

Series de Tiempo Univariadas

Lección 3:

**Inferencia sobre las funciones de medias y de autocorrelación de un
proceso estacionario**

Manuel Dario Hernandez

Universidad Nacional de Colombia

Presentación basada en las notas de clase del profesor Sergio Calderon (2023).

Inferencia sobre las funciones de medias y de autocorrelación de un proceso estacionario

Filtros infinitos Lineales Invariantes y sus Propiedades.

ARMA para estudiar procesos estacionarios.

La clase de modelos de series temporales lineales, que incluye la clase de modelos autorregresivos de promedio móvil (ARMA), proporciona un marco general para estudiar los procesos estacionarios.

- ✓ De hecho, todo proceso estacionario de segundo orden es un proceso lineal o puede transformarse en un proceso lineal restando un componente determinista. Este resultado se conoce como descomposición de Wold (Brokwell y Davis (2016)).

Filtros infinitos Lineales Invariantes y sus Propiedades.

Basado en Brockwell & Davis (2016) : Sección 2.2.

Definición: Considere un filtro lineal invariante en el tiempo definido como

The time series $\{X_t\}$ is a **linear process** if it has the representation

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \quad \left. \begin{array}{l} \text{es una combinación lineal infinita} \\ \text{de variable de un proceso de RB.} \end{array} \right\} \quad (2.2.1)$$

for all t , where $\{Z_t\} \sim WN(0, \sigma^2)$ and $\{\psi_j\}$ is a sequence of constants with $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$.

filtro lineal invariante con proceso estocástico de entrada $\{Z_t\} \sim RB$.

Nota: A la condición $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ se le conoce como que la serie de valores reales $\{\psi_j\}$ sea **absolutamente convergente**.

Filtros infinitos Lineales Invariantes y sus Propiedades.

tenemos $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} = \sum_{j=-\infty}^{\infty} \psi_j B^j Z_t = \psi(B) Z_t$

✓ In terms of the backward shift operator B , (2.2.1) can be written more compactly as

✓ $X_t = \psi(B) Z_t, \quad (2.2.2)$

where $\psi(B) = \sum_{j=-\infty}^{\infty} \psi_j B^j$.

A linear process is called a moving average or $MA(\infty)$ if $\psi_j = 0$ for all $j < 0$, i.e., if

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \sim MA(\infty)$$

Under some conditions, AR and ARMA are special cases of the above linear process (will be shown later).

recordemos un $MA(1)$ \xrightarrow{RB}

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

Filtros infinitos Lineales Invariantes y sus Propiedades.

tenemos : $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$, $\Rightarrow |X_t| = |\sum \psi_j Z_{t-j}| \leq \underbrace{\sum |\psi_j| |Z_{t-j}|}_{|a+b| \leq |a| + |b|} \xrightarrow{\text{desigualdad triangular}}$

Remark 1. The condition $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ ensures that the infinite sum in (2.2.1) converges (with probability one), since $E |Z_t| \leq \sigma$ and

$$E |X_t| \leq \sum_{j=-\infty}^{\infty} (|\psi_j| E |Z_{t-j}|) \leq \left(\sum_{j=-\infty}^{\infty} |\psi_j| \right) \sigma < \infty.$$

It also ensures that $\sum_{j=-\infty}^{\infty} \psi_j^2 < \infty$ and hence converges in mean square, i.e., that X_t is the mean square limit of the partial sums $\sum_{j=-n}^n \psi_j Z_{t-j}$.

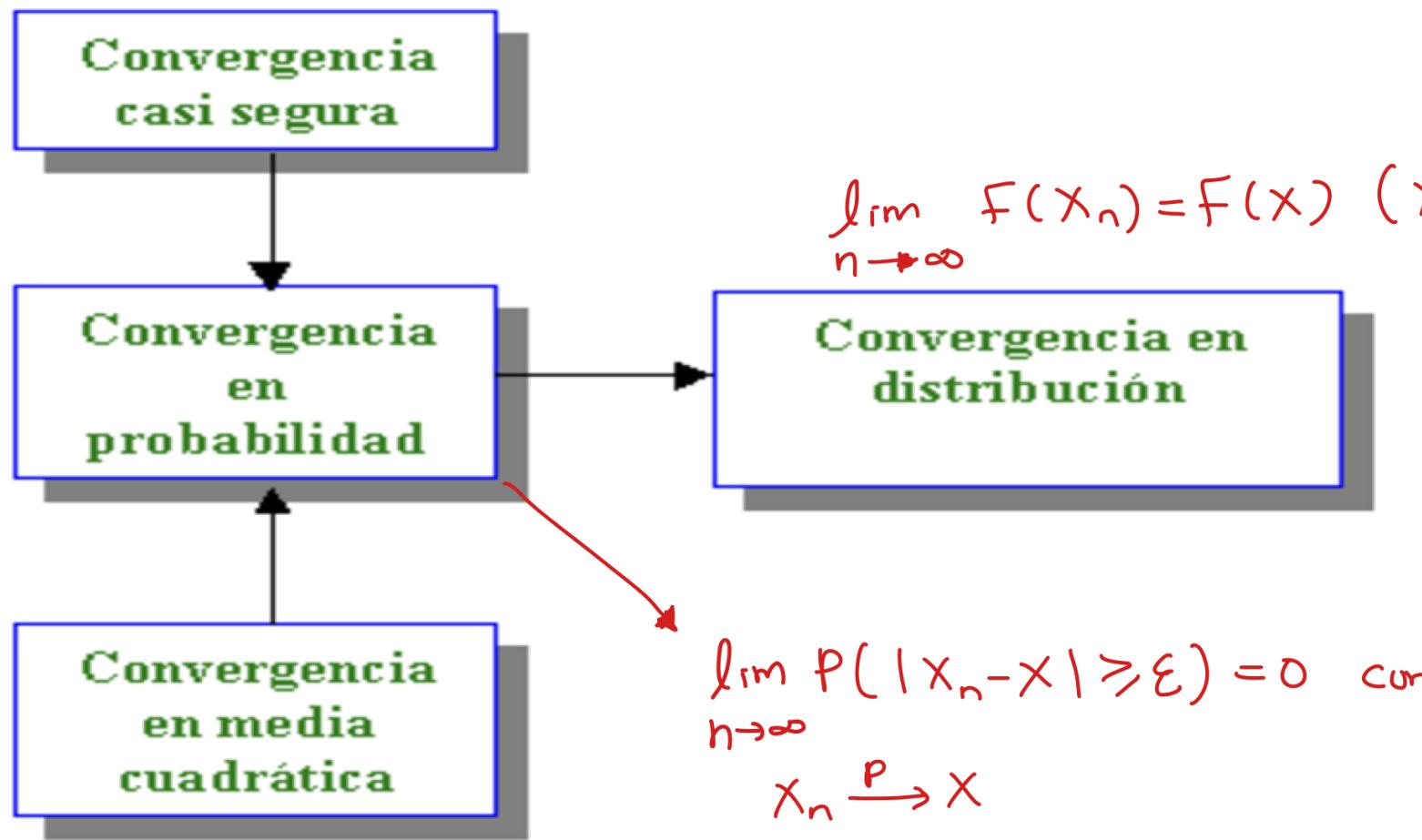
The condition $\sum_{j=-n}^n |\psi_j| < \infty$ also ensures convergence in both senses of the more general series (2.2.3) considered in the following Proposition (slide 10).

Cuando Z_t no es RB sino un proceso estacionario

Convergencia absoluta significa
que la serie converge incluso
al tomar el cuadrado

Review: convergence criteria

$$P\left(\lim_{n \rightarrow \infty} X_n = x\right) = 1 \quad (X_n \xrightarrow{cs} x)$$



$$\lim_{n \rightarrow \infty} f(X_n) = f(x) \quad (X_n \xrightarrow{d} x)$$

$$\lim_{n \rightarrow \infty} P(|X_n - x| \geq \varepsilon) = 0 \quad \text{con } \varepsilon > 0$$
$$X_n \xrightarrow{P} x$$

Review: Mean Square Convergence (Convergencia en media cuadrática)



The sequence S_n of random variables is said to converge in mean square to the random variable S if

$$E(S_n - S)^2 \rightarrow 0 \text{ as } n \rightarrow \infty.$$

$$S_n = \sum_{k=1}^n X_k$$

In particular, we say that the sum $\sum_{k=1}^n X_k$ converges (in mean square) if there exists a random variable S such that $E\left(\sum_{k=1}^n X_k - S\right)^2 \rightarrow 0$ as $n \rightarrow \infty$. If this is the case, then we use the notation $S = \sum_{k=1}^{\infty} X_k$.

