SERIES DE TIEMPO
LICENCIATURA EN CIENCIA DE DATOS
IIMAS-UNAM 5

RAMSÉS H. MENA

CLASE 5: MODELOS LINEALES II

17. ACF ARMA

Utilizando la Definición 9 se puede ver que en caso de un proceso ARMA(p,q) la función de autocovarianza de $\{Y_t\}$ en (47) se puede escribir como

$$\gamma(h) \;=\; \text{Cov}(Y_{t+h}, Y_t) = \sigma_\omega^2 \sum_{j=0}^{\infty} \psi_j \, \psi_{t+h}, \quad h \geq 0, \quad (59)$$

donde los pesos ψ_j's se pueden encontrar utilizando (50).

De manera alternativa, en vez de utilizar la representación causal, se podría proceder como lo hemos hecho anteriormente, es decir

$$\begin{aligned}\gamma(h) &= \text{Cov}(Y_{t+h}, Y_t) \\ &= \mathbb{E}\left[\left(\sum_{j=1}^{p} \phi_j \, Y_{t+h-j} + \sum_{j=0}^{q} \theta_j \, \omega_{t+h-j}\right) Y_t\right] \\ &= \sum_{j=1}^{p} \phi_j \, \gamma(h-j) + \sigma_\omega^2 \sum_{j=h}^{q} \theta_j \, \psi_{j-h}, \qquad (60)\end{aligned}$$

para $h \geq 0$. La segunda igualdad se sigue del hecho de que

$$\mathbb{E}(Y_{t+h}\,\omega_t) = \mathbb{E}\left[\left(\sum_{j=0}^{\infty} \psi_j \, \omega_{t+h-j}\right) \omega_t\right] = \psi_h \, \sigma_\omega^2. \quad (61)$$

De la expresión (60) se puede escribir la siguiente ecuación recurrente para las autocovarianzas

$$\gamma(h) - \phi_1 \gamma(h-1) - \cdots - \phi_p \gamma(h-p) = 0, \qquad (62)$$

para $h \geq \text{máx}(p, q+1)$ y con condiciones iniciales

$$\gamma(h) - \sum_{j=1}^{p} \phi_j \gamma(h-j) = \sigma_\omega^2 \sum_{j=h}^{q} \theta_j \, \psi_{j-h}, \qquad (63)$$

para $0 \leq h \leq \text{máx}(p, q+1)$ y donde $\theta_0 := 1$. Las ecuaciones correspondientes para determinar la ACF se encuentran dividiendo (62) y (63) por $\gamma(0)$, es decir $\rho(h) = \gamma(h)/\gamma(0)$.

Ejemplo 4. *Supóngase que se tiene un modelo ARMA$(1,1)$ causal, es decir $Y_t = \phi Y_{r-1} + \theta \omega_{t-1} + \omega_t$ donde $|\theta| < 1$, entonces (62) se reduce a*

$$\gamma(h) - \phi\gamma(h-1) = 0, \quad h = 2, 3, \ldots,$$

de donde se puede ver que

$$\gamma(h) = \phi\gamma(h-1) = \phi^2\gamma(h-2) = \cdots = \phi^{h-1}\gamma(1).$$

Por otro lado, de (63) se tiene que

$$\gamma(h) - \phi\gamma(h-1) = \sigma_\omega^2 \sum_{j=h}^{1} \theta_j \, \psi_{j-h}$$

y de (52) se ve que $\psi_0 = 1$ y $\psi_1 = \theta + \phi$. Así pues, se tiene

$$\begin{aligned}\gamma(0) &= \phi\gamma(1) + \sigma_\omega^2(1 + \theta(\theta + \phi)) \\ \gamma(1) &= \phi\gamma(0) + \sigma_\omega^2\theta.\end{aligned}$$

Resolviendo para $\gamma(0)$ y $\gamma(1)$ se obtiene

$$\begin{aligned}\gamma(0) &= \sigma_\omega^2 \frac{1 + 2\theta\phi + \theta^2}{1 - \phi^2} \\ \gamma(1) &= \sigma_\omega^2 \frac{(1 + \theta\phi)(\phi + \theta)}{1 - \phi^2}.\end{aligned}$$

