

Índice

1. Series temporales vs datos de sección cruzada	2
1.1. Correlación serial vs muestreo aleatorio simple	6
1.2. Simplificación del escenario	7
2. Procesos estocásticos de segundo orden	8
2.1. Un poco de geometría	8
2.2. Primeros momentos de procesos estocásticos de segundo orden	8
2.3. Procesos estocásticos (débilmente) estacionarios y la ACF	9
3. Notación: convolución y el operador retardo	10
4. Ejemplos de procesos (débilmente) estacionarios	11
4.1. Proceso de ruido blanco	11
4.2. Procesos lineales	11
4.2.1. Media móvil infinita. MA(∞)	12
4.2.2. Proceso de media móvil de orden q . MA(q)	12
4.2.3. Proceso autorregresivo de orden p . AR(p)	13
4.2.4. Proceso autorregresivo de media móvil. ARMA(p, q)	14
4.2.5. Proceso autorregresivo de media móvil con media no nula	15
5. Primeros momentos de procesos lineales causales	15
5.1. Esperanza y autocovarianzas de un proceso lineal causal	15
5.2. Covarianza cruzada entre dos procesos lineales causales	16
5.3. Las Ecuaciones de Yule-Walker para un AR(p) estacionario	16
5.4. Función de autocovarianzas para un ARMA(p, q)	17

Lección 3. Modelos lineales

Marcos Bujosa

18 de agosto de 2025

Resumen

Analizaremos las dificultades que plantea la **correlación serial** y veremos cómo la estacionariedad en sentido débil los simplifica. Nos centraremos en los **procesos lineales**, estudiando su valor esperado, su función de autocovarianza, las covarianzas cruzadas entre ellos y las ecuaciones de YuleWalker.

- ([slides](#)) — ([html](#)) — ([pdf](#)) — ([mybinder](#))

Carga de algunos módulos de python y creación de directorios auxiliares

```
# Para trabajar con los datos y dibujarlos necesitamos cargar algunos módulos de python
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import matplotlib as mpl
# definimos parámetros para mejorar los gráficos
mpl.rcParams['text', usetex=False]
import matplotlib.pyplot as plt # data visualization
```

- Creación del directorio auxiliar para albergar las figuras de la lección. Para publicar la lección como pdf o página web, necesito los gráficos como ficheros .png alojados algún directorio específico:

```
imagenes_leccion = "./img/lecc03" # directorio para las imágenes de la lección
import os
os.makedirs(imagenes_leccion, exist_ok=True) # crea el directorio si no existe
```

1. Series temporales vs datos de sección cruzada

Corresponden a observaciones de un mismo objeto a lo largo del tiempo. El índice indica el instante de cada medición. *El orden cronológico puede ser crucial* al modelar los datos.

- El motivo es que frecuentemente el valor medido en un instante de tiempo está relacionado con otras mediciones próximas en el tiempo (*correlación serial*).
- Si es así, ya no deberíamos asumir que las variables aleatorias del proceso estocástico subyacente, $\mathbf{X} = (X_t \mid t \in \mathbb{Z})$, son independientes entre sí.

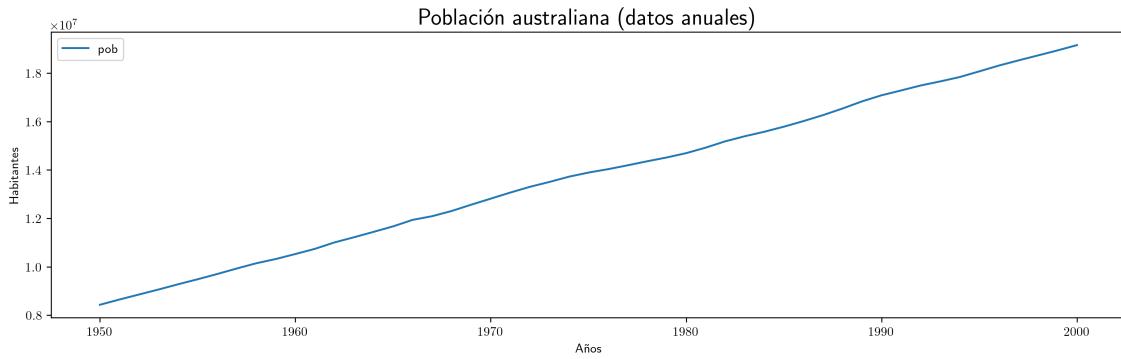
Esto tiene importantes implicaciones en las técnicas de análisis y los modelos a utilizar.

Veamos algunos ejemplos de series temporales...

```
path = '../datos/'  
df1 = pd.read_csv(path+'PoblacionAustralia.csv')  
df2 = pd.read_csv(path+'PIB_UEM.csv')  
df3 = pd.read_csv(path+'Retiro.txt')  
df4 = pd.read_csv(path+'IBEX35.csv')  
df5 = pd.read_csv(path+'ProduccionCemento.csv')
```

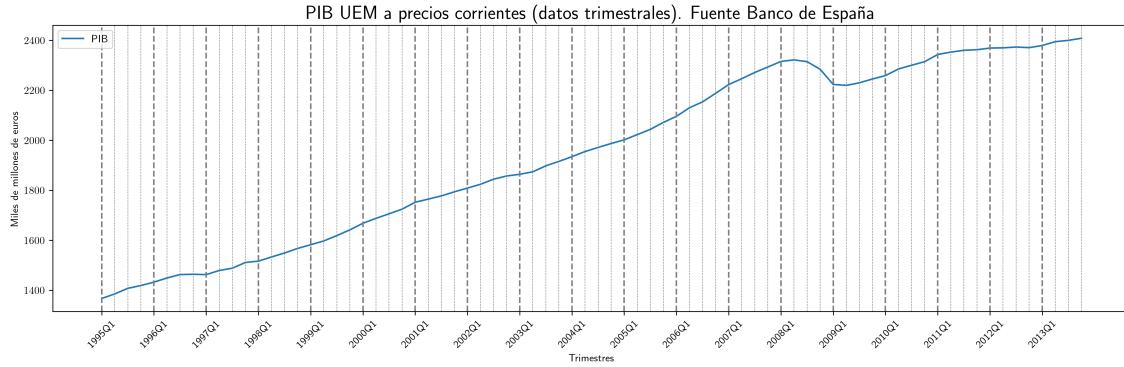
1. Población en Australia

```
df1.plot(x='obs', xlabel='Años', ylabel='Habitantes', figsize=(15, 4)).set_title('Población australiana (datos anuales)', fontweight='bold')  
plt.savefig('../img/lecc03/PoblacionAustralia.png', dpi=300, bbox_inches='tight')
```



