

Series de tiempo

Apunte de clases

Italo Cipriano

Universidad Técnica Federico Santa María
Segundo semestre 2024

Resumen

Se espera que al final del curso el alumno sea capaz de: *identificar-estimar-validar modelos para modelar series de tiempo, aplicarlos para predicción y control, y finalmente, usar R para modelar series de tiempo usando data real.*

Los contenidos del curso son:

1. Procesos estocásticos estacionarios
2. Modelo ARMA, ARIMA
3. Modelos ARIMA estacionales
4. Análisis espectral
5. Variables de estado y filtro de Kalman
6. Aplicaciones en economía y negocios

Las referencias del curso son:

- Chatfield, C.: “The analysis of time series. Theory and practice”, Chapman and Hall, 1975.
- Abraham, B.; Ledolter, J.: “Statistical Methods for Forecasting”, John Wiley, 1983.
- Box, G.; Jenkins, : “Time series. Forecasting and control”, Holden Day, 1976.
- Granger, C.: “Forecasting in Business and Economics”, Academic Press, 2^a Edition, 1989.
- Brockwell, P.; Davis, R.: “Time Series: Theory and Methods”, Springer Verlag, Berlin, 1987.
- Shumway, H.R; Stoffer, D.S.: “Time Series Analysis and Its Applications“, Fourth Edition, Springer Texts in Statistics, 2017.

Proceso estocásticos

[clase 1, 6 de Agosto 2024]

Palabras claves: *Proceso estocástico, realización, función de distribución, Teorema de Kolmogorov.*

Una serie de tiempo es un conjunto de observaciones x_t , cada una obtenida en el tiempo t . Una serie de tiempo discreta, es cuando las observaciones son hechas en un conjunto discreto de tiempos. Cuando las observaciones son hechas en continuamente, se usa la notación $x(t)$. Ejemplos, son el precio de una acción al cierre de los mercados durante un mes, la población de Chile medida en los últimos 10 sensores, etc.

En este curso se desarrollaran técnicas para obtener información de tales series. Y posteriormente, poder separar ruido, predecir y controlar valores futuros. Con este objetivo, lo primero es escoger (o construir) un modelo matemático que modele los datos: procesos estocásticos.

Definición 0.1 (Proceso estocástico) *Un proceso estocástico es una familia de variables aleatorias $\{X_t, t \in T\}$ definidas en un espacio de probabilidad (Ω, \mathcal{F}, P) .*

Definición 0.2 (Realización) *Las funciones $\{X_t(\omega), \omega \in \Omega\}$ con dominio T se llaman realizaciones del proceso estocástico $\{X_t, t \in T\}$.*

Veremos el teorema de Kolmogorov que sirve para garantizar la existencia de procesos estocásticos, para esto necesitamos introducir el concepto de función de distribución.

Definición 0.3 (Función de distribución) *Sea $T \subset \mathbb{R}$ y*

$$\mathcal{F} = \{t = (t_1, \dots, t_n)^t \in T^n : t_1 < t_2 < \dots < t_n, n = 1, 2, \dots\}.$$

Las funciones de distribución de $\{X_t, t \in T\}$ son las funciones $\{F_t(\cdot), t \in \mathcal{F}\}$ definidas para $t = (t_1, \dots, t_n)$ por

$$F_t(x) = P(X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n)$$

donde $x = (x_1, \dots, x_n)^t \in \mathbb{R}^n$.

El teorema siguiente dice que ciertas funciones son distribuciones ssi las distribuciones marginales en cada variable coinciden con la distribución de una dimensión menor sin la variable.

Teorema 1 (Teorema de Kolmogorov) *Las funciones $\{F_t(\cdot), t \in \mathcal{F}\}$ son distribuciones ssi para todo $n \in \{1, 2, \dots\}$, $t = (t_1, \dots, t_n) \in \mathcal{F}$ y $1 \leq i \leq n$, se tiene*

$$\lim_{x_i \rightarrow \infty} F_t(x) = F_{t(i)}(x(i)),$$

donde $t(i)$ y $x(i)$ son los vectores de $n - 1$ componentes que se obtienen al borrar la componente i -ésima de t y x respectivamente.

A continuación veremos algunos ejemplos básicos de procesos estocásticos, cuya existencia se deduce del Teorema de Kolmogorov (demostrarlo como ejercicio).

Ejemplo 1 (Proceso binario) *Un proceso binario esta dado por una secuencia de variables aleatorias independientes $\{X_t, t \in T\}$ tales que*

$$P(X_t = 1) = P(X_t = -1) = \frac{1}{2}.$$

Una realización de este proceso corresponde a secuencia de 1's y -1's.

Ejemplo 2 (Camino aleatorio) *Un proceso de camino aleatorio esta dado por una secuencia de variables aleatorias $\{S_t, t = 1, 2, 3, \dots\}$ donde $S_0 = 0$ y*

$$S_t = \sum_{i=1}^t X_i, t \geq 1,$$

donde $\{X_t, t \in T\}$ es un proceso binario.

Ejemplo 3 (Proceso de ramificación) *Un proceso de ramificación viene dado por $\{X_t, t = 1, 2, 3, \dots\}$ donde $X_0 = x$ y*

$$X_{t+1} = \sum_{j=1}^{X_t} Z_{t,j}, t = 0, 1, \dots,$$

donde $Z_{t,j}$, $t = 0, 1, \dots$, $j = 1, 2, \dots$, son variables aleatorias independientes e idénticamente distribuidas no-negativas y con valores enteros. Este modelo usualmente se usa para modelar el número de individuos de una población en el tiempo, donde x es el el número de individuos en la población en el tiempo inicial, y $Z_{t,j}$ es el número de individuos descendientes del j -ésimo individuo nacido en el tiempo t .

Durante el curso, denotaremos por serie de tiempo ya sea al conjunto de observaciones, como se menciona al inicio, o también al proceso estocástico del que son realización las observaciones.

Recuerdos de Probabilidades

[clase 2, 8 de Agosto 2024]

Palabras claves: *Variable aleatoria real X , función de distribución F_X , función de densidad f_X , función característica φ_X , esperanza E , desigualdad de Markov, desigualdad de Chebyshev, desigualdad de Jensen, varianza Var , función de distribución conjunta, función de densidad conjunta, función de covarianza Cov , función de correlación ρ , variables aleatorias independientes (i.), distribución normal $\mathcal{N}(\mu, \sigma^2)$, teorema central del límite TCL*

En esta clase recordaremos los conceptos básicos de probabilidades para variables aleatorias reales que son absolutamente continuas con respecto a la medida de Lebesgue, ya que este será habitualmente el caso en este curso, ya que es la hipótesis usual para las variables aleatorias que definen el proceso estocástico que modela la serie de tiempo.

Para esto recordemos que una variable aleatoria real es una función medible $X : \Omega \rightarrow \mathbb{R}$ definida en un espacio de probabilidad (Ω, \mathcal{F}, P) .

La definición de función medible es simplemente para poder usar la probabilidad P para asociar un comportamiento aleatorio a la función X mediante la definición

$$P(X \in S) = P\{\omega \in \Omega : X(\omega) \in S\}$$

para todo conjunto medible $S \subset \mathbb{R}$. Con esta definición (y los teoremas del curso de medida) se puede 'codificar' la probabilidad P mediante ciertas funciones que dependen de X , y que son fundamentales en Teoría de Probabilidades, tales como: las funciones de distribución, y las funciones características, y hay otras como los momentos, que no recordaremos acá.

Definición 0.4 (Función de distribución) *Una variable aleatoria $X : \Omega \rightarrow \mathbb{R}$ en un espacio de probabilidad (Ω, \mathcal{F}, P) tiene densidad f_X , donde f_X es una función Lebesgue integrable no negativa, si*

$$P(X \in [a, b]) = \int_a^b f_X(x) dx.$$

Definición 0.5 (Esperanza) *La esperanza (o primer momento) de una variable aleatoria real X con distribución f_X viene dada por*

$$E(X) := EX := \int x f_X(x) dx.$$

Recordemos algunas propiedades importantes de la esperanza.

Proposición 1

- $X \geq 0 \Rightarrow EX \geq 0$.
- $\forall X, Y$ v.a. y a constante, se tiene

$$E(X + Y) = EX + EY,$$

$$E(aX) = aEX.$$

- $X \leq Y$ c.s. $\Rightarrow EX \leq EY$.
- $X = Y$ c.s. $\Rightarrow EX = EY$.
- $|EX| \leq E|X|$.
- $EX = \int_{\mathbb{R}} x dF_X(x)$ donde $F_X(x) = P(X \leq x)$.
- Si $g : \mathbb{R} \rightarrow \mathbb{R}$ es medible, entonces

$$E(g(X)) = \int_{\mathbb{R}} g(x) f_X(x) dx.$$

- *Desigualdad de Markov.*

$$P(X \geq a) \leq \frac{EX}{a}.$$

- *Desigualdad de Chebyshev.*

$$P(|X - EX| \geq a) \leq \frac{\text{Var}(X)}{a^2},$$

donde $\text{Var}(X) = E((X - EX)^2)$.

- *Desigualdad de Jensen.* Si $f : \mathbb{R} \rightarrow \mathbb{R}$ es convexa y $EX < \infty$, entonces

$$f(EX) \leq E(f(X)).$$

Definición 0.6 (Función característica) Una variable aleatoria $X : \Omega \rightarrow \mathbb{R}$ en un espacio de probabilidad (Ω, \mathcal{F}, P) tiene función característica

$$\varphi_X(t) = E[e^{itX}].$$

Recordemos las definiciones de varianza y covarianza de una variable aleatoria.

Definición 0.7 (Varianza) La varianza de una variable aleatoria real X viene dada por

$$Var(x) = E((X - EX)^2).$$

Definición 0.8 (Función de distribución acumulada conjunta) Dadas dos variables aleatorias X, Y sobre un mismo espacio de probabilidad (Ω, \mathcal{F}, P) , la función de distribución acumulada conjunta viene dada por

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y).$$

Definición 0.9 (Función de densidad conjunta) Dadas dos variables aleatorias X, Y sobre un mismo espacio de probabilidad (Ω, \mathcal{F}, P) , la función de densidad conjunta viene dada por

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y}.$$

Definición 0.10 Dos variables aleatorias X, Y son independientes ssi

$$F_{X,Y}(x, y) = F_X(x)F_Y(y).$$

Y si son absolutamente continuas (como usualmente en este curso), X, Y son independientes ssi

$$f_{X,Y}(x, y) = f_X(x)f_Y(y).$$

Luego, para describir como varían dos variables aleatorias entre ellas se introduce el concepto de covarianza.

Definición 0.11 (Covarianza) *La covarianza de dos variables aleatorias X y Y con distribución conjunta a valores reales viene dada por*

$$\sigma_{X,Y} := Cov(X, Y) := E[(X - EX)(Y - EY)].$$

Otra medida de como varían dos variables aleatorias es la correlación entre X e Y , que viene dada por

$$\rho_{X,Y} := Corr(X, Y) := \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}.$$

Observar que $Var(X) = Cov(X, X)$, entonces la covarianza generaliza el concepto de varianza. También recordemos la siguiente propiedad útil para calcular la covarianza.

Proposición 2

$$Cov(X, Y) = \int_{\mathbb{R}} \int_{\mathbb{R}} (F_{X,Y}(x, y) - F_X(x)F_Y(y)) dx dy.$$

Recordemos ahora un ejemplo básico de distribución, la distribución normal.

Ejemplo 4 (Distribución normal) *Una variable aleatoria X a valores reales tiene distribución normal de parámetros μ, σ ssi su distribución de probabilidad esta dada por*

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

En este caso se denota $X \sim \mathcal{N}(\mu, \sigma^2)$. Si $X \sim \mathcal{N}(\mu, \sigma^2)$ entonces

$$E(X) = \int_{\mathbb{R}} x f_X(x) dx = \mu$$

y

$$Var(X) = E[(X - EX)^2] = \int_{\mathbb{R}} (x - EX)^2 f_X(x) dx = \sigma^2,$$

además la función característica viene dada por

$$\varphi_X(t) = \exp\left(it\mu - \frac{1}{2}\sigma^2 t^2\right).$$

Cuando $\mu = 0$ y $\sigma = 1$ se dice distribución normal estandar.

Finalmente, recordemos el teorema central del límite (TCL), en sus versiones de la ley fuerte y ley débil de los grandes números, y el teorema de Lindeberg–Lévy.

Teorema 2 (Ley fuerte de los grandes números) *Si X_1, X_2, \dots es una secuencia de variables aleatorias independientes i.i.d. (independientes e idénticamente distribuidas), Lebesgue integrables y tales que $E(X_1) = \mu$, entonces*

$$P\left(\lim_{n \rightarrow \infty} \frac{X_1 + \dots + X_n}{n} = \mu\right) = 1.$$

Teorema 3 (Ley débil de los grandes números) *Si X_1, X_2, \dots es una secuencia de variables aleatorias independientes i.i.d. (independientes e idénticamente distribuidas), Lebesgue integrables y tales que $E(X_1) = \mu$ y $\text{Var}(X_1) = \sigma^2$, entonces*

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| < \epsilon\right) = 1.$$

Teorema 4 (Teorema de Lindeberg–Lévy) *Si X_1, X_2, \dots es una secuencia de variables aleatorias independientes i.i.d. (independientes e idénticamente distribuidas), Lebesgue integrables y tales que $E(X_1) = \mu$ y $\text{Var}(X_1) = \sigma^2$, entonces*

$$\sqrt{n} \left(\frac{X_1 + \dots + X_n}{n} - \mu \right) \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

donde \xrightarrow{d} significa que la convergencia es en distribución.

Vectores aleatorios y normal multivariada

[clase 3, 13 de Agosto 2024]

El objetivo de esta clase es definir la normal multivariada, que es un vector aleatorio a \mathbb{R}^n que queda determinado de manera única por un vector $\mu \in \mathbb{R}^n$ y por una matriz simétrica no-definida negativa $\Sigma \in \mathbb{R}^{n \times n}$. La normal multivariada generaliza las variables aleatorias normales (que son uni-dimensionales) a dimensiones mayores, y tiene buenas propiedades, tales como que el hecho que condicionar algunas coordenadas con respecto al complemento preserva el hecho de ser una normal multivariada. Las demostraciones de las propiedades pasan por conocer la función característica de una normal multivariada.

