

Capítulo 6

Predicción

6.1. Predictores lineales

Consideramos ahora el problema de predecir los valores X_{n+h} , $h > 0$ de una serie de tiempo estacionaria con media μ y función de autocovarianza γ conocidas, a partir de los valores X_1, \dots, X_n hasta el instante n . Nos restringimos a buscar un predictor lineal: queremos hallar la combinación lineal de $1, X_n, \dots, X_1$ que predice X_{n+h} minimizando el error cuadrático medio. A este estimador lo denotamos por $P_n X_{n+h}$ y tiene la forma

$$P_n X_{n+h} = a_0 + a_1 X_n + \dots + a_n X_1. \quad (6.1)$$

Para determinar los valores de los coeficientes queremos minimizar

$$S(a_0, \dots, a_n) = E(X_{n+h} - a_0 - a_1 X_n - \dots - a_n X_1)^2. \quad (6.2)$$

Como S es una función cuadrática y está acotada inferiormente por 0, está claro que hay al menos un valor de (a_0, a_1, \dots, a_n) que minimiza S y el mínimo (a_0, a_1, \dots, a_n) satisface las ecuaciones

$$\frac{\partial S(a_0, a_1, \dots, a_n)}{\partial a_j} = 0, \quad j = 0, \dots, n \quad (6.3)$$

Al evaluar las derivadas de (6.3) obtenemos las ecuaciones

$$E\left(X_{n+h} - a_0 - \sum_{i=1}^n a_i X_{n+1-i}\right) = 0 \quad (6.4)$$

$$E\left((X_{n+h} - a_0 - \sum_{i=1}^n a_i X_{n+1-i})X_{n+1-j}\right) = 0 \quad j = 1, \dots, n. \quad (6.5)$$

Estas ecuaciones se conocen como las *ecuaciones de predicción* y sirven para obtener los valores de a_0, a_1, \dots, a_n . Se pueden escribir con notación vectorial de la siguiente manera

$$a_0 = \mu\left(1 - \sum_{i=1}^n a_i\right) \quad (6.6)$$

y

$$\Gamma_n \mathbf{a}_n = \gamma_n(h), \quad (6.7)$$

donde

$$\mathbf{a}_n = (a_1, \dots, a_n)', \quad \Gamma_n = [\gamma(i-j)]_{i,j=1}^n,$$

y

$$\gamma_n(h) = (\gamma(h), \gamma(h+1), \dots, \gamma(h+n-1))'. \quad (6.8)$$

En consecuencia

$$P_n X_{n+h} = \mu + \sum_{i=1}^n a_i (X_{n+1-i} - \mu), \quad (6.9)$$

donde \mathbf{a}_n satisface (6.7). A partir de (6.9) el valor esperado del error de predicción $X_{t+h} - P_n X_{n+h}$ es cero, y el error de predicción en media cuadrática es

$$\begin{aligned} E(X_{n+h} - P_n X_{n+h})^2 &= \gamma(0) - 2 \sum_{i=1}^n a_i \gamma(h+i-1) + \sum_{i=1}^n \sum_{j=1}^n a_i \gamma(j-i) a_j \\ &= \gamma(0) - \mathbf{a}_n' \gamma_n(h), \end{aligned} \quad (6.10)$$

donde la última igualdad sigue de (6.7).

Observación 6.1 Para ver que las ecuaciones (6.4) y (6.5) determinan $P_n X_{n+h}$ de manera única, sean $\{a_j^1, j = 0, \dots, n\}$ y $\{a_j^2, j = 0, \dots, n\}$ dos soluciones y sea Z la diferencia entre los predictores correspondientes:

$$Z = a_0^1 - a_0^2 + \sum_{j=1}^n (a_j^1 - a_j^2) X_{n+1-j}.$$

Entonces

$$Z^2 = Z(a_0^1 - a_0^2 + \sum_{j=1}^n (a_j^1 - a_j^2) X_{n+1-j}).$$

Pero a partir de (6.4) y (6.5) vemos que $E(Z) = 0$ y $E(ZX_{n+1-j}) = 0$ para $j = 1, \dots, n$. En consecuencia $E(Z^2) = 0$ y $Z = 0$.

Propiedades de $P_n X_{n+h}$

1. $P_n X_{n+h} = \mu + \sum_{i=1}^n a_i (X_{n+1-i} - \mu)$, donde $\mathbf{a}_n = (a_1, \dots, a_n)'$ satisface (6.7).
2. $E(X_{n+h} - P_n X_{n+h})^2 = \gamma(0) - \mathbf{a}_n' \gamma_n(h)$, donde $\gamma_n(h) = (\gamma(h), \dots, \gamma(h+n-1))'$.
3. $E(X_{n+h} - P_n X_{n+h}) = 0$
4. $E[(X_{n+h} - P_n X_{n+h})X_j] = 0, \quad j = 1, \dots, n$.

Observación 6.2 Observamos que las propiedades 3 y 4 son equivalentes a (6.4) y (6.5) y pueden escribirse de manera resumida como

$$E[(\text{Error}) \times (\text{Variable predictora})] = 0. \quad (6.11)$$

Estas ecuaciones determinan, por lo tanto, $P_n X_{n+h}$ de manera única.