Por lo tanto,

$$\gamma(h) = \sigma_\omega^2 \frac{(1 + \theta\phi)(\phi + \theta)}{1 - \phi^2} \, \phi^{h-1}. \qquad (64)$$

Dividendo entre $\gamma(0)$ se tiene que la ACF esta dada por

$$\rho(h) = \frac{(1 + \theta\phi)(\phi + \theta)}{1 + 2\theta\phi + \theta^2} \phi^{h-1}, \; h \geq 1. \qquad (65)$$

En el caso de un proceso AR(p), es decir un ARMA$(p,0)$ se verifica fácilmente, de la ecuación (62) que la ACF está determinada mediante la solución a la siguiente ecuación de diferencias

$$\rho(h) - \phi\rho(h-1) - \cdots - \phi_p \rho(h-p) = 0, \; h \geq p. \quad (66)$$

18. Identificación y PACF

En la Definición 8 se presentaron los procesos ARMA como una alternativa de modelos lineales. Un punto importante en el ajuste de estos modelos es la identificación del orden de los componentes tanto autoregresivo, "p", como de promedios móviles, "q". En el caso de un proceso MA(q), la situación se simplifica considerablemente ya que la ACF esta dada por

$$\rho(h) = \begin{cases} \frac{\sum_{j=0}^{q-h} \theta_j \, \theta_{j+h}}{1+\theta_1^2+\cdot+\theta_q^2}, & 1 \leq h \leq q \\ 0 & h > q, \end{cases} \tag{67}$$

y por lo tanto se vale cero después de q rezagos. Es decir si una vez eliminada la tendencia m_t y la ciclicidad de los datos, decidimos modelar mediante un proceso MA(q) entonces una manera de identificar el orden es mediante el uso de la ACF. En la Figura se presenta la ACF " muestral " correspondiente a una simulación de un proceso MA(2). Claramente la ACF se vuelve " despreciable " después del rezago $q = 2$.

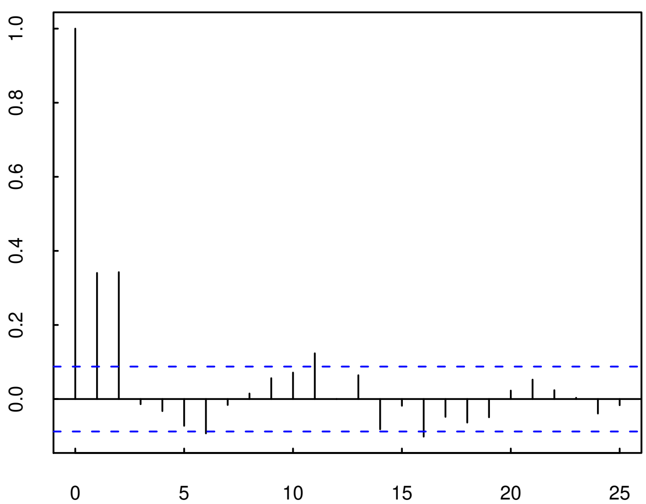

Figura 9: ACF muestral de una realización simulada del proceso $Y_t = 0.8\,\varepsilon_{t-1} + 2.5\,\varepsilon_{t-2} + \varepsilon_t$, donde $\varepsilon_t \sim$ N$(0, 0.8)$.

En términos de la ACF, la situación en el caso de un proceso AR(p) no es tan clara ya que la autocorrelación persiste incluso en aquellos rezagos mayores a p. Por ejemplo en la Figura 3 se presenta la ACF muestral de un modelo AR(1) en donde se ve una clara dependencia hasta el rezago tres y por lo tanto no nos da información contundente acerca del orden verdadero. Una situación similar ocurre para procesos ARMA.

18.1. Función de autocorrelación parcial PACF

De forma ideal se busca una función que se comporte como la ACF de un MA(q), pero para modelos AR(p) y ARMA(p, q). Una función que sirve de manera exitosa para modelos lineales esta dada por la ***función de autocorrelación parcial*** (PACF).