Partiremos la clase definiendo un vector aleatorio a \mathbb{R}^n , su esperanza, y su matriz de covarianza. Decimos que un vector $X = (X_1, \dots, X_n)'$ es un vector aleatorio n -dimensional si cada componente es una variable aleatoria. Si $E|X_i| < \infty$ para cada i , se puede definir el valor esperado de X por

$$\mu_X = EX = (EX_1, \dots, EX_n)'$$

Si $X = (X_1, \dots, X_n)'$ y $Y = (Y_1, \dots, Y_m)'$ son vectores aleatorios con

$$E|X_i|^2, E|Y_j|^2 < \infty \text{ para todo } i, j,$$

se define la matriz de covarianza de X e Y como la matriz

$$\Sigma_{X,Y} = Cov(X, Y) = E[(X - EX)(Y - EY)'] = E(XY') - (EX)(EY)'$$

La componente (i, j) -ésima de $\Sigma_{X,Y}$ corresponde a

$$(\Sigma_{X,Y})_{i,j} = Cov(X_i, Y_j) = E(X_i Y_j) - E(X_i)E(Y_j).$$

Para un vector aleatorio $X = (X_1, \dots, X_n) \in \mathbb{R}^n$, la función de densidad conjunta corresponde a una función $f_{X_1, \dots, X_n}(x_1, \dots, x_n)$ tal que

$$P[X_1, \dots, X_n \in D] = \int \cdots \int_D f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \cdots dx_n.$$

La siguiente proposición será clave para estudiar las propiedades de la normal multivarida más adelante.

Proposición 3 Si $a \in \mathbb{R}^m$, $B \in \mathbb{R}^{m \times n}$ y se tiene el vector aleatorio $X = (X_1, \dots, X_n)'$ donde $E|X_i|^2 < \infty$ para todo $i = 1, \dots, n$, entonces el vector aleatorio

$$Y = a + BX$$

tiene media

$$EY = a + BEX$$

y matriz de covarianza

$$\Sigma_{Y,Y} = B\Sigma_{X,X}B'.$$

Ejercicio 1 Demuestre la Proposición 3.

Con respecto a lo necesario de algebra lineal, partamos recordando el concepto de matriz diagonalizable y una propiedad que garantiza diagonalizabilidad. En este curso generalmente usaremos una versión muy particular de matriz diagonalizable, de todos modos por completitud recordemos la definición general.

Sea F un cuerpo y n, m enteros positivos. Denotamos por $F^{n \times m}$ el conjunto de matrices de $n \times m$. Y denotamos por $GL_n(F)$, llamado grupo general lineal de matrices cuando se considera con la operación de producto de matrices, el conjunto de matrices invertibles en $F^{n \times n}$.

Definición 0.12 (Matriz diagonalizable) Una matriz $A \in F^{n \times n}$ se dice diagonalizable ssi existe $P \in GL_n(F)$ tal que $P^{-1}AP$ es una matriz diagonal.

Un ejemplo importante de matrices diagonalizables reales son las matrices reales simétricas ($M \in \mathbb{R}^{n \times n}$ con $M = M'$). Más aún, si adicionalmente la matriz es no definida negativa (una matriz real simétrica M se dice no definida negativa si $x'Mx \geq 0$ para todo vector $x \in \mathbb{R}^n$) entonces P es una matriz ortogonal, i.e., $P' = P^{-1}$ y la matriz diagonal viene dada por los valores propios. Esto es el contenido de la siguiente proposición.

Proposición 4 Sea $\Sigma \in \mathbb{R}^{n \times n}$ simétrica y no definida negativa. Entonces

$$\Sigma = P\Lambda P'$$

donde $P' = P^{-1}$ (P es ortogonal) y $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, donde λ_i 's son los valores propios de Σ . En particular,

$$\det \Sigma = \prod_{i=1}^n \lambda_i.$$

Proposición 5 La matriz de covarianza $\Sigma_{X,X}$ es simétrica y no-definida negativa.

Dem. La simetría es clara de la definición. Sea $x \in \mathbb{R}^n$ un vector arbitrario, entonces

$$\begin{aligned} x'\Sigma_{X,X}x &= x'[E(XX') - (EX)(EX)']x \\ &= x'E(XX')x - x'(EX)(EX)'x \\ &= E[(x'X)(x'X)'] - E(x'X)E(x'X)' \\ &= \text{Var}(x'X) \\ &= E[(x'X - E(x'X))^2] \geq 0. \end{aligned}$$

□

Definición 0.13 (Distribución normal multivariada) *Un vector aleatorio $Y = (Y_1, \dots, Y_n)'$ se dice normal multivariado o que tiene distribución normal multivariada ssi existe un vector columna a , una matriz B y un vector aleatorio $X = (X_1, \dots, X_n)'$ donde las componentes son variables aleatorias independientes con distribución normal estandard, tales que*

$$Y = a + BX. \quad (1)$$

Observación 1 *De la Proposición 3, si Y es normal multivariada definida como en (1) entonces*

$$\begin{aligned} EY &= a, \\ \Sigma_{Y,Y} &= BB'. \end{aligned}$$

Para estudiar la distribución normal multivariada, hay dos resultados que son muy útiles, el primero es la Proposición 3, el segundo es su función característica, que veremos a continuación.

Proposición 6 (Función característica de la normal multivariada) *Si Y es normal multivariada definida como en (1) entonces su función característica $\phi_Y : \mathbb{R}^n \rightarrow \mathbb{R}$ viene dada por*

$$\phi_Y(u) := E(e^{iu'Y}) = \exp\left(iu'\mu - \frac{1}{2}u'\Sigma_{Y,Y}u\right), u \in \mathbb{R}^n.$$

Finalmente, enunciemos las propiedades siguientes, que en particular permiten caracterizar las normales multivariadas como mencionamos en la introducción de esta clase.

Proposición 7 *Para todo $\mu \in \mathbb{R}^n$, para toda matriz simétrica no-definida negativa $\Sigma \in \mathbb{R}^{n \times n}$, existe una normal multivariada con media μ y matriz de covarianza Σ .*

Proposición 8 *Una normal multivariada queda únicamente determinada por su media y su matriz de covarianza.*

Observación 2 *Si Y es normal multivariada con $EY = \mu$ y $\Sigma_{Y,Y} = \Sigma$, anotamos*

$$Y \sim \mathcal{N}(\mu, \Sigma).$$

Proposición 9 $Y \sim \mathcal{N}(\mu, \Sigma)$ ssi $\forall a \in \mathbb{R}^n$, $a'Y \sim \mathcal{N}(a'\mu, a'\Sigma a)$.

Finalmente, la última proposición la escribimos en palabras, en estas notas incluimos los detalles.

Proposición 10 Si Y es normal multivariada y particionamos el vector Y de la forma $Y = (Y^{(1)}, Y^{(2)})$, donde $E(Y^{(i)}) = \mu_i$ y

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

para $\Sigma_{ij} = \text{Cov}(Y^{(i)}, Y^{(j)})$. Entonces

1. $Y^{(1)}$ e $Y^{(2)}$ son independientes ssi $\Sigma_{12} = 0$.
2. Si $\det(\Sigma_{22}) > 0$, entonces la distribución de $Y^{(1)}$ dado $Y^{(2)}$ es

$$\mathcal{N}\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(Y^{(2)} - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right).$$

Motivaciones, Filtros lineales y series de tiempo estacionarias

[clase 4, 20 de Agosto 2024]

En esta clase nos enfocaremos en una primera mirada práctica para trabajar con series de tiempo. Por eso partiremos con ejemplos que muestran porque las series de tiempo aparecen naturalmente en cualquier proceso predictivo cuando hay disponibilidad de datos. Cuando uno trabaja con una serie de tiempo el primer paso es graficar los datos y el primer objetivo natural es suavizar los datos (a veces se dice suavizar la señal) para remover las fluctuaciones rápidas. Con este objetivo se definen los filtros lineales. Al final de la clase definiremos las series de tiempos estacionarias que serán importantes más adelante en el curso.

Motivaciones prácticas

Ejemplo 5 (Finanzas) Para contener la inflación el banco central decide una tasa de politica monetaria, para obtener tal número es necesario hacer varias predicciones, tales como predecir el IPC, el consumo, el cambio del precio del dolar, las inversiones futuras en diversas areas, el nivel de endeudamiento de las personas y las empresas, etc.

Ejemplo 6 (Stock) *En el problema de tener stock suficiente para satisfacer la demanda, hace necesario predecir la demanda, ya que el exceso y el deficit de stock son no deseados.*

Ejemplo 7 (Fallas en cadena de producción) *En una cadena de producción en que se utilizan máquinas, es necesario predecir la probabilidad con que se produzcan fallas en las máquinas, de modo de poder anticiparse y tomar decisiones considerando los costos estimados de tener que detener la producción así como el de tener presupuesto y para el reemplazo de las máquinas, o máquinas compradas para reemplazo con anticipación.*

Ejemplos prácticos en Chile actual son:

- Podemos predecir el número de médicos en los próximos 10 años? Esta predicción permitiría por ejemplo elaborar políticas para fomentar o disminuir el número de vacantes disponibles para estudiar medicina, así como planificar la infraestructura para que todos los médicos puedan trabajar adecuadamente, y que tal número sea acorde a la demanda estimada de acuerdo a la población de Chile en los próximos 10 años. Una falla en tal predicción significa no solo un costo, sino que también eventualmente no poder satisfacer una demanda, ya sea porque no habrá un número suficiente de médicos, o porque el sistema no dará abasto para poder dar trabajo a todos los médicos disponibles. El problema es más general, ya que todas las carreras necesitan ser analizadas.
- Podemos predecir el impacto de la inteligencia artificial en los puestos de trabajo en los próximos 10 años? Que puestos de trabajo se verán más afectados? Como debemos tomar las decisiones hoy para enfrentar de mejor forma la demanda y oferta de puestos de trabajo en 10 años?
- Podemos predecir cuantos postes de luz se caeran en el próximo temporal o terremoto en Chile? Las compañías de luz seguramente tendrán que hacer un estudio utilizando series de tiempo para poder disponibilizar de métodos para poder acortar los tiempos de reparaciones.

Filtros lineales

Definición 0.14 (Filtros lineales) Dada una serie de tiempo $\{Z_t, t \in \mathbb{Z}\}$ un filtro lineal consiste en una serie $\{Y_t, t \in \mathbb{Z}\}$ obtenida a partir de la fórmula

$$Y_t = \sum_{j=-\infty}^{\infty} \tau_j Z_{t-j},$$

donde $\tau_j \in \mathbb{R}_{\geq 0}, \forall j \in \mathbb{Z}$ y satisface

$$\sum_{j=-\infty}^{\infty} \tau_j = 1.$$

La idea del filtro lineal es preservar la media y disminuir la varianza de la serie de tiempo, ver siguiente observación.

Observación 1 Si $\{Z_t, t \in \mathbb{Z}\}$ es un proceso donde Z_t son independientes, $EZ_t = \mu$ y $VarZ_t = \sigma^2$ para todo t , entonces

$$EY_t = \mu$$

y además

$$VarY_t = \sum_j Var(\tau_j Z_{t-j}) = \sum_j \tau_j^2 Var(Z_{t-j}) = \sigma^2 \sum_j \tau_j^2 \leq \sigma^2.$$

A continuación veremos ejemplos de filtros lineales.

Ejemplo 8 (Filtro de media móvil) En este filtro

$$\tau_j = \begin{cases} 1/N & \text{si } -N \leq j \leq 0, \\ 0 & \text{en otro caso.} \end{cases}$$

En este caso Y_t corresponde al promedio de $Z_{t-N-1}, Z_{t-N}, \dots, Z_{t-1}, Z_t$.

Ejemplo 9 (Filtro de media móvil simétrico) En este filtro

$$\tau_j = \begin{cases} 1/(2N+1) & \text{si } -N \leq j \leq N, \\ 0 & \text{en otro caso.} \end{cases}$$

En este caso Y_t corresponde al promedio de $Z_{t-N}, \dots, Z_t, \dots, Z_{t+N}$.

Ejemplo 10 (Filtro de media móvil ponderada) *En este filtro*

$$\tau_j = \begin{cases} w_j & \text{si } -N \leq j \leq N, \\ 0 & \text{en otro caso .} \end{cases}$$

En este caso Y_t corresponde al promedio ponderado por w_{-N}, \dots, w_N de $Z_{t-N}, \dots, Z_t, \dots, Z_{t+N}$.

Ejemplo 11 (Filtro de suavizamiento exponencial) *En este filtro*

$$\tau_j = \begin{cases} \alpha(1 - \alpha)^j & \text{si } 0 \leq j, \\ 0 & \text{en otro caso ,} \end{cases}$$

donde $0 < \alpha < 1$. En este caso

$$Y_t = \alpha Z_t + \alpha(1 - \alpha)Z_{t-1} + \alpha(1 - \alpha)^2 Z_{t-2} + \dots .$$

De manera recursiva, el suavizamiento exponencial puede definirse como

$$Y_t = \alpha Z_t + (1 - \alpha)Y_{t-1},$$

donde $Y_0 = Z_1$.

Ejercicio 2 *Defina el filtro de Hodrick–Prescott. Implementelo en Python, incluya el código que utilizó. Incluya un gráfico que muestre como funciona en un ejemplo con data ficticia.*

Series de tiempo estacionarias

El concepto de matriz de covarianza se puede generalizar para un proceso estocástico $\{X_t, t \in T\}$, de modo de obtener información de la dependencia entre las variables.

Definición 0.15 (Función de autocovarianza) *Si un proceso $\{X_t, t \in T\}$ tiene $\text{Var}(X_t) < \infty$ para cada $t \in T$, entonces la función de autocovarianza asociada $\gamma(\cdot, \cdot)$ se define como*

$$\gamma(r, s) = \text{Cov}(X_r, X_s) = E[(X_r - EX_r)(X_s - EX_s)], r, s \in T.$$

A continuación definiremos los conceptos de estacionaridad (usualmente llamada también estacionaridad débil) y estacionaridad estricta. La hipótesis de estacionaridad es útil en algoritmos de procesamiento de señales, que veremos más adelante. Cuando los datos no son estacionarios, veremos técnicas para modelarlos como procesos estacionarios.

