MATEMÁTICAS

ESTADÍSTICA III

SEXTO SEMESTRE

UNIDAD 1. PROCESOS Y SERIES DE TIEMPO

Clave
05143633

Universidad Abierta y a Distancia de México

DCEIT

Índice

Unidad 1. Procesos y series de tiempo

Presentación de la Unidad

En esta unidad se revisarán las definiciones de proceso estocástico de de segundo orden, procesos estacionarios y procesos estacionarios de segundo orden. Se verá que esta última clase de procesos tienen la característica de tener valor esperado y sucesión de autocovarianza que no cambian con el índice de tiempo.

Teniendo en mente esta característica de los momentos de primer y segundo orden para un proceso estacionario de segundo orden, se revisarán los procesos de ruido blanco, autorregresivo y de promedios móviles, estudiando las condiciones bajo las cuales estos procesos son estacionarios de segundo orden. Este análisis permitirá identificar las propiedades estadísticas de los modelos lineales de series de tiempo, conocidos en literatura como procesos ARMA(p, q).

Competencia específica

Identificar las propiedades estadísticas de los modelos de series de tiempo estacionarias para caracterizar los modelos pertenecientes a la clase ARMA (p,q), mediante la definición de las estructuras: autorregresiva, de promedios móviles y de ruido blanco.

Logros

- **Identificar** las nociones de proceso estocástico estacionario y estacionario de segundo orden.
- **Identificar** las relaciones entre los diferentes tipos de estacionariedad.
- **Determinar** las condiciones sobre los parámetros de cada proceso, que hacen que éste sea estacionario de segundo orden.

1.1. Procesos de segundo orden y series de tiempo

La medición de una cantidad de interés a lo largo de varios instantes en el tiempo es un proceso que emerge en forma natural en diferentes ámbitos de estudio. En economía, si te interesa la evolución de algún índice, se consideran mediciones de este registradas, por ejemplo, cada mes durante varios años. Para este ejemplo pueden tenerse en mente varios objetivos como saber si hay un patrón que se repite en intervalos de tiempo con la misma longitud (existencia de ciclos). También puede ser de interés predecir el valor del índice una unidad de tiempo después de la última observación. Otro ejemplo se da en estudios ecológicos, donde se tienen mediciones anuales de la temperatura global del planeta tierra y la pregunta a contestar es si existe una tendencia de estas observaciones a tener mayor magnitud a lo largo de diferentes tiempos.

La metodología estadística para trabajar con observaciones de algún fenómeno, hechas en varios tiempos, se conoce como series de tiempo.

Desde un punto de vista teórico, el modelo matemático para pensar en una serie de observaciones indexadas en el tiempo es un proceso estocástico. A continuación estudiarás brevemente este tema.

En esta unidad, T denota al **conjunto índice de tiempo,** conjunto que se toma como un modelo de la colección de tiempos en los que se observan mediciones con respecto a un fenómeno de intéres. Para tus objetivos, el conjunto T puede estar dado por los números enteros $T = \{\dots, -1,0,1, \dots\}$ o bien, por el conjunto de los números naturales $T = \{0,1,2, \dots\}$.

Un proceso estocástico es una colección de variables aleatorias indexadas en el conjunto T

$$\{X_t : t \in T\}.$$

Para cada $t \in T$, la variable aleatoriaX_t está definida en el espacio de probabilidad $(\Omega, F, \mathbb{P})$.

Una serie de tiempo es una realización del proceso $\{X_t : t \in T\}$, es decir, para $w \in \Omega$

$$x_{t_1} = x_{t_1}(w), x_{t_2} = x_{t_2}(w), \ldots$$

y las observaciones $\{x_t : t \in T\}$ son la serie de tiempo.

Recuerda que la información necesaria para describir el comportamiento de una variable aleatoria X_t, está dada en su función de distribución de probabilidades

$$F_t(a) = I = \mathbb{P}\ (X_t \leq a).$$

1.1.1. Procesos de segundo orden

Los procesos estocásticos con los cuales vas a trabajar tienen caractersticas basadas en sus momentos de primer y segundo orden. Por esta razón, resultará importante revisar algunas nociones referentes variables aleatorias con segundo momento finito.

Ahora, supóngase que para cada variable aleatoria indexada en T, existen su valor medio $\mathbb{E}\ (X_t)$ y su varianza $VAR(X_t)$. Usando notación de integral de Stieltjes

$$\mathbb{E}(X_t) = \int_{-\infty}^{\infty} x dF_t(x) \equiv \mu_t \text{ y} \tag{1}$$

$$VAR\ (X_t) = \int_{-\infty}^{\infty} (x - \mu_t)^2 dF_t(x) \equiv \sigma_t^2. \tag{2}$$

Para ti serán de interés los procesos estocásticos para los cuales la media y la varianza, definidas por estas expresiones, existen. Recuerda que una condición equivalente a la existencia de (1) y de (2) es la existencia de segundo momento finito

Lema 1.1 $VAR\ (X_t)$ y $\mathbb{E}\ (X_t)$ existen si y sólo si $\mathbb{E}\ (X_t^2)$ existe.

Demostración: Recuerda que si $IE\ (X_t^2)$ es finita, entonces $IE\ (X_t)$ es finita, ya que

$$\infty > \int_{-\infty}^{\infty} x^2 f_{X_t}(x)dx \geq \int_{|x|>1} x^2 f_{X_t}(x)dx > \int_{|x|>1} x f_{X_t}(x)dx$$

y como $\int_{|x|\leq 1} x f_{X_t}(x)dx \leq \int_{|x|\leq 1} f_{X_t}(x)dx < \infty$, entonces $\mathbb{E}\ (X_t)$ es finita.

Por otra parte,

$$VAR\ (X_t) = \mathbb{E}\ (X_t^2) - (\ \mathbb{E}(X_t))^2,$$

es decir, la varianza de X_t es la suma de dos cantidades finitas. Entonces $VAR\ (X_t)$ y $\mathbb{E}\ (X_t)$ son finitos. Inversamente, si $VAR\ (X_t)$ y $\mathbb{E}(X_t)$ son finitos, entonces, como $\mathbb{E}[X_t^2] = \quad VAR\ (X_t) + (\ \mathbb{E}\ (X_t))^2$, se tiene que $\mathbb{E}\ (X_t^2)$ es finito.

Se dice que el proceso $\{X_t\}_t$ es un *proceso de* 2^o *orden* si $\forall t \in T, \mathbb{E}(X_t^2) < \infty$.

1.1.2. Procesos estacionarios

Será difícil trabajar con cualquier proceso estocástico arbitrario para tratar de describir fenómenos en la naturaleza. Por esta razón, en una primera forma de restringir la búsqueda, se propondrá que los procesos con los que se trate tengan una característica de "invarianza" de sus propiedades estocásticas en el tiempo. Esta forma de invarianza se conoce como estacionariedad, y básicamente se tiene que los procesos estacionarios son aquellos para los cuales las distribuciones de probabilidades asociadas a cada variable dentro del proceso no cambian con el índice de tiempo.

Sea $\{X_t\}_{t\in T}$ un proceso estocástico, y para cada $n \in \{1,2, \dots\}$, tomes $t_1, t_2, \dots, t_n \in T$. Las distribuciones de probabilidad conjuntas del vector aleatorio $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ están dadas por

$$F_{t_1,t_2,\ldots,t_n}(a_1,a_2,\ldots,a_n) = \mathbb{P}(X_{t_1} \leq a_1;\ldots;X_{t_n} \leq a_n). \text{ y} \tag{3}$$

Estas distribuciones se llaman **distribuciones de dimensión finita** del proceso $\{X_t\}_t$.

Definición 1. [Proceso estacionario]

Un proceso $\{X_t\}_{t\in T}$ se llama **estacionario** si para todo $n \geq 1$, para cualesquiera $t_1, t_2, \ldots, t_n$ elementos de T, y para todo τ, tal que $t_1+\tau, t_2+\tau, \ldots, t_n+\tau \in T$, la función de distribución conjunta del vector $(X_{t_1},\ldots,X_{t_n})$ es igual a la distribución conjunta del vector $(X_{t_1+\tau}, X_{t_2+\tau}, \ldots, X_{t_n+\tau})$

$$\begin{aligned} F_{t_1,\ldots,t_n}(a_1,\ldots,a_n) &= \mathbb{P}\,(X_{t_1} \leq a_1;\ldots;X_{t_n} \leq a_n) \\ &= \mathbb{P}\,(X_{t_1+\tau} \leq a_1;\ldots;X_{t_n+\tau} \leq a_n) \\ &= F_{t_1+\tau,\ldots,t_n+\tau}(a_1,\ldots,a_n). \end{aligned}$$

En otras palabras, las propiedades estocásticas de un proceso estacionario no cambian al pasar el tiempo. Para comprender mejor esta definición, considera $n=1$ y $\tau=1$. Entonces, si $T = \{0,1,\ldots\}$ y el proceso es estacionario, la distribución de las variables X_0 y X_1 es la misma, pero tambien la distribución de X_1 y X_2 es la misma. Esta situación se ilustra en la figura 1.1, en donde el eje horizontal representa el tiempo t, y el eje vertical el valor de x_t. Asimismo, se representan las parejas (t, x_t) con un carácter $'\star'$. Para cada tiempo t, la distribución de X_t es la misma (como se muestra en la figura 2).

Algunos autores usan el nombre **completamente estacionario o estrictamente estacionario** para este tipo de procesos.

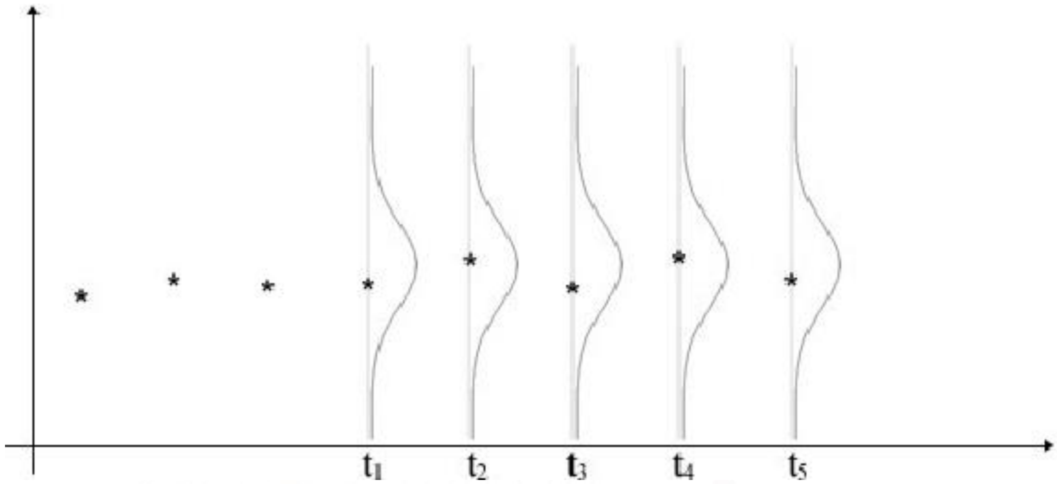

Figura 1: Proceso completamente estacionario

1.1.3. Procesos estacionarios de segundo orden

Como un modelo para datos reales, los procesos estacionarios resultan restrictivos, ya que se está pidiendo que la distribución de cada variable X_t sea la misma, independientemente de en qué tiempo t se observa. A continuación se relaja la definición de proceso completamente estacionario al pedir condiciones que pueden ser cumplidas más fácilmente por datos reales.