The Cauchy Criterion

Para chegar Convergencia en
media cuadrática

For a given sequence S_n of random variables to converge in mean square to *some* random variable, it is necessary and sufficient that

$$E(S_m - S_n)^2 \rightarrow 0 \text{ as } m, n \rightarrow \infty$$

(for a proof of this see Brockwell and Davis (1991), Chapter 2). The point of the criterion is that it permits checking for mean square convergence without having to identify the limit of the sequence.

Review: Mean Square Convergence (Example)

Consider the sequence of partial sums $S_n = \sum_{t=-n}^n a_t Z_t$, $n = 1, 2, \dots$, where $\{Z_t\} \sim WN(0, \sigma^2)$. Under what conditions on the coefficients a_i does this sequence converge in mean square? To answer this question we apply the Cauchy criterion as follows. For $n > m > 0$,

$$\begin{aligned} E(S_n - S_m)^2 &= E \left(\sum_{i=-n}^n a_i Z_i - \sum_{i=-m}^m a_i Z_i \right)^2 = E \left(\sum_{m < |i| \leq n} a_i Z_i \right)^2 \\ &= E \left[\sum_{m < |i| \leq n} \sum_{m < |j| < n} a_i a_j Z_i Z_j \right] \quad \text{as } \left(\sum_{i=1}^N a_i \right)^2 = \sum_{i=1}^N \sum_{j=1}^N a_i a_j \\ &= \sum_{m < |i| \leq n} \sum_{m < |j| < n} a_i a_j E[Z_i Z_j] \\ &= \sigma^2 \sum_{m < |i| \leq n} a_i^2. \end{aligned}$$

Definición S_n y S_m

ya que $n > m > 0$

Review: Mean Square Convergence (Example)

Consequently, $E(S_n - S_m)^2 \rightarrow 0$ if and only if $\sum_{m < |i| \leq n} a_i^2 \rightarrow 0$. Since the Cauchy criterion applies also to real-valued sequences, this last condition is equivalent to convergence of the sequence $\sum_{i=-n}^n a_i^2$, or equivalently to the condition

$$\sum_{i=-\infty}^{\infty} a_i^2 < \infty$$

Properties of Mean Square Convergence:

If $X_n \rightarrow X$ and $Y_n \rightarrow Y$, in mean square as $n \rightarrow \infty$, then

- (a) $E(X_n^2) \rightarrow E(X^2)$,
- (b) $E(X_n) \rightarrow E(X)$,

and

- (c) $E(X_n Y_n) \rightarrow E(XY)$.

Review: Mean Square Convergence (Example)

recordemos

$$E \left(\sum_{i=-n}^n a_i Z_i - \sum_{i=-m}^m a_i Z_i \right)^2 = E \left(\sum_{m < |i| \leq n} a_i Z_i \right)^2$$

Suppose $n > m > 0$, with $n = 3$ and $m = 2$. What do we have for
 $\sum_{i=-n}^n a_i Z_i - \sum_{i=-m}^m a_i Z_i$?

$$\begin{aligned} \sum_{i=-3}^3 a_i Z_i - \sum_{i=-2}^2 a_i Z_i &= a_{-3} Z_{-3} + a_{-2} Z_{-2} + a_{-1} Z_{-1} + a_0 Z_0 + a_1 Z_1 + a_2 Z_2 + a_3 Z_3 \\ &\quad - (a_{-2} Z_{-2} + a_{-1} Z_{-1} + a_0 Z_0 + a_1 Z_1 + a_2 Z_2) \\ &= a_{-3} Z_{-3} + a_3 Z_3 \end{aligned}$$

Note que quedan las colas, es decir, los valores mas grandes que m y mas pequeños o iguales a n .

Filtros infinitos Lineales Invariantes y sus Propiedades.

The operator $\psi(B)$ can be thought of as a linear filter, which when applied to the white noise "input" series $\{Z_t\}$ produces the "output" $\{X_t\}$.

As established in the following proposition, a linear filter, when applied to any stationary input series, produces a stationary output series.

Proposition

Let $\{Y_t\}$ be a stationary time series with mean 0 and covariance function γ_Y . If $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, then the time series

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j} = \psi(B) Y_t \quad (2.2.3)$$

is stationary with mean 0 and autocovariance function

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_Y(h+k-j). \quad (2.2.4)$$

In the special case where $\{X_t\}$ is the linear process (2.2.1),

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} \sigma^2. \quad (2.2.5)$$



Proof: The argument used in Remark 1, with σ replaced by $\sqrt{\gamma_Y(0)}$, shows that the series in (2.2.3) is convergent. Since $EY_t = 0$, we have

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

and

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

which shows that $\{X_t\}$ is stationary with covariance function (2.2.4). (The interchange of summation and expectation operations in the above calculations can be justified by the absolute summability of ψ_j .)

Filtros infinitos Lineales Invariantes y sus Propiedades.

Finally, if $\{Y_t\}$ is the white noise sequence $\{Z_t\}$ in (2.2.1), then

$\gamma_Y(h - j + k) = \sigma^2$ if $k = j - h$ and 0 otherwise, from which (2.2.5) follows. Other way:

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

Remark 2. The absolute convergence of (2.2.3) implies that filters of the form $\alpha(B) = \sum_{j=-\infty}^{\infty} \alpha_j B^j$ and $\beta(B) = \sum_{j=-\infty}^{\infty} \beta_j B^j$ with absolutely summable coefficients can be applied successively to a stationary series $\{Y_t\}$ to generate a new stationary series

$$W_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j},$$

where

$$\psi_j = \sum_{k=-\infty}^{\infty} \alpha_k \beta_{j-k} = \sum_{k=-\infty}^{\infty} \beta_k \alpha_{j-k}. \quad (2.2.6)$$

These relations can be expressed in the equivalent form

$$W_t = \psi(B) Y_t,$$

where

$$\psi(B) = \alpha(B)\beta(B) = \beta(B)\alpha(B), \quad (2.2.7)$$

and the products are defined by (2.2.6) or equivalently by multiplying the series $\sum_{j=-\infty}^{\infty} \alpha_j B^j$ and $\sum_{j=-\infty}^{\infty} \beta_j B^j$ term by term and collecting powers of B . It is clear from (2.2.6) and (2.2.7) that the order of application of the filters $\alpha(B)$ and $\beta(B)$ is immaterial.

Filtros infinitos Lineales Invariantes y sus Propiedades (Ejercicio).

Consider the AR(1) model: $X_t = \phi X_{t-1} + Z_t$, with $\{Z_t\} \sim WN(0, \sigma^2)$ and ϕ is a real-valued constant. Assume that $|\phi| < 1$. Show that this AR(1) model can be expressed as a linear process ($X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$):

Filtros infinitos Lineales Invariantes.

Nota: Al proceso lineal $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$ (Ec 2.2.1 en diapositiva 3) se le puede incluir una constante tal que

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$$

así, el proceso $\{X_t\}$ tiene media μ , y su covarianza no se ve afectada.

Filtros infinitos Lineales Invariantes.

Nota: Al proceso lineal $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$ (Ec 2.2.1 en diapositiva 3) se le puede incluir una constante tal que

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$$

así, el proceso $\{X_t\}$ tiene media μ , y su covarianza no se ve afectada.

Nota:. Un proceso lineal $\{X_t\}$ está completamente caracterizado por su media μ y su función de autocovarianza $\gamma(h)$ o autocorrelación $\rho(h)$ puesto que todas las relaciones entre las variables del proceso son lineales. Sin embargo, un proceso lineal visto como un modelo estadístico para los datos $\{x_1, \dots, x_T\}$ es inconveniente, por qué?

Propiedades de la media muestral y de la función de autocorrelación

Un proceso estacionario $\{X_t\}$ se caracteriza, al menos desde un punto de vista de segundo orden, por su media μ y su función de autocovarianza $\gamma(\cdot)$.

