text{Var}(X_t) = \sigma^2$. Sin pérdida de generalidad, en la mayoría de los modelos que se exponen del capítulo 3 en adelante, las series de tiempo estacionarias que se consideran tienen como supuesto que $E(X_t) = 0$.

Definición 2.1 *Sea* $\{X_t : t \in \mathbb{Z}\}$ *una serie de tiempo con función media* $\mu_t = E(X_t)$, $t \in \mathbb{Z}$. *La serie centralizada es*

$$Y_t = X_t - \mu_t, \quad t \in \mathbb{Z}. \tag{2.1}$$

Siguiendo la idea anterior, cuando se tengan los valores numéricos $x_1, \ldots, x_n$ de una serie de tiempo que se intente modelar como serie estacionaria, primero se centralizan los datos restando a cada observación la media muestral, es decir, $x_t - \overline{x}$ y el análisis para ajustar algún modelo se hace con la serie transformada. La definición es entonces como se enuncia a continuación.

Definición 2.2 *Sea* $x_1, \ldots, x_n$ *una serie de tiempo con media muestral* $\bar{x} = (x_1 + \cdots + x_n)/n$. *La transformación de centralización se define como*

$$y_t = x_t - \overline{x}, \quad t = 1, \ldots, n. \tag{2.2}$$

Se puede comprobar fácilmente que la media de la serie transformada es cero, es decir, $\bar{y} = 0$. Los datos originales se obtienen a través de la transformación inversa $x_t = y_t + \bar{x}$, para $t = 1, \ldots, n$.

Transformación de Box-Cox[1]

La transformación de Box-Cox es en realidad una colección de varias transformaciones. Su definición se muestra en el siguiente recuadro.

[1] George Edward Pelham Box (1919–2013), estadístico inglés.
David Roxbee Cox (1924–2022), estadístico inglés.

Definición 2.3 *Sea $x_1, \ldots, x_n$ una serie de tiempo y sea λ cualquier número real. La transformación de Box-Cox de parámetro λ se define como*

$$y_t = \begin{cases} \dfrac{x_t^{\lambda} - 1}{\lambda} & si \;\; \lambda \neq 0, \\ \ln x_t & si \;\; \lambda = 0 \;\; y \;\; x_t > 0, \end{cases} \tag{2.3}$$

para valores de $t = 1, \ldots, n$.

De esta manera, para cada valor de λ se tiene una transformación distinta. Observe que en el primer renglón de la igualdad (2.3) cada dato x_t se eleva a la potencia λ.

	Valor de λ	Transformación	Nombre
(1)	-2	$y = -\dfrac{1}{2x^2} + \dfrac{1}{2}$	Inversa cuadrada
(2)	-1	$y = -\dfrac{1}{x} + 1$	Recíproca
(3)	$-1/2$	$y = -\dfrac{2}{\sqrt{x}} + 2$	Inversa raíz cuadrada
(4)	0	$y = \ln x$	Logaritmo natural
(5)	$1/2$	$y = 2\sqrt{x} - 2$	Raíz cuadrada
(6)	1	$y = x - 1$	Identidad
(7)	2	$y = \dfrac{1}{2}x^2 - \dfrac{1}{2}$	Cuadrada

Tabla 2.1: Transformaciones de Box-Cox.

En la Tabla 2.1 se muestran algunas transformaciones que forman parte de la familia de transformaciones de Box-Cox. Se observa que una transformación lineal adicional hace que las transformaciones (1)-(7) adquieran las siguientes expresiones: $1/x^2$, $1/x$, $1/\sqrt{x}$, $\ln x$, $\sqrt{x}$, x, x^2. En la figura 2.8 se pueden observar las distintas curvas que se obtienen a partir de las transformaciones (1)-(7) para valores positivos de x.

Se puede comprobar que la transformación de Box-Cox dada en (2.3) es invertible cuando se aplica sobre números positivos. En tal caso, a partir de los datos transformados $y_1, \ldots, y_n$, se puede recuperar la serie original a través de la transformación inversa dada por

$$x_t = \begin{cases} (\lambda y_t + 1)^{1/\lambda} & \text{si } \lambda \neq 0, \\ e^{y_t} & \text{si } \lambda = 0. \end{cases} \tag{2.4}$$

Para una serie de tiempo teórica $\{X_t : t \in \mathbb{Z}\}$, la transformación (2.3) también puede ser aplicada, y si la serie original es positiva, la transformación inversa es como en la igualdad (2.4).

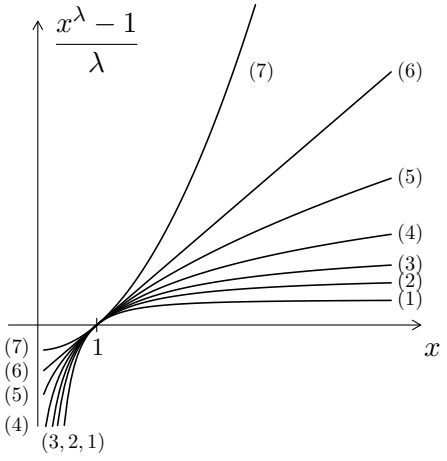

Figura 2.8: Transformación de Box-Cox para diferentes valores de λ.

Esta transformación se utiliza con distintos propósitos. En las series de tiempo es común aplicarla para intentar que se cumplan supuestos de normalidad o varianza constante. Además, se puede usar para suavizar el comportamiento de una serie. Por ejemplo, supongamos que la gráfica de una serie presenta un comportamiento creciente en las amplitudes de las oscilaciones, como se muestra en la figura 2.9.

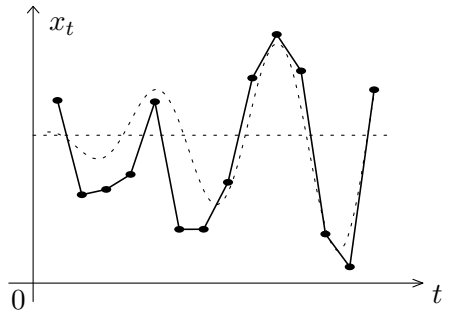

Figura 2.9: Serie de tiempo donde se sugiere realizar una transformación logarítmica.

En estos casos, transformar los datos con la función $\ln x_t$ puede ayudar a atenuar el crecimiento de las oscilaciones. Si es necesario, se puede sumar una constante $c > 0$ a cada observación para hacerla positiva y así poder aplicar el logaritmo. En la gráfica de la figura 2.10 se puede observar la serie original x_t y su transformación $y_t = \ln x_t + k$, donde k es un traslado para poder comparar más fácilmente las series.

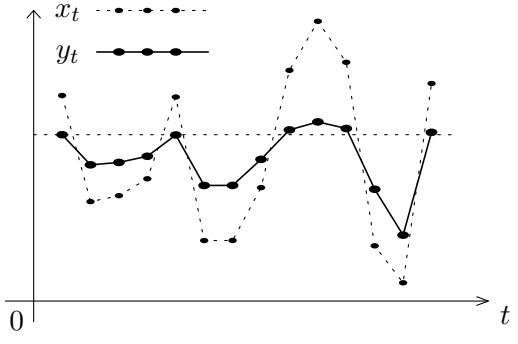

Figura 2.10: Serie de tiempo x_t y su transformación $y_t = \ln x_t + k$.

Después de aplicar la transformación de Box-Cox (2.3), la distribución de la muestra $y_1, \ldots, y_n$ depende de la distribución original de $x_1, \ldots, x_n$ y

del valor de λ. En general, no se puede especificar una distribución para los valores transformados $\{y_t\}$ si no se tiene mayor información sobre la distribución original de los valores de la serie $\{x_t\}$. Sin embargo, en algunos casos específicos, es posible que la distribución de la serie $\{y_t\}$ sea aproximadamente normal o siga una distribución conocida. En este sentido, se puede analizar si existe normalidad en los nuevos datos aplicando pruebas estadísticas no paramétricas. Por ejemplo, se puede utilizar la prueba de Shapiro-Wilk o la prueba de Kolmogorov-Smirnov, véase por ejemplo el libro de Conover (1999) [13].

Según sea el caso, existen distintos métodos para estimar el valor del parámetro λ más adecuado. Por ejemplo, el método gráfico que consiste en realizar gráficas con distintos valores de λ hasta ajustar visualmente los datos al comportamiento deseado. Este método sencillo, aunque subjetivo, sirve para tener una primera idea de la elección de λ. También se pueden usar los métodos de mínimos cuadrados, o bien, de máxima verosimilitud suponiendo que la transformación tiene distribución normal. Un método sencillo usando un programa de cómputo consiste en realizar varias pruebas de bondad de ajuste con distintos valores de λ hasta encontrar un buen ajuste estadístico a la distribución que se busca. A continuación se muestra un ejemplo de este último caso.

Ejemplo 2.1 *Se obtuvo una muestra* $x_1, \ldots, x_{1000}$ *de datos provenientes de una distribución* Exp(1)*. Se aplicó una transformación de Box-Cox con distintos valores de* λ *hasta obtener una muestra con el mayor p-value en la prueba de Kolmogorov-Smirnov de dos colas, donde la hipótesis nula establece que la muestra proviene de una distribución normal estándar. Se determinó un valor de* $\lambda = 0.2735$ *obteniéndose un estadístico* $D = 0.016336$*, y un p-value*$= 0.9523$*. En la figura 2.11 se puede observar la serie original* $\{x_t\}$ *y la serie* $\{y_t\}$*, donde para cada valor de t primero se calculó*

$$z_t = \frac{x_t^{\lambda} - 1}{\lambda},$$

para después calcular

$$y_t = \frac{z_t - \widehat{\mu}}{\widehat{\sigma}},$$

donde $\widehat{\mu}$ y $\widehat{\sigma}$ son la media muestral y la desviación estándar muestral de $\{z_t\}$, respectivamente. En la figura 2.12 se muestra el histograma sin frecuencias y el diagrama de cuantiles de la serie $\{y_t\}$. El código utilizado en este ejemplo se puede consultar en el Apéndice. ∎

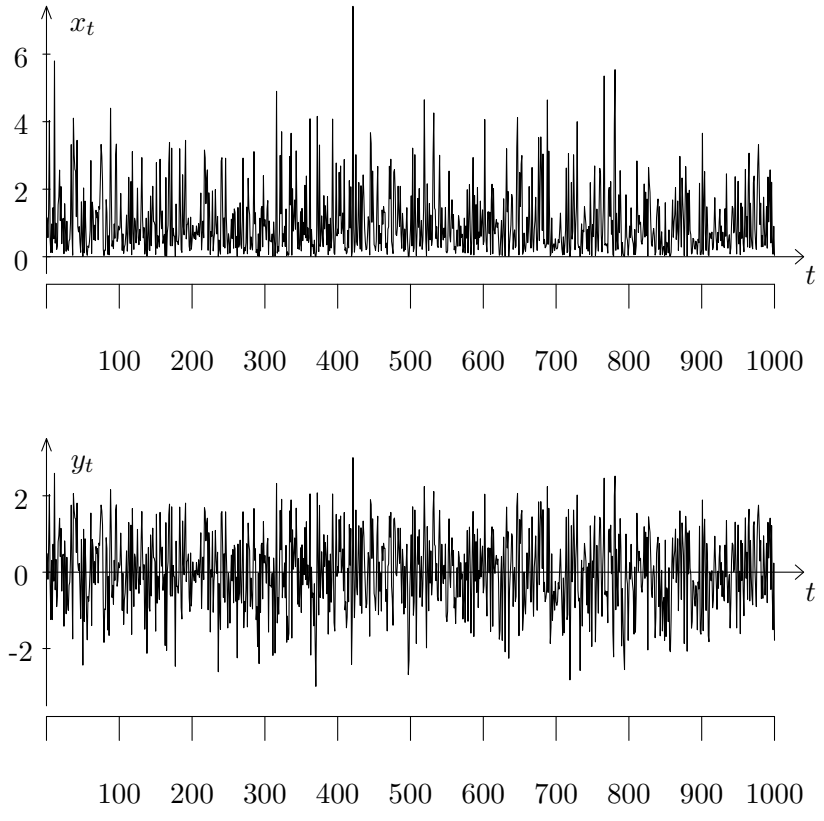

Figura 2.11: Serie de tiempo original y su transformación.

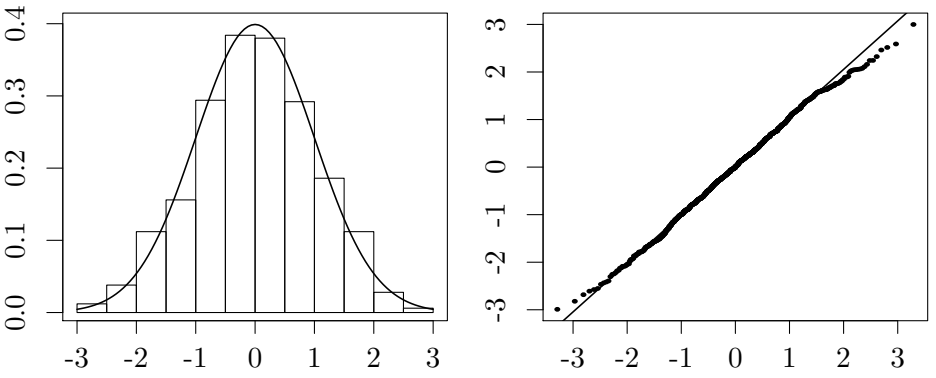

Figura 2.12: Histograma y diagrama de cuantiles de los datos transformados y estandarizados.

Transformación tasa de retorno

Cuando se tienen series de tiempo generadas en temas económicos y financieros, es relevante estudiar el comportamiento de los cambios de un periodo a otro. Para tal labor puede ser útil la siguiente transformación.

Definición 2.4 *Sea $x_1, \ldots, x_n$ una serie de tiempo. La transformación tasa de retorno se define como*

$$r_t = \ln \frac{x_t}{x_{t-1}}, \quad \text{para} \quad t = 2, \ldots, n, \tag{2.5}$$

suponiendo $x_t/x_{t-1} > 0$.

Observe que la fórmula (2.5) representa una medida para comparar los valores x_t y x_{t-1}. La transformación (2.5) se puede escribir también como

$$x_t = x_{t-1}\, e^{r_t}, \quad \text{para} \quad t = 2, \ldots, n. \tag{2.6}$$

De esta manera, a partir de x_1, de los valores $r_2, \ldots, r_n$ y usando (2.6) se puede reconstruir la serie original $x_1, x_2, \ldots, x_n$.

Para una serie de tiempo teórica $\{X_t : t \in \mathbb{Z}\}$, también se puede aplicar la transformación (2.5) cuando la serie es estrictamente positiva, produciendo la serie $\{R_t : t \in \mathbb{Z}\}$. La transformación inversa se calcula de manera iterativa como en (2.6), aunque se necesita un término inicial, X_1 por ejemplo, y las variables $R_2, R_3, \ldots$ para recuperar las variables originales $X_1, X_2, \ldots$

Por otro lado, en algunos casos es útil la siguiente transformación que guarda cierta similitud con la anterior.

Definición 2.5 *Sea $x_1, \ldots, x_n$ una serie de tiempo. La transformación tasa de crecimiento o tasa de retorno discreta se define como*

$$y_t = \frac{x_t - x_{t-1}}{x_{t-1}} = \frac{x_t}{x_{t-1}} - 1, \quad \text{para} \quad t = 2, \ldots, n, \tag{2.7}$$

suponiendo $x_{t-1} \neq 0$.

La expresión (2.7) representa otra medida de comparación de los valores x_t y x_{t-1}. La transformación (2.7) se puede escribir también como

$$x_t = x_{t-1}(y_t + 1), \quad \text{para} \quad t = 2, \ldots, n. \tag{2.8}$$

De esta manera, a partir de x_1, de los valores $y_2, \ldots, y_n$ y usando (2.8) se puede reconstruir la serie original $x_1, x_2, \ldots, x_n$.

Para una serie de tiempo teórica $\{X_t : t \in \mathbb{Z}\}$, la transformación (2.7) también se puede aplicar en el caso cuando la serie nunca se anula, produciendo así la serie $\{Y_t : t \in \mathbb{Z}\}$. La transformación inversa se calcula de manera iterativa como en (2.8), aunque se necesita un término inicial, X_1 por ejemplo, y las variables $Y_2, Y_3, \ldots$ para recuperar las variables originales $X_1, X_2, \ldots$

Las transformaciones (2.5) y (2.7) tienden a ser muy parecidas cuando las diferencias $x_t - x_{t-1}$ son relativamente pequeñas, pues tenemos la aproxi-

mación $\ln(1+x) \approx x$ para $-\delta < x < \delta$. En tal caso,

$$\begin{aligned} y_t &= \frac{x_t - x_{t-1}}{x_{t-1}} \\ &\approx \ln\,(1 + \frac{x_t - x_{t-1}}{x_{t-1}}) \\ &= \ln \frac{x_t}{x_{t-1}} \\ &= r_t. \end{aligned}$$

Lo anterior significa que cuando una serie de tiempo presenta tasas de crecimiento cercanas a cero, estas se asemejan a las tasas de retorno continuas.

2.3. Análisis de series con tendencia

En esta sección se presentan algunos métodos para estimar la tendencia de una serie de tiempo. Una vez estimada la posible tendencia, se puede restar de la serie original para producir una serie sin tendencia. Con el fin de que la exposición sea más clara, se supondrá que la serie de tiempo que se analiza presenta únicamente un componente de tendencia y un proceso irregular o aleatorio.

Regresión lineal

Cuando visualmente la serie presenta algunas desviaciones respecto de un patrón lineal de crecimiento o decrecimiento, se puede considerar el siguiente modelo aditivo de dos componentes:

$$X_t = m_t + Z_t, \quad t \in \mathbb{Z}, \tag{2.9}$$

en donde la tendencia es de la forma $m_t = a_0 + a_1\, t$, con a_0 y a_1 constantes desconocidas y $\{Z_t : t \in \mathbb{Z}\}$ es el componente aleatorio. En estos casos es posible estimar las constantes a_0 y a_1 por el método de mínimos cuadrados, para lo cual se buscan aquellos valores que minimizan la función

$$g(a_0, a_1) = \sum_{t=1}^{n} (x_t - m_t)^2 = \sum_{t=1}^{n} (x_t - a_0 - a_1 t)^2,$$

en donde $x_1, \ldots, x_n$ son n valores numéricos de la serie observada. Se puede comprobar que los estimadores para a_0 y a_1 están dados por

$$\widehat{a}_0 = \frac{1}{n}\sum_{t=1}^{n} x_t - \widehat{a}_1 \frac{1}{n}\sum_{t=1}^{n} t, \tag{2.10}$$

$$\widehat{a}_1 = \frac{n\sum_{t=1}^{n} t x_t - (\sum_{t=1}^{n} t)(\sum_{t=1}^{n} x_t)}{n\sum_{t=1}^{n} t^2 - (\sum_{t=1}^{n} t)^2}, \tag{2.11}$$

suponiendo $n \geqslant 2$. De esta manera se obtiene la estimación de la tendencia

$$\widehat{m}_t = \widehat{a}_0 + \widehat{a}_1 t, \quad t = 1, \ldots, n.$$

Los valores estimados del componente irregular Z_t se obtienen del cálculo de las diferencias

$$\widehat{z}_t = x_t - \widehat{m}_t, \quad \text{para} \;\; t = 1, \ldots, n.$$

Si se desea pronosticar un valor futuro x_{n+m}, con $m \in \mathbb{N}$, se puede usar la tendencia estimada $\widehat{m}_t$ evaluada en $t = n + m$. Esto es razonable cuando la serie del componente irregular muestra signos de estar compuesta por variables aleatorias no correlacionadas de media 0. La posible correlación no nula de los residuos puede ayudar a proponer un mejor pronóstico. En la figura 2.13 se puede observar un ejemplo de trayectoria de los residuos $\{\widehat{z}_t : t = 1, \ldots, n\}$, la cual se generó simulando un ruido blanco.

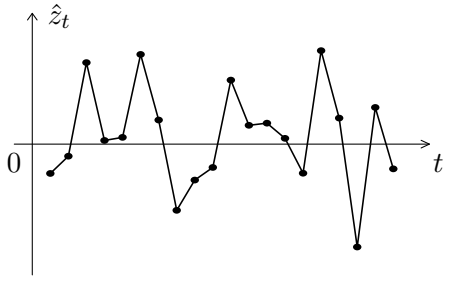

Figura 2.13: Ejemplo de una trayectoria de residuos.

Regresión polinomial

En algunos casos la tendencia de una serie de tiempo puede, aparentemente, seguir la trayectoria suave de un polinomio de un determinado grado, de modo que el ajuste de tal modelo puede parecer natural. En esta sección se considera la siguiente descomposición

$$X_t = m_t + Z_t, \quad t \in \mathbb{Z},$$

en donde se propone como la tendencia m_t al polinomio en una variable (el tiempo t) dado por

$$m_t = a_0 + a_1 t + a_2 t^2 + \cdots + a_k t^k,$$

para algún valor entero $k \geqslant 1$ (el grado del polinomio) y en donde, nuevamente, las constantes $a_0, a_1, \ldots, a_k$ son desconocidas y es necesario estimarlas. A esto se le llama regresión polinomial en una variable. En la figura 2.2 se muestra una tendencia no lineal sugerida por un conjunto de observaciones. Es claro que conocer el comportamiento gráfico de los polinomios de distintos grados puede ayudar a proponer un valor particular del parámetro k. Supondremos que tal valor ha sido especificado y nos concentraremos en estimar las constantes $a_0, a_1, \ldots, a_k$.

Sea $x_1, x_2, \ldots, x_n$ el conjunto de observaciones. Para considerar el desarrollo matemático un poco más general, supondremos que estas observaciones se realizaron en los tiempos $t_1 < t_2 < \cdots < t_n$. Entonces, el modelo que se propone es

$$x_t = a_0 + a_1 t + a_2 t^2 + \cdots + a_k t^k + z_t, \quad \text{para} \quad t = t_1, \ldots, t_n,$$

en donde $z_1, z_2, \ldots, z_n$ son las diferencias entres los datos observados y el modelo polinomial, una vez que se estiman los valores de $a_0, a_1, \ldots, a_k$. Se tiene así un sistema de ecuaciones que se puede escribir en términos matriciales de la forma siguiente:

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^k \\ 1 & t_2 & t_2^2 & \cdots & t_2^k \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_n & t_n^2 & \cdots & t_n^k \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{pmatrix} + \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}. \tag{2.12}$$

Denotemos por A a la matriz que aparece en el sistema (2.12), la cual es una matriz de $n \times (k+1)$, en donde n es el número de datos y k es el grado del polinomio, se le conoce con el nombre de matriz rectangular de Vandermonde. Cuando $k = 1$ se obtiene la regresión lineal mencionada en la sección anterior.

El sistema (2.12) se puede escribir como

$$\underline{\mathrm{x}} = A\underline{\mathrm{a}} + \underline{\mathrm{z}},$$

en donde $\underline{\mathrm{x}}$, $\underline{\mathrm{a}}$ y $\underline{\mathrm{z}}$ son los vectores columna que aparecen en (2.12). Para estimar los valores de $a_0, a_1, \ldots, a_k$ se utiliza usualmente el método de mínimos cuadrados. El método consiste en encontrar aquellos valores que minimizan la función

$$\begin{aligned} g(a_0, a_1, \ldots, a_k) &= \sum_{t=1}^{n} (x_t - m_t)^2 \qquad (2.13) \\ &= \sum_{t=1}^{n} (x_t - a_0 - a_1 t - a_2 t^2 - \cdots - a_k t^k)^2. \end{aligned}$$

Se puede comprobar que el sistema de ecuaciones que resulta del método de mínimos cuadrados para estimar $\underline{\mathrm{a}} = (a_0, a_1, \ldots, a_k)^t$ se puede escribir como

$$A^t\underline{\mathrm{x}} = (A^tA)\underline{\mathrm{a}},$$

de modo que se tiene la solución explícita

$$\underline{\mathrm{a}} = (A^tA)^{-1}A^t\underline{\mathrm{x}}, \qquad (2.14)$$

cuando la matriz cuadrada A^tA es invertible. Observe que la matriz A^tA es de tamaño $(k+1) \times (k+1)$. La solución (2.14) también se puede obtener a partir de la ecuación $\underline{\mathrm{x}} = A\underline{\mathrm{a}}$ multiplicando por la izquierda por A^t y después invirtiendo A^tA.

Una vez estimados los valores para $a_0, a_1, \ldots, a_k$, se obtiene la tendencia polinomial estimada $\widehat{m}_t = \widehat{a}_0 + \widehat{a}_1 t + \widehat{a}_2 t^2 + \cdots + \widehat{a}_k t^k$ y se puede estimar el componente irregular mediante las diferencias

$$\widehat{z}_t = x_t - \widehat{m}_t, \quad \text{para} \;\; t = t_1, \ldots, t_n.$$

Las características de la serie de residuos $\{\widehat{z}_t : t = 1, \ldots, n\}$ pueden ser de utilidad para generar un estimador distinto al caracterizado por la tendencia estimada.

Suavizamiento por promedios móviles

El suavizamiento por promedios móviles es una forma no paramétrica de estimar el componente de tendencia de una serie para la que se presume que sigue el modelo aditivo de dos componentes:

$$X_t = m_t + Z_t, \quad t \in \mathbb{Z}.$$

La definición de la estimación es la que se enuncia a continuación.

Definición 2.6 *Sea $x_1, \ldots, x_n$ una serie que es observada en los tiempos $t = 1, \ldots, n$. Sea $q \geqslant 1$ un entero. Se define la serie de promedios móviles como*

$$\widehat{m}_t = \frac{1}{2q+1} \sum_{j=-q}^{q} x_{t-j} \quad \textit{para} \quad t = 1, \ldots, n. \tag{2.15}$$

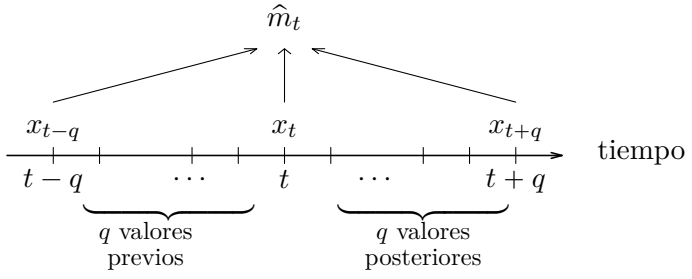

Figura 2.14: Promedio móvil de orden q.

La expresión (2.15) corresponde al promedio de las $2q + 1$ observaciones vecinas o adjuntas a la observación x_t. Estos valores son: el valor x_t (cuando $j = 0$), los q valores previos a x_t (cuando $j = 1, 2, \ldots, q$) y los q valores posteriores a x_t (cuando $j = -q, \ldots, -1$). Se trata de un promedio móvil

de dos lados y simétrico. Se le conoce también con el nombre de filtro lineal de orden q, y a la serie $\{\widehat{m}_t : t = 1, \ldots, n\}$ se le llama proceso filtrado. Este esquema se muestra en la figura 2.14.

Observe que la nueva serie de tiempo resultante $\{\widehat{m}_t\}$ está definida para aquellos tiempos t donde haya q valores previos y q valores posteriores en la serie original. Estos valores son $t = 1 + q, \ldots, n - q$. Sin embargo, se puede calcular $\widehat{m}_t$ para $t = 1, 2, \ldots, n$ extendiendo artificialmente la serie $\{x_t\}$ de la siguiente manera:

$$x_t := \begin{cases} x_1 & \text{si } \ t \leqslant 0 \qquad\quad \text{(constante } x_1 \text{ antes de } t = 1), \\ x_n & \text{si } \ t \geqslant n+1 \quad \text{(constante } x_n \text{ después de } t = n). \end{cases}$$

Es decir, se repite el primer valor x_1 para tiempos menores a $t = 1$ tantas veces como sea necesario según el valor de q, y también se repite el último valor x_n para tiempos mayores a $t = n$ tantas veces como sea necesario. De esta manera la expresión (2.15) queda ahora bien definida para $t = 1, \ldots, n$.

El suavizamiento por promedios móviles atenúa el componente aleatorio en series de tiempo de la forma $X_t = m_t + Z_t$. En efecto, sustituyendo la descomposición $x_t = m_t + z_t$ en la definición de $\widehat{m}_t$, se tiene que

$$\begin{aligned} \widehat{m}_t &= \frac{1}{2q+1} \sum_{j=-q}^{q} x_{t-j} \\ &= \frac{1}{2q+1} \sum_{j=-q}^{q} m_{t-j} + \underbrace{\frac{1}{2q+1} \sum_{j=-q}^{q} z_{t-j}}_{\approx 0} \\ &\approx \frac{1}{2q+1} \sum_{j=-q}^{q} m_{t-j} \\ &\approx m_t. \end{aligned}$$

La primera aproximación es válida si se considera que $z_1, \ldots, z_n$ son valores de un ruido o desviación aleatoria que, en promedio, toma el valor cero. La segunda aproximación es factible para valores pequeños de q, o bien en periodos de un comportamiento lineal de la tendencia.

De esta manera, los valores $\widehat{m}_t$ se pueden tomar como una estimación del componente de tendencia desconocido m_t de la serie. Más generalmente, se puede proponer un filtro lineal de la forma

$$\widehat{m}_t = \sum_{j=-q}^{q} a_j \, x_{t-j},$$

para ciertas constantes (ponderaciones) a_j, o bien el filtro no simétrico

$$\widehat{m}_t = \sum_{j=-q_1}^{q_2} a_j \, x_{t-j},$$

para ciertos enteros $q_1 \geqslant 0$ y $q_2 \geqslant 0$, en donde, nuevamente, la serie original se debe extender para que estas expresiones estén bien definidas para $t = 1, \ldots, n$. De la misma manera que en los casos anteriores, una vez estimada la tendencia, el componente irregular estimado se define como la serie de residuos

$$z_t = x_t - \widehat{m}_t, \quad \text{para} \quad t = 1, \ldots, n.$$

Suavizamiento exponencial

Otra técnica no paramétrica para estimar la tendencia de una serie de tiempo de la forma $X_t = m_t + Z_t$, es atenuar la dependencia que pudiera existir entre los valores que sean lejanos dentro de la serie. Esto se puede realizar como se enuncia a continuación.

Definición 2.7 *Sea $x_1, \ldots, x_n$ una serie que es observada en los tiempos $t = 1, \ldots, n$. Sea $\alpha \in [0, 1]$ un número fijo. Se define la serie por suavizamiento exponencial $\{\widehat{m}_t : t = 1, \ldots, n\}$ como aparece a continuación. Al número α se le llama constante de suavizamiento.*

$$\widehat{m}_t = \begin{cases} x_1 & si \quad t = 1, \\ \alpha x_t + (1 - \alpha)\, \widehat{m}_{t-1} & si \quad t = 2, \ldots, n. \end{cases} \tag{2.16}$$

Se trata de un promedio ponderado de dos valores: el valor de la serie al tiempo t y el valor estimado de la tendencia al tiempo anterior $t - 1$ con

ponderadores α y $(1-\alpha)$, respectivamente. Observemos que cuando $\alpha = 1$, se obtiene la serie original $\widehat{m}_t = x_t$.

Si se desarrolla explícitamente la fórmula recursiva (2.16) se obtiene lo siguiente:

$$\begin{aligned}
\widehat{m}_t &= \alpha x_t + (1-\alpha)\widehat{m}_{t-1} \\
&= \alpha x_t + (1-\alpha)(\alpha x_{t-1} + (1-\alpha)\widehat{m}_{t-2}) \\
&= \alpha x_t + \alpha(1-\alpha)x_{t-1} + (1-\alpha)^2\widehat{m}_{t-2} \\
&= \alpha x_t + \alpha(1-\alpha)x_{t-1} + (1-\alpha)^2(\alpha x_{t-2} + (1-\alpha)\widehat{m}_{t-3}) \\
&= \alpha x_t + \alpha(1-\alpha)x_{t-1} + \alpha(1-\alpha)^2 x_{t-2} + (1-\alpha)^3\widehat{m}_{t-3} \\
&\vdots
\end{aligned}$$

Siguiendo con la substitución hacia atrás se llega a la expresión siguiente, la cual sólo depende de los valores de la serie,

$$\widehat{m}_t = \sum_{j=0}^{t-2} \alpha(1-\alpha)^j x_{t-j} + (1-\alpha)^{t-1} x_1, \tag{2.17}$$

para valores de $t = 2, \ldots, n$. En la fórmula (2.17) se puede observar la dependencia explícita de $\widehat{m}_t$ de los valores $x_1, x_2, \ldots, x_t$. Esta dependencia se hace cada vez menor (decrecimiento exponencial) para aquellos valores más alejados en el tiempo. Es decir, la historia reciente tiene una mayor preponderancia que la historia lejana. En otras palabras, el suavizamiento exponencial es un proceso de promedios móviles de un solo lado (se utilizan todos los valores previos de la serie, incluyendo el tiempo t) y con ponderadores cambiantes en el tiempo, decayendo a cero exponencialmente. Una vez obtenida la tendencia estimada $\widehat{m}_t$, el componente irregular estimado estará dado por la serie de residuos

$$z_t = x_t - \widehat{m}_t, \quad \text{para} \quad t = 1, \ldots, n.$$

Diferenciación

Varios modelos de series de tiempo incluyen la hipótesis de estacionariedad y, como se ha comentado en el primer capítulo, se puede eliminar la

tendencia en una serie estacionaria. Por lo anterior, no sólo son importantes los métodos para estimar la tendencia, aún más importante es tener un mecanismo para atenuar o eliminar tal tendencia. Para lograrlo, una idea sugerida por Box y Jenkins (1994) [5] es aplicar un filtro llamado diferenciación. Cabe mencionar que desde 1970 estos autores establecieron toda una metodología para modelar series de tiempo siguiendo las etapas de identificación, estimación y aplicación de pruebas de correcta especificación del modelo. En la etapa de identificación es común aplicar la diferenciación y posiblemente alguna transformación adicional para lograr la estacionariedad de la serie. Recordemos que el operador diferencia de orden k se definió en la página 21 de la siguiente manera

$$\nabla^k X_t = (1 - B)^k X_t, \quad t \in \mathbb{Z},$$

en donde $BX_t = X_{t-1}$ es el operador retraso de un tiempo. Como se menciona en el ejercicio 18 del capítulo 1, si el modelo para la tendencia de una serie es de la forma $m_t = a_0 + a_1 t + \cdots + a_k t^k$, entonces la aplicación del operador diferencia de orden k produce la serie

$$\nabla^k X_t = k!\, a_k + \nabla^k Z_t, \quad \text{para} \;\; t \geqslant k.$$

Es decir, el operador ∇^k elimina la tendencia polinomial de grado k de la serie. Por ser de uso más común, a continuación se escriben explícitamente las expresiones de las dos primeras diferencias. La primera diferencia es

$$\nabla X_t = X_t - X_{t-1}, \quad \text{para} \;\; t = 1, 2, \ldots \tag{2.18}$$

y la segunda diferencia es

$$\nabla^2 X_t = X_t - 2X_{t-1} + X_{t-2}, \quad \text{para} \;\; t = 2, 3, \ldots \tag{2.19}$$

2.4. Análisis de series con estacionalidad

En ocasiones las series de tiempo pueden mostrar un patrón que se repite o que es similar cada cierto tiempo. Esto es lo que hemos definido como la estacionalidad de una serie. Por ejemplo, es común que las series que emanan de datos económicos sean estacionales debido a que muchos fenómenos económicos se ligan a meses, semestres, años, etc. Otros fenómenos, por

ejemplo la demanda de gas o la venta de boletos de avión, exhiben estacionalidad anual, ya que tienen una fuerte relación con las estaciones que se repiten año con año. Este comportamiento cíclico, que es parecido en periodos de tiempo fijo, representa un componente de estacionalidad que impide que la serie sea estacionaria. Por lo tanto, se busca eliminar el componente estacional para poder aplicar alguno de los modelos que requieren que la serie sea estacionaria. Más adelante se retoma el tema acerca de la forma en la que puede intentarse eliminar la estacionalidad en una serie de tiempo.

Estimación del componente estacional

Supongamos que una serie carece de tendencia y es de la forma

$$X_t = s_t + Z_t, \tag{2.20}$$

en donde el término s_t es un componente estacional y Z_t es el componente irregular. La estacionalidad de la función s_t se expresa solicitando que sea una función periódica de periodo d, es decir, que se cumpla la identidad $s_{t+d} = s_t$, para $t \in \mathbb{Z}$, en donde el número $d \geqslant 2$ es un entero que indica la longitud (número de unidades de tiempo) del intervalo de repetición.

Existen distintas maneras de intentar modelar y estimar el componente s_t. Por ejemplo, para un patrón de repetición en tiempo discreto se puede proponer la función

$$s_t = \sum_{i=1}^{d} \alpha_i \, C_{t,i}, \tag{2.21}$$

en donde $\alpha_1, \ldots, \alpha_d$ son parámetros desconocidos que sirven para ponderar el efecto de que suceda o no suceda un evento de interés. El término $C_{t,i}$ es una variable que sólo toma los valores cero o uno, que en el contexto de series de tiempo se conoce como variable *dummy*. Modela la ocurrencia de cierto fenómeno cada cierto tiempo. Por ejemplo, puede modelar la ocurrencia de aumento de turismo cada semestre con respecto a los meses anteriores. Si se mide el tiempo en semestres, entonces se tendrá que $d = 2$, por lo cual

$$s_t = \alpha_1 \, C_{t,1} + \alpha_2 \, C_{t,2},$$

en donde $C_{t,1}$ tomará el valor de uno cuando aumente el turismo en julio (cuando el valor de t sea impar) y $C_{t,2}$ será uno cuando aumente el turismo en enero (cuando el valor de t sea par).

Si la serie exhibe una forma cíclica continua, entonces se puede usar la siguiente combinación lineal de senos y cosenos

$$s_t = a_0 + \sum_{j=1}^{k} \left[a_j \cos(\lambda_j t) + b_j \operatorname{sen}(\lambda_j t) \right], \tag{2.22}$$

en donde $k \in \mathbb{N}$, $a_0, a_1, \ldots, a_k, b_1, \ldots, b_k$ son constantes que indican las amplitudes de las ondas, y las constantes $\lambda_1, \ldots, \lambda_k$ modifican las frecuencias de los senos y cosenos. Notemos que si d es par y $k = d/2$, entonces haciendo $\lambda_j = 2\pi j/d$ los modelos (2.21) y (2.22) son equivalentes.

Para el modelo (2.22), y suponiendo que el valor de k es conocido, se busca encontrar el valor de los otros parámetros que ajusten, usando mínimos cuadrados, el modelo teórico a los datos observados. A este proceso de estimación se le llama regresión armónica. Por ejemplo, la función $s_t = 2 + \cos(3t) + \cos(6t)$ presenta el comportamiento cíclico que se muestra en la figura 2.15.

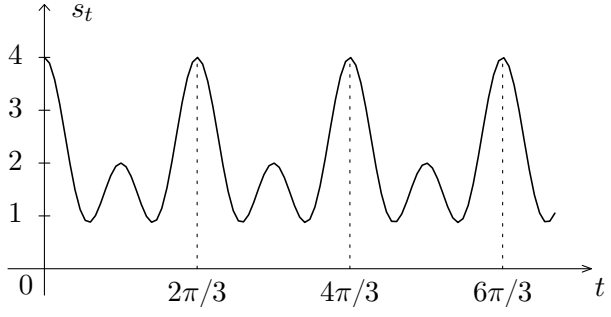

Figura 2.15: Comportamiento cíclico $s_t = 2 + \cos(3t) + \cos(6t)$.

Estimación y eliminación de la tendencia y la estacionalidad

Sean $x_1, \ldots, x_n$ los valores numéricos observados de una serie de tiempo para la cual se ha adoptado el modelo

$$X_t = m_t + s_t + Z_t, \quad t \in \mathbb{Z}, \tag{2.23}$$

en donde m_t es el componente de tendencia, s_t es el componente de estacionalidad de periodo d, es decir, $s_{t+d} = s_t$, y Z_t es el componente irregular. A continuación veremos dos métodos para estimar los componentes m_t y s_t del modelo (2.23).

Método de promedios móviles

a) Se estima preliminarmente la tendencia por el método de los promedios móviles distinguiendo los siguientes dos casos, dependiendo de la paridad del periodo d:

- Si d es par, se toman los promedios móviles de orden $q = d/2$ con las siguientes ponderaciones

$$\widehat{m}_t = \frac{1}{2d}\, x_{t-q} + \frac{1}{d}\, x_{t-q+1} + \cdots + \frac{1}{d}\, x_{t+q-1} + \frac{1}{2d}\, x_{t+q},$$

para $t = q+1, \ldots, n-q$. De esta manera, el promedio móvil se efectúa sobre periodos de longitud total $d+1$ de manera simétrica: $d/2$ términos hacia adelante y $d/2$ términos hacia atrás, e incluyendo el término central x_t. Observe que se tiene la ponderación $1/(2d)$ en los extremos izquierdo y derecho de esta expresión, y se tiene la ponderación distinta $1/d$ para el resto de los términos; sin embargo, la suma de estas ponderaciones es $2\,(1/2d) + 2\,(q-1)/d + 1/d = 1$.

- Si d es impar, se toman promedios móviles de orden $q = (d-1)/2$ con idénticas ponderaciones, es decir,

$$\widehat{m}_t = \frac{1}{d}\, x_{t-q} + \frac{1}{d}\, x_{t-q+1} + \cdots + \frac{1}{d}\, x_{t+q-1} + \frac{1}{d}\, x_{t+q},$$

para $t = q+1, \ldots, n-q$. El total de sumandos es $2q+1 = d$ y, por lo tanto, las ponderaciones también suman uno.

Como antes, los promedios móviles recién indicados se pueden definir para cualquier tiempo $t = 1, \ldots, n$, extendiendo las observaciones x_t más allá de estos tiempos, de la siguiente manera

$$x_t = \begin{cases} x_1 & \text{si } t \leqslant 0 \qquad \text{(constante } x_1 \text{ antes de } t = 1\text{)}, \\ x_n & \text{si } t \geqslant n+1 \qquad \text{(constante } x_n \text{ después de } t = n\text{)}. \end{cases}$$

b) Una vez estimada la tendencia, se estima el componente estacional. Primero se definen los promedios de las desviaciones de la serie y la tendencia estimada. Para $t = 1, \ldots, d$,

$$w_t := \frac{1}{N} \sum_j (x_{t+jd} - \widehat{m}_{t+jd}), \tag{2.24}$$

en donde la suma se efectúa para valores de j tales que $q < t+jd \leqslant n-q$, en donde N es el total de estos valores. Se define la estimación para el componente estacional s_t de la forma siguiente

$$\widehat{s}_t = \begin{cases} w_t - \dfrac{1}{d} \displaystyle\sum_{i=1}^{d} w_i, & \text{si } \ t = 1, \ldots, d, \\ \widehat{s}_{t-d} & \text{si } \ t \geqslant d+1. \end{cases}$$

La serie desestacionalizada estará dada por

$$d_t = x_t - \widehat{s}_t, \quad \text{para } \ t = 1, \ldots, n.$$

c) Se vuelve a estimar la tendencia pero ahora de los datos desestacionalizados $\{d_t : t = 1, \ldots, n\}$ mediante alguno de los métodos vistos antes para series que sólo presentan tendencia y ruido. Esta es la versión final de $\widehat{m}_t$. Finalmente, el componente aleatorio estimado es

$$\widehat{z}_t := x_t - \widehat{m}_t - \widehat{s}_t, \quad \text{para } \ t = 1, \ldots, n.$$

Método por diferenciación

Este es un método sencillo que hace uso del operador ∇_d definido en la sección 1.3, en donde ahora el parámetro es el entero $d \geqslant 0$. Veremos que

aplicar este operador a una serie de tiempo que sigue el modelo (2.23) elimina el componente estacional. Sea $x_1, \ldots, x_n$ la serie observada. Recordemos que

$$\nabla_d x_t = x_t - x_{t-d} = (1 - B^d)x_t,$$

para $t = d + 1, \ldots, n$. Estamos suponiendo que $2 \leqslant d < n$, es decir, el número de datos n excede el periodo d del componente estacional, el cual es por lo menos 2. Al aplicar el operador ∇_d a la serie $\{x_t\}$ se obtiene

$$\begin{aligned} \nabla_d x_t &= \nabla_d (m_t + s_t + Z_t) \\ &= \nabla_d m_t + \nabla_d s_t + \nabla_d Z_t \\ &= (m_t - m_{t-d}) + \underbrace{(s_t - s_{t-d})}_{=0} + (Z_t - Z_{t-d}) \\ &= (m_t - m_{t-d}) + (Z_t - Z_{t-d}), \end{aligned}$$

para $t = d + 1, \ldots, n$. Es decir, la serie transformada $\nabla_d x_t$ sólo tiene el componente de tendencia dado por la función $m_t - m_{t-d}$, y el componente aleatorio dado por $Z_t - Z_{t-d}$. De esta manera se ha eliminado el componente estacional.

Ahora se puede estimar y eliminar la tendencia por alguno de los métodos vistos antes y obtener una serie que es función únicamente del componente aleatorio.

Suavizamiento exponencial de Holt-Winters

Este método se puede usaro para realizar pronósticos cuando se tiene una tendencia y un comportamiento estacional con un periodo conocido. Se puede ver como una generalización del algoritmo recursivo usado en el suavizamiento exponencial considerando la tendencia y la estacionalidad. Se ponderan las observaciones del pasado teniendo en cuenta los patrones tanto de la serie como de su componente estacional y de su tendencia. Este método se utiliza con frecuencia en series de tiempo con datos relacionados con inventarios y ventas. El lector interesado en profundizar en este tema puede consultar, por ejemplo, Kalekar (2004) [23], Lima (2009) [28] o Goodwin (2010) [16]. Estos son trabajos breves que muestran la manera de usar este método de manera práctica.

2.5. Interpretación del correlograma

Recordemos que el correlograma es la gráfica de la función de autocorrelación $\rho(\tau)$ de una serie de tiempo. En la definición 1.9 de la página 18 se proporciona una fórmula para la autocorrelación muestral $r(\tau)$ de una serie de observaciones. En esta sección se dan algunos lineamientos para la interpretación del correlograma muestral. El gráfico del correlograma muestral $r(\tau)$ puede ser de utilidad para realizar un primer análisis de algunos componentes de la serie de tiempo o para proponer alguno de los modelos teóricos.

Debemos considerar que si una serie de observaciones $x_1, \ldots, x_n$ se toma de un modelo teórico $\{X_t : t \in \mathbb{Z}\}$, entonces el correlograma muestral presentará características semejantes al correlograma del modelo. Sin embargo, en las aplicaciones lo único que conocemos es el conjunto de observaciones $x_1, \ldots, x_n$. De esta manera, si las observaciones presentan un correlograma muestral semejante al correlograma de algún modelo teórico conocido, entonces es natural pensar que los datos provienen de tal modelo. Por otro lado, es importante considerar que la función de autocorrelación de una serie no necesariamente caracteriza de manera única al modelo, es decir, en general, no hay una correspondencia biunívoca entre estos dos objetos.

En ocasiones puede ser complicado dar una interpretación del correlograma. Sin embargo, los criterios propuestos por Chatfield (1996) [11] que se mencionan a continuación pueden resultar de ayuda. En la figura 2.16 se pueden observar algunos ejemplos de este tipo de correlogramas.

Serie de variables aleatorias independientes

Para el modelo de serie de tiempo compuesto de variables aleatorias independientes, la función de autocorrelación $\rho(\tau)$ es 1 para $\tau = 0$ y es 0 para $\tau = 1, 2, \ldots$ Por lo tanto, si las observaciones muestran un correlograma muestral con un valor igual a 1 cuando $\tau = 0$ y con valores cercanos a 0 para $\tau = 1, 2, \ldots$, entonces es muy posible que las observaciones provengan del modelo compuesto por variables aleatorias independientes.

Correlación a corto plazo

En algunas ocasiones una serie de tiempo estacionaria muestra un correlograma en el cual aparece cierta correlación en los primeros valores de τ (en los primeros rezagos) y tiende a cero rápidamente en los rezagos siguientes. Esto se debe a pequeñas rachas que se presenten dentro de la trayectoria. Se dice que hay una racha cuando dos o más incrementos (o decrementos) son consecutivos.

Serie alternada

Se dice que una serie de tiempo es alternante si sus valores van cambiando hacia arriba y hacia abajo con respecto a su media muestral. En este caso, el correlograma tiende a alternar sus valores entre positivos y negativos, por ejemplo, $r(1)$ negativo, $r(2)$ positivo, $r(3)$ negativo, etc.

Series de tiempo no estacionarias

Si el correlograma no toma valores cercanos a cero a partir de alguno de los primeros rezagos, es muy posible que la serie no sea estacionaria. En el caso cuando una serie exhibe tendencia, habrá varias observaciones consecutivas que estén por arriba o por abajo de la media. Esto provocará que los valores del correlograma no disminuyan durante varios rezagos.

Series con componente estacional

El correlograma de las series de tiempo con un componente estacional, en algunos casos podría heredar tal estacionalidad, en el sentido de que habrá una autocorrelación con un patrón de repetición con valores más grandes cada d rezagos, donde d es el periodo estacional de la serie. Por otro lado, al aplicar algún filtro que remueva este componente, el nuevo correlograma puede aportar información valiosa sobre la serie resultante.

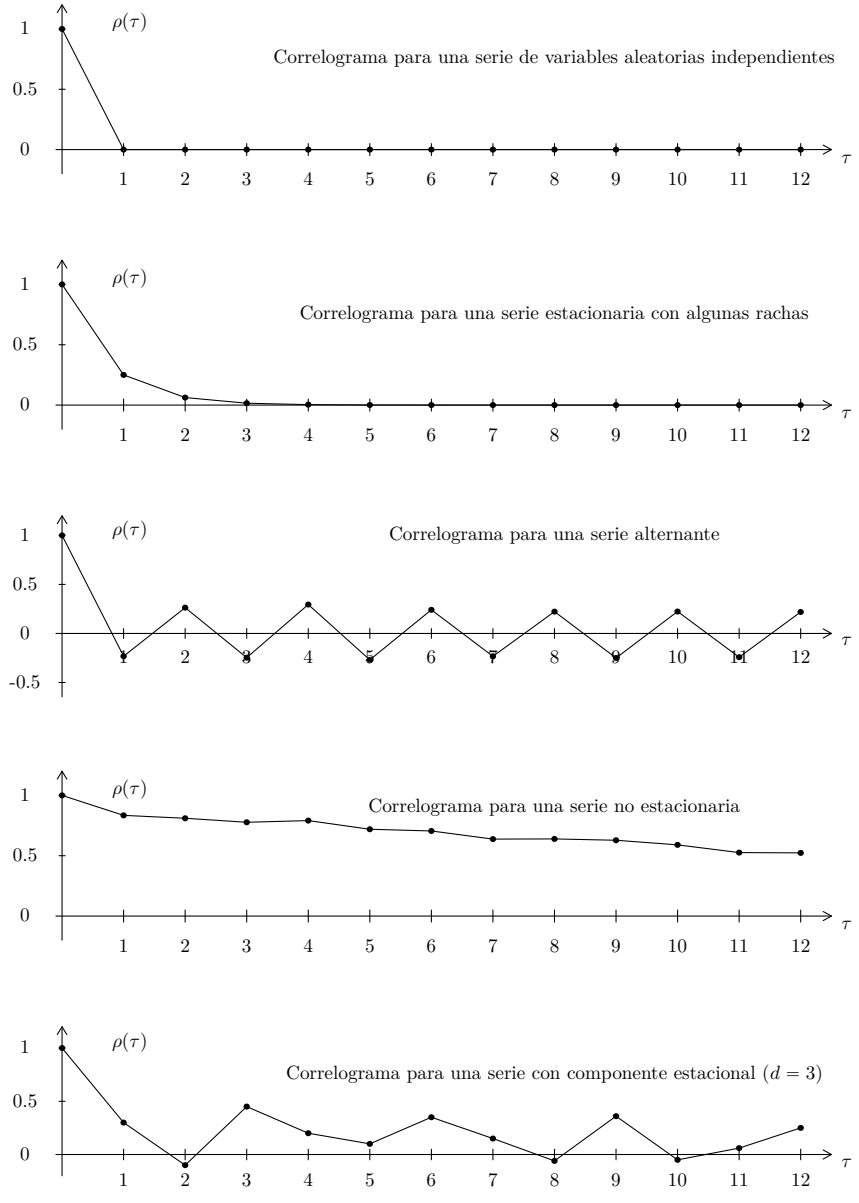

Figura 2.16: Ejemplos de correlogramas según las características de la serie.

Valores atípicos

Cuando se presentan valores atípicos (*outliers*) en una serie, esto provoca que el correlograma muestre información engañosa. Por lo anterior, para hacer un análisis de la serie se puede intentar atenuar o remover el valor atípico. No existe una regla sobre cómo substituir un valor atípico, pero un mecanismo práctico puede ser reemplazar esa observación por el promedio de otras observaciones. Por ejemplo, en el caso cuando la serie presenta cierta tendencia, se podría reemplazar el valor atípico x_t por el promedio $(x_{t-1} + x_{t+1})/2$. Por otro lado, si la serie presenta un comportamiento periódico cada p observaciones, se podría reemplazar el valor de x_t por el promedio $(x_{t-p} + x_{t+p})/2$.

Como se mencionó, los coeficientes de correlación $r(\tau)$ (ACF muestral) aportan información de la serie, pero esta herramienta no es suficiente para decidir sobre algunas características o componentes de la serie. En el capítulo siguiente se especifican algunos modelos particulares comunes en las aplicaciones, y se calcula la función de autocorrelación en cada caso. Observando dichos resultados, se podrán apreciar otras formas de aprovechar e interpretar el correlograma muestral. Por último, también en el capítulo siguiente se definirá la llamada función de autocorrelación parcial (denotada como PACF en la mayoría de la literatura afín) que genera un gráfico alterno parecido al correlograma. Este gráfico normalmente se analiza junto al gráfico de la función ACF y es de gran importancia en la etapa de identificación de modelos.

2.6. Ejercicios

Gráficas con respecto al tiempo

34. Explique la diferencia entre el componente estacional y el cíclico en una serie de tiempo.

35. Investigue los precios diarios de alguna acción financiera durante un año. Debido a que los fines de semana no funcionan los mercados accionarios, después de cada 5 datos (lunes a viernes) asigne a las siguientes 2 observaciones (sábado y domingo) el valor promedio de los precios de los días viernes y lunes adyacentes. Realice y analice el

gráfico de la serie de tiempo resultante y observe si existe algún patrón de comportamiento. ¿Se percibe alguna tendencia? Responda si cree que existe algún componente estacional, cíclico o aleatorio.

36. Investigue el consumo de energía eléctrica mensual de algún periodo de 10 años para algún país que usted elija. Realice y analice el gráfico de la serie de tiempo resultante y observe si existe algún patrón de comportamiento. ¿Se percibe alguna tendencia? Responda si cree que existe algún componente estacional, cíclico o aleatorio.

37. Investigue las temperaturas promedio mensuales de algún periodo de 10 años para algún país que usted elija. Realice y analice el gráfico de la serie de tiempo resultante y observe si existe algún patrón de comportamiento. ¿Se percibe alguna tendencia? Responda si cree que existe algún componente estacional, cíclico o aleatorio.

Transformaciones

38. Sea $y_1, \ldots, y_n$ la serie obtenida a través de la transformación de centralización definida en (2.2). Demuestre que $\bar{y} = 0$.

39. Demuestre que la transformación de Box-Cox es continua como función del parámetro λ, es decir, compruebe que

$$\lim_{\lambda \to 0} \frac{x^\lambda - 1}{\lambda} = \ln x, \quad \text{para } x > 0.$$

40. Demuestre que la transformación de Box-Cox es invertible encontrando la expresión de la transformación inversa.

41. Investigue los precios diarios de alguna acción financiera durante un año. Construya una serie de tiempo $x_1, \ldots, x_n$ considerando sólo los datos de días hábiles (el valor de n debería ser aproximadamente entre 250 y 260 datos). Calcule los log-retornos (tasa de retorno continua) dados por la transformación

$$r_t = \ln \frac{x_t}{x_{t-1}}, \quad t = 2, \ldots, n.$$

Realice el gráfico de la serie de tiempo $\{r_t : t = 2, \ldots, n\}$. Analice si se percibe que la serie presenta las siguientes propiedades:

a) Media constante igual a cero.

b) Varianza constante.

c) Distribución normal (además puede realizar el histograma).

d) Autocorrelación nula.

42. En el análisis de series de tiempo frecuentemente se hace uso de una transformación que genera variables indicadoras, conocidas en inglés como variables *dummy*. Estas son variables que sólo toman los valores cero o uno dependiendo de algún criterio. Usando la serie de log-retornos del ejercicio anterior defina

$$y_t = \begin{cases} 1 & \text{si } r_t > 0, \\ 0 & \text{si } r_t \leqslant 0. \end{cases}$$

Si las variables indicadoras $\{y_t : t = 2, \ldots, n\}$ se pudieran modelar con una distribución Bernoulli con probabilidad de éxito $1/2$, entonces su media y varianza muestral deberían ser $1/2$ y $1/4$, respectivamente. Además, la suma $\sum_{t=2}^{n} y_t$ debería ser aproximadamente $n/2$. Verifique si su serie cumple lo anterior.

Análisis de series con tendencia

43. *Regresión lineal.*

Compruebe que los coeficientes a_0 y a_1 en la regresión lineal para un conjunto de datos $x_1, \ldots, x_n$ están dados por las ecuaciones (2.10) y (2.11).

44. *Regresión lineal.*

Considere el siguiente conjunto de 8 datos observados en los tiempos $t = 1, \ldots, 8$.

$x_1 = 5$	$x_3 = 9$	$x_5 = 15$	$x_7 = 16$
$x_2 = 9$	$x_4 = 14$	$x_6 = 16$	$x_8 = 21$

a) Grafique $\{x_t : t = 1, \ldots, 8\}$.

b) Encuentre la estimación de los coeficientes a_0 y a_1 del modelo lineal $x_t = a_0 + a_1 t$ por el método de mínimos cuadrados.

c) Grafique la recta estimada sobre la misma gráfica de los datos observados.

45. *Regresión lineal para datos colineales.*

Suponga que una serie observada $x_1, \ldots, x_n$ tiene el comportamiento lineal $x_t = a_0 + a_1 t$, para $t = 1, \ldots, n$, para ciertas constantes conocidas a_0 y a_1. Compruebe que la regresión lineal produce lo esperado, es decir, $\widehat{a}_0 = a_0$ y $\widehat{a}_1 = a_1$.

46. *Tendencia por promedios móviles.*

Suponga que una serie de tiempo está compuesta por las siguientes cinco observaciones:

$$x_1 = 3,\ x_2 = 2.5,\ x_3 = 3.8,\ x_4 = 4,\ x_5 = 5$$

a) Grafique $\{x_t : t = 1, \ldots, 5\}$.

b) Suponga que estos datos siguen el modelo $X_t = m_t + Z_t$. Encuentre la tendencia estimada $\widehat{m}_t$ por promedios móviles con $q = 1$ y muestre su gráfica en el mismo plano de la gráfica del inciso anterior.

c) Encuentre y grafique el componente aleatorio. Una los puntos con líneas rectas.

