

UNIVERSIDAD NACIONAL DE CÓRDOBA

FACULTAD DE MATEMÁTICA, ASTRONOMÍA, FÍSICA Y COMPUTACIÓN



# MEDIDAS DE DISIMILITUD EN SERIES TEMPORALES

AUTORA: LAURA A. RODRÍGUEZ ASTRAIN

DIRECTORA: DRA. SILVIA MARÍA OJEDA

MARZO DE 2017



MEDIDAS DE DISIMILITUD EN SERIES TEMPORALES POR LAURA A. RODRÍGUEZ ASTRAIN SE DISTRIBUYE BAJO UNA LICENCIA CREATIVE COMMONS ATRIBUCIÓN-NOComercial-SINDerivadas 2.5.

[HTTP://creativecommons.org/licenses/by-nc-nd/2.5/ar/](http://creativecommons.org/licenses/by-nc-nd/2.5/ar/)

---

# Agradecimientos

En primer lugar, es mi deseo agradecerle a mi familia por su apoyo incondicional; en especial, a mi mamá que ha hecho siempre hasta lo imposible por intentar facilitarme el camino para yo poder dedicarme exclusivamente a mi carrera. En segundo lugar, y no por eso menos importante, le agradezco profundamente a mi novio por entender mis locas y largas horas de estudio y acompañarme siempre en los buenos y malos momentos. También quisiera agradecerle a mi directora por dedicarme su tiempo, ser tan paciente conmigo e interesarse siempre por mis dudas e inquietudes. Por último, es mi deseo agradecerles a mis compañeros con los que he compartido tantos momentos; son en verdad muy importantes para mi. Nuevamente, muchas gracias a todos.

---

# Resumen

El análisis de series de tiempo surge de la necesidad de estudiar el comportamiento de observaciones ordenadas que se obtienen en períodos regulares de tiempo y que, en general, son dependientes entre sí.

El interés por un método que permita identificar el grado de disimilitud entre dos o mas series, surge en muchas áreas del conocimiento o disciplinas tales como biología, criminalística, climatología, entre otras; en donde se suscitan diferentes problemáticas en las que resulta indispensable comparar dichas series.

Podemos distinguir al menos dos enfoques centrales para evaluar la proximidad entre series de tiempo; por un lado, un enfoque paramétrico que consiste en la proyección de series en un determinado espacio de forma funcional; y por otro lado, un enfoque no paramétrico que consiste en evaluar la proximidad entre series de tiempo en base a los datos propios de las mismas. Hay autores que han desarrollado medidas de comparación con un enfoque no paramétrico y es el estudio de estas medidas nuestro principal objetivo.

**Palabras claves:** Series de tiempo, comparación, disimilitud, proceso estacionario, índice.

**Clasificación:**(Mathematics Subject Classification)  
60G10,62M10,62P10.

---

# Abstract

The analysis of time series arises from the need to study the behavior of ordered observations that are obtained in regular periods of time and that, in general, are dependent on each other. The interest in a method that identifies the degree of dissimilarity between two or more time series, arises in many areas of knowledge or disciplines such as biology, criminality, climatology, among others; In which different problems arise in which it is essential to compare time series. We can distinguish at least two central approaches to assess the proximity between time series; On the one hand, a parametric approach that consists in the projection of time series in a given functional space; And on the other hand, a non-parametric approach that consists in evaluating the proximity between time series based on their own data. There are authors who have developed measures of comparison with a non-parametric approach and it is the study of these measures our main objective.

---

# Índice general

<b>Agradecimientos</b>	<b>2</b>
<b>Resumen</b>	<b>3</b>
<b>1. Series de tiempo</b>	<b>7</b>
1.1. Introducción . . . . .	7
1.2. Conceptos básicos . . . . .	8
1.3. Series de tiempo estacionarias . . . . .	9
1.4. Estimación de las características de una serie de tiempo estacionaria.	20
1.4.1. Estimación de la media $\mu$ . . . . .	20
1.4.2. Estimación de las autocovarianzas. . . . .	22
1.4.3. Estimación de las autocorrelaciones. . . . .	24
1.4.4. Estimación de las autocorrelaciones parciales. . . . .	25
<b>2. Modelos de series de tiempo estacionarias</b>	<b>26</b>
2.1. Representaciones de una serie de tiempo. . . . .	26
2.1.1. Representación de una serie de tiempo como un proceso de promedios móviles de orden infinito. . . . .	26
2.1.2. Representación de una serie de tiempo como un proceso autorregresivo de orden infinito . . . . .	29
2.2. Modelos autorregresivos de orden finito. . . . .	29
2.2.1. Modelo autorregresivo de orden 1 . . . . .	30
2.3. Modelos de promedios móviles de orden finito. . . . .	34
2.3.1. Modelo de promedios móviles de orden 1. . . . .	36
2.4. Modelos Autorregresivos de Promedios Móviles. . . . .	42

---

2.4.1. Modelo ARMA(1,1) . . . . .	44
<b>3. Comparación de series de tiempo</b>	<b>53</b>
3.1. Distancia euclídea como medida de proximidad entre series temporales	54
3.2. Correlación temporal para medir proximidad con respecto a comportamiento . . . . .	55
3.3. Índice de disimilitud para series de tiempo . . . . .	59
3.4. Aplicación y comportamiento del índice de disimilitud . . . . .	61
<b>Conclusión y trabajo futuro</b>	<b>67</b>
<b>Anexo</b>	<b>68</b>
<b>Bibliografía</b>	<b>69</b>

---

# Capítulo 1

## Series de tiempo

### 1.1. Introducción

El análisis de series de tiempo surge de la necesidad de estudiar el comportamiento de observaciones ordenadas que se obtienen en períodos regulares de tiempo y que, en general, son dependientes entre sí, a diferencia de datos que provienen de una muestra aleatoria, es decir, de variables aleatorias independientes idénticamente distribuidas. Como dichos datos aparecen tan frecuentemente en la realidad, es de allí la importancia de encontrar métodos de estudio para ellos. Algunos campos en donde se obtienen este tipo de observaciones son, por ejemplo, economía, en el comportamiento de índices de precios; agronomía, en el registro de la producción anual de ciertos productos agrícolas; y otros campos, como hidrología, ciencias sociales, electrónica, etc.

Entre los objetivos principales de los métodos que se desarrollan para analizar series de tiempo se encuentran:

1. generar modelos probabilísticos que permitan entender y describir los procesos aleatorios que generan las observaciones;
2. la predicción de valores futuros;
3. el control óptimo de esta clase de procesos.

## 1.2. Conceptos básicos

**Definición 1.2.1:** Sea  $(\Omega, A, P)$  un espacio de probabilidades y sea  $T$  un conjunto no vacío de índices. Se llama **serie de tiempo** o **proceso estocástico** con valores reales a una función  $X: T \times \Omega \rightarrow \mathbb{R}$  tal que para cada valor fijo de  $t$  es  $X(t, \cdot): \Omega \rightarrow \mathbb{R}$  una variable aleatoria a la que usualmente se la denota como  $X_t$  o  $X(t)$ , según el conjunto de índices  $T$  sea numerable o no, respectivamente. En el caso en que el conjunto de índices sea numerable, se dirá que se trata de una serie de tiempo discreto; en caso contrario, se dirá que se trata de una serie de tiempo continuo. La notación usualmente empleada para las series de tiempo discreto es  $\{X_t: t \in T\}$  (en general,  $T=\mathbb{Z}$  o  $T=\mathbb{N}$ ), mientras que para las series de tiempo continuo se usa la notación  $\{X(t): t \in T\}$ , donde generalmente  $T=\mathbb{R}$ .

**Definición 1.2.2:** Para cada valor  $\omega \in \Omega$ , la función  $X(\cdot, \omega): T \rightarrow \mathbb{R}$ , que usualmente es indicada por  $\{x_t: t \in T\}$  o bien  $\{x(t): t \in T\}$ , constituye un valor de cada una de las variables aleatorias  $X_t$  o  $X(t)$ , y se llama **función muestral** correspondiente a  $\omega$  o **realización** de la serie de tiempo.

En lo que sigue nos referiremos exclusivamente a series de tiempo discreto donde  $T=\mathbb{N}$  o  $T=\mathbb{Z}$ . Para conocer completamente a una serie de tiempo  $\{X_t: t \in T\}$ , es necesario conocer la función de distribución conjunta de cada uno de los subconjuntos finitos posibles de variables aleatorias  $X_t$  tal que  $t \in T$ . Es decir, si  $n$  es cualquier número natural y si  $t_1, t_2, \dots, t_n$  son elementos cualesquiera de  $T$ , debemos conocer totalmente a  $F_{X_{t_1}, \dots, X_{t_n}} = P[X_{t_1} \leq x_{t_1}, X_{t_2} \leq x_{t_2}, \dots, X_{t_n} \leq x_{t_n}]$ , para todo  $x_{t_1}, x_{t_2}, \dots, x_{t_n} \in \mathbb{R}$ . En el caso de variables aleatorias independientes e idénticamente distribuidas basta con conocer la distribución  $F_X$  común a cada una de las variables. Sin embargo, en el caso general de series de tiempo, por la dependencia que existe entre las variables aleatorias  $X_t$  tal que  $t \in T$ , se requiere, para poder hacer inferencias estadísticas, de suposiciones adicionales respecto al mecanismo aleatorio que ha generado los datos observados, pues es importante notar que en la práctica solo se cuenta en general con un solo dato de cada variable aleatoria de un conjunto finito del total de las variables aleatorias que constituyen la serie de tiempo.

**Definición 1.2.3:** Una serie de tiempo  $\{X_t: t \in T\}$  se dice **estrictamente estacionaria** o **completamente estacionaria** si  $F_{X_{t_1}, \dots, X_{t_n}} = F_{X_{t_1+k}, \dots, X_{t_n+k}}$ , para todo  $k \in \mathbb{N}$  y  $n \in \mathbb{N}$ , y para todo  $t_1, t_2, \dots, t_n, t_1+k, \dots, t_n+k$  en el conjunto de índices



T.

**Proposición 1.2.1:** Si  $\{X_t : t \in T\}$  es una serie de tiempo estrictamente estacionaria tal que  $E[X_t^2] < \infty$  para todo  $t \in T$ , entonces todas las variables aleatorias  $X_t$  tal que  $t \in T$  tienen la misma media y la misma varianza. Mas aún, la covarianza entre un par de variables  $X_t$  y  $X_s$  será la misma que la covarianza entre cualquier otro par  $X_p$  y  $X_q$  tal que  $p$  y  $q$  satisfacen  $|p - q| = |t - s|$ .

**Demostración:** Todas las variables aleatorias  $X_t$  tal que  $t \in T$  tienen la misma media y la misma varianza pues  $F_{X_t} = F_{X_s}$  para todo  $t, s \in T$ . Por otro lado, si suponemos que  $p = t+k$  y  $q = s+k$  con  $k \in \mathbb{N}$ , luego  $F_{X_p, X_q} = F_{X_{t+k}, X_{s+k}} = F_{X_t, X_s}$ , de modo que,  $E[X_p X_q] = E[X_t X_s]$  y, por lo tanto,  $cov(X_p, X_q) = E[X_p X_q] - E[X_p]E[X_q] = E[X_t X_s] - E[X_t]E[X_s] = cov(X_t, X_s)$ , es decir,  $cov(X_p, X_q) = cov(X_t, X_s)$ . ■

Este concepto de estacionariedad estricta, si bien relaja la exigencia de tener variables aleatorias independientes e idénticamente distribuidas, es en la práctica muy difícil de verificar. Sin embargo, como muchas de las propiedades lineales que interesan conocer en una serie de tiempo, pueden ser descritas en función de sus momentos de primero y segundo orden, no es necesario conocer sus funciones de distribuciones conjuntas y, por lo tanto, el concepto de estacionariedad puede ser relajado exigiendo únicamente ciertas condiciones en sus momentos de primero y segundo orden. De esto hablaremos precisamente en la siguiente sección.

### 1.3. Series de tiempo estacionarias

**Definición 1.3.1:** Se dice que una serie de tiempo  $\{X_t : t \in T\}$  es **estacionaria** (o también **estacionaria en sentido amplio** o **estacionaria en segundo orden**) si se verifica que:

1.  $E[X_t] = \mu < \infty$  para todo  $t \in T$
2.  $cov(X_t, X_{t+h}) = E[(X_t - \mu)(X_{t+h} - \mu)] = \gamma_h < \infty$  para todo  $t, t+h \in T$

Notar que la propiedad 2 de la definición anterior no hace mas que decir que la covarianza entre un par de variables  $X_t$  y  $X_s$  será la misma que la covarianza entre cualquier otro par  $X_p$  y  $X_q$  tal que  $p$  y  $q$  satisfacen  $|p - q| = |t - s|$ , lo mismo que ocurría en una serie de tiempo estrictamente estacionaria. En particular, cuando  $h = 0$ , esta propiedad afirma que las variables aleatorias que componen la serie de tiempo estacionaria tienen todas la misma varianza, digamos  $\gamma_0 = \sigma^2$ .

Mas aún, con lo ya dicho podemos afirmar que toda serie de tiempo estrictamente estacionaria con momentos de segundo orden finitos es entonces una serie de tiempo estacionaria, siendo la recíproca no verdadera en general, salvo que se asuma que el proceso estocástico es Gaussiano, es decir, que todas sus distribuciones conjuntas son normales multivariadas, pues en este caso estas distribuciones quedan determinadas totalmente por su media y su matriz de varianza-covarianza.

**Definición 1.3.2:** Dada una serie de tiempo estacionaria  $\{X_t : t \in T\}$  se llama **función de autocovarianza teórica** a la función  $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$  que a cada entero  $k$  le asigna el valor  $\gamma_k = cov(X_t, X_{t+k}) = E[(X_t - \mu)(X_{t+k} - \mu)]$ .

**Definición 1.3.3:** Dada una serie de tiempo estacionaria  $\{X_t : t \in T\}$  se llama **función de autocorrelación teórica** a la función  $\rho : \mathbb{Z} \rightarrow \mathbb{R}$  que a cada entero  $k$  le hace corresponder  $\rho_k = corr(X_t, X_{t+k}) = \frac{cov(X_t, X_{t+k})}{\sqrt{var(X_t)var(X_{t+k})}} = \frac{\gamma_k}{\gamma_0}$ .

**Proposición 1.3.1:** Dada una serie de tiempo estacionaria  $\{X_t : t \in T\}$ , las funciones de autocovarianza y autocorrelación teóricas satisfacen las siguientes propiedades:

1. Las funciones de autocovarianza y de autocorrelación teóricas son semidefinidas positivas, es decir, para todo  $n \in \mathbb{N}$  y para todo  $t_1, t_2, \dots, t_n \in T$  se verifica que:  $\sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma_{|t_i - t_j|} \geq 0$  y  $\sum_{i=1}^n \sum_{j=1}^n a_i a_j \rho_{|t_i - t_j|} \geq 0$  para cualesquiera  $a_1, \dots, a_n \in \mathbb{R}$ .
2.  $|\gamma_k| \leq \gamma_0 = var(X_t)$  y  $|\rho_k| \leq \rho_0 = 1 \forall k \in \mathbb{Z}$ .
3.  $\gamma_k = \gamma_{-k}$  y  $\rho_k = \rho_{-k}$ .

**Demostración:**

1. Consideremos la variable  $Z = \sum_{i=1}^n a_i X_{t_i}$  y calculemos su varianza la cual es no negativa, es decir,  $0 \leq var(Z) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j cov(X_{t_i}, X_{t_j}) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma_{|t_i - t_j|}$  como se quería probar. Para probar la segunda desigualdad basta con dividir miembro a miembro de la desigualdad anterior por  $\gamma_0$ , pues  $\rho_{|t_i - t_j|} = \frac{\gamma_{|t_i - t_j|}}{\gamma_0}$  para todo  $i, j \in \{1, \dots, n\}$ .
2. Si consideramos la primera desigualdad dada en la propiedad anterior para el caso particular  $n = 2$  y  $|t_1 - t_2| = k$  se tiene la relación  $0 \leq a_1^2 \gamma_0 + a_2^2 \gamma_0 +$

$2a_1a_2\gamma_k$  válida para todo  $a_1, a_2 \in \mathbb{R}$ , de donde si elegimos en primer lugar  $a_1 = 1$  y  $a_2 = -1$  resulta que  $2\gamma_0 - 2\gamma_k \geq 0$ , y por lo tanto,  $\gamma_0 \geq \gamma_k$ . Si en cambio elegimos  $a_1 = a_2 = 1$ , resulta que  $\gamma_0 \geq -\gamma_k$ , y así podemos concluir de estas dos últimas desigualdades que  $\gamma_0 \geq |\gamma_k|$ . Dividiendo esta última desigualdad por  $\gamma_0 > 0$  resulta que  $\rho_0 = 1 \geq |\rho_k|$ .

3. Para probar que  $\gamma_k = \gamma_{-k}$ , dado  $t \in T$ , definimos  $s = t - k$ , luego  $\gamma_{-k} = \text{cov}(X_t, X_{t-k}) = \text{cov}(X_{t-k}, X_t) = \text{cov}(X_s, X_{s+k}) = \gamma_k$ . Para probar que  $\rho_k = \rho_{-k}$ , basta dividir miembro a miembro de la igualdad recién probada por  $\gamma_0$ . ■

**Definición 1.3.4:** Dada una serie de tiempo estacionaria  $\{X_t : t \in T\}$  se llama **correlograma teórico** de la serie de tiempo, al gráfico de la función de autocorrelación teórica en un sistema de coordenadas cartesianas.

Como la función de autocorrelación teórica es una función par, pues  $\rho_k = \rho_{-k}$  para todo  $k \in \mathbb{Z}$ , el correlograma teórico se realiza únicamente para valores de  $k \in \mathbb{N}$ , pues, además, para  $k = 0$  siempre es cierto que  $\rho_0 = 1$ .

**Definición 1.3.5:** Sea  $\{A_t : t \in T\}$  una serie de tiempo tal que las variables aleatorias  $A_t$  son no correlacionadas entre sí, es decir,  $\text{cov}(A_t, A_s) = 0$  para todo  $t, s \in T$ ; con media común  $\mu_A = E[A_t] = 0$  para todo  $t \in T$  y con varianza común  $\sigma_A^2 = \text{var}(A_t)$  para todo  $t \in T$ . Una serie de tiempo con estas características es llamada **proceso de ruido blanco**.

**Proposición 1.3.2:** Si  $\{A_t : t \in T\}$  es un proceso de ruido blanco, entonces dicha serie de tiempo es estacionaria.

**Demostración:**  $\{A_t : t \in T\}$  es una serie de tiempo estacionaria, pues cumple con las siguientes propiedades:

1.  $\mu_A = E[A_t] = 0 < \infty$  para todo  $t \in T$
2.  $\text{cov}(A_t, A_s)$  es “invariante en el tiempo”, en efecto:

$$\text{cov}(A_t, A_s) = \begin{cases} \sigma_A^2 & \text{si } t = s \\ 0 & \text{si } t \neq s \end{cases} = \gamma_{|t-s|}.$$

Así se verifica que todo proceso de ruido blanco es una serie de tiempo estacionaria. ■

Notemos que si  $\{A_t : t \in T\}$  es un proceso de ruido blanco entonces su función de autocovarianza teórica esta definida por  $\gamma_k = \begin{cases} \sigma_A^2 & \text{si } k = 0 \\ 0 & \text{si } k \neq 0 \end{cases}$  y su función de autocorrelación teórica esta definida por  $\rho_k = \begin{cases} 1 & \text{si } k = 0 \\ 0 & \text{si } k \neq 0 \end{cases}$  (ver Figura 1.1).

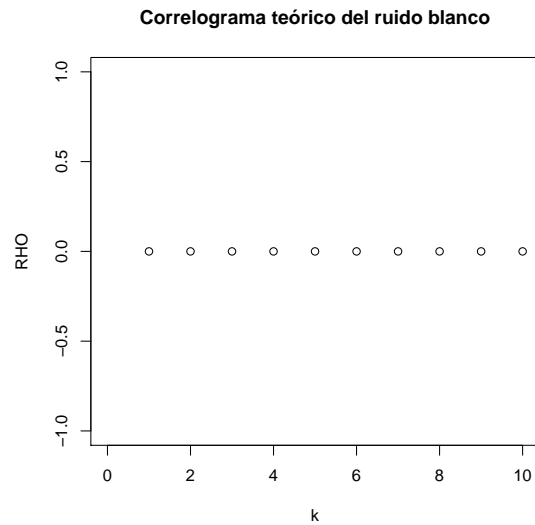


Figura 1.1: Correlograma teórico del proceso de ruido blanco

Si bien el proceso de ruido blanco rara vez se presenta en la práctica, este proceso juega un papel fundamental en la construcción de modelos de series de tiempo.