La estimación de μ , $\gamma(\cdot)$ y la función de autocorrelación $\rho(\cdot) = \gamma(\cdot)/\gamma(0)$ a partir de las observaciones x_1, \dots, x_n desempeñan, por tanto, un papel crucial en los problemas de inferencia y, en particular, en el problema de construir un modelo apropiado para los datos.

En esta sección examinamos algunas de las propiedades de las estimaciones muestrales \bar{x} y $\hat{\rho}(\cdot)$ de μ y $\rho(\cdot)$, respectivamente.

Adicionalmente, indicaremos como hacer inferencia acerca de la función de autocorrelación parcial. Note que sólo tenemos disponible una muestra x_1, \dots, x_n , es por esto que debemos suponer que el proceso es estacionario.

Estimación de la Media

El estimador de momento de la media μ de un proceso estacionario es la media muestral

$$\bar{X}_n = \frac{1}{n} (X_1 + X_2 + \cdots + X_n).$$

Es un estimador insesgado de μ , ya que

$$E(\bar{X}_n) = n^{-1} (EX_1 + \cdots + EX_n) = \mu.$$

Veamos el comportamiento del error cuadrático medio, es decir, $E[(\bar{X}_n - \mu)^2]$ cuando $n \rightarrow \infty$.

Estimación de la Media

Proposición

Si $\{X_t\}$ es una serie temporal estacionaria con media μ y función de autocovarianza $\gamma(\cdot)$, entonces cuando $n \rightarrow \infty$,

$$\text{Var}(\bar{X}_n) = E(\bar{X}_n - \mu)^2 \rightarrow 0 \quad \text{si } \gamma(n) \rightarrow 0,$$

$$nE(\bar{X}_n - \mu)^2 \rightarrow \sum_{|h|<\infty} \gamma(h) \quad \text{si } \sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty.$$

Estimación de la Media

Demostración

$$\text{Var}(\bar{X}_n) = E(\bar{X}_n - \mu)^2$$

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2} \text{Var}(X_1 + \cdots + X_n)$$

Entonces

$$n \text{Var}(\bar{X}_n) = \frac{1}{n} \text{Cov}(X_1 + \cdots + X_n, X_1 + \cdots + X_n)$$

$$n \text{Var}(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(X_i, X_j)$$

Miremos la intuición del último término:

Estimación de la Media

si $n = 2$

$$\begin{aligned}\frac{1}{2} \sum_{i,j=1}^2 Cov(X_i, X_j) &= \frac{1}{2}[Cov(X_1, X_1) + Cov(X_1, X_2) + Cov(X_2, X_1) + Cov(X_2, X_2)] \\ &= \frac{1}{2}[2\gamma(0) + \gamma(-1) + \gamma(1)]\end{aligned}$$

$n = 3$

$$\begin{aligned}\frac{1}{3} \sum_{i,j=1}^3 Cov(X_i, X_j) &= \frac{1}{3}[Cov(X_1, X_1) + Cov(X_1, X_2) + Cov(X_1, X_3) + Cov(X_2, X_1) + \\ &\quad Cov(X_2, X_2) + Cov(X_2, X_3) + Cov(X_3, X_1) + Cov(X_3, X_2) + Cov(X_3, X_3)] \\ &= \frac{1}{3}[\gamma(0) + \gamma(-1) + \gamma(-2) + \gamma(1) + \gamma(0) + \gamma(-1) + \gamma(2) + \gamma(1) + \gamma(0)] \\ &= \frac{1}{3} \sum_{h=-2}^2 (3 - |h|)\gamma(h)\end{aligned}$$

En general:

$$\begin{aligned} n \text{Var}(\bar{X}_n) &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(X_i, X_j) \\ &= \frac{1}{n} \sum_{h=-n+1}^{n-1} (n - |h|) \gamma(h) \\ &= \sum_{|h| < n} \left(1 - \frac{|h|}{n}\right) \gamma(h) \end{aligned}$$

Ahora bien, si $\gamma(h) \rightarrow 0$ cuando $h \rightarrow \infty$, el lado derecho de la última expresión converge a cero, de modo que \bar{X}_n converge en media cuadrática a μ y así en probabilidad al mismo límite.

Estimación de la Media

$$nE(\bar{X}_n - \mu)^2 \rightarrow \sum_{|h|<\infty} \gamma(h) \quad \text{si } \sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty.$$

Para la segunda parte de la demostración, expandiendo la sumatoria

$$\sum_{|h|< n} \left(1 - \frac{|h|}{n}\right) \gamma(h)$$

$$\begin{aligned} &= \gamma(0) + \underbrace{\left(1 - \frac{1}{n}\right)}_{<1} \gamma(1) + \underbrace{\left(1 - \frac{1}{n}\right)}_{<1} \gamma(-1) + \underbrace{\left(1 - \frac{2}{n}\right)}_{<1} \gamma(2) + \cdots + \underbrace{\left(1 - \frac{n-1}{n}\right)}_{<1} \gamma(n-1) \\ &+ \underbrace{\left(1 - \frac{n-1}{n}\right)}_{<1} \gamma(-(n-1)) \end{aligned}$$

por propiedad de valor absoluto y dado que cada $\left(1 - \frac{n-1}{n}\right) < 1$, entonces

$$\leq |\gamma(0)| + |\gamma(1)| + |\gamma(-1)| + |\gamma(2)| + |\gamma(-2)| + \cdots + |\gamma(n-1)| + |\gamma(-(n-1))|$$

$$= \sum_{|h|< n} |\gamma(h)|$$

Entonces: $\sum_{|h|< n} \left(1 - \frac{|h|}{n}\right) \gamma(h) \leq \sum_{|h|< n} |\gamma(h)|$

Así, si $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$ entonces:

$$\begin{aligned}
 \lim_{n \rightarrow \infty} nVar(\bar{X}_n) &= \lim_{n \rightarrow \infty} \sum_{|h|< n} \left(1 - \frac{|h|}{n}\right) \gamma(h) \\
 &= \lim_{n \rightarrow \infty} \sum_{|h|< n} \gamma(h) - \lim_{h \rightarrow \infty} \sum_{|h|< n} \frac{|h|}{n} \gamma(h) \\
 &= \sum_{h=-\infty}^{\infty} \gamma(h).
 \end{aligned}$$

Estimación de la Media

Nota Si $X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$ con $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, entonces $\sum_{j=-\infty}^{\infty} |\gamma(h)| < \infty$ y así

$$nVar(\bar{X}_n) \rightarrow \sum_{h=-\infty}^{\infty} \gamma(h) = \sigma^2 \left(\sum_{j=-\infty}^{\infty} \psi_j \right)^2.$$

Nota. Lo que se ha probado básicamente es que \bar{X}_n converge en media cuadrática a μ y así en probabilidad al mismo límite. Asimismo, es posible ver que se cumple el teorema del límite central para proceso estocásticos estacionarios con la condición que $\sum_{j=-\infty}^{\infty} |\gamma(h)| < \infty$, y no sólo para procesos que son I.I.D.

Estimación de la Media

Definición

Una sucesión de VA's $\{X_n\}$ se dice que es asintóticamente normal con "media" μ_n y "desviación estándar" σ_n , si $\sigma_n > 0$, para n suficientemente grande y

$$\sigma_n^{-1}(X_n - \mu_n) \Rightarrow Z \quad Z \sim N(0, 1).$$

Teorema

Si $\{X_t\}$ es un proceso estocástico estacionario tal que $X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$ donde $\{Z_t\} \sim IID(0, \sigma^2)$ y $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ y $\sum_{j=-\infty}^{\infty} \psi_j \neq 0$ entonces \bar{X}_n es $AN(\mu, n^{-1}v)$ con $v = \sum_{h=-\infty}^{\infty} \gamma(h) = \sigma^2 \left(\sum_{j=-\infty}^{\infty} \psi_j \right)^2$

Estimación de la Media

Nota. Del teorema anterior tenemos que:

- Con este teorema uno puede encontrar intervalos de confianza aproximados para μ cuando el tamaño de la muestra es suficientemente grande.
- Por supuesto si $\{X_n\}$ es estacionario y Gaussiano tenemos que la distribución de \bar{X}_n es exacta y del segundo resultado de la proposición (Diapositiva 18) tenemos que

$$n^{-1/2} (\bar{X}_n - \mu) \sim N \left(0, \sum_{|h| < n} \left(1 - \frac{|h|}{n} \right) \gamma(h) \right).$$

Sin embargo, si no se conoce a $\gamma(h)$, se puede cambiar una estimación $\hat{\gamma}(h)$. En vez de incluir todos los rezagos hasta n en la suma se restringe a \sqrt{n} o $n/4$.