47. Demuestre que el filtro lineal de orden q no altera al componente de tendencia cuando éste es lineal. Es decir, compruebe que si $m_t = a_0 + a_1 t$, entonces

$$\frac{1}{2q+1} \sum_{j=-q}^{q} m_{t-j} = m_t.$$

48. Demuestre que las ponderaciones que aparecen en la expansión (2.17) del suavizamiento exponencial suman 1.

Análisis de series con estacionalidad

49. Suponga que se tiene el siguiente modelo con tendencia lineal y componente estacional tal que $s_t = s_{t-12}$,

$$X_t = (a_0 + a_1 t) + s_t + Z_t, \quad t \in \mathbb{Z}.$$

Demuestre que al aplicar el operador $\nabla_{12} = 1 - B^{12}$ a la serie $\{X_t\}$ se obtiene una serie estacionaria.

50. Suponga que se tiene el siguiente modelo de serie de tiempo

$$X_t = (a_0 + a_1 t)\, s_t + Z_t, \quad t \in \mathbb{Z},$$

en donde el componente estacional s_t es tal que $s_t = s_{t-12}$, para cualquier $t \in \mathbb{Z}$. Observe que el componente estacional se incorpora al componente de tendencia $a_0 + a_1 t$ de forma multiplicativa.

a) ¿Se obtiene una serie estacionaria al aplicar el operador ∇_{12} a la serie $\{X_t\}$?

b) ¿Se podrá aplicar alguna otra transformación que vuelva a la serie estacionaria?

Interpretación del correlograma

51. Con ayuda de algún software estadístico, genere al azar 100 valores de la distribución normal estándar. Por ejemplo, en el software `R` se puede usar el comando `rnorm(100)`. A continuación grafique las series de tiempo y los correlogramas muestrales respectivos de los siguientes modelos, considerando que los valores de las variables aleatorias Z_t son reemplazados con la muestra generada. En todos los casos suponga que $X_0 = 0$.

a) $X_t = Z_t$, para $t = 1, \ldots, 100$.
Este es el proceso de ruido blanco.

b) $X_t = 0.5X_{t-1} + Z_t$, para $t = 1, \ldots, 100$.
Este es un modelo autorregresivo de orden 1.

c) $X_t = 0.1t - 0.25X_{t-1} + Z_t$, para $t = 1, \ldots, 100$.
Este es un modelo autorregresivo de orden 1 con tendencia lineal.

d) $X_t = 0.35Z_{t-1} + Z_t$, para $t = 1, \ldots, 100$,
(suponga que $Z_0 = 0.5$.)
Este es un modelo de promedios móviles de orden 1.

e) $X_t = -0.5X_{t-1} + 0.35Z_{t-1} + Z_t$, para $t = 1, \ldots, 100$,
(suponga que $Z_0 = 0.5$.)
Este es un modelo ARMA$(1, 1)$.

52. A partir de los gráficos del ejercicio anterior, determine en qué casos se tiene una serie estacionaria.

Capítulo 3

Algunos modelos para series de tiempo

En este capítulo se estudian algunos modelos básicos de series de tiempo univariadas que se utilizan con frecuencia en las aplicaciones. Estos modelos son procesos estocásticos a tiempo discreto $\{X_t : t \in \mathbb{Z}\}$ con espacio de estados el conjunto de números reales.

3.1. Modelo MA

El modelo de medias o promedios móviles (MA) se construye a partir de un proceso de ruido $\{Z_t : t \in \mathbb{Z}\}$ con media 0 y varianza σ^2. A continuación se establece la definición.

Definición 3.1 *Sea $q \geqslant 0$ un entero y sean $\beta_1, \ldots, \beta_q$ constantes dadas, con $\beta_q \neq 0$. Una serie de tiempo $\{X_t : t \in \mathbb{Z}\}$ sigue un modelo de medias móviles de orden q, y se le denota por las letras MA(q), si es de la forma*

$$\begin{aligned} X_t &= Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q} \qquad (3.1) \\ &= \sum_{k=0}^{q} \beta_k Z_{t-k}, \quad t \in \mathbb{Z}. \qquad (\beta_0 = 1) \end{aligned}$$

Observe que la expresión (3.1) es una combinación lineal de los q ruidos inmediatos previos al tiempo t dados por el proceso aleatorio puro. La condición $\beta_q \neq 0$ garantiza que el último sumando que aparece en (3.1) no se anula y que, efectivamente, el orden es el entero q. Los distintos procesos MA(q) están determinados por los distintos valores de los coeficientes $\beta_1, \ldots, \beta_q$. Se define $\beta_0 = 1$.

A partir de (3.1), es claro que la serie de tiempo consta de variables aleatorias independientes cuando $q = 0$, ya que en tal caso, $X_t = Z_t$, $t \in \mathbb{Z}$. En cambio, cuando $q \geqslant 1$, dos variables aleatorias de la serie que se encuentren suficientemente cercanas en el tiempo comparten valores del proceso de ruido y, por lo tanto, tienen cierto grado de dependencia. Por otro lado, dos variables aleatorias de la serie suficientemente lejanas en el tiempo una de la otra no compartirán valores del proceso de ruido y, por lo tanto, serán independientes. Estas afirmaciones se tratarán con mayor precisión a través de las fórmulas que se presentan más adelante.

Ejemplo 3.1 *En la Tabla 3.1 se muestran algunos ejemplos particulares de series de tiempo que siguen el modelo de medias móviles* (3.1) *con distintos valores del parámetro* q.

Serie	Modelo
$X_t = Z_t$	MA(0)
$X_t = Z_t - 5Z_{t-1}$	MA(1)
$X_t = Z_t - Z_{t-3} + 2Z_{t-4}$	MA(4)
$X_t = Z_t + 8Z_{t-2}$	MA(2)

Tabla 3.1: Ejemplos de modelos MA.

■

Las letras MA provienen del término en inglés *moving average*, que se puede traducir como medias móviles o promedios móviles. Este es un término técnico que se utiliza en la literatura sobre series de tiempo, aunque no

parece haber una interpretación lógica de lo que estas palabras puedan significar literalmente.

Primeras propiedades del modelo MA(q)

Usando la propiedad de independencia de las variables del proceso de ruido, se pueden calcular con facilidad las siguientes propiedades elementales que cumple todo proceso de medias móviles.

Proposición 3.1 *Sea* $\{X_t : t \in \mathbb{Z}\}$ *el proceso* MA(q) *especificado en* (3.1). *Entonces*

1. $E(X_t) = 0$.
2. $\mathrm{Var}(X_t) = \sigma^2 \sum_{k=0}^{q} \beta_k^2$.
3. $\gamma(\tau) = \mathrm{Cov}\,(X_t, X_{t+\tau}) = \begin{cases} \sigma^2 \sum_{k=0}^{q-\tau} \beta_k\, \beta_{k+\tau} & si \;\; \tau = 0, 1, \ldots, q, \\ 0 & si \;\; \tau \geqslant q+1. \end{cases}$
4. $\{X_t : t \in \mathbb{Z}\}$ *es estacionario.*

Demostración. Las identidades (1) y (2) se siguen directamente de calcular la esperanza y la varianza a partir de (3.1). Para la identidad (3), tenemos que, para $\tau = 0, 1, \ldots, q$,

$$\begin{aligned} \mathrm{Cov}\,(X_t, X_{t+\tau}) &= E(X_t \cdot X_{t+\tau}) \\ &= E[\,(\sum_{k=0}^{q} \beta_k\, Z_{t-k})(\sum_{j=0}^{q} \beta_j\, Z_{(t+\tau)-j})\,] \\ &= \sum_{k=0}^{q} \sum_{j=0}^{q} \beta_k\, \beta_j\, E(Z_{t-k} \cdot Z_{(t+\tau)-j}). \end{aligned}$$

La última esperanza es distinta de cero sólo cuando los dos subíndices

de las variables aleatorias coinciden, esto es, cuando $t - k = (t + \tau) - j$. Esto ocurre cuando $j = k + \tau$. El valor de esta esperanza es σ^2 y de aquí se obtiene el resultado anunciado. El valor máximo para j es q, de modo que ahora el valor máximo para k es $q - \tau$. Cuando $\tau \geqslant q + 1$, no existen variables Z comunes entre X_t y $X_{t+\tau}$, de modo que la esperanza del producto de estas dos últimas variables aleatorias es cero.

Finalmente, para la afirmación (4), y por los resultados anteriores, tenemos que, como la esperanza es constante y la covarianza $\text{Cov}\,(X_t, X_{t+\tau})$ no depende de t, se concluye que todo proceso MA(q) es estacionario. ∎

Es decir, el modelo MA(q) representa una serie de tiempo que carece de tendencia, pues $E(X_t) = 0$ y su varianza es constante. Observe que la varianza del proceso de ruido se denota por σ^2, mientras que la varianza de X_t se escribe explícitamente como $\text{Var}(X_t)$.

Recordemos que, por simetría, para valores negativos de τ, la covarianza $\gamma(\tau)$ coincide con $\gamma(-\tau)$, en donde $-\tau$ es positivo. Más explícitamente, la función de autocovarianza de un proceso MA(q) es, para $\tau = 0, 1, \ldots, q$,

$$\begin{aligned} \gamma(\tau) &= \text{Cov}\,(X_t, X_{t+\tau}) \\ &= \sigma^2(\beta_0\,\beta_\tau + \beta_1\,\beta_{1+\tau} + \cdots + \beta_{q-\tau}\,\beta_q). \end{aligned}$$

En esta suma aparecen $q - \tau + 1$ sumandos, de modo que conforme τ toma los valores $0, 1, \ldots, q$, el número de sumandos disminuye. En particular, al evaluar en $\tau = 0$ se tiene el máximo de sumandos, que es $q+1$, y se recupera la expresión para la varianza, esto es,

$$\gamma(0) = \sigma^2 \sum_{k=0}^{q} \beta_k^2.$$

La evaluación en $\tau = q$ involucra sólo un sumando, el cual es

$$\gamma(q) = \sigma^2(\beta_0\,\beta_q).$$

Autocorrelación del modelo MA(q)

Dividiendo $\gamma(\tau)$ entre $\gamma(0)$ se obtiene la expresión para la autocorrelación. Esto se muestra en el siguiente recuadro.

Proposición 3.2 *Sea* $\{X_t : t \in \mathbb{Z}\}$ *el proceso* MA(q) *especificado en* (3.1)*. La función de autocorrelación es*

$$\rho(\tau) = \mathrm{Corr}(X_t, X_{t+\tau}) = \begin{cases} \dfrac{\sum_{k=0}^{q-\tau} \beta_k \beta_{k+\tau}}{\sum_{k=0}^{q} \beta_k^2} & si \quad \tau = 0, 1, \ldots, q, \\ 0 & si \quad \tau \geqslant q+1. \end{cases}$$

Recordemos nuevamente que, para valores negativos de τ, la autocorrelación $\rho(\tau)$ coincide con $\rho(-\tau)$, en donde $-\tau > 0$. Por lo tanto, se usa la misma fórmula que aparece en la proposición anterior para $\tau < 0$, es decir, para cualquier valor entero de τ, $\rho(\tau) = \rho(|\tau|)$. Observe que la fórmula para $\rho(0)$ se reduce a la identidad conocida $\rho(0) = 1$.

Como se ha observado antes, cuando los subíndices de las variables X_t y $X_{t+\tau}$ se encuentran separados en más de q unidades, las variables son independientes y la correlación es cero. Esta es la razón por la que a estos procesos se les denomina procesos con memoria corta, o procesos con dependencia de corto alcance.

Ejemplo 3.2 *En la figura 3.1 se muestra el correlograma de un proceso* MA(q) *con* $q = 10$ *y coeficientes* $\beta_1 = \cdots = \beta_{10} = 1$*. Las autocorrelaciones decaen a cero de manera lineal. A partir de* $\tau = 11$ *en adelante, la autocorrelación es cero. Se ha dibujado una línea continua para enfatizar el comportamiento decreciente.* ∎

Ejemplo 3.3 *En la figura 3.2 se muestra el correlograma de un proceso* MA(q) *con* $q = 10$ *y coeficientes* $\beta_1 = \cdots = \beta_5 = -1$ *y* $\beta_6 = \cdots = \beta_{10} = 1$*. A partir de* $\tau = 11$ *en adelante, la autocorrelación es, nuevamente, cero. La línea continua sirve para visualizar mejor el comportamiento de la función.* ∎

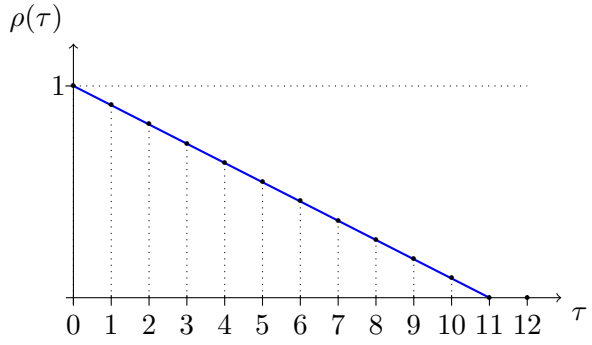

Figura 3.1: Correlograma de un proceso MA(q)
con parámetros $q = 10$ y $\beta_1 = \cdots = \beta_{10} = 1$.

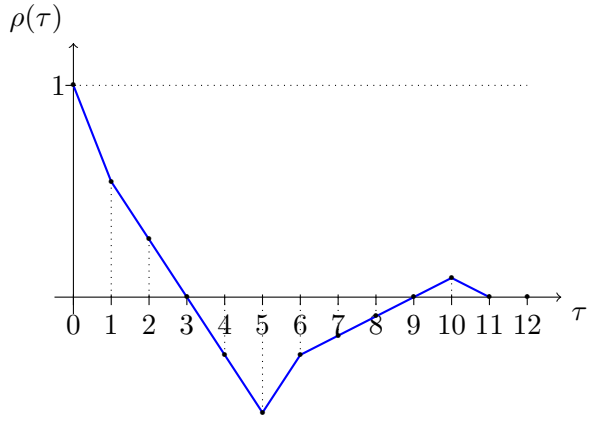

Figura 3.2: Correlograma de un proceso MA(q)
con parámetros $q = 10$, $\beta_1 = \cdots = \beta_5 = -1$ y $\beta_6 = \cdots = \beta_{10} = 1$.

El modelo MA(∞)

Es conveniente considerar también modelos de medias móviles de orden infinito. Se trata de una serie de tiempo de la forma

$$X_t = \sum_{k=0}^{\infty} \beta_k \, Z_{t-k}, \quad t \in \mathbb{Z}, \tag{3.2}$$

en donde las constantes $\beta_0 = 1, \beta_1, \beta_2, \ldots$ son números tales que la suma indicada es convergente en algún sentido. Por ejemplo, no es difícil comprobar que bajo la condición

$$\sum_{k=0}^{\infty} \beta_k^2 < \infty, \tag{3.3}$$

la serie infinita de variables aleatorias (3.2) es convergente en media cuadrática. Como antes, $\{Z_t : t \in \mathbb{Z}\}$ es un proceso de ruido de media 0 y varianza σ^2. A la serie (3.2) se le conoce también con el nombre de modelo lineal generalizado.

Suponiendo que se cumple la condición (3.3), se puede comprobar que la serie del modelo (3.2) satisface las siguientes propiedades:

- $E(X_t) = 0$.
- $\text{Var}(X_t) = \sigma^2 \sum_{k=0}^{\infty} \beta_k^2$.
- $\gamma(\tau) = \text{Cov}\,(X_t, X_{t+\tau}) = \sigma^2 \sum_{k=0}^{\infty} \beta_k\, \beta_{k+\tau}$, para $\tau = 0, 1, \ldots$
- $\{X_t : t \in \mathbb{Z}\}$ es estacionario.
- $\rho(\tau) = \text{Corr}(X_t, X_{t+\tau}) = \dfrac{\sum_{k=0}^{\infty} \beta_k\, \beta_{k+\tau}}{\sum_{k=0}^{\infty} \beta_k^2}$, para $\tau = 0, 1, \ldots$

En las siguientes secciones se presentan ejemplos de modelos MA(q) con $q = \infty$.

El modelo MA(1)

En esta sección se estudian las características del modelo particular MA(1), es decir, cuando $q = 1$. El objetivo es entender algunos conceptos y resultados de este proceso simple para después extenderlos al modelo general

cuando $q \geqslant 2$. Recordemos que el proceso MA(1) está definido por la relación

$$X_t = \beta_0\, Z_t + \beta_1\, Z_{t-1}, \quad \text{para} \;\; t \in \mathbb{Z}, \tag{3.4}$$

en donde $\beta_0 = 1$ y $\beta_1 \neq 0$ son dos constantes, y $\{Z_t : t \in \mathbb{Z}\}$ es un proceso de ruido con media 0 y varianza σ^2.

Es evidente que la media de la serie es $E(X_t) = 0$ y la varianza es $\text{Var}(X_t) = \sigma^2\,(\beta_0^2 + \beta_1^2)$, en donde $\beta_0 = 1$. La función de autocovarianza es

$$\gamma(\tau) = \text{Cov}(X_t, X_{t+\tau}) = \begin{cases} \sigma^2\,(\beta_0^2 + \beta_1^2) & \text{si} \;\; \tau = 0, \\ \sigma^2\,\beta_0\,\beta_1 & \text{si} \;\; \tau = 1, \\ 0 & \text{si} \;\; \tau \geqslant 2. \end{cases}$$

Como en el caso general en el que se tiene cualquier orden q, las funciones $E(X_t)$ y $\text{Cov}(X_t, X_{t+\tau})$ no dependen de la variable tiempo t y, por lo tanto, el proceso es estacionario.

En el caso particular cuando $Z_t \sim \text{N}(0, \sigma^2)$, se puede comprobar que $X_t \sim \text{N}(0, \sigma^2\,(\beta_0^2 + \beta_1^2))$. Al ser $\{X_t : t \in \mathbb{Z}\}$ un proceso gausiano, la estacionariedad débil implica la estacionariedad fuerte. Esta última afirmación es el contenido del ejercicio 25 en la página 36. La función de autocorrelación del modelo MA(1) es

$$\rho(\tau) = \text{Corr}(X_t, X_{t+\tau}) = \begin{cases} 1 & \text{si} \;\; \tau = 0, \\ \dfrac{\beta_0\,\beta_1}{\beta_0^2 + \beta_1^2} & \text{si} \;\; \tau = 1, \\ 0 & \text{si} \;\; \tau \geqslant 2. \end{cases}$$

Invertibilidad del modelo MA(1)

Es evidente que el proceso MA(1) depende únicamente de la constante β_1 y del proceso de ruido $\{Z_t : t \in \mathbb{Z}\}$. Recíprocamente, bajo ciertas condiciones se puede escribir el proceso $\{Z_t : t \in \mathbb{Z}\}$ en términos de la serie $\{X_t : t \in \mathbb{Z}\}$ y de algunas otras constantes. A continuación se hace un análisis llevando

a cabo una serie de substituciones sucesivas. A partir de (3.4),

$$\begin{aligned} Z_t &= \frac{1}{\beta_0} X_t - \frac{\beta_1}{\beta_0} Z_{t-1} \\ &= \frac{1}{\beta_0} X_t - \frac{\beta_1}{\beta_0} \left(\frac{1}{\beta_0} X_{t-1} - \frac{\beta_1}{\beta_0} Z_{t-2} \right) \\ &= \frac{1}{\beta_0} X_t - \frac{\beta_1}{\beta_0^2} X_{t-1} + \frac{\beta_1^2}{\beta_0^2} Z_{t-2} \\ &= \frac{1}{\beta_0} X_t - \frac{\beta_1}{\beta_0^2} X_{t-1} + \frac{\beta_1^2}{\beta_0^2} \left(\frac{1}{\beta_0} X_{t-2} - \frac{\beta_1}{\beta_0} Z_{t-3} \right) \\ &= \frac{1}{\beta_0} X_t - \frac{\beta_1}{\beta_0^2} X_{t-1} + \frac{\beta_1^2}{\beta_0^3} X_{t-2} - \frac{\beta_1^3}{\beta_0^3} Z_{t-3} \qquad (3.5) \\ &\vdots \\ &= \sum_{k=0}^{\infty} \frac{1}{\beta_0} \left(-\frac{\beta_1}{\beta_0} \right)^k X_{t-k}. \qquad (3.6) \end{aligned}$$

De esta manera, para las constantes

$$\pi_k = \frac{1}{\beta_0} \left(-\frac{\beta_1}{\beta_0} \right)^k, \quad \text{para} \quad k = 0, 1, \ldots \qquad (3.7)$$

se cumple que

$$Z_t = \sum_{k=0}^{\infty} \pi_k \, X_{t-k} = (\sum_{k=0}^{\infty} \pi_k \, B^k) \, X_t. \qquad (3.8)$$

Cuando esto es así, es decir, cuando se cumple la identidad (3.8) para ciertas constantes π_k, se dice que la serie de tiempo $\{X_t : t \in \mathbb{Z}\}$ dada en el modelo (3.4) es invertible. Observe que la fórmula (3.8) encontrada para $\{Z_t : t \in \mathbb{Z}\}$ tiene la forma de un promedio móvil de orden infinito en donde la parte del ruido está dada por las variables aleatorias X_t, las cuales no son necesariamente independientes.

Definición 3.2 *Para el modelo* MA(1) *especificado en* (3.4) *se define:*

1. *El operador de retraso* $\theta(B) = \beta_0 + \beta_1 \, B$.
2. *El polinomio* $\theta(x) = \beta_0 + \beta_1 \, x, \quad x \in \mathbb{R}$.

Haciendo uso del operador $\theta(B)$, el modelo MA(1) dado por (3.4) se puede escribir como

$$X_t = \beta_0\, Z_t + \beta_1\, Z_{t-1} = (\beta_0 + \beta_1\, B)\, Z_t = \theta(B)\, Z_t.$$

Resultará conveniente poder obtener el operador inverso de $\theta(B)$ y pasarlo del lado izquierdo en la identidad anterior. Esto lleva a la siguiente definición.

Definición 3.3 *Se dice que el modelo* MA(1) *es invertible cuando el operador de retraso asociado* $\theta(B) = \beta_0 + \beta_1\, B$ *es invertible, de modo que se puede escribir*

$$Z_t = (\theta(B))^{-1}\, X_t, \quad t \in \mathbb{Z}.$$

Como hemos visto, bajo ciertas condiciones, el operador inverso $(\theta(B))^{-1}$ está dado por

$$(\theta(B))^{-1} = \sum_{k=0}^{\infty} \pi_k\, B^k,$$

en donde las constantes $\pi_0, \pi_1, \ldots$ están indicadas en (3.7). Es posible realizar un procedimiento similar para el proceso de promedios móviles MA(q) con $q \geqslant 2$. En la siguiente sección se hace mención de algunos resultados relacionados.

Para concluir esta sección, y a manera de resumen, en el siguiente recuadro se escriben nuevamente las propiedades establecidas en las proposiciones 3.1 y 3.2 para el modelo MA(q), ahora en el caso $q = 1$.

Proposición 3.3 *Sea* $\{X_t : t \in \mathbb{Z}\}$ *el proceso* MA(1) *dado por*

$$X_t = \beta_0 Z_t + \beta_1 Z_{t-1}, \quad \textit{para} \quad t \in \mathbb{Z},$$

con $\beta_0 = 1$. *Entonces*

1. $E(X_t) = 0$.
2. $\mathrm{Var}(X_t) = \sigma^2(\beta_0^2 + \beta_1^2)$.
3. $\gamma(\tau) = \mathrm{Cov}\,(X_t, X_{t+\tau}) = \begin{cases} \sigma^2(\beta_0^2 + \beta_1^2) & si \ \ \tau = 0, \\ \sigma^2(\beta_0\beta_1) & si \ \ \tau = 1, \\ 0 & si \ \ \tau \geqslant 2. \end{cases}$
4. $\{X_t : t \in \mathbb{Z}\}$ *es estacionario.*
5. $\rho(\tau) = \mathrm{Corr}\,(X_t, X_{t+\tau}) = \begin{cases} 1 & si \ \ \tau = 0, \\ \dfrac{\beta_0\beta_1}{\beta_0^2 + \beta_1^2} & si \ \ \tau = 1, \\ 0 & si \ \ \tau \geqslant 2. \end{cases}$

Invertibilidad del modelo MA(q)

A continuación se define el concepto de invertibilidad para un proceso MA(q). Empezaremos definiendo el operador de retraso y el polinomio asociado al proceso MA(q) de manera análoga a como se hizo en el caso $q = 1$.

Definición 3.4 *Para el modelo* MA(q) *especificado en* (3.1) *se define:*

1. *El operador de retraso* $\theta(B) = 1 + \beta_1 B + \cdots + \beta_q B^q$.
2. *El polinomio* $\theta(x) = 1 + \beta_1 x + \cdots + \beta_q x^q, \quad x \in \mathbb{R}$.

Al operador $\theta(B)$ se le llama operador de promedios móviles, pues está asociado a ese modelo. Así, el modelo (3.1) se puede escribir como $X_t = \theta(B)\, Z_t$, y tenemos la misma definición de invertibilidad que en el caso $q = 1$.

Definición 3.5 *Se dice que el modelo* MA(q) *es invertible cuando el operador de retraso asociado* $\theta(B) = \beta_0 + \beta_1\, B + \cdots + \beta_q\, B^q$ *es invertible, de modo que se puede escribir como aparece abajo para ciertas constantes* $\pi_0, \pi_1, \ldots$

$$Z_t = (\theta(B))^{-1}\, X_t = (\sum_{k=0}^{\infty} \pi_k\, B^k)\, X_t, \quad t \in \mathbb{Z}.$$

Más adelante extenderemos el concepto de invertibilidad para procesos más generales que incluyen el modelo MA(q) como caso particular.

A continuación se enuncia un criterio para la invertibilidad del modelo MA(q) en términos de las raíces del polinomio asociado $\theta(x)$. Recordemos que el número real o complejo z es una raíz de $\theta(x)$ si

$$\theta(z) = \beta_0 + \beta_1\, z + \beta_2\, z^2 + \cdots + \beta_q\, z^q = 0.$$

El criterio que aparece a continuación como teorema 3.1 distingue si tales raíces $z \in \mathbb{C}$ se encuentran dentro o fuera del círculo unitario cerrado

$$D = \{\, z \in \mathbb{C} : |z| \leqslant 1 \,\},$$

en donde $|z| = \sqrt{a^2 + b^2}$ cuando $z = a + ib$.

Teorema 3.1 (Criterio para la invertibilidad del modelo MA(q)**)** *Un proceso* MA(q) *es invertible* $\Leftrightarrow$ *toda raíz* $z \in \mathbb{C}$ *del polinomio asociado* $\theta(x)$ *es tal que* $|z| > 1$.

Este resultado es un caso particular del teorema 3.1.2 que aparece en el libro de Brockwell y Davis (1987) [7]. Su demostración se puede consultar

en dicha fuente. Observe que el teorema establece una condición necesaria y suficiente para tener la invertibilidad en un proceso MA(q). Además, es claro que esta condición también se puede expresar solicitando que $\theta(x) \neq 0$ para cualquier $x \in \mathbb{C}$ tal que $|x| \leqslant 1$, es decir, ningún número complejo dentro del círculo unitario es solución de la ecuación $\theta(x) = 0$.

De esta manera, determinar si un modelo MA(q) es invertible usando el teorema 3.1 requiere conocer la norma de las q raíces del polinomio $\theta(x)$. A continuación se presentan algunos ejemplos sencillos en donde el polinomio $\theta(x)$ es lineal y, por lo tanto, es inmediato encontrar sus raíces.

Ejemplo 3.4 *Sea* $\{X_t : t \in \mathbb{Z}\}$ *un proceso* MA(1) *con* $\beta_0 = 1$, *es decir,*

$$X_t = Z_t + \beta_1 Z_{t-1}, \quad t \in \mathbb{Z}.$$

El polinomio asociado es $\theta(x) = 1 + \beta_1 x$, *cuya única raíz es* $x = -1/\beta_1$. *Este número tiene norma mayor a* 1 *si* $|\beta_1| < 1$. *Así, para valores de* $\beta_1 \neq 0$ *que satisfagan la condición indicada, este proceso* MA(1) *es invertible.* ▪

El modelo MA[Q] estacional

Este es un modelo de promedios móviles que incorpora un patrón estacional. El número entero $s \geqslant 2$ denotará la longitud del periodo de la estación.

Definición 3.6 *Sean* $Q \geqslant 0$ *y* $s \geqslant 2$ *dos enteros. Sean* $\beta'_0 = 1$ *y* $, \beta'_1, \ldots, \beta'_Q$ *constantes dadas, con* $\beta'_Q \neq 0$. *Una serie de tiempo* $\{X_t : t \in \mathbb{Z}\}$ *sigue un modelo* MA[Q] *estacional de periodo* s, *si es de la forma*

$$\begin{aligned} X_t &= Z_t + \beta'_1 Z_{t-s} + \beta'_2 Z_{t-2s} + \cdots + \beta'_Q Z_{t-Qs}, \qquad (3.9) \\ &= \sum_{k=0}^{Q} \beta'_k Z_{t-ks}, \quad t \in \mathbb{Z}. \end{aligned}$$

De esta forma, la variable X_t depende del proceso de ruido $\{Z_t : t \in \mathbb{Z}\}$ en los tiempos $t, t-s, t-2s, \ldots, t-Qs$, es decir, depende del tiempo t y los Q tiempos anteriores con rezagos múltiplos de s.

Es claro que el proceso (3.9) es un caso particular de un modelo MA(q) con $q = Qs$, en donde los únicos coeficientes distintos de cero son aquellos con subíndice con rezago $0, s, 2s, \ldots, Qs$. Por lo tanto, todos los conceptos y resultados vistos antes para procesos autorregresivos se pueden aplicar también para el proceso (3.9). Por ejemplo, el operador de retraso asociado se denota por $\Theta(B^s)$ y es

$$\Theta(B^s) = 1 + \beta_1' B^s + \beta_2' B^{2s} + \cdots + \beta_Q' B^{Qs}.$$

La expresión (3.9) se puede escribir entonces como aparece abajo. En particular, sabemos que este proceso es estacionario.

$$X_t = \Theta(B^s) Z_t, \quad t \in \mathbb{Z}. \tag{3.10}$$

De manera un poco más general, se puede considerar el proceso de promedios móviles estacional

$$X_t = \Theta(B^s)\theta(B) Z_t, \quad t \in \mathbb{Z}, \tag{3.11}$$

en donde $\theta(B)$ es el operador usual de orden q considerado antes en la definición 3.4 que aparece en la página 87. El término $\theta(B)Z_t$ establece la dependencia de X_t respecto del vector de variables $(Z_t, \ldots, Z_{t-q})$ y el operador $\Theta(B^s)$ indica la dependencia de estas mismas variables, pero con rezagos tales que se encuentren en cada una de las Q estaciones previas. Por ejemplo, para un periodo estacional previo, el vector de variables es $(Z_{t-s}, \ldots, Z_{t-q-s})$, para 2 periodos estacionales previos el vector de variables es $(Z_{t-2s}, \ldots, Z_{t-q-2s})$, etc. Esto se puede expresar de la siguiente forma

$$X_t = \theta(B)Z_t + \beta_1'\theta(B)Z_{t-s} + \beta_2'\theta(B)Z_{t-2s} + \cdots + \beta_Q'\theta(B)Z_{t-Qs}. \tag{3.12}$$

Observe que los operadores $\theta(B)$ y $\Theta(B^s)$ conmutan, de modo que su producto se puede escribir en cualquier orden.

El modelo MA(q) con media no cero

En ocasiones es conveniente considerar modelos MA(q) con media constante pero distinta de cero. Esto se establece en la siguiente definición.

Definición 3.7 *Sea μ una constante. Se dice que la serie $\{X_t : t \in \mathbb{Z}\}$ sigue un modelo* MA(q) *con media μ si el proceso centrado $\{X_t - \mu : t \in \mathbb{Z}\}$ es un proceso* MA(q)*, es decir,*

$$X_t - \mu = Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q}, \quad t \in \mathbb{Z}. \tag{3.13}$$

De la ecuación (3.13) es evidente que la serie $\{X_t - \mu : t \in \mathbb{Z}\}$ es la que sigue el modelo MA(q). Puesto que una serie que sigue el modelo MA(q) tiene media cero, se cumple que $E(X_t) = \mu$, para cualquier $t \in \mathbb{Z}$.

Concluimos esta sección señalando que todo proceso MA(q) tiene una única función de autocorrelación $\rho(\tau)$, calculada a partir de su definición. Por otro lado, dada una función de autocorrelación que sabemos que proviene de un proceso MA(q), esta función no queda unívocamente determinado, ya que pueden existir dos procesos MA(q) con la misma función $\rho(\tau)$. Como ejemplos, en los ejercicios 67 y 68 se muestran dos series distintas del modelo MA(1) que tienen la misma función de correlación.

Sin embargo, se puede comprobar que bajo la hipótesis de invertibilidad, se puede garantizar cierta unicidad del modelo MA(q) a partir de su función de autocorrelación. Por ejemplo, en el ejercicio 69 aparece una guía para demostrar tal afirmación en el caso $q = 1$. En general, la condición de invertibilidad garantiza que existe una relación uno a uno entre el espacio de los procesos MA(q) invertibles y sus funciones de autocorrelación cuando el proceso de ruido $\{Z_t : t \in \mathbb{Z}\}$ es el ruido blanco WN$(0, \sigma^2)$. La unicidad es en el sentido de que las variables aleatorias tienen la misma distribución.

3.2. Modelo AR

Para definir este modelo se toma nuevamente como base un proceso de ruido $\{Z_t : t \in \mathbb{Z}\}$ con media 0 y varianza σ^2. Su definición se presenta a continuación.

Definición 3.8 *Sea $p \geqslant 1$ un entero y sean $\alpha_1, \ldots, \alpha_p$ constantes dadas tal que $\alpha_p \neq 0$. Se dice que una serie de tiempo $\{X_t : t \in \mathbb{Z}\}$ sigue un modelo autorregresivo de orden p, y se le denota por las letras AR(p), si es de la forma*

$$\begin{aligned} X_t &= \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + Z_t \\ &= \sum_{k=1}^{p} \alpha_k X_{t-k} + Z_t, \quad t \in \mathbb{Z}. \end{aligned} \tag{3.14}$$

La expresión (3.14) es una combinación lineal de los p valores del proceso inmediatos previos al tiempo t y un ruido adicional dado por la variable Z_t. Las ponderaciones son las constantes $\alpha_1, \ldots, \alpha_p$, las cuales no necesariamente suman 1, pueden ser positivas, negativas, y algunas de ellas pueden ser cero. Este proceso se puede definir para $t = p+1, p+2, \ldots$ cuando los valores iniciales $x_1, \ldots, x_p$ se suponen conocidos. Las letras AR provienen del término en inglés *autoregressive*.

Ejemplo 3.5 *En la Tabla 3.2 se muestran algunos ejemplos particulares de series de tiempo que siguen el modelo autorregresivo* (3.14) *con distintos valores del parámetro p.* ▪

Serie	Modelo
$X_t = X_{t-1} + Z_t$	AR(1)
$X_t = 2\, X_{t-4} + Z_t$	AR(4)
$X_t = X_{t-1} - X_{t-3} + Z_t$	AR(3)
$X_t = -2\, X_{t-1} + 5\, X_{t-2} + Z_t$	AR(2)

Tabla 3.2: Ejemplos de modelos AR.

En esta sección se estudian algunas características del modelo autorregresivo. Se presentan primero los resultados para el caso más simple AR(1), y después se regresa al caso general AR(p).

El modelo AR(1)

Este es el modelo autorregresivo más sencillo, está dado por la relación

$$X_t = \alpha_1 X_{t-1} + Z_t, \quad t \in \mathbb{Z}, \tag{3.15}$$

en donde $\alpha_1 \neq 0$ es un número real. Observe que este modelo depende sólo del parámetro α_1 y de la varianza σ^2 del proceso de ruido, de modo que todas las características numéricas del modelo se pueden escribir en términos de estas dos cantidades. Usando el operador de retraso B, la identidad (3.15) se puede escribir como

$$(1 - \alpha_1 B) X_t = Z_t, \quad t \in \mathbb{Z}.$$

El operador que aparece en el lado izquierdo es el operador de retraso asociado a este modelo. En el siguiente recuadro se detalla esta definición.

Definición 3.9 *Para el proceso* AR(1) *especificado en* (3.15) *se define:*

1. *El operador de retraso* $\phi(B) = 1 - \alpha_1 B$.
2. *El polinomio* $\phi(x) = 1 - \alpha_1 x, \quad x \in \mathbb{R}$.

Por lo tanto, la identidad (3.15) se puede escribir como $\phi(B) X_t = Z_t$, para $t \in \mathbb{Z}$. Veremos que es de utilidad poder calcular el operador inverso de $\phi(B)$. A continuacióonse explican dos procedimientos a través de los cuales se puede encontrar este operador inverso.

- Recordemos que, para cualquier número real x tal que $-1 < x < 1$, se cumple la identidad

$$\frac{1}{1-x} = 1 + x + x^2 + \cdots$$

Tratando al término $\alpha_1 B$ como si fuera un número real en el intervalo

$(-1, 1)$, la expansión anterior sugiere el siguiente desarrollo:

$$\begin{aligned}(\phi(B))^{-1} &= \frac{1}{1-\alpha_1 B} \\ &= 1 + \alpha_1 B + \alpha_1^2 B^2 + \cdots \\ &= \sum_{k=0}^{\infty} \alpha_1^k B^k.\end{aligned}$$

- Alternativamente, se puede proponer la expansión $(\phi(B))^{-1} = \psi_0 + \psi_1 B + \psi_2 B^2 + \cdots$ para ciertos coeficientes desconocidos $\psi_0, \psi_1, \ldots$ y plantear la ecuación

$$1 = (\phi(B))^{-1}\phi(B) = (\psi_0 + \psi_1 B + \psi_2 B^2 + \cdots)(1 - \alpha_1 B).$$

  Desarrollando el producto que aparece en la última expresión, e igualando los coeficientes de uno y otro lado de esta identidad, se encuentra nuevamente que los coeficientes ψ_k que aparecen en $(\phi(B))^{-1}$ son α_1^k, para $k = 0, 1, \ldots$

Por lo tanto, tenemos la siguiente representación de un modelo AR(1):

$$\begin{aligned}X_t &= (\phi(B))^{-1} Z_t \\ &= (\sum_{k=0}^{\infty} \alpha_1^k B^k) Z_t \\ &= \sum_{k=0}^{\infty} \alpha_1^k Z_{t-k}, \quad t \in \mathbb{Z}, \end{aligned} \tag{3.16}$$

en donde, recordemos, esta serie infinita de variables aleatorias es convergente en media cuadrática cuando $-1 < \alpha_1 < 1$. De esta manera, encontramos nuevamente en la expresión (3.16) a un modelo de promedios móviles de orden infinito. En la demostración de la siguiente proposición se muestra la utilidad de la representación (3.16) para encontrar directamente algunas propiedades del modelo AR(1).

Proposición 3.4 *Sea* $\{X_t : t \in \mathbb{Z}\}$ *un proceso* AR(1) *especificado en* (3.15)*, en donde* $-1 < \alpha_1 < 1$*. Entonces*

1. $E(X_t) = 0$.
2. $\text{Var}(X_t) = \dfrac{\sigma^2}{1-\alpha_1^2}$.
3. $\gamma(\tau) = \text{Cov}\,(X_t, X_{t+\tau}) = \sigma^2 \cdot \dfrac{\alpha_1^{|\tau|}}{1-\alpha_1^2}, \quad \tau \in \mathbb{Z}$.
4. $\{X_t : t \in \mathbb{Z}\}$ *es estacionario.*
5. $\rho(\tau) = \rho(X_t, X_{t+\tau}) = \alpha_1^{|\tau|}, \quad \tau \in \mathbb{Z}$.

Demostración.

(1)(2) Estos dos primeros resultados son una consecuencia de la representación (3.16). Aplicando esperanza y varianza en (3.16), e intercambiando estas operaciones con el límite, se llega a las expresiones anunciadas.

(3) Supongamos $\tau \geqslant 0$. Entonces

$$\begin{aligned}
\text{Cov}\,(X_t, X_{t+\tau}) &= E(X_t \cdot X_{t+\tau}) \\
&= E(\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \alpha_1^{j+k}\, Z_{t-j} \cdot Z_{t+\tau-k}) \\
&= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \alpha_1^{j+k}\, E(Z_{t-j} \cdot Z_{t+\tau-k}) \\
&= \sigma^2 \sum_{k=\tau}^{\infty} \alpha_1^{2k-\tau} \qquad (3.17) \\
&= \frac{\sigma^2}{\alpha_1^{\tau}} \sum_{k=\tau}^{\infty} (\alpha_1^2)^k.
\end{aligned}$$

Utilizando la fórmula para una suma geométrica,

$$\begin{aligned}\operatorname{Cov}(X_t, X_{t+\tau}) &= \frac{\sigma^2}{\alpha_1^\tau} \cdot \frac{(\alpha_1^2)^\tau}{1-\alpha_1^2} \\ &= \sigma^2 \cdot \frac{\alpha_1^\tau}{1-\alpha_1^2}.\end{aligned}$$

La doble suma se reduce a sólo una suma, pues la esperanza indicada se anula cuando los subíndices de las variables Z son distintos. Los subíndices son iguales cuando $t-j = t+\tau-k$, lo que lleva a la condición $j = k-\tau$. Como $j \geqslant 0$, el subíndice k debe satisfacer la desigualdad $k \geqslant \tau$.

Para $\tau < 0$, el cálculo es el mismo, pero ahora la condición $k \geqslant \tau$ siempre se cumple y la suma (3.17) se calcula para $k \geqslant 0$, es decir,

$$\begin{aligned}\operatorname{Cov}(X_t, X_{t+\tau}) &= \frac{\sigma^2}{\alpha_1^\tau} \sum_{k=0}^{\infty} (\alpha_1^2)^k \\ &= \frac{\sigma^2}{\alpha_1^\tau} \cdot \frac{1}{1-\alpha_1^2} \\ &= \sigma^2 \cdot \frac{\alpha_1^{-\tau}}{1-\alpha_1^2}.\end{aligned}$$

De esta manera se obtiene una fórmula para la covarianza. Se utiliza la expresión $|\tau|$ para reducir en un solo caso los valores positivos y negativos de τ. Observe que se recupera la expresión para la varianza cuando $\tau = 0$.

(4) Puesto que la esperanza y la covarianza no dependen de la variable tiempo t, concluimos que todo proceso AR(1), con $-1 < \alpha_1 < 1$, es estacionario.

(5) Por las fórmulas anteriores, para $\tau \in \mathbb{Z}$,

$$\rho(\tau) = \operatorname{Corr}(X_t, X_{t+\tau}) = \frac{\operatorname{Cov}(X_t, X_{t+\tau})}{\sqrt{\operatorname{Var}(X_t)\operatorname{Var}(X_{t+\tau})}} = \alpha_1^{|\tau|}.$$

■

Por lo recién demostrado, el modelo AR(1) especificado en la proposición anterior representa una serie de tiempo estacionaria y sin tendencia pues $E(X_t) = 0$ para cualquier $t \in \mathbb{Z}$. El correlograma general de un proceso AR(1) presenta dos comportamientos distintos: cuando $\alpha_1 > 0$ y cuando $\alpha_1 < 0$. Estas dos posibilidades se muestran en los siguientes ejemplos gráficos.

Ejemplo 3.6 *En la figura 3.3 se muestra el correlograma de un proceso* AR(1) *con* $\alpha_1 = 0.85$. *La dependencia lineal de las variables inicia en el valor* 1, *siempre es positiva y decrece geométricamente conforme el desfase o rezago* τ *crece. Esto es evidente a partir de examinar la expresión* $\rho(\tau) = \alpha_1^\tau$, *para* $\tau \geqslant 0$.

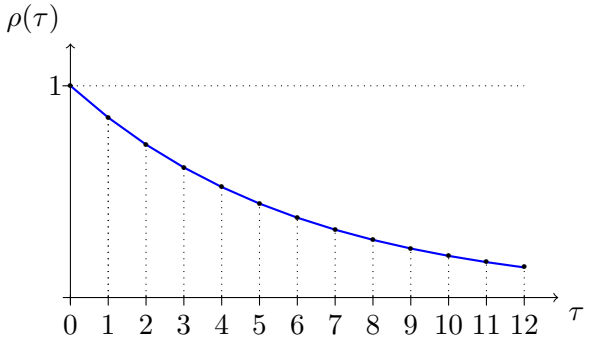

Figura 3.3: Correlograma de un proceso AR(1) con $\alpha_1 = 0.85$.

■

Ejemplo 3.7 *En la figura 3.4 se muestra el correlograma de un proceso* AR(1) *con* $\alpha_1 = -0.7$. *La dependencia lineal de las variables inicia en el valor* 1, *pero como* $\alpha_1 < 0$, *las potencias* $\rho(\tau) = \alpha_1^\tau$, *para* $\tau \geqslant 0$, *presentan oscilaciones de valores positivos y negativos. Se observa que el comportamiento decreciente siempre se mantiene.*

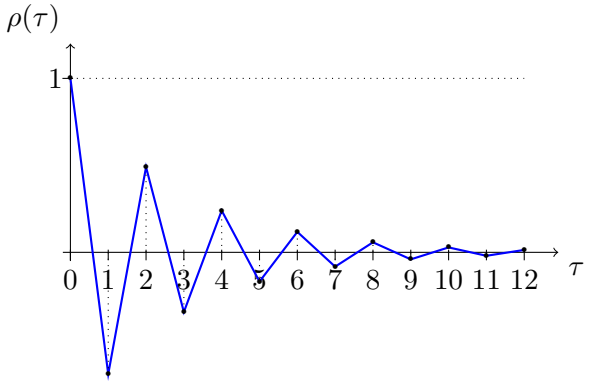

Figura 3.4: Correlograma de un proceso AR(1) con $\alpha_1 = -0.7$.

■

En cualquier caso, como $-1 < \alpha_1 < 1$, se verifica el decaimiento exponencial de la función $\rho(\tau) = \alpha_1^\tau$, para $\tau \geqslant 0$. Cuando α_1 es negativa, se presentan oscilaciones por el cambio de signo. Las gráficas presentadas sólo muestran las correlaciones para $\tau \geqslant 0$; ambas gráficas se pueden reflejar hacia la parte negativa del eje a manera de espejo.

Si un correlograma muestral presenta alguna similitud con las características de las figuras 3.3 ó 3.4, entonces podría ser factible ajustar el modelo AR(1) a los datos observados.

El modelo AR(2)

Este modelo autorregresivo corresponde a la serie de tiempo

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + Z_t, \quad t \in \mathbb{Z}, \tag{3.18}$$

en donde α_1 y α_2 son dos constantes con $\alpha_2 \neq 0$. Se definen ahora el operador de retraso y el polinomio asociados como

$$\begin{aligned} \phi(B) &= 1 - \alpha_1 B - \alpha_2 B^2, \\ \phi(x) &= 1 - \alpha_1 x - \alpha_2 x^2, \quad x \in \mathbb{R}. \end{aligned}$$

El análisis no es tan sencillo como en el caso AR(1). Bajo condiciones sobre los coeficientes se puede comprobar que $E(X_t) = 0$ y que el proceso es estacionario. Las expresiones que surgen en el cálculo de las covarianzas son complicadas y es mejor expresarlas de forma recursiva. Encontraremos estas fórmulas como un caso particular de los resultados generales que se presentan en la siguiente sección. Véanse los ejemplos 3.13 y 3.15.

El modelo AR(p)

A continuación se estudia el caso general $p \geqslant 1$. Usando el operador de retraso B, el modelo AR(p) especificado en (3.14) se puede escribir como

$$(1 - \sum_{k=1}^{p} \alpha_k B^k) X_t = Z_t, \quad t \in \mathbb{Z}. \tag{3.19}$$

Como antes, el término que aparece entre paréntesis es un operador de retraso y se define a continuación junto con su polinomio asociado.

Definición 3.10 *Para el modelo* AR(p) *especificado en* (3.19) *se define:*

1. *El operador de retraso* $\phi(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_p B^p$.
2. *El polinomio* $\phi(x) = 1 - \alpha_1 x - \alpha_2 x^2 - \cdots - \alpha_p x^p, \quad x \in \mathbb{R}$.

Al operador $\phi(B)$ se le llama operador autorregresivo, pues está asociado a ese modelo. De esta manera, la identidad (3.19) se puede escribir como

$$\phi(B) X_t = Z_t, \quad t \in \mathbb{Z}. \tag{3.20}$$

En el caso cuando $\phi(B)$ sea invertible, podremos escribir el modelo AR(p)

de la manera más directa

$$\begin{aligned} X_t &= (\phi(B))^{-1} Z_t \\ &= \Big(\sum_{k=0}^{\infty} \pi_k B^k\Big) Z_k \\ &= \sum_{k=0}^{\infty} \pi_k Z_{t-k}, \end{aligned} \tag{3.21}$$

para ciertas constantes $\pi_0, \pi_1, \ldots$ tales que la suma infinita es convergente. En particular, la condición que aparece abajo garantiza la convergencia de (3.21) en media cuadrática,

$$\sum_{k=0}^{\infty} \pi_k^2 < \infty. \tag{3.22}$$

Cuando la representación (3.21) es válida, se dice que el proceso AR(p) es causal. Observe que esta representación es nuevamente la de un modelo de medias móviles de orden infinito e implica que el proceso es estacionario. Esto lleva a la siguiente definición.

Definición 3.11 *El proceso* AR(p) *especificado en* (3.14) *es causal cuando el operador asociado* $\phi(B)$ *es invertible.*

Resulta nuevamente que la invertibilidad de $\phi(B)$ depende de la norma de las raíces del polinomio asociado $\phi(x)$. Esto es similar a lo estudiado antes sobre la invertibilidad de $\theta(B)$ en el caso de procesos de promedios móviles. En este caso tenemos que el número real o complejo z es una raíz de $\phi(x)$ si

$$\phi(z) = 1 - \alpha_1 z - \alpha_2 z^2 - \cdots - \alpha_p z^p = 0,$$

y el criterio que se enuncia a continuación como teorema 3.2 distingue nuevamente si tales raíces se encuentran dentro o fuera del círculo unitario cerrado

$$D = \{ z \in \mathbb{C} : |z| \leqslant 1 \},$$

en donde $|z| = \sqrt{a^2 + b^2}$ cuando $z = a + ib$.

Teorema 3.2 (Criterio para la causalidad) *Un proceso* AR(p) *es causal* $\Leftrightarrow$ *toda raíz* $z \in \mathbb{C}$ *del polinomio asociado* $\phi(x)$ *es tal que* $|z| > 1$.

La demostración de este resultado se puede consultar en Brockwell y Davis (1987) [7]. El teorema aquí enunciado es un caso particular de un resultado más general que aparece en el libro citado como teorema 3.1.1 . Observe que el teorema aquí enunciado establece una condición necesaria y suficiente para tener la causalidad en un proceso AR(p). Es evidente que esta condición también se puede expresar solicitando que $\phi(x) \neq 0$ para cualquier $x \in \mathbb{C}$ tal que $|x| \leqslant 1$, es decir, ningún número complejo dentro del círculo unitario es solución de la ecuación $\phi(x) = 0$.

El siguiente ejemplo muestra el caso AR(1), el cual ya fue estudiado. Haciendo uso del teorema 3.2, corroboraremos que el proceso AR(1) es causal cuando $-1 < \alpha_1 < 1$.

Ejemplo 3.8 *Sea* $\{X_t : t \in \mathbb{Z}\}$ *un proceso* AR(1)*, es decir,*

$$X_t = \alpha_1 X_{t-1} + Z_t, \quad t \in \mathbb{Z}.$$

El polinomio asociado es $\phi(x) = 1 - \alpha_1 x$*, cuya única raíz es* $z = 1/\alpha_1$*. Esta raíz tiene una norma mayor a* 1 *cuando* $|\alpha_1| < 1$*. Por el teorema 3.2, encontramos nuevamente que el proceso* AR(1) *es causal cuando* $-1 < \alpha_1 < 1$*.* ∎

Ejemplo 3.9 *Sea* $\{X_t : t \in \mathbb{Z}\}$ *un proceso* AR(2)*, es decir,*

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + Z_t, \quad t \in \mathbb{Z}.$$

El polinomio asociado es $\phi(x) = 1 - \alpha_1 x - \alpha_2 x^2$*, cuyas dos raíces son*

$$z = \frac{\alpha_1 \pm \sqrt{\alpha_1^2 + 4\alpha_2}}{-2\alpha_2}.$$

Así, el proceso AR(2) *es causal cuando los coeficientes* α_1 *y* α_2 *son tales que* $|z| > 1$*.* ∎

Regresando al caso general, vamos a suponer que el operador $\phi(B)$ es invertible, y a partir de esta hipótesis encontraremos una manera de determinar las constantes $\pi_0, \pi_1, \ldots$ de la expansión (3.21).

Proposición 3.5 *Sea $\phi(B)$ el operador de retraso asociado a un proceso* AR(p) *que es causal. Entonces*

$$(\phi(B))^{-1} = \sum_{k=0}^{\infty} \pi_k B^k, \tag{3.23}$$

en donde los coeficientes $\pi_0, \pi_1, \ldots$ están dados por

$$\begin{aligned} \pi_0 &= 1, \\ \pi_k &= \alpha_1\pi_{k-1} + \alpha_2\pi_{k-2} + \cdots + \alpha_p\pi_{k-p}, \quad \textit{para} \quad k \geqslant 1. \end{aligned} \tag{3.24}$$

Se define $\pi_k = 0$ para $k < 0$.

Demostración. Como el proceso AR(p) en cuestión es causal, el operador $\phi(B)$ es invertible. Así, el operador inverso debe cumplir que

$$\begin{aligned} 1 &= (\phi(B))^{-1}\phi(B) \\ &= (\pi_0 + \pi_1 B + \pi_2 B^2 + \cdots)(1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_p B^p) \\ &= \pi_0 + (\pi_1 - \alpha_1\pi_0)B + (\pi_2 - \alpha_1\pi_1 - \alpha_2\pi_0)B^2 + \cdots \\ &= \pi_0 + \sum_{k=1}^{\infty} [\pi_k - \sum_{i=1}^{p} \alpha_i \pi_{k-i}] B^k, \end{aligned}$$

en donde $\pi_k = 0$ para $k < 0$. Esta es una igualdad de polinomios en B y sus coeficientes deben coincidir. De aquí se obtiene que $\pi_0 = 1$ y que para $k \geqslant 1$, la igualación de los coeficientes de B^k, de uno y otro lado, produce la ecuación

$$0 = \pi_k - \sum_{i=1}^{p} \alpha_i \pi_{k-i}, \quad k \geqslant 1.$$

Esta es la relación (3.24) anunciada, en donde, recordemos, se define $\pi_k = 0$ para valores enteros de k tales que $k < 0$. ∎

En consecuencia, este resultado garantiza la validez de la representación (3.21) como un modelo de promedios móviles de orden infinito. En los siguientes ejemplos se muestran los valores de los coeficientes π_k para los dos primeros modelos autorregresivos.

Ejemplo 3.10 *Para el modelo* AR(1)*, los coeficientes* π_k *dados por la ecuación recién demostrada* (3.24) *son:*

$$\begin{aligned} \pi_0 &= 1, \\ \pi_k &= \alpha_1 \pi_{k-1}, \quad \textit{para} \quad k \geqslant 1. \end{aligned}$$

De aquí se obtiene que $\pi_k = \alpha_1^k$*, para* $k \geqslant 0$*. Se confirma lo que se había encontrado para este modelo.* ∎

Ejemplo 3.11 *Para el modelo* AR(2)*, los coeficientes* π_k *que aparecen en la ecuación recursiva* (3.24) *se calculan de manera sucesiva:*

$$\begin{aligned} \pi_0 &= 1, \\ \pi_1 &= \alpha_1 \pi_0 = \alpha_1, \\ \pi_2 &= \alpha_1 \pi_1 + \alpha_2 \pi_0 = \alpha_1^2 + \alpha_2, \\ \pi_3 &= \alpha_1 \pi_2 + \alpha_2 \pi_1 = \alpha_1(\alpha_1^2 + \alpha_2) + \alpha_1\alpha_2, \\ &\vdots \end{aligned}$$

En este caso no se cuenta con una fórmula cerrada para el término general π_k*.* ∎

Con ayuda de la expansión (3.21) ahora se pueden demostrar las siguientes primeras propiedades generales de todo modelo autorregresivo.

Proposición 3.6 *Sea* $\{X_t : t \in \mathbb{Z}\}$ *un proceso* AR(p) *que es causal, tiene la representación* (3.21) *y se cumple la condición* (3.22). *Entonces*

1. $E(X_t) = 0$.
2. $\mathrm{Var}(X_t) = \sigma^2 \sum_{k=0}^{\infty} \pi_k^2$.
3. $\mathrm{Cov}(X_t, X_{t+\tau}) = \sigma^2 \sum_{k=0}^{\infty} \pi_k \pi_{k+\tau}$.
4. $\{X_t : t \in \mathbb{Z}\}$ *es estacionario.*

Demostración. Las primeras dos afirmaciones se siguen de manera inmediata de (3.21). Para obtener la varianza se hace uso de la condición (3.22), la cual garantiza que la suma infinita que aparece en el enunciado es convergente en media cuadrática. El cálculo de la covarianza es el siguiente

$$\begin{aligned} \mathrm{Cov}(X_t, X_{t+\tau}) &= \mathrm{Cov}(\sum_{k=0}^{\infty} \pi_k \, Z_{t-k}, \sum_{k=0}^{\infty} \pi_k \, Z_{t+\tau-k}) \\ &= \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \pi_k \pi_j \, \mathrm{Cov}(Z_{t-k}, Z_{t+\tau-j}). \end{aligned}$$

La última covarianza es distinta de cero, e igual a σ^2, cuando los subíndices k y j coinciden, es decir, cuando $t - k = t + \tau - j$, o bien $j = \tau + k$. Esto reduce la doble suma a una sola suma, que es la fórmula que aparece en el enunciado. Finalmente, puesto que $E(X_t)$ y $\mathrm{Cov}(X_t, X_{t+\tau})$ no dependen del tiempo t, se sigue que el proceso es estacionario. ∎

Observe que, en general, el proceso AR(p) no es estacionario, pero loo será cuando sea causal, es decir, cuando los coeficientes $\alpha_1, \ldots, \alpha_p$ senn tales que el polinomio $\phi(x) = 1 - \alpha_1 x - \alpha_2 x^2 - \cdots - \alpha_p x^p$ tenga todas sus raíces fuera del círculo unitario, de modo que $\phi(B)$ es invertible y la expansión (3.21) es válida. El proceso AR(p) es un modelo de serie de tiempo sin tendencia, pues $E(X_t) = 0$ para cualquier $t \in \mathbb{Z}$.