Definición 2. [Proceso estacionario de $2^{\mathbf{o}}$ orden]

Un proceso de 2^o orden $\{X_t\}_{t\in T}$ se llama **débilmente estacionario o estacionario de $2^{\mathbf{o}}$ orden** si:

Para cada $n \in \{1,2,\dots\}$, para cualesquiera $t_1, t_2, \dots, t_n \in T$ y para todo τ, tal que $t_1 + \tau, \dots, t_n + \tau \in T$, todos los momentos de orden 1 y 2 del vector $(X_{t_1}, \dots, X_{t_n})$ son iguales a los correspondientes momentos de orden 1 y 2 del vector $(X_{t_1+\tau}, X_{t_2+\tau}, \dots, X_{t_n+\tau})$, es decir,

$$\mathbb{E}\,[X_{t_1}^{r_1} \cdots X_{t_n}^{r_n}] = \mathbb{E}[X_{t_1+\tau}^{r_1} \cdots X_{t_n+\tau}^{r_n}], \tag{4}$$

donde $r_i \in \{0,1,2\}$ y $\sum_{i=1}^{n} r_i \leq 2$.

La definición 2 debilita la definición 1, en el sentido de que para un proceso estacionario de segundo orden la distribución de X_t puede ser diferente a la distribución de X_{t+1}, pero los momentos $\mathbb{E}(X_{t_1} X_{t_2}), \mathbb{E}(X_t^2)$ y $\mathbb{E}(X_t)$ son los mismos que los momentos $\mathbb{E}\,(X_{t_1+\tau} X_{t_2+\tau}), \mathbb{E}(X_{t+\tau}^2)$ y $\mathbb{E}\,(X_{t+\tau})$. Entonces, si por ejemplo, $\tau = 1,2, \dots$, se tendrá que

$$\mathbb{E}(X_t) = \mathbb{E}\,(X_{t+1}) = \mathbb{E}\,(X_{t+2}) = \cdots, \tag{5}$$

es decir, la media del proceso es constante. Análogamente, para $IE\,(X_{t+\tau}^2)$ la estacionariedad de segundo orden implica que

$$\mathbb{E}(X_t^2) = \mathbb{E}\,(X_{t+1}^2) = \mathbb{E}\,(X_{t+2}^2) = \cdots \tag{6}$$

Nota que de (5) y (6) se sigue que, para un proceso estacionario de segundo orden $\{X_t\}_t$

$$\mathbb{E}\,(X_t) = \mu, \quad \forall t \in T \text{ y } VAR(X_t) = \sigma^2, \quad \forall t \in T,$$

es decir, media y varianza son cantidades que no dependen de t. En forma análoga, se sigue de (4) que

$$\begin{aligned} COV[X_t, X_s] &= \mathbb{E}(X_t X_s) - \mathbb{E}(X_t)\mathbb{E}(X_s) = \mathbb{E}(X_t X_s) - \mu^2 \\ &= \mathbb{E}\,(X_{t+\tau} X_{s+\tau}) - \mathbb{E}(X_{t+\tau})\mathbb{E}(X_{s+\tau}) = COV[X_{t+\tau}, X_{s+\tau}]. \end{aligned}$$

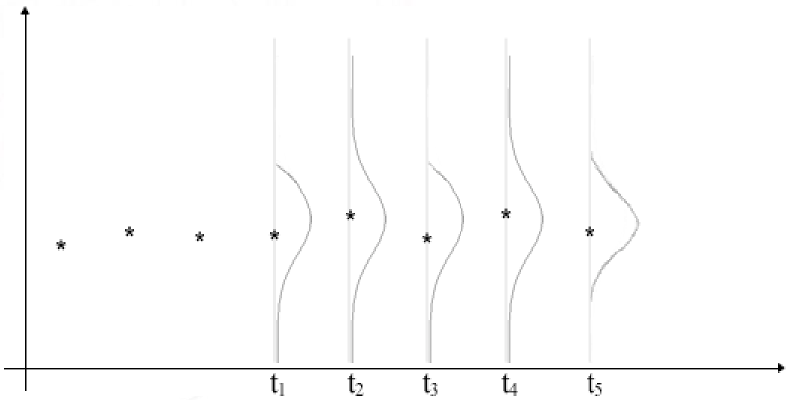

Figura 2: Proceso estacionario de segundo orden

1.1.4. Ejemplos

Ejemplo (Variables aleatorias independientes e idénticamente distribuidas con segundo momento finito)

Sea $X_1, X_2, \dots$ una sucesión de variables aleatorias independientes e idénticamente distribuidas, tal que $\mathbb{E}[X_i^2] < \infty, \forall i = 1,2,3, \dots$, entonces $\{X_i\}_{i=1}^{\infty}$ es fuertemente estacionario y estacionario de 2^o orden.

Para el siguiente ejemplo se necesita una definición:

Definición 3. [Matriz no negativo definida]
Una matriz simétrica Σ de dimensiones $n \times n$ se llama no negativo definida si para cada a'=$(a_1, \dots, a_n) \in \mathbb{R}^n$ se tiene que $a'\Sigma a \geq 0$.

Dado un vector aleatorio $(X_{t_1}, \dots, X_{t_n})$ de dimensión n, su **matriz de varianzas covarianzas** es la matriz Σ_X de dimensiones $n \times n$, simétrica y con entrada i, j dada por $COV[X_{t_i}, X_{t_j}]$ (la covarianza entre X_{t_i} y X_{t_j}) para $i, j = 1,2, \dots n$.

Los procesos Gaussianos son una clase muy importante de procesos de segundo orden. Se caracterizan por el hecho de que sus distribuciones de dimensión finita son distribuciones normales multivariadas. Como la distribución normal multivariada esta caracterizada por un vector de medias μ y una matriz de varianzas covarianzas Σ, al establecer condiciones sobre estas dos cantidades se pueden encontrar ejemplos de procesos estacionarios

Ejemplo (Proceso Gaussiano)
Sea $X_1, X_2, \dots, X_n, \dots$ un proceso Gaussiano, es decir, $\{X_t\}_{t=1}^{\infty}$ es un proceso de 2^o orden tal que para cada $n = 1,2, \dots$ y para cada $t_1, t_2, \dots, t_n \in \{1,2, \dots\} = T$, existen $\mu \in \mathbb{R}^n$ y Σ una matriz de dimensiones $n \times n$, simétrica y no negativo definida tales que

$$\mathbb{P}\left[X_{t_1} \leq a_1; X_{t_2} \leq a_2; \dots; X_{t_n} \leq a_n\right] =$$

$$= \int_{-\infty}^{a_1} \cdots \int_{-\infty}^{a_n} \frac{1}{(2\pi)^{n/2}} |\Sigma|^{-\frac{1}{2}} \exp(-\frac{1}{2}(y - \mu\mu)'\Sigma^{-1}(y - \mu\mu))dy_1 \cdots dy_n \qquad (7)$$

donde $y' = (y_1, \dots, y_n)$ y $\mu' = (\mu_1, \dots, \mu_n)$.

Para construir un segundo ejemplo de un proceso que es estacionario fuerte y estacionario de 2^o orden se fija

$$\mathbb{E}\,[X_t] = 0\ \forall t,$$

y

$$COV[X_{t_i}, X_{t_j}] = \begin{cases} 1 & i = j; \\ \frac{1}{2} & i \neq j,, \end{cases} \tag{8}$$

de donde $\mu\ = 0$ y Σ es una matriz con todos los elementos de la diagonal igual a 1 y todos los elementos fuera de la diagonal igual a $1/2$. El proceso $\{X_t\}_t$ construido así es fuertemente estacionario en virtud de que μ y Σ caracterizan a la distribución Normal y en este ejemplo no dependen de $t_1, \dots t_n, \forall n = 1,2, \dots$, por lo que se sigue que las distribuciones de $(X_{t_1}, \dots, X_{t_n})$ y de $(X_{t_1+\tau}, \dots, X_{t_n+\tau})$ son iguales.

Por otra parte, $\{X_t\}_t$ es estacionario de 2^o orden, ya que $IE\ [X_t] = 0, \forall t$ y de la ecuación (8)

$$\begin{aligned} \mathbb{E}\,[X_{t_1+\tau} X_{t_2+\tau}] &= COV[X_{t_1+\tau}, X_{t_2+\tau}] + \mathbb{E}[X_{t_1+\tau}]\, IE\,[X_{t_2+\tau}] \\ &= COV[X_{t_1+\tau}, X_{t_2+\tau}] \\ &= COV[X_{t_1}, X_{t_2}] \\ &= COV[X_{t_1}, X_{t_2}] + IE\,[X_{t_1}]\, IE\,[X_{t_2}] \\ &= IE\,[X_{t_1} X_{t_2}]. \end{aligned}$$

Nota que Σ, definida a través de la ecuación (8), es no negativa definida; un argumento para verificarlo se da a continuación. Si $T = \{0,1, \dots\}$ y para $n \geq 1$ tomas $a \in \mathbb{R}^n$, puedes calcular

$$
\begin{aligned}
a^t\Sigma a &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i\Sigma_{ij}a_j \\
&= \sum_{i=1}^{n}\{a_i^2 + \frac{1}{2}\sum_{i\neq j} a_ia_j\} \\
&= \sum_{i=1}^{n} a_i^2 + \frac{1}{2}\cdot 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_ia_j \\
&= \sum_{i=1}^{n} a_i^2 + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_ia_j.
\end{aligned}
$$

Se tienen dos escenarios

Caso 1: Si $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_ia_j \geq 0$, entonces $a^t\Sigma a \geq 0$.

Caso 2: Si $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_ia_j < 0$, entonces $2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_ia_j < \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_ia_j < 0$, de donde

$$
\begin{aligned}
0 \leq (a_1 + \cdots + a_n)(a_1 + \cdots + a_n) &= \sum_{i=1}^{n} a_i^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_ia_j \\
&< \sum_{i=1}^{n} a_i^2 + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_ia_j.
\end{aligned}
$$

1.1.5. Relaciones entre las formas de estacionariedad

Es natural que te preguntes, ¿cuál es la relación entre estacionariedad y estacionariedad de segundo orden?

Para responder a esta pregunta, se puede mencionar lo siguiente:

En general se tendrá que:

(i) Si $\{X_t\}_t$ es estacionario fuerte y además es un proceso de 2^o orden, entonces $\{X_t\}_t$ es estacionario de 2^o orden, o bien, estacionariedad fuerte y existencia de momentos de 2^o orden implica estacionariedad débil.

Sin embargo,

(ii) Si $\{X_t\}_t$ es estacionario de 2^o orden, entonces no necesariamente $\{X_t\}_t$ es fuertemente estacionario. Se probará (i) en el caso de variables aleatorias absolutamente continuas.

