Análisis de Series de Tiempo

Camilo Enrique Argoty Pulido

Especialización en Inteligencia Artificial

13 de marzo de 2025

Laboratorio de Sistemas Embebidos

Introducción

Plan del curso

En este curso esperamos completar el siguiente contenido:

1. Conceptos básicos de procesos estocásticos y series de tiempo.
2. Tipos de dato datetime, procesamiento básico de datos, visualización de series de tiempo.
3. Ruido blanco y caminata aleatoria. Estacionaridad. Prueba de Dickey-Fuller. Filtro de Holt-Winters. Descomposición ETS. Suavizado exponencial. Criterios LLH, AIC y BIC para bondad de modelos.
4. Pronósticos.
5. Modelos autorregresivos y de media móvil: AR, MA, ARIMA.
6. Modelos integrados, estacionales y con variables externas: ARIMA, SARIMA, SARIMAX.
7. Modelos de heteroscedasticidad condicional: ARCH y GARCH.
8. Librerás autoarima de Python y Prophet de Facebook.
9. Redes neuronales LSTM para series de tiempo

Materiales

El material está disponible en el aula virtual del campus, excepto las grabaciones de las clases.

Materiales

El material está disponible en el aula virtual del campus, excepto las grabaciones de las clases.
También hay un repositorio de Github:
https://github.com/FIUBA-Posgrado-Inteligencia-Artificial/CEIA-AST.git

Materiales

El material está disponible en el aula virtual del campus, excepto las grabaciones de las clases.
También hay un repositorio de Github:
https://github.com/FIUBA-Posgrado-Inteligencia-Artificial/CEIA-AST.git

Bibliografía recomendada:

Materiales

El material está disponible en el aula virtual del campus, excepto las grabaciones de las clases.
También hay un repositorio de Github:
https://github.com/FIUBA-Posgrado-Inteligencia-Artificial/CEIA-AST.git

Bibliografía recomendada:

- Kitagawa G. *Introduction to Time Series Modeling*. Taylor and Francis Group, 2010.

Materiales

El material está disponible en el aula virtual del campus, excepto las grabaciones de las clases.
También hay un repositorio de Github:
https://github.com/FIUBA-Posgrado-Inteligencia-Artificial/CEIA-AST.git

Bibliografía recomendada:

- Kitagawa G. *Introduction to Time Series Modeling*. Taylor and Francis Group, 2010.
- Cryer J., Chan K-S. *Time Series Analysis with applications in R*. Springer Science+Business Media, 2010.

Materiales

El material está disponible en el aula virtual del campus, excepto las grabaciones de las clases.
También hay un repositorio de Github:
https://github.com/FIUBA-Posgrado-Inteligencia-Artificial/CEIA-AST.git

Bibliografía recomendada:

- Kitagawa G. *Introduction to Time Series Modeling*. Taylor and Francis Group, 2010.
- Cryer J., Chan K-S. *Time Series Analysis with applications in R*. Springer Science+Business Media, 2010.
- Montenegro A. *Análisis de Series de Tiempo*. Editorial Pontificia Universidad Javeriana, Bogotá, Colombia, 2010.

Materiales

El material está disponible en el aula virtual del campus, excepto las grabaciones de las clases.
También hay un repositorio de Github:
https://github.com/FIUBA-Posgrado-Inteligencia-Artificial/CEIA-AST.git

Bibliografía recomendada:

- Kitagawa G. *Introduction to Time Series Modeling*. Taylor and Francis Group, 2010.
- Cryer J., Chan K-S. *Time Series Analysis with applications in R*. Springer Science+Business Media, 2010.
- Montenegro A. *Análisis de Series de Tiempo*. Editorial Pontificia Universidad Javeriana, Bogotá, Colombia, 2010.
- Nualart D. *Cálculo estocástico*. Documento libre en la web.

Materiales

El material está disponible en el aula virtual del campus, excepto las grabaciones de las clases.
También hay un repositorio de Github:
https://github.com/FIUBA-Posgrado-Inteligencia-Artificial/CEIA-AST.git

Bibliografía recomendada:

- Kitagawa G. *Introduction to Time Series Modeling*. Taylor and Francis Group, 2010.
- Cryer J., Chan K-S. *Time Series Analysis with applications in R*. Springer Science+Business Media, 2010.
- Montenegro A. *Análisis de Series de Tiempo*. Editorial Pontificia Universidad Javeriana, Bogotá, Colombia, 2010.
- Nualart D. *Cálculo estocástico*. Documento libre en la web.
- Ratanov N. *Modelos estocásticos de mercados financieros*. Editorial Universidad del Rosario, Bogotá, Colombia, 2009.

Evaluación

La nota final es un promedio ponderado de las notas siguientes:

1. Ejercicios teóricos individuales con fecha de entrega a definir. Peso: 30 %

Evaluación

La nota final es un promedio ponderado de las notas siguientes:

1. Ejercicios teóricos individuales con fecha de entrega a definir. Peso: 30 %
2. Trabajo Final de análisis de una o varias series de tiempo en grupos de 3 integrantes. Peso: 70 %

Procesos estocásticos

Procesos estocásticos

Un **proceso estocástico** es una familia de variables aleatorias reales $\{X_t, \mid t \in T\}$ definidas sobre un espacio de probabilidad $(\Omega, \mathcal{F}, P)$, donde T puede ser un intervalo contenido en $\mathbb{N}_{\geq 0}$ o en $\mathbb{R}_{\geq 0}$.

Procesos estocásticos

Un **proceso estocástico** es una familia de variables aleatorias reales $\{X_t, \mid t \in T\}$ definidas sobre un espacio de probabilidad $(\Omega, \mathcal{F}, P)$, donde T puede ser un intervalo contenido en $\mathbb{N}_{\geq 0}$ o en $\mathbb{R}_{\geq 0}$.