Suponiendo que un proceso AR(p) es estacionario, encontraremos ahora fórmulas recursivas para las funciones de autocovarianza y de autocorrelación. En este caso, resulta conveniente calcular primero la autocorrelación y después la autocovarianza. Debe observarse que la autocorrelación no se obtiene de manera explícita, sino que se determina a través de ciertas ecuaciones recursivas llamadas ecuaciones de Yule-Walker[1].

Proposición 3.7 (Ecuaciones de Yule-Walker) *Sea* $\{X_t : t \in \mathbb{Z}\}$ *un proceso* AR(p) *que es causal. La función de autocorrelación está dada de manera recursiva por*

$$\rho(\tau) = \text{Corr}(X_t, X_{t+\tau}) = \begin{cases} 1 & si \quad \tau = 0, \\ \sum_{k=1}^{p} \alpha_k \, \rho(\tau - k) & si \quad \tau \geqslant 1. \end{cases} \tag{3.25}$$

Demostración. Por definición, $\rho(0) = 1$. Veamos el caso $\tau \geqslant 1$. Recordemos que

$$X_t = \alpha_1 \, X_{t-1} + \cdots + \alpha_p \, X_{t-p} + Z_t, \quad t \in \mathbb{Z}.$$

Multiplicando esta igualdad por $X_{t-\tau}$ y tomando esperanza se obtiene

$$\gamma(\tau) = \alpha_1 \, \gamma(\tau - 1) + \cdots + \alpha_p \, \gamma(\tau - p) + E(X_{t-\tau} Z_t),$$

en donde la última esperanza es cero, pues para $\tau \geqslant 1$, las variables $X_{t-\tau}$ y Z_t son independientes. Dividiendo entre $\gamma(0)$ se obtiene el resultado anunciado. ∎

Como la autocorrelación es una función par, es decir $\rho(\tau) = \rho(-\tau)$, tenemos que, más explícitamente, la función de autocorrelación se encuentra de la forma siguiente: se resuelve primero el sistema de p ecuaciones

[1] George Udny Yule (1871–1951) estadístico escocés.
Gilbert Walker (1868–1958) físico y estadístico inglés.

$$\begin{aligned}
\rho(1) &= \alpha_1\,\rho(0) + \alpha_2\,\rho(1) + \cdots + \alpha_p\,\rho(p-1)\\
\rho(2) &= \alpha_1\,\rho(1) + \alpha_2\,\rho(0) + \cdots + \alpha_p\,\rho(p-2)\\
\rho(3) &= \alpha_1\,\rho(2) + \alpha_2\,\rho(1) + \cdots + \alpha_p\,\rho(p-3)\\
&\vdots\\
\rho(p) &= \alpha_1\,\rho(p-1) + \alpha_2\,\rho(p-2) + \cdots + \alpha_p\,\rho(0),
\end{aligned}$$

en donde $\rho(1), \ldots, \rho(p)$ son p incógnitas. Recordemos que $\rho(0) = 1$. A partir de estos primeros p valores, se calculan los subsiguientes valores usando la fórmula recursiva (3.25), es decir,

$$\begin{aligned}
\rho(p+1) &= \alpha_1\,\rho(p) + \cdots + \alpha_p\,\rho(1)\\
\rho(p+2) &= \alpha_1\,\rho(p+1) + \cdots + \alpha_p\,\rho(2)\\
&\vdots
\end{aligned}$$

Ejemplo 3.12 *Las ecuaciones de Yule-Walker para el proceso* AR(1) *dado por* $X_t = \alpha_1 X_{t-1} + Z_t$, $t \in \mathbb{Z}$, *son*

$$\begin{aligned}
\rho(0) &= 1,\\
\rho(\tau) &= \alpha_1\,\rho(\tau-1), \quad \tau \geqslant 1.
\end{aligned}$$

De aquí se obtiene de manera inmediata que $\rho(\tau) = \alpha_1^\tau$, *para* $\tau \geqslant 0$. *Esto corrobora lo encontrado en la proposición 3.4.* ∎

Ejemplo 3.13 *Las ecuaciones de Yule-Walker para el proceso* AR(2) *dado por* $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + Z_t$, $t \in \mathbb{Z}$, *son*

$$\begin{aligned}
\rho(1) &= \alpha_1 + \alpha_2\,\rho(1),\\
\rho(2) &= \alpha_1\,\rho(1) + \alpha_2.
\end{aligned}$$

De aquí se obtiene $\rho(1) = \alpha_1/(1-\alpha_2)$ *y* $\rho(2) = \alpha_1^2/(1-\alpha_2) + \alpha_2$. *A partir de estos dos valores se pueden obtener los siguientes valores de* $\rho(\tau)$ *usando la fórmula*

$$\rho(\tau) = \alpha_1\,\rho(\tau-1) + \alpha_2\,\rho(\tau-2), \quad \tau \geqslant 3.$$

Esta es una ecuación recursiva de orden 2 *con valores iniciales* $\rho(1)$ *y* $\rho(2)$. *En el ejercicio 80 se plantea un procedimiento para obtener la solución general de esta ecuación.* ∎

A continuación encontraremos la función de autocovarianza del modelo AR(p). Esta función quedará expresada en términos de la función de autocorrelación.

Proposición 3.8 *Sea* $\{X_t : t \in \mathbb{Z}\}$ *un proceso* AR(p) *que es causal y con función de autocorrelación* $\rho(\tau)$. *La función de autocovarianza es*

$$\gamma(\tau) = \mathrm{Cov}(X_t, X_{t+\tau}) \begin{cases} \dfrac{\sigma^2}{1 - \displaystyle\sum_{k=1}^{p} \alpha_k \, \rho(k)} & si \;\; \tau = 0, \\ \gamma(0)\, \rho(\tau) & si \;\; \tau = 1, 2 \ldots \end{cases}$$

Demostración. Recordemos que $\rho(\tau) = \gamma(\tau)/\gamma(0)$, para $\tau \in \mathbb{Z}$. Habiendo encontrado los valores $\rho(\tau)$ mediante las ecuaciones de Yule-Walker, sólo falta encontrar $\gamma(0)$. Partimos nuevamente de la ecuación

$$X_t = \alpha_1 \, X_{t-1} + \cdots + \alpha_p \, X_{t-p} + Z_t, \quad t \in \mathbb{Z}.$$

Multiplicando por X_t y tomando esperanza se obtiene

$$\gamma(0) = \alpha_1 \, \gamma(1) + \cdots + \alpha_p \, \gamma(p) + E(X_t \, Z_t).$$

Usando la representación (3.21) y el hecho de que $\pi_0 = 1$, se comprueba que el último sumando es $E(X_t \, Z_t) = E(Z_t^2) = \sigma^2$. Substituyendo la expresión $\rho(\tau) = \gamma(t)/\gamma(0)$ para $\tau = 1, 2, \ldots, p$, se obtiene

$$\gamma(0) = \alpha_1 \, \gamma(0) \, \rho(1) + \cdots + \alpha_p \, \gamma(0) \, \rho(p) + \sigma^2.$$

Por lo tanto,

$$\gamma(0) = \frac{\sigma^2}{1 - \alpha_1 \, \rho(1) - \cdots - \alpha_p \, \rho(p)}.$$

∎

En resumen, para un proceso AR(p) que es causal, las funciones de autocovarianza y autocorrelación se calculan de la manera siguiente:

- Recordemos que $\rho(0) = 1$. Resolviendo las ecuaciones de Yule-Walker se obtienen los valores $\rho(1), \ldots, \rho(p)$. A partir de estos valores iniciales se encuentran sucesivamente los siguientes valores:

$$\begin{aligned} \rho(p+1) &= \alpha_1 \rho(p) + \cdots + \alpha_p \rho(1) \\ \rho(p+2) &= \alpha_1 \rho(p+1) + \cdots + \alpha_p \rho(2) \\ &\vdots \end{aligned}$$

- Usando los valores $\rho(1), \ldots, \rho(p)$ se calcula $\gamma(0)$ según la fórmula mostrada en la proposición 3.8.
- Se calcula $\gamma(\tau) = \gamma(0)\,\rho(\tau)$, para cualquier $\tau \in \mathbb{Z}$.

Ejemplo 3.14 *Para el proceso* AR(1) *dado por* $X_t = \alpha_1 X_{t-1} + Z_t$, $t \in \mathbb{Z}$, *sabemos que su función de autocorrelación es* $\rho(\tau) = \alpha_1^\tau$, *para* $\tau \geqslant 0$. *Por la proposición 3.8 recién demostrada, tenemos que*

$$\gamma(0) = \frac{\sigma^2}{1 - \alpha_1 \rho(1)} = \frac{\sigma^2}{1 - \alpha_1^2}.$$

Por lo tanto,

$$\gamma(\tau) = \gamma(0)\,\rho(\tau) = \frac{\sigma^2}{1 - \alpha_1^2}\,\alpha_1^\tau, \quad \tau \geqslant 1.$$

Esto corrobora lo encontrado antes en la proposición 3.4. ∎

Ejemplo 3.15 *Para el proceso* AR(2) *dado por* $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + Z_t$, $t \in \mathbb{Z}$, *habíamos encontrado que* $\rho(1) = \alpha_1/(1-\alpha_2)$ *y* $\rho(2) = \alpha_1^2/(1-\alpha_2) + \alpha_2$. *A partir de estos valores se pueden obtener los siguientes valores de* $\rho(\tau)$ *usando la relación*

$$\rho(\tau) = \alpha_1\,\rho(\tau - 1) + \alpha_2\,\rho(\tau - 2), \quad \tau \geqslant 3.$$

Por la proposición 3.8,

$$\begin{aligned}\gamma(0) &= \frac{\sigma^2}{1-\alpha_1\rho(1)-\alpha_2\rho(2)} \\ &= \frac{\sigma^2}{1-\alpha_1^2/(1-\alpha_2)-\alpha_2(\alpha_1^2/(1-\alpha_2)+\alpha_2)}.\end{aligned}$$

A partir de este número y de las correlaciones $\rho(\tau)$, se pueden calcular las covarianzas $\gamma(\tau) = \gamma(0)\,\rho(\tau)$, $\tau \geqslant 0$. ∎

Recordemos que $\gamma(0) = \mathrm{Cov}(X_t, X_t) = \mathrm{Var}(X_t)$, en donde esta varianza se encontró previamente en términos de los coeficientes π_k de la representación (3.21). De esta manera, tenemos dos expresiones para la varianza de un proceso AR(p) que es causal:

$$\mathrm{Var}(X_t) = \sigma^2 \sum_{k=0}^{\infty} \pi_k^2 = \frac{\sigma^2}{1-\alpha_1\,\rho(1)-\cdots-\alpha_p\,\rho(p)}.$$

El modelo AR[P] estacional

Este es un modelo autorregresivo que incorpora un patrón estacional. El número entero $s \geqslant 2$ denota la longitud del periodo de la estación.

Definición 3.12 *Sean $P \geqslant 0$ y $s \geqslant 2$ dos enteros. Sean $\alpha'_1, \ldots, \alpha'_P$ constantes dadas, con $\alpha'_P \neq 0$. Una serie de tiempo $\{X_t : t \in \mathbb{Z}\}$ sigue un modelo AR[P] estacional de periodo s, si es de la forma*

$$\begin{aligned}X_t &= \alpha'_1\,X_{t-s} + \alpha'_2\,X_{t-2s} + \cdots + \alpha'_P\,X_{t-Ps} + Z_t \qquad (3.26)\\ &= \sum_{k=1}^{P} \alpha'_k\,X_{t-ks} + Z_t, \quad t \in \mathbb{Z}.\end{aligned}$$

Es claro que la variable X_t depende de los P valores previos X_{t-s}, X_{t-2s},..., X_{t-Ps}. Estos valores previos tienen un número de rezago que es múltiplo de s.

También es claro que el proceso (3.26) es un caso particular de un modelo AR(p) con $p = Ps$, en donde los únicos coeficientes distintos de cero son aquellos con subíndice con rezago $s, 2s, \ldots, Ps$. En consecuencia, se cumplen todas las propiedades que conocemos para el modelo AR(p); en particular, sabemos que el proceso (3.26) es causal y, por lo tanto, estacionario cuando el operador de retraso asociado tiene todas sus raíces fuera del círculo unitario. Este operador se denota por $\Phi(B^s)$ y es

$$\Phi(B^s) = 1 - \alpha'_1 B^s - \alpha'_2 B^{2s} - \cdots - \alpha'_P B^{Ps}.$$

La expresión (3.26) se puede escribir de la siguiente forma:

$$\Phi(B^s)X_t = Z_t, \quad t \in \mathbb{Z}. \tag{3.27}$$

Un poco más general, se puede considerar el proceso autorregresivo estacional

$$\Phi(B^s)\phi(B)X_t = Z_t, \quad t \in \mathbb{Z}, \tag{3.28}$$

en donde $\phi(B)$ es el operador usual de orden p considerado antes en la definición 3.10 que aparece en la página 99. En el término $\phi(B)X_t$ se establece la dependencia de X_t respecto de las variables $(X_{t-1}, \ldots, X_{t-p})$ y el operador $\Theta(B^s)$ indica la dependencia de estas mismas variables con rezagos tales que correspondan a los P periodos estacionales previos. Observe que los operadores $\phi(B)$ y $\Phi(B^s)$ conmutan, de modo que su producto se puede escribir en cualquier orden.

El modelo AR(p) con media no cero

Concluimos esta sección definiendo una extensión del modelo AR(p), en donde la media no necesariamente es cero.

Definición 3.13 *Sea μ una constante. Se dice que la serie $\{X_t : t \in \mathbb{Z}\}$ sigue un modelo* AR(p) *con media μ si el proceso centrado $\{X_t - \mu : t \in \mathbb{Z}\}$ es un proceso* AR(p)*, es decir,*

$$X_t - \mu = \alpha_1 (X_{t-1} - \mu) + \cdots + \alpha_p (X_{t-p} - \mu) + Z_t, \quad t \in \mathbb{Z}. \tag{3.29}$$

De la ecuación (3.29) es claro que la serie trasladada $Y_t = X_t - \mu$, para $t \in \mathbb{Z}\}$ sigue el modelo AR(p). Puesto que una serie que sigue ese modelo tiene media cero, tenemos que $E(X_t) = \mu$, para cualquier $t \in \mathbb{Z}$. Las siguientes identidades se pueden comprobar fácilmente:

- $\text{Var}(X_t) = \text{Var}(Y_t)$.
- $\text{Cov}(X_t, X_{t+\tau}) = \text{Cov}(Y_t, Y_{t+\tau})$.
- $\text{Corr}(X_t, X_{t+\tau}) = \text{Corr}(Y_t, Y_{t+\tau})$.

Es decir, aunque el proceso $\{X_t : t \in \mathbb{Z}\}$ dado por (3.29) no sigue propiamente el modelo AR(p), tiene las mismas características numéricas de segundo orden que el proceso $\{Y_t : t \in \mathbb{Z}\}$, y será estacionario cuando $\{Y_t : t \in \mathbb{Z}\}$ lo sea.

3.3. Modelo ARMA

Este modelo es una combinación de un proceso autorregresivo AR(p) y de un proceso de promedios móviles MA(q). Su definición aparece a continuación.

Definición 3.14 *Sean $p \geqslant 0$ y $q \geqslant 0$ dos enteros, no ambos cero. Sean $\alpha_1, \ldots, \alpha_p$ y $\beta_1, \ldots, \beta_q$ constantes tales que $\alpha_p \neq 0$ y $\beta_q \neq 0$. Sea $\{Z_t : t \in \mathbb{Z}\}$ un proceso de ruido con media 0 y varianza σ^2. Una serie de tiempo $\{X_t : t \in \mathbb{Z}\}$ es un proceso ARMA de orden (p, q) si es de la forma*

$$\begin{aligned} X_t &= \underbrace{\alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p}} + Z_t + \underbrace{\beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q}} \\ &= \sum_{k=1}^{p} \alpha_k X_{t-k} + Z_t + \sum_{k=1}^{q} \beta_k Z_{t-k}, \quad t \in \mathbb{Z}. \end{aligned} \tag{3.30}$$

Cuando $p = 0$ y $q \geqslant 1$, la expresión (3.30) se reduce al modelo MA(q), y cuando $p \geqslant 1$ y $q = 0$, se obtiene el modelo AR(p). De esta manera, el estudio del modelo ARMA(p, q) incluye los dos modelos particulares

anteriores.

Definiendo $\beta_0 = 1$, y en términos del operador de retraso B, un proceso ARMA(p, q) dado por la expresión (3.30) se puede escribir como

$$\underbrace{(1 - \sum_{k=1}^{p} \alpha_k B^k)}_{\phi(B)} X_t = \underbrace{(\sum_{k=0}^{q} \beta_k B^k)}_{\theta(B)} Z_t, \quad t \in \mathbb{Z}. \tag{3.31}$$

Es decir,

$$\phi(B)\, X_t = \theta(B)\, Z_t, \quad t \in \mathbb{Z}, \tag{3.32}$$

en donde $\phi(B)$ y $\theta(B)$ son los operadores de retraso definidos antes y a los que también se les refiere como operador autorregresivo y operador de promedios móviles, respectivamente, pues corresponden a esos dos aspectos del modelo ARMA. A continuación se recuerda su definición.

Definición 3.15 *Para el modelo* ARMA(p, q) *especificado en* (3.30) *se definen:*

1. *Los operadores de retraso*

$$\begin{aligned} \phi(B) &= 1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_p B^p, \\ \theta(B) &= 1 + \beta_1 B + \beta_2 B^2 + \cdots + \beta_q B^q. \end{aligned}$$

2. *Los polinomios*

$$\begin{aligned} \phi(x) &= 1 - \alpha_1 x - \alpha_2 x^2 - \cdots - \alpha_p x^p, \quad x \in \mathbb{R}, \\ \theta(x) &= 1 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_q x^q, \quad x \in \mathbb{R}. \end{aligned}$$

Como en los casos de los modelos AR(p) y MA(q), nos interesa la invertibilidad de los operadores $\phi(B)$ y $\theta(B)$ pues, en tal situación, se puede escribir la serie $\{X_t : t \in \mathbb{Z}\}$ en términos del proceso de ruido $\{Z_t : t \in \mathbb{Z}\}$ y viceversa. Tal característica de invertibilidad está relacionada con las normas de las raíces de los polinomios asociados $\phi(x)$ y $\theta(x)$ como ya se ha indicado.

Causalidad del modelo ARMA(p, q)

La siguiente definición es idéntica a la que aparece en la página 100 sobre la causalidad para procesos AR(p).

Definición 3.16 *Se dice que un proceso* $\{X_t : t \in \mathbb{Z}\}$ *que sigue un modelo* ARMA(p, q) *dado por* (3.32) *es causal si el operador* $\phi(B)$ *es invertible.*

En este caso, el proceso definido en (3.30) se puede escribir como

$$X_t = (\phi(B))^{-1}\,\theta(B)\,Z_t, \quad t \in \mathbb{Z}, \tag{3.33}$$

es decir, la serie $\{X_t : t \in \mathbb{Z}\}$ se ha escrito exclusivamente en términos del proceso de ruido $\{Z_t : t \in \mathbb{Z}\}$. Recordemos que $\phi(B)$ es invertible si y sólo si la ecuación $\phi(x) = 0$ tiene todas sus raíces fuera del círculo unitario. La causalidad implica que existen constantes $\pi_0, \pi_1, \ldots$ tales que el proceso $\{X_t : t \in \mathbb{Z}\}$ adquiere la representación

$$X_t = \sum_{k=0}^{\infty} \pi_k\, Z_{t-k}, \quad t \in \mathbb{Z}, \tag{3.34}$$

en donde $\sum_{k=0}^{\infty} \pi_k^2 < \infty$. Esta representación corresponde a un modelo de promedios móviles de orden infinito e implica que el proceso es estacionario. En particular, la representación (3.34) ayuda a encontrar fácilmente algunas características del modelo ARMA(p, q), como la media y la varianza. Esos cálculos se hacen más adelante. En particular, bajo la hipótesis de causalidad, el modelo ARMA(p, q) corresponde a una serie de tiempo sin tendencia, pues comprobaremos que $E(X_t) = 0$, para cualquier $t \in \mathbb{Z}$.

Invertibilidad del modelo ARMA(p, q)

Definición 3.17 *Se dice que un proceso* $\{X_t : t \in \mathbb{Z}\}$ *que sigue un modelo* ARMA(p, q) *dado por* (3.32) *es invertible si el operador* $\theta(B)$ *es invertible.*

En este caso, tenemos la relación

$$Z_t = (\theta(B))^{-1}\,\phi(B)\,X_t, \quad t \in \mathbb{Z}, \tag{3.35}$$

es decir, el proceso de ruido $\{Z_t : t \in \mathbb{Z}\}$ es función de la serie $\{X_t : t \in \mathbb{Z}\}$. Recordemos que $\theta(B)$ es invertible si y sólo si la ecuación $\theta(x) = 0$ tiene todas sus raíces fuera del círculo unitario. La definición anterior es la misma que se dio para el modelo MA(q), e implica que existen constantes $\pi_0, \pi_1, \ldots$ tales que el proceso de ruido $\{Z_t : t \in \mathbb{Z}\}$ adquiere la representación

$$Z_t = \sum_{k=0}^{\infty} \pi_k\, X_{t-k}, \quad t \in \mathbb{Z}, \tag{3.36}$$

en donde $\sum_{k=0}^{\infty} |\pi_k| < \infty$.

Media y varianza del modelo ARMA(p, q)

En esta sección se encuentran la media y la varianza del modelo general ARMA(p, q). Estas cantidades son idénticas a las del modelo AR(p). El cálculo hace uso de la representación (3.34).

Proposición 3.9 *Sea $\{X_t : t \in \mathbb{Z}\}$ un proceso* ARMA(p, q) *que es causal y tiene la representación* (3.34)*. Entonces*

1. $E(X_t) = 0$.
2. $\text{Var}(X_t) = \sigma^2 \sum_{k=0}^{\infty} \pi_k^2$.

Demostración. Suponiendo que es válido el intercambio de suma infinita y esperanza, tenemos que

1. $E(X_t) = E(\sum_{k=0}^{\infty} \pi_k\, Z_{t-k}) = \sum_{k=0}^{\infty} \pi_k\, E(Z_{t-k}) = 0.$

2. $\text{Var}(X_t) = \text{Var}(\sum_{k=0}^{\infty} \pi_k Z_{t-k}) = \sum_{k=0}^{\infty} \pi_k^2 \text{Var}(Z_{t-k}) = \sigma^2 \sum_{k=0}^{\infty} \pi_k^2.$

■

Autocovarianza y autocorrelación del modelo ARMA(p, q)

A continuación encontraremos una forma de calcular las funciones de autocovarianza y de autocorrelación para un modelo ARMA(p, q) que es causal y, por lo tanto, estacionario.

Proposición 3.10 *Sea* $\{X_t : t \in \mathbb{Z}\}$ *un proceso* ARMA(p, q) *dado por* (3.30). *Suponga que el proceso es causal y tiene la representación* (3.34). *Entonces*

1. *La función de autocovarianza* $\gamma(\tau)$ *está dada por el sistema de ecuaciones*

$$\gamma(\tau) - \alpha_1 \gamma(\tau - 1) - \cdots - \alpha_p \gamma(\tau - p) = \sigma^2 \sum_{k=0}^{\infty} \pi_k \beta_{\tau+k}, \quad (3.37)$$

   para $\tau = 0, 1, 2, \ldots$ *en donde se define* $\beta_{q+1} = \beta_{q+2} = \cdots = 0$, *de modo que la suma indicada siempre es finita.*

2. *La función de autocorrelación es*

$$\rho(\tau) = \gamma(\tau)/\gamma(0), \quad \textit{para} \quad \tau = 0, 1, 2, \ldots$$

Demostración.

1. Sea $\tau \geqslant 0$ un entero. Conviene recordar la ecuación que define al modelo ARMA(p, q),

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q}, \quad t \in \mathbb{Z}.$$

Multiplicando esta identidad por $X_{t-\tau}$ y tomando esperanza se obtiene

$$\gamma(\tau) - \alpha_1\,\gamma(\tau-1) - \cdots - \alpha_p\,\gamma(\tau-p) = \sum_{k=0}^{\infty} \pi_k \sum_{j=0}^{q} \beta_j\, E(Z_{t-\tau-k} \cdot Z_{t-j}),$$

en donde el lado derecho se obtiene al substituir la expansión (3.36) para $X_{t-\tau}$. Este lado derecho es distinto de 0 sólo cuando $j = \tau + k$. Se obtiene así la relación

$$\gamma(\tau) - \alpha_1\,\gamma(\tau-1) - \cdots - \alpha_p\,\gamma(\tau-p) = \sigma^2 \sum_{k=0}^{\infty} \pi_k\,\beta_{\tau+k}.$$

2. Conociendo los valores de $\gamma(\tau)$ para $\tau = 0, 1, 2, \ldots$, la función de autocorrelación se calcula de la forma usual, es decir, $\rho(\tau) = \gamma(\tau)/\gamma(0)$, para $\tau = 0, 1, 2, \ldots$

■

Ejemplo 3.16 *Para comprender mejor el conjunto de ecuaciones* (3.37)*, las escribiremos de manera explícita para el caso* $p = 2$ *y* $q = 3$.

$$\begin{aligned}
\tau = 0 \quad & \gamma(0) - \alpha_1\gamma(1) - \alpha_2\gamma(2) = \sigma^2(\pi_0\beta_0 + \pi_1\beta_1 + \pi_2\beta_2 + \pi_3\beta_3) \\
\tau = 1 \quad & \gamma(1) - \alpha_1\gamma(0) - \alpha_2\gamma(1) = \sigma^2(\pi_0\beta_1 + \pi_1\beta_2 + \pi_2\beta_3) \\
\tau = 2 \quad & \gamma(2) - \alpha_1\gamma(1) - \alpha_2\gamma(0) = \sigma^2(\pi_0\beta_2 + \pi_1\beta_3) \\
\\
\tau = 3 \quad & \gamma(3) - \alpha_1\gamma(2) - \alpha_2\gamma(1) = \sigma^2(\pi_0\beta_3) \\
\tau = 4 \quad & \gamma(4) - \alpha_1\gamma(3) - \alpha_2\gamma(2) = 0 \\
\vdots \quad & \vdots
\end{aligned}$$

Las ecuaciones para $\tau = 0, 1, 2$ *forman un sistema de ecuaciones de* 3×3 *que se resuelve para conocer los valores de* $\gamma(0)$*,* $\gamma(1)$ *y* $\gamma(2)$*. Los siguientes valores:* $\gamma(3), \gamma(4), \ldots$ *se encuentran de manera recursiva usando la fórmula* (3.37)*, en donde los valores base serán* $\gamma(1)$ *y* $\gamma(2)$*, como se muestra en las ecuaciones indicadas arriba para* $\tau = 3, 4, \ldots$ ■

En el caso $p > q$, el lado derecho de las ecuaciones es igual a cero, excepto tal vez para algunos casos, cuando $\tau = 0, 1, 2$.

Concluimos esta sección con el enunciado de un resultado que establece condiciones necesarias y suficientes que caracterizan la causalidad y la invertibilidad en un proceso ARMA(p, q). Estos resultados son similares a los que se presentaron como teorema 3.2 en la página 101 para la causalidad de un proceso AR(p), y teorema 3.1 en la página 88 para la invertibilidad de un proceso MA(q).

Teorema 3.3 (Criterio para la causalidad y la invertibilidad del modelo ARMA) *Considere un proceso* ARMA(p, q) *tal que los polinomios asociados* $\phi(x)$ *y* $\theta(x)$ *no tienen raíces comunes. Entonces*

1. *El proceso es invertible* $\Leftrightarrow$
   toda raíz $z \in \mathbb{C}$ *del polinomio* $\theta(x)$ *tiene norma* $|z| > 1$.

2. *El proceso es causal* $\Leftrightarrow$
   toda raíz $z \in \mathbb{C}$ *del polinomio* $\phi(x)$ *tiene norma* $|z| > 1$.

La demostración de este resultado se puede encontrar en Brockwell y Davis (1987) [7]. En esta referencia aparecen los enunciados separados como teorema 3.1.1 y teorema 3.1.2 .

Como se ha señalado, la condición para la invertibilidad que se presenta en el teorema 3.3 se puede expresar de manera equivalente solicitando que $\theta(x) \neq 0$ para cualquier $x \in \mathbb{C}$ dentro del círculo unitario $\{x \in \mathbb{C} : |x| \leqslant 1\}$. De manera similar, la condición para la causalidad también se puede escribir como $\phi(x) \neq 0$ para cualquier $x \in \mathbb{C}$ dentro del círculo unitario alrededor del origen.

El modelo ARMA$(1,1)$

En esta sección se revisan los detalles del modelo particular ARMA$(1,1)$ definido por la relación

$$X_t = \alpha_1 X_{t-1} + Z_t + \beta_1 Z_{t-1}, \quad t \in \mathbb{Z}, \tag{3.38}$$

en donde $\alpha_1 \neq 1$ y $\beta_1 \neq 1$. Como en el caso del modelo general, este proceso se puede escribir en términos de los operadores de retraso como

$$\phi(B) X_t = \theta(B) Z_t, \quad t \in \mathbb{Z}, \tag{3.39}$$

en donde $\phi(B) = 1 - \alpha_1 B$ y $\theta(B) = 1 + \beta_1 B$.

Estacionariedad del modelo ARMA$(1,1)$

El proceso ARMA$(1,1)$ es estacionario cuando el operador de retraso $\phi(B)$ es invertible, de tal manera que la serie admite la siguiente representación de medias móviles de orden infinito

$$X_t = (\phi(B))^{-1} \theta(B) Z_t, \quad t \in \mathbb{Z}.$$

El operador $\phi(B)$ es invertible cuando todas las raíces del polinomio $\phi(x) = 1 - \alpha_1 x$ quedan fuera del círculo unitario. La única raíz es $x = 1/\alpha_1$, de modo que el proceso es estacionario cuando $|1/\alpha_1| > 1$, es decir, cuando $-1 < \alpha_1 < 1$.

Causalidad del modelo ARMA$(1,1)$

De acuerdo con la definición, la causalidad requiere la invertibilidad del operador de retraso $\phi(B) = 1 - \alpha_1 B$ como en el párrafo anterior. Por lo tanto, el proceso es causal cuando $|\alpha_1| < 1$. A continuación se encuentra la expansión de la serie $\{X_t : t \in \mathbb{Z}\}$ en términos del proceso de ruido $\{Z_t : t \in \mathbb{Z}\}$.

$$\begin{aligned} X_t &= (\phi(B))^{-1}\,\theta(B)\,Z_t \\ &= (1-\alpha_1 B)^{-1}(1+\beta_1 B)\,Z_t \\ &= (1+\alpha_1 B+\alpha_1^2 B^2+\cdots)(1+\beta_1 B)\,Z_t \\ &= [1+(\alpha_1+\beta_1)B+(\alpha_1+\beta_1)\alpha_1 B^2+(\alpha_1+\beta_1)\alpha_1^2 B^3+\cdots]\,Z_t \\ &= Z_t+(\alpha_1+\beta_1)\sum_{k=1}^{\infty}\alpha_1^{k-1}\,Z_{t-k} \qquad (3.40) \\ &= \sum_{k=0}^{\infty}\pi_k\,Z_{t-k}, \end{aligned}$$

en donde los coeficientes de la última serie son

$$\pi_0 = 1 \quad \text{y} \quad \pi_k = (\alpha_1+\beta_1)\alpha_1^{k-1} \quad \text{para} \quad k = 1, 2, \ldots$$

De esta manera, la serie $\{X_t : t \in \mathbb{Z}\}$ queda expresada únicamente en términos del proceso $\{Z_t : t \in \mathbb{Z}\}$ como un modelo de medias móviles de orden infinito. A partir de (3.40) se puede verificar fácilmente la media y la varianza de este modelo, a saber,

- $E(X_t) = 0$.
- $\text{Var}(X_t) = [1+\dfrac{(\alpha_1+\beta_1)^2}{1-\alpha_1^2}]\,\sigma^2$.

Invertibilidad del modelo ARMA$(1,1)$

La parte de promedio móviles del modelo ARMA$(1,1)$ tiene asociado el operador de retraso $\theta(B) = 1+\beta_1 B$. Este operador es invertible cuando toda raíz $z \in \mathbb{C}$ del polinomio $\theta(x) = 1+\beta_1 x$ es tal que $|z| > 1$. La única raíz de esta ecuación es $z = -1/\beta_1$. Por lo tanto, $\theta(B)$ es invertible cuando $-1 < \beta_1 < 1$. En tal situación, usando la expansión

$$\frac{1}{1+x} = 1-x+x^2-x^3+\cdots \quad \text{para} \quad -1 < x < 1,$$

y para $-1 < \beta_1 < 1$, tenemos que

$$\begin{aligned} Z_t &= (\theta(B))^{-1}\,\phi(B)\,X_t \\ &= (1+\beta_1 B)^{-1}\,(1-\alpha_1 B)\,X_t \\ &= (1-\beta_1 B+\beta_1^2 B^2-\beta_1^3 B^3+\cdots)\,(1-\alpha_1 B)\,X_t \\ &= [1-(\beta_1+\alpha_1)B+(\beta_1+\alpha_1)\beta_1 B^2-(\beta_1+\alpha_1)\beta_1^2 B^3+\cdots]\,X_t \\ &= X_t-(\beta_1+\alpha_1)\sum_{k=1}^{\infty}(-\beta_1)^{k-1}\,X_{t-k}, \quad t\in\mathbb{Z}. \end{aligned}$$

Así, hemos expresado el proceso de ruido $\{Z_t : t \in \mathbb{Z}\}$ en términos de la serie de tiempo $\{X_t : t \in \mathbb{Z}\}$.

Autocovarianza y autocorrelación del modelo ARMA$(1,1)$

A continuación se calculan las funciones de autocovarianza y de autocorrelación del modelo ARMA$(1,1)$ bajo la hipótesis de estacionariedad.

Proposición 3.11 *Sea* $\{X_t : t \in \mathbb{Z}\}$ *un proceso* ARMA$(1,1)$ *especificado en* (3.38) *y que es estacionario. Entonces*

1. *La función de autocovarianza es*

$$\gamma(\tau)=\mathrm{Cov}(X_t,X_{t+\tau})=\begin{cases}\left[1+\dfrac{(\alpha_1+\beta_1)^2}{1-\alpha_1^2}\right]\sigma^2 & si\ \ \tau=0,\\[2ex] \alpha_1^{\tau-1}\dfrac{(1+\alpha_1\beta_1)(\alpha_1+\beta_1)}{1-\alpha_1^2}\,\sigma^2 & si\ \ \tau\geqslant 1.\end{cases}$$

2. *La función de autocorrelación es*

$$\rho(\tau)=\mathrm{Corr}(X_t,X_{t+\tau})=\begin{cases}1 & si\ \ \tau=0,\\[2ex] \alpha_1^{\tau-1}\dfrac{(1+\alpha_1\beta_1)(\alpha_1+\beta_1)}{1+2\alpha_1\beta_1+\beta_1^2} & si\ \ \tau\geqslant 1.\end{cases}$$

Demostración.

1. El caso $\tau = 0$ corresponde a la varianza que fue indicada antes, y ahora se presentan los detalles. Usando la representación (3.40),

$$\begin{aligned}
\gamma(0) &= \mathrm{Cov}(X_t, X_t) \\
&= \mathrm{Var}(X_t) \\
&= \mathrm{Var}(Z_t + (\alpha_1 + \beta_1) \sum_{k=1}^{\infty} \alpha_1^{k-1} Z_{t-k}) \\
&= \mathrm{Var}(Z_t) + (\alpha_1 + \beta_1)^2 \sum_{k=1}^{\infty} \alpha_1^{2(k-1)} \mathrm{Var}(Z_{t-k}) \\
&= \sigma^2 + (\alpha_1 + \beta_1)^2 \sigma^2 \sum_{k=0}^{\infty} (\alpha_1^2)^k \\
&= [1 + \frac{(\alpha_1 + \beta_1)^2}{1 - \alpha_1^2}] \sigma^2.
\end{aligned}$$

Para $\tau = 1, 2, \ldots$, usando nuevamente (3.40),

$$\begin{aligned}
\gamma(\tau) &= \mathrm{Cov}(X_t, X_{t+\tau}) \\
&= \mathrm{Cov}(Z_t + (\alpha_1 + \beta_1) \sum_{k=1}^{\infty} \alpha_1^{k-1} Z_{t-k}, \\
&\qquad Z_t + (\alpha_1 + \beta_1) \sum_{k=1}^{\infty} \alpha_1^{k-1} Z_{t-k}) \\
&= \underbrace{\mathrm{Cov}(Z_t, Z_{t+\tau})}_{=0} + \underbrace{(\alpha_1 + \beta_1) \sum_{k=1}^{\infty} \alpha_1^{k-1} \mathrm{Cov}(Z_t, Z_{t+\tau-k})}_{=(\alpha_1+\beta_1)\,\alpha_1^{\tau-1}\,\sigma^2} \\
&\quad + \underbrace{(\alpha_1 + \beta_1) \sum_{k=1}^{\infty} \alpha_1^{k-1} \mathrm{Cov}(Z_{t-k}, Z_{t+\tau})}_{=0} \\
&\quad + (\alpha_1 + \beta_1)^2 \sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \alpha_1^{k+\ell-2} \mathrm{Cov}(Z_{t-k}, Z_{t+\tau-\ell})
\end{aligned}$$

En la última expresión se tienen 4 sumandos. El primero y el tercero son cero, el segundo es $(\alpha_1 + \beta_1)\,\alpha_1^{\tau-1}\,\sigma^2$ y es resultado del término de la suma cuando $k = \tau$; finalmente, la covarianza que aparece en el cuarto sumando es

$$\begin{cases} \sigma^2 & \text{si } \ \ell = \tau + k, \\ 0 & \text{en otro caso.} \end{cases}$$

Por lo tanto,

$$\begin{aligned} \gamma(\tau) &= (\alpha_1 + \beta_1)\,\alpha_1^{\tau-1}\,\sigma^2 + (\alpha_1 + \beta_1)^2 \sum_{k=1}^{\infty} \alpha_1^{\tau+2k-2}\,\sigma^2 \\ &= (\alpha_1 + \beta_1)\,\alpha_1^{\tau-1}\,\sigma^2 + (\alpha_1 + \beta_1)^2\,\alpha_1^{\tau} \sum_{k=1}^{\infty} (\alpha_1^2)^{k-1}\,\sigma^2 \\ &= (\alpha_1 + \beta_1)\,\alpha_1^{\tau-1}\,\sigma^2 + \frac{(\alpha_1 + \beta_1)^2}{1 - \alpha_1^2}\,\alpha_1^{\tau}\,\sigma^2 \\ &= \alpha_1^{\tau-1}\frac{(1 + \alpha_1\beta_1)(\alpha_1 + \beta_1)}{1 - \alpha_1^2}\,\sigma^2. \end{aligned}$$

2. Usando el inciso anterior, para $\tau = 1, 2, \ldots$

$$\begin{aligned} \rho(\tau) &= \frac{\gamma(\tau)}{\gamma(0)} \\ &= \alpha_1^{\tau-1}\frac{(1 + \alpha_1\beta_1)(\alpha_1 + \beta_1)}{1 - \alpha_1^2}\,\sigma^2 \cdot \frac{1 - \alpha_1^2}{(1 + 2\alpha_1\beta_1 + \beta_1^2)\sigma^2} \\ &= \alpha_1^{\tau-1}\frac{(1 + \alpha_1\beta_1)(\alpha_1 + \beta_1)}{1 + 2\alpha_1\beta_1 + \beta_1^2}. \end{aligned}$$

■

El modelo ARMA$[P, Q]$ estacional

Este es un modelo ARMA que permite incorporar un patrón estacional en la estructura del modelo. La estacionalidad aparece tanto en la parte de promedios móviles como en la parte autorregresiva, y supondremos

que la longitud del periodo estacional es el entero $s \geqslant 2$. Se trata de la combinación de los procesos MA[Q] y AR[P] estacionales ya mencionados. El resultado será, nuevamente, un caso particular de un modelo ARMA.

Además de los operadores usuales $\theta(B)$ de orden q, y $\phi(B)$ de orden p, asociados a los modelo MA(q) y AR(p), respectivamente, conviene recordar la definición de los siguiente operadores de retrasos estacionales:

$$\begin{aligned} \Theta(B^s) &= 1 + \beta_1' B^s + \beta_2' B^{2s} + \cdots + \beta_Q' B^{Qs}, \\ \Phi(B^s) &= 1 - \alpha_1' B^s - \alpha_2' B^{2s} - \cdots - \alpha_P' B^{Ps}, \end{aligned}$$

en donde $Q \geqslant 0$ y $P \geqslant 0$ son enteros. Estos operadores son semejantes a los operadores $\theta(B)$ y $\phi(B)$, pero sus retrasos son múltiplos del periodo estacional s y sus coeficientes se denotan por las mismas letras, pero tienen un apóstrofe ya que pueden ser distintos a los de los operadores $\theta(B)$ y $\phi(B)$.

Definición 3.18 *Una serie de tiempo* $\{X_t : t \in \mathbb{Z}\}$ *sigue un modelo estacional ARMA de periodo* s*, y se denota por* ARMA$(p,q)[P,Q]_s$*, si es de la forma*

$$\Phi(B^s)\phi(B)X_t = \Theta(B^s)\theta(B)Z_t, \quad t \in \mathbb{Z}. \tag{3.41}$$

El operador $\Phi(B^s)$ lleva la parte regresiva $\phi(B)X_t$ a los P periodos estacionales previos y el operador $\Theta(B^s)$ establece la dependencia de X_t respecto de $\theta(B)Z_t, \theta(B)Z_{t-s}, \ldots, \theta(B)Z_{t-Qs}$. Esta es la interpretación que se indicó antes cuando se mencionaron por separado los modelos MA[Q] y AR[P] estacionales. Véanse las páginas 89 y 109.

El modelo ARMA(p,q) con media no cero

Finalmente, a continuación se define el proceso ARMA(p,q) con media no necesariamente cero.

Definición 3.19 *Sea μ una constante. Se dice que la serie $\{X_t : t \in \mathbb{Z}\}$ sigue un modelo* ARMA(p, q) *con media μ si el proceso centrado $\{X_t - \mu : t \in \mathbb{Z}\}$ es un proceso* ARMA(p, q)*, es decir,*

$$\begin{aligned} X_t - \mu &= \alpha_1 (X_{t-1} - \mu) + \cdots + \alpha_p (X_{t-p} - \mu) + Z_t \\ &\quad + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q}, \qquad t \in \mathbb{Z}. \end{aligned} \tag{3.42}$$

De la ecuación (3.42) se puede ver de inmediato que la serie trasladada $\{X_t - \mu : t \in \mathbb{Z}\}$ sigue el modelo ARMA(p, q). Como una serie que sigue ese modelo tiene media cero, tenemos que $E(X_t) = \mu$, para cualquier $t \in \mathbb{Z}$.

3.4. Modelo ARIMA

Esta sección contiene una breve introducción al modelo ARIMA. Este tipo de procesos es una generalización del modelo ARMA(p, q). La letra I añadida se refiere a la palabra en inglés *Integrated*, e indica que cierta operación de diferenciación de la serie se ha integrado al modelo.

Una serie de tiempo ARIMA es un modelo que se puede utilizar para modelar datos que presentan tendencia, y que tal tendencia se puede eliminar a través de la aplicación del operador diferencia $\nabla^d = (1 - B)^d$, para cierto número entero $d \geqslant 0$. A continuación se establece la definición.

Definición 3.20 *Sean $p, d, q \geqslant 0$ números enteros. Una serie de tiempo $\{X_t : t \in \mathbb{Z}\}$ sigue un modelo* ARIMA(p, d, q) *si la serie $\{W_t : t \in \mathbb{Z}\}$ que aparece a continuación es un proceso* ARMA(p, q)*,*

$$W_t = (1 - B)^d X_t, \quad t \in \mathbb{Z}.$$

Esto quiere decir que la serie diferenciada $W_t = \nabla^d X_t = (1 - B)^d X_t$ se

puede expresar como

$$W_t = \sum_{k=1}^{p} \alpha_k W_{t-k} + Z_t + \sum_{k=1}^{q} \beta_k Z_{t-k}, \quad t \in \mathbb{Z},$$

para ciertas constantes $\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q$, en donde $\{Z_t : t \in \mathbb{Z}\}$ es un proceso de ruido con media 0 y varianza σ^2. En el ejercicio 18 se señala que las tendencias de tipo polinomial en una serie de tiempo se pueden eliminar mediante la aplicación del operador $\nabla^d = (1-B)^d$.

En términos de los operadores de retraso $\phi(B)$ y $\theta(B)$ definidos antes, tenemos la representación $\phi(B)W_t = \theta(B)Z_t$ para un proceso ARIMA o, más explícitamente,

$$\phi(B)(1-B)^d X_t = \theta(B)Z_t, \quad t \in \mathbb{Z}. \tag{3.43}$$

Observemos que el modelo ARIMA$(p, 0, q)$ es el modelo ARMA(p, q), pues $(1-B)^0 = 1$. Además, recordemos que diferenciar una serie $\{X_t : t \in \mathbb{Z}\}$ significa aplicar el operador diferencia $\nabla^d = (1-B)^d$ a cada variable aleatoria de la serie. Se produce así la nueva serie de tiempo. Por ejemplo,

$$\begin{aligned}
&(1-B)^0 X_t = X_t && \text{Identidad} \\
&(1-B)^1 X_t = X_t - X_{t-1} && \text{Primera diferencia} \\
&(1-B)^2 X_t = X_t - 2X_{t-1} + X_{t-2} && \text{Segunda diferencia} \\
&\quad \vdots && \quad \vdots
\end{aligned}$$

Modelo ARIMA sin alguna de sus partes

Cuando un proceso ARIMA(p, d, q) no contiene la parte regresiva se le llama modelo IMA(d, q). Así, tenemos que $\phi(B) = 1$, y la representación (3.43) toma la expresión reducida que aparece a continuación en la ecuación (3.44), que corresponde a un modelo ARMA(d, q),

$$(1-B)^d X_t = \theta(B)Z_t, \quad t \in \mathbb{Z}. \tag{3.44}$$

Por ejemplo, el modelo IMA$(1, q)$ está dado por la expresión (3.45), el cual es un modelo ARMA$(1, q)$,

$$X_t = X_{t-1} + \theta(B)Z_t, \quad t \in \mathbb{Z}. \tag{3.45}$$

También se pueden considerar modelos ARIMA(p, d, q) sin la parte de promedios móviles, es decir, modelos ARIMA en donde $\theta(B) = 1$. En este caso se obtiene el modelo que aparece en (3.46). No es difícil darse cuenta que este aparente nuevo modelo es en realidad un proceso AR$(p + d)$.

$$\phi(B)(1 - B)^d X_t = Z_t, \quad t \in \mathbb{Z}. \tag{3.46}$$

Por ejemplo, cuando $p = 1$ y $d = 1$ se tiene el modelo autorregresivo

$$X_t - (1 + \alpha_1)X_{t-1} + \alpha_1 X_{t-2} = Z_t, \quad t \in \mathbb{Z}. \tag{3.47}$$

Causalidad, invertibilidad y estacionariedad del modelo ARIMA(p, d, q)

- Un proceso ARIMA(p, d, q) no es causal cuando $d \geqslant 1$, pues el polinomio asociado al operador $\phi(B)(1 - B)^d$ es $\phi(x)(1 - x)^d$, y la ecuación $\phi(x)(1 - x)^d = 0$ tiene d raíces sobre el círculo unitario, vea el teorema 3.3. Así, como no se tiene la causalidad, no se puede garantizar la estacionariedad del proceso ARIMA(p, d, q) para $d \geqslant 1$.

- Cuando $d = 0$ tenemos la situación del modelo ARMA(p, q), el cual es causal, y por lo tanto estacionario cuando el operador $\phi(B)$ es invertible. De esta manera, la diferenciación $\nabla^d X_t$ hace que esta nueva serie siga el modelo ARMA(p, q), es decir, ahora $d = 0$ y la serie diferenciada puede ser estacionaria cuando $\phi(B)$ sea invertible. De esta manera el modelo ARIMA se puede utilizar para representar tanto series estacionarias como series no estacionarias.

- A partir de la representación (3.43), es claro que un proceso que sigue el modelo ARIMA(p, d, q) es invertible cuando el operador $\theta(B)$ es invertible; esta es la misma situación que se tenía antes para el proceso ARMA(p, q).

A continuación se muestran algunos ejemplos.

Ejemplo 3.17 *El modelo* ARIMA$(0, 0, 0)$ *es el proceso de ruido* $\{Z_t : t \in \mathbb{Z}\}$*, el cual es estacionario. Ver sus propiedades en la página 10.* ▪

Ejemplo 3.18 *El modelo* ARIMA$(0,1,0)$ *es la caminata aleatoria*

$$X_t - X_{t-1} = Z_t, \quad t \in \mathbb{Z}.$$

En el ejercicio 27 se pide demostrar que este proceso no es estacionario. Tal afirmación no es difícil de verificar, ya que se puede comprobar que el proceso admite la representación $X_t = Z_1 + \cdots + Z_t$, *para* $t \geqslant 1$. *A partir de esta expresión los cálculos son sencillos. Aunque* $E(X_t) = 0$, *la función de autocovarianza* $\gamma(t, t+\tau)$ *depende de* t, *pues*

$$\begin{aligned}
\gamma(t, t+\tau) &= \operatorname{Cov}(X_t, X_{t+\tau}) \\
&= \operatorname{Cov}(\sum_{k=1}^{t} Z_k, \sum_{k=1}^{t+\tau} Z_k) \\
&= \operatorname{Cov}(\sum_{k=1}^{t} Z_k, \sum_{k=1}^{t} Z_k) \\
&= t\,\sigma^2.
\end{aligned}$$

Se concluye que la caminata aleatoria no es una serie de tiempo estacionaria. Por otro lado, el proceso también se puede escribir como $\phi(B)(1-B)X_t = \theta(B)Z_t$, *en donde los operadores de retraso son los operadores identidad, es decir, en este caso,* $\phi(B) = 1$ *y* $\theta(B) = 1$. *Es inmediato ver que el proceso diferenciado es el proceso de ruido* $W_t = Z_t$, $t \geqslant 1$, *el cual evidentemente es estacionario. Ver sus propiedades en la página 10.* ▪

Ejemplo 3.19 *El modelo* ARIMA$(1,1,0)$ *es*

$$(1-\alpha_1 B)(1-B)X_t = Z_t, \quad t \in \mathbb{Z}.$$

Este proceso no es estacionario, pero el proceso diferenciado una vez, $W_t = X_t - X_{t-1} = (1_B)X_t$ *es estacionario cuando* $-1 < \alpha_1 < 1$. *Para el proceso diferenciado tenemos la representación usual*

$$\begin{aligned}
W_t &= X_t - X_{t-1} \\
&= (1-B)X_t \\
&= (1-\alpha_1 B)^{-1} Z_t \\
&= (1 + \alpha_1 B + \alpha_1^2 B^2 + \cdots) Z_t \\
&= \sum_{k=0}^{\infty} \alpha_1^k \, Z_{t-k}.
\end{aligned}$$

Alternativamente, se puede escribir

$$\begin{aligned} X_t &= X_{t-1} + W_t \\ &= X_{t-2} + W_{t-1} + W_t \\ &= X_{t-3} + W_{t-2} + W_{t-1} + W_t \\ &\vdots \\ &= X_0 + \sum_{k=1}^{t} W_k. \end{aligned}$$

■

Ejemplo 3.20 *Considere la caminata aleatoria $X_t = \alpha_0 + \alpha_1 t + Z_t$. Esta serie no es estacionaria pues su media es $\alpha_0 + \alpha_1 t$, lo cual depende del tiempo. La primera diferencia de la serie es*

$$\begin{aligned} W_t^{(1)} &= (1-B)X_t \\ &= (\alpha_0 + \alpha_1 t + Z_t) - (\alpha_0 + \alpha_1(t-1) + Z_{t-1}) \\ &= \alpha_1 + Z_t + Z_{t-1}, \end{aligned}$$

la cual es ahora estacionaria. Este no es un modelo ARIMA debido a la constante α_1. La segunda diferencia es

$$\begin{aligned} W_t^{(2)} &= (1-B)^2 X_t \\ &= (\alpha_1 + Z_t + Z_{t-1}) - (\alpha_1 + Z_{t-1} + Z_{t-2}) \\ &= Z_t - Z_{t-2}, \end{aligned}$$

Tenemos ahora un modelo ARIMA$(0, 2, 2)$. ■

Por definición, el modelo ARIMA(p, d, q) es tal que el proceso asociado $W_t = (1-B)^d X_t$ sigue un modelo ARMA(p, q), y por lo tanto, $E(W_t) = 0$ para cualquier $t \in \mathbb{Z}$. Una extensión del modelo ARIMA consiste en solicitar que $\{W_t - \mu : t \in \mathbb{Z}\}$ siga el modelo ARMA para algún valor μ. Esta condición hace que $E(W_t) = \mu$.

El modelo ARIMA$[P, D, Q]$ estacional

Este es un modelo que sugiere una manera de incorporar un patrón estacional en la estructura del modelo ARIMA. Como antes, supondremos que la longitud del periodo estacional es el entero $s \geqslant 2$. Se trata de una extensión del caso ARIMA(p, d, q) y recibe el nombre de SARIMA por el término en inglés *Seasonal ARIMA*.

La definición de este modelo se realiza de la misma manera en que se hizo para definir el modelo ARMA estacional en la página 122. Recordemos nuevamente que los operadores no estacionales son:

$$\begin{aligned} \theta(B) &= 1 + \beta_1 B + \cdots + \beta_q B^q, \\ \phi(B) &= 1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_p B^p. \end{aligned}$$

Estos son los operadores que definen la parte de promedios móviles de orden q y la parte regresiva de orden p, respectivamente. Por otro lado, los operadores estacionales son:

$$\begin{aligned} \Theta(B^s) &= 1 + \beta'_1 B^s + \beta'_2 B^{2s} + \cdots + \beta'_Q B^{Qs}, \\ \Phi(B^s) &= 1 - \alpha'_1 B^s - \alpha'_2 B^{2s} - \cdots - \alpha'_P B^{Ps}. \end{aligned}$$

Estos operadores se encargan de establecer la dependencia de los promedios móviles en Q periodos estacionales previos, y la dependencia de la parte regresiva en los P periodos estacionales previos, respectivamente. Estos operadores se aplican de manera multiplicativa:

$$\Theta(B^s)\theta(B) \quad y \quad \Phi(B^s)\phi(B).$$

Como estos operadores conmutan, el producto se puede escribir en el otro orden sin cambio en el resultado.

Además, tenemos el operador diferencia no estacional $(1-B)^d$ y el operador diferencia estacional $(1 - B^s)^D$, en donde $d \geqslant 0$ y $D \geqslant 0$ son dos números enteros. Con estos elementos es posible ahora establecer una definición del modelo SARIMA.

Definición 3.21 *Sean $d \geqslant 0$ y $D \geqslant 0$ dos números enteros. Una serie de tiempo $\{X_t : t \in \mathbb{Z}\}$ sigue un modelo estacional ARIMA de periodo $s \geqslant 2$, y se denota por* ARIMA$(p, d, q)[P, D, Q]_s$*, si la serie $\{W_t : t \in \mathbb{Z}\}$ dada por*

$$W_t = (1 - B^s)^D (1 - B)^d X_t, \quad t \in \mathbb{Z}.$$

sigue un modelo ARMA *de la forma*

$$\Phi(B^s)\phi(B)W_t = \Theta(B^s)\theta(B)Z_t, \quad t \in \mathbb{Z}. \tag{3.48}$$

Observe que el proceso integrado $\{W_t : t \in \mathbb{Z}\}$ se puede escribir como

$$W_t = (\nabla_s)^D \nabla^d X_t, \quad t \in \mathbb{Z},$$

en donde los dos operadores diferencia indicados conmutan y su multiplicación se puede escribir en cualquier orden. No es difícil darse cuenta que la multiplicación de todos los operadores involucrados en la definición del modelo SARIMA produce un modelo ARMA de ciertos parámetros y que, por lo tanto, todos los resultados que se aplican al modelo ARMA se heredan al modelo SARIMA.

Ejemplo 3.21 *El modelo* ARIMA$(p, d, q)[0, 0, 0]_s$ *es simplemente el modelo* ARIMA(p, d, q). ∎

Ejemplo 3.22 *El modelo* ARIMA$(0, 0, 0)[0, 0, Q]_s$ *es el modelo de promedios móviles estacional* MA$[Q]$ *especificado en la definición 3.6 de la página 89. Explícitamente,*

$$X_t = Z_t + \beta_1' Z_{t-s} + \beta_2' Z_{t-2s} + \cdots + \beta_Q' Z_{t-Qs}, \quad t \in \mathbb{Z}.$$

∎

Ejemplo 3.23 *El modelo* ARIMA$(0, 0, 0)[P, 0, 0]_s$ *es el modelo autorregresivo estacional* AR$[P]$ *especificado en la definición 3.12 de la página 109. Explícitamente,*

$$X_t = \alpha_1' X_{t-s} + \alpha_2' X_{t-2s} + \cdots + \alpha_P' X_{t-Ps} + Z_t, \quad t \in \mathbb{Z}.$$

∎

Ejemplo 3.24 *Sea* $\{X_t : t \in \mathbb{Z}\}$ *una serie de tiempo que sigue el modelo estacional* ARIMA$(1,0,0)[1,0,1]_s$*, es decir,* $(p,d,q) = (1,0,0)$ *y* $[P,D,Q] = [1,0,1]$*. Como* $d = 0$ *y* $D = 0$*, la serie se puede escribir como*

$$\Phi(B^s)\phi(B)X_t = \Theta(B^s)\theta(B)Z_t, \quad t \in \mathbb{Z}.$$

Como $p = 1$ *y* $q = 0$*,*

$$\Phi(B^s)(1 - \alpha_1 B)X_t = \Theta(B^s)Z_t, \quad t \in \mathbb{Z}.$$

Como $P = 1$ *y* $Q = 1$*,*

$$(1 - \alpha_1' B^s)(X_t - \alpha_1 X_{t-1}) = (1 + \beta_1' B^s)Z_t, \quad t \in \mathbb{Z}.$$

Es decir, el modelo indicado se puede escribir de manera explícita como aparece en la ecuación siguiente. En esta expresión se señalan los componentes de la serie,

$$\underbrace{(X_t - \alpha_1 X_{t-1})}_{\substack{\text{Autorregresión} \\ \text{no estacional}}} - \alpha_1' \underbrace{(X_{t-s} - \alpha_1 X_{t-s-1})}_{\substack{\text{Autorregresión} \\ \text{estacional}}} = \underbrace{Z_t + \beta_1' Z_{t-s}}_{\substack{\text{Promedio móvil} \\ \text{estacional}}}, \quad t \in \mathbb{Z}.$$

■

3.5. Resumen sobre causalidad e invertibilidad

La Tabla 3.3 muestra un resumen de los conceptos de causalidad e invertibilidad para los modelos MA(q), AR(p) y ARMA(p,q). Recordemos que de estos tres modelos, el más general es ARMA(p,q), el cual se puede escribir como

$$\phi(B)\, X_t = \theta(B)\, Z_t, \quad t \in \mathbb{Z},$$

en donde el operador de promedios móviles es

$$\theta(B) = \sum_{k=0}^{q} \beta_k\, B^k = \beta_0 + \beta_1\, B + \cdots + \beta_q\, B^q,$$

con $\beta_0 = 1$. El proceso es invertible cuando $\theta(B)$ es invertible, y esto ocurre si, y sólo si, todas las raíces del polinomio $\theta(x)$ quedan fuera del círculo

unitario, es decir, si tienen norma mayor a 1. Por otro lado, el operador autorregresivo es

$$\phi(B) = 1 - \sum_{k=1}^{p} \alpha_k B^k = 1 - \alpha_1 B - \cdots - \alpha_p B^p.$$

El proceso es causal cuando $\phi(B)$ es invertible, lo cual ocurre si, y sólo si, todas las raíces del polinomio $\phi(x)$ quedan fuera del círculo unitario, es decir, si tienen norma mayor a 1. La causalidad garantiza que el proceso sea estacionario. Las constantes $\pi_0, \pi_1, \ldots$ que aparecen en las sumas infinitas de la tabla son distintas para cada modelo. Como hemos visto, estas representaciones de los modelos son útiles para encontrar algunas de sus propiedades matemáticas.