Ejemplo 6.1 (AR(1))

Consideremos un proceso AR(1), definido por la ecuación

$$X_t = \phi X_{t-1} + w_t, \quad t \in \mathbb{Z},$$

con $|\phi| < 1$. Usando (6.7) y (6.8), el mejor predictor lineal de X_{n+1} a partir de $1, X_n, \dots, X_1$ es

$$P_n X_{n+1} = \mathbf{a}'_n \mathbf{X}_n,$$

donde $\mathbf{X}_n = (X_n, \dots, X_1)'$ y

$$\begin{pmatrix} 1 & \phi & \phi^2 & \dots & \phi^{n-1} \\ \phi & 1 & \phi & \dots & \phi^{n-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \phi^{n-3} & \dots & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \phi \\ \phi^2 \\ \vdots \\ \phi^n \end{pmatrix}. \quad (6.12)$$

Una solución de este sistema es

$$\mathbf{a}_n = (\phi, 0, \dots, 0)'$$

y por lo tanto el mejor predictor lineal de X_{n+1} a partir de X_1, \dots, X_n es

$$P_n X_{n+1} = \mathbf{a}'_n \mathbf{X}_n = \phi X_n$$

con error cuadrático medio

$$E(X_{n+1} - P_n X_{n+1})^2 = \gamma(0) - \mathbf{a}'_n \gamma_n(1) = \frac{\sigma^2}{1 - \phi^2} - \phi \gamma(1) = \sigma^2.$$

Otra manera de resolver este problema es adivinar, por inspección de la ecuación que define el proceso, que el mejor predictor de X_{n+1} es ϕX_n . Para verificar esta conjetura basta verificar que (6.11) vale para cada una de las variables predictoras $1, X_n, \dots, X_1$. El error de predicción del predictor ϕX_n es $X_{n+1} - \phi X_n = w_{n+1}$. Pero $E(w_{n+1} Y) = 0$ para $Y = 1$ y para $Y = X_j, j = 1, \dots, n$. En consecuencia ϕX_n es el mejor predictor lineal de X_{n+1} a partir de $1, X_1, \dots, X_n$. \blacktriangle

6.1.1. Espacios de Hilbert y el Teorema de Proyección

Recordemos que un Espacio de Hilbert es un espacio con un producto interno que es completo respecto a la norma definida por el producto interno. Veamos que quiere decir todo esto.

Si \mathcal{H} es un conjunto, un producto interno sobre \mathcal{H} es una función $\langle \cdot, \cdot \rangle : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}$ que satisface las siguientes propiedades

- Para $a, b \in \mathcal{H}$, $\langle a, b \rangle = \langle b, a \rangle$.
- Para $a_1, a_2, b \in \mathcal{H}$, $\alpha_1, \alpha_2, \beta \in \mathbb{R}$, $\langle \alpha_1 a_1 + \alpha_2 a_2, b \rangle = \alpha_1 \langle a_1, b \rangle + \alpha_2 \langle a_2, b \rangle$.
- $\langle a, a \rangle = 0 \Leftrightarrow a = 0$.

A partir del producto interno podemos definir una norma sobre \mathcal{H} ,

$$\|a\| = \langle a, a \rangle$$

es decir, la función $\|\cdot\| : \mathcal{H} \rightarrow \mathbb{R}$ satisface las siguientes propiedades:

- $\|a\| \geq 0$; $\|a\| = 0 \Leftrightarrow a = 0$.
- $\|\alpha a\| = |\alpha| \|a\|$.
- $\|a + b\| \leq \|a\| + \|b\|$, que se conoce como la *desigualdad triangular*.

Decimos que \mathcal{H} es un *Espacio de Hilbert* si es completo respecto a la convergencia definida por la norma, es decir, si toda sucesión de Cauchy en \mathcal{H} converge a un elemento del mismo espacio. Los ejemplos más conocidos son los espacios euclídeos \mathbb{R}^n con producto interno $\langle x, y \rangle = \sum_i x_i y_i$. Otro ejemplo fundamental es el espacio de las variables aleatorias de cuadrado integrable, que se conoce como L^2 , con producto interno

$$\langle X, Y \rangle = E(XY).$$

Teorema 6.1 (Teorema de Proyección) *Sea \mathcal{H} un espacio de Hilbert, \mathcal{M} un subespacio lineal cerrado de \mathcal{H} y $y \in \mathcal{H}$, entonces existe un punto $Py \in \mathcal{M}$, que se conoce como la proyección de y sobre \mathcal{M} , que satisface las siguientes propiedades:*

1. $\|Py - y\| \leq \|z - y\|$ para todo $z \in \mathcal{H}$.
2. $\|Py - y\| < \|z - y\|$ para todo $z \in \mathcal{H}, z \neq Py$.
3. $\langle y - Py, z \rangle = 0$ para todo $z \in \mathcal{M}$.

La propiedad 3 con frecuencia resulta útil para obtener la proyección de un vector. La norma del error se puede escribir

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

por la ortogonalidad.

Si el subespacio \mathcal{M} está generado por los vectores x_1, \dots, x_n , la proyección de $y \in \mathcal{H}$ sobre \mathcal{M} es única y se puede escribir como

$$Py = a_1 x_1 + \dots + a_n x_n$$

donde los coeficientes a_1, \dots, a_n se pueden hallar usando la ortogonalidad

$$\langle y - Py, x_j \rangle = 0, \quad j = 1, \dots, n.$$

Las ecuaciones que hay que resolver para obtener los a_i son

$$\sum_{i=1}^n a_i \langle x_i, x_j \rangle = \langle y, x_j \rangle, \quad j = 1, \dots, n. \quad (6.13)$$

Si los elementos de \mathcal{H} son vectores, éste es el problema de regresión lineal.