2. PIB UEM

```
# Crear el gráfico  
fig, ax = plt.subplots(figsize=(15, 5))  
df2.plot(x='obs', xlabel='Trimestres', ylabel='Miles de millones de euros', ax=ax)  
ax.set_title('PIB UEM a precios corrientes (datos trimestrales). Fuente Banco de España', fontsize=18)  
  
# Agregar líneas verticales con diferente grosor  
for i, obs in enumerate(df2['obs']):  
    if obs.endswith('Q1'): # Primer trimestre  
        ax.axvline(x=i, color='gray', linestyle='--', linewidth=1.5) # Más gruesa  
    else: # Otros trimestres  
        ax.axvline(x=i, color='gray', linestyle='--', linewidth=0.5) # Más fina  
  
# Personalizar las etiquetas del eje X  
ticks = []  
labels = []  
for i, obs in enumerate(df2['obs']):  
    if obs.endswith('Q1'): # Etiquetas solo en los primeros trimestres  
        ticks.append(i) # Usar el índice como posición  
        labels.append(obs)  
  
ax.set_xticks(ticks) # Configurar las posiciones de las etiquetas  
ax.set_xticklabels(labels, rotation=45) # Configurar las etiquetas correspondientes  
  
# Asegurar que todas las líneas y etiquetas se muestren  
plt.tight_layout()  
plt.savefig('../img/lecc03/PIB_UEM.png', dpi=300, bbox_inches='tight')
```



3. Temperatura media en el Parque del Retiro. Madrid

```
# Generar una columna de fechas desde enero de 1985
df3['Fecha'] = pd.date_range(start='1985-01', periods=len(df3), freq='M') # cambiar por 'ME' en el futuro

# Configurar la columna de fechas como índice
df3.set_index('Fecha', inplace=True)

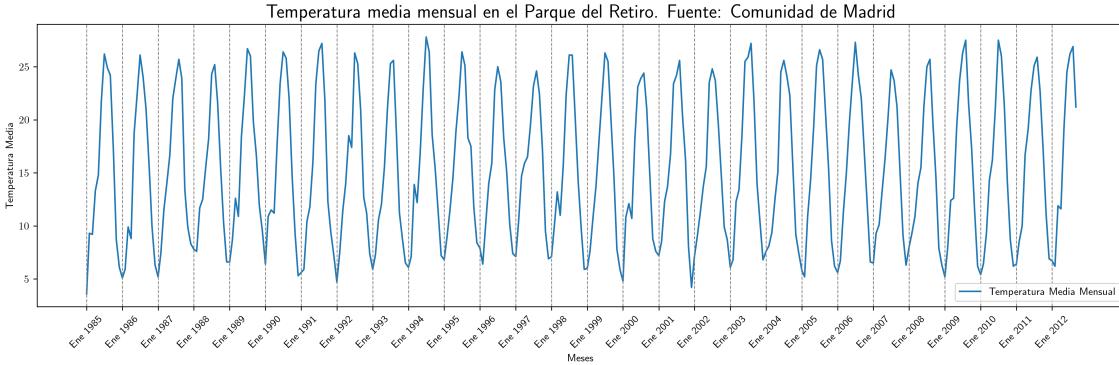
# Crear el gráfico
fig, ax = plt.subplots(figsize=(15, 5))
ax.plot(df3.index, df3['TemperaturaMedia'], label='Temperatura Media Mensual')
ax.set_title('Temperatura media mensual en el Parque del Retiro. Fuente: Comunidad de Madrid', fontsize=18)
ax.set_xlabel('Meses')
ax.set_ylabel('Temperatura Media')
ax.legend()

# Agregar barras verticales en enero de cada año
for fecha in df3.index:
    if fecha.month == 1: # Mes de enero
        ax.axvline(x=fecha, color='gray', linestyle='--', linewidth=0.8)

# Agregar etiquetas en todos los eneros
ticks = []
labels = []
for fecha in df3.index:
    if fecha.month == 1: # Enero
        ticks.append(fecha)
        labels.append(f'Ene {fecha.year}') # Formato: "Ene Año" (ej. "Ene 1985")

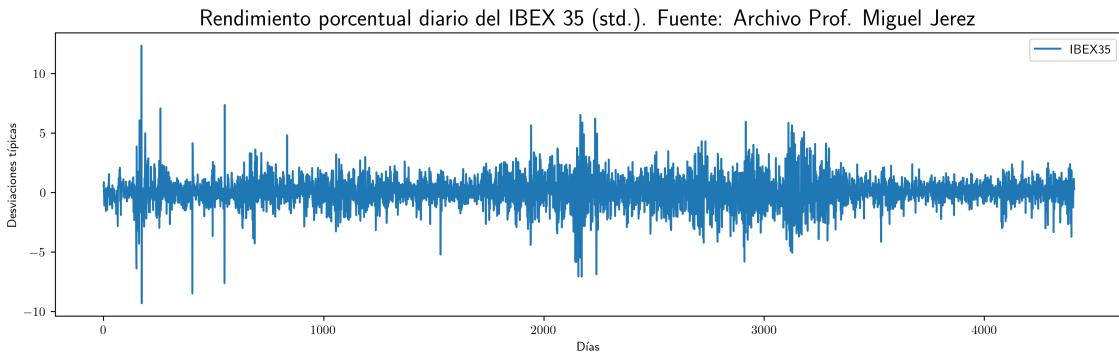
ax.set_xticks(ticks) # Configurar las posiciones de las etiquetas
ax.set_xticklabels(labels, rotation=45) # Configurar las etiquetas correspondientes

# Asegurar que todas las líneas y etiquetas se muestren
plt.tight_layout()
# Guardar el gráfico
plt.savefig('./img/lecc03/TemperaturaRetiro.png', dpi=300, bbox_inches='tight')
```