Modelo	CAUSALIDAD $\phi(B)$ invertible X_t en términos de $\{Z_t\}$ (Implica estacionariedad)	INVERTIBILIDAD $\theta(B)$ invertible Z_t en términos de $\{X_t\}$
MA(q): $X_t = \theta(B)\, Z_t$	✓ $X_t = \theta(B)\, Z_t$	$Z_t = (\theta(B))^{-1}\, X_t$ $= \sum_{k=0}^{\infty} \pi_k\, X_{t-k}$
AR(p): $\phi(B)\, X_t = Z_t$	$X_t = (\phi(B))^{-1}\, Z_t$ $= \sum_{k=0}^{\infty} \pi_k\, Z_{t-k}$	✓ $\phi(B)\, X_t = Z_t$
ARMA(p, q): $\phi(B)\, X_t = \theta(B)\, Z_t$	$X_t = (\phi(B))^{-1}\, \theta(B)\, Z_t$ $= \sum_{k=0}^{\infty} \pi_k\, Z_{t-k}$	$Z_t = (\theta(B))^{-1}\, \phi(B)\, X_t$ $= \sum_{k=0}^{\infty} \pi_k\, X_{t-k}$

Tabla 3.3: Resumen de los conceptos de causalidad e invertibilidad.

3.6. Función de autocorrelación parcial (PACF)

La función de autocorrelación parcial se obtiene al aplicar regresiones lineales a un modelo autorregresivo que tendrá un número de parámetros igual al argumento de la función. A continuación se presenta la definición formal.

Definición 3.22 *Supongamos que $\mu, \phi_{k1}, \ldots, \phi_{kk}$ son parámetros desconocidos. El coeficiente de autocorrelación parcial en el rezago $k > 0$, denotado por ϕ_k, se define como el parámetro ϕ_{kk} en la regresión lineal*

$$X_t = \mu + \phi_{k1}X_{t-1} + \phi_{k2}X_{t-2} + \cdots + \phi_{kk}X_{t-k} + \epsilon_t, \tag{3.49}$$

en donde $\{\epsilon_t\}$ es una variable aleatoria independiente de X_{t-i} para todo $i \geqslant 1$. La función de autocorrelación parcial (PACF) muestral se define como la relación $k \mapsto \widehat{\phi}_k$, para $k = 1, 2, \ldots$

El coeficiente de autocorrelación parcial en el rezago k, ϕ_k, mide la correlación entre X_t y X_{t+k} después de eliminar los efectos de $X_{t+1}, \ldots, X_{t+k-1}$. En otras palabras, ϕ_k mide la correlación simple entre X_t y X_{t-k} después de extraer la influencia de los rezagos intermedios, $X_{t-1}, \ldots, X_{t-(k-1)}$ de la variable.

Por lo anterior, una forma de estimar los valores $\widehat{\phi}_k$, es utilizar el método de mínimos cuadrados en la regresión dada por (3.49) de X_t respecto a los rezagos $X_{t-1}, \ldots, X_{t-k}$ y un término constante.

Así, para estimar ϕ_1 debemos estimar la regresión

$$X_t = \phi_{11}X_{t-1} + \epsilon_t,$$

entonces $\widehat{\phi}_1 := \widehat{\phi}_{11}$. Para estimar ϕ_2 debemos estimar la regresión

$$X_t = \phi_{21}X_{t-1} + \phi_{22}X_{t-2} + \epsilon_t,$$

en este caso, $\widehat{\phi}_2 := \widehat{\phi}_{22}$. Para estimar ϕ_3 debemos estimar la regresión

$$X_t = \phi_{31}X_{t-1} + \phi_{32}X_{t-2} + \phi_{33}X_{t-3} + \epsilon_t,$$

por lo tanto, $\widehat{\phi}_3 := \widehat{\phi}_{33}$, y así sucesivamente. Siempre se estima la regresión lineal múltiple a pesar de que nos interesa conocer únicamente la estimación del último coeficiente. Por lo anterior, comúnmente el cálculo de estos valores se realiza con la ayuda de un programa de cómputo estadístico.

La serie de coeficientes $\widehat{\phi}_1, \widehat{\phi}_2, \ldots, \widehat{\phi}_k$ estimados con las observaciones de la serie de tiempo conforman la función de autocorrelación parcial muestral de orden k, PACF muestral. Bajo ciertas condiciones, $\widehat{\phi}_k$ se distribuye asintóticamente como una variable aleatoria $N(0, 1/n)$. Queda como ejercicio para el lector investigar tales condiciones.

3.7. Selección, ajuste y evaluación de modelos

Sea $x_1, x_2, \ldots, x_n$ una serie de tiempo observada. Habiendo revisado algunos modelos matemáticos de series de tiempo, el problema natural que se plantea es el de encontrar un modelo $\{X_t : t \in \mathbb{Z}\}$ que represente de mejor manera los datos obtenidos. Este es un problema difícil que, en general, no tiene una repuesta única. La posible respuesta que se puede dar depende de varios factores como, por ejemplo, la definición de lo que se entienda por una mejor representación de los datos, del número de observaciones disponibles, de las propiedades estadísticas de las observaciones, de las hipótesis inherentes del modelo a ajustar, y depende también del uso que se hará del modelo.

Más aún, el proceso de selección y ajuste de modelos puede ser dinámico cuando se considera que nuevas observaciones pueden ser agregadas al conjunto de datos. Esto sucede con mucha frecuencia y la nueva información puede modificar las condiciones bajo las cuales se tomaron las decisiones anteriores. En general, se pueden distinguirse 3 etapas en el proceso de adecuación de un modelo para un conjunto de datos observados, las cuales se revisan brevemente a continuación.

Selección del modelo

En esta etapa se debe especificar el modelo matemático a utilizar. Seleccionar un modelo que represente adecuadamente un conjunto de datos

observados no es una tarea fácil, especialmente si se considera que el conjunto de modelos disponibles es grande, en donde algunos de ellos tienen características similares. En la práctica, es de mucha ayuda comenzar haciendo una inspección visual del gráfico de la serie, así como de los correlogramas ACF y PACF. En general, se busca que las características de la serie observada sean similares a las características del modelo matemático.

Estimación de los parámetros del modelo

Esta etapa no es tan complicada, se le llama también ajuste del modelo. Consiste en estimar los parámetros que aparecen en las expresiones del modelo seleccionado. Por ejemplo, si se desea ajustar un modelo AR(1) dado por

$$X_t = \alpha_1 X_{t-1} + Z_t, \quad t \in \mathbb{Z},$$

el problema es estimar el parámetro desconocido α_1, pero también la varianza σ^2 del proceso de ruido $\{Z_t : t \in \mathbb{Z}\}$. Por supuesto, el problema se puede complicar si se decide ajustar un modelo en el que aparece una gran cantidad de parámetros. Más adelante se muestran algunos ejemplos de estimaciones puntuales de estos parámetros.

Verificación del modelo

Esta etapa es una continuación de la primera. Aquí se analiza si la serie observada cumple las hipótesis inherentes al modelo seleccionado. Por ejemplo, si el modelo contempla un proceso de ruido $\{Z_t : t \in \mathbb{Z}\}$, se puede verificar si efectivamente, el proceso de residuos cumple propiedades como las siguientes: tener media cero, tener varianza constante, presentar la propiedad de independencia entre las variables, o tener distribución normal cuando tal hipótesis es adoptada. En general, todos los modelos presentan deficiencias en su verificación pues difícilmente los datos observados siguen el comportamiento teórico de uno de los modelos. El objetivo es obtener el mejor modelo dentro de un programa flexible de selección.

3.8. Estimación de parámetros

En esta sección utilizaremos el método de mínimos cuadrados para estimar los parámetros de los modelos MA, AR y ARMA, cuando la situación lo permite. Con mucha frecuencia es necesario aplicar algún método numérico para realizar estas estimaciones.

Estimación de parámetros para un modelo AR(p)

Sean $x_1, \ldots, x_n$ observaciones de una serie de tiempo $\{X_t : t \in \mathbb{Z}\}$ para la cual se ha considerado adecuado ajustar un modelo AR(p) de media μ. Consideraremos principalmente el caso $p = 1$. El modelo AR(1) de media μ se define como

$$X_t - \mu = \alpha_1(X_{t-1} - \mu) + Z_t, \quad t \in \mathbb{Z},$$

en donde los parámetros μ y α_1 son desconocidos. Supondremos que $-1 < \alpha_1 < 1$ y definiremos a continuación las medias muestrales de las primeras y las últimas $n-1$ observaciones:

$$\bar{x}_{(1)} := \frac{1}{n-1}\sum_{i=1}^{n-1} x_i,$$

$$\bar{x}_{(2)} := \frac{1}{n-1}\sum_{i=2}^{n} x_i.$$

Estos promedios parciales aparecen en el siguiente resultado.

Proposición 3.12 *Los estimadores por mínimos cuadrados para los parámetros μ y α_1 en un modelo* AR(1) *de media μ están dados por*

$$\widehat{\mu} = \frac{\bar{x}_{(2)} - \widehat{\alpha}_1 \bar{x}_{(1)}}{1 - \widehat{\alpha}_1}, \tag{3.50}$$

$$\widehat{\alpha}_1 = \frac{\sum_{i=1}^{n-1}(x_i - \widehat{\mu})(x_{i+1} - \widehat{\mu})}{\sum_{i=1}^{n-1}(x_i - \widehat{\mu})^2}. \tag{3.51}$$

Demostración. La función a minimizar es la suma de los residuos al cuadrado, es decir,

$$S = \sum_{i=2}^{n} z_i^2 = \sum_{i=2}^{n} [x_i - \mu - \alpha_1(x_{i-1} - \mu)]^2.$$

Derivando respecto de μ y de α_1 e igualando a cero, se obtienen las ecuaciones

$$\begin{aligned} \frac{\partial S}{\partial \mu} &= 2 \sum_{i=2}^{n} [x_i - \mu - \alpha_1(x_{i-1} - \mu)](-1 + \alpha_1), \\ \frac{\partial S}{\partial \alpha_1} &= 2 \sum_{i=2}^{n} [x_i - \mu - \alpha_1(x_{i-1} - \mu)](-(x_{i-1} - \mu)). \end{aligned}$$

Distribuyendo las sumas y simplificando se encuentran las ecuaciones

$$\begin{aligned} [(n-1)\bar{x}_{(2)} - (1-\alpha_1)(n-1)\mu - \alpha_1(n-1)\bar{x}_{(1)}](1-\alpha_1) &= 0, \\ -\sum_{i=2}^{n} (x_i - \mu)(x_{i-1} - \mu) + \alpha_1 \sum_{i=2}^{n} (x_{i-1} - \mu)^2 &= 0. \end{aligned}$$

De la primera ecuación se obtiene la expresión para el estimador $\widehat{\mu}$ y de la segunda ecuación se obtiene la fórmula buscada para $\widehat{\alpha}_1$. ∎

En la sección de ejercicios se pide demostrar que los estimadores por máxima verosimilitud para α_1 y μ coinciden con los estimadores por mínimos cuadrados (3.50) y (3.51) encontrados en el caso en que $Z_t \sim \text{N}(0, \sigma^2)$. Es decir, que en este caso, los dos métodos de estimación producen los mismos resultados.

En general, es posible llevar a cabo un análisis similar de estimación para un modelo AR(p) de media μ, en cuyo caso se tendrían los siguientes $p+1$ parámetros a estimar: $\mu, \alpha_1, \ldots, \alpha_p$.

Regresando al caso $p = 1$, observe que las fórmulas para $\widehat{\mu}$ y $\widehat{\alpha}_1$ no son explícitas. Estos valores son una solución de las dos ecuaciones (3.50) y (3.51).

Sin embargo, existen las siguientes aproximaciones que resultan interesantes y que simplifican el cálculo de la solución: como $\bar{x}_{(1)} \approx \bar{x}_{(2)} \approx \bar{x}$, las expresiones (3.50) y (3.51) se pueden escribir de manera aproximada en la forma siguiente:

$$\begin{aligned} \widehat{\mu} &= \bar{x}, \\ \widehat{\alpha}_1 &= \frac{\sum_{i=1}^{n-1}(x_i - \bar{x})(x_{i+1} - \bar{x})}{\sum_{i=1}^{n-1}(x_i - \bar{x})^2}. \end{aligned}$$

Más aún, el denominador de $\widehat{\alpha}_1$ puede ser aproximado por $\sum_{i=1}^{n}(x_i - \bar{x})^2$, encontrando así, como valor aproximado, el coeficiente de correlación muestral con desfase $\tau = 1$, esto es,

$$\widehat{\alpha}_1 \approx \frac{\mathrm{cov}_x(1)}{\mathrm{cov}_x(0)} = r(1).$$

Llegamos así al siguiente resultado.

Proposición 3.13 *Sean $x_1, \ldots, x_n$ observaciones de una serie de tiempo $\{X_t : t \in \mathbb{Z}\}$ que sigue el modelo* AR(1) *de media μ. Una aproximación para la estimación de los parámetros α_1 y μ es*

$$\begin{aligned} \widehat{\mu} &\approx \bar{x}, \\ \widehat{\alpha}_1 &\approx r(1). \end{aligned}$$

Anteriormente se encontró que, para un modelo AR(1), $\rho(1) = \alpha_1$. Por otro lado, $r(1)$ es una aproximación para $\rho(1)$. Así, el estimador encontrado tiene sentido pues $\widehat{\alpha}_1 \approx r(1) \approx \rho(1) = \alpha_1$.

Estimación de parámetros para un modelo MA(q)

Sean $x_1, \ldots, x_n$ observaciones de una serie de tiempo $\{X_t : t \in \mathbb{Z}\}$ para la cual se ha considerado adecuado ajustar un modelo MA(q) de media μ.

Consideraremos principalmente el caso $q = 1$. Este modelo está definido por la identidad

$$X_t - \mu = Z_t + \beta_1 Z_{t-1}, \quad t \in \mathbb{Z},$$

en donde μ y β_1 son parámetros desconocidos. Se buscan valores para μ y β_1 que minimicen la función suma de los residuos al cuadrado, es decir, que minimicen la función

$$S = \sum_{i=1}^{n} z_i^2 = \sum_{i=1}^{n} (x_i - \mu - \beta_1 z_{i-1})^2.$$

La situación es ahora diferente, pues debido a las características del modelo, esta función no queda expresada únicamente en términos de las observaciones $x_1, \ldots, x_n$ y los parámetros μ y β_1. Aparecen también los valores z_i del proceso de ruido. Una posible forma numérica de resolver este problema de estimación es por inspección evaluando la función S para distintos valores de μ y β_1. El procedimiento es el siguiente:

- Sean μ y β_1 valores factibles y seleccionados.

- Suponga $z_0 = 1$. Se calcula la sucesión de valores $z_1, \ldots, z_n$ usando la expresión

$$z_i = x_i - \mu - \beta_1 z_{i-1}, \quad \text{para } i = 1, 2, \ldots, n.$$

- Ahora se puede calcular el valor de $S = \sum_{i=1}^{n} z_i^2$ asociado al par (μ, β_1) previamente seleccionado.

Sobre la región acotada de posibles valores para (μ, β_1) se puede crear una malla con los puntos a evaluar. Véase la figura 3.5. Se escoge el punto $(\widehat{\mu}, \widehat{\beta}_1)$ que produzca la evaluación más pequeña.

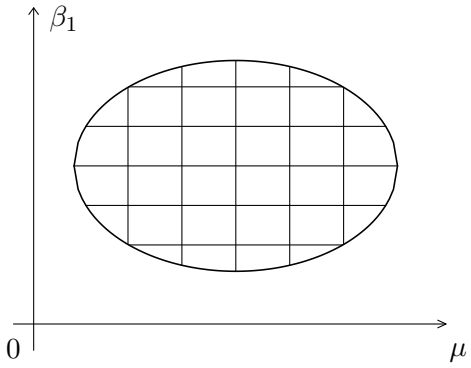

Figura 3.5: Malla de valores para μ y β_1.

Este procedimiento numérico se puede aplicar también para cualquier orden del modelo MA(q). Por ejemplo, para $q = 2$ tenemos el modelo

$$X_t - \mu = Z_t + \beta_1 Z_{t-1} + \beta_2 Z_{t-2}, \quad t \in \mathbb{Z}.$$

Dados valores particulares de μ, β_1 y β_2, a partir de $z_0 = z_1 = 0$ se generan los valores

$$z_i = x_i - \mu - \beta_1 z_{i-1} - \beta_2 z_{i-2}, \quad \text{para } i = 2, \ldots, n,$$

y con ellos se calcula el valor de la función $S = \sum_{i=1}^{n} z_i^2$. Sobre alguna región acotada para los valores de μ, β_1 y β_2 se coloca una malla, y para cada terna (μ, β_1, β_2) se calcula S. La terna $(\hat{\mu}, \hat{\beta}_1, \hat{\beta}_2)$ que produzca el valor más pequeño para S es la estimación que se busca.

Estimación de parámetros para un modelo ARMA(p, q)

Sean $x_1, \ldots, x_n$ observaciones de una serie de tiempo $\{X_t : t \in \mathbb{Z}\}$ para la cual se ha considerado adecuado ajustar un modelo ARMA(p, q) de media desconocida μ pero donde los parámetros p y q se han seleccionado con anterioridad. En esta sección sólo mencionaremos que el algoritmo numérico de exploración mencionado antes para la estimación de los parámetros puede

aplicarse también para este método. Por ejemplo, un modelo ARMA$(1,1)$ de media μ es de la forma

$$X_t - \mu = \alpha_1(X_{t-1} - \mu) + Z_t + \beta_1 Z_{t-1}, \quad t \in \mathbb{Z}.$$

Se toman ciertos valores de prueba para μ, α_1 y β_1, y se calculan los residuos de manera recursiva usando la siguiente fórmula

$$z_i = (x_i - \mu) - \alpha_1(x_{i-1} - \mu) - \beta_1 z_{i-1}, \quad i = 1, 2, \ldots, n,$$

en donde $x_1, x_2, \ldots, x_n$ son los valores observados de la serie y $z_0 := 0$. De esta manera se puede calcular el valor

$$S = \sum_{i=1}^{n} z_i^2 = \sum_{i=1}^{n} (x_i - \mu - \alpha_1(x_{i-1} - \mu) - \beta_1 z_{i-1})^2.$$

Este procedimiento se repite para cualquier otro conjunto de valores posibles para μ, α_1, β_1, y por inspección se obtienen valores aproximados para $\widehat{\mu}$, $\widehat{\alpha}_1$, $\widehat{\beta}_1$ que minimizan el error cuadrático medio. El mismo procedimiento numérico se puede aplicar para el modelo general ARMA(p,q), en donde el uso de una computadora es necesario. Finalmente, mencionamos que si se supone una distribución para Z_t, por ejemplo $\mathrm{N}(0, \sigma^2)$, los parámetros se pueden estimar también por máxima verosimilitud. Es importante señalar que la optimización requiere normalmente de métodos numéricos para encontrar estimaciones aproximadas. Lo anterior se debe a que el problema de optimización puede derivar en la necesidad de resolver algún sistema de ecuaciones no lineal cuya solución analítica puede ser difícil de obtener. Afortunadamente, en la mayoría del software estadístico, así como en varios lenguajes de programación existen librerías o paquetes con funciones especiales para realizar las estimaciones de los parámetro, e internamente usan algoritmos numéricos, tanto para realizar estimaciones por mínimos cuadrados como por el método de máxima verosimilitud. En el capítulo siguiente, se expone en forma breve este último método de estimación.

3.9. Ejercicios

Modelo MA

53. Determine el orden q de los siguientes procesos MA(q) y diga si son invertibles:

a) $X_t = Z_t$.
b) $X_t = Z_t - 5\,Z_{t-1}$.
c) $X_t = Z_t - Z_{t-1} + Z_{t-2}$.
d) $X_t = Z_t + 2\,Z_{t-4}$.
e) $X_t = Z_t - 7\,Z_{t-2}$.
f) $X_t = Z_t + Z_{t-20}$.

54. Considere la serie de tiempo $\{X_t : t \in \mathbb{Z}\}$ dada por

$$X_t = (Z_t + Z_{t-1})/2, \quad t \in \mathbb{Z},$$

en donde $\{Z_t : t \in \mathbb{Z}\}$ es el proceso de ruido blanco $\text{WN}(0, \sigma^2)$ con $\sigma^2 = 1$. Demuestre directamente que:

a) $E(X_t) = 0$.
b) $\text{Var}(X_t) = 1/2$.
c) $\gamma(0) = 1/2$.
d) $\gamma(1) = 1/4$.
e) $\gamma(\tau) = 0, \quad \tau \geqslant 2$.
f) $\rho(0) = 1$.
g) $\rho(1) = 1/2$.
h) $\rho(\tau) = 0, \quad \tau \geqslant 2$.

55. Considere el modelo MA(1) dado por

$$X_t = \beta_0 Z_t + \beta_1 Z_{t-1}, \quad t \in \mathbb{Z},$$

en donde $\beta_0 = 1$ y $\beta_1 \neq 1$. Demuestre directamente que:

a) $E(X_t) = 0$.
b) $\text{Var}(X_t) = \sigma^2(\beta_0^2 + \beta_1^2)$.
c) $\gamma(0) = \sigma^2(\beta_0^2 + \beta_1^2)$.
d) $\gamma(1) = \sigma^2\beta_0\beta_1$.
e) $\gamma(\tau) = 0, \quad \tau \geqslant 2$.
f) $\rho(0) = 1$.
g) $\rho(1) = \beta_0\beta_1/(\beta_0^2 + \beta_1^2)$.
h) $-1 \leqslant \rho(1) \leqslant 1$.
i) $\rho(\tau) = 0, \quad \tau \geqslant 2$.

56. Demuestre que la expresión encontrada para la función de autocorrelación del proceso MA(q) en la proposición 3.2 es tal que

$$-1 \leqslant \rho(\tau) \leqslant 1.$$

57. *Convergencia en media cuadrática.*
Sea $\{X_t : t \in \mathbb{Z}\}$ una serie de tiempo que sigue el modelo de promedios móviles de orden infinito definido en (3.2). Demuestre que tal

suma infinita de variables aleatorias es convergente en media cuadrática cuando

$$\sum_{k=0}^{\infty} \beta_k^2 < \infty.$$

58. *Comportamiento monótono.*
Considere el modelo MA(q) en donde los coeficientes $\beta_1, \ldots, \beta_q$ son positivos. Demuestre que las funciones de autocovarianza $\gamma(\tau)$ y de autocorrelación $\rho(\tau)$ son decrecientes para $\tau \geqslant 0$.

59. Considere el modelo MA(1) especificado abajo, en donde $\beta_0 = 1$. Demuestre que si $Z_t \sim \mathrm{N}(0, \sigma^2)$, entonces $X_t \sim \mathrm{N}(0, \sigma^2\,(\beta_0^2 + \beta_1^2))$.

$$X_t = \beta_0\, Z_t + \beta_1\, Z_{t-1}, \quad t \in \mathbb{Z}.$$

60. *Experimentos en computadora.*
Considere el modelo MA(q) en donde el proceso aleatorio puro $\{Z_t : t \in \mathbb{Z}\}$ es tal que $P(Z_t = 1) = P(Z_t = -1) = 1/2$. Con ayuda de una computadora obtenga una trayectoria del modelo especificado abajo. Muestre, además, el correlograma correspondiente.

   a) $X_t = Z_t + 4Z_{t-6}, \quad t \geqslant 0.$
   b) $X_t = Z_t + Z_{t-1} + Z_{t-2} + Z_{t-3}, \quad t \geqslant 0.$
   c) $X_t = Z_t + 10Z_{t-20}, \quad t \geqslant 0.$

61. Sea $\{X_t : t \in \mathbb{Z}\}$ la serie de tiempo especificada abajo. Observe que los coeficientes son todos iguales y que suman 1.

$$X_t = \frac{1}{q+1}\, Z_t + \frac{1}{q+1}\, Z_{t-1} + \cdots + \frac{1}{q+1}\, Z_{t-q}, \quad t \in \mathbb{Z}.$$

   a) Demuestre que la función de autocorrelación es

$$\rho(\tau) = \begin{cases} \dfrac{q+1-\tau}{q+1} & \text{si } \tau = 0, 1, \ldots, q, \\ 0 & \text{si } \tau \geqslant q+1. \end{cases}$$

   b) Grafique y describa con palabras el comportamiento del correlograma para esta serie de tiempo.

62. Para el proceso MA(1) indicado abajo, se señaló que $\rho(\tau) = \beta_0\,\beta_1/(\beta_0^2 + \beta_1^2)$, para $\tau = \pm 1$, con $\beta_0 = 1$. Demuestre que esta cantidad es, efectivamente, menor o igual a 1 en valor absoluto.

$$X_t = \beta_0 Z_t + \beta_1 Z_{t-1}, \quad t \in \mathbb{Z}.$$

63. Determine si el proceso MA(2) especificado abajo es invertible.

$$X_t = Z_t + Z_{t-1} + Z_{t-2}, \quad t \in \mathbb{Z}.$$

64. Considere el proceso MA(2) especificado abajo. Encuentre condiciones sobre los coeficientes β_1 y β_2 para que el proceso sea invertible.

$$X_t = Z_t + \beta_1 Z_{t-1} + \beta_2 Z_{t-2}, \quad t \in \mathbb{Z}.$$

65. Sea $q \geqslant 1$ un entero. Determine los valores de la constante $\beta_q \neq 0$ para que el siguiente proceso MA(q) sea invertible. Obtenga, además, la función de autocorrelación y el correlograma.

$$X_t = Z_t + \beta_q\, Z_{t-q}, \quad t \in \mathbb{Z}.$$

66. Sea $\{X_t : t \in \mathbb{Z}\}$ la serie de tiempo dada por

$$X_t = \frac{1}{2} Z_t + \frac{1}{2} Z_{t-1}, \quad t \in \mathbb{Z}.$$

   a) Compruebe que el proceso es estacionario.

   b) Encuentre la función de autocovarianza $\gamma(\tau), \quad \tau \geqslant 0$.

   c) Encuentre y grafique la función de autocorrelación $\rho(\tau), \quad \tau \geqslant 0$.

67. *Dos series* MA(1) *distintas con la misma función de autocorrelación.* Sea $\beta \neq 0$ una constante y considere las dos series de promedios móviles que aparecen abajo. Demuestre que las dos series tienen la misma función de autocorrelación. Compruebe que si una de ellas es invertible, la otra no lo es.

$$\begin{aligned} X_t &= Z_t + \beta\, Z_{t-1}, \quad t \in \mathbb{Z}, \\ X_t &= Z_t + \frac{1}{\beta}\, Z_{t-1}, \quad t \in \mathbb{Z}. \end{aligned}$$

68. *Otro ejemplo de dos series de tiempo distintas con la misma función de autocorrelación.*

Considere las dos series de tiempo especificadas abajo. Observe que un modelo tiene los coeficientes intercambiados respecto del otro modelo. Demuestre que las dos series tienen la misma función de autocorrelación. Compruebe que si una de ellas es invertible, la otra no lo es.

$$\begin{aligned} X_t &= \beta_0 Z_t + \beta_1 Z_{t-1}, \quad t \in \mathbb{Z}, \\ X_t &= \beta_1 Z_t + \beta_0 Z_{t-1}, \quad t \in \mathbb{Z}. \end{aligned}$$

69. *Unicidad del modelo* MA(1) *dada su función de autocorrelación bajo la hipótesis de invertibilidad.*

Considere los siguientes dos procesos MA(1) invertibles:

$$\begin{aligned} X_t &= Z_t + \beta_1 Z_{t-1}, \quad t \in \mathbb{Z}, \\ Y_t &= Z_t + \beta_1' Z_{t-1}, \quad t \in \mathbb{Z}, \end{aligned}$$

con operadores asociados $\theta(B)$ y $\theta'(B)$, y funciones de autocorrelación $\rho(\tau)$ y $\rho'(\tau)$, respectivamente. Demuestre que:

a) $\rho(\tau) = \rho'(\tau) \Leftrightarrow \beta_1 = \beta_1'$.

b) $\theta(B)\,(\theta'(B))^{-1} = 1$.

c) $\theta'(B)\,(\theta(B))^{-1} = 1$.

Concluya que las dos series son la misma ($\beta_1 = \beta_1'$), o bien una es el negativo de la otra ($\beta_1 = -\beta_1'$). Cuando el proceso $\{Z_t : t \in \mathbb{Z}\}$ es el ruido blanco WN$(0, \sigma^2)$, el cambio de signo de Z_t es irrelevante y la unicidad se cumple en el sentido de la igualdad en distribución, esto es, las variables X_t y Y_t tienen la misma distribución normal.

70. *Propiedades del modelo lineal generalizado.*

Sea $\{X_t : t \in \mathbb{Z}\}$ una serie que sigue el modelo lineal generalizado (3.2) bajo la condición (3.3). Demuestre que:

a) $E(X_t) = 0$.

b) $\text{Var}(X_t) = \sigma^2 \sum_{k=0}^{\infty} \beta_k^2$.

c) $\gamma(\tau) = \text{Cov}\,(X_t, X_{t+\tau}) = \sigma^2 \sum_{k=0}^{\infty} \beta_k\, \beta_{k+\tau}, \quad \text{para } \tau = 0, 1, \ldots$

d) $\{X_t : t \in \mathbb{Z}\}$ es estacionario.

e) $\rho(\tau) = \text{Corr}(X_t, X_{t+\tau}) = \sum_{k=0}^{\infty} \beta_k\, \beta_{k+\tau} / \sum_{k=0}^{\infty} \beta_k^2, \quad \text{para } \tau = 0, 1, \ldots$

71. *Modelo* MA(∞) *de media* μ.

Sea $\{X_t : t \in \mathbb{Z}\}$ una serie que sigue el modelo lineal generalizado (3.2) bajo la condición (3.3). Sea μ una constante y defina la serie $Y_t = X_t - \mu$, para $t \in \mathbb{Z}$. Es evidente que $E(Y_t) = \mu$. Compruebe que las propiedades $(b), (c), (d)$ y (e) del ejercicio anterior permanecen sin cambio, es decir, el proceso es también estacionario y la varianza, covarianza y correlación son las mismas.

72. Sea $\{X_t : t \in \mathbb{Z}\}$ una serie que sigue el modelo lineal generalizado (3.2) bajo la condición (3.3), en donde los coeficientes son tales que $\beta_k = \alpha^k$ para $k = 0, 1, \ldots$, en donde $\alpha \neq 0$ satisface $-1 < \alpha < 1$. Demuestre que:

a) $E(X_t) = 0$.

b) $\text{Var}(X_t) = \dfrac{\sigma^2}{1 - \alpha^2}$.

c) $\gamma(\tau) = \text{Cov}\,(X_t, X_{t+\tau}) = \sigma^2 \dfrac{\alpha^\tau}{1 - \alpha^2}, \quad \text{para } \tau = 0, 1, \ldots$

d) $\{X_t : t \in \mathbb{Z}\}$ es estacionario.

e) $\rho(\tau) = \text{Corr}(X_t, X_{t+\tau}) = \alpha^\tau, \quad \text{para } \tau = 0, 1, \ldots$

Modelo AR

73. Determine el orden p de los siguientes procesos AR(p) y verifique si son causales:

a) $X_t = 2X_{t-1} + Z_t, \quad t \in \mathbb{Z}$.

b) $X_t = X_{t-1} + X_{t-2} + Z_t, \quad t \in \mathbb{Z}$.

c) $X_t = X_{t-1} - X_{t-3} + Z_t, \quad t \in \mathbb{Z}$.

d) $X_t = -X_{t-1} + X_{t-2} + (1/2)X_{t-3} + Z_t, \quad t \in \mathbb{Z}$.

74. Sea $\{X_t : t \in \mathbb{Z}\}$ el proceso AR(2) dado por

$$X_t = X_{t-1} - \frac{1}{2}X_{t-2} + Z_t, \quad t \in \mathbb{Z}.$$

a) Determine si el proceso es causal.

b) Compruebe que el proceso es estacionario.

c) Plantee y resuelva las ecuaciones de Yule-Walker para $\rho(1)$ y $\rho(2)$.

d) Encuentre $\gamma(0)$ y $\gamma(1)$.

75. Considere el proceso AR(1) dado por (3.15) en donde $-1 < \alpha_1 < 1$. El operador de retraso asociado es $\phi(B) = 1 - \alpha_1 B$. Compruebe directamente que la expansión $(\phi(B))^{-1} = 1 + \alpha_1 B + \alpha_1^2 B^2 + \cdots$ satisface las identidades

$$(\phi(B))^{-1} \cdot \phi(B) = \phi(B) \cdot (\phi(B))^{-1} = 1.$$

76. Sea $\{X_t : t \in \mathbb{Z}\}$ un proceso AR(1) que es estacionario. A partir de las ecuaciones de Yule-Walker demuestre las fórmulas que aparecen abajo. Estas expresiones ya se habían encontrado de manera directa.

a) $\rho(\tau) = \alpha_1^{|\tau|}, \quad \tau \in \mathbb{Z}.$

b) $\gamma(\tau) = \dfrac{\sigma^2}{1-\alpha_1^2}\,\alpha_1^{|\tau|}, \quad \tau \in \mathbb{Z}.$

77. Sea $\{X_t : t \in \mathbb{Z}\}$ un proceso AR(2) que es estacionario. A partir de las ecuaciones de Yule-Walker demuestre que:

a) $\rho(\tau) = \begin{cases} 1 & \text{si } \tau = 0, \\ \alpha_1/(1-\alpha_2) & \text{si } \tau = 1, \\ \alpha_1\rho(\tau-1) + \alpha_2\rho(\tau-2) & \text{si } \tau \geqslant 2. \end{cases}$

b) $\gamma(\tau) = \dfrac{(1-\alpha_2)\sigma^2}{[(1-\alpha_2)^2 - \alpha_1^2](1+\alpha_2)} \cdot \rho(\tau) \quad \text{para } \tau \in \mathbb{Z}.$

78. Determine si el proceso $\{X_t : t \in \mathbb{Z}\}$ que sigue el modelo AR(2) especificado abajo es estacionario.

$$X_t = \frac{1}{3}X_{t-1} + \frac{2}{9}X_{t-2} + Z_t, \quad t \in \mathbb{Z}.$$

79. *Condición equivalente para la causalidad de un modelo AR.*
Sea $\phi(B) = 1-\alpha_1 B-\alpha_2 B^2-\cdots-\alpha_p B^p$ el operador de retraso de un proceso AR(p). Este operador es invertible cuando se cumple alguna de las siguientes dos condiciones. Demuestre que las dos afirmaciones son equivalentes:

   a) Toda raíz $z \in \mathbb{C}$ del polinomio $\phi(x) = 1-\alpha_1 x-\alpha_2 x^2-\cdots-\alpha_p x^p$ es tal que $|z| > 1$, es decir, z se encuentra fuera del círculo unitario.

   b) Toda raíz $z \in \mathbb{C}$ del polinomio $\phi_0(x) = x^p - \alpha_1 x^{p-1} - \alpha_2 x^{p-2} - \cdots - \alpha_p$ es tal que $|z| \leqslant 1$, es decir, z se encuentra dentro del círculo unitario.

80. *Función de autocorrelación para un proceso* AR(2).
Considere el modelo general AR(2) dado por

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + Z_t, \quad t \in \mathbb{Z}.$$

con $\alpha_2 \neq 0$. Recordemos que la función de autocorrelación $\rho(\tau)$ está dada por las ecuaciones de Yule-Walker

$$\rho(\tau) = \begin{cases} 1 & \text{si } \tau = 0, \\ \alpha_1\,\rho(\tau-1) + \alpha_2\,\rho(\tau-2) & \text{si } \tau = 1, 2, \ldots \end{cases}$$

   a) Substituyendo directamente, compruebe que la solución a las ecuaciones de Yule-Walker tiene la forma

$$\rho(\tau) = a_1 r_1^\tau + a_2 r_2^\tau,$$

   en donde a_1 y a_2 son ciertas constantes, y r_1 y r_2 son las raíces (reales o complejas) de la ecuación característica $r^2-\alpha_1 r-\alpha_2 = 0$ y son tales que $-1 < |r_1| < 1$ y $-1 < |r_2| < 1$. Véase el texto de R. Sedgewick y P. Flajolet [32] para conocer más acerca de este método de solución de una ecuación recursiva de orden dos.

   b) Suponga el caso en que $r_1 = r_2 = r$, con $-1 < |r| < 1$. Muestre el comportamiento del correlograma para este proceso.

81. Demuestre que el modelo AR(2) especificado abajo es estacionario si y sólo si los coeficientes α_1 y α_2 son tales que $\alpha_1+\alpha_2 < 1$, $\alpha_2-\alpha_1 < 1$

y $|\alpha_2| < 1$. Dibuje en un plano cartesiano la región que determinan estas desigualdades.

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + Z_t, \quad t \in \mathbb{Z}.$$

82. Sea $d \geqslant 1$ un entero. Demuestre que el operador $\nabla^k = (1-B)^d$ no es invertible.

83. Determine si el proceso AR(1) dado por la caminata aleatoria (1.1) es causal.

84. *Un modelo* AR(p) *particular.*
Sea $p \geqslant 1$ un entero y considere el modelo autorregresivo

$$X_t = \alpha X_{t-p} + Z_t, \quad t \in \mathbb{Z}. \tag{3.52}$$

Se trata del modelo AR(p) con $\alpha_1 = \cdots = \alpha_{p-1} = 0$ y $\alpha_p = \alpha \neq 0$. El operador asociado es $\phi(B) = 1 - \alpha B^p$.

a) Demuestre que el proceso es causal, es decir, $\phi(B)$ es invertible $\Leftrightarrow -1 < \alpha < 1$.

b) Compruebe que los coeficientes π_k de la expansión (3.24) para $(\phi(B))^{-1}$ son

$$\pi_k = \begin{cases} 1 & \text{si } \ k = 0, \\ \alpha^{k/p} & \text{si } \ k = p, 2p, \ldots, \\ 0 & \text{en otro caso.} \end{cases}$$

c) Concluya que el operador $(\phi(B))^{-1}$ toma la siguiente forma

$$(\phi(B))^{-1} = \sum_{k=0}^{\infty} \alpha^k b^{kp} = 1 + \alpha B^p + \alpha^2 B^{2p} + \cdots$$

85. *El modelo* AR(p) *de media* μ.
Sea $\{X_t : t \in \mathbb{Z}\}$ una serie de tiempo que sigue el modelo AR(p) de media μ, definido en (3.29). Sea $Y_t = X_t - \mu$, para $t \in \mathbb{Z}$. Demuestre que:

a) $\text{Var}(X_t) = \text{Var}(Y_t)$.
b) $\text{Cov}(X_t, X_{t+\tau}) = \text{Cov}(Y_t, Y_{t+\tau})$.
c) $\text{Corr}(X_t, X_{t+\tau}) = \text{Corr}(Y_t, Y_{t+\tau})$.

Modelo ARMA

86. Determine el orden (p, q) de los siguientes procesos ARMA. Para cada uno de ellos establezca si son causales e invertibles. Señale aquellos que sean estacionarios.

    a) $X_t + (6/10)X_{t-1} = Z_t + (6/5)Z_{t-1}, \quad t \in \mathbb{Z}.$

    b) $X_t + (8/5)X_{t-1} = Z_t - (2/5)Z_{t-1} + (1/25)Z_{t-2}, \quad t \in \mathbb{Z}.$

    c) $X_t - (1/5)X_{t-2} = Z_t + (1/5)Z_{t-3}, \quad t \in \mathbb{Z}.$

87. Escriba explícitamente la expresión de una serie de tiempo que sigue el modelo estacional:

    a) ARMA$(0,0)[1,0]_s$.

    b) ARMA$(0,1)[1,0]_s$.

    c) ARMA$(1,0)[0,1]_s$.

    d) ARMA$(1,0)[1,0]_s$.

    e) ARMA$(1,1)[1,1]_s$.

    f) ARMA$(2,1)[1,0]_s$.

88. *Suma de dos modelos ARMA independientes.*

    Sean $\{X_t : t \in \mathbb{Z}\}$ y $\{Y_t : t \in \mathbb{Z}\}$ dos series de tiempo independientes que siguen los modelos ARMA(p_1, q_1) y ARMA(p_2, q_2), respectivamente. Demuestre que la suma de estas series sigue un modelo ARMA(p, q), con $p \leqslant p_1 + p_2$ y $q \leqslant \max\{p_1 + q_2, p_2 + q_1\}$.

Modelo ARIMA

89. Escriba explícitamente la expresión de una serie de tiempo que sigue el modelo estacional:

    a) ARIMA$(0,0,1)[1,0,1]_s$.

    b) ARIMA$(0,1,0)[0,1,0]_s$.

    c) ARIMA$(1,0,1)[1,0,1]_s$.

    d) ARIMA$(1,1,1)[1,0,1]_s$.

    e) ARIMA$(1,0,1)[1,1,1]_s$.

90. Piense en fenómenos que por sus característica tiendan a tener alguna tendencia o bien algún comportamiento oscilatorio y consiga una serie de tiempo de tal fenómeno. Grafique la serie original y su función

ACF. A partir de la gráfica generada, ¿parece tener las características de una serie estacionaria? Después, repita los gráficos con la primera diferencia de la serie original y observe si ahora parece haber estacionariedad. Repita el mecanismo si es necesario.

Función de autocorrelación parcial (PACF)

91. Para las series que utilizó en el ejercicio 90, realice los correspondientes gráficos de la función PACF y analice si tiene un comportamiento distinto en cada caso.

92. En la sección 3.6 se establece la definición de la función PACF y sus valores estimados fueron denotados como $\widehat{\phi}_k$, para $k = 1, 2, \ldots$ Investigue bajo qué condiciones la estimación $\widehat{\phi}_k$, vista como variable aleatoria, se distribuye asintóticamente $N(0, 1/n)$.

Selección, ajuste y evaluación de modelos

93. Con lo realizado en los ejercicios 90 y 91, sugiera modelos ARMA para la serie que considere visualmente estacionaria.

Estimación de parámetros

94. Suponga que el proceso de ruido $\{Z_t : t \in \mathbb{Z}\}$ es tal que $Z_t \sim \mathrm{N}(0, \sigma^2)$. Demuestre que los estimadores por máxima verosimilitud para α_1 y μ en un modelo AR(1) de media μ coinciden con los estimadores por mínimos cuadrados (3.50) y (3.51).

95. Sean $x_1, \ldots, x_n$ observaciones de una serie de tiempo $\{X_t : t \in \mathbb{Z}\}$ que sigue un modelo AR(2) de media μ, es decir,

$$X_t - \mu = \alpha_1(X_{t-1} - \mu) + \alpha_2(X_{t-2} - \mu) + Z_t, \quad t \in \mathbb{Z},$$

en donde los parámetros μ, α_1 y α_2 son desconocidos. Suponga que α_1 y α_2 son tales que $\{X_t : t \in \mathbb{Z}\}$ es estacionario.

a) Encuentre los estimadores por mínimos cuadrados para μ, α_1 y α_2.

b) Muestre que los estimadores del inciso anterior pueden ser aproximados de la siguiente manera:

$$\begin{aligned} \widehat{\mu} &\approx \bar{x}, \\ \widehat{\alpha}_1 &\approx r(1) \cdot \frac{1 - r(2)}{1 - r^2(1)}, \\ \widehat{\alpha}_2 &\approx \frac{r(2) - r^2(1)}{1 - r^2(1)}. \end{aligned}$$

c) Para un modelo AR(1) sabemos que $\rho(2) = \alpha_1^2 \approx r^2(1)$. Compruebe que los valores aproximados para $\widehat{\mu}$, $\widehat{\alpha}_1$ y $\widehat{\alpha}_2$ del inciso anterior se reducen a los del modelo AR(1) cuando $r(2) \approx r^2(1)$, es decir,

$$\begin{aligned} \widehat{\mu} &\approx \bar{x}, \\ \widehat{\alpha}_1 &\approx r(1), \\ \widehat{\alpha}_2 &\approx 0. \end{aligned}$$

Capítulo 4

Construcción de modelos

Construir un modelo matemático es útil para intentar entender un fenómeno o sistema y poder explicar su funcionamiento. En el caso de las series de tiempo, uno de los objetivos principales es generar pronósticos de un fenómeno aleatorio que evoluciona en el tiempo. Por lo anterior, es de gran importancia tener una metodología que sirva como guía para la correcta construcción de un modelo y hacer pronósticos posteriores a partir de él.

La modelación estadística que se emplea en la investigación se puede realizar mediante diversas metodologías. La selección de modelos estadísticamente adecuados ha tenido un gran desarrollo en los últimos años. Sin embargo, debido a la obra de Box y Jenkins (1970) [4] se estableció lo que actualmente se conoce como la metodología de Box-Jenkins (1994) [5] para construir modelos ARIMA, los cuales han sido aplicados ampliamente en los últimos años. En este capítulo se expone una breve introducción a este método para poder modelar series de tiempo estacionarias, o bien, que se puedan volver estacionarias después de alguna transformación. La metodología de Box-Jenkins se resume en cuatro fases: (1) identificación del modelo, (2) estimación de los parámetros, (3) verificación y diagnóstico y (4) pronóstico. En la fase (3) se evalúa que los residuos sigan un proceso de ruido blanco, y en la última fase se realizan las predicciones a partir del modelo creado. En este capítulo se comenta sobre las tres primeras fases, y en el siguiente se estudia la realización de pronósticos.

4.1. Identificación

Los procesos ARIMA de orden (p, d, q) que se presentaron en el capítulo anterior representan la posibilidad de construir una infinidad de modelos ya que es posible elegir combinaciones distintas de valores para p, d y q. Sin embargo, cada uno de estos parámetros puede tener valores candidatos de acuerdo con ciertos criterios y distintas propiedades de la serie de tiempo observada que se quiera modelar. Por ejemplo, es importante recordar que el parámetro d indica el número de veces que se aplica el operador diferencia a una serie no estacionaria hasta convertirla en serie estacionaria. Por otro lado, analizando las gráficas de la función de autocorrelación (ACF) y de la función de autocorrelación parcial (PACF), se proponen valores para p y q y, por tanto, se tendría un punto de partida para seleccionar un conjunto de modelos candidatos del modelo estadísticamente más adecuado para los datos de una serie de tiempo. Cada modelo ARIMA candidato puede ser *calificado* respecto de algunos criterios y compararlo con otros modelos candidatos para finalmente determinar el más adecuado. Uno de los criterios de selección más conocidos, y que se expone en la sección 4.3, es el criterio de Akaike (AIC).

Comencemos por recordar que el proceso ARIMA(p, d, q), donde los valores de p, d y q están en el conjunto $\{0, 1, 2, \ldots\}$, se puede escribir de acuerdo con la siguiente ecuación simplificada

$$\phi(B)(1-B)^d X_t = \theta(B) Z_t, \quad t \in \mathbb{Z}, \tag{4.1}$$

en donde se usa la notación introducida en (3.32), es decir,

$$\begin{aligned} \phi(B) &= 1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_p B^p, \\ \theta(B) &= 1 + \beta_1 B + \beta_2 B^2 + \cdots + \beta_q B^q. \end{aligned}$$

Ejemplo 4.1 *Para un modelo* ARIMA$(2, 0, 0)$*, el lado izquierdo de* (4.1) *se escribe como*

$$\begin{aligned} \phi(B)(1-B)^0 X_t &= \left(1 - \alpha_1 B - \alpha_2 B^2\right) X_t \\ &= X_t - \alpha_1 B X_t - \alpha_2 B^2 X_t \\ &= X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2}, \end{aligned}$$

y el lado derecho de (4.1) *se escribe como*

$$\theta(B)Z_t = Z_t.$$

Uniendo ambos lados y despejando el valor de X_t *se tiene la expresión conocida de un proceso* AR(2),

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + Z_t, \quad t \in \mathbb{Z}.$$

■

Ahora, supongamos que se tiene una serie de tiempo que inicialmente no es estacionaria, pero que después de aplicar una vez el operador diferencia se obtiene una nueva serie que sí lo es. Supongamos que esta serie $Y_t = (1-B)X_t$ se modela con un proceso ARMA(2, 2), entonces decimos que la serie original X_t se modela con un proceso ARIMA(2, 1, 2). Sin embargo, si se desarrolla el lado izquierdo de (4.1) con los valores de $(p, d, q) = (2, 1, 2)$, se tiene lo siguiente:

$$\begin{aligned} \phi(B)(1-B)X_t &= \left(1 - \alpha_1 B - \alpha_2 B^2\right)(X_t - X_{t-1}) \\ &= (X_t - X_{t-1}) - \alpha_1(X_{t-1} - X_{t-2}) - \alpha_2(X_{t-2} - X_{t-3}) \\ &= X_t - (1+\alpha_1)X_{t-1} + (\alpha_1 - \alpha_2)X_{t-2} + \alpha_2 X_{t-3}, \end{aligned}$$

y el lado derecho de (4.1) se escribe como

$$\begin{aligned} \theta(B)Z_t &= \left(1 + \beta_1 B + \beta_2 B^2\right) Z_t \\ &= Z_t + \beta_1 Z_{t-1} + \beta_2 Z_{t-2}. \end{aligned}$$

Uniendo ambos lados y despejando el valor de X_t se llega a la expresión

$$X_t = (1+\alpha_1)X_{t-1} - (\alpha_1 - \alpha_2)X_{t-2} - \alpha_2 X_{t-3} + Z_t + \beta_1 Z_{t-1} + \beta_2 Z_{t-2}.$$

Notemos que si llamamos $\alpha_1^* = 1 + \alpha_1$ y $\alpha_2^* = -(\alpha_1 - \alpha_2)$ y $\alpha_3^* = -\alpha_2$, entonces el proceso anterior puede ser escrito como

$$X_t = \alpha_1^* X_{t-1} + \alpha_2^* X_{t-2} + \alpha_3^* X_{t-3} + Z_t + \beta_1 Z_{t-1} + \beta_2 Z_{t-2}. \tag{4.2}$$

Por lo tanto, el modelo que inicialmente era ARIMA(2, 1, 2), parecería que se puede ver como un modelo ARMA(3, 2). Sin embargo, esto no es

correcto debido a que la serie original no es estacionaria, lo cual ocasionaría errores al intentar estimar los parámetros. Por lo tanto, lo más adecuado para evitar confusiones es estudiar sólo los procesos ARMA que se obtienen después de diferenciar de manera apropiada las series originales que no son estacionarias.

Como se menciona al inicio de esta sección, la fase de identificación consiste en sugerir uno o varios modelos a partir de la observación del comportamiento de una serie de tiempo observada. En esta fase, para poder usar modelos ARMA se presupone que la serie original es estacionaria; si no lo es, entonces para usar esta clase de modelos se debe utilizar primero alguna transformación apropiada. En el caso de series que se observen con un componente estacional, será necesario primero transformar los datos para remover ese componente. Una transformación sugerida consiste en aplicar el operador rezago de acuerdo con el periodo de repetición s, es decir,

$$\nabla_s x_t = (1 - B^s)x_t = x_t - x_{t-s}.$$

En otros casos será necesario aplicar el operador diferencia de orden d,

$$\nabla^d x_t = (1 - B)^d x_t = \sum_{j=0}^{d} \binom{d}{j} (-1)^j x_{t-j},$$

en donde el valor de d puede ser $1, 2, 3, \ldots$ hasta que la serie resultante sea estacionaria. La pregunta natural es ¿cómo determinar si una serie de tiempo proviene de un proceso estacionario?

Una primera respuesta emana de las características que se observan en la gráfica de la serie de tiempo. Cuando se percibe que se presenta un componente de tendencia o ciclo, es muy posible que la serie no sea estacionaria. Otro método gráfico es el correlograma de la serie de tiempo. Como se menciona en la sección 2.5, cuando el correlograma muestra una tendencia a cero muy lenta, es probable que la serie sea no estacionaria ya que ese es el comportamiento común de las series no estacionarias.

También existen algunas pruebas de hipótesis para evaluar estadísticamente si hay evidencia de estacionariedad. La más frecuente es la llamada prueba

de Dickey-Fuller, en donde la hipótesis alternativa establece que la serie $\{X_t\}$ es estacionaria. A continuación se describen los detalles de esta prueba.

Prueba de Dickey-Fuller

La prueba de Dickey-Fuller Aumentada (ADF) se usa comúnmente para evaluar si una serie de tiempo es estacionaria.

Sea $\{X_t : t \in \mathbb{Z}\}$ un proceso autorregresivo de primer orden, es decir,

$$X_t = \mu + \rho X_{t-1} + Z_t, \quad t \in \mathbb{Z}, \tag{4.3}$$

en donde $-1 < \rho < 1$. Restando X_{t-1} a ambos lados de la ecuación (4.3) se tiene que

$$\nabla X_t = \mu + (\rho - 1)X_{t-1} + Z_t. \tag{4.4}$$

Si $\rho = 1$, entonces se dice que el proceso $\{X_t : t \in \mathbb{Z}\}$ tiene raíz unitaria y es integrado de orden 1, denotado como $I(1)$. Se dice que un proceso $\{X_t : t \in \mathbb{Z}\}$ es integrado de orden d ($I(d)$) cuando el proceso que resulta después de aplicar el operador diferencia en d ocasiones es débilmente estacionario. Nótese, entonces, que las series integradas son originalmente son un caso particular de series no estacionarias.

El modelo que la prueba de Dickey-Fuller considera para determinar si el proceso $\{X_t : t \in \mathbb{Z}\}$ tiene raíces unitarias está dado por

$$\nabla X_t = \mu + \beta X_{t-1} + \delta t + \gamma \nabla X_{t-1} + Z_t.$$

En el caso de la prueba aumentada de Dickey-Fuller, ADF(s), el modelo es aumentado con s rezagos de la variable ∇X_t,

$$\nabla X_t = \mu + \beta X_{t-1} + \delta t + \sum_{i=1}^{s} \gamma_i \nabla X_{t-i} + Z_t. \tag{4.5}$$

Observe que el parámetro β representa el término $\rho - 1$ que aparece en la ecuación (4.4). Por otro lado, los parámetros γ_i para valores de $i = 1, 2, \ldots, s$ se estiman durante la prueba y aportan información sobre la necesidad de incluir cada rezago ∇X_{t-i} en el modelo de ∇X_t. La hipótesis nula de la prueba establece que la serie tiene una raíz unitaria ($\beta = 0$) y, por lo tanto, no es estacionaria.

H_0: *La ecuación* (4.5) *posee raíz unitaria.*
vs
H_a: *Existe estacionariedad.*

El estadístico de prueba τ, el cual es difícil de calcular y que comúnmente se puede obtener en cualquier software estadístico, toma valores negativos y no sigue una distribución estándar, está dado por el siguiente cociente,

$$\tau = \frac{\widehat{\beta}^*}{\widehat{\mathrm{EE}}(\widehat{\beta}^*)},$$

en donde $\widehat{\beta}^*$ es la estimación por mínimos cuadrados de β en la ecuación (4.5) y $\widehat{\mathrm{EE}}(\widehat{\beta}^*)$ representa el error estándar estimado de $\widehat{\beta}^*$, calculado como

$$\widehat{\mathrm{EE}}(\widehat{\beta}^*) = \sqrt{\frac{S^2}{\sum_{t=2}^{n}(x_{t-1} - \overline{x})^2}},$$

en donde

$$S^2 = \sum_{t=2}^{n}(\nabla x_t - \mu^* - \beta^* x_{t-1})^2/(n-3).$$

La hipótesis nula H_0 se rechaza si el estadístico τ es menor que el valor crítico de la distribución correspondiente para un nivel de significancia dado. Por lo que, entre más grande con signo negativo sea el valor de τ, mayor será la probabilidad de rechazar H_0.

Casos distintos para la prueba ADF

- **Sin término constante**. Indica que en la especificación del modelo (4.5) el término constante es eliminado y que el proceso, bajo la hipótesis nula, es una caminata aleatoria que inicia en cero.
- **Con término constante**. Indica que el proceso (4.5), bajo la hipótesis nula, es una caminata aleatoria con un valor inicial distinto de cero.
- **Tendencia lineal**. Especifica que el modelo (4.5) incluye un término de tendencia en la regresión asociada y que el proceso bajo la hipótesis nula es una caminata aleatoria con tendencia.

En temas de economía, pocas variables muestran tendencia cuadrática, pero a menudo es recomendable incluir el término de tendencia en la ecuación (4.5) cuando se observa que la variable es estacionaria alrededor de una tendencia lineal. Igualmente, siempre se recomienda incluir un término constante, a menos que sea claro que la variable tiene una media cero.

Tipo de prueba	Cambio en (4.5)
Sin término constante	$\mu = \delta = 0$.
Con término constante	$\mu > 0$ pero $\delta = 0$.
Con tendencia lineal	$\mu \geqslant 0$ y $\delta > 0$.

Tabla 4.1: Tipos de prueba de hipótesis de Dickey-Fuller.

Análisis de las funciones ACF y PACF

Una vez que se ha mostrado evidencia estadística de que las observaciones provienen de una serie estacionaria, se utilizan las gráficas de la función de autocorrelación (ACF) y de la función de autocorrelación parcial (PACF) muestrales para identificar un proceso ARMA(p, q), que se propondrá como proceso generador de los datos observados de la serie de tiempo.

Los siguientes resultados relacionados con las funciones ACF y PACF de los procesos estacionarios AR(p), MA(q) y ARMA(p, q) sirven como criterios guía para una identificación inicial.

- Proceso estacionario autorregresivo AR(p)

  En los procesos estacionarios autorregresivos, los valores de la función ACF decaen gradualmente hacia cero, pero no en todos los casos son iguales a cero. Estos valores pueden alternar de signo y en ocasiones se observa un patrón ondulatorio. Ahora bien, la función PACF toma valores cercanos a cero después del rezago p. Por ejemplo, en series de tiempo generadas por procesos AR(1), el coeficiente

de autocorrelación parcial respecto al primer rezago, $k = 1$, es estadísticamente distinto de cero. Sin embargo, los valores que toma la autocorrelación parcial respecto a observaciones rezagadas más de un periodo, es decir $k \geqslant 1$, son estadísticamente iguales a cero.

