



Universidad de Valladolid

Facultad de Ciencias

TRABAJO FIN DE GRADO

Grado en Matemáticas

**Fundamentos matemáticos del análisis
de series temporales**

Autora: Sonia Carrillo Grande

Tutor: Eustasio del Barrio Tellado

Índice general

Introducción	5
1. Procesos estacionarios	9
1.1. Proyecciones y autocovarianzas	10
1.2. Procesos lineales	13
2. Teoría espectral	29
2.1. El Teorema de Herglotz	31
2.2. Análisis espectral de procesos ARMA	37
2.2.1. Invertibilidad y densidad espectral	40
2.2.2. Aproximaciones racionales para densidades espectrales	43
3. Predicción de procesos estacionarios	49
3.0.1. Algoritmo de Durbin-Levinson	51
3.0.2. Algoritmo de innovaciones	54
3.1. Predicción de un proceso ARMA(p,q)	58
4. Inferencia en modelos ARMA	61
4.1. Preprocesado: estacionaridad y estacionalidad	62
4.1.1. Periodograma	64
4.2. Estimación preliminar	69
4.2.1. Estimación de Yule-Walker	70
4.2.2. Estimación basada en el algoritmo de innovaciones	71
4.3. Estimación de máxima verosimilitud	73
4.4. Diagnóstico y validación	75
4.5. Selección del orden	78
Conclusiones	83
Bibliografía	85

Introducción

En cualquier campo de la ciencia nos encontramos con ejemplos de series temporales. En Economía, algunos ejemplos son los precios diarios de las acciones, los beneficios trimestrales de una empresa, el producto interior bruto anual, la tasa de desempleo, etc. En Demografía, la tasa de natalidad, la población total, etc. En Meteorología, la temperatura, las precipitaciones diarias en una ciudad, etc. En Medio Ambiente, las emisiones anuales de CO_2 , etc. En Medicina, los electrocardiogramas, etc. Esto nos sugiere la necesidad de desarrollar técnicas específicas para el análisis estadístico de series temporales. Con este fin es necesario disponer de modelos estadísticos apropiados para describir la dependencia temporal. Los modelos naturales en este ámbito son los procesos estacionarios. Estos son procesos que mantienen una estructura probabilística constante en el tiempo y esto posibilita la estimación de características asociadas y la predicción. En apariencia muchas series no son estacionarias pero pueden transformarse en series que pueden modelarse como realizaciones de algún proceso estacionario.

El objetivo de este Trabajo de Fin de Grado es estudiar los principales resultados de la teoría de procesos estacionarios en tiempo discreto a intervalos iguales con vistas a su aplicación en el análisis, ajuste y predicción de series temporales. Los procesos estacionarios Gaussianos juegan un papel fundamental en la teoría. Bajo la hipótesis de normalidad la distribución de un proceso estacionario está totalmente determinada por la media y la función de autocovarianza, lo que permite una descripción bastante simple de la estructura de un proceso. Uno de los resultados principales estudiados en esta memoria es el conocido como Teorema de Herglotz, que establece que las funciones de autocovarianza admiten una representación en términos de la conocida como distribución espectral.

Junto con la teoría espectral, uno de los elementos principales en el análisis de series temporales es el modelado con procesos lineales, que, esencialmente, pueden aproximarse mediante la clase de modelos ARMA (autorregresivos de media móvil). El origen y la base de estos modelos se establece en los traba-

jos de George Udny Yule (1871-1951) con las primeras especificaciones sobre procesos autoregresivos (AR) y Eugen Slutsky (1880-1948) de medias móviles (MA) aparecidos entre 1921 y 1937 (ver [6]). En 1970, George E. P. Box (1919-2013) y G. M. Jenkins (1932-1982) generalizaron los modelos ARMA a través de la publicación en 1970 del libro 'Time Series Analysis: Forecasting and Control' (ver [3]) sobre los modelos ARIMA (autorregresivos integrados de media móvil). A partir de ese momento, los procesos ARMA han tenido un desarrollo espectacular estableciendo lo que hoy conocemos como análisis moderno de series temporales.

Esta memoria consta de cuatro capítulos. En el primer capítulo comenzamos introduciendo algunas definiciones básicas sobre procesos estacionarios. A continuación introducimos otras herramientas fundamentales en el análisis de series temporales como la función de autocovarianza y las funciones de autocorrelación simple y parcial. Por último presentamos la clase de procesos lineales, que contiene la clase de los procesos ARMA. También se definen propiedades importantes de estos procesos, como son la causalidad y la invertibilidad, caracterizamos su función de autocovarianza y mostramos algunos ejemplos.

En el segundo capítulo abordamos el estudio de la teoría espectral. El capítulo empieza introduciendo algunas definiciones como la representación espectral de los procesos estacionarios y de la función de autocovarianza, la función de distribución espectral o la densidad espectral. A continuación se exponen resultados que caracterizan las funciones de autocovarianza, como el Teorema de Herglotz, y que nos aportan una vía más simple para el cálculo de la densidad espectral. Por último dedicamos una sección al análisis espectral de procesos ARMA, donde obtenemos una forma más sencilla de calcular su densidad espectral y una mejor caracterización de la invertibilidad de un proceso ARMA. También veremos la importancia de estos procesos, que pueden aproximar a cualquier otro proceso estacionario con un error pequeño.

En el tercer capítulo estudiaremos los principales algoritmos de predicción de procesos estacionarios: el algoritmo de Durbin-Levinson y el algoritmo de innovaciones. A continuación veremos que en el caso de los procesos ARMA el algoritmo de innovaciones es mejor para predecir porque nos lleva a expresiones más sencillas.

El cuarto y último capítulo lo dedicaremos a estudiar métodos de inferencia en modelos ARMA. El capítulo comienza describiendo algunas transformaciones previas que debemos efectuar a la serie para que se pueda modelar mediante un modelo ARMA. El resto del capítulo lo dedicamos a exponer el procedimiento a seguir para encontrar el modelo ARMA que mejor se ajuste a los datos observados. En primer lugar, describiremos el procedimiento que

debemos seguir para hacer una estimación preliminar de los parámetros del modelo con los órdenes p y q fijados. A continuación, estos parámetros nos servirán para encontrar los estimadores de máxima verosimilitud, y calcular sus errores estándar y los residuales del modelo. En la siguiente fase del procedimiento, se exponen unas pruebas prácticas para comprobar que los residuales no tienen estructura de dependencia y siguen un proceso de ruido blanco. Este procedimiento se realiza para diferentes valores de p y q y, al final de este capítulo, se discuten distintos métodos para elegir los órdenes adecuados para la construcción del mejor modelo ARMA que nos ayude a realizar pronósticos de valores futuros de la serie observada.

Los libros principales que se han usado a lo largo de la memoria son [4] y [5], ambos escritos por Brockwell y Davis. El primero se centra en la teoría matemática de procesos estacionarios mientras que el segundo ofrece, sobre todo, un análisis práctico de estimación. Por otra parte, el libro de Box y Jenkins ([3]) es una referencia clásica para el modelado ARMA y propone un marco general de análisis de series temporales. El libro de Peña ([7]) se ha utilizado para comprender el funcionamiento del periodograma, que veremos en el Capítulo 4. Las otras referencias que aparecen en la bibliografía tienen que ver con aspectos secundarios: [1] se ha utilizado para entender con más claridad el por qué de la descripción de técnicas de selección del orden del modelo al final del Capítulo 4, [6] para establecer un contexto histórico de los modelos ARMA y [2] como referencia para un resultado utilizado en el Capítulo 2.

Capítulo 1

Procesos estacionarios

Este capítulo tiene como objetivo presentar algunas ideas básicas sobre procesos estacionarios. La estacionaridad es una propiedad casi necesaria para que se puede hacer inferencia sobre series de tiempo, puesto que se refiere, en este contexto, a la permanencia de la distribución de la serie a lo largo del tiempo.

El objeto básico que manejamos en este capítulo es una sucesión $\{X_t\}_{t \in \mathbb{Z}}$ de variables aleatorias con valores complejos definidas en un mismo espacio de probabilidad (Ω, F, P) . A estos objetos les llamamos *series temporales*. El espacio de estas variables, satisfaciendo $E|X_t|^2 < \infty$, con el producto interno dado por $\langle X, Y \rangle = E(X\bar{Y})$ es $L^2(\Omega, F, P)$ y tiene estructura de espacio de Hilbert. Este será el espacio natural en el que trabajaremos en esta memoria.

La serie es estacionaria si

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

para cualquier elección de t_1, \dots, t_k , $h \in \mathbb{Z}$ y $k \geq 1$. A veces se alude a esta propiedad como estacionaridad *fuerte*, en contraste con la estacionaridad *débil*, que consiste en la permanencia en el tiempo de los momentos de primer y segundo orden, es decir, $\{X_t\}_{t \in \mathbb{Z}}$ es débilmente estacionaria si

$$E(X_t) = E(X_{t+h}) \quad \text{y} \quad Cov(X_t, X_{t+h}) = Cov(X_{t+k}, X_{t+k+h})$$

para toda elección de índices. Estas fórmulas se utilizan tanto en el caso real como en el complejo, pero tenemos que aclarar que en el caso complejo la covarianza es $Cov(X_t, X_{t+h}) = E((X_{t+h} - E(X_{t+h}))(X_t - E(X_t)))$.

La estacionaridad fuerte implica obviamente la estacionaridad débil (asumiendo momentos de orden dos finitos). Además, existe un caso importante en el que ambas propiedades son equivalentes: las series temporales Gaussianas. Una serie temporal es *Gaussiana* si

$$(X_{t_1}, \dots, X_{t_k}) \sim N_k(\mu_k, \Sigma_k)$$

para todo $k \geq 1$, es decir, si las funciones de distribución de $\{X_t\}$ son todas normales multivariantes.

Los procesos estacionarios juegan un papel crucial en el análisis de series temporales. Es claro que muchas series temporales observadas no son estacionarias en apariencia pero pueden transformarse en series que pueden modelarse razonablemente como realizaciones de algún proceso estacionario. La teoría de los procesos estacionarios se utiliza luego para el análisis, ajuste y predicción de las series resultantes.

En el resto de este capítulo se presenta la función de autocovarianza, principal herramienta en el análisis, ajuste y predicción de series, y la clase de los procesos lineales, que contiene la clase de los modelos autorregresivos de media móvil (ARMA).

1.1. Proyecciones y autocovarianzas

Antes de presentar la función de autocovarianza y otros conceptos útiles para conocer mejor los procesos estacionarios, necesitamos introducir la notación que se va a emplear a lo largo de la memoria. Utilizaremos $\overline{\text{sp}}\{X\}$ para referirnos al mínimo subespacio cerrado que contiene al objeto X y $P_M(X)$ a la proyección ortogonal del objeto X sobre el subespacio M .

La distribución de un proceso estacionario Gaussiano está totalmente determinada por la media y la función de autocovarianza. Por este motivo, la función de autocovarianza tiene un papel fundamental.

Bajo el supuesto de que la serie $\{X_t\}$ es estacionaria la media $\mu = E(X_t)$ es constante. A la función

$$\gamma(h) = \text{Cov}(X_{t+h}, X_t)$$

se la conoce como *función de autocovarianza* de la serie (ACVF, del inglés 'autocovariance function') y a la función

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}$$

como *función de autocorrelación* o *función de autocorrelación simple* (FAS o ACF, del inglés 'autocorrelation function') para todo entero h . En el caso estacionario ambas funciones solo dependen de h y, además, verifican las siguientes propiedades. La función de autocorrelación verifica:

- (i) $\rho(0) = 1$.

- (ii) $|\rho(h)| \leq 1$, para cada h .
- (iii) $\rho(\cdot)$, es una función Hermítica, es decir, $\rho(h) = \overline{\rho(-h)}$ para todo h . En el caso real, la función es par.

Las propiedades de la función de autocovarianza son las siguientes:

- (i) $\gamma(0) \geq 0$.
- (ii) $|\gamma(h)| \leq \gamma(0)$ para todo h .
- (iii) $\gamma(\cdot)$ es una función Hermítica, es decir, $\gamma(h) = \overline{\gamma(-h)}$ para todo h . En el caso real, la función es par.

La comprobación de estas propiedades es elemental, tal como se expone a continuación.

- (i) $\gamma(0) = Cov(X_t, X_t) = E((X_t - EX_t)(\overline{X_t - EX_t})) = E(|X_t - EX_t|^2) \geq 0$.
- (ii) Sabemos que $|\rho(h)| \leq 1$ para todo h , luego $|\rho(h)| = \left| \frac{\gamma(h)}{\gamma(0)} \right| = \frac{|\gamma(h)|}{\gamma(0)} \leq 1$. Entonces $|\gamma(h)| \leq \gamma(0)$.
- (iii) $\overline{\gamma(-h)} = \overline{Cov(X_{t-h}, X_t)} = \overline{E((X_t - E(X_t))(X_{t-h} - E(X_{t-h})))} = E((X_t - E(X_t))(X_{t-h} - E(X_{t-h}))) = Cov(X_t, X_{t-h}) = Cov(X_{t+h}, X_t) = \gamma(h)$.

La función de autocorrelación mide el nivel de asociación lineal entre las observaciones de dos instantes del proceso estacionario en función de la separación que hay entre ellos en el tiempo. Otra medida de asociación que se suele utilizar es la función de autocorrelación parcial, que mide la asociación entre las variables una vez que hemos eliminado la dependencia respecto de las variables intermedias. En términos de proyecciones sobre subespacios lineales, la *función de autocorrelación parcial* (PACF o FAP) se define como

$$\alpha(1) = Corr(X_2, X_1) = \rho(1),$$

$$\alpha(k) = Corr(X_{k+1} - P_{\overline{sp}\{1, X_2, \dots, X_k\}} X_{k+1}, X_1 - P_{\overline{sp}\{1, X_2, \dots, X_k\}} X_1), \quad k \geq 2.$$

El valor $\alpha(k)$ es conocido como la *autocorrelación parcial en el retardo k* . A continuación veremos otra definición equivalente de esta función que nos proporcionará un método más fácil para calcular estos coeficientes. Supongamos que ϕ_{kj} , $j = 1, \dots, k$, $k = 1, 2, \dots$, son los coeficientes en la representación

$$P_{\overline{sp}\{X_1, \dots, X_k\}} X_{k+1} = \phi_{k1} X_k + \phi_{k2} X_{k-1} + \dots + \phi_{kk} X_1.$$

Por ortogonalidad se satisfacen las ecuaciones

$$\langle X_{k+1} - P_{\overline{\text{span}}\{X_1, \dots, X_k\}} X_{k+1}, X_j \rangle = 0, \quad j = k, \dots, 1,$$

a partir de las cuales obtenemos, por la linealidad del producto interno, las llamadas *ecuaciones de Yule-Walker*, que veremos con más detalle en el Capítulo 3,

$$\sum_{i=1}^k \phi_{ki} \gamma(i-j) = \gamma(j), \quad j = 1, \dots, k,$$

o en forma matricial

$$\Gamma_k \Phi_k = \gamma_k,$$

es decir,

$$\begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \cdots & \gamma(k-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \cdots & \gamma(k-2) \\ \gamma(2) & \gamma(1) & \gamma(0) & \cdots & \gamma(k-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma(k-1) & \gamma(k-2) & \gamma(k-3) & \cdots & \gamma(0) \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \phi_{k3} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \gamma(3) \\ \vdots \\ \gamma(k) \end{bmatrix}.$$

El Teorema de la Proyección asegura que existe al menos una solución. Además, esta solución es única si Γ_k es no singular, en cuyo caso la solución es $\Phi_k = \Gamma_k^{-1} \gamma_k$. Equivalentemente, dividiendo entre $\gamma(0)$ ambos lados de las ecuaciones, podemos resolver el sistema de ecuaciones

$$R_k \Phi_k = \rho_k, \quad (1.1)$$

es decir,

$$\begin{bmatrix} 1 & \rho(1) & \rho(2) & \cdots & \rho(k-1) \\ \rho(1) & 1 & \rho(1) & \cdots & \rho(k-2) \\ \rho(2) & \rho(1) & 1 & \cdots & \rho(k-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho(k-1) & \rho(k-2) & \rho(k-3) & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \phi_{k3} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \rho(3) \\ \vdots \\ \rho(k) \end{bmatrix},$$

una vez conocidas las autocorrelaciones. Entonces la autocorrelación parcial en el retardo k es

$$\alpha(k) = \phi_{kk}, \quad k \geq 1.$$

Demostremos la equivalencia de estas dos definiciones de la función de autocorrelación parcial cuando veamos el algoritmo de Durbin-Levinson (Capítulo 3).

1.2. Procesos lineales

Un modelo de uso frecuente en el análisis de series temporales es el llamado 'ruido blanco'. Formalmente $\{Z_t\}_{t \in \mathbb{Z}}$ es un *ruido blanco* de media cero y varianza σ^2 si Z_t son variables aleatorias independientes e igualmente distribuidas $N(0, \sigma^2)$. Usaremos la notación $\{Z_t\}_{t \in \mathbb{Z}} \sim WN(0, \sigma^2)$ para representar esta situación (las iniciales vienen de 'white noise', ruido blanco en inglés). La terminología procede del campo del análisis de señales. El ruido blanco representa la ausencia de estructura o la máxima impredecibilidad. El ruido blanco no es un modelo de gran interés en sí, pero permite construir muchos modelos interesantes a partir de él. Entre ellos juegan un papel fundamental los procesos lineales, que incluyen la clase de los modelos autorregresivos de media móvil (ARMA). En el resto de la sección asumiremos que $\{Z_t\}_{t \in \mathbb{Z}}$ es un ruido blanco dado y estudiaremos la construcción de otros procesos estacionarios a partir de él.

Un proceso $\{X_t\}_{t \in \mathbb{Z}}$ es *lineal* si se puede representar como

$$X_t = \sum_{j \in \mathbb{Z}} \psi_j Z_{t-j}$$

para todo t , donde $\{Z_t\} \sim WN(0, \sigma^2)$ y $\{\psi_j\}$ es una sucesión de constantes con $\sum_{j \in \mathbb{Z}} |\psi_j| < \infty$. Observamos que la condición $\sum_{j \in \mathbb{Z}} |\psi_j| < \infty$ implica que $\sum_{j \in \mathbb{Z}} |\psi_j|^2 < \infty$. En particular, la expresión $\sum_{j \in \mathbb{Z}} \psi_j Z_{t-j} < \infty$ converge en $L^2(\Omega, F, P)$ y si $X_t = \sum_{j \in \mathbb{Z}} \psi_j Z_{t-j}$ entonces $EX_t = 0$ y $E|X_t|^2 = \sigma^2 \sum_{j \in \mathbb{Z}} |\psi_j|^2 < \infty$. Un proceso lineal de este tipo necesariamente es un proceso estacionario Gaussiano. Si se suprime la hipótesis de normalidad en el ruido blanco las conclusiones son las mismas, salvo que el proceso resultante será estacionario (en el sentido débil) pero no Gaussiano.

Si el proceso lineal depende solo del ruido blanco en instantes anteriores, es decir, si

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j},$$

con $\psi_j = 0$ para todo $j < 0$, se dice que el proceso es *de media móvil* (MA(∞), de su nombre en inglés 'moving average'). Su función de autocovarianza es

$$\begin{aligned} \gamma(h) &= Cov(X_{t+h}, X_t) = Cov\left(\sum_{k=0}^{\infty} \psi_k Z_{t+h-k}, \sum_{j=0}^{\infty} \psi_j Z_{t-j}\right) \\ &= \sum_{k,j=0}^{\infty} \psi_k \psi_j Cov(Z_{t+h-k}, Z_{t-j}) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}. \end{aligned}$$

Además, si solo depende de q instantes anteriores se dice que el proceso $\{X_t\}$ es un proceso *de media móvil de orden q* (MA(q)), es decir,

$$X_t = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} \quad (1.2)$$

donde $\{Z_t\} \sim WN(0, \sigma^2)$ y $\theta_1, \dots, \theta_q$ son constantes. Es evidente que si el proceso X_t se define por (1.2) entonces es un proceso estacionario centrado y su función de autocovarianza es

$$\gamma(h) = \sigma^2 \sum_{j=0}^{q-|h|} \theta_j \theta_{j+|h|}, \quad (1.3)$$

donde consideramos $\theta_0 = 1$, y deducimos que $\gamma(h) = 0$ si $|h| > q$.

Se puede caracterizar cuando un proceso es un MA(q) mirando únicamente la media y la función de autocovarianza, y esto lo recogemos en la siguiente proposición.

Proposición 1.2.1. Si $\{X_t\}$ es un proceso estacionario de media cero con función de autocovarianza $\gamma(\cdot)$ tal que $\gamma(h) = 0$ para $|h| > q$ y $\gamma(q) \neq 0$, entonces $\{X_t\}$ es un proceso MA(q), es decir, existe un proceso de ruido blanco $\{Z_t\}$ tal que $X_t = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}$.

Demostración. Sea, para cada t , $M_t = \overline{\text{sp}}\{X_s, -\infty < s \leq t\}$ subespacio de L^2 y

$$Z_t = X_t - P_{M_{t-1}} X_t. \quad (1.4)$$

Claramente, $Z_t \in M_t$ y por definición de $P_{M_{t-1}}$, $Z_t \in M_{t-1}^\perp$. Por lo tanto si $s < t$, $Z_s \in M_s \subset M_{t-1}$ y entonces $E Z_s Z_t = 0$. Además

$$P_{\overline{\text{sp}}\{X_s, t-n \leq s \leq t-1\}} X_t \xrightarrow{L^2} P_{M_{t-1}} X_t$$

cuando $n \rightarrow \infty$ luego

$$\begin{aligned} \|Z_{t+1}\| &= \|X_{t+1} - P_{M_t} X_{t+1}\| = \lim_{n \rightarrow \infty} \|X_{t+1} - P_{\overline{\text{sp}}\{X_s, t+1-n \leq s \leq t\}} X_{t+1}\| \\ &= \lim_{n \rightarrow \infty} \|X_t - P_{\overline{\text{sp}}\{X_s, t-n \leq s \leq t-1\}} X_t\| = \|X_t - P_{M_{t-1}} X_t\| = \|Z_t\| \end{aligned}$$

por ser $\{X_t\}$ estacionaria y por la continuidad de la norma de L^2 . Definiendo $\sigma^2 = \|Z_t\|^2$, concluimos que $\{Z_t\} \sim WN(0, \sigma^2)$.

Ahora por (1.4) tenemos que $M_{t-1} = \overline{\text{sp}}\{X_s, s < t-1, Z_{t-1}\} = \overline{\text{sp}}\{X_s, s < t-q, Z_{t-q}, \dots, Z_{t-1}\}$ y entonces podemos descomponer M_{t-1} en dos subespacios

ortogonales M_{t-q-1} y $\overline{sp}\{Z_{t-q}, \dots, Z_{t-1}\}$. Por hipótesis $\gamma(h) = 0$ para $|h| > q$, luego $X_t \perp M_{t-q-1}$ y por las propiedades de la proyección tenemos que

$$\begin{aligned} P_{M_{t-1}}X_t &= P_{M_{t-q-1}}X_t + P_{\overline{sp}\{Z_{t-q}, \dots, Z_{t-1}\}}X_t \\ &= 0 + \sigma^{-2}E(X_t Z_{t-1})Z_{t-1} + \dots + \sigma^{-2}E(X_t Z_{t-q})Z_{t-q} \\ &= \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} \end{aligned}$$

donde $\theta_j = \sigma^{-2}E(X_t Z_{t-j})$, que por estacionariedad es independiente de t para $j = 1, \dots, q$. Sustituyendo $P_{M_{t-1}}X_t$ en (1.4) concluimos la demostración. \square

En la demostración se ha usado que las proyecciones sobre subespacios crecientes convergen hacia la proyección sobre el subespacio generado. Más concretamente, que

$$P_{\overline{sp}\{X_s, t-n \leq s \leq t-1\}}X_t \xrightarrow{L^2} P_{\overline{sp}\{X_s, -\infty < s \leq t-1\}}X_t$$

cuando $n \rightarrow \infty$, si $\{X_t\}$ es un proceso estacionario. Este es un resultado estándar en la teoría de procesos estacionarios. Se puede encontrar por ejemplo en [4], página 75.

Otra clase importante de procesos estacionarios es la de procesos autorregresivos. $\{X_t\}$ es un proceso *autoregresivo de orden p* (AR(p), de su nombre en inglés 'autoregressive') si

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t \quad (1.5)$$

donde $\{Z_t\} \sim WN(0, \sigma^2)$ y ϕ_1, \dots, ϕ_p son constantes.

Los procesos autorregresivos son muy interesantes desde el punto de vista de la predicción. De la definición se deduce que si $\{X_t\}_{t \in \mathbb{Z}}$ es un AR(p), entonces, conocidos X_{t-1}, \dots, X_{t-p} la observación X_t es una función lineal de los valores observados en instantes anteriores más una innovación aleatoria. A diferencia de lo descrito en el caso de procesos de media móvil, la existencia de procesos AR(p) estacionarios, es decir, de soluciones estacionarias de (1.5) requiere condiciones algo más complicadas.