Definición 0.16 (Estacionaridad) *Una serie de tiempo (en este caso nos referimos a un proceso estocástico) $\{X_t, t \in \mathbb{Z}\}$ se dice estacionario si*

1. $E|X_t|^2 < \infty$ para todo $t \in \mathbb{Z}$,
2. $EX_t = m$ para todo $t \in \mathbb{Z}$,
3. $\gamma(r, s) = \gamma(r + t, s + t)$ para todo $r, s, t \in \mathbb{Z}$.

En lo siguiente definiremos estacionaridad estricta.

Definición 0.17 (Estacionaridad estricta) *Una serie de tiempo $\{X_t, t \in \mathbb{Z}\}$ se dice estacionaria estricta si (X_1, \dots, X_k) y $(X_{1+h}, \dots, X_{k+h})$ tienen la misma distribución para todo $k \in \mathbb{N}, h \in \mathbb{Z}$.*

Es importante notar que un proceso estrictamente estacionario con $E|X_t|^2 < \infty$ es estacionario. Sin embargo el converso en general no es cierto, basta considerar el proceso $\{X_t, t \in \mathbb{Z}\}$ donde X_1 y X_2 tienen distintas distribuciones con $E|X_t|^2 < \infty$ y $EX_t = m$ para todo t , la distribución de X_t para t impar es igual a la distribución de X_1 y son independientes y la distribución de X_t para t par es igual a la distribución de X_2 y son independientes.

Propiedades de la función de autocovarianza de procesos estacionarios

[clase 5, 22 de Agosto 2024]

Recordar que cuando $\{X_t, t \in \mathbb{Z}\}$ es un proceso estacionario entonces la función de autocovarianza

$$\gamma(r, s) := Cov(X_r, X_s) := E((X_r - EX_r)(X_s - EX_s))$$

satisface

$$\gamma(r, s) = \gamma(r + t, s + t) \quad \forall r, s, t \in \mathbb{Z},$$

luego para un procesos estacionario la autocovarianza se puede definir como una función en una variable

$$\gamma(h) := \gamma(h, 0) = \text{Cov}(X_{t+h}, X_t) \quad \forall t, h \in \mathbb{Z}.$$

La función $\gamma(h)$ se conoce como autocovarianza del proceso estacionario $\{X_t, t \in \mathbb{Z}\}$. También se define la función de autocorrelación

$$\rho(h) := \gamma(h)/\gamma(0) = \text{Corr}(X_{t+h}, X_t) \quad \forall t, h \in \mathbb{Z}.$$

La función de autocovarianza de un proceso estacionario satisface la siguiente propiedad.

Proposición 11 *Si γ es la función de autocovarianza de un proceso estacionario, entonces*

1. $\gamma(0) \geq 0$,
2. $|\gamma(h)| \leq \gamma(0), \forall h \in \mathbb{Z}$, y
3. γ es par, i.e. $\gamma(h) = \gamma(-h), \forall h \in \mathbb{Z}$.

Dem.

1. $\gamma(0) = \text{Cov}(X_t, X_t) = \text{Var}(X_t) \geq 0$.
2. Por desigualdad de Cauchy-Schwartz

$$|\gamma(h)| = |\text{Cov}(X_{t+h}, X_t)| \leq (\text{Var}(X_{t+h}))^{1/2} (\text{Var}(X_t))^{1/2}.$$

3. $\gamma(-h) = \text{Cov}(X_{t-h}, X_t) = \text{Cov}(X_t, X_{t+h}) = \gamma(h)$. \square

Más aún, pueden caracterizarse las funciones de autocovarianza de procesos estacionarios, similarmente a como se hizo con las matrices asociadas a una normal multivariada, ver Proposición 7.

Definición 0.18 *Una función $k : \mathbb{Z} \rightarrow \mathbb{R}$ se dice que es no definida negativa si satisface $\forall n \in \mathbb{N}, \forall (a_1, \dots, a_n)' \in \mathbb{R}^n, \forall (t_1, \dots, t_n)' \in \mathbb{Z}^n$,*

$$\sum_{i,j=1}^n a_i k(t_i - t_j) a_j \geq 0.$$

Teorema 5 Una función $k : \mathbb{Z} \rightarrow \mathbb{R}$ es una función de autocovarianza de un proceso estacionario si y solo si k es par y es no definida negativa.

Dem. La condición de paridad es necesaria por Propiedad 11, parte 3. Veamos que la condición de no definida negativa también es necesaria. Asumamos para esto que k es la función de autocovarianza de un proceso estacionario $\{X_t\}$ y sean $n \in \mathbb{N}$ $a = (a_1, \dots, a_n)' \in \mathbb{R}^n$, $t = (t_1, \dots, t_n)' \in \mathbb{R}^n$, y

$$Z_t = (X_{t_1} - EX_{t_1}, \dots, X_{t_n} - EX_{t_n})',$$

entonces

$$\begin{aligned} 0 &\leq \text{Var}(a'Z_t) \\ &= a'E(Z_t(Z_t)')a \\ &= a'Ka \\ &= \sum_{i,j=1}^n a_i k(t_i - t_j) a_j, \end{aligned}$$

donde $K = [k(t_i - t_j)]_{i,j}$ es la matriz de covarianza de $(X_{t_1}, \dots, X_{t_n})'$. Para demostrar que las condiciones son también suficiente construiremos con la ayuda del teorema de Kolmogorov un proceso estacionario con función de autocovarianza k . Para esto, sea k par y no definida negativa. Sea $n \in \mathbb{N}$ y $t = (t_1, \dots, t_n)' \in \mathbb{R}^n$ tal que $t_1 < t_2 < \dots < t_n$. Sea F_t la distribución en \mathbb{R}^n con función característica (recordar Proposición 6)

$$\phi_t(u) = \exp(-u'Ku), u \in \mathbb{R}^n,$$

para $K = [k(t_i - t_j)]_{i,j}$. Ya que k es no definida negativa, entonces K es no definida negativa, luego ϕ_t es la función característica de una distribución normal multivariada $\mathcal{N}(0, K)$. Es fácil observar que la condición del teorema de Kolmogorov se satisface, por lo que el Teorema de Kolmogorov garantiza la existencia de tal proceso estocástico, que llamaremos $\{X_t\}$. Queda demostrar que el proceso obtenido es estacionario. La distribución conjunta de X_i y X_j es una normal multivariada con media 0 y matriz de covarianza

$$\begin{pmatrix} k(0) & k(i-j) \\ k(j-i) & k(0) \end{pmatrix}$$

luego $\text{Cov}(X_i, X_j) = k(i, j)$, lo que concluye la demostración. \square

Ejemplos procesos estacionarios

[clase 6, 27 de Agosto 2024]

En esta clase veremos 3 ejemplos en el contexto de procesos estacionarios. Primero veremos ejemplos de procesos que son y otros que no son estacionarios. Luego veremos un ejemplo para estimar la población de Chile usando series estacionarias, y finalmente mostraremos un ejemplo que aplica el teorema de Proyección en espacios de Hilbert para hacer estimaciones en series de tiempos estacionarias.

Ejemplos de procesos estacionarios y procesos no estacionarios

Ejemplo 12 Sea $\{Z_t\}$ iid con $EZ_t = 0, \text{Var}Z_t = \sigma^2$. Sea $X_t = Z_t + \theta Z_{t-1}$, para θ una constante. Este proceso es estacionario, ya que se tiene por linealidad de la esperanza que $EX_t = 0$ para todo t , además

$$EX_t^2 = E(Z_t^2 + \theta^2 Z_{t-1}^2 + 2\theta Z_{t-1}) = E(Z_t^2) + \theta^2 E(Z_{t-1}^2) + 2\theta E(Z_{t-1}),$$

donde $E(Z_t^2) = E(Z_{t-1}^2)$ y $E(Z_{t-1}) = 0$. Además, si σ^2 es finito, entonces $\text{Var}(Z_t) = E(Z_t^2) - E(Z_t)^2 = E(Z_t^2) = \sigma^2$, luego

$$EX_t^2 = \sigma^2(1 + \theta^2) < \infty,$$

y finalmente, la función de autocovarianza viene dada por

$$\begin{aligned} \text{Cov}(X_{t+h}, X_t) &= \text{Cov}(Z_{t+h} + \theta Z_{t+h-1}, Z_t + \theta Z_{t-1}) \\ &= \begin{cases} (1 + \theta^2)\sigma^2 & \text{si } h = 0, \\ \theta\sigma^2 & \text{si } h = \pm 1, \\ 0 & \text{si } |h| > 1, \end{cases} \end{aligned}$$

que no depende de t , por lo que $\{X_t\}$ es estacionario.

Ejemplo 13 Sea $\mu \in \mathbb{R}_{>0}$ una constante, $\{Y_t\}$ un proceso estacionario y

$$X_t = \begin{cases} Y_t & \text{si } t \text{ es múltiplo de } 5, \\ Y_t + \mu & \text{en otro caso,} \end{cases}$$

luego $\{X_t\}$ no es estacionario, pues EX_t no es constante.

Ejemplo 14 Sea $\{Y_t\}$ un proceso iid con $EY_t = 0$ y $VarY_t\sigma^2$. Luego el proceso $\{X_t\}$ definido por $X_t = Y_1 + \dots + Y_t$ no es estacionario, ya que la función de autocovarianza depende de t, h como se ve a continuación:

$$\begin{aligned} Cov(X_{t+h}, X_t) &= Cov(X_1 + \dots + X_{t+h}, X_1 + \dots + X_t) \\ &= Cov(X_1 + \dots + X_t, X_1 + \dots + X_t) \\ &= t\sigma^2. \end{aligned}$$

Ejemplo 15 En este ejemplo definiremos un caso particular de procesos en que si el proceso es estacionario, entonces es también estrictamente estacionario. Esta implicancia en general no se tiene, como vimos anteriormente.

Definición 0.19 (Proceso Gaussiano) Un proceso $\{X_t\}$ se dice Gaussiano si cada X_t es una normal multivariada.

Si $\{X_t\}$ es un proceso estacionario Gaussiano, entonces el proceso es estrictamente estacionario. En efecto, sean $h, t_1 < t_2 \in \mathbb{Z}$, entonces $(X_{t_1}, \dots, X_{t_2})'$ y $(X_{t_1+h}, \dots, X_{t_2+h})'$ tienen la misma media y matriz de covarianza, entonces la misma distribución.

Estimaciones simples para cierta serie de tiempo idealizadas

Supongamos queremos modelar la población de Chile en el tiempo t . Consideramos el modelo en que

$$X_t = m_t + Y_t,$$

donde m_t es una función de t que “varía poco” en t , e Y_t es un proceso estacionario con $EY_t = 0$. Supongamos que tenemos datos $\{y_t\}_{t=t_0}^T$ de la población de Chile en los tiempos t_0, \dots, T . Ajustamos una curva $m_t = a_0 + a_1t + a_2t^2$ a los datos, minimizando

$$\left\{ \sum_t |y_t - m_t|^2 : a_0, a_1, a_2 \in \mathbb{R} \right\}.$$

Una vez encontrados $\hat{a}_0, \hat{a}_1, \hat{a}_2$ que alcanzan el mínimo, usamos $\hat{m}_t = \hat{a}_0 + \hat{a}_1t + \hat{a}_2t^2$ como predictor de futuros valores de X_t . Así, para estimar la población en el tiempo $t + h$, podemos estimar el valor esperado de la población con nuestro modelo, obteniendo $X_{t+h} = \hat{m}_{t+h}$.

Aplicación del Teorema de proyección en espacios de Hilbert para estimaciones en series de tiempo estacionarias

Sea $\{X_t, t \in \mathbb{Z}\}$ un proceso estocástico estacionario con media $E(X_t) = 0$ y función de autocovarianza $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$. Consideremos el problema de encontrar la combinación lineal de X_1, \dots, X_n que mejor aproxima a X_{n+1} , en el sentido que minimiza

$$E|X_{n+1} - \sum_{j=1}^n \alpha_j X_j|^2,$$

donde $\alpha_1, \dots, \alpha_n \in \mathbb{R}$.

Recordemos el teorema de proyección en espacios de Hilbert.