4. Rendimiento porcentual diario del IBEX 35 (std)

```
df4.plot(x='obs', xlabel='Días', ylabel='Desviaciones típicas', figsize=(15,4)).set_title('Rendimiento porcentual diario del IBEX 35 (std.)')
plt.savefig('./img/lecc03/IBEX35.png', dpi=300, bbox_inches='tight')
```



5. Producción de cemento

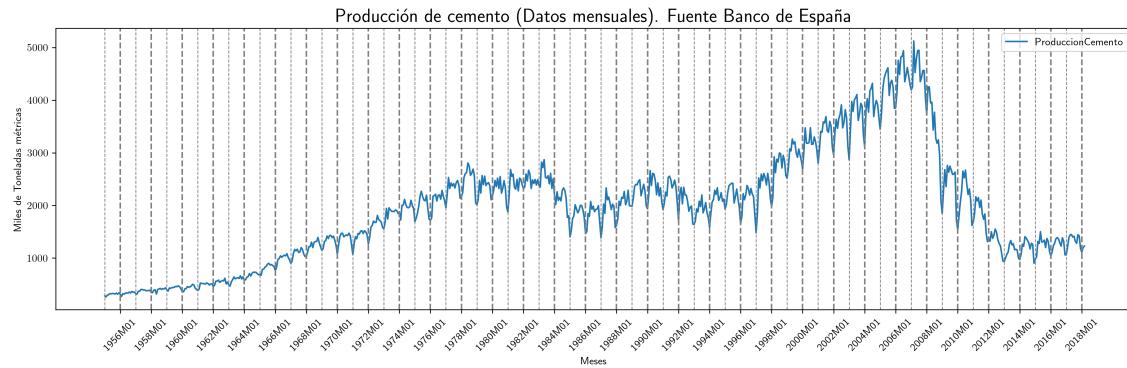
```
fig, ax = plt.subplots(figsize=(15, 5))
df5.plot(x='obs', xlabel='Meses', ylabel='Miles de Toneladas métricas', ax=ax)
ax.set_title('Producción de cemento (Datos mensuales). Fuente Banco de España', fontsize=18)

# Agregar líneas verticales con diferente grosor para años pares e impares
for i, obs in enumerate(df5['obs']):
    if obs.endswith('M01'): # Enero
        year = int(obs.split('M')[0]) # Extraer el año
        if year % 2 == 0: # Años pares
            ax.axvline(x=i, color='gray', linestyle='--', linewidth=1.5) # Más gruesa
        else: # Años impares
            ax.axvline(x=i, color='gray', linestyle='--', linewidth=0.8) # Más fina

# Personalizar las etiquetas del eje X para eneros de años pares
ticks = []
labels = []
for i, obs in enumerate(df5['obs']):
    if obs.endswith('M01'): # Enero
        year = int(obs.split('M')[0]) # Extraer el año
        if year % 2 == 0: # Solo años pares
            ticks.append(i) # Usar el índice como posición
            labels.append(obs) # Etiqueta completa (ej. "1986M01")

ax.set_xticks(ticks) # Configurar las posiciones de las etiquetas
ax.set_xticklabels(labels, rotation=45) # Configurar las etiquetas correspondientes
```

```
# Asegurar que todas las líneas y etiquetas se muestren
plt.tight_layout()
plt.savefig('./img/lecc03/ProduccionCemento.png', dpi=300, bbox_inches='tight')
```



Las anteriores series temporales presentan características que impiden su consideración como realizaciones de procesos estocásticos estacionarios.

- En tres de ellas —la población en Australia, el PIB de la zona euro y la producción de cemento— se observa una tendencia clara, lo que implica una evolución de la media local a medio y largo plazo.
- Además, se identifican patrones estacionales en las temperaturas medias del Parque del Retiro y en la producción de cemento. Estos patrones son el resultado de la **traslación constante de la Tierra** alrededor del Sol. Tales oscilaciones periódicas generan ciclos regulares, de manera que, por ejemplo, las temperaturas invernales son sistemáticamente más bajas que las estivales.
- En la producción de cemento, se observa que la amplitud de la variación estacional tiende a aumentar conforme se incrementa el nivel de la serie.
- Respecto a los rendimientos del IBEX35, se aprecian períodos de alta volatilidad intercalados con fases menos volátiles, lo que indica la presencia de heterocedasticidad.

Todas estas características resaltan la complejidad de los procesos subyacentes y la necesidad de emplear modelos que capten adecuadamente estas propiedades dinámicas.

1.1. Correlación serial vs muestreo aleatorio simple

Generalmente cuando disponemos de datos de

sección cruzada asumimos que proceden de un el muestreo es aleatorio simple

- i.e., los datos son realizaciones de variables aleatorias i.i.d.

series temporales dicha asunción resulta errónea

- con frecuencia el nivel esperado (o la volatilidad) parece cambiar con t
- con frecuencia hay dependencia temporal (correlación serial).

Ejemplo: no parece aceptable asumir que $ProdCemento_{1960M01}$ se distribuye igual que $ProdCemento_{2000M04}$ (ni que sea independiente de $ProdCemento_{1959M01}$).

Veamos por qué esto genera dificultades en su tratamiento estadístico...

Consideremos el proceso estocástico

$$\mathbf{X} = (X_t \mid t = 0, \pm 1, \pm 2, \dots).$$

Caracterizar su distribución conjunta (**sus infinitos momentos**) resulta ser demasiado ambicioso. Limitemos a los dos primeros momentos (esperanzas y covarianzas):

$$E(X_t) = \mu_t \quad \text{y} \quad Cov(X_t, X_k) = E[(X_t - \mu_t)(X_k - \mu_k)] = \gamma_{t,k}; \quad t, k \in \mathbb{Z}$$

(donde si $k = t$, entonces $\gamma_{t,t} = Var(X_t) = \sigma_t^2$).