Si $T \subseteq \mathbb{N}_{\geq 0}$, se dice que el proceso estocástico es **discreto**.

Procesos estocásticos

Un **proceso estocástico** es una familia de variables aleatorias reales $\{X_t, \mid t \in T\}$ definidas sobre un espacio de probabilidad $(\Omega, \mathcal{F}, P)$, donde T puede ser un intervalo contenido en $\mathbb{N}_{\geq 0}$ o en $\mathbb{R}_{\geq 0}$.

Si $T \subseteq \mathbb{N}_{\geq 0}$, se dice que el proceso estocástico es **discreto**.

De otro lado, si $T \subseteq \mathbb{R}_{\geq 0}$, se dice que el proceso estocástico es **continuo**.

Procesos estocásticos

Un **proceso estocástico** es una familia de variables aleatorias reales $\{X_t, \mid t \in T\}$ definidas sobre un espacio de probabilidad $(\Omega, \mathcal{F}, P)$, donde T puede ser un intervalo contenido en $\mathbb{N}_{\geq 0}$ o en $\mathbb{R}_{\geq 0}$.

Si $T \subseteq \mathbb{N}_{\geq 0}$, se dice que el proceso estocástico es **discreto**.

De otro lado, si $T \subseteq \mathbb{R}_{\geq 0}$, se dice que el proceso estocástico es **continuo**. Una

Serie de Tiempo es un proceso estocástico discreto donde T es finito.

Media, varianza y autocovarianza de un proceso estocástico

A partir de un proceso estocástico se pueden definir varias funciones:

Media, varianza y autocovarianza de un proceso estocástico

A partir de un proceso estocástico se pueden definir varias funciones:

Media Se define como:

Media, varianza y autocovarianza de un proceso estocástico

A partir de un proceso estocástico se pueden definir varias funciones:

Media Se define como:

$$m_X(t) = E[X_t]$$

Media, varianza y autocovarianza de un proceso estocástico

A partir de un proceso estocástico se pueden definir varias funciones:

Media Se define como:

$$m_X(t) = E[X_t]$$

Varianza Se define como:

Media, varianza y autocovarianza de un proceso estocástico

A partir de un proceso estocástico se pueden definir varias funciones:

Media Se define como:

$$m_X(t) = E[X_t]$$

Varianza Se define como:

$$\sigma_X^2(t) = Var(X_t)$$

Autocovarianza Se define como:

Media, varianza y autocovarianza de un proceso estocástico

A partir de un proceso estocástico se pueden definir varias funciones:

Media Se define como:

$$m_X(t) = E[X_t]$$

Varianza Se define como:

$$\sigma_X^2(t) = Var(X_t)$$

Autocovarianza Se define como:

$$\Gamma_X(s,t) = Cov(X_s, X_t) = E[(X_s - m_x(s))(X_t - m_x(t))]$$

Media, varianza y autocovarianza de un proceso estocástico

A partir de un proceso estocástico se pueden definir varias funciones:

Media Se define como:

$$m_X(t) = E[X_t]$$

Varianza Se define como:

$$\sigma_X^2(t) = Var(X_t)$$

Autocovarianza Se define como:

$$\Gamma_X(s,t) = Cov(X_s, X_t) = E[(X_s - m_x(s))(X_t - m_x(t))]$$

Es de notar que

$$\sigma_X^2(t) = \Gamma_x(t,t)$$

Ejemplo 1: Proceso estocástico trigonométrico

Sea el proceso estocástico:

Ejemplo 1: Proceso estocástico trigonométrico

Sea el proceso estocástico:

$$X_t = A\cos\left(\phi + \lambda t\right)$$

Ejemplo 1: Proceso estocástico trigonométrico

Sea el proceso estocástico:

$$X_t = A\cos\left(\phi + \lambda t\right)$$

donde A y ϕ son variables aleatorias independientes tales que $E[A^2] < \infty$ y ϕ se distribuye uniforme sobre $[0, 2\pi]$.

Ejemplo 2: Proceso gaussiano

Un proceso estocástico se dice **gaussiano** si X_t se distribuye normal para todo $t \in T$.

Ejemplo 2: Proceso gaussiano

Un proceso estocástico se dice **gaussiano** si X_t se distribuye normal para todo $t \in T$.

Sea $\{X_t,\ t \in [0,1]$ un proceso estocástico gaussiano tal que las variables X_t son independientes e identicamente distribuidas $\mathcal{N}(0,\sigma^2)$

Procesos de Poisson

Un **proceso de Poisson** $\{N_t \; t \geq 0\}$ es un proceso estocástico con las siguientes propiedades:

Procesos de Poisson

Un **proceso de Poisson** $\{N_t \; t \geq 0\}$ es un proceso estocástico con las siguientes propiedades:

1. $N_0 = 0$.

Procesos de Poisson

Un **proceso de Poisson** $\{N_t \; t \geq 0\}$ es un proceso estocástico con las siguientes propiedades:

1. $N_0 = 0$.
2. Si $s \leq t$, entonces $N_s \leq N_t$.

Procesos de Poisson

Un **proceso de Poisson** $\{N_t \; t \geq 0\}$ es un proceso estocástico con las siguientes propiedades:

1. $N_0 = 0$.
2. Si $s \leq t$, entonces $N_s \leq N_t$.
3. Dados n instantes de tiempo $0 < t_1 < \cdots < t_n$, las variables aleatorias N_{t_1}, $N_{t_2} - N_{t_1}, \cdots, N_{t_n} - N_{t_{n-1}}$ son independientes.