Teorema 6 (Teorema de proyección) *Sea \mathcal{H} un espacio de Hilbert, \mathcal{M} un subespacio cerrado y $x \in \mathcal{H}$. Entonces*

1. *existe un único $\hat{x} \in \mathcal{M}$ tal que $\|x - \hat{x}\| = \inf_{y \in \mathcal{M}} \|x - y\|$, y*
2. *$\hat{x} \in \mathcal{M}$ tal que $\|x - \hat{x}\| = \inf_{y \in \mathcal{M}} \|x - y\|$ ssi $\hat{x} \in \mathcal{M}$ y $(x - \hat{x}) \in \mathcal{M}^\perp$.*

Recordemos que $L^2 = L^2(\Omega, \mathcal{F}, P)$, el espacio de variables aleatorias con segundo momento finito módulo $X \sim Y$ ssi $P(X = Y) = 1$, es un espacio de Hilbert con el producto interno $\langle x, y \rangle := E(xy)$. Como $\{X_t, t \in \mathbb{Z}\}$ es estacionario, en particular $E|X_t|^2 < \infty$, entonces $X_t \in L^2$, para todo $t \in \mathbb{Z}$. Observemos que

$$\inf \left\{ E|X_{n+1} - \sum_{j=1}^n \alpha_j X_j|^2 : \alpha_1, \dots, \alpha_n \in \mathbb{R} \right\} = \inf_{\hat{X}_{n+1} \in \mathcal{M}} \|X_{n+1} - \hat{X}_{n+1}\|^2$$

para $\mathcal{M} = \{\sum_{j=1}^n \alpha_j X_j : \alpha_1, \dots, \alpha_n \in \mathbb{R}\}$. Luego el Teorema 6 garantiza que existe un único $\hat{X}_{n+1} \in \mathcal{M}$ que alcanza el ínfimo. Para encontrar \hat{X}_{n+1} podemos usar caracterización del ínfimo en la segunda implicancia del Teorema 6, observando que por definición de ortogonalidad $(x - \hat{x}) \in \mathcal{M}^\perp$ ssi $\langle x - \hat{x}, y \rangle = 0$ para todo $y \in \mathcal{M}$. Ya que \mathcal{M} es el subespacio generado por

$\{X_1, \dots, X_n\}$, luego la condición $\langle x - \hat{x}, y \rangle = 0$ para todo $y \in \mathcal{M}$ se puede escribir en este caso como un sistema de n ecuaciones:

$$\langle X_{n+1} - \sum_{j=1}^n \alpha_j X_j, X_k \rangle = 0, k = 1, 2, \dots, n.$$

Donde

$$\langle X_{n+1} - \sum_{j=1}^n \alpha_j X_j, X_k \rangle = \langle X_{n+1}, X_k \rangle - \sum_{j=1}^n \alpha_j \langle X_j, X_k \rangle,$$

y con el supuesto que $E(X_t) = 0$ se obtiene que para γ la función de autocovarianza que $\langle X_{n+1}, X_k \rangle = Cov(X_{n+1}, X_k) = \gamma(n+1-k)$ y $\langle X_j, X_k \rangle = Cov(X_j, X_k) = \gamma(j-k)$, luego el sistema de ecuaciones puede reescribirse como

$$\gamma(n+1-k) = \sum_{j=1}^n \alpha_j \gamma(j-k), k = 1, 2, \dots, n,$$

que se puede escribir como la ecuación matricial

$$\Gamma \alpha = \gamma$$

para $\Gamma = [\gamma(i-j)]_{i,j=1}^n \in \mathbb{R}^{n \times n}$, $\alpha = (\alpha_1, \dots, \alpha_n)' \in \mathbb{R}^n$, $\gamma = (\gamma(1), \dots, \gamma(n))' \in \mathbb{R}^n$. Si Γ es invertible entonces existe una única solución α , en el caso que sea no invertible hay infinitas soluciones para α , pero todas producen la misma aproximación \hat{X}_{n+1} de X_{n+1} por el teorema de proyección.

Procesos estacionarios ARMA

[clase 7, 29 de Agosto 2024]

Definiremos los procesos (o modelo) autoregresivos de medias móviles, abreviados ARMA (autorregressive moving average), que son un caso particular de procesos estacionarios que se pueden usar para modelos de predicción y control. Una limitante inmediata es la necesidad de estacionalidad, a los que muchas series de tiempo reales no se ajustan, sin embargo, constituyen un modelo clásico y muy usado de series de tiempo que es importante conocer. Para esto necesitamos definir los procesos de ruido blanco.

Definición 0.20 (Ruido blanco) *Un proceso estacionario $\{Z_t\}$ se dice ruido blanco si $EZ_t = 0$ y su función de autocovarianza satisface*

$$\gamma(h) := Cov(Z_{t+h}, Z_t) = \begin{cases} \sigma^2 & \text{si } h = 0, \\ 0 & \text{si } h \neq 0, \end{cases}$$

donde σ^2 es finito.

En simulaciones es usual modelar los procesos de ruido blanco como procesos iid de distribuciones normales con esperanza igual a cero y varianza finita σ^2 .

Observación 2 Para abreviar que un proceso $\{Z_t\}$ es ruido blanco, se usa la notación $\{Z_t\} \sim RB(0, \sigma^2)$. Para abreviar que un proceso $\{Z_t\}$ es de variables aleatorias iid de esperanza 0 y varianza σ^2 , se usa la notación $\{Z_t\} \sim IID(0, \sigma^2)$. Notar que si $\{Z_t\} \sim IID(0, \sigma^2)$, entonces $\{Z_t\} \sim RB(0, \sigma^2)$.

Definición 0.21 (ARMA(p, q)) Un proceso $\{X_t, t \in \mathbb{Z}\}$ se dice ARMA(p, q), donde $p, q \in \mathbb{N} \cup \{0\}$, si $\{X_t\}$ es estacionario y satisface la ecuación

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}, t \in \mathbb{Z} \quad (2)$$

donde $\{Z_t\}$ es un proceso de ruido blanco.

La ecuación (2) usualmente se escribe usando operadores de la siguiente manera. Definimos el polinomio autoregresivo

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p,$$

el polinomio de media móvil

$$\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q,$$

y el operador B tal que $BX_t = X_{t-1}$. Luego el operador B^j para $j \in \mathbb{N} \cup \{0\}$ satisface $B^j X_t = X_{t-j}$, donde definimos B^0 tal que $B^0 X_t = X_t$. Luego la ecuación (2) se reescribe por

$$\phi(B)X_t = \theta(B)Z_t, t \in \mathbb{Z}. \quad (3)$$

A continuación veremos dos casos particulares de modelos ARMA, aquellos en que $\phi(z) = 1$, llamados procesos de media móvil MA(q), y aquellos en que $\theta(z) = 1$, llamados procesos autoregresivos AR(p).

Ejemplo 16 (Modelo MA(q)) Sea $\{X_t\}$ un proceso ARMA($0, q$) donde $\phi(z) = 1$. Este proceso satisface la ecuación

$$X_t = \theta(B)Z_t.$$

En este caso la estacionalidad es automática de la ecuación anterior, ya que por linealidad de la esperanza $EX_t = 0$, además $EX_t^2 = \sigma^2(1 + \sum_{i=1}^p \theta_i^2) < \infty$, y la función de autocovarianza satisface

$$\gamma(h) = \text{Cov}(X_{t+h}, X_t) = \begin{cases} \sigma^2 \sum_{j=0}^{q-|h|} \theta_j \theta_{j+|h|} & \text{si } |h| \leq q, \\ 0 & \text{si } |h| > q. \end{cases}$$

Ejemplo 17 (Modelo AR(p)) Sea $\{X_t\}$ un proceso ARMA(p, 0) donde $\theta(z) = 1$. En este caso la ecuación que satisface el proceso

$$\phi(B)X_t = Z_t.$$

Aquí no es directo que la ecuación tenga una única solución y tampoco que la solución sea estacionaria.

Ejercicio 3 Demuestre que no existe una solución estacionaria de la ecuación

$$X_t = \pm X_{t-1} + Z_t, t \in \mathbb{Z}.$$

Estudiamos el caso particular en que $\phi(z) = 1 - \phi_1 z$. En este caso la ecuación (2) viene dada por

$$X_t - \phi_1 X_{t-1} = Z_t,$$

i.e. $X_t = Z_t + \phi_1 X_{t-1}$, e iterando,

$$X_t = Z_t + \phi_1 Z_{t-1} + \cdots + \phi_1^k Z_{t-k} + \phi_1^{k+1} Z_{t-k-1}.$$

Consideremos los casos $|\phi_1| < 1$ y $|\phi_1| > 1$. El caso $|\phi_1| = 1$ no tiene solución estacionaria, hacer ejercicio.

- Caso $|\phi_1| < 1$. Observar que si $\{X_t\}$ es estacionario, entonces $\gamma(0) = \text{Cov}(X_t, X_t) = \text{Var}(X_t) = EX_t^2 + (EX_t)^2$ es independiente de t , y $(EX_t)^2 = m^2$ es constante y $EX_t^2 < \infty$, entonces EX_t^2 es constante, i.e., existe $\tau^2 \in \mathbb{R}_{\geq 0}$ tal que $EX_t^2 = \tau^2$. Luego para $\|\cdot\| = \|\cdot\|_{L^2}$, se tiene que

$$\|X_t - \sum_{j=0}^k \phi_1^j Z_{t-j}\|^2 = \phi_1^{2k+2} \|X_{t-k-1}\|^2 = \phi_1^{2k+2} \tau^2 \rightarrow 0, \text{ cuando } k \rightarrow \infty.$$

Como $\sum_{j=0}^{\infty} \phi_1^j Z_{t-j}$ converge en L^2 , entonces $X_t = \sum_{j=0}^{\infty} \phi_1^j Z_{t-j}$ en L^2 . Esta solución es estacionaria, pues $EX_t = \sum_{j=0}^{\infty} \phi_1^j EZ_{t-j} = 0$ y la función de autocovarianza satisface $\gamma(h) = \frac{\sigma^2 \phi_1^{|h|}}{1 - \phi_1^2}$, además como satisface la ecuación (2), es la única solución.

- Caso $|\phi_1| > 1$. En este caso la serie $\sum_{j=0}^{\infty} \phi_1^j Z_{t-j}$ no converge en L^2 , pero se tiene que X_t satisface la ecuación

$$X_t = -\phi_1^{-1} Z_{t+1} + \phi_1^{-1} X_{t+1},$$

luego, iterando, obtenemos que

$$X_t = -\phi_1^{-1} Z_{t+1} + \cdots + \phi_1^{-k-1} Z_{t+k+1} + \phi_1^{-k-1} X_{t+k+1},$$

entonces

$$X_t = -\sum_{j=1}^{\infty} \phi_1^{-j} Z_{t+j}$$

es la única solución. Sin embargo, es poco natural que la serie este correlacionada con $\{Z_s, s > t\}$, por esta razón se tiende a restringir $|\phi_1| < 1$. En este caso el proceso se llama causal o independiente del futuro.

Procesos ARMA causales

[clase 8, 3 de Septiembre 2024]

Motivados por el ejemplo de la clase pasada, en que el que hay tres casos para el proceso que resuelve la ecuación

$$X_t = \phi_1 X_{t-1} + Z_t, t \in \mathbb{Z},$$

con $\{Z_t\} \sim RB(0, \sigma^2)$, dependiendo del valor de $\phi_1 \in \mathbb{C}$. En el caso que $|\phi_1| < 1$, se tiene una solución estacionaria $\{X_t\}$, donde X_t depende de $\{Z_s : s \leq t\}$. En el caso $|\phi_1| = 1$ no existe solución estacionaria, por lo tanto la solución no es un proceso ARMA. Y en el caso $|\phi_1| > 1$, se tiene una solución estacionaria $\{X_t\}$, donde X_t depende de $\{Z_s : s > t\}$. El primer caso es deseable para modelar usando un proceso ARMA, el segundo no lo permite, y el tercero es indeseable, por ser poco natural. El objetivo de esta clase es definir una familia más general de procesos, conocida como procesos causales, donde la ecuación pueda resolverse obteniendo como solución un proceso estacionario $\{X_t\}$, donde X_t depende de $\{Z_s : s \leq t\}$, así como demostrar un criterio que permite, a partir de las propiedades algebraicas de los polinomios ϕ y θ definir procesos causales.

Definición 0.22 (Procesos ARMA causales) *Un proceso ARMA(p, q) definido por las ecuaciones*

$$\phi(B)X_t = \theta(B)Z_t, t \in \mathbb{Z},$$

se dice causal si existe una secuencia de constantes $\{\psi_j\}$ tal que

$$\sum_{j=0}^{\infty} |\psi_j| < \infty$$

y

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, t \in \mathbb{Z}.$$

Antes de caracterizar la causalidad, veremos algunas propiedades útiles, que permiten demostrar propiedades de operadores sobre X_t usando propiedades de convergencia absoluta de series en los complejos.

Proposición 12 *Si $\{X_t\}$ es un proceso estocástico tal que $\sup_t E|X_t| < \infty$, y $\{\psi_j : j \in \mathbb{Z}\} \subset \mathbb{C}$ tal que $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$. Entonces*

$$\psi(B)X_t := \sum_{j=-\infty}^{\infty} \psi_j B^j X_t := \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} \quad (4)$$

converge absolutamente con probabilidad 1. Si además, $\sup_t E|X_t|^2 < \infty$, entonces converge en L^2 , y más aún, al mismo límite.

Dem. Para demostrar que converge absolutamente con probabilidad uno, es necesario demostrar que

$$E \left[\sum_{j=-\infty}^{\infty} |\psi_j| |X_{t-j}| \right] < \infty.$$

Usando el teorema de convergencia monotonía se tiene

$$E \left[\sum_{j=-\infty}^{\infty} |\psi_j| |X_{t-j}| \right] = \lim_{n \rightarrow \infty} E \left[\sum_{j=-n}^n |\psi_j| |X_{t-j}| \right].$$

Usando las condiciones $\sup_t E|X_t| < \infty$ y $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ se concluye el resultado, ya que

$$\lim_{n \rightarrow \infty} E \left[\sum_{j=-n}^n |\psi_j| |X_{t-j}| \right] \leq \lim_{n \rightarrow \infty} \left(\sup_t E|X_t| \right) \left(\sum_{j=-n}^n |\psi_j| \right) = \left(\sup_t E|X_t| \right) \left(\sum_{j=-\infty}^{\infty} |\psi_j| \right).$$

Lo que demuestra que con probabilidad uno $\sum_{j=-\infty}^{\infty} |\psi_j| |X_{t-j}|$ y $\psi(B)X_t$ son finitos.