Ejemplo 6.2 (Regresión Lineal)

Para el modelo de regresión queremos hallar los coeficientes de regresión β_i en la expresión

$$X_t = \beta_1 Z_{t1} + \beta_2 Z_{t2} + \dots + \beta_q Z_{tq} + w_t \quad (6.14)$$

de modo de minimizar la suma de los cuadrados de los residuales. Consideramos los vectores $\mathbf{y} = (y_1, \dots, y_n)'$ y $\mathbf{z}_i = (z_{1i}, \dots, z_{ni})'$ para $i = 1, \dots, q$ con producto interno

$$\langle \mathbf{z}_i, \mathbf{y} \rangle = \sum_{j=1}^n z_{ji} y_j = \mathbf{z}'_i \mathbf{y}.$$

Para hallar la proyección del vector \mathbf{y} sobre el subespacio generado por $\mathbf{z}_1, \dots, \mathbf{z}_q$ usamos el principio de ortogonalidad:

$$\langle \mathbf{y} - \sum_{i=1}^q \beta_i \mathbf{z}_i, \mathbf{z}_j \rangle = 0, \quad j = 1, \dots, q.$$

Esta condición se puede escribir vectorialmente como

$$\mathbf{y}' \mathbf{z}_j = \sum_{i=1}^q \beta_i \mathbf{z}'_i \mathbf{z}_j \quad j = 1, \dots, q,$$

Si consideramos la matriz $Z = (\mathbf{z}_1, \dots, \mathbf{z}_q)$, que suponemos que tiene rango completo, podemos escribir la ecuación anterior como

$$\mathbf{y}' Z = \boldsymbol{\beta}' (Z' Z),$$

con $\boldsymbol{\beta} = (\beta_1, \dots, \beta_q)'$. Trasponiendo los términos de esta ecuación y multiplicando por la inversa obtenemos la solución para los coeficientes

$$\hat{\boldsymbol{\beta}} = (Z' Z)^{-1} Z' \mathbf{y}$$

que se conocen como las ecuaciones normales.

6.1.2. Predicción de variables en L^2

Consideremos ahora variables Y y W_1, \dots, W_n variables cualesquiera con segundo momento finito, de modo que las medias $\mu = E(Y)$, $\mu_i = E(W_i)$ y las covarianzas $\text{Cov}(Y, Y)$, $\text{Cov}(Y; W_i)$ y $\text{Cov}(W_i, W_j)$ son todas conocidas. Usaremos la siguiente notación: $\mathbf{W} = (W_n, \dots, W_1)'$, $\boldsymbol{\mu}_W = (\mu_n, \dots, \mu_1)$,

$$\boldsymbol{\gamma} = \text{Cov}(Y, \mathbf{W}) = (\text{Cov}(Y, W_n), \text{Cov}(Y, W_{n-1}), \dots, \text{Cov}(Y, W_1))'$$

y

$$\Gamma = \text{Cov}(\mathbf{W}, \mathbf{W}) = [\text{Cov}(W_{n+1-i}, W_{n+1-j})]_{i,j=1}^n$$

Usando la misma argumentación que usamos para el cálculo de $P_n X_{n+h}$, el mejor predictor lineal de Y es términos de $\{1, W_n, \dots, W_1\}$ es

$$P(Y|\mathbf{W}) = \mu_Y + \mathbf{a}'(\mathbf{W} - \boldsymbol{\mu}_W), \quad (6.15)$$

donde $\mathbf{a} = (a_1, \dots, a_n)'$ es cualquier solución de

$$\Gamma \mathbf{a} = \boldsymbol{\gamma} \quad (6.16)$$

El error medio cuadrático del predictor es

$$E((Y - P(Y|W))^2) = \text{Var}(Y) - \mathbf{a}' \boldsymbol{\gamma}. \quad (6.17)$$

Ejemplo 6.3 (Estimación de un valor faltante)

Consideremos de nuevo un modelo AR(1)

$$X_t = \phi X_{t-1} + w_t, \quad t = 0, \pm 1, \dots$$

con $|\phi| < 1$. Supongamos que observamos la serie en los instantes 1 y 3 y queremos usar estas observaciones para hallar la combinación lineal de $1, X_1$ y X_3 que estima X_2 con mínimo error cuadrático medio. La solución a este problema la podemos obtener a partir de (6.15) y (6.16) poniendo $Y = X_2$ y $\mathbf{W} = (X_1, X_3)'$. Esto nos da las ecuaciones

$$\begin{pmatrix} 1 & \phi^2 \\ \phi^2 & 1 \end{pmatrix} \mathbf{a} = \begin{pmatrix} \phi \\ \phi \end{pmatrix},$$

con solución

$$\mathbf{a} = \frac{1}{1 + \phi^2} \begin{pmatrix} \phi \\ \phi \end{pmatrix}.$$

Por lo tanto el mejor estimador de X_2 es

$$P(X_2|\mathbf{W}) = \frac{\phi}{1 + \phi^2} (X_1 + X_3)$$

con error cuadrático medio

$$E((X_2 - P(X_2|\mathbf{W}))^2) = \frac{\sigma_w^2}{1 - \phi^2} - \mathbf{a}' \begin{pmatrix} \frac{\phi \sigma_w^2}{1 - \phi^2} \\ \frac{\phi \sigma_w^2}{1 - \phi^2} \end{pmatrix} = \frac{\sigma^2}{1 + \phi^2}.$$

▲

6.1.3. El operador de predicción $P(\cdot|\mathbf{W})$

Dados un vector $\mathbf{W} = (W_n, \dots, W_1)'$ y una variable Y con segundos momentos finitos, hemos visto como obtener el mejor predictor lineal $P(Y|\mathbf{W})$ de Y a partir de $1, W_n, \dots, W_1$ usando de las ecuaciones (6.15) y (6.16). La función $P(\cdot|\mathbf{W})$ que lleva Y a $P(Y|\mathbf{W})$ se conoce como un operador de predicción. Estos operadores tienen una serie de propiedades útiles que pueden usarse para simplificar el cálculo de los mejores predictores lineales.

Propiedades del Operador de Predicción $P(\cdot|\mathbf{W})$:

Suponemos que $E(U^2) < \infty, E(V^2) < \infty, \Gamma = \text{Cov}(\mathbf{W}, \mathbf{W})$ y $\alpha_1, \dots, \alpha_n, \beta$ son constantes.