Asumiendo estacionariedad fuerte, se tiene:

$$\begin{aligned} F_t(a) \quad &:= \mathbb{P}\left(X_t \leq a\right) \\ &= \mathbb{P}\left(X_{t+\tau} \leq a\right) \\ &:= F_{t+\tau}(a), \end{aligned}$$

donde la segunda igualdad se sigue de la definición de estacionariedad fuerte. Sean $f_{X_t}(x)$ la densidad de la $v.a.$ X_t y t, τ tales que $t, t+\tau \in T$, entonces

$$\begin{aligned} \mu_t = \mathbb{E}\left(X_t\right) \quad &= \int_{-\infty}^{\infty} x f_{X_t}(x) dx \\ &= \int_{-\infty}^{\infty} x f_{X_{t+\tau}}(x) dx, \end{aligned}$$

ya que al tener X_t y $X_{t+\tau}$ la misma función de distribución, entonces X_t y $X_{t+\tau}$ tienen la misma función de densidad. Por lo tanto,

$$\begin{aligned} \mu_t \quad &= \int_{-\infty}^{\infty} x f_{X_{t+\tau}}(x) dx \\ &= \mathbb{E}\left[X_{t+\tau}\right] = \mu_{t+\tau}. \end{aligned}$$

Como lo anterior se vale para cualesquiera t y τ tales que t y $t+\tau \in T$, se puede concluir que $\mu_t = \mu \forall t \in T$. Ahora hay que demostrar que la varianza no cambia con traslaciones del tiempo, es decir:

$$\begin{aligned}\sigma_t^2 &= \mathbb{E}\left[(X_t-\mu)^2\right]\\ &= \int_{-\infty}^{\infty}(x-\mu)^2 f_{X_t}(x)dx\\ &= \int_{-\infty}^{\infty}(x-\mu)^2 f_{X_{t+\tau}}(x)dx\\ &= \mathbb{E}\left[(X_{t+\tau}-\mu)^2\right] = \sigma_{t+\tau}^2,\end{aligned}$$

por lo tanto $\sigma_t^2 = \sigma^2 \ , \ \forall t \in T$.

Por último, denota por $f_{X_{t_1},X_{t_2}}$ a la función de densidad conjunta de (X_{t_1}, X_{t_2})

$$\begin{aligned}\mathbb{E}[X_{t_1}X_{t_2}] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X_{t_1},X_{t_2}}(x,y)dxdy\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X_{t_1+\tau},X_{t_2+\tau}}(x,y)dxdy,\end{aligned}$$

ya que la definición de estacionariedad fuerte dice que

$$\begin{aligned}\mathbb{P}\left(X_{t_1} \le a; X_{t_2} \le b\right) &:= F_{t_1 t_2}(a,b)\\ &= F_{t_1+\tau, t_2+\tau}(a,b)\\ &= \mathbb{P}\left(X_{t_1+\tau} \le a; X_{t_2+\tau} \le b\right)\end{aligned}$$

y por lo tanto, las correspondientes densidades bivariadas son iguales.

Entonces

$$\mathbb{E}\left[X_{t_1}X_{t_2}\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X_{t_1+\tau},X_{t_2+\tau}}(x,y)dxdy = \mathbb{E}\left[X_{t_1+\tau}X_{t_2+\tau}\right].$$

La anterior discusión se puede elaborar en forma análoga en el caso de variables aleatorias $\{X_t\}_t$ discretas.

Ejemplo

(De acuerdo a ii, estacionarionariedad de 2^{o} orden, no necesariamente implica estacionariedad)

Sea $\{X_n\}_{n\geq 1}$ una sucesión de variables aleatorias independientes tales que $X_{2n}\sim\ exponencial(1)$ y $X_{2n+1}\sim\ \mathrm{N}ormal(1{,}1)$. Entonces se tiene que $\mathbb{E}[X_{2n}] = 1 = VAR\ (X_{2n})$ y $\mathbb{E}[X_{2n+1}] = 1$

$$= \quad VAR\ (X_{2n+1}).$$

Claramente $\{X_n\}_n$ no es fuertemente estacionario, ya que para cada $n, F_{X_{2n}} \neq F_{X_{2n+1}}$, pero se tiene que $\mathbb{E}\ [X_{2n+1}] = \ \mathbb{E}\ [X_{2n}]$; tambien que $VAR\ (X_{2n+1}) = \quad VAR\ (X_{2n})$ y que $\mathbb{E}\ [X_n X_{n+1}] = \ \mathbb{E}\ [X_n]\mathbb{E}[X_{n+1}] = 1 \cdot 1 = 1$, por ser $\{X_n\}_n$ variables independientes. De esta forma $COV[X_{t_1}, X_{t_2}] = COV[X_{t_1+\tau}, X_{t_2+\tau}]$. Por lo tanto, $\{X_n\}_{n\geq 1}$ si es estacionario de segundo orden.

1.2. La sucesión de autocorrelación

Por ser menos restrictivos en sus supuestos, se trbajarán los procesos estacionarios de segundo orden. Por lo cual será de interés analizar las características de los momentos de segundo orden para esta clase de procesos. En particular, un momento que ayuda a cuantificar el grado de dependencia (en un sentido lineal) entre dos variables del proceso estocástico, es la covarianza entre estas dos variables. Para procesos estacionarios de segundo orden esta covarianza tiene propiedades que se estudian a continuación.

1.2.2. La sucesión de autocorrelación

Considera $T = \{\dots, -1{,}0{,}1, \dots\}$ y sea $\{X_t\}_t$ un proceso estacionario de 2^{o}orden. Entonces, tomando $\tau = -s$

$$\begin{aligned} COV[X_t, X_s] &= COV[X_{t+\tau}, X_{s+\tau}] \\ &= COV[X_{t-s}, X_0], \end{aligned}$$

si ahora tomas $\tau = -t$

$$\begin{aligned} COV[X_t, X_s] &= COV[X_{t+\tau}, X_{s+\tau}] \\ &= COV[X_0, X_{s-t}]. \end{aligned} \tag{9}$$

Es decir

$$\begin{aligned} COV[X_t, X_s] &= COV[X_0, X_{s-t}] \\ &= COV[X_{t-s}, X_0] \end{aligned} \tag{10}$$

y las últimas dos cantidades dependen de s y t sólo a través de $|s-t|$. En otras palabras, para un proceso estacionario de 2^o orden la covarianza entre X_s y X_t es función de $|t-s|$. Así, la covarianza entre X_1 y X_5 vale lo mismo que la covarianza entre X_3 y X_7.

De la anterior observación, se define la sucesión de autocovarianza del proceso $\{X_t\colon\ t \in T\}$ como

$$S_\tau \equiv COV[X_s, X_{s+\tau}]\ \forall s \in T, \text{y}\ \tau \in T\ \text{ tales que } s+\tau \in T.$$

Nota que para $\tau = 0, S_0 = VAR(X_t) = \sigma^2$. Por otra parte, y en virtud de las ecuaciones (9) y (10), esta sucesión es una función par de τ, es decir,

$$S_\tau = S_{-\tau},$$

véase la figura 3.

Asimismo, se define la sucesión de autocorrelación (ACF) del proceso $\{X_t\}_t$ como

$$\rho_\tau = \frac{S_\tau}{S_0}; \tau = 0, \pm 1, \pm 2, \dots$$

La ACF $\{\rho_\tau, \tau \in T\}$ es tal que $\rho_0 = 1$. Además, para cada $\tau \in T$, como consecuencia de la desigualdad de Cauchy Schwarz (desigualdad (a) en la sección 1.3), se tiene $|\rho_\tau| \leq 1$.
Aunque puede suceder que dos procesos estacionarios de segundo orden tengan la misma sucesión de autocovarianza (autocorrelación), dentro de una clase particular de procesos que se introdujeron a continuación, la sucesión $\{\rho_\tau, \tau \in T\}$ se usará para diagnosticar qué tipo de proceso no resulta inadecuado como modelo para un conjunto de datos.

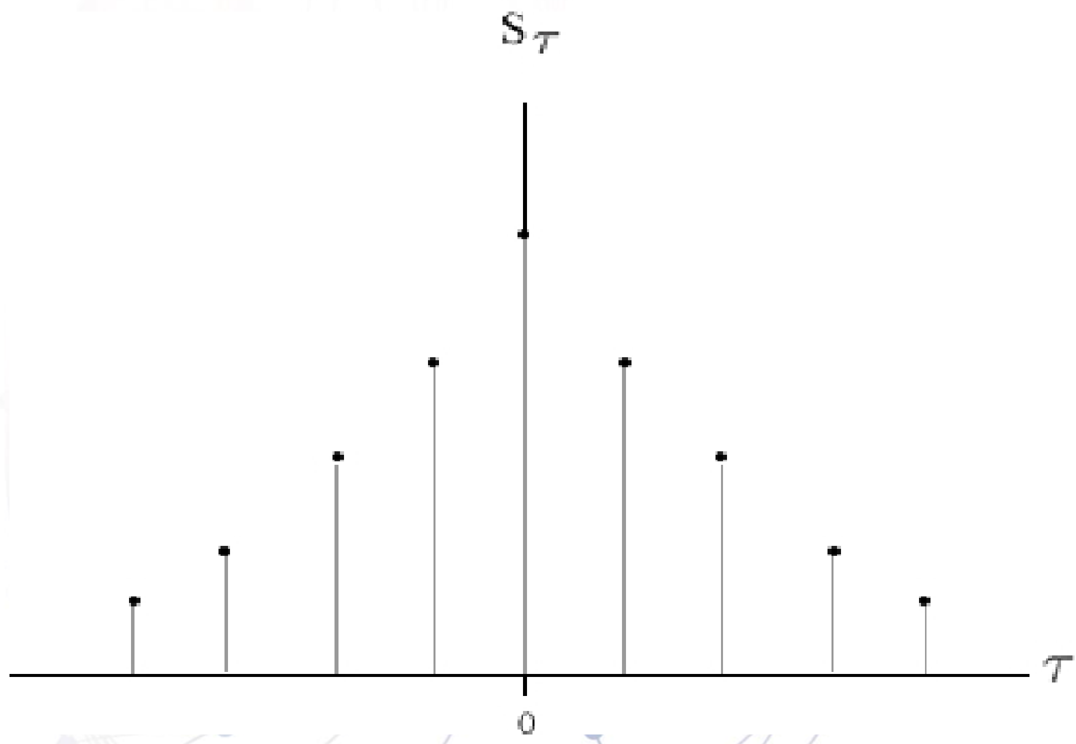

Figura 3: Gráfica de una sucesión de covarianza

1.3. El espacio de variables aleatorias con segundo momento finito

Es menester revisar sin demostrar algunas propiedades matemáticas del entorno en el cual se puede concebir (para fines teóricos) a los procesos de segundo orden, la clase de variables aleatorias dentro de la cual yacen los procesos estacionarios de segundo orden. Las demostraciones de varios de estas propiedades se pueden consultar, por ejemplo, en Laha y Rohatgi (1979), o Royden (1968). Las proyecciones ortogonales juegan un papel importante tanto en inferencia estadística (teorema de Rao-Blackwell) como en regresión. Para el contexto de series de tiempo, observarás que estos conceptos nos ayudan a dar una solución al problema de predicción

1.3.1. Variables aleatorias con segundo momento finito

Definición 4 [Espacio de variables aleatorias con segundo momento finito]

Sea $(\Omega, \ F, IP)$ un espacio de probabilidad, se define

$$L^2(\Omega, \ F, \mathbb{P}) \equiv \{X: \Omega \to \mathbb{R} | X \text{ es variable aleatoria y } \mathbb{E}\,(X^2) < \infty\}.$$

El conjunto $L^2(\Omega, \ F, \mathbb{P})$ tiene estructura de espacio vectorial sobre el campo de los números reales. Por ejemplo, dadas X, Y en $L^2(\Omega, \ F, \mathbb{P})$ se tiene que casi seguramente con respecto a $\mathbb{P}$

$$0 \leq (X+Y)^2 \leq 2X^2 + 2Y^2, \tag{11}$$

ya que para $a, b \in \mathbb{R}$ se tiene $(a+b)^2 \leq 2a^2 + 2b^2$.