Ejemplo 1.2.2. Supongamos que $\{X_t\}$ es un proceso AR(1), es decir,

$$X_t = \phi_1 X_{t-1} + Z_t. \quad (1.6)$$

De la misma forma tenemos que $X_{t-1} = \phi_1 X_{t-2} + Z_{t-1}$ y sustituyendo en la ecuación anterior tenemos que

$$X_t = \phi_1(\phi_1 X_{t-2} + Z_{t-1}) + Z_t = Z_t + \phi_1 Z_{t-1} + \phi_1^2 X_{t-2}.$$

Iterando k veces obtenemos

$$X_t = Z_t + \phi_1 Z_{t-1} + \phi_1^2 Z_{t-2} + \dots + \phi_1^k Z_{t-k} + \phi_1^{k+1} X_{t-k-1}.$$

Ahora si $|\phi_1| < 1$ y $\{X_t\}$ es estacionario entonces $\|X_t\|^2 = E(X_t^2)$ es constante y

$$\left\| X_t - \sum_{j=0}^k \phi_1^j Z_{t-j} \right\|^2 = \phi_1^{2k+2} \|X_{t-k-1}\|^2 \rightarrow 0 \text{ cuando } k \rightarrow \infty.$$

Entonces tenemos que

$$X_t = \sum_{j=0}^{\infty} \phi_1^j Z_{t-j}. \quad (1.7)$$

La condición $|\phi_1| < 1$ garantiza que $\sum_{j=0}^{\infty} \phi_1^j Z_{t-j}$ es convergente en media cuadrática. Concluimos que el proceso definido por (1.7) es un AR(1), es decir, es solución estacionaria de la ecuación (1.6). Este proceso es también un MA(∞), es decir, es un proceso lineal. Por el contrario, si $|\phi_1| > 1$ el comportamiento de la serie (1.2.2) es explosivo, es decir, no converge en L^2 . Sin embargo ahora podemos escribir $X_t = \phi_1^{-1}(-Z_{t+1} + X_{t+1}) = -\phi_1^{-1}Z_{t+1} - \phi_1^{-2}Z_{t+2} + \phi_1^{-2}X_{t+2}$. Iterando concluimos que

$$X_t = -\sum_{j=1}^{\infty} \phi_1^{-j} Z_{t+j}.$$

La expresión anterior define también un proceso estacionario, solución de (1.6). Sin embargo, este tipo de proceso es menos interesante desde el punto de vista de la predicción, puesto que depende de los valores ‘futuros’ del ruido blanco.

La clase de modelos autorregresivos de media móvil (ARMA) que definimos a continuación es más amplia y flexible ya que está formada por dos partes, una parte autorregresiva (AR) y otra de media móvil (MA), y es útil para representar una gran variedad de series utilizando menos parámetros que un proceso AR o MA por sí solo.

Definición 1.2.3. El proceso $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$ es un proceso *autoregresivo de media móvil* (ARMA(p, q)) si $\{X_t\}$ es estacionario y si para cada t

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}, \quad (1.8)$$

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

Además, $\{X_t\}$ es un proceso ARMA(p, q) con media μ si $\{X_t - \mu\}$ es un proceso ARMA(p, q).

Las ecuaciones anteriores se pueden escribir de manera más compacta a partir del operador de retardo B . Este operador B funciona de la siguiente manera: si lo aplicamos una vez a X_t obtenemos $BX_t = X_{t-1}$ y aplicándolo j veces, $j = \pm 2, \pm 3, \dots$, obtenemos $B^j X_t = X_{t-j}$. Otras propiedades son (siendo a y b constantes): $Ba = a$; $B(aX_t) = aBX_t$; y B lineal. Con esta notación, las ecuaciones se convierten en $\phi(B)X_t = \theta(B)Z_t$, $t = 0, \pm 1, \pm 2, \dots$, donde $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ es un polinomio de grado p (llamado *autorregresivo*) y $\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$ es un polinomio de grado q (llamado *de media móvil*). Además, si $\phi(z) \equiv 1$ entonces $X_t = \theta(B)Z_t$, $t = 0, \pm 1, \pm 2, \dots$, es un proceso MA(q), y si $\theta(z) \equiv 1$ entonces $\phi(B)X_t = Z_t$, $t = 0, \pm 1, \pm 2, \dots$, es un proceso AR(p). Esta notación se empleará de forma reiterada en el resto de esta memoria.

En la literatura de series temporales se suele emplear el término causal para los procesos MA(∞), es decir, un proceso ARMA(p, q) definido por las ecuaciones $\phi(B)X_t = \theta(B)Z_t$ es *causal* (o más específicamente una función causal de $\{Z_t\}$) si existe una secuencia de constantes $\{\psi_j\}$ tal que $\sum_{j=0}^{\infty} |\psi_j| < \infty$ y

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad t = 0, \pm 1, \dots \quad (1.9)$$

Tal como se observó en el Ejemplo 1.2.2, los procesos interesantes son, precisamente, los procesos causales: si un proceso no es causal entonces su valor en el instante t depende de valores futuros del ruido blanco y el modelo resultante no es apropiado para la predicción.

La siguiente proposición nos da una condición necesaria y suficiente para que un proceso ARMA sea causal y también una representación de X_t en términos de los antecedentes $\{Z_s, s \leq t\}$.

Proposición 1.2.4. Sea $\{X_t\}$ un proceso ARMA(p, q) donde los polinomios $\phi(\cdot)$ y $\theta(\cdot)$ no tienen ceros comunes. Entonces $\{X_t\}$ es causal si y sólo si las raíces de $\phi(z)$ están fuera del círculo unidad, es decir, $\phi(z) \neq 0$ para todo $z \in \mathbb{C}$ tal que $|z| \leq 1$. Los coeficientes $\{\psi_j\}$ de (1.9) están determinados por

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, \quad |z| \leq 1. \quad (1.10)$$

Demostración. Probamos que la condición es suficiente. El recíproco se puede encontrar en [4] (Teorema 3.1.1). Como ϕ no tiene raíces en la bola unidad cerrada, existe un $\varepsilon > 0$ tal que la función $\psi(z) = \frac{\theta(z)}{\phi(z)}$ es holomorfa en $B(0, 1 + \varepsilon)$ y, por lo tanto, admite el desarrollo (1.10). El proceso definido por

$Y_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ es un proceso estacionario, de acuerdo con los resultados anteriores. Podemos aplicar el mismo argumento a $\xi(z) = \frac{1}{\phi(z)}$ y a los procesos $\phi(B)X_t$ y $\theta(B)Z_t$. Si X_t satisface la ecuación (1.8) entonces $\xi(B)\theta(B)Z_t$ es estacionario (por serlo Z_t) y

$$X_t = \xi(B)\phi(B)X_t = \xi(B)\theta(B)Z_t = \psi(B)Z_t = Y_t.$$

Concluimos que X_t es causal. □

En el ejemplo 1.2.2 el polinomio autorregresivo es $\phi(z) = 1 - \phi_1 z$ cuya única raíz es $\frac{1}{\phi_1}$. En este caso, la condición de la proposición anterior es equivalente a la condición $|\phi_1| < 1$ que ya vimos que garantizaba que X_t fuera causal.

Una interpretación equivalente de la Proposición 1.2.4 es que las ecuaciones (1.8) tienen una única solución estacionaria causal cuando ϕ y θ no tiene ceros comunes y θ no tiene ceros en la bola unidad cerrada. Entonces la única solución causal está dada por

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j},$$

donde ψ_j son los coeficientes en el desarrollo (1.10).

Por analogía al concepto de causalidad, se dice que un proceso ARMA(p, q) definido por las ecuaciones $\phi(B)X_t = \theta(B)Z_t$ es *invertible* (o más específicamente una función invertible de $\{Z_t\}$) si existe una secuencia de constantes $\{\pi_j\}$ tal que $\sum_{j=0}^{\infty} |\pi_j| < \infty$ y

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \quad t = 0, \pm 1, \dots \quad (1.11)$$

En ocasiones se usa la terminología AR(∞) para referirse a estos procesos.

La condición de proceso invertible se puede caracterizar de forma análoga a la de proceso causal. Recogemos el resultado a continuación (esta proposición es el Teorema 3.1.2 en [4]).

Proposición 1.2.5. Sea $\{X_t\}$ un proceso ARMA(p, q) donde los polinomios $\phi(\cdot)$ y $\theta(\cdot)$ no tienen ceros comunes. Entonces $\{X_t\}$ es invertible si y sólo si las raíces de $\theta(z)$ están fuera del círculo unidad, es decir, $\theta(z) \neq 0$ para todo $z \in \mathbb{C}$ tal que $|z| \leq 1$. Los coeficientes $\{\pi_j\}$ de (1.11) están determinados por

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

Ahora vamos a caracterizar las funciones de autocovarianza de los procesos ARMA. Como hemos visto antes, si $\{X_t\}$ es un proceso ARMA(p, q) causal definido por las ecuaciones $\phi(B)X_t = \theta(B)Z_t$ entonces su función de autocovarianza es

$$\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|}. \quad (1.12)$$

donde $\psi(z) = \sum_{j \in \mathbb{Z}} \psi_j z^j = \frac{\theta(z)}{\phi(z)}$. Para determinar los coeficientes ψ_j reescribimos $\psi(z)\phi(z) = \theta(z)$ e igualamos los coeficientes de z^j . Los primeros valores son

$$\begin{aligned} \psi_0 &= 1 \\ \psi_1 - \phi_1 \psi_0 &= \theta_1 \\ \psi_2 - \phi_1 \psi_1 - \phi_2 \psi_0 &= \theta_2 \\ \psi_3 - \phi_1 \psi_2 - \phi_2 \psi_1 - \phi_3 \psi_0 &= \theta_3 \\ &\vdots \end{aligned}$$

donde suponemos $\phi_j = 0$ para $j > p$ y $\theta_j = 0$ para $j > q$. Las constantes ψ_j satisfacen la ecuación en diferencias homogénea

$$\psi_j - \sum_{k=1}^p \phi_k \psi_{j-k} = 0, \quad j \geq \max(p, q+1), \quad (1.13)$$

con condiciones iniciales

$$\psi_j - \sum_{k=1}^j \phi_k \psi_{j-k} = \theta_j, \quad 0 \leq j < \max(p, q+1). \quad (1.14)$$

La solución general depende de las raíces del polinomio autorregresivo ϕ y la solución específica de las condiciones iniciales. Otra forma de obtener los coeficientes es encontrar la solución general que es

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

donde ξ_i , $i = 1, \dots, k$, son las raíces del polinomio $\phi(\cdot)$ y r_i sus respectivas multiplicidades. Los coeficientes ψ_j , $0 \leq j < \max(p, q+1) - p$, están determinadas por las condiciones iniciales (1.14) y determinan las p constantes α_{ij} . Esta forma de determinar la función de autocovarianza también es válida para los procesos autorregresivos causales (basta tomar $q = 0$).

Para concluir esta sección veamos algunos gráficos de las funciones de autocorrelación simple y parcial de algunos procesos AR, MA y ARMA. La representación gráfica de ambas, en función de los distintos retardos, se denomina *correlograma*.

La tarea de identificar el orden de un proceso autorregresivo a partir de su función de autocorrelación simple es complicada. Estos procesos presentan efectos directos de observaciones separadas por $1, 2, \dots, p$ retardos, y los efectos directos de las observaciones separadas por más de p retardos son nulos. Por esta razón, hemos introducido el concepto de función de autocorrelación parcial. Veamos algunos ejemplos.

Ejemplo 1.2.6 (AR(1)). Sea $\{X_t\}$ un proceso autorregresivo de orden uno definido por las ecuaciones $X_t = \phi_1 X_{t-1} + Z_t$ con $\{Z_t\} \sim WN(0, \sigma^2)$. Primero calcularemos su función de autocovarianza utilizando la ecuación (1.12). En este caso, $\max(p, q+1) = 1$ luego a partir de (1.14) obtenemos $\psi_0 = 1$. Por (1.13), para $j \geq 1$, los coeficientes ψ_j satisfacen la ecuación $\psi_j - \phi_1 \psi_{j-1} = 0$, que tiene solución general $\psi_j = C(\phi_1)^j$. Para hallar una solución particular usamos la condición inicial $\psi_0 = 1$, de donde $C = 1$. Finalmente obtenemos $\psi_j = (\phi_1)^j$ para $j \geq 0$. Ahora utilizando la ecuación (1.12) obtenemos

$$\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} (\phi_1)^j (\phi_1)^{j+|h|} = \sigma^2 \phi_1^{|h|} \sum_{j=0}^{\infty} (\phi_1^2)^j = \phi_1^{|h|} \frac{\sigma^2}{1 - \phi_1^2} = \phi_1^{|h|} \gamma(0),$$

ya que $\phi_1^2 < 1$ por ser $|\phi_1| < 1$ (recordamos que esta última condición garantiza que un proceso autorregresivo sea causal). Entonces la función de autocorrelación es

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \phi_1^{|h|}.$$

Esta función tiene un decrecimiento exponencial (ya que $|\phi_1| < 1$), más lento cuanto mayor sea $|\phi_1|$, pero no nos proporciona un método para identificar el orden del proceso. Calculemos la función de autocorrelación parcial. A partir de las ecuaciones (1.1) obtenemos:

$$\alpha(1) = \phi_{11} = \frac{|\rho(1)|}{|1|} = \rho(1) = \phi_1,$$

$$\alpha(2) = \phi_{22} = \frac{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & \rho(2) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{vmatrix}} = \frac{\begin{vmatrix} 1 & \phi_1 \\ \phi_1 & \phi_1^2 \end{vmatrix}}{\begin{vmatrix} 1 & \phi_1 \\ \phi_1 & 1 \end{vmatrix}} = 0,$$

$$\alpha(3) = \phi_{33} = \frac{\begin{vmatrix} 1 & \rho(1) & \rho(1) \\ \rho(1) & 1 & \rho(2) \\ \rho(2) & \rho(1) & \rho(3) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & \rho(2) \\ \rho(1) & 1 & \rho(1) \\ \rho(2) & \rho(1) & 1 \end{vmatrix}} = \frac{\begin{vmatrix} 1 & \phi_1 & \phi_1 \\ \phi_1 & 1 & \phi_1^2 \\ \phi_1^2 & \phi_1 & \phi_1^3 \end{vmatrix}}{\begin{vmatrix} 1 & \phi_1 & \phi_1^2 \\ \phi_1 & 1 & \phi_1 \\ \phi_1^2 & \phi_1 & 1 \end{vmatrix}} = 0,$$

$$\vdots$$

Hallando las autocorrelaciones parciales en los sucesivos retardos obtenemos que la función de autocorrelación parcial de un proceso AR(1) tiene el primer valor distinto de cero y el resto de valores iguales a cero, es decir,

$$\alpha(1) = \rho(1)$$

y

$$\alpha(h) = 0, \quad h > 1.$$

En general, se deduce que un proceso AR(p) tendrá los p primeros coeficientes de autocorrelación parcial distintos de cero, y, por tanto, en la PACF el número de coeficientes distintos de cero indica el orden del proceso AR.

Las figuras 1.1 y 1.2 resumen los correlogramas de la ACF y la PACF de distintos procesos AR(1). Observamos que si el parámetro ϕ_1 es mayor que cero, los valores de la ACF y de la PACF son mayores o iguales a cero, mientras que si ϕ_1 es menor que cero, los valores de la ACF van alternando el signo mientras que la PACF toma valores negativos o iguales a cero.

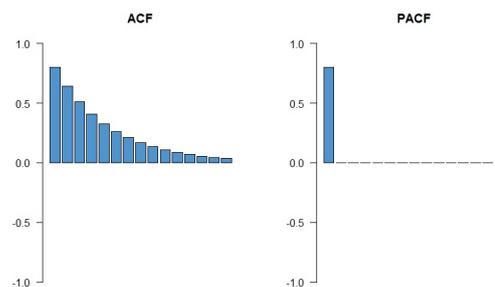
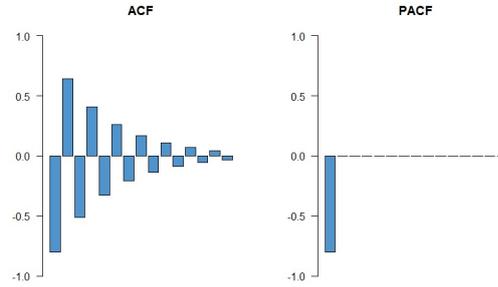


Figura 1.1: $\phi_1 = 0,8$

En el siguiente ejemplo veremos que existe una dualidad entre procesos AR y MA, de forma que la PACF de un MA(q) tiene la estructura de la ACF de un AR(q), y la ACF de un MA(q) tiene la estructura de la PACF de un AR(q).

Figura 1.2: $\phi_1 = -0,8$

Ejemplo 1.2.7. MA(1). Sea $\{X_t\}$ un proceso de media móvil de orden uno definido por las ecuaciones $X_t = Z_t + \theta_1 Z_{t-1}$ con $\{Z_t\} \sim WN(0, \sigma^2)$. La ecuación (1.3) nos dice que la función de autocovarianza es

$$\gamma(h) = \sigma^2 \sum_{j=0}^{1-|h|} \theta_j \theta_{j+|h|},$$

donde consideramos $\theta_0 = 1$. Entonces obtenemos

$$\gamma(0) = \sigma^2(1 + \theta_1^2),$$

$$\gamma(1) = \gamma(-1) = \sigma^2 \theta_1$$

y deducimos que

$$\gamma(h) = 0, \quad |h| > 1.$$

Como consecuencia, la función de autocorrelación está definida por las ecuaciones

$$\rho(1) = \rho(-1) = \frac{\theta_1}{1 + \theta_1^2}$$

y

$$\rho(h) = 0, \quad |h| > 1.$$

Esto quiere decir que un proceso MA(1) solo tiene el primer valor de la función de autocorrelación distinto de cero. En general, podemos identificar el orden de un proceso MA(q) observando el número de valores distintos de cero de su función de autocorrelación. En el caso de procesos MA es la función de autocorrelación parcial la que no nos proporciona información sobre su orden. A partir de las ecuaciones (1.1) obtenemos las autocorrelaciones parciales de un proceso MA(1):

$$\alpha(1) = \phi_{11} = \frac{|\rho(1)|}{|1|} = \rho(1) = \frac{\theta_1}{1 + \theta_1^2},$$

$$\begin{aligned}\alpha(2) &= \phi_{22} = \frac{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & \rho(2) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{vmatrix}} = \frac{\begin{vmatrix} 1 & \frac{\theta_1}{1+\theta_1^2} \\ \frac{\theta_1}{1+\theta_1^2} & 0 \end{vmatrix}}{\begin{vmatrix} 1 & \frac{\theta_1}{1+\theta_1^2} \\ \frac{\theta_1}{1+\theta_1^2} & 1 \end{vmatrix}} = \frac{-\left(\frac{\theta_1}{1+\theta_1^2}\right)^2}{1 - \left(\frac{\theta_1}{1+\theta_1^2}\right)^2} \\ &= \frac{-\theta_1^2}{1 + \theta^2 + \theta^4},\end{aligned}$$

$$\begin{aligned}\alpha(3) &= \phi_{33} = \frac{\begin{vmatrix} 1 & \rho(1) & \rho(1) \\ \rho(1) & 1 & \rho(2) \\ \rho(2) & \rho(1) & \rho(3) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & \rho(2) \\ \rho(1) & 1 & \rho(1) \\ \rho(2) & \rho(1) & 1 \end{vmatrix}} = \frac{\begin{vmatrix} 1 & \frac{\theta_1}{1+\theta_1^2} & \frac{\theta_1}{1+\theta_1^2} \\ \frac{\theta_1}{1+\theta_1^2} & 1 & 0 \\ 0 & \frac{\theta_1}{1+\theta_1^2} & 0 \end{vmatrix}}{\begin{vmatrix} 1 & \frac{\theta_1}{1+\theta_1^2} & 0 \\ \frac{\theta_1}{1+\theta_1^2} & 1 & \frac{\theta_1}{1+\theta_1^2} \\ 0 & \frac{\theta_1}{1+\theta_1^2} & 1 \end{vmatrix}} \\ &= \frac{\left(\frac{\theta_1}{1+\theta_1^2}\right)^3}{1 - 2\left(\frac{\theta_1}{1+\theta_1^2}\right)^2} = \frac{\theta_1^3}{1 + \theta_1^2 + \theta_1^4 + \theta_1^6},\end{aligned}$$

⋮

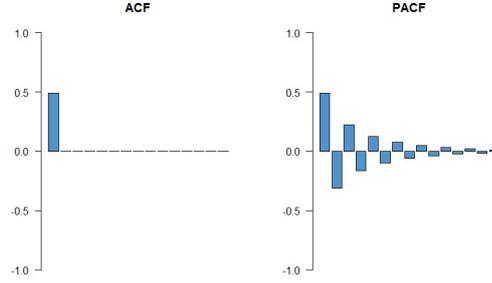
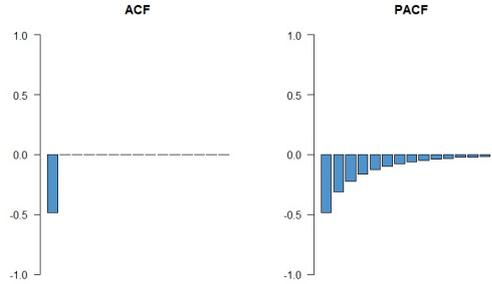
Entonces obtenemos que la función de autocorrelación parcial es

$$\alpha(h) = \frac{(-\theta_1)^h}{\sum_{j=0}^h \theta_1^{2j}},$$

que es distinta de cero para todo valor de h y decreciente, luego no nos proporciona información sobre el orden.

Las figuras 1.3 y 1.4 presentan los correlogramas de estas funciones para distintos procesos MA(1). Analizando los gráficos observamos que si θ_1 es mayor que cero entonces los valores de la ACF son mayores o iguales que cero y los valores de la PACF alternan el signo, mientras que si el parámetro θ_1 es menor que cero, ambas funciones toman valores menores o iguales que cero.

En el caso de procesos ARMA, la ACF y la PACF dependen de sus propiedades AR y MA, por lo que en la práctica será difícil identificar su orden. A pesar de esto, son los más frecuentes en la práctica. Una razón es que al sumar procesos AR resulta un proceso ARMA entonces cuando observemos procesos que sean suma de otros y alguno tenga estructura de proceso AR es esperable observar procesos ARMA. En cambio, con la suma de procesos MA independientes obtenemos nuevos procesos MA.