- Si el proceso $\{X_t\}$ no es Gaussiano, entonces la distribución de \bar{X}_n es aproximada y se usa la misma expresión anterior.

Estimación de $\gamma(\cdot)$ y $\rho(\cdot)$

Vamos a considerar los siguientes estimadores de $\gamma(h)$ y $\rho(h)$

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (X_t - \bar{X}_n)(X_{t+h} - \bar{X}_n) \quad 0 \leq h \leq n-1$$

y

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

$\hat{\gamma}(h)$ es un estimador sesgado de $\gamma(h)$. Note que para $h = 0$, el estimador es

$$\hat{\gamma}(0) = \frac{1}{n} \sum_{t=1}^n (X_t - \bar{X}_n)^2,$$

el cual tiene denominador n y no $n - 1$ como es usual. Dividir entre $n - 1$ asegura que el estimador sea insesgado.

Estimación de $\gamma(\cdot)$ y $\rho(\cdot)$

Se puede verificar que para cada $n \geq 1$ la matriz

$$\hat{\Gamma}_n = \begin{bmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \cdots & \hat{\gamma}(n-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \cdots & \hat{\gamma}(n-2) \\ \vdots & \vdots & \cdots & \vdots \\ \hat{\gamma}(n-1) & \hat{\gamma}(n-2) & \cdots & \hat{\gamma}(0) \end{bmatrix}$$

es definida no negativa ya que se puede verificar que

$$\hat{\Gamma}_n = n^{-1}TT'$$

donde T es una matriz $n \times 2n$

$$T = \begin{bmatrix} 0 & \cdots & 0 & Y_1 & Y_2 & \cdots & Y_n \\ 0 & \cdots & 0 & Y_1 & Y_2 & \cdots & Y_n & 0 \\ \vdots & & & & & & & \vdots \\ 0 & Y_1 & Y_2 & \cdots & Y_n & 0 & \cdots & 0 \end{bmatrix}$$

con $Y_i = X_i - \bar{X}_n \quad i = 1, 2, 3, \dots, n.$

Estimación de $\gamma(\cdot)$ y $\rho(\cdot)$

Luego, para cada vector $\underset{\sim}{a} \in \mathbb{R}^n$ se puede verificar que

$$\underset{\sim}{a}' \hat{\Gamma}_n \underset{\sim}{a} = n^{-1} (\underset{\sim}{a}' T) (\underset{\sim}{a}' T)' \geq 0.$$

También la matriz de correlación

$$\hat{R}_n = \frac{\hat{\Gamma}_n}{\hat{\gamma}(0)}$$

es definida no negativa.

Nota 5.8. En ocasiones, en la definición de $\hat{\gamma}(h)$, es reemplazado el factor n^{-1} por $(n - h)^{-1}$, pero las matrices $\hat{\Gamma}_n$ y \hat{R}_n pueden no ser definidas no-negativas.

Nota 5.10. *Desde el punto de vista práctico tenemos lo siguiente:*

- *Basado en solo la información X_1, X_2, \dots, X_n , no es posible estimar $\gamma(k)$ para $k \geq n$.*
- *Para valores de k un poco mas pequeños que n los estimadores $\hat{\gamma}(k)$ no son muy confiables ya que de sólo intervienen un poco de pares (X_t, X_{t+k}) , por ejemplo para $k = n - 1$ solo se usa un único par.*
- *Box-Jenkins sugieren que son útiles las estimaciones de la autocorrelación $\rho(k)$ puede ser hechas cuando n es mayor a 50 y para valores de $k \leq \frac{n}{4}$ o $k \leq \sqrt{n}$.*

Estimación de $\gamma(\cdot)$ y $\rho(\cdot)$

Una de las cosas importantes al identificar un modelo ARMA(p,q) para un conjunto de observaciones, es que se identifiquen cuales autocorrelaciones poblacionales(basadas en las muestrales) son significativamente diferentes de 0. Para eso es útil el siguiente resultado:

Teorema 5.11. Si $\{X_t\}$ es un proceso estocástico estacionario tal que

$$X_t - \mu = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} \quad \{Z_t\} \sim IID(0, \sigma^2)$$

donde $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ y $EZ_t^4 < \infty$, entonces para cada $h = 1, 2, 3, \dots$ tenemos que

$$\hat{\rho}(h) \text{ es } AN(\rho(h), n^{-1}W)$$

donde

$$\hat{\rho}(h) = (\hat{\rho}(1), \hat{\rho}(2), \dots, \hat{\rho}(h))$$

$$\rho(h) = (\rho(1), \rho(2), \dots, \rho(h))$$

y W es la matriz de covarianza, tiene entradas

$$\begin{aligned} W_{ij} = \sum_{k=-\infty}^{\infty} & \{ \rho(k+i)\rho(k+j) + \rho(k-i)\rho(k+j) + 2\rho(i)\rho(j)\rho^2(k) \\ & - 2\rho(i)\rho(k)\rho(k+j) - 2\rho(j)\rho(k)\rho(k+i) \} \end{aligned}$$

Una forma de relajar el supuesto que $E[Z_t^4] < \infty$ es dada en el siguiente teorema.

Ver demostración en Brockwell and Davis (2009), sección 7.3.

Estimación de $\gamma(\cdot)$ y $\rho(\cdot)$

Teorema 5.12. Si $\{X_t\}$ es estacionario tal que

$$X_t - \mu = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} \quad \{Z_t\} \sim RB(0, \sigma^2)$$

donde $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ y $\sum_{j=-\infty}^{\infty} \psi_j^2 |j| < \infty$, entonces para cada $h \in \{1, 2, \dots\}$

$$\hat{\rho}(h) \quad \text{es} \quad AN(\rho(h), n^{-1}W)$$

Ver demostración en Brockwell and Davis (2009), sección 7.3.

Nota 5.13. (Fórmula de Bartlett)

Se puede verificar que

$$W_{ij} = \sum_{k=1}^{\infty} \{\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)\} \times \{\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)\}$$

Ejemplo 1:

Usando la expresión de la diapositiva anterior para W_{ij} , demuestre lo siguiente:

iid Noise

If $\{X_t\} \sim \text{IID}(0, \sigma^2)$, then $\rho(h) = 0$ for $|h| > 0$, so from (2.4.10) we obtain

$$w_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

Para n suficientemente grande $\hat{\rho}(1), \dots, \hat{\rho}(h)$ son variables aleatorias aproximadamente independientes e idénticamente distribuidas normales de media cero y varianza n^{-1} . Así un IC aproximado para cada $\rho(1), \dots, \rho(h)$ es

Este resultado es la base para la prueba de que los datos se generan a partir del ruido iid utilizando la ACF muestral descrita en la Lección 2 (Diapositiva 32).

$$\hat{\rho}(h) \pm Z_{1-\alpha/2} n^{-1/2}.$$

El anterior procedimiento puede ser usado para comprobar si las observaciones en verdad provienen de un proceso IID, o más específicamente si es no autocorrelacionado hasta el rezago h . La siguiente figura muestra un ejemplo de ACF de una serie IID en R usando `acf`.