Siguiendo con el análisis de la función de autocorrelación teórica, es importante resaltar que cuando para un cierto  $k \in \mathbb{Z}$  sucede que  $\rho_k \neq 0$ , esto nos muestra que hay una cierta relación o dependencia lineal entre  $X_t$  y  $X_{t+k}$ , pues si  $\rho_k \neq 0$ , luego  $cov(X_t, X_{t+k}) \neq 0$ . Esto significa que por ejemplo, si  $\rho_2 \neq 0$ , el valor obtenido por la serie en el tiempo  $t$  “influye” en el valor obtenido por la serie en el tiempo  $t+2$ , para todo  $t \in T$ . Sería interesante entonces poder evaluar para el caso general de una serie de tiempo estacionaria  $\{X_t : t \in T\}$ , cual es la dependencia lineal entre  $X_t$  y  $X_{t+k}$ , para cada  $k \in \mathbb{Z}$  tal que  $\rho_k \neq 0$ . Para evaluar por ejemplo la dependencia lineal entre  $X_t$  y  $X_{t+2}$ , existe como medida el coeficiente de correlación parcial teórica  $\pi_2 = corr(X_t, X_{t+2}/X_{t+1})$ , que se obtiene realizando los siguientes pasos:

1. Se le “extrae” a  $X_t$  la dependencia lineal que tiene con  $X_{t+1}$ : para

esto se define una nueva variable  $X_t^*$  que se obtiene de  $X_t$  restándole la “influencia” lineal de  $X_{t+1}$ , es decir, se define  $X_t^* = (X_t - \alpha X_{t+1})$  donde  $\alpha$  se elige de tal forma que  $var(X_t^*)$  se haga mínima. Pero  $var(X_t^*) = var(X_t - \alpha X_{t+1}) = E[((X_t - \mu) - \alpha(X_{t+1} - \mu))^2] = E[(X_t - \mu)^2] + \alpha^2 E[(X_{t+1} - \mu)^2] - 2\alpha E[(X_t - \mu)(X_{t+1} - \mu)] = \gamma_0 + \alpha^2 \gamma_0 - 2\alpha \gamma_1$ . Luego minimizar  $var(X_t^*)$  es equivalente a minimizar la función cuadrática de  $\alpha$ ,  $\gamma_0 + \alpha^2 \gamma_0 - 2\alpha \gamma_1$ , con coeficiente principal  $\gamma_0 > 0$ , para lo cual es suficiente con hallar la raíz de su derivada primera respecto de  $\alpha$ . Como  $\frac{d}{d\alpha} var(X_t^*) = 2\gamma_0 \alpha - 2\gamma_1$ , luego  $\alpha = \frac{\gamma_1}{\gamma_0} = \rho_1$  y, en consecuencia, la  $var(X_t^*) = \gamma_0 + \frac{\gamma_1^2}{\gamma_0} \gamma_0 - 2\frac{\gamma_1}{\gamma_0} \gamma_1 = \gamma_0 + \frac{\gamma_1^2}{\gamma_0} - 2\frac{\gamma_1^2}{\gamma_0} = \gamma_0 - \frac{\gamma_1 \gamma_1}{\gamma_0} = \gamma_0 - \rho_1 \gamma_1$ . Notemos que si no hay relación lineal entre  $X_t$  y  $X_{t+1}$ , luego tendría sentido tomar  $\alpha = 0$ , y esto resulta ser así pues  $\alpha = \rho_1 = 0$  si no hay dependencia lineal entre  $X_t$  y  $X_{t+1}$ .

2. **Se le “extrae” a  $X_{t+2}$  la dependencia lineal que tiene con  $X_{t+1}$ :** para ello se define, del mismo modo que en inciso anterior,  $X_{t+2}^* = X_{t+2} - \beta X_{t+1}$  y se elige  $\beta$  de tal forma que se minimice la varianza de  $X_{t+2}^*$ . Pero, repitiendo el cálculo hecho en el inciso anterior, resulta que  $var(X_{t+2}^*) = \gamma_0 + \beta^2 \gamma_0 - 2\beta \gamma_1$ , de modo que el valor de  $\beta$  que se obtiene es el mismo que para  $\alpha$ , es decir,  $\beta = \alpha = \frac{\gamma_1}{\gamma_0} = \rho_1$  y, en consecuencia,  $var(X_{t+2}^*) = var(X_t^*) = \gamma_0 - \rho_1 \gamma_1$ .

3. **Se calcula el coeficiente de correlación parcial  $\pi_2$  entre  $X_{t+2}$  y  $X_t$ :**

$\pi_2 = corr(X_t, X_{t+2}/X_{t+1})$  es por definición el coeficiente de correlación entre  $X_t^*$  y  $X_{t+2}^*$ , es decir,  $\pi_2 = corr(X_t^*, X_{t+2}^*) = \frac{cov(X_t^*, X_{t+2}^*)}{\sqrt{var(X_{t+2}^*)} \sqrt{var(X_t^*)}} = \frac{cov(X_t^*, X_{t+2}^*)}{\gamma_0 - \rho_1 \gamma_1}$ .

Pero como  $cov(X_t^*, X_{t+2}^*) = cov(X_t - \rho_1 X_{t+1}, X_{t+2} - \rho_1 X_{t+1}) = cov(X_t, X_{t+2}) - \rho_1 cov(X_{t+1}, X_{t+2}) - \rho_1 cov(X_t, X_{t+1}) + \rho_1^2 cov(X_{t+1}, X_{t+1}) = \gamma_2 - \rho_1 \gamma_1 - \rho_1 \gamma_1 + \rho_1^2 \gamma_0 = \gamma_2 - 2\rho_1 \frac{\gamma_1}{\gamma_0} \gamma_0 + \rho_1^2 \gamma_0 = \gamma_2 - \rho_1^2 \gamma_0 = \gamma_2 - \frac{\gamma_1^2}{\gamma_0} \gamma_0 = \gamma_2 - \rho_1 \gamma_1$ , resulta que

$\pi_2 = \frac{\gamma_2 - \rho_1 \gamma_1}{\gamma_0 - \rho_1 \gamma_1}$ , o, dividiendo numerador y denominador por  $\gamma_0$ , obtenemos que,

$$\pi_2 = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}.$$

Así, en general, puede probarse que para un cierto  $k \in \mathbb{N}$ , la fórmula general del coeficiente de correlación parcial teórica  $\pi_k = corr(X_t, X_{t+k}/X_{t+1}, \dots, X_{t+k-1})$ , esta dada por:

$$\pi_k = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{vmatrix}}$$

Otra manera de calcular el coeficiente de correlación parcial  $\pi_2$  entre  $X_t$  y  $X_{t+2}$  es considerando una regresión lineal entre  $X_{t+2} - \mu$  como variable dependiente y entre  $X_{t+1} - \mu$  y  $X_t - \mu$  como variables independientes (o variables regresoras), es decir:

$$X_{t+2} - \mu = \phi_{21}(X_{t+1} - \mu) + \phi_{22}(X_t - \mu) + E_{t+2} \text{ para } t \in \mathbb{Z} \quad (1.3.1)$$

donde se supone que  $\{E_t : t \in \mathbb{Z}\}$  es un proceso de ruido blanco tal que las variables  $E_{t+2}$  no están correlacionadas con  $X_t$  y  $X_{t+1}$ , para todo  $t$ . Multiplicando a ambos miembros de (1.3.1) por  $(X_{t+1} - \mu)$  y tomando esperanzas, resulta que

$$E[(X_{t+2} - \mu)(X_{t+1} - \mu)] = \phi_{21}E[(X_{t+1} - \mu)^2] + \phi_{22}E[(X_t - \mu)(X_{t+1} - \mu)] + 0$$

osea que,  $\gamma_1 = \phi_{21}\gamma_0 + \phi_{22}\gamma_1$ . Similarmente, si se multiplica a ambos miembros de (1.3.1) por  $(X_t - \mu)$  y se toma esperanza, el resultado que se obtiene es  $\gamma_2 = \phi_{21}\gamma_1 + \phi_{22}\gamma_0$ . Dividiendo a ambos miembros de las últimas dos ecuaciones obtenidas por  $\gamma_0$ , se obtiene el siguiente sistema de ecuaciones en las incógnitas  $\phi_{21}$  y  $\phi_{22}$ :

$$\begin{cases} 1\phi_{21} + \rho_1\phi_{22} = \rho_1 \\ \rho_1\phi_{21} + 1\phi_{22} = \rho_2 \end{cases}, \text{ o bien, } \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} \begin{bmatrix} \phi_{21} \\ \phi_{22} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix}.$$

Luego usando la regla de Cramer para  $\phi_{22}$ , resulta que:

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} \doteq \pi_2$$

Luego si reescribimos la fórmula (1.3.1) como

$$X_{t+2} = \mu(1 - \phi_{21} - \phi_{22}) + \phi_{21}X_{t+1} + \phi_{22}X_t + E_{t+2}$$

y denotamos por  $\phi_{20}$  al término  $\mu(1 - \phi_{21} - \phi_{22})$  resulta que

$$X_{t+2} = \phi_{20} + \phi_{21}X_{t+1} + \phi_{22}X_t + E_{t+2}$$

que es un modelo de regresión lineal de  $X_{t+2}$  respecto de  $X_{t+1}$  y  $X_t$ , en el que resulta que el coeficiente  $\phi_{22}$  de  $X_t$  es el coeficiente de correlación parcial teórico entre  $X_{t+2}$  y  $X_t$ . Similarmente si se realiza una regresión lineal de  $X_{t+k}$  respecto de  $X_{t+k-1}$ ,  $X_{t+k-2}, \dots, X_t$ ; es decir,

$$X_{t+k} = \phi_{k0} + \phi_{k1}X_{t+k-1} + \phi_{k2}X_{t+k-2} + \dots + \phi_{kk}X_t + E_{t+k}$$

el coeficiente  $\phi_{kk}$  de  $X_t$  es el coeficiente  $\pi_k$  de correlación parcial teórico entre  $X_{t+k}$  y  $X_t$ .

**Definición 1.3.6:** Sea  $\{X_t : t \in \mathbb{Z}\}$  una serie de tiempo estacionaria. La función  $\pi : \mathbb{Z} \rightarrow \mathbb{R}$  que a cada valor  $k$  le asigna el coeficiente de correlación parcial teórico  $\pi_k$  entre  $X_t$  y  $X_{t+k}$ , se denomina **función de autocorrelación parcial teórica** de la serie de tiempo  $\{X_t : t \in \mathbb{Z}\}$ .

**Ejemplo 1.3.1:** Sea  $\{A_t : t \in T\}$  un proceso de ruido blanco, luego como  $\mu_A = 0$ ,

$$\delta_k \doteq cov(A_t, A_{t+k}) = E[A_t \cdot A_{t+k}] = \begin{cases} \sigma_A^2 & \text{si } k = 0 \\ 0 & \text{si } k \neq 0 \end{cases}.$$

Definimos a partir de este proceso, la serie de tiempo  $\{X_t : t \in \mathbb{Z}\}$  mediante la relación:

$$X_t = A_t + \theta A_{t-1}$$

donde  $\theta$  es un número fijo. A una serie de tiempo así definida se la denomina **Promedio móvil de orden 1**, indicándose ésto con la notación **MA(1)**.

Notemos que esta serie de tiempo es estacionaria, en efecto:

1.  $E[X_t] = E[A_t + \theta A_{t-1}] = E[A_t] + \theta E[A_{t-1}] = 0 \quad \forall t \in \mathbb{Z}$ .
2.  $cov(X_t, X_{t+k}) = E[X_t X_{t+k}] = E[(A_t + \theta A_{t-1})(A_{t+k} + \theta A_{t+k-1})] = E[A_t \cdot A_{t+k}] + \theta^2 E[A_{t-1} \cdot A_{t+k-1}] + \theta E[A_t \cdot A_{t+k-1}] + \theta E[A_{t-1} \cdot A_{t+k}] = \delta_k + \theta^2 \delta_k + \theta \delta_{k-1} + \theta \delta_{k+1} = \gamma_k$  (independiente de t).

Luego la serie  $\{X_t : t \in \mathbb{Z}\}$  es estacionaria.

Por otro lado, notemos que si reemplazamos  $\delta_k$  por su expresión, obtenemos que:

$$\gamma_k = \begin{cases} (1 + \theta^2)\sigma_A^2 & \text{si } k = 0 \\ \theta\sigma_A^2 & \text{si } |k| = 1 \\ 0 & \text{si } |k| > 1 \end{cases}$$

por lo que,

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} 1 & \text{si } k = 0 \\ \frac{\theta}{1+\theta^2} & \text{si } |k| = 1 \\ 0 & \text{si } |k| > 1 \end{cases} .$$

Notar que el valor  $\rho_1$  de esta serie está comprendido entre  $-0,5$  y  $0,5$ . En efecto,  $\rho_1 = \frac{\theta}{1+\theta^2}$  y su derivada respecto de  $\theta$  igualada a 0 es  $\frac{d\rho_1}{d\theta} = \frac{1+\theta^2-2\theta^2}{(1+\theta^2)^2} = 0$  si, y sólo si,  $\theta^2 = 1$  si, y sólo si,  $\theta = \pm 1$ . Es decir,  $\theta = 1$  y  $\theta = -1$  son los puntos críticos que resultan ser donde la función  $\rho_1$  alcanza su máximo y su mínimo absoluto, siendo éstos los que se obtienen al reemplazar  $\theta$  en la fórmula  $\rho_1 = \frac{\theta}{1+\theta^2}$  por  $\theta = 1$  y  $\theta = -1$  respectivamente, obteniendo que  $\max_{\theta \in \mathbb{R}} \rho_1 = \frac{1}{2}$  y  $\min_{\theta \in \mathbb{R}} \rho_1 = -\frac{1}{2}$ . El correlograma teórico de  $X_t = A_t + \theta A_{t-1}$  responde a una de las dos formas que se muestran en la Figura 1.2.

**Ejemplo 1.3.2:** Sea  $\{A_t : t \in T\}$  un proceso de ruido blanco, luego como  $\mu_A = 0$ ,

$$\delta_k \doteq cov(A_t, A_{t+k}) = E[A_t \cdot A_{t+k}] = \begin{cases} \sigma_A^2 & \text{si } k = 0 \\ 0 & \text{si } k \neq 0 \end{cases} .$$

Definimos a partir de éste proceso, la serie de tiempo  $\{X_t : t \in \mathbb{Z}\}$  mediante la relación:

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

donde  $\phi$  es un número fijo. A una tal serie de tiempo se la denomina **Proceso autorregresivo de orden 1**, indicándose ésto con la notación **AR(1)**. Como la relación anterior es válida para todo  $t \in \mathbb{Z}$ , se puede sustituir  $X_{t-1} = \phi X_{t-2} + A_{t-1}$  en ella obteniéndose,  $X_t = \phi X_{t-1} + A_t = \phi(\phi X_{t-2} + A_{t-1}) + A_t = \phi^2 X_{t-2} + \phi A_{t-1} + A_t$ . Reemplazando ahora  $X_{t-2}$  por  $\phi X_{t-3} + A_{t-2}$  se tiene  $X_t = \phi^3 X_{t-3} + \phi^2 A_{t-2} + \phi A_{t-1} + A_t$ . Así, repitiendo este proceso  $k - 1$  veces, se obtiene:



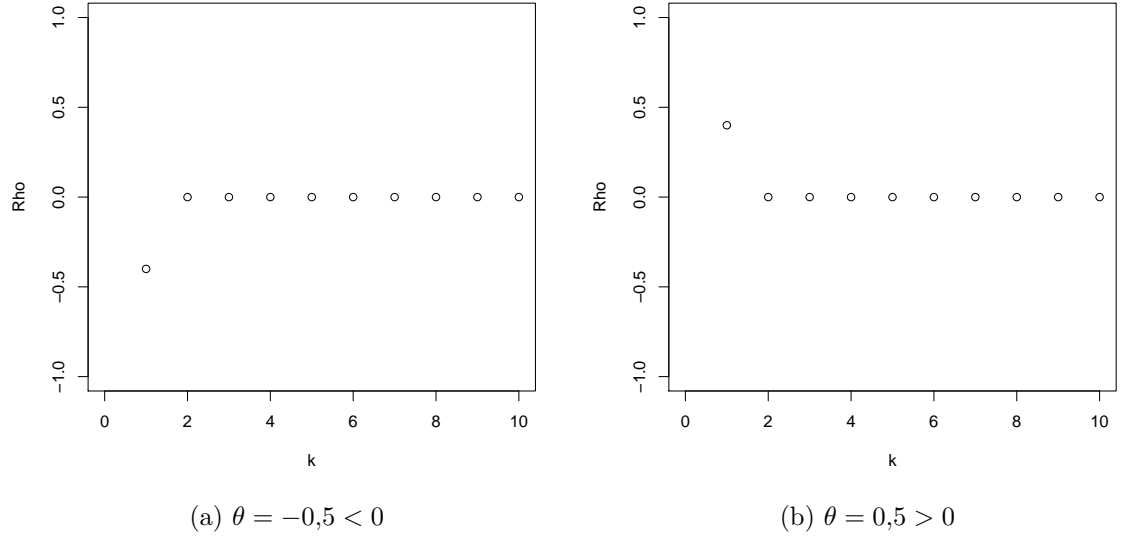


Figura 1.2: Ejemplo de correlograma teórico de  $X_t = A_t + \theta A_{t-1}$  para un  $\theta < 0$  y un  $\theta > 0$

$$X_t = \phi^k X_{t-k} + \phi^{k-1} A_{t-(k-1)} + \dots + \phi A_{t-1} + A_t$$

es decir que,

$$X_t = \phi^k X_{t-k} + \sum_{i=0}^{k-1} \phi^i A_{t-i}$$

para  $k = 1, 2, \dots$ . Si se supone que  $|\phi| < 1$  y que  $E[X_t^2] < M < \infty$  para todo  $t \in \mathbb{Z}$  con  $M \in \mathbb{N}$ , y se continúa con este proceso indefinidamente, se tiene que

$$X_t = \sum_{n=0}^{\infty} \phi^n A_{t-n}$$

donde esta igualdad debe entenderse como convergencia en cuadrados medios, es decir,

$$E\left[\left(X_t - \sum_{i=0}^{k-1} \phi^i A_{t-i}\right)^2\right] \rightarrow 0 \text{ cuando } k \rightarrow \infty.$$

Luego,  $X_t$  queda expresado como un promedio móvil de orden infinito del proceso de ruido blanco  $\{A_t : t \in \mathbb{Z}\}$ , de lo que hablaremos con mas detalle en el capítulo siguiente. Con este resultado, es fácil ver que  $E[X_t] = \sum_{n=0}^{\infty} \phi^n E[A_{t-n}] = 0$ , en efecto, dado  $\epsilon > 0$ ,

$$|E[X_t] - \sum_{i=0}^k \phi^i E[A_{t-i}]| = |E[X_t] - E[\sum_{i=0}^k \phi^i A_{t-i}]| = |E[X_t - \sum_{i=0}^k \phi^i A_{t-i}]| \leq$$

$$E[|X_t - \sum_{i=0}^k \phi^i A_{t-i}|] \leq (E[(X_t - \sum_{i=0}^k \phi^i A_{t-i})^2])^{\frac{1}{2}} \leq \epsilon \text{ para } k \gg 0.$$

Además, dado un  $k \in \mathbb{N}$  fijo, se tiene que,

$$\text{cov}(X_t, X_{t+k}) = E[X_t \cdot X_{t+k}] = E[(\sum_{n=0}^{\infty} \phi^n A_{t-n})(\sum_{m=0}^{\infty} \phi^m A_{t+k-m})] =$$

$$\sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \phi^n \phi^m E[A_{t-n} \cdot A_{t+k-m}] = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \phi^{n+m} E[A_{t-n} \cdot A_{t-(m-k)}]$$

pero, como  $E[A_{t-n} \cdot A_{t-(m-k)}] = \begin{cases} \sigma_A^2 & \text{si } n = m - k \\ 0 & \text{si } n \neq m - k \end{cases}$ , resulta que para cada valor de  $n$  fijo de la primera sumatoria, el único término de la segunda sumatoria que es distinto de 0 es el que corresponde a  $m = n + k$ ; en consecuencia,

$$\text{cov}(X_t, X_{t+k}) = \sum_{n=0}^{\infty} \phi^{n+n+k} E[A_{t-n} \cdot A_{t-[(n+k)-k]}] = \sum_{n=0}^{\infty} \phi^{2n+k} E[A_{t-n} \cdot A_{t-n}] =$$

$$\sum_{n=0}^{\infty} \phi^{2n} \phi^k \sigma_A^2 = \phi^k \sigma_A^2 \sum_{n=0}^{\infty} \phi^{2n}.$$

Para que este valor sea un número finito, se debe verificar que la suma infinita  $\sum_{n=0}^{\infty} \phi^{2n}$  converja a un número finito, lo cual ocurre si, y sólo si,  $\phi^2 < 1$  o, equivalentemente, si  $|\phi| < 1$ ; en cuyo caso,  $\sum_{n=0}^{\infty} \phi^{2n} = \frac{1}{1-\phi^2}$ , con lo cual se obtiene que la  $\text{cov}(X_t, X_{t+k})$  depende solo de  $k$  y está dada por

$$\gamma_k = \text{cov}(X_t, X_{t+k}) = \sigma_A^2 \frac{\phi^k}{1-\phi^2} \text{ si } |\phi| < 1$$

Como en particular,  $\gamma_0 = \frac{\sigma_A^2}{1-\phi^2}$ , el resultado anterior se puede expresar como

$$\gamma_k = \phi^k \gamma_0 \text{ si } |\phi| < 1, \text{ donde } \gamma_0 = \frac{\sigma_A^2}{1-\phi^2}.$$

Este resultado también se puede obtener multiplicando a ambos miembros de  $X_t = \phi^k X_{t-k} + \sum_{i=0}^{k-1} \phi^i A_{t-i}$ , fórmula obtenida anteriormente, por  $X_{t+k}$  para cada  $k \in \{1, 2, \dots\}$  y tomando esperanza miembro a miembro, recordando que  $X_{t-k}$  es función de los  $A_{t-k-j}$  con  $j \in \{0, 1, 2, \dots\}$  y, en consecuencia, es no correlacionado con  $A_t, A_{t-1}, \dots, A_{t-(k-1)}$ . De este modo, resulta que  $E[X_t \cdot X_{t+k}] = \phi^k E[X_{t-k}^2] + \sum_{i=0}^{k-1} \phi^i E[A_{t-i} \cdot X_{t+k}] = \phi^k E[X_{t-k}^2]$ , osea que,  $\gamma_k = \phi^k \gamma_0$  si  $|\phi| < 1$  para  $k \in \{1, 2, \dots\}$ .

Notemos que la función de autocorrelación teórica, que se halla dividiendo a la función de autocovarianza por  $\gamma_0$ , esta dada por  $\rho_k = \phi^k$  para todo  $k \in \mathbb{N}$  si  $|\phi| < 1$ . Por lo tanto, el correlograma teórico de  $X_t = \phi X_{t-1} + A_t$  responde a una de las dos formas que se muestran en la Figura 1.3.