- Proceso de media móvil MA(q)

  En los procesos estacionarios de medias móviles, la gráfica de la función ACF teórica toma el valor de cero después del rezago q. Por otro lado, la función PACF teórica decae muy lentamente hacia cero pero nunca es igual a cero.

- Proceso estacionario ARMA(p, q)

  Los procesos estacionarios ARMA pueden mostrar una mezcla de las características de los procesos AR(p) y MA(q), es decir, los valores de la función PACF son cercanos a cero después de p rezagos y los valores de la función ACF son cercanos a cero después de q rezagos. En general, ambas funciones teóricas, ACF y PACF, decaen lentamente hacia cero después de ciertos valores q y p, respectivamente, pero no toman el valor de cero.

Ejemplos generados por simulación

A continuación se presenta una variedad de ejemplos para ilustrar la fase de identificación de un modelo ARMA y/o ARIMA.

Las gráficas y los cálculos relacionados con las funciones ACF y PACF se realizaron con la ayuda del software estadístico `R`. Se supondrá que el proceso $\{Z_t : t \in \mathbb{Z}\}$ es un ruido blanco WN$(0, 1)$ (consultar el ejemplo 1.1), es decir, Z_t tiene una distribución N$(0, 1)$. El lector interesado puede consultar los códigos en el Apéndice para reproducir lo mostrado o para analizar algunos otros ejemplos.

Ejemplo 4.2 *Se simula el proceso* AR(1) *dado por*

$$X_t = 0.5X_{t-1} + Z_t, \qquad t = 1, \ldots, 1000,$$

en donde se define $X_0 = 0$. *Los resultados obtenidos se muestran en la figura 4.1. Se puede observar lo siguiente:*

- *Los valores que toma la función PACF muestral son estadísticamente iguales a cero después del rezago* $\tau = 1$.
- *Respecto del comportamiento de la función ACF muestral, se observa que conforme va aumentando el tamaño del rezago con respecto al cual se calcula, esta función toma valores absolutos cada vez más pequeños. Los valores de las funciones ACF muestral y teórica se mantienen cercanos.*

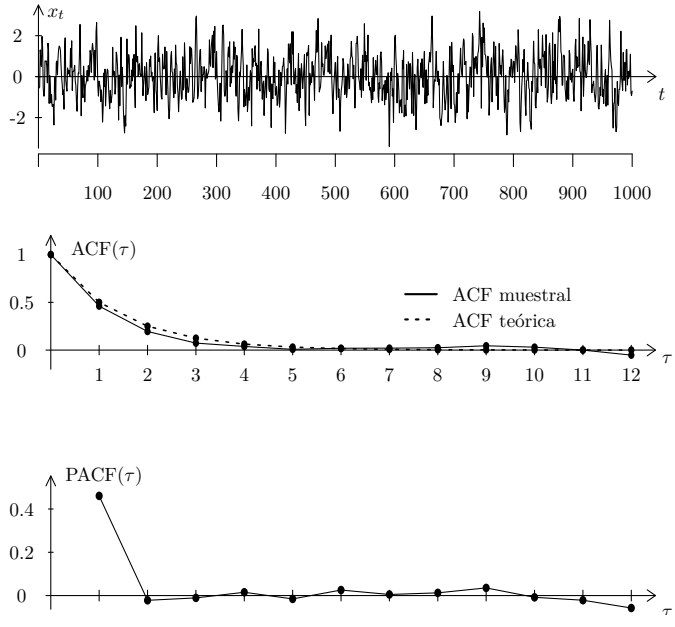

Figura 4.1: Proceso AR(1) dado por $X_t = 0.5X_{t-1} + Z_t$.

■

Ejemplo 4.3 *Se simula el proceso* AR(2) *dado por*

$$X_t = 0.6X_{t-1} - 0.3X_{t-2} + Z_t, \qquad t = 2, \ldots, 1000,$$

en donde se define $X_0 = 0$ *y* $X_1 = Z_1$. *Los resultados de la simulación se muestran en la figura 4.2. Se puede observar lo siguiente:*

- *Los valores que toma la función PACF muestral son estadísticamente iguales a cero después del rezago* $\tau = 2$.
- *Nuevamente se observa que el gráfico de la función ACF muestral toma valores absolutos cada vez más pequeños, conforme aumenta el tamaño del rezago respecto del cual se calcula. También fueron muy parecidos los valores de las funciones ACF muestral y teórica.*

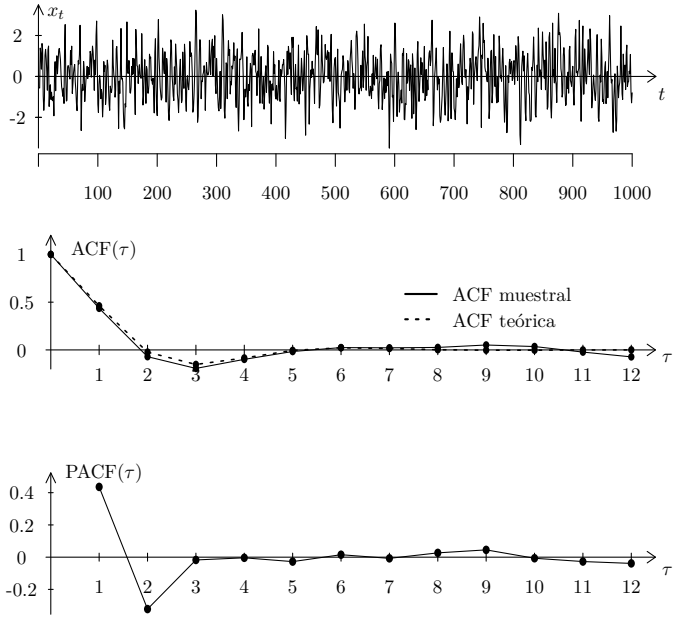

Figura 4.2: Proceso AR(2) dado por $X_t = 0.6X_{t-1} - 0.3X_{t-2} + Z_t$.

■

Ejemplo 4.4 *Se simula el proceso* MA(1) *dado por*

$$X_t = Z_t + 0.7Z_{t-1}, \qquad t = 1, \ldots, 1000,$$

en donde se define $X_0 = Z_0 := 0$. *Los resultados obtenidos de la simulación se muestran en la figura 4.3. Se observa lo siguiente:*

- *Los valores que toma la función ACF muestral prácticamente coinciden con los valores que toma la función ACF teórica y son estadísticamente iguales a cero después del rezago* $\tau = 1$.
- *En la función PACF muestral, los valores absolutos cada vez son más pequeños conforme aumentan los rezagos.*

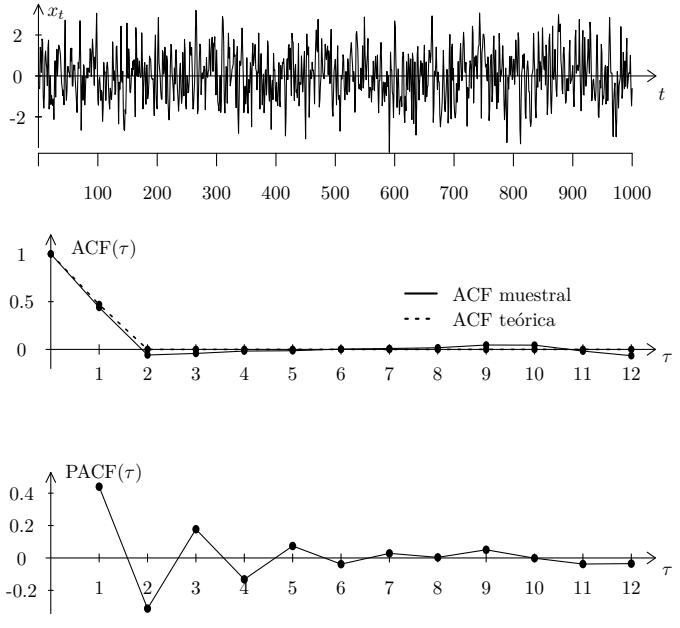

Figura 4.3: Proceso MA(1) dado por $X_t = Z_t + 0.7Z_{t-1}$.

■

Ejemplo 4.5 *Se simula el proceso* MA(2) *dado por*

$$X_t = Z_t - 0.8Z_{t-1} + 0.6Z_{t-2}, \qquad t = 2, \ldots, 1000,$$

en donde se define $X_0 = Z_0 := 0$ *y* $X_1 = Z_1$. *Los resultados obtenidos se muestran en la figura 4.4. Se observa lo siguiente:*

- *Los valores que toma la función ACF muestral son estadísticamente iguales a cero después del segundo rezago. Nuevamente se observa una gran coincidencia de la función ACF muestral con la función ACF teórica.*
- *En la función PACF muestral, de nuevo los valores absolutos cada vez son más pequeños conforme crece el rezago.*

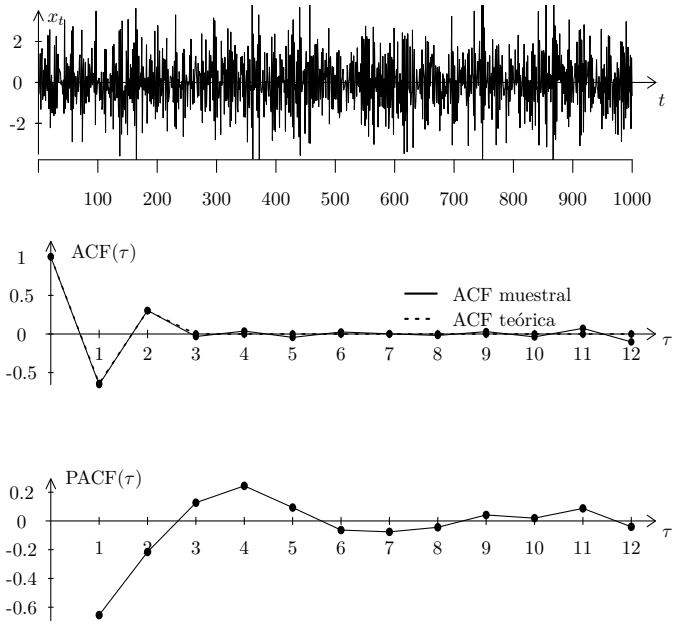

Figura 4.4: Proceso MA(2) dado por $X_t = Z_t + 0.8Z_{t-1} - 0.6Z_{t-2}$.

■

Ejemplo 4.6 *Se simula el proceso* ARMA(1, 1) *dado por*

$$X_t = 0.3X_{t-1} + Z_t + 0.5Z_{t-1}, \qquad t = 1, \ldots, 1000,$$

en donde se define $X_0 = Z_0 := 0$. *Los resultados de la simulación se muestran en la figura 4.5. Se puede observar que, tanto los valores que toma la función ACF muestral como los valores de la función PACF muestral, tienden rápidamente hacia cero después del primer rezago. Los gráficos de las funciones ACF muestral y teórica se mantienen cercanos en todos los rezagos.*

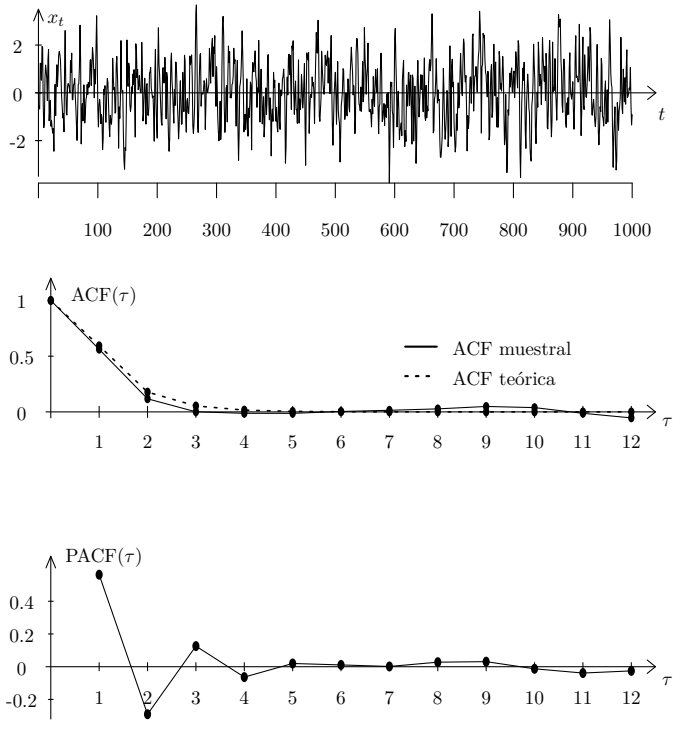

Figura 4.5: Proceso ARMA(1, 1) dado por $X_t = 0.3X_{t-1} + 0.5Z_{t-1} + Z_t$.

■

Ejemplo 4.7 *Se simula el proceso* ARMA(2, 2) *dado por*

$$X_t = -0.3X_{t-1} + 0.2X_{t-2} + Z_t + 0.1Z_{t-1} + 0.4Z_{t-2}, \qquad t = 2, \ldots, 1000,$$

con valores iniciales $X_0 = Z_0 := 0$ *y* $X_1 = Z_1$. *Los resultados obtenidos se muestran en la figura 4.6. Los valores de las funciones ACF muestral y teórica tienen un gran parecido. Se puede observar que tanto los valores que toma la función ACF muestral como los valores de la función PACF muestral, tienden rápidamente hacia cero después del segundo rezago.*

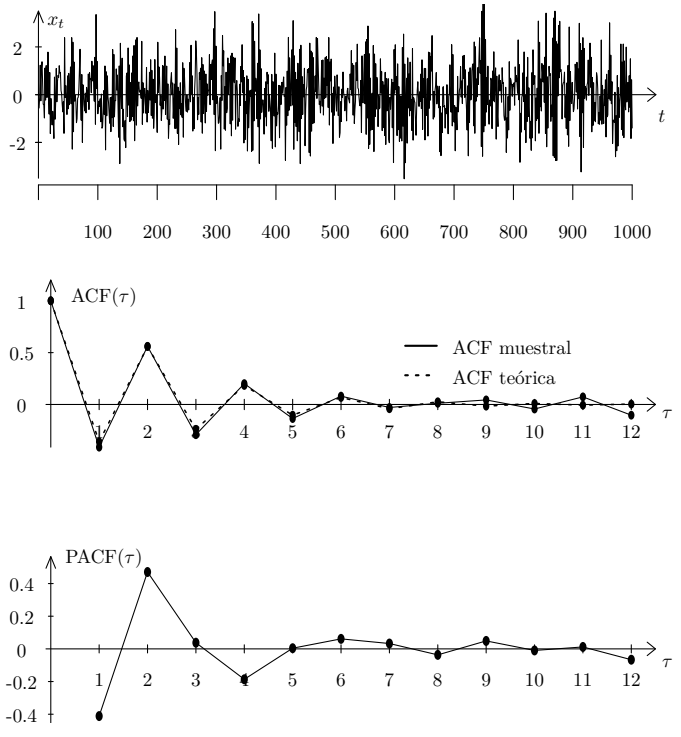

Figura 4.6: Proceso ARMA(2, 2) dado por
$X_t = -0.3X_{t-1} + 0.2X_{t-2} + Z_t + 0.1Z_{t-1} + 0.4Z_{t-2}$.

■

Ejemplo 4.8 (Proceso no estacionario con tendencia lineal)
Se simula el proceso

$$X_t = 0.5X_{t-1} + 0.1t + Z_t, \qquad t = 1, \ldots, 100,$$

con valor inicial $X_0 = 0$. *Este es un ejemplo de una serie de tiempo generada por un proceso* AR(1) *con tendencia lineal. Los resultados de la simulación se muestran en la figura 4.7 . Se puede observar que la serie de tiempo no tiene las características de un proceso estacionario alrededor de una tendencia nula o constante. La gráfica de la función de autocorrelación muestral debería converger a cero geométricamente si la serie fuera estacionaria. Sin embargo, en este caso se observa que la función ACF muestral decae muy lentamente. En la gráfica de la función PACF muestral se puede identificar la parte autoregresiva de primer orden de la serie.*

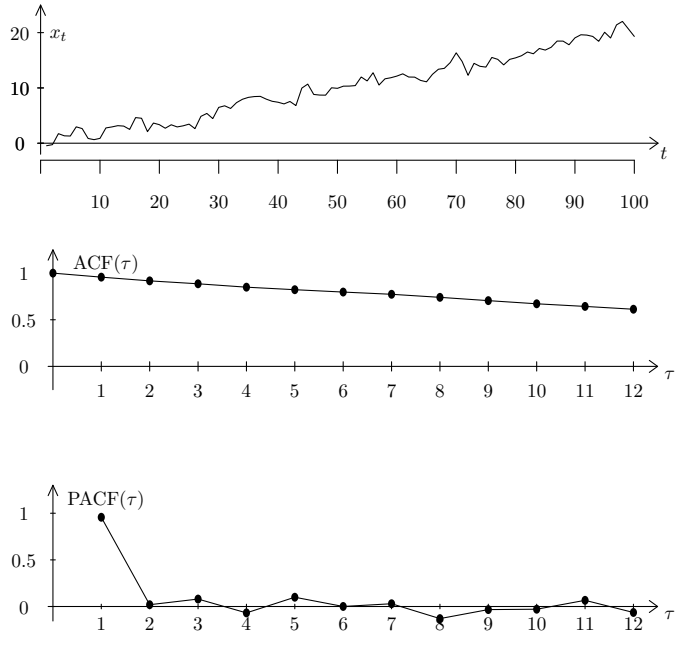

Figura 4.7: Proceso no estacionario AR(1) con tendencia lineal.

■

Ejemplo 4.9 (Proceso ARIMA) *En el ejemplo anterior se analizó la serie no estacionaria* AR(1) *con tendencia lineal*

$$X_t = 0.5X_{t-1} + 0.1t + Z_t,$$

para valores de $t = 1, \ldots, 100$ *y* $X_0 = 0$. *Ahora se aplica la transformación*

$$Y_t = \nabla X_t = X_t - X_{t-1},$$

con lo cual se obtiene un componente de tendencia constante que se resta para tener el proceso centralizado

$$X_t^* = Y_t - 0.1 = 0.5(X_{t-1} - X_{t-2}) + Z_t^*,$$

en donde $Z_t^* \sim N(0, 2)$. *En la figura 4.8 se puede observar que la serie* X_t^* *ya parece comportarse como un proceso estacionario alrededor de cero (tendencia constante), y la gráfica de la función de autocorrelación converge rápidamente a cero. Se observa también que la función PACF muestral decae a cero rápidamente. En este caso es posible ajustar un modelo* ARMA *a la nueva serie.*

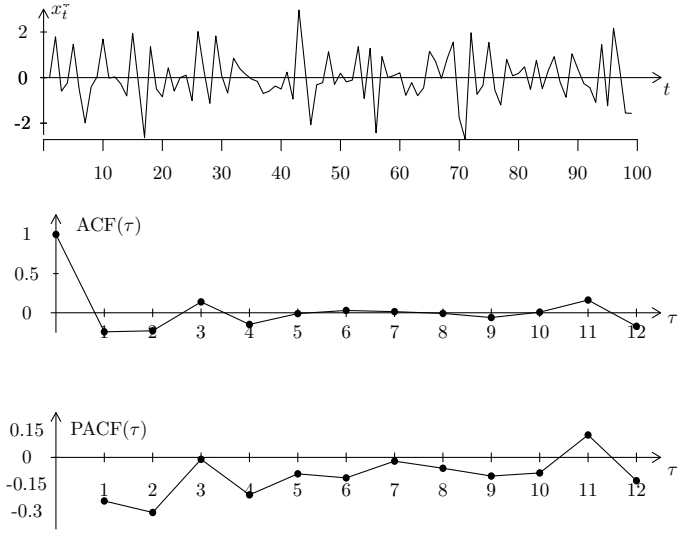

Figura 4.8: Proceso diferenciado.

■

4.2. Estimación

Como se ha mencionado, los modelos AR(p), MA(q) y ARMA(p, q) son casos especiales del modelo ARIMA(p, d, q). En su caso, en esta sección supondremos que el proceso de diferenciación ya se ha dado y se tiene una serie estacionaria. El siguiente paso natural en la construcción de modelos ARMA(p, q), después de haber elegido valores candidatos de p y de q, es hacer la estimación de los parámetros $\alpha_1, \ldots, \alpha_p$, $\beta_1, \ldots, \beta_q$ y σ^2. Se tienen entonces $p + q + 1$ parámetros desconocidos para ajustar el modelo ARMA con media cero,

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q}, \quad t \in \mathbb{Z}, \quad (4.6)$$

en donde $\{Z_t : t \in \mathbb{Z}\}$ es un ruido blanco con media cero y varianza desconocida σ^2. Existe más de una forma de realizar esta estimación. De hecho, en la sección 3.8 se describe una posible estimación por mínimos cuadrados y en esta sección se mencionan algunos métodos alternativos.

Ecuaciones de Yule-Walker

Es posible volver a utilizar las ecuaciones de Yule-Walker del capítulo 3 para estimar los parámetros de modelos autorregresivos. Recordemos que estas ecuaciones recursivas para calcular las autocorrelaciones de un modelo AR(p) (proposición 3.7) están dadas por

$$\rho(\tau) = \begin{cases} 1 & \text{si } \tau = 0, \\ \displaystyle\sum_{k=1}^{p} \alpha_k \, \rho(\tau - k) & \text{si } \tau \geqslant 1. \end{cases}$$

Estas ecuaciones son útiles para establecer igualdades que sirvan como estimadores de momentos. La idea es usar las ecuaciones recursivas pero sustituyendo la autocorrelación (poblacional) del modelo por la autocorrelación muestral que se presenta en el primer capítulo, y que se calcula como

$$r(\tau) = \frac{\displaystyle\sum_{t=1}^{n-\tau} (x_t - \bar{x})(x_{t+\tau} - \bar{x})}{\displaystyle\sum_{t=1}^{n} (x_t - \bar{x})^2}, \quad (4.7)$$

para $\tau = 0, 1, \ldots, n-1$. El siguiente es un ejemplo simple para ilustrar esta idea.

Ejemplo 4.10 *Supongamos que se tiene una serie $x_1, \ldots, x_n$ para la cual se desea ajustar el modelo* AR(1) *dado por*

$$X_t = \alpha_1 X_{t-1} + Z_t, \quad t \in \mathbb{Z}. \tag{4.8}$$

A partir de las ecuaciones recursivas se tiene que $\rho(1) = \alpha_1 \rho(0) = \alpha_1$. El parámetro α_1 se puede estimar a partir de igualar la autocorrelación muestral de la serie $x_1, \ldots, x_n$ con la autocorrelación del modelo (4.8)*, es decir,*

$$r(1) = \rho(1).$$

Por lo tanto,

$$\widehat{\alpha}_1 = r(1) = \frac{\sum_{t=1}^{n-1} (x_t - \overline{x})(x_{t+1} - \overline{x})}{\sum_{t=1}^{n} (x_t - \overline{x})^2}.$$

Por otro lado, para estimar la varianza de Z_t, primero se calcula la varianza muestral de la serie $x_1, \ldots, x_n$ como se define en (1.3)*, es decir,*

$$\operatorname{var}(x) = \frac{1}{n} \sum_{t=1}^{n} (x_t - \bar{x})^2.$$

De la proposición 3.4 se tiene que

$$\operatorname{Var}(X_t) = \frac{\sigma^2}{1 - \alpha_1^2}.$$

Substituyendo $\operatorname{Var}(X_t)$ *por la varianza muestral y utilizando la estimación previa de α_1, se propone la estimación*

$$\widehat{\sigma}^2 = (1 - \widehat{\alpha}_1^2) \operatorname{var}(x).$$

■

Con la misma idea, en el siguiente ejemplo se estima el valor de los parámetros α_1, α_2 y σ^2 en el modelo AR(2).

Ejemplo 4.11 *Supongamos que se tiene una serie $x_1, \ldots, x_n$ para la cual se desea ajustar el modelo* AR(2) *dado por*

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + Z_t, \quad t \in \mathbb{Z}.$$

Por las ecuaciones de Yule-Walker, se tiene el siguiente sistema de ecuaciones:

$$\begin{aligned} \rho(1) &= \alpha_1 \rho(0) + \alpha_2 \rho(1), \\ \rho(2) &= \alpha_1 \rho(1) + \alpha_2 \rho(0), \end{aligned}$$

en donde $\rho(0) = 1$. Si se toman las correlaciones muestrales como estimaciones de las correlaciones teóricas, entonces las incógnitas α_1 y α_2 deben satisfacer de manera aproximada el mismo sistema de ecuaciones:

$$\begin{aligned} r(1) &= \alpha_1 + \alpha_2 r(1), \\ r(2) &= \alpha_1 r(1) + \alpha_2. \end{aligned}$$

Resolviendo este sistema se obtiene

$$\begin{aligned} \widehat{\alpha}_1 &= \frac{r(1)(1 - r(2))}{1 - r(1)^2}, \\ \widehat{\alpha}_2 &= \frac{r(2) - r(1)^2}{1 - r(1)^2}. \end{aligned}$$

Recordemos que los valores de $r(1)$ y $r(2)$ se calculan según la fórmula (4.7). *Para estimar el valor de σ^2 se procede de la siguiente forma. A partir de la proposición 3.8 y suponiendo que el proceso es causal, se tiene que*

$$\gamma(0) = \frac{\sigma^2}{1 - \alpha_1 \rho(1) - \alpha_2 \rho(2)},$$

en donde $\gamma(0) = \mathrm{Var}(X_t)$. Despejando σ^2 se tiene que

$$\sigma^2 = (1 - \alpha_1 \rho(1) - \alpha_2 \rho(2)) \, \mathrm{Var}(X_t).$$

Se propone entonces la siguiente estimación para σ^2,

$$\widehat{\sigma}^2 = (1 - \widehat{\alpha}_1 r(1) - \widehat{\alpha}_2 r(2)) \, \mathrm{var}(x),$$

en donde, como antes, $\mathrm{var}(x)$ *es la varianza muestral.* ∎

En general se puede usar el mismo esquema para estimar los parámetros de un modelo AR(p). Para escribir formalmente la fórmula para estos estimadores para la matriz de autocovarianzas muestrales es conveniente usar la siguiente notación:

$$\widehat{\Gamma} = \begin{pmatrix} \widehat{\gamma}(0) & \widehat{\gamma}(1) & \cdots & \widehat{\gamma}(p-1) \\ \widehat{\gamma}(1) & \widehat{\gamma}(0) & \cdots & \widehat{\gamma}(p-2) \\ \vdots & \vdots & \cdots & \vdots \\ \widehat{\gamma}(p-1) & \widehat{\gamma}(p-2) & \cdots & \widehat{\gamma}(0) \end{pmatrix}.$$

La matriz de autocorrelaciones muestrales será denotada por $\widehat{\mathbf{R}}$ y, en consecuencia,

$$\widehat{\mathbf{R}} = \frac{1}{\widehat{\gamma}(0)}\,\widehat{\Gamma}. \tag{4.9}$$

Se deja como ejercicio para el lector demostrar que esta matriz es invertible.

Definición 4.1 *Sea $x_1, \ldots, x_n$ una serie de tiempo a la que se desea ajustar un modelo* AR(p). *Supongamos que $\widehat{\rho} = (r(1), \ldots, r(p))'$ es el vector transpuesto de las autocorrelaciones muestrales. Entonces los estimadores de Yule-Walker de $\alpha = (\alpha_1, \ldots, \alpha_p)$ y σ^2 están dados por*

$$\begin{aligned} \widehat{\alpha} &= \widehat{\mathbf{R}}^{-1}\widehat{\rho}, \\ \widehat{\sigma}^2 &= \widehat{\gamma}(0)\left(1 - \widehat{\rho}'\widehat{\mathbf{R}}^{-1}\widehat{\rho}\right). \end{aligned}$$

S puede demostrar que los estimadores de α en un proceso AR(p) tienden a tener una distribución normal cuando el tamaño de la muestra n es grande; en específico,

$$\widehat{\alpha} \approx N\left(\alpha, n^{-1}\sigma^2\Gamma^{-1}\right).$$

En el caso de los modelos MA(q) se puede utilizar la misma idea de igualar los valores de las autocorrelaciones teóricas y muestrales. Sin embargo, el sistema de ecuaciones resultante es no lineal y podría tener soluciones restringidas sólo para ciertos valores de las autocorrelaciones muestrales como en el siguiente caso.

Ejemplo 4.12 *Consideremos el modelo* MA(1) *dado por*

$$X_t = Z_t + \beta_1 Z_t, \quad t \in \mathbb{Z}.$$

Por la proposición 3.2,

$$\rho(1) = \frac{\beta_1}{1+\beta_1^2}.$$

Sustituyendo $\rho(1)$ *por* $r(1)$ *en la ecuación anterior, se llega a la ecuación cuadrática*

$$r(1)\beta_1^2 - \beta_1 + r(1) = 0.$$

Entonces se tienen dos posibles soluciones para β_1*,*

$$\beta_1 = \frac{1 \pm \sqrt{1-4(r(1))^2}}{2\,r(1)}.$$

Si $|r(1)| = 1/2$*, entonces* $|\widehat{\beta}_1| = 1$ *y no se satisface la condición de invertibilidad* $|\beta_1| < 1$*. Si* $|r(1)| > 1/2$*, entonces no existe solución real. Sólo cuando* $|r(1)| < 1/2$*,* $\widehat{\beta}_1$ *será la solución que cumpla* $|\beta_1| < 1$*.* ∎

Cuando se tiene un orden $q \geqslant 2$ el sistema no lineal podría requerir una solución numérica, y además de que seguirá el problema de si hay una solucón única, múltiple o si no hay solución.

Para estimar la varianza, como se establece en la proposición 3.1,

$$\mathrm{Var}(X_t) = \sigma^2 \sum_{k=0}^{q} \beta_k^2,$$

lo que implica

$$\sigma^2 = \frac{\mathrm{Var}(X_t)}{\sum_{k=0}^{q} \beta_k^2}.$$

Por lo tanto, para el modelo MA(1) mostrado en el ejemplo, el estimador de la varianza de Z_t será

$$\widehat{\sigma}^2 = \frac{\widehat{\gamma}(0)}{1+\widehat{\beta}_1^{\,2}}.$$

En el caso de los modelos ARMA(p,q), nuevamente este método por momentos puede resultar en un sistema de ecuaciones complicado. En las siguientes

subsecciones se muestran algunos métodos alternativos. Sin embargo, en el siguiente ejemplo se presentan los cálculos para estimar por momentos los parámetros de un modelo ARMA$(1,1)$.

Ejemplo 4.13 *Consideremos el modelo* ARMA$(1,1)$ *dado por*

$$X_t = \alpha_1 X_{t-1} + Z_t + \beta_1 Z_{t-1}, \quad t \in \mathbb{Z}.$$

Por la proposición 3.11 se tiene que

$$\rho(1) = \frac{(1+\alpha_1\beta_1)(\alpha_1+\beta_1)}{1+2\alpha_1\beta_1+\beta_1^2}, \tag{4.10}$$

y

$$\rho(2) = \alpha_1 \frac{(1+\alpha_1\beta_1)(\alpha_1+\beta_1)}{1+2\alpha_1\beta_1+\beta_1^2}.$$

Por lo tanto, $\rho(2)/\rho(1) = \alpha_1$ *y*

$$\widehat{\alpha}_1 = \frac{r(2)}{r(1)}.$$

Para estimar el valor de β_1 *se puede usar la ecuación* (4.10) *con los estimadores de* $\rho(1)$ *y* α_1*,*

$$r(1) = \frac{(1+\widehat{\alpha}_1\beta_1)(\widehat{\alpha}_1+\beta_1)}{1+2\widehat{\alpha}_1\beta_1+\beta_1^2}.$$

De esta última igualdad se puede resolver la ecuación cuadrática para encontrar dos posibles valores de β_1 *y analizar si alguno cumple la condición* $|\beta_1| < 1$*. Por último, para estimar el valor de la varianza de* Z_t *se tiene que*

$$\mathrm{Var}(X_t) = \left[1 + \frac{(\alpha_1+\beta_1)^2}{1-\alpha_1^2}\right]\sigma^2.$$

Por lo tanto,

$$\widehat{\sigma}^2 = \widehat{\gamma}(0) \Big/ \left[1 + \frac{(\widehat{\alpha}_1+\widehat{\beta}_1)^2}{1-\widehat{\alpha}_1^2}\right].$$

■

Algoritmo de innovaciones

Otro de los métodos para estimar los parámetros de un modelo ARMA(p, q), en donde $q \geqslant 1$, es el llamado algoritmo de innovaciones. Este método puede ser más útil que las ecuaciones de Yule-Walker para las estimaciones del modelo MA. El método consiste en calcular de manera recursiva una sucesión de estimaciones convergentes a la estimación que buscamos.

Comenzaremos exponiendo el algoritmo para el modelo MA(q) dado por

$$X_t = Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q}, \quad t \in \mathbb{Z},$$

cuando

1. $E(X_t^4) < \infty$.
2. El modelo es invertible, es decir, las raíces $z \in \mathbb{C}$ del polinomio que aparece abajo son tales que $|z| > 1$ (ver el teorema 3.1),

$$1 + \beta_1 z + \cdots + \beta_q z^q = 0.$$

Estamos suponiendo que el proceso es estacionario y que $E(X_t) = 0$, así que para $0 < i < j$ se tiene que

$$E(X_{t+\tau} X_t) = \mathrm{Cov}(X_{t+\tau}, X_t) = \gamma(\tau).$$

Si $\tau > q$, entonces $\gamma(\tau) = 0$. Además,

$$\begin{aligned} \gamma(0) &= \mathrm{Var}(X_t) = \sigma^2 \sum_{k=0}^{q} \beta_k^2, \qquad (\beta_0 = 1) \\ \gamma(\tau) &= \sigma^2 \sum_{k=0}^{q-\tau} \beta_k \, \beta_{k+\tau} \quad \text{si } \tau = 1, \ldots, q. \end{aligned}$$

La autocovarianza muestral, que será denotada como $\widehat{\gamma}(\tau)$, se definió como

$$\widehat{\gamma}(\tau) = \frac{1}{n-\tau} \sum_{t=1}^{n-\tau} (x_t - \bar{x})(x_{t+\tau} - \bar{x}).$$

En el siguiente algoritmo se calculan de forma recursiva $\beta_{1,1}$, v_1; $\beta_{2,2}$, $\beta_{2,1}$, v_2; $\beta_{3,3}$, $\beta_{3,2}$, $\beta_{3,1}$, v_3; .... Al final, los valores de interés serán

$$\beta_{m,1}, \beta_{m,2}, \ldots, \beta_{m,q},$$

para un valor de m donde las fluctuaciones de los parámetros se estabilicen, sabiendo que convergen a los verdaderos valores cuando $m \to \infty$.

Algoritmo de innovaciones

1. Establecer como valor inicial a

$$v_0 = \widehat{\gamma}(0) = \frac{1}{n}\sum_{t=1}^{n}(x_t - \overline{x})^2.$$

2. Calcular

$$\beta_{1,1} = \widehat{\gamma}(1)/v_0; \quad v_1 = v_0(1 - \beta_{1,1}^2).$$

3. Escoger un valor inicial de m que será el número de iteraciones a realizar. Para $i = 2, \ldots, m$ y $j = 0, 1, \ldots, i-1$,

$$\begin{aligned} \beta_{i,i-j} &= v_j^{-1}\left(\widehat{\gamma}(i-j) - (\beta_{j,j}\,\beta_{i,i}\,v_0 + \cdots + \beta_{j,1}\beta_{i,i-j+1}v_{j-1})\right) \\ &= v_j^{-1}\left(\widehat{\gamma}(i-j) - \sum_{k=0}^{j-1}\beta_{j,j-k}\beta_{i,i-k}v_k\right), \\ v_i &= v_0 - \sum_{k=0}^{i-1}\beta_{i,i-k}^2 v_k. \end{aligned}$$

4. Si los valores de $\beta_{m,j}$ obtenidos tuvieron valores muy parecidos (se estabilizaron) a los valores de $\beta_{i,j}$ para $i = m-1, m-2, m-3, m-4$ y $j = 1, \ldots, q$, entonces los estimadores serán

$$\widehat{\beta}_1 = \beta_{m,1},\ \widehat{\beta}_2 = \beta_{m,2},\ \ldots,\ \widehat{\beta}_q = \beta_{m,q}, \quad \text{y} \quad \widehat{\sigma}^2 = v_m. \tag{4.11}$$

   En caso contrario, continuar el proceso del paso 3 hasta que los valores de (4.11) se estabilicen.

Consideremos ahora el modelo ARMA dado por

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + Z_t + \beta_1 Z_{t-1} + \cdots \beta_q Z_{t-q}, \quad t \in \mathbb{Z}.$$

Se tiene la siguiente forma de expresar al proceso X_t debido a (3.34) cuando se tiene causalidad,

$$X_t = \sum_{k=0}^{\infty} \pi_k Z_{t-k}, \quad \in \mathbb{Z},$$

en donde $\sum_{k=0}^{\infty} |\pi_k| < \infty$.

Los coeficientes π_k van a satisfacer la relación

$$\pi_k = \beta_k + \sum_{i=1}^{k \wedge p} \alpha_i \pi_{k-i}, \tag{4.12}$$

para $k = 0, 1, \ldots$, en donde se define $\beta_0 := 1$ y $\beta_k := 0$ para $k > q$. Para estimar los valores de $\pi_1, \ldots, \pi_{p+q}$ se hace uso de los valores

$$\widehat{\beta}_{m,1}, \ldots, \widehat{\beta}_{m,p+q},$$

estimados con el algoritmo de innovaciones dados en (4.11), teniendo en cuenta un modelo MA$(p+q)$.

Se reemplazan los valores de π_k por los valores de $\widehat{\beta}_{m,k}$ en la relación (4.12) para establecer el siguiente sistema de ecuaciones en el que las incógnitas son los parámetros α y β,

$$\widehat{\beta}_{m,k} = \beta_k + \sum_{i=1}^{k \wedge p} \alpha_i \widehat{\beta}_{m,k-i}, \quad k = 1, \ldots, p+q. \tag{4.13}$$

Primero se obtienen las estimaciones de $\alpha_1, \ldots, \alpha_p$ de las últimas p ecuaciones de (4.13) debido a que en esas ecuaciones $\beta_k = 0$ y entonces se tienen

las ecuaciones:

$$\begin{aligned}
\widehat{\beta}_{m,q+1} &= \sum_{i=1}^{(q+1)\wedge p} \alpha_i \widehat{\beta}_{m,q+1-i}, \\
\widehat{\beta}_{m,q+2} &= \sum_{i=1}^{(q+2)\wedge p} \alpha_i \widehat{\beta}_{m,q+2-i}, \\
&\vdots \\
\widehat{\beta}_{m,q+p} &= \sum_{i=1}^{p} \alpha_i \widehat{\beta}_{m,q+p-i}.
\end{aligned}$$

Después de encontrar los valores de $\alpha_1, \ldots, \alpha_p$, sólo resta resolver el sistema siguiente que ahora es de q incógnitas y q ecuaciones:

$$\begin{aligned}
\beta_1 &= \widehat{\beta}_{m,1}, \\
\beta_2 &= \widehat{\beta}_{m,2} - \sum_{i=1}^{2\wedge p} \alpha_i \widehat{\beta}_{m,2-i}, \\
&\vdots \\
\beta_q &= \widehat{\beta}_{m,q} - \sum_{i=1}^{q\wedge p} \alpha_i \widehat{\beta}_{m,q-i}.
\end{aligned}$$

Estimación por máxima verosimilitud

La estimación por máxima verosimilitud (EMV) del vector de parámetros

$$(\alpha, \beta, \sigma^2) = (\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q, \sigma^2),$$

consiste en calcular los valores $\widehat{\alpha}$, $\widehat{\beta}$ y $\widehat{\sigma^2}$ que maximicen la función de verosimilitud. Recordemos que esta función es igual a la función de densidad conjunta de las variables aleatorias $X_1, \ldots, X_n$

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n; \alpha, \beta, \sigma^2),$$

suponiendo a los parámetros α, β y σ^2 como variables y teniendo en cuenta los valores numéricos $x_1, \ldots, x_n$. Cuando las variables aleatorias $X_1, \ldots, X_n$

son independientes, el cálculo de la función de verosimilitud resulta sencillo, porque la función de densidad conjunta es simplemente el producto de las funciones de densidad marginales. Sin embargo, en los modelos ARMA existe dependencia y habrá mayor dificultad en el cálculo de esta función. Resolver de manera analítica el problema de optimización puede ser muy complicado. Es necesario aplicar algún método numérico para maximizar la función de verosimilitud con respecto a los parámetros, y afortunadamente, este procedimiento está implementado en varios programas de cómputo estadístico.

A continuación se describen dos enfoques para la obtención de la función de verosimilitud. Consideremos el siguiente modelo ARMA,

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q}. \qquad (4.14)$$

Como sabemos, t puede ser cualquier valor entero, pero supondremos que es mayor a cero para considerar valores numéricos de la serie. Es necesario especificar una distribución para el proceso aleatorio $\{Z_t\}$. De manera típica y conveniente, supondremos que

$$Z_t \sim N(0, \sigma^2). \qquad (4.15)$$

Por conveniencia, usaremos la siguiente notación:

$$\underline{X}_n = (X_1, \ldots, X_n) \quad \text{y} \quad \underline{x}_n = (x_1, \ldots, x_n).$$

Además, usaremos la siguiente notación para referirnos a la función de densidad conjunta:

$$f(x_1, \ldots, x_n) = f_{X_1,\ldots,X_n}(x_1, \ldots, x_n; \alpha, \beta, \sigma^2),$$

y cuando la tratemos como función de verosimilitud, se escribirá

$$L(\alpha, \beta, \sigma^2) = f(x_1, \ldots, x_n),$$

en donde supondremos que se conocen n observaciones de la serie de tiempo estudiada.

Enfoque 1. Condicionando

Supongamos que en el modelo ARMA(p,q), $r = \max\{p,q\}$. Debido a que la variable aleatoria X_n depende de r variables anteriores, supongamos un valor de $n > r$. Condicionando en forma repetida, la función de densidad conjunta se puede escribir como

$$\begin{aligned} L(\alpha,\beta,\sigma^2) &= f(x_1,\ldots,x_n) \\ &= f(x_n \mid \underline{x}_{n-1}) f(\underline{x}_{n-1}) \\ &= f(x_n \mid \underline{x}_{n-1}) f(x_{n-1} \mid \underline{x}_{n-2}) f(\underline{x}_{n-2}) \\ &\vdots \\ &= f(x_n \mid \underline{x}_{n-1}) f(x_{n-1} \mid \underline{x}_{n-2}) \cdots f(x_2 \mid x_1) f(x_1) \\ &= f(x_1) \prod_{t=2}^{n} f(x_t \mid \underline{x}_{t-1}). \end{aligned} \tag{4.16}$$

Observemos que la variable aleatoria X_t condicionada a los valores anteriores es igual a una suma de variables aleatorias normales, por lo que tiene distribución normal. Calculamos esperanza y varianza de la expresión (4.14) para conocer los parámetros de la distribución normal. Primero, consideremos los casos en los que $t > r$:

$$E(X_t \mid \underline{X}_{t-1} = \underline{x}_{t-1}) = \sum_{i=1}^{p} \alpha_i \, x_{t-i},$$

$$\mathrm{Var}(X_t \mid \underline{X}_{t-1}) = \sigma^2 \sum_{j=0}^{q} \beta_j^2,$$

en donde $\beta_0 := 1$. Por lo tanto,

$$X_t \mid \underline{X}_{t-1} = \underline{x}_{t-1} \sim N\left(\sum_{i=1}^{p} \alpha_i \, x_{t-i} \, , \, \sigma^2 \sum_{j=0}^{q} \beta_j^2 \right). \tag{4.17}$$

Ahora, en los casos en que $t = 2,\ldots,r$ definimos $p^* = \min\{p, t-1\}$ y $q^* = \min\{q, t-1\}$. Por ejemplo, si $p = 2$ y $q = 4$, entonces para $t = 4$ se tendrá que $p^* = \min\{2,3\} = 2$ y $q^* = \min\{4,3\} = 3$. Estos valores

indican cuántas observaciones y errores hay disponibles desde 1 hasta $t-1$. Entonces, el modelo ARMA que usaremos será el modelo incompleto

$$X_t = \sum_{i=1}^{p^*} \alpha_i \, X_{t-i} + Z_t + \sum_{j=1}^{q^*} \beta_j \, Z_{t-j}. \tag{4.18}$$

En este caso,

$$E(X_t \mid \underline{X}_{t-1} = \underline{x}_{t-1}) = \sum_{i=1}^{p^*} \alpha_i \, x_{t-i}.$$

$$\mathrm{Var}(X_t \mid \underline{X}_{t-1}) = \sigma^2 \sum_{j=0}^{q^*} \beta_j^2,$$

donde $\beta_0 := 1$. Por lo tanto,

$$X_t \mid \underline{X}_{t-1} = \underline{x}_{t-1} \sim N\left(\sum_{i=1}^{p^*} \alpha_i \, x_{t-i} \, , \, \sigma^2 \sum_{j=0}^{q^*} \beta_j^2 \right). \tag{4.19}$$

Cuando el tamaño de la serie observada es muy grande con respecto a los valores de p y q, la función de verosimilitud incompleta que se obtiene tendrá una optimización muy parecida a la que tendría la función de verosimilitud exacta, considerando datos desconocidos del pasado.

Por último, la distribución no condicionada de X_1 será normal, y como estamos suponiendo que $E(X_t) = 0$, entonces

$$X_1 \sim N(0, \mathrm{Var}(X_1)). \tag{4.20}$$

Como se ha visto, el cálculo de $\gamma(0) = \mathrm{Var}(X_1)$ depende de cada modelo y puede ser de difícil obtención.

Finalmente, la función de verosimilitud (4.16) se obtiene al substituir las funciones de densidad de acuerdo con las distribuciones dadas en (4.17), (4.19) y (4.20).

Ejemplo 4.14 (Modelo AR(1)**)**
En este caso $p = 1$, $q = 0$ *y* $r = 1$, *por lo cual sólo se necesitan los casos* (4.17) *y* (4.20). *Para* $t = 2, 3, \ldots, n$, *se tiene que*

$$X_t \mid \underline{X}_{t-1} = \underline{x}_{t-1} \sim N\left(\alpha_1 \, x_{t-1} \, , \, \sigma^2\right).$$

Por otro lado, $X_1 \sim N(0, \sigma^2/(1-\alpha_1^2))$, *donde el cálculo de* $\mathrm{Var}(X_1)$ *se muestra en la proposición 3.4. Por lo tanto,*

$$L(\alpha, \beta, \sigma^2) = f(x_1) \prod_{t=2}^{n} f(x_t \mid x_{t-1}),$$

en donde

$$f(x_1) = \frac{1}{\sqrt{2\pi\sigma^2/(1-\alpha_1^2)}} \exp\left(-\frac{1}{2} \frac{x_1^2}{\sigma^2/(1-\alpha_1^2)}\right)$$

y

$$f(x_t \mid x_{t-1}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{(x_t - \alpha_1 \, x_{t-1})^2}{\sigma^2}\right).$$

■

Ejemplo 4.15 (Modelo ARMA(1, 1)**)**
En este caso $p = 1$, $q = 1$ *y* $r = 1$, *nuevamente, sólo se necesitan los casos* (4.17) *y* (4.20). *Para* $t = 2, 3, \ldots, n$, *se tiene que*

$$X_t \mid \underline{X}_{t-1} = \underline{x}_{t-1} \sim N\left(\alpha_1 \, x_{t-1} \, , \, \sigma^2(1+\beta_1^2)\right).$$

Por otro lado, $X_1 \sim N(0, (1 + (\alpha_1 + \beta_1)^2/(1-\alpha_1^2))\sigma^2)$, *donde el cálculo de* $\mathrm{Var}(X_1)$ *se muestra en la proposición 3.11. Por lo tanto,*

$$L(\alpha, \beta, \sigma^2) = f(x_1) \prod_{t=2}^{n} f(x_t \mid x_{t-1}),$$

en donde usando la notación $\sigma_{x_1}^2 = (1 + (\alpha_1 + \beta_1)^2/(1-\alpha_1^2))\sigma^2$, *se tiene que*

$$f(x_1) = \frac{1}{\sqrt{2\pi\sigma_{x_1}^2}} \exp\left(-\frac{1}{2} \frac{x_1^2}{\sigma_{x_1}^2}\right)$$

y

$$f(x_t \mid x_{t-1}) = \frac{1}{\sqrt{2\pi\sigma^2(1+\beta_1^2)}} \exp\left(-\frac{1}{2} \frac{(x_t - \alpha_1 \, x_{t-1})^2}{\sigma^2(1+\beta_1^2)}\right).$$

■

Ejemplo 4.16 (Modelo MA(1)**)**
En este caso $p = 0$, $q = 1$ y $r = 1$, y las esperanzas y varianzas de X_t coinciden cuando son condicionadas y no condicionadas a los valores anteriores de la serie. Para $t = 2, 3, \ldots, n$, se tiene que

$$X_t \mid \underline{X}_{t-1} = \underline{x}_{t-1} \sim N\left(0,\, \sigma^2(1+\beta_1^2)\right).$$

Y $X_1 \sim N\left(0,\, \sigma^2(1+\beta_1^2)\right)$. Por lo tanto,

$$L(\alpha, \beta, \sigma^2) = f(x_1) \prod_{t=2}^{n} f(x_t \mid x_{t-1}),$$

en donde

$$f(x_1) = \frac{1}{\sqrt{2\pi\sigma^2(1+\beta_1^2)}} \exp\left(-\frac{1}{2}\frac{x_1^2}{\sigma^2(1+\beta_1^2)}\right)$$

y

$$f(x_t \mid x_{t-1}) = \frac{1}{\sqrt{2\pi\sigma^2(1+\beta_1^2)}} \exp\left(-\frac{1}{2}\frac{x_t^2}{\sigma^2(1+\beta_1^2)}\right).$$

■

Enfoque 2. Usando distribución normal multivariada

En el enfoque anterior se tiene la desventaja de que para realizar cálculos con valores grandes de p y/o q se tienen que manejar demasiados términos y es difícil explorar más propiedades de la función de verosimilitud, razón por la cual es útil conocer otra forma de obtenerla aprovechando propiedades del álgebra lineal y de la distribución normal de los errores, que implican una distribución normal de las variables aleatorias que forman la serie.

Como $X_1, \ldots, X_n$ son variables aleatorias, tales que cada una tiene distribución normal, su media es cero y $\mathrm{Cov}(X_t, X_{t-k}) = E(X_t X_{t-k}) = \gamma(k)$, entonces la densidad conjunta sigue una distribución normal multivariada, donde el vector de medias es un vector de n ceros y la matriz de varianzas y covarianzas está dada por

$$\mathbf{\Sigma} = \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(n-2) \\ \vdots & \vdots & \cdots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(0) \end{pmatrix}. \tag{4.21}$$

Debemos suponer que esta matriz tiene inversa (es no singular) y la denotamos por $\mathbf{\Sigma}^{-1}$. Además, se tiene que $\mathbf{X}$ representa el vector columna cuyas entradas son $X_1, \ldots, X_n$ y $\mathbf{X}'$ representa el vector transpuesto de $\mathbf{X}$. Nótese que $\mathbf{\Sigma} = \mathbf{E}(\mathbf{X}\,\mathbf{X}')$. Entonces, la función de verosimilitud está dada por

$$L(\alpha, \beta, \sigma^2) = (2\pi)^{-n/2}(\det \mathbf{\Sigma})^{-1/2} \exp\left(-\frac{1}{2}\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X}\right). \tag{4.22}$$

Aquí no se presentan los detalles debido al enfoque introductorio de este libro, pero la expresión anterior puede tener una forma más compacta realizando ajustes adicionales. El lector interesado puede consultar ya sea en el libro de Hamilton (1994) [18] o en el libro de Brockwell y Davis (1987) [7] para una exposición más detallada, y además consultar algunos de los métodos numéricos para realizar la optimización de esta función.

Cálculo de la serie de residuos

Suponiendo que los parámetros ya han sido estimados, es posible realizar el cálculo de la serie de residuos, también llamados residuales. Los residuos o errores $\widehat{z}_t$, en muchas ocasiones también denotados como ϵ_t, son los valores numéricos de las variables aleatorias Z_t después de la estimación de los parámetros. A continuación se describe el procedimiento para calcularlos.

Supongamos una serie de n observaciones conocidas (disponibles) expresada por $x_1, x_2, \ldots, x_n$ para la cual hemos ajustado un modelo ARMA(p, q), obviamente con $n > \text{máx}\{p, q\}$. Para calcular la serie de los residuos $\widehat{z}_1, \widehat{z}_2, \ldots, \widehat{z}_n$, se puede comenzar suponiendo $\widehat{x}_1 = \overline{x}_n$, entonces $\widehat{z}_1 = x_1 - \widehat{x}_1$, y para valores de $t = 2, \ldots, n$,

$$\widehat{x}_t = \sum_{i=1}^{p^*} \widehat{\alpha}_i x_{t-i} + \sum_{j=1}^{q^*} \widehat{\beta}_j \widehat{z}_{t-j}, \tag{4.23}$$

y

$$\widehat{z}_t = x_t - \widehat{x}_t. \tag{4.24}$$

Los valores de p^* y q^* se refieren al número de observaciones y/o residuos disponibles para el cálculo y las podemos definir para cada valor de t como

$$p^* = \text{mín}\{t-1, p\} \quad \text{y} \quad q^* = \text{mín}\{t-1, q\}. \tag{4.25}$$

Conociendo los valores de los residuos se aplican las pruebas de correcta especificación y se determina si los parámetros estimados del modelo son estadísticamente significativos.

4.3. Verificación

En esta sección se exponen algunas de las pruebas de hipótesis más conocidas para verificar la correcta especificación del modelo. Lo anterior consiste en probar que se cumplen las hipótesis de los modelos ARMA (y/o ARIMA) sobre los residuos. Además, se habla sobre un criterio para elegir un modelo sobre otro cuando ambos parecen ser adecuados para una misma serie de tiempo.

Prueba de no heterocedasticidad

La prueba de no heterocedasticidad de White (1980) [36] se basa en una regresión auxiliar del cuadrado de los residuos (o errores) $\hat{z}^2$ sobre todos los regresores originales x_{it}, sus cuadrados x_{it}^2 y los productos cruzados de los regresores. La prueba contrasta la hipótesis de homocedasticidad H_0 contra la hipótesis alternativa de heterocedasticidad H_a de la siguiente manera:

$$H_0 : E(Z_i^2) = \sigma^2 \text{ para todo valor de } i$$
$$\text{vs}$$
$$H_a : E(Z_i^2) \neq \sigma^2 \text{ al menos para un valor de } i.$$

El estadístico de prueba de White se calcula como nR^2 donde R^2 es el coeficiente de determinación de la regresión auxiliar y n es el tamaño de la muestra.

Rechazamos H_0 en el nivel α si el estadístico de prueba excede el valor crítico de una distribución χ^2 con grados de libertad igual al número de regresores incluidos en la regresión auxiliar.

De manera más explícita, consideremos el modelo

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \widehat{z}_i, \quad i = 1, \ldots, n. \tag{4.26}$$

La hipótesis a contrastar es

$$H_0 : E(\hat{z}_i^2) = \sigma^2 \text{ vs. } H_a : E(\hat{z}_i^2) = \sigma_i^2.$$

El modelo para la varianza está dado por

$$\sigma_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2i} X_{3i} + u_i.$$

La hipótesis que se contrasta es

$$\begin{gathered} H_0 : \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = 0, \\ \text{vs.} \\ H_a : \alpha_i \neq 0 \text{ para alguna } i = 2, \ldots, 6. \end{gathered}$$

Notemos que bajo H_0 la varianza de los errores es constante,

$$\sigma_i^2 = \alpha_1.$$

Pasos para realizar el contraste

- Estimar por MC la regresión de interés y calcular los errores.
- Estimar por MC la siguiente regresión auxiliar,

  $$e_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2i} X_{3i} + u_i$$

  y calcular el coeficiente de determinación R^2.
- Calcular el estadístico de prueba como nR^2, que sigue asintóticamente una distribución χ^2 con $k-1$ grados de libertad, donde k es el número de parámetros de la regresión auxiliar.
- La hipótesis H_0 se rechaza al nivel de significancia α, si $nR^2 > c$, donde c es el cuantil $1-\alpha$ de una distribución χ^2 con $k-1$ grados de libertad.

Adicional a lo anterior, la gráfica de la serie de tiempo de los residuos es una herramienta útil para visualizar la variabilidad de los errores y formarnos una idea de si el supuesto de varianza constante se cumple o no. Si visualmente va cambiando la amplitud de la serie en el tiempo, habrá indicios de heterocedasticidad.

Prueba de normalidad

Habitualmente las pruebas de normalidad de los errores se realizan usando la información de los coeficientes de asimetría y curtosis.

Consideremos una variable aleatoria X_t con media μ y varianza σ_x^2. Además, definimos y denotamos el i-ésimo momento centrado como $\mu_i := E[X_t - \mu]^i$. Nótese, por ejemplo, que usando esta notación se tiene que $\sigma_x^2 = \mu_2$. Los coeficientes de asimetría y curtosis se definen como

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} \quad y \quad \gamma_2 = \frac{\mu_4}{\mu_2^2}.$$

Dado que el supuesto sobre los errores es que $Z_t \sim N(0, \sigma^2)$, entonces, teóricamente, $\gamma_1 = 0$ y $\gamma_2 = 3$. Bajo la hipótesis nula de normalidad de los errores, es posible comparar los momentos de los errores estimados con los momentos teóricos de la distribución normal. Considerando estos aspectos teóricos, la prueba de Jarque y Bera (1980, 1987) [21, 22] tiene el siguiente estadístico:

$$\text{jb} = n\left(\frac{(g_1)^2}{6} + \frac{(g_2 - 3)^2}{24}\right),$$

en donde las estimaciones muestrales están dadas por

$$g_1 = \frac{m_3}{m_2^{3/2}} \quad \text{y} \quad g_2 = \frac{m_4}{m_2^2}; \quad \text{con } m_i = \frac{1}{n}\sum_{t=1}^{n}(x_t - \overline{x})^i.$$

Bajo la hipótesis nula de que la variable se distribuye normal, el estadístico de prueba jb se distribuye asintóticamente como una variables aleatoria χ^2 con dos grados de libertad, $\chi^2_{(2)}$. Lo anterior debido a que jb se calcula como la suma de los cuadrados de dos variables independientes asintóticamente normales estandarizadas, véase Bowman y Shenton (1975) [3].

El criterio de decisión de la prueba es que si el estadístico de prueba jb es mayor al valor del cuantil $(1 - \alpha)$ de la variable $\chi^2_{(2)}$, es decir, $\chi^2_{(2)}(1 - \alpha)$, se rechaza la hipótesis nula al nivel de significancia α. Análogamente, si el $p - value = P(\chi^2_{(2)} \geqslant \text{jb})$ es menor a α, entonces la hipótesis nula H_0 se rechaza al nivel de significancia α.