1. $P(U|\mathbf{W}) = E(U) + \mathbf{a}'(\mathbf{W} - E(\mathbf{W}))$, donde $\Gamma \mathbf{a} = \text{Cov}(U, \mathbf{W})$.
2. $E((U - P(U|\mathbf{W}))\mathbf{W}) = \mathbf{0}$ y $E(U - P(U|\mathbf{W})) = 0$.
3. $E((U - P(U|\mathbf{W}))^2) = \text{Var}(U) - \mathbf{a}' \text{Cov}(U, \mathbf{W})$.
4. $P(\alpha_1 U + \alpha_2 V + \beta|\mathbf{W}) = \alpha_1 P(U|\mathbf{W}) + \alpha_2 P(V|\mathbf{W}) + \beta$.
5. $P(\sum_{i=1}^n \alpha_i W_i + \beta|\mathbf{W}) = \sum_{i=1}^n \alpha_i W_i + \beta$.
6. $P(U|\mathbf{W}) = E(U)$ si $\text{Cov}(U, \mathbf{W}) = \mathbf{0}$.

7. $P(U|\mathbf{W}) = P(P(U|\mathbf{W}, \mathbf{V})|\mathbf{W})$ si \mathbf{V} es un vector aleatorio tal que las componentes de $E(\mathbf{V}\mathbf{V}')$ son todas finitas.

Ejemplo 6.4 (AR(p))

Supongamos que X_t es un proceso AR(p) causal que satisface la ecuación

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + w_t, \quad t \in \mathbb{Z}$$

Si $n > p$ podemos aplicar el operador de predicción P_n a cada lado de la ecuación y usando las propiedades 4, 5 y 6 obtenemos

$$P_n X_{n+1} = \phi_1 X_n + \cdots + \phi_p X_{n+1-p}. \quad (6.18)$$

▲

Ejemplo 6.5 (AR(1))

El proceso Y_t es un proceso AR(1) con media μ si $X_t = Y_t - \mu$ es un proceso AR(1) con media cero. Definimos X_t por la ecuación

$$X_t = \phi X_{t-1} + w_t$$

y ponemos $Y_t = X_t + \mu$ vemos que Y_t satisface la ecuación

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + w_t. \quad (6.19)$$

Si $P_n Y_{n+h}$ es el mejor predictor lineal de Y_{n+h} a partir de $1, Y_n, \dots, Y_1$, una aplicación de P_n a la ecuación (6.19) con $t = n+1, n+2, \dots$ nos da las ecuaciones recursivas

$$P_n Y_{n+h} - \mu = \phi(P_n Y_{n+h-1} - \mu), \quad h \geq 1.$$

Observamos que $P_n Y_n = Y_n$ y podemos resolver estas ecuaciones de manera recursiva para $P_n Y_{n+h}, h \geq 1$. Obtenemos

$$P_n Y_{n+h} = \mu + \phi^h (Y_n - \mu). \quad (6.20)$$

El error cuadrático medio correspondiente es

$$E(Y_{n+h} - P_n Y_{n+h})^2 = \gamma(0)(1 - \mathbf{a}'_n \boldsymbol{\rho}(h)). \quad (6.21)$$

En la clase anterior vimos que $\gamma(0) = \sigma_w^2 / (1 - \phi^2)$ y $\rho(h) = \phi^h, h \geq 0$. Reemplazando $\mathbf{a}_n = (\phi^h, 0, \dots, 0)'$ en (6.21) obtenemos

$$E(Y_{n+h} - P_n Y_{n+h})^2 = \sigma_w^2 (1 - \phi^{2h}) / (1 - \phi^2). \quad (6.22)$$

▲

Observación 6.3 En general, si Y_t es una serie de tiempo estacionaria con media μ y si X_t es la serie centrada definida por $X_y = Y_t - \mu$ entonces la clase de las combinaciones lineales de $1, Y_n, \dots, Y_1$ coincide con la de las combinaciones lineales de $1, X_n, \dots, X_1$ y en consecuencia el predictor lineal de cualquier v.a. W a partir de $1, Y_n, \dots, Y_1$ es igual al predictor lineal que se obtiene a partir de $1, X_n, \dots, X_1$. Si llamamos a este predictor $P_n W$ y aplicamos P_n a la ecuación $Y_{n+h} = X_{n+h} + \mu$ obtenemos

$$P_n Y_{n+h} = \mu + P_n X_{n+h} \quad (6.23)$$

Por lo tanto el mejor predictor lineal de Y_{n+h} se obtiene hallando el mejor predictor lineal de X_{n+h} y sumándole μ . A partir de (6.8) observamos que como $E(X_t) = 0$, $P_n X_{n+h}$ es igual al mejor predictor lineal de X_{n+h} a partir de X_n, \dots, X_1 .

6.2. El algoritmo de Durbin-Levinson

A partir de ahora nos restringimos a la consideración de series de tiempo centradas. Si X_t es una serie de tiempo centrada y estacionaria, las ecuaciones (6.15) y (6.16) resuelven completamente el problema de determinar el mejor predictor lineal $P_n X_{n+h}$ de X_{n+h} a partir de X_n, \dots, X_1 . Sin embargo, este enfoque directo requiere resolver un sistema de n ecuaciones lineales, que puede resultar computacionalmente costoso para n grande.

En general, para procesos estacionarios cualesquiera sería útil si el predictor a un paso $P_n X_{n+1}$ basado en las n observaciones previas se pudiera utilizar para simplificar el cálculo de $P_{n+1} X_{n+2}$, que es el predictor a un paso basado en las $n+1$ observaciones previas. Este tipo de algoritmos de predicción se conocen como algoritmos recursivos. Dos ejemplos importantes son los algoritmos de Durbin-Levinson y de innovación.

A partir de (6.15) y (6.16) sabemos que si la matriz Γ_n no es singular, entonces

$$P_n X_{n+1} = \phi'_n \mathbf{X}_n = \phi_{n1} X_n + \dots + \phi_{nn} X_1, \quad (6.24)$$

donde

$$\phi_n = \Gamma_n^{-1} \gamma_n, \quad \gamma_n = (\gamma(1), \dots, \gamma(n))'$$

y el error medio cuadrático correspondiente es

$$\nu_n = E(X_{n+1} - P_n X_{n+1})^2 = \gamma(0) - \phi'_n \gamma_n.$$

Una condición suficiente para que ninguna de las matrices de autocovarianza $\Gamma_1, \Gamma_2, \dots$ sea singular es que $\gamma(0) > 0$ y $\gamma(h) \rightarrow 0$ cuando $h \rightarrow \infty$.