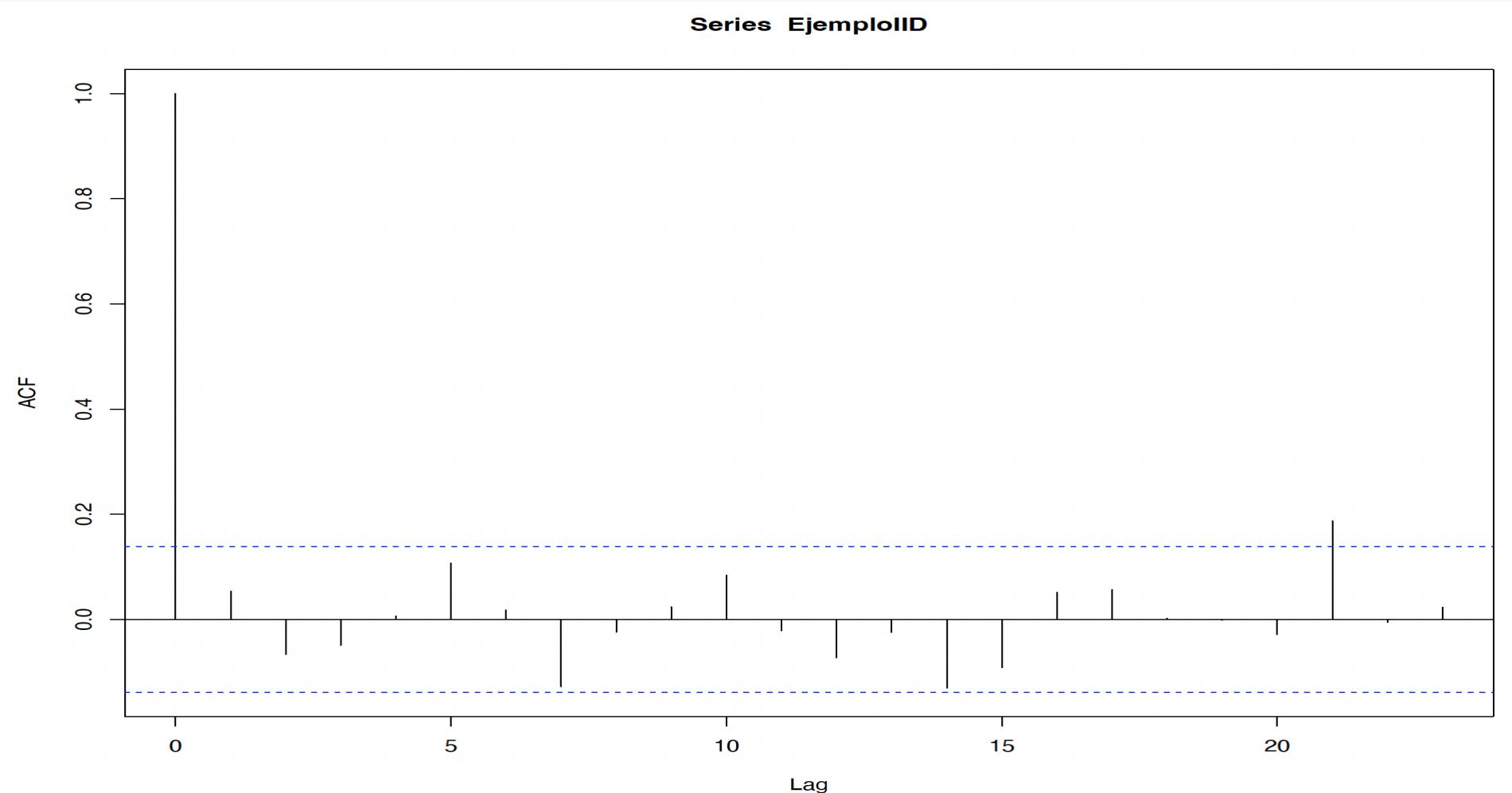


Figura : Gráfica ACF de una serie IID en R usando `acf`

Ejemplo 2:

Consideremos que $\{X_t\}$ es un proceso MA(1) tal que

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

donde $\{Z_t\} \sim WN(0, \sigma^2)$ y θ es una cte de valor real. Para este modelo se probó en tarea 1 que la FACV era:

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

Así que la FAC (ACF en inglés) es:

$$\rho_X(h) = \begin{cases} 1, & \text{if } h = 0, \\ \theta / (1 + \theta^2), & \text{if } h = \pm 1, \\ 0, & \text{if } |h| > 1. \end{cases}$$

Ejemplo 2: Cont.

Usando nuevamente la ecuación para w_{ij} (Diapositiva 33), se puede verificar que las entradas de la diagonal principal de la matriz W quedan determinadas como sigue:

$$w_{ii} = \begin{cases} 1 - 3\rho^2(1) + 4\rho^4(1), & \text{si } i = 1, \\ 1 + 2\rho^2(1), & \text{si } i > 1, \end{cases}$$

lo que quiere decir, es que estas son las varianzas aproximadas de $n^{-1/2}(\hat{\rho}(i) - \rho(i))$.

Para ver si los datos de la muestra X_1, \dots, X_n son compatibles con un MA(1), primero verificamos que no sea IID, usando la metodología anterior. Después, como se sabe que en un MA(1) : $\rho(1) \neq 0$ y $\rho(i) = 0$ para $i > 1$, entonces procedemos a probar que la autocorrelación $\rho(1)$ sea significativa, mientras que las autocorrelaciones para rezagos más altos no lo sean.

Por lo tanto, usaremos las varianzas asintóticas para probar su significancia.

Ejemplo 2: Cont.

Para probar la hipótesis nula $H_0 : \rho(1) = 0$, entonces calculamos el estadístico de prueba bajo la hipótesis nula

$$Z = \frac{\hat{\rho}(1) - 0}{\frac{1}{n^{1/2}} \sqrt{1 - 3\rho^2(1) + 4\rho^4(1)}} \quad \text{es} \quad AN(0, 1),$$

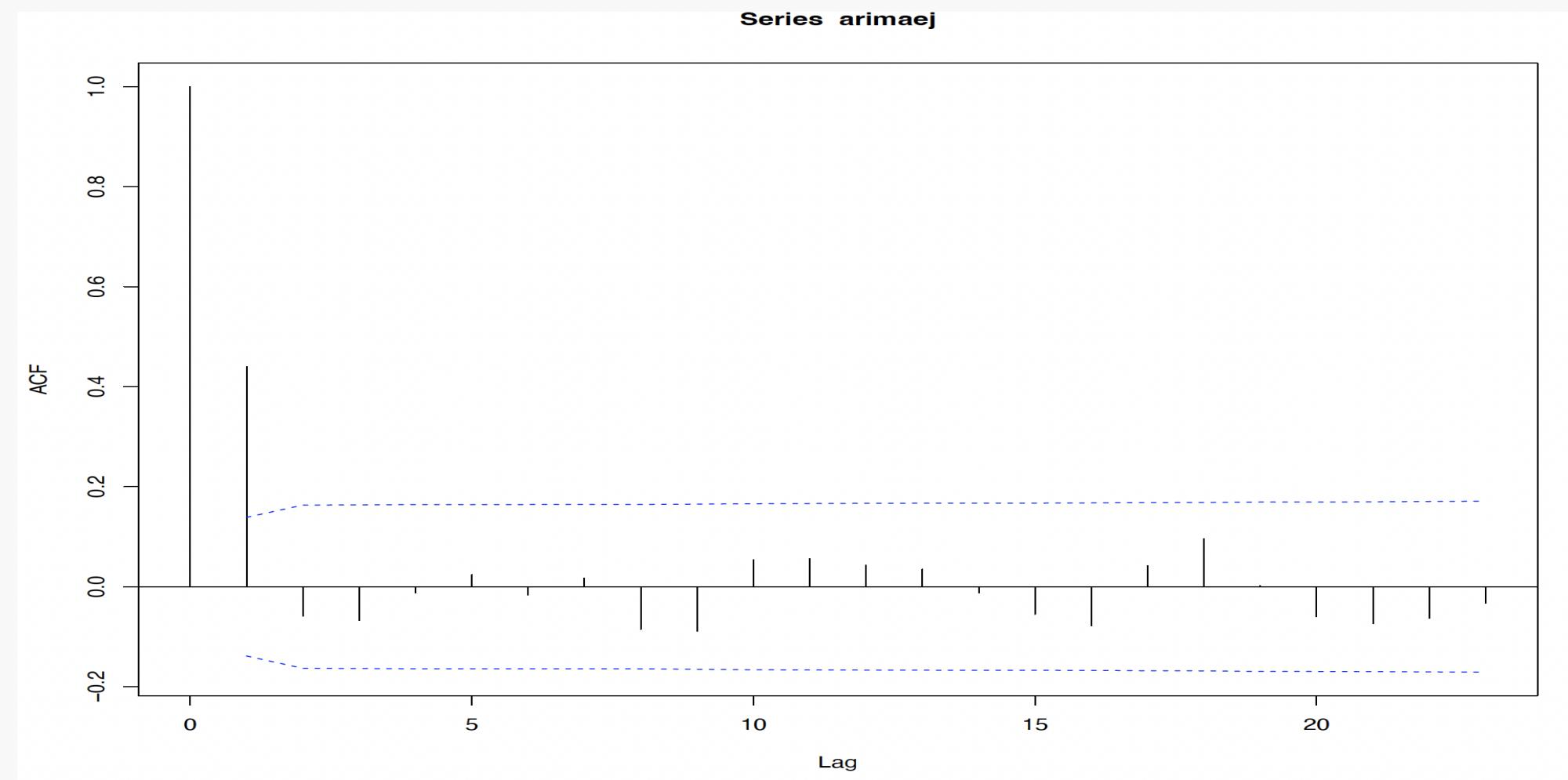
con lo cual se debería rechazar H_0 si $|z| > z_{1-\alpha/2}$. Como los valores de las correlaciones en el denominador son desconocidos para computar la varianza, entonces usamos sus estimaciones. Lo cual es equivalente a probar que