Prueba de no autocorrelación

La prueba de Ljung-Box (Portmanteau) consiste en determinar si las m primeras autocorrelaciones simples de X_t son conjuntamente significativas. El estadístico de la prueba de Ljung-Box para la hipótesis nula de no autocorrelación está dado por

$$Q_{LB} = n(n+2) \sum_{k=1}^{m} \frac{\widehat{\rho}_k^2}{n-k}, \tag{4.27}$$

donde n es el número de observaciones en la serie de tiempo y cada coeficiente de autocorrelación simple $\widehat{\rho}_k^2$ está ponderado por el número $n-k$. El estadístico de prueba Q_{LB} converge en distribución a una χ^2 con m grados de libertad. Bajo la hipótesis nula, $H_0 : \rho_1 = \rho_2 = \ldots = \rho_k = 0$, la prueba Portmanteau de Box y Pierce (1970) [6] de ruido blanco fue retomada y refinada en la prueba de Ljung y Box (1978) [27].

Otra manera de detectar posible correlación es observando los gráficos de las funciones de autocorrelación simple (ACF) y parcial (PACF), aunque como todo criterio gráfico, se debe tener cuidado en formalizar con las pruebas estadísticas las conjeturas visuales.

El criterio de Akaike

Supongamos que mediante las funciones ACF y PACF se ha identificado una colección de modelos ARMA para diferentes órdenes de p y q, y que además cumplen con los supuestos probabilísticos de no autocorrelación de los errores, homocedasticidad y distribución normal. Para decidir cuál modelo describe mejor la serie de tiempo se procede a comparar los modelos en competencia mediante criterios de información y medidas de precisión de pronóstico.

Si dos de los modelos seleccionados son anidados, en el sentido de que un modelo es una restricción del otro (por ejemplo, un modelo AR(2) versus un modelo ARMA$(2,2)$), se puede seleccionar el mejor modelo aplicando el principio de modelación de lo general a lo específico (*general-to-specific modelling*). Dicha metodología consiste en simplificar un modelo general

que caracteriza adecuadamente la serie observada, mediante pruebas estadísticas que justifican la reducción de parámetros estimados en el modelo inicial. Se pueden consultar los detalles de esta metodología en Campos et al. (2005) [9].

Si los modelos en competencia no están anidados, entonces se pueden utilizar criterios de información como el número AIC y el número BIC, en los que toma relevancia el número de parámetros k de un modelo. Por ejemplo, para un modelo ARMA(p, q) se tiene que $k = p + q + 1$; si el modelo se estima con constante, entonces $k = p + q + 2$, y después se elige el modelo que minimiza el número AIC o el número BIC.

El criterio de información de Akaike (AIC, por sus siglas en inglés) fue desarrollado en 1974 [1] para series temporales, el cual está dado por

$$\text{AIC} = -2\,\text{LL} + k, \tag{4.28}$$

donde LL denota el logaritmo de la verosimilitud del modelo estimado y k es el número de parámetros estimados. Entre menor es el AIC, mayor es la verosimilitud del modelo y el ajuste es mejor.

Otro criterio de ajuste es el llamado criterio de información bayesiano de Schwarz (1978) [34]. Este valor, denotado por las siglas BIC, está definido como

$$\text{BIC} = -2\,\text{LL} + k\,\ln T, \tag{4.29}$$

en donde LL y k tienen el mismo significado que en (4.28) y T representa el tamaño de la muestra. De la misma forma que el AIC, entre menor es el BIC, mayor es la verosimilitud del modelo y mejor el ajuste.

Error de pronóstico

La bondad de predicción de un modelo se puede evaluar por medio de las magnitudes de los errores de predicción, definidos como las diferencias entre el valor observado y su pronóstico estimado. Adicional a los criterios anteriores, la medición de estos errores constituye un criterio para elegir entre dos

modelos. La decisión de cuál es el mejor modelo para una serie de tiempo requiere considerar la precisión con la que cada modelo calcula nuevos datos.

En la práctica, cuando se ajusta un modelo, es común dividir los datos disponibles en dos conjuntos: el primero denominado datos de entrenamiento y el segundo conocido como datos de prueba. Los datos de entrenamiento se utilizan para estimar los parámetros del modelo, mientras que los datos de prueba sirven para evaluar la precisión de las predicciones estimadas mediante dicho modelo. La comparación entre los datos de prueba y las predicciones debería proporcionar un indicador confiable de la bondad del modelo para pronosticar nuevos datos, debido a que los datos de prueba no se emplean para determinar dichas predicciones.

Uno de los criterios más comunes para decidir el mejor modelo de predicción entre varios candidatos, consiste en elegir aquel modelo con menor error cuadrático medio (ECM) de los errores de pronóstico. Más adelante, en la sección 5 se comenta más ampliamente sobre los pronósticos y el ECM.

4.4. Ejemplo con datos reales

En esta sección se desarrolla un modelo para la serie de tiempo de la temperatura promedio mensual de una región tropical en el periodo de enero de 1975 a diciembre de 2000, medida en grados centígrados. Primero se presenta un breve análisis descriptivo de los datos y posteriormente se aplica la metodología de Box-Jenkins para seleccionar el modelo estadísticamente más adecuado.

En el Apéndice se pueden consultar los códigos utilizados. El análisis se hizo con ayuda del software estadístico `R`, donde se hicieron las pruebas estadísticas sobre los residuos, estimaciones y pronósticos con la ayuda de los siguientes paquetes de funciones: `aTSA`, `astsa`, `forecast`, `tsoutliers` y `whitestrap`.

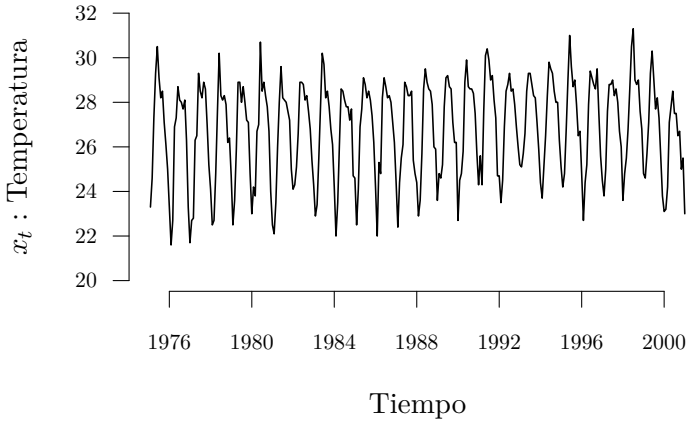

Figura 4.9: Serie de tiempo de las temperaturas medias mensuales x_t de enero de 1975 a diciembre de 2000.

Análisis de los datos

La temperatura promedio durante el mes t se denotará como x_t, para $t = 1, 2, \ldots, 312$, en donde $t = 1$ corresponde al mes de enero de 1975 y $t = 312$ corresponde al mes de diciembre del año 2000.

En el periodo de estudio la menor temperatura media mensual fue de 21.6°C y se registró en enero de 1976 y la más alta fue de 31.3°C en el mes de junio de 1998. Finalmente, la temperatura media mensual promedio del periodo de estudio fue de 26.8°C.

En la figura 4.9 se observa que la serie de tiempo aparentemente tiene un valor medio y una varianza constantes. También se puede observar que no presenta un componente de tendencia, aunque sí se observa un componente de estacionalidad anual de 12 periodos, inducido por la periodicidad mensual de los datos. Tales características nos permiten suponer que es probable que los datos no se comporten como un proceso estacionario.

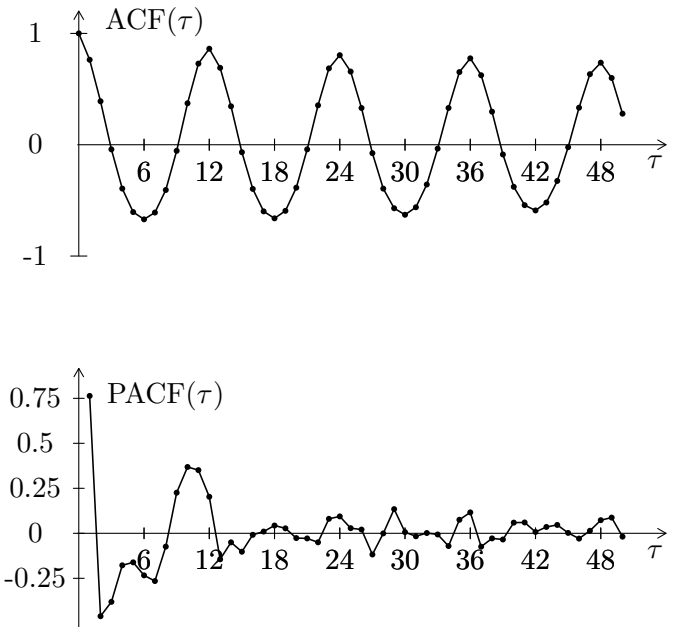

Figura 4.10: Funciones ACF y PACF de la serie $\{x_t\}$.

Con la finalidad de reconocer evidencia acerca de la estacionariedad de la serie de tiempo, se analiza la gráfica de la función de autocorrelación muestral (ACF). En la figura 4.10 se presenta la gráfica de la función ACF muestral de las primeras 50 autocorrelaciones de la serie $\{x_t\}$. Se observa que las autocorrelaciones disminuyen lentamente hacia cero, evidenciando que la serie no se comporta como un proceso ruido blanco, ya que la autocorrelación de una serie estacionaria decae hacia cero relativamente rápido, mientras que para una serie no estacionaria disminuirá lentamente. También es útil observar el comportamiento de la función de autocorrelación parcial (PACF) muestral, que permite notar que las primeras autocorrelaciones parciales no parecen ser estadísticamente igual a cero.

Ahora bien, con la finalidad de atenuar el componente cíclico de la serie $\{x_t\}$, aplicamos una diferencia estacional, es decir, una diferenciación respecto del

rezago 12, ya que la periodicidad de la serie de tiempo es mensual. La serie de tiempo diferenciada se calcula como

$$y_t = \nabla_{12} x_t,$$

en donde

$$\nabla_{12} x_t = x_t - x_{t-12} = (1 - B^{12}) x_t.$$

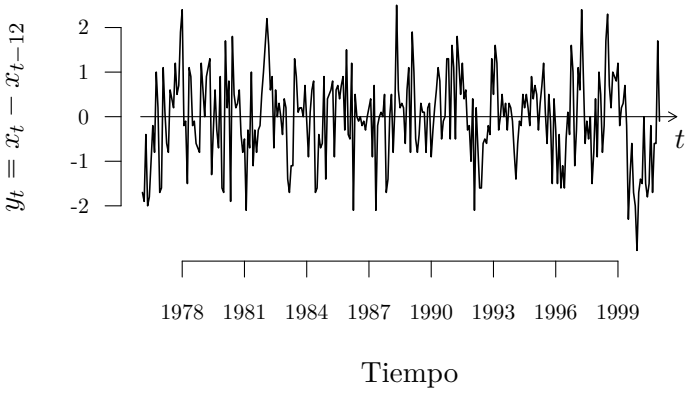

Figura 4.11: Serie de tiempo de la diferencia estacional de la temperatura media mensual y_t de enero de 1971 a diciembre de 2000.

La variable y_t, para $t = 1, 2, \ldots, 300$ representa la variación de la temperatura con respecto a la observada en el mismo mes del año anterior. El comportamiento de la serie $\{y_t\}$ se muestra en la figura 4.11, donde se puede observar que se asemeja al de una variable estacionaria, es decir, con media y varianza constantes y sin comportamientos cíclicos evidentes.

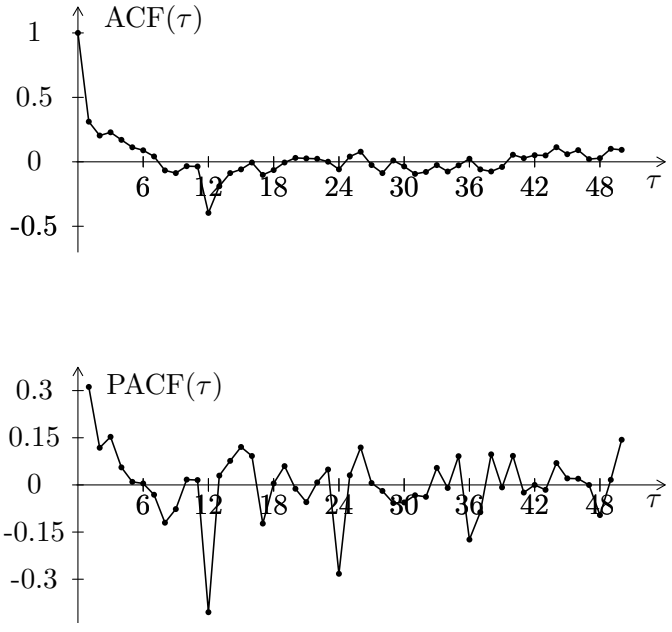

Figura 4.12: Funciones ACF y PACF de la serie $\{y_t\}$.

El comportamiento de las funciones ACF y PACF de la serie diferenciada se muestra en la gráfica 4.12. En la función ACF se puede observar que los valores que se toman respecto a rezagos más lejanos disminuyen hacia cero de manera rápida. También se observa que únicamente las autocorrelaciones respecto de los primeros rezagos y respecto del rezago 12 son grandes. En la gráfica de la función PACF se observa que las autocorrelaciones parciales respecto de los primeros rezagos y de los rezagos 12 y 24 son grandes.

Los pocos saltos observados en los gráficos anteriores no son un impedimento para sospechar que la serie $\{y_t\}$ se comporta como una variable estacionaria.

Para determinar si la serie de tiempo se comporta como un proceso estacionario, además de analizar el comportamiento de las funciones de autocorrelación y autocorrelación parcial, realizamos pruebas para determinar el

orden de integración de la serie. A continuación se presentan los resultados de la prueba de hipótesis Dickey-Fuller aumentada (ADF), descrita en la sección 4.1 de las series $\{x_t\}$ y $\{y_t\}$, donde

H_0: La serie tiene una raíz unitaria,

es decir, la serie no es estacionaria.

Lo primero que hay que decidir es con cuántos rezagos realizar la prueba ADF, es decir, cuántos rezagos de la diferencia ∇x_{t-i} se deben incluir en la ecuación

$$\nabla x_t = \mu + \beta x_{t-1} + \delta t + \sum_{i=1}^{s} \gamma_i \nabla x_{t-i} + \widehat{z}_t, \tag{4.30}$$

y si los parámetros μ y δ son distintos de cero en el modelo de la prueba.

La decisión de cuántos rezagos incluir en la ecuación (4.30) (el valor de s) está relacionada con el orden autorregresivo que suponemos tiene la serie $\{x_t\}$, así como del tamaño del ciclo presente en los datos para realizar la prueba.

La decisión acerca de los parámetros μ y δ de la ecuación (4.30) de la prueba ADF genera las siguientes especificaciones:

- Sin constante y sin tendencia ($\mu = 0$ y $\delta = 0$).
- Con constante y sin tendencia ($\mu \neq 0$ y $\delta = 0$).
- Con constante y con tendencia ($\mu \neq 0$ y $\delta \neq 0$).

De acuerdo con la información de la gráfica de la serie de tiempo de las temperaturas esbozada en la figura 4.9, en donde se observa que no tiene un componente de tendencia determinística, y que su valor medio es distinto de cero, la especificación de la ecuación (4.30) con constante y sin tendencia parece la más adecuada. Sin embargo, es adecuado considerar también los resultados de las otras especificaciones para concluir acerca de la estacionariedad de la serie de tiempo.

	$\mu = 0$ y $\delta = 0$		$\mu \neq 0$ y $\delta = 0$		$\mu \neq 0$ y $\delta \neq 0$	
Rezago	ADF	*p-value*	ADF	*p-value*	ADF	*p-value*
0	−0.485	0.50	−6.310	0.01	−6.280	0.01
1	−0.747	0.41	−10.01	0.01	−10.04	0.01
2	−0.865	0.37	−12.90	0.01	−13.06	0.01
3	−0.737	0.42	−12.38	0.01	−12.66	0.01
4	−0.635	0.45	−11.73	0.01	−12.18	0.01
5	−0.507	0.50	−12.27	0.01	−12.99	0.01
6	−0.457	0.51	−13.17	0.01	−14.60	0.01
7	−0.425	0.52	−11.37	0.01	−13.50	0.01
8	−0.308	0.55	−7.150	0.01	−8.790	0.01
9	−0.279	0.56	−4.170	0.01	−5.120	0.01
10	−0.384	0.53	−2.450	0.15	−2.810	0.24
11	−0.460	0.51	−1.670	0.46	−1.630	0.73
12	−0.286	0.56	−2.010	0.32	−1.980	0.58
13	−0.158	0.60	−2.220	0.24	−2.120	0.52

Nota: *p-value* = 0.01 significa *p-value* $\leqslant 0.01$.

Tabla 4.2: Prueba de Dickey-Fuller aumentada para la serie $\{x_t\}$.

Los resultados de la prueba ADF para la serie de tiempo de las temperaturas medias se muestran en la tabla 4.2. La conclusión de la prueba ADF no es la misma con las tres especificaciones que se muestran en la tabla. Los resultados de la prueba que modela los datos sin tendencia y sin constante ($\mu = 0$ y $\delta = 0$) y que considera términos de ∇x_{t-i} para $i = 1, \ldots, 13$, no rechazan la hipótesis nula de raíz unitaria al 95 % de confianza. Sin embargo, los resultados de las pruebas con constante y sin tendencia ($\mu \neq 0$ y $\delta = 0$), y de la prueba con constante y con tendencia ($\mu \neq 0$ y $\delta \neq 0$) que incluyen los términos de ∇x_{t-i} para $i = 1, \ldots, 9$ rechazan la hipótesis de raíz unitaria, pero cuando se consideran los términos de ∇x_{t-i} para $i = 10, 11, 12, 13$, los resultados tampoco rechazan la hipótesis de raíz unitaria. Por lo tanto, debemos concluir que existe evidencia estadística para no rechazar la hipótesis de raíz unitaria, es decir, no se puede decir que la serie $\{x_t\}$ sea estacionaria.

	$\mu = 0$ y $\delta = 0$		$\mu \neq 0$ y $\delta = 0$		$\mu \neq 0$ y $\delta \neq 0$	
Rezago	ADF	*p-value*	ADF	*p-value*	ADF	*p-value*
0	−12.57	0.01	−12.55	0.01	−12.57	0.01
1	−9.10	0.01	−9.08	0.01	−9.12	0.01
2	−6.85	0.01	−6.84	0.01	−6.88	0.01
3	−6.17	0.01	−6.16	0.01	−6.22	0.01
4	−5.85	0.01	−5.83	0.01	−5.92	0.01
5	−5.52	0.01	−5.50	0.01	−5.60	0.01
6	−5.31	0.01	−5.29	0.01	−5.39	0.01
7	−5.72	0.01	−5.71	0.01	−5.81	0.01
8	−5.70	0.01	−5.69	0.01	−5.76	0.01
9	−5.32	0.01	−5.31	0.01	−5.40	0.01
10	−5.17	0.01	−5.16	0.01	−5.29	0.01
11	−7.91	0.01	−7.90	0.01	−8.06	0.01
12	−6.81	0.01	−6.80	0.01	−6.94	0.01
13	−5.84	0.01	−5.83	0.01	−6.01	0.01

Nota: *p-value* = 0.01 significa *p-value* $\leqslant$ 0.01.

Tabla 4.3: Prueba de Dickey-Fuller aumentada para la serie $\{y_t\}$.

En la tabla 4.3 se presentan los resultados de la prueba ADF para la serie diferenciada con respecto al rezago 12 de la temperatura: $y_t = \nabla x_{t-12}$. En las distintas especificaciones de la prueba, se rechaza la hipótesis nula de raíz unitaria al 99 % de confianza. Este resultado, junto con el comportamiento de las funciones ACF y PACF de la serie de tiempo $\{y_t\}$, nos permite concluir que esta serie es estacionaria.

Es importante tener en cuenta que, dado que la serie de tiempo de la temperatura es integrada de orden 1, la transformación $w_t = (1-B)(1-B^{12})x_t$ implicaría una sobrediferenciación.

Es frecuente modelar las series de tiempo estacionales (que se caracterizan por exhibir correlaciones más fuertes en los rezagos estacionales) como procesos SARIMA porque contienen relativamente pocos parámetros para modelarlas. Sin embargo, en la práctica se debe elegir el modelo adecuado para las características de la serie observada.

Por lo anterior, para considerar las características estacionales de la serie de tiempo $\{x_t\}$, se explora un modelo para la diferencia estacional $y_t = (1 - B^{12})x_t$ de las temperaturas medias, es decir, $y_t = \nabla_{12} x_t$. Puesto que la serie $\{y_t\}$ cumple con ser estacionaria, es adecuado representarla mediante un modelo ARMA(p, q).

Identificación

El objetivo principal en la fase de identificación de la metodología Box-Jenkins es reconocer el comportamiento de procesos estacionarios ARMA(p, q) a través del comportamiento de las funciones ACF y PACF muestrales.

Es importante señalar que un valor grande en la función ACF o en la función PACF con respecto al k-ésimo rezago puede no ser relevante si se encuentra aislado. En el caso de la serie $\{y_t\}$, tanto en la función ACF como en la función PACF se observa que se reducen rápidamente sus valores hacia cero; en particular en la función ACF a partir del rezago cuatro y en la función PACF a partir del rezago tres.

En la gráfica de la función ACF se puede observar que el valor de la autocorrelación con respecto al rezago 12 es alto, pero aislado. En el caso de la función PACF los valores con respecto a los rezagos 12, 24 y 36 son grandes, pero decrecientes y aislados. Los valores altos en las funciones ACF y PACF con respecto al rezago 12 nos lleva a considerar empíricamente el modelo ARMA$(12, 12)$ para la serie $\{y_t\}$. Además, se deben considerar alternativas con menos parámetros que posiblemente tengan un mejor ajuste. Por tanto, también se estudian dos modelos ARMA con algún valor de p o q igual a 12 y el otro con un valor menor a 12. El modelo que aquí se prueba considerando la información del rezago 12 de la función PACF, y reduciendo los términos de medias móviles, es el modelo ARMA$(12, 4)$. El modelo que se especifica para considerar el rezago 12 de la función ACF, y que además reduce los parámetros autorregresivos, es el modelo ARMA$(3, 12)$.

A partir de la información anterior, se presentan los siguientes modelos

ARMA para la serie $\{y_t\}$ y un modelo SARIMA para la serie $\{x_t\}$, con la finalidad de modelar mejor el efecto estacional.

1. El primer modelo que se analiza para la serie $\{y_t\}$ es un modelo ARMA con orden autorregresivo igual a 12 ($p = 12$) y con orden en medias móviles también de 12 ($q = 12$). Se hace énfasis en que el valor $p = 12$ se identifica debido a que en la gráfica de la función PACF de la serie $\{y_t\}$ que se muestra en la figura 4.12, se observa que la autocorrelación parcial respecto del rezago número 12 es el más grande. De manera similar, observando la gráfica de la función ACF de la serie $\{y_t\}$ (figura 4.12) se observa que la autocorrelación respecto del rezago 12 es la más grande, por lo que se elige $q = 12$. En este caso, el modelo ARMA$(12, 12)$ a estimar para la serie $\{y_t\}$ es

$$Y_t = \alpha + \sum_{i=1}^{12} (\alpha_i Y_{t-i} + \beta_i Z_{t-i}) + Z_t.$$

2. El segundo modelo propuesto es un ARMA$(12, 4)$, autorregresivo de orden 12 y medias móviles de orden 4. El modelo a estimar es

$$Y_t = \alpha + \sum_{i=1}^{12} \alpha_i Y_{t-i} + \sum_{j=1}^{4} \beta_j Z_{t-j} + Z_t.$$

3. Un tercer modelo propuesto es un ARMA$(3, 12)$ autorregresivo de orden 3 con medias móviles de orden 12. El modelo a estimar es

$$Y_t = \alpha + \sum_{i=1}^{3} \alpha_i Y_{t-i} + \sum_{j=1}^{12} \beta_j Z_{t-j} + Z_t.$$

4. Una cuarta propuesta es un modelo SARIMA, en la notación de la Definición 3.18 ARIMA$(3, 0, 0)[0, 1, 1]_{12}$ para la serie $\{x_t\}$. Es decir, se modelan los datos de las temperaturas mensuales mediante un modelo que considera (con respecto al rezago $s = 12$) una diferenciación estacional ($D = 1$), medias móviles estacionales de orden uno ($Q = 1$) y orden autorregresivo (no estacional) igual a tres ($p = 3$).

ARMA(12, 12)			ARMA(12, 4)			ARMA(3, 12)		
	coef.	e.e.		coef.	e.e.		coef.	e.e.
ar1	0.362	0.097	ar1	0.036	0.126	ar1	0.348	0.070
ar2	0.058	0.107	ar2	−0.168	0.086	ar2	0.072	0.079
ar3	0.244	0.108	ar3	−0.244	0.084	ar3	0.207	0.073
ar4	−0.002	0.116	ar4	−0.210	0.090	ar4		
ar5	0.054	0.114	ar5	0.225	0.064	ar5		
ar6	0.012	0.110	ar6	0.167	0.060	ar6		
ar7	−0.019	0.113	ar7	0.121	0.060	ar7		
ar8	−0.144	0.099	ar8	−0.050	0.058	ar8		
ar9	−0.049	0.105	ar9	−0.038	0.058	ar9		
ar10	0.071	0.101	ar10	−0.005	0.056	ar10		
ar11	0.137	0.102	ar11	0.072	0.054	ar11		
ar12	−0.014	0.082	ar12	−0.423	0.058	ar12		
ma1	−0.084	0.091	ma1	0.197	0.138	ma1	−0.052	0.050
ma2	0.100	0.098	ma2	0.367	0.071	ma2	0.070	0.059
ma3	−0.152	0.102	ma3	0.590	0.088	ma3	−0.123	0.060
ma4	0.024	0.098	ma4	0.520	0.119	ma4	0.031	0.052
ma5	−0.126	0.100	ma5			ma5	−0.068	0.053
ma6	0.043	0.118	ma6			ma6	0.042	0.049
ma7	−0.040	0.116	ma7			ma7	−0.043	0.049
ma8	0.081	0.105	ma8			ma8	−0.028	0.053
ma9	0.029	0.110	ma9			ma9	−0.035	0.052
ma10	0.038	0.100	ma10			ma10	0.004	0.055
ma11	0.008	0.105	ma11			ma11	0.086	0.057
ma12	−0.924	0.110	ma12			ma12	−0.884	0.048
cte	0.023	0.016	cte	−0.026	0.081	cte	0.027	0.012

Tabla 4.4: Parámetros (coef.) y sus respectivos errores estándar (e.e.) para los modelos ARMA propuestos para la serie de tiempo $\{y_t\}$.

Las estimaciones del parámetro α y de los coeficientes α_i y β_j (redondeados a tres decimales) se encuentran en la tabla 4.4 para los modelos ARMA. La estimación del modelo SARIMA se muestra en la tabla 4.5. En esta última, la abreviación `sma1` se refiere al parámetro estacional de medias móviles.

	coef.	error estándar
ar1	0.2595	0.0573
ar2	0.1713	0.059
ar3	0.1233	0.059
sma1	-0.8873	0.0445
cte	0.002	0.0013
AIC = 675.16		
Prueba Jarque Bera χ^2 (Ji-cuadrada) = 3.4093, g.l. = 2, *p-value* = 0.1818		
Prueba de White χ^2 (Ji-cuadrada) = 16.26, g.l. = 2, *p-value* = 0.0003		
Prueba de Ljung-Box rezagos = 24, estadístico = 24.69, *p-value* = 0.213 rezagos = 36, estadístico = 40.78, *p-value* = 0.137 rezagos = 42, estadístico = 43.61, *p-value* = 0.245		

Tabla 4.5: Parámetros estimados y pruebas del modelo SARIMA: ARIMA$(3,0,0)[0,1,1]_{12}$ para la serie $\{x_t\}$.

Estas cuatro especificaciones de modelos de series de tiempo nos permitirán evaluar estadísticamente si es necesario o no, incorporar los términos autorregresivos y de promedios móviles de orden 12. También permiten evaluar cuántos de estos términos autorregresivos y de medias móviles son significativos para modelar la serie, y si es mejor considerar un modelo que capture los efectos estacionales. Estas decisiones se toman en la fase de verificación de la metodología Box-Jenkins que se presenta en la siguiente sección, considerando los criterios de información y las pruebas de correcta especificación de los modelos estimados.

Modelo	ARMA(12, 12)	ARMA(12, 4)	ARMA(3, 12)
No Autocorrelación: Prueba de Ljung-Box			
estadístico	10.341	33.867	14.657
p-value	0.0158	$4.295e-05$	0.1008
rezagos	27	24	24
estadístico	15.839	43.17	23.241
p-value	0.1987	0.00194	0.3312
rezagos	36	36	36
estadístico	18.762	45.099	24.5
p-value	0.4066	0.0115	0.6025
rezagos	42	42	42
Normalidad: Prueba de Jarque-Bera			
estadístico	2.2603	0.5046	2.3379
p-value	0.323	0.777	0.3107
g.l.	2	2	2
No Heterocedasticidad: Prueba de White			
estadístico	0.56	1.57	0.62
p-value	0.754	0.455	0.733
g.l.	2	2	2
AIC	689.26	749.45	679.65

Tabla 4.6: Pruebas de correcta especificación para los modelos ARMA de la serie $\{y_t\}$.

Verificación

En esta fase se evalúan los residuos (errores estimados) de los modelos propuestos de acuerdo con las pruebas estadísticas para verificar que cumplan con los supuestos del modelo. En la tabla 4.6 se presentan los resultados de la prueba Ljung-Box para evaluar si los errores se comportan como un proceso de ruido blanco, la prueba de Jarque-Bera para evaluar

la normalidad, y la prueba de White para evaluar la homocedasticidad de los errores y el coeficiente AIC de cada modelo. Los resultados de las pruebas de correcta especificación del modelo SARIMA se muestra en la tabla 4.5. Finalmente, los resultados de las pruebas ADF de los errores de cada modelo se presentan las Tablas 4.7 - 4.10. Nota: en los resultados que se muestran en las tablas, *p-value* = 0.01 significa *p-value* $\leqslant 0.01$.

En la tabla 4.6 se puede ver que en los tres modelos ARMA existe evidencia estadística suficiente para no rechazar la hipótesis de homocedasticidad ni la hipótesis de la distribución normal de los errores con valores del *p-value* mayores que 0.10.

En el caso de la prueba Ljung-Box, se presentan los resultados de 3 pruebas que consideran cantidades distintas de rezagos, los cuales en conjunción con el número de parámetros estimados $(p + q + 2)$ determinan los grados de libertad del estadístico de la prueba.

	α=0 y δ=0		$\alpha \neq 0$ y $\delta = 0$		$\alpha \neq 0$ y $\delta \neq 0$	
Rezago	ADF	*p-value*	ADF	*p-value*	ADF	*p-value*
0	−17.4	0.01	−17.4	0.01	−17.3	0.01
1	−12.3	0.01	−12.3	0.01	−12.3	0.01
2	−9.86	0.01	−9.85	0.01	−9.83	0.01
3	−8.82	0.01	−8.81	0.01	−8.79	0.01
4	−8.07	0.01	−8.06	0.01	−8.04	0.01
5	−7.24	0.01	−7.23	0.01	−7.21	0.01
6	−6.48	0.01	−6.47	0.01	−6.45	0.01
7	−6.02	0.01	−6.01	0.01	−5.99	0.01
8	−5.49	0.01	−5.48	0.01	−5.46	0.01
9	−5.07	0.01	−5.05	0.01	−5.03	0.01
10	−4.86	0.01	−4.86	0.01	−4.82	0.01
11	−4.80	0.01	−4.80	0.01	−4.75	0.01
12	−4.60	0.01	−4.60	0.01	−4.56	0.01

Tabla 4.7: Prueba de Dickey-Fuller. Errores del modelo ARMA$(12, 12)$.

	α=0 y δ=0		$\alpha \neq 0$ y $\delta = 0$		$\alpha \neq 0$ y $\delta \neq 0$	
Rezago	ADF	*p-value*	ADF	*p-value*	ADF	*p-value*
0	−17.5	0.01	−17.5	0.01	−17.4	0.01
1	−12.6	0.01	−12.6	0.01	−12.6	0.01
2	−10.4	0.01	−10.4	0.01	−10.4	0.01
3	−9.35	0.01	−9.34	0.01	−9.40	0.01
4	−8.35	0.01	−8.35	0.01	−8.43	0.01
5	−7.24	0.01	−7.24	0.01	−7.34	0.01
6	−6.48	0.01	−6.47	0.01	−6.57	0.01
7	−6.06	0.01	−6.06	0.01	−6.17	0.01
8	−5.50	0.01	−5.49	0.01	−5.59	0.01
9	−4.96	0.01	−4.96	0.01	−5.06	0.01
10	−4.67	0.01	−4.67	0.01	−4.82	0.01
11	−5.47	0.01	−5.49	0.01	−5.66	0.01
12	−5.41	0.01	−5.42	0.01	−5.58	0.01

Tabla 4.8: Prueba de Dickey-Fuller. Errores del modelo ARMA$(12, 4)$.

	$\alpha = 0$ y $\delta = 0$		$\alpha \neq 0$ y $\delta = 0$		$\alpha \neq 0$ y $\delta \neq 0$	
Rezago	ADF	*p-value*	ADF	*p-value*	ADF	*p-value*
0	−17.4	0.01	−17.3	0.01	−17.3	0.01
1	−12.5	0.01	−12.4	0.01	−12.4	0.01
2	−10.2	0.01	−10.2	0.01	−10.2	0.01
3	−9.03	0.01	−9.01	0.01	−8.99	0.01
4	−7.93	0.01	−7.92	0.01	−7.89	0.01
5	−6.81	0.01	−6.80	0.01	−6.77	0.01
6	−6.08	0.01	−6.07	0.01	−6.05	0.01
7	−5.98	0.01	−5.97	0.01	−5.94	0.01
8	−5.63	0.01	−5.62	0.01	−5.59	0.01
9	−4.90	0.01	−4.89	0.01	−4.85	0.01
10	−4.42	0.01	−4.42	0.01	−4.35	0.01
11	−4.45	0.01	−4.46	0.01	−4.37	0.01
12	−4.34	0.01	−4.34	0.01	−4.27	0.01

Tabla 4.9: Prueba de Dickey-Fuller. Errores del modelo ARMA$(3, 12)$.

	α=0 y δ=0		$\alpha \neq 0$ y $\delta = 0$		$\alpha \neq 0$ y $\delta \neq 0$	
Rezago	ADF	*p-value*	ADF	*p-value*	ADF	*p-value*
0	−17.7	0.01	−17.7	0.01	−17.6	0.01
1	−12.6	0.01	−12.6	0.01	−12.6	0.01
2	−10.4	0.01	−10.4	0.01	−10.4	0.01
3	−8.73	0.01	−8.72	0.01	−8.71	0.01
4	−7.70	0.01	−7.69	0.01	−7.68	0.01
5	−6.56	0.01	−6.55	0.01	−6.54	0.01
6	−5.87	0.01	−5.86	0.01	−5.84	0.01
7	−5.99	0.01	−5.98	0.01	−5.96	0.01
8	−5.75	0.01	−5.74	0.01	−5.72	0.01
9	−4.97	0.01	−4.96	0.01	−4.93	0.01
10	−4.06	0.01	−4.05	0.01	−4.01	0.01
11	−3.62	0.01	−3.60	0.01	−3.55	0.04
12	−4.07	0.01	−4.06	0.01	−3.99	0.01

Tabla 4.10: Prueba de Dickey-Fuller. Errores del modelo SARIMA: ARIMA$(3,0,0)[0,1,1]_{12}$ de la serie $\{x_t\}$.

En el caso del modelo ARMA$(12,12)$, cuando se considera un total de 27 rezagos, el estadístico de prueba tiene una distribución χ^2 con 3 grados de libertad $(27-(p+q))$ y un *p-value* de 0.0158, lo cual implica que la prueba rechaza el supuesto de no autocorrelación al 95 % de confianza, sin embargo, con un total de rezagos de 36 y 42, el supuesto se satisface al 95 % de confianza.

Ahora, en el caso del modelo ARMA$(12,4)$, aplicando la misma prueba no se cumple con el supuesto de no autocorrelación al 95 % de confianza en ninguno de los tres casos. Esta conclusión también se evidencia en la gráfica de las autocorrelaciones de los residuos del modelo, en donde se observa que las autocorrelaciones con respecto al rezago 12 y 24 toman valores estadísticamente diferente de cero. A diferencia de los modelos ARMA$(12,12)$ y ARMA$(12,4)$, el modelo ARMA$(3,12)$ satisface el supuesto de no autocorrelación y es el que presenta el segundo menor número AIC de los tres modelos ARMA estimados. De la información en la tabla 4.5 se puede concluir con un 95 % de confianza, que los errores del modelo

ARIMA$(3,0,0)[0,1,1]_{12}$ no rechazan la hipótesis de no autocorrelación ni la hipótesis de distribución normal: sin embargo, se rechaza la hipótesis de no heterocedasticidad.

Las figuras 4.13, 4.14, 4.15, 4.16, 4.17 y 4.18 muestran las gráficas de la serie, el diagrama de cuantiles, la función ACF y la función PACF de los errores de los modelos ARMA$(12,12)$, ARMA$(12,4)$ y ARMA$(3,12)$. Las figuras 4.19 y 4.20 muestran las correspondientes gráficas del modelo SARIMA, es decir, del modelo ARIMA$(3,0,0)[0,1,1]_{12}$.

Ahora se analiza la estacionariedad de los residuos de los modelos ARMA$(12,12)$, ARMA$(12,4)$, ARMA$(3,12)$ y del modelo SARIMA representado como ARIMA$(3,0,0)[0,1,1]_{12}$. Observando los resultados en las Tablas 4.7 a 4.10, se puede concluir que en todos los casos se rechaza la hipótesis nula de raíces unitarias con un 99 % de confianza, es decir, los residuos de los cuatro modelos estimados se comportan como variables estacionarias.

Por otro lado, debido a que los modelos ARMA$(12,12)$ y ARMA$(12,4)$ no cumplieron el supuesto de autocorrelación, y el modelo ARIMA$(3,0,0)[0,1,1]_{12}$ rechaza el supuesto de no heterocedasticidad, el modelo preferible es el ARMA$(3,12)$.

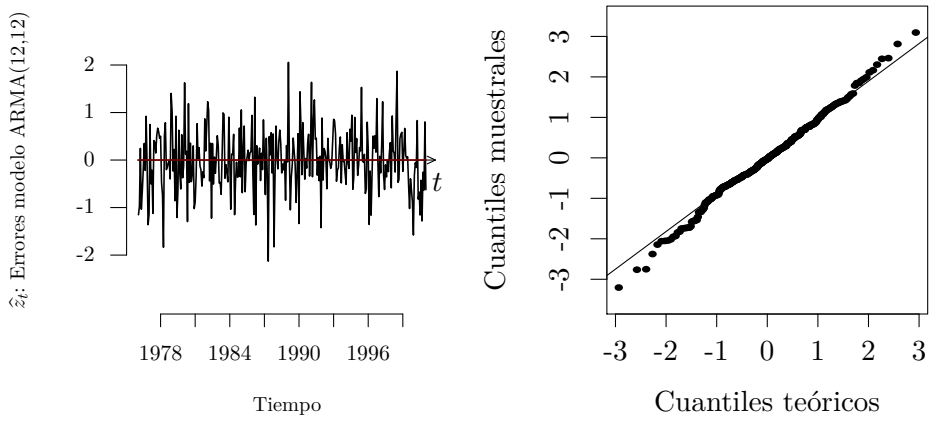

Figura 4.13: Serie de los errores del modelo ARMA(12, 12) y su diagrama de cuantiles.

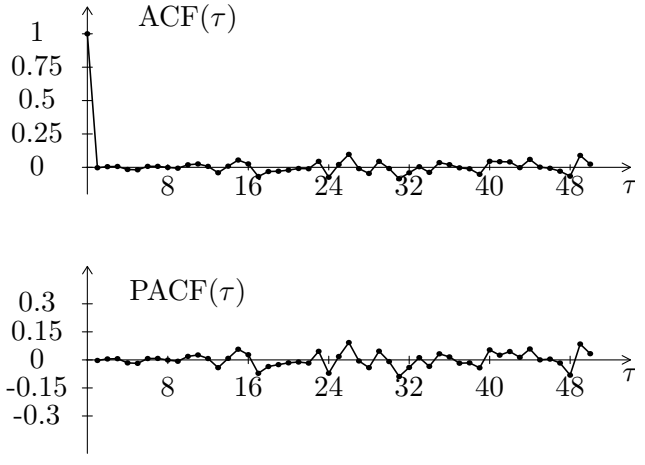

Figura 4.14: Funciones ACF y PACF de los errores del modelo ARMA(12, 12).

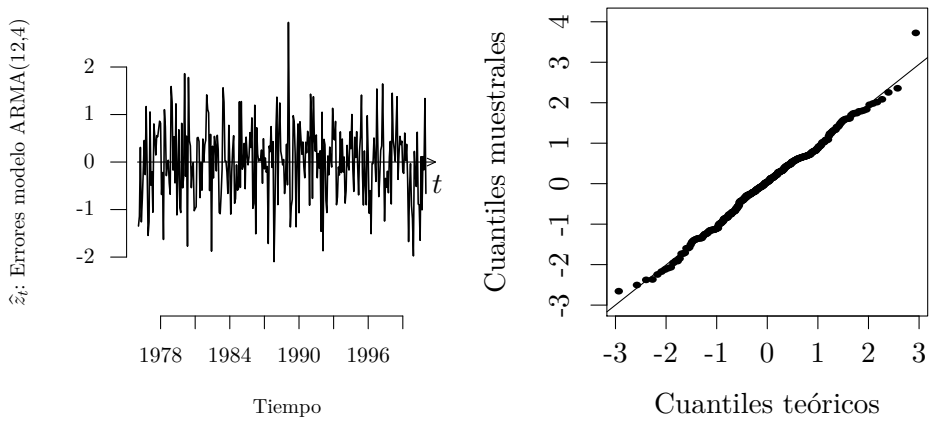

Figura 4.15: Serie de los errores del modelo ARMA$(12, 4)$ y su diagrama de cuantiles.

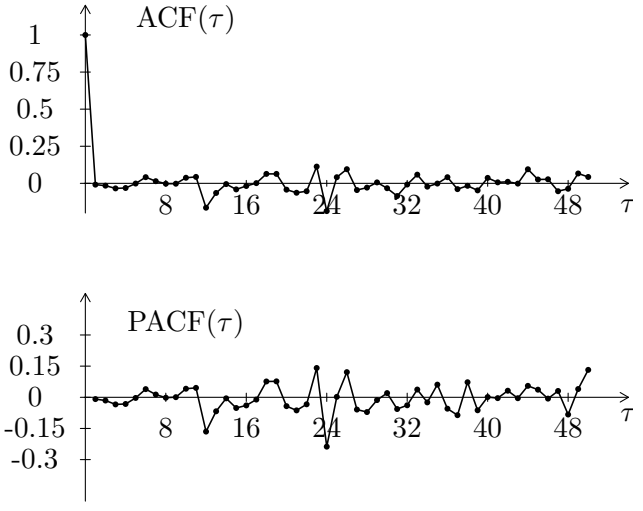

Figura 4.16: Funciones ACF y PACF de los errores del modelo ARMA$(12, 4)$.

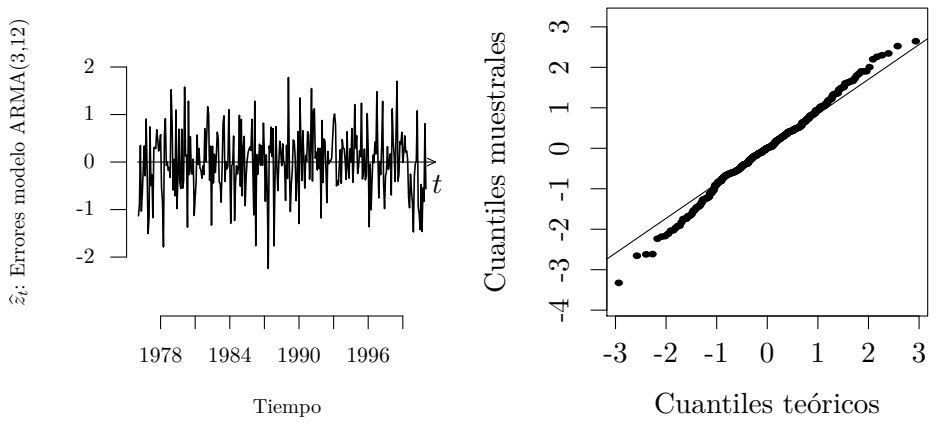

Figura 4.17: Serie de los errores del modelo ARMA$(3, 12)$ y su diagrama de cuantiles.

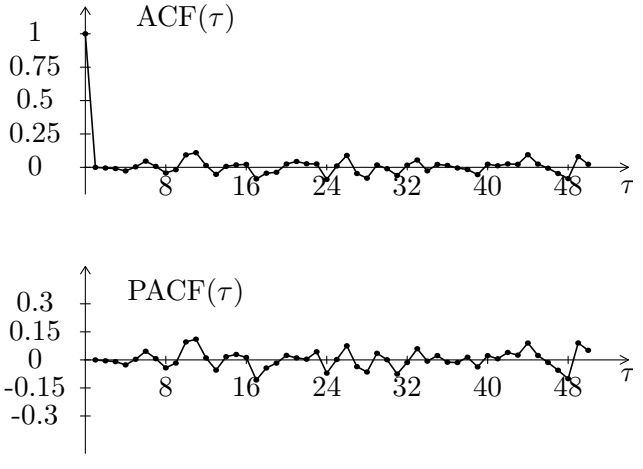

Figura 4.18: Funciones ACF y PACF de los errores del modelo ARMA$(3, 12)$.

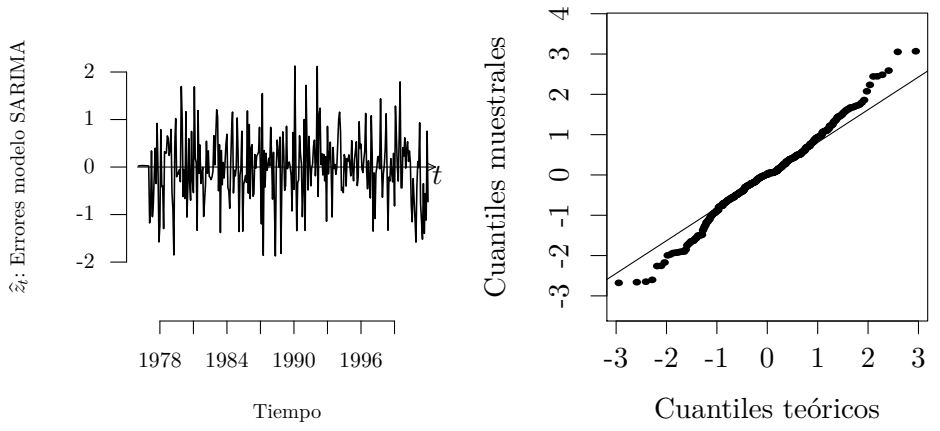

Figura 4.19: Serie de los errores del modelo SARIMA y su diagrama de cuantiles.

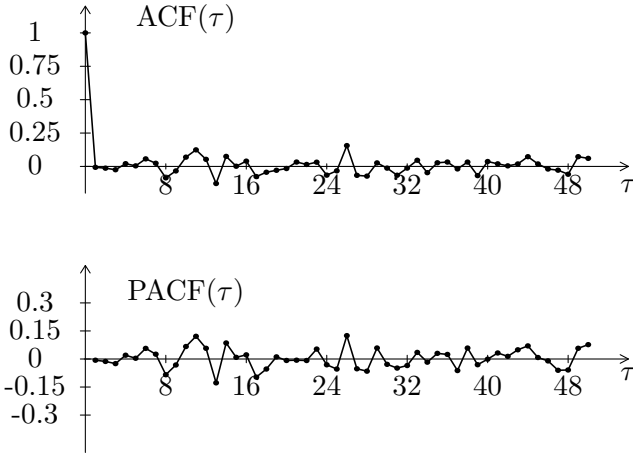

Figura 4.20: Funciones ACF y PACF de los errores del modelo SARIMA.

Pronóstico muestral

Después de haber realizado la estimación de los parámetros de nuestro modelo y de haber verificado que se cumplan los supuestos estadísticos de los errores o residuos, podemos hacer un pronóstico muestral, es decir, realizar una predicción de los valores $\widehat{x}_{n+1}, \widehat{x}_{n+2}, \ldots, \widehat{x}_{n+s}$ de la serie aprovechando las observaciones disponibles. Conforme el valor de s sea más grande, se terminará la información real de la serie y comenzaremos a usar las mismas predicciones para alimentar el modelo. Lo anterior provocará que entre más grande sea el valor de s, más incierta será la predicción $\widehat{x}_s$.

La idea básica es avanzar paso por paso. Por ejemplo, para la serie $x_1, x_2, \ldots, x_n$ supongamos el modelo ARMA(p, q) con término constante α. Ya conociendo los valores de los residuos, entonces

$$\widehat{x}_{n+1} = \widehat{\alpha} + \sum_{i=1}^{p} \widehat{\alpha}_i \, x_{n+1-i} + \sum_{j=1}^{q} \widehat{\beta}_j \, \widehat{z}_{n+1-j}, \tag{4.31}$$

y

$$\widehat{z}_{n+1} = 0. \tag{4.32}$$

Para valores de $t = 2, \ldots, s$,

$$\widehat{x}_{n+t} = \widehat{\alpha} + \sum_{i=1}^{p} \widehat{\alpha}_i {x^*}_{n+t-i} + \sum_{j=1}^{q} \widehat{\beta}_j \widehat{z}_{n+t-j}, \tag{4.33}$$

y

$$\widehat{z}_{n+t} = 0. \tag{4.34}$$

Observe que se combinan tanto observaciones conocidas como observaciones pronosticadas a partir del tiempo $n + 2$. Usamos la notación ${x^*}_{n+t-i}$ para indicar que se debe usar $\widehat{x}_{n+t-i}$ si $t - i > 0$ y se debe usar la observación x_{n+t-i} si $t - i \leqslant 0$. Por otro lado, conforme s crece, la parte autoregresiva del modelo se vuelve más importante que la parte de promedios móviles.

Retomando nuestro ejemplo, en nuestra serie ya diferenciada $y_1, y_2, \ldots, y_n$, donde $n = 300$, evaluaremos $s = 12$ pronósticos. Para calcular el error cuadrático medio de predicción del modelo ARMA$(3, 12)$, utilizaremos los valores conocidos de las observaciones $y_{301}, \ldots, y_{312}$. Como se menciona

al final de la sección anterior, estos datos son conocidos como datos de prueba, mientras que las primeras 300 observaciones sirven como datos de entrenamiento. Cabe señalar que, en la serie de temperaturas original, es necesario conocer los datos de un año a partir del último dato de entrenamiento para tener 12 observaciones conocidas de la serie diferenciada.

Así, el error cuadrático medio se calculará como

$$\text{ECM} = \frac{1}{12}\sum_{t=1}^{12}(y_{300+t} - \widehat{y}_{300+t})^2, \tag{4.35}$$

donde los valores de $\widehat{y}_{300+t}$ para $t = 1, 2, \ldots, 12$ se calculan con las fórmulas (4.31) a (4.34).

Mes	Observación real	ARMA$(12, 12)$	ARMA$(12, 4)$	ARMA$(3, 12)$
Enero	-1.2	0.388	0.303	0.381
Febrero	-0.4	0.523	0.504	0.468
Marzo	-1.4	-0.687	-0.573	-0.597
Abril	0.2	0.365	0.273	0.586
Mayo	-0.8	0.888	0.820	1.160
Junio	0.7	1.057	0.918	1.401
Julio	0.5	0.511	-0.034	0.889
Agosto	1.4	1.383	0.160	1.787
Septiembre	0.3	1.024	-0.204	1.343
Octubre	1.1	1.346	0.243	1.566
Noviembre	-0.8	0.279	-0.370	0.307
Diciembre	0.8	1.099	0.239	1.146

Tabla 4.11: Pronósticos para los modelos ARMA estimados.

Para los otros modelos ARMA es posible seguir el mismo procedimiento que para el cálculo de los pronósticos y del ECM. Para el modelo SARIMA, el cálculo se realizó únicamente mediante el software `R`. Los pronósticos muestrales estimados para cada modelo ARMA se muestran en la tabla 4.11, para el modelo SARIMA en la tabla 4.12, y el error cuadrático medio del pronóstico (y de los datos de entrenamiento considerando $t = 1, \ldots, 300$)

para cada modelo se muestra en la tabla 4.13.

Mes	Real	Pronóstico	Mes	Real	Pronóstico
Enero	22.0	23.298	Julio	28.0	28.502
Febrero	23.8	24.441	Agosto	27.9	28.423
Marzo	25.7	26.351	Septiembre	27.0	28.080
Abril	28.0	28.362	Octubre	26.1	26.812
Mayo	27.7	29.579	Noviembre	24.7	25.785
Junio	28.2	29.142	Diciembre	23.8	24.213

Tabla 4.12: Pronósticos para el modelo SARIMA: ARIMA$(3,0,0)[0,1,1]_{12}$.

La información del error cuadrático medio indica que con los datos de entrenamiento, el modelo con mejor desempeño es el que satisfase todos los supuestos sobre los errores, es decir, el ARMA$(3,12)$. El modelo con menor ECM con los datos de prueba es el ARMA$(12,12)$. Sin embargo, este modelo no cumple con el supuesto de no autocorrelación de los errores en el rezago 27. De los modelos analizados, el único que satisface todos los supuestos sobre los errores es el ARMA$(3,12)$. Aunque no es el de menor ECM con los datos de prueba, sí es el mejor en los datos de entrenamiento.

Modelo	Datos	ECM
ARMA$(12,12)$	Entrenamiento	0.439
	Prueba	0.727
ARMA$(12,4)$	Entrenamiento	0.620
	Prueba	0.812
ARMA$(3,12)$	Entrenamiento	0.460
	Prueba	0.944
ARIMA$(3,0,0)[0,1,1]_{12}$	Entrenamiento	0.482
	Prueba	0.884

Tabla 4.13: Error cuadrático medio de los datos de entrenamiento y de los pronósticos.

Una vez presentadas las especificaciones de los modelos que cumplen los supuestos estadísticos para los residuos, como en el caso del modelo

ARMA$(3, 12)$, es posible proceder a realizar pruebas de significancia de los parámetros estimados. Por ejemplo, como se puede observar en la tabla 4.4, en el modelo ARMA$(3, 12)$ la estimación de β_{10} relacionada con el término ma10 (Z_{t-10}) es de 0.004, lo cual parece no ser significativo en el modelo. Para verificar estadísticamente si un parámetro es significativo, se puede aplicar la prueba t, donde el estadístico T se calcula dividiendo el valor del parámetro estimado entre su error estándar. Si $|T| > t_{n-(p+q),1-\alpha/2}$ (cuantil $1-\alpha/2$ con $n-(p+q)$ grados de libertad de una distribución t de *Student*), se rechaza la hipótesis nula (el parámetro es igual a cero) con un nivel de significancia de α y se concluye que el parámetro es significativo. Se podría volver a realizar la estimación omitiendo ese término del modelo y proceder nuevamente a verificar los supuestos sobre los residuos. Es posible que siga siendo un buen modelo en el sentido de que no rechazar los supuestos probabilísticos de los residuos y con la ventaja de tener menos parámetros. Sin embargo, este análisis extra no se incluye porque está más allá de los objetivos de nuestro texto.

4.5. Ejercicios

Identificación

96. Mencione algún fenómeno de la naturaleza que considere que genera una serie de tiempo estacionaria. ¿Cree usted que se podría modelar con un modelo autorregresivo? Explique su respuesta.

97. Mencione algún fenómeno de la naturaleza que usted considere que genera una serie de tiempo no estacionaria. Explique su respuesta.

98. Mencione algún fenómeno económico que considere que genera una serie de tiempo estacionaria. ¿Cree usted que se podría modelar con un modelo autorregresivo? Explique su respuesta.

99. Mencione algún fenómeno económico que usted considere que genera una serie de tiempo no estacionaria. Explique su respuesta.

100. Intente explicar por qué para una serie de tiempo no estacionaria, el gráfico de la función ACF puede mostrar alturas que van decreciendo muy lentamente.

101. Intente explicar por qué para una serie de tiempo estacionaria cuyo gráfico de la función PACF muestra alturas muy cercanas a cero después del rezago k, se sugiere como posible elección un modelo AR(k).

102. ¿Es posible que una serie de tiempo con tendencia lineal sea estacionaria?

Estimación

103. Demostrar que la matriz de autocorrelaciones dada en (4.9) es una matriz invertible.

104. Dado el siguiente modelo ARMA$(2,1)$,

$$X_t = 0.5X_{t-1} - 0.3X_{t-2} + Z_t - 0.7Z_{t-1},$$

donde los residuos son normales: $Z_t \sim N(0,1)$,

a) Calcule la varianza de X_t.

b) Calcule la densidad de X_t condicionada a los valores de $X_{t-1} = x_{t-1}$ y $X_{t-2} = x_{t-2}$.

105. En el análisis de una serie de tiempo se ha identificado que un modelo ARMA$(1,1)$ con media cero es el más apropiado para describir un conjunto de datos. Los parámetros estimados son $\widehat{\alpha}_1 = 0.8$ y $\widehat{\beta}_1 = -0.5$. Obtenga la expresión de la función de verosimilitud suponiendo que $Z_t \sim N(0,\sigma^2)$. Obtenga el estimador máximo verosímil del parámetro σ^2.

106. Dada una serie de tiempo $\{Y_t\}$, se ha ajustado un modelo ARMA$(2,1)$ con los siguientes parámetros estimados: $\widehat{\alpha}_1 = 0.6$, $\widehat{\alpha}_2 = 0.2$ y $\widehat{\beta}_1 = -0.4$. ¿Cuál es la expresión de la función de verosimilitud suponiendo que $Z_t \sim N(0,\sigma^2)$? Obtenga el estimador máximo verosímil del parámetro σ^2.

107. Supongamos que tenemos la siguiente serie de tiempo $\{x_t\}$ con 24 observaciones:

$$4.2, 4.4, 4.1, 4.7, 4.9, 4.8, 4.6, 4.3,$$

$$4.2, 4.5, 4.6, 4.8, 4.9, 5.1, 5.3, 5.5,$$

$$5.7, 5.9, 6.1, 6.3, 4.2, 4.4, 4.1, 4.7.$$

a) Utilice el método de las ecuaciones de Yule-Walker para estimar los 3 parámetros $(\alpha_1, \alpha_2, \sigma^2)$ de un modelo AR(2) con media cero para modelar la serie dada por $y_t = x_t - \overline{x}_{24}$.

b) Ahora, utilice nuevamente el método de las ecuaciones de Yule-Walker para estimar los 4 parámetros $(\alpha_1, \alpha_2, \alpha_3, \sigma^2)$ de un modelo AR(3) con media cero para modelar la serie $\{y_t\}$.