Para motivar la idea de esta función consideremos un modelo AR(1), $Y_t = \phi Y_{t-1} + \omega_t$, entonces

$$\begin{aligned} \gamma(2) &= \text{Cov}(Y_t, Y_{t-2}) = \text{Cov}(\phi Y_{t-1} + \omega_t, Y_{t-2}) \\ &= \text{Cov}(\phi^2 Y_{t-2} + \phi\omega_{t-1} + \omega_t, Y_{t-2}) = \phi^2\gamma(0), \end{aligned}$$

ya que por causalidad $\omega_{t-2}, \omega_{t-3}, \ldots$ son no correlacionados con ω_t y ω_{t-1}. La correlación entre Y_t y Y_{t-2} no es cero como lo sería para un MA(1) ya que Y_t es dependiente de Y_{t-2} a través de Y_{t-1}.

Una manera de romper la dependencia sobre Y_{t-1} de forma " parcial " se puede hacer considerando la correlación entre $Y_t - \phi Y_{t-1}$ y $Y_{t-2} - \phi Y_{t-1}$, es decir la correlación entre Y_t y Y_{t-2} removiendo la dependencia " lineal " sobre Y_{t-1} de cada uno de ellos. De hecho

$$\text{Cov}(Y_t - \phi Y_{t-1}, Y_{t-2} - \phi Y_{t-1}) = \text{Cov}(\omega_t, Y_{t-2} - \phi Y_{t-1}) = 0$$

De manera más general la PACF de un proceso " débilmente " estacionario consiste en remover la " mejor predicción lineal " de Y_h basada en $\{Y_1, Y_2, \ldots, Y_{h-1}\}$. Por tendencia lineal entiéndase

$$Y_h^{h-1} = \beta_1\, Y_{h-1} + \beta_2\, Y_{h-2} + \cdots + \beta_{h-1}\, Y_1, \tag{68}$$

donde las β's están escogidas de manera que $\mathbb{E}(Y_h - Y_h^{h-1})^2$ sea minimizado. De forma análoga se define la mejor predicción lineal de Y_0 basada en $\{Y_1, Y_2, \ldots, Y_{h-1}\}$ denotada por

$$Y_0^{h-1} = \beta_1\, Y_1 + \beta_2\, Y_2 + \cdots + \beta_{h-1}\, Y_{h-1}. \tag{69}$$

Las ecuaciones (68) y (69) se pueden ver como las regresiones lineales de Y_h sobre su pasado y Y_0 sobre su futuro respectivamente.

Definición 11. La ***función de autocorrelación parcial*** de un proceso débilmente estacionario está definida mediante ϕ_{hh}, $h = 1, 2, \ldots$, donde

$$\phi_{hh} := \begin{cases} \text{Corr}(Y_1, Y_2), & h = 1 \\ \text{Corr}(Y_h - Y_h^{h-1}, Y_0 - Y_0^{h-1}), & h \geq 2. \end{cases} \tag{70}$$

Notemos que tanto $Y_h - Y_h^{h-1}$ como $Y_0 - Y_0^{h-1}$ no están correlacionados con $\{Y_1, Y_2, \ldots, Y_{h-1}\}$ y debido a la estacionariedad subyacente a $\{Y_t\}$ la PACF, ϕ_{hh}, es la correlación entre Y_t y Y_{t-h} con la dependencia lineal en $\{Y_1, Y_2, \ldots, Y_{h-1}\}$ removida.

Así pues, si una vez hechas las transformaciones a estacionaridad se decide modelar mediante un proceso AR(p) entonces la PACF muestral se debería de " eliminar " para rezagos mayores a p. El Cuadro (1) describe el comportamiento de la ACF y de la PACF para la diferentes opciones de modelos ARMA causales.