Figura 1.3: $\theta_1 = 0,8$ Figura 1.4: $\theta_1 = -0,8$

Ejemplo 1.2.8. ARMA(1,1). Sea $\{X_t\}$ un proceso autorregresivo de media móvil de órdenes $p = 1$ y $q = 1$ definido por las ecuaciones $X_t - \phi_1 X_{t-1} = Z_t + \theta_1 Z_{t-1}$ con $\{Z_t\} \sim WN(0, \sigma^2)$. Primero calcularemos su función de autocovarianza utilizando la ecuación (1.12). En este caso, $\max(p, q + 1) = 2$ luego a partir de (1.14) obtenemos $\psi_0 = 1$ y $\psi_1 = \phi_1 + \theta_1$. Por (1.13), para $j \geq 2$, los coeficientes ψ_j satisfacen la ecuación $\psi_j - \phi_1 \psi_{j-1} = 0$, que tiene solución general $\psi_j = C(\phi_1)^j$. Para hallar una solución particular usamos la condición inicial $\psi_1 = \phi_1 + \theta_1$, de donde $C = \frac{\phi_1 + \theta_1}{\phi_1}$. Finalmente obtenemos $\psi_j = (\phi_1 + \theta_1)(\phi_1)^{j-1}$ para $j \geq 1$. Ahora utilizando la ecuación (1.12) obtenemos

$$\begin{aligned}
 \gamma(0) &= \sigma^2 \sum_{j=0}^{\infty} \psi_j^2 = \sigma^2 \left[1 + (\phi_1 + \theta_1)^2 \sum_{j=1}^{\infty} \phi_1^{2j-2} \right] \\
 &= \sigma^2 \left[1 + \frac{(\phi_1 + \theta_1)^2}{\phi_1^2} \left(\frac{1}{1 - \phi_1^2} - 1 \right) \right] = \sigma^2 \left[1 + \frac{(\phi_1 + \theta_1)^2}{1 - \phi_1^2} \right] \\
 &= \sigma^2 \left(\frac{1 + \theta_1^2 + 2\phi_1\theta_1}{1 - \phi_1^2} \right)
 \end{aligned}$$

y

$$\begin{aligned}
\gamma(h) &= \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|} \\
&= \sigma^2 \left[(\phi_1 + \theta_1) \phi_1^{|h|-1} + (\phi_1 + \theta_1)^2 \sum_{j=1}^{\infty} \phi_1^{j-1} \phi_1^{j+|h|-1} \right] \\
&= \sigma^2 \left[(\phi_1 + \theta_1) \phi_1^{|h|-1} + (\phi_1 + \theta_1)^2 \phi_1^{|h|-2} \left(\frac{1}{1 - \phi_1^2} - 1 \right) \right] \\
&= \sigma^2 \left[(\phi_1 + \theta_1) \phi_1^{|h|-1} + \frac{(\phi_1 + \theta_1)^2 \phi_1^{|h|}}{1 - \phi_1^2} \right] \\
&= \sigma^2 \phi_1^{|h|-1} \left[\phi_1 + \theta_1 + \frac{(\phi_1 + \theta_1)^2 \phi_1}{1 - \phi_1^2} \right] = \sigma^2 \phi_1^{|h|-1} \left(\frac{\phi_1 \gamma(0)}{\sigma^2} + \theta_1 \right) \\
&= \phi_1^{|h|-1} (\phi_1 \gamma(0) + \theta_1 \sigma^2)
\end{aligned}$$

para $|h| \geq 1$, donde hemos utilizado que $\phi_1^2 < 1$ por ser $|\phi_1| < 1$. Escrito de forma más compacta tenemos que la función de autocovarianza es

$$\gamma(0) = \sigma^2 \left(\frac{1 + \theta_1^2 + 2\phi_1\theta_1}{1 - \phi_1^2} \right),$$

$$\gamma(1) = \phi_1 \gamma(0) + \theta_1 \sigma^2$$

y para $|h| > 1$

$$\gamma(h) = \phi_1^{|h|-1} \gamma(1).$$

A partir de esta función obtenemos la función de autocorrelación. Para $|h| = 1$ obtenemos

$$\rho(1) = \rho(-1) = \frac{\gamma(1)}{\gamma(0)} = \phi_1 + \frac{\theta_1(1 - \phi_1^2)}{1 + \theta_1^2 + 2\phi_1\theta_1} = \frac{(1 + \phi_1\theta_1)(\phi_1 + \theta_1)}{1 + \theta_1^2 + 2\phi_1\theta_1}$$

y para $|h| > 1$

$$\rho(h) = \phi_1^{|h|-1} \rho(1),$$

función cuyo decrecimiento está dictado por ϕ_1 , es decir, por la parte autorregresiva. Calculemos ahora la función de autocorrelación parcial. A partir de las ecuaciones (1.1) obtenemos:

$$\alpha(1) = \phi_{11} = \frac{|\rho(1)|}{|1|} = \rho(1) = \frac{(1 + \phi_1\theta_1)(\phi_1 + \theta_1)}{1 + \theta_1^2 + 2\phi_1\theta_1},$$

$$\begin{aligned}\alpha(2) &= \phi_{22} = \frac{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & \rho(2) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{vmatrix}} = \frac{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & \phi_1\rho(1) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{vmatrix}} = \frac{\phi_1\rho(1) - \rho(1)^2}{1 - \rho(1)^2} \\ &= \frac{\phi_1 - \rho(1)}{1 - \rho(1)^2}\rho(1),\end{aligned}$$

$$\begin{aligned}\alpha(3) &= \phi_{33} = \frac{\begin{vmatrix} 1 & \rho(1) & \rho(1) \\ \rho(1) & 1 & \rho(2) \\ \rho(2) & \rho(1) & \rho(3) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & \rho(2) \\ \rho(1) & 1 & \rho(1) \\ \rho(2) & \rho(1) & 1 \end{vmatrix}} = \frac{\begin{vmatrix} 1 & \rho(1) & \rho(1) \\ \rho(1) & 1 & \phi_1\rho(1) \\ \phi_1\rho(1) & \rho(1) & \phi_1^2\rho(1) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & \phi_1\rho(1) \\ \rho(1) & 1 & \rho(1) \\ \phi_1\rho(1) & \rho(1) & 1 \end{vmatrix}} \\ &= \frac{(\phi_1 - \rho(1))^2}{1 + \phi_1\rho(1)^3 - \phi_1^2\rho(1)^2 - 2\rho(1)^2}\rho(1),\end{aligned}$$

⋮

Aunque con estos resultados no se aprecia claramente, la función de autocorrelación parcial es decreciente, y lo veremos en los correlogramas de los siguientes ejemplos de distintos procesos ARMA(1,1) representados en las figuras 1.5, 1.6, 1.7, 1.8, 1.9 y 1.10. También está representado el correlograma de la ACF. Por analogía con los procesos AR y MA podemos observar en los gráficos que las funciones ACF y PACF toman valores positivos y negativos dependiendo del signo de los parámetros ϕ_1 y θ_1 . También se tiene en cuenta si $\phi_1 < \theta_1$ o $\phi_1 > \theta_1$.

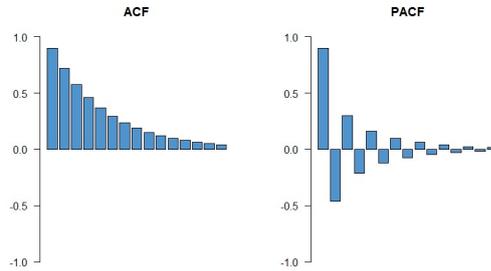


Figura 1.5: $\phi_1 = \theta_1 = 0,8$

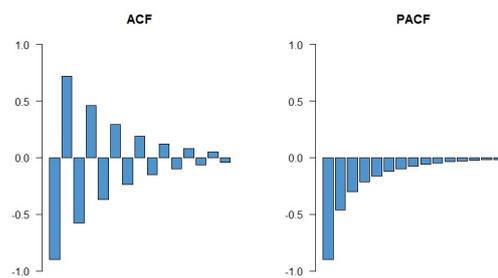


Figura 1.6: $\phi_1 = \theta_1 = -0,8$

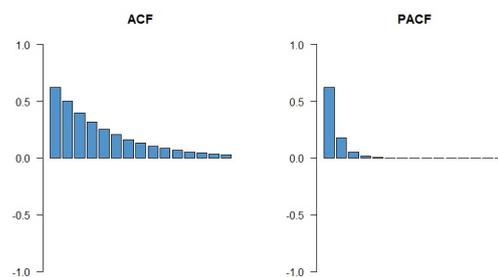


Figura 1.7: $\phi_1 = 0,8; \theta_1 = -0,3$

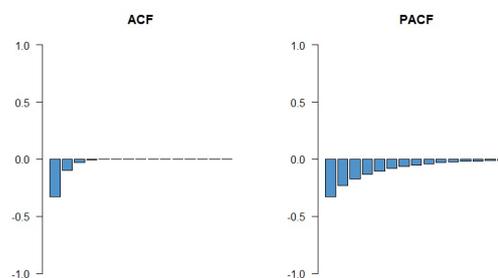


Figura 1.8: $\phi_1 = 0,3; \theta_1 = -0,8$

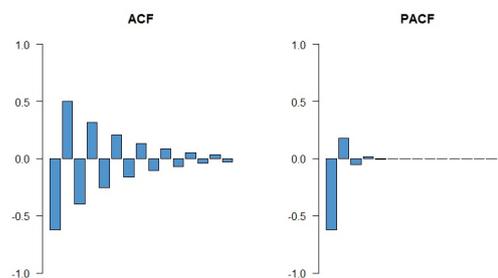
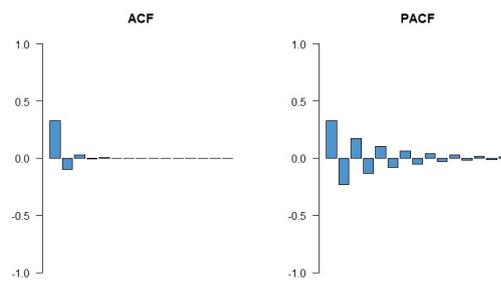


Figura 1.9: $\phi_1 = -0,8; \theta_1 = 0,3$

Figura 1.10: $\phi_1 = -0,3; \theta_1 = 0,8$

Capítulo 2

Teoría espectral

El análisis espectral es una herramienta muy útil en la fase de diagnóstico del proceso de ajuste estacional, para detectar la presencia de componentes estacionales en la serie ajustada.

La *representación espectral* de un proceso $\{X_t\}$ es adecuada para sacar a la luz componentes periódicas ya que consiste en expresar el proceso como una combinación lineal de funciones trigonométricas o exponenciales complejas, ponderadas por coeficientes aleatorios, es decir,

$$X_t = \sum_{j=1}^n A(\lambda_j) e^{it\lambda_j} \quad (2.1)$$

con $-\pi < \lambda_1 < \lambda_2 < \dots < \lambda_n = \pi$ y $A(\lambda_1), \dots, A(\lambda_n)$ coeficientes aleatorios incorrelados con valores complejos tal que

$$E(A(\lambda_j)) = 0 \quad \text{y} \quad E(A(\lambda_j) \overline{A(\lambda_j)}) = \sigma_j^2, \quad j = 1, \dots, n.$$

Si un proceso admite la representación (2.1) y si además $A(\lambda_j)$ son coeficientes reales para todo $j = 1, \dots, n$ y se cumplen las igualdades $\lambda_j = \lambda_{n-j}$ y $A(\lambda_j) = \overline{A(\lambda_{n-j})}$ para todo $j = 1, \dots, n-1$ entonces $\{X_t\}$ es un proceso real.

Además, un proceso X_t con la representación espectral (2.1) es estacionario en sentido débil. Para comprobarlo, probaremos que EX_t y $E(X_{t+h} \overline{X}_t)$ son independientes de t y que $E|X_t|^2 < \infty$.

Por una parte,

$$EX_t = E\left(\sum_{j=1}^n A(\lambda_j) e^{it\lambda_j}\right) = \sum_{j=1}^n E(A(\lambda_j)) E(e^{it\lambda_j}) = 0$$

por la linealidad de la esperanza y $E(A(\lambda_j)) = 0, j = 1, \dots, n$.
Por otra parte,

$$\begin{aligned}
E(X_{t+h}\overline{X}_t) &= E\left(\sum_{j=1}^n A(\lambda_j)e^{i(t+h)\lambda_j} \cdot \overline{\sum_{k=1}^n A(\lambda_k)e^{it\lambda_k}}\right) \\
&= E\left(\sum_{j=1}^n A(\lambda_j)e^{i(t+h)\lambda_j} \cdot \sum_{k=1}^n \overline{A(\lambda_k)}e^{-it\lambda_k}\right) \\
&= E\left(\sum_{j=1}^n \sum_{k=1}^n A(\lambda_j)\overline{A(\lambda_k)}e^{ih\lambda_j}e^{it(\lambda_j-\lambda_k)}\right) \\
&= \sum_{j=1}^n \sum_{k=1}^n E(A(\lambda_j)\overline{A(\lambda_k)})e^{ih\lambda_j}E(e^{it(\lambda_j-\lambda_k)}) \\
&= \sum_{j=1}^n E(A(\lambda_j)\overline{A(\lambda_j)})e^{ih\lambda_j} = \sum_{j=1}^n \sigma_j^2 e^{ih\lambda_j}
\end{aligned}$$

por la linealidad de la esperanza y $E(A(\lambda_j)\overline{A(\lambda_j)}) = \sigma_j^2, j = 1, \dots, n$. En particular, si tomamos $h = 0$ en la ecuación anterior tenemos que

$$E|X_t|^2 = E(X_t\overline{X}_t) = \sum_{j=1}^n \sigma_j^2 < \infty,$$

y con esto concluimos el argumento.

Si los coeficientes de la representación espectral (2.1) fueran Gaussianos entonces el proceso sería Gaussiano, y la estacionariedad se daría en sentido fuerte.

Un proceso con una representación como la de (2.1) admite una representación particular de la función de autocovarianza a la que se denomina *representación espectral de la función de autocovarianza*. De acuerdo con el cálculo anterior,

$$\begin{aligned}
\gamma(h) &= E(X_h\overline{X}_0) - E(X_h)E(\overline{X}_0) = E(X_h\overline{X}_0) \\
&= E\left(\sum_{j=1}^n A(\lambda_j)e^{ih\lambda_j} \cdot \overline{\sum_{k=1}^n A(\lambda_k)}\right) = E\left(\sum_{j=1}^n A(\lambda_j)e^{ih\lambda_j} \cdot \sum_{k=1}^n \overline{A(\lambda_k)}\right) \\
&= E\left(\sum_{j=1}^n \sum_{k=1}^n A(\lambda_j)\overline{A(\lambda_k)}e^{ih\lambda_j}\right) = \sum_{j=1}^n \sum_{k=1}^n E(A(\lambda_j)\overline{A(\lambda_k)})e^{ih\lambda_j} \\
&= \sum_{j=1}^n E(A(\lambda_j)\overline{A(\lambda_j)})e^{ih\lambda_j} = \sum_{j=1}^n \sigma_j^2 e^{ih\lambda_j} = \int_{(-\pi, \pi]} e^{ih\nu} dF(\nu),
\end{aligned}$$

donde $F(\lambda) = \sum_{j:\lambda_j \leq \lambda} \sigma_j^2$ y la última integral representa la integral de Riemann-Stieltjes. Esta integral tiene el mismo sentido que el valor esperado respecto de una función de distribución tal como se estudia en los cursos de probabilidad. La función F asigna toda su masa al intervalo de frecuencia $(-\pi, \pi]$. Además, es no decreciente (por saltos de magnitud $\sigma_j^2 = E(A(\lambda_j)\overline{A(\lambda_j)}) = E|A(\lambda_j)|^2 \geq 0$), continua por la derecha y $F(-\pi) = 0$, pero $F(\pi) = \gamma(0) = E|X_t|^2 = \sum_{j=1}^n \sigma_j^2$ que no tiene por qué ser uno. Entonces salvo por ese factor de normalización F es una función de distribución y es conocida como la *función de distribución espectral*.

El siguiente teorema (Teorema 4.1.1 en [4]) caracteriza las funciones de autocovarianza como aquellas funciones que son Hermíticas y definidas no negativas. En la siguiente sección, a través del Teorema de Herglotz, se encontrará una caracterización equivalente para estas funciones.

Teorema 2.0.1. Una función $K(\cdot)$ definida en los enteros es una función de autocovarianza de una serie de tiempo estacionaria si, y sólo si, $K(\cdot)$ es Hermítica y definida no negativa, es decir, si, y sólo si, $K(n) = \overline{K(-n)}$ y

$$\sum_{i,j=1}^n a_i K(i-j) \bar{a}_j \geq 0,$$

para todo entero positivo n y vectores $a = (a_1, \dots, a_n)' \in \mathbb{C}^n$.

2.1. El Teorema de Herglotz

El cálculo anterior ha mostrado que la función de autocovarianza de un proceso con representación espectral finita como en (2.1) admite una representación como transformada de Fourier de cierta función de distribución espectral. El Teorema de Herglotz, que incluimos en esta sección, prueba que esta propiedad se da de forma general, lo que permite un análisis de series temporales en el dominio de la frecuencia o análisis espectral.

Teorema 2.1.1 (Herglotz). Una función $\gamma(\cdot)$ con valores reales definida en los enteros es definida no negativa si, y sólo si,

$$\gamma(h) = \int_{(-\pi, \pi]} e^{ih\nu} dF(\nu) \quad (2.2)$$

para todo $h \in \mathbb{Z}$, donde $F(\cdot)$ es continua por la derecha, no decreciente, acotada en $[-\pi, \pi]$ y con $F(-\pi) = 0$. Esta función F es la función de distribución espectral de γ . Igual que antes F no es exactamente una función

de distribución, porque $F(\pi) = \gamma(0)$ no es necesariamente 1. Si F admite la representación $F(\lambda) = \int_{-\pi}^{\lambda} f(\nu) d\nu$, $-\pi \leq \lambda \leq \pi$ entonces f es la *densidad espectral* de γ .

Demostración. Suponemos que $\gamma(\cdot)$ es de la forma (2.2). Tenemos que

$$\overline{\gamma(h)} = \overline{\int_{(-\pi, \pi]} e^{ih\nu} dF(\nu)} = \int_{(-\pi, \pi]} \overline{e^{ih\nu}} dF(\nu) = \int_{(-\pi, \pi]} e^{-ih\nu} dF(\nu) = \gamma(-h)$$

luego la función $\gamma(\cdot)$ es Hermítica. Además, si $a_r \in \mathbb{C}$, $r = 1, \dots, n$, entonces

$$\sum_{r,s=1}^n a_r \gamma(r-s) \overline{a_s} = \int_{-\pi}^{\pi} \sum_{r,s=1}^n a_r \overline{a_s} e^{i\nu(r-s)} dF(\nu) = \int_{-\pi}^{\pi} \left| \sum_{r=1}^n a_r e^{i\nu r} \right|^2 dF(\nu) \geq 0$$

lo que implica que $\gamma(\cdot)$ es definida no negativa. Entonces, por el teorema (2.0.1), $\gamma(\cdot)$ es una función de autocovarianza.

Para la otra implicación suponemos que $\gamma(\cdot)$ es una función definida no negativa en los enteros. Entonces

$$f_N(\nu) = \frac{1}{2\pi N} \sum_{r,s=1}^N e^{-ir\nu} \gamma(r-s) e^{is\nu} = \frac{1}{2\pi N} \sum_{|m| < N} (N - |m|) e^{-im\nu} \gamma(m) \geq 0$$

para todo $\nu \in (-\pi, \pi]$.

Sea $F_N(\cdot)$ la función de distribución correspondiente a la densidad $f_N(\cdot)I_{(-\pi, \pi]}(\cdot)$ cuya integral no va a ser uno pero sí finita y positiva. Entonces

$$F_N(\lambda) = \begin{cases} 0 & \text{si } \lambda \leq -\pi \\ \int_{-\pi}^{\lambda} f_N(\nu) d\nu & \text{si } -\pi \leq \lambda \leq \pi \\ F_N(\pi) & \text{si } \lambda \geq \pi \end{cases}$$

Para cada entero h tenemos que

$$\begin{aligned} \int_{(-\pi, \pi]} e^{ih\nu} dF_N(\nu) &= \int_{(-\pi, \pi]} e^{ih\nu} f_N(\nu) d\nu \\ &= \frac{1}{2\pi N} \sum_{|m| < N} (N - |m|) \gamma(m) \int_{-\pi}^{\pi} e^{i(h-m)\nu} d\nu \\ &= \frac{1}{2\pi} \sum_{|m| < N} \left(1 - \frac{|m|}{N}\right) \gamma(m) \int_{-\pi}^{\pi} e^{i(h-m)\nu} d\nu. \end{aligned}$$

Como, además, sabemos que

$$\int_{-\pi}^{\pi} e^{ikx} dx = \begin{cases} 2\pi & \text{si } k = 0, \\ 0 & \text{en otro caso.} \end{cases}$$

Entonces

$$\int_{(-\pi, \pi]} e^{ih\nu} dF_N(\nu) = \begin{cases} \left(1 - \frac{|h|}{N}\right) \gamma(h) & \text{si } |h| < N, \\ 0 & \text{en otro caso.} \end{cases} \quad (2.3)$$

Por otra parte, $F_N(\pi) = \int_{(-\pi, \pi]} dF_N(\nu) = \gamma(0) < \infty$ para todo N , luego aplicando el Teorema de Helly (ver, por ejemplo, el Capítulo 25 de [2]) deducimos que existe una función de distribución F y una subsucesión $\{F_{N_k}\}$ de la sucesión $\{F_N\}$ tal que para toda función g continua y acotada en $[-\pi, \pi]$,

$$\int_{(-\pi, \pi]} g(\nu) dF_{N_k}(\nu) \rightarrow \int_{(-\pi, \pi]} g(\nu) dF(\nu)$$

cuando $k \rightarrow \infty$. Cambiando N por N_k en (2.3) y haciendo k tender a ∞ concluimos que

$$\gamma(h) = \int_{(-\pi, \pi]} e^{ih\nu} dF(\nu).$$

□

La función de distribución espectral del Teorema de Herglotz es única, tal como se recoge en la siguiente proposición.

Proposición 2.1.2. Si F y G son dos funciones de distribución concentradas en $[-\pi, \pi]$ y tales que

$$\gamma(h) = \int_{(-\pi, \pi]} e^{ih\nu} dF(\nu) = \int_{(-\pi, \pi]} e^{ih\nu} dG(\nu), \quad h = 0, \pm 1, \dots \quad (2.4)$$

entonces $F = G$.

Demostración. Suponemos que se verifica (2.4). Sea f una función continua en $[-\pi, \pi]$ tal que $f(\pi) = f(-\pi)$ y sea

$$S_n f = \sum_{|j| \leq n} \langle f, e_j \rangle e_j(x)$$

su aproximación de Fourier de orden n , donde $\langle f, e_j \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ijx} dx$ son los coeficientes de Fourier de la función f y $e_j(x) = e^{ijx}$. Por el Teorema de Fejér (ver, por ejemplo, el Teorema 2.11 en [4]), dado $\varepsilon > 0$ existe n_0 tal que si $n \geq n_0$,

$$\sup_{x \in (-\pi, \pi]} \left| \frac{1}{n} \sum_{k=0}^{n-1} S_k f(x) - f(x) \right| \leq \varepsilon.$$

Por otra parte, tenemos que

$$\begin{aligned}
\int_{(-\pi, \pi]} S_n f(x) dF(x) &= \sum_{|j| \leq n} \langle f, e_j \rangle \int_{(-\pi, \pi]} e^{ijx} dF(x) \\
&= \sum_{|j| \leq n} \langle f, e_j \rangle \int_{(-\pi, \pi]} e^{ijx} dG(x) \\
&= \int_{(-\pi, \pi]} S_n f(x) dG(x)
\end{aligned}$$

lo que implica que

$$\int_{(-\pi, \pi]} \frac{1}{n} \sum_{k=0}^{n-1} S_k f(x) dF(x) = \int_{(-\pi, \pi]} \frac{1}{n} \sum_{k=0}^{n-1} S_k f(x) dG(x).$$

Entonces para cada $\varepsilon > 0$ y $n \geq n_0$ tenemos que

$$\begin{aligned}
&\left| \int_{(-\pi, \pi]} f(x) dF(x) - \int_{(-\pi, \pi]} f(x) dG(x) \right| \\
&= \left| \left(\int_{(-\pi, \pi]} f(x) dF(x) - \int_{(-\pi, \pi]} \frac{1}{n} \sum_{k=0}^{n-1} S_k f(x) dF(x) \right) \right. \\
&\quad \left. + \left(\int_{(-\pi, \pi]} \frac{1}{n} \sum_{k=0}^{n-1} S_k f(x) dG(x) - \int_{(-\pi, \pi]} f(x) dG(x) \right) \right| \\
&\leq \int_{(-\pi, \pi]} \varepsilon dF(x) + \int_{(-\pi, \pi]} \varepsilon dG(x) \\
&= \varepsilon \cdot (F(\pi) - F(-\pi)) + \varepsilon \cdot (G(\pi) - G(-\pi)) \\
&= \varepsilon \cdot (F(\pi) + G(\pi)).
\end{aligned}$$

Por lo tanto,

$$\int_{(-\pi, \pi]} f(x) dF(x) = \int_{(-\pi, \pi]} f(x) dG(x)$$

si f es una función continua con $f(\pi) = f(-\pi)$. Entonces tenemos que $F(\lambda) = G(\lambda)$ para todo λ . \square

Al principio de este capítulo ya comentamos que el teorema 2.0.1 y el Teorema de Herglotz eran caracterizaciones equivalentes de la función de autocovarianza. Entonces combinando esos dos resultados concluimos que una función $\gamma(\cdot)$ con valores complejos definida en los enteros es la función de autocovarianza de un proceso estacionario $\{X_t, t = 0, \pm 1, \dots\}$ si, y sólo si, ya sea

- (1) $\gamma(h) = \int_{(-\pi, \pi]} e^{ihv} dF(v)$ para todo $h = 0, \pm 1, \dots$, donde F es una función continua por la derecha, no decreciente, acotada en $[-\pi, \pi]$ con $F(-\pi) = 0$,

ó

- (2) $\sum_{i,j=1}^n a_i \gamma(i-j) \bar{a}_j \geq 0$ para todo entero positivo n y todo $a = (a_1, \dots, a_n)' \in \mathbb{C}^n$,

donde $F(\cdot)$ es la función de distribución espectral de $\gamma(\cdot)$ y $\{X_t\}$.

El siguiente teorema sobre el desarrollo de funciones en serie de Fourier lo incluimos aquí porque es útil para la teoría.

Teorema 2.1.3. Si $K(\cdot)$ es una función en los enteros con valores complejos tal que $\sum_{n=-\infty}^{\infty} |K(n)| < \infty$, entonces

$$K(h) = \int_{-\pi}^{\pi} e^{ihv} f(v) dv, \quad h = 0, \pm 1, \dots$$

donde $f(\lambda) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{-in\lambda} K(n)$.