Tomando valor esperado en (11), se obtiene

$$\mathbb{E}\,((X+Y)^2) \leq 2\,\mathbb{E}\,(X^2) + 2\mathbb{E}\,(Y^2) < \infty,$$

de donde se tienes que si $\alpha \in \mathbb{R}$

$$\mathbb{E}\,((\alpha X + Y)^2) \leq 2\,\mathbb{E}\,((\alpha X)^2) + 2\,\mathbb{E}\,(Y^2)$$
$$= 2\alpha^2 \mathbb{E}(X^2) + 2\mathbb{E}\,(Y^2) < \infty.$$

Se puede, además, definir un producto interno para $L^2(\Omega, \ F, \mathbb{P}\ )$. Éste está dado por

$$\langle X, Y \rangle \equiv \ \mathbb{E}\,(XY)$$

y tiene las siguientes propiedades para X, Y, Z en $L^2(\Omega, \ F, \mathbb{P}\,)$

- $\langle X, Y \rangle = \langle Y, X \rangle$,
- $\langle \alpha X + \beta Y, Z \rangle = \alpha \langle X, Z \rangle + \beta \langle Y, Z \rangle$,
- $\langle X, \gamma Y \rangle = \gamma \langle X, Y \rangle$,
- $\langle X, X \rangle = \ \mathbb{E}\,(X^2) \geq 0$.

Con este producto interno se define la siguiente norma para $L^2(\Omega, \ F, \mathbb{P}\,)$

$$||X||_2 = \sqrt{\langle X, X \rangle} = \sqrt{\mathbb{E}\,(X^2)}.$$

Esta norma caracteriza la llamada convergencia en media cuadrática, para variables aleatorias con segundo momento finito.

Nota que si $IE\ (X) = \mathbb{E}\ (Y) = 0$, entonces $\langle X, Y \rangle = COV[X, Y]$, correspondientemente $||X||_2 = \sqrt{VAR\ (X)}$.

1.3.2. Desigualdades importantes

Se tienen, además, dos desigualdades cruciales, las cuales no se demostrarán.

a) • Desigualdad de Cauchy-Schwarz: Dadas X, Y en $L^2(\Omega, \ F, IP)$

$$|\langle X, Y\rangle| \leq \big||X|\big|_2 \big||Y|\big|_2, \tag{12}$$

b) Desigualdad de Minkowski: Dadas X, Y en $L^2(\Omega, \ F, IP)$

$$||X+Y||_2 \leq ||X||_2 + ||Y||_2.$$

A continuación se presenta una operación que se usará frecuentemente en la descripción y tratamiento de los procesos estocásticos y métodos a estudiar en el curso.

1.3.3. Operadores de retraso

Definición 5 [Operadores de retraso y diferencia]

Sea $\{X_t\}_t$ un proceso de segundo orden. El operador de retraso $B: L^2 \to L^2$ está dado por la relación

$$BX_t = X_{t-1}.$$

El operador de diferencia $\nabla: L^2 \to L^2$ queda definido por

$$\nabla X_t = X_t - X_{t-1}.$$

Se tiene que $\nabla X_t = (1-B)X_t$ y estas operaciones son lineales, dados dos procesos de segundo orden $\{X_t\}_t$ y $\{Y_t\}_t$

$$B(\alpha X_t + Y_t) = \alpha X_{t-1} + Y_{t-1} = \alpha B X_t + B Y_t.$$

Se puede definir la aplicación iterada (composición) de estos operadores, por ejemplo

$$\begin{aligned} B^k X_t &= B^{k-1} B X_t = B^{k-1} X_{t-1} = B^{k-2} B X_{t-1} \\ &= B^{k-2} X_{t-2} = \cdots \\ &= X_{t-k}, \end{aligned}$$

con la convención de que $B^0 = I$, donde $IX_t = X_t$ es el operador identidad.

Lo anterior permite trabajar con operadores definidos como polinomios, por ejemplo, $P(B) = 1 + \frac{5}{2}B - \frac{3}{2}B^2$, el cual actua sobre X_t como

$$P(B)X_t = (1 + \frac{5}{2}B - \frac{3}{2}B^2)X_t = X_t + \frac{5}{2}X_{t-1} - \frac{3}{2}X_{t-2}.$$

Para z un número complejo, considera el polinomio obtenido al reemplazar a B por z, $P(z) = 1 + \frac{5}{2}z - \frac{3}{2}z^2$. El teorema fundamental del álgebra (véase, por ejemplo, Kurosh, 1972, o Lang, 1974), dice si $z_{0,1}$ y $z_{0,2}$ son las raíces de $P(z)$, entonces éste se puede escribir como

$$P(z) = (1 - \frac{1}{z_{0,1}}z)(1 - \frac{1}{z_{0,2}}z)$$

y en este caso $z_{0,1} = 2$ y $z_{0,2} = -\frac{1}{3}$, es decir

$$P(z) = (1 - \frac{1}{2}z)(1 + 3z).$$

Ahora, aplicando el operador $(1 - \frac{1}{2}B)(1 + 3B)$ a X_t

$$\begin{aligned} \left(1 - \frac{1}{2}B\right)(1 + 3B)X_t &= \left(1 - \frac{1}{2}B\right)(X_t + 3X_{t-1}) \\ &= X_t + 3X_{t-1} - \frac{1}{2}X_{t-1} - \frac{3}{2}X_{t-2} \\ &= X_t + \frac{5}{2}X_{t-1} - \frac{3}{2}X_{t-2} = \left(1 + \frac{5}{2}B - \frac{3}{2}B^2\right)X_t \end{aligned} \tag{13}$$

$$= P(B)X_t.,$$

De (13) ves que $P(B)$ y $(1 - \frac{1}{2}B)(1 + 3B)$ son el mismo operador.

Esta correspondencia entre operaciones algebráicas con polinomios de variable compleja z y los correspondientes polinomios donde la variable es B será usada más adelante.
En particular, para α un número real tal que $|\alpha| < 1$, será de interés el operador inverso del operador $1 - \alpha B$, dado por

$$(1 - \alpha B)^{-1} = \sum_{k=0}^{\infty} \alpha^k B^k., \qquad (14)$$

En Radjavi y Rosenthal (2003), se establece el lado derecho de (14) es otro operador de L^2 en L^2 que está bien definido. Nota que formalmente (14) es un inverso de $1 - \alpha B$, ya que

$$\begin{aligned}(1 - \alpha B)\sum_{k=0}^{\infty} \alpha^k B^k &= \sum_{k=0}^{\infty} \alpha^k B^k - \sum_{k=1}^{\infty} \alpha^k B^k \\ &= I + \sum_{k=1}^{\infty} \alpha^k B^k - \sum_{k=1}^{\infty} \alpha^k B^k \\ &= I.\end{aligned}$$

Un problema que emerge en la práctica es el problema de la proyección ortogonal. Por ejemplo, en el espacio vectorial euclidiano $H = \mathbb{R}^3$, con el producto interno $\langle\ x\ ,\ \ y\ \rangle = x_1y_1 + x_2y_2 + x_3y_3$, para x , y en H. Considera vectores y , x_1 y x_2 en H y el subespacio V de H dado por

$$V\ = \{x\ \in H|\ \ x\ = \alpha x_1 + \beta x_2;\ \alpha, \beta \in \mathbb{R}\}.$$

Supón que tienes el problema de encontrar el vector y en V que minimiza la distancia entre y y V , el vector y se conoce como la proyección ortogonal de y sobre V y, además de ser un elemento de V, cumple la condición de ortogonalidad

$$\forall \quad x \in V, \langle \quad y - \quad y, \quad x \rangle = 0 \tag{15}$$

Ahora supón que $y \ = (\frac{1}{4}, \frac{1}{4}, 1)', x_1 = (1{,}0, \frac{1}{4})'$ y $x_2 = (0{,}1, \frac{1}{4})'$. Como $y \ \in \quad V$, entonces existen a y b en $\mathbb{R}$ tales que $y \ = ax_1 + bx_2$, además, la condición (15) implica que $\langle \quad y - ax_1 - bx_2, \ \ x_2 \rangle = 0$ y que $\langle \quad y - ax_1 - bx_2, \ \ x_2 \rangle = 0$, estas son dos ecuaciones lineales con incógnitas a y b que se pueden escribir

$$\begin{aligned} a\langle \ x_1, \ x_1 \rangle + b\langle \ x_2, \ x_1 \rangle &= \langle \ y, \ x_1 \rangle \ y \\ a\langle \ x_1, \ x_2 \rangle + b\langle \ x_2, \ x_2 \rangle &= \langle \ y, \ x_2 \rangle \end{aligned} \tag{16}$$

Para los vectores especificados arriba se tiene $\langle \ y, \ x_1 \rangle = 1/2, \langle \ y, \ x_2 \rangle = 1/2, \langle \ x_1, \ x_1 \rangle = 17/16, \langle \ xx_2, \ xx_2 \rangle = 17/16$ y $\langle \ x_1, \ x_2 \rangle = 1/16$, el sistema de ecuaciones lineales (16) tiene solución $a = b = 4/9$, y entonces $y\,' = (\frac{4}{9}, \frac{4}{9}, \frac{2}{9})'$. La notación $y' \ = Proy_V(\ y \ )$ ayudará a ser más específicos posteriormente, ya que indica el espacio sobre el cual se proyecta. Ahora piensa en el mismo problema de encontrar una proyección ortogonal, pero considerando $H = L^2(\Omega, \ \ F, \mathbb{P})$ y variables aleatorias Y, X_1 y X_2 en H. El teorema de la proyección ortogonal (véase Shumway y Stoffer, 2006, apéndice B, o también Brockwell y Davis, 2009, capítulo 2) dice que si $V = \{X \in L^2(\Omega, \ \ F, IP)|X = \alpha X_1 + \beta X_2; \ \alpha, \beta \in \mathbb{R}\}$, entonces existe $\hat{Y} = Proy_V(Y)$ en V que minimiza la distancia entre Y y V

$$||Y - \hat{Y}||_2 = \inf_{X \in V}\{||Y - X||_2\}.$$

La variable aleatoria $\hat{Y}$ pertenece a V, por lo cual $\hat{Y} = aX_1 + bX_2$, además se satisface la condición de ortogonalidad

$$\forall X \in V, \qquad \langle Y - aX_1 - bX_2, X \rangle = 0. \tag{17}$$

Supón ahora que las variables aleatorias Y, X_1 y X_2 tienen media 0, en tal caso $\langle Y, X_i \rangle = COV[Y, X_i], i = 1{,}2$ y $\langle Y, X \rangle = COV[Y, X]$ y debido a la condición de ortogonalidad (17) se tiene el sistema de ecuaciones lineales

$$\begin{aligned} a \cdot COV[X_1, X_1] + b \cdot COV[X_2, X_1] &= COV[Y, X_1] \\ a \cdot COV[X_1, X_2] + b \cdot COV[X_2, X_2] &= COV[Y, X_2]. \end{aligned} \tag{18}$$

Si conoces las covarianzas $COV[Y, X_i], i = 1{,}2, COV[X_1, X_2], COV[X_1, X_1]$ y $COV[X_2, X_2]$, entonces puedes dar solución al sistema para conocer a y b. Por último, verás cómo estas ideas ayudan a resolver el problema de predicción.