$$|\hat{\rho}(1)| > z_{1-\alpha/2} \frac{1}{n^{1/2}} \sqrt{1 - 3\hat{\rho}^2(1) + 4\hat{\rho}^4(1)}.$$

Lo que quiere decir, es que si $\hat{\rho}(1)$ está por fuera de las bandas de confianza $\pm z_{1-\alpha/2} \frac{1}{n^{1/2}} \sqrt{1 - 3\hat{\rho}^2(1) + 4\hat{\rho}^4(1)}$ entonces es significativamente diferente de cero. Con un razonamiento igual se puede verificar la significancia de $\rho(i)$ para $i \geq 2$ usando unas bandas $\pm z_{1-\alpha/2} \frac{1}{n^{1/2}} \sqrt{1 + 2\hat{\rho}^2(1)}$.

Ejemplo 2: Cont.

La siguiente gráfica nos muestra el ACF simple de bandas consistente con un proceso MA, para una serie simulada que proviene de un proceso MA(1). Note que la gráfica es consiste para identificar el proceso MA(1).



Estimación de $\gamma(\cdot)$ y $\rho(\cdot)$

Nota: si

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} \quad \{Z_t\} \sim IID(0, \sigma^2)$$

se puede verificar de la formula de Barlett que

$$w_{ii} = [1 + 2\rho^2(1) + 2\rho^2(2) + \cdots + 2\rho^2(q)] \quad i > q$$

por lo cual la varianza asintótica de $n^{-1/2}(\hat{\rho}(i) - 0)$ n es suf. grande. Lo anterior permite diseñar un procedimiento para identificar un proceso $MA(q)$ usando la función de autocorrelación simple.

Ejemplo 3:

Podemos aplicar la fórmula de Barlett (Diap. 33) a un proceso AR(1)

$$X_t = \phi X_{t-1} + Z_t \quad \{Z_t\} \sim \text{IID } (0, \sigma^2)$$

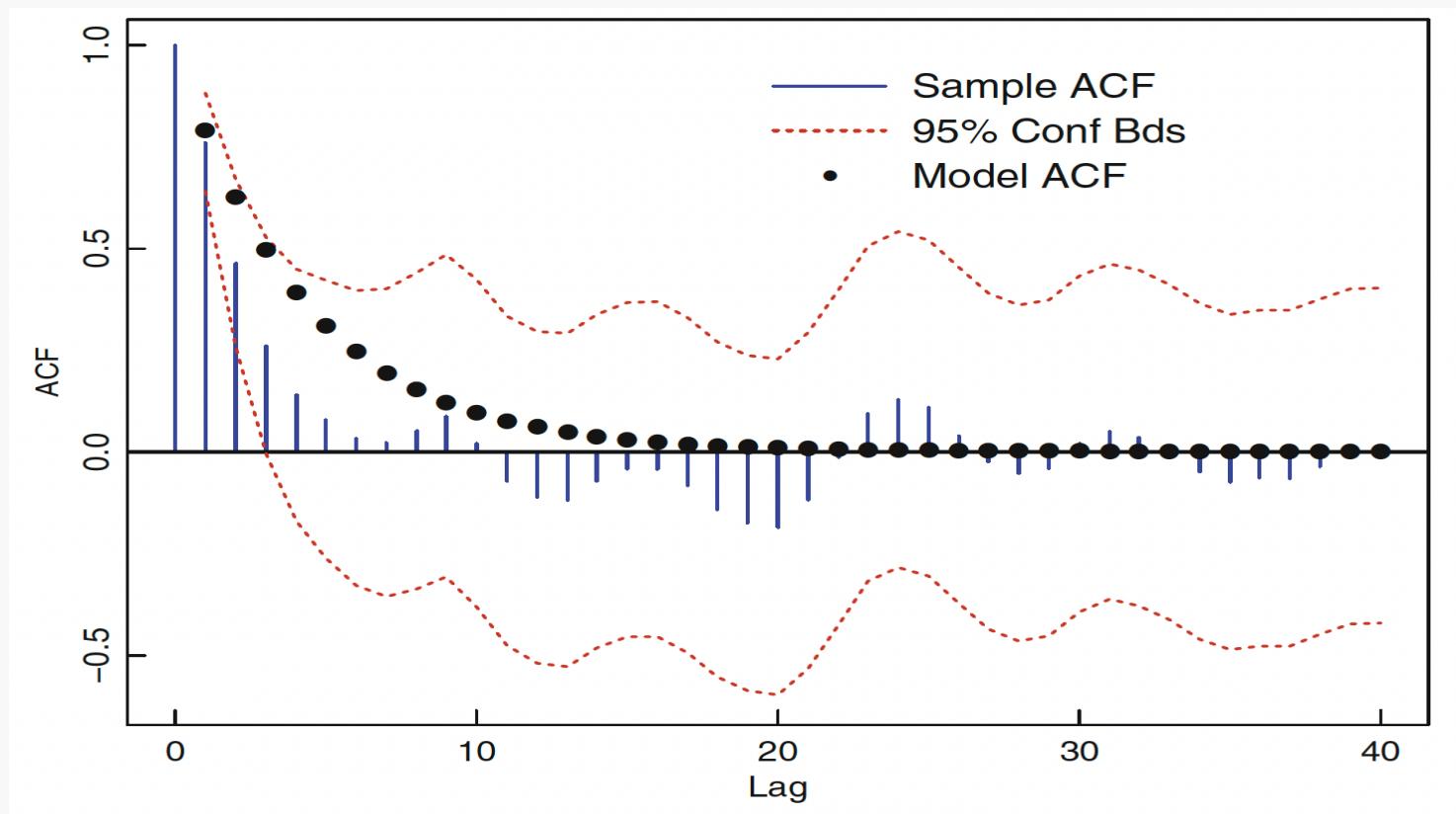
Cómo se verá en la siguiente lección $\rho(i) = \phi^{|i|}$, entonces la varianza asintótica de $n^{-1/2}(\hat{\rho}(i) - \rho(i))$ es

$$\begin{aligned} w_{ii} &= \sum_{k=1}^i \phi^{2i} (\phi^{-k} - \phi^k)^2 + \sum_{k=i+1}^{\infty} \phi^{2k} (\phi^i - \phi^i)^2 \\ &= (1 - 2\phi^{2i})(1 + \phi^2)(1 - \phi^2)^{-1} - 2i\phi^{2i} \\ &\simeq (1 + \phi^2)/(1 - \phi^2) \quad \text{para } i \text{ grande} \end{aligned}$$

No podemos sugerir una metodología igual que antes por que para ningún i : $\rho(i) = 0$, lo cual hace que no tenga sentido proponer bandas para probar $\rho(i) = 0$. Esto se hará usando la función de autocorrelación parcial FACP.

Ejemplo 3 (cont):

Suponga que simulamos el modelo $X_t = 0.791X_{t-1} + Z_t$, con $\phi = 0.791$. Los intervalos de confianza son calculados a partir de $\hat{\rho}(i) \pm 1.96n^{-1/2}w_{ii}^{1/2}$, con w_{ii} como en la diapositiva anterior (Diap. 41), el ACF teórico es $\rho(i) = (0.791)^i$. Así tenemos:



Observe que el modelo ACF apenas toca los límites de confianza en los rezagos 2 a 4. Esto sugiere cierta incompatibilidad de los datos con el modelo AR(1) simulado.