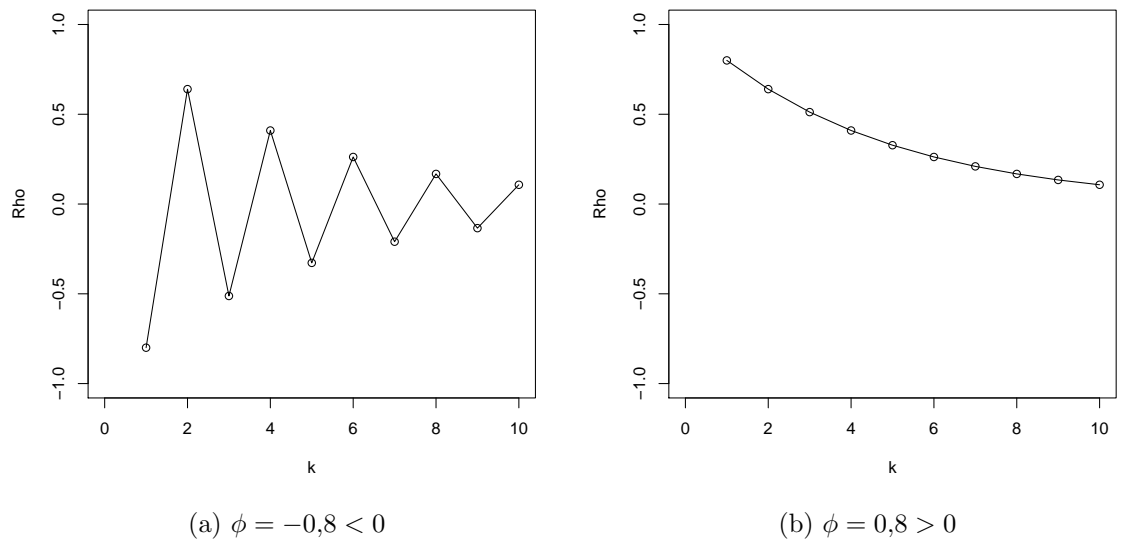


Figura 1.3: Ejemplo de correlograma teórico de  $X_t = \phi X_{t-1} + A_t$  para un  $\phi < 0$  y un  $\phi > 0$ .

Por otro lado, la función de autocorrelación parcial teórica se obtiene calculando los valores de  $\pi_k$  para  $k \in \{1, 2, \dots\}$  a partir de la fórmula general, esto es

$$\pi_1 = \rho_1 = \phi; \pi_2 = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\begin{vmatrix} 1 & \phi \\ \phi & \phi^2 \end{vmatrix}}{\begin{vmatrix} 1 & \phi \\ \phi & 1 \end{vmatrix}} = \frac{\phi^2 - \phi^2}{1 - \phi^2} = 0$$

y en general,

$$\pi_k = \frac{\begin{vmatrix} 1 & \phi & \dots & \phi^{k-2} & \phi \\ \phi & 1 & \dots & \phi^{k-3} & \phi^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \phi^{k-1} & \phi^{k-2} & \dots & \phi & \phi^k \end{vmatrix}}{\begin{vmatrix} 1 & \phi & \dots & \phi^{k-2} & \phi^{k-1} \\ \phi & 1 & \dots & \phi^{k-3} & \phi^{k-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \phi^{k-1} & \phi^{k-2} & \dots & \phi & 1 \end{vmatrix}} = 0 \text{ para } k > 1,$$

pues es fácil ver que el determinante del numerador es igual a cero, ya que la última columna es igual a la primera multiplicada por  $\phi$ , y también se puede demostrar que el determinante del denominador es igual a  $(1 - \phi^2)^{k-1} \neq 0$ , por lo que resulta

$$\begin{cases} \phi & \text{si } k = 1 \\ 0 & \text{si } k \neq 1 \end{cases}.$$

En consecuencia, la representación gráfica de la función de autocorrelación parcial teórica corresponde a alguna de las dos formas que se muestran en la Figura 1.4.

## 1.4. Estimación de las características de una serie de tiempo estacionaria.

Una serie de tiempo estacionaria está caracterizada por su media  $\mu$ , su varianza  $\sigma^2$ , sus autocorrelaciones  $\rho_k$  y sus autocorrelaciones parciales  $\pi_k$ . Estos parámetros son en general desconocidos y es necesario estimarlos a partir de los datos muestrales, usando promedios a lo largo del tiempo.

### 1.4.1. Estimación de la media $\mu$ .

Dada una serie de tiempo estacionaria  $\{X_t : t \in \mathbb{Z}\}$ , y considerando que en general se contará solamente con  $n$  valores de una única realización de dicha serie de tiempo, digamos  $x_1, x_2, \dots, x_n$ , el estimador natural de la media común  $\mu = E[X_t]$  es la media muestral

$$\bar{X}_n = \frac{1}{n} \sum_{t=1}^n X_t$$

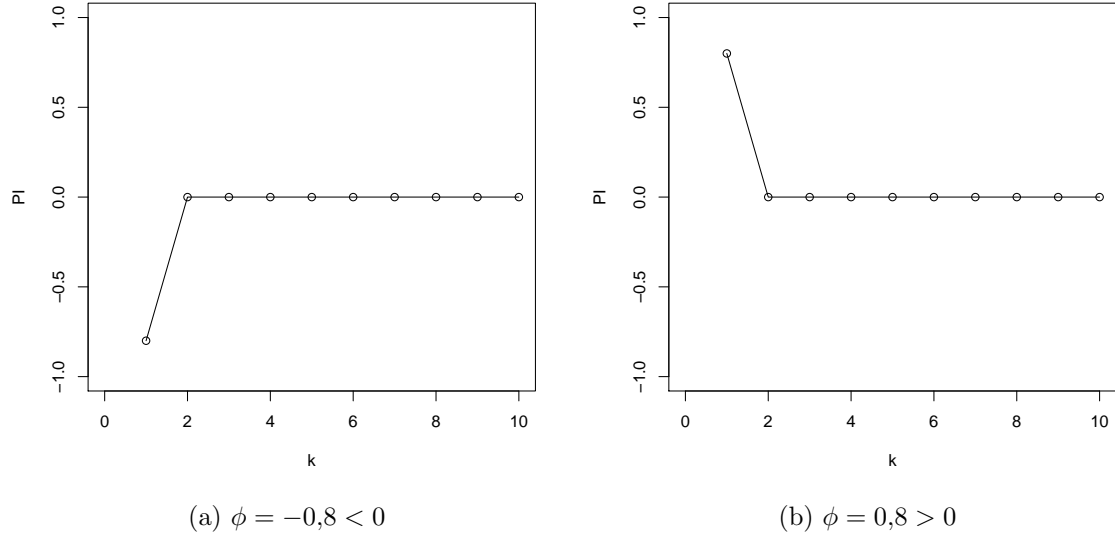


Figura 1.4: Función de autocorrelación parcial teórica de  $X_t = \phi X_{t-1} + A_t$  para un  $\phi < 0$  y un  $\phi > 0$ .

Es claro que  $\bar{X}_n$  es un estimador insesgado para  $\mu$ , ya que  $E[\bar{X}_n] = \frac{1}{n} \sum_{t=1}^n E[X_t] = \mu$ , pero ¿será  $\bar{X}_n$  un estimador consistente en cuadrados medios para  $\mu$ ? Más precisamente nos interesaría conocer bajo que condiciones se cumplirá que  $\lim_{n \rightarrow \infty} E[(\bar{X}_n - \mu)^2] = \lim_{n \rightarrow \infty} var(\bar{X}_n) = 0$ . Calculemos para ello la  $var(\bar{X}_n)$ :

$$var(\bar{X}_n) = var\left(\frac{1}{n} \sum_{t=1}^n X_t\right) = \frac{1}{n^2} \sum_{t=1}^n \sum_{s=1}^n cov(X_t, X_s) = \frac{1}{n^2} \sum_{t=1}^n \sum_{s=1}^n \gamma_{|t-s|}$$

Realizando el cambio de variable  $k = t - s$ , esta doble suma sobre el reticulado de puntos  $\{(t, s) : t \in \{1, 2, \dots, n\}, s \in \{1, 2, \dots, n\}\}$ , es igual a la suma realizada sobre las diagonales de dicho reticulado que corresponden a  $t - s = -(n - 1), \dots, -1, 0, 1, \dots, (n - 1)$ , es decir que:

$$var(\bar{X}_n) = \frac{1}{n^2} \sum_{k=-(n-1)}^{n-1} (n - |k|)\gamma_k = \frac{\gamma_0}{n} \sum_{k=-(n-1)}^{n-1} \left(1 - \frac{|k|}{n}\right)\rho_k,$$

si multiplicamos y dividimos por  $\gamma_0$ . Por lo tanto, si  $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=-(n-1)}^{n-1} \left(1 - \frac{|k|}{n}\right)\rho_k = 0$ , entonces se verificará que  $\lim_{n \rightarrow \infty} var(\bar{X}_n) = 0$  (esta propiedad se denomina ergódica).

Una condición para que esto se verifique es que  $\lim_{n \rightarrow \infty} \rho_k = 0$ , ya que si vale esta igualdad, entonces para cualquier  $\epsilon > 0$  se puede elegir un  $N \in \mathbb{N}$  tal que para todo  $k > N$  se verifique que  $|\rho_k| < \frac{\epsilon}{4}$  y, por lo tanto,  $\frac{2}{n} \sum_{k=N+1}^{n-1} |\rho_k| \leq \frac{2}{n}(n-1-N-1)\frac{\epsilon}{4} \leq \frac{2}{n}n\frac{\epsilon}{4} = \frac{\epsilon}{2}$ ; luego, dado  $n > N$ ,  $|\frac{1}{n} \sum_{k=-(n-1)}^{n-1} (1 - \frac{|k|}{n})\rho_k| \leq \frac{2}{n} \sum_{k=0}^{n-1} (1 - \frac{|k|}{n})|\rho_k| \leq \frac{2}{n} \sum_{k=0}^{n-1} |\rho_k| = \frac{2}{n} \sum_{k=0}^N |\rho_k| + \frac{2}{n} \sum_{k=N+1}^{n-1} |\rho_k| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$  si elegimos  $n$  suficientemente grande como para que  $\frac{2}{n} \sum_{k=0}^N |\rho_k| \leq \frac{\epsilon}{2}$ . Intuitivamente, el hecho de que,  $\lim_{n \rightarrow \infty} \rho_k = 0$  implica que  $\lim_{n \rightarrow \infty} var(\bar{X}_n) = 0$ , afirma que, si los valores del proceso que están suficientemente separados entre sí son prácticamente no correlacionados, entonces ellos contribuyen adicionando nueva información útil acerca de  $\mu$ , cuando uno realiza su promedio a lo largo del tiempo, lográndose que cuando  $n \rightarrow \infty$ ,  $\bar{X}_n \rightarrow \mu$ .

### 1.4.2. Estimación de las autocovarianzas.

Dadas  $n$  observaciones  $x_1, x_2, \dots, x_n$  de una serie de tiempo discreto estacionaria, uno puede formar pares de observaciones donde la primera componente de cada par sea una observación realizada  $k$  tiempos antes que la observación que aparece como segunda componente. Se puede formar para un  $k$  fijo, con  $k \in \{0, 1, \dots, n-1\}$ ,  $n-k$  de tales pares, digamos

$$(x_1, x_{1+k}), (x_2, x_{2+k}), \dots, (x_{n-k}, x_n)$$

Mirando a la primera observación de cada par como correspondiente a una variable, y a la segunda como a una observación de otra variable, se puede calcular una estimación de la covarianza  $\gamma_k$  mediante la fórmula

$$\frac{1}{n-k} \sum_{t=1}^{n-k} (x_t - \bar{x}_{(1)})(x_{t+k} - \bar{x}_{(2)})$$

donde  $\bar{x}_{(1)} = \frac{1}{n-k} \sum_{t=1}^{n-k} x_t$  y  $\bar{x}_{(2)} = \frac{1}{n-k} \sum_{t=k+1}^n x_t$ . Como esta fórmula es demasiado complicada y, en general,  $\bar{x}_{(1)} \cong \bar{x}_{(2)}$ , se conviene en reemplazar ambas por  $\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t$ . Surgen entonces dos posibles estimadores de  $\gamma_k$ :

1.  $\hat{\gamma}_k = \frac{1}{n} \sum_{t=1}^{n-k} (X_t - \bar{X}_n)(X_{t+k} - \bar{X}_n)$

$$2. \tilde{\gamma}_k = \frac{1}{n-k} \sum_{t=1}^{n-k} (X_t - \bar{X}_n)(X_{t+k} - \bar{X}_n)$$

Ahora bien, como

$$\begin{aligned} \sum_{t=1}^{n-k} (X_t - \bar{X}_n)(X_{t+k} - \bar{X}_n) &= \sum_{t=1}^{n-k} ((X_t - \mu) - (\bar{X}_n - \mu))((X_{t+k} - \mu) - (\bar{X}_n - \mu)) \\ &= \sum_{t=1}^{n-k} (X_t - \mu)(X_{t+k} - \mu) - (\bar{X}_n - \mu) \sum_{t=1}^{n-k} (X_t - \mu) - (\bar{X}_n - \mu) \sum_{t=1}^{n-k} (X_{t+k} - \mu) + \\ &\quad (n-k)(\bar{X}_n - \mu)^2 \\ &= \sum_{t=1}^{n-k} (X_t - \mu)(X_{t+k} - \mu) - (\bar{X}_n - \mu) \left( \frac{(\sum_{t=1}^{n-k} X_t) - (n-k)\mu}{(n-k)} \right) (n-k) - \\ &\quad (\bar{X}_n - \mu)(n-k) \left( \frac{(\sum_{t=1}^{n-k} X_{t+k}) - (n-k)\mu}{(n-k)} \right) + (n-k)(\bar{X}_n - \mu)^2 \\ &= \sum_{t=1}^{n-k} (X_t - \mu)(X_{t+k} - \mu) - (n-k)(\bar{X}_n - \mu)(\bar{X}_{(1)} - \mu) - \\ &\quad (n-k)(\bar{X}_n - \mu)(\bar{X}_{(2)} - \mu) + (n-k)(\bar{X}_n - \mu)^2 \\ &\cong \sum_{t=1}^{n-k} (X_t - \mu)(X_{t+k} - \mu) - (n-k)(\bar{X}_n - \mu)^2, \end{aligned}$$

donde en la última aproximación hemos reemplazado a  $\bar{X}_{(1)}$  y a  $\bar{X}_{(2)}$  por  $\bar{X}_n$ , entonces tomando esperanzas a  $\hat{\gamma}_k$  y a  $\tilde{\gamma}_k$ , resulta que

$$E[\hat{\gamma}_k] = \left(\frac{n-k}{n}\right) \cdot \gamma_k - \left(\frac{n-k}{n}\right) \cdot \text{var}(\bar{X}_n)$$

y que

$$E[\tilde{\gamma}_k] = \gamma_k - \text{var}(\bar{X}_n).$$

Es claro entonces que ambos estimadores son sesgados; sin embargo, cuando se verifica la propiedad ergódica ( $\lim_{n \rightarrow \infty} \text{var}(\bar{X}_n) = 0$ ) ambos estimadores resultan asintóticamente insesgados para  $\gamma_k$ . Por otra parte, dos características de  $\hat{\gamma}_k$  hace que se lo prefiera en lugar de  $\tilde{\gamma}_k$ ; puede probarse que para ciertos tipos de procesos el error cuadrático medio de  $\hat{\gamma}_k$  es menor al de  $\tilde{\gamma}_k$ , y que además,  $\hat{\gamma}_k$  es siempre una función

de  $k$  semidefinida positiva como lo es  $\gamma_k$ , mientras que  $\tilde{\gamma}_k$  no necesariamente cumple con esta propiedad.

**Definición 1.4.1:** La función  $\hat{\gamma} : \mathbb{Z} \rightarrow \mathbb{R}$  que a cada  $k \in \mathbb{Z}$  le asigna el valor  $\hat{\gamma}_k$  recibe el nombre de **función de autocovarianza muestral**.

### 1.4.3. Estimación de las autocorrelaciones.

Como las autocorrelaciones teóricas  $\rho_k$  están relacionadas con las covarianzas teóricas mediante la relación  $\rho_k = \frac{\gamma_k}{\gamma_0}$  un estimador natural de  $\rho_k$  es

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{n-k} (X_t - \bar{X}_n)(X_{t+k} - \bar{X}_n)}{\sum_{t=1}^n (X_t - \bar{X}_n)^2} \text{ para } k = 0, \pm 1, \pm 2, \dots, \pm(n-1).$$

**Definición 1.4.2:** La función  $\hat{\rho} : \mathbb{Z} \rightarrow \mathbb{R}$  que a cada valor  $k \in \mathbb{Z}$  le asigna el valor  $\hat{\rho}_k$  recibe el nombre de **función de autocorrelación muestral**, y el gráfico de los puntos pertenecientes al conjunto  $\{(k, \hat{\rho}_k) : k \in \{1, \dots, n-1\}\}$ , donde  $\hat{\rho}_k$  es la estimación de  $\rho_k$  calculada a partir de la muestra  $x_1, x_2, \dots, x_n$  de tamaño  $n$  es llamado **correlograma muestral**.

Las propiedades muestrales de  $\hat{\rho}_k$  son muy complicadas de obtener, pero para  $n$  suficientemente grande se pueden hallar expresiones aproximadas para sus propiedades de segundo orden. En particular, si la serie de tiempo estacionaria es gaussiana, Bartlett (1946) ha demostrado que

$$cov(\hat{\rho}_k, \hat{\rho}_{k+j}) \cong \frac{1}{n} \sum_{i=-\infty}^{+\infty} (\rho_i \rho_{i+j} + \rho_{i+k+j} \rho_{i-k} - 2\rho_k \rho_i \rho_{i-k-j} - 2\rho_{k+j} \rho_i \rho_{i-k} + 2\rho_k \rho_{k+j} \rho_i^2)$$

y, en particular,

$$var(\hat{\rho}_k) \cong \frac{1}{n} \sum_{i=-\infty}^{\infty} (\rho_i^2 + \rho_{i-k} \rho_{i+k} - 4\rho_k \rho_i \rho_{i-k} + 2\rho_k^2 \rho_i^2).$$

Además para  $n$  grande,  $\hat{\rho}_k$  está distribuido aproximadamente normal con media  $\rho_k$  y varianza dada por la expresión anterior.

Si la serie de tiempo es tal que  $\rho_k = 0$  para  $k > m$ , entonces la fórmula de la varianza se reduce a

$$var(\hat{\rho}_k) \cong \frac{1}{n} (1 + 2\rho_1 + 2\rho_2^2 + \dots + 2\rho_m^2) \text{ para } k > m.$$

Como en la práctica los parámetros  $\rho_i$  tal que  $i \in \{1, \dots, m\}$  son desconocidos, ellos son reemplazados por sus estimaciones muestrales, obteniéndose como estimación del error estándar de  $\hat{\rho}_k$  a



$$S_{\hat{\rho}_k} \cong \sqrt{\frac{1}{n}(1 + 2\hat{\rho}_1^2 + \dots + 2\hat{\rho}_m^2)} \text{ para } k > m.$$

**Ejemplo 1.4.1:** Sea  $\{A_t : t \in \mathbb{Z}\}$  un proceso de ruido blanco. Recordemos que en la sección 1.3 de este capítulo habíamos calculado la función de autocorrelación teórica de ruido blanco  $\rho_k = \begin{cases} 1 & \text{si } k = 0 \\ 0 & \text{si } k \neq 0 \end{cases}$ ; en consecuencia, para  $0 \leq k$  y para  $j \in \mathbb{Z}$  es

$$\text{cov}(\hat{\rho}_k, \hat{\rho}_{k+j}) \cong \begin{cases} \frac{1}{n} & \text{si } j = 0 \\ 0 & \text{si } j \neq 0 \end{cases}$$

osea que,  $S_{\hat{\rho}_k} = \sqrt{\frac{1}{n}}$ . Este valor es el que se usará luego para decidir, en base a los datos muestrales, si un proceso es de ruido blanco.

#### 1.4.4. Estimación de las autocorrelaciones parciales.

Los estimadores naturales de  $\pi_k = \phi_{kk}$  son los que se obtienen reemplazando en la fórmula que expresa a  $\pi_k$  como un cociente de dos determinantes, a los  $\rho_j$  por sus estimadores  $\hat{\rho}_j$ .

Para valores de  $k$  grandes el cálculo de estos determinantes es complicado, es por eso conveniente usar un método recursivo para su evaluación que, comenzando por  $\hat{\pi}_1 = \hat{\phi}_{11} = \hat{\rho}_1$ , usa la fórmula:

$$\hat{\pi}_{k+1} = \frac{\hat{\rho}_{k+1} - \sum_{j=1}^k \hat{\phi}_{k,j} \hat{\rho}_{k+1-j}}{1 - \sum_{j=1}^k \hat{\phi}_{k,j} \hat{\rho}_j}$$

donde,

$$\hat{\phi}_{k,j} = \hat{\phi}_{k-1,j} - \hat{\phi}_{k-1,k-j} \cdot \hat{\pi}_k \text{ para } j \in \{1, \dots, k\}.$$

Este método es debido a Durbin (1960). Para el caso de un proceso de ruido blanco, se tiene que la  $\text{var}(\hat{\pi}_k) \cong \frac{1}{n}$ .

---

## Capítulo 2

# Modelos de series de tiempo estacionarias

### 2.1. Representaciones de una serie de tiempo.

#### 2.1.1. Representación de una serie de tiempo como un proceso de promedios móviles de orden infinito.

En 1938, Herman Wold probó que todo proceso estacionario  $\{X_t : t \in \mathbb{Z}\}$  que sea no determinístico puro, puede siempre expresarse bajo la forma de un **promedio móvil de orden infinito**, es decir,

$$X_t = \mu + A_t + \alpha_1 A_{t-1} + \alpha_2 A_{t-2} + \dots = \mu + \sum_{i=0}^{\infty} \alpha_i A_{t-i}$$

con  $\alpha_0 = 1$  y donde  $\{A_t : t \in \mathbb{Z}\}$  es un proceso de ruido blanco  $(0, \sigma_A^2)$  y  $\sum_{i=0}^{\infty} \alpha_i^2 < \infty$ .