Sea $\{X_t\}_t$ un proceso estacionario de segundo orden tal que $\mathbb{E}\,(X_t) = 0$. Supón que dadas las variables $X_1, X_2, \dots, X_n$, quieres predecir el valor de una observación futura X_{n+1} en función de $X_1, X_2, \dots, X_n$. Para lo anterior se denota por $\hat{X}_{n+1}$ a esta función de la historia del proceso. En un sentido práctico te concentras en el caso de que $\hat{X}_{n+1}$ es una función lineal del pasado, es decir

$$\hat{X}_{n+1} = \sum_{j=1}^{n} \phi_{n,j} X_{n+1-j} = \phi_{n,1} X_n + \phi_{n,2} X_{n-1} + \phi_{n,n} X_1 \tag{19}$$

A continuación vas a escribir esta tarea en el contexto del problema de la proyección ortogonal, sean $\mathcal{H} = L^2(\Omega, \mathcal{F}, \mathbb{P}), Y = X_{n+1}$ y quieres encontrar la variable aleatoria $\hat{X}_{n+1}$ en el subespacio $V = \{X \in H \mid X = \sum_{i=1}^{n} \alpha_i X_i;\ \alpha_i \in \mathbb{R}, i = 1, \dots, n\}$ que minimiza la distancia entre X_{n+1} y V. Usando el teorema de la proyección ortogonal puedes encontrar los coeficientes $\{\phi_{n,j}, j = 1, \dots, n\}$ en (19) que definen a $\hat{X}_{n+1} = Proy_V(X_{n+1})$, al resolver el sistema lineal (ecuaciones de predicción)

$$\phi_{n,1} \cdot COV[X_n, X_1] + \phi_{n,2} \cdot COV[X_{n-1}, X_1] + \cdots + \phi_{n,n} \cdot COV[X_1, X_1] = COV[X_{n+1}, X_1]$$

$$\phi_{n,1} \cdot COV[X_n, X_2] + \phi_{n,2} \cdot COV[X_{n-1}, X_2] + \cdots + \phi_{n,n} \cdot COV[X_1, X_2] = COV[X_{n+1}, X_2]$$
$$\vdots \quad \vdots \qquad \vdots$$
$$\phi_{n,1} \cdot COV[X_n, X_n] + \phi_{n,2} \cdot COV[X_{n-1}, X_n] + \cdots + \phi_{n,n} \cdot COV[X_1, X_n] = COV[X_{n+1}, X_n],$$

o equivalentemente (dado que $S_{i-j} = COV[X_i, X_j]$)

$$\begin{aligned} \phi_{n,1}S_0 + \phi_{n,2}S_1 + \cdots + \phi_{n,n}S_{n-1} &= S_1 \\ \phi_{n,1}S_1 + \phi_{n,2}S_0 + \cdots + \phi_{n,n}S_{n-2} &= S_2 \\ \vdots \quad \vdots \vdots & \\ \phi_{n,1}S_{n-1} + \phi_{n,2}S_{n-2} + \cdots + \phi_{n,n}S_0 &= S_n \end{aligned} \tag{20}$$

La proposición 5.1.1 en el libro de Brockwell y Davis (2009) establece si $S_0 = VAR(X_t) > 0$ y $S_h \to 0$, cuando $h \to \infty$, entonces las ecuaciones de predicción (??) tienen solución. En la siguiente unidad verás que estas condiciones sobre $\{S_\tau\}_\tau$ son válidas para un proceso ARMA(p, q) causal.

1.4. Procesos ARMA (p,q)

Es momento de observar algunos procesos estacionarios de 2^o orden que son utilizados para modelar datos. Estos procesos forman parte de la familia de **procesos lineales,** y en su definición se recuperan ideas de teoría de regresión lineal, en el sentido de que para cada tiempo t, el valor observado X_t puede pensarse como la respuesta en un modelo de regresión. Para este modelo de regresión las variables de control son: los valores observados a tiempos pasados (antes de t) y valores de otro proceso a tiempos pasados.

Primeramente se estudian las propiedades de este otro proceso al cual se llamará ruido blanco. En lo que sigue, se asumirá que $T = \{\ldots, -1, 0, 1, \ldots\}$.

1.4.1. Proceso de ruido blanco

Sea $\{X_t\}_t$ una sucesión de variables aleatorias no-correlacionadas, tal que $IE\ [X_t] = \mu$ y $VAR\ [X_t] = \sigma^2$, $\forall t$. Entonces $COV[X_t, X_{t+\tau}] = 0 \forall t \in T$ y $\tau \neq 0$, luego

$$S_\tau = \begin{cases} \sigma^2, & si\ \tau = 0, \\ 0, & si\ \tau \neq 0, \end{cases} \quad \rho_\tau = \begin{cases} 1, & si\ \tau = 0, \\ 0, & si\ \tau \neq 0. \end{cases} \tag{21}$$

El ruido blanco es un proceso estacionario de segundo orden, ya que la media, la varianza y la covarianza entre cualesquiera dos variables del proceso son constantes independientes del tiempo, y por ende, se satisface la condición (4) revisada en el subtema 1.1.3.

Este proceso parece muy simple, pero servirá como el material principal para construir casos más sofisticados.

1.4.2. Proceso de promedios móviles

Para q en $\{1,2,\ldots\}$, se dirá que el proceso $\{X_t\}_t$ se llama de promedios móviles con orden q (MA(q)), si se puede escribir como

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$

donde μ y θ_j son constantes reales, $j = 1,2,\ldots,q$. El proceso $\{\varepsilon_t\}_t$ es ruido blanco con $\mathbb{E}\ (\varepsilon_t) = 0$, $\forall t$ y $VAR\ (\varepsilon_t) = \sigma_\varepsilon^2$.

Debido a las propiedades del valor esperado, se tiene $IE\ (X_t) = \mu$. Vas a calcular la sucesión de autocovarianza de $\{X_t\}_t$, para lo cual, primeramente asume $\tau \geq 0$ y observa que el proceso $Y_t = X_t - \mu$ es tal que

$$\begin{aligned} COV[X_t, X_{t+\tau}] &= \mathbb{E}\,[(X_t - \mu)(X_{t+\tau} - \mu)] \\ &= \mathbb{E}\,[Y_t \cdot Y_{t+\tau}]. \end{aligned}$$

Sea $\theta_0 \equiv 1$, se puede escribir

$$\begin{aligned} Y_t \cdot Y_{t+\tau} &= \left(\sum_{j=0}^{q} \theta_j \varepsilon_{t-j}\right)\left(\sum_{l=0}^{q} \theta_l \varepsilon_{t+\tau-l}\right) \\ &= \sum_{j=0}^{q}\sum_{l=0}^{q} \theta_j \theta_l \varepsilon_{t-j}\varepsilon_{t+\tau-l}, \end{aligned}$$

y luego, por linealidad de la esperanza

$$OV[X_t, X_{t+\tau}] = \mathbb{E}\,[Y_t \cdot Y_{t+\tau}] = \sum_{j=0}^{q}\sum_{l=0}^{q} \theta_j \theta_l \mathbb{E}\,[\varepsilon_{t-j}\varepsilon_{t+\tau-l}].$$

Usando que $\{\varepsilon_t\}_t$ es ruido blanco, de la ecuación (21) se concluye que el valor esperado en cada término de la suma se anula siempre que no se de la condición $t - j = t + \tau - l$, es decir, cuando $l = \tau + j$ y en cuyo caso el valor esperado vale σ_ε^2. Pero, como $0 \leq l \leq q$, se debe tener que $\tau + j \leq q$, o bien, $j \leq q - \tau$. Entonces, $0 \leq j \leq q - \tau$ y, por lo tanto, $\tau \leq q$. De estas observaciones se obtiene que

$$COV[X_t, X_{t+\tau}] = \begin{cases} \left(\sum_{j=0}^{q-\tau} \theta_j \theta_{\tau+j}\right)\sigma_\varepsilon^2, & si \quad 0 \leq \tau \leq q, \\ 0, & si\ \tau > q. \end{cases}$$

Dentro de las actividades a desarrollar está probar la siguiente afirmación: Si $\tau < 0$, entonces

$$COV[X_t, X_{t+\tau}] = \begin{cases} \left(\sum_{j=0}^{q+\tau} \theta_j \theta_{j-\tau}\right)\sigma_\varepsilon^2, & si \quad -q \leq \tau < 0, \\ \\ 0, & si\ \tau < -q. \end{cases}$$

Al comparar los 2 casos, el caso $\tau \geq 0$ y el caso $\tau < 0$, se concluye que

$$COV[X_t, X_{t+\tau}] = \begin{cases} \left(\sum_{j=0}^{q-|\tau|} \theta_j \theta_{j+|\tau|}\right)\sigma_\varepsilon^2, & si\ \ 0 \leq |\tau| \leq q, \\ \\ 0, & si \quad |\tau| > q. \end{cases}$$

Nota que $\mathbb{E}\ (X_t)$ no depende de t y $COV[X_t, X_{t+\tau}]$; sólo depende de τ y no de t. De lo anterior se puede establecer que si $\{X_t\}_t$ es un proceso MA(q), entonces $\mathbb{E}(X_t) = \mu,\ \ \forall t \in T$, y, además, $COV[X_t, X_s] = S(|\tau|), \tau = t - s$, es decir, se satisface la condición (4). Por lo tanto, no hay restricciones sobre los valores $\theta_j, j = 1,2, \dots, q$ y μ para que $\{X_t\}_t$ sea estacionario de segundo orden y se puede escribir $S_\tau = COV[X_t, X_{t+\tau}]$ y observando que $S_0 = \sigma_\varepsilon^2 \sum_{j=0}^{q} \theta_j^2 > 0$, se deduce que

$$\rho_\tau = \begin{cases} \dfrac{\sum_{j=0}^{q-|\tau|} \theta_j \theta_{j+|\tau|}}{\sum_{j=0}^{q} \theta_j^2}, & si\ \ 0 \leq |\tau| \leq q, \\ 0, & si \quad |\tau| > q. \end{cases} \tag{22}$$

Observa que para el caso de los procesos de promedios móviles, de acuerdo a sucesión de autocorrelación o función de autocorrelación (ACF) (22), si $|\tau| > q$, entonces $\rho_\tau = 0$. Es decir, para un proceso MA(q), su ACF correspondiente $\{\rho_\tau |\ \tau \in \{\dots, -2, -1, 0, 1, 2, \dots\}\}$ es diferente de 0 sólo para valores de τ en el rango $-q, -q+1, \dots, 0, \dots, q-1, q$, y se cancela para los valores de τ que no sean estos números enteros. Como podrás observar en el siguiente ejemplo, este comportamiento de la ACF no sucede en los procesos autorregresivos, y por esta razón será importante recordar que para los procesos MA(q), la ACF se comporta como en la ecuación (22), la cual da información del valor de q. Dicho de otra forma, si se tiene un proceso del cual sólo se conoce ρ_τ, y ésta se comporta como la ecuación (22), entonces, dentro de la clase de procesos de promedios móviles el proceso MA(q) es un candidato para usarse como modelo.