	AR(p)	MA(q)	ARMA(p,q)
ACF	persiste	se corta	persiste
PACF	se corta	se corta	persiste

Cuadro 1: Comportamiento de la ACF y PACF en modelos ARMA causales

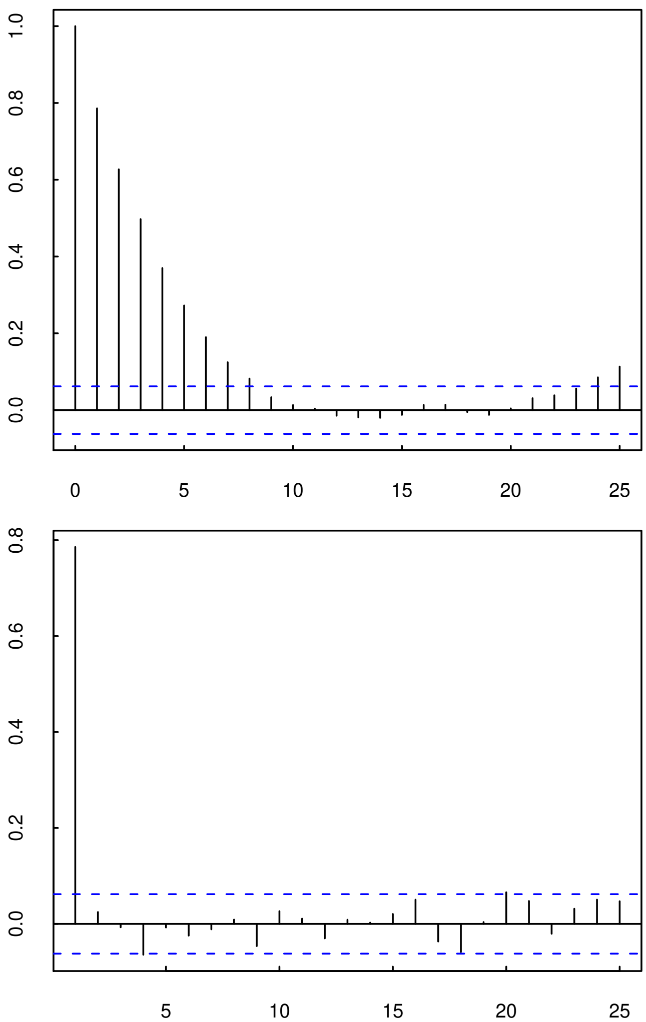

Figura 10: ACF y PACF muestral de una realización simulada del proceso $Y_t = 0.8\,Y_{t-1} + \varepsilon_t$, donde $\varepsilon_t \sim \mathrm{N}(0,1)$ y $Y_0 = 1$.

19. Predicción

En ***predicción***, el principal objetivo es conocer una estimación del valor futuro Y_{n+m} dado una serie de observaciones $\mathbf{Y} = \{Y_n, Y_{n-1}, \ldots, Y_1\}$. Cuando el modelo es estacionario, una manera de conocer dicha predicción es mediante el ***mínimo error cuadrático medio*** dado por

$$Y_{n+m}^n = \mathbb{E}(Y_{n+m} \mid Y_n, Y_{n-1}, \ldots, Y_1). \qquad (71)$$

Si Y_{n+m}^n es lineal, es decir,

$$Y_{n+m}^n = \alpha_0 + \sum_{k=1}^{n} \alpha_k Y_k, \qquad (72)$$

mejor conocido como la ***mejor predicción lineal***, se puede encontrar mediante la solución de las siguientes ecuaciones

$$\begin{aligned} \mathbb{E}[Y_{n+m} - Y_{n+m}^n] &= 0 \\ \mathbb{E}[(Y_{n+m} - Y_{n+m}^n)Y_k] &= 0, \quad k = 1, \ldots, n. \end{aligned} \qquad (73)$$

A las ecuaciones (73) se le conoce como ***ecuaciones de predicción*** y se utilizan para encontrar los coeficientes $\{\alpha_0, \alpha_1, \ldots, \alpha_n\}$. El caso particular de la ***predicción a un paso a futuro***, se puede ver que los coeficientes $\{\phi_{nj}\}_{j=1,\ldots,n}$ en

$$Y_{n+1}^n = \phi_{n1} Y_n + \phi_{n2} Y_{n-2} + \cdot + \phi_{nn} Y_1$$

satisfacen

$$\sum_{j=1}^{n} \phi_{nj} \gamma(k-j) = \gamma(k), \quad k = 1, \ldots, n.$$

Esta ecuación se puede escribir en forma matricial como

$$\Gamma_n \, \Phi_n = \overline{\gamma}_n, \qquad (74)$$

donde $\Gamma_n = \{\gamma(k-j)\}_{j,k=1}^n$ es una matrix de $n \times n$, $\Phi_n = (\phi_{n1}, \ldots, \phi_{nn})'$ y $\overline{\gamma}_n = (\gamma(1), \ldots, \gamma(n))'$ son vectores de dimensión $n \times 1$.