**Definición 2.1.1:** Sea  $\{X_t : t \in \mathbb{Z}\}$  una serie de tiempo. Definimos el operador  $B: \overline{\langle \{X_t : t \in \mathbb{Z}\} \rangle} \rightarrow \mathcal{L}^2(\Omega, A, P)$  dado por  $B(X_t) = X_{t-1}$  y lo llamaremos **operador de cambio hacia atrás**; donde se considera que  $\overline{\langle \{X_t : t \in \mathbb{Z}\} \rangle}$  es la clausura de  $\langle \{X_t : t \in \mathbb{Z}\} \rangle$  en  $\mathcal{L}^2(\Omega, A, P)$ .

**Nota 2.1.1:** Adoptaremos la convención  $B^0(X_t) = 1$  para todo  $t \in \mathbb{Z}$ .

**Nota 2.1.2:** Puede probarse que el operador  $B$  es una isometría de norma 1 en el espacio  $\mathcal{B}(\overline{\langle \{X_t : t \in \mathbb{Z}\} \rangle}) \doteq \{A : \overline{\langle \{X_t : t \in \mathbb{Z}\} \rangle} \rightarrow \overline{\langle \{X_t : t \in \mathbb{Z}\} \rangle} : A \text{ es lineal y continua}\}$ .

**Proposición 2.1.1:** Dada  $\{X_t : t \in \mathbb{Z}\}$  serie de tiempo,  $B$  operador de cambio hacia atrás y  $k \in \mathbb{N} \cup \{0\}$ , entonces  $B^k(X_t) = X_{t-k}$ .

**Demostración:** Lo haremos por inducción. Sea  $P(n)$  la propiedad  $B^n(X_t) = X_{t-n}$  para cada  $n \in \mathbb{N}$ . Si  $n = 2$ , luego  $B^2(X_t) \doteq B(B(X_t)) \doteq B(X_{t-1}) \doteq X_{(t-1)-1} = X_{t-2}$ , de modo que,  $P(2)$  es verdadera. Supongamos ahora que  $P(k)$  es verdadera para algún  $k \geq 2$ , luego  $B^{k+1}(X_t) \doteq B(B^k(X_t)) = B(X_{t-k})$  por hipótesis inductiva, y como  $B(X_{t-k}) \doteq X_{(t-k)-1} = X_{t-(k+1)}$ , luego  $P(k+1)$  es verdadera. Por lo tanto,  $P(n)$  es verdadera para todo  $n \in \mathbb{N}$ , y como  $B^0(X_t) = 1$  por convención, luego  $B^k(X_t) = X_{t-k}$  para todo  $k \in \mathbb{N} \cup \{0\}$ . ■

Con este operador de cambio hacia atrás, el proceso de promedios móviles de orden infinito

$$X_t = \mu + A_t + \alpha_1 A_{t-1} + \alpha_2 A_{t-2} + \dots$$

se puede escribir como

$$X_t = \mu + (B^0 + \alpha_1 B^1 + \alpha_2 B^2 + \dots)A_t = \mu + \alpha(B)A_t$$

con  $\alpha(B) = B^0 + \alpha_1 B^1 + \alpha_2 B^2 + \dots$ . Si además denotamos por  $\dot{X}_t$  a la diferencia  $X_t - \mu$ , la expresión anterior se reduce a

$$\dot{X}_t = \alpha(B)A_t \text{ donde } \alpha(B) = \sum_{i=0}^{\infty} \alpha_i B^i \text{ con } \alpha_0 = 1.$$

**Proposición 2.1.2:** Sea  $\{X_t : t \in \mathbb{Z}\}$  un proceso de promedios móviles de orden infinito dado por  $X_t = \mu + \alpha(B)A_t$  con  $\alpha(B) = \sum_{i=0}^{\infty} \alpha_i B^i$  tal que  $\alpha_0 = 1$  y  $\sum_{i=0}^{\infty} \alpha_i^2 < \infty$ . Entonces  $\{X_t : t \in \mathbb{Z}\}$  es estacionario.

**Demostración:** Veamos que  $\{X_t : t \in \mathbb{Z}\}$  es estacionario verificando que se dan las dos condiciones necesarias para que esto ocurra:

1. Veamos que  $E[X_t] = \mu$  para todo  $t \in \mathbb{Z}$ , o lo que es lo mismo, veamos que  $E[\dot{X}_t] = 0$  para todo  $t \in \mathbb{Z}$ . Para ello veamos que  $E[\dot{X}_t] = \sum_{i=0}^{\infty} \alpha_i E[A_{t-i}]$ . Dado  $\epsilon > 0$  y dado  $k \in \mathbb{N}$ ,  $|E[\dot{X}_t] - \sum_{i=0}^k \alpha_i E[A_{t-i}]| = |E[\dot{X}_t] - E[\sum_{i=0}^k \alpha_i A_{t-i}]| = |E[\dot{X}_t - \sum_{i=0}^k \alpha_i A_{t-i}]| \leq$

$E[|\dot{X}_t - \sum_{i=0}^k \alpha_i A_{t-i}|] \leq (E[(\dot{X}_t - \sum_{i=0}^k \alpha_i A_{t-i})^2])^{\frac{1}{2}}$ . Como  $\dot{X}_t = \sum_{i=0}^{\infty} \alpha_i A_{t-i}$  significa que  $\left\{ \sum_{i=0}^k \alpha_i A_{t-i} \right\}_k$  converge en cuadrados medios a  $\dot{X}_t$ , luego existe  $N \in \mathbb{N}$  tal que si  $k > N$ , luego  $(E[(\dot{X}_t - \sum_{i=0}^k \alpha_i A_{t-i})^2])^{\frac{1}{2}} < \epsilon$ . Así queda probado que  $E[\dot{X}_t] = \sum_{i=0}^{\infty} \alpha_i E[A_{t-i}]$  y como  $E[A_t] = 0$  para todo  $t \in \mathbb{Z}$ , luego  $E[\dot{X}_t] = 0$  para todo  $t \in \mathbb{Z}$ .

Por otro lado, dado  $k \in \mathbb{N}$ ,

$$\begin{aligned}
 \text{cov}(X_t, X_{t+k}) &= E[\dot{X}_t \dot{X}_{t+k}] \\
 &= E\left[\left(\sum_{n=0}^{\infty} \alpha_n A_{t-n}\right)\left(\sum_{m=0}^{\infty} \alpha_m A_{t+k-m}\right)\right] \\
 &= \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \alpha_n \alpha_m E[A_{t-n} A_{t+k-m}]
 \end{aligned}$$

pero como  $E[A_{t-n} A_{t-(m-k)}] = \begin{cases} \sigma_A^2 & \text{si } n = m - k \\ 0 & \text{si } n \neq m - k \end{cases}$ , resulta que para cada valor de  $n$  fijo de la primera sumatoria, el único término de la segunda sumatoria que es distinto de 0 es el que corresponde a  $m = n + k$ ; en consecuencia,

$$\begin{aligned}
 \text{cov}(X_t, X_{t+k}) &= \sum_{n=0}^{\infty} \alpha_n \alpha_{n+k} E[A_{t-n} A_{t-(n+k)-k}] \\
 &= \sum_{n=0}^{\infty} \alpha_n \alpha_{n+k} E[A_{t-n} A_{t-n}] \\
 &= \sum_{n=0}^{\infty} \alpha_n \alpha_{n+k} \sigma_A^2 \\
 &= \sigma_A^2 \sum_{n=0}^{\infty} \alpha_n \alpha_{n+k} \\
 &= \gamma_k
 \end{aligned}$$

que depende solo de  $k$ , y además como siempre vale que  $|\gamma_k| \leq \gamma_0$  y  $\gamma_0 = \sigma_A^2 \sum_{i=0}^{\infty} \alpha_i^2 < \infty$ , luego  $\text{cov}(X_t, X_{t+k}) = \gamma_k < \infty$ .

Luego  $\{X_t : t \in \mathbb{Z}\}$  resulta ser estacionario. ■

**Nota 2.1.3:** Podemos deducir de la demostración anterior que si  $\{X_t : t \in \mathbb{Z}\}$  es un proceso de promedios móviles de orden infinito dado por  $X_t = \mu + \alpha(B)A_t$

con  $\alpha(B) = \sum_{i=0}^{\infty} \alpha_i B^i$  y  $\alpha_0 = 1$ , y tal que  $\sum_{i=0}^{\infty} \alpha_i^2 < \infty$ , luego  $\gamma_k = \sigma_A^2 \sum_{n=0}^{\infty} \alpha_n \alpha_{n+k}$  y  $\rho_k = \frac{\sum_{i=0}^{\infty} \alpha_i \alpha_{i+k}}{\sum_{i=0}^{\infty} \alpha_i^2}$ , para todo  $k \in \mathbb{N}$ .

### 2.1.2. Representación de una serie de tiempo como un proceso autorregresivo de orden infinito

**Definición 2.1.2:** Se dice que una serie de tiempo  $\{X_t : t \in \mathbb{Z}\}$  es **invertible** o puede expresarse como un **proceso autorregresivo de orden infinito** si  $\dot{X}_t = X_t - \mu$  puede escribirse del siguiente modo:

$$\dot{X}_t = A_t + \beta_1 \dot{X}_{t-1} + \beta_2 \dot{X}_{t-2} + \dots \text{ con } \sum_{i=1}^{\infty} |\beta_i| < \infty$$

o lo que es lo mismo, utilizando el operador de cambio hacia atrás,

$$\dot{X}_t - \beta_1 B(\dot{X}_t) - \beta_2 B^2(\dot{X}_t) - \dots = A_t$$

es decir,

$$\beta(B)\dot{X}_t = A_t \text{ donde } \beta(B) = 1 - \beta_1 B - \beta_2 B^2 - \dots$$

Una pregunta natural que podríamos hacernos es bajo que condiciones un proceso invertible es estacionario. Si un proceso es invertible, podemos utilizar el resultado de Wold antes mencionado, y establecer bajo que condiciones dicho proceso invertible de la forma  $\beta(B)\dot{X}_t = A_t$ , es un proceso de promedios móviles de orden infinito de la forma  $\dot{X}_t = \alpha(B)A_t$ , donde  $\alpha(B) = \sum_{i=0}^{\infty} \alpha_i B^i$  con  $\alpha_0 = 1$  y tal que  $\sum_{i=0}^{\infty} \alpha_i^2 < \infty$ . Para ello, se desea averiguar cuando se verifica que  $\frac{1}{\beta(B)} = \alpha(B)$ . Se puede demostrar que la condición para que esto ocurra, es que las soluciones de  $\beta(B) = 0$  deben ser tales que sus módulos sean mayores que 1. Recíprocamente, para que un proceso estacionario de la forma  $\dot{X}_t = \alpha(B)A_t$  sea invertible, se debe expresar a  $\frac{1}{\alpha(B)}$  como  $\beta(B)$  para lo cual se requiere que las soluciones de  $\alpha(B) = 0$  sean todas de módulo mayor a 1.

## 2.2. Modelos autorregresivos de orden finito.

**Definición 2.2.1:** Se dice que  $\{X_t : t \in \mathbb{Z}\}$  es un **modelo autorregresivo de orden p**, donde  $p$  es un natural fijo, si  $\dot{X}_t = X_t - \mu$  puede expresarse como:

$$\dot{X}_t = \phi_1 \dot{X}_{t-1} + \phi_2 \dot{X}_{t-2} + \dots + \phi_p \dot{X}_{t-p} + A_t \text{ para } t \in \mathbb{Z}$$

o, análogamente, como  $\Phi_p(B)\dot{X}_t = A_t$ , donde  $\Phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$  con B operador de cambio hacia atrás; donde  $\{A_t : t \in \mathbb{Z}\}$  es un proceso de ruido blanco con  $var(A_t) = \sigma_A^2 < \infty$ . Denotaremos a este proceso como **AR(p)**.

**Proposición 2.2.1:** Todo modelo AR(p) es un proceso invertible.

**Demotración:** El modelo AR(p) es un caso particular del modelo autorregresivo de orden infinito  $\dot{X}_t = A_t + \sum_{i=1}^{\infty} \beta_i \dot{X}_{t-i}$  con  $\beta_1 = \phi_1, \dots, \beta_p = \phi_p$  y  $\beta_i = 0$  para todo  $i > p$ , verificándose que  $\sum_{i=1}^{\infty} |\beta_i| = \sum_{i=1}^p |\phi_i| < \infty$ . ■

**Proposición 2.2.2:** Un proceso AR(p) de la forma  $\Phi_p(B)\dot{X}_t = A_t$ , es estacionario si las raíces de  $\Phi_p(B)$  tienen módulo mayor a uno.

**Demostración:** Surge del hecho de que un proceso AR(p) es un caso particular de un proceso autorregresivo de orden infinito. ■

### 2.2.1. Modelo autorregresivo de orden 1

**Definición 2.2.2:** Una serie de tiempo  $\{X_t : t \in \mathbb{Z}\}$  es un **proceso autorregresivo de orden 1**, y lo denotaremos por **AR(1)**, si  $\dot{X}_t = X_t - \mu$  se puede expresar bajo la forma

$$\dot{X}_t = \phi_1 \dot{X}_{t-1} + A_t,$$

donde  $\{A_t : t \in \mathbb{Z}\}$  es un proceso de ruido blanco con  $var(A_t) = \sigma_A^2 < \infty$ .

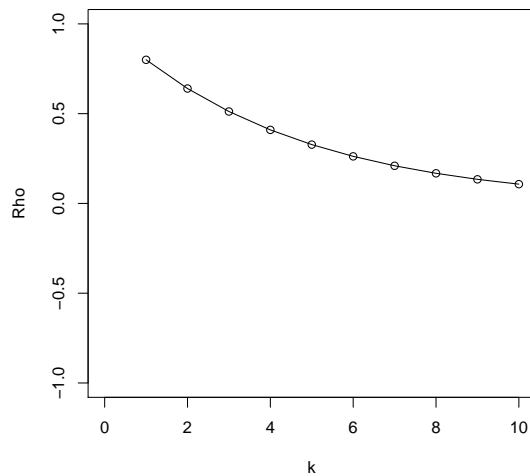
Notemos que si en el ejemplo 1.3.2 reemplazamos a  $X_t$  por  $\dot{X}_t$ , las mismas características que habíamos probado para  $\{X_t : t \in \mathbb{Z}\}$  valen ahora para  $\{\dot{X}_t : t \in \mathbb{Z}\}$ . Gracias a esto, podemos afirmar que:

1.  $E[X_t] = \mu$  para todo  $t \in \mathbb{Z}$ , en efecto,  $E[\dot{X}_t] = 0$  para todo  $t \in \mathbb{Z}$ .
2.  $\gamma_k = cov(X_t, X_{t+k}) = \sigma_A^2 \frac{\phi_1^k}{1-\phi_1^2}$  si  $|\phi_1| < 1$ , pues  $\gamma_k = cov(X_t, X_{t+k}) = E[(X_t - \mu)(X_{t+k} - \mu)] = cov(\dot{X}_t, \dot{X}_{t+k}) = \sigma_A^2 \frac{\phi_1^k}{1-\phi_1^2}$  si  $|\phi_1| < 1$ .
3.  $\rho_k = \frac{\gamma_k}{\gamma_0} = \phi_1^k$  si  $|\phi_1| < 1$ .
4.  $\pi_k = \begin{cases} \phi_1 & \text{si } k = 1 \\ 0 & \text{si } k > 1 \end{cases}$ .

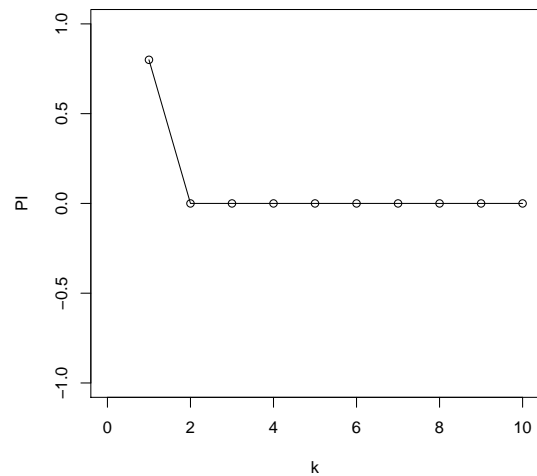
Recordemos que la condición  $|\phi_1| < 1$  la pedíamos para asegurarnos de que el proceso  $\{\dot{X}_t : t \in \mathbb{Z}\}$  fuese estacionario y, como  $cov(X_t, X_{t+k}) = cov(\dot{X}_t, \dot{X}_{t+k})$ , esta condición también me asegura que  $\{X_t : t \in \mathbb{Z}\}$  es estacionario, que es equivalente a la condición de estacionalidad vista en el modelo AR(p) con  $p=1$ .

**Ejemplo 2.2.1:** Analicemos el modelo  $X_t = 0,8X_{t-1} + A_t$  autorregresivo de orden 1. En este caso,  $\phi_1 = 0,8$  y como  $|\phi_1| < 1$ , luego este proceso es estacionario. Además:

1.  $\mu = E[X_t] = 0$ ,
2.  $\gamma_k = cov(X_t, X_{t+k}) = \sigma_A^2 \frac{(0,8)^k}{1-0,64}$  para todo  $k \in \mathbb{N} \cup \{0\}$ ,
3.  $\rho_k = (0,8)^k$  para todo  $k \in \mathbb{N} \cup \{0\}$  (ver Figura 2.1 (a)),
4.  $\pi_k = \phi_{kk} = \begin{cases} 0,8 & \text{si } k = 1 \\ 0 & \text{si } k > 1 \end{cases}$  (ver Figura 2.1 (b)).



(a) Correlograma teórico



(b) Función de autocorrelación parcial

Figura 2.1: Funciones teóricas de autocorrelación y autocorrelación parcial del proceso autorregresivo  $X_t = 0,8X_{t-1} + A_t$ .

Con el objeto de visualizar como serían las observaciones generadas por este modelo, simulamos 200 observaciones utilizando la fórmula  $X_t = 0,8X_{t-1} + A_t$  partiendo

de un valor inicial de  $X_0$  y generando los valores de  $A_1, A_2, \dots, A_{200}$  de una distribución  $N(0, 1)$  en forma independiente. Para realizar dicha simulación utilizamos la función `arima.sim` del programa R. Los datos  $x_1, x_2, \dots, x_{200}$  así obtenidos pueden verse en la Figura 2.2. Las funciones de autocorrelación y autocorrelación parcial muestral se muestran en la Figura 2.3, que es la representación gráfica de los valores del cuadro 2.1; obtenidos a partir de las funciones `acf` y `pacf` del programa R. El comportamiento de estos datos es más evidente cuando se grafican los pares de puntos  $(x_{t-1}, x_t)$  para  $t \in \{2, \dots, 200\}$  en un sistema de coordenadas cartesianas. La nube de puntos obtenida de este modo, mostrará una tendencia a concentrarse alrededor de la recta  $x_t = 0,758x_{t-1}$ , pues la estimación de la función de autocorrelación teórica obtenida a partir de los datos simulados satisface que  $\hat{\rho}_1 = 0,758$  (esto puede observarse en la Figura 2.4). Si en cambio miramos el gráfico de los pares de puntos  $(x_{t-2}, x_t)$ , podremos ver que, aunque con menor intensidad, se concentran alrededor de la recta  $X_t = 0,602X_{t-2}$ , ya que el coeficiente de correlación entre  $X_{t-2}$  y  $X_t$  es  $\hat{\rho}_2 = 0,602$  (ver Figura 2.5). Para comprobar que la autocorrelación parcial  $\pi_2 = 0$ , es conveniente graficar los pares de puntos  $(y_t, z_t)$ , donde  $y_t = x_t - 0,758x_{t+1}$  y  $z_t = x_{t+2} - 0,758x_{t+1}$ , y ver que la nube de puntos efectivamente no marca ninguna tendencia a concentrarse alrededor de una recta. (ver Figura 2.6).



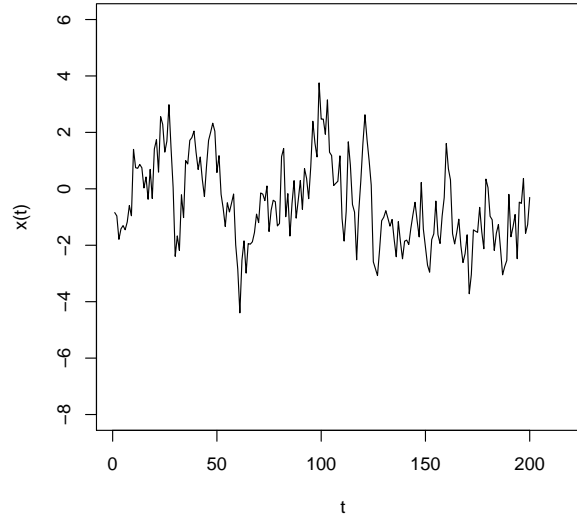
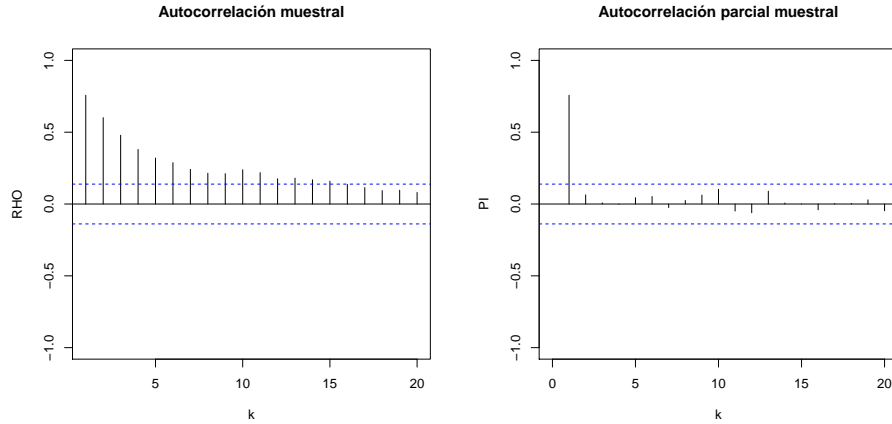


Figura 2.2: Doscientos datos simulados con el modelo  $X_t = 0,8X_{t-1} + A_t$ .