Función de Autocorrelación Parcial (PACF) (Shumway and Stoffer (2016))

- Hemos visto que, para un MA(1) el ACF será cero para rezagos mayores que 1.
- En general, para los modelos MA (q), la ACF será cero para los rezagos mayores que q . Además, debido a que $\theta_q \neq 0$, la ACF no será cero en el rezago q .

Función de Autocorrelación Parcial (PACF) (Shumway and Stoffer (2016))

- Hemos visto que, para un MA(1) el ACF será cero para rezagos mayores que 1.
- En general, para los modelos MA (q), la ACF será cero para los rezagos mayores que q . Además, debido a que $\theta_q \neq 0$, la ACF no será cero en el rezago q .
- Así, el ACF proporciona una cantidad considerable de información sobre el orden de la dependencia cuando el proceso es un proceso de promedio móvil. Sin embargo, si el proceso es ARMA o AR, la ACF por sí sola nos dice poco sobre los órdenes de dependencia.
- Por lo tanto, vale la pena buscar una función que se comporte como la ACF de los modelos MA, pero para los modelos AR, a saber, la función de autocorrelación parcial (Partial Autocorrelation Function PACF).

Función de Autocorrelación Parcial (PACF)

Recall that if X , Y , and Z are random variables, then the partial correlation between X and Y given Z is obtained by regressing X on Z to obtain \hat{X} , regressing Y on Z to obtain \hat{Y} , and then calculating

$$\rho_{XY|Z} = \text{corr}\{X - \hat{X}, Y - \hat{Y}\}.$$

The idea is that $\rho_{XY|Z}$ measures the correlation between X and Y with the linear effect of Z removed (or partialled out). If the variables are multivariate normal, then this definition coincides with $\rho_{XY|Z} = \text{corr}(X, Y | Z)$.

Función de Autocorrelación Parcial (PACF)

To motivate the idea for time series, consider a causal AR(1) model, $x_t = \phi x_{t-1} + w_t$. Then,

$$\begin{aligned}\gamma_x(2) &= \text{cov}(x_t, x_{t-2}) = \text{cov}(\phi x_{t-1} + w_t, x_{t-2}) \\ &= \text{cov}(\phi^2 x_{t-2} + \phi w_{t-1} + w_t, x_{t-2}) = \phi^2 \gamma_x(0)\end{aligned}$$

This result follows from causality because x_{t-2} involves $\{w_{t-2}, w_{t-3}, \dots\}$, which are all uncorrelated with w_t and w_{t-1} .

The correlation between x_t and x_{t-2} is not zero, as it would be for an MA(1), because x_t is **dependent** on x_{t-2} through x_{t-1} .

Función de Autocorrelación Parcial (PACF)

Suppose we break this chain of dependence by removing (or partial out) the effect x_{t-1} . That is, we consider the correlation between

$$x_t - \phi x_{t-1} \text{ and } x_{t-2} - \phi x_{t-1},$$

because it is the correlation between x_t and x_{t-2} with the linear dependence of each on x_{t-1} removed.

In this way, we have broken the dependence chain between x_t and x_{t-2} . In fact,

$$\text{cov}(x_t - \phi x_{t-1}, x_{t-2} - \phi x_{t-1}) = \text{cov}(w_t, x_{t-2} - \phi x_{t-1}) = 0.$$

Hence, the tool we need is partial autocorrelation, which is the correlation between x_s and x_t with the linear effect of everything "*in the middle*" removed.

Función de Autocorrelación Parcial (PACF)

To formally define the PACF for mean-zero stationary time series, let \hat{x}_{t+h} , for $h \geq 2$, denote the regression^{3.3} of x_{t+h} on $\{x_{t+h-1}, x_{t+h-2}, \dots, x_{t+1}\}$, which we write as

$$\hat{x}_{t+h} = \beta_1 x_{t+h-1} + \beta_2 x_{t+h-2} + \cdots + \beta_{h-1} x_{t+1}. \quad (3.53)$$

No intercept term is needed in (3.53) because the mean of x_t is zero (otherwise, replace x_t by $x_t - \mu_x$ in this discussion). In addition, let \hat{x}_t denote the regression of x_t on $\{x_{t+1}, x_{t+2}, \dots, x_{t+h-1}\}$, then

$$\hat{x}_t = \beta_1 x_{t+1} + \beta_2 x_{t+2} + \cdots + \beta_{h-1} x_{t+h-1}. \quad (3.54)$$

Because of stationarity, the coefficients, $\beta_1, \dots, \beta_{h-1}$ are the same in (3.53) and (3.54); we will explain this result in the next section, but it will be evident from the examples.

^{3.3} The term regression here refers to regression in the population sense. That is, \hat{x}_{t+h} is the linear combination of $\{x_{t+h-1}, x_{t+h-2}, \dots, x_{t+1}\}$ that minimizes the mean squared error $E(x_{t+h} - \sum_{j=1}^{h-1} \alpha_j x_{t+j})^2$.

Función de Autocorrelación Parcial (PACF)

Definition 3.9 *The partial autocorrelation function (PACF) of a stationary process, x_t , denoted ϕ_{hh} , for $h = 1, 2, \dots$, is*

$$\phi_{11} = \text{corr}(x_{t+1}, x_t) = \rho(1) \quad (3.55)$$

and

$$\phi_{hh} = \text{corr}(x_{t+h} - \hat{x}_{t+h}, x_t - \hat{x}_t), \quad h \geq 2. \quad (3.56)$$

The PACF, ϕ_{hh} , is the correlation between x_{t+h} and x_t with the linear dependence of $\{x_{t+1}, \dots, x_{t+h-1}\}$ on each, removed.

If the process x_t is Gaussian, then $\phi_{hh} = \text{corr}(x_{t+h}, x_t \mid x_{t+1}, \dots, x_{t+h-1})$; that is, ϕ_{hh} is the correlation coefficient between x_{t+h} and x_t in the bivariate distribution of (x_{t+h}, x_t) conditional on $\{x_{t+1}, \dots, x_{t+h-1}\}$.

Example 1: The PACF of an AR(1)

Consider the PACF of the AR(1) process given by $x_t = \phi x_{t-1} + w_t$, with $|\phi| < 1$. By definition, $\phi_{11} = \rho(1) = \phi$. To calculate ϕ_{22} , consider the regression of x_{t+2} on x_{t+1} , say, $\hat{x}_{t+2} = \beta x_{t+1}$. We choose β to minimize

$$E(x_{t+2} - \hat{x}_{t+2})^2 = E(x_{t+2} - \beta x_{t+1})^2 = \gamma(0) - 2\beta\gamma(1) + \beta^2\gamma(0).$$

Taking derivatives with respect to β and setting the result equal to zero, we have $\beta = \gamma(1)/\gamma(0) = \rho(1) = \phi$. Next, consider the regression of x_t on x_{t+1} , say $\hat{x}_t = \beta x_{t+1}$. We choose β to minimize

$$E(x_t - \hat{x}_t)^2 = E(x_t - \beta x_{t+1})^2 = \gamma(0) - 2\beta\gamma(1) + \beta^2\gamma(0).$$

This is the same equation as before, so $\beta = \phi$. Hence,

$$\begin{aligned}\phi_{22} &= \text{corr}(x_{t+2} - \hat{x}_{t+2}, x_t - \hat{x}_t) = \text{corr}(x_{t+2} - \phi x_{t+1}, x_t - \phi x_{t+1}) \\ &= \text{corr}(w_{t+2}, x_t - \phi x_{t+1}) = 0\end{aligned}$$

by causality. Thus, $\phi_{22} = 0$. In the next example, we will see that in this case, $\phi_{hh} = 0$ for all $h > 1$.