Demostrar la convergencia en L^2 , también conocida como convergencia en media cuadrática, corresponde a demostrar que existe $S \in L^2$, llamado límite cuadrático medio o límite en L^2 , tal que

$$E \left| \sum_{j=-n}^n \psi_j X_{t-j} - S \right|^2 \rightarrow 0 \text{ cuando } n \rightarrow \infty.$$

Usando el criterio de Cauchy, basta demostrar que para $n > m > 0$

$$E \left| \sum_{m < |j| \leq n} \psi_j X_{t-j} \right|^2 \rightarrow 0 \text{ cuando } m, n \rightarrow \infty. \quad (5)$$

Rescribiendo $E \left| \sum_{m < |j| \leq n} \psi_j X_{t-j} \right|^2$ obtenemos

$$E \left| \sum_{m < |j| \leq n} \psi_j X_{t-j} \right|^2 = \sum_{m < |j| \leq n} \sum_{m < |k| \leq n} \psi_j \bar{\psi}_k E[X_{t-j} \bar{X}_{t-k}] \leq \sup_t E|X_t|^2 \left(\sum_{m < |j| \leq n} |\psi_j| \right)^2.$$

Luego, asumiendo que $\sup_t E|X_t|^2 < \infty$ y usando que $\sum_{j \in \mathbb{Z}} |\psi_j| < \infty$, se concluye (5). Para terminar la demostración, es necesario demostrar que

$$S$$

y $\psi(B)X_t$ son iguales con probabilidad uno. Para esto es suficiente demostrar que

$$E|S - \psi(B)X_t|^2 = 0.$$

Para esto, notar que

$$E|S - \psi(B)X_t|^2 = E \liminf_{n \rightarrow \infty} \left| S - \sum_{j=-n}^n \psi_j X_{t-j} \right|^2.$$

Usando el lemma de Fatou se obtiene que

$$E \liminf_{n \rightarrow \infty} |S - \sum_{j=-n}^n \psi_j X_{t-j}|^2 \leq \liminf_{n \rightarrow \infty} E |S - \sum_{j=-n}^n \psi_j X_{t-j}|^2 = 0.$$

lo que concluye la demostración. \square

Proposición 13 *Si $\{X_t\}$ es un proceso estacionario con función de autocovarianza γ y $\{\psi_j : j \in \mathbb{Z}\} \subset \mathbb{C}$ tal que $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$. Entonces*

$$\psi(B)X_t = \sum_{j=-\infty}^{\infty} \psi_j B^j X_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} \quad (6)$$

converge absolutamente con probabilidad 1 y en L^2 al mismo límite. Si $\{Y_t\}$ es el proceso que resuelve la ecuación

$$Y_t = \psi(B)X_t, t \in \mathbb{Z},$$

entonces $\{Y_t\}$ es estacionario con función de autocovarianza

$$\gamma_Y = \sum_{i,k=-\infty}^{\infty} \psi_j \psi_k \gamma(h - j + k).$$

Dem. Si $\{X_t\}$ es estacionario, entonces

$$E|X_t| \leq (E|X_t|^2)^{1/2} = C < \infty,$$

para C una constante independiente de t . Luego la convergencia queda demostrada por Propiedad 12. Para demostrar que $\{Y_t\}$ es estacionario, usar la convergencia en L^2 de Propiedad 12 y continuidad del producto interno en L^2 para concluir que

$$EY_t = \lim_{n \rightarrow \infty} \sum_{j=-n}^n \psi_j EX_{t-j} = \left(\sum_{j=-\infty}^{\infty} \psi_j \right) EX_t,$$

y

$$\begin{aligned} E(Y_{t+h}Y_t) &= \lim_{n \rightarrow \infty} E \left[\left(\sum_{j=-n}^n \psi_j X_{t+h-j} \right) \left(\sum_{k=-n}^n \psi_k X_{t-k} \right) \right] \\ &= \sum_{j,k=-\infty}^{\infty} \psi_j \psi_k (\gamma(h - j + k) + (EX_t)^2), \end{aligned}$$

lo que demuestra que EY_t y $E(Y_{t+h}Y_t)$ son ambas finitas e independientes de t , por lo tanto es un proceso estacionario. Finalmente, la función de autocovarianza de $\{Y_t\}$ viene dada por

$$\gamma_Y(h) = E(Y_{t+h}Y_t) - E(Y_{t+h})E(Y_t) = \sum_{j,k=-\infty}^{\infty} \psi_j\psi_k\gamma(h-j+k).$$

□

Observación 3 *Los operadores*

$$\psi(B) = \sum_{j=-\infty}^{\infty} \psi_j B^j \quad (7)$$

con $\{\psi_j : j \in \mathbb{Z}\} \subset \mathbb{C}$ tal que $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, aplicados a procesos estacionarios, heredan las propiedades algebraicas de las series de potencia. Veamos un ejemplo. Si $\{\alpha_j : j \in \mathbb{Z}\} \subset \mathbb{C}$ tal que $\sum_{j=-\infty}^{\infty} |\alpha_j| < \infty$, y $\{\beta_j : j \in \mathbb{Z}\} \subset \mathbb{C}$ tal que $\sum_{j=-\infty}^{\infty} |\beta_j| < \infty$, entonces $\alpha(B)\beta(B)X_t$ está bien definido y

$$\alpha(B)\beta(B)X_t = \beta(B)\alpha(B)X_t = \psi(B)X_t,$$

donde

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

para

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

Finalmente, veamos un teorema que permite caracterizar la causalidad.

Teorema 7 (Caracterización de causalidad) *Sea $\{X_t\}$ un proceso ARMA(p, q) que resuelve la ecuación*

$$\phi(B)X_t = \theta(B)Z_t,$$

para $\{Z_t\} \sim RB(0, \sigma^2)$, tal que los polinomios ϕ y θ no tienen ceros en común en \mathbb{C} . Entonces $\{X_t\}$ es causal ssi ϕ no tiene ceros en $\{z \in \mathbb{C} : |z| \leq 1\}$. Más aún, los coeficientes $\{\psi_j\}$ quedan determinados por

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, |z| \leq 1. \quad (8)$$

Dem. Asumiendo que $\phi(z) \neq 0$ si $|z| \leq 1$, se tiene que existe $\epsilon > 0$ tal que $\frac{1}{\phi(z)}$ tiene expansión en serie de potencias

$$\frac{1}{\phi(z)} = \sum_{j=0}^{\infty} \xi_j z^j = \xi(z), \quad |z| < 1 + \epsilon.$$

Luego $\xi_j(1 + \epsilon/2)^j \rightarrow 0$ cuando $j \rightarrow \infty$, y por lo tanto existe $K \in (0, \infty)$ tal que $|\xi_j| < K(1 + \epsilon/2)^{-j}$ para $j = 0, 1, \dots$. Por consiguiente, se concluye que $\sum_{j=0}^{\infty} |\xi_j| < \infty$ y que $\xi(z)\phi(z) \equiv 1$ en $|z| \leq 1$. Luego usando Proposición 13, se aplica el operador $\xi(B)$ a ambos lados de la ecuación $\phi(B)X_t = \theta(B)Z_t$, obteniendo

$$X_t = \xi(B)\theta(B)Z_t,$$

lo que demuestra que X_t tiene la representación

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j},$$

donde la secuencia $\{\psi_j\}$ queda únicamente determinada por (8). Para la otra implicancia, asumiendo que $\{X_t\}$ es causal, \square

Observación 3 *Más adelante veremos como calcular numericamente los coeficientes $\{\psi_j\}$.*

Procesos ARMA invertibles

[clase 9, 5 de Septiembre 2024]

Primero definiremos el concepto de procesos ARMA invertibles, que es análogo al de causalidad, intercambiando los roles de $\{X_t\}$ y $\{Z_t\}$, i.e., en este caso estaremos interesados en poder escribir Z_t en función de $\{X_s : s \leq t\}$.

Definición 0.23 (Procesos ARMA invertibles) *Un proceso ARMA(p, q) definido por las ecuaciones*

$$\phi(B)X_t = \theta(B)Z_t, \quad t \in \mathbb{Z},$$

se dice invertible si existe una secuencia de constantes $\{\pi_j\}$ tal que

$$\sum_{j=0}^{\infty} |\pi_j| < \infty$$

y

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, t \in \mathbb{Z}.$$

Teorema 8 (Caracterización de invertibilidad) Sea $\{X_t\}$ un proceso ARMA(p, q) que resuelve la ecuación

$$\phi(B)X_t = \theta(B)Z_t,$$

para $\{Z_t\} \sim RB(0, \sigma^2)$, tal que los polinomios ϕ y θ no tienen ceros en común en \mathbb{C} . Entonces $\{X_t\}$ es invertible ssi θ no tiene ceros en $\{z \in \mathbb{C} : |z| \leq 1\}$. Más aún, los coeficientes $\{\pi_j\}$ quedan determinados por

$$\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)}, |z| \leq 1.$$

Dem. Análoga a la demostración del teorema de caracterización de causalidad. \square

Observación 4 De los dos teoremas de caracterización de causalidad e invertibilidad se deduce que las soluciones estacionarias de la ecuación $\phi(B)X_t = \theta(B)Z_t, t \in \mathbb{Z}$ para $\{Z_t\} \sim RB(0, \sigma^2)$, y para ϕ, θ polinomios tales que $\phi(z)\theta(z)$ no tiene ceros en $\{z \in \mathbb{C} : |z| \leq 1\}$, entonces el proceso es causal e invertible, donde los coeficientes de la causalidad e invertibilidad quedan determinados por los coeficientes de las series de potencias de $\frac{\theta(z)}{\phi(z)}$ y $\frac{\phi(z)}{\theta(z)}$ para $|z| \leq 1$, respectivamente.

En general estudiaremos procesos ARMA causales e invertibles, cuando causalidad e invertibilidad no se asumen, tenemos el siguiente teorema.

Teorema 9 Si $\phi(z) \neq 0$ para todo $z \in \mathbb{C}$ tal que $|z| = 1$, entonces

$$\phi(B)X_t = \theta(B)Z_t, t \in \mathbb{Z},$$

tiene la solución única estacionaria,

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \quad (9)$$

donde los coeficientes quedan determinados por

$$\theta(z)\phi(z)^{-1} = \sum_{j=-\infty}^{\infty} \psi_j z^j = \psi(z), r^{-1} < |z| < r,$$

para $r > 1$.

Dem. El proceso $\{X_t\}$ definido por (9) es estacionario por Teorema 8. Luego se puede aplicar a ambos lados de la ecuación (9) el operador $\phi(B)$, y denuevo por Teorema 8 se obtiene que $\phi(B)\psi(B)Z_t = \theta(B)Z_t$, y por lo tanto $\phi(B)X_t = \theta(B)Z_t$. Luego $\{X_t\}$ es una solución estacionari de la ecuación ARMA. Para probar la implicancia en el otro sentido, sea $\{X_t\}$ cualquier solución estacionaria de la ecuación ARMA $\phi(B)X_t = \theta(B)Z_t$. Ta que $\phi(z) \neq 0$ para todo $z \in \mathcal{C}$ tal que $|z| = 1$, entonces existe $\delta > 1$ tal que la serie de potencias

$$\sum_{j=-\infty}^{\infty} \xi_j z^j = \phi(z)^{-1} = \xi(z)$$

converge absolutamente para $\delta^{-1} < |z| < \delta$. Luego, aplicando el operador $\xi(B)$ a ambos lados de la ecuación ARMA, se obtiene

$$\xi(B)\phi(B)X_t = \xi(B)\theta(B)Z_t,$$

i.e. $X_t = \psi(B)Z_t$, lo que concluye la demostración. \square

Generalizacion de procesos ARMA causales y su función de autocovarianza

[clase 10, 10 de Septiembre 2024]

Generalizaremos los precesos ARMA causales y veremos un teorema que permite obtener la función de autocovarianza en terminos de los coeficientes $\{\psi_j\}$ de causalidad, recordar definición 0.22.

Definición 0.24 *Un proceso $\{X_t\}$ se dice $MA(\infty)$ si*

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, t \in \mathbb{Z}, \{Z_t\} \sim RB(0, \sigma^2) \quad (10)$$

y $\{\psi_j\} \subset \mathbb{R}$ es tal que $\sum_{j=0}^{\infty} |\psi_j| < \infty$.

Ejemplo 18 $MA(q)$, $AR(1)$ con $|\phi_1| < 1$, y procesos $ARMA(p, q)$ causales, son todos procesos $MA(\infty)$.

La siguiente proposición entrega una herramienta rápida de usar para construir procesos $MA(q)$. La demostración, sin embargo, es un poco técnica, y no la vimos en clases.

Proposición 14 Sea $\{X_t\}$ un proceso estacionario con $E(X_t) = 0$ y función de autocovarianza $\gamma(h) = 0$ para $|h| > q$ y $\gamma(h) \neq 0$. Entonces $\{X_t\}$ es $MA(q)$.

Teorema 10 (Función de autocovarianza de procesos $MA(\infty)$) Si $\{X_t\}$ es un proceso $MA(\infty)$ definido por (10). Entonces es estacionario con $E(X_t) = 0$ y tiene función de autocovarianza

$$\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|} \quad (11)$$

Dem. Aplicación de Proposición 13. \square

Observación 4 La función de autocovarianza de un proceso $ARMA(p, q)$ causal queda completamente determinada por (11).

Cálculo de la función de autocovarianza de procesos ARMA

[clases 11 y 12, 12 y 24 de Septiembre 2024]

Usando como base el teorema 10, investigaremos 3 métodos para calcular la función de autocovarianza de procesos ARMA causales.

Método I o directo

Este método consiste en combinar el Teorema 7 que caracteriza la causalidad y permite obtener la ecuación

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, |z| \leq 1, \quad (12)$$

y el Teorema 10 que permite escribir la función de autocovarianza en términos de los coeficientes de causalidad $\{\psi_j\}$

$$\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|}, \quad (13)$$

estas dos fórmulas son muy útiles, y las utilizaremos varias veces en el curso.

De la ecuación (12) podemos reemplazar

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$$

y

$$\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q,$$

obteniendo

$$\psi(z) = \frac{\theta(z)}{\phi(z)} \Rightarrow \psi(z)\phi(z) = \theta(z) \Rightarrow (1 - \phi_1 z - \dots - \phi_p z^p) \sum_{j=0}^{\infty} \psi_j z^j = 1 + \theta_1 z + \dots + \theta_q z^q.$$

Igualando coeficientes obtenemos

$$\begin{aligned} \psi_0 &= 1 && \text{coeficiente de } z^0, \\ \psi_1 - \phi_1 \psi_0 &= \theta_1 && \text{coeficiente de } z^1, \\ \psi_2 - \phi_1 \psi_1 - \phi_2 \psi_0 &= \theta_2 && \text{coeficiente de } z^2, \\ &\vdots && \end{aligned} \tag{14}$$

Observar que las ecuaciones (14) corresponden a las ecuaciones para determinar los coeficientes de causalidad $\{\psi_j\}$ en función de los coeficientes de $\{\phi_i\}_{i=1}^p$ y $\{\theta_i\}_{i=1}^q$ para un proceso ARMA(p, q) causal. Estas ecuaciones se pueden escribir alternativamente como

$$\psi_j - \sum_{0 < k \leq j} \phi_k \psi_{j-k} = \theta_j, \text{ para } 0 \leq j < \text{máx}(p, q + 1) \tag{15}$$

y

$$\psi_j - \sum_{0 < k \leq p} \phi_k \psi_{j-k} = 0, \text{ para } j \geq \text{máx}(p, q + 1). \tag{16}$$

Una vez resueltas las ecuaciones (14) o alternativamente, las ecuaciones (15) y (16), uno obtiene los coeficientes de causalidad $\{\psi_j\}$, finalmente, uno usa la ecuación (13) para reemplazar los valores obtenidos y así calcular la función de autocovarianza γ . Más adelante en el curso, veremos que los coeficientes de causalidad pueden obtenerse automáticamente usando el software de estadística R y el paquete `astsa`. Sin embargo, con la fórmula recién obtenida, no es difícil programar una función que haga lo mismo en Python, por ejemplo. Por último, veamos un ejemplo de cómo este método puede usarse para encontrar la función de autocovarianza en un caso concreto.