Capítulo 5

Algunos resultados sobre pronósticos

Uno de los objetivos principales en el análisis de las series de tiempo es el de pronosticar o estimar el valor de una serie observada para algún tiempo futuro. Este tiempo futuro puede ser cualquier tiempo que se encuentre después de la última observación o dato registrado. Supondremos que se conocen los valores de las primeras n variables $X_1, \ldots, X_n$ y deseamos encontrar un pronóstico para X_{n+s}, para algún número entero $s \geqslant 0$, incluimos el valor $s = 0$ como un caso especial. Véase la figura 5.1. Junto a tal estimación se busca idealmente también contar con una cuantificación de la precisión de la estimación.

Este capítulo contiene una breve introducción al tema de los pronósticos en las series de tiempo. Supondremos que conocemos el modelo teórico $\{X_t : t \in \mathbb{Z}\}$ de la serie observada y que todos sus parámetros son conocidos, o bien han sido estimados.

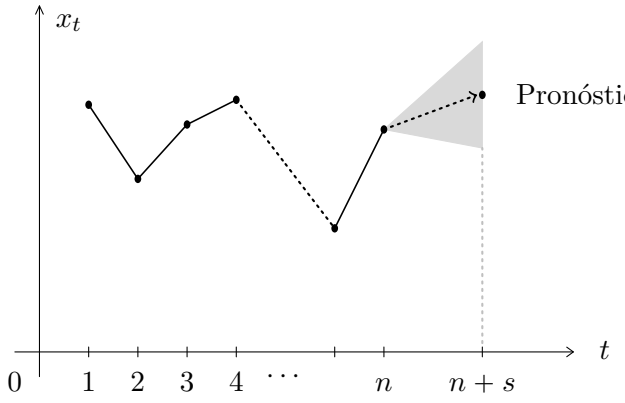

Figura 5.1: Pronóstico al tiempo futuro $n + s$.

5.1. La esperanza condicional como pronóstico

Supondremos que se tienen n observaciones o registros $X_1, \ldots, X_n$ de los primeros tiempos de una serie y que deseamos pronosticar el valor de X_{n+s}, con $s \geqslant 1$ un entero fijo. Denotaremos por $\hat{X}_n(s)$ a tal estimador y lo llamaremos también pronóstico. Una manera de definir a $\hat{X}_n(s)$ es a través de la siguiente esperanza condicional.

Definición 5.1 *Dada una serie de tiempo* $\{X_t : t \in \mathbb{Z}\}$ *con esperanza finita y dados números enteros* $n \geqslant 1$ *y* $s \geqslant 0$, *se define el estimador*

$$\hat{X}_n(s) := E(X_{n+s} \,|\, X_1, \ldots, X_n). \tag{5.1}$$

La variable aleatoria $\hat{X}_n(s)$ es la esperanza condicional de la observación futura X_{n+s} dadas las variables aleatorias $X_1, \ldots, X_n$. El número n indica el número de datos observados y $s \geqslant 0$ es un número entero tal que $n + s$ representa el tiempo futuro para el cual se desea hacer el pronóstico.

En términos generales, la esperanza condicional (5.1) se puede escribir

como $E(X \mid Y)$, en donde X y Y son dos variables aleatorias. En este caso, X es X_{n+s} y Y es el vector aleatorio $(X_1, \ldots, X_n)$. Uno de sus posibles valores es el número $E(X \mid Y = y)$, en donde y es cualquier valor de Y. A esta última cantidad también se le llama esperanza condicional y se estudia en los cursos básicos de probabilidad.

Parte de la dificultad para entender el término $E(X \mid Y)$ es que su definición se expresa en términos de tres propiedades que identifican de manera única a esta esperanza condicional. A continuación se expone esta definición, aunque se debe advertir al lector que la comprensión de este concepto no es sencilla, pues se encuentran involucrados varios conceptos técnicos: la esperanza condicional $E(X \mid Y)$ es una variable aleatoria que satisface las siguientes tres propiedades:

1. Es $\sigma(Y)$-medible.

2. Tiene esperanza finita.

3. Para cualquier evento $G \in \sigma(Y)$,

$$E(X \cdot 1_G) = E(\, E(X \mid Y) \cdot 1_G \,).$$

Se puede comprobar que las tres propiedades anteriores determinan de manera única (en el sentido casi seguro) a la variable $E(X \mid Y)$, es decir, si $\tilde{E}(X \mid Y)$ es otra variable aleatoria que satisface las propiedades anteriores, entonces $\tilde{E}(X \mid Y) = E(X \mid Y)$, casi seguramente. Para la primera propiedad, $\sigma(Y)$ es la mínima σ-álgebra respecto de la cual Y es variable aleatoria. En la tercera propiedad el símbolo 1_G indica la función indicadora del evento G.

En general, no es fácil encontrar $E(X \mid Y)$, o su distribución o características; su manejo se lleva a cabo a través de las propiedades que satisface, adicionales a las que aparecen en la definición. Algunas de estas propiedades se enuncian a continuación:

- $E(c \mid Y) = c$, en donde c es una constante.
- $E(cX \mid Y) = c\, E(X \mid Y)$, c constante.

- Si $X \geqslant 0$ entonces $E(X \,|\, Y) \geqslant 0$.
- $E(X_1 + X_2 \,|\, Y) = E(X_1 \,|\, Y) + E(X_2 \,|\, Y)$.
- Las variables aleatorias X y $E(X \,|\, Y)$ tienen la misma esperanza, es decir,
$$E(E(X \,|\, Y)) = E(X).$$
- Si X y Y son independientes entonces $E(X \,|\, Y) = E(X)$.
- Si Z es $\sigma(Y)$-medible, entonces $E(XZ \,|\, Y) = Z\, E(X \,|\, Y)$.

Las igualdades anteriores se cumplen en el sentido casi seguro, aunque esta condición no se indique. Se puede encontrar mayor información sobre $E(X \,|\, Y)$ en los libros de A. Gut [17] o en D. Williams [37]. La esperanza condicional $E(X \,|\, Y)$ también se denota por el símbolo $E(X \,|\, \sigma(Y))$, es decir, la esperanza condicional de X dada la σ-álgebra $\sigma(Y)$. Este es un caso particular del término $E(X \,|\, \mathscr{G})$, la esperanza condicional de X dada una σ-álgebra $\mathscr{G}$. Este concepto general se estudia en las referencias indicadas.

Error del estimador

Regresando al estimador (5.1), el error se define como la diferencia

$$e_n(s) := X_{n+s} - \hat{X}_n(s).$$

Este término es una variable aleatoria que no la conocemos, pues X_{n+s} es desconocida; sin embargo, se pueden calcular sus momentos. Por ejemplo, el error cuadrático medio de un estimador $g(X_1, \ldots, X_n)$ para X_{n+s}, basado en las observaciones $X_1, \ldots, X_n$, es el número no negativo

$$\text{ECM} = E(X_{n+s} - g(X_1, \ldots, X_n))^2.$$

La función g debe ser tal que $g(X_1, \ldots, X_n)$ es una variable aleatoria con segundo momento finito. Cuando tal función g es $E(X_{n+s} \,|\, X_1, \ldots, X_n)$, como aparece en (5.1), el error cuadrático medio es $\text{ECM} = E(e_n^2(s))$.

El siguiente resultado interesante establece que el estimador $\hat{X}_n(s)$, definido en (5.1), minimiza el error cuadrático medio cuando se le compara con

cualquier otro estimador basado en las variables aleatorias $X_1, \ldots, X_n$.

Proposición 5.1 *Sea $\{X_t : t \in \mathbb{Z}\}$ una serie de tiempo con segundo momento finito. Sea $\hat{X}_n(s)$ el estimador definido en (5.1). Para cualquier variable aleatoria $g(X_1, \ldots, X_n)$ se cumple que*

$$E(X_{n+s} - \hat{X}_n(s))^2 \leqslant E(X_{n+s} - g(X_1, \ldots, X_n))^2.$$

La demostración de este resultado no se presenta en este trabajo, sin embargo, se puede consultar en el texto de D. Williams [37].

En general no se puede hacer el cálculo del estimador (5.1) sin contar con la manera en la que están relacionadas las variables aleatorias $X_1, \ldots, X_n$ y X_{n+s}. Considerando varios modelos particulares para $\{X_t : t \in \mathbb{Z}\}$, a continuación se presentan algunas fórmulas para $\hat{X}_n(s)$, los cuales son casos en los que es relativamente fácil encontrar la esperanza condicional (5.1).

Pronóstico en una serie con tendencia

Consideremos el modelo de serie $\{X_t : t \in \mathbb{Z}\}$ dado por

$$X_t = m_t + Z_t, \quad t \in \mathbb{Z},$$

en donde m_t es una función determinista del tiempo y $\{Z_t : t \in \mathbb{Z}\}$ es un proceso de ruido de media cero y varianza $\sigma^2 > 0$. La función m_t representa el componente de tendencia, puede ser lineal, cuadrática o presentar cualquier otro tipo de comportamiento. Para cualquier entero $s \geqslant 0$ se puede calcular el estimador (5.1) de la siguiente manera:

$$\begin{aligned} \hat{X}_n(s) &= E(X_{n+s} \,|\, X_1, \ldots, X_n) \\ &= E(m_{n+s} + Z_{n+s} \,|\, X_1, \ldots, X_n) \\ &= m_{n+s} + E(Z_{n+s} \,|\, X_1, \ldots, X_n) \\ &= m_{n+s} + E(Z_{n+s}) \\ &= m_{n+s}. \end{aligned} \tag{5.2}$$

De esta manera, la estimación para X_{n+s} es simplemente la tendencia m_t evaluada en el tiempo futuro $t = n + s$. A partir de la serie observada $x_1, \ldots, x_n$, se puede estimar la tendencia m_t y el resultado anterior corresponde a la hipótesis de que la tendencia permanece válida en el tiempo futuro señalado. Por supuesto, tal hipótesis es muy aventurada y difícilmente contaremos con suficientes elementos para justificarla. Por otro lado, el error del pronóstico (5.2) es

$$\begin{aligned} e_n(s) &= X_{n+s} - \hat{X}_n(s) \\ &= m_{n+s} + Z_{n+s} - m_{n+s} \\ &= Z_{n+s}, \end{aligned}$$

de modo que, $E(e_n(s)) = E(Z_{n+s}) = 0$. La varianza del error es la siguiente cantidad, la cual no depende del parámetro s,

$$\text{Var}(e_n(s)) = \text{Var}(Z_{n+s}) = \sigma^2.$$

El error cuadrático medio del estimador (5.2) es

$$\begin{aligned} \text{ECM} &= E(X_{n+s} - \hat{X}_n(s))^2 \\ &= E(m_{n+s} + Z_{n+s} - m_{n+s})^2 \\ &= E(Z_{n+s})^2 \\ &= \sigma^2. \end{aligned}$$

Pronóstico en el modelo AR(1)

Sea $\{X_t : t \in \mathbb{Z}\}$ una serie de tiempo que sigue el modelo AR(1), es decir,

$$X_t = \alpha X_{t-1} + Z_t, \quad t \in \mathbb{Z}. \tag{5.3}$$

No utilizaremos la hipótesis $-1 < \alpha < 1$, la cual garantiza que el proceso es estacionario. Supongamos primero que deseamos pronosticar el valor de la serie un tiempo después del último momento observado, es decir, $s = 1$. Substituyendo $t = n + 1$ en el modelo (5.3), tenemos que

$$X_{n+1} = \alpha X_n + Z_{n+1}.$$

Por lo tanto,

$$\begin{aligned}\hat{X}_n(1) &= E(X_{n+1} \,|\, X_1, \ldots, X_n) \\ &= E(\alpha X_n + Z_{n+1} \,|\, X_1, \ldots, X_n) \\ &= \alpha E(X_n \,|\, X_1, \ldots, X_n) + E(Z_{n+1} \,|\, X_1, \ldots, X_n) \\ &= \alpha X_n + E(Z_{n+1}) \\ &= \alpha X_n.\end{aligned}$$

Esto significa que el pronóstico es un factor de la última observación X_n. Un cálculo similar muestra que el estimador para el caso $s = 2$ es

$$\hat{X}_n(2) = \alpha \hat{X}_n(1) = \alpha^2 X_n.$$

De manera más general, el pronóstico para el caso $s \geqslant 1$ se muestra en el siguiente recuadro. Los detalles de la demostración se dejan como ejercicio.

Proposición 5.2 *Sea* $\{X_t : t \in \mathbb{Z}\}$ *una serie de tiempo que sigue un modelo* AR(1) *especificado en* (5.3). *Sea* $s \geqslant 0$ *un número entero. El pronóstico para* X_{n+s} *basado en* $X_1, \ldots, X_n$ *y definido en* (5.1) *es*

$$\hat{X}_n(s) = \alpha^s X_n. \tag{5.4}$$

De esta manera, el pronóstico sólo hace uso de la última variable observada X_n. Se puede comprobar que el error del pronóstico (5.4) es

$$\begin{aligned}e_n(s) &= X_{n+s} - \hat{X}_n(s) \\ &= Z_{n+s} + \alpha\, Z_{n+s-1} + \alpha^2\, Z_{n+s-2} + \cdots + \alpha^{s-1}\, Z_{n+1}.\end{aligned}$$

Es claro que $E(e_n(s)) = 0$. A partir de lo anterior se puede demostrar que

$$\begin{aligned}\text{Var}(e_n(s)) &= E(Z_{n+s} + \alpha\, Z_{n+s-1} + \alpha^2\, Z_{n+s-2} + \cdots + \alpha^{s-1}\, Z_{n+1})^2 \\ &= \sigma^2\, \frac{1-\alpha^{2s}}{1-\alpha^2}, \quad \alpha^2 \neq 1.\end{aligned}$$

El error cuadrático medio del estimador (5.4) es, para $\alpha^2 \neq 1$,

$$\text{ECM} = E(X_{n+s} - \hat{X}_n(s))^2 = E(e_n^2(s)) = \text{Var}(e_n(s)) = \sigma^2\, \frac{1-\alpha^{2s}}{1-\alpha^2}.$$

En la sección de ejercicios se pide comprobar las fórmulas anteriores en el caso de este mismo modelo AR(1), pero de media μ. Se puede comprobar que únicamente en la expresión para $\hat{X}_n(s)$ aparece el parámetro μ, en el resto de las fórmulas tal parámetro no aparece.

Pronóstico en el modelo MA(1)

Consideremos ahora una serie de tiempo $\{X_t : t \in \mathbb{Z}\}$ que sigue el modelo MA(1) dado por

$$X_t = Z_t + \beta Z_{t-1}, \quad t \in \mathbb{Z}. \tag{5.5}$$

Deseamos estimar el valor de X_{n+s} a partir de $X_1, \ldots, X_n$ usando el pronóstico dado por la esperanza condicional (5.1). Primero consideraremos el caso $s = 1$ y después el caso $s \geqslant 2$.

- *Caso* $s = 1$. Substituyendo t por $n+1$ en el modelo (5.5) tenemos que

$$X_{n+1} = Z_{n+1} + \beta\, Z_n.$$

Tomando esperanza condicional respecto de $X_1, \ldots, X_n$, obtenemos que

$$\hat{X}_n(1) = E(Z_{n+1} \,|\, X_1, \ldots, X_n) + \beta\, E(Z_n \,|\, X_1, \ldots, X_n).$$

Como Z_{n+1} es independiente de las variables $X_1, \ldots, X_n$, la primera esperanza se reduce a $E(Z_{n+1}) = 0$. Para la segunda esperanza se explica una manera de obtener una aproximación suponiendo que el proceso es invertible. Bajo tal hipótesis, por la ecuación (3.6) que aparece en la página 85,

$$\begin{aligned} Z_n &= \sum_{k=0}^{\infty} (-\beta)^k X_{n-k} \\ &\approx \sum_{k=0}^{n-1} (-\beta)^k X_{n-k}. \end{aligned}$$

Observe que se ha recortado la suma infinita a sólo los primeros n sumandos y por ello sólo se tiene un valor aproximado. A partir de esta

expresión aproximada tenemos que Z_n es una función de $X_1, \ldots, X_n$ y la esperanza condicional respecto de estas variables es la misma Z_n, es decir,

$$E(Z_n \,|\, X_1, \ldots, X_n) = \sum_{k=0}^{n-1} (-\beta)^k X_{n-k} \approx Z_n.$$

Por lo tanto, el estimador que se propone es

$$\hat{X}_n(1) := \beta\, Z_n.$$

- *Caso* $s \geqslant 2$. Comprobaremos que es suficiente analizar el caso $s = 2$. Tomando $t = n + 2$ en el modelo (5.5), tenemos que

$$X_{n+2} = Z_{n+2} + \beta Z_{n+1}.$$

Como $s \geqslant 2$, las variables Z_{n+s} y Z_{n+1} son independientes de las observaciones $X_1, \ldots, X_n$, de modo que al tomar esperanza condicional respecto de $X_1, \ldots, X_n$, obtenemos la ecuación

$$\begin{aligned} \hat{X}_n(s) &= E(Z_{n+s} \,|\, X_1, \ldots, X_n) + \beta E(Z_{n+s-1} \,|\, X_1, \ldots, X_n) \\ &= E(Z_{n+s}) + \beta E(Z_{n+s-1}) \\ &= 0. \end{aligned}$$

Incorporando los dos casos anteriores, tenemos el siguiente resultado.

Proposición 5.3 *Sea* $\{X_t : t \in \mathbb{Z}\}$ *una serie que sigue un modelo* MA(1) *especificado en* (5.5). *Sea* $s \geqslant 1$ *un entero. El pronóstico para* X_{n+s} *basado en* $X_1, \ldots, X_n$ *y definido en* (5.1) *es*

$$\hat{X}_n(s) = \begin{cases} \beta Z_n & si \;\; s = 1, \\ 0 & si \;\; s \geqslant 2. \end{cases} \tag{5.6}$$

El error del estimador es

$$\begin{aligned} e_n(s) &= X_{n+s} - \hat{X}_n(s) \\ &= Z_{n+s} + \beta Z_{n+s-1} - \hat{X}_n(s) \\ &= \begin{cases} Z_{n+1} & \text{si} \;\; s = 1, \\ Z_{n+s} + \beta Z_{n+s-1} & \text{si} \;\; s \geqslant 2. \end{cases} \end{aligned}$$

Se puede comprobar que $E(e_n(s)) = 0$ y que el error cuadrático medio del estimador (5.6) es

$$\begin{aligned} \text{ECM} &= E(X_{n+s} - \hat{X}_n(s))^2 \\ &= E(e_n^2(s)) \\ &= \text{Var}(e_n(s)) \\ &= \begin{cases} \sigma^2 & \text{si } s = 1, \\ (1+\beta^2)\sigma^2 & \text{si } s \geqslant 2. \end{cases} \end{aligned}$$

En la sección de ejercicios se pide comprobar las fórmulas anteriores en el caso de este mismo modelo MA(1) pero de media μ. Es posible comprobar que únicamente en la expresión para $\hat{X}_n(s)$ aparece el parámetro μ, en el resto de las fórmulas no aparece.

§

Como es evidente en los ejemplos presentados, encontrar el pronóstico mediante la esperanza condicional (5.1) depende en gran medida de la facilidad con la que tal esperanza se pueda calcular. A continuación se estudia el caso restringido en el que al pronóstico se le pide que sea una función lineal de las observaciones.

5.2. Pronóstico lineal para series estacionarias

En esta sección supondremos que la serie $\{X_t : t \in \mathbb{Z}\}$ es estacionaria, consideraremos que se conocen los valores iniciales $X_1, \ldots, X_n$ y que deseamos encontrar el mejor pronóstico lineal para X_{n+s}, con $s \geqslant 1$ un número entero. A tal estimador lo denotaremos por el símbolo $\ell\hat{X}_n(s)$, y lo llamaremos pronóstico lineal. La letra ℓ indica que el estimador debe ser una función lineal de los datos $X_1, \ldots, X_n$, es decir, para ciertas constantes $a_0, a_1, \ldots, a_n$, la expresión general del estimador debe ser

$$\ell\hat{X}_n(s) := a_0 + a_1X_1 + \cdots + a_nX_n.$$

Dentro de todos estos estimadores lineales buscamos aquel que minimice el error cuadrático medio dado por

$$\text{ECM} = E(X_{n+s} - \ell\hat{X}_n(s))^2.$$

Al estimador lineal con error cuadrático mínimo se le llama el mejor estimador lineal por mínimos cuadrados. De esta manera, si $\hat{X}_n(s)$ es el estimador $E(X_{n+s} \,|\, X_1, \ldots, X_n)$ definido en la sección anterior y $\hat{\ell}$ es cualquier estimador lineal para X_{n+s} basado en $X_1, \ldots, X_n$, entonces se cumplen las desigualdades

$$E(X_{n+s} - \hat{X}_n(s))^2 \leqslant E(X_{n+s} - \ell\hat{X}_n(s))^2 \leqslant E(X_{n+s} - \hat{\ell})^2.$$

Debe observarse que si el estimador general (esperanza condicional) $\hat{X}_n(s)$ resulta ser una función lineal de las variables predictoras $X_1, \ldots, X_n$, entonces coincide con el estimador lineal $\ell\hat{X}_n(s)$ y el error cuadrático medio es el mismo para ambos estimadores.

El resultado que aparece a continuación garantiza la existencia y unicidad del mejor estimador lineal $\ell\hat{X}_n(s)$. Por brevedad en la escritura, a la función de autocovarianza $\gamma(\tau) = \text{Cov}(X_t, X_{t+\tau})$, para una serie estacionaria, la escribiremos como γ_τ o bien $\gamma(\tau)$.

Proposición 5.4 *Sea $\{X_t : t \in \mathbb{Z}\}$ una serie de tiempo estacionaria de media μ y función de autocovarianza γ_τ, ambas conocidas. Sea $s \geqslant 1$ un número entero. El estimador lineal por mínimos cuadrados para el valor X_{n+s}, basado en $X_1, \ldots, X_n$, es*

$$\ell\hat{X}_n(s) = a_0 + a_1X_1 + \cdots + a_nX_n, \tag{5.7}$$

en donde el primer coeficiente a_0 está dado por

$$a_0 = (1 - \sum_{i=1}^{n} a_i)\,\mu, \tag{5.8}$$

y los otros coeficientes $a_1, \ldots, a_n$ son la solución al sistema de ecuaciones

$$\begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{n-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \gamma_{n-3} & \cdots & \gamma_0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \gamma_{n+s-1} \\ \gamma_{n+s-2} \\ \gamma_{n+s-3} \\ \vdots \\ \gamma_s \end{pmatrix}. \tag{5.9}$$

Demostración. El error cuadrático medio del estimador (5.7) es la siguiente función de varias variables,

$$\text{ECM}(a_0, \ldots, a_n) = E(X_{n+s} - a_0 - a_1 X_1 - \cdots - a_n X_n)^2.$$

Esta es una función cuadrática de $n+1$ variables y es no negativa, es decir, está acotada inferiormente por el valor 0. Por lo tanto, su valor mínimo se alcanza por lo menos para algún valor del vector $(a_0, a_1, \ldots, a_n)$, el cual satisface las ecuaciones que se obtienen al igualar a cero la derivada de la función $\text{ECM}(a_0, \ldots, a_n)$.

Derivando respecto de $a_0, a_1, \ldots, a_n$, e igualando a cero, se encuentran las siguientes ecuaciones:

$$\frac{\partial}{\partial a_0}\text{ECM}(a_0, \ldots, a_n) = -2\, E(X_{n+s} - a_0 - a_1 X_1 - \cdots - a_n X_n) = 0,$$

y para $j = 1, \ldots, n$,

$$\frac{\partial}{\partial a_j}\text{ECM}(a_0, \ldots, a_n) = -2\, E[(X_{n+s} - a_0 - a_1 X_1 - \cdots - a_n X_n)\, X_j] = 0.$$

Simplificando la primera ecuación se obtiene $\mu - a_0 - a_1\,\mu - \cdots - a_n\,\mu = 0$, lo cual se puede escribir como aparece en (5.8). Usando la propiedad de estacionariedad, la segunda ecuación se reduce a lo siguiente: para $j = 1, \ldots, n$,

$$\begin{aligned}
0 &= E(X_{n+s}\, X_j) - a_0\, E(X_j) - a_1\, E(X_1\, X_j) - \cdots - a_n\, E(X_n\, X_j) \\
&= \gamma_{n+s-j} + \mu^2 - a_0\,\mu - a_1(\gamma_{1-j} + \mu^2) - \cdots - a_n(\gamma_{n-j} + \mu^2) \\
&= \gamma_{n+s-j} - a_0\,\mu - a_1\,\gamma_{1-j} - \cdots - a_n\,\gamma_{n-j} + \underbrace{(1 - \sum_{i=1}^{n} a_i)\,\mu^2}_{a_0\,\mu} \\
&= \gamma_{n+s-j} - a_1\,\gamma_{1-j} - \cdots - a_n\,\gamma_{n-j}. \qquad (5.10)
\end{aligned}$$

Esto produce el sistema de n ecuaciones lineales para las incógnitas $a_1, \ldots, a_n$ que aparece en la igualdad (5.9) del enunciado. ∎

Más adelante encontraremos una expresión para el error cuadrático medio y otras propiedades del pronóstico lineal $\ell\hat{X}_n(s)$. Como un primer ejemplo, a continuación se encuentra este estimador en el modelo AR(1) cuando $s = 1$.

Pronóstico lineal en el modelo AR(1)

Consideremos el proceso AR(1) dado por

$$X_t = \alpha X_{t-1} + Z_t, \quad t \in \mathbb{Z}, \tag{5.11}$$

en donde $\{Z_t : t \in \mathbb{Z}\}$ es un proceso de ruido con media 0 y varianza σ^2. Sabemos que la serie $\{X_t : t \in \mathbb{Z}\}$ es causal y, por lo tanto, que es estacionaria para valores de α tales que $-1 < \alpha < 1$. En esta sección se encuentra el estimador lineal por mínimos cuadrados para el valor X_{n+1} (es decir, sólo una unidad de tiempo adelante en el futuro), tomando como datos los valores de las variables $X_1, \ldots, X_n$.

La esperanza del proceso (5.11) es $E(X_t) = \mu = 0$ y la función de autocovarianza es

$$\gamma(\tau) = \text{Cov}\,(X_t, X_{t+\tau}) = \frac{\alpha^{|\tau|}}{1-\alpha^2}\,\sigma^2, \quad \text{para } \tau \in \mathbb{Z}.$$

El estimador lineal por mínimos cuadrados para X_{n+1} está dado por la expresión (5.7), en donde los coeficientes $a_0, a_1, \ldots, a_n$ están determinados por las ecuaciones (5.8) y (5.9). Como $\mu = 0$, de (5.8) se obtiene que $a_0 = 0$. Por otro lado, el sistema de ecuaciones (5.8) se reduce a

$$\begin{pmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^{n-1} \\ \alpha & 1 & \alpha & \cdots & \alpha^{n-2} \\ \alpha^2 & \alpha & 1 & \cdots & \alpha^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha^{n-1} & \alpha^{n-2} & \alpha^{n-3} & \cdots & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \alpha^n \\ \alpha^{n-1} \\ \alpha^{n-2} \\ \vdots \\ \alpha \end{pmatrix}.$$

Se puede comprobar que la solución a este sistema de ecuaciones es

$$(a_1, \ldots, a_{n-1}, a_n) = (0, \ldots, 0, \alpha).$$

Por lo tanto, el estimador buscado adquiere la forma simple

$$\ell\hat{X}_n(1) = \alpha X_n.$$

Como era de esperarse, esta expresión coincide con la fórmula encontrada, $\hat{X}_n(s) = \alpha^s X_n$, que aparece en (5.4) cuando $s = 1$. El error cuadrático

medio, nuevamente, coincide con la expresión encontrada antes, cuando $s = 1$, el cual resulta ser igual a σ^2. El cálculo es sencillo y se puede hacer directamente de la siguiente manera

$$E(X_{n+1} - \ell\hat{X}_n(1))^2 = E(\alpha X_n + Z_{n+1} - \alpha X_n)^2 = E(Z_{n+1}^2) = \sigma^2.$$

Propiedades del pronóstico lineal para series estacionarias

En esta sección se demuestran varias propiedades del estimador lineal por mínimos cuadrados $\ell\hat{X}_n(s)$ que aparece en el enunciado de la proposición 5.4.

Proposición 5.5 *Bajo la notación e hipótesis de la proposición 5.4, el estimador lineal por mínimos cuadrados $\ell\hat{X}_n(s)$ satisface las siguientes propiedades:*

1. $\ell\hat{X}_n(s) = \mu + \sum_{i=1}^{n}(X_i - \mu)\, a_i.$
2. $E(\ell\hat{X}_n(s)) = E(X_{n+s}) = \mu.$
3. *El valor mínimo del error cuadrático medio es*

$$E(X_{n+s}-\ell\hat{X}_n(s))^2 = \gamma(0)-2\sum_{i=1}^{n} a_i\,\gamma(n+s-i)+\sum_{i=1}^{n}\sum_{j=1}^{n} a_i\,a_j\,\gamma(j-i).$$

4. $E[\,(X_{n+s} - \ell\hat{X}_n(s)) \cdot X_j\,] = 0, \quad$ *para* $\quad j = 1, \ldots, n.$
5. *El estimador $\ell\hat{X}_n(s)$ es único en el sentido casi seguro.*

Demostración.

1. Se substituye la fórmula (5.8) para a_0 en la expresión lineal del pro-

nóstico (5.7) y se obtiene que

$$\begin{aligned} \ell\hat{X}_n(s) &= a_0 + a_1 X_1 + \cdots + a_n X_n \\ &= (1 - \sum_{i=1}^{n} a_i)\,\mu + \sum_{i=1}^{n} a_i X_i \\ &= \mu + \sum_{i=1}^{n} (X_i - \mu)\, a_i. \end{aligned}$$

2. Por el inciso anterior,

$$E(\ell\hat{X}_n(s)) = E(\,\mu + \sum_{i=1}^{n} (X_i - \mu)\, a_i\,) = \mu.$$

3. El error cuadrático medio del pronóstico es

$$\begin{aligned} & E(X_{n+s} - \ell\hat{X}_n(s))^2 \\ &= E(X_{n+s}^2) - 2\,E(X_{n+s} \cdot \ell\hat{X}_n(s)) + E(\ell\hat{X}_n(s))^2 \\ &= \mathrm{Var}(X_{n+s}) + \mu^2 - 2\,[a_0\,\mu + \sum_{i=1}^{n} a_i\,E(X_{n+s} \cdot X_i)] \\ &\quad + E(\mu + \sum_{i=1}^{n} (X_i - \mu)\, a_i)^2 \\ &= \gamma(0) + \mu^2 - 2[(1 - \sum_{i=1}^{n} a_i)\,\mu^2 + \sum_{i=1}^{n} a_i\,[\gamma(n+s-1) + \mu^2]\,] \\ &\quad + \mu^2 + 0 + E(\sum_{i=1}^{n} (X_i - \mu)\, a_i)^2 \\ &= \gamma(0) - 2\,\sum_{i=1}^{n} a_i\,\gamma(n+s-i) + \sum_{i=1}^{n}\sum_{j=1}^{n} a_i\,a_j\,\underbrace{E(X_i - \mu)(X_j - \mu)}_{\gamma(j-i)}. \end{aligned}$$

4. El resultado se obtiene al desarrollar el producto y distribuir la esperanza. En efecto,

$$\begin{aligned}
&E[\,(X_{n+s} - \ell\hat{X}_n(s)) \cdot X_j\,] \\
&\quad = E[\,X_{n+s} \cdot X_j - (a_0 \cdot X_j + \sum_{i=1}^{n} a_i \cdot X_i \cdot X_j)\,] \\
&\quad = \gamma(n+s-j) + \mu^2 - a_0 \cdot \mu - \sum_{i=1}^{n} a_i\,[\gamma(j-i) + \mu^2] \\
&\quad = \gamma(n+s-j) - \sum_{i=1}^{n} a_i\,\gamma(i-j) \\
&\quad = 0.
\end{aligned}$$

La última igualdad se obtiene utilizando la identidad (5.10) que aparece en la página 228.

5. Sean $a_0, a_1, \ldots, a_n$ y $a'_0, a'_1, \ldots, a'_n$ dos soluciones a las ecuaciones

$$\frac{\partial}{\partial\, a_j}\mathrm{ECM}(a_0, \ldots, a_n) = 0, \quad \text{para} \;\; j = 0, 1, \ldots, n.$$

Se denotará por $\ell\hat{X}_n(s)$ y $\ell'\hat{X}_n(s)$ a los pronósticos lineales construidos a partir de cada conjunto de constantes, respectivamente. Definamos la variable aleatoria diferencia

$$\begin{aligned}
Z &= \ell\hat{X}_n(s) - \ell'\hat{X}_n(s) \\
&= (a_0 + \sum_{j=1}^{n} a_j\, X_j) - (a'_0 + \sum_{i=1}^{n} a'_j\, X_j) \\
&= (a_0 - a'_0) + \sum_{j=1}^{n} (a_j - a'_j)\, X_j.
\end{aligned}$$

Entonces tenemos que $E(Z) = 0$ y el segundo momento es

$$\begin{aligned} E(Z^2) &= E(Z \cdot Z) \\ &= E(Z \cdot [(a_0 - a_0') + \sum_{j=1}^{n} (a_j - a_j')\, X_j]) \\ &= \sum_{j=1}^{n} (a_j - a_j')\, E(Z \cdot X_j) \\ &= \sum_{j=1}^{n} (a_j - a_j')\, E(\ell \hat{X}_n(s) \cdot X_j - \ell' \hat{X}_n(s) \cdot X_j) \\ &= \sum_{j=1}^{n} (a_j - a_j')\, (E(X_{n+s}) - E(X_{n+s})) \\ &= 0. \end{aligned}$$

La penúltima identidad es consecuencia del inciso (4), por lo tanto, Z es una variable aleatoria con segundo momento nulo. En consecuencia, Z debe ser constante. Este valor constante es su media, que es cero, es decir, $\ell \hat{X}_n(s) = \ell' \hat{X}_n(s)$ c.s.

■

5.3. Pronóstico lineal sin estacionariedad

Primero se estudia este problema en un contexto general y después se aplican los resultados al caso de series de tiempo no necesariamente estacionarias. Sean $W_1, \ldots, W_n, Y$ variables aleatorias con segundo momento finito. El problema que se plantea aquí es el de pronosticar el valor de Y a través de una combinación lineal de las variables $W_1, \ldots, W_n$. A diferencia de lo desarrollado en la sección anterior, lo relevante ahora es que no supondremos la hipótesis de estacionariedad sobre la sucesión de variables aleatorias predictoras.

Por simplicidad en la escritura, supondremos que tenemos sólo dos variables aleatorias predictoras: W_1 y W_2. El procedimiento es similar al del caso de n variables. Procediendo así, tenemos la ventaja de que las expresiones

no aparecen demasiado largas. Por otro lado, para que el pronóstico sea posible, es necesario suponer que la variable a pronosticar Y y el vector predictor (W_1, W_2) están relacionados de alguna manera. Supondremos que esta relación se conoce a través de las siguientes cantidades:

- $E(Y)$.
- $\text{Var}(Y) = \text{Cov}\,(Y, Y)$.
- $\text{Cov}\,(Y, W_1)$.
- $\text{Cov}\,(Y, W_2)$.
- $(E(W_1), E(W_2))$.
- $\text{Var}(W_1, W_2) = \begin{pmatrix} \text{Cov}(W_1, W_1) & \text{Cov}(W_1, W_2) \\ \text{Cov}\,(W_2, W_1) & \text{Cov}\,(W_2, W_2) \end{pmatrix}$.

Repetiremos el procedimiento expuesto antes para encontrar el estimador buscado, el cual denotaremos ahora de manera explícita por el símbolo $\ell\hat{Y}(W_1, W_2)$. Recordemos que el estimador debe ser una función lineal de W_1 y W_2, es decir, para ciertas constantes a_0, a_1, a_2,

$$\ell\hat{Y}(W_1, W_2) = a_0 + a_1\, W_1 + a_2\, W_2,$$

y además debe minimizar el error cuadrático medio, el cual se define como $\text{ECM}(a_0, a_1, a_2) = E(Y - a_0 - a_1 W_1 - a_2 W_2)^2$. El resultado es el que se expone a continuación.

Proposición 5.6 *Sean Y, W_1, W_2 tres variables aleatorias con segundo momento finito. El estimador lineal por mínimos cuadrados para Y basado en W_1 y W_2 es*

$$\hat{\ell Y}(W_1, W_2) = a_0 + a_1 W_1 + a_2 W_2, \tag{5.12}$$

en donde el coeficiente a_0 está dado por

$$a_0 = E(Y) - a_1 E(W_1) - a_2 E(W_2), \tag{5.13}$$

y los coeficientes a_1 y a_2 son la solución al sistema de ecuaciones

$$\begin{pmatrix} \text{Cov}(W_1, W_1) & \text{Cov}(W_1, W_2) \\ \text{Cov}\,(W_2, W_1) & \text{Cov}\,(W_2, W_2) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \text{Cov}(Y, W_1) \\ \text{Cov}\,(Y, W_2) \end{pmatrix}. \tag{5.14}$$

Demostración. Derivando el error cuadrático medio respecto de a_0, a_1 y a_2, e igualando a cero, se obtienen las ecuaciones lineales

$$\begin{aligned} \frac{\partial}{\partial a_0}\text{ECM}(a_0, a_1, a_2) &= -2\,E(Y - a_0 - a_1 W_1 - a_2 W_2) = 0, \\ \frac{\partial}{\partial a_j}\text{ECM}(a_0, a_1, a_2) &= -2\,E[(Y - a_0 - a_1 W_1 - a_2 W_2)\,W_j] = 0, \quad j = 1, 2. \end{aligned}$$

Distribuyendo la esperanza en la primera ecuación se obtiene (5.13). De la segunda ecuación se obtiene

$$E(Y\,W_j) - a_0\,E(W_j) - a_1\,E(W_1\,W_j) - a_2\,E(W_2)\,W_j) = 0, \quad j = 1, 2.$$

Substituyendo el valor de a_0 y usando la definición $\text{Cov}\,(X, Y) = E(XY) - E(X)\,E(Y)$, se obtiene el siguiente sistema de ecuaciones para a_1 y a_2: para $j = 1, 2$,

$$\begin{aligned} 0 \;=\; & \text{Cov}\,(Y, W_j) + E(Y)\,E(W_j) \\ & -[\,E(Y) - a_1\,E(W_1) - a_2\,E(W_2)\,]\,E(W_j) \\ & -a_1[\text{Cov}\,(W_1, W_j) + E(W_1)\,E(W_j)] \\ & -a_2[\text{Cov}\,(W_2, W_j) + E(W_2)\,E(W_j)]. \end{aligned}$$

Simplificando se obtiene

$$\text{Cov}\,(Y, W_j) - a_1\,\text{Cov}\,(W_1, W_j) - a_2\,\text{Cov}\,(W_2, W_j) = 0, \quad j = 1, 2.$$

Estas dos ecuaciones se pueden escribir en términos matriciales y es así como aparecen en (5.14). ▪

El resultado recién demostrado es similar al encontrado antes en el contexto de series de tiempo estacionarias y se puede extender sin dificultad el caso $Y, W_1, \ldots, W_n$. Se puede comprobar que los valores de los coeficientes $a_0, a_1, \ldots, a_n$ son

$$a_0 = E(Y) - \sum_{j=1}^{n} a_j \, E(W_j),$$

y escribiendo $\text{cov}(i,j)$ en lugar de $\text{Cov}(W_i, W_j)$, las constantes $a_1, \ldots, a_n$ son la solución al sistema de ecuaciones

$$\begin{pmatrix} \text{cov}(1,1) & \text{cov}(1,2) & \cdots & \text{cov}(1,n) \\ \text{cov}\,(2,1) & \text{cov}\,(2,2) & \cdots & \text{cov}\,(2,n) \\ \vdots & \vdots & & \vdots \\ \text{cov}\,(n,1) & \text{cov}\,(n,2) & \cdots & \text{cov}\,(n,n) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \text{Cov}(Y, W_1) \\ \text{Cov}\,(Y, W_2) \\ \vdots \\ \text{Cov}\,(Y, W_n) \end{pmatrix}.$$

Esta ecuación matricial también se puede escribir como

$$\sum_{i=1}^{n} a_j \, \text{Cov}\,(W_i, W_j) = \text{Cov}\,(Y, W_i), \quad i = 1, \ldots, n. \tag{5.15}$$

Observemos que, en el caso de series estacionarias, los términos $\text{Cov}\,(W_i, W_j)$ dependen de i y de j sólo a través de la diferencia $j - i$. Regresando al caso de dos variables predictoras, encontraremos ahora el valor mínimo que toma el error cuadrático medio.

Proposición 5.7 *El error cuadrático medio del estimador* $\ell\hat{Y}(W_1, W_2)$ *definido en* (5.12)*, es*

$$E[(Y - \ell\hat{Y}(W_1, W_2))^2] = \text{Var}(Y) - a_1 \, \text{Cov}\,(Y, W_1) - a_2 \, \text{Cov}\,(Y, W_2).$$

Demostración. Substituyendo el valor a_0 que aparece en (5.13), tenemos

que

$$\begin{aligned} & E[(Y - \ell\hat{Y}(W_1, W_2))^2] \\ & \quad = E[(Y - a_0 - a_1\, W_1 - a_2\, W_2)^2] \\ & \quad = E[(Y - E(Y) + a_1\, E(W_1) + a_2\, E(W_2) - a_1\, W_1 - a_2\, W_2)^2] \\ & \quad = E[(Y - E(Y) - a_1(W_1 - E(W_1)) - a_2(E(W_2) - W_2))^2]. \end{aligned}$$

Desarrollando el cuadrado sobre los tres términos que aparecen entre paréntesis y distribuyendo la esperanza, la expresión anterior se reduce a la que aparece en el enunciado. ∎

En el caso general de n variables predictoras, se puede comprobar que el error cuadrático medio del pronóstico lineal por mínimos cuadrados es

$$E[(Y - \ell\hat{Y}(W_1, \ldots, W_n))^2] = \text{Var}(Y) - \sum_{j=1}^{n} a_j \,\text{Cov}\,(Y, W_j).$$

Propiedades del pronóstico lineal sin estacionariedad

En esta sección se revisan algunas propiedades del pronóstico lineal por mínimos cuadrados $\ell\hat{Y}(W_1, \ldots, W_n)$ para series no necesariamente estacionarias. Recordemos que este pronóstico tiene la forma

$$\ell\hat{Y}(W_1, \ldots, W_n) = a_0 + a_1\, W_1 + \cdots + a_n\, W_n, \tag{5.16}$$

en donde $a_1, a_1, \ldots, a_n$ son constantes, $W_1, \ldots, W_n$ son las variables predictoras y Y es la variable a pronosticar. En la demostración de las propiedades de (5.16) es conveniente usar el siguiente cambio de notación,

$$\begin{aligned} \mathbb{P}_\ell(Y \,|\, W_1, \ldots, W_n) &:= \ell\hat{Y}(W_1, \ldots, W_n) \\ &= a_0 + a_1\, W_1 + \cdots + a_n\, W_n. \end{aligned} \tag{5.17}$$

La letra $\mathbb{P}$ indica pronóstico y el subíndice ℓ indica que se usan las variables predictoras $W_1, \ldots, W_n$ de manera lineal. Esta notación alternativa es conveniente pues permite escribir en la variable a pronosticar Y expresiones más elaboradas. Véanse, por ejemplo, los enunciados de la siguiente proposición.

Proposición 5.8 *El pronóstico lineal* $\mathbb{P}_\ell(X_{n+s} \,|\, W_1, \ldots, W_n)$ *definido en* (5.17) *satisface las siguientes propiedades:*

1. $\mathbb{P}_\ell(Y \,|\, W_1, \ldots, W_n) = E(Y) + \sum_{i=1}^{n} a_i \, (W_j - E(W_j))$.
2. $E(\mathbb{P}_\ell(Y \,|\, W_1, \ldots, W_n)) = E(Y)$.
3. $E[(Y - \mathbb{P}_\ell(Y \,|\, W_1, \ldots, W_n)) \cdot W_j] = 0, \quad j = 1, \ldots, n$.
4. $E[(Y - \mathbb{P}_\ell(Y \,|\, W_1, \ldots, W_n))^2] = \mathrm{Var}(Y) - \sum_{i=1}^{n} a_i \, \mathrm{Cov}\,(Y, W_i)$.
5. $\mathbb{P}_\ell(X + Y \,|\, W_1, \ldots, W_n) = \mathbb{P}_\ell(X \,|\, W_1, \ldots, W_n) + \mathbb{P}_\ell(Y \,|\, W_1, \ldots, W_n)$.
6. *Si* b *y* c *son constantes, entonces*
$$\mathbb{P}_\ell(bY + c \,|\, W_1, \ldots, W_n) = b \, \mathbb{P}_\ell(Y \,|\, W_1, \ldots, W_n) + c.$$
7. *Si* $c_0, \ldots, c_n$ *constantes, entonces*
$$\mathbb{P}_\ell(c_0 + c_1 \, W_1 + \cdots + c_n \, W_n \,|\, W_1, \ldots, W_n) = c_0 + c_1 \, W_1 + \cdots + c_n \, W_n.$$
8. *Si* $\mathrm{Cov}\,(Y, W_j) = 0$ *para* $j = 1, \ldots, n$, *entonces*
$$\mathbb{P}_\ell(Y \,|\, W_1, \ldots, W_n) = E(Y).$$
9. $\mathbb{P}_\ell(\mathbb{P}_\ell(Y \,|\, W_1, \ldots, W_n, W_{n+1}) \,|\, W_1, \ldots, W_n) = \mathbb{P}_\ell(Y \,|\, W_1, \ldots, W_n)$.

Antes de demostrar estas afirmaciones haremos algunas observaciones e interpretaciones. La propiedad (1) proporciona una representación del pronóstico que sólo depende de $a_1, \ldots, a_n$ y no de a_0. La identidad (2) dice que el valor esperado del pronóstico coincide con el valor esperado de la variable a estimar, es decir, el estimador es insesgado. En la igualdad (4) se muestra el error cuadrático medio del pronóstico. Las propiedades (5) y (6) indican que el operador pronóstico es lineal en la variable a estimar.

Finalmente, en la propiedad (7) se muestra el caso particular cuando la variable a estimar es una combinación lineal de las variables predictoras.

Demostración.

1. Ya se demostró que $a_0 = E(Y) - \sum_{i=1}^{n} a_i\, E(W_i)$; Por lo tanto,

$$\begin{aligned}
\mathbb{P}_\ell(Y \,|\, W_1, \ldots, W_n) &= a_0 + \sum_{i=1}^{n} a_i\, W_i \\
&= E(Y) - \sum_{i=1}^{n} a_i\, E(W_i) + \sum_{i=1}^{n} a_i\, W_i \\
&= E(Y) + \sum_{i=1}^{n} a_i\, (W_j - E(W_j)).
\end{aligned}$$

2. Esta identidad es evidente de la propiedad anterior al tomar esperanza.

3. Por la propiedad (1),

$$\begin{aligned}
&E[(Y - \mathbb{P}_\ell(Y \,|\, W_1, \ldots, W_n)) \cdot W_j] \\
&\quad = E(Y\, W_j) - E(Y)\, E(W_j) - \sum_{i=1}^{n} a_i\, E[(W_i - E(W_i))\, W_j] \\
&\quad = \mathrm{Cov}\,(Y, W_j) - \sum_{i=1}^{n} a_i\, \mathrm{Cov}\,(W_i, W_j) \\
&\quad = 0.
\end{aligned}$$

La última identidad se debe a que los coeficientes $a_0, a_1, \ldots, a_n$ satisfacen ese sistema de ecuaciones, el cual se encontró en (5.15).

4. Usando nuevamente la representación de la propiedad (1),

$$\begin{aligned}
&E[(Y - \mathbb{P}_\ell(Y \mid W_1, \ldots, W_n))^2] \\
&\quad = E[(Y - E(Y) - \sum_{i=1}^{n} a_i\,(W_i - E(W_i)))^2] \\
&\quad = E[(Y - E(Y))^2] - 2\,E[(Y - E(Y))\,\sum_{i=1}^{n} a_i\,(W_i - E(W_i))] \\
&\qquad + E(\sum_{i=1}^{n} a_i\,(W_i - E(W_i)))^2 \\
&\quad = \mathrm{Var}(Y) - 2\,\sum_{i=1}^{n} a_i\,\mathrm{Cov}\,(Y, W_i) + \sum_{i=1}^{n}\,\sum_{j=1}^{n} a_i\,a_j\,\mathrm{Cov}\,(W_i, W_j) \\
&\quad = \mathrm{Var}(Y) - 2\,\sum_{i=1}^{n} a_i\,\mathrm{Cov}\,(Y, W_i) + \sum_{i=1}^{n} a_i\,\underbrace{\sum_{j=1}^{n} \mathrm{Cov}\,(W_i, W_j)\,a_j}_{\mathrm{Cov}(Y, W_i)} \\
&\quad = \mathrm{Var}(Y) - \sum_{i=1}^{n} a_i\,\mathrm{Cov}\,(Y, W_i).
\end{aligned}$$

5. Denotemos por $a_0(Y), a_1(Y), \ldots, a_n(Y)$ a los coeficientes de la combinación lineal del pronóstico para la variable Y, es decir,

$$\mathbb{P}_\ell(Y \mid W_1, \ldots, W_n) = a_0(Y) + a_1(Y)\,W_1 + \cdots + a_n(Y)\,W_n.$$

Los términos $a_0(X), a_1(X), \ldots, a_n(X)$ tienen un significado análogo para la variable X. Se requiere demostrar que $a_i(X+Y) = a_i(X) + a_i(Y)$, para $i = 1, \ldots, n$. Así, tenemos que el sistema de ecuaciones para los coeficientes $a_1, \ldots, a_n$ asociados a $X+Y$ es, para $i = 1, \ldots, n$,

$$\begin{aligned}
\sum_{j=1}^{n} \mathrm{Cov}\,(W_i, W_j)\,a_j(X+Y) &= \mathrm{Cov}\,(X+Y, W_i) \\
&= \mathrm{Cov}\,(X, W_i) + \mathrm{Cov}\,(Y, W_i) \\
&= \sum_{j=1}^{n} \mathrm{Cov}\,(W_i, W_j)\,a_j(X) \\
&\quad + \sum_{j=1}^{n} \mathrm{Cov}\,(W_i, W_j)\,a_j(Y).
\end{aligned}$$

Es decir, para $i = 1, \ldots, n$,

$$\sum_{j=1}^{n} \mathrm{Cov}\,(W_i, W_j)\,[\,a_j(X+Y) - a_j(X) - a_j(Y)\,] = 0.$$

Por ejemplo, para el caso de dos variables W_1 y W_2, tenemos el sistema de ecuaciones

$$\begin{pmatrix} \mathrm{Cov}(W_1,W_1) & \mathrm{Cov}(W_1,W_2) \\ \mathrm{Cov}\,(W_2,W_1) & \mathrm{Cov}\,(W_2,W_2) \end{pmatrix} \begin{pmatrix} a_1(X+Y) - a_1(X) - a_1(Y) \\ a_2(X+Y) - a_2(X) - a_2(Y) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Dado que la matriz de covarianzas $(\mathrm{Cov}\,(W_i, W_j))_{i,j=1}^{n}$ es positiva definida, se concluye que

$$a_j(X+Y) - a_j(X) - a_j(Y) = 0, \quad \text{para} \;\; j = 1, \ldots, n.$$

Finalmente,

$$\begin{aligned} a_0(X+Y) &= E(X+Y) - \sum_{i=1}^{n} a_i(X+Y)\,E(W_i) \\ &= E(X) + E(Y) - \sum_{i=1}^{n} [\,a_i(X) + a_i(Y)\,]\,E(W_i) \\ &= E(X) - \sum_{i=1}^{n} a_i(X)\,W_i + E(Y) - \sum_{i=1}^{n} a_i(Y)\,E(W_i) \\ &= a_0(X) + a_0(Y). \end{aligned}$$

6. El sistema de ecuaciones para los coeficientes $a_1, \ldots, a_n$ asociados a la variable a estimar bY es, para $i = 1, \ldots, n$,

$$\begin{aligned} \sum_{j=1}^{n} \mathrm{Cov}\,(W_i, W_j)\,a_j(bY) &= \mathrm{Cov}\,(bY, W_i) \\ &= b\,\mathrm{Cov}\,(Y, W_i) \\ &= b \sum_{j=1}^{n} \mathrm{Cov}\,(W_i, W_j)\,a_j(Y) \\ &= \sum_{j=1}^{n} \mathrm{Cov}\,(W_i, W_j)\,[b\,a_j(Y)]. \end{aligned}$$

Es decir,

$$\sum_{j=1}^{n} \mathrm{Cov}\,(W_i, W_j)\,[a_j(bY) - b\,a_j(Y)] = 0.$$

Nuevamente, la propiedad de la matriz de covarianzas de ser positiva definida implica que

$$a_j(bY) - b\,a_j(Y) = 0.$$

Para el coeficiente a_0,

$$\begin{aligned} a_0(bY) &= E(bY) - \sum_{i=1}^{n} a_i(bY)\,E(W_i) \\ &= b\,E(Y) - b\,\sum_{i=1}^{n} a_i(Y)\,E(W_i) \\ &= b\,a_0(Y). \end{aligned}$$

Finalmente, añadir una constante c a la variable bY es un caso particular del inciso anterior. Por lo tanto,

$$a_j(bY + c) = b\,a_j(Y) + a_j(c), \quad \text{para} \;\; j = 0, 1, \ldots, n,$$

en donde se puede comprobar que

$$a_j(c) = \begin{cases} c & \text{si} \;\; j = 0, \\ 0 & \text{si} \;\; j = 1, \ldots, n. \end{cases}$$

Por lo tanto,

$$\begin{aligned} \mathbb{P}_\ell(bY + c \,|\, W_1, \ldots, W_n) &= a_0(bY + c) + \sum_{i=1}^{n} a_i(bY + c)\,W_i \\ &= b\,a_0(Y) + c + \sum_{i=1}^{n} b\,a_i(Y)\,W_i \\ &= b[a_0(Y) + \sum_{i=1}^{n} a_i(Y)\,W_i] + c \\ &= b\,\mathbb{P}_\ell(Y \,|\, W_1, \ldots, W_n) + c. \end{aligned}$$

7. Se puede comprobar que, para $j \in \{0, 1, \ldots, n\}$ fijo, las constantes

$$a_i(W_j) = \begin{cases} 1 & \text{si } \ i = j, \\ 0 & \text{si } \ i \neq j, \end{cases}$$

satisfacen las ecuaciones

$$a_0(W_j) = E(W_j) + \sum_{i=1}^{n} a_i(W_j)\, E(W_i),$$

y

$$\begin{pmatrix} \operatorname{cov}(1,1) & \operatorname{cov}(1,2) & \cdots & \operatorname{cov}(1,n) \\ \operatorname{cov}(2,1) & \operatorname{cov}(2,2) & \cdots & \operatorname{cov}(2,n) \\ \vdots & \vdots & & \vdots \\ \operatorname{cov}(n,1) & \operatorname{cov}(n,2) & \cdots & \operatorname{cov}(n,n) \end{pmatrix} \begin{pmatrix} a_1(W_j) \\ a_2(W_j) \\ \vdots \\ a_n(W_j) \end{pmatrix} = \begin{pmatrix} \operatorname{Cov}(W_j, W_1) \\ \operatorname{Cov}(W_j, W_2) \\ \vdots \\ \operatorname{Cov}(W_j, W_n) \end{pmatrix},$$

en donde $\operatorname{cov}(i,j)$ es $\operatorname{Cov}(W_i, W_j)$. Por lo tanto,

$$\mathbb{P}_\ell(W_j \,|\, W_1, \ldots, W_n) = W_j, \quad \text{para } \ j = 1, \ldots, n.$$

Entonces, por la propiedad de linealidad, propiedades (5) y (6), tenemos que

$$\begin{aligned} \mathbb{P}_\ell(c_0 + \sum_{j=1}^{n} c_j\, W_j \,|\, W_1, \ldots, W_n) &= c_0 + \sum_{j=1}^{n} c_j\, \mathbb{P}_\ell(W_j \,|\, W_1, \ldots, W_n) \\ &= c_0 + \sum_{j=1}^{n} c_j\, W_j. \end{aligned}$$

8. Cuando $\operatorname{Cov}(Y, W_i) = 0$ para $i = 1, \ldots, n$, el lado derecho del sistema de ecuaciones para los coeficientes $a_0, a_1, \ldots, a_n$ es el vector cero y la solución al sistema es el vector cero. Por la representación que aparece en la primera propiedad, se concluye que

$$\mathbb{P}_\ell(Y \,|\, W_1, \ldots, W_n) = E(Y).$$

9. Sean $a_1, \ldots, a_{n+1}$ los coeficientes del pronóstico para Y basado en las variables $W_1, \ldots, W_{n+1}$, y sean $b_1, \ldots, b_n$ los coeficientes del pronóstico para W_{n+1} basado en $W_1, \ldots, W_n$. Por la representación de la

propiedad (1) y por la propiedad de linealidad,

$$\begin{aligned}
&\mathbb{P}_\ell(\mathbb{P}_\ell(Y \,|\, W_1, \ldots, W_{n+1}) \,|\, W_1, \ldots, W_n) \\
&= \mathbb{P}_\ell(E(Y) + \sum_{j=1}^{n+1} a_j \,(W_j - E(W_j)) \,|\, W_1, \ldots, W_n) \\
&= E(Y) + \sum_{j=1}^{n+1} a_j \,(\mathbb{P}_\ell(W_j \,|\, W_1, \ldots, W_n) - E(W_j)) \\
&= E(Y) + \sum_{j=1}^{n} a_j \,(W_j - E(W_j)) \\
&\qquad + a_{n+1} \,[\mathbb{P}_\ell(W_{n+1} \,|\, W_1, \ldots, W_n) - E(W_{n+1})] \\
&= E(Y) + \sum_{j=1}^{n} a_j \,(W_j - E(W_j)) + a_{n+1} \sum_{j=1}^{n} b_j \,(W_j - E(W_j)) \\
&= E(Y) + \sum_{j=1}^{n} (a_j + a_{n+1}\, b_j) \,(W_j - E(W_j)) \\
&= \mathbb{P}_\ell(Y \,|\, W_1, \ldots, W_n).
\end{aligned}$$

La última igualdad se cumple ya que se puede comprobar que los números $a_j + a_{n+1}b_j$, para $j = 1, \ldots, n$, son los coeficientes del pronóstico para Y con base en las variables $W_1, \ldots, W_n$. Concluimos la demostración verificando esta afirmación

$$\begin{aligned}
\sum_{j=1}^{n} \operatorname{Cov}(W_i, W_j)(a_j + a_{n+1}b_j) &= \operatorname{Cov}(Y, W_i) - a_{n+1} \operatorname{Cov}(W_i, W_{n+1}) \\
&\quad + a_{n+1} \operatorname{Cov}(W_{n+1}, W_i) \\
&= \operatorname{Cov}(Y, W_i).
\end{aligned}$$

∎

A continuación se muestran dos ejemplos en los cuales se aplican los resultados anteriores. Usaremos indistintamente las expresiones $\ell\hat{X}_n(s)$ y $\mathbb{P}_\ell(X_{n+s} \,|\, X_1, \ldots, X_n)$ para el mejor estimador lineal por mínimos cuadrados para X_{n+s} basado en el historial $X_1, \ldots, X_n$ de una serie de tiempo.