Demostración. Por una parte, si $f(\lambda) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{-in\lambda} K(n)$ entonces

$$\int_{-\pi}^{\pi} e^{ihv} f(v) dv = \int_{-\pi}^{\pi} \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{i(h-n)v} K(n) dv \quad (2.5)$$

Por otra parte,

$$\begin{aligned} \left| \int_{-\pi}^{\pi} \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{i(h-n)v} K(n) dv \right| &\leq \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{n=-\infty}^{\infty} |e^{i(h-n)v} K(n)| dv \\ &\leq \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{n=-\infty}^{\infty} |e^{i(h-n)v}| |K(n)| dv \\ &\leq \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{n=-\infty}^{\infty} |K(n)| dv < \infty \end{aligned}$$

por ser la integral en el intervalo $[-\pi, \pi]$ de $\sum_{n=-\infty}^{\infty} |K(n)| < \infty$. Entonces podemos aplicar el Teorema de Fubini en (2.5) e intercambiar el orden del sumatorio y la integral obteniendo

$$\begin{aligned} \int_{-\pi}^{\pi} e^{ihv} f(v) dv &= \int_{-\pi}^{\pi} \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{i(h-n)v} K(n) dv \\ &= \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} K(n) \int_{-\pi}^{\pi} e^{i(h-n)v} dv = K(h) \end{aligned}$$

ya que como sabemos

$$\int_{-\pi}^{\pi} e^{ikx} dx = \begin{cases} 2\pi & \text{si } k = 0, \\ 0 & \text{en otro caso.} \end{cases}$$

□

La principal aplicación del teorema 2.1.3 al estudio de procesos estacionarios es el siguiente resultado, que proporciona una caracterización alternativa de las funciones de autocovarianza (bajo una hipótesis adicional) y, sobre todo, aporta una vía más simple para el cálculo de la densidad espectral.

Corolario 2.1.4. Una función $\gamma(\cdot)$ definida en los enteros con valores complejos absolutamente sumable ($\sum_{h \in \mathbb{Z}} |\gamma(h)| < \infty$) es la función de autocovarianza de un proceso estacionario si, y sólo si,

$$f(\lambda) := \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{-in\lambda} \gamma(n) \geq 0,$$

para todo $\lambda \in [-\pi, \pi]$, en cuyo caso $f(\cdot)$ es la densidad espectral de $\gamma(\cdot)$.

Demostración. Primero suponemos que $\gamma(\cdot)$ es una función de autocovarianza. Como $\gamma(\cdot)$ es definida no negativa,

$$\begin{aligned} 0 &\leq f_N(\lambda) := \frac{1}{2\pi N} \sum_{r,s=1}^N e^{-ir\lambda} \gamma(r-s) e^{is\lambda} \\ &= \frac{1}{2\pi} \sum_{|m| < N} \left(1 - \frac{|m|}{N}\right) e^{-im\lambda} \gamma(m) \rightarrow f(\lambda) \end{aligned}$$

cuando $N \rightarrow \infty$. Entonces $f(\lambda) \geq 0$, $-\pi \leq \lambda \leq \pi$. Además, como $\gamma(\cdot)$ es absolutamente sumable, por el teorema 2.1.3, tenemos que

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ihv} f(v) dv, \quad h = 0, \pm 1, \dots$$

y, por lo tanto, $f(\cdot)$ es la densidad espectral de $\gamma(\cdot)$.

Recíprocamente, supongamos que

$$f(\lambda) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{-in\lambda} \gamma(n) \geq 0$$

para todo $\lambda \in [-\pi, \pi]$. Como $\gamma(\cdot)$ es absolutamente sumable, entonces por el teorema 2.1.3

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ihv} f(v) dv.$$

Observamos que $f(\lambda) \geq 0$ y $\gamma(h) = \int_{-\pi}^{\pi} e^{ihv} dF(v)$ con $F(\lambda) = \int_{-\pi}^{\lambda} f(v) dv$. Esto implica, por los comentarios posteriores a la proposición 2.1.2, que $\gamma(\cdot)$ es una función de autocovarianza con densidad espectral f . \square

Este corolario sirve para verificar si una función absolutamente sumable en los enteros es definida negativa o no. Es una herramienta más simple que la verificación directa usando la definición establecida en el teorema 2.0.1.

2.2. Análisis espectral de procesos ARMA

Como consecuencia del corolario 2.1.4 obtenemos la siguiente representación de la densidad espectral de una clase amplia de procesos que incluye a los procesos ARMA.

Teorema 2.2.1. Si $\{Y_t\}$ es un proceso estacionario de media cero y valores complejos con función de distribución espectral $F_Y(\cdot)$ y $\{X_t\}$ es el proceso

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j} \quad \text{donde} \quad \sum_{j=-\infty}^{\infty} |\psi_j| < \infty,$$

entonces $\{X_t\}$ es estacionario con función de distribución espectral

$$F_X(\lambda) = \int_{(-\pi, \lambda]} \left| \sum_{j=-\infty}^{\infty} \psi_j e^{-ijv} \right|^2 dF_Y(v), \quad -\pi \leq \lambda \leq \pi. \quad (2.6)$$

Demostración. Primero veamos que $\{X_t\}$ es estacionario. Para ello, comprobaremos que EX_t y $E(X_{t+h}\bar{X}_t)$ son independientes de t y que $E|X_t|^2 < \infty$.

$$EX_t = E \left(\sum_{j=-\infty}^{\infty} \psi_j Y_{t-j} \right) = \sum_{j=-\infty}^{\infty} \psi_j E(Y_{t-j}) = 0,$$

por ser $\{Y_t\}$ un proceso de media cero. Nótese que el intercambio en el orden de sumas y esperanzas está justificado por el Teorema de Fubini y la

convergencia $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$.

$$\begin{aligned} E(X_{t+h}\bar{X}_t) &= E\left(\sum_{j=-\infty}^{\infty} \psi_j Y_{t+h-j} \cdot \overline{\sum_{k=-\infty}^{\infty} \psi_k Y_{t-k}}\right) \\ &= E\left(\sum_{j,k=-\infty}^{\infty} \psi_j \bar{\psi}_k Y_{t+h-j} \bar{Y}_{t-k}\right) = \sum_{j,k=-\infty}^{\infty} \psi_j \bar{\psi}_k E(Y_{t+h-j} \bar{Y}_{t-k}) \\ &= \sum_{j,k=-\infty}^{\infty} \psi_j \bar{\psi}_k \gamma_Y(h-j+k), \end{aligned}$$

independiente de t . En particular,

$$E|X_t|^2 = \sum_{j,k=-\infty}^{\infty} \psi_j \bar{\psi}_k E(Y_{t-j} \bar{Y}_{t-k}) \leq \left(\sum_{j=-\infty}^{\infty} |\psi_j|\right)^2 \sup_t E|Y_t|^2 < \infty$$

ya que por hipótesis $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ y $E|Y_t|^2 < \infty$ por ser $\{Y_t\}$ un proceso estacionario. Luego $\{X_t\}$ es estacionario y su función de autocovarianza es

$$\gamma_X(h) = \sum_{j,k=-\infty}^{\infty} \psi_j \bar{\psi}_k \gamma_Y(h-j+k).$$

Utilizando la representación espectral de $\{\gamma_Y(\cdot)\}$ tenemos que

$$\begin{aligned} \gamma_X(h) &= \sum_{j,k=-\infty}^{\infty} \psi_j \bar{\psi}_k \int_{(-\pi,\pi]} e^{i(h-j+k)\nu} dF_Y(\nu) \\ &= \int_{(-\pi,\pi]} \left(\sum_{j=-\infty}^{\infty} \psi_j e^{-ij\nu}\right) \left(\sum_{k=-\infty}^{\infty} \bar{\psi}_k e^{ik\nu}\right) e^{ih\nu} dF_Y(\nu) \\ &= \int_{(-\pi,\pi]} e^{ih\nu} \left|\sum_{j=-\infty}^{\infty} \psi_j e^{-ij\nu}\right|^2 dF_Y(\nu), \end{aligned}$$

lo que nos asegura que $F_X(\cdot)$ definido como en (2.6) es la función de distribución espectral de $\{X_t\}$. \square

Como consecuencia del resultado anterior, si $\{Y_t\}$ tiene densidad espectral $f_Y(\cdot)$ y si $\{X_t\}$ está definido por $X_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j}$ donde $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, entonces $\{X_t\}$ tiene densidad espectral

$$f_X(\lambda) = |\psi(e^{-i\lambda})|^2 f_Y(\lambda),$$

donde $\psi(e^{-i\lambda}) = \sum_{j=-\infty}^{\infty} \psi_j e^{-ij\lambda}$.

En el caso particular de modelos ARMA esto proporciona una expresión simple de la densidad espectral y el proceso estacionario de referencia es un ruido blanco, $\{Z_t\}_{t \in \mathbb{Z}} \sim WN(0, \sigma^2)$. Por definición, $\{Z_t\}$ tiene función de autocovarianza

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

Por otra parte, $\gamma_Z(h) = \int_{[-\pi, \pi]} e^{ihx} dF_Z(x) = \int_{[-\pi, \pi]} e^{ihx} f_Z(x) dx$. Entonces juntando las dos igualdades anteriores tenemos que

$$\int_{[-\pi, \pi]} e^{ihx} f_Z(x) dx = \begin{cases} \sigma^2 & \text{si } h = 0, \\ 0 & \text{si } h \neq 0 \end{cases}$$

y como sabemos que

$$\int_{-\pi}^{\pi} e^{ihx} dx = \begin{cases} 2\pi & \text{si } h = 0, \\ 0 & \text{si } h \neq 0 \end{cases}$$

podemos concluir que

$$f_Z(x) = \frac{\sigma^2}{2\pi}, \quad -\pi \leq x \leq \pi.$$

Teorema 2.2.2. Sea $\{X_t\}$ un proceso ARMA(p, q) que satisface

$$\phi(B)X_t = \theta(B)Z_t, \quad \{Z_t\} \sim WN(0, \sigma^2),$$

donde $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ y $\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$ no tienen ceros comunes y $\phi(z)$ no tiene ceros en el círculo unidad. Entonces $\{X_t\}$ tiene densidad espectral

$$f_X(\lambda) = \frac{\sigma^2 |\theta(e^{-i\lambda})|^2}{2\pi |\phi(e^{-i\lambda})|^2}, \quad -\pi \leq \lambda \leq \pi.$$

Demostración. Por hipótesis, como los polinomios $\phi(z)$ y $\theta(z)$ no tienen ceros comunes y $\phi(z) \neq 0$ si $|z| \leq 1$ entonces por la proposición 1.2.4 el proceso $\{X_t\}$ es causal. Por tanto podemos escribir

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \tag{2.7}$$

con $\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}$, $|z| \leq 1$, y $\sum_{j=0}^{\infty} |\psi_j| < \infty$. Combinando este cálculo con los comentarios posteriores al teorema 2.2.1 obtenemos

$$f_X(\lambda) = |\psi(e^{-i\lambda})|^2 f_Z(\lambda) = \left| \frac{\theta(e^{-i\lambda})}{\phi(e^{-i\lambda})} \right|^2 \frac{\sigma^2}{2\pi} = \frac{\sigma^2 |\theta(e^{-i\lambda})|^2}{2\pi |\phi(e^{-i\lambda})|^2}$$

para $-\pi \leq \lambda \leq \pi$, como queríamos probar. \square

2.2.1. Invertibilidad y densidad espectral

Como aplicación de los resultados anteriores sobre la densidad espectral, se puede mejorar la caracterización de la invertibilidad de un proceso ARMA descrita en el Capítulo 1. Este es el contenido de esta subsección.

Por la proposición 1.2.5 sabemos que un proceso ARMA es invertible si su polinomio de media móvil no tiene raíces en el círculo unidad (bajo el supuesto de que los polinomios autorregresivo y de media móvil no tienen ceros comunes). En la siguiente proposición se trata el caso en el que el polinomio de media móvil sí tiene ceros en el círculo unidad.

Proposición 2.2.3. Sea $\{X_t\}$ un proceso ARMA(p, q) satisfaciendo

$$\phi(B)X_t = \theta(B)Z_t, \quad \{Z_t\} \sim WN(0, \sigma^2),$$

donde $\phi(z)$ y $\theta(z)$ no tienen ceros comunes, $\phi(z) \neq 0$ para $|z| = 1$ y $\theta(z) \neq 0$ para $|z| < 1$. Entonces el proceso $\{X_t\}$ es invertible, es decir, $Z_t \in \overline{sp}\{X_s, -\infty < s \leq t\}$.

Demostración. Escribamos $\theta(z) = \prod_{j=1}^q (1 - b_j^{-1}z)$ donde $\{b_j\}_{j=1}^q$ son las raíces del polinomio de media móvil $\theta(\cdot)$. Por hipótesis $\theta(\cdot)$ no tiene raíces en el interior del círculo unidad luego factorizando obtenemos $\theta(z) = \theta^+(z)\theta^*(z)$, donde

$$\theta^+(z) = \prod_{j=1}^s (1 - b_j^{-1}z), \quad \text{con } |b_j| > 1$$

y

$$\theta^*(z) = \prod_{j=s+1}^q (1 - b_j^{-1}z), \quad \text{con } |b_j| = 1.$$

Consideramos el proceso MA($q - s$), $Y_t = \theta^*(B)Z_t$. Es claro que

$$\phi(B)X_t = \theta^+(B)Y_t$$

ya que

$$\phi(B)X_t = \theta(B)Z_t = \theta^+(B)\theta^*(B)Z_t = \theta^+(B)Y_t.$$

El polinomio $\theta^+(z)$ no tiene ceros en $B(0, 1 + \varepsilon)$ para un $\varepsilon > 0$ y es una función holomorfa, luego $\frac{1}{\theta^+(z)}$ es una función holomorfa en $B(0, 1 + \varepsilon)$ y se puede desarrollar en serie de potencias

$$\frac{1}{\theta^+(z)} = \sum_{j=0}^{\infty} \xi_j z^j = \xi(z), \quad |z| < 1 + \varepsilon.$$

En consecuencia $\xi_j(1 + \frac{\varepsilon}{2})^j \xrightarrow{j \rightarrow \infty} 0$ y esto implica que $\exists K \in (0, \infty)$ tal que $|\xi_j| < K(1 + \frac{\varepsilon}{2})^{-j}, \forall j = 0, 1, 2, \dots$. En particular,

$$\sum_{j=0}^{\infty} |\xi_j| < \infty \quad (2.8)$$

y

$$\xi(z)\theta^+(z) \equiv 1, \quad \text{para todo } |z| \leq 1. \quad (2.9)$$

Por otra parte, llamando $V_t = \phi(B)X_t = \theta(B)Z_t$ y teniendo en cuenta que $\{Z_t\} \sim WN(0, \sigma^2)$ tenemos que $E|V_t|^2 = (1 + |\theta_1|^2 + \dots + |\theta_q|^2)\sigma^2 < \infty$. Entonces $\sup_t E|V_t|^2 < \infty$ y, en consecuencia, $\sup_t E|V_t| < \infty$. Luego por el Teorema de la Convergencia Monótona,

$$\begin{aligned} E \left(\sum_{j=0}^{\infty} |\xi_j| |V_{t-j}| \right) &= \lim_{n \rightarrow \infty} E \left(\sum_{j=0}^n |\xi_j| |V_{t-j}| \right) \\ &\leq \lim_{n \rightarrow \infty} \left(\sum_{j=0}^n |\xi_j| \right) \sup_t E|V_t| < \infty, \end{aligned}$$

donde hemos utilizado (2.8). Entonces $\sum_{j=0}^{\infty} |\xi_j| |V_{t-j}|$ es finita con probabilidad 1 y, en consecuencia, $\sum_{j=0}^{\infty} \xi_j V_{t-j} = \xi(B)V_t$ es finita con probabilidad 1. Entonces podemos aplicar $\xi(B)$ a ambos lados de la ecuación $\phi(B)X_t = \theta^+(B)Y_t$ y obtenemos

$$\xi(B)\phi(B)X_t = \xi(B)\theta^+(B)Y_t = Y_t,$$

utilizando la ecuación (2.9). Por lo tanto,

$$Y_t = \sum_{j=0}^{\infty} \psi_j X_{t-j} \quad (2.10)$$

donde $\{\psi_j\}$ satisface $\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \xi(z)\phi(z)$ con $|z| \leq 1$. Entonces $\overline{sp}\{Y_k, -\infty < k \leq t\} \subseteq \overline{sp}\{X_k, -\infty < k \leq t\}$ y bastaría demostrar que

$Z_t \in \overline{sp}\{Y_k, -\infty < k \leq t\}$. Replicando el argumento usado en la demostración de la proposición 1.2.1 tenemos que si

$$U_t = Y_t - P_{\overline{sp}\{Y_k, -\infty < k < t\}} Y_t. \quad (2.11)$$

entonces $\{U_t\} \sim WN(0, \sigma_U^2)$ y además

$$Y_t = U_t + \alpha_1 U_{t-1} + \dots + \alpha_{q-s} U_{t-q+s}. \quad (2.12)$$

Por los comentarios posteriores a la proposición 2.2.1 tenemos que la densidad espectral de $\{Y_t\}$ es

$$f_Y(\lambda) = \frac{\sigma_U^2}{2\pi} |\alpha(e^{-i\lambda})|^2 = \frac{\sigma^2}{2\pi} |\theta^*(e^{-i\lambda})|^2,$$

donde hemos utilizado las dos representaciones de $\{Y_t\}$ (ecuaciones (2.12) y (2.10) respectivamente). Los polinomios $\alpha(z)$ y $\theta^*(z)$ tienen el mismo grado, $q - s$. Además, todos los ceros de $\theta^*(z)$ están en el círculo unidad luego todos los ceros con su multiplicidad de $\theta^*(z)$ son ceros de $\alpha(z)$ pero $\alpha(z)$ no puede tener más ceros. Entonces $\alpha(z)$ y $\theta^*(z)$ tienen los mismos ceros. Esto implica que

$$\theta^*(z) = \alpha(z)$$

y

$$\sigma^2 = \sigma_U^2.$$

Entonces los vectores $(U_t, Y_t, \dots, Y_{t-n})'$ y $(Z_t, Y_t, \dots, Y_{t-n})'$ tienen la misma matriz de covarianza y

$$P_{\overline{sp}\{Y_k, t-n < k \leq t\}} Z_t = P_{\overline{sp}\{Y_k, t-n < k \leq t\}} U_t.$$

Tomando límites cuando n tiende a infinito tenemos que

$$P_{\overline{sp}\{Y_k, -\infty < k \leq t\}} Z_t = P_{\overline{sp}\{Y_k, -\infty < k \leq t\}} U_t,$$

y como por (2.11) $U_t \in \overline{sp}\{Y_k, -\infty < k \leq t\}$ entonces

$$P_{\overline{sp}\{Y_k, -\infty < k \leq t\}} Z_t = U_t.$$

Por otra parte,

$$\begin{aligned} E(Z_t^2) &= \|Z_t\|^2 = \|P_{\overline{sp}\{Y_k, -\infty < k \leq t\}} Z_t\|^2 + \|Z_t - P_{\overline{sp}\{Y_k, -\infty < k \leq t\}} Z_t\|^2 \\ &= \|U_t\|^2 + \|Z_t - U_t\|^2 = E(U_t^2) + E((Z_t - U_t)^2). \end{aligned} \quad (2.13)$$

Entonces $E((Z_t - U_t)^2) = E(Z_t^2) - E(U_t^2) = 0$ debido a que

$$E(Z_t^2) = E(Z_t^2) - E(Z_t)^2 = \sigma^2 = \sigma_U^2 = E(U_t^2) - E(U_t)^2 = E(U_t^2)$$

por ser $\{Z_t\} \sim WN(0, \sigma^2)$ y $\{U_t\} \sim WN(0, \sigma_U^2)$. Sustituyendo en (2.13) tenemos que $Z_t = U_t = P_{\overline{sp}\{Y_k, -\infty < k \leq t\}} Z_t$, es decir, $Z_t \in \overline{sp}\{Y_k, -\infty < k \leq t\}$, como queríamos probar. \square

Asumiendo que ϕ y θ no tienen ceros comunes y que ϕ no se anula en la circunferencia unidad, la Proposición 2.2.3 nos dice que una condición suficiente para garantizar que $\{X_t\}$ es invertible es que θ no se anule en el interior del círculo unidad. La condición es también necesaria. La demostración de este hecho es técnicamente más complicada y puede consultarse en [4]

2.2.2. Aproximaciones racionales para densidades espectrales

En este apartado veremos que la función de densidad espectral de un proceso autorregresivo causal o un proceso de media móvil invertible puede aproximar, con un error pequeño, a cualquier densidad espectral continua y simétrica de un proceso estacionario. Estos resultados son muy importantes para entender la importancia de los procesos ARMA, ya que nos indican que pueden aproximar cualquier otro proceso estacionario con un error pequeño.

Para la demostración de esos resultados (corolarios 2.2.5 y 2.2.6) necesitamos el siguiente teorema de aproximación. Recordamos que f es la densidad espectral de un proceso estacionario de valor real si y sólo si f es simétrica, no negativa e integrable en $[-\pi, \pi]$.

Teorema 2.2.4. Si f es una densidad espectral simétrica y continua en $[-\pi, \pi]$, entonces para cada $\varepsilon > 0$ existe un entero no negativo p y un polinomio $a(z) = \prod_{j=1}^p (1 - \eta_j^{-1}z) = 1 + a_1z + \dots + a_pz^p$ con $|\eta_j| > 1, j = 1, \dots, p$, y coeficientes con valores reales a_1, \dots, a_p , tal que

$$|A \cdot |a(e^{-i\lambda})|^2 - f(\lambda)| < \varepsilon \quad \text{para todo } \lambda \in [-\pi, \pi]$$

donde $A = (2\pi)^{-1}(1 + a_1^2 + \dots + a_p^2)^{-1} \int_{-\pi}^{\pi} f(v)dv$.

Demostración. Si $f(\lambda) \equiv 0$, el resultado es cierto tomando $p = 0$. Por lo tanto, suponemos que $M = \sup_{\lambda \in [-\pi, \pi]} f(\lambda) > 0$. Definimos para algún $\varepsilon > 0$

$$\delta = \min \left\{ M, \frac{\varepsilon}{4\pi M \left(\int_{-\pi}^{\pi} f(u)du \right) + 2} \right\} \quad \text{y} \quad f^\delta(\lambda) = \max\{f(\lambda), \delta\}.$$

Es claro que $f^\delta(\lambda)$ es una densidad espectral simétrica y continua, ya que $f(\lambda)$ lo es, con

$$f^\delta(\lambda) \geq \delta \quad \text{y} \quad 0 \leq f^\delta(\lambda) - f(\lambda) \leq \delta \quad \text{para todo } \lambda \in [-\pi, \pi]. \quad (2.14)$$

Ahora por el Teorema de Fejér (ver, por ejemplo, el Teorema 2.11 en [4]) existe un entero r tal que

$$\left| \frac{1}{r} \cdot \sum_{j=0}^{r-1} \sum_{|k| \leq j} b_k e^{-ik\lambda} - f^\delta(\lambda) \right| < \delta \quad (2.15)$$

para todo $\lambda \in [-\pi, \pi]$, donde $b_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f^\delta(v) e^{ivk} dv$. Intercambiando el orden de los sumatorios y utilizando el hecho de que f^δ es una función simétrica, tenemos que

$$\frac{1}{r} \cdot \sum_{j=0}^{r-1} \sum_{|k| \leq j} b_k e^{-ik\lambda} = \sum_{|k| < r} \left(1 - \frac{|k|}{r}\right) b_k e^{-ik\lambda}.$$

Esta función es estrictamente positiva para todo λ por definición de $f^\delta(\lambda)$ y la ecuación (2.14).

Sea $C(z) = \sum_{|k| < r} \left(1 - \frac{|k|}{r}\right) b_k z^k$. Observemos que

$$\begin{aligned} C(z^{-1}) &= \sum_{|k| < r} \left(1 - \frac{|k|}{r}\right) b_k z^{-k} = \sum_{|s| < r} \left(1 - \frac{|s|}{r}\right) b_{-s} z^s \\ &= \sum_{|s| < r} \left(1 - \frac{|s|}{r}\right) b_s z^s = C(z) \end{aligned}$$

haciendo el cambio de variable $s = -k$ y utilizando que

$$\begin{aligned} b_{-s} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f^\delta(v) e^{-ivs} dv = \frac{1}{2\pi} \int_{-\pi}^{\pi} f^\delta(-u) e^{ius} du \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f^\delta(-u) e^{ius} du = b_s \end{aligned}$$

donde hemos realizado el cambio de variable $u = -v$ y utilizado la simetría de la función f^δ .

