

apuntes predicción

Valentina Valdez Vega

2025-04-27

Modelos de predicción de series de tiempo

Una serie de tiempo se define como:

$$Y_t \in \mathbb{R}^m$$

Si:

- $m = 1$ es una serie de tiempo univariante
- $m > 1$ es una serie de tiempo multivariante

El número de *periodos observados* por año se llama *frecuencia* (s).

- Anual: s=1
- Semestral: s=2
- Trimestral: s=4
- Mensual: s=12
- Diaria: s=365

Diferencias y tasas de crecimiento

La primera diferencia de una serie se define como:

$$\Delta Y_t = Y_t - Y_{t-1}$$

$$\Delta^2 Y_t = \Delta Y_t - \Delta Y_{t-1}$$

$$\Delta_s Y_t = Y_t - Y_{t-s}$$

Tasa de crecimiento de un periodo (t, t-1) en porcentaje se define como:

$$Q_t = 100 \left(\frac{\Delta Y_t}{Y_{t-1}} \right) = 100 \left(\frac{Y_t}{Y_{t-1}} - 1 \right)$$

Cuando la frecuencia es mayor a 1 (s>1) se recomienda:

$$G_t = 100 \left(\frac{\Delta_s Y_t}{Y_{t-s}} \right) = 100 \left(\frac{Y_t}{Y_{t-s}} - 1 \right)$$

$$Q_t \simeq 100 \Delta \log Y_t$$

Estacionariedad

Propiedad estadística clave en la que las características de la serie no cambian a lo largo del tiempo.

$$Y_t \sim (.)$$

- Media constante:

$$E[Y_t] = \mu$$

- Varianza constante:

$$V(Y_t) = \sigma^2$$

- Autocovarianza es independiente del tiempo:

$$\text{cov}(Y_t, Y_{t-k}) = \gamma(k)$$

Ergodicidad

Las características de una serie de tiempo pueden ser inferidas a partir de una única realización de la serie en el tiempo, sin necesidad de observar múltiples realizaciones independientes.

Una serie es ergódica si:

- Es estacionaria
- Asintóticamente independientes

$$\lim_{k \rightarrow \infty} \text{Corr}(Y_t, Y_{t+k}) = 0$$

Aplicación en R

API Banco Mundial

- *wbstats*
- WDI

Definiendo una serie de tiempo en R

Verificando estacionariedad

Test de Dickey-Fuller aumentada

- H0: Los datos *no son estacionarios*
- H1: Los datos son estacionarios

Función de autocorrelación

Modelos de suavizado exponencial

$$Y_t = f(Nivel, Tendencia, Estacionalidad, Ruido)$$

- Modelo ETS en R
 - Error: Aditivo (A), multiplicativo (M)
 - Tendencia: Sin tendencia (N), aditivo (Ad), multiplicativo (M)
 - Estacional: Ninguna (N), aditivo (Ad), multiplicativo (M)

Adecuación del Dataset y Preprocesamiento

Preprocesar y transformar los datos para que puedan ser utilizados en el análisis de series de tiempo. Esto implica convertir fechas, ordenar registros y estructurar el dataset en un formato temporal.

La transformación de datos incluye:

- Adecuación del dataset mediante un pivot
- Conversión de la columna de fecha al formato `Date`.
- Transformación a un objeto de series de tiempo

Ejemplo de pivot: RECAUDACIÓN RENTA INTERNA POR TIPO DE IMPUESTOS (UDAPE) ANUAL

Ejemplo de pivot: Exportaciones según Actividad Económica y Producto por Año y Mes, 1992 - 2024

Modelo ARIMA

Es un modelo univariante que permite trabajar con 2 componetes; parte autoregresiva (AR) y medias móviles (MA).

Modelo AR(p)

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \epsilon_t$$
$$\epsilon_t = y_t - \sum_{i=1}^p \phi_i y_{t-i}$$

Donde ϕ corresponde a los parámetros a estimar, ϵ_t el ruido blanco, con:

$$E[\epsilon_t] = 0 \quad V(\epsilon_t) = \sigma^2 \quad Cov(\epsilon_t, \epsilon_{t-h}) = 0$$

Por último, pero no estricto:

$$\epsilon_t \sim N(0, \sigma^2)$$

p es el número de términos autoregresivos (AR) que se incluirán en el modelo. Es decir, cuántos rezagos de la serie se utilizan para predecir el valor actual de la serie. Para determinar el valor de p :

- Análisis de la ACF (Autocorrelation Function): El número de rezagos significativos en la ACF.
- Criterios de información: AIC (Akaike Information Criterion) o BIC (Bayesian Information Criterion) son utilizados para seleccionar el valor de p que minimiza la función de pérdida.
- Criterio de la parcimonia: El valor de p debe ser el menor posible que aún capture la dinámica de la serie temporal. Es decir, agregar más rezagos no debe mejorar significativamente el ajuste.

Modelo MA(q)

$$y_t = \epsilon_t + \sum_{j=1}^q \theta_j \epsilon_{t-j}$$

$$\epsilon_t = y_t - \sum_{j=1}^q \theta_j \epsilon_{t-j}$$

q es el número de términos de media móvil (MA) que se incluirán en el modelo. Es decir, cuántos rezagos de los errores (residuos) se utilizarán para modelar la serie temporal. La elección de q se hace:

- Análisis de la PACF (Partial Autocorrelation Function): Muestra la autocorrelación de la serie temporal después de eliminar el efecto de los rezagos intermedios.
- Criterios de información: AIC o BIC

Modelo ARIMA (p,d,q)

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \epsilon_t + \sum_{j=1}^q \theta_j \epsilon_{t-j}$$

$$\epsilon_t = y_t - \sum_{i=1}^p \phi_i y_{t-i} - \sum_{j=1}^q \theta_j \epsilon_{t-j}$$

$$E(\epsilon_t) = 0 \quad Var(\epsilon_t) = \sigma^2 \quad Cov(\epsilon_t, \epsilon_{t-h}) = 0 \quad \text{para todo } h \neq 0$$

La d en un modelo ARIMA(p, d, q) representa el grado de diferenciación de la serie temporal, que se utiliza para hacerla estacionaria.

Notas:

- El modelo ARIMA(p,d,q) se usa para series sin componentes estacional
- El modelo ARIMA(p,d,q)SARIMA(P,D,Q) se usa para series con componente estacional

Modelo ARIMAX

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \epsilon_t + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \sum_{k=1}^K \beta_k X_{t-k}$$

Nota: Para las predicciones del modelo ARIMAX se requiere predicciones de los regresores X .

Aplicación en R