Procesos de Poisson

Un **proceso de Poisson** $\{N_t \; t \geq 0\}$ es un proceso estocástico con las siguientes propiedades:

1. $N_0 = 0$.
2. Si $s \leq t$, entonces $N_s \leq N_t$.
3. Dados n instantes de tiempo $0 < t_1 < \cdots < t_n$, las variables aleatorias N_{t_1}, $N_{t_2} - N_{t_1}$,$\cdots$, $N_{t_n} - N_{t_{n-1}}$ son independientes.
4. Si $s \leq t$, entonces el incremento $N_t - N_s$ se distribuye Poisson con parámetro $\lambda(t-s)$, es decir

$$P(N_t - N_s = k) = e^{-\lambda(t-s)} \frac{(\lambda(t-s))^k}{k!}$$

Procesos de Poisson

Un **proceso de Poisson** $\{N_t \; t \geq 0\}$ es un proceso estocástico con las siguientes propiedades:

1. $N_0 = 0$.
2. Si $s \leq t$, entonces $N_s \leq N_t$.
3. Dados n instantes de tiempo $0 < t_1 < \cdots < t_n$, las variables aleatorias N_{t_1}, $N_{t_2} - N_{t_1}, \cdots, N_{t_n} - N_{t_{n-1}}$ son independientes.
4. Si $s \leq t$, entonces el incremento $N_t - N_s$ se distribuye Poisson con parámetro $\lambda(t-s)$, es decir

   $$P(N_t - N_s = k) = e^{-\lambda(t-s)} \frac{(\lambda(t-s))^k}{k!}$$

   para $k = 0, \, 2, \, 3, \cdots$

Construcción de un proceso de Poisson

Sea $\{Y_n \mid n \geq 1\}$ una sucesión de variables aleatorias independientes, con distribución geométrica de parámetro λ, es decir,

Construcción de un proceso de Poisson

Sea $\{Y_n \mid n \geq 1\}$ una sucesión de variables aleatorias independientes, con distribución geométrica de parámetro λ, es decir,

$$P(Y_n \geq x) = \lambda e^{-\lambda x}$$

Construcción de un proceso de Poisson

Sea $\{Y_n \mid n \geq 1\}$ una sucesión de variables aleatorias independientes, con distribución geométrica de parámetro λ, es decir,

$$P(Y_n \geq x) = \lambda e^{-\lambda x}$$

Sea $T_0 = 0$ y dado $n \geq 1$,

Construcción de un proceso de Poisson

Sea $\{Y_n \mid n \geq 1\}$ una sucesión de variables aleatorias independientes, con distribución geométrica de parámetro λ, es decir,

$$P(Y_n \geq x) = \lambda e^{-\lambda x}$$

Sea $T_0 = 0$ y dado $n \geq 1$,

$$T_n = Y_1 + \cdots + Y_n$$

Construcción de un proceso de Poisson

Sea $\{Y_n \mid n \geq 1\}$ una sucesión de variables aleatorias independientes, con distribución geométrica de parámetro λ, es decir,

$$P(Y_n \geq x) = \lambda e^{-\lambda x}$$

Sea $T_0 = 0$ y dado $n \geq 1$,

$$T_n = Y_1 + \cdots + Y_n$$

Entonces, el proceso definido por:

Construcción de un proceso de Poisson

Sea $\{Y_n \mid n \geq 1\}$ una sucesión de variables aleatorias independientes, con distribución geométrica de parámetro λ, es decir,

$$P(Y_n \geq x) = \lambda e^{-\lambda x}$$

Sea $T_0 = 0$ y dado $n \geq 1$,

$$T_n = Y_1 + \cdots + Y_n$$

Entonces, el proceso definido por:

$$N_t = n \text{ si } T_n \leq t < T_{n+1}$$

Construcción de un proceso de Poisson

Sea $\{Y_n \mid n \geq 1\}$ una sucesión de variables aleatorias independientes, con distribución geométrica de parámetro λ, es decir,

$$P(Y_n \geq x) = \lambda e^{-\lambda x}$$

Sea $T_0 = 0$ y dado $n \geq 1$,

$$T_n = Y_1 + \cdots + Y_n$$

Entonces, el proceso definido por:

$$N_t = n \text{ si } T_n \leq t < T_{n+1}$$

es un proceso de Poisson.

Ruido Blanco

Una serie de tiempo se dice de **ruido blanco**, si todas las X_t son independientes entre sí y son idénticamente distribuidas con media y varianza finita.

El movimiento browniano: Origen

En 1828, el botánico inglés Robert Brown observó que los granos de polen en suspensión se movían de forma irregular.

El movimiento browniano: Origen

En 1828, el botánico inglés Robert Brown observó que los granos de polen en suspensión se movían de forma irregular.

Más adelante se descubrió que esto se debe a los choques aleatorios de las partículas de polen con las moléculas del líquido.

El movimiento browniano: Origen

En 1828, el botánico inglés Robert Brown observó que los granos de polen en suspensión se movían de forma irregular.

Más adelante se descubrió que esto se debe a los choques aleatorios de las partículas de polen con las moléculas del líquido. En los años 20 del siglo XX,

Norbert Wiener presentó un modelo matemático para este movimiento como un proceso estocástico.

Movimiento browinano: Modelo como proceso estocástico

Un proceso estocástico $\{B_t \mid t \geq 0\}$ se denomina **movimiento browinano** si:

Movimiento browinano: Modelo como proceso estocástico

Un proceso estocástico $\{B_t \mid t \geq 0\}$ se denomina **movimiento browinano** si:

1. $B_0 = 0$.

Movimiento browinano: Modelo como proceso estocástico

Un proceso estocástico $\{B_t \mid t \geq 0\}$ se denomina **movimiento browinano** si:

1. $B_0 = 0$.
2. Dados n instantes fijos $0 \leq t_1 < \cdots < t_n$, los incrementos $B_{t_n} - B_{t_{n-1}}$, ..., $B_2 - B_{t_1}$, son variables aleatorias independientes.