1.4.3. Proceso autorregresivo

Para p en $\{1,2,\dots\}$, se dirá que el proceso $\{X_t\}_t$ se llama **autorregresivo** de orden p (AR(p)) si se puede escribir como

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t, \tag{23}$$

donde μ y ϕ_j, $j = 1,2,\dots,p$, son constantes reales y el proceso $\{\varepsilon_t\}_t$ es ruido blanco con $\mathbb{E}\ (\varepsilon_t) = 0, \forall t$ y $VAR\ (\varepsilon_t) = \sigma_\varepsilon^2$.

Con la notación del operador de retraso B definido en la sección anterior se puede escribir (23) como:

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) X_t = \mu + \varepsilon_t \tag{24}$$

A diferencia de los procesos MA(q) que siempre son estacionarios, los procesos AR(p) no resultan estacionarios para cualesquiera valores de ϕ_j, $j = 1,2,\dots,p$. Se comenzará por estudiar el caso $p = 1$.

Considera el siguiente proceso

$$(1 - \phi_1 B) X_t = \mu + \varepsilon_t,$$

con $\{\varepsilon_t\}_t$ ruido blanco con media 0 y varianza σ_ε^2.

De la discusión respecto al operador de retraso B en el subtema 1.3.3, siempre que $|\phi_1| < 1$ se puede escribir

$$\begin{aligned} X_t &= (1-\phi_1 B)^{-1}(\mu+\varepsilon_t) = \sum_{k=0}^{\infty} \phi_1^k B^k(\mu+\varepsilon_t) \\ &= \sum_{k=0}^{\infty} \phi_1^k B^k \mu + \sum_{k=0}^{\infty} \phi_1^k B^k \varepsilon_t \\ &= \sum_{k=0}^{\infty} \phi_1^k \mu + \sum_{k=0}^{\infty} \phi_1^k \varepsilon_{t-k} \\ &= (\frac{\mu}{1-\phi_1}) + \sum_{k=0}^{\infty} \phi_1^k \varepsilon_{t-k} \end{aligned} \tag{25}$$

Esta es la conocida representación MA(∞) para un proceso autorregresivo; escribir así al proceso AR(1) te permitirá calcular sus momentos en forma análoga a cómo se hizo para el proceso MA(q) en el inciso (b). Con este fin, sea $\mu_0 = \frac{\mu}{1-\phi_1}$ y al tomar el valor esperado en ambos lados de (25)

$$\begin{aligned} \mathbb{E}(X_t) &= \mathbb{E}\left(\mu_0 + \sum_{k=0}^{\infty} \phi_1^k \varepsilon_{t-k}\right) \\ &= \mu_0 + \mathbb{E}\left(\sum_{k=0}^{\infty} \phi_1^k \varepsilon_{t-k}\right) \\ &= \mu_0 + \sum_{k=0}^{\infty} \phi_1^k \mathbb{E}\,(\varepsilon_{t-k}) \\ &= \mu_0. \end{aligned} \tag{26}$$

En la penúltima igualdad, el intercambio entre valor esperado y suma se puede justificar con los teoremas de convergencia monótona y convergencia dominada (Royden, 1968).

Como en el caso MA(q), se pueden volver a tomar $Y_t = X_t - \mu_0$ y se tiene

$$COV[X_t, X_{t+\tau}] = \mathbb{E}[(X_t-\mu_0)(X_{t+\tau}-\mu_0)]$$

$$
\begin{aligned}
&= \mathbb{E}\,[Y_t \cdot Y_{t+\tau}] \\
&= \mathbb{E}\left[\left(\sum_{k=0}^{\infty} \phi_1^k \varepsilon_{t-k}\right)\left(\sum_{l=0}^{\infty} \phi_1^l \varepsilon_{t+\tau-l}\right)\right] \\
&= \mathbb{E}\left[\sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \phi_1^k \phi_1^l \varepsilon_{t-k}\varepsilon_{t+\tau-l}\right] \\
&= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \phi_1^k \phi_1^l IE\,[\varepsilon_{t-k}\varepsilon_{t+\tau-l}].
\end{aligned}
$$

De nueva cuenta, el intercambio entre valor esperado y suma se puede justificar con los teoremas de convergencia dominada y monótona. Luego

$$
COV[X_t, X_{t+\tau}] = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \phi_1^k \phi_1^l IE\,[\varepsilon_{t-k}\varepsilon_{t+\tau-l}],
$$

Pero, por ser $\{\varepsilon_t\}_t$ un proceso de ruido blanco,

$$
\mathbb{E}\,[\varepsilon_{t-k}\varepsilon_{t+\tau-l}] = \begin{cases} 0, & sil \neq k+\tau, \\ \sigma_\varepsilon^2, & sil = k+\tau. \end{cases}
$$

Entonces, si $\tau \geq 0$,

$$
\begin{aligned}
COV[X_t, X_{t+\tau}] &= \sum_{k=0}^{\infty} \phi_1^k \phi_1^{k+\tau} \sigma_\varepsilon^2 \\
&= \sum_{k=0}^{\infty} \phi_1^{2k+\tau} \sigma_\varepsilon^2 \\
&= \phi_1^\tau \sigma_\varepsilon^2 \sum_{k=0}^{\infty} \phi_1^{2k} \\
&= \sigma_\varepsilon^2 \left(\frac{\phi_1^\tau}{1-\phi_1^2}\right),
\end{aligned}
\tag{27}
$$

toda vez que $|\phi_1| < 1$. Por otra parte, si $\tau < 0$, entonces $l = k + \tau$ es no negativo cuando $k \geq -\tau$, en tal caso

$$\begin{aligned} COV[X_t, X_{t+\tau}] &= \sum_{k=-\tau}^{\infty} \phi_1^k \phi_1^{k+\tau} \sigma_\varepsilon^2 \\ &= \sum_{k=0}^{\infty} \phi_1^{2k-\tau} \sigma_\varepsilon^2 \\ &= \phi_1^{-\tau} \sigma_\varepsilon^2 \sum_{k=0}^{\infty} \phi_1^{2k} \\ &= \sigma_\varepsilon^2 \left(\frac{\phi_1^{-\tau}}{1-\phi_1^2} \right). \end{aligned} \tag{28}$$

Combinando (27) y (28) se concluye que en general (para cualquier τ en T tal que $t + \tau \in T$)

$$COV[X_t, X_{t+\tau}] = \sigma_\varepsilon^2 \left(\frac{\phi_1^{|\tau|}}{1-\phi_1^2} \right) \tag{29}$$

De (26) y (29) se observa que $\mathbb{E}\ (X_t)$ y $COV[X_t, X_{t+\tau}]$ no dependen de t y que $COV[X_t, X_{t+\tau}]$ sólo depende de τ. Entonces el proceso AR(1) es estacionario de 2^o orden bajo el supuesto de que $|\phi_1| < 1, \mathbb{E}(X_t) = \mu_0, S_\tau = \sigma_\varepsilon^2 (\frac{\phi_1^{|\tau|}}{1-\phi_1^2})$ y $VAR(X_t) = \frac{\sigma_\varepsilon^2}{1-\phi_1^2}$. Claramente, para un proceso AR(1), el comportamiento de la sucesión de autocorrelación ρ_τ es radicalmente diferente al de la correspondiente sucesión de autocorrelación de un proceso MA(1), en el sentido de que para el proceso AR(1) esta sucesión no se cancela a partir de un valor entero como sucede en la ecuacion (22). Por esta razón, cuando se está tratando de identificar si un modelo autorregresivo se podría usar para los datos, la ACF no es una herramienta útil ya que no da una idea de lo que vale p. Como se verá mas adelante en el tema 1.5, para identificar modelos autorregresivos para los datos se requiere usar otra sucesión de autocorrelación conocida como autocorrelación parcial (PACF).

Para z en el campo de los números complejos $\mathbb{C}$, sea $\phi(z) \equiv 1 - \phi_1 z$ el polinomio obtenido al reemplazar B por z en el operador $1 - \phi_1 B$. Una observación que será importante es que restringir a $\phi(z)$ a que su raíz esté afuera del conjunto $\{z \in \mathbb{C}: |z| \leq 1\}$, conduce a la condición de estacionariedad $|\phi_1| < 1$.

Trata ahora el caso $p = 2$.

Considérese ahora X_t un proceso de segundo orden tal que para cada t

$$(1 - \phi_1 B - \phi_2 B^2) X_t = \mu + \varepsilon_t, \tag{30}$$

donde ϕ_1 y ϕ_2 son números reales y $\{\varepsilon_t\}_t$ es un proceso de ruido blanco con $IE\ (\varepsilon_t) = 0$ y $VAR\ (\varepsilon_t) = \sigma_\varepsilon^2$. Sea $\phi(z) = 1 - \phi_1 z - \phi_2 z^2$, con $z \in \mathbb{C}$. Vas a usar el mismo argumento que lleva a la ecuación (13). Con este fin, escribe

$$\begin{aligned} 1 - \phi_1 z - \phi_2 z^2 &= (az - b)(cz - d) \\ &= acz^2 - (ad + bc)z + bd, \end{aligned}$$

en donde $ac = -\phi_2,\ ad + bc = \phi_1 y bd = 1$. Equivalentemente

$$\begin{aligned} \phi(z) = 1 - \phi_1 z - \phi_2 z^2 &= (b - az)(d - cz) \\ &= b(1 - \frac{a}{b}z)d(1 - \frac{c}{d}z) \\ &= (1 - \frac{a}{b}z)(1 - \frac{c}{d}z), \end{aligned}$$

con las raíces de $\phi(z)$ dadas por $z_1 = \frac{b}{a} y z_2 = \frac{d}{c}$.