En conclusión, si $C(m) = 0$ entonces tenemos que $C(m^{-1}) = 0$ y, definiendo $p = \max\{k : b_k \neq 0\}$, podemos escribir

$$z^p C(z) = K_1 \prod_{j=1}^p (1 - \eta_j^{-1} z)(1 - \eta_j z)$$

para algunas $K_1, \eta_1, \dots, \eta_p$ tales que $|\eta_j| > 1$, $j = 1, \dots, p$. Si llamamos $a(z)$

al polinomio $1 + a_1z + \dots + a_pz^p = \prod_{j=1}^p (1 - \eta_j^{-1}z)$ podemos escribir

$$\begin{aligned}
C(z) &= z^{-p} K_1 \prod_{j=1}^p (1 - \eta_j^{-1}z)(1 - \eta_j z) = z^{-p} K_1 a(z) \prod_{j=1}^p (1 - \eta_j z) \\
&= K_1 a(z) \prod_{j=1}^p z^{-1} (1 - \eta_j z) = K_1 a(z) \prod_{j=1}^p (z^{-1} - \eta_j) \\
&= K_1 a(z) (-1)^p \prod_{j=1}^p (\eta_j - z^{-1}) = K_1 a(z) (-1)^p \prod_{j=1}^p \eta_j (1 - \eta_j^{-1} z^{-1}) \\
&= (-1)^p \eta_1 \cdots \eta_p K_1 a(z) a(z^{-1}) = K_2 a(z) a(z^{-1})
\end{aligned}$$

donde $K_2 = (-1)^p \eta_1 \cdots \eta_p K_1$. Entonces, igualando los coeficientes de z^0 en cada lado de la ecuación $K_2 a(z) a(z^{-1}) = C(z)$ tenemos $K_2(1 + a_1^2 + \dots + a_p^2) = b_0$ y despejando la constante K_2 tenemos que

$$K_2 = b_0(1 + a_1^2 + \dots + a_p^2)^{-1} = (2\pi)^{-1}(1 + a_1^2 + \dots + a_p^2)^{-1} \int_{-\pi}^{\pi} f^\delta(v) dv.$$

Además, por la ecuación (2.15) tenemos

$$|K_2 |a(e^{-i\lambda})|^2 - f^\delta(\lambda)| < \delta \quad \text{para todo } \lambda,$$

es decir,

$$\left[(2\pi)^{-1}(1 + a_1^2 + \dots + a_p^2)^{-1} \int_{-\pi}^{\pi} f^\delta(v) dv \right] |a(e^{-i\lambda})|^2 < f^\delta(\lambda) + \delta.$$

Y utilizando la ecuación (2.14) obtenemos la acotación

$$\begin{aligned}
(1 + a_1^2 + \dots + a_p^2)^{-1} |a(e^{-i\lambda})|^2 &\leq (f^\delta(\lambda) + \delta) 2\pi \left(\int_{-\pi}^{\pi} f^\delta(v) dv \right)^{-1} \\
&\leq 4\pi M \left(\int_{-\pi}^{\pi} f(v) dv \right)^{-1}.
\end{aligned}$$

Ahora definiendo A como en el enunciado del teorema, $A = (2\pi)^{-1}(1 + a_1^2 + \dots + a_p^2)^{-1} \int_{-\pi}^{\pi} f(v) dv$, y utilizando la acotación anterior y la ecuación (2.14)

tenemos

$$\begin{aligned}
|A|a(e^{-i\lambda})|^2 - K_2|a(e^{-i\lambda})|^2| &= |(A - K_2)|a(e^{-i\lambda})|^2| \\
&= \left| (2\pi)^{-1}(1 + a_1^2 + \dots + a_p^2)^{-1} \left(\int_{-\pi}^{\pi} f(v)dv - \int_{-\pi}^{\pi} f^\delta(v)dv \right) |a(e^{-i\lambda})|^2 \right| \\
&= (2\pi)^{-1} \left(\int_{-\pi}^{\pi} (f(v) - f^\delta(v))dv \right) (1 + a_1^2 + \dots + a_p^2)^{-1} |a(e^{-i\lambda})|^2 \\
&\leq (2\pi)^{-1} \left(\int_{-\pi}^{\pi} |f(v) - f^\delta(v)|dv \right) 4\pi M \left(\int_{-\pi}^{\pi} f(v)dv \right)^{-1} \\
&\leq (2\pi)^{-1} (2\pi\delta) 4\pi M \left(\int_{-\pi}^{\pi} f(v)dv \right)^{-1} \leq 4\pi M\delta \left(\int_{-\pi}^{\pi} f(v)dv \right)^{-1}.
\end{aligned}$$

En conclusión, utilizando todos los resultados obtenidos anteriormente tenemos

$$\begin{aligned}
|A|a(e^{-i\lambda})|^2 - f(\lambda)| &\leq |A|a(e^{-i\lambda})|^2 - K_2|a(e^{-i\lambda})|^2| + |K_2|a(e^{-i\lambda})|^2 - f^\delta(\lambda)| + |f^\delta(\lambda) - f(\lambda)| \\
&\leq 4\pi M\delta \left(\int_{-\pi}^{\pi} f(v)dv \right)^{-1} + \delta + \delta = \left[4\pi M \left(\int_{-\pi}^{\pi} f(v)dv \right)^{-1} + 2 \right] \delta \\
&\leq \left[4\pi M \left(\int_{-\pi}^{\pi} f(v)dv \right)^{-1} + 2 \right] \frac{\varepsilon}{\left[4\pi M \left(\int_{-\pi}^{\pi} f(v)dv \right)^{-1} + 2 \right]} = \varepsilon,
\end{aligned}$$

por la definición de δ . □

Como consecuencia del Teorema 2.2.4, estamos en condiciones de justificar que cualquier proceso estacionario real regular (con densidad espectral) es, aproximadamente, en cierto sentido, un proceso MA(q) invertible o un proceso AR(p) causal.

Corolario 2.2.5. Si f es una densidad espectral continua y simétrica y $\varepsilon > 0$, entonces existe un proceso invertible MA(q)

$$X_t = Z_t + a_1 Z_{t-1} + \dots + a_q Z_{t-q}, \quad \{Z_t\} \sim WN(0, \sigma^2),$$

tal que $|f_X(\lambda) - f(\lambda)| < \varepsilon$ para todo $\lambda \in [-\pi, \pi]$, donde $\sigma^2 = \frac{\int_{-\pi}^{\pi} f(u)du}{1+a_1^2+\dots+a_q^2}$.

Demostración. f es una densidad espectral continua y simétrica en $[-\pi, \pi]$ luego por el Teorema 2.2.4, para cada $\varepsilon > 0$, existe un entero q no negativo

y un polinomio $a(z) = \prod_{j=1}^q (1 - \eta_j^{-1}z) = 1 + a_1z + \dots + a_qz^q$ con $|\eta_j| > 1$, $j = 1, \dots, q$, y coeficientes reales a_1, \dots, a_q tal que

$$|A \cdot |a(e^{-i\lambda})|^2 - f(\lambda)| < \varepsilon \quad \text{para todo } \lambda \in [-\pi, \pi] \quad (2.16)$$

donde $A = \frac{1}{2\pi \cdot (1+a_1^2+\dots+a_q^2)} \cdot \int_{-\pi}^{\pi} f(u)du$. Entonces por el Teorema 2.2.2 el proceso invertible MA(q),

$$X_t = a(B)Z_t, \quad \{Z_t\} \sim WN(0, 2\pi A),$$

tiene densidad espectral $f_X(\lambda) = A \cdot |a(e^{-i\lambda})|^2$ y se cumple $|f_X(\lambda) - f(\lambda)| < \varepsilon$ para todo $\lambda \in [-\pi, \pi]$ por (2.16).

Además,

$$\sigma^2 = 2\pi A = \frac{\int_{-\pi}^{\pi} f(u)du}{(1 + a_1^2 + \dots + a_q^2)}.$$

□

Corolario 2.2.6. Si f es una densidad espectral continua y simétrica y $\varepsilon > 0$, entonces existe un proceso causal AR(p)

$$X_t + a_1X_{t-1} + \dots + a_pX_{t-p} = Z_t, \quad \{Z_t\} \sim WN(0, \sigma^2),$$

tal que $|f_X(\lambda) - f(\lambda)| < \varepsilon$ para todo $\lambda \in [-\pi, \pi]$.

Demostración. Sea $f^\varepsilon(\lambda) = \max\{f(\lambda), \frac{\varepsilon}{2}\}$. Entonces

$$f^\varepsilon(\lambda) \geq \frac{\varepsilon}{2} \quad \text{y} \quad 0 \leq f^\varepsilon(\lambda) - f(\lambda) \leq \frac{\varepsilon}{2} \quad \text{para todo } \lambda \in [-\pi, \pi]. \quad (2.17)$$

Sea $M = \max_{\lambda} f^\varepsilon(\lambda)$ y $\delta = \min\left\{\frac{\varepsilon}{(2M)^2}, \frac{1}{2M}\right\}$. Por como hemos definido la función $\frac{1}{f^\varepsilon(\lambda)}$, es una densidad espectral continua y simétrica en $[-\pi, \pi]$. Entonces por el Teorema 2.2.4, y tomando el δ definido antes que era mayor que 0, existe un entero no negativo p y un polinomio $a(z) = \prod_{j=1}^p (1 - \eta_j^{-1}z) = 1 + a_1z + \dots + a_pz^p$ con $|\eta_j| > 1$, $j = 1, \dots, p$, y coeficientes reales a_1, \dots, a_p tal que

$$\left| A \cdot |a(e^{-i\lambda})|^2 - \frac{1}{f^\varepsilon(\lambda)} \right| < \delta \quad \text{para todo } \lambda \in [-\pi, \pi]$$

donde $A = \frac{1}{2\pi \cdot (1+a_1^2+\dots+a_p^2)} \cdot \int_{-\pi}^{\pi} \frac{1}{f^\varepsilon(u)} du$.

Por la desigualdad anterior tenemos que

$$A \cdot |a(e^{-i\lambda})|^2 - \frac{1}{f^\varepsilon(\lambda)} > -\delta.$$

Es decir,

$$A \cdot |a(e^{-i\lambda})|^2 > \frac{1}{f^\varepsilon(\lambda)} - \delta = \frac{1 - \delta f^\varepsilon(\lambda)}{f^\varepsilon(\lambda)}.$$

Entonces,

$$A^{-1} \cdot |a(e^{-i\lambda})|^{-2} \leq \frac{f^\varepsilon(\lambda)}{1 - \delta f^\varepsilon(\lambda)} \leq \frac{M}{1 - \delta M} \leq 2M,$$

por definición de M y δ . Por lo tanto, utilizando las acotaciones obtenidas anteriormente, y las definiciones de M y de δ obtenemos

$$\begin{aligned} |A^{-1}|a(e^{-i\lambda})|^{-2} - f^\varepsilon(\lambda)| &= [A^{-1}|a(e^{-i\lambda})|^{-2}f^\varepsilon(\lambda)] \cdot \left| A|a(e^{-i\lambda})|^2 - \frac{1}{f^\varepsilon(\lambda)} \right| \\ &< 2M^2\delta \leq \frac{\varepsilon}{2}. \end{aligned} \quad (2.18)$$

Utilizando las desigualdades (2.18) y (2.17) obtenemos

$$\begin{aligned} |A^{-1}|a(e^{-i\lambda})|^{-2} - f(\lambda)| &\leq |A^{-1}|a(e^{-i\lambda})|^{-2} - f^\varepsilon(\lambda)| + |f^\varepsilon(\lambda) - f(\lambda)| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \end{aligned} \quad (2.19)$$

para todo $\lambda \in [-\pi, \pi]$. Entonces por el teorema 2.2.2 el proceso causal AR(p),

$$a(B)X_t = Z_t, \quad \{Z_t\} \sim WN(0, 2\pi A^{-1}),$$

tiene densidad espectral $f_X(\lambda) = A^{-1} \cdot |a(e^{-i\lambda})|^{-2}$ y se cumple $|f_X(\lambda) - f(\lambda)| < \varepsilon$ para todo $\lambda \in [-\pi, \pi]$ por (2.19). \square

Capítulo 3

Predicción de procesos estacionarios

El problema de la predicción es un problema de interés general en estadística. En este capítulo, consideramos el problema de la predicción desde un punto de vista teórico. El objetivo es predecir, es decir, dar valores aproximadamente correctos para $\{X_t, t \geq n + 1\}$ a partir de $\{X_1, \dots, X_n\}$. Fijándonos en $t = n + 1$ (predicción de un paso) se trata de encontrar la función de X_1, \dots, X_n que mejor aproxima a X_{n+1} , es decir, que minimice el error cuadrático medio. Cuando se mide la calidad de la predicción con este error entonces la mejor función de X_1, \dots, X_n para predecir X_{n+1} es la esperanza condicionada $E(X_{n+1}|X_1, \dots, X_n)$. Si la serie es Gaussiana entonces es un resultado bien conocido que la esperanza condicionada coincide con la mejor predicción lineal, $P_{\overline{sp}\{X_1, \dots, X_n\}}X_{n+1}$. Asumiendo $\mu = 0$ tenemos que $P_{\overline{sp}\{1, X_1, \dots, X_n\}}X_{n+h} = P_{\overline{sp}\{X_1, \dots, X_n\}}X_{n+h}$ luego nuestro objetivo es encontrar la función afín de X_1, \dots, X_n más próxima a X_{n+1} (y en general a X_{n+h}) en el sentido de los mínimos cuadrados.

El carácter teórico de los resultados en este capítulo viene dado por el hecho de que las funciones óptimas de predicción en el sentido anterior dependen en general de la función de autocovarianza del proceso. Por ello, la utilización en la práctica de estas fórmulas de predicción requiere la estimación de la función de autocovarianza a partir de los datos. Este problema se pospone al Capítulo 4.

Ecuaciones para predictores de un paso

De acuerdo con la discusión anterior, denotamos $H_n = \overline{\text{sp}}\{X_1, \dots, X_n\}$, $n \geq 1$, subespacio lineal cerrado, y definimos el predictor de un paso

$$\hat{X}_{n+1} = \begin{cases} 0 & \text{si } n = 0 \\ P_{H_n} X_{n+1} & \text{si } n \geq 1. \end{cases}$$

Por definición, para $n \geq 1$ tenemos que $\hat{X}_{n+1} \in H_n$, luego podemos escribir $\hat{X}_{n+1} = \phi_{n1}X_n + \dots + \phi_{nn}X_1$, satisfaciendo por ortogonalidad $\langle X_{n+1} - \hat{X}_{n+1}, X_{n+1-j} \rangle = 0$, $j = 1, \dots, n$. Por la linealidad del producto interno estas ecuaciones llamadas *ecuaciones de predicción de un paso* o *ecuaciones de Yule-Walker*, que ya vimos en el Capítulo 1, se pueden reescribir de la forma

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

o en forma matricial $\Gamma_n \Phi_n = \gamma_n$, es decir,

$$\begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \cdots & \gamma(n-2) \\ \gamma(2) & \gamma(1) & \gamma(0) & \cdots & \gamma(n-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \gamma(n-3) & \cdots & \gamma(0) \end{bmatrix} \begin{bmatrix} \phi_{n1} \\ \phi_{n2} \\ \phi_{n3} \\ \vdots \\ \phi_{nn} \end{bmatrix} = \begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \gamma(3) \\ \vdots \\ \gamma(n) \end{bmatrix}.$$

Recordamos que la solución es $\Phi_n = \Gamma_n^{-1} \gamma_n$ y es única si Γ_n es no singular. Junto con el mejor predictor de un paso podemos calcular el error (cuadrático) de predicción, es decir, $v_n = E(X_{n+1} - \hat{X}_{n+1})^2$. Escribiendo $X_{n+1} = \hat{X}_{n+1} + (X_{n+1} - \hat{X}_{n+1})$ y utilizando que \hat{X}_{n+1} y $X_{n+1} - \hat{X}_{n+1}$ son ortogonales tenemos que $EX_{n+1}^2 = E\hat{X}_{n+1}^2 + E(X_{n+1} - \hat{X}_{n+1})^2$, luego

$$v_n = EX_{n+1}^2 - E\hat{X}_{n+1}^2 = \gamma(0) - E\hat{X}_{n+1}^2.$$

Por otra parte, sabemos que $\hat{X}_{n+1} = \Phi_n' X$, donde $X = (X_n, \dots, X_1)'$, luego

$$\begin{aligned} E\hat{X}_{n+1}^2 &= \text{Var}(\hat{X}_{n+1}) = \text{Var}(\Phi_n' X) = \Phi_n' \text{Cov}(X) \Phi_n = \gamma_n' \Gamma_n^{-1} \Gamma_n \Gamma_n^{-1} \gamma_n \\ &= \gamma_n' \Gamma_n^{-1} \gamma_n, \end{aligned}$$

utilizando que $\text{Cov}(X) = \Gamma_n$ y que Γ_n es simétrica. Entonces, el error cuadrático medio es

$$v_n = E(X_{n+1} - \hat{X}_{n+1})^2 = \gamma(0) - \gamma_n' \Gamma_n^{-1} \gamma_n.$$

Ecuaciones para predictores de h pasos, $h \geq 1$

El mejor predictor lineal de X_{n+h} , $h \geq 1$, en términos de X_1, \dots, X_n se halla de la misma forma que en el caso de un solo paso. Por tanto,

$$\hat{X}_{n+h} = P_{H_n} X_{n+h} = \phi_{n1}^{(h)} X_n + \dots + \phi_{nn}^{(h)} X_1, \quad n, h \geq 1, \quad (3.1)$$

donde $\Phi_n^{(h)} = (\phi_{n1}^{(h)}, \dots, \phi_{nn}^{(h)})'$ es solución (única si Γ_n es no singular) de $\Gamma_n \Phi_n^{(h)} = \gamma_n^{(h)}$, con $\gamma_n^{(h)} = (\gamma(h), \gamma(h+1), \dots, \gamma(n+h-1))'$.

En los dos casos el cálculo de la predicción requiere la solución directa del sistema lineal

$$\Gamma_n \Phi_n^{(h)} = \gamma_n^{(h)},$$

que para un n grande es un trabajo muy laborioso. Para no tener que resolver un sistema en cada paso podemos utilizar un algoritmo recursivo como los dos que presentamos a continuación, cuya idea general es utilizar los predictores basados en n observaciones para computar los predictores basados en $n+1$ observaciones, $n = 1, 2, \dots$

3.0.1. Algoritmo de Durbin-Levinson

El algoritmo de Durbin-Levinson permite el cálculo de los coeficientes ϕ_{nj} , $j = 1, \dots, n$, de forma recursiva. El fundamento es el siguiente resultado.

Teorema 3.0.1. Si $\{X_t\}$ es un proceso estacionario de media cero, con función de autocovarianza γ y Γ_n es no singular, entonces los coeficientes ϕ_{nj} , $j = 1, \dots, n$, definidos por

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

y los errores cuadráticos medios

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

satisfacen

$$\phi_{nn} = \left[\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j} \gamma(n-j) \right] v_{n-1}^{-1} \quad \text{con} \quad \phi_{11} = \frac{\gamma(1)}{\gamma(0)},$$

$$\begin{bmatrix} \phi_{n1} \\ \vdots \\ \phi_{n,n-1} \end{bmatrix} = \begin{bmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{bmatrix} - \phi_{nn} \begin{bmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{bmatrix}$$

y

$$v_n = v_{n-1} [1 - \phi_{nn}^2] \quad \text{con} \quad v_0 = \gamma(0).$$

Demostración. Por definición de P_{K_1} , $K_1 = \overline{\text{sp}}\{X_2, \dots, X_n\}$ y $K_2 = \overline{\text{sp}}\{X_1 - P_{K_1}X_1\}$ son subespacios ortogonales de $H_n = \overline{\text{sp}}\{X_1, \dots, X_n\}$. Entonces podemos escribir

$$\hat{X}_{n+1} = P_{K_1}X_{n+1} + P_{K_2}X_{n+1} = P_{K_1}X_{n+1} + a(X_1 - P_{K_1}X_1) \quad (3.2)$$

donde

$$a = \frac{\langle X_{n+1}, X_1 - P_{K_1}X_1 \rangle}{\|X_1 - P_{K_1}X_1\|^2}. \quad (3.3)$$

Ahora por ser $\{X_t\}$ un proceso estacionario, $(X_1, \dots, X_n)'$ tiene la misma matriz de covarianza que $(X_n, X_{n-1}, \dots, X_1)'$ y $(X_2, \dots, X_{n+1})'$ por lo que

$$P_{K_1}X_1 = \sum_{j=1}^{n-1} \phi_{n-1,j}X_{j+1}, \quad P_{K_1}X_{n+1} = \sum_{j=1}^{n-1} \phi_{n-1,j}X_{n+1-j} \quad (3.4)$$

y

$$\|X_1 - P_{K_1}X_1\|^2 = \|X_{n+1} - P_{K_1}X_{n+1}\|^2 = \|X_n - \hat{X}_n\|^2 = v_{n-1}. \quad (3.5)$$

De las ecuaciones (3.2) y (3.4) obtenemos

$$\hat{X}_{n+1} = aX_1 + \sum_{j=1}^{n-1} (\phi_{n-1,j} - a\phi_{n-1,n-j})X_{n+1-j} \quad (3.6)$$

donde, a partir de la definición de a (ecuación (3.3)) y de (3.4),

$$\begin{aligned} a &= \left(\langle X_{n+1}, X_1 \rangle - \sum_{j=1}^{n-1} \phi_{n-1,j} \langle X_{n+1}, X_{j+1} \rangle \right) v_{n-1}^{-1} \\ &= \left(\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j} \gamma(n-j) \right) v_{n-1}^{-1}. \end{aligned}$$

Además, como por hipótesis Γ_n es no singular, la representación

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

es única y comparando coeficientes con (3.6) deducimos que

$$\phi_{nn} = a, \quad (3.7)$$

$$\phi_{nj} = \phi_{n-1,j} - a\phi_{n-1,n-j}, \quad j = 1, \dots, n-1.$$

Por otra parte, el error cuadrático medio del predictor \hat{X}_{n+1} es

$$\begin{aligned} v_n &= E(X_{n+1} - \hat{X}_{n+1})^2 = \|X_{n+1} - \hat{X}_{n+1}\|^2 \\ &= \|X_{n+1} - P_{K_1}X_{n+1} - P_{K_2}X_{n+1}\|^2 \\ &= \|X_{n+1} - P_{K_1}X_{n+1}\|^2 + \|P_{K_2}X_{n+1}\|^2 - 2\langle X_{n+1} - P_{K_1}X_{n+1}, P_{K_2}X_{n+1} \rangle \\ &= v_{n-1} + a^2v_{n-1} - 2a\langle X_{n+1}, X_1 - P_{K_1}X_1 \rangle, \end{aligned}$$

donde hemos utilizado la igualdad (3.5), la ortogonalidad de K_1 y K_2 , y $P_{K_2}X_{n+1} = a(X_1 - P_{K_1}X_1)$. Finalmente a partir de la definición de a (ecuación (3.3)) obtenemos $v_n = v_{n-1}(1 - a^2)$. \square

Este algoritmo procede de la siguiente manera: cuando $n = 1$ la predicción de X_2 es

$$\hat{X}_2 = \phi_{11}X_1,$$

donde

$$\phi_{11} = \Phi_1 = \Gamma_1^{-1}\gamma_1 = \frac{\gamma(1)}{\gamma(0)}.$$

El error cuadrático medio es

$$v_1 = v_0[1 - \phi_{11}^2] = \gamma(0) \left[1 - \left(\frac{\gamma(1)}{\gamma(0)} \right)^2 \right] = \gamma(0) - \frac{\gamma(1)^2}{\gamma(0)}.$$

A partir de estos resultados, puedo calcular los coeficientes para $n = 2$ aplicando las fórmulas anteriores sin tener que resolver ningún sistema. La predicción de X_3 es

$$\hat{X}_3 = \phi_{21}X_2 + \phi_{22}X_1,$$

con

$$\begin{aligned} \phi_{22} &= [\gamma(2) - \phi_{11}\gamma(1)]v_1^{-1} = \left[\gamma(2) - \frac{\gamma(1)}{\gamma(0)}\gamma(1) \right] \left[\gamma(0) - \frac{\gamma(1)^2}{\gamma(0)} \right]^{-1} \\ &= \left[\frac{\gamma(0)\gamma(2) - \gamma(1)^2}{\gamma(0)} \right] \left[\frac{\gamma(0)^2 - \gamma(1)^2}{\gamma(0)} \right]^{-1} = \frac{\gamma(0)\gamma(2) - \gamma(1)^2}{\gamma(0)^2 - \gamma(1)^2} \end{aligned}$$

y

$$\phi_{21} = \phi_{11} - \phi_{22}\phi_{11} = \frac{\gamma(1)}{\gamma(0)} \left[1 - \frac{\gamma(0)\gamma(2) - \gamma(1)^2}{\gamma(0)^2 - \gamma(1)^2} \right] = \frac{\gamma(1)[\gamma(0) - \gamma(2)]}{\gamma(0)^2 - \gamma(1)^2}.$$

El error cuadrático medio es

$$v_2 = v_1[1 - \phi_{22}^2] = \gamma(0) - \frac{\gamma(1)^2}{\gamma(0)} \left[1 - \left(\frac{\gamma(0)\gamma(2) - \gamma(1)^2}{\gamma(0)^2 - \gamma(1)^2} \right)^2 \right].$$

Entonces de forma recursiva calculamos las predicciones sin tener que hacer grandes cálculos. Tal como se mencionó antes todos estos cálculos son útiles si conociésemos la función de autocovarianza. En el Capítulo 4 se discuten distintos métodos de estimación de esta función.