Movimiento browinano: Modelo como proceso estocástico

Un proceso estocástico $\{B_t \mid t \geq 0\}$ se denomina **movimiento browinano** si:

1. $B_0 = 0$.
2. Dados n instantes fijos $0 \leq t_1 < \cdots < t_n$, los incrementos $B_{t_n} - B_{t_{n-1}}$, ..., $B_2 - B_{t_1}$, son variables aleatorias independientes.
3. Si $s < t$, el incremento $B_t - B_s$ se distribuye $\mathcal{N}(0, t-s)$

Operador de rezago

La **función de rezago** L se define como:

Operador de rezago

La **función de rezago** L se define como:

$$LX_t = X_{t-1}$$

Operador de rezago

La **función de rezago** L se define como:

$$LX_t = X_{t-1}$$

De esta manera, si $\tau > 0$,

Operador de rezago

La **función de rezago** L se define como:

$$LX_t = X_{t-1}$$

De esta manera, si $\tau > 0$,

$$L^\tau X_t = X_{t-\tau}$$

Operadores básicos sobre series de tiempo

Operador de diferencia

La **función de diferenciación** Δ se define como:

Operador de diferencia

La **función de diferenciación** Δ se define como:

$$\Delta X_t = X_t - X_{t-1}$$

Operador de diferencia

La **función de diferenciación** Δ se define como:

$$\Delta X_t = X_t - X_{t-1}$$

De esta manera, si $\tau > 1$,

Operador de diferencia

La **función de diferenciación** Δ se define como:

$$\Delta X_t = X_t - X_{t-1}$$

De esta manera, si $\tau > 1$,

$$\Delta^\tau X_t = \Delta^{\tau-1} X_t - \Delta^{\tau-1} X_{t-1}$$

Estacionaridad

Estacionaridad

Un proceso estocástico $\{X_t \mid t \in T\}$ se dice **estrictamente estacionario** si sus propiedades estadísticas o probabilísticas no cambian con el tiempo, esto es, si su función de distribución acumulativa es independiente del tiempo:

Estacionaridad

Un proceso estocástico $\{X_t \mid t \in T\}$ se dice **estrictamente estacionario** si sus propiedades estadísticas o probabilísticas no cambian con el tiempo, esto es, si su función de distribución acumulativa es independiente del tiempo:

$$F(X_{t_1}, X_{t_2}, \ldots, X_{t_n}) = F(X_{t_1+\tau}, X_{t_2+\tau}, \ldots, X_{t_n+\tau})$$

Estacionaridad

Un proceso estocástico $\{X_t \mid t \in T\}$ se dice **estrictamente estacionario** si sus propiedades estadísticas o probabilísticas no cambian con el tiempo, esto es, si su función de distribución acumulativa es independiente del tiempo:

$$F(X_{t_1}, X_{t_2}, \ldots, X_{t_n}) = F(X_{t_1+\tau}, X_{t_2+\tau}, \ldots, X_{t_n+\tau})$$

para todo $\tau \geq 0$

Estacionaridad

Un proceso estocástico $\{X_t \mid t \in T\}$ se dice **estrictamente estacionario** si sus propiedades estadísticas o probabilísticas no cambian con el tiempo, esto es, si su función de distribución acumulativa es independiente del tiempo:

$$F(X_{t_1}, X_{t_2}, \ldots, X_{t_n}) = F(X_{t_1+\tau}, X_{t_2+\tau}, \ldots, X_{t_n+\tau})$$

para todo $\tau \geq 0$ En otras palabras, un procesos estocástico es estrictamente estacionario si las distribuciones conjuntas de tuplas finitas de observaciones no dependen del tiempo.

Estacionaridad débil

La estacionaridad estricta es muy difícil de encontrar en la naturaleza.

Estacionaridad débil

La estacionaridad estricta es muy difícil de encontrar en la naturaleza.

Por lo anterior, lo que realmente se utiliza es el concepto de estacionaridad débil.

Estacionaridad débil

La estacionaridad estricta es muy difícil de encontrar en la naturaleza.

Por lo anterior, lo que realmente se utiliza es el concepto de estacionaridad débil. Un proceso estocástico $\{X_t \mid t \in T\}$ se dice **débilmente estacionario** si:

a) m_{X_t} es constante.

Estacionaridad débil

La estacionaridad estricta es muy difícil de encontrar en la naturaleza.

Por lo anterior, lo que realmente se utiliza es el concepto de estacionaridad débil. Un proceso estocástico $\{X_t \mid t \in T\}$ se dice **débilmente estacionario** si:

a) m_{X_t} es constante.
b) $\Gamma_x(t, t-\tau)$ depende sólo de τ.

Estacionaridad débil

La estacionaridad estricta es muy difícil de encontrar en la naturaleza.

Por lo anterior, lo que realmente se utiliza es el concepto de estacionaridad débil. Un proceso estocástico $\{X_t \mid t \in T\}$ se dice **débilmente estacionario** si:

a) m_{X_t} es constante.

b) $\Gamma_x(t, t-\tau)$ depende sólo de τ.
En este caso, se define la función:

Estacionaridad débil

La estacionaridad estricta es muy difícil de encontrar en la naturaleza.

Por lo anterior, lo que realmente se utiliza es el concepto de estacionaridad débil. Un proceso estocástico $\{X_t \mid t \in T\}$ se dice **débilmente estacionario** si:

a) m_{X_t} es constante.
b) $\Gamma_x(t, t-\tau)$ depende sólo de τ.
En este caso, se define la función:

$$R(\tau) = \Gamma_x(t, t-\tau)$$

Esta función se conoce como la **función de autocovarianza** de $\{X_t, \mid t \in T\}$.

Estacionaridad débil

La estacionaridad estricta es muy difícil de encontrar en la naturaleza.