Ejemplo 19 Calculemos la función de autocovarianza del proceso ARMA(2, 1) determinado por las ecuaciones

$$X_t - X_{t-1} + 0,25X_{t-2} = Z_t + Z_{t-1}, \{Z_t\} \sim RB(0, \sigma^2), t \in \mathbb{Z}.$$

En este caso, el proceso ARMA(2, 1) resuelve las ecuaciones $\phi(B)X_t = \theta(B)Z_t$ para $\phi(z) = 1 - \phi_1z - \psi_2z^2$ con $\psi_1 = 1, \psi_2 = -0,25$, y $\theta(z) = 1 + \theta_1z$ con $\theta_1 = 1, = .$ Luego, reescribiendo las ecuaciones (14), donde despejamos los coeficientes de causalidad $\{\psi_j\}$ en función de los coeficientes $\phi_1 = 1, \phi_2 = -0,25$ y $\theta_1 = 1$, obtenemos

$$\begin{aligned} \psi_0 &= 1, \\ \psi_1 &= \theta_1 + \psi_0\phi_1 = \theta_1 + \phi_1 = 2, \\ \psi_j - \psi_{j-1} + 0,25\psi_{j-2} &= 0, j \geq 2. \end{aligned}$$

Para resolver estas ecuaciones existe un método general que veremos más adelante. Por el momento, asumamos que obtenemos la solución

$$\psi_j = (1 + 3j)2^{-j}, j = 0, 1, \dots$$

Finalmente, usando la fórmula para escribir la función de autocovarianza en función de los coeficientes de causalidad (13), podemos calcular la función de autocovarianza

$$\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} (1 + 3j)(1 + 3(j+k))2^{-2j-k} = \sigma^2 2^{-k} \left(\frac{32}{3} + 8k \right).$$

Método II o de fórmula cerrada

Este método consiste en considerar el proceso ARMA(p, q) que resuelve las ecuaciones

$$\phi(B)X_t = \theta(B)Z_t,$$

multiplicar cada lado de la ecuación por X_{t-k} , y tomar esperanza, de modo de obtener las ecuaciones

$$E [(\phi(B)X_t) X_{t-k}] = E [(\theta(B)Z_t) X_{t-k}]. \quad (17)$$

Reemplazando con

$$\phi(z) = 1 - \phi_1z - \dots - \phi_pz^p$$

y

$$\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$$

se obtiene para el lado izquierdo de la ecuación (17), usando la linealidad de la esperanza y el hecho que en un proceso estacionario con $E[X_t] = 0$ se tiene que $E[X_t X_{t-h}] = E[X_t X_{t-h}] - E[X_t]E[X_{t-h}] = Cov(X_t, X_{t-h}) = \gamma(h)$, lo siguiente

$$\begin{aligned} & E[(\phi(B)X_t) X_{t-k}] \\ &= E[X_t X_{t-k} - \phi_1 X_{t-1} X_{t-k} - \phi_2 X_{t-2} X_{t-k} - \dots - \phi_p X_{t-p} X_{t-k}] \\ &= E[X_t X_{t-k}] - \phi_1 E[X_{t-1} X_{t-k}] - \phi_2 E[X_{t-2} X_{t-k}] - \dots - \phi_p E[X_{t-p} X_{t-k}] \\ &= \gamma(k) - \phi_1 \gamma(k-1) - \phi_2 \gamma(k-2) - \dots - \phi_p \gamma(k-p). \end{aligned}$$

Mientras que para el lado derecho de la ecuación (17), usando causalidad del proceso con respecto a los coeficientes $\{\psi_j\}$ y el hecho que $\{Z_t\} \sim RB(0, \sigma^2)$, se obtiene

$$\begin{aligned} & E[(\theta(B)Z_t) X_{t-k}] = \\ &= E[Z_t X_{t-k} + \theta_1 Z_{t-1} X_{t-k} + \theta_2 Z_{t-2} X_{t-k} + \dots + \theta_q Z_{t-q} X_{t-k}] \\ &= E[Z_t X_{t-k}] + \theta_1 E[Z_{t-1} X_{t-k}] + \theta_2 E[Z_{t-2} X_{t-k}] + \dots + \theta_q E[Z_{t-q} X_{t-k}] \\ &= E \left[Z_t \sum_{j=0}^{\infty} \psi_j Z_{t-k-j} \right] + \theta_1 E \left[Z_{t-1} \sum_{j=0}^{\infty} \psi_j Z_{t-k-j} \right] + \dots + \theta_q E \left[Z_{t-q} \sum_{j=0}^{\infty} \psi_j Z_{t-k-j} \right] \\ &= \sum_{j=0}^{\infty} \psi_j E[Z_t Z_{t-k-j}] + \theta_1 \sum_{j=0}^{\infty} \psi_j E[Z_{t-1} Z_{t-k-j}] + \dots + \theta_q \sum_{j=0}^{\infty} \psi_j E[Z_{t-q} Z_{t-k-j}] \\ &= \begin{cases} \sigma^2 \sum_{k \leq j \leq q} \theta_j \psi_{j-k} & \text{para } 0 \leq k < \max(p, q+1), \\ 0 & \text{para } k \geq \max(p, q+1). \end{cases} \end{aligned}$$

Volviendo a la ecuación (17), igualando el lado derecho con el lado izquierdo, hemos obtenido las siguientes dos ecuaciones

$$\gamma(k) - \phi_1 \gamma(k-1) - \phi_2 \gamma(k-2) - \dots - \phi_p \gamma(k-p) = \sigma^2 \sum_{k \leq j \leq q} \theta_j \psi_{j-k} \quad (18)$$

para $0 \leq k < \max(p, q+1)$, y las ecuaciones

$$\gamma(k) - \phi_1 \gamma(k-1) - \phi_2 \gamma(k-2) - \dots - \phi_p \gamma(k-p) = 0 \quad (19)$$

para $k \geq \max(p, q+1)$. Las ecuaciones (19) se pueden resolver en general, usando una técnica que veremos más adelante. Por el momento, veamos el siguiente ejemplo de como usar este método para encontrar la función de autocovarianza del ejemplo anterior.

Ejemplo 20 *Calculemos la función de autocovarianza del proceso ARMA(2, 1) determinado por las ecuaciones*

$$X_t - X_{t-1} + 0,25X_{t-2} = Z_t + Z_{t-1}, \{Z_t\} \sim RB(0, \sigma^2), t \in \mathbb{Z}.$$

En este caso, el proceso ARMA(2, 1) resuelve las ecuaciones $\phi(B)X_t = \theta(B)Z_t$ para $\phi(z) = 1 - \phi_1z - \psi_2z^2$ con $\psi_1 = 1, \psi_2 = -0,25$, y $\theta(z) = 1 + \theta_1z$ con $\theta_1 = 1, = .$ En este caso $\max(p, q+1) = \max(2, 1+1) = 2$, luego las ecuaciones (19) son

$$\gamma(k) - \gamma(k-1) + 0,25\gamma(k-2) = 0, k \geq 2.$$

Asumamos por el momento que tenemos un modo de demostrar que las solución general de esta última ecuación viene dada por

$$\gamma(k) = (\beta_1 + \beta_2k)2^{-k}, k \geq 0. \quad (20)$$

Por otra parte, las ecuaciones (19) en este caso se pueden escribir de la siguiente manera

$$\begin{aligned} \gamma(0) - \gamma(-1) - 0,25\gamma(-2) &= \sigma^2(\psi_0 + \psi_1) \\ \gamma(1) - \gamma(0) - 0,25\gamma(-1) &= \sigma^2\psi_0 \end{aligned}$$

Usando la paridad de γ , las últimas dos ecuaciones podemos escribirlas como

$$\begin{aligned} \gamma(0) - \gamma(1) - 0,25\gamma(2) &= \sigma^2(\psi_0 + \psi_1) \\ \gamma(1) - \gamma(0) - 0,25\gamma(1) &= \sigma^2\psi_0. \end{aligned}$$

Observando que $\psi_0 = 1$ y $\psi_1 = \theta_1 + \psi_1 = 2$, podemos reemplazar $\gamma(0), \gamma(1), \gamma(2)$ usando la formula general obtenida en (20) obtenemos el siguiente sistema de ecuaciones

$$\begin{aligned} 3\beta_1 - 2\beta_2 &= 16\sigma^2, \\ -3\beta_1 + 5\beta_2 &= 8\sigma^2. \end{aligned}$$

De donde obtenemos $\beta_1 = 32\sigma^2/3$ y $\beta_2 = 8\sigma^2$. Finalmente, obtenemos para γ la misma fórmula que en el ejemplo anterior, como tenía que ser.

Método III o numérico

Este método consiste en encontrar γ usando las ecuaciones (18) y (19) de la siguiente manera:

- Primero, encontrar $\gamma(0), \dots, \gamma(p)$ de las ecuaciones (18) y (19) combinadas, para $k = 0, 1, \dots, p$.
- Segundo, usar el resto de las ecuaciones para determinar recursivamente $\gamma(p+1), \gamma(p+2), \dots$.

Calculemos la función de autocovarianza del mismo ejemplo que hemos estado estudiando a lo largo de esta clase.

Ejemplo 21 *En este caso $p = 2$, y el primer paso es encontrar $\gamma(0), \gamma(1), \gamma(2)$. Para $k = 0, 1, 2$ las ecuaciones son :*

$$\begin{aligned}\gamma(0) - \gamma(1) + 0,25\gamma(2) &= 3\sigma^2, \\ \gamma(1) - \gamma(0) + 0,25\gamma(1) &= \sigma^2, \\ \gamma(2) - \gamma(1) + 0,25\gamma(0) &= 0.\end{aligned}$$

Resolviendo el sistema, uno obtiene $\gamma(0) = 32\sigma^2/3$, $\gamma(1) = 28\sigma^2/3$ y $\gamma(2) = 20\sigma^2/3$. Luego, el segundo paso es usar estas soluciones para recursivamente encontrar $\gamma(3), \gamma(4), \dots$. En este caso uno obtiene la fórmula recursiva

$$\gamma(k) = \gamma(k-1) - 0,25\gamma(k-2)$$

para $3, 4, \dots$. Así se pueden encontrar recursivamente los valores de $\gamma(3), \gamma(4), \dots$, y se obtiene el mismo resultado anterior (propuesto).

Solución general de una relación de recurrencia lineal

Los métodos arriba expuestos necesitan en algún momento resolver una relación de recurrencia lineal de la forma

$$h_t + \alpha_1 h_{t-1} + \dots + \alpha_k h_{t-k} = 0, t = n, n+1, \dots \quad (21)$$

donde $k \in \mathbb{N}$, $n \geq k$, $\{h_t\}$ es una secuencia que se desea encontrar (i.e. son las variables de la ecuación, para efectos de este curso pensar en que son los coeficientes de causalidad $\{\psi_j\}_{j=n}^{\infty}$ o que son los valores de una función de autocovarianza $\{\gamma(k)\}_{k=n}^{\infty}$ y recordar en que momento de los ejemplos del

método I y del método II se asumió conocida una solución general a una ecuación de este tipo), y α_j son coeficientes reales conocidos donde $\alpha_k \neq 0$.

Por el momento no hemos hecho una reflexión de como resolver estas ecuaciones en general. De hecho, para resolver estas ecuaciones existe una técnica de variable compleja, que en clase solo enunciamos, que se muestra a continuación.

Teorema 11 (Solución general de la recurrencia lineal) *Dados los coeficientes reales $\{\alpha_j\}_{j=1}^k$ con $\alpha_k \neq 0$. La solución general de la ecuación*

$$h_t + \alpha_1 h_{t-1} + \cdots + \alpha_k h_{t-k} = 0, \quad t = k, k+1, \dots$$

viene dada por

$$h_t = \sum_{i=1}^j \sum_{n=0}^{r_i-1} \beta_{i,n} t^n \xi_i^{-t},$$

donde ξ_i para $i = 1, \dots, j$, son los ceros distintos (en \mathbb{C}) de $\alpha(z) := 1 + \alpha_1 z + \cdots + \alpha_k z^k$, r_i es la multiplicidad de ξ_i , y $\{\beta_{i,n}\} \subset \mathbb{C}$ es un conjunto de coeficientes.

Las recurrencias lineales surgen de manera natural cuando uno trabaja con procesos ARMA, pero están presentes en diversas áreas de las matemáticas, un caso famoso es el de la secuencia de Fibonacci F_0, F_1, F_2, \dots que satisface la relación de recurrencia lineal $F_k = F_{k-1} + F_{k-2}, k \geq 2$ con condiciones iniciales $F_0 = F_1 = 1$. Con el teorema anterior uno puede encontrar una fórmula explícita para cada F_k , de hecho,

$$F_k = \frac{1}{\sqrt{5}} \left(\frac{1 + \sqrt{5}}{2} \right)^{k+1} - \frac{1}{\sqrt{5}} \left(\frac{1 - \sqrt{5}}{2} \right)^{k+1}.$$

Series de tiempo estacionarias complejas, distribución espectral y el teorema de Herglotz

[clase 13, 26 de Septiembre 2024]

Con el objetivo de poder introducir el concepto de distribución espectral, que es una herramienta útil para estudiar series de tiempo, extenderemos la definición de procesos estacionarios a procesos estacionarios complejos,

los que nos permitiran recuperar en particular los resultados anteriores para el caso real, y adicionalmente enriquecer la maquinaria para trabajar con series de tiempo, al introducir herramientas de series de potencia, que son útiles para estudiar operadores, y que permiten relacionar ceros de funciones complejas con decaimiento de correlaciones, por ejemplo. Otro concepto importante para estudiar series de tiempo, pero que no estará en estos apuntes (ni en las clases) es el de representación espectral, ya que necesita la definición de integración estocástica. Este concepto queda propuesto como tema de presentación para el final del curso.