Example 2: The PACF of an AR(p)

The model implies $x_{t+h} = \sum_{j=1}^p \phi_j x_{t+h-j} + w_{t+h}$, where the roots of $\phi(z)$ are outside the unit circle. When $h > p$, the regression of x_{t+h} on $\{x_{t+1}, \dots, x_{t+h-1}\}$, is

$$\hat{x}_{t+h} = \sum_{j=1}^p \phi_j x_{t+h-j}.$$

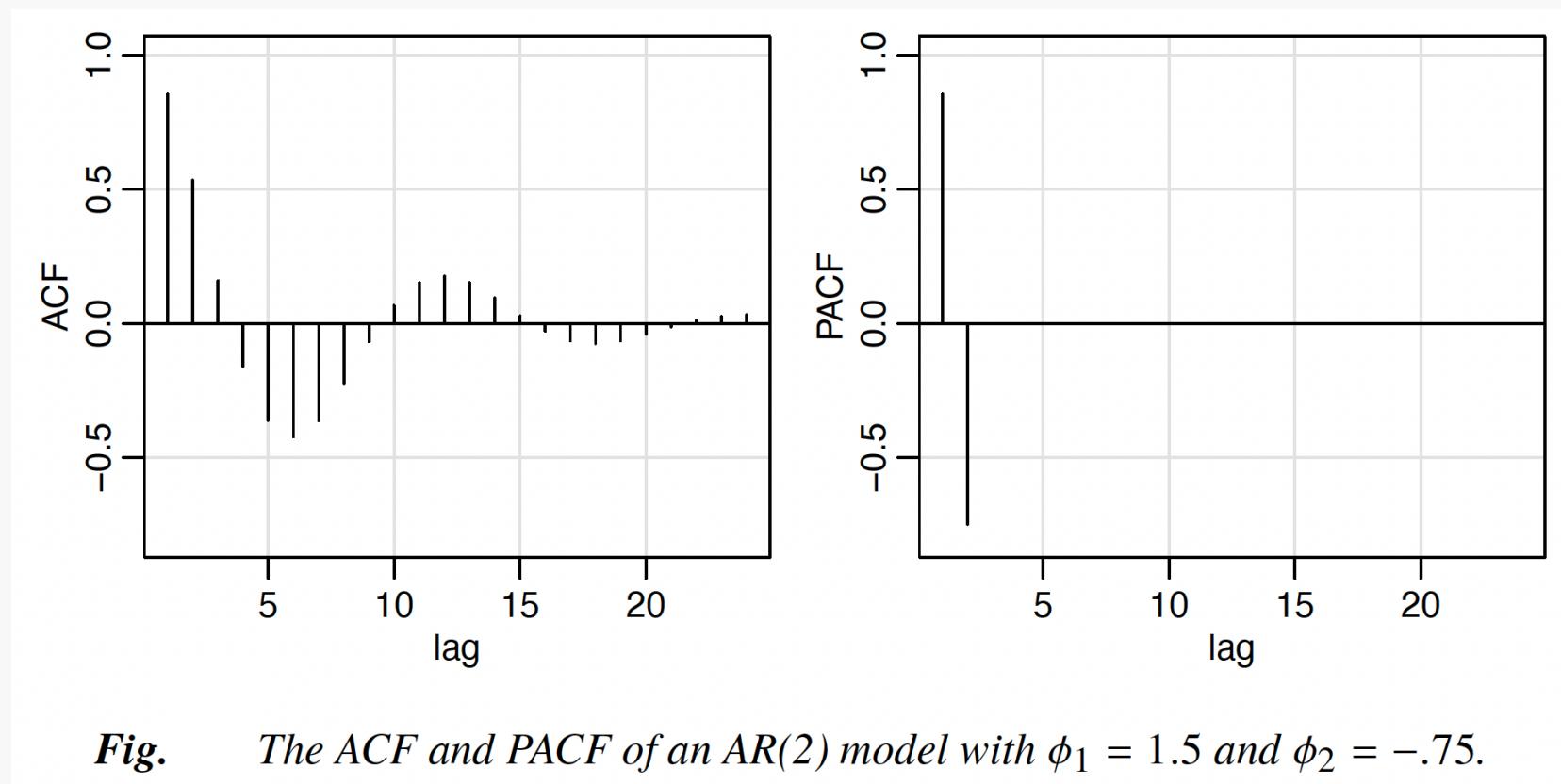
Thus, when $h > p$,

$$\phi_{hh} = \text{corr}(x_{t+h} - \hat{x}_{t+h}, x_t - \hat{x}_t) = \text{corr}(w_{t+h}, x_t - \hat{x}_t) = 0,$$

because, by causality, $x_t - \hat{x}_t$ depends only on $\{w_{t+h-1}, w_{t+h-2}, \dots\}$; recall equation (3.54). When $h \leq p$, ϕ_{pp} is not zero, and $\phi_{11}, \dots, \phi_{p-1,p-1}$ are not necessarily zero. We will see later that, in fact, $\phi_{pp} = \phi_p$.

Example 3: The PACF of an AR(2)

The following Figure shows the ACF and the PACF of the AR(2) model with $\phi_1 = 1.5$ and $\phi_2 = -0.75$.



Example 4: The PACF of an MA(1)

In general, for an invertible (we see this later) MA(q), the PACF will never cut off, as in the case of an AR(p).

For an MA(1), $x_t = w_t + \theta w_{t-1}$, with $|\theta| < 1$, calculations similar to Example 1 (PACF of the AR(1) model) will yield

$$\phi_{22} = -\theta^2 / (1 + \theta^2 + \theta^4)$$

For the MA(1) in general, we can show that

$$\phi_{hh} = -\frac{(-\theta)^h (1 - \theta^2)}{1 - \theta^{2(h+1)}}, \quad h \geq 1.$$

Example 4: The PACF of an MA(1)

In general, for an invertible (we see this later) MA(q), the PACF will never cut off, as in the case of an AR(p).

For an MA(1), $x_t = w_t + \theta w_{t-1}$, with $|\theta| < 1$, calculations similar to Example 1 (PACF of the AR(1) model) will yield

$$\phi_{22} = -\theta^2 / (1 + \theta^2 + \theta^4)$$

For the MA(1) in general, we can show that

$$\phi_{hh} = -\frac{(-\theta)^h (1 - \theta^2)}{1 - \theta^{2(h+1)}}, \quad h \geq 1.$$

- The PACF for MA models behaves much like the ACF for AR models.
- Also, the PACF for AR models behaves much like the ACF for MA models.

PACF

In general, we have:

Table 3.1. Behavior of the ACF and PACF for ARMA Models

	AR(p)	MA(q)	ARMA(p, q)
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

See R code: "P2 - ACF y PACF para AR y MA.R"

The Sample PACF of an AR(p) Series

- If $\{X_t\}$ is an AR(p) series, then the sample PACF based on observations $\{x_1, \dots, x_n\}$ should reflect (with sampling variation) the properties of the PACF itself.
- In particular, if the sample PACF $\hat{\phi}_{hh}$ is significantly different from zero for $0 \leq h \leq p$ and negligible for $h > p$, an AR(p) model might provide a good representation of the data.

The Sample PACF of an AR(p) Series

- If $\{X_t\}$ is an AR(p) series, then the sample PACF based on observations $\{x_1, \dots, x_n\}$ should reflect (with sampling variation) the properties of the PACF itself.
- In particular, if the sample PACF $\hat{\phi}_{hh}$ is significantly different from zero for $0 \leq h \leq p$ and negligible for $h > p$, an AR(p) model might provide a good representation of the data.
- To decide what is meant by "negligible" we can use the result that for an AR(p) process the sample PACF values at lags greater than p are approximately independent $N(0, 1/n)$ random variables. This means that roughly 95% of the sample PACF values beyond lag p should fall within the bounds $\pm 1.96/\sqrt{n}$.
- If we observe a sample PACF satisfying $|\hat{\phi}_{hh}| > 1.96/\sqrt{n}$ for $0 \leq h \leq p$ and $|\hat{\phi}_{hh}| < 1.96/\sqrt{n}$ for $h > p$, this suggests an AR(p) model for the data.

Lecturas sugeridas:

- Brockwell and Davis (2016): Capítulo 2, sección 2.2 (hasta pág. 46).
- Shumway and Stoffer (2016): PACF páginas 98-101.
- Sergio Calderon (2023): Capítulo 5 y Sección 4.8.

Aviso:

Examen 1: Miercoles 21 de Mayo de 2025.
Más adelante daré detalles más específicos.