El algoritmo de Durbin-Levinson también sirve para calcular los coeficientes de autocorrelación parcial de orden n de un proceso $\{X_t\}$ que son los ϕ_{nn} , $n \geq 1$, es decir, probemos que

$$\phi_{nn} = \alpha(n), \quad n \geq 1.$$

Por las ecuaciones (3.7), (3.3) y (3.5) tenemos que

$$\begin{aligned} \phi_{nn} &= \frac{\langle X_{n+1}, X_1 - P_{K_1} X_1 \rangle}{\|X_1 - P_{K_1} X_1\|^2} = \frac{\langle X_{n+1} - P_{K_1} X_{n+1}, X_1 - P_{K_1} X_1 \rangle}{\|X_1 - P_{K_1} X_1\|^2} \\ &= \frac{\langle X_{n+1} - P_{K_1} X_{n+1}, X_1 - P_{K_1} X_1 \rangle}{\|X_{n+1} - P_{K_1} X_{n+1}\| \cdot \|X_1 - P_{K_1} X_1\|} \\ &= \text{Corr}(X_{n+1} - P_{K_1} X_{n+1}, X_1 - P_{K_1} X_1) = \alpha(n), \end{aligned}$$

donde en la segunda igualdad hemos utilizado que $P_{K_1} X_{n+1} \perp (X_1 - P_{K_1} X_1)$.

3.0.2. Algoritmo de innovaciones

El algoritmo de innovaciones es otro algoritmo recursivo que calcula los predictores basados en $n+1$ observaciones a partir de los predictores basados en n observaciones, $n \geq 1$.

Este algoritmo se aplica a procesos $\{X_t\}$ con media cero y función de autocovarianza $\kappa(i, j) = \langle X_i, X_j \rangle = E(X_i X_j)$ no necesariamente estacionarios. Definimos para $n \geq 1$ como antes, $H_n = \overline{\text{span}}\{X_1, \dots, X_n\}$, los predictores de un paso

$$\hat{X}_n = \begin{cases} 0 & \text{si } n = 1 \\ P_{H_{n-1}} X_n & n = 2, 3, \dots \end{cases}$$

y sus errores cuadráticos medios $v_{n-1} = E(X_n - \hat{X}_n)^2$. Se definen los errores de predicción de un paso o *innovaciones* como $U_n = X_n - \hat{X}_n$. Como $\hat{X}_n \in H_{n-1}$ entonces $U_n \in H_n$ y podemos escribir

$$U_n = a_{n-1, n-1} X_1 + a_{n-1, n-2} X_2 + \dots + a_{n-1, 1} X_{n-1} + a_{n-1, 0} X_n,$$

ecuaciones que pueden escribirse de forma matricial de la forma

$$\mathbf{U}_n = A_n \mathbf{X}_n$$

donde $\mathbf{U}_n = (U_1, \dots, U_n)'$, $\mathbf{X}_n = (X_1, \dots, X_n)'$ y la matriz A_n tiene estructura

$$A_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ a_{11} & 1 & 0 & \cdots & 0 \\ a_{22} & a_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n-1,n-1} & a_{n-1,n-2} & a_{n-1,n-3} & \cdots & 1 \end{bmatrix}.$$

Si denotamos por C_n a la matriz inversa de A_n , cuya estructura también es triangular,

$$C_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \theta_{11} & 1 & 0 & \cdots & 0 \\ \theta_{22} & \theta_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \cdots & 1 \end{bmatrix},$$

y el vector de proyecciones $\hat{\mathbf{X}}_n = (X_1, P_1 X_2, \dots, P_{n-1} X_n)'$ entonces tenemos que

$$\hat{\mathbf{X}}_n = \mathbf{X}_n - \mathbf{U}_n = C_n \mathbf{U}_n - \mathbf{U}_n = (C_n - I) \mathbf{U}_n = \Theta_n (\mathbf{X}_n - \hat{\mathbf{X}}_n),$$

donde I es la matriz identidad de orden n y

$$\Theta_n = C_n - I = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ \theta_{11} & 0 & 0 & \cdots & 0 \\ \theta_{22} & \theta_{21} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \cdots & 0 \end{bmatrix}.$$

Entonces podemos escribir

$$\hat{X}_{n+1} = \begin{cases} 0, & \text{si } n = 0, \\ \sum_{j=1}^n \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}), & \text{si } n = 1, 2, \dots \end{cases} \quad (3.8)$$

Si conociésemos los θ_{nj} , $j = 1, \dots, n$, podríamos utilizar (3.8) para calcular \hat{X}_{n+1} de forma recursiva. El siguiente resultado proporciona un medio para calcular esos coeficientes $\{\theta_{nj}, j = 1, \dots, n\}$, $n \geq 1$, y los errores de predicción $\{v_n\}$, $n \geq 1$, de forma recursiva a partir de la función de autocovarianza del proceso.

Teorema 3.0.2. Si el proceso $\{X_t\}$ tiene media cero y la matriz de covarianzas $[\kappa(i, j)]_{i, j=1}^n$, $\kappa(i, j) = E(X_i X_j)$, es no singular para cada $n \geq 1$, entonces

$$\begin{cases} v_0 = \kappa(1, 1) \\ \theta_{n, n-k} = v_k^{-1} \left(\kappa(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k, k-j} \theta_{n, n-j} v_j \right), & k = 0, 1, \dots, n-1, \\ v_n = \kappa(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n, n-j}^2 v_j. \end{cases}$$

Demostración. El conjunto $\{X_1 - \hat{X}_1, X_2 - \hat{X}_2, \dots, X_n - \hat{X}_n\}$ es ortogonal ya que $(X_i - \hat{X}_i) \in H_{j-1}$ para $i < j$ y $(X_j - \hat{X}_j) \perp H_{j-1}$ por definición de \hat{X}_j . Entonces haciendo el producto interno en ambos lados de la igualdad (3.8) con $X_{k+1} - \hat{X}_{k+1}$, $0 \leq k < n$, tenemos $\langle \hat{X}_{n+1}, X_{k+1} - \hat{X}_{k+1} \rangle = \theta_{n, n-k} v_k$, o equivalentemente, $\theta_{n, n-k} = v_k^{-1} \langle X_{n+1}, X_{k+1} - \hat{X}_{k+1} \rangle$. Haciendo uso de la representación (3.8) reemplazando n por k obtenemos

$$\begin{aligned} \theta_{n, n-k} &= v_k^{-1} \left(\kappa(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k, k-j} \langle X_{n+1}, X_{j+1} - \hat{X}_{j+1} \rangle \right) \\ &= v_k^{-1} \left(\kappa(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k, k-j} \theta_{n, n-j} v_j \right). \end{aligned}$$

Por el Teorema de la Proyección,

$$\begin{aligned} v_n &= E(X_{n+1} - \hat{X}_{n+1})^2 = \|X_{n+1} - \hat{X}_{n+1}\|^2 = \|X_{n+1}\|^2 - \|\hat{X}_{n+1}\|^2 \\ &= \kappa(n+1, n+1) - \sum_{k=0}^{n-1} \theta_{n, n-k}^2 v_k, \end{aligned}$$

concluyendo la demostración. □

Con este teorema obtenemos los coeficientes θ_{nj} , $j = 1, \dots, n$, resolviendo las ecuaciones recursivamente en el orden: $v_0; \theta_{11}, v_1; \theta_{22}, \theta_{21}, v_2; \theta_{33}, \theta_{32}, \theta_{31}, v_3; \dots$

El algoritmo de innovaciones procede de la siguiente forma. Para $n = 1$ tenemos por (3.8) que

$$\hat{X}_2 = \theta_{11}(X_1 - \hat{X}_1) = \theta_{11}X_1,$$

donde el coeficiente θ_{11} lo obtenemos utilizando el teorema anterior, es decir,

$$\theta_{11} = v_0^{-1} \kappa(2, 1) = \frac{\kappa(2, 1)}{\kappa(1, 1)}.$$

Si $n = 2$ por (3.8) tenemos que

$$\hat{X}_3 = \theta_{21}(X_2 - \hat{X}_2) + \theta_{22}(X_1 - \hat{X}_1)$$

donde los coeficientes θ_{21} y θ_{22} los obtenemos a partir del teorema anterior en el orden antes indicado. Primero calculamos el error cuadrático medio para $n = 1$,

$$v_1 = \kappa(2, 2) - \theta_{1,1}^2 v_0 = \kappa(2, 2) - \frac{\kappa(2, 1)^2}{\kappa(1, 1)^2} \kappa(1, 1) = \kappa(2, 2) - \frac{\kappa(2, 1)^2}{\kappa(1, 1)}.$$

A continuación calculamos los dos coeficientes,

$$\theta_{2,2} = v_0^{-1} \kappa(3, 1) = \frac{\kappa(3, 1)}{\kappa(1, 1)},$$

y

$$\begin{aligned} \theta_{2,1} &= v_1^{-1} (\kappa(3, 2) - \theta_{1,1} \theta_{2,2} v_0) \\ &= \left(\kappa(2, 2) - \frac{\kappa(2, 1)^2}{\kappa(1, 1)} \right)^{-1} \left(\kappa(3, 2) - \frac{\kappa(2, 1) \kappa(3, 1)}{\kappa(1, 1) \kappa(1, 1)} \kappa(1, 1) \right) \\ &= \frac{\kappa(1, 1) [\kappa(3, 2) - \kappa(2, 1) \kappa(3, 1)]}{\kappa(1, 1) \kappa(2, 2) - \kappa(2, 1)^2}. \end{aligned}$$

Luego el algoritmo de innovaciones nos permite calcular los predictores sin tener que resolver grandes sistemas de ecuaciones.

Cálculo recursivo de los predictores de h pasos, $h \geq 1$

A partir de ahora utilizaremos la notación P_n para referirnos a la proyección P_{H_n} . Por las propiedades de la proyección, para $h \geq 1$, y utilizando que $(X_{n+h-j} - \hat{X}_{n+h-j}) \perp H_n$ para $j < h$ tenemos que

$$\begin{aligned} \hat{X}_{n+h} &= P_n X_{n+h} = P_n P_{n+h-1} X_{n+h} = P_n \hat{X}_{n+h} \\ &= P_n \left(\sum_{j=1}^{n+h-1} \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j}) \right) \\ &= \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j}), \end{aligned} \quad (3.9)$$

donde los coeficientes θ_{nj} están determinados como en el algoritmo de innovaciones. Además, el error cuadrático medio es

$$\begin{aligned} v_n &= E(X_{n+h} - \hat{X}_{n+h})^2 = \|X_{n+h}\|^2 - \|\hat{X}_{n+h}\|^2 \\ &= \kappa(n+h, n+h) - \sum_{j=1}^{n+h-1} \theta_{n+h-1,j}^2 v_{n+h-j-1}. \end{aligned}$$

Ambos procedimientos, Durbin-Levinson e innovaciones, nos permiten calcular la predicción del proceso en el instante $n + 1$ a partir de los valores anteriores de forma recursiva. El algoritmo de Durbin-Levinson requiere almacenar menos coeficientes pero, al menos en el caso de los procesos ARMA, el algoritmo de innovaciones es más fácil de utilizar porque obtenemos expresiones más sencillas como veremos a continuación. En todo caso, la elección de uno u otro algoritmo lo debatiremos en el siguiente capítulo.

3.1. Predicción de un proceso ARMA(p, q)

El algoritmo de innovaciones nos proporciona un método recursivo para predecir el proceso causal ARMA(p, q)

$$\phi(B)X_t = \theta(B)Z_t, \quad \{Z_t\} \sim WN(0, \sigma^2), \quad (3.10)$$

conocidos los polinomios $\phi(z)$ y $\theta(z)$, y la varianza σ^2 . Para simplificar los cálculos el algoritmo no se aplica directamente al proceso $\{X_t\}$ sino al proceso transformado

$$W_t = \begin{cases} \sigma^{-1}X_t, & t = 1, \dots, m, \\ \sigma^{-1}\phi(B)X_t, & t > m, \end{cases} \quad \text{donde } m = \max(p, q). \quad (3.11)$$

Sin pérdida de generalidad, definimos $\theta_0 = 1$ y asumimos $p \geq 1$ y $q \geq 1$. Denotamos, como antes, para $n \geq 1$, $H_n = \overline{\text{sp}}\{X_1, \dots, X_n\} = \overline{\text{sp}}\{W_1, \dots, W_n\}$, \hat{X}_{n+1} y \hat{W}_{n+1} las proyecciones de X_{n+1} y W_{n+1} en H_n , y $\hat{X}_1 = \hat{W}_1 = 0$. Las autocovarianzas $\kappa(i, j) = E(W_i W_j)$ son

$$\kappa(i, j) = \begin{cases} \sigma^{-2}\gamma_X(i - j), & 1 \leq i, j \leq m, \\ \sigma^{-2}[\gamma_X(i - j) - \sum_{r=1}^p \phi_r \gamma_X(r - |i - j|)] & \min(i, j) \leq m < \max(i, j) \leq 2m, \\ \sum_{r=0}^q \theta_r \theta_{r+|i-j|}, & \min(i, j) > m, \\ 0, & \text{en otro caso,} \end{cases} \quad (3.12)$$

donde convenimos $\theta_j = 0$ para $j > q$ y γ_X es la función de autocovarianza del proceso $\{X_t\}$ calculada en el Capítulo 1 (ecuación (1.12)).

Aplicando el algoritmo de innovaciones al proceso $\{W_t\}$ obtenemos

$$\hat{W}_{n+1} = \begin{cases} \sum_{j=1}^n \theta_{nj}(W_{n+1-j} - \hat{W}_{n+1-j}), & 1 \leq n < m, \\ \sum_{j=1}^q \theta_{nj}(W_{n+1-j} - \hat{W}_{n+1-j}), & n \geq m, \end{cases} \quad (3.13)$$

donde los coeficientes θ_{nj} y los errores cuadráticos medios $r_n = E(W_{n+1} - \hat{W}_{n+1})^2$ se hallan recursivamente con el algoritmo de innovaciones con κ definida como antes. Una característica notable de estos predictores es la desaparición de θ_{nj} cuando $n \geq m$ y $j > q$. Esto es consecuencia del algoritmo de innovaciones y del hecho de que $\kappa(n, j) = 0$ si $n > m$ y $|n - j| > q$.

Proyectando cada lado de la ecuación (3.11) sobre el subespacio H_{t-1} y utilizando la linealidad de P_n obtenemos

$$\hat{W}_t = \begin{cases} \sigma^{-1} \hat{X}_t, & t = 1, \dots, m, \\ \sigma^{-1}(\hat{X}_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p}), & t > m, \end{cases} \quad (3.14)$$

que, junto con la ecuación (3.11), muestra que $X_t - \hat{X}_t = \sigma(W_t - \hat{W}_t)$ para todo $t \geq 1$. Reemplazando $(W_j - \hat{W}_j)$ por $\sigma^{-1}(X_t - \hat{X}_t)$ en (3.13) y sustituyendo en (3.14) obtenemos

$$\hat{X}_{n+1} = \begin{cases} \sum_{j=1}^n \theta_{nj}(X_{n+1-j} - \hat{X}_{n+1-j}), & 1 \leq n < m, \\ \phi_1 X_n + \dots + \phi_p X_{n+1-p} + \sum_{j=1}^q \theta_{nj}(X_{n+1-j} - \hat{X}_{n+1-j}), & n \geq m, \end{cases} \quad (3.15)$$

y $v_n = E(X_{n+1} - \hat{X}_{n+1})^2 = \sigma^2 E(W_{n+1} - \hat{W}_{n+1})^2 = \sigma^2 r_n$, donde θ_{nj} y r_n se han obtenido anteriormente aplicando el algoritmo de innovaciones a $\{W_t\}$. Estas últimas ecuaciones determinan los predictores de un paso recursivamente.

Predictores de h pasos, $h \geq 1$, de un proceso ARMA(p,q)

De la ecuación (3.9) obtenemos

$$\begin{aligned} \hat{W}_{n+h} &= P_n W_{n+h} = \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}(W_{n+h-j} - \hat{W}_{n+h-j}) \\ &= \sigma^{-1} \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}(X_{n+h-j} - \hat{X}_{n+h-j}). \end{aligned}$$

Utilizando este resultado y aplicando el operador proyección P_n a cada lado de las ecuaciones (3.11), concluimos que los predictores de h pasos satisfacen

$$\hat{X}_{n+h} = \begin{cases} \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}(X_{n+h-j} - \hat{X}_{n+h-j}), & 1 \leq h \leq m - n, \\ \sum_{i=1}^p \phi_i P_n X_{n+h-i} + \sum_{j=h}^q \theta_{n+h-1,j}(X_{n+h-j} - \hat{X}_{n+h-j}), & h > m - n. \end{cases}$$

Si, como ocurre siempre en la práctica, $n > m = \max(p, q)$, entonces para todo $h \geq 1$ tenemos

$$\hat{X}_{n+h} = \sum_{i=1}^p \phi_i \hat{X}_{n+h-i} + \sum_{j=h}^q \theta_{n+h-1,j}(X_{n+h-j} - \hat{X}_{n+h-j}).$$

donde el segundo término es cero si $h > q$.

Entonces, una vez calculados los predictores $\hat{X}_1, \dots, \hat{X}_n$ con (3.15), es sencillo determinar, con n fijo, los predictores $\hat{X}_{n+1}, \hat{X}_{n+2}, \dots$, recursivamente a partir de las ecuaciones anteriores.

Capítulo 4

Inferencia en modelos ARMA

Se dedica un capítulo a estudiar inferencia en modelos ARMA porque según se ha visto en el Capítulo 2 son un conjunto de modelos suficientemente rico como para aproximar cualquier otro proceso regular con un error pequeño (corolarios 2.2.5 y 2.2.6).

Antes de empezar a modelar una serie temporal mediante un modelo ARMA se deben cumplir ciertas condiciones, como la estacionaridad y la ausencia de componentes estacionales. Si no se cumplen debemos hacer una transformación previa de la serie, la cual estudiaremos en la Sección 4.1. Una vez transformada (si es necesario) podemos empezar a identificar el tipo de modelo más adecuado para predecir los datos observados.

Lo primero es llevar a cabo una estimación preliminar de los coeficientes $\{\phi_i, i = 1, \dots, p\}$ y $\{\theta_i, i = 1, \dots, q\}$ del modelo $\text{ARMA}(p, q)$ y la varianza σ^2 del ruido blanco, con los órdenes p y q fijados (Sección 4.2). Estas estimaciones nos servirán para inicializar el algoritmo de maximización de la función de verosimilitud, es decir, para encontrar los estimadores de máxima verosimilitud. Para que la búsqueda del mejor modelo sea más rápida, las estimaciones preliminares tienen que estar lo más cerca posible de las estimaciones de máxima verosimilitud. La evaluación de la función de verosimilitud es complicada y en la Sección 4.3 se describe como hacerlo utilizando los algoritmos de predicción descritos en el Capítulo 3. La maximización de esta función requiere métodos de optimización numérica que están fuera del objetivo de esta memoria.

Una vez obtenidas las estimaciones de los coeficientes del modelo se utilizan distintos procedimientos para valorar la adecuación del modelo. Parte de los procedimientos se basan en el análisis de los residuales, tal como se describe en la Sección 4.4.

Todos este procedimiento se efectúa para distintos valores de p y q . En

general, un modelo más complejo (con valores altos de p y q) se ajustará mejor a los datos observados, pero puede tener peor capacidad predictiva. En la Sección 4.5 se discuten métodos orientados a una selección adecuada de los órdenes p y q .

Por último, el modelo seleccionado se utilizará para realizar pronósticos de valores futuros de la serie temporal observada. Esta predicción se lleva a cabo mediante las técnicas descritas en el Capítulo 3.

Este capítulo tiene un carácter diferente de los anteriores. Aquí se presentan de forma resumida algunas de las técnicas de uso frecuente en el modelado de series temporales, sin discutir los principales resultados matemáticos que los justifican, más allá de los presentados en capítulos anteriores.

4.1. Preprocesado: estacionaridad y estacionariedad

Para modelar una serie temporal x_1, \dots, x_n mediante un modelo ARMA de órdenes p y q tienen que ser asumibles ciertas condiciones. Como ya conocemos, los modelos ARMA son estacionarios. Esto quiere decir que si quiero modelar una serie temporal mediante un modelo ARMA, la serie no puede presentar tendencia en media o en varianza. En caso contrario tenemos que aplicar a la serie alguna transformación para conseguir la estacionariedad. Un procedimiento habitual es diferenciar la serie. El operador de diferencia ∇ funciona de la siguiente manera. Si lo aplicamos una vez al proceso X_t , es decir, aplicamos una diferenciación, obtenemos

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t,$$

donde B es el operador de retardo definido en el Capítulo 1. Si lo aplicamos j veces obtenemos

$$\nabla^j X_t = (1 - B)^j X_t.$$

Algunas de sus propiedades son $\nabla^j X_t = \nabla(\nabla^{j-1} X_t)$ y $\nabla^0 X_t = X_t$. La diferenciación puede corregir algunas formas de no estacionariedad. Por ejemplo, si \tilde{X}_t un proceso formado por una parte ARMA, X_t , y una parte con tendencia lineal, es decir,

$$\tilde{X}_t = X_t + \mu t,$$

y aplicamos una diferenciación entonces obtenemos

$$\nabla \tilde{X}_t = \tilde{X}_t - \tilde{X}_{t-1} = \mu + X_t - X_{t-1},$$

que es un proceso estacionario. Si en vez de tener tendencia lineal fuera cuadrática, es decir,

$$\tilde{X}_t = X_t + \mu t^2,$$

donde X_t es un proceso ARMA aplicaremos dos diferenciaciones. Con la primera diferenciación obtenemos

$$\nabla \tilde{X}_t = \mu(2t - 1) + X_t - X_{t-1},$$

y con la segunda

$$\nabla^2 \tilde{X}_t = 2\mu + X_t - 2X_{t-1} + X_{t-2},$$

que es estacionario, como ocurría en el primer ejemplo.

Una tarea, de la cual no hablamos en esta memoria, es tratar de encontrar el grado de diferenciación apropiado que nos lleve a un proceso estacionario.

Por la discusión inicial en el Capítulo 2 sabemos que la función de autocovarianza de un proceso con componentes estacionales, es decir, un proceso de la forma (2.1) no tiene densidad espectral porque su función de distribución es de salto. Además, si tenemos dos procesos estacionarios independientes, su suma tiene como función de autocovarianza la suma de las funciones de autocovarianza de cada proceso. Entonces si a un proceso ARMA le sumamos una componente estacional independiente, el resultado no tendrá densidad espectral porque el proceso ARMA sí tiene pero la componente estacional no. Luego si queremos modelar una serie temporal mediante un proceso ARMA tenemos que desestacionalizarla, es decir, eliminar las componentes estacionales. El periodograma, que trataremos más adelante en esta sección, es una herramienta útil para este fin. Tal como veremos el periodograma (suavizado) es un estimador consistente de la densidad espectral. Los puntos de no diferenciabilidad de la función de distribución espectral asociados a la presencia de componentes estacionales producen 'picos' en el periodograma y esto ayuda a encontrar el tipo de diferenciación estacional que debemos aplicar. El operador de diferenciación estacional de orden d , ∇_d , se define por

$$\nabla_d(X_t) = X_t - X_{t-d} = (1 - B^d)X_t,$$

donde B es el operador de retardo definido en el Capítulo 1. Un ejemplo sobre como desestacionalizar una serie es el siguiente. Sea \tilde{X}_t una serie mensual formada por una parte ARMA, X_t , y otra parte con periodicidad anual, es decir, de la forma

$$\tilde{X}_t = X_t + A \cos\left(\frac{\pi}{6}t\right), \quad (4.1)$$

donde A es una constante. Como la serie es mensual y la parte estacional tiene periodicidad anual, el proceso se repite cada 12 meses y entonces tenemos que aplicar una diferenciación estacional de orden 12. Así obtenemos un proceso

$$\nabla_{12}(\tilde{X}_t) = X_t - X_{t-12},$$

en el que hemos eliminado la componente estacional, por lo que puede ser modelado por un proceso ARMA.

4.1.1. Periodograma

La palabra periodograma fue inventada por Arthur Schuster en 1898. Originalmente se utilizó para detectar y estimar la amplitud de una componente sinusoidal, de frecuencia conocida, 'enterrada' en ruido (ver [3], página 35).

En principio el periodograma se plantea como un estimador de la densidad espectral pero en esta subsección veremos que no es un estimador consistente. Una manera de solucionar este problema es promediando las estimaciones del periodograma (periodograma suavizado).

El periodograma de una secuencia de observaciones x_1, \dots, x_n consiste simplemente en su transformada discreta de Fourier. Sea $x = [x_1, \dots, x_n]'$ el vector de valores complejos formado por las observaciones. Se definen las *frecuencias de Fourier* como

$$\omega_k = \frac{2\pi k}{n},$$

con

$$k \in F_n = \{j \in \mathbb{Z} : -\pi < \omega_j \leq \pi\} = \left\{ -\left[\frac{n-1}{2} \right], \dots, \left[\frac{n}{2} \right] \right\}$$

donde $[y]$ denota la parte entera de y . Correspondientemente, sean

$$e_k = \frac{1}{\sqrt{n}} [\exp(i\omega_k), \exp(2i\omega_k), \dots, \exp(ni\omega_k)]', \quad k \in F_n,$$

vectores ortonormales que forman una base de \mathbb{C}^n . Entonces podemos expresar el vector x como

$$x = \sum_{s=-[(n-1)/2]}^{[n/2]} a_s e_s, \quad (4.2)$$

donde los coeficientes a_k se obtienen multiplicando a ambos lados de esta igualdad por e_k' y utilizando que los $\{e_k\}$ forman una base ortonormal. Luego $a_k = e_k' x = \frac{1}{\sqrt{n}} \sum_{t=1}^n x_t \exp(-it\omega_k)$. La secuencia $\{a_k\}$ es conocida como la *transformada discreta de Fourier* de $\{x_1, \dots, x_n\}$.