Por lo anterior, lo que realmente se utiliza es el concepto de estacionaridad débil.
Un proceso estocástico $\{X_t \mid t \in T\}$ se dice **débilmente estacionario** si:

a) m_{X_t} es constante.

b) $\Gamma_x(t, t-\tau)$ depende sólo de τ.
En este caso, se define la función:

$$R(\tau) = \Gamma_x(t, t-\tau)$$

Esta función se conoce como la **función de autocovarianza** de $\{X_t, \mid t \in T\}$.

Aquí, la variable τ se conoce como **rezago**.

Autocorrelación de un proceso estocástico débilmente estacionario

Si $\{X_t \mid t \in T\}$ es un proceso débilmente estacionario, la función de autocorrelación se define como:

Autocorrelación de un proceso estocástico débilmente estacionario

Si $\{X_t \mid t \in T\}$ es un proceso débilmente estacionario, la función de autocorrelación se define como:

$$r(\tau) = \frac{R(\tau)}{R(0)}$$

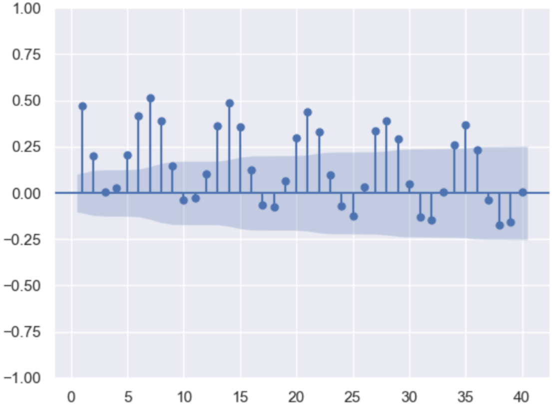
donde $R(\tau)$ es la función de autocovarianza.

Correlograma

La gráfica de esta función se denomina **correlograma**, y representa el grado de relación lineal entre un valor del proceso y sus valores en el pasado.

Correlograma

La gráfica de esta función se denomina **correlograma**, y representa el grado de relación lineal entre un valor del proceso y sus valores en el pasado.



Función de autocorrelación parcial

A partir de la función de autocorrelación, se define otra función similar que descuenta las relaciones lineales entre los valores anteriores y que pueden magnificarse al calcular la autocorrelación usual.

Función de autocorrelación parcial

A partir de la función de autocorrelación, se define otra función similar que descuenta las relaciones lineales entre los valores anteriores y que pueden magnificarse al calcular la autocorrelación usual.
Esta función $\alpha(\tau)$ se define de la siguiente manera:

Función de autocorrelación parcial

A partir de la función de autocorrelación, se define otra función similar que descuenta las relaciones lineales entre los valores anteriores y que pueden magnificarse al calcular la autocorrelación usual.
Esta función $\alpha(\tau)$ se define de la siguiente manera:

$$\alpha(1) = Corr(X_t, X_{t+1}))$$

y

$$\alpha(\tau) = Corr(x_{t+\tau} - P_{t,\tau}(x_{t+\tau}), x_t - P_{t,\tau}(x_t)),$$

Función de autocorrelación parcial

A partir de la función de autocorrelación, se define otra función similar que descuenta las relaciones lineales entre los valores anteriores y que pueden magnificarse al calcular la autocorrelación usual.
Esta función $\alpha(\tau)$ se define de la siguiente manera:

$$\alpha(1) = Corr(X_t, X_{t+1}))$$

y

$$\alpha(\tau) = Corr(x_{t+\tau} - P_{t,\tau}(x_{t+\tau}), x_t - P_{t,\tau}(x_t)),$$

donde $P_{t,\tau}(x)$ es la proyección de x sobre el espacio generado por X_{t+1}, ..., $X_{t+\tau-1}$

Función de autocorrelación parcial

A partir de la función de autocorrelación, se define otra función similar que descuenta las relaciones lineales entre los valores anteriores y que pueden magnificarse al calcular la autocorrelación usual.
Esta función $\alpha(\tau)$ se define de la siguiente manera:

$$\alpha(1) = Corr(X_t, X_{t+1}))$$

y

$$\alpha(\tau) = Corr(x_{t+\tau} - P_{t,\tau}(x_{t+\tau}), x_t - P_{t,\tau}(x_t)),$$

donde $P_{t,\tau}(x)$ es la proyección de x sobre el espacio generado por X_{t+1}, ..., $X_{t+\tau-1}$ Esta función se conoce como **Función de Autocorrelación Parcial**.

Prueba de estacionaridad de Dickey-Fuller

Existe una prueba de estacionaridad, debida a Dickey-Fuller.

Prueba de estacionaridad de Dickey-Fuller

Existe una prueba de estacionaridad, debida a Dickey-Fuller.
Aprenderemos a usarla en la práctica desde ahora, pero sus fundamentos teóricos los veremos cuando veamos los modelos autorregresivos.

Filtros y suavizados

Filtro de Hodrick-Prescott

Un **filtro** de una serie de tiempo, es un procedimiento para extraer una serie de tiempo S_t, a partir de otra X_t. La idea es que la serie S_t satisfaga algunas propiedades, en general de suavidad.

Filtro de Hodrick-Prescott

Un **filtro** de una serie de tiempo, es un procedimiento para extraer una serie de tiempo S_t, a partir de otra X_t. La idea es que la serie S_t satisfaga algunas propiedades, en general de suavidad.

El **filtro de Hodrick-Prescott** es una serie S_t, calculada a partir de una serie X_t de forma que se minimiza el siguiente valor:

Filtro de Hodrick-Prescott

Un **filtro** de una serie de tiempo, es un procedimiento para extraer una serie de tiempo S_t, a partir de otra X_t. La idea es que la serie S_t satisfaga algunas propiedades, en general de suavidad.