Teorema 6.2 (Algoritmo de Durbin-Levinson) *Los coeficientes $\phi_{n1}, \dots, \phi_{nn}$ pueden calcularse recursivamente usando las ecuaciones*

$$\phi_{nn} = \left(\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j} \gamma(n-j) \right) \nu_{n-1}^{-1} \quad (6.25)$$

$$\begin{pmatrix} \phi_{n1} \\ \vdots \\ \phi_{n,n-1} \end{pmatrix} = \begin{pmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{pmatrix} - \phi_{nn} \begin{pmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{pmatrix} \quad (6.26)$$

y

$$\nu_n = \nu_{n-1} (1 - \phi_{nn}^2) \quad (6.27)$$

donde $\phi_{11} = \gamma(1)/\gamma(0) = \rho(1)$ y $\nu_0 = \gamma(0)$.

Demostración. Sea R_n la matriz de correlaciones, la definición de ϕ_{11} asegura que la ecuación

$$R_n \phi_n = \boldsymbol{\rho}_n \quad (6.28)$$

donde $\boldsymbol{\rho}_n = (\rho(1), \dots, \rho(n))'$ se satisface para $n = 1$. El primer paso de la demostración es probar que ϕ_n , que se define recursivamente por (6.25) y (6.26), satisface (6.28) para todo n . Supongamos que esto es cierto para $n = k$. Hacemos una partición de R_{k+1} y definimos

$$\boldsymbol{\rho}_k^{(r)} = (\rho(k), \rho(k-1), \dots, \rho(1))'$$

y

$$\phi_k^{(r)} = (\phi_{kk}, \phi_{k,k-1}, \dots, \phi_{k1})'$$

vemos que las recursiones implican

$$\begin{aligned} R_{k+1}\phi_{k+1} &= \begin{pmatrix} R_k & \rho_k^{(r)} \\ \rho_k^{(r)'} & 1 \end{pmatrix} \begin{pmatrix} \phi_k - \phi_{k+1,k+1}\phi_k^{(r)} \\ \phi_{k+1,k+1} \end{pmatrix} \\ &= \begin{pmatrix} \rho_k - \phi_{k+1,k+1}\rho_k^{(r)} + \phi_{k+1,k+1}\rho_k^{(r)} \\ \rho_k^{(r)'}\phi_k - \phi_{k+1,k+1}\rho_k^{(r)'}\phi_k^{(r)} + \phi_{k+1,k+1} \end{pmatrix} \\ &= \rho_{k+1} \end{aligned}$$

como queríamos. Hemos usado el hecho de que si $R_k\phi_k = \rho_k$ entonces $R_k\phi_k^{(r)} = \rho_k^{(r)}$. Esto se puede verificar fácilmente escribiendo las componentes de la ecuación en orden inverso. Como (6.28) se satisface para $n = 1$, por inducción vemos que los coeficientes ϕ_n definidos recursivamente por (6.25) y (6.26) satisfacen (6.28) para todo n .

Sólo falta demostrar que los errores medios cuadráticos

$$\nu_n = E(X_{n+1} - \phi_n' \mathbf{X}_n)^2$$

satisfacen $\nu_0 = \gamma(0)$ y (6.27). El hecho de que $\nu_0 = \gamma(0)$ es consecuencia inmediata de la definición $P_0 X_1 = E(X_1) = 0$. Como hemos mostrado que $\phi_n' \mathbf{X}_n$ es el mejor predictor lineal de X_{n+1} , podemos escribir, a partir de (6.9) y (6.26),

$$\nu_n = \gamma(0) - \phi_n' \gamma_n = \gamma(0) - \phi_{n-1}' \gamma_{n-1} + \phi_{nn} \phi_{n-1}^{(r)'} \gamma_{n-1} - \phi_{nn} \gamma(n).$$

Usando (6.9) de nuevo obtenemos

$$\nu_n = \nu_{n-1} + \phi_{nn} (\phi_{n-1}^{(r)'} \gamma_{n-1} - \gamma(n)),$$

y por (6.25)

$$\nu_n = \nu_{n-1} + \phi_{nn}^2 (\gamma(0) - \phi_{n-1}' \gamma_{n-1}) = \nu_{n-1} (1 - \phi_{nn}^2).$$

■

Observación 6.4 Bajo las condiciones de la proposición, la función α definida por $\alpha(0) = 1$ y $\alpha(n) = \phi_{nn}$, $n \geq 1$ es la función de autocorrelación parcial (PACF) que vimos en la clase anterior.

Ejemplo 6.6

Veamos como podemos usar el algoritmo de Durbin-Levinson para hallar los coeficientes. Comenzamos con

$$\phi_{11} = \rho(1), \quad \nu_1 = \nu_0(1 - \phi_{11}^2) = \gamma(0)(1 - \phi_{11}^2)$$