Pronóstico a un tiempo para un modelo AR(p)

Sea $\{X_t : t \in \mathbb{Z}\}$ una serie de tiempo que sigue un modelo AR(p) dado por

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + Z_t, \quad t \in \mathbb{Z}. \tag{5.18}$$

Encontraremos una expresión para el pronóstico lineal por mínimos cuadrados de la variable X_{n+1} con base en las variables $X_1, \ldots, X_n$. Para ello supondremos lo siguiente:

- El número de datos n es tal que $n \geqslant p$.
- Los coeficientes $\alpha_1, \ldots, \alpha_p$ son tales que la serie es estacionaria.
- La variable Z_t no está correlacionada con X_s para $s < t$, es decir, $\text{Cov}(Z_t, X_s) = 0$, para $s = t-1, t-2, \ldots$

Proposición 5.9 *Sea $\{X_t : t \in \mathbb{Z}\}$ una serie de tiempo que sigue un modelo* AR(p) *especificado en* (5.18). *Bajo las hipótesis anteriores, el mejor estimador lineal para X_{n+1} basado en $X_1, \ldots, X_n$ es*

$$\ell\hat{X}_n(1) = \alpha_1 X_n + \cdots + \alpha_p X_{n+1-p},$$

con error cuadrático medio

$$E[(X_{n+1} - \ell\hat{X}_n(1))^2] = \gamma(0) - \sum_{i=n+1-p}^{n} \alpha_{n+1-i}\, \gamma(n+1-i).$$

Demostración. Usando las propiedades generales demostradas antes sobre el pronóstico lineal, y por la especificación (5.18), tenemos que

$$\begin{aligned}
\ell\hat{X}_n(1) &= \mathbb{P}_\ell(X_{n+1} \,|\, X_1, \ldots, X_n) \\
&= \mathbb{P}_\ell(\alpha_1 X_n + \cdots + \alpha_p X_{n+1-p} + Z_{n+1} \,|\, X_1, \ldots, X_n) \\
&= \alpha_1 \mathbb{P}_\ell(X_n \,|\, X_1, \ldots, X_n) + \cdots + \alpha_p \mathbb{P}_\ell(X_{n+1-p} \,|\, X_1, \ldots, X_n) \\
&\quad + \mathbb{P}_\ell(Z_{n+1} \,|\, X_1, \ldots, X_n) \\
&= \alpha_1 X_n + \cdots + \alpha_p X_{n+1-p}.
\end{aligned}$$

Observe que aquí sólo aparecen las p variables previas inmediatas a la variable a pronosticar X_{n+1}, en donde $p \leqslant n$. De esta manera, hemos encontrado una fórmula explícita para pronosticar un paso adelante una serie AR(p) por mínimos cuadrados a través de combinaciones lineales de los valores previos observados. Como la expresión general de este estimador es la combinación lineal

$$\mathbb{P}_\ell(X_{n+1}) = a_0 + a_1 X_1 + \cdots + a_n X_n,$$

concluimos que las constantes son: $a_0 = 0$, $a_i = \alpha_{n+1-i}$ para $i = 1, \ldots, p$, y $a_i = 0$ para $i = p+1, \ldots, n$.

Por otro lado, recordemos que la fórmula general del error cuadrático medio para el pronóstico $\mathbb{P}_\ell(Y \,|\, W_1, \ldots, W_n) = a_0 + a_1 W_1 + \cdots + a_n W_n$ es

$$E[(Y - \mathbb{P}_\ell(Y \,|\, W_1, \ldots, W_n))^2] = \text{Var}(Y) - \sum_{i=1}^{n} a_i \text{Cov}(Y, W_i).$$

Por lo tanto,

$$\begin{aligned} E[(X_{n+1} - \mathbb{P}_\ell(X_{n+1} \,|\, X_1, \ldots, X_n))^2] & \\ &= \text{Var}(X_{n+1}) - \sum_{i=n+1-p}^{n} \alpha_{n+1-i} \,\text{Cov}(X_{n+1}, X_i) \\ &= \gamma(0) - \sum_{i=n+1-p}^{n} \alpha_{n+1-i} \,\gamma(n+1-i), \end{aligned}$$

en donde

$$\begin{aligned} \gamma(0) &= \frac{\sigma^2}{1 - \alpha_1\rho(1) - \alpha_2\rho(2) - \cdots - \alpha_p\rho(p)}, \\ \gamma(\tau) &= \gamma(0) \cdot \rho(\tau), \quad \text{para } \tau \in \mathbb{Z}, \end{aligned}$$

y $\rho(1), \ldots, \rho(p)$ están determinadas por las ecuaciones de Yule-Walker. ▪

Pronóstico a cualquier tiempo para un modelo AR(1)

A continuación se considera una serie de tiempo $\{X_t : t \in \mathbb{Z}\}$ que sigue un modelo AR(1) de media μ, es decir,

$$X_t - \mu = \alpha(X_{t-1} - \mu) + Z_t, \quad t \in \mathbb{Z}. \tag{5.19}$$

El siguiente resultado establece una fórmula para estimar el valor de la variable X_{n+s} para cualquier entero $s \geqslant 1$ de manera lineal usando las variables $X_1, \ldots, X_n$.

Proposición 5.10 *Sea $\{X_t : t \in \mathbb{Z}\}$ una serie que sigue un modelo* AR(1) *con media μ especificado en* (5.19). *Sea $s \geqslant 1$ un número entero. El pronóstico lineal por mínimos cuadrados para X_{n+s} basado en $X_1, \ldots, X_n$, es*

$$\ell\hat{X}_n(s) = (1 - \alpha^s)\mu + \alpha^s X_n,$$

con error cuadrático medio

$$E[(X_{n+s} - \ell\hat{X}_n(s))^2] = \frac{1 - \alpha^{2s}}{1 - \alpha^2}\,\sigma^2.$$

Demostración. Por brevedad en la escritura omitiremos indicar las variable predictoras $X_1, \ldots, X_n$ y utilizaremos la notación $\mathbb{P}_\ell$. Aplicando el operador pronóstico $\mathbb{P}_\ell$ a la variable $X_{n+s} - \mu$ se obtiene

$$\begin{aligned} \mathbb{P}_\ell(X_{n+s} - \mu) &= \mathbb{P}_\ell(\alpha(X_{n+s-1} - \mu) + Z_{n+s}) \\ &= \alpha\,\mathbb{P}_\ell(X_{n+s-1} - \mu) + \alpha\,\mathbb{P}_\ell(Z_{n+s}). \end{aligned}$$

Como $s \geqslant 1$, se cumple que $\mathrm{Cov}(Z_{n+s}, X_k) = 0$ para $k = 1, \ldots, n$ y, por lo

tanto, $\mathbb{P}_\ell(Z_{n+s}) = E(Z_{n+s}) = 0$. En consecuencia,

$$\begin{aligned}
\mathbb{P}_\ell(X_{n+s} - \mu) &= \alpha\, \mathbb{P}_\ell(X_{n+s-1} - \mu) \\
&= \alpha^2\, \mathbb{P}_\ell(X_{n+s-2} - \mu)] \\
&\vdots \\
&= \alpha^s\, \mathbb{P}_\ell(X_n - \mu) \\
&= \alpha^s\, (X_n - \mu),
\end{aligned}$$

de donde se obtiene que

$$\mathbb{P}_\ell(X_{n+s}) = (1 - \alpha^s)\mu + \alpha^s X_n.$$

Este es el estimador lineal para X_{n+s} por mínimos cuadrados basado en las variables $X_1, \ldots, X_n$. Como la expresión general de este estimador es la combinación lineal

$$\mathbb{P}_\ell(X_{n+s}) = a_0 + a_1 X_1 + \cdots + a_n X_n,$$

concluimos que las constantes son: $a_0 = (1 - \alpha^s)\mu$, $a_1 = \cdots = a_{n-1} = 0$ y $a_n = \alpha^s$. Por otro lado, el error cuadrático medio es

$$\begin{aligned}
E[(X_{n+s} - \mathbb{P}_\ell(X_{n+s}))^2] &= \mathrm{Var}(X_{n+s}) - \sum_{i=1}^{n} a_i\, \mathrm{Cov}(X_{n+s}, X_i) \\
&= \mathrm{Var}(X_{n+s}) - \alpha^s \mathrm{Cov}(X_{n+s}, X_n) \\
&= \gamma(0) - \alpha^s\, \gamma(s) \\
&= \gamma(0)(1 - \alpha^s \frac{\gamma(s)}{\gamma(0)}) \\
&= \gamma(0)(1 - \alpha^s\, \rho(s)).
\end{aligned}$$

Substituyendo las expresiones $\gamma(0) = \sigma^2/(1 - \alpha^2)$ y $\rho(s) = \alpha^s$ encontradas para un proceso AR(1), tenemos que el error cuadrático medio es

$$E[(X_{n+s} - \mathbb{P}_\ell(X_{n+s}))^2] = \frac{1 - \alpha^{2s}}{1 - \alpha^2}\, \sigma^2.$$

■

§

Con esto concluimos nuestra breve introducción a algunos métodos de pronósticos para series de tiempo. Se pueden encontrar exposiciones más completas en Brockwell y Davis [7], [8], Cryer y Chan [15], o en Shumway y Stoffer [33].

5.4. Ejercicios

La esperanza condicional como pronóstico

108. Considere la serie de tiempo con tendencia $\{X_t : t \in \mathbb{Z}\}$ dada por

$$X_t = (a_0 + a_1 t + a_2 t^2) + Z_t, \quad t \in \mathbb{Z},$$

en donde a_0, a_1 y a_2 son constantes y $m(t) = a_0 + a_1 t + a_2 t^2$ representa la tendencia cuadrática de la serie. Suponga que se conocen los valores de $X_1, \ldots, X_n$. Sea $s \geqslant 1$ un número entero constante.

a) Encuentre el pronóstico $\hat{X}_n(s)$ definido por la esperanza condicional (5.1).

b) Encuentre el error cuadrático medio del estimador.

109. Considere la serie de tiempo $\{X_t : t \in \mathbb{Z}\}$ que sigue el modelo AR(1) dado por

$$X_t = \frac{1}{2} X_{t-1} + Z_t, \quad t \in \mathbb{Z}.$$

Suponga que se han observado los siguientes valores $x_1 = 4$, $x_2 = 1.70$, $x_3 = 1.15$, $x_4 = 2.11$ y $x_5 = 3.64$.

a) Encuentre el valor del estimador $\hat{X}_5(s)$, para $s = 1, 2$.

b) Grafique las observaciones $x_1, \ldots, x_5$ y los valores estimados $\hat{x}_6$ y $\hat{x}_7$.

110. *Modelo* AR(1).

Sea $\{X_t : t \in \mathbb{Z}\}$ una serie de tiempo que sigue un modelo AR(1) especificado en (5.3). Sea $s \geqslant 0$ un número entero. Demuestre que

$$\hat{X}_n(s) = \alpha^s X_n.$$

111. Sea $\{X_t : t \in \mathbb{Z}\}$ una serie de tiempo que sigue el modelo AR(1) no estacionario dado por

$$X_t = X_{t-1} + Z_t, \quad t \in \mathbb{Z}.$$

a) Encuentre $\hat{X}_n(s)$.

b) Encuentre $e_n(s)$.

c) Encuentre el ECM.

112. *Modelo* AR(1) *de media* μ.

Sea $s \geqslant 0$ un entero y considere el pronóstico $\hat{X}_n(s)$ definido como la esperanza condicional que aparece en (5.1) para una serie de tiempo que sigue el modelo AR(1) de media μ, es decir,

$$X_t - \mu = \alpha(X_{t-1} - \mu) + Z_t, \quad t \in \mathbb{Z}.$$

Suponga que $-1 < \alpha < 1$. Compruebe que:

a) $\hat{X}_n(s) = \mu + \alpha^s(X_n - \mu)$.

b) $e_n(s) = \sum_{k=0}^{s-1} \alpha^k Z_{n+s-k}$.

c) $E(e_n(s)) = 0$.

d) $\text{Var}(e_n(s)) = \sigma^2 \dfrac{1-\alpha^{2s}}{1-\alpha^2}$.

e) $\lim\limits_{s\to\infty} \text{Var}(e_n(s)) = \dfrac{\sigma^2}{1-\alpha^2} = \text{Var}(X_t)$.

f) $\text{ECM} = \sigma^2 \dfrac{1-\alpha^{2s}}{1-\alpha^2}$.

113. *Modelo* MA(1) *de media* μ.

Sea $s \geqslant 1$ un número entero y considere el pronóstico $\hat{X}_n(s)$ definido como la esperanza condicional que aparece en (5.1) para una serie de tiempo que sigue el modelo MA(1) de media μ, es decir,

$$X_t - \mu = Z_t + \beta Z_{t-1}, \quad t \in \mathbb{Z}.$$

Compruebe que:

a) $\hat{X}_n(s) = \begin{cases} \mu + \beta Z_n & \text{si } s = 1, \\ \mu & \text{si } s \geqslant 2. \end{cases}$

b) $e_n(s) = \begin{cases} Z_{n+1} & \text{si } s = 1, \\ Z_{n+s} + \beta Z_{n+s-1} & \text{si } s \geqslant 2. \end{cases}$

c) $E(e_n(s)) = 0.$

d) $\text{Var}(e_n(s)) = \begin{cases} \sigma^2 & \text{si } s = 1, \\ (1+\beta^2)\sigma^2 & \text{si } s \geqslant 2. \end{cases}$

e) $\lim_{s\to\infty} \text{Var}(e_n(s)) = \text{Var}(X_t) = (1+\beta^2)\sigma^2.$

f) $\text{ECM} = \begin{cases} \sigma^2 & \text{si } s = 1, \\ (1+\beta^2)\sigma^2 & \text{si } s \geqslant 2. \end{cases}$

Pronóstico lineal para series estacionarias

114. Sea $\{X_t : t \in \mathbb{Z}\}$ una serie de tiempo estacionaria que sigue el modelo MA(1) dado por

$$X_t = Z_t - \beta Z_{t-1}, \quad t \in \mathbb{Z},$$

en donde $-1 < \beta < 1$ es una constante y $\{Z_t : t \in \mathbb{Z}\}$ es un proceso aleatorio puro de media 0 y varianza σ^2. Sea $n \in \mathbb{Z}$ fijo. Denote por $\mathbb{P}_\ell(X_{n+1})$ al mejor pronóstico lineal para X_{n+1} por mínimos cuadrados basado en todas las variables previas $X_n, X_{n-1}, \ldots$ Demuestre que:

a) $\mathbb{P}_\ell(X_{n+1}) = -\sum_{k=1}^{\infty} \beta^k X_{n+1-k}.$

b) El error cuadrático medio del estimador es σ^2.

115. Sea $\{X_t : t \in \mathbb{Z}\}$ una serie de tiempo estacionaria que sigue el modelo AR(1) dado por

$$X_t = \alpha X_{t-1} + Z_t, \quad t \in \mathbb{Z},$$

en donde $-1 < \alpha < 1$. Sea $s \geqslant 0$ un número entero. Defina el pronóstico para X_{n+s} a partir de $X_n, X_{n-1}, \ldots$ como la variable aleatoria $E(X_{n+s} \,|\, X_n, X_{n-1}, \ldots)$. Demuestre que

$$E(X_{n+s} \,|\, X_n, X_{n-1}, \ldots) = \alpha^s X_n.$$

Pronóstico lineal para series no estacionarias

116. Sea $\mathbb{P}_\ell(Y \,|\, W_1, W_2)$ el estimador lineal por mínimos cuadrados para Y basado en W_1 y W_2. Diga falso o verdadero:

   a) $\mathbb{P}_\ell(W_1 \,|\, W_1, W_2) = W_1$.

   b) $\mathbb{P}_\ell(W_1 + W_2 \,|\, W_1, W_2) = W_1 + W_2$.

   c) $\mathbb{P}_\ell(4W_2 - 1 \,|\, W_1, W_2) = 4W_2 - 1$.

117. El estimador lineal por mínimos cuadrados para Y basado en W_1 y W_2 es $\ell\hat{Y}(W_1, W_2) = a_0 + a_1 W_1 + a_2 W_2$, en donde la constante a_0 se obtiene de manera explícita, como se indica en la proposición 5.6, y las constantes a_1 y a_2 se obtiene usando la matriz

$$\begin{pmatrix} \operatorname{Cov}(W_1, W_1) & \operatorname{Cov}(W_1, W_2) \\ \operatorname{Cov}(W_2, W_1) & \operatorname{Cov}(W_2, W_2) \end{pmatrix}.$$

   Demuestre que esta matriz es:

   a) Simétrica, es decir, es idéntica a su transpuesta.

   b) Positiva definida, es decir, para cualquier $(u, v) \in \mathbb{R}^2$ se cumple que

$$(u, v) \begin{pmatrix} \operatorname{Cov}(W_1, W_1) & \operatorname{Cov}(W_1, W_2) \\ \operatorname{Cov}(W_2, W_1) & \operatorname{Cov}(W_2, W_2) \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} \geqslant 0.$$

118. Suponga que las variables predictoras W_1 y W_2 son independientes en el estimador lineal

$$\ell\hat{Y}(W_1, W_2) = a_0 + a_1\, W_1 + a_2\, W_2. \tag{5.20}$$

   Demuestre que:

   a) $a_0 = E(Y) - a_1\, E(W_1) - a_2\, E(W_2)$.

   b) $a_1 = \dfrac{\operatorname{Cov}(Y, W_1)}{\operatorname{Var}(W_1)}$.

   c) $a_2 = \dfrac{\operatorname{Cov}(Y, W_2)}{\operatorname{Var}(W_2)}$.

119. Sea $\hat{\ell Y}(W_1, W_2)$ el pronóstico lineal de la variable Y a partir de las variables W_1, W_2, es decir,

$$\hat{\ell Y}(W_1, W_2) = a_0 + a_1\, W_1 + a_2\, W_2,$$

en donde a_1, a_1 y a_n son ciertas constantes. Compruebe que el error cuadrático medio del estimador es

$$E[(Y - \hat{\ell Y}(W_1, W_2))^2] = \mathrm{Var}(Y) - a_1\, \mathrm{Cov}\,(Y, W_1) - a_2\, \mathrm{Cov}\,(Y, W_2).$$

120. Sea $\hat{\ell Y}(W_1, \ldots, W_n)$ el pronóstico lineal de la variable Y a partir de las variables $W_1, \ldots, W_n$, es decir,

$$\hat{\ell Y}(W_1, \ldots, W_n) = a_0 + a_1\, W_1 + \cdots + a_n\, W_n.$$

Demuestre que:

a) El coeficiente a_0 está dado por

$$a_0 = E(Y) - \sum_{j=1}^{n} a_j\, E(W_j).$$

b) Los coeficientes $a_1, \ldots, a_n$ son la solución al sistema de ecuaciones

$$\begin{pmatrix} \mathrm{cov}(1,1) & \mathrm{cov}(1,2) & \cdots & \mathrm{cov}(1,n) \\ \mathrm{cov}\,(2,1) & \mathrm{cov}\,(2,2) & \cdots & \mathrm{cov}\,(2,n) \\ \vdots & \vdots & & \vdots \\ \mathrm{cov}\,(n,1) & \mathrm{cov}\,(n,2) & \cdots & \mathrm{cov}\,(n,n) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \mathrm{Cov}(Y, W_1) \\ \mathrm{Cov}\,(Y, W_2) \\ \vdots \\ \mathrm{Cov}\,(Y, W_n) \end{pmatrix},$$

en donde $\mathrm{cov}(i, j) := \mathrm{Cov}(W_i, W_j)$.

121. Demuestre que la matriz que aparece en el ejercicio anterior es:

a) Simétrica, es decir, es idéntica a su transpuesta.

b) Positiva definida, es decir, para cualquier $(u_1, \ldots, u_n) \in \mathbb{R}^n$ se cumple que

$$(u_1, \ldots, u_n) \begin{pmatrix} \mathrm{cov}(1,1) & \mathrm{cov}(1,2) & \cdots & \mathrm{cov}(1,n) \\ \mathrm{cov}\,(2,1) & \mathrm{cov}\,(2,2) & \cdots & \mathrm{cov}\,(2,n) \\ \vdots & \vdots & & \vdots \\ \mathrm{cov}\,(n,1) & \mathrm{cov}\,(n,2) & \cdots & \mathrm{cov}\,(n,n) \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \geqslant 0.$$

122. Suponga que las variables predictoras $W_1, \ldots, W_n$ son independientes en el estimador lineal

$$\ell\hat{Y}(W_1, W_2) = a_0 + a_1\, W_1 + \cdots + a_n\, W_n. \tag{5.21}$$

Demuestre que

a) $a_0 = E(Y) - \sum_{j=1}^{n} a_j\, E(W_j)$.

b) $a_j = \dfrac{\mathrm{Cov}(Y, W_j)}{\mathrm{Var}(W_j)}$ para $j = 1, \ldots, n$.

123. Sea $\ell\hat{Y}(W_1, \ldots, W_n)$ el pronóstico lineal de la variable Y a partir de las variables $W_1, \ldots, W_n$, es decir,

$$\ell\hat{Y}(W_1, \ldots, W_n) = a_0 + a_1\, W_1 + \cdots + a_n\, W_n,$$

en donde $a_1, a_1, \ldots, a_n$ son ciertas constantes. Compruebe que el error cuadrático medio del estimador es

$$E[(Y - \ell\hat{Y}(W_1, \ldots, W_n))^2] = \mathrm{Var}(Y) - \sum_{j=1}^{n} a_j\, \mathrm{Cov}\,(Y, W_j).$$

Apéndice A

Códigos en lenguaje R

En este apartado se presentan los códigos utilizados en el ejemplo 2.1, en los ejemplos de la sección 4.1 y en el ejemplo con datos reales de la sección 4.4. Se pretende que el lector tome como referencia estos códigos para reproducirlos, modificarlos y/o mejorarlos para que ajuste sus propios modelos a la serie de tiempo que desee analizar.

A.1. Ejemplo Box-Cox

Recordemos lo que se menciona en el ejemplo 2.1: allí se obtuvo una muestra $x_1, \ldots, x_{1000}$ de datos provenientes de una distribución Exp(1). Se aplicó una transformación de Box-Cox con distintos valores de λ hasta obtener una muestra con el mayor *p-value* en la prueba de Kolmogorov-Smirnov de dos colas donde la hipótesis nula establece que la muestra proviene de una distribución normal estándar. Se determinó un valor de $\lambda = 0.2735$ obteniéndose un estadístico $D = 0.016336$, y un *p-value*$= 0.9523$. La serie original se representa por $\{x_t\}$, y la serie transformada se representa por $\{y_t\}$, donde para cada valor de t primero se calculó

$$z_t = \frac{x_t^{\lambda} - 1}{\lambda},$$

para después calcular

$$y_t = \frac{z_t - \widehat{\mu}}{\widehat{\sigma}},$$

donde $\widehat{\mu}$ y $\widehat{\sigma}$ son la media muestral y la desviación estándar muestral de $\{z_t\}$, respectivamente.

```
##########################################
# Ejemplo de búsqueda del valor de lambda
# en una transformación de Box-Cox
##########################################

##########################################
# La función t_boxcox siguiente aplica
# la transformación a los elementos del
# vector 'x' y evaluando el valor 'lambda'
##########################################
t_boxcox <- function(x,lambda){
    aux <- (x^lambda-1)/lambda
    return(aux)
}
#######################################
set.seed(1981)      # semilla para reproducibilidad
x <- rexp(1000,1)   # obtención de una muestra exponencial
#######################################
m <- 1000           # número de valores que serán probados
results <- numeric(m)
lambdas <- seq(0.05,1,length=m)
# se probarán m valores equidistantes entre 0.05 y 1
for(i in 1:m){
    y <- t_boxcox(x,lambdas[i])
    muy <- mean(y)
    sdy <- sd(y)
    # estandarización de la transformación:
    y_std <- (y-muy)/sdy
    # prueba de Kolmogorov-Smirnov (K-S)
    aux <- ks.test(y_std,'pnorm')
    results[i] <- aux$p.value
}
# En las siguientes tres líneas buscamos el valor de lambda
```

```
# que obtuvo un mejor resultado en la prueba con respecto
# al valor del p-value
max(results)
```

```
## [1] 0.9522975
```

```
maximo <- which(results==max(results))
lambdas[maximo] # valor elegido
```

```
## [1] 0.2734735
```

```
#################################
# las siguientes líneas almacenan la transformación con el
# valor de lambda elegido y posteriormente se hacen los
# gráficos del histograma y del diagrama de cuantiles para
# observar el correcto ajuste de los datos estandarizados
# a la distribución N(0,1)
#################################
y <- t_boxcox(x,lambdas[maximo])
muy <- mean(y); sdy <- sd(y)
y_std <- (y-muy)/sdy
par(mfrow=c(1,2),mar=c(2,2,1,1))
hist(y_std,freq = F,main = '',ylim = c(0,0.4))
curve(dnorm(x),add=T,col='blue',lwd=1.5)
qqnorm(y_std,pch=16,cex=0.5,main='')
qqline(y_std,col='blue',lwd=1.5)
```

```
# Finalmente, se muestran los valores del estadístico
# de prueba y del p-value de la prueba de K-S
ks.test(y_std,'pnorm')
```

```
##
##  One-sample Kolmogorov-Smirnov test
```

```
##
## data:  y_std
## D = 0.016337, p-value = 0.9523
## alternative hypothesis: two-sided
```

A.2. Ejemplos generados por simulación

En esta sección se muestran los códigos utilizados en la sección 4.1 y en ellos se establecen valores semilla para su reproducibilidad. En el primer caso se presentan las funciones utilizadas para obtener los gráficos. En el resto de ejemplos los gráficos se obtienen de forma similar.

Modelo AR(1)

```
library(simts) #para valores de ACF teórica
n <- 1000
set.seed(123)
z <- rnorm(n)
x <- numeric(n)
x0 <- 0
x[1] <- 0.5*x0+z[1]
for(t in 2:n){x[t] <- 0.5*x[t-1]+z[t]}
par(mfrow=c(3,1),mar=c(2,1,1,2),oma=c(0,0,0,0))
lags <- 12
a <- acf(x,lag.max=lags,col='blue',lwd=1.5,
         xlab='rezago',ylab='ACF',plot=F)
b <- pacf(x,lag.max=lags,col='blue',lwd=1.5,
          xlab='rezago',ylab='PACF',plot=F)
aa <- a$acf
aa_teo <- theo_acf(0.5,lagmax = 13)[1:13]
bb <- b$acf
y1 <- -3.5; y2 <- -y1
############################
plot(x,type='n',axes = F,xlim=c(-20,1080),ylim=c(y1,y2))
arrows(0,0,1040,0,length = 0.05)
arrows(0,y1,0,y2,length = 0.05)
```

```
lines(1:1000,x,col='blue')
axis(1,at=seq(0,1000,100),c('',seq(100,1000,100)),
     cex.axis=1 )
text(rep(-35,3),c(-2,0,2),c(-2,0,2))
segments(rep(-5,5),c(-2,0,2),rep(5,5),c(-2,0,2))
text(30,3.2,'$x_t$')
text(1050,-0.8,'$t$')

########################################
plot(aa,type='n',axes = F,
     xlim=c(-0.5,13),ylim=c(-0.3,1.2*max(aa)))
lines(0:lags,aa,col='blue')
points(0:lags,aa,pch=16,cex=0.8)
lines(0:lags,aa_teo,lty=3,lwd=1.5)
points(0:lags,aa_teo,pch=16,cex=0.8)
arrows(-0.1,0,12.5,0,length = 0.05)
arrows(0,-0.2,0,1.2*max(aa),length = 0.05)
text(1:12,rep(-0.25,12),c(1:12))
segments(1:12,rep(-0.05,12),1:12,rep(0.05,12))
text(rep(-0.6,3),c(0,0.5,1),c(0,0.5,1))
segments(rep(-0.05,2),c(0.5,1),rep(0.05,2),c(0.5,1))
text(1.1,1.05*max(aa),'ACF($\\tau$)')
text(12.75,-0.075,'$\\tau$')
legend(7,0.85,
       c('ACF muestral','ACF teórica'),
       bty = 'n', lwd=c(2,2), lty = c(1,3),
       col=c('blue',1))


########################################
mm <- min(bb)
plot(bb,type='n',axes = F,
     xlim=c(-0.5,13),ylim=c(1.1*mm,1.3*max(bb)))
lines(1:lags,bb,col='blue')
points(1:lags,bb,pch=16,cex=0.9)
```

```
arrows(-0.1,0,12.5,0,length = 0.05)
arrows(0,1.1*mm,0,1.2*max(bb),length = 0.05)
text(1:12,rep(-0.18,12),c(1:12))
segments(1:12,rep(-0.025,12),1:12,rep(0.025,12))
text(1.1,1.2*max(bb),'PACF($\\tau$)')
text(12.75,-0.075,'$\\tau$')
text(rep(-0.6,3),c(0,0.2,0.4),c(0,0.2,0.4))
segments(rep(-0.05,2),c(0.2,0.4),rep(0.05,2),c(0.2,0.4))
##############################################
##############################################
```

Modelo AR(2)

```
##AR(2)
library(simts) #para valores de ACF teórica
n <- 1000
set.seed(123)
z <- rnorm(n)
x <- numeric(n)
x0 <- 0
x[1] <- z[1]
x[2] <- 0.6*x[1]-0.3*x0+z[2]
for(t in 3:n){x[t] <- 0.6*x[t-1]-0.3*x[t-2]+z[t]}
lags <- 12
a <- acf(x,lag.max=lags,col='blue',lwd=1.5,
         xlab='rezago',ylab='ACF',plot=F)
b <- pacf(x,lag.max=lags,col='blue',lwd=1.5,
          xlab='rezago',ylab='PACF',plot=F)
aa <- a$acf  # Valores de la función ACF
aa_teo <- theo_acf(c(0.6,-0.3),lagmax = 13)[1:13]
bb <- b$acf  # Valores de la función PACF
```

Modelo MA(1)

```
##MA(1)
library(simts) #para valores de ACF teórica
n <- 1000
set.seed(123)
z <- rnorm(n)
x <- numeric(n)
z0 <- 0
x[1] <- 0.7*z0+z[1]
for(t in 2:n){x[t] <- 0.7*z[t-1]+z[t]}
lags <- 12
a <- acf(x,lag.max=lags,col='blue',lwd=1.5,
         xlab='rezago',ylab='ACF',plot=F)
b <- pacf(x,lag.max=lags,col='blue',lwd=1.5,
          xlab='rezago',ylab='PACF',plot=F)
aa <- a$acf  # Valores de la función ACF
aa_teo <- theo_acf(NULL,0.7,lagmax = 13)[1:13]
bb <- b$acf  # Valores de la función PACF
```

Modelo MA(2)

```
##MA(2)
library(simts) #para valores de ACF teórica
n <- 1000
set.seed(123)
z <- rnorm(n)
x <- numeric(n)
z0 <- 0
x[1] <- z[1]
x[2] <- -0.8*z[1]+0.6*z0+z[2]
for(t in 3:n){x[t] <- -0.8*z[t-1]+0.6*z[t-2]+z[t]}
lags <- 12
a <- acf(x,lag.max=lags,col='blue',lwd=1.5,
         xlab='rezago',ylab='ACF',plot=F)
b <- pacf(x,lag.max=lags,col='blue',lwd=1.5,
          xlab='rezago',ylab='PACF',plot=F)
```

```
aa <- a$acf  # Valores de la función ACF
aa_teo <- theo_acf(NULL,c(-0.8,0.6),lagmax = 13)[1:13]
bb <- b$acf  # Valores de la función PACF
```

Modelo ARMA$(1,1)$

```
##ARMA(1,1)
library(simts) #para valores de ACF teórica
n <- 1000
set.seed(123)
z <- rnorm(n)
x <- numeric(n)
x_0 <- z0 <- 0
x[1] <- 0.3*x_0+0.5*z0+z[1]
for(t in 2:n){x[t] <- 0.3*x[t-1]+0.5*z[t-1]+z[t]}
lags <- 12
a <- acf(x,lag.max=lags,col='blue',lwd=1.5,
         xlab='rezago',ylab='ACF',plot=F)
b <- pacf(x,lag.max=lags,col='blue',lwd=1.5,
          xlab='rezago',ylab='PACF',plot=F)
aa <- a$acf  # Valores de la función ACF
aa_teo <- theo_acf(0.3,0.5,lagmax = 13)[1:13]
bb <- b$acf  # Valores de la función PACF
```

Modelo ARMA$(2,2)$

```
##ARMA(2,2)
library(simts) #para valores de ACF teórica
n <- 1000
set.seed(123)
z <- rnorm(n)
x <- numeric(n)
x_0 <- z0 <- 0
x[1] <- z[1]
```

```
x[2] <- -0.3*x[1]+0.2*x_0+0.4*z0+0.1*z[1]+z[2]
for(t in 3:n){
 x[t] <- -0.3*x[t-1]+0.2*x[t-2]+0.4*z[t-2]+0.1*z[t-1]+z[t]
}
lags <- 12
a <- acf(x,lag.max=lags,col='blue',lwd=1.5,
         xlab='rezago',ylab='ACF',plot=F)
b <- pacf(x,lag.max=lags,col='blue',lwd=1.5,
          xlab='rezago',ylab='PACF',plot=F)
aa <- a$acf  # Valores de la función ACF
aa_teo <- theo_acf(c(-0.3,0.2),c(0.1,0.4),lagmax = 13)[1:13]
bb <- b$acf  # Valores de la función PACF
```

Modelo AR(1) con tendencia

```
##AR(1) con tendencia
n <- 100
set.seed(123)
z <- rnorm(n)
x <- numeric(n)
x_0 <- z0 <- 0
x[1] <- 0.5*x_0+0.1+ z[1]
for(t in 2:n){x[t] <- 0.5*x[t-1]+0.1*t+ z[t]}
lags <- 12
a <- acf(x,lag.max=lags,col='blue',lwd=1.5,
         xlab='rezago',ylab='ACF',plot=F)
b <- pacf(x,lag.max=lags,col='blue',lwd=1.5,
          xlab='rezago',ylab='PACF',plot=F)
aa <- a$acf  # Valores de la función ACF
bb <- b$acf  # Valores de la función PACF
```

Proceso diferenciado ARIMA$(2,1,0)$

```
##Proceso diferenciado
n <- 100
set.seed(123)
z <- rnorm(n)
x <- numeric(n)
x_0 <- z0 <- 0
x[1] <- 0.5*x_0+0.1+ z[1]
for(t in 2:n){x[t] <- 0.5*x[t-1]+0.1*t+ z[t]}
yt <- x[2:n]-x[1:(n-1)]
xx <- yt-mean(yt)
lags <- 12
a <- acf(x,lag.max=lags,col='blue',lwd=1.5,
         xlab='rezago',ylab='ACF',plot=F)
b <- pacf(x,lag.max=lags,col='blue',lwd=1.5,
          xlab='rezago',ylab='PACF',plot=F)
aa <- a$acf  # Valores de la función ACF
bb <- b$acf  # Valores de la función PACF
```

A.3. Ejemplo con datos reales

En esta sección se presenta el código utilizado en la sección 4.4, incluyendo la llamada a las librerías necesarias.

```
library(aTSA) # necesario para la función 'adf.test'
library(astsa)  # necesario para la función 'sarima'
library(whitestrap) # necesario para la prueba de White
library(tsoutliers) # necesario para la prueba de Jarque-Bera
library(forecast) # necesario para las funciones
                  # 'checkresiduals' y 'Arima'


# Se leen los datos en la carpeta de trabajo
# ~/Documentos = ubicación del archivo tempmedia.csv
```

```
setwd("~/Documentos")
tempmedia <- read.csv("tempmedia.csv",header=T)

# Formatea la base de datos como matriz
tempmedia<-as.matrix(tempmedia)

# Ahora como vector
vtempmedia<- as.vector(t(tempmedia))

# Formato de serie de tiempo
ymedia <- ts(vtempmedia, start = c(1971,1),end=c(2001,12),
             frequency = 12)
summary(ymedia)

# Seleccion de datos para modelar y pronóstico
yimed = window(ymedia,start=c(1975,1),end=c(2000,12))
yfmed = window(ymedia,start=c(2001,1),end=c(2001,12))

# Estadísticos descriptivos simples
summary(yimed)
Maximo_en <- which(yimed==max(yimed))
Minimo_en <- which(yimed==min(yimed))
meses <- c('ene','feb','mar','abr','may','jun',
           'jul','ago','sep','oct','nov','dic')
aux1 <- Maximo_en %/% 12
aux2 <- Maximo_en %% 12
aux3 <- Minimo_en %/% 12
aux4 <- Minimo_en %% 12
añoMax <- 1975+aux1
mesMax <- meses[aux2]
maxima <- yimed[Maximo_en] #Temperatura Máxima
añoMin <- 1975+aux3
mesMin <- meses[aux4]
minima <- yimed[Minimo_en] #Temperatura Mínima
cat('Temperatura máxima de',maxima,'en',mesMax,
```

```
    'del año',añoMax)
cat('Temperatura mínima de',minima,'en',mesMin,
    'del año',añoMin)

# ACF Y PACF DE SERIE DE TIEMPO ORIGINAL
plot(ymedia, ylim=c(20,33), xlim=c(1975,2000), col=2)
acf(yimed,50,main='')
pacf(yimed,50,main='')

# SERIE DE TIEMPO CON UNA DIFERENCIA DE 12 REZAGOS
difymedia=diff(ymedia,lag=12)
difymed = window(difymedia,start=c(1976,1),end=c(2000,12))
difymed_f= window(difymedia,start=c(2001,1),end=c(2001,12))

# Grafica de serie diferenciada, acf y pacf
plot(difymed)
acf(difymed,50)
pacf(difymed,50)

# PRUEBAS DE RAÍCES UNITARIAS PARA SERIE ORIGINAL
adf.test(yimed,14)
#DICKEY-FULLER AUMENTADA serie original

# PRUEBAS DE RAÍCES UNITARIAS PARA SERIE DIFERENCIADA
adf.test(difymed, 14)
#DICKEY-FULLER AUMENTADA serie diferenciada

##############################################
###### PROPUESTAS DE MODELOS ARIMA ########
##############################################
###### ARIMA(12,0,12) = ARMA(12,12)
###### Estimación de parámetros
dt_12_0_12 <- arima(difymed,c(12,0,12))
#Si se desea estimar un modelo ARMA con media de cero,
#entonces usar arima(difymed,c(12,0,12),include.mean=F)
```

```
dt_12_0_12
aic12012 <- -2*-318.63+2*26
aic12012

###### (12,0,4)
dt_12_0_4 <- arima(difymed,c(12,0,4))
dt_12_0_4
aic1204 <- -2*-356.73+2*18
aic1204

###### (3,0,12)
dt_3_0_12 <- arima(difymed,c(3,0,12))
dt_3_0_12
aic3012 <- -2*-322.83+2*17
aic3012

#########  Modelo SARIMA  ##############
####sarima(3,0,0)(0,0,1)[12] usa variable yimed
t_300_001_12<- Arima(yimed ,
                     order=c(3,0,0),
                     seasonal=c(0,1,1),
                     include.constant=TRUE,
                     include.mean=TRUE)
#estimación con comando Arima
t_300_001_12
aic300011_12 <- -2*-331.58+2*6
aic300011_12

###################################
#Mismos resultados modelando la variable
#yimed con SARIMA
##############################
sarima(yimed,12,0,12,P=0,D=1,Q=0,S=12)
sarima(yimed,12,0,4,P=0,D=1,Q=0,S=12,gg=TRUE)
#Mejor opción:
```

```
sarima(yimed,3,0,12,P=0,D=1,Q=0,S=12,gg=TRUE)
#(3,0,0)(0,0,1)[12]:
sarima(yimed,3,0,0,P=0,D=1,Q=1,S=12,gg=TRUE)


##############################
# Pruebas para  ARMA(12_0_12)#
##############################
# Cálculo de los residuos o errores
e_12_0_12<-dt_12_0_12$residuals
# Estandarización
e_st_12_0_12 <-dt_12_0_12$residuals/sd(e_12_0_12)
###Gráficas###
plot.ts(e_12_0_12)
acf(e_12_0_12,50)
pacf(e_12_0_12,50)
qqnorm(e_st_12_0_12, asp=1)
qqline(e_st_12_0_12)
##### Pruebas diagnóstico
adf.test(e_12_0_12, 13) # raíces unitarias
white_test(dt_12_0_12) #heterocedasticidad
JB <- JarqueBera.test(e_12_0_12) #normalidad
JB[[1]]
####################################
Bt <- Box.test(e_12_0_12,lag = 27, type = "Ljung-Box")
Bt
#p-valuesLB con gl correctos
pchisq(Bt$statistic, 3, lower.tail=FALSE)
#El p-value se calcula con los df del test
# (lag-(p+q) = 27-24 = 3)
checkresiduals(dt_12_0_12, lag=27)
####################################
Bt <- Box.test(e_12_0_12,lag = 36, type = "Ljung-Box")
Bt
#p-valuesLB con gl correctos
```

```
pchisq(Bt$statistic, 12, lower.tail=FALSE)
#El p-value se calcula con los df del test
# (lag-(p+q) = 36-24 = 12)
checkresiduals(dt_12_0_12, lag=36)
####################################
Bt <- Box.test(e_12_0_12,lag = 42, type = "Ljung-Box")
Bt
#p-valuesLB con gl correctos
pchisq(Bt$statistic, 18, lower.tail=FALSE)
#El p-value se calcula con los df del test
# (lag-(p+q) = 42-24 = 18)
checkresiduals(dt_12_0_12, lag=42)
# Si las autocorrelaciones provinieran de una serie
# de ruido blanco, entonces el estadísitico de la
# prueba Ljung-Box tendría una \chi^2 con (h-k)
# grados de libertad, donde k es el número de
# parámetros en el modelo. Si se calculan a partir
# de datos sin procesar (en lugar de los residuos
# de un modelo), establezca k=0.
####################################
# Criterios gráficos
hist(e_st_12_0_12,prob = TRUE, main = "")
curve(dnorm(x),add=T)
plot(dt_12_0_12) #raíces

##############################
# Pruebas para  ARMA(12_0_4)#
##############################
# Cálculo de los residuos o errores
e_12_0_4<-dt_12_0_4$residuals
# Estandarización
e_st_12_0_4 <-dt_12_0_4$residuals/sd(e_12_0_4)
###Gráficas###
plot.ts(e_12_0_4)
acf(e_12_0_4,50)
```

```
pacf(e_12_0_4,50)
qqnorm(e_st_12_0_4, asp=1)
qqline(e_st_12_0_4)
##### Pruebas diagnóstico
adf.test(e_12_0_4, 13) # raíces unitarias
white_test(dt_12_0_4) #heterocedasticidad
JB <- JarqueBera.test(e_12_0_4) #normalidad
JB[[1]]
####################################
Bt <- Box.test(e_12_0_4,lag = 27, type = "Ljung-Box")
Bt
#p-valuesLB con gl correctos
pchisq(Bt$statistic, 3, lower.tail=FALSE)
#El p-value se calcula con los df del test
# (lag-(p+q) = 27-24 = 3)
checkresiduals(dt_12_0_4, lag=27)
####################################
Bt <- Box.test(e_12_0_4,lag = 36, type = "Ljung-Box")
Bt
#p-valuesLB con gl correctos
pchisq(Bt$statistic, 12, lower.tail=FALSE)
#El p-value se calcula con los df del test
# (lag-(p+q) = 36-24 = 12)
checkresiduals(dt_12_0_4, lag=36)
####################################
Bt <- Box.test(e_12_0_4,lag = 42, type = "Ljung-Box")
Bt
#p-valuesLB con gl correctos
pchisq(Bt$statistic, 18, lower.tail=FALSE)
#El p-value se calcula con los df del test
# (lag-(p+q) = 42-24 = 18)
checkresiduals(dt_12_0_4, lag=42)
####################################
# Criterios gráficos
hist(e_st_12_0_4,prob = TRUE, main = "")
```

```
curve(dnorm(x),add=T)
plot(dt_12_0_4) #raíces

##############################
# Pruebas para  ARMA(3_0_12)#
##############################
# Cálculo de los residuos o errores
e_3_0_12<-dt_3_0_12$residuals
# Estandarización
e_st_3_0_12 <-dt_3_0_12$residuals/sd(e_3_0_12)
###Gráficas###
plot.ts(e_3_0_12)
acf(e_3_0_12,50)
pacf(e_3_0_12,50)
qqnorm(e_st_3_0_12, asp=1)
qqline(e_st_3_0_12)
##### Pruebas diagnóstico
adf.test(e_3_0_12, 13) # raíces unitarias
white_test(dt_3_0_12) #heterocedasticidad
JB <- JarqueBera.test(e_3_0_12) #normalidad
JB[[1]]
###################################
Bt <- Box.test(e_3_0_12,lag = 27, type = "Ljung-Box")
Bt
#p-valuesLB con gl correctos
pchisq(Bt$statistic, 3, lower.tail=FALSE)
#El p-value se calcula con los df del test
# (lag-(p+q) = 27-24 = 3)
checkresiduals(dt_3_0_12, lag=27)
###################################
Bt <- Box.test(e_3_0_12,lag = 36, type = "Ljung-Box")
Bt
#p-valuesLB con gl correctos
pchisq(Bt$statistic, 12, lower.tail=FALSE)
#El p-value se calcula con los df del test
```

```
# (lag-(p+q) = 36-24 = 12)
checkresiduals(dt_3_0_12, lag=36)
####################################
Bt <- Box.test(e_3_0_12,lag = 42, type = "Ljung-Box")
Bt
#p-valuesLB con gl correctos
pchisq(Bt$statistic, 18, lower.tail=FALSE)
#El p-value se calcula con los df del test
# (lag-(p+q) = 42-24 = 18)
checkresiduals(dt_3_0_12, lag=42)
####################################
# Criterios gráficos
hist(e_st_3_0_12,prob = TRUE, main = "")
curve(dnorm(x),add=T)
plot(dt_3_0_12) #raíces


######### Pruebas para ARIMA(3,0,0)(0,1,1)[12]
e_300011<-t_300_001_12$residuals
e_st_e_300011 <-t_300_001_12$residuals/sd(e_300011)
###Gráficas###
plot.ts(e_300011)
acf(e_300011,50)
pacf(e_300011,50)
qqnorm(e_st_e_300011, asp=1)
qqline(e_st_e_300011)
hist(e_st_e_300011,prob = TRUE, main = "")
curve(dnorm(x),add=T)
##### Pruebas diagnóstico
adf.test(e_300011, 13) # raíces unitarias
white_test(t_300_001_12) #heterocedasticidad
JarqueBera.test(e_300011) #normalidad
Box.test(e_300011,lag = 36, type = "Ljung-Box")
#p-valuesLB con gl correctos
pchisq(40.78, 32, lower.tail=FALSE)
```

```
#gráfico de residuales, acf y histograma
checkresiduals(e_300011, lag=36)
# En la última función, no se ejecuta la
# prueba de Ljung-Box porque la función
# checkresiduals no localiza la información
# de los grados de libertad en el vector
# t_300_001_12$residuals generado por la
# función 'sarima'.

############## PRONÓSTICO #################
#arima 12,0,12
fdt_12_0_12 <- forecast(dt_12_0_12,12)
plot(fdt_12_0_12,
     xlim=c(1999.0,2002.1),ylim = c(-5.0,5),
     main="Pronóstico a un tiempo ARIMA(12,0,12)")
par(new=TRUE)
plot(difymed_f,
     xlim=c(1999.0,2002.1),ylim = c(-5.0,5), lwd=2)
# se genera una tabla con los pronósticos muestrales:
fdt_12_0_12

#arima 12,0,4
fdt_12_0_4=forecast(dt_12_0_4,12)
plot(fdt_12_0_4,
     xlim=c(1999,2002.1),ylim = c(-5.0,5),
     main="Pronóstico a un tiempo ARIMA(12,0,4)")
par(new=TRUE)
plot(difymed_f,
     ylim=c(-5,5), xlim=c(1999.0,2002.1), lwd=2)
# se genera una tabla con los pronósticos muestrales:
fdt_12_0_4

#arima 3,0,12
fdt_3_0_12=forecast(dt_3_0_12,12)
plot(fdt_3_0_12,
```

```
     xlim=c(1999,2002.1),ylim = c(-5.0,5),
     main="Pronóstico a un tiempo ARIMA(3,0,12)")
par(new=TRUE)
plot(difymed_f,
     ylim=c(-5,5), xlim=c(1999.0,2002.1), lwd=2)
# se genera una tabla con los pronósticos muestrales.
fdt_3_0_12

####sarima(3,0,0)(0,0,1)[12]
fyt_300_011=forecast(t_300_001_12, 12)
plot(fyt_300_011,
     xlim=c(1999,2002.1),ylim = c(20.0, 35),
     main="Pronóstico SARIMA(3,0,0)(0,1,1)[12]")
par(new=TRUE)
plot(yfmed ,
     ylim=c(20,35), xlim=c(1999.0,2002.1), lwd=2)
# se genera una tabla con los pronósticos muestrales:
fyt_300_011

### Precisión del pronóstico ############
#evaluación usando los datos de prueba (observados de 2001)
# RMSE = Raíz cuadrada del ECM
accuracy(fdt_12_0_12,difymed_f)
accuracy(fdt_12_0_4,difymed_f)
accuracy(fdt_3_0_12,difymed_f)
accuracy(fyt_300_011,yfmed)
```

Bibliografía

[1] Akaike H. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6), 716–723, 1974.

[2] Bisgaard S., Kulahci M. *Time series analysis and forecasting by example.* Wiley, 2011.

[3] Bowman K., Shenton L. Omnibus test contours for departures from normality based on b1 and b2. *Biometrika* 62, 243–250, 1975.

[4] Box G. E. P., Jenkins G. M. *Time series analysis: forecasting and control.* Holden-Day, San Francisco, 1970.

[5] Box G. E. P., Jenkins G. M., Reinsel G. C. *Time series analysis: forecasting and control.* Prentice-Hall, 1994. Third edition.

[6] Box G. E. P., Pierce D.A. Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. *Journal of the American Statistical Association* 65 (332), 1509–26, 1970.

[7] Brockwell P. J., Davis R. A. *Time series: theory and methods.* Springer-Verlag, 1987.

[8] Brockwell P. J., Davis R. A. *Introduction to time series and forecasting.* Springer, 2016. Third edition.

[9] Campos J., Ericsson N. R., Hendry D. F. General-to-specific modeling: an overview and selected bibliography. *FRB International Finance Discussion Paper*, (838).

[10] Chan N. H. *Times series: applications to finance with* `R` *and* `S-Plus`. Wiley, 2010.

[11] Chatfield C. *The analysis of time series: an introduction.* Chapman & Hall/CRC, 1996. Fifth edition.

[12] Coghlan A. *A little book of* R *for time series.* Release 0.2, september 10, 2018.

[13] Conover W. J. *Practical nonparametric statistics.* John Wiley & Sons, 1999.

[14] Cowpertwait P. S. P., Metcalfe A. V. *Introductory time series with* R. Springer, 2009.

[15] Cryer J. D., Chan K. S. *Time series analysis with applications in* R. Springer, 2008. Second edition.

[16] Goodwin P. The Holt-Winters approach to exponential smoothing: 50 years old and going strong. *Foresight*, 19 (19), 30–33, 2010.

[17] Gut A. *Probability: a graduate course.* Springer, 2005.

[18] Hamilton J. D. *Time series analysis.* Princeton University Press, 1994.

[19] Hanke J. E., Reitsch A. G. *Pronósticos en los negocios.* Prentice-Hall, 1996. Quinta edición.

[20] Hanke J. E., Wichern D. *Business forecasting.* Pearson, 2014. Ninth edition.

[21] Jarque C., Bera A. Efficient tests for normality homoscedasticity and serial independence of regression residuals. *Econometric Letters* 6, 255–259, 1980.

[22] Jarque C., Bera A. A test for normality of observations and regression residuals. *International Statistical Review* 55, 163–172, 1987.

[23] Kalekar P. S. Time series forecasting using Holt-Winters exponential smoothing. *Kanwal Rekhi school of information Technology*, 4329008(13), 1–13, 2004.

[24] Kirchgässner G., Wolters J. *Introduction to modern time series analysis.* Springer, 2007.

[25] Kitagawa G. *Introduction to time series modeling.* CRC Press, 2010.

[26] Lamigueiro O. P. *Displaying time series, spatial, and space-time data with* R. CRC Press, 2014.

[27] Ljung G. M., Box G. E. On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297–303, 1978.

[28] Lima S., Gonçalves A. M., Costa M. Time series forecasting using Holt-Winters exponential smoothing: An application to economic data. *AIP Conference Proceedings*, Vol. 2186, No. 1, AIP Publishing LLC, 2019.

[29] Pankratz A. *Forecasting with dynamic regression models.* John Wiley & Sons, 1991.

[30] Pfaff B. *Analysis of integrated and cointegrated time series with* R. Springer, 2008.

[31] Sampson M. *Time series analysis.* Loglinear Publishing, 2001.

[32] Sedgewick R., Flajolet P. *An introduction to the analysis of algorithms.* Addison-Wesley, 2013.

[33] Shumway R. H., Stoffer D. S. *Time series analysis and its applications with* R *examples.* Springer, 2017. Fourth edition.

[34] Schwarz G. Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464, 1978.

[35] Wei W. W. S. *Time series analysis: univariate and multivariate methods.* Pearson, 2006.

[36] White H. L., Jr. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48 (4), 817-838, 1980.

[37] Wlliams D. *Probability with martingales.* Cambridge University Press, 1991.

Índice alfabético

White noise, 10

ACF, 17
ACVF, 16
AR, 91
AR(p)
  autocorrelación, 105
  autocovarianza, 107
  causal, 100
  de media μ, 110
  operador de retraso, 99
  polinomio, 99
  propiedades, 104
ARIMA, 124
ARMA, 111
  autocorrelación, 115
  autocovarianza, 115
  causal, 113
  de media μ, 124
  invertible, 113
  media y varianza, 114, 115
Autocorrelación
  de una serie de tiempo, 17
  muestral, 18
Autocovarianza
  de una serie de tiempo, 16
  muestral, 16

Box-Jenkins, 153

Caminata aleatoria, 11
  con ruido blanco, 13
  simétrica simple, 13
Causalidad
  AR(p), 100
  ARMA(p,q), 113
  tabla resumen, 131
Componente
  cíclico, 41
  de tendencia, 40
    eliminación de, 65
    estimación de, 65
  estacional, 42
    estimación, 63
  integración de –s, 44
  irregular o aleatorio, 43
Construcción
  de modelos, 153
    estimación, 169
    identificación, 154
    verificación, 185
Correlograma, 27

interpretación del, 68
Covarianza
de una serie de tiempo, 16

Diferenciación, 61
Distribuciones
finito-dimensionales, 8, 9

Ecuaciones
de Yule-Walker, 105
Estacionariedad, 23
de segundo orden, 25
débil, 25
estricta, 24
fuerte, 24
Estimación
de máxima verosimilitud, 178

Filtro lineal, 59
Función
ACF, 17
ACVF, 16
de autocorrelación, 17
de autocorrelación parcial, 133
de autocovarianza, 16
de correlación, 17
de covarianza, 16
media, 14
PACF, 133
varianza, 15

Integración
de componentes, 44
Invertibilidad
ARMA(p,q), 113
MA(q), 87
tabla resumen, 131

MA(1), 83
invertibilidad, 86
operador de retraso, 85
polinomio, 85
propiedades, 87
MA(q), 77
correlograma, 81
de media μ, 91
de orden infinito, 82
invertibilidad, 87
operador de retraso, 87
polinomio, 87
propiedades, 79
Matriz
positiva semidefinida, 36
Media
–s móviles, 77
de una serie de tiempo, 14
muestral, 15
Metodología
de Box-Jenkins, 153
Modelo
ARMA(p, q) media no cero, 123
AR(p) media no cero, 110
MA(q) media no cero, 90
–s estimación de los parámetros, 135
–s selección, 134
–s verificación, 135
aditivo, 44
AR, 91
autocorrelación, 105
autocovarianza, 107
causal, 100
de media μ, 110

propiedades, 104
AR(1), 93
AR(2), 98
AR(p), 99
AR[P] estacional, 109
ARIMA, 124
ARMA, 111
autocorrelación, 115
autocovarianza, 115
causal, 113
de media μ, 124
invertible, 113
media y varianza, 114, 115
ARMA(1,1), 118
ARMA[P,D,Q] estacional, 129
ARMA[P,Q] estacional, 122
autorregresivo, 91
de medias móviles integrado, 124
construcción de –s, 153
de medias móviles, 77
de orden infinito, 82
de promedios móviles, 77
IMA, 125
lineal generalizado, 83
MA, 77
de media μ, 91
MA(1), 83
propiedades, 87
MA(q)
correlograma, 81
de orden infinito, 82
propiedades, 79
MA[Q] estacional, 89
multiplicativo, 44
SARIMA, 129
Momento
–s de una serie de tiempo, 9
primer –, 9
segundo –, 9

Operador
autorregresivo, 99
de promedios móviles, 88
diferencia, 21
diferencia de orden k, 21
retraso, 19

PACF, 133
Polinomio de retrasos, 20
Proceso
–s con dependencia de corto alcance, 81
–s con memoria corta, 81
aleatorio puro, 10
de ruido, 10
filtrado, 59
Promedios móviles, 77
Pronóstico, 217
esperanza condicional, 218
lineal para series estacionarias, 226
lineal sin estacionariedad, 233
lineal, propiedades, 230

Regresión
armónica, 64
lineal, 54
polinomial, 56
Residuales, 184
Ruido blanco, 11

Serie de tiempo, 7, 8
Simulación, 160
Suavizamiento
    exponencial, 60
    por promedios móviles, 58

Transformación
    de Box-Cox, 47
    de centralización, 45, 46
    tasa de crecimiento, 53
    tasa de retorno, 52
      discreta, 53

Variables
    dummy, 73
Varianza
    de una serie de tiempo, 15
    muestral, 15

Yule-Walker, 105

Una introducción a las series de tiempo
fue editado por la Facultad de Ciencias
de la Universidad Nacional Autónoma de México.
Se terminó de editar el 20 de marzo de 2025.

Publicación electrónica.
En su composición se utilizó tipografía
LM roman 10/12 pt.

El cuidado de la edición estuvo a cargo de
Leticia Pacheco Gasca.