Sean $\mu_1 = \frac{1}{z_1}, \mu_2 = \frac{1}{z_2}$. Usando la correspondencia entre $\phi(z)$ y $\phi(B)$ ilustrada en la ecuación (31), escribe el polinomio de retraso $\phi(B)$ como

$$1 - \phi_1 B - \phi_2 B^2 = (1 - \mu_1 B)(1 - \mu_2 B). \tag{31}$$

Así, usando (31) escribe (30) como

$$(1-\mu_1 B)(1-\mu_2 B)X_t = \mu + \varepsilon_t$$

y puedes pensar en los operadores $(1-\mu_1 B)^{-1}$ y $(1-\mu_2 B)^{-1}$ de acuerdo a (14), siempre que

$$|\mu_1| < 1 \;\; y \;\; |\mu_2| < 1. \tag{32}$$

Asumiendo estas condiciones

$$\begin{aligned} X_t &= (1-\mu_1 B)^{-1}(1-\mu_2 B)^{-1}(\mu+\varepsilon_t) \\ &= \frac{1}{\mu_1-\mu_2}[\mu_1(1-\mu_1 B)^{-1} - \mu_2(1-\mu_2 B)^{-1}](\mu+\varepsilon_t). \end{aligned}$$

Para justificar la última igualdad, sean $a_1 \equiv (1-\mu_1 B), a_2 \equiv (1-\mu_2 B), \alpha \equiv \frac{\mu_1}{\mu_1-\mu_2}$ y $\beta \equiv \frac{-\mu_2}{\mu_1-\mu_2}$.

Tienes que

$$a_1^{-1}\alpha + a_2^{-1}\beta = a_1^{-1}a_2^{-1}(a_2\alpha + a_1\beta) = a_1^{-1}a_2^{-1},$$

ya que $a_2\alpha + a_1\beta = 1$.

Consecuentemente, usando (14)

$$\begin{aligned} X_t &= \frac{1}{\mu_1-\mu_2}\left[\mu_1 \sum_{k=0}^{\infty}(\mu_1 B)^k - \mu_2 \sum_{k=0}^{\infty}(\mu_2 B)^k\right](\mu+\varepsilon_t) \\ &= \frac{1}{\mu_1-\mu_2}\left[\sum_{k=0}^{\infty}(\mu_1^{k+1} - \mu_2^{k+1})B^k\right](\mu+\varepsilon_t) \\ &= \sum_{k=0}^{\infty}\left\{\frac{\mu_1^{k+1}-\mu_2^{k+1}}{\mu_1-\mu_2}\right\}(\mu+\varepsilon_{t-k}) \qquad (33) \\ &= \sum_{k=0}^{\infty}\left\{\frac{\mu_1^{k+1}-\mu_2^{k+1}}{\mu_1-\mu_2}\right\}\mu + \sum_{k=0}^{\infty}\left\{\frac{\mu_1^{k+1}-\mu_2^{k+1}}{\mu_1-\mu_2}\right\}\varepsilon_{t-k} \\ &= \frac{\mu}{(1-\mu_1)(1-\mu_2)} + \sum_{k=0}^{\infty}\psi_k\varepsilon_{t-k}, \end{aligned}$$

donde se denota $\psi_k = \frac{\mu_1^{k+1} - \mu_2^{k+1}}{\mu_1 - \mu_2}$. Al igual que en el caso AR(1) (donde se usó la condición $|\phi_1| < 1$), en este caso bajo las restricciones (32) has escrito al proceso AR(2) en una representación MA(∞). Lo anterior sera útil para calcular media y sucesión de autocovarianza del proceso; para hacer esto, sea $\mu_0 = \frac{\mu}{(1-\mu_1)(1-\mu_2)}$ y calculando valor esperado en ambos lados de (33)

$$\mathbb{E}\,(X_t) = \mathbb{E}\left(\mu_0 + \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}\right) = \mu_0 + \sum_{k=0}^{\infty} \psi_k \mathbb{E}\,(\varepsilon_{t-k}) = \mu_0. \tag{34}$$

Toma $Y_t = X_t - \mu_0$ y vas a calcular la covarianza

$$\begin{aligned} COV[X_t, X_{t+\tau}] &= IE\,[(X_t - \mu_0)(X_{t+\tau} - \mu_0)] \\ &= IE\,[Y_t \cdot Y_{t+\tau}] \\ &= IE\left[\left(\sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}\right)\left(\sum_{l=0}^{\infty} \psi_l \varepsilon_{t+\tau-l}\right)\right] \\ &= IE\left[\sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \psi_k \psi_l \varepsilon_{t-k} \varepsilon_{t+\tau-l}\right] \\ &= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \psi_k \psi_l IE\,[\varepsilon_{t-k} \varepsilon_{t+\tau-l}]. \end{aligned}$$

El valor esperado en cada sumando se anula siempre que no suceda que $l = k + \tau$. Si $\tau \geq 0$; entonces $COV[X_t, X_{t+\tau}] = (\sum_{k=0}^{\infty} \psi_k \psi_{k+\tau})\sigma_\varepsilon^2$. Por otra parte, si $\tau < 0$, entonces

$$COV[X_t, X_{t+\tau}] = \left(\sum_{k=-\tau}^{\infty} \psi_k \psi_{k+\tau}\right)\sigma_\varepsilon^2 = \left(\sum_{k=0}^{\infty} \psi_k \psi_{k-\tau}\right)\sigma_\varepsilon^2.$$

Ambos casos se pueden escribir en la expresión

$$COV[X_t, X_{t+\tau}] = \left(\sum_{k=0}^{\infty} \psi_k \psi_{k+|\tau|}\right)\sigma_\varepsilon^2 \tag{35}$$

Nota que $\sum_{k=0}^{\infty} \psi_k \psi_{k+|\tau|}$ es finita. Para ver esto, bajo las condiciones (32) se tiene que $\sum_{k=0}^{\infty} |\psi_k| < \infty$, luego, dada $\varepsilon \in (0,1)$, existe k_0 tal que si $k \geq k_0$, entonces $|\psi_k| < \varepsilon < 1$. De aquí se sigue que si $k \geq k_0 |\psi_k||\psi_{k+|\tau|}| < |\psi_k|$, de donde $\sum_{k=0}^{\infty} \psi_k \psi_{k+|\tau|}$ es finita.

De (34) y (35) observa que $IE\ (X_t)$ y $COV[X_t, X_{t+\tau}]$ no dependen de t, y que $COV[X_t, X_{t+\tau}]$ sólo depende de τ. Entonces el proceso AR(2) es estacionario de 2^oorden bajo las condiciones (32), y en tal caso, $IE\ (X_t) = \mu_0$ y $S_\tau = \sigma_\varepsilon^2(\sum_{k=0}^{\infty} \psi_k \psi_{k+|\tau|})$.

Observa que en la condición de estacionariedad para el proceso AR(2), dada en la ecuación (32), se restringe a que las raíces del polinomio autorregresivo $\phi(z) = 1 - \phi_1 z - \phi_2 z^2$ estén fuera del circulo unitario en $\mathbb{C}$.

Se podría proseguir en forma análoga para estudiar la estacionariedad de los procesos AR(p), $p > 2$, pero el teorema 1.7 que aparece al final de esta sección tiene el objetivo de llevar a cabo esta generalizacion. A continuación se define la clase general de procesos que serán tu objeto de estudio y aplicación en el resto del curso.

1.4.4. Proceso ARMA (p,q)

Para p y q en $\{1,2,\ldots\}$, se dirá que el proceso $\{X_t\}_t$ se llama **autorregresivo y de promedios móviles** de ordenes p y q (ARMA(p,q)), si se puede escribir como

$$X_t - \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \tag{36}$$

donde μ y ϕ_j, $j = 1,2,\dots,p$, θ_j, $j = 1,2,\dots,q$, son constantes reales y el proceso $\{\varepsilon_t\}_t$ es ruido blanco con $\mathbb{E}(\varepsilon_t) = 0, \forall t$ y $VAR\ (\varepsilon_t) = \sigma_\varepsilon^2$.

Definiendo los polinomios autorregresivo $\phi(z)$ y de promedios móviles $\theta(z)$ como

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \quad , \qquad z \in \mathbb{C},$$

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q \quad , \qquad z \in \mathbb{C},$$

Se puede escribir (36) como

$$\phi(B)X_t = \theta(B)\varepsilon_t \quad ; \quad t \in T,$$

Nota que

• Si $\phi(z) \equiv 1$ $(p = 0)$, entonces $X_t = \theta(B)\varepsilon_t$ es un modelo MA(q).

• Si $\theta(z) \equiv 1$ $(q = 0)$, entonces $\phi(B)X_t = \varepsilon_t$ es un modelo AR(p).

Como ya has revisado anteriormente, para estudiar la estacionariedad de los procesos ARMA(p, q), usarás sin demostrar un teorema del libro de Brockwell y Davis (1991), para lo cual necesitas algunos resultados y definiciones previos

Definición 6. [Causalidad]

Un proceso ARMA(p, q), definido por la ecuación (36), se llama causal (función causal de $\{\varepsilon_t\}_t$), si existe una sucesión de constantes $\{\psi_j\}_j$ tales que $\sum_{j=0}^{\infty} |\psi_j| < \infty$ y

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \quad , \quad t = 0, \pm 1, \pm 2, \dots \tag{37}$$

La ecuación anterior se conoce en la literatura como **Modelo lineal general o Promedios moviles de orden +∞.**

Ejemplo:

• El proceso de ruido blanco es causal considerando en la definición de causalidad $\psi_0 = 1, \psi_j = 0, j = 1,2, \ldots$.

• El proceso MA(q) es causal considerando en la definición de causalidad $\psi_0 = 1, \psi_j = \theta_j, j = 1,2, \ldots, q$ y $\psi_j = 0$, si $j > q$.

• El proceso AR(1) es causal siempre que se tenga la condición $|\phi_1| < 1$, en cuyo caso $\psi_j = \phi_1^j, j = 0,1,2, \ldots$.

Proposición 1.5 Sea $\{Y_t\}_t$ un proceso estacionario de segundo orden con función de autocovarianza $\{S_k: k = 0, \pm 1, \pm 2, \ldots\}$. Asumiendo que $\sum_{j=0}^{\infty} |\psi_j| < \infty$, entonces la serie $\sum_{j=0}^{\infty} \psi_j Y_{t-j}$ converge absolutamente con probabilidad uno y en media cuadrática. Además, el proceso

$$X_t = \sum_{j=0}^{\infty} \psi_j Y_{t-j}, \tag{38}$$

es un proceso estacionario de segundo orden con función de autocovarianza

$$S_h^X = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \psi_j \psi_k S_{h-j+k} \tag{39}$$

La demostración de este resultado puede encontrarse en el libro de Brockwell y Davis (1991), capítulo 3, lo cual no se demuestra, dado la complejidad y corresponde a un nivel mas elevado.

Respecto a esta proposición, una observación importante es que, si se toma $Y_t = \varepsilon_t$, con $\{\varepsilon_t\}_t$ un proceso de ruido blanco en la ecuación (38), y si se conoce $\{\psi_k\}_k$, tal que $\sum_k |\psi_k| < \infty$ y el proceso $\{X_t\}_t$ satisface la ecuación (38), se obtiene que $\{X_t\}_t$ es estacionario de segundo orden y su autocovarianza está dada en (39). Entonces, si se tiene $\{X_t\}_t$ un proceso ARMA(p, q) tal que

$\{X_t\}_t$ es causal, se sigue de la proposición 1.5 que $\{X_t\}_t$ es estacionario de segundo orden con sucesión de autocovarianza $S_h^X = (\sum_{k=0}^{\infty} \psi_k \psi_{h+k})\sigma_\varepsilon^2$.