Definición 4.1.1. Se define el *periodograma* de $\{x_1, \dots, x_n\}$ como la función

$$I_n(\lambda) = \frac{1}{n} \left| \sum_{t=1}^n x_t \exp(-it\lambda) \right|^2. \quad (4.3)$$

Ahora bien, si λ es una de las frecuencias de Fourier, entonces $I_n(\omega_k) = |a_k|^2$ y, a partir de la ecuación (4.2) y utilizando que $\{e_k\}$ es una base ortonormal obtenemos

$$\sum_{t=1}^n |x_t|^2 = x'x = \sum_{k=-[(n-1)/2]}^{[n/2]} |a_k|^2 = \sum_{k=-[(n-1)/2]}^{[n/2]} I_n(\omega_k).$$

El valor del periodograma en la frecuencia ω_k es por lo tanto la contribución a esta suma de cuadrados del término $a_k e_k$ de la 'frecuencia ω_k ' en (4.2).

En la siguiente proposición veremos que el periodograma de una secuencia de observaciones x_1, \dots, x_n se puede escribir en términos de la *función de autocovarianza muestral* que se define como

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

donde $\bar{x}_n = n^{-1}(x_1 + x_2 + \dots + x_n)$ es la *media muestral*. Esta forma de escribirlo nos ayudará a compararlo con la densidad espectral de un proceso estacionario.

Proposición 4.1.2. Si x_1, \dots, x_n son números reales y ω_k es una frecuencia de Fourier distinta de cero, entonces

$$I_n(\omega_k) = \sum_{|h|<n} \hat{\gamma}(h) \exp(-ih\omega_k), \quad (4.4)$$

donde $\hat{\gamma}(h)$ es la función de autocovarianza muestral de x_1, \dots, x_n .

Demostración. Por la definición 4.1.1,

$$I_n(\omega_k) = \frac{1}{n} \left| \sum_{t=1}^n x_t \exp(-it\omega_k) \right|^2 = \frac{1}{n} \sum_{t=1}^n x_t \exp(-it\omega_k) \sum_{s=1}^n x_s \exp(is\omega_k).$$

Ahora $\sum_{t=1}^n \exp(-it\omega_k) = \sum_{s=1}^n \exp(is\omega_k) = 0$ si $\omega_k \neq 0$ y entonces

$$\begin{aligned} I_n(\omega_k) &= \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n (x_t - \bar{x})(x_s - \bar{x}) \exp(-i(t-s)\omega_k) \\ &= \sum_{|h|<n} \hat{\gamma}(h) \exp(-ih\omega_k). \end{aligned}$$

□

La similitud entre (4.4) y la densidad espectral de un proceso estacionario con $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$,

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) \exp(-ih\lambda), \quad \lambda \in (-\pi, \pi],$$

sugiere que $\frac{I_n(\lambda)}{2\pi}$ se puede emplear como estimador de la densidad espectral $f(\lambda)$.

Para una gran clase de series temporales $\{X_t\}$ con densidad espectral estrictamente positiva, se puede mostrar que para frecuencias fijas $\lambda_1, \dots, \lambda_m$ tal que $0 < \lambda_1 < \dots < \lambda_m < \pi$, la función de distribución conjunta $F_n(x_1, \dots, x_m)$ de los valores del periodograma $(I_n(\lambda_1), \dots, I_n(\lambda_m))$ converge, cuando $n \rightarrow \infty$, a $F(x_1, \dots, x_m)$, donde

$$F(x_1, \dots, x_m) = \begin{cases} \prod_{i=1}^m \left(1 - \exp\left(\frac{-x_i}{2\pi f(\lambda_i)}\right)\right), & \text{si } x_1, \dots, x_m > 0, \\ 0, & \text{en otro caso.} \end{cases}$$

Por lo tanto para n grande las ordenadas del periodograma $(I_n(\lambda_1), \dots, I_n(\lambda_m))$ son aproximadamente distribuidas como variables aleatorias exponenciales independientes con medias $2\pi f(\lambda_1), \dots, 2\pi f(\lambda_m)$, respectivamente. En particular, para cada $\lambda \in (0, \pi)$ fijo y $\epsilon > 0$,

$$P(|I_n(\lambda) - 2\pi f(\lambda)| > \epsilon) \rightarrow p > 0, \quad \text{cuando } n \rightarrow \infty,$$

así que la probabilidad de una estimación del error mas grande que ϵ no puede hacerse arbitrariamente pequeña eligiendo un n suficientemente grande. Por lo tanto, $I_n(\lambda)$ no es un estimador consistente de $2\pi f(\lambda)$, pero podemos 'suavizarlo' para que sí lo sea. Esto se consigue promediando las estimaciones del periodograma en un pequeño intervalo de frecuencia que contenga λ bajo ciertas condiciones que exponemos a continuación.

Definición 4.1.3. El *periodograma suavizado* es

$$\hat{f}(\lambda) = \frac{1}{2\pi} \sum_{|j| \leq m_n} W_n(j) I_n(g(n, \lambda) + 2\pi j/n), \quad (4.5)$$

donde los anchos de banda m_n satisfacen $m_n \rightarrow \infty$ y $m_n/n \rightarrow 0$ cuando $n \rightarrow \infty$, y las funciones de peso $W_n(\cdot)$ satisfacen $W_n(j) = W_n(-j)$, $W_n(j) \geq 0$ para todo j , $\sum_{|j| \leq m_n} W_n(j) = 1$, y $\sum_{|j| \leq m_n} W_n^2(j) \rightarrow 0$ cuando $n \rightarrow \infty$, y $g(n, \lambda)$ es el múltiplo de $2\pi/n$ más cercano a λ .

Las condiciones de $\{m_n\}$ significan que, cuando n crece, el número de términos en el promedio ponderado (4.5) tiende a infinito mientras que el ancho de frecuencia sobre el cual se toma el promedio tiende a cero. Las condiciones de $\{W_n(\cdot)\}$ aseguran que la media y la varianza de $\hat{f}(\lambda)$ converjan, cuando $n \rightarrow \infty$, a $f(\lambda)$ y 0, respectivamente. Como consecuencia, con las condiciones anteriores $\hat{f}(\lambda) \rightarrow f(\lambda)$ en probabilidad si $n \rightarrow \infty$.

En la práctica, cuando el tamaño de muestra n es un número finito fijo, la elección de $\{m_n\}$ y $\{W_n(\cdot)\}$ implica un compromiso entre lograr un sesgo pequeño y una varianza pequeña para el estimador $\hat{f}(\lambda)$. Una función de peso que asigna pesos aproximadamente iguales a una amplia banda de frecuencias producirá una estimación de $f(\lambda)$ que, aunque uniforme, puede tener un gran sesgo, ya que la estimación de $f(\lambda)$ depende de los valores de I_n en frecuencias distantes de λ . Por otro lado, una función de peso que asigna la mayor parte de su peso a una banda de frecuencia estrecha centrada en cero proporcionará un estimador con un sesgo relativamente pequeño, pero con una varianza mayor. En la práctica, es aconsejable experimentar con un rango de funciones de peso y seleccionar el que parece lograr un equilibrio satisfactorio entre el sesgo y la varianza.

Para entender mejor el funcionamiento del periodograma, vamos a aplicar estos resultados con un ejemplo. Supongamos que tenemos un proceso AR(1) de la forma

$$X_t - 0,8X_{t-1} = Z_t,$$

con $\{Z_t\} \sim WN(0, 1)$. Por el teorema 2.2.2 del Capítulo 2 tenemos que el proceso $\{X_t\}$ tiene densidad espectral

$$\begin{aligned} f(\lambda) &= \frac{1}{2\pi} |1 - 0,8 \exp(-i\lambda)|^{-2} = \frac{1}{2\pi} (1 - 1,6 \cos(\lambda) + 0,64)^{-1} \\ &= \frac{1}{2\pi(1,64 - 1,6 \cos(\lambda))}. \end{aligned}$$

Primero, en la figura 4.1, representamos el periodograma con diferentes tamaños muestrales ($n=10$, $n=100$, $n=1000$, $n=10000$, respectivamente). Observamos que a medida que aumentamos el tamaño de la muestra, se producen más oscilaciones en el periodograma.

Ahora representamos los periodogramas en crudo y suavizado junto con la densidad espectral para comprobar si son buenos estimadores (figura 4.2). Los gráficos se han realizado con tamaño muestral $n=10000$. En el primer gráfico está representado el periodograma en crudo de la serie (color negro) y la densidad espectral (color rojo). Se puede observar que el periodograma en crudo tiene oscilaciones consecuencia del carácter exponencial.

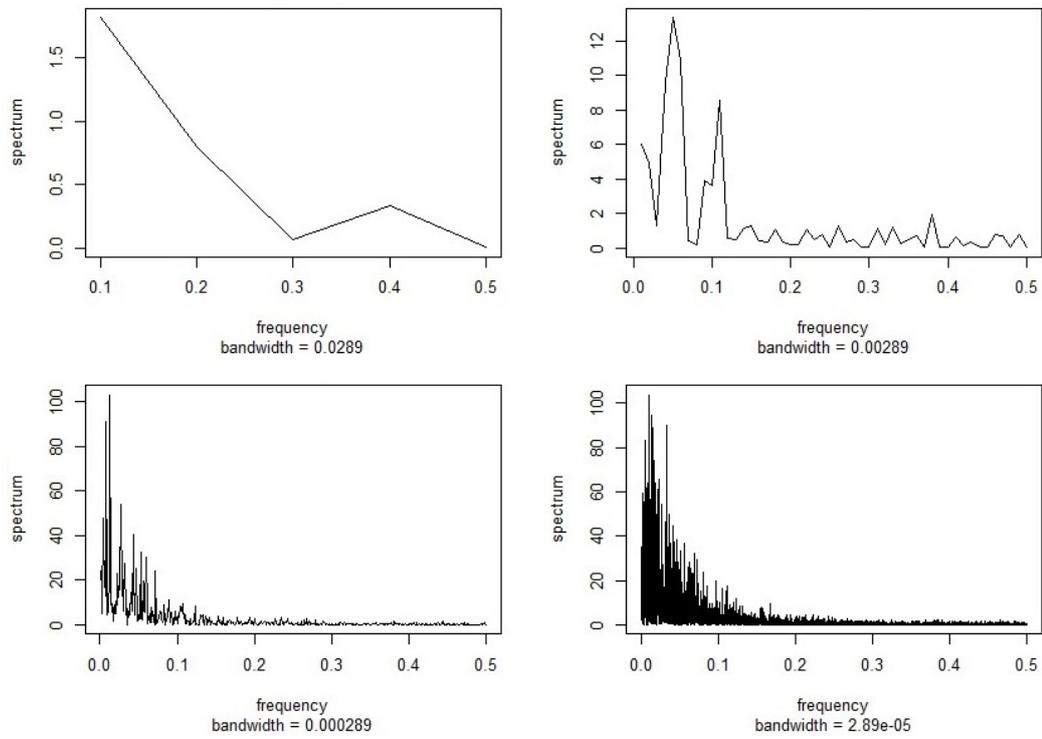


Figura 4.1: Periodograma en crudo con diferentes tamaños muestrales

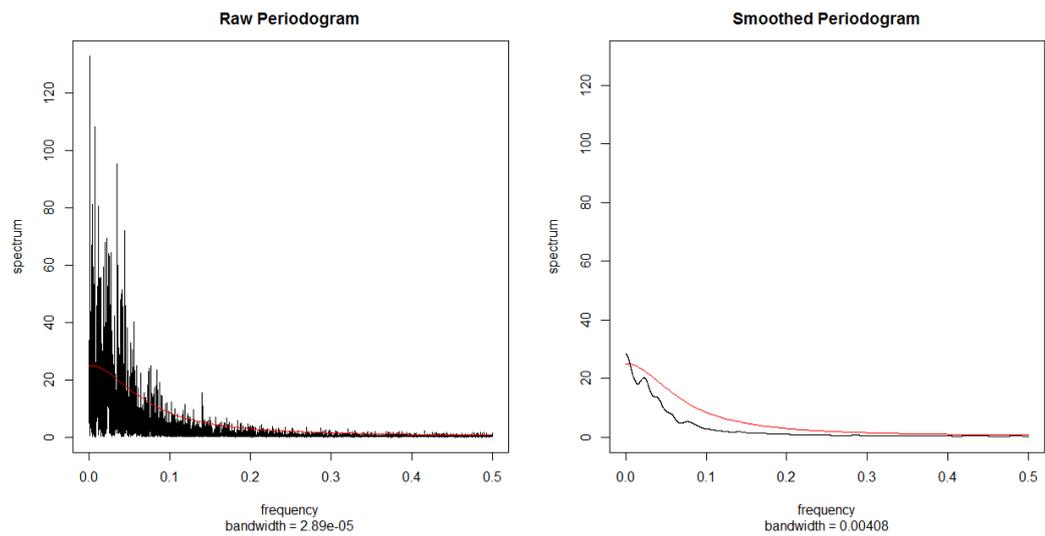


Figura 4.2: Periodograma en crudo y periodograma suavizado

En cuanto se suaviza el periodograma (segundo gráfico), observamos como el periodograma suavizado (color negro) se aproxima, quizá no tan rápido como era esperable, a la densidad espectral (color rojo). En todo caso, es claro que el suavizado no presenta tantas irregularidades como el periodograma en crudo y nos ofrece una mejor estimación de la densidad espectral.

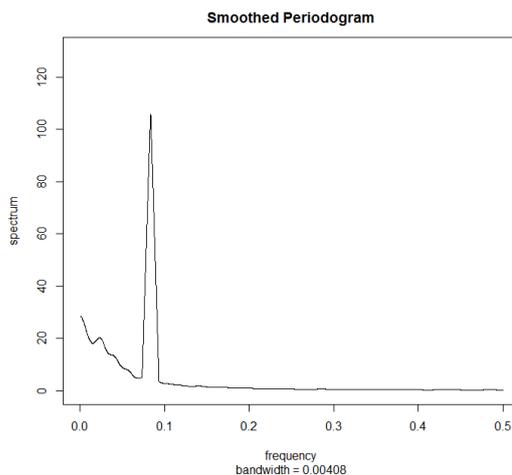


Figura 4.3: Periodograma suavizado

Ahora generamos una serie que tenga una frecuencia de periodo 12 y representamos su periodograma suavizado en la figura 4.3. Observamos que el periodograma detecta esta frecuencia ('pico'). Esto nos anuncia que tenemos que desestacionalizar la serie aplicando el operador de diferencias en esta frecuencia.

Las siguientes secciones las dedicaremos a estudiar inferencia en modelos ARMA sobre datos que ya han sido preprocesados de forma que podemos suponer que hay estacionaridad y ausencia de componentes estacionales. Además, asumiremos que los datos están centrados (si la media de los datos observados no es cero, corregimos centrandolo con la media muestral). Empezaremos con la estimación preliminar de los parámetros.

4.2. Estimación preliminar

En esta sección, vamos a considerar dos técnicas para la estimación preliminar de los parámetros $\{\phi_i, i = 1, \dots, p\}$, $\{\theta_i, i = 1, \dots, q\}$ y σ^2 a partir de la observación de x_1, \dots, x_n , que supondremos parte de una realización del proceso causal ARMA(p, q) definido por $\phi(B)X_t = \theta(B)Z_t$, $\{Z_t\} \sim WN(0, \sigma^2)$.

El objetivo de estas dos técnicas es aplicar un procedimiento simple que nos proporcione buenos valores iniciales para calcular el estimador máximo verosímil en la forma que se describe más adelante.

El primer procedimiento que se describe se llama la estimación de Yule-Walker y se aplica para el ajuste de modelos autorregresivos. El segundo procedimiento está basado en el algoritmo de innovaciones y es válido en general, también cuando existe una parte de media móvil.

4.2.1. Estimación de Yule-Walker

Sea $\{X_t\}$ un proceso autorregresivo de orden p , es decir,

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t \quad (4.6)$$

donde $\{Z_t\} \sim WN(0, \sigma^2)$ y ϕ_1, \dots, ϕ_p son constantes. Si asumimos causalidad podemos escribir $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$, donde $\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{1}{\phi(z)}$. Entonces

$$\begin{aligned} E(X_t | X_1, \dots, X_{t-1}) &= \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + E(Z_t | X_1, \dots, X_{t-1}) \\ &= \phi_1 X_{t-1} + \dots + \phi_p X_{t-p}, \end{aligned}$$

es decir, ϕ_1, \dots, ϕ_p son los coeficientes de $P_{\overline{\text{sp}}\{X_1, \dots, X_{t-1}\}} X_t$ y, de acuerdo con lo estudiado en el Capítulo 3, satisfacen

$$\sigma^2 = \gamma(0) - \phi' \gamma_p$$

y las ecuaciones de Yule-Walker

$$\Gamma_p \phi = \gamma_p,$$

donde

$$\Gamma_p = \begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \dots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \dots & \gamma(p-2) \\ \gamma(2) & \gamma(1) & \gamma(0) & \dots & \gamma(p-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma(p-1) & \gamma(p-2) & \gamma(p-3) & \dots & \gamma(0) \end{bmatrix},$$

$$\phi = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \vdots \\ \phi_p \end{bmatrix}, \quad \gamma_p = \begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \gamma(3) \\ \vdots \\ \gamma(p) \end{bmatrix}.$$

Entonces dadas las observaciones x_1, \dots, x_n podemos calcular su vector y matriz de covarianzas muestrales,

$$\hat{\gamma}_p = \begin{bmatrix} \hat{\gamma}(1) \\ \hat{\gamma}(2) \\ \hat{\gamma}(3) \\ \vdots \\ \hat{\gamma}(p) \end{bmatrix} \quad \text{y} \quad \hat{\Gamma}_p = \begin{bmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \hat{\gamma}(2) & \cdots & \hat{\gamma}(p-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \hat{\gamma}(1) & \cdots & \hat{\gamma}(p-2) \\ \hat{\gamma}(2) & \hat{\gamma}(1) & \hat{\gamma}(0) & \cdots & \hat{\gamma}(p-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}(p-1) & \hat{\gamma}(p-2) & \hat{\gamma}(p-3) & \cdots & \hat{\gamma}(0) \end{bmatrix},$$

y, si $\hat{\Gamma}_p$ es no singular, podemos obtener las estimaciones preliminares de ϕ y σ^2 a partir de las ecuaciones

$$\hat{\Gamma}_p \hat{\phi} = \hat{\gamma}_p$$

y

$$\hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\phi}' \hat{\gamma}_p.$$

Estas se pueden resolver de forma recursiva, como se describió en el Capítulo 3.

Se pueden formular ecuaciones parecidas para modelos ARMA(p, q) con $q > 0$ pero el resultado son ecuaciones no lineales en los coeficientes desconocidos y, por lo tanto, no se utiliza este procedimiento en este caso.

4.2.2. Estimación basada en el algoritmo de innovaciones

Según se ha visto en el Capítulo 3 (ecuación (3.15)), si X_t es un proceso de media móvil de orden q de la forma

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

con $\{Z_t\} \sim WN(0, \sigma^2)$ y $\theta_1, \dots, \theta_q$ constantes, entonces

$$\hat{X}_{n+1} = \sum_{j=1}^n \theta_{n,j} (X_{n+1-j} - \hat{X}_{n+1-j}).$$

Podemos interpretar que $X_{n+1-j} - \hat{X}_{n+1-j}$ es una 'estimación' de Z_{n+1-j} y tomar entonces los coeficientes $\theta_{n,j}$ como estimadores de θ_j . En realidad, los coeficientes $\theta_{n,j}$ no son conocidos, por lo que debemos estimarlos. Para ello se aplica el algoritmo de innovaciones, descrito en el Capítulo 3, a partir de

las covarianzas muestrales. Se obtienen de esta forma los coeficientes.

Este algoritmo también se puede utilizar para estimar cuando $p > 0$. Sea $\{X_t\}$ un proceso ARMA(p, q) causal de media cero, $\phi(B)X_t = \theta(B)Z_t$, $\{Z_t\} \sim WN(0, \sigma^2)$. Por ser causal podemos escribir $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$, satisfaciendo

$$\begin{cases} \psi_0 = 1, \\ \psi_j = \theta_j + \sum_{i=1}^{\min(j,p)} \phi_i \psi_{j-i}, \quad j = 1, 2, \dots \end{cases} \quad (4.7)$$

y por convención, $\theta_j = 0$ para $j > q$ y $\phi_j = 0$ para $j > p$. Sea m un entero grande tal que $m < n$ y $m \cdot n^{-1/3} \rightarrow 0$. Para estimar $\psi_1, \dots, \psi_{p+q}$ podemos utilizar las estimaciones de innovación $\hat{\theta}_{m1}, \dots, \hat{\theta}_{m,p+q}$ a partir de las covarianzas muestrales. Reemplazando ψ_j por $\hat{\theta}_{mj}$ en (4.7) y resolviendo las ecuaciones resultantes,

$$\hat{\theta}_{mj} = \theta_j + \sum_{i=1}^{\min(j,p)} \phi_i \hat{\theta}_{m,j-i}, \quad j = 1, 2, \dots, p+q, \quad (4.8)$$

obtenemos estimaciones iniciales de los coeficientes $\phi = (\phi_1, \dots, \phi_p)'$ y $\theta = (\theta_1, \dots, \theta_q)'$. Para resolver estas ecuaciones primero debemos hallar la estimación de ϕ a partir de las últimas p ecuaciones,

$$\begin{bmatrix} \hat{\theta}_{m,q+1} \\ \hat{\theta}_{m,q+2} \\ \vdots \\ \hat{\theta}_{m,q+p} \end{bmatrix} = \begin{bmatrix} \hat{\theta}_{m,q} & \hat{\theta}_{m,q-1} & \cdots & \hat{\theta}_{m,q+1-p} \\ \hat{\theta}_{m,q+1} & \hat{\theta}_{m,q} & \cdots & \hat{\theta}_{m,q+2-p} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\theta}_{m,q+p-1} & \hat{\theta}_{m,q+p-2} & \cdots & \hat{\theta}_{m,q} \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix},$$

donde hemos utilizado que $\theta_j = 0$ para $j > q$. Una vez obtenida la solución del sistema, $\hat{\phi}$, (quizá no causal), la estimación de θ se obtiene de las primeras q ecuaciones de (4.8) sustituyendo ϕ por su estimación $\hat{\phi}$, es decir,

$$\hat{\theta}_j = \hat{\theta}_{mj} - \sum_{i=1}^{\min(j,p)} \hat{\phi}_i \hat{\theta}_{m,j-i}, \quad j = 1, 2, \dots, q.$$

La estimación de la varianza σ^2 del ruido blanco es $\hat{\sigma}^2 = \hat{v}_m$, obtenido del algoritmo de innovaciones.

Estos dos algoritmos tienen una diferencia importante. Mientras que para un proceso autorregresivo AR(p) los estimadores de Yule-Walker $\hat{\phi}$ son estimadores consistentes de ϕ cuando n tiende a infinito, para un proceso de media móvil MA(q) el estimador $\hat{\theta}$ no es consistente para θ . La condición para que sí lo sea es que m sea un entero grande tal que $m < n$ y $m \cdot n^{-1/3} \rightarrow 0$.