$k$	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	0.758	0.602	0.479	0.380	0.320	0.288	0.241	0.214	0.212	0.239
$\hat{\pi}_k$	0.758	0.064	0.008	-0.002	0.044	0.052	-0.025	0.025	0.062	0.103
$k$	11	12	13	14	15	16	17	18	19	20
$\hat{\rho}_k$	0.219	0.176	0.180	0.169	0.159	0.136	0.114	0.093	0.096	0.081
$\hat{\pi}_k$	-0.049	-0.061	0.090	0.007	0.001	-0.040	0.004	0.004	0.029	-0.04

Cuadro 2.1: Valores de las funciones muestrales de autocorrelación y autocorrelación parcial obtenidos a partir de la simulación de los 200 datos del modelo  $X_t = 0,8X_{t-1} + A_t$ .



(a) Función de autocorrelación muestral (b) Función de autocorrelación parcial muestral

Figura 2.3: Funciones muestrales de autocorrelación y autocorrelación parcial obtenidas a partir de los 200 datos generados por la simulación del modelo  $X_t = 0,8X_{t-1} + A_t$ .

### 2.3. Modelos de promedios móviles de orden finito.

**Definición 2.3.1:** Se dice que  $\{X_t : t \in \mathbb{Z}\}$  es un **proceso de promedios móviles de orden  $q$** , con  $q \in \mathbb{N}$  fijo, si  $\dot{X}_t = X_t - \mu$  puede expresarse como:

$$\dot{X}_t = A_t + \theta_1 A_{t-1} + \theta_2 A_{t-2} + \dots + \theta_q A_{t-q} \text{ para } t \in \mathbb{Z}$$

o análogamente, utilizando el operador  $B$  de cambio hacia atrás, bajo la forma  $\dot{X}_t = \theta(B)A_t$  con  $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$  para todo  $t \in \mathbb{Z}$ ; donde  $\theta_q \neq 0$  y donde  $\{A_t : t \in \mathbb{Z}\}$  es un proceso de ruido blanco con  $var(A_t) = \sigma_A^2 < \infty$ . Denotaremos a este proceso como **MA( $q$ )**.

**Proposición 2.3.1:** Sea  $\{X_t : t \in \mathbb{Z}\}$  un proceso MA( $q$ ) dado por  $\dot{X}_t = \theta(B)A_t$  con  $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$ , luego dicho proceso es estacionario. Mas aún,  $\{X_t : t \in \mathbb{Z}\}$  es invertible si las raíces de  $\theta(B)$  son de módulo mayor a 1.

**Demostración:** Notemos que este modelo MA( $q$ ), es un caso particular del modelo de promedios móviles de orden infinito  $\dot{X}_t = \sum_{i=0}^{\infty} \alpha_i A_{t-i}$  con  $\alpha_0 = 1$ , donde

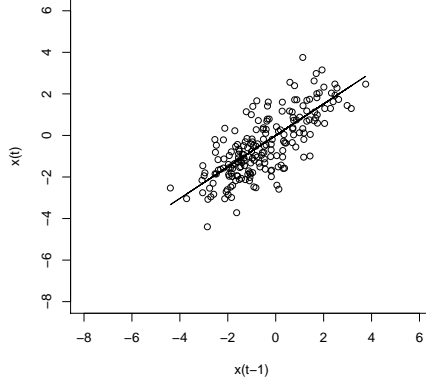


Figura 2.4: Nube de puntos  $(x_{t-1}, x_t)$  de 200 datos simulados a partir del modelo  $X_t = 0,8X_{t-1} + A_t$ . La recta dibujada corresponde a la ecuación  $X_t = 0,758X_{t-1}$

$\alpha_1 = \theta_1, \alpha_2 = \theta_2, \dots, \alpha_q = \theta_q$  y  $\alpha_m = 0$  para todo  $m > q$ . Como este último modelo es estacionario si  $\sum_{i=0}^{\infty} \alpha_i^2 < \infty$  y, en este caso,  $\sum_{i=0}^{\infty} \alpha_i^2 = 1 + \theta_1^2 + \dots + \theta_q^2 < \infty$ , luego  $\{X_t : t \in \mathbb{Z}\}$  es estacionario. Por último, como el modelo de promedio móviles de orden infinito es invertible si las raíces de  $\theta(B)$  son de módulo mayor a 1, queda probado que, en dicho caso,  $\{X_t : t \in \mathbb{Z}\}$  es invertible. ■

Como el modelo MA(q) es un caso particular del modelo de promedios móviles de orden infinito  $\dot{X}_t = \sum_{i=0}^{\infty} \alpha_i A_{t-i}$  con  $\alpha_0 = 1$ , donde  $\alpha_1 = \theta_1, \alpha_2 = \theta_2, \dots, \alpha_q = \theta_q$  y  $\alpha_m = 0$  para todo  $m > q$ , como ya lo dijimos en la demostración de la proposición anterior, este modelo satisface las siguientes propiedades, que surgen a partir de la demostración de la Proposición 2.1.2 para modelos de promedios móviles de orden infinito:

1.  $E[X_t] = \mu$
2.  $var(X_t) = \sigma_A^2 \sum_{i=0}^{\infty} \alpha_i^2 = \sigma_A^2(1 + \theta_1^2 + \dots + \theta_q^2)$
3.  $\gamma_k = cov(X_t, X_{t+k}) = \sigma_A^2 \sum_{i=0}^{\infty} \alpha_i \alpha_{i+k} = \begin{cases} \sigma_A^2(\theta_k + \theta_1\theta_{k+1} + \dots + \theta_{q-k}\theta_q) & si \quad k \leq q \\ 0 & si \quad k > q \end{cases}$
4.  $\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} \frac{\theta_k + \theta_1\theta_{k+1} + \dots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \dots + \theta_q^2} & si \quad 0 \leq k \leq q \\ 0 & si \quad k > q \end{cases}$

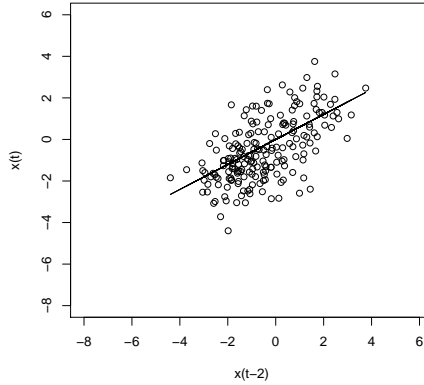


Figura 2.5: Nube de puntos  $(x_{t-2}, x_t)$  de 200 datos simulados a partir del modelo  $X_t = 0,8X_{t-1} + A_t$ . La recta dibujada corresponde a la ecuación  $X_t = 0,602X_{t-2}$ .

Podemos notar que la función de autocovarianza teórica y, por lo tanto, la función de autocorrelación teórica de un proceso MA(q), tiene la propiedad de que  $\gamma_k$  se anula para  $k > q$ . Mas aún, esta propiedad caracteriza a los procesos MA(q). En efecto, puede probarse que si  $\{X_t : t \in \mathbb{Z}\}$  es un proceso estacionario con media 0 y función de autocovarianza  $\gamma$  tal que  $\gamma_q \neq 0$  y  $\gamma_k = 0$  para todo  $k > q$ , entonces  $\{X_t : t \in \mathbb{Z}\}$  es un proceso MA(q), es decir, existen un proceso de ruido blanco  $\{A_t : t \in \mathbb{Z}\}$  y constantes  $\theta_1, \dots, \theta_q$  con  $\theta_q \neq 0$  tales que  $X_t = \sum_{i=0}^q \theta_i A_{t-i}$  para todo  $t \in \mathbb{Z}$ .

### 2.3.1. Modelo de promedios móviles de orden 1.

**Definición 2.3.2:** Una serie de tiempo  $\{X_t : t \in \mathbb{Z}\}$  es un **proceso de promedios móviles de orden 1** si  $\dot{X}_t = X_t - \mu$  se puede expresar bajo la forma

$$\dot{X}_t = \theta_1 A_{t-1} + A_t,$$

donde  $\theta_1$  es una constante y  $\{A_t : t \in \mathbb{Z}\}$  es un proceso de ruido blanco con  $var(A_t) = \sigma_A^2 < \infty$ . Denotaremos a este modelo por **MA(1)**.

Notemos que si en el ejemplo 1.3.1 reemplazamos a  $X_t$  por  $\dot{X}_t$ , las mismas características que habíamos probado para  $\{X_t : t \in \mathbb{Z}\}$  valen ahora para  $\{\dot{X}_t : t \in \mathbb{Z}\}$ . Gracias a esto, podemos afirmar que:

1.  $E[X_t] = \mu$  para todo  $t \in \mathbb{Z}$ , en efecto  $E[\dot{X}_t] = 0$  para todo  $t \in \mathbb{Z}$ .

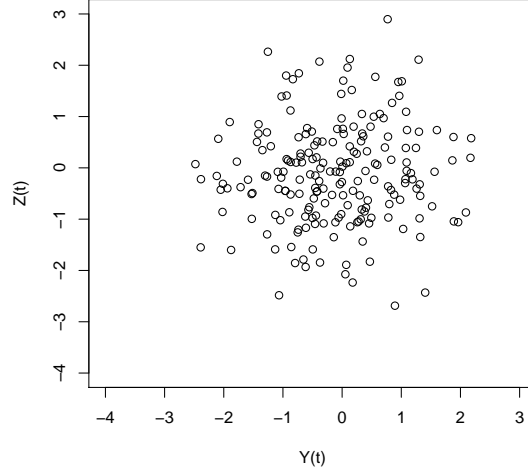


Figura 2.6: Nube de puntos  $(y_t, z_t)$ , donde  $y_t = x_t - 0,758x_{t+1}$  y  $z_t = x_{t+2} - 0,758x_{t+1}$ , para los 200 datos generados por el modelo  $X_t = 0,8X_{t-1} + A_t$ .

$$2. \gamma_k = cov(X_t, X_{t+k}) = E[(X_t - \mu)(X_{t+k} - \mu)] = E[\dot{X}_t \cdot \dot{X}_{t+k}] = cov(\dot{X}_t, \dot{X}_{t+k}) = \begin{cases} (1 + \theta_1^2)\sigma_A^2 & si \quad k = 0 \\ \theta_1\sigma_A^2 & si \quad |k| = 1 \\ 0 & si \quad |k| > 1 \end{cases}$$

$$3. \rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} 1 & si \quad k = 0 \\ \frac{\theta_1}{1+\theta_1^2} & si \quad |k| = 1 \\ 0 & si \quad |k| > 1 \end{cases} .$$

Lo que no habíamos analizado en el ejemplo 1.3.1 era la función de autocorrelación parcial teórica. En general, se puede demostrar que  $\pi_k = \frac{-(-\theta_1)^k}{1+\theta_1^2+\dots+\theta_1^{2k}}$  si  $k \geq 1$ . En este trabajo lo demostraremos únicamente para los casos 1 y 2:

- Caso k=1:  $\pi_1 = \rho_1 = \frac{\theta_1}{1+\theta_1^2}$ .

- Caso k=2:  $\pi_2 = \frac{\begin{vmatrix} 1 & \frac{\theta_1}{1+\theta_1^2} \\ \frac{\theta_1}{1+\theta_1^2} & 0 \end{vmatrix}}{\begin{vmatrix} 1 & \frac{\theta_1}{1+\theta_1^2} \\ \frac{\theta_1}{1+\theta_1^2} & 1 \end{vmatrix}} = \frac{-\left(\frac{\theta_1}{1+\theta_1^2}\right)^2}{1 - \left(\frac{\theta_1}{1+\theta_1^2}\right)^2} = \frac{-1}{\frac{(1+\theta_1^2)^2}{\theta_1^2} - 1} = \frac{-\theta_1^2}{1+2\theta_1^2+\theta_1^4-\theta_1^2} = \frac{-\theta_1^2}{1+\theta_1^2+\theta_1^4}$ .

En la Figura 2.7 se muestra el gráfico de la función de autocorrelación parcial teórica de  $\dot{X}_t = \theta_1 A_{t-1} + A_t$  para un  $\theta_1 > 0$  y un  $\theta_1 < 0$ .

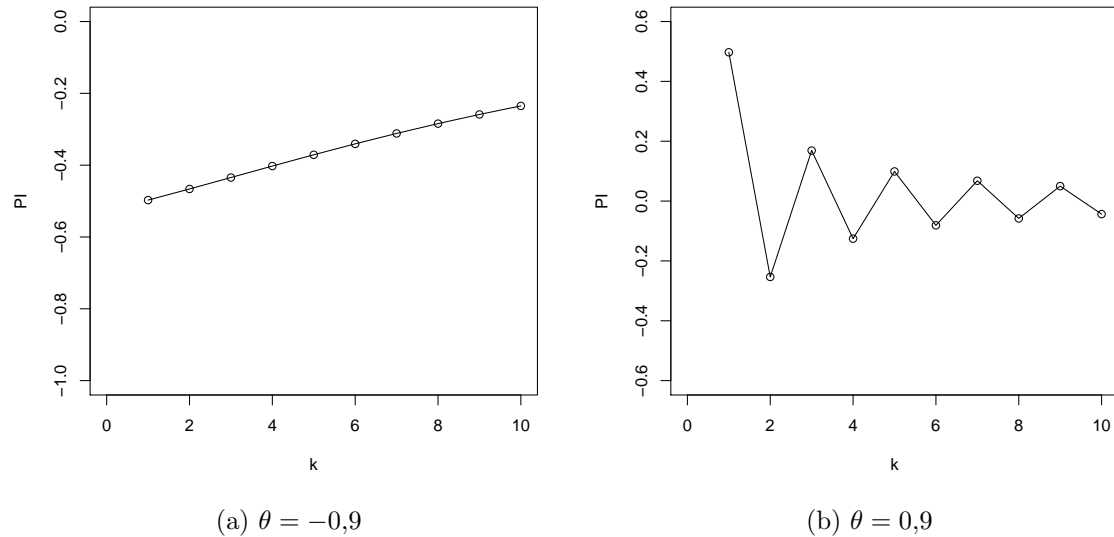


Figura 2.7: Funciones teóricas de autocorrelación parcial del proceso de promedios móviles dado por  $\hat{X}_t = \theta A_{t-1} + A_t$  para un  $\theta < 0$  y un  $\theta > 0$ .

**Ejemplo 2.3.1:** Como ejemplo de un proceso MA(1), y para ver como se comportarían los datos generados por este modelo, simulamos 200 datos del proceso dado por  $X_t = A_t - 0,9A_{t-1}$ , donde los valores de  $A_t$  fueron generados en forma independiente de una distribución  $N(0,1)$ . Estos datos están representados en la Figura 2.8. Este proceso tiene como función de autocorrelación teórica a  $\rho_k = \begin{cases} \frac{-0,9}{1+(-0,9)^2} = -0,497 & \text{si } |k| = 1 \\ 0 & \text{si } |k| > 1 \end{cases}$  y como función de autocorrelación parcial teórica a  $\pi_k = \frac{-(-0,9)^k}{1+(-0,9)^2+\dots+(-0,9)^{2k}}$ .

En el cuadro 2.2, se muestran algunos valores de las funciones muestrales de autocorrelación y autocorrelación parcial generados por la simulación del modelo  $X_t = A_t - 0,9A_{t-1}$ , y sus gráficos están representados en la Figura 2.9.

La correlación negativa entre observaciones adyacentes en el tiempo, indicadas por el valor  $\rho_1 = -0,497$ , puede observarse en la Figura 2.10 donde se han representado los puntos  $(x_{t-1}, x_t)$  tal que  $t = 2, \dots, 200$ , junto con la recta de ecuación  $x_t = -0,459x_{t-1}$ , donde -0.459 es la estimación obtenida con los datos del valor teórico  $\rho_1 = -0,497$ . Esta recta claramente indica la tendencia de la nube de puntos.

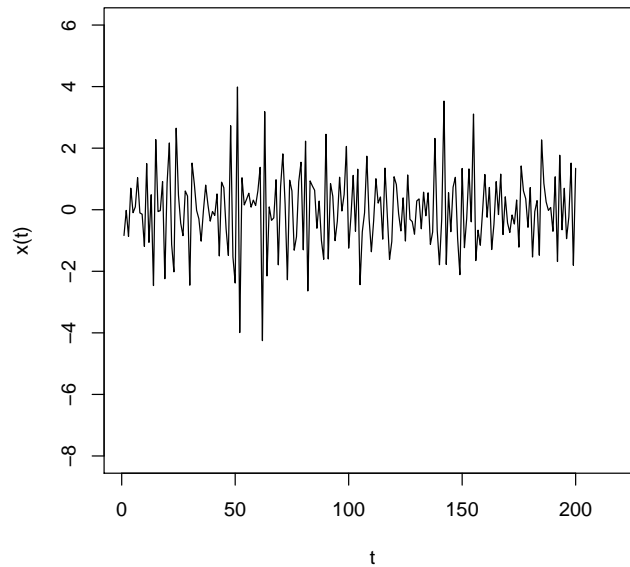


Figura 2.8: Doscientos datos generados por el proceso MA(1) de ecuación  $X_t = A_t - 0,9A_{t-1}$

$k$	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	-0.459	-0.045	0.040	-0.013	0.034	-0.099	0.059	-0.062	0.114	-0.086
$\hat{\pi}_k$	-0.459	-0.323	-0.191	-0.138	-0.046	-0.147	-0.094	-0.169	-0.006	-0.068
$k$	11	12	13	14	15	16	17	18	19	20
$\hat{\rho}_k$	-0.042	0.086	0.047	-0.076	0.023	-0.031	0.016	0.064	-0.118	0.077
$\hat{\pi}_k$	-0.140	-0.074	0.060	-0.001	0.037	-0.041	-0.032	0.071	-0.031	0.028

Cuadro 2.2: Valores de las funciones muestrales de autocorrelación y autocorrelación parcial obtenidos a partir de la simulación de los 200 datos del modelo  $X_t = A_t - 0,9A_{t-1}$ .



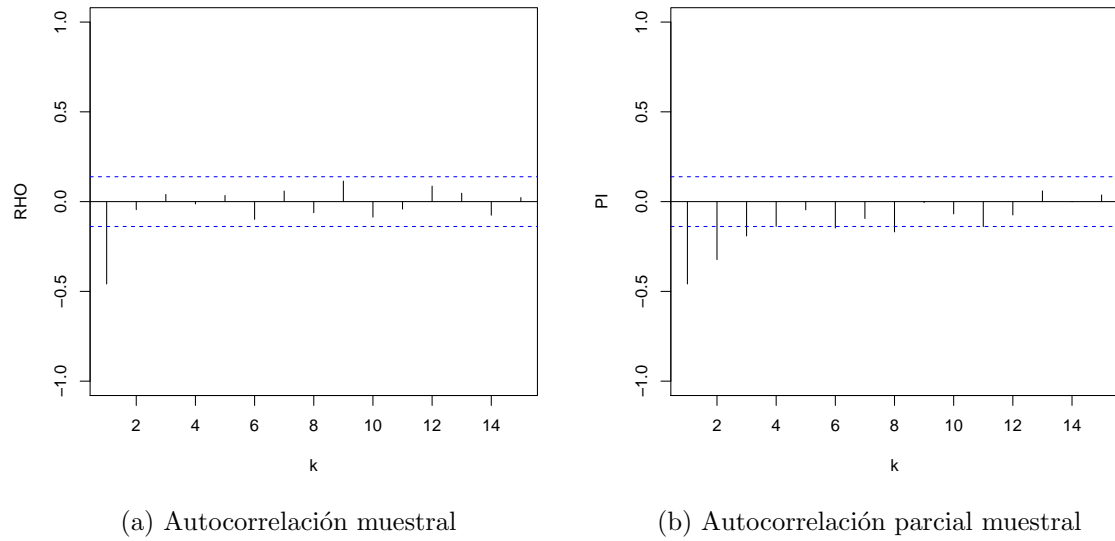


Figura 2.9: Funciones muestrales de autocorrelación y autocorrelación parcial del proceso de promedios móviles dado por  $X_t = -0,9A_{t-1} + A_t$ .

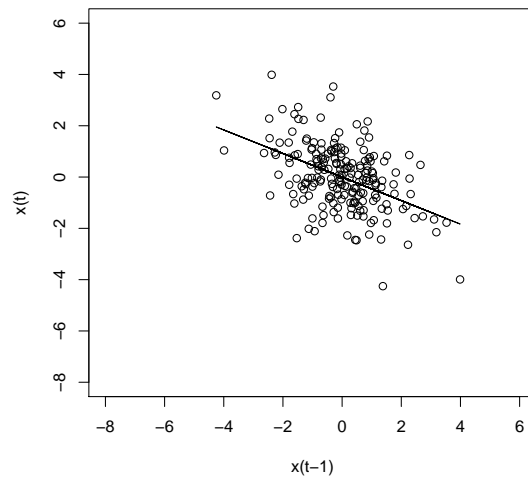


Figura 2.10: Nube de puntos  $(x_{t-1}, x_t)$  correspondiente a la simulación de 200 datos del modelo  $X_t = A_t - 0,9A_{t-1}$ . La recta dibujada corresponde a la ecuación  $x_t = -0,459x_{t-1}$

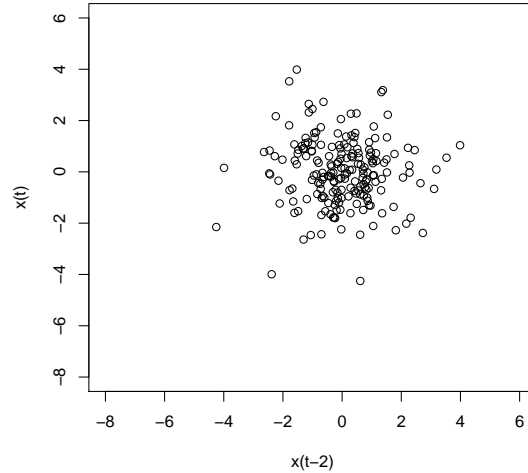


Figura 2.11: Nube de puntos  $(x_{t-2}, x_t)$  correspondiente a la simulación de 200 datos del modelo  $X_t = A_t - 0,9A_{t-1}$ .