Corolario 1.6 Si $\{X_t\}_t$ es un proceso ARMA(p, q) causal, entonces $\{X_t\}_t$ es estacionario de segundo orden.

Otra aplicación de la proposición 1.5 es, si se toma $X_t = \phi_1 X_{t-1} + \varepsilon_t$ un proceso $AR(1)$ con $|\phi| < 1$. Como ya se estableció en el último ejemplo que sigue a la definición 6, $\{X_t\}_t$ es causal con $\psi_j = \phi_1^j$, entonces, de la proposición $\{X_t\}_t$ es estacionario de 2^o orden con

$$S_\tau^X = \sigma_\varepsilon \cdot \frac{\phi^{|\tau|}}{1-\phi^2},$$

que es un resultado que ya se había desarrollado cuando estudiaste el proceso AR(1).
Para terminar esta sección, se enuncia sin demostrar un teorema del libro de Brockwell y Davis (1991) que, a la luz de la última observación, establece cuándo un proceso ARMA(p, q) es estacionario de segundo orden. Sean $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ y $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$ los polinomios autorregresivo y de promedios móviles que definen al proceso ARMA(p, q).

Teorema 1.7 [Causalidad para procesos ARMA]
Sea $\{X_t\}_t$ un proceso ARMA(p,q) para el cual los polinomios $\phi(\cdot)$ y $\theta(\cdot)$ no tienen ceros en común.

Entonces $\{X_t\}_t$ es un proceso causal si y sólo si $\phi(z) \neq 0 \ \ \forall z \in \mathbb{C}$ tal que $|z| \leq 1$.

Otra condición importante en los procesos ARMA(p, q) es conocida como invertibilidad:

Definición 7 [Invertibilidad]
Un proceso ARMA(p, q) definido por la ecuación (36) se llama invertible si existe una sucesión de constantes $\{\pi_j\}_j$ tales que $\sum_{j=0}^{\infty} |\pi_j| < \infty$ y

$$\varepsilon_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} \quad , \quad t = 0, \pm 1, \pm 2, \dots \tag{40}$$

Ejemplo

Un proceso AR(p), $X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p} = \varepsilon_t$ siempre es invertible. Para ver lo anterior, sean $\pi_0 = 1, \pi_j = -\phi_j, j = 1,2, \dots, p$ y $\pi_1 = 0$ si $j \geq p + 1$.

Ejemplo

Un proceso MA(1), $X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$ tal que $|\theta_1| < 1$ es invertible. En efecto, se escribe $X_t = (1 - \alpha B)\varepsilon_t$ con $\alpha = -\theta_1$, y de la relación (14) se tiene que

$$\varepsilon_t = \sum_{k=0}^{\infty} \alpha^k B^k X_t = \sum_{k=0}^{\infty} (-\theta_1)^k X_{t-k}$$

con $\pi_j = (-\theta_1)^j, j = 0,1, \dots$, tal que $\sum_{k=0}^{\infty} |\pi_k| < \infty$.

El siguiente teorema da condiciones necesarias y suficientes para invertibilidad.

Teorema 1.8 [Invertibilidad para procesos ARMA)]

Sea $\{X_t\}_t$ un proceso ARMA(p,q) para el cual los polinomios $\phi(\cdot)$ y $\theta(\cdot)$ no tienen ceros en común, entonces $\{X_t\}_t$ es invertible si y sólo si $\theta(z) \neq 0 \;\; \forall z \in \mathbb{C}$ tal que $|z| \leq 1$. .

En el capitulo 3 del libro de Guerrero (2009) se establece que la importancia del concepto de invertibilidad radica en que todo proceso invertible está determinado de manera única por su ACF.

1.5. La sucesión de autorrelación parcial

Tal como se observó en la sección anterior, la sucesión de autocorrelación ρ_τ no resulta una herramienta útil en la identificación de un modelo AR(p) para unos datos, ya que no proporciona una idea del valor del orden de autorregresión p. Por esta razón se requiere estudiar otra estructura de dependencia del proceso estacionario de segundo orden $\{X_t\}_t$, la cual ayude a identificar posibles modelos autorregresivos. A menos que se indique lo contrario, se supondrá que $\{X_t\}_t$ es estacionario de segundo orden con media 0.

1.5.1. Procesos autorregresivos y el proceso de sucesión de autocorrelación parcial

Para definir la sucesión de autocorrelación parcial, recuerda del tema 1.3 que dada la historia del proceso $X_1, \ldots, X_k$ puedes encontrar el predictor $Proy_V(X_{k+1})$ de X_{k+1} al resolver las ecuaciones de predicción (??), donde $V = \{X \in L^2(\Omega, \ F, IP\ ) | X = \sum_{i=1}^{k} \alpha_i X_i;\ \alpha_i \in \mathbb{R}, i = 1, \ldots, k\}$.

En general, usando los mismos resultados sobre proyecciones ortogonales, dadas variables aleatorias Y y $Y_1, \ldots Y_m$ en $L^2(\Omega, \ F, IP\ )$ y considerando $U = \{X \in L^2(\Omega, \ F, IP\ ) | X = \sum_{i=1}^{m} \alpha_i Y_i;\ \alpha_i \in \mathbb{R}, i = 1, \ldots, m\}$ es posible aproximar a Y, encontrando la mejor combinación lineal $\sum_{j=1}^{m} \alpha_{m,j} Y_{m+1-j}$ en U. El adjetivo mejor es en el sentido de que $Proy_U(Y) = \sum_{j=1}^{m} \alpha_{m,j} Y_{m+1-j}$ es tal que

$$||Y - \sum_{j=1}^{m} \alpha_{m,j} Y_{m+1-j}||_2 = \inf_{X \in U} \{||Y - X||_2\}.$$

Para cada $k \in \{2,3, \ldots\}$, sean $U_k = \{X \in L^2(\Omega, \ F, IP\ ) | X = \sum_{i=2}^{k} \alpha_i X_i;\ \alpha_i \in \mathbb{R}, i = 1, \ldots, m\}$ el subespacio de $L^2(\Omega, \ F, IP\ )$ generado por las combinaciones lineales de $X_2, X_3, \ldots, X_k$.

Considera los elementos de U_k dados por $Proy_{U_k}(X_{k+1})$ y $Proy_{U_k}(X_1)$. Estos son las mejores combinaciones lineales para explicar a X_1 y X_{k+1}.

Definición 8 [Sucesión de autocorrelación parcial]

La sucesión de autocorrelación parcial (PACF) del proceso $\{X_t\}_t$ es la sucesión de correlaciones

$$\alpha(1) = Corr(X_2, X_1) = \rho_1$$

$$\text{y} \tag{41}$$

$$\alpha(k) = Corr(X_{k+1} - Proy_{U_k}(X_{k+1}), X_1 - Proy_{U_k}(X_1)), k \geq 2$$

Asumiendo que el proceso $\{X_t\}_t$ es Gaussiano, en el apéndice B del libro de Madsen (2008), se demuestra que la definición 8 de la PACF y la definición 9 abajo son equivalentes. Para cada $k = 1,2,\ldots$, considera el sistema lineal

$$\begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \vdots & & & & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{pmatrix} \begin{pmatrix} \phi_{k,1} \\ \phi_{k,2} \\ \vdots \\ \phi_{k,k} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{pmatrix} \tag{42}$$

Nota que al dividir el sistema de ecuaciones (??) por S_0 se obtiene el sistema de ecuaciones (42). En Brockwell y Davis (2009), proposición 5.1.1, se establece que este sistema tiene solución si $\{X_t\}$ es un proceso estacionario de segundo orden con sucesión de autocovarianza $\{S_\tau\}_\tau$ tal que si $h \to \infty, S_h \to 0$.

Definición 9 [Sucesión de autocorrelación parcial]

La sucesión de autocorrelación parcial (PACF) del proceso $\{X_t\}_t$ es la sucesión

$$\alpha(k) = \phi_{k,k}, \quad k \geq 1, \tag{43}$$

donde $\phi_{k,k}$ está determinada en forma única por (42).

Recordando la regla de Cramer para sistemas lineales y utilizando esta equivalencia de las definiciones 8 y 9, la PACF $\{\phi_{k,k,}|k = 1,2, \ldots\}$ del proceso $\{X_t\}$ se calcula como

$$\phi_{1,1,} = \rho_1$$

$$\phi_{2,2} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}$$

$$\phi_{3,3,} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}$$

$$etcétera\ldots$$

Se puede demostrar que cuando $\{X_t\}_t$ es un proceso AR(p) causal $X_t = \phi_1 X_{t-1} + \cdots \phi_p X_{t-p} + \varepsilon_t$, entonces

$$\phi_{k,k} = 0 \text{ para cada } k \text{ tal que } k \geq p + 1. \tag{44}$$

Es decir, para los procesos autorregresivos, la PACF se hace cero para cada lag $k$después del valor del orden p. Este comportamiento es el análogo al comportamiento de la ACF para los procesos de promedios móviles. En la siguiente unidad se dará un argumento para ver que (44) es cierta.

Por esta razón, la PACF será la correlación que resulta útil para identificar un modelo AR(p) adecuado para los datos.

Cierre de la Unidad

Los procesos estacionarios de segundo orden tienen la caracteirstica de tener momentos como media, varianza y sucesión de autocovarianza tales que no dependen del índice del tiempo t. Bajo algunas condiciones, los modelos ARMA(p, q) son de la clase de prosesos estacionaros de 2° orden cuyos elementos han sido utilizados para modelar fenómenos rales. Por ejemplo, en econometría, para describir y predecir algún indice ecónomico; en ecología, para estudiar tendencias en la temperatura global del planeta; y en general, en fenómenos en donde hay observaciones indexadas en el tiempo. El dominio de este tema te permitirá proponer modelos para datos que guardan dependencia en el tiempo.

Para saber más

La sección de estas notas trata el tema de los procesos estacionarios de segundo orden, los procesos ARMA(p, q), las sucesiones de autocovarianza y autocorrelación y su relación con la densidad espectral del proceso. Las secciones 1 y 2 de las notas de la profesora Gesine Reinert también explican los temas referentes a procesos ARMA(p, q) y sus propiedades de momentos de segundo orden.

Reinert, G. (2010). Time Series Analysis. Recuperado de http://www.stats.ox.ac.uk/~reinert/time/notesht10short.pdf

Por ultimo, un muy buen archivo de documentos que tratan el tema del aprendizaje de series de tiempo utilizando el paquete R, se encuentra en:

Revoluciones Hitos en IA, aprendizaje automático, ciencia de datos y visualización con R y Python desde 2008 (27 de junio de 2013). Aprendizaje de series temporales con R. http://blog.revolutionanalytics.com/2013/06/learning-time-series-with-r.html

Fuentes de consulta

Básica

- Brockwell, J.D. y Davis, A. (2009). *Time series: Theory and Methods*. Nueva York:
- Cowpertwait, P.S.P. (2010). *Introductory Time Series with R*. Nueva York: Springer-Verlag.
- Guerrero, V.M. (2009). *Análisis Estadístico de Series de Tiempo Económicas.* México: Just in Time press.
  R examples. Nueva York: Springer-Verlag.
- Shumway, R.H. y Stoffer, D.S. (2010). *Time Series Analysis and Its Applications: with* Springer-Verlag.