4.3. Estimación de máxima verosimilitud

La estimación de máxima verosimilitud es el método en el que para estimar los parámetros se maximiza la función de verosimilitud. Aquí describimos como utilizar este método para estimar los parámetros en una serie temporal Gaussiana $\{X_t\}$ con media cero y función de autocovarianza $\kappa(i, j) = E(X_i X_j)$. Asumimos que observamos los primeros n instantes de la serie $\mathbf{X}_n = (X_1, \dots, X_n)'$ y denotamos por $\hat{\mathbf{X}}_n = (\hat{X}_1, \dots, \hat{X}_n)'$ los predictores de un paso de los que hemos hablado en el Capítulo 3,

$$\hat{X}_n = \begin{cases} 0, & \text{si } n = 1, \\ E(X_n | X_1, \dots, X_{n-1}) = P_{n-1} X_n, & \text{si } n = 2, 3, \dots \end{cases}$$

Sea $\Gamma_n = E(\mathbf{X}_n \mathbf{X}_n')$ la matriz de covarianza que suponemos no singular. Entonces la verosimilitud se puede escribir en términos del vector de medias y de la matriz de covarianzas y, como la serie es centrada, tenemos que la verosimilitud de \mathbf{X}_n es

$$L(\Gamma_n) = (2\pi)^{-n/2} (\det \Gamma_n)^{-1/2} \exp\left(-\frac{1}{2} \mathbf{X}_n' \Gamma_n^{-1} \mathbf{X}_n\right). \quad (4.9)$$

Para poder evaluar esta función es necesario calcular Γ_n y su inversa Γ_n^{-1} pero podemos evitar el cálculo directo reescribiendo la verosimilitud en términos de los errores de predicción de un paso $X_j - \hat{X}_j$ y sus varianzas v_{j-1} , $j = 1, \dots, n$, los cuales se calculan recursivamente a partir del algoritmo de innovaciones descrito en el Capítulo 3 (ver subsección 3.0.2). Utilizando la notación de la subsección 3.0.2 tenemos que

$$\mathbf{X}_n = C_n (\mathbf{X}_n - \hat{\mathbf{X}}_n). \quad (4.10)$$

Recordamos del Capítulo 3 que las componentes de $\mathbf{X}_n - \hat{\mathbf{X}}_n$ son incorreladas. Entonces por la definición de v_j tenemos que la matriz de covarianzas de $\mathbf{X}_n - \hat{\mathbf{X}}_n$ es

$$D_n = \text{diag}\{v_0, \dots, v_{n-1}\}.$$

Entonces por (4.10) tenemos que $\Gamma_n = C_n D_n C_n'$ y, por lo tanto,

$$\mathbf{X}_n' \Gamma_n^{-1} \mathbf{X}_n = (\mathbf{X}_n - \hat{\mathbf{X}}_n)' D_n^{-1} (\mathbf{X}_n - \hat{\mathbf{X}}_n) = \sum_{j=1}^n \frac{(X_j - \hat{X}_j)^2}{v_{j-1}}$$

y

$$\det \Gamma_n = (\det C_n)^2 (\det D_n) = v_0 v_1 \cdots v_{n-1},$$

(recordamos que $\det C_n = 1$). Entonces la verosimilitud de \mathbf{X}_n (4.9) se reduce a

$$L(\Gamma_n) = \frac{1}{\sqrt{(2\pi)^n v_0 \cdots v_{n-1}}} \exp \left(-\frac{1}{2} \sum_{j=1}^n \frac{(X_j - \hat{X}_j)^2}{v_{j-1}} \right). \quad (4.11)$$

La expresión (4.11) reduce el problema de evaluar la verosimilitud a la evaluación de los predictores, \hat{X}_j , y las varianzas de los errores de predicción, v_j , a partir de los datos observados y los parámetros. En el caso de los modelos ARMA(p, q) esto puede hacerse mediante la forma especializada del algoritmo de innovaciones descrito en la Sección 3.1, es decir,

$$\hat{X}_{n+1} = \begin{cases} \sum_{j=1}^n \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}), & 1 \leq n < m, \\ \phi_1 X_n + \cdots + \phi_p X_{n+1-p} + \sum_{j=1}^q \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}), & n \geq m, \end{cases}$$

y $v_n = E(X_{n+1} - \hat{X}_{n+1})^2 = \sigma^2 E(W_{n+1} - \hat{W}_{n+1})^2 = \sigma^2 r_n$, donde θ_{nj} y r_n se hallan a partir del algoritmo de innovaciones con κ en (3.12) y $m = \max(p, q)$. Entonces sustituyendo en (4.11) la verosimilitud de \mathbf{X}_n quedaría

$$L(\phi, \theta, \sigma^2) = \frac{1}{\sqrt{(2\pi\sigma^2)^n r_0 \cdots r_{n-1}}} \exp \left(-\frac{1}{2\sigma^2} \sum_{j=1}^n \frac{(X_j - \hat{X}_j)^2}{r_{j-1}} \right),$$

donde $\phi = (\phi_1, \dots, \phi_p)$ y $\theta = (\theta_1, \dots, \theta_q)$ son los vectores de coeficientes autorregresivos y de media móvil. A partir de esta función obtenemos la log-verosimilitud, es decir,

$$\begin{aligned} l(\phi, \theta, \sigma^2) &= \ln L(\phi, \theta, \sigma^2) \\ &= -\frac{1}{2\sigma^2} \sum_{j=1}^n \frac{(X_j - \hat{X}_j)^2}{r_{j-1}} - \frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2} \sum_{j=1}^n \ln r_{j-1}. \end{aligned}$$

Para ϕ y θ fijados y notando que \hat{X}_j y r_j son independientes de σ^2 , maximizamos en σ^2 la función

$$-\frac{1}{2\sigma^2} \sum_{j=1}^n \frac{(X_j - \hat{X}_j)^2}{r_{j-1}} - \frac{n}{2} \ln \sigma^2$$

obteniendo que el estimador de máxima verosimilitud de σ^2 es

$$\hat{\sigma}^2 = n^{-1} S(\hat{\phi}, \hat{\theta}),$$

donde

$$S(\phi, \theta) = \sum_{j=1}^n \frac{(X_j - \hat{X}_j)^2}{r_{j-1}}. \quad (4.12)$$

Los estimadores de máxima verosimilitud $\hat{\phi}$ y $\hat{\theta}$ son los valores de ϕ , θ que minimizan

$$l(\phi, \theta) = \ln(n^{-1}S(\phi, \theta)) + n^{-1} \sum_{j=1}^n \ln r_{j-1}. \quad (4.13)$$

Esta función es no lineal por lo que su optimización requiere métodos de optimización numérica que están fuera del objetivo de esta memoria. Por esta razón es importante contar con buenos estimadores preliminares.

Una forma alternativa de obtener los estimadores es la estimación de mínimos cuadrados. Los estimadores de mínimos cuadrados $\tilde{\phi}$ y $\tilde{\theta}$ de ϕ y θ se obtienen minimizando la función (4.12) en lugar de (4.13), y el estimador de σ^2 es $\tilde{\sigma}^2 = \frac{S(\tilde{\phi}, \tilde{\theta})}{n-p-q}$.

Los estimadores de máxima verosimilitud, bajo ciertas condiciones, tienen algunas propiedades estadísticas. Si el proceso es causal e invertible entonces el estimador $\hat{\beta} = (\hat{\phi}, \hat{\theta})$ formado por las estimaciones de máxima verosimilitud de $\beta = (\phi, \theta)$ es consistente y asintóticamente normal, es decir,

$$\sqrt{n}(\hat{\beta} - \beta) \rightarrow N(0, V(\beta)),$$

donde la matriz de covarianza $V(\beta)$ se puede estimar de forma consistente.

4.4. Diagnóstico y validación

La bondad de un modelo estadístico ajustado a un conjunto de datos se puede valorar comparando valores observados con valores predichos a partir del modelo ajustado. Si el modelo ajustado es apropiado, entonces los *residuales*, definidos como

$$\hat{W}_t = \frac{X_t - \hat{X}_t(\hat{\phi}, \hat{\theta})}{\sqrt{r_{t-1}(\hat{\phi}, \hat{\theta})}}, \quad t = 1, \dots, n,$$

deben comportarse de una manera consistente con el modelo, es decir, deben tener propiedades similares a las de la secuencia de ruido blanco

$$W_t(\phi, \theta) = \frac{X_t - \hat{X}_t(\phi, \theta)}{\sqrt{r_{t-1}(\phi, \theta)}}, \quad t = 1, \dots, n.$$

Entonces, si el modelo es correcto esperamos que los *residuales reescalados*,

$$\hat{R}_t = \frac{\hat{W}_t}{\hat{\sigma}}, \quad t = 1, \dots, n,$$

donde $\hat{\sigma} = \sqrt{\left(\sum_{t=1}^n \hat{W}_t^2\right)/n}$ es la estimación de la desviación estándar del ruido blanco, tengan propiedades similares a las de una secuencia $WN(0, 1)$ o de una secuencia *i.i.d.*(0, 1) (independientes e igualmente distribuidas de media cero y varianza uno) si asumimos que el ruido blanco $\{Z_t\}$ del proceso ARMA es ruido blanco independiente. Las siguientes comprobaciones de diagnóstico se basan todas en las propiedades esperadas de los residuales o residuales reescalados bajo la suposición de que el modelo ajustado es correcto y que $\{Z_t\} \sim i.i.d.(0, \sigma^2)$.

La gráfica de $\{\hat{R}_t, t = 1, \dots, n\}$

Si el modelo ajustado es apropiado, entonces el gráfico de los residuales reescalados $\{\hat{R}_t, t = 1, \dots, n\}$ debe parecerse al de una secuencia de ruido blanco con varianza uno. En el gráfico, la media es cero si observamos que el error se mueve en torno al valor cero y la varianza es constante si no observamos fluctuaciones cuya magnitud dependa fuertemente de t .

El siguiente paso es verificar que la función de autocorrelación muestral de $\{\hat{W}_t\}$ (o de manera equivalente de $\{\hat{R}_t\}$) se comporte como debería bajo el supuesto de que el modelo ajustado es apropiado.

La función de autocorrelación muestral de los residuos

Para n grande, las autocorrelaciones muestrales de una secuencia Y_1, \dots, Y_n independiente e igualmente distribuida con varianza finita son aproximadamente independientes e igualmente distribuidas con distribución $N(0, 1/n)$. Entonces rechazaremos la hipótesis de ruido *i.i.d.* si el 5% de las autocorrelaciones muestrales de los residuales quedan fuera de los límites $\pm 1,96/\sqrt{n}$ o si uno cae muy lejos de los límites.

Prueba de aleatoriedad de los residuos

Esta prueba consiste en comprobar si se verifica la hipótesis de que los residuos son valores observados de variables aleatorias independientes e idénticamente distribuidas. Existen muchas pruebas de aleatoriedad de los residuos pero en esta memoria solo vamos a exponer el test portmanteau y su variante de Ljung-Box, una prueba basada en ajustar un modelo autorregresivo y la prueba de normalidad. Otras son la prueba de punto de inflexión, la del signo de diferencia y la de rango. La estrategia general al aplicar estas pruebas es verificarlas todas pero, como a medida que aumenta el número de pruebas la

probabilidad de que al menos una rechace la hipótesis nula cuando es verdadera aumenta, no se debe rechazar necesariamente la hipótesis nula si alguna prueba no la verifica.

El test portmanteau

Consideramos el estadístico

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

donde $\hat{\rho}(j)$ son las autocorrelaciones muestrales. Si Y_1, \dots, Y_n es una secuencia de variables independientes e igualmente distribuidas con varianza finita, Q se distribuye aproximadamente como la suma de cuadrados de las variables aleatorias independientes $N(0, 1)$, $\sqrt{n}\hat{\rho}(j)$, $j = 1, \dots, h$, es decir, como chi-cuadrado con h grados de libertad. Entonces rechazamos la hipótesis en el nivel α si $Q > \chi_{1-\alpha}^2(h)$, donde $\chi_{1-\alpha}^2(h)$ es el cuantil $1 - \alpha$ de la distribución chi-cuadrado con h grados de libertad. Existe un refinamiento de esta prueba, formulado por Ljung y Box a finales de la década de los 70, en la que Q se reemplaza por

$$Q_{LB} = n(n+2) \sum_{j=1}^h \frac{\hat{\rho}^2(j)}{n-j}$$

cuya distribución se aproxima mejor por la distribución chi-cuadrado con h grados de libertad.

Ajustando un modelo autorregresivo

Una prueba adicional es adaptar un modelo autorregresivo a los datos usando el algoritmo Yule-Walker y elegir el orden que minimiza la estadística AICC que se describe en la Sección 4.5. Un orden seleccionado igual a cero sugiere que los residuos son ruido blanco.

Prueba de normalidad

Si el proceso de ruido es Gaussiano, es decir, si todas sus distribuciones conjuntas son normales, se pueden extraer conclusiones más sólidas cuando se ajusta un modelo a los datos. La siguiente prueba nos permite verificar si es razonable suponer que las observaciones de una secuencia de variables independientes e igualmente distribuidas son también Gaussianas. Sean

$Y_{(1)} < Y_{(2)} < \dots < Y_{(n)}$ las estadísticas de orden de una muestra aleatoria Y_1, \dots, Y_n de la distribución $N(\mu, \sigma^2)$. Si $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ son las estadísticas de orden de una muestra $N(0, 1)$ de tamaño n , entonces $EY_{(j)} = \mu + \sigma m_j$, donde $m_j = EX_{(j)}$, $j = 1, \dots, n$. El gráfico de los puntos $(m_1, Y_{(1)}), \dots, (m_n, Y_{(n)})$ se llama gráfica qq-Gaussiana. Si la suposición normal es correcta, la gráfica qq-Gaussiana debería ser aproximadamente lineal. En consecuencia, la correlación al cuadrado de los puntos $(m_i, Y_{(i)})$, $i = 1, \dots, n$, debe estar cerca de uno. Por lo tanto, se rechaza la asunción de la normalidad si la correlación al cuadrado R^2 es suficientemente pequeña. Si aproximamos m_i por $\Phi^{-1}((i-.5)/n)$, entonces R^2 se reduce a

$$R^2 = \frac{(\sum_{i=1}^n (Y_{(i)} - \bar{Y}) \Phi^{-1}(\frac{i-.5}{n}))^2}{\sum_{i=1}^n (Y_{(i)} - \bar{Y})^2 \sum_{i=1}^n (\Phi^{-1}(\frac{i-.5}{n}))^2}$$

donde $\bar{Y} = n^{-1}(Y_1 + \dots + Y_n)$. Los puntos porcentuales para la distribución de R^2 , asumiendo la normalidad de los valores muestrales, son dados por Shapiro y Francia en los años 70 para tamaños de muestra $n < 100$. Para valores mayores de n se puede utilizar la prueba Jarque-Bera para la normalidad $n[m_3^2/(6m_2^3) + (m_4/m_2^2 - 3)^2/24]$, donde $m_r = \sum_{j=1}^n (Y_j - \bar{Y})^r/n$, se distribuye asintóticamente como $\chi^2(2)$ si $\{Y_t\} \sim i.i.d.N(\mu, \sigma^2)$. Esta hipótesis se rechaza si la estadística es suficientemente grande (en el nivel α si el valor p de la prueba es menor que α).

4.5. Selección del orden

Una vez que podemos ajustar una serie $\{X_t\}$ por un modelo ARMA de media cero, tenemos que seleccionar los valores apropiados para los órdenes p y q .

Es un hecho general en estadística que el ajuste de un modelo a los datos mejora, aparentemente, cuando aumenta la complejidad del modelo. Esta mejora aparente puede conducir, sin embargo, a un deterioro de la capacidad predictiva del modelo estadístico. Este es un principio general, aplicable también al modelado de series temporales. Por simplicidad, vamos a tratar de explicar esta idea en el contexto más sencillo de regresión. Supongamos que $(Y_1, X_1), \dots, (Y_n, X_n)$ son unas observaciones iid de un par (Y, X) con $Y \in \mathbb{R}$ y $X \in \mathbb{R}^k$, y queremos predecir Y a partir de X . La mejor predicción posible en el sentido de que se minimice el error cuadrático medio $E(Y - f(X))^2$ es

$$f_B(X) = E(Y|X).$$

Esta función no es fácil de estimar así que podemos buscar f en una clase \mathcal{F} , es decir,

$$\bar{f} = \underset{f \in \mathcal{F}}{\operatorname{argmin}} E(Y - f(X))^2.$$

Esta función \bar{f} no puede ser mejor que f_B luego

$$\varepsilon(\mathcal{F}) = E(Y - \bar{f}(X))^2 - E(Y - f_B(X))^2 \geq 0$$

donde $\varepsilon(\mathcal{F})$ se denomina el *exceso de riesgo asociado a la clase \mathcal{F}* . Ahora estimamos \bar{f} a partir de $(Y_1, X_1), \dots, (Y_n, X_n)$ mediante

$$\bar{f}_n = \underset{f \in \mathcal{F}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n (Y_i - f(X_i))^2.$$

Entonces se puede comprobar que

$$\begin{aligned} & E(Y - \bar{f}_n(X))^2 - E(Y - f_B(X))^2 \\ & \leq \sqrt{n} E \left(\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n (Y_i - f(X_i))^2 - E(Y - f(X))^2 \right| \right) + \varepsilon(\mathcal{F}), \end{aligned}$$

(esta es la ecuación (2.2), página 204 de [1]). El término

$$\sqrt{n} E \left(\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n (Y_i - f(X_i))^2 - E(Y - f(X))^2 \right| \right)$$

se puede considerar como un término de varianza mientras que a $\varepsilon(\mathcal{F})$ nos referimos como término de sesgo. Si escogemos una clase \mathcal{F} pequeña tenemos menos opciones de encontrar una función que se aproxime a f_B y, por lo tanto, más probabilidad de que el término de sesgo sea grande. Pero tampoco es conveniente que la clase \mathcal{F} sea grande porque aumentaría el término de varianza.

En general, en cualquier contexto estadístico, hay que limitar la complejidad de los modelos que consideramos para evitar un sobreajuste. Para ello, necesitamos introducir un "factor de penalización". Los siguientes criterios están basados en esos factores de penalización. El Criterio de FPE, del inglés 'Final Prediction Error', desarrollado por Akaike en los años 70 (ver [4], página 301) para seleccionar el orden apropiado de un proceso AR para ajustar una serie temporal $\{X_1, \dots, X_n\}$, se basa en la idea de escoger el modelo para $\{X_t\}$ de forma que se minimice el error cuadrático medio de un solo paso cuando el modelo ajustado a $\{X_t\}$ se usa para predecir una realización

independiente $\{Y_t\}$ del mismo proceso que generó $\{X_t\}$. El orden se obtiene eligiendo el valor de p que minimice

$$FPE_p = \hat{\sigma}^2 \frac{n+p}{n-p}.$$

Observamos que si tomamos órdenes p altos, el criterio de FPE aumenta.

Otro criterio más general también desarrollado por Akaike es el Criterio de Información de Akaike (AIC) (ver [4], página 302). Consiste en elegir el valor de p y q que minimice

$$AIC(\beta) = -2 \ln L_X \left(\beta, \frac{S_X(\beta)}{n} \right) + 2(p+q+1),$$

donde S_X es la función (4.12) y $\beta = (\phi, \theta)$ si el proceso ARMA(p, q) está descrito por las ecuaciones $\phi(B)X_t = \theta(B)Z_t$. El factor de penalización de este criterio es $2(p+q+1)$ que no tiene en cuenta el número de observaciones del modelo. Por ello, Hurvich y Tsai (ver [4], página 302) crearon una versión corregida por sesgo de este criterio conocida como AICC. Para aplicar el criterio de AICC se elige el valor de p y q que minimice

$$AICC(\beta) = -2 \ln L_X \left(\beta, \frac{S_X(\beta)}{n} \right) + \frac{2(p+q+1)n}{n-p-q-2}.$$

El factor de penalización en este caso es $\frac{2(p+q+1)n}{n-p-q-2}$, que es asintóticamente equivalente al de AIC cuando $n \rightarrow \infty$ pero para los modelos de orden grande tiene una penalización más extrema, lo que contrarresta el sobreajuste del AIC. Tanto $AIC(\beta, \sigma^2)$ como $AICC(\beta, \sigma^2)$ se pueden definir para σ^2 arbitrario reemplazando $S_X(\beta)/n$ en las definiciones anteriores por σ^2 . Los estudios de Monte Carlo (ver [4], Capítulo 9, página 304) sugieren que el AIC tiende a sobreestimar p en el ajuste de los modelos AR.

Existen otros métodos de penalización para seleccionar el orden de un modelo ARMA(p, q) como, por ejemplo, el Criterio de Información Bayesiano (BIC) definido por Akaike (ver [4], página 304) como

$$\begin{aligned} BIC &= (n-p-q) \ln \left(\frac{n\hat{\sigma}^2}{n-p-q} \right) + n \left(1 + \ln \sqrt{2\pi} \right) \\ &\quad + (p+q) \ln \left(\frac{\sum_{t=1}^n X_t^2 - n\hat{\sigma}^2}{p+q} \right), \end{aligned}$$

para un proceso ARMA(p, q) causal e invertible de media cero, donde $\hat{\sigma}^2$ es la estimación de máxima verosimilitud de la varianza del ruido blanco. La

selección de orden para un proceso AR minimizando el BIC no es asintóticamente eficiente (páginas 304-305 de [4]) y, por ello, no incluimos nada más sobre este procedimiento.

Por otra parte, existen métodos menos generales para la selección del orden de un modelo adecuado para los datos.

En el caso de los modelos autorregresivos, si tenemos $\{X_t\}$ un proceso causal AR(p) y ajustamos un modelo con orden $m > p$ utilizando las ecuaciones de Yule-Walker, entonces el último coeficiente del vector $\hat{\phi}_m$ tiene aproximadamente una distribución $N(0, 1/n)$. Como las autocorrelaciones parciales ϕ_{mm} , $m > p$, son cero, si un modelo AR(p) es apropiado para los datos, entonces los valores $\hat{\phi}_{kk}$, $k > p$, deberían ser compatibles con las observaciones de la distribución $N(0, 1/n)$. En particular, para $k > p$ $\hat{\phi}_{kk}$ caerá en los límites $\pm 1,96n^{-1/2}$ con una probabilidad cercana a 0,95. Esto sugiere utilizar como un estimador preliminar de p el valor m más pequeño de manera que $|\hat{\phi}_{kk}| < 1,96n^{-1/2}$ para $k > m$.

En el caso de los modelos de media móvil, una estimación preliminar del orden q se obtiene tomando el valor más pequeño de m tal que $\hat{\rho}(k)$ no es significativamente diferente de cero para todo $k > m$ (en la práctica se interpreta como mayor que $1,96/\sqrt{n}$ en valor absoluto), ya que las autocorrelaciones $\rho(m)$ de un proceso MA(q) son cero si $m > q$. Si además examinamos los vectores de coeficientes $\hat{\theta}_m$, $m = 1, 2, \dots$, podemos obtener estimaciones preliminares $\hat{\theta}_{m1}, \dots, \hat{\theta}_{mq}$ de los coeficientes del algoritmo de innovaciones. Al inspeccionar los coeficientes estimados $\hat{\theta}_{m1}, \dots, \hat{\theta}_{mm}$ para $m = 1, 2, \dots$ y la relación de cada estimación del coeficiente $\hat{\theta}_{mj}$ a 1,96 veces su desviación estándar aproximada $\sigma_j = n^{-1/2} \left(\sum_{i=0}^{j-1} \hat{\theta}_{mi}^2 \right)^{1/2}$, podemos ver cuáles de las estimaciones de los coeficientes son significativamente más diferentes de cero, estimar el orden del modelo que se ajustará como el mayor retardo j para el cual la relación es mayor que 1 en valor absoluto, y al mismo tiempo leer los valores estimados para cada uno de los coeficientes.

Conclusiones

La infinidad de series temporales aparecidas en diversos campos científicos ha hecho necesario el desarrollo de modelos matemáticos apropiados que nos permitan reproducir el comportamiento de la serie para predecir valores futuros. En esta tarea, los procesos estacionarios tienen un papel fundamental. La permanencia en el tiempo de su estructura probabilística posibilita que se empleen como modelos en los que una observación prolongada puede conducir a un aprendizaje consistente, es decir, a una estimación consistente de los parámetros del modelo.

En este trabajo se han estudiado algunos de los resultados teóricos más importantes de estos procesos. El marco de trabajo es el del espacio de variables con cuadrado integrable. La estructura de Hilbert de estos espacios ha permitido manejar fácilmente conceptos como el de proyección ortogonal sobre subespacios. Las proyecciones sobre ciertos subespacios son en este contexto las principales herramientas de predicción.

Entre los resultados principales estudiados, destacamos las caracterizaciones de la función de autocovarianza, en especial el Teorema de Herglotz, que conecta con la teoría espectral. La función de autocovarianza determina completamente la dependencia temporal de un proceso centrado, Gaussiano y estacionario, lo que permite restringir el modelado a el de la función de autocovarianza.

Por otra parte, se ha estudiado la importancia de los modelos ARMA que permiten aproximar cualquier proceso regular con un error pequeño. Por ello, se ha especificado un procedimiento general para seleccionar el modelo ARMA más adecuado para realizar pronósticos de valores futuros de una serie temporal.

El modelado de series temporales no se limita al uso de modelos ARMA, sino que en el análisis de datos temporales se emplean muchos otros modelos no considerados en esta memoria (modelos ARCH, GARCH, dependencia larga, etc). Con todo, el conocimiento adquirido sobre la teoría de procesos estacionarios y de modelos ARMA es una herramienta útil para el manejo futuro de nuevas técnicas de análisis de series temporales.

Bibliografía

- [1] Del Barrio, E.; Deheuvels, P. y Van de Geer, S. (2007). *Lectures on Empirical Processes: Theory and Statistical Applications*. European Mathematical Society.
- [2] Billingsley, P. (1986). *Probability and Measure. 2nd Ed.* Wiley.
- [3] Box, G. E. P.; Jenkins, G. M. y Reinsel, G. C. (2008). *Time Series Analysis: Forecasting and Control, 4th Ed..* John Wiley & Sons.
- [4] Brockwell, P. J. y Davis, R. A. (1991). *Time Series: Theory and Methods, 2nd Ed..* Springer.
- [5] Brockwell, P. J. y Davis, R. A. (2002). *Introduction to Time Series and Forecasting, 2nd Ed..* Springer.
- [6] Mentz, R. P. (1988). Estimación en los modelos autorregresivos y de promedios móviles. *Revista Estadística Española*, núm. 116, págs. 87-106.
- [7] Peña, D. (2010). *Análisis de series temporales*. Alianza Editorial.