De la misma forma, el valor  $\rho_2 = 0$  es claro al observarse la Figura 2.11, en donde representamos los puntos  $(x_{t-2}, x_t)$  tal que  $t = 3, \dots, 200$ . Esta nube de puntos, no muestra ninguna tendencia.

Ahora si vemos el valor  $\hat{\pi}_2$  en el cuadro 2.2, la estimación de  $\pi_2$  obtenida a partir de los datos, este valor resulta ser  $\hat{\pi}_2 = -0,323$ . Para representar este valor y su significado, graficamos la nube de puntos  $(y_t, z_t)$  donde  $y_t = x_t + 0,459x_{t+1}$  y  $z_t = x_{t+2} + 0,459x_{t+1}$ , y la recta  $z_t = -0,323y_t$ , donde -0.459 es el valor de  $\hat{\rho}_1$ , lo cual puede verse en la Figura 2.12.

## 2.4. Modelos Autorregresivos de Promedios Móviles.

**Definición 2.4.1:** Se dice que una serie de tiempo  $\{X_t : t \in \mathbb{Z}\}$  es un **proceso autorregresivo de promedios móviles de órdenes p y q** respectivamente, y se lo denota como **ARMA(p,q)**, si se verifica que  $\dot{X}_t = X_t - \mu$  se puede expresar como

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) \dot{X}_t = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q) A_t,$$

o, análogamente, si se puede expresar como

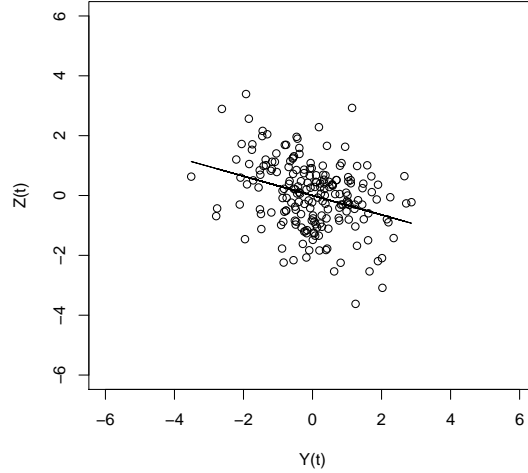


Figura 2.12: Nube de puntos  $(y_t, z_t)$  donde  $y_t = x_t + 0,459x_{t+1}$  y  $z_t = x_{t+2} + 0,459x_{t+1}$ , correspondiente a la simulación de 200 datos del modelo  $X_t = A_t - 0,9A_{t-1}$ . La recta corresponde a la ecuación  $z_t = -0,323y_t$

$$\Phi(B)\dot{X}_t = \theta(B)A_t, \text{ con } \Phi(B) = 1 - \phi_1B - \phi_2B^2 - \dots - \phi_pB^p \text{ y}$$

$$\theta(B) = 1 + \theta_1B + \theta_2B^2 + \dots + \theta_qB^q,$$

donde  $\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q$  son constantes con  $\phi_p \neq 0$  y  $\theta_q \neq 0$ , B es el operador de cambio hacia atrás y  $\{A_t : t \in \mathbb{Z}\}$  es un proceso de ruido blanco con varianza  $\sigma_A^2$ .

Es claro que el modelo ARMA(p,q) se reduce al modelo AR(p) cuando  $\theta_1 = \theta_2 = \dots = \theta_q = 0$ , o, dicho de otro modo, cuando  $q = 0$ ; y se reduce al modelo MA(q) cuando  $\phi_1 = \phi_2 = \dots = \phi_p = 0$ , o, de otro modo, cuando  $p = 0$ .

Una condición suficiente para que un tal proceso sea estacionario es que las raíces de  $\Phi(B)$  sean todas de módulo mayor que uno, pues en dicho caso, se puede demostrar que podemos escribir al proceso de la forma:

$$\dot{X}_t = \frac{\theta(B)}{\Phi(B)}A_t = \alpha(B)A_t \text{ siendo } \alpha(B) = 1 + \alpha_1B + \alpha_2B^2 + \dots,$$

donde los coeficientes  $\alpha_i$  con  $i \in \mathbb{N}$  son funciones de los coeficientes  $\theta_1, \dots, \theta_q, \phi_1, \dots, \phi_p$  verificándose que  $\sum_{i=1}^{\infty} \alpha_i^2 < \infty$ ; y así representado, sabemos por la Proposición 2.1.2 que el proceso es estacionario.

Por otro lado, una condición necesaria para que el proceso  $\{X_t : t \in \mathbb{Z}\}$  sea invertible es que las raíces del polinomio  $\theta(B)$  sean todas de módulo mayor que uno, ya que, en dicho caso, puede demostrarse que se puede representar al proceso de la forma:

$$A_t = \frac{\Phi(B)}{\theta(B)} \dot{X}_t = \beta(B) \dot{X}_t \text{ con } \beta(B) = 1 - \beta_1 B - \beta_2 B^2 - \dots,$$

donde los coeficientes  $\beta_t$  con  $t \in \mathbb{N}$  son funciones de los coeficientes  $\theta_1, \dots, \theta_q, \phi_1, \dots, \phi_p$ , verificándose que  $\sum_{i=1}^{\infty} |\beta_i| < \infty$ .

**Definición 2.4.2:** Sea  $\{X_t : t \in \mathbb{Z}\}$  un proceso **ARMA(p,q)** dado por  $\Phi(B)\dot{X}_t = \theta(B)A_t$ , con  $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ ,  $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$  y  $\dot{X}_t = X_t - \mu$ . Luego si las raíces de  $\Phi(B)$  son de módulo mayor a 1, dicho proceso se llama **estacionario causal**.

### 2.4.1. Modelo ARMA(1,1)

**Definición 2.4.3:** Se dice que una serie de tiempo  $\{X_t : t \in \mathbb{Z}\}$  es un **proceso autorregresivo de promedios móviles de órdenes p=1 y q=1** respectivamente, y se lo denota como **ARMA(1,1)**, si se verifica que  $\dot{X}_t = X_t - \mu$  se puede expresar como:

$$\Phi(B)\dot{X}_t = \theta(B)A_t, \text{ con } \Phi(B) = 1 - \phi_1 B \text{ y } \theta(B) = 1 + \theta_1 B,$$

con  $\theta_1$  y  $\phi_1$  constantes no nulas.

Por ser este proceso un caso particular del proceso ARMA(p,q), luego es estacionario (causal) cuando la raíz  $\frac{1}{\phi_1}$  de  $\Phi(B)$  es de módulo mayor a uno, es decir, cuando  $|\phi_1| < 1$ ; y es invertible cuando la raíz  $-\frac{1}{\theta_1}$  de  $\theta(B)$  es de módulo mayor a 1, es decir, cuando  $|\theta_1| < 1$ .

**Proposición 2.4.1:** Sea  $\{X_t : t \in \mathbb{Z}\}$  un proceso ARMA(1,1) estacionario (causal) dado por  $\Phi(B)\dot{X}_t = \theta(B)A_t$ , con  $\Phi(B) = 1 - \phi_1 B$  y  $\theta(B) = 1 + \theta_1 B$ . Luego  $\{X_t : t \in \mathbb{Z}\}$  se puede representar como un proceso de promedios móviles de orden infinito de la forma  $\dot{X}_t = \alpha(B)A_t$ .

**Demostración:** Es importante tener en cuenta para esta demostración que es válida la igualdad  $\frac{1}{1-\phi_1 B} = \sum_{i=0}^{\infty} \phi_1^i B^i$ , que surge del hecho de que la norma del operador retroceso es 1 en el espacio  $\mathcal{B}(\overline{\langle \{X_t : t \in \mathbb{Z}\} \rangle})$  y del hecho de que  $|\phi_1| < 1$  por ser

$\{X_t : t \in \mathbb{Z}\}$  estacionario causal. Luego de la fórmula  $\Phi(B)\dot{X}_t = \theta(B)A_t$ , se deduce que:

$$\begin{aligned}\dot{X}_t &= \frac{\theta(B)}{\Phi(B)}A_t \\ &= (1 + \theta_1 B) \frac{1}{1 - \phi_1 B} A_t \\ &= (1 + \theta_1 B) \left( \sum_{i=0}^{\infty} \phi_1^i B^i \right) A_t \\ &= \left( \sum_{i=0}^{\infty} \phi_1^i B^i + \sum_{i=0}^{\infty} \theta_1 \phi_1^i B^{i+1} \right) A_t \\ &= \left( 1 + \sum_{i=1}^{\infty} (\phi_1^i + \theta_1 \phi_1^{i-1}) B^i \right) A_t,\end{aligned}$$

es decir,

$$\dot{X}_t = \left( 1 + \sum_{i=1}^{\infty} ((\phi_1 + \theta_1) \phi_1^{i-1}) B^i \right) A_t,$$

de donde se deduce que  $\alpha_0 = 1$  y  $\alpha_i = (\phi_1 + \theta_1) \phi_1^{i-1}$  para  $i \geq 1$ . ■

Análogamente, puede probarse que si  $\{X_t : t \in \mathbb{Z}\}$  es un proceso ARMA(1,1) invertible dado por  $\Phi(B)\dot{X}_t = \theta(B)A_t$ , con  $\Phi(B) = 1 - \phi_1 B$  y  $\theta(B) = 1 + \theta_1 B$  con  $|\theta_1| < 1$ , luego  $\{X_t : t \in \mathbb{Z}\}$  se puede representar como un proceso autorregresivo de orden infinito de la forma  $\beta(B)\dot{X}_t = A_t$  donde  $\beta(B) = 1 + \sum_{i=1}^{\infty} \beta_i B^i = \frac{\Phi(B)}{\theta(B)}$ , usando la igualdad  $\frac{1}{1+\theta_1 B} = \sum_{i=0}^{\infty} (-\theta_1)^i B^i$ . Mas aún, resulta que  $\beta_i = \theta_1^{i-1}(\phi_1 + \theta_1)(-1)^i$  para  $i \geq 1$ .

Suponiendo ahora que  $\{X_t : t \in \mathbb{Z}\}$  es un proceso ARMA(1,1) estacionario causal, podemos calcular la autocovarianza  $\gamma_k$  del proceso, representando a éste último como un promedio móvil de orden infinito

$$\dot{X}_t = \left( \sum_{i=0}^{\infty} \alpha_i B^i \right) A_t,$$

donde  $\alpha_0 = 1$  y  $\alpha_i = (\phi_1 + \theta_1) \phi_1^{i-1}$  si  $i > 0$ , recordando que para estos procesos si

$k > 0$  se verifica que

$$\begin{aligned}
\gamma_k &= \sigma_A^2 \sum_{i=0}^{\infty} \alpha_i \alpha_{i+k} \\
&= \sigma_A^2 \left( \alpha_k + \sum_{i=1}^{\infty} \alpha_i \alpha_{i+k} \right) \\
&= \sigma_A^2 \left( (\phi_1 + \theta_1) \phi_1^{k-1} + \sum_{i=1}^{\infty} (\phi_1 + \theta_1)^2 \phi_1^{i-1} \cdot \phi_1^{i+k-1} \right) \\
&= \sigma_A^2 \left( (\phi_1 + \theta_1) \phi_1^{k-1} + (\phi_1 + \theta_1)^2 \phi_1^{k-2} \sum_{i=1}^{\infty} (\phi_1^2)^i \right) \\
&= \sigma_A^2 \left( (\phi_1 + \theta_1) \phi_1^{k-1} + (\phi_1 + \theta_1)^2 \phi_1^{k-2} \cdot \frac{\phi_1^2}{1 - \phi_1^2} \right) \\
&= \sigma_A^2 (\phi_1 + \theta_1) \phi_1^{k-1} \left[ \frac{(1 - \phi_1^2) + (\phi_1 + \theta_1) \phi_1}{1 - \phi_1^2} \right] \\
&= \sigma_A^2 \frac{(\phi_1 + \theta_1) \phi_1^{k-1} (1 + \phi_1 \theta_1)}{1 - \phi_1^2}
\end{aligned}$$

y si  $k = 0$  se verifica que

$$\begin{aligned}
\gamma_0 &= \text{var}(\dot{X}_t) \\
&= \sigma_A^2 \sum_{i=0}^{\infty} \alpha_i^2 \\
&= \sigma_A^2 \left( 1 + \sum_{i=1}^{\infty} (\phi_1 + \theta_1)^2 \phi_1^{2(i-1)} \right) \\
&= \sigma_A^2 \left( 1 + (\phi_1 + \theta_1)^2 \cdot \frac{1}{1 - \phi_1^2} \right) \\
&= \sigma_A^2 \left( \frac{1 - \phi_1^2 + (\phi_1 + \theta_1)^2}{1 - \phi_1^2} \right) \\
&= \sigma_A^2 \left( \frac{1 + 2\theta_1 \phi_1 + \theta_1^2}{1 - \phi_1^2} \right)
\end{aligned}$$

Es decir,

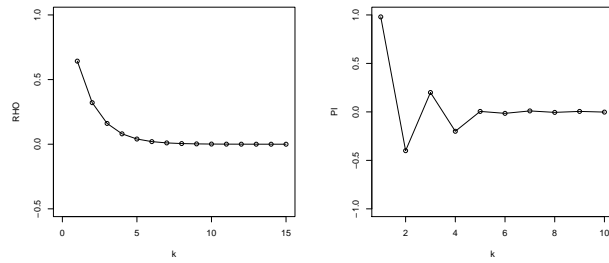
$$\gamma_k = \begin{cases} \sigma_A^2 \left( \frac{1+\theta_1^2+2\phi_1\theta_1}{1-\phi_1^2} \right) & \text{si } k = 0 \\ \sigma_A^2 \left( \frac{(\phi_1+\theta_1)\phi_1^{k-1}(1+\phi_1\theta_1)}{1-\phi_1^2} \right) & \text{si } k \geq 1 \end{cases}$$

En consecuencia, las autocorrelaciones teóricas  $\rho_k$  están dadas por

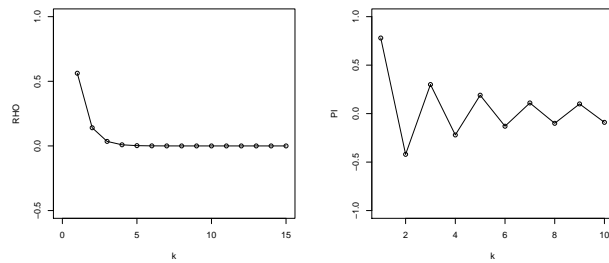
$$\rho_k = \frac{(\phi_1+\theta_1)\phi_1^{k-1}(1+\phi_1\theta_1)}{1+\theta_1^2+2\phi_1\theta_1} \text{ para } k = 1, 2, \dots$$

Notar que  $\rho_k = \phi_1 \rho_{k-1} = \phi_1^{k-1} \cdot \rho_1$  donde  $\rho_1 = \frac{(\phi_1+\theta_1)(1+\phi_1\theta_1)}{1+\theta_1^2+2\phi_1\theta_1} = \frac{(\phi_1+\theta_1)(1+\phi_1\theta_1)}{(1-\phi_1^2)+(\phi_1+\theta_1)^2}$ , donde se advierte que la función de autocorrelación de un modelo ARMA(1,1) tiene un comportamiento similar al de un proceso AR(1), sólo que el valor inicial  $\rho_1$  depende de  $\phi_1$  y de  $\theta_1$ .

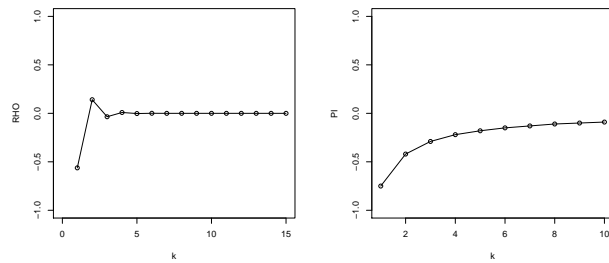
Por otro lado, la función de autocorrelación parcial teórica de un proceso ARMA(1,1) es complicada. Los valores de  $|\pi_k|$  decrecen en forma exponencial cuando  $k \rightarrow \infty$  (en forma similar a lo que ocurre con  $\rho_k$ ) dependiendo del signo y magnitud de  $\phi_1$  y  $\theta_1$ . La Figura 2.13 muestra las funciones teóricas de autocorrelación y de autocorrelación parcial de un proceso ARMA(1,1) de ecuación  $\dot{X}_t - \phi_1 \dot{X}_{t-1} = A_t + \theta_1 A_{t-1}$  para distintos valores de  $\phi_1$  y  $\theta_1$ .



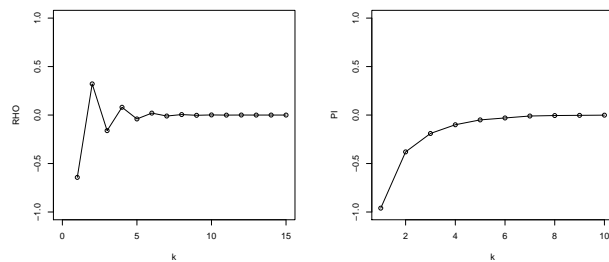
(a) Caso  $\phi_1 > \theta_1 > 0$



(b) Caso  $\theta_1 > \phi_1 > 0$

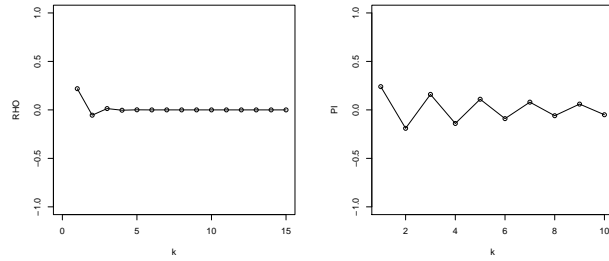


(c)  $\theta_1 < \phi_1 < 0$

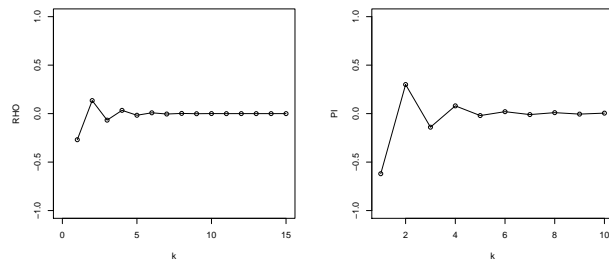


(d) Caso  $\phi_1 < \theta_1 < 0$

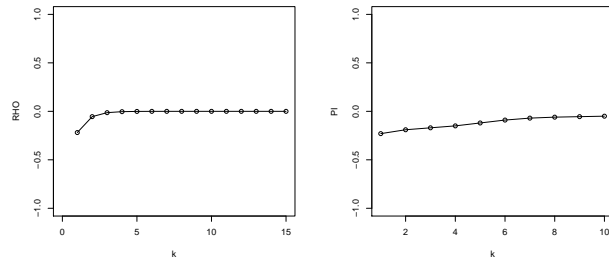




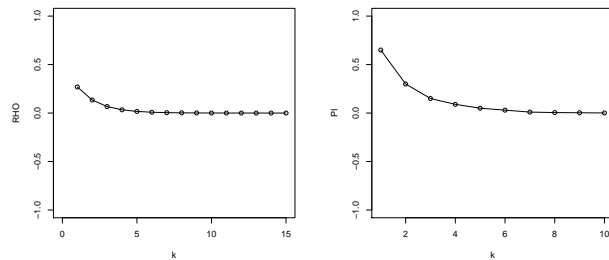
(e) Caso  $\phi_1 < 0 < \theta_1$  y  $\phi_1 + \theta_1 > 0$



(f) Caso  $\phi_1 < 0 < \theta_1$  y  $\phi_1 + \theta_1 < 0$



(g) Caso  $\theta_1 < 0 < \phi_1$  y  $\phi_1 + \theta_1 < 0$



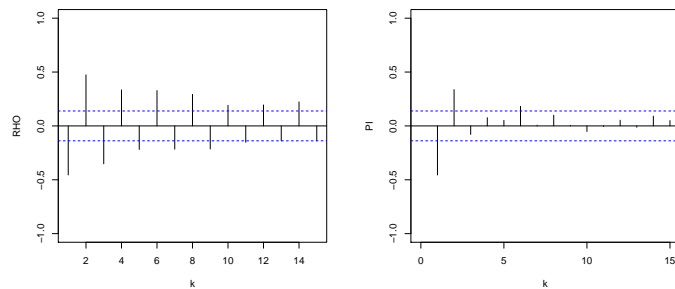
(h) Caso  $\theta_1 < 0 < \phi_1$  y  $\phi_1 + \theta_1 > 0$

Figura 2.13: Funciones teóricas de autocorrelación y autocorrelación parcial del proceso ARMA(1,1) de ecuación  $\dot{X}_t - \phi_1 \dot{X}_{t-1} = A_t + \theta_1 A_{t-1}$

**Ejemplo 2.4.1:** Consideremos el modelo ARMA(1,1), dado por  $X_t + 0,9X_{t-1} = A_t + 0,6A_{t-1}$ . Luego en este caso  $\phi_1 = -0,9$  y  $\theta_1 = 0,6$ . Para ver que comportamiento tendrían las funciones de autocorrelación y autocorrelación parcial del modelo, simulamos 200 datos, donde los valores de  $A_t$  para  $t \in \{1, \dots, 200\}$  fueron generados de manera independiente de una variable con distribución  $N(0, 1)$ . En el cuadro 2.3 se presentan los primeros 20 valores muestrales de las funciones de autocorrelación y autocorrelación parcial, obtenidos de la simulación de los 200 datos; y en la Figura 2.14 están representados gráficamente estos valores.