La matrix Γ_n debe de ser no negativa definida, lo cual se cumple en los modelos ARMA. En tal caso se puede ver que

$$\Phi_n = \Gamma_n^{-1} \overline{\gamma}_n.$$

Por ejemplo, en el caso de un proceso AR(p) la mejor predicción a un paso a futuro está dada por

$$Y_{n+1}^n = \phi_1 Y_n + \phi_2 Y_{n-1} + \cdots + +\phi_p Y_{n-p+1},$$

es decir a través de los coeficientes del mismo. Sin embargo, en el caso de un proceso ARMA la situación no es tan sencilla y existen algunos métodos iterativos. Por ejemplo el ***algoritmo de Durbin-Levinson*** ó el ***algoritmo de innovaciones***.

20. Estimación

Para modelos ARMA(p, q) causales e invertibles, la estimación de los parámetros $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ y σ_ω^2 se puede hacer de varias formas, una vez dado el orden de p y q.

20.1. Método de momentos

El método de momentos consiste en igualar los momentos poblacionales a los momentos muestrales para entonces resolver para los parámetros en términos de los momentos muestrales. En el caso de un proceso AR(p) las primeras $p+1$ ecuaciones de (62) y (63), conocidas como las ***ecuaciones de Yule-Walker***

$$\begin{aligned} \gamma(0) &= \phi_1\gamma(1) + \cdots + \phi_p\gamma(p) + \sigma_\omega^2 \qquad (75) \\ \gamma(h) &= \phi_1\gamma(h-1) + \cdots + \phi_p\gamma(h-p), \quad h = 1, \ldots, p \end{aligned}$$

se pueden utilizar para obtener estimadores de las ϕ's. Las ecuaciones (75) se pueden escribir de forma matricial como

$$\Gamma_p\Phi = \overline{\gamma}_p, \quad \sigma_\omega^2 = \gamma(0) - \Phi'\overline{\gamma}_p.$$

Así pues, le método de momentos consiste en reemplazar $\gamma(h)$ por $\hat{\gamma}(h)$ y resolver

$$\hat{\Phi} = \hat{\Gamma}_p^{-1}\hat{\overline{\gamma}}_p, \quad \hat{\sigma}_\omega^2 = \hat{\gamma}(0) - \hat{\overline{\gamma}}_p'\hat{\Gamma}_p^{-1}\hat{\overline{\gamma}}_p.$$

A los estimadores resultantes se les conoce como los ***estimadores de Yule-Walker***.

En el caso de los modelos MA y ARMA los estimadores de momentos no son óptimos y por lo tanto se utiliza máxima verosimilitud, es decir,

$$L(\Theta, \sigma_\omega) = \prod_{t=1}^{n} f(y_t \mid y_{t-1}, \ldots, y_1)$$

En el caso de un proceso ARMA se puede utilizar la mejor predicción lineal a un paso para obtener

$$L(\Theta, \sigma_\omega) = \frac{(2\pi\sigma_\omega^2)^{-n/2}}{\sqrt{r_1^0 r_2^1 \cdot r_n^{n-1}}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{n}\frac{(y_t - y_t^{t-1})^2}{r_t^{t-1}}\right\} \quad (76)$$

donde r_t^{t-1} son los residuales estandarizados de la mejor predicción lineal.

20.2. Criterio AICC

Un problema para la estimación de parámetros es la identificación del orden (p, q). La verosimilitud evaluada en los estimadores se puede utilizar para escoger dicho orden mediante el siguiente criterio; Escoger p,q, ϕ's y θ's tal que se minimize

$$AICC = -2\log L(\hat{\Theta}, \hat{\sigma}_\omega) + \frac{2(p+q+1)n}{n-p-q-2} \qquad (77)$$

21. Modelos ARIMA

En las Secciones 7-10 revisamos varios métodos para transformar datos no estacionarios a datos estacionarios. Un método que revisamos es el de diferenciación mediante el operador (∇), mencionamos también que una tendencia polinomial de orden k, $\mu_t = \sum_{j=0}^{k}\beta_j t^j$, se podía remover mediante la aplicación del operador de diferencias k veces, $\nabla^k X_t$.