Si el proceso \mathbf{X} fuera gaussiano, estos parámetros bastarían para caracterizar la distribución conjunta completamente. No obstante...

- Necesitaríamos una muestra suficiente de cada X_t para estimar los parámetros;

A cada X_t le corresponden infinitos parámetros ($\mu_t, \sigma_t^2, \gamma_{t,k}$; donde $k \in \mathbb{Z}$);

Pero en una serie temporal \mathbf{x} solo disponemos de *una* realización para cada X_t .

1.2. Simplificación del escenario

Cuando \mathbf{X} es **débilmente estacionario** se reduce *drásticamente* el número de parámetros (aunque siga habiendo infinitos):

$$E(X_t) = \mu \tag{1}$$

$$Cov(X_t, X_{t-k}) = \gamma_k \quad k \in \mathbb{Z} \tag{2}$$

El desafío para el analista es (y nótese el abuso de lenguaje)

primero transformar los datos para lograr que sean "**débilmente estacionarios**".

- (Algo vimos en la lección 1)

después transformar los datos estacionarios en "**ruido blanco**"

- (Es lo que veremos en esta lección y las siguientes)

Este proceso constituye la especificación y ajuste de un modelo ARIMA a los datos.

Antes de atacar los temas de especificación y ajuste de modelos, debemos estudiar las propiedades de la familia de procesos estocásticos débilmente estacionarios.

2. Procesos estocásticos de segundo orden

El ambiente natural para estudiar las propiedades de segundo orden de una colección de variables aleatorias es el espacio de variables aleatorias X definidas en un espacio de probabilidad tales que

$$E(X) = 0 \quad \text{y} \quad E(X^2) < \infty$$

donde E es el operador esperanza. Denotaremos este espacio con H .

2.1. Un poco de geometría

El espacio, dotado de producto escalar y norma

$$\langle X | Y \rangle = E(XY), \quad \|X\| = \sqrt{E(X^2)}, \quad X, Y \in H,$$

es un espacio de Hilbert.

Nótese que como las variables de H tienen esperanza cero, el producto escalar entre $X, Y \in H$ también es

$$\langle X | Y \rangle = Cov(X, Y).$$

Por tanto, en este espacio H la noción geométrica de ortogonalidad coincide con la noción estadística de *no correlación*. Por tanto, en este contexto los términos producto escalar, covarianza y esperanza del producto serán intercambiables (esto deja de ser cierto cuando hay variables aleatorias con esperanza no nula).

Una colección de variables aleatorias pertenecientes a H

$$\mathbf{X} = (X_t | t \in \mathbb{Z}) \quad \text{con} \quad X_t \in H$$

se denomina *proceso estocástico de segundo orden*.

Si $\mathbf{Y} = (Y_t | t \in \mathbb{Z})$ es tal que $E(Y_t) = \mu \neq 0$, entonces \mathbf{Y} no es de segundo orden.

Pero basta restar μ de cada Y_t para tener un proceso $(\mathbf{Y} - \mu\mathbf{1})$ de segundo orden.

Por ello siempre asumiremos (sin pérdida de generalidad) *que las variables aleatorias de los procesos estocásticos de esta lección (y la siguiente) tienen esperanza cero.*

2.2. Primeros momentos de procesos estocásticos de segundo orden

Si $E(X_t) < \infty$ para $t \in \mathbb{Z}$, entonces $E(\mathbf{X})$ es la secuencia

$$E(\mathbf{X}) = (E(X_t) | t \in \mathbb{Z}) = \sum_{t \in \mathbb{Z}} E(X_t)z^t = (\dots, E(X_{-1}), E(X_0), E(X_1), \dots)$$

Si \mathbf{X} tiene segundos momentos finitos, la secuencia de autocovarianzas de orden k es

$$\begin{aligned} \left(Cov(X_t, X_{t-k}) \mid t \in \mathbb{Z} \right) &= (\gamma_{k,t} \mid t \in \mathbb{Z}) \\ &= (\dots, \gamma_{k,-1}, \gamma_{k,0}, \gamma_{k,1}, \gamma_{k,2}, \dots). \end{aligned}$$

(la secuencia solo contiene la covarianza de un orden k fijo... pero en distintos instantes t).

Así, para cada par (k, t) , tenemos la covarianza $\gamma_{k,t}$ entre X_t y X_{t-k} . Por tanto, en general, tenemos una esperanza para cada t y una covarianza de orden k para cada t . Dado que t recorre todos los números enteros (son muchos momentos). Por eso necesitamos reducir el número de parámetros restringiéndonos a procesos estocásticos débilmente estacionarios.

2.3. Procesos estocásticos (débilmente) estacionarios y la ACF

Un proceso estocástico de segundo orden \mathbf{X} se dice que es *débilmente estacionario* (estacionario en covarianza o, sencillamente, estacionario) si:

- $E(X_t) = \mu$ para todo t y
- la covarianza entre X_s y X_t solo depende de la diferencia $s - t$.

En tal caso, llamamos función de autocovarianzas a la siguiente secuencia:

$$\boldsymbol{\gamma} = (\gamma_k \mid k \in \mathbb{Z}) = (\dots, \gamma_{-1}, \color{blue}{\gamma_0}, \gamma_1, \gamma_2, \dots) = \sum_{-\infty}^{\infty} \gamma_k z^k.$$

(nótese que esta secuencia sí incluye covarianzas para todos los órdenes $k \in \mathbb{Z}$).

