

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.

1. Proceso estocásticos [clase 1]

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 1.1 (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 1.2 (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 1.3 (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.

2. Recuerdos de Probabilidades [clase 2]

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 2.1 (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 2.2 (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 2.3 (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 2.4 (Varianza) *La varianza de una variable aleatoria real X viene dada por*

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

Definición 2.5 (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 2.6 (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 2.7 *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 2.8 (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

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

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

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

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.

3. Vectores aleatorios y normal multivariada [clase 3]

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 3.1 (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 3.2 (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

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

4. Motivaciones, Filtros lineales y series de tiempo estacionarias [clase 4]

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.

4.1. Motivaciones prácticas

Ejemplo 5 (Finanzas) *Para contener la inflación el banco central decide una tasa de política 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 maquinas 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? La compañías de luz seguramente tendran que hacer un estudio utilizando series de tiempo para poder disponibilizar de métodos para poder acortar los tiempos de reparaciones.

4.2. Filtros lineales

Definición 4.1 (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 observacion.

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.*

4.3. 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 4.2 (Función de autocovarianza) Si un proceso $\{X_t, t \in T\}$ tiene $Var(X_t) < \infty$ para cada $t \in T$, entonces la función de autocovarianza asociada $\gamma(\cdot, \cdot)$ se define como

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

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 4.3 (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 4.4 (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 distinta 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.

5. Propiedades de la función de autocovarianza de procesos estacionarios [clase 5]

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

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

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

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

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

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

3. $\gamma(-h) = Cov(X_{t-h}, X_t) = Cov(X_t, X_{t+h}) = \gamma(h)$. \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 5.1 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 $Cov(X_i, X_j) = k(i, j)$, lo que concluye la demostración. \square

6. Ejemplos procesos estacionarios [clase 6]

En esta clase veremos 3 para 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.

6.1. Ejemplos de procesos estacionarios y procesos no estacionarios

Ejemplo 12 Sea $\{Z_t\}$ iid con $EZ_t = 0, VarZ_t = \sigma^2$. 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 $Var(Z_t) = E(Z_t^2) - E(Z_t)^2 = E(Z_t^2) = \sigma^2$, luego

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

y finalmente, la función de autocovarianza viene dada por

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

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

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

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 6.1 (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.

6.2. 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 $X_{t+h} = \hat{m}_{t+h}$.

6.3. 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.

7. Procesos estacionarios ARMA [clase 7]

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

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

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 7.2 (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

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

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 $EX_t^2 = \sigma^2(1 + \sum_{i=1}^q \theta_i^2) < \infty$, y la función de autocovarianza satisface

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

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.