El **filtro de Hodrick-Prescott** es una serie S_t, calculada a partir de una serie X_t de forma que se minimiza el siguiente valor:

$$\sum_{t=1}^{n}(X_t - S_t)^2 + \lambda \sum_{t=1}^{n}((S_{t+1} - S_t) - (S_t - S_{t-1}))^2$$

Filtro de Hodrick-Prescott

Un **filtro** de una serie de tiempo, es un procedimiento para extraer una serie de tiempo S_t, a partir de otra X_t. La idea es que la serie S_t satisfaga algunas propiedades, en general de suavidad.

El **filtro de Hodrick-Prescott** es una serie S_t, calculada a partir de una serie X_t de forma que se minimiza el siguiente valor:

$$\sum_{t=1}^{n}(X_t - S_t)^2 + \lambda \sum_{t=1}^{n}((S_{t+1} - S_t) - (S_t - S_{t-1}))^2$$

donde $\lambda > 0$ es un parámetro que determina el grado de suavidad deseado.

Estacionalidad

Una serie de tiempo se dice **estacional** si tiene patrones que se repiten con una frecuencia fija conocida.

La descomposición ETS

Similar a la descomposición de Hodrick-Prescott, la **Descomposición ETS** (Error-Trend-Stationality) descompone una serie de tiempo en una componente de tendencia, una componente estacional y una componente de error.

La descomposición ETS

Similar a la descomposición de Hodrick-Prescott, la **Descomposición ETS** (Error-Trend-Stationality) descompone una serie de tiempo en una componente de tendencia, una componente estacional y una componente de error.

Si el modelo es aditivo, estas componentes se suman. Si el modelo es multiplicativo, las componentes se multiplican.

Media móvil simple

La **Media Móvil Simple** es el promedio de los anteriores k datos:

Media móvil simple

La **Media Móvil Simple** es el promedio de los anteriores k datos:

$$SMA_k(i) = \frac{1}{k} \sum_{j=i-k+1}^{i} y_j$$

Exponentially Weighted Moving Average (EWMA)

El SMA tiene algunas desventajas:

Exponentially Weighted Moving Average (EWMA)

El SMA tiene algunas desventajas:

- No alcanza los picos ni los valles de la señal

Exponentially Weighted Moving Average (EWMA)

El SMA tiene algunas desventajas:

- No alcanza los picos ni los valles de la señal
- No es muy bueno para hacer predicciones. Sólo describe los datos

Exponentially Weighted Moving Average (EWMA)

El SMA tiene algunas desventajas:

- No alcanza los picos ni los valles de la señal
- No es muy bueno para hacer predicciones. Sólo describe los datos
- Valores extremos lo afectan significativamente

Exponentially Weighted Moving Average (EWMA)

El SMA tiene algunas desventajas:

- No alcanza los picos ni los valles de la señal
- No es muy bueno para hacer predicciones. Sólo describe los datos
- Valores extremos lo afectan significativamente

El **Promedio Móvil Ponderado Exponencial** (EWMA por sus siglas en inglés) supera algunos de estos problemas.

Exponentially Weighted Moving Average (EWMA)

El SMA tiene algunas desventajas:

- No alcanza los picos ni los valles de la señal
- No es muy bueno para hacer predicciones. Sólo describe los datos
- Valores extremos lo afectan significativamente

El **Promedio Móvil Ponderado Exponencial** (EWMA por sus siglas en inglés) supera algunos de estos problemas. La fórmula del EWMA es:

Exponentially Weighted Moving Average (EWMA)

El SMA tiene algunas desventajas:

- No alcanza los picos ni los valles de la señal
- No es muy bueno para hacer predicciones. Sólo describe los datos
- Valores extremos lo afectan significativamente

El **Promedio Móvil Ponderado Exponencial** (EWMA por sus siglas en inglés) supera algunos de estos problemas. La fórmula del EWMA es:

$$y_t = \frac{\sum_{i=0}^{t} w_i x_{t-i}}{\sum_{i=0}^{t} w_i},$$

Exponentially Weighted Moving Average (EWMA)

El SMA tiene algunas desventajas:

- No alcanza los picos ni los valles de la señal
- No es muy bueno para hacer predicciones. Sólo describe los datos
- Valores extremos lo afectan significativamente

El **Promedio Móvil Ponderado Exponencial** (EWMA por sus siglas en inglés) supera algunos de estos problemas. La fórmula del EWMA es:

$$y_t = \frac{\sum_{i=0}^{t} w_i x_{t-i}}{\sum_{i=0}^{t} w_i},$$

donde $w_i = (1 - \alpha)^i$, para algún α.

Exponentially Weighted Moving Average (EWMA)

El SMA tiene algunas desventajas:

- No alcanza los picos ni los valles de la señal
- No es muy bueno para hacer predicciones. Sólo describe los datos
- Valores extremos lo afectan significativamente

El **Promedio Móvil Ponderado Exponencial** (EWMA por sus siglas en inglés) supera algunos de estos problemas. La fórmula del EWMA es:

$$y_t = \frac{\sum_{i=0}^{t} w_i x_{t-i}}{\sum_{i=0}^{t} w_i},$$

donde $w_i = (1-\alpha)^i$, para algún α.Este número se denomina **factor de suavizamiento**

Suavizado exponencial de Holt-Winters I

Similar al filtro de Hodrick-Prescott, el suavizado de Holt-Winters, busca una serie suave S_t, a partir de una serie dada X_t.

Suavizado exponencial de Holt-Winters I

Similar al filtro de Hodrick-Prescott, el suavizado de Holt-Winters, busca una serie suave S_t, a partir de una serie dada X_t.

Se elige un $0 < A < 1$ y se construye la siguiente serie:

Suavizado exponencial de Holt-Winters I

Similar al filtro de Hodrick-Prescott, el suavizado de Holt-Winters, busca una serie suave S_t, a partir de una serie dada X_t.