Y se denomina matriz de autocovarianzas de \mathbf{X} a la matriz simétrica

$$\boldsymbol{\Gamma} = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Tanto la secuencia $\boldsymbol{\gamma}$ como la matriz $\boldsymbol{\Gamma}$ son **definidas positivas**; es decir, para todos los enteros $n \geq 1$ y escalares c_1, c_2, \dots, c_n

$$\sum_{i=1}^n \sum_{j=1}^n c_i c_j \gamma_{i-j} \geq 0$$

ya que

$$\sum_{i=1}^n \sum_{j=1}^n c_i c_j Cov(X_i, X_j) = Cov\left(\sum_{i=1}^n c_i X_i, \sum_{j=1}^n c_j X_j\right) = \left\langle \sum_{i=1}^n c_i X_i \mid \sum_{i=1}^n c_i X_i \right\rangle = \left\| \sum_{i=1}^n c_i X_i \right\|^2 \geq 0.$$

Esto es equivalente a que las submatrices principales de $\boldsymbol{\Gamma}$ son definidas positivas.

Es más, una secuencia $\boldsymbol{\gamma}$ es definida positiva si y solo si existe un espacio de Hilbert H y un proceso estocástico estacionario \mathbf{X} con $X_t \in H$ tales que $\gamma_k = Cov(X_t, X_{t-k})$ para todo $t, k \in \mathbb{Z}$ (Kolmogorov, 1941).

Propiedades de la función de autocovarianzas γ (ACF):

- $\gamma_0 \geq 0$
- la secuencia γ es definida positiva; y por tanto,
 - γ es simétrica: $\gamma_k = \gamma_{-k}$
 - γ es acotada: $|\gamma_k| \leq \gamma_0$

Y, si $\gamma_0 > 0$, llamamos función de autocorrelación (ACF) a la secuencia: $\rho = \frac{1}{\gamma_0}(\gamma) = \sum_{k \in \mathbb{Z}} \frac{\gamma_k}{\gamma_0} z^k$.

3. Notación: convolución y el operador retardo

Los procesos estocásticos se pueden sumar elemento a elemento y se pueden multiplicar por escalares. Si \mathbf{X} e \mathbf{Y} son dos procesos estocásticos y $a \in \mathbb{R}$, entonces

$$\mathbf{X} + \mathbf{Y} = (X_t + Y_t \mid t \in \mathbb{Z}) \quad \text{y} \quad a\mathbf{X} = (a(X_t) \mid t \in \mathbb{Z}).$$

El conjunto de procesos estocásticos con la suma y el producto por escalares es un espacio vectorial.

Si \mathbf{a} es una secuencia de números y \mathbf{X} un proceso estocástico tales que, para todo t ,

$$\text{la suma } \sum_{k=-\infty}^{\infty} a_k X_{t-k} \text{ converge;}$$

definimos el producto convolución (*) de \mathbf{a} con \mathbf{X} como el proceso estocástico:

$$\mathbf{a} * \mathbf{X} = \left(\sum_{k+s=t} a_k X_s \mid t \in \mathbb{Z} \right)$$

es decir

$$(\mathbf{a} * \mathbf{X})_t = \sum_{k+s=t} a_k X_s, \quad \text{para } t \in \mathbb{Z}.$$

Por tanto, cada elemento de $(\mathbf{a} * \mathbf{X})$ es una combinación de variables aleatorias de \mathbf{X} .

Si aplicamos el operador \mathbf{B} sobre un elemento de \mathbf{X} obtenemos el anterior:

$$\mathbf{B}X_t = X_{t1}, \quad \text{para } t \in \mathbb{Z}.$$

Por tanto, aplicando el operador retardo \mathbf{B} repetidamente tenemos

$$\mathbf{B}^k X_t = X_{tk}, \quad \text{para } t, z \in \mathbb{Z}.$$

Así, para el polinomio $\mathbf{a}(z) = a_0 + a_1 z + a_2 z^2 + a_3 z^3$, y el proceso estocástico \mathbf{Y} :

$$\begin{aligned} \mathbf{a}(\mathbf{B})Y_t &= (a_0 + a_1 \mathbf{B} + a_2 \mathbf{B}^2 + a_3 \mathbf{B}^3)Y_t \\ &= a_0 Y_t + a_1 Y_{t-1} + a_2 Y_{t-2} + a_3 Y_{t-3} \\ &= (\mathbf{a} * \mathbf{Y})_t, \quad \text{para } t \in \mathbb{Z} \end{aligned}$$

Y en general, si la suma $\sum_{k=-\infty}^{\infty} a_k Y_{t-k}$ converge para todo t , entonces

$$\begin{aligned}\mathbf{a}(\mathbf{B})Y_t &= (\cdots + a_{-2}\mathbf{B}^{-2} + a_{-1}\mathbf{B}^{-1} + a_0 + a_1\mathbf{B} + a_2\mathbf{B}^2 + \cdots)Y_t \\ &= \cdots + a_{-1}Y_{t+1} + a_0Y_t + a_1Y_{t-1} + a_2Y_{t-2} + \cdots = \sum_{k=-\infty}^{\infty} a_k Y_{t-k} \\ &= (\mathbf{a} * \mathbf{Y})_t, \quad \text{para } t \in \mathbb{Z}\end{aligned}$$

4. Ejemplos de procesos (débilmente) estacionarios

4.1. Proceso de ruido blanco

Una secuencia $\mathbf{U} = (U_t \mid t \in \mathbb{Z})$ de variables aleatorias **incoreladas** y tales que

$$E(U_t) = 0 \quad \text{y} \quad \text{Var}(U_t) = E(U_t^2) = \sigma^2$$

para $t \in \mathbb{Z}$ y $0 < \sigma^2 < \infty$ se llama *proceso de ruido blanco*. $\mathbf{U} \sim WN(0, \sigma^2)$.

Al ser variables aleatorias incorreladas, su función de autocovarianzas es

$$\gamma(z) = \sigma^2 z^0 = (\dots, 0, 0, \sigma^2, 0, 0, \dots)$$

- Es el proceso estacionario (no trivial) más sencillo.
- Este proceso es el pilar sobre el que definiremos el resto de ejemplos.