Para $n = 2$,

$$\begin{aligned} \phi_{22} &= \frac{\gamma(2) - \phi_{11}\gamma(1)}{\gamma(0)(1 - \phi_{11}^2)} = \frac{\rho(2) - \rho(1)\phi_{11}}{1 - \phi_{11}^2} \\ &= \frac{\rho(2) - \phi_{11}^2}{1 - \phi_{11}^2} \end{aligned}$$

$$\phi_{21} = \phi_{11}(1 - \phi_{22})$$

$$\nu_2 = \nu_1(1 - \phi_{22}^2) = \gamma(0)(1 - \phi_{11}^2)(1 - \phi_{22}^2)$$

Para $n = 3$,

$$\begin{aligned}\phi_{33} &= \frac{\gamma(3) - \phi_{21}\gamma(2) - \phi_{22}\gamma(1)}{\gamma(0)(1 - \phi_{11}^2)(1 - \phi_{22}^2)} = \frac{\rho(2) - \rho(1)\phi_{11}}{1 - \phi_{11}^2} \\ &= \frac{\rho(3) - \phi_{21}\rho(2) - \phi_{22}\rho(1)}{(1 - \phi_{11}^2)(1 - \phi_{22}^2)} \\ &= \frac{\rho(3) - \phi_{21}\rho(2) - \phi_{22}\rho(1)}{1 - \phi_{21}\rho(1) - \phi_{22}\rho(2)}, \\ \phi_{31} &= \phi_{21} - \phi_{22}\phi_{33}, \quad \phi_{32} = \phi_{22} - \phi_{33}\phi_{21} \\ \nu_3 &= \nu_2(1 - \phi_{33}^2) = \gamma(0)(1 - \phi_{11}^2)(1 - \phi_{22}^2)(1 - \phi_{33}^2)\end{aligned}$$

y continuamos este procedimiento hasta el último valor deseado. ▲

Una consecuencia importante del algoritmo de Durbin-Levinson es la siguiente

Cálculo iterativo de la PACF

La función de autocorrelación parcial PACF de un proceso estacionario X_t se puede obtener iterativamente a partir de (6.25).

Ejemplo 6.7 (AR(p))

Usando el resultado anterior y poniendo $n = p$ en las ecuaciones (6.18) y (6.24) vemos que para un modelo AR(p)

$$P_p X_{p+1} = \phi_{p1} X_p + \phi_{p2} X_{p-1} + \cdots + \phi_{pp} X_1 = \phi_1 X_p + \phi_2 X_{p-1} + \cdots + \phi_p X_1. \quad (6.29)$$

Este resultado muestra que para un modelo AR(p), el coeficiente de autocorrelación parcial con retardo p , ϕ_{pp} , también es el último coeficiente en el modelo, ϕ_p .

Ejemplo 6.8 (PACF de un AR(2))

Usaremos los resultados del ejemplo 6.6 para calcular los tres primeros valores, ϕ_{11} , ϕ_{22} y ϕ_{33} de la PACF. Usaremos la relación

$$\rho(h) - \phi_1 \rho(h-1) - \phi_2 \rho(h-2) = 0, \quad h \geq 1.$$

para la función de correlación de este proceso, que obtuvimos en clases anteriores.

Para $h = 1, 2, 3$ tenemos

$$\rho(1) = \frac{\phi_1}{1 - \phi_2}, \quad \rho(2) = \phi_1 \rho(1) + \phi_2, \quad \rho(3) = \phi_1 \rho(2) + \phi_2 \rho(1).$$

Por lo tanto,

$$\begin{aligned}\phi_{11} &= \rho(1) = \frac{\phi_1}{1 - \phi_2} \\ \phi_{22} &= \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2} = \frac{\phi_1 \left(\frac{\phi_1}{1 - \phi_2} \right) + \phi_2 - \left(\frac{\phi_1}{1 - \phi_2} \right)^2}{1 - \left(\frac{\phi_1}{1 - \phi_2} \right)^2} \\ \phi_{21} &= \rho(1)(1 - \phi_2) = \phi \\ \phi_{33} &= \frac{\rho(3) - \phi_1 \rho(2) - \phi_2 \rho(1)}{1 - \phi_1 \rho(1) - \phi_2 \rho(2)} = 0.\end{aligned}$$

6.3. El algoritmo de innovación

Este algoritmo se puede aplicar a todas las series con segundo momento finito, aunque no sean estacionarias. Supongamos que X_t es una serie de tiempo centrada con $E|X_t|^2 < \infty$ para todo t y

$$E(X_i X_j) = \kappa(i, j) \quad (6.30)$$

Introducimos la siguiente notación para el mejor predictor a un paso y su error medio cuadrático

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

y

$$\nu_n = E(X_{n+1} - P_n X_{n+1})^2.$$

Introducimos también las innovaciones o errores de predicción a un paso

$$U_n = X_n - \widehat{X}_n.$$

En términos de los vectores $\mathbf{U}_n = (U_1, \dots, U_n)'$ y $\mathbf{X}_n = (X_1, \dots, X_n)$ las últimas ecuaciones se pueden escribir como

$$\mathbf{U}_n = A_n \mathbf{X}_n, \quad (6.31)$$

donde A_n es de la forma

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

Si X_t es estacionario entonces $a_{ij} = -a_j$ con a_j dado por (6.7) con $h = 1$. Esta ecuación implica que A_n no es singular y tiene inversa C_n de la forma

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

El vector de predictores a un paso $\widehat{\mathbf{X}}_n = (X_1, P_1 X_2, \dots, P_{n-1} X_n)'$ puede escribirse como

$$\widehat{\mathbf{X}}_n = \mathbf{X}_n - \mathbf{U}_n = C_n \mathbf{U}_n - \mathbf{U}_n = \Theta_n (\mathbf{X}_n - \widehat{\mathbf{X}}_n), \quad (6.32)$$

donde

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

y \mathbf{X}_n satisfice

$$\mathbf{X}_n = C_n(\mathbf{X}_n - \widehat{\mathbf{X}}_n). \quad (6.33)$$

La ecuación (6.32) se puede escribir como

$$\widehat{X}_{n+1} = \begin{cases} 0, & \text{si } n = 0, \\ \sum_{j=1}^n \theta_{nj}(X_{n+1-j} - \widehat{X}_{n+1-j}), & \text{si } n = 1, 2, \dots \end{cases} \quad (6.34)$$

a partir de la cual podemos calcular recursivamente los predictores a un paso $\widehat{X}_1, \widehat{X}_2, \dots$ una vez que los coeficientes θ_{ij} hayan sido determinados. El siguiente algoritmo genera estos coeficientes y los errores medios cuadráticos $\nu_i = E(X_{i+1} - \widehat{X}_{i+1})^2$ a partir de las covarianzas $\kappa(i, j)$.