Se elige un $0 < A < 1$ y se construye la siguiente serie:

$$S_t = AX_t + (1-A)AX_{t-1} + (1-A)^2AX_{t-2} + (1-A)^3AX_{t-3} + \cdots$$

Suavizado exponencial de Holt-Winters I

Similar al filtro de Hodrick-Prescott, el suavizado de Holt-Winters, busca una serie suave S_t, a partir de una serie dada X_t.

Se elige un $0 < A < 1$ y se construye la siguiente serie:

$$S_t = AX_t + (1-A)AX_{t-1} + (1-A)^2 AX_{t-2} + (1-A)^3 AX_{t-3} + \cdots$$

Esto se puede expresar como:

Suavizado exponencial de Holt-Winters I

Similar al filtro de Hodrick-Prescott, el suavizado de Holt-Winters, busca una serie suave S_t, a partir de una serie dada X_t.

Se elige un $0 < A < 1$ y se construye la siguiente serie:

$$S_t = AX_t + (1-A)AX_{t-1} + (1-A)^2 AX_{t-2} + (1-A)^3 AX_{t-3} + \cdots$$

Esto se puede expresar como:

$$S_t = \big(1 + (1-A)L + (1-A)^2 L^2 + (1-A)^3 L^3 + \cdots\big) AX_t$$

Suavizado exponencial de Holt-Winters I

Similar al filtro de Hodrick-Prescott, el suavizado de Holt-Winters, busca una serie suave S_t, a partir de una serie dada X_t.

Se elige un $0 < A < 1$ y se construye la siguiente serie:

$$S_t = AX_t + (1 - A)AX_{t-1} + (1 - A)^2 AX_{t-2} + (1 - A)^3 AX_{t-3} + \cdots$$

Esto se puede expresar como:

$$S_t = \big(1 + (1 - A)L + (1 - A)^2 L^2 + (1 - A)^3 L^3 + \cdots\big) AX_t$$

Lo cual se puede llevar a que:

Suavizado exponencial de Holt-Winters I

Similar al filtro de Hodrick-Prescott, el suavizado de Holt-Winters, busca una serie suave S_t, a partir de una serie dada X_t.

Se elige un $0 < A < 1$ y se construye la siguiente serie:

$$S_t = AX_t + (1-A)AX_{t-1} + (1-A)^2 AX_{t-2} + (1-A)^3 AX_{t-3} + \cdots$$

Esto se puede expresar como:

$$S_t = \big(1 + (1-A)L + (1-A)^2 L^2 + (1-A)^3 L^3 + \cdots\big) AX_t$$

Lo cual se puede llevar a que:

$$S_t = \frac{AX_t}{1-(1-A)L}$$

Suavizado exponencial de Holt-Winters II

De lo anterior se obtiene:

Suavizado exponencial de Holt-Winters II

De lo anterior se obtiene:

$$S_t - (1 - A)S_{t-1} = AX_t$$

Suavizado exponencial de Holt-Winters II

De lo anterior se obtiene:

$$S_t - (1 - A)S_{t-1} = AX_t$$

de donde,

Suavizado exponencial de Holt-Winters II

De lo anterior se obtiene:

$$S_t - (1 - A)S_{t-1} = AX_t$$

de donde,

$$S_t = AX_t + (1 - A)S_{t-1}$$

Suavizado exponencial de Holt-Winters II

De lo anterior se obtiene:

$$S_t - (1 - A)S_{t-1} = AX_t$$

de donde,

$$S_t = AX_t + (1 - A)S_{t-1}$$

De aquí que el suavizado exponencial de Holt-Winters se calcula en cada etapa como un promedio ponderado entre X_t y S_{t-1}.

Suavizado exponencial de Holt-Winters II

De lo anterior se obtiene:

$$S_t - (1 - A)S_{t-1} = AX_t$$

de donde,

$$S_t = AX_t + (1 - A)S_{t-1}$$

De aquí que el suavizado exponencial de Holt-Winters se calcula en cada etapa como un promedio ponderado entre X_t y S_{t-1}.El valor de A determina si se le da más importancia al pasado o al presente.

Criterios de bondad de modelos

Función de verosimilitud

Una función de densidad de probabilidad f suele depender de un parámetro θ.

Función de verosimilitud

Una función de densidad de probabilidad f suele depender de un parámetro θ. Así, f se puede entender como una función que depende tanto de x como de θ. $f(x;\theta)$.

Función de verosimilitud

Una función de densidad de probabilidad f suele depender de un parámetro θ. Así, f se puede entender como una función que depende tanto de x como de θ. $f(x;\theta)$.

Cuando el parámetro θ es conocido o se considera fijo, entonces f pasa a entenderse como dependiente sólo de x y se f se entiende como una función de densidad de probabilidad.

Función de verosimilitud

Una función de densidad de probabilidad f suele depender de un parámetro θ. Así, f se puede entender como una función que depende tanto de x como de θ. $f(x;\theta)$.

Cuando el parámetro θ es conocido o se considera fijo, entonces f pasa a entenderse como dependiente sólo de x y se f se entiende como una función de densidad de probabilidad.

Ahora bien, cuando se desconoce el parámetro θ, pero se tiene observación a, la función $f(a;\theta)$ pasa a considerarse con variable θ y se denomina **función de verosimilitud**, e indica el grado de "credibilidad"de un valor posible de un parámetro, dada la observación a.

Función de verosimilitud

Una función de densidad de probabilidad f suele depender de un parámetro θ. Así, f se puede entender como una función que depende tanto de x como de θ. $f(x;\theta)$.

Cuando el parámetro θ es conocido o se considera fijo, entonces f pasa a entenderse como dependiente sólo de x y se f se entiende como una función de densidad de probabilidad.