4.2. Procesos lineales

Sea $\mathbf{U} \sim WN(0, \sigma^2)$ y $\mathbf{b} \in \ell^2$; es decir, una secuencia de cuadrado sumable $\sum_{j \in \mathbb{Z}} b_j^2 < \infty$.

Denominamos *proceso lineal* al proceso estocástico $\mathbf{X} = \mathbf{b} * \mathbf{U}$ cuyos elementos son

$$X_t = (\mathbf{b} * \mathbf{U})_t = \mathbf{b}(B)U_t = \sum_{j=-\infty}^{\infty} b_j U_{t-j}, \quad \text{con } t \in \mathbb{Z}.$$

$\mathbf{b}(B)$ se denomina *función de transferencia* del filtro lineal que relaciona X_t con U_t .

El proceso está bien definido puesto que la serie infinita converge en norma por el Teorema de Riesz-Fisher (Pourahmadi, M. 2001, Teorema 9.7). Y el proceso es estacionario porque, usando la

continuidad de los productos escalares (Pourahmadi, M. 2001, Teorema 9.2),

$$\begin{aligned}
\gamma_k &= Cov(X_{t+k}, X_t) = \langle X_{t+k} | X_t \rangle = \lim_{m,n \rightarrow \infty} \left\langle \sum_{i=-m}^m b_i U_{t+k-i} | \sum_{j=-n}^n b_j U_{t-j} \right\rangle \\
&= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} b_i b_j \langle U_{t+k-i} | U_{t-j} \rangle \\
&= \sigma^2 \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} b_i b_j \delta_{t+k-i, t-j} \\
&= \sigma^2 \sum_{i=-\infty}^{\infty} b_i b_{i+k} = \sigma^2 (\mathbf{b}(z) * \mathbf{b}(z^{-1}))_k
\end{aligned} \tag{3}$$

que solo depende de k (en la tercera línea, $\delta_{p,q}$ es la delta de Kronecker; y en la cuarta hemos usado la última ecuación de la Lección 2).

El proceso lineal es “causal” si además \mathbf{b} es una serie formal (i.e., $cogrado(\mathbf{b}) \geq 0$)

$$X_t = \sum_{j=0}^{\infty} b_j U_{t-j}; \quad t \in \mathbb{Z}$$

(pues cada X_t es una suma de variables “del presente y/o el pasado”).

La clase de **procesos lineales causales** incluye muchas e importantes subclases de procesos, algunas de las cuales son objeto principal de estudio de este curso.

4.2.1. Media móvil infinita. MA(∞)

Sea $\mathbf{U} \sim WN(0, \sigma^2)$ y sea $\psi \in \ell^2$ una serie formal con infinitos términos NO nulos; entonces el proceso estocástico $\psi * \mathbf{U}$, cuyos elementos son

$$X_t = (\psi * \mathbf{U})_t = \psi(B)U_t = \sum_{j=0}^{\infty} \psi_j U_{t-j}; \quad t \in \mathbb{Z}$$

se denomina proceso de *media móvil infinita* MA(∞).

Algunas clases de procesos lineales causales poseen una representación parsimoniosa, pues basta un número finito de parámetros para describirlos completamente. Por ejemplo, cuando ψ tiene un número finito de términos no nulos...

4.2.2. Proceso de media móvil de orden q . MA(q)

Sea $\mathbf{U} \sim WN(0, \sigma^2)$ y sea θ un polinomio de grado (q) con $\theta_0 = 1$; entonces el proceso estocástico $\theta * \mathbf{U}$, cuyos elementos son

$$X_t = (\theta * \mathbf{U})_t = \theta(B)U_t = \sum_{j=0}^q \theta_j U_{t-j}; \quad t \in \mathbb{Z}$$

se denomina proceso de *media móvil* MA(q).

Es decir, si $\boldsymbol{\theta} = 1 - \theta_1 z - \cdots - \theta_q z^q$, tenemos que

$$X_t = U_t - \theta_1 U_{t-1} - \cdots - \theta_q U_{t-q}.$$

Hay otros procesos lineales con representación parsimoniosa.

4.2.3. Proceso autorregresivo de orden p . AR(p)

Sea $\mathbf{U} \sim WN(0, \sigma^2)$, se denomina *proceso autorregresivo de orden p* a aquel proceso estocástico *estacionario* \mathbf{X} que es solución de la siguiente ecuación

$$\phi * \mathbf{X} = \mathbf{U}$$

donde ϕ un polinomio de grado (p) con $\phi_0 = 1$.

Por tanto,

$$(\phi * \mathbf{X})_t = \phi(\mathbf{B})X_t = \sum_{j=0}^p \phi_j X_{t-j} = U_t.$$

Si $\phi = 1 - \phi_1 z - \cdots - \phi_p z^p$, entonces $\mathbf{X} = (X_t \mid t \in \mathbb{Z})$ es solución de la ecuación en diferencias:

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = U_t.$$

El problema con la anterior definición es que la ecuación $\phi * \mathbf{X} = \mathbf{U}$ no tiene solución única (y en algunos casos ninguna solución es estacionaria). Despejemos \mathbf{X} para verlo.

Multiplicando ambos lados de la ecuación por una inversa de ϕ tenemos

$$\mathbf{X} = \text{inversa}(\phi) * \mathbf{U}.$$

Y si denotamos la secuencia $\text{inversa}(\phi)$ con \mathbf{a} entonces

$$X_t = \mathbf{a}(\mathbf{B})U_t = \sum_{j \in \mathbb{Z}} a_j U_{t-j}.$$

Pero... ¿qué secuencia \mathbf{a} usamos como inversa de ϕ ? Recuerde que hay infinitas y la mayoría no son sumables (si el polinomio ϕ tiene raíces unitarias ninguna lo es).

En tal caso la expresión $\mathbf{a}(\mathbf{B})U_t = \sum_{j=-\infty}^{\infty} a_j U_{t-j}$ carece de sentido (pues la suma no converge).

Requisitos sobre el polinomio autorregresivo ϕ : para que el proceso AR exista y sea

1. lineal y estacionario, exigiremos que ϕ no tenga raíces de módulo 1.