$k$	1	2	3	4	5	6	7	8	9	10
$\widehat{\rho}_k$	-0.455	0.474	-0.352	0.335	-0.218	0.328	-0.216	0.293	-0.214	0.191
$\widehat{\pi}_k$	-0.455	0.337	-0.079	0.076	0.052	0.182	0.007	0.100	0.005	-0.052
$k$	11	12	13	14	15	16	17	18	19	20
$\widehat{\rho}_k$	-0.152	0.195	-0.138	0.223	-0.136	0.201	-0.139	0.121	-0.099	0.135
$\widehat{\pi}_k$	-0.005	0.053	-0.012	0.091	0.050	0.061	0.011	-0.051	-0.017	0.013

Cuadro 2.3: Valores de las funciones muestrales de autocorrelación y autocorrelación parcial obtenidos a partir de la simulación de los 200 datos del modelo  $X_t + 0,9X_{t-1} = A_t + 0,6A_{t-1}$ .



(a) Autocorrelación muestral (b) Autocorrelación parcial muestral

Figura 2.14: Autocorrelación y autocorrelación parcial muestrales del proceso de promedios móviles dado por  $X_t + 0,9X_{t-1} = A_t + 0,6A_{t-1}$ .

Es importante notar que si  $\phi_1 + \theta_1$  es cercano a cero, entonces los valores de

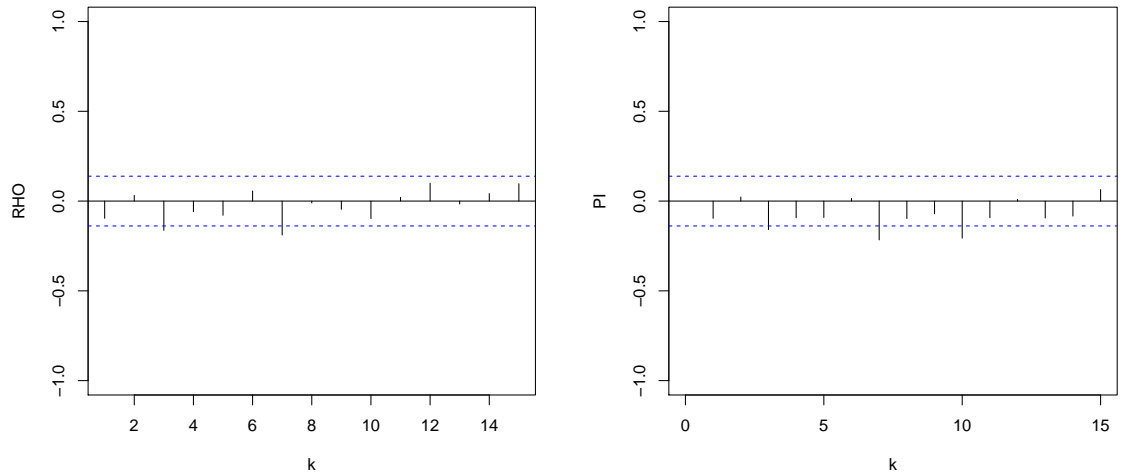
$$\rho_k = \frac{(\phi_1 + \theta_1)\phi_1^{k-1}(1 + \phi_1\theta_1)}{1 + \theta_1^2 + 2\phi_1\theta_1}$$

serán próximos a cero y, en consecuencia, las autocorrelaciones muestrales no se diferenciarán de las que corresponden a un proceso de ruido blanco. Esto es notable si se expresa al proceso ARMA(1,1) como  $\dot{X}_t = \frac{1+\theta_1 B}{1-\phi_1 B} A_t \cong A_t$  cuando  $\theta_1 \cong -\phi_1$ , o de otro modo, cuando  $\theta_1 + \phi_1 \cong 0$ . Veamos el siguiente ejemplo que servirá para ilustrar dicha situación.

**Ejemplo 2.4.2:** Consideremos el modelo ARMA(1,1) dado por  $X_t - 0,7X_{t-1} = A_t - 0,8A_{t-1}$ . Luego  $\phi = 0,7$  y  $\theta_1 = -0,8$ , de modo que,  $\theta_1 + \phi_1 = -0,1 \cong 0$ . Simulamos 200 datos de este modelo, donde los valores de  $A_t$  los generamos de forma independiente de una distribución  $N(0, 1)$ . En el cuadro 2.4 se presentan los primeros 20 valores muestrales de las funciones de autocorrelación y autocorrelación parcial, obtenidos de la simulación de los 200 datos; y en la Figura 2.15 estan representados graficamente estos valores.

$k$	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	-0.096	0.032	-0.163	-0.059	-0.079	0.056	-0.189	-0.010	-0.046	-0.097
$\hat{\pi}_k$	-0.096	0.022	-0.159	-0.093	-0.091	0.014	-0.216	-0.097	-0.071	-0.207
$k$	11	12	13	14	15	16	17	18	19	20
$\hat{\rho}_k$	0.021	0.100	-0.016	0.042	0.096	-0.045	0.036	-0.044	-0.053	0.077
$\hat{\pi}_k$	-0.093	0.010	-0.094	-0.084	0.064	-0.069	-0.056	-0.056	-0.071	0.030

Cuadro 2.4: Valores de las funciones muestrales de autocorrelación y autocorrelación parcial obtenidos a partir de la simulación de los 200 datos del modelo  $X_t - 0,7X_{t-1} = A_t - 0,8A_{t-1}$ .



(a) Autocorrelación muestral

(b) Autocorrelación parcial muestral

Figura 2.15: Autocorrelación y autocorrelación parciales muestrales del proceso autorregresivo de promedios móviles dado por  $X_t - 0,7X_{t-1} = A_t - 0,8A_{t-1}$ .

---

## Capítulo 3

# Comparación de series de tiempo

La búsqueda de un método que permita identificar el grado de similitud entre dos o mas series de tiempo ha sido abordado por múltiples autores en los últimos años [2, 4–6]. Ello se debe a que en la práctica, es frecuente enfrentar situaciones en las que se requiere comparar dos o mas series temporales, o analizar un gran número de estas series para separarlas en grupos tan homogéneos como sea posible. Normalmente, se generan bases de datos temporales que luego se estudian para identificar concordancias y disimilitudes. Así, desde la matemática y la estadística, es necesario desarrollar metodologías que permitan no solo analizar, sino también comparar, datos obtenidos a partir de la evolución de sistemas que se modifican en el tiempo.

El interés en una medida de similitud de series de tiempo surge en múltiples disciplinas, tales como la biología, criminalística, climatología, economía, etc. En estas áreas se suscitan diferentes problemáticas en las que resulta indispensable comparar series de tiempo. Solo a modo de ejemplo planteamos muy brevemente algunos de los problemas:

1. establecer zonas geográficas similares desde una perspectiva biológica, a partir del estudio de series de índices de vegetación, generadas por imágenes satelitales (biología);
2. comparar series de tiempo de las rentabilidades de las administradoras de fondos de pensiones, a fin de seleccionar aquellas con mayor rentabilidad a mediano y largo plazo (economía);
3. conseguir patrones temporales de actividad cerebral, para caracterizar en fun-

ción de la similitud/disimilitud de las series diferentes patologías (medicina);

4. desarrollar mapas climáticos a partir de la comparación y análisis de series temporales de precipitaciones, temperatura y presión atmosférica (climatología).

Podemos distinguir al menos dos enfoques centrales para evaluar la proximidad entre series de tiempo[2]. Primero, un enfoque paramétrico, que consiste en la proyección de series de tiempo en un determinado espacio de forma funcional, que corresponde, por ejemplo, a aproximar una serie con un modelo ARMA o un polinomio. El segundo enfoque es no paramétrico y consiste en evaluar la proximidad entre series de tiempo en base a los datos propios de las mismas. Dentro del alcance del enfoque no paramétrico, una de las medidas convencionales de proximidad es la distancia euclídea. Si bien estas medidas son muy utilizadas, ignoran la interdependencia entre las mediciones de una serie temporal y basan la similitud entre dos series únicamente en la proximidad entre valores observados de ambas en los mismos instantes de tiempo. Para paliar esta limitación, se han planteado diferentes propuestas que estudian el comovimiento entre dos series de tiempo, además de la cercanía entre sus valores. Entre estas iniciativas se encuentra el índice elaborado por Chouakria y Nagabhushan en el año 2007[2], que se presenta en la sección 3.3 de este capítulo. En las secciones 3.1 y 3.2, expondremos los preliminares necesarios para definir dicho índice.

### 3.1. Distancia euclídea como medida de proximidad entre series temporales

**Definición 3.1.1:** Sean  $S_1 = (u_1, \dots, u_p)$  y  $S_2 = (v_1, \dots, v_p)$  dos series de tiempo observadas en los instantes  $(t_1, \dots, t_p)$ . Definimos la **distancia euclídea**  $\delta_E$  entre  $S_1$  y  $S_2$  como:

$$\delta_E(S_1, S_2) = \left( \sum_{i=1}^p (u_i - v_i)^2 \right)^{\frac{1}{2}}.$$

Se puede observar directamente de la definición anterior que la cercanía entre dos series de tiempo, medida con la distancia euclídea, depende exclusivamente de la proximidad de los valores observados en los correspondientes instantes de tiempo. Mas aún, las observaciones son independientes para la distancia euclídea, en el sentido

de que  $\delta_E$  resulta invariante con respecto a las transformaciones que modifican el orden de las observaciones a través del tiempo. En otras palabras, es importante notar que  $\delta_E$  ignora la información de interdependencia entre los valores observados.

### 3.2. Correlación temporal para medir proximidad con respecto a comportamiento

En esta sección, estamos interesados en definir un coeficiente de correlación temporal entre dos series, que considere el comportamiento de la relación vecina entre los valores de cada serie de tiempo. En otras palabras, nos interesa poder medir de algún modo la similitud en las tasas de crecimiento de dos series en períodos de tiempo de la forma  $[t_i, t_{i+1}]$ . Para ello recordemos algunos conceptos básicos.

**Definición 3.2.1:** Sea  $S = \{u_1, \dots, u_p\}$  un conjunto de  $p$  observaciones de una variable aleatoria obtenidas en forma independiente. La **varianza muestral** dentro de  $S$  se define como:

$$\widehat{Var}(S) = \frac{1}{p-1} \sum_{i=1}^p (u_i - m)^2 \text{ donde } m = \frac{1}{p} \sum_{i=1}^p u_i.$$

**Proposición 3.2.1:** Sea  $S = \{u_1, \dots, u_p\}$  un conjunto de  $p$  observaciones de una variable aleatoria obtenidas en forma independiente.

Entonces  $\widehat{Var}(S) = \frac{1}{2p(p-1)} \sum_{i=1}^p \sum_{j=1}^p (u_i - u_j)^2$ .

**Demostración:**

$$\begin{aligned}
\widehat{Var}(S) &= \frac{1}{p-1} \sum_{i=1}^p \left( u_i - \frac{1}{p} \sum_{j=1}^p u_j \right)^2 \\
&= \frac{1}{p-1} \sum_{i=1}^p \left( u_i^2 - 2u_i \frac{1}{p} \sum_{j=1}^p u_j + \left( \sum_{j=1}^p \frac{u_j}{p} \right)^2 \right) \\
&= \frac{1}{p-1} \left( \sum_{i=1}^p u_i^2 - 2p \left( \sum_{i=1}^p \frac{u_i}{p} \right)^2 + \frac{1}{p} \left( \sum_{j=1}^p u_j \right)^2 \right) \\
&= \frac{1}{p-1} \left( \sum_{i=1}^p u_i^2 - \frac{1}{p} \left( \sum_{j=1}^p u_j \right)^2 \right) \\
&= \frac{2p}{2p} \frac{1}{p-1} \left( \sum_{i=1}^p u_i^2 - \frac{1}{p} \left( \sum_{j=1}^p u_j \right)^2 \right) \\
&= \frac{1}{2p(p-1)} \left( 2p \sum_{i=1}^p u_i^2 - 2 \left( \sum_{j=1}^p u_j \right)^2 \right) \\
&= \frac{1}{2p(p-1)} \left( p \sum_{i=1}^p u_i^2 - 2 \sum_{i=1}^p \sum_{j=1}^p u_i u_j + p \sum_{j=1}^p u_j^2 \right) \\
&= \frac{1}{2p(p-1)} \sum_{i=1}^p \sum_{j=1}^p (u_i^2 - 2u_i u_j + u_j^2) \\
&= \frac{1}{2p(p-1)} \sum_{i=1}^p \sum_{j=1}^p (u_i - u_j)^2. \blacksquare
\end{aligned}$$

**Definición 3.2.2:** Sean  $S_1 = \{u_1, \dots, u_p\}$  y  $S_2 = \{v_1, \dots, v_p\}$  dos conjuntos de  $p$  observaciones de dos variables aleatorias obtenidas en forma independiente. La **covarianza muestral** entre los valores de  $S_1$  y  $S_2$  se define como:

$$Cov(\widehat{S}_1, \widehat{S}_2) = \frac{1}{p-1} \sum_{i=1}^p (u_i - m_1)(v_i - m_2) \text{ donde } m_1 = \frac{1}{p} \sum_{i=1}^p u_i \text{ y } m_2 = \frac{1}{p} \sum_{i=1}^p v_i.$$

**Proposición 3.2.2:** Sean  $S_1 = \{u_1, \dots, u_p\}$  y  $S_2 = \{v_1, \dots, v_p\}$  dos conjuntos de  $p$  observaciones de dos variables aleatorias obtenidas en forma independiente.

Entonces  $Cov(\widehat{S}_1, \widehat{S}_2) = \frac{1}{2p(p-1)} \sum_{j=1}^p \sum_{i=1}^p (u_i - u_j)(v_i - v_j)$ .



**Demostración:**

$$\begin{aligned}
\widehat{Cov}(S_1, S_2) &= \frac{1}{p-1} \sum_{i=1}^p \left( u_i - \frac{1}{p} \sum_{j=1}^p u_j \right) \left( v_i - \frac{1}{p} \sum_{j=1}^p v_j \right) \\
&= \frac{1}{p-1} \sum_{i=1}^p \left( u_i v_i - \frac{1}{p} u_i \sum_{j=1}^p v_j - \frac{1}{p} v_i \sum_{j=1}^p u_j + \frac{1}{p^2} \left( \sum_{j=1}^p u_j \right) \left( \sum_{j=1}^p v_j \right) \right) \\
&= \frac{1}{p-1} \left( \sum_{i=1}^p u_i v_i - \frac{1}{p} \left( \sum_{i=1}^p u_i \right) \left( \sum_{j=1}^p v_j \right) - \frac{1}{p} \left( \sum_{i=1}^p v_i \right) \left( \sum_{j=1}^p u_j \right) + \frac{1}{p} \left( \sum_{j=1}^p u_j \right) \left( \sum_{j=1}^p v_j \right) \right) \\
&= \frac{1}{p-1} \left( \sum_{i=1}^p u_i v_i - \frac{2}{p} \left( \sum_{i=1}^p u_i \right) \left( \sum_{j=1}^p v_j \right) + \frac{1}{p} \left( \sum_{j=1}^p u_j \right) \left( \sum_{j=1}^p v_j \right) \right) \\
&= \frac{1}{p-1} \left( \sum_{i=1}^p u_i v_i - \frac{1}{p} \left( \sum_{i=1}^p u_i \right) \left( \sum_{j=1}^p v_j \right) \right) \\
&= \frac{2p-1}{2p(p-1)} \left( \sum_{i=1}^p u_i v_i - \frac{1}{p} \left( \sum_{i=1}^p u_i \right) \left( \sum_{j=1}^p v_j \right) \right) \\
&= \frac{1}{2p(p-1)} \left( 2p \sum_{i=1}^p u_i v_i - 2 \left( \sum_{i=1}^p u_i \right) \left( \sum_{j=1}^p v_j \right) \right) \\
&= \frac{1}{2p(p-1)} \left( p \sum_{i=1}^p u_i v_i - \sum_{j=1}^p v_j \sum_{i=1}^p u_i - \sum_{j=1}^p u_j \sum_{i=1}^p v_i \right) \\
&= \frac{1}{2p(p-1)} \left( \sum_{j=1}^p \sum_{i=1}^p u_i v_i - u_i v_j - u_j v_i + u_j v_j \right) \\
&= \frac{1}{2p(p-1)} \sum_{j=1}^p \sum_{i=1}^p (u_i - u_j)(v_i - v_j). \blacksquare
\end{aligned}$$

**Definición 3.2.3:** Sean  $S_1 = \{u_1, \dots, u_p\}$  y  $S_2 = \{v_1, \dots, v_p\}$  dos conjuntos de  $p$  observaciones de dos variables aleatorias obtenidas en forma independiente. La **correlación muestral** entre los valores de  $S_1$  y  $S_2$  se define como:

$$\widehat{Corr}(S_1, S_2) = \frac{\widehat{Cov}(S_1, S_2)}{\sqrt{\widehat{Var}(S_1)} \sqrt{\widehat{Var}(S_2)}}$$

**Proposición 3.2.3:** Sean  $S_1 = \{u_1, \dots, u_p\}$  y  $S_2 = \{v_1, \dots, v_p\}$  dos conjuntos de  $p$  observaciones de dos variables aleatorias obtenidas en forma independiente. En-

$$\text{tonces } \widehat{Corr}(S_1, S_2) = \frac{\sum_{j=1}^p \sum_{i=1}^p (u_i - u_j)(v_i - v_j)}{\sqrt{\sum_{i=1}^p \sum_{j=1}^p (u_i - u_j)^2} \sqrt{\sum_{i=1}^p \sum_{j=1}^p (v_i - v_j)^2}}.$$

**Demostración:** Inmediata de las proposiciones 3.2.1 y 3.2.2. ■

En el caso general de mediciones independientes, se tiene que la varianza, covarianza y correlación se basan en la contribución de todas las parejas de mediciones. Por el contrario, en el caso de mediciones interdependientes que definen una relación vecina, las expresiones de la varianza y covarianza clásicas podrían ser descompuestas como:

1.  $\widehat{Var}(S) = \frac{1}{2p(p-1)} \left( \sum_{i \text{ vecino de } j} (u_i - u_j)^2 + \sum_{i \text{ no vecino de } j} (u_i - u_j)^2 \right)$
2.  $\widehat{Cov}(S_1, S_2) = \frac{1}{2p(p-1)} \left( \sum_{i \text{ vecino de } j} (u_i - u_j)(v_i - v_j) + \sum_{i \text{ no vecino de } j} (u_i - u_j)(v_i - v_j) \right)$

donde consideramos que  $i < j$  y que  $i$  es vecino de  $j$  si  $j = i + 1$  debido a que nos interesa comparar tasas de crecimiento en períodos de tiempo  $[t_i, t_{i+1}]$ . La idea principal para la inclusión de la información sobre la dependencia entre los valores vecinos es restringir la expresión de la correlación clásica solo para aquellas parejas de valores. A tales efectos, definimos un nuevo coeficiente de correlación al que llamaremos coeficiente de correlación temporal.

**Definición 3.2.4:** Sean  $S_1 = (u_1, \dots, u_p)$  y  $S_2 = (v_1, \dots, v_p)$  dos series de tiempo observadas en los instantes  $(t_1, \dots, t_p)$ . Definimos el **coeficiente de correlación temporal** entre  $S_1$  y  $S_2$  como:

$$CorT(S_1, S_2) = \frac{\sum_{i=1}^{p-1} (u_i - u_{i+1})(v_i - v_{i+1})}{\sqrt{\sum_{i=1}^{p-1} (u_i - u_{i+1})^2} \sqrt{\sum_{i=1}^{p-1} (v_i - v_{i+1})^2}}$$

Notemos que  $CorT \in [-1, 1]$ . El valor  $CorT(S_1, S_2) = 1$  significa que en cualquier período observado  $[t_i, t_{i+1}]$ , las series  $S_1$  y  $S_2$  han aumentado o disminuído simultáneamente con la misma tasa de crecimiento. El valor  $CorT(S_1, S_2) = -1$  significa que en cualquier período observado  $[t_i, t_{i+1}]$ , donde  $S_1$  ha aumentado  $S_2$  ha disminuído con la misma tasa de crecimiento, y viceversa. Por último, el valor  $CorT(S_1, S_2) = 0$  expresa que no hay monotonicidad entre  $S_1$  y  $S_2$ , y que sus tasas de crecimiento son estocásticamente linealmente independientes.

### 3.3. Índice de disimilitud para series de tiempo

En el año 2007, Chouakria y Nagabhushan propusieron un nuevo índice de disimilitud entre series temporales que tiene en cuenta tanto el comportamiento o comovimiento de las series, como los valores de las medidas de proximidad convencionales. Es importante para el estudio de un índice de disimilitud, especificar cuales son las características de las series de tiempo que compara el mismo y entender que información nos brinda esta medida de comparación. A tales efectos, introducimos dos definiciones; la definición 3.3.1 que establece cuando dos series se acompañan a lo largo del tiempo, es decir, son similares en comportamiento, y la definición 3.3.2, que establece cuando dos series son similares en valores.

**Definición 3.3.1:** Sean  $S_1 = (u_1, \dots, u_p)$  y  $S_2 = (v_1, \dots, v_p)$  dos series de tiempo observadas en los instantes  $t_1, \dots, t_p$ . Decimos que  $S_1$  y  $S_2$  son **similares en comportamiento**, si en cualquier período de tiempo observado, digamos  $[t_i, t_{i+1}]$ , aumentan o disminuyen de forma simultánea (monotonicidad), con una misma tasa de crecimiento (cercanía de las tasas de crecimiento); en otras palabras, si el valor de  $CorT(S_1, S_2)$  se aproxima a 1.