Ahora bien, cuando se desconoce el parámetro θ, pero se tiene observación a, la función $f(a;\theta)$ pasa a considerarse con variable θ y se denomina **función de verosimilitud**, e indica el grado de "credibilidad"de un valor posible de un parámetro, dada la observación a. Así, mientras más alta la verosimilitud, mayor es la confianza en que la distribución f, junto con el valor de θ modela la observación.

Función de verosimilitud

Una función de densidad de probabilidad f suele depender de un parámetro θ. Así, f se puede entender como una función que depende tanto de x como de θ. $f(x;\theta)$.

Cuando el parámetro θ es conocido o se considera fijo, entonces f pasa a entenderse como dependiente sólo de x y se f se entiende como una función de densidad de probabilidad.

Ahora bien, cuando se desconoce el parámetro θ, pero se tiene observación a, la función $f(a;\theta)$ pasa a considerarse con variable θ y se denomina **función de verosimilitud**, e indica el grado de "credibilidad"de un valor posible de un parámetro, dada la observación a. Así, mientras más alta la verosimilitud, mayor es la confianza en que la distribución f, junto con el valor de θ modela la observación.

Más aún, si se tienen n observaciones experimentales $a_1, a_2, \ldots, a_n$, para n experimentos independientes distribuidos según f, la función de verosimilitud asociada a estos valores es:

Función de verosimilitud

Una función de densidad de probabilidad f suele depender de un parámetro θ. Así, f se puede entender como una función que depende tanto de x como de θ. $f(x;\theta)$.

Cuando el parámetro θ es conocido o se considera fijo, entonces f pasa a entenderse como dependiente sólo de x y se f se entiende como una función de densidad de probabilidad.

Ahora bien, cuando se desconoce el parámetro θ, pero se tiene observación a, la función $f(a;\theta)$ pasa a considerarse con variable θ y se denomina **función de verosimilitud**, e indica el grado de "credibilidad"de un valor posible de un parámetro, dada la observación a. Así, mientras más alta la verosimilitud, mayor es la confianza en que la distribución f, junto con el valor de θ modela la observación.

Más aún, si se tienen n observaciones experimentales $a_1, a_2, \ldots, a_n$, para n experimentos independientes distribuidos según f, la función de verosimilitud asociada a estos valores es:

$$L(\theta) = f(\theta|X_1 = a_1, X_2 = a_2, \ldots, X_n = a_n) = \Pi_{i=1}^{n} f(a_i;\theta)$$

Función loglikelihood

La función de verosimilitud $L(\theta)$ suele ser dificil de calcular, así que en su lugar suele usarse el logaritmo natural de la misma, la cual cumple el mismo criterio de, a mayor valor, mejor ajuste del modelo representado en la función f.

Función loglikelihood

La función de verosimilitud $L(\theta)$ suele ser dificil de calcular, así que en su lugar suele usarse el logaritmo natural de la misma, la cual cumple el mismo criterio de, a mayor valor, mejor ajuste del modelo representado en la función f.

$$\ell(\theta) = \ln L(\theta)$$

Función loglikelihood

La función de verosimilitud $L(\theta)$ suele ser dificil de calcular, así que en su lugar suele usarse el logaritmo natural de la misma, la cual cumple el mismo criterio de, a mayor valor, mejor ajuste del modelo representado en la función f.

$$\ell(\theta) = \ln L(\theta)$$

Esta función se denomina función ℓ o **loglikelihood function**.

Criterios de información

A partir de la función l, se contruyen los denominados **criterios de información**.

Criterios de información

A partir de la función l, se contruyen los denominados **criterios de información**.

A diferencia de ℓ, para los criterios de información, a mejor modelo, menor valor de dicho criterio.

Criterios de información

A partir de la función l, se contruyen los denominados **criterios de información**.

A diferencia de ℓ, para los criterios de información, a mejor modelo, menor valor de dicho criterio.

Los criterios de información más conocidos y utilizados son:

Criterios de información

A partir de la función l, se contruyen los denominados **criterios de información**.

A diferencia de ℓ, para los criterios de información, a mejor modelo, menor valor de dicho criterio.

Los criterios de información más conocidos y utilizados son:

- Criterio de información de Akaike:

Criterios de información

A partir de la función l, se contruyen los denominados **criterios de información**.

A diferencia de ℓ, para los criterios de información, a mejor modelo, menor valor de dicho criterio.

Los criterios de información más conocidos y utilizados son:

- Criterio de información de Akaike:

$$AIC = 2k - 2\ell(\theta)$$

Criterios de información

A partir de la función l, se contruyen los denominados **criterios de información**.

A diferencia de ℓ, para los criterios de información, a mejor modelo, menor valor de dicho criterio.

Los criterios de información más conocidos y utilizados son:

- Criterio de información de Akaike:

$$AIC = 2k - 2\ell(\theta)$$

- Criterio de información de Bayes:

Criterios de información

A partir de la función l, se contruyen los denominados **criterios de información**.

A diferencia de ℓ, para los criterios de información, a mejor modelo, menor valor de dicho criterio.

Los criterios de información más conocidos y utilizados son:

- Criterio de información de Akaike:

$$AIC = 2k - 2\ell(\theta)$$

- Criterio de información de Bayes:

$$BIC = k \ln n - 2\ell(\theta)$$

Criterios de información

A partir de la función l, se contruyen los denominados **criterios de información**.

A diferencia de ℓ, para los criterios de información, a mejor modelo, menor valor de dicho criterio.

Los criterios de información más conocidos y utilizados son:

- Criterio de información de Akaike:

$$AIC = 2k - 2\ell(\theta)$$

- Criterio de información de Bayes:

$$BIC = k \ln n - 2\ell(\theta)$$

Aquí, k es el número de parámetros del modelo y n es el número de datos.