El algoritmo de innovación:

Los coeficientes $\theta_{n1}, \dots, \theta_{nm}$ pueden calcularse a partir de las ecuaciones

$$\begin{aligned} \nu_0 &= \kappa(1, 1), \\ \theta_{n, n-k} &= \nu_k^{-1} \left(\kappa(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k, k-j} \theta_{n, n-j} \nu_j \right), \quad 0 \leq k < n. \end{aligned}$$

y

$$\nu_n = \kappa(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n, n-j}^2 \nu_j.$$

La demostración se puede ver en el libro de Brockwell y Davis.

Observación 6.5 Mientras el algoritmo de Durbin y Levinson nos da los coeficientes de X_n, \dots, X_1 en la representación $\widehat{X}_{n+1} = \sum_{j=1}^n \phi_{nj} X_{n+1-j}$, el algoritmo de innovación nos da los coeficientes de $(X_n - \widehat{X}_n), \dots, (X_1 - \widehat{X}_1)$ en el desarrollo $\widehat{X}_{n+1} = \sum_{j=1}^n \theta_{nj}(X_{n+1-j} - \widehat{X}_{n+1-j})$. Este último desarrollo tiene ventajas porque las innovaciones no están correlacionadas y además puede simplificarse considerablemente en el caso de un proceso ARMA(p,q). Una consecuencia inmediata de (6.33) es la representación de X_{n+1} en términos de las innovaciones. Si ponemos $\theta_{n0} = 1$,

$$X_{n+1} = X_{n+1} - \widehat{X}_{n+1} + \widehat{X}_{n+1} = \sum_{j=0}^n \theta_{nj}(X_{n+1-j} - \widehat{X}_{n+1-j}),$$

para $n = 0, 1, 2, \dots$

Ejemplo 6.9 (Predicción recursiva de un MA(1))

Si X_t está definido por $X_t = w_t + \theta w_{t-1}$, entonces

$$\kappa(i, j) = \begin{cases} \sigma^2(1 + \theta^2), & \text{si } j = i, \\ \theta\sigma^2, & \text{si } j = i + 1, \\ 0, & \text{si } |i - j| > 1. \end{cases}$$

Su usamos el algoritmo de innovación obtenemos las relaciones recursivas

$$\begin{aligned} \theta_{n1} &= \nu_{n-1}^{-1} \theta \sigma^2, & \theta_{nj} &= 0, \text{ para } 2 \leq j \leq n, \\ \nu_0 &= (1 + \theta^2) \sigma^2, & \nu_n &= (1 + \theta^2 - \nu_{n-1}^{-1} \theta^2 \sigma^2) \sigma^2 \end{aligned}$$

▲

Observación 6.6 El algoritmo de innovación funciona muy bien para la predicción de procesos MA(q), porque para estos procesos $\theta_{nj} = 0$ para $n - j > q$. Para procesos AR(p) el algoritmo de Durbin-Levinson es más conveniente, porque $\phi_{nj} = 0$ para $n - j > p$.

Predicción a h pasos

Para obtener una predicción a h pasos partimos del resultado

$$P_n(X_{n+k} - P_{n+k-1}N_{n+k}) = 0, \quad k \geq 1. \quad (6.35)$$

Esta relación sigue de (6.11) y el hecho de que

$$E((X_{n+k} - P_{n+k-1}X_{n+k} - 0)X_{n+j-1}) = 0, \quad j = 1, \dots, n.$$

Por lo tanto

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

De nuevo usamos (6.35) y la linealidad de P_n para obtener

$$P_n X_{n+h} = \sum_{j=h}^{n+h-1} \sum_{j=1}^{n+h-1} \theta_{n+h-1,j} (X_{n+h-j} - \widehat{X}_{n+h-j}), \quad (6.36)$$

donde los coeficientes θ_{nj} se obtienen igual que antes a partir del algoritmo de innovación. Podemos expresar el el error medio cuadrático como

$$\begin{aligned} E(X_{n+h} - P_n X_{n+h})^2 &= E(X_{n+h}^2 - E(P_n X_{n+h})^2) \\ &= \kappa(n+h, n+h) - \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}^2 \nu_{n+h-j-1}. \end{aligned} \quad (6.37)$$

6.3.1. Predicción de procesos ARMA

La aplicación del algoritmo de innovación puede simplificarse considerablemente cuando consideramos procesos ARMA causales de la forma

$$\phi(B)X_t = \theta(B)w_t.$$

La idea es aplicar el algoritmo a una transformación del proceso

$$Z_t = \begin{cases} X_t/\sigma, & t = 1, \dots, m, \\ \phi(B)X_t/\sigma, & t > m, \end{cases} \quad (6.38)$$

con $m = \max(p, q)$.

Para facilitar la notación definimos $\theta_0 = 1$ y $\theta_j = 0$ para $j > q$. También supondremos que $p \geq 1$ y $q \geq 1$.