Entonces existe una única inversa absolutamente sumable: $\phi^{-1} \in \ell^1 \subset \ell^2$.

La inversa ϕ^{-1} corresponde a la única solución *estacionaria* de $\phi * \mathbf{X} = \mathbf{U}$. (Si ϕ tuviera raíces de módulo 1 no existiría ni $\phi^{-1} \in \ell^1$, ni solución estacionaria).

$$X_t = \phi^{-1}(\mathbf{B})U_t = \sum_{j=-\infty}^{\infty} a_j U_{t-j}$$

2. causal, exigiremos que las raíces de ϕ sean mayores que 1 en valor absoluto (raíces fuera del círculo unidad): $\phi^{-1} = \phi^{-\triangleright}$ (**serie formal** $\in \ell^1 \subset \ell^2$).

$$X_t = \phi^{-1}(\mathbf{B})U_t = \sum_{j=0}^{\infty} a_j U_{t-j}; \quad \text{donde } \mathbf{a} = \phi^{-1}.$$

(¡de nuevo un proceso lineal causal!)

El siguiente modelo lineal es una generalización de los dos anteriores.

4.2.4. Proceso autorregresivo de media móvil. ARMA(p, q)

Sea $\mathbf{U} \sim WN(0, \sigma^2)$, se denomina *proceso autorregresivo de media móvil* (p, q) al proceso estocástico estacionario \mathbf{X} que es solución de la ecuación en diferencias:

$$\phi * \mathbf{X} = \theta * \mathbf{U}$$

donde el polinomio *autorregresivo* ϕ tiene grado p con $\phi_0 = 1$ y con todas sus raíces fuera del círculo unidad (*por los motivos anteriormente vistos*); y el polinomio *de media móvil* θ es de grado q con $\theta_0 = 1$;

$$\text{es decir, } \mathbf{X} = \frac{\theta}{\phi} * \mathbf{U}; \quad \text{donde } \frac{\theta}{\phi} \equiv \phi^{-1} * \theta$$

Tanto ϕ^{-1} como θ son series formales absolutamente sumables. Dado que tanto ℓ^1 como las series formales son anillos: $\phi^{-1} * \theta$ es otra una serie formal absolutamente sumable (y, por lo tanto, de cuadrado sumable). Consecuentemente, \mathbf{X} es, nuevamente, un proceso estocástico **lineal causal**.

$$X_t = \frac{\theta}{\phi}(\mathbf{B})U_t = \sum_{j=0}^{\infty} a_j U_{t-j}; \quad \text{donde } \mathbf{a} = \frac{\theta}{\phi}.$$

4.2.5. Proceso autorregresivo de media móvil con media no nula

Consideremos un proceso \mathbf{Y} con media distinta de cero, es decir,

$$E(Y_t) = \mu \neq 0$$

y definamos la secuencia constante $\boldsymbol{\mu} = \sum_{j \in \mathbb{Z}} \mu z^j = (\dots, \mu, \mu, \mu, \dots)$.

Decimos que \mathbf{Y} es un proceso ARMA(p, q) con media distinta de cero si \mathbf{X} es ARMA(p, q)

$$\boldsymbol{\phi} * \mathbf{X} = \boldsymbol{\theta} * \mathbf{U}$$

donde $\mathbf{X} = \mathbf{Y} - \boldsymbol{\mu}$ es evidentemente un proceso de media cero. Por tanto

$$\begin{aligned}\boldsymbol{\phi} * (\mathbf{Y} - \boldsymbol{\mu}) &= \boldsymbol{\theta} * \mathbf{U} \\ \boldsymbol{\phi} * \mathbf{Y} - \boldsymbol{\phi} * \boldsymbol{\mu} &= \boldsymbol{\theta} * \mathbf{U} \\ \boldsymbol{\phi} * \mathbf{Y} &= \boldsymbol{\phi} * \boldsymbol{\mu} + \boldsymbol{\theta} * \mathbf{U}\end{aligned}$$

Es decir, si $\boldsymbol{\phi}(B)$ es $1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$, entonces

$$\boldsymbol{\phi}(B)Y_t = c + \boldsymbol{\theta}(B)U_t$$

donde

$$c = (1 - \phi_1 - \phi_2 - \dots - \phi_p)\mu$$

y donde $\mu = E(Y_t)$, es un proceso autorregresivo de media móvil ARMA(p, q) *con media no nula*.

5. Primeros momentos de procesos lineales causales

5.1. Esperanza y autocovarianzas de un proceso lineal causal

Sea $\mathbf{X} = \boldsymbol{\psi} * \mathbf{U}$, donde $\boldsymbol{\psi}$ es una serie formal de cuadrado sumable ($\boldsymbol{\psi} \in \ell^2$) y donde $\mathbf{U} \sim WN(0, \sigma^2)$. Recordando que la convolución es una operación lineal:

$$E(\mathbf{X}) = E(\boldsymbol{\psi} * \mathbf{U}) = \boldsymbol{\psi} * E(\mathbf{U}) = \boldsymbol{\psi} * \mathbf{0} = \mathbf{0}.$$

Así, la covarianza de orden k para cada X_t es (demo en el [pdf](#) —ecuación (3)):

$$\gamma_{k,t} = E[(\boldsymbol{\psi}(B)X_t) \cdot (\boldsymbol{\psi}(B)X_{t-k})] = \sigma^2 (\boldsymbol{\psi}(z) * \boldsymbol{\psi}(z^{-1}))_k$$

que no depende de t (\mathbf{X} es estacionario), es decir, $\gamma_{k,t} = \gamma_k$. Y, por tanto

$$\gamma = \sigma^2 \boldsymbol{\psi}(z) * \boldsymbol{\psi}(z^{-1}). \tag{4}$$

5.2. Covarianza cruzada entre dos procesos lineales causales

Sean $\mathbf{W} = \boldsymbol{\theta} * \mathbf{U}$ e $\mathbf{Y} = \boldsymbol{\psi} * \mathbf{U}$, donde $\boldsymbol{\theta}$ y $\boldsymbol{\psi}$ son series formales de cuadrado sumable y donde $\mathbf{U} \sim WN(0, \sigma^2)$.