Procesos estacionarios complejos

Definición 0.25 *Una serie de tiempo a valores complejos estacionaria (o un proceso estocástico a valores complejos estacionario) se define como un proceso estocástico a valores complejos $\{X_t : t \in \mathbb{Z}\}$ tal que $E(X_t)$ es independiente de t ,*

$$E|X_t|^2 := E(X_t \overline{X_t}) < \infty \forall t,$$

y $E(X_{t+h} \overline{X_t})$ es independiente de t .

Observar que las variables aleatorias complejas X en un espacio de probabilidad (Ω, \mathcal{F}, P) tal que $E|X|^2 < \infty$ forman un espacio de Hilbert con el producto interno $\langle X, Y \rangle := E(X \overline{Y})$.

Ahora podemos definir la función de autocovarianza de un proces estacionario complejo de manera análoga a como se hizo para el caso real, tal que mide como se de-correlacionan las variables aleatorias X_t y X_{t+h} .

Definición 0.26 *La función de autocovarianza de un proceso estacionario complejo $\{X_t\}$ se define por*

$$\gamma(h) := Cov(X_{t+h}, X_t) := E(X_{t+h} \overline{X_t}) - E(X_{t+h})E(\overline{X_t}).$$

La función de autocovarianza de un proceso estacionario complejo en principio se podría esperar que fuese una función a valores en \mathbb{C} , sin embargo es una función a valores reales, más aún se tiene la siguiente propiedad.

Proposición 15 *La función de autocovarianza γ de un proceso estacionario complejo satisface las siguientes propiedades.*

1. $\gamma(0) \geq 0$.
2. $|\gamma(h)| \leq \gamma(0)$.
3. $\gamma(h) = \overline{\gamma(-h)}$.

También obtenemos una caracterización de las funciones de autocovarianza análoga a la del caso real.

Teorema 12 *Una función $k : \mathbb{Z} \rightarrow \mathbb{C}$ es la función de autocovarianza de un proceso estacionario complejo ssi*

1. $k(h) = \overline{k(-h)}, \forall h \in \mathbb{Z}$, y
2. $\sum_{i,j=1}^n a_i k(i-j) \overline{a_j} \geq 0 \forall n \in \mathbb{N}, \forall a = (a_1, \dots, a_n)' \in \mathbb{C}^n$.

La primera condición en el teorema anterior se conoce como Hermiticidad de γ , y la segunda condición se conoce como γ es no-definida negativa.

Antes de proseguir con el teorema de Herglotz y su demostración, queremos motivar un concepto fundamental en tal teorema, que es el de distribución espectral.

Motivación de la distribución espectral

Consideremos el caso particular en que tenemos un proceso estocástico $\{X_t\}$ tal que

$$X_t = \sum_{j=1}^n Y_{\lambda_j} e^{it\lambda_j}, t \in \mathbb{Z},$$

donde $-\pi < \lambda_1 < \dots < \lambda_n = \pi$ y $Y_{\lambda_1}, \dots, Y_{\lambda_n}$ son variables aleatorias a valores complejos tales que $E[Y_{\lambda_j}] = 0$ y $E[Y_{\lambda_j} \overline{Y_{\lambda_j}}] = \sigma_j^2$ para todo j . En este caso $\{X_t\}$ es estacionario ya que $E[X_t] = 0$ para todo t y $E[X_{t+h} \overline{X_t}] = \sum_{j=1}^n \sigma_j^2 e^{it\lambda_j}$ no depende de t . Además, la función de autocovarianza γ satisface

$$\gamma(h) = \sum_{j=1}^n \sigma_j^2 e^{it\lambda_j} = \int_{-\pi}^{\pi} e^{ihv} dF(v),$$

donde la última igualdad es una integral de Riemann-Stieltjes con respecto a la función de distribución

$$F(v) = \sum_{j:\lambda_j \leq v} \sigma_j^2.$$

A esta función F se le llama distribución espectral del proceso $\{X_t\}$. A la ecuación

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ihv} dF(v)$$

se le conoce como representación espectral de γ . Además,

$$F(\lambda) = \sum_{\lambda_j \leq \lambda} E|Y_{\lambda_j}|^2.$$

En lo siguiente generalizaremos la representación espectral para un proceso estacionario a valores complejos en general. Este es esencialmente el contenido del teorema de Herglotz que demostraremos a continuación.

El teorema de Herglotz

Terminaremos la clase con el teorema de Herglotz y su demostración.

Teorema 13 (Teorema de Herglotz) *Una función $k : \mathbb{Z} \rightarrow \mathbb{C}$ es no-definida negativa ssi*

$$k(h) = \int_{-\pi}^{\pi} e^{ihv} dF(v), \forall h \in \mathbb{Z}, \quad (22)$$

donde F es continua por la derecha, no-decreciente, acotada en $[-\pi, \pi]$ y $F(-\pi) = 0$.

Dem. El implica hacia la izquierda se obtiene fácilmente a partir del siguiente cálculo. Sean $a_r \in \mathbb{C}$, $r = 1, \dots, n$, entonces

$$\begin{aligned} \sum_{r,s=1}^n a_r \gamma(r-s) \bar{a}_s &= \int_{-\pi}^{\pi} \sum_{r,s=1}^n a_r \bar{a}_s \exp[iv(r-s)] dF(v) \\ &= \int_{-\pi}^{\pi} \left| \sum_{r,s=1}^n a_r \exp[ivr] \right|^2 dF(v) \\ &\geq 0, \end{aligned}$$

lo que demuestra que γ es no-definida negativa.

Para demostrar el implica hacia la derecha, el plan es aproximar la distribución F , usando una sucesión de densidades f_N que sirven para construir

distribuciones F_N , y finalmente usar el teorema de Helly para concluir que esta sucesión tiene una subsección convergente a la distribución F .

Definimos f_N de la siguiente manera:

$$\begin{aligned} f_N(v) &:= \frac{1}{2\pi N} \sum_{r,s=1}^N e^{-irv} \gamma(r-s) e^{isv} \\ &= \frac{1}{2\pi N} \sum_{|m|<N} (N-|m|) e^{-imv} \gamma(m), \\ &\geq 0 \text{ para todo } v \in (-\pi, \pi], \text{ ya que } \gamma \text{ es no-def. neg.} \end{aligned}$$

Definimos F_N la distribución con densidad $f_N 1_{(-\pi, \pi]}$. Luego

$$F_N(\lambda) = \begin{cases} 0 & \text{si } \lambda \leq -\pi, \\ \int_{-\pi}^{\lambda} f_N(v) dv & \text{si } -\pi \leq \lambda \leq \pi, \\ F_N(\pi) & \text{si } \lambda \geq \pi. \end{cases}$$

Entonces para todo entero h ,

$$\begin{aligned} \int_{(\pi, \pi]} e^{ihv} dF_N(v) &= \frac{1}{2\pi} \sum_{|m|<N} \left(1 - \frac{|m|}{N}\right) \gamma(m) \int_{-\pi}^{\pi} e^{i(h-m)v} dv \\ &= \begin{cases} \left(1 - \frac{|h|}{N}\right) \gamma(h) & \text{si } |h| < N, \\ 0 & \text{si no.} \end{cases} \end{aligned} \tag{23}$$

Como $F_N(\pi) = \gamma(0) < \infty$ para todo N , entonces por Teorema de Helly, existe una distribución F que es límite de una subsección de $\{F_N\}$. Consideremos tal subsecuencia, y renombrémosla, abusando notación, por $\{F_N\}$. Luego para toda función continua g con $g(-\pi) = g(\pi)$ se tiene que

$$\int_{-\pi}^{\pi} g(v) dF_N(v) \rightarrow \int_{-\pi}^{\pi} g(v) dF(v), \text{ cuando } N \rightarrow \infty.$$

Recordando la segunda igualdad de la ecuación (23), y haciendo $N \rightarrow \infty$ obtenemos que

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ihv} dF(v),$$

lo que concluye la demostración. \square

A la función F del Teorema 13 se le llama distribución espectral. Y a la ecuación (22) se le conoce como representación espectral. Si es que existe una función f tal que $F(\lambda) = \int_{-\pi}^{\lambda} f(v)dv$, $-\pi \leq \lambda \leq \pi$, entonces f se llama densidad espectral. En general, tanto la distribución espectral como la densidad espectral, se dicen tanto distribución y densidad espectral de γ o distribución y densidad espectral de $\{X_t\}$, para $\{X_t\}$ un proceso con función de autocovarianza γ .

Ejercicio 4 Demuestre que si $0 < a < \pi$, la función

$$\gamma(h) = \begin{cases} \sin(ah)/h & \text{si } h \in \mathbb{Z} \setminus \{0\}, \\ a & \text{si } h = 0, \end{cases}$$

es la función de autocovarianza de un proceso estacionario y encuentre su densidad espectral.

Aplicaciones del teorema de Herglotz

[clase 14, 1 Octubre 2024]

En esta clase veremos como el teorema de Herglotz permite obtener criterios prácticos para determinar por ejemplo cuando una función $k : \mathbb{Z} \rightarrow \mathbb{C}$ es no definida negativa, lo que podremos usarlo para determinar si k puede o no ser la función de autocovarianza de un proceso estacionario.

Un primer corolario del Teorema 13 nos entrega una condición alternativa a la del Teorema 12 para determinar cuando una función $k : \mathbb{Z} \rightarrow \mathbb{C}$ es la función de autocovarianza de un proceso estacionario complejo.

Corolario 1 Una función $k : \mathbb{Z} \rightarrow \mathbb{C}$ es la función de autocovarianza de un proceso estacionario complejo ssi alguna de las condiciones siguientes se cumple:

1. $k(h) = \int_{-\pi}^{\pi} e^{ihv} dF(v), \forall h \in \mathbb{Z}$. Donde F es continua por la derecha, no-decreciente, acotada en $[-\pi, \pi]$ y $F(-\pi) = 0$.
2. $\sum_{i,j=1}^n a_i k(i-j) \bar{a}_j \geq 0 \forall n \in \mathbb{N}, \forall a = (a_1, \dots, a_n)' \in \mathbb{C}^n$.

Dem. El Teorema 13 dice que 1. y 2. son equivalente. De 1. sin embargo puede demostrarse que γ es Hermitica. Esto garantiza que γ satisface

las condiciones del Teorema 12, luego cualquiera de estas dos condiciones implica que γ es una función de autocovarianza de un proceso estacionario complejo. \square

En otras palabras el Corolario 1 se puede escribir de la siguiente manera.

Observación 5 (Manera alternativa de escribir Corolario 1) *Una función $k : \mathbb{Z} \rightarrow \mathbb{C}$ es la función de autocovarianza de un proceso estacionario complejo ssi k tiene representación espectral o si k es no-definida negativa.*

En lo siguiente veremos un criterio útil para determinar cuando una función $k : \mathbb{Z} \rightarrow \mathbb{C}$ es no definida negativa.

Teorema 14 *Si $k : \mathbb{Z} \rightarrow \mathbb{C}$ es tal que*

$$\sum_{n=-\infty}^{\infty} |k(n)| < \infty.$$

Entonces

$$k(h) = \int_{-\pi}^{\pi} e^{ihv} f(v) dv, \quad h \in \mathbb{Z},$$

donde

$$f(v) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{-inv} k(n).$$

Dem. La condición $\sum_{n=-\infty}^{\infty} |k(n)| < \infty$ garantiza que

$$\int_{-\pi}^{\pi} \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} |e^{i(h-n)v} K(n)| dv < \infty.$$

Luego, por teorema de Fubini

$$\begin{aligned} \int_{-\pi}^{\pi} e^{ihv} f(v) dv &= \int_{-\pi}^{\pi} \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{i(h-v)v} K(n) dv \\ &= \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} K(n) \int_{-\pi}^{\pi} e^{i(h-n)v} dv. \end{aligned}$$

Finalmente, observar que

$$\int_{-\pi}^{\pi} e^{i(h-n)v} dv = \begin{cases} 2\pi & \text{si } n = h, \\ 0 & \text{si no,} \end{cases}$$

lo que demuestra que $\int_{-\pi}^{\pi} e^{ihv} f(v) dv = K(h)$. \square

Como consecuencia obtenemos el siguiente corolario.

Corolario 2 Sea $k : \mathbb{Z} \rightarrow \mathbb{C}$ tal que $\sum_{n \in \mathbb{Z}} |k(n)| < \infty$. Entonces k es la función de autocovarianza de un proceso estacionario ssi

$$f(v) := \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{-inv} k(n) \geq 0, \forall \lambda \in [-\pi, \pi].$$

Además, en este caso f es la densidad espectral de k .

Dem. Primero demostramos el implica hacia la derecha. Sea k una función de autocovarianza. Entonces k es no-def. negativa y absolutamente sumable. Definamos

$$\begin{aligned} f_N(\lambda) &:= \frac{1}{2\pi N} \sum_{r,s=1}^N e^{-ir\lambda} k(r-s) e^{is\lambda} \\ &= \frac{1}{2\pi} \sum_{|m| < N} \left(1 - \frac{|m|}{N}\right) e^{-im\lambda} k(m). \end{aligned}$$

Luego $f_N \geq 0$ por definición, y además, tomando el límite cuando $N \rightarrow \infty$ se obtiene que puntualmente $f_N(\lambda) \rightarrow f(\lambda)$. Esto prueba que $f(\lambda) \geq 0$, para todo $-\pi \leq \lambda \leq \pi$. Usando Teorema obtenemos que $k(h) = \int_{-\pi}^{\pi} e^{ihv} f(v) dv$, $h \in \mathbb{Z}$. Luego f es la densidad espectral de k .