**Definición 3.3.2:** Sean  $S_1 = (u_1, \dots, u_p)$  y  $S_2 = (v_1, \dots, v_p)$ ,  $p$  valores de dos series de tiempo observados en los instantes de tiempo  $t_1, \dots, t_p$ . Decimos que  $S_1$  y  $S_2$  son **similares en valores**, si  $\delta_E(S_1, S_2)$  es próxima a 0.

A continuación definimos el índice de Chouakria y Nagabhushan, que denotaremos como  $D$ .

**Definición 3.3.3:** Sean  $S_1 = (u_1, \dots, u_p)$  y  $S_2 = (v_1, \dots, v_p)$  dos series de tiempo observadas en los instantes  $t_1, \dots, t_p$ . Definimos el **índice de disimilitud**  $D$  como

$$D(S_1, S_2) = f(CorT(S_1, S_2)) \cdot \delta_E(S_1, S_2)$$

donde  $f$  es la función dada por  $f(x) = \frac{2}{1+exp(k \cdot x)}$ ,  $k \geq 0$  entero fijo.

Este índice, como ya señalamos, considera la similitud en comportamiento y la similitud en valores al comparar dos series de tiempo. Además es capaz de ajustar y modular la contribución de cada uno de estos dos aspectos de las series a comparar, a través de una función de modulación  $f$  que presenta las siguientes propiedades:

1. aumenta el grado de aporte al índice de la medida convencional de distancia entre las series, cuando la correlación temporal disminuye de 0 a -1;
2. el valor de  $D$  se aproxima al valor de la medida convencional de distancia entre las series, si la correlación temporal es próxima a cero;
3. disminuye el grado de aporte al índice de la medida convencional de distancia entre las series, cuando la correlación temporal aumenta de 0 a 1.

En la figura 3.1, se muestra el comportamiento de la función  $f(CorT)$  para distintos valores de  $k$ .

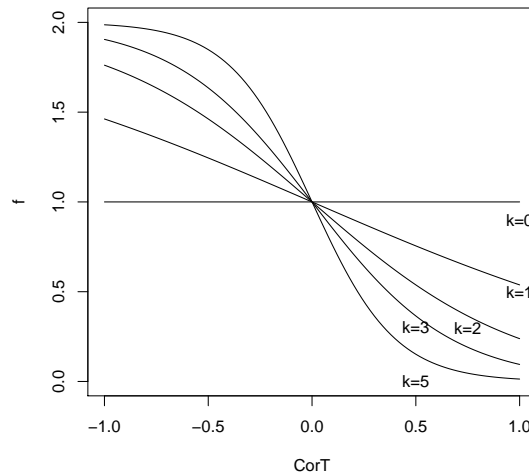


Figura 3.1: Gráfico de la función  $f(x) = \frac{2}{1+\exp(k.x)}$  para distintos valores de  $k$

Notemos que para cualquier valor de  $k$ ,  $f(0) = 1$ . En el caso de series de tiempo con comportamientos diferentes, es decir,  $CorT$  cerca de cero,  $f(x)$  es cercano a 1 sea cual sea el valor de  $k$ ; y en este caso,  $D$  es aproximadamente igual a  $\delta_E$ . Sin embargo, en el caso de series de tiempo con un comportamiento opuesto o similar, es decir, cuando  $|CorT|$  es cercano a 1, el parámetro  $k$  modula para el índice de disimilitud  $D$ , las contribuciones del comportamiento y de la cercanía en valores de las series.

### 3.4. Aplicación y comportamiento del índice de disimilitud

En esta sección presentaremos una aplicación del uso del índice de disimilitud  $D$  al problema de distinguir diferentes coberturas de suelo. Para ello compararemos series temporales del índice de vegetación NDVI (Normalized Difference Vegetation Index), que puede ser captado a partir de diferentes sensores remotos.

Este índice es usado para estimar la cantidad, calidad y desarrollo de la vegetación en base a la medición de la intensidad de la radiación que la vegetación emite o refleja, por medio de sensores remotos instalados comúnmente desde una plataforma espacial[7]. En nuestro caso, las series de NDVI fueron obtenidas a partir de imágenes captadas por el sensor MODIS (Moderate Resolution Imaging Spectroradiometer) a bordo del satélite Terra. El sensor MODIS es un instrumento que viaja a bordo de los satélites Terra y Aqua. La órbita de Terra alrededor de la tierra viaja del norte al sur cruzando el Ecuador por la mañana, mientras que el Aqua viaja del sur al norte cruzando el Ecuador por la tarde. Terra-MODIS y Aqua-MODIS cubren la superficie de la tierra, adquiriendo datos en 36 bandas espectrales. Estos datos ayudan a comprender la dinámica global de la Tierra y el comportamiento en la superficie terrestre, en los océanos y en la atmósfera. MODIS está desempeñando un papel vital en el desarrollo de modelos validados y globales, capaces de predecir el cambio global que está sufriendo el planeta Tierra, colaborando con los gobiernos en la toma de decisiones sanas referentes a la protección de nuestro ambiente. Para mayores detalles puede consultarse el sitio (<http://crepadweb.cec.inta.es/es/plataformas/modis.html>).

Los datos con los que trabajaremos se encuentran disponibles de forma libre y gratuita en [https://lpdaac.usgs.gov/data\\_access/](https://lpdaac.usgs.gov/data_access/); corresponden al producto MODQ13 elaborado por el grupo MODLAND perteneciente a la NASA, y fueron obtenidos desde el 18 de febrero de 2000 hasta el 14 de septiembre de 2012. El sensor capta una imagen cada 16 días, por lo que las series de datos tendrá un valor de NDVI cada 16 días.

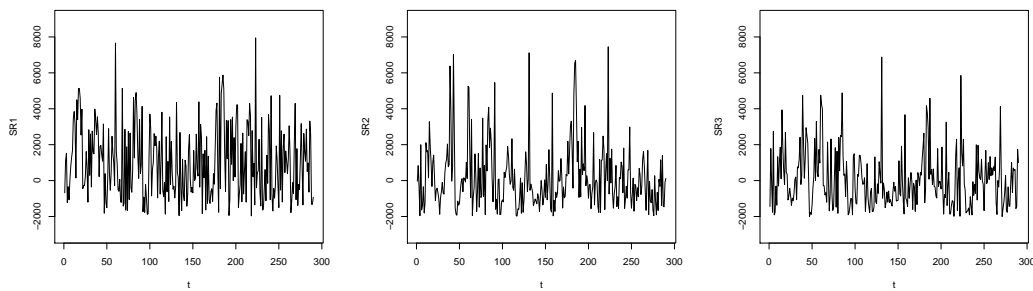
Los sitios geográficos elegidos para la extracción de los valores de NDVI se corresponden con sitios cuyas coberturas se clasifican en:

1. Ciudad (ciudad de Córdoba)

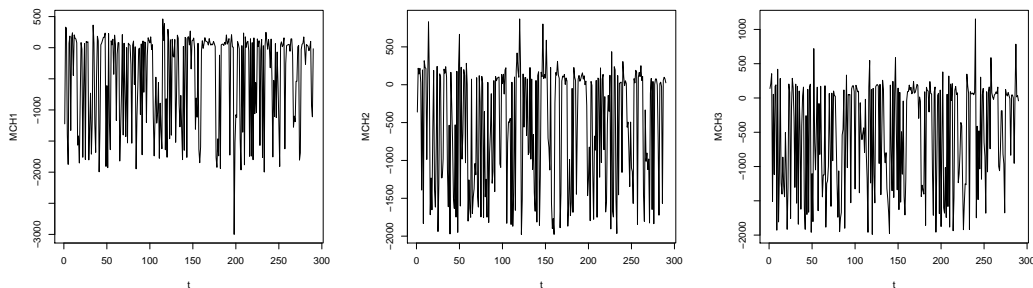
2. Suelo de explotación agronómica (ubicado al este de la provincia de Córdoba)

3. Región que contiene agua (Lago San Roque y Mar Chiquita)

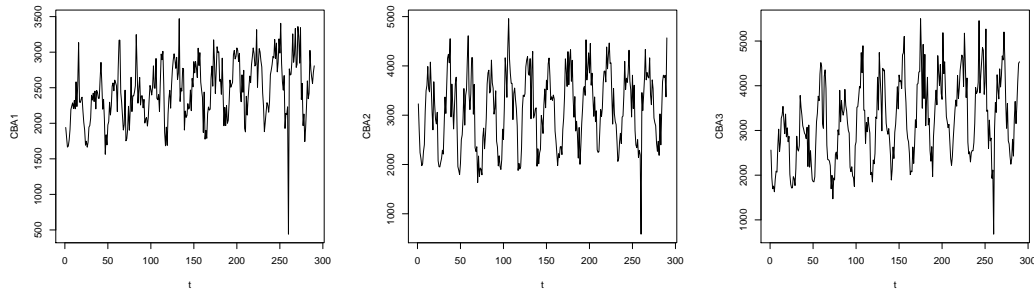
y fueron seleccionados por un biólogo especialista. El agrupamiento real de los datos lo consideramos conocido a priori. En esta ocasión contamos con tres series de datos de la zona del lago San Roque (SR1, SR2, SR3), tres series de la zona de la laguna Mar Chiquita (MC1, MC2, MC3), tres series de la zona de la ciudad de Córdoba (CBA1, CBA2, CBA3) y tres series de una zona de explotación agronómica al este de la provincia de Córdoba (AG1, AG2, AG3). La Figura 3.2 muestra el gráfico de las 12 series recién mencionadas.



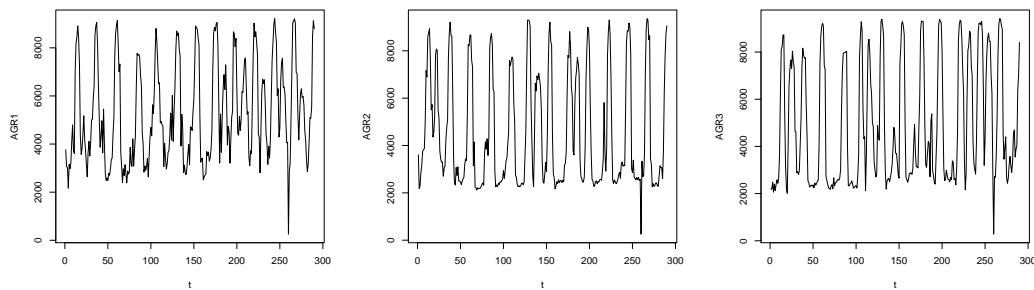
(a) Series del lago San Roque



(b) Series de la laguna Mar Chiquita



(c) Series de ciudad de Córdoba



(d) Series de zona de explotación agronómica

Figura 3.2: Gráfico de series de NDVI

Las series fueron comparadas de a pares a partir del índice  $D$ , con  $k = 3$ . La implementación computacional del índice se realizó en el software libre R, la cual puede verse en Anexo. El cuadro de la Figura 3.3 muestra los resultados obtenidos y puede verse por dichos resultados como el índice parece capaz de distinguir las series de regiones geográficas que pertenecen a una misma cobertura de suelo. El cuadro 3.1 muestra en promedio el resultado del índice al comparar las series de NDVI entre los grupos y la Figura 3.4 ilustra dicha situación.

	SR				MCH				CBA				AGR			
	1	2	3		1	2	3		1	2	3		1	2	3	
SR	1	0	18780	20987	1	37048	49080	41245	1	35207	50107	53946	1	82736	81783	88430
	2	18780	0	5895	2	35114	39463	34945	2	43053	44924	53507	2	85777	74774	90968
	3	20987	5895	0	3	33692	33444	30593	3	46239	48322	46183	3	83020	71947	75291
MCH	1	37048	35114	33692	1	0	8369	6306	1	52294	69001	74853	1	111631	97122	104139
	2	49080	39463	33444	2	8369	0	9062	2	58442	68933	65428	2	104182	95342	101933
	3	41245	34945	30593	3	6306	9062	0	3	53536	71590	62036	3	91376	96145	108180
CBA	1	35207	43053	46239	1	52294	58442	53536	1	0	7360	9352	1	41227	48084	50650
	2	50107	44924	48322	2	69001	68933	71590	2	7360	0	4911	2	25554	26996	37671
	3	53946	53507	46183	3	74853	65428	62036	3	9352	4911	0	3	30952	31156	38796
AGR	1	82736	85777	83020	1	111631	104182	91376	1	41227	25554	30952	1	0	14361	8847
	2	81783	74774	71947	2	97122	95342	96145	2	48084	26996	31156	2	14361	0	12283
	3	88430	90968	75291	3	104139	101933	108180	3	50650	37671	38796	3	8847	12283	0

Figura 3.3: Resultado de índice D para los distintos pares de series, con un valor de  $k = 3$

	SR	MCH	CBA	AGR
SR	<b>15220</b>	37180	46832	81636
MCH	37180	<b>7912</b>	64012	101116
CBA	46832	64012	<b>7207</b>	36787
AGR	81636	101116	36787	<b>11830</b>

Cuadro 3.1: Promedio del resultado del índice al comparar las series de NDVI entre los grupos.



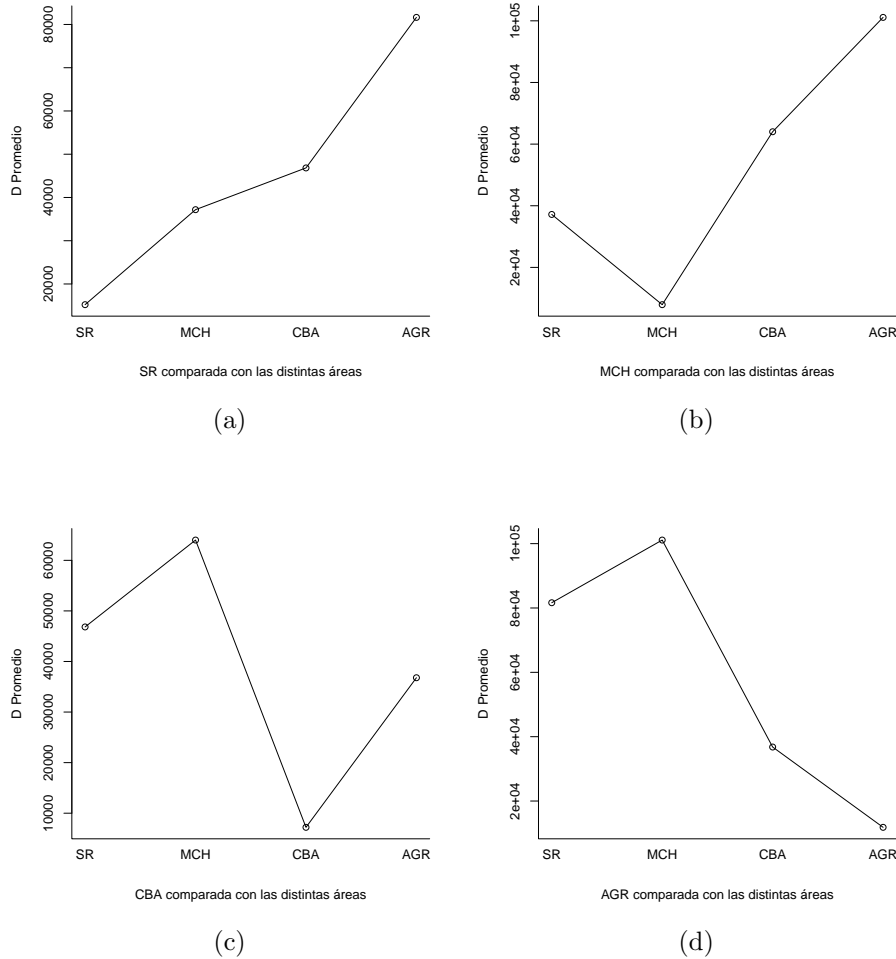


Figura 3.4: Promedio del resultado del índice  $D$  al comparar las series de NDVI entre los distintos grupos de cobertura de suelo.

En general, el índice asumió valores mas bajos cuando comparamos dos series de la misma región geográfica; al comparar dos series de regiones geográficas con presencia de agua, pero de distinto grupo, una serie de agua del lago San Roque, y otra de la laguna Mar Chiquita, el índice arrojó valores inferiores con relación a la comparación de una serie de agua y una serie que no fuera de agua; sin embargo los valores resultaron superiores a comparar dos series de agua de la misma región. Este resultado nos muestra que el índice, al menos en los datos analizados, podría ser capaz de distinguir entre dos grupos de agua. Sin embargo, es fundamental destacar

la importancia de la elección del parámetro  $k$ , que interviene en el peso o ponderación que se da a la similitud de las series en cuanto a su comovimiento y a su similitud en valores. En nuestro caso, elegimos un valor de  $k = 3$ , siguiendo las indicaciones de los autores del índice  $D$ , permitiendo un mayor aporte del comovimiento de las series con respecto al aporte de la similitud en valores. En las series analizadas los valores de las series representan una cierta intensidad o nivel de gris, de modo que, la información de la cercanía en valores es muy importante en el análisis de comparación de las series si se quiere, por ejemplo, distinguir si las características de las áreas que representan son similares. Pero, asimismo, la relación entre los valores vecinos de cada serie puede permitir, por ejemplo, analizar como pueden modificarse las áreas con el correr del tiempo o, si se modifican con la misma tasa de crecimiento. Esta última información con respecto a los valores vecinos de las series, complementa enormemente a la información de la cercanía en valores.

---

## Conclusión y trabajo futuro

En este trabajo se realizó en primer lugar un estudio de las definiciones y resultados básicos del Análisis de Series de Tiempo[1]. En el Capítulo 3 se abordó el problema de comparar dos series temporales, destacando la importancia de incorporar no solo una comparación entre los valores de las series, sino también un análisis del comovimiento (comportamiento) de las mismas. En particular, se estudió el paper de Chouakria y Nagabhushan del año 2007 “Adaptive dissimilarity index for measuring time series proximity”[2], y se implementó computacionalmente en el software libre R el índice  $D$  que los autores proponen en el trabajo, para el caso de la distancia euclídea. Finalmente se aplicó este índice al problema de distinguir coberturas de suelo, utilizando series de tiempo del índice de vegetación NDVI. El análisis descriptivo de los resultados determinó que el índice de disimilitud  $D$ , puede ser una herramienta capaz de agrupar y distinguir diferentes coberturas de suelo, a partir de la comparación de series temporales de NDVI de dichas coberturas. Como trabajo futuro, se plantea realizar un estudio de cluster a partir de los resultados que arroje el índice sobre una base de datos de series de NDVI con mayor cantidad de series por grupo de cobertura. Se plantea también el estudio de la dependencia del índice con relación al parámetro  $k$ , pues si bien Chouakria y Nagabhushan sugieren utilizar  $k = 3$ , no hay estudios formales sobre esta decisión. Finalmente se plantea a futuro, estudiar como se comporta el índice utilizando otra distancia diferente de la euclídea, por ejemplo, la distancia de Fréchet y la de Minkowski[2].

---

# Anexo

## Implementación en R del índice de disimilitud D

```
CORT = function(r, s){m = length(r, s); y = 0; z = 0; w = 0;
  for(i in 1 : (m - 1)){y ← y + (r[i + 1] - r[i]) * (s[i + 1] - s[i])};
  for(i in 1 : (m - 1)){z ← z + (r[i + 1] - r[i]) ^ 2};
  for(i in 1 : (m - 1)){w ← w + (s[i + 1] - s[i]) ^ 2};
  return(y/sqrt(z * w))}
% r y s son dos vectores en  $\mathbb{R}^n$ , para algún  $n \in \mathbb{N}$ .
```

```
delta = function(r, s){m = length(r); y = 0;
  for(i in 1 : m){y ← y + (r[i] - s[i])^2};
  return(y ^ (1/2))}
% r y s son dos vectores en  $\mathbb{R}^n$ , para algún  $n \in \mathbb{N}$ .
```

```
f = function(x, k){return(2/(1 + exp(k * x)))}
% x es un valor de  $\mathbb{R}$  y  $k \in \mathbb{N} \cup \{0\}$ .
```

```
D = function(r, s, k){return(f(CORT(r, s), k) * delta(r, s))}
% r y s son vectores en  $\mathbb{R}^n$ .
% k  $\in \mathbb{N} \cup \{0\}$ .
```

---

# Bibliografía

- [1] Leiva, R. 1995. Introducción al Análisis de las series de tiempo. Facultad de Ciencias Económicas de la Universidad Nacional de Cuyo. ISBN 9-503-90057-3 e-ISBN 978-9-503-90057-4.
- [2] Chouakria, A. D. y Nagabhushan, P. N. (2007). Adaptive dissimilarity index for measuring time series proximity. *Advances in Data Analysis and Classification*, 1(1), 5-21.
- [3] Morettin P. A. y Toloí C. M. de Castro. *Previsão de sries temporais. 2a Edición.* Atual editora.
- [4] Ronny V. y Ojeda. S. M. (2012). *Image Segmentation and Time Series Clustering Based on Spatial and Temporal ARMA Processes.* INTECH.
- [5] Chouakria Douzal A.(2003). *Compression technique preserving correlations of a multivariate temporal sequence.*. Springer, Berlin Heidelberg, pp 566-577.
- [6] Francisco J. Garcia-Lopez, Ildar Batyrshin, Alexander Gelbukh. *Similitud de series de tiempo basada en longitud de patrones de la transformada por aproximación móvil.* *Research in Computing Science* 115 (2016), pp. 7992.
- [7] Genesis T. Yengoh, David Dent, Lennart Olsson, Anna E. Tengberg, Compton J. Tucker III (2015). *Use of the Normalized Difference Vegetation Index (NDVI) to assess land degradation at multiple scales.* Springer.
- [8] Jonathan D. Cryer, Kung-Sik Chan (2008). *Time series analysis with applications in R.* Springer.