Repetiendo los mismos pasos que en el caso de la autocovarianza, llegamos a que la función de covarianzas cruzadas es la secuencia

$$\gamma_{\mathbf{W}, \mathbf{Y}} = \sigma^2 \boldsymbol{\theta}(z) * \boldsymbol{\psi}(z^{-1}) \quad (5)$$

con grado igual al grado de $\boldsymbol{\theta}$ y cogrado igual a menos el grado de $\boldsymbol{\psi}$.

5.3. Las Ecuaciones de Yule-Walker para un AR(p) estacionario

Por una parte (lado izquierdo):

Si \mathbf{X} es un proceso (débilmente) estacionario con $E(\mathbf{X}) = \mathbf{0}$ y ϕ es una serie formal absolutamente sumable; entonces para $t, k \in \mathbb{Z}$

$$E[(\phi(\mathbf{B})X_t) \cdot X_{t-k}] = \phi(\mathbf{B})E(X_t \cdot X_{t-k}) = \phi(\mathbf{B})\gamma_k \quad (6)$$

que no depende de t , por ser \mathbf{X} es un proceso (débilmente) estacionario.

Por otra parte (lado derecho):

Si \mathbf{X} tiene representación $\mathbf{X} = \boldsymbol{\psi} * \mathbf{U}$ donde $\mathbf{U} \sim WN(0, \sigma^2)$ y $\boldsymbol{\psi} \in \ell^2$ es una serie formal con $\psi_0 = 1$; es decir, si es un proceso lineal causal

$$X_t = U_t + \sum_{j=1}^{\infty} \psi_j U_{t-j},$$

entonces para $t, k \in \mathbb{Z}$

$$E[U_t \cdot X_{t-k}] = E[U_t \left(U_{t-k} + \sum_{j=1}^{\infty} \psi_j U_{t-k-j} \right)] = \begin{cases} \sigma^2 & \text{cuando } k = 0 \\ 0 & \text{cuando } k \neq 0 \end{cases} \quad (7)$$

Sea un AR(p) estacionario: $\phi(\mathbf{B})X_t = U_t$ donde $\phi(z) = 1 - \phi_1 z^1 - \cdots - \phi_p z^p$. Multiplicando por X_{t-k} y tomando esperanzas:

$$E[(\phi(\mathbf{B})X_t) \cdot X_{t-k}] = E[U_t \cdot X_{t-k}]$$

para $k = 0$: (por 6 y 7)

$$\boxed{\phi(\mathbf{B})\gamma_0 = \sigma^2} \Rightarrow \gamma_0 - \phi_1\gamma_1 - \cdots - \phi_p\gamma_p = \sigma^2 \Rightarrow \sigma^2 = \gamma_0 - \sum_{j=1}^p \phi_j\gamma_j.$$

Dividiendo por γ_0 (y recordando que $\rho_0 = 1$):

$$\phi(\mathbf{B})\rho_0 = \frac{\sigma^2}{\gamma_0} \Rightarrow \boxed{\gamma_0 = \frac{\sigma^2}{\phi(\mathbf{B})\rho_0}} \Rightarrow \gamma_0 = \frac{\sigma^2}{1 - \sum_{j=1}^p \phi_j\rho_j}.$$

para $k > 0$: (por 6 y 7)

$$\boxed{\phi(\mathbf{B})\gamma_k = 0} \Rightarrow \gamma_k - \phi_1\gamma_{k-1} - \cdots - \phi_p\gamma_{k-p} = 0 \Rightarrow \gamma_k = \sum_{j=1}^p \phi_j\gamma_{k-j}.$$

Dividiendo por γ_0 :

$$\boxed{\phi(\mathbf{B})\rho_k = 0} \Rightarrow \rho_k - \phi_1\rho_{k-1} - \cdots - \phi_p\rho_{k-p} = 0 \Rightarrow \rho_k = \sum_{j=1}^p \phi_j\rho_{k-j}.$$

Por tanto, la estructura autorregresiva del proceso impone que las autocovarianzas (y las autocorrelaciones) verifiquen las ecuaciones de Yule-Walker.

5.4. Función de autocovarianzas para un ARMA(p, q)

Sea un ARMA(p, q) estacionario: $\phi(\mathbf{B})X_t = \theta(\mathbf{B})U_t$ donde ϕ y θ no tienen raíces comunes. Multiplicando por X_{t-k} , tomando esperanzas y sustituyendo X_{t-k} por su representación MA(∞), donde $\psi = \frac{\theta}{\phi}$:

$$\underbrace{E\left[\left(\phi(\mathbf{B})X_t\right) \cdot X_{t-k}\right]}_{\phi(\mathbf{B})\gamma_k \text{ (por 6)}} = E\left[\left(\theta(\mathbf{B})U_t\right) \cdot X_{t-k}\right] = \underbrace{E\left[\left(\theta(\mathbf{B})U_t\right) \cdot \left(\psi(\mathbf{B})U_{t-k}\right)\right]}_{\gamma_{W,Y}(k)}$$

Donde hemos usando (6) y renombrando $\theta(\mathbf{B})U_t = \mathbf{W}$ y $\psi(\mathbf{B})U_t = \mathbf{Y}$. Así:

$$\begin{aligned} \phi(\mathbf{B})\gamma_k &= \gamma_{W,Y}(k) \\ &= \sigma^2 \left(\theta(z) * \psi(z^{-1}) \right)_k \quad \text{por (5)} \end{aligned}$$

Y como $\theta(z) * \psi(z^{-1})$ tiene grado q y cogrado $-\infty$

$$\phi(\mathbf{B})\gamma_k = \begin{cases} 0 & k > q \quad (\text{como en un AR}) \\ \sigma^2 \left(\theta(z) * \psi(z^{-1}) \right)_k & k \leq q \quad (\text{que depende de } \theta \text{ y } \phi) \end{cases} \quad (8)$$