Segundo, demostramos el implica hacia la izquierda. Por Teorema , se tiene que $k(h) = \int_{-\pi}^{\pi} e^{ihv} f(v) dv$. Si $f(\lambda) \geq 0$, entonces $F(\lambda) = \int_{-\pi}^{\lambda} f(v) dv$ es continua por la derecha, no-decreciente, acotada en $[-\pi, \pi]$ y $F(-\pi) = 0$, y satisface

$$k(h) = \int_{\pi}^{-\pi} e^{ihv} dF(v), h \in \mathbb{Z}.$$

Luego por Corolario 1, se satisface la condición 2. por lo tanto k es la función de autocovarianza de un proceso estacionario complejo. \square

Los teoremas y corolarios anteriores aplican a procesos estacionarios complejos, y en particular, a procesos estacionarios reales. Con respecto a las densidades espectrales, en general, se tiene que la densidad espectral de un proceso estacionario complejo es una función cuyo dominio es $[-\pi, \pi]$ y que

toma valores reales, ya que es Hermitica. La siguiente observación, cuya demostración queda propuesta, nos entrega condiciones necesarias y suficientes para que la densidad espectral sea la de un proceso estacionario real.

Observación 6 f definido en $[-\pi, \pi]$ es la densidad espectral de un proceso estacionario real ssi $f(\lambda) = f(-\lambda)$, $f(\lambda) \geq 0$ y $\int_{-\pi}^{\pi} f(\lambda) d\lambda < \infty$.

Terminaremos la clase con la definición de filtro lineal invariante con respecto al tiempo. Estudiaremos la densidad espectral de estos procesos en la clase siguiente.

Definición 0.27 Dado un proceso estacionario $\{Y_t\}$ y una secuencia $\{\psi_j\}$ tal que $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$. Al proceso $\{X_t\}$ que se obtiene de resolver las ecuaciones

$$X_t = \phi(B)Y_t := \sum_{j=-\infty}^{\infty} \psi_j B^j Y_t := \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j}, t \in \mathbb{Z}$$

se le llama filtro lineal invariante con respecto al tiempo.

Densidad espectral de procesos ARMA

[clase 15, 3 Octubre 2024]

En esta clase demostraremos una fórmula que permite obtener la densidad espectral de un proceso ARMA, a partir de los polinomios autorregresivos y de media móvil que lo definen. El resultado principal de la clase no es más que una aplicación de los teoremas 11 y 13.

Primero veremos un teorema que permite obtener la densidad espectral de un filtro lineal invariante con respecto al tiempo.

Teorema 15 Sea $\{Y_t\}$ un procesos estacionario complejo tal que $E[Y_t] = 0$ y con distribución espectral F_Y . Sea $\{X_t\}$ el proceso que resuelve las ecuaciones

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j}, t \in \mathbb{Z},$$

para $\{\psi_j\}$ una secuencia tal que $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$. Entonces $\{X_t\}$ es un proceso estacionario con distribución espectral

$$F_x(\lambda) = \int_{-\pi}^{\lambda} \left| \sum_{j=-\infty}^{\infty} \psi_j e^{-ijv} \right|^2 dF_Y(v), -\pi \leq \lambda \leq \pi.$$

Dem. Los filtros lineales invariantes con respecto al tiempo ya los hemos estudiado, sin haberle puesto nombre, en Proposición 12. De hecho esta proposición prueba que $\{X_t\}$ es estacionario con media cero y función de autocovarianza

$$E(X_{t+h}\bar{X}_t) = \sum_{j,k=-\infty}^{\infty} \psi_j \bar{\psi}_k \gamma_Y(h-j+k), h \in \mathbb{Z}.$$

Luego, usando la representación espectral de γ_Y , obtenemos

$$\begin{aligned} \gamma_X(h) &= \sum_{j,k=-\infty}^{\infty} \psi_j \bar{\psi}_k \int_{-\pi}^{\pi} e^{i(h-j+k)v} dF_Y(v) \\ &= \int_{-\pi}^{\pi} \left(\sum_{j=-\infty}^{\infty} \psi_j e^{-ijv} \right) \left(\sum_{k=-\infty}^{\infty} \bar{\psi}_k e^{-ikv} \right) e^{ihv} dF_Y(v) \\ &= \int_{-\pi}^{\pi} e^{ihv} \left| \sum_{j=-\infty}^{\infty} \psi_j e^{-ijv} \right|^2 dF_Y(v), \end{aligned}$$

lo que demuestra que F_X es la distribución espectral de $\{X_t\}$. \square

Se tiene la siguiente observación para las densidades espectrales.

Observación 7 *En general no es cierto que un proceso tiene densidad espectral. Sin embargo, aplicando el teorema anterior, si $\{Y_t\}$ tiene densidad espectral, digamos f_Y , entonces el proceso filtrado $\{X_t\}$ también tiene densidad espectral, digamos f_X . Más aún,*

$$f_X(\lambda) = |\psi(e^{-i\lambda})|^2 f_Y(\lambda),$$

donde

$$\psi(e^{-i\lambda}) = \sum_{j=-\infty}^{\infty} \psi_j e^{-ij\lambda}.$$

Finalmente, podemos aplicar este teorema para encontrar la densidad espectral de un proceso ARMA, bajo ciertas hipótesis.

Teorema 16 *Sea $\{X_t\}$ un proceso ARMA(p, q) tal que*

$$\phi(B)X_t = \theta(B)Z_t, \{Z_t\} \sim RB(0, \sigma^2),$$

con $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$, $\theta(z) = 1 - \theta_1 z - \dots - \theta_q z^q$, tal que ϕ y θ no tienen ceros en común y ϕ no tiene ceros en $\{z \in \mathbb{C} : |z| \leq 1\}$. Entonces $\{X_t\}$ tiene densidad espectral

$$f_X(\lambda) = \frac{\sigma^2 |\theta(e^{-i\lambda})|^2}{2\pi |\phi(e^{-i\lambda})|^2}, \quad -\pi \leq \lambda \leq \pi.$$

Dem. Después de un cálculo es posible concluir que $\{Z_t\}$ tiene densidad espectral $\frac{\sigma^2}{2\pi}$. Recordando que bajo las hipótesis de este teorema se tiene que $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$ donde $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$. Luego, aplicando Teorema 15 se concluye que $\{X_t\}$ tiene densidad espectral. Luego, podemos escribir

$$U_t = \phi(B)X_t = \theta(B)Z_t$$

y aplicar Teorema 15 nuevamente para obtener que la densidad espectral de $\{U_t\}$ viene dada por

$$f_U(\lambda) = |\phi(e^{-i\lambda})|^2 f_X(\lambda) = |\theta(e^{-i\lambda})|^2 f_Z(\lambda). \quad (24)$$

Finalmente, ya que $\phi(e^{-i\lambda}) \neq 0$ para todo $\lambda \in [-\pi, \pi]$ se puede dividir en la ecuación (24) por $|\phi(e^{-i\lambda})|^2$ para obtener el resultado. \square

Técnicas para trabajar con datos

[clase 16, 15 Octubre 2024]

Desde este momento las clases estarán enfocadas en un aspecto más práctico de series de tiempo. El énfasis estará en identificar - estimar - validar modelos para modelar series de tiempo, y en su aplicación para predicción y control. Esto lo haremos usando R.

Para comenzar necesitaremos introducir algunos conceptos tales como la función de autocovarianza parcial, el operador de diferencias $\nabla X_t = X_t - X_{t-1}$, las funciones de autocovarianza de muestreo y la función de autocovarianza parcial de muestreo. Además, expondremos un método de predicción para modelos $ARMA(p, q)$, definiremos el modelo $ARIMA$, que es una extensión del modelo $ARMA$ para trabajar con datos con tendencia y dos herramientas útiles para disminuir la oscilación de los datos. Finalmente, y a modo de introducción, enunciaremos una serie de pasos que usualmente

es necesario seguir para, a partir de un conjunto de datos, escoger el modelo apropiado para poder hacer predicciones con los datos. El detalle de cada uno de estos pasos los veremos en las clases siguientes.

Escribir clase. Pendiente.

Predicción en procesos estacionarios

[clase 17, 17 Octubre 2024]

En esta clase nos centraremos en el problema de predicción en un proceso estacionario. Para los cálculos, supondremos que tenemos certeza que los datos provienen de un proceso estacionario donde los parametros necesarios para definir la función de autocovarianza del proceso son conocidos, i.e. vamos a suponer que tenemos datos $\{x_i\}_{i=1}^n$ que provienen de un proceso estacionario $\{X_i\}_{i \in \mathbb{Z}}$ con media cero, y donde $\gamma(h) := E(X_{t+h}X_t)$, $h \in \mathbb{Z}$ es conocida. Como vimos la clases anterior, si bien es cierto estas son hipótesis que en la realidad usualmente no son cierta, es decir, dado un conjunto de datos, al graficarlos, generalmente veremos una serie de tiempo que no proviene de un proceso estacionario (puede presentar por ejemplo una tendencia o periodicidad), y por otra parte, los datos de por si no nos entregan γ . Sin embargo, vimos métodos para transformar los datos originales de modo de obtener una muestra que pueda suponerse que proviene de un proceso estacionario. Tales como por ejemplo la suavización de los datos por log o la transformación de Box-Cox. Y vimos el modelo ARIMA que permite eliminar la tendencia aplicando reiteradas veces el operador ∇ . Más adelante, veremos un modelo SARIMA que permite adicionalmente eliminar la estacionalidad de los datos. Y por otra parte, puede usarse el γ de muestreo, en caso de no conoce γ . Por estas razones, si bien es cierto las hipótesis de estacionalidad y de conocimiento de γ son fuertes, no son siempreun impedimento para trabajar con datos que no los cumplan, ya que hay métodos para transformar la data, de modo que las hipótesis si se satisfagan, con cierto nivel de confianza.

En particular nos centraremos en el problema de encontrar el mejor estimador lineal.

Escribir clase. Pendiente.

Ajuste de parámetros en un modelo ARIMA

[clase 18, 22 Octubre 2024]

En esta clase veremos en detalle un método de ajuste de parámetros usando el método de los momentos. En el problema de ajustar los parámetros uno dispone de datos $\{x_t\}_{t=1}^n$, y uno hace el supuesto que sabe que estos datos provienen de una distribución conocida, de hecho para la clase de hoy asumiremos que provienen de proceso ARMA(p, q) donde p y q son conocidos. Esto será suficiente para nosotros, ya que el modelo más general que veremos son los modelos SARIMA, y en este caso lo que veamos hoy puede directamente extenderse a esos casos. Bajo el supuesto que provienen de este proceso ARMA(p, q), nuestro objetivo es estimar los parámetros del autorregresivo ϕ y del polinomio de media móvil θ . Este método, como veremos, consiste en escribir unas ecuaciones para los parámetros, estas ecuaciones serán llamadas de Yule-Walker, y para poder resolverlas se necesita conocer la función de autocovarianza γ del proceso que origina los datos. Como esto es en nuestro caso desconocido, lo que uno hace es $\gamma = \hat{\gamma}$, donde $\hat{\gamma}$ es la función de autocovarianza de muestreo. Con esto, seremos capaces de escribir matricialmente las ecuaciones para encontrar los parámetros del modelo, e involucrará invertir de una matriz de autocovarianzas de muestreo de $n \times n$, para esto podemos usar el algoritmo de Durbin-Levinson visto la clase anterior, y así evitar invertir la matriz. Hecho esto, obtendremos dos predictores, uno para los coeficientes $(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$ y el otro para la varianza σ^2 del ruido blanco del modelo, estos predictores recibirán el nombre de predictores de Yule-Walker. Los predictores de Yule-Walker para procesos SARIMA, que generalizan los que veremos hoy, están implementados en la librería 'ASTSA' de R, que usaremos en este curso. Hay otras maneras de ajustar los parámetros, tales como usar el criterio de máxima verosimilitud, que consiste en suponer que los parámetros vienen de una distribución conocida que depende de los parámetros a estimar $\beta = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma^2)'$, y encontrar usando algún método numérico, por ejemplo Gauss-Raphson, el parámetro vectorial $\hat{\beta}$ que maximizan la función de densidad conjunta de $\{x_t\}_{t=1}^n$ condicionada a β , llamada función de verosimilitud. Para esto, el primer paso es escribir la función de densidad conjunta usando el hecho de que el proceso es ARMA, y el segundo en utilizar algún algoritmo de cálculo numérico para aproximar el argmax de la función de verosimilitud. Este método no alcanzamos a verlo en detalle en el curso, está también implementado en la librería 'ASTSA' de R.

Escribir clase. Pendiente.

Procesos SARIMA y Diagnóstico

[clase 19, 24 Octubre 2024]

En esta clase veremos el modelo SARIMA, que sirve para eliminar tendencia y periodicidad de los datos. Es el modelo más general que veremos en este capítulo, y será nuestra herramienta más poderosa para hacer predicciones del estilo estimar x_{t+1} dado que conocemos $\{x_t\}_{t=1}^n$. Además, veremos de un modo intuitivo 3 gráficos, que nos sirvan para evaluar la calidad de un modelo. Esto se conoce como diagnóstico, y sirve para, dado un conjunto de distintos modelos que uno construye que se ajustan a los datos observados, poder evaluar la calidad de cada uno, de modo de quedarse con el mejor de ellos, y hacer predicciones usando el modelo escogido.

Escribir clase. Pendiente.

Modelo de Espacio de estados

[clase 20, 29 Octubre 2024]

Motivamos la formulación del modelo de espacio de estados de Kalman, y escribimos el postulado del filtro de Kalman.

Escribir clase. Pendiente.

Filtro de Kalman

[clase 21, 5 Noviembre 2024]

Demostremos la propiedad del filtro de Kalman.

Escribir clase. Pendiente.

Suavizador de Kalman

[clase 22, 7 Noviembre 2024]

Enunciamos y demostramos la propiedad del suavizador de Kalman.

Escribir clase. Pendiente.

Estimación de parámetros en modelo de Espacio de estados

[clase 23, 12 Noviembre 2024]

Introducimos el problema de estimación de parámetros del modelo de espacio de estados de Kalman. Mostramos dos algoritmos para estimar los parámetros: Algoritmo de Schweppe 1965, y algoritmo de Shumway-Stoffer 1982.