La covarianza γ_X del proceso X_t puede obtenerse por alguno de los métodos que describimos en clases anteriores. A partir de esta función podemos hallar las autocovarianzas $\kappa(i, j) = E(Z_i Z_j)$, $i, j \geq 1$:

$$\kappa(i, j) = \begin{cases} \gamma_X(i - j)/\sigma^2, & 1 \leq i, j \leq m, \\ (\gamma_X(i - j) - \sum_{r=1}^p \phi_r \gamma_X(r - |i - j|))/\sigma^2, & \min(i, j) \leq m < \max(i, j) \leq 2m, \\ \sum_{r=0}^q \theta_r \theta_{r+|i-j|}, & \min(i, j) > m, \\ 0, & \text{en otro caso.} \end{cases} \quad (6.39)$$

Si aplicamos el algoritmo de innovación al proceso Z_t obtenemos

$$\begin{cases} \widehat{Z}_{n+1} = \sum_{j=1}^n \theta_{nj}(Z_{n+1-j} - \widehat{Z}_{n+1-j}), & 1 \leq n < m, \\ \widehat{Z}_{n+1} = \sum_{j=1}^q \theta_{nj}(Z_{n+1-j} - \widehat{Z}_{n+1-j}), & n \geq m, \end{cases} \quad (6.40)$$

donde los coeficientes θ_{nj} y los errores medios cuadráticos $r_n = E(Z_{n+1} - \widehat{Z}_{n+1})^2$ se obtienen recursivamente a partir del algoritmo de innovación con κ definido en (6.39). Lo interesante de los predictores (6.40) es que θ_{nj} se anula cuando $n \geq m$ y $j > q$. Esto es consecuencia del algoritmo de innovación y el hecho de que $\kappa(r, s) = 0$ si $r > m$ y $|r - s| > q$.

Observamos que las ecuaciones (6.38) permiten escribir el proceso $X_n, n \geq 1$ como combinación lineal de $Z_n, n \geq 1$ y recíprocamente, cada Z_n puede escribirse como combinación lineal de $X_n, n \geq 1$. Esto quiere decir que el mejor predictor lineal de cualquier variable Y a partir de $\{1, X_1, \dots, X_n\}$ es el mismo que el mejor predictor lineal de Y a partir de $\{1, Z_1, \dots, Z_n\}$. Denotamos este predictor por $P_n Y$. En particular las predicciones a un paso de Z_{n+1} y X_{n+1} son

$$\widehat{Z}_{n+1} = P_n Z_{n+1} \quad \text{y} \quad \widehat{X}_{n+1} = P_n X_{n+1}.$$

Usando la linealidad de P_n y las ecuaciones (6.38) vemos que

$$\begin{cases} \widehat{Z}_t = \widehat{X}_t/\sigma, & t = 1, \dots, m, \\ \widehat{Z}_t = (\widehat{X}_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p})/\sigma, & t > m, \end{cases} \quad (6.41)$$

que junto a (6.38) muestra que

$$X_t - \widehat{X}_t = \sigma(Z_t - \widehat{Z}_t), \quad \text{para } t \geq 1. \quad (6.42)$$

Si reemplazamos $Z_j - \widehat{Z}_j$ por $(X_j - \widehat{X}_j)/\sigma$ en (6.39) y luego sustituimos en (6.40) obtenemos finalmente

$$\widehat{X}_{n+1} = \begin{cases} \sum_{j=1}^n \theta_{nj}(X_{n+1-j} - \widehat{X}_{n+1-j}), & 1 \leq n < m, \\ \sum_{j=1}^p \phi_j X_{n+1-j} + \sum_{j=1}^q \theta_{nj}(X_{n+1-j} - \widehat{X}_{n+1-j}) & n \geq m, \end{cases} \quad (6.43)$$

y

$$E(X_{n+1} - \widehat{X}_{n+1})^2 = \sigma^2 E(Z_{n+1} - \widehat{Z}_{n+1})^2 = \sigma^2 r_n, \quad (6.44)$$

donde θ_{nj} y r_n se obtienen a partir del algoritmo de innovación con κ usando (6.39). Las ecuaciones (6.43) determinan los predictores a un paso $\widehat{X}_2, \widehat{X}_3, \dots$ de manera recursiva.

Ejemplo 6.10 (Predicción de un AR(p))

Para un proceso AR(p) las ecuaciones se reducen a

$$\widehat{X}_{n+1} = \phi_1 X_n + \dots + \phi_p X_{n+1-p}, \quad n \geq p.$$

Ejemplo 6.11 (Predicción de un MA(q))

Para un proceso de este tipo tenemos

$$\widehat{X}_{n+1} = \sum_{j=1}^{\min(n,q)} \theta_{nj} (X_{n+1-j} - \widehat{X}_{n+1-j}), \quad n \geq 1.$$

donde los coeficientes θ_{nj} se obtienen aplicando el algoritmo de innovación a las covarianza $\kappa(i, j)$ definidas en (6.39). Como los procesos X_t y Z_t/σ son iguales, estas covarianzas se simplifican:

$$\kappa(i, j) = \frac{\gamma_X(i-j)}{\sigma^2} = \sum_{r=0}^{q-|i-j|} \theta_r \theta_{r+|i-j|}.$$

Ejemplo 6.12 (Predicción de un ARMA(1,1))

Si

$$X_t - \phi X_{t-1} = w_t + \theta w_{t-1}$$

y $|\phi| < 1$, las ecuaciones (6.43) se reducen a

$$\widehat{X}_{n+1} = \phi X_n + \theta_{n1} (X_n - \widehat{X}_n), \quad n \geq 1.$$

Para hallar θ_{n1} recordamos que

$$\gamma_X(0) = \sigma^2 \frac{1 + 2\theta\phi + \theta^2}{1 - \phi^2}$$

Sustituyendo en (6.39) obtenemos para $i, j \geq 1$,

$$\kappa(i, j) = \begin{cases} (1 + 2\theta\phi + \theta^2)/(1 - \phi^2), & i = j = 1, \\ 1 + \theta^2, & i = j \geq 2, \\ \theta, & |i - j| = 1, i \geq 1, \\ 0 & \text{en otro caso.} \end{cases}$$

Con estos valores de $\kappa(i, j)$, la recursiones del algoritmo de innovación se reducen a

$$\begin{aligned} r_0 &= (1 + 2\theta\phi + \theta^2)/(1 - \phi^2), \\ \theta_{n1} &= \theta/r_{n-1}, \\ r_n &= 1 + \theta^2 - \theta^2/r_{n-1} \end{aligned} \tag{6.45}$$