Los ***modelos ARMA integrados*** (ARIMA(p, d, q)) son una versión de los modelos ARMA que a su vez incluyen diferenciación. Un proceso se dice ARIMA(p, d, q), si $\nabla^d X_t = (1-B)^d Y_t$ es un proceso ARMA(p, q). En general, el modelo se puede escribir como

$$\phi(B)(1-B)^d X_t = \theta(B)\omega_t. \qquad (78)$$

El nombre integrated viene del hecho que la operación inversa a la diferenciación es la " integración ". Por ejemplo si $Y_t = \nabla X_t$ entonces $X_t = \sum_{i=1}^{t} Y_i + X_0$.

22. Pasos para el ajuste de un modelo ARIMA

De forma genérica se pueden considerar los siguientes pasos

- Inspección gráfica de los datos.
- Transformación de los datos a estacionariedad.
- Identificación del orden.
- Estimación de parámetros.
- Diagnóstico de residuales.

22.1. Inspección gráfica de los datos y transformación de los datos a estacionariedad

El primer paso es dibujar la serie de los datos x_t vs. t para inspeccionar posibles anomalías que puedan afectar la estacionariedad de los datos. Si, por ejemplo, la " variabilidad " cambia en el tiempo entonces un cambio el cual estabilice la varianza, e.g. Box-Cox, se debe de aplicar. Por ejemplo, en aplicaciones económicas y financieras se estila aplicar $\nabla[\log(x_t)]$ el cual tiene sentido práctico al considerar

$x_t = (1 + i_t)x_{t-1}$, ya que $\nabla[\log(x_t)] \approx p_t$, conocido como el ***rendimiento*** o ***tasa de crecimiento***. La mejor manera, tal vez, de medir la relevancia de los datos transformados y saber si más diferenciación es necesaria, es haciéndolo de manera secuencial. Es decir, inspeccionar (e.g. gráficamente) los datos diferenciados y discernir si se necesita otra diferenciación. Esto nos puede llevar a varias opciones de orden de la diferenciación d. Es importante ***no sobrediferenciar*** ya que esto puede llevar a dependencias no deseadas. Otro indicador de que se necesita diferenciación es la tasa de decremento de la ACF, si ésta es muy lenta entonces posiblemente se necesite de más diferenciación.

Por otro lado si los datos tienen características claras, e.g. tendencia, que se pudiesen modelar mediante alguna función paramétrica, entonces se pueden transformar los datos mediante la substracción de dicha función ajustada como se explicó en la Sección 8.

22.2. Identificación del orden y estimación de parámetros

Una vez que algunos valores preliminares de d se han escogido, entonces se pueden utilizar las ACF y PACF para dar valores preliminares de p y q. Por ejemplo, si la ACF(2) y PACF(1) de los datos resultantes de ∇X_t son significativamente diferentes de cero, entonces un modelo tentativo podría ser ARIMA$(1, 1, 2)$. Una vez existentes algunas opciones de p, q y d se procede a la estimación de los parámetros subyacentes. Este último paso se puede realizar para varias opciones de p, q, y d y así escoger el modelo que minimize el coeficiente de AICC.

22.3. Diagnóstico de residuales

Un ajuste adecuado, no necesariamente implica que el modelo ARIMA ajustado es apropiado para la serie ajustada, es por esto que una vez que un cierto modelo se ha ajustado se debe de verificar que los residuales del modelo coincidan con la definición de ruido blanco. De manera gráfica se pueden utilizar los diagramas de Q-Q (Normal o Student-t) y la ACF de los residuales las cuales no deben de presentar ningún rezago significativo.

Tarea proyecto: En equipos de máximo 3 personas, se deberá elaborar el análisis de una serie de datos asignada siguiendo los pasos arriba mencionados. Cada uno de los pasos a utilizar deberán estar justificados. El trabajo deberá ser entregado el 